Large Deviations and Transient Multiplexing at a Buffered Resource

Ken Duffy and Fergal Toomey *

April 7, 2000

Abstract

In this paper we discuss asymptotics associated with a large number of sources using a resource in a compact time interval. A large deviations condition is placed on the sum of the vectors that describe the stochastic behaviour of the sources and large deviations results deduced about the probability of exhaustion of the resource. This approach allows us to consider sources which are highly non-stationary in time. The examples in mind are a single server queue and a form of the Cramer-Lundburg model from risk theory. Connection is made with past work on stability of queues and effective bandwidths. A number of examples are presented to illustrate the strengths of this approach.

*Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Ireland.
K.D's work is supported by a grant from Microsoft Research, Cambridge

1 Introduction

Large deviation asymptotics have been applied in many areas to calculate probabilities of exhaustion of a system resource as the size of the system becomes large. This body of work includes large buffer asymptotics in queueing theory with linear scalings such as Glynn and Whitt [13], and with non-linear scalings such as Duffield and O'Connell [8]; large initial capital asymptotics in risk theory such as Martin-Lof [20] and Nyrhinen [22] and references therein; and also large number of lines asymptotics in queueing theory with linear number of lines/linear time scaling such as Botvitch and Duffield [3], Courcoubetis and Weber [5] and with non-linear time scalings in Duffield [7].

The work on large number of lines asymptotics aims to capture the effect of statistical multiplexing, see Kelly [16] for a general reference. This is where superpositions of bursty traffic lead to a smoother, less bursty, multiplex and hence to economies of scale. The approach taken by other authors in their work on large number of lines asymptotics is to work with homogeneous and heterogeneous superpositions of traffic whose workload processes display long time-scale large deviations behaviour. This connects their work with the work on large buffer asymptotics. Specifically Duffield [7] finds that in the case where the large deviation behaviour of the underlying sources is on a non-linear time scale, economies of scale can be much greater than in the linear case.

We will not approach the problem in this manner. We assume large deviation behaviour on non-linear scalings in the number of lines and consider only compact time intervals. This allows us to consider processes which are highly non-stationary in time and still deduce large deviation results. We have two applications in mind, one is the single server queue with many sources each of whom has an associated arrival and service process; the resource in question being the buffer space at the queue. The other application is to a version of the Cramer-Lundburg model in risk theory (for a general introduction to risk theory see Grandell [14]), here the resource is an insurance companies capital and the sources are the clients who make claims and pay premiums.

In section two we introduce the setup under which the results will be developed. In section three the main results are presented. In section four the notion of stability which was introduced in the early queueing literature (see Loynes [19]) is revisited in the current setting. In section five the results are considered in the presence of convex structure and a connection is made with work on effective bandwidths (see Kelly [16]). In section six references to the literature are given to conditions under which the main assumption of this work is true. In section seven we present a range of examples to highlight the features of the approach adopted in this paper.

The first example illustrates how even in a simple i.i.d. setting the appropriate large deviation scale may be non-linear. The second example displays many-scale behaviour; the scale on which the large deviations are seen depend upon the size of the deviation. The third example is designed to show how the asymptotics of the two models, the single server queue and the Cramer-Lundburg model, can differ. It too displays many-scale behaviour.

2 Setup

We consider L sources using a resource in the discrete time interval $[1, T]$ where T is finite. Ultimately we will take the limit as the number of sources using the resource becomes large. Each source i has an associated vector $X(i)$ in $\mathbb{R}^T$ which is almost surely finite. $X_t(i)$ describes the input imparted to the resource at time t by source i. We define the vector $Z(L) := X(1) + \cdots + X(L) \in \mathbb{R}^T$. $Z_t(L)$ is the total input imparted to the resource at time t from the first L sources. We consider two cases.

(I) The resource can not grow with time. This is the case with queues where $X_t(i)$ represents the arrivals less service from source i at time t and the resource is the buffer space at the queue.

(II) The resource can grow with time. This is the case with the Cramer-Lundburg model of risk theory where $X_t(i)$ represents the claims made less premium received from customer i by an insurance company at time t. Here the resource is the company's capital which acts as a buffer to bankruptcy.

We define a scale to be a non-decreasing sequence of real numbers diverging to infinity.

We assume that the size of the resource, $\{B(L)\}$, grows as the number of sources attached to the resource grows. That is, there exists a scale $a(L)$ such that the size of the resource is given for $b > 0$ by $B(L) = a(L)b$.

We keep b free as we shall be interested in what effect b, the resource space-scale, plays in the multiplexing.

We only consider discrete time although it is possible to deal with continuous time if a growth condition is placed on $X_t(\cdot)$ as a function of t such that for any compact interval $[1,T]$ it is sufficient to know the processes at some finite number of times within the interval, this sort of approach is taken by Duffield and O'Connell in [8] for large buffer asymptotics.

In case (I) the resource experiences overflow in the interval $[1,T]$ if the total arrivals less total service imparted to the resource by the sources exceeds $a(L)b$ in any interval $[r,s] \subseteq [1,T]$. That is if

$$\max_{1\le r\le s\le T} \sum_{t=r}^{s} Z_t(L) > a(L)b.$$

We define the maximum queue length operator

$$Q(\vec{y}) = \max_{1\le r\le s\le T} \sum_{t=r}^{s} y_t$$

so that overflow occurs if and only if

$$Q\left(\frac{Z(L)}{a(L)}\right) > b.$$

In case (II) the resource is exhausted in the interval $[1,T]$ if the total claims less total premium imparted to the resource by the sources exceeds $a(L)b$ in any interval $[1,s] \subseteq [1,T]$. That is if

$$\max_{1\le s\le T} \sum_{t=1}^{s} Z_t(L) > a(L)b.$$

We define the worst negative bank balance operator

$$C(\vec{y}) = \max_{1 \le s \le T} \sum_{t=r}^{s} y_t$$

so that bankruptcy occurs if and only if

$$C\left(\frac{Z(L)}{a(L)}\right) > b.$$

We will calculate large deviation probabilities for these events in terms of an assumed underlying large deviations principle for $\{Z(L)\}$ as the number of sources using the resource tends towards infinity.

3 Main results

The proofs for the results for cases (I) and (II) are almost identical, so although the results are stated in terms of both they are only proved in (the more general) case (I).

Assumption one: $\{Z(L)/a(L)\}$ satisfies a large deviation principle on the external scale $\{V(L)\}$ with rate function $I(\cdot)$ which is continuous on the interior of the set upon which it is finite. That is, there exists a function $I : \mathbb{R}^T \to [0, \infty]$, which is continuous on the interior of the set upon which it is finite, such that for all F closed in $\mathbb{R}^T$

$$\limsup_{L \to \infty} \frac{1}{V(L)} \log \mathbb{P}\left[\frac{Z(L)}{a(L)} \in F\right] \le -\inf_{\vec{x} \in F} I(\vec{x})$$

and for all G open in $\mathbb{R}^T$

$$\liminf_{L \to \infty} \frac{1}{V(L)} \log \mathbb{P}\left[\frac{Z(L)}{a(L)} \in G\right] \ge -\inf_{\vec{x} \in G} I(\vec{x}).$$

References to the literature giving sufficient conditions for this assumption to hold shall be presented in a later section. For a superb review of large deviation theory see Lewis and Pfister [17].

