Loss systems with slow retrials in the Halfin-Whitt regime

F. Avram* A.J.E.M. Janssen† J.S.H. van Leeuwaarden‡

June 6, 2011

Abstract

The Halfin-Whitt regime, or Quality-and-Efficiency-Driven (QED) regime, for multi-server systems refers to a situation with many servers, a critical load, and yet favorable system performance. We apply this regime to the classical multi-server loss system with slow retrials. We derive non-degenerate limiting expressions for the main steady-state performance measures including the retrial rate and the blocking probability. It is shown that the economies-of-scale associated with the QED regime persists for systems with retrials, although in situations when the load becomes *extremely* critical the retrials cause deteriorated performance, and the system starts behaving as in the Efficiency-Driven (ED) regime instead of the QED regime. Most of our results are obtained by a detailed analysis of *Cohen's equation* that defines the retrial rate in an implicit way. The limiting expressions are established by studying pre-limit behavior and exploiting the connection between Cohen's equation and Mills' ratio for the Gaussian and Poisson distribution.

Keywords: retrial system, loss system, Erlang B model, Cohen's equation, Mills' ratio, Halfin-Whitt regime, QED regime

AMS 2000 Subject Classification: 60K25, 68M10, 41A60.

*Département de Mathématiques, Avenue de l'Université - BP 1155, 64013 Pau Cedex, France. E-mail: florin.avram@univ-pau.fr

†Eindhoven University of Technology and Eurandom, Department of Mathematics and Computer Science and Department of Electrical Engineering, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail a.j.e.m.janssen@tue.nl

‡Eindhoven University of Technology, Department of Mathematics and Computer Science, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. E-mail j.s.h.v.leeuwaarden@tue.nl

1 Introduction

Customers of call centers that obtain a busy signal usually repeat calls until the required connection is made. A call center is therefore facing two flows of incoming calls: *primary calls* of those customers attempting for the first time, and *repeated calls* generated by previously blocked customers. Such processes can be studied using retrial systems. It is widely accepted that the phenomenon of repeated calls, in which customers keep calling until being successful, is one of the crucial factors for call center performance. In this paper, we investigate the basic multi-server loss system with repeated calls, or retrials, and study this system in a regime with many servers under heavy-traffic conditions. The modeling of retrials is quite challenging, see e.g. [5, 6], which is why one often resorts to computational approaches [2]. These numerical approaches face increasing numerical difficulties when the number of servers becomes large, which is precisely the regime we are interested in. Therefore, we combine a many-server regime with a limit theorem of Cohen [5] for slow retrials, meaning that blocked customers repeat their calls only after a relatively (compared to the time scale of the system) long time. The combination of these two asymptotic regimes leads to a tractable model.

There is by now a vast literature on the asymptotic analysis of multi-server systems, in which a finite-size system is seen as one in a sequence of systems, and the limiting behavior of this sequence is used to approximate the performance of this finite-size system. Depending on how this sequence is parameterized, its limiting behavior is different, giving rise to different approximations [4]. Among the most effective approximations arises in the Quality-and-Efficiency-Driven (QED) regime, in which the number of servers s and the offered workload λ are related according to a square-root principle, namely $\lambda = s - \gamma\sqrt{s}$ for some fixed constant γ. The latter is asymptotically equivalent with setting $s = \lambda + \beta\sqrt{\lambda}$ (square-root staffing) for some fixed constant β. Square-root staffing and the QED limiting regime for multi-server systems (without retrials) were brought to the center of attention by the work of Halfin and Whitt [7], and therefore the QED regime goes also by the name Halfin-Whitt regime.

In this paper we consider the multi-server loss system ($M/M/s/s$ queue) with retrials. We analyse this system in the Halfin-Whitt regime, in a similar spirit as was done for the $M/M/s$ queue [7, 9], the $M/M/s/s$ queue [8] and the Erlang A model ($M/M/s$ queue with abandonments) [13]. Compared to these earlier studies, the system with retrials brings about additional mathematical challenges, mainly because the retrial rate of returning customers is given implicitly as the solution of what we call *Cohen's equation*; cf. (2.1). In short, we make the following contributions:

(i) Within the realm of the Halfin-Whitt regime, the retrial phenomenon was relatively unexplored. This paper presents the first analytical results in this direction. We derive new QED approximations for the retrial rate and the blocking probability. We show that the additional arrival rate due to retrials is of the same order as the overcapacity: both are $O(\sqrt{s})$. Therefore, this additional load on the system can cause serious capacity problems, causing the system's behavior to become much less favorable than perhaps expected in the Halfin-Whitt regime. We further investigate the rate of convergence to the limiting regime by undertaking an in-depth study of the pre-limit or true retrial rate. It is shown that the difference between the true retrial rate and its QED approximation diminishes rapidly as a function of system size, which provides evidence for the appropriateness of square-root staffing for call centers, even in the case of retrials.

(ii) A major effort is put in the study of Cohen's equation and the analysis of its solution (retrial rate), both for the case that s is finite and the limiting form of this equation and solution when $s \to \infty$ as in the Halfin-Whitt regime. In the latter case, Cohen's equation comprises the well-known Mills' ratio of the Gaussian distribution. Existence and uniqueness of the solution of Cohen's equation follows from monotonicity results of this Mills' ratio as given by Sampford [11]. For the case of finite s, Cohen's equation comprises a ratio of Mills type as well, and a sizeable part of this paper is devoted to the case s is fixed. This yields Cohen's existence and uniqueness result for his equation with finite s, as well as analytic and asymptotic results for the solution as $\gamma \downarrow 0$.

(iii) When the overcapacity given by $\gamma\sqrt{s}$ becomes small ($\gamma \downarrow 0$), the retrial rate grows as $\sqrt{s}/\gamma$ and completely dominates the system's performance. The case $\gamma \downarrow 0$ can be interpreted as a *double* heavy-traffic limit, in the sense that we not only let $\rho = \lambda/s = 1 - \gamma/\sqrt{s}$ approach one by making s large, but also by making γ small. The resulting extremely heavily loaded system then, at some point, tunnels from the QED regime to the Efficiency-Driven (ED) regime (see [10, 12] for more background on these regimes). While the blocking probability in the QED regime approaches zero as s becomes large, in the ED regime the blocking probability approaches one. We present several results that help understand this crucial transition between regimes.

Section 2 introduces the multi-server system with slow retrials. Section 3 presents the main results, and all proofs are given in Section 4.

2 Description of retrial system

We now describe the classical multi-server loss system with retrials (see e.g. [6, Chapter 2] and [2]). Consider a group of s servers to which calls arrive according to a Poisson process with rate λ. These calls are referred to as *primary calls*. A primary call that finds, upon arrival, a free server, immediately occupies this server and leaves the system after service. If all servers are occupied, the blocked primary call leaves the system but reattempts to obtain service after some time. Hence, each blocked primary call starts producing retrials until it is served.

Assume that periods between successive retrials are exponentially distributed with mean $1/\mu$, service times are exponentially distributed with mean one, and interarrival times, service times and retrial times are mutually independent. The system state can then be described by means of a bivariate process $\{(C(t), N(t)); t \geqslant 0\}$ with $C(t)$ the number of occupied servers and $N(t)$ the number of retrial sources at time t. Under the above assumptions this process is a continuous-time Markov chain on the lattice infinite strip $\{0, 1, \ldots, s\} \times \{0, 1, \ldots\}$.

Since the transition rates of this process clearly depend on the second coordinate, even deriving the stationary distribution poses analytical difficulties, and no closed-form solutions exist for cases with more than four servers. Due to the lack of analytical formulas for the main performance measures, limit theorems fulfill an important role in understanding the influence of the repeated attempts in some domains of the system parameters.

The main result in this direction was obtained by Cohen [5] who showed that the retrial queue, in the limit as $\mu \downarrow 0$, behaves as an Erlang loss system, except with an increased arrival intensity. More specifically, for the $M/M/s/s$ loss system with retrials, as $\mu \downarrow 0$, the steady-state distribution of the number of busy servers converges to the corresponding distribution of the standard Erlang loss system $M/M/s/s$ (which is a truncated Poisson distribution), but with increased arrival rate

$\lambda + \Omega$, where Ω is the unique positive root of the polynomial equation

$$\Omega = (\lambda + \Omega)B(s, \lambda + \Omega). \tag{2.1}$$

Here $B(s, \lambda)$ is the Erlang B formula, representing the steady-state blocking probability in the Erlang loss system, and given by

$$B(s, \lambda) \;=\; \frac{\lambda^s/s!}{\sum_{k=0}^{s} \lambda^k/k!} = \frac{\mathrm{e}^{-\lambda}(\lambda/s)^s}{\int_{\lambda}^{\infty} \mathrm{e}^{-\lambda'}(\lambda'/s)^s d\lambda'} \tag{2.2}$$

with $\lambda > 0$ and $s = 1, 2, \ldots$. The form in (2.2) comprising the integral allows us to consider $B(s, \lambda)$ for arbitrary $s > 0$.

Equation (2.1), written as $\lambda = (\lambda + \Omega)(1 - B(s, \lambda + \Omega))$, is intuitively clear as it expresses equality of arrivals and carried traffic. However, in order for this heuristic to be justified, one needs to assume that the flow of repeated calls does not depend on the flow of primary calls. In that case, the total flow of calls is a Poisson process with rate $\lambda + \Omega$, a fact that was rigorously proved to be true when $\mu \downarrow 0$ by Cohen [5]. Indeed, in the case of infinitely long retrial times, it seems plausible that the flow of repeated calls is independent from the flow of primary calls. For retrial systems with finite retrial times, the independence assumption on the two arrival processes gave rise to the so-called constant retrial rate approximation, which has proved useful for many retrial systems (see [2]).

The additional arrival rate Ω can be thought of as the load formed by the sources of repeated calls. This result shows that it is important to distinguish between the cases $\mu = 0$ and $\mu \downarrow 0$. If $\mu = 0$, then the blocked customers are lost (do not send repeated attempts at all) and the retrial queue becomes the standard Erlang loss system with the same arrival rate λ and stationary distribution

$$p_k(0) = \frac{\lambda^k/k!}{\sum_{k=0}^{s} \lambda^k/k!}, \quad k = 0, 1, \ldots, s. \tag{2.3}$$

In contrast, if $\mu \downarrow 0$, then the retrial model in steady state can be viewed as the standard Erlang loss system but with the increased arrival rate $\lambda + \Omega$. The limit behavior of retrial queues as $\mu \downarrow 0$ is of interest on account of the weak dependence of the stationary distribution $\{p_i(\mu); 0 \leqslant i \leqslant s\}$ of the number of busy servers. Because $\lim_{\mu \to 0} p_i(\mu)$ has a beautiful closed-form solution, it is common practice to use this limit as an approximation of $p_i(\mu)$ for all $\mu > 0$ (see [2]). The results presented in the next section are all for this limiting regime of slow retrials.

