IEOR 6707: ADVANCED TOPICS IN QUEUEING THEORY: FOCUS ON CUSTOMER CONTACT CENTERS

Fall 2002, Ward Whitt

HOMEWORK 1e. Solutions for the Erlang B and C Formulas

The Erlang B and C formulas are true probability classics. Indeed, much of the theory was developed by A. K. Erlang and his colleagues prior to 1925; see Brockmeyer, Halstrom and Jensen (1948). The subject has been extensively studied and applied by telecommunications engineers and mathematicians ever since; e.g., see Syski (1960), Riordan (1962) and Kosten (1973). A substantial record is available in the Proceedings of the International Teletraffic Congress, a sample being Labetoulle and Roberts (1994). A nice introductory account, including some of the telecommunications subtleties (e.g., the equivalent random method), is provided by Cooper (1981), who wrote his book shortly after spending several years at Bell Laboratories.

Please let me know about possible errors I have made. Suggestions (e.g., related problems of interest) are welcome.

1. The Erlang B Formula (due on September 16)

The Erlang B (or loss) formula gives the (steady-state) blocking probability in the Erlang loss model, i.e., in the $M/M/s/0$ model. This model has s homogenous servers working in parallel and no extra waiting space. Customers arriving when all s servers are busy are blocked (lost) without affecting future arrivals; e.g., there are no customer retrials. This model has a Poisson arrival process and IID (independent and identically distributed) service times (which are also independent of the arrival process) with an exponential distribution having finite mean. (The two M's in $M/M/s/0$ are for Markov, referring to the "lack-of-memory" property of the exponential distribution. Both the interarrival times and the service times have exponential distributions.)

Following convention, let the arrival rate be denoted by λ and let the mean service time be denoted by $1/\mu$. Thus, the (individual) service rate is μ. Since at most s customers can be in the system at any time, the stochastic process representing the number of busy servers as a function of time has a proper steady-state distribution for all (positive) values of the parameters λ and μ. The Erlang loss model has an *insensitvity property* implying that the blocking probability is independent of the service-time distribution beyond its mean. Thus the blocking probability is the same in an $M/GI/s/0$ model ("GI" for IID service times with a general distribution) as it is in the $M/M/s/0$ model, providing that the service-time distribution has a finite mean. This attractive insensitivity property is lost if the arrival process is extended to either nonstationary Poisson ($M_t/GI/s/0$) or stationary renewal with nonexponential interarrival times ($GI/GI/s/0$); e.g., see Davis et al. (1995). Here we only consider the $M/GI/s/0$ model.

The steady-state distribution of the number of busy servers also does not depend on the units we use to measure time. Thus the blocking probability depends on the arrival rate λ and the service rate μ only through their ratio, the *offered load*,

$$a \equiv \lambda/\mu \ . \tag{1.1}$$

(Here $\equiv$ denotes equality by definition.) Closely related to the offered load is the offered load per server, called the *server utilization* or *traffic intensity*,

$$\rho \equiv \lambda/s\mu = a/s \ . \tag{1.2}$$

Both the offered load a and the traffic intensity ρ are dimensionless quantities.

Let N denote the steady-state number of busy servers at an arbitrary time. We obtain the distribution of N from basic birth-and-death-process theory; e.g., see Cooper (1981). It turns out that N has a truncated Poisson distribution; i.e.,

$$P(N=j) = \frac{a^j/j!}{\sum_{k=0}^{k=s} a^k/k!}, \quad 0 \le j \le s \ ; \tag{1.3}$$

see pp. 4 and 80 of Cooper (1981). For background on the basic Markov-process theory, see Chapter 8 of Cinlar (1975).

As indicated above, the Erlang B formula gives the steady-state blocking probability of a typical arrival. We obtain the Erlang B formula from the probability $P(N=s)$ by applying the PASTA (Poisson Arrivals See Time Averages) property; see p. 77 of Cooper (1981), Wolff(1982, 1989), Melamed and Whitt (1990) and El-Taha and Stidham (1999).

The Erlang B formula is

$$B \equiv B(s,a) = P(N=s) = \frac{a^s/s!}{\sum_{k=0}^{k=s} a^k/k!} \ ; \tag{1.4}$$

e.g., see pp. 5 and 79 of Cooper (1981).

An important related quantity (because it tends to be easier to analyze) is the *reciprocal*,

$$R \equiv R(s,a) \equiv \frac{1}{B(s,a)} \ . \tag{1.5}$$

Here are some exercises on the Erlang B formula:

1. **recursion for the reciprocal**

   Show that the reciprocal satisfies the recursion

   $$R(s,a) = \frac{sR(s-1,a)}{a} + 1 \ , \quad s \ge 1, \tag{1.6}$$

   where $R(0,a) \equiv 1$.

   Answer: Let

   $$S(s) \equiv \sum_{k=0}^{k=s} a^k/k!$$

   Then

   $$R(s) = \frac{S(s)}{a^s/s!} = \frac{S(s-1) + a^s/s!}{a^s/s!} = \frac{sR(s-1,a)}{a} + 1 \ .$$

2. **recursion for B**

   Apply Problem 1 to establish the recursion

$$B(s,a) = \frac{aB(s-1,a)}{s + aB(s-1,a)} \; , \qquad (1.7)$$

   where $B(0,a) \equiv 1$ and a is the offered load defined in (1.1), which is understood to be fixed. The recursion can be written as

$$B(s,a) = \frac{\rho B(s-1,a)}{1 + \rho B(s-1,a)} \; , \qquad (1.8)$$

   provided that we understand $\rho = a/s$ on the righthand side.

   Answer: Combine (1.5) and (1.6).

3. **computation**

   Why is the recursion in (1.7) appealing for computing the blocking probability for very large s compared to the explicit formula in (1.4)? Write a computer program to compute $B(s,a)$ based on (1.7). Compute $B(10^k, 10^k)$ for $k =$ 1, 2, 3 and 4. Please turn in a copy of your code as well as your numerical results.

   Answer: By Problem 6 below, $B(s,a)$ is decreasing in s. We start with $B(0,a) \equiv 1$ and $B(1,a) = a/(1+a)$. Then $B(s,a)$ decreases toward 0 as s increases for fixed a. Thus the recursion leads to a stable algorithm. In contrast, for large s, direct calculation from the explicit formula in (1.4) involves very large powers and factorials. For specific numerical values, $B(10,10) \approx 0.215$.