Lemma 1 *For each pair* $[r,s]$ *such that* $1 \leq r \leq s \leq T$ *the sequence* $\{S_{r,s}(L)/a(L)\}$, *with* $S_{r,s}(L) := \sum_{t=r}^{s} Z_t(L)$, *satisfies a large deviation principle in* $\mathbb{R}$ *on the exterior scale* $\{V(L)\}$ *with rate function* $J_{r,s}(y)$ *defined by*

$$J_{r,s}(y) := \inf\left\{ I(\vec{x}) : \sum_{t=r}^{s} x_t = y \right\}.$$

PROOF As the function $\phi(x_r, \ldots, x_s) = x_r + \cdots + x_s$ is continuous it follows directly from the contraction principle (see Theorem 6.4 of Lewis and Pfister[17]) and assumption one that the image measures $\mathbb{M}'_L[B] := \mathbb{P}\left[\frac{Z(L)}{a(L)} \in \phi^{-1}(B)\right] = \mathbb{P}\left[\frac{S_{r,s}(L)}{a(L)} \in B\right]$ satisfy a large deviation principle in $\mathbb{R}$ on the external scale $V(L)$ with rate function

$$J_{r,s}(y) := \inf\left\{ I(\vec{x}) : \sum_{t=r}^{s} x_t = y \right\}.$$

$J_{r,s}(\cdot)$ is continuous as $I(\cdot)$ and $\phi(\cdot)$ are.

Theorem 2 $\{Q(Z(L)/a(L))\}$ *and* $\{C(Z(L)/a(L))\}$ *satisfy large deviation principles in* $\mathbb{R}$ *on the exterior scale* $\{V(L)\}$ *with rate functions* $I_Q(y)$ *and* $I_C(y)$, *respectively, which are defined by*

$$\begin{aligned} I_Q(y) &:= \min_{\{1 \leq r \leq s \leq T\}} J_{r,s}(y) \\ I_C(y) &:= \min_{\{1 \leq s \leq T\}} J_{1,s}(y). \end{aligned}$$

PROOF A simple application of the principle of the largest term (see lemma 2.3 of Lewis and Pfister [17]) suffices for the upper bound. We have that

$$Q(Z(L)/a(L)) = \max_{1\le r\le s\le T} \sum_{t=r}^{s} Z_t(L)/a(L).$$

For F closed

$$\max_{\{1\le r\le s\le T\}} \mathbb{P}\left[\sum_{t=r}^{s} \frac{Z_t(L)}{a(L)} \in F\right] \le \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) \in F\right]$$

and

$$\mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) \in F\right] \le T \max_{\{1\le r\le s\le T\}} \mathbb{P}\left[\sum_{t=r}^{s} \frac{Z_t(L)}{a(L)} \in F\right].$$

Taking limits and using the fact that for any finite collection of sequences $\{a_n^{(1)}\},\ldots,\{a_n^{(N)}\}$ in $[-\infty,\infty]$

$$\limsup_{n\to\infty} \max_{i\in\{1,\ldots,N\}} a_n^{(i)} = \max_{i\in\{1,\ldots,N\}} \limsup_{n\to\infty} a_n^{(i)}$$

it follows that

$$\limsup_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[Q\left(\tfrac{Z(L)}{a(L)}\right) \in F\right] =$$
$$\max_{\{1\le r\le s\le T\}} \limsup_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[\textstyle\sum_{t=r}^{s} \tfrac{Z_t(L)}{a(L)} \in F\right].$$

Hence

$$\limsup_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) \in F\right] \le -\inf_{x\in F}\left\{\min_{\{1\le r\le s\le T\}} J_{r,s}(x)\right\}.$$

Now for the lower bound. For G open we have that

$$\mathbb{P}\left[\sum_{t=r}^{s} Z_t(L)/a(L) \in G\right] \le \mathbb{P}[Q(Z(L)/a(L)) \in G]$$

for any $[r,s] \in [1,T]$ and thus

$$\liminf_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[\sum_{t=r}^{s} \frac{Z_t(L)}{a(L)} \in G\right] \le \liminf_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) \in G\right]$$

for any $[r,s] \in [1,T]$ and hence

$$\max_{\{1\le r\le s\le T\}} \liminf_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[\sum_{t=r}^{s} \tfrac{Z_t(L)}{a(L)} \in G\right] \le \liminf_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[Q\left(\tfrac{Z(L)}{a(L)}\right) \in G\right].$$

Thus for all G open

$$-\inf_{x\in G}\left\{\min_{\{1\le r\le s\le T\}} J_{r,s}(x)\right\} \le \liminf_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) \in G\right].$$

$I_Q(\cdot)$ and $I_C(\cdot)$ are clearly continuous as for all $[r,s] \in [1,T]$ we know $J_{r,s}(\cdot)$ is continuous and the minimum or maximum of a finite collection of continuous functions is continuous.

Corollary 3 *The rate of decay of the probability of overflow satisfies*

$$\lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) > b\right] = -\inf_{x>b} I_Q(x).$$

The rate of decay of the probability of bankruptcy satisfies

$$\lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[C\left(\frac{Z(L)}{a(L)}\right) > b\right] = -\inf_{x>b} I_C(x).$$

PROOF Using the large deviation principle and the fact that for all a $\mathbb{P}[Q(\vec{y}) > a] \le \mathbb{P}[Q(\vec{y}) \ge a]$, we have that

$$\begin{aligned}
-\inf_{x>b} I_Q(x) &\le \liminf_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[Q\left(\tfrac{Z(L)}{a(L)}\right) > b\right] \\
&\le \limsup_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[Q\left(\tfrac{Z(L)}{a(L)}\right) > b\right] \\
&\le \limsup_{L\to\infty} \tfrac{1}{V(L)} \log \mathbb{P}\left[Q\left(\tfrac{Z(L)}{a(L)}\right) \ge b\right] \\
&\le -\inf_{x\ge b} I_Q(x) \\
&= -\inf_{x>b} I_Q(x),
\end{aligned}$$

where the last line uses the continuity of $I_Q(\cdot)$. Hence

$$\lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) > b\right] = -\inf_{x>b} I_Q(x).$$

4 Stability

Traditionally with time stationary results for queues (see Loynes [19]) a queue is said to be stable if the mean arrival rate is less than the mean service rate. In this case there exists a minimum stationary sequence of random variables that satisfy the queueing recursion (Lindley's equation) and under a regularity condition every other solution of Lindley's equation couples to the stationary one in almost surely finite time.

Assumption two: For each $[r,s] \subseteq [1,T]$ we have

$$\lim_{L\to\infty} \frac{S_{r,s}(L)}{a(L)} = M_{r,s} \in \mathbb{R}$$

in probability.