3 Main results and their implications

We have divided our contributions into three parts. In Subsection 3.1 we present new QED approximations for the retrial rate Ω and the blocking probability $B(s, \lambda + \Omega)$ of the retrial system in the Halfin-Whitt regime. In Subsection 3.2 we present a series of results for Ω, both in the case of finite s and infinite s (Halfin-Whitt regime). In Subsection 3.3 we give several new results for the key function that governs Cohen's equation (2.1). As it turns out, this key function is a slight adaptation of the Erlang B formula that can be interpreted in terms of Mills' ratio for the Poisson distribution. Hence, all results presented in Subsection 3.3 are in fact new results for the Erlang B formula and Mills' ratio for the Poisson distribution. The proofs of all results are given in Section 4.

3.1 Halfin-Whitt regime

The Halfin-Whitt regime for multi-server systems refers to the scaling of the arrival rate λ and the numbers of servers s such that, while both λ and s increase toward infinity, the traffic intensity $\rho = \lambda/s$ approaches one and

$$(1 - \rho)\sqrt{s} \to \gamma, \tag{3.1}$$

where γ is a fixed constant. The scaling combines large capacity with high utilization. For the Erlang loss system, this kind of scaling leads to the following classical result due to Erlang (see e.g. [8]).

Lemma 1. *For $\lambda = s - \gamma\sqrt{s}$, with $\gamma < \sqrt{s}$ fixed,*

$$\lim_{s\to\infty} \sqrt{s}B(s, \lambda) = \frac{\phi(\gamma)}{\Phi(\gamma)}, \tag{3.2}$$

where $\Phi(x)$ and $\phi(x)$ denote the standard normal cumulative distribution function and density, respectively.

We now apply the same scaling to the multi-server retrial system, for which we need to assume that $\gamma \in (0, \sqrt{s})$ because of the stability condition $\lambda < s$.

Theorem 2. *For $\lambda = s - \gamma\sqrt{s}$, with $\gamma \in (0, \sqrt{s})$ fixed, and Ω defined as in (2.1),*

$$\lim_{s\to\infty} \frac{\Omega}{\sqrt{s}} = a \tag{3.3}$$

and

$$\lim_{s\to\infty} \sqrt{s}B(s, \lambda + \Omega) = \frac{\phi(\gamma - a)}{\Phi(\gamma - a)} \tag{3.4}$$

with $a = a_\infty(\gamma)$ the unique positive solution to

$$a = \frac{\phi(\gamma - a)}{\Phi(\gamma - a)}. \tag{3.5}$$

Theorem 2 shows that the additional load Ω, for a system with many servers, is of the order $\sqrt{s}$. In particular, as the number of servers grows large, Ω is well approximated by $a\sqrt{s}$, where a is a constant that no longer depends on s. This also means that for the overall retrial system the arrival rate $\lambda + \Omega$ is approximately $s - (\gamma - a)\sqrt{s}$, which gives (3.4). Theorem 2 thus says that the blocking probability in the retrial system in the Halfin-Whitt regime is $O(1/\sqrt{s})$. When the number of servers is large enough, the blocking probability is well approximated by some constant divided by $\sqrt{s}$, where the constant a only depends on γ. This then gives the QED approximations

$$\Omega \approx a\sqrt{s}, \quad B(s, \lambda + \Omega) \approx \frac{\phi(\gamma - a)}{\sqrt{s}\Phi(\gamma - a)}. \tag{3.6}$$

Theorem 2 follows from the more general Theorem 15 presented below. The key idea behind the proof of Theorem 2 is the following. Writing $\Omega = a\sqrt{s}$ and using (2.1) gives

$$\begin{aligned} \lim_{s\to\infty} \frac{\Omega}{\sqrt{s}} &= \lim_{s\to\infty} \frac{s - (\gamma - a)\sqrt{s}}{\sqrt{s}} B(s, s - (\gamma - a)\sqrt{s}) \\ &= \lim_{s\to\infty} \sqrt{s}B(s, s - (\gamma - a)\sqrt{s}), \end{aligned} \tag{3.7}$$

so that the result follows from (3.2). Notice that in (3.7) we ignore the fact that, for finite s, the factor a is not only a function of γ (through (3.5)) but also of s. Therefore, the steps in (3.7) are only serving as a heuristic. The formal proof of Theorem 2 shall take into account this dependence on s and starts from a transformed version of (2.1) that is easier to work with in the Halfin-Whitt regime. Using $\lambda = s - \gamma\sqrt{s}$ with $\gamma < \sqrt{s}$ (negative values of γ are allowed) and $\Omega = a\sqrt{s}$ with $a > 0$, Equation (2.1) is turned into

$$a = f_s(\gamma - a), \tag{3.8}$$

in which, for $\delta < \sqrt{s}$

$$f_s(\delta) := (1 - \delta/\sqrt{s})g_s(\delta) := \sqrt{s}(1 - \delta/\sqrt{s})B(s, s - \delta\sqrt{s}). \tag{3.9}$$

See Figure 3 for an illustration of the function f_s.

Theorem 3 (Cohen [5])**.** *For $\gamma \in (0, \sqrt{s})$ the equation (3.8) has a unique solution $a > 0$.*

We give a separate proof of Theorem 3 in Section 4. We denote the unique positive solution of (3.8) by $a_s(\gamma)$ and refer to it as *retrial factor*, because $\Omega = a_s(\gamma)\sqrt{s}$. In order to understand the behavior of our retrial system we need to understand the dependencies of $a_s(\gamma)$ on s and γ (Subsection 3.2). Since $a_s(\gamma)$ is defined implicitly in (3.8), it is crucial to study the function f_s (Subsection 3.3).

3.2 Properties of the retrial factor

We now present several results for the retrial factor $a_s(\gamma)$.

Theorem 4. *$a_s(\gamma) : (0, \sqrt{s}) \to (0, \infty)$ is a positive, decreasing and convex function of $\gamma \in (0, \sqrt{s})$.*

See Figure 1. Theorem 4 can be understood by interpreting γ as the inverse load on the system. Indeed, the load is given by $1 - \gamma/\sqrt{s}$ and hence decreases from one for $\gamma = 0$ to zero for $\gamma = \sqrt{s}$. We would expect the retrial factor to increase with the load, since an increased load leads to more blocked calls.

The next result describes the asymptotic behavior of the retrial factor in heavy traffic (when γ is small).

Theorem 5. *For $s \geqslant 1$ and $0 < \gamma < \sqrt{s}$,*

$$a_s(\gamma) = \frac{1}{\gamma} - \frac{2}{\sqrt{s}} - \Big(1 - \frac{2}{s}\Big)\gamma + \varepsilon_a \tag{3.10}$$

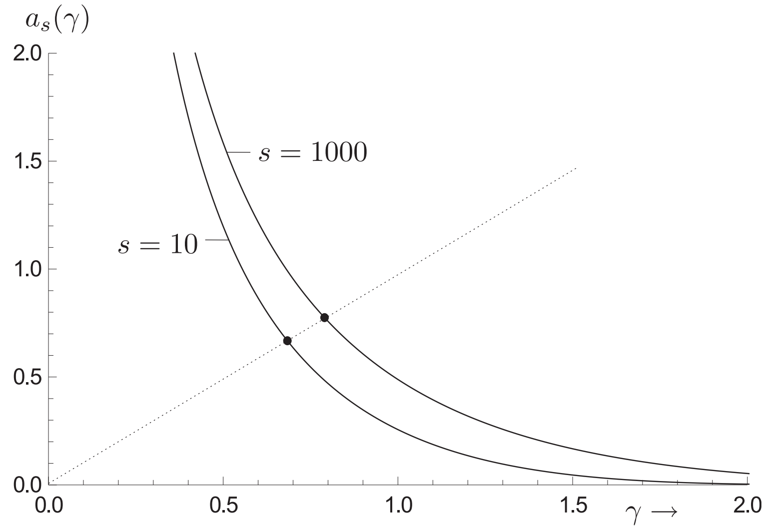

Figure 1: The function $a_s(\gamma)$ for $s = 10$ and $s = 1000$; the • indicates the point $(\gamma_s^*, a_s(\gamma_s^*))$, see Proposition 7.

with $\varepsilon_a = O(\gamma^2/\sqrt{s}) + O(\gamma^3)$.

The approximation $a_s(\gamma) \approx 1/\gamma$ can be anticipated from Theorem 2 and the well-known result

$$\frac{\phi(\gamma)}{\Phi(\gamma)} = -\gamma + O(\gamma^{-1}), \quad \gamma \to -\infty. \tag{3.11}$$

To identify the higher-order terms in (3.10), and to understand better the behavior of the retrial rate a_s, a detailed study of the function f_s is required (Subsection 3.3). A comparison of f_s and Mills' ratio is presented is Subsection 3.3.

We next complement the asymptotic result in (3.11) with some basic inequalities satisfied by $a_s(\gamma)$.

Proposition 6. *For* $0 < \gamma < \sqrt{s}$, $\gamma a_s(\gamma)$ *is a monotonically decreasing function, and*

$$\frac{1}{\gamma} - \frac{2}{\sqrt{s}} - \gamma < a_s(\gamma) < \frac{1}{\gamma} - \frac{1}{\sqrt{s}}. \tag{3.12}$$

See Figure 2. Notice that the additional arrival rate $a_s(\gamma)\sqrt{s}$ due to retrials is of the same order as the overcapacity $\gamma\sqrt{s}$, and the additional arrivals start causing serious capacity problems when $a_s(\gamma) > \gamma$, and in particular when $\gamma \downarrow 0$. Indeed, for large but fixed s, and γ approaching zero, the system makes the transition from the QED regime tot the ED regime. In the QED regime, the blocking probability is $O(1/\sqrt{s})$ as in (3.6), while in the ED regime the blocking probability approaches one. To see this we can use the fact that $a_s(\gamma)$ tends to infinity, like $1/\gamma - 2/\sqrt{s}$, and hence, for fixed s and $\gamma \downarrow 0$ ($a \to \infty$),

$$B(s, s-(\gamma-a)\sqrt{s}) = \frac{1}{a-\gamma+\sqrt{s}} f_s(\gamma-a)$$
$$= 1 - \frac{\sqrt{s}}{a-\gamma} + O\Big(\Big(\frac{1}{a-\gamma}\Big)^2\Big), \tag{3.13}$$

where we also use, see Theorem 10, that $f_s(\delta) = -\delta + O(1/\delta)$, $\delta \to -\infty$. In comparing $a_s(\gamma)$ and γ we have the following result:

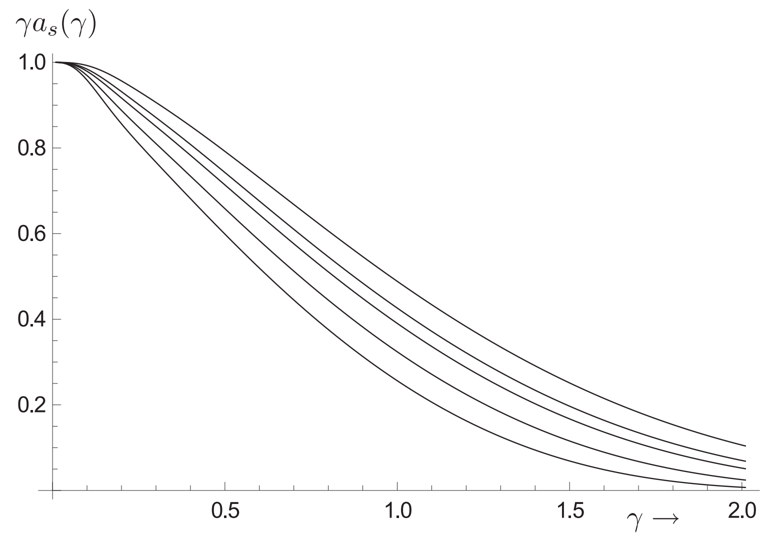

Figure 2: The function $\gamma a_s(\gamma)$ for $s = 10, 20, 50, 100, 1000$ (ordered upwards).

