RELAXATION TIME FOR THE DISCRETE D/G/1 QUEUE

A.J.E.M. Janssen

Digital Signal Processing Group, Philips Research, WO-02, 5656 AA Eindhoven, The Netherlands

E-mail: a.j.e.m.janssen@philips.com

J.S.H. van Leeuwaarden

Eurandom, P.O. Box 513, 5600 MB, Eindhoven, The Netherlands

E-mail: j.s.h.v.leeuwaarden@tue.nl

August 23, 2004

Abstract

For the discrete $D/G/1$ queue, the stationary waiting time can be expressed in terms of infinite series that follow from Spitzer's identity. These series involve convolutions of the probability distribution of a discrete random variable, which makes them suitable for computation. For practical purposes, though, the infinite series should be truncated. We therefore seek for some means to characterize the speed at which these series converge. Such a characterization is related to the notion of *relaxation time* in queueing theory, a generic term for the time required for a transient system to reach its stationary regime.

We derive relaxation time asymptotics for the discrete $D/G/1$ queue in a purely analytical way, mostly relying on the saddle point method. We present a simple and useful approximate upper bound which may serve as a stopping criterium for the number of convolutions to calculate in case the load on the system is not very high. A sharpening of this upper bound, which involves the complementary error function, is then developed and this covers both the cases of low and high loads.

Keywords: D/G/1 queue, relaxation time, Spitzer's identity, saddle point method, asymptotics.

AMS 2000 Subject Classification: 30E15, 41A60, 60G50, 60K25.

1 Introduction and motivation

The discrete $D/G/1$ queue refers to a single server queue at which customers arrive with discrete and deterministic interarrival times. We assume that customers are served on a first-come-first-served basis and that their service requirements are i.i.d. according to a discrete random variable A. The waiting time of the nth customer, denoted by W_n, then satisfies (see e.g. [17])

$$W_{n+1} = (W_n + A_n - s)^+, \quad n = 0, 1, \ldots. \tag{1}$$

Here, $x^+ = \max\{0, x\}$, A_n denotes the service time of customer n and the integer s denotes the fixed interarrival time between two consecutive customers. When $\mathbb{E}A < s$, the stationary waiting time denoted by W, $W = \lim_{n\to\infty} W_n$, exists.

To derive the probability generating function (pgf) of W, denoted by $W(z)$, it is common practice to apply an explicit factorization that requires the s roots of a characteristic equation on and inside the unit circle (see e.g [6, 17]). We denote by $A(z)$ the pgf of A, which is assumed to be an analytic function in a disk $|z| \leq 1 + \epsilon$ with $\epsilon > 0$. We further assume throughout (without loss of generality) that $A(0) = \mathbb{P}(A = 0) > 0$. The stationary waiting time in the discrete $D/G/1$ queue as defined by (1) is then fully specified by its pgf

$$W(z) = \frac{s - \mathbb{E}A}{z^s - A(z)}(z - 1)\prod_{k=1}^{s-1}\frac{z - z_k}{1 - z_k}, \tag{2}$$

where $z_0 = 1, z_1, \ldots, z_{s-1}$ are the s roots of $z^s = A(z)$ in $|z| \leq 1$.

For the continuous $G/G/1$ queue, such an explicit factorization is often not available. Obtaining waiting time characteristics then involves the evaluation of either a contour integral or an infinite series of convolutions of the probability distributions of continuous random variables, the latter approach being based on Spitzer's identity (see [18]). Several authors [14, 16, 19] have suggested to approximate the $G/G/1$ queue by its discrete $D/G/1$ counterpart. This can be done as follows. Denote by B_n the service time of customer n and by C_n the interarrival time between customer n and $n+1$. Choose B_n and C_n i.i.d. according to discrete random variable B and C, respectively. Moreover, assume $C \leq s$. Then W_n satisfies

$$W_{n+1} \;=\; (W_n + B_n - C_n)^+ \;=\; (W_n + A_n - s)^+, \quad n = 0, 1, \ldots, \tag{3}$$

where A_n assumed i.i.d. as $A = B - C + s$, and (2) gives the pgf of the stationary waiting time.

This solution still requires the determination of the roots, which in most cases should not give problems (see [7]). Evidently, (2) provides no information on the transient waiting time.

We now show that Spitzer's identity, for the continuous $G/G/1$ queue often leading to unwieldy expressions, results for the discrete $D/G/1$ queue in manageable expressions for both transient and stationary waiting time characteristics.

Using (3), the distribution of W_{n+1} follows from the convolution of the distribution of W_n and that of $A-s$, corrected for the maximum operator. Again, it is favorable to work with discrete random variables, since discrete convolutions are easy to compute (see e.g. [3]). The idea of iterating (3) to obtain transient waiting time characteristics can be made more rigorous using random walk theory. When we assume that the first customer (referred to by subscript 0) arrives at an empty queue ($W_0 = 0$), the joint probability generating function of all W_n is given by Spitzer's identity. That is, for $0 \le t < 1$, $|z| \le 1$,

$$\sum_{n=0}^{\infty} t^n \mathbb{E} z^{W_n} = \exp\Big\{ \sum_{l=1}^{\infty} t^l l^{-1} \mathbb{E} z^{S_l^+} \Big\}, \tag{4}$$

with S_l the lth partial sum of random variables i.i.d. as $A-s$. It follows from manipulating (4) that the mean of W_n is given by

$$\begin{aligned} \mathbb{E}W_n &= \sum_{l=1}^{n} l^{-1} \mathbb{E} S_l^+ \\ &= \sum_{l=1}^{n} l^{-1} \sum_{i=ls}^{\infty} (i-ls)\mathbb{P}(A^{*l} = i), \end{aligned} \tag{5}$$

where A^{*l} denotes the l-fold convolution of A, i.e. $A^{*l} = \sum_{i=1}^{l} A_i$ with A_i i.i.d. as A.

From (4) the stationary waiting time distribution can be obtained as well. When we write (4) as

$$(1-t)\sum_{n=0}^{\infty} t^n \mathbb{E} z^{W_n} = \exp\Big\{ \sum_{l=1}^{\infty} t^l l^{-1} (\mathbb{E} z^{S_l^+} - 1) \Big\}, \tag{6}$$

it follows from Abel's theorem (see [18], p. 207, [8], p. 650), that $W(z)$ is given by

$$\begin{aligned} W(z) &= \lim_{t\uparrow 1}(1-t)\sum_{n=0}^{\infty} t^n \mathbb{E} z^{W_n} \;=\; \exp\Big\{ \sum_{l=1}^{\infty} l^{-1} (\mathbb{E} z^{S_l^+} - 1) \Big\} \\ &= \exp\Big\{ -\sum_{l=1}^{\infty} l^{-1} \mathbb{P}(S_l > 0) \Big\} \exp\Big\{ \sum_{l=1}^{\infty} l^{-1} \mathbb{E}(z^{S_l} \mathbf{1}\{S_l > 0\}) \Big\}, \end{aligned} \tag{7}$$

where $\mathbf{1}\{x\}$ equals 1 if x true and 0 otherwise. Introducing the short-hand notation $C_{z^j}[f(z)]$ for the coefficient of z^j in $f(z)$, and w_j for $\mathbb{P}(W = j)$, it is readily seen that the stationary waiting time distribution is given by

$$\frac{w_j}{w_0} = C_{z^j}\left[\exp\left\{\sum_{l=1}^{\infty} l^{-1} \sum_{i=ls}^{\infty} \mathbb{P}(A^{*l} = i) z^{i-ls}\right\}\right], \quad j = 0, 1, \ldots, \tag{8}$$

where

$$w_0 = \exp\left\{-\sum_{l=1}^{\infty} l^{-1} \sum_{i=ls}^{\infty} \mathbb{P}(A^{*l} = i)\right\}, \tag{9}$$

and A^{*l} denotes the l-fold convolution of A.

Expressions (5), (8) and (9) provide explicit representations of waiting time characteristics solely in terms of infinite series of convolutions of A. Calculating these characteristics is a matter of brute force and the applicability indisputably depends on the ability of computing the discrete convolutions involved. An easy way would be to determine the distribution of A^{*l} from the distribution of $A^{*(l-1)}$. As suggested in [3], it is better, though, to apply a fast Fourier transform algorithm. In that way, given the pgf $A(z)$, the probability distribution of the l-fold convolution can be obtained directly from its pgf $A^l(z)$. In [3] it is shown that the computational speed gained is considerable. For a description of the fast Fourier transform approach to invert a pgf we refer to [2].

1.1 Relaxation time

Irrespective of the method used to compute the convolutions, the issue of truncating the infinite series should be addressed. It is therefore that we seek for some means to characterize the speed at which these series converge. Such a characterization is known in the queueing literature as *relaxation time*, a generic term for time required for transient system characteristics to tend to their steady-state values. When the relaxation time would be defined in terms of the mean waiting time, it could be expressed as the speed at which the difference

$$\begin{aligned} \mathbb{E}W - \mathbb{E}W_{L-1} &= \sum_{l=L}^{\infty} l^{-1} \mathbb{E}S_l^+ \\ &= \sum_{l=L}^{\infty} l^{-1} \sum_{i=ls}^{\infty} (i - ls) \mathbb{P}(A^{*l} = i) \end{aligned} \tag{10}$$

tends to zero for increasing values of L.