   By scaling, $\sqrt{a}B(a + c\sqrt{a}, a)$ approaches a proper limit as $a \to \infty$; see Challenge Problem 29 below. (The limit of $\sqrt{a}B(a + c\sqrt{a})$ as $a \to \infty$ is the same as the limit of $\sqrt{s}B(s, s - c\sqrt{s})$ as $s \to \infty$.) Hence $\sqrt{a}B(a,a)$ should approach a limit as $a \to \infty$. Hence, we should have

$$B(10^6, 10^6) \approx 0.1 B(10^4, 10^4) \approx 0.01 B(10^2, 10^2) \; .$$

4. **monotonicity in a**

   Use Problem 2 to show that $B(s,a)$ is increasing in a.

   Answer: Use mathematical induction on s and the recursion (1.7).

5. **lower bound**

   Use Little's law ($L = \lambda W$) to establish the lower bound

$$B(s,a) \ge \max\{0, 1 - \rho^{-1}\} \; , \qquad (1.9)$$

   where $\rho \equiv a/s$.

   Answer: Let the system be the set of servers. The arrival rate, excluding blocked customers, is $\lambda(1 - B(s,a))$ and the expected waiting time in this instance is just the mean service time $1/\mu$. Hence, by Little's law ,

$$EN = \lambda(1 - B(s,a)/\mu = a(1 - B(s,a) \; .$$

   The lower bound follows from the trivial inequality $EN \le s$. This inequality is established by Sobel (1980).

6. **monotonicity in s**

   Use Problems 2 and 5 to show that $B(s,a)$ is decreasing in s as s runs through the positive integers.

   Answer: From (1.7),

   $$B(s-1,a) - B(s,a) = B(s-1,a)[1 - a/(s + aB(s-1,a)] > 0$$

   because, by (1.9),

   $$B(s-1,a) \geq max\{0, 1-(s-1)/a\} \geq max\{0, 1-s/a\} \ ,$$

   which implies that the term in brackets is nonnegative.

7. **partial derivative with respect to a**

   Show that the partial derivative of $B(s,a)$ with repect to a can be expressed directly in terms of ρ and $B(s,a)$ by

   $$\frac{\partial B(s,a)}{\partial a} = B(s,a)[\rho^{-1} - 1 + B(s,a)] \ . \tag{1.10}$$

   Answer: Use the derivative-of-a-ratio formula

   $$\frac{d\,(U(a)/V(a))}{da} = \frac{VU' - UV'}{V^2}$$

   with $V = S(s)$ defined in Problem 1 above. Get

   $$\begin{aligned} B' &= \frac{S(s,a)sa^{s-1}/s! - (a^s/s!)S(s-1,a)}{S(s,a)^2} \\ &= (s/a)B(s,a) - B(s,a)S(s-1,a)/S(s,a) \\ &= (s/a)B(s,a)[1 - B(s,a)/B(s-1,a)] \ , \end{aligned} \tag{1.11}$$

   which equals the displayed formula by virtue of (1.7).

8. **retake on monotonicity**

   Use Problems 7 and 5 to provide an alternative proof of the monotonicity result in Problem 4.

   Answer: The derivative in (1.10) is nonnegative because the lower bound in (1.9) implies that the term in brackets is nonnegative. (Actually all terms are strictly positive.)

9. **marginal analysis**

   The *load carried by the last server* is defined as

   $$F_B(s,a) \equiv a[B(s-1,a) - B(s,a)] \ . \tag{1.12}$$

   By Challenge Problem 17 below, F_B is decreasing in s. Suppose that there is a cost rate of \$$c$ per minute for each server used and a revenue rate of \$$r$ per minute for each unit of carried load. How can you use the load carried by the last server, F_B to determine the optimal number of servers to maximize profit (revenue - cost)?

   Answer: To find the optimal s, it suffices to find the largest value of s such that $F_B(s,a) \geq c/r$. Let s^* be that value of s. Then the revenue added by server s^* exceeds the cost of adding that server, whereas that property fails for $s^* + 1$.

Historically, this marginal economic analysis is the essential feature of *Moe's Principle*, going back to the time of Erlang (1910-1930); see Jensen (1950). Indeed, both Moe and the marginal analysis are discussed by Erlang (1924). The theoretical justification for this marginal analysis is more recent, however; see Problem 6 in the second part on Erlang C and Challenge Problems 17, 18 and 24.

10. **large systems**

Consider B as a function of ρ, i.e., $B(s, s\rho)$ (where $a = s\rho$) with s fixed. What should $B(s, 3s)$ be approximately when s is very large? Plot what $B(s, s\rho)$ should look like as a function of ρ for very large s (as $s \to \infty$).

Answer: When $\rho \le 1$, $B(s, s\rho) \to 0$ as $s \to \infty$. (The case $\rho = 1$ is more delicate.) When $\rho > 1$, a proportion $(s\rho - s)/s\rho$ of the input rate $s\rho$ should be blocked. That shows that the lower bound in Problem 5 is asymptotically correct as $s \to \infty$. Thus $B(s, 3s)$ should be approximately $2/3$ for large s.

Can $B(s, s\rho)$ be concave or convex as a function of ρ for all fixed s? (And, thus, can $B(s, a)$ be concave or convex as a function of a for all fixed s?)

Answer: No. This is so because the limiting function (as $s \to \infty$) and its derivative are 0 for $\rho < 1$, whereas the derivative for $\rho > 1$ is $1/\rho^2$. Hence the left derivative at $\rho = 1$ is 0, while the right derivative at $\rho = 1$ is 1. The derivative is first constant, then jumps up, and then decreases. The limit is first constant, then increasing concave.

