Large Deviations of Generalised Jackson Networks

vorgelegt von Diplom Wirtschaftsmathematikerin
Silke Meiner
aus Berlin

Von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften
Dr. rer. nat.

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Reinahrd Nabben
Gutachter: Prof. Dr. Jean-Dominique Deuschel
Gutachter: Prof. Adam Shwartz, Ph.D.

Tag der wissenschaftlichen Aussprache: 3. September 2008

Berlin, 2008
D 83

This research has been supported in parts by the Minerva Foundation of the Max Planck Society and the Berliner Programm zur Förderung der Chancengleichheit von Frauen in Forschung und Lehre.

Abstract

In dieser Arbeit entwickeln wir die lokalen großen Abweichungen von verallgemeinerten Jackson Netzwerken. Im Unterschied zum Jackson Netzwerk sind Zwischenankunfts- und Servicezeiten allgemeinen Verteilungen unterworfen und nicht auf Exponentialverteilungen beschränkt. Die daraus resultierenden stochastischen Prozesse sind nicht Markovsch, was eine Herausforderung an die zur Verfügung stehende mathematische Technik bedeutet.

Im ersten Teil der Arbeit untersuchen wir, inwieweit und mit welchen Mitteln die verlorene Markoveigenschaft aufgewogen werden kann. Die verallgemeinerten Prozesse, die wir betrachten, sind Erneuerungsprozesse. Es gelingt uns, die Prozesse, mit denen wir das generalisierte Jackson Netzwerk beschreiben werden, so abzuändern, dass sie unabhängige stationäre Inkremente haben und im Sinne der großen Abweichungen nicht von den ursprünglichen Prozessen zu unterscheiden sind. Weiter entwickeln wir einen exponentiellen Maßwechsel für die Erneuerungsprozesse, so dass die Erneuerungseigenschaft erhalten bleibt. Der resultierende Maßwechsel für den Netzwerkprozess verändert nur die Raten des Netzwerkes, nicht aber seine grundlegenden Eigenschaften.

Im Ergebnis erhalten wir ein lokales Prinzip großer Abweichungen mit einer Ratenfunktion, die fast die Fenchel Legendre Transformierte der logarithmischen Momenterzeugendenfunktion Ψ des freien Prozesses ist, der dem generalisierten Jackson Netzwerk zugeordnet ist:

$$L(x,v) = \sup_{\alpha \in \mathcal{B}_{K(x,v)}} \langle \alpha, v \rangle - \Psi(\alpha) \qquad (1)$$

Die lokale Ratenfunktion $L(\cdot,\cdot)$ unterscheidet sich von einer Fenchel Legendre Transformierten durch die Einschränkung auf Elemente aus $\mathcal{B}_{K(x,v)}$. Diese Menge beschreibt die unterschiedlichen Verhaltensweisen des Netzwerkprozesses in Abhängigkeit vom derzeitigen Zustand des Netzwerkes - repräsentiert durch x - und dem zukünftigen Verlauf - repräsentiert durch v. Ist eine zukünftige Entwicklung des Netzwerkes in Richtung v ein seltenes Ereignis und α der Optimierer in (1), so ändert sich die Situation unter dem Maßwechsel mit Parameter α dahingehend, dass die Entwicklung in Richtung v zum erwarteten Verhalten des Netzwerkes wird.

Contents

Notation

$\|\| \cdot \|\|$	for $f \in \mathbb{R}^{[0,T]}$: $\|\|f\|\| = \sup_{t\in[0,T]} \|f(t)\|$ for $f \in (\mathbb{R}^d)^{[0,T]}$: $\|\|f\|\| = \max_{i=1,\dots,d} \sup_{t\in[0,T]} \|f_i(t)\|$
$\|\| \cdot \|\|_{[a,b]}$	$\|\|f\|\|_{[a,b]} = \|\|f\, 1\!\!1_{[a,b]}\|\|$, $[a,b] \subseteq [0,T]$
AC, $AC([0,T],\mathbb{R}^d)$	$\{f \in (\mathbb{R}^d)^{[0,T]} \mid f$ is absolutely continuous $\}$
$\mathcal{B}_\Lambda$, $\mathcal{B}_K$	set of restrictions, claims 6.1.1, 6.1.4
$C([0,T],\mathbb{R}^d)$ $D([0,T],\mathbb{R}^d)$	$\{f \in (\mathbb{R}^d)^{[0,T]} \mid f$ is continuous at each $t \in [0,T]\}$ $\{f \in (\mathbb{R}^d)^{[0,T]} \mid f$ with each f_i right-continuous at each $t \in [0,T)$ and with left limits for each $t \in (0,T]\}$
d	number of nodes of a network / graph
$\mathbb{E}^{(\alpha)}$, $\mathbb{E}^{[\alpha]}$	expectation wrt a twisted distribution, (α): inter event time twisted with parameter α; $[\alpha]$ counting process twisted with parameter α, section 3.6
F_β F_{+a}	exponential tranfform / twist, def 2.3.1 distribution function of $\tau - a$ conditional on $\tau > a$, def 2.1.7
$g_i(\cdot)$ $g_\gamma(\cdot)$	restriction, above claim 6.3.2 def 6.2.2
$K(\cdot)$ $K(\cdot)$	lmgf of splitting probabilities, def 4.4.4 set of nodes / indices, claim 6.1.1

$\mathcal{L}(X)$	law / distribution of a random variable X
$\mathcal{L}(\alpha)$, $\mathcal{L}(\alpha, k)$	a level set, the level set of a function identified through k
$\Lambda(\cdot)$	lmgf of an inter event time, def 2.2.1
$\Lambda(\cdot)$	set of nodes / indices, proof of 5.2.15, claim 6.1.1
$L(\cdot, \cdot)$	local rate function
$\lambda \in \mathbb{R}^d$, λ_i	vector of arrival rates, arrival rate at node i
λ_M	vector of arrival rates to nodes $i \in M$, claim 5.2.10
$\lambda^\Lambda \in \mathbb{R}^{\lvert\Lambda^c\rvert}$	arrival rates to Λ^c-nodes in the subnetwork of Λ^c nodes with Λ-nodes free, proof of 5.2.15
μ, μ_i	(vector of) service rate(s), same indexing as with λ
N	counting process, def 3.1.1
$N^\sigma, N^{\tilde{\tau}}$	cp with indicated initial inter event time, def 3.1.5
$\check{N}$	counting process as a memeber of a coupling
$\hat{N}$	interpolated counting process
$N^{\mathrm{re},s}, N^{\mathrm{re},(s_1,\ldots,s_k)}$	restarted counting process, s (or $s_1, \ldots, s_k$) indicating the time(s) of restarting, defs 3.4.11, 3.4.16
N^{sp}	a split counting proces, def 4.4.1
P , $p^{(i)}$	routing matrix, row of the routing matrix P
π_k	projection onto span $\{e_k\}$
π_M, M a set	projection onto span $\{e_k \mid k \in M\}$
$\Psi(\cdot)$	lmgf of the free process, def 5.3.9
$\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$	$\mathbb{R} \cap (0, \infty)$, $\mathbb{R} \cap [0, \infty)$
$R^{(i)}$	runtime process of the server at node i
$\mathbf{R}(\cdot, \cdot)$	operator to describe / define a runtime process, def 5.3.21
$\mathbb{T}$, $\mathbb{T}^{(i)}$	linear transformation, def 4.4.7
$A^\top$	transpose of a matrix A
T	often a time-index $\in \mathbb{R}_{>0}$

Chapter 1

Introduction

We develop local large deviations for the generalised Jackson network. We work with a continuous time model and with light tail distributions for inter arrival and service times. Using classical large deviation theory of logarithmic moment generating functions and exponential changes of measure we get a local rate function that is almost a Fenchel Legendre transform of the free process' logarithmic moment generating function Ψ.

$$L(x,v) = \sup_{\alpha \in \mathcal{B}_{K(x,v)}} \langle \alpha, v \rangle - \Psi(\alpha) \tag{1.1}$$

What keeps the local rate function from being a full Fenchel Legendre transform is the restrictions $\mathcal{B}_{K(x,v)}$ reflecting nodes not-empty when the state of the network is x and nodes filling up when the network evolves in direction v.

The way we develop the local large deviation will allow to get a weak and full large deviation principle for the generalised Jackson network quite easily. We also give a representation of the almost Fenchel Legendre transform as a Fenchel Legendre transform in lower dimension.

The approach to apply classical large deviation theory to stochastic networks is inspired by "Large Deviations of Jackson Networks" of Irina Ignatiouk-Robert published in 2000 [12].

1.1 Queues, networks, and rare events

A queueing network is a collection of inter connected service stations that customers arrive to, travel through and leave. At each service station a customer occupies the server in order to receive service. Whenever a customer arriving at a server finds it busy serving another customer, the arriving customer will queue.

We like to think about networks in a stochastic way: the time a customer occupies a server may vary from customer to customer and from server to server; the travelling through the network may be along different possible routes and a customer leaving a node may choose between different nodes to go to next. We perceive these decisions as random.

Assuming that the network has resources enough to finish servicing customers at each station in a reasonable time we are interested in the rare events of long queue sizes at some nodes and in probabilities of different evolutions of large queue sizes over time.

In a network with sufficient resources large queue sizes will occur only rarely - but they will. A manager of a network will have to find the balance between increasing resources and the tolerance to rarely occurring large queue sizes and when necessary come up with actions to reduce large queue sizes quickly. The present thesis gives a guideline of how often to expect any kind of large queue sizes for given resources in a network under its regular operating conditions.

Before turning to the main object of interest of this thesis which are networks of queues, let us look at a single network node in isolation. The simplest setting is the single server queue with Poisson arrivals and exponential service times, the so called M/M/1 queue. A generalised M/M/1 would be denoted a GI/GI/1 queue where the stream of arriving customers is a renewal counting process and service times are independent with a general distribution (M is for Markovian and the Poisson process is Markovian, GI is for General Independent).

To get a first impression of typical results consider a queue that generally can serve arriving customers without the queue size becoming too large. Provided that the queue has been running for a long time, the probability for a M/M/1 queue to be of size $x \in \mathbb{N}$ or larger is

$$\mathbb{P}(Q \geq x) = \rho^x \tag{1.2}$$

where $\rho < 1$ is the traffic intensity, the ratio of arrival rate and service rate. For the GI/GI/1 queue we can make the following approximation for the probability of an unusually large size of at least nx: For $x \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Q \geq nx) = -\delta x \tag{1.3}$$

for some $\delta > 0$ (for example [6] in 1994). This can be equivalently expressed as: For any small $\epsilon > 0$ and n large enough

$$e^{-n(\delta x - \epsilon)} \leq \mathbb{P}(Q \geq nx) \leq e^{-n(\delta x + \epsilon)} \tag{1.4}$$

The approach via steady state probabilities works well if the network has been running for a long time under stable conditions and is usually followed in situations where we have no information other than that. If however we are able to monitor the system and observe its state x_0 at time $t = 0$ say, then we might want to know how the queue size evolves from now on. In a stable system we will most likely observe the queue size decreasing but even so rarely other evolutions may occur too. In our approximative setting we do not observe the queue size exactly but for some n the smoothed version Q_n as a function of time: $Q_n(t) = \frac{1}{n} Q_{nt}$. We are then interested in following probability

$$\mathbb{P}(||Q_n - \psi|| < \epsilon \,|\, Q_n(0) = x_0). \tag{1.5}$$

with $||\cdot||$ denoting the supremum norm over the interval of interest.

This and more general questions can be answered with a large deviation result by Anatolii Puhalskii [15] from 1995: under some technical conditions for a set of functions A

$$\begin{aligned} -\inf_{\phi\in A^\circ} I(\phi) \quad \leq \quad & \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Q_n \in A \,|\, Q_n(0) = x_0) \\ & \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Q_n \in A \,|\, Q_n(0) = x_0) \;\leq\; -\inf_{\phi\in \overline{A}} I(\phi) \end{aligned} \tag{1.6}$$

Here the so called rate function $I(\phi)$ at ϕ is either infinite or can be calculated from an explicitly known local rate function L by

$$I(\phi) = \int_{s=0}^{t_1} L(\phi(s), \phi'(s))\, ds. \tag{1.7}$$

The rate function $I(\cdot)$ allows to generalise the fixed rate δ of (1.3) to bound the steady state probabilities: We get a lower rate $\inf_{\phi\in A^\circ} I(\phi)$ and an upper rate $\inf_{\phi\in\overline{A}} I(\phi)$ that naturally depend on the event A in which we are interested in.

In particular the choice $A = \{\phi \,|\, ||\phi - \psi|| < \epsilon\}$ is feasible and we can approximate (1.5) by applying (1.6). We rephrase (1.6) in terms of small $\epsilon > 0$ and large n:

$$e^{-n(\inf_{\phi\in A^\circ} I(\phi)-\epsilon)} \leq \mathbb{P}(||Q_n - \psi|| < \epsilon \,|\, Q_n(0) = x_0) \leq e^{-n(\inf_{\phi\in \overline{A}} I(\phi)+\epsilon)} \quad (1.8)$$

So far we have stated results for the single queue only. We now proceed to networks of queues and state similar results. We are interested in approximating probabilities for specified rarely occurring behaviour of queue sizes, now jointly for the queues at each node of the network. We start with steady state probabilities for queue sizes at a fixed time and then move to probabilities over intervals of time and with fixed starting positions.

Consider a network of d M/M/1 queues and now let $Q \in \mathbb{N}^d$ be the vector of queue sizes at some fixed time when assuming that the system has been running for a long time. For $d = 1$ we are back with the isolated queue.

This network of M/M/1 queues, the so called Jackson network, was introduced by James R. Jackson in 1957 [14] to model machine shops with goods to be processed travelling the network of machines. As a machine is busy processing, newly arriving goods have to wait to be processed later and in the mean time queue. The Jackson network has also been applied to documents travelling a network of offices of a business or administration where they are processed by clerks [25] ; Q is then the height of the stack of documents piling on each desk. More recently Jackson networks are applied in the analysis and design of computer networks where there are now data packages being routed through a local area network or some larger network like the internet. Q is then the number of packages in the buffer of each routing device.

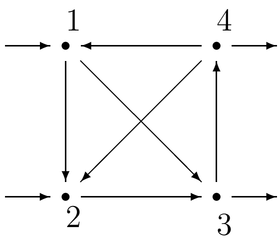


Figure 1.1: A network with $d = 4$ nodes

As an example consider the network of $d = 4$ nodes in figure 1.1. Let

- $\lambda \in \mathbb{R}^4$ be non-negative with coordinates $\lambda_1, \lambda_2 > 0$ for the rate at which customers arrive at nodes 1 and 2. There are no customers arriving at nodes 3 and 4, so $\lambda_3 = \lambda_4 = 0$;

- $\mu \in \mathbb{R}^4$ be positive with μ_i the rate at which a server releases customers while busy;

- $P \in \mathbb{R}^{4,4}$ be a routing matrix: the i-th row of P gives probabilities of which node to go to next and of leaving the network. In the example network leaving the network is possible only after finishing service at nodes 3 and 4.

The result of Jackson published in [14] is the following: If there is a unique ν solving the traffic equation

$$\nu = \lambda + P^\top \nu \tag{1.9}$$

and $\nu_i < \mu_i$ for all nodes $i = 1, \ldots, 4$ then the steady state probability for the queue sizes to be of size $x \in \mathbb{N}^4$ is

$$\mathbb{P}(Q = x) = \prod_{i=1}^{4} (1 - \frac{\nu_i}{\mu_i}) \left(\frac{\nu_i}{\mu_i}\right)^{x_i}. \tag{1.10}$$

Note that $\nu_i < \mu_i$ makes precise the sufficient resources (all μ_i large enough) in the network to generally allow to serve all customers in a reasonable time.

Applications of Jackson's results are manifold: it helps us to decide about the design of a network system where the probability of queue sizes to exceed some $x_{\max}$ has to be below some fixed level. Similarly, in a network with given service-resources we can now decide how to invest additional resources to get a maximum reduction of the probability to exceed $x_{\max}$. To a network service provider probabilities of large queue sizes and the empirical occurrence of large queue sizes are a measure for the quality of the provided service, and presumably for customer satisfaction.

In computer networks conditions will vary throughout the day and a network service provider will observe and thus fix the present state of the network and ask about probabilities for future evolution starting from the present state. For this purpose steady state probabilities are not enough and we turn to sample path large deviations. As before the large deviations concern the scaled, smoothed process Q_n that for t in some interval is defined as $Q_n(t) = \frac{1}{n} Q(nt) \in \mathbb{R}^d$.

1.2 Large deviations of Jackson and generalised Jackson networks

Large deviations for a wide class of Markovian models including the Jackson network have in principle been obtained by Paul Dupius and Richard S. Ellis in 1995 [7]. The analysis of stochastic networks has been perceived as difficult due to discontinuity of their behaviour as queue sizes change from empty to full and back.

A result of Irina Ignatiouk-Robert of 2000 [12] gives the explicit form of the rate function for the Jackson network. We can make the same kind of bounding as in (1.6) and (1.8) now with A a set of d-dimensional functions and the rate function $I(\cdot)$ again in integral form over a local rate function $L(\cdot,\cdot)$:

$$\begin{aligned} I(\phi) &= \int_{s=0}^{T} L(\phi(s), \phi'(s))\, ds \\ L(x,v) &= l^{\Lambda(x)}(v) = \sup_{\alpha \in \mathcal{B}_{\Lambda(x)}} \langle \alpha, v\rangle - R(\alpha) \end{aligned} \tag{1.11}$$

for some explicitly known R and set of restrictions $\mathcal{B}_{\Lambda(x)}$. It is quite remarkable that given the complexity of a network the result comes in such handy format.

The next step is of course to generalise the sample path large deviations from networks of M/M/1 queues to networks of GI/GI/1 queues. This is what we do in this thesis.

During the work on this thesis, in 2007, Anatolii Puhalskii [16] has proved existence of a large deviation principle for the generalised Jackson network. He gives a rate function in integral form with the local rate function a high-dimensional convex optimisation problem. Complementing this we give the explicit representation of the local rate function.

Generally our approach is very different from that of Puhalskii. It is closer to ideas found in the work of Ignatiouk-Robert [12] and highlights the classical large deviation theory of finding the exponential change of measure that turns the deviating behaviour into regular behaviour.

As a generalised Jackson network is not a Markov process we cannot profit from the rich theory developed around them in the last 100 years and we have to develop our own tools. The thesis is structured as follows:

- **Chapter 2** is on inter event times. Inter event times in a queueing network are times between arrivals of customers or the service times of customers at a node. We introduce inter event times and make assumptions on their distributions. Further, we introduce the hazard rate function. From inter event times we build

- renewal counting processes in **chapter 3**. In the network setting an arrival counting process will count the number of arrivals at a node over an interval of time. We generally work with non-Markovian counting processes and we prove in this section that some implications of the Markov property can still be obtained for these non-Markovian counting processes. This will be done through exponential equivalence.

  We then introduce a change of measure for the renewal counting process such that under the changed measure the process stays a renewal counting process. Then of course we need to know about the

- large deviations of the renewal counting process: We develop them in **chapter 4**. We start with local large deviations applying the change of measure developed in the previous chapter. Building on the local large deviations we prove weak large deviations and strengthen these to a full large deviation principle.

  The large deviations of renewal counting processes are on the one hand required to develop the local large deviations of the generalised Jackson network. On the other hand we think that in a similar and relatively easy way this allows to strengthen the local large deviations of the generalised Jackson network to a full large deviation principle for the generalised Jackson network.

- **Chapter 5** introduces stochastic networks and stochastic processes that describe (aspects of) such networks: the free, the network, and the local process.

  We first define drifts for networks based on deterministic rates for the network starting empty and the network starting with some nodes initially non empty, and we give formulations of network drifts in terms of the solution to the linear complementary problem and the Skorohod problem. When introducing the stochastic processes we show that these initially defined drifts are drifts of these processes, and thus describe the regular behaviour of theses processes.

We then investigate rare events and develop sample path large deviations for the free process.

- In **chapter 6** we prove the local large deviations for the generalised Jackson network. We follow a classical approach here that uses an exponential change of measure (developed over chapters 3, 4, 5) that gives the upper bound as the almost Fenchel Legendre transform that will be the local rate function. The optimiser of the almost Fenchel Legendre transform corresponds to a change of measure which is then applied to obtain the lower bound.

  We close with identifying the rate function of Puhalskii with our local large deviation rate function and with an example of how to calculate the local rate function.

Also there are further applications of these sample path large deviations by the contraction principle that allow approximating probabilities for events that continuously depend on the queue sizes Q in [9], [23] that require an analytical form of the rate function.

As Jackson networks are Markov processes and generalised Jackson networks are not our technique has to be fundamentally different from that in [7] and [12]. In terms of results the difference of (1.1) and (1.11) is small:

- Ψ versus R: Both are logarithmic moment generating functions of the associated free process and $\Psi = R$ if the generalised Jackson network is a Jackson network.

- optimising over $\mathcal{B}_{\Lambda(x)}$ versus $\mathcal{B}_{K(x,v)}$: For an absolutely continuous ϕ we have $\Lambda(\phi(t)) = K(\phi(t), \phi'(t))$ for almost all t.

We cite related work we are aware of at the beginning of chapters and give reference to alternative proofs in the main text.

Chapter 2

Inter event times

Looking at a stochastic network times between arrivals of customers at a node, as well as the time a customer occupies the server at a node are random.

In this thesis we generalise the sample path large deviation principle for Jackson networks [12]: In a Jackson network times between arrivals of customers at a node, as well as the time a customer occupies the server at a node are iid exponentially distributed. We generalise this to arbitrary light tailed distributions. Independence assumptions of the Jackson network are not challenged. We start - in this chapter - with investigating general inter event times and comparing them to exponential ones: How they are different and what angle can be chosen to highlight similarities.

We will see that there are basically two classes of inter event times: those that stay relatively small (we call them LD-bounded) and those that may be large. The exponential distribution is one that produces not LD-bounded inter event times and here we will apply general properties of not LD-bounded inter event times to make up for the lost Markov property.

2.1 Inter event time

In a GI/GI/1 queue times between consecutive arrivals are iid and so are the required lengths of service for each customer. We will refer to times between consecutive arrivals and to the lengths of service as inter event times.

Definition 2.1.1. *An inter event time is a non-negative random variable.*

Inter event times are often denoted by τ and variations of it. The distribution function of the inter event time τ will be denoted F. Throughout this thesis we follow the convention that for F a distribution function $F^c = 1 - F$.

Definition 2.1.2 ($\sim$-transform of F)**.** *For a distribution function F of an inter event time with finite mean define the distribution function $\tilde{F}$ as*

$$\tilde{F}(x) := \int_{s=0}^{x} \frac{F^c(s)}{\int_{t=0}^{\infty} F^c(t)\,dt}\,ds$$

Note that $\tilde{F}$ has density

$$\tilde{f}(x) = \frac{F^c(x)}{\int_{s=0}^{\infty} F^c(s)\,ds} = \frac{F^c(x)}{\mathbb{E}[\tau]} \tag{2.1}$$

and is the distribution function of an inter event time. The inter event time with distribution function $\tilde{F}$ will be denoted $\tilde{\tau}$. We see how the mean of τ and $\tilde{\tau}$ relate:

Claim 2.1.3. *Let τ be an inter event time and $\tilde{\tau}$ associated with it. If $\mathbb{E}[\tau^{n+1}] < \infty$ then $\mathbb{E}[\tilde{\tau}^n] < \infty$ and $\mathbb{E}[\tilde{\tau}^n] = \frac{\mathbb{E}[\tau^{n+1}]}{(n+1)\mathbb{E}[\tau]}$.*

Proof of 2.1.3:

$$\begin{aligned}
\mathbb{E}[\tau^{n+1}] &= \int_{x=0}^{\infty} x^{n+1}\,dF(x) \\
&= \lim_{z\to\infty} \Big(z^{n+1}F(z) - \int_{x=0}^{z} F(x-)\,dx^{n+1} \Big) \\
&= \lim_{z\to\infty} \Big((n+1)\int_{x=0}^{z} x^n\,dx\,F(z) - (n+1)\int_{x=0}^{z} F(x)\,x^n\,dx \Big) \\
&= \lim_{z\to\infty} (n+1)\int_{x=0}^{z} (F(z) - F(x))\,x^n\,dx \\
&= (n+1)\int_{x=0}^{\infty} (1 - F(x))\,x^n\,dx \\
&= (n+1)\mathbb{E}[\tau]\int_{x=0}^{\infty} \frac{F^c(x)}{\mathbb{E}[\tau]}\,x^n\,dx \\
&= (n+1)\mathbb{E}[\tau]\,\mathbb{E}[\tilde{\tau}^n]
\end{aligned}$$

$\square$
2.1.3

We construct another inter event time from τ:

Definition 2.1.4 (τ° associated with τ, G)**.** *Let $\tau, \tau_1, \tau_2, \ldots$ be iid with distribution function F, $p \in (0,1)$ fixed and G the geometrically distributed random* *variable with mass function* $P(G = g) = p^{g-1}(1-p)$ *for* $g = 1, 2, \ldots$. *Let* $G, \tau_1, \tau_2, \ldots$ *be independent. Define* τ° *as*

$$\tau^\circ = \sum_{k=1}^{G} \tau_k.$$

Claim 2.1.5. *If* τ *is an inter event time and* τ° *is associated with* τ *and* G *with parameter* p *then the mean relate as:* $\mathbb{E}[\tau^\circ] = \frac{1}{1-p}\mathbb{E}[\tau]$.

Proof of 2.1.5: From independence of $\{G, \tau_1, \tau_2, \ldots,\}$

$$\begin{aligned} \mathbb{E}[\tau^\circ] &= \mathbb{E}[\sum_{k=1}^{G} \tau_k] = \sum_{g=1}^{\infty} \underbrace{\mathbb{E}[\sum_{k=1}^{g} \tau_k]}_{=g\mathbb{E}[\tau]} P(G = g) = \mathbb{E}[G\mathbb{E}[\tau]] = \mathbb{E}[G]\,\mathbb{E}[\tau] \\ &= \frac{1}{1-p}\mathbb{E}[\tau] \end{aligned}$$

□
2.1.5

Generally, τ° is not qualitatively different from τ and whenever working with τ it might be of the form τ°.

Remark 2.1.6. *We will later see that* $\tilde{\tau}$ *as the time to the first event makes the renewal counting process have stationary increments and that* τ° *is the inter event time at a service node in a network that allows the leaving customer to immediately join again the queue just left.*

A few times in this thesis we will need the distribution of an inter event time τ conditional on $\tau > a$ for some positive a.

Definition 2.1.7. *For a distribution function* F *of an inter event time and* $a \in \mathbb{R}_{\geq 0}$ *such that* $F(a) \in [0, 1)$ *define*

$$F_{+a} : [0, \infty) \to [0, \infty) \quad , \quad x \mapsto \frac{F(x + a) - F(a)}{F^c(a)}$$

Claim 2.1.8. *If* τ *has distribution function* F *and* $a \in \mathbb{R}$ *is such that* $F(a) \in [0, 1)$ *then* F_{+a} *is the distribution function of* $\tau - a$ *conditional on* $\tau > a$.

Proof of 2.1.8:

$$\begin{aligned}
\mathbb{P}(\tau - a > x \,|\, \tau > a) &= \frac{\mathbb{P}(\tau - a > x\,,\, \tau > a)}{\mathbb{P}(\tau > a)} = \frac{\mathbb{P}(\tau > x + a)}{\mathbb{P}(\tau > a)} \\
&= \frac{F^c(x+a)}{F^c(a)} \\
\mathbb{P}(\tau - a \le x \,|\, \tau > a) &= 1 - \frac{F^c(x+a)}{F^c(a)} = \frac{F^c(a) - F^c(x+a)}{F^c(a)} \\
&= \frac{F(x+a) - F(a)}{F^c(a)}
\end{aligned}$$

□ 2.1.8

2.2 The logarithmic moment generating function

The logarithmic moment generating function is an essential in classical large deviation theory.

Definition 2.2.1 $(\Lambda, \mathcal{D}(\Lambda))$**.** *For an inter event time τ the logarithmic moment generating function Λ is defined as*

$$\Lambda \quad : \quad \mathbb{R} \to \mathbb{R} \cup \{\infty\} \quad , \quad \theta \mapsto \log \mathbb{E}[e^{\theta\tau}]$$

and abbreviated as lmgf. The domain of Λ is defined as

$$\mathcal{D}(\Lambda) := \{\theta \in \mathbb{R} \,|\, \Lambda(\theta) < \infty\}.$$

For a distribution function F of an inter event time we may also say that Λ is the lmgf of F; similarly for a density f of an inter event time.

Throughout this thesis we make the following

Assumption 2.2.2. *$\mathcal{D}(\Lambda)$ is open and $\inf_{\theta\in\mathbb{R}} \Lambda(\theta) = -\infty$.*

The following two claims are to interprete the assumption:

Claim 2.2.3. *τ has point mass at 0 iff its lmgf Λ is bounded from below.*

Proof of 2.2.3. We first establish $\log F(0)$ as a lower bound of Λ. Let $\theta < 0$ and $\epsilon > 0$.

$$\begin{aligned}
E[e^{\theta\tau}] &= \mathbb{E}[e^{\theta\tau}\, \mathbb{1}_{[0,\epsilon]}(\tau) + e^{\theta\tau}\mathbb{1}_{(\epsilon,\infty)}(\tau)] \ge e^{\theta\epsilon}F(\epsilon) \\
E[e^{\theta\tau}] &\ge \lim_{\epsilon\searrow 0} e^{\theta\epsilon}F(\epsilon) = F(0) \qquad (F \text{ right-continuous})
\end{aligned}$$

If $F(0) = 0$ we found a trivial lower bound for Λ and we have to show that there is no other, non-trivial lower bound. Let $\epsilon > 0$ again and first choose a such that $0 < a \leq F^{-1}(\frac{\epsilon}{2})$ and then θ such that $\theta < \frac{1}{a}\log\frac{\epsilon}{2} < 0$. With these

$$\mathbb{E}[e^{\theta\tau}] \quad \leq \quad e^{\theta a}\, F^c(a) + F(a) \leq \epsilon.$$

$\square$
2.2.3

Note that by definition of Λ we have $\Lambda(0) = 0$ and by the assumed (in 2.2.2) openness of $\mathcal{D}(\Lambda)$ there is $\theta > 0$ such that $\Lambda(\theta) < \infty$.

Claim 2.2.4. *If $\Lambda(\theta) < \infty$ for some $\theta > 0$ then all moments of τ exist.*

Proof of 2.2.4: For $\theta > 0$ and $x \in \mathbb{R}_{\geq 0}$ we have non-negative $(l, x) \mapsto \frac{(\theta x)^l}{l!}$ and are allowed to interchange summation and integration as an application of the Tonelli theorem (cf [21] theorem 20 of chapter 12, p. 270).

$$\begin{aligned} \mathbb{E}[e^{\theta\tau}] &= \int_{x=0}^{\infty} e^{\theta x}\, dF(x) = \int_{x=0}^{\infty} \sum_{l=0}^{\infty} \frac{(\theta x)^l}{l!}\, dF(x) = \sum_{l=0}^{\infty} \frac{\theta^l}{l!} \int_{x=0}^{\infty} x^l\, dF(x) \\ &= \sum_{l=0}^{\infty} \frac{\theta^l}{l!}\, \mathbb{E}[\tau^l] \end{aligned} \tag{2.2}$$

$\square$
2.2.4

We get $\Lambda \in \mathcal{C}^\infty(\mathcal{D}(\Lambda)^\circ)$ from $e^{\Lambda(\theta)}$ being a power series. e^Λ being a power series we are also allowed to interchange differentiation and summation in the following

$$\frac{d}{d\theta}\mathbb{E}[e^{\theta\tau}] \;=\; \sum_{l=0}^{\infty} \frac{d}{d\theta}\frac{\theta^l}{l!}\,\mathbb{E}[\tau^l] \;=\; \sum_{l=0}^{\infty} \frac{\theta^l}{l!}\,\mathbb{E}[\tau^{l+1}] \;=\; \mathbb{E}[\tau \sum_{l=0}^{\infty} \frac{\theta^l}{l!}\tau^l] \;=\; \mathbb{E}[\tau e^{\theta\tau}]$$

Openness of the domain allows this easy differentiation for all $\theta \in \mathcal{D}(\Lambda)$.

Claim 2.2.5. *Λ is convex and strictly increasing. If τ is not deterministic we have strict convexity of Λ on its domain.*

Proof of 2.2.5: Convexity follows from the Hölder inequality. By differentiating in the open domain

$$\frac{d}{d\theta}\Lambda(\theta) = \mathbb{E}[\tau\frac{e^{\theta\tau}}{e^{\Lambda(\theta)}}] > 0 \quad , \quad \frac{d^2}{d\theta^2}\Lambda(\theta) = \mathbb{E}[\tau^2\frac{e^{\theta\tau}}{e^{\Lambda(\theta)}}] - \mathbb{E}[\tau\frac{e^{\theta\tau}}{e^{\Lambda(\theta)}}]^2$$

While $\frac{d}{d\theta}\Lambda(\theta) > 0$ follows from $P(\tau > 0) > 0$ as implied by assumption 2.2.2 we have $\frac{d^2}{d\theta^2}\Lambda(\theta) > 0$ only for $P(\tau = x) < 1$ for all $x \in \mathbb{R}_{\geq 0}$. $\square$
2.2.5

Remark 2.2.6. *Note that*

$$\frac{e^{\theta\tau}}{e^{\Lambda(\theta)}} > 0 \qquad , \qquad \mathbb{E}[\frac{e^{\theta\tau}}{e^{\Lambda(\theta)}}] = 1$$

and we can write $\Lambda'(\theta)$ and $\Lambda''(\theta)$ as expectation and variance of τ unter a changed measure.

Openness of the domain $\mathcal{D}(\Lambda)$ also implies that Λ is not bounded from above. So 2.2.2 implies $\Lambda(\mathcal{D}(\Lambda)) = \mathbb{R}$ and the following is a feasible definition.

Definition 2.2.7 (Γ associated with Λ)**.** *Let Λ be the lmgf of an inter event time τ for which assumption 2.2.2 holds. Then define Γ associated with Λ as*

$$\Gamma \quad : \quad \mathbb{R} \to \mathbb{R} \quad , \quad \theta \mapsto -\Lambda^{-1}(-\theta)$$

Claim 2.2.8. *$\mathcal{D}(\Gamma) = \mathbb{R}$, Γ is strictly increasing, and (if τ is not deterministic) strictly convex on $\mathbb{R}$.*

Proof of 2.2.8: Finiteness of Γ on all of $\mathbb{R}$ should be immediate from the definition. Strict convexity of Λ was argued for in 2.2.5 and implies strict convexity for Γ:

$$\begin{aligned} \frac{d}{d\theta}\Gamma(\theta) &= -\frac{d}{d\theta}\Lambda^{-1}(-\theta) = \frac{1}{\Lambda'(-\Gamma(\theta))} > 0 \\ \frac{d^2}{d\theta^2}\Gamma(\theta) &= \frac{d}{d\theta}\frac{1}{\Lambda'(-\Gamma(\theta))} = \frac{\Lambda''(-\Gamma(\theta))}{\Lambda'(-\Gamma(\theta))^3} > 0 \end{aligned}$$

$\square$ 2.2.8

So properties of Λ translate into properties of Γ. For example

- $\lim_{\theta\to-\infty} \Lambda'(\theta) = 0 \Leftrightarrow \lim_{\theta\to\infty} \Gamma'(\theta) = \infty$
- $\mathcal{D}(\Lambda) = (-\infty, L_C(h))$ is equivalent to $\Gamma(\mathbb{R}) = (-L_C(h), \infty)$.
- $\Gamma'(\theta) = \frac{1}{\mathbb{E}^{(-\Gamma(\theta))}[\tau]}$

We will now investigate properties of the lmgf of inter event times $\tilde{\tau}$ and τ°.

Definition 2.2.9. *The lmgf of inter event time $\tilde{\tau}$ with distribution function $\tilde{F}$ is denoted $\tilde{\Lambda}$:*

$$\tilde{\Lambda}(\theta) = \log \mathbb{E}[e^{\theta\tilde{\tau}}] = \log \int_{t=0}^{\infty} e^{\theta t}\, d\tilde{F}(t)$$

The lmgf of inter event time τ° is denoted Λ°:

$$\Lambda^\circ(\theta) = \log \mathbb{E}[e^{\theta\tau^\circ}]$$

Claim 2.2.10. $\mathcal{D}(\Lambda) = \mathcal{D}(\tilde{\Lambda})$ *and*

$$\tilde{\Lambda}(\theta) = \begin{cases} \log \frac{e^{\Lambda(\theta)}-1}{\mathbb{E}[\tau]\,\theta} & ,\theta \neq 0 \\ 0 & ,\theta = 0. \end{cases} \tag{2.3}$$

Proof of 2.2.10: If for τ all moments exist the same holds for $\tilde{\tau}$ (cf claim 2.1.3). And we can write the lmgf of $\tilde{\tau}$ in terms of the lmgf of τ. While $\log \mathbb{E}[e^{0\tilde{\tau}}] = 0$ for $\theta \neq 0$ we get

$$\begin{aligned} \mathbb{E}[e^{\theta\tilde{\tau}}] &= \sum_{l=0}^{\infty} \frac{\theta^l}{l!}\,\mathbb{E}[\tilde{\tau}^l] \\ &= \sum_{l=0}^{\infty} \theta^l \underbrace{\frac{1}{l!}\frac{1}{l+1}}_{=\frac{1}{(l+1)!}} \frac{1}{\mathbb{E}[\tau]}\,\mathbb{E}[\tau^{l+1}] \\ &= \frac{1}{\mathbb{E}[\tau]}\frac{1}{\theta}\sum_{l=1}^{\infty} \frac{\theta^l\,\mathbb{E}[\tau^l]}{l!} \\ &= \frac{1}{\mathbb{E}[\tau]\,\theta}\Big(\sum_{l=0}^{\infty} \frac{\theta^l\,\mathbb{E}[\tau^l]}{l!} - 1\Big) \\ &= \frac{1}{\mathbb{E}[\tau]\theta}(\mathbb{E}[e^{\theta\tau}] - 1) \end{aligned}$$

□
2.2.10

We see that $\tilde{\tau}$ falls under assumption 2.2.2. $\tilde{\Lambda} \in \mathcal{C}^\infty(\mathcal{D}(\tilde{\Lambda})^\circ) = \mathcal{C}^\infty(\mathcal{D}(\tilde{\Lambda}))$ and we need not worry about continuity of $\tilde{\Lambda}$ in $\theta = 0$.

Claim 2.2.11. *If τ° is associated with τ and geometric G of parameter $p \in (0,1)$ then*

$$\Lambda^\circ(\theta) = \Lambda(\theta) + \log\frac{1-p}{1-p\,e^{\Lambda(\theta)}} \quad , \quad \mathcal{D}(\Lambda^\circ) = \big(-\infty, \Lambda^{-1}(-\log p)\big)$$

Proof of 2.2.11:

$$\mathbb{E}[e^{\alpha G}] = \frac{1-p}{1-p\,e^{\alpha}}\,e^{\alpha}$$

is finite only for $\alpha < \log\frac{1}{p}$. For $\theta < \Lambda^{-1}(-\log p)$

$$\begin{aligned} \mathbb{E}[e^{\theta\tau^\circ}] &= \mathbb{E}[e^{\theta\sum_{k=1}^{G}\tau_k}] \;=\; \mathbb{E}[\,\mathbb{E}[e^{\theta\sum_{k=1}^{G}\tau_k}\,|\,G]\,] \\ &= \mathbb{E}[\,\mathbb{E}[e^{\theta\tau_1}\,|\,G]^G\,] \;=\; \mathbb{E}[\,\mathbb{E}[e^{\theta\tau_1}]^G\,] \\ &= \mathbb{E}[\big(e^{\Lambda(\theta)}\big)^G] \;=\; \mathbb{E}[e^{\Lambda(\theta)\,G}] \end{aligned}$$

is finite and

$$\log \mathbb{E}[e^{\theta\tau^\circ}] \;=\; \log \mathbb{E}[e^{\Lambda(\theta)\, G}] \;=\; \log \frac{1-p}{1-p\, e^{\alpha}}\, e^{\alpha}\Big|_{\alpha=\Lambda(\theta)}$$

as claimed. Openness of the domain of Λ° follows from openness of the domain of the lmgf of G. $\square$ 2.2.11

Corollary 2.2.12. *If for τ assumption 2.2.2 holds then this assumption holds for τ° of definition 2.1.4, too.*

Proof of 2.2.12: Openness of the domain of Λ° comes from 2.2.11. Unboundedness from below is immediate from the explicit form of Λ°:

$$\begin{aligned}\lim_{\theta\to-\infty} \Lambda^\circ(\theta) &= \lim_{\theta\to-\infty} \Lambda(\theta) + \log \lim_{\theta\to-\infty} \frac{1-p}{1-p\, e^{\Lambda(\theta)}} \\ &= -\infty + \log(1-p) = -\infty\end{aligned}$$

$\square$ 2.2.12

Throughout this thesis we make the following

Assumption 2.2.13. *τ has a density f and $\limsup_{a\to\infty} \mathbb{E}[\tau - a \,|\, t > a] < \infty$.*

This assumption is mainly for technical convenience and to ease proofs. Existence of the density of course implies the unboundedness from below of Λ of assumption 2.2.2 - only $\mathbb{P}(\tau = 0) = 0$ was required for this. The boundedness of the conditional expectation is a property of many inter event times. Also, as an interpretation note that $\lim_{a\to\infty} \mathbb{E}[\tau - a \,|\, \tau > a]$ $(= \lim_{a\to\infty} \int_{x=0}^{\infty} F^c_{+a}(x)\, dx\,) = \infty$ is a harsh case of used better than new - excluding it should do no harm.

2.3 Exponential twist of inter event times

Definition 2.3.1 (Twisted distribution, exponential transform of F)**.** *Let F be the distribution function of an inter event time with lmgf Λ. Let $\beta \in \mathcal{D}(\Lambda)$ then the twisted distribution F_β is defined as*

$$F_\beta(x) := \int_{s=0}^{x} e^{\beta s - \Lambda(\beta)}\, dF(s)$$

and β will be called the twist parameter.

We can say that F_β has density $s \mapsto e^{\beta s - \Lambda(\beta)}$ wrt F and that if F has density f (wrt Lebesgue measure) then F_β has density (wrt Lebesgue measure)

$$f_\beta(x) := f(x)\, e^{\beta x - \Lambda(\beta)}. \tag{2.4}$$

The exponential transform of an exponential distribution is again an exponential distribution, the parameter changes as $\mu \mapsto \mu - \beta$ for β the parameter in the exponential transform. In this section we describe properties of exponential transforms applied to general inter event times.

Claim 2.3.2. *If τ falls under assumption 2.2.2 and τ_β is associated with τ through the exponential transform with parameter $\beta \in \mathcal{D}(\Lambda)$ then τ_β falls under assumption 2.2.2, too.*

Proof of 2.3.2: We calculate Λ_β, the lmgf of τ_β with distribution function F_β.

$$\begin{aligned}
\mathbb{E}[e^{\theta\tau_\beta}] &= \int_{s=0}^{\infty} e^{\theta s}\, dF_\beta(s) \;=\; \int_{s=0}^{\infty} e^{\theta s}\, e^{\beta s - \Lambda(\beta)}\, dF(s) \\
&= \int_{s=0}^{\infty} e^{(\theta+\beta)\, s}\; dF(s)\, e^{-\Lambda(\beta)} \;=\; \mathbb{E}[e^{(\theta+\beta)\tau}]\frac{1}{\mathbb{E}[e^{\beta\tau}]} \\
\Lambda_\beta(\theta) &= \log E^{(\beta)}[e^{\theta\tau}] = \Lambda(\beta+\theta) - \Lambda(\beta)
\end{aligned} \tag{2.5}$$

which we might want to write as $\mathcal{D}(\Lambda_\beta) = \mathcal{D}(\Lambda) - \beta$. So $\beta \in \mathcal{D}(\Lambda)$ is equivalent to $0 \in \mathcal{D}(\Lambda_\beta)$ and generally openness of the domain is not changed by its translation. Unboundedness from below for Λ_β is immediate. □ 2.3.2

Lemma 2.3.3. *Exponential transforms can be inverted: If $\beta \in \mathcal{D}(\Lambda)$ then*

$$(F_\beta)_{-\beta} = F \qquad , \qquad (f_\beta)_{-\beta} = \beta$$

Proof of 2.3.3: The second twist with parameter $-\beta$ has to be relative to f_β. We write the twists explicitly.

$$\begin{aligned}
f_\beta(x) &= f(x)\, e^{\beta x - \Lambda(\beta)} \\
(f_\beta)_{-\beta}(x) &= f_\beta(x)\, e^{-\beta x - \Lambda_\beta(-\beta)} \\
e^{\Lambda_\beta(\theta)} &= E^{(\beta)}[e^{\theta\tau}] \;=\; \int_{x=0}^{\infty} e^{\theta x}\, f_\beta(x)\, dx \\
&= \int_{x=0}^{\infty} e^{\theta x}\, f(x)\, e^{\beta x - \Lambda(\beta)} dx \;=\; \int_{x=0}^{\infty} e^{(\theta+\beta)\, x}\, f(x)\, e^{-\Lambda(\beta)} dx \\
e^{\Lambda_\beta(-\beta)} &= \int_{x=0}^{\infty} f(x)\, e^{-\Lambda(\beta)} dx \;=\; e^{-\Lambda(\beta)} \\
(f_\beta)_{-\beta}(x) &= f(x)\, \underbrace{e^{(\beta-\beta)\, x}}_{=1}\, \underbrace{e^{-\Lambda(\beta)+\Lambda(\beta)}}_{=1}
\end{aligned}$$

We check that everything is well defined.

$$-\beta \in \mathcal{D}(\Lambda_\beta) \;=\; \mathcal{D}(\Lambda) - \beta \;\Leftrightarrow\; 0 \in \mathcal{D}(\Lambda)$$

Since we calculated $\Lambda_\beta(-\beta) = -\Lambda(\beta)$ the lhs of this expression is finite whenever $\beta \in \mathcal{D}(\Lambda)$, a condition we started with. So all's well. $\square$ 2.3.3

We close the section with some remarks

Remark 2.3.4.

- *The $\sim$-transform and the exponential transform of F do not generally commute: $(\tilde{F})_\beta \neq \widetilde{(F_\beta)}$.*
  Writing down $(\tilde{f})_\beta$ and $\widetilde{(f_\beta)}$ equality requires that $x \mapsto e^{\beta x}\frac{F^c}{F^c_\beta}(x)$ is constant. This holds for the exponential but not for the uniform distribution.

- *Denote expectation and variance wrt the exponentially transformed distributions of 2.3.1 with parameter θ by indexing with (θ). Then derivatives of Λ can be written as $\Lambda'(\theta) = \mathbb{E}^{(\theta)}[\tau]$ and $\Lambda''(\theta) = \mathbb{V}^{(\theta)}[\tau]$ (cf 2.2.5, 2.2.6).*

2.4 Hazard

We introduce the hazard rate and make some mild technical assumptions on it. We'll see how the hazard rate relates to the domain of the lmgf, especially in terms of boundedness. Also, we give examples for the hazard rate - for well known distributions and when constructing inter event times from the hazard rate.

We also investigate how hazard rates change under exponentially twisting the distribution, and how hazard rates of τ and $\tilde{\tau}$ relate.

We will again see an analogy to the exponential distribution: When defined the right way the mean rate of the hazard function equals the bound of the domain.

2.4.1 Introduction of the hazard function

For inter event time τ with distribution function F define $H := -\log F^c$ on the support of τ. Assuming absolute continuity of F also H has a derivative almost everywhere and $h(x) = \frac{d}{dx}H(x) = \frac{\frac{d}{dx}F(x)}{F^c(x)} = \frac{f}{F^c}(x)$ where it exists.

Interpreting h in terms of probability:

$$h(x) = \lim_{\epsilon \to 0} \frac{\frac{1}{\epsilon} P(\tau \in (x, x+\epsilon])}{P(\tau > x)} = \lim_{\epsilon \to 0} \frac{1}{\epsilon} P(\tau \in (x, x+\epsilon] \,|\, \tau > x)$$

which is the infinitesimal probability for τ to take a value in the infinitesimal interval $(x, x+dx]$ given that $\tau > x$. h is called the hazard function of τ.

Definition 2.4.1 (Hazard function, $L_C(h)$)**.** *For an inter event time τ with density f the hazard function h is*

$$h : [0, \infty) \to [0, \infty] \quad , \quad x \mapsto h(x) = \begin{cases} \frac{f}{F^c}(x) & , F^c(x) > 0 \\ 0 & , else \end{cases}$$

and $L_C(h) := \lim_{x \to \infty} \frac{1}{x} \int_{s=0}^{x} h(s)\, ds$ is the Cesaro mean of h.

In this thesis we always make the following

Assumption 2.4.2. *For an inter event time τ with hazard function h the Cesaro-mean $L_C(h)$ exists and $L_C(h) \in (0, \infty]$.*

Note that given a distribution function F the density f and the hazard function h are not uniquely defined. However, we have

$$F^c(x) = \exp\{-\int_{s=0}^{x} h(s)\, ds\} \text{ holds for all } x$$

$$-\frac{d}{dx} \log F^c(x) = h(x) \text{ holds for almost all } x \in \operatorname{supp}(f).$$

Thus we see that given a hazard function we can construct the distribution function. But, what is a hazard function in general?

Claim 2.4.3. *Let $h : [0, \infty) \to [0, \infty]$ be measurable and $H = \int h$. If $\lim_{\epsilon \to 0} H(\epsilon) = 0$ and $\lim_{x \to \infty} H(x) = \infty$ then $F = 1 - e^{-H}$ is a distribution function with a density. On the support of F the hazard function is a.s equal to h.*

Proof of 2.4.3: We prove that $F := 1 - e^{-H}$ is a distribution function.

- F is non-negative and continuous since H is.
- $\lim_{x \to \infty} F(x) = 1 - e^{-\lim_{x \to \infty} H(x)} = 1 - 0 = 1$
- F is increasing since $h \geq 0$ and H is increasing.

This distribution has a density iff F is an absolutely continuous function.

H is absolutely continuous by definition. And $g(z) = 1 - e^{-z}$ is Lipschitz continuous with $g'(z) = e^{-z} \leq 1$ for $z \geq 0$. Since $H \geq 0$ we get absolute continuity for $F = g \circ H$. Let x be such that $F(x) < 1$ ($\Leftrightarrow H(x) < \infty$):

$$\frac{d}{dx}F(x) \;=\; \frac{d}{dx}(1 - e^{-H(x)}) \;=\; -e^{-H(x)}(-h(x)) \;=\; h(x)\,\exp\{-\int_{s=0}^{x} h(s)\,ds\}$$

(with $H' = h$ where H is finite a.s. from absolute continuity) is the density of F.
How does the support of F relate to H? If $H(x) = \infty$ for some finite x, then $H(z) = \infty$ and $F^c(z) = 0$ for all $z > x$. The hazard of F is thus defined to be $= 0$ on $[x, \infty)$, though h may take any value there (from the claim).
Now on the support of F: if $F^c(x) < 1$ then $H(x) < \infty$ and applying the density f as calculated $\frac{f}{F^c} = \frac{he^{-H}}{F^c} = h$. $\square$ 2.4.3

Had we assumed continuity of h all of the above would have been immediate.

How does assumption 2.2.2 translate into H? We will discuss this in the following examples sections.
For a bounded inter event time $\tau \leq b$ it is necessary ($F(b) = 1$) that $\lim_{t\to b} H(t) = \lim_{t\to b}\int_{s=0}^{t} h(s)\,ds = \infty$ and $\limsup_{t\to b} h(t) = \infty$. This agrees with the following interpretation for the hazard: If an event has to happen before b then the force for it to happen increases without bound as time approaches b.

2.4.2 The hazard function and the domain of lmgf

For an exponential distribution with density $f(x) = \mu\, e^{-\mu x}$ the parameter μ is the boundary for the domain of its logarithmic moment generating function: $\mathcal{D}(\Lambda) = (-\infty, \mu)$ and the constant hazard rate $h \equiv \mu$. We generalise this.

Claim 2.4.4. *If the Cesaro-mean of the hazard rate diverges to ∞ the lmgf has an unbounded domain: $L_C(h) = \infty \Rightarrow \mathcal{D}(\Lambda) = \mathbb{R}$.*

Proof of 2.4.4: Let $M > \theta$ and x_0 be large enough for $\frac{H(x)}{x} > M$ for all

$x \geq x_0$.

$$\begin{aligned}
\mathbb{E}[e^{\theta\tilde{\tau}}] &= \frac{1}{\mathbb{E}[\tau]}\Big(\int_{x=0}^{x_0} e^{\theta x}\, F^c(x)\, dx + \int_{x=x_0}^{\infty} \underbrace{e^{\theta x}\, F^c(x)}_{=e^{x(\theta-\frac{H(x)}{x})}}\, dx\Big) \\
&\leq \frac{1}{\mathbb{E}[\tau]}\Big(\int_{x=0}^{x_0} e^{\theta x}\, F^c(x)\, dx + \int_{x=x_0}^{\infty} e^{x(\theta-M)}\, dx\Big) \;<\; \infty
\end{aligned}$$

By 2.2.10 this is equivalent to unboundedness of $\mathcal{D}(\Lambda)$. □
2.4.4

Claim 2.4.5. *If the Cesaro-limit $L_C(h)$ exists in $(0, \infty)$ then the domain of the lmgf of τ is bounded and $\mathcal{D}(\Lambda) = (-\infty, L_C(h))$.*

Proof of 2.4.5: For $\theta < L_C(h)$ set $\epsilon := \frac{L_C(h)-\theta}{2} > 0$. Let x_0 be large enough for

$$\frac{H(x)}{x} > L_C(h) - \epsilon \quad \forall x \geq x_0.$$

Then

$$\begin{aligned}
\theta x - H(x) &= x(\theta - \frac{H(x)}{x}) \;\leq\; x(\theta - (L_C(h) - \epsilon)) \;\leq\; x(-2\epsilon + \epsilon) \\
&= -\epsilon x
\end{aligned}$$

and

$$\int_{x=x_0}^{\infty} e^{\theta x} F^c(x)\, dx \;\leq\; \int_{x=x_0}^{\infty} e^{-\epsilon x}\, dx \;\stackrel{x_0\to\infty}{\to}\; 0$$

and by 2.2.10 also $\theta \in \mathcal{D}(\Lambda)$.
Now let $\theta > L_C(h)$ and $\epsilon = \frac{\theta - L_C(h)}{2} > 0$ and x_0 large enough for $\frac{H(x)}{x} < L_C(h) + \epsilon$. Then

$$\theta x - H(x) \;=\; x(\theta - \frac{H(x)}{x}) \;>\; x(\theta - L_C(h) - \epsilon) > \epsilon x$$

and $\int_{x=x_0}^{\infty} e^{\theta x} F^c(x)\, dx = \infty$ and $E[e^{\theta\tilde{\tau}}] = \infty$ implying $\theta \notin \mathcal{D}(\tilde{\Lambda})$ and $L_C(h)$ thus has to be on the boundary of $\overline{\mathcal{D}(\tilde{\Lambda})}$. Since $\mathcal{D}(\Lambda) = \mathcal{D}(\tilde{\Lambda})$ the claimed statement is justified. □
2.4.5

Corollary 2.4.6. *From 2.4.4 and 2.4.5 we can completely generalise the property of the exponential distribution: If $L_C(h)$ exists in $(0, \infty]$ then $\mathcal{D}(\Lambda) = (-\infty, L_C(h))$.*

Similarly for the exponential distribution $L_C(h) = \mu$ is the exact exponential decay rate of the tail of the distribution function $t \mapsto e^{-\mu t}$. We generalise this expression to the more general F we allow for inter event time distributions:

Remark 2.4.7. *For $L_C(h) = \infty$ we say the tails of F decay superexponentially and for $L_C(h) < \infty$ we'll say that F^c decays exponentially with rate $L_C(h)$.*

Corollary 2.4.8.

- $L_C(h) = L_C(\tilde{h})$
  due to $(-\infty, L_C(h)) \overset{2.4.6}{=} \mathcal{D}(\Lambda) \overset{2.2.10}{=} \mathcal{D}(\tilde{\Lambda}) \overset{2.4.6}{=} (-\infty, L_C(\tilde{h}))$.
- $L_C(h_\beta) = L_C(h) - \beta$
  due to $(-\infty, L_C(h)) \overset{2.4.6}{=} \mathcal{D}(\Lambda)$ *and* $\mathcal{D}(\Lambda_\beta) \overset{(2.5)}{=} \mathcal{D}(\Lambda) - \beta$.

In 2.2.2 we have assumed openness of the domain. The following is a sufficient condition for openness of $\mathcal{D}(\Lambda)$.

Lemma 2.4.9. *If $L_C(h) < \infty$ and $\limsup_{x\to\infty} H(x) - L_C(h)x < \infty$ then the domain of Λ is open as assumed in 2.2.2.*

Proof of 2.4.9: If $\limsup_{x\to\infty} H(x) - x\, L_C(h) < \infty$ then $\liminf_{x\to\infty} e^{x\, L_C(h) - H(x)} > 0$ and

$$\begin{aligned}\mathbb{E}[e^{L_C(h)\tilde{\tau}}] &= \frac{1}{\mathbb{E}[\tau]} \int_{x=0}^{\infty} e^{L_C(h)x}\, F^c(x)\, dx = \frac{1}{\mathbb{E}[\tau]} \int_{x=0}^{\infty} e^{L_C(h)x - H(x)}\, dx \\ &= \infty\end{aligned}$$

Then $\mathcal{D}(\tilde{\Lambda})$ does not contain its right boundary $L_C(\tilde{h}) = L_C(h)$ and since $\mathcal{D}(\tilde{\Lambda}) = \mathcal{D}(\Lambda)$ neither does $\mathcal{D}(\Lambda)$. □ 2.4.9

The equivalent condition for an open domain is obvious from the proof of 2.4.9:

$$\mathbb{E}[e^{L_C(h)\tilde{\tau}}] < \infty \quad \Leftrightarrow \quad e^{L_C(h)x - H(x)} \to 0 \,\text{integrably fast.}$$

2.4.3 Examples

We give several examples of hazard rates with different properties in boundedness and monotonicity. We discuss assumptions 2.2.2 and 2.4.2.

Example 2.4.10. *The exponential distribution is characterised by its constant hazard rate: $f(x) = \mu\, e^{-\mu x} \Leftrightarrow h(x) = \mu = L_C(h)$.*

Example 2.4.11. *The Erlang distribution $E_k(\mu)$ where $k \geq 1$ has a hazard rate that is monotonically increasing to μ.*

$$\begin{aligned} F^c(x) &= e^{-\mu x} \sum_{j=0}^{k-1} \frac{(\mu x)^j}{j!} \quad , \quad H(x) = \mu x - \log \sum_{j=0}^{k-1} \frac{(\mu x)^j}{j!} \\ f(x) &= \frac{\mu^k\, x^{k-1}}{(k-1)!} e^{-\mu x} \quad , \quad h(x) \overset{x\geq 0}{=} \mu\Big(1 + \sum_{j=1}^{k-1} (\mu\, x)^{-j} \frac{(k-1)!}{(k-1+j)!}\Big)^{-1} \\ L_C(h) &= \mu \end{aligned}$$

The domain of a hazard function with $h(x) < L_C(h)$ is generally open as implied by 2.4.9. Erlang distributed inter event times fall under both assumptions.

Example 2.4.12 (Oscillating hazard, discrete)**.** *If $h(x) = \kappa 1\!\!1_{\lfloor x \rfloor\ odd} + \mu 1\!\!1_{\lfloor x \rfloor\ even}$ then $L_C(h) = \frac{1}{2}(\mu + \kappa)$.*

We have

$$\begin{aligned} H(x) &= 2\lfloor \frac{x}{2} \rfloor \frac{\kappa + \mu}{2} + \int_{s=2\lfloor \frac{x}{2} \rfloor}^{x} h(s)\, ds \\ L_C(h)x - H(x) &= \int_{s=2\lfloor \frac{x}{2} \rfloor}^{x} \Big(\frac{\kappa+\mu}{2} - \kappa\Big) 1\!\!1_{\lfloor s \rfloor \text{ odd}} + \Big(\frac{\kappa+\mu}{2} - \mu\Big) 1\!\!1_{\lfloor s \rfloor \text{ even}}\, ds \\ &= \int_{s=2\lfloor \frac{x}{2} \rfloor}^{x} -\frac{\kappa-\mu}{2} 1\!\!1_{\lfloor s \rfloor \text{ odd}} + \frac{\kappa-\mu}{2} 1\!\!1_{\lfloor s \rfloor \text{ even}}\, dx \\ &\not\to -\infty \end{aligned}$$

Again from 2.4.9 we get an open $\mathcal{D}(\Lambda)$.

Example 2.4.13 (Oscillating hazard, continuous)**.** *If $h(x) = 1 + \sin(x)$ then $L_C(h) = 1$ and $F^c(x) = \exp\{-x - 1 + \cos x\}$. More generally for $a, b > 0$ define $h_{a,b}(x) := a + b(1 + sin(x))$.*

Again $L_C(h)x - H(x) \not\to -\infty$ and $\exp\{L_C(x)(h) - H(x)\}$ is not integrable. 2.4.9 applies.

Example 2.4.14. *If $h(x) = a + \frac{c}{x+b}$ for $a, b > 0$ and $c \in (0, 1]$ then $L_C(h) = a$.*

We have

$$\begin{aligned} L_C(h) &= a \quad , \quad H(x) = ax + c \log \frac{x+b}{b} \\ e^{L_C(h)x - H(x)} &= \Big(\frac{b}{x+b}\Big)^c \end{aligned}$$

And $\mathbb{E}[e^{L_C(h)\tilde{\tau}}] = \infty$ for the parameters given in the example.

$c > 1$ is excluded since otherwise the domain was not open. Similarly $b > 0$ is necessary since a function $x \mapsto a + \frac{1}{x}$ does not qualify as a hazard function as $\int_{x=0}^{z} a + \frac{1}{x}\, dx = \infty$ for any $z > 0$ or equivalently $\lim_{\epsilon\to 0} H(\epsilon) = \infty \neq 0$. Also the case of $a = 0$ does not fall under assumption 2.4.2

While $h(x) = 2 + \cos\log(1 + x)$ is a hazard function it does not fall under assumption 2.4.2.

Example 2.4.15 (Exponential hazard)**.** *If $h(x) = 1 - e^{-x}$ then $L_C(h) = 1$ and $F^c(x) = \exp\{-x - e^{-x} - 1\}$. More generally for $a > 1$, b define $h_{a,b}(x) = a + \operatorname{sign}(b)e^{bx}$.*

For $b < 0$ we have $L_C(h) = a$ and an open domain from 2.4.9. For $b > 0$ we get $L_C(h) = \infty$ and openness is not an issue. Existence of a density is obvious in both cases and so 2.2.2 holds.

Example 2.4.16. *The uniformly distributed random variable has*

$$\begin{aligned} f(x) &= \mathbb{1}_{[0,1]}(x) \qquad , \qquad F(x) = x\mathbb{1}_{[0,1]}(x) + \mathbb{1}_{(1,\infty)}(x) \\ h(x) &\overset{x\in[0,1]}{=} \frac{1}{1-x} \overset{(x\to 1)}{\to} \infty = L_C(h) \end{aligned}$$

Starting from the hazard function:

Example 2.4.17 (Affine hazard)**.** *If $h(x) = x$ then $L_C(h) = \infty$ and $F^c(x) = \exp\{-\frac{x^2}{2}\}$. More generally for $a, b > 0$ define $h_{a,b}(x) := a\,x + b$.*

Generalising this to polynomial hazards $h(x) = (ax)^k + b$ with $k \geq 1$ we arrive at Weibull-distributions.

2.5 Coupling

This section is on coupling inter event times.

Definition 2.5.1 (Coupled inter event times)**.** *Two inter event times are coupled if they have a joint distribution.*

Any two inter event times σ_1, σ_2 have a joint distribution as a tuple of independent random variables. Another possibility to construct a joint distribution while keeping individual (=marginal) distributions of σ_1, σ_2 is the quantile coupling. It is nicely explained in section 3 of chapter 1 in the book of Hermann Thorisson [24].

Definition 2.5.2 (Quantile coupling). *If inter event times σ_1, σ_2 have respective distribution functions F_1, F_2 and U is uniformly distributed on* $(0,1)$ *then $(F_1^{-1}(U), F_2^{-1}(U))$ is the quantile coupling of (σ_1, σ_2).*

It is obvious that $(F_1^{-1}(U), F_2^{-1}(U))$ are coupled. Furthermore the association of $(F_1^{-1}(U), F_2^{-1}(U))$ with (σ_1, σ_2) is in the fact that $\mathcal{L}(F_i^{-1}(U)) = \mathcal{L}(\sigma_i)$ for $i = 1, 2$:

$$\mathbb{P}(F_1^{-1}(U) \leq t) \;=\; \mathbb{P}(U \leq F_1(t)) \;=\; F_1(t) \;=\; \mathbb{P}(\sigma_1 \leq t). \tag{2.6}$$

A joint distribution is required if we want to consider a function of two random variables, for example their difference. This is required in the following definition of exponential equivalence.

Definition 2.5.3 (Exponential equivalence in $\mathbb{R}$). *The sequences of real valued random variables $(Y_n\,;\, n \in \mathbb{N})$ and $(Z_n\,;\, n \in \mathbb{N})$ are exponentially equivalent if for each $n \in \mathbb{N}$ there is a coupling $(\check{Y}_n, \check{Z}_n)$ of (Y_n, Z_n) such that the sequence $(|\check{Y}_n - \check{Z}_n|\,;\, n \in \mathbb{N})$ decays super exponentially: For any $\delta > 0$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|\check{Y}_n - \check{Z}_n| > \delta) = -\infty.$$

This definition is a special case of the more general definition 4.2.10 of [5]. And the importance of this definition is in theorem 4.2.13 of [5] which tells us that if one of two exponentially equivalent sequences satisfies a large deviation principle with a good rate function then the other does too, and with the same rate function.

We introduce a new property for the inter event times.

Definition 2.5.4. *A inter event time is LD-bounded if its lmgf is finite on all of* $\mathbb{R}$.

An alternative definition is that an inter event time is LD-bounded if its hazard function satisfies $L_C(h) = \infty$, by 2.4.6.

Thus any bounded inter event time will be LD-bounded. In this section we connect LD-boundedness to exponential equivalence.

Very generally we can scale a single inter event time τ and investigate its large deviation behaviour. Doing so, any nice non-negative random variable falls in one of two classes: It's either exponentially equivalent to 0 or to an exponentially distributed random variable with the correct parameter.

The correct parameter for the exponential distribution is the Cesaro mean of the hazard function $L_C(h)$ with h the hazard function of τ.

In the following we prove the claimed exponential equivalence separately for the LD-bounded and the non-LD bounded inter event times.

Claim 2.5.5. *If the inter event time σ is LD-bounded then the sequences $(\frac{\sigma}{n}\,;\, n \in \mathbb{N})$ and $(0\,;\, n \in \mathbb{N})$ are exponentially equivalent.*

As a shorter way to express 2.5.5 we might say that σ and 0 are exponentially equivalent.

Proof of 2.5.5: Let $L_C(h)$ be the Cesaro mean of the hazard rate of σ. We generally assumed that $L_C(h)$ exists in $(0, \infty]$ (cf. 2.4.2) and for LD-bounded σ we have $L_C(h) = \infty$ by 2.4.6. Thus

$$\begin{aligned}\frac{1}{n}\log P(\frac{1}{n}|\sigma - 0| > x) &= \frac{1}{n}\log F^c(nx) = \frac{1}{n}\log\exp\{-\int_{s=0}^{nx} h(s)\,ds\} \\ &= -x\,\underbrace{\frac{1}{x\,n}\int_{s=0}^{nx} h(s)\,ds}_{\to L_C(h)=\infty} \overset{n\to\infty}{\to} -\infty \qquad \square \; 2.5.5\end{aligned}$$

This immediately implies that any two LD-bounded random variables are exponentially equivalent.

Claim 2.5.6. *If σ is an inter event time with distribution function F and $L_C(h) \in (0, \infty)$ and if X is exponentially distributed with parameter $L_C(h)$ then the sequences $(\frac{\sigma}{n}\,;\, n \in \mathbb{N})$ and $(\frac{X}{n}\,;\, n \in \mathbb{N})$ are exponentially equivalent.*

Again, as a shorter way to express 2.5.6 we might say that if τ and exponential X have the same finite Cesaro mean for their respective hazard functions then they are exponentially equivalent.