If $X(i)$ is stationary then with $a(L) = L$, assumption two corresponds to a weak law of large numbers.

Lemma 4 *If $M_{r,s} \geq b$ for any $[r,s] \subseteq [1,T]$ then the probability of overflow does not decay exponentially on the scale $V(L)$. If $M_{1,s} \geq b$ for any $s \in [1,T]$ then the probability of bankruptcy does not decay exponentially on the scale $V(L)$.*

PROOF As $M_{r,s} \geq b$ for some $[r,s] \subseteq [1,T]$ we have that

$$\inf_{x>b} J_{r,s}(x) = J_{r,s}(M_{r,s}) = 0.$$

By corollary one

$$\lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) > b\right] = -\inf_{x>b} I_Q(x)$$

so that

$$\lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) > b\right] = -\inf_{x>b}\left(\min_{\{0\le r\le s\le T\}} J_{r,s}(x)\right) = 0.$$

Similarly for the rate of decay of the probability of bankruptcy.

Assumption 3: For each $[r,s] \subseteq [1,T]$ we have that $J_{r,s}(x)$ is non-decreasing to the right of $M_{r,s}$. Furthermore we assume there exists $\overline{m}_{r,s} \geq M_{r,s}$ such that for all $x > \overline{m}_{r,s}$ we know that $J_{r,s}(x) > 0$. We call this the monotone property.

We note that if $I(\cdot)$ is strictly convex then $J_{r,s}(\cdot)$ is also strictly convex and hence satisfies the *monotone property* with $\overline{m}_{r,s} = M_{r,s}$.

If one proves the joint LDP in assumption one via the Gartner [12]–Ellis [10] theorem then $I(\cdot)$ is automatically strictly convex. Other simple examples of where the *monotone property* holds but the underlying rate function is not strictly convex come from models in statistical mechanics where flat spots in the rate function around the mean correspond to phase transitions. We will provide an example adapted from Gantert [11] where the underlying rate function is concave but still satisfies the *monotone property* with $\overline{m}_{r,s} = M_{r,s}$.

Lemma 5 *If $\overline{m}_{r,s} < b$ for all $[r,s] \subseteq [1,T]$ then the probability of overflow decays exponentially on the scale $V(L)$. If $\overline{m}_{1,s} < b$ for all $s \in [1,T]$ then the probability of bankruptcy decays exponentially on the scale $V(L)$.*

PROOF In this case we know that for all $[r,s] \subseteq [1,T]$

$$\inf_{x>b} J_{r,s}(x) = J_{r,s}(b) > 0.$$

Hence

$$\lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{P}\left[Q\left(\frac{Z(L)}{a(L)}\right) > b\right] = -\inf_{x>b}\left(\min_{\{0\le r\le s\le T\}} J_{r,s}(x)\right) > 0.$$

We note that if $X_t(i)$ is stationary in both t and i then

$$M_{r,s} = (s-r+1)\mathbb{E}[X_1(1)]$$

and hence, taking $r=1, s=T$ we see $M_{1,T} = T\mathbb{E}[X_1(1)]$. Letting T be large we end up with the usual condition, $\mathbb{E}[X_1(1)] < 0$, for a stable queue (or insurance company) to exist.

5 Convexity and Effective Bandwidths

In the presence of convex structure large deviation theory becomes substantially more powerful. It becomes possible to deal with the Legendre-Fenchel transform of the rate function, the scaled cumulant generating function (sCGF) $\lambda(\theta)$

$$I(x) = \sup_{\theta}\{x\theta - \lambda(\theta)\}.$$

For a general reference to convex functions see Rockafellar [23] (specifically on convex conjugates section 12). Often large deviation principles are proved under conditions on the sCGF, the Gartner [12]–Ellis [10] conditions, which ensure that not only does the rate function exist but also that it is strictly convex. There is however another way to use the convex structure. If we know a large deviation principle holds with a convex (but not necessarily strictly convex) rate function and that the sCGF exists in a neighbourhood of the origin then we know that they are dual to each other (see Theorem 7.1 [17]).

Assumption four: The rate function $I(\cdot)$ is convex and

$$\lambda\left(\vec{\theta}\right) := \lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{E}\left[\exp\left(\frac{V(L)}{a(L)}\langle\vec{\theta}, Z(L)\rangle\right)\right]$$

exists as an extended real number for all $\vec{\theta} \in \mathbb{R}^T$ and is finite in an open ball containing the origin.

By Theorem 7.1 [17] $I(\cdot)$ and $\lambda(\cdot)$ are convex duals.

For each pair $[r,s] \subseteq [1,T]$ define $\lambda_{r,s}(\theta)$ for $\theta \in \mathbb{R}$ by

$$\lambda\left(\theta\right) := \lim_{L\to\infty} \frac{1}{V(L)} \log \mathbb{E}\left[\exp\left(\frac{V(L)}{a(L)}\theta S_{r,s}(L)\right)\right].$$

Lemma 6 *For all $[r,s] \subseteq [1,T]$ we have that $\lambda_{r,s}(\theta)$ exists as an extended real number, is finite in a neigbourhood of the origin and that $J_{r,s}(\cdot)$ and $\lambda_{r,s}(\cdot)$ are convex dual to each other. Moreover $I_Q(\cdot)$ and $I_C(\cdot)$ are convex functions and*

$$I_Q(x) = \min_{\{1\le r\le s\le T\}} \sup_\theta \{x\theta - \lambda_{r,s}(\theta)\}$$

and

$$I_C(x) = \min_{\{1\le s\le T\}} \sup_\theta \{x\theta - \lambda_{1,s}(\theta)\}$$

PROOF We know that $\lambda_{r,s}(\theta)$ exists as an extended real number for all θ as $S_{r,s}(L)$ is a linear function of $Z(L)$. $\lambda_{r,s}(\cdot)$ is finite in a neigbourhood of the origin as $\lambda(\cdot)$ is. $I(\cdot)$ being convex implies that $J_{r,s}(\cdot)$ is convex thus, using Theorem 7.1 of [17], we see that $J_{r,s}(\cdot)$ and $\lambda_{r,s}(\cdot)$ are convex dual to each other. Thus for each $[r,s] \subseteq [1,T]$

$$J_{r,s}(x) = \sup_{\theta}\{x\theta - \lambda_{r,s}(\theta)\}.$$

Hence $I_Q(\cdot)$ is given by

$$I_Q(x) = \min_{\{1 \le r \le s \le T\}} \sup_{\theta}\{x\theta - \lambda_{r,s}(\theta)\}.$$

Now for the connection to Effective Bandwidths. For a thorough review of effective bandwidth functions see Kelly [16]. The notion of Effective Bandwidth has become widely accepted as a measure of the resource requirements of traffic in a queueing network. Only linear scalings are dealt with in this context.

Consider a queue with fixed buffer-space per source and fixed service-rate per source being driven by independent sources. The stochastic behaviour of the queue length arises from randomness within the sources. The objective is to find a service rate per source per unit time which ensures that the probability of overflow is below some prescribed threshold.

