Adv. Appl. Prob. **45,** 274–294 (2013)
Printed in Northern Ireland

LOSS SYSTEMS WITH SLOW RETRIALS IN THE HALFIN–WHITT REGIME

F. AVRAM,* *Université de Pau et des Pays de l'Adour*
A. J. E. M. JANSSEN,** *Eindhoven University of Technology and EURANDOM*
J. S. H. VAN LEEUWAARDEN,*** *Eindhoven University of Technology*

Abstract

The Halfin–Whitt regime, or the quality-and-efficiency-driven (QED) regime, for multiserver systems refers to a situation with many servers, a critical load, and yet favorable system performance. We apply this regime to the classical multiserver loss system with slow retrials. We derive nondegenerate 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 persist for systems with retrials, although in situations when the load becomes *extremely* critical the retrials cause deteriorated performance. 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 prelimit behavior and exploiting the connection between Cohen's equation and Mills' ratio for the Gaussian and Poisson distributions.

Keywords: Retrial system; loss system; Erlang B model; Cohen's equation; Mills' ratio; Halfin–Whitt regime, QED regime

2010 Mathematics Subject Classification: Primary 60K25; 68M10; 41A60

1. Introduction

Customers of call centers that obtain a busy signal usually repeat calls until the required connection is made. A call center therefore receives two types of incoming calls: *primary calls* received from customers calling 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 successful, is one of the crucial factors for call center performance; see, for instance, [2]. In this paper we investigate the basic multiserver 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] and [6], which is why one often resorts to computational approaches [3]. 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 long time (compared to the time

Received 14 June 2011; revision received 30 May 2012.
* Postal address: Département de Mathématiques, Université de Pau et des Pays de l'Adour, Avenue de l'Université - BP 1155, 64013 Pau Cedex, France. Email address: florin.avram@univ-pau.fr
** Postal address: Department of Mathematics and Computer Science and Department of Electrical Engineering, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands.
Email address: a.j.e.m.janssen@tue.nl
*** Postal address: Department of Mathematics and Computer Science, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands. Email address: j.s.h.v.leeuwaarden@tue.nl

scale of the system). The combination of these two asymptotic regimes leads to a tractable model.

There is by now a vast literature on the asymptotic analysis of multiserver 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]. A quite effective approximation 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 to setting $s = \lambda + \beta\sqrt{\lambda}$ (square-root staffing) for some fixed constant β. Square-root staffing and the QED limiting regime for multiserver systems (without retrials) were brought to the center of attention by the work of Halfin and Whitt [7], and, therefore, the QED regime also goes by the name Halfin–Whitt regime.

In this paper we consider the multiserver loss system (M/M/s/s queue) with retrials. We analyze this system in the Halfin–Whitt regime, in a similar spirit as was done for the M/M/s queue [7], [10], the M/M/s/s queue [9], and the Erlang A model (M/M/s queue with abandonments) [12]. 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. (1) below. In short, we make the following contributions.

(i) Within the realm of the Halfin–Whitt regime, the retrial phenomenon was relatively unexplored. In this paper we present 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 lead to 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 prelimit or true retrial rate. It is shown that the difference between the true retrial rate and its QED approximation tends to 0 as the system size tends to ∞, which provides evidence for the appropriateness of square-root staffing for call centers, even in the case of retrials.

(ii) A major effort is devoted to 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 $s \to \infty$ 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 in which 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 1 by making s large, but also by making γ small. We present several results that help in understanding the effects of both scalings (see Theorem 3 and Proposition 1).

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

In Section 2 we introduce the multiserver system with slow retrials. In Section 3 we present the main results. In Section 4 we study Cohen's equation and give the work plan for the proofs of the main results. Finally, Section 5 contains all the proofs.

2. Description of the retrial system

We now describe the classical multiserver loss system with retrials (see, e.g. [3] and [6, Chapter 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 1, and interarrival times, service times, and retrial times are mutually independent. The system state can then be described by means of a bivariate process $\{(C(t), N(t)); t \geq 0\}$ with $C(t)$ the number of busy 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 $\{0, 1, \ldots, s\} \times \{0, 1, \ldots\}$. We assume that $\lambda < s$, which is a necessary and sufficient condition for ergodicity (see [6, Chapter 2]). Denote by $(C(\infty), N(\infty))$ the random variables having the joint stationary distribution of the process $\{(C(t), N(t)); t \geq 0\}$.

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 formulae 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 equation

$$\Omega = (\lambda + \Omega) B(s, \lambda + \Omega). \tag{1}$$

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

$$B(s, \lambda) = \frac{\lambda^s / s!}{\sum_{k=0}^{s} \lambda^k / k!} = \frac{\mathrm{e}^{-\lambda} (\lambda/s)^s}{\int_{\lambda}^{\infty} \mathrm{e}^{-\lambda'} (\lambda'/s)^s \, \mathrm{d}\lambda'} \tag{2}$$

with $\lambda > 0$ and $s = 1, 2, \ldots$. The form in (2) comprising the integral allows us to consider $B(s, \lambda)$ for arbitrary $s > 0$. The result of Cohen for $\mu \downarrow 0$ is also contained in the more general result of Falin [6, Theorem 2.6], which says that, in the limit of $\mu \downarrow 0$, (i) $C(\infty)$ and $N(\infty)$ are independent, (ii) $C(\infty)$ has a truncated Poisson distribution with rate $\lambda + \Omega$, and (iii) an appropriately scaled version of $N(\infty)$ is Gaussian distributed with a certain mean and variance that can be specified. In this paper we will focus entirely on assertion (ii).

Equation (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,

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

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 extremely 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 [3]).

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 $\mu \downarrow 0$. Let $\{p_i(\mu); 0 \le i \le s\}$ denote the stationary distribution of the number of busy servers in the system with retrial rate μ. 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

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

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 $\lambda + \Omega$. 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 [3]). 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 Section 4 we give several new results for the key function that governs Cohen's equation (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 Section 4 are in fact new results for the Erlang B formula and Mills' ratio for the Poisson distribution. The proofs of all the results are given in Section 5.

3.1. Halfin–Whitt regime

The Halfin–Whitt regime for multiserver systems refers to the scaling of the arrival rate λ and the number of servers s such that, while both λ and s increase toward ∞, the traffic intensity $\rho = \lambda/s$ approaches 1 and

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

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. [9]).

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

$$\lim_{s \to \infty} \sqrt{s} B(s, \lambda) = \frac{\varphi(\gamma)}{\Phi(\gamma)}, \tag{3}$$

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

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

We now apply the same scaling to the multiserver retrial system.

Theorem 1. *For $\lambda = s - \gamma\sqrt{s}$, with fixed $\gamma > 0$, and Ω defined as in (1),*

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

and

$$\lim_{s\to\infty} \sqrt{s} B(s, \lambda + \Omega) = a \tag{4}$$

with a the unique positive solution of the equation

$$a = \frac{\varphi(\gamma - a)}{\Phi(\gamma - a)}. \tag{5}$$

Theorem 1 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 (4). Theorem 1 thus says that the blocking probability of 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 depends only on γ. This then gives the QED approximations

$$\Omega \approx a\sqrt{s}, \qquad B(s, \lambda + \Omega) \approx \frac{\varphi(\gamma - a)}{\sqrt{s}\Phi(\gamma - a)}.$$

Theorem 1 follows from the stronger result given in Subsection 5.11. The key idea behind the proof of Theorem 1 is the following. Writing $\Omega = a\sqrt{s}$ and using (1) gives

$$\begin{aligned} \lim_{s\to\infty} \frac{\Omega}{\sqrt{s}} &= \lim_{s\to\infty} \frac{s - (\gamma - a)\sqrt{s}}{\sqrt{s}} B(s, s - (\gamma - a)\sqrt{s}) \\ &= \lim_{s\to\infty} \sqrt{s} B(s, s - (\gamma - a)\sqrt{s}), \end{aligned} \tag{6}$$

so that the result follows from (3). Note that in (6) we ignore the fact that, for finite s, the factor a is not only a function of γ (through (5)) but also of s. Therefore, the steps in (6) are only serving as a heuristic. The formal proof of Theorem 1 will take into account this dependence on s and starts from a transformed version of (1) that is easier to work with in the Halfin–Whitt regime (see Section 4). That is, we let $\lambda = s - \gamma\sqrt{s}$ in (1), and we write the unique solution Ω as $a_s(\gamma)\sqrt{s}$, where $a_s(\gamma)$ is referred to as the *retrial factor*. In terms of retrial factors, Theorem 1 can be reformulated as $a_s(\gamma) \to a_\infty(\gamma)$ as $s \to \infty$, where $a = a_\infty(\gamma)$ is the unique positive solution of (5) when $\gamma > 0$. In Subsection 3.2 we present several properties of the retrial factor that help in understanding the behavior of the retrial system.

