Ph.D. Thesis

Buffer Engineering for M|G|infinity Input Processes

by Minothi Parulekar
Advisor: Armand Makowski

Ph.D. 2000-2

ABSTRACT

Title of Dissertation: BUFFER ENGINEERING FOR $M|G|\infty$ INPUT PROCESSES

Minothi A. Parulekar, Doctor of Philosophy, 1999

Dissertation directed by: Professor Armand Makowski
Department of Electrical Engineering and
Institute for Systems Research

We suggest the $M|G|\infty$ input process as a viable model for representing the heavy correlations observed in network traffic. Originally introduced by Cox, this model represents the busy–server process of an $M|G|\infty$ queue with Poisson inputs and general service times distributed according to G, and provides a large and versatile class of traffic models. We examine various properties of the $M|G|\infty$ process, focusing particularly on its rich correlation structure. The process is shown to effectively portray short or long–range dependence simply by controlling the tail of the distribution G.

In an effort to understand the dynamics of a system supporting $M|G|\infty$ traffic, we study the large buffer asymptotics of a multiplexer driven by an $M|G|\infty$ input process. Using the large deviations framework developed by Duffield and

O'Connell, we investigate the tail probabilities for the steady–state buffer content. The key step in this approach is the identification of the appropriate large deviations scaling. This scaling is shown to be closely related to the forward recurrence time of the service time distribution, and a closed form expression is derived for the corresponding limiting log–moment generating function associated with the input process. Three different regimes are identified.

The results are then applied to obtain the large buffer asymptotics under a variety of service time distributions. In each case, the derived asymptotics are compared with simulation results.

While the general functional form of buffer asymptotics may be derived via large deviations techniques, direct arguments often provide a more precise description when the input traffic is heavily correlated. Even so, several significant inferences may be drawn from the functional dependencies of the tail buffer probabilities. The asymptotics already indicate a sub–exponential behavior in the case of heavily–correlated traffic, in sharp contrast to the geometric decay usually observed for Markovian input streams. This difference, along with a shift in the explicit dependence of the asymptotics on the input and output rates r_{in} and c, from $\rho = r_{in}/c$ when G is exponential, to $\Delta = c - r_{in}$ when G is sub–exponential, clearly delineates the heavy and light tailed cases. Finally, comparison with similar asymptotics for a different class of input processes indicates that buffer sizing cannot be adequately determined by appealing solely to the short versus long–range dependence characterization of the input model used.

BUFFER ENGINEERING FOR $M|G|\infty$ INPUT PROCESSES

by

Minothi A. Parulekar

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
1999

Advisory Committee:

Professor Armand Makowski, Chairman/Advisor
Professor Prakash Narayan
Professor Adrian Papamarcou
Professor Leandros Tassiulas
Professor Udaya Shankar

DEDICATION

To my very patient family

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor Armand Makowski for his expert guidance and counsel over the years, and for his generosity and encouragement. The lessons I have learnt under his schooling have extended well beyond the realm of $M|G|\infty$ processes.

The able and cheerful assistance of Sanjeev Khudanpur throughout my studies at UMD is greatly appreciated. I would especially like to thank him for introducing me to the world of communications networks, for providing clear–sighted (if sometimes unsolicited) advice, and for being generally indispensable. I also wish to thank Arnab Das for providing much in the way of stimulating discussions regarding both work and play, as well as invaluable, sanity-saving, stress–busting "griping" sessions, and Konstantinos Tsoukatos for his suggestions and for sharing relevant information.

I am truly thankful for my husband Shekar's inexhaustable reserves of patience and unflappable good humor. I have sorely tried them both and rarely, if ever, succeeded in disturbing either. My brother Rahul, though sadly lacking in these areas has more than made up with his infectious, fiery enthusiasm and a constant willingness to battle on my behalf.

Finally, I owe an eternal debt of gratitude to my parents and my loving aunt Sheila Panjabi for their many sacrifices on my behalf. Their love, enthusiastic support and implicit acceptance of my choices, has greatly enriched my life.

The support of the NSF under Grant NSFD CDR-88-03012, NASA under Grant NAGW77S, the Army Research Laboratory under Cooperative Agreement No. DAAL01-96-2-0002 and the Maryland Procurement Office under Grant No. MDA90497C3015 is gratefully acknowledged.

TABLE OF CONTENTS

List of Tables **x**

List of Figures **xi**

1 Introduction **1**

1.1 Failure of Poisson modeling 1

1.2 Why the $M|G|\infty$ process ? 3

1.3 System Implications 5

1.3.1 Motivation for Research 7

1.4 Overview 9

2 $M|G|\infty$ input processes **11**

2.1 The discrete–time $M|G|\infty$ system 12

2.1.1 System Description 12

2.1.2 Mathematical Representation of $b_t,\ t = 0, 1, \ldots$ 13

2.1.3 The log–moment generating function 15

2.2 $M|G|\infty$ process: stationary version 17

2.2.1 Stationarity 17

2.2.2 Ergodicity 20

2.2.3 Reversibility 20

2.2.4 $M|G|\infty$ process: stationary representation 21
2.3 Correlation structure 25
2.3.1 Association 25
2.3.2 Covariance 27
2.3.3 Long and Short–range Dependence 30
2.3.4 Self-similarity 32

3 General Buffer Asymptotics for a multiplexer 37
3.1 The buffer sizing problem 38
3.2 The Theory of Large Deviations: A brief overview 40
3.2.1 The Large Deviations Principle 43
3.2.2 The Gärtner–Ellis Theorem 45
3.3 The lower bound 46
3.4 The upper bound 48
3.4.1 The basic upper bound 49
3.4.2 An upper bound on $\alpha(y)$ 51
3.4.3 An upper bound on $\beta(y)$ 54
3.4.4 The upper bound 57
3.5 Special Cases 60

4 Evaluation of $\Lambda(\theta)$, $(\theta \in \mathbb{R})$ for the $M|G|\infty$ process 64
4.1 Evaluation of $\Lambda(\theta)$, $(\theta \in \mathbb{R})$ 65
4.2 Evaluation of $\Lambda_{b,n}(\theta)$ $(n = 1, 2, \ldots, \theta \in \mathbb{R})$ 67
4.3 Selection of the sequence v_n, $(n = 1, 2, \ldots)$ 69
4.3.1 Preliminary Results 69
4.4 The linear scaling 73

4.4.1 Finiteness of exponential moments 75

4.5 Sub–linear scaling sequences 80

4.5.1 General Results 80

4.5.2 Scaling $v_n = v_n^\star,\ (n = 1, 2, \ldots)$ 82

4.5.3 Scaling $v_n = o(v_n^\star),\ (n = 1, 2, \ldots)$ 84

4.6 Equivalent scaling sequences 86

4.7 Review and discussion 91

4.7.1 Comparison with the instantaneous input model 95

5 Buffer Asymptotics for the $M|G|\infty$ process 97

5.1 Selection of h and g 97

5.2 $\gamma_\star$ and $\gamma^\star$ 100

5.2.1 The lower bound 100

5.2.2 The upper bound 101

5.3 Buffer Asymptotics for the $M|G|\infty$ process 104

5.3.1 Auxiliary scalings 105

5.3.2 Cases I and II: $R > 0$ 107

5.3.3 Case III: $R = 0$ 108

5.3.4 Beyond Large Deviations Techniques 110

5.4 Alternate Bounds 111

5.4.1 Improved upper bounds 112

5.4.2 General lower bounds 113

6 Examples and Simulation Results 115

6.1 Super–exponential distributions 115

6.1.1 The Deterministic case 115

6.1.2 The Rayleigh case 119
6.2 Exponential distributions 120
6.2.1 The geometric case 120
6.2.2 The Gamma case 126
6.3 Sub–exponential distributions 128
6.3.1 The Weibull case 128
6.3.2 The log–normal case 136
6.3.3 The Pareto case 141
6.4 Discussion 143

7 Concluding Remarks **147**

7.1 Alternate asymptotics 150
7.2 Directions for future research 151

Appendix **153**

A **153**

A.1 Proof of Proposition 2.1.1 153
A.2 Proof of Proposition 2.2.1 158
A.3 Proof of Proposition 2.2.2 161
A.4 Proof of Proposition 2.2.3 165
A.5 Proof of Proposition 2.2.4 167
A.6 Proof of Proposition 2.3.2 169

B **171**

B.1 Proof of Lemma 3.4.3 171

C **174**

C.1 Proof of Lemma 4.2.1 174

C.2 Proof of Theorem 4.2.2 175

C.3 Proof of Lemmas 4.5.4 and 4.5.5 177

D **181**

D.1 Proof of Proposition 5.2.2 181

D.2 Proof of Lemma 5.3.2 183

Bibliography **186**

LIST OF TABLES

4.1 Three cases defined by the tail of distribution G 72

6.1 $\gamma_{\text{Deterministic}}$ 117

6.2 γ_{Rayleigh} 121

6.3 $\gamma_{\text{Geometric}}$ 124

6.4 $\gamma_{\text{Weibull}}, c = 1$ 133

6.5 $\gamma_{\text{Weibull}}, c = 2$ 133

6.6 $\gamma_{\text{Lognormal}}, c = 2$ 139

6.7 $\gamma_{\text{Pareto}}, c = 2$ 144

LIST OF FIGURES

1 $f_{\mathrm{D}}^{-1}(\rho)$ versus ρ 117

2 Tail probability vs. buffer size: Deterministic ($\zeta = 4$) 118

3 Tail probability vs. buffer size: Deterministic ($\zeta = 5$) 118

4 $f_{\alpha}^{-1}(\rho)$ versus ρ, $\alpha = 2.0$ and 6.0. 121

5 Tail probability vs. buffer size: Rayleigh ($\alpha = 2$) 122

6 Tail probability vs. buffer size: Rayleigh ($\alpha = 6$) 122

7 $f_{G}^{-1}(\rho)$ versus ρ, $q = 0.5$ and 0.75. 124

8 Tail probability vs. buffer size: Geometric ($q = 0.5$) 125

9 Tail probability vs. buffer size: Geometric ($q = 0.75$) 125

10 Tail probability vs. buffer size: Weibull ($a = 1.0, \beta = 0.25$) 134

11 Tail probability vs. buffer size: Weibull ($a = 0.5, \beta = 0.5$) 134

12 Tail probability vs. buffer size: Weibull ($a = 1.0, \beta = 0.25$) 135

13 Tail probability vs. buffer size: Weibull ($a = 1.0, \beta = 0.25$) 135

14 Tail probability vs. buffer size: Lognormal ($\delta = 1.414$) 139

15 Tail probability vs. buffer size: Lognormal ($\delta = 1.732$) 140

16 Tail probability vs. buffer size: Lognormal ($\delta = 1.732$) 140

17 Tail probability vs. buffer size: Pareto ($\alpha = 2.5$) 144

18 Tail probability vs. buffer size: Pareto ($\alpha = 1.5$) 145

19 Tail probability vs. buffer size: Pareto ($\alpha = 1.5$) 145

20 Pareto ($\alpha = 1.5$) 151

Chapter 1

Introduction

1.1 Failure of Poisson modeling

The last fifty years have seen a remarkable increase in the number and complexity of available communications services. Networks today support a wide variety of applications, ranging from FTP and TELNET to video and the World Wide Web. The statistical profile of network traffic has also undergone considerable change.

In recent years, traffic measurement studies in a wide range of currently working packet networks, e.g., Ethernet LANs [24,41,42], VBR traffic [26], WAN traffic [54], WWW traffic [13], have uncovered striking differences between traditional and modern traffic patterns. Unlike conventional voice traffic which begins to resemble white noise upon aggregation, modern traffic traces show no signs of "smoothing "at larger time–scales. Instead, the traces remain persistently "bursty"over multiple time–scales, displaying a property of statistical invariance called *self–similarity*. Furthermore, the correlations observed in the accumulated data are significantly heavier than the weak, exponentially decaying correlations seen in traditional telephony. In fact, the underlying correlation structure shows time–

dependencies characteristic of *long–range dependent* processes.

Long–range dependence (LRD) is inherently a non–Markovian property by which the long–term correlations in a process, though individually small, exhibit a slow, hyperbolic decay and as a result are non–summable. This behavior is in sharp contrast to the exponentially decaying (thus summable) correlations traditionally observed in short–range dependent (SRD), mostly Markovian, models [7, 11].

Classical traffic models, based almost exclusively on Poisson–like assumptions about traffic arrival patterns and on exponential assumptions about resource holding requirements, are singularly ill-equipped to account for time dependencies recently observed in network traffic. Superposing several such SRD processes to model LRD is a poor option, akin to expressing a hyperbolic function as the sum of several exponentials. Such models require an increasing number of parameters in order to incorporate an even larger number of time–scales. This "failure of Poisson modeling" along with the need for parsimonious, yet accurate, traffic models has generated an increased interest in a number of alternate traffic models which capture observed (long–range) dependencies.

Proposed models include fractional Brownian motion (FBM) input processes [6, 47, 49], fractional Auto-regressive Integrated Moving Average (F-ARIMA) processes [30, 31], fractal shot–noise driven (FSN) processes [55, 56], as well as several others [3, 19, 61]. In this dissertation, we focus our attention on a model that is extremely versatile, yet remains mathematically convenient: The $M|G|\infty$ input process.

1.2 Why the $M|G|\infty$ process ?

The $M|G|\infty$ input model is the busy server process of a discrete–time $M|G|\infty$ system. Customers, generated according to a (discrete–time) Poisson process with rate λ, are offered to an infinite server group. The required service times are i.i.d. random variables (rvs), with σ denoting the generic service time (expressed in number of time slots). The process $\{b_t,\ t = 0, 1, \ldots\}$ that counts the number of busy servers at the beginning of a time slot is referred to as the $M|G|\infty$ input process.

The process was studied early on by Cox as a model for textile yarn processing [11,12]. His analysis indicated that it was extremely versatile, capable of exhibiting correlations over a wide range of time scales simply by controlling the tail behavior of the distribution of the service time σ. If the autocovariance of lag h for a stationary $M|G|\infty$ process is denoted by $\Gamma(h)$, then

$$\Gamma(h) = \lambda \mathbf{E}\left[\sigma\right] e^{-\upsilon_h}, \quad h = 0, 1, \ldots \tag{1.2.1}$$

where $\upsilon_h = -\ln \mathbf{P}\left[\widehat{\sigma} > h\right]$ and $\widehat{\sigma}$ is the forward recurrence time associated with σ (Proposition 2.3.2). This relation already indicates the tremendous amount of flexibility in modeling positive correlation structures. The degree of positive correlation exhibited by an $M|G|\infty$ input process can be further characterized by the sum of the autocovariances (1.2.1), or index of dispersion of counts (IDC). As is shown later in Proposition 2.3.4, we have

$$\text{IDC} \equiv \sum_{h=0}^{\infty} \Gamma(h) = \frac{\lambda}{2} \mathbf{E}\left[\sigma(\sigma+1)\right], \tag{1.2.2}$$

leading to the simple conclusion that the process is SRD (i.e., IDC finite) if and only if $\mathbf{E}\left[\sigma^2\right]$ is finite. In addition, the correlation in the process uniquely determines the distribution G of the service time σ (Proposition 2.3.3), a valuable property for simple parametric modeling.

The class of $M|G|\infty$ input processes also has other features desirable in a model, in that it is tractable and parsimonious, being completely defined by the pair (λ, G). Further it is stable under multiplexing, i.e., the superposition of several independent $M|G|\infty$ processes can be represented by an $M|G|\infty$ input process. In addition to these natural advantages, research investigating the $M|G|\infty$ model for some wide area applications reports a good fit to TELNET and FTP data using a (integer) log–normal service time [54].

However, its relevance to modern–day traffic modeling is perhaps best explained through its connection to an attractive model for aggregate packet streams proposed by Likhanov, Tsybakov and Georganas [43]. They combine traffic generated by several ON–OFF sources with a Pareto distributed activity period, and show that under appropriate conditions, increasing the number of sources yields a limiting behavior identical to the $M|G|\infty$ input stream with a Pareto distributed σ. The limiting argument used is similar to the Palm–Khintchin Theorem used to justify the Poisson model for interactive data traffic, and is easily seen to hold for arbitrary activity period distributions. This identification of the $M|G|\infty$ model of Cox as the limiting regime of a large number of ON–OFF sources might help explain its success in modeling packet traffic stream in certain applications [54], and points to the $M|G|\infty$ input process as a natural alternative to present traffic models, at least for certain multiplexed applications.

1.3 System Implications

The presence of strong correlations in network traffic is certainly expected to have a serious impact on various aspects of network design including storage requirements, resource allocation, scheduling policies and congestion control. In particular, effective buffer provisioning must now take into account the statistical properties of the traffic supported by the network, or run the risk of increased congestion, packet loss and delay. To gain some insights into this fundamental issue, we analyze the steady–state buffer content at a multiplexer fed by a heavily correlated traffic stream.

For the sake of concreteness, we consider a discrete–time single server queue with infinite capacity and constant release rate of c (cells/slot), as a surrogate for a multiplexer. The number of customers in the input buffer at time t is denoted by q_t. The input stream is assumed to be stationary with rate $r_{in} < c$, in which case the buffer content admits a stationary regime, say q_∞. The steady–state buffer tail probability $\mathbf{P}[q_\infty > b]$ then provides a reasonable performance index, as this quantity is indicative of the buffer overflow probability in a corresponding finite buffer system with b positions.

Results involving the asymptotic behavior of the tail probability $\mathbf{P}[q_\infty > b]$ have been the focus of several researchers, in view of their role in creating effective bandwidths for admission control, and other resource allocation policies [27, 29, 37, 38, 63]. In particular, Glynn and Whitt derived asymptotics of the form

$$\lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}[q_\infty > b] = -\gamma, \qquad \gamma > 0, \tag{1.3.1}$$

under fairly general conditions [27]. These asymptotics naturally bring to mind approximations of the form

$$\mathbf{P}\left[q_\infty > b\right] \sim e^{-b\gamma}, \qquad (b \to \infty); \tag{1.3.2}$$

of course, such extrapolations must be approached with caution [10]. Nevertheless, both (1.3.1) and (1.3.2) are useful in providing qualitative insights into the queueing behavior at the multiplexer, and could in principle provide effective guidelines for buffer sizing.

Unfortunately, the general conditions under which (1.3.1) was derived did not cover input processes with a high degree of correlation, and hence could not be applied to LRD input processes. One of the earliest available results on the queueing behavior of LRD processes, due to Norros, showed that the tail probability of buffer occupancy for a fractional Brownian motion (FBM) input does not exhibit the exponential decay evident in (1.3.1), and is in fact Weibullian in nature [49]. Similar results examining the buffer asymptotics for other LRD input processes have followed since, including [4, 8, 43].

In [16], using large deviations techniques, Duffield and O'Connell developed a generalised version of Glynn and Whitt's classical result [27]. Under this extension several input processes could now be analyzed, including those with a high degree of correlation, that had previously been inadmissible under the stricter requirements imposed by Glynn and Whitt. Furthermore, the extended result was in complete agreement with Glynn and Whitt's predictions of exponential decay under classical conditions, i.e., for lightly correlated input processes.

However, in the case of highly–correlated input streams, the linear scaling evident in (1.3.1) was now replaced by a generalised mapping $h : \mathbb{R}_+ \to \mathbb{R}_+$, giving rise to asymptotics of the kind

$$\lim_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] = -\gamma, \qquad \gamma > 0. \tag{1.3.3}$$

Several applications were also provided in [16] as illustrations of the generalised Duffield and O'Connell result. Of particular interest was the instance when the input stream was FBM in nature. The resulting asymptotics in this case were consistent with those previously derived by Norros [49].

1.3.1 Motivation for Research

The research presented in this dissertation was initiated in an effort to understand, and if possible, isolate the impact of long–range dependence in traffic on buffer dynamics, specifically on the tail probability $\mathbf{P}\left[q_\infty > b\right]$. To this effect, the queueing behavior was examined under two differing sets of traffic assumptions, namely, the $M|G|\infty$ input model with Pareto service times, and the fractional Gaussian noise (FGN) input model (essentially the discrete–time analog of Norros' FBM input model) [51]. Both models were selected in view of their being LRD, mathematically convenient, and able to provide a statistically good fit in diverse applications.

The steady state buffer asymptotics in the FGN case were derived as an application of the Duffield and O'Connell results discussed earlier, and were shown to have the same Weibull-like characteristics visible in the qualitatively similar FBM case. However, applying the Duffield and O'Connell results to the $M|G|\infty$ input process proved a challenging task, as some of the required conditions failed to hold. In place of the equality (1.3.3), we were only able to establish the lower bound

$$\liminf_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \geq -\gamma_\star, \tag{1.3.4}$$

for positive constant $\gamma_\star$, and mapping $h : \mathbb{R}_+ \to \mathbb{R}_+$.

The obvious next step then, was to derive the counterpart to (1.3.4), i.e., the upper bound

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \leq -\gamma^\star, \tag{1.3.5}$$

for some positive constant $\gamma^\star \leq \gamma_\star$, under the same mapping $h : \mathbb{R}_+ \to \mathbb{R}_+$, possibly by an extension of the original Duffield and O'Connell result. To complicate matters, we discovered at this point that their proof as given in [16], was in fact incomplete.

We therefore proceeded to correct the arguments provided in [16], with the intention of adapting them to include the $M|G|\infty$ process with Pareto service times and eventually, extending the derived asymptotics to apply in a broader context, i.e., for a general distribution G [52, 53].

In the interim, Duffield drew our attention to the fact that the lower bound established in [51] was erroneous, as the terms of Gärtner-Ellis theorem, essential to the proof, did not hold under Pareto service times. A version of the lower bound was later provided by Towsley et. al. [45] using direct arguments instead of the usual large deviations approach.

However, the most relevant contribution of this thesis, the identification of the functional form of the tail probability $\mathbf{P}\left[q_\infty > b\right]$, in itself remains significant and offers valuable insights into the impact of heavy correlations in input traffic on queueing behavior. Furthermore, the observation made in [51], that buffer–sizing cannot be adequately determined simply through the LRD versus SRD nature of traffic is still both valid and pertinent.

1.4 Overview

The dissertation is organized as follows: Chapter 2 provides a mathematical representation for the $M|G|\infty$ input process. The necessary conditions for stationarity and ergodicity are investigated via its generating function. The effect of the tail behavior of the service distribution G on the correlation structure is examined in detail. Particularly relevant in the context of practical traffic modeling is the invertibility result presented in Proposition 2.3.3 which claims that an $M|G|\infty$ process is uniquely defined by its correlation structure [39].

Chapter 3 analyzes the buffer dynamics of a single server system driven by a *general* input process via large deviations techniques. Applying extensions of the Duffield and O'Connell results [16], we establish asymptotics of the form (1.3.4) and (1.3.5) under reasonably general conditions for positive constants $\gamma_\star$ and $\gamma^\star$, and mapping $h : \mathbb{R}_+ \to \mathbb{R}_+$.

The key step in the identification of the function h and the constants $\gamma_\star$ and $\gamma^\star$, is the selection of two monotone increasing, $\mathbb{R}_+$–valued scaling sequences $\{v_n,\ n = 0, 1, \ldots\}$ and $\{a_n,\ n = 0, 1, \ldots\}$, so that the limit

$$\Lambda(\theta) \quad \equiv \quad \lim_{n\to\infty} \frac{1}{v_n} \ln \mathbf{E}\left[\exp\left(\frac{\theta v_n}{a_n} S_n\right)\right], \qquad \theta \in \mathbb{R},$$

exists and is non–trivial for some $\theta > 0$.

In Chapters 4 and 5, the general buffer asymptotics derived previously are applied in the particular context of the $M|G|\infty$ input process. Chapter 4 focuses exclusively on the selection of appropriate scalings a and v and the subsequent form taken by the limiting log-moment generating function Λ.

As in the determination of the correlation structure, we find that the tail of the distribution G plays a vital role both in identifying the scalings, and in predicting the resulting form of the function Λ. The consequent derivation of the constants $\gamma^\star$ and $\gamma_\star$, and of the mapping $h : \mathbb{R}_+ \to \mathbb{R}_+$ follows in Chapter 5.

Chapter 6 presents a comparison between theory and simulation for a number of distributions G, in order of increasing tails, in other words, in order of increasing time dependence in the input process. Finally, in Chapter 7, we close with a short discussion and a few suggestions regarding future avenues of research.

A few words on the notation used in this dissertation: For any scalar x in $\mathbb{R}$, we write $\lfloor x \rfloor$ to denote the integer part or floor of x and $\lceil x \rceil$ to denote its ceiling. All rvs are defined on some probability triple $(\Omega, \mathcal{F}, \mathbf{P})$, with $\mathbf{E}$ denoting the corresponding expectation operator. Finally two rvs X and Y are said to be *equal in law* if they have the same distribution, a fact we denote by $X =_{st} Y$. Weak convergence is denoted by $\Longrightarrow$.

Chapter 2

$M|G|\infty$ input processes

As mentioned earlier, an $M|G|\infty$ *input process* is the busy server process of a discrete–time infinite server system fed by a discrete–time Poisson process of rate λ (customers/slot) and with generic service time σ (expressed in number of time slots) distributed according to G. It is an example of a marked process where the underlying point process is Poisson, and the marks associated with each arrival are *i.i.d.* rvs.

In this chapter, we present various facts concerning the busy server process of a discrete–time $M|G|\infty$ system. In Section 2.1, we formally define this system and develop a mathematical representation for the $M|G|\infty$ process. In Section 2.2, we investigate the conditions under which the process is stationary, ergodic and reversible. Finally, we discuss various expressions of the correlations exhibited by the $M|G|\infty$ model, including its covariance, long versus short–range dependence and self–similarity.

2.1 The discrete–time $M|G|\infty$ system

2.1.1 System Description

Consider a system with infinitely many servers. During time slot $[t-1,t)$, $t = 1,2,\ldots$, β_t new customers arrive into the system. Customer i, $i = 1,\ldots,\beta_t$, is presented to its own server and begins service by the start of slot $[t,t+1)$; its service time has duration $\sigma_{t,i}$ (expressed in number of slots). The number of customers initially present in the system at $t = 0$ is denoted by b; customer i, $i = 1,\ldots,b$, brings $\sigma_{0,i}$ units of work to its server. Let b_t denote the number of busy servers, or equivalently of customers still present in the system, at the beginning of slot $[t,t+1)$.

The $\mathbb{N}$–valued rvs b, $\{\beta_t,\ t = 1,2,\ldots\}$ and $\{\sigma_{t,i},\ t = 0,1,\ldots;\ i = 1,2,\ldots\}$ satisfy the following assumptions: (i) The rvs are mutually independent; (ii) The rvs $\{\beta_t,\ t = 1,2,\ldots\}$ are *i.i.d.* Poisson rvs with parameter $\lambda > 0$; (iii) The rvs $\{\sigma_{t,i},\ t = 1,2,\ldots;\ i = 1,2,\ldots\}$ are *i.i.d.* with common pmf G on $\{1,2,\ldots\}$. We denote by σ a generic $\mathbb{N}$–valued rv distributed according to the pmf G. Throughout we shall assume that this pmf G has a finite first moment, or equivalently, that $\mathbf{E}[\sigma] < \infty$.

No additional assumptions are made on the rvs $\{\sigma_{0,i},\ i = 1,2,\ldots\}$ which represent the service durations of the b customers present in the system at the beginning of the slot $[0,1)$, so that various scenarios can in principle be accommodated: If the initial customers start their service at time $t = 0$, then it is appropriate to assume that the rvs $\{\sigma_{0,i},\ i = 1,2,\ldots\}$ are also i.i.d. rvs distributed according to the pmf G. On the other hand, if we take the viewpoint that the system has been in operation for a long time, then these rvs $\{\sigma_{0,i},\ i = 1,2,\ldots\}$ may be interpreted as the residual work (expressed in time slots) that the b "initial" customers require from their respective servers before service is completed. In general, the statistics of the rvs $\{\sigma_{0,i},\ i = 1, 2, \ldots\}$ cannot be specified in any meaningful way, except for the situation when the system is in statistical equilibrium or steady state.

To that end, we find it useful to introduce the forward recurrence time $\widehat{\sigma}$ associated with the rv σ; the pmf of $\widehat{\sigma}$ is given by

$$\mathbf{P}\left[\widehat{\sigma} = t\right] = \widehat{g}_t = \frac{\mathbf{P}\left[\sigma \geq t\right]}{\mathbf{E}\left[\sigma\right]}, \qquad t = 1, 2, \ldots. \tag{2.1.1}$$

or alternatively,

$$\mathbf{P}\left[\widehat{\sigma} \leq t\right] = \frac{\mathbf{E}\left[\min(\sigma, t)\right]}{\mathbf{E}\left[\sigma\right]} = 1 - \frac{\mathbf{E}\left[(\sigma - t)^+\right]}{\mathbf{E}\left[\sigma\right]}.$$

2.1.2 Mathematical Representation of $b_t,\ t = 0, 1, \ldots$

Fix $t = 0, 1, \ldots$. We note that

$$b_t = b_t^{(0)} + b_t^{(a)}, \tag{2.1.2}$$

where the rvs $b_t^{(0)}$ and $b_t^{(a)}$ describe the contributions to the number of customers in the system at the beginning of slot $[t, t+1)$ from those initially present (at $t = 0$) and from the new arrivals, respectively.

Customer i present initially in the system at $t = 0$, survives at time t **iff** its service time $\sigma_{0,i} > t,\ i = 1, 2, \ldots, b$. This gives rise to the representation

$$b_t^{(0)} = \sum_{i=1}^{b} \mathbf{1}\left[\sigma_{0,i} > t\right]. \tag{2.1.3}$$

The rv $b_t^{(a)}$ can also be interpreted as the number of busy servers in the system at time t, given that the system was initially empty (i.e., $b = 0$). Of the β_s arrivals at time s, only those with service requirement exceeding $t-s$ time slots remain in the system at time $t \geq s$. In other words, if $b_t^{(a,s)}$ denotes the number of customers present in the system at time t, having arrived during the interval $[s-1, s)$, we have

$$b_t^{(a,s)} = \sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > t-s\right], \qquad s = 1, 2, \ldots, t.$$

Summing over all such contributions for $s = 1, 2, \ldots, t$, gives the total number of survivors at time t as

$$b_t^{(a)} = \sum_{s=1}^{t} \sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > t-s\right]. \tag{2.1.4}$$

Fixing $s = 1, 2, \ldots, t$, we note that

$$\begin{aligned}
\mathbf{E}\left[\exp\left(\theta b_t^{(a,s)}\right)\right] &= \mathbf{E}\left[\exp\left(\theta \sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > t-s\right]\right)\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\exp\left(\theta \sum_{i=1}^{n} \mathbf{1}\left[\sigma_{s,i} > t-s\right]\right)\right]_{n=\beta_s}\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\exp\left(\theta \mathbf{1}\left[\sigma > t-s\right]\right)\right]^{\beta_s}\right] \\
&= \mathbf{E}\left[\left(\left(e^{\theta}-1\right)\mathbf{P}\left[\sigma > t-s\right]+1\right)^{\beta_s}\right], \qquad \theta \in \mathbb{R}
\end{aligned}$$

upon invoking the $i.i.d$ nature of rvs $\{\beta_t,\ t = 1, 2, \ldots\}$ and $\{\sigma_{t,i},\ t = 0, 1, \ldots;\ i = 1, 2, \ldots\}$, and their mutual independence.

As β_s is a Poisson rv with rate λ, it holds that

$$\mathbf{E}\left[\chi^{\beta_s}\right] = e^{\lambda(\chi-1)}, \qquad \chi > 0, \tag{2.1.5}$$

and substituting $\chi = \left(e^{\theta}-1\right)\mathbf{P}\left[\sigma > t-s\right]+1$ in (2.1.5), we find

$$\mathbf{E}\left[\exp\left(\theta b_t^{(a,s)}\right)\right] = \exp\left(\lambda\left(e^{\theta}-1\right)\mathbf{P}\left[\sigma > t-s\right]\right),$$

i.e., the rv $b_t^{(a,s)}$ is Poisson with rate $\lambda \mathbf{P}\left[\sigma > t-s\right]$.

Under the enforced assumptions, the Poisson rvs $b_t^{(a,1)}, b_t^{(a,2)}, \ldots, b_t^{(a,t)}$ are independent, hence $b_t^{(a)}$ is also Poisson by virtue of (2.1.4) with rate $\lambda \sum_{u=1}^{t} \mathbf{P}[\sigma \geq u]$.

As the distribution of $\{b_t^{(0)},\ t = 0, 1, \ldots\}$ remains unspecified, we are as yet unable to characterize the distribution of the $M|G|\infty$ input process $\{b_t,\ t = 0, 1, \ldots\}$. Section 2.2 sees the emergence of one possible characterization in a fairly natural fashion.

2.1.3 The log–moment generating function

A useful tool in establishing several properties and results concerning any random process is its moment generating function or Laplace transform; we now introduce the notation necessary for its computation in the context of the $M|G|\infty$ process $\{b_t,\ t = 0, 1, \ldots\}$.

For every $n = 1, 2, \ldots$, let $\mathcal{T}^n$ denote the set of all sequences $T_n = (t_1, t_2, \ldots, t_n)$, where $\{t_i,\ i = 1, 2, \ldots, n\}$ are finite, non-negative, non–decreasing integers that divide the interval $(0, \infty]$ into $n+1$ non-overlapping intervals $\{I_j,\ j = 0, 1, 2, \ldots, n\}$. In other words,

$$I_j = \begin{cases} (0, t_1]\,, & j = 0 \\ (t_j, t_{j+1}]\,, & j = 1, 2, \ldots, n-1 \\ (t_n, \infty]\,, & j = n. \end{cases}$$

With the convention $t_0 = 0$ and $t_{n+1} = \infty$, we can rewrite the previous definition in the more convenient form

$$I_j = (t_j, t_{j+1}], \qquad j = 0, 1, \ldots, n. \tag{2.1.6}$$

Let $\mathcal{Q}^n$ be the set of all real–valued sequences $Q_n = (\theta_1, \theta_2, \ldots, \theta_n)$. The log–moment generating function of the random vector $(b_{t_1}, b_{t_2}, \ldots, b_{t_n})$ is then given by

$$\mathcal{L}(T_n, Q_n) = \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}\right)\right], \qquad T_n \in \mathcal{T}^n, Q_n \in \mathcal{Q}^n. \tag{2.1.7}$$

From (2.1.2), we have

$$\mathcal{L}(T_n, Q_n) = \mathcal{L}^{(0)}(T_n, Q_n) + \mathcal{L}^{(a)}(T_n, Q_n), \tag{2.1.8}$$

where

$$\mathcal{L}^{(0)}(T_n, Q_n) = \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b^{(0)}_{t_i}\right)\right]$$

and

$$\mathcal{L}^{(a)}(T_n, Q_n) = \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b^{(a)}_{t_i}\right)\right]$$

for every T_n in $\mathcal{T}^n$ and Q_n in $\mathcal{Q}^n$.

Employing representations (2.1.3) and (2.1.4), we derive alternate expressions for $\mathcal{L}^{(0)}(T_n, Q_n)$ and $\mathcal{L}^{(a)}(T_n, Q_n)$; the details of their derivation can be found in Appendix A.1.

Proposition 2.1.1 *Fix $n = 0, 1, \ldots$. For every T_n in $\mathcal{T}^n$ and Q_n in $\mathcal{Q}^n$, we have*

$$\mathcal{L}^{(0)}(T_n, Q_n) \;=\; \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b}\sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j} \in I_r\right]\right)\right] \tag{2.1.9}$$

and

$$\mathcal{L}^{(a)}(T_n, Q_n) \;=\; \lambda \mathbf{E}\left[\sigma\right] \sum_{j=1}^{n} \Phi_j(T_n, Q_n), \tag{2.1.10}$$

with

$$\Theta_0 \;=\; 0, \qquad \text{and} \qquad \Theta_r \;\equiv\; \sum_{i=1}^{r} \theta_i, \qquad r = 1, 2, \ldots, n, \tag{2.1.11}$$

and

$$\Phi_j(T_n, Q_n) \equiv \left(1 - e^{\Theta_j}\right) \mathbf{P}\left[\widehat{\sigma} \in I_j\right] + \sum_{k=1}^{j} e^{\Theta_j - \Theta_k} \left(e^{\theta_k} - 1\right) \mathbf{P}\left[\widehat{\sigma} + t_k \in I_j\right], \qquad (2.1.12)$$

where the forward recurrence time $\widehat{\sigma}$ *associated with* σ *is given by (2.1.1).*

From Proposition 2.1.1 and (2.1.8) we conclude for each $n = 1, 2, \ldots$ that

$$\mathcal{L}(T_n, Q_n) = \lambda \mathbf{E}\left[\sigma\right] \sum_{j=1}^{n} \Phi_j(T_n, Q_n) + \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j} \in I_r\right]\right)\right] \qquad (2.1.13)$$

for every T_n in $\mathcal{T}^n$ and Q_n in $\mathcal{Q}^n$.

2.2 $M|G|\infty$ process: stationary version

2.2.1 Stationarity

In general, the busy server process $\{b_t,\ t = 0, 1, \ldots\}$ given by

$$b_t = \sum_{i=1}^{b} \mathbf{1}\left[\sigma_{0,i} > t\right] + \sum_{s=1}^{t} \sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > t - s\right], \qquad t = 0, 1, \ldots.$$

is *not* a (strictly) stationary process. However it does have a stationary version denoted from now on by $\{b^\star_t,\ t = 0, 1, \ldots\}$, and obtained by appropriately selecting the initial conditions $b = b^\star_0$ and $\sigma_{0,i} = \sigma^\star_{0,i},\ i = 1, 2, \ldots$.

This section analyzes the restrictions placed on the initial conditions in order to achieve stationarity; we continue to use the notation of Section 2.1.3. We introduce a translation operator $\oplus$ such that for each $n = 1, 2, \ldots$,

$$T_n \oplus h \equiv \{t_i + h,\ i = 1, 2, \ldots, n\}$$

and

$$I_j \oplus h \equiv (t_j + h, t_{j+1} + h], \qquad j = 0, 1, \ldots, n$$

for any T_n in $\mathcal{T}^n$ and $h = 0, 1, \ldots$.

In order for the busy server process to be strictly stationary, it is **necessary** and **sufficient** that

$$(b^\star_{t_1+h}, b^\star_{t_2+h}, \ldots, b^\star_{t_n+h}) \quad =_{st} \quad (b^\star_{t_1}, b^\star_{t_2}, \ldots, b^\star_{t_n}), \qquad h = 0, 1, \ldots; \quad T_n \in \mathcal{T}^n$$

for every $n = 1, 2, \ldots$. As a process is uniquely defined by its log–moment generating function, we have equivalently that a **necessary** and **sufficient** condition for stationarity is given by

$$\mathcal{L}^\star(T_n \oplus h, Q_n) = \mathcal{L}^\star(T_n, Q_n), \qquad h = 0, 1, \ldots; \quad T_n \in \mathcal{T}^n; \quad Q_n \in \mathcal{Q}^n$$

for every $n = 1, 2, \ldots$, where

$$\mathcal{L}^\star(T_n, Q_n) \equiv \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^n \theta_i b^\star_{t_i}\right)\right], \qquad T_n \in \mathcal{T}^n; \quad Q_n \in \mathcal{Q}^n.$$

Further analysis allows this second condition to be re-expressed in the simpler yet equivalent forms given in the following Proposition:

Proposition 2.2.1 *Fix $n = 0, 1, \ldots$. For each pair (T_n, Q_n) in $(\mathcal{T}^n, \mathcal{Q}^n)$ the following requirements are equivalent:*

(i)

$$\mathcal{L}^\star(T_n \oplus h, Q_n) = \mathcal{L}^\star(T_n, Q_n), \qquad h = 0, 1, \ldots; \tag{2.2.1}$$

(ii)

$$\mathcal{L}^{(0)^\star}(T_n, Q_n) \quad = \quad \lambda \mathbf{E}[\sigma]\left(-1 + \sum_{j=0}^n e^{\Theta_j} \mathbf{P}[\widehat{\sigma} \in I_j]\right); \tag{2.2.2}$$

(iii)

$$\mathcal{L}^\star(T_n, Q_n) \quad = \quad \lambda \mathbf{E}\left[\sigma\right] \sum_{k=1}^{n} \left(e^{\theta_k} - 1\right) \sum_{j=k}^{n} e^{\Theta_j - \Theta_k} \mathbf{P}\left[\widehat{\sigma} + t_k \in I_j\right]. \qquad (2.2.3)$$

Detailed arguments substantiating these equivalences are provided in Appendix A.2.

If we select $n = 1$ in (2.2.2) we find

$$\mathbf{E}\left[e^{\theta b_0^\star}\right] \quad = \quad \exp\left(\lambda \mathbf{E}\left[\sigma\right]\left(e^{\theta} - 1\right)\right), \qquad \theta \in \mathbb{R},$$

thus characterizing $b_0^\star$ as a Poisson variable of mean $\lambda \mathbf{E}\left[\sigma\right]$.

By the independence of rvs $b_0^\star$ and $\{\sigma_{0,i}^\star \; i = 1, 2, \ldots\}$, we see from (2.1.9) that

$$\begin{aligned} \mathcal{L}^{(0)^\star}(T_n, Q_n) \quad &= \quad \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r\right]\right)\right] \\ &= \quad \ln\left(\sum_{k=0}^{\infty} I_{T_n,Q_n}(k, \lambda) \mathbf{P}\left[b_0^\star = k\right]\right), \qquad n = 1, 2, \ldots \end{aligned}$$

where

$$I_{T_n,Q_n}(k, \lambda) \equiv \mathbf{E}\left[\exp\left(\sum_{j=1}^{k} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r\right]\right)\right], \qquad n = 1, 2, \ldots. \qquad (2.2.4)$$

By the Poisson nature of rv $b_0^\star$, Proposition 2.2.1 (ii) is equivalent to the condition

$$\sum_{k=0}^{\infty} \frac{(\lambda \mathbf{E}\left[\sigma\right])^k}{k!} I_{T_n,Q_n}(k, \lambda) \quad = \quad \exp\left(\lambda \mathbf{E}\left[\sigma\right]\left(\sum_{j=0}^{n} e^{\Theta_j} \mathbf{P}\left[\widehat{\sigma} \in I_j\right]\right)\right) \qquad (2.2.5)$$

Expanding the right–hand side in the form of a power series, we conclude that the process $\{b_t^\star, t = 0, 1, \ldots\}$ will be stationary **if and only if** for each $n = 1, 2, \ldots$, it holds that

$$\sum_{k=0}^{\infty} \frac{(\lambda \mathbf{E}\left[\sigma\right])^k}{k!} \left(\left(\sum_{j=0}^{n} e^{\Theta_j} \mathbf{P}\left[\widehat{\sigma} \in I_j\right]\right)^k - I_{T_n,Q_n}(k, \lambda)\right) = 0, \qquad (2.2.6)$$

for every pair (T_n, Q_n) in $(\mathcal{T}^n, \mathcal{Q}^n)$.

2.2.2 Ergodicity

A traditional approach often used in investigating the ergodicity of a stationary process involves a form of asymptotic independence known as *mixing*. A stationary process $\{X_n,\ n = 0, 1, \ldots\}$ is said to be *mixing* or *strongly mixing* if for all $k = 1, 2, \ldots$, and all Borel subsets A and B in $\mathcal{B}(\mathbb{R}^k)$, it holds that

$$\begin{aligned}\lim_{n\to\infty} \mathbf{P}\left[(X_1, X_2, \ldots, X_k) \in A, \quad (X_{n+1}, X_{n+2}, \ldots, X_{n+k}) \in B\right] \\ = \mathbf{P}\left[(X_1, X_2, \ldots, X_k) \in A\right] \mathbf{P}\left[(X_1, X_2, \ldots, X_k) \in B\right]. \qquad (2.2.7)\end{aligned}$$

The following proposition makes use of the well-known result that any strongly mixing process is necessarily ergodic [36, p. 489]; the proof is outlined in Appendix A.3.

Proposition 2.2.2 *The stationary process* $\{b^\star_n,\ n = 0, 1, \ldots\}$ *is strongly mixing and therefore ergodic.*

2.2.3 Reversibility

Proposition 2.2.3 *The stationary and ergodic version* $\{b^\star_n,\ n = 0, 1, \ldots\}$ *of the busy server process is reversible in that*

$$(b^\star_0, b^\star_1, \ldots, b^\star_n) =_{st} (b^\star_n, b^\star_{n-1}, \ldots, b^\star_0)$$

for all $n = 0, 1, \ldots$

Proof. As defined earlier in Section 2.2.1, consider any $Q_n \equiv (\theta_1, \theta_2, \ldots, \theta_n)$ in $\mathbb{R}^n$ and denote its mirror image by $Q^r_n \equiv (\theta_n, \theta_{n-1}, \ldots, \theta_1)$.

The process $\{b_n^\star,\ n = 0, 1, \ldots\}$ is reversible **iff** for all $n = 0, 1, \ldots$, we have

$$\mathcal{L}^\star(H_n, Q_n) \quad = \quad \mathcal{L}^\star(H_n, Q_n^r) \tag{2.2.8}$$

where $H_n = (1, 2, \ldots, n)$.

In Appendix A.4 we show that this condition does indeed hold true for the stationary and ergodic version $\{b_n^\star,\ n = 0, 1, \ldots\}$ of the $M|G|\infty$ input process. ■

2.2.4 $M|G|\infty$ process: stationary representation

A number of distributions of the rvs $\{\sigma_{0,j}^\star,\ j = 1, 2, \ldots\}$ satisfy condition (2.2.6), thereby ensuring that the corresponding $M|G|\infty$ process is both stationary and ergodic. In order to narrow down our selection, we can introduce additional restrictions as we deem convenient.

One possible criterion for selection is that the candidate distribution have the desirable property of being **independent** of the arrival rate λ. In other words, for a fixed $n = 1, 2, \ldots$, and (T_n, Q_n) in $(\mathcal{T}^n, \mathcal{Q}^n)$, the expression I_{T_n,Q_n} may be rewritten as a function of just one variable, say

$$I_{T_n,Q_n}(k, \lambda) = I_{T_n,Q_n}(k), \qquad k = 0, 1, \ldots \tag{2.2.9}$$

for all $\lambda > 0$. Under this criterion, the requirement (2.2.6) takes the form

$$\sum_{k=0}^{\infty} \frac{(\lambda \mathbf{E}[\sigma])^k}{k!} \left(\left(\sum_{r=0}^{n} e^{\Theta_r} \mathbf{P}\left[\widehat{\sigma} \in I_r\right] \right)^k - I_{T_n,Q_n}(k) \right) \quad = \quad 0,$$

for **every** $\lambda > 0$, or equivalently,

$$I_{T_n,Q_n}(k) \quad = \quad \left(\sum_{r=1}^{n} e^{\Theta_r} \mathbf{P}\left[\widehat{\sigma} \in I_r\right] \right)^k \tag{2.2.10}$$

for **each** $k = 0, 1, \ldots$.

By the definition of I_{T_n,Q_n} provided in (2.2.4), we may rewrite (2.2.10) as

$$\mathbf{E}\left[\exp\left(\sum_{r=0}^{n} \Theta_r A_r^k\right)\right] = \left(\sum_{r=1}^{n} e^{\Theta_r} \mathbf{P}\left[\widehat{\sigma} \in I_r\right]\right)^k \tag{2.2.11}$$

for each $k = 0, 1, \ldots$, where the rvs

$$A_r^k \equiv \sum_{j=1}^{k} \mathbf{1}\left[\sigma_{0,j}^{\star} \in I_r\right], \quad r = 0, 1, \ldots, n, \tag{2.2.12}$$

constitute an $(n+1)$–length random vector given by $A^k \equiv (A_0^k, A_1^k, \ldots, A_n^k)$.

We note that the value of rv A_r^k depends **only** on **how many** of the k customers present initially in the system have service times that lie in the interval I_r, $r = 0, 1, \ldots, n$; given this information, it is then entirely unaffected by the service time of any particular customer.

To present this idea more rigorously, we fix $k = 0, 1, \ldots$, and introduce $\mathcal{A}_n^k$, the set of all $(n+1)$–length vectors $\alpha^k \equiv (\alpha_0^k, \alpha_1^k, \ldots, \alpha_n^k)$, such that

$$0 \le \alpha_r^k \le k, \quad r = 0, 1, \ldots, n, \quad \text{and} \quad \sum_{r=0}^{n} \alpha_r^k = k.$$

The set $\mathcal{A}_n^k$ in fact, constitutes the range of the random vector A^k introduced earlier. The statement $A^k = \alpha^k$ may then be taken to mean that exactly α_r^k of the initial k customers have service requirements in the interval I_r, $r = 0, 1, \ldots, n$. Of course, summing over all the intervals will give the total number of customers present in the system at time 0, whereby we have $\sum_{r=0}^{n} A_r^k = \sum_{r=0}^{n} \alpha_r^k = k$.

Using this terminology, the left–hand side of condition (2.2.11) may be rewritten as

$$\mathbf{E}\left[\exp\left(\sum_{r=1}^{n} \Theta_r A_r^k\right)\right] = \sum_{\alpha^k \in \mathcal{A}_n^k} \exp\left(\sum_{r=1}^{n} \Theta_r \alpha_r^k\right) \mathbf{P}\left[A^k = \alpha^k\right]. \tag{2.2.13}$$

Turning our attention to the right-hand side of (2.2.11) we have

$$\left(\sum_{r=0}^{n} e^{\Theta_r}\mathbf{P}\left[\widehat{\sigma} \in I_r\right]\right)^k = \sum_{\beta^k \in \mathcal{A}_n^k} \binom{k}{\beta_0^k\ \beta_1^k\ \ldots\ \beta_n^k} \prod_{r=0}^{n} \left(e^{\Theta_r}\mathbf{P}\left[\widehat{\sigma} \in I_r\right]\right)^{\beta_r^k} \qquad (2.2.14)$$

via the multinomial theorem. Equating (2.2.13) and (2.2.14), we derive an alternate representation of condition (2.2.11) given by

$$\sum_{\alpha^k \in \mathcal{A}_n^k} \exp\left(\sum_{r=1}^{n} \Theta_r \alpha_r^k\right) \left(\mathbf{P}\left[A^k = \alpha^k\right] - \binom{k}{\alpha_0^k\ \alpha_1^k\ \ldots\ \alpha_n^k} \prod_{r=0}^{n} \left(\mathbf{P}\left[\widehat{\sigma} \in I_r\right]\right)^{\alpha_r^k}\right) = 0. \qquad (2.2.15)$$

As (2.2.15) holds under all selections of $Q_n \in \mathcal{Q}^n$, it then follows that

$$\mathbf{P}\left[A^k = \alpha^k\right] = \binom{k}{\alpha_0\ \alpha_1\ \ldots\ \alpha_n} \prod_{r=1}^{n} \mathbf{P}\left[\widehat{\sigma} \in I_r\right]^{\alpha_r^k}, \qquad k = 0, 1, \ldots \quad (2.2.16)$$

It is merely a matter of algebra to deduce from this result that

$$\mathbf{P}\left[\sigma_{0,k}^{\star} \in I_r\right] = \mathbf{P}\left[\widehat{\sigma} \in I_r\right], \qquad r = 0, 1, \ldots, n;\ k = 0, 1, \ldots.$$

(The result may be inductively proved for each case $k = 1, 2, \ldots$; the details are left to the reader).

As the selection of the intervals I_r, $r = 0, 1, \ldots, n$, is arbitrary and as our analysis is valid for all $n = 1, 2, \ldots$, we conclude that $\{\sigma_{0,i}^{\star},\ i = 1, 2, \ldots\}$ are identical in distribution to that of the forward recurrence time $\widehat{\sigma}$ associated with σ.

At this point we pause to recapitulate the ground covered so far in this section. Our requirement that the $M|G|\infty$ process be stationary, imposed a restriction on the initial conditions $b_0^{\star}$ and $\{\sigma_{0,j}^{\star},\ j = 1, 2, \ldots\}$: The rv $b_0^{\star}$ was necessarily Poisson with rate $\lambda\mathbf{E}\left[\sigma\right]$ and $\{\sigma_{0,j}^{\star},\ j = 1, 2, \ldots\}$ were required to satisfy condition (2.2.6).

Along with stationarity, the $M|G|\infty$ process automatically inherited properties of ergodicity and reversibility without imposing additional constraints.

Various distributions of $\{\sigma^\star_{0,j},\ j = 1, 2, \ldots\}$ satisfy (2.2.6); we selected only those that did not have any functional dependence on λ. This characterized the rvs $\{\sigma^\star_{0,j},\ j = 1, 2, \ldots\}$ as having marginal densities identical to the rv $\widehat{\sigma}$, but still did not isolate their joint distribution.