Proposition 7. *Let* $\gamma_s^* = f_s(0) = \sqrt{s}B(s,s)$. *Then,*

- $a_s(\gamma) = \gamma$ *when* $\gamma = \gamma_s^* \in (0, \sqrt{s})$.
- $a_s(\gamma) > \gamma$ *when* $\gamma \in (0, \gamma_s^*)$.
- $a_s(\gamma) < \gamma$ *when* $\gamma \in (\gamma_s^*, \sqrt{s})$.

Sharp bounds exist for γ_s^* (cf. (3.24)) and from (3.2) we see that

$$\gamma_s^* \to \gamma_\infty^* = \frac{\phi(0)}{\Phi(0)} = \frac{2}{\sqrt{2\pi}} = 0.79788\dots, \quad s \to \infty. \tag{3.14}$$

The next result describes the retrial factor in light-traffic conditions (when γ is large).

Theorem 8. *For $s \geqslant 1$ and $\gamma \uparrow \sqrt{s}$,*

$$a_s(\gamma) = \frac{s^{s+1/2}}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s+1}(1 + \Delta_a) \tag{3.15}$$

with $\Delta_a = O(\sqrt{s}\mathrm{e}^s(1 - \gamma/\sqrt{s})^s) + O(1 - \gamma/\sqrt{s})$.

For the computation of $a_s(\gamma)$ a simple Newton iteration works quite well due to convexity of f_s (see Theorem 11 below). When γ is not too small initialization can be taken as $a^{(0)} = 0$. When γ is close to 0, the initialization should be done using (3.10). In all cases, convergence is monotonic after one step. The Newton step

$$a^{(n+1)} = a^{(n)} - \frac{a^{(n)} - f_s(\gamma - a^{(n)})}{1 - f_s'(\gamma - a^{(n)})} \tag{3.16}$$

is implemented conveniently using

$$f_s'(\delta) = -\frac{f_s(\delta)}{1 - \delta/\sqrt{s}}\Big(\delta + \frac{1}{\sqrt{s}} + f_s(\delta)\Big). \tag{3.17}$$

Theorem 9. *$a_s(\gamma)$, with $\gamma \in (0, \sqrt{s})$ fixed, increases monotonically in $s \geqslant 1$.*

See Figure 1. There seems no easy explanation for Theorem 9. Note that the traffic intensity, with $\gamma \in (0, \sqrt{s})$ fixed,

$$\rho = \frac{\lambda}{s} = 1 - \gamma/\sqrt{s} \tag{3.18}$$

increases monotonically in s, but there is no obvious ordering between two systems (indexed by s), making a stochastic comparison difficult.

3.3 The function f_s and Mills' ratio

The analysis of equation (3.8) and f_s is conducted with in mind known results for the with Mills' ratio related quantity

$$f_\infty(\delta) = \frac{\mathrm{e}^{-\delta^2/2}}{\int_{-\infty}^{\delta} \mathrm{e}^{-(\delta')^2/2} d\delta'} = \frac{\phi(\delta)}{\Phi(\delta)}, \quad \delta \in \mathbb{R}, \tag{3.19}$$

of the normal distribution. The quantity in (3.19) is actually the reciprocal of Mills' ratio evaluated at $-\delta$, see [11, 3]. Let

$$\alpha_s(\delta) = \Big(-2s\big(\delta/\sqrt{s} + \ln(1-\delta/\sqrt{s})\big)\Big)^{1/2}, \quad \delta < \sqrt{s}, \tag{3.20}$$

where the square root is chosen such that $\mathrm{sgn}(\alpha_s(\delta)) = \mathrm{sgn}(\delta)$. Using this in the integral form in (2.2) with $\lambda = s - \delta\sqrt{s}$ and the substitution $\lambda' = s - \delta'\sqrt{s}$, yields

$$f_s(\delta) = \frac{(1-\delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}}{\int_{-\infty}^{\delta} \mathrm{e}^{-\alpha_s^2(\delta')/2} d\delta'}, \quad \delta < \sqrt{s}. \tag{3.21}$$

It is convenient to define $f_s(\delta) = 0$ for $\delta \geqslant \sqrt{s}$. From

$$\alpha_s(\delta) = \delta\Big(1 + \frac{2\delta}{3\sqrt{s}} + \frac{\delta^2}{2s} + \dots\Big)^{1/2} \tag{3.22}$$

it follows that $f_s(\delta)$ converges pointwise to $f_\infty(\delta)$ as $s \to \infty$.

One can interpret f_s as a version of Mills' ratio of the Poisson distribution. Indeed, let $\mathrm{Pois}(\lambda)$ denote a Poisson random variable with mean λ. Then, for $\lambda = s - \delta\sqrt{s}$ with $0 < \delta < \sqrt{s}$ fixed,

$$f_s(\delta) = \frac{\lambda}{\sqrt{s}} \frac{\mathbb{P}\left(\mathrm{Pois}(\lambda) = s\right)}{\mathbb{P}\left(\mathrm{Pois}(\lambda) \leqslant s\right)} = \Big(1 - \frac{\delta}{\sqrt{s}}\Big)\frac{\phi(\alpha_s(\delta))}{\Phi(\alpha_s(\delta))} \to f_\infty(\delta), \qquad s \to \infty. \tag{3.23}$$

It turns out that including the factor $(1-\delta/\sqrt{s})$ in the definition of $f_s(\delta)$ turns $f_s(\delta)$ into a function with monotonicity and asymptotic properties of similar nature as those possessed by $f_\infty(\delta)$ in (3.19). Hence, in this sense, one should consider $f_s(\delta)$, and not $g_s(\delta)$ of (3.9), in connection with Mills' ratio.

In [8] the quasi-Gaussian form in (3.21) was exploited to derive a whole arsenal of asymptotic expansions and bounds for $f_s(\delta)$ and the closely related Erlang B formula. For instance, one result from [8] is

$$\Big(\frac{1}{2}\sqrt{2\pi} + \frac{2}{3\sqrt{s}} + \frac{1}{24s}\sqrt{2\pi}\Big)^{-1} < f_s(0) < \Big(\frac{1}{2}\sqrt{2\pi} + \frac{2}{3\sqrt{s}} + \frac{1}{24s}\sqrt{2\pi} - \frac{4}{135s\sqrt{s}}\Big)^{-1}. \tag{3.24}$$

In the present paper, much of the interest is for the situation when $\delta \to -\infty$, and for this purpose the results in [8] that are best when δ stays bounded are less useful. We therefore derive a set of new

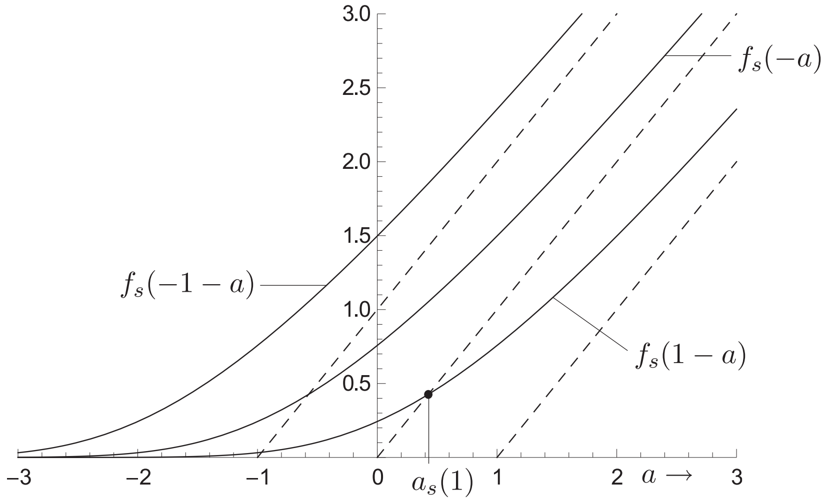

Figure 3: The function $f_s(\gamma - a)$ for $s = 100$ and $\gamma = -1, 0, 1$.

results for $f_s(\delta)$, again using the quasi-Gaussian form, but now our manipulations are specifically geared at the regime $\delta \to -\infty$.

Theorem 10. *For $\delta < 0$ and $s \geqslant 1$,*

$$f_s(\delta) = -\delta - \frac{1}{\delta} - \frac{2}{\delta^2\sqrt{s}} + \left(2 - \frac{6}{s}\right)\frac{1}{\delta^3} + \varepsilon_f \tag{3.25}$$

with $\varepsilon_f = O(\delta^{-4}s^{-1/2}) + O(\delta^{-5})$.

Theorem 11. *For $\delta < \sqrt{s}$,*

$$f_s(\delta) > -\delta, \quad f_s'(\delta) > -1, \quad f_s''(\delta) > 0. \tag{3.26}$$

The last two inequalities in (3.26) are equivalent with the following upper and lower bound for $f_s(\delta)$.

Corollary 12. *For $\delta < \sqrt{s}$,*

$$L(\delta) < f_s(\delta) < U(\delta), \tag{3.27}$$

where

$$L(\delta) = -\frac{3}{4}\delta - \frac{1}{2\sqrt{s}} + \frac{1}{4}\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2}, \tag{3.28}$$

$$U(\delta) = -\frac{1}{2}\delta - \frac{1}{2\sqrt{s}} + \frac{1}{2}\Big(\Big(\delta - \frac{1}{\sqrt{s}}\Big)^2 + 4\Big)^{1/2}. \tag{3.29}$$

Theorem 13. *$f_s(\delta)$, with $\delta \in (-\infty, \sqrt{s})$ fixed, increases monotonically in $s \geqslant 1$.*

We now list results for f_∞ of the type we have obtained for f_s above. We have

$$f_\infty(\delta) > -\delta, \quad f'_\infty(\delta) > -1, \quad f''_\infty(\delta) > 0, \quad \delta \in \mathbb{R}, \tag{3.30}$$

and

$$-\frac{3}{4}\delta + \frac{1}{4}(\delta^2 + 8)^{1/2} < f_\infty(\delta) < -\frac{1}{2}\delta + \frac{1}{2}(\delta^2 + 4)^{1/2}. \tag{3.31}$$

These results were obtained by Sampford [11]. From [1, 7.1.13] we get the sharper lower bound $-\frac{1}{2}\delta + \frac{1}{2}(\delta^2 + 8/\pi)^{1/2}$ for $f_\infty(\delta)$ with equality at $\delta = 0$. The (non-strict versions of the) inequalities in (3.30) and (3.31) follow also from Theorem 11 and Corollary 12 by letting $s \to \infty$.