Proof of 2.5.6: Let G be the distribution function of exponentially distributed X: $G(x) = 1 - e^{-L_C(h)x}$ and assume that X, σ are already quantile-coupled: that there is U uniform on $(0, 1)$ such that $X = G^{-1}(U)$ and $\sigma = F^{-1}(U)$. We first observe that a large value for σ implies a large value for X. Let $0 < s < t$.

$$\begin{aligned}\mathbb{P}(\sigma > nt\,,\, X \le ns) &= \mathbb{P}(U > F(nt)\,,\, U < G(ns)) \\ &= \mathbb{P}(U \in [F(nt)\,,\, G(ns)]) \\ &= \big(G(ns) - F(ns)\big)\mathbb{1}_{F(nt)<G(ns)}\end{aligned}$$

Let $\epsilon > 0$ be small enough for $\frac{t}{s} > \frac{L_C(h)}{L_C(h)-\epsilon}$ to hold and n large enough for $\frac{H(nt)}{nt} > L_C(h) - \epsilon$ to hold.

$$\begin{array}{rrcl} & \frac{t}{s} & > & \frac{L_C(h)}{L_C(h)-\epsilon} \\ \Rightarrow & nt(L_C(h)-\epsilon) & > & nsL_C(h) \\ \Rightarrow & H(nt) & > & nsL_C(h) \\ \Leftrightarrow & e^{-H(nt)} & < & e^{-nsL_C(h)} \\ \Leftrightarrow & F^c(nt) & < & G^c(ns) \end{array}$$

Thus for n large enough

$$\mathbb{P}(\sigma > nt\,,\, X \leq ns) = 0$$

and the same works very similarly for $P(X > nt\,,\, \tau < ns)$. Put another way we have for n large enough (depending on $t - s$)

$$\mathbb{P}(\sigma > nt) = \mathbb{P}(\sigma > nt\,,\, X \geq ns) \text{ or } \mathbb{P}(X \geq ns \,|\, \tau > nt) = 1. \tag{2.7}$$

We apply (2.7) in the following that will show the super exponential decay of $\sigma - X$:

$$\left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & 0 \\ X & > & 0 \end{array}\right\} \Leftrightarrow \left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & na \\ X & > & 0 \end{array}\right\} \overset{(2.7)}{\Leftrightarrow} \left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & na \\ X & > & n(a-\epsilon) \end{array}\right\}$$

We see that for two positive random variables to be large *and* their difference to be large, too, the subtrahend has to be even larger. With the help of $\sigma - X > na \Rightarrow \sigma > na + X > n(2a - \epsilon)$ the equivalence

$$\left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & na \\ X & > & n(a-\epsilon) \end{array}\right\} \Leftrightarrow \left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & n\,(2a-\epsilon) \\ X & > & n(a-\epsilon) \end{array}\right\}$$

is easy to see. We can iterate and again use that given σ is large, X will almost surely be similarly large for n large enough.

$$\left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & n\,(2a-\epsilon) \\ X & > & n(a-\epsilon) \end{array}\right\} \Leftrightarrow \cdots \Leftrightarrow \left\{\begin{array}{ccc} \sigma - X & > & na \\ \sigma & > & n\,((k+1)a - k\epsilon) \\ X & > & k\,n\,(a-\epsilon) \end{array}\right\}$$

Thus

$$\begin{array}{rcl} \mathbb{P}(\sigma - X > na) & = & \mathbb{P}(\sigma - X > na\,,\, \sigma > n((k+1)a - k\epsilon)\,,\, X > nk(a-\epsilon)) \\ & \leq & \mathbb{P}\big(X > nk(a-\epsilon)\big) \\ & = & e^{-L_C(h)\,n\,k\,(a-\epsilon)} \\ & \overset{\epsilon<\frac{a}{2}}{\leq} & e^{-L_C(h)\,n\,k\,\frac{a}{2}} \end{array}$$

and the last expression has an exponential decay rate in n of $L_C(h)k\frac{a}{2}$ which can be made arbitrarily large by increasing k (we need our assumption 2.4.2 here: that $L_C(h) > 0$). So the decay is faster than exponential.

Similarly for $X - \sigma$. $\square$ 2.5.6

The following is fairly general.

Claim 2.5.7. *If σ_1 , σ_2 are inter event times with distribution functions F_1, F_2 and integrated hazard functions $H_i = -\log F_i^c$ (for $i = 1, 2$) with a common Cesaro limit $L_C = \lim_{x\to\infty} \frac{H_i(x)}{x}$ (for $i = 1, 2$) then the sequences $(\frac{\sigma_1}{n}\,;\, n \in \mathbb{N})$ and $(\frac{\sigma_2}{n}\,;\, n \in \mathbb{N})$ are exponentially equivalent.*

Proof of 2.5.7: For $L_C < \infty$ as in 2.5.6 let X be exponentially distributed with parameter L_C and distribution function G and assume that σ_1, σ_2, G are quantile coupled, that is $\sigma_i = F_i^{-1}(U)$ (for $i = 1, 2$) and $X = G^{-1}(U)$ for some U uniformly distributed on $(0, 1)$. Then

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|\sigma_1 - \sigma_2| > n\delta) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log \Big(\mathbb{P}(|\sigma_1 - X| > n\frac{\delta}{2}) + \mathbb{P}(|\sigma_2 - X| > n\frac{\delta}{2})\Big) \\
&= \max\{\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|\sigma_1 - X| > n\frac{\delta}{2})\,,\ \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(|X - \sigma_2| > n\frac{\delta}{2})\} \\
&= \max\{-\infty\,,\, -\infty\} \;=\; -\infty
\end{aligned}$$

While for $L_C = \infty$ we have already argued that exponential equivalence holds. We might as well do the same calculation as above replacing X by 0. $\square$ 2.5.7

Corollary 2.5.8. *τ and $\tilde{\tau}$ are exponentially equivalent.*

So far we have applied the distinction between LD-bounded and not LD-bounded random variables in the different proofs for exponential equivalence of inter event times in 2.5.5 and 2.5.6. We now connect them directly:

Claim 2.5.9. *If inter event times σ, τ have hazard functions with the same Cesaro mean then their difference under the quantile coupling is LD-bounded.*

Proof of 2.5.9: For σ, τ themselves LD-bounded and $\theta > 0$

$$\begin{aligned}
\mathbb{E}[e^{|\sigma-\tau|\,(\theta+\epsilon)}] &\le \mathbb{E}[e^{\sigma\,(\theta+\epsilon)}\, e^{\tau\,(\theta+\epsilon)}] \\
&\le \mathbb{E}[e^{\sigma\,p\,(\theta+\epsilon)}]^{\frac{1}{p}}\, \mathbb{E}[e^{\tau\,q\,(\theta+\epsilon)}]^{\frac{1}{q}} \;<\; \infty
\end{aligned}$$

And if σ, τ are not LD-bounded: Let uniform U be such that $\sigma_i = F_i^{-1}(U)$ for $i = 1, 2$ and set $X = G^{-1}(U)$ for G the exponential distribution function with parameter L_C. We proved that quantile coupled σ_i and X have an LD-bounded difference. For the difference of X and σ_i under the quantile coupling we have bounded

$$\mathbb{P}(\sigma_i - X > na) \leq e^{-L_C\, k\, \frac{na}{2}}$$

with $a > 0$ and arbitrary $k \in \mathbb{N}$ for n large enough with reference to a, k. Let $\mathbb{F}_i$ be the distribution function of $|\sigma_i - X|$ and $\mathbb{H}$ the hazard function associated with $\mathbb{F}$ (that is: $\mathbb{F}_i^c = e^{-\mathbb{H}_i}$). Then

$$-\log \mathbb{F}_i^c(x) \geq L_C\, k\, \frac{x}{2} \qquad \text{and} \qquad \frac{\mathbb{H}_i(x)}{x} = \frac{-\log \mathbb{F}_i^c(x)}{x} \geq L_C\, k\, \frac{1}{2}$$

for x large enough. The limit takes care of x being large enough for any fixed k and

$$\forall k \in \mathbb{N} \,:\, \lim_{x \to \infty} \frac{\mathbb{H}_i(x)}{x} \geq L_C\, k\, \frac{1}{2} \qquad \Rightarrow \qquad \lim_{x \to \infty} \frac{\mathbb{H}_i(x)}{x} = \infty$$

This tells us that $\mathbb{E}[e^{\theta|\sigma_i - X|}] < \infty$ for all $\theta \in \mathbb{R}$. For $p, q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$ we get

$$\mathbb{E}[e^{\theta|\sigma_1 - \sigma_2|}] \leq \mathbb{E}[e^{\theta|\sigma_1 - X|}\, e^{\theta|\sigma_2 - X|}] \leq \mathbb{E}[e^{\theta p|\sigma_1 - X|}]^{\frac{1}{p}}\, \mathbb{E}[e^{\theta q|\sigma_2 - X|}]^{\frac{1}{q}} \,<\, \infty$$

□
2.5.9

2.6 Fenchel-Legendre transforms

Definition 2.6.1. *Given a logarithmic moment generating function $H : \mathbb{R}^m \to \mathbb{R}$ its Fenchel Legendre transform is defined as*

$$H^* \,:\, \mathbb{R}^m \to \mathbb{R} \quad , \quad x \mapsto \sup_{\theta \in \mathbb{R}^m} \langle \theta, x \rangle - H(\theta).$$

H^* is again convex and continuous on the interior of its domain.

Remark 2.6.2. *Sufficient criteria for a Fenchel-Legendre transform H^* of H to have compact level sets are*

- *for $m = 1$: $0 \in \mathcal{D}(H)^\circ$*
- *for arbitrary $m \in \mathbb{N}$: $\mathcal{D}(H) = \mathbb{R}^m$ or H is essentially smooth lower semicontinuous.*

(cf [5] lemma 2.2.20, theorem 2.3.6). Thus the Fenchel-Legendre transforms Λ^* *and* Γ^* *of lmgfs* Λ *and* Γ *are good rate functions.*

For an inter event time τ we have denoted its lmgf Λ. From Λ we defined an associated Γ (cf (2.2.7)). Their Fenchel-Legendre transforms relate, too.

Claim 2.6.3. $\Gamma^*(x) = x\Lambda^*(\frac{1}{x})$ *for* $x > 0$ *holds and* $\Gamma^*(0) = \lim_{x\to 0} x\Lambda^*(\frac{1}{x})$.

Proof of 2.6.3: First let $x > 0$.

$$\begin{aligned}
\Gamma^*(x) &= \sup_\theta \theta\, x - \Gamma(\theta) \;=\; \sup_\theta \theta\, x + \Lambda^{-1}(-\theta) \\
&= \sup_{\theta=-\Lambda(\gamma)\,;\,\gamma\in\mathbb{R}} \theta\, x + \Lambda^{-1}(-\theta) \\
&= \sup_{\gamma\in\mathbb{R}} -\Lambda(\gamma)\, x + \Lambda^{-1}\circ\Lambda(\gamma) \\
&= x \sup_{\gamma\in\mathbb{R}} -\Lambda(\gamma) + \frac{1}{x}\gamma \\
&= x\, \Lambda^*(\frac{1}{x})
\end{aligned}$$

While for $x = 0$

$$\Gamma^*(0) \;=\; \sup_{\theta\in\mathbb{R}} -\Gamma(\theta) = \sup_{\theta\in\mathbb{R}} \Lambda^{-1}(-\theta) \;=\; \lim_{\theta\to\infty} \Lambda^{-1}(\theta) \;=\; L_C(h)$$

If $\Gamma^*(0) = L_C(h) < \infty$ we get continuity of Γ^* at 0 from convexity and

$$\Gamma^*(0) = \lim_{x\to 0} \Gamma^*(x) = \lim_{x\to 0} x\Lambda^*(\frac{1}{x})$$

Otherwise, if $\Gamma^*(0) = L_C(h) = \infty$ we get simultaneous divergence to ∞ from lower semicontinuity of Γ^*:

$$\begin{aligned}
\infty &= \Gamma^*(0) \le \liminf_{x\to 0} \Gamma^*(x) = \liminf_{x\to 0} x\Lambda^*(\frac{1}{x}) \\
\Rightarrow \qquad \infty &= \Gamma^*(0) = \lim_{x\to 0} \Gamma^*(x) = \lim_{x\to 0} x\Lambda^*(\frac{1}{x})
\end{aligned}$$

$\square_{2.6.3}$

Corollary 2.6.4.

- $\Gamma^*(0) = \lim_{x\to 0} \Gamma^*(x)$
- $\Gamma^*(0) = L_C(h)$ *and the domain of* Γ^* *contains its left boundary point* $\{0\}$ *iff* τ *is LD-bounded.*

Similarly we are interested in how Γ^* behaves for large arguments.

Claim 2.6.5. Γ^* *behaves superlinearly at infinity.*

Proof of 2.6.5: We see the correspondence of unboundedness from below of Λ (left) and superlinearity of Γ^* (right).

$$\lim_{\theta\to-\infty} -\Lambda(\theta) = \Lambda^*(0) = \lim_{x\to\infty} \Lambda^*(\frac{1}{x}) = \lim_{x\to\infty} \frac{x}{x}\Lambda^*(\frac{1}{x}) = \lim_{x\to\infty} \frac{\Gamma^*(x)}{x}$$

where $\lim_{x\to\infty} \Lambda^*(\frac{1}{x}) = \Lambda^*(0)$ follows from lower semi continuity and $\Lambda^*(0) = \infty$).

$\square$
2.6.5

2.7 Extreme twists

What happens as the twist parameter θ for an inter event time tends to $\pm\infty$ or, if the domain of Λ is bounded, to $L_C(h)$? With the assumption 2.2.2 we have interpreted the derivative of the lmgf as the twisted mean ($\Lambda'(\theta) = \mathbb{E}^{(\theta)}[\tau]$) and so the range of Λ' is the interval of attainable means under exponential twists. I assume the inter event time under the twisted distribution tends to the essential infimum and the essential supremum of the untwisted inter event time as the twist parameter tends to $-\infty$ and $+\infty/L_C(h)$ respectively.

Claim 2.7.1. *If $F(0) = 0$ and there is some $k \in \mathbb{N}$ such that $f^{(0)}(0) = \cdots = f^{(k-1)}(0) = 0 < f^{(k)}(0)$ then $lim_{\theta\to\infty}\Lambda'(-\theta) = 0$.*

Proof of 2.7.1: First for $f(0) > 0$.

$$\begin{aligned}
\theta\mathbb{E}[e^{-\theta\tau}] &= \theta\int_{x=0}^{\infty} e^{-\theta\, x} f(x)\,dx = \underbrace{\theta^2\int_{x=0}^{\infty} e^{-\theta x} F(x)\,dx}_{(1)} \\
&\overset{y=\theta\, x}{=} \int_{y=0}^{\infty} e^{-y}\, F(\frac{y}{\theta})\,\frac{\theta^2}{\theta}\,dy \\
\lim_{\theta\to\infty}\theta\mathbb{E}[e^{-\theta\tau}] &= \lim_{\theta\to\infty}\int_{y=0}^{\infty} e^{-y}\,\theta\, F(\frac{y}{\theta})\,dy = \int_{y=0}^{\infty} e^{-y}\, y \lim_{\theta\to\infty}\frac{\theta}{y}\, F(\frac{y}{\theta})\,dy \\
&= f(0)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\tau e^{-\theta\tau}] &= \int_{x=0}^{\infty} x\, e^{-\theta x} f(x)\,dx \\
&= x\, e^{-\theta x} F(x)\big|_{x=0}^{\infty} - \int_{x=0}^{\infty}\left(e^{-\theta x} - x\,\theta\, e^{-\theta x}\right) F(x)\,dx \\
&= 0 - 0 - \underbrace{\int_{x=0}^{\infty} e^{-\theta\, x}\, F(x)\,dx}_{\text{see above}} + \theta\int_{x=0}^{\infty} x\, e^{-\theta x}\, F(x)\,dx
\end{aligned}$$

and

$$\theta^2 \int_{x=0}^{\infty} x\, e^{-\theta x}\, f(x)\, dx \;=\; -\,\underbrace{\theta^2 \int_{x=0}^{\infty} e^{-\theta\, x}\, F(x)\, dx}_{=f(0)\text{by (1) above}} + \underbrace{\theta^3 \int_{x=0}^{\infty} x\, e^{-\theta x}\, F(x)\, dx}_{=2f(0)\text{ by(2) below}}$$
$$= \; -f(0) + 2f(0) \;=\; f(0)$$

$$(2) : \lim_{\theta\to\infty} \theta^3 \int_{y=0}^{\infty} x\, e^{-\theta\, x} F(x)\, dx \;\overset{y=\theta\, x}{=}\; \lim_{\theta\to\infty} \theta^3 \int_{y=0}^{\infty} \frac{y}{\theta} e^{-y}\, F(\frac{y}{\theta})\, \frac{1}{\theta}\, dy$$
$$= \int_{y=0}^{\infty} y^2\, e^{-y} \lim_{\theta\to\infty} \frac{\theta}{y}\, F(\frac{y}{\theta})\, dy$$
$$= f(0) \int_{y=0}^{\infty} y^2\, e^{-y}\, dy \;=\; 2\, f(0)$$

All this we apply in

$$\lim_{\theta\to\infty} \theta\Lambda'(-\theta) \;=\; \frac{\lim_{\theta\to\infty} \theta^2 \mathbb{E}[\tau e^{-\theta\tau}]}{\lim_{\theta\to\infty} \theta \mathbb{E}[e^{-\theta\tau}]} \;=\; \frac{f(0)}{f(0)} \;=\; 1$$

If $f(0) = 0$ but $f'(0) > 0$ then we do a similar calculation with $\lim_{\epsilon\to 0} \frac{F(\epsilon)}{\epsilon^2} = \frac{1}{2} f'(0) > 0$ instead of the above $\lim_{\epsilon\to 0} \frac{F(\epsilon)}{\epsilon} = f(0) > 0$. This works generally as claimed. $\square$ 2.7.1

This claim is a generalisation of an assumption in [16], theorem 2.2. It is about being able to twist τ to have positive mean values as small as we like since we identified Λ' with the expectation under the twisted distribution.

Corollary 2.7.2. *If an inter event time τ is bounded with least upper bound $b > 0$, no point mass on 0 and density f such that there is some $k \in \mathbb{N}$ such that $f^{(0)}(b) = \cdots = f^{(k-1)}(b) = 0 < f^{(k)}(b)$ then $lim_{\theta\to\infty}\Lambda'(\theta) = b$.*

This is by considering $b - \tau$ as an inter event time and applying 2.7.1.

We note that we don't really need the density on the whole support but just in a neighbourhood of 0 and - for bounded τ - on the least upper bound.

Claim 2.7.3. *For τ with support $[0, b]$ we have $\Lambda'(\mathbb{R}) = (0, b)$ under assumption 2.7.1 for the density at both 0 and b.*

Proof of 2.7.3: Openness of $\Lambda'(\mathbb{R}) = (0, b)$ remains to be proved. Assume the contrary and let $\theta_0 \in \mathbb{R}$ be such that $\Lambda'(\theta_0) = b$. Then for $\theta > \theta_0$ we'd have $\Lambda'(\theta) = b$ since Λ' cannot decrease due to convexity and not increase due to $\lim_{\theta\to\infty} \Lambda'(\theta) = b$. $\square$ 2.7.3

Also we can argue for unbounded τ.

Claim 2.7.4. *If τ satisfies the if part of 2.7.1 and is unbounded but LD-bounded, then $\Lambda'(\mathbb{R}) = \mathbb{R}_{>0}$.*

Proof of 2.7.4: Fix some $M > 0$ and let $\theta > 0$.

$$\begin{aligned} \Lambda'(\theta) &> \frac{\Lambda(\theta)}{\theta} \geq \frac{\log\left(\int_{s=0}^{M} e^{\theta s}\, dF(s) + e^{\theta M}\, F^c(M)\right)}{\theta} \\ &\geq \frac{\log(e^{\theta M}\, F^c(M))}{\theta} = \frac{\theta M - \int_{s=0}^{M} h(s)\, ds}{\theta} \\ &= M - \frac{1}{\theta} \underbrace{\int_{s=0}^{M} h(s)\, ds}_{=-\log F^c(M)} \end{aligned}$$

But for unbounded τ we have $F^c > 0$ always and thus $-\log F^c(M) < \infty$. Therefore

$$\lim_{\theta\to\infty} \Lambda'(\theta) \quad > \quad M$$

$\square$ 2.7.4

For $\mathcal{D}(\Lambda) = (-\infty, L_C(h))$ for $L_C(h) < \infty$ we have already seen that Λ' can't be bounded and that Λ is essentially smooth / steep. With 2.7.1 we again get $\Lambda'(\mathcal{D}(\Lambda)) = \mathbb{R}_{>0}$.

2.8 Generalisations

Up to now we have assumed that inter event times could be arbitrarily small. We now investigate the more general inter event times $\tau \in (a, b)$ for $a \geq 0$ and $b \leq \infty$.

Claim 2.8.1. *If $\tau \in (a, b)$ with $a \geq 0$ and $b \leq \infty$ and $\Lambda'(\mathcal{D}(\Lambda)) = (a, b)$ then $\mathcal{D}(\Lambda^*) \subseteq [a, b]$ and for $x \in (a, b)$ an optimising $\theta = \theta(x) \in \mathbb{R}$ exists such that*

$$\Lambda^*(x) = \theta x - \Lambda(\theta) \qquad , \qquad \theta = (\Lambda')^{-1}(x).$$

From the assumption of no-point mass on boundary points we get $\mathcal{D}(\Lambda^*) = (a, b)$.

Proof of 2.8.1:

$$\begin{aligned} \log \mathbb{E}[e^{\theta\tau}] &= a\theta + \log \mathbb{E}[e^{\theta(\tau-a)}] \\ \Lambda^*(a) &= \sup_{\theta\in\mathbb{R}} \theta a - \log \mathbb{E}[e^{\theta\tau}] = \sup_{\theta\in\mathbb{R}} -\log \mathbb{E}[e^{\theta(\tau-a)}] \end{aligned}$$

and thus $\Lambda^*(a) = \infty$ if τ has no point mass on a. Similarly $\Lambda^*(b) = \infty$. $\square$ 2.8.1

Remark 2.8.2. *This affects the associated* Γ^* *in the following way:*

- *If* $a > 0$, $b < \infty$ *then* $\mathcal{D}(\Gamma^*) = (\frac{1}{b}, \frac{1}{a})$.
- *If* $a = 0$, $b < \infty$ *then* $\mathcal{D}(\Gamma^*) = (\frac{1}{b}, \infty)$.
- *If* $a > 0$, $b = \infty$ *then* $\mathcal{D}(\Gamma^*) = (0, \frac{1}{a})$ *or* $\mathcal{D}(\Gamma^*) = [0, \frac{1}{a})$ *depending on* $L_C(h)$.

And $\mathcal{D}(\Gamma^*) = [0, \frac{1}{a})$ *has* $\lim_{x\to 0} \Gamma^{*\prime}(x) = \infty$.

Chapter 3

The counting process

In this chapter we move from inter event times to renewal counting processes that we will define through these inter event times. For exponentially distributed inter event times the resulting renewal counting process will be the Markovian Poisson process. For the more general inter event times of chapter 2 we will obtain non-Markovian processes. In this chapter we will argue that many implications of the Markov-property that are desirable when developing a large deviation principle on the process level can be obtained for renewal counting processes as well.

The Poisson process as a Markovian renewal counting process has stationary increments starting at any fixed moment. This does not hold true for the general renewal counting process. For the general case of a counting process with iid inter event times we show how to construct the associated process with stationary increments in section 3.2: Choosing a different first inter event time will be enough and we will make the definition of renewal counting process such as to include these kinds of processes. In section 3.3 we calculate the lmgf of the general renewal counting process and we find that it relates to the lmgf of the inter event times as $\Gamma(\theta) = -\Lambda^{-1}(-\theta)$. This has been proved in 1994 by Peter Glynn and Ward Whitt for more general counting processes and by other means (cf [10]).

The fourth section is about exponential equivalence of different kinds of counting processes: A Markovian renewal counting process has a first inter event time with the same distribution as all following inter event times *and* has stationary increments. For a general renewal counting process these are mutually exclusive properties. Luckily, in terms of large deviations the process with stationary increments and the process with the first and all following inter event times identically distributed are indistinguishable. In a similar way we have independence of increments over disjoint intervals for the Markovian counting process while this does not hold for the general renewal counting process. We will construct a process that does have independent increments when observed over finitely many, fixed, disjoint intervals. We say the process restarts at the beginning of each interval. The restarted process will be indistinguishable from the general renewal counting process in terms of large deviations.

The exponential change of measure is classical in large deviation theory. It is central to the proof of the large deviation principle for Jackson networks of Irina Ignatiouk-Robert [12]. Preparing for a similar approach we develop the exponential change of measure for the single (undelayed) renewal counting process of general inter event times in the sixth section. We derive the change of measure for the counting process from exponentially twisting its inter event times. We will see that this does not change the process' renewal property and we stay in the same class of processes. This way we know about the process' expected behaviour under the changed measure.

Before starting the first section we remind the reader of assumptions and some notation from the previous chapter:

- Inter event times are denoted by τ and variations of it with distribution function F and density (wrt Lebesgue measure) f and lmgf Λ with an open domain (cf 2.2.2 , 2.2.13).

- For a non-deterministic inter event time τ the hazard function is denoted h and existence of a positive (possibly infinite) Cesaro mean $L_C(h)$ is assumed (cf 2.4.2).

- All Λ and Γ are strictly convex as soon as the associated inter event time is not deterministic.

- For the domain of Λ we have $\mathcal{D}(\Lambda) = (-\infty, L_C(h))$ (cf 2.4.6) and if $L_C(h) = \infty$ we say that τ is LD-bounded (cf 2.5.4) and that F^c decays super exponentially (cf 2.4.7).

3.1 Introducing the counting process

With a sequence of inter event times $\tau_1, \tau_2, \ldots$ we associate a process that for every $t \geq 0$ counts how many events have happened in $[0, t]$. This process should at $t = 0$ take the value 0 and increase by 1 at each occurrence of an event. Since events are separated by inter event times $\tau_1, \tau_2, \ldots$ events happen at $\tau_1, \tau_1 + \tau_2, \sum_{l=1}^{3} \tau_l, \ldots$.

Definition 3.1.1 (Counting process)**.** *Let $\tau_1, \tau_2, \ldots$ be inter event times. Then the associated counting process N is defined as*

$$N_t \le k \qquad \Leftrightarrow \qquad t < \sum_{l=1}^{k+1} \tau_l$$

for any $t \in \mathbb{R}_{\ge 0}$.

Claim 3.1.2. *N is piecewise constant and jumps at $\left(\sum_{l=1}^{k} \tau_l \,;\, k \in \mathbb{N}\right)$.*

Proof of 3.1.2:

$$N(t) = k \Leftrightarrow \left\{ \begin{array}{c} N(t) \le k \\ N(t) \not\le k-1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{c} S_{k+1} > t \\ S_k \le t \end{array} \right\} \Leftrightarrow t \in [S_k, S_{k+1}) \qquad \square_{3.1.2}$$

Claim 3.1.3. *If $\tau_1, \tau_2, \ldots$ are independent and fall under assumption 2.2.2 then all jump-sizes of N are equal to 1 and $N(t) < \infty$ for any $t \in \mathbb{R}_{\ge 0}$.*

Proof of 3.1.3: Jump sizes are equal to 1 because intervals $[S_k, S_{k+1})$ are a.s. not degenerate due to $\mathbb{P}(\tau_k = 0) = 0$ (which is necessary for 2.2.2 to hold). For finiteness:

$$\begin{aligned} N(t) = \infty \;\Leftrightarrow\; N(t) \le k \text{ for no } k \in \mathbb{N} \;&\Leftrightarrow\; t < S_{k+1} \text{ for no } k \in \mathbb{N} \\ &\Leftrightarrow\; t \ge S_{k+1}\, \forall k \in \mathbb{N}. \end{aligned}$$

But the partial sums process grows a.s. unboundedly. $\square_{3.1.3}$

Definition 3.1.4 (Renewal counting process, rcp)**.** *If inter event times $\tau_1, \tau_2, \ldots$ of a counting process N are independent and $\tau_2, \tau_3, \ldots$ are identically distributed then N is a renewal counting process or - abbreviated - a rcp. If*

- *τ_1, τ_2 are identically distributed then N is an undelayed rcp.*
- *τ_1, τ_2 are not identically distributed then N is a delayed rcp.*

For a rcp we say that τ_1 is the initial inter event time and τ_2 is a typical inter event time.

Of the delayed renewal counting processes one is particularly important.

Definition 3.1.5 ($\tilde{N}$)**.** *The delayed renewal counting process with typical inter event time distribution F and initial inter event time distribution $\tilde{F}$, the $\sim$-transform of F defined in 2.1.2, is denoted $\tilde{N}$.*

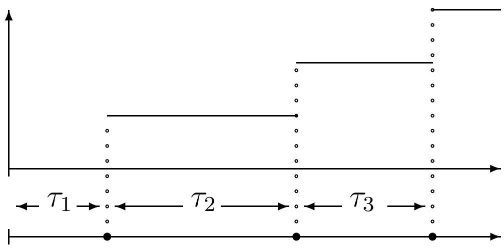


Figure 3.1: A sequence of inter event times and the associated counting process

The sequence of inter event times and the associated counting process contain the same information. We give a graphical representation of both in figure 3.1.

We want to describe the distribution of the discrete valued delayed renewal counting process N. For a convenient way to describe the mass function we make the following

Definition 3.1.6 (Convolution).

- *For two density functions f and g on $\mathbb{R}_{\geq 0}$ their convolution $f * g$ is defined as $f * g(t) = \int_{s=0}^{t} f(t-s)\, g(s)\, ds$.*
- *For two distribution functions F and G on $\mathbb{R}_{\geq 0}$ their convolution $F \circledast G$ is defined as $F \circledast G(t) = \int_{s=0}^{t} F(t-s)\, dG(s)$.*
- *For $F : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ not necessarily a distribution function the convolution is again $F \circledast G = \int_{s=0}^{t} F(t-s)\, dG(s)$.*

Convoluting f or F with itself we write $f * f = f^{*2}$ and $F \circledast F = F^{\circledast 2}$. More generally f^{*k} is the k-time convolution recursively defined for $k \geq 2$ and for a uniform notation we write $f^{*1} = f$ and $f^{*0} = \delta_0$. Analogously for F and $F^{\circledast k}$ with $F^{\circledast 0} = \mathbb{1}_{[0,\infty)}$.

Defined this way the convolution of densities and distributions match: $\int_{s=0}^{t} f * g(s)\, ds \; = \; F \circledast G(t)$.

Remark 3.1.7. *Convolutions of densities and distribution functions describe densities and distribution functions of sums of independent inter event times: e.g. if τ, σ are independent and have densities f and g then $\tau + \sigma$ has density $f * g$.*

Claim 3.1.8 (Mass function for delayed rcp)**.** *If N is a delayed rcp with initial inter event time distribution G and typical inter event time distribution F then*

$$\mathbb{P}(N_t = k) \;=\; \begin{cases} G^c(t) & \textit{if } k = 0 \\ F^c \circledast G \circledast F^{\circledast(k-1)}(t) & \textit{if } k \geq 1 \end{cases}$$

Proof of 3.1.8: Let $\tau_1, \tau_2, \ldots$ be the inter event times of N, τ_1 with distribution function G and $\tau_2, \tau_3, \ldots$ each with distribution function F. For $k \geq 1$ the distribution function of $S_k = \sum_{l=1}^{k} \tau_l$ is the convolution $G \circledast F^{\circledast(k-1)}$. We calculate the mass function:

$$\begin{aligned} \mathbb{P}(N_t = 0) &= \mathbb{P}(\tau_1 > t) = G^c(t) \\ \mathbb{P}(N_t = k) &= \int_{s=0}^{t} \mathbb{P}(N_t = k \mid S_k = s)\, dG \circledast F^{\circledast(k-1)}(s) \\ &= \int_{s=0}^{t} \mathbb{P}(\tau_k > t - s)\, dG \circledast F^{\circledast(k-1)}(s) \\ &= F^c \circledast G \circledast F^{\circledast(k-1)}(t) \end{aligned} \tag{3.1}$$

and for the last line we need to allow the convolution of the not-a-distribution function F^c.

□
3.1.8

Corollary 3.1.9. *For $G = F$ we get the following mass function for the undelayed counting process*

$$P(N_t = k) \;=\; F^c \circledast F^{\circledast k}(t) \;=\; \mathbb{E}[\tau]\, \tilde{f} * f^{*k}(t) \qquad (k \in \mathbb{N})$$

and $\sum_{k=0}^{\infty} F^c \circledast F^{\circledast k}(t) = 1$ or $\sum_{k=0}^{\infty} \tilde{f} * f^{*k}(t) = \frac{1}{\mathbb{E}[\tau]}$.

We have introduced exponential transformation of distribution functions in definition 2.3.1 and we want to know how this relates to convolutions.

Claim 3.1.10. *Let Λ be the lmgf of interevent time τ with density f and distribution function F. If $k \geq 1$ then*

(1) $\log \int_{x=0}^{\infty} e^{\beta x}\, dF^{\circledast k}(x) = k\Lambda(\beta)$*;*

(2) $\left(f^{*k}\right)_\beta = \left(f_\beta\right)^{*k}$ *or equivalently* $\left(F^{\circledast k}\right)_\beta = \left(F_\beta\right)^{\circledast k}$*.*

Remark 3.1.11.

- *Claim 3.1.10 and the definition of the exponential transform 2.3.1 imply that $\left(f^{*k}\right)_\beta(x) = e^{\beta x - k\Lambda(\beta)}\, f^{*k}(x)$ or equivalently $\left(F^{\circledast k}\right)_\beta(x) = \int_{s=0}^{x} e^{\beta s - k\Lambda(\beta)}\, dF^{\circledast k}(s)$.*

- *As a consequence we can omit parentheses:* $F_\beta^{\circledast k} = (F_\beta)^{\circledast k} = (F^{\circledast k})_\beta$.

Proof of 3.1.10: (1) is clear from $F^{\circledast k}$ being the distribution function of $\sum_{l=1}^k \tau_l$ for iid $\tau_1, \ldots, \tau_k$. We prove (2) inductively for the densities. Initially for $k = 1$

$$(2), k = 1 \qquad (f_\beta)^{*1}(x) = f_\beta(x) \; = \; (f^{*1})_\beta(x).$$

And generally for $k \geq 1$: If (2) holds for k and (1) generally holds then (2) holds for $k+1$, too.

$$\begin{aligned}
(f_\beta)^{*(k+1)}(x) &= \big((f_\beta)^{*k} * f_\beta\big)(x) \; = \; \int_{s=0}^x (f_\beta)^{*k}(x-s) f_\beta(s)\, ds \\
&\overset{(2)}{=} \int_{s=0}^x (f^{*k})_\beta(x-s) f_\beta(s)\, ds \\
&\overset{(1)}{=} \int_{s=0}^x f^{*k}(x-s)\, e^{\beta(x-s)-k\Lambda(\beta)}\, f(s)\, e^{\beta s-\Lambda(\beta)}\, ds \\
&= \int_{s=0}^x f^{*k}(x-s) f(s)\, ds\, e^{\beta x-(k+1)\Lambda(\beta)} \\
&\overset{(1)}{=} (f^{*(k+1)})_\beta(x)
\end{aligned}$$

$\square$
3.1.10

3.1.1 Joint distributions

At any fixed time t it might be interesting to know how much time passed since the last event and how far the next event is.

Definition 3.1.12 (Age, residual lifetime, spread)**.** *For a counting process N with associated partial sums of inter event times S we define*

- *the age at time t as $B(N,t) = t - S_{N(t)}$.*
- *the residual lifetime at time t as $C(N,t) = S_{N(t)+1} - t$.*
- *the spread at time t as $B(N,t)+C(N,t)$. This is the length of the inter event time covering t.*

We sometimes abbreviate $B(N,t)$ and $C(N,t)$ by omitting N and write $B(t)$ and $C(t)$ instead.

Figure 3.2 illustrates the definition.

Remark 3.1.13. *From construction of the process we immediately get: Given the age, the residual life is independent of the state.*

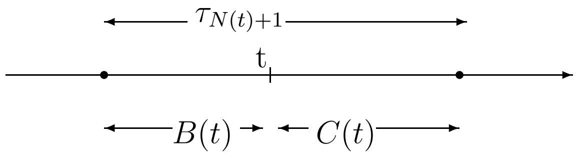


Figure 3.2: Age, residual lifetime, and spread

Claim 3.1.14. *$P(C(t) \leq x \,|\, B(t)) = F_{+B(t)}(x)$ with $F_{+\cdot}$ of definition 2.1.7.*

Proof of 3.1.14: First note that

$$\begin{aligned} B(t) = a \quad &\Leftrightarrow \quad \left\{\Delta N(t-a) = 1\,,\, \tau_{N(t)+1} > a\right\} \\ \{B(t) = a\,,\, N(t) = k\} \quad &\Leftrightarrow \quad \{S_k = t-a\,,\, \tau_{k+1} > a\} \end{aligned}$$

and apply this in the following.

$$\begin{aligned} &\mathbb{P}(\tau_{N(t)+1} > x + a \,|\, B(t) = a) \\ = \quad & \sum_{k=0}^{\infty} \mathbb{P}(\tau_{N(t)+1} > x + a \,|\, N(t) = k\,,\, B(t) = a)\, \mathbb{P}(N(t) = k) \\ = \quad & \sum_{k=0}^{\infty} \mathbb{P}(\tau_{k+1} > x + a \,|\, N(t) = k\,,\, S_k = t-a\,,\, \tau_{k+1} > a)\, \mathbb{P}(N(t) = k) \\ = \quad & \sum_{k=0}^{\infty} \mathbb{P}(\tau_{k+1} > x + a \,|\, S_k = t-a\,,\, \tau_{k+1} > a)\, \mathbb{P}(N(t) = k) \\ \overset{S_k \perp\!\!\!\perp \tau_{k+1}}{=} \quad & \sum_{k=0}^{\infty} \underbrace{\mathbb{P}(\tau_{k+1} > x + a \,|\, \tau_{k+1} > a)}_{=\mathbb{P}(\tau_1 > x+a \,|\, \tau_1 > a)}\, \mathbb{P}(N(t) = k) \\ = \quad & \mathbb{P}(\tau_1 > x + a \,|\, \tau_1 > a) \sum_{k=0}^{\infty} \mathbb{P}(N(t) = k) \\ = \quad & F^c_{+a}(x) \end{aligned}$$

$\square$
3.1.14

Corollary 3.1.15. *By symmetry of age and residual lifetime similarly to 3.1.14 holds: $P(B(t) \leq x \,|\, C(t)) = F_{+C(t)}(x)$ with $F_{+\cdot}$ of definition 2.1.7.*

Let N be a delayed renewal counting process with initial inter event time distribution function G and typical inter event time distribution function F. We give the joint distribution of state and residual lifetime.

Claim 3.1.16. *The joint mass of $(N(t), C(t))$, the state and residual lifetime at t, on $\{k\} \times (s, \infty)$ for $k \in \mathbb{N}$ and $s \geq 0$ is*

$$\mathbb{P}(N(t) = k, C(t) > s) = \begin{cases} G^c(t+s) & k = 0 \\ \int_{r=0}^{t} F^c(t+s-r)\, dG \circledast F^{\circledast(k-1)}(r) & k \geq 1. \end{cases}$$

The joint mass of $(N(t), B(t))$, the state and the age at t, on $\{k\} \times [0, s]$ for $k \in \mathbb{N}$ and $s \geq 0$ is

$$\mathbb{P}(N(t) = k\,,\, B(t) \leq s) = \begin{cases} 0 & k = 0\,,\, s < t \\ G^c(t) & k = 0\,,\, s \geq t \\ \begin{pmatrix} F^c \circledast G \circledast F^{\circledast(k-1)}(t) \\ -F^c \circledast G \circledast F^{\circledast(k-1)}(t-s) \end{pmatrix} & k > 0\,,\, s \leq t \\ F^c \circledast G \circledast F^{\circledast(k-1)}(t) & k > 0\,,\, s > t. \end{cases}$$

Proof of 3.1.16: Let $S_k = \sum_{l=1}^{k} \tau_l$.

$$\begin{aligned} &\mathbb{P}(N(t) = k, C(t) > s) \\ &= \mathbb{P}(S_k \leq t < S_{k+1}\,,\, S_{k+1} > t+s) \\ &= \mathbb{P}(S_k \leq t\,,\, \underbrace{S_{k+1}}_{=S_k+\tau_{k+1}} > t+s) \\ &= \int_{r=0}^{t} P(r \leq t\,,\, r + \tau_{k+1} > t+s \,|\, S_k = r)\, dG \circledast F^{\circledast(k-1)}(r) \\ &= \int_{r=0}^{t} P(\tau_{k+1} > t+s-r)\, dG \circledast F^{\circledast(k-1)}(r) \\ &= \int_{r=0}^{t} F^c(t+s-r)\, dG \circledast F^{\circledast(k-1)}(r) \end{aligned}$$

Now the joint mass of state and age and first for $k = 0$.

$$P(N(t) = 0\,,\, B(t) \leq s) \;=\; \begin{cases} 0 & , s < t \\ G^c(t) & , s \geq t \end{cases}$$

Now for the general $k \geq 1$ and $s \leq t$

$$
\begin{aligned}
&P(N(t) = k\,,\, B(t) \leq s) \\
&= P(\underbrace{N(t) = k}_{t \in [S_k\,,\, S_k + \tau_{k+1})}, t - S_{N(t)} \leq s) \\
&= P(S_k \in [t-s, t]\,,\, \tau_{k+1} > t - S_k) \\
&= \int_{r=t-s}^{t} P(r \in [t-s,t]\,,\, \tau_{k+1} > t - r \,|\, S_k = r)\, g * f^{*(k-1)}(r)\, dr \\
&= \int_{r=t-s}^{t} F^c(t-r)\, g * f^{*(k-1)}(r)\, dr \qquad (3.2) \\
&= \int_{r=t-s}^{t} F^c(t-r)\, d\, G \circledast F^{\circledast(k-1)}(r) \\
&= F^c \circledast G \circledast F^{\circledast(k-1)}(t) - F^c \circledast G \circledast F^{\circledast(k-1)}(t-s)
\end{aligned}
$$

whereas for $k \geq 1\,,\, t < s$

$$
\begin{aligned}
&\{\underbrace{N(t) = k}_{\Rightarrow B(t) < t}, B(t) \leq s\} &=& \{N(t) = k\} \\
\Rightarrow \quad &P(N(t) = k\,,\, B(t) \leq s) &=& F^c \circledast G \circledast F^{\circledast(k-1)}(t) \qquad \text{(cf. 3.1)}
\end{aligned}
$$

□
3.1.16

Note that the joint distribution of state and age has a density for $k \geq 1$ fixed and $s \leq t$ (easily seen from (3.2)):

$$s \mapsto F^c(s)\, g * f^{*(k-1)}(t-s) \qquad (3.3)$$

Corollary 3.1.17. *The undelayed renewal counting process with inter event time distribution function F has as joint distribution of state and age*

$$
\mathbb{P}(N(t) = k\,,\, B(t) \leq s) = \begin{cases} 0 & k = 0\,,\, s < t \\ F^c(t) & k = 0\,,\, s \geq t \\ F^c \circledast F^{\circledast k}(t) - F^c \circledast F^{\circledast k}(t-s) & k > 0\,,\, s \leq t \\ F^c \circledast F^{\circledast k}(t) & k > 0\,,\, s > t \end{cases}
$$

with the following density for $k > 0\,,\, s \in [0,t]$

$$s \mapsto \frac{d}{ds}\mathbb{P}(N(t) = k\,,\, B(t) \leq s) = F^c(s)\, f^{*k}(t-s)$$

3.2 Stationary increments

Consider the delayed renewal counting process $\tilde{N}$ defined in 3.1.5. In this section we prove that $\tilde{N}$ has strictly / strongly stationary increments.

The mass function of $\tilde{N}$ at fixed t is derived from claim 3.1.8 as

$$k \mapsto \mathbb{P}(\tilde{N}(t) = k) = \begin{cases} \tilde{F}^c(t) & k = 0 \\ F^c \circledast \tilde{F} \circledast F^{\circledast(k-1)}(t) & k \geq 1 \end{cases} \tag{3.4}$$

The following is weak stationarity of increments.

Claim 3.2.1. *$t \mapsto \mathbb{E}[\tilde{N}(t)]$ is linear with slope $\frac{1}{\mathbb{E}[\tau]}$.*

Proof of 3.2.1:

$$\mathbb{E}[\tilde{N}(t)] \;=\; \sum_{k=1}^{\infty} \mathbb{P}(\tilde{N}(t) \geq k) \;=\; \sum_{k=1}^{\infty} \mathbb{P}(\tilde{\tau} + \sum_{l=2}^{k} \tau_l \leq t) \;=\; \sum_{k=1}^{\infty} \tilde{F} \circledast F^{\circledast(k-1)}(t)$$

and for the derivative

$$\begin{aligned} \frac{d}{dt}\tilde{F} \circledast F^{\circledast k}(t) \;&=\; \underbrace{\tilde{F}(0)}_{=0}\, F^{\circledast k}(t) + \int_{s=0}^{t} \tilde{f}(t-s)\, dF^{\circledast k}(s) \;=\; \frac{1}{\mathbb{E}[\tau]} F^c \circledast F^{\circledast k}(t) \\ &\overset{3.1.9}{=} \frac{1}{\mathbb{E}[\tau]}\, \mathbb{P}(N(t) = k) \end{aligned}$$

and thus

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[\tilde{N}(t)] \;&=\; \frac{d}{dt}\big(\tilde{F}(t) + \sum_{k=1}^{\infty} \tilde{F} \circledast F^{\circledast k}(t)\big) \\ &=\; \tilde{f}(t) + \sum_{k=1}^{\infty} \frac{1}{E[\tau]} \mathbb{P}(N(t) = k) \\ &=\; \frac{1}{\mathbb{E}[\tau]}\big(\underbrace{F^c(t)}_{=\mathbb{P}(N(t)=0)} + P(N(t) \geq 1)\big) \;=\; \frac{1}{\mathbb{E}[\tau]} \end{aligned}$$

□ 3.2.1

We will continue by arguing for strong stationarity of increments of the delayed counting process $\tilde{N}$.

Lemma 3.2.2. *For the delayed renewal counting process $\tilde{N}$ with initial inter event time $\tilde{\tau}$ holds: $\mathcal{L}(C(\tilde{N}, t)) = \mathcal{L}(\tilde{\tau})$ for any t.*

Proof of 3.2.2: We work with the joint distribution of $\tilde{N}(t)$ and $C(\tilde{N},t)$ of 3.1.16. We have to set $g = \tilde{f}$.

$$\begin{aligned}
\mathbb{P}(C(t) > s) &= \mathbb{P}(N(t) = 0\,,\, C(t) > s) + \sum_{k=1}^{\infty} \mathbb{P}(N(t) = k\,,\, C(t) > s) \\
&= \tilde{F}^c(t+s) + \int_{r=0}^{t} F^c(t+s-r) \sum_{k=1}^{\infty} \tilde{f} * f^{*(k-1)}(r)\, dr \\
&= \int_{r=t+s}^{\infty} \frac{F^c(r)}{\mathbb{E}[\tau]}\, dr + \int_{r=0}^{t} F^c(t+s-r) \underbrace{\sum_{k=0}^{\infty} \tilde{f} * f^{*k}(r)}_{=\frac{1}{\mathbb{E}[\tau]} \text{ by } 3.1.9}\, dr \\
&= \int_{r=t+s}^{\infty} \frac{F^c(r)}{\mathbb{E}[\tau]}\, dr + \int_{r=s}^{t+s} \frac{F^c(r)}{\mathbb{E}[\tau]}\, dr \\
&= \tilde{F}^c(s) \;=\; \mathbb{P}(\tilde{\tau} > s) \qquad \underset{3.2.2}{\square}
\end{aligned}$$

From this strong stationarity of increments of $\tilde{N}$ is immediate:

Claim 3.2.3. *Let $N' : s \mapsto \tilde{N}(t+s) - \tilde{N}(t)$ for $s \geq 0$. Then $\mathcal{L}(\tilde{N}) = \mathcal{L}(N')$.*

Proof of 3.2.3: The process N' is piecewise constant and if it jumps at s then

$$N'(s) - N'(s-) = \tilde{N}(t+s) - \tilde{N}(t+s-) = 1.$$

Typical inter event times of N' are typical inter event times of $\tilde{N}$, so they are independent. Thus N' is a delayed renewal counting process. The initial inter event time distribution of N' is $\tau_1' = C(\tilde{N},t)$ and we have just seen that $\mathcal{L}(C(\tilde{N},t)) = \mathcal{L}(\tilde{\tau})$. So inter event times of N' have the same distribution as those of $\tilde{N}$ and the distributions of the associated counting processes coincide. $\underset{3.2.3}{\square}$

As a graphical presentation of the proof of 3.2.3 consider figure 3.3. Inter event times of $\tilde{N}$ are on the upper and of N' on the lower timeline.

For a different kind of proof of strong stationarity of increments of $\tilde{N}$ see [26] (ch. 2.16 especially theorem 17. on p. 112).

We can similarly argue for reversibility of the counting process with stationary increments. The following lemma is a preparation.

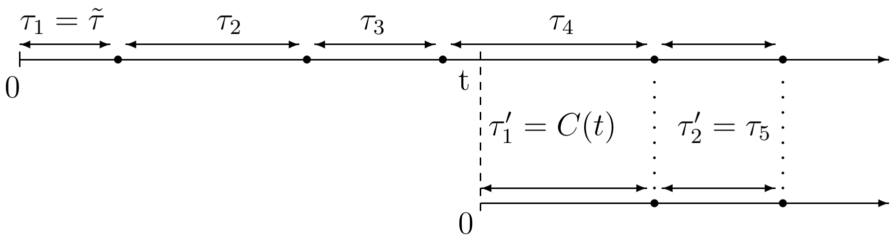


Figure 3.3: Inter event times of $\tilde{N}$ and N' of claim 3.2.3

Lemma 3.2.4.

$$\mathbb{P}(B(t) \leq x) \;=\; \begin{cases} 1 & \text{if } x \geq t \\ \tilde{F}(x) & \text{if } x < t \end{cases}$$

Proof of 3.2.4: We apply the conditional distribution 3.1.15 and the knowledge of the distribution of $C(\tilde{N}, t)$ from 3.2.2.

$$\begin{aligned}
\mathbb{P}(B(t) \leq x) &= \int_{y=0}^{\infty} \underbrace{\mathbb{P}(B(t) \leq x \,|\, C(t) = y)}_{=F_{+y}(x)} \tilde{f}(y)\, dy \\
&= \int_{y=0}^{\infty} (1 - \frac{F^c(x+y)}{F^c(y)})\, \frac{F^c(y)}{\mathbb{E}[\tau]}\, dy \\
&= 1 - \int_{y=0}^{\infty} \frac{F^c(x+y)}{\mathbb{E}[\tau]}\, dy \\
&= 1 - \int_{y=x}^{\infty} \tilde{f}(y)\, dy \\
&= \tilde{F}(x)
\end{aligned}$$

For the point mass: $P(B(\tilde{N}, t) = t) = \mathbb{P}(\tilde{\tau} > t) = \tilde{F}^c(t)$ which suits just fine.

$\square$
3.2.4

Claim 3.2.5 (Reversibility of increments)**.** *For $\tilde{N}$ and fixed $t > 0$ let N'' : $s \mapsto \tilde{N}(t) - \tilde{N}(t-s)$ for $s \in [0, t]$. Then $\mathcal{L}(\tilde{N}\big|_{[0,t]}) = \mathcal{L}(N'')$.*

Proof of 3.2.5: $B(\tilde{N}, t)$ has the required distribution by lemma 3.2.4. Denote inter event times of N'' by $\tau''_1, \tau''_2, \ldots$. The remaining proof is in figure 3.4.

$\square$
3.2.5

3.3 Lmgf for the undelayed rcp

The logarithmic moment generating function is an essential in classical large deviation theory.

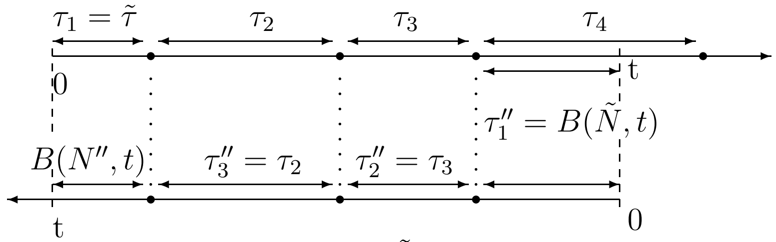


Figure 3.4: Inter event times of $\tilde{N}$ and the reversed process N''

Claim 3.3.1. *Let N be the undelayed rcp with inter event time distribution F with lmgf Λ (cf 2.2.1) and associated Γ (cf 2.2.7). Then for any $\theta \in \mathbb{R}$*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \Gamma(\theta).$$

Proof of 3.3.1: We calculate exactly. For $\theta \in \mathbb{R}$ set $\rho = -\Gamma(\theta)$ which is equivalent to $-\Lambda(\rho) = \theta$ by definition of Γ.

$$\begin{aligned}
\mathbb{E}[e^{\theta N_t}] &\overset{3.1.9}{=} F^c(t) + \sum_{k=1}^{\infty} e^{\theta k}\, F^c \circledast F^{\circledast k}(t) \\
&= F^c(t) + \sum_{k=1}^{\infty} e^{\theta k} \int_{s=0}^{t} F^c(t-s) f^{*k}(s)\, ds \\
&= F^c(t) + \int_{s=0}^{t} F^c(t-s) \sum_{k=1}^{\infty} e^{\theta k}\, f^{*k}(s)\, ds \\
&\overset{\theta=-\Lambda(\rho)}{=} F^c(t) + \int_{s=0}^{t} F^c(t-s) \sum_{k=1}^{\infty} e^{-\rho s}\, e^{\rho s - k\Lambda(\rho)}\, f^{*k}(s)\, ds \\
&\overset{3.1.10}{=} F^c(t) + \int_{s=0}^{t} F^c(t-s)\, e^{-\rho s} \sum_{k=1}^{\infty} f_\rho^{*k}(s)\, ds \\
&= F^c(t) + e^{-t\rho} \int_{s=0}^{t} F^c(t-s)\, e^{\rho(t-s)} \underbrace{\Big(\frac{1}{\mathbb{E}^{(\rho)}[\tau]} - \widetilde{(f_\rho)}(s) \Big)}_{= \frac{1}{\mathbb{E}^{(\rho)}[\tau]} - \frac{F_\rho^c(s)}{\mathbb{E}^{(\rho)}[\tau]}} ds \\
&= F^c(t) + e^{-t\rho} \int_{s=0}^{t} F^c(t-s)\, e^{\rho(t-s)} \frac{F_\rho(s)}{\mathbb{E}^{(\rho)}[\tau]}\, ds
\end{aligned}$$

And the integral in the last line converges to some value in $(0, \infty)$. Which implies that there is no exponential decay or growth and the integral does not contribute to the exponential rate of the expectation.

$$
\begin{aligned}
\int_{s=0}^{t} F^c(t-s)\, e^{\rho(t-s)}\, F_\rho(s)\, ds &\overset{r=t-s}{=} \int_{r=0}^{t} \underbrace{F^c(r)}_{=\mathbb{E}[\tau]\tilde{f}(r)}\; e^{r\rho}\, F_\rho(t-r)\, dr \\
&= \int_{r=0}^{t} (\tilde{f})_\rho(r)\, e^{\tilde{\Lambda}(\rho)}\, F_\rho(t-r)\, dr \\
&= e^{\tilde{\Lambda}(\rho)} \int_{r=0}^{t} F_\rho(t-r)\, d(\tilde{F})_\rho(r) \\
&= e^{\tilde{\Lambda}(\rho)}\, F_\rho \circledast (\tilde{F})_\rho(t) \\
&\to e^{\tilde{\Lambda}(\rho)} \quad (\text{as } t \to \infty)
\end{aligned}
$$

Thus under the exponential scaling

$$
\begin{aligned}
\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] &\leq \max\{\lim_{t\to\infty} \frac{1}{t} \log F^c(t)\,,\, -\rho + \limsup_{t\to\infty} \frac{1}{t} \log \frac{e^{\tilde{\Lambda}(\rho)}}{\mathbb{E}^{(\rho)}[\tau]}\} \\
&= \max\{-L_C(h)\,,\, -\rho\} \;=\; -\rho
\end{aligned}
$$

where we applied 2.4.7 (for decay rate of F^c as $L_C(h)$) and $\rho \in \mathcal{D}(\Lambda) = (-\infty, L_C(h)))$ and

$$
\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \;\geq\; \liminf_{t\to\infty} \frac{1}{t} \log \left(e^{-t\rho} \frac{e^{\tilde{\Lambda}(\rho)}}{\mathbb{E}^{(\rho)}[\tau]}\right) \;=\; -\rho
$$

Since $\rho = -\Gamma(\theta)$ the claim is proved. $\square$ 3.3.1

3.4 Exponential equivalence for cps

Doing Large Deviations for counting processes it will be convenient to make some small alterations from the original process from time to time. In this section we describe these alterations and prove that they do not affect the Large Deviation behaviour of the process.

Definition 3.4.1 (Scaling)**.** *For N a renewal counting process define*

$$
N_n \,:\, \mathbb{R}_{\geq 0} \to \mathbb{R} \quad , \quad t \mapsto \frac{1}{n} N(nt).
$$

For every $n \in \mathbb{N}$ and $T > 0$ the scaled counting processes on $[0, T]$ are elements of $D([0, T], \mathbb{R})$. The sup-norm of any N_n restricted to $[0, T]$ is finite and for two counting processes the sup-norm induced distance will be finite:

$$
||N_n - N_n'|| \leq ||N_n|| + ||N_n'|| = \frac{1}{n}(N(nT) + N'(nT)) < \infty
$$

Definition 3.4.2 (Exponential equivalence in $\big(D([0,T],\mathbb{R}),||.||\big)$). *The sequences of processes $(Y_n\,;\,n\in\mathbb{N})$ and $(Z_n\,;\,n\in\mathbb{N})$ in $D([0,T],\mathbb{R})$ equipped with the sup-norm are exponentially equivalent if for each $n\in\mathbb{N}$ there is a coupling $(\check{Y}_n,\check{Z}_n)$ of (Y_n,Z_n) such that the sequence of sup-norm distances $(||\check{Y}_n-\check{Z}_n||\,;\,n\in\mathbb{N})$ decays super exponentially: For any $\delta>0$*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(||\check{Y}_n-\check{Z}_n||>\delta)=-\infty.$$

If the sequence of processes used in the exponential equivalence is obvious and for example relates to a counting process under the scaling 3.4.1, then we may say that two processes N , N' are exponentially equivalent when indeed we should be saying that $(N_n;n\in\mathbb{N})$ and $(N'_n;n\in\mathbb{N})$ are exponentially equivalent.

3.4.1 Initial inter event time

Here we argue for the exponential equivalence of counting processes that differ only in the distribution of the time to the first event if these distributions are exponentially equivalent. This will imply exponential equivalence of the undelayed rcp and the associated rcp with stationary increments.

Definition 3.4.3 (N,N^σ).

- *N is an undelayed renewal counting process and F is the distribution function of each inter event time. τ is the first inter event time of N.*
- *N^σ is a delayed renewal counting process with typical inter event time distribution F and initial inter event time σ with distribution function G.*

Claim 3.4.4. *If the initial inter event times τ,σ of N,N^σ are exponentially equivalent then the counting processes N,N^σ are exponentially equivalent.*

To prove this we have to give a coupling $\check{N},\check{N}^\sigma$ of N,N^σ such that the difference $||\check{N}_n-\check{N}^\sigma_n||$ decays faster than exponentially in n.

Definition 3.4.5 (Coupled $\check{N},\check{N}^\sigma$). *Let $\{U,\tau_2,\tau_3,\dots\}$ be independent random variables where U is uniform on $[0,1]$ and $\tau_2,\tau_3,\dots$ have density f and distribution function F. Define the counting process $\check{N}$ through its inter event times $F^{-1}(U),\tau_2,\tau_3,\dots$ and the counting process $\check{N}^\sigma$ through its inter event times $G^{-1}(U),\tau_2,\tau_3,\dots$ (cf def 3.1.1 of a counting process).*

Figure 3.5 is a graphical representation of the coupled $\check{N}$ and $\check{N}^\sigma$.

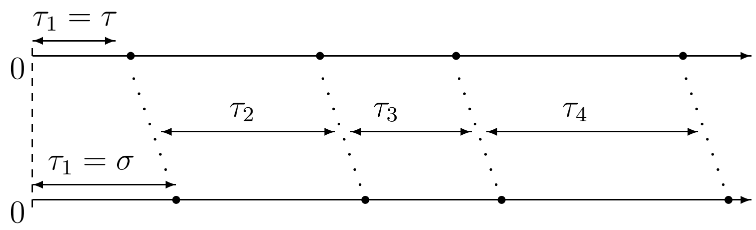


Figure 3.5: Inter event times of coupled $\check{N}$ and $\check{N}^\sigma$

Claim 3.4.6 (Marginal distributions)**.** *$\mathcal{L}(N) = \mathcal{L}(\check{N})$ and $\mathcal{L}(N^\sigma) = \mathcal{L}(\check{N}^\sigma)$.*

Proof of 3.4.6: Inter event times $F^{-1}(U), \tau_2, \tau_3, \dots$ and $G^{-1}(U), \tau_2, \tau_3, \dots$ are independent because $U, \tau_2, \tau_3, \dots$ are. Thus $\check{N}$ and $\check{N}_\sigma$ are rcps. All inter event times of $\check{N}$ have distribution function F: the $\tau_2, \tau_3, \dots$ by definition and $F^{-1}(U)$ by the quantile coupling and (2.6). Thus $\check{N}$ is an undelayed renewal counting process. By the definition 3.1.1 of a counting process its distribution is determined by the distribution of its inter event times: so distributions of $\check{N}$ and N coincide. Similarly N^σ is a delayed renewal counting process and its first inter event time has the required distribution: $\mathcal{L}(\sigma) = \mathcal{L}(G^{-1}(U))$ again by the quantile coupling and (2.6). $\square$ 3.4.6

Lemma 3.4.7. *For coupled $\check{N}, \check{N}^\sigma$ of 3.4.5: $\check{N}^\sigma(t) = \check{N}(t + \tau - \sigma)$ (with $\check{N}(s) = 0$ for $s \leq 0$).*

Proof of 3.4.7: Set $S_k = \tau + \sum_{l=2}^k \tau_k$ and $S_k^\sigma = \sigma + \sum_{l=2}^k \tau_k$. For $k \geq 1$

$$\begin{aligned} \check{N}(t + \tau - \sigma) = k \quad &\Leftrightarrow \quad S_k \leq t + \tau - \sigma < S_{k+1} \\ &\overset{-\tau+\sigma}{\Leftrightarrow} \quad S_k^\sigma \leq t < S_{k+1}^\sigma \quad \Leftrightarrow \quad \check{N}^\sigma(t) = k \end{aligned}$$

and for $k = 0$

$$\check{N}(t + \tau - \sigma) = 0 \;\Leftrightarrow\; t + \tau - \sigma < \tau \;\Leftrightarrow\; t < \sigma \;\Leftrightarrow\; \check{N}^\sigma(t) = 0$$

Consider the case $\tau < \sigma$ and let $0 < s < -\tau + \sigma$. Since $s < \sigma$ we have $N^\sigma(s) = 0$ and $N(s + \tau - \sigma) = 0$ since $s + \tau - \sigma < 0$. $\square$ 3.4.7

Proof of 3.4.4: The claim is proved if we can prove

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(||\check{N}_n - \check{N}_n^\sigma|| > a) = -\infty \tag{3.5}$$

for $\check{N}$, $\check{N}^\sigma$ of definition 3.4.5 and any $a > 0$. We introduce a small parameter

$\delta > 0$.

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in[0,T]} |\check{N}_n^\sigma(t) - \check{N}_n(t)| > a) \\
&= \mathbb{P}(\sup_{t\in[0,T]} |N(nt+\tau-\sigma) - N(nt)| > n\,a\,,\, |\tau-\sigma| < \delta n) \\
&\quad + \mathbb{P}(\sup_{t\in[0,T]} |N(nt+\tau-\sigma) - N(nt)| > n\,a\,,\, |\tau-\sigma| \geq \delta n)
\end{aligned}$$

(We write N instead of $\check{N}$ again since $\mathcal{L}(N) = \mathcal{L}(\check{N})$). From the quantile coupling of τ and σ (such that they are exponentially equivalent, cf claim 2.5.7) we already know that the second probability decays superexponentially. We further investigate the first event. Under the condition of a small distance $|\tau - \sigma|$ and in light of the monotonicity of N

$$\begin{aligned}
&|N(nt+\tau-\sigma) - N(nt)| \\
&\leq \max\{|N(nt+n\delta) - N(nt)|\,,\, |N((nt-n\delta)^+) - N(nt)|\} \\
&= \max\{N(nt+n\delta) - N(nt)\,,\, N(nt) - N((nt-n\delta)^+)\}
\end{aligned}$$

and taking the sup over all t makes the max{...} unnecessary.

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in[0,T]} |\check{N}^\sigma(nt) - \check{N}(nt)| > n\,a\,,\, |\tau-\sigma| < \delta n) \\
&\leq \mathbb{P}(\sup_{t\in[0,T]} N(nt+n\delta) - N(nt) > n\,a\,,\, |\tau-\sigma| < \delta n)
\end{aligned}$$

And

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in[0,T]} |\check{N}_n^\sigma(t) - \check{N}_n(t)| > a) \\
&\leq \mathbb{P}(\sup_{t\in[0,T]} N(nt+n\delta) - N(nt) > n\,a) + \mathbb{P}(|\tau-\sigma| \geq n\delta)
\end{aligned}$$

Exponential equivalence as stated in 3.5 has now become equivalent to

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sup_{t\in[0,T]} N_n(t+\delta) - N_n(t) > a) = -\infty \qquad (3.6)$$

Now fix a $\delta < \frac{a}{\lambda}$ (and small relative to T) and divide the interval $[0,T]$ into many intervals of size δ. If there is t such that $N_n(t+\delta) - N_n(t) > a$ then this t will be in one of these intervals.

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in[0,T]} N_n(t+\delta) - N_n(t) \geq a) \\
&= \mathbb{P}(\max_{m=0,\ldots,\lfloor\frac{T}{\delta}\rfloor} \sup_{t\in[m\delta,(m+1)\delta]} N_n(t+\delta) - N_n(t) \geq a) \\
&\leq \sum_{m=0}^{\lfloor\frac{T}{\delta}\rfloor} \mathbb{P}(\sup_{t\in[m\delta,(m+1)\delta]} N_n(t+\delta) - N_n(t) \geq a) \\
&= \sum_{m=0}^{\lfloor\frac{T}{\delta}\rfloor} \mathbb{P}(C(mn\delta) < n\delta\,, \sup_{t\in[m\delta,(m+1)\delta]} N_n(t+\delta) - N_n(t) \geq a) \\
&= \sum_{m=0}^{\lfloor\frac{T}{\delta}\rfloor} \mathbb{P}(\sup_{t\in[m\delta,(m+1)\delta]} \underbrace{\underbrace{N_n(t+\delta)}_{\leq N_n((m+2)\delta)} - \underbrace{N_n(t)}_{\substack{\geq N_n(m\delta) \\ =\frac{1}{n}(N(nm\delta+C(nm\delta))-1)}}}_{\leq\frac{1}{n}+N_n^\tau(2\delta)} \geq a \\
&\qquad\qquad | \, C(mn\delta) < n\delta) \, \underbrace{\mathbb{P}(C(mn\delta) < n\delta)}_{\leq 1} \\
&\leq \sum_{m=0}^{\lfloor\frac{T}{\delta}\rfloor} \mathbb{P}(N_n^\tau(2\delta) \geq a - \frac{1}{n})
\end{aligned}$$

where we introduced N^τ as an undelayed rcp. to bounds increments of N_n (It should be a stochastic domination only since we don't want to do another explicit coupling again ...). Since N defined in 3.4.3 was undelayed too, we can even omit the τ.

$$\begin{aligned}
\mathbb{P}(\sup_{t\in[0,T]} N_n(t+\delta) - N_n(t) \geq a) &\leq (\lfloor\frac{T}{\delta}\rfloor + 1)\mathbb{P}(N_n(2\delta) > a - \frac{1}{n}) \\
&\overset{a'<a}{\leq} (\lfloor\frac{T}{\delta}\rfloor + 1)\mathbb{P}(N_n(2\delta) > a')
\end{aligned}$$

With in the last inequality $n > \frac{1}{a-a'}$ which is required for $a - \frac{1}{n} > a'$ to hold for $a' < a$. To apply the scaling we move from counting processes to partial sums since we already have the large deviation for their mean (cf [5]

Cramér's theorem 2.2.3).

$$\begin{aligned} \frac{1}{n}\log\mathbb{P}(N_n(2\delta)\ge a') &= \frac{1}{n}\log\mathbb{P}(N(n2\delta)\ge na') \\ &= \frac{1}{n}\log\mathbb{P}(S_{na'}\le n2\delta) \\ &= a'\frac{1}{na'}\log\mathbb{P}(\frac{1}{na'}S_{na'}\le\frac{2\delta}{a'}) \overset{n\to\infty}{\to} -a'\,\Lambda^*(\frac{2\delta}{a'}) \end{aligned}$$

As we have arbitrary $a' < a$ and Λ^* is continuous we now have checked (3.6):

$$\begin{aligned} &\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sup_{t\in[0,T]} N_n(t+\delta)-N_n(t) > a) \\ &\le\ -\lim_{\delta\to 0} a\,\Lambda^*(\frac{2\delta}{a})\ =\ -a\Lambda^*(0) = -\infty \end{aligned}$$

Note that $\Lambda^*(0)=\infty$ is a consequence of the no-point-mass at zero property of assumption 2.2.2 and that $\lim_{\delta\to 0}\Lambda^*(\delta)=\infty$ follows from lower semi continuity of Λ^*.