11. **integral representation of the reciprocal**

Show that the reciprocal defined in (1.5) can be represented as

$$R(s, a) = \sum_0^s \frac{s^{(k)}}{a^k} \; , \tag{1.13}$$

where

$$s^{(k)} \equiv s(s-1)\cdots(s-k+1), \quad k \ge 1 \; , \tag{1.14}$$

and $s^{(0)} \equiv 1$. Use the form of the probability density function for an Erlang-distributed random variable (p. 11 of Feller (1971) or p. 65 of Cooper (1981)) or the Euler integral in the representation of the gamma function to show that

$$\frac{s^{(k)}}{a^k} = a \binom{s}{k} \int_0^\infty t^k e^{-at} \, dt \; . \tag{1.15}$$

Then apply (1.13)–(1.15) to obtain the integral representation

$$R(s, a) = a \int_0^\infty (1+t)^s e^{-at} \, dt \; . \tag{1.16}$$

Answer: Equation (1.13) is obtained dividing each term in the numerator $S(s, a)$ by $a^s/s!$. The Erlang (special case of gamma) probability density function (pdf) is

$$g_k(t) \equiv \frac{a(at)^k}{k!} e^{-at}, \quad t \ge 0 \; .$$

The integral of the Erlang pdf is 1. Formula (1.15) is a minor modification of that integral. Thus, we have

$$R(s, a) = \sum_{k=0}^{k=s} a \binom{s}{k} \int_0^\infty t^k e^{-at} \, dt = a \int_0^\infty (1+t)^s e^{-at} \, dt \; ,$$

after interchanging the order of summation and integration, which is justified because everything is nonnegative; apply Tonelli's theorem (the nonnegative version of Fubini's theorem); e.g., p. 270 of Royden (1968).

The reciprocal $R(s,a)$ and its extension to a function of complex variables are exploited extensively by Jagerman (1974).

12. **extension to a function of complex variables**

Use the integral representation in (1.16) to determine (or guess) the extension of the functions $R(s,a)$ and $B(s,a)$ to functions of the complex variables z and α (z for s and α for a) with $Re(\alpha) > 0$, where $Re(\alpha)$ is the real part of the complex number α.

As a special case, this construction extends $B(s,a)$ for positive integer s to positive real s, which is useful for approximations and further analysis.

Answer: With the integral representation in (1.16), it simply suffices to substitute the complex variable z for the positive integer s and the complex variable α with $Re(\alpha) > 0$ for the positive real number a. Then we make the definition $B(z,\alpha) \equiv 1/R(z,\alpha)$. This step allows us to apply established results for special functions; e.g., the asymptotic expansion in Challenge Problem 20 follows because the generalized reciprocal $R(z,\alpha)$ can be expressed directly in terms of the Whittaker functions; see p. 534 of Jagerman (1974). This approach brings the subject into the domain of classical applied mathematics.

2. The Erlang C Formula (due on September 23)

The Erlang C (or delay) formula gives the (steady-state) probability of delay (that an arrival must wait before beginning service) in the Erlang delay model, i.e., in the $M/M/s \equiv M/M/s/\infty$ model. Just like the Erlang loss model, the Erlang delay model has a Poisson arrival process with arrival rate λ, IID service times (which are also independent of the arrival process) with an exponential distribution having finite mean $1/\mu$ and s homogenous servers working in parallel. The difference is that the Erlang delay model has extra waiting space for customers to wait when all servers are busy. Indeed, in the Erlang delay model there is unlimited waiting room, so that no customers are lost.

Since there is no limit to the number of customers that can be in the system at any time, the stochastic process representing the number of customers in the system at time t need not have a proper steady-state distribution for all (positive) values of the parameters λ and μ. Indeed, with probability one, the number in system grows without bound if $\lambda > \mu$. The number of customers in the system has a proper steady-state distribution if and only if the traffic intensity satisfies the stability condition

$$\rho < 1 \; . \tag{2.1}$$

For our discussion of the Erlang C formula, we always assume that stability condition (2.1) holds.

Unlike the Erlang loss model, *the Erlang delay model does not have an insensitvity property*, so that the steady-state distribution for the number of customers in the system depends upon the service-time distribution beyond its mean. While there is quite a bit of theory for the more general $M/GI/s/\infty$ model, it remains difficult to analyze. Our discussion of the Erlang C formula thus applies directly only to the $M/M/s/\infty$ model with exponential service times. However, the Erlang C formula plays an important role in approximations for more general systems.

Just as for the Erlang loss model, the steady-state distribution of the number of customers in the system does not depend on the units we use to measure time. Thus the delay probability (the Erlang C formula) depends on the arrival rate λ and the (individual) service rate μ only through their ratio, the offered load a, or equivalently through the traffic intensity ρ.

As indicated above, the Erlang C formula gives the steady-state delay probability of a typical arrival. We obtain the steady-state distribution for the number of customers in the system at an arbitrary time from basic birth-and-death-process theory. We then obtain the steady-state delay probability for an arbitrary arrival from the PASTA property.

The Erlang C formula is

$$C \equiv C(s,a) = \frac{\frac{a^s}{s!(1-\rho)}}{\sum_{k=0}^{k=s-1} a^k/k! + \frac{a^s}{s!(1-\rho)}} \; ; \tag{2.2}$$

e.g., see pp. 7 and 91 of Cooper (1981).

To appreciate the overall unity of the theory, it is important to recognize that the Erlang C formula is intimately related to the Erlang B formula. Indeed, as we see below, we can express the Erlang C formula in terms of the Erlang B formula, and vice versa provided that $0 < a < s$. Since we may well have $a \geq s$ in loss systems, the Erlang B formula should be considered more fundamental. Even though properties of one can be deduced from the other, the two formulas have somewhat different structure.

Moreover, many other steady-state performance measures of interest for the $M/M/s/\infty$ model can be defined in terms of the delay probability C. Let N and Q be the steady-state numbers of customers in the system and in the queue, respectively. Let W and S be the steady-state waiting time (until beginning service) and sojourn time (waiting time plus service time), respectively. By Little's law,

$$EQ = \lambda EW \quad \text{and} \quad EN = \lambda ES \; . \tag{2.3}$$

The key links to the Erlang C formula are

$$EN = a + C(s,a)\rho/(1-\rho) \tag{2.4}$$

and

$$EQ = C(s,a)\rho/(1-\rho) \; . \tag{2.5}$$

Here are some exercises on the Erlang C formula:

1. **relating $C(s,a)$ to $B(s-1,a)$**

   Show that

   $$C(s,a) = \frac{\frac{\rho B(s-1,a)}{(1-\rho)}}{1 + \frac{\rho B(s-1,a)}{(1-\rho)}} \; , \tag{2.6}$$

   where ρ on the righthand side is a/s.