The properties of $f_\infty(\delta)$ of course play a crucial role in the QED approximations obtained in Theorem 2, where in (3.5) we have defined $a = a_\infty(\gamma)$ as the solution of

$$a = f_\infty(\gamma - a). \tag{3.32}$$

The next theorem collects several results regarding (3.32).

Theorem 14. *For any $\gamma > 0$ fixed, equation (3.32) has a unique positive solution $a_\infty(\gamma)$. Moreover,*

$$f_\infty(\delta) = -\delta - \frac{1}{\delta} + \frac{2}{\delta^3} - \frac{10}{\delta^5} + \frac{74}{\delta^7} + O(\delta^{-9}), \quad \delta < 0, \tag{3.33}$$

and

$$a_\infty(\gamma) = \frac{1}{\gamma} - \gamma + 2\gamma^3 - 10\gamma^5 + 82\gamma^7 + O(\gamma^9), \quad \gamma > 0. \tag{3.34}$$

Comparing Theorem 14 and Theorems 10 and 5 shows that the series in (3.33) and (3.34) omit all even powers of δ and γ, respectively. These even powers do occur in (3.25) and (3.10), but the coefficients decay, as far as we can see, with a deviation of $O(1/\sqrt{s})$ with the corresponding coefficients in (3.33) and (3.34). We further see that

$$f_\infty(\delta) - f_s(\delta) = \frac{2}{\delta^2\sqrt{s}} + \frac{6}{\delta^3 s} + O(s^{-1/2}\delta^{-4}) + O(\delta^{-5}), \quad \delta \to -\infty, \tag{3.35}$$

$$a_\infty(\gamma) - a_s(\gamma) = \frac{2}{\sqrt{s}} - \frac{2\gamma}{s} + O(\gamma^2 s^{-1/2}) + O(\gamma^3), \quad \gamma \downarrow 0. \tag{3.36}$$

This would suggest that $f_s(\delta) \to f_\infty(\delta)$, $a_s(\gamma) \to a_\infty(\gamma)$ at a uniform $1/\sqrt{s}$–rate when $\delta < 0$, $0 < \gamma < \sqrt{s}$, respectively. However, we have not been able to find a proof of such a statement. Below we give the best result we have found in this respect.

Theorem 15.

$$f_\infty(\delta) - f_s(\delta) = O(1/\sqrt{s}), \quad a_\infty(\gamma) - a_s(\gamma) = O(1/\sqrt{s}), \tag{3.37}$$

in which the O's hold uniformly in any set $\delta_0 \leqslant \delta$, $\gamma \geqslant \gamma_0$ where $\delta_0 > -\infty$ and $\gamma_0 > 0$.

Note that Theorem 15 implies Theorem 2.

4 Proofs

4.1 Proof of Theorem 10

Let

$$I_s(\delta) = \int_{-\infty}^{\delta} \mathrm{e}^{-\frac{1}{2}\alpha_s^2(\delta')} d\delta'. \tag{4.1}$$

We want to find the asymptotic behavior of $I_s(\delta)$ as $\delta \to -\infty$ and $s \geqslant 1$, and to this end we adopt the approach of [8]. That is, we bring $I_s(\delta)$ into quasi-Gaussian form, see [8, Sec. 2], by letting $y(x)$ for $x \in \mathbb{R}$ be the solution $y \in (-\infty, 1)$ of the equation

$$-y - \ln(1-y) = \tfrac{1}{2}x^2, \tag{4.2}$$

where we take $y = y(x)$ such that $\mathrm{sgn}(y(x)) = \mathrm{sgn}(x)$. With this y, we get by the substitution $x = \alpha(\delta')$,

$$y(\alpha(\delta')/\sqrt{s}) = \delta'/\sqrt{s}, \quad y'(\alpha(\delta')/\sqrt{s})\alpha'(\delta') = 1 \tag{4.3}$$

so that

$$I_s(\delta) = \int_{-\infty}^{\alpha(\delta)} \mathrm{e}^{-\frac{1}{2}x^2} y'(x/\sqrt{s})dx, \quad \delta < \sqrt{s}. \tag{4.4}$$

An asymptotic result as $\delta \to -\infty$ can now be obtained for $I_s(\delta)$ by repeated partial integration and using (4.3) together with

$$\frac{1}{x}y'(x) = \frac{1}{y(x)} - 1. \tag{4.5}$$

Due to the properties (4.3) and (4.5), the administration caused by these partial integrations remains within reasonable bounds.

Lemma 16. *For $k = 0, 1, \dots$ and $\alpha < 0$,*

$$\int_{-\infty}^{\alpha} \mathrm{e}^{-\frac{1}{2}x^2} y'(x/\sqrt{s})dx = \sum_{l=0}^{k} E_l(\alpha) + \int_{-\infty}^{\alpha} \frac{d}{dx}\left[\left(\frac{1}{x}\frac{d}{dx}\right)^k \left(\frac{1}{x}y'(x/\sqrt{s})\right)\right] \mathrm{e}^{-x^2/2}dx. \tag{4.6}$$

Here

$$E_0(\alpha) = -\frac{1 - y(\alpha/\sqrt{s})}{s^{1/2}y(\alpha/\sqrt{s})}\mathrm{e}^{-\alpha^2/2}, \tag{4.7}$$

and for $l = 1, 2, \dots$,

$$E_l(\alpha) = \frac{1 - y(\alpha/\sqrt{s})}{s^{l+1/2}} \sum_{j=0}^{l-1} \frac{(l+j)c_{l-1,j}}{y^{l+2+j}(\alpha/\sqrt{s})}\mathrm{e}^{-\alpha^2/2}. \tag{4.8}$$

The c_{lj} are defined recursively by

$$c_{00} = 1; \quad c_{0j} = 0, \ j < 0 \text{ or } j > 0 \tag{4.9}$$

$$c_{l+1,j} = (l+1+j)c_{lj} - (l+j)c_{l,j-1}, \ j = 0, 1, \dots, l+1; \quad c_{l+1,j} = 0, \ j < 0 \text{ or } j > l+1, \tag{4.10}$$

where $l = 0, 1, \dots$. The remainder

$$R_k(\alpha) = \int_{-\infty}^{\alpha} \frac{d}{dx}\left[\Big(\frac{1}{x}\frac{d}{dx}\Big)^k \Big(\frac{1}{x}y'(x/\sqrt{s})\Big)\right] \mathrm{e}^{-x^2/2} dx \tag{4.11}$$

has the same sign as and smaller modulus than $E_{k+1}(\alpha)$.

Proof. We have by partial integration

$$\begin{aligned}\int_{-\infty}^{\alpha} \mathrm{e}^{-\frac{1}{2}x^2} y'(x/\sqrt{s})dx &= -\int_{-\infty}^{\alpha} \frac{1}{x} y'(x/\sqrt{s}) d(\mathrm{e}^{-\frac{1}{2}x^2}) \\ &= -\frac{1}{\alpha} y'(\alpha/\sqrt{s}) e^{-\frac{1}{2}\alpha^2} + \int_{-\infty}^{\alpha} \frac{d}{dx}\Big(\frac{1}{x}y'(x/\sqrt{s})\Big)\mathrm{e}^{-\frac{1}{2}x^2}dx, \end{aligned} \tag{4.12}$$

and repeating this yields

$$\begin{aligned}\int_{-\infty}^{\alpha} \mathrm{e}^{-\frac{1}{2}x^2} y'(x/\sqrt{s})dx = -\sum_{l=0}^{k} \left[\Big(\frac{1}{x}\frac{d}{dx}\Big)^l \Big(\frac{1}{x}y'(x/\sqrt{s})\Big)\right]_{x=\alpha} \mathrm{e}^{-\alpha^2/2} \\ + \int_{-\infty}^{\alpha} \frac{d}{dx}\left[\Big(\frac{1}{x}\frac{d}{dx}\Big)^k \Big(\frac{1}{x}y'(x/\sqrt{s})\Big)\right] \mathrm{e}^{-x^2/2} dx. \end{aligned} \tag{4.13}$$

The form (4.7) for $E_0(\alpha)$ follows from (4.5). Next, we have from (4.5) that

$$\frac{1}{x}\frac{d}{dx}\Big(\frac{1}{x}y'(x/\sqrt{s})\Big) = \frac{-1}{xs}\frac{y'(x/\sqrt{s})}{y^2(x/\sqrt{s})} = \frac{1}{s^{3/2}}\Big(\frac{1}{y^2(x/\sqrt{s})} - \frac{1}{y^3(x/\sqrt{s})}\Big) = -\frac{1-y(x/\sqrt{s})}{s^{3/2}}\frac{1}{y^3(x/\sqrt{s})}, \tag{4.14}$$

and this establishes

$$\Big(\frac{1}{x}\frac{d}{dx}\Big)^l \Big(\frac{1}{x}y'(x/\sqrt{s})\Big) = \frac{1}{s^{l+1/2}} \sum_{j=0}^{l} \frac{c_{lj}}{y^{l+1+j}(x/\sqrt{s})} \tag{4.15}$$

for $l = 1$ (we have $c_{10} = 1$, $c_{11} = -1$). Having established (4.15) for a certain $l = 1, 2, \dots$, we get

$$\Big(\frac{1}{x}\frac{d}{dx}\Big)^{l+1} \Big(\frac{1}{x}y'(x/\sqrt{s})\Big) = \frac{1}{s^{l+1/2}} \sum_{j=0}^{l} \frac{-(l+1+j)c_{lj}}{y^{l+2+j}(x/\sqrt{s})}\frac{1}{x\sqrt{s}} y'(x/\sqrt{s}). \tag{4.16}$$

Then again using (4.5) we obtain (4.15) for $l+1$ instead of l in which the required $c_{l+1,j}$ are expressed in terms of c_{lj} as in (4.10). From (4.16) and (4.5) there also follows the form (4.8) for $E_{l+1}(\alpha)$.