Another common way to define the relaxation time is in terms of the probability that a customer has zero waiting time (see e.g. [5]), since determining w_0 is often the bottleneck. For the discrete

$D/G/1$ queue this can be seen as follows. Denote by $w_j(L)$ the estimated value of w_j that results from truncating the series over l at $l = L-1$ in (8) and (9), respectively. The relative error made in estimating w_0 then equals

$$\begin{aligned} \frac{w_0(L) - w_0}{w_0} &= \exp\Big\{\sum_{l=L}^{\infty} l^{-1} \sum_{i=ls}^{\infty} \mathbb{P}(A^{*l} = i)\Big\} - 1 \\ &\approx \sum_{l=L}^{\infty} l^{-1} \sum_{i=ls}^{\infty} \mathbb{P}(A^{*l} = i), \end{aligned} \tag{11}$$

where the far right-hand side of (11) sums all truncation errors $\sum_{l=L}^{\infty} l^{-1}\mathbb{P}(A^{*l} = i)$ that appear in (8) when estimating w_j by $w_j(L)$. Hence, when the left-hand side of (11) is small enough, the accuracy of the estimated values of all w_j seems guaranteed.

A third way to define the relaxation time is in terms of the variance of the waiting time, whose stationary value σ_W^2 follows from (7) by $\sigma_W^2 = W''(1) + W'(1) - W'(1)^2$ yielding

$$\sigma_W^2 = \sum_{l=1}^{\infty} l^{-1} \sum_{i=ls}^{\infty} (i - ls)^2 \mathbb{P}(A^{*l} = i). \tag{12}$$

When we denote by $\sigma_W^2(L)$ the series (12) over l truncated at $l = L-1$, the relaxation time can be expressed as the speed at which the difference

$$\sigma_W^2 - \sigma_W^2(L) = \sum_{l=L}^{\infty} l^{-1} \sum_{i=ls}^{\infty} (i - ls)^2 \mathbb{P}(A^{*l} = i), \tag{13}$$

tends to zero.

In order to extract information from the above measures on the relaxation time, we need insight in the behavior as $L \to \infty$ of the tail series

$$R_m(L) := \sum_{l=L}^{\infty} S_m(l), \quad m = 0, 1, 2, \tag{14}$$

where

$$S_m(l) = l^{-1} \sum_{i=ls}^{\infty} (i - ls)^m \mathbb{P}(A^{*l} = i). \tag{15}$$

Formally, the relaxation time $T(R_m; \epsilon)$ for R_m at level $\epsilon > 0$ can be defined as

$$T(R_m; \epsilon) = \min\{L \mid R_m(L) < \epsilon\}, \tag{16}$$

although this definition is not very practical since it requires computation of all terms in the series defining $R_m(L)$. In this paper we present easily computable asymptotic approximations of $R_m(L)$ and sharp upper bounds on these. A possibility is then to replace the $R_m(L)$ in the definition of $T(R_m;\epsilon)$ in (16) by these upper bounds.

For the continuous $G/G/1$ queue, the relaxation time in terms of the virtual waiting time has been studied extensively by Cohen [8], p. 600, based on analytic continuation of a Laplace transform and the saddle point method. An overview and continuation of this work is given in [5]. In terms of moments of the actual waiting time, expressions for the relaxation time using a change of measure or large deviations technique are obtained in [11, 12] (also see [4], p. 355).

The main contribution in this paper is that we derive relaxation time asymptotics for the discrete $D/G/1$ queue in a concise and purely analytical way. We start from a simple asymptotic approximation of the $\mathbb{P}(A^{*l}=i)$ that appear in (15) using the saddle point method. From this result, we derive asymptotic expressions for $S_m(l)$ and $R_m(L)$, where the latter will allow us to calculate a good approximation of $T(R_m;\epsilon)$ in (16). As a first result, we present an asymptotic expression for $R_m(L)$ based on this asymptotic approximation. This expression admits a simple and useful upper bound which may serve as a stopping criterium for the number of convolutions L to calculate in case the load on the system is not too high. A sharpening of this upper bound, which involves the complementary error function, is then developed and this covers both the cases of low and high loads.

In Sec. 2 we present the main results, that are proved in Secs. 3, 4 and 5. Examples are provided in Sec. 6.

2 Results

Denote $\mathbb{P}(A=n)$ by a_n. Let z_∞ be the radius of convergence of the series $\sum_{n=0}^{\infty} a_n z^n$, and let

$$L_A = \lim_{z\uparrow z_\infty} \frac{zA'(z)}{A(z)}. \tag{17}$$

In Appendix A we show that the limit in (17) always exists as a finite or infinite number, and that $A'(1) < L_A$ unless A is a monomial. In Sec. 3 we obtain an asymptotic approximation of

$$\mathbb{P}(A^{*l}=i) = \frac{1}{2\pi\iota}\oint_{\mathcal{C}} \frac{A^l(z)}{z^{i+1}}\mathrm{d}z, \tag{18}$$

where $\iota = \sqrt{-1}$ and $\mathcal{C}$ is any contour around 0 within the analyticity region of $A(z)$. Using the saddle point method, see De Bruijn [9], we find the following result:

Theorem 2.1. *Assume that* $|A(e^{\iota\theta})|$ *is strictly maximal at* $\theta = 0$ *as a function of* $\theta \in [-\pi, \pi]$. *Let* $i, l \geq 0$ *be integers such that* $A'(1) \leq i/l < L_A$, *and denote* $h(z) = l \ln A(z) - i \ln z$. *Then there is a unique* $z = z_0 \in [1, z_\infty)$ *of the equation* $h'(z) = 0$, *we have* $h''(z_0) > 0$ *and*

$$\mathbb{P}(A^{*l} = i) \approx \frac{1}{z_0\sqrt{2\pi h''(z_0)}} \frac{A^l(z_0)}{z_0^i}. \tag{19}$$

We shall be more precise about the $\approx$ in Sec. 3.

The assumption $A'(1) \leq i/l < L_A$ in Thm. 2.1 ensures the existence of a saddle point on the positive real axis for the integral in (18). The fact that we have to consider integers $i, l \geq 0$ such that $A'(1) \leq i/l < L_A$ does not pose a strong restriction in the present context. To see this, first note that the series defining $S_m(l)$ in (15) involve $i \geq ls$ while we have made the assumption $A'(1) < s$. Secondly, we have in many cases that $L_A = \infty$, see the examples in Sec. 6. Finally, the cases where the load $\rho = A'(1)/s$ is not far away from the maximum sustainable value 1 are the more interesting ones. We shall see in Sec. 4 that, for an accurate approximation of $S_m(l)$ in (15), it is sufficient to consider i for which i/l is not much larger than s. Hence, even in the case of finite L_A, the more interesting cases allow one to restrict to s and i, l satisfying $A'(1) < s \leq i/l < L_A$.

The assumption that $|A(e^{\iota\theta})|$, $\theta \in [-\pi, \pi]$, is strictly maximal at $\theta = 0$ allows us to restrict attention to the immediate vicinity of the saddle point on the positive real axis when the contour $\mathcal{C}$ in (18) is taken to be the circle around zero passing through the saddle point. This condition is not restrictive either. Due to the non-negativity of the a_n and the fact that $a_0 > 0$, the condition is contravened only for $A(z)$ of the form $B(z^p) = \sum_{l=0}^{\infty} b_l z^{lp}$, where

$$p = \min\{|n_1 - n_2| \mid n_1, n_2 = 0, 1, \ldots, n_1 \neq n_2, a_{n_1} \neq 0 \neq a_{n_2}\} > 1. \tag{20}$$

This B is a pgf, just like A, and it does satisfy the condition that $|B(e^{\iota\theta})|$, $\theta \in [-\pi, \pi]$, is strictly maximal at $\theta = 0$. If s is a multiple of p, it suffices to consider B and s/p instead of A and s. Much of the analysis given in this paper applies when $A(z) = B(z^p)$ where s is not a multiple of p, but the administration required for the series over i in (8), (9), (10) and (12) becomes somewhat complicated due to the fact that $\mathbb{P}(A^{*l} = i) \neq 0$ only when i is a multiple of p; we shall exclude such A's.

In all cases, irrespective whether the conditions in Thm. 2.1 on A and i/l are satisfied or not, we have the following bound. For $i, l = 0, 1, \ldots$ there holds

$$\mathbb{P}(A^{*l} = i) \le \inf_{1 \le z < z_\infty} \frac{A^l(z)}{z^i}. \tag{21}$$

In case that the conditions in Thm. 2.1 on A and i/l are satisfied, the number at the right-hand side of (21) equals $A^l(z_0)/z_0^i$ with z_0 as in Thm. 2.1. We note that the right-hand sides of (19) and (21) basically differ by the factor $1/\sqrt{2\pi z_0^2 h''(z_0)}$. Normally, this factor is quite innocent, the key features of the bounds and approximations being determined by the crucial quantity $A^l(z_0)/z_0^i$.

For simplicity we shall assume now that $L_A = \infty$, and we denote

$$\hat{S}_m(l) = \sum_{i=ls}^{\infty} \frac{1}{l}(i - ls)^m \frac{1}{z_0\sqrt{2\pi h''(z_0)}} \frac{A^l(z_0)}{z_0^i}. \tag{22}$$

Theorem 2.2. *We have*

$$S_m(l) \;\approx\; \hat{S}_m(l) \;\approx\; \frac{l^{-3/2}}{\sqrt{2\pi \hat{z}\phi'(\hat{z})}} \left(\frac{A(\hat{z})}{\hat{z}^s}\right)^l \sum_{i=0}^{\infty} i^m \hat{z}^{-i}, \tag{23}$$

where $\phi(z) = zA'(z)/A(z)$ *and* $\hat{z}$ *is the unique* $z \ge 1$ *such that* $\phi(z) = s$.

This result is proved in Sec. 4 where we will be more precise about the $\approx$ in (23). In fact, in many cases, the second $\approx$ in (23) holds as an upper bound on $\hat{S}_m(l)$. Moreover, we briefly consider the issue of how to modify Thm. 2.2 for the case that L_A is finite.