For the rest of this section $V(L) = L$, $a(L) = L$ and $X_t(i) := Y_t(i) - c$, where the $Y_\cdot(\cdot)$ are almost surely non-negative, and $c > 0$ is the capacity of the resource per source per unit time. For each $[r,s] \subseteq [1,T]$ define $S^Y_{r,s}(L) := \sum_{t=r}^{s} Y_t(i)$.

Lemma 7 *If $\{Y_t(i)\}$ is stationary in t, and if for each $i \neq j$ $Y(i)$ and $Y(j)$ are mutually independent then*

$$\lambda(\theta) = \log \mathbb{E}\left[\exp\left(\theta S^Y_{1,s-r+1}(1)\right)\right] - (s-r+1)\theta c.$$

PROOF

$$\begin{aligned}
\lambda(\theta) &= \lim_{L\to\infty} \tfrac{1}{L} \log \mathbb{E}\left[\exp\left(\theta S_{r,s}(L)\right)\right] \\
&= \lim_{L\to\infty} \tfrac{1}{L} \log \mathbb{E}\left[\exp\left(\theta S^Y_{r,s}(L)\right)\right] - (s-r+1)\theta c \\
&= \lim_{L\to\infty} \tfrac{1}{L} \log \mathbb{E}\left[\exp\left(\theta S^Y_{1,s-r+1}(1)\right)\right]^L - (s-r+1)\theta c \\
&= \log \mathbb{E}\left[\exp\left(\theta S^Y_{1,s-r+1}(1)\right)\right] - (s-r+1)\theta c.
\end{aligned}$$

The function

$$\alpha(\theta, s) := \frac{1}{\theta s} \log \mathbb{E}[\exp(\theta S^Y_{1,s}(1))]$$

is the *effective bandwidth* of a source (see Kelly [16]). Then, in this setting

$$\lambda_{r,s}(\theta) = (s-r)\theta\alpha(\theta, s-r) - (s-r+1)\theta c.$$

Therefore in this scenario we have that the probability of overflow and the probability of bankruptcy both satisfy

$$I_Q(b) = I_C(b) = \min_{0\leq s\leq T} \sup_{\theta} \left\{\theta b - s\theta\alpha(\theta, s-r) + (s-r+1)\theta c\right\}.$$

For an article on measuring effective bandwidths and its uses see Györfi *et al.* [15] and references therein.

6 Joint Large Deviation Principles

The main underlying assumption (assumption one) has been the existence of a joint large deviation principle for the underlying sources, that is, a large deviation principle for their vectors. For a general reference to large deviation techniques see Dembo and Zeitouni [6]. Here we will try to illustrate the conditions under which this holds. Essentially we expect that it will hold when for each $t \in [1, T]$ the partial sums of $X_t(\cdot)$ satisfy large deviation principles.

It is difficult to pin down a simple set of general mixing conditions for which the assumption is true. Ellis [10] proved a large deviation principle for random vectors under what would ultimately be called the Gartner-Ellis conditions. These conditions essentially amount to assuming the sCGF exists, is finite in a neighbourhood of the origin and has steepness properties. There are plently of examples where large deviation principles hold but these conditions are not true; see section 2.3 of Dembo and Zeitouni [6] for ones motivated from applied probability. In statistical mechanics any Ising model which allows a phase transition fails the steepness property. In heavy tailed distributions of risk theory (see for example T. Mikosch and A. V. Nagaev [21]) the sCGF's are not finite in the neighbourhood of the origin.

There are other general mixing conditions under which large deviation principles can be deduced. See Bryc and Dembo [4] for conditions motivated by mixing conditions under which central limit theorems are proved and Lewis *et al.* [18] for a condition motivated by Gibbs measures of statistical mechanics.

It is certainly trivially true that if for each $t \in [1, T]$ the partial sums of $X_t(\cdot)$ satisfy large deviation principles on the scales $a(\cdot)$ and $V(\cdot)$ with rate functions $I_t(\cdot)$; and if $X_r(\cdot)$ and $X_s(\cdot)$ are independent for all $r \neq s$ then the partial sums of the $X(\cdot)$ satisfy a large deviation principle on the scales $a(\cdot)$ and $V(\cdot)$ with rate function $I(\vec{x}) = I_1(x_1) + \cdots + I_T(x_T)$.

7 Examples

7.1 Example 1

This example presents a simple set of independent sources consisting of i.i.d. random variables which satisfy a large deviation principle on a non-linear scale.

We consider fixed capacity (or premium) $c > 0$ per source per unit time and model the arrivals (or claims) process.

The behaviour is described by a heavy tailed i.i.d. sequence adapted from Gantert [11]. Gantert has extended these results to include dependent random variables satisfying a mixing condition along the lines of of that found in Bryc and Dembo [4]. Heavy tailed distributions are often studied in risk theory (see Mikosch and Nagaev [21] and references therein) and more recently in queueing theory (see Asmussen and

Collamore [1]). For these models it is possible to get much more information than is available on a logarithmic scale. Interesting features are still displayed even after taking logs.

For each $t \in [1, T]$, we define the sequence $\{X_t(L)\}$ to be an independently and identically distributed sequence of random variables each with distribution equal to that of $Y_1(1) - c$ where

$$\mathbb{P}[Y_1(1) \geq x] := d(x) \exp(-E(x)x^z),$$

$z \in (0, 1)$ and $d(\cdot)$ and $E(\cdot)$ are slowly varying, that is for all $\eta \in (0, \infty)$

$$\lim_{x \to \infty} \frac{d(\eta x)}{d(x)} = \lim_{x \to \infty} \frac{E(\eta x)}{E(x)} = 1$$

(see Bingham *et al.* [2]). Define $M_{t,t} := \mathbb{E}[Y_t(1)] - c$. This example belongs to the class of semi-exponential distributions where all moments are finite but the cumulant generating function $\lambda(\theta)$ is infinite for all $\theta > 0$.

The sCGF for $X_t(\cdot)$ is not finite in a neighbourhood of the origin so the Gartner-Ellis conditions are not satisfied. It is possible to show using a sub-additivity argument that a large deviation principle is satisfied on the internal scale $a(L) = L$ and external scale $V(L) = L$. The rate function however is trivial; it is zero above it's mean and infinity below it. The scales on which the rate function is non-trivial are $a(L) = L$ and $V(L) = E(L)L^z$. On these scales

$$J_{t,t}(x) = \begin{cases} \infty & x < M_{t,t} \\ (x - M_{t,t})^z & x \geq M_{t,t}. \end{cases}$$

Note as $z \in (0, 1)$ that $J_{t,t}(\cdot)$ is concave hence it is not surprising it fails the Gartner-Ellis conditions. We note also that in this case $\overline{m}_{t,t}$

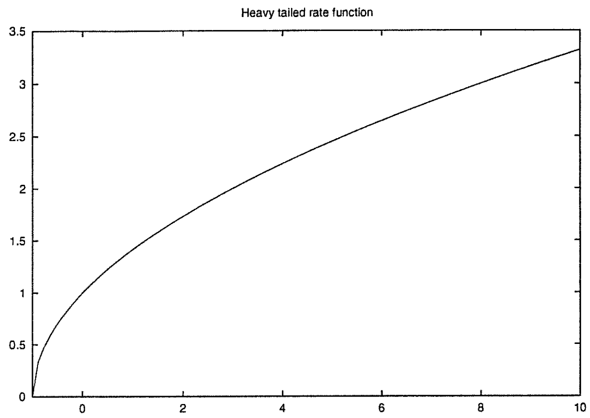

Figure 1: $J_{t,t}(y)$ vs. y for a heavy tailed r.v.'s on the scale $V(L) = \sqrt{L}$.

as defined in section four can be set equal to $M_{t,t}$. For example if $E(L) = d(L) = 1$, $c = 3$ and $z = 1/2$, then $M_{t,t} = -1$. See figure 1 for a graph of $J_{t,t}(\cdot)$.