□ 3.4.4

The following generalisation of 3.4.4 is immediate.

Corollary 3.4.8. *Any two renewal counting processes with the same typical inter event time distribution and exponentially equivalent initial inter event times are exponentially equivalent.*

3.4.2 Independence of increments

For the Markovian renewal counting process the state at a fixed time $s\in[0,T]$ and future increments $(N(s)\,,\,N(T)-N(s))$ are independent. This does not hold for non-deterministic inter event distributions different from the exponential.

The non-independence of $N(s)$ and $N(T)-N(s)$ for a general rcp N is through the age $B(s)$ at time s and affects the initial distribution of the process of increments on $[s,T]$. We have seen in the last section that a single initial inter event time may be changed without affecting the large deviation behaviour of the process. This will be a tool when replacing a renewal counting process by a similar process with a certain independence of increments.

Claim 3.4.9. *Let N be a rcp and N_n the associated scaled process. Fix a $k\in\mathbb{N}$ and fix $0<s_1<\cdots<s_k<T$. Then there is a sequence of scaled counting process $(N'_n\,;\,n\in\mathbb{N})$ such that $(N_n\,;\ n\in\mathbb{N})\,,\,(N'_n\,;\,n\in\mathbb{N})$ are*

exponentially equivalent in $(D([0,T],\mathbb{R})\,,\,||\cdot||)$ *and processes of increments over disjoint intervals for* N'_n

$$\left\{\,\big(N'_n(t)-N'_n(s_m)\,;\,t\in[s_m,s_{m+1}]\big)\,|\,m=0,\ldots,k\,\right\}$$

(with $s_0=0\,,\,s_{k+1}=T$*) are independent.*

Since rcps with the same Cesaro mean for their initial inter event time distribution are exponentially equivalent we can pick one for the proof of 3.4.9 and we pick $\tilde{N}$ of definition 3.1.5 and section 3.2. The proof will be by induction and we start with the base case $k=1$.

Claim 3.4.10. *Let* $\tilde{N}$ *be the delayed rcp with stationary increments and for some* $T>0$ *let* $s\in[0,T]$*. Then there is a sequence of counting process* $(N'_n\,;\,n\in\mathbb{N})$ *such that* $(N'_n\,;\,n\in\mathbb{N})$ *and* $(\tilde{N}_n\,;\,n\in\mathbb{N})$ *are exponentially equivalent in* $(D([0,T],\mathbb{R})\,,\,||\cdot||)$ *and for each* $n\in\mathbb{N}$

$$(N'_n(t)\,;\,t\in[0,s])\,,\,(N'_n(t)-N'_n(s)\,;\,t\in[s,T])$$

are independent.

Definition 3.4.11 (Restarted process $N^{\mathrm{re},s}$)**.** *Given* $s>0$ *and rcps* $N,N^{(1)}$ *define*

$$N^{re,s}\;:\;\mathbb{R}_{\geq 0}\to\mathbb{R}\quad,\quad t\mapsto\begin{cases}N(t) & \text{if } t\leq s\\ N(s)+N^{(1)}(t-s) & \text{if } t>s.\end{cases}$$

Note that $N^{\mathrm{re},s}$ is a counting process and if $N\big|_{[0,s]},N^{(1)}$ are independent, then $N^{\mathrm{re},s}(s)$ and increments of $N^{\mathrm{re},s}$ after s are independent.

We want a scaling for the restarted process $N^{\mathrm{re},s}$ that also scales the associated epoche s.

Definition 3.4.12 (Scaled restarted process $N_n^{\mathrm{re},s}$)**.** *Fix* $s>0$*. Let* N *be a renewal counting process and* $(N^{re,ns}\,;\,n\in\mathbb{N})$ *a sequence of restarted process associated with* $ns,N,N^{(1)}$*. Then define*

$$N_n^{re,s}\;:\;\mathbb{R}_{\geq 0}\to\mathbb{R}\quad,\quad t\mapsto\frac{1}{n}N^{re,ns}(nt).$$

For some fixed s we now give a coupling of $\tilde{N}$ and a restarted process $N^{\mathrm{re},s}$.

Definition 3.4.13 (Coupling M, M')**.** *Let $s > 0$ be fixed. Let $U_1, U_2, \tau_2', \tau_2'', \tau_3', \tau_3'' \ldots$ be independent random variables where U_1, U_2 are uniform on $[0,1]$ and all τ_i', τ_i'' have distribution function F. Let $F_{+\cdot}$ be associated with F (cf definition 2.1.7) and define $(B, C) = (\tilde{F}^{-1}(U_1), F_{+B}^{-1}(U_2))$ through the quantile coupling (cf definition 2.5.2).*

Finally define M as the counting process associated with the following sequence of inter event times:

- $B(M, s) = B$
- $C(M, s) = C$
- *inter event times after $s + C$ are $\tau_2', \tau_3', \ldots$*
- *inter event times before $s - B$ are $\tau_2'', \tau_3'', \ldots$*

And define M' as the counting process with the following inter event times

- *inter event time covering s: $C(M', s) = G^{-1}(U_2)$, $B(M', s) = B(M, s)$*
- *all other inter event times the same as in M.*

Figure 3.6 is a graphical representation of the definition of inter event times for the coupled M, M'.

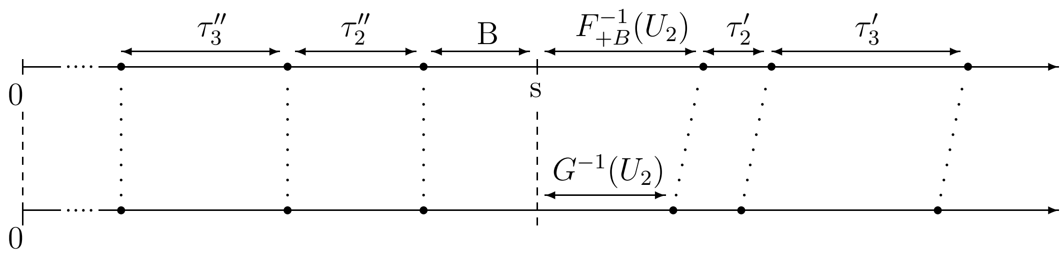


Figure 3.6: Inter event times of coupled M (top line) and M' (bottom line)

Claim 3.4.14. $\mathcal{L}(M) = \mathcal{L}(\tilde{N})$.

Proof of 3.4.14: The construction of M is a combination of those in section 3.2: We construct the counting process starting in t for $s \geq t$ as in 3.2.3 and also construct its past as the reverse of a stationary counting process in 3.2.5.

We make sure that $C = F^{-1}_{+B}(U_2)$ has the correct marginal distribution: density $\tilde{f}$.

$$
\begin{aligned}
\mathbb{P}(C \le x) &= \int_{b=0}^{\infty} \mathbb{P}(\underbrace{C \le x}_{\Leftrightarrow F^{-1}_{+b}(U_2) \le x} \mid B = b)\, \tilde{f}(b)\, db = \int_{b=0}^{\infty} F_{+b}(x)\, \tilde{f}(b)\, db \\
&\overset{2.1.7}{=} \int_{b=0}^{\infty} \frac{F(x+b) - F(b)}{F^c(b)} \frac{F^c(b)}{\mathbb{E}[\tau]}\, db = \int_{b=0}^{\infty} \frac{F(x+b) - F(b)}{\mathbb{E}[\tau]}\, db \\
\frac{d}{dx}\mathbb{P}(C \le x) &= \frac{\int_{b=0}^{\infty} f(x+b)\, db}{\mathbb{E}[\tau]} = \frac{F^c(x)}{\mathbb{E}[\tau]} = \tilde{f}(x)
\end{aligned}
$$

So we can really use B as the age $B(\tilde{N}, t)$ and C as the residual lifetime $C(\tilde{N}, t)$. The inter event times of M and $\tilde{N}$ that cover t have the same distribution. Since all other inter event times of of M and $\tilde{N}$ have the same distribution, too, the distributions of the counting processes coincide. $\square$ 3.4.14

Claim 3.4.15. $\mathcal{L}(M') = \mathcal{L}(N^{re,s})$

Proof of 3.4.15: We have the process of increments of M' after s with inter event times $G^{-1}(U_2), \tau'_2, \tau'_3, \dots$ which is independent of M' on $[0, s]$ and we can identify increments of M' after s - distribution wise - with $N^{(1)}$ of the definition of $N^{\mathrm{re},s}$. $\square$ 3.4.15

Proof of claim 3.4.10 with $N'_n = N^{\mathrm{re},s}_n$: For each $n \in \mathbb{N}$ let M, M' be the coupled processes defined above associated with ns. Then for fixed n

$$
\begin{aligned}
\sup_{t \in [0, nT]} |M(t) - M'(t)| &= \sup_{t \in [ns, nT]} |M(t) - M'(t)| \\
&= \sup_{t \in [0, n(T-s)]} |N^{F^{-1}_{+B}(U_2)}(t) - N^{G^{-1}(U_2)}(t)|
\end{aligned}
$$

where $N^{\cdot}$ are rcp with typical inter event time distribution function F and indicated initial inter event times: and $F^{-1}_{+b}(U_2)$ and $G^{-1}(U_2) = \tilde{F}^{-1}(U_2)$ have the same Cesaro mean for their hazard functions uniformly in $b \in \mathbb{R}_{\ge 0}$. Thus by section 3.4.1 / corollary 3.4.8 these processes are exponentially equivalent and

$$
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\sup_{t \in [0, n(T-s)]} |N^{F^{-1}_{+B}(U_2)}(t) - N^{G^{-1}(U_2)}(t)| > na) = -\infty
$$

$\square$ 3.4.10

We now prepare for the inductive step:

Definition 3.4.16 (Restarted process $N^{\mathrm{re},(s_1,\dots,s_k)}$)**.** *Given an ordered sequence $0 < s_1 < \cdots < s_k$ and rcps $N, N^{(1)}, \dots, N^{(k)}$ define*

$$N^{re,(s_1,\dots,s_k)} : \mathbb{R}_{\geq 0} \to \mathbb{R}$$
$$t \mapsto \begin{cases} N(t) & \text{if } t \in [0, s_1] \\ N(s_1) + N^{(1)}(t - s_1) & \text{if } t \in (s_1, s_2] \\ \vdots & \vdots \\ N(s_1) + \sum_{m=1}^{k-1} N^{(m)}(s_{m+1} - s_m) + N^{(k)}(t - s_k) & \text{if } t \in (s_k, \infty). \end{cases}$$

Definition 3.4.16 generalises definition 3.4.11 in the number of restarted epochs.

Lemma 3.4.17. *The following is an equivalent recursive definition for 3.4.16: Given an ordered sequence $0 < s_1 < \cdots < s_k$ and rcps $N, N^{(1)}, \dots, N^{(k)}$ define*

- *For $k = 1$ apply definition 3.4.11.*
- *For $k \geq 2$ and an ordered sequence $0 < s_1 < \cdots < s_k < \infty$ and rcps $N, N^{(1)}, \dots, N^{(k)}$ let $N^{re,(s_1,\dots,s_{k-1})}$ be the restarted process associated with $N, N^{(1)}, \dots, N^{(k-1)}$ and the ordered sequence $s_1, \dots, s_{k-1}$*

$$N^{re,(s_1,\dots,s_k)}(t) = \begin{cases} N^{re,(s_1,\dots,s_{k-1})}(t) & \text{if } t \in [0, s_k] \\ N^{re,(s_1,\dots,s_{k-1})}(s_k) + N^{(k)}(t - s_k) & \text{if } t \in (s_k, \infty) \end{cases}$$

The proof of 3.4.17 is just by inductively spelling out the definition. $\square$ 3.4.17

Claim 3.4.18. *Let $\tilde{N}$ be the delayed rcp with stationary increments. Fix $T > 0$ and $k \in \mathbb{N}$ and let $0 < s_1, \dots, s_{k+1} < T$ be an ordered sequence. Assume that for any ordered sequence $0 < s'_1, \dots, s'_k$ the sequences of processes $(\tilde{N}_n; n \in \mathbb{N})$ and $(N_n^{re,(s'_1,\dots,s'_k)}; n \in \mathbb{N})$ are exponentially equivalent in $(D([0,T],\mathbb{R}), ||\cdot||)$. Then $N'_n = N_n^{re,(s_1,\dots,s_{k+1})}$ is such that the sequences of processes $(\tilde{N}_n\,;\, n \in \mathbb{N})$ and $(N'_n\,;\, n \in \mathbb{N})$ are exponentially equivalent in $(D([0,T],\mathbb{R}), ||\cdot||)$ and for each $n \in \mathbb{N}$ the processes of increments*

$$(N'_n(t)\,;\, t \in [0, s_1])\,,\, (N'_n(t) - N'_n(s_1)\,;\, t \in [s_1, s_2])\,, \dots$$
$$\dots\,,\, (N'_n(t) - N'_n(s_{k+1})\,;\, t \in [s_{k+1}, T])$$

are independent.

Proof of 3.4.18: For each $n \in \mathbb{N}$ we couple the sequence of $k+2$ processes

$$\{\tilde{N}_n, \tilde{N}_n^{(1)}, \tilde{N}_n^{(2)}, \ldots, \tilde{N}_n^{(k)}, \tilde{N}_n^{(k+1)}\} \tag{3.7}$$

the following way:

- $\tilde{N}_n, \tilde{N}_n^{(1)}, \tilde{N}_n^{(2)}, \ldots, \tilde{N}_n^{(k)}$ are coupled such that the exponential equivalence of $(\tilde{N}_n\,;\, n \in \mathbb{N})$ and $(N_n^{\mathrm{re},(s_1,\ldots,s_k)}\,;\, n \in \mathbb{N})$ holds when $N_n^{\mathrm{re},(s_1,\ldots,s_k)}$ is build from $\tilde{N}_n$, $\tilde{N}_n^{(1)}$,$\tilde{N}_n^{(2)}, \ldots, \tilde{N}_n^{(k)}$.
- $\tilde{N}_n^{(k+1)}$ and $\tilde{N}_n^{(k)}$ are coupled as in definition 3.4.13 for $s = n(s_{k+1} - s_k)$.

We have for any process N

$$\begin{aligned}
||\tilde{N}_n - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}|| &\leq ||\tilde{N}_n - N|| + ||N - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}|| \\
\mathbb{P}(||\tilde{N}_n - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}|| > \delta) &\leq \mathbb{P}(||\tilde{N}_n - N|| + ||N - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}|| > \delta) \\
&\leq \mathbb{P}(||\tilde{N}_n - N|| > \frac{\delta}{2}) \qquad (3.8) \\
&\qquad + \mathbb{P}(||N - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}|| > \frac{\delta}{2}) \qquad (3.9)
\end{aligned}$$

For fixed $n \in \mathbb{N}$ choose $N = N_n^{\mathrm{re},(s_1,\ldots,s_k)}$ associated with the $\tilde{N}$, $\tilde{N}^{(1)}$, ..., $\tilde{N}^{(k)}$ of (3.7). Then $(\tilde{N}_n\,;\, n \in \mathbb{N})$ and $(\tilde{N}_n^{\mathrm{re},(s_1,\ldots,s_k)}\,;\, n \in \mathbb{N})$ have an LD bounded sup-norm distance by construction and the induction assumption: The probability (3.8) decays super exponentially. For (3.9) first notice that

$$t \in [0, s_{k+1}] \quad : \quad N_n^{\mathrm{re},(s_1,\ldots,s_k)}(t) = N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}(t)$$

since both processes are associated with the same sequence of (3.7). And on $[s_{k+1}, T]$ first apply the representation of $N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}$ of lemma 3.4.17 and then the plain definition of the restarted process.

$$\begin{aligned}
t \geq s_{k+1} \quad : \quad & N_n^{\mathrm{re},(s_1,\ldots,s_k)}(t) - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}(t) \\
&= N_n^{\mathrm{re},(s_1,\ldots,s_k)}(t) - \left(N_n^{\mathrm{re},(s_1,\ldots,s_k)}(s_{k+1}) + \tilde{N}_n^{(k+1)}(t - s_{k+1}) \right) \\
&= \tilde{N}_n^{(k)}(t - s_k) - \tilde{N}_n^{(k)}(s_{k+1} - s_k) - \tilde{N}_n^{(k+1)}(t - s_{k+1})
\end{aligned}$$

and the event of the probability in (3.9) can be simplified:

$$\begin{aligned}
& ||N_n^{\mathrm{re},(s_1,\ldots,s_k)} - N_n^{\mathrm{re},(s_1,\ldots,s_{k+1})}|| \\
= & \sup_{t \in [s_{k+1}, T]} |\tilde{N}_n^{(k)}(t - s_k) - \tilde{N}_n^{(k)}(s_{k+1} - s_k) - \tilde{N}_n^{(k+1)}(t - s_{k+1})| \\
= & \sup_{t \in [s_{k+1} - s_k, T - s_k]} |\tilde{N}_n^{(k)}(t) - \tilde{N}_n^{(k)}(s_{k+1} - s_k) - \tilde{N}_n^{(k+1)}(t + s_k - s_{k+1})| \\
= & \sup_{t \in [s_{k+1} - s_k, T - s_k]} |\tilde{N}_n^{(k)}(t) - \tilde{N}_n^{(k)\mathrm{re},s_{k+1} - s_k}(t)| \\
= & ||\tilde{N}_n^{(k)} - \tilde{N}_n^{(k)\mathrm{re},s_{k+1} - s_k}||_{[0, T - s_{k+1}]}
\end{aligned}$$

where $\tilde{N}^{(k)\mathrm{re},s_{k+1}-s_k}$ is the once restarted process associated with $\tilde{N}^{(k)}$,$\tilde{N}^{(k+1)}$ both of (3.7). By the the base case 3.4.10 of the once restarted process we have exponential equivalence and the probability (3.9) decays super exponentially.

$\square$
3.4.18

Proof of 3.4.9: by induction with the base case 3.4.10 and inductive step 3.4.18. N' is the restarted process of definition 3.4.16.

$\square$
3.4.9

Corollary 3.4.19. *Let N be a rcp and N_n the associated scaled process. Fix $k \in \mathbb{N}$ and fix $0 < s_1 < \cdots < s_k < T$. Then there is a sequence of scaled counting process $(N'_n\,;\, n \in \mathbb{N})$ such that $(N_n\,;\, n \in \mathbb{N})\,,\,(N'_n\,;\, n \in \mathbb{N})$ are exponentially equivalent in $(D([0,T],\mathbb{R}), ||\cdot||)$ and for each $n \in \mathbb{N}$ increments*

$$\{N'_n(s_1)\,,\, N'_n(s_2) - N'_n(s_1)\,, \ldots,\, N'_n(T) - N'_n(s_k)\}$$

of N'_n are independent.

3.4.3 Interpolation

The counting process is piecewise constant and each realisation is a rightcontinuous function with left limits, an element of $D([0,T],\mathbb{R})$. Interpolating N and denoting the interpolated counting process $\hat{N}$ we get an element of $C([0,T],\mathbb{R})$. For any t we have $\hat{N}(t) - N(t) \in [0,1]$ so $||\hat{N} - N||$ is bounded and the counting process and the interpolated counting process are exponentially equivalent.

3.5 Conclusions from previous proofs

In this section we apply the exponential equivalences of the last section to obtain finite dimensional large deviations for the renewal counting process. Subsection 3.5.2 is on the limiting distribution for the scaled renewal counting process being concentrated on the space of continuous functions. Technically it is more a reinterpretation of a statement in the proof of 3.4.4.

3.5.1 Finite dimensional large deviations

We develop finite dimensional large deviation principles for the delayed and undelayed renewal counting process.

Having calculated the lmgf for the undelayed rcp in 3.3 the following theorem is now immediately clear.

Theorem 3.5.1. *Let N be the undelayed renewal counting process with typical inter event time density f and lmgf Λ. Then for $\theta \in \mathbb{R}$*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \Gamma(\theta)$$

with $\Gamma(\theta) = -\Lambda^{-1}(-\theta)$. Furthermore a one-dimensional Large Deviation principle holds: for any $s > 0$ and open set G and closed set F

$$\begin{aligned} -s \inf_{x\in G} \Gamma^*(x) \;\; &\leq \;\; \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n(s) \in G) \\ & \;\; \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n(s) \in F) \leq -s \inf_{x\in F} \Gamma^*(x) \end{aligned}$$

with Γ^* the Fenchel-Legendre transform of Γ.

Our theorem 3.5.1 is theorem 1 in [10].

Proof of 3.5.1: By the Gärtner-Ellis theorem, cf [5] Theorem 2.3.6 on p. 44 or 7.5.3 of the appendix. □

In terms of large deviation we need not distinguish between the delayed and the undelayed counting process any more. The same theorem holds for undelayed rcp that are exponentially equivalent to undelayed N; we have proved exponential equivalence for delayed and undelayed rcp in section 3.4.1.

We state and prove the more general finite dimensional large deviations for the delayed process $\tilde{N}$. We apply exponential equivalence of $\tilde{N}$ and the restarted process. Again the theorem immediately generalises to all other rcp exponentially equivalent to $\tilde{N}$.

Theorem 3.5.2. *Let $\tilde{N}$ be the delayed renewal counting process with stationary increments and inter event time of density f and lmgf Λ. Then for any $k \geq 1$ a k-dimensional Large Deviation principle holds: for any ordered sequence $s_1, \ldots, s_k$ (and for easier notation $s_0 = x_0 = 0$), and open set $G \subseteq \mathbb{R}^k$*

$$-\inf_{x\in G} \sum_{r=1}^{k} (s_r - s_{r-1})\, \Gamma^*(\frac{x_r - x_{r-1}}{s_r - s_{r-1}}) \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\tilde{N}_n(s_1, \ldots, s_k) \in G)$$

while for any closed set $F \subseteq \mathbb{R}^k$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\tilde{N}_n(s_1, \ldots, s_k) \in F) \leq -\inf_{x\in F} \sum_{r=1}^{k} (s_k - s_{k-1})\, \Gamma^*(\frac{x_k - x_{k-1}}{s_k - s_{k-1}})$$

with Γ^* the Fenchel-Legendre transform of Γ.

Proof of 3.5.2: Again by the Gärtner-Ellis theorem. In the scope of this proof abbreviate $N_n^{\mathrm{re}} = N_n^{\mathrm{re},(s_1,\dots,s_{k-1})}$ for the restarted process defined in 3.4.16 associated with $\tilde{N}, \tilde{N}^{(1)}, \dots, \tilde{N}^{(k-1)}$. We calculate the following relatively simple lmgf:

$$
\begin{aligned}
&\mathbb{E}[\exp\{\langle\theta\,,\begin{pmatrix} N_n^{\mathrm{re}}(s_1) \\ \vdots \\ N_n^{\mathrm{re}}(s_k)\end{pmatrix}\rangle\}] \\
\overset{(1)}{=}\quad &\mathbb{E}[\exp\{\langle\begin{pmatrix}\theta_1+\cdots+\theta_k \\ \theta_2+\cdot\cdot+\theta_k \\ \vdots \\ \theta_k\end{pmatrix}, \begin{pmatrix} N_n^{\mathrm{re}}(s_1) \\ N_n^{\mathrm{re}}(s_2)-N_n^{\mathrm{re}}(s_1) \\ \vdots \\ N_n^{\mathrm{re}}(s_k)-N_n^{\mathrm{re}}(s_{k-1})\end{pmatrix}\rangle\}] \\
=\quad &\mathbb{E}[\exp\{\langle\begin{pmatrix}\theta_1+\cdots+\theta_k \\ \theta_2+\cdot\cdot+\theta_k \\ \vdots \\ \theta_k\end{pmatrix}, \begin{pmatrix} \tilde{N}_n(s_1) \\ \tilde{N}_n^{(1)}(s_2-s_1) \\ \vdots \\ \tilde{N}^{(k-1)}(s_k-s_{k-1})\end{pmatrix}\rangle\}] \\
\overset{\tilde{N}^{(0)}:=\tilde{N}}{=}\quad &\prod_{r=1}^{k}\mathbb{E}[\exp\{(\theta_r+\cdots+\theta_k)\,\tilde{N}_n^{(r-1)}(s_r-s_{r-1})\}]
\end{aligned}
$$

where in (1) we applied

$$
(1) \qquad : \qquad \langle\theta, y\rangle = \langle\mathbb{T}^\top\theta, \mathbb{T}^{-1}y\rangle \qquad , \qquad \mathbb{T} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{bmatrix}
$$

and thus

$$
\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\exp\{\langle\theta\,,\begin{pmatrix} N_n^{\mathrm{re}}(s_1) \\ \vdots \\ N_n^{\mathrm{re}}(s_k)\end{pmatrix}\rangle\}] \\
=\quad &\sum_{r=1}^{k}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}[\exp\{(\theta_r+\cdots+\theta_k)\,\tilde{N}_n^{(r-1)}(s_r-s_{r-1})\}] \\
=\quad &\sum_{r=1}^{k}(s_r-s_{r-1})\,\Gamma(\theta_r+\cdots+\theta_k)
\end{aligned}
$$

And the Fenchel-Legendre transform of this lmgf in some $x \in \mathbb{R}^k$ becomes (first apply to the inner product $\langle \theta, x \rangle$ the same transformation as in (1), then change the variable θ)

$$
\begin{aligned}
&\sup_{\theta \in \mathbb{R}^k} \langle \theta, x \rangle - \sum_{r=1}^{k} (s_r - s_{r-1})\, \Gamma(\theta_r + \cdots + \theta_k) \\
= \quad & \sup_{\theta \in \mathbb{R}^k} \langle \begin{pmatrix} \theta_1 + \cdots + \theta_k \\ \theta_2 + \cdot\cdot + \theta_k \\ \vdots \\ \theta_k \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 - x_1 \\ \vdots \\ x_k - x_{k-1} \end{pmatrix} \rangle \\
& \qquad - \sum_{r=1}^{k} (s_r - s_{r-1})\, \Gamma(\theta_r + \cdots + \theta_k) \\
\overset{\xi_r = \theta_r + \cdots + \theta_k}{=} \quad & \sup_{\xi \in \mathbb{R}^k} \langle \xi \,, \begin{pmatrix} x_1 \\ x_2 - x_1 \\ \vdots \\ x_k - x_{k-1} \end{pmatrix} \rangle - \sum_{r=1}^{k} (s_r - s_{r-1})\, \Gamma(\xi_r) \\
= \quad & \sup_{\xi \in \mathbb{R}^k} \sum_{r=1}^{k} \xi_r (x_r - x_{r-1}) - (s_r - s_{r-1})\, \Gamma(\xi_r) \\
= \quad & \sum_{r=1}^{k} (s_r - s_{r-1}) \sup_{\xi \in \mathbb{R}} \left(\frac{x_r - x_{r-1}}{s_r - s_{r-1}}\, \xi - \Gamma(\xi) \right) \\
= \quad & \sum_{r=1}^{k} (s_r - s_{r-1})\, \Gamma^* (\frac{x_r - x_{r-1}}{s_r - s_{r-1}})
\end{aligned}
$$

Now the restarted process is exponentially equivalent to $\tilde{N}$. By 7.2.3 they have the same lmgf. Alternatively from the large deviation principle proved for the restarted process N^{re} follows the large deviation principle for $\tilde{N}$ and with the same rate function by [5] theorem 4.2.13. Thus the claim is proved. $\square$
3.5.2

Corollary 3.5.3. *Any delayed renewal counting process exponentially equivalent to $\tilde{N}$ satisfies a finite dimensional large deviation principle with the rate function of 3.5.2.*

3.5.2 Continuous paths

Recall section 3.4.1 and equation (3.6). This equation is really about the modulus of continuity for N_n. The modulus of continuity is defined as

$$\omega_\delta(f) := \sup\{|f(s) - f(t)| \, : \, s,t \in [0,T]\, , \, |s-t| \le \delta\}$$

and for monotonously increasing $f = N_n$ we get

$$\omega_\delta(N_n) = \sup_{t\in[0,T]} N_n(t+\delta) - N_n(t)$$

Claim 3.5.4. $\lim_{\delta\to 0}\lim_{n\to\infty}\omega_\delta(N_n) = 0$ *a.s.*

Proof of 3.5.4: In the proof of 3.4.4 we showed that (3.6) holds for any fixed $a > 0$. We restate (3.6) in terms of the modulus of continuity:

$$(3.6) \quad \Leftrightarrow \quad \lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log \mathbb{P}(\omega_\delta(N_n) > a) = -\infty$$

This implies that for any fixed $a > 0, M \in \mathbb{R}$

$$\exists(\delta_0, n_0) \ : \ \forall(\delta < \delta_0\, , \, n \ge n_0) \ : \ \mathbb{P}(\omega_\delta(N_n) > a) \ \le \ e^{-nM}$$

but then $\sum_{n=1}^\infty \mathbb{P}(\omega_\delta(N_n) > a) < \infty$ and by the Borel Cantelli lemma $\omega_\delta(N_n) > a$ happens only for finitely many n for almost any fixed path N in the scaling $(N_n; n \in \mathbb{N})$. Thus $\limsup_{n\to\infty}\omega_\delta(N_n) \le a$ for $\delta < \delta_0$ a.s. Thus

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\omega_\delta(N_n) \ \le \ \sup_{\delta<\delta_0}\limsup_{n\to\infty}\omega_\delta(N_n) \ \le \ a$$

Since a was arbitrary the claim is proved. $\square$ 3.5.4

So we know that as $n \to \infty$ the counting process tends to a continuous function. We can look at interpolations $\hat{N}_n$ of N_n and their distribution on $C([0,T],\mathbb{R})$ for finite n. We now know that any limiting distribution will be concentrated on $C([0,T],\mathbb{R})$.

3.6 Change of measure

We start with an intuitive way of defining a change of measure for the process N over compact intervals and then rigorously prove that it is a mean one martingale and has the intuitively clear properties. Since delayed and undelayed counting processes are exponentially equivalent (as long as initial inter event times share the same Cesaro mean of hazard rates) we are free to chose an initial distribution and we will work with the undelayed counting process throughout this section.

Fix a path of the counting process N over $[0, t]$. The path is defined through the now fixed value at t and the fixed inter event times $\tau_1, \ldots, \tau_{N_t}$. The likelihood of the path under the inter event times density f is

$$f(\tau_1) \cdot \cdots \cdot f(\tau_{N_t})\, F^c(B(t))$$

and the likelihood ratio for two different inter event times densities f and g (with distribution function G)

$$\frac{g(\tau_1) \cdot \cdots \cdot g(\tau_{N_t})\, G^c(B(t))}{f(\tau_1) \cdot \cdots \cdot f(\tau_{N_t})\, F^c(B(t))} = \prod_{k=1}^{N_t} \frac{g}{f}(\tau_k) \frac{G^c}{F^c}(B(t))$$

In the case of $g = f_\beta$

$$\prod_{k=1}^{N_t} \frac{f_\beta}{f}(\tau_k) \frac{F^c_\beta}{F^c}(B(t)) = \prod_{k=1}^{N_t} e^{\beta\tau_k - \Lambda(\beta)} \frac{F^c_\beta}{F^c}(B(t)) = e^{\beta(t-B(t)) - N_t\,\Lambda(\beta)} \frac{F^c_\beta}{F^c}(B(t))$$

and this will be our density process.

3.6.1 Martingale property

We prove the martingale property directly relying only on the renewal property of the counting process.

Claim 3.6.1. *For $\beta \in \mathcal{D}(\Lambda)$ and any $T \in \mathbb{R}$ the process*

$$L(\beta, \cdot) \quad : \quad [0, T] \to [0, \infty) \quad , \quad t \mapsto e^{\beta(t-B(t)) - N_t\,\Lambda(\beta)} \frac{F^c_\beta}{F^c}(B(t)) \tag{3.10}$$

is a martingale with respect to the natural filtration of N.

Proof of 3.6.1: $L(\beta, t) < \infty$ always for unbounded τ. If τ is bounded and has no point mass on its least upper bound, b say, then $\mathbb{P}(F^c(B(t)) = 0) = \mathbb{P}(\tau = b) = 0$ and $L(\beta, t) < \infty$ a.s. Changing the density of τ to $f(b) = 0$ makes $L(\beta, t) < \infty$ always.

Measurability wrt the filtration generated by N is immediate: Observing N up to time t we know N_t and $B(t)$ which $L(\beta, t)$ is a function of. We prove integrability in the following claim 3.6.2 by calculating the mean and then innovation in claim 3.6.3. $\square$ 3.6.1

Claim 3.6.2 (Integrability). *$\mathbb{E}[L(\beta, t)] = 1$ for all β and $t \geq 0$.*

Proof of 3.6.2: We need to know about the distribution of the age when $N_t = k$ is fixed. If $k = 0$ then $B(t) = t$ for sure. For specified $k \geq 1$ we have an explicit density for the age (cf 3.1.17)

$$\begin{aligned}
&\mathbb{E}[L(\beta,t)] \\
&= \sum_{k=0}^{\infty} \mathbb{E}[e^{\beta(t-B(t))-N_t\Lambda(\beta)} \frac{F^c_\beta}{F^c}(B(t))\, 1\!\!1_{N_t=k}] \\
&= \frac{F^c_\beta}{F^c}(t)\, F^c(t) \\
&\quad + \sum_{k=1}^{\infty} \int_{x=0}^{t} \mathbb{E}[e^{\beta(t-B(t))-k\Lambda(\beta)} \frac{F^c_\beta}{F^c}(B(t))\, 1\!\!1_{N_t=k} \,|\, B(t) = x]\, F^c(x)\, f^{*k}(t-x)\, dx \\
&= F^c_\beta(t) + \sum_{k=1}^{\infty} \int_{x=0}^{t} \underbrace{e^{\beta(t-x)-k\Lambda(\beta)} \frac{F^c_\beta}{F^c}(x)\, F^c(x)\, f^{*k}(t-x)}_{=F^c_\beta(x) f^{*k}_\beta(t-x)}\, dx \\
&= F^c_\beta(t) + \sum_{k=1}^{\infty} F^c_\beta \circledast F^{\circledast k}_\beta(t) \\
&= \mathbb{P}^{(\beta)}(N_t = 0) + \sum_{k=1}^{\infty} \mathbb{P}^{(\beta)}(N_t = k) \;= 1
\end{aligned}$$

□
3.6.2

Claim 3.6.3 (Innovation). *$\mathbb{E}[L(\beta, t) \,|\, (N_r; r \leq s)] = L(\beta, s)$ for any $\beta \in \mathcal{D}(\Lambda)$ and all t, s with $s \leq t$.*

Proof of 3.6.3: Let us first investigate the conditional expectation of the claim.

$$\begin{aligned}
&\mathbb{E}[L(\beta,t) \,|\, (N_r; r \leq s)] \\
&= \mathbb{E}[\prod_{k=1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(t)) \,|\, (N_r; r \leq s)] \\
&= \prod_{k=1}^{N_s} \exp\{\beta\tau_k - \Lambda(\beta)\} \mathbb{E}[\prod_{k=N_s+1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(t)) \,|\, (N_r; r \leq s)]
\end{aligned}$$

For the remaining conditional expectation we do not need the condition on the whole past of the process: All the information the integrand requires is in the state $N(s)$ in s and the age $B(s)$ in s. Claim 3.6.3 is equivalent to the following

$$\mathbb{E}[\prod_{k=N_s+1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B_t) \,|\, B_s\,,\, N_s] = \frac{F^c_\beta}{F^c}(B_s) \tag{3.11}$$

for any $\beta \in \mathcal{D}(\Lambda)$ and all t, s with $s \leq t$. In (3.11) there is $\tau_{N(s)+1}$, the inter event time covering s and we condition on $B(s)$, the age of the process at time s. We in fact condition on $\tau_{N(s)+1} > b$ for $b = B(s)$ which allows us to apply the distribution function F_{+a} (cf 2.1.7) with $a = B(s)$ to $\tau_{N(s)+1}$. We denote the density of $F_{+B(s)}$ by $f_{+B(s)}$.

Now introduce the indicator $\mathbb{1}_{N_t=N_s+l}$ with $l \in \mathbb{N}$. For $l = 0$ we have an empty product in the following

$$\begin{aligned} &\mathbb{E}[\prod_{k=N_s+1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B_t)\, \mathbb{1}_{N_t=N_s} \,|\, B_s\,,\, N_s] \\ &= \frac{F^c_\beta(t-s+B_s)}{F^c(B_s)} \end{aligned} \tag{3.12}$$

And we have applied $N_t = N_s \Rightarrow B_t = B_s + t - s$. Now for general $l \geq 1$.

$$\begin{aligned} &\mathbb{E}[\prod_{k=N_s+1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(t))\, \mathbb{1}_{N_t=N_s+l} \,|\, B(s)\,,\, N_s] \\ &= \mathbb{E}[e^{\beta\tau_{N_s+1}-\Lambda(\beta)} \prod_{k=N_s+2}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(t))\, \mathbb{1}_{N_t=N_s+l} \,|\, B(s)\,,\, N_s] \end{aligned} \tag{3.13}$$

Now $\tau_{N_s+1} = B_s + C_s$ where B_s is known. We solve for the unknown

$$C_s = \tau_{N_s+1} - B_s$$

which has density f_{+a} of definition 2.1.7 with $a = B_s$ (cf 3.1.14).

$$\begin{aligned} (3.13) &= \int_{x=0}^{\infty} \mathbb{E}[e^{\beta(B_s+C_s)-\Lambda(\beta)} \prod_{k=N_s+2}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B_t) \\ &\qquad \mathbb{1}_{N_t=N_s+l} \,|\, B_s\,,\, N_s\,,\, C_s = x]\, f_{+B_s}(x)\, dx \\ &= \int_{x=0}^{t-s} e^{\beta(B_s+x)-\Lambda(\beta)} \mathbb{E}[\prod_{k=N_s+2}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B_t) \\ &\qquad \mathbb{1}_{N_t=N_s+l} \,|\, B_s\,,\, N_s\,,\, C_s = x] f_{+B_s}(x)\, dx \end{aligned} \tag{3.14}$$

In the remaining conditional expectation we have inter event times τ_{N_s+2}, $\tau_{N_s+3}, \ldots$ that are independent of N_s and identically distributed to $\tau_1, \tau_2, \ldots$. The age $B(t) = B(N,t)$ conditional on N having an event at $s+x$ has the same distribution as the age of the renewal counting process $N^{\tau_{N(s)+2}}$ defined through its inter event times $\tau_{N(s)+2}, \tau_{N(s)+3}, \ldots$ at time $t-(s+x)$:

$$\mathcal{L}\left(B(N,t) | s, N_s, C_s = x \right) = \mathcal{L}\left(B(N^{\tau_{N_s+2}}, t-(s+x)) \,|\, s, N_s \right)$$

And from the just mentioned independence

$$\mathcal{L}\left(B(N^{\tau_{N_s+2}}, t-(s+x)) \,|\, s, N_s \right) \;=\; \mathcal{L}\left(B(N^{\tau_1}, t-(s+x)) \right)$$

The event $N_t = N_s + l$ translates into $1 + N^{\tau_{N_s+2}}(t-(s+x)) = l$ or equivalently $N^{\tau_1}(t-(s+x)) = l-1$. In the product we had $k = N_s, \ldots, N_t$ for $N_t = N_s + l$, so we did take the product over all $l-1$ inter event times of N between $[s + C(s), t]$. These inter event times now become $\tau_1, \ldots \tau_{l-1}$ or $\tau_1, \ldots, \tau_{N^{\tau_1}_{t-(s+x)}}$.

We summarise all just mentioned changes:

$$\begin{aligned}(3.14) \;=\; &\int_{x=0}^{t-s} e^{\beta(B(s)+x)-\Lambda(\beta)} \\ &\qquad \mathbb{E}[\prod_{k=1}^{N^{\tau_1}(t-(s+x))} \exp\{\beta\tau_k - \Lambda(\beta)\} \\ &\qquad \frac{F^c_\beta}{F^c}(B(N^{\tau_1}, t-(s+x)))) \mathbb{1}_{N^{\tau_1}(t-(s+x))=l-1}] \\ &f_{+B(s)}(x)\, dx\end{aligned}$$

and summing expressions (3.13) over $l \geq 1$

$$\begin{aligned}&\sum_{l=1}^{\infty}(3.13) \\ &= \int_{x=0}^{t-s} e^{\beta(B(s)+x)-\Lambda(\beta)} \mathbb{E}[\prod_{k=1}^{N^{\tau_1}(t-(s+x))} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(N^{\tau_1}, t-(s+x)))) \\ &\qquad \underbrace{\sum_{l=1}^{\infty} \mathbb{1}_{N^{\tau_1}(t-(s+x))=l-1}}_{=1}] f_{+B(s)}(x)\, dx\end{aligned}$$

and applying 3.6.2 to N^{τ_1} at $t-(s+x)$

$$\int_{x=0}^{t-s} e^{\beta(B(s)+x)-\Lambda(\beta)} \underbrace{\mathbb{E}[\prod_{k=1}^{N^{\tau_1}(t-(s+x))} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(N^{\tau_1}, t-(s+x))))]}_{=1} f_{+B(s)}(x)\, dx$$

we can simplify

$$
\begin{aligned}
\sum_{l=1}^{\infty}(3.13) &= \int_{x=0}^{t-s} e^{\beta(B(s)+x)-\Lambda(\beta)} \underbrace{f_{+B(s)}(x)}_{=\frac{f(x+B(s))}{F^c(B(s))}} \, dx \\
&= \frac{1}{F^c(B(s))} \int_{x=0}^{t-s} e^{\beta(B(s)+x)-\Lambda(\beta)} f(x+B(s)) \, dx \\
&= \frac{1}{F^c(B(s))} \int_{x=B(s)}^{t-s+B(s)} f_\beta(x) \, dx \\
&= \frac{1}{F^c(B(s))} \left(F_\beta(t-s+B(s)) - F_\beta(B(s))\right) \\
&= \frac{1}{F^c(B(s))} \left(-F^c_\beta(t-s+B(s)) + F^c_\beta(B(s))\right) \\
&= \frac{-F^c_\beta(t-s+B(s))}{F^c(B(s))} + \frac{F^c_\beta(B(s))}{F^c(B(s))}
\end{aligned}
$$

which finally results in

$$
\begin{aligned}
&\mathbb{E}[\prod_{k=N_s+1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(t)) \,|\, B(s)\,,\, N_s] \\
&= \sum_{l=0}^{\infty} E[\prod_{k=N_s+1}^{N_t} \exp\{\beta\tau_k - \Lambda(\beta)\} \frac{F^c_\beta}{F^c}(B(t)) \, \mathbb{1}_{N_t=N_s+l} \,|\, B(s)\,,\, N_s] \\
&= (3.12) + \sum_{l=1}^{\infty}(3.13) \\
&= \frac{F^c_\beta(t-s+B(s))}{F^c(B(s))} + \frac{-F^c_\beta(t-s+B(s))}{F^c(B(s))} + \frac{F^c_\beta(B(s))}{F^c(B(s))} \\
&= \frac{F^c_\beta(B(s))}{F^c(B(s))}
\end{aligned}
$$

and we got (3.11). □ 3.6.3

For an alternative proof cf. [4] proposition 13.3.V on p. 535.

3.6.2 The twisted distribution

We rearrange the non-negative mean one martingale $L(\beta, \cdot)$ a little and change the parameter from β to γ such that $\beta = -\Gamma(\gamma)$. Since $\mathcal{D}(\Gamma) = \mathbb{R}$ and $\Lambda(\mathcal{D}(\Lambda)) = \mathbb{R}$ we can start with any γ and find the matching β. The following is well defined.

Definition 3.6.4. $\mathcal{M}(\gamma,t) := L(-\Gamma(\gamma),t)$ *for* $\gamma \in \mathbb{R}, t \in \mathbb{R}_{\geq 0}$.

We get a simplification in $L(\beta,t)$ from $\Lambda(-\Gamma(\gamma)) = \Lambda(\Lambda^{-1}(-\gamma)) = -\gamma$

$$\begin{aligned} L(-\Gamma(\gamma),t) &\overset{(3.10)}{=} e^{-\Gamma(\gamma)\,(t-B(t))-N_t\Lambda(-\Gamma(\gamma))}\,\frac{F^c_{-\Gamma(\gamma)}}{F^c}(B(t)) \\ &= e^{\gamma N_t - t\Gamma(\gamma)}\,\frac{F^c_{-\Gamma(\gamma)}}{F^c}(B(t))e^{\Gamma(\gamma)B(t)} \end{aligned}$$

and will write

$$\mathcal{M}(\gamma,t) = \exp\{\gamma N_t - t\Gamma(\gamma)\}r(-\Gamma(\gamma),t) \tag{3.15}$$

with

Definition 3.6.5. *For* $t \geq 0$ *and* β *such that* F_β *is well defined (cf 2.3.1) set* $r(\beta,t) := \frac{F^c_\beta}{F^c}(B(t))\,e^{-\beta B(t)}$.

Note that r is measurable by continuity of $(\beta,x) \mapsto \frac{F^c_\beta}{F^c}(x)\,e^{-\beta x}$ and measurability of $t \mapsto B(t)$.

Claim 3.6.6. *The one dimensional distributions of the counting process under the change of measure* $\mathcal{M}(\gamma,\cdot)$ *coincide with the one dimensional distributions of a renewal counting process with inter event times densities* f_β *for* $\beta = -\Gamma(\gamma)$.

Proof of 3.6.6: It holds for $k=0$ and arbitrary $t \geq 0$:

$$\mathbb{P}^{[\gamma]}(N_t = 0) = \mathbb{E}[\mathbb{1}_{N_t=0}\,e^{\gamma 0-(t-t)\Gamma(\gamma)}\frac{F^c_\beta}{F^c}(t)] = \mathbb{E}[\mathbb{1}_{N_t=0}]\,\frac{F^c_\beta}{F^c}(t) = F^c_\beta(t)$$

and for $k \geq 1$ and $t > 0$, applying $\beta = -\Gamma(\gamma) \Leftrightarrow \Lambda(\beta) = -\gamma$:

$$\begin{aligned} &\mathbb{P}^{[\gamma]}(N_t = k) \\ &= \mathbb{E}[\mathbb{1}_{N_t=k}\,e^{\gamma N_t-(t-B(t))\Gamma(\gamma)}\frac{F^c_\beta}{F^c}(B(t))] \\ &\overset{3.3}{=} \int_{s=0}^t \mathbb{E}[\mathbb{1}_{N_t=k}\,e^{\gamma N_t-(t-B(t))\Gamma(\gamma)}\frac{F^c_\beta}{F^c}(B(t))\,|\,B(t)=s, N_t=k]\,F^c(s)f^{*k}(t-s)\,ds \\ &= \int_{s=0}^t e^{\gamma k-(t-s)\Gamma(\gamma)}\frac{F^c_\beta}{F^c}(s)\,F^c(s)f^{*k}(t-s)\,ds \\ &= \int_{s=0}^t \underbrace{e^{\gamma k-(t-s)\Gamma(\gamma)}\;f^{*k}(t-s)}_{=f^{*k}_\beta(t-s)}\,F^c_\beta(s)\,ds \\ &= F^c_\beta \circledast F^{\circledast k}_\beta(t) \qquad \square_{3.6.6} \end{aligned}$$

Where we did not fix the order of convoluting and exponentially twisting due to 3.1.10.

Claim 3.6.7. *The two dimensional distributions of the counting process under the change of measure $\mathcal{M}(\gamma, \cdot)$ coincide with two dimensional distributions of a renewal counting process with inter event times densities f_β for $\beta = -\Gamma(\gamma)$.*

Proof of 3.6.7: First the untwisted counting process. We apply the joint distribution of state and residual lifetime introduced in 3.1.16 for $k \geq 0$, $l \geq 1$. Remember that $F^{\circledast 0} = \mathbb{1}_{(0,\infty)}$.

$$\begin{aligned}
&\mathbb{P}(N_{t_1} = k, N_{t_2} = k + l) \\
&\quad = \int_{r=0}^{t_2-t_1} \mathbb{P}(N_{t_1} = k, N_{t_2} = k + l \,|\, N_{t_1} = k, C(t_1) = r) \\
&\qquad\qquad \int_{p=0}^{t_1} f(t_1 + r - p)\, f^{*k}(p)\, dp\, dr \\
&\quad = \int_{r=0}^{t_2-t_1} \underbrace{\mathbb{P}(N_{t_2} = k + l \,|\, N_{t_1} = k, C(t_1) = r)}_{=\mathbb{P}(N'_{t_2-t_1-r}=l-1)} \int_{p=0}^{t_1} f(t_1 + r - p)\, f^{*k}(p)\, dp\, dr
\end{aligned}$$

With N' counting increments of N after $t_1 + C(t_1)$ and N' an undelayed renewal counting process (cf section 3.2). Now for $k \geq 0, l = 0$.

$$\begin{aligned}
\mathbb{P}(N_{t_1} = k, N_{t_2} = k) &= \mathbb{P}(N_{t_1} = k, C(t_1) > t_2 - t_1) \\
&= \int_{r=0}^{t_1} F^c(t_1 + (t_2 - t_1) - r)\, dF^{\circledast k}(r) \\
&= \int_{r=0}^{t_1} F^c(t_2 - r)\, dF^{\circledast k}(r)
\end{aligned}$$

Now we investigate the twisted process. $k \geq 0, l \geq 1$

$$\begin{aligned}
&\mathbb{P}^{[\gamma]}(N_{t_1} = k, N_{t_2} = k + l) \\
&\quad = \int_{r=0}^{t_2-t_1} \int_{s=0}^{t_1} \mathbb{E}[\mathbb{1}_{N_{t_1}=k}\, \mathbb{1}_{N_{t_2}=k+l}\, e^{\gamma N_{t_2} - (t_2 - B(t_2))\Gamma(\gamma)}\, \frac{F^c_\beta}{F^c}(B(t_2)) \\
&\qquad\qquad | \, N_{t_1} = k, B(t_1) = s, C(t_1) = r] f(t_1 + r - s)\, f^{*k}(s)\, ds\, dr \qquad (3.16)
\end{aligned}$$

Let again N' count increments of N after $t_1 + C(t_1)$. Then the notation changes:

$$\begin{aligned}
N_{t_2} - N_{t_1} \,|\, C(t_1) = r &\;\rightarrow\; 1 + N'_{t_2-t_1-r} \\
N_{t_2} &= N_{t_1} + N_{t_2} - N_{t_1} = N_{t_1} + 1 + N'_{t_2-t_1-r} \\
B(N, t_2) &\;\rightarrow\; B(N', t_2 - t_1 - r) \\
t_2 - B(t_2) &= t_1 + r - s \quad + t_2 - t_1 - r - B(N, t_2) \quad + s \\
&= t_1 + r - s \quad + t_2 - t_1 - r - B(N', t_2 - t_1 - r) \quad + s
\end{aligned}$$

Continue

(3.16)

$$
\begin{aligned}
&= \int_{r=0}^{t_2-t_1} \int_{s=0}^{t_1} \mathbb{E}[1\!\!1_{N'_{t_2-t_1-r}=l-1}\, e^{\gamma N'_{t_2-t_1-r} - (t_2-t_1-r-B(N',t_2-t_1-r))\Gamma(\gamma)} \\
&\qquad\qquad \frac{F^c_\beta}{F^c}(B(N',t_2-t_1-r)) \,|\, N_{t_1}=k, B(t_1)=s, C(t_1)=r] \\
&\qquad\qquad e^{(k+1)\gamma-(t_1+r)\Gamma(\gamma)}\, f(t_1+r-s)\, f^{*k}(s)\, ds\, dr \\
&= \int_{r=0}^{t_2-t_1} \int_{s=0}^{t_1} \mathbb{E}^{[\gamma]}[1\!\!1_{N'_{t_2-t_1-r}=l-1}]\, f_\beta(t_1+r-s)\, f^{*k}_\beta(s)\, ds\, dr
\end{aligned}
$$

to get the twisted analogue of the untwisted finite dimensional distribution of t_1, t_2. The case of $l = 0$ for the twisted process is omitted. □ 3.6.7

Iterating this all finite dimensional distributions of the counting process under the change of measure $\mathcal{M}(\gamma, \cdot)$ are those of a renewal counting process with inter event time densities f_β (for $\beta = -\Gamma(\gamma)$):

Conclusion 3.6.8. *Let N be a renewal counting process with typical inter event time density f and lmgf Λ. Let $t \mapsto \mathcal{M}(t,\gamma)$ be the change of measure process defined in 3.6.4. Under this change of measure the counting process remains renewal and inter event times now have density f_β with $\beta = -\Gamma(\gamma) = \Lambda^{-1}(-\gamma)$.*

We now investigate the change of measure process further.

Lemma 3.6.9. *Let τ be an inter event time with density f and distribution function F and $\beta \in \mathcal{D}(\Lambda)$. Then $\inf_{x:F(x)<1} \frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} > 0$.*

Proof of 3.6.9: The function $x \mapsto \frac{F^c_\beta}{F^c}(x)\, e^{-\beta x}$ is continuous on $\{x \,|\, F(x) < 1\}$ from the existence of the density f for F. If τ is unbounded the infimum is over $\mathbb{R}_{\geq 0}$ otherwise - if $\tau \in (0,b)$, say - over some bounded interval. The function is positive for each fixed x and thus has a positive minimum over any compact interval whithin $\{x \,|\, F(x) < 1\}$. We still need a positive liminf as $x \to \infty$ or $x \to b$. We do a calculation for τ unbounded, but the same holds for bounded τ.

$$
\begin{aligned}
\frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} &= \int_{s=x}^{\infty} f_\beta(s)\, ds\, \frac{1}{F^c(x)} e^{-\beta x} \\
&= \int_{s=x}^{\infty} e^{\beta s - \Lambda(\beta)}\, \frac{f(s)}{F^c(x)}\, ds\, e^{-\beta x} \\
&= \int_{s=x}^{\infty} e^{\beta s}\, \frac{f(s)}{F^c(x)}\, ds\, e^{-\Lambda(\beta)-\beta x} \\
&= \mathbb{E}[e^{\beta(\tau-x)} \,|\, \tau > x] e^{-\Lambda(\beta)}
\end{aligned}
$$

Thus for $\beta > 0$

$$\frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} = \mathbb{E}[e^{\beta(\tau - x)} \,|\, \tau > x] e^{-\Lambda(\beta)} > e^{-\Lambda(\beta)} > 0$$

(the $e^{-\Lambda(\beta)} > 0$ requires $\beta \in \mathcal{D}(\Lambda)$). If on the other hand $\beta < 0$ by an application of Jensen's inequality

$$\frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} = \mathbb{E}[e^{\beta(\tau - x)} \,|\, \tau > x] e^{-\Lambda(\beta)} \;\geq\; e^{\beta \mathbb{E}[\tau - x \,|\, \tau > x]} e^{-\Lambda(\beta)}$$

and

$$\begin{aligned} \epsilon > \frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} \quad &\Rightarrow \quad \epsilon > e^{\beta \mathbb{E}[\tau - x \,|\, \tau > x]} e^{-\Lambda(\beta)} \\ &\Leftrightarrow \quad \frac{1}{\beta} \log\left(\epsilon\, e^{\Lambda(\beta)}\right) < \mathbb{E}[\tau - x \,|\, \tau > x] \end{aligned}$$

and the last inequality holds for arbitrarily small ϵ only if $\limsup_{x\to\infty} \mathbb{E}[\tau - x \,|\, \tau > x] = \infty$. We have assumed in 2.2.13 that this does not happen. $\square$ 3.6.9

We can often do without the cited assumption 2.2.13 but then have to concider different cases:

- For bounded τ and $\beta < 0$ we could have argued directly:

  $$\tau > x \quad \Rightarrow \quad \tau \in (x, b) \quad \overset{\beta<0}{\Rightarrow} \quad \beta\,(\tau - x) > \beta\,(b - x)$$

  Implying $\mathbb{E}[e^{\beta(\tau-x)} \,|\, \tau > x] > e^{\beta(b-x)} > e^{\beta b} > 0$ which makes $\frac{F^c_\beta}{F^c}(x)\, e^{-\beta x}$ strictly positive for all $x \in (0, b)$.

- If for unbounded τ the limit $\lim_{x\to\infty} h(x)$ exists in $(0, \infty]$ then $\lim_{x\to\infty} \mathbb{E}[\tau - x \,|\, \tau > x] = \frac{1}{\lim_{x\to\infty} h(x)} < \infty$. (cf (7.4) of the appendix with $\mathbb{E}[\tau_{+x}]$ the expectation under the distribution F_{+x} (cf definition 2.1.7), so $\mathbb{E}[\tau_{+x}] = \mathbb{E}[\tau - x \,|\, \tau > x]$)

Lemma 3.6.10. *Let τ be an inter event time with density f and distribution function F and $\gamma \in \mathcal{D}(\Lambda)$. Then $\sup_{x:F(x)<1} \frac{F^c_\gamma}{F^c}(x)\, e^{-\gamma x} < \infty$.*

Proof of 3.6.10: We have seen that $\inf_x \frac{F^c_\beta}{F^c}(x) e^{-\beta x} > 0$ for any $\beta \in \mathcal{D}(\Lambda)$. Since $0 \in \mathcal{D}(\Lambda)$ we also have $\alpha \in \mathcal{D}(\Lambda) + \alpha = \mathcal{D}(\Lambda_{-\alpha})$ for any $\alpha \in \mathbb{R}$ (cf 2.5). For $-\alpha \in \mathcal{D}(\Lambda)$ we apply claim 3.6.9 to the inter event time with distribution function $F_{-\alpha}$[1] (and parameter α in the place of β):

$$\begin{aligned} 0 \;&<\; \inf_{x\,:\,F_{-\alpha}(x)<1} \frac{(F_{-\alpha})^c_\alpha}{F^c_{-\alpha}}(x)\, e^{-\alpha x} \;=\; \inf_{x\,:\,F(x)<1} \frac{F^c}{F^c_{-\alpha}}(x)\, e^{-\alpha x} \\ &=\; \frac{1}{\sup_{x\,:\,F(x)<1} \frac{F^c_{-\alpha}}{F^c}(x)\, e^{\alpha x}} \end{aligned}$$

(With $\{x \,|\, F_{-\alpha}(x) < 1\} = \{x \,|\, F(x) < 1\}$ from equivalence of measures / distribution functions under the exponential twist; for $(F_{-\alpha})_\alpha = F$ cf lemma 2.3.3.) So we got the claim with $\gamma = -\alpha$ and our only requirement was $-\alpha \in \mathcal{D}(\Lambda)$. □ 3.6.10

Claim 3.6.11. *$t \mapsto r(\beta, t)$ defined in 3.6.5 is bounded and strictly positive.*

Proof of 3.6.11: For unbounded τ

$$\inf_{t\in\mathbb{R}_{\geq 0}} r(\beta,t) = \inf_{t\in\mathbb{R}_{\geq 0}} \frac{F^c_\beta}{F^c}(B(t))\, e^{-\beta B(t)} \overset{B(t)\in[0,t]\subseteq\mathbb{R}_{\geq 0}}{\geq} \inf_{x\in\mathbb{R}_{\geq 0}} \frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} > 0$$

$$\sup_{t\in\mathbb{R}_{\geq 0}} r(\beta,t) = \sup_{t\in\mathbb{R}_{\geq 0}} \frac{F^c_\beta}{F^c}(B(t))\, e^{-\beta B(t)} \overset{B(t)\in[0,t]\subseteq\mathbb{R}_{\geq 0}}{\leq} \sup_{x\in\mathbb{R}_{\geq 0}} \frac{F^c_\beta}{F^c}(x)\, e^{-\beta x} < \infty$$

while for bounded $\tau \in (0, b)$ we have $B(t) \in (0, b)$ and we apply the infimum and supremum over $(0, b) = \{x \,:\, F(x) < 1\}$.
Positivity of the infimum was proved in 3.6.9 and finiteness of the supremum in 3.6.10. □ 3.6.11

We have now introduced a change of measure $\mathcal{M}(\gamma, \cdot)$ for the renewal counting process and have proved how it affects the inter event times of the counting process. In future notation we will distinguish exponentially twisting inter event times with some parameter β in parentheses: $\mathbb{E}^{(\beta)}[\tau]$. If we explicitly refer to a counting process we write $\mathbb{E}^{(\beta)}[e^{\theta N_t}]$ to express the twist of its inter event times with parameter β; or we might write $\mathbb{E}^{[\gamma]}[e^{\theta N_t}]$ to denote the twisting of the counting process with the change of measure process $\mathcal{M}(\gamma, \cdot)$. Both are the same as soon as $\beta = -\Gamma(\gamma)$.

Further more about notation: Under the exponentially transformed inter event densities with parameter β the mean changes from $\mathbb{E}[\tau] = \Lambda'(0)$ to $\mathbb{E}^{(\beta)}[\tau] = \Lambda'(\beta)$ (cf proof of 2.2.8 and 2.3.4). This corresponds to a change of the renewal point process' rate from $\lambda = \frac{1}{\mathbb{E}[\tau]}$ to $\frac{1}{\mathbb{E}^{(\beta)}[\tau]}$. The following is a

[1] $F_{-\alpha}$ is an exponential transform as defined in 2.1.2 and not of the kind of definition 2.1.7. If it was it would have to be $F_{+(-\alpha)}$ for $\alpha \leq 0$

translation of this fact into the changed parameters ($\beta = -\Gamma(\gamma)$) and more in terms of the counting process.

Definition 3.6.12. *The rate of the point process N under the twist $\mathcal{M}(\gamma, \cdot)$ is denoted $\lambda(\gamma)$ and is defined as $\lambda(\gamma) := \Gamma'(\gamma)$.*

Alternatively we could have defined $\lambda(\gamma) = \frac{1}{\mathbb{E}^{(-\Gamma(\gamma))}[\tau]}$.

To conclude: We have developed our basic tools to be working with renewal counting processes. This allows us to make up for the lost Markov property. We have

- in terms of large deviations identified the undelayed renewal counting process, the renewal counting process with stationary increments, and the counting process with independent increments over disjoint intervals. For the Poisson process all these three properties hold generally and truly - not only in terms of large deviation.
- We have defined an exponential change of measure for the undelayed renewal counting process such that under the changed distribution the process remains renewal. Changing a Markovian jump process this holds true immediately.

Chapter 4

Large deviations of the ren. counting process

In this chapter we develop a sample path large deviation principle for the renewal counting process and we get a rate function in integral form with the integrand, the so-called local rate function, the Fenchel-Legendre transform of the lmgf of the counting process. The integral form fits the claimed closeness to the Markov property while the local rate function reflects the generalised distribution of inter event times.

The importance of this chapter is twofold: On the one hand the sample path large deviations for a one dimensional counting process are much simpler than the sample path large deviations for a higher dimensional process with non independent coordinates and discontinuous statistics. A stochastic process describing a stochastic network will have such undesirable properties. So aiming at the large deviations for a network process we want to develop and test our tools by proving the large deviations of the counting process. On the other hand we will later directly apply the sample path large deviations of the arrival and service process to obtain local large deviations for the network. Thus it is also a matter of completeness that we include the sample path large deviations for the counting process.

The large deviation principle for the counting process is not a new result: it was proved in 1997 by Anatolii Puhalskii and Ward Whitt in the space $D([0,\infty),\mathbb{R})$. Under our assumption 2.2.2 they equip the space with the Skorohod J_1-topology [17], theorem 6.1. They use weak convergence analogs in large deviations and apply an extended contraction principle. It was not immediately clear to us how to apply their techniques in the setting of a stochastic network.

There may be another interesting point in this chapter: The sample path large deviations of the partial sums process can be obtained for LD-bounded iid summands by Mogulskii's theorem, 5.1.2 in [5]. It may be intuitive that the partial sums process of inter event times contains the same information as the counting process constructed from these inter event times and that rare events of one process can be translated into rare events of the other. In recent work Raymond Russel [22] and Mark Rodgers-Lee [20] prove how to obtain a large deviation principle for the counting process from the large deviation principle of the partial sums process and vice versa. This would be one approach to develop the large deviations for the counting process of LD-bounded inter event times.
However, one would still have to develop the large deviations for partial sums processes with not LD-bounded summands for which the Mogulskii theorem does not hold - for example for exponentially distributed summands. In this case, one would rather start from the counting process-side: Sample path large deviations for the Poisson process and other more general Markovian jump processes have been developed in the book of Shwartz and Weiss [23]. We decided to directly develop the sample path large deviations for the counting process and will do so for LD-bounded and not LD-bounded inter event times at the same time. One could then like to transfer the large deviation principle for the renewal counting process with not LD-bounded inter event times back to the partial sums process.

Our approach of development of the sample path large deviations for the renewal counting process starts with local large deviations and calculating exponential decay rates on sup-norm balls around piecewise linear functions with diminishing radii applying an exponential change of measure. In the change of measure we need to find the suitable change of measure that makes the deviating event become the expected behaviour. We get a weak large deviation principle and identify the integral form for the rate function. Then we strengthen the weak to a full large deviation principle by exponential tightness.

Apart from the existence of a large deviation principle and the explicit form of the rate function this chapter contributes in allowing standard large deviation interpretation of rare events: that the rare event happens like a regular (common, non-rare) event under a different distribution, the twisted distribution. And this allows for standard applications as in fast simulation.

4.1 The space

We have introduced the counting process as a stochastic process with piecewise constant paths. Its paths are elements of $D([0,T],\mathbb{Z})$. As we interpolate $\hat{N}$ its paths become elements of $C([0,T],\mathbb{R}) \subseteq D([0,T],\mathbb{R})$. By exponential equivalence of N and $\hat{N}$ we need not distinguish between N and $\hat{N}$ in terms of large deviations when working in the sup-norm induced topology, cf 3.4.3. As before we choose the process that we are most convenient to work with. In this section in terms of large deviation theorems it will be the interpolated process $\hat{N}$ living in the continuous functions $C([0,T],\mathbb{R})$ while for direct calculations of exponentially scaled limits of probabilities it will be the not-interpolated undelayed process N with realisations in $D([0,T],\mathbb{R})$.

The choice of $C([0,T],\mathbb{R})$ equipped with the sup-norm $||\cdot||$ may seem natural in that we have already seen that limiting distributions of scaled counting processes will be in the continuous functions, cf 3.5.2. The sup-norm induces a metric on $C([0,T],\mathbb{R})$ and on $D([0,T],\mathbb{R})$ and thus a topology on these sets of functions. We use the same notation for sup-norm balls in $C([0,T],\mathbb{R})$ as on $D([0,T],\mathbb{R})$:

Definition 4.1.1. *For $\psi \in C([0,T],\mathbb{R}), \epsilon > 0$*

$$\mathcal{U}_\epsilon(\psi) = \{f \,:\, ||f-\psi|| < \epsilon\}$$

where $\mathcal{U}_\epsilon(\psi)$ is understood to be a subset of $D([0,T],\mathbb{R})$ or $C([0,T],\mathbb{R})$.

Remark 4.1.2. *The space $(C([0,T],\mathbb{R}), ||\cdot||)$ of continuous functions over the compact interval $[0,T]$ with the sup-norm is a complete seperable metric space.*

While the scaled process N_n (cf definition 3.4.1) is a function on $[0,T]$ it contains information about N over $[0,nT]$.

4.1.1 A base of the topology

In this section we give a base of the sup-norm induced topology of $C([0,T],\mathbb{R})$ the space of continuous functions. It will consist of sup-norm balls around piecewise linear functions.

Definition 4.1.3 (Piecewise linear functions)**.** *For $J \in \mathbb{N}$ set*

$$\begin{aligned}\mathcal{P}^J \;&=\; \{f \in C([0,T],\mathbb{R}) \,|\, f \text{ linear on } [k\,T\,2^{-J}\,,\,(k+1)\,T\,2^{-J}] \\ &\qquad \text{for } k = 0,\ldots,2^J-1\}.\end{aligned}$$

Claim 4.1.4.

$$\mathcal{J} \;=\; \{\,\mathcal{U}_\epsilon(\psi)\,|\,\epsilon > 0\,,\; J \in \mathbb{N}\,,\; \psi \in \mathcal{P}^J\,\}$$

is a base of the sup-norm induced topology of the real valued continuous functions $C([0,T],\mathbb{R})$.

Proof of 4.1.4: We argue with [21] (chapter 8, section 2, proposition 3, p. 146)

- The base elements cover the set of continuous functions: Let f be a continuous function. Then there is an element of $\mathcal{J}$ that contains f: There is for any $\epsilon > 0$ a $J \in \mathbb{N}$ and a piecewise linear function $\psi \in \mathcal{P}^J$ such that $||f - \psi|| < \epsilon$.
- Let B_1 and B_2 be in $\mathcal{J}$. There is for any $g \in B_1 \cap B_2$ a $B_3 = B_3(g) \in \mathcal{J}$ such that $g \in B_3$ and $B_3 \subseteq B_1 \cap B_2$.