The ϕ of Thm. 2.2 is considered in some detail in Appendix A and is related to z_0 of Thm. 2.1 as follows. When $A'(1) \le t < L_A$ and $z_0(t)$ denotes the unique $z \in [1, z_\infty)$ of $\phi(z) = t$, then $z_0 = z_0(i/l)$ for integer $i, l \ge 0$ such that $A'(1) \le i/l < L_A$.

The series $K_m(v) = \sum_{i=0}^{\infty} i^m v^i,\ m = 0, 1, 2, \ldots$, have been studied in some detail, see [15, 20]. We only need the first few K_m's. We have

$$K_0(v) = \frac{1}{1-v}, \quad K_1(v) = \frac{v}{(1-v)^2}, \quad K_2(v) = \frac{v^2}{(1-v)^3} + \frac{v}{(1-v)^2}. \tag{24}$$

Theorem 2.3. *Let* $0 < \delta < 1$. *When* $A'(1) \le (1-\delta)s$ *and* $L \to \infty$, *there holds for the* $R_m(L)$ *in (14) that*

$$R_m(L) \;\approx\; \hat{R}_m(L) := \sum_{l=L}^{\infty} \hat{S}_m(l) \;\approx\; \frac{K_m(\hat{z}^{-1})}{\sqrt{2\pi \hat{z}\phi'(\hat{z})}} \frac{x^L}{L^{3/2}(1-x)}, \tag{25}$$

where $x = A(\hat{z})/\hat{z}^s$.

For the case that $A'(1)$ *is close to* s *(so that both* $\hat{z}$ *and* x *are close to 1) , there is the more* *precise result*

$$\hat{R}_m(L) \;\approx\; \frac{K_m(\hat{z}^{-1})}{\sqrt{2\pi\hat{z}\phi'(\hat{z})}}\Big[\frac{2x^{L-1}}{\sqrt{L}} - x^{L-3/2}\sqrt{(1-x)\pi}e^{\beta^2}\mathrm{erfc}(\beta)\Big], \tag{26}$$

where $\beta = \sqrt{(1-x)L/x}$*, and*

$$\mathrm{erfc}(\beta) = \frac{2}{\sqrt{\pi}}\int_\beta^\infty e^{-t^2}dt, \quad \beta \geq 0, \tag{27}$$

denotes the complementary error function.

In Sec. 5 we present the proof of this result, and we pay more attention to the $\approx$ in (25) and (26). The sharpening in (26) requires a detailed study of the function $\sum_{l=L}^\infty l^{-3/2}x^l$ in which $L \to \infty$ and x is allowed vary through all values of $[0,1]$.

3 Details for Theorem 2.1

In this section we use the saddle point method to prove Thm. 2.1, and we discuss the conditions on A and i,l that appear in the formulation of Thm. 2.1.

$A(z)$ is assumed to be an analytic function in a disk $|z| < z_\infty$, where the radius of convergence z_∞ of $\sum_{n=0}^\infty a_n z^n$ exceeds 1. Hence

$$\mathbb{P}(A^{*l} = i) = \frac{1}{2\pi\iota}\oint_{\mathcal{C}_r}\frac{A^l(z)}{z^{i+1}}\mathrm{d}z, \tag{28}$$

where $\iota = \sqrt{-1}$ and $\mathcal{C}_r$ is any contour around 0 with radius $r \in [1, z_\infty)$. On such a circle we have by non-negativity of all a_n that $|A^l(z)/z^i|$ is maximal at $z = r$. Hence we get at once that for all $i, l = 0, 1, \ldots$

$$\mathbb{P}(A^{*l} = i) \leq \frac{1}{2\pi}\cdot 2\pi r \cdot \frac{1}{r}\cdot\frac{A^l(r)}{r^i} = \frac{A^l(r)}{r^i} \tag{29}$$

for any $r \in [1, z_\infty)$.

We consider now i, l such that $i/l \geq A'(1)$. Under this assumption we have that $\frac{d}{dz}[A^l(z)/z^i] \leq 0$ at $z = 1$. The information on $\mathbb{P}(A^{*l} = i)$ can now be made more precise when the infimum over $r \in [1, z_\infty)$ of the numbers at the right-hand side of (29) is actually assumed as a minimum at a point $r = z_0 \in [1, z_\infty)$. In that case the point z_0 is a saddle point of

$$h(z) := l\ln A(z) - i\ln z\ , \tag{30}$$

and it is tempting to apply the saddle point method, see De Bruijn [9], Ch. 5, as to obtain an approximation of $\mathbb{P}(A^{*l} = i)$ of the form

$$\mathbb{P}(A^{*l} = i) \approx \frac{e^{h(z_0)}}{z_0\sqrt{2\pi h''(z_0)}}. \tag{31}$$

In order that this saddle point approach is valid, we must make some assumptions. First of all, we need to bother about the existence of the saddle point z_0. Thus, with $\phi(z) = zA'(z)/A(z)$ we want

$$h'(z) = \frac{l}{z}\Big(\phi(z) - \frac{i}{l}\Big) \tag{32}$$

to have one zero $z = z_0(i/l) \in [1, z_\infty)$ at which we have $h''(z) > 0$. In Appendix A it will be shown that (unless A is monomial) $\phi(z)$ is strictly increasing in $z \in [1, z_\infty)$. Hence $h'(z)$ has exactly one zero in $[1, z_\infty)$, provided that

$$\frac{i}{l} \in [\phi(1), \lim_{z\uparrow z_\infty} \phi(z)) = [A'(1), L_A) \tag{33}$$

with L_A as in (17). Furthermore, since

$$h''(z) = \frac{l}{z}\phi'(z) - \frac{1}{z^2}\Big(\phi(z) - \frac{i}{l}\Big), \tag{34}$$

we have that

$$h''(z_0(i/l)) = \frac{l}{z_0(i/l)}\phi'(z_0(i/l)) > 0. \tag{35}$$

The second issue in validating the saddle point approach as embodied by (31) is the fact that we should be allowed to restrict attention to a only a small portion of the integration contour $\mathcal{C}_{z_0}$ in (28) around the saddle point z_0. To that end we make the assumption that $|A(e^{\iota\theta})|$, $\theta \in [-\pi, \pi]$, is strictly maximal at $\theta = 0$. Due to the non-negativity of a_n, this assumption is not really restrictive. Indeed, assuming that for some $\theta \neq 0$, $\theta \in [-\pi, \pi]$

$$|A(e^{\iota\theta})| = \Big|\sum_{n=0}^{\infty} a_n e^{\iota n\theta}\Big| = \sum_{n=0}^{\infty} a_n = A(1), \tag{36}$$

we see that there is a $\gamma \in [-\pi, \pi]$ such that $e^{\iota n\theta} = e^{\iota\gamma}$ for all $n = 0, 1, \ldots$ such that $a_n \neq 0$.

It is easily seen that this strict maximality of $|A(e^{\iota\theta})|$, $\theta \in [-\pi, \pi]$, at $\theta = 0$ is equivalent with strict maximality of$|A(re^{\iota\theta})|$, $\theta \in [-\pi, \pi]$, at $\theta = 0$ for any $r \in [1, z_\infty)$. As a consequence we can replace the integral along $\mathcal{C}_{z_0}$ by an integral along small circle segments $\{z_0e^{\iota\theta}||\theta| \leq \delta\}$ at the expense of exponentially small errors of order

$$\max_{\delta\leq|\theta|\leq\pi}\Big|\frac{A(z_0)e^{\iota\theta}}{A(z_0)}\Big|^l. \tag{37}$$

The term "exponentially small" used here can be somewhat deceptive as the following example shows. Choose

$$A(z)=\frac{\cosh\lambda z+\epsilon z}{\cosh\lambda+\epsilon}, \tag{38}$$

where $\lambda>0$ is large and $\epsilon>0$ is small. The ratio $A(-z_0)/A(z_0)$ is extremely close to 1, and it is only for very large l that one can ignore the contribution to the integral of z's near $-z_0$.

The further details of applying the saddle point method for the present case follow to a large extent the discussion in [9], p. 92 on the range of the saddle point. Here, it is important to note that considerations in Sec. 4 show that we can restrict attention to integers $i,l\geq 0$ such that i/l is not much larger than s. Thus we can write

$$h(z)=lg_t(z);\quad g_t(z)=\ln A(z)-t\ln z, \tag{39}$$

where $t\in[s,(1+\delta)s]$ with $\delta>0$ not large and certainly such that $(1+\delta)s<L_A$. This implies that the z_0 are in a compact interval in $[1,z_\infty)$, whence the $g_t''(z_0)$ are uniformly bounded away from 0 while higher derivatives, such as $g_t'''(z_0)$ and $g_t''''(z_0)$, are uniformly bounded away from ∞. Following the discussion in [9], p.92, we then replace the integral along $\mathcal{C}_{z_0}$ by an integral along the line segment between $z_0-\iota l^{-\gamma}$ and $z_0+\iota l^{-\gamma}$, where γ is a real number between 1/3 and 1/2, at the expense of an exponentially small error like $\exp(-\frac{1}{2}l^{1-2\gamma}g_t''(z_0))$. On this line segment the remainder of $lg_t(z)$, after splitting off the constant and quadratic term $-lg_t(z_0)+\frac{1}{2}lg_t''(z_0)(z-z_0)^2$,

$$l(\tfrac{1}{6}g_t'''(z_0)(z-z_0)^3+\tfrac{1}{24}g_t''''(z_0)(z-z_0)^4+\ldots) \tag{40}$$

tends to zero as $l\to\infty$. Hence, at the expense of smaller errors, we can linearize $\exp(h(z))$ on the line segment $z_0+\iota u$, $|u|\leq l^{-\gamma}$, as

$$e^{h(z)}=e^{h(z_0)-\frac{1}{2}u^2h''(z_0)}(1-\tfrac{1}{6}\iota lg_t'''(z_0)u^3+\tfrac{1}{24}lg_t''''(z_0)u^4). \tag{41}$$