The simplest and most convenient of all the qualifying distributions that satisfy (2.2.16) is undoubtedly the one for which $\{\sigma^\star_{0,j},\ j = 1, 2, \ldots\}$ are *i.i.d.* with common distribution $\widehat{G}$; therefore we proceed with this selection for the remainder of this thesis.

Having made these observations, we now state the following proposition that renders the selection of the initial conditions inconsequential during asymptotic or steady–state analysis; its proof can be found in Appendix A.5.

Proposition 2.2.4 *For any choice of the initial condition rv b and of the service times $\{\sigma_{0,i},\ i = 1, 2, \ldots\}$,*

$$\{b_{t+k},\ t = 0, 1, \ldots\} \Longrightarrow \{b^\star_t,\ t = 0, 1, \ldots\} \qquad (k \to \infty) \tag{2.2.17}$$

where $\{b^\star_t,\ t = 0, 1, \ldots\}$ is the stationary and ergodic version of the busy–server process.

In light of Proposition 2.2.4, we assume the stationary version of the $M|G|\infty$ process for the remainder of this document, denoted hereafter without its demarcating asterisk and given by

$$b_t \;=\; \sum_{i=1}^{b} \mathbf{1}\left[\widehat{\sigma}_i > t\right] \quad + \quad \sum_{s=1}^{t}\sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > t - s\right], \quad t = 0, 1, \ldots, \tag{2.2.18}$$

where

(i) The rv b is Poisson with rate $\lambda\mathbf{E}[\sigma]$,

(ii) The rvs $\{\widehat{\sigma}_i,\ i = 1, 2, \ldots\}$ are independent of b and *i.i.d.* with distribution $\widehat{G}$ given by (2.1.1).

2.3 Correlation structure

As mentioned before, one of the most attractive features of the $M|G|\infty$ process is its rich correlation structure. It is capable of exhibiting a remarkable range of time-dependencies simply by changing the distribution G. In particular, the correlation is keenly sensitive to the tail of the distribution $\widehat{G}$ (thus of G). With this in mind, let $v^\star : \mathbb{R}_+ \to \mathbb{R}_+$ denote a mapping such that

$$e^{-v_n^\star} = \mathbf{P}\left[\widehat{\sigma} > n\right], \qquad n = 0, 1, \ldots. \tag{2.3.1}$$

In this section we examine the influence of $v^\star$ in determining the strength of correlation shown by the process. To this end, we review notions of association, long and short–range dependence and self–similarity, all of which can be used to characterize correlations in time.

2.3.1 Association

The first indication that the rvs $\{b_t,\ t = 0, 1, \ldots\}$ exhibit some form of dependence can already be traced to the fact that these rvs are indeed positively correlated in a strong sense: For all $t = 0, 1, \ldots$, we write $b^t \equiv (b_0, b_1, \ldots, b_t)$.

Proposition 2.3.1 *For any choice of the initial condition rv b and of the service times $\{\sigma_{0,i},\ i = 1, 2, \ldots\}$, the rvs $\{b_t,\ t = 0, 1, \ldots\}$ are associated, in that for any*

$t = 0, 1, \ldots$ *and any pair of non–decreasing mappings* $f, g : I\!N^{t+1} \to I\!R$, *we have*

$$\mathbf{E}\left[f(b^t)g(b^t)\right] \geq \mathbf{E}\left[f(b^t)\right]\mathbf{E}\left[g(b^t)\right] \tag{2.3.2}$$

provided that the expectations exist and are finite.

The notion of association was introduced by Esary, Proschan and Walkup in [22]. For the sake of completeness, we include some properties of association derived in [22, p. 1467].

(P1) Any subset of associated random variables is associated;

(P2) The union of two independent sets of associated rvs is also associated;

(P3) The set consisting of a single rv is associated;

(P4) Non–decreasing functions of associated rvs are associated.

Proof. Recall that the collections of rvs $\{b_t^{(0)},\ t = 0, 1, \ldots\}$ and $\{b_t^{(a)},\ t = 0, 1, \ldots\}$ are independent. Hence, in view of (P2) and (2.1.2), we need only show the association (2.3.2) for each of these two collections.

Fix $s = 0, 1, \ldots$ and $i = 1, 2, \ldots$. The rv $\sigma_{s,i}$ being associated by itself by (P3), we see that the rvs $\mathbf{1}\left[\sigma_{s,i} > u\right]$, $u = 0, 1, \ldots$ are also associated by (P4). Further, by independence (P2), the rvs $\left\{\sum_{i=1}^{n} \mathbf{1}\left[\sigma_{s,i} > u\right],\ u = 0, 1, \ldots\right\}$ are associated for each $n = 1, 2, \ldots$, or putting it differently, the rvs $\left\{\sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > u\right],\ u = 0, 1, \ldots\right\}$ are conditionally associated given β_s (where β_0 denotes b).

Using the notation $b^{(s),u} = \sum_{i=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,i} > u\right]$ for $u = 0, 1, \ldots$, we conclude from this last remark that for any pair of non–decreasing mappings $f, g : I\!N^{u+1} \to I\!R$,

$$\mathbf{E}\left[f(b^{(s),u})g(b^{(s),u})\right] = \mathbf{E}\left[\mathbf{E}\left[f(b^{(s),u})g(b^{(s),u})|\beta_s\right]\right]$$

$$\geq \mathbf{E}\left[\mathbf{E}\left[f(b^{(s),u})|\beta_s\right]\mathbf{E}\left[g(b^{(s),u})|\beta_s\right]\right]$$
$$\geq \mathbf{E}\left[\mathbf{E}\left[f(b^{(s),u})|\beta_s\right]\right]\mathbf{E}\left[\mathbf{E}\left[g(b^{(s),u})|\beta_s\right]\right] \quad (2.3.3)$$

thus implying that the rvs $\{b^{(s),u},\ u = 0, 1, \ldots\}$ are associated. The passage to (2.3.3) is a consequence of the rv β_s being associated by (P3), and of the non–decreasing nature of the mappings $n \to \mathbf{E}\left[f(b^{(s),u})|\beta_s = n\right]$ and $n \to \mathbf{E}\left[g(b^{(s),u})|\beta_s = n\right]$.

The desired conclusion on the rvs $\{b_t^{(0)},\ t = 0, 1, \ldots\}$ follows directly from (2.3.3) when $s = 0$. In the case of the collection $\{b_t^{(a)},\ t = 0, 1, \ldots\}$, we arrive at the same conclusion by the independence of the rvs $\{b^{(s),u},\ s = 0, 1, \ldots\}$ and the fact that $b_t^{(a)} = \sum_{s=1}^{t} b^{(s),u}$ when $u = t - s$. ■

Proposition 2.3.1 already suggests that the covariance structure of $\{b_t,\ t = 0, 1, \ldots\}$ satisfies

$$\text{cov}[b_t, b_{t+h}] \geq 0, \qquad t, h = 0, 1, \ldots. \quad (2.3.4)$$

We now proceed to compute its exact form.

2.3.2 Covariance

The covariance structure of the $M|G|\infty$ process can be calculated directly from (2.2.18). However, it is simpler to derive it by the partial differentiation of the generating function of process $\{b_t,\ t = 1, 2, \ldots\}$ given by (2.2.3). The following proposition gives the final form of the covariance function $\Gamma(h)$; its proof is available for perusal in Appendix A.6.

Proposition 2.3.2 *The covariance structure of the stationary and ergodic version* $\{b_t,\ t = 0, 1, \ldots\}$ *of the busy server process is given by*

$$\Gamma(h) \equiv \mathrm{cov}[b_t, b_{t+h}] = \lambda \mathbf{E}\left[\sigma\right] \mathbf{P}\left[\widehat{\sigma} > h\right] = \lambda \mathbf{E}\left[\sigma\right] e^{-v_h^\star}, \qquad t, h = 0, 1, \ldots \quad (2.3.5)$$

An alternative expression is given in the following lemma.

Lemma 2.3.1 *We have*

$$\Gamma(h) = \lambda \mathbf{E}\left[(\sigma - h)^+\right], \qquad h = 1, 2, \ldots \quad (2.3.6)$$

Proof. Fix $h = 1, 2, \ldots$, and note that

$$\begin{aligned} \Gamma(h) &= \lambda \mathbf{E}\left[\sigma\right] \mathbf{P}\left[\widehat{\sigma} > h\right] \\ &= \lambda \mathbf{E}\left[\sigma\right] \sum_{r=h+1}^{\infty} \mathbf{P}\left[\widehat{\sigma} = r\right] \\ &= \lambda \sum_{r=h+1}^{\infty} \mathbf{P}\left[\sigma \geq r\right] \\ &= \lambda \sum_{r=0}^{\infty} \mathbf{P}\left[\sigma > h + r\right] \\ &= \lambda \sum_{r=0}^{\infty} \mathbf{P}\left[(\sigma - h)^+ > r\right] \\ &= \lambda \mathbf{E}\left[(\sigma - h)^+\right]. \end{aligned}$$

■

In the case of the stationary $M|G|\infty$ input process, the autocorrelation function $\gamma : I\!N \to \mathbb{R}$, obtained from (2.3.5) is given by

$$\gamma(h) \equiv \frac{\Gamma(h)}{\Gamma(0)} = \mathbf{P}\left[\widehat{\sigma} > h\right], \qquad h = 0, 1, \ldots \quad (2.3.7)$$

The correlation structure is completely determined by the pmf of $\widehat{\sigma}$ (thus of σ). It turns out that the inverse is true as well. Indeed,

$$\begin{aligned} \gamma(h) - \gamma(h+1) &= \mathbf{P}\left[\widehat{\sigma} > h\right] - \mathbf{P}\left[\widehat{\sigma} > h+1\right] \\ &= \frac{1}{\mathbf{E}\left[\sigma\right]}\mathbf{P}\left[\sigma > h\right], \qquad h = 0, 1, \ldots \end{aligned} \tag{2.3.8}$$

so that the mapping $h \to \gamma(h)$ is necessarily decreasing and integer–convex. We conclude from (2.3.8) (with $h = 0$) that

$$\mathbf{E}\left[\sigma\right]^{-1} = 1 - \gamma(1) \tag{2.3.9}$$

with $\gamma(1) < 1$ necessarily by the finiteness of $\mathbf{E}\left[\sigma\right]$. Combining (2.3.8) and (2.3.9) we find that

$$\mathbf{P}\left[\sigma > h\right] = \frac{\gamma(h) - \gamma(h+1)}{1 - \gamma(1)}, \qquad h = 0, 1, \ldots \tag{2.3.10}$$

Note also from (2.3.10) that

$$\mathbf{E}\left[\sigma\right] = \sum_{h=0}^{\infty} \mathbf{P}\left[\sigma > h\right] = \frac{1 - \lim_{h\to\infty} \gamma(h)}{1 - \gamma(1)}$$

and (2.3.9) imposes $\lim_{h\to\infty} \gamma(h) = 0$. A moment of reflection readily leads to the following invertibility result.

Proposition 2.3.3 *An $\mathbb{R}_+$–valued sequence $\{\gamma(h),\ h = 0, 1, \ldots\}$ is the auto-correlation function of the $M|G|\infty$ process (λ, σ) with integrable σ if and only if the corresponding mapping $h \to \gamma(h)$ is decreasing and integer–convex with $\gamma(0) = 1 > \gamma(1)$ and $\lim_{h\to\infty} \gamma(h) = 0$, in which case the pmf G of σ is given by (2.3.10).*

2.3.3 Long and Short–range Dependence

The strength of the positive correlation exhibited by the sequence $\{b_t,\ t = 0, 1, \ldots\}$ can be formalized as follows: We say that the sequence $\{b_t,\ t = 0, 1, \ldots\}$ exhibits *short range dependence* if

$$\sum_{h=0}^{\infty} \Gamma(h) < \infty.$$

Otherwise, the sequence $\{b_t,\ t = 0, 1, \ldots\}$ is said to be *long range dependent* [6, 7]. As we now show, for $M|G|\infty$ processes this dependence can be partially characterized through the scaling $\{v_t^\star,\ t = 1, 2, \ldots\}$.

Proposition 2.3.4 *We have the relation*

$$\sum_{h=0}^{\infty} \Gamma(h) = \lambda \mathbf{E}[\sigma] \mathbf{E}[\widehat{\sigma}] = \frac{\lambda}{2} \mathbf{E}[\sigma(\sigma + 1)], \tag{2.3.11}$$

so that the stationary sequence $\{b_t,\ t = 0, 1, \ldots\}$ *is short range dependent (SRD) (resp. long range dependent (LRD)) if and only if* $\mathbf{E}[\widehat{\sigma}]$ *is finite (resp. infinite).*

Proof. From (2.3.6), we see that

$$\begin{aligned}
\sum_{h=0}^{\infty} \Gamma(h) &= \lambda \mathbf{E}[\sigma] \sum_{h=0}^{\infty} \mathbf{P}[\widehat{\sigma} > h] \\
&= \lambda \mathbf{E}[\sigma] \mathbf{E}[\widehat{\sigma}] \\
&= \lambda \mathbf{E}[\sigma] \sum_{r=1}^{\infty} r \mathbf{P}[\widehat{\sigma} = r] \\
&= \lambda \mathbf{E}[\sigma] (\mathbf{E}[\sigma])^{-1} \sum_{r=1}^{\infty} r \mathbf{P}[\sigma \geq r] \\
&= \lambda \sum_{r=1}^{\infty} r \sum_{t=r}^{\infty} \mathbf{P}[\sigma = t]
\end{aligned}$$

$$
\begin{aligned}
&= \lambda \sum_{t=1}^{\infty} \mathbf{P}[\sigma = t] \left(\sum_{r=1}^{t} r \right) \\
&= \frac{\lambda}{2} \sum_{t=1}^{\infty} t(t+1) \mathbf{P}[\sigma = t]
\end{aligned}
$$

and the conclusion (2.3.11) is now immediate. ■

Proposition 2.3.4 gives rise to a simple test to check if the process $\{b_t,\ t = 0, 1, \ldots\}$ is SRD; we present it in the form of the following corollary.

Corollary 2.3.1 *If* $\lim_{n\to\infty} \frac{v_n^\star}{\ln n} = K$, *then the process* $\{b_t,\ t = 0, 1, \ldots\}$ *is SRD (resp. LRD) if* $K > 1$ *(resp.* $K < 1$*).*

The proof rests on the well–known result

$$
\sum_{n=1}^{\infty} n^{-\delta} < \infty, \qquad \delta > 1, \tag{2.3.12}
$$

and on the lemma stated below, the proof of which is fairly simple and has not been included.

Lemma 2.3.2 *Consider two* $\mathbb{R}_+$*–valued sequences* $\{\alpha_n,\ n = 0, 1, \ldots\}$ *and* $\{\beta_n,\ n = 0, 1, \ldots\}$ *such that* $\lim_{n\to\infty} \frac{\alpha_n}{\beta_n} = \infty$*. Then* $\sum_{n=1}^{\infty} \alpha_n < \infty$ *necessarily implies* $\sum_{n=1}^{\infty} \beta_n < \infty$, *while* $\sum_{n=1}^{\infty} \beta_n = \infty$ *implies* $\sum_{n=1}^{\infty} \alpha_n = \infty$.

Proof of Corollary 2.3.1. Set $\alpha_n = n^{-\delta}$ with $\delta > 0$ and $\beta_n = e^{-v_n^\star}$, $n = 1, 2, \ldots$, in which case

$$
\lim_{n\to\infty} \frac{\alpha_n}{\beta_n} = \lim_{n\to\infty} \frac{e^{-\delta \ln n}}{e^{-v_n^\star}}
$$

$$
\begin{aligned}
&= \lim_{n\to\infty} e^{-\delta \ln n + v_n^\star} \\
&= \lim_{n\to\infty} e^{(\frac{v_n^\star}{\ln n} - \delta)\ln n} \\
&= \begin{cases} \infty & K > \delta \\ 0 & K < \delta. \end{cases}
\end{aligned}
\tag{2.3.13}
$$

When $K > 1$, select δ such that $1 < \delta < K$, so that $\sum_{n=1}^{\infty} \alpha_n < \infty$ by (2.3.12), and $\lim_{n\to\infty} \frac{\alpha_n}{\beta_n} = \infty$ by (2.3.13). Therefore, applying Lemma 2.3.2, we have

$$
\sum_{n=0}^{\infty} e^{-v_n^\star} = \mathbf{E}\left[\widehat{\sigma}\right] < \infty,
$$

and the related process $\{b_t,\ t = 0, 1, \ldots\}$ is SRD.

Next, set $\alpha_n = e^{-v_n^\star}$ and $\beta_n = n^{-\delta}$, $n = 1, 2, \ldots$, with $\delta > 0$, giving

$$
\lim_{n\to\infty} \frac{\alpha_n}{\beta_n} = \begin{cases} 0 & K > \delta \\ \infty & K < \delta. \end{cases}
\tag{2.3.14}
$$

When $K < 1$, select δ such that $K < \delta < 1$, and note that $\sum_{n=1}^{\infty} \beta_n = \infty$ by (2.3.12), and $\lim_{n\to\infty} \frac{\alpha_n}{\beta_n} = \infty$ by (2.3.14). Therefore, applying Lemma 2.3.2, we conclude that

$$
\sum_{n=0}^{\infty} e^{-v_n^\star} = \infty,
$$

and the process $\{b_t,\ t = 0, 1, \ldots\}$ is LRD. ■

For the case $K = 1$, the test fails and no conclusion may be drawn without knowing the exact form of $v_n^\star$.

2.3.4 Self-similarity

An interesting class of processes can be identified through the notion of second–order *self–similarity*. We briefly review some relevant definitions and properties concerning self–similar processes; detailed information regarding the subject can be found in [11] and [60].

Consider a process $\{a_t,\ t = 0, 1, \ldots\}$. For each $m = 1, 2, \ldots$, we introduce the associated *m–averaged* process $\{a_t^{(m)},\ t = 0, 1, \ldots\}$, given by

$$a_t^{(m)} \equiv \frac{1}{m} \sum_{k=0}^{m-1} a_{mt+k}, \qquad t = 0, 1, \ldots, \tag{2.3.15}$$

and the *m–normalized* process $\{\breve{a}_t^m, t = 0, 1, \ldots\}$ achieved through the transformation

$$\breve{a}_t^{(m)} \quad \equiv \quad m^{1-H} a_t^{(m)} \quad = \quad \frac{1}{m^H} \sum_{k=0}^{m-1} a_{mt+k}, \qquad t = 0, 1, \ldots, \tag{2.3.16}$$

where $0 < H < 1$ is the index of normalization.

The motivation behind naming a process *self–similar* is evident from the following definition:

Definition 2.3.1 *A strictly stationary process $\{a_t,\ t = 0, 1, \ldots\}$ is called strictly self–similar with Hurst parameter H, if for each $m = 1, 2, \ldots$ we have*

$$\breve{a}_t^{(m)} \quad =_{st} \quad a_t, \qquad t = 0, 1, \ldots, \tag{2.3.17}$$

where $\{\breve{a}_t^{(m)},\ t = 0, 1, \ldots\}$ is the m-normalized process (2.3.16) with index of normalization H [60].

Clearly, a strictly self–similar process maintains its probabilistic structure when scaled and appropriately normalized.

Definition 2.3.1 is far too restrictive to be more than theoretically useful. As is evident via (2.3.16) and (2.3.17), a strictly self–similar process must necessarily have a zero–mean. Further, if the process $\{a_t,\ t = 0, 1, \ldots\}$ is positive and non–degenerate, neither $\{a_t,\ t = 0, 1, \ldots\}$, nor $\{a_t - \mathbf{E}[a_t],\ t = 0, 1, \ldots\}$ can be strictly self–similar [60]. We therefore shift our focus to the broader class of *exactly second–order self–similar* processes.

Definition 2.3.2 *A wide–sense stationary process* $\{a_t,\ t = 0, 1, \ldots\}$ *is said to be exactly second–order self–similar with Hurst parameter* H*, if for each* $m = 1, 2, \ldots$ *we have*

$$\mathrm{var}[\breve{a}_t^{(m)}] \quad = \quad \mathrm{var}[a_t], \qquad t = 0, 1, \ldots, \tag{2.3.18}$$

where $\{\breve{a}_t^{(m)},\ t = 0, 1, \ldots\}$ *is the m-normalized process (2.3.16) with index of normalization* H *[60].*

Under the assumption that the process $\{a_t,\ t = 0, 1, \ldots\}$ is wide–sense stationary, it is plain that for each $m = 1, 2, \ldots$, the rvs $\{a_t^{(m)},\ t = 0, 1, \ldots\}$ and $\{\breve{a}_t^{(m)},\ t = 0, 1, \ldots\}$ also form wide–sense stationary sequences with correlations

$$\Gamma^{(m)}(h) \equiv \mathrm{cov}[a_t^{(m)}, a_{t+h}^{(m)}] \quad \text{and} \quad \gamma^{(m)}(h) \equiv \frac{\Gamma^{(m)}(h)}{\Gamma^{(m)}(0)}, \quad h = 0, 1, \ldots,$$

and

$$\breve{\Gamma}^{(m)}(h) \equiv \mathrm{cov}[\breve{a}_t^{(m)}, \breve{a}_{t+h}^{(m)}] \quad \text{and} \quad \breve{\gamma}^{(m)}(h) \equiv \frac{\breve{\Gamma}^{(m)}(h)}{\breve{\Gamma}^{(m)}(0)}, \quad h = 0, 1, \ldots,$$

respectively.

We may now rewrite (2.3.18) as

$$\breve{\Gamma}^{(m)}(0) \quad = \quad \Gamma(0), \qquad m = 1, 2, \ldots.$$

or alternatively as

$$\Gamma^{(m)}(0) \quad = \quad \Gamma(0) m^{-2(1-H)}, \qquad m = 1, 2, \ldots. \tag{2.3.19}$$

A requirement equivalent to (2.3.19) is given by

$$\Gamma(h) = \Gamma(0)\gamma_H(h), \qquad h = 0, 1, \ldots \tag{2.3.20}$$

where the Hurst parameter of the process is H and the mapping $\gamma_H : \mathbb{N} \to \mathbb{R}_+$ is given by

$$\gamma_H(h) \equiv \frac{1}{2}\left(|h+1|^{2H} - 2|h|^{2H} + |h-1|^{2H}\right), \qquad h = 0, 1, \ldots \tag{2.3.21}$$

The parameter H being in the range $(.5, 1)$, the mapping γ_H is strictly decreasing and integer–convex with $\gamma_H(0) = 1$, and behaves asymptotically as

$$\gamma_H(h) \sim H(2H-1)h^{2H-2} \qquad (h \to \infty). \tag{2.3.22}$$

By Proposition 2.3.3 we can interpret γ_H as the autocorrelation function of the $M|G|\infty$ input process (λ, σ_H) with

$$\mathbf{P}\left[\sigma_H > r\right] = \frac{|r+2|^{2H} - 3|r+1|^{2H} + 3|r|^{2H} - |r-1|^{2H}}{4(1-2^{2H-2})}, \qquad r = 1, 2, \ldots$$

so that the $M|G|\infty$ input process (λ, σ_H) is exactly second–order self–similar with Hurst parameter H.

For convenient application, even the relatively larger class of exactly second order self–similar process proves too narrow. We therefore relax our definition as follows:

Definition 2.3.3 *The sequence $\{a_t,\ t = 0, 1, \ldots\}$ is said to be asymptotically second–order self–similar if*

$$\lim_{m\to\infty} \gamma^{(m)}(h) = \gamma_H(h), \qquad h = 1, 2, \ldots, \tag{2.3.23}$$

in which case H is still referred to as the Hurst parameter of the process.

For the $M|G|\infty$ process $(\lambda, \mathbf{E}[\sigma])$ to be *asymptotically (second–order) self–similar* it suffices to have

$$\mathbf{P}[\sigma > r] \sim r^{-\alpha} L(r),$$

with $1 < \alpha < 2$, for some slowly varying function $L : \mathbb{R}_+ \to \mathbb{R}_+$, in which case $H = \dfrac{3-\alpha}{2}$.

Chapter 3

General Buffer Asymptotics for a multiplexer

Section 1.3 emphasized the importance of taking into account the statistical nature of the traffic indigenous to a network in its design. In this chapter we tackle a problem of particular interest in network design, namely, buffer provisioning. To gain some insights into this fundamental issue, we analyze the content of an infinite-sized buffer, at a multiplexer being fed by a random traffic stream. Assuming the system achieves statistical equilibrium, the steady–state queue size q_∞ provides valuable guidelines in estimating the size of buffer in various practical applications. In particular, the tail probability $\mathbf{P}[q_\infty > b]$ provides a good indicator to the cell loss probability in the corresponding finite buffer system with b positions.

Computing these tail probabilities, either analytically or numerically, represents a challenging problem in the absence of any underlying Markov property for a general input process. Instead, we focus on the simpler task of determining the tail behaviour of the queue–length distribution in some asymptotic sense. More precisely, we seek asymptotic results of the kind

$$\lim_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}[q_\infty > b] = -\gamma \tag{3.0.1}$$

for some positive constant γ and mapping $h : \mathbb{R}_+ \to \mathbb{R}_+$.

In order to obtain (3.0.1), our focus has primarily been on Large Deviations techniques. This approach has already been adopted by a number of authors [16, 27, 38], and typically involves computing the upper and lower bounds

$$\liminf_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \geq -\gamma_\star, \tag{3.0.2}$$

and

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \leq -\gamma^\star, \tag{3.0.3}$$

for some constants $\gamma_\star > 0$ and $\gamma^\star > 0$, and analyzing conditions under which the two coincide.

We begin the discussion with a brief description and mathematical representation of the system in Section 3.1. Section 3.2 covers the background necessary for later analysis and presents a brief overview of Large Deviations techniques that apply in our context. The subsequent sections develop asymptotic estimates of $\mathbf{P}\left[q_\infty > b\right]$ via extensions of existing LD results.

3.1 The buffer sizing problem

The multiplexer is modeled as a discrete–time single server queue with infinite buffer capacity and constant release rate of c cells/slot under the first–come first–serve discipline. The input stream is represented by $\{b_n,\ n = 1, 2, \ldots\}$ where b_{n+1} is the number of new cells arriving at the start of time slot $[n, n+1)$. Let q_n denote the number of cells remaining in the buffer by the end of slot $[n-1, n)$, so that $q_n + b_{n+1}$ cells are ready for transmission in the next slot. If the multiplexer output link can transmit c cells/slot, then the buffer content sequence $\{q_n,\ n = 0, 1, \ldots\}$

evolves according to the Lindley recursion

$$q_0 = q; \quad q_{n+1} = [q_n + b_{n+1} - c]^+, \qquad n = 0, 1, \ldots \tag{3.1.1}$$

for some initial condition q.

With the notation $\xi_n = b_n - c$, the recursion (3.1.1) becomes

$$q_0 = q; \quad q_{n+1} = [q_n + \xi_{n+1}]^+, \qquad n = 0, 1, \ldots \tag{3.1.2}$$

yielding the solution

$$\begin{aligned} q_n &= \max\left(0, q + \sum_{i=1}^{n} \xi_i, \sup_{j=2,\ldots,n} \left(\sum_{i=j}^{n} \xi_i\right)\right) \\ &= \max\left(0, q + \sum_{i=1}^{n} \xi_i, \sup_{j=1,2,\ldots,n-1} \left(\sum_{i=1}^{j} \xi_{n+1-i}\right)\right), \qquad n = 0, 1, \ldots \end{aligned}$$

If the process $\{b_n,\ n = 1, 2, \ldots\}$ is reversible, then we have

$$\{b_i,\ i = 1, 2, \ldots, n\} \quad =_{st} \quad \{b_{n+1-i},\ i = 1, 2, \ldots, n\},$$

for each $n = 1, 2, \ldots$, and it holds that

$$\begin{aligned} q_n &=_{st} \max\left(0, q + \sum_{i=1}^{n} \xi_i, \sup_{j=1,2,\ldots,n-1} \left(\sum_{i=1}^{j} \xi_i\right)\right) \\ &= \max\left(q + S_n, \sup_{j=0,1,\ldots,n-1} S_j\right) \end{aligned}$$

with

$$S_0 = 0; \quad S_j = \sum_{i=1}^{j} \xi_i, \qquad j = 1, 2, \ldots. \tag{3.1.3}$$

Under certain conditions, the multiplexer will reach statistical equilibrium, that is $q_n \Longrightarrow_n q_\infty$, where q_∞ represents the steady–state buffer content at the multiplexer.

The following proposition, due to Loynes, outlines these conditions [46].

Proposition 3.1.1 *If the sequence* $\{\xi_n,\ n = 1, 2, \ldots\}$ *is stationary and ergodic, with average rate* $\mathbf{E}[\xi_1] < 0$*, then (3.1.2) admits a steady state regime, in that* $q_n \Longrightarrow_n q_\infty$ *with*

$$q_\infty =_{st} \sup\{S_j,\ j = 0, 1, \ldots\} \tag{3.1.4}$$

where $\{S_j,\ j = 0, 1, \ldots\}$ *is given by (3.1.3). Further, if* $\mathbf{E}[\xi_1] > 0$*, then* $q_n \Longrightarrow_n \infty$*.*

In other words, if the process $\{b_n,\ n = 1, 2, \ldots\}$ is stationary and ergodic with $r_{in} = \mathbf{E}[b_1] < c$, then the system will eventually reach steady–state for any choice of the initial condition q and will then be described as stable. Though sufficient, the stationarity and ergodicity of the input process is by no means necessary for the system to reach stability. In fact, it has been shown [5] that Loynes' result can be extended to Lindley recursions driven by sequences which *couple* with their stationary and ergodic versions.

3.2 The Theory of Large Deviations: A brief overview

To begin with, we introduce two *scaling* sequences $\{v_n,\ n = 0, 1, \ldots\}$ and $\{a_n,\ n = 0, 1, \ldots\}$. A scaling sequence is any monotone increasing $\mathbb{R}$–valued sequence $\{v_n,\ n = 0, 1, \ldots\}$ such that $\lim_{n\to\infty} v_n = \infty$. Throughout we use the notation (v, a) to denote a particular selection of scalings $v, a : \mathbb{N} \to \mathbb{R}_+$.

The generalized inverses of the sequence $\{a_n,\ n = 0, 1, \ldots\}$ are the mappings $a_l^{-1}, a_r^{-1} : \mathbb{R}_+ \to \mathbb{N}$ given by

$$a_l^{-1}(x) \equiv \inf\{s \in \mathbb{N} :\ a_s \geq x\}, \qquad x \geq 0 \tag{3.2.1}$$

and

$$a_r^{-1}(x) \equiv \sup\{s \in I\!N : \ a_s \leq x\}, \qquad x \geq 0 \tag{3.2.2}$$

with the convention $a_0 = 0$. We refer to the mappings a_l^{-1} and a_r^{-1} as the left and right generalized inverses, as they are left and right continuous, respectively. By their definition, it is plain that

$$a_{a_r^{-1}(x)} \leq x \leq a_{a_l^{-1}(x)}, \qquad x \geq 0. \tag{3.2.3}$$

For every pair (v, a) we postulate the existence of functions $g, h : \mathbb{R}_+ \to \mathbb{R}_+$ such that h is monotone increasing with $\lim_{b\to\infty} h(b) = \infty$, and the limits

$$\lim_{b\to\infty} \frac{v_{a_l^{-1}(b/y)}}{h(b)} = \lim_{b\to\infty} \frac{v_{a_r^{-1}(b/y)}}{h(b)} = g(y), \qquad y > 0 \tag{3.2.4}$$

all exist and are finite. The mapping $g : (0, \infty) \to \mathbb{R}_+$ is necessarily monotone decreasing.

Condition (3.2.4) is not as stringent as it seems at first. In fact, it holds broadly in applications and often has a rather simple form.

Lemma 3.2.1 *For every pair (v, a), the monotone mappings $g, h : \mathbb{R}_+ \to \mathbb{R}_+$ satisfy (3.2.4)* **iff**

$$\lim_{n\to\infty} \frac{v_n}{h(ya_n)} \quad = \quad g(y), \qquad y > 0. \tag{3.2.5}$$

Proof. Fix $y > 0$. We introduce the variables $n \equiv a_l^{-1}(b/y)$, and $m \equiv a_r^{-1}(b/y)$, $b > 0$. Clearly, by definitions (3.2.1) and (3.2.2) the values taken by n and m traverse the entire range $\{0, 1, 2, \ldots\}$. Furthermore, as b increases to ∞, so do n and m.

Via (3.2.3), we have the inequalities

$$ya_m \le b \le ya_n, \qquad b > 0,$$

whence

$$\frac{v_n}{h(ya_n)} \le \frac{v_{a_l^{-1}(b/y)}}{h(b)} \qquad \text{and} \qquad \frac{v_{a_r^{-1}(b/y)}}{h(b)} \le \frac{v_m}{h(ya_m)} \tag{3.2.6}$$

for *any* monotone increasing function $h : \mathbb{R}_+ \to \mathbb{R}_+$.

Assume that there exist mappings $g, h : \mathbb{R}_+ \to \mathbb{R}_+$ such that (3.2.5) holds. Then, by (3.2.6) we get

$$g(y) \le \liminf_{b\to\infty} \frac{v_{a_l^{-1}(b/y)}}{h(b)} \qquad \text{and} \qquad \limsup_{b\to\infty} \frac{v_{a_r^{-1}(b/y)}}{h(b)} \le g(y).$$

The scaling v being monotone increasing, (3.2.4) follows directly upon noting that $a_r^{-1}(x) \le a_l^{-1}(x)$ for all $x > 0$.

Conversely, assume the existence of mappings $g, h : \mathbb{R}_+ \to \mathbb{R}_+$ in (3.2.4). We have

$$\limsup_{n\to\infty} \frac{v_n}{h(ya_n)} \le g(y) \qquad \text{and} \qquad g(y) \le \liminf_{m\to\infty} \frac{v_m}{h(ya_m)},$$

and (3.2.5) is obtained. ■

Recall from [23, p. 269] that a positive Borel function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is *regularly varying* if the limits

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = F(x), \qquad x \ge 0 \tag{3.2.7}$$

all exist (possibly as an extended real number). It is well known [23] that the limiting function F is necessarily of the form

$$F(x) = x^\rho, \qquad x \ge 0 \tag{3.2.8}$$

for some ρ in $[-\infty, \infty]$, in which case f is said to belong to $RV(\rho)$.

Define

$$f_l(x) \equiv v_{a_l^{-1}(x)} \qquad \text{and} \qquad f_r(x) \equiv v_{a_r^{-1}(x)}, \qquad x > 0. \tag{3.2.9}$$

Then the following lemma holds.

Lemma 3.2.2 *If f_l (resp. f_r) is regularly varying with parameter ρ in $[0, \infty]$, then $h(b) = f_l(b)$ (resp. $f_r(b)$) will satisfy condition (3.2.4) with $g(y) = y^{-\rho}$.*

In the important special case $a_n = n, \ n = 0, 1, \ldots$, the inverse functions are given by

$$a_l^{-1}(x) \equiv \lceil x \rceil \qquad \text{and} \qquad a_r^{-1}(x) \equiv \lfloor x \rfloor, \qquad x \geq 0, \tag{3.2.10}$$

with $\lceil x \rceil$ (resp. $\lfloor x \rfloor$) denoting the ceiling (resp. floor) of x. In that case, if the function v is regularly varying with parameter ρ, then h and g can immediately be identified as

$$h(b) = v_{\lceil b \rceil} \ (\text{or } v_{\lfloor b \rfloor}) \qquad \text{and} \qquad g(y) = y^{-\rho}, \qquad b, y > 0. \tag{3.2.11}$$

3.2.1 The Large Deviations Principle

The sequence $\{S_n/a_n, \ n = 1, 2, \ldots\}$ is said to satisfy the *Large Deviations Principle under scaling* v_n if there exists a lower–semicontinuous function $I : \mathbb{R} \to [0, \infty]$ such that for every open set G,

$$-\inf_{x \in G} I(x) \leq \liminf_{n \to \infty} \frac{1}{v_n} \ln \mathbf{P}\left[S_n/a_n \in G\right] \tag{3.2.12}$$

and for every closed set F,

$$\limsup_{n \to \infty} \frac{1}{v_n} \ln \mathbf{P}\left[S_n/a_n \in F\right] \leq -\inf_{x \in F} I(x). \tag{3.2.13}$$

We refer to (3.2.12) and (3.2.13) as the Large Deviations lower and upper bounds, respectively. The *rate function* I is said to be *good* if for each $r > 0$, the level set $\{x \in \mathbb{R} : I(x) \leq r\}$ is a compact subset of $\mathbb{R}$. Additional information on Large Deviations can be found in [15].

The existence of (3.2.12), and for that matter of (3.2.13), is typically validated through the Gärtner–Ellis Theorem [15, Thm. 2.3.6, p. 45]: In that context, for each $n = 1, 2, \ldots$, we introduce

$$\Lambda_n(\theta) \equiv \frac{1}{v_n} \ln \mathbf{E} \left[\exp \left(\theta S_n \right) \right], \qquad \theta \in \mathbb{R}. \tag{3.2.14}$$

The limit

$$\Lambda(\theta) \equiv \lim_{n \to \infty} \Lambda_n(\theta(n)), \qquad \theta \in \mathbb{R} \tag{3.2.15}$$

with $\theta(n) \equiv \theta v_n / a_n$, is assumed to exist (possibly as an extended real number). We also define the *Legendre–Fenchel transform* of the mapping $\Lambda : \mathbb{R} \to (-\infty, \infty]$ by

$$\Lambda^\star(z) \equiv \sup_{\theta \in \mathbf{R}} \left(\theta z - \Lambda(\theta) \right), \qquad z \in \mathbb{R}. \tag{3.2.16}$$

A few relevant properties of mappings Λ and $\Lambda^\star$ are stated next; the corresponding proofs can be found in [15, Lemma 2.2.5, p. 27].

Lemma 3.2.3 *Consider a stationary sequence* $\{\xi_i,\ i = 1, 2, \ldots\}$ *with partial sums* $S_n = \sum_{i=1}^{n} \xi_i,\ n = 1, 2, \ldots$. *Let* Λ *and* $\Lambda^\star$ *be defined by identities (3.2.14)-(3.2.16). The following properties then hold.*

(i) Λ *and* $\Lambda^\star$ *are both convex;*

(ii) (a) *If* $\Lambda(\theta) < \infty$ *for some* $\theta > 0$ *, then* $\mathbf{E}[\xi_1] < \infty$*, and*

$$\Lambda^\star(z) = \sup_{\theta \geq 0} \left(\theta z - \Lambda(\theta) \right), \qquad z \geq \mathbf{E}[\xi_1] \tag{3.2.17}$$

with $\Lambda^\star(z)$ *non–decreasing in the range* $z \geq \mathbf{E}[\xi_1]$*;*

(b) *If* $\Lambda(\theta) < \infty$ *for some* $\theta < 0$ *, then* $\mathbf{E}[\xi_1] > -\infty$*, and*

$$\Lambda^\star(z) = \sup_{\theta \leq 0} (\theta z - \Lambda(\theta)), \qquad z \leq \mathbf{E}[\xi_1] \tag{3.2.18}$$

with $\Lambda^\star(z)$ *non–increasing in the range* $z \leq \mathbf{E}[\xi_1]$*;*

(iii) *When* $\mathbf{E}[\xi_1]$ *is finite,*

$$\inf_{z \in \mathbf{R}} \Lambda^\star(z) \;=\; \Lambda^\star(\mathbf{E}[\xi_1]) \;=\; 0. \tag{3.2.19}$$

When $\mathbf{E}[\xi_1] < 0$, the convexity of Λ, along with Jensen's inequality and the fact $\Lambda(0) = 0$, ensure that Λ only has non–negative roots, and that it subscribes to one of the following patterns of behavior:

(F1) It has a double root at $\theta = 0$;

(F2) It has a root at some finite $\theta' > 0$;

(F3) $\Lambda(\theta) = \infty,\ \theta > 0$;

(F4) $\Lambda(\theta) < 0,\ 0 < \theta < \theta'$, and $\Lambda(\theta) = \infty,\ \theta > \theta'$ for some finite $\theta' > 0$.

3.2.2 The Gärtner–Ellis Theorem

Under certain conditions, the process $\{S_n/a_n,\ n = 1, 2, \ldots\}$ satisfies the Large Deviations Principle under scaling v_n and with good rate function $\Lambda^\star : \mathbb{R} \to [0, \infty]$. The Gärtner–Ellis Theorem, reproduced below from [15], provides one such set of conditions.

Theorem 3.2.1 *The process $\{S_n/a_n,\ n = 1, 2, \ldots\}$ satisfies the Large Deviations Principle under scaling v_n and with good rate function $\Lambda^\star : \mathbb{R} \to [0, \infty]$, if the following conditions hold.*

GE 1] *The limit (3.2.15) exists as an extended real number;*

GE 2] *The origin belongs to the interior D_Λ^o of the set $D_\Lambda = \{\theta \in \mathbb{R} : \Lambda(\theta) < \infty\}$;*

GE 3] *Λ is essentially smooth, i.e.,*

- **(i)** *D_Λ^o is non-empty,*
- **(ii)** *Λ is differentiable throughout D_Λ^o,*
- **(iii)** *Λ is steep, i.e.,*

$$\lim_{i\to\infty} \left| \frac{\partial \Lambda(\theta_i)}{\partial \theta_i} \right| = \infty \tag{3.2.20}$$

for every sequence $\{\theta_i : \theta_i \in D_\Lambda^o,\ i = 1, 2, \ldots\}$ that converges to a boundary point of D_Λ.

We are now ready to derive results of the form (3.0.2) and (3.0.3), and we do so forthwith.

3.3 The lower bound

The following theorem is essentially due to Duffield and O'Connell [16]; a proof is included here for the sake of completeness:

Proposition 3.3.1 *If the process $\{S_n/a_n,\ n = 1, 2, \ldots\}$ satisfies the Large Deviations Principle with good rate function $I : \mathbb{R} \to [0, \infty]$ under scaling v_n, then, for* *each $y > 0$ we have*

$$\liminf_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \geq -g(y) \inf_{x>y} I(x). \tag{3.3.1}$$

Proof. Fix $b > 0$ and $y > 0$. From the definition of q_∞, we have

$$\begin{aligned}
\mathbf{P}\left[q_\infty > b\right] &= \mathbf{P}\left[\sup_{n=0,1,\ldots} S_n > b\right] \\
&\geq \mathbf{P}\left[\frac{S_{a_l^{-1}(b/y)}}{a_{a_l^{-1}(b/y)}} > \frac{b}{a_{a_l^{-1}(b/y)}}\right] \\
&\geq \mathbf{P}\left[\frac{S_{a_l^{-1}(b/y)}}{a_{a_l^{-1}(b/y)}} > y\right],
\end{aligned}$$

where we have employed the second inequality in (3.2.3) with $x = b/y$. Consequently

$$\frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \geq \frac{v_{a_l^{-1}(b/y)}}{h(b)} \cdot \frac{1}{v_{a_l^{-1}(b/y)}} \ln \mathbf{P}\left[\frac{S_{a_l^{-1}(b/y)}}{a_{a_l^{-1}(b/y)}} > y\right], \tag{3.3.2}$$

and letting b go to infinity in this last inequality, we readily get (3.3.1) from (3.2.4) and the lower bound (3.2.12) (with $G = (-\infty, y)$). ■

This is essentially Theorem 2.1 of [16], and shows the local nature of the lower bound. As the best lower bound is the largest, we can immediately sharpen (3.3.1) into the lower bound (3.0.2) as stated in the following Proposition.

Proposition 3.3.2 *Under the assumptions of Proposition 3.3.1, we have the lower bound (3.0.2) with*

$$\gamma_\star \equiv \inf_{y>0} \left(g(y) \inf_{x>y} I(x) \right). \tag{3.3.3}$$

The expression (3.3.3) simplifies when the Large Deviations Principle for the process $\{S_n/a_n,\ n = 1, 2, \ldots\}$ holds with a good rate function $I : \mathbb{R} \to [0, \infty]$ which is *convex*. Indeed, the relation

$$\inf_{x \in \mathbf{R}} I(x) = I(\mathbf{E}[\xi_1]) \tag{3.3.4}$$

follows readily from the goodness of I and the fact that $\lim_{n\to\infty} n^{-1}S_n = \mathbf{E}[\xi_1] < 0$ a.s. under the ergodic assumption. However, by convexity we have I increasing (resp. decreasing) on $(\mathbf{E}[\xi_1], \infty)$ (resp. on $(-\infty, \mathbf{E}[\xi_1])$, and the conclusion

$$\inf_{x>y} I(x) = I(y+) \tag{3.3.5}$$

holds for all $y > 0$. Thus (3.3.3) becomes

$$\gamma_\star = \inf_{y>0} g(y)I(y+) = \inf_{0<y<y^\star} g(y)I(y), \tag{3.3.6}$$

for some $0 \leq y^\star \leq \infty$. The non–degeneracy condition $y^\star > 0$ holds in most applications.

For the special case when the Gärtner–Ellis Theorem applies, the rate function I is given by the function $\Lambda^\star$, which, being convex, allows the following representation of the lower bound.

Proposition 3.3.3 *If the process $\{S_n/a_n,\ n = 1, 2, \ldots\}$ satisfies the conditions of the Gärtner–Ellis Theorem under scaling v_n, then, we have the lower bound (3.0.2) with*

$$\gamma_\star = \inf_{y>0} g(y)\Lambda^\star(y). \tag{3.3.7}$$

3.4 The upper bound

We now turn to the derivation of the companion upper bound (3.0.3). In [16], such an upper bound, was derived under a set of conditions which, though reasonably general, do not apply in certain situations of interest to us. Furthermore, the proof provided was not entirely accurate. Upon refining the arguments of [16], we have established a similar asymptotic upper bound which we present in this section. More recently, an alternative approach was given in [17]; however the expressions obtained here being simpler in form, are rather more conveniently applied.

The discussion that follows is considerably more involved than that of the lower bound as shown in the previous section. To simplify matters, we present the derivation in three parts: In the first, we establish the basic asymptotic upper bound; its various terms are then studied in greater details in Sections 3.4.2 and 3.4.3.

3.4.1 The basic upper bound

For each $m = 1, 2, \ldots$ and $b > 0$, we define the quantities

$$A(m,b) \equiv m \max_{n=1,\ldots,m} \mathbf{P}\left[S_n > b\right] \tag{3.4.1}$$

and

$$B(m,b) \equiv \sum_{n=m+1}^{\infty} \mathbf{P}\left[S_n > b\right]. \tag{3.4.2}$$

From the representation (3.1.4) we readily see that

$$\begin{aligned} \mathbf{P}\left[q_\infty > b\right] &= \mathbf{P}\left[\sup_{n=0,1,\ldots} S_n > b\right] \\ &\leq m \max_{n=1,\ldots,m} \mathbf{P}\left[S_n > b\right] + \sum_{n=m+1}^{\infty} \mathbf{P}\left[S_n > b\right] \\ &= A(m,b) + B(m,b) \\ &\leq 2 \max\Big(A(m,b), B(m,b)\Big). \end{aligned} \tag{3.4.3}$$

For a fixed $y > 0$, we substitute $m = a_r^{-1}(b/y)$ in (3.4.3), thus obtaining

$$\frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \quad \leq \quad \frac{1}{h(b)} \ln \left(2 \max\Big(A(a_r^{-1}(b/y), b), B(a_r^{-1}(b/y), b)\Big)\right).$$

Letting b go to infinity in the resulting inequality, we achieve the basic asymptotic upper bound

$$\limsup_{b \to \infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \leq \max\left(\alpha(y), \beta(y)\right) \tag{3.4.4}$$

where we have used the notation

$$\alpha(y) \equiv \limsup_{b \to \infty} \frac{1}{h(b)} \ln A(a_r^{-1}(b/y), b) \tag{3.4.5}$$

and

$$\beta(y) \equiv \limsup_{b \to \infty} \frac{1}{h(b)} \ln B(a_r^{-1}(b/y), b). \tag{3.4.6}$$

We conclude this section with a couple of assumptions required for further analysis:

Assumption A1: *For each θ in $\mathbb{R}$, the limit (3.2.15) exists (possibly as an extended real number) with*

$$-\infty < \inf_{\theta > 0} \Lambda(\theta) < \infty;$$

Assumption A2: *For some finite $\kappa \geq 0$, we have*

$$\lim_{n \to \infty} \frac{\ln n}{v_n} = \kappa. \tag{3.4.7}$$

If $\kappa = 0$, we further assume that the sequence $\{\ln n / v_n,\ n = 1, 2, \ldots\}$ is eventually decreasing, i.e., there exists a finite integer $N^\star$ such that

$$\frac{\ln(n+1)}{v_{n+1}} \leq \frac{\ln n}{v_n}, \qquad n \geq N^\star.$$

We note that the case $\kappa = 0$ is equivalent to Hypothesis 2.2(iv) in [16].

In the next two sections we develop upper bounds on each of the terms $\alpha(y)$ and $\beta(y)$ in terms of of quantities which are conveniently derived from the statistics of $\{\xi_n,\ n = 1, 2, \ldots\}$, and which can be easily related to the expressions for the lower bound.

3.4.2 An upper bound on $\alpha(y)$

Going back to (3.4.1) we note that

$$\ln A(a_r^{-1}(b/y), b) \;=\; \ln a_r^{-1}(b/y) + \max_{n=1,\ldots,a_r^{-1}(b/y)} \ln \mathbf{P}\left[S_n > b\right] \tag{3.4.8}$$

for each $b > 0$ and $y > 0$.

Lemma 3.4.1 *Under Assumption* **A2** *we have*

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln a_r^{-1}(b/y) = \kappa g(y), \qquad y > 0. \tag{3.4.9}$$

Proof. For each $y > 0$ we have

$$\begin{aligned}
\limsup_{b\to\infty} \frac{1}{h(b)} \ln a_r^{-1}(b/y) &= \limsup_{b\to\infty} \left(\frac{v_{a_r^{-1}(b/y)}}{h(b)} \cdot \frac{1}{v_{a_r^{-1}(b/y)}} \ln a_r^{-1}(b/y) \right) \\
&= \lim_{b\to\infty} \frac{v_{a_r^{-1}(b/y)}}{h(b)} \cdot \limsup_{b\to\infty} \frac{1}{v_{a_r^{-1}(b/y)}} \ln a_r^{-1}(b/y) \\
&= g(y) \lim_{n\to\infty} \frac{\ln n}{v_n}
\end{aligned}$$

and the desired conclusion follows from Assumption **A2**. ∎

For the second term of (3.4.8), fixing $y > 0$ and $b > 0$, we note that

$$\max_{n=1,\ldots,a_r^{-1}(b/y)} \ln \mathbf{P}\left[S_n > b\right] \;\le\; \sup_{x>y} \ln \mathbf{P}\left[S_{a_r^{-1}(b/x)} > b\right]$$

$$
\begin{aligned}
&= \sup_{x>y} \ln \mathbf{P}\left[\frac{S_{a_r^{-1}(b/x)}}{a_{a_r^{-1}(b/x)}} > \frac{b}{a_{a_r^{-1}(b/x)}}\right] \\
&\le \sup_{x>y} \ln \mathbf{P}\left[\frac{S_{a_r^{-1}(b/x)}}{a_{a_r^{-1}(b/x)}} > x\right] \\
&= \sup_{x>y} \left(v_{a_r^{-1}(b/x)} \cdot \frac{1}{v_{a_r^{-1}(b/x)}} \ln \mathbf{P}\left[\frac{S_{a_r^{-1}(b/x)}}{a_{a_r^{-1}(b/x)}} > x\right]\right)
\end{aligned}
$$

where the last inequality comes about through the first inequality in (3.2.3). It is now easy to see for each $y > 0$, that

$$
\begin{aligned}
&\limsup_{b\to\infty} \frac{1}{h(b)} \max_{n=1,\ldots,a_r^{-1}(b/y)} \ln \mathbf{P}\left[S_n > b\right] \\
\le\; &\limsup_{b\to\infty} \sup_{x>y} \frac{v_{a_r^{-1}(b/x)}}{h(b)} \cdot \frac{1}{v_{a_r^{-1}(b/x)}} \ln \mathbf{P}\left[\frac{S_{a_r^{-1}(b/x)}}{a_{a_r^{-1}(b/x)}} > x\right] \\
\le\; &\limsup_{m\to\infty} \sup_{x>y} \frac{v_m}{h(a_m x)} \cdot \frac{1}{v_m} \ln \mathbf{P}\left[\frac{S_m}{a_m} > x\right],
\end{aligned}
$$

where the last step follows on substituting $m = a_r^{-1}(b/x)$, through the inequality (3.2.3) and monotone increasing nature of the function h.