Nothing to say about the first bullet. About the second: Let $B_1\,,\, B_2 \in \mathcal{J}$ be such that they have a non-empty intersection. Let ψ_i be the centre of B_i for $i = 1,2$ and ϵ_i be such that $B_i = \mathcal{U}_{\epsilon_i}(\psi_i)$ and J_i such that $\psi_i \in \mathcal{P}^{J_i}$. Set $J := \max\{J_1, J_2\}$ then $\psi_i \in \mathcal{P}^J$ for $i = 1,2$.

Let g be some continuous function (not necessarily piecewise linear!) in $B_1 \cap B_2$. On any of the 2^J intervals of $[0,T]$ this g has a positive distance to the boundary of the intersection $B_1 \cap B_2\big|_{[T\,k\,2^{-J}\,,\,T\,(k+1)\,2^{-J}]}$.

In terms of pointwise restrictions any function in the intersection $B_1 \cap B_2$ maps $x \in [0,T]$ to y with the following properties.

$$\begin{aligned} x &\in [T\,k\,2^{-J}\,,\,T\,(k+1)\,2^{-J}] \\ y &\in (\max\{\psi_1(x) - \epsilon_1\,,\,\psi_2(x) - \epsilon_2\}\,,\,\min\{\psi_1(x) + \epsilon_1\,,\,\psi_2(x) + \epsilon_2\}) \end{aligned} \tag{4.1}$$

and for fixed g define the strictly positive, continuous functions h, i as

$$\begin{aligned} h(x) &:= g(x) - \max\{\psi_1(x) - \epsilon_1\,,\,\psi_2(x) - \epsilon_2\} \\ i(x) &:= \min\{\psi_1(x) + \epsilon_1\,,\,\psi_2(x) + \epsilon_2\} - g(x) \end{aligned}$$

for $x \in [0,T]$. Each has a strictly positive minimum over $[0,T]$ and we can define $0 < \epsilon := \min_{x\in[0,T]}\{\min\{h(x)\,,\,i(x)\}\}$ and uniformly on $[0,T]$ the distance of $g(x)$ to the bound $B_1 \cap B_2$ is at least ϵ. Now there is $K \in \mathbb{N}$ and a piecewise linear $\phi \in \mathcal{P}^K$ that is $\frac{\epsilon}{4}$-close to g in the sup-norm. The neighbourhood $\mathcal{U}_{\frac{\epsilon}{2}}(\phi)$ lies within $B_1 \cap B_2$ and contains g. $\square$ 4.1.4

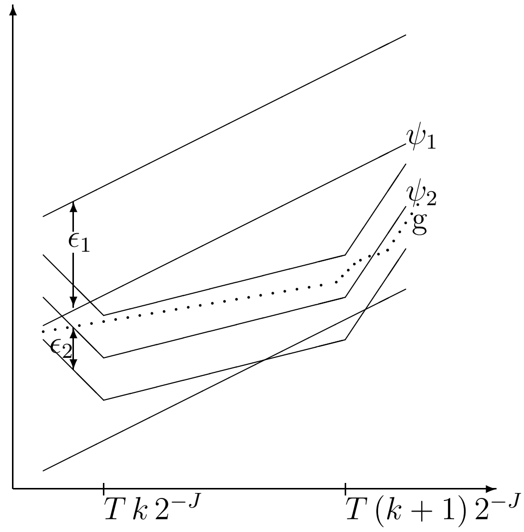


Figure 4.1: Feasible $B_1, B_2 \in \mathcal{J}$ and $g \in B_1 \cap B_2$

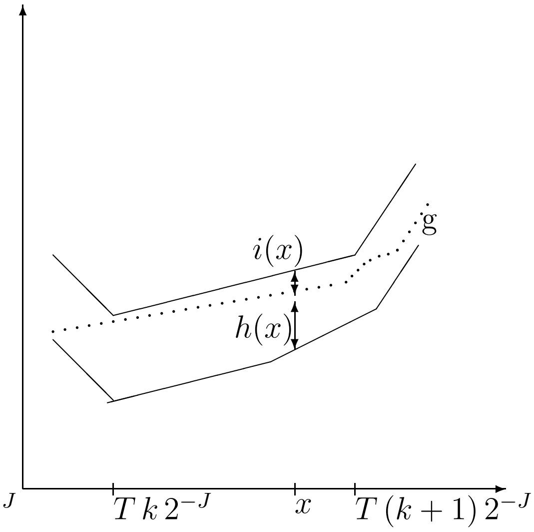


Figure 4.2: Restrictions (4.1) for functions in $B_1 \cap B_2$, g and $h(x), i(x)$ for a fixed x and the fixed g

The base $\mathcal{J}$ consists of convex sets (not-compact sup-norm balls). This agrees with $C([0,T],\mathbb{R})$ being locally convex.

We get a similar countable $\mathcal{J}$ if we allow only piecewise linear functions with slopes in $\mathbb{Q}$, starting in $\mathbb{Q}$, and sup norm balls of rational radii. This agrees with $C([0,T],\mathbb{R})$ being separable. As a complete separable space $(C([0,T]),\mathbb{R})$ is denoted a Polish space.

If we fix the initial values and work with $(C([0,T],\mathbb{R}) \cap \{f \mid f(0) = x\}\,,\, ||.||)$ we get a base of the induced topology as $\mathcal{J} \cap \{f \mid f(0) = x\}$.

4.2 Local large deviations

We directly compute decay-rates for probabilities of the event that the scaled renewal counting process N_n (cf definition 3.4.1) stays close to a piecewise linear function $\psi \in \mathcal{P}^J$ with some $J \in \mathbb{N}$:

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \tag{4.2}$$

We do this in steps, first for linear $\psi \in \mathcal{P}^0$ and then for the general case $\psi \in \mathcal{P}^J$, $J \geq 1$.

Remember that Λ is the lmgf of a typical inter event time τ of N and $\Gamma = -\Lambda^{-1}(-\,\cdot)$ has been defined in 2.2.7. We will apply the density process $\mathcal{M}$ defined in 3.6.4 (for an explicit form: (3.15)). The r appearing in the density process has been defined in 3.6.5. We repeat the change of measure process with twist parameter $\alpha = -\Lambda(\beta)$ for $\beta \in \mathcal{D}(\Lambda)$:

$$\mathcal{M}(\alpha, t) \;=\; \exp\{\alpha N_t - t\,\Gamma(\alpha)\}\, r(-\Gamma(\alpha), t)$$

This change of measure process applied to the counting process corresponds to the exponential twist of inter event times with parameter $\beta = -\Gamma(\alpha)$ (cf 3.6.8).

4.2.1 Local large deviations upper bound

We calculate the limsup for the expression (4.2) for $t \mapsto tv \in \mathcal{P}^0$ with some $v \geq 0$.

Claim 4.2.1. *For a renewal counting process N with lmgf Γ and Γ^* the Fenchel-Legendre transform of Γ (cf section 2.6)*

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(N_n \in \mathcal{U}_\epsilon(t\mapsto tv)) \leq -T\Gamma^*(v)$$

Proof of 4.2.1: We start with a change of measure with $\mathcal{M}(\alpha,\cdot)$ for some $\alpha \in \mathbb{R}$.

$$\begin{aligned}
&\mathbb{P}(N_n \in \mathcal{U}_\epsilon(t\mapsto tv))\\
&\quad= \mathbb{E}[\mathbb{1}_{N_n\in\mathcal{U}_\epsilon(t\mapsto tv)}] = \mathbb{E}[\mathbb{1}_{N_n\in\mathcal{U}_\epsilon(t\mapsto tv)}\frac{\mathcal{M}(\alpha,nT)}{\mathcal{M}(\alpha,nT)}]\\
&\quad= \mathbb{E}^{[\alpha]}[\mathbb{1}_{N_n\in\mathcal{U}_\epsilon(t\mapsto tv)}\frac{1}{\mathcal{M}(\alpha,nT)}]\\
&\quad= \mathbb{E}^{[\alpha]}[\mathbb{1}_{N_n\in\mathcal{U}_\epsilon(t\mapsto tv)}\;\exp\{-\alpha\,N(nT) + n\,T\,\Gamma(\alpha)\}\,\frac{1}{r(-\Gamma(\alpha),nT)}]
\end{aligned}$$

We now add the zero $\alpha\,nTv - \alpha\,nTv$ and bound by applying closeness of

$N(nT)$ to nTv as enforced by the indicator. For $\alpha > 0$

$$\begin{aligned}
&\mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \\
&= \mathbb{E}^{[\alpha]}[\mathbb{1}_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \\
&\qquad \exp\{\underbrace{-\alpha\, N(nT) + \alpha\, nTv}_{=-\alpha\, n\, T(N_n(T)-Tv) \le \alpha\, n\, T\, \epsilon}\ \underbrace{-\alpha\, nTv + n\, T\, \Gamma(\alpha)}_{=-nT(\alpha\, v - \Gamma(\alpha))}\} \frac{1}{r(-\Gamma(\alpha), nT)}] \\
&\le \mathbb{E}^{[\alpha]}[\mathbb{1}_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha), nT)}] \exp\{-n\, T\, \big(\alpha\, (v-\epsilon) - \Gamma(\alpha)\big)\}
\end{aligned}$$

The expectation $\mathbb{E}^{[\alpha]}[\mathbb{1}_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha),nT)}]$ is finite and uniformly bounded in n: Bound $\mathbb{1}_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \le 1$ and apply boundedness of $\frac{1}{r(-\Gamma(\alpha),\cdot)}$ proven in 3.6.11. There is no exponential growth in the expectation and it vanishes in the exponentially scaled upper bound limit. Thus we finish for $\alpha > 0$

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \\
&\le \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}^{[\alpha]}[\mathbb{1}_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha), nT)}] \exp\{-n\, T\, \big(\alpha\, (v-\epsilon) - \Gamma(\alpha)\big)\} \\
&\le -T\, \big(\alpha\, (v-\epsilon) - \Gamma(\alpha)\big) \qquad (4.3)
\end{aligned}$$

while for $\alpha < 0$ we get

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \ \le\ -\alpha\, T\, (v+\epsilon) + T\, \Gamma(\alpha)$$

Optimising the bound

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \\
&\qquad \le -T \max\{\sup_{\alpha>0} \alpha\, (v-\epsilon) - \Gamma(\alpha)\, ,\, \sup_{\alpha<0} \alpha\, (v+\epsilon) - \Gamma(\alpha)\}
\end{aligned}$$

and letting $\epsilon \to 0$

$$\begin{aligned}
&\lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \\
&\qquad \le -T \max\{\sup_{\alpha>0} \alpha\, v - \Gamma(\alpha)\, ,\, \sup_{\alpha<0} \alpha\, v - \Gamma(\alpha)\} \\
&\qquad = -T\Gamma^*(v).
\end{aligned}$$

$\square$ 4.2.1

4.2.2 Local large deviations lower bound

We calculate the liminf for the expression (4.2) for $t \mapsto tv \in \mathcal{P}^0$ and $v \geq 0$.

Claim 4.2.2. *For a renewal counting process N with lmgf Γ and Γ^* the Fenchel-Legendre transform of Γ (cf section 2.6)*

$$\lim_{\epsilon \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \geq -T\Gamma^*(v)$$

To prove the claim we will apply the following

Lemma 4.2.3. *If $\Gamma'(\alpha) = v$ then*

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}^{[\alpha]}[1\!\!1_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha), t))}] = 0.$$

Proof of 4.2.3: We have for $\beta = -\Gamma(\alpha)$ bounded $x \mapsto \frac{F^c_\beta}{F^c}(x)\, e^{-\beta x}$ and thus

$$\mathbb{E}^{[\alpha]}[1\!\!1_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha), nT)}] \;\; \geq \;\; \underbrace{\mathbb{E}^{[\alpha]}[1\!\!1_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)}]}_{\to 1 \text{ by } 7.1.3} \inf_{x \in \mathbb{R}} \frac{F^c}{F^c_\beta}(x)\, e^{\beta x} \; > \; 0$$

The convergence in 7.1.3 of the appendix is almost surely. From boundedness the convergence is in the mean, too. $\square$ 4.2.3

Proof of 4.2.2: Consider the case $v > 0$ first. We start similarly as for the upper bound. We again apply that $N(nT)$ is close to nTv. Let $\alpha > 0$.

$$\begin{aligned}
&\mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \\
&= \; \mathbb{E}^{[\alpha]}[1\!\!1_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \\
&\qquad \exp\{ \underbrace{-\alpha\, N(nT) + \alpha\, nTv}_{= -\alpha\, n\, T(N_n(T) - Tv) \geq -\alpha\, n\, T\, \epsilon} \; \underbrace{-\alpha\, nTv + n\, T\, \Gamma(\alpha)}_{= -nT(\alpha\, v - \Gamma(\alpha))}\} \frac{1}{r(-\Gamma(\alpha), nT)}] \\
&\geq \; \mathbb{E}^{[\alpha]}[1\!\!1_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha), nT)}] \exp\{-n\, T \left(\alpha\,(v + \epsilon) - \Gamma(\alpha)\right)\}
\end{aligned}$$

Fix α such that $\Gamma'(\alpha) = v$ and lemma 4.2.3 is applicable. For $v > \Gamma'(0)$ we have $\alpha > 0$ (by $\alpha = (\Gamma')^{-1}(v) > (\Gamma')^{-1}(\Gamma'(0)) = 0$) and lower bound

$$\begin{aligned}
&\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(t \mapsto tv)) \\
&\qquad \geq \; \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{E}^{[\alpha]}[1\!\!1_{N_n \in \mathcal{U}_\epsilon(t \mapsto tv)} \frac{1}{r(-\Gamma(\alpha), nT)}] \\
&\qquad\qquad\qquad\qquad \exp\{-\alpha\, n\, T\,(v + \epsilon) + n\, T\, \Gamma(\alpha)\} \\
&\qquad = \; 0 - \alpha\, T\,(v + \epsilon) + T\, \Gamma(\alpha).
\end{aligned}$$

This results in the now accurate

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}(N_n\in\mathcal{U}_\epsilon(t\mapsto tv)) \overset{(1)}{=} -T\left(\alpha(v)\,v-\Gamma(\alpha(v))\right) \overset{(2)}{=} -T\,\Gamma^*(v).$$

where we wrote $\alpha(v)$ since α is the twist with $\Gamma'(\alpha)=v$ from lemma 4.2.3 needed in (1). The same property $\Gamma'(\alpha)=v$ justifies (2).

Similarly for $v\in(0,\Gamma'(0))$ and $\Gamma'(\alpha)=v$ which implies $\alpha\le 0$.

Since $\Gamma'>0$ always lemma 4.2.3 does not apply to $v=0$. However, for $v=0$ we may write

$$N_n\in\mathcal{U}_\epsilon(t\mapsto 0\cdot t) \qquad \Leftrightarrow \qquad N_n<\epsilon$$

and calculate the probability of the event directly.

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sup_{t\in[0,T]}|N_n(t)|<\epsilon) = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(N_n(T)<\epsilon) = -T\Gamma^*(\epsilon)$$

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(\sup_{t\in[0,T]}|N_n(t))|<\epsilon) = -T\Gamma^*(0)$$

With $\Gamma^*(0)=L_C(h)$ and $=\infty$ for LD-bounded inter event times. We applied continuity of Γ^* (or simultaneous divergence to ∞) in 0 from the corollary 2.6.4.

$\square$
4.2.2

4.2.3 Generalisation

Up to now we have started the scaled counting process as $N_n(0)=0$ and calculated decay rates for the counting process to stay close to some linear function starting in 0, too. Upper and lower bound work exactly the same way for counting processes starting in $\frac{\lfloor nx\rfloor}{n}$ and an affine function $t\mapsto x+tv$. We make this explicit in the following notation and state the generalised result.

Definition 4.2.4 (Scaling)**.** *For a counting process N and a fixed $x>0$ set*

$$N_n(\cdot,x): t\mapsto \frac{\lfloor nx\rfloor}{n}+\frac{1}{n}N(nt)$$

Corollary 4.2.5. *Let N be a counting process, $x>0$ and $\psi\in\mathcal{P}^0$, $\psi(t)=x+tv$. Let $N_n(\cdot,x)$ be the scaled process of definition 4.2.4. Then*

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(N_n(\cdot,x)\in\mathcal{U}_\epsilon(t\mapsto x+tv)) = -T\Gamma^*(v)$$

4.2.4 Piecewise linear functions

We calculate (4.2) for general $J \geq 1$ applying heuristics learned from [23] (chapter 5, p. 73) developed for Markovian processes.

Claim 4.2.6. *If $\psi \in \mathcal{P}^J$ with non-negative slopes $v_1, v_2, \ldots, v_{2^J}$ and N an undelayed renewal counting process with lmgf Γ and N_n the scaled process associated with N. Then*

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \;=\; -T2^{-J} \sum_{k=1}^{2^J} \Gamma^*(v_k)$$

Proof of 4.2.6: Let $N_n^{\mathrm{re},(T2^{-J},2T2^{-J},\ldots,(2^J-1)T2^{-J})}$ be the scaled restarted process defined in 3.4.16 and 3.4.12 for $k = 2^J - 1$ with the equidistant

$$s_1 = T\,2^{-J}\,,\; s_2 = 2\,T\,2^{-J}\;\ldots\; s_{2^J-1} = T\,(2^J-1)\,2^{-J}$$

and identically distributed rcps

$$N^{(0)}\,,\; N^{(1)}\,,\;\ldots\;,\; N^{(2^J-1)}$$

We abbreviate $N_n^{\mathrm{re}} := N_n^{\mathrm{re},(T2^{-J},2T2^{-J},\ldots,(2^J-1)T2^{-J})}$ for the rest of this section 4.2.4.

On the interval $[kT2^{-J}, (k+1)T2^{-J}]$ the restarted process starts in $N_n^{\mathrm{re}}(kT2^{-J})$ and increases as the renewal counting process $N^{(k)}$. Note that N_n^{re} and ψ are defined with respect to the same cases. Let $v_1, \ldots, v_{2^J}$ be the piecewise constant slopes of $\psi \in \mathcal{P}^J$ and set $s_0 = 0, s_{2^J} = T$ for notational convenience.

$$\begin{aligned}
&N_n^{\mathrm{re}}(t) - \psi(t)\\
&= \begin{cases}
N(t) - v_1 t & \text{if } t \in [0, s_1]\\
N(s_1) + N^{(1)}(t - s_1) - v_1 s_1 - v_2(t - s_1) & \text{if } t \in (s_1, s_2]\\
\vdots & \\
\begin{aligned}&\sum_{j=0}^{2^J-2} N^{(j)}(s_{j+1} - s_j) - v_{j+1}(s_{j+1} - s_j)\\ &+N^{(2^J-1)}(t - s_{2^J-1}) - v_{2^J}(t - s_{2^J-1})\end{aligned} & \text{if } t \in (s_{2^J-1}, s_{2^J}] = (s_{2^J-1}, T]
\end{cases}\\
&= \sum_{k=0}^{2^J-1} \mathbb{1}_{t \in (s_k, s_{k+1}]}\\
&\qquad \Big(\sum_{j=0}^{k-1} N^{(j)}(s_{j+1} - s_j) - v_{j+1}(s_{j+1} - s_j) + N^{(k)}(t - s_k) - v_{k+1}(t - s_k) \Big)
\end{aligned}$$

and from the triangle inequality

$$||N_n^{\mathrm{re}} - \psi|| \;\le\; \sum_{k=0}^{2^J-1} \mathbb{1}_{t\in(s_k,s_{k+1}]} \tag{4.4}$$
$$\Big(\sum_{j=0}^{k-1} |N^{(j)}(s_{j+1}-s_j) - v_{j+1}(s_{j+1}-s_j)| + ||N^{(k)} - (t\mapsto tv_{k+1})||_{[0,s_{k+1}-s_k]}\Big).$$

Now if

$$||N^{(j)} - (t\mapsto tv_{j+1})||_{[0,s_{j+1}-s_j]} \le \frac{\epsilon}{2^J}$$

then especially for the end point of the interval

$$|N^{(j)}(s_{j+1}-s_j) - v_{j+1}(s_{j+1}-s_j)| \le \frac{\epsilon}{2^J}$$

and

$$\begin{aligned}
||N_n^{\mathrm{re}} - \psi|| \;&\le\; \sum_{k=0}^{2^J-1} \mathbb{1}_{t\in(s_k,s_{k+1}]} \\
&\quad \Big(\sum_{j=0}^{k-1} |N^{(j)}(s_{j+1}-s_j) - v_{j+1}(s_{j+1}-s_j)| + ||N^{(k)} - (t\mapsto tv_{k+1})||_{[0,s_{k+1}-s_k]}\Big) \\
&\le\; \sum_{k=0}^{2^J-1} \mathbb{1}_{t\in(s_k,s_{k+1}]}\Big(\sum_{j=0}^{k-1}\frac{\epsilon}{2^J} + \frac{\epsilon}{2^J}\Big) = \sum_{k=0}^{2^J-1} \mathbb{1}_{t\in(s_k,s_{k+1}]}(k+1)\frac{\epsilon}{2^J} \\
&\le\; \epsilon.
\end{aligned}$$

Thus, by independence of increments

$$\begin{aligned}
\mathbb{P}(N_n^{\mathrm{re}} \in \mathcal{U}_\epsilon(\psi)) \;&\ge\; \mathbb{P}(||N_n^{(k)} - (t\mapsto tv_{k+1})|| < \epsilon\, 2^{-J} \quad \forall k) \\
&=\; \prod_{k=0}^{2^J-1} \mathbb{P}(||N_n^{(k)} - (t\mapsto tv_{k+1})|| < \epsilon\, 2^{-J})
\end{aligned}$$

$$\begin{aligned}
&\lim_{\epsilon\to 0}\liminf_{n\to\infty} \frac{1}{n}\log \mathbb{P}(N_n^{\mathrm{re}} \in \mathcal{U}_\epsilon(\psi)) \\
&\quad\ge\; \sum_{k=0}^{2^J-1} \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log \mathbb{P}(||N_n^{(k)} - (t\mapsto tv_{k+1})|| < \epsilon\, 2^{-J}) \\
&\quad\ge\; \sum_{k=0}^{2^J-1} T2^{-J}\Gamma^*(v_{k+1}) \;=\; T2^{-J}\sum_{k=1}^{2^J}\Gamma^*(v_k)
\end{aligned}$$

From exponential equivalence of N and the restarted process limits in the form of the claim are the same for both processes by application of claim 7.2.1 of the appendix. This finishes the lower bound. The upper bound is quite similar with blowing up radii from constant ϵ to an increasing sequence of radii $\epsilon, 2\epsilon, \ldots, 2^J\epsilon$. $\square$ 4.2.6

4.2.5 Towards linear geodesics

In sample path large deviations linear geodesics is the property that the rate function is in integral form and the integrand, the so called local rate function, is convex (cf [9], definition 6.1). In their book [9] Ayalvadi Ganesh, Neil O'Connell, and Damon Wischik approach large deviations through proving a sample path large deviation principle once and then deducing further large deviation principles by application of the contraction principle. To get an explicit rate function from the contraction principle a variational problem has to be solved - and here linear geodesics helps. Technically linear geodesics is the settting where Jensen's inequality can be applied.

In the queueing setting a nice application of the contraction principle and linear geodesics is the large deviations of the queue size at some fixed time $t > 0$ if the smoothed queue has started in $t = 0$ with size x_0. In this and many other cases we can observe a most likely path to the event of interest which is piecewise linear.

While linear geodesics starts with a rate function in integral form and can prove the implication that the process "likes" to move along piecewise linear functions, we argue the other way around: Aiming to get a large deviation principle with a rate function in integral form (like in the Markovian case) we can already prove that the scaled renewal counting process deliberately stays close to linear functions if it can.

Claim 4.2.7. *For $\psi \in \mathcal{P}^0$ and $\phi \in \mathcal{P}^1$ with $\phi(0) = \psi(0)$, $\phi(T) = \psi(T)$ and $||\phi - \psi|| > 0$*

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\phi)) \quad < \quad \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi))$$

Proof of 4.2.7: We name slopes of ψ and ϕ first:

$$\psi : [0, T] \to [0, \infty) \quad , \quad t \mapsto tv$$

$$\phi : [0, T] \to [0, \infty) \quad , \quad t \mapsto \begin{cases} tw_1 & \text{for } t \in [0, \frac{T}{2}] \\ \frac{T}{2} w_1 + (t - \frac{T}{2})\, w_2 & \text{for } t \in (\frac{T}{2}, T] \end{cases}$$

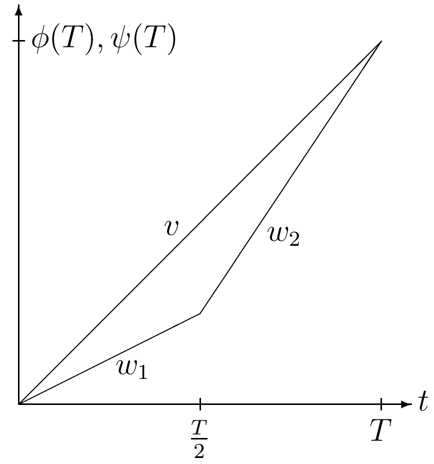


Figure 4.3: ψ with slope v and ϕ with slopes w_1, w_2

The paths' different slopes relate as

$$Tv \;=\; \frac{T}{2}\,w_1 + (T - \frac{T}{2})\,w_2 \;=\; T\,(\frac{1}{2}w_1 + \frac{1}{2}w_2)$$

And from strict convexity of the Fenchel-Legendre transform Γ^*

$$\Gamma^*(v) = \Gamma^*(\frac{1}{2}w_1 + \frac{1}{2}w_2) \;<\; \frac{1}{2}\Gamma^*(w_1) + \frac{1}{2}\Gamma^*(w_2)$$

which in terms of the decay rate for the tubes is

$$\begin{aligned} T\Gamma^*(v) \quad &< \quad \frac{T}{2}\Gamma^*(w_1) + \frac{T}{2}\,\Gamma^*(w_2) \\ \Leftrightarrow\; \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \quad &> \quad \lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\phi)) \qquad \square_{4.2.7} \end{aligned}$$

Comparing two elements of $\mathcal{P}^1$ the one that is close to a linear function is asymptotically preferred by the scaled process N_n. In preparation for claim 4.2.9 we make the following

Definition 4.2.8. *Let $\psi, \phi \in \mathcal{P}^1$ be defined as*

$$\psi : [0,T] \to [0,\infty) \quad , \quad t \mapsto \begin{cases} tv_1 & \text{for } t \in [0, \frac{T}{2}] \\ \frac{T}{2}v_1 + (t - \frac{T}{2})\,v_2 & \text{for } t \in (\frac{T}{2}, T] \end{cases}$$

$$\phi : [0,T] \to [0,\infty) \quad , \quad t \mapsto \begin{cases} tw_1 & \text{for } t \in [0, \frac{T}{2}] \\ \frac{T}{2}w_1 + (t - \frac{T}{2})\,w_2 & \text{for } t \in (\frac{T}{2}, T] \end{cases}$$

with non-negative v_1, v_2, w_1, w_2 and $\phi(0) = \psi(0)$, $\phi(T) = \psi(T)$.

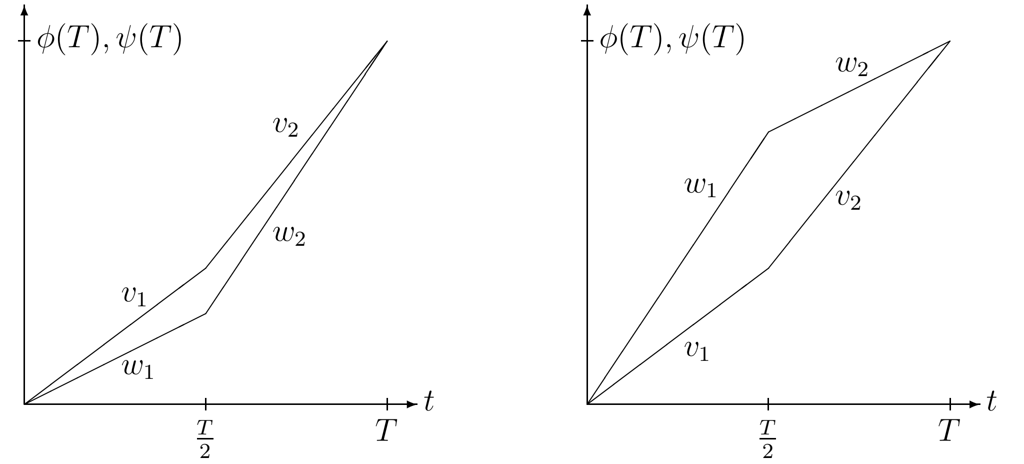


Figure 4.4: Two sets of $\{\psi, \phi\}$ suiting definition 4.2.8

Note that v_1, v_2 have the same distance to $\frac{\psi(T)}{T}$.

$$\begin{aligned} \frac{\psi(T)}{T} = \frac{1}{2}(v_1 + v_2) \quad &\Leftrightarrow \quad \frac{\psi(T)}{T} - v_2 = v_1 - \frac{\psi(T)}{T} \\ &\Rightarrow \quad |\frac{\psi(T)}{T} - v_2| = |\frac{\psi(T)}{T} - v_1| \end{aligned}$$

Writing slopes as a vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ we get

$$\begin{aligned} \frac{\psi(T)}{T} = \frac{1}{2}(v_1 + v_2) \quad &\Leftrightarrow \quad v_2 = 2\frac{\psi(T)}{T} - v_1 \\ &\Leftrightarrow \quad \vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \underbrace{\frac{\psi(T)}{T}}_{\text{mean slope}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \underbrace{(v_1 - \frac{\psi(T)}{T}) \begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\text{off-set from mean slope}} \end{aligned}$$

which is visualised in figure 4.5. Since all the above considerations apply to ϕ and its vector of slopes $\vec{w}$ as well, the figure shows again a $\vec{v}$ corresponding to ψ and a $\vec{w}$ corresponding to ϕ.

Claim 4.2.9. *For ϕ, ψ of definition 4.2.8 If*

$$||\vec{v} - \frac{\psi(T)}{T} \begin{bmatrix} 1 \\ 1 \end{bmatrix}|| < ||\vec{w} - \frac{\psi(T)}{T} \begin{bmatrix} 1 \\ 1 \end{bmatrix}||$$

then

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\phi)) \quad < \quad \lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi))$$

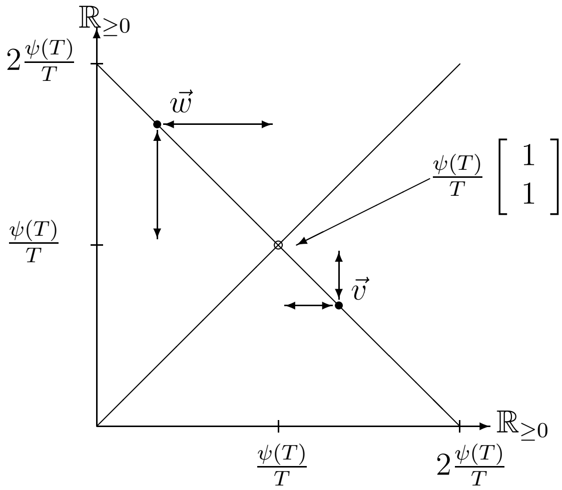


Figure 4.5: Slopes $\vec{v}$ for ψ and $\vec{w}$ for ϕ suiting definition 4.2.8.

Proof of 4.2.9: Consider the vectors of slopes $\vec{v}, \vec{w}$ in $\mathbb{R}^2_{\geq 0}$. Note that

$$||\vec{v} - \frac{\psi(T)}{T}\begin{bmatrix}1\\1\end{bmatrix}|| = |v_1 - \frac{\psi(T)}{T}| \cdot ||\begin{bmatrix}1\\1\end{bmatrix}||$$

and thus

$$||\vec{v} - \frac{\psi(T)}{T}\begin{bmatrix}1\\1\end{bmatrix}|| \;<\; ||\vec{w} - \frac{\psi(T)}{T}\begin{bmatrix}1\\1\end{bmatrix}|| \quad \Leftrightarrow \quad |v_1 - \frac{\psi(T)}{T}| < |w_1 - \frac{\psi(T)}{T}|$$

which, due to symmetry, is equivalent to

$$|v_2 - \frac{\psi(T)}{T}| \;<\; |w_2 - \frac{\psi(T)}{T}|$$

Now we have an ordering

$$w_{i_1} < v_{j_1} < \frac{\psi(T)}{T} < v_{j_2} < w_{i_2}$$

with $i_1, i_2, j_1, j_2 \in \{1, 2\}$ and $i_1 \neq i_2$, $j_1 \neq j_2$. Lets assume that $w_1 < v_1 < \frac{\psi(T)}{T} < v_2 < w_2$. Then

$$v_i = \frac{v_i - w_1}{w_2 - w_1}\, w_1 + \frac{w_2 - v_i}{w_2 - w_1}\, w_2 \qquad (i = 1, 2)$$

and by convexity

$$\begin{aligned}
\Gamma^*(v_i) &\leq \frac{v_i - w_1}{w_2 - w_1}\,\Gamma^*(w_1) + (1 - \frac{v_i - w_1}{w_2 - w_1})\,\Gamma^*(w_2) \qquad (i = 1, 2) \\
\Rightarrow \Gamma^*(v_1) &+ \Gamma^*(v_2) \\
&\leq \left(\frac{v_1 - w_1}{w_2 - w_1} + \frac{v_2 - w_1}{w_2 - w_1}\right)\Gamma^*(w_1) + \left(\frac{w_2 - v_1}{w_2 - w_1} + \frac{w_2 - v_2}{w_2 - w_1}\right)\Gamma^*(w_2) \\
&= \frac{v_1 + v_2 - 2w_1}{w_2 - w_1}\,\Gamma^*(w_1) + \frac{2w_2 - v_1 - v_2}{w_2 - w_1}\,\Gamma^*(w_2) \\
&= \Gamma^*(w_1) + \Gamma^*(w_2)
\end{aligned}$$

Thus from

$$\begin{aligned}
\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &= -\frac{T}{2}(\Gamma^*(v_1) + \Gamma^*(v_2)) \\
\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\phi)) &= -\frac{T}{2}(\Gamma^*(w_1) + \Gamma^*(w_2))
\end{aligned}$$

follows the claim. $\square$ 4.2.9

4.2.6 A limit

We can write the rhs of 4.2.6 in integral form as

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = -\int_{s=0}^{T} \Gamma^*(\psi'(s))\, ds$$

While for finite J this integral is artifical, it gives us a uniform expression for finite J and the limit of $J \to \infty$.

In the following we will apply a standard way of approximating absolutely continuous functions.

Definition 4.2.10. *For $f \in AC[0,T]$ with $f(s) = f(0) + \int_{r=0}^{s} g(r)\, dr$ for some $g \in \mathcal{L}^1$ define its piecewise linear approximation $f^J \in \mathcal{P}^J$ through a piecewise constant approximation of the almost derivative g.*

$$\begin{aligned}
g_J(s) &= \frac{1}{T}\, 2^J \int_{r=\lfloor \frac{s}{T} 2^J \rfloor\, T\, 2^{-J}}^{\lceil \frac{s}{T} 2^J \rceil\, T\, 2^{-J}} g(r)\, dr \\
f_J &= f(0) + \int g_J
\end{aligned}$$

Claim 4.2.11. *The f^J of 4.2.10 have a limit "in the rate function":*

$$\lim_{J\to\infty}\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log P(N_n\in\mathcal{U}_\epsilon(f^J)) = \int_{s=0}^{T}\Gamma^*(f'(s))\,ds$$

Proof of 4.2.11: The g_J approximate g in a pointwise fashion (cf [5] C.13, a Lebesgue theorem). An application of the pointwise convergence of $g_J\to g$ and the Fatou-lemma tells us

$$\begin{aligned}
\liminf_{J\to\infty}\int_{s=0}^{T}\Gamma^*(f^{J\prime}(s))\,ds &= \liminf_{J\to\infty}\sum_{l=0}^{2^J-1}2^{-J}\,T\,\Gamma^*(g_J(l\,2^{-J}\,T))\\
&= \liminf_{J\to\infty}\int_{s=0}^{T}\Gamma^*(g_J(s))\,ds\\
&\overset{\text{Fatou}}{\geq} \int_{s=0}^{T}\liminf_{J\to\infty}\Gamma^*(g_J(s))\,ds\\
&\overset{\text{lsc}}{\geq} \int_{s=0}^{T}\Gamma^*(g(s))\,ds\\
&= \int_{s=0}^{T}\Gamma^*(f'(s))\,ds
\end{aligned}$$

We get the other direction, an upper bound, immediately from Jensen's inequality. We only need this if the lower bound is finite.

$$\Gamma^*(g_J(lT2^{-J})) \;=\; \Gamma^*\Big(\frac{2^J}{T}\int_{r=lT2^{-J}}^{(l+1)T2^{-J}}g(r)\,dr\Big) \;\leq\; \frac{2^J}{T}\int_{r=lT2^{-J}}^{(l+1)T2^{-J}}\Gamma^*(g(r))\,dr$$

thus

$$\begin{aligned}
\int_{s=0}^{T}\Gamma^*(f^{J\prime}(s))\,ds &= \sum_{k=0}^{2^J-1}2^{-J}T\Gamma^*(g_J(kT2^{-J})) \;\leq\; \sum_{k=0}^{2^J-1}2^{-J}T\frac{2^J}{T}\int_{r=kT2^{-J}}^{(k+1)T2^{-J}}\Gamma^*(g(r))\,dr\\
&= \int_{r=0}^{T}\Gamma^*(g(r))\,dr \;=\; \int_{s=0}^{T}\Gamma^*(f'(s))\,ds
\end{aligned}$$

What we have now is a well defined limit for any $f\in\mathcal{AC}$.

$$\lim_{J\to\infty}\int_{s=0}^{T}\Gamma^*(f^{J\prime}(s))\,ds = \int_{s=0}^{T}\Gamma^*(f'(s))\,ds \tag{4.5}$$

So the claimed equality is proved. $\square$ 4.2.11

4.3 The LDP in sample space

In this section we develop a full large deviation principle for the counting process. We will apply the local large deviations, the tube limits, found in previous sections and repeat some of the techniques already developed. The main object of this section are ϵ-neighbourhoods of piecewise linear functions: the technical difference to the tubes in the local large deviations is that we have fixed $\epsilon > 0$ instead of a limit $\epsilon \to 0$.

4.3.1 The weak large deviation principle

The following is an application of theorem 4.1.11 of [5]. We state it here in the notation fitting our context and as implied by the remark following the theorem.

Theorem 4.3.1 (Dembo and Zeitouni)**.** *In the space of continuous functions over a compact interval equipped with the sup-norm ($C[0,T], ||.||$) let $\mathcal{A}$ be a base of the topology. If for every $U \in \mathcal{A}$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{N}_n \in U) \tag{4.6}$$

exists in $\mathbb{R} \cup \{-\infty\}$ then $\hat{N}_n$ satisfies the weak LDP with the rate function I defined as

$$I(f) := \sup_{U\in\mathcal{A}\,:\,f\in U} -\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{N}_n \in U).$$

The theorem takes place in $(C[0,T], ||.||)$ and the large deviation object has to be the interpolated counting process $\hat{N}_n \in C[0,T]$. To calculate the limit (4.6) we will be working with the undelayed and not-interpolated counting process N since limits for both processes are the same (cf claim 7.2.1 of the appendix).

Claim 4.3.2. *For $\psi \in \mathcal{P}^0$, $\psi(t) = vt$ for some $v \geq 0$, and $\epsilon > 0$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = -T \inf_{w\in[v-\frac{\epsilon}{T}\,,\,v+\frac{\epsilon}{T}]} \Gamma^*(w)$$

Proof of 4.3.2: For the lower bound we apply our tubes limit. Let $\delta > 0$ and $w \in [v - \frac{\epsilon}{T} + \frac{\delta}{T}\,,\, v + \frac{\epsilon}{T} - \frac{\delta}{T}]$.

$$\begin{aligned} \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\geq \mathbb{P}\big(N_n \in \mathcal{U}_\delta(t \mapsto tw)\big) \qquad (\delta \in (0,\epsilon]) \\ \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\geq -T\,\Gamma^*(w) \end{aligned}$$

As we let $\delta \to 0$ the restrictions for w become $w \in (v - \frac{\epsilon}{T}\,,\, v + \frac{\epsilon}{T})$ resulting in

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\geq -T \inf_{w \in (v-\frac{\epsilon}{T}\,,\, v+\frac{\epsilon}{T})} \Gamma^*(w) \\
&\geq -T \inf_{w \in [v-\frac{\epsilon}{T}\,,\, v+\frac{\epsilon}{T}]} \Gamma^*(w).
\end{aligned}$$

The upper bound is simplified to the large deviation of the mean (one-dimensional LDP in claim 3.5.1).

$$\begin{aligned}
\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\leq \mathbb{P}\big(N_n(T) \in [Tv - \epsilon\,,\, Tv + \epsilon]\big) \\
&= \mathbb{P}\big(\frac{1}{nT} N(nT) \in [v - \frac{\epsilon}{T}\,,\, v + \frac{\epsilon}{T}]\big) \\
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\leq -T \inf_{w \in [v-\frac{\epsilon}{T}\,,\, v+\frac{\epsilon}{T}]} \Gamma^*(w)
\end{aligned}$$

□

We can rephrase 4.3.2 the following way: There is a $\phi \in \mathcal{P}^0 \cap \mathcal{U}_\epsilon(\psi)$ such that the probability to stay within $\mathcal{U}_\epsilon(\psi)$ is carried by some ϕ:

$$\lim_{n\to\infty} \frac{1}{n} \log P(N_n \in \mathcal{U}_\epsilon(\psi)) = \lim_{\delta\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\delta(\phi))$$

That ϕ is linear agrees with linear geodesics (cf section 4.2.5).

Claim 4.3.3. *For $\psi \in \mathcal{P}^1$ and $\epsilon > 0$*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = - \inf_{\phi \in \mathcal{P}^1 \cap \mathcal{U}_\epsilon(\psi)} \int_{t=0}^{T} \Gamma^*(\phi'(t))\, dt$$

Proof of 3.4.7: Let $v_1, v_2 \geq 0$ be such that

$$\psi(t) = \begin{cases} v_1\, t & \text{for } t \in [0, \frac{T}{2}] \\ v_1\, \frac{T}{2} + v_2\, (t - \frac{T}{2}) & \text{for } t \in (\frac{T}{2}, T] \end{cases}$$

For a lower bound: if $\phi \in \mathcal{P}^1$ with non-negative slopes w_1, w_2 is such that

$$\begin{aligned}
w_1 &\in (v_1 - \frac{2\epsilon}{T}\,,\, v_1 + \frac{2\epsilon}{T}) \\
\text{and} \qquad w_1 + w_2 &\in (v_1 + v_2 - \frac{2\epsilon}{T}\,,\, v_1 + v_2 + \frac{2\epsilon}{T})
\end{aligned}$$

then for $\delta > 0$ small enough $\mathcal{U}_\delta(\phi) \subseteq \mathcal{U}_\epsilon(\psi)$. Figure 4.6 shows such ψ and ϕ. Now we can bound applying claim 4.2.6 to $J = 1$, ϕ and $\mathcal{U}_\delta(\phi)$.

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\geq \lim_{\delta\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\delta(\phi)) \\ &= -\frac{T}{2}\big(\Gamma^*(w_1) + \Gamma^*(w_2)\big)\end{aligned}$$

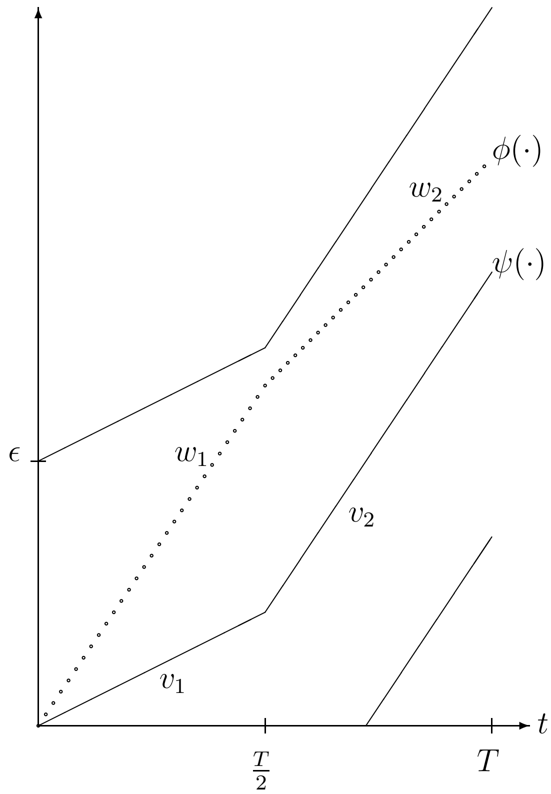


Figure 4.6: Solid ψ with slopes v_1, v_2 and dashed ϕ with slopes w_1, w_2, $\phi \in \mathcal{U}_\epsilon(\psi)$

Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ and to abbreviate the condition on $\vec{w}$ set

$$\begin{aligned}\mathcal{S}^1(v,\epsilon) := \{\, w \in \mathbb{R}^2 \,:\, & w_1 \in [v_1 - \frac{2\epsilon}{T}\,,\, v_1 + \frac{2\epsilon}{T}]\,, \\ & w_1 + w_2 \in [v_1 + v_2 - \frac{2\epsilon}{T}\,,\, v_1 + v_2 + \frac{2\epsilon}{T}]\,\} \\ \mathbb{V}_1 \,:\, \mathcal{S}^1(v,\epsilon) \,\to\, \mathbb{R} \quad , \quad & \vec{w} \,\mapsto\, \Gamma^*(w_1) + \Gamma^*(w_2)\end{aligned}$$

and optimise the lower bound:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \quad \geq \quad -\frac{T}{2} \inf_{\vec{w}\in\mathcal{S}^1(\vec{v},\epsilon)} \big(\Gamma^*(w_1) + \Gamma^*(w_2)\big)$$

To get an upper bound we apply the finite dimensional large deviation principle of 3.5.1: Consider the process only at fixed epochs $\frac{T}{2}$ and T:

$$\begin{aligned}
\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \quad &\leq \quad \mathbb{P}\big(N_n(\frac{T}{2}, T) \in \\
&\qquad (\psi(\frac{T}{2}) - \epsilon\,,\, \psi(\frac{T}{2}) + \epsilon) \times (\psi(T) - \epsilon\,,\, \psi(T) + \epsilon)\big) \\
&\leq \quad \mathbb{P}\big(N_n(\frac{T}{2}, T) \in \\
&\qquad [\psi(\frac{T}{2}) - \epsilon\,,\, \psi(\frac{T}{2}) + \epsilon] \times [\psi(T) - \epsilon\,,\, \psi(T) + \epsilon]\big)
\end{aligned}$$

Now apply the upper bound for closed sets of claim 3.5.2:

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \\
&\qquad \leq \quad -\frac{T}{2} \inf_{\vec{z}\in F} \Gamma^*(\frac{2z_1}{T}) + \Gamma^*(\frac{2(z_2 - z_1)}{T}) && (4.7) \\
\text{with } F \quad &= \quad [\psi(\frac{T}{2}) - \epsilon\,,\, \psi(\frac{T}{2}) + \epsilon] \times [\psi(T) - \epsilon\,,\, \psi(T) + \epsilon] && (4.8)
\end{aligned}$$

Define $\mathbb{V}_2$ the following way.

$$\mathbb{V}_2 \,:\, F \,\to\, \mathbb{R} \quad , \quad \vec{z} \,\mapsto\, \Gamma^*(\frac{2z_1}{T}) + \Gamma^*(\frac{2(z_2 - z_1)}{T})$$

To match lower and upper bound we need

$$\inf_{\vec{w}\in\mathcal{S}^1(\vec{v},\epsilon)} \mathbb{V}_1(\vec{w}) = \inf_{\vec{z}\in F} \mathbb{V}_2(\vec{z})$$

We prove a little more in lemma 4.3.4 which finishes the proof. $\square$ 4.3.3

Lemma 4.3.4. $\mathbb{V}_1 = \mathbb{V}_2 \circ \mathbb{U}$ *for a regular transformation* $\mathbb{U} : \mathcal{S}^1(\vec{v}, \epsilon) \to F$.

Proof of 4.3.4: Define $\mathbb{U}$ as

$$\mathbb{U} \,:\, \mathcal{S}^1(\vec{v}, \epsilon) \to F \quad , \quad \vec{w} \mapsto \frac{T}{2} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \vec{w}.$$

It is well defined since

$$
\begin{aligned}
\mathbb{U}(\vec{w}) \in F \quad &\Leftrightarrow \quad \left\{ \begin{array}{ccc} \frac{T}{2} w_1 & \in & [\psi(\frac{T}{2}) - \epsilon\,,\, \psi(\frac{T}{2}) + \epsilon] \\ \frac{T}{2}(w_1 + w_2) & \in & [\psi(T) - \epsilon\,,\, \psi(T) + \epsilon] \end{array} \right\} \\
&\overset{\cdot \frac{2}{T}}{\Leftrightarrow} \quad \left\{ \begin{array}{ccc} w_1 & \in & [\underbrace{\frac{2\psi(\frac{T}{2})}{T}}_{=v_1} - \frac{2\epsilon}{T}\,,\, \frac{2\psi(\frac{T}{2})}{T} + \frac{2\epsilon}{T}] \\ w_1 + w_2 & \in & [\underbrace{\frac{2\psi(T)}{T}}_{=v_1+v_2} - \frac{2\epsilon}{T}\,,\, \frac{2\psi(T)}{T} + \frac{2\epsilon}{T}] \end{array} \right\} \\
&\Leftrightarrow \quad \vec{w} \in \mathcal{S}^1(\vec{v}, \epsilon).
\end{aligned}
$$

Regularity is immediate from $\mathcal{S}^1(\vec{v}, \epsilon)$ having relative dimension 2 and the matrix representation. Further note that mapping of $\vec{z} \in F$ onto the arguments of Γ^* in (4.7)

$$
\vec{z} \quad \mapsto \quad \begin{bmatrix} \frac{2z_1}{T} \\ \frac{2(z_2 - z_1)}{T} \end{bmatrix} = \frac{2}{T} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \vec{z} = \mathbb{U}^{-1}(\vec{z})
$$

is the inverse transformation. Which already implies the claim. To be very exact

$$
\begin{aligned}
\mathbb{V}_2 \circ \mathbb{U}(\vec{w}) &= \mathbb{V}_2\Big(\begin{bmatrix} \frac{T}{2} w_1 \\ \frac{T}{2}(w_1 + w_2) \end{bmatrix} \Big) \\
&= \Gamma^*(\frac{2}{T}\frac{T}{2} w_1) + \Gamma^*(\frac{2}{T}(\frac{T}{2}(w_1 + w_2) - \frac{T}{2} w_1)) \\
&= \Gamma^*(w_1) + \Gamma^*(w_2) \\
&= \mathbb{V}_1(\vec{w})
\end{aligned}
$$

$\square$
4.3.4

Claim 4.3.3 holds in the general case of $J \in \mathbb{N}$, too. We state it and give the definition of the generalised objects, but do not prove the general case.

Remark 4.3.5. *For $\psi \in \mathcal{P}^J$ for some fixed $J \in \mathbb{N}$ and $\epsilon > 0$*

$$
\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = - \inf_{\phi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)} \int_{t=0}^{T} \Gamma^*(\phi'(t))\, dt
$$

The following is the definition analogue to the case $J = 1$ and an application of this definition.

Remark 4.3.6. *Given $\psi \in \mathcal{P}^J$ with slopes $v_1, \ldots, v_{2^J}$ forming $\vec{v} \in \mathbb{R}^{2^J}$*

$$\begin{aligned}
\mathcal{S}^J(\vec{v}, \epsilon) &:= \{w \in \mathbb{R}^{2^J}_{\geq 0} : \forall k \in \{1, \ldots, 2^J\} : \\
&\qquad \sum_{l=1}^{k} w_l \in [\sum_{l=1}^{k} v_l - \frac{2^J \epsilon}{T}, \sum_{l=1}^{k} v_l + \frac{2^J \epsilon}{T}]\} \\
F &= \times_{k=1}^{2^J} [\psi(k\,T\,2^{-J}) - \frac{2^J \epsilon}{T}, \psi(k\,T\,2^{-J}) + \frac{2^J \epsilon}{T}] \\
\mathbb{U} &: \mathcal{S}^J(\vec{v}, \epsilon) \to F \quad , \quad \vec{w} \mapsto T\,2^{-J} \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 & 0 \\ 1 & 1 & \ldots & 1 & 1 \end{bmatrix} \vec{w}
\end{aligned}$$

If ψ, $\phi \in \mathcal{P}^J$ with slopes v_i (for ψ) and w_i (for ϕ) for $i = 1, \ldots, 2^J$ forming $\vec{v}, \vec{w} \in \mathbb{R}^{2^J}$ then

$$\phi \in \mathcal{U}_\epsilon(\psi) \qquad \Leftrightarrow \qquad \vec{w} \in \mathcal{S}^J(\vec{v}, \epsilon) \qquad \Leftrightarrow \qquad \mathbb{U}(\vec{w}) \in F$$

According to 4.3.1 we have a rate function for the weak LDP. We want to identify the rate function as the decay rate on tubes.

Claim 4.3.7 (Rate function identification)**.** *If $\psi \in \mathcal{P}^K$ then*

$$I(\psi) = \int_{s=0}^{T} \Gamma^*(\psi'(s))\, ds.$$

Proof of 4.3.7: Let $\psi \in \mathcal{P}^K$ and K be minimal in that $\psi \notin \mathcal{P}^{K-1}$. From 4.3.1 we have $I(f) = \sup_{U \in \mathcal{A}, f \in U} - \lim_{n\to\infty} \frac{1}{n} \log P(N_n \in U)$. Applying the form of our base as $U = \mathcal{U}_\epsilon(\psi)$ for some $\psi \in \mathcal{P}^J$ with $J \in \mathbb{N}, \epsilon \in \mathbb{R}$.

$$I(f) \quad = \quad \sup_{\substack{\epsilon \in \mathbb{R}\,,\, J \in \mathbb{N} \\ \psi \in \mathcal{P}^J\,,\, f \in \mathcal{U}_\epsilon(\psi)}} \; - \underbrace{\lim_{n\to\infty} \frac{1}{n} \log P(N_n \in \mathcal{U}_\epsilon(\psi))}_{= -\inf_{\xi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)} \int \Gamma^* \circ \xi'}$$

In the infimum we need to concider all $\psi \in \mathcal{U}_\epsilon(f)$ and $\mathcal{U}_\epsilon(\psi) \cap \mathcal{P}^J$ will never be empty since it contains ψ by construction - we will not see the infimum over the empty set.

We immediately have $I(f) \geq \int \Gamma^* \circ f'$ (fix $J = K$ and let $\epsilon \to 0$).

To get the opposite inequality fix a feasible combination of J, ϵ, ψ: such that $\psi \in \mathcal{P}^J$ and $f \in \mathcal{U}_\epsilon(\psi)$. If $K \leq J$ then $f \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)$ and

$$\inf_{\xi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)} \int \Gamma^* \circ \xi' \leq \int \Gamma^* \circ f'$$

If $J < K$ optimising over $\mathcal{P}^K \cap \mathcal{U}_\epsilon(\psi)$ instead of $\mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)$ would generally decrease the infimum. However, from section 4.2.5 on linear geodesics we know that the functions $\xi \in \mathcal{P}^K \cap (\mathcal{P}^j)^c$ we added to the set of restrictions do not decrease but increase the decay rate $\int \Gamma^* \circ \xi'$. Thus

$$\inf_{\xi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)} \int \Gamma^* \circ \xi' = \inf_{\xi \in \mathcal{P}^K \cap \mathcal{U}_\epsilon(\psi)} \int \Gamma^* \circ \xi' \leq \int \Gamma^* \circ f'$$

We have a uniform bound for the infimum (uniform in ϵ, J, ψ) which implies

$$I(f) = \sup_{\substack{\epsilon \in \mathbb{R},\, J \in \mathbb{N} \\ \psi \in \mathcal{P}^J,\, f \in \mathcal{U}_\epsilon(\psi)}} \inf_{\xi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)} \int \Gamma^* \circ \xi' \leq \sup_{\substack{\epsilon \in \mathbb{R},\, J \in \mathbb{N} \\ \psi \in \mathcal{P}^J,\, f \in \mathcal{U}_\epsilon(\psi)}} \int \Gamma^* \circ f' \leq \int \Gamma^* \circ f'$$

matching the lower bound of $I(f)$. □ 4.3.7

We want this to be the general form of the rate function.

Claim 4.3.8 (Rate function identification). *$I(\psi) = \int_{s=0}^T \Gamma^*(\psi'(s))\, ds$ for $\psi \in AC[0, T]$.*

Proof of 4.3.8: Let ψ be absolutely continuous and $\psi \notin \bigcup_{J \in \mathbb{N}} \mathcal{P}^J$. We investigate again

$$\begin{aligned} I(f) &= \sup_{U \in \mathcal{A}, f \in U} - \lim_{n \to \infty} \frac{1}{n} \log P(N_n \in U) \\ &= \sup_{\substack{\epsilon \in \mathbb{R},\, J \in \mathbb{N} \\ \psi \in \mathcal{P}^J,\, f \in \mathcal{U}_\epsilon(\psi)}} \inf_{\xi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(\psi)} \int \Gamma^* \circ \xi' \end{aligned}$$

Let f^J be approximations of f in $\mathcal{P}^J$ (cf 4.2.10) and assume that $\lim_{J \to \infty} I(f^J) < \infty$ and let $\gamma > 0$ be some small number. Choose J large enough for

$$|I(f^J) - \lim_{K \to \infty} I(f^K)| < \frac{\gamma}{2}$$

to hold and $\epsilon > 0$ small enough for

$$|I(f^J) - \inf_{\xi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(f^J)} I(\xi)| < \frac{\gamma}{2}$$

to hold. We then get

$$|\inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(f^K)} I(\xi) - \lim_{K\to\infty} I(f^K)| < \gamma$$

With J, ϵ chosen according to γ and now fixed we have

$$I(f) \geq \inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(f^J)} I(\xi) \geq \lim_{K\to\infty} I(f^K) - \gamma$$

And since γ was arbitrarily small

$$I(f) \geq \lim_{K\to\infty} I(f^K) \qquad (=: I)$$

Now lets assume the inequality was strict: that is the following defines a positive number.

$$\gamma' := \sup_{\substack{\epsilon\in\mathbb{R}\,,\,J\in\mathbb{N}\\ \psi\in\mathcal{P}^J\,,\,\psi\in\mathcal{U}_\epsilon(f)}} \inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I = \sup_{\substack{\epsilon\in\mathbb{R}\,,\,J\in\mathbb{N}\\ \psi\in\mathcal{P}^J\,,\,\psi\in\mathcal{U}_\epsilon(f)}} \big(\inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I\big)$$

There are basically two possibilities to choose ϵ, J, ψ in the supremum. We investigate the argument of the supremum in both cases.
1st case. ϵ, J, ψ are chosen such that $f^J \in \mathcal{U}_\epsilon(\psi)$ then $\inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(\psi)} I(\xi) \leq I(f^J)$ and

$$\inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I = \underbrace{\inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I(f^J)}_{\leq 0} + \underbrace{I(f^J) - I}_{\leq 0} \leq 0$$

2nd case. If on the other hand ϵ, J, ψ are such that $f^J \notin \mathcal{U}_\epsilon(\psi)$ then we have $\psi \in \mathcal{P}^J \cap \mathcal{U}_\epsilon(f) \cap \mathcal{U}_\epsilon(f^J)^c$. That is ψ is an element of $\mathcal{P}^J$ that is closer in the sup-norm to f than f^J.
Let $||f-\psi|| =: \delta < \epsilon$ and K large enough for $||f^K - f|| < \epsilon - \delta$ to hold. Then $||f^K - \psi|| < \epsilon$ and

$$\begin{aligned} \inf_{\xi\in\mathcal{P}^J\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I &= \inf_{\xi\in\mathcal{P}^K\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I \\ &= \underbrace{\inf_{\xi\in\mathcal{P}^K\cap\mathcal{U}_\epsilon(\psi)} I(\xi) - I(f^K)}_{\leq 0} + \underbrace{I(f^K) - I}_{\leq 0} \leq 0 \end{aligned}$$

So $\gamma' \leq 0$. □ 4.3.8

Since Γ^* is convex the rate function I is convex, too.

Looking at [5] and their proof of the sample path LDP for the partial sums process in lemma 5.1.6 (p. 181, top) we see that the rate function I is concentrated on absolutely continuous functions.

4.3.2 The full large deviation principle

We can now strengthen the weak large deviation principle to a full one.

Claim 4.3.9. *In the space of continuous functions $C([0,T],\mathbb{R})$ equipped with the sup-norm induced topology the interpolated renewal counting process $\hat{N}$ (under the scaling $\hat{N}_n : t \mapsto \frac{1}{n}\hat{N}(nt)$) satisfies: for any open set G and any closed set F*

$$\begin{aligned} -\inf_{f\in G} I(f) \leq \quad & \liminf_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}(\hat{N}_n \in G) \\ & \limsup_{n\to\infty} \tfrac{1}{n}\log \mathbb{P}(\hat{N}_n \in F) \quad \leq -\inf_{f\in F} I(f) \end{aligned}$$

with the good convex rate function

$$I(f) = \begin{cases} \int_{t=0}^{T} \Gamma^* \circ f'(t)\,dt & \text{if } f \in AC([0,T],\mathbb{R}), f(0) = 0 \\ \infty & \text{else.} \end{cases}$$

Proof of 4.3.9: From the weak LDP the full LDP follows as soon as we have a good rate function in the weak LDP. As we are in a Polish space goodness of the rate function is the compactness of its level sets.

We apply the Arzela-Ascoli theorem (cf 7.5.1) to identify level sets

$$\mathcal{L}(c) = \{f \mid I(f) \leq c\}$$

with $c \geq 0$ as compact subsets of $C([0,T],\mathbb{R})$.

- Closedness of the level set is given as a property of I being a rate function.
- $I(f) \leq c$ implies $f(0) = 0$, so initial points of elements of $\mathcal{L}(c)$ are bounded.
- It remains to be shown that

$$\forall(t,\epsilon)\,\exists\,\delta > 0 \,:\, |t-s| < \delta \,\Rightarrow\, \sup_{f\in\mathcal{L}(c)} |f(t) - f(s)| < \epsilon$$

Remember 4.2.11 and let $f \in \mathcal{L}(c)$, $J \in \mathbb{N}$, and f^J its approximation with piecewise constant derivatives $v_1^J, \ldots, v_{2^J}^J$: $v_k = \frac{1}{T}\,2^J \int_{t=T\,2^{-J}\,(k-1)}^{T\,2^{-J}\,k} f'(t)\,dt$.

(Note that only 2^{-J} is a power and the J in f^J, v_k^J is an index)

$$\begin{aligned}
& T\,2^{-J}\sum_{k=1}^{2^J}\Gamma^*(v_k^J) = I(f^J) \le I(f) \le c \text{ for each } J \\
\Rightarrow \quad & \max_{k=1,\dots,2^J} T\,2^{-J}\,\Gamma^*(v_k^J) \le c \text{ for each } J \\
\Rightarrow \quad & \sup_{J\in\mathbb{N},k=1,\dots,2^J} T\,2^{-J}\,\Gamma^*(v_k^J) \le c \\
\Leftrightarrow \quad & \sup_{t,s<t}(t-s)\,\Gamma^*(\frac{f(t)-f(s)}{t-s}) \le c
\end{aligned}$$

Fix $t,\epsilon > 0$ and let $m \ge \frac{c}{\epsilon}$. From $\lim_{x\to\infty}\frac{\Gamma^*(x)}{x} = \infty$ let M be such that $\frac{\Gamma^*(x)}{x} > m$ for $x \ge M$. Set $\delta = \frac{c}{\epsilon}$

$$\frac{f(t)-f(s)}{t-s} < M \quad \Rightarrow \quad f(t)-f(s) \le M(t-s) \le \epsilon \text{ for } |t-s| < \delta$$

otherwise

$$\begin{aligned}
\frac{f(t)-f(s)}{t-s} &\ge M \\
\Rightarrow \quad (t-s)\,\Gamma^*(\frac{f(t)-f(s)}{t-s}) &> m(f(t)-f(s)) \\
\overset{f\in\mathcal{L}(c)}{\Rightarrow} \quad c \ge (t-s)\,\Gamma^*(\frac{f(t)-f(s)}{t-s}) &> m(f(t)-f(s)) \\
\Rightarrow \quad c &\ge m(f(t)-f(s)) \\
\Gamma^*(\frac{f(t)-f(s)}{t-s}) &> m(f(t)-f(s)) \\
\Rightarrow \quad \epsilon = \frac{c}{m} &\ge f(t)-f(s)
\end{aligned}$$

$\square$
4.3.9

For an alternative proof see [23], lemma 5.18. There compactness of level sets is proved for the Poisson process, which is the renewal process with exponential inter event times. It works exactly the same for all rcp in the scope of this thesis; especially the no-point mass in $\{0\}$-property of assumption 2.2.2 is important since it is equivalent to $\Lambda^*(0) = \infty$ which is equivalent to $\lim_{x\to\infty}\frac{\Gamma^*(x)}{x} = \infty$.

Corollary 4.3.10. *In the space $D([0,T],\mathbb{R})$ equipped with the sup-norm induced topology the sequence of scaled renewal counting processes $(N_n\,;\, n\in\mathbb{N})$ satisfies a sample path large deviation principle with the good, convex rate function $I(.)$ of 4.3.9.*

The corollary follows from claim 4.3.9 by an application of lemma 4.1.5 and theorem 4.2.13 of Dembo and Zeitouni [5].

4.3.3 Interpretation

We denote Γ^* the local rate function for the large deviation principle of the counting process. Note that with reference to *Big Queues* [9] (section 6.2 definition 6.1, p .99) we now do have linear geodesics for the counting process.

Given some fixed $\psi \in AC[0,T]$ with $\psi' \geq 0$ where it exists, what can we do?

- Approximate ψ by ψ^J for some large J and denote $v \in \mathbb{R}^{2^J}$ the vector of slopes of ψ^J. The rate function $I(\psi)$ is well approximated by the finite sum $T\,2^{-J}\sum_{k=1}^{2^J}\Gamma^*(v_k)$. Each summand has $\Gamma^*(v_k) = \theta(v_k)\cdot v_k - \Gamma^*(\theta(v_k))$ where $\theta(v_k)$ as a twist parameter makes v_k the expectation of $\lim_{n\to\infty} N_n(1)$ under the twisted measure $\theta(v_k)$.
- Simulate a counting proccess that (over a long time) behaves different from what would be expected in terms of its empirical mean.
- Find the most likely paths for certain events. The most likely path to a “too large” or “too small” value at some fixed time will be along some piecewise linear function with only two different slopes. This is a typical application of linear geodesics.
- Apply the contraction principle and solve the associated variational problem.

4.4 Split counting processes

This is a first step towards networks. In a network of $d \in \mathbb{N}$ nodes customers leaving a node $i \in \{1,\ldots,d\}$ may be routed to another network node or may leave the network, cf figure 4.7. The queue of customers waiting for service at node i may be non-empty over a period of time. During this time times between subsequent departures are the customers service times. Equivalently: if over an interval of time the queue at node i is never empty then increments of the departure process from this queue are increments of the service process at this queue.

When describing the network and how it evolves in time we want to document where customers departing from a node go to next. As customers

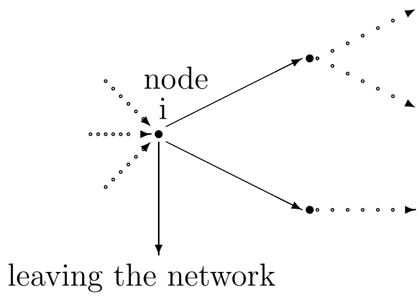


Figure 4.7: Possible routing of customers leaving node i

leaving queue i are routed into different directions on a technical level we are splitting the service process.

We will then need a linear transformation to describe the number of customers leaving node i *and* the number of customers being routed to other network nodes as a vector-valued process in time. Customers leaving the network will not be counted they only appear as leaving node i. With an additional condition the linear transformation we apply is a bijection.

In this section we develop the large deviations for a split renewal counting process under a linear transformation. We start in 4.4.1with constructing the split process and calculating its lmgf. In another subsection we give the explicit linear transformation we will later apply in the generalised Jackson network. We continue in 4.4.2 with an exponential change of measure that transforms the split rcp into another split rcp. We also give the change of measure explicitly for the linearly transformed split process we will apply in the network setting. Finally, in 4.4.3 we develop the full sample path large deviations principle for the split and the linearly transformed split counting process..

The examples we give in this section fit the example-network we will work with in chapter 6.

4.4.1 Construction of the split process

Let us split a counting process N into m processes $N^{(1)}, \ldots, N^{(m)}$ with the property that $\sum_{j=1}^{m} N_t^{(j)} = N_t$ for all $t \geq 0$ and that $(N^{(1)}, \ldots, N^{(m)})$ as an m-tupel changes state iff N does. If it changes state it will be by increasing one coordinate by 1.