Now note that the term involving $g_t'''(z_0)$ cancels upon integration over $u\in[-l^{-\gamma},l^{-\gamma}]$ since u^3

is odd. The integral of the term involving $g_t''''(z_0)$ can be estimated at

$$e^{h(z_0)}\tfrac{l}{24}|g_t''''(z_0)|\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2h''(z_0)}u^4du = \Gamma(5/2)\tfrac{l}{24}|g_t''''(z_0)|e^{h(z_0)}\Big(\frac{2}{h''(z_0)}\Big)^{5/2}. \tag{42}$$

It follows that the relative error due to this latter term has the order

$$l\Big(\frac{2}{h''(z_0)}\Big)^2 \;=\; \frac{4}{l(g_t''(z_0))^2} \;=\; \mathcal{O}(1/l) \tag{43}$$

uniformly in $t \in [s,(1+\delta)s]$. Similarly, the lowest order deleted quadratic term $-\frac{1}{36}l^2(g_t'''(z_0))^2u^6$ at the right-hand side of (41) produces a relative error $\mathcal{O}(1/l)$ as well, and higher order terms produce smaller errors, etc. In all this, the additional factor $1/z$ that appears in the integral in (28) according to

$$\frac{A^l(z)}{z^{i+1}} = \frac{1}{z}e^{h(z)} \tag{44}$$

has been considered as a constant $1/z_0$. As in the above, this can be shown to be allowed, at the expense of a relative error of order $\mathcal{O}(1/l)$. We conclude that when we restrict i/l to a range in $[s,(1+\delta)s] \subset [A'(1), L_A)$, the relative error for the approximation in (31) is $\mathcal{O}(1/l)$ uniformly in i.

We conclude this section by a consideration of A for which L_A is finite (in many cases $L_A = \infty$ so that the assumption $i/l < L_A$ in Thm. 2.1 presents no restriction). First assume that $z_\infty = \infty$, so that $A(z)$ is an entire function. From $L_A < \infty$ and the fact that $a_n \geq 0$ it then follows that $A(z)$ is a polynomial of degree L_A. Hence in this case $\mathbb{P}(A^{*l} = i) = 0$ when $i > lL_A$. Next consider the case that $z_\infty < \infty$ and $L_A < \infty$. It is easy to see that then

$$A(z_\infty) := \lim_{z\uparrow z_\infty} A(z), \quad A'(z_\infty) := \lim_{z\uparrow z_\infty} A'(z) \tag{45}$$

exist as finite numbers. When now $i/l > L_A$, a precise approximation is feasible only when additional information about the nature of the singular point z_0 is available. However, the bound in (29) remains valid, and this is normally enough for our purposes where we may restrict to integers i/l such that i/l is not much larger than s while $s < L_A$.

4 Details for Theorem 2.2

In this section we present the proof of Thm. 2.2 and detail some of its claims. We exclude the case that A is a polynomial (only for the sake of a smoother presentation with z_∞ below). Hence, when $L_A < \infty$, we assume that $z_\infty < \infty$ so that $A(z_\infty), A'(z_\infty)$ are given by (45) as finite numbers, while

$$L_A = \frac{z_\infty A'(z_\infty)}{A(z_\infty)} = \phi(z_\infty). \tag{46}$$

Here, with L_A finite or not,

$$\phi(z) = zA'(z)/A(z), \quad |z| < z_\infty, \tag{47}$$

as in (32). We show in Appendix A that ϕ is strictly increasing in $z \in [0, z_\infty)$, unless A is a monomial.

We let for $t > A'(1)$

$$z_0(t) = \begin{cases} \text{unique } z \geq 1 \text{ such that } \phi(z) = t, & A'(1) \leq t < L_A, \\ z_\infty, & t \geq L_A. \end{cases} \tag{48}$$

Thus $z_0(t)$ is strictly increasing in $t \in [A'(1), L_A)$ and constant z_∞ for $t \geq L_A$.

In terms of ϕ and z_0 we can express the saddle point approximation of $\mathbb{P}(A^{*l} = i)$ in Thm. 2.1 as

$$\mathbb{P}(A^{*l} = i) \approx \frac{1}{\sqrt{2\pi l z_0(i/l)\phi'(z_0(i/l))}} \left(\frac{A(z_0(i/l))}{z_0(i/l)^{i/l}}\right)^l \tag{49}$$

when $A'(1) \leq i/l < L_A$. Also, the bound (29) can be expressed in terms of z_0 as

$$\mathbb{P}(A^{*l} = i) \leq \left(\frac{A(z_0(i/l))}{z_0(i/l)^{i/l}}\right)^l, \quad i/l \geq A'(1). \tag{50}$$

For the analysis that follows we introduce the function

$$G(t) := \ln\left[\frac{A(z_0(t))}{z_0(t)^t}\right], \quad t \geq A'(1). \tag{51}$$

Note that $G(t) = g_t(z_0(t))$, see (39). The function G is considered in some detail in Appendix A. It is shown that G is a non-positive, strictly decreasing, concave function of $t \geq A'(1)$ for which the t-axis is a tangent of the graph $(t, G(t))$, $t \geq A'(1)$ at the point $(t = A'(1), G(A'(1)) = 0)$. Moreover, it is shown that

$$G'(t) = -\ln z_0(t), \quad t \geq A'(1). \tag{52}$$

In particular, G is strictly concave on $[A'(1), L_A)$ with

$$G''(t) = \frac{-1}{z_0(t)\phi'(z_0(t))}, \quad t \in [A'(1), L_A), \tag{53}$$

and G is linear on $[L_A, \infty)$ with $G'(t) = -\ln z_\infty$ (when $L_A < \infty$). Also see Fig. 1.

We restrict for the moment to $L_A = \infty$, and we consider

$$\begin{aligned} \hat{S}_m(l) &= \frac{1}{l\sqrt{2\pi l}} \sum_{i=ls}^{\infty} \frac{(i-ls)^m}{\sqrt{z_0(i/l)\phi'(z_0(i/l))}} \left(\frac{A(z_0(i/l))}{(z_0(i/l))^{i/l}}\right)^l \\ &= \frac{1}{l\sqrt{2\pi l}} \sum_{i=ls}^{\infty} (i-ls)^m \sqrt{-G''(i/l)} e^{lG(i/l)} \end{aligned} \tag{54}$$

as an approximation of $S_m(l)$, see (15). In the first line of (54) we have inserted the saddle point approximation (49) of $\mathbb{P}(A^{*l} = i)$ into the series (15) at the expense of a relative error $\mathcal{O}(1/l)$. Next, the concave function G in the exponential is replaced by its linearization around $(s, G(s))$, and $-G''(i/l)$ is replaced by $-G''(s)$. We thus obtain the approximation

$$\begin{aligned} \hat{S}_m(l) &\approx \frac{\sqrt{-G''(s)}}{l\sqrt{2\pi l}} \sum_{i=ls}^{\infty} (i-ls)^m e^{l[G(s)+(i/l-s)G'(s)]} \\ &= \frac{\sqrt{-G''(s)}}{l\sqrt{2\pi l}} e^{lG(s)} \sum_{i=ls}^{\infty} (i-ls)^m e^{(i-ls)G'(s)}. \end{aligned} \tag{55}$$

When we now use (51), (52), (53), we get (23) in Thm. 2.2. The crucial step in getting the approximation (55) is the linearization of the function $G(t)$ around $t = s$. In Fig. 1 we display this linearization for the case that $A(z) = \exp(\lambda(z-1))$ with $\lambda = 9$ and $s = 9, 15, 20$.

We note that in many cases the approximation (55) holds as an upper bound on $\hat{S}_m(l)$. This is certainly so when $z\phi'(z)$ is an increasing function of z (as often happens, see the examples in Sec. 6). For then both replacing $-G''(i/l)$ by $-G''(s)$ and G by its linearization in (54) comes with a $\leq$-sign. The condition $(z\phi'(z))' = \phi'(z) + z\phi''(z) \geq 0$ is not very restrictive; it excludes functions A that grow slower than $z^{a+b\ln z}$ with some $a > 0$, $b > 0$.

We next make a brief error assessment for the approximation in (55). We note that by Taylor's formula

$$l[G(i/l) - G(s) - (i/l - s)G'(s)] = \frac{1}{2l}(i-ls)^2 G''(\zeta) < 0, \tag{56}$$

where ζ is a number $\in [s, i/l]$. Also, $G''(i/l) - G''(s) = \mathcal{O}(i/l - s)$. Then, due to exponential decay, one can show that relative errors of order $1/l$ occur.