At this stage, the authors of [16] argue that

$$
\begin{aligned}
&\limsup_{n\to\infty} \sup_{x>y} \frac{v_n}{h(a_n x)} \cdot \frac{1}{v_n} \ln \mathbf{P}\left[\frac{S_n}{a_n} > x\right] \\
\le\; &\limsup_{n\to\infty} \sup_{x>y} \frac{v_n}{h(a_n x)} \cdot (\delta - \inf_{z>x} I(z))
\end{aligned}
\tag{3.4.10}
$$

for any $\delta > 0$, if the process $\{S_n/a_n,\ n = 1, 2, \ldots\}$ satisfies a Large Deviations Principle under scaling v_n with good rate function $I : \mathbb{R} \to [0, \infty]$. It is our contention that (3.4.10) may not be concluded directly via (3.2.13), as is implied in [16]. Indeed, from (3.2.13) we know that for any $\delta > 0$, there exists an integer $n^\star$ such that

$$
\frac{1}{v_n} \ln \mathbf{P}\left[\frac{S_n}{a_n} > x\right] \le -\inf_{z>x} I(z) + \delta, \qquad n > n^\star. \tag{3.4.11}
$$

It is important to realize that $n^\star = n^\star(x,\delta)$. This dependence of $n^\star$ on x precludes taking $\sup_{x>y}$ on both sides of (3.4.11) to obtain (3.4.10). We overcome this technicality through a slightly different approach.

Lemma 3.4.2 *Assume* **A1**. *Then, for each $y > 0$ we have*

$$\begin{aligned}
&\limsup_{b\to\infty} \frac{1}{h(b)} \max_{n=1,\ldots,a_r^{-1}(b/y)} \ln \mathbf{P}\left[S_n > b\right] \\
\leq \quad & -\sup_{\theta>0} \liminf_{n\to\infty} \inf_{x>y} \left(\frac{v_n}{h(xa_n)} (\theta x - \Lambda(\theta)) \right). \qquad (3.4.12)
\end{aligned}$$

Proof. Fix $x > 0$. For each $\theta > 0$, the usual Chernoff bound argument gives

$$\begin{aligned}
\frac{1}{v_n} \ln \mathbf{P}\left[\frac{S_n}{a_n} > x\right] &\leq \frac{1}{v_n} \ln\left(\mathbf{E}\left[e^{\theta v_n \frac{S_n}{a_n}}\right] e^{-\theta x v_n}\right) \\
&= -\theta x + \Lambda_n(\theta), \qquad n = 1, 2, \ldots \qquad (3.4.13)
\end{aligned}$$

Under Assumption **A1**, if $\Lambda(\theta)$ is finite, then for each $\delta > 0$, there exists an integer $n^\star = n^\star(\theta,\delta)$ such that

$$\Lambda(\theta) - \delta \leq \Lambda_n(\theta) \leq \Lambda(\theta) + \delta, \qquad n \geq n^\star. \qquad (3.4.14)$$

Reporting this fact into the Chernoff bound (3.4.13), we get

$$\frac{1}{v_n} \ln \mathbf{P}\left[\frac{S_n}{a_n} > x\right] \leq -\theta x + \Lambda(\theta) + \delta, \qquad n \geq n^\star.$$

Consequently, for $y > 0$ and $n \geq n^\star$, we see that

$$\begin{aligned}
\sup_{x>y} \frac{1}{h(xa_n)} \ln \mathbf{P}\left[\frac{S_n}{a_n} > x\right] &\leq \sup_{x>y} \frac{v_n}{h(xa_n)} \left(\delta - \theta x + \Lambda(\theta)\right) \\
&\leq \sup_{x>y} \frac{v_n}{h(xa_n)} \delta - \inf_{x>y} \frac{v_n}{h(xa_n)} (\theta x - \Lambda(\theta)) \\
&\leq \frac{v_n}{h(ya_n)} \delta - \inf_{x>y} \frac{v_n}{h(xa_n)} (\theta x - \Lambda(\theta)). \qquad (3.4.15)
\end{aligned}$$

It is now apparent via Lemma 3.2.1 that

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{x>y}\frac{1}{h(xa_n)}\ln\mathbf{P}\left[\frac{S_n}{a_n}>x\right]\\ \le\ & g(y)\delta-\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{v_n}{h(xa_n)}(\theta x-\Lambda(\theta))\right),\end{aligned}$$

whence

$$\begin{aligned}&\limsup_{n\to\infty}\sup_{x>y}\frac{1}{h(xa_n)}\ln\mathbf{P}\left[\frac{S_n}{a_n}>x\right]\\ \le\ & -\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{v_n}{h(xa_n)}(\theta x-\Lambda(\theta))\right)\end{aligned}$$

because δ can be made arbitrarily small. The least upper bound being the sharpest, we readily conclude (3.4.12). ■

Combining Lemmas 3.4.1 and 3.4.2 we conclude from (3.4.8) to the following upper bound on $\alpha(y)$.

Proposition 3.4.1 *Under Assumptions* **A1** *and* **A2** *we have* $\alpha(y)\le\alpha_U(y)$ *for each* $y>0$ *with*

$$\alpha_U(y)\ \equiv\ g(y)\kappa-\sup_{\theta>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{v_n}{h(xa_n)}(\theta x-\Lambda(\theta))\right).\qquad(3.4.16)$$

3.4.3 An upper bound on $\beta(y)$

As in the proof of Lemma 3.4.2, we begin with a simple Chernoff bound argument. Fix $b>0$ and $\theta>0$. This time, we have

$$\mathbf{P}\left[S_n>b\right]\le e^{-\theta\frac{v_n}{a_n}b}\mathbf{E}\left[e^{\theta\frac{v_n}{a_n}S_n}\right]\le e^{v_n\Lambda_n(\theta)},\qquad n=1,2,\ldots.\qquad(3.4.17)$$

Under Assumption **A1**, if $\Lambda(\theta)$ is finite, then for each $\delta > 0$, there exists a finite integer $n^\star = n^\star(\theta, \delta)$ such that (3.4.14) holds. Hence, for each $y > 0$, it follows from (3.4.2) that

$$\begin{aligned}
\beta(y) &= \limsup_{b\to\infty} \frac{1}{h(b)} \ln \left(\sum_{n=a_r^{-1}(b/y)+1}^{\infty} \mathbf{P}\left[S_n > b\right] \right) \\
&\leq \limsup_{b\to\infty} \frac{1}{h(b)} \ln \left(\sum_{n=a_r^{-1}(b/y)+1}^{\infty} e^{(\Lambda(\theta)+\delta)v_n} \right) \\
&= \limsup_{b\to\infty} \frac{v_{a_r^{-1}(b/y)}}{h(b)} \frac{1}{v_{a_r^{-1}(b/y)}} \cdot \ln \left(\sum_{n=a_r^{-1}(b/y)+1}^{\infty} e^{(\Lambda(\theta)+\delta)v_n} \right) \\
&= g(y) \limsup_{b\to\infty} \frac{1}{v_{a_r^{-1}(b/y)}} \ln \left(\sum_{n=a_r^{-1}(b/y)+1}^{\infty} e^{(\Lambda(\theta)+\delta)v_n} \right) \\
&= g(y) \limsup_{m\to\infty} \frac{1}{v_m} \ln \left(\sum_{n=m+1} e^{(\Lambda(\theta)+\delta)v_n} \right). \qquad (3.4.18)
\end{aligned}$$

Setting

$$L(\gamma) \equiv \limsup_{m\to\infty} \frac{1}{v_m} \ln \left(\sum_{n=m+1}^{\infty} e^{-\gamma v_n} \right), \qquad \gamma \in \mathbb{R} \qquad (3.4.19)$$

we can rephrase (3.4.18) as

$$\beta(y) \leq g(y) L(-(\Lambda(\theta) + \delta)), \qquad y > 0. \qquad (3.4.20)$$

The remainder of the discussion hinges on the following expression for (3.4.19) which shows that in (3.4.20) we need only be concerned with the situation $\Lambda(\theta) + \delta < 0$:

Lemma 3.4.3 *Under Assumption* **A2**, *we have*

$$L(\gamma) = \begin{cases} \kappa - \gamma & \text{if } \gamma > \kappa \\ \infty & \text{if } \gamma < \kappa. \end{cases} \qquad (3.4.21)$$

When $\kappa = 0$, (3.4.21) reads $L(\gamma) = -\gamma$ for all $\gamma > 0$. When $\kappa > 0$, the boundary case $\gamma = \kappa$ depends on the finer structure of the sequence $\{v_n/\ln n,\ n = 1, 2, \ldots\}$. The proof of Lemma 3.4.3 is given in Appendix B.1.

Proposition 3.4.2 *Assumptions* **A1** *and* **A2** *are in force. If* $\Lambda^\star(0) > \kappa$*, then for each* $y > 0$ *we have* $\beta(y) \le \beta_U(y)$*, with*

$$\beta_U(y) \equiv g(y)\left(\kappa - \Lambda^\star(0)\right). \tag{3.4.22}$$

If $\Lambda^\star(0) < \kappa$*, then* $\beta_U(y) = \infty$*, yielding a trivial bound.*

For $\kappa = 0$, (3.4.22) becomes

$$\beta_U(y) = -g(y)\Lambda^\star(0), \qquad y > 0. \tag{3.4.23}$$

Proof. We first note that

$$-\Lambda^\star(0) = \inf_{\theta \in \mathbf{R}} \Lambda(\theta) = \inf_{\theta > 0} \Lambda(\theta), \tag{3.4.24}$$

where the last equality made use of the fact that $\mathbf{E}[\xi_1] < 0$ [15].

Lemma 3.4.3 yields

$$L(-(\Lambda(\theta)+\delta)) = \begin{cases} \kappa + (\Lambda(\theta)+\delta) & \text{if } -(\Lambda(\theta)+\delta) > \kappa \\ \infty & \text{if } -(\Lambda(\theta)+\delta) < \kappa. \end{cases} \tag{3.4.25}$$

Consequently, for each $y > 0$, we see from (3.4.20) that

$$\beta(y) \le g(y) \inf_{\delta > 0} \left(\inf\{\kappa + \Lambda(\theta) + \delta :\ \theta \in \Theta(\delta)\}\right) \tag{3.4.26}$$

where

$$\Theta(\delta) \equiv \{\theta > 0 : \kappa + \Lambda(\theta) + \delta < 0\}, \qquad \delta > 0. \tag{3.4.27}$$

In view of (3.4.24), if $\Lambda^\star(0) \leq \kappa$, then the sets $\Theta(\delta)$ are all empty and the right hand–side of (3.4.26) is ∞, justifying the comment that the bound so obtained is vacuous. When $\Lambda^\star(0) > \kappa$, $\Theta(\delta)$ is not empty for δ in some non–degenerate interval $(0, \delta^\star)$. Therefore, (3.4.26) becomes

$$\beta(y) \leq g(y) \inf_{0<\delta<\delta^\star} \left(\inf_{\theta>0} (\kappa + \Lambda(\theta) + \delta) \right), \qquad y > 0 \tag{3.4.28}$$

and the conclusion of Proposition 3.4.2 follows. ■

3.4.4 The upper bound

We are now ready to combine Propositions 3.4.1 and 3.4.2.

Proposition 3.4.3 *Assuming conditions* **A1** *and* **A2***, we have*

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \leq -\gamma^\star(y), \tag{3.4.29}$$

for each $y > 0$*, with*

$$\gamma^\star(y) = \min\left(\sup_{\theta>0} \liminf_{n\to\infty} \inf_{x>y} \left(\frac{v_n}{h(xa_n)} (\theta x - \Lambda(\theta)) \right), \Lambda^\star(0) g(y) \right) - \kappa g(y). \tag{3.4.30}$$

Proof. The proof follows from (3.4.4) and from Propositions 3.4.1 and 3.4.2 with

$$\gamma^\star(y) \quad = \quad \max\left(\alpha(y), \beta(y)\right).$$

■

As the least upper bound is the sharpest, under the assumptions of Proposition 3.4.3 we immediately get (3.0.3) with

$$\gamma^\star \quad = \quad -\sup_{y>0} \gamma^\star(y). \tag{3.4.31}$$

In many situations of interest, the function h appearing in (3.2.4) satisfies the following condition:

Assumption A3: *The mapping $x \to h(x)/x$ is eventually non–increasing on $(0,\infty)$, i.e., there exists an $x^\star > 0$ such that*

$$\frac{h(x)}{x} \leq \frac{h(y)}{y}, \qquad x \geq y \geq x^\star. \tag{3.4.32}$$

In these cases, the upper bound given by (3.4.30) may be weakened, as indicated by the the following Proposition.

Proposition 3.4.4 *Assuming conditions* **A1**, **A2** *and* **A3** *are in force, we have*

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \quad \leq \quad -g(y)\left(\min\Big(\theta_0 y, \Lambda^\star(0)\Big) - \kappa\right), \tag{3.4.33}$$

for each $y > 0$, with

$$\theta_0 = \sup\{\theta > 0 : \Lambda(\theta) < 0\}. \tag{3.4.34}$$

Proof. Fix $y > 0$. Then, for every $\theta > 0$, we have

$$\begin{aligned}
&\inf_{x>y} \frac{v_n}{h(xa_n)}\Big(\theta x - \Lambda(\theta)\Big) \\
= \;& \frac{v_n}{a_n} \inf_{x>y} \frac{\theta x a_n - \Lambda(\theta) a_n}{h(xa_n)} \\
\geq \;& \theta \frac{v_n}{a_n} \inf_{x>y} \frac{xa_n}{h(xa_n)} + v_n \inf_{x>y} \frac{-\Lambda(\theta)}{h(xa_n)}
\end{aligned}$$

$$= \theta \frac{v_n}{a_n} \inf_{x>y} \frac{x a_n}{h(x a_n)} + v_n \begin{cases} \frac{-\Lambda(\theta)}{h(x a_n)}\Big|_{x=y} & \text{if } \Lambda(\theta) \geq 0 \\ 0 & \text{if } \Lambda(\theta) \leq 0 \end{cases} \tag{3.4.35}$$

$$= \theta \frac{v_n}{a_n} \inf_{x>y} \frac{x a_n}{h(x a_n)} - (\Lambda(\theta))^+ \frac{v_n}{h(y a_n)}, \tag{3.4.36}$$

where (3.4.35) follows through the monotone increasing nature of the mapping $h : (0, \infty) \to \mathbb{R}_+$. Applying Assumption **A3**, we may rewrite (3.4.36) with $y > x^\star$ (with $x^\star$ appearing in (3.4.32)) as

$$\begin{aligned} \inf_{x>y} \frac{v_n}{h(x a_n)} \Big(\theta x - \Lambda(\theta)\Big) &\geq \theta \frac{v_n}{a_n} \frac{y a_n}{h(y a_n)} - (\Lambda(\theta))^+ \frac{v_n}{h(y a_n)} \\ &= \Big(\theta y - (\Lambda(\theta))^+\Big) \frac{v_n}{h(y a_n)}, \end{aligned}$$

whence it follows that

$$\begin{aligned} &\liminf_{n\to\infty} \inf_{x>y} \frac{v_n}{h(x a_n)} \Big(\theta x - \Lambda(\theta)\Big) \\ \geq\ &\liminf_{n\to\infty} \Big(\theta y - (\Lambda(\theta))^+\Big) \frac{v_n}{h(y a_n)} \\ =\ &\Big(\theta y - (\Lambda(\theta))^+\Big) \begin{cases} \liminf_{n\to\infty} \frac{v_n}{h(y a_n)} & \text{if } \theta y \geq (\Lambda(\theta))^+ \\ \limsup_{n\to\infty} \frac{v_n}{h(y a_n)} & \text{if } \theta y \leq (\Lambda(\theta))^+ \end{cases} \\ =\ &\Big(\theta y - (\Lambda(\theta))^+\Big) g(y) \end{aligned}$$

via Lemma 3.2.1.

The desired result now follows through Proposition 3.4.3 with

$$\gamma^\star(y) \ \geq\ g(y) \left(\min \left(\sup_{\theta>0} \Big(\theta y - (\Lambda(\theta))^+\Big), \Lambda^\star(0) \right) - \kappa \right), \tag{3.4.37}$$

and through the fact that

$$\begin{aligned} \sup_{\theta>0} \Big(\theta y - (\Lambda(\theta))^+\Big) &= \max \left(\sup_{\theta>0, \Lambda(\theta)<0} \theta y, \sup_{\theta>0, \Lambda(\theta)\geq 0} (\theta y - \Lambda(\theta)) \right) \\ &= \max \left(\theta_0 y, \sup_{\theta>0, \Lambda(\theta)\geq 0} (\theta y - \Lambda(\theta)) \right) \\ &\geq \theta_0 y, \end{aligned}$$

where θ_0 is given by (3.4.34). ■

3.5 Special Cases

Under additional conditions the bounds can be still further modified. We address two special cases: In the first case we have $g(y) = 1,\ y > 0$, suggesting a slowly-varying scaling function v_n, while the second is the familiar and well-researched case $v_n = a_n = n$. We assume **A3** holds in both cases.

Case 1: **A3** holds; $g(y) = 1,\ y > 0$. Applying Proposition 3.4.4, and maximizing over $y > 0$ in order to achieve the smallest upper bound, we have

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \quad \leq \quad \kappa - \sup_{y>0} \Big(\min\Big(\theta_0 y, \Lambda^\star(0)\Big)\Big) \, ,$$

with θ_0 given by (3.4.34). The function $\min\Big(\theta_0 y, \Lambda^\star(0)\Big)$ being non–decreasing in y, we easily conclude to the bound

$$\limsup_{b\to\infty} \frac{1}{h(b)} \ln \mathbf{P}\left[q_\infty > b\right] \quad \leq \quad \kappa - \Lambda^\star(0). \tag{3.5.38}$$

Further, if the conditions of the Gärtner–Ellis Theorem hold, then by Proposition 3.3.3, we have the lower bound

$$\gamma_\star \quad = \quad \inf_{y>0} \Lambda^\star(y) \quad = \quad \Lambda^\star(0). \tag{3.5.39}$$

Case 2: $v_n = a_n = n,\ n = 1, 2, \ldots$. This condition automatically suggests the selection $h(b) = b,\ b > 0$, whereupon (3.2.4) yields $g(y) = \frac{1}{y},\ y > 0$. Both Assumptions **A2** and **A3** hold, and $\kappa = 0$. Applying Proposition 3.4.4, we have the upper bound

$$\gamma^\star(y) \quad \geq \quad \frac{1}{y} \min\Big(\theta_0 y, \Lambda^\star(0)\Big), \qquad y > 0 \tag{3.5.40}$$

where θ_0 is given by (3.4.34). Maximizing over $y > 0$ in order to achieve the smallest upper bound yields

$$\gamma_\star \;=\; \sup_{y>0} \gamma^\star(y) \;\geq\; \min\left\{\theta_0, \sup_{y>0} \frac{\Lambda^\star(0)}{y}\right\} \;=\; \theta_0. \tag{3.5.41}$$

The corresponding lower bound may also be derived under certain conditions. By Proposition 3.3.3, we know that (3.0.2) holds with

$$\gamma_\star \;=\; \inf_{y>0} \frac{\Lambda^\star(y)}{y}, \tag{3.5.42}$$

assuming that **GE 1-3** of the Gärtner–Ellis Theorem hold. We already know via (3.5.41) that

$$\gamma_\star \;\geq\; \gamma^\star \;\geq\; \theta_0. \tag{3.5.43}$$

From the remarks following Lemma 3.2.3, we know that Λ exhibits one of four possible forms of behavior, **F1 - F4**. Of these, **F3** is ruled out by assumption **GE 2** of the Gärtner–Ellis Theorem.

When Λ obeys **F4**, $\Lambda(\theta) = \infty,\ \theta' < \theta$, and we have

$$\begin{aligned}
\sup_{\theta>0}\left(\theta - \frac{\Lambda(\theta)}{y}\right) &= \max\left(\sup_{0<\theta\leq\theta'}\left(\theta - \frac{\Lambda(\theta)}{y}\right),\ \sup_{\theta'<\theta}\left(\theta - \frac{\Lambda(\theta)}{y}\right)\right) \\
&= \sup_{0<\theta\leq\theta'}\left(\theta - \frac{\Lambda(\theta)}{y}\right),
\end{aligned}$$

in which case (3.5.42) easily simplifies as

$$\begin{aligned}
\gamma_\star &= \inf_{y>0}\ \sup_{\theta>0}\left(\theta - \frac{\Lambda(\theta)}{y}\right) \\
&= \inf_{y>0}\ \sup_{0<\theta\leq\theta'}\left(\theta - \frac{\Lambda(\theta)}{y}\right).
\end{aligned}$$

As $\Lambda(\theta) < 0$ in the range $0 < \theta \leq \theta'$, the term $\sup_{0<\theta\leq\theta'}\left(\theta - \frac{\Lambda(\theta)}{y}\right)$ is decreasing in y, whence,

$$\gamma_\star \;=\; \lim_{y\to\infty}\ \sup_{0<\theta\leq\theta'}\left(\theta - \frac{\Lambda(\theta)}{y}\right)$$

$$
\begin{aligned}
&= \sup_{0<\theta\leq\theta'} \theta \\
&= \theta', \qquad (3.5.44)
\end{aligned}
$$

where the second inequality follows through the *uniform* convergence of the function $1/y$ to 0, as y tends to infinity. Comparing **F4** with (3.4.34), we immediately conclude that $\theta_0 = \theta'$ in this case, therefore implying that $\gamma_\star = \theta_0$. Referring to (3.5.43) we realize in fact that $\gamma_\star = \gamma^\star = \theta_0$.

In case **F1**, we know via (3.4.34) that $\theta_0 = 0$. $\Lambda(\theta)$ being positive for all $\theta > 0$, the function $\sup_{0<\theta<\theta'} \left(\theta - \frac{\Lambda(\theta)}{y}\right)$ is now increasing in y. Therefore, $\gamma_\star = 0$ trivially, and via (3.5.43) we arrive at the same result as before: $\gamma_\star = \gamma^\star = \theta_0$.

We now only need investigate the case **F2**, where the equation $\Lambda(\theta) = 0$ has roots 0 and θ_0. By (3.5.42), we know that

$$
\gamma_\star \leq \sup_{\theta>0} \left(\theta - \frac{\Lambda(\theta)}{y}\right)
$$

for all values of $y > 0$.

Consider the particular selection $y = y_0$, where y_0 is the slope of the function $\Lambda(\theta)$ at $\theta = \theta_0$ and is given by

$$
y_0 = \left.\frac{d\Lambda(\theta)}{d\theta}\right|_{\theta=\theta_0}. \qquad (3.5.45)
$$

Under assumption **GE 3 (ii)**, the function Λ is differentiable wherever it takes finite values.

By the convexity of Λ and by (3.5.45), we know that

$$
\arg\sup_{\theta>0} \left(\theta - \frac{\Lambda(\theta)}{y_0}\right) = \theta_0,
$$

and

$$
\sup_{\theta>0} \left(\theta - \frac{\Lambda(\theta)}{y_0}\right) = \theta_0 - \frac{\Lambda(\theta_0)}{y_0} = \theta_0, \qquad (3.5.46)
$$

whereby we conclude that $\gamma_\star \leq \theta_0$. A quick comparison with (3.5.43) again yields the same final result as before, i.e., $\gamma_\star = \gamma^\star = \theta_0$.

In other words, when assumptions **A1** and **GE 1-3** are satisfied for selections $v_n = a_n = n, \; n = 1, 2, \ldots$, we have

$$\lim_{b \to \infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty > b\right] \;\; = \;\; -\theta_0 \tag{3.5.47}$$

where θ_0 is given by (3.4.34), and is identical to the expression derived by Glynn and Whitt in [27].

Chapter 4

Evaluation of $\Lambda(\theta)$, $(\theta \in \mathbb{R})$ for the $M|G|\infty$ process

The previous chapter addressed the issue of buffer–sizing at a switch in a network. Having introduced the $M|G|\infty$ process as a suitable model for network traffic, the next logical step consists in applying the results derived in Chapter 3, to estimate quality of service parameters such as cell–loss and buffer overflow probabilities, in a network supporting $M|G|\infty$ traffic. In this and the following chapters we seek results of the type (3.0.2) and (3.0.3) when the traffic input $\{b_t,\ t = 0, 1, \ldots\}$ into the multiplexer is an $M|G|\infty$ input process.

The log–moment generating function $\Lambda(\theta)$ being central to our analysis, we focus exclusively on its computation in this chapter. To do so, we identify the scaling functions v and a, so that the limit (3.2.15) exists (possibly as an extended real number) for every θ in $\mathbb{R}$. Of course, if $\Lambda(\theta) = \infty$ for *all* $\theta > 0$, then the Legendre–Fenchel transform $\Lambda^\star$ of Λ vanishes on the entire positive half–line and (3.3.6) and (3.4.29) yield vacuous bounds on the probabilities of interest. To guard against such an eventuality, we require $\Lambda(\theta) < \infty$ for some $\theta > 0$, as per assumption **A1** in the previous chapter.

Finally we note that the discussion is carried out under the premise that the limits $\lim_{n\to\infty} \frac{v_n}{n}$ and $\lim_{n\to\infty} \frac{a_n}{n}$ exist; this is a very mild assumption which holds in most situations of interest, in fact in all situations known to the author.

4.1 Evaluation of $\Lambda(\theta)$, $(\theta \in \mathbb{R})$

To state the results more conveniently, we fix $n = 1, 2, \ldots$ and set

$$\Lambda_{b,n}(\theta) \equiv \frac{1}{v_n} \ln \mathbf{E}\left[\exp\left(\theta S_n^b\right)\right], \qquad \theta \in \mathbb{R}, \tag{4.1.1}$$

where $S_n^b = \sum_{i=1}^{n} b_i$. Now, using definition (3.1.3), with $\xi_i = b_i - c$, we rewrite (3.2.14) as

$$\begin{aligned} \Lambda_n(\theta) &= \frac{1}{v_n} \ln \mathbf{E}\left[\exp\left(\theta(S_n^b - cn)\right)\right] \\ &= \Lambda_{b,n}(\theta) - \frac{\theta cn}{v_n}, \qquad \theta \in \mathbb{R}, \end{aligned}$$

whence it follows via (3.2.15) that

$$\begin{aligned} \Lambda(\theta) &\equiv \lim_{n\to\infty} \Lambda_n(\theta(n)) \\ &= \Lambda_{b,n}(\theta(n)) - \theta(n)\frac{cn}{v_n} \\ &= \Lambda_{b,n}(\theta(n)) - \theta c\frac{n}{a_n}, \qquad \theta \in \mathbb{R}, \end{aligned} \tag{4.1.2}$$

where as before, we have used the notation $\theta(n) = \theta v_n / a_n$.

Taking into consideration the fact that the output rate c of the multiplexer would certainly influence the queue–size, we can rule out selections of a that yield $\lim_{n\to\infty} \frac{n}{a_n} = 0$ or ∞. In other words, meaningful bounds only result when

$$\lim_{n\to\infty} \frac{n}{a_n} = \alpha, \qquad 0 < \alpha < \infty, \tag{4.1.3}$$

in which case, if the limit

$$\Lambda_b(\theta) \equiv \lim_{n\to\infty} \Lambda_{b,n}(\theta(n)), \qquad \theta \in \mathbb{R} \tag{4.1.4}$$

exists, so does (4.1.2) with

$$\Lambda(\theta) = \Lambda_b(\theta) - \alpha c\theta, \qquad \theta \in \mathbb{R} \tag{4.1.5}$$

and it suffices to concentrate on finding (4.1.4). As we will see later in Section 4.6, so long as (4.1.3) holds, the particular form of the function a does not alter the essence of our result, nor does the actual value taken by α. We therefore proceed for the remainder of this thesis, with the convenient assumption that $a_n = n,\ n = 1, 2, \ldots$, so that $\alpha = 1$, and $\theta(n) = \theta v_n/n$.

Under this selection, we can already predict via Jensen's inequality and Proposition 2.3.2 that

$$\Lambda_{b,n}(\theta(n)) \geq \lambda \mathbf{E}[\sigma]\,\theta, \quad \theta \in \mathbb{R}$$

for each $n = 1, 2, \ldots$, so that $\Lambda_b(\theta) > -\infty$ (though possibly ∞) when it exists. The mapping $\theta \to \Lambda_b(\theta)$ is non–decreasing and convex, so that $\{\theta \in \mathbb{R} : \Lambda_b(\theta) = \infty\}$ is an interval of the form $(\theta_\star, \infty)$ or $[\theta_\star, \infty)$ for some $\theta_\star$ in $\mathbb{R} \cup \{\infty\}$.

Further, if $\Lambda_b(\theta) < \infty$ for some $\theta > 0$, then, by (3.2.16) and by Lemma 3.2.3 (ii) (a), we have

$$\begin{aligned} \Lambda^\star(x) &= \sup_{\theta\in\mathbf{R}}(\theta x - \Lambda(\theta)) \\ &= \sup_{\theta>0}(\theta x - \Lambda(\theta)) \\ &= \sup_{\theta>0}(\theta(x + c) - \Lambda_b(\theta)) \end{aligned} \tag{4.1.6}$$

for every $x > \mathbf{E}[\xi_1]$, where $\xi_1 = b_1 - c$ as specified shortly after (4.1.1). However, $\mathbf{E}[\xi_1] = \lambda\mathbf{E}[\sigma] - c = r_{in} - c < 0$ by the stability requirement, and we conclude that for any $x > 0$, $\Lambda^\star(x)$ is given by (4.1.6), as long as $\Lambda_b(\theta) < \infty$ for some $\theta > 0$.

4.2 Evaluation of $\Lambda_{b,n}(\theta)$ ($n = 1, 2, \ldots, \theta \in \mathbb{R}$)

As per the notation introduced in Section 2.2.1, for a sequence $T_n = \{t_i,\ i = 1, 2, \ldots, n\}$ of finite, non-negative, non–decreasing integers, let

$$\mathcal{L}(T_n, Q_n) = \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}\right)\right] < \infty \tag{4.2.1}$$

where $Q_n = \{\theta_i,\ i = 1, 2, \ldots, n\}$ is a sequence of real–valued numbers. (This is essentially (2.1.7) reproduced here for convenience.)

Consider the special case $H_n = (1, 2, \ldots, n)$ and $\widetilde{Q}_n = \{\theta, \ldots, \theta\}$ for some θ in $\mathbb{R}$, so that (4.2.1) now becomes

$$\mathcal{L}(H_n, \widetilde{Q}_n) = \ln \mathbf{E}\left[\exp\left(\theta \sum_{i=1}^{n} b_i\right)\right].$$

Comparison with (4.1.1) easily gives

$$\mathcal{L}(H_n, \widetilde{Q}_n) = v_n \Lambda_{b,n}(\theta), \qquad \theta \in \mathbb{R} \tag{4.2.2}$$

for each $n = 1, 2, \ldots$.

For the results of the Chapter 3 to be applicable, we restrict our attention to the stationary and ergodic version of the busy server process $\{b_t,\ t = 0, 1, \ldots\}$. The form of $\Lambda_{b,n}(\theta)$ (θ in $\mathbb{R},\ n = 1, 2, \ldots$) is then given by Theorem 4.2.1.

Theorem 4.2.1 *Fix θ in $\mathbb{R}$. For each $n = 1, 2, \ldots$, we have*

$$\Lambda_{b,n}(\theta) = \lambda \mathbf{E}[\sigma]\left(e^{\theta} - 1\right)\frac{n}{v_n}\left(1 + (1 - e^{-\theta})\Delta(n, \theta)\right) \tag{4.2.3}$$

where

$$\Delta(n, \theta) = \frac{1}{n}\sum_{r=1}^{n}(n - r)e^{\theta r}\mathbf{P}\left[\widehat{\sigma} > r\right]. \tag{4.2.4}$$

Theorem 4.2.1 follows directly from Lemma 4.2.1 below, which derives an expression for $\mathcal{L}(H_n, \widetilde{Q}_n)$ in the special case $H_n = (1, 2, \ldots, n)$ and $\widetilde{Q}_n = \{\theta, \ldots, \theta\}$ for some θ in $\mathbb{R}$. The proof is simple and has been included in Appendix C.1.

Lemma 4.2.1 *For each $n = 1, 2, \ldots$, and θ in $\mathbb{R}$, we have*

$$\mathcal{L}(H_n, \widetilde{Q}_n) = \lambda \mathbf{E}\left[\sigma\right](e^{\theta} - 1)n\left(1 + (1 - e^{-\theta})\Delta(n, \theta)\right) \tag{4.2.5}$$

with $\Delta(n, \theta)$ defined by (4.2.4).

The expression $\Delta(n, \theta)$, when coaxed into a different form, gives rise to an alternate formulation for $\Lambda_{b,n}(\theta)$ (θ in $\mathbb{R}$, $n = 1, 2, \ldots$), which proves both insightful and convenient for further analysis. This expression is given in Theorem 4.2.2, with the details of the proof available in Appendix C.2.

Theorem 4.2.2 *Fix θ in $\mathbb{R}$. For each $n = 1, 2, \ldots$, we have*

$$\Lambda_{b,n}(\theta) = \frac{\lambda}{v_n}\left(\mathcal{D}_n(\sigma, \theta) - 2\mathbf{E}\left[\min(n, \sigma)\right]\right),$$

where

$$\mathcal{D}_n(\sigma, \theta) \equiv \mathbf{E}\left[(n - \sigma)^+\left(e^{\theta\sigma} - 1\right)\right] + \left(\frac{e^{\theta} + 1}{e^{\theta} - 1}\right)\mathbf{E}\left[e^{\theta \min(n, \sigma)} - 1\right] + \mathbf{E}\left[\sigma\right]\mathbf{P}\left[\widehat{\sigma} > n\right]\left(e^{\theta n} - 1\right). \tag{4.2.6}$$

With definition (4.1.4) in mind, for each θ in $\mathbb{R}$, Lemma 4.2.1 gives

$$\Lambda_b(\theta) = \lambda \mathbf{E}\left[\sigma\right]\lim_{n\to\infty}\left(e^{\theta\frac{v_n}{n}} - 1\right)\frac{n}{v_n}\left(1 + \left(1 - e^{-\theta\frac{v_n}{n}}\right)\Delta\left(n, \theta\frac{v_n}{n}\right)\right), \tag{4.2.7}$$

while Theorem 4.2.2 yields

$$\Lambda_b(\theta) = \lambda \lim_{n\to\infty}\frac{1}{v_n}\mathcal{D}_n\left(\sigma, \theta\frac{v_n}{n}\right), \tag{4.2.8}$$

provided these limits exist, since for *any* scaling v, we have

$$0 \leq \lim_{n\to\infty} \frac{1}{v_n}\mathbf{E}\left[\min(n,\sigma)\right] \leq \lim_{n\to\infty} \frac{1}{v_n}\mathbf{E}\left[\sigma\right] = 0$$

by the finiteness of $\mathbf{E}[\sigma]$.

As we shall see in the remainder of this chapter, both the representations (4.2.7) and (4.2.8) prove useful in determining the scaling function v.

4.3 Selection of the sequence v_n, $(n = 1, 2, \ldots)$

It is quite obvious that the function $\theta \to \Lambda_b(\theta)$ critically depends on the selection of the pair (a, v). The concluding portion of Section 4.1 having already narrowed down the selection of a to linear sequences, we now focus on the challenging task of identifying the scaling v.

4.3.1 Preliminary Results

A glance at (4.2.7) or (4.2.8) suggests that the limiting value of the ratio $\frac{v_n}{n}$ as n goes to infinity might have a substantial effect on the form of $\Lambda_b(\theta)$, θ in $\mathbb{R}$. This effect is demonstrated by Theorem 4.3.1, which indicates at the very outset the ineligibility of scalings v with $\lim_{n\to\infty} \frac{v_n}{n} = \infty$.

Theorem 4.3.1 *If the scaling sequence* $\{v_n,\ n = 1, 2, \ldots\}$ *is such that* $\lim_{n\to\infty} \frac{v_n}{n} = \infty$, *then we have*

$$\Lambda_b(\theta) = \begin{cases} 0, & \theta \leq 0 \\ \infty, & \theta > 0. \end{cases}$$

Proof. Fix θ in $\mathbb{R}$: By the definition (4.2.4), we have

$$\Delta(n,\theta) \;=\; \sum_{r=1}^{n}\left(1-\frac{r}{n}\right)e^{r\theta}\mathbf{P}\left[\widehat{\sigma}>r\right] \;\geq\; 0.$$

Recognizing that $1-e^{-\theta}>0$ for $\theta>0$, we use (4.2.7) to provide the bound

$$\begin{aligned}\Lambda_b(\theta) &\geq \lambda\mathbf{E}\left[\sigma\right]\lim_{n\to\infty}\left(\frac{e^{\theta\frac{v_n}{n}}-1}{\frac{v_n}{n}}\right)\\ &= \lambda\mathbf{E}\left[\sigma\right]\lim_{n\to\infty}\frac{e^{\theta\frac{v_n}{n}}-1}{\frac{v_n}{n}}\\ &= \infty.\end{aligned}$$

When $\theta<0$,

$$\begin{aligned}0 \;\leq\; \Delta(n,\theta) &= \sum_{r=1}^{n}\left(1-\frac{r}{n}\right)e^{r\theta}\mathbf{P}\left[\widehat{\sigma}>r\right]\\ &\leq \sum_{r=1}^{n}e^{r\theta}\\ &\leq \sum_{r=1}^{\infty}e^{r\theta}\\ &= \frac{e^{\theta}}{1-e^{\theta}},\end{aligned}$$

and

$$0 \;\geq\; \left(1-e^{-\theta}\right)\Delta(n,\theta) \;\geq\; \left(1-e^{-\theta}\right)\frac{e^{\theta}}{1-e^{\theta}},$$

implying

$$0 \;\leq\; 1+\left(1-e^{-\theta}\right)\Delta(n,\theta) \;\leq\; 1.$$

Hence referring to (4.2.7) with $\theta<0$, we have

$$\lambda\mathbf{E}\left[\sigma\right]\lim_{n\to\infty}\frac{n}{v_n}\left(e^{\theta\frac{v_n}{n}}-1\right) \;\leq\; \Lambda_b(\theta) \;\leq\; 0,$$

leading inevitably to the conclusion that $\Lambda_b(\theta) = 0$ when $\theta \leq 0$. ■

Due to the linear nature of the scaling function a, Theorem 4.3.1 clearly indicates that only linear and sub–linear forms may be considered in selecting v, *irrespective* of the form of distribution G.

Next, we provide the first clue that the dependence of v on the distribution G occurs through the function $v^\star$, in the form of Theorem 4.3.2. Acting in a capacity similar to Theorem 4.3.1, it further streamlines the set of acceptable sequences by dismissing those that asymptotically increase faster than $v^\star$.

Theorem 4.3.2 *Consider a scaling sequence* $\{v_n,\ n = 1, 2, \ldots\}$ *such that*

$$\lim_{n\to\infty} \frac{v_n^\star}{v_n} = K \qquad (4.3.1)$$

for some finite constant $K \geq 0$. *We then have*

$$\lim_{n\to\infty} \Lambda_{b,n}(\theta) = \infty, \qquad \theta > K. \qquad (4.3.2)$$

Proof. Fix $\theta > 0$. From (4.2.6) we have

$$\begin{aligned} \mathcal{D}_n(\sigma,\theta) &\geq \left(e^{\theta n} - 1\right) \mathbf{E}\left[\sigma\right] \mathbf{P}\left[\widehat{\sigma} > n\right] \\ &= \left(e^{\theta n} - 1\right) \mathbf{E}\left[\sigma\right] e^{-v_n^\star}, \quad n = 1, 2, \ldots, \end{aligned}$$

where the final step follows by the definition (2.3.1). Applying the previous inequality to (4.2.8), we get

$$\begin{aligned} \Lambda_b(\theta) &\geq \lambda \mathbf{E}\left[\sigma\right] \lim_{n\to\infty} \frac{e^{-v_n^\star}}{v_n} \left(e^{n\theta \frac{v_n}{n}} - 1\right) \\ &= \lambda \mathbf{E}\left[\sigma\right] \lim_{n\to\infty} \frac{e^{-v_n^\star + \theta v_n}}{v_n} \end{aligned}$$

$$
\begin{aligned}
&= \lambda \mathbf{E}[\sigma] \lim_{n\to\infty} \frac{e^{v_n\left(\theta - \frac{v_n^\star}{v_n}\right)}}{v_n} \\
&= \lambda \mathbf{E}[\sigma] \lim_{n\to\infty} \frac{e^{v_n(\theta-K)}}{v_n} \\
&= \infty, \qquad \theta > K.
\end{aligned}
$$

■

Summarizing the ground covered so far, we can state definitely, that the scaling v may at most be linear in case $v^\star$ is asymptotically linear or super–linear, but can only be sub–linear when $v^\star$ is asymptotically sub–linear. In other words, it seems reasonable to expect that the selection of v relies on the limiting value of $\frac{v_n^\star}{n}$. With this in mind, we make the reasonable assumption that the limit

$$
\lim_{n\to\infty} \frac{v_n^\star}{n} = R, \qquad 0 \le R \le \infty \tag{4.3.3}
$$

exists, and identify three distinct cases as shown in Table 4.1.

Category	$\lim_{n\to\infty} \frac{v_n^\star}{n} = R$	Tail of G	$v^\star$	v
I	$R = \infty$	super–exponential	super–linear	linear/sub–linear
II	$0 < R < \infty$	exponential	linear	linear/sub–linear
III	$R = 0$	sub–exponential	sub–linear	sub–linear

Table 4.1: Three cases defined by the tail of distribution G

Having laid the necessary foundation, we now proceed to study $\Lambda_b(\theta)$ in each of the three cases, under both linear and sub–linear scalings.

4.4 The linear scaling

We have already seen via Theorem 4.3.2 that a linear scaling v is not suitable when $v^\star$ is sub–linear, i.e., when G falls under category III; we now investigate its eligibility for cases I and II.

To assist our calculations, we begin with the simplifying assumption that $v_n = n,\ n = 0, 1, \ldots$. In this case, Theorem 4.4.1 gives the form of $\Lambda_b(\theta)$ for cases I, II and III.

Theorem 4.4.1 *Fix $\theta \neq R$, in $\mathbb{R}$. For the scaling sequence $v_n = n,\ n = 0, 1, \ldots$, we always have*

$$\Lambda_b(\theta) \;=\; \begin{cases} \lambda \mathbf{E}\left[e^{\theta\sigma} - 1\right] & \theta < R \\ \infty & \theta > R, \end{cases} \tag{4.4.1}$$

for $R \geq 0$.

Proof. Fix θ in $\mathbb{R}$. For the particular scaling sequence $v_n = n,\ n = 0, 1, \ldots$, the relations (4.2.6) and (4.2.8) yield

$$\begin{aligned}
\Lambda_b(\theta) &= \lim_{n\to\infty} \frac{\lambda}{n}\, \mathcal{D}_n\left(\sigma, \theta \frac{v_n}{n}\right) \\
&= \lim_{n\to\infty} \frac{\lambda}{n}\, \mathcal{D}_n\left(\sigma, \theta\right) \\
&= \lambda \left(L_1(\theta) + \left(\frac{e^\theta + 1}{e^\theta - 1}\right) L_2(\theta) + \mathbf{E}\left[\sigma\right] L_3(\theta) \right),
\end{aligned}$$

where

$$L_1(\theta) \;\equiv\; \lim_{n\to\infty} \frac{1}{n} \mathbf{E}\left[(n-\sigma)^+ \left(e^{\theta\sigma} - 1\right)\right],$$

$$L_2(\theta) \;\equiv\; \lim_{n\to\infty} \frac{1}{n} \mathbf{E}\left[e^{\theta \min(n,\sigma)} - 1\right],$$

and

$$L_3(\theta) \equiv \lim_{n\to\infty} \frac{1}{n}\mathbf{P}\left[\widehat{\sigma} > n\right]\left(e^{\theta n} - 1\right).$$

We have already evaluated $L_3(\theta)$ in the course of proving Theorem 4.3.2; to reiterate,

$$\begin{aligned} L_3(\theta) &= \lim_{n\to\infty} \frac{e^{-v_n^\star}}{n}\left(e^{\theta n} - 1\right) \\ &= \lim_{n\to\infty} \frac{e^{\theta n - v_n^\star}}{n} \\ &= \lim_{n\to\infty} \frac{e^{n(\theta - R + o(1))}}{n} \\ &= \begin{cases} \infty & \theta > R \\ 0 & \theta < R. \end{cases} \end{aligned} \tag{4.4.2}$$

The value taken by $L_3(\theta)$ when $\theta = R$, cannot be ascertained without additional information about the behaviour of the sequence $v^\star$; we postpone this discussion for later.

As $L_1(\theta), L_2(\theta)$, and $L_3(\theta)$ are non–negative for $\theta \geq 0$, and hence for $\theta > R$, (4.4.2) already implies

$$\Lambda_b(\theta) = \infty, \qquad \theta > R.$$

We therefore restrict our attention to $L_1(\theta)$ and $L_2(\theta)$ in the region $\theta < R$.

For $\theta > 0$, monotone convergence ensures that

$$\begin{aligned} L_1(\theta) &= \lim_{n\to\infty} \mathbf{E}\left[\left(1 - \frac{\sigma}{n}\right)^+ \left(e^{\theta\sigma} - 1\right)\right] \\ &= \mathbf{E}\left[e^{\theta\sigma} - 1\right]. \end{aligned}$$

Since $e^{\theta\sigma} - 1 = -(1 - e^{\theta\sigma})$, the same argument can be applied when $\theta \leq 0$, yielding the same result.

For the case $\theta > 0$, observe that the limit

$$L_2(\theta) \;\leq\; \lim_{n\to\infty} \frac{1}{n}\mathbf{E}\left[e^{\theta\sigma} - 1\right] \;=\; 0,$$

as long as $\mathbf{E}\left[e^{\theta\sigma}\right] < \infty$. When $\mathbf{E}\left[e^{\theta\sigma}\right] = \infty$, the limit $L_1(\theta)$ is infinite in any case, rendering the value taken by $L_2(\theta)$ irrelevant. When $\theta < 0$, $L_2(\theta) = 0$, since $0 < e^{\theta\min(\sigma,n)} < 1$.

Finally, recombining the limits $L_1(\theta)$, $L_2(\theta)$, and $L_3(\theta)$, we get the desired result. ■

This last result is compatible with Theorem 4.3.2 for the cases II and III, since $K = R$ with $v_n = n,\ n = 0, 1, \ldots$.

Theorem 4.4.1 suggests the linear scaling as suitable in those situations when $R > 0$, *and* $\mathbf{E}\left[e^{\theta\sigma}\right] < \infty,\ 0 < \theta < R$. Naturally, it would be advantageous to study these situations in greater detail; we address the issue promptly in the next sub–section.

4.4.1 Finiteness of exponential moments

We shall find it useful to relate the exponential moments of σ to those of $\widehat{\sigma}$, and to characterize their finiteness in terms of the properties of the sequence $\{v^\star_n,\ n = 1, 2, \ldots\}$. To do so, we define

$$\Sigma(\theta) \equiv \sum_{r=1}^{\infty} \exp\left(r(\theta - \frac{v^\star_r}{r})\right), \quad \theta \in \mathbb{R}\,. \tag{4.4.3}$$

Cauchy's convergence criterion readily yields the following fact used in several places.

Lemma 4.4.1 *Assume (4.3.3) with $R > 0$, possibly infinite. Then, the quantity $\Sigma(\theta)$ is finite (resp. infinite) if $\theta < R$ (resp. $\theta > R$).*

The main result of this section is contained in the following

Proposition 4.4.1 *For each θ in $\mathbb{R}$, the quantities $\mathbf{E}\left[e^{\theta\widehat{\sigma}}\right]$, $\mathbf{E}\left[e^{\theta\sigma}\right]$ and $\Sigma(\theta)$ are all finite (resp. infinite) simultaneously.*

The proof of this result passes by the next two technical lemmas.

Lemma 4.4.2 *For each θ in $\mathbb{R}$, the relation*

$$\mathbf{E}\left[e^{\theta\widehat{\sigma}}\right] = \frac{1}{\mathbf{E}\left[\sigma\right]}\frac{1}{1-e^{-\theta}}\mathbf{E}\left[e^{\theta\sigma}-1\right] \tag{4.4.4}$$

holds.

We obviously have $\mathbf{E}\left[e^{\theta\widehat{\sigma}}\right] = 1$ for $\theta = 0$, a fact which is easily seen to be consistent with (4.4.4) by applying L'Hospital's rule on its right hand–side.

Proof. Fix $\theta \neq 0$ in $\mathbb{R}$. We see that

$$\begin{aligned}
\mathbf{E}\left[e^{\theta\widehat{\sigma}}\right] &= \sum_{r=1}^{\infty} \widehat{g}_r e^{r\theta} \\
&= \frac{1}{\mathbf{E}\left[\sigma\right]}\sum_{r=1}^{\infty} \mathbf{P}\left[\sigma \geq r\right] e^{r\theta} \\
&= \frac{1}{\mathbf{E}\left[\sigma\right]}\sum_{r=1}^{\infty} e^{r\theta}\sum_{s=r}^{\infty} g_s \\
&= \frac{1}{\mathbf{E}\left[\sigma\right]}\sum_{s=1}^{\infty} g_s\sum_{r=1}^{s} e^{r\theta} \\
&= \frac{1}{\mathbf{E}\left[\sigma\right]}\frac{e^{\theta}}{e^{\theta}-1}\sum_{s=1}^{\infty} g_s(e^{s\theta}-1)
\end{aligned}$$

and the conclusion (4.4.4) follows. ■

Lemma 4.4.3 *For each θ in $\mathbb{R}$, we have the relation*

$$\Sigma(\theta) = \left(\frac{1}{1-e^{-\theta}}\right)\left(\frac{\mathbf{E}\left[e^{\theta\sigma}-1\right]}{\mathbf{E}\left[\sigma\right](e^{\theta}-1)} - 1\right). \tag{4.4.5}$$

When $\theta = 0$, the arguments leading to (4.4.5) can be modified to yield

$$\Sigma(\theta) = \sum_{h=1}^{\infty} \mathbf{P}\left[\widehat{\sigma} > h\right] = \frac{1}{2}\mathbf{E}\left[\sigma(\sigma-1)\right]. \tag{4.4.6}$$

This last expression is easily seen to coincide with (4.4.5) via L'Hospital's rule.

Proof. Fix $\theta \neq 0$ in $\mathbb{R}$, and note from (C.2.1) that

$$\begin{aligned} \Sigma(\theta) &= \sum_{r=1}^{\infty} e^{r\theta}\mathbf{P}\left[\widehat{\sigma} > r\right] \\ &= \frac{1}{\mathbf{E}\left[\sigma\right]}\sum_{r=1}^{\infty} e^{r\theta}\mathbf{E}\left[(\sigma - r)^{+}\right]. \end{aligned} \tag{4.4.7}$$

Next,

$$\begin{aligned} \sum_{r=1}^{\infty} e^{r\theta}\mathbf{E}\left[(\sigma-r)^{+}\right] &= \sum_{r=1}^{\infty} e^{r\theta}\sum_{j=r}^{\infty}(j-r)g_j \\ &= \sum_{j=1}^{\infty} g_j \sum_{r=1}^{j}(j-r)e^{r\theta} \\ &= \sum_{j=1}^{\infty} g_j e^{j\theta}\sum_{h=0}^{j-1} he^{-h\theta}, \end{aligned} \tag{4.4.8}$$

by the substitution $h = j - r$.

Using the standard identity

$$\sum_{r=a}^{\infty} re^{-\theta r} = \frac{e^{-a\theta}}{1-e^{-\theta}}\left(a + \frac{1}{e^{\theta}-1}\right), \qquad a = 0, 1, \ldots$$

we evaluate

$$\begin{aligned} \sum_{h=0}^{j-1} he^{-h\theta} &= \sum_{h=0}^{\infty} he^{-h\theta} - \sum_{h=j}^{\infty} he^{-h\theta} \\ &= \frac{1}{1-e^{-\theta}}\left(\frac{1-e^{-j\theta}}{e^{\theta}-1} - je^{-j\theta}\right), \qquad j = 1, 2, \ldots. \end{aligned} \tag{4.4.9}$$

Reporting (4.4.9) into (4.4.8) gives

$$\begin{aligned}\sum_{r=1}^{\infty} e^{r\theta}\mathbf{E}\left[(\sigma-r)^{+}\right] &= \frac{1}{1-e^{-\theta}}\sum_{j=1}^{\infty} g_j\left(\frac{e^{j\theta}-1}{e^{\theta}-1}-j\right)\\ &= \frac{1}{1-e^{-\theta}}\left(\frac{\mathbf{E}\left[e^{\sigma\theta}-1\right]}{e^{\theta}-1}-\mathbf{E}\left[\sigma\right]\right),\end{aligned}$$

and the result emerges through a simple comparison with (4.4.7). ■

When $R>0$, Proposition 4.4.1 and Lemma 4.4.1 readily imply $\mathbf{E}\left[e^{\theta\sigma}\right]$ finite (resp. infinite) if $\theta<R$ (resp. $\theta>R$), allowing Theorem 4.4.1 to be rephrased as follows:

Theorem 4.4.2 *Fix $\theta \neq R in\, \mathbb{R}$. For the scaling sequence $v_n = n,\ n=0,1,\ldots,$ we have*

$$\Lambda_b(\theta) = \lambda \mathbf{E}\left[e^{\theta\sigma}-1\right].$$

with $\mathbf{E}\left[e^{\theta\sigma}\right]$ finite (resp. infinite) if $\theta<R$ (resp. $R<\theta$).