Definition 4.4.1 (Split rcp N^{sp} associated with N, p)**.** *Let N be a rcp and $p = (p_1, \ldots, p_m)$ for some $m \in \mathbb{N}$ such that $\mathbb{P}(r = e_i) = p_i$ defines the distribution of r on $\mathbb{R}^m$ (e_i is the i-th standard base vector in $\mathbb{R}^m$). Let $r, r_1, r_2, \ldots$ be iid and define*

$$N_t^{sp} := \sum_{i=1}^{N_t} r_i \quad (\in \mathbb{R}^m).$$

The coordinates of the split process N^{sp} are renewal counting processes and we will identify $t \mapsto \sum_{i=1}^{N_t} r_i$ and $\begin{bmatrix} N^{(1)} \\ \vdots \\ N^{(m)} \end{bmatrix}$.

Definition 4.4.2 (Scaled split process)**.** *For the split process N^{sp} defined in 4.4.1 define the scaled split process N_n^{sp} and*

$$N_n^{sp} \quad : \quad t \mapsto \frac{1}{n} N_{nt}^{sp} = \frac{1}{n} \sum_{i=1}^{N_{nt}} r_i$$

And we can identify the scaled split process with the split process of all its coordinates scaled as in 3.4.1:

$$N_n^{\mathrm{sp}} = \begin{bmatrix} N_n^{(1)} \\ \vdots \\ N_n^{(m)} \end{bmatrix} \quad , \quad N_n^{\mathrm{sp}}(t) = \begin{bmatrix} \frac{1}{n} N^{(1)}(nt) \\ \vdots \\ \frac{1}{n} N^{(m)}(nt) \end{bmatrix}$$

Example 4.4.3. *Let $m = 5$ and $p = (0, \frac{1}{2}, \frac{1}{2}, 0, 0)$. Then $N^{sp} = \begin{bmatrix} 0 \\ N^{(2)} \\ N^{(3)} \\ 0 \\ 0 \end{bmatrix}$.*
Figure 4.8 shows a realisation of N and the coordinate processes $N^{(2)}$ and $N^{(3)}$ of N^{sp}. Note that $N = N^{(2)} + N^{(3)}$ in the figure. Inter event times of N are uniformly distributed on $(0, 2)$.

We define the lmgf for a discrete probability measure p and then the lmgf for the split counting process.

Definition 4.4.4. *The logarithmic moment generating function of r with $\mathbb{P}(r = e_i) = p_i$ for $i = 1, \ldots, m$ is denoted K (as a capital Greek letter). For $\theta \in \mathbb{R}^m$*

$$K(\theta) = \log E[e^{\langle \theta, r \rangle}] \;=\; \log \sum_{j=1}^{m} e^{\langle \theta, e_j \rangle} \, p_j \;=\; \log \sum_{j=1}^{m} e^{\theta_j} \, p_j.$$

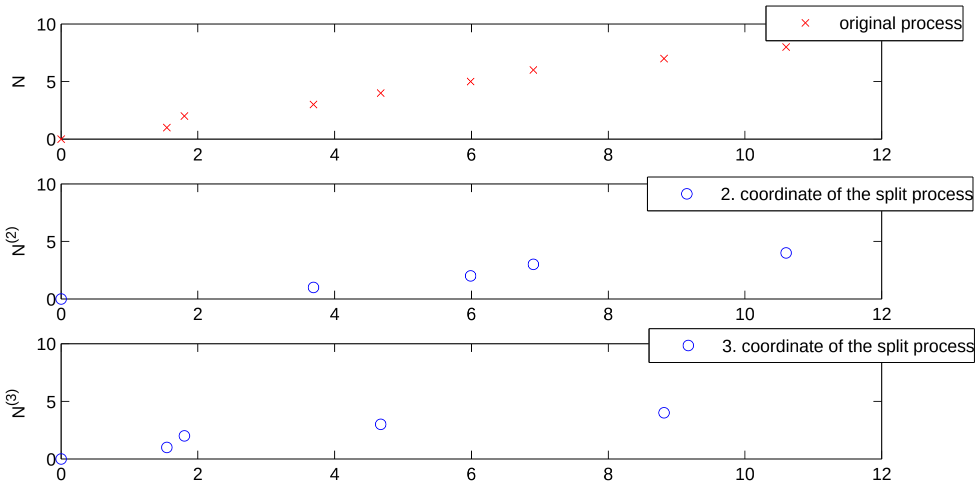


Figure 4.8: Realisation of N and coordinates of N^{sp} of example 4.4.3.

It is $\mathcal{D}(K) = \mathbb{R}^m$ either from the explicit form of the lmgf or from boundedness of r and by 2.6.2 the Fenchel-Legendre transform K^* has compact level sets.

Claim 4.4.5. *If the counting process N with lmgf Γ is split wrt the probability measure p with lmgf K then the lmgf of the split process N^{sp} associated with N and p is*

$$\lim_{t\to\infty} \frac{1}{t} \log E[e^{\langle \theta \, , \, N_t^{sp} \rangle}] \;=\; \Gamma \circ K(\theta).$$

Proof of 4.4.5: The split process is a time change of the partial sum $n \mapsto \sum_{k=1}^{n} r_k$ replacing n by N_t resulting in $t \mapsto \sum_{k=1}^{N_t} r_k$. Note that coordinates of $\sum_{k=1}^{n} r_k$ are binomially distributed.

We calculate exponential moments for the split process. Let $\theta \in \mathbb{R}^m$.

$$\begin{aligned}
E[e^{\langle \theta\,,\,N_t^{\mathrm{sp}}\rangle}] &= E[e^{\langle \theta\,,\,\sum_{k=1}^{N_t} r_k\rangle}] \\
&= E[e^{\langle \theta\,,\,\sum_{k=1}^{n} r_k\rangle} \sum_{n=0}^{\infty} \mathbb{1}_{N_t=n}] \\
&= \sum_{n=0}^{\infty} \sum_{\substack{i_1,\dots,i_m \\ \sum_{l=1}^m i_l = n}} \underbrace{P(N_t = n\,,\, \sum_{k=1}^{n} r_k = \begin{pmatrix} i_1 \\ \vdots \\ i_m \end{pmatrix})}_{=P(N_t=n)\,P(\sum_{k=1}^n r_k = \dots)}\, e^{\sum_{j=1}^m \theta_j\, i_j} \\
&= \sum_{n=0}^{\infty} P(N_t = n) \sum_{\substack{i_1,\dots,i_m \\ \sum_{l=1}^m i_l = n}} \frac{n!}{i_1!\, i_2! \cdots i_m!} \underbrace{\prod_{j=1}^{m} p_j^{i_j}\, e^{\sum_{j=1}^m \theta_j\, i_j}}_{=\prod_{j=1}^m (p_j\, e^{\theta_j})^{i_j}} \\
&= \sum_{n=0}^{\infty} P(N_t = n) (\sum_{j=1}^{m} p_j\, e^{\theta_j})^n \\
&= E[\exp\{N_t \log \sum_{j=1}^{m} p_j\, e^{\theta_j}\}]
\end{aligned}$$

We take the scaled limit to obtain the lmgf for the split process.

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \log E[e^{\langle \theta\,,\,N_{nt}^{\mathrm{sp}}\rangle}] &= \lim_{n\to\infty} \frac{1}{n} \log E[\exp\{N_{nt} \log \sum_{j=1}^{m} p_j\, e^{\theta_j}\}] \\
= \; t\,\Gamma\big(\log \sum_{j=1}^{m} p_j\, e^{\theta_j}\big) &= t\,\Gamma \circ K(\theta)
\end{aligned}$$

$\square$
4.4.5

We'll later need a linear transformation of the split process N^{sp}.

Remark 4.4.6. *Generally, given a random variable $X \in \mathbb{R}^m$ with lmgf K and $\mathbb{T}X$ a linear transformation of X the lmgf of $\mathbb{T}X$ is $\theta \mapsto K(\mathbb{T}^\top \theta)$. Especially we might use a linear transformation to only work with a subset $M \subseteq \{1, \dots, m\}$ of coordinates. Then*

$$\mathbb{T} = diag(\mathbb{1}_{1\in M}, \dots, \mathbb{1}_{m\in M}) = \mathbb{T}^\top$$

and thus

$$\begin{aligned} K \circ \mathbb{T}^\top(\theta) &= K(\sum_{k\in M} \theta_k\, e_k) = \log\Big(\sum_{k\in M} p_k\, e^{\theta_k} + \sum_{k\notin M} p_k\Big) \\ &= \log\Big(\sum_{k\in M} p_k\, e^{\theta_k} + 1 - \sum_{k\in K} p_k\Big) \end{aligned}$$

which is the Ξ-style lmgf of a sub-probability measure $p\big|_M$.

Application to the network

We now give the explicit linear transformation we need in the network setting and calculate the lmgf of the linearly transformed split counting process.

If a vector $y \in \mathbb{R}^{d+1}$ describes the next destinations of customers leaving node $i = 1$ over an interval of time

- with y_{d+1} the customers that have left the network at departure from node $i = 1$
- and customers cannot immediately join the same queue again

then

$$\begin{pmatrix} -\sum_{k=1}^{d+1} y_k \\ y_2 \\ \vdots \\ y_d \end{pmatrix} \in \mathbb{R}^d$$

has as its first coordinate the number of customers that have left the first node $i = 1$. Remaining coordinates $j = 2, \dots, d$ are the number of customers who have gone from node $i = 1$ to node j. This formalises as a linear transformation

$$\mathbb{T} = \begin{pmatrix} -1 & \dots & \dots & -1 \\ 0 & 1 & \dots & \dots \\ \dots & & & \dots \\ 0 & \dots & 1 & 0 \end{pmatrix} \quad , \quad \mathbb{T} : \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_{d+1} \end{pmatrix} \mapsto \begin{pmatrix} -\sum_{k=1}^{d+1} y_k \\ y_2 \\ \vdots \\ y_d \end{pmatrix}$$

When working with a network of d nodes we need a family $(\mathbb{T}^{(i)})_{i=1,\dots,d}$ of such transformations, one at each node. The $\mathbb{T}$ just given would be $\mathbb{T}^{(1)}$.

Definition 4.4.7 (Transformation $\mathbb{T}^{(i)}$). *Let $d, i \in \mathbb{Z}$, $d \geq 2$, $i \leq d$ and $\{e_1, \ldots, e_d\}$ the standard base of $\mathbb{R}^d$.*

$$\mathbb{T}^{(i)} : \mathbb{R}^{d+1} \to \mathbb{R}^d \quad , \quad y \mapsto \sum_{\substack{k=1 \\ k \neq i}}^{d} y_k \, e_k - \Big(\sum_{k=1}^{d+1} y_k \Big) \, e_i$$

For example $\mathbb{T}^{(2)}$ transforms

$$\mathbb{T}^{(2)} : \quad \begin{pmatrix} y_1 \\ \vdots \\ y_{d+1} \end{pmatrix} \mapsto \sum_{\substack{k=1 \\ k \neq 2}}^{d} y_k \, e_k - \Big(\sum_{k=1}^{d+1} y_k \Big) \, e_2 = \begin{pmatrix} y_1 \\ -\sum_{k=1}^{d+1} y_k \\ y_3 \\ \vdots \\ y_d \end{pmatrix}$$

and $\mathbb{T}^{(2)}$ can be identified with the $d \times (d+1)$ matrix

$$\begin{bmatrix} 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ -1 & -1 & \ldots & -1 & -1 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix} .$$

As we did split the counting process it was with respect to a probability measure p with lmgf K: all departing customers were considered. Now, after the transformation we only consider customers *staying in the network* and how they are split onto the other nodes: The measure we use becomes a subprobability measure. In the following we define a lmgf Ξ for a sub-probability measure.

Definition 4.4.8. *Let p be a sub-probability distribution on $\{1, \ldots, d\}$ set $p_0 = 1 - \sum_{k=1}^{d} p_k$ and define*

$$\Xi : \mathbb{R}^d \to \mathbb{R} \quad , \quad \xi \mapsto \log \Big(\sum_{k=1}^{d} p_k \, e^{\theta_k} + p_0 \Big)$$

Since $\Xi(\theta) < \infty$ for all $\theta \in \mathbb{R}^d$ the Fenchel-Legendre transform Ξ^* has compact level sets (cf 2.6.2).

Claim 4.4.9. *Let $\Xi^{(i)}$ be associated with the sub-probability measure $p^{(i)}$ on $\{e_1,\dots,e_d\}$ with $p_{ii}=0$. Let N be a counting process with lmgf Γ and let N^{sp} be the split process associated with N and the unique probability measure on $\{e_1,\dots,e_d,0\}$ associated with $p^{(i)}$. The lmgf of the linearly transformed split process is*

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\langle\xi\,,\,\mathbb{T}^{(i)}N_t^{sp}\rangle}]=\Gamma(-\xi_i+\Xi^{(i)}(\xi))$$

Proof of 4.4.9: If $p^{(i)}=(p_{i1},\dots,p_{id})$ is a sub-probability measure then set $p_{i0}=1-\sum_{j=1}^d p_{ij}$ and let K be associated with this probability measure. Applying $p_{ii}=0$ we get

$$\Xi^{(i)}(\xi)=\xi_i+K(\mathbb{T}^{(i)\top}\xi)$$

from calculating

$$\begin{aligned}
K(\mathbb{T}^{(1)\top}\xi) \quad &= \quad K\left(-\xi_1\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}+\begin{bmatrix}0\\ \xi_2\\ \vdots\\ \xi_d\\ 0\end{bmatrix}\right)\\
&\overset{p_{ii}=p_{11}=0}{=} \quad \log\Big(\sum_{j=1}^d p_{ij}\,e^{-\xi_1+\xi_j}+p_{i0}\,e^{-\xi_1}\Big)\\
&= \quad -\xi_1+\log\Big(\sum_{j=1}^d p_{ij}\,e^{\xi_j}+p_{i0}\Big)\\
&= \quad -\xi_1+\Xi^{(1)}(\xi)
\end{aligned}$$

and apply the adaption of the lmgf to a linear transformation of the split process 4.4.5 to obtain

$$\begin{aligned}
\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\langle\xi\,,\,\mathbb{T}^{(i)}N_t^{\mathrm{sp}}\rangle}] \;&=\; \lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\langle\mathbb{T}^{(i)\top}\xi\,,\,N_t^{\mathrm{sp}}\rangle}]\\
&=\; \Gamma\circ K(\mathbb{T}^{(i)\top}\xi)\;=\;\Gamma(-\xi_i+\Xi^{(i)}(\xi)).
\end{aligned}$$

□
4.4.9

Remark 4.4.10. *The transformation $\mathbb{T}^{(i)}$ defined in 4.4.7 is a regular transformation between d dimensional spaces when defined on*

$$\mathbb{T}^{(i)}\,:\,\mathbb{R}^{d+1}\cap\{x\,|\,x_i=0\}\to\mathbb{R}^d.$$

The inverse transformation is

$$\mathbb{R}^d \to \mathbb{R}^{d+1} \cap \{x \mid x_i = 0\} \ , \ \xi \mapsto \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_d \\ -\sum_{k=1}^d \xi_k \end{bmatrix} - \xi_i \, e_i .$$

Conclusion 4.4.11. *If N is split wrt $p^{(i)} = (p_{1i}, \dots, p_{di})$ with $p_{ii} = 0$ then the transformation between N^{sp} and $\mathbb{T}^{(i)} N^{sp}$ is a linear bijection.*

When splitting a counting process wrt some probability measure $p^{(i)}$ with $p_{ii} = 0$ we'll have $N_t^{(i)} \equiv 0$ and the split process $N_t^{\mathrm{sp}} \in \mathbb{R}^{d+1} \cap \{x \mid x_i = 0\}$.

We continue the example 4.4.3 where we had split a counting process: We now apply a linear transformation.

Example 4.4.12. *The transformation $\mathbb{T} = \mathbb{T}^{(1)}$ applied to the split process* $\begin{bmatrix} 0 \\ N^{(2)} \\ N^{(3)} \\ 0 \\ 0 \end{bmatrix}$ *results in* $\begin{bmatrix} -N \\ N^{(2)} \\ N^{(3)} \\ 0 \end{bmatrix}$.

The transformation works a little different if the last coordinate process is not $\equiv 0$.

Example 4.4.13. *Let $m = 5$, $p = (\frac{1}{4}, \frac{1}{4}, 0, 0, \frac{1}{2})$, and the transformation $\mathbb{T} = \mathbb{T}^{(3)}$*

$$N \mapsto N^{sp} \mapsto \mathbb{T}^{(3)} N^{sp} \qquad : \qquad N \overset{split}{\mapsto} \begin{bmatrix} N^{(1)} \\ N^{(2)} \\ 0 \\ 0 \\ N^{(5)} \end{bmatrix} \overset{\mathbb{T}^{(3)}}{\mapsto} \begin{bmatrix} N^{(1)} \\ N^{(2)} \\ -N \\ 0 \end{bmatrix}$$

All examples will appear in the later example network of $d = 4$ nodes.

4.4.2 Change of measure for the split process

Changing the split processes means changing the unsplit process and the Bernoulli random variable reigning the routing. Since we constructed the counting process N and the routing variables $r_1, r_2, \dots$ to be independent we multiply mass functions. First each variable is twisted exponentially.

We apply the exponential twist of 2.3.1 to r with the lmgf K defined in 4.4.4. For measurable $A \subset \mathbb{R}^m$

$$\begin{aligned}\mathbb{P}(r \in A) &= \sum_{k\,:\,e_k \in A} p_k \\ \mathbb{P}^{(\theta)}(r \in A) &= \sum_{k\,:\,e_k \in A} e^{\langle \theta, e_k\rangle - K(\theta)}\, p_k \;=\; \sum_{k\,:\,e_k \in A} \frac{e^{\theta_k}}{\mathbb{E}[e^{\langle \theta, r\rangle}]}\, p_k\end{aligned}$$

We may also write

$$\mathbb{P}^{(\theta)}(r \in A) \;=\; \mathbb{E}[\mathbb{1}_{r \in A} \frac{e^{\langle r, \theta\rangle}}{\mathbb{E}[e^{\langle \theta, r\rangle}]}]$$

and we identify the change of measure

$$r \mapsto e^{\langle r,\theta\rangle - K(\theta)} \;=\; \frac{e^{\langle r,\theta\rangle}}{\mathbb{E}[e^{\langle \theta, r\rangle}]}$$

The sum $\sum_{k=1}^n r_k$ has lmgf nK by independence of $r_1, \ldots, r_n$ and the density of the twisted distribution wrt the original distribution is

$$\sum_{k=1}^n r_k \mapsto \prod_{k=1}^n e^{\langle r_k, \theta\rangle - K(\theta)} \;=\; e^{\langle \sum_{k=1}^n r_k\,,\, \theta\rangle - nK(\theta)}$$

and summands remain independent under the new measure.

The change of measure for the counting process was given in definition 3.6.4 and (3.15). We combine both changes of measure on the product space for N_t and $\sum_{k=1}^n r_k$:

Claim 4.4.14. *Let N be a counting process with lmgf Γ and $r \in \mathbb{R}^m$ a routing variable with mass function p and lmgf K. Then for $\theta \in \mathcal{D}(K) = \mathbb{R}^m$*

$$(\theta, t) \mapsto \exp\{\langle N_t^{sp}, \theta\rangle - t\Gamma \circ K(\theta)\}\, r(t, -\Gamma \circ K(\theta))$$

is a change of measure process for the split process constructed from N and p. The $r(\cdot,\cdot)$ appearing in the change of measure is a random function defined in (3.6.5) referring to the distribution function F of inter event times of N and the age B (and $r \neq r(\cdot,\cdot)$).

Proof of 4.4.14: Since N and $\sum_{k=1}^{n} r_k$ are independent we twist them individually with $\zeta \in \mathbb{R}$ the twist parameter for N and $\theta \in \mathbb{R}^m$ the twist parameter for $\sum_{k=1}^{n} r_k$. Let $j \in \mathbb{N}$ and $x \in \mathbb{N}^m$.

$$\begin{aligned}
&\mathbb{P}^{\theta[\zeta]}(N_t = j\,,\ \sum_{k=1}^{n} r_k = x) \\
&= \mathbb{E}^{\theta[\zeta]}[\,\mathbb{1}_{N_t=j}\,\mathbb{1}_{\sum_{k=1}^{n} r_k = x}\,] \\
&= \mathbb{E}^{\theta}[\,\mathbb{1}_{N_t=j}\,e^{\zeta j - t\Gamma(\zeta)}\frac{F^c_{-\Gamma(\zeta)}}{F^c}(B(t))\,e^{B(t)\,\Gamma(\zeta)}\ \mathbb{1}_{\sum_{k=1}^{n} r_k = x}\,] \\
&= \mathbb{E}[\,\mathbb{1}_{N_t=j}\,e^{\zeta j - t\Gamma(\zeta)}\frac{F^c_{-\Gamma(\zeta)}}{F^c}(B(t))\,e^{B(t)\,\Gamma(\zeta)}\ \mathbb{1}_{\sum_{k=1}^{n} r_k = x}\,e^{\langle\theta,x\rangle - nK(\theta)}\,]
\end{aligned}$$

The following is for the case of $n = j$ and $\zeta = K(\theta)$.

$$\begin{aligned}
&\mathbb{P}^{\theta[K(\theta)]}(N_t = n\,,\ \sum_{k=1}^{n} r_k = x) \\
&= \mathbb{E}[\,\mathbb{1}_{N_t=n}\,e^{K(\theta)\,n - t\Gamma\circ K(\theta)}\frac{F^c_{-\Gamma\circ K(\theta)}}{F^c}(B(t))\,e^{B(t)\,\Gamma\circ K(\theta)}\ \mathbb{1}_{\sum_{k=1}^{n} r_k = x}\,e^{\langle\theta,x\rangle - nK(\theta)}\,] \\
&= \mathbb{E}[\,\mathbb{1}_{N_t=n}\,\mathbb{1}_{\sum_{k=1}^{n} r_k = x}\,e^{\langle\theta,x\rangle - t\Gamma\circ K(\theta)}\frac{F^c_{-\Gamma\circ K(\theta)}}{F^c}(B(t))\,e^{B(t)\,\Gamma\circ K(\theta)}\,]
\end{aligned}$$

We apply this to the split counting process $N^{\mathrm{sp}}(t) = \sum_{k=1}^{N(t)} r_k$. Fix an arbitrary $x \in \mathbb{N}^m$ and $n = n(x) = \sum_{k=1}^{m} x_k$. Then

$$\mathbb{P}(N^{\mathrm{sp}}(t) = x) \;=\; \mathbb{P}(N_t = n\,,\ \sum_{k=1}^{n} r_k = x)$$

and

$$\begin{aligned}
&\mathbb{P}^{\theta,[K(\theta)]}(N_t = n\,,\ \sum_{k=1}^{n} r_k = x) \\
&= \mathbb{E}[\,\mathbb{1}_{N_t^{\mathrm{sp}}=x}\,e^{\langle\theta,x\rangle - t\Gamma\circ K(\theta)}\frac{F^c_{-\Gamma\circ K(\theta)}}{F^c}(B(t))\,e^{B(t)\,\Gamma\circ K(\theta)}\,]
\end{aligned}$$

Which identifies the claimed density process. $\square$
4.4.14

Application to the network

We rewrite the change of measure process to fit our transformed split counting process. Let $\theta \in \mathbb{R}^{d+1}$ and $\mathbb{T} = \mathbb{T}^{(1)}$ for notational simplicity.

$$\begin{aligned} \langle N_t^{\mathrm{sp}}, \theta \rangle &= \langle \mathbb{T}\, N_t^{\mathrm{sp}}\,,\, (\mathbb{T}^\top)^{-1}\theta \rangle \\ \xi &:= (\mathbb{T}^\top)^{-1}\theta \;=\; (\mathbb{T}^{-1})^\top \theta \;=\; \begin{bmatrix} 0 \\ \theta_2 \\ \vdots \\ \theta_d \end{bmatrix} - \theta_{d+1} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \end{aligned}$$

with this definition $\theta = \mathbb{T}^\top \xi$ and $K(\theta) = \Xi^{(1)}(\xi) - \xi_1$ from definition 4.4.8. We rewrite the change of measure 4.4.14 and denote it $\mathcal{G}^{(i)}$.

Definition 4.4.15 ($\mathcal{G}^{(i)}$). *For*

- *a renewal counting process N*
- *a subprobability $p^{(i)} = (p_{i1}, \ldots, p_{id})$ with $p_{ii} = 0$ and lmgf Ξ*
- *the split process N^{sp} associated with N and $(p_{i1}, \ldots, p_{id}, 1 - \sum_{j=1}^{d} p_{ij})$*
- *a linear transformation $\mathbb{T}^{(i)}$*

we define the change of measure process for $\mathbb{T}^{(i)} N^{sp}$ with parameter $\xi \in \mathbb{R}^d$ as

$$\mathcal{G}^{(i)}(\xi, t) = \exp\{\langle \mathbb{T}^{(i)} N_t^{sp}, \xi \rangle - t\Gamma(-\xi_i + \Xi^{(i)}(\xi))\}\, r(t, -\Gamma(-\xi_i + \Xi^{(i)}(\xi))).$$

We give a summary of this subsection 4.4.2:

Corollary 4.4.16. *Let N be a counting process with inter event times density f and lmgf Γ that is split into $d+1$ processes wrt a probability measure p. Let $(p_1, \ldots, p_d)$ be the sub-probability measure associated with p with lmgf Ξ. Then under the change of measure $\mathcal{G}^{(i)}$ the new process is distributed like the split process of iid inter event times with density $f_{-\Gamma(-\xi_1 + \Xi^{(1)}(\xi))}$ and mean $\Gamma'(-\xi_1 + \Xi^{(1)}(\xi))$ and with routing (sub)probabilities $\nabla \Xi^{(i)}(\xi) = p(\xi)$.*

Definition 4.4.17. *For a rcp N with lmgf Γ and fixed i and $\xi \in \mathbb{R}^d$ the function*

$$\xi \mapsto \Gamma'(-\xi_i + \Xi^{(i)}(\xi))$$

gives the rate of N under the change of measure $\mathcal{G}^{(i)}$.

4.4.3 Sample path LDP for the split process

Claim 4.4.18. *Let N^{sp} be the split process associated with the counting process N and the probability measure p. Let N have lmgf Γ and p lmgf K. Then under the scaling of 4.4.2 a sample path large deviation principle holds for N^{sp} in $D([0,T],\mathbb{R}^m)$ equipped with the sup-norm induced topology. The rate function is good, convex and*

$$h \mapsto \int_{t=0}^{T} (\Gamma \circ K)^*(h'(t))\, dt$$

for $h \in AC([0,T],\mathbb{R}^m)$, $h(0)=0$ and $h \mapsto \infty$ otherwise.

Proof of 4.4.18: Set $S_n = \sum_{k=1}^{n} r_k$ and we get a sample path LDP for the split process $t \mapsto N_t^{\mathrm{sp}} = \sum_{k=1}^{N_t} = S \circ N(t)$ similarly to the partial sums process of iid summands in Mogulskii's theorem (cf [5], theorem 5.1.2 with $K(\theta) < \infty$ for all $\theta \in \mathbb{R}^m$). We only sketch it in the following. We already have the lmgf $\Gamma \circ K$ and from restarting the counting process we easily get the finite dimensional lmgfs for $(\frac{1}{n}S \circ N(0), \frac{1}{n}S \circ N(nt_1), \ldots, \frac{1}{n}S \circ N(nT))$. From the Gärtner-Ellis theorem we get the finite dimensional large deviations principle with the rate function as the sum over expressions of $(\Gamma \circ K)^*$.

Applying the projective limit theorem we get the large deviation principle in the continuous functions with the rate function in integral form, the local rate function the Fenchel-Legendre transform $(\Gamma \circ K)^*$. The topology we get is that of pointwise convergence. One can then deduce that the rate function is concentrated on absolutely continuous functions. To obtain the large deviation principle in the sup-norm induced topology we have to prove exponential tightness. This is done in 4.4.19. □
4.4.18

Claim 4.4.19. *The distributions of the scaled split counting process $(N_n^{sp}; n \in \mathbb{N})$ are exponentially tight under the sup-norm induced topology.*

Proof of 4.4.19: The i-th coordinate process for the split process has inter event times $\tau^\circ = \sum_{k=1}^{G} \tau_k$ with geometric G with mass function $P(G=g) = p_i(1-p_i)^{g-1}$. To fit our notation: $p = 1 - p_i$.

By 2.2.11 the lmgf is $\Lambda^\circ(\theta) = \Lambda(\theta) + \log \frac{1-p}{1-pe^{\Lambda(\theta)}}$ and we need the associated Γ, that is $-\Lambda^{\circ -1}(-\theta)$. The inverse of Λ° is

$$x \mapsto \Lambda^{-1}\big(-\log(p + (1-p)\, e^{-x})\big)$$

and we get on the level of the counting process

$$\Gamma^\circ(x) \;=\; -\Lambda^{\circ -1}(-x) \;=\; -\Lambda^{-1}\big(-\log(p + (1-p)\, e^{x})\big).$$

Substituting $p = 1 - p_i$ and applying the definition of Γ through Λ (cf (2.2.7))

$$\Gamma^\circ(x) \quad = \quad \Gamma\big(\log(1 - p_i + p_i\, e^x)\big).$$

This perfectly coincides with the coordinate-wise lmgf of the split counting process: Consider $x\, e_i$ for some $x \in \mathbb{R}$

$$\begin{aligned} \Gamma \circ K \circ \pi_i(x\, e_i) &= \Gamma \circ K(x\, e_i) \;=\; \Gamma\big(\log(p_i\, e^x + \sum_{m \neq i} p_m)\big) \\ &= \Gamma\big(\log(p_i\, e^x + (1 - p_i))\big) \;=\; \Gamma^\circ(x) \end{aligned}$$

For each coordinate process $N^{(k)}$ (for $k = 1, \ldots, d$) a sample path large deviation principle holds (cf. 4.3.9). We apply these to construct a compact set from the level sets of the coordinate functions. Let $\epsilon > 0$.

$$\begin{aligned} \mathcal{L}(\alpha, k) &:= \{f \in C([0,T], \mathbb{R}) \,|\, \int_{t=0}^{T} \Gamma^{\circ(k)*}(f'(t))\, dt < \alpha\} \\ \mathcal{K}(\alpha) &:= \{f \in C([0,T], \mathbb{R}^m) \,|\, f_k \in \mathcal{L}(\alpha + \epsilon, k) \;\text{ for } k = 1, \ldots, m\} \end{aligned}$$

Each $\mathcal{L}(\alpha + \epsilon, k)$ is a compact set in $C([0,T], \mathbb{R})$ by the large deviation principle 4.3.9 with good rate function. $\mathcal{K}(\alpha)$ is a Cartesian product of compact sets and itself compact in the product space $\big(C([0,T], \mathbb{R})\big)^m = C([0,T], \mathbb{R}^m)$.

$$\begin{aligned} P(\hat{N}_n \notin \mathcal{K}(\alpha)) &= P(\hat{N}_n^{(k)} \notin \mathcal{L}(\alpha + \epsilon) \text{ for some } k) \\ &\leq m \max_{k=1,\ldots,m} P(\hat{N}_n^{(k)} \notin \mathcal{L}(\alpha + \epsilon, k)) \end{aligned}$$

The closure of $\mathcal{L}(\alpha + \epsilon, k)^c$ is a subset of $\mathcal{L}(\alpha, k)^c$ and we can apply an alternative formulation of the LDP (cf [DZ] (1.2.7), p.6).

$$\begin{aligned} & \limsup_{n\to\infty} \frac{1}{n} \log P(\hat{N}_n^{(k)} \notin \mathcal{L}(\alpha + \epsilon)) \leq -\alpha \\ \Rightarrow \quad & \limsup_{n\to\infty} \frac{1}{n} \log P(\hat{N}_n^{(k)} \notin \mathcal{K}(\alpha)) \leq -\alpha \end{aligned}$$

which is exponential tightness. □ 4.4.19

Application to the network

We simply restate the claim of the large deviation principle 4.4.18 in terms of the process transformed by $\mathbb{T}$. As usual $d = m - 1$.

Corollary 4.4.20. *There is a sample path large deviation principle for the transformed split process $t \mapsto \mathbb{T}N_t^{sp}$ in $D([0,T],\mathbb{R}^d)$ equipped with the sup-norm induced topology with the good convex rate function*

$$h \mapsto \int_{t=0}^{T} \big(\Gamma \circ (\Xi^{(i)} - \pi_i)\big)^*(h'(t))\,dt$$

for $h \in AC([0,T],\mathbb{R}^d)$, $h(0) = 0$ and $h \mapsto \infty$ otherwise.

Proof of 4.4.20: From 4.4.18 and from $\mathbb{T}^{(i)}$ being a continuous bijection we immediately have the sample path large deviation principle in the same space with rate function

$$h \mapsto \int_{t=0}^{T} \big(\Gamma \circ K\big)^* \circ \mathbb{T}^{-1}(h'(t))\,dt$$

and we only have to identify the local rate functions. Abbreviate $\mathbb{T} = \mathbb{T}^{(i)}$.

$$\begin{aligned}
(\Gamma \circ K)^* \circ \mathbb{T}^{-1}(x) \quad &= \quad \inf_z \Gamma^*(z) + z\,K^*(\frac{\mathbb{T}^{-1}x}{z}) \\
&\overset{\text{linearity of } \mathbb{T}}{=} \quad \inf_z \Gamma^*(z) + z\,K^*(\mathbb{T}^{-1}\frac{x}{z}) \\
&\overset{\text{def. 4.4.8}}{=} \quad \inf_z \Gamma^*(z) + z\,(\Xi^{(i)} - \pi_i)^*(\frac{x}{z}) \\
&= \quad (\Gamma \circ (\Xi^{(i)} - \pi_i))^*(x) \qquad (4.9)
\end{aligned}$$

$\square$
4.4.20

Lemma 4.4.21 (Rate function identification)**.** *For the local rate function in 4.4.20 for the transformed split counting process holds*

$$\big(\Gamma \circ (\Xi^{(i)} - \pi_i)\big)^*(x) \;=\; \inf_{\substack{r^{(i)},\gamma \\ \gamma(r^{(i)} - e_i) = x}} \Gamma^*(\gamma) + \gamma\,\Xi^{(i)*}(r^{(i)})$$

Proof of 4.4.21: We make explicit the Fenchel-Legendre transform (4.9).

$$\begin{aligned}
(\Gamma \circ (\Xi^{(i)} - \pi_i))^*(x) \;&=\; \inf_\gamma \Gamma^*(\gamma) + \gamma\,(\Xi^{(i)} - \pi_i)^*(\frac{1}{\gamma}x) \\
&=\; \inf_\gamma \Gamma^*(\gamma) + \gamma \inf_{b \in \mathbb{R}^d} \Xi^{(i)*}(\frac{1}{\gamma}x - b) + (-\pi_i)^*(b)
\end{aligned}$$

but for the projection π_i and $b \neq -e_i$ we get

$$(-\pi_i)^*(b) \stackrel{(7.1)}{=} -\pi_i^*(-b) = \infty.$$

So the projection enforces $b = -e_i$ and then vanishes. We continue

$$(\Gamma \circ (\Xi^{(i)} - \pi_i))^*(x) \quad = \quad \inf_{\gamma} \Gamma_S^{(i)*}(\gamma) + \gamma\, \Xi^{(i)*}(\frac{1}{\gamma}x + e_i)$$

Writing the restriction differently as

$$r^{(i)} \quad = \quad \frac{x}{\gamma} + e_i \qquad \Leftrightarrow \qquad \gamma(r^{(i)} - e_i) = x$$

we have proved the claim. □ 4.4.21

Claim 4.4.22. *If Ξ is associated with the subprobability measure p on $\mathbb{R}^d$ then $\Xi^*(r) < \infty$ for r representing a strictly positive sub-probability measure in $\mathbb{R}^d$ with*

$$p_0 = 0 \Leftrightarrow r_0 = 0 \quad \textit{and} \quad p_i = 0 \Rightarrow r_i = 0.$$

Proof of 4.4.22: We have for $\theta \in \mathbb{R}^d$

$$\nabla\Xi(\theta) \quad = \quad \left[\frac{p_i\, e^{\theta_i}}{\sum_{j=1}^{d} p_j\, e^{\theta_j} + p_0}\right]_{i=1,\dots,d}$$

and $\nabla\Xi(\theta)$ is a probability measure iff p is ($p_0 = 0$). For $p_0 > 0$ and $r_i = 0 \Rightarrow p_i = 0$ set

$$\theta_i = \log\big(\frac{r_i}{r_0}\frac{p_0}{p_i}\big) \qquad (i : p_i > 0).$$

In i such that $p_i = 0$ the θ_i never shows up in $\Xi(\theta)$ at all and can be set to an arbitrary value. We get $\nabla\Xi(\theta) = r$ and an optimiser in the transform is found. The coordinates θ_i with $i \in \mathrm{supp}(p)$ are uniquely defined.

If there are i such that $r_i = 0, p_i > 0$ then setting this $\theta_i = -\infty$ also makes $\nabla\Xi(\theta) = r$. Formally an optimiser of the transform does not exist (not in $\mathbb{R}^d$). We nevertheless get an $\Xi^*(r) < \infty$ and explain it more formally:

Set $z = \sum_{i:r_i>0} p_i + p_0$ which will be $z < 1$ if there are i such that $r_i =$

$0, p_i > 0$. If $z = 0$ then p, r are orthogonal / their supports have an empty intersection. So we assume $z > 0$ in the following:

$$\begin{aligned}
\sup_{\theta\in\mathbb{R}^d} \langle\theta, r\rangle - \Xi(\theta) &= \sup_{\theta_i : r_i > 0} \langle\theta, r\rangle + \sup_{\theta_i : r_i = 0} -\Xi(\theta) \\
&= \sup_{\theta_i : r_i > 0} \langle\theta, r\rangle - \lim_{\theta_i \to -\infty : r_i = 0} -\Xi(\theta) \\
&= \sup_{\theta_i : r_i > 0} \langle\theta, r\rangle - \log\big(\sum_{\substack{j=1 \\ r_j > 0}}^{d} p_j e^{\theta_j} + p_0\big) \\
&= \sup_{\theta_i : r_i > 0} \langle\theta, r\rangle - \Big(\log\frac{1}{z} + \log\big(\sum_{\substack{j=1 \\ r_j > 0}}^{d} p_j e^{\theta_j} + p_0\big) - \log\frac{1}{z}\Big) \\
&= \sup_{\theta_i : r_i > 0} \langle\theta, r\rangle - \log\big(\sum_{\substack{j=1 \\ r_j > 0}}^{d} \frac{p_j}{z} e^{\theta_j} + \frac{p_0}{z}\big) + \log\frac{1}{z}
\end{aligned}$$

which is now of the kind $\Xi(r)$ with Ξ associated with the submeasure $p' = (\frac{p_i}{z}\,; i \in \{1, \ldots, d\} \cap \mathrm{supp}(r))$. r and p' are now equivalent and our initial reasoning applies. $\square$ 4.4.22

So if $\Xi^*(r) < \infty$ for r, p that are not equivalent there is no optimising $\theta \in \mathbb{R}^d$. However, we can still associate with $\Xi^*(r)$ a change of measure for the sub-probability measure p that changes p into r. Note that Ξ is different from the inter event times lmgfs in that it comes from a random variable with point mass.

Corollary 4.4.23. *If Ξ is associated with the sub-probability measure p on $\mathbb{R}^d$ (cf definition 4.4.8) and $\theta \in \mathbb{R}^d$ then*

$$\nabla\Xi(\theta) = \left[\frac{p_i\, e^{\theta_i}}{\sum_{j=1}^{d} p_j\, e^{\theta_j} + p_0}\right]_{i=1,\ldots,d}$$

is a subprobability measure. $\nabla\Xi^{(i)}(\theta)$ is a probability measure if p is and

$$\big(\nabla\Xi(\theta)\big)_i > 0 \quad \Leftrightarrow \quad p_i > 0$$

Chapter 5

Stochastic networks and associated processes

This chapter introduces stochastic networks and the processes we work with in the next chapter.

We give a formal definition of stochastic networks starting from random walks on graphs. We will point out how the generalised Jackson network and the Jackson network are instances of stochastic networks.

As we are concerned with rare events for generalised Jackson networks where we can observe initial queue sizes we introduce tools to describe the expected behaviour of such a network with an initial condition. The initial condition will tell us which queues are initially full with a large queue size and which are about empty. In order to describe the expected future behaviour of the network with a given present starting point we will apply the Skorohod map.

Processes we work with are the free, the network, and the local process and we develop their change of measure. For the free process we will develop sample path large deviations.

We repeat our notation for counting processes.

inter event time	rcp, renewal counting process	lmgf of rcp	rate
τ	$N : t \mapsto N(t)$	$\Gamma(\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$	$\Gamma'(0) = \frac{1}{\mathbb{E}[\tau]}$
chapter 2	definition 3.1.1	section 3.3	definition 3.6.12

We have further defined the scaled process N_n in 3.4.1, the split process

N^{sp} in 4.4.1 with N_n^{sp} the scaled split process, and the restarted process $N^{\text{re},(s_1,\dots,s_k)}$ in 3.4.16 with the scaling specified in 3.4.12.

And we repeat our assumptions

Assumption 5.0.2. *If $\tau < \infty$ is a non-deterministic inter event time then the assumptions on inter event times of chapter 2 should hold: 2.2.2, 2.2.13, 2.4.2.*

In this chapter we will have different counting processes in the setting of a stochastic network: processes counting external arrivals at nodes of the network, we will call them arrival processes, and those counting how many customer can be served over any period of time, these will be denoted service processes.

5.1 Stochastic networks

Definition 5.1.1 (Graph)**.** *A graph is a collection of nodes and edges. For a graph of finitely many nodes we denote them as $\{1, \dots, d\}$ for some $d \in \mathbb{N}$. Edges will be directed and are written*

$$\{(i, j) \,|\, i, j \in \{1, \dots, d\}\}.$$

In this thesis we work with finite directed graphs only.

Definition 5.1.2 (Random walk on a graph)**.** *Given*

- *a graph with nodes $\{1, \dots, d\}$ for some $d \in \mathbb{N}$ and edges $Y \subseteq \{1, \dots, d\}^{\times 2}$*
- *a substochastic matrix $P \in \mathbb{R}^{d,d}$ with components $p_{ij} > 0 \Leftrightarrow (i, j) \in Y$ and rows $p^{(i)}$*
- *a fixed node $i_A \in \{1, \dots, d\}$*
- *a fixed $t_0 \geq 0$*
- *for each $i \in \{1, \dots, d\}$ a sequence of inter event times $\tau_1^{(i)S}, \tau_2^{(i)S}, \dots$ denoted service times*

a random walk z on the graph is a stochastic process giving at each time t the position of a customer who enters the graph at time t_0 at node i_A and travels the graph according to the following rules:

- *when at node i for the k-th time occupy the server for time $\tau_k^{(i)S}$*

- *at the end of the server occupation time / service time leave node i immediately and either go to node j with probability p_{ij} or leave the network with probability $1 - \sum_{j=1}^{d} p_{ij}$.*

If $t_0 = 0$ the process starts as $z(0) = i_A$, if $t_0 > 0$ the process starts as $z(t) = 0$ for $t \in [0, t_0)$ and $z(t_0) = i_A$. As the customer leaves the network, the state of the process is fixed at 0.

This random walk on a graph will also be referred to as isolated random walk on a graph.

Definition 5.1.3 (Stochastic network)**.** *A stochastic network is a joint random walk on a graph with inter action: If d is the number of nodes of the network and $A^{(i)}$ are counting processes modelling the arrivals to node i such that*

- *for every jumptime t' of $A^{(j)}$ and jump size $\Delta A^{(j)}(t') = k$ of any $j \in \{1, \ldots, d\}$ there are k random walks $z_{j,A^{(j)}(t'-)+1}, \ldots, z_{j,A^{(j)}(t')}$ with $i_A = j$ and $t_0 = t'$*

then define the stochastic network process Z as

$$Z^{(i)}(t) = \sum_{j=1}^{d} \sum_{k=1}^{\infty} 1\!\!1_{z_{j,k}(t)=i} \quad , \quad Z(t) = \begin{bmatrix} Z^{(1)}(t) \\ \vdots \\ Z^{(d)}(t) \end{bmatrix}.$$

A selection of typical interaction of randomly walking customers in the network is

- queueing at single server nodes: if at arrival at a node the server is occupied, an arriving customer has to wait until they can occupy the server;

- state dependent routing: the matrix $P_{j,n}$ of the random walk $z_{j,n}$ may be time-inhomogeneous and $P_{j,n}(t)$ may depend on $Z(t-)$;

- modified service: for each random walk the service times $\tau_1^{(i)}, \tau_2^{(i)}, \ldots$ at node i may depend on the value of Z at the time the customer occupies the server at node i.

If the isolated random walks forming the stochastic network are independent (no interaction) the stochastic network represents a network with a pool of infinitely many servers at each node. Generally queueing occurs whenever there are more customers at a node than servers, as a server can only serve one customer at a time. The number of customers exceeding the number of servers forms the queue at that node. Possible permutations of the sequence of customers arriving at a node and of customers starting service are specified in the queueing discipline (most well known are FIFO, LIFO which are explained in every book on queueing). Examples for state dependent routing are "join the shortest queue" and modifications of it, like: "joint the shortest queue of a randomly sampled subset". Modified service might be: if queue i is empty over time $[t_1, t_2)$ then the server at node i becomes an additional server at some node j with a non-empty queue ([8] of Robert Foley and David McDonald).

Definition 5.1.4 (Generalised Jackson network)**.** *A generalised Jackson network is a stochastic network with all random walks sharing*

- *the same graph,*
- *the same time-homogeneous, deterministic routing matrix P,*
- *the same distribution of service times at each fixed node; and service times are independent of how often a customer returns to the node.*

Further more

- *inter arrival times at each node are iid and independent of everything else,*
- *service times at each node are iid and independent of everything else,*
- *there is (only) queueing interaction with a single server at each node.*

In this thesis we assume that all service times and all inter arrival times satisfy assumption 5.0.2.

Note that under the general assumption 5.0.2 the counting process $A^{(i)}$ associated with inter arrival times at node i is a renewal counting process with all jumpsizes equal to unity, and regular (finitely many jumps over finite intervals). The stochastic network process for the generalised Jackson network Z satisfies

$$Z^{(i)}(t) \; = \sum_{j=1}^{d} \sum_{k=1}^{A^{(j)}(t)} 1\!\!1_{z_{j,k}(t)=i} \leq \sum_{j=1}^{d} A^{(j)}(t) \; < \; \infty$$

which makes $Z : \mathbb{R}_{\geq 0} \to \mathbb{N}^d$.

For the generalised Jackson network we have for each random walk on the graph $z_{j,n}$ a sequence of inter event times at each node. Due to independence of all service times of the isolated random walks we can replace service times at each node i by just a single sequence of iid service times. These service times will form the renewal counting processes $S^{(1)}, \ldots, S^{(d)}$.

We summarise the counting processes introduced so far:

inter event time	rcp	lmgf	rate
$\tau^{(i)A}$	$A^{(i)} \, : \, t \mapsto A^{(i)}(t)$	$\Gamma_A^{(i)}$	$\lambda_i = \Gamma_A^{(i)\prime}(0) = \frac{1}{\mathbb{E}[\tau^{(i)A}]}$
$\tau^{(i)S}$	$S^{(i)} \, : \, t \mapsto S^{(i)}(t)$	$\Gamma_S^{(i)}$	$\mu_i = \Gamma_S^{(i)\prime}(0) = \frac{1}{\mathbb{E}[\tau^{(i)S}]}$

The following definition names the processes required to describe the generalised Jackson network the networks primitives. We will later, in 5.3.4, give another equivalent definition for the network primitives that will be technically more convenient and will apply results from section 4.4.

Definition 5.1.5 (Network primitives I)**.** *For a generalised Jackson network with d nodes the following processes are denoted the networks primitives:*

- *arrival processes $A^{(1)}, \ldots, A^{(d)}$ with inter event times $\tau^{(i)A}$ and lmgf $\Gamma_A^{(i)}$ for $A^{(i)}$;*
- *service processes $S^{(1)}, \ldots, S^{(d)}$ with inter event times $\tau^{(i)S}$ and lmgf $\Gamma_S^{(i)}$ for $S^{(i)}$;*
- *the processes of routing decisions*

$$n \mapsto \sum_{k=1}^{n} r_k^{(i)}$$

with $r^{(i)}, r_1^{(i)}, r_2^{(i)}, \ldots$ iid with values in $\{e_1, \ldots, e_d, \vec{0}\}$ and $\mathbb{P}(r^{(i)} = e_j) = p_{ij}$ for $i = 1, \ldots, d$ and $\Xi^{(i)}$ the lmgf of $r^{(i)}$, cf 4.4.8.

Definition 5.1.6 (Rates λ, μ, P))**.** *For a generalised Jackson network with network primitives as in 5.1.5 and typical inter arrival times $\tau^{(i)A}$ for the primitive arrival process $A^{(i)}$ and typical service times $\tau^{(i)S}$ for the primitive service process $S^{(i)}$ (for $i = 1, \ldots, d$) let*

- $\lambda_i = \Gamma_A^{(i)\prime}(0) = \frac{1}{\mathbb{E}[\tau^{(i)A}]}$ *and $\lambda \in \mathbb{R}^d_{\geq 0}$ have coordinates λ_i for $i = 1, \ldots, d$;*
- $\mu_i = \Gamma_S^{(i)\prime}(0) = \frac{1}{\mathbb{E}[\tau^{(i)S}]}$ *and $\mu \in \mathbb{R}^d_{>0}$ have coordinates μ_i for $i = 1, \ldots, d$;*

- $p^{(i)} = \nabla\Xi^{(i)}(0) = \mathbb{E}[r^{(i)}]$ *and* $P \in \mathbb{R}^{d\times d}$ *has rows* $p^{(i)}$ *for* $i = 1, \ldots, d$.

Then we denote by (λ, μ, P) *the arrival rates, the service rates, and the routing matrix of the generalised Jackson network. In short we say that* (λ, μ, P) *are the rates of the network.*

In the following section we will investigate stochastic networks through their rates only. Results we obtain there are applicable to stochastic networks that are not generalised Jackson networks: They for example apply to networks where arrival processes are sums of independent renewal counting processes.

Definition 5.1.7 (Jackson Network)**.** *A Jackson Network is a generalised Jackson network with all inter arrival and service times being exponentially distributed.*

James Jackson's definition in [14] of what he then called a network of waiting lines allowed for finitely many servers at each node and did specify the queueing discipline as “first come first serve”. In the context of large deviation for the queue sizes multiple servers are modeled as a single server and the service times at that node are decreased correspondingly. Since we do not distinguish different classes of customers and investigate queue sizes only and not delay the queueing discipline is not relevant here. Our definition 5.1.7 of Jackson networks corresponds to the definition of Ignatiouk-Robert in [12].

The following definition is reasonable in a network where customer share the same underlying graph as they do in the generalised Jackson network.

Definition 5.1.8 (Path)**.** *In a network with routing matrix* P *we say there is a path from node* i *to node* $j \neq i$ $(i, j \in \{1, \ldots, d\})$ *if there is a sequence* $k_1, \ldots, k_m$ *of nodes in* $\{1, \ldots, d\}$ *such that*

$$p_{ik_1} p_{k_1k_2} \ldots p_{k_mj} > 0$$

or equivalently $P^{m+1}(i, j) > 0$ *with* P^{m+1} *the* $m + 1$*-times product of* P*. We say that such a path has length* $m+1$ *as it moves along* $m+1$ *(not necessarily different) edges.*

We make the following assumption for the rest of this thesis:

Assumption 5.1.9 (Open, no immediate feedback)**.** *Networks are open and without immediate feedback:*

- *For each node* j *there is a node* i *with* $\lambda_i > 0$ *and a path from* i *to* j*. If* $\lambda_i > 0$ *node* i *is called an entry node.*

- *For each node i there is a node j with $p_{j0} > 0$ and a path from i to j. If $p_{i0} > 0$ node i is called an exit node.*

- *When leaving a node customers are not immediately fed back into the same node.*

The assumed properties are reflected in the routing matrix P: P is strictly substochastic and $-1, 1$ are no eigenvalues. Then $(\mathrm{id} - P^\top)^{-1}$ exists, and $(\mathrm{id} - P^\top)^{-1}\lambda > 0$ (coordinate-wise). Also $(\mathrm{id}_M - P_M^\top)^{-1}$ exists for any $M \subseteq \{1, \ldots, d\}$. No immediate feedback in the network is equivalent to $p_{ii} = 0$ for $i = 1, \ldots, d$.

Satisfying assumption 5.1.9 is a property of a network that depends on the network topology or the adjacency matrix associated with P and the entry nodes $\{i \mid \lambda^{(i)} > 0\}$. Different networks with equivalent distributions for inter arrival times $\tau^{(i)A}$ and inter service times $\tau^{(i)S}$ and equivalent distributions for all routing decisions $r^{(i)}$ are either all open and without feedback or none of them is.

Example 5.1.10. *The network of figure 5.1 satisfies 5.1.9. Nodes* $1, 2$ *are entry-nodes, nodes* $3, 4$ *are exit nodes.*

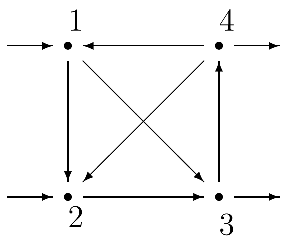


Figure 5.1: An open network without immediate feedback for $d = 4$

The graph is even strongly connected which is not a general assumption.

Removing immediate feedback

The assumption 5.1.9 of no feedback is no restriction at all. For an open network with feedback we can remove the feedback and remodel the network such that the theory developed in this thesis is applicable.

In a feedback queue customers finishing service have two possibilities to proceed: either leave the network or join the queue again. If decisions to join the queue again are iid with p the probability to do so then a typical total service time is the τ° defined in 2.1.4. We repeat the definition:

$$\tau^\circ = \sum_{k=1}^{G} \tau_k \quad \text{with geometrically distributed } G \geq 1 \text{ and iid } \tau, \tau_1, \tau_2, \ldots$$

The $\circ$ above the τ is a reference to the loop and the parameter p is dropped in the notation.

We can now remodel the queue as one without feedback but with inter event times $\tau_1^\circ, \tau_2^\circ, \ldots$.

Claim 5.1.11. *Queue sizes of the single queue with inter event times $\tau, \tau_1, \tau_2, \ldots$ and feedback and the GI/GI/1 with service times $\tau^\circ, \tau_1^\circ, \tau_2^\circ, \ldots$ and without feedback have the same distribution.*

Proof of 5.1.11 by coupling: Let $\tau_1, \tau_2, \ldots$ be the iid sequence of service times and $r, r_1, r_2, \ldots$ the sequence of routing decisions $r \in \{0, 1\}$ and $r = 1$ representing a customer just having finished service rejoins the feedback-queue.
We model an associated queue without feedback the following way: With the sequence of routing decisions associate a sequence $G_1, G_2, \ldots$ that counts the length of runs of 1-s:

- $G_1 = 1 + \max\{j \geq 1 \,|\, 1 = r_1 = \cdots = r_j, \}$
- $K(i) := \sum_{k=1}^{i-1} G_k$, $G_i = 1 + \max\{j \geq 1 \,|\, 1 = r_{K(i)+1} = \cdots = r_{K(i)+j}\}$.

(with $\max \emptyset = 0$) Then the $G_1, G_2, \ldots$ are iid geometric and $\tau_i^\circ = \sum_{g=1}^{G_i} \tau_{K(i)+g}$ defines service times of the associated queue without feedback.

Starting from one set of sequences $\tau_1, \tau_2, \ldots$ and $r_1, r_2, \ldots$ we have customers leaving the respective queues with or without feedback at exactly the same epochs: at $\sum_{j=1}^{k} \tau_k$ with a k such that $r_k = 0$. With the same arrival process both queues are always of the same size. $\square$
5.1.11

The same remodelling of inter event times can be applied to a queue in a network. If at node i customers are routed internally wrt $\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_d \end{bmatrix}$ with $\alpha_i > 0$ we can change inter event times at node i from $\tau^{(i)}$ to $\tau^{(i)\circ}$ (with $p = \alpha_i$) and routing to $p^{(i)} = \begin{bmatrix} p_{i1} \\ \vdots \\ p_{id} \end{bmatrix}$ with $p_{ii} = 0, p_{ij} = \frac{\alpha_j}{1-\alpha_i}$.

5.2 Deterministic descriptions of stochastic networks

In this section we focus on the deterministic rates of a stochastic network. This section generally applies to stochastic networks with deterministic rates, including the generalised Jackson network. In terms of only the rates the Jackson and the generalised Jackson network are not distinguishable.

Definition 5.2.1 (Flow). *If (λ, μ, P) are the deterministic rates of an open stochastic network then*

$$\nu = \lambda + P^\top \min\{\nu, \mu\}$$

is called the traffic equation (in ν) and its unique solution is denoted the flow in the network.

Uniqueness of ν is proved in [11]. Given the network flow one can decide if or if not the network is ergodic.

Definition 5.2.2 (Ergodic network). *A network with flow ν with $\nu_i < \mu_i$ for all i is called ergodic.*

From ergodicity we get the existence of an equilibrium distribution for the number of customers in each queue of the Jackson network. We have seen this in the introduction. This implies that in the limit of the scaling 3.4.1 all queue sizes will uniformly on compacts stay small: The a.s. limit of the network process is the function $t \mapsto 0$. That is we know the deterministic limiting behaviour, the expected behaviour. We expect the same for the generalised Jackson network.

An easy to check criterion for ergodicity is: Calculate the flow rates as if all service rates were $= \infty$.

$$\nu = \lambda + P^\top \min\{\nu, \infty\} = \lambda + P^\top \nu \quad \Leftrightarrow \quad \nu = (\mathrm{id} - P^\top)^{-1} \lambda$$

and check if $\nu_i < \mu_i$ for all nodes i of the network. If not, the network is not ergodic. This can be written as an (coordinate wise) "equilibrium inequality"

$$(\mathrm{id} - P^\top)^{-1} \lambda < \mu \tag{5.1}$$

If a network is not ergodic solving the traffic equation is still important. For a nice way to solve the traffic equation see [11] of Jonathan Goodman and William Massey where they give an algorithm to calculate ν with at most d matrix vector multiplications in $\mathbb{R}^d$.

Definition 5.2.3 (Traffic intensity)**.** *In a network with flow ν and service rates μ the traffic intensity ρ_i at the i-th node is defined as $\rho_i = \frac{\nu_i}{\mu_i}$.*

Definition 5.2.4. *In a network with traffic intensity ρ we say node i is*

- *a bottleneck if $\rho_i \geq 1$*
- *a strict bottleneck if $\rho_i > 1$*
- *ergodic if $\rho_i < 1$.*

Definition 5.2.5 (Loss rate)**.** *In a stochastic network with deterministic rates (λ, μ, P) and flow ν define the loss rate $y \in \mathbb{R}^d_{\geq 0}$*

$$y := \max\{0, \mu - \nu\}. \tag{5.2}$$

For ergodic nodes we have $y_i = \mu_i - \nu_i$ and in an ergodic network

$$y = \mu - (\mathrm{id} - P^\top)^{-1} \lambda \qquad (\in \mathbb{R}^d_{>0}) \tag{5.3}$$

Definition 5.2.6 (Free drift, equilibrium network drift)**.** *In a stochastic network with deterministic rates (λ, μ, P) define the free drift of the network as*

$$\lambda + P^\top \mu - \mu.$$

Additionally, for ν the flow of the network (cf 5.2.1) define the network drift as the coordinate wise maximum

$$\max\{0, \lambda + P^\top \min\{\nu, \mu\} - \mu\}$$

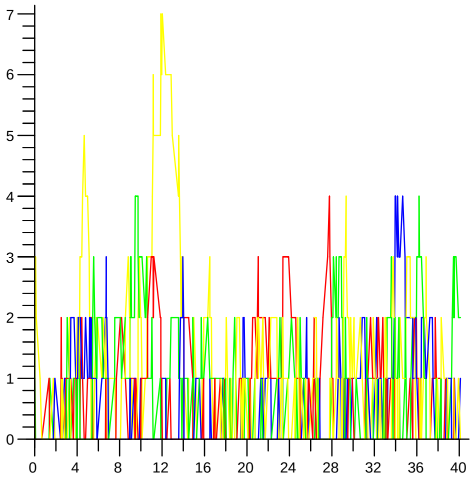
Remark 5.2.7.

- *The equilibrium drift of an ergodic network is 0.*
- *The network drift can equivalently be expressed as $\max\{\nu - \mu\}$.*
- *$\nu_i - \mu_i = \mu_i (\rho_i - 1)$ for each $i = 1, \ldots, d$ from expressing ν_i in terms of ρ_i.*

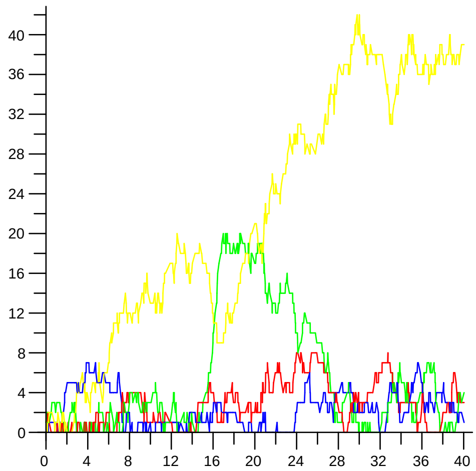
The following is an example on rates and drifts in a network with fixed routing probabilities and service resources at the nodes. The network drift is qualitatively different wrt arrival processes with the different rates λ.

Example 5.2.8. *Consider the network 5.1 with λ, μ, P where*

$$\mu = \begin{bmatrix} 3 \\ 4 \\ 4 \\ 3 \end{bmatrix} \quad , \quad P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}.$$



Ergodic network



Non-ergodic network,
bottleneck node 3 (yellow)

Figure 5.2: Realisation of the queue sizes in the networks of example 5.2.8.

Then the free and network drift and the set of bottlenecks depend on the arrival rates

arrival rates	*free drift*	*network drift*	*bottlenecks*
$\lambda = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\frac{1}{4} \begin{bmatrix} -5 \\ -3 \\ 6 \\ -4 \end{bmatrix}$	0 *since* $(id - P^{\top})^{-1}\lambda = \frac{1}{3} \begin{bmatrix} 4 \\ 6 \\ 8 \\ 4 \end{bmatrix} < \mu$	*none*
$\lambda = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}$	$\frac{1}{4} \begin{bmatrix} -1 \\ 1 \\ 6 \\ -4 \end{bmatrix}$	$\max\{ \begin{bmatrix} 2.5 \\ 3.75 \\ 5 \\ 2 \end{bmatrix} - \mu \, , \, 0\} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$	*3 (strict)*

Figure 5.2 gives two simulations of the network of example 5.2.8 and for the different arrival rates. Note the different scales for the y-axis: In the ergodic network all queues stay small, in the non-ergodic network the queue size at the bottleneck grows.

In the following two subsections we interprete the solution of the traffic equation as actual flow and we will look at non-ergodic networks and investigate ergodic subnetworks.

5.2.1 Fluid network

We generally look at a network that is travelled by customers or packages, things that can be counted. We will now look at an associated sewer like network model.

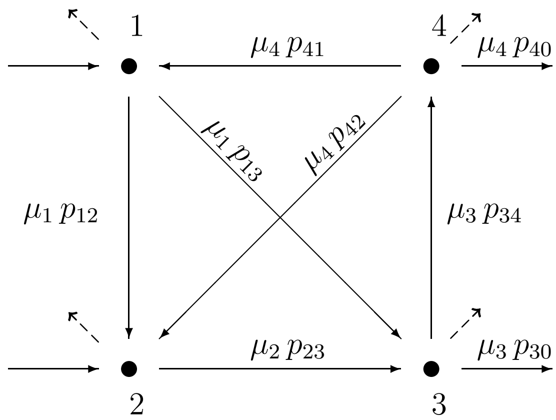


Figure 5.3: The fluid network of definition 5.2.9 associated with the network of figure 5.1. $\dashrightarrow$ represents an outlet.

Definition 5.2.9 (Fluid model associated with a stochastic network)**.** *Given a stochastic network with deterministic rates* (λ, μ, P) *the associated deterministic fluid model is based on the same graph where edges now represent pipes:*

- *At node* i *with* $\lambda_i > 0$ *some kind of fluid flows into the network at rate* λ_i*.*
- *The edge / pipe* (i, j) *has capacity* $\mu_i\, p_{ij}$*. The joint capacity of edges leaving from node* i *is* $\mu_i = \sum_{j=0}^{d} \mu_i\, p_{ij}$*.*
- *At each node there is an outlet.*
- *Propagation: If at node* i *the sum of incoming flow is strictly less than* μ_i *then the incoming flow is divided into outgoing flow to nodes* $1, \ldots, d$ *and the outside world according to* $p_{i1}, \ldots, p_{id}$ *and* p_{i0}*. Otherwise all outgoing pipes / edges will get flow equal to their capacity and the non-negative surplus leaves the network through the outlet.*

In this setting the solution ν to the traffic equation is the actually observed equilibrium flow in the network: ν_i is the total flow into node i and

$\min\{\mu_i, \nu_i\}\, p_{ij}$ is the flow through the edge / pipe connecting nodes i and j. In a non-ergodic network there is a node where at rate $\nu_i - \mu_i \geq 0$ fluid leaves the network through the outlet. All nodes that can route all incoming flow to nodes $j \in \text{supp}(p^{(i)})$ and still have some spare capacity left are ergodic nodes.

5.2.2 Subnetworks

Distinguishing bottlenecks and ergodic nodes in a network allows to find an ergodic subnetwork. We can move the bottleneck nodes to the outside world and calculate flow rates in the network of remaining nodes. In the flow network moving node i "to the outside world" means removing node i from the network and increasing for all nodes j in the support of $p^{(i)}$ the arrival rate from λ_j to $\lambda_j + \mu_i\, p_{ij}$. From the intuition of the flow network it is clear that flow rates in the ergodic subnetwork remain the same. However, we do the explicit calculations.

The network starting empty

Let ν be the flow of the network with rates (λ, μ, P) as defined in 5.2.1. Partition nodes into two sets: the ergodic nodes E and the bottleneck nodes B as defined in 5.2.4.

Claim 5.2.10. *In a network with rates (λ, μ, P) and flow ν partition nodes into ergodic nodes E and bottleneck nodes B and consider the subnetwork of nodes E with rates*

$$\lambda_E + (P^\top)_{E\,B}\, \mu_B\,,\ \mu_E\,,\ P_E \tag{5.4}$$

Then $\nu_E = [\nu_i]_{i \in E}$ is the flow in this subnetwork.

Proof of 5.2.10: For the network of all d nodes the following initial statement is true.

$$\begin{aligned} & \nu = \lambda + P^\top \min\{\nu, \mu\} \\ \overset{\text{any partition}}{\Leftrightarrow} & \left\{ \begin{array}{lcl} \nu_B & = & \lambda_B + P_B^\top \min\{\nu_B, \mu_B\} + (P^\top)_{B\,E} \min\{\nu_E, \mu_E\} \\ \nu_E & = & \lambda_E + (P^\top)_{E\,B} \min\{\nu_B, \mu_B\} + P_E^\top \min\{\nu_E, \mu_E\} \end{array} \right\} \\ \overset{\text{our partition}}{\Leftrightarrow} & \left\{ \begin{array}{lcl} \nu_B & = & \lambda_B + P_B^\top\, \mu_B + (P^\top)_{B\,E}\, \nu_E \\ \nu_E & = & \lambda_E + (P^\top)_{E\,B}\, \mu_B + P_E^\top\, \nu_E \end{array} \right\} \end{aligned} \tag{5.5}$$

The second equation of (5.5) can be rearranged:

$$\nu_E = \lambda_E + \ (P^\top)_{E\,B}\, \mu_B + P_E^\top\, \nu_E \ = \quad \underbrace{\lambda_E + (P^\top)_{E\,B}\, \mu_B}_{\text{arrivals to the ergodic subnetwork}} \quad + P_E^\top\, \underbrace{\min\{\nu_E\,,\, \mu_E\}}_{=\nu_E}$$

So the ν_E we extracted coordinate wise from the flow of the network of all d, bottleneck and ergodic, nodes is the solution for the traffic equation in the remodelled smaller network of ergodic nodes with rates (5.4). $\square$ 5.2.10

Let us now turn again to the bottleneck nodes

Claim 5.2.11. *In a network with rates (λ, μ, P) and flow ν partition nodes into ergodic nodes E and bottleneck nodes B. Then the bottleneck nodes B have equilibrium network drift*

$$\big(\,\lambda_B + (P^\top)_{BE}\,\nu_E\,\big) + (P_B^\top - id)\,\mu_B$$

with $\nu_E = [\nu_i]_{i\in E}$ the flow in the ergodic subnetwork. The equilibrium network drift equals the free drift.