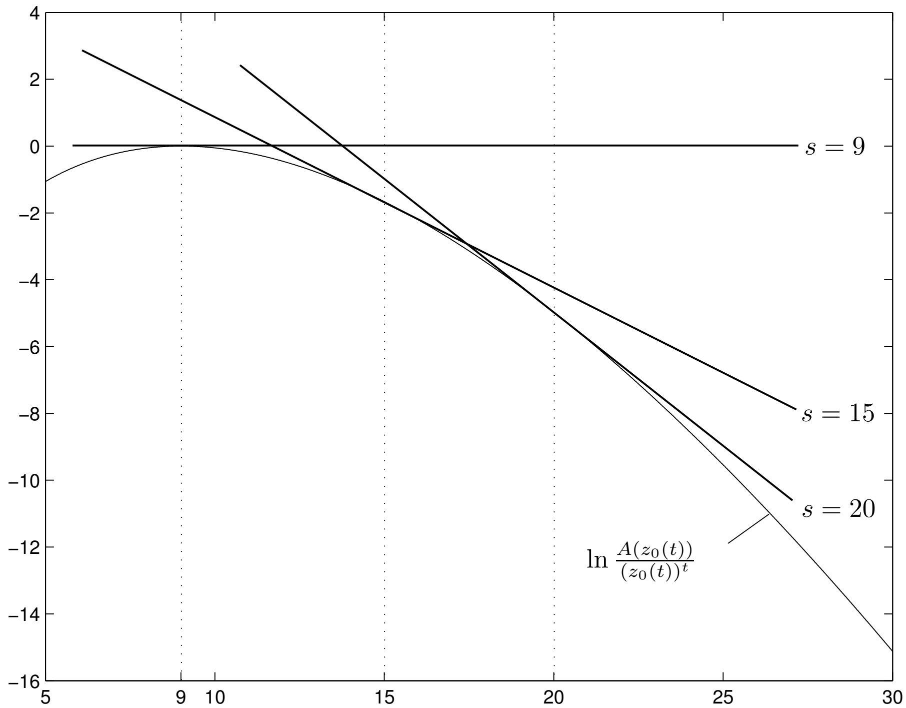

Figure 1: Linearization of $\ln \frac{A(z_0(t))}{(z_0(t))^t}$ at $t = s$ for $A(z) = e^{\lambda(z-1)}$ (Poisson distribution), $\lambda = 9$, $s = 9$ (t-axis), 15 and 20.

Finally, in the case of finite L_A and $s < L_A$, the above argument to approximate and bound $S_m(l)$ remains basically the same (due to the bound in (50) and concavity of G) at the expense of exponentially small relative errors.

5 Proof of Theorem 2.3

We shall now approximate and bound the quantity

$$\hat{R}_m(L) \; = \frac{K_m(\hat{z}^{-1})}{\sqrt{2\pi\hat{z}\phi'(\hat{z})}} \sum_{l=L}^{\infty} l^{-3/2} \Big(\frac{A(\hat{z})}{\hat{z}^s}\Big)^l, \quad L \to \infty, \tag{57}$$

as it occurs as an upper bound of the approximation $\sum_{l=L}^{\infty} \hat{S}_m(l)$ on $R_m(L) = \sum_{l=L}^{\infty} S_m(l)$. It suffices to study the quantity

$$F_L(x) := \sum_{l=L}^{\infty} l^{-3/2} x^l, \tag{58}$$

for large L and $x \in [0, 1]$. It is interesting to note that $F_L(x) = x^L \Phi(z = x, s = 3/2, v = L)$, where $\Phi(z, s, v)$ is Lerch's transcendent as occurs in [10], §1.11 on pp. 27-31. Of the many formulas and representations developed in [10], §1.11 for Φ, the one in §1.11 (3) is particularly convenient for getting a simple and accurate approximation of $F_L(x)$ when L gets large and $x \in [0,1]$. When x is away from 1, one simply has

$$\begin{aligned} F_L(x) &= \frac{x^L}{L^{3/2}} \sum_{l=0}^{\infty} \frac{1}{(1+l/L)^{3/2}} x^l, \\ &\approx \frac{x^L}{L^{3/2}} \sum_{l=0}^{\infty} x^l \;=\; \frac{x^L}{L^{3/2}(1-x)}, \end{aligned} \tag{59}$$

with a relative error that is of the order $-3x/(2L(1-x))$. The right-hand side of (59) is in fact an upper bound for $F_L(x)$. This gives the result in (25).

When x may get close to 1 while $L \to \infty$, we have to proceed more carefully: While $F_L(x)$ is evidently bounded for $x \in [0,1]$, the last member of (59) tends to infinity as x tends to 1. From

$$y^{-3/2} = \frac{2}{\sqrt{\pi}} \int_0^{\infty} u^{1/2} e^{-yu} du, \tag{60}$$

we obtain [10], §1.11(3),

$$F_L(x) = \frac{2x^L}{\sqrt{\pi}} \int_0^{\infty} \frac{u^{1/2} e^{-Lu}}{1 - xe^{-u}} du. \tag{61}$$

With $L \to \infty$, we may restrict attention in the integral in (61) to small $u \geq 0$, and we expand

$$\begin{aligned} \frac{1}{1-xe^{-u}} &= \frac{1}{1-x+xu-x(e^{-u}-1+u)} \\ &= \frac{1}{1-x+xu}\Big(1 + \frac{x(e^{-u}-1+u)}{1-x+xu} + \Big(\frac{x(e^{-u}-1+u)}{1-x+xu}\Big)^2 + \ldots\Big). \end{aligned} \tag{62}$$

Since

$$0 \leq \frac{x(e^{-u}-1+u)}{1-x+xu} \leq \frac{\frac{1}{2}xu^2}{1-x+xu} \leq \frac{1}{2}u, \tag{63}$$

we see that the leading term of $F_L(x)$ is given as

$$\frac{2x^L}{\sqrt{\pi}} \int_0^{\infty} \frac{u^{1/2} e^{-Lu}}{1-x+xu} du, \tag{64}$$

while the error is accurately estimated at

$$\frac{2x^L}{\sqrt{\pi}} \int_0^{\infty} \frac{\frac{1}{2}xu^{5/2} e^{-Lu}}{(1-x+xu)^2} du. \tag{65}$$

The analysis given above can be made more precise as follows. We restrict the integration range in (61) to $u \in [0, L^{-1/2}]$ at the expense of exponentially small errors $\mathcal{O}(\exp(-L^{-1/2}))$. On the relevant integration range we have from (62) and (63)

$$\frac{1}{1-xe^{-u}}=\frac{1}{1-x+xu}+\frac{\frac{1}{2}xu^2}{(1-x+xu)^2}\ (1+\mathcal{O}(L^{-1/2})),\tag{66}$$

where the $\mathcal{O}$ holds uniformly in $x\in[0,1]$, $u\in[0,L^{-1/2}]$. Then the integration range for the two functions of u at the right-hand side of (66) is restored to $[0,\infty)$, again at the expense of exponentially small errors $\mathcal{O}(\exp(-L^{-1/2}))$.

We shall now express the integrals in (64), (65) in terms of the complementary error function, [1], p. 297,

$$\operatorname{erfc}(x):=\frac{2}{\sqrt{\pi}}\int_x^\infty e^{-y^2}dy,\quad x\geq 0.\tag{67}$$

There holds

$$\int_0^\infty\frac{u^{1/2}e^{-Lu}}{1-x+xu}du=\frac{1}{xL^{1/2}}\int_0^\infty\frac{t^{1/2}e^{-t}}{\beta^2+t}dt,\tag{68}$$

where we have set

$$\beta=\Big(\frac{1-x}{x}L\Big)^{1/2}.\tag{69}$$

Then, by [1], p. 302

$$\begin{aligned}\int_0^\infty\frac{t^{1/2}e^{-t}}{\beta^2+t}dt &= \int_0^\infty t^{-1/2}e^{-t}dt-\beta^2\int_0^\infty\frac{e^{-t}}{\sqrt{t}(\beta^2+t)}dt\\ &= \sqrt{\pi}(1-\sqrt{\pi}\beta e^{\beta^2}\operatorname{erfc}(\beta)).\end{aligned}\tag{70}$$

Therefore, with β given in (69), we have

$$\int_0^\infty\frac{u^{1/2}e^{-Lu}}{1-x+xu}du=\frac{\sqrt{\pi}}{xL^{1/2}}(1-\sqrt{\pi}\ \beta\ e^{\beta^2}\operatorname{erfc}(\beta)).\tag{71}$$

Similarly, we have by partial integration

$$\begin{aligned}\int_0^\infty\frac{\frac{1}{2}xu^{5/2}e^{-Lu}}{(1-x+xu)^2}du &= \frac{1}{2}\int_0^\infty\frac{1}{1-x+xu}\Big(\frac{5}{2}u^{3/2}-u^{5/2}L\Big)e^{-Lu}du\\ &= -\frac{1}{2}\Big\{\frac{5}{2}\Big(\frac{d}{dL}\Big)+L\Big(\frac{d}{dL}\Big)^2\Big\}\int_0^\infty\frac{u^{1/2}e^{-Lu}}{1-x+xu}du.\end{aligned}\tag{72}$$

Then, using (71) and the definition of β in (69), we see that the integral in (65) can be expressed in terms of elementary functions and the erfc, but the resulting expression is rather unwieldy, i.e.

$$\int_0^\infty \frac{\frac{1}{2}xu^{5/2}e^{-Lu}}{(1-x+xu)^2}du = \frac{\sqrt{\pi}}{8xL^{3/2}}\Big\{(5+2\beta^2)(1-2\beta^2(1-\sqrt{\pi}\beta e^{\beta^2}\mathrm{erfc}(\beta)))-3\Big\}. \tag{73}$$

The function $\exp(\beta^2)\mathrm{erfc}(\beta)$ is known as Mills' ratio, see [1], 7.1.13 on p.298. Using the asymptotic series, [1], p. 298,

$$\sqrt{\pi}\beta e^{\beta^2}\mathrm{erfc}(\beta) \sim 1 - \frac{1}{2\beta^2} + \frac{3}{4\beta^4} - \frac{15}{8\beta^6} + \frac{105}{16\beta^8} - \ldots, \quad \beta\to\infty, \tag{74}$$

we get from (71)

$$\int_0^\infty \frac{u^{1/2}e^{-Lu}}{1-x+xu}du \sim \frac{\sqrt{\pi}}{x}\Big(\frac{1}{2\frac{1-x}{x}L^{3/2}} - \frac{3}{4(\frac{1-x}{x})^2L^{5/2}} + \frac{15}{8(\frac{1-x}{x})^3L^{7/2}} - \ldots\Big), \tag{75}$$

as $(1-x)L/x\to\infty$. Furthermore, from (72) and (75) (repeated differentiation of the asymptotic series is allowed)

$$\int_0^\infty \frac{\frac{1}{2}xu^{5/2}e^{-Lu}}{(1-x+xu)^2}du \sim \frac{\sqrt{\pi}}{x}\Big(\frac{15}{16(\frac{1-x}{x})^2L^{7/2}} - \frac{105}{16(\frac{1-x}{x})^3L^{9/2}} + \ldots\Big), \tag{76}$$

as $(1-x)L/x\to\infty$. Note that the first term at the right-hand side of (75) agrees with the approximation given in (59). Also note that the leading term in (76) is a factor $\frac{8}{15}\frac{1-x}{x}L^2$ smaller than the leading term in the right-hand side of (75).