The finiteness of the limiting value at the boundary $\theta = R$ depends on the value of R: If $R=\infty$, then the issue is moot as $\Lambda_b(\theta)$ is finite for all θ in $\mathbb{R}$, while if $R=0$, then the boundary point $\theta=0$ yields a zero limit. When $0<R<\infty$, the result at the boundary point $\theta=R$ is highly dependent on the finer structure of the sequence $\{v_n^\star,\ n=1,2,\ldots\}$. Lemma 4.4.4 presents results along these lines, complementing Theorem 4.4.2 with a simple characterization of the finiteness of $\mathbf{E}\left[e^{R\sigma}\right]$ in Case II.

Lemma 4.4.4 *Assume $0<R<\infty$. We have $\mathbf{E}\left[e^{R\sigma}\right]$ infinite if either (i) $v_n^\star \leq Rn$ infinitely often or (ii) $v_n^\star > Rn$ for $n=N,N+1,\ldots$ for some finite N and $\limsup_{n\to\infty}(v_n^\star - Rn) = L$ for some finite $L\geq 0$.*

Proof. By Proposition 4.4.1, the finiteness of $\mathbf{E}\left[e^{R\sigma}\right]$ is equivalent to that of $\Sigma(R)$. Under (i), the set $\mathcal{N} \equiv \{n = 1, 2, \ldots : \ v_n^\star \leq Rn\}$ is countably infinite, so that

$$\Sigma(R) \geq \sum_{n \in \mathcal{N}} \exp\left(Rn - v_n^\star\right) \geq \sum_{n \in \mathcal{N}} 1 = \infty.$$

Under (ii), the condition $\limsup_{n\to\infty}(v_n^\star - Rn) = L$ for some finite $L \geq 0$ implies for any $\varepsilon > 0$, the existence of an integer $n^\star = n^\star(\varepsilon)$ such that $0 \leq v_n^\star - Rn \leq L + \varepsilon$ for all $n \geq n^\star$, whence

$$\Sigma(R) \geq \sum_{n=n^\star}^{\infty} e^{-(L+\varepsilon)} = \infty.$$

■

Conditions (i) and (ii) are non overlapping, and do cover most distributions of interest in case II. However, Lemma 4.4.4 does not cover the situation in (ii) with $\limsup_{n\to\infty}(v_n^\star - Rn) = \infty$. Indeed, with $R = 1$, for $v_n^\star = n + \sqrt{n}$ we find $\Lambda_b(1) = \infty$, while for $v_n^\star = n + \frac{n}{\ln n}$, we have $\Lambda_b(1) < \infty$.

Theorem 4.4.2 indicates that the linear scaling

$$v_n = n, \quad n = 1, 2, \ldots \tag{4.4.10}$$

is a suitable candidate in Cases I and II. It also concurs with Theorem 4.3.2 in deeming the linear scaling inappropriate for Case III, thus paving the way towards a discussion on the possibilities offered by the class of sub–linear scaling functions.

4.5 Sub–linear scaling sequences

4.5.1 General Results

We begin with a Lemma that holds in general for any choice of sub–linear scaling, irrespective of the value taken by R, and for all θ in $\mathbb{R}$. The result, though straightforward is quite significant, as it indicates that $\Lambda_b(\theta)$ is *only* determined by the limit $\lim_{n\to\infty} \frac{v_n}{n}\Delta\left(n, \theta\frac{v_n}{n}\right)$, when the selected scaling v is sub–linear.

Lemma 4.5.1 *If the scaling sequence* $\{v_n,\ n = 1, 2, \ldots\}$ *is selected so that*

$$\lim_{n\to\infty} \frac{v_n}{n} = 0,$$

then for all θ *in* $\mathbb{R}$ *we have*

$$\Lambda_b(\theta) = \lambda \mathbf{E}[\sigma]\,\theta\,(1 + M^v(\theta)), \tag{4.5.1}$$

where

$$M^v(\theta) \equiv \lim_{n\to\infty} \frac{v_n}{n}\Delta\left(n, \theta\frac{v_n}{n}\right). \tag{4.5.2}$$

Proof. Fix θ in $\mathbb{R}$ and set $x_n = \theta\frac{v_n}{n}$ in (4.2.7). Noting that $\lim_{n\to\infty} x_n = 0$, we have

$$\begin{aligned}
\Lambda_b(\theta) &= \lambda \mathbf{E}[\sigma] \lim_{n\to\infty} (e^{x_n} - 1)\frac{n}{v_n}\left(1 + \left(1 - e^{-x_n}\right)\Delta(n, x_n)\right) \\
&= \lambda \mathbf{E}[\sigma]\,\theta \lim_{n\to\infty} \left(\frac{e^{x_n} - 1}{x_n}\right)\left(1 + x_n e^{-x_n}\left(\frac{e^{x_n} - 1}{x_n}\right)\Delta(n, x_n)\right) \\
&= \lambda \mathbf{E}[\sigma]\,\theta\left(1 + \lim_{n\to\infty} x_n \Delta(n, x_n)\right),
\end{aligned}$$

achieved by invoking the limit

$$\lim_{x\to 0} \frac{e^x - 1}{x} = 1,$$

and the result follows. ■

Lemma 4.5.1 clearly declares that evaluating $\Lambda_b(\theta)$ under a sub–linear scaling really amounts to evaluating the limit $M^v(\theta)$. With this in mind, we evaluate this limit under various conditions in the next few Lemmas.

Lemma 4.5.2 *Assume $R > 0$. Under any sub–linear scaling, the limit (4.5.2) satisfies $M^v(\theta) = 0, \theta$ in $\mathbb{R}$.*

Proof. When $\theta > 0$, the definitions (4.2.4) and (4.5.2) yield

$$\begin{aligned} 0 \leq M^v(\theta) &\leq \lim_{n\to\infty} \frac{v_n}{n} \sum_{r=1}^{n} e^{\theta \frac{v_n}{n} r - v_r^\star} \\ &\leq \lim_{n\to\infty} \frac{v_n}{n} \sum_{r=1}^{\infty} e^{\theta \frac{v_n}{n} r - v_r^\star} \\ &= \lim_{n\to\infty} \frac{v_n}{n} \Sigma\left(\theta \frac{v_n}{n}\right), \end{aligned} \tag{4.5.3}$$

by definition (4.4.3). By Lemma 4.4.1, $\Sigma\left(\theta \frac{v_n}{n}\right)$ is finite for any finite θ (since $\theta \frac{v_n}{n} < R$ for n large enough), thus allowing us to conclude by monotonicity that $M^v(\theta) = 0$ when $\theta > 0$. A similar argument holds when $\theta \leq 0$ with a reversal of the inequalities, thereby concluding the proof. ■

Here, we pause to comment briefly that though Theorem 4.4.2 had declared the linear scaling as acceptable in Cases I and II, it did not rule out the existence of *other* suitable scaling functions. In fact, as has been pointed out by Lemma 4.5.2, when $R > 0$, *any* sub–linear scaling yields a non–trivial value for $\Lambda_b(\theta)$ for every $\theta \neq 0$ in $\mathbb{R}$. The corresponding result for the case $R = 0$ is given next:

Lemma 4.5.3 *Assume $R = 0$. Under any sub–linear scaling, the limit (4.5.2) satisfies $M^v(\theta) = 0, \theta \leq 0$.*

Proof. For the case $\theta < 0$, we have

$$0 \;\leq\; M^v(\theta) \;\leq\; \lim_{n\to\infty} \frac{v_n}{n} \Sigma\left(\theta \frac{v_n}{n}\right)$$

from (4.2.4) and (4.5.2). By Lemma 4.4.1 we know that $\Sigma\left(\theta \frac{v_n}{n}\right)$ is finite for $\theta < 0$, therefore implying by monotonicity that $\lim_{n\to\infty} \frac{v_n}{n} \Sigma\left(\theta \frac{v_n}{n}\right) = 0$ and providing the result. ■

4.5.2 Scaling $v_n = v_n^\star$, $(n = 1, 2, \ldots)$

The first intimation that $v^\star$ might have a pivoting role in selecting the scaling v was given by Theorem 4.3.2, which indicated the inadequacy of any scaling that asymptotically increased infinitely faster than $v^\star$. This result is all the more significant in the case when $R = 0$, leading one to suspect the natural choice $v_n = v_n^\star$ to work out.

In this situation, Theorem 4.3.2 and Lemma 4.5.3 allow us to identify the limit $\Lambda_b(\theta)$ for all θ outside the interval $(0, 1]$. We now present two Lemmas that evaluate $\Lambda_b(\theta)$, via $M^v(\theta)$ defined in (4.5.2), in the missing interval $(0, 1]$, under the scaling $v_n = v_n^\star$. The computation of $\Lambda_b(\theta)$ within this region is somewhat more involved, as it seems to depend on the finer distributional properties of the rv σ. This being the case, we introduce some additional assumptions to aid our calculations.

We say that a sequence $\{q_n,\ n = 1, 2, \ldots\}$ is monotone decreasing (resp. increasing) in the limit if there exists a finite integer N such that the tail $\{q_n,\ n = N+1, N+2, \ldots\}$ is monotone decreasing (resp. increasing).

Assumption C1: *The sequence* $\{v^\star_n/n,\ n = 1, 2, \ldots\}$ *is monotone decreasing in the limit.*

Assumption C2a: *We assume*

$$\sum_{r=1}^{\infty} e^{-(1-\theta)v^\star_r} < \infty, \quad 0 < \theta < 1. \tag{4.5.4}$$

Lemma 4.5.4 *Assume* $R = 0$*. If the sequence* $\{v^\star_n,\ n = 1, 2, \ldots\}$ *satisfies Assumptions* **C1** *and* **C2a***, then*

$$M^{v^\star}(\theta) = \lim_{n\to\infty} \frac{v^\star_n}{n} \Delta\left(n, \theta \frac{v^\star_n}{n}\right) = 0 \tag{4.5.5}$$

for each $0 \leq \theta \leq 1$.

Noting that

$$\sum_{r=1}^{\infty} e^{-v^\star_r} \leq \sum_{r=1}^{\infty} e^{-(1-\theta)v^\star_r}, \quad 0 < \theta < 1.$$

we see from (4.4.6) that a necessary condition for (4.5.4) to hold is that σ have a finite moment of order two, i.e., $\mathbf{E}\left[\sigma^2\right] < \infty$. As this finite moment assumption is not satisfied in the important case when the $M|G|\infty$ process is long–range dependent, we now present another criterion that ensures the conclusion of Lemma 4.5.4.

Assumption C2b: *There exists a mapping* $Z : \mathbb{N} \to \mathbb{N}$ *for the sequence* $\{v^\star_n/n,\ n = 1, 2, \ldots\}$ *such that*

(i) $Z(n) < n$ *for large* $n = 1, 2, \ldots$*;*

(ii) $\lim_{n\to\infty} v_n^\star \dfrac{Z(n)}{n} = \infty;$

(iii) $\lim_{n\to\infty} \dfrac{v_n^\star}{n} \dfrac{Z(n)}{v_{Z(n)}^\star} = 0.$

Lemma 4.5.5 *Assume $R = 0$. If the sequence $\{v_n^\star,\ n = 1, 2, \ldots\}$ satisfies Assumptions* **C1** *and* **C2b***, then the result (4.5.5) still holds for $0 \le \theta < 1$.*

The proofs to Lemmas 4.5.4 and 4.5.5 are provided in Appendix C.3.

The assumptions of Lemma 4.5.5 are satisfied in all cases known to the author, and are easy to check for broad classes of distributions: If $v_n^\star \sim n^\beta$ $(0 < \beta < 1)$, we can take $Z(n) = n^\gamma$ with $1 - \beta < \gamma < 1$. If $v_n^\star \sim (\ln n)^\beta$ $(\beta > 0)$, then the choice $Z(n) = n(\ln n)^{-\gamma}$ with $0 < \gamma < \beta$ is convenient.

Lemmas 4.5.1, 4.5.3, 4.5.4, and 4.5.5, when combined with Theorem 4.3.2, identify $\Lambda_b(\theta)$ under scaling $v = v^\star$, for all $\theta \neq 1$ in $\mathbb{R}$. The resulting expression is presented in the following Theorem.

Theorem 4.5.1 *Assume $R = 0$. If a sequence $\{v_n^\star,\ n = 1, 2, \ldots\}$ satisfies condition* **C1,** *and* **any one** *of conditions* **C2a** *and* **C2b***, then the limit $\Lambda_b(\theta)$ exists for all $\theta \neq 1$ in $\mathbb{R}$, under scaling $v = v^\star$ and is given by*

$$\Lambda_b(\theta) = \begin{cases} \lambda \mathbf{E}[\sigma]\,\theta & \text{if } \theta < 1 \\ \infty & \text{if } \theta > 1. \end{cases}$$

4.5.3 Scaling $v_n = o(v_n^\star)$, $(n = 1, 2, \ldots)$

We have considered $v^\star$, and sub–linear functions asymptotically increasing faster than $v^\star$, as potential scaling functions. For the sake of completeness, we now address sub–linear forms that increase slower than $v^\star$ in the limit, i.e., those sequences

$\{v_n,\ n = 1, 2, \ldots\}$ for which $\lim_{n\to\infty} \frac{v_n}{n} = 0$, and $\lim_{n\to\infty} \frac{v_n^\star}{v_n} = \infty$. Lemma 4.5.6 indicates that under certain conditions, such scaling functions yield the limit $M^v(\theta) = 0$ for every $\theta > 0$.

Lemma 4.5.6 *Consider a scaling sequence* $\{v_n,\ n = 1, 2, \ldots\}$ *such that*

$$\lim_{n\to\infty} \frac{v_n^\star}{v_n} = \infty.$$

Assume $R = 0$. *If the sequence* $\{v_n^\star/n,\ n = 1, 2, \ldots\}$ *satisfies Assumptions* **C1** *and any one of* **C2a** *and* **C2b**, *then* $M^v(\theta) = 0,\ \theta > 0$.

Proof. As $\lim_{n\to\infty} \frac{v_n}{v_n^\star} = 0$, for ever $\delta > 0$ we can find an integer $N(\delta)$, such that

$$\frac{v_n}{v_n^\star} < \delta, \qquad n > N(\delta).$$

Fix $\theta > 0$. By (4.5.2), we get

$$\begin{aligned} 0 \ \leq \ M^v(\theta) \ &\leq \ \lim_{n\to\infty} \frac{v_n}{n} \sum_{r=1}^{n} e^{\left(\theta \frac{v_n}{n} - \frac{v_r^\star}{r}\right) r} \\ &= \ \lim_{n\to\infty} \frac{v_n}{v_n^\star} \frac{v_n^\star}{n} \sum_{r=1}^{n} e^{\left(\theta \frac{v_n}{v_n^\star} \frac{v_n^\star}{n} - \frac{v_r^\star}{r}\right) r} \\ &\leq \ \delta \lim_{n\to\infty} \frac{v_n^\star}{n} \sum_{r=1}^{n} e^{\left(\theta \delta \frac{v_n^\star}{n} - \frac{v_r^\star}{r}\right) r}. \end{aligned}$$

Select $\delta < \frac{1}{\theta}$. The required result then follows under Assumptions **C1** and any one of **C2a** and **C2b** through arguments provided in Appendix C.3. ■

4.6 Equivalent scaling sequences

The scaling sequence $\{v_n,\ n = 1.2, \ldots\}$ that guarantees a non–trivial limit (4.1.4) is obviously *not* unique. Indeed, consider two scaling sequences $\{v_n,\ n = 1, 2, \ldots\}$ and $\{w_n,\ n = 1, 2, \ldots\}$ such that

$$\lim_{n\to\infty} \frac{w_n}{v_n} = K \tag{4.6.1}$$

for some finite positive constant K. Using the superscript v or w in order to explicitly indicate the dependence on the scaling sequence, we denote the quantity defined in (4.1.4), by $\Lambda_b^v(\theta)$ and $\Lambda_b^w(\theta)$, respectively, for each θ in $\mathbb{R}$. We now examine their relationship with the help of Lemma 4.6.1.

Lemma 4.6.1 *Consider two scaling sequences $\{v_n,\ n = 1, 2, \ldots\}$ and $\{w_n,\ n = 1, 2, \ldots\}$ satisfying (4.6.1) for some positive constant K. Then,*

$$\Lambda_b^w(\theta) = \frac{1}{K}\Lambda_b^v(K\theta), \qquad \theta \in \mathbb{R} \tag{4.6.2}$$

except possibly at an isolated point $\theta^\star$ where $\Lambda_b^v((K\theta^\star)-) < \Lambda_b^v((K\theta^\star)+) = \infty$.

Proof. Equation (4.6.1) implies that for any choice of $\delta > 0$, there exists an integer $N(\delta)$ such that

$$K - \delta \leq \frac{w_n}{v_n} \leq K + \delta$$

for every $n > N(\delta)$.

For any $\theta > 0$, the inequalities

$$\mathbf{E}\left[\exp\left(\theta(K-\delta)\frac{v_n}{n}S_n^b\right)\right] \ \leq\ \mathbf{E}\left[\exp\left(\frac{\theta w_n}{n}S_n^b\right)\right] \ \leq\ \mathbf{E}\left[\exp\left(\theta(K+\delta)\frac{v_n}{n}S_n^b\right)\right],$$

then hold for all $n > N(\delta)$. Taking the natural logarithm and dividing by v_n and letting n go to infinity gives

$$\Lambda_b^v\left(\theta(K-\delta)\right) \leq \lim_{n\to\infty} \frac{w_n}{v_n} \frac{1}{w_n} \ln \mathbf{E}\left[\exp\left(\frac{\theta w_n}{n} S_n^b\right)\right] \leq \Lambda_b^v\left(\theta(K+\delta)\right),$$

leading to the conclusion that

$$\Lambda_b^v\left(\theta(K-\delta)\right) \leq K\Lambda_b^w(\theta) \leq \Lambda_b^v\left(\theta(K+\delta)\right), \qquad \theta > 0. \tag{4.6.3}$$

The case $\theta \leq 0$ can be dealt with in an identical fashion, and also yields the inequality (4.6.3), but with the inequalities reversed. The result follows directly by letting δ go to zero in (4.6.3). ■

A similar argument can be made with regard to the scaling sequence $\{a_n,\ n = 1, 2, \ldots\}$. The following result, proved as a Corollary to Lemma 4.6.1, confirms the claim made in Section 4.1, that the form of the function Λ_b derived under scaling $a_n = n,\ n = 1, 2, \ldots$ does not change under a different selection of a, (except for a multiplicative factor) as long as (4.1.3) holds.

We introduce the notation $\Lambda_b^{(a,v)}$ to identify the pair of scalings (a, v) under which the function Λ_b is derived. For the particular selection $a_n = n,\ n = 1, 2, \ldots$, we drop the double superscript and continue with our previous terminology, allowing Λ_b^v to denote the use of the implicit scaling $a_n = n,\ n = 1, 2, \ldots$ and of course $\{v_n,\ n = 1, 2, \ldots\}$.

Corollary 4.6.1 *Consider a scaling sequence $\{a_n,\ n = 1, 2, \ldots\}$ satisfying (4.1.3) for some finite, non–zero α. Then under any scaling v, we have*

$$\Lambda_b^{(a,v)}(\theta) \quad = \quad \Lambda_b^v(\theta\alpha), \qquad \theta \in \mathbb{R}, \tag{4.6.4}$$

except possibly at an isolated $\theta^\star \in \mathbb{R}$ where $\Lambda_b^v((\alpha\theta^\star)-) < \Lambda_b^v((\alpha\theta^\star)+) = \infty$.

Proof. Fix θ in $\mathbb{R}$. By (4.2.7), it holds that

$$\begin{aligned}\Lambda_b^{(a,v)}(\theta) &= \lambda \mathbf{E}[\sigma] \lim_{n\to\infty} \left(e^{\theta \frac{v_n}{a_n}} - 1\right) \frac{n}{v_n} \left(1 + \left(1 - e^{-\theta \frac{v_n}{a_n}}\right) \Delta\left(n, \theta \frac{v_n}{a_n}\right)\right) \\ &= \lambda \mathbf{E}[\sigma] \lim_{n\to\infty} \left(e^{\theta \frac{w_n}{n}} - 1\right) \frac{n}{w_n} \frac{n}{a_n} \left(1 + \left(1 - e^{-\theta \frac{w_n}{n}}\right) \Delta\left(n, \theta \frac{w_n}{n}\right)\right),\end{aligned}$$

where we have set $w_n = v_n \dfrac{n}{a_n}$. Hence,

$$\Lambda_b^{(a,v)}(\theta) = \alpha \Lambda_b^{w}(\theta),$$

and we have (4.6.4) upon applying Lemma 4.6.1 with $\lim_{n\to\infty} \dfrac{w_n}{v_n} = K = \alpha$. ■

Lemma 4.6.1 and Corollary 4.6.1 when combined, provide quick conversion equations relating the function Λ_b under different pairs of scalings (a, v). Corresponding relations can also be derived for Λ and $\Lambda^\star$, and are presented below in Theorem 4.6.1.

Theorem 4.6.1 *Consider scalings (a, v) and $(\widehat{a}, \widehat{v})$ such that*

$$\lim_{n\to\infty} \frac{\widehat{v}_n}{v_n} = \widehat{K}, \qquad \lim_{n\to\infty} \frac{n}{a_n} = \alpha, \quad \text{and} \quad \lim_{n\to\infty} \frac{n}{\widehat{a}_n} = \widehat{\alpha}, \tag{4.6.5}$$

with $0 < \widehat{K}, \alpha, \widehat{\alpha} < \infty$. The following conversion relations then apply.

(i) *For all θ in $\mathbb{R}$, except possibly for an isolated point $\theta^\star$ where*

$$\Lambda_b^{(a,v)}\left(\left(\widehat{K}\, \frac{\widehat{\alpha}}{\alpha}\, \theta^\star\right) -\right) < \Lambda_b^{(a,v)}\left(\left(\widehat{K}\, \frac{\widehat{\alpha}}{\alpha}\, \theta^\star\right) +\right) = \infty, \tag{4.6.6}$$

we have

(a)

$$\Lambda_b^{(\widehat{a},\widehat{v})}(\theta) = \frac{1}{\widehat{K}} \Lambda_b^{(a,v)}\left(\widehat{K}\, \frac{\widehat{\alpha}}{\alpha}\, \theta\right); \tag{4.6.7}$$

(b)

$$\Lambda^{(\widehat{a},\widehat{v})}(\theta) \;=\; \frac{1}{\widehat{K}}\Lambda^{(a,v)}\left(\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta\right). \qquad (4.6.8)$$

(ii) *If the limit* $\Lambda^{(\widehat{a},\widehat{v})}(\theta^\star)$ *exists,*

$$\Lambda^{\star(\widehat{a},\widehat{v})}(z) \;=\; \frac{1}{\widehat{K}}\Lambda^{\star(a,v)}\left(z\,\frac{\alpha}{\widehat{\alpha}}\right), \qquad z > 0. \qquad (4.6.9)$$

Proof. Fix θ in $\mathbb{R}$. By applying methods similar to those used in the proof of Lemma 4.6.1, it is possible to show that

$$\frac{1}{\widehat{K}}\Lambda^{(a,v)}\left(\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta-\varepsilon\right) \;\le\; \Lambda_b^{(\widehat{a},\widehat{v})}(\theta) \;\le\; \frac{1}{\widehat{K}}\Lambda_b^{(a,v)}\left(\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta+\varepsilon\right) \qquad (4.6.10)$$

for any $\varepsilon > 0$, and (4.6.7) follows directly for all $\theta \neq \theta^\star$ in $\mathbb{R}$.

Applying this result to (4.1.5) gives

$$\begin{aligned}\Lambda^{(\widehat{a},\widehat{v})}(\theta) &= \frac{1}{\widehat{K}}\Lambda_b^{(a,v)}\left(\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta\right) - c\widehat{\alpha}\theta \\ &= \frac{1}{\widehat{K}}\left(\Lambda^{(a,v)}\left(\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta\right) + c\alpha\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta\right) - c\widehat{\alpha}\theta, \qquad \theta \neq \theta^\star,\ \theta \in \mathbb{R},\end{aligned}$$

easily yielding (4.6.8).

Finally, to prove (4.6.9), we turn to (3.2.16), by which

$$\Lambda^{\star(\widehat{a},\widehat{v})}(z) \;\equiv\; \sup_{\theta>0}\left(\theta z - \Lambda^{(\widehat{a},\widehat{v})}(\theta)\right)$$

for all $z > 0$. By the non–decreasing nature of Λ_b, we infer from (4.6.6) and (4.6.10), that $\Lambda^{(\widehat{a},\widehat{v})}(\theta) = \infty$ for all $\theta > \theta^\star$, thereby implying that

$$\Lambda^{\star(\widehat{a},\widehat{v})}(z) \;=\; \sup_{0<\theta\le\theta^\star}\left(\theta z - \Lambda^{(\widehat{a},\widehat{v})}(\theta)\right), \qquad z > 0.$$

As long as the limit $\Lambda^{(\widehat{a},\widehat{v})}(\theta^\star)$ exists, its *specific* value is of no consequence in evaluating $\Lambda^{\star(\widehat{a},\widehat{v})}$. The existence of $\Lambda^{(\widehat{a},\widehat{v})}(\theta^\star)$, together with the fact

$$\lim_{\theta\uparrow\theta^\star} \Lambda^{(\widehat{a},\widehat{v})}(\theta) \leq \Lambda^{(\widehat{a},\widehat{v})}(\theta^\star),$$

is enough to conclude that

$$\begin{aligned}
\Lambda^{\star(\widehat{a},\widehat{v})}(z) &= \sup_{0<\theta<\theta^\star}\left(\theta z - \Lambda^{(\widehat{a},\widehat{v})}(\theta)\right) \\
&= \sup_{0<\theta<\theta^\star}\left(\theta z - \frac{1}{\widehat{K}}\Lambda^{(a,v)}\left(\widehat{K}\,\frac{\widehat{\alpha}}{\alpha}\,\theta\right)\right) \\
&= \frac{1}{\widehat{K}}\sup_{0<\theta'<\widehat{K}\frac{\widehat{\alpha}}{\alpha}\theta^\star}\left(\theta' z\,\frac{\alpha}{\widehat{\alpha}} - \Lambda^{(a,v)}(\theta')\right) \\
&= \frac{1}{\widehat{K}}\sup_{\theta>0}\left(\theta z\,\frac{\alpha}{\widehat{\alpha}} - \Lambda^{(a,v)}(\theta)\right) \\
&= \frac{1}{\widehat{K}}\Lambda^{\star(a,v)}\left(z\,\frac{\alpha}{\widehat{\alpha}}\right), \qquad z>0,
\end{aligned}$$

where the last but one step follows because $\Lambda^{(a,v)}(\theta) = \infty$ for all $\theta > \widehat{K}\,\dfrac{\widehat{\alpha}}{\alpha}\,\theta^\star$. ■

As the functions $\Lambda^{(a,v)}$ and $\Lambda^{(\widehat{a},\widehat{v})}$ have the same form except for multiplicative factors K and α, it follows that if the conditions of the Gärtner-Ellis Theorem hold for one, they must necessarily hold for the other. The same is true for requirement **A1** outlined in Section 3.4.1.

A primary advantage of Theorem 4.6.1 is its easy application in Case III, when $R = 0$. Often, the function $v^\star$ takes on a rather cumbersome form, making its use as a scaling function inconvenient. Theorem 4.6.1 allows a switch to any alternate scaling v that obeys $\lim_{n\to\infty} \dfrac{v_n}{v_n^\star} = 1$, with no modification whatsoever, to the form of Λ. Even so, in order to use such a scaling, we still have to ensure that the conditions laid out in Theorem 4.5.1 are satisfied. As this calculation involves the function $v^\star$, it too could prove a challenging task. The following Lemma suggests a way around this problem; its proof is trivial and has not been included.

Lemma 4.6.2 *Consider a sequence* $\{v_n,\ n = 1, 2, \ldots\}$ *such that* $\lim_{n\to\infty} \frac{v_n^\star}{v_n} = K$, *with* $0 < K < \infty$. *Any of the three conditions* **C1**, **C2a**, *or* **C2b** *is satisfied by the sequence* $v^\star$ **iff** *it is simultaneously satisfied by the scaling* v *in place of* $v^\star$.

The upshot of this discussion is that for Case III, we can blithely adopt any convenient scaling v that obeys $\lim_{n\to\infty} \frac{v_n}{v_n^\star} = 1$, and satisfies a modified version of the requirements stated in Theorem 4.5.1, the modification in the conditions being a simple replacement of the function $v^\star$ by v.

4.7 Review and discussion

We have seen a fair number of theorems and lemmas in the last few sections, each evaluating the function $\Lambda_b(\theta)$ under a particular scaling v, for θ belonging to some subset of $\mathbb{R}$. In this section we organize the results and present them in a simple and concise form so as to facilitate their application. We visit each of the three categories, i.e., I, II, and III, separately and recount the behavior of $\Lambda_b(\theta)$ in the various regimes of θ. We always assume the selection $a_n = n,\ n = 1, 2, \ldots$.

Case I: $R = \infty$

For Case I, G has a super–exponential tail. Grouping together the results of Theorems 4.3.1, 4.4.2 and 4.6.1, and Lemmas 4.5.1 and 4.5.2, we arrive at the following result.

Theorem 4.7.1 *Assume $R = \infty$ and select the scaling $\{v_n,\ n = 1, 2, \ldots\}$ so that $\lim_{n\to\infty} \frac{v_n}{n} = K,\ 0 \le K \le \infty$. Then, for each θ in $\mathbb{R}$, the limit $\Lambda_b(\theta)$ is given by*

$$\Lambda_b(\theta) = \begin{cases} \infty & \theta > 0, \\ 0 & \theta \le 0, \end{cases} \qquad \text{if} \quad K = \infty,$$

$$\Lambda_b(\theta) = \frac{\lambda}{K}\mathbf{E}\left[e^{K\theta\sigma} - 1\right] \qquad \theta \in \mathbb{R}, \qquad \text{if} \quad 0 < K < \infty,$$

$$\Lambda_b(\theta) = \lambda\mathbf{E}\left[\sigma\right]\theta \qquad \theta \in \mathbb{R}, \qquad \text{if} \quad K = 0.$$

Case II: $0 < R < \infty$

It is easy to check that the condition $0 < R < \infty$, or equivalently, $v_t^\star = O(t)$, is tantamount to G having an exponential tail. The next theorem is derived from Theorems 4.3.1, 4.4.2 and 4.6.1, and Lemmas 4.5.1 and 4.5.2.

Theorem 4.7.2 *Assume $0 < R < \infty$ and select the scaling $\{v_n,\ n = 1, 2, \ldots\}$ so that $\lim_{n\to\infty} \frac{v_n}{n} = K,\ 0 \le K \le \infty$. Then, for each $\theta \neq \frac{R}{K}$ in $\mathbb{R}$, the limit $\Lambda_b(\theta)$ is given by*

$$\Lambda_b(\theta) = \begin{cases} \infty & \theta > 0, \\ 0 & \theta \le 0, \end{cases} \qquad \text{if} \quad K = \infty,$$

$$\Lambda_b(\theta) = \begin{cases} \infty & \theta > \frac{R}{K}, \\ \frac{\lambda}{K}\mathbf{E}\left[e^{K\theta\sigma} - 1\right] & \theta < \frac{R}{K}, \end{cases} \qquad \text{if} \quad 0 < K < \infty,$$

$$\Lambda_b(\theta) = \lambda\mathbf{E}\left[\sigma\right]\theta \qquad \theta \in \mathbb{R}, \qquad \text{if} \quad K = 0.$$

Case III: $R = 0$

When $R = 0$, i.e., $v_t^\star = o(t)$, the situation is technically more involved, and additional growth assumptions are required on the scaling sequence $\{v_t^\star,\ t = 1, 2, \ldots\}$. Combining Theorems 4.3.1, 4.5.1, and 4.6.1, and Lemmas 4.5.1, 4.5.3 and 4.5.6, we construct the following theorem.

Theorem 4.7.3 *Assume $R = 0$, and that condition* **C1** *holds together with at least one of conditions* **C2a** *and* **C2b**. *Select the scaling $\{v_n,\ n = 1, 2, \ldots\}$ so that $\lim_{n\to\infty} \frac{v_n}{n} = K,\ 0 \le K \le \infty$ and $\lim_{n\to\infty} \frac{v_n^\star}{v_n} = C$, where $0 \le C \le \infty$. Then, for each $\theta \ne C$ in $\mathbb{R}$, the limit $\Lambda_b(\theta)$ is given by*

$$\Lambda_b(\theta) = \begin{cases} \infty & \theta > 0, \\ 0 & \theta \le 0, \end{cases} \qquad \text{if} \quad K = \infty,$$

$$\Lambda_b(\theta) = \begin{cases} \infty & \theta > 0, \\ \lambda \mathbf{E}[\sigma]\,\theta & \theta \le 0, \end{cases} \qquad \text{if} \quad K < \infty,\ C = 0,$$

$$\Lambda_b(\theta) = \begin{cases} \infty & \theta > C, \\ \lambda \mathbf{E}[\sigma]\,\theta & \theta < C, \end{cases} \qquad \text{if} \quad 0 < C < \infty,$$

$$\Lambda_b(\theta) = \lambda \mathbf{E}[\sigma]\,\theta \qquad \theta \in \mathbb{R}, \qquad \text{if} \quad C = \infty.$$

For cases I and II, when $K = 0$, we have from (4.1.6),

$$\Lambda^\star(x) = \sup_{\theta > 0}(x + c - r_{in})\theta = \infty, \qquad x > 0$$

with $r_{in} = \lambda \mathbf{E}[\sigma] < c$. This leads to the trivial upper bound $\gamma^\star = -\infty$ in (3.0.3), indicating that such a selection of v does not increase to infinity fast enough.

The same argument applies for case III when $C = \infty$.

When $K = \infty$ for cases I and II, and when $C = 0$ for case III, we have the opposite problem, i.e.,

$$\begin{aligned}
\Lambda^\star(x) &= \sup_{\theta\in\mathbf{R}}\Big((x+c)\theta - \Lambda_b(\theta)\Big) \\
&= \max\left(\sup_{\theta>0}\Big((x+c)\theta - \Lambda_b(\theta)\Big),\ \sup_{\theta\le 0}\Big((x+c)\theta - \Lambda_b(\theta)\Big)\right) \\
&= \sup_{\theta\le 0}\Big((x+c)\theta - \Lambda_b(\theta)\Big) \\
&= 0, \qquad x > 0
\end{aligned}$$

as the scaling v increases too fast to ∞.

Having disqualified these alternatives, we are left with the following straightforward selection for scaling v:

- Cases I and II:

$$\lim_{n\to\infty}\frac{v_n}{n} = K, \qquad 0 < K < \infty; \tag{4.7.1}$$

- Case III:

$$\lim_{n\to\infty}\frac{v_n^\star}{v_n} = C, \qquad 0 < C < \infty. \tag{4.7.2}$$

In case II, as $\lim_{n\to\infty}\frac{v_n^\star}{n} = R,\ 0 < R < \infty$, the selected scaling could alternatively be required to satisfy the condition (4.7.2), with $C = \frac{R}{K}$. We now see that Cases II and III could in fact have been grouped together for the purpose of presenting the results. This would have highlighted the influence of the service time rv σ as the scaling (4.7.2) applies to both cases. However, the original presentation has the advantage of suggesting a probabilistic viewpoint which further emphasizes the role played by the distribution of σ.

To see that, we study an input process closely related to the $M|G|\infty$ input process (λ, σ), according to which the work associated with a session is offered *instantaneously* to the buffer, rather than gradually as was the case for the $M|G|\infty$ input model.

4.7.1 Comparison with the instantaneous input model

Consider the instantaneous input process, $\{a_n,\ n = 0, 1, \ldots\}$ given by

$$a_0 = \sum_{i=1}^{b} \widehat{\sigma}_i; \quad a_n = \sum_{i=1}^{\beta_n} \sigma_{n,i}, \quad n = 1, 2, \ldots$$

where the families of i.i.d. rvs b, $\{\beta_n,\ n = 1, 2, \ldots\}$, $\{\sigma_{n,i},\ n = 1, 2, \ldots,\ i = 1, 2, \ldots\}$ and $\{\widehat{\sigma}_i,\ i = 1, 2, \ldots\}$ are described in Section 2.1.

The partial sum sequence $\{S^a_n,\ n = 0, 1, \ldots\}$ associated with $\{a_n,\ n = 0, 1, \ldots\}$ and given by

$$S^a_n = \sum_{t=0}^{n} a_t, \quad n = 0, 1, \ldots \tag{4.7.3}$$

is in essence the workload process offered to the infinite server queue.

Under the enforced independence assumptions, we readily get for all $n = 1, 2, \ldots$ that

$$\mathbf{E}\left[e^{\theta S^a_n}\right] = \mathbf{E}\left[\mathbf{E}\left[e^{\theta\widehat{\sigma}}\right]^b\right]\mathbf{E}\left[\mathbf{E}\left[e^{\theta\sigma}\right]^\beta\right]^n, \quad \theta \in \mathbb{R} \tag{4.7.4}$$

where β is a $\mathbb{N}$–valued rv which is Poisson distributed with parameter λ. Hence,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \ln \mathbf{E}\left[e^{\theta S^a_n}\right] &= \ln \mathbf{E}\left[\mathbf{E}\left[e^{\theta\sigma}\right]^\beta\right] \\ &= \lambda \mathbf{E}\left[e^{\theta\sigma} - 1\right], \quad \theta \in \mathbb{R} \end{aligned} \tag{4.7.5}$$

and going back to (4.1.1) and to Theorem 4.4.2, we conclude to the equality of the limiting logarithmic moment generating functions of the processes $\{S^b_n,\ n = 0, 1, \ldots\}$ and $\{S^a_n,\ n = 0, 1, \ldots\}$ (under the linear scaling (4.4.10)). This equality suggests a possible connection between these processes at the level of large deviations properties, and a natural way to formulate such a relationship passes through the notion of *exponential equivalence* [15, p. 114]. As pointed out in [15, Thm. 4.2.13, p. 114], such an exponential equivalence, once established, readily leads to the aforementioned equality. Therefore, a straightforward way to derive Theorem 4.4.2 would be to simply establish by a direct argument that the processes $\{n^{-1}(S_n^a - cn),\ t = 1, 2, \ldots\}$ and $\{n^{-1}(S_n^b - cn),\ n = 1, 2, \ldots$ are indeed exponentially equivalent. This is easily done with the help of the identities

$$\sum_{s=1}^{t} b_s^{\star(0)} = \sum_{n=1}^{b} \min(t, \widehat{\sigma}_n - 1) \tag{4.7.6}$$

and

$$\sum_{r=1}^{t} b_r^{(a)} = \sum_{s=1}^{t} \sum_{i=1}^{\beta_s} \min(t - s + 1, \sigma_{s,i}), \tag{4.7.7}$$

derived through the relations (2.1.3) and (2.1.4) for a stationary $M|G|\infty$ process.

As a consequence of this exponential equivalence, the sequence $\{n^{-1}(S_n^a - cn),\ n = 1, 2, \ldots\}$ also satisfies the Large Deviations Principle with rate functional identical to that of $\{n^{-1}(S_n^b - cn),\ n = 1, 2, \ldots\}$. This then implies that

$$\lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty^a > b\right] = \lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty > b\right] = -\gamma, \tag{4.7.8}$$

where

$$q_\infty^a =_{st} \sup\left(S_n^a - cn,\ n = 0, 1, \ldots\right), \tag{4.7.9}$$

represents the steady–state buffer content given by the Lindley recursion

$$q_0^a = \sum_{i=1}^{b} \widehat{\sigma}_i; \quad q_{n+1}^a = [q_n^a + a_{n+1} - c]^+, \quad n = 0, 1, \ldots,$$

under the obvious stability condition $\mathbf{E}\left[a_1\right] < c$.

Chapter 5

Buffer Asymptotics for the $M|G|\infty$ process

The key element in computing the upper and lower bounds $\gamma^\star$ and $\gamma_\star$, lies in the pair of scalings (a, v) which provide a non–trivial Λ. The role of a may be interpreted as that of the law of large numbers as opposed to v which is representative of a large deviations scaling.

Having resolved this question in the previous chapter, we now proceed to derive upper and lower bounds to $\ln \mathbf{P}\left[q_\infty > b\right]$, as outlined in Chapter 3.

5.1 Selection of h and g

We have seen in the previous chapter that the scalings a and v are **not** unique. Any pair of scalings (a, v) such that

- a is asymptotically linear, i.e., it satisfies (4.1.3),
- v satisfies (4.7.2) for sub–linear $v^\star$, and (4.7.1) otherwise,

is acceptable. Moreover, as indicated by Theorem 4.6.1, the corresponding functions Λ and $\Lambda^\star$ do not change their basic form for all acceptable scaling pairs, except for multiplicative scale factors. In what follows, we see that though these scale factors *do* influence the functions g and h, defined by (3.2.4), they *do not* appear in the final expressions for $\gamma^\star$ and $\gamma_\star$.

As before, we use the superscript (a, v) to identify the scalings used.

Lemma 5.1.1 *Let v and w be scaling functions obeying (4.6.1) with $0 < K < \infty$. For every pair of mappings $h^{(a,v)}, g^{(a,v)} : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying (3.2.4) with scalings (a, v), there exists a corresponding pair of mappings $h^{(a,w)}, g^{(a,w)} : \mathbb{R}_+ \to \mathbb{R}_+$ given by*

$$h^{(a,w)} = h^{(a,v)} \qquad \text{and} \qquad g^{(a,w)} = K g^{(a,v)} \tag{5.1.1}$$

that satisfy (3.2.4) with scalings (a, w).

Proof. For each $y > 0$, we note from (3.2.4) and (4.6.1) that

$$\begin{aligned} \lim_{b\to\infty} \frac{w_{a_l^{-1}(b/y)}}{h^{(a,v)}(b)} &= \lim_{b\to\infty} \frac{w_{a_l^{-1}(b/y)}}{v_{a_l^{-1}(b/y)}} \cdot \lim_{b\to\infty} \frac{v_{a_l^{-1}(b/y)}}{h^{(a,v)}(b)} \\ &= K g^{(a,v)}(y). \end{aligned} \tag{5.1.2}$$

A similar equality holds for the right inverse function a_r^{-1}, concluding the proof. ■

We now consider the particular selection $a_n = n$, $n = 1, 2, \ldots$. In keeping with our earlier convention, we drop the scaling a from the superscript and only specify the scaling v. The generalized inverses in this case are given by (3.2.10). In reference to (3.2.4), suppose now we can find mappings $h^v, g^v : \mathbb{R}_+ \to \mathbb{R}_+$ such that

$$\lim_{b\to\infty} \frac{v_{\lceil b/y \rceil}}{h^v(b)} = \lim_{b\to\infty} \frac{v_{\lfloor b/y \rfloor}}{h^v(b)} = g^v(y), \qquad y > 0. \tag{5.1.3}$$

A companion result to Lemma 5.1.1 can then be derived under the fairly general conditions described below:

Assumption E1: *For every pair of non–decreasing mappings* $n_1, n_2 : \mathbb{R}_+ \to \mathbb{I}_+$, *the condition* $\lim_{x\to\infty} \frac{n_1(x)}{n_2(x)} = 1$, *implies* $\lim_{x\to\infty} \frac{v_{n_1(x)}}{v_{n_2(x)}} = 1$.

Lemma 5.1.2 *Consider any scaling pair* (a, v) *such that* a *obeys (4.1.3) with* $0 < \alpha < \infty$, *and* v *satisfies Assumption* **E1**. *Then, for every pair of mappings* $h^v, g^v : \mathbb{R}_+ \to \mathbb{R}_+$ *satisfying (5.1.3), there exists a corresponding pair of mappings* $h^{(a,v)}, g^{(a,v)} : \mathbb{R}_+ \to \mathbb{R}_+$ *given by*

$$h^{(a,v)}(b) \;=\; h^v(b) \qquad \text{and} \qquad g^{(a,v)}(y) \;=\; g^v\left(\frac{y}{\alpha}\right), \tag{5.1.4}$$

that satisfy (3.2.4) with scalings (a, v).

Proof. Fix $x \geq 0$. By (3.2.1) we have $a_{a_l^{-1}(x)-1} \leq x \leq a_{a_l^{-1}(x)}$ Simple algebraic manipulations yield

$$\alpha a_{a_l^{-1}(x)-1} \leq \lceil \alpha x \rceil \leq \alpha a_{a_l^{-1}(x)} + 1,$$

so that

$$\alpha \cdot \frac{a_{a_l^{-1}(x)-1}}{a_l^{-1}(x)} \leq \frac{\lceil \alpha x \rceil}{a_l^{-1}(x)} \leq \alpha \cdot \frac{a_{a_l^{-1}(x)}}{a_l^{-1}(x)} + \frac{1}{a_l^{-1}(x)}.$$

By virtue of (4.1.3) and by the increasing nature of a_l^{-1}, we easily conclude that

$$\lim_{x\to\infty} \frac{\lceil \alpha x \rceil}{a_l^{-1}(x)} \;=\; 1. \tag{5.1.5}$$

Similarly, it can be shown that

$$\lim_{x\to\infty} \frac{\lfloor \alpha x \rfloor}{a_r^{-1}(x)} \;=\; 1. \tag{5.1.6}$$

Under Assumption **E1**, (5.1.5) and (5.1.6) imply

$$\lim_{x\to\infty} \frac{v_{\lceil \alpha x\rceil}}{v_{a_l^{-1}(x)}} \;=\; \lim_{x\to\infty} \frac{v_{\lfloor \alpha x\rfloor}}{v_{a_r^{-1}x)}} \;=\; 1,$$

which, on setting $x = b/y$ for a fixed $y > 0$, yields

$$\lim_{b\to\infty} \frac{v_{a_l^{-1}(b/y)}}{h^v(b)} \;=\; \lim_{b\to\infty} \frac{v_{a_r^{-1}(b/y)}}{h^v(b)} \;=\; g^v\left(\frac{y}{\alpha}\right),$$

via (5.1.3), thus completing the proof. ∎

We end this section with a generalized version of Lemmas 5.1.1 and 5.1.2:

Lemma 5.1.3 *Consider scalings* (a, v) *and* $(\widehat{a}, \widehat{v})$ *satisfying (4.6.5) with* $0 < \widehat{K}, \alpha, \widehat{\alpha} < \infty$. *Assume* v *(and therefore* $\widehat{v}$*) satisfies* **E1**. *Then, for every pair of mappings* $h^{(a,v)}, g^{(a,v)} : \mathbb{R}_+ \to \mathbb{R}_+$ *satisfying (3.2.4) with scalings* (a, v), *there exists a corresponding pair of mappings* $h^{(\widehat{a},\widehat{v})}, g^{(\widehat{a},\widehat{v})} : \mathbb{R}_+ \to \mathbb{R}_+$, *given by*

$$h^{(\widehat{a},\widehat{v})}(b) \;=\; h^{(a,v)}(b) \qquad \text{and} \qquad g^{(\widehat{a},\widehat{v})}(y) \;=\; \widehat{K} g^{(a,v)}\left(y\,\frac{\alpha}{\widehat{\alpha}}\right), \tag{5.1.7}$$

that satisfy (3.2.4) with scalings $(\widehat{a}, \widehat{v})$.

5.2 $\gamma_\star$ and $\gamma^\star$

Having computed $\Lambda^{(a,v)}$ as directed in Chapter 4, and selected appropriate mappings $h^{(a,v)}, g^{(a,v)}\, \mathbb{R}_+ \to \mathbb{R}_+$, we now realize our primary objective of providing asymptotic bounds of the form (3.0.2) and (3.0.3).

5.2.1 The lower bound

In order to compute the lower bound via Proposition 3.3.3, we require $\Lambda^{(a,v)}$ to satisfy conditions **GE 1** - **GE 3** of the Gärtner–Ellis Theorem (Theorem 3.2.1).

As per our remarks in Section 4.6, if the above–mentioned conditions hold for Λ computed under scalings (a, v), then, by Theorem 4.6.1, they must hold under any scaling pair $(\widehat{a}, \widehat{v})$ satisfying

$$0 < \lim_{n\to\infty} \frac{\widehat{v}_n}{v_n},\ \lim_{n\to\infty} \frac{\widehat{a}_n}{a_n} < \infty, \tag{5.2.1}$$

as long as the limit $\Lambda^{(\widehat{a},\widehat{v})}(\theta)$ exists for all θ in $\mathbb{R}$. Under these circumstances, Proposition 3.3.3 applies under both scaling pairs (a, v) and $(\widehat{a}, \widehat{v})$, and we have

$$\begin{aligned}
\gamma_\star^{(\widehat{a},\widehat{v})} &= \inf_{y>0} g^{(\widehat{a},\widehat{v})}(y)\Lambda^{\star(\widehat{a},\widehat{v})}(y) \\
&= \inf_{y>0} \widehat{K} g^{(a,v)}\left(y\ \frac{\alpha}{\widehat{\alpha}}\right)\ \frac{1}{\widehat{K}}\Lambda^{\star(a,v)}\left(y\ \frac{\alpha}{\widehat{\alpha}}\right) && (5.2.2) \\
&= \inf_{y>0} g^{(a,v)}(y)\Lambda^{\star(a,v)}(y) \\
&= \gamma_\star^{(a,v)}, && (5.2.3)
\end{aligned}$$

where (5.2.2) follows under assumption **E1**, via Lemma 5.1.3 and Theorem 4.6.1.

We note in passing that in our context, conditions **GE 1** and **GE 2** together, are equivalent to Assumption **A1** stated in Section 3.4.1.

5.2.2 The upper bound

To compute the upper bound $\gamma^\star$, and establish an equivalence similar to (5.2.3), we must first investigate conditions **A1**, **A2** and **A3** under various scalings.

Consider the scalings (a, v) and $(\widehat{a}, \widehat{v})$ described in Lemma 5.1.3. Further, assume that $\Lambda^{(a,v)}(\theta)$ and $\Lambda^{(\widehat{a},\widehat{v})}(\theta)$ exist for all θ in $\mathbb{R}$. In case **A1** holds under scalings (a, v), then, by Theorem 4.6.1, it must also hold under $(\widehat{a}, \widehat{v})$. Further, if (a, v) satisfies assumption **A2** with constant $\kappa^{(a,v)}$, it follows that $(\widehat{a}, \widehat{v})$ must also satisfy **A2** albeit with a different constant $\kappa^{(\widehat{a},\widehat{v})}$ given by

$$\kappa^{(\widehat{a},\widehat{v})} = \frac{\kappa^{(a,v)}}{\widehat{K}}. \tag{5.2.4}$$

By Lemma 5.1.3, the function h remains unaffected by the transformation from (a,v) to $(\widehat{a},\widehat{v})$. As assumption **A3** solely depends on the form of the function h, it must simultaneously hold for both pairs of scalings or none at all.

By these observations it is clear that if Propositions 3.4.3 and 3.4.4 apply under one pair of scalings, they also apply under the other. Of course, there still remains the possibility that the corresponding upper bounds $\gamma^{\star(a,v)}$ and $\gamma^{\star(\widehat{a},\widehat{v})}$ are not identical, a question we now speedily dismiss.

Proposition 5.2.1 *Consider the scalings (a,v) and $(\widehat{a},\widehat{v})$ described in Lemma 5.1.3. Assume that $\Lambda^{(a,v)}(\theta)$ and $\Lambda^{(\widehat{a},\widehat{v})}(\theta)$ exist for all θ in $\mathbb{R}$. Further, assume conditions* **A1**–**A3** *hold under (a,v). Then,*

(i) **A1**–**A3** *also hold under $(\widehat{a},\widehat{v})$;*

(ii) *The inequalities*

$$\limsup_{b\to\infty} \frac{1}{h^{(a,v)}(b)} \ln \mathbf{P}\left[q_\infty > b\right] \leq -\gamma^{\star(a,v)}, \tag{5.2.5}$$

and

$$\limsup_{b\to\infty} \frac{1}{h^{(\widehat{a},\widehat{v})(b)}} \ln \mathbf{P}\left[q_\infty > b\right] \leq -\gamma^{\star(\widehat{a},\widehat{v})} \tag{5.2.6}$$

hold, with $\gamma^{\star(\widehat{a},\widehat{v})} = \gamma^{\star(a,v)}$ given by Proposition 3.4.4, and $h^{(\widehat{a},\widehat{v})}(b) = h^{(a,v)}(b)$, for all $b > 0$.