Next we do an administration of the signs of the c_{lj} and the left-hand sides of (4.15). We have, as can be seen by induction, that for $l = 0, 1, \ldots$

$$\operatorname{sgn}(c_{lj}) = (-1)^j, \quad j = 0, 1, \ldots, l. \tag{4.17}$$

Then from (4.15) and the fact that $y(x) < 0$ when $x < 0$, it is seen that for $l = 0, 1, \ldots$ and $x < 0$

$$\operatorname{sgn}\left[\left(\frac{1}{x}\frac{d}{dx}\right)^l \left(\frac{1}{x}y'(x/\sqrt{s})\right)\right] = (-1)^{l+1}. \tag{4.18}$$

By differentiating (4.15) and using that $y'(x) > 0$ it also follows that for $x < 0$

$$\operatorname{sgn}\left[\frac{d}{dx}\left[\left(\frac{1}{x}\frac{d}{dx}\right)^l \left(\frac{1}{x}y'(x/\sqrt{s})\right)\right]\right] = (-1)^{l+1}. \tag{4.19}$$

As a consequence we have $\operatorname{sgn}(R_k(\alpha)) = (-1)^{k+1}$. From this the statement about the modulus follows upon doing one more partial integration in (4.6). □

Proposition 17. *For $\delta < 0$ and $s \geqslant 1$,*

$$I_s(\delta) = -(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}\left[\frac{1}{\delta} - \frac{1}{\delta^3} - \frac{2}{\delta^4\sqrt{s}} + \left(3 - \frac{6}{s}\right)\frac{1}{\delta^5} + \varepsilon_I\right], \tag{4.20}$$

where $\varepsilon_I = O(\delta^{-6}s^{-1/2}) + O(\delta^{-7})$.

Proof. We apply Lemma 16 with $k = 3$ and $\alpha = \alpha(\delta)$ using, see (4.3), that $y(\alpha_s(\delta)/\sqrt{s}) = \delta/\sqrt{s}$. This gives

$$E_0(\alpha_s(\delta)) = -\frac{1}{\delta}(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}, \tag{4.21}$$

$$E_1(\alpha_s(\delta)) = \frac{1}{\delta^3}(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}, \tag{4.22}$$

$$E_2(\alpha_s(\delta)) = -\left(\frac{3}{\delta^5} - \frac{2}{\delta^4\sqrt{s}}\right)(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}, \tag{4.23}$$

$$E_3(\alpha_s(\delta)) = \left(\frac{15}{\delta^7} - \frac{20}{\delta^6\sqrt{s}} + \frac{6}{\delta^5 s}\right)(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}, \tag{4.24}$$

and for the sake of estimating the remainder $R_3(\alpha_s(\delta))$ we have

$$E_4(\alpha_s(\delta)) = -\Big(\frac{105}{\delta^9} - \frac{210}{\delta^8\sqrt{s}} + \frac{130}{\delta^7 s} - \frac{24}{\delta^6 s\sqrt{s}}\Big)(1-\delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}. \tag{4.25}$$

Now the result follows upon deleting the terms $-20/\delta^6\sqrt{s}$ and $15/\delta^7$ from $E_3(\alpha_s(\delta))$ and estimating $|R_3(\alpha_s(\delta))|$ by $|E_4(\alpha_s(\delta))|$, all this at the expense of an error which is $O(\delta^{-6}s^{-1/2}) + O(\delta^{-7})$. □

Using (3.21) and (4.21) we have

$$f_s(\delta) = -\delta\Big(1 - \frac{1}{\delta^2} - \frac{2}{\delta^3\sqrt{s}} + \Big(3 - \frac{6}{s}\Big)\frac{1}{\delta^4} + \delta\varepsilon_I\Big)^{-1}. \tag{4.26}$$

Now expanding $(1-x)^{-1} = 1 + x + x^2 + O(x^3)$ finally gives the result in Theorem 10.

4.2 Proof of Theorem 11

We suppress s in $f_s(\delta)$, $g_s(\delta)$ and $\alpha_s(\delta)$ until Subsection 4.9.

Lemma 18. *With the prime denoting differentiation with respect to δ, for $\delta < \sqrt{s}$,*

$$(-\tfrac{1}{2}\alpha^2(\delta))' = \frac{-\delta}{1-\delta/\sqrt{s}}, \tag{4.27}$$

$$f'(\delta) = -g(\delta)\Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big), \tag{4.28}$$

$$g'(\delta) = -g(\delta)\Big(\frac{\delta}{1-\delta/\sqrt{s}} + g(\delta)\Big), \tag{4.29}$$

$$f''(\delta) = g(\delta)\Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big)\Big(2g(\delta) + \frac{\delta}{1-\delta/\sqrt{s}}\Big) - g(\delta). \tag{4.30}$$

Proof. Straightforward verification from (3.21) and (3.22). □

We have $f(\delta) > 0 \geqslant -\delta$ when $\delta > 0$ and for $\delta < 0$ we have

$$f(\delta) > -\delta \quad \Leftrightarrow \quad -\frac{1}{\delta}\mathrm{e}^{-\alpha^2(\delta)/2}\Big(1 - \frac{\delta}{\sqrt{s}}\Big) > I(\delta), \tag{4.31}$$

where $I = I_s$ is the integral given in (4.4). From Theorem 10 it is seen that $f(\delta) > -\delta$ holds for large negative δ. We compute, using (4.27),

$$\Big(-\frac{1}{\delta}\mathrm{e}^{-\alpha^2(\delta)/2}\Big(1-\frac{\delta}{\sqrt{s}}\Big)\Big)' = \Big(1+\frac{1}{\delta^2}\Big)\mathrm{e}^{-\alpha^2(\delta)/2} > \mathrm{e}^{-\alpha^2(\delta)/2} = I'(\delta). \tag{4.32}$$

Therefore, the two inequalities in (4.31) hold for all $\delta < 0$.

We next show that $f'(\delta) > -1$. Using $g(\delta) = f(\delta)/(1-\delta/\sqrt{s})$ and (4.28), we have for $\delta < \sqrt{s}$

$$f'(\delta) > -1 \quad \Leftrightarrow \quad f(\delta)\Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big) < 1 - \frac{\delta}{\sqrt{s}}. \tag{4.33}$$

Using $f(\delta) = -\delta - \delta^{-1} - 2\delta^{-2}s^{-1/2} + O(\delta^{-3})$, see Theorem 10, we get

$$f(\delta)\Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big) = 1 - \frac{\delta}{\sqrt{s}} + \frac{1}{\delta\sqrt{s}} + O(\delta^{-2}) < 1 - \frac{\delta}{\sqrt{s}} \tag{4.34}$$

for large negative δ. Hence the two statements in (4.33) hold for large negative δ. The inequality in the second statement in (4.33) can be written as

$$\Big|f(\delta) + \frac{1}{2}\Big(\delta + \frac{1}{\sqrt{s}}\Big)\Big| < \Big(1 + \frac{1}{4}\Big(\delta - \frac{1}{\sqrt{s}}\Big)^2\Big)^{1/2}. \tag{4.35}$$

Now $f(\delta) > 0$ and $\frac{1}{2}(\delta + 1/\sqrt{s}) - (1 + \frac{1}{4}(\delta - 1/\sqrt{s})^2)^{1/2} < 0$ for $\delta < \sqrt{s}$, and so we have for $\delta < \sqrt{s}$ that

$$f'(\delta) > -1 \quad \Leftrightarrow \quad f(\delta) < \Big(1 + \frac{1}{4}\Big(\delta - \frac{1}{\sqrt{s}}\Big)^2\Big)^{1/2} - \frac{1}{2}\Big(\delta + \frac{1}{\sqrt{s}}\Big). \tag{4.36}$$

Using the fact that either statement in (4.36) holds for large negative δ, we shall show in Subsection 4.12 that second inequality in (4.36) holds for all $\delta < \sqrt{s}$.

We finally show that $f''(\delta) > 0$. It follows from (4.30), positivity of $g(\delta)$ when $\delta < \sqrt{s}$ and $g(\delta) = f(\delta)/(1-\delta/\sqrt{s})$ that for $\delta < \sqrt{s}$

$$f''(\delta) > 0 \quad \Leftrightarrow \quad \varphi(\delta) := \Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big)(2f(\delta) + \delta) > 1 - \frac{\delta}{\sqrt{s}}. \tag{4.37}$$

Using now the full strength of Theorem 10 (with $\varepsilon_f = O(\delta^{-4})$) we get

$$\begin{aligned}\varphi(\delta) &= \Big(-\frac{1}{\delta} + \frac{1}{\sqrt{s}} - \frac{2}{\delta^2\sqrt{s}} + \Big(2 - \frac{6}{s}\Big)\frac{1}{\delta^3} + O(\delta^{-4})\Big)\Big(-\delta - \frac{2}{\delta} - \frac{4}{\delta^2\sqrt{s}} + O(\delta^{-3})\Big) \\ &= 1 - \frac{\delta}{\sqrt{s}} + \frac{2}{s\delta^2} + O(\delta^{-3}) > 1 - \frac{\delta}{\sqrt{s}} \end{aligned} \tag{4.38}$$

for large negative δ. Hence, the two statements in (4.37) hold for large negative δ. The second inequality in (4.37) can be written as

$$\Big|f(\delta) + \frac{3}{4}\delta + \frac{1}{2\sqrt{s}}\Big| > \frac{1}{4}\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2}. \tag{4.39}$$

Using the fact that either statement in (4.37) holds for large negative δ, we shall show in Subsection 4.12 that, in fact, for $\delta < \sqrt{s}$,

$$f(\delta) > \frac{1}{4}\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2} - \frac{3}{4}\delta - \frac{1}{2\sqrt{s}}. \tag{4.40}$$

Hence, (4.39) holds for $\delta < \sqrt{s}$ and so $f''(\delta) > 0$ for $\delta < \sqrt{s}$.

4.3 Proof of Theorem 3

Assume that $\gamma \leqslant 0$. We have from Theorem 11 that

$$f(\gamma - a) > -(\gamma - a) \geqslant a \tag{4.41}$$

for any $a > 0$. Hence (3.8) does not have a solution.

Next assume that $\gamma \geqslant \sqrt{s}$. Recalling the definition $f(\delta) = 0$, $\delta \geqslant \sqrt{s}$, we have $f(\gamma - a) = 0$ at $a = 0$ while $\frac{d}{da}[f(\gamma - a)] < 1$ for all $a > 0$. Again it follows that (3.8) has no solution.

Finally, assume that $0 < \gamma < \sqrt{s}$. It follows from Theorem 10 that

$$f(\gamma - a) = -(\gamma - a) + O(a^{-1}) < a \tag{4.42}$$

for large positive a. Also, $f(\gamma - a) > 0$ at $a = 0$. Therefore, (3.8) has at least one solution. This solution is unique since $\frac{d}{da}f(\gamma - a) < 1$ by Theorem 11.

4.4 Proof of Theorem 4

Positivity is clear. We compute from (3.8) by implicit differentiation with respect to γ

$$a'(\gamma) = \frac{f'(\gamma - a(\gamma))}{1 + f'(\gamma - a(\gamma))} < 0, \qquad 0 < \gamma < \sqrt{s}, \tag{4.43}$$

since $f'(\delta) \in (-1, 0)$ for $\delta < \sqrt{s}$ by Theorem 11. Hence $a(\gamma)$ is strictly decreasing in $\gamma \in (0, \sqrt{s})$, and so $\gamma - a(\gamma)$ is strictly increasing in $\gamma \in (0, \sqrt{s})$. By convexity of f, see Theorem 11, we thus have that $f'(\gamma - a(\gamma))$ is strictly increasing in $\gamma \in (0, \sqrt{s})$. Since $f'(\gamma - a(\gamma)) \in (-1, 0)$ and $x/(1+x)$ is strictly increasing in $x \in (-1, 0)$ it then follows from (4.43) that $a'(\gamma)$ is strictly increasing in $\gamma \in (0, \sqrt{s})$. That is, $a(\gamma)$ is strictly convex.