3.2. Properties of the retrial factor

We first present several results for the retrial factor $a_s(\gamma)$ with finite s.

Theorem 2. *The retrial factor $a_s(\gamma)\colon (0, \sqrt{s}) \to (0, \infty)$ is a positive, decreasing, and convex function of $\gamma \in (0, \sqrt{s})$.*

See Figure 1. Theorem 2 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 1 for $\gamma = 0$ to 0 for $\gamma = \sqrt{s}$. We expect the retrial factor to increase with the load, since an increased load leads to more blocked calls.

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

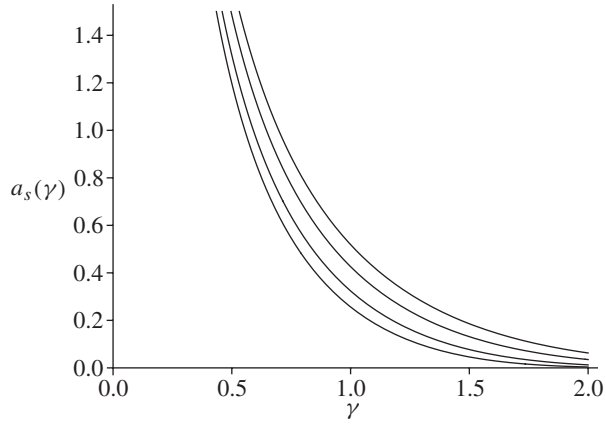

FIGURE 1: The function $a_s(\gamma)$ for $s = 10, 20, 100, \infty$ (ordered upwards).

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

Theorem 3. *Let $s \geq 1$ be fixed. Then, for $0 < \gamma < \sqrt{s}$,*

$$a_s(\gamma) = \frac{1}{\gamma} - \frac{2}{\sqrt{s}} - \left(1 - \frac{2}{s}\right)\gamma + O(\gamma^2). \tag{7}$$

That $a_s(\gamma) \to \infty$ as $\gamma \downarrow 0$ is intuitively clear. Furthermore, that $a_s(\gamma)$ is of the order $1/\gamma$ as $\gamma \downarrow 0$ can be anticipated from Theorem 1 and the well-known result

$$\frac{\varphi(\delta)}{\Phi(\delta)} = -\delta + O(\delta^{-1}) \quad \text{as } \delta \to -\infty.$$

We next complement the asymptotic result in (7) with some basic inequalities satisfied by $a_s(\gamma)$.

Proposition 1. *For $0 < \gamma < \sqrt{s}$, $\gamma a_s(\gamma)$ is a monotonically decreasing function, and*

$$\frac{1}{\gamma} - \frac{2}{\sqrt{s}} - \gamma < a_s(\gamma) < \frac{1}{\gamma} - \frac{1}{\sqrt{s}}. \tag{8}$$

See Figure 2. Note that the additional arrivals $a_s(\gamma)\sqrt{s}$ due to retrials is of the same order as the overcapacity $\gamma\sqrt{s}$, and these 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 0, the blocking probability approaches 1. To see this, we can use the fact that $a_s(\gamma)$ tends to ∞, like $1/\gamma$, and, hence, for fixed s and $\gamma \downarrow 0$ ($a \to \infty$),

$$B(s, s - (\gamma - a)\sqrt{s}) = 1 - \frac{\sqrt{s}}{a - \gamma} + O\left(\left(\frac{1}{a - \gamma}\right)^2\right),$$

where we have also used the well-known fact that $B(s, s - \delta\sqrt{s}) = 1 + \sqrt{s}/\delta + O(1/\delta^2)$ as $\delta \to -\infty$.

Proposition 1 indicates that additional arrivals $a_s(\gamma)\sqrt{s}$ and overcapacity are of the same order of magnitude when γ is of order unity. A precise result for this is that

$$a_s(\gamma) = \gamma \quad \text{when} \quad \gamma = \gamma_s^* = \sqrt{s}\, B(s, s). \tag{9}$$

This γ_s^* lies between $\frac{1}{2}$ ($s = 1$) and $(2/\pi)^{1/2} = 0.79788\ldots$ ($s = \infty$). See Figure 1.

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

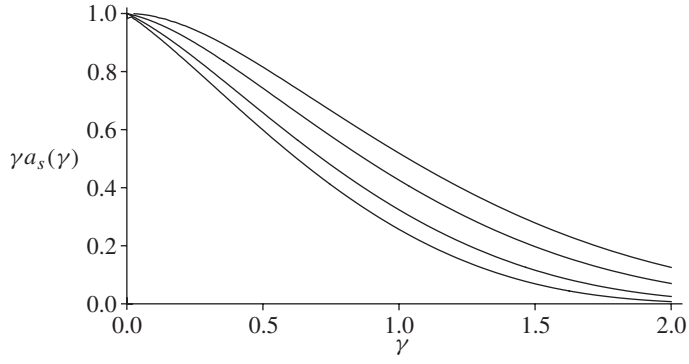

FIGURE 2: The function $\gamma a_s(\gamma)$ for $s = 10, 20, 100, \infty$ (ordered upwards).

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

Theorem 4. *Let $s \geq 1$ be fixed. Then, for $\gamma \uparrow \sqrt{s}$,*

$$a_s(\gamma) = \frac{s^{s+1/2}}{s!}\left(1 - \frac{\gamma}{\sqrt{s}}\right)^{s+1}\left(1 + O\left(\left(1 - \frac{\gamma}{\sqrt{s}}\right)\right)\right). \tag{10}$$

There is the following monotonicity result.

Theorem 5. *The retrial factor $a_s(\gamma)$, with $\gamma \in (0, \sqrt{s})$ fixed, increases monotonically in $s \geq 1$.*

See Figures 1 and 2. There seems no easy explanation for Theorem 5. Note that the traffic intensity, with $\gamma \in (0, \sqrt{s})$ fixed,

$$\rho = \frac{\lambda}{s} = 1 - \frac{\gamma}{\sqrt{s}},$$

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

Theorem 5 gives useful complementary information on Theorem 1 in that the asymptotic retrial factor $a_\infty(\gamma)$ is approximated monotonically from below by the retrial factor $a_s(\gamma)$ as $s \to \infty$. We now present several properties on this $a_\infty(\gamma)$.

Theorem 6. *For any $\gamma > 0$, (5) has a unique solution $a = a_\infty(\gamma)$. This $a_\infty(\gamma)$ is a strictly decreasing and convex function of $\gamma > 0$. Moreover,*

$$a_\infty(\gamma) = \frac{1}{\gamma} - \gamma + 2\gamma^3 - 20\gamma^5 + 82\gamma^7 + O(\gamma^9), \qquad \gamma > 0, \tag{11}$$

and there is the inequality

$$\frac{1}{\gamma} - \gamma < a_\infty(\gamma) < \frac{1}{\gamma}, \qquad \gamma > 0. \tag{12}$$

Finally,

$$a_\infty(\gamma) = O(\mathrm{e}^{-\gamma^2/2}), \qquad \gamma \geq 1. \tag{13}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

4. Cohen's equation 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 $a > 0$, Cohen's equation (1) takes the form

$$a = f_s(\gamma - a), \tag{14}$$

where, for $\delta < \sqrt{s}$,

$$f_s(\delta) = \sqrt{s}\left(1 - \frac{\delta}{\sqrt{s}}\right)B(s, s - \delta\sqrt{s}). \tag{15}$$

For solving (14) and understanding its solution $a = a_s(\gamma)$, it is necessary to study the function f_s. In Subsection 4.1 we present a number of results for f_s that are finite s-versions of known results for (the reciprocal of) Mills' ratio

$$f_\infty(\delta) = \frac{\varphi(\delta)}{\Phi(\delta)} = \frac{\mathrm{e}^{-\delta^2/2}}{\int_{-\infty}^{\delta} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta'}, \qquad \delta \in \mathbb{R}, \tag{16}$$

for the normal distribution. With regard to analytic properties, one can view $f_s^{-1}(\delta)$ as the appropriate version of Mills' ratio for the Poisson distribution. In Subsection 4.2 we indicate how the various results for f_s and f_∞ can be used to prove the results in Section 3 on a_s and a_∞, and we comment on the interrelations between the various results on f_s. From this, a technical workplan for proving the results on a_s and a_∞ in Section 3 and those on f_s and f_∞ emerges that is carried out in detail in Section 5. Finally, in Subsection 4.3 we present a Newton-type computational method, based on the results on f_s and f_∞, for a_s and a_∞.