Proof. The arguments leading to part (i) of Proposition 5.2.1 have already been discussed earlier in this section. Proposition 3.4.4 applies for both pairs of scalings, yielding (5.2.5) and (5.2.6), with

$$\gamma^{\star(\widehat{a},\widehat{v})} \quad = \quad \sup_{y>0} g^{(\widehat{a},\widehat{v})}(y) \left(\min\left(\theta_0^{(\widehat{a},\widehat{v})} y, \Lambda^{\star(\widehat{a},\widehat{v})}(0)\right) - \kappa^{(\widehat{a},\widehat{v})} \right), \tag{5.2.7}$$

and

$$\begin{aligned}
\theta_0^{(\widehat{a},\widehat{v})} &= \sup\{\theta > 0 : \Lambda^{(\widehat{a},\widehat{v})}(\theta) < 0\} \\
&= \sup\left\{\theta > 0 : \Lambda^{(a,v)}\left(\widehat{K}\, \frac{\widehat{\alpha}}{\alpha}\, \theta\right) < 0\right\} \qquad (5.2.8) \\
&= \sup\left\{\frac{1}{\widehat{K}}\, \frac{\alpha}{\widehat{\alpha}}\, \theta' > 0 : \Lambda^{(a,v)}(\theta') < 0\right\} \\
&= \frac{1}{\widehat{K}}\, \frac{\alpha}{\widehat{\alpha}}\, \theta_0^{(a,v)},
\end{aligned}$$

where (5.2.8) follows on account of Theorem 4.6.1. Incorporating this result into (5.2.7) and employing (5.2.4), we have

$$\begin{aligned}
\gamma^{\star(\widehat{a},\widehat{v})} &= \sup_{y>0} g^{(\widehat{a},\widehat{v})}(y) \left(\min\left(\frac{1}{\widehat{K}}\, \frac{\alpha}{\widehat{\alpha}}\, \theta_0^{(a,v)} y, \Lambda^{\star(\widehat{a},\widehat{v})}(0)\right) - \frac{1}{\widehat{K}}\, \kappa^{(a,v)}\right) \\
&= \frac{1}{\widehat{K}} \sup_{y>0} g^{(\widehat{a},\widehat{v})}(y) \left(\min\left(\frac{\alpha}{\widehat{\alpha}}\, \theta_0^{(a,v)} y, \Lambda^{\star(a,v)}(0)\right) - \kappa^{(a,v)}\right)
\end{aligned}$$

by Theorem 4.6.1. Finally, using Lemma 5.1.3, we have

$$\begin{aligned}
\gamma^{\star(\widehat{a},\widehat{v})} &= \frac{1}{\widehat{K}} \sup_{y>0} \widehat{K}\, g^{(a,v)}\left(\frac{\alpha}{\widehat{\alpha}}\, y\right) \left(\min\left(\theta_0^{(a,v)}\, \frac{\alpha}{\widehat{\alpha}}\, y, \Lambda^{\star(a,v)}(0)\right) - \kappa^{(a,v)}\right) \\
&= \sup_{y'>0} g^{(a,v)}(y') \left(\min\left(\theta_0^{(a,v)} y', \Lambda^{\star(a,v)}(0)\right) - \kappa^{(a,v)}\right) \\
&= \gamma^{\star(a,v)}.
\end{aligned}$$

■

In the next Proposition, proved in Appendix D.1, we establish that $\gamma^{\star(\widehat{a},\widehat{v})} = \gamma^{\star(a,v)}$, even in the case when **A3** fails to hold.

Proposition 5.2.2 *Consider the scalings (a, v) and $(\widehat{a}, \widehat{v})$ described in Lemma 5.1.3. Assume that $\Lambda^{(a,v)}(\theta)$ and $\Lambda^{(\widehat{a},\widehat{v})}(\theta)$ exist for all θ in $\mathbb{R}$. Further, assume conditions* **A1**–**A2** *hold under (a, v). Then,*

(i) **A1**–**A2** *also hold under* $(\widehat{a}, \widehat{v})$;

(ii) *The inequalities (5.2.5) and (5.2.6) still hold, with* $\gamma^{\star(\widehat{a},\widehat{v})} = \gamma^{\star(a,v)}$ *given by Proposition 3.4.3, and* $h^{(\widehat{a},\widehat{v})}(b) = h^{(a,v)}(b)$, *for all* $b > 0$.

The fact that the bounds $\gamma^\star$ and $\gamma_\star$ remain unaffected by the transformation from (a, v) to $(\widehat{a}, \widehat{v})$ proves quite convenient. In conjunction with the results of Section 4.6, it implies that our entire analysis so far is dictated solely by the *asymptotic* behavior of scalings a and v. Therefore, in applying the results of Chapter 4, the appropriate conditions can all be checked by replacing (a, v) by any other pair $(\widehat{a}, \widehat{v})$, asymptotically equivalent to it in the sense of (4.6.5), and hopefully more tractable analytically. We shall refer to any such scalings as auxiliary scalings, and will employ them while checking the necessary conditions.

5.3 Buffer Asymptotics for the $M|G|\infty$ process

We are now close to realizing our initial objective of deriving asymptotic bounds of the kind (3.0.2) and (3.0.3) for a queue fed by $M|G|\infty$ traffic. For any particular distribution G, we begin by identifying the appropriate pair of auxiliary scalings (a, v) and the corresponding function Λ, via Theorems 4.7.1, 4.7.2 or 4.7.3. Next, we select functions h and g as outlined earlier in this chapter. Finally, applying the results derived in Chapter 3, we achieve the asymptotic bounds to the tail probability of buffer exceedance.

5.3.1 Auxiliary scalings

As we have already seen, beyond being asymptotically linear, there are no restrictions on the selection of sequence a. We therefore make the most convenient choice, and set $a_n = n, \; n = 1, 2, \ldots$ without further comment.

The selection of an auxiliary scaling in lieu of the large deviations scaling v, is closely linked with the form of $v^\star$. However, as we shall now see, the determination of such a scaling is possible even without the explicit computation of the function $v^\star$.

With this in mind, let $w^\star : \mathbb{R}_+ \to \mathbb{R}_+$ denote a mapping such that

$$e^{-w_n^\star} \equiv \mathbf{P}\left[\sigma > n\right], \qquad n = 0, 1, \ldots, \tag{5.3.1}$$

We make the reasonable assumption that the limit

$$\lim_{n\to\infty} \frac{\ln n}{w_n^\star} \equiv W, \tag{5.3.2}$$

exists for some non–negative W (possibly infinite). The following lemma then holds.

Lemma 5.3.1 $\mathbf{E}\left[\sigma\right]$ *is finite (resp. infinite) if* $W < 1$ *(resp.* > 1*).*

Proof.

$$\begin{aligned} \mathbf{E}\left[\sigma\right] &= \sum_{n=0}^{\infty} \mathbf{P}\left[\sigma > n\right] \\ &= \sum_{n=0}^{\infty} e^{-w_n^\star}. \end{aligned} \tag{5.3.3}$$

Assume $0 < W < \infty$, and pick any $\delta > 0$ such that $\delta W < 1$. Then, (5.3.2) implies that there exists an integer $N(\delta)$ so that

$$\frac{1}{W} - \delta \le \frac{w_n^\star}{\ln n} \le \frac{1}{W} + \delta, \qquad n > N(\delta), \tag{5.3.4}$$

and

$$\sum_{n=N(\delta)+1}^{\infty} n^{-(\frac{1}{W}+\delta)} \leq \sum_{n=N(\delta)+1}^{\infty} e^{-w_n^\star} \leq \sum_{n=N(\delta)+1}^{\infty} n^{-(\frac{1}{W}-\delta)}.$$

Clearly, all three sums are infinite for $W > 1$, and finite for $W < 1$, proving the result for all non–zero, finite $W \neq 1$. The cases $W = 0$ and ∞ are proved along similar lines. ■

As we do require $\mathbf{E}[\sigma]$ to be finite, we proceed with the assumption that $0 \leq W < 1$. We ignore the disagreeable case $W = 1$, for which the finiteness of $\mathbf{E}[\sigma]$ may not be established without additional information.

Lemma 5.3.2 *Assume that $0 \leq W < 1$. In case $W = 0$, further assume that the sequence $\dfrac{\ln n}{w_n^\star}$ is eventually decreasing. Then the limit*

$$\lim_{n\to\infty} \frac{v_n^\star}{w_n^\star} = 1 - W \tag{5.3.5}$$

always holds.

This result immediately suggests a simple procedure by which the auxiliary scaling may be identified at a glance. If the limit

$$\lim_{n\to\infty} \frac{w_n^\star}{n} = L, \qquad 0 \leq L \leq \infty, \tag{5.3.6}$$

exists, then we have

$$R = (1 - W)L \tag{5.3.7}$$

under the assumptions of Lemma 5.3.2, where R and W are defined by (2.3.1) and (5.3.2), respectively. Having computed R, we already know that the auxiliary scaling $v_n = n,\ n = 1, 2, \ldots$ works for Cases I and II, while in Case III we may set $v_n = (1 - W)w_n^\star,\ n = 1, 2, \ldots$, in accordance with Lemma 5.3.2 and Theorem 4.7.3.

5.3.2 Cases I and II: $R > 0$

We have the auxiliary scaling $v_n = n$, hence we set $h(b) = b$, and correspondingly $g(x) = 1/x$. Then by Theorems 4.7.1 and 4.7.2, we have

$$\Lambda(\theta) \quad = \quad \begin{cases} \lambda \mathbf{E}\left[e^{\theta\sigma} - 1\right] - c\theta & \theta < R, \\ \infty & \theta > R, \end{cases} \tag{5.3.8}$$

indicating that condition **A1**, and by their equivalence, conditions **GE 1** and **GE 2** are satisfied. As $\dfrac{\ln n}{n}$ is asymptotically monotonically decreasing, **A2** also holds with $\kappa = 0$.

The bounds $\gamma^\star$ and $\gamma_\star$ are now available without further calculation, as we recall that the details have already been worked out as a special case in Section 3.5. The upper bound given by

$$\begin{aligned} \gamma^\star &= \sup\{\theta > 0 : \; \Lambda(\theta) < 0\} \\ &= \sup\left\{0 < \theta < R : \; \mathbf{E}\left[e^{\theta\sigma}\right] < 1 + \theta\,\frac{c}{\lambda}\right\}, \end{aligned} \tag{5.3.9}$$

holds under no additional assumptions, while for the lower bound, $\gamma_\star = \gamma^\star$ in the event that assumption **GE 3** is satisfied.

Denoting the input rate to the multiplexer by $r_{in} = \lambda \mathbf{E}[\sigma]$, and the utilization factor by

$$\rho = \frac{r_{in}}{c} \quad = \quad \frac{\lambda \mathbf{E}[\sigma]}{c}, \tag{5.3.10}$$

(5.3.9) may be alternatively expressed as

$$\gamma^\star \quad = \quad \sup\{0 < \theta < R : \; \rho < f(\theta)\}, \tag{5.3.11}$$

where the mapping $f : (0, \infty) \to (0, 1)$ is defined by

$$f(\theta) \quad \equiv \quad \frac{\mathbf{E}[\theta\sigma]}{\mathbf{E}\left[e^{\theta\sigma}\right] - 1}, \qquad \theta > 0. \tag{5.3.12}$$

The function f being strictly decreasing, we can define the corresponding inverse function $f^{-1} : (0,1) \to (0,\infty)$, allowing (5.3.11) to be rewritten as

$$\begin{aligned} \gamma^\star &= \sup\left\{0 < \theta < R : \ \theta < f^{-1}(\rho)\right\} \\ &= f^{-1}(\rho). \end{aligned} \tag{5.3.13}$$

We present these results more formally in the following proposition:

Proposition 5.3.1 *Assume $R > 0$ and $\lambda \mathbf{E}[\sigma] < c$. Then, under condition* **GE 3**, *we have*

$$\lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}[q_\infty > b] = -\gamma, \tag{5.3.14}$$

with

$$\gamma = f^{-1}(\rho), \tag{5.3.15}$$

where the function $f^{-1} : (0,1) \to (0,\infty)$ is as defined in (5.3.12) and ρ is given by (5.3.10).

It is noteworthy that while the computation of the limit $\gamma^\star$ is closely linked with the distribution G, its functional dependence on the parameters c and λ is *only* through the utilization factor ρ.

5.3.3 Case III: $R = 0$

Select $v_n = (1 - W)w_n^\star, \ n = 1, 2, \ldots$. Then, under conditions **C1** and at least one of **C2a** and **C2b**,

$$\Lambda(\theta) = \begin{cases} (\lambda \mathbf{E}[\sigma] - c)\theta & \theta < 1 \\ \infty & \theta > 1 \end{cases} \tag{5.3.16}$$

by Theorem 4.7.3 and Lemma 4.6.2, thus meeting the requirements imposed by **A1**, **GE 1** and **GE 2**.

Unfortunately, the same cannot be said of assumption **GE 3**, which fails to hold as

$$\lim_{\theta \to 1} \left| \frac{\partial \Lambda(\theta)}{\partial \theta} \right| = c - \lambda \mathbf{E}\left[\sigma\right] \neq \infty, \tag{5.3.17}$$

indicating that Λ is *not* steep, and precluding the use of Proposition 3.3.3 to provide the lower bound.

However, the upper bound may still be computed under the premise that $\kappa \neq \infty$, and **A2** holds. By (3.4.7) and (5.3.2), we have

$$\begin{aligned} \kappa &= \frac{1}{1-W} \lim_{n \to \infty} \frac{\ln n}{w_n^\star} \\ &= \frac{W}{1-W}. \end{aligned} \tag{5.3.18}$$

Lemma 5.3.1 dictates that $0 \leq W < 1$ for finite $\mathbf{E}\left[\sigma\right]$, whence $0 \leq \kappa < \infty$. Furthermore, $\kappa = 0$ **iff** $W = 0$, in which case under the assumptions of Lemma 5.3.2, the sequence $\dfrac{\ln n}{w_n^\star}$, and hence $\dfrac{\ln n}{v_n}$ is eventually decreasing. In other words, condition **A2** always holds under the assumptions of Lemma 5.3.2.

For a number of distributions G, the function v belongs to the class of *regularly varying functions*. As mentioned earlier in Lemma 3.2.2, this suggests a natural choice for h as the piecewise–continuous interpolation of the auxiliary scaling sequence v, in which case

$$g(y) = y^{-\rho}, \quad \rho \geq 0. \tag{5.3.19}$$

Finally, we recall that the condition $\Lambda^\star(0) > \kappa$ must hold in order to ensure a non–trivial upper bound. Reference to (3.2.16) and (5.3.18) translates this requirement to $c - \lambda \mathbf{E}\left[\sigma\right] > \frac{W}{1-W}$.

5.3.4 Beyond Large Deviations Techniques

At this point, we have seen that in the non–exponential case a non–trivial upper bound (3.0.2) always holds (under the assumption $\Lambda^\star(0) > \kappa$), whereas the lower bound (3.0.3) is in doubt, at least if one insists on going through Proposition 3.3.3. However, the possibility that the process $\{t^{-1}(S_t^b - ct),\ t = 1, 2, \ldots\}$ satisfies the Large Deviations Principle with rate functional $\Lambda^\star$ *still* remains, though we cannot look to the Gärtner–Ellis Theorem for its proof. In other words, the lower bound suggested by (3.3.7) may yet prove valid, but cannot be derived under the terms of Proposition 3.3.3.

Supposing for the sake of argument that this were indeed the case, would the bounds thus provided be tight ? Could we then replace (3.0.2) and (3.0.3) by the stronger limiting equality (3.0.1) ?

Consider a heavy tailed distribution G for which $v_n^\star \sim K \ln n$ is a slowly varying function of n. In this case, $g(y) = 1$ and the upper and lower bounds derived earlier in Section 3.5 and given by $\Lambda^\star(0) - \kappa$ and $\Lambda^\star(0)$ respectively, are clearly *not* tight.

Although this inequality is not sufficient by itself to reach any negative conclusion concerning the existence of the lower bound (3.3.7), it strongly suggests that in cases where G is heavy tailed, the investigation of the buffer asymptotics will require that we look beyond large deviations techniques. Going back to the heuristics given in [38], we attribute this to the fact that now buffer exceedances cannot be explained entirely by large deviations excursions in the arrival stream, as there is a need to take into consideration the effect of a single customer with a large workload – the tail of the distribution has become too heavy to neglect such a customer! Hence, any argument based on large deviations techniques *alone* is bound to fall short. However, we conjecture that (3.0.1) still holds with scaling $h(b) = \ln b$ as specified through (3.2.4) but of course with a different value for γ.

5.4 Alternate Bounds

The suspicion that the failure of Proposition 3.3.3 is linked to the methodology based on Large Deviations was confirmed in [45] where the issue of devising asymptotics for $\mathbf{P}[q_\infty > b]$ was revisited by means of basic principles; both lower and upper bound asymptotics were proposed and in some case the latter are tighter than the ones given here. This section is devoted to a discussion of the results of [45], inclusive of a comparison with the bounds obtained here.

The approach of [45] is most informative when applied to the subset of distributions with non–exponential tails known as *sub–exponential* distributions [9, 21]: An $\mathbb{R}_+$–valued rv X is said to be *sub–exponential*, written $X \in \mathcal{S}$, if

$$\lim_{x\to\infty} \frac{\mathbf{P}[X + X' > x]}{\mathbf{P}[X > x]} = 2 \qquad (5.4.20)$$

where X' is an independent copy of X. The terminology is substantiated by the fact that under (5.4.20) [21], we have

$$\lim_{x\to\infty} e^{\delta x}\mathbf{P}[X > x] = \infty, \quad \delta > 0. \qquad (5.4.21)$$

The clue that the bounds obtained thus far could indeed be improved comes via from the following well–known result of Pakes [50]:

Proposition 5.4.1 *Consider a $GI/GI/1$ queue with i.i.d. service times $\{\mu_n,\ n = 1, 2, \ldots\}$ distributed according to G, and i.i.d. inter–arrival times $\{\tau_n,\ n = 1, 2, \ldots\}$. Let μ and τ denote the generic service time and inter–arrival time rvs, respectively.*

Assume $\mathbf{E}[\mu] < \mathbf{E}[\tau]$. If the forward recurrence time $\widehat{\mu} \in \mathcal{S}$, it holds that the steady-state queue-size $q_\infty^P \in \mathcal{S}$ with

$$\mathbf{P}\left[q_\infty^P > b\right] \sim \frac{\mathbf{E}[\mu]}{\mathbf{E}[\tau] - \mathbf{E}[\mu]} \mathbf{P}[\widehat{\mu} > b]. \tag{5.4.22}$$

When applied to the model with instantaneous inputs introduced earlier in Section 4.7.1, Proposition 5.4.1 states that if $\widehat{a} \in \mathcal{S}$,

$$\mathbf{P}[q_\infty^a > b] \sim \frac{r_{in}}{c - r_{in}} \mathbf{P}[\widehat{a} > b] \tag{5.4.23}$$

under the stability condition $r_{in} < c$.

If $\sigma \in \mathcal{S}$, then $a \in \mathcal{S}$ with $\mathbf{P}[a > t] \sim \mathbf{E}[\beta]\mathbf{P}[\sigma > t]$ by [21], and standard arguments now yield

$$\int_b^\infty \mathbf{P}[a > t]\,dt \sim \mathbf{E}[\beta] \int_b^\infty \mathbf{P}[\sigma > t]\,dt \tag{5.4.24}$$

so that $\mathbf{P}[\widehat{a} > b] \sim \mathbf{P}[\widehat{\sigma} > b]$ and $\widehat{a} \in \mathcal{S}$ whenever $\widehat{\sigma} \in \mathcal{S}$. Combining these comments, we immediately get

Proposition 5.4.2 *If $\sigma \in \mathcal{S}$ and $\widehat{\sigma} \in \mathcal{S}$ with $r_{in} < c$, then $q_\infty^a \in \mathcal{S}$ with*

$$\mathbf{P}[q_\infty^a > b] \sim \frac{r_{in}}{c - r_{in}} \mathbf{P}[\widehat{\sigma} > b]. \tag{5.4.25}$$

5.4.1 Improved upper bounds

The expressions (4.7.6) and (4.7.7) readily lead to the bound

$$q_\infty \le \sum_{n=1}^{b} \widehat{\sigma}_n + q_\infty^a. \tag{5.4.26}$$

This observation, when coupled with the asymptotics (5.4.25), forms the basis for the following asymptotics:

Proposition 5.4.3 *If $\sigma \in \mathcal{S}$ and $\widehat{\sigma} \in \mathcal{S}$ with $r_{in} < c$, then*

$$\limsup_{b\to\infty} \frac{\mathbf{P}[q_\infty > b]}{\mathbf{P}[\widehat{\sigma} > b]} \le r_{in} + \frac{r_{in}}{c - r_{in}}, \tag{5.4.27}$$

whence

$$\limsup_{b\to\infty} \frac{1}{w^\star_{\lceil b\rceil}} \ln \mathbf{P}[q_\infty > b] \le -(1 - W). \tag{5.4.28}$$

The details of the derivation of (5.4.27), available in [45], rely on well–known properties of sub–exponential rvs [21], while (5.4.28) follows directly through (2.3.1) and (5.3.2) via Lemma 5.3.2. The release rate c does *not* appear in (5.4.28), thereby suggesting that (5.4.28) will not always improve on upper asymptotics obtained previously.

5.4.2 General lower bounds

The following lower bound holds in great generality and is essentially Proposition 3.1 in [45] couched in the notation used here. Details of the calculations are omitted in the interest of brevity:

Proposition 5.4.4 *For any $\{1, 2, \ldots\}$-valued rv σ, it holds that*

$$-\gamma_\star(1 - W) \le \liminf_{b\to\infty} \frac{1}{w^\star_{\lfloor b\rfloor}} \ln \mathbf{P}[q_\infty > b] \tag{5.4.29}$$

with

$$\gamma_\star = \inf_{y>0} \left((\lfloor c - r_{in} + y\rfloor + 1) \limsup_{b\to\infty} \frac{v^\star_{\lfloor b\rfloor}}{v^\star_{\lfloor by\rfloor}} \right). \tag{5.4.30}$$

In the situation when $v^\star$ (and therefore by Lemma 5.3.2, $w^\star$) is regularly varying, (5.4.30) lends itself to further simplification, and we have

$$\gamma_\star = \inf_{y>0} \Big(\lfloor c - r_{in} + y + 1\rfloor g(y) \Big). \tag{5.4.31}$$

It is noteworthy that in Case II with $0 < R < \infty$, we have $w_t^\star \sim Rt$ so that

$$\gamma_\star = \inf_{y>0} \frac{\lfloor c - r_{in} + y \rfloor + 1}{y} = 1 \tag{5.4.32}$$

and by Proposition 5.4.4 we get

$$-R \leq \liminf_{b\to\infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty > b\right]. \tag{5.4.33}$$

Interestingly enough, this lower bound is not as good as the one obtained in Proposition 5.3.1 by applying the general buffer asymptotics based on large deviations arguments.

Chapter 6

Examples and Simulation Results

We now proceed to various examples which illustrate the details of each of the three cases. The examples considered here are constructed by taking the $\{1, 2, \ldots\}$–valued rv σ to be of the form $\sigma =_{st} \lceil X \rceil$ (or $\sigma =_{st} \lfloor X \rfloor$), where X is an integrable $\mathbb{R}_+$–valued rv with $\mathbf{P}[X = 0] = 0$. The function $w^\star$, defined in (5.3.1), is then given by

$$w^\star_n \quad = \quad - \ln \mathbf{P}[X > n], \qquad n = 0, 1, \ldots. \tag{6.0.1}$$

The examples are presented in order of increasing tail in G; in other words in order of increasing time dependence in the input process. We always assume $\rho < 1$, where $\rho \equiv \frac{r_{in}}{c} = \frac{\lambda \mathbf{E}[\sigma]}{c}$.

6.1 Super–exponential distributions

6.1.1 The Deterministic case

We begin with the simplest case of all, where $\sigma = \zeta$ for some constant ζ in $\mathbb{I}^+$.

Proposition 6.1.1 *If $\sigma = \zeta$, where ζ is a constant in I_+, then*

$$\lim_{b \to \infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty > b\right] = -\gamma_{\text{Deterministic}}, \tag{6.1.1}$$

where

$$\gamma_{\text{Deterministic}} = \frac{1}{\zeta} \cdot f_D^{-1}(\rho), \tag{6.1.2}$$

with $f_{\mathrm{D}}(\theta) \equiv \dfrac{\theta}{e^{\theta} - 1}, \ \theta > 0.$

Proof. By (5.3.1), we have

$$w_n^\star = \begin{cases} 0 & n \le \zeta \\ \infty & n > \zeta. \end{cases} \tag{6.1.3}$$

Hence $R = \infty$ by Lemma 5.3.2 and (4.3.3), and selecting the auxiliary scaling $v_n = n, \ n = 1, 2, \ldots$, we have

$$\Lambda(\theta) = \lambda\left(e^{\theta\zeta} - 1\right) - c\theta, \qquad \theta \in \mathbb{R}, \tag{6.1.4}$$

via Theorem 4.7.1.

As condition **GE 3** clearly holds, Proposition 5.3.1 applies, yielding (6.1.1) with $\gamma_{\text{Deterministic}}$ given by (6.1.2). ■

A plot of $f_{\mathrm{D}}^{-1}(\rho)$ versus ρ is presented in Figure 1, using which $\gamma_{\text{Deterministic}}$ may be calculated for various values of ρ and ζ.

The simulation results for $\zeta = 4$ and 5 displayed in Figure 3, are as predicted by Proposition 6.1.1. The corresponding values of $f_{\mathrm{D}}^{-1}(\rho)$ and $\gamma_{\text{Deterministic}}$ are provided in Table 6.1 for easy reference.

ρ	$f_{\mathrm{D}}^{-1}(\rho)$	$\gamma_{\text{Deterministic}}$	
		$\zeta = 4$	$\zeta = 5$
0.5	1.2564	0.3141	0.2513
0.7	0.6755	0.1689	0.1351
0.9	0.2071	0.0518	0.0414

Table 6.1: $\gamma_{\text{Deterministic}}$

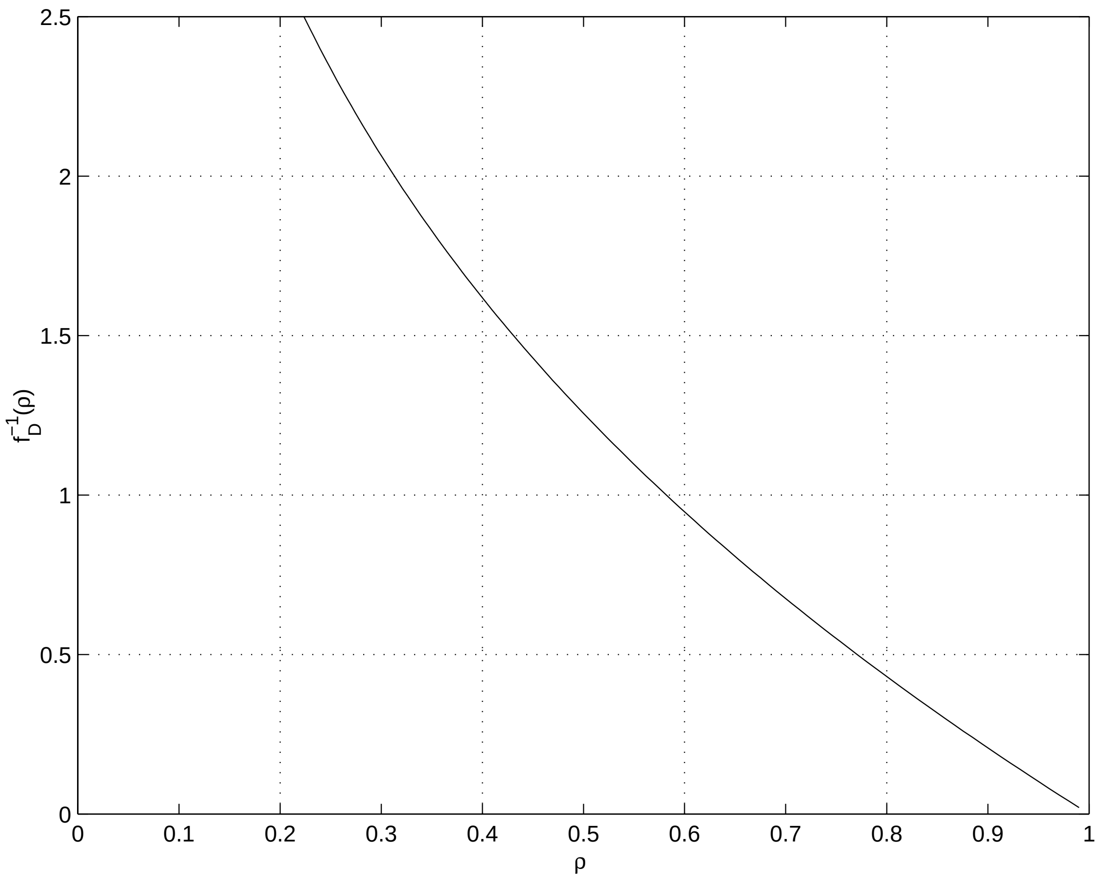

Figure 1: $f_{\mathrm{D}}^{-1}(\rho)$ versus ρ

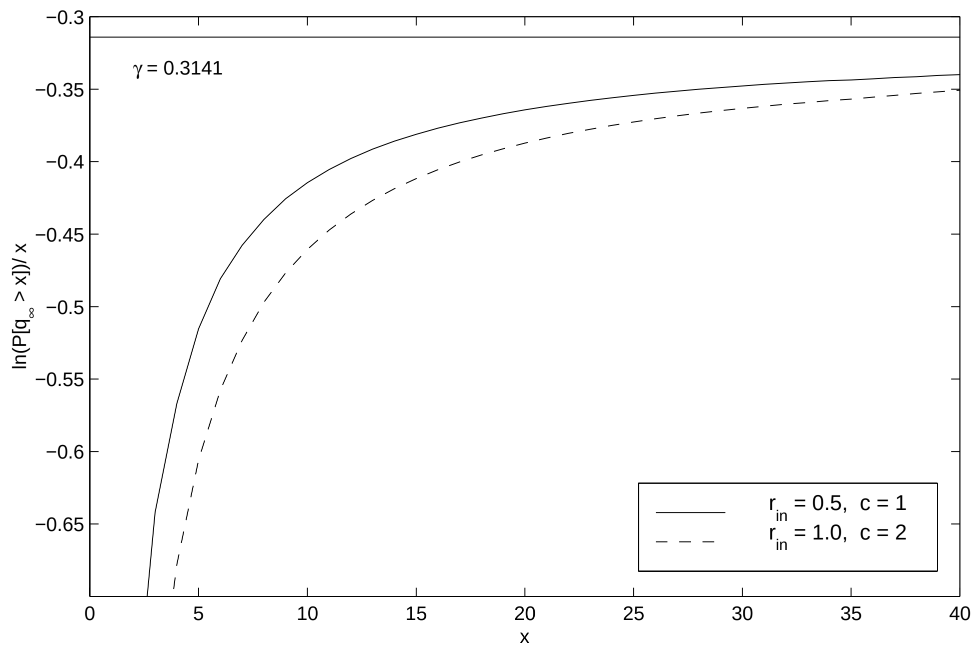

Figure 2: Tail probability vs. buffer size: Deterministic ($\zeta = 4$)

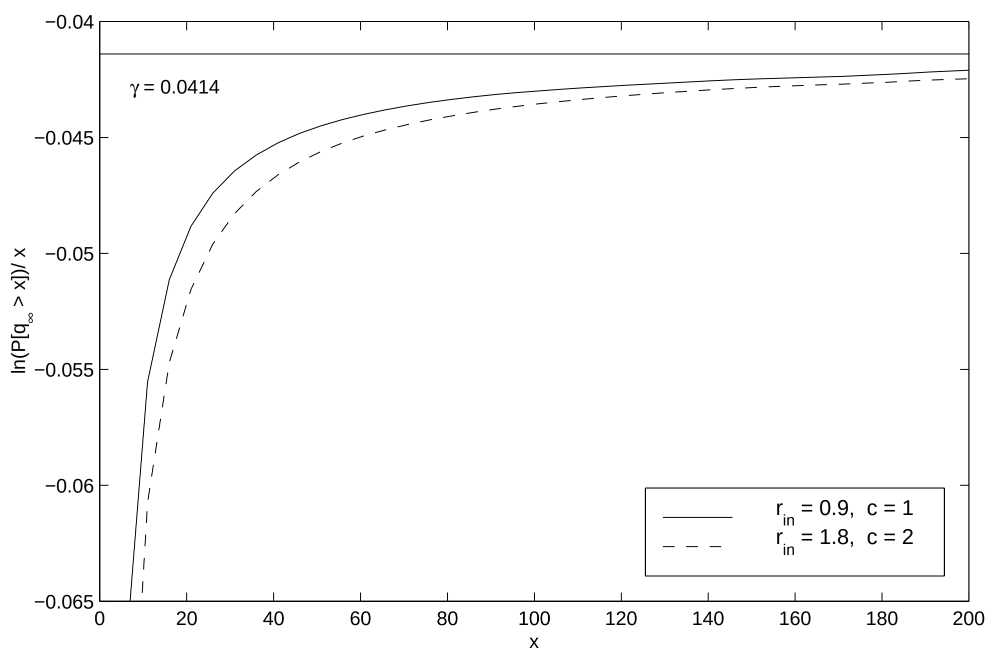

Figure 3: Tail probability vs. buffer size: Deterministic ($\zeta = 5$)

6.1.2 The Rayleigh case

A continuous rv X is said to be a Rayleigh rv with parameter $\alpha > 0$, if

$$\mathbf{P}\left[X \leq x\right] = 1 - e^{-\frac{x^2}{2\alpha^2}}, \qquad x \geq 0. \tag{6.1.5}$$

The rv $\sigma =_{st} \lceil X \rceil$ is then said to have a discrete Rayleigh distribution with parameter $\alpha > 0$, and pmf $G = \{g_r,\ r = 1, 2, \ldots\}$, where

$$g_r = e^{-\frac{(r-1)^2}{2\alpha^2}} - e^{-\frac{r^2}{2\alpha^2}}, \qquad r = 1, 2, \ldots. \tag{6.1.6}$$

Proposition 6.1.2 *If G is a discrete Rayleigh distribution with parameter $\alpha > 0$ then*

$$\lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty > b\right] = -\gamma_{\text{Rayleigh}}, \tag{6.1.7}$$

where

$$\gamma_{\text{Rayleigh}} \;=\; f_\alpha^{-1}(\rho), \tag{6.1.8}$$

and

$$f_\alpha(\theta) \;=\; \left(\frac{e^\theta - 1}{\theta} \cdot \frac{1}{\mathbf{E}\left[\sigma\right]} \sum_{r=0}^{\infty} e^{\theta r - \frac{r^2}{2\alpha^2}}\right)^{-1}. \tag{6.1.9}$$

Proof. Fix $n > 0$. By (6.0.1) and (6.1.5), we know that $w_n^\star = \dfrac{n^2}{2\alpha^2}$, implying via Lemma 5.3.2 and (4.3.3), that $R = \infty$. Applying Theorem 4.7.1 with the auxiliary scaling $v_n = n,\ n = 1, 2, \ldots$, we have

$$\Lambda(\theta) \;=\; -c\theta + \lambda\left(e^\theta - 1\right) \sum_{r=0}^{\infty} e^{\theta r - \frac{r^2}{2\alpha^2}}, \qquad \theta > 0.$$

Clearly, **GE 3** holds, hence the desired result follows via Proposition 5.3.1 upon noting that

$$\mathbf{E}\left[e^{\sigma\theta}\right] \;=\; 1 + \left(e^\theta - 1\right) \sum_{r=0}^{\infty} e^{\theta r - \frac{r^2}{2\alpha^2}}, \qquad \theta > 0.$$

■

A closed form expression for γ_{Rayleigh} is not easily calculated. However, a numerical solution is readily available and is provided in Table 6.2 for select values of ρ, with $\alpha = 2.0$ and 6.0. Figure 4 displays plots of $f_\alpha^{-1}(\rho)$ vs. ρ, for $\alpha = 2.0$ and 6.0. The corresponding simulated results presented in Figure 5, are in accordance with Proposition 6.1.2.

6.2 Exponential distributions

6.2.1 The geometric case

The geometric pmf $G = \{g_r,\ r = 1, 2, \ldots\}$ of parameter q $(0 < q < 1)$, is given by

$$g_r \equiv \mathbf{P}[\sigma = r] = (1-q)q^{r-1}, \qquad r = 1, 2, \ldots \tag{6.2.1}$$

Proposition 6.2.1 *If G is a geometric pmf of parameter q, with $0 < q < 1$, given by (6.2.1), then*

$$\lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}[q_\infty > b] = -\gamma^\star_{\text{Geometric}} \tag{6.2.2}$$

where

$$\gamma_{\text{Geometric}} = f_G^{-1}(\rho), \tag{6.2.3}$$

with

$$f_G(\theta) = \frac{\theta}{e^\theta - 1} \cdot \frac{1 - qe^\theta}{1-q}, \qquad 0 < \theta < -\ln q. \tag{6.2.4}$$

Proof. Fix $n > 0$. From (6.0.1) and (6.2.1) we remark that $w_n^\star = (-\ln q)n$,

ρ	$\gamma_{\text{Rayleigh}} = f_\alpha^{-1}(\rho)$	
	$\alpha = 2.0$	$\alpha = 6.0$
0.5	0.3267	0.1173
0.7	0.1811	0.0653
0.9	0.0569	0.0206

Table 6.2: γ_{Rayleigh}

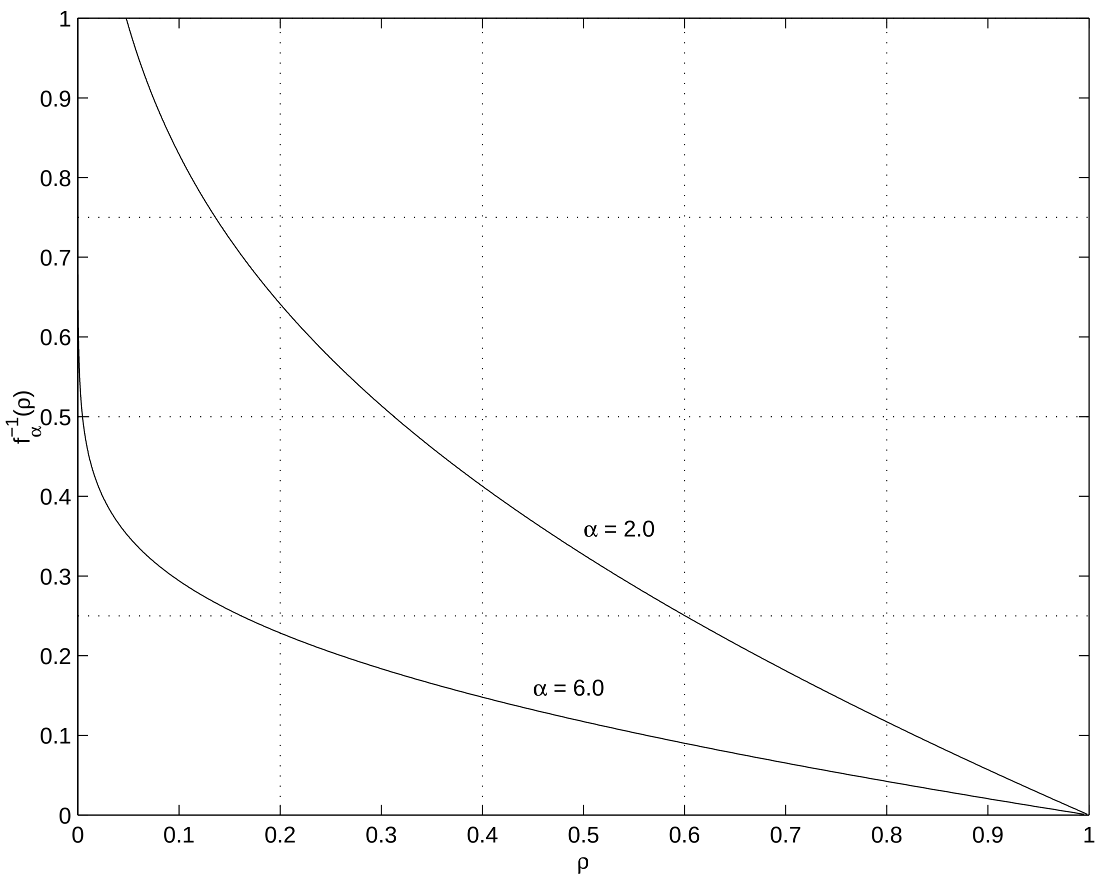

Figure 4: $f_\alpha^{-1}(\rho)$ versus ρ, $\alpha = 2.0$ and 6.0.

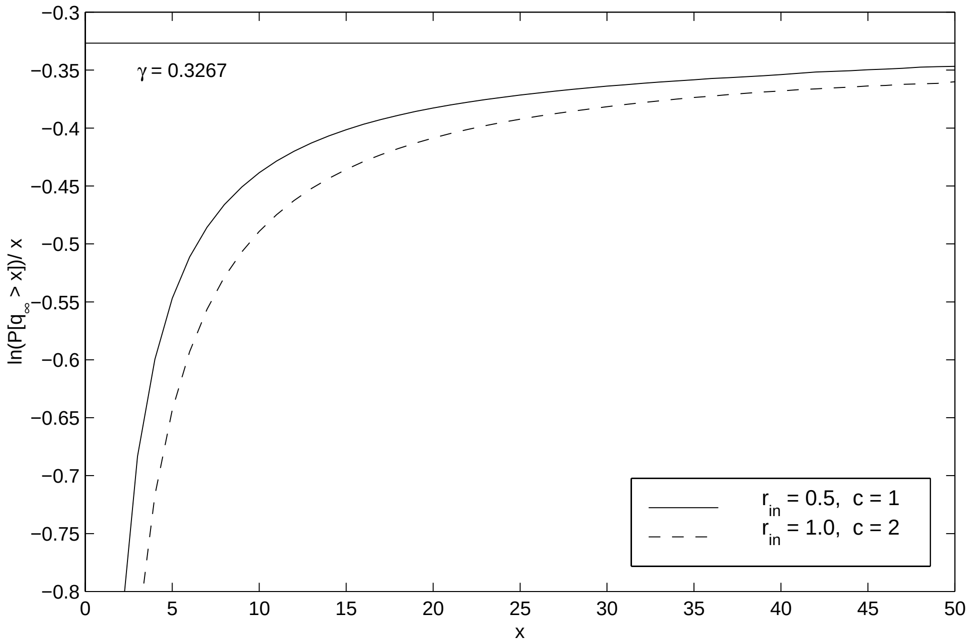

Figure 5: Tail probability vs. buffer size: Rayleigh ($\alpha = 2$)

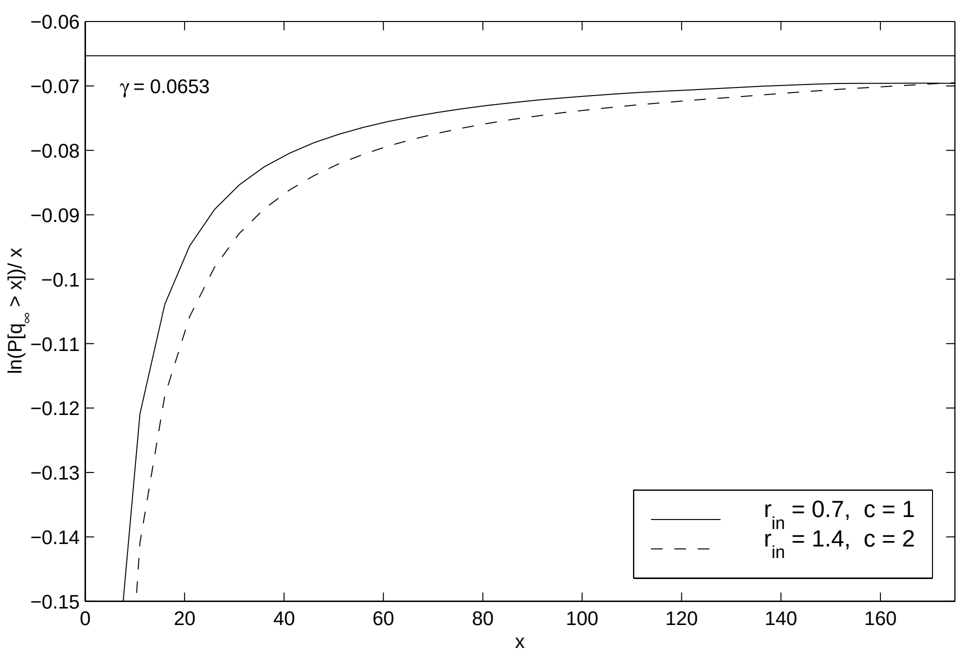

Figure 6: Tail probability vs. buffer size: Rayleigh ($\alpha = 6$)

implying via Lemma 5.3.2 and (4.3.3), that we are in Case II with $R = -\ln q$. Applying Theorem 4.7.2 with the auxiliary scaling $v_n = n,\ n = 1, 2, \ldots$, we have

$$\Lambda(\theta) = \begin{cases} \lambda \cdot \dfrac{e^\theta - 1}{1 - qe^\theta} - c\theta & \text{if } \theta < -\ln q \\ \infty & \text{if } \theta \geq -\ln q. \end{cases} \tag{6.2.5}$$

Noting that **GE 3** is indeed satisfied for this case, we invoke Proposition 5.3.1, and the result follows through the fact that

$$\mathbf{E}\left[e^{\sigma\theta}\right] = \begin{cases} 1 + \dfrac{e^\theta - 1}{1 - qe^\theta} & \text{if } \theta < -\ln q \\ \infty & \text{if } \theta \geq -\ln q. \end{cases} \tag{6.2.6}$$

■

As in the Rayleigh case, an exact solution to (6.2.3) is not available. However the numerical solution is easily achieved as shown in Figure 7 and Table 6.3 for $q = 0.5$ and 0.75. The numerical results plotted in Figure 8 follow the asymptotic behavior outlined by Proposition 6.2.1.

For values of q close to 1, $\ln q \sim q - 1$, and we find that

$$f_G(\theta) = \frac{\theta}{e^\theta - 1} \cdot \frac{1 - qe^\theta}{1 - q}, \qquad 0 < \theta < 1 - q.$$

from (6.2.4). Using the notation $q' \equiv 1 - q$, we have

$$\begin{aligned} f_G(\theta) &= \frac{\theta}{e^\theta - 1} \cdot \frac{1 - (1 - q')e^\theta}{q'} \\ &= \frac{\theta}{q'\left(e^\theta - 1\right)} \cdot \left(q'e^\theta - \left(e^\theta - 1\right)\right) \\ &= \frac{\theta e^\theta}{e^\theta - 1} - \frac{\theta}{q'}, \qquad 0 < \theta < q'. \end{aligned}$$

As $0 < \theta < q'$ and $q' \to 0$, we know that $e^\theta \sim 1 + \theta + \dfrac{\theta^2}{2}$ and

$$f_G(\theta) \sim \frac{\theta(1 + \theta + \frac{\theta^2}{2})}{\theta + \frac{\theta^2}{2}} - \frac{\theta}{q'}$$

ρ	$f_G^{-1}(\rho)$		$2(1-\rho)\frac{1-q}{1+q}$	
	$q = 0.5$	$q = 0.75$	$q = 0.5$	$q = 0.75$
0.5	0.3397	0.1433	0.3333	0.1429
0.7	0.2023	0.0859	0.2000	0.0857
0.9	0.0669	0.0286	0.0667	0.0286

Table 6.3: $\gamma_{\text{Geometric}}$

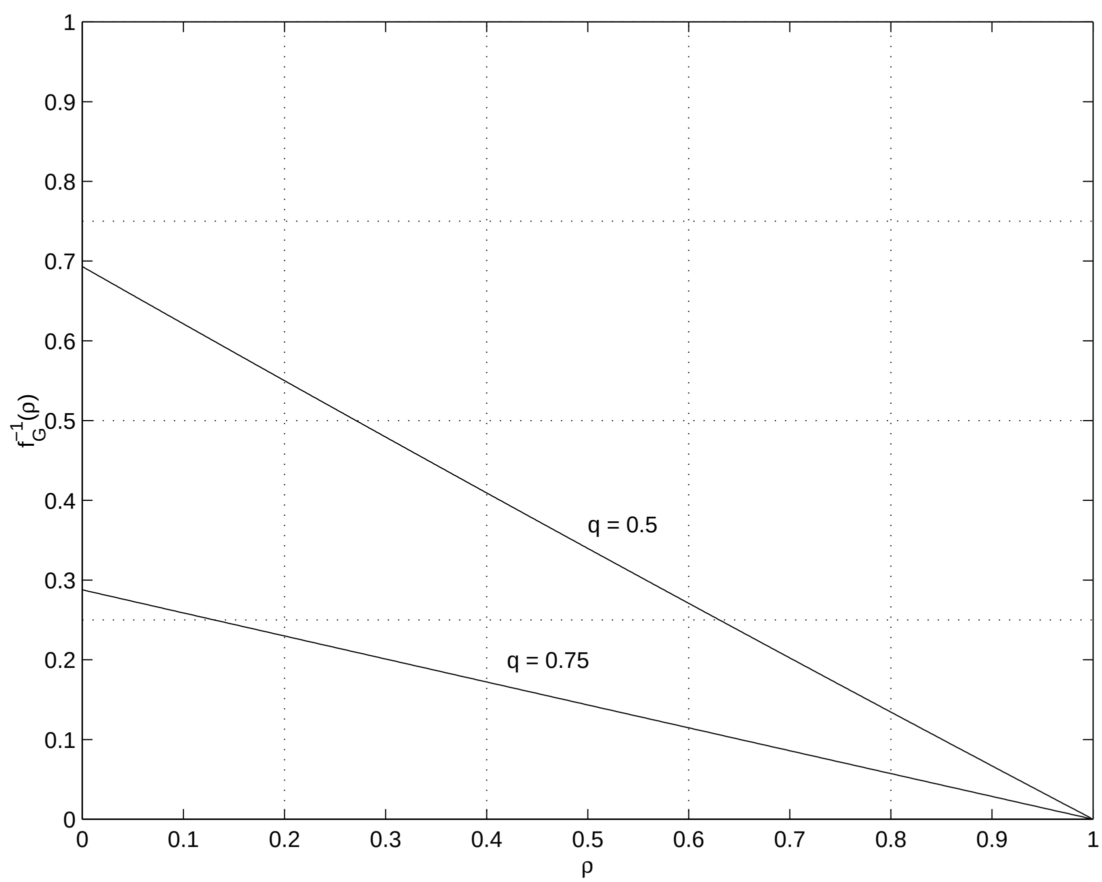

Figure 7: $f_G^{-1}(\rho)$ versus ρ, $q = 0.5$ and 0.75.