4.5 Proof of Theorem 5

We first show that $b := \lim_{\gamma \downarrow 0} a(\gamma) = \infty$. Indeed, when $b < \infty$, we would have $b = \lim_{\gamma \downarrow 0} f(\gamma - a(\gamma)) = f(-b)$, contradicting Theorem 11. Hence, we can use Theorem 10 to see that

$$a = f(\gamma - a) = -(\gamma - a) - \frac{1}{\gamma - a} - \frac{2}{(\gamma - a)^2\sqrt{s}} + \left(2 - \frac{6}{s}\right)\frac{1}{(\gamma - a)^3} + \varepsilon_f(\gamma - a) \tag{4.44}$$

in which we have temporarily written $a = a(\gamma)$. Thus

$$\begin{aligned} \gamma &= \frac{1}{a - \gamma} - \frac{2}{(a - \gamma)^2\sqrt{s}} - \left(2 - \frac{6}{s}\right)\frac{1}{(a - \gamma)^3} + \varepsilon_f(\gamma - a) \\ &= \frac{1}{a - \gamma}\,(1 + o(1)), \qquad \gamma \downarrow 0. \end{aligned} \tag{4.45}$$

It follows that $a = \gamma^{-1}(1 + o(1))$, $\gamma \downarrow 0$. We now write the first line of (4.45) as

$$a - \gamma = \frac{1}{\gamma}\left(1 - \frac{2}{(a - \gamma)\sqrt{s}} - \left(2 - \frac{6}{s}\right)\frac{1}{(a - \gamma)^2} + (a - \gamma)\,\varepsilon_f(\gamma - a)\right) \tag{4.46}$$

noting that $(a - \gamma)^{-1} = O(\gamma)$, $(a - \gamma)\,\varepsilon_f(\gamma - a) = O(\gamma^3/\sqrt{s}) + O(\gamma^4)$. The form (4.46) is appropriate for getting ever more precise asymptotic information on $a - \gamma$ by iteration. Thus it is first found that

$$a - \gamma = \frac{1}{\gamma}\,(1 + O(\gamma/\sqrt{s}) + O(\gamma^2)), \tag{4.47}$$

and next it is found from (4.46) that

$$a - \gamma = \frac{1}{\gamma}\Big(1 - \frac{2\gamma}{\sqrt{s}} + O(\gamma^2)\Big). \tag{4.48}$$

One more iteration yields

$$a - \gamma = \frac{1}{\gamma}\Big(1 - \frac{2\gamma}{\sqrt{s}} - \Big(2 - \frac{2}{s}\Big)\gamma^2 + O(\gamma^3/\sqrt{s}) + O(\gamma^4)\Big), \tag{4.49}$$

and this is already (3.10).

4.6 Proof of Proposition 6

We have

$$a'(\gamma) = \frac{-a(\gamma)(\gamma + 1/\sqrt{s})}{1 - \gamma(a(\gamma) + 1/\sqrt{s})}, \quad 0 < \gamma < \sqrt{s}. \tag{4.50}$$

From Theorem 4 it follows that $a(\gamma) > 0$, $a'(\gamma) < 0$, so that by (4.50), $1 - \gamma(a(\gamma) + 1/\sqrt{s}) > 0$.

Let L as in (3.28). From $f(\delta) > L(\delta)$ we get that $a(\gamma) > a_L(\gamma)$ with $a_L(\gamma)$ the solution $a > 0$ of $a = L(\gamma - a)$ for $\gamma \in (0, \sqrt{s})$. Solving this equation yields $\gamma a_L(\gamma) = 1 - 2\gamma/\sqrt{s} - \gamma^2$ and this gives the lower bound in (3.12)

Finally, note that

$$(\gamma a(\gamma))' = a(\gamma) + \gamma a'(\gamma) = a(\gamma)\Big(1 - \frac{\gamma(\gamma + 1/\sqrt{s})}{1 - \gamma(a(\gamma) + 1/\sqrt{s})}\Big), \tag{4.51}$$

and this is negative if and only if $a(\gamma) > \frac{1}{\gamma} - \frac{2}{\sqrt{s}} - \gamma$.

4.7 Proof of Proposition 7

Since $a(\gamma)$ decreases in $\gamma > 0$, it is sufficient to show that $a(f(0)) = f(0)$. This follows from

$$f(0) = f(f(0) - f(0)), \tag{4.52}$$

see equation (3.8).

4.8 Proof of Theorem 8

Let $c := \lim_{\gamma\uparrow\sqrt{s}} a(\gamma)$. We shall show that $c = 0$. Indeed, from $c = f(\sqrt{s} - c)$ we get for some d, $\sqrt{s} - c \leqslant d \leqslant \sqrt{c}$, that

$$c = f(\sqrt{s} - c) = f(\sqrt{s}) - cf'(d) = 0 - cf'(d), \tag{4.53}$$

and $f'(d) > -1$ by Theorem 11.

Next, we have from (2.2) and (3.9) that

$$f(\delta) = \frac{s^{s+1/2}(1 - \delta/\sqrt{s})^{s+1}/s!}{\sum_{k=0}^{s} \frac{s^k}{k!}(1 - \delta/\sqrt{s})^k}, \qquad \delta \leqslant \sqrt{s}, \tag{4.54}$$

so that, in particular, $f'(\delta) \to 0$ as $\delta \uparrow \sqrt{s}$. Writing temporarily $a = a(\gamma)$, we have from $a = f(\gamma - a)$ that there is a $\xi \in [\gamma - a, \gamma]$ such that

$$a = f(\gamma) - af'(\xi). \tag{4.55}$$

Since $f'(\xi) \to 0$ as $\gamma \uparrow \sqrt{s}$ (which follows from $\xi \geqslant \gamma - a$ and $a \to 0$ as $\gamma \uparrow \sqrt{s}$), we thus see that

$$a = O(f(\gamma)) = O\Big(\frac{s^{s+1/2}}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s+1}\Big), \qquad \gamma \uparrow \sqrt{s}. \tag{4.56}$$

It then follows from $a = f(\gamma - a)$ and (4.54) that

$$\begin{aligned} a &= \frac{\frac{s^{s+1/2}}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}} + O\Big(\frac{s^s}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s+1}\Big)\Big)^{s+1}}{1 + O\Big(1 - \frac{\gamma}{\sqrt{s}}\Big) + O\Big(\frac{s^s}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s}\Big)} \\ &= \frac{s^{s+1/2}}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s+1} \frac{\Big(1 + O\Big(\frac{e^s}{\sqrt{s}}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s}\Big)\Big)^{s+1}}{1 + O\Big(1 - \frac{\gamma}{\sqrt{s}}\Big) + O\Big(\frac{e^s}{s!}\Big(1 - \frac{\gamma}{\sqrt{s}}\Big)^{s}\Big)}, \quad \gamma \uparrow \sqrt{s}. \end{aligned} \tag{4.57}$$

From this the result follows.

4.9 Proofs of Theorems 9 and 13

We start by showing monotonicity of $g_s(\delta)$ in $s \geqslant 1$. With the notation of (3.9), (3.20) and (4.4) we have for a fixed δ with $\delta < \sqrt{s}$ that

$$\frac{\partial}{\partial s}\,(g_s(\delta)) = \frac{\partial}{\partial s}\,[e^{-\frac{1}{2}\alpha_s^2(\delta)}/I_s(\delta)]$$
$$= \frac{e^{-\frac{1}{2}\alpha_s^2(\delta)}}{I_s^2(\delta)} \int_{-\infty}^{\delta} e^{-\frac{1}{2}\alpha_s^2(\delta')} \Big[\frac{\partial}{\partial s}\,(-\tfrac{1}{2}\alpha_s^2(\delta)) - \frac{\partial}{\partial s}\,(-\tfrac{1}{2}\alpha_s^2(\delta'))\Big]\, d\delta'. \tag{4.58}$$

Now

$$\frac{\partial}{\partial s}\,(-\tfrac{1}{2}\alpha_s^2(\delta)) = \frac{\partial}{\partial s}\,(\delta\sqrt{s} + s\ln(1-\delta/\sqrt{s}))$$
$$= \frac{\delta}{2\sqrt{s}}\Big(1 + \frac{1}{1-\delta/\sqrt{s}}\Big) + \ln(1-\delta/\sqrt{s}) = \tfrac{1}{2}\int_0^{\delta/\sqrt{s}} \Big(\frac{x}{1-x}\Big)^2 dx. \tag{4.59}$$

Hence, when $-\infty < \delta' < \delta < \sqrt{s}$, we have

$$\frac{\partial}{\partial s}\,(-\tfrac{1}{2}\,\alpha_s^2(\delta)) - \frac{\partial}{\partial s}\,(-\tfrac{1}{2}\,\alpha_s^2(\delta')) = \tfrac{1}{2}\int_{\delta'/\sqrt{s}}^{\delta/\sqrt{s}} \Big(\frac{x}{1-x}\Big)^2 dx > 0. \tag{4.60}$$

Using this in (4.58) yields $\frac{\partial}{\partial s}\,(g_s(\delta)) > 0$.