Proof of 5.2.11: we prove that

$$\nu_B - \mu_B = \lambda_B + (P^\top)_{BE}\,\nu_E + (P_B^\top - \mathrm{id})\,\mu_B$$

where the lhs is the network equilibrium drift by definition 5.2.6. We apply the same partitioning of $\{1,\dots,d\} = E \cup B$ we solve the second equation of (5.5) for ν_E

$$\nu_E = \lambda_E + (P^\top)_{E\,B}\,\mu_B + P_E^\top\,\nu_E \quad\Leftrightarrow\quad \nu_E = (\mathrm{id} - P_E^\top)^{-1}\big(\lambda_E + (P^\top)_{E\,B}\,\mu_B\big)$$

and plug it into the first equation of (5.5). Substraction of μ_B gives the claimed equilibrium drift. Which is a free drift as in definition 5.2.6. $\square$ 5.2.11

We can also get an expression for ν_E , ν_B (And then for the drift of bottleneck nodes) with on the rhs only network primitives:

$$\begin{aligned}\nu_E &= (\mathrm{id} - P_E^\top)^{-1}\,(\lambda_E + (P^\top)_{EB}\,\mu_B &\quad (5.6)\\ \nu_B &= \lambda_B + P_B^\top\,\mu_B + (P^\top)_{B\,E}\,(\mathrm{id} - P_E^\top)^{-1}\big(\lambda_E + (P^\top)_{E\,B}\,\mu_B\big) &\quad (5.7)\end{aligned}$$

We remember that the partition of nodes into sets E and B was such that $\nu_E < \mu_E$ and $\nu_B \geq \mu_B$ coordinate wise.

Another approach to finding the ergodic subnetwork is in Chen and Mandelbaum, [3] p. 411. LCP is short for "linear complementary problem".

Definition 5.2.12 (LCP in $\mathbb{R}^d$)**.** *Let x , P be given with $x \in \mathbb{R}^d$ and P a substochastic matrix in $\mathbb{R}^{d\times d}$ and 1 not an eigenvalue of P. The linear complementary problem is to find (y, z) satisfying*

$$z = x + (id - P^\top)y \quad , \quad y, z \geq 0 \quad , \quad \langle y, z\rangle = 0 \tag{5.8}$$

A solution to the LCP exists and is unique for any x if all the principal minors of $\mathrm{id} - P$ are positive (cf [1] p. 271). This is what we get from 1 not being an eigenvalue of P. The following claim is from [3] p. 412.

Claim 5.2.13. *In a stochastic network with deterministic rates (λ, μ, P) and free drift $x = \lambda + P^\top \mu - \mu$ let (y, z) be the solution to the LCP in $\mathbb{R}^d$ for (x, P). Then z is the equilibrium network drift and y is the loss rate.*

Proof of 5.2.13:

- $y = \max\{0, \mu - \nu\} \geq 0$, $z = \max\{0, \nu - \mu\} \geq 0$
- $\max\{0, \mu - \nu\} \cdot \max\{0, \nu - \mu\} = 0 \geq 0$
- to show $z = x + (\mathrm{id} - P^\top)y$ we start with partitioning nodes $\{1, \ldots, d\}$ into ergodic nodes E and bottleneck nodes B. The network drift is then $\begin{bmatrix} 0 \\ \nu_B - \mu_B \end{bmatrix}$. We show that the drift equals $x + (\mathrm{id} - P^\top)y$.

$$\begin{aligned} \big(x + (\mathrm{id} - P^\top)y\big)_E &= \lambda_E + \big((P^T - \mathrm{id})\,\mu\big)_E + \big((\mathrm{id} - P^\top)\,y\big)_E \\ &= \lambda_E + \left[\ P_E^\top - \mathrm{id} \quad (P^\top)_{EB}\ \right] (\mu - y) \\ &= \lambda_E + (P_E^\top - \mathrm{id})\,(\underbrace{\mu_E - y_E}_{=\nu_E}) + (P^\top)_{EB}\,(\mu_B - \underbrace{y_B}_{=0}) \\ &\overset{(5.6)}{=} 0 \end{aligned}$$

For the bottleneck nodes

$$\begin{aligned} &\big(x + (\mathrm{id} - P^\top)y\big)_B \\ &\quad = \quad \lambda_B + (P_B^\top - \mathrm{id})\,(\mu_B - \underbrace{y_B}_{=0}) + (P^\top)_{BE}\,(\underbrace{\mu_E - y_E}_{=\nu_E}) \\ &\quad \overset{5.2.11}{=} \quad \nu_B - \mu_B \end{aligned}$$

So we feed the LCP the free drift and get the loss rate and the equilibrium network drift. We can identify bottlenecks and ergodic nodes from the solution of the LCP by

- i is a bottleneck if $z_i \geq 0$
- a strict bottleneck if if $z_i > 0$
- an ergodic node if $z_i = 0$ and $y_i > 0$.

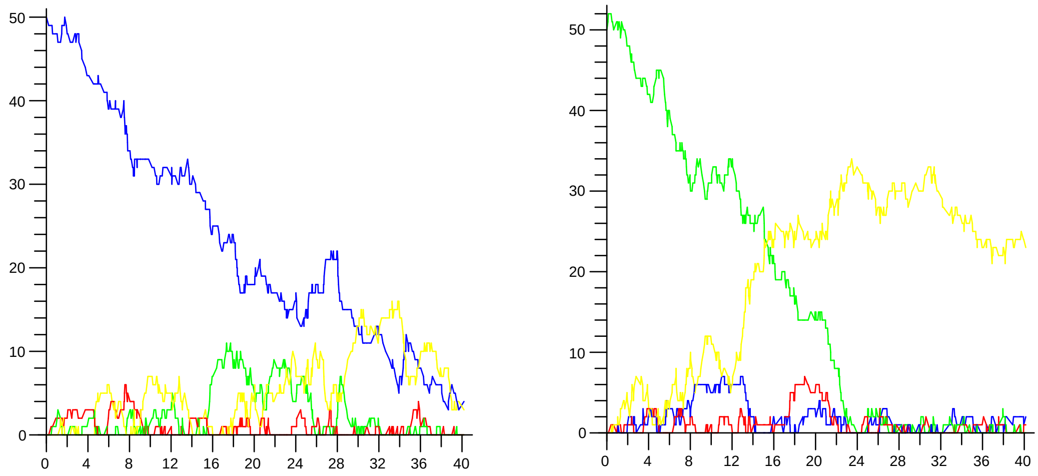


Figure 5.4: The ergodic network of 5.2.8 with a queue starting non-empty. This may or may not change ergodic nodes to become bottlenecks.

The network starting non-empty

Similar to observing the subnetwork of ergodic nodes we can investigate a network where some nodes are initially non-empty. We would investigate the subnetwork of initially empty nodes and the initially non-empty nodes separately.

The reason why bottleneck nodes and full nodes are treated similarly is: bottlenecks tend to fill up if started empty (positive network drift), so the initial difference vanishes immediately: We see this in figure 5.5 which shows simulations of the non-ergodic network of example 5.2.8: Qualitatively the evolution of queue sizes is the same if or if not the bottleneck node (green) starts empty or not.

The following is a generalisation of the LCP to function space. It will lead to a definition of a unique network drift and loss rate in a network where initial nodes may be non-empty. The Skorohod problem is generally investigated in the space $D([0,\infty),\mathbb{R}^d)$ functions (cf [18], appendix D) but we only need it for (piecewise) linear input functions. In [3], section 5.2 p. 431 the Skorohod problem is stated under the name “oblique reflection mapping”. We give a simplified version:

Theorem 5.2.14 (Skorohod problem for linear input functions)**.** *If*

- $P \in \mathbb{R}^{d\times d}$ *is substochastic with* $\rho(P) < 1$

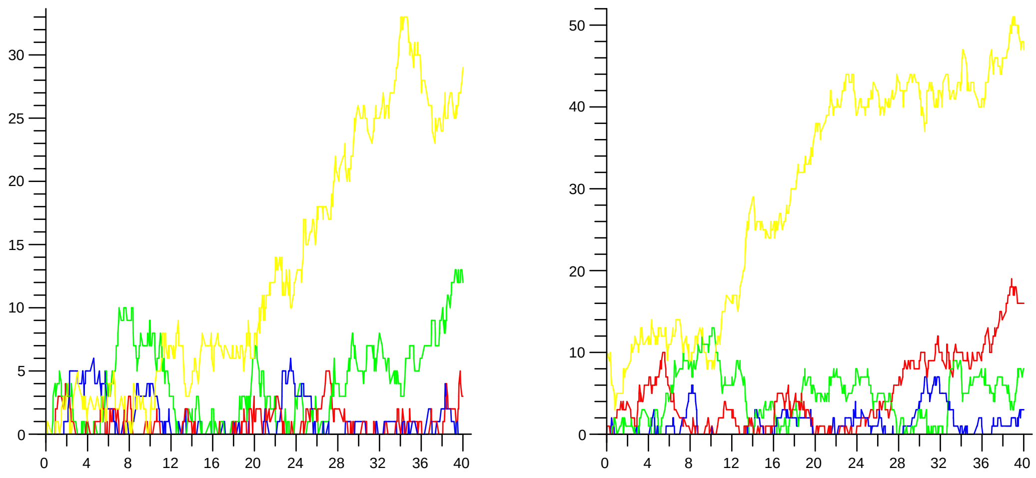


Figure 5.5: The non-ergodic network of 5.2.8, the bottleneck node starting empty or not does not affect the qualitative behaviour of the remaining ergodic nodes

- $z_0 \in \mathbb{R}^d_{\geq 0}$, $\theta \in \mathbb{R}^d$

and $X : X(t) = z_0 + t\,\theta$ *then there are unique functions* Y, Z *continuously depending on* X *such that*

- $Z = X + (id - P^\top)\,Y$
- $Y, Z \geq 0$
- $Y(0) = 0$; $Z(0) = z_0$
- Y *coordinate wise increasing and for each* $i \in \{1, \dots, d\}$*:* Y_i *increases at times* $t \geq 0$ *when* $Z_i(t) = 0$*.*

We consider the Skorohod problem with the linear input function with the free network drift as slope.

Claim 5.2.15. *Consider the Skorohod problem 5.2.14 for* P, $X(t) = z_0 + t\,(\lambda + (P^\top - id)\,\mu)$*. Then* Y, Z *of the solution of the Skorohod problem are linear over* $[0, T]$ *for some* $T > 0$*.*

Proof of 5.2.15: Let $\Lambda = \{i \,|\, z_{0,i} > 0\}$ and ν^Λ the flow in the Λ^c-subnetwork with rates

$$(\lambda^\Lambda\,,\, \mu_{\Lambda^c}\,,\, P_{\Lambda^c}) \quad , \quad \lambda^\Lambda = \lambda_{\Lambda^c} + (P^\top)_{\Lambda^c\Lambda}\mu_\Lambda$$

Let $K = \{i \,|\, z_{0,i} > 0 \text{ or for } i \in \Lambda^c : \nu_i^\Lambda > \mu_i\}$ and consider the subnetwork of K^c-nodes with rates

$$(\lambda^K \,,\, \mu_{K^c} \,,\, P_{K^c}) \quad , \quad \lambda^K = \lambda_{K^c} + (P^\top)_{K^c K}\mu_K$$

Note that $\nu_i^\Lambda = \nu_i^K$ for $i \in K^c$ by 5.2.10 and that K^c nodes are ergodic or non-strict bottlenecks: the flow ν^K satisfies $\nu_i^K \leq \mu_i$ for all $i \in K^c$.

Rearrange nodes such that $\Lambda = \{1, \ldots, |\Lambda|\}$ and $\Lambda\,,\, K = \{1, \ldots, |K|\}$ (this works since $K \supseteq \Lambda$ in the above construction).

Now set

- z_{K^c} to the network drift of the K^c subnetwork and $y' \in \mathbb{R}^{|K^c|}$ to the loss rates:

  $$z_{K^c} \;=\; \lambda_{K^c} + (P^\top)_{K^c K}\,\mu_K + (P_{K^c}^\top - \mathrm{id})\,(\mu_{K^c} - y')$$

  which is $= 0$ since K^c nodes are no strict bottlenecks. Equivalently let (z_{K^c}, y') be the solution of the LCP in $\mathbb{R}^{|K^c|}$ with (x, P) of definition 5.2.12 as

  $$x = \lambda_{K^c} + (P^\top)_{K^c K}\,\mu_K + (P_{K^c}^\top - \mathrm{id})\,\mu_{K^c} \quad , \quad P = P_{K^c}$$

- z_K to the drift of K-nodes: $z_K = \lambda_K + (P^\top)_{KK^c}\,\nu^K + (P_K^\top - \mathrm{id})\,\mu_K$. Note that z_i may be negative for $i \in \Lambda \subseteq K$.

Then let $Z(t) = z_0 + tz$ and $Y(t) = ty$ with $y = \begin{bmatrix} 0 \\ y' \end{bmatrix}$. We rearrange

$$\begin{aligned} z \;&=\; \begin{bmatrix} z_K \\ z_{K^c} \end{bmatrix} = \lambda + \begin{bmatrix} (P_K^\top - \mathrm{id})\,\mu_K & + & (P^\top)_{KK^c}\,(\mu_{K^c} - y^K) \\ (P^\top)_{K^c K}\,\mu_K & + & (P_{K^c}^\top - \mathrm{id})\,(\mu_{K^c} - y^K) \end{bmatrix} \\ &=\; \lambda + (P^\top - \mathrm{id})\,(\mu - y) \\ &=\; \lambda + (P^\top - \mathrm{id})\,\mu + (\mathrm{id} - P^\top)\,y \end{aligned}$$

We have

- $i \in K^c \;\Rightarrow\; z_{0,i} = 0, z_i = 0$
- $i \in K \setminus \Lambda \;\Rightarrow\; z_{0,i} = 0, z_i > 0$
- $i \in \Lambda \;\Rightarrow\; z_{0,i} > 0, z_i \in \mathbb{R}$

which implies that there is $T > 0$ such that $z_0 + Tz \geq 0$ coordinatewise. We now give the explicit solution of the Skorohod problem on $[0, T]$:

$$\begin{aligned}
Y(t) &= t\,y \\
Z(t) &= z_0 + tz \\
&= z_0 + t\big(\lambda + (P^\top - \mathrm{id})\,\mu + (\mathrm{id} - P^\top)\,y\big) \\
&= z_0 + t\big(\lambda + (P^\top - \mathrm{id})\,\mu\big) + (\mathrm{id} - P^\top)\,t\,y \\
&= X(t) + (\mathrm{id} - P^\top)\,Y(t)
\end{aligned}$$

$\square$
5.2.15

Claim 5.2.15 can be generalised to arbitrary T and piecewise linear Y, Z.

Definition 5.2.16 (Network drift and loss rate, non-empty network)**.** *Consider a stochastic network with deterministic rates (λ, μ, P). Let $z_0 \in \mathbb{R}^d_{\geq 0}$ and consider the Skorohod problem with*

- *P,*
- *z_0,*
- *$X(t) = z_0 + t\big(\,\lambda + (P^\top - id)\,\big)\mu$.*

If linear processes $Z(t) = z_0 + tz, Y(t) = ty$ are the solution of this Skorohod problem then define

- *z as the network drift*
- *y as the loss rate.*

For well-defined-ness note that from the Skorohod problem $z_{0,i} = 0 \Rightarrow y_i \geq 0$ and $z_{0,i} > 0 \Rightarrow y_i = 0$. The network drift and loss rate are unique from uniqueness of the solution to the Skorohod problem.

Remark 5.2.17. *If a network with rates (λ, μ, P) starts in z_0 and $\Lambda = \{i \,|\, z_{0,i} > 0\}$ is the set of indices of queues starting nonempty and z is the network drift then*

- *$z_\Lambda = [z_i]_{i\in\Lambda}$ is the free drift of the initially non-empty nodes $i \in \Lambda$;*
- *$z_{\Lambda^c} = [z_i]_{i\in\Lambda^c}$ is the equilibrium drift of the subnetwork of Λ^c-nodes, the initially empty nodes.*

5.3 Processes

We now move from vectors of real numbers to sample paths. This section is about different kind of processes we need to describe the behaviour of the network, especially with reference to it starting non-empty and being non-ergodic.

We define stochastic processes to analyse the generalised Jackson network in terms of large deviations. We will emphasise the changes of measure that maintain desirable properties of the processes.

5.3.1 The free process

We describe the free process associated with a generalised Jackson network and investigate its large deviation behaviour. The free process is interesting since it behaves like a network in a certain way and will be easy to analyse in terms of large deviations. We gain valuable insights we will apply to obtain the local large deviations of the generalised Jackson network.

We apply results from chapter 4 on the sample paths large deviations of the networks primitives to obtain relatively easily a sample path large deviation principle for the free process. We present it in 5.3.13. Similarly the change of measure for the free process can be given easily relying on chapter 4 and independence of the network primitives. We define and interprete a change of measure in 5.3.14.

Given a generalised Jackson network of d nodes with primitive processes of 5.1.5 we make some definitions:

Definition 5.3.1 (Vectorial arrival process)**.** *In a stochastic network with d nodes, network primitives 5.1.5 where we denote the arrival processes $A^{(1)}, \dots, A^{(d)}$ set*

$$A = \begin{pmatrix} A^{(1)} \\ \vdots \\ A^{(d)} \end{pmatrix}.$$

Definition 5.3.2. *In a stochastic network with d nodes, network primitives 5.1.5 we define for the i-th service process $S^{(i)}$ and the i-th process of routing decisions $n \mapsto \sum_{k=1}^{n} r_k^{(i)}$*

$$\mathbf{S}^{(i)} \, : \, t \mapsto \sum_{k=1}^{S^{(i)}(t)} r_k^{(i)} - e_i S^{(i)}.$$

Remark 5.3.3. *For the primitive service process $S^{(i)}$ and the process $S^{(i)sp}$ split wrt $S^{(i)}$ and $(p_{i1}, \dots, p_{id}, 1 - \sum_{j=1}^{d} p_{ij})$ as in 4.4.1. For $\mathbb{T}^{(i)}$ of definition 4.4.7 we have*

$$\mathbb{T}^{(i)} S^{(i)sp} = \mathbb{T}^{(i)} \begin{bmatrix} S^{(i1)} \\ \vdots \\ S^{(id)} \\ S^{(i,d+1)} \end{bmatrix} = -e_i S^{(i)} + \sum_{k=1}^{d} e_k \, S^{(ik)} = \mathbf{S}^{(i)}$$

and for this process we have developed the sample path large deviations in section 4.4.3.

We will refer to $\mathbf{S}^{(i)}$ as a renewal counting process, too. Its coordinates count customers leaving node i and customers moving from node i to the other network nodes and times between changes of state are independent. We now give a definition for the network primitives that is equivalent to that in 5.1.5

Definition 5.3.4 (Network primitives II)**.** *If for $i = 1, \dots, d$*

$$A^{(i)}, S^{(i)}, n \mapsto \sum_{k=1}^{n} r_k^{(i)}$$

are the network primitives of a generalised Jackson network as defined in 5.1.5 and A is defined as in 5.3.1 and $\mathbf{S}^{(i)}$ for $i = 1, \dots, d$ as in 5.3.2 then

$$A, \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(d)}$$

represent the same generalised Jackson network. These processes will be denoted the networks primitives.

In [3], section 2.7 these processes together with the initial starting point $Z(0)$ of the network are denoted the networks primitives. In a generalised Jackson network the primitives are independent. With these we define the

Definition 5.3.5 (Free process)**.** $X = A + \sum_{i=1}^{d} \mathbf{S}^{(i)}$.

The process X increases in its i-th coordinate at the occurrence of an event of the arrival process $A^{(i)}$. When $\mathbf{S}^{(i)}$ changes state we can interprete this as a customer leaving node i and being routed to some other network node or leaving the network. This change of state also happens in the free process X. In these aspects X describes the stochastic network well. However, the free process does not suitably describe a stochastic network because coordinates of X may be of negative values.

The free process is nice to work with: We will develop its large deviations in the following sections. Before that we want to connect the free process with the free drift defined in section 5.2.

Claim 5.3.6. *Network primitives in a generalised Jackson network converge a.s uniformly on compacts with a linear limit.*

Proof of 5.3.6: This follows as a functional strong law of large numbers for the renewal counting processes, cf 7.1.3 of the appendix, or from the sample path large deviations proved in chapter 4. $\square$ 5.3.6

Definition 5.3.7 (Drift of $\mathbf{S}^{(i)}$)**.** $\mathbb{T}^{(i)}\mu_i\, p^{(i)} = \mu_i \big(- e_i + p^{(i)} \big)$

Claim 5.3.8. *The free drift $\lambda + (P^\top - id)\,\mu$ defined in 5.2.6 is the drift of the free process.*

Proof of 5.3.8: The drift of the process as its a.s limit under the scaling and wrt the supnorm over some compact interval.

$$\lim_{t\to\infty} \frac{1}{t} X_t \;=\; \lambda + \sum_{i=1}^d \lim_{t\to\infty} \frac{1}{t} \mathbf{S}_t^{(i)} \;=\; \lambda + \sum_{i=1}^d \mu_i \big(- e_i + p^{(i)} \big) \;=\; \lambda + (P^\top - \mathrm{id})\,\mu$$

$\square$ 5.3.8

For the network primitives and the free process the rates of section 5.2 describe the expected behaviour of the process in the limit of the scaling 3.4.1.

Lmgf for free process

We calculate the lmgf Ψ of the free process X and prove its strict convexity.

Definition 5.3.9 (Lmgf Ψ of the free process)**.** *If X is the free process build from the network primitives of a generalised Jackson network as defined in 5.1.5, 5.3.4 then the lmgf of X is*

$$\Psi(\theta) \;=\; \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\langle \theta, X_t \rangle}] \;=\; \sum_{i=1}^d \Gamma_A^{(i)}(\theta_i) + \Gamma_S^{(i)}(-\theta_i + \Xi^{(i)}(\theta)).$$

For the form of the lmgf we have applied independence of network primitives and the form of the lmgf of $\mathbf{S}^{(i)}$, in the representation of 5.3.3, of claim 4.4.9:

$$\mathbb{E}[e^{\langle \theta, X_t \rangle}] \;=\; \prod_{i=1}^d \mathbb{E}[e^{\theta_i\, A_t^{(i)}}]\, \mathbb{E}[e^{\langle \theta, \mathbf{S}_t^{(i)} \rangle}].$$

Claim 5.3.10. $\mathcal{D}(\Psi) = \mathbb{R}^d$ *and* Ψ *is strictly convex.*

Proof of 5.3.10: Finiteness of Ψ is a direct consequence of finiteness of all involved Γ's and Ξ's. Convexity of Ψ is immediate since it is the limit (in t) of strictly convex (in θ) functions (cf 2.2.8). In this approach in the limit strictness is lost, so we choose another: Assume all network primitives are non-deterministic. Let $\alpha, \beta \in \mathbb{R}^d$, $\alpha \neq 0$. We show that along the half line $\{\beta + c\alpha \,|\, c \geq 0\}$ the function $c \mapsto \Psi(\beta + c\,\alpha)$ is strictly convex.

We calculate derivatives in direction α.

$$\begin{aligned}
&\frac{d}{dc}\Psi(\beta + c\alpha) \\
&= \sum_{i=1}^{d} \alpha_i \Gamma_A^{(i)\prime}(\beta_i + c\alpha_i) \\
&\qquad + \Gamma_S^{(i)\prime}(-(\beta + c\alpha)_i + \Xi^{(i)}(\beta + c\alpha))\,(-\alpha_i + \frac{d}{dc}\Xi^{(i)}(\beta + c\alpha)) \\
&\frac{d^2}{dc^2}\Psi(\beta + c\alpha) \qquad\qquad (5.9)\\
&= \sum_{i=1}^{d} \alpha_i^2 \Gamma_A^{(i)\prime\prime}(\beta_i + c\alpha_i) \\
&\qquad + \Gamma_S^{(i)\prime\prime}(-(\beta + c\alpha)_i + \Xi^{(i)}(\beta + c\alpha))\,(-\alpha_i + \frac{d}{dc}\Xi^{(i)}(\beta + c\alpha))^2 \\
&\qquad + \Gamma_S^{(i)\prime}(-(\beta + c\alpha)_i + \Xi^{(i)}(\beta + c\alpha))\,\frac{d^2}{dc^2}\Xi^{(i)}(\beta + c\alpha)
\end{aligned}$$

where the Γ'' are non-negative (cf 2.2.5). Convexity of the $\Xi^{(i)}$ follows from their definition 4.4.8 (and convexity of $K^{(i)}$ and linearity of $\mathbb{T}$) or alternatively from the following application of Jensen's inequality.

$$\begin{aligned}
&\frac{d^2}{dc^2}\Xi^{(i)}(\beta + c\alpha) \qquad\qquad (5.10)\\
&= \frac{d}{dc}\frac{\sum_{j=1}^{d} \alpha_j\, p_{ij}\, \exp\{(\beta + c\alpha)_j\}}{\exp\{\Xi^{(i)}(\beta + c\alpha)\}} \\
&= \sum_{j=1}^{d} \alpha_j^2 \frac{p_{ij}\exp\{(\beta + c\alpha)_j\}}{\exp\{\Xi^{(i)}(\beta + c\alpha)\}} - \Big(\sum_{j=1}^{d} \alpha_j\, \frac{p_{ij}\exp\{(\beta + c\alpha)_j\}}{\exp\{\Xi^{(i)}(\beta + c\alpha)\}}\Big)^2 \\
&\geq\ 0
\end{aligned}$$

We have (5.10)$= 0$ for $p^{(i)}$ a point measure (when routing from node i is deterministic) or when the $\alpha_1, \ldots, \alpha_d$ are constant on the support of $p^{(i)}$.

We now argue for strict convexity. We show that for $\alpha, \beta \in \mathbb{R}^d$, $\alpha \neq 0$ the second derivative (5.9) is strictly positive.

Strict positivity of (5.9) follows if there is i such that $\lambda_i \, \alpha_i \neq 0$ since then $\sum_{i=1}^d \alpha_i^2 \, \Gamma_A^{(i)\prime\prime}(\alpha_i) > 0$ from strict convexity of the $\Gamma_A(\cdot)$'s. Otherwise $\sum_{i=1}^d \alpha_i^2 \, \Gamma_A^{(i)\prime\prime}(\alpha_i) = 0$ and we need the Γ_S's to argue for strict positivity. Let's assume that

- $\alpha, \beta \in \mathbb{R}^d$, $\alpha \not\equiv 0$
- $\forall i \,:\, \lambda_i > 0 \,\Rightarrow\, \alpha_i = 0$
- $\frac{d^2}{dc^2} \Psi(\beta + c\alpha) = 0$

and produce a contradiction. The derivatives $\Gamma_S^{(i)\prime}$ and $\Gamma_S^{(i)\prime\prime}$ are strictly positive by 2.2.8. So the other sums in (5.9) are $= 0$ iff

$$- \alpha_i + \frac{d}{dc} \Xi^{(i)}(\beta + c\alpha) = 0 \quad , \quad \frac{d^2}{dc^2} \Xi^{(i)}(\beta + c\alpha) = 0 \;\; \forall i \tag{5.11}$$

First and second derivatives of $\Xi^{(i)}$ on the half line can be written as expectation and variance under exponentially twisted $p^{(i)}$:

$$\begin{aligned} \frac{d}{dc} \Xi^{(i)}(\beta + c\alpha) \quad &= \quad \sum_{j=1}^d \alpha_j \frac{p_{ij} e^{(\beta + c\alpha)_j}}{e^{\Xi^{(i)}(\beta + c\alpha)}} \; = \; \mathbb{E}_{p^{(i)}(\beta + c\alpha)}[\alpha] \\ \frac{d^2}{dc^2} \Xi^{(i)}(\beta + c\alpha) \quad &\overset{(5.10)}{=} \quad \mathbb{V}_{p^{(i)}(\beta + c\alpha)}[\alpha] \end{aligned}$$

and (5.11) becomes

$$\alpha_i = \mathbb{E}_{p^{(i)}(\beta + c\alpha)}[\alpha] \quad , \quad \mathbb{V}_{p^{(i)}(\beta + c\alpha)}[\alpha] = 0 \;\; \forall i \tag{5.12}$$

From openess of our network we know there is at least one $\lambda_i > 0$ and by our assumption (2nd bullet) for this i we have $\alpha_i = 0$. From (5.11) we know that all α_j with $j \in \mathrm{supp}(p^{(i)})$ take the same value (from $\mathbb{V} = 0$) and that this value is again $= 0$ (from $0 = \alpha_i = \mathbb{E}$). So moving along a directed spanning forest with each tree of the spanning forest rooted at an entry node, we find that $\alpha_i = 0$ at each node we pass. From the spanning property we get $\alpha \equiv 0$, the contradiction.

If in the network there are deterministic primitives and some $\Gamma_A'' \equiv 0$ the proof works the same way if there is non-deterministic flow at each node / if there is a spanning forest of the network with only those entry nodes as roots that have non-deterministic arrival processes; all service processes have to be non-deterministic. This condition (non-det. flow at each node) compares to condition $(P, \mathcal{S})$ applied in theorem 2.2. in Puhalskii's [16]. $\square$ 5.3.10

Since we did not properly define a spanning forest we at least give examples.

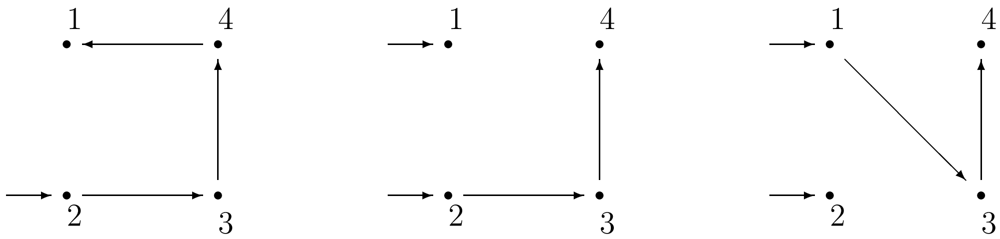


Figure 5.6: Three spanning forests for the network in figure 5.1

The first spanning forest in figure 5.6 is a spanning tree. If the network is not strongly connected a spanning tree does not necessarily exist.

Large deviations of the free process

We start with a one dimensional large deviation principle and proceed with the sample path large deviation principle.

Claim 5.3.11. *For the mean of the free process a large deviation principle holds: For open $G \subseteq \mathbb{R}$ and closed $F \subseteq \mathbb{R}$*

$$\begin{aligned} -\inf_{x\in G} \Psi^*(x) \leq \quad & \liminf_{n\to\infty} \tfrac{1}{n} \log P(\tfrac{1}{n} X_n \in G) \\ & \limsup_{n\to\infty} \tfrac{1}{n} \log P(\tfrac{1}{n} X_n \in F) \quad \leq -\inf_{x\in F} \Psi^*(x) \end{aligned}$$

with Ψ the lmgf of the free process of definition 5.3.9 and Ψ^* its Fenchel-Legendre transform.

Proof of 5.3.11: By application of the Gärtner-Ellis theorem. Strict convexity of Ψ and $\mathcal{D}(\Psi) = \mathbb{R}^d$ are important here (no regularisation in the Gärtner-Ellis theorem needed). The rate function is the convex Ψ^*, the Fenchel-Legendre transform of the lmgf. $\square$ 5.3.11

Claim 5.3.12. Ψ^* *is a good rate function and*

$$\Psi^*(v) \quad = \quad \inf_{\substack{a,r,\gamma \\ a+(r^\top - id)\gamma = v}} \sum_{i=1}^d \Gamma_A^{(i)*}(a_i) + \Gamma_S^{(i)*}(\gamma_i) + \gamma_i\, \Xi^{(i)*}(r^{(i)}).$$

Proof of 5.3.12: We have one dimensional large deviations with good rate functions for the arrival and split service processes that form the free process, cf 3.5.1. As summing is continuous in $\mathbb{R}$ we can apply the contraction principle (cf 7.5.2 in the appendix) to get a large deviation principle for the mean of the free process. As rate function we get

$$v \quad \mapsto \quad \inf_{\substack{a,s_1,\dots,s_d \in \mathbb{R}^d \\ a+\sum_{i=1}^d s_i = v}} \sum_{i=1}^d \Gamma_A^{(i)*}(a_i) + (\Gamma_S^{(i)} \circ (\Xi^{(i)} - \pi_i))^*(s_i)$$

Since the rate function of a large deviation object is unique this rate function has to equal Ψ^*.

$$\begin{aligned}
\Psi^*(v) \quad &= \quad \inf_{\substack{a,s_1,\dots,s_d \in \mathbb{R}^d \\ a+\sum_{i=1}^d s_i = v}} \sum_{i=1}^d \Gamma_A^{(i)*}(a_i) + (\Gamma_S^{(i)} \circ (\Xi^{(i)} - \pi_i))^*(s_i) \\
&\overset{4.4.21}{=} \quad \inf_{\substack{a,s_1,\dots,s_d \in \mathbb{R}^d \\ a+\sum_{i=1}^d s_i = v}} \sum_{i=1}^d \Gamma_A^{(i)*}(a_i) + \inf_{\substack{r^{(i)},\gamma_i \\ \gamma_i(r^{(i)} - e_i) = s_i}} \Gamma_S^{(i)*}(\gamma_i) + \gamma_i \Xi^{(i)*}(r^{(i)})
\end{aligned}$$

Now in the set of restrictions we write r as the matrix with rows $r^{(i)}$. Note that for $\Psi^*(v)$ to be finite rows have to be subprobabilities concentrated on the support of $p^{(i)}$, cf 4.4.22. We rewrite the restriction as

$$v = a + \sum_{i=1}^d s_i = a + \sum_{i=1}^d \gamma_i(r^{(i)} - e_i) \; = \; a + (r^\top - \mathrm{id})\gamma$$

and we have obtained the desired representation. From goodness of the primitives' rate functions (cf 2.6.2) all of the above representations of the rate function are good, too. $\square$ 5.3.12

Since we do have sample path large deviation principles for the arrival and split service processes we can apply the contraction principle to obtain sample path large deviations for the free process.

Claim 5.3.13. *The free process X obeys a sample path large deviation principle in $D([0,T],\mathbb{R}^d)$ equipped with the sup-norm induced topology under the scaling 3.4.1. The good rate function is*

$$\psi \mapsto \int_{t=0}^{T} \Psi^*(\psi'(t))\,dt$$

for $\psi \in AC[0,T]$, $\psi(0)=0$. The rate function equals ∞ for all other ψ.

Proof of 5.3.13: We have sample path large deviation principles for each primitive A and $\mathbf{S}^{(i)}$ (for $i=1,\dots,d$) in $(D([0,T],\mathbb{R}^d)$ equipped with the supremum norm induced topology. For the definition of $A,\mathbf{S}^{(i)}$ cf 5.3.1, 5.3.2 and for the sample path large deviation principles 4.3.10, 4.4.20. As primitives are independent we have the joint large deviation principle of

$$(A,\mathbf{S}^{(1)},\dots,\mathbf{S}^{(d)}) \in \left(D([0,T],\mathbb{R}^d)\right)^{\times(d+1)}$$

with the supremum norm

$$f \mapsto \max_{\substack{i=1,\dots,d\\ j=1,\dots,d+1}} ||f_{i,j}|| \qquad \text{for } f \in \left(D([0,T],\mathbb{R}^d)\right)^{\times(d+1)},\ ||f_{ij}|| = \sup_{t\in[0,T]} |f_{ij}(t)|$$

The rate function of the joint large deviation principle is the sum of individual rate functions by independence of the primitives. It is infinite for any f with an $f_{ij} \notin AC([0,T],\mathbb{R})$ as each rate function is concentrated on absolutely continuous functions. If all f_{ij} are absolutely continuous then for this element

$$f \in \left(\, D([0,T],\mathbb{R}^d)\,\right)^{\times(d+1)}$$

the rate function has the representation

$$f \mapsto \sum_{i=1}^{d} \int_{t=0}^{T} \Gamma_A^{(i)*}(f'_{i1}(t))\,dt + \sum_{j=2}^{d+1} \int_{t=0}^{T} \left(\Gamma_S^{(j)} \circ (-\pi_j + \Xi^{(j)})\right)^*(f'_{\cdot j}(t))\,dt$$

Now addition on $D([0,T],\mathbb{R}^d) \times \cdots \times D([0,T],\mathbb{R}^d)$ is a continuous map wrt the sup-norm induced topology. Applying the contraction principle we get a large deviation principle for the free process with rate function

$$\psi \quad \mapsto \quad \inf_{\substack{a,s_1,\dots,s_d\\ a+s_1+\cdots+s_d=\psi}} \sum_{i=1}^{d} \int_{t=0}^{T} \Gamma_A^{(i)*}(a'_i(t)) + \left(\Gamma_S^{(i)} \circ (-\pi_i + \Xi^{(i)})\right)^*(s'_i(t))\,dt$$

where the infimum can be taken over $a,s_1,\dots,s_d \in AC([0,T],\mathbb{R}^d)$. Note that this makes the rate function infinite as soon as $\psi \notin AC([0,T],\mathbb{R}^d)$.

We get a lower bound for the rate function by interchanging summation and integration and then optimisation and integration. With the infimum inside the integral it changes from an optimisation in function space to an optimisation in $\mathbb{R}^d$.

$$\begin{aligned}
&\inf_{\substack{a,s_1,\dots,s_d \in AC\\ a+s_1+\dots+s_d=\psi}} \sum_{i=1}^{d} \int_{t=0}^{T} \Gamma_A^{(i)*}(a_i'(t)) + \left(\Gamma_S^{(i)} \circ (-\pi_i + \Xi^{(i)})\right)^*(s_i'(t))\,dt \\
\geq\ & \int_{t=0}^{T} \sum_{i=1}^{d} \inf_{\substack{a,s_1,\dots,s_d \in \mathbb{R}^d\\ a+s_1+\dots+s_d=\psi'(t)}} \Gamma_A^{(i)*}(a_i) + \left(\Gamma_S^{(i)} \circ (-\pi_i + \Xi^{(i)})\right)^*(s_i)\,dt \\
=\ & \int_{t=0}^{T} \Psi^*(\psi'(t))\,dt
\end{aligned}$$

where the last equality is due to the different representations of Ψ^* we have already seen in the proof of 5.3.12.

Since there are no continuity restrictions for the derivatives and ψ is absolutely continuous, we get absolutely continuous functions from integrating infimisers found inside the integral. The $\geq$ is in fact an equality. $\square$ 5.3.13

Change of measure

We define a change of measure process for the free process associated with a generalised Jackson network.

Definition 5.3.14. *Consider a generalised Jackson network with*

- *primitives defined in 5.1.5, 5.3.4.*
- *Let X be the free process defined in 5.3.5*
- *with lmgf Ψ defined in 5.3.9.*

Define for $t \geq 0$ and $\alpha \in \mathbb{R}^d$

$$\mathcal{M}(\alpha, t) \;=\; \exp\{\langle \alpha\,,\, X_t\rangle - t\Psi(\alpha)\}\, r(\alpha, t)$$

where

$$\begin{aligned}
r(\alpha, t) &= \prod_{i=1}^{d} r^{(i)S}(\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)), t)\, r^{(i)A}(-\Gamma_A^{(i)}(\alpha_i), t) \\
r^{(i)D}(\beta, t) &= \frac{F_\beta^c}{F^c}(B(t))\, e^{-\beta B(t)}
\end{aligned}$$

for $D \in \{A, S\}$ and F the distribution function of inter event times of the renewal counting process $D^{(i)}$ and $B(t)$ the age of $D^{(i)}$ at time t (for the definition of $r^{(i)D}$ cf 3.6.5).

Claim 5.3.15.

- *$t \mapsto \mathcal{M}(\alpha, t)$ is a change of measure process for the free process.*
- *Under the change of measure $\mathcal{M}(\alpha, \cdot)$ the process $X = A + \sum_{i=1}^{d} \mathbf{S}^{(i)}$ is again a free process associated with the primitives of a generalised Jackson network. Each primitive changes its distribution in the following way:*
  - *if $A^{(i)}$ had inter event densities f then under the changed measure $A^{(i)}$ remains a renewal counting process and now has inter event times density $f_{-\Gamma_A^{(i)}(\alpha_i)}$, cf 2.3.1, (2.4).*
  - *if the routing decision $r^{(i)}$ was distributed on $\{e_1, \ldots, e_d, 0\}$ with probability measure $(p_{i1}, \ldots, p_{id}, 1 - \sum_{j=1}^{d} p_{ij})$ associated with the sub-probability measure $p^{(i)}$ then under the change of measure routing decisions remain iid. $r^{(i)}$ now has the distribution associated with the sub-probability measure $\nabla \Xi^{(i)}(\alpha)$, cf 4.4.23.*
  - *if $S^{(i)}$ had inter event densities f then under the changed measure $S^{(i)}$ remains a renewal counting process and now has inter event times density $f_{-\Gamma_S^{(i)} \circ (-\pi_i + \Xi^{(i)})(\alpha)}$, cf 2.3.1, (2.4).*

Corollary 5.3.16. *Under the change of measure $\mathcal{M}(\alpha, \cdot)$ the rates of the primitives change from λ, μ, P to*

- $\lambda_i(\alpha_i) = \Gamma_A^{(i)\prime}(\alpha_i)$ *for* $i = 1, \ldots, d$
- $\mu_i(\alpha) = \Gamma_S^{(i)\prime}(-\alpha_i + \Xi^{(i)}(\alpha))$ *for* $i = 1, \ldots, d$
- $P(\alpha)^\top = [\nabla \Xi^{(1)}(\alpha) \cdots \nabla \Xi^{(d)}(\alpha)]$

The Corollary fits 5.3.15 and is a summary of definitions 3.6.12, 4.4.17, (4.4.23).

Proof of 5.3.15: For the primitives $A^{(i)}, \mathbf{S}^{(i)}$ for $i = 1, \ldots, d$ we have in previous sections developed individual change of measure processes. For the arrival processes (definition 3.6.4 and representation (3.15)) with $\gamma \in \mathbb{R}$

$$(\gamma, t) \mapsto \exp\{\gamma A_t^{(i)} - t\Gamma_A^{(i)}(\gamma)\} r^{(i)A}(-\Gamma_A^{(i)}(\gamma), t)$$

and in vectorial notation with $\alpha \in \mathbb{R}^d$

$$(\alpha, t) \mapsto \exp\{\langle \alpha\,,\ A_t\rangle - t\sum_{i=1}^{d} \Gamma_A^{(i)}(\alpha_i)\} \prod_{i=1}^{d} r^{(i)A}(-\Gamma_A^{(i)}(\alpha_i), t)$$

and for $\mathbf{S}^{(i)}$ (cf definition 4.4.15)

$$(\alpha, t) \mapsto \exp\{\langle \mathbf{S}_t^{(i)}, \alpha\rangle - t\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\}\, r^{(i)S}(t, -\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)))$$

Since primitives are independent we can multiply changes of measure processes.

$$\begin{aligned}
&\prod_{i=1}^{d} \exp\{\langle \alpha\,,\ \mathbf{S}_t^{(i)}\rangle - t\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\} \times r^{S(i)}(\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)), t) \\
&\qquad \times \exp\{\langle \alpha\,,\ A_t\rangle - t\sum_{i=1}^{d} \Gamma_A^{(i)}(\alpha_i)\} \prod_{i=1}^{d} r^{A(i)}(-\Gamma_A^{(i)}(\alpha_i), t) \\
&= \quad \exp\{\langle \alpha\,,\ \underbrace{A_t + \sum_{i=1}^{d} \mathbf{S}_t^{(i)}}_{=X_t}\rangle - t\underbrace{\big(\sum_{i=1}^{d} \Gamma_A(\alpha_i) + \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\big)}_{=\Psi(\alpha)}\} \\
&\qquad \underbrace{\prod_{i=1}^{d} r^{S(i)}(\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)), t)\, r^{A(i)}(-\Gamma_A^{(i)}(\alpha_i), t)}_{=r(\alpha,t)} \\
&= \quad \exp\{\langle \alpha\,,\ X_t\rangle - t\Psi(\alpha)\}\, r(\alpha, t).
\end{aligned}$$

So $\mathcal{M}(\alpha, \cdot)$ is the compound change of measure we get when changing each primitive process separately. For the translation of the change of measure for $\mathbf{S}^{(i)}$ into that of $S^{(i)}$ and $p^{(i)}$ cf 4.4.15 and 4.4.16. $\square$

5.3.15

Application of the change of measure in the large deviation

The rate function for the large deviation for the mean of the free process is the Fenchel Legendre transform of the free process' lmgf.

$$\Psi^*(v) \quad = \quad \sup_{\alpha \in \mathbb{R}^d} \langle \alpha, v\rangle - \Psi(v)$$

We want to point out here that the drift of the free process under the change of measure $\mathcal{M}(\alpha,\cdot)$ with some $\alpha \in \mathbb{R}^d$ coincides with $\nabla\Psi(\alpha)$:

$$\begin{aligned}
&\frac{d}{d\theta_i}\Psi(\alpha) \\
&= \Gamma_A^{(i)\prime}(\alpha_i) - \Gamma_S^{(i)\prime}(-\alpha_i + \Xi^{(i)}(\alpha)) + \sum_{j=1}^{d} \Gamma_S^{(j)\prime}(-\alpha_j + \Xi^{(i)}(\alpha))\frac{d}{d\alpha_i}\Xi^{(j)}(\alpha) \\
&= \lambda_i(\alpha) - \mu_i(\alpha) + \sum_{j=1}^{d} \mu_j(\alpha)\, p_{ji}(\alpha)
\end{aligned}$$

Where for the last equality we applied corollary 5.3.16. We get

$$\nabla\Psi(\alpha) \;=\; \lambda(\alpha) + (P(\alpha)^\top - \mathrm{id})\,\mu(\alpha) \tag{5.13}$$

Claim 5.3.17. *If for $v \in \mathbb{R}^d$ there is $\overline{\alpha} \in \mathbb{R}^d$ such that under the change of measure $\mathcal{M}(\overline{\alpha},\cdot)$ the free process has drift v then*

$$\Psi^*(v) = \langle \overline{\alpha}, v\rangle - \Psi(\overline{\alpha}).$$

Proof of 5.3.17: From the definition of the Fenchel-Legendre transform and the above

$$\begin{aligned}
\Psi^*(v) \;=\; \sup_{\alpha\in\mathbb{R}^d} \langle \alpha, v\rangle - \Psi(\alpha) \quad &= \quad \langle \overline{\alpha}, v\rangle - \Psi(\overline{\alpha}) \\
&\Leftrightarrow \quad v \;=\; \nabla\Psi(\overline{\alpha})
\end{aligned}$$

so $\nabla\Psi(\alpha)$ is the free drift (cf 5.2.6) under the changed measure and if this equals v then $\theta = \alpha$ is the optimiser in the Fenchel-Legendre transform. □ 5.3.17

It is therefore interesting under which conditions an optimising $\overline{\alpha} \in \mathbb{R}^d$ in the Fenchel-Legendre transform $\Psi^*(v)$ exists.

Remark 5.3.18.

- *From identifying rate functions in 5.3.12 we have finiteness of $\Psi^*(v)$ iff there are a, r, γ such that $a + (r^\top - id)\gamma = v$. And from goodness of Ψ^* we have existence of an optimiser whenever $\Psi^*(v) < \infty$:*

$$\begin{aligned}
\exists a,\gamma,r \,:\, a + (r^\top - id)\,\gamma \;&=\; v \\
\Leftrightarrow \quad \exists \alpha \,:\, \lambda(\alpha) + (P(\alpha)^\top - id)\,\mu(\alpha) \;&=\; v
\end{aligned}$$

  so while $\alpha \in \mathbb{R}^d$ has d coordinates or degrees of freedom and a, γ, r have $2\,d + d^2$ this seems not to be relevant in terms of existence of optimisers.

- *From the form of the rate function in 5.3.12 and openness of the domains $\mathcal{D}(\Gamma_D^{(i)*})$ for all $D = S, i \in \{1, \dots, d\}$ and $D = A, i \in \{1, \dots, d\}$ with $\lambda_i > 0$ (cf 2.8.2): if for $v \in \mathbb{R}^d$ exists α such that $\nabla\Psi(\alpha) = v$ then for v' close enough to v there is α' such that $\nabla\Psi(\alpha') = v'$. This also implies that the domain $\mathcal{D}(\Psi^*)$ is open.*

5.3.2 The network process

This section is on the stochastic process that describes the generalised Jackson network. We will construct it from the network primitives in a similar way as the free process in such a way that its coordinates cannot become negative. We give the network process' expected behaviour.

The large deviations of the network process need a more thorough theoretical background. The local large deviations for the network process will be developed in chapter 6. The following theorem 2.1 of Chen and Mandelbaum [3] allows to move from the networks primitives defined in 5.3.4 to the network process that we define in 5.3.20.

Theorem 5.3.19. *Let $z_0 \in \mathbb{R}$ and consider a network of d nodes and primitives $A, \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(d)}$ that satisfy*

- *$A^{(1)}, \dots, A^{(d)}$ are in $D([0,\infty), \mathbb{R})$, non-decreasing and $A^{(i)}(0) = 0$ for $i = 1, \dots, d$;*
- *$\mathbf{S}^{(i)}, \dots, \mathbf{S}^{(d)}$, $\mathbf{S}^{(i)} = -e_i S^{(i)} + \sum_{j=1}^d S^{(ij)} e_j$ with $S^{(i)}, S^{(ij)}$ are in $D([0,\infty), \mathbb{R})$, nondecreasing, and $S^{(ij)}(0) = S^{(i)}(0) = 0$.*

If further there is $\epsilon > 0$ such that for each $D \in \{A^{(i)}, S^{(i)}, S^{(ij)} \,|\, i,j = 1, \dots, d\}$

$$D(t) \leq D(t-) + \epsilon$$

then there exists a unique pair of d-dimensional processes (Z, R) satisfying

- $Z(t) = z_0 + A(t) + \sum_{i=1}^d \mathbf{S}^{(i)} \circ R^{(i)}(t)$
- $Z^{(j)} \geq 0$
- $R^{(j)}(t) = \int_0^t \mathbb{1}_{Z^{(j)}(s)>0}\, ds$

Note that this theorem applies to the unscaled ($\epsilon = 1$) and scaled primitives ($\epsilon = \frac{1}{n}$).

Definition 5.3.20 (Network process). *Given network primitives $A, \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(d)}$ and some fixed $z_0 \in \mathbb{N}^d$ such that theorem 5.3.19 applies set*

$$\begin{aligned} Z(t, z_0) &= z_0 + A(t) + \sum_{i=1}^{d} \mathbf{S}^{(i)} \circ R^{(i)}(t) \\ R^{(i)}(t) &= \int_0^t 1\!\!1_{Z^{(i)}(s,z_0)>0}\, ds \quad (j = 1, \dots, d) \end{aligned}$$

and denote (Z, R) the network process. When addressed in isolation Z is denoted the queue size process and R the runtime process of the network.

The difference of the free and the network process is in the runtime process $R^{(i)}$ defined at each node. If queue i is empty during $[t_0, t_1]$ then during this time $Z^{(i)} = 0$ and $R^{(i)\prime} = 0$. Thus $\mathbf{S}^{(i)}$ cannot change state in $[t_0, t_1]$ and no customer leaves the empty queue at node i. The queue size always stays non-negative.

We have just described the different behaviour of a queue when empty and when nonempty. This is termed "discontinuous statistics" of the network process and it is the reason why the network process is more difficult to work with than the free process.

Definition 5.3.21 ($\mathbf{R}(\cdot,\cdot)$). *For $t \in \mathbb{R}_{>0}$ and Z a well behaved non-negative process on $[0, t]$ with values in $\mathbb{R}$ let*

$$\mathbf{R}(Z, t) = \int_{s=0}^{t} 1\!\!1_{Z(s)>0}\, ds.$$

Definition 5.3.22 (Scaled network process). *For $z_0 \in \mathbb{R}^d_{\geq 0}$ define the processes*

$$Z_n : \quad \mathbb{R}_{\geq 0} \;\to\; \mathbb{R}^d_{\geq 0} \qquad , \qquad t \;\mapsto\; \frac{1}{n} Z(nt, \lfloor n z_0 \rfloor)$$

and R_n with coordinate processes $R_n^{(i)}$ defined as

$$R_n^{(i)}(t) = \frac{1}{n} \mathbf{R}(Z^{(i)}(\cdot, \lfloor n z_0 \rfloor), nt)$$

Then (Z_n, R_n) is denoted the scaled network process starting in z_0. If the starting point is $z_0 = 0$ it may be omitted. When addressed in isolation Z_n may be referred to as the scaled queue size process and R_n as the scaled runtime process of the network.

From Chen and Mandelbaum (part of their theorem 5.1. Their assumed uniform convergence on compacts holds for the generalised Jackson network under our assumptions.) we get a statement paralleling claims 5.3.6, 5.3.8.

Theorem 5.3.23 (Drift of the network process)**.** *Consider the scaled network process (Z_n, R_n) of the generalised Jackson network with rates (λ, μ, P). Consider the Skorohod problem 5.2.14 with*

- *starting point z_0*
- *linear input functions X , $X(t) = z_0 + t\big(\lambda + (P^\top - id)\,\mu\big)$*
- *P*

and let (Z, Y) be the linear solution of the form $Z(t) = z_0 + tz$, $Y(t) = ty$ where z is the network drift and y is the loss rate. Then the network process converges almost surely uniformly on compacts

$$\lim_{n\to\infty}(Z_n, R_n) = \big(t \mapsto z_0 + tz\,,\; t \mapsto t(1-\rho)\,\big).$$

Chen and Mandelbaum denote the network drift z the fluid limit of the queue size process. And indeed it can be identified with the deterministic flow in the associated fluid network of section 5.2.1.

5.3.3 The local process

Observe that in the scaled network process $t \mapsto Z_n(t, z_0)$ (z_0 fixed) initially non empty nodes stay non-empty for some positive length of time and since $R^{(i)}(t) = t$ over this interval the network process looks like the free process in this i-th coordinate and over this interval. Figure 5.7 gives a realisation of the non-ergodic network of example 5.1, 5.2.8 over some short time span $[0, T]$, $T = 0.1$ with a scaling parameter $n = 500$. This motivates the local process which is somewhere in between the free and the network process. We work with the local process to prove local large deviations in chapter 6.

Definition 5.3.24 (W^Λ, R^Λ)**.** *Given network primitives $A, \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(d)}$ and some fixed $z_0 \in \mathbb{R}^d_{\geq 0}$ let*

$$\Lambda = \Lambda(z_0) = \{i \,|\, z_{0,i} > 0\}$$

and define the local process W^Λ as

$$\begin{aligned} W^\Lambda: \quad \mathbb{R}_{\geq 0} \times \mathbb{R}^d_{\geq 0} &\to \mathbb{Z}^d \\ (t, z_0) &\mapsto z_0 + A(t) + \sum_{i\in\Lambda} \mathbf{S}^{(i)}(t) + \sum_{i\in\Lambda^c} \mathbf{S}^{(i)} \circ R^{(i)\Lambda}(t) \end{aligned}$$

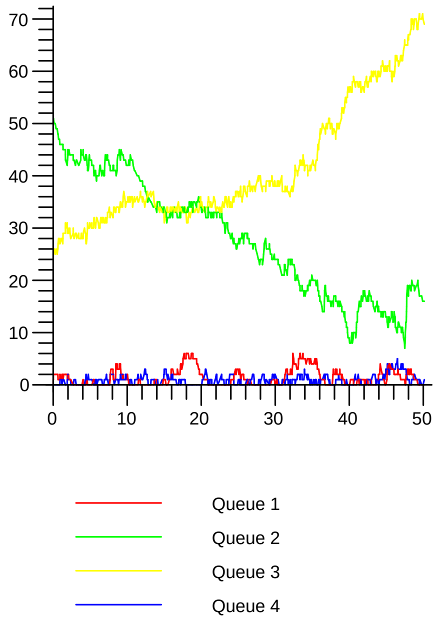


Figure 5.7: Realisation of the non-ergodic network or associated local process with $\Lambda = \{2, 3\}$, cf examples 5.2.8 ($T = 0.1$ and scaling $n = 500$)

and the local runtime process for $i \in \Lambda^c$

$$R^{(i)\Lambda}(t) = \mathbf{R}(W_i^\Lambda(\cdot, z_0), t) = \int_{s=0}^{t} 1\!\!1_{W^{(i)\Lambda}(s,z_0)>0}\, ds$$

The local process W^Λ will have non-negative Λ^c-coordinates: $W_i^\Lambda \geq 0$ for $i \in \Lambda^c$ since the runtime $R^{(i)\Lambda}$ will stop $\mathbf{S}^{(i)}$ if $W_i^\Lambda = 0$. Λ-coordinates of the local process are allowed to have negative values. We denote $i \in \Lambda$ a free node and $i \in \Lambda^c$ a restricted node.

We split the local process into the sum of a free and a network process:

Claim 5.3.25. *Let* $A, \mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(d)}$ *be network primitives of a generalised Jackson network,* $z_0 \in \mathbb{R}^d_{\geq 0}$ *and* (Z, R) *the associated network process starting in* z_0*. Set* $\Lambda = \{i \,|\, z_{0,i} > 0\}$ *and let* π_M *be the projection such that* $\pi_\Lambda(\alpha) = \sum_{i\in\Lambda} \alpha_i e_i$*. Then*

$$W^\Lambda = z_0 + X^\Lambda + Z^\Lambda$$

for the following Z^Λ, X^Λ*:*

- Z^Λ *is the network process of the* Λ^c*-nodes with all nodes starting empty.* $Z^\Lambda(t) \in \mathbb{R}^d_{\geq 0}, Z^{(i)\Lambda} \equiv 0$ *for* $i \in \Lambda$

$$\begin{aligned} Z^\Lambda &= \pi_{\Lambda^c}(W^\Lambda) \\ &= \sum_{i\in\Lambda^c} \big(A^{(i)} + \sum_{k\in\Lambda} S^{(ki)} - S^{(i)} \circ \boldsymbol{R}(S^{(i)}, t) + \sum_{k\in\Lambda^c} S^{(ki)} \circ \boldsymbol{R}(S^{(ki)}, t) \big) \, e_i \end{aligned}$$

- X^Λ *is the free process of the* Λ*-nodes with* $X^\Lambda(t) \in \mathbb{R}^d, X^\Lambda(0) = 0$

$$\begin{aligned} X^\Lambda &= \pi_\Lambda(W^\Lambda) - z_0 \\ &= \sum_{i\in\Lambda} \big(A^{(i)} + \sum_{k\in\Lambda^c} S^{(ki)} \circ \boldsymbol{R}(S^{(ki)}, t) - S^{(i)} + \sum_{k\in\Lambda^c} S^{(ki)} \big) \, e_i \end{aligned}$$

Proof of 5.3.25: We only have to check that the projections $\pi_\Lambda(W^\Lambda)$ and $\pi_{\Lambda^c}(W^\Lambda)$ are correct.

The interpretation as a network process for Z^Λ is correct since 5.3.20 applies to the primitives

- arrival processes $A^{(i)\Lambda} = A^{(i)} + \sum_{k\in\Lambda} S^{(ki)}$ for $i \in \Lambda^c$;
- service processes $-e_i S^{(i)} + \sum_{k\in\Lambda^c} S^{(ki)}$.

Similarly X^Λ is a free process as in 5.3.5 with

- arrival processes $t \mapsto A^{(i)}(t) + \sum_{k\in\Lambda^c} S^{(ki)} \circ \mathbf{R}(S^{(ki)}, t)$ for $i \in \Lambda$ where $\mathbf{R}$ makes distributions of these arrival processes difficult;
- service processes $-e_i S^{(i)} + \sum_{k\in\Lambda} S^{(ki)}$.

□
5.3.25

The new arrival processes $A^{(i)\Lambda}$, $i \in \Lambda^c$ of Z^Λ defined in the proof above as

$$A^{(i)\Lambda} = A^{(i)} + \sum_{k\in\Lambda} S^{(ki)}$$

for $i \in \Lambda^c$ are no renewal counting processes but the sum of independent renewal counting processes. The rates of the network process Z^Λ are

$$\sum_{i\in\Lambda^c} \big(\lambda_i + \sum_{k\in\Lambda} \mu_k p_{ki} \big) \, e_i \quad , \quad \mu \quad , \quad [\, p_{ij} \, \mathbb{1}_{i,j\in\Lambda^c} \,]_{i,j=1,\dots,d}$$

since $Z^{(i)\Lambda} = 0$ for $i \in \Lambda$ nodes in Λ are not relevant in the Z^Λ network and we move to $\mathbb{R}^{|\Lambda^c|}$, describing the (sub)network of Λ^c-nodes as

$$\lambda_{\Lambda^c} + (P^\top)_{\Lambda^c\Lambda}\, \mu_\Lambda = \Big[\, \lambda_i + \sum_{k\in\Lambda} \mu_k p_{ki} \,\Big]_{i\in\Lambda^c} \;\; , \;\; \mu_{\Lambda^c} = [\, \mu_i \,]_{i\in\Lambda^c} \;\; , \;\; P_{\Lambda^c} = [\, p_{ij} \,]_{i,j\in\Lambda^c}$$

These rates are like in (5.4) with Λ instead of B and Λ^c instead of E. Λ^c nodes are not necessarily ergodic.

Exit nodes of the Λ^c-network are $\{i \in \Lambda^c \,|\, \exists j \in \Lambda \cup \{0\} \, : \, p_{ij} > 0\}$ and for $p^{(i)\Lambda}$ a row of P_{Λ^c} and a sub-probability measure $p^\Lambda_{i0} = p_{i0} + \sum_{k\in\Lambda} p_{ik}$.

Claim 5.3.26. *Z^Λ satisfies assumption 5.1.9.*

Proof of 5.3.26: If Z is open then Z^Λ is, too: In a picture of an open network there is a sequence of arrows from the outside world to an arbitrary node i and from that node to the outside world. By remodelling a subset of nodes as a network we get "more outside world" which makes sequences from the outside world to node i shorter but never looses the required accessability. Feedback cannot be created by removing edges from the network. $\square$ 5.3.26

We can now investigate if the subnetwork of Λ^c-nodes described through Z^Λ is ergodic or find the ergodic subnetwork. The network has deterministic rates and definitions and claims of section 5.2 apply.

Change of measure

We define the change of measure process for the local process associated with a generalised Jackson network for a starting point $z_0 \in \mathbb{R}^d_{\geq 0}$ and set $\Lambda = \{i \,|\, z_{0,i} > 0\}$ of initially non-empty nodes.

Observing the local process W^Λ up to some fixed t we observe all arrival processes and the service processes of Λ-nodes up to time t. For restricted nodes $i \in \Lambda^c$ we only have observed the service process and the routing decisions up to time $R^{(i)\Lambda}(t) \leq t$. If node i is ergodic with high probability $R^{(i)\Lambda}(t) < t$. We want to argue that the change of measure process for the i-th service process we developed in section 4.4.2, esp. definition 4.4.15, is still a change of measure process under the time change $t \mapsto R^{(i)\Lambda}(t)$.

Definition 5.3.27 (Local filtration).

$$\mathcal{F}^\Lambda_t = \sigma\big(W^\Lambda_s \,;\, s \leq t\big) \quad , \quad \mathcal{F}^\Lambda = \big(\mathcal{F}^\Lambda_t \,;\, t \geq 0\big)$$

Then $(W_t^\Lambda; t \geq 0)$ is adapted to the filtration $(\mathcal{F}_t^\Lambda)_{t\geq 0}$ and the runtime $t \mapsto R^{(i)\Lambda}(t)$ is adapted to this filtration, since $R^{(i)\Lambda}(t) = \mathbf{R}(W^\Lambda(\cdot, z_0), t)$ is a measurable function of W^Λ.

Let $\mathcal{G}^{(i)}$ for some fixed $i \in \Lambda^c$ be the change of measure process for the counting process $\mathbf{S}^{(i)}$ (cf 5.3.2, 4.4.15):

$$\mathcal{G}^{(i)}(\alpha, t) = \exp\{\langle \alpha, \mathbf{S}_t^{(i)} \rangle - t\, \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\}\, r(-\alpha_i + \Xi^{(i)}(\alpha), t)$$

Claim 5.3.28. *$t \mapsto \mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(t))$ is a martingale wrt $\mathcal{F}^\Lambda$.*

Proof of 5.3.28: $\mathcal{G}^{(i)}(\alpha, \cdot)$ is a martingale wrt the filtration

$$\Big(\sigma\big(\mathbf{S}_s^{(i)}\, ;\, s \leq t \,\big) \Big)_{t \geq 0}$$

generated by $\mathbf{S}^{(i)}$ and for fixed $t \geq 0$ the runtime $R^{(i)\Lambda}(t)$ is a stopping time wrt $\mathcal{F}^\Lambda$. Thus $\mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(t))$ is a regular random variable with mean 1 (the same mean as $\mathcal{G}^{(i)}(\alpha, t)$) and measurable wrt $\sigma(\mathbf{S}_s^{(i)}\, ;\, s \leq R^{(i)\Lambda}(t))) \subseteq \mathcal{F}_t^\Lambda$.

If $s < t$ we have two ordered, bounded stopping times $R^{(i)\Lambda}(s) \leq R^{(i)\Lambda}(t)$ and

$$\mathbb{E}[\mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(t)) \,|\, \mathcal{F}_s^\Lambda] = \mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(s))$$

From this we have the martingale property for any $s_1 < \cdots < s_n$ or $0 = s_0 < s_1 < \cdots < s_{n-1} < s_n = T$ of the following finite dimensional vector:

$$\big(1, \mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(s_1)), \ldots, \mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(s_{n-1})), \mathcal{G}^{(i)}(\alpha, T)\big)$$

And $t \mapsto \mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(t))$ is a local martingale. We now show that for any $t \geq 0$ we have $\mathbb{E}[\sup_{s\in[0,t]} \mathcal{G}^{(i)}(\alpha, s)] < \infty$ and then the martingale property follows (cf proposition A.7 in [18]).

$$\begin{array}{rcl}
\sup\limits_{t\in\mathbb{R}} r(-\alpha_i + \Xi^{(i)}(\alpha), t) & < & \infty \\
-s\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) & \leq & \max\{0\,,\, -t\,\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\,\} \\
\mathbb{E}[\exp\{\langle \alpha\,,\, \mathbf{S}_s^{(i)} \rangle\}] & = & \mathbb{E}[\exp\{\langle \alpha\,,\, \mathbb{T}^{(i)} S_s^{(i)\mathrm{sp}} \rangle\}] \\
 & \overset{5.3.3}{=} & \mathbb{E}[\exp\{\langle \mathbb{T}^{(i)\top}\alpha\,,\, S_s^{(i)\mathrm{sp}} \rangle\}] \\
 & \overset{\beta = \mathbb{T}^{(i)\top}\alpha}{=} & \mathbb{E}[\exp\{\langle \beta\,,\, S_s^{(i)\mathrm{sp}} \rangle\}]
\end{array}$$

Replace β_i by β_i^+ then the last expression on the rhs is monotone increasing in s and

$$\mathbb{E}[\sup_{s\in[0,t]} \mathcal{G}^{(i)}(\alpha, s)] \leq \mathbb{E}[\exp\{\langle \beta^+\,,\, S_t^{(i)\mathrm{sp}} \rangle\}] \,<\, \infty.$$

$\square$
5.3.28

Corollary 5.3.29. *$t \mapsto \mathcal{G}^{(i)}(\alpha, R^{(i)\Lambda}(t))$ is the change of measure process for $t \mapsto \mathbf{S}^{(i)} \circ R^{(i)\Lambda}(t)$, the primitive service process at node i in the setting of the local process.*

The corollary brings together claim 5.3.28 and the additional property of a mean equal to unity.

Definition 5.3.30 ($\mathcal{M}^{\Lambda}(\alpha, \cdot)$). *Consider a generalised Jackson network with*

- *primitives defined in 5.1.5, 5.3.4.*
- *Let X be the free process defined in 5.3.5*
- *with lmgf Ψ defined in 5.3.9.*
- *Let $(W^{\Lambda}, R^{\Lambda})$ be the local process for some $\Lambda \subseteq \{1, \ldots, d\}$.*

Define for $t \geq 0$ and $\alpha \in \mathbb{R}^d$

$$\begin{aligned}
\mathcal{M}^{\Lambda}(\alpha, t) = \exp\{\langle \alpha \,, W_t^{\Lambda} \rangle \quad &- \quad \sum_{i=1}^{d} t\Gamma_A^{(i)}(\alpha_i) \\
&- \quad \sum_{i \in \Lambda} t\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
&- \quad \sum_{i \in \Lambda^c} R^{(i)\Lambda}(t)\, \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\} \times r(\alpha, t)
\end{aligned}$$

where

$$\begin{aligned}
r(\alpha, t) \quad &= \quad \prod_{i=1}^{d} r^{(i)S}(\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)), t)\, r^{(i)A}(-\Gamma_A^{(i)}(\alpha_i), t) \\
r^{(i)D}(\beta, t) \quad &= \quad \frac{F_\beta^c}{F^c}(B(t))\, e^{-\beta B(t)}
\end{aligned}$$

for $D \in \{A, S\}$ and F the distribution function of inter event times of the renewal counting process $D^{(i)}$ and $B(t) = B(D^{(i)}, t)$ the age of $D^{(i)}$ at time t (for the definition of $r^{(i)D}$ cf 3.6.5).

Note that stochasticity of $\mathcal{M}^{\Lambda}(\alpha, t)$ is in the inner product of W^{Λ} and in the runtimes $R^{(i)\Lambda}$ for $i \in \Lambda^c$.