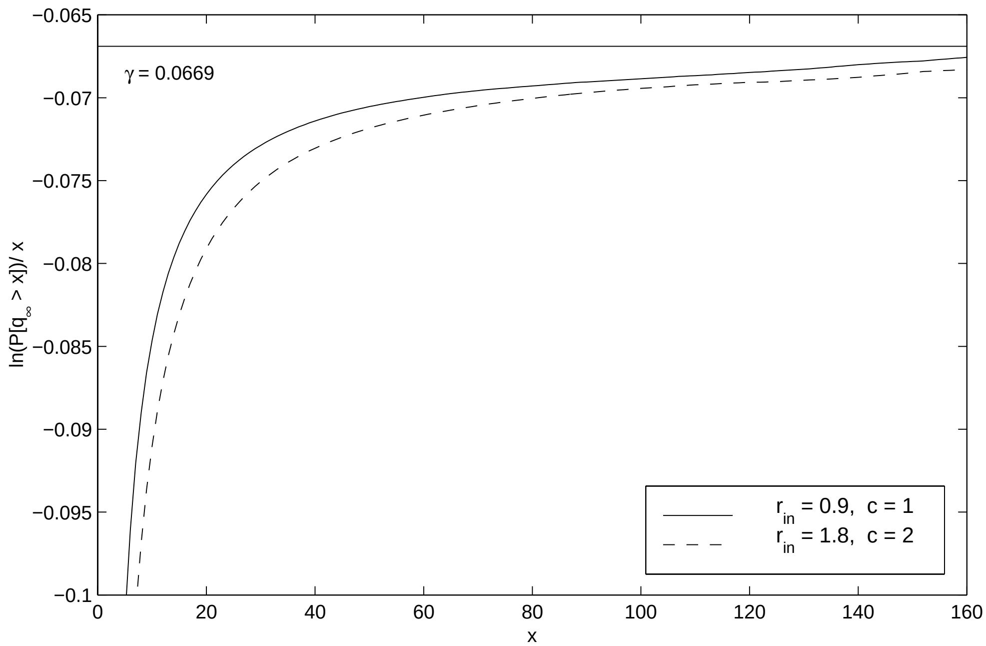

Figure 8: Tail probability vs. buffer size: Geometric ($q = 0.5$)

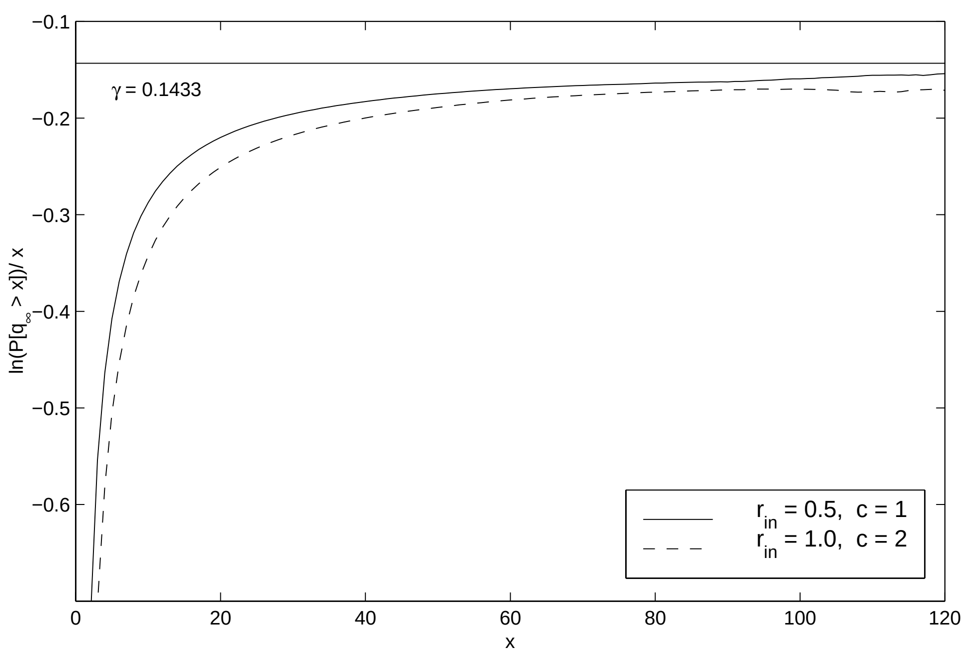

Figure 9: Tail probability vs. buffer size: Geometric ($q = 0.75$)

$$
\begin{aligned}
&= \frac{1+\theta+\frac{\theta^2}{2}}{1+\frac{\theta}{2}} - \frac{\theta}{q'} \\
&= 1+\frac{\theta}{2}\cdot\frac{1+\theta}{1+\frac{\theta}{2}} - \frac{\theta}{q'} \\
&\sim 1+\theta\left(\frac{1}{2}-\frac{1}{q'}\right), \qquad 0<\theta<q',
\end{aligned}
$$

or in terms of the original parameter q,

$$
f_G(\theta) \;\sim\; 1-\theta\left(\frac{1}{1-q}-\frac{1}{2}\right), \qquad 0<\theta<1-q. \tag{6.2.7}
$$

By (6.2.7) and (6.2.3) we now have

$$
f_G^{-1}(\rho) \;\sim\; 2(1-\rho)\cdot\frac{1-q}{1+q}, \qquad 0<\rho<1. \tag{6.2.8}
$$

This linear relationship between $f_G^{-1}(\rho)$ and ρ is clearly reflected in the plots displayed in Figure 7.

6.2.2 The Gamma case

A continuous rv X is said to be a Gamma rv with parameters $\mu>0$ and $a>0$, if

$$
\mathbf{P}\left[X\le x\right] = 1-\frac{\Gamma(\mu,ax)}{\Gamma(\mu)}, \qquad x\ge 0 \tag{6.2.9}
$$

where

$$
\Gamma(\eta,x) \equiv \int_x^\infty e^{-t}t^{\eta-1}dt, \qquad \eta\ge 0,\; x>0 \tag{6.2.10}
$$

is the incomplete Γ–function, and $\Gamma(\eta)\equiv\Gamma(\eta,0)$. The pmf $G=\{g_r,\; r=1,2,\ldots\}$ of the rv $\sigma =_{st} \lceil X\rceil$ is then given by

$$
g_r = \frac{1}{\Gamma(\mu)}\left(\Gamma(\mu,a(r-1))-\Gamma(\mu,ar)\right), \qquad r=1,2,\ldots \tag{6.2.11}
$$

and is said to be a discrete Gamma distribution with parameters $\mu>0$ and $a>0$.

Proposition 6.2.2 *If G is a discrete Gamma distribution with parameters $\mu > 0$ and $a > 0$, then*

$$\lim_{b\to\infty} \frac{1}{b} \ln \mathbf{P}\left[q_\infty > b\right] = -\gamma^\star_{\text{Gamma}} \tag{6.2.12}$$

where

$$\gamma^\star_{\text{Gamma}} \quad = \quad f^{-1}_{(\mu,a)}(\rho), \tag{6.2.13}$$

and

$$f_{(\mu,a)}(\theta) \quad = \quad \left(\frac{e^\theta - 1}{\theta} \cdot \frac{1}{\mathbf{E}\left[\sigma\right]} \cdot \frac{1}{\Gamma(\mu)} \sum_{r=0}^{\infty} e^{\theta r}\Gamma(\mu, ar)\right)^{-1}, \qquad 0 < \theta < a. \tag{6.2.14}$$

Proof. From (6.0.1) and (6.2.9) we remark that

$$w^\star_n = \ln\Gamma(\mu) - \ln\Gamma(\mu, an), \qquad n \geq 0. \tag{6.2.15}$$

Proceeding with the well-known asymptotics [1]

$$\Gamma(\eta, x) \sim e^{-x} x^{\eta - 1}(1 + o(x)), \qquad \eta > 0 \ (x \to \infty) \tag{6.2.16}$$

we find that

$$\begin{aligned} \lim_{n\to\infty} \frac{w^\star_n}{n} &= -\lim_{n\to\infty} \frac{\ln\Gamma(\mu, an)}{n} \\ &= \lim_{n\to\infty} \frac{an - (\mu - 1)\ln n}{n} \\ &= a. \end{aligned} \tag{6.2.17}$$

Therefore $R = a$ by Lemma 5.3.2 and (4.3.3), and Theorem 4.7.2 applies with the auxiliary scaling $v_n = n, \ n = 1, 2, \ldots$ yielding

$$\Lambda(\theta) = \begin{cases} -c\theta + \lambda\left(e^\theta - 1\right) \sum_{r=0}^{\infty} e^{\theta r} \frac{\Gamma(\mu, ar)}{\Gamma(\mu)} & \text{if } \theta < a \\ \infty & \text{if } \theta > a. \end{cases} \tag{6.2.18}$$

Assumption **GE 3** being satisfied, the result now follows through Proposition 5.3.1 upon noting that

$$\mathbf{E}\left[e^{\sigma\theta}\right] = \begin{cases} \left(e^{\theta}-1\right)\sum_{r=0}^{\infty} e^{\theta r}\frac{\Gamma(\mu, ar)}{\Gamma(\mu)} & \text{if } \theta < a \\ \infty & \text{if } \theta > a. \end{cases} \tag{6.2.19}$$

■

For the special case when $\mu = 1$ we have

$$\Gamma(\mu, ax) = e^{-ax}, \qquad x > 0,$$

and the pmf G defined by (6.2.11) is now given by

$$g_r = e^{-a(r-1)}(1-e^{-a}), \qquad r = 1, 2, \ldots. \tag{6.2.20}$$

A quick glance at (6.2.1) indicates that G in this case is identical to the geometric distribution with parameter $q = e^{-a}$. As expected, both Propositions 6.2.2 and 6.2.1 yield the same asymptotic results.

6.3 Sub–exponential distributions

6.3.1 The Weibull case

A rv X is said to be a Weibull rv with parameters a and β ($a > 0$ and $0 < \beta < 1$) if

$$\mathbf{P}\left[X \leq x\right] = 1 - e^{-ax^{\beta}}, \qquad x \geq 0. \tag{6.3.1}$$

The pmf $G = \{g_r,\ r = 1, 2, \ldots\}$ of the rv σ is said to be an (integer–valued) Weibull distribution with parameters a and β if $\sigma =_{st} \lceil X \rceil$, in which case we have

$$g_r = e^{-a(r-1)^{\beta}} - e^{-ar^{\beta}}, \qquad r = 1, 2, \ldots \tag{6.3.2}$$

Proposition 6.3.1 *If G is a discrete Weibull distribution with parameters a and β ($a > 0$ and $0 < \beta < 1$), then*

$$\lim_{b\to\infty} \frac{1}{ab^\beta} \ln \mathbf{P}\left[q_\infty > b\right] \;\; \leq \;\; -\gamma^\star_{\text{Weibull}} \tag{6.3.3}$$

where

$$\gamma^\star_{\text{Weibull}} = e^{(1-\beta)\ln(c-r_{in})+H(\beta)}, \tag{6.3.4}$$

and $H(\beta) = -\beta \ln \beta - (1-\beta)\ln(1-\beta)$ denotes the natural entropy of the pmf $(\beta, 1-\beta)$.

Proof. From (6.0.1) and (6.3.1) we have

$$w^\star_n \;\; = \;\; an^\beta, \qquad n = 1, 2, \ldots.$$

Hence by Lemma 5.3.2 we conclude that $R = 0$ and select $w^\star$ as the auxiliary scaling.

The fact that the condition **C1** holds under scaling $w^\star$ is easily verified. Selecting $Z(n) = n^\alpha$ with $1-\beta < \alpha < 1$, we see that the condition **C2b** is also satisfied. Therefore Theorem 4.7.3 applies, yielding (5.3.16).

We note that the function $w^\star$ satisfies (3.2.7), hence it is regularly varying. By Lemma 3.2.2, we may then select the function h to be the piecewise–continuous interpolation of $w^\star$, i.e. $h(b) = ab^\beta,\ b > 0$, and consequently, $g(y) = y^{-\beta},\ y > 0$.

Finally, noting that the assumption **A2** holds with $\kappa = W = 0$, we invoke Proposition 3.4.3, which yields (6.3.3) with

$$\gamma^\star_{\text{Weibull}} \;\; = \;\; \sup_{y>0} \gamma^\star(y), \tag{6.3.5}$$

where

$$\begin{aligned}
\gamma^\star(y) &= \min\left(\sup_{\theta>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{w_n^\star}{h(xn)}(\theta x - \Lambda(\theta))\right),\ \Lambda^\star(0)g(y)\right) \\
&= \min\left(\sup_{0<\theta<1}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{n^\beta}{(xn)^\beta}(\theta(x+c-r_{in}))\right),\ (c-r_{in})y^{-\beta}\right) \\
&= \min\left(\sup_{0<\theta<1}\ \theta\ \inf_{x>y}\left(\frac{x+c-r_{in}}{x^\beta}\right),\ (c-r_{in})y^{-\beta}\right) \\
&= \min\left(\inf_{x>y}\left(\frac{x+c-r_{in}}{x^\beta}\right),\ (c-r_{in})y^{-\beta}\right). \qquad (6.3.6)
\end{aligned}$$

Define the mapping $T : (0,\infty) \to \mathbb{R}_+$ by

$$T(x) \equiv \frac{x+c-r_{in}}{x^\beta}, \qquad x > 0. \qquad (6.3.7)$$

Differentiating with respect to x gives

$$\begin{aligned}
\frac{dT}{dx}(x) &= \frac{x(1-\beta)-\beta(c-r_{in})}{x^{1+\beta}} \\
&= (1-\beta)\,\frac{x-x^\star}{x^{1+\beta}}
\end{aligned}$$

with

$$x^\star \equiv (c-r_{in})\,\frac{\beta}{1-\beta}. \qquad (6.3.8)$$

As

$$\frac{dT}{dx}(x)\ \begin{cases} > 0 & \text{if } x > x^\star \\ = 0 & \text{if } x = x^\star \\ < 0 & \text{if } x < x^\star, \end{cases}$$

we conclude that

$$\inf_{x>y} T(x) = \begin{cases} T(y) & \text{if } y > x^\star \\ T(x^\star) & \text{if } y \le x^\star. \end{cases}$$

We may then rewrite (6.3.6) as

$$
\begin{aligned}
\gamma^\star(y) &= \min\left(\inf_{x>y} T(x),\ (c-r_{in})y^{-\beta}\right) \\
&= \begin{cases} \min\left((y+c-r_{in})y^{-\beta},\ (c-r_{in})y^{-\beta}\right) & \text{if } \ y > x^\star \\ \min\left((x^\star+c-r_{in})(x^\star)^{-\beta},\ (c-r_{in})y^{-\beta}\right) & \text{if } \ y \le x^\star \end{cases} \\
&= \begin{cases} (c-r_{in})y^{-\beta} & \text{if } \ y > x^\star \\ (c-r_{in})\min\left(\dfrac{(x^\star)^{-\beta}}{1-\beta},\ y^{-\beta}\right) & \text{if } \ y \le x^\star \end{cases}
\end{aligned}
$$

where the final step follows via (6.3.8). Reporting this to (6.3.5) we have

$$
\begin{aligned}
\gamma^\star_{\text{Weibull}} &= (c-r_{in})\max\left(\sup_{y>x^\star} y^{-\beta},\ \sup_{0<y\le x^\star}\min\left(\frac{(x^\star)^{-\beta}}{1-\beta},\ y^{-\beta}\right)\right) \\
&= (c-r_{in})\max\left((x^\star)^{-\beta}, \min\left(\frac{(x^\star)^{-\beta}}{1-\beta},\ \sup_{0<y\le x^\star} y^{-\beta}\right)\right) \\
&= (c-r_{in})\max\left((x^\star)^{-\beta}, \frac{(x^\star)^{-\beta}}{1-\beta}\right) \\
&= (c-r_{in})\frac{(x^\star)^{-\beta}}{1-\beta},
\end{aligned}
$$

and the expression (6.3.4) follows directly through (6.3.8). ■

The alternate upper bound achieved through Proposition 5.4.3 and given by

$$\limsup_{b\to\infty} \frac{1}{ab^\beta} \ln \mathbf{P}\left[q^b_\infty > b\right] \le -1, \tag{6.3.9}$$

improves upon (6.3.3) if $\gamma^\star_{\text{Weibull}} < 1$, i.e., if $(1-\beta)\ln(c-r_{in}) + H(\beta) < 0$. This only occurs in the instance when $c - r_{in} < 1$, and even then, not for all values of β in the interval $(0, 1)$.

Applying Proposition 5.4.4, we arrive at the lower bound

$$-\gamma^{\text{W}}_\star \le \liminf_{b\to\infty} \frac{1}{ab^\beta} \ln \mathbf{P}\left[q^b_\infty > b\right], \tag{6.3.10}$$

with

$$\gamma_\star^{\mathrm{W}} = \inf_{y>0} \left((1 + \lfloor c - r_{in} + y \rfloor) y^{-\beta} \right). \tag{6.3.11}$$

Explicit expressions for $\gamma_\star^{\mathrm{W}}$ are given in [45, Section 3].

Tables 6.4 and 6.5 list values taken by the upper and lower bounds for output rates $c = 1, 2$. The upper bounds (6.3.3) and (6.3.9) are easily compared; the one showing to advantage is highlighted. Though the bounds do not coincide for all values of r_{in}, they are reasonably close, in many of the cases shown.

Figure 10 features the simulation results for selected parameters $a = 1$ and $\beta = 0.25$, with $c = 1$ and $r_{in} = 0.9$. The associated upper and lower bounds, available in Table 6.4, are depicted in the graph. The continuous curve, representing the quantity of interest, i.e., $\ln \mathbf{P}\left[q_\infty > x\right]/ax^\beta$, is clearly outside the predicted bounds in the plotted regime, though it does hold out the promise of eventually satisfying (6.3.9) and (6.3.10).

One possible explanation for this apparent incongruence could be that terms of the order $o(x^\beta)$, neglected so far, provide a significant contribution to the asymptotics *in the plotted range*. This argument is validated by the second (dashed) curve, which converges much faster to the predicted bounds. The second curve represents the log–tail buffer probability, now scaled by the function $v^\star$, which accounts for smaller order terms, and in the Weibull case takes the form

$$v_n^\star \quad \sim \quad an^\beta - (1-\beta)\ln n, \quad (n \to \infty). \tag{6.3.12}$$

This relation follows directly from (2.3.1); its proof is not included.

Similar conclusions may be drawn from the simulation plots depicted in Figures 11 and 12. In Figure 13, we compare the buffer asymptotics for a fixed distribution G, under identical utilization factors $\rho = r_{in}/c$, but differing values of r_{in} and c.

r_{in}	$\beta = 0.25$			$\beta = 0.5$		
	$\gamma^{\star}_{\text{Weibull}}$	$\gamma^{\text{W}}_{\text{U}}$	$\gamma^{\text{W}}_{\text{L}}$	$\gamma^{\star}_{\text{Weibull}}$	$\gamma^{\text{W}}_{\text{U}}$	$\gamma^{\text{W}}_{\text{L}}$
0.5	**1.04**	1.00	**1.19**	**1.41**	1.00	**1.41**
0.7	0.71	**1.00**	**1.09**	**1.10**	1.00	**1.20**
0.9	0.31	**1.00**	**1.03**	0.63	**1.00**	**1.06**

Table 6.4: $\gamma_{\text{Weibull}}, c = 1$

r_{in}	$\beta = 0.25$			$\beta = 0.5$		
	$\gamma^{\star}_{\text{Weibull}}$	$\gamma^{\text{W}}_{\text{U}}$	$\gamma^{\text{W}}_{\text{L}}$	$\gamma^{\star}_{\text{Weibull}}$	$\gamma^{\text{W}}_{\text{U}}$	$\gamma^{\text{W}}_{\text{L}}$
0.8	**2.01**	1.00	**2.11**	**1.10**	1.00	**2.24**
1.0	**1.75**	1.00	**2.00**	**2.00**	2.00	**2.00**
1.4	**1.20**	1.00	**1.26**	**1.54**	1.00	**1.58**
1.8	0.52	**1.00**	**1.50**	0.90	**1.00**	**1.82**

Table 6.5: $\gamma_{\text{Weibull}}, c = 2$

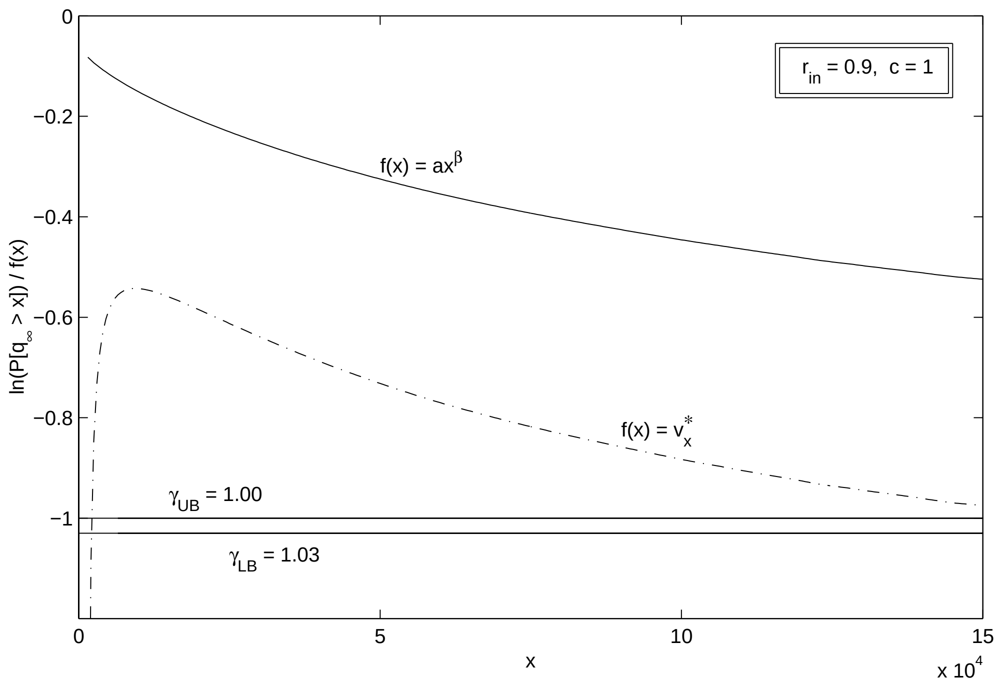

Figure 10: Tail probability vs. buffer size: Weibull ($a = 1.0, \beta = 0.25$)

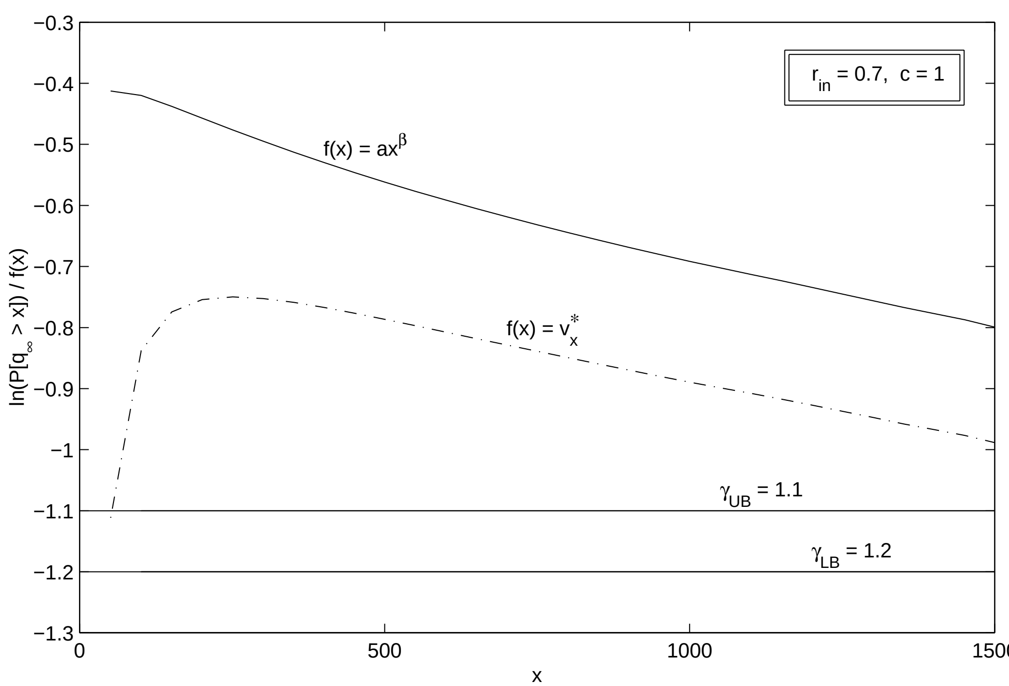

Figure 11: Tail probability vs. buffer size: Weibull ($a = 0.5, \beta = 0.5$)

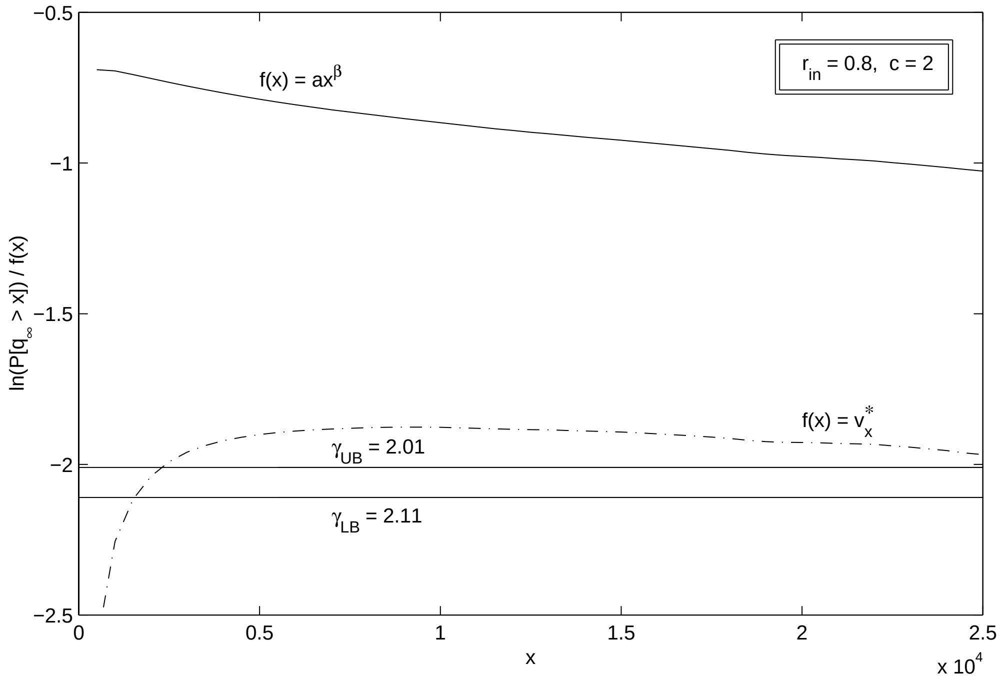

Figure 12: Tail probability vs. buffer size: Weibull ($a = 1.0, \beta = 0.25$)

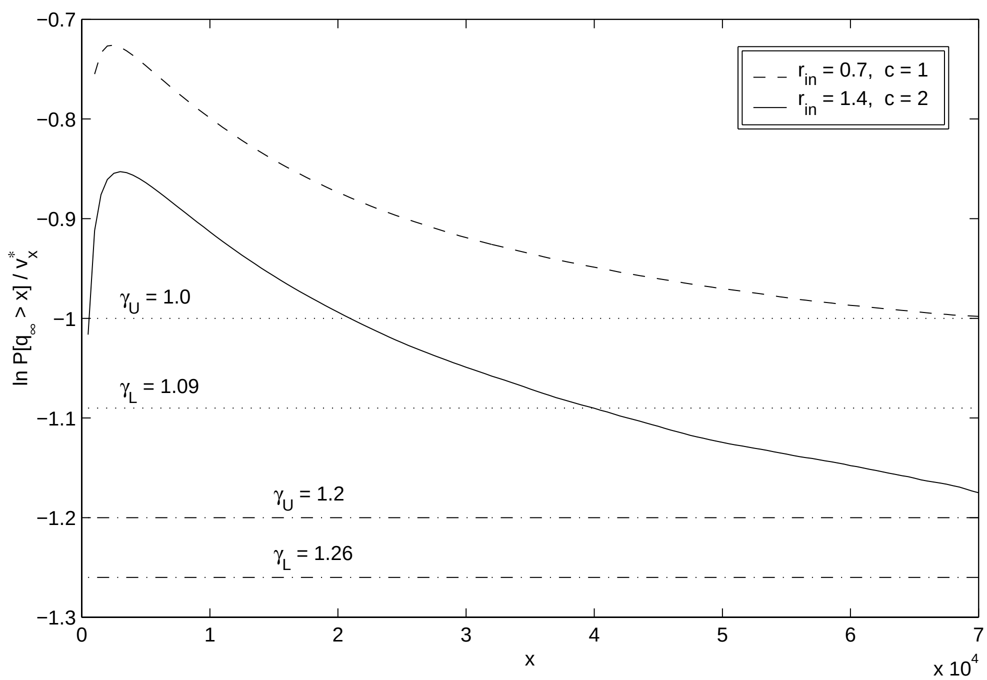

Figure 13: Tail probability vs. buffer size: Weibull ($a = 1.0, \beta = 0.25$)

For the example illustrated, it is apparent that the two curves converge to different limits. This is a distinct departure from the earlier exponential plots, where, under a fixed distribution G, the dependence of the limiting value $\gamma^\star$ on the input and output rates c and r_{in}, is strictly through their ratio ρ.

6.3.2 The log–normal case

A rv X is said to be a log–normal rv if $X =_{st} \exp(Y)$ where Y is a Gaussian rv with mean μ and variance δ^2. The pmf $G = \{g_r,\ r = 1, 2, \ldots\}$ of the rv σ is said to be an (integer–valued) log–normal distribution with parameters μ and δ if $\sigma =_{st} \lceil X \rceil$. It is easy to check that

$$\begin{aligned} g_r &= \mathbf{P}\left[r-1 < X \leq r\right] \\ &= \Phi\left(\frac{1}{\delta}\ln\left(\frac{r}{m}\right)\right) - \Phi\left(\frac{1}{\delta}\ln\left(\frac{r-1}{m}\right)\right), \qquad r = 1, 2, \ldots \end{aligned} \tag{6.3.13}$$

where $m \equiv e^\mu$, and Φ is the cumulative distribution function of a Gaussian rv with zero mean and unit variance.

Proposition 6.3.2 *If G is a discrete log–normal distribution with parameters μ and δ as described above, then*

$$\liminf_{b\to\infty} \frac{2\delta^2}{(\ln b)^2} \ln \mathbf{P}\left[q_\infty > b\right] \leq -\gamma^\star_{\text{Lognormal}} \tag{6.3.14}$$

where

$$\gamma^\star_{\text{Lognormal}} = c - r_{in}. \tag{6.3.15}$$

Proof. Fix $n = 1, 2, \ldots$. We begin by noting from (6.3.13) that

$$\mathbf{P}\left[\sigma > n\right] = 1 - \Phi\left(\frac{1}{\delta}\ln\left(\frac{n}{m}\right)\right).$$

Hence, by (5.3.1),

$$w_n^\star = -\ln\left(1 - \Phi\left(\frac{\ln\left(\frac{n}{m}\right)}{\delta}\right)\right). \tag{6.3.16}$$

Using the well known asymptotics [1]

$$1 - \Phi(n) \sim \frac{1}{\sqrt{2\pi}} \cdot \frac{e^{-\frac{n^2}{2}}}{n}, \qquad (n \to \infty) \tag{6.3.17}$$

we find that

$$\begin{aligned} \lim_{n\to\infty} \frac{w_n^\star}{(\ln n)^2} &= \lim_{n\to\infty} \frac{-\ln\left(1 - \Phi\left(\frac{\ln\left(\frac{n}{m}\right)}{\delta}\right)\right)}{(\ln n)^2} \\ &= \lim_{n\to\infty} \frac{\left(\frac{\ln\left(\frac{n}{m}\right)}{\delta}\right)^2}{2(\ln n)^2} \\ &= \frac{1}{2\delta^2} \lim_{n\to\infty} \frac{(\ln n - \ln m)^2}{(\ln n)^2} \\ &= \frac{1}{2\delta^2}. \end{aligned} \tag{6.3.18}$$

Hence $R = W = 0$ via Lemma 5.3.2, and definitions (4.3.3) and (5.3.2).

As argued earlier in Section 5.3.3, we may use the auxiliary scaling $w_n = \dfrac{(\ln n)^2}{2\delta^2}$. Clearly, condition **C1** holds under this choice of scaling, as does **C2**, for the selection $Z(n) = \dfrac{n}{\ln n}$. Therefore Theorem 4.7.3 applies, yielding (5.3.16).

The function w_n being regularly varying according to the definition (3.2.7), we have a ready candidate for the function h in the piecewise–continuous interpolation of w_n, as proved in Lemma 3.2.2. In other words, we select $h(b) = \dfrac{(\ln b)^2}{2\delta^2},\ b > 0$, and $g(y) = 1,\ y > 0$.

Finally, verifying that Assumption **A2** is also satisfied (with $\kappa = W = 0$), we apply Proposition 3.4.3 which yields (6.3.14) with

$$\gamma^\star_{\text{Lognormal}} = \sup_{y>0} \gamma^\star(y), \tag{6.3.19}$$

where $\gamma^\star(y),\ y > 0$, is given by (3.4.30).

Referring to **Case 1** in Section 3.5, we note that $\sup_{y>0} \gamma^\star(y)$ has already been computed for the special case $g(y) = 1,\ y > 0$, and is given by (3.5.38), thus concluding the proof of Proposition 6.3.2. ■

Proposition 5.4.3 provides the alternate asymptotic upper bound

$$\limsup_{b\to\infty} \frac{2\delta^2}{(\ln b)^2} \ln \mathbf{P}\left[q_\infty^b > b\right] \le -1, \tag{6.3.20}$$

which proves tighter than (6.3.14) only if $c - r_{in} < 1$.

The corresponding lower bound, derived via Proposition 5.4.4, is given by

$$-\lfloor c - r_{in} + 1 \rfloor \le \liminf_{b\to\infty} \frac{2\delta^2}{(\ln b)^2} \ln \mathbf{P}\left[q_\infty^b > b\right]. \tag{6.3.21}$$

A striking difference from the cases observed so far, is the fact that the limits $\gamma^\star_{\text{Lognormal}}$ and $\gamma_\star^L$ show no dependence on the parameters (μ, δ) characterizing the lognormal distribution G.

In the case $c - r_{in} < 1$, we observe that the bounds given by (6.3.20) and (6.3.21) coincide, yielding

$$\lim_{b\to\infty} \frac{1}{(\ln b)^2} \ln \mathbf{P}\left[q_\infty^b > b\right] \quad = \quad -\frac{1}{2\delta^2}. \tag{6.3.22}$$

The limit (6.3.22) is always true when $c = 1$; the calculated bounds for the case $c = 2$ are listed in Table 6.6.

Figures 14, 15 and 16 present the tail buffer asymptotics when G is lognormal with parameters 1.414 and 1.732, for $c = 1, 2$ and varying values of r_{in}.

Learning from our experience in the Weibull case, we take care to plot the log–tail probability $\ln \mathbf{P}\left[q_\infty > x\right]$ using the exact scaling $v^\star$, instead of its asymptotic

r_{in}	$\gamma^{\star}_{\text{Lognormal}}$	$\gamma^{\text{L}}_{\text{U}}$	$\gamma^{\text{L}}_{\text{L}}$
0.8	**1.2**	1.0	**2.0**
1.0	**1.0**	**1.0**	**1.0**
1.4	0.6	**1.0**	**1.0**
1.8	0.2	**1.0**	**1.0**

Table 6.6: $\gamma_{\text{Lognormal}}, c = 2$

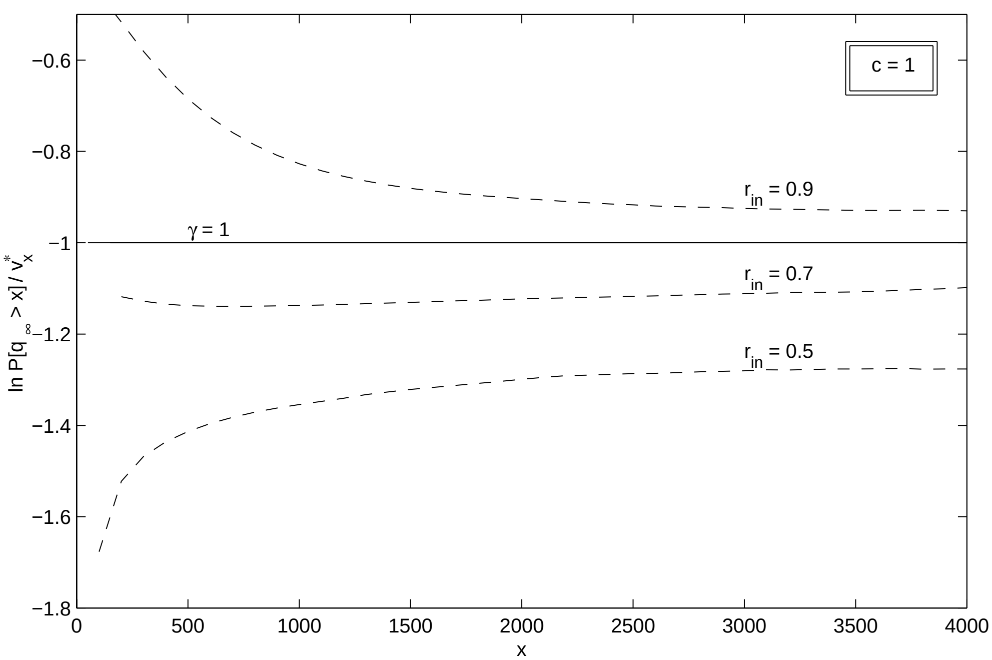

Figure 14: Tail probability vs. buffer size: Lognormal ($\delta = 1.414$)

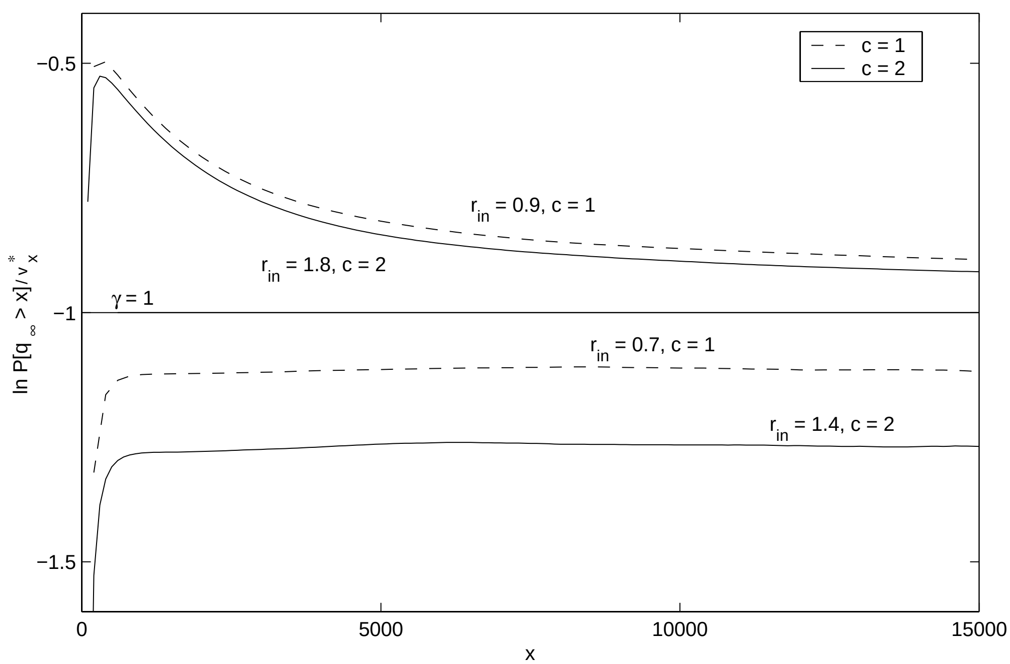

Figure 15: Tail probability vs. buffer size: Lognormal ($\delta = 1.732$)

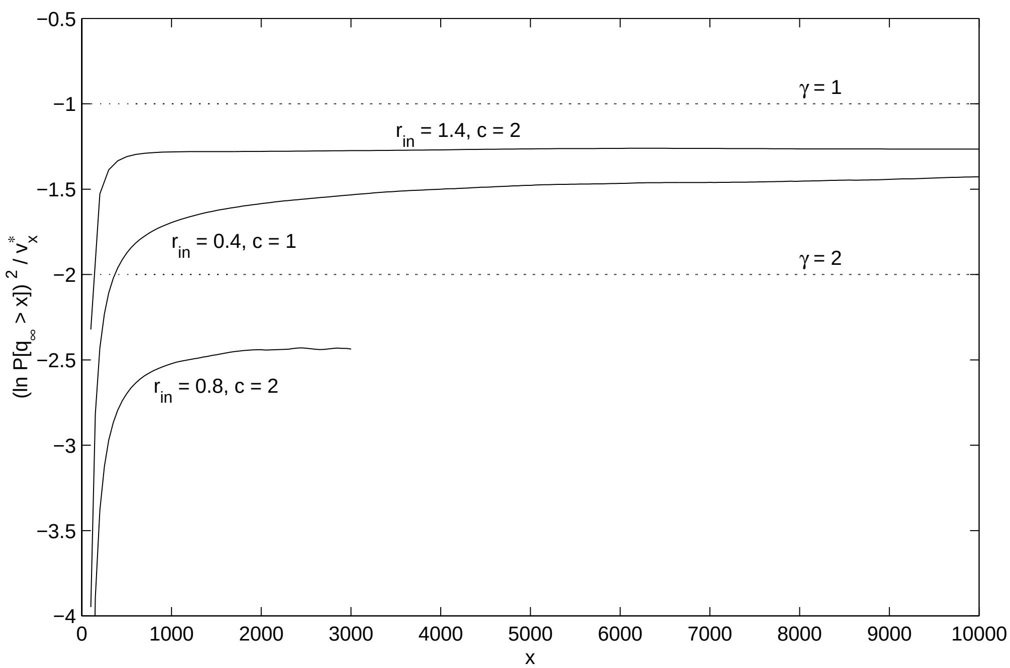

Figure 16: Tail probability vs. buffer size: Lognormal ($\delta = 1.732$)

equivalent w, where

$$v_n^\star \;\sim\; \frac{1}{2}\left(\frac{\ln n}{\delta}\right)^2 - \ln n + \ln(\ln n) + \ln(\ln n - \delta^2) \quad (n \to \infty). \quad (6.3.23)$$

The convergence of the plots to their predicted limiting values however remains slow. This could very possibly be due to neglected terms of still smaller order, and perhaps even constants, which gain importance when the dominating scaling (in this case $(\ln n)^2$) increases to infinity at a very slow rate.

6.3.3 The Pareto case

A rv X is said to be Pareto with parameters $A, \alpha > 0$,

$$\mathbf{P}[X \leq x] = 1 - \left(\frac{x}{A}\right)^{-\alpha}, \qquad x \geq A. \qquad (6.3.24)$$

The pmf $G = \{g_r,\ r = 1, 2, \ldots\}$ of the rv σ is said to be an (integer–valued) Pareto distribution with parameter $\alpha > 0$ if $\sigma =_{st} \left\lfloor \frac{X}{A} \right\rfloor$, in which case we have

$$\begin{aligned} \mathbf{P}[\sigma > r] &= \mathbf{P}\left[\left\lfloor \frac{X}{A} \right\rfloor > r\right] \\ &= \mathbf{P}[X \geq A(r+1)] \\ &= (r+1)^{-\alpha} \qquad r = 1, 2, \ldots, \end{aligned} \qquad (6.3.25)$$

and

$$g_r = \mathbf{P}[\sigma = r] = r^{-\alpha} - (r+1)^{-\alpha}, \qquad r = 1, 2, \ldots. \qquad (6.3.26)$$

Having defined the distribution, we note that the requirement $\mathbf{E}[\sigma] < \infty$ is equivalent to the constraint $\alpha > 1$.

Proposition 6.3.3 *If G is a discrete Pareto distribution with parameter α, ($\alpha > 1$), then*

$$\liminf_{b \to \infty} \frac{1}{(\alpha - 1)\ln b} \ln \mathbf{P}[q_\infty > b] \leq -\gamma^\star_{\text{Pareto}} \qquad (6.3.27)$$

where

$$\gamma^\star_{\text{Pareto}} = (c - r_{in}) - \frac{1}{\alpha - 1}. \tag{6.3.28}$$

Proof. Fix $n = 1, 2, \ldots$, and note from (6.3.25) that

$$w_n^\star \;=\; \alpha \ln n. \tag{6.3.29}$$

Therefore, by Lemma 5.3.2, and definitions (4.3.3) and (5.3.2), we conclude that $R = 0$, and $W = \alpha^{-1}$. We mention briefly that our constraint $\alpha > 1$ is seconded by Lemma 5.3.1, which requires that $W < 1$.

Referring to Lemma 5.3.2 gives

$$\lim_{n\to\infty} \frac{v_n^\star}{w_n^\star} \;=\; 1 - \alpha^{-1},$$

leading to the choice of $w_n = (1 - W)w_n^\star = (\alpha - 1)\ln n$ as the auxiliary scaling.

Noting that condition **C1** is trivially satisfied, while **C2** holds for the selection $Z(n) = \dfrac{n}{(\ln n)^{1/2}}$, we derive (5.3.16) via Theorem 4.7.3.

A quick glance at (3.2.7) verifies that w_n is indeed regularly varying, and invoking Lemma 3.2.2 as before, we choose $h(b) = (\alpha - 1)\ln b,\ b > 0$, and $g(y) = 1,\ y > 0$.

As Assumption **A2** is satisfied (with $\kappa = (\alpha - 1)^{-1}$), Proposition 3.4.3 applies, yielding (6.3.27) with

$$\gamma^\star_{\text{Pareto}} \;=\; \sup_{y>0} \gamma^\star(y),$$

and (6.3.28) follows by (3.5.38). ■

Of course, the bound derived in Proposition 6.3.3 is entirely superfluous if $c - r_{in} < (\alpha - 1)^{-1}$, as had been remarked upon earlier in Section 5.3.3.

The alternate upper bound provided by Proposition 5.4.3 and given by

$$\limsup_{b\to\infty} \frac{1}{(\alpha-1)(\ln b)} \ln \mathbf{P}\left[q_\infty^b > b\right] \le -1, \tag{6.3.30}$$

fares better than (6.3.27) only if $(\alpha - 1)(c - r_{in}) < \alpha$.

By Proposition 5.4.4 we have the lower bound

$$\gamma_\star^P \le \liminf_{b\to\infty} \frac{1}{(\alpha-1)(\ln b)} \ln \mathbf{P}\left[q_\infty^b > b\right], \tag{6.3.31}$$

with

$$\gamma_\star^P = \inf_{y>0} \lfloor 1 + c - r_{in} + y \rfloor = \lfloor c - r_{in} + 1 \rfloor. \tag{6.3.32}$$

We note that the bounds provided by (6.3.30) and (6.3.31) are tight under the condition $c - r_{in} < 1$, in which case

$$\lim_{b\to\infty} \frac{1}{(\alpha-1)(\ln b)} \ln \mathbf{P}\left[q_\infty^b > b\right] = -1. \tag{6.3.33}$$

Figures 17–19 present the simulated tail buffer probabilities for G Pareto with $\alpha = 1.5$ and 2.5 under the scaling

$$v_n^\star \sim (\alpha - 1)\ln n, \quad (n \to \infty). \tag{6.3.34}$$

As in the Lognormal case (and probably for similar reasons), the simulation plots do not provide conclusive evidence either verifying or denying the derived bounds.

6.4 Discussion

The difference in buffer asymptotics for heavy–tailed, sub–exponential distributions versus their lighter–tailed, exponential counterparts extends beyond their

r_{in}	$\alpha = 1.5$			$\alpha = 2.5$		
	$\gamma^{\star}_{\text{Pareto}}$	$\gamma^{\text{P}}_{\text{U}}$	$\gamma^{\text{P}}_{\text{L}}$	$\gamma^{\star}_{\text{Pareto}}$	$\gamma^{\text{P}}_{\text{U}}$	$\gamma^{\text{P}}_{\text{L}}$
0.8	–	**1.0**	**2.0**	0.53	**1.0**	**2.0**
1.0	–	**1.0**	**2.0**	0.33	**1.0**	**2.0**
1.4	–	**1.0**	**1.0**	–	**1.0**	**1.0**
1.8	–	**1.0**	**1.0**	–	**1.0**	**1.0**

Table 6.7: $\gamma_{\text{Pareto}}, c = 2$

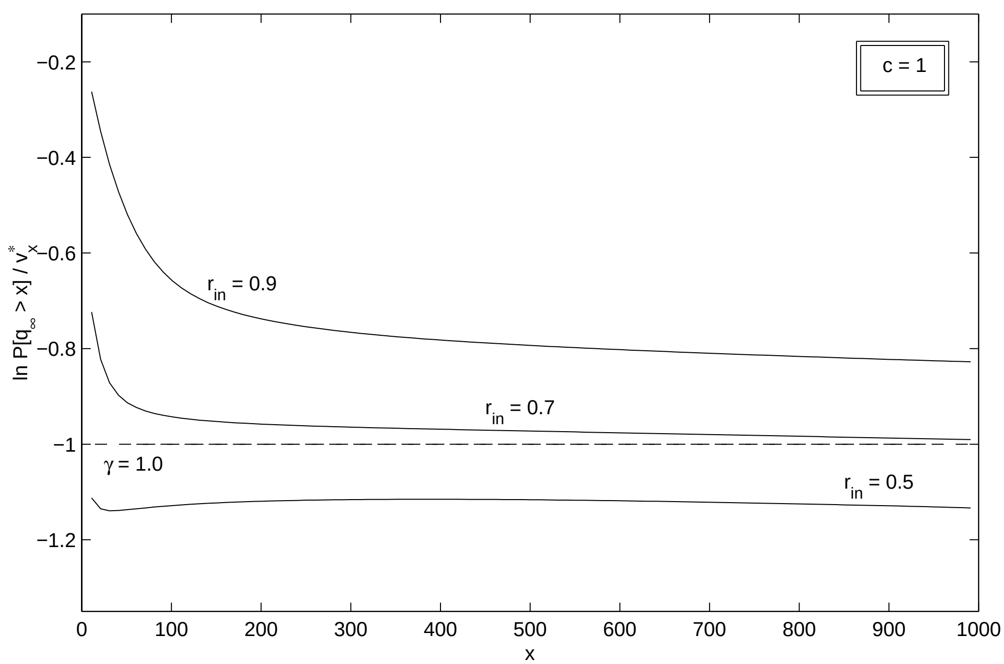

Figure 17: Tail probability vs. buffer size: Pareto ($\alpha = 2.5$)

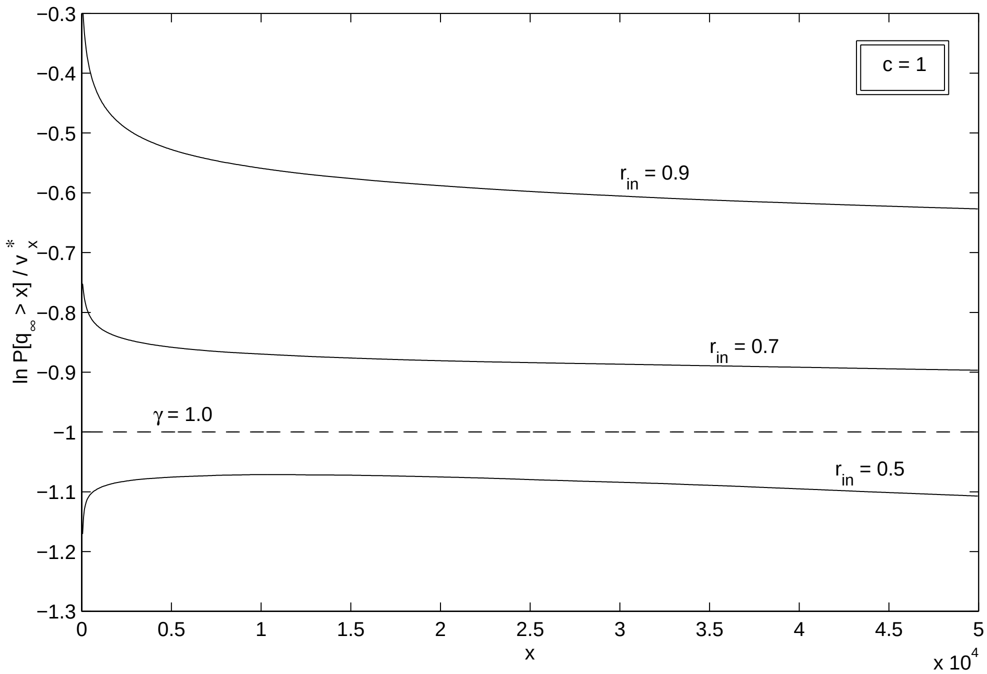

Figure 18: Tail probability vs. buffer size: Pareto ($\alpha = 1.5$)

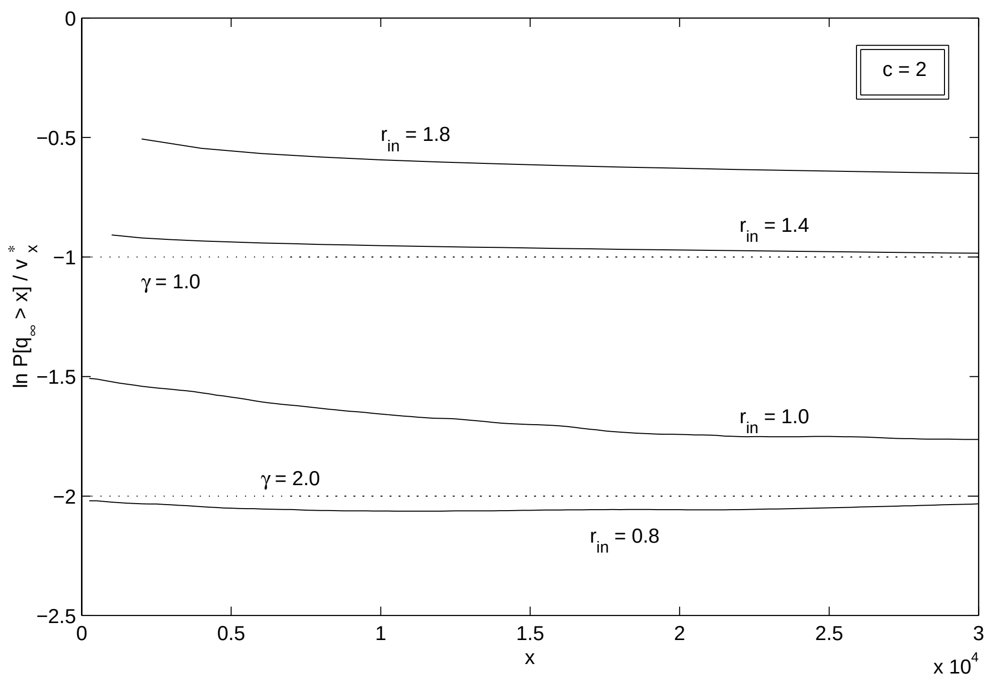

Figure 19: Tail probability vs. buffer size: Pareto ($\alpha = 1.5$)

obviously different scalings. While the utilization $\rho = r_{in}/c$ is of prime importance in predicting the tail probability behavior for exponential and super–exponential distributions, it plays a relatively minor role when G is sub–expon–ential. Instead, the governing factor in this case seems to be the difference $c - r_{in}$. This could perhaps be attributed to the bursty nature of heavy–tailed processes which causes the buffer to grow extremely rapidly in size upon the arrival of a single burst. In such a scenario, the rate at which the queue empties is of great significance. Light–tailed processes in contrast, present a more even supply of packets to the buffer; therefore it is reasonable to expect that the average time that the server is busy, i.e., the utilization ρ, plays a key role in the system dynamics.