By the comments at the end of the last section, on the scale $V(L) = E(L)L^z$ the partial sums of $X(\cdot)$ satisfy a large deviation principle with rate function, $I(\cdot)$ given by

$$I(\vec{x}) = \begin{cases} \infty & \text{if } x_t < M_{t,t} \quad \text{for any } t \in [1,T] \\ \sum_{t=1}^{T}(x_t - M_{t,t})^z & \text{if } x_t \geq M_{t,t} \quad \text{for all } t \in [1,T]. \end{cases}$$

For each $[r,s] \subseteq [1,T]$ by lemma one,

$$\begin{aligned} J_{r,s}(y) &= \inf\{I(\vec{x}) : \textstyle\sum_{t=r}^{s} x_t = y\} \\ &= \inf\{\textstyle\sum_{t=r}^{s} J_{t,t}(x_t) : \sum_{t=r}^{s} x_t = y\} \end{aligned}$$

but as for each t, $J_{t,t}(\cdot)$ is concave and unbounded we have that

$$J_{t,t}(y) = \inf\left\{\sum_{t=r}^{s} J_{t,t}(x_t) : \sum_{t=r}^{s} x_t = y\right\}.$$

Therefore on the scale $V(L) = E(L)L^z$, by corollary one and lemma three the rate of decay of the probability of overflow is given by

$$\inf_{y>b} I_Q(y) = \inf_{y>b} \min_{\{1\le r\le s\le T\}} J_{r,s}(y) = \inf_{y>b} J_{t,t}(y) = \begin{cases} \infty & b < M_{t,t} \\ (b - M_{t,t})^z & b \ge M_{t,t}. \end{cases}$$

The rate of decay of the probability of bankruptcy is also

$$\inf_{y>b} I_C(y) = \inf_{y>b} \min_{\{1\le s\le T\}} J_{1,s}(y) = \inf_{y>b} J_{1,t}(y) = \begin{cases} \infty & b < M_{t,t} \\ (b - M_{t,t})^z & b \ge M_{t,t}. \end{cases}$$

7.2 Example 2

The purpose of this example is to illustrate how many-scale behaviour can arise. The scale on which large deviations are observed depends upon the size of the deviation.

Again we consider constant service rate (or premium), $c > 0$, per source per unit time. The stochastic driving force is a sequence of almost surely non-negative random vectors $Y(L)$ that describe the arrivals (or claims) process, that is $X_t(L) := Y_t(L) - c$. Consider time stationary sources which are made up of two independent parts. A large part, $U_t(L)$, which has no correlation across the sources, and a small part, $W_t(L)$, which is highly correlated accross the sources. $Y_t(L) := U_t(L) + W_t(L)$.

We model $U_t(L)$ by i.i.d. bernoulli random variables which take the values $\{0, A\}$ with $\mathbb{P}[U_1(1) = A] = p$ and $\mathbb{P}[U_1(1) = 0] = 1 - p$. The mean of $U_t(L)$ is $M^u_{t,t} := Ap$. The partial sums of $U_t(L)$ satisfy a large deviation principle on the scale $V(L) = L$ with rate function $J^u_{t,t}(\cdot)$. This

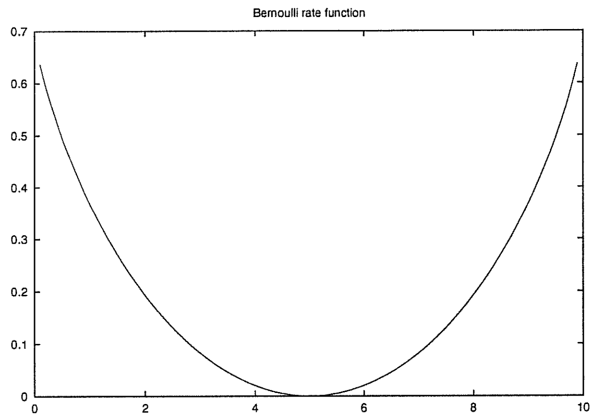

Figure 2: $J^u_{t,t}(y)$ vs. y for a bernoulli r.v.'s on the scale $V(L) = L$.

rate function is simple to calculate by means of the sCGF.

$$J^u_{t,t}(y) = \begin{cases} \left(1 - \frac{x}{A}\right) \log\left(\frac{a-A}{1-p}\right) + \frac{x}{A} \log \frac{x}{p} - \log A & \text{if } y \in [0, A] \\ \infty & \text{otherwise.} \end{cases}$$

For example, with $A = 10$ and $p = 1/2$ see figure 2 for a graph of $J^u_{t,t}(\cdot)$. On the scale $V(L) = \sqrt{L}$ the rate function for $U_t(\cdot)$ is trivial

$$K^u_{t,t}(y) = \begin{cases} 0 & \text{if } x = M^u_{t,t} \\ \infty & \text{otherwise.} \end{cases}$$

We model $W_t(L)$ as follows. Define the discrete heavy tailed distribution W by

$$\mathbb{P}[W \geq m] = e^{\sqrt{m}}$$

for $m \in \mathbb{Z}^+$, and define $\{W_t(\cdot)\}$ to be a stationary sequence of two state random variables taking values in $\{0, B\}$, whose sojourn 'times' spent in the 0 and B states are distributed by an i.i.d. sequence with distribution W. On the scale $V(L) = L$ the partial sums of $W_t(\cdot)$ satisfy a large deviation principle with a rate function, $J^w_{t,t}(\cdot)$, which is trivial,

$$J^w_{t,t}(x) = \begin{cases} 0 & \text{if } y \in [0, B] \\ \infty & \text{otherwise.} \end{cases}$$

In [24] Russell lays down a prescription to calculate the large deviations rate function for the partial sums of a two state source which can be described in terms of the sojourn times it spends in the 'on' and 'off' states. Under technical conditions, Russell proves that the large deviations of a randomly sampled partial sums process is a simple functional of the large deviations of partial sums process itself, and of the large deviations of the random sampling. Hence setting random sampling to occur at the end of sojourn times, one can calculate the rate function for the two state process by way of a functional of the rate function for the sojourn times.