Next, we show monotonicity of $f_s(\delta)$ in $s \geqslant 1$. From $f_s(\delta) = (1-\delta/\sqrt{s})\,g_s(\delta)$, we have

$$\frac{\partial}{\partial s}\,(f_s(\delta)) = \frac{\delta}{2s\sqrt{s}}\,g_s(\delta) + (1-\delta/\sqrt{s})\,\frac{\partial}{\partial s}\,(g_s(\delta)), \tag{4.61}$$

and so $\frac{\partial}{\partial s}\,(f_s(\delta)) > 0$ when $0 \leqslant \delta < \sqrt{s}$. It remains to show that

$$\frac{\partial}{\partial s}\,(g_s(\delta))/g_s(\delta) > \frac{-1}{2s}\,\frac{\delta/\sqrt{s}}{1-\delta/\sqrt{s}}, \qquad \delta < 0. \tag{4.62}$$

We compute, using (4.58) and (4.60),

$$\frac{\partial}{\partial s}\,(g_s(\delta))/g_s(\delta) = \frac{1}{2I_s(\delta)}\int_{-\infty}^{\delta} e^{-\frac{1}{2}\alpha_s^2(\delta')} \left(\int_{\delta'/\sqrt{s}}^{\delta/\sqrt{s}} \Big(\frac{x}{1-x}\Big)^2 dx\right) d\delta'. \tag{4.63}$$

Using (4.27), we get by partial integration (noting that $\delta' < \delta < 0$)

$$\int_{-\infty}^{\delta} e^{-\frac{1}{2}\alpha_s^2(\delta')} \left(\int_{\delta'/\sqrt{s}}^{\delta/\sqrt{s}} \left(\frac{x}{1-x}\right)^2 dx \right) d\delta'$$

$$= \frac{-1}{2s} \int_{-\infty}^{\delta} e^{-\frac{1}{2}\alpha_s^2(\delta')} \frac{d}{dz} \left[\frac{1-z}{-z} \int_{z}^{\delta/\sqrt{s}} \left(\frac{x}{1-x}\right)^2 dx \right] (z = \delta'/\sqrt{s})\, d\delta'. \tag{4.64}$$

We shall show that for $z < w < 0$

$$-\frac{d}{dz} \left[\frac{1-z}{-z} \int_{z}^{w} \left(\frac{x}{1-x}\right)^2 dx \right] > \frac{-w}{1-w}. \tag{4.65}$$

Using this with $w = \delta/\sqrt{s}$ in (4.63) and using the definition (4.4) of $I_s(\delta)$ then yields (4.62). The inequality in (4.65) with $z < w < 0$ is equivalent with

$$\frac{d}{db} \left[\frac{1+b}{b} \int_{a}^{b} \left(\frac{x}{1+x}\right)^2 dx \right] > \frac{a}{1+a}, \qquad 0 < a < b. \tag{4.66}$$

Now (4.66) holds if and only if

$$\frac{1}{b^2} \int_{a}^{b} \left(\frac{x}{1+x}\right)^2 dx < \frac{b}{1+b} - \frac{a}{1+a}, \qquad 0 < a < b, \tag{4.67}$$

and since

$$\frac{b}{1+b} - \frac{a}{1+a} = \int_{a}^{b} \left(\frac{1}{1+x}\right)^2 dx, \qquad 0 < a < b, \tag{4.68}$$

we obviously have (4.67). Hence $\frac{\partial}{\partial s}\left(f_s(\delta)\right) > 0$.

We finally show monotonicity of $a_s(\gamma)$ in $s \geqslant 1$, where now γ, $0 < \gamma < \sqrt{s}$, is fixed. By implicity differentiation of the equation $a_s(\gamma) = f_s(\gamma - a_s(\gamma))$ with respect to s, we get

$$\frac{\partial}{\partial s}\left(a_s(\gamma)\right) = \frac{\dfrac{\partial f_s}{\partial s}\left(\gamma - a_s(\gamma)\right)}{1 + \dfrac{\partial f_s}{\partial \gamma}\left(\gamma - a_s(\gamma)\right)} > 0, \qquad 0 < \gamma < \sqrt{s}, \tag{4.69}$$

by Theorem 11 and $\frac{\partial f_s}{\partial s}(\delta) > 0$.

4.10 Proof of Theorem 14

The matter of existence and uniqueness of the solution of equation (3.32) is settled in a similar way as this was done for equation (3.8) in the proof of Theorem 3 (see Subsection 4.3).

By [1], 7.2.14 on p. 300, case $n = 0$, we have

$$\begin{aligned}\int_{-\infty}^{\delta} \exp(-\tfrac{1}{2}\,(\delta')^2)\,d\delta' &= \sqrt{\frac{\pi}{2}}\,\mathrm{erfc}(-\delta/\sqrt{2}) \sim -e^{-\frac{1}{2}\delta^2} \sum_{m=0}^{\infty} \frac{(-1)^m (2m)!}{m!\,2^m\,\delta^{2m+1}} \\ &= -e^{-\frac{1}{2}\delta^2}\Big(\frac{1}{\delta} - \frac{1}{\delta^3} + \frac{3}{\delta^5} - \frac{15}{\delta^7} + \frac{105}{\delta^9} + O\Big(\frac{1}{\delta^{11}}\Big)\Big), \quad \delta < 0. \end{aligned} \tag{4.70}$$

Then (3.33) follows from (4.70) and some administration. Next, (3.33) is used in a similar way as (3.25) was used to prove (3.10) (this requires including an additional term $-945/\delta^{11}$ in the expansion in (4.70)).

4.11 Proof of Theorem 15

We start by analyzing the function

$$J_s(\delta) := \int_{-\infty}^{\delta} e^{-\frac{1}{2}(\delta')^2}\,d\delta' - \int_{-\infty}^{\delta} e^{-\frac{1}{2}\alpha_s^2(\delta')}\,d\delta', \qquad \delta \leqslant \sqrt{s}. \tag{4.71}$$

We have $\alpha_s(0) = 0$ and $\alpha_s(\delta) > \delta$, $0 \neq \delta < \sqrt{s}$ and so

$$e^{-\frac{1}{2}\alpha_s^2(\delta)} > e^{-\frac{1}{2}\delta^2}\,, \quad \delta < 0\,; \qquad e^{-\frac{1}{2}\alpha_s^2(\delta)} < e^{-\frac{1}{2}\delta^2}\,, \quad 0 < \delta \leqslant \sqrt{s}. \tag{4.72}$$

It follows that $J_s(\delta)$ decreases from 0 to the value

$$\tfrac{1}{2}\sqrt{2\pi} - (f_s(0))^{-1} = \frac{-2}{3\sqrt{s}} + O\Big(\frac{1}{s}\Big) \tag{4.73}$$

at $\delta = 0$, and increases from this value (4.73) at $\delta = 0$ to

$$\begin{aligned} J_s(\sqrt{s}) &= \int_{-\infty}^{\sqrt{s}} e^{-\frac{1}{2}(\delta')^2}\, d\delta' - \int_{-\infty}^{\sqrt{s}} e^{-\frac{1}{2}\alpha_s^2(\delta')}\, d\delta' \\ &= \sqrt{2\pi} + O(e^{-\frac{1}{2}s}) - \frac{e^s\, s!}{s^{s+1/2}} \\ &= -\frac{\sqrt{2\pi}}{12s} + O\Big(\frac{1}{s}\Big) \end{aligned} \tag{4.74}$$

by Stirling's formula at $\delta = \sqrt{s}$.

Now let $\delta_0 < \sqrt{s}$. We have that both $\int_{-\infty}^{\delta} e^{-\frac{1}{2}(\delta')^2}\, d\delta'$ and $\int_{-\infty}^{\delta} e^{-\frac{1}{2}\alpha_s^2(\delta')}\, d\delta'$ are bounded away from 0 when $\delta_0 \leqslant \delta \leqslant \sqrt{s}$ while their difference $J_s(\delta) = O(1/\sqrt{s})$ uniformly on $[\delta_0, \sqrt{s}]$. From (3.21) and (3.22), considered in the uniformly bounded interval $[\delta_0, 0]$, we see that $f_\infty(\delta) - f_s(\delta) = O(1/\sqrt{s})$ uniformly in $\delta \in [\delta_0, 0]$. The interval $[0, \sqrt{s}]$ presents, however, a complication since $\sqrt{s} \to \infty$ as $s \to \infty$.

To deal with the interval $[0, \sqrt{s}]$, we let

$$G_s(\delta) := e^{-\frac{1}{2}\delta^2} - e^{-\frac{1}{2}\alpha_s^2(\delta)}, \qquad 0 \leqslant \delta \leqslant \sqrt{s}. \tag{4.75}$$

Now $G_s(\delta) \geqslant 0$ with equality only when $\delta = 0$, and

$$G_s'(\delta) = \delta\Big(1 - \frac{\delta}{\sqrt{s}}\Big)^{s-1} e^{\delta\sqrt{s}} - \delta\, e^{-\frac{1}{2}\delta^2} \tag{4.76}$$

vanishes when $\delta = 0$ and when

$$\varphi_s(\delta) := \tfrac{1}{2}\,\delta^2 + \delta\,\sqrt{s} + (s-1)\,\ln(1 - \delta/\sqrt{s}) = 0. \tag{4.77}$$

There holds

$$\varphi_s(\delta) = \sum_{k=1}^{\infty} \frac{\delta^k}{k\, s^{k/2}} \Big(1 - \frac{k\,\delta^2}{k+2}\Big) \tag{4.78}$$

which shows that $\varphi_s(\delta) > 0$ when $0 < \delta \leqslant 1$ and that $\varphi_s(\delta) < 0$ when $\delta \geqslant \sqrt{3}$. Accordingly, $G_s'(\delta) > 0$ for $0 < \delta \leqslant 1$ and $G_s'(\delta) < 0$ for $\delta \geqslant \sqrt{3}$.

In view of (3.21), we should, however, consider

$$H_s(\delta) := e^{-\frac{1}{2}\delta^2} - \Big(1 - \frac{\delta}{\sqrt{s}}\Big)\, e^{-\frac{1}{2}\alpha_s^2(\delta)}, \qquad 0 \leqslant \delta \leqslant \sqrt{s}. \tag{4.79}$$

We have $H(\delta) \geqslant G(\delta) > 0$, and we compute

$$H_s'(\delta) = \Big(\delta + \frac{1}{\sqrt{s}}\Big)\Big(1 - \frac{\delta}{\sqrt{s}}\Big)\Big(1 - \frac{\delta}{\sqrt{s}}\Big)^{s-1} e^{\delta\sqrt{s}} - \delta\, e^{-\frac{1}{2}\delta^2}, \qquad 0 \leqslant \delta \leqslant \sqrt{s}. \tag{4.80}$$

From

$$\Big(\delta + \frac{1}{\sqrt{s}}\Big)\Big(1 - \frac{\delta}{\sqrt{s}}\Big) \leqslant \delta, \qquad \delta \geqslant 1, \tag{4.81}$$

we see that $H_s'(\delta) \leqslant G_s'(\delta)$ when $\delta \geqslant 1$, and so $H_s(\delta)$ decreases when $\delta \geqslant \sqrt{3}$. It follows that $H_s(\delta)$ takes its maximum in the interval $[0, \sqrt{3}]$, and so this maximum is $O(1/\sqrt{s})$. Then, as in the case of the interval $[\delta_0, 0]$, it is concluded that $f_\infty(\delta) - f_s(\delta) = O(1/\sqrt{s})$ uniformly in $\delta \in [0, \sqrt{s}]$.