As the tail of the distribution G increases in weight, the rate at which the leading term of the scaling $v^\star$ increases to infinity becomes progressively slower, and terms of smaller order begin to gather importance. Eventually, in the very heavy–tailed Lognormal and Pareto cases, we expect even the constant terms to prove significant in establishing reasonably accurate asymptotics. This could very well imply that results of the kind (1.3.3) are truly relevant only at impractically large buffer sizes and that we would be better occupied deriving asymptotics of the form (1.3.2).

Finally, a word of caution regarding simulating heavy–tailed processes: With increasing correlations, the variance exhibited by the process also increases [Proposition 2.3.4]. In fact, when G is Pareto with $1 < \alpha < 2$, the generated process is LRD and exhibits *infinite* variance [Corollary 2.3.1]. Different simulation runs for the same set of parameters could now display vastly different behavior! Longer simulation runs then become imperative in order to ensure meaningful results.

Chapter 7

Concluding Remarks

In Chapter 1 the $M|G|\infty$ process was proposed as a versatile model for packet traffic. We have since examined its various properties in careful detail, paying particular attention to its rich correlation structure. In an effort to understand the dynamics of a system supporting such traffic, the buffer asymptotics of a single–server queue fed by an $M|G|\infty$ traffic stream have been investigated and results of the form (1.3.3) derived, using large deviations techniques.

For a large class of distributions, we have seen that the asymptotics take the compact form

$$\mathbf{P}\left[q_\infty > b\right] \sim \mathbf{P}\left[\widehat{\sigma} > b\right]^{\gamma^\star} \quad (b \to \infty), \tag{7.0.1}$$

implying thereby that q_∞ and $\widehat{\sigma}$ belong to the same distributional class as characterized by tail behavior.

Sometimes, in lieu of (1.3.3), large deviations techniques yield only weaker asymptotics of the form (3.0.2) and (3.0.3). This situation typically occurs when σ is heavy–tailed, in which case large deviations excursions are only one of several causes for buffer exceedances. While the basic functional form of the tail probability is still preserved, it now becomes necessary to pursue alternate methods in order to deduce $\gamma^\star$.

Nonetheless, knowledge of the functional form in itself offers valuable insights into the complex and subtle impact that heavy correlations have on the tail probability $\mathbf{P}[q_\infty > b]$.

Extension to finite–buffered systems: The asymptotics already indicate that the tail probabilities $\mathbf{P}[q_\infty > b]$ display a sub–exponential behavior in the case of heavily–correlated traffic, in sharp contrast to the geometric decay that is usually observed for Markovian input streams. The implications for the corresponding finite–buffered system would then be that lightly correlated traffic gains more (in terms of a decrease in cell-loss) by increasing the buffer size, than does traffic with heavy correlations. This "buffer ineffectiveness "phenomenon has already been observed [28], [32], and clearly indicates the wisdom of improving system performance by multiplexing streams and by investing in faster servers rather than larger buffers.

Parsimonious modeling: In [41] and [42], Leland et al. have stressed the need for parsimonious models for self–similar traffic, using only the Hurst parameter to typify long–range dependence. However, a comparison of results derived earlier with those from [49] clearly points to the inefficacy of such a model in characterizing buffer asymptotics.

Indeed, in [49] the input stream to the multiplexer was modeled as a fractional Gaussian noise process exhibiting long–range dependence (in fact, self–similarity), and the buffer asymptotics displayed Weibull–like characteristics. On the other hand, by the results described in this thesis, an $M|G|\infty$ input process with a Weibull service time also yields Weibull–like buffer asymptotics although the input process is now short–range dependent. Hence, the *same* asymptotic buffer behavior can be induced by two vastly different input streams, one long–range dependent and the other short–range dependent! To make matters worse, if the pmf G were selected to be Pareto instead of Weibull, the input process would be long–range dependent, in fact asymptotically self–similar, but the buffer distribution would now exhibit Pareto–like asymptotics.

This comparison clearly reveals the insufficiency of the Hurst parameter in characterizing buffer asymptotics. Furthermore, buffer sizing cannot be adequately determined by appealing solely to the short versus long–range dependence characterization of the input model used, be it of the $M|G|\infty$ type or otherwise. Of course, long–range dependence (and its close cousin, self–similarity) are determined by second–order properties of the input process, while asymptotics of the form (3.0.1) invoke much finer probabilistic properties. The finiteness of $\mathbf{E}[\sigma^2]$ (needed in (1.2.2)) is obviously a poor marker for predicting the behavior of the sequence $\{v_t^\star,\ t = 1, 2, \ldots\}$ (which drives (3.0.1)).

Utilization versus Difference : Even without precisely identifying the constant $\gamma^\star$, the very form taken by the limiting log–moment function Λ clearly delineates the heavy and light tailed cases. Indeed, the shift in its explicit dependence on the input and output rates r_{in} and c, from $\rho = r_{in}/c$ when G is exponential to $\Delta = c - r_{in}$ when G is sub–exponential, already provides some understanding of the difference in the system dynamics for the two cases. The traditional role of the utilization factor in defining the load in a network must now be re-evaluated in the context of highly correlated traffic.

Having said this, we now briefly visit existing results on the buffer asymptotics in question. Many of the results are surprisingly accurate. The fact that not one of these has been derived via large deviations techniques justifies our earlier intuition that the forces at play extend beyond the realm of large deviations theory.

7.1 Alternate asymptotics

The following result, independently derived by Jelenkovic and Lazar [33], and Daniels and Blondia [14], for a continuous–time $M|G|\infty$ system, applies to the Pareto case. The proofs in both cases rely heavily on Karamata's Tauberian/Abelian theorems.

Proposition 7.1.1 *If G is regularly varying with non–integral exponent $\alpha > 1$, and $c = 1$, the asymptotics*

$$\lim_{b\to\infty} \frac{\mathbf{P}\left[q_\infty > b\right]}{\int_{b/\rho}^{\infty} \mathbf{P}\left[\sigma > u\right] du} \quad = \quad \frac{\lambda}{1-\rho} \tag{7.1.1}$$

hold for $\rho = \lambda \mathbf{E}\left[\sigma\right] < c$.

When extrapolated to the discrete–time $M|G|\infty$ system with G Pareto, the asymptotics described above, translate to

$$\ln \mathbf{P}\left[q_\infty > b\right] \quad \sim \quad -(1-\alpha)\ln\left(\frac{b}{\rho}\right) + \ln\left(\frac{\lambda}{(1-\rho)(1-\alpha)}\right). \tag{7.1.2}$$

Figure 20 displays excellent agreement between the asymptotics described above and the simulated results (denoted by the points), even at smaller time scales.

A more general result due to Likhanov [44] provides bounds for the tail probability $\mathbf{P}\left[q_\infty > b\right]$, which though not quite tight, are remarkably close. The proof involves viewing the input process as a sum of two processes, one of which contributes the long–range behavior, while the other, comprising of the bulk of the inputs, provides the short–range characteristics. Unlike the earlier Proposition, the

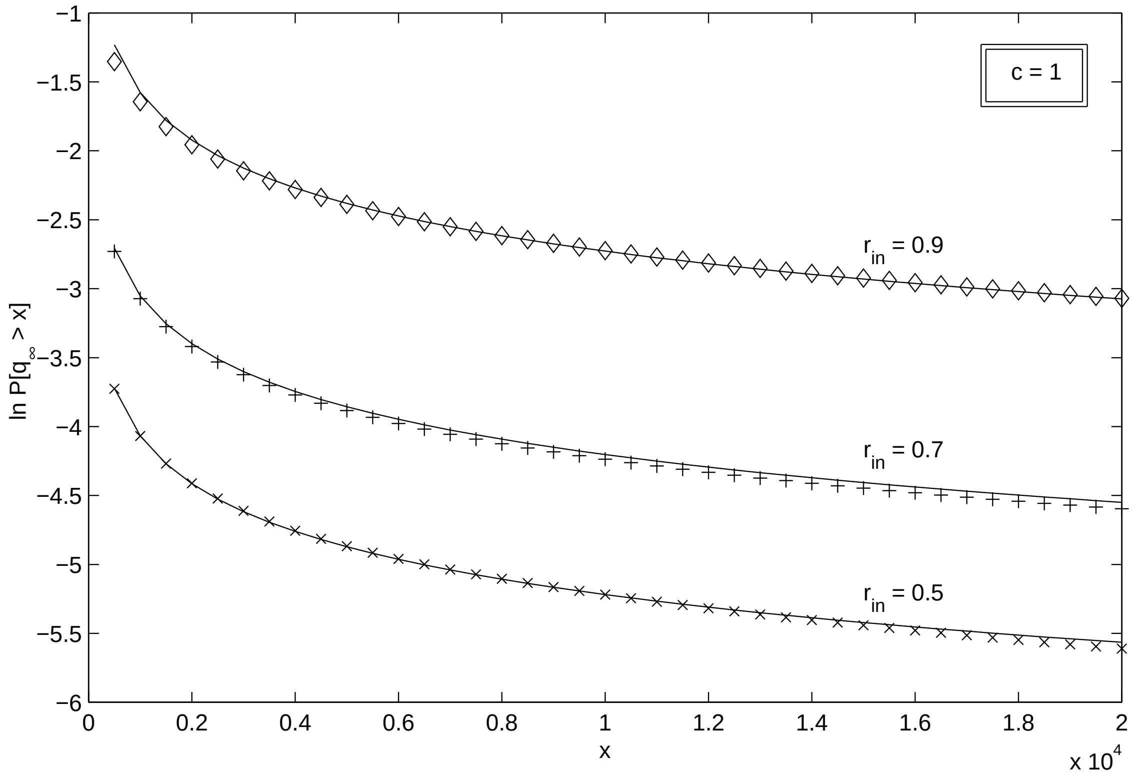

Figure 20: Pareto ($\alpha = 1.5$)

result applies for all values of c, and although derived specifically for the Pareto case can be extended to other cases as well. Work in this direction is currently in progress.

7.2 Directions for future research

The diverse queueing behavior and rich correlation structure demonstrated here confirms the versatility of $M|G|\infty$ inputs as network traffic models. However, several issues important to network design as well as dynamic control have yet to be addressed. Notable amongst these are the first order statistics evident in traffic, and the computation of cell–loss probabilities and buffer dynamics for finite buffered systems.

The heavy correlations inherent in network traffic have occupied most researchers to the degree that first–order characteristics have been side–lined. These have a crucial effect on buffer dynamics and must be accounted for [28, 57]. As far as the $M|G|\infty$ process is concerned, its Poisson marginals may be suitably adapted by a simple transformation described in [39] and [40].

There are researchers who believe that the solutions to many modeling questions lie in the underlying physical mechanism that causes the long–range dependence. On the opposite end of the spectrum, there are those who claim that in the scramble to provide accurate models for LRD traffic, the practical significance of these representations has been overlooked. While it is important for a model to provide a close fit to the data, the superiority of a model is decided by the quality of decisions it makes *in the regime of interest*. The fact that buffers in real systems are finite in size creates a very different scenario from the one studied in this thesis, by setting a hard limit on the memory of the system. The system in the latter case is reset only when the the buffer is empty; in the former case however, this happens when the buffer is empty *or full*. Predictably then, correlations for lag greater than the buffer size will be of little consequence in the finite–queued system. This finite "correlation horizon " [28] explains why literature on Markov modeling reports good performance prediction for finite buffer systems even when input traffic streams are correlated over many time–scales [18, 58].

In conclusion, one can only state the very obvious, namely, that the problem of modeling network traffic when time dependencies are either observed or suspected must be approached with caution.

Appendix A

A.1 Proof of Proposition 2.1.1

We present the proof of Proposition 2.1.1 in the form of the two following Lemmas. Throughout we employ the notation of Section 2.1.3.

Lemma A.1.1 *For each pair* (T_n, Q_n) *in* $(\mathcal{T}_n, \mathcal{Q}_n),\ n = 1, 2, \ldots,$ *it holds that*

$$\ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}^{(0)}\right)\right] = \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b}\sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j} \in I_r\right]\right)\right]$$

where $\Theta_r,\ r = 1, 2, \ldots, n,$ *is given by (2.1.11).*

Proof. Using (2.1.3) we note that

$$\begin{aligned}
\sum_{i=1}^{n} \theta_i b_{t_i}^{(0)} &= \sum_{j=1}^{b}\sum_{i=1}^{n} \theta_i \mathbf{1}\left[\sigma_{0,j} > t_i\right] \\
&= \sum_{j=1}^{b}\sum_{i=1}^{n} \theta_i \mathbf{1}\left[\sigma_{0,j} \in \bigcup_{r=i}^{n} I_r\right] \\
&= \sum_{j=1}^{b}\sum_{i=1}^{n} \theta_i \sum_{r=i}^{n} \mathbf{1}\left[\sigma_{0,j} \in I_r\right] \\
&= \sum_{j=1}^{b}\sum_{r=1}^{n} \left(\sum_{i=1}^{r} \theta_i\right) \mathbf{1}\left[\sigma_{0,j} \in I_r\right]
\end{aligned}$$

and the result follows. ■

Lemma A.1.2 *For each pair* (T_n, Q_n) *in* $(\mathcal{T}_n, \mathcal{Q}_n)$, $n = 1, 2, \ldots,$ *we have*

$$\ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}^{(a)}\right)\right] \quad = \quad \lambda \mathbf{E}\left[\sigma\right] \sum_{j=1}^{n} \Phi_j(T_n, Q_n),$$

where $\Phi_j(T_n, Q_n),\ j = 1, 2, \ldots, n,$ *is given by (2.1.12).*

Proof. We attempt to write the exponent as a sum of independent rvs. To do so we use (2.1.4) to get

$$\begin{aligned} b_{t_i}^{(a)} &= \sum_{s=1}^{t_i} \sum_{m=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,m} > t_i - s\right] \\ &= \sum_{k=0}^{i-1} \sum_{s \in I_k} \sum_{m=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,m} > t_i - s\right], \qquad i = 1, 2, \ldots, n \end{aligned}$$

upon expressing the interval $(0, t_i]$ as $\bigcup_{k=0}^{i-1} I_k$. This gives

$$\begin{aligned} \sum_{i=1}^{n} \theta_i b_{t_i}^{(a)} &= \sum_{i=1}^{n} \theta_i \left(\sum_{k=0}^{i-1} \sum_{s \in I_k} \sum_{m=1}^{\beta_s} \mathbf{1}\left[\sigma_{s,m} > t_i - s\right] \right) \\ &= \sum_{k=0}^{n-1} \sum_{s \in I_k} \sum_{m=1}^{\beta_s} \sum_{i=k+1}^{n} \theta_i \mathbf{1}\left[\sigma_{s,m} > t_i - s\right] \\ &= \sum_{k=0}^{n-1} \sum_{s \in I_k} \sum_{m=1}^{\beta_s} U_{k,s,m} \end{aligned} \tag{A.1.1}$$

where

$$U_{k,s,m} \equiv \sum_{i=k+1}^{n} \theta_i \mathbf{1}\left[\sigma_{s,m} > t_i - s\right]$$

$$
\begin{aligned}
&= \sum_{i=k+1}^{n} \theta_i \mathbf{1}\left[\sigma_{s,m} + s \in \bigcup_{j=i}^{n} I_j\right] \\
&= \sum_{i=k+1}^{n} \theta_i \sum_{j=i}^{n} \mathbf{1}\left[\sigma_{s,m} + s \in I_j\right] \\
&= \sum_{j=k+1}^{n} \left(\sum_{i=k+1}^{j} \theta_i\right) \mathbf{1}\left[\sigma_{s,m} + s \in I_j\right] \\
&= \sum_{j=k+1}^{n} (\Theta_j - \Theta_k) \mathbf{1}\left[\sigma_{s,m} + s \in I_j\right].
\end{aligned}
$$

Recall that the rvs $\{\sigma_{s,m},\ m = 1, 2, \ldots;\ s = 1, 2, \ldots\}$ are $i.i.d.$ with common distribution G. Hence, it follows from (A.1.1) that

$$
\begin{aligned}
\mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}^{(a)}\right)\right] &= \prod_{k=0}^{n-1} \prod_{s \in I_k} \mathbf{E}\left[\prod_{m=1}^{\beta_s} e^{U_{k,s,m}}\right] \\
&= \prod_{k=0}^{n-1} \prod_{s \in I_k} \mathbf{E}\left[\left(\exp\left(\sum_{j=k+1}^{n} (\Theta_j - \Theta_k) \mathbf{1}\left[\sigma + s \in I_j\right]\right)\right)^{\beta_s}\right] \\
&= \prod_{k=0}^{n-1} \prod_{s \in I_k} \mathbf{E}\left[\left(\chi_{k,s}\right)^{\beta_s}\right] \\
&= \prod_{k=0}^{n-1} \prod_{s \in I_k} e^{\lambda(\chi_{k,s} - 1)}, \qquad \text{(A.1.2)}
\end{aligned}
$$

with

$$
\chi_{k,s} \equiv \mathbf{E}\left[\exp\left(\sum_{j=k+1}^{n} (\Theta_j - \Theta_k) \mathbf{1}\left[\sigma + s \in I_j\right]\right)\right], \qquad \text{(A.1.3)}
$$

for $k = 0, 1, \ldots, n-1$, and s in I_k, and upon using the independence of the rvs $\{\sigma_{s,m},\ m = 1, 2, \ldots;\ s = 1, 2, \ldots\}$ and $\{\beta_s,\ s = 1, 2, \ldots\}$, together with the fact that the rv β_s is Poisson with rate λ. Simplifying (A.1.3) we have

$$
\begin{aligned}
\chi_{k,s} &= 1 - \mathbf{P}\left[\sigma + s > t_{k+1}\right] + \sum_{j=k+1}^{n} e^{\Theta_j - \Theta_k} \mathbf{P}\left[\sigma + s \in I_j\right] \\
&= 1 - \sum_{j=k+1}^{n} \mathbf{P}\left[\sigma + s \in I_j\right] + \sum_{j=k+1}^{n} e^{\Theta_j - \Theta_k} \mathbf{P}\left[\sigma + s \in I_j\right]
\end{aligned}
$$

$$= \quad 1 + \sum_{j=k+1}^{n} \left(e^{\Theta_j - \Theta_k} - 1\right) \mathbf{P}\left[\sigma + s \in I_j\right]. \tag{A.1.4}$$

Combining (A.1.2) and (A.1.4) gives

$$\begin{aligned}
\mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}^{(a)}\right)\right] &= \exp\left(\lambda \sum_{k=0}^{n-1} \sum_{s \in I_k} (\chi_{k,s} - 1)\right) \\
&= \exp\left(\lambda \sum_{k=0}^{n-1} \sum_{j=k+1}^{n} \left(e^{\Theta_j - \Theta_k} - 1\right) \sum_{s \in I_k} \mathbf{P}\left[\sigma + s \in I_j\right]\right) \\
&= \exp\left(\lambda \mathbf{E}\left[\sigma\right] \sum_{j=1}^{n} \Phi_j(T_n, Q_n)\right) \qquad \text{(A.1.5)}
\end{aligned}$$

where

$$\Phi_j(T_n, Q_n) \quad = \quad \frac{1}{\mathbf{E}\left[\sigma\right]} \sum_{k=0}^{j-1} \left(e^{\Theta_j - \Theta_k} - 1\right) \sum_{s \in I_k} \mathbf{P}\left[\sigma + s \in I_j\right] \tag{A.1.6}$$

for $j = 1, \ldots, n$.

All that remains to be done now, is to rewrite (A.1.6) in order to show its equivalence to representation (2.1.12): Using (2.1.6) to expand the summation in (A.1.6), we have

$$\begin{aligned}
\sum_{s \in I_k} \mathbf{P}\left[\sigma + s \in I_j\right] &= \sum_{s=t_k+1}^{t_{k+1}} \mathbf{P}\left[t_j < \sigma + s \leq t_{j+1}\right] \\
&= \sum_{s=t_k+1}^{t_{k+1}} \mathbf{P}\left[\sigma + s \leq t_{j+1}\right] - \mathbf{P}\left[\sigma + s \leq t_j\right].
\end{aligned}$$

The fact that $k < j$ in (A.1.6) ensures that $t_{k+1} \leq t_j$, thus allowing the previous expression to be re–organized using the substitution $u = t_{j+1} - s + 1$ in the first summation, and $u = t_j - s + 1$ in the second. This gives

$$\begin{aligned}
\sum_{s \in I_k} \mathbf{P}\left[\sigma + s \in I_j\right] &= \sum_{u=t_{j+1}-t_{k+1}+1}^{t_{j+1}-t_k} \mathbf{P}\left[\sigma \leq u - 1\right] - \sum_{u=t_j-t_{k+1}+1}^{t_j-t_k} \mathbf{P}\left[\sigma \leq u - 1\right] \\
&= \sum_{u=t_{j+1}-t_{k+1}+1}^{\infty} \mathbf{P}\left[\sigma < u\right] - \sum_{u=t_{j+1}-t_k+1}^{\infty} \mathbf{P}\left[\sigma < u\right]
\end{aligned}$$

$$- \sum_{u=t_j-t_{k+1}+1}^{\infty} \mathbf{P}[\sigma < u] + \sum_{u=t_j-t_k+1}^{\infty} \mathbf{P}[\sigma < u]$$

$$= \sum_{u=t_j-t_k+1}^{t_{j+1}-t_k} \mathbf{P}[\sigma < u] - \sum_{u=t_j-t_{k+1}+1}^{t_{j+1}-t_{k+1}} \mathbf{P}[\sigma < u].$$

Finally, using the substitution $r = u + t_k$ in the first summation and $r = u + t_{k+1}$ in the second, we conclude that

$$\begin{aligned}\sum_{s\in I_k} \mathbf{P}[\sigma + s \in I_j] &= \sum_{r=t_j+1}^{t_{j+1}} \mathbf{P}[\sigma < r - t_k] - \sum_{r=t_j+1}^{t_{j+1}} \mathbf{P}[\sigma < r - t_{k+1}] \\ &= \sum_{r\in I_j} \mathbf{P}[t_k < r - \sigma] - \mathbf{P}[t_{k+1} < r - \sigma] \\ &= \sum_{r\in I_j} \mathbf{P}[\sigma \geq r - t_{k+1}] - \mathbf{P}[\sigma \geq r - t_k]\end{aligned}$$

for $k < j$. Upon expressing σ in terms of $\widehat{\sigma}$ via (2.1.1), we have

$$\begin{aligned}\frac{1}{\mathbf{E}[\sigma]} \sum_{s\in I_k} \mathbf{P}[\sigma + s \in I_j] &= \sum_{r=t_j+1}^{t_{j+1}} \mathbf{P}[\widehat{\sigma} + t_{k+1} = r] - \mathbf{P}[\widehat{\sigma} + t_k = r] \\ &= \mathbf{P}[\widehat{\sigma} + t_{k+1} \in I_j] - \mathbf{P}[\widehat{\sigma} + t_k \in I_j] \qquad \text{(A.1.7)}\end{aligned}$$

for $j = 1, 2, \ldots, n$ and $k = 0, 1, \ldots, j-1$. Incorporating (A.1.7) in (A.1.6) gives

$$\begin{aligned}\Phi_j(T_n, Q_n) &= \sum_{k=0}^{j-1} \left(e^{\Theta_j - \Theta_k} - 1\right) \Big(\mathbf{P}[\widehat{\sigma} + t_{k+1} \in I_j] - \mathbf{P}[\widehat{\sigma} + t_k \in I_j]\Big) \\ &= \sum_{k=1}^{j} \left(e^{\Theta_j - \Theta_{k-1}} - 1\right) \mathbf{P}[\widehat{\sigma} + t_k \in I_j] \\ &\qquad - \sum_{k=0}^{j-1} \left(e^{\Theta_j - \Theta_k} - 1\right) \mathbf{P}[\widehat{\sigma} + t_k \in I_j] \\ &= \left(1 - e^{\Theta_j}\right) \mathbf{P}[\widehat{\sigma} \in I_j] + \sum_{k=1}^{j} e^{\Theta_j - \Theta_k} \left(e^{\theta_k} - 1\right) \mathbf{P}[\widehat{\sigma} + t_k \in I_j]\end{aligned}$$

which in conjunction with (A.1.5) concludes the proof. ■

A.2 Proof of Proposition 2.2.1

Fix $n = 1, 2, \ldots,$ T_n in $\mathcal{T}_n$ and Q_n in $\mathcal{Q}_n$. We establish Proposition 2.2.1 through the following series of Lemmas.

Lemma A.2.1 *For every $h = 0, 1, \ldots,$ we have*

$$
\begin{aligned}
& \mathcal{L}^\star(T_n \oplus h, Q_n) - \mathcal{L}^\star(T_n, Q_n) \\
= \;& \lambda \mathbf{E}[\sigma] \sum_{j=1}^{n} \left(1 - e^{\Theta_j}\right) \Big(\mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] - \mathbf{P}\left[\widehat{\sigma} \in I_j\right] \Big) \\
& + \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r \oplus h\right] \right) \right] \\
& - \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r\right] \right) \right]. \qquad \text{(A.2.1)}
\end{aligned}
$$

Proof. Fix $h = 0, 1, \ldots$. Using (2.1.13) to compute $\mathcal{L}^\star(T_n \oplus h, Q_n)$, we have

$$
\begin{aligned}
\mathcal{L}^\star(T_n \oplus h, Q_n) &= \lambda \mathbf{E}[\sigma] \sum_{j=1}^{n} \Phi_j(T_n \oplus h, Q_n) \\
& \quad + \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r \oplus h\right] \right) \right] \qquad \text{(A.2.2)}
\end{aligned}
$$

with

$$
\begin{aligned}
& \Phi_j(T_n \oplus h, Q_n) \\
= \;& \left(1 - e^{\Theta_j}\right) \mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] + \sum_{k=1}^{j} e^{\Theta_j - \Theta_k} \left(e^{\theta_k} - 1\right) \mathbf{P}\left[\widehat{\sigma} + t_k + h \in I_j \oplus h\right] \\
= \;& \left(1 - e^{\Theta_j}\right) \mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] + \sum_{k=1}^{j} e^{\Theta_j - \Theta_k} \left(e^{\theta_k} - 1\right) \mathbf{P}\left[\widehat{\sigma} + t_k \in I_j\right] \\
= \;& \Phi_j(T_n, Q_n) + \left(1 - e^{\Theta_j}\right) \Big(\mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] - \mathbf{P}\left[\widehat{\sigma} \in I_j\right] \Big), \qquad j = 1, \ldots, n.
\end{aligned}
$$

Using this last relation to rewrite (A.2.2), we have

$$\begin{aligned}
&\mathcal{L}^\star(T_n \oplus h, Q_n) \\
=\ & \lambda \mathbf{E}[\sigma] \sum_{j=1}^{n} \left(\Phi_j(T_n, Q_n) + \left(1 - e^{\Theta_j}\right) \left(\mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] - \mathbf{P}\left[\widehat{\sigma} \in I_j\right] \right) \right) \\
& \quad + \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r \oplus h\right] \right) \right] \\
=\ & \mathcal{L}^\star(T_n, Q_n) - \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r\right] \right) \right] \\
& \quad + \lambda \mathbf{E}[\sigma] \sum_{j=1}^{n} \left(1 - e^{\Theta_j}\right) \left(\mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] - \mathbf{P}\left[\widehat{\sigma} \in I_j\right] \right) \\
& \quad + \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r \oplus h\right] \right) \right],
\end{aligned}$$

and the result is established. ■

Lemma A.2.2 *Proposition 2.2.1 (i) implies Proposition 2.2.1 (ii).*

Proof. Fix $n = 1, 2, \ldots$. Under Proposition 2.2.1 (i), we can rewrite (A.2.1) as

$$\begin{aligned}
& \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r \oplus h\right] \right) \right] \\
& - \ln \mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star} \sum_{r=1}^{n} \Theta_r \mathbf{1}\left[\sigma_{0,j}^\star \in I_r\right] \right) \right] \\
=\ & \lambda \mathbf{E}[\sigma] \sum_{j=1}^{n} \left(1 - e^{\Theta_j}\right) \left(\mathbf{P}\left[\widehat{\sigma} \in I_j\right] - \mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right] \right)
\end{aligned} \tag{A.2.3}$$

for every $h = 0, 1, \ldots$. Let h go to infinity. This forces

$$\mathbf{P}\left[\widehat{\sigma} \in I_j \oplus h\right]\Big|_{h=\infty} = 0, \qquad j = 1, 2, \ldots, n$$

and

$$\sum_{j=1}^{b_0^\star}\sum_{r=1}^{n}\Theta_r\mathbf{1}\left[\sigma_{0,j}^\star\in I_r\oplus h\right]\Bigg|_{h=\infty} \quad = \quad 0,$$

thus obliterating any effect of the initial conditions, and (A.2.3) in the limit becomes

$$\begin{aligned}&\ln\mathbf{E}\left[\exp\left(\sum_{j=1}^{b_0^\star}\sum_{r=1}^{n}\Theta_r\mathbf{1}\left[\sigma_{0,j}^\star\in I_r\right]\right)\right]\\ = \quad &\lambda\mathbf{E}\left[\sigma\right]\sum_{j=1}^{n}\left(e^{\Theta_j}-1\right)\mathbf{P}\left[\widehat{\sigma}\in I_j\right].\end{aligned} \tag{A.2.4}$$

We conclude the proof by recognizing the equivalence of (A.2.4) and Proposition 2.2.1 (ii) via (2.1.9). ■

Lemma A.2.3 *Proposition 2.2.1 (ii) implies Proposition 2.2.1 (iii).*

Proof. Assume that Proposition 2.2.1 (ii) holds. Then, as noted in the proof of the previous lemma, (A.2.4) holds through (2.1.9). This allows (2.1.13) to be modified to

$$\begin{aligned}\mathcal{L}^\star(T_n,Q_n) \quad &= \quad \lambda\mathbf{E}\left[\sigma\right]\sum_{j=1}^{n}\left(\Phi_j(T_n,Q_n)+\left(e^{\Theta_j}-1\right)\mathbf{P}\left[\widehat{\sigma}\in I_j\right]\right)\\ &= \quad \lambda\mathbf{E}\left[\sigma\right]\sum_{j=1}^{n}\sum_{k=1}^{j}e^{\Theta_j-\Theta_k}\left(e^{\theta_k}-1\right)\mathbf{P}\left[\widehat{\sigma}+t_k\in I_j\right] &\text{(A.2.5)}\\ &= \quad \lambda\mathbf{E}\left[\sigma\right]\sum_{k=1}^{n}\left(e^{\theta_k}-1\right)\sum_{j=k}^{n}e^{\Theta_j-\Theta_k}\mathbf{P}\left[\widehat{\sigma}+t_k\in I_j\right],\end{aligned}$$

with (A.2.5) following on comparison with (2.1.12). The final step is achieved by interchanging the order of summation and proves Proposition 2.2.1 (iii). ■

We conclude with Lemma A.2.4 which provides the final link in proving Proposition 2.2.1.

Lemma A.2.4 *Proposition 2.2.1 (iii) implies Proposition 2.2.1 (i).*

Proof. Fix $n = 1, 2, \ldots$ and $h = 0, 1, \ldots$. The result follows directly through the observation that

$$\begin{aligned}\mathbf{P}\left[\widehat{\sigma} + t_k + h \in I_j \oplus h\right] &= \mathbf{P}\left[t_j + h < \widehat{\sigma} + t_k + h < t_{j+1} + h\right] \\ &= \mathbf{P}\left[\widehat{\sigma} + t_k \in I_j\right],\end{aligned}$$

for $j = k, k+1, \ldots, n;\ k = 1, 2, \ldots, n.$ ■

A.3 Proof of Proposition 2.2.2

Fix $n,\ m = 1, 2, \ldots$. Consider sequences $U_n \equiv \{u_i,\ i = 1, 2, \ldots, n\}$ in $\mathcal{T}^n$ and $V_m \equiv \{v_j,\ j = 1, 2, \ldots, m\}$ in $\mathcal{T}^m$, and sequences $\Psi_n \equiv \{\psi_i,\ i = 1, 2, \ldots, n\}$ in $\mathcal{Q}^n$ and $\Phi_m \equiv \{\phi_j,\ j = 1, 2, \ldots, m\}$ in $\mathcal{Q}^m$.

Define

$$\mathcal{L}^\star\Big((U_n, \Psi_n), (V_m, \Phi_m)\Big) \equiv \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \psi_i b^\star_{u_i} + \sum_{j=1}^{m} \phi_j b^\star_{v_j}\right)\right]. \qquad \text{(A.3.1)}$$

In order to prove that the process $\{b^\star_t,\ t = 0, 1, \ldots\}$ is strongly mixing via the property described by (2.2.7), it suffices to prove the following Lemma.

Lemma A.3.1 *For all pairs* (U_n, V_m) *in* $(\mathcal{T}^n, \mathcal{T}^m)$ *and* (Ψ_n, Φ_m) *in* $(\mathcal{Q}^n, \mathcal{Q}^m)$*, we have*

$$\lim_{h\to\infty} \mathcal{L}^\star\Big((U_n, \Psi_n), (V_m \oplus h, \Phi_m)\Big) \quad = \quad \mathcal{L}^\star(U_n, \Psi_n) + \mathcal{L}^\star(V_m, \Phi_m),$$

where $h = 1, 2, \ldots,$ *and the function* $\mathcal{L}^\star$ *is as defined in (2.1.7).*

Proof. Fix $h = 1, 2, \ldots,$ U_n in $\mathcal{T}^n, V_m$ in $\mathcal{T}^m$ and Ψ_n in $\mathcal{Q}^n, \Phi_m$ in $\mathcal{Q}^m$. Using the notation

$$\Delta_{m+n} = \{\delta_i,\ i = 1, 2, \ldots, m+n\}, \quad \delta_i = \begin{cases} \psi_i & 1 \le i \le n \\ \phi_{i-n} & n+1 \le i \le n+m, \end{cases}$$

and

$$W_{m+n} = \{w_i,\ i = 1, 2, \ldots, m+n\}, \quad w_i = \begin{cases} u_i & 1 \le i \le n \\ v_{i-n} + h & n+1 \le i \le n+m, \end{cases}$$

we rewrite (A.3.1) in the more familiar form

$$\mathcal{L}^\star\Big((U_n, \Psi_n), (V_m \oplus h, \Phi_m)\Big) \quad = \quad \mathcal{L}^\star(W_{m+n}, \Delta_{m+n}). \tag{A.3.2}$$

By the stationarity of the process $\{b_t^\star,\ t = 0, 1, \ldots\}$, Proposition 2.2.1 (iii) applies to both pairs (U_n, V_m) and (Ψ_n, Φ_m), yielding

$$\mathcal{L}^\star(U_n, \Psi_n) \quad = \quad \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n} \Big(e^{\psi_k} - 1\Big) \sum_{j=k}^{n} e^{\mathcal{Y}_j - \mathcal{Y}_k} \mathbf{P}\left[u_j < \widehat{\sigma} + u_k \le u_{j+1}\right] \tag{A.3.3}$$

and

$$\mathcal{L}^\star(V_m, \Phi_m) \quad = \quad \lambda \mathbf{E}[\sigma] \sum_{k=1}^{m} \Big(e^{\phi_k} - 1\Big) \sum_{j=k}^{m} e^{\mathcal{P}_j - \mathcal{P}_k} \mathbf{P}\left[v_j < \widehat{\sigma} + v_k \le v_{j+1}\right] \tag{A.3.4}$$

where $\mathcal{Y}_j \equiv \sum_{i=1}^{n} \psi_i,\ j = 1, 2, \ldots, n,\ \mathcal{P}_j \equiv \sum_{i=1}^{n} \phi_i,\ j = 1, 2, \ldots, m,$ and $u_{n+1} = v_{m+1} = \infty$ by convention.

In order to write a similar equation for the pair (W_{m+n}, Δ_{m+n}), we require that W_{m+n} be an element of $\mathcal{T}^{n+m}$, a condition always fulfilled when $h > u_n$. In that case we have

$$\begin{aligned} & \mathcal{L}^{\star}(W_{m+n}, \Delta_{m+n}) \\ = \quad & \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n+m} \left(e^{\delta_k} - 1\right) \sum_{j=k}^{n+m} e^{\mathcal{D}_j - \mathcal{D}_k} \mathbf{P}\left[w_j < \widehat{\sigma} + w_k \leq w_{j+1}\right] \end{aligned} \tag{A.3.5}$$

where

$$\mathcal{D}_j \equiv \sum_{i=1}^{j} \delta_i = \begin{cases} \mathcal{Y}_j & 1 \leq j \leq n \\ \mathcal{Y}_n + \mathcal{P}_{j-n} & n+1 \leq j \leq n+m \end{cases}$$

and $w_{n+m+1} = \infty$.

We now attempt to evaluate $\lim_{h\to\infty} \mathcal{L}^{\star}(W_{m+n}, \Delta_{m+n})$. For this purpose, we split the double summation in (A.3.5) into three component sums, namely

$$\begin{aligned} & \lim_{h\to\infty} \mathcal{L}^{\star}(W_{m+n}, \Delta_{m+n}) \\ = \quad & \lambda \mathbf{E}[\sigma] \lim_{h\to\infty} \sum_{k=1}^{n+m} \sum_{j=k}^{n+m} \mathcal{A}_{j,k} \\ = \quad & \lambda \mathbf{E}[\sigma] \lim_{h\to\infty} \left(\sum_{k=1}^{n} \sum_{j=k}^{n} \mathcal{A}_{j,k} + \sum_{k=1}^{n} \sum_{j=n+1}^{n+m} \mathcal{A}_{j,k} + \sum_{k=n+1}^{n+m} \sum_{j=k}^{n+m} \mathcal{A}_{j,k} \right) \end{aligned} \tag{A.3.6}$$

where

$$\mathcal{A}_{j,k} \quad \equiv \quad e^{\mathcal{D}_j - \mathcal{D}_k} \left(e^{\delta_k} - 1\right) \mathbf{P}\left[w_j < \widehat{\sigma} + w_k \leq w_{j+1}\right]$$

for $j = k, \ldots n+m; k = 1, \ldots n+m$.

When $1 \leq k \leq j \leq n$, we have

$$\mathcal{A}_{j,k} = \begin{cases} e^{\mathcal{Y}_j - \mathcal{Y}_k} \left(e^{\psi_k} - 1\right) \mathbf{P}\left[u_j < \widehat{\sigma} + u_k \leq u_{j+1}\right] & 1 \leq j < n \\ e^{\mathcal{Y}_n - \mathcal{Y}_k} \left(e^{\psi_k} - 1\right) \mathbf{P}\left[u_n < \widehat{\sigma} + u_k \leq v_1 + h\right] & j = n. \end{cases} \tag{A.3.7}$$

For the region $1 \leq k \leq n < j \leq n+m$, we have

$$\begin{aligned}\mathcal{A}_{j,k} &= e^{\mathcal{Y}_n+\mathcal{P}_{j-n}-\mathcal{Y}_k}\Big(e^{\psi_k}-1\Big)\mathbf{P}\left[v_{j-n}+h<\widehat{\sigma}+u_k\leq v_{j+1-n}+h\right]\\ &= e^{\mathcal{Y}_n+\mathcal{P}_{j'}-\mathcal{Y}_k}\Big(e^{\psi_k}-1\Big)\mathbf{P}\left[v_{j'}+h<\widehat{\sigma}+u_k\leq v_{j'+1}+h\right] \qquad \text{(A.3.8)}\end{aligned}$$

where $j'=j-n$. Finally, when $n+1\leq k\leq j\leq n+m$,

$$\begin{aligned}\mathcal{A}_{j,k} &= e^{\mathcal{P}_{j-n}-\mathcal{P}_{k-n}}\Big(e^{\phi_{k-n}}-1\Big)\mathbf{P}\left[v_{j-n}+h<\widehat{\sigma}+v_{k-n}+h\leq v_{j+1-n}+h\right]\\ &= e^{\mathcal{P}_{j'}-\mathcal{P}_{k'}}\Big(e^{\phi_{k'}}-1\Big)\mathbf{P}\left[v_{j'}<\widehat{\sigma}+v_{k'}\leq v_{j'+1}\right] \qquad \text{(A.3.9)}\end{aligned}$$

using the substitution $j'=j-n$ and $k'=k-n$.

We report (A.3.7), (A.3.8), and (A.3.9) in each of the components of (A.3.6) and let h go to infinity. We first find

$$\begin{aligned}&\lim_{h\to\infty}\sum_{k=1}^{n}\sum_{j=k}^{n}\mathcal{A}_{j,k}\\ &= \sum_{k=1}^{n}\sum_{j=k}^{n-1}e^{\mathcal{Y}_j-\mathcal{Y}_k}\Big(e^{\psi_k}-1\Big)\mathbf{P}\left[u_j<\widehat{\sigma}+u_k\leq u_{j+1}\right]\\ &\quad +\lim_{h\to\infty}\sum_{k=1}^{n}e^{\mathcal{Y}_n-\mathcal{Y}_k}\Big(e^{\psi_k}-1\Big)\mathbf{P}\left[u_n<\widehat{\sigma}+u_k\leq v_1+h\right]\\ &= \frac{1}{\lambda\mathbf{E}\left[\sigma\right]}\mathcal{L}^\star(U_n,\Psi_n) \qquad \text{(A.3.10)}\end{aligned}$$

with the help of (A.3.3). Next,

$$\begin{aligned}&\lim_{h\to\infty}\sum_{k=1}^{n}\sum_{j=n+1}^{n+m}\mathcal{A}_{j,k}\\ &= \lim_{h\to\infty}\sum_{k=1}^{n}\sum_{j'=1}^{m}e^{\mathcal{Y}_n+\mathcal{P}_{j'}-\mathcal{Y}_k}\Big(e^{\psi_k}-1\Big)\mathbf{P}\left[v_{j'}<\widehat{\sigma}+u_k-h\leq v_{j'+1}\right]\\ &= 0, \qquad \text{(A.3.11)}\end{aligned}$$

while

$$\lim_{h\to\infty}\sum_{k=n+1}^{n+m}\sum_{j=k}^{n+m}\mathcal{A}_{j,k}$$

$$= \lim_{h\to\infty} \sum_{k'=1}^{m} \sum_{j'=k'}^{n} e^{\mathcal{P}_{j'}-\mathcal{P}_{k'}} \left(e^{\phi_{k'}} - 1\right) \mathbf{P}\left[v_{j'} < \widehat{\sigma} + v_{k'} \le v_{j'+1}\right]$$

$$= \frac{1}{\lambda \mathbf{E}[\sigma]} \mathcal{L}^\star(V_n, \Phi_n) \qquad (A.3.12)$$

by consulting (A.3.4).

Combining (A.3.10), (A.3.11) and (A.3.12) with (A.3.6), we conclude that

$$\lim_{h\to\infty} \mathcal{L}^\star(W_{m+n}, \Delta_{m+n}) = \mathcal{L}^\star(U_n, \Psi_n) + \mathcal{L}^\star(V_n, \Phi_n),$$

and the required result follows on comparison with (A.3.2). ■

A.4 Proof of Proposition 2.2.3

We begin by proving the following lemma which essentially computes $\mathcal{L}^\star(H_n, Q_n)$ for $H_n = \{1, 2, \ldots, n\}$ and Q_n in $\mathcal{Q}^n$ for each $n = 1, 2, \ldots$.

Lemma A.4.1 *For all $n = 1, 2, \ldots,$ and all Q_n in $\mathcal{Q}^n$, it holds that*

$$\mathcal{L}^\star(H_n, Q_n) = \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n} \left(e^{\theta_k} - 1\right) \left(1 + \sum_{j=k+1}^{n} e^{\Theta_{j-1}-\Theta_k} \left(e^{\theta_j} - 1\right) \mathbf{P}\left[\widehat{\sigma} > j-k\right]\right) \qquad (A.4.1)$$

where $H_n = \{1, 2, \ldots, n\}$.

Proof. Fix $n = 1, 2, \ldots,$ and Q_n in $\mathcal{Q}^n$. Specializing Proposition 2.2.1 (iii) to the case $T_n = H_n = \{1, 2, \ldots, n\}$, we get

$$\mathcal{L}^\star(H_n, Q_n) = \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n} \left(e^{\theta_k} - 1\right) \sum_{j=k}^{n} e^{\Theta_j - \Theta_k} \mathbf{P}\left[\widehat{\sigma} + k \in I_j\right] \qquad (A.4.2)$$

where

$$I_j = \begin{cases} (j, j+1], & j = 1, 2, \ldots, n-1 \\ (n, \infty], & j = n. \end{cases} \tag{A.4.3}$$

By (A.4.3), we have

$$\begin{aligned} &\sum_{j=k}^{n} e^{\Theta_j - \Theta_k} \mathbf{P}[\widehat{\sigma} + k \in I_j] \\ = \;& e^{\Theta_n - \Theta_k} \mathbf{P}[\widehat{\sigma} + k > n] + \sum_{j=k}^{n-1} e^{\Theta_j - \Theta_k} \mathbf{P}[j < \widehat{\sigma} + k \leq j+1] \\ = \;& \sum_{j=k}^{n} e^{\Theta_j - \Theta_k} \mathbf{P}[\widehat{\sigma} > j-k] - \sum_{j=k+1}^{n} e^{\Theta_{j-1} - \Theta_k} \mathbf{P}[\widehat{\sigma} > j-k] \\ = \;& 1 + \sum_{j=k+1}^{n} \left(e^{\Theta_j - \Theta_k} - e^{\Theta_{j-1} - \Theta_k}\right) \mathbf{P}[\widehat{\sigma} > j-k] \\ = \;& 1 + \sum_{j=k+1}^{n} e^{\Theta_{j-1} - \Theta_k} \left(e^{\theta_j} - 1\right) \mathbf{P}[\widehat{\sigma} > j-k], \end{aligned}$$

which on comparison with (A.4.2) easily yields the required result. ■

To proceed with the proof of Proposition 2.2.3, we apply Lemma A.4.1 to the reversed sequence $Q_n^r = (\theta_n, \theta_{n-1}, \ldots, \theta_1)$. This gives

$$\begin{aligned} &\mathcal{L}^\star(H_n, Q_n^r) \\ = \;& \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n} \left(e^{\theta_k^r} - 1\right) \left(1 + \sum_{j=k+1}^{n} e^{\Theta_{j-1}^r - \Theta_k^r} \left(e^{\theta_j^r} - 1\right) \mathbf{P}[\widehat{\sigma} > j-k]\right), \end{aligned} \tag{A.4.4}$$

where $\theta_i^r = \theta_{n+1-i}$, $i = 1, 2, \ldots, n$ so that $\Theta_j^r \equiv \sum_{i=1}^{j} \theta_i^r = \Theta_n - \Theta_{n-j}$, $j = 1, \ldots, n$. Using the substitution $k' = n+1-j$ and $j' = n+1-k$ in (A.4.4) yields

$$\mathcal{L}^\star(H_n, Q_n^r)$$

$$= \lambda \mathbf{E}[\sigma] \sum_{j'=1}^{n} \left(e^{\theta^r_{n+1-j'}} - 1\right) \left(1 + \sum_{k'=1}^{j'-1} e^{\Theta^r_{n-k'} - \Theta^r_{n+1-j'}} \left(e^{\theta^r_{n+1-k'}} - 1\right) \mathbf{P}[\widehat{\sigma} > j' - k']\right)$$

$$= \lambda \mathbf{E}[\sigma] \sum_{j'=1}^{n} \left(e^{\theta_{j'}} - 1\right) \left(1 + \sum_{k'=1}^{j'-1} e^{\Theta_{j'-1} - \Theta_{k'}} \left(e^{\theta_{k'}} - 1\right) \mathbf{P}[\widehat{\sigma} > j' - k']\right)$$

$$= \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n} \left(e^{\theta_k} - 1\right) + \sum_{j=1}^{n} \sum_{k=1}^{j-1} e^{\Theta_{j-1} - \Theta_k} \left(e^{\theta_j} - 1\right) \left(e^{\theta_k} - 1\right) \mathbf{P}[\widehat{\sigma} > j - k].$$

Interchanging the order of summation and comparing with (A.4.1) establishes (2.2.8), thus concluding the proof. ■

A.5 Proof of Proposition 2.2.4

As before, we view the $M|G|\infty$ process $\{b_t,\ t = 0, 1, \ldots\}$ as a sum of two independent processes, $\{b_t^{(0)},\ t = 0, 1, \ldots\}$ and $\{b_t^{(a)},\ t = 0, 1, \ldots\}$, the former representing the contribution from the initial conditions, and the latter, that from the new arrivals.

Let Ω_I denote the underlying sample space on which the initial conditions $(b, \{\sigma_{0,i},\ i = 1, 2, \ldots\})$, and therefore the process $\{b_t^{(0)},\ t = 0, 1, \ldots\}$, are defined.

For any arbitrary realization of $\{b_t^{(a)},\ t = 0, 1, \ldots\}$, consider initial conditions ω_1 and ω_2 in Ω_I giving rise to two *distinct* realizations of the busy server process, denoted by $\{b_t^{(j)},\ t = 0, 1, \ldots\},\ j = 1, 2$.

For such a pair ω_1, ω_2 in Ω_I, there exists an integer $k(\omega_1, \omega_2)$ such that

$$b_{t+k}^{(1)} = b_{t+k}^{(2)}, \qquad k > k(\omega_1, \omega_2), \quad t = 0, 1, \ldots, \tag{A.5.1}$$

indicating that no matter what the initial conditions, the two processes do eventually couple.

Let P_{EQ} denote the probability measure on Ω_I, under which the rv b is Poisson with rate $\lambda\mathbf{E}[\sigma]$, and the rvs $\{\sigma_{0,i},\ i=1,2,\ldots\}$, independent of b, are *i.i.d.* with distribution $\widehat{G}$ given by (2.1.1).

From the comments preceding Proposition 2.2.4 it is clear that the initial conditions, when selected under probability measure P_{EQ}, give rise to an $M|G|\infty$ process satisfying

$$\{b_t,\ t=0,1,\ldots\} =_{st} \{b_t^\star,\ t=0,1,\ldots\},$$

where $\{b_t^\star,\ t=0,1,\ldots\}$ is the stationary and ergodic version of the busy server process. In other words, when the initial conditions are selected under probability measure P_{EQ}, we have

$$\{b_{t+k},\ t=0,1,\ldots\} \quad =_{st} \quad \{b_t^\star,\ t=0,1,\ldots\}, \tag{A.5.2}$$

for *each* $k=0,1,\ldots$.