In [9] Duffy uses this prescription to calculate the large deviations rate function, $K^w_{t,t}(\cdot)$, for the partial sums of $\{W_t(\cdot)\}$ on the scale $V(L) = \sqrt{L}$. $K^w_{t,t}(\cdot)$ is defined by

$$K^w_{t,t}(y) = \begin{cases} \left(1 - \frac{2y}{B}\right)^{1/2} & \text{if} \quad 0 \le y \le B/2 \\ \left(\frac{2y}{B} - 1\right)^{1/2} & \text{if} \quad B/2 \le y \le 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Note that $\{W_t(\cdot)\}$ has mean $M^w_{t,t} := B/2$ as it's sojourn times spent in the 'on' and 'off' states have finite and equal, expectation. For example if $B = 1$, see figure 3 for a graph of $K^w_{t,t}(\cdot)$.

On the scale $V(L) = L$ by the contraction principle,

$$J_{t,t}(y) = \inf\{J^u_{t,t}(y_1) + J^w_{t,t}(y_2) : y_1 + y_2 = y + c\}$$

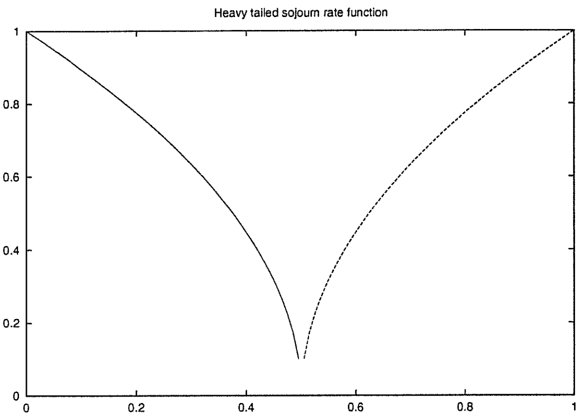

Figure 3: $K^{w}_{t,t}(y)$ vs. y for heavy tailed sojourn times on the scale $V(L) = \sqrt{L}$.

but $J^{w}_{t,t}(x) = 0$ for $x \in [0, B]$, thus

$$J_{t,t}(y) = \begin{cases} 0 & \text{if } y \in \left[0, B + M^{u}_{t,t} - c\right] \\ J^{u}_{t,t}(y + c - B) & \text{otherwise.} \end{cases}$$

As $J^{u}_{t,t}(\cdot)$ is convex and the sources are time independent

$$\begin{aligned} J_{r,s}(y) &= \inf\{\textstyle\sum_{i=r}^{s} J_{i,i}(y_i) : \sum_{i=r}^{s} y_i = y\} \\ &= (s-r) J_{t,t}\left(\tfrac{y}{s-r}\right) \\ &= \begin{cases} 0 & \text{if } y \in [0, (s-r)(B + M^{u}_{t,t} - c)] \\ (s-r) J^{u}_{t,t}\left(\tfrac{y}{s-r} + c - B\right) & \text{otherwise.} \end{cases} \end{aligned}$$

For each r, s we have $J^{u}_{r,r}(\cdot) = J^{u}_{s,s}(\cdot)$ and we know that $J^{u}_{r,r}(\cdot)$ is convex, hence for $y > 0$

$$J_{r,s}(y) = (s-r) J_{t,t}\left(\frac{y}{s-r}\right) \le J_{t,t}(y).$$

Thus on the scale $V(L) = L$, by corollary one and lemma three, the rate of decay of the probability of overflow is given by

$$\inf_{y>b} I_Q(y) = \inf_{y>b} \min_{\{1 \le r \le s \le T\}} J_{r,s}(y) = \inf_{y>b} J_{1,T}(y) = J_{1,T}(b).$$

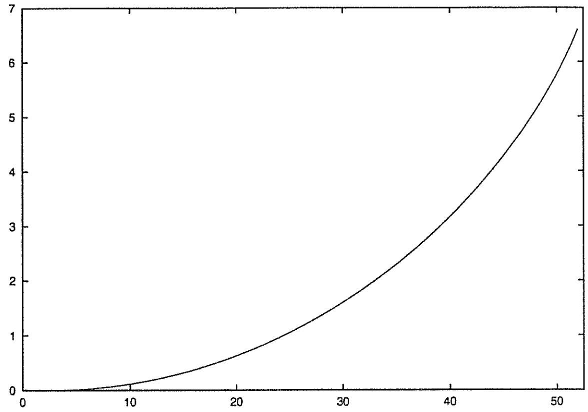

Figure 4: $\inf_{x>b} I_Q(x)$ vs. b on the scale L.

The rate of decay of the probability of bankruptcy is also

$$\inf_{y>b} I_C(y) = \inf_{y>b} \min_{\{1\le s\le T\}} J_{1,s}(y) = \inf_{y>b} J_{1,T}(y) = J_{1,T}(b).$$

For example if $A = 10$, $p = 1/2$, $B = 1$, $c = 5\frac{3}{4}$ and $T = 11$, see figure 4 for a graph of $\inf_{x>b} I_Q(x)$ vs. b on the scale $V(L) = L$. Note that the rate function is zero for b below $(s-r)(B+M^u_{t,t}-c) = 2\frac{1}{2}$. On this scale exhaustion of the resource is a concentration set if the resource space scale is below $2\frac{1}{2}$.

By the contraction principle on the scale $V(L) = \sqrt{L}$,

$$J_{t,t}(y) = \inf\{K^u_{t,t}(y_1) + K^w_{t,t}(y_2) : y_1 + y_2 = y + c\},$$

but $K^u_{t,t}$ is infinite except at $M^u_{t,t}$, where it is zero, therefore

$$J_{t,t}(y) = K^w_{t,t}(y + c - M^u_{t,t}).$$

Note that $J_{t,t}(y)$ is infinite for $y > B + M^u_{t,t} - c$ as $K^w_{t,t}(x)$ is infinite for $x > B$.

By the comments at the end of the last section

$$J_{r,s}(y) = \inf\left\{\sum_{i=r}^{s} J_{i,i}(y_i) : \sum_{i=r}^{s} y_i = y\right\}.$$

However as $K^w_{t,t}$ is concave

$$\begin{aligned} J_{r,s}(y) &= \left\lfloor \tfrac{y}{B+M^u_{t,t}-c} \right\rfloor J_{t,t}(B+M^u_{t,t}-c) \\ &\quad + J_{t,t}\left(y - \left\lfloor \tfrac{y}{B+M^u_{t,t}-c} \right\rfloor (B+M^u_{t,t}-c)\right), \end{aligned}$$

and as $J_{t,t}(B+M^u_{t,t}-c) = K^w_{t,t}(B) = 1$, we have that

$$J_{r,s}(y) = \left\lfloor \frac{y}{B+M^u_{t,t}-c} \right\rfloor + J_{t,t}\left(y - \left\lfloor \frac{y}{B+M^u_{t,t}-c} \right\rfloor (B+M^u_{t,t}-c)\right).$$

$J_{r,s}(\cdot) \leq J_{1,T}(\cdot)$ as for all $y \leq (s-r)(B+M^u_{t,t}-c)$, $J_{r,s}(y) = J_{1,T}(y)$ and $J_{r,s}(y)$ is infinity elsewhere.