4.1. Properties of f_s and f_∞

The properties of f_s given in this subsection can be derived from the integral representation of $B(s, \lambda)$ in (2) that can be used to represent f_s in quasi-Gaussian form. For this, we let

$$\alpha_s(\delta) = \left(-2s\left(\frac{\delta}{\sqrt{s}} + \ln\left(1 - \frac{\delta}{\sqrt{s}}\right)\right)\right)^{1/2}, \qquad \delta < \sqrt{s}, \tag{17}$$

where the square root is taken such that $\operatorname{sgn}(\alpha_s(\delta)) = \operatorname{sgn}(\delta)$. With $\lambda = s - \delta\sqrt{s}$ and the substitution $\lambda' = s - \delta'\sqrt{s}$ in the integral in (2), we can write $f_s(\delta)$ in terms of α_s as

$$f_s(\delta) = \frac{(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha_s^2(\delta)/2}}{\int_{-\infty}^{\delta} \mathrm{e}^{-\alpha_s^2(\delta')/2}\,\mathrm{d}\delta'}, \qquad \delta < \sqrt{s}; \tag{18}$$

compare (16). A power series expansion in (17) yields, for $\delta \in \mathbb{R}$,

$$\alpha_s(\delta) = \delta\left(1 + \frac{2\delta}{3\sqrt{s}} + \cdots\right)^{1/2} = \delta + O\left(\frac{\delta^2}{\sqrt{s}}\right) \quad \text{as } s \to \infty, \tag{19}$$

and this can be used in showing that $f_s(\delta) \to f_\infty(\delta)$ as $s \to \infty$ for $\delta \in \mathbb{R}$; see Proposition 6 below.

There is a number of (known) properties of f_∞ that turn out to have finite s-versions. These are the asymptotic expansion

$$f_\infty(\delta) = -\delta - \frac{1}{\delta} + \frac{2}{\delta^3} - \frac{10}{\delta^5} + \frac{74}{\delta^7} + O\left(\frac{1}{\delta^9}\right) \quad \text{as } \delta \to -\infty, \tag{20}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

the inequalities for derivatives

$$f_\infty(\delta) > -\delta, \qquad -1 < f'_\infty(\delta) < 0, \qquad f''_\infty(\delta) > 0, \qquad \delta \in \mathbb{R}, \tag{21}$$

and the bounds

$$-\tfrac{3}{4}\delta + \tfrac{1}{4}(\delta^2+8)^{1/2} < f_\infty(\delta) < -\tfrac{1}{2}\delta + \tfrac{1}{2}(\delta^2+4)^{1/2}, \qquad \delta \in \mathbb{R}. \tag{22}$$

The results in (21) and (22) were obtained by Sampford [11], while (20) is obtained from the asymptotic expansion of the error function; see Subsection 5.11. For f_s, we show the following results in Section 5.

Proposition 2. *Let $s \geq 1$ be fixed. We have, for $\delta < 0$,*

$$f_s(\delta) = -\delta - \frac{1}{\delta} - \frac{2}{\delta^2\sqrt{s}} + \left(2 - \frac{6}{s}\right)\frac{1}{\delta^3} + O\left(\frac{1}{\delta^4}\right). \tag{23}$$

Proposition 3. *For $s \geq 1$ and $\delta < \sqrt{s}$, we have*

$$f_s(\delta) > -\delta, \qquad -1 < f'_s(\delta) < 0, \qquad f''_s(\delta) > 0. \tag{24}$$

Proposition 4. *For $\delta \in \mathbb{R}$, let*

$$L_s(\delta) = -\frac{3}{4}\delta - \frac{1}{2\sqrt{s}} + \frac{1}{4}\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2}, \tag{25}$$

$$U_s(\delta) = -\frac{1}{2}\delta - \frac{1}{2\sqrt{s}} + \frac{1}{2}\left(\left(\delta - \frac{1}{\sqrt{s}}\right)^2 + 4\right)^{1/2}. \tag{26}$$

For $s \geq 1$ and $\delta < \sqrt{s}$, we have

$$L_s(\delta) < f_s(\delta) < U_s(\delta). \tag{27}$$

Furthermore, with regard to dependence on s of $f_s(\delta)$, we have the following results. Define $f_s(\delta) = 0$, $\delta \geq \sqrt{s}$.

Proposition 5. *$f_s(\delta)$ increases to $f_\infty(\delta)$ for any $\delta \in \mathbb{R}$.*

Proposition 6. *We have*

$$f_s(\delta) - f_\infty(\delta) = O\left(\frac{1}{\sqrt{s}}\right),$$

uniformly in any compact set of $\delta \in \mathbb{R}$.

4.2. Technical workplan

We now indicate how the results of Propositions 2–6 on f_s are used to prove the results on a_s given in Section 3; the results on a_∞ as given in Theorem 6 follow in a similar way from the results in (20)–(22) on f_∞.

Proposition 3 gives, for general $s \geq 1$, the existence and uniqueness of the solution of $a = a_s(\gamma)$ of Cohen's equation (14) in the Halfin–Whitt regime. See Figure 3. Furthermore, Theorem 2 on the monotonicity and convexity of $a_s(\gamma)$ as a function of γ is shown from Proposition 3. Next, Theorem 3 on the behavior of $a_s(\gamma)$ as $\gamma \downarrow 0$ can be established from Proposition 2; Theorem 4 on the behavior of $a_s(\gamma)$ as $\gamma \uparrow \sqrt{s}$ follows from more elementary considerations. The two bounds on $a_s(\gamma)$ in Proposition 1 follow from the two bounds on $f_s(\delta)$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

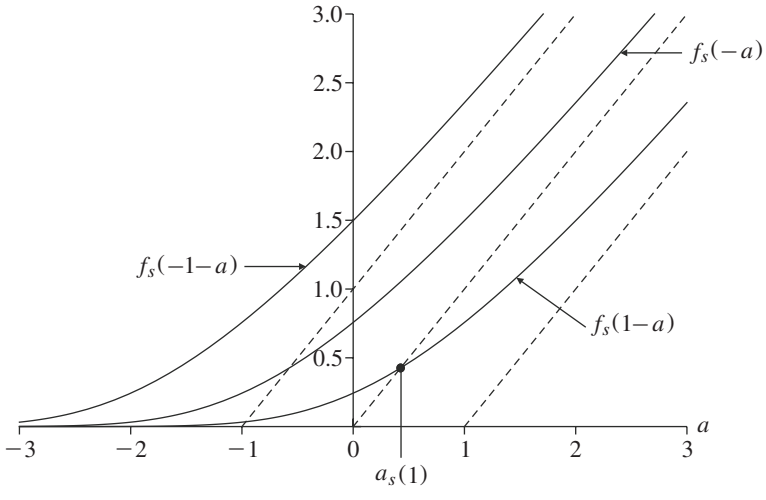

FIGURE 3: The function $f_s(\gamma - a)$ for $s = 100$ and $\gamma = -1, 0, 1$.

in Proposition 4. Finally, Theorem 5 on the monotonic dependence of $a_s(\gamma)$ on s is shown by using Propositions 3 and 5.