   Answer: Divide by $a^{s-1}/(s-1)!$ in (2.2) to get

   $$C(s,a) = \frac{\rho/(1-\rho)}{(1/B(s-1,a)) + \rho/(1-\rho)} \; ,$$

   from which (2.6) follows easily.

2. **monotonicity in a and s**

   Apply (2.6) and the established monotonicity for $B(s,a)$ in a in Problem 4 of the first part of these exercises to show that $C(s,a)$ is increasing in a.

   Apply (2.6) and the established monotonicity of $B(s,a)$ in s in Problem 6 of the first part to show that $C(s,a)$ is decreasing in s.

   Answer: immediate

3. **relating C to B**

   Apply equations (2.6) and (1.7) to deduce that

$$C(s,a) = \frac{B(s,a)}{1-\rho[1-B(s,a)]}, \quad 0 < a < s \ . \tag{2.7}$$

   As a consequence,

$$B(s,a) = \frac{(1-\rho)C(s,a)}{1-\rho C(s,a)}, \quad 0 < a < s \ . \tag{2.8}$$

   Answer: elementary

4. **comparing C and B**

   Apply (2.7) to deduce that $C(s,a) > B(s,a)$ for all $s \geq 1$ and all a, $0 < a < s$.

   Answer: elementary

5. **relating C to R**

   Apply (1.6) and (2.7) to show that

$$C(s,a) = \frac{1}{R(s,a) - R(s-1,a)} \ .$$

   Answer: By (1.6),

$$\frac{\rho}{B(s,a)} = \frac{1}{B(s-1,a)} + \rho \ .$$

   By (2.7),

$$C(s,a) = \frac{1}{\frac{1-\rho}{B(s,a)} + \rho} = \frac{1}{\frac{1-\rho}{B(s,a)} + \frac{\rho}{B(s,a)} - \frac{1}{B(s-1,a)}} \ ,$$

   which equals the claimed formula.

6. **more marginal analysis**

   Paralleling Problem 9 in the first part, find an expression for the reduction in expected delay provided by the last server (s servers as opposed to $s-1$). Suppose that there is a cost rate of $\$c$ per minute for each server and a cost rate $\$d$ per minute for each customer delayed. How can you determine the optimal number of servers to minimize the long-run average cost?

   Answer: Let $EW(s,a)$ be the expected steady-state delay as a function of s and a. Then the reduction in expected delay provided by server s is

$$F_D(s,a) = \lambda[EW(s-1,a) - EW(s,a)] \ . \tag{2.9}$$

Using the previous problem,

$$\begin{aligned} F_D(s,a) &= \frac{aC(s-1,a)}{s-1-a} - \frac{aC(s,a)}{s-a} \\ &= \frac{a}{s-1-a}\frac{1}{R(s-1,a)-R(s-2,a)} - \frac{a}{s-a}\frac{1}{R(s,a)-R(s-1,a)} \end{aligned} \qquad (2.10)$$

Using the property that $F_D(s,a)$ is decreasing in s over the positive integers such that $a < s$, choose the largest value of s such that $F_D(s,a) \geq c/d$.

This second form of marginal analysis is also discussed by Erlang (1924). Theoretical support has been provided by Fox (1966), Rolfe (1971), Dyer and Proll (1977), Weber (1980) and Jagers and van Doorn (1991).

7. **allocation of servers to multiple facilities**

   Suppose that there are m $M/M/s/\infty$ service facilities with specified arrival rates and service rates, but yet-to-be-specified numbers of servers. Suppose that the total number of servers for all m facilities has been specified as n. Develop an efficient algorithm to allocate the n servers to the m service facilities in order to minimize the average delay among all arriving customers.

   Answer: First the number of servers assigned to facility i must be such that the traffic intensity ρ_i is less than one. That gives an initial number $n_{i,0}$ of servers for each facility i. We require that $n \geq \sum_{i=1}^{i=m} n_{i,0}$. We now can allocate the remaining available servers to the facilities one at a time. We allocate the next server to the facility which provides the greatest improvement in weighted expected delay. The expected delay at facility i should be weighted by $\lambda_i / \sum_{j=1}^{j=m} \lambda_j$.

   The theoretical basis is demonstrated by Fox (1966), assuming discrete convexity of the component weighted expected delays. The discrete convexity is demonstrated for the $M/D/s$ model by Rolphe (1971), for the $M/M/s$ model by Dyer and Proll (1977) and for the $G/GI/s$ model by Weber (1980). An alternate proof for $M/M/s$ follows from Jagers and van Doorn (1991).

8. **convexity as a function of a**

   Consistent with intuition and numerical evidence, it turns out that $C(s,a)$ is a convex function of a (see Challenge Problem 13 below). Use this property with equations (2.3)–(2.5) to deduce that EN, EQ, EW, and ES are all convex functions of a.

   Answer: The product of two nondecreasing nonnegative convex functions is necessarily nondecreasing nonnegative and convex. To see that, differentiate twice: If f and g are two such functions, then

   $$(fg)'' = f''g + 2f'g' + fg'' \ ,$$

   where f, f', f'', g, g', g'' are all nonnegative.

9. **large systems**

   Paralleling Problem 10 in the first part on the Erlang B formula, consider C as a function of ρ, i.e., consider $C(s, s\rho)$ where ρ is constrained to satisfy $0 < \rho < 1$. What should $C(s, 0.95s)$ be for very large s (as $s \rightarrow \infty$)? Plot what $C(s,\rho)$ should look like as s gets large (as $s \rightarrow \infty$)?

   Answer: Just like B, $C(s, s\rho) \rightarrow 0$ as $s \rightarrow \infty$ for all ρ with $0 < \rho < 1$.

10. **recursion for C**

Establish the recursion

$$C(s,a) = \frac{1}{1 + \frac{(s-a)(s-1-aC(s-1,a))}{a(s-1-a)C(s-1,a)}}, \quad s \ge 1 \ , \tag{2.11}$$

where $C(0,a) \equiv 1$. Why is this direct recursion much less useful for computation than the recursion for B in Problem 2 of the first part?

Answer: Divide through by the numerator in (2.6) to obtain

$$C(s,a) = \frac{1}{1 + \frac{1-\rho}{\rho B(s-1,a)}} \ .$$

Now apply (2.8) (recalling that here s is $s-1$) to obtain the displayed formula.