Claim 5.3.31.
- *$t \mapsto \mathcal{M}^{\Lambda}(\alpha, t)$ is a change of measure process for the local process.*

- *Under the change of measure $\mathcal{M}^{\Lambda}(\alpha,\cdot)$ the process W^{Λ} is again a local process associated with the primitives of a generalised Jackson network and the same set Λ. Each primitive changes its distribution in the following way:*
  - *if $A^{(i)}$ had inter event densities f then under the changed measure $A^{(i)}$ remains a renewal counting process and now has inter event times density $f_{-\Gamma_A^{(i)}(\alpha_i)}$, cf 2.3.1, (2.4).*
  - *if the routing decision $r^{(i)}$ was distributed on $\{e_1,\ldots,e_d,0\}$ with probability measure $(p_{i1},\ldots,p_{id},1-\sum_{j=1}^{d}p_{ij})$ associated with the sub-probability measure $p^{(i)}$ then under the change of measure routing decisions remain iid. $r^{(i)}$ now has the distribution associated with the sub-probability measure $\nabla\Xi^{(i)}(\alpha)$, cf 4.4.23.*
  - *if $S^{(i)}$ had inter event densities f then under the changed measure $S^{(i)}$ remains a renewal counting process and now has inter event times density $f_{-\Gamma_S^{(i)}\circ(-\pi_i+\Xi^{(i)})(\alpha)}$, cf 2.3.1, (2.4).*

Proof of 5.3.31: Can we sequentially change the distributions of the primitive processes. We do it over different times: over $[0,t]$ for the arrival processes and the $\mathbf{S}^{(i)}$ with $i\in\Lambda$. For the restricted nodes $i\in\Lambda^c$ we change the distribution of $\mathbf{S}^{(i)}$ over $[0,R^{(i)}(t)]$. This still works in the new setting due to 5.3.29. Since $R^{(i)}(t)$ is a random variable in our setting and a stopping time we can work with the stopped martingale as the change of measure. $\square$ 5.3.31

Since the change of measure $\mathcal{M}^{\Lambda}(\alpha,\cdot)$ has the same effect on the network primitives as in the case of the free process, the rates of the network's primitives change in the same way and corollary 5.3.16 holds under the change $\mathcal{M}^{\Lambda}(\alpha,\cdot)$.

Definition 5.3.32 ($\mathbb{E}^{[\alpha]}$). *For any fixed $t\geq 0$ and set $A\in\mathcal{F}_t^{\Lambda}$*

$$\mathbb{E}[\mathbb{1}_{(W_s^{\Lambda}\,;\,s\leq t)\in A}\mathcal{M}^{\Lambda}(\alpha,t)]=\mathbb{E}^{[\alpha]}[\mathbb{1}_{(W_s^{\Lambda}\,;\,s\leq t)\in A}].$$

Remark 5.3.33.

- *The local process has been defined with explicit reference to a starting point z_0. This suits our following application of the local process associated with a network process $Z=Z(\cdot,z_0)$. However, local processes can be defined more generally, not requiring that $\Lambda=\Lambda(z_0)$.*
- *If $\Lambda=\emptyset$ then $\mathcal{M}^{\Lambda}(\alpha,\cdot)=\mathcal{M}^{\emptyset}(\alpha,\cdot)$ is a change of measure process for the network process.*

Chapter 6

Local large deviations of the generalised Jackson network

For the local sample path large deviations for a Markovian network we cite the definition of Irina Ignatiouk-Robert in the introduction of [13], "Large Deviations for Processes with Discontinuous Statistics". The paper is concerned with how to develop full large deviations for Markovian processes with discontinuous statistics starting from local large deviations.

Definition 6.0.2 (Local large deviations [13])**.** *Let $x \in \mathbb{R}^d_{\geq 0}$ and $(X(t,x))$ be a Markov process on $E \subseteq \mathbb{R}^d_{\geq 0}$ with initial state $X(0,x) = x$. For $n \in \mathbb{N}$, $(Z_n(t,z))$ is the rescaled Markov process on $\mathcal{E}_n = \frac{1}{n}E$ and having initial state $Z_n(0,z) = z \in \mathcal{E}_n$:*

$$Z_n(t,z) = \frac{1}{n}X(nt,nz).$$

A local sample path large deviation principle with a rate function $J_{[0,T]}$ is said to hold when the following inequalities are satisfied:

$$(3)\ :\ \lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\inf_{\substack{z\in\mathcal{E}_n\\ |z-\psi(0)|<\epsilon}}\frac{1}{n}\log\mathbb{P}_z(||\psi - Z_n||_\infty < \delta) \geq -J_{[0,T]}(\psi)$$

$$(4)\ :\ \lim_{\delta\to 0}\limsup_{n\to\infty}\sup_{\substack{z\in\mathcal{E}_n\\ |z-\psi(0)|<\delta}}\frac{1}{n}\log\mathbb{P}_z(||\psi - Z_n||_\infty < \delta) \leq -J_{[0,T]}(\psi)$$

for every piecewise linear function $\psi\ :\ [0,T] \to \mathbb{R}^d_{\geq 0}$.

From the Markov property it is deduced in [13] that one only needs to consider linear functions. We will work only with linear functions for the non-Markovian generalised Jackson network since we have shown that its primitives are exponentially equivalent to primitive proceses that have independent increments over finitely many, deterministic, disjoint intervals, cf section 3.4.2. This can be generalised to independent evolution of the network over finitely many, deterministic, disjoint intervals of time, over which the process stays close to some linear function over each interval (we can thereby bound runtimes and will obtain deterministic intervals for the service processes of the restricted nodes.).

Since a piecewise linear function over a compact interval will hit and leave boundaries at a finite number of fixed instances of time, independence of the network evolution over such intervals should be enough to move from linear to piecewise linear functions. We work with a slightly weaker definition of local large deviations.

Definition 6.0.3 (Local large deviations)**.** *For each $n \in \mathbb{N}$ and fixed z_0 let $Z_n(\cdot, z_0)$ be a scaled network process starting in z_0, cf 5.3.22. A local sample path large deviation principle is said to hold when for any x, v, T such that*

- $x_i = 0 \Rightarrow v_i \geq 0$
- $x_i > 0 \Rightarrow x_i + Tv_i > 0$

the following inequalities are satisfied:

$$(1) : \lim_{\delta\to 0}\lim_{\epsilon\to 0}\liminf_{n\to\infty}\inf_{|z_0-x|<\epsilon}\frac{1}{n}\log\mathbb{P}(||(t\mapsto x+tv)-Z_n(\cdot,z_0)||<\delta) \geq -TL(x,v)$$

$$(2) : \lim_{\delta\to 0}\lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{|z_0-x|<\epsilon}\frac{1}{n}\log\mathbb{P}(||(t\mapsto x+tv)-Z_n(\cdot,z_0)||<\delta) \leq -TL(x,v)$$

with $||\cdot||$ the supremum norm over the interval $[0, T]$.

We prove the following local large deviation principle for the non-Markovian generalised Jackson network. We explicitly allow linear functions that leave a boundary. Such a situation is given if there is i such that $x_i = 0, v_i > 0$.

Claim 6.0.4 (Local large deviations for the generalised Jackson network)**.** *Consider the generalised Jackson network with d nodes and primitives 5.1.5. Let $\Gamma_A^{(i)}, \Gamma_S^{(i)}, \Xi^{(i)}$ be the the lmgfs for the primitives and Ψ the lmgf for the free process. Under assumptions 5.0.2 for the inter event times and 5.1.9 for the network a sample path local large deviation principle holds with rate function*

$$L(x, v) = \sup_{\alpha\in\mathcal{B}_K}\langle\alpha, v\rangle - \Psi(\alpha)$$

with

$$\begin{aligned} K &= \{i \,|\, x_i > 0 \text{ or } v_i > 0\} \\ \mathcal{B}_K &= \{\alpha \in \mathbb{R}^d \,|\, -\alpha_i + \Xi^{(i)}(\alpha) \geq 0 \,\forall i \in K^c\}. \end{aligned}$$

From the definition of the local large deviation we make the general assumption

Assumption 6.0.5. *x, v, T are such that*

- *$x_i = 0 \Rightarrow v_i \geq 0$ and*
- *$x_i > 0 \Rightarrow x_i + T\,v_i > 0$.*

In the local large deviation we bound the probability that the queue size process stays close to a specified linear function. Since queue sizes will always be non-negative this assumption picks linear functions that a network process may have positive probability of staying close to; it is not restrictive.

In the next section we will prove the upper bound of the local large deviation principle and obtain a candidate for the local rate function $L(\cdot,\cdot)$ as an inequality constrained optimisation problem. We will then investigate existence of an optimiser in this optimisation problem; an optimiser can be interpreted as the parameter of a change of measure for the network process. We will further investigate properties of the network under the change of measure with the optimising α as parameter and finally prove the lower bound. As the lower and upper bound coincide the candidate local rate function is the local rate function and 6.0.4 will be proved.

6.1 Local large deviations upper bound

We are interested in the event

$$Z_n(\cdot, z) \in \mathcal{U}_\delta(t \mapsto x + t\,v) \tag{6.1}$$

over some interval $[0, T]$ for the scaled network process Z_n starting in $z \in \mathbb{R}^d$ and the asymptotic decay of the probability that the process stays in the neighbourhood over an interval of positive length as $n \to \infty$.

We start with linear functions where empty queues stay empty and the scaled network process has the same starting point as the linear function $t \mapsto x + tv$.

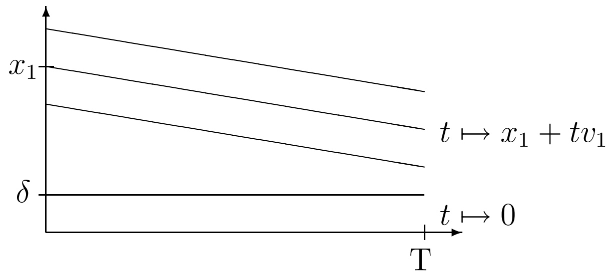


Figure 6.1: $d = 2$, $x_2 = 0$, $v_2 = 0$

Claim 6.1.1. *Consider the generealised Jackson network with d nodes and primitives 5.1.5. Let $\Gamma_A^{(i)}, \Gamma_S^{(i)}, \Xi^{(i)}$ be the the lmgfs for the primitives and Ψ the lmgf for the free process. Let Z_n be the scaled queue size process of the generalised Jackson network. Given x, v, T such that assumption 6.0.5 holds and $x_i = 0 \Rightarrow v_i = 0$*

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+t\,v)) \;\le\; -T\sup_{\alpha\in\mathcal{B}_\Lambda}\langle\alpha,v\rangle-\Psi(\alpha)$$

with the set $\mathcal{B}_\Lambda$ defined as

$$\begin{aligned}\Lambda &= \{i\,|\,x_i>0\}\\ \mathcal{B}_\Lambda &= \{\alpha\in\mathbb{R}^d\,|\,-\alpha_i+\Xi^{(i)}(\alpha)\ge 0\,\forall i\in\Lambda^c\}.\end{aligned}$$

Proof of 6.1.1: By the assumption on x, v, T we have $R^{(i)}(t) = t$ for all $i \in \Lambda$ and $t \le nT$ while $Z_n(\cdot, x) \in \mathcal{U}_\delta(t \mapsto x + tv)$. Thus we can exchange the network process for the local process:

$$Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+t\,v) \quad\Leftrightarrow\quad W_n^\Lambda(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+t\,v)$$

We uniformise in x:

$$\begin{aligned}&W_n^\Lambda(t,x)-(x+tv) \qquad (6.2)\\ &= Z_n(0,x)+A_n(t)+\sum_{i\in\Lambda}\mathbf{S}_n^{(i)}(t)+\frac{1}{n}\sum_{i\in\Lambda^c}^{d}\mathbf{S}^{(i)}\circ R^{(i)}(nt)-x-tv\\ &= \underbrace{\frac{\lfloor nx\rfloor}{n}-x}_{\in[-\frac{1}{n},0]}+\underbrace{A_n(t)+\sum_{i\in\Lambda}\mathbf{S}_n^{(i)}(t)+\frac{1}{n}\sum_{i\in\Lambda^c}^{d}\mathbf{S}^{(i)}\circ R^{(i)}(nt)-tv}_{=W_n^\Lambda(t,0)-tv}\end{aligned}$$

As we have removed x from (6.2) we remove it in the notation and write $W_n^\Lambda(t)$ instead of $W_n^\Lambda(t,0)$. The difference $\frac{\lfloor nx \rfloor}{n} - x$ forces us to change the δ of our neighbourhood to some δ' with $|\delta - \delta'| \leq \frac{1}{n}$ but we choose to ignore this notational nuisance. We are now investigating the event

$$W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)$$

and will apply the change of measure $\mathcal{M}^\Lambda$ (cf 5.3.24) with parameter $\alpha \in \mathbb{R}^d$.

$$\begin{aligned}\mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)) &= \mathbb{E}[1\!\!1_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)}] = \mathbb{E}[1\!\!1_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)} \frac{\mathcal{M}^\Lambda(\alpha, nT)}{\mathcal{M}^\Lambda(\alpha, nT)}] \\ &= \mathbb{E}^{[\alpha]}[1\!\!1_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)} \frac{1}{\mathcal{M}^\Lambda(\alpha, nT)}] \qquad (6.3)\end{aligned}$$

We bound

$$\begin{aligned}\frac{1}{\mathcal{M}^\Lambda(\alpha, nT)} &= \exp\{- \underbrace{\langle \alpha, W_{nT}^\Lambda \rangle}_{=\langle \alpha, W_{nT}^\Lambda - nTv\rangle + \langle \alpha, nTv\rangle} + nT\big(\sum_{i=1}^d \Gamma_A^{(i)}(\alpha_i) \\ &\quad + \sum_{i \in \Lambda} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\ &\quad + \sum_{i \in \Lambda^c} \underbrace{\frac{R^{(i)}(nT)}{nT}}_{\leq 1} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\big)\} \frac{1}{r(\alpha, R(nT))} \qquad (6.4)\end{aligned}$$

To get an upper bound we restrict α such that

$$\mathcal{B}_\Lambda := \{\alpha \in \mathbb{R}^d \,|\, -\alpha_i + \Xi^{(i)}(\alpha) \geq 0 \quad \forall i \in \Lambda^c\} \qquad (6.5)$$

resulting in

$$\frac{1}{\mathcal{M}^\Lambda(\alpha, nT)} \leq \exp\Big\{ - \langle \alpha, W_{nT}^\Lambda - nTv\rangle + nT\big(\Psi(\alpha) - \langle \alpha, v\rangle\big)\Big\} \frac{1}{r(\alpha, R(nT))}$$

We go on with the bound

$$\begin{aligned}&\mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)) \\ &\overset{(6.4)}{\leq} \mathbb{E}^{[\alpha]}[1\!\!1_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)} \exp\{ \underbrace{\langle \alpha, nTv - W_{nT}^\Lambda \rangle}_{\leq ||\alpha|| \cdot ||nTv - W_{nT}^\Lambda|| < ||\alpha|| nT\delta} + nT\big(\Psi(\alpha) - \langle \alpha, v\rangle\big)\} \\ &\quad \frac{1}{r(\alpha, R(nT))}] \qquad (6.6) \\ &\leq \underbrace{\mathbb{E}^{[\alpha]}[1\!\!1_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)} \frac{1}{r(\alpha, R(nT))}]}_{\leq 1 \sup \frac{1}{r} < \infty} \exp\{||\alpha||\, n\,T\,\delta + nT\big(\Psi(\alpha) - \langle \alpha, v\rangle\big)\end{aligned}$$

With the expectation finite uniformly in n and δ by $\mathbb{1} \leq 1$ and $\frac{1}{r(\alpha, R(nT))}$ a product of bounded terms independent of δ and n, cf claim 3.6.11.

For fixed $\alpha \in \mathcal{B}_\Lambda$ we have

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)) \quad \leq \quad T\big(\Psi(\alpha) - \langle \alpha, v \rangle\big)$$

Optimising over $\alpha \in \mathcal{B}_\Lambda$ we get the desired upper bound

$$\inf_{\alpha \in \mathcal{B}_\Lambda} T(\Psi(\alpha) - \langle \alpha, v \rangle) \;=\; -T \sup_{\alpha \in \mathcal{B}_\Lambda} \langle \alpha, v \rangle - \Psi(\alpha)$$

□
6.1.1

The upper bound $-T \sup_{\alpha \in \mathcal{B}_\Lambda} \langle \alpha, v \rangle - \Psi(\alpha)$ in 6.1.1 looks similar to a Fenchel Legendre transform, the usual candidate for a rate function. We will sometimes refer to this upper bound as an almost Fenchel-Legendre transform.

Interpretation 6.1.2. *The $\alpha \in \mathcal{B}_\Lambda$ are twist parameter in the change of measure process $\mathcal{M}^\Lambda(\alpha, \cdot)$ and by definition 4.4.17 the change of measure for the service process changes the rate of the counting process from μ_i to*

$$\mu_i(\alpha) = \Gamma_S^{(i)\prime}(-\alpha_i + \Xi^{(i)}(\alpha)).$$

Thus we can interprete $\mathcal{B}_\Lambda$ as the twist parameters that do not allow a decrease of service rates at the Λ^c-nodes.

Corollary 6.1.3. *It should be immediate that we can similarly bound the event (6.1) with $z \neq x$ if we have $z \to x$ before $\delta \to 0$:*

$$\begin{aligned}\lim_{\delta \to 0} \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \sup_{z : |z - x| < \epsilon} \mathbb{P}(Z_n(\cdot, z) \in \mathcal{U}_\delta(t \mapsto x + t\,v)) \\ \leq -T \sup_{\alpha \in \mathcal{B}_\Lambda} \big(\Psi(\alpha) - \langle \alpha, v \rangle\big)\end{aligned}$$

The situation $x \neq z$ affects the proof in (6.2) as we get $\frac{\lfloor nx \rfloor}{n} - z = \frac{\lfloor nx \rfloor}{n} - x + x - z$ instead of $\frac{\lfloor nx \rfloor}{n} - x$. It is again just a matter of changing δ to some δ'. The order of limits as $\epsilon \to 0$ before $\delta \to 0$ is important.

6.1.1 Leaving a boundary

In this section we investigate the event that a network process starting in $Z_n(0) = \frac{\lfloor nx \rfloor}{n}$ stays close to some affine function $t \mapsto x + vt$ and we allow $v_i > 0$ for $i \notin \Lambda(x)$. That is: an initially empty node $x_i = 0$ increases over $[0, T]$ and becomes non-empty. Figure 6.2 is an example of this situation.

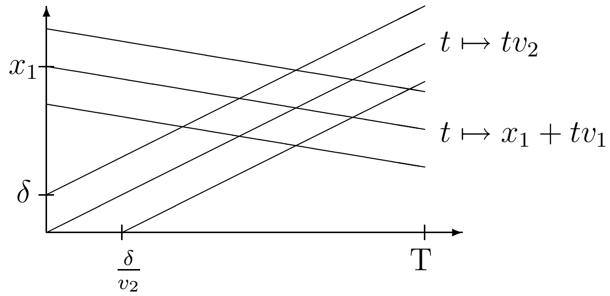


Figure 6.2: $d = 2$, $x_2 = 0$, $v_2 > 0$

Claim 6.1.4. *Consider the generalised Jackson network with d nodes and primitives 5.1.5. Let $\Gamma_A^{(i)}, \Gamma_S^{(i)}, \Xi^{(i)}$ be the the lmgfs for the primitives and Ψ the lmgf for the free process. Let assumption 6.0.5 hold for x, v, T. Then*

$$\lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}(Z_n(\cdot, x) \in \mathcal{U}_\delta(t \mapsto x + t\,v)) \;\le\; -T \sup_{\alpha \in \mathcal{B}_K} \langle \alpha, v \rangle - \Psi(\alpha).$$

for

$$\begin{aligned} K &= \{i \,|\, x_i > 0 \text{ or } v_i > 0\} \\ \mathcal{B}_K &= \{\alpha \in \mathbb{R}^d \,|\, -\alpha_i + \Xi^{(i)}(\alpha) \ge 0 \,\forall i \in K^c\}. \end{aligned}$$

The difference of claims 6.1.1 and 6.1.4 is in $v_{\Lambda^c} = 0$ vs $v_{\Lambda^c} \ge 0$ and the optimisation on the right hand sides over $\mathcal{B}_\Lambda$ vs $\mathcal{B}_K$. The difference in the result stems only from the additional assumption in 6.1.1.

Lemma 6.1.5. *For i with $x_i = 0$, $v_i > 0$*

$$Z_n(\cdot, z) \in \mathcal{U}_\delta(t \mapsto x + t\,v) \quad \Rightarrow \quad R_n^{(i)}(T) \;\ge\; T - \frac{\delta}{v_i}$$

Proof of 6.1.5: The claim may be obvious from figure 6.2. Nevertheless, we apply the definition of the runtime and bound. In this proof we abbreviate $Z_n(\cdot, z) \in \mathcal{U}_\delta(t \mapsto x + t\,v)$ as $Z_n \in \mathcal{U}$.

$$\begin{aligned} 1\!\!1_{Z_n \in \mathcal{U}}\, R^{(i)}(nT) &= 1\!\!1_{Z_n \in \mathcal{U}} \int_{t=0}^{nT} 1\!\!1_{Z_t^{(i)} > 0}\, dt \\ &\ge 1\!\!1_{Z_n \in \mathcal{U}} \int_{t=0}^{nT} 1\!\!1_{x_i + tv_i - n\delta > 0}\, dt \\ &\overset{x_i = 0}{=} 1\!\!1_{Z_n \in \mathcal{U}} \int_{t=0}^{nT} 1\!\!1_{t > \frac{n\delta}{v_i}}\, dt \;=\; 1\!\!1_{Z_n \in \mathcal{U}}\, (nT - \frac{n\delta}{v_i}) \end{aligned}$$

$\square$
6.1.5

Proof of 6.1.4: In this proof we abbreviate $W_n(\cdot,0) \in \mathcal{U}_\delta(t \mapsto t\,v)$ as $W_n^\Lambda \in \mathcal{U}$. The proof goes unchanged up to (6.4) where we bound differently the summands for $i \in K \cap \Lambda^c$:

$$\begin{aligned}
&\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \frac{R^{(i)}(nT)}{nT} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
= \;& \mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)>0} \underbrace{\frac{R^{(i)}(nT)}{nT}}_{\leq 1} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
& \quad + \mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)\leq 0} \underbrace{\mathbb{1}_{W^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \frac{R^{(i)}(nT)}{nT}}_{\geq 1-\frac{\delta}{Tv_i}} \underbrace{\Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))}_{\leq 0} \\
\leq \;& \mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \big(\mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)>0} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
& \quad + \mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)\leq 0} (1 - \frac{\delta}{Tv_i}) \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\big) \\
= \;& \mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \, (1 - \frac{\delta}{Tv_i} \mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)\leq 0})
\end{aligned}$$

and then analog to (6.4)

$$\begin{aligned}
&\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \frac{1}{\mathcal{M}^\Lambda(\alpha, nT)} \\
\leq \;& \mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \exp\{-\langle \alpha, W_{nT}^\Lambda \rangle + nT\big(\sum_{i=1}^d \Gamma_A^{(i)}(\alpha_i) \\
& + \sum_{i\in\Lambda} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
& + \sum_{i\in K\cap\Lambda^c} \underbrace{\frac{R^{(i)}(nT)}{nT}}_{\leq 1} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
& + \sum_{i\in K^c} \underbrace{\frac{R^{(i)}(nT)}{nT}}_{\leq 1 - \frac{\delta}{Tv_i}\mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)}} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\big)\} \frac{1}{r(\alpha, R(nT))}
\end{aligned}$$

The upper bound then becomes

$$\begin{aligned}\mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)) \quad \leq \quad & \mathbb{E}^{[\alpha]}[\mathbb{1}_{W^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \frac{1}{r}] \\ & \exp\{||\alpha|| nT\delta - nT\langle \alpha, v\rangle + nT\Big(\sum_{i=1}^d \Gamma_A^{(i)}(\alpha_i) \\ & \quad + \sum_{i\in\Lambda^c\cap K} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\,(1 - \frac{\delta}{Tv_i}\mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)\leq 0}) \\ & \quad + \sum_{i\in\Lambda} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\ & \quad + \sum_{i\in K^c} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\Big)\}\end{aligned} \tag{6.7}$$

Where we need the restriction $-\alpha_i+\Xi^{(i)}(\alpha) \geq 0$ only for $i \in K^c$ for bounding the relative runtime in (6.7) by 1. These restrictions define $\mathcal{B}_K$ analogue to (6.5). We continue

$$\begin{aligned}\mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)) \quad \leq \quad & \mathbb{E}^{[\alpha]}[\underbrace{\mathbb{1}_{W^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)}}_{\leq 1} \underbrace{\frac{1}{r}}_{\text{bounded by } 3.6.11}] \\ & \exp\{||\alpha|| nT\delta - nT\langle \alpha, v\rangle + nT\Psi(\alpha) \\ & - \sum_{i\in\Lambda^c\cap K} \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \underbrace{\frac{\delta}{Tv_i}}_{\to 0 \text{ as } \delta\to 0} \mathbb{1}_{-\alpha_i+\Xi^{(i)}(\alpha)\leq 0}\}\end{aligned}$$

and under the scaling limit

$$\lim_{\delta\to 0}\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)) \quad \leq \quad -T\langle \alpha, v\rangle + T\Psi(\alpha)$$

$\square$
6.1.4

6.2 Existence and uniqueness of an optimiser

We investigate the optimisation problem found in claim 6.1.4

$$\sup_{\alpha\in\mathcal{B}_K} \langle \alpha, v\rangle - \Psi(\alpha) \tag{6.8}$$

$$\begin{aligned}\text{with} \quad & K \supseteq \{i \,|\, v_i > 0\} \\ & \mathcal{B}_K = \{\alpha \in \mathbb{R}^d \,|\, -\alpha_i + \Xi^{(i)}(\alpha) \geq 0 \quad \forall i \in K^c\}\end{aligned}$$

and we will argue for the existence of a unique optimiser in the following. Uniqueness is not really an issue since we have seen that Ψ is strictly convex (if all inter event times are non-deterministic or if at least there is nondeterministic flow reaching each node).

We start with a simple condition for existence of an optimiser and then develop a second one, more elaborate and less restrictive.

Claim 6.2.1. *If the Fenchel Legendre transform Ψ^* is finite on all of $\mathbb{R}^d$ then an optimising $\overline{\alpha}$ in (6.8) exists.*

Proof of 6.2.1: We prove that For any v, K such that $v_{K^c} = 0$ the level sets $\{\alpha \in \mathcal{B}_K \mid \Psi(\alpha) - \langle \alpha\,,\, v\rangle \leq c\}$ are compact. Similar to [12] we construct a finite norm-ball including the level set.

Let $||\cdot||$ denote some norm in $\mathbb{R}^d$. and define the norm $|\cdot|_1$ as

$$|\alpha|_1 = \sup_{||v'||\leq 1} \langle \alpha, v'\rangle$$

We give a finite bound for $|\alpha|_1$ uniform in α of the level set.

$$\sup_{\substack{\alpha\in\mathcal{B}_K:\\ \Psi(\alpha)-\langle v,\alpha\rangle\leq c}} |\alpha|_1 = \sup_{\substack{\alpha\in\mathcal{B}_K:\\ \Psi(\alpha)-\langle v,\alpha\rangle\leq c}} \sup_{v'\,:\,||v'||\leq 1} \langle \alpha, v'\rangle$$

For α from the level set $\{\alpha \in \mathcal{B}_K \mid \Psi(\alpha) - \langle \alpha\,,\, v\rangle \leq c\}$

$$\begin{aligned} \langle \alpha, v'\rangle &= \langle \alpha, v'+v\rangle - \langle \alpha, v\rangle \\ &= \langle \alpha, v'+v\rangle - \Psi(\alpha) + \underbrace{\Psi(\alpha) - \langle \alpha, v\rangle}_{\leq c} \end{aligned}$$

and thus

$$\begin{aligned} |\alpha|_1 &\leq c + \sup_{v':||v'||\leq 1} \langle \alpha\,,\, v'+v\rangle - \Psi(\alpha) \\ &\leq c + \sup_{v':||v'||\leq 1} \sup_{\alpha\in\mathbb{R}^d} \langle \alpha\,,\, v'+v\rangle - \Psi(\alpha) \\ &\leq c + \sup_{v':||v'||\leq 1} \Psi^*(v'+v) \end{aligned} \tag{6.9}$$

which is finite if Ψ^* is finite on $\mathbb{R}^d$ and the supremum is over a convex set. The bound uniform in α: The level set is bounded. Closedness is immediate and only needs continuity of Ψ. From compactness of level sets and finiteness and continuity of the objective $\alpha \mapsto \Psi(\alpha) - \langle \alpha, v\rangle$ follows existence of

an infimiser. $\square$
6.2.1

We formulate our second criterion below in 6.2.7 and we believe that its if-part holds if the rate functions $\Gamma_D^{(i)*}$ for all $i = 1, \dots, d$ and $D \in \{A, S\}$ are open, cf 2.8. Technically we need to replace Ψ^* in the bound (6.9) in the case that that (6.8) is finite but there is no neighbourhood of v on which Ψ^* is finite. We will define the replacement G_K in 6.2.4.

We start with basic convex analysis and then give an upper bound for (6.8). Also we will argue for finiteness of this upper bound.

Consider the generalised Jackson network with primitives $A^{(i)}, S^{(i)}, n \mapsto \sum_{k=1}^{n} r_k^{(i)}$ for $i = 1, \dots, d$ (cf 5.1.5) where the arrival and service processes have lmgfs $\Gamma_A^{(i)}, \Gamma_S^{(i)}$ and $r_k^{(i)}$ has lmgf $\Xi^{(i)}$, cf 4.4.8. The $\Gamma_D^{(i)}$ are strictly convex as soon as they are not deterministic, the $\Xi^{(i)}$ are strictly convex if the routing measure they are build from are no point measures.

Let $\gamma \in \mathbb{R}_{\geq 0}^d$ and $w \in \mathbb{R}^d$ be fixed for the moment and denote by π_j the projection from $\mathbb{R}^d$ onto $\operatorname{span}\{e_j\}$, that is $\pi_j(\alpha) = \alpha_j \, e_j$. Then $\alpha \mapsto \sum_{i=1}^{d} \Gamma_A^{(i)} \circ \pi_i(\alpha) + \gamma_i \, \Xi^{(i)}(\alpha)$ is a convex function and we investigate its Fenchel-Legendre transform.

Definition 6.2.2. *For $\gamma \in \mathbb{R}_{\geq 0}^d$ define*

$$g_\gamma : \quad \mathbb{R}^d \to \mathbb{R} \cup \{\infty\} \quad , \quad \theta \mapsto \Big(\sum_{i=1}^{d} \Gamma_A^{(i)} \circ \pi_i + \gamma_i \, \Xi^{(i)} \Big)(\theta).$$

Claim 6.2.3. *g_γ is convex and*

$$g_\gamma^*(w) \quad = \quad \inf_{\substack{a \in \mathbb{R}^r, r \in \mathbb{R}^{d \times d} \\ a + r^\top \gamma = w}} \sum_{i=1}^{d} \Gamma_A^{(i)*}(a_i) + \gamma_i \, \Xi^{(i)*}(r^{(i)})$$

We interprete $g_\gamma(w)$ as the joint decay rate for the probability that at node i an empirical arrival rate a_i can be observed instead of the expected λ_i and that routing happens at empirical rates $r^{(i)}$ instead of $p^{(i)}$ over time γ_i. Additionally there is the condition that $a_i, r^{(i)}$ have to be such as to produce total flow into each node i of rate $w_i = a_i + (r^\top \gamma)_i$ in the associated fluid network. If there are no a, r in the domains of $\Gamma_A^{(i)*}, \Xi^{(i)*}$ such a flow of a into the fluid network and a splitting of flow at each node i wrt $r^{(i)}$ would produce input flow w into the nodes then $g_\gamma^*(w) = \infty$.

Proof of 6.2.3: We start with some convex analysis (cf (7.2) of the appendix).

$$\Big(\sum_{i=1}^{d} \Gamma_A^{(i)} \circ \pi_i + \gamma_i\, \Xi^{(i)}\Big)^*(w) \quad = \quad \inf_{\substack{c_j, d_j \in \mathbb{R}^d \\ \sum_j c_j + d_j = w}} \sum_{j=1}^{d} (\Gamma_A^{(j)} \circ \pi_j)^*(c_j) + (\gamma_j\, \Xi^{(j)})^*(d_j)$$

Due to 7.3.1 we do not increase the infimum when restricting c_j to $\pi_j(\mathbb{R}^d)$. In the following we optimise over $a \in \mathbb{R}^d$ with $a_j = \langle c_j, e_j \rangle$. This is OK since $(c_1, \ldots, c_d) \mapsto a$ is a bijection for those c_j for which $(\Gamma_A^{(j)} \circ \pi_j)^*(c_j) < \infty$. We also apply (7.1)

$$g_\gamma^*(w) = \inf_{\substack{a, d_j \in \mathbb{R}^d \\ a + \sum d_j = w}} \sum_{j=1}^{d} \Gamma_A^{(j)*}(a_j) + \gamma_j\, \Xi^{(j)*}(\frac{1}{\gamma_j}\, d_j) \tag{6.10}$$

Changing variables from $\frac{1}{\gamma_j}\, d_j$ to $r^{(j)}$ in the argument of $\Xi^{(j)*}$ we need to change the restriction, too.

$$r^{(j)} = \frac{1}{\gamma_j} d_j \quad \Rightarrow \quad \sum_{j=1}^{d} d_j = \sum_{j=1}^{d} \gamma_j\, r^{(j)} = r^\top \gamma$$

This completes the proof. $\square$ 6.2.3

From the new representation of g_γ^* in 6.2.3 we see that g_γ^* is finite if there are $(a, r) \in \mathbb{R}^d \times \mathbb{R}^{d \times d}$ with $a_i \geq 0$, $\lambda_i = 0 \Rightarrow a_i = 0$, and $r^{(i)}$ equivalent to $p^{(i)}$. It is also possible that there is $r^{(i)}$ not equivalent to $p^{(i)}$ but with $\mathrm{supp}(r^{(i)}) \subsetneq \mathrm{supp}(p^{(i)})$. In that case there'd be no finite optimiser in $\Xi^{(i)*}(r^{(i)})$ (cf 4.4.22).

Note that the Fenchel-Legendre transform in claim 6.2.3 satisfies

- $g_\gamma^* \geq 0$ from $\Gamma_A^{(i)*}$, $\Xi^{(i)*}$, $\gamma_i \geq 0$ and thus $g_\gamma^* > -\infty$ always,
- $g_\gamma^*(\lambda + P^\top \mu) < \infty$ since $(\lambda, P) \in \mathcal{X}(\lambda + P^\top \mu, \mu)$. This is also a minimiser of the Fenchel-Legendre transform: $g_\gamma^*(\lambda + P^\top \mu) = 0$.

Thus g_γ^* is a proper convex function (cf [19] p. 24, definition of "proper").

As a next prep-step

Definition 6.2.4. *Let $K \subseteq \{1, \dots, d\}$.*

$$\begin{aligned} G_K : \mathbb{R}^d &\to \mathbb{R} \cup \{\infty\} \\ v &\mapsto \inf_{\substack{a,r,\gamma: \\ a+(r^\top - id)\gamma = v}} \sum_{i=1}^{d} \Gamma_A^{(i)*}(\alpha_i) + \gamma_i \, \Xi^{(i)*}(r^{(i)}) \\ &\qquad + \sum_{j \in K} \Gamma_S^{(j)*}(\gamma_i) + \sum_{j \in K^c} 1\!\!1_{\gamma_j > \mu^{(j)}} \Gamma_S^{(j)*}(\gamma_j) \end{aligned}$$

$G_K(v) < \infty$ if there is one set of $\{a, r, \gamma\}$ such that the respective rate functions are finite and the fluid network with

- non empty nodes K
- arrival rate a_i at node i
- service rate γ_i at each node $i \in K$
- service rate $\max\{\gamma_i, \mu_i\}$ at each node $i \in K^c$
- routing matrix r

has network drift v (cf definition 5.2.16). The difference $\mu_i - \gamma_i$ would be the loss rate usually denoted y_i at the initially empty subnetwork K^c.

Claim 6.2.5. *If the set of restrictions $\{a, r, \gamma \, : \, a + (r^\top - id)\gamma = v\}$ has a non-empty inter section with the domains of the respective individual rate functions $\Gamma_A^{(i)*}, \Gamma_S^{(i)*}, \Xi^{(i)*}$ then $G_K(v) < \infty$ and the infimum is a minimum (optimiser exists).*

Proof of 6.2.5: We argue with compactness of level sets.

$$\sum_{j=1}^{d} \Gamma_A^{(j)*}(a_j) + \gamma_j \, \Xi^{(j)*}(r^{(j)}) + \sum_{i \in \Lambda^c} 1\!\!1_{\gamma_i > \mu^{(i)}} \Gamma_S^{(i)*}(\gamma_i) + \sum_{i \in \Lambda} \Gamma_S^{(i)*}(\gamma_i) \le M$$

$$\Rightarrow \left\{ \begin{array}{cc} \Gamma_A^{(j)*}(a_j) \le M & , j = 1, \dots d \\ \Gamma_S^{(i)*}(\gamma_i) \le M & , i \in \Lambda \\ \left\{ \begin{array}{c} \gamma_i \in [0, \mu^{(i)}] \\ \text{or} \\ \Gamma_S^{(i)*}(\gamma_i) \le M \end{array} \right\} & , i \in \Lambda^c \end{array} \right\}$$

From goodness of the Γ^*-rate functions the a, γ are in compact sets of $\mathbb{R}^d$, forming a bounded set themselves. Also the $r^{(i)}$ are sub-probability measures and their max-norm is ≤ 1 so each is in a bounded set of $\mathbb{R}^d$. From continuity of all involved rate functions the level set is closed. Thus all parameters are in compact sets forming a compact set in product space. Therefore G_K that was defined as an infimum is actually a minimum whenever it is finite.

$\square$
6.2.5

Note that G_K can be written as a composition of g^*_γ that is constant in K and the rate functions for the service processes.

$$G_K(v) = \inf_\gamma g^*_\gamma(v+\gamma) + \sum_{i\in K^c} 1\!\!1_{\gamma_i>\mu^{(i)}} \Gamma_S^{(i)*}(\gamma_i) + \sum_{i\in K} \Gamma_S^{(i)*}(\gamma_i).$$

Further note that $G_{\{1,\ldots,d\}} = \Psi^*$ so $G_{\{1,\ldots,d\}}$ is a tight upper bound for the Fenchel-Legendre transform Ψ^* and this generalises as:

Claim 6.2.6. *G_K bounds the almost Fenchel-Legendre transform (6.8).*

Proof of 6.2.6: We start with some transformations that will allow us to apply 6.2.3.

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{B}_K} \langle \alpha\,,\,v\rangle - \Psi(\alpha)\\
&\quad = \sup_{\alpha\in\mathcal{B}_K} \langle \alpha\,,\,v\rangle - \sum_{i=1}^d \Gamma_A^{(i)}(\alpha_i) + \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))
\end{aligned}$$

we add a zero (introducing the non-negative parameters $\gamma_1,\ldots,\gamma_d$) and rearrange.

$$\begin{aligned}
= \quad & \sup_{\alpha\in\mathcal{B}_K} \langle \alpha\,,\,v\rangle + \sum_{i=1}^d -\gamma_i\,(-\alpha_i + \Xi^{(i)}(\alpha)) - \Gamma_A^{(i)}(\alpha_i)\\
& \qquad\qquad + \gamma_i\,(-\alpha_i + \Xi^{(i)}(\alpha)) - \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))\\
= \quad & \sup_{\alpha\in\mathcal{B}_K} \langle \alpha\,,\,v+\gamma\rangle - \sum_{i=1}^d \gamma_i\,\Xi^{(i)}(\alpha) + \Gamma_A^{(i)}(\alpha_i)\\
& \qquad\qquad + \gamma_i\,(-\alpha_i + \Xi^{(i)}(\alpha)) - \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha))
\end{aligned}$$

and take suprema separately. We loose the restriction in the first supremum.

The expression may increase:

$$
\begin{aligned}
\leq\ & \sup_{\alpha\in\mathbb{R}^d} \langle \alpha\,,\, v+\gamma\rangle - \sum_{i=1}^{d} \gamma_i\, \Xi^{(i)}(\alpha) + \Gamma_A^{(i)}(\alpha_i) \\
& \qquad + \sum_{i=1}^{d} \sup_{\alpha\in\mathcal{B}_K} \gamma_i\left(-\alpha_i + \Xi^{(i)}(\alpha)\right) - \Gamma_S^{(i)}(-\alpha_i + \Xi^{(i)}(\alpha)) \\
\leq\ & \Big(\sum_{i=1}^{d} \Gamma_A^{(i)} \circ \pi_i + \gamma_i\, \Xi^{(i)}\Big)^*(v+\gamma) \\
& \qquad + \sum_{i\in K^c} 1\!\!1_{\gamma_i > \mu^{(i)}} \Gamma_S^{(i)*}(\gamma_i) + \sum_{i\in K} \Gamma_S^{(i)*}(\gamma_i)
\end{aligned}
\tag{6.11}
$$

Optimise over γ and the claim follows: we got

$$
\sup_{\alpha\in\mathcal{B}_K} \langle \alpha, v\rangle - \Psi(\alpha) \leq \inf_{\gamma\in\mathbb{R}^d} (6.11) = G_K(v)
$$

□
6.2.6

In the rest of the subsection we apply G_K to prove existence of an optimiser in the almost Fenchel-Legendre transform (6.8).

Claim 6.2.7. *If $v \in \mathcal{D}^\circ(G_K)$ then for any v, K such that $v_{K^c} = 0$ the level sets $\{\alpha \in \mathcal{B}_K \,|\, \Psi(\alpha) - \langle \alpha\,,\, v\rangle \leq c\}$ are compact and an optimiser in (6.8) exists.*

Proof of 6.2.7: Let $a > 0$ be small enough for G_K to be finite in an $||\cdot||$-ball of radius a around v and let $|\alpha|_a = \sup_{v':||v'||\leq a} \langle \alpha, v\rangle$. Then as in the proof of 6.2.1 we obtain

$$
\begin{aligned}
|\alpha|_a\ &\leq\ c + \sup_{v':||v'||\leq a} \langle \alpha\,,\, v'+v\rangle - \Psi(\alpha) \\
&\leq\ c + \sup_{v':||v'||\leq a} \sup_{\alpha\in\mathcal{B}_\Lambda} \langle \alpha\,,\, v'+v\rangle - \Psi(\alpha) \\
&\leq\ c + \sup_{v':||v'||\leq a} G_\Lambda(v'+v)
\end{aligned}
$$

which is a finite bound by choice of a and uniform in α: The level set is bounded. Closedness is immediate from continuity of Ψ. □
6.2.7

6.3 Network drift under the changed measure

From the local large deviations upper bound in 6.1.1 and more general 6.1.4 we got the candidate for the local rate function $L(x,v)$ as the following optimisation problem

$$\sup_{\alpha \in \mathcal{B}_K} \langle \alpha, v \rangle - \Psi(\alpha). \tag{6.12}$$

In section 6.2 we have given conditions under which an optimiser $\overline{\alpha}$ exists. We will now investigate the behaviour of the network after the change of measure $\mathcal{M}^{\emptyset}(\overline{\alpha}, \cdot)$.

For any α the change of measure can be translated back into the individual changes of measure for each network primitive, cf 5.3.31. Thus once we have identified an $\overline{\alpha}$ we are not restricted to work with the local process W^{Λ} we started with, we just switch from $\mathbb{P}$ to $\mathbb{P}^{[\overline{\alpha}]}$ and work with the network primitives and the free, the network, and the local process as before.

Claim 6.3.1. *Under assumption 6.0.5 for x, v, T and under $\mathbb{P}^{[\overline{\alpha}]}$ for $\overline{\alpha}$ the optimiser in 6.1.4 the fluid limit of $Z_n(\cdot, z_0)$ is $t \mapsto z_0 + tv$ and*

$$\lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(Z_n \in \mathcal{U}_\delta(t \mapsto z_0 + tv)) = 1.$$

The proof of 6.3.1 requires elements of optimisation theory we state before we begin the proof. For this section we rephrase the almost Fenchel-Legendre transform (6.12) as an inequality constrained minimisation problem (ICM).

$$-\langle \alpha\,,\, v\rangle + \Psi(\alpha) \to \min \quad \text{subj. to } \alpha_i - \Xi^{(i)}(\alpha) \le 0 \quad \forall i \in K^c \tag{ICM}$$

For future reference set $g_i = \pi_i - \Xi^{(i)}$.

If for the ICM the Slater condition holds the optimiser satisfies the Karush-Kuhn-Tucker (KKT) condition. The KKT condition will help us to identify v as the network drift under $\mathbb{P}^{[\overline{\alpha}]}$.

Claim 6.3.2 (Slater condition)**.** *There is $\alpha \in \mathbb{R}^d$ such that $g_i(\alpha) < 0$ for each $i \in \Lambda^c$.*

Proof of 6.3.2: We assume that $\Lambda^c \neq \emptyset$. We step by step fix the value of each α_i such that the condition $g_i(\alpha) < 0$ always (and finally) holds for all

$i \in \Lambda^c$.
Fix values for α_Λ and let $B := \Lambda$ be the set of indices with α_i already fixed. The proof stops as $B = \{1, \dots, d\}$ and $g_i(\alpha) < 0$ has been checked for all $i \in \Lambda^c$.

We define a partition of the Λ^c-nodes relative to the length of the shortest path from a node $i \in \Lambda^c$ to $\Lambda \cup \{0\}$. From assumption 5.1.9 there is a finite length path from i to $\{0\}$ making the shortest path from i to $\Lambda \cup \{0\}$ have finite length. For fixed Λ set

$$A_0 \;=\; \Lambda \cup \{0\}$$

and for $k = 1, 2, \dots$ and while A_{k-1} is not empty set

$$A_k \;=\; \{i \in \{1, \dots, d\} \setminus (A_0 \cup \dots \cup A_{k-1}) \,|\, \exists j \in A_{k-1} \,:\, p_{ij} > 0\}$$

Then A_k is the subset of Λ^c nodes with the shortest path to $\Lambda \cup \{0\}$ consisting of k edges. Note that there is at most $d + 1$ such sets.

Set α_i for $i \in \Lambda$ to an arbitrary value in $\mathbb{R}$. In the following we fix the values of α_i , $i \in A_1$: We omit the α's not to be fixed now and get an inequality.

$$\begin{aligned}
\Xi^{(i)}(\alpha) \;&\geq\; \log\Big(\sum_{j \in A_1} p_{ij}\, e^{\alpha_j} + \underbrace{\sum_{j \in B} p_{ij}\, e^{\alpha_j} + p_{i0}}_{=:p'_{i0}}\Big) \\
\Rightarrow \qquad g_i(\alpha) = \alpha_i - \Xi^{(i)}(\alpha) \;&\leq\; \alpha_i - \log\Big(\sum_{j \in A_1} p_{ij}\, e^{\alpha_j} + p'_{i0}\Big)
\end{aligned}$$

We now have an upper bound for $g_i(\alpha)$ and we choose α_{A_1} to make the upper bound negative.

$$\begin{aligned}
0 \;&>\; \alpha_i - \log\Big(\sum_{j \in A_1} p_{ij}\, e^{\alpha_j} + p'_{i0}\Big) \\
\Leftrightarrow \quad e^{\alpha_i} \;&<\; \sum_{j \in A_1} p_{ij}\, e^{\alpha_j} + p'_{i0} \\
\Leftrightarrow \quad \hat{\alpha}_i \;&<\; \sum_{j \in A_1} p_{ij}\, \hat{\alpha}_j + p'_{i0} \qquad (6.13)
\end{aligned}$$

where we set $\hat{\alpha}_i := e^{\alpha_i}$ and will have to observe the condition $\hat{\alpha}_i > 0$ to get an $\alpha_i \in \mathbb{R}$. From construction of A_1 we have $p'_{i0} > 0$ - either due to $p_{i0} > 0$ or from $p_{ij} > 0$ for some $j \in B = \Lambda$ and the α_Λ fixed as some real numbers

- the rhs of (6.13) is positive. In the linear notation $p'_{i0} = \sum_{j \in B} p_{ij}\hat{\alpha}_j + p_{i0}$. We get a system of $|\Lambda^c|$ linear inequalities.

$$\begin{aligned} (6.13)\,\forall i \in \Lambda^c \;\Leftrightarrow\; \hat{\alpha}_{\Lambda^c} &< P_{\Lambda^c A_1}\hat{\alpha}_{A_1} + P_{\Lambda^c B}\hat{\alpha}_B + P_{\Lambda^c\{0\}} \\ \Rightarrow\; \hat{\alpha}_{A_1} &< P_{A_1}\hat{\alpha}_{A_1} + P_{A_1 B}\hat{\alpha}_B + P_{A_1\{0\}} \\ \Leftrightarrow\; (\mathrm{id} - P_{A_1})\,\hat{\alpha}_{A_1} &< P_{A_1 B}\hat{\alpha}_B + P_{A_1\{0\}} \end{aligned}$$

The inverse of id $-P_{A_1}$ exists and is strictly positive. Thus we can multiply with (id $-P_{A_1})^{-1}$ and keep the coordinate wise inequality.

$$\hat{\alpha}_{A_1} < (\mathrm{id} - P_{A_1})^{-1}\left(P_{A_1 B}\hat{\alpha}_B + P_{A_1\{0\}}\right)$$

which leaves a non-degenerate positive interval for each $\hat{\alpha}_i$, $i \in A_1$. We can fix α_{A_1} and thus update $B := B \cup A_1$ $(= \Lambda \cup A_1)$. If $|B| = d$ we are done. Else:

We can iterate this. For $k \geq 2$ the k-th iteration is to be done only if $|B| = |\Lambda \cup A_1 \cup \cdots \cup A_{k-1}| < d$ which is equivalent to $A_k \neq \emptyset$. Up to the $k-1$-st iteration α_i are known for all $i \in B$. As before

$$\begin{aligned} \Xi^{(i)}(\alpha) &\geq \log\Big(\sum_{j \in A_k} p_{ij}\, e^{\alpha_j} + \underbrace{\sum_{j \in B} p_{ij}\, e^{\alpha_j} + \underbrace{p_{i0}}_{=0}}_{=:p'_{i0}}\Big) \\ \Leftrightarrow \qquad \alpha_i - \Xi^{(i)}(\alpha) &\leq \alpha_i - \log\Big(\sum_{j \in A_k} p_{ij}\, e^{\alpha_j} + p'_{i0}\Big) \end{aligned}$$

and

$$g_i(\alpha) < 0 \qquad \Leftarrow \qquad \hat{\alpha}_i \;<\; \sum_{j \in A_k} p_{ij}\,\hat{\alpha}_j + p'_{i0}$$

with strictly positive p'_{i0} (from $B \cap \mathrm{supp}(p^{(i)}) \neq \emptyset$ by construction of A_k). Putting all $i \in A_k$ in one inequality

$$\hat{\alpha}_{A_k} \;<\; (\mathrm{id} - P_{A_k})^{-1}\, P_{A_k B}\hat{\alpha}_B$$

Positivity of the p'_{i0} grants solvability of the inequality and we can fix real coordinates for α_{A_k} and update $B := B \cup A_k$. If necessary iterate again. $\square$ 6.3.2

Claim 6.3.3 (KKT). *In $\overline{\alpha}$ the Karush-Kuhn-Tucker condition holds:*

- $\nabla(-\langle\overline{\alpha}\,,\,v\rangle + \Psi(\overline{\alpha})) + \sum_{i \in K^c} \eta_i\,\nabla g_i(\overline{\alpha}) = 0$

- $\eta \in \mathbb{R}_{\geq 0}^{|K^c|}$
- $\langle \eta, g(\overline{\alpha}) \rangle = 0$

with a unique $\eta = \eta(\overline{\alpha})$.

Proof of 6.3.3: The KKT condition holds since we proved that the Slater condition holds; existence of the optimiser $\overline{\alpha}$ was already proved. For uniqueness of η it remains to be shown that the gradients $\{\nabla g_i(\overline{\alpha}), \,|\, i \in K^c\}$ are linearly independent.

From our general assumption 5.1.9 one is not an eigenvalue of P and $\mathrm{id} - P^\top$ is a regular matrix. Then also the sub-matrix $\mathrm{id}_{K^c} - P_{K^c}^\top$ is regular. Since rows of P_{K^c} and $P_{K^c}(\overline{\alpha})$ are both sub-probability measures (cf 4.4.16, 4.4.23) also $\mathrm{id}_{K^c} - P_{K^c}^\top(\overline{\alpha})$ is a regular matrix. And from regularity of $\mathrm{id} - P_{K^c}^\top(\overline{\alpha})$ follows linear independence of their columns $e_i - p_{K^c}(\overline{\alpha})^{(i)} \in \mathbb{R}^{|K^c|}$:

$$\{e_i - p_{K^c}^{(i)}(\overline{\alpha}) \,|\, i \in K^c\} \subseteq \mathbb{R}^{|K^c|}$$

and thus of the longer columns

$$\{\nabla g_i(\overline{\alpha}) \,|\, i \in K^c\} = \{e_i - p^{(i)}(\overline{\alpha}) \,|\, i \in K^c\} \subseteq \mathbb{R}^d.$$

$\square$
6.3.3

We can now prove the statement about the fluid limit under the change of measure.

Proof of 6.3.1: Restate the first bullet from KKT:

$$\sum_{i \in K^c} \eta_i \nabla g_i(\overline{\alpha}) \;=\; \sum_{i \in K^c} (e_i - \nabla \Xi^{(i)}(\overline{\alpha}))\, \eta_i$$

each $\nabla \Xi^{(i)}$ is an exponential twist of the row $p^{(i)}$ of P (cf 4.4.16) and we have under the change of measure with $\overline{\alpha}$ from 5.3.16)

$$P^\top(\overline{\alpha}) = [\, \nabla \Xi^{(1)}(\overline{\alpha}), \ldots, \nabla \Xi^{(d)}(\overline{\alpha}) \,]$$

$$\sum_{i \in K^c} \eta_i \nabla g_i(\overline{\alpha}) \;=\; (\mathrm{id}_{\{1,\ldots,d\},K^c} - P^\top(\overline{\alpha})_{\{1,\ldots,d\}K^c}\eta \;=\; (\mathrm{id} - P^\top(\overline{\alpha})) \begin{bmatrix} 0 \\ \eta \end{bmatrix}$$

and the KKT first bullet becomes

$$\begin{aligned} v &= \nabla \Psi(\overline{\alpha}) + \sum_{i \in K^c} \eta_i (e_i - \nabla \Xi^{(i)}(\overline{\alpha})) \\ &= \lambda(\overline{\alpha}) + (P^\top(\overline{\alpha}) - \mathrm{id})\, \mu(\overline{\alpha}) + (\mathrm{id} - P^\top(\overline{\alpha})) \begin{bmatrix} 0 \\ \eta \end{bmatrix} \\ &= \lambda(\overline{\alpha}) + (P^\top(\overline{\alpha}) - \mathrm{id}) \left(\mu(\overline{\alpha}) - \begin{bmatrix} 0 \\ \eta \end{bmatrix} \right) \end{aligned}$$

and on the right hand side we have the network drift defined in 5.2.16 for the network process starting in some z_0 with $z_{0,i} > 0$ for $i \in K$ and $z_{0,i} = 0$ for $i \in K^c$. The $\begin{bmatrix} 0 \\ \eta \end{bmatrix}$ is the loss rate. And by 5.3.23 the network drift is the expected, normal behaviour of the network process, its fluid limit. $\square$ 6.3.1

Interpretation 6.3.4. *We have interpreted $\mathcal{B}_K$ in 6.1.2 as twist parameters that do not decrease service rates at K^c-nodes. From $g_i(\overline{\alpha}) < 0 \Rightarrow \eta_i = 0$ (complementarity in the KKT condition) and the identification of η as the loss rate in 6.3.1 we know that if the service rate of a node is strictly increased under the twist $\overline{\alpha}$ then this node is a bottleneck.*

We have written the local process W^Λ as the sum of a free subprocess X^Λ and the network process of nodes Λ^c denoted Z^Λ.

Corollary 6.3.5. *From 6.3.1 and for $K = \Lambda$: Under $\mathbb{P}^{[\overline{\alpha}]}$ there are no strict bottlenecks in the Λ^c-nodes. Ergodic nodes in the Λ^c-subnetwork are identified through the network drift (v, η) obtained from the KKT (via $\eta_i > 0$).*

Proof of 6.3.5:

$$\begin{aligned} 1 &= \lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(Z_n \in \mathcal{U}_\delta(t \mapsto z_0 + tv)) \\ &= \lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto z_0 + tv)) \\ &= \lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(X_n^\Lambda \in \mathcal{U}_\delta(t \mapsto z_0 + t\pi_\Lambda(v)) \,,\, Z_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\pi_{\Lambda^c}(v))) \end{aligned}$$

From $\Lambda = K$ we have $\pi_{\Lambda^c}(v) = 0$ and

$$\begin{aligned} 1 &= \lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(X_n^\Lambda \in \mathcal{U}_\delta(t \mapsto z_0 + t\pi_\Lambda(v)) \,,\, Z_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\pi_{\Lambda^c}(v))) \\ &\leq \lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(Z_n^\Lambda \in \mathcal{U}_\delta(t \mapsto 0)) \;=\; \lim_{n\to\infty} \mathbb{P}^{[\overline{\alpha}]}(||Z_n^\Lambda|| < \delta) \end{aligned}$$

$\square$ 6.3.5

6.4 Local large deviations lower bound

In this section we give a lower bound for the exponential decay rate of the event that the local process W^Λ starting in $W_n^\Lambda(0, x) = \frac{\lfloor nx \rfloor}{n}$ follows the affine function $t \mapsto x + tv$. Let $\Lambda = \{i \,|\, x_i > 0\}$ and then uniformise over x and only investigate the following event:

$$\{ \sup_{t\in[0,nT]} |\, W_t^\Lambda - tv| < n\delta\} \;=\; \{W^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)\}.$$

Further let again $K = \{i \,|\, x_i > 0 \text{ or } v_i > 0\} \supseteq \Lambda$.

Claim 6.4.1. *If $\overline{\alpha}$ is the optimiser in the almost Fenchel-Legendre transform*

$$\sup_{\alpha \in \mathcal{B}_K} \langle \alpha, v \rangle - \Psi(\alpha) \;=\; \langle \overline{\alpha}, v \rangle - \Psi(\overline{\alpha})$$

then the lower local large deviation bound holds:

$$\lim_{\delta \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(||Z_n(\cdot, x) - (t \mapsto x + tv)|| < \delta) \geq -T \left(\langle \overline{\alpha}, v \rangle - \Psi(\overline{\alpha}) \right).$$

We will apply the change of measure with parameter $\overline{\alpha}$ that was found as the optimiser in the upper bound 6.1.4. We start the same way as for the upper bound.

$$\begin{aligned}
\mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)) &= \mathbb{E}[\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)}] \\
&= \mathbb{E}[\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto t\,v)} \frac{\mathcal{M}^\Lambda(\overline{\alpha}, nT)}{\mathcal{M}^\Lambda(\overline{\alpha}, nT)}] \\
&= \mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)} \frac{1}{\mathcal{M}^\Lambda(\overline{\alpha}, nT)}]
\end{aligned}$$

From the change of measure applied here that was defined in 5.3.30, 5.3.32:

$$\begin{aligned}
\frac{1}{\mathcal{M}^\Lambda(\overline{\alpha}, nT)} &= \exp\Big\{ - \langle \overline{\alpha}, W_{nT}^\Lambda \rangle + nT \Big(\sum_{i=1}^{d} \Gamma_A^{(i)}(\overline{\alpha}_i) \\
&\quad + \sum_{i \in \Lambda} \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha})) \\
&\quad + \sum_{i \in \Lambda^c} \frac{1}{nT} R^{(i)}(nT) \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha})) \Big) \Big\} \frac{1}{r(\overline{\alpha}, nT)}
\end{aligned}$$

We apply the change of measure to our event (uniformise over x already).

$$
\begin{aligned}
&\mathbb{P}(W_n^\Lambda \in \mathcal{U}_\delta(t \mapsto tv)) \\
&= \mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \frac{1}{\mathcal{M}^\Lambda(\overline{\alpha}, nT)}] \\
&= \mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{W_n^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \exp\Big\{ - \langle \overline{\alpha}\,,\, W_{nT}^\Lambda \rangle + nT \sum_{i=1}^d \Gamma_A^{(i)}(\overline{\alpha}_i) \\
&\qquad + nT \sum_{i\in\Lambda} \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha})) \\
&\qquad + nT \sum_{i\in\Lambda^c} \frac{R^{(i)}(nT)}{nT} \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha})) \Big\} \frac{1}{r(\overline{\alpha}, nT)}]
\end{aligned}
$$

$$
\geq \exp\{-||\overline{\alpha}||nT\delta - \langle \overline{\alpha}, nTv \rangle + nT\Psi(\overline{\alpha})\} \tag{6.14}
$$

$$
\mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{W^\Lambda \in \mathcal{U}_\delta(t\mapsto tv)} \exp\{-nT \sum_{i\in\Lambda^c} (1 - \frac{R^{(i)}(nT)}{nT}) \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\} \frac{1}{r(\overline{\alpha}, nT)}] \tag{6.15}
$$

Inequality in (6.14) is due only to the minus in $-||\overline{\alpha}||nT\delta$. It applies the definition of Ψ of 5.3.9 from the primitives lmgfs. The proof of 6.4.1 is thus equivalent to the proof of

Lemma 6.4.2.

$$
\lim_{\delta\to 0} \liminf_{n\to\infty} \frac{1}{n} \log(\mathit{6.15}) = 0
$$

Proof of 6.4.2: We have joint uniform convergence of the queue size and the runtime process.

$$
\begin{aligned}
&\mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+tv)} \exp\{nT \sum_{i\in\Lambda^c} (\frac{R^{(i)}(nT)}{nT} - 1) \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\}] \\
&\geq \mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+tv)}\, \mathbb{1}_{R_n\in\mathcal{U}_\delta(t\mapsto t\rho)} \\
&\qquad \exp\{nT \sum_{i\in\Lambda^c} (\frac{R^{(i)}(nT)}{nT} - 1) \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\}] \\
&= \mathbb{E}^{[\overline{\alpha}]}[\mathbb{1}_{Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+tv)}\, \mathbb{1}_{R_n\in\mathcal{U}_\delta(t\mapsto t\rho)} \\
&\qquad \exp\{nT \sum_{\substack{i\in\Lambda^c \\ -\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}) = 0}} (\frac{R^{(i)}(nT)}{nT} - 1) \underbrace{\Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))}_{=0}\} \\
&\qquad \exp\{nT \sum_{\substack{i\in\Lambda^c \\ -\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}) \neq 0}} (\frac{R^{(i)}(nT)}{nT} - 1) \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\}]
\end{aligned}
$$

We have started with $\overline{\alpha} \in \mathcal{B}_K$, thus for $i \in K^c$

$$-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}) \neq 0 \Rightarrow -\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}) > 0$$

For nodes $i \in K^c$ we have found in 6.3.4 that under $\overline{\alpha}$

$$-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}) > 0 \Rightarrow \rho_i = 1.$$

and from the indicator for the runtime process

$$\frac{R^{(i)}(nT)}{nT} - 1 \geq \rho_i - \delta - 1 = -\delta$$

$K \cap \Lambda^c$ was the set of nodes with $x_i = 0, v_i > 0$ and from the runtime bound in 6.1.5 we have

$$\frac{R^{(i)}(nT)}{nT} - 1 \geq 1 - \frac{\delta}{tv_i} - 1 = -\frac{\delta}{tv_i}$$

We finally have

$$\begin{aligned}
& 1\!\!1_{Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+tv)}\, 1\!\!1_{R_n\in\mathcal{U}_\delta(t\mapsto t\rho)} \\
& \qquad \exp\{nT \sum_{\substack{i\in\Lambda^c \\ -\overline{\alpha}_i+\Xi^{(i)}(\overline{\alpha})>0}} (\frac{R^{(i)}(nT)}{nT} - 1)\Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\} \\
\geq \; & 1\!\!1_{Z_n(\cdot,x)\in\mathcal{U}_\delta(t\mapsto x+tv)}\, 1\!\!1_{R_n\in\mathcal{U}_\delta(t\mapsto t\rho)} \\
& \qquad \exp\{nT \sum_{\substack{i\in\Lambda^c\cap K \\ -\overline{\alpha}_i+\Xi^{(i)}(\overline{\alpha})>0}} \underbrace{(\frac{R^{(i)}(nT)}{nT} - 1)}_{\geq -\frac{\delta}{Tv_i}} \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\} \\
& \qquad \exp\{nT \sum_{\substack{i\in K^c \\ -\overline{\alpha}_i+\Xi^{(i)}(\overline{\alpha})>0}} \underbrace{(\frac{R^{(i)}(nT)}{nT} - 1)}_{\geq -\delta} \Gamma_S^{(i)}(-\overline{\alpha}_i + \Xi^{(i)}(\overline{\alpha}))\}
\end{aligned}$$

So we only have deterministic exponential expressions left, and under the scaling in 6.4.2 they tend to 0. The indicators are such that their expectation tends to 1. □
6.4.2

We can also have $Z_n(\cdot, z)$ with $z \neq x$, as soon as $z \to x$ before $\delta \to 0$.

6.5 Rate function identification

For the generalised Jackson network Anatolii Puhalskii proved a sample path large deviation principle in [16]. The rate function is infinite on not absolute continuous functions, for absolutely continuous functions $\mathbf{q}$ it is defined as

$$\begin{aligned}
I^Q_{q_0}(\mathbf{q}) &= \int_{t=0}^{\infty} \sum_{J\subseteq\{1,\dots,d\}} 1\!\!1_{\mathbf{q}(t)\in F_J}\, \mathcal{R}_J(\dot{\mathbf{q}}(t))\, dt \quad (= \int_{t=0}^{\infty} L(\mathbf{q}(t)\,,\,\dot{\mathbf{q}}(t))\, dt) \\
F_J &= \{x \in \mathbb{R}^K_+ \mid x_k = 0, k \in J\,;\, x_k > 0, k \notin J\} \\
\mathcal{R}_J(v) &= \inf_{(a,d,r):v=a+(r^\top - I)d} \psi_J(a,d,r) \\
\psi_J(a,d,r) &= \psi^A(a) + \sum_{k\in J^c} \psi^S_k(d_k) + \sum_{k\in J} \psi^S_k(d_k)\, 1\!\!1_{d_k>\hat{\mu}_k} + \sum_{k=1}^{d} d_k \psi^R_k(r_k)
\end{aligned}$$

In our notation (definition 6.2.4)

$$\mathcal{R}_J(v) \;=\; G_{J^c}(v)$$

Claim 6.5.1 (Rate function identification)**.** *$\mathcal{R}_J(v) = L(x,v)$ for $J^c = \Lambda(x)$.*

Proof of 6.5.1: The set J is the set of initially ($t = 0$) empty nodes and F_J is the face with J^c-coordinates strictly positive. This is just the other way around compared to the definition of Λ (and the face B_Λ in [12]).

Puhalskii's large deviation principle is on $D([0,\infty),\mathbb{R}^d)$ equipped with the extended J_1-topology of Skorohod. It implies a sample path large deviation principle on $D([0,T],\mathbb{R}^d)$ equipped with the J_1-topology with the same local rate function.

Skorohod's J_1-topology is a metric topology, denote by $d_{J_1}(\cdot,\cdot)$ a metric inducing this topology (cf d_d in display (A.2), (A.3) below theorem A.53 of [23]). Convergence in $D([0,T],\mathbb{R}^d)$ to affine functions in the supremum norm induced metric and d_{J_1} is equivalent, since: for $\psi, f \in D([0,T],\mathbb{R})$ by definition of the metrices

$$d_{J_1}(f,\psi) \leq ||f-\psi||$$

And if $||\psi'|| = \sup_{t\in[0,T]} |\psi'(t)| < \infty$ then

$$||f-\psi|| \leq (||\psi'||+1)\, d_{J_1}(\psi,f)$$

Thus, open balls around affine functions wrt these metrices can be nested. Define the open ball U around ψ with $\psi(t) = x + tv$ wrt d_{J_1}

$$U(\delta) := \{f \in D([0,T],\mathbb{R}^d) \mid d_{J_1}(f, t \mapsto x+tv) < \delta\}$$

then

$$\mathcal{U}_\delta(t \mapsto x + tv) \subseteq U(\delta) \subseteq \overline{U(\delta)} \subseteq \mathcal{U}_{\delta\,(1+|v|)}(t \mapsto x + tv)$$

and

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Z_n \in \mathcal{U}_\delta(t \mapsto x + tv)) &\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Z_n \in \overline{U(\delta)}) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(Z_n \in \mathcal{U}_{(1+|v|)\delta}(t \mapsto x + tv))\end{aligned}$$

Then let $\delta \to 0$. From goodness of the rate function $I^Q_{q_0}$, closedness of $\{\psi\}$, and [5], p.119

$$T\,\mathcal{R}_{J(x)}(v) = I^Q_{q_0}(\psi) = \inf_{g\in\{\psi\}} I^Q_{q_0}(g) = \lim_{\delta\to 0} \inf_{g\in\overline{U(\delta)}} I^Q_{q_0}(g)$$

and we finally obtain

$$-TL(x,v) \leq -T\mathcal{R}_{J(x)}(v) \leq -TL(x,v)$$

which identifies the local rate functions. □ 6.5.1

Corollary 6.5.2. *The upper bound for the almost Fenchel Legendre transform in 6.2.4 is always a tight bound and*

$$G_K(v) = \sup_{\alpha\in\mathcal{B}_K} \langle \alpha, v\rangle - \Psi(\alpha)$$

6.6 Calculating the local rate function

We have identified the local large deviation rate function $L(\cdot,\cdot)$ for the generalised Jackson network as a restricted optimisation problem 6.0.4, an almost Fenchel-Legendre transform. For the Jackson network the rate function can be expressed as a (real, true) Fenchel-Legendre transform in lower dimensional space. We cite proposition 10.2 of [12] of Ignatiouk-Robert that we generalise here. We give it in our notation. For $L(x,v)$ and $\overline{\alpha}$ the optimiser in the restricted optimisation problem is such that under the change of measure with paramter $\overline{\alpha}$ the network drift becomes v and the deviating event that the scaled network process stays close to the linear function $t \mapsto x + tv$ becomes the expected behaviour, the network's fluid limit. Knowing the change of measure that makes $t \mapsto x + tv$ the fluid limit of the network is knowing the rate function.