Due to the point–wise equivalence evident in (A.5.1) for large values of k, it now follows that (A.5.2) holds under *all* other probability measures on Ω_I (assuming of course that the distribution $\widehat{G}$ is not defective, i.e. that $\mathbf{P}[\widehat{\sigma}=\infty]=0$). In other words, (A.5.2) holds as $k\to\infty$, irrespective of the initial conditions $(b,\{\sigma_{0,i},\ i=1,2,\ldots\})$, thus concluding our proof. ■

A.6 Proof of Proposition 2.3.2

We derive (2.3.5) from the log–moment generating function (2.2.3) of the process $\{b_t,\ t = 1, 2, \ldots\}$: For each $n = 1, 2, \ldots$, we have

$$\begin{aligned}\mathcal{L}^\star(T_n, Q_n) &= \ln \mathbf{E}\left[\exp\left(\sum_{i=1}^{n} \theta_i b_{t_i}\right)\right] \\ &= \lambda \mathbf{E}[\sigma]\left(\sum_{j=1}^{n}\sum_{k=1}^{j} e^{\Theta_j - \Theta_k}\left(e^{\theta_k} - 1\right)\mathbf{P}\left[t_j < \widehat{\sigma} + t_k \leq t_{j+1}\right]\right)\end{aligned}$$

for all T_n in $\mathcal{T}_n$ and Q_n in $\mathcal{Q}_n$, with the convention $t_{n+1} = \infty$.

Now, take $n = 2$ and denote the difference $t_2 - t_1$ by h. As is well known, the covariance $\Gamma(h)$ is given by

$$\begin{aligned}\Gamma(h) &= \left.\frac{\partial^2 \mathbf{E}\left[\exp\left(\theta_1 b_{t_1} + \theta_2 b_{t_2}\right)\right]}{\partial\theta_1 \partial\theta_2}\right|_{\theta_1 = \theta_2 = 0} - (\lambda \mathbf{E}[\sigma])^2 \\ &= \left.\frac{\partial^2 \exp\left(\mathcal{L}^\star(T_2, Q_2)\right)}{\partial\theta_1 \partial\theta_2}\right|_{\theta_1 = \theta_2 = 0} - (\lambda \mathbf{E}[\sigma])^2. \qquad \text{(A.6.1)}\end{aligned}$$

We note that

$$\begin{aligned}\mathcal{L}^\star(T_2, Q_2) &= \lambda \mathbf{E}[\sigma]\left(\sum_{j=1}^{2}\sum_{k=1}^{j} e^{\Theta_j - \Theta_k}\left(e^{\theta_k} - 1\right)\mathbf{P}\left[t_j < \widehat{\sigma} + t_k \leq t_{j+1}\right]\right) \\ &= \lambda \mathbf{E}[\sigma]\left(\left(e^{\theta_1} - 1\right)\left(\mathbf{P}[\widehat{\sigma} \leq h] + e^{\Theta_2 - \Theta_1}\mathbf{P}[\widehat{\sigma} > h]\right) + \left(e^{\theta_2} - 1\right)\right) \\ &= \lambda \mathbf{E}[\sigma]\left(\left(e^{\theta_1} - 1\right)\left(1 + \left(e^{\theta_2} - 1\right)\mathbf{P}[\widehat{\sigma} > h]\right) + \left(e^{\theta_2} - 1\right)\right).\end{aligned}$$

Define $C_i = e^{\theta_i} - 1,\ i = 1, 2$, in which case

$$\mathcal{L}^\star(T_2, Q_2) = \mathcal{L}^\star_C(h) = \lambda \mathbf{E}[\sigma]\left(C_1 + C_2 + C_1 C_2 \mathbf{P}[\widehat{\sigma} > h]\right). \qquad \text{(A.6.2)}$$

We use the relations

$$\begin{aligned}\frac{\partial \mathcal{L}^\star(T_2, Q_2)}{\partial\theta_1} &= \lambda \mathbf{E}[\sigma]\left(1 + C_2 \mathbf{P}[\widehat{\sigma} > h]\right) \cdot \frac{dC_1}{d\theta_1} \\ &= \lambda \mathbf{E}[\sigma]\left(1 + C_2 \mathbf{P}[\widehat{\sigma} > h]\right)(C_1 + 1),\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{L}^\star(T_2,Q_2)}{\partial \theta_2} &= \lambda \mathbf{E}[\sigma]\Big(1 + C_1 \mathbf{P}[\widehat{\sigma} > h]\Big) \cdot \frac{dC_2}{d\theta_2} \\ &= \lambda \mathbf{E}[\sigma]\Big(1 + C_1 \mathbf{P}[\widehat{\sigma} > h]\Big)(C_2+1),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial^2 \mathcal{L}^\star(T_2,Q_2)}{\partial \theta_1 \partial \theta_2} &= \lambda \mathbf{E}[\sigma]\,\mathbf{P}[\widehat{\sigma} > h] \cdot \frac{dC_1}{d\theta_1} \cdot \frac{dC_2}{d\theta_2} \\ &= \lambda \mathbf{E}[\sigma]\,\mathbf{P}[\widehat{\sigma} > h]\,(C_1+1)(C_2+1)\end{aligned}$$

to evaluate

$$\begin{aligned}&\frac{\partial^2}{\partial \theta_1 \partial \theta_2} \exp\Big(\mathcal{L}^\star(T_2,Q_2)\Big)\Big|_{\theta_1=0,\theta_2=0} \\ =\ & \exp\Big(\mathcal{L}^\star(T_2,Q_2)\Big) \cdot \left(\frac{\partial \mathcal{L}^\star(T_2,Q_2)}{\partial \theta_1} \cdot \frac{\partial \mathcal{L}^\star(T_2,Q_2)}{\partial \theta_2} + \frac{\partial^2 \mathcal{L}^\star(T_2,Q_2)}{\partial \theta_1 \partial \theta_2}\right)\Big|_{\theta_1=0,\theta_2=0} \\ =\ & (\lambda \mathbf{E}[\sigma])^2 \Big(1 + C_2 \mathbf{P}[\widehat{\sigma} > h]\Big)\Big(1 + C_1 \mathbf{P}[\widehat{\sigma} > h]\Big)(C_1+1)(C_2+1)\Big|_{C_1=0,C_2=0} \\ & + \lambda \mathbf{E}[\sigma]\,\mathbf{P}[\widehat{\sigma} > h] \cdot (C_1+1)(C_2+1)\Big|_{C_1=0,C_2=0} \\ =\ & (\lambda \mathbf{E}[\sigma])^2 + \lambda \mathbf{E}[\sigma]\,\mathbf{P}[\widehat{\sigma} > h],\end{aligned}$$

and the result now follows via (A.6.1).

Appendix B

B.1 Proof of Lemma 3.4.3

Fix γ in $\mathbb{R}$, and define

$$L_m(\gamma) \equiv \frac{1}{v_m} \ln \left(\sum_{n=m+1}^{\infty} e^{-\gamma v_n} \right), \quad m = 1, 2, \ldots$$

For $\gamma \leq 0$, we obviously have $L_m(\gamma) = \infty$ for all $m = 1, 2, \ldots$, and the desired conclusion $L(\gamma) = \infty$ trivially follows. For $\gamma > 0$, the proof proceeds along the two cases $\kappa > 0$ and $\kappa = 0$.

Case I $-$ $\kappa > 0$**:** Write $C = \kappa^{-1}$. For every $\varepsilon > 0$ with $0 < \varepsilon < C$, there exists a finite integer $n^\star = n^\star(\varepsilon)$ such that $(C - \varepsilon) \ln n < v_n < (C + \varepsilon) \ln n$ whenever $n \geq n^\star$. Therefore,

$$\left(\frac{\ln m}{v_m} \right) L_m^{+\varepsilon}(\gamma) \leq L_m(\gamma) \leq \left(\frac{\ln m}{v_m} \right) L_m^{-\varepsilon}(\gamma), \quad m \geq n^\star \tag{B.1.1}$$

where we have used the notation

$$\begin{aligned} L_m^{\pm\varepsilon}(\gamma) &\equiv \frac{1}{\ln m} \ln \left(\sum_{n=m+1}^{\infty} e^{-\gamma(C \pm \varepsilon) \ln n} \right) \\ &= \frac{1}{\ln m} \ln \left(\sum_{n=m+1}^{\infty} n^{-\gamma(C \pm \varepsilon)} \right), \quad m = 1, 2, \ldots \end{aligned} \tag{B.1.2}$$

If $\gamma > \kappa$ (or equivalently, $\gamma C > 1$), then $\gamma(C \pm \varepsilon) > 1$ for all $\varepsilon > 0$ small enough, in which case all the quantities (B.1.2) are finite with

$$\lim_{m\to\infty} L_m^{\pm\varepsilon}(\gamma) = 1 - \gamma(C \pm \varepsilon) \tag{B.1.3}$$

by standard arguments. Letting m go to infinity in (B.1.1) and using the limiting values (B.1.3), we get

$$\begin{aligned} \liminf_{m\to\infty} L_m(\gamma) &\geq \liminf_{m\to\infty} \left(\frac{\ln m}{v_m}\right) L_m^{+\varepsilon}(\gamma) \\ &= \kappa(1 - \gamma(C + \varepsilon)) \end{aligned}$$

and

$$\begin{aligned} \limsup_{m\to\infty} L_m(\gamma) &\leq \limsup_{m\to\infty} \left(\frac{\ln m}{v_m}\right) L_m^{-\varepsilon}(\gamma) \\ &= \kappa(1 - \gamma(C - \varepsilon)) \end{aligned}$$

and the desired conclusion follows because $\varepsilon > 0$ can be selected arbitrarily small.

If $\gamma < \kappa$ (or equivalently, $\gamma C < 1$), then $\gamma(C + \varepsilon) < 1$ for some $\varepsilon > 0$, in which case $L_m^{+\varepsilon}(\gamma) = \infty$ for all $m = 1, 2, \ldots$, yielding the required result via (B.1.1).

Case II – $\kappa = 0$**:** Noting that

$$L_m(\gamma) \geq \frac{1}{v_m} \ln e^{-\gamma v_m} = -\gamma, \quad m = 1, 2, \ldots$$

so that $L(\gamma) \geq -\gamma$, we see that the desired conclusion will follow if we can show the reversed inequality

$$L_m(\gamma) \leq -\gamma. \tag{B.1.4}$$

To do so, we write $\phi(m) \equiv v_m / \ln m$ for all $m = 1, 2, \ldots$. The fact $\kappa = 0$ translates into $\lim_{m\to\infty} \phi(m) = \infty$ with the sequence $\{\phi(m),\ m = 1, 2, \ldots\}$ being monotone increasing in the tail by virtue of **A1**. This implies that for m large enough, we have the inequality $\frac{1}{\gamma} < \phi(m) \le \phi(n)$ for all $n \ge m$, so that

$$\begin{aligned}\sum_{n=m+1}^{\infty} e^{-\gamma v_n} &= \sum_{n=m+1}^{\infty} n^{-\gamma\phi(n)} \\ &\le \sum_{n=m+1}^{\infty} n^{-\gamma\phi(m)} \\ &\le \int_m^{\infty} x^{-\gamma\phi(m)} dx \\ &= \frac{m^{-\gamma\phi(m)+1}}{\gamma\phi(m)-1}.\end{aligned}$$

In that range, we see that

$$\begin{aligned}L_m(\gamma) &\le \frac{1}{\phi(m)\ln m} \ln\left(\frac{m^{-\gamma\phi(m)+1}}{\gamma\phi(m)-1}\right) \\ &= \frac{-\gamma\phi(m)+1}{\phi(m)} - \frac{\ln(\gamma\phi(m)-1)}{\phi(m)\ln m}\end{aligned}$$

and taking the limit as m goes to infinity, we finally obtain (B.1.4).

Appendix C

C.1 Proof of Lemma 4.2.1

Fix $n = 1, 2, \ldots$. Applying Lemma A.4.1 for the particular case $H_n = (1, 2, \ldots, n)$ and $\widetilde{Q}_n = \{\theta_i = \theta,\ i = 1, 2, \ldots, n\}$, we get

$$\begin{aligned}
\mathcal{L}(H_n, \widetilde{Q}_n) &= \lambda \mathbf{E}[\sigma] \sum_{k=1}^{n} \left(e^{\theta_k} - 1\right) \left(1 + \sum_{j=k+1}^{n} e^{\Theta_{j-1} - \Theta_k} \left(e^{\theta_j} - 1\right) \mathbf{P}[\widehat{\sigma} > j - k]\right) \\
&= \lambda \mathbf{E}[\sigma] \left(e^{\theta} - 1\right) \sum_{k=1}^{n} 1 + \sum_{j=k+1}^{n} e^{\theta(j-1-k)} \left(e^{\theta} - 1\right) \mathbf{P}[\widehat{\sigma} > j - k] \\
&= \lambda \mathbf{E}[\sigma] \left(e^{\theta} - 1\right) \left(n + \sum_{k=1}^{n} \sum_{j=k+1}^{n} e^{\theta(j-1-k)} \left(e^{\theta} - 1\right) \mathbf{P}[\widehat{\sigma} > j - k]\right).
\end{aligned}$$

Simplifying the double summation above, we get

$$\begin{aligned}
\sum_{k=1}^{n} \sum_{j=k+1}^{n} e^{\theta(j-1-k)} \left(e^{\theta} - 1\right) \mathbf{P}[\widehat{\sigma} > j - k] &= \sum_{k=1}^{n} \sum_{r=1}^{n-k} e^{\theta r} \left(1 - e^{-\theta}\right) \mathbf{P}[\widehat{\sigma} > r] \\
&= \left(1 - e^{-\theta}\right) \sum_{r=1}^{n-1} \sum_{k=1}^{n-r} e^{\theta r} \mathbf{P}[\widehat{\sigma} > r] \\
&= \left(1 - e^{-\theta}\right) \sum_{r=1}^{n} (n - r) e^{\theta r} \mathbf{P}[\widehat{\sigma} > r],
\end{aligned}$$

and the proof is now completed. ■

C.2 Proof of Theorem 4.2.2

The tail probability $\mathbf{P}[\widehat{\sigma} > r]$ can be expressed in terms of the distribution G through the relation

$$\mathbf{E}[\sigma]\,\mathbf{P}[\widehat{\sigma} > r] \;=\; \mathbf{E}[(\sigma - r)^+] \;=\; \sum_{j=r+1}^{\infty} g_j (j - r), \qquad r = 1, 2, \ldots, \tag{C.2.1}$$

which was derived in the proving of Lemma 2.3.1. Fixing $n = 1, 2, \ldots,$ and θ in $\mathbb{R}$, we may now rewrite definition (4.2.4) as

$$\begin{aligned}
\Delta(n,\theta) &= \frac{1}{n}\sum_{r=1}^{n-1}(n-r)e^{r\theta}\mathbf{P}[\widehat{\sigma} > r] \\
&= \frac{1}{n\mathbf{E}[\sigma]}\sum_{r=1}^{n-1}\sum_{j=r+1}^{\infty} g_j (n-r)(j-r)e^{r\theta} \\
&= \frac{1}{n\mathbf{E}[\sigma]}\sum_{j=2}^{\infty} g_j \sum_{r=1}^{\min(j,n)-1} (n-r)(j-r)e^{r\theta} \\
&= \frac{1}{n\mathbf{E}[\sigma]}\sum_{j=1}^{\infty} g_j \sum_{r=1}^{\min(j,n)} (n-r)(j-r)e^{r\theta} \\
&= \frac{1}{n\mathbf{E}[\sigma]}\mathbf{E}\left[\sum_{r=1}^{\min(\sigma,n)} (n-r)(\sigma-r)e^{r\theta}\right].
\end{aligned} \tag{C.2.2}$$

Further simplification of the sum in (C.2.2) gives

$$\begin{aligned}
&\sum_{r=1}^{\min(\sigma,n)} (n-r)(\sigma-r)e^{r\theta} \\
&= \sum_{r=1}^{\min(\sigma,n)} \left(n\sigma - r(n+\sigma) + r^2\right) e^{r\theta} \\
&= \sum_{r=1}^{\min(\sigma,n)} \left[\left(r - \frac{n+\sigma}{2}\right)^2 - \left(\frac{n+\sigma}{2}\right)^2 + n\sigma\right] e^{r\theta} \\
&= \sum_{r=1}^{\min(\sigma,n)} \left(r - \frac{n+\sigma}{2}\right)^2 e^{r\theta} \;-\; \left(\frac{n-\sigma}{2}\right)^2 \sum_{r=1}^{\min(\sigma,n)} e^{r\theta}
\end{aligned}$$

$$= \quad e^{\left(\frac{n+\sigma}{2}\right)\theta} \sum_{q=\left|\frac{n-\sigma}{2}\right|}^{\frac{n+\sigma}{2}-1} q^2 e^{-q\theta} \;-\; \left(\frac{n-\sigma}{2}\right)^2 e^{\theta} \left(\frac{e^{\theta \min(\sigma,n)} - 1}{e^{\theta}-1}\right) \tag{C.2.3}$$

where we have set $q = \frac{n+\sigma}{2} - r$. Using the standard expansion

$$\sum_{r=a}^{\infty} r^2 e^{-r\theta} \;=\; \frac{e^{-\theta a}}{1-e^{-\theta}} \left(a^2 + \frac{(2a+1)\,e^{-\theta}}{(1-e^{-\theta})} + \frac{2e^{-2\theta}}{(1-e^{-\theta})^2}\right), \quad a = 1, 2, \ldots,$$

we find

$$\begin{aligned}
\sum_{q=\left|\frac{n-\sigma}{2}\right|}^{\frac{n+\sigma}{2}-1} q^2 e^{-q\theta} \;&=\; \sum_{q=\left|\frac{n-\sigma}{2}\right|}^{\infty} q^2 e^{-q\theta} - \sum_{q=\frac{n+\sigma}{2}}^{\infty} q^2 e^{-q\theta} \\
&=\; \frac{e^{-\theta\left|\frac{n-\sigma}{2}\right|}}{1-e^{-\theta}} \left[\left(\frac{n-\sigma}{2}\right)^2 + \frac{(|n-\sigma|+1)\,e^{-\theta}}{(1-e^{-\theta})} + \frac{2e^{-2\theta}}{(1-e^{-\theta})^2}\right] \\
&\quad - \frac{e^{-\theta\left(\frac{n+\sigma}{2}\right)}}{1-e^{-\theta}} \left[\left(\frac{n+\sigma}{2}\right)^2 + \frac{(n+\sigma+1)\,e^{-\theta}}{(1-e^{-\theta})} + \frac{2e^{-2\theta}}{(1-e^{-\theta})^2}\right],
\end{aligned}$$

which, when reported into (C.2.3) yields

$$\begin{aligned}
&\left(1-e^{-\theta}\right) \sum_{r=1}^{\min(\sigma,n)} (n-r)(\sigma-r)e^{r\theta} \\
=\;& e^{\theta \min(\sigma,n)} \left[\left(\frac{n-\sigma}{2}\right)^2 + \frac{|n-\sigma|+1}{e^{\theta}-1} + \frac{2}{(e^{\theta}-1)^2}\right] \\
&\quad - \left(\frac{n+\sigma}{2}\right)^2 - \frac{n+\sigma+1}{e^{\theta}-1} - \frac{2}{(e^{\theta}-1)^2} \\
&\quad - \left(\frac{n-\sigma}{2}\right)^2 \left(e^{\theta \min(\sigma,n)} - 1\right) \\
=\;& e^{\theta \min(\sigma,n)} \left(\frac{|n-\sigma|+1}{e^{\theta}-1} + \frac{2}{(e^{\theta}-1)^2}\right) - n\sigma - \frac{n+\sigma+1}{e^{\theta}-1} - \frac{2}{(e^{\theta}-1)^2}.
\end{aligned}$$

Incorporating this final step into (C.2.2), we get

$$\begin{aligned}
\Delta(n,\theta) \;=\; \frac{1}{n\mathbf{E}[\sigma]\left(1-e^{-\theta}\right)} \cdot \mathbf{E}\Bigg[& e^{\theta \min(\sigma,n)} \left(\frac{|n-\sigma|+1}{e^{\theta}-1} + \frac{2}{(e^{\theta}-1)^2}\right) \\
& - n\sigma - \frac{n+\sigma+1}{e^{\theta}-1} - \frac{2}{(e^{\theta}-1)^2}\Bigg],
\end{aligned}$$

or put differently,

$$
\begin{aligned}
& n\mathbf{E}[\sigma]\left(e^{\theta}-1\right)\left(1+\left(1-e^{-\theta}\right)\Delta(n,\theta)\right) \\
= \; & \mathbf{E}\left[e^{\theta\min(\sigma,n)}\left(|n-\sigma|+1+\frac{2}{e^{\theta}-1}\right)- (n+\sigma+1)-\frac{2}{e^{\theta}-1}\right] \\
= \; & \mathbf{E}\left[\left(n+\sigma+1+\frac{2}{e^{\theta}-1}-2\min(n,\sigma)\right)\left(e^{\theta\min(n,\sigma)}-1\right)-2\min(n,\sigma)\right] \\
= \; & \mathbf{E}\left[\left((n-\sigma)^{+}+(\sigma-n)^{+}+\frac{e^{\theta}+1}{e^{\theta}-1}\right)\left(e^{\theta\min(n,\sigma)}-1\right)\right]-2\mathbf{E}\left[\min(n,\sigma)\right]
\end{aligned}
\tag{C.2.4}
$$

upon using the identities

$$
\sigma=(\sigma-n)^{+}+\min(n,\sigma) \qquad \text{and} \qquad n=(n-\sigma)^{+}+\min(n,\sigma).
$$

The result follows on comparing the left–hand side of (C.2.4) with (4.2.3), and upon noting that

$$
\mathbf{E}\left[(n-\sigma)^{+}\left(e^{\theta\min(n,\sigma)}-1\right)\right] \;=\; \mathbf{E}\left[(n-\sigma)^{+}\left(e^{\theta\sigma}-1\right)\right],
$$

and that

$$
\begin{aligned}
\mathbf{E}\left[(\sigma-n)^{+}\left(e^{\theta\min(n,\sigma)}-1\right)\right] &= \left(e^{\theta n}-1\right)\mathbf{E}\left[(\sigma-n)^{+}\right] \\
&= \left(e^{\theta n}-1\right)\mathbf{E}[\sigma]\,\mathbf{P}\left[\widehat{\sigma}>n\right],
\end{aligned}
$$

where the final step ensues from (C.2.1).

C.3 Proof of Lemmas 4.5.4 and 4.5.5

In the interest of clarity, we discuss only the case when the sequence $\{v_n^\star/n,\ n = 1,2,\ldots\}$ is monotone decreasing, and leave it to the reader to extend the arguments to the asymptotically monotone case, an easy but tedious exercise. Moreover, for each $n = 1, 2, \ldots$,

$$0 \le \Delta\left(n, \theta\frac{v_n^\star}{n}\right) \le \sum_{r=1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star}, \qquad \theta \in \mathbb{R},$$

so that we need only establish

$$\lim_{n\to\infty} \frac{v_n^\star}{n}\sum_{r=1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} \; = \; 0,$$

in the range $0 < \theta < 1$.

A proof of Lemma 4.5.4: Fixing θ in the interval $(0, 1)$ and $n = 1, 2, \ldots$, we note that

$$r\theta\frac{v_n^\star}{n} - v_r^\star = r\left(\theta\frac{v_n^\star}{n} - \frac{v_r^\star}{r}\right) \le r\left(\theta\frac{v_r^\star}{r} - \frac{v_r^\star}{r}\right), \qquad r = 1, 2, \ldots, n$$

so that

$$0 \le \frac{v_n^\star}{n}\sum_{r=1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} \le \frac{v_n^\star}{n}\sum_{r=1}^{n} e^{-(1-\theta)v_r^\star},$$

and the conclusion immediately follows from the finiteness assumption (4.5.4) and the fact that $R = 0$. ■

A proof of Lemma 4.5.5: This time, with θ in the interval $(0, 1)$ and $n = 1, 2, \ldots$, we begin with the decomposition

$$\frac{v_n^\star}{n}\sum_{r=1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} = \frac{v_n^\star}{n}\sum_{r=1}^{Z(n)} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} + \frac{v_n^\star}{n}\sum_{r=Z(n)+1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} \qquad \text{(C.3.1)}$$

where $Z(n)$ is as described by Assumption **C2b**. The analysis successively considers the two terms in this last expression.

We first discuss the second term of (C.3.1): From the monotonicity of the sequence $\{v_n^\star/n,\ n=1,2,\ldots\}$, we get

$$\theta\frac{v_n^\star}{n}r - v_r^\star = (\theta-1)\frac{v_n^\star}{n}r - \left(\frac{v_r^\star}{r} - \frac{v_n^\star}{n}\right) r \le (\theta-1)\frac{v_n^\star}{n}r, \qquad r=1,\ldots,n$$

and it is now plain that

$$\begin{aligned}
\frac{v_n^\star}{n}\sum_{r=Z(n)+1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} &\le \frac{v_n^\star}{n}\sum_{r=Z(n)+1}^{n} e^{(\theta-1)\frac{v_n^\star}{n}r} \\
&= \frac{v_n^\star}{n}\left(\frac{e^{(\theta-1)v_n^\star\frac{(n+1)}{n}} - e^{(\theta-1)\frac{v_n^\star}{n}(Z(n)+1)}}{e^{(\theta-1)\frac{v_n^\star}{n}} - 1}\right) \\
&= \left(\frac{e^{(\theta-1)\frac{v_n^\star}{n}} - 1}{\frac{v_n^\star}{n}}\right)^{-1}\left(e^{(\theta-1)v_n^\star\frac{(n+1)}{n}} - e^{(\theta-1)\frac{v_n^\star}{n}(Z(n)+1)}\right).
\end{aligned}$$

Using Assumption **C2b**, we readily conclude

$$\lim_{n\to\infty}\frac{v_n^\star}{n}\sum_{r=Z(n)+1}^{n} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} = 0. \tag{C.3.2}$$

Next, going back to the first term of (C.3.1), we note for $0<\theta<1$ that

$$\theta\frac{v_n^\star}{n}r - v_r^\star \le \theta\frac{v_r^\star}{r}r - v_r^\star \le (\theta-1)\frac{v_{Z(n)}^\star}{Z(n)}r, \qquad r=1,\ldots,Z(n)$$

by the monotonicity of the sequence $\{v_n^\star/n,\ n=1,2,\ldots\}$. Therefore,

$$\begin{aligned}
\frac{v_n^\star}{n}\sum_{r=1}^{Z(n)} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} &\le \frac{v_n^\star}{n}\sum_{r=1}^{Z(n)} e^{(\theta-1)\frac{v_{Z(n)}^\star}{Z(n)}r} \\
&= \frac{v_n^\star}{n}e^{(\theta-1)\frac{v_{Z(n)}^\star}{Z(n)}}\left(\frac{e^{(\theta-1)v_{Z(n)}^\star} - 1}{e^{(\theta-1)\frac{v_{Z(n)}^\star}{Z(n)}} - 1}\right) \\
&= e^{(\theta-1)\frac{v_{Z(n)}^\star}{Z(n)}}\left(\frac{e^{(\theta-1)\frac{v_{Z(n)}^\star}{Z(n)}} - 1}{\frac{v_{Z(n)}^\star}{Z(n)}}\right)^{-1}\frac{v_n^\star}{n}\frac{Z(n)}{v_{Z(n)}^\star}\left(e^{(\theta-1)v_{Z(n)}^\star} - 1\right).
\end{aligned}$$

Again by Assumption **C2b** we have

$$\lim_{n\to\infty}\frac{v_n^\star}{n}\sum_{r=1}^{Z(n)} e^{\theta\frac{v_n^\star}{n}r - v_r^\star} = 0. \tag{C.3.3}$$

Combining (C.3.1), (C.3.2) and (C.3.3) readily gives the result. ■

Appendix D

D.1 Proof of Proposition 5.2.2

Proposition 5.2.2, (i) follows through the discussion prior to the statement of Proposition 5.2.1. Inequalities (5.2.5) and (5.2.6) come about via Proposition 3.4.3 and Theorem 4.6.1, with

$$\begin{aligned}
\gamma^{\star(\widehat{a},\widehat{v})} &= \sup_{y>0}\ \min\Bigg(\sup_{\theta>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(\widehat{a},\widehat{v})}(x\widehat{a}_n)}(\theta x-\Lambda^{(\widehat{a},\widehat{v})}(\theta))\right), \\
&\qquad \Lambda^{\star(\widehat{a},\widehat{v})}(0)g^{(\widehat{a},\widehat{v})}(y)\Bigg)-\kappa^{(\widehat{a},\widehat{v})}g^{(\widehat{a},\widehat{v})}(y) \\
&= \sup_{y>0}\ \min\Bigg(\sup_{\theta>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(\widehat{a},\widehat{v})}(x\widehat{a}_n)}\left(\theta x-\frac{1}{\widehat{K}}\Lambda^{(a,v)}\left(\widehat{K}\ \frac{\widehat{\alpha}}{\alpha}\ \theta\right)\right)\right), \\
&\qquad \frac{1}{\widehat{K}}\Lambda^{\star(a,v)}(0)g^{(\widehat{a},\widehat{v})}(y)\Bigg)-\kappa^{(\widehat{a},\widehat{v})}g^{(\widehat{a},\widehat{v})}(y).
\end{aligned}$$

Using (5.2.4) and the substitution $\theta'=\widehat{K}\ \dfrac{\widehat{\alpha}}{\alpha}\ \theta$, gives

$$\begin{aligned}
\gamma^{\star(\widehat{a},\widehat{v})} &= \frac{1}{\widehat{K}}\sup_{y>0}\ \min\Bigg(\sup_{\theta'>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(\widehat{a},\widehat{v})}(x\widehat{a}_n)}\left(\theta'\ \frac{\alpha}{\widehat{\alpha}}\ x-\Lambda^{(a,v)}(\theta')\right)\right), \\
&\qquad \Lambda^{\star(a,v)}(0)g^{(\widehat{a},\widehat{v})}(y)\Bigg)-\kappa^{(a,v)}g^{(\widehat{a},\widehat{v})}(y) \\
&= \frac{1}{\widehat{K}}\sup_{y>0}\ \min\Bigg(\sup_{\theta>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(a,v)}(x\widehat{a}_n)}\left(\theta\ \frac{\alpha}{\widehat{\alpha}}\ x-\Lambda^{(a,v)}(\theta)\right)\right), \\
&\qquad \Lambda^{\star(a,v)}(0)\widehat{K}g^{(a,v)}\left(\frac{\alpha}{\widehat{\alpha}}y\right)\Bigg)-\kappa^{(a,v)}\widehat{K}g^{(a,v)}\left(\frac{\alpha}{\widehat{\alpha}}y\right)
\end{aligned}$$

by Lemma 5.1.3. Setting $x' = \frac{\alpha}{\widehat{\alpha}}\, x$ and $y' = \frac{\alpha}{\widehat{\alpha}}\, y$, we have

$$\begin{aligned}\gamma^{\star(\widehat{a},\widehat{v})} \;=\; \sup_{y'>0} \min\Bigg(& \sup_{\theta>0} \liminf_{n\to\infty} \inf_{x'>y'} \left(\frac{1}{\widehat{K}} \frac{\widehat{v}_n}{h^{(a,v)}\left(x'\frac{\widehat{\alpha}}{\alpha}\widehat{a}_n\right)} \left(\theta x' - \Lambda^{(a,v)}(\theta)\right) \right), \\ & \Lambda^{\star(a,v)}(0) g^{(a,v)}(y') \Bigg) - \kappa^{(a,v)} g^{(a,v)}(y') \qquad \text{(D.1.1)}\end{aligned}$$

By (4.6.5), for any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that

$$1-\varepsilon \;\leq\; \frac{\widehat{\alpha}\widehat{a}_n}{\alpha a_n} \;\leq\; 1+\varepsilon, \quad n > N(\varepsilon),$$

and the monotone increasing nature of h yields

$$h(xa_n(1-\varepsilon)) \;\leq\; h\left(xa_n \frac{\widehat{\alpha}}{\alpha}\frac{\widehat{a}_n}{a_n}\right) \;\leq\; h(xa_n(1+\varepsilon)), \qquad n > N(\varepsilon),$$

thereby implying

$$\frac{\widehat{v}_n}{h^{(a,v)}\left(x\frac{\widehat{\alpha}}{\alpha}\widehat{a}_n\right)} \;\leq\; \frac{\widehat{v}_n}{h^{(a,v)}(xa_n(1-\varepsilon))}, \qquad n > N(\varepsilon).$$

Now fix $\theta > 0$. Setting $x'' = x(1-\varepsilon)$, we have

$$\begin{aligned} & \inf_{x>y} \left(\frac{\widehat{v}_n}{h^{(a,v)}\left(x\frac{\widehat{\alpha}}{\alpha}\widehat{a}_n\right)} \left(\theta x - \Lambda^{(a,v)}(\theta)\right) \right) \\ \leq\; & \inf_{x''>y(1-\varepsilon)} \left(\frac{\widehat{v}_n}{h^{(a,v)}(x''a_n)} \left(\frac{\theta}{(1-\varepsilon)} x'' - \Lambda^{(a,v)}(\theta) \right) \right) \\ \leq\; & \inf_{x>y} \left(\frac{\widehat{v}_n}{h^{(a,v)}(xa_n)} \left(\frac{\theta}{(1-\varepsilon)} x - \Lambda^{(a,v)}(\theta) \right) \right) \\ \leq\; & \inf_{x>y} \left(\frac{\widehat{v}_n}{h^{(a,v)}(xa_n)} \left(\theta x - \Lambda^{(a,v)}(\theta)\right) \right), \qquad n > N(\varepsilon). \qquad \text{(D.1.2)} \end{aligned}$$

A lower bound can be similarly derived and is given by

$$\begin{aligned} & \inf_{x>y} \left(\frac{\widehat{v}_n}{h^{(a,v)}\left(x\frac{\widehat{\alpha}}{\alpha}\widehat{a}_n\right)} \left(\theta x - \Lambda^{(a,v)}(\theta)\right) \right) \\ \geq\; & \inf_{x>y} \left(\frac{\widehat{v}_n}{h^{(a,v)}(xa_n)} \left(\frac{\theta}{(1+\varepsilon)} x - \Lambda^{(a,v)}(\theta) \right) \right), \qquad n > N(\varepsilon). \quad \text{(D.1.3)} \end{aligned}$$

Combining (D.1.2) and (D.1.3) we have

$$\begin{aligned}
&\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(a,v)}(xa_n)}\left(\frac{\theta}{(1+\varepsilon)}x-\Lambda^{(a,v)}(\theta)\right)\right)\\
\le\ &\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(a,v)}\left(x\frac{\widehat{\alpha}}{\alpha}\widehat{a}_n\right)}\left(\theta x-\Lambda^{(a,v)}(\theta)\right)\right)\\
\le\ &\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(a,v)}(xa_n)}\left(\theta x-\Lambda^{(a,v)}(\theta)\right)\right).
\end{aligned}$$

This being true for every $\varepsilon > 0$, we conclude that

$$\begin{aligned}
&\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(a,v)}\left(x\frac{\widehat{\alpha}}{\alpha}\widehat{a}_n\right)}\left(\theta x-\Lambda^{(a,v)}(\theta)\right)\right)\\
=\ &\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{\widehat{v}_n}{h^{(a,v)}(xa_n)}\left(\theta x-\Lambda^{(a,v)}(\theta)\right)\right)\\
=\ &\widehat{K}\liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{v_n}{h^{(a,v)}(xa_n)}\left(\theta x-\Lambda^{(a,v)}(\theta)\right)\right).
\end{aligned}$$

Reporting this conclusion back into (D.1.1), gives

$$\begin{aligned}
\gamma^{\star(\widehat{a},\widehat{v})} &= \sup_{y>0}\ \min\Bigg(\sup_{\theta>0}\ \liminf_{n\to\infty}\ \inf_{x>y}\left(\frac{v_n}{h^{(a,v)}\left(xa_n\right)}\left(\theta x-\Lambda^{(a,v)}(\theta)\right)\right),\\
&\qquad\qquad \Lambda^{\star(a,v)}(0)g^{(a,v)}(y)\Bigg)-\kappa^{(a,v)}g^{(a,v)}(y)\\
&= \gamma^{\star(a,v)},
\end{aligned}$$

and the proof is complete. ■

D.2 Proof of Lemma 5.3.2

From the definitions (2.1.1) and (2.3.1), we have

$$\begin{aligned}
v_n^\star &= -\ln\left(\sum_{i=n}^{\infty}\frac{\mathbf{P}\left[\sigma>i\right]}{\mathbf{E}\left[\sigma\right]}\right)\\
&= \ln\mathbf{E}\left[\sigma\right]-\ln\left(\sum_{i=n}^{\infty}e^{-w_i^\star}\right),\qquad n=1,2,\ldots,
\end{aligned}$$

whence

$$\limsup_{n\to\infty} \frac{v_n^\star}{w_n^\star} \;=\; -\limsup_{n\to\infty} \frac{\ln\left(\sum_{i=n}^{\infty} e^{-w_i^\star}\right)}{w_n^\star}, \tag{D.2.1}$$

and

$$\liminf_{n\to\infty} \frac{v_n^\star}{w_n^\star} \;=\; -\liminf_{n\to\infty} \frac{\ln\left(\sum_{i=n}^{\infty} e^{-w_i^\star}\right)}{w_n^\star}. \tag{D.2.2}$$

By virtue of the simple lower bound

$$e^{-w_n^\star} \;\leq\; \sum_{i=n}^{\infty} e^{-w_i^\star}, \qquad n = 1, 2, \ldots,$$

we immediately infer that

$$\limsup_{n\to\infty} \frac{v_n^\star}{w_n^\star} \;\leq\; 1, \tag{D.2.3}$$

irrespective of the value taken by W.

Case I: $W = 0$.

Assuming that the sequence $\frac{\ln n}{w_n^\star}$ is asymptotically monotone decreasing, we find that

$$\begin{aligned}
\sum_{i=n}^{\infty} e^{-w_i^\star} &= \sum_{i=n}^{\infty} e^{-\left(\frac{w_i^\star}{\ln i}\right)\ln i} \\
&\leq \sum_{i=n}^{\infty} e^{-\left(\frac{w_n^\star}{\ln n}\right)\ln i} \\
&\leq \int_{i=n-1}^{\infty} x^{-\left(\frac{w_n^\star}{\ln n}\right)} dx \\
&= \frac{(n-1)^{1-\left(\frac{w_n^\star}{\ln n}\right)}}{\frac{w_n^\star}{\ln n} - 1}
\end{aligned}$$

for n large enough. Therefore

$$\liminf_{n\to\infty} \frac{v_n^\star}{w_n^\star} \;\geq\; \liminf_{n\to\infty} \left[\frac{\left(\frac{w_n^\star}{\ln n} - 1\right)\ln(n-1)}{w_n^\star} + \frac{\ln\left(\frac{w_n^\star}{\ln n} - 1\right)}{w_n^\star} \right]$$

$$
\begin{aligned}
&= \liminf_{n\to\infty} \left[\frac{\ln(n-1)}{\ln n} - \frac{\ln(n-1)}{w_n^\star} + \frac{\ln\left(\frac{w_n^\star}{\ln n} - 1\right)}{\left(\frac{w_n^\star}{\ln n}\right)\ln n} \right] \\
&= 1,
\end{aligned}
$$

which, in combination with (D.2.3), yields the required result.

Case II: $0 < W < 1$.

By (5.3.2), we know that for any $\delta > 0$, there exists an integer $N(\delta)$ so that

$$
\frac{1}{W} - \delta \le \frac{w_n^\star}{\ln n} \le \frac{1}{W} + \delta, \qquad n > N(\delta). \tag{D.2.4}
$$

Therefore, for $n > N(\delta)$ we have

$$
\begin{aligned}
\sum_{i=n}^{\infty} e^{-w_i^\star} &\le \sum_{i=n}^{\infty} i^{-\left(\frac{1}{W}-\delta\right)} \\
&\le \int_{n-1}^{\infty} x^{-\left(\frac{1}{W}-\delta\right)} dx \\
&= \frac{(n-1)^{1-\left(\frac{1}{W}-\delta\right)}}{\frac{1}{W}-\delta-1},
\end{aligned}
$$

and referring to (D.2.2) we conclude that

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{v_n^\star}{w_n^\star} &\ge \left(\frac{1}{W} - \delta - 1\right) \liminf_{n\to\infty} \frac{\ln(n-1)}{w_n^\star} + \liminf_{n\to\infty} \frac{\ln\left(\frac{1}{W} - \delta - 1\right)}{w_n^\star} \\
&= W\left(\frac{1}{W} - \delta - 1\right).
\end{aligned} \tag{D.2.5}
$$

The corresponding upper bound, similarly derived, is given by

$$
\limsup_{n\to\infty} \frac{v_n^\star}{w_n^\star} \le W\left(\frac{1}{W} + \delta - 1\right). \tag{D.2.6}
$$

The desired result now follows through the observation that the bounds (D.2.5) and (D.2.6) hold for any $\delta > 0$.

■

BIBLIOGRAPHY

[1] Abramowitz and I. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, J. Wiley & Sons, New York (NY), 1972.

[2] A. Adas, "Traffic models in broadband networks," *IEEE Communications Magazine* **35** (1997), pp. 82–89.

[3] R.G. Addie, M. Zukerman and T. Neame, "Fractal traffic: Measurements, modeling and performance evaluation," *Proceedings of INFOCOM '95*, Boston (MA), April 1995, pp. 985–992.

[4] V. Anantharam, "On the sojourn time of sessions at an ATM buffer with long-range dependent input traffic," *Proceedings of the IEEE Conference on Decision and Control*, December 1995, pp. 859–864.

[5] F. Baccelli and P. Bremaud, *Elements of Queueing Theory: Palm–Martingale Calculus and Stochastic Recurrences*, Applications of Mathematics **26**, Springer–Verlag, Berlin Heidelberger, 1994.

[6] J. Beran, *Statistics for Long-Memory Processes*, Chapman and Hall, New York (NY), 1994.

[7] J. Beran, R. Sherman, M. S. Taqqu and W. Willinger, "Long-range dependence in variable bit–rate video traffic," *IEEE Transactions on Communications* **COM–43** (1995), pp. 1566–1579.

[8] F. Brichet, J. W. Roberts, A. Simonian and D. Veitch, "Heavy traffic analysis of a storage model with long–range dependent On/Off sources," *Queueing Systems: Theory and Applications* **23** (1996), pp. 197–215.

[9] V. P. Chistakov, "A theorem on sums of independent positive random variables and its application to branching random processes," *Theory of Probability and its Applications* **9** (1964), pp. 640-648.

[10] G.L. Choudhury, D.M. Lucantoni and W. Whitt, "Squeezing the most out of ATM," *IEEE Transactions on Communications* **COM–44** (1996), pp. 203–217.

[11] D. R. Cox, "Long–Range Dependence: A Review," *Statistics: An Appraisal*, H. A. David and H. T. David, Eds., The Iowa State University Press, Ames (IA), 1984, pp. 55–74.

[12] D. R. Cox and V. Isham, *Point Processes*, Chapman and Hall, New York (NY), 1980.

[13] M.E. Crovella and A. Bestavros, "Self–similarity in World Wide Web traffic: Evidence and possible causes," *Performance Evaluation Review* **24** (1996), Proceedings of the ACM SIGMETRICS 96 Conference, Philadelphia (PA), May 1996, pp. 160–169.

[14] T. Daniels and C. Blondia, "Asymptotic Behavior of a Discrete–Time Queue with Long Range Dependent Input," *Proceedings of INFOCOM '99*, New York (NY), March 1999, pp. 633-640.

[15] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, Jones and Bartlett, Boston (MA), 1993.

[16] 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), pp. 363–374.

[17] N. G. Duffield, "On the relevance of long–tailed durations for the statistical multiplexing of large aggregations," in *Proceedings of the 34th Annual Allerton Conference on Communications, Control and Computing*, October 1996.

[18] A. Elwalid, D. Heyman, T. V. Lakshman, D. Mitra and A. Weiss, "Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing,"*IEEE Journal on Selected Areas in Communications* **JSAC–13** (1995), pp. 1004–1016.

[19] A. Erramilli, R. P. Singh and P. Pruthi, "An Application of Chaotic Maps to Model Packet Traffic," *Queueing Systems: Theory and Applications* **20** (1995), pp. 171–206.

[20] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long range dependent traffic," *IEEE/ACM Transactions on Networking* **4** (1996), pp. 209-223.

[21] P. Embrechts, C. Klüppelberg and T. Mikosch, *Modeling Extremal Events*, Springer-Verlag, New York (NY), 1997.

[22] J.D. Esary, F. Proschan, and D.W. Walkup, "Association of random variables, with applications," *Annals of Mathematical Statistics* **38** (1967), pp. 166–1474.

[23] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II*, J. Wiley & Sons, New York (NY), 1966.

[24] H. J. Fowler and W. E. Leland, "Local area network traffic characteristics, with implications for broadband network congestion management," *IEEE Journal on Selected Areas in Communications* **JSAC–9** (1991), pp. 1139–1149.

[25] V. S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Communications Magazine* **32** (1994), pp. 70–81.

[26] M. Garrett and W. Willinger, "Analysis, modeling and generation of self–similar VBR video traffic," *Proceedings of SIGCOMM '94*, September 1994, pp. 269–280.

[27] P.W. Glynn and W. Whitt, "Logarithmic asymptotics for steady–state tail probabilities in a single–server queue," *Journal of Applied Probability* **31** (1994), pp. 131–159.

[28] M. Grossglauser and J. Bolot, "On the relevance of long–range dependence in network traffic," *Computer Communications Review*, **26 (4)**, October 1996, pp. 15–24.

[29] R. Guérin, H. Ahmadi and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high–speed networks," *IEEE Journal on Selected Areas in Communications* **JSAC–9** (1991), pp. 968–981.

[30] J. R. M. Hosking, "Fractional differencing," *Biometrika* **68(1)** (1981), pp. 165–176.

[31] J. R. M. Hosking, "Modeling persistence in hydrological time series using fractional differencing," *Water Resources Research* **20**(1984), pp. 1898–1908.

[32] C. L. Hwang and S. Q. Li, "On input state space reductions and buffer non–effective region," in *Proceedings of INFOCOM '94*, Toronto, Canada, June 1994, pp. 1018–1028.

[33] P.R. Jelenković and A.A. Lazar, "Asymptotic results for multiplexing sub–exponential on-off processes," *Self–similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds., J. Wiley & Sons, New York (NY), 1999, to appear.

[34] P.R. Jelenković, A.A. Lazar and N. Semret, "Multiple time scales and sub–exponentiality in MPEG video streams," in *Broadband Communications: International IFIP–IEEE Conference on Broadband Communications*, Montreal (PQ), April 1996, pp. 64–75.

[35] P.R. Jelenković and A.A. Lazar, "Multiplexing on–off sources with sub–exponential on periods: Part II," in *Proceedings of the 15th International Teletraffic Congress*, Washington, D.C., June 1997, pp. 965–974.

[36] Samuel Karlin and Howard M. Taylor, *A First Course in Stochastic Processes*, Academic Press, Inc., San Diego (CA), 1975.

[37] F.P. Kelly, "Effective bandwidths at multi–class queues," *Queueing Systems: Theory and Applications* **9** (1991), pp. 5-16.

[38] G. Kesidis, J. Walrand and C.S. Chang, "Effective bandwidths for multi–class Markov fluids and other ATM sources," *IEEE/ACM Transactions on Networking* **1** (1993), pp. 424–428.

[39] M. Krunz and A.M. Makowski, "A source model for VBR video traffic based on $M|G|\infty$ input processes," in *Proceedings of INFOCOM '98*, San Francisco (CA), April 1998, pp. 1441–1448.

[40] M. Krunz and A.M. Makowski, "Modeling video traffic using $M|G|\infty$ input processes: A compromise between Markovian and LRD models," *IEEE Journal on Selected Areas in Communications* **JSAC–16** (1998) pp. 733–748.

[41] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self–similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking* **2** (1994), pp. 1–15.

[42] W. Leland, M. Taqqu, W. Willinger and D. Wilson, "Self–similarity in high–speed packet traffic: Analysis and modeling of Ethernet traffic measurements," *Statistical Science*, **10** (1995), pp. 67–85.

[43] N. Likhanov, B. Tsybakov and N.D. Georganas, "Analysis of an ATM buffer with self–similar ("fractal") input traffic," in *Proceedings of INFOCOM '95*, Boston (MA), April 1995, pp. 985–992.

[44] N. Likhanov "Bounds on the buffer occupancy probability with self-similar input traffic," *Self–similar Network Traffic and Performance Evaluation*, K.

Park and W. Willinger, Eds., J. Wiley & Sons, New York (NY), 1999, to appear.

[45] Z. Liu, Ph. Nain, D. Towsley and Z.–L. Zhang, "Asymptotic behavior of a multiplexer fed by a long–range dependent process," *Journal of Applied Probability* **36**(1999), pp. 105–118.

[46] R.M. Loynes, "The stability of a queue with non–independent inter–arrival and service times," *Proceedings of the Cambridge Philosophical Society* **58** (1962), pp. 497–520.

[47] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review* **10** (1968), pp. 422–437.

[48] H. Michiel and K. Laevens, "Teletraffic engineering in a broadband era," *Proceedings of the IEEE* **85** (1997), pp. 2007–2033.

[49] I. Norros, "A storage model with self–similar input," *Queueing Systems: Theory and Applications* **16** (1994), pp. 387-396.

[50] A. G. Pakes, "On the tails of waiting time distributions," *Journal of Applied Probability* **12**(1975), pp. 555–564.

[51] M. Parulekar and A. Makowski, "Tail probabilities for a multiplexer with self–similar traffic," *Proceedings of INFOCOM '96*, San Fransisco (CA), March 1996, pp. 1452–1459.

[52] M. Parulekar and A. Makowski, "$M|G|\infty$ input processes: A versatile class of models for traffic network," *Proceedings of INFOCOM '97*, Kobe (Japan), April 1997.

[53] M. Parulekar and A. Makowski, "Tail probabilities for $M|G|\infty$ processes (I): Preliminary asymptotics," *Queueing Systems: Theory and Applications* **27** (1997), pp. 271–296.

[54] V. Paxson and S. Floyd, "Wide area traffic: The failure of Poisson modeling," *IEEE/ACM Transactions on Networking* **3** (1993), pp. 226–244.

[55] B. K. Ryu and S. Lowen, "Modeling Self–Similar Traffic with the Fractal–Shot–Noise–Driven–Poisson process," Center for Telecommunications Research, Technical Report, **392-94-39**, Columbia University, New York, (NY), 1994

[56] B. K. Ryu and S. Lowen, "Point Process Approaches to Modeling and Analysis of Self–Similar Traffic. Part I: Model Construction," *Proceedings of INFOCOM '96*, San Fransisco (CA), March 1996, pp. 1468–1475.

[57] B. Ryu and A. Elwalid, "The importance of long–range dependence of VBR video traffic in ATM traffic engineering: Myths and realities," *Proceedings of the ACM SIGCOMM 96 Conference*, Stanford University, Stanford (CA), August 1996, pp. 3–14.

[58] P. Skelly, M. Schwartz and S. Dixit, "A histogram based model for video traffic behavior in an ATM multiplexer,"*IEEE/ACM Transactions on Networking* **1**(4) (1993), pp. 446–458.

[59] B. Tsybakov and N.D. Georganas, "On self–similar traffic in ATM queues: Definitions, overflow probability bound, and cell delay distribution," *IEEE/ACM Transactions on Networking* **TON–5** (1997), pp. 397–409.

[60] B. Tsybakov and N.D. Georganas, "Self–similar Processes in Communication Networks," *IEEE Transactions on Information Theory* **5** (1998) pp. 1713–1725.

[61] D. Veitch, "Novel Models of Broadband Traffic," *Proceedings of GLOBECOM '93* Houston (TX), December 1993, pp. 362–368.

[62] N. Veraverbeke, "Asymptotic behaviour of Wiener-Hopf factors of a random walk," *Stochastic Processes and Their Applications* **5** (1977), pp. 27–37.

[63] W. Whitt, "Tail probability with statistical multiplexing and effective bandwidths in multi–class queues," *Telecommunication Systems* **2**(1) (1994), pp. 71–167.

[64] W. Willinger, M. Taqqu and A. Erramilli, "A Bibliographical Guide to Self–Similar Traffic and Performance Modeling for Modern High–Speed Networks," *Stochastic networks: Theory and Applications*, F.P. Kelly, S. Zachary and I. Ziedins, Eds., Clarendon Press (Oxford University Press), Oxford, 1996, pp. 339-366.