This recursion is much less useful than the recursion for B in Problem 2 of the first part here because $C(s,a)$ is only defined for $a < s$. Hence, if we start the recursion with $C(1,a)$, then we are constrained to have $a < 1$. When s is large, we typically are interested in $a > 1$. Thus, for a of interest we have difficulty initializing the recursion. Hence this recursion for C is almost totally useless.

11. **computation**

Develop a computer program to compute $C(s,a)$. What is $C(10^2, 10^2 - 10^1)$? What is $C(10^4, 10^4 - 10^2$? What is $C(10^6, 10^6 - 10^3)$? Please turn in a copy of your code as well as the numerical results.

Comment As indicated in the previous problem, we do not want to use the recursion for C just developed. Instead, we can exploit the relation between C and B in (2.7), and use the recursion for B established in the first part.

By scaling, to be discussed later, the three numerical values $C(10^2, 10^2 - 10^1)$, $C(10^4, 10^4 - 10^2$ and $C(10^6, 10^6 - 10^3)$ should be very close. These are approximately 0.22.

12. **asymptotics**

It is known (by Challenge Problem 29 below) that $\sqrt{s}B(s, s - c\sqrt{s}) \to \phi(c)/\Phi^c(c)$ as $s \to \infty$. Use (2.7) to establish a corresponding limit for $C(s, s - c\sqrt{s})$ as $s \to \infty$ when $c > 0$. Show that $C(s, s - c\sqrt{s})$ has a limit as $s \to \infty$ without scaling (multiplying or dividing by $\sqrt{s}$).

Answer: Multiply on the right in (2.7) above and below by $\sqrt{s}$ and then let $s \to \infty$. Observe that $\sqrt{s}(1-\rho) \to c$ as $s \to \infty$. Thus,

$$C(s, s - c\sqrt{s}) \to \frac{\phi(c)/\Phi^c(c)}{c + (\phi(c)/\Phi^c(c))} = \frac{1}{1 + c\Phi^c(c)/\phi(c)}$$

as $s \to \infty$. This limit was established by different reasoning in Halfin and Whitt (1981). Halfin and Whitt also go further and establish a stochastic-process limit involving convergence of the scaled queue-length process to a diffusion process.

3. Challenge Problems (due on September 23 if at all)

Below are **optional** harder (sometimes much harder) problems (from research papers). You are not expected to do any of these problems, but if you want, go ahead and surprise me!

13. **convexity of C as a function of a**

    Show that $C(s,a)$ is a convex function of a for $0 < a < s$.

    Answer: This result surely must have a long history, but it is hard to trace. A proof appears in Lee and Cohen (1983). Direct proofs of convexity for the related performance measures EN and EQ appear in Grassman (1983) together in the same journal.

14. **convexity of B as a function of a^{-1}**

    Show that $B(s,a)$ is a convex function of a^{-1}, but not of a. From that immediately deduce that B is a convex function of the service rate μ for fixed s and λ, and a convex function of the mean interarrival time λ^{-1} for fixed μ and s.

    Answer: This is Proposition 3 of Harel (1990).

15. **concavity of the carried load as a function of a**

    The carried load in the Erlang loss model represents the time-average amount of work done by the system. Equivalently, it represents the portion of the offered load that is actually served. Equivalently, it is the expected number of servers that are busy at any time. Mathematically, the carried load is defined as

    $$a' \equiv a'(s,a) \equiv a[1 - B(s,a)] \; . \qquad (3.1)$$

    Show that the carried load is a concave function of a.

    Answer: This result surely has a long history too. This is Corollary 1 (a) on p. 504 of Harel (1990). Harel refers to an unpublished manuscript by K. R. Krishnan (1988) that also establishes this result. There is a large body of related work for this model and more general models. Much can be traced from Chapter 3 of Chen and Yao (2001) and Müller and Stoyan (2002).

16. **joint concavity of the throughput as a function of (λ, μ)**

    The throughput is defined as the rate at which customers complete service. Mathematically, the throughput for the Erlang loss model is defined as

    $$\theta \equiv \theta(s,a) \equiv \lambda[1 - B(s,a)] = \lambda[1 - B(s,\lambda/\mu)] \; . \qquad (3.2)$$

    Show that the throughput θ is jointly concave in (λ, μ).

    Answer: This is Corollary 1 (b) on p. 504 of Harel (1990).

17. **discrete convexity of B as a function of s**

    Show that $B(s-1,a) - B(s,a)$ is decreasing in s for positive integers s. This property is the basis for marginal analysis to determine the optimal number of servers, discussed in Problem 9 in the first part.

    Answer: This property is evidently first established by Messerli (1972). See the next problem for an extension to all real s. See Wolff and Wang (2002) for an extension to $G/GI/s/0$ models. (That extension parallels the results of Weber (1980) for the delay model.)

18. **convexity of $B(s,a)$ as a function of real s**

    Show that $B(s,a)$ extended to real positive s by equation (1.5) and Problem 12 is convex in s.

    Answer: This is Theorem 1 of Jagers and van Doorn (1986).

19. **partial derivative of $B(z,\alpha)$ with respect to α**

    Use the extension of B to a function of two complex variables to compute the partial derivative of $B(z,\alpha)$ with respect to α for $Re(\alpha) > 0$.

    Answer: Theorem 15 of Jagerman (1974) shows that the derivative with respect to α is the direct complex analog of equation (1.10), i. e.,

    $$\frac{\partial B(z,\alpha)}{\partial \alpha} = \left(\frac{z}{\alpha} - 1 + B(z,\alpha)\right) B(z,\alpha)$$

    for $Re(\alpha) > 0$.

20. **logconvexity of $R(x,a)$ as a function of x and a**

    Show that $R(x,a)$ is logconvex as a function of real positve x and a, where the case of non-integral x is obtained from the extension in Problem 12 in the first part.

    Answer: This is Theorem 19 on p. 545 of Jagerman (1974).

21. **subadditivity of the loss function**

    For the Erlang loss model, let the *loss rate* be defined by

    $$L \equiv L(s,a) \equiv aB(s,a) \ . \tag{3.3}$$

    Show that the loss rate is subadditive in (s,a) for integer values of s; i.e.,

    $$L(s_1 + s_2, a_1 + a_2) \leq L(s_1, a_1) + L(s_2, a_2) \tag{3.4}$$

    for all positve integers s_1 and s_2 and all positive real numbers a_1 and a_2. What does this say about the loss rate when two $M/M/s/0$ models are combined (the numbers of servers are added and the offered loads are added)?