For the proof of Theorem 1, which states that $a_s(\gamma) \to a_\infty(\gamma)$ as $s \to \infty$ for any $\gamma > 0$, Propositions 5 and 6 as well as the inequalities $f_s'(\delta) > -1$ and $f_\infty'(\delta) > -1$ are used. It also follows from these results that $a_s(\gamma) = a_\infty(\gamma) + O(1/\sqrt{s})$, uniformly in any compact set of $\gamma > 0$, and that $a_s(\gamma)$ increases to $a_\infty(\gamma)$ for any $\gamma > 0$.

The details of the proofs for the above results on a_s and a_∞ are given in Section 5, together with the proofs of Propositions 2–6 for f_s. The results and proofs of Propositions 2–4 are strongly interrelated. The inequality $f_s(\delta) < U_s(\delta)$ in (27) is equivalent to $f_s'(\delta) > -1$, see (24), and the inequality $f_s(\delta) > L_s(\delta)$ in (27) is equivalent to $f_s''(\delta) > 0$ in (24). A similar situation occurs for the inequalities in (21) and (22) involving f_∞. While the inequalities in (22) follow rather easily from the inequalities in (21), see [11], it turns out that, for f_s, one better proceeds by proving Propositions 3–4 in a combined effort. In the latter approach, a crucial role is played by the asymptotic result for $f_s(\delta)$ as $\delta \to -\infty$ in Proposition 2, yielding validity of the bounds on f_s in (27) for large negative δ. Hence, Section 5 starts with proving Proposition 2; see Subsection 5.1. Next, in Subsection 5.2, the equivalence of the inequalities on derivatives of f_s in Proposition 3 with bounds on f_s in Proposition 4 is shown, together with validity of the bounds in (27) for large negative δ from Proposition 2. Finally, the proofs of both Proposition 3 and 4 are completed by showing corresponding inequalities to (27) for the derivatives of the occurring functions for all $\delta < \sqrt{s}$.

In Section 5 we then proceed with establishing the properties of the retrial factors as given in Subsection 3.2. Propositions 5 and 6 on the (monotonic) approximation of $f_\infty(\delta)$ by $f_s(\delta)$ as $s \to \infty$ are shown, and for this the quasi-Gaussian representation (18) of f_s is crucial. Then the proof of the results in Theorem 6 on $a_\infty(\gamma)$ is briefly indicated, and, finally, we show from Propositions 5–6 and the inequalities $f_s'(\delta) > -1$ and $f_\infty'(\delta) > -1$ that $a_s(\gamma) = a_\infty(\gamma) + O(1/\sqrt{s})$ as $s \to \infty$, uniformly in any compact set of $\gamma > 0$.

4.3. Computation scheme for $a_s(\gamma)$ and $a_\infty(\gamma)$

For the computation of $a_s(\gamma)$ and $a_\infty(\gamma)$, a simple Newton iteration works quite well due to convexity of $f = f_s$ and f_∞; see Theorem 6 and (24). When $\gamma > 0$ is not too small,

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

initialization can be done by setting $a^{(0)} = 0$. When γ is close to 0, the initialization should be done using (7) and (11) for the respective cases. In all cases, convergence is quadratic and monotonic after one step. The Newton step

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

is implemented conveniently, using

$$f_s'(\delta) = \frac{-f_s(\delta)}{1 - \delta/\sqrt{s}}\left(\delta + \frac{1}{\sqrt{s}} + f_s(\delta)\right),$$

see (33) below, and

$$f_\infty'(\delta) = -f_\infty(\delta)(\delta + f_\infty(\delta)),$$

respectively.

5. Proofs

In this section we present the proofs of all the results of this paper, where we follow the workplan as outlined in Subsection 4.2.

5.1. Proof of Proposition 2

With the substitution $t = (\lambda'/\lambda) - 1$, we have, from (2),

$$\lambda B(s, \lambda) = \left(\int_0^\infty \mathrm{e}^{-\lambda t}(1+t)^s \,\mathrm{d}t\right)^{-1}; \tag{28}$$

see [8]. From (28) with $\lambda = s - \delta\sqrt{s}$, we obtain

$$f_s(\delta) = \frac{1}{\sqrt{s}}(s - \delta\sqrt{s})B(s, s - \delta\sqrt{s}) = \left(\sqrt{s}\int_0^\infty \mathrm{e}^{-st+\delta t\sqrt{s}}(1+t)^s \,\mathrm{d}t\right)^{-1}.$$

Next, with the substitution $t = x/s$, we obtain

$$(f_s(\delta))^{-1} = \frac{1}{\sqrt{s}}\int_0^\infty \mathrm{e}^{\delta x/\sqrt{s}}\mathrm{e}^{-x}\left(1 + \frac{x}{s}\right)^s \mathrm{d}x. \tag{29}$$

We expand

$$\mathrm{e}^{-x}\left(1 + \frac{x}{s}\right)^s = \sum_{j=0}^\infty c_j(s)x^j, \qquad |x| < s.$$

On account of

$$\mathrm{e}^{-x}\left(1 + \frac{x}{s}\right)^s = \sum_{j=0}^J c_j(s)x^j + O(x^{J+1}), \qquad x > 0,$$

for any $J = 0, 1, \ldots$, we then obtain

$$(f_s(\delta))^{-1} = \frac{1}{\sqrt{s}}\sum_{j=0}^J (-1)^{j+1} j!\, c_j(s)\left(\frac{\sqrt{s}}{\delta}\right)^{j+1} + O\left(\left(\frac{1}{\delta}\right)^{J+2}\right) \quad \text{as } \delta \to -\infty. \tag{30}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

The $c_j(s)$ can be found in finite terms by multiplying the power series of $\exp(-x)$ and that of $(1+x/s)^s$. Thus,

$$c_j(s) = \sum_{i,l\ge 0,\, i+l=j} \frac{(-1)^i}{i!} \frac{s(s-1)\cdots(s-l+1)}{l!\, s^l}.$$

The first few $c_j(s)$ are given by $c_0 = 1$, $c_1 = 0$, and

$$c_2(s) = \frac{-1}{2s}, \qquad c_3(s) = \frac{1}{3s^2}, \qquad c_4(s) = \frac{1}{8s^2} - \frac{1}{4s^3}.$$

It then follows from (30) with $J = 4$ that

$$(f_s(\delta))^{-1} = -\frac{1}{\delta} + \frac{1}{\delta^3} + \frac{2}{\delta^4\sqrt{s}} - \left(3 - \frac{6}{s}\right)\frac{1}{\delta^5} + O\left(\frac{1}{\delta^6}\right) \quad \text{as } \delta \to -\infty.$$

Therefore,

$$f_s(\delta) = -\delta\left(1 - \frac{1}{\delta^2} - \frac{2}{\delta^3\sqrt{s}} + \left(3 - \frac{6}{s}\right)\frac{1}{\delta^4} + O\left(\frac{1}{\delta^5}\right)\right)^{-1} \quad \text{as } \delta \to -\infty, \tag{31}$$

and the results of Proposition 2 are obtained from (31) by expanding $(1-x)^{-1} = 1 + x + x^2 + O(x^3)$, $|x| < \frac{1}{2}$.

5.2. Proofs of Propositions 3 and 4

We suppress s in $f_s(\delta)$ and $\alpha_s(\delta)$ until Subsection 5.9. Let $g(\delta) = g_s(\delta) = f(\delta)/(1-\delta/\sqrt{s})$.

Lemma 2. *With the prime denoting differentiation with respect to δ, for $\delta < \sqrt{s}$,*

$$\left(-\tfrac{1}{2}\alpha^2(\delta)\right)' = \frac{-\delta}{1-\delta/\sqrt{s}}, \tag{32}$$

$$f'(\delta) = -g(\delta)\left(\delta + \frac{1}{\sqrt{s}} + f(\delta)\right), \tag{33}$$

$$g'(\delta) = -g(\delta)\left(\frac{\delta}{1-\delta/\sqrt{s}} + g(\delta)\right),$$

$$f''(\delta) = g(\delta)\left(\delta + \frac{1}{\sqrt{s}} + f(\delta)\right)\left(2g(\delta) + \frac{\delta}{1-\delta/\sqrt{s}}\right) - g(\delta). \tag{34}$$

Proof. Straightforward verification from (17) and (18).