The asymptotics in (75), (76) are valid when $(1-x)L/x\to\infty$. We complement this info by presenting lower and upper bounds for the integrals in (64), (65) that show that the second integral is roughly a factor $L(1+\frac{1}{2}\beta^2)$ smaller than the first integral for all values $\beta=\sqrt{(1-x)L/x}\geq 0$. As to the first integral we have

$$\begin{aligned}\int_0^\infty \frac{u^{1/2}e^{-Lu}}{1-x+xu}du &= \frac{1}{xL^{1/2}}\int_0^\infty \frac{t^{1/2}e^{-t}}{\beta^2+t}dt \\ &= \frac{\sqrt{\pi}}{2xL^{1/2}}\int_0^\infty \frac{e^{-\beta^2 v}}{(1+v)^{3/2}}dv.\end{aligned} \tag{77}$$

Here we have inserted

$$\Gamma(\alpha)\Big(\frac{1}{t+\beta^2}\Big)^\alpha = \int_0^\infty v^{\alpha-1}e^{-v(t+\beta^2)}dv \tag{78}$$

with $\alpha=1$ into the second integral in (77), interchanged the order of integration and used

$\sqrt{\pi}/2 = \Gamma(3/2) = \int_0^\infty t^{1/2}e^{-t}dt$. Then by the inequality

$$e^{-(\beta^2+3/2)v} \leq \frac{e^{-\beta^2 v}}{(1+v)^{3/2}} \leq \Big(\frac{1}{1+v}\Big)^{3/2+\beta^2}, \tag{79}$$

we immediately get

$$\frac{\sqrt{\pi}}{2xL^{1/2}}\frac{1}{3/2+\beta^2} \leq \int_0^\infty \frac{u^{1/2}e^{-Lu}}{1-x+xu}du \leq \frac{\sqrt{\pi}}{2xL^{1/2}}\frac{1}{1/2+\beta^2}. \tag{80}$$

In an entirely similar way, using (78) with $\alpha = 2$, we get

$$\int_0^\infty \frac{\frac{1}{2}xu^{5/2}e^{-Lu}}{(1-x+xu)^2}du = \frac{15\sqrt{\pi}}{16xL^{3/2}}\int_0^\infty \frac{ve^{-\beta^2 v}}{(1+v)^{7/2}}dv, \tag{81}$$

from which it follows that

$$\frac{15\sqrt{\pi}}{16xL^{3/2}}\frac{1}{(7/2+\beta^2)^2} \leq \int_0^\infty \frac{\frac{1}{2}xu^{5/2}e^{-Lu}}{(1-x+xu)^2}du \leq \frac{15\sqrt{\pi}}{16xL^{3/2}}\frac{1}{(3/2+\beta^2)(5/2+\beta^2)}. \tag{82}$$

We may, finally, note that continued fraction expansions for the integrals in (64) and (65) can be obtained from [23], pp. 352-355; also see [21], Sec. 11.2 where asymptotics of integrals of type (64) and (65) are considered in connection with the incomplete Gamma function.

6 Examples

In this section we consider several examples for which we determine characteristics of the relaxation time. For each example, the load on the system is defined as $\rho = A'(1)/s$ and assumed to be less than one.

Example 6.1. Poisson case. $A(z) = e^{\lambda(z-1)}$, $A'(1) = \lambda$, $\phi(z) = \lambda z$, $z_\infty = \infty$, $L_A = \infty$, $z\phi'(z) = \lambda z$ increasing and

$$z_0 = \frac{i}{l\lambda}, \quad \hat{z} = z_0(s) = \frac{s}{\lambda}, \quad x = \frac{A(z_0(s))}{(z_0(s))^s} = \Big(\frac{\lambda}{s}e^{1-\lambda/s}\Big)^s. \tag{83}$$

From Thm. 2.1 we thus have

$$\mathbb{P}(A^{*l} = i) \approx \frac{1}{i/(l\lambda)} \cdot \frac{1}{\sqrt{2\pi\frac{1}{i}(l\lambda)^2}} \cdot \frac{\exp(l\lambda(\frac{i}{l\lambda}) - 1)}{(\frac{i}{l\lambda})^i} = \frac{1}{\sqrt{2\pi i}}\Big(\frac{l\lambda}{i}\cdot e^{1-\frac{l\lambda}{i}}\Big)^i. \tag{84}$$

Observe that $te^{1-t} \in [0,1)$ when $t \in [0,1)$. In the present case we have, explicitly,

$$\mathbb{P}(A^{*l} = i) = \frac{e^{-l\lambda}}{i!}(l\lambda)^i \approx e^{-l\lambda}(i^n e^{-i}\sqrt{2\pi i})^{-1}(l\lambda)^i = \frac{1}{\sqrt{2\pi i}}\Big(\frac{l\lambda}{i}\cdot e^{1-\frac{l\lambda}{i}}\Big)^i, \tag{85}$$

where Stirling's formula $i! \approx i^{i+1/2}e^{-i}\sqrt{2\pi}$ has been used. It is thus seen that the approximation as obtained per Thm. 2.1 amounts to replacing $i!$ in the exact expression (85) by its Stirling approximation. Accordingly, the approximation given by (84) has relative error $\mathcal{O}(1/i)$ independent of λ.

Example 6.2. Geometric case. $A(z) = (1-p)/(1-pz)$, $A'(1) = p/(1-p)$, $\phi(z) = pz/(1-pz)$, $z_\infty = 1/p$, $L_A = \infty$, $z\phi'(z) = pz/(1-pz)^2$ increasing for $z \le 1$ and decreasing for $z \ge 1$, and

$$z_0 = \frac{i}{p(i+l)}, \quad \hat{z} = z_0(s) = \frac{1}{p}\frac{s}{s+1}, \quad x = \frac{A(z_0(s))}{(z_0(s))^s} = (1-p)p^s\frac{(s+1)^{s+1}}{s^s}. \tag{86}$$

From Thm. 2.1 it follows that

$$\mathbb{P}(A^{*l} = i) \approx \frac{1}{\sqrt{2\pi}}(1-p)^l p^i (i+l)^{i+l-1/2} i^{-i-1/2} l^{-l+1/2}. \tag{87}$$

From the explicit representation

$$\mathbb{P}(A^{*l} = i) = (1-p)^l p^i \frac{(i+l)!}{i!\ l!}\ \frac{l}{i+l}, \tag{88}$$

we obtain by Stirling's formula exactly (87). Accordingly, as in Example 6.1, the approximation given by (87) has relative error $\mathcal{O}(1/i)$ independent of p.

Example 6.3. Binomial case. $A(z) = (p+qz)^n$, $p+q = 1$, $A'(1) = nq$, $\phi(z) = nqz/(p+qz)$, $z_\infty = \infty$, $L_A = n$, $z\phi'(z) = npqz/(p+qz)^2$ increasing for $z \le q/p$ and decreasing for $z \ge q/p$, and

$$z_0 = \frac{1}{q}\frac{pi}{nl-i}, \quad \hat{z} = z_0(s) = \frac{1}{q}\frac{ps}{n-s}, \quad x = \frac{A(z_0(s))}{(z_0(s))^s} = q^s p^{n-s} n^n s^{-s}(n-s)^{-(n-s)}. \tag{89}$$

From Thm. 2.1 it follows that

$$\mathbb{P}(A^{*l} = i) \approx \frac{1}{\sqrt{2\pi}}\frac{(nl)^{nl+1/2}}{(nl-i)^{nl-i+1/2} i^{i+1/2}} p^{nl-i} q^i. \tag{90}$$

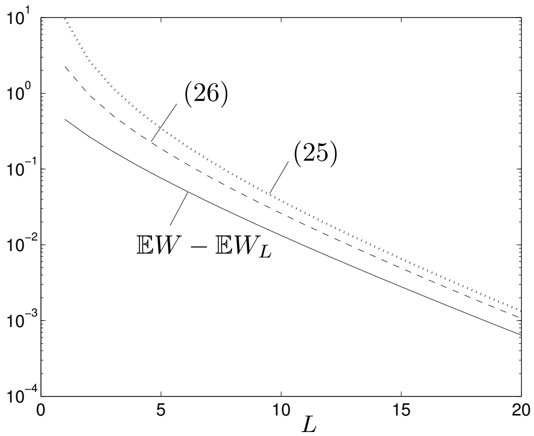

Figure 2: Relaxation time for the Poisson case with $s = 10$, $\lambda = 8$.

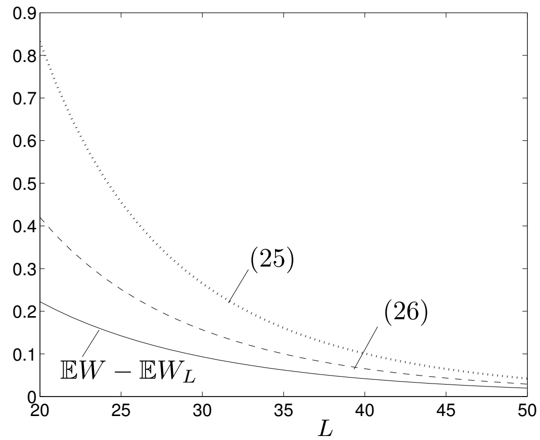

Figure 3: Relaxation time for the Poisson case with $s = 10$, $\lambda = 9$.