    Answer: This is Theorem 1 of Smith and Whitt (1981), which is proved by stochastic-comparison methods. A short alternative proof based on the convexity established in Challenge Problem 18 above (due to Jagers and van Doorn (1986)) is given on p. 54 of Smith and Whitt (1981). This result shows that when two $M/M/s/0$ systems with common service-time distributions are combined, the overall average (per arrival) steady-state blocking probability decreases. Thus, this result provides a mathematical demonstration of the economy of scale.

22. **subadditivity of the expected waiting time**

    Let $EW(s,\lambda,\mu)$ be the expected waiting time for the Erlang delay model. Show that $\lambda EW(s,\lambda,\mu)$ is subadditive in (s,λ) for common fixed μ and integer values of s; i.e.,

    $$(\lambda_1 + \lambda_2) EW(s_1 + s_2, \lambda_1 + \lambda_2, \mu) \leq \lambda_1 EW(s_1, \lambda_1, \mu) + \lambda_2 EW(s_2, \lambda_2, \mu) \tag{3.5}$$

    for all positve integers s_1 and s_2 and all positive real numbers λ_1 and λ_2. What does this subadditivity property imply about the average expected waiting time when two

$M/M/s/\infty$ models with common service-time distributions are combined (the numbers of servers are added and the arrival rates are added)?

Answer: This is Theorem 2 of Smith and Whitt (1981). The subadditivity property shows that when two $M/M/s/\infty$ systems with common service rate are combined that the average (per-arrival) expected delay decreases. Thus the subadditivity is a mathematical demonstration of the economy of scale.

23. **convexity threshold for B**

Show that, for each $s \geq 2$, there exists a positive threshold $a^* \equiv a^*(s)$ such that $B(s,a)$ is a convex function of a for all $a < a^*$ and $B(s,a)$ is a concave function of a for all $a > a^*$. If possible, determine how $a^*(s)$ depends upon s.

Answer: This is Proposition 1 of Harel (1990). Let $\rho^* \equiv a^* s$. Harel shows that $\rho^* < 1$ for all $s \leq 18$ and that $1 < \rho^* < 1.5$ for all $s \geq 19$. He calculates ρ^* as a function of s for $1 \leq s \leq 270$. He shows that ρ^* first increases and then decreases in this range. It is not difficult to show that $\rho^* \rightarrow 1$ as $s \rightarrow \infty$.

24. **convexity of $C(s,a)$ as a function of s**

Develop an extension for $C(s,a)$ as a function of real positive s and a and show that function $C(s,a)$ is convex and decreasing as a function of s.

Answer: This is contained in Jagers and van Doorn (1991).

25. **joint convexity for the sojourn time**

As before, let S denote the steady-state sojourn time (waiting time plus service time) in the $M/M/s/\infty$ model. Show that the mean ES and the standard deviation $SD(S)$ are both jointly convex in the arrival rate λ and the service rate μ.

Answer: See Harel and Zipkin (1987). This is established by working with the reciprocal.

26. **bounds for B**

Show that

$$\begin{aligned}
B &< U1 < U3 \quad \text{for all} \quad s \geq 2 \quad \text{and} \quad \rho > 0, \\
B &< U1 < U2 < U3 \quad \text{for all} \quad s \geq 2 \quad \text{and} \quad \rho > 1, \\
B &> L0 \geq L4 \quad \text{for all} \quad s \geq 1 \quad \text{and} \quad 0 < \rho \leq 1, \\
B &> L1 \geq L4 \quad \text{for all} \quad s \geq 1 \quad \text{and} \quad 0 < \rho \leq 1, \\
B &> L1 > L2 > L3 > L4 \quad \text{for all} \quad s \geq 3 \quad \text{and} \quad \rho \geq 1,
\end{aligned} \tag{3.6}$$

where

$$\begin{aligned}
U1 &\equiv \frac{Num_1}{Den_1}, \quad \rho > 0, \\
U2 &\equiv \frac{s(1-\rho)^2 + 2\rho - 1}{-s(1-\rho) + 2\rho}, \quad \rho > 1, \\
U3 &\equiv \frac{\rho}{1+\rho}, \quad \rho > 0
\end{aligned} \tag{3.7}$$

with

$$\begin{aligned}
Num_1 &\equiv s(1-\rho)^2 + 2\rho - (1-\rho)\sqrt{4s\rho + s^2(1-\rho)^2} \\
Den_1 &\equiv -s\rho(1-\rho) + 2\rho + \rho\sqrt{4s\rho + s^2(1-\rho)^2}
\end{aligned} \tag{3.8}$$

and

$$\begin{aligned}
L0 &\equiv [N_{L0}/D_L0]^+, \quad 0 < \rho \leq 1, \\
L1 &\equiv [N_{L1}/D_{L1}]^+, \quad \rho > 0, \\
L2 &\equiv N_{L2}/D_{L2}, \quad \rho > 1, \\
L3 &\equiv \frac{s\rho(1-\rho)}{1-s\rho^2}, \quad \rho > 1, \\
L4 &\equiv max\{0, 1-(1/\rho)\}, \quad \rho > 0\ .
\end{aligned} \tag{3.9}$$

with

$$\begin{aligned}
N_{L0} &\equiv 1-(1-\rho)\sqrt{\pi s/2}, \\
D_{L0} &\equiv 1+\rho\sqrt{\pi s/2}, \\
N_{L1} &\equiv 2\rho + s(1-\rho)^2 - (1-\rho)\sqrt{4s\rho + 4(s+1) + s^2(1-\rho)^2}, \\
D_{L1} &\equiv 2\rho - s\rho(1-\rho) + \rho\sqrt{4s\rho + 4(s+1) + s^2(1-\rho)^2}, \\
N_{L2} &\equiv s(1-\rho)^3 + 2\rho(1-\rho) + (1-\rho), \\
D_{L2} &\equiv -s\rho(1-\rho)^2 - 2\rho^2\ .
\end{aligned} \tag{3.10}$$

Answer: See Harel (1988).