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

$$f(\delta) > -\delta \quad \Longleftrightarrow \quad -\frac{1}{\delta}\,\mathrm{e}^{-\alpha^2(\delta)/2}\left(1 - \frac{\delta}{\sqrt{s}}\right) > I(\delta), \tag{35}$$

where, see (18),

$$I(\delta) = \int_{-\infty}^{\delta} \mathrm{e}^{-\alpha^2(\delta')/2}\,\mathrm{d}\delta', \qquad \delta < \sqrt{s}. \tag{36}$$

From Proposition 2, it is seen that $f(\delta) > -\delta$ holds for large negative δ. We compute, using (32),

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

Therefore, the two inequalities in (35) hold for all $\delta < 0$.

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

We next show that $f'(\delta) > -1$. Using $g(\delta) = f(\delta)/(1-\delta/\sqrt{s})$ and (33), we have, for $\delta < \sqrt{s}$,

$$f'(\delta) > -1 \quad \Longleftrightarrow \quad f(\delta)\left(\delta + \frac{1}{\sqrt{s}} + f(\delta)\right) < 1 - \frac{\delta}{\sqrt{s}}. \tag{37}$$

The inequality in the second statement in (37) can be written as

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

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 $\delta < \sqrt{s}$,

$$f'(\delta) > -1 \quad \Longleftrightarrow \quad f(\delta) < \left(1 + \frac{1}{4}\left(\delta - \frac{1}{\sqrt{s}}\right)^2\right)^{1/2} - \frac{1}{2}\left(\delta + \frac{1}{\sqrt{s}}\right). \tag{38}$$

The inequality in the second member of (38) is the second inequality in Proposition 4 which will be proved below. Therefore, $f'(\delta) > -1$ holds for all $\delta < \sqrt{s}$. The inequality $f'(\delta) < 0$ follows from (33), and so we have shown that $-1 < f'(\delta) < 0$ for $\delta < \sqrt{s}$.

We finally show that $f''(\delta) > 0$. It follows from (34), the positivity of $g(\delta)$ when $\delta < \sqrt{s}$, and $g(\delta) = f(\delta)/(1-\delta/\sqrt{s})$ that, for $\delta < \sqrt{s}$,

$$f''(\delta) > 0 \quad \Longleftrightarrow \quad \left(\delta + \frac{1}{\sqrt{s}} + f(\delta)\right)(2f(\delta) + \delta) > 1 - \frac{\delta}{\sqrt{s}}. \tag{39}$$

The second inequality in (39) can be written as

$$\left| f(\delta) + \frac{3}{4}\delta + \frac{1}{2\sqrt{s}}\right| > \frac{1}{4}\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2}. \tag{40}$$

Now $f(\delta) > -\delta > -\frac{3}{4}\delta - 1/2\sqrt{s} - \frac{1}{4}(\delta - 2/\sqrt{s})^2$ for $\delta < \sqrt{s}$, and so we have, for $\delta < \sqrt{s}$,

$$f''(\delta) > 0 \quad \Longleftrightarrow \quad f(\delta) > \frac{1}{4}\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2} - \frac{3}{4}\delta - \frac{1}{2\sqrt{s}}. \tag{41}$$

The inequality in the second member of (41) is the first inequality in Proposition 4 which will be proved below. Hence, (40) holds for $\delta < \sqrt{s}$ and so $f''(\delta) > 0$ for $\delta < \sqrt{s}$. This completes the proof of Proposition 3.

We now turn to the proof of Proposition 4. We first show that, for $\delta < \sqrt{s}$,

$$f(\delta) < -\frac{1}{2}\left(\delta + \frac{1}{\sqrt{s}}\right) + \frac{1}{2}\left(\left(\delta - \frac{1}{\sqrt{s}}\right)^2 + 4\right)^{1/2} =: F(\delta). \tag{42}$$

From (18) we have

$$f(\delta) < F(\delta) \quad \Longleftrightarrow \quad I(\delta) > \frac{(1-\delta/\sqrt{s})\mathrm{e}^{-\alpha^2(\delta)/2}}{F(\delta)} =: S(\delta), \tag{43}$$

where I is the integral in (36). From (32) we compute

$$S'(\delta) = \frac{-\delta}{F(\delta)}\mathrm{e}^{-\alpha^2(\delta)/2} - \frac{F(\delta)/\sqrt{s} + (1-\delta/\sqrt{s})F'(\delta)}{F^2(\delta)}\mathrm{e}^{-\alpha^2(\delta)/2}. \tag{44}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

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

$$F^2(\delta) > -\left(\delta + \frac{1}{\sqrt{s}}\right)F(\delta) - \left(1 - \frac{\delta}{\sqrt{s}}\right)F'(\delta) \tag{45}$$

by (44). Now, from the definition of $F(\delta)$ given in (42), we have

$$F^2(\delta) + \left(\delta + \frac{1}{\sqrt{s}}\right)F(\delta) = F(\delta)\left(F(\delta) + \delta + \frac{1}{\sqrt{s}}\right) = 1 - \frac{\delta}{\sqrt{s}},$$

and so (45) is equivalent to $F'(\delta) > -1$. We compute

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

and so $I'(\delta) > S'(\delta)$ for all $\delta < \sqrt{s}$. Therefore, it is enough to show that $I(\delta) > S(\delta)$ for large negative δ. From (42), (43) and (38), (37), we have, for $\delta < \sqrt{s}$,

$$I(\delta) > S(\delta) \quad \Longleftrightarrow \quad \varphi(\delta) := f(\delta)\left(\delta + \frac{1}{\sqrt{s}} + f(\delta)\right) < 1 - \frac{\delta}{\sqrt{s}}. \tag{46}$$

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

$$\varphi(\delta) = 1 - \frac{\delta}{\sqrt{s}} + \frac{1}{\delta\sqrt{s}} + O(\delta^{-2}) < 1 - \frac{\delta}{\sqrt{s}}$$

for large negative δ. Hence, the two statements in (46) hold for large negative δ and the proof of (42) is complete.

We next show that, for $\delta < \sqrt{s}$,

$$f(\delta) > -\left(\frac{3}{4}\delta + \frac{1}{2\sqrt{s}}\right) + \frac{1}{4}\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2} =: E(\delta). \tag{47}$$

From (18) we have

$$f(\delta) > E(\delta) \quad \Longleftrightarrow \quad I(\delta) < \frac{(1 - \delta/\sqrt{s})\mathrm{e}^{-\alpha^2(\delta)/2}}{E(\delta)} =: R(\delta). \tag{48}$$

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

$$I'(\delta) < R'(\delta) \quad \Longleftrightarrow \quad E^2(\delta) < -\left(\delta + \frac{1}{\sqrt{s}}\right)E(\delta) - \left(1 - \frac{\delta}{\sqrt{s}}\right)E'(\delta).$$

We now compute

$$E^2(\delta) + \left(\delta + \frac{1}{\sqrt{s}}\right)E(\delta) = -\frac{1}{8}\delta^2 - \frac{3\delta}{4\sqrt{s}} + \frac{1}{2} - \frac{1}{8}\delta\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2}.$$

Then using the fact that

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

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

we find that

$$I'(\delta) < R'(\delta) \iff -\frac{1}{8}\delta^2 - \frac{3\delta}{4\sqrt{s}} + \frac{1}{2} - \frac{1}{8}\delta\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2} < -\left(1 - \frac{\delta}{\sqrt{s}}\right)\left(-\frac{3}{4} + \frac{1}{4}\,\frac{\delta - 2/\sqrt{s}}{((\delta - 2/\sqrt{s})^2 + 8)^{1/2}}\right).$$

With some algebra, this becomes

$$I'(\delta) < R'(\delta) \iff (\delta^2 + 2)\left(\left(\delta - \frac{2}{\sqrt{s}}\right)^2 + 8\right)^{1/2} > -\delta^3 + \frac{2}{\sqrt{s}}\delta^2 - 6\delta - \frac{4}{\sqrt{s}}. \quad (49)$$

Setting $x = -\delta$ and taking squares, the inequality in the second proposition of (49) is implied by

$$(x^2 + 2)^2\left(\left(x + \frac{2}{\sqrt{s}}\right)^2 + 8\right) > \left(x^3 + \frac{2}{\sqrt{s}}x^2 + 6x - \frac{4}{\sqrt{s}}\right)^2.$$