Claim 6.6.1. *For (x, v), $K = \{i \,|\, x_i > 0$ or $v_i > 0\}$ there is $\Theta \supseteq K$ and convex $\Psi^\Theta : \mathbb{R}^{|\Theta|} \to \mathbb{R}$ such that*

$$\sup_{\alpha \in \mathcal{B}_K} \langle \alpha, v \rangle - \Psi(\alpha) = \Psi^{\Theta *}(v_\Theta)$$

Let $M \supset K = \{i \,|\, x_i > 0 \text{ or } v_i > 0\}$ and consider the equality constrained optimisation problem.

$$\inf_{\alpha \in \mathcal{D}_M} -\langle \alpha \,,\, v \rangle + \Psi(\alpha) \quad , \quad \mathcal{D}_M := \{\alpha \in \mathbb{R}^d \,|\, \alpha_i = \Xi^{(i)}(\alpha)\, i \in M^c\} \tag{6.16}$$

To better describe elements of $\mathcal{D}_M$ we need the following

Definition 6.6.2. *For strictly substochastic P and $M \subseteq \{1, \dots, d\}$ define the matrix $Q \in \mathbb{R}^{d \times |M|}$*

$$Q \;=\; P_{\{1,\dots,d\}M} + P_{\{1,\dots,d\}M^c}\,(id - P_{M^c})^{-1}\,P_{M^c\,M}$$

Note that Q simplifies when splitting $\{1, \dots, d\}$ into M and M^c.

$$\begin{aligned} Q_{M^C M} &= (\mathrm{id} - P_{M^c})^{-1}\,P_{M^c M} \\ Q_M &= P_M + P_{M M^c}\,(\mathrm{id} - P_{M^c})^{-1}\,P_{M^c\,M} \end{aligned}$$

Remark 6.6.3. *Q is substochastic and 1 is not an eigenvalue of Q_M.*

The remark was proved in [3] Lemma 4.3. Thus the rows $q^{(i)}$ of Q are measures with total mass ≤ 1. The $q^{(i)}$ are not generally equivalent to the $p^{(i)}$ when restricted to M: there may be $j \in M$ such that $q_{ij} > 0 = p_{ij}$. We define the lmgf of $q^{(i)}$ parallel to $\Xi^{(i)}$ for the $p^{(i)}$:

Definition 6.6.4. *In parallel to the definition of Ξ in 4.4.8 define for the sub-probability measure $q^{(j)}$, the j-th row of Q and $\beta \in \mathbb{R}^{|M|}$*

$$\Upsilon^{(j)}(\beta) = \log\Big(\sum_{k=1}^{|M|} q_{jk}\, e^{\beta_k} + \underbrace{(1 - q_{j1} - \cdots - q_{j|M|})}_{=:q_{j0}}\Big).$$

If $q^{(j)}$ is a subprobability measure on $\{e_1, \dots, e_d\}$ then $\Upsilon^{(j)}$ is as in 4.4.6, representing the restriction to M.

Lemma 6.6.5. *If $P \in \mathbb{R}^{d \times d}$ is a substochastic matrix and Q associated with P as in 6.6.2 and for $j \in \{1, \dots, d\}$ $\Xi^{(j)}$ is associated with the j-th row $p^{(j)}$ of P and $\Upsilon^{(j)}$ with the j-th row $q^{(j)}$ of Q then $\Xi^{(j)}(\alpha) = \Upsilon^{(j)}(\alpha_M)$ for any $\alpha \in \mathcal{D}_M$.*

Proof of 6.6.5: We transform e^{α_i} such that expressions become linear: $\breve{\alpha}_i = e^{\alpha_i} - 1$. Definition of $\Upsilon^{(i)}$ and $\Xi^{(i)}$ then become

$$\left(e^{\Xi^{(i)}(\alpha)}\right)_{i=1,\dots,d} = \left(\sum_{j=1}^{d} p_{ij}\, e^{\alpha_j} + (1 - \sum_{j=1}^{d} p_{ij})\right)_{i=1,\dots,d} = P\,\breve{\alpha} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\left(e^{\Upsilon^{(i)}(\alpha_M)}\right)_{i=1,\dots,d} = \left(\sum_{j=1}^{|M|} q_{ij}\, e^{\alpha_j} + (1 - \sum_{j=1}^{|M|} q_{ij})\right)_{i=1,\dots,d} = Q\,\breve{\alpha}_M + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

rewriting the condition $\alpha \in \mathcal{D}_M$ in a similar fashion

$$\alpha_i = \Xi^{(i)}(\alpha)\,\forall i \in M^c \quad \Leftrightarrow \quad \breve{\alpha}_{M^c} = P_{M^c\{1,\dots,d\}}\breve{\alpha} \tag{6.17}$$

and iterating

$$\begin{aligned}
\breve{\alpha}_{M^c} &= P_{M^c\{1,\dots,d\}}\breve{\alpha} \\
&= P_{M^c}\breve{\alpha}_{M^c} + P_{M^cM}\breve{\alpha}_M \\
&= P_{M^c}\big(P_{M^c}\breve{\alpha}_{M^c} + P_{M^cM}\breve{\alpha}_M\big) + P_{M^cM}\breve{\alpha}_M \\
&= (P_{M^c})^2\breve{\alpha}_{M^c} + (P_{M^c} + \mathrm{id})P_{M^cM}\breve{\alpha}_M \\
&= \dots = (P_{M^c})^{n+1}\breve{\alpha}_{M^c} + \sum_{k=0}^{n}(P_{M^c})^k\, P_{M^cM}\breve{\alpha}_M
\end{aligned}$$

which converges

$$\breve{\alpha}_{M^c} = \lim_{n\to\infty}\sum_{k=0}^{n}(P_{M^c})^k\, P_{M^cM}\breve{\alpha}_M = (\mathrm{id} - P_{M^c})^{-1}P_{M^cM}\breve{\alpha}_M$$

Applying this in P split into M and M^c indices

$$\begin{aligned}
P\,\breve{\alpha} &= P_{\{1,\dots,d\}M}\breve{\alpha}_M + P_{\{1,\dots,d\}M^c}\breve{\alpha}_{M^c} \\
&= P_{\{1,\dots,d\}M}\breve{\alpha}_M + P_{\{1,\dots,d\}M^c}(\mathrm{id} - P_{M^c})^{-1}P_{M^cM}\breve{\alpha}_M \\
&= Q\,\breve{\alpha}_M
\end{aligned}$$

Thus $P\,\breve{\alpha} + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = Q\,\breve{\alpha}_M + \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ and we get the claim. $\square$ 6.6.5

Corollary 6.6.6. $\mathcal{D}_M = \{\alpha \,|\, \alpha_i = \Upsilon^{(i)}(\alpha_M)\,\forall i \in M^c\}$.

Claim 6.6.7. *The equality constraint optimisation problem is a Fenchel-Legendre transform.*

$$\inf_{\alpha\in\mathcal{D}_M} -\langle\alpha,v\rangle+\Psi(\alpha) \;=\; -\Psi^{M*}(v_M)$$

Proof of 6.6.7: For $\alpha\in\mathcal{D}_M$ write $\alpha=\begin{bmatrix}\alpha_M\\ \alpha_{M^c}\end{bmatrix}$ and $u_i=\Upsilon^{(i)}(\alpha_M)$ for $i\in M^c$. Then $\alpha=\begin{bmatrix}\alpha_M\\ u\end{bmatrix}$ by 6.6.6. Thus

$$\inf_{\alpha\in\mathcal{D}_M} -\langle\alpha,v\rangle+\Psi(\alpha) \;=\; \inf_{\alpha_M\in\mathbb{R}^{|M|}} -\langle v,\begin{bmatrix}\alpha_M\\ u\end{bmatrix}\rangle+\Psi(\begin{pmatrix}\alpha_M\\ u\end{pmatrix}).$$

By choice of M we have $v_{M^c}=0$ and $\langle v,\begin{bmatrix}\alpha_M\\ u\end{bmatrix}\rangle=\langle\alpha_M,v_M\rangle$. Also Ψ simplifies:

$$\begin{aligned}
\Psi\big|_{\alpha\in\mathcal{D}_M}(\alpha) &= \sum_{j\in M}\Gamma_A^{(j)}(\alpha_j)+\Gamma_S^{(j)}(-\alpha_j+\underbrace{\Xi^{(j)}(\alpha)}_{=\Upsilon^{(j)}(\alpha_M)})\\
&\quad+\sum_{j\in M^c}\Gamma_A^{(j)}(\underbrace{\alpha_j}_{=\Xi^{(j)}(\alpha)=\Upsilon^{(j)}(\alpha_M)})+\Gamma_S^{(j)}(\underbrace{-\alpha_j+\Xi^{(j)}(\alpha)}_{=0})\\
&= \sum_{j\in M}\Gamma_A^{(j)}(\alpha_j)+\Gamma_S^{(j)}(-\alpha_j+\Upsilon^{(j)}(\alpha_M))+\sum_{j\in M^c}\Gamma_A^{(j)}(\Upsilon^{(j)}(\alpha_M))\\
&=: \Psi^M(\alpha_M)
\end{aligned}$$

and we now have

$$\inf_{\alpha\in\mathcal{D}_M} -\langle\alpha,v\rangle+\Psi(\alpha) \;=\; \inf_{\alpha_M\in\mathbb{R}^{|M|}} -\langle v_M,\alpha_M\rangle+\Psi^M(\alpha_M)$$

and the claim is proved. □ 6.6.7

Proof of 6.6.1: Let $\overline{\alpha}$ be the optimiser of 6.1.4 , (6.8) and

$$A=\{i\in K^c\,|\,\overline{\alpha}_i=\Xi^{(i)}(\overline{\alpha})\}$$

the set of indices of restrictions active in $\overline{\alpha}$. Set

$$\Theta=K\cup A^c=\{i\,|\,x_i>0 \text{ or } v_i>0 \text{ or } \overline{\alpha}_i-\Xi^{(i)}(\overline{\alpha})<0\}.$$

Then $\overline{\alpha}$ is also the optimiser in the equality constrained optimisation problem with restrictions only in $M^c=A$ (cf [2] chapter 3 on Lagrange Multiplier Theory, section 3.3) and

$$\inf_{\alpha\in\mathcal{B}_K} -\langle\alpha,v\rangle+\Psi(\alpha)=\inf_{\alpha\in\mathcal{D}_\Theta} -\langle\alpha,v\rangle+\Psi(\alpha)-\Psi^{\Theta*}(v_\Theta)$$

□ 6.6.1

Remark 6.6.8. *Let $A^{(i)}, S^{(i)}, \ldots, S^{(d)}$ be primitive processes of a generalised Jackson network with lmgfs $\Gamma_A^{(i)}, \Gamma_S^{(i)}$. Let P be the routing matrix of the network. Fix $M \subseteq \{1, \ldots, d\}$ and define Q as in 6.6.2 and let its i-th row define the lmgf $\Upsilon^{(i)}$ as in 6.6.4. For $i \in M^c$ split the i-th arrival process into $\left[\ A^{(ij)}\ \right]_{j\in M}$ wrt $q^{(i)}$ and define the compound arrival process for $i \in M$*

$$\mathbf{A}^{(i)} = A^{(i)} + \sum_{j\in M^c} A^{(ji)}.$$

For $i \in M$ split $S^{(i)}$ wrt $q^{(i)}$ and define the new $\mathbf{S}^{(i)}$

$$\mathbf{S}^{(i)} = \sum_{k\in M} S^{(ik)} - S^{(i)}$$

Then

$$\mathbf{A} + \sum_{j\in M} \mathbf{S}^{(j)}$$

is a free process, its routing matrix is Q_M and it has lmgf Ψ^M as in 6.6.1.

It seems immediate that the network represented by Q_M is open and that Ψ^M is finite on $\mathbb{R}^{|M|}$. General theory for the large deviations of the free process apply. We can now look for a suitable superset Θ in 6.6.1 in the following way:

Algorithm 6.6.9. *To calculate the local large deviation rate function for the generalised Jackson network (cf 6.0.4), especially to find the optimiser $\overline{\alpha} \in \mathcal{B}_K$ and a suitable set Θ of 6.6.1*

1. *Set $M := K$, $\mathcal{S} := \emptyset$.*
2. *Renumber nodes such that $\{1, \ldots, d\} = \{1, \ldots, |M|, \ldots, d\}$.*
3. *Find the optimiser $\tilde{\alpha}_M \in \mathbb{R}^{|M|}$ in $\Psi^{M*}(v_M)$.*
4. *Calculate $\tilde{\alpha}_i = \Upsilon^{(i)}(\tilde{\alpha}_M)$ for $i \in M^c$ such that $\tilde{\alpha} = \begin{bmatrix} \tilde{\alpha}_M \\ \tilde{\alpha}_{M^c} \end{bmatrix} \in \mathcal{D}_M$.*
5. *If under the change of measure $\mathcal{M}^{\emptyset}(\tilde{\alpha}, \cdot)$ there are no strict bottlenecks in the M^c-nodes then add $(M\ ,\ \tilde{\alpha})$ to $\mathcal{S}$. Else, for each strict bottleneck $i \in M^c$ set $M := M \cup \{i\}$ and iterate from 3. on.*
6. *$(\Theta, \overline{\alpha}) \in \mathcal{S}$ such that $\Psi^{\Theta*}(v_\Theta) = \min\{\Psi^{M*}(v_M) \,|\, (M, \tilde{\alpha}) \in \mathcal{S}\}$.*

This compares to theorem 2 of [12].

6.6.1 Interpretation and possible improvement

We now want to give an interpretation of the local rate function of the generalised Jackson network in terms of the associated free rate function $\Psi^{\Theta *}$, cf 6.6.1.

From interpretation 6.1.3 we know that any feasible twist (any $\alpha \in \mathcal{B}_K$) does not decrease service rates at K^c nodes. And from 6.3.4 we have that if the optimal twist parameter strictly increases a service rate of a K^c-node then this node will be a bottleneck. Both interpretations make sense as minimum-cost (in terms of service rates $\Gamma_S^{(i)*}$ and routing $\Xi^{(i)*}$) to allow a certain flow through K^c-nodes: to reduce flow through a restricted node one only has to reduce its input, the service rate does not have to change ($\Gamma_S^{(i)*}(0) = 0$). And increasing the service rate of a restricted node is reasonable only if all of the service capacity is required to get a certain flow through this node.

In steps 3.-5. of the algorithm 6.6.9 the network is partitioned into M and M^c. The free process of M-nodes as in 6.6.8 is twisted to have drift v_M, The twist $\tilde{\alpha}_M$ is the optimiser in $\Psi^{M*}(v_M)$, thus an optimal twist wrt service at and routing between M-nodes and original arrival processes at M^c-nodes. Sevice rates at M^c nodes are not considered in the model of the free process of M-nodes of 6.6.8; cost for these service rates $\Gamma_S^{(i)*}$, $i \in M^c$ do not appear in Ψ^{M*}. It may now happen that as the arrival to M^c nodes is split to become arrivals at M nodes the flow through the M^c-subnetwork does not go as smoothly as assumed in 6.6.8: Nodes in M^c overflow when they cannot handle the flow into M and / or out of M.

If this happens, then in the network the drift v is not realised under the changed measure: $\tilde{\alpha} \neq \overline{\alpha}$ and $M \neq \Theta$. Capacities and cost for increasing capacities of M^c nodes has not been considered in the choice of $\tilde{\alpha}$ but should have been. In the next iteration 3.-5. an increased set M and cost at this increased set of nodes is considered.

In the proof of 6.6.1 we have characterised $(\Theta\, , \overline{\alpha})$ as $\Theta = K \cup \{i \in K^c \,|\, \mu_i(\overline{\alpha}) > \mu_i\}$. Thus if $(M\, , \tilde{\alpha})$ with $i \in M \setminus K$ and $\mu_i(\tilde{\alpha}) < \mu_i$ then $M \neq \Theta$. This allows us to remove a node $i \in M \setminus K$ from the set of free nodes. It would be an advantage if one could change step 5. of algorithm 6.6.9 to become

5'. If under the change of measure $\mathcal{M}^{\emptyset}(\tilde{\alpha}, \cdot)$ there are no strict bottlenecks in the M^c subnetwork and $\mu_i(\tilde{\alpha}) \geq \mu_i$ for all $i \in M \setminus K$ then $\overline{\alpha} = \tilde{\alpha}$ and $\Theta = M$. Else

- If for some $i \in M \setminus K \,:\, \mu_i(\tilde{\alpha}) < \mu_i$ then restrict this node: set $M := M \setminus \{i\}$ and iterate from 3. on.
- If for all $i \in M \backslash K \,:\, \mu_i(\tilde{\alpha}) \geq \mu_i$ and some $i \in M^c$ is a bottleneck in the subnetwork of M^c-nodes then free this node: set $M := M \cup \{i\}$ and iterate from 3. on.

However, it is not obvious that this algorithm terminates or whether the sequence in which nodes are freed and/or restricted influences the final set of free nodes. In the best of all cases step 6. of algorithm 6.6.9 could be omitted.

6.6.2 Example

We give a simple example of calculating the decay rate / local rate function with algorithm 6.6.9. The result is of course the same as when calculating it from the restricted optimisation problem of the almost Fenchel-Legendre transform of 6.0.4. The following example is simple as there will be only one iteration in the algorithm and only one feasible choice for adding a bottleneck in 5.

We work with the network of $d = 4$ nodes as introduced in figure 5.1 of chapter 5. We chose exponential inter event times for all arrival and service events for simplicity. Let the rates be

$$\lambda = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad , \quad \mu = \begin{bmatrix} 3 \\ 6 \\ 4 \\ 5 \end{bmatrix} \quad , \quad P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} \\ \frac{3}{10} & \frac{3}{10} & 0 & 0 \end{bmatrix}.$$

Let the network process start in x end investigate the probability that it evolves in direction v

$$x = \begin{bmatrix} 0.1 \\ 0.2 \\ 0 \\ 0 \end{bmatrix} \quad , \quad v = \begin{bmatrix} -1 \\ -\frac{1}{2} \\ 0 \\ 0 \end{bmatrix}$$

v is not the network drift and we will calculate the decay rate for the scaled network process to stay close to the function $t \mapsto x + tv$.

We have $K = \Lambda = \{1,2\}$ and the local process $W^{\Lambda} = W^{\{1,2\}}$. Figure

6.3 is an adaption of the network to represent the local process: Λ-nodes are circled, indicating the behaviour as always non-empty.

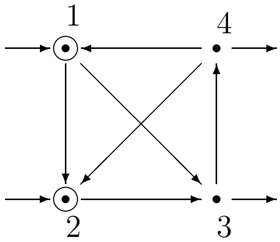


Figure 6.3: Adaption of the network to represent the local process of $\Lambda = \{1,2\}$.

For step 3. of algorith 6.6.9 we transform the network process into the associated free process in $|\Lambda| = 2$ dimensions. We do this in steps. We remove nodes $3,4$ but keep the flow through these nodes. Flow indirectly leaving the network through $\{3,4\}$-nodes is now documented as exits from the $\{1,2\}$ nodes: both nodes 1 and 2 become exit nodes. Similarly there is flow comming back to each node (via $p_{13}\,p_{34}\,p_{41} > 0$ and $p_{23}\,p_{34}\,p_{42} > 0$) and indirect flow from node 1 to node 2 (via $p_{23}\,p_{34}\,p_{41} > 0$).

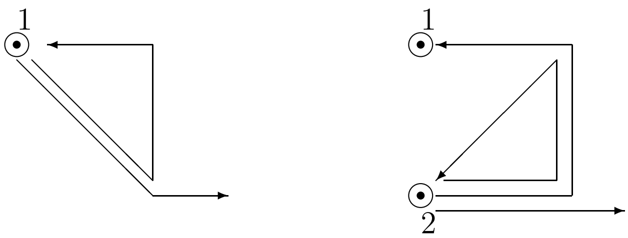


Figure 6.4: Some flows through $\{3,4\}$-nodes

Since immediate feedback is not allowed in our model we have to remodel inter event times at these feedback nodes as in section 5.1.

We calculate Q with $\Lambda = \{1,2\}$ and $\Lambda^c = \{3,4\}$.

$$\begin{aligned} Q_{\Lambda} &= P_{\Lambda} + P_{\Lambda\Lambda^c}\,(\mathrm{id} - P_{\Lambda^c})^{-1}\,P_{\Lambda^c\Lambda} = \begin{bmatrix} \frac{3}{40} & \frac{23}{40} \\ \frac{3}{20} & \frac{3}{20} \end{bmatrix} \\ Q_{\Lambda^c\Lambda} &= \sum_{k=0}^{\infty} P_{\Lambda^c}^k\,P_{\Lambda^c\Lambda} = \begin{bmatrix} \frac{3}{20} & \frac{3}{20} \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Removing immediate feedback we have to remodel inter event times (and their lmgfs): In this setting we now have inter event times $\tau^{(i)S\Lambda} = \tau^{\circ(i)S}$

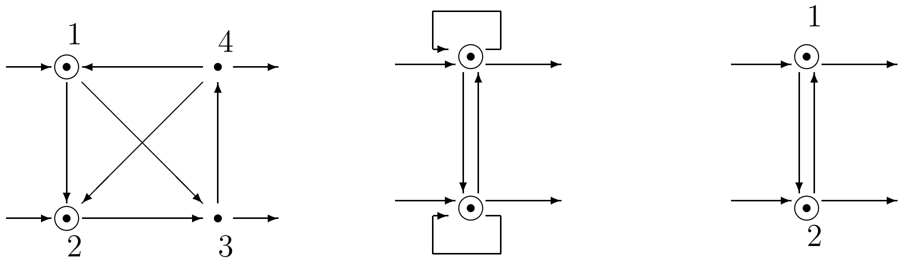


Figure 6.5: Construction of the free subprocess for $M = \Lambda = \{1,2\}$

with $\mathbb{E}[\tau^{\circ(i)S}] = \frac{1}{\mu\,(1-q_{ii})}$. Generally the new free processes primitives have the rates

	free process of Λ-nodes (with immediate feedback)	free process remodelled (without immediate feedback, cf 5.1)
arrival rates	$\lambda_\Lambda + (Q_{\Lambda^c\Lambda})^\top \lambda_{\Lambda^c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$
service rates	$\mu_\Lambda = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$	$[\mu_i(1-q_{ii})]_{i\in\Lambda} = \begin{bmatrix} 2.775 \\ 5.1 \end{bmatrix}$
routing	$Q_\Lambda = \begin{bmatrix} \frac{3}{40} & \frac{23}{40} \\ \frac{3}{20} & \frac{3}{20} \end{bmatrix}$	$\begin{bmatrix} 0 & \frac{23}{37} \\ \frac{3}{17} & 0 \end{bmatrix}$

We now investigate the free process with rates $(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2.775 \\ 5.1 \end{bmatrix}, \begin{bmatrix} 0 & \frac{23}{37} \\ \frac{3}{17} & 0 \end{bmatrix})$.

$$\Psi^{\{1,2\}*}(\begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix}) = 0.2694 \quad , \quad \tilde{\alpha}_{\{1,2\}} = \begin{bmatrix} 0.1457 \\ 0.3028 \end{bmatrix}$$

Step 4: We use rows of $Q_{\Lambda^c\Lambda}$ to calculate remaining coordinates of $\tilde{\alpha}$. For example $\tilde{\alpha}_3 = \log(q_{31}\, e^{\tilde{\alpha}_1} + q_{32}\, e^{\tilde{\alpha}_2} + q_{30})$.

$$\tilde{\alpha} = \begin{bmatrix} 0.1457 \\ 0.3028 \\ 0.0738 \\ 0 \end{bmatrix}$$

Step 5: We check if under the twist $\tilde{\alpha}$ the Λ^c-nodes have strict bottlenecks.

twisted arrival rates	twisted service rates	twisted routing
$\begin{bmatrix} 1.2 \\ 1.4 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 3.2 \\ 4.8 \\ 3.7 \\ 5.8 \end{bmatrix}$	$\begin{bmatrix} 0 & 0.55 & 0.44 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 \\ 0.3 & 0.35 & 0 & 0 \end{bmatrix}$

Node 3 is a strict bottleneck. Also note that with these rates the evolution at node 1 is less than the required $v_1 = -1$ as flow that was supposed to reach node 1 via node 3 is held back at the bottleneck node 3. We free the bottleneck node 3, set $M := \{1, 2, 3\}$, and continue with step 3.

Step 3: For $M = \{1, 2, 3\}$ construct the associated free process. We give flows of node 3 that are affected by removing node 4.

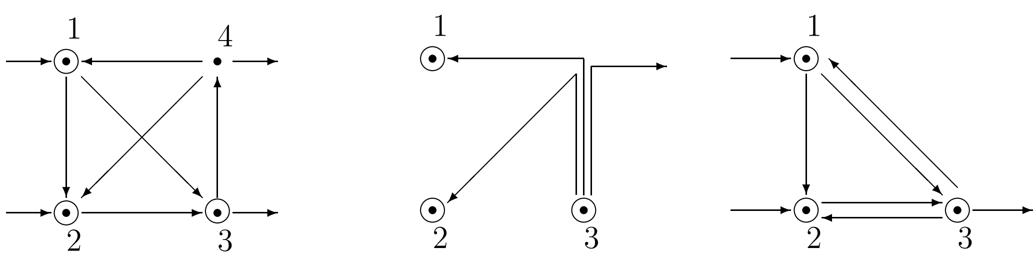


Figure 6.6: Construction of the free subprocess for $M = \{1, 2, 3\}$

We calculate Q for the free subprocess. We now have $M^c = \{4\}$.

$$\begin{aligned} Q_M &= P_M + P_{M\{4\}} P_{\{4\}M} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{3}{20} & \frac{3}{20} & 0 \end{bmatrix} \\ Q_{\{4\}M} &= \underbrace{\sum_{k=0}^{\infty} P_{\{4\}}^{k}}_{=\mathrm{id}=[1]} P_{\{4\}M} = \begin{bmatrix} \frac{3}{10} & \frac{3}{10} & 0 \end{bmatrix} \end{aligned}$$

now consider the free process with ates $\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{3}{20} & \frac{3}{20} & 0 \end{bmatrix} \right)$.
We get the free rate function and its optimiser.

$$\Psi^{\{1,2,3\}*}\left(\begin{bmatrix} -1 \\ -\frac{1}{2} \\ 0 \end{bmatrix} \right) = 0.5754 \quad , \quad \tilde{\alpha}_{\{1,2,3\}} = \begin{bmatrix} 0.0121 \\ 0.1110 \\ -0.2595 \end{bmatrix}$$

Step 4:

$$\tilde{\alpha} = \begin{bmatrix} 0.0121 \\ 0.1110 \\ -0.2595 \\ 0.0381 \end{bmatrix}$$

twisted arrival rates	twisted service rates	twisted routing
$\begin{bmatrix} 1.0 \\ 1.1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 2.8 \\ 4.1 \\ 5.3 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 0 & 0.59 & 0.41 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.51 \\ 0.29 & 0.32 & 0 & 0 \end{bmatrix}$

These rates have node 3 a bottleneck but not a strickt bottleneck, and node 4 in equilibrium. The network drift is v.

Step 5: $\Theta = \{1, 2, 3\}$ and $\overline{\alpha} = \begin{bmatrix} 0.0121 \\ 0.1110 \\ -0.2595 \\ 0.0381 \end{bmatrix}$. The algorithm terminates with $L(x, v) = 0.5754$.

The following are 4 simulations of the 4-nodes network under the twisted distribution with $n = 3000$ and $T = .05$.

The exponential decay rate (per time) is about .5754: The probability that the scaled network process follows $t \mapsto x + tv$ for an interval of scaled time length 0.2 has a decay rate in n of about $.2 * .5754 = .1151$.

Chapter 7

Appendix

7.1 Functional strong law of large numbers

Let $X_1, X_2, \dots$ be iid and centred with $E[X_1] = 0$. Let S_n be the n-th partial sum and $0, S_1, S_2, \dots$ a path of partial sums. Let Z_n be the scaled process under the law of large numbers scaling.

$$Z_n : \mathbb{R}_{\geq 0} \to \mathbb{R} \quad , \quad t \mapsto \frac{1}{n} S_{\lfloor nt \rfloor}$$

From the strong law of large numbers for almost every path of partial sums

$$\forall \epsilon > 0 \, \exists n_0 \in \mathbb{N} : \forall n \geq n_0 \, :< \left| \frac{1}{n} S_n \right| < \frac{\epsilon}{T}$$

where the n_0 depends on the path. Fix $\epsilon > 0$ and for each path the associated n_0 where it exists. Then for an arbitrary path where such a finite n_0 exists, there is also a maximum of the partial sums before n_0 attained (again depending on the path).

$$M := \max_{m \in \{1, \dots, n_0\}} | \, Z_m \, |$$

When choosing n large enough the scaled partial sums process Z_n before and after $\frac{n_0}{n}$ will be arbitrarily small. We do an exact calculation: Fix $\delta > 0$. Fix for $\epsilon := \frac{\delta}{T}$ for each path the n_0 and M as above. Choose $n \geq n_0$ and $n \geq \frac{M}{\delta}$. Both lower bounds depend on the path. They are finite for almost all paths. Let $t \in [0, T]$. For $n \geq \max\{n_0, \frac{M}{\delta}, \frac{1}{T}\}$ we either have $nt \geq n_0$ or not. First consider the first case.

$$\begin{aligned}
\big|Z_n(t)\big| &\leq \frac{1}{n}\max\{|S_{\lfloor nt\rfloor}|\,,\,|S_{\lfloor nt\rfloor+1}|\} \\
&\leq \max\{\frac{\lfloor nt\rfloor}{n}\underbrace{\frac{1}{\lfloor nt\rfloor}|S_{\lfloor nt\rfloor}|}_{<\frac{\delta}{T}}\,,\,\frac{\lfloor nt\rfloor+1}{n}\underbrace{\frac{1}{\lfloor nt\rfloor+1}|S_{\lfloor nt\rfloor+1}|}_{<\frac{\delta}{T}}\} \\
&\leq \frac{\lfloor nt\rfloor+1}{n}\frac{\delta}{T} = \frac{\lfloor nt\rfloor}{n}\frac{\delta}{T}+\frac{1}{n}\frac{\delta}{T} \leq 2\delta
\end{aligned}$$

And now the second with $nt < n_0$.

$$\big|Z_n(t)\big| = \frac{1}{n}\max\{|S_{\lfloor nt\rfloor}|\,,\,|S_{\lfloor nt\rfloor+1}|\} \leq \frac{1}{n}M \leq \delta$$

So almost all paths converge under the scaling and in the sup-norm to the function $t \mapsto 0$.

$$1 = P(\lim_{n\to\infty}\sup_{t\in[0,T]}|Z_n(t)| = 0)$$

In another notation with $X_1 := \tau_1 - E[\tau]$ we have $S_n = \sum_{k=1}^{n}\tau_k - nE[\tau]$ and $Z_n(t) = \frac{1}{n}S_{\lfloor nt\rfloor} = \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\tau_k - \frac{\lfloor nt\rfloor}{n}E[\tau]$. Since the difference between $\frac{\lfloor nt\rfloor}{n}E[\tau]$ and $tE[\tau]$ is bounded by $\frac{E[\tau]}{n}$ we get the almost sure convergence:

$$1 = P(\lim_{n\to\infty}\sup_{t\in[0,T]}|\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\tau_k - tE[\tau]| = 0)$$

7.1.1 Implication for the counting process

Define the partial sums and the interpolated partial sums process wrt $\tau_1, \tau_2, \ldots$

$$\begin{aligned}
Y_n &: t \mapsto \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor}\tau_k \\
\hat{Y}_n &: t \mapsto Y_n(t) + \frac{nt-\lfloor nt\rfloor}{n}\tau_{\lfloor nt\rfloor+1}
\end{aligned}$$

And note that

- $Y_n(\frac{k}{n}) = \hat{Y}_n(\frac{k}{n})$ for all $k \in \mathbb{N}$.
- $\hat{Y}_n(t) \geq Y_n(t)$ and $\hat{Y}_n(t) > Y_n(t)$ for $t \notin \{\frac{k}{n}\,|\,k\in\mathbb{N}\}$.

Claim 7.1.1. $Y_n \in \mathcal{U}_\epsilon(t \mapsto t\lambda) \quad \Rightarrow \quad \hat{Y}_n \in \mathcal{U}_\epsilon(t \mapsto t\lambda)$.

Proof of 7.1.1: Note that if Y_n is close to $t \mapsto t\lambda$ then necessarily Y_n is close to $t \mapsto t\lambda$ whenever it jumps, that is in $\{\frac{k}{n} \,|\, k = 0, 1, \ldots, nT\}$:

$$Y_n \in \mathcal{U}_\epsilon(t \mapsto t\lambda) \qquad \Rightarrow \qquad \epsilon > |Y_n(\frac{k}{n}) - \frac{k}{n}\lambda| \qquad (k = 1, \ldots, nT)$$

But this is enough for $\hat{Y}_n$ to be close to $t \mapsto t\lambda$: $(t, \hat{Y}_n(t))$ is the interpolation between $\left(\frac{k}{n}, Y_n(\frac{k}{n})\right)$ and $(\frac{k+1}{n}, Y_n(\frac{k+1}{n}))$ for $k = \frac{\lfloor nt \rfloor}{n}$. And since $\mathcal{U}_\epsilon(t \mapsto t\lambda)$ is a convex subset of $\mathbb{R}^2$ it has to contain $(t, \hat{Y}_n(t))$. $\square$ 7.1.1

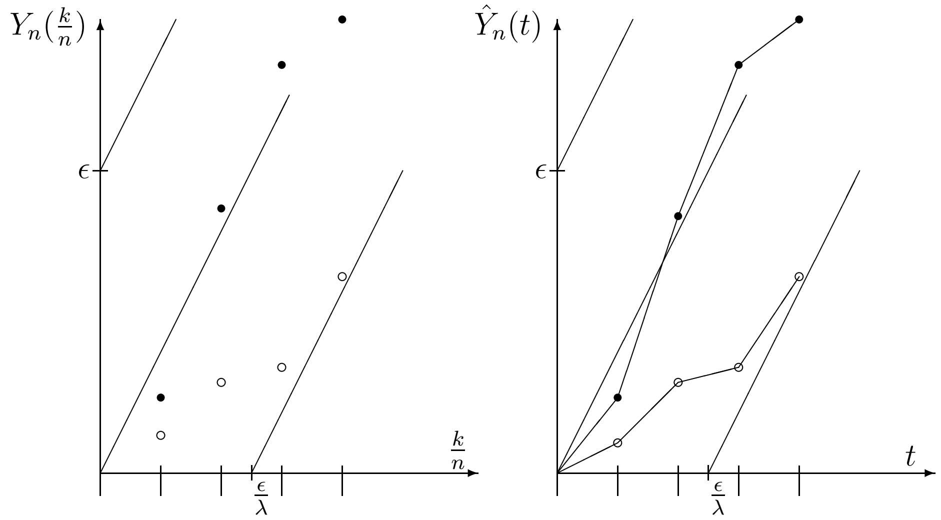


Figure 7.1: 2 realisations of Y_n at jumptimes $\{\frac{k}{n} \,|\, k = 1, \ldots, 4\}$ and as interpolated functions $\hat{Y}_n$

Claim 7.1.2. *For the interpolated counting process a functional strong law of large numbers holds.*

Proof of 7.1.2: From 7.1.1 the functional strong law of large numbers holds for the interpolated partial sums process too. $\hat{Y}^{-1} = \hat{N}$ so

$$\begin{aligned}\sup_{t\in[0,T]} |\hat{Y}(t) - t\,\mathbb{E}[\tau]| < \epsilon \;\;&\Rightarrow\;\; \sup_{t\in[0,T\mathbb{E}[\tau]-\epsilon]} |\hat{Y}^{-1}(t) - t\,\frac{1}{\mathbb{E}[\tau]}| < \frac{\epsilon}{\mathbb{E}[\tau]} \\ &\Leftrightarrow\;\; \sup_{t\in[0,T\mathbb{E}[\tau]-\epsilon]} |\hat{N}(t) - t\,\frac{1}{\mathbb{E}[\tau]}| < \frac{\epsilon}{\mathbb{E}[\tau]}\end{aligned}$$

For an ϵ_0 smaller than ϵ we get

$$\begin{aligned}
1 &= \mathbb{P}(\lim_{n\to\infty} \sup_{t\in[0,T]} |\frac{1}{n}\hat{Y}(nt) - t\,\mathbb{E}[\tau]| < \epsilon) \\
&\leq \mathbb{P}(\lim_{n\to\infty} \sup_{t\in[0,T\mathbb{E}[\tau]-\epsilon]} |\frac{1}{n}\hat{Y}^{-1}(nt) - t\,\frac{1}{\mathbb{E}[\tau]}| < \frac{\epsilon}{\mathbb{E}[\tau]}) \\
&\leq \mathbb{P}(\lim_{n\to\infty} \sup_{t\in[0,T\mathbb{E}[\tau]-\epsilon_0]} |\frac{1}{n}\hat{N}(nt) - t\,\frac{1}{\mathbb{E}[\tau]}| < \frac{\epsilon}{\mathbb{E}[\tau]})
\end{aligned}$$

Claim 7.1.3. *A functional strong law of large numbers holds for the uninterpolated counting process.*

Let $0 < \epsilon' < \epsilon$ and $n \geq \frac{1}{\epsilon-\epsilon'}$.

$$\begin{aligned}
||N_n - (t \mapsto t\lambda)|| < \epsilon &\Leftarrow ||\hat{N}_n - (t \mapsto t\lambda)|| + \underbrace{||N_n - \hat{N}_n||}_{\leq\frac{1}{n}} < \epsilon \\
&\Leftarrow ||\hat{N}_n - (t \mapsto t\lambda)|| < \epsilon - \frac{1}{n} \\
&\Leftarrow ||\hat{N}_n - (t \mapsto t\lambda)|| < \epsilon'
\end{aligned}$$

and

$$\mathbb{P}(\lim_{n\to\infty} ||N_n - (t \mapsto t\lambda)|| < \epsilon) \quad \geq \quad \mathbb{P}(\lim_{n\to\infty} ||N_n - (t \mapsto t\lambda)|| < \epsilon') = 1$$

7.2 Implications from exponential equivalence

Claim 7.2.1. *If N and N' are exponentially equivalent then*

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = \lim_{\epsilon\to 0}\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}(N'_n \in \mathcal{U}_\epsilon(\psi)).$$

If $\epsilon > 0$ is fixed and we have

$$\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = f(\epsilon)$$

for some f continuous in ϵ then

$$\lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) = \lim_{n\to\infty} \frac{1}{n}\log\mathbb{P}(N'_n \in \mathcal{U}_\epsilon(\psi)).$$

Proof of 7.2.1: Assume N and N' are coupled such that their difference decays super exponentially fast as required for exponential equivalence. We have for any $\delta > 0$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(||N_n - N'_n|| > \delta) = -\infty$$

And we want for U convex and open or closed and $U^\circ \neq \emptyset$

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in U) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in U)$$

Let's try. Let U^δ be the closed blow up of U

$$\begin{aligned} \mathbb{P}(N_n \in U) &= \mathbb{P}(N_n \in U\,,\, ||N_n - N'_n|| \le \delta) + \mathbb{P}(N_n \in U\,,\, ||N_n - N'_n|| > \delta) \\ &\le \mathbb{P}(N'_n \in U^\delta\,,\, ||N_n - N'_n|| \le \delta) + \mathbb{P}(N_n \in U\,,\, ||N_n - N'_n|| > \delta) \\ &\le \mathbb{P}(N'_n \in U^\delta) + \mathbb{P}(||N_n - N'_n|| > \delta) \end{aligned}$$

and thus

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in U) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \left(\mathbb{P}(N'_n \in U^\delta) + \mathbb{P}(||N_n - N'_n|| > \delta) \right) \\ &= \max\{\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in U^\delta)\,,\, -\infty\} \\ &= \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in U^\delta) \end{aligned}$$

and since $\delta > 0$ was arbitrary and for $U = \mathcal{U}_\epsilon(\psi)$ and for the limit of $\epsilon \to 0$.

$$\begin{aligned} \lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\le \lim_{\epsilon\to 0} \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_{\epsilon+\delta}(\psi)) \\ &= \lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_\epsilon(\psi)) \end{aligned}$$

By symmetry we are done.

In the case of $\epsilon > 0$ fixed with a bound continuous in ϵ

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) &\le \lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_{\epsilon+\delta}(\psi)) \\ &= \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_\epsilon(\psi)) \end{aligned}$$

For a lower bound we need the δ-interior of U, that is a non-empty subset of U such that its δ-blow up is still contained in U. Let this be $U^{-\delta}$.

$$\begin{aligned}
P(N_n \in U) &= \mathbb{P}(N'_n + (N_n - N'_n) \in U) \\
&\geq \mathbb{P}(N'_n + (N_n - N'_n) \in U \,,\, ||N_n - N'_n|| \leq \delta) \\
&\geq \mathbb{P}(N'_n \in U^{-\delta} \,,\, ||N_n - N'_n|| \leq \delta)
\end{aligned}$$

and for $U = \mathcal{U}_\epsilon(\psi)$, $U^{-\delta} = \mathcal{U}_{\epsilon-\delta}(\psi)$

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N_n \in \mathcal{U}_\epsilon(\psi)) \\
&= \quad \lim_{\delta\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_{\epsilon-\delta}(\psi) \,,\, ||N_n - N'_n|| \leq \delta) \\
&= \quad \lim_{\delta\to 0} \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_{\epsilon-\delta}(\psi)) \\
&\overset{\text{continuity}}{\underset{\text{in } \epsilon}{=}} \quad \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(N'_n \in \mathcal{U}_\epsilon(\psi))
\end{aligned}$$

$\square$
7.2.1

Lemma 7.2.2. *For $\theta > 0$ and $|| \cdot ||$ the supremum norm over $[0,1]$:*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{n\theta ||N_n - N'_n||}] = 0$$

Proof of 7.2.2: Let $\delta > 0$.

$$\mathbb{E}[e^{n\theta||N_n - N'_n||}] \quad = \quad \mathbb{E}[e^{n\theta||N_n - N'_n||} \, 1\!\!1_{||N_n - N'_n|| > \delta}] + \mathbb{E}[e^{n\theta||N_n - N'_n||} \, 1\!\!1_{||N_n - N'_n|| \leq \delta}]$$

Investigate summands separately, starting with the first. Apply Hölder

$$\begin{aligned}
\mathbb{E}[e^{n\theta||N_n - N'_n||} \, 1\!\!1_{||N_n - N'_n|| > \delta}] &\leq \mathbb{E}[e^{np\theta||N_n - N'_n||}]^{\frac{1}{p}} \, \mathbb{E}[1\!\!1_{||N_n - N'_n|| > \delta}]^{\frac{1}{q}} \\
&\leq \mathbb{E}[e^{np\theta(N_n(1)+\delta)}]^{\frac{1}{p}} \, \mathbb{E}[1\!\!1_{||N_n - N'_n|| > \delta}]^{\frac{1}{q}}
\end{aligned}$$

with some $p \geq 1$. Since $\Gamma(\cdot)$ is finite on all of $\mathbb{R}$:

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\theta||N_n - N'_n||} \, 1\!\!1_{||N_n - N'_n|| > \delta}] \\
&\leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{p\theta N(n)}]^{\frac{1}{p}} + \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[1\!\!1_{||N_n - N'_n|| > \delta}]^{\frac{1}{q}} \\
&= \frac{1}{p} \Gamma(p\theta) - \infty \\
&= -\infty
\end{aligned}$$

Continuing for the second.

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{n\theta||N_n - N'_n||} \, 1\!\!1_{||N_n - N'_n|| \le \delta}] \;\le\; \limsup_{n\to\infty} \frac{1}{n} \log e^{n\theta\delta} \;=\; \theta\delta$$

Thus

$$\lim_{\delta\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\theta||N_n - N'_n||}] \;=\; \lim_{\delta\to 0} \max\{-\infty\,,\, \delta\theta\} = 0$$

But $\theta > 0$ and $||\cdot|| \ge 0$ makes $\mathbb{E}[e^{\theta||N_n - N'_n||}] \ge 1$ and we cannot have decay. So the lim has to be $= 0$.

$\square$
7.2.2

Claim 7.2.3. *If N, N' are exponentially equivalent then they have the same lmgfs:*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N'_t}]$$
$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\langle \theta, N_t\rangle}] = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\langle \theta, N'_t\rangle}]$$

Proof of 7.2.3: For $\theta > 0$. Upper bound: Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$

$$\begin{aligned}
\mathbb{E}[e^{\theta N_t}] \;&=\; \mathbb{E}[e^{\theta N'_t + \theta(N_t - N'_t)}] \;\le\; \mathbb{E}[e^{\theta N'_t + \theta||N - N'||_{[0,t]}}] \\
&\le\; \mathbb{E}[e^{p\theta N'_t}]^{\frac{1}{p}} \, \mathbb{E}[e^{q\theta||N-N'||_{[0,t]}}]^{\frac{1}{q}}
\end{aligned}$$

under the exponential scaling

$$\begin{aligned}
\frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \;&\le\; \frac{1}{tp} \log \mathbb{E}[e^{p\theta N'_t}] + \frac{1}{tq} \log \mathbb{E}[e^{tq\theta \frac{1}{t}||N-N'||_{[0,t]}}] \\
&\to\; \frac{1}{p}\Gamma(p\theta) + 0 \qquad (t\to\infty)
\end{aligned}$$

by application of 7.2.2. As p is arbitrary with only $p > 1$ we let $p \to 1$. We have $\mathcal{D}(\Gamma) = \mathbb{R}$ and continuity of $\Gamma(\cdot)$ from finiteness and convexity. We get the upper bound $\Gamma(\theta)$. Lower bound, still $\theta > 0$:

$$\begin{aligned}
\mathbb{E}[e^{\theta N_t}] \;&=\; \mathbb{E}[e^{\theta N'_t + \theta(N_t - N'_t)}] \\
&\ge\; \mathbb{E}[e^{\theta N'_t - \theta||N-N'||_{[0,t]}}] \\
&\ge\; \mathbb{E}[e^{\theta N'_t - \theta||N-N'||_{[0,t]}} \, 1\!\!1_{||N-N'|| \le t\delta}] + \mathbb{E}[e^{\theta N'_t - \theta||N-N'||_{[0,t]}} \, 1\!\!1_{||N-N'|| > t\delta}] \\
&\ge\; \mathbb{E}[e^{\theta N'_t - \theta t\delta} \, 1\!\!1_{||N-N'|| \le t\delta}]
\end{aligned}$$

under the exponential scaling

$$\begin{aligned}
\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \;&\ge\; \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N'_t} \, 1\!\!1_{||N-N'|| \le t\delta}] - \theta\delta \\
&=\; \Gamma(\theta) - \delta\,\theta
\end{aligned}$$

as we had $\delta > 0$ arbitrarily small, the lower bound is done.

For $\theta < 0$ we apply basically the same tool and in general dimensions for $\theta \in \mathbb{R}^d$ and $|\cdot|$ a norm in $\mathbb{R}^d$, $||\cdot||_{[0,t]}$ the supremum norm over $[0,t]$. Bounds work similarly starting from

$$\begin{aligned}\mathbb{E}[e^{\langle\theta,N_t\rangle}] &= \mathbb{E}[e^{\langle\theta,N_t'\rangle+\langle\theta,N_t-N_t'\rangle}] \leq \mathbb{E}[e^{\langle\theta,N_t'\rangle+|\theta|\,|N_t-N_t'|}] \\ &\leq \mathbb{E}[e^{\langle\theta,N_t'\rangle+|\theta|\,||N_t-N_t'||_{[0,t]}}]\end{aligned}$$

□
7.2.3

7.3 Fenchel-Legendre transforms

Some simple transformations. All functions are assumed to be convex.

$$(cf)^*(x) = cf^*(\frac{x}{c}) \tag{7.1}$$
$$(f+g)^*(x) = \inf_\alpha f^*(x-\alpha) + g^*(\alpha) \tag{7.2}$$
$$(f \circ g)^*(x) = \inf_{\alpha\in\mathbb{R}} \alpha\, g^*(\frac{x}{\alpha}) + f^*(\alpha) \tag{7.3}$$

And as a combination of the above

$$\begin{aligned}(h + f\circ g)^*(x) &\leq \inf_{\alpha\in\mathbb{R}} h^*(x-\alpha) + (f\circ g)^*(\alpha) \\ &\leq \inf_{\alpha\in[0,1]} \inf_\beta h^*(x-\alpha) + \beta\, g^*(\frac{\alpha}{\beta}) + f^*(\beta)\end{aligned}$$

Should we prove them?

$$(c\,f)^*(x) = \sup_\theta \theta\, x - c\, f(\theta) = c\,\sup_\theta \theta\frac{x}{c} - f(\theta) = c\, f^*(\frac{x}{c})$$

$$\begin{aligned}(f+g)^*(x) &= \sup_\theta \theta\, x - f(\theta) - g(\theta) \\ &= \sup_\theta \theta\,(x-\alpha) - f(\theta) + \alpha\,\theta - g(\theta) \\ &\leq \sup_\theta \theta\,(x-\alpha) - f(\theta) + \sup_\theta \alpha\,\theta - g(\theta) \\ &= f^*(x-\alpha) + g^*(\alpha)\end{aligned}$$

With equality if the optimiser is the same in f^* and g^*. Also we can optimise the bound.

$$(f+g)^*(x) \leq \inf_\alpha f^*(x-\alpha) + g^*(\alpha)$$

If there is an optimal θ for $(f+g)^*(x)$ then

$$f'(\theta)+g'(\theta)=x \quad \Leftrightarrow \quad \left\{ \begin{array}{ccc} f'(\theta) & = & x-g'(\theta) \\ g'(\theta) & = & g'(\theta) \end{array} \right\} \quad \Leftrightarrow \quad \underbrace{\left\{ \begin{array}{ccc} f'(\theta) & = & x-\alpha \\ g'(\theta) & = & \alpha \end{array} \right\}}_{\text{for some } \alpha}$$

and the same θ is the optimiser in $f^*(x-\alpha)$ and $g^*(\alpha)$. We got the equality of (7.2). If an optimal θ does not exist for finite $(f+g)^*(x)$ we probably get equality by approximation. If $(f+g)^*(x)=\infty$ there is nothing to do. The claim and the proof do not rely on $x\in\mathbb{R}$, it is for general $f,g:\mathbb{R}^m\to\mathbb{R}$ and $x,\alpha\in\mathbb{R}^m$.

Next one. We have $g:\mathbb{R}^m\to\mathbb{R}$ and $f:\mathbb{R}\to\mathbb{R}$.

$$\begin{aligned} (f\circ g)^*(x) &= \sup_\theta \theta\, x - f(g(\theta)) \\ &= \sup_\theta \theta\, x - \alpha\, g(\theta) + \alpha\, g(\theta) - f(g(\theta)) \\ &\leq \sup_\theta \theta\, x - \alpha\, g(\theta) + \sup_\theta \alpha\, g(\theta) - f(g(\theta)) \\ &= \alpha \sup_\theta \theta\, \frac{x}{\alpha} - g(\theta) + \sup_{\xi\in g(\mathbb{R})} \alpha\,\xi - f(\xi) \\ &\leq \alpha\, g^*(\frac{x}{\alpha}) + f^*(\alpha) \end{aligned}$$

We argue as before: If there is an optimising θ for $(f\circ g)^*(x)$ it will satisfy $(f\circ g)'(\theta)=x$ and if we set $\alpha := f'(g(\theta))\ (\in\mathbb{R})$ then we get $g'(\theta)=\frac{x}{\alpha}$ and have obtains optimisers for $f^*(\alpha)$ and $g^*(\frac{x}{\alpha})$. Again we got equality as in (7.3). ☐

Claim 7.3.1. *For $g=\pi_j$ in (7.3):* $\big(f\circ\pi_j\big)^*(\alpha) = \begin{cases} f^*\circ\pi_j(\alpha) & ,\alpha=\pi_j(\alpha) \\ \infty & ,\textit{else} \end{cases}$.

Define $\Pi_j^\perp$ as the subspace of $\mathbb{R}^d$ perpendicular to $\pi_j(\mathbb{R}^d)$ such that $\mathbb{R}^d = \pi_j(\mathbb{R}^d)\oplus\Pi_j^\perp$ and write

$$\begin{aligned} \alpha = \pi_j(\alpha)+\alpha^\perp \quad &, \quad \alpha\perp\in\Pi_j^\perp \\ \theta = \pi_j(\theta)+\theta^\perp \quad &, \quad \theta\perp\in\Pi_j^\perp . \end{aligned}$$

This implies $\langle\alpha\,,\,\theta\rangle = \langle\pi_j(\alpha)\,,\,\pi_j(\theta)\rangle + \langle\alpha^\perp\,,\,\theta^\perp\rangle$ which we apply in the definition of the F-L transform:

$$\begin{aligned} \big(f\circ\pi_j\big)^*(\alpha) &= \sup_{\theta\in\mathbb{R}^d} \langle\pi_j(\alpha)\,,\,\pi_j(\theta)\rangle - f\circ\pi_j(\theta) + \langle\alpha^\perp\,,\,\theta^\perp\rangle \\ &= \underbrace{\sup_{\theta^{(1)}\in\pi_j(\mathbb{R}^d)} \langle\pi_j(\alpha)\,,\,\theta^{(1)}\rangle - f(\theta^{(1)})}_{=f^*(\pi_j(\alpha))} + \underbrace{\sup_{\theta^{(2)}\in\Pi_j^\perp} \langle\alpha^\perp\,,\,\theta^{(2)}\rangle}_{\in\{0,\infty\}} \end{aligned}$$

Iff $\alpha = \pi_j(\alpha)$ then $\alpha^\perp = 0$ implying the claimed statement. $\square$

7.4 The shifted inter event time

Let τ be an inter event time with distribution function F and density f. We denote f_{+x} the shifted density and its distribution function F_{+x}.

$$\begin{aligned}
f_{+x}(t) &= \frac{f(x+t)}{F^c(x)} \\
F^c_{+x}(t) &= \int_{s=t}^{\infty} f_{+x}(s)\,ds = \int_{s=t}^{\infty} \frac{f(s+x)}{F^c(x)}\,ds = \int_{s=t+x}^{\infty} \frac{f(s)}{F^c(x)}\,ds = \frac{F^c(t+x)}{F^c(x)} \\
h_{+x}(t) &= \frac{f_{+x}}{F^c_{+x}}(t) = \frac{f(x+t)}{F^c(x+t)} = h(x+t) \\
H_{+x}(t) &= \int_{s=0}^{t} h_{+x}(s)\,ds = \int_{s=0}^{t} h(s+x)\,ds = \int_{s=x}^{t+x} h(s)\,ds = H(t+x) - H(x)
\end{aligned}$$

and F_{+x} matches H_{+x} by $F^c_{+x} = e^{-H_{+x}}$. The Cesaro limit for the shifted distributions hazard function does not change:

$$L_C(h_{+x}) = \lim_{t\to\infty} \frac{H_{+x}(t)}{t} = \lim_{t\to\infty} \frac{H(t+x) - H(x)}{t} = \lim_{t\to\infty} \frac{H(t+x)}{t} - 0 = \lim_{t\to\infty} \frac{H(t)}{t}$$

and immediately $L_C(h) = L_C(h_{+x})$ and $\mathcal{D}(\Lambda) = \mathcal{D}(\Lambda_{+x})$.

We get finiteness of all moments of τ_{+x}. We calculate the mean for unbounded τ

$$\mathbb{E}[\tau_{+x}] = \int_{s=0}^{\infty} F^c_{+x}(s)\,ds = \int_{s=0}^{\infty} e^{-H(x+s)+H(x)}\,ds = e^{H(x)} \int_{s=x}^{\infty} e^{-H(s)}\,ds$$

If τ is bounded by b , say, then $x < b$ is required. $\tau_{+x} \in (0, b-x)$

$$\mathbb{E}[\tau_{+x}] = \int_{s=0}^{b-x} F^c_{+x}(s)\,ds = \int_{s=0}^{b-x} e^{-H(x+s)+H(x)}\,ds = e^{H(x)} \int_{s=x}^{b} e^{-H(s)}\,ds$$

In both cases (τ bounded or not) we got a product of something large and something small. What happens as $x \to \infty$ (or $x \to b$)?

$$\lim_{x\to ..} \mathbb{E}[\tau_{+x}] = \lim_{x\to ..} \frac{\int_{s=x}^{b} e^{-H(s)}\,ds}{e^{-H(x)}} = \lim_{x\to ..} \frac{-e^{-H(x)}}{-h(x)\,e^{-H(x)}} = \lim_{x\to ..} \frac{1}{h(x)} \tag{7.4}$$

which only makes sense if $\lim_{t\to\infty} h(t)$ exists.

Claim 7.4.1. *If* $\lim_{x\to\infty} h(x)$ *exists then* $\lim_{x\to\infty} \mathbb{E}[\tau_{+x}] = \frac{1}{L_C(h)}$.

- *For* τ *unbounded but LD-bounded:* $L_C(h) = \infty$. *If* $\liminf_{x\to\infty} h(x) = \infty$ *then* $\lim_{x\to\infty} \mathbb{E}[\tau_{+x}] = 0$.
- *If* τ *is bounded by* b*:* $L_C(h) = \infty$ *and if* $\liminf_{x\to b} h(x) = \infty$ *then* $\lim_{x\to b} \mathbb{E}[\tau_{+x}] = 0$.
- *If* τ *is not LD-bounded and* $\lim_{x\to\infty} h(x)$ *exists in* $(0,\infty)$ *(and is* $=$ $L_C(h)$*) then* $\lim_{x\to\infty} \mathbb{E}[\tau_{+x}] = \frac{1}{L_C(h)}$.

$$
\begin{aligned}
\mathbb{E}[e^{\theta\tau_{+x}}] &= \int_{s=0}^{\infty} e^{\theta s}\,\frac{f(s+x)}{F^c(x)}\,ds \;=\; \frac{1}{F^c(x)}\,e^{-\theta x}\int_{s=0}^{\infty} e^{\theta(s+x)}\,f(s+x)\,ds \\
&= \frac{1}{F^c(x)}\,e^{-\theta x}\int_{s=x}^{\infty} e^{\theta s}\,f(s)\,ds \;=\; \frac{1}{F^c(x)\,e^{\theta x}}\int_{s=x}^{\infty} e^{\theta s}\,f(s)\,ds \\
&= \frac{1}{F^c(x)\,e^{\theta x-\Lambda(\theta)}}\int_{s=x}^{\infty} \underbrace{e^{\theta s-\Lambda(\theta)}\,f(s)}_{=f_\theta(s)}\,ds \\
&= \frac{1}{F^c(x)\,e^{\theta x-\Lambda(\theta)}} F^c_\theta(x) \;=\; \frac{F^c_\theta}{F^c}(x)\,e^{-\theta x+\Lambda(\theta)}
\end{aligned} \tag{7.5}
$$

Claim 7.4.2. *Shifting and exponentially twisting commute.*

$$
\begin{aligned}
\big(f_{+x}\big)_\beta(x) &= f_{+x}(t)\,e^{\beta t-\Lambda_{+x}(\beta)} \;=\; \frac{f(t+x)}{F^c(x)}e^{\beta t-\Lambda_{+x}(\beta)} \\
&= \frac{1}{F^c(x)}e^{-\beta x-\Lambda_{+x}(\beta)+\Lambda(\beta)}\,f(t+x)e^{\beta(t+x)-\Lambda(\beta)} \\
&= \frac{1}{F^c(x)e^{\beta x-\Lambda(\beta)}}e^{-\Lambda_{+x}(\beta)}\,f_\beta(t+x) \\
&\overset{(7.5)}{=} \frac{1}{F^c(x)e^{\beta x-\Lambda(\beta)}}\,\frac{F^c}{F^c_\beta}(x)\,e^{\beta x-\Lambda(\beta)}\,f_\beta(t+x) \\
&= \frac{f_\beta(x+t)}{F^c_\beta(x)} \;=\; (f_\beta)_{+x}(t) \qquad \square
\end{aligned}
$$

Claim 7.4.3. *If* $\lim_{x\to\infty} h(x) = \infty$ *then* $\lim_{x\to\infty} \mathbb{E}[e^{\theta\tau_{+x}}] = 1$.

We generally have $e^{\tilde{\Lambda}(\theta)} = \frac{e^{\Lambda(\theta)}-1}{\Lambda'(\theta)\,\theta}$ and

$$\begin{aligned}
&\lim_{x\to\infty} \mathbb{E}[e^{\theta\widetilde{(\tau_{+x})}}] \\
&= \lim_{x\to\infty} \frac{\int_{s=x}^{\infty} e^{\theta s - H(s)}\,ds}{e^{\theta x}\int_{s=x}^{\infty} e^{-H(s)}\,ds} = \lim_{x\to\infty} \frac{-e^{\theta x - H(x)}}{\theta\, e^{\theta x}\int_{s=x}^{\infty} e^{-H(s)}\,ds - e^{\theta x - H(x)}} \\
&= \lim_{x\to\infty} \frac{-1}{\theta e^{H(x)}\int_{s=x}^{\infty} e^{-H(s)}\,ds - 1} = \lim_{x\to\infty} \frac{-1}{\theta\mathbb{E}[\tau_{+x}] - 1}
\end{aligned} \tag{7.6}$$

If $\lim_{x\to\infty} h(x) = \infty$ then $\lim_{x\to\infty} \mathbb{E}[\tau_{+x}] = 0$ and $\lim_{x\to\infty} \mathbb{E}[e^{\theta\widetilde{(\tau_{+x})}}] \overset{(7.6)}{=} 1$ for all $\theta \in \mathbb{R}$.

$$\begin{aligned}
\mathbb{E}[e^{\theta\tau_{+x}}] &= 1 + \mathbb{E}[\tau_{+x}]\theta\mathbb{E}[e^{\theta\widetilde{(\tau_{+x})}}] \\
\lim_{x\to\infty} \mathbb{E}[e^{\theta\tau_{+x}}] &= 1 + \lim_{x\to\infty} \mathbb{E}[\tau_{+x}]\theta\mathbb{E}[e^{\theta\widetilde{(\tau_{+x})}}] = 1 + 0\cdot\theta\cdot 1 = 1
\end{aligned}$$

□

7.5 Large deviations and other tools

Theorem 7.5.1 (Arzelà-Ascoli, theorem A.51 of [23]). *The set A has compact closure in $C([0,T],\mathbb{R}^d)$ equipped with the sup-norm if and only if*

- *The initial points are bounded:* $\sup_{\vec{x}\in A} |\vec{x}(0)| < \infty$*, and*
- *The functions in A are equicontinuous, that is, for every t and ϵ there exists a δ so that, whenever $|t-s| < \delta$ we have $|\vec{x}(t) - \vec{x}(s)| < \epsilon$ for all $\vec{x} \in A$.*

Theorem 7.5.2 (Contraction principle, theorem 4.2.1 of [5]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces and $\mathcal{X} \to \mathcal{Y}$ a continuous function. Consider a good rate function $I : \mathcal{X} \to [0,\infty]$.*

(a) For each $y \in \mathcal{Y}$, define

$$I'(y) = \inf\{I(x) \,:\, x \in \mathcal{X}, y = f(x)\}.$$

Then I' is a good rate function on $\mathcal{Y}$, where as usual the infimum over the empty set is taken as ∞.

(b) If I controls the LDP associated with a family of probability measures $\{\mu_\epsilon\}$ on $\mathcal{X}$, then I' controls the LDP associated with the family of probability measures $\{\mu_\epsilon \circ f^{-1}\}$ on $\mathcal{Y}$.

The following version of the Gärtner-Ellis theorem is equivalent to theorem 2.3.6 of [5].

Theorem 7.5.3 (Gärtner and Ellis). *Let* $(Z_n\,;\,n \in \mathbb{N})$ *be a sequence of random vectors in* $\mathbb{R}^d$ *with* μ_n *the law of* Z_n. *Assumue that the limit*

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\langle \lambda, Z_n \rangle}]$$

exists as an extended real number. Assume further that $0 \in \mathcal{D}(\Lambda)^\circ$.
(a) For any closed set F,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq - \inf_{x\in F} \Lambda^*(x).$$

(b) For any open set G,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(F) \geq - \inf_{x\in G\cap\mathcal{F}} \Lambda^*(x),$$

where $\mathcal{F}$ *is the set of exposed points of* Λ^* *whose exposing hyperplane belongs to* $\mathcal{D}(\Lambda)^\circ$.
(c) If Λ *is an essentially smooth, lower semicontinuous function, then a large deviation principle holds for* $(Z_n\,;\,n \in \mathbb{N})$ *with the good rate function* $\Lambda^*(\cdot)$.

7.6 Assumptions

Chapter 2
2.2.2: Inter event times a.s. strictly positive. Lmgfs open and not bounded from below.
2.2.13: Existence of density, no harsh used better than new.
2.4.2 : Existence of positive Cesaro mean for the hazard rate.
Chapter 5
5.0.2: Summarises previous assumptions.
5.1.9: Networks are open and have no immediate feedback.

Bibliography

[1] Abraham Berman and Robert J. Plemmons. *Non-negative matrices in the mathematical sciences.* Academic Press, 1979.

[2] Dimitri P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1995.

[3] Hong Chen and Avishai Mandelbaum. Discrete flow networks: Bottleneck analysis and fluid approximations. *Mathematics of Operations Research*, 16:408–446, 1991.

[4] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes.* Springer, 1988.

[5] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications.* Springer, 1993.

[6] Nicholas G. Duffield and Neil O'Connell. Large deviations and overflow probabilities for the general single-server queue with applications. *Math. Proc. Cam. Phil. Soc.*, 1995.

[7] Paul Dupuis and Richard S. Ellis. The large deviation principle for a general class of queueing systems. *Transactions of the American Mathematical Society*, 347(8), 1995.

[8] Robert Foley and David McDonald. Large deviations of a modified jackson network: Stability and rough asymptotics. *The Annals of Applied Probability*, 15, 2005.

[9] Ayalvadi Ganesh, Neil O'Connell, and Damon Wischik. *Big Queues.* Springer, 2004.

[10] Peter Glynn and Ward Whitt. Large deviations behaviour of counting processes and their inverses. *Queueing Systems*, 17:107–128, 1994.

[11] J.B. Goodman and William Massey. The non-ergodic jackson-network. *Journal of Applied Probability*, 21:860–869, 1984.

[12] Irina Ignatiouk-Robert. Large deviations of jackson networks. *Annals of Applied Probability*, 10:962–1001, 2000.

[13] Irina Ignatiouk-Robert. Large deviations for processes with discontinuous statistics. *Annals of Probability*, 33:1479–1508, 2005.

[14] James R. Jackson. Networks of waiting lines. *Operations Research*, 5:518–521, 1957.

[15] Anatolii Puhalskii. Large deviation analysis of the single server queue. *Queueing Systems*, 21:5–66, 1995.

[16] Anatolii Puhalskii. The action functional for the jackson network. *Markov Processes and Related Fields*, 13:99–136, 2007.

[17] Anatolii Puhalskii and Ward Whitt. Functional large deviation principles for first-passage-time processes. *The Annals of Applied Probability*, 7(2):362–381, 1997.

[18] Philippe Robert. *Stochastic Networks and Queues*. Springer, 2003.

[19] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[20] Mark Rodgers-Lee. The large deviations of random time-changes in a metric topology, 2003. Master Thesis, University of Dublin.

[21] H. L. Royden. *Real Analysis*. The Macmillan Company, 1968.

[22] Raymond Russell. The large deviations of random timechanges, 1998. PhD thesis, Trinity College Dublin.

[23] Adam Shwartz and Alan Weiss. *Large Deviation for Performance Analysis*. Chapman and Hall, 1995.

[24] Hermann Thorisson. *Coupling, Stationarity, and Regeneration*. Springer, 2000.

[25] James S. Vandergraft. A fluid flow model of networks of queues. *Management Science*, 29(10), 1983.

[26] Ronald W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice-Hall, 1989.