From the explicit representation

$$\mathbb{P}(A^{*l} = i) = \frac{(nl)!}{(nl-i)!i!} p^{nl-i} q^i, \tag{91}$$

we obtain by Stirling's formula exactly (90). For the remainder of this section we set $n = 4s$.

Thm. 2.3 gives asymptotic expressions for $\hat{R}_m(L)$ from which we can extract information on the relaxation time. Expression (25) yields an upper bound on $\hat{R}_m(L)$, which is expected to be sharp for loads well below one. Expression (26) sharpens (25) and should be useful when ρ tends to 1. The complementary error function $\mathrm{erfc}(\beta)$ needed to calculate (26) is a standard function available in most software packages.

For the Poisson case, Figs. 2 and 3 depict characteristics of these asymptotic approximations for $m = 1$ (corresponding to the mean waiting time), for $s = 10$ and $\rho = 0.8, 0.9$, respectively. The true value of $R_1(L)$ results from $\mathbb{E}W - \mathbb{E}W_L$, where we approximate $\mathbb{E}W$ using extremely high truncation levels (something we want to avoid). For $\rho = 0.8$, the relaxation time decreases rapidly, indicating that the transient behavior of the waiting time converges rapidly to its steady-state. (26) improves upon (25), although the improvement is marginal relative to the true value $\mathbb{E}W - \mathbb{E}W_L$.

For $\rho = 0.9$, the relaxation time again decreases rapidly, although we need a much higher value of L in order to achieve the same accuracy as for $\rho = 0.8$. Again, (26) improves upon (25), where now the absolute improvement is much larger.

For each of the three examples, we calculate

$$T(\hat{R}_m;\epsilon) = \min\{L \mid \hat{R}_m(L) < \epsilon\}, \tag{92}$$

where we replace $\hat{R}_m(L)$ in (92) by either (25) or (26). Remember that (25) is an upper bound on $\hat{R}_m(L)$, where (26) is, although asymptotically sharp, an asymptotic approximation of $\hat{R}_m(L)$. We set ϵ equal to 0.001. When we want, for example, to determine the mean waiting time, we could approximate $\mathbb{E}W$ using (5) with $n = T(\hat{R}_1; 0.001)$, knowing that $\mathbb{E}W - \mathbb{E}W_n$ is of order 0.001. Of course, (5) still contains an infinite series over i, but truncating this series at $ls + M$ for some large value M gives a truly negligible error, for reasons addressed in Sec. 3.

Results are displayed in Table 1, 2 and 3. We first make some general observations. For low values of ρ, a small value of $T(\hat{R}_m;\epsilon)$ is sufficient. For high values of ρ, though, the $T(\hat{R}_m;\epsilon)$ required increases enormously. Using (26) instead of (25) leads to moderate reductions in $T(\hat{R}_m;\epsilon)$, mostly for high values of ρ. Further, $T(\hat{R}_0;\epsilon) \leq T(\hat{R}_1;\epsilon) \leq T(\hat{R}_2;\epsilon)$, which is obvious from the K_m functions given in (24).

Table 1: $T(\hat{R}_m;\epsilon)$ for $m = 0$, $\epsilon = 0.001$, using either (25) or (26), for the binomial, Poisson and geometric case.

	$s = 10$						$s = 30$					
	binomial		Poisson		geometric		binomial		Poisson		geometric	
ρ	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)
0.5	2	2	3	3	14	14	1	1	1	1	13	12
0.6	4	4	4	4	27	26	2	2	2	2	25	23
0.7	6	6	7	7	58	54	3	3	3	3	53	50
0.8	14	13	17	16	152	143	5	5	6	6	141	133
0.9	54	51	70	66	707	664	19	18	24	23	660	621

Table 2: $T(\hat{R}_m;\epsilon)$ for $m = 1$, $\epsilon = 0.001$, using either (25) or (26), for the binomial, Poisson and geometric case.

	$s = 10$						$s = 30$					
	binomial		Poisson		geometric		binomial		Poisson		geometric	
ρ	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)
0.5	2	2	3	3	21	21	1	1	1	1	22	21
0.6	4	4	5	4	43	41	2	2	2	2	45	44
0.7	7	7	9	8	96	94	3	3	3	3	101	99
0.8	16	16	21	21	273	267	6	6	8	8	287	281
0.9	75	72	101	98	1395	1367	25	25	34	33	1460	1436

Table 3: $T(\hat{R}_m;\epsilon)$ for $m=2$, $\epsilon=0.001$, using either (25) or (26), for the binomial, Poisson and geometric case.

	$s=10$						$s=30$					
	binomial		Poisson		geometric		binomial		Poisson		geometric	
ρ	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)	(25)	(26)
0.5	3	3	3	3	29	28	1	1	1	1	32	32
0.6	4	4	5	5	60	59	2	2	2	2	67	66
0.7	8	8	10	10	139	137	3	3	4	4	154	153
0.8	20	20	27	26	406	402	7	7	10	9	448	444
0.9	98	96	135	133	2156	2136	33	33	46	45	2347	2330

6.1 On the impact of the distribution of A

The geometric distribution results in much higher values of $T(\hat{R}_m;\epsilon)$ than does the binomial and Poisson distribution. The reason for this is that the geometric distribution has a much heavier tail. To be more precise, the crucial quantity (as it appears in (21))

$$\frac{A(z_0(t))}{(z_0(t))^t} = \exp\Big(\min_{z\geq 1}[\ln A(z) - t\ln z]\Big), \tag{93}$$

is far larger for the geometric distribution. To give a comparison with a relatively light-tailed distribution, we introduce a fourth distribution of A.

Example 6.4. Subexponential case.

$$\mathbb{P}(A=n) = \frac{\theta^{2n}}{(2n)!\cosh\theta}, \quad n=0,1,\ldots, \tag{94}$$

and

$$A(z) = \frac{\cosh\theta\sqrt{z}}{\cosh\theta}, \quad A'(1) = \frac{1}{2}\theta\tanh\theta, \quad \phi(z) = \frac{1}{2}\theta\sqrt{z}\tanh(\theta\sqrt{z}), \tag{95}$$

$z_\infty = \infty$, $L_A = \infty$ and $z\phi'(z)$ increasing. Also, let $z_0^{(1)}(t) \equiv z_0(t)$ denote the solution of $\theta\sqrt{z}\tanh(\theta\sqrt{z}) = 2t$.

We denote by $z_0^{(2)}(t)$ and $z_0^{(3)}(t)$ the $z_0(t)$ for the Poisson and geometric case, respectively, i.e. $z_0^{(2)}(t) = t/\lambda$ and $z_0^{(3)}(t) = t/(p(1+t))$. Fig. 4 displays $z_0^{(i)}(t)$, $i=1,2,3$, for a common value $A'(1)$ of unity (i.e. $\theta=2.065$, $\lambda=1$, $p=1/2$) and $t=3$. The three heavy line segments above $z_0^{(i)}(t)$ indicate the difference between $\ln A(z)$ and $t\ln z$ at the minimizing $z=z_0(t)$, see (93). It is thus seen that the magnitude of the quantity in (93) strongly depends on the type of distribution. This effect becomes even more manifest when we increase s and keep ρ fixed, as we discuss in the next subsection.

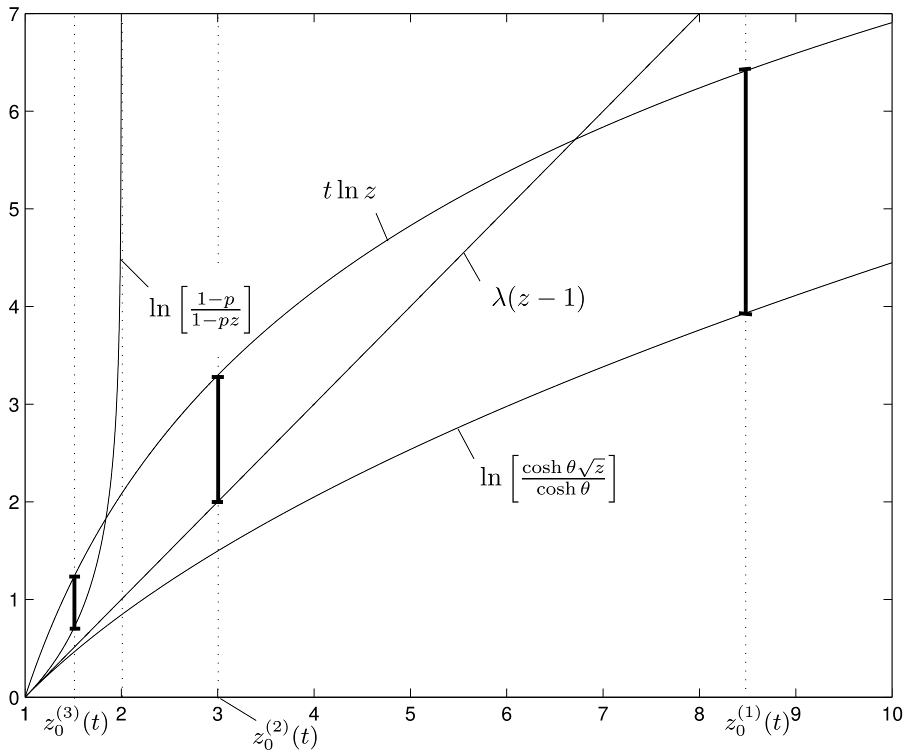

Figure 4: Picture of $z_0^{(i)}(t)$, $i = 1, 2, 3$, values of $z_0(t)$ for the subexponential, Poisson and geometric case, respectively, for a common value $A'(1) = 1$ and $t = 3$. The three heavy line segments above $z_0^{(i)}(t)$ indicate the difference between $\ln A(z)$ and $t \ln z$ at the minimizing $z = z_0(t)$, see (93).