Working this out and simplifying finally leads to the condition $(x + \sqrt{s})^2 > 0$, which obviously holds. Hence, $I'(\delta) < R'(\delta)$ holds for all $\delta < \sqrt{s}$. Therefore, it is enough to show that $I(\delta) < R(\delta)$ for large negative δ. From (47), (48) and (41), (39), we have, for $\delta < \sqrt{s}$,

$$I(\delta) < R(\delta) \iff \psi(\delta) := \left(\delta + \frac{1}{\sqrt{s}} + f(\delta)\right)(2f(\delta) + \delta) > 1 - \frac{\delta}{\sqrt{s}}. \quad (50)$$

Now using the full strength of Proposition 2, we obtain

$$\begin{aligned}\psi(\delta) &= \left(-\frac{1}{\delta} + \frac{1}{\sqrt{s}} - \frac{2}{\delta^2\sqrt{s}} + \left(2 - \frac{6}{s}\right)\frac{1}{\delta^3} + O(\delta^{-4})\right)\left(-\delta - \frac{2}{\delta} - \frac{4}{\delta^2\sqrt{s}} + O(\delta^{-3})\right)\\ &= 1 - \frac{\delta}{\sqrt{s}} + \frac{2}{s\delta^2} + O(\delta^{-3})\\ &> 1 - \frac{\delta}{\sqrt{s}}\end{aligned}$$

for large negative δ. Hence, the two statements in (50) hold for large negative δ, and the proof of (47) is complete. This completes the proof of Proposition 4.

5.3. Solving (14): existence and uniqueness

Assume that $\gamma \le 0$. We have, from Proposition 3,

$$f(\gamma - a) > -(\gamma - a) \ge a$$

for any $a > 0$. Hence, (14) does not have a solution.

Next assume that $\gamma \ge \sqrt{s}$. Since $f(\delta) = 0$, $\delta \ge \sqrt{s}$, we have $f(\gamma - a) = 0$ at $a = 0$ while $\mathrm{d}f(\gamma - a)/\mathrm{d}a < 1$ for all $a > 0$. Again, it follows that (14) has no solution.

Finally, assume that $0 < \gamma < \sqrt{s}$. It follows from Proposition 2 that

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

for large positive a. Also, $f(\gamma - a) > 0$ at $a = 0$. Therefore, (14) has at least one solution. This solution is unique since $\mathrm{d}f(\gamma - a)/\mathrm{d}a < 1$ by Proposition 3. See also Figure 3.

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

We further note the following. When $0 < \gamma < \sqrt{s}$ and $a > 0$, $a - a(\gamma)$ and $a - f(\gamma - a)$ have the same sign. Thus, $a < a(\gamma)$ implies that $a < f(\gamma - a)$, and vice versa; see Figure 3. Indeed, we have, by the mean-value theorem,

$$a - f(\gamma - a) = a - f(\gamma - a) - (a(\gamma) - f(\gamma - a(\gamma))) = (a - a(\gamma))(1 + f'(\gamma - b))$$

for some b between a and $a(\gamma)$. Hence, $a - f(\gamma - a)$ and $a - a(\gamma)$ have the same sign since $1 + f'(\gamma - b) > 0$ by Proposition 3.

5.4. Proof of Theorem 2

Positivity is clear. From (14), by implicit differentiation with respect to γ, we compute

$$a'(\gamma) = \frac{f'(\gamma - a(\gamma))}{1 + f'(\gamma - a(\gamma))} < 0, \qquad 0 < \gamma < \sqrt{s}, \tag{51}$$

since $f'(\delta) \in (-1, 0)$ for $\delta < \sqrt{s}$ by Proposition 3. 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 the convexity of f, see Proposition 3, $f'(\gamma - a(\gamma))$ 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 (51) that $a'(\gamma)$ is strictly increasing in $\gamma \in (0, \sqrt{s})$. That is, $a(\gamma)$ is strictly convex.

5.5. Proof of Theorem 3

We first show that $b := \lim_{\gamma \downarrow 0} a(\gamma) = \infty$ (from Theorem 2 we know that b indeed exists). Indeed, if $b < \infty$, we would have, by continuity, $b = \lim_{\gamma \downarrow 0} f(\gamma - a(\gamma)) = f(-b)$, contradicting Proposition 3. Hence, since $\gamma - a(\gamma) \to -\infty$ as $\gamma \downarrow 0$, we can use Proposition 2 to see that

$$a = f(\gamma - a) = -(\gamma - a) - \frac{1}{\gamma - a} - \frac{2}{(\gamma - a)^2\sqrt{s}} + \left(2 - \frac{6}{s}\right)\frac{1}{(\gamma - a)^3} + O\left(\frac{1}{(\gamma - a)^4}\right)$$

as $\gamma \downarrow 0$, in which we have temporarily written $a = a(\gamma)$. Thus,

$$\begin{aligned} \gamma &= \frac{1}{a - \gamma} - \frac{2}{(a - \gamma)^2\sqrt{s}} - \left(2 - \frac{6}{s}\right)\frac{1}{(a - \gamma)^3} + O\left(\frac{1}{(\gamma - a)^4}\right) \\ &= \frac{1}{a - \gamma}(1 + o(1)) \quad \text{as } \gamma \downarrow 0. \end{aligned} \tag{52}$$

Multiplying the first and last members of (52) by $a - \gamma$ and dividing them by γ, it follows that $a = \gamma^{-1}(1 + o(1))$ as $\gamma \downarrow 0$. We now write the first line of (52) as

$$a - \gamma = \frac{1}{\gamma}\left(1 - \frac{2}{(a - \gamma)\sqrt{s}} - \left(2 - \frac{6}{s}\right)\frac{1}{(a - \gamma)^2} + O\left(\frac{1}{(\gamma - a)^3}\right)\right), \tag{53}$$

noting that $(a - \gamma)^{-1} = O(\gamma)$ as $\gamma \downarrow 0$. The form (53) is appropriate for getting ever more precise asymptotic information on $a - \gamma$ by iteration. Thus, we find first that

$$a - \gamma = \frac{1}{\gamma}\left(1 + O\left(\frac{\gamma}{\sqrt{s}}\right) + O(\gamma^2)\right)$$

and then from (53) that

$$a - \gamma = \frac{1}{\gamma}\left(1 - \frac{2\gamma}{\sqrt{s}} + O(\gamma^2)\right).$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

One more iteration yields

$$a-\gamma=\frac{1}{\gamma}\left(1-\frac{2\gamma}{\sqrt{s}}-\left(2-\frac{2}{s}\right)\gamma^2+O(\gamma^3)\right),$$

and this is (7).

5.6. Proof of Proposition 1

Let $L=L_s$ and $U=U_s$ as in (25) and (26). We have $a_L(\gamma)<a(\gamma)<a_U(\gamma)$ when $a_L(\gamma)$ and $a_U(\gamma)$ are the solutions $a>0$ of $a=L(\gamma-a)$ and $a=U(\gamma-a)$, respectively. Indeed, from (14) and (27), we have

$$\begin{aligned}a_L(\gamma)-f(\gamma-a_L(\gamma))&<a_L(\gamma)-L(\gamma-a_L(\gamma))\\&=0\\&=a_U(\gamma)-U(\gamma-a_U(\gamma))<a_U(\gamma)-f(\gamma-a_U(\gamma)),\end{aligned}$$

and so $a_L(\gamma)<a(\gamma)<a_U(\gamma)$ follows from the comment at the end of Subsection 5.3. Solving the equations $a=L(\gamma-a)$ and $a=U(\gamma-a)$ gives $\gamma\,a_L(\gamma)=1-2\gamma/\sqrt{s}-\gamma^2$ and $\gamma\,a_U(\gamma)=1-\gamma/\sqrt{s}$, respectively. This shows (8).

We have, from (33) and (51),

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

where the facts that $f(\delta)=(1-\delta/\sqrt{s})\,g(\delta)$ and $f(\gamma-a)=a$ have been used. Therefore,

$$(\gamma a(\gamma))'=a(\gamma)+\gamma a'(\gamma)=a(\gamma)\left(1-\frac{\gamma(\gamma+1/\sqrt{s})}{1-\gamma(a(\gamma)+1/\sqrt{s})}\right).$$

From (8), it is then seen that $a'(\gamma)$ and $(\gamma\,a(\gamma))'$ are positive.