27. **Bounds for C**

Show that

$$\begin{aligned}
C &< UC < \rho \quad \text{for all} \quad s \geq 2 \quad \text{and} \quad 0 < \rho < 1, \\
C &> LC1 > LC2 \quad \text{for all} \quad s \geq 2 \quad \text{and} \quad 0 < \rho < 1, \\
\rho &> AC1 > LC1 \quad \text{for all} \quad s \geq 2 \quad \text{and} \quad 0 < \rho < 1, \\
UC &> AC2 > LC1 \quad \text{for all} \quad s \geq 2 \quad \text{and} \quad 0 < \rho < 1\ ,
\end{aligned} \tag{3.11}$$

where

$$\begin{aligned}
UC &\equiv 1 + \frac{s(1-\rho)^2}{2\rho} - \frac{(1-\rho)}{2\rho}\sqrt{4s\rho + s^2(1-\rho)^2)}, \\
LC1 &\equiv [1-(1-\rho)\sqrt{\pi s/2}]^+, \\
LC2 &\equiv [1-(1-\rho)s]^+, \\
AC1 &\equiv [1-(1-\rho)\sqrt{\pi s/4}]^+, \\
AC2 &\equiv [1 + \frac{s(1-\rho)^2}{2\rho} - \frac{(1-\rho)}{2\rho}\sqrt{5s\rho + s^2(1-\rho)^2}]^+\ .
\end{aligned} \tag{3.12}$$

Note that $C = \rho = UC = LC1 > LC2$ for $s = 1$.

Answer: See Harel (1988).

28. **asymptotic expansion for $R(z, z + c\sqrt{z})$ as $z \to \infty$**

For the Erlang loss model, show that the function $R(z, z + c\sqrt{z})$ for fixed real c and complex z has an asymptotic expansion as $z \to \infty$ (with $|arg(z)| < \pi/2$) of the form

$$R(z, z + c\sqrt{z}) \sim a_0(c)\sqrt{z} + a_1(c) + \frac{a_2(c)}{\sqrt{z}} + \frac{a_3(c)}{z} + \ldots \tag{3.13}$$

where

$$\begin{aligned}
a_0(c) &= \frac{\Phi^c(c)}{\phi(c)}, \\
a_1(c) &= \frac{2}{3} + \frac{1}{3}c^2 - \frac{1}{3}c^3 a_0(c) \\
a_2(c) &= -\frac{1}{18}c^5 - \frac{7}{36}c^3 + \frac{1}{12}c + \left(\frac{1}{18}c^6 + \frac{1}{4}c^4 + \frac{1}{12}\right) a_0(c) \ . \qquad (3.14)
\end{aligned}$$

Answer: This asymptotic expansion was established by Jagerman (1974), exploiting connections to the Whittaker functions; see Theorem 14 on p. 539.

29. **asymptotic expansion for $B(z, z + c\sqrt{z})$ as $z \to \infty$**

Apply the asymptotic expansion for $R(z, z + c\sqrt{z})$ as $z \to \infty$ to develop an associated asymptotic expansion for $B(z, z + c\sqrt{z})$ as $z \to \infty$.

Answer: Use the series expansion of $1/s_1$, where $s_1 = 1 + a_1 * z + a_2 * z^2 + \cdots$; see 3.6.16 on p. 15 of Abramowitz and Stegun (1972) to get

$$\begin{aligned}
B(z, z + c\sqrt{z}) \quad &\sim \quad \frac{1}{a_0(c)\sqrt{z}} - \frac{a_1(c)}{a_0(c)^2 z} + \left(\frac{a_1(c)^2}{a_0(c)^3} - \frac{a_2(c)}{a_0(c)^2}\right)\frac{1}{z^{3/2}} \\
&+ \left(\frac{2a_1(c)a_2(c)}{a_0(c)^3} - \frac{a_3(c)}{a_0(c)^3} - \frac{a_3(c)^3}{a_0(c)^4}\right)\frac{1}{z^2} \qquad (3.15)
\end{aligned}$$

as $z \to \infty$ (with $|arg(z)| < \pi/2$) for c real.

The first-order asymptotics will be discussed associated with scaling. Then we show the weaker result that

$$\sqrt{a}B(a - c\sqrt{a}, a) \to \phi(c)/\Phi^c(c)$$

as $a \to \infty$.

30. **asymptotic expansion for $C(s, s - c\sqrt{s})$ as $s \to \infty$**

Use (2.7) and the results above to show that $C(s, s - c\sqrt{s})$ has an asymptotic expansion for $c > 0$ and display the first two terms. Hint: Remember that ρ is a function of s.

Answer: First note that

$$\rho \equiv \rho(s) = 1 - \frac{c}{\sqrt{s}}$$

as $s \to \infty$. Then apply (2.7) to get

$$C(s, s - c\sqrt{s}) = \frac{B(s, s - c\sqrt{s})}{\frac{c}{\sqrt{s}} + (1 - \frac{c}{\sqrt{s}})B(s, s - c\sqrt{s})} \ . \qquad (3.16)$$

Now apply The asymptotic expansion for B plus relations for series, as on p. 15 of Abramowitz and Stegun (1972), to get an associated asymptotic expansion for C. It takes the form

$$C(s, s - c\sqrt{s}) \sim \frac{Num_C}{Den_C} \qquad \text{as} \quad s \to \infty \ , \qquad (3.17)$$

where

$$\begin{aligned}
Num_C &= \frac{A_0}{s^{1/2}} + \frac{A_1}{s} + \frac{A_2}{s^{3/2}} + \ldots \\
Den_C &= \frac{c}{\sqrt{s}} + \left(\frac{A_0}{\sqrt{s}} + \frac{A_1 - cA_0}{s} + \frac{A_2 - cA_1}{s^{3/2}} + \ldots\right), \qquad (3.18)
\end{aligned}$$

from which we deduce that

$$C(s, s - c\sqrt{s}) \sim \left(\frac{1}{1 + \frac{c}{A_0}}\right)\left(1 + \left(\frac{A_1}{A_0} - \frac{A_2 - cA_0}{A_0 + c}\right)\frac{1}{\sqrt{s}}\right) \ . \tag{3.19}$$

Finally, substituting the known expressions for A_0, A_1, etc., we obtain

$$C(s, s - c\sqrt{s}) \sim c_0 + \frac{c_1}{\sqrt{s}} + \frac{c_2}{s} + \cdots \ , \tag{3.20}$$

where

$$c_0 = \frac{1}{1 + ca_0(c)} = [1 + (c\Phi^c(c)/\phi(c))]^{-1} \tag{3.21}$$

and

$$c_1 = c_0\left(\frac{a_1(c)}{a_0(c)} - \frac{1}{a_0(c)^2}\left(\frac{a_1(c)^2 - a_0(c)a_2(c) - a_0(c)^2c}{1 + a_0(c)c}\right)\right) \ , \tag{3.22}$$

where the coefficients $a_i(c)$ are given in (3.14).