6.2 On the effect of increasing s at fixed load ρ

We now compare the values of $T(\hat{R}_m;\epsilon)$ in Table 1, 2 and 3 for $s=10$ and $s=30$. Observe that by increasing s from 10 to 30, the values $T(\hat{R}_m;\epsilon)$ decrease for the binomial and Poisson distribution (and the geometric distribution for $m=0$), while for the geometric distribution the opposite is true (for $m=1,2$). Again this has to do with the heavier tail of the geometric distribution. In this case there holds, for $\alpha\geq 1$ and a fixed value of $\rho=A'(1)/s$,

$$\begin{aligned}\frac{A(z_0(\alpha s))}{(z_0(\alpha s))^{\alpha s}} &= \frac{1}{1+\rho s}\Big(1+\frac{1}{\rho s}\Big)^{-\alpha s}\Big(1+\frac{1}{\alpha s}\Big)^{\alpha s}(1+\alpha s)\\ &\to \frac{\alpha}{\rho}e^{1-\alpha/\rho},\quad s\to\infty.\end{aligned}\tag{96}$$

As a consequence, we see that it does not help to increase s to substantially decrease the value of the crucial quantity (93). In the Poisson case we find, for $\alpha\geq 1$ and a fixed value of $\rho=A'(1)/s$,

$$\frac{A(z_0(\alpha s))}{(z_0(\alpha s))^{\alpha s}}=\Big(\frac{\rho}{\alpha}e^{1-\rho/\alpha}\Big)^{\alpha s},\tag{97}$$

and this decays exponentially in s since $te^{1-t}<1$ for $t\in(0,1)$. Thus in the Poisson case it does pay to increase s. This observation continues to be valid for distributions with lighter tails, such as the binomial distribution or the distribution in Example 6.4.

A Appendix

A.1 The function ϕ

We consider the function $\phi(z)=zA'(z)/A(z)$, and we show that ϕ strictly increases on $[0,z_\infty)$ unless A is a monomial

Let $z>0$ and set

$$t:=\phi(z)=zA'(z)/A(z).\tag{98}$$

Using $A'(z)/A(z)=t/z$ we compute

$$\phi'(z)=\frac{A'(z)}{A(z)}+z\frac{A''(z)}{A(z)}-z\Big(\frac{A'(z)}{A(z)}\Big)^2=\frac{z^2A''(z)-t(t-1)A(z)}{zA(z)}\tag{99}$$

With $A(z) = \sum_{j=0}^{\infty} a_j z^j$ we can write

$$\begin{aligned} z^2 A''(z) - t(t-1)A(z) &= \sum_{j=0}^{\infty} (j(j-1) - t(t-1)) a_j z^j \\ &= \sum_{j=0}^{\infty} (j-t)(j+t-1) a_j z^j. \end{aligned} \tag{100}$$

Subtracting

$$0 = (2t-1)(zA'(z) - tA(z)) = (2t-1) \sum_{j=0}^{\infty} (j-t) a_j z^j \tag{101}$$

from either side of (100) we obtain

$$z^2 A''(z) - t(t-1)A(z) = \sum_{j=0}^{\infty} (j-t)^2 a_j z^j. \tag{102}$$

Hence

$$\phi'(z) = \frac{1}{zA(z)} \sum_{j=0}^{\infty} (j-t)^2 a_j z^j \geq 0. \tag{103}$$

There is equality in (103) only when t is a non-negative integer and $A(z) = z^t$.

As a consequence we have that (excluding the monomial case)

$$A'(1) = \phi(1) < \lim_{z \uparrow z_\infty} \phi(z) = L_A. \tag{104}$$

We observe furthermore from (103) that $\phi'(1) = \sigma_A^2$ since $t = A'(1) = \mathbb{E}A$ when $z = 1$.

Interestingly, one computes in a similar fashion as above that

$$(z\phi'(z))' = \frac{1}{zA(z)} \sum_{j=0}^{\infty} (j-t)^3 a_j z^j, \tag{105}$$

with t as in (98). This is of some relevance to the approximation made going from (54) to (55): one cannot assert monotonicity of $z\phi'(z)$ in general.

A.2 The function G

We consider the function

$$G(t) = \ln[A(z_0(t))/(z_0(t))^t], \quad t \geq A'(1), \tag{106}$$

where $z_0(t)$ is given by (48) for which we assume that A is not a polynomial. In terms of g_t in (39) we have $G(t) = g_t(z_0(t))$.

We compute for $t \in [A'(1), L_A)$ from

$$\phi(z_0(t)) = z_0(t)\frac{A'(z_0(t))}{A(z_0(t))} = t, \quad \phi'(z_0(t))z_0'(t) = 1 \tag{107}$$

that

$$G'(t) = \frac{A'(z_0(t))}{A(z_0(t))}z_0'(t) - t\frac{z_0'(t)}{z_0(t)} - \ln z_0(t) = -\ln z_0(t). \tag{108}$$

Note that the identity $G'(t) = -\ln z_0(t)$ continues to hold for $t \geq L_A$ by the definition of G, z_0 on $[L_A, \infty)$, and there holds $G'(t) = -\ln z_\infty$ for $t \geq L_A$.

When $t \in [A'(1), L_A)$ it furthermore follows from (107) and (108) that

$$G''(t) = -\frac{z_0'(t)}{z_0(t)} = \frac{-1}{z_0(t)\phi'(z_0(t))} < 0. \tag{109}$$

Observe that $\phi(1) = A'(1)$, whence $z_0(A'(1)) = 1$, and that $\phi'(1) = \sigma_A^2$. It then follows that

$$G(A'(1)) = 1, \quad G'(A'(1)) = 0, \quad G''(A'(1)) = -1/\sigma_A^2. \tag{110}$$

From (109) and (110) all further claims made about G in the main text follow.

References

[1] Abramowitz, M, I.A. Stegun (1970). *Handbook of Mathematical Functions*, ninth printing, Dover, New York.

[2] Abate, J., W. Whitt (1992). Numerical inversion of probability generating functions, *Operations Research Letters* **12**: 245-251.

[3] Ackroyd, M. (1980). Computing the waiting time distribution for the $G/G/1$ queue by signal processing methods, *IEEE Transactions on Communications* **28**: 52-58.

[4] Asmussen, S. (2003). *Applied Probability and Queues*, second edition, Springer, New York.

[5] Blanc, J.P.C., E.A. van Doorn (1984). Relaxation times for queueing systems. In: J.W. de Bakker, M. Hazewinkel, J.K. Lenstra (eds.), *Mathematics and Computer Science* (C.W.I. Monograph 1, North-Holland, Amsterdam), 139-162.

[6] Bruneel, H. (1984). A general model for the behavior of infinite buffers with periodic service opportunities. *European Journal of Operational Research* **16**: 98-106.

[7] Chaudhry, M.L., C.M. Harris, W.G. Marchal (1990). Robustness of rootfinding in single-server queueing models, *ORSA Journal on Computing* **3**: 273-286.

[8] Cohen, J.W. (1982). *The Single Server Queue*, North-Holland, Amsterdam.

[9] De Bruijn, N.G. (1981). *Asymptotic Methods in Analysis*, Dover Publications, New York.

[10] Erdélyi, A., W. Magnus, F. Oberheffinger, F.G. Tricomi (1953). *Higher transcendental functions*, Vol. I, McGraw-Hill, New York.

[11] Heathcote. C.R. (1967). Complete exponential convergence and some related issues. *Journal of Applied Probability* **4**: 217-256.

[12] Heathcote. C.R., P. Winer (1969). An aprroximation for the moments of waiting times. *Operations Research* **17**: 175-186.

[13] Janssen, A.J.E.M., J.S.H. van Leeuwaarden (2003). A discrete queue, Fourier sampling on Szegö curves, and Spitzer's formula. *Eurandom report series* 018; submitted for publication.

[14] Konheim, A.G. (1975). An elementary solution of the queueing system G/G/1. *SIAM Journal on Computing* **4**: 540-545.

[15] Lawden, D.F. (1951). The function $\sum_{n=1}^{\infty} n^r z^n$ and associated polynomials. *Proc. Cambridge Phil. Soc.* **47**: 309-314.

[16] Murata, M., H. Miyahara (1991). An analytic solution of the waiting time distribution for the discrete-time $G/G/1$ queue. *Performance Evaluation* **13**: 87-95.

[17] Servi, L.D. (1986). $D/G/1$ queues with vacations. *Operations Research* **34**: 619-629.

[18] Spitzer, F.L. (1956), A combinatorial lemma and its application to probability theory, Trans. Amer. Math. Soc. **82**: 323-339.

[19] Stadje, W. (1997). A new approach to the Lindley recursion. *Statistics & Probability Letters* **31**: 169-175.

[20] Stalley, R. (1949). A generalization of the geometric series. *Am. Math. Monthly* **56**: 325-327.

[21] Temme, N.M. (1996). *Special Functions*, Wiley, New York.

[22] Titchmarsh, E.C. (1939). *The Theory of Functions*, 2^{nd} edition, Oxford University Press, New York.

[23] Wall, H.S. (1948). *Continued Fractions*, Van Nostrand, Princeton.

[24] Whittaker, E.T., G.N. Watson (1963). *A Course of Modern Analysis*, 4^{th} edition, Cambridge University Press, Cambridge.