5.7. Proof of (9)

From $f(0)=f(f(0)-f(0))$, it is seen that $a=f(0)$ solves the equation $a=f(\gamma-a)$ when $\gamma=f(0)$. Hence, $a(f(0))=f(0)$.

5.8. Proof of Theorem 4

Let $c:=\lim_{\gamma\uparrow\sqrt{s}}a(\gamma)$ (from Theorem 2 we know that c indeed exists). We will show that $c=0$. From $c=f(\sqrt{s}-c)$, we obtain, for some d, $\sqrt{s}-c\le d\le\sqrt{c}$, by the mean value theorem,

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

Now $f'(d)>-1$ by Proposition 3, and so $c=0$.

Next, we have, from (2) and (15),

$$\begin{aligned}f(\delta)&=\frac{s^{s+1/2}(1-\delta/\sqrt{s})^{s+1}/s!}{\sum_{k=0}^{s}s^k(1-\delta/\sqrt{s})^k/k!}\\&=\frac{s^{s+1/2}}{s!}\left(1-\frac{\delta}{\sqrt{s}}\right)^{s+1}\left(1+O\left(1-\frac{\delta}{\sqrt{s}}\right)\right),\qquad\delta\le\sqrt{s}.\end{aligned}\tag{54}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

The first form for f in (54) shows that f has a zero of order $s+1$ at $\delta = \sqrt{s}$, and so $f'(\delta) \to 0$ as $\delta \uparrow \sqrt{s}$. Writing, temporarily, $a = a(\gamma)$, we see, from $a = f(\gamma - a)$, by the mean value theorem, that there exists a $\xi \in [\gamma - a, \gamma]$ such that

$$a = f(\gamma) - af'(\xi). \tag{55}$$

Since $f'(\xi) \to 0$ as $\gamma \uparrow \sqrt{s}$ (which follows from $\xi \geq \gamma - a$ and $a \to 0$ as $\gamma \uparrow \sqrt{s}$), we thus see from (54)–(55) that

$$a = \frac{f(\gamma)}{1 + f'(\xi)} = O\left(\left(1 - \frac{\gamma}{\sqrt{s}}\right)^{s+1}\right) \quad \text{as } \gamma \uparrow \sqrt{s}.$$

It then follows from $a = f(\gamma - a)$ and (54) that

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

and this gives (10).

5.9. Proofs of Propositions 5 and 6

We have from (29), by the substitution $x = y\sqrt{s}$,

$$\frac{1}{f_s(\delta)} = \int_0^\infty \mathrm{e}^{\delta y}\mathrm{e}^{-y\sqrt{s}}\left(1 + \frac{y}{\sqrt{s}}\right)^s \mathrm{d}y.$$

Therefore, it suffices to show that, for $y > 0$,

$$\mathrm{e}^{-y\sqrt{s}}\left(1 + \frac{y}{\sqrt{s}}\right)^s = \exp\left(-y\sqrt{s} + s\ln\left(1 + \frac{y}{\sqrt{s}}\right)\right)$$

decreases in $s \geq 1$. We have

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[-y\sqrt{s} + s\ln\left(1 + \frac{y}{\sqrt{s}}\right)\right] = -\frac{1}{2}\left(\frac{y}{\sqrt{s}} + \frac{y/\sqrt{s}}{1 + y/\sqrt{s}}\right) + \ln\left(1 + \frac{y}{\sqrt{s}}\right), \tag{56}$$

and we will show that this is negative for $y > 0$. With $a = y/\sqrt{s} > 0$, we have, from $(u\ln u)' = 1 + \ln u$,

$$(1+a)\ln(1+a) = \int_0^a (1 + \ln(1+v))\,\mathrm{d}v < \int_0^a (1+v)\,\mathrm{d}v = a + \frac{1}{2}a^2,$$

i.e. that

$$\ln(1+a) < \frac{a + a^2/2}{1+a} = \frac{1}{2}\left(a + \frac{a}{1+a}\right).$$

This implies negativity of (56), and Proposition 5 is proved.

For the proof of Proposition 6, we start by analyzing the function

$$J_s(\delta) := \int_{-\infty}^{\delta} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta' - \int_{-\infty}^{\delta} \mathrm{e}^{-\alpha_s^2(\delta')/2}\,\mathrm{d}\delta', \qquad \delta \leq \sqrt{s},$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

where we recall the quasi-Gaussian representation of $f_s(\delta)$ in (18). We have $\alpha_s(0) = 0$ and $\alpha_s(\delta) > \delta$, $0 \neq \delta < \sqrt{s}$, and so $\mathrm{e}^{-\alpha_s^2(\delta)/2} > \mathrm{e}^{-\delta^2/2}$, $\delta < 0$, $\mathrm{e}^{-\alpha_s^2(\delta)/2} < \mathrm{e}^{-\delta^2/2}$, $0 < \delta \leq \sqrt{s}$. It follows that $J_s(\delta)$ decreases from the value 0 at $\delta = -\infty$ to its minimum value

$$\frac{1}{2}\sqrt{2\pi} - (f_s(0))^{-1} = \frac{-2}{3\sqrt{s}} + O\left(\frac{1}{s}\right) \tag{57}$$

at $\delta = 0$. Here we have used the fact that $f_s(0) = \sqrt{s}B(0, 0)$ and [9, Theorem 11]. Furthermore, $J_s(\delta)$ increases from value (57) at $\delta = 0$ to the value

$$\begin{aligned} J_s(\sqrt{s}) &= \int_{-\infty}^{\sqrt{s}} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta' - \int_{-\infty}^{\sqrt{s}} \mathrm{e}^{-\alpha_s^2(\delta')/2}\,\mathrm{d}\delta' \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta' - \int_{\sqrt{s}}^{\infty} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta' - \int_{-\infty}^{\sqrt{s}} \mathrm{e}^{\delta'\sqrt{s}}\left(1 - \frac{\delta'}{\sqrt{s}}\right)\mathrm{d}\delta' \\ &= \sqrt{2\pi} + O(\mathrm{e}^{-s}) - \sqrt{s}\int_0^{\infty} \mathrm{e}^{(1-t)s} t^s\,\mathrm{d}t \\ &= \sqrt{2\pi} + O(\mathrm{e}^{-s}) - \frac{\mathrm{e}^s \Gamma(s+1)}{s^{s+1/2}}. \end{aligned}$$

Here, the definition of α_s given in (17) has been used and subsequently the substitution $\delta' = (1-t)\sqrt{s}$ has been used to write the integral involving α_s in terms of the gamma integral. By Stirling's formula, $\Gamma(s+1) = s^{s+1/2}\mathrm{e}^{-s}\sqrt{2\pi}(1 + O(1/s))$, it is seen that $J_s(\sqrt{s}) = O(1/s)$. It follows that $J_s(\delta) = O(1/\sqrt{s})$ uniformly in $\delta \leq \sqrt{s}$.

Now let $-\infty < \delta_0 < \delta_1 < \sqrt{s}$. Both $\int_{-\infty}^{\delta} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta'$ and $\int_{-\infty}^{\delta} \mathrm{e}^{-\alpha_s^2(\delta')/2}\,\mathrm{d}\delta'$ are bounded away from 0 when $\delta_0 \leq \delta \leq \delta_1$, while their difference $J_s(\delta) = O(1/\sqrt{s})$ uniformly on $[\delta_0, \delta_1]$. From (18) and (19), considered on the compact interval $[\delta_0, \delta_1]$, we see that $f_\infty(\delta) - f_s(\delta) = O(1/\sqrt{s})$ uniformly in $\delta \in [\delta_0, \delta_1]$. This proves Proposition 6.

5.10. Proof of Theorem 5

Let $\gamma \in (0, \sqrt{s})$ be fixed. By implicit differentiation of the equation $a_s(\gamma) = f_s(\gamma - a_s(\gamma))$ with respect to s, we obtain

$$\frac{\partial}{\partial s}(a_s(\gamma)) = \frac{\partial f_s(\gamma - a_s(\gamma))/\partial s}{1 + \partial f_s(\gamma - a_s(\gamma))/\partial \gamma} > 0, \qquad 0 < \gamma < \sqrt{s},$$

by Proposition 3 and Proposition 5.