If $y > (s-r)(B+M^u_{t,t}-c)$ then it is not possible for the sources to cause a large deviation on this scale and hence $J_{r,s} = \infty$. As $K^w_{t,t}(\cdot)$ is concave we have the following structure, the minimum amount of time is used to cause the deviation. If it is possible for one time instance to cause all of the deviation then due to the concavity this has a lower rate than sharing the deviation over time. This is in stark contrast to the convex case where the deviation is shared equally over time.

Thus on the scale $V(L) = \sqrt{L}$, by corollary one and lemma three, the rate of decay of the probability of overflow is given by

$$\inf_{y>b} I_Q(y) = \inf_{y>b} \min_{\{1\leq r\leq s\leq T\}} J_{r,s}(y) = \inf_{y>b} J_{1,T}(y) = J_{1,T}(b)$$

The rate of decay of the probability of bankruptcy is also

$$\inf_{y>b} I_C(y) = \inf_{y>b} \min_{\{1\leq s\leq T\}} J_{1,s}(y) = \inf_{y>b} J_{1,T}(y) = J_{1,T}(b).$$

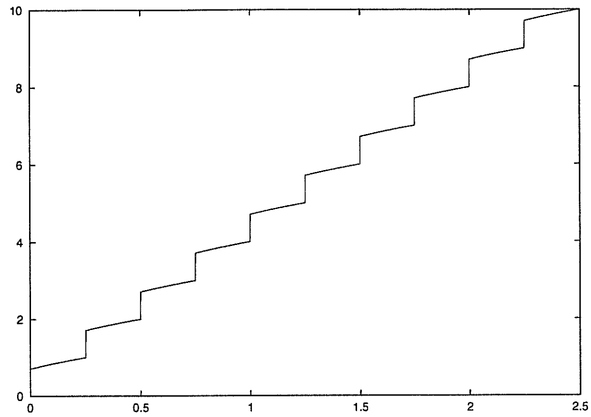

Figure 5: $\inf_{x>b} I_Q(x)$ vs. b on the scale $\sqrt{L}$.

For example if $A = 10$, $p = 1/2$, $B = 1$, $c = 5\frac{3}{4}$ and $T = 11$, see figure 5 for a graph of $\inf_{x>b} I_Q(x)$ vs. b on the scale $V(L) = \sqrt{L}$. Note that the rate function is infinity for b above $(T-1)(B + M^u_{t,t} - c) = 2\frac{1}{2}$. On this scale exhaustion of the resource will not happen (in probability) if the resource space scale is above $2\frac{1}{2}$.

Hence if $A > B$ we observe many-scale dehaviour. The scale on which large deviations are observed depends upon the size of the deviation in question.

7.3 Example 3

The purpose of this example is to highlight the effect of non-stationarity. Many-scale behavior also appears. The rate of decay of probability of overflow and rate of decay of probability of bankruptcy differ on the slower scale.

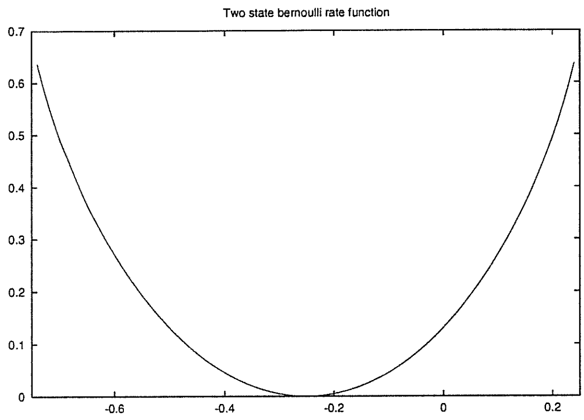

Figure 6: $J_{r,r}(y)$ vs. y.

We consider fixed service rate (or premium), c per source per unit time, and model the arrivals (or claims) process $\{Y_{\cdot}(\cdot)\}$, that is $X_t(L) := Y_t(L) - c$. At all times bar one the sources are uncorrelated. At one instant they are highly correlated.

$Y_t(L)$ will take one of two values, $\{0, A\}$, for all t and all L. Fix $t \in [1, T]$. At all times $r \neq t$ we model $Y_r(L)$ by bernoulli random variables taking the values $\{0, A\}$, with $\mathbb{P}[Y_r(L) = A] = p$ and $\mathbb{P}[Y_r(L) = 0] = 1 - p$. $M_{r,r} = Ap - c$. The partial sums of $X_r(L)$ satisfy a large deviation principle on with non-trivial rate function $J_{r,r}(\cdot)$ on the scale $V(L) = L$, where $J_{r,r}(y) = J^u_{t,t}(y + c)$ and $J^u_{t,t}(y)$ is defined in the previous example. With $c = 3/4$ and $A = 1$, see figure 6 for a graph of $J_{r,r}(y)$ vs. y. On the scale $V(L) = \sqrt{L}$ the rate function is trivial, it is zero at $M_{r,r}$ and infinity elsewhere.

At time t we model the source by the heavy tailed sojourn source described in the previous example, setting $B := A$. On the scale $V(L) = L$

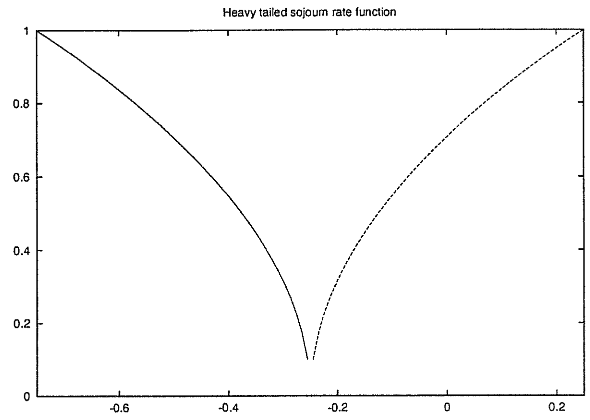

Figure 7: $J_{t,t}(y)$ vs. y.

the rate function for $X_t(\cdot)$ is trivial, it is zero in $[-c, A-c]$ and infinity elsewhere. On the scale $V(L) = \sqrt{L}$ the rate function is non-trivial, $J_{t,t}(y) = K^w_{t,t}(y+c)$, where $K^w_{t,t}(y)$ is defined in the previous example. With $c = 3/4$ and $B := A = 1$, see figure 7 for a graph of $J_{t,t}(y)$ vs. y.