31. **limit for improvement in loss provided by the last server**

Show that the improvement provided by server s, $F_B(s, a)$ defined in (1.12) has a limit as $a \to \infty$ when we let $s = a + c\sqrt{a}$.

Answer: Interestingly, this limit is given by Erlang (1924). The limit is

$$\lim_{a \to \infty} F_B(a + c\sqrt{a}, a) = \frac{c\phi(c)}{\Phi^c(c)} + \frac{\phi(c)^2}{\Phi^c(c)^2}$$

The shift of index from s to $s - 1$ makes the limit involve the derivative of the limit function $\phi(c)/\Phi^c$. Use the fact that $\phi(c)' = -c\phi(c)$.

32. **limit for the improvement in delay provided by the last server**

Show that the improvement in delay provided by server s, $F_C(s, a)$ established in Problem 6 of the second part, has a limit as $a \to \infty$ when we let $s = a + c\sqrt{a}$ for $c > 0$.

Answer: This limiting form is also given by Erlang (1924). It involves the first and second derivatives of the limit function $\Phi^c(c)/\phi(c)$ for $R(a + c\sqrt{a}, a)$.

References

Abramowitz, M. and Stegun, I. A. (1972) *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D. C.

Borst, S. C., Mandelbaum, A. and Reiman, M. I. 2000. Dimensioning large call centers. Working paper.

Brockmeyer, E., Halstrom, H. L. and Jensen, A. 1948. *The Life and Works of A. K. Erlang*, Danish Academy of Technical Sciences, Copenhagen. (second ed., *Acta Polytechnica Scandinavica*, AP287, 1960)

Chen, H. and Yao, D. D. 2001. *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization*, Springer, New York.

Cinlar, E. 1975. *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ.

Cooper, R. B. 1981. *Introduction to Queueing Theory*, second edition, North Holland, New York.

Davis, J. L., Massey, W. A. and Whitt, W. 1995. Sensitivity to the service-time distribution in the nonstationary Erlang loss model. *Management Science* 41, 1107-1116.

Dyer, M. E. and Proll, L. G. 1977. On the validity of marginal analysis for allocating servers in $M/M/c$ queues. *Management Science* 23, 1019-1022.

El-Taha, M. and Stidham, Jr., S. 1999. *Sample-Path Analysis of Queueing Systems*, Kluwer, Boston.

Erlang, A. K. 1924. On the rational determination of the number of circuits, In *The Life and Works of A. K. Erlang*, E. Brockmeyer, H. L. Halstrom and A. Jensen (eds.), Danish Academy of Technical Sciences, 1948, 216-221.

Feller, W. 1971. *An Introduction to Probability Theory and its Applications*, volume II, second edition, Wiley, New York.

Fox, B. L. 1966. Discrete optimization via marginal analysis. *Management Science* 13 (3), 210-216.

Grassman, W. 1983. The convexity of the mean queue size of the $M/M/c$ queue with respect to the arrival rate. *J. Appl. Prob.* 20, 916-919.

Halfin, S. and Whitt, W. 1981. Heavy-traffic limits for queues with many exponential servers. *Operations research* 29, 567-588.

Harel, A. 1988. Sharp bounds and simple approximations for the Erlang delay and loss formulas. *Management Science* 34, 959-972.

Harel, A. 1990. Convexity properties of the Erlang loss formula. *Operations Research* 38, 499-505.

Harel, A. and Zipkin, P. H. 1987. Strong convexity results for queueing systems. *Operations Research* 35, 405–418.

Jagerman, D. L. 1974. Some properties of the Erlang loss function. *Bell System Technical Journal* 53, 525-551.

Jagers, A. A. and van Doorn, E. A. 1986. On the continued Erlang loss function. *Operations research Letters* 5, 43-46.

Jagers, A. A. and van Doorn, E. A. 1991. Convexity of functions which are generalizations of the Erlang los function and the Erlang delay function. *SIAM Review* 33, 281-282.

Jensen, A. 1950. *Moe's Principle*, The Copenhagen Telephone Company (K. T. A. S.), Copenhagen.

Kosten, L. 1973. *Stochastic Theory of Service Systems*, Pergamon Press, New York.

Krishnan, K. R. 1988. The convexity of loss rate in an $M/M/n/n$ queue with repect to arrival and service rates. Bell Communications research, Morristown, NJ.

Labetoulle, J. and Roberts, J. W. 1994. *The Fundamental Role of Teletraffic in the Evolution of Telecommunication Networks: Proceedings of the* 14th *International Teletraffic Congress - ITC 14*, Elsevier, Amsterdam.

Lee, H. L. and Cohen, M. A. 1983. A note on the convexity of performance measures of $M/M/c$ queueing systems. *J. Appl. Prob.* 20, 920-923.

Melamed, B. and Whitt, W. 1990. On arrivals that see time averages. *Operations Research* 38, 156-172.

Messerli, E. J. 1972. Proof of a convexity property of the Erlang B formula. *Bell System Tech. J.* 51, 951-953.

Müller, A. and Stoyan, D. 2002. Comparison Methods for Stochastic Models and Risks, Wiley, New York.

Riordan, J. 1962. *Stochastic Service Systems*, Wiley, New York.

Rolfe, A. J. 1971. A note on marginal allocation in multiple-server service systems. *Management Science* 17 (9), 656-658.

Royden, H. L. 1968. *Real Analysis*, second ed., Macmillan, London.

Smith, D. R. and Whitt, W. 1981. Resource sharing for efficiency in traffic systems. *Bell System Tech. J.* 60, 39-55.

Sobel, M. J. 1980. Simple inequalities for multiserver queues. *Management Science* 26 (9) 951-956.

Syski, R. 1960. *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd, Edinburgh. (second ed., North-Holland Studies in Telecommunications 