5.11. Proof of Theorem 6

The matter of existence and uniqueness of solutions of (5) is settled in a similar way as was done for (14); see Subsection 5.3.

By [1, Item 7.2.14, p. 300], case $n = 0$, we have the asymptotic expansion

$$\begin{aligned} \int_{-\infty}^{\delta} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta' &= \sqrt{\frac{\pi}{2}}\mathrm{erfc}\left(-\frac{\delta}{\sqrt{2}}\right) \\ &\sim -\mathrm{e}^{-\delta^2/2}\sum_{m=0}^{\infty} \frac{(-1)^m (2m)!}{m!\,2^m \delta^{2m+1}} \\ &= -\mathrm{e}^{-\delta^2/2}\left(\frac{1}{\delta} - \frac{1}{\delta^3} + \frac{3}{\delta^5} - \frac{15}{\delta^7} + \frac{105}{\delta^9} + O\left(\frac{1}{\delta^{11}}\right)\right) \quad \text{as } \delta \to -\infty. \end{aligned} \tag{58}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

Then (20) follows from (58) and some additional computations; cf. end of Subsection 5.1. Next, (20) is used for the proof of (11) in a similar way as (23) was used to prove (7) in Subsection 5.5 (this requires including an additional term $-945/\delta^{11}$ in the expansion in (58)).

Next, the inequalities in (12) follow from the inequality (22) in a similar way as (8) follows from (25) and (26); see Subsection 5.6.

Finally, we have, from $a = a_\infty(\gamma) < 1/\gamma$,

$$a = \frac{\exp(-(\gamma-a)^2/2)}{\int_{-\infty}^{\gamma-a} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta'} < \frac{\exp(-\gamma^2/2+1)}{\int_{-\infty}^{0} \mathrm{e}^{-(\delta')^2/2}\,\mathrm{d}\delta'}, \qquad \gamma \geq 1,$$

and this shows (13).

5.12. Proof of Theorem 1

We will show that $a_\infty(\gamma) - a_s(\gamma) = O(1/\sqrt{s})$ uniformly in $\gamma \in [\gamma_0, \gamma_1]$ when $0 < \gamma_0 < \gamma_1 < \infty$. We start by noting that $0 < a_s(\gamma) < a_\infty(\gamma)$ when $0 < \gamma < \sqrt{s}$. Indeed, $a_s(\gamma) > 0$ when $0 < \gamma < \sqrt{s}$. Moreover, $f_s(\gamma) < f_\infty(\gamma)$ by Propositions 5 and 6, and so $a_\infty(\gamma) - f_s(\gamma - a_\infty(\gamma)) > a_\infty(\gamma) - f_\infty(\gamma - a_\infty(\gamma)) = 0$. Therefore, see the end of Subsection 5.3, we have $a_\infty(\gamma) > a_s(\gamma)$.

Take $s \geq \gamma_1^2$, and let

$$\delta_0 = \gamma_0 - \alpha_\infty(\gamma_0), \qquad \delta_1 = \gamma_1 - a_s(\gamma_1).$$

Since $\gamma - a_s(\gamma)$ and $\gamma - a_\infty(\gamma)$ are increasing in γ by Theorem 2 and Theorem 6, we have $-\infty < \delta_0 < \delta_1 < \gamma$. By Propositions 5 and 6, there exists a $K > 0$ such that

$$0 < f_\infty(\delta) - f_s(\delta) < \frac{K}{\sqrt{s}}, \qquad s \geq \gamma_1^2,\ \delta \in [\delta_0, \delta_1]. \tag{59}$$

Also, for any $\delta \in [\delta_0, \delta_1]$, by Proposition 3 and (21),

$$f_s'(\delta) \geq f_s'(\delta_0) > -1, \qquad f_s'(\delta_0) \to f_\infty'(\delta_0) > -1 \quad \text{as } s \to \infty. \tag{60}$$

It follows from (60) that there exists a $\varepsilon > 0$ such that

$$f_s'(\delta) \geq -(1-\varepsilon), \qquad s \geq \gamma_1^2,\ \delta \in [\delta_0, \delta_1]. \tag{61}$$

Now take any $\gamma \in [\gamma_0, \gamma_1]$, and let

$$h_s(a) := a - f_s(\gamma - a), \qquad h_\infty(a) := a - f_\infty(\gamma - a), \qquad \gamma - \delta_1 \leq a \leq \gamma - \delta_0.$$

Then $a_s(\gamma), a_\infty(\gamma) \in [\gamma - \delta_1, \gamma - \delta_0]$, and

$$h_s(a_s(\gamma)) = 0 = h_\infty(a_\infty(\gamma)), \tag{62}$$

while

$$0 < h_s(a) - h_\infty(a) < \frac{K}{\sqrt{s}}, \qquad \gamma - \delta_1 \leq a \leq \gamma - \delta_0 \tag{63}$$

by (59). Furthermore, by (61),

$$h_s'(a) \geq \varepsilon, \qquad h_\infty'(a) \geq \varepsilon, \qquad \gamma - \delta_1 \leq a \leq \gamma - \delta_0. \tag{64}$$

https://doi.org/10.1239/aap/1363354111 Published online by Cambridge University Press

By the mean value theorem, there exists a $b \in [a_s(\gamma), a_\infty(\gamma)]$ such that

$$h_s(a_\infty(\gamma)) - h_s(a_s(\gamma)) = (a_\infty(\gamma) - a_s(\gamma))h_s'(b).$$

Since

$$h_s(a_\infty(\gamma)) - h_s(a_s(\gamma)) = h_s(a_\infty(\gamma)) - h_\infty(a_\infty(\gamma))$$

by (62), it follows from (63) with $a = a_\infty(\gamma)$ that

$$(a_\infty(\gamma) - a_s(\gamma))\, h_s'(b) < \frac{K}{\sqrt{s}}.$$

Then, finally, from $a_s(\gamma) < a_\infty(\gamma)$ and (64) with $a = b$, we obtain

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

as required.

Acknowledgements

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

References

[1] Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* US Government Printing Office, Washington, DC.
[2] Aguire, M. S., Aksin, O. Z., Karaesmen, F. and Dallery, Y. (2008). On the interaction of retrials and sizing of call centers. *Europ. J. Operat. Res.* **191,** 398–408.
[3] Artalejo, J. R. and Gómez-Corral, A. (2008). *Retrial Queueing Systems*. Springer, Berlin.
[4] Borst, S., Mandelbaum, A. and Reiman, M. I. (2004). Dimensioning large call centers. *Operat. Res.* **52,** 17–34.
[5] Cohen, J. W. (1957). Basic problems of telephone traffic theory and the influence of repeated calls. *Philips Telecommun. Rev.* **18,** 49–100.
[6] Falin, G. I. and Templeton, J. G. C. (1997). *Retrial Queues*. Chapman & Hall, London.
[7] Halfin, S. and Whitt, W. (1981). Heavy-traffic limits for queues with many exponential servers. *Operat. Res.* **29,** 567–588.
[8] Jagers, A. A. and van Doorn, E. A. (1986). On the continued Erlang loss function. *Operat. Res. Lett.* **5,** 43–46.
[9] Janssen, A. J. E. M., van Leeuwaarden, J. S. H. and Zwart, B. (2008). Gaussian expansions and bounds for the Poisson distribution applied to the Erlang B formula. *Adv. Appl. Prob.* **40,** 122–143.
[10] Janssen, A. J. E. M., van Leeuwaarden, J. S. H. and Zwart, B. (2011). Refining square-root safety staffing by expanding Erlang C. *Operat. Res.* **59,** 1512–1522.
[11] Sampford, M. R. (1953). Some inequalities on Mill's ratio and related functions. *Ann. Math. Statist.* **24,** 130–132.
[12] Zhang, B., van Leeuwaarden, J. S. H. and Zwart, B. (2013). Refined square-root staffing for call centers with impatient customers. To appear in *Operat. Res.*