If $t \notin [r,s]$ then on the scale $V(L) = L$,

$$\begin{aligned} J_{r,s}(y) &= \inf\{\textstyle\sum_{i=r}^{s} J_{i,i}(y_i) : \sum_{i=r}^{s} y_i = y\} \\ &= (s-r)J_{r,r}\left(\tfrac{y}{s-r}\right) \end{aligned}$$

as $J_{r,r}(\cdot)$ is convex and the sources are time independent. If $t \in [r,s]$ then on the scale $V(L) = L$,

$$J_{r,s}(y) = \begin{cases} 0 & \text{if } y \in [0, A-c] \\ (s-r)J_{r,r}\left(\frac{y-A+c}{s-r}\right) & \text{otherwise,} \end{cases}$$

as at t a deviation of $A-c$ has rate zero. Note that $J_{r,s}(y)$ is infinite if $y > (s-r+1)(A-c)$ as it is not possible for the sources to create a deviation that large. Thus, as $J_{r,r}(\cdot)$ is convex,

$$J_{1,T} \leq J_{r,s}(y)$$

for all $[r,s] \in [1,T]$. Thus on the scale $V(L) = L$, by corollary one and lemma three, the rate of decay of the probability of overflow is given by

$$\inf_{y>b} I_Q(y) = \inf_{y>b} \min_{\{1\leq r\leq s\leq T\}} J_{r,s}(y) = \inf_{y>b} J_{1,T}(y) = J_{1,T}(b).$$

The rate of decay of the probability of bankruptcy is also

$$\inf_{y>b} I_C(y) = \inf_{y>b} \min_{\{1\leq s\leq T\}} J_{1,s}(y) = \inf_{y>b} J_{1,T}(y) = J_{1,T}(b).$$

On the scale $V(L) = \sqrt{L}$ if $t \notin [r,s]$,

$$J_{r,s}(y) = \begin{cases} 0 & \text{if } y = (s-r)M_{r,r} \\ \infty & \text{otherwise.} \end{cases}$$

If $t \in [r,s]$ then,

$$J_{r,s}(y) = \begin{cases} J_{t,t}\left(y - (s-r-1)M_{r,r}\right) & \text{if } y \geq (s-r-1)M_{r,r} \\ \infty & \text{otherwise.} \end{cases}$$

If on average there is more arrivals than service for $r \neq t$ then $M_{r,r} = Ap - c < 0$; but $J_{t,t}(x) = \infty$ for $x > A - c$, hence if $(s-r-1) > (A-c)/(Ap-c)$ then $J_{r,s}(y) = \infty$.

Note that $J_{t,t}(\cdot) \leq J_{r,s}(\cdot)$ for all $[r,s]$ and $J_{1,t}(\cdot) \leq J_{1,s}(\cdot)$ for all s.

Thus on the scale $V(L) = \sqrt{L}$, by corollary one and lemma three, the rate of decay of the probability of overflow is given by

$$\inf_{y>b} I_Q(y) = \inf_{y>b} \min_{\{1\leq r\leq s\leq T\}} J_{r,s}(y) = \inf_{y>b} J_{t,t}(y) = J_{t,t}(b).$$

Note that $J_{t,t}(x)$ is infinite if $x > A-c$ thus this scale is only appropriate for deviations in the range $[0, A-c]$. The rate of decay of the probability of bankruptcy is given by

$$\inf_{y>b} I_C(y) = \inf_{y>b} \min_{\{1\leq s\leq T\}} J_{1,s}(y) = \inf_{y>b} J_{1,t}(y) = J_{1,t}(b).$$

However, if t is greater than $(A-c)/(Ap-c)$ then the heavytailed sojourn effect is not enough to cause a large deviation on this scale as it can not compensate for the downward pull of the eariler bernoulli effects. Hence no deviation would be seen on the slower scale.

References

[1] S. Asmussen and J. F. Collamore, *Exact asymptotics for a large deviations problem for the GI/G/1 queue*, Markov Processes and Related Fields **5** (1999), no. 4, 451–476.

[2] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, Cambridge University Press, 1987.

[3] D. D. Botvich and N. G. Duffield, *Large deviations, the shape of the loss curve, and economies of scale in large multiplexers*, Queueing Systems **20** (1996), 293–320.

[4] W. Bryc and A. Dembo, *Large deviations and strong mixing*, Annales de l'I.H.P. Probabilitiés and Statistiques **32** (1996), 549–569.

[5] C. Courcoubetis and R. Weber, *Buffer overflow asymptotics for a switch handling many traffic sources*, Journal of Applied Probability **33** (1996), 886–903.

[6] A. Dembo and O. Zeitouni, *Large deviation techniques and applications*, Springer, 1998.

[7] N. G. Duffield, *Economies of scale in queues with sources having power-law large deviation scalings*, Journal of Applied Probability **33** (1996), 840–857.

[8] N. G. Duffield and N. O'Connell, *Large deviations and overflow probabilities for the general single server queue, with applications*, Mathematical Proceedings of the Cambridge Philosophical Society **118** (1995), no. 2, 363–374.

[9] K. Duffy, *On the large deviations of a class of stationary on/off sources which exhibit the Hurst phenomenon*, Tech. report, DIAS, 2000.

[10] R. Ellis, *Large deviations for a general class of random vectors*, Annals of Probability **12** (1984), 1–12.

[11] N. Gantert, *Large deviations for a heavy tailed mixing sequence*, Preprint, Technische Universitat Berlin (1996).

[12] J. Gartner, *On large deviations from the invariant measure*, Theor Prob. Appl. **22** (1977), 24–39.

[13] P. Glynn and W. Whitt, *Large deviations behaviour of counting processes and their inverses*, Queueing Systems **17** (1994), 107–128.

[14] J. Grandell, *Aspects of risk theory*, Springer, Berlin, 1991.

[15] L. Györfi, A. Rácz, K. Duffy, J. T. Lewis, and F. Toomey, *Distribution-free confidence intervals for measurement of effective bandwidths*, Journal of Applied Probability **37** (2000), 1–12.

[16] F. P. Kelly, *Notes on effective bandwidths*, Stochastic Networks: Theory and Applications (F. P. Kelly, S. Zachary, and I. B. Ziedins, eds.), Oxford University Press, 1996, pp. 141–168.

[17] J. T. Lewis and C. E. Pfister, *Thermodynamic probability theory: some aspects of large deviations.*, Russian Mathematical Surveys **50** (1995), no. 2, 279–317.

[18] J. T. Lewis, C. E. Pfister, and W. G. Sullivan, *Entropy, concentration of probability and conditional limit theorems*, Markov Processes and Related Fields **1** (1995), 319–386.

[19] R. M. Loynes, *The stability of a queue with non-independent interarrival and service times*, Proceedings of the cambridge philosphical society **58** (1962), 497–520.

[20] A. Martin-Löf, *Entropy, a useful concept in risk theory.*, Scandinavian Actuarial Journal **3-4** (1986), 223–235.

[21] T. Mikosch and A. V. Nagaev, *Large deviations of heavy-tailed sums with applications in insurance*, Extremes **1** (1998), no. 1, 81–110.

[22] H. Nyrhinen, *Large deviations for the time of ruin.*, Journal of Applied Probability **36** (1999), 733–746.

[23] R. T. Rockafellar, *Convex analysis*, Princeton University Press, 1970.

[24] Raymond Russell, *The large deviations of random time changes*, Ph.D. thesis, Trinity College, Dublin, 1997.