Now we consider $a_\infty - a_s$. Let $\gamma_0 > 0$ and consider $s \geqslant \gamma_0^2$. We have by Theorem 13 that for all $\gamma \in [\gamma_0, \sqrt{s}]$

$$\gamma - a_s(\gamma) \geqslant \gamma_0 - a_\infty(\gamma_0) =: \delta_0 > -\infty. \tag{4.82}$$

Furthermore, by what we have shown already, there is a $K > 0$ such that

$$0 < f_\infty(\delta) - f_s(\delta) < \frac{K}{\sqrt{s}}, \qquad s \geqslant 1\,, \quad \delta \in [\delta_0, \sqrt{s}]. \tag{4.83}$$

Also, for any $\delta \in [\delta_0, \sqrt{s}]$, by Theorem 11 and (3.30)

$$f_s'(\delta) \geqslant f_s'(\delta_0) > -1\,, \quad f_s'(\delta_0) \to f_\infty'(\delta_0) > -1, \qquad s \to \infty. \tag{4.84}$$

It follows that there is an $\varepsilon > 0$ such that

$$f_s'(\delta) \geqslant -(1 - \varepsilon), \qquad s \geqslant 1\,, \quad \delta \in [\delta_0, \sqrt{s}] \tag{4.85}$$

Now take any $\gamma \in [\gamma_0, \sqrt{s}]$, and let

$$h_s(a) := a - f_s(\gamma - a)\,, \quad h_\infty(a) := a - f_\infty(\gamma - a), \qquad a \leqslant \delta_0 - \gamma. \tag{4.86}$$

Then

$$h_s(a_s(\gamma)) = 0 = h_\infty(a_\infty(\gamma)), \tag{4.87}$$

and

$$0 < h_s(a) - h_\infty(a) < \frac{K}{\sqrt{s}}, \qquad a \leqslant \delta_0 - \gamma\,, \quad s \geqslant 1, \tag{4.88}$$

by Theorem 13 and (4.83), while

$$h_s'(a) = 1 + f_s'(\gamma - a) \geqslant \varepsilon\,, \quad h_\infty'(a) = 1 + f_\infty'(\gamma - a) \geqslant \varepsilon,$$

$$a \leqslant \delta_0 - \gamma\,, \quad s \geqslant 1. \tag{4.89}$$

By the mean-value theorem there is a b, $a_s(\gamma) \leqslant b \leqslant a_\infty(\gamma) \leqslant \delta_0 - \gamma$, such that

$$\begin{aligned}(a_\infty(\gamma) - a_s(\gamma))\, h_s'(b) &= h_s(a_\infty(\gamma)) - h_s(a_s(\gamma)) \\ &= h_s(a_\infty(\gamma)) = h_s(a_\infty(\gamma)) - h_\infty(a_\infty(\gamma)).\end{aligned} \tag{4.90}$$

Then from Theorem 13 and (4.88), (4.89)

$$0 < a_\infty(\gamma) - a_s(\gamma) < \frac{K}{\varepsilon\,\sqrt{s}}, \tag{4.91}$$

and the result follows.

4.12 Proof of Corollary 12

The inequalities in Corollary 12 were needed in the proofs of convexity of f (lower bound) and of $f'(\delta) > -1$ (upper bound). In the proof of Theorem 11, these inequalities were shown to be equivalent with

$$\Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big)\,(2f(\delta) + \delta) > 1 - \delta/\sqrt{s}, \qquad \delta < \sqrt{s}, \tag{4.92}$$

and

$$f(\delta)\Big(\delta + \frac{1}{\sqrt{s}} + f(\delta)\Big) < 1 - \delta/\sqrt{s}, \qquad \delta < \sqrt{s}, \tag{4.93}$$

respectively. It was already shown in the proof of Theorem 11 that (4.92) and (4.93) hold for large negative δ.

We first show that

$$f(\delta) < -\frac{1}{2}\Big(\delta + \frac{1}{\sqrt{s}}\Big) + \frac{1}{2}\Big(\Big(\delta - \frac{1}{\sqrt{s}}\Big)^2 + 4\Big)^{1/2} =: F(\delta). \tag{4.94}$$

From (3.21) we have

$$f(\delta) < F(\delta) \Leftrightarrow I(\delta) > (1 - \delta/\sqrt{s})\, e^{-\frac{1}{2}\alpha^2(\delta)}/F(\delta) =: S(\delta), \tag{4.95}$$

where $I = I_s$ is the integral in (4.4). From (4.27) we compute

$$S'(\delta) = \frac{-\delta}{F(\delta)}\, e^{-\frac{1}{2}\alpha^2(\delta)} - \frac{\dfrac{1}{\sqrt{s}}\, F(\delta) + (1 - \delta/\sqrt{s})\, F'(\delta)}{F^2(\delta)}\, e^{-\frac{1}{2}\alpha^2(\delta)}. \tag{4.96}$$

We shall show that $I'(\delta) = \exp(-\frac{1}{2}\,\alpha^2(\delta)) > S'(\delta)$, and this is equivalent with

$$F^2(\delta) > -\Big(\delta + \frac{1}{\sqrt{s}}\Big)\, F(\delta) - (1 - \delta/\sqrt{s})\, F'(\delta) \tag{4.97}$$

by (4.96). Now from the definition (4.94) of $F(\delta)$, we have

$$\begin{aligned} F^2(\delta) + \Big(\delta + \frac{1}{\sqrt{s}}\Big)\, F(\delta) &= F(\delta)\Big(F(\delta) + \delta + \frac{1}{\sqrt{s}}\Big) \\ &= 1 - \delta/\sqrt{s}, \end{aligned} \tag{4.98}$$

and so (4.97) is equivalent with $F'(\delta) > -1$. We compute

$$F'(\delta) = -\frac{1}{2} + \frac{1}{2}\,\frac{\delta - 1/\sqrt{s}}{((\delta - 1/\sqrt{s})^2 + 4)^{1/2}} > -1, \qquad \delta \in \mathbb{R}, \tag{4.99}$$

and so $I'(\delta) > S'(\delta)$ for all $\delta < \sqrt{s}$. Since $I(\delta) > S(\delta)$ holds for large negative δ, it follows that $I(\delta) > S(\delta)$ for all $\delta < \sqrt{s}$. Hence (4.94) and $f'(\delta) > -1$ hold for all $\delta < \sqrt{s}$.

We next show that

$$f(\delta) > -\Big(\frac{3}{4}\,\delta + \frac{1}{2\sqrt{s}}\Big) + \frac{1}{4}\,\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2} =: E(\delta). \tag{4.100}$$

From (3.21) we have

$$f(\delta) > E(\delta) \Leftrightarrow I(\delta) < (1 - \delta/\sqrt{s})\, e^{-\frac{1}{2}\alpha^2(\delta)}/E(\delta) =: R(\delta). \tag{4.101}$$

We shall show that $I'(\delta) < R'(\delta)$. As above, we have

$$I'(\delta) < R'(\delta) \Leftrightarrow E^2(\delta) < -\Big(\delta + \frac{1}{\sqrt{s}}\Big)\, E(\delta) - (1 - \delta/\sqrt{s})\, E'(\delta). \tag{4.102}$$

We now compute

$$E^2(\delta) + \Big(\delta + \frac{1}{\sqrt{s}}\Big)\, E(\delta) = -\frac{1}{8}\,\delta^2 - \frac{3\delta}{4\sqrt{s}} + \frac{1}{2} - \frac{1}{8}\,\delta\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2}. \tag{4.103}$$

Then using that

$$E'(\delta) = -\frac{3}{4} + \frac{1}{4}\,\frac{\delta - 2/\sqrt{s}}{((\delta - 2/\sqrt{s})^2 + 8)^{1/2}}, \tag{4.104}$$

it is found that

$$\begin{aligned} I'(\delta) < R'(\delta) \quad \Leftrightarrow \quad & -\frac{1}{8}\,\delta^2 - \frac{3\delta}{4\sqrt{s}} + \frac{1}{2} - \frac{1}{8}\,\delta\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2} \\ & < \; -(1 - \delta/\sqrt{s})\Big(-\frac{3}{4} + \frac{1}{4}\,\frac{\delta - 2/\sqrt{s}}{((\delta - 2/\sqrt{s})^2 + 8)^{1/2}}\Big). \end{aligned} \tag{4.105}$$

With some algebra, this works out to

$$I'(\delta) < R'(\delta) \Leftrightarrow (\delta^2 + 2)\Big(\Big(\delta - \frac{2}{\sqrt{s}}\Big)^2 + 8\Big)^{1/2} > -\delta^3 + \frac{2}{\sqrt{s}}\,\delta^2 - 6\delta - \frac{4}{\sqrt{s}}. \tag{4.106}$$

Setting $x = -\delta$ and taking squares, the inequality in the second proposition of (4.106) is implied by

$$(x^2 + 2)^2\Big(\Big(x + \frac{2}{\sqrt{s}}\Big)^2 + 8\Big) > \Big(x^3 + \frac{2}{\sqrt{s}}\,x^2 + 6x - \frac{4}{\sqrt{s}}\Big)^2. \tag{4.107}$$

Working this out and simplifying finally leads to the condition $(x + \sqrt{s})^2 > 0$ which obviously holds when $-x = \delta < \sqrt{s}$. Hence $I'(\delta) < R'(\delta)$ holds for all $\delta < \sqrt{s}$ while $I(\delta) < R(\delta)$ holds for all large negative δ. We conclude that $I(\delta) < R(\delta)$ holds for all $\delta < \sqrt{s}$, and so (4.100) and $f''(\delta) > 0$ hold for all $\delta < \sqrt{s}$.

Acknowledgments. FA would like to thank CNRS and EURANDOM for providing wonderful working conditions during a CNRS detachment at EURANDOM in the year 2010. JvL is supported by an ERC Starting Grant.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions (9th printing)*. Dover, 1970.

[2] J.R. Artalejo and A. Gómez-Corral. *Retrial Queueing Systems*. Springer-Verlag, 2008.

[3] A. Baricz. Mills' ratio: Monotonicity patterns and functional inequalities. *Journal of Mathematical Analysis and Applications*, 340:1362–1370, 2008.

[4] S. Borst, A. Mandelbaum, and M. Reiman. Dimensioning large call centers. *Oper. Res.*, 52:17–34, 2004.

[5] J.W. Cohen. Basic problems of telephone traffic theory and the influence of repeated calls. *Philips Telecommunication Review*, 18:49–100, 1957.

[6] G.I. Falin and J.G.C. Templeton. *Retrial Queues*. Chapman & Hall, London, 1997.

[7] S. Halfin and W. Whitt. Heavy-traffic limits for queues with many exponential servers. *Oper. Res.*, 29:567–588, 1981.

[8] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, and B. Zwart. Gaussian expansions and bounds for the Poisson distribution applied to the Erlang B formula. *Adv. in Appl. Probab.*, 40, 2008.

[9] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, and B. Zwart. Refining square root safety staffing by expanding Erlang C. *To appear in Operations Research*, 2011.

[10] A. Mandelbaum and S. Zeltyn. Service engineering in action: The Palm/Erlang-A queue, with applications to call centers. In *Advances in Services Innovations*, pages 17–48. Springer-Verlag, 2007.

[11] M.R. Sampford. Some inequalities on Mill's ratio and related functions. *Annals of Mathematical Statistics*, 24:130–132, 1953.

[12] S. Zeltyn and A. Mandelbaum. Call centers with impatient customers: Many-server asymptotics of the M/M/n + G queue. *Queueing Syst. Theory Appl.*, 51:361–402, 2005.

[13] B. Zhang, J.S.H. van Leeuwaarden, and B. Zwart. Refined square-root staffing for call centers with impatient customers. *To appear in Operations Research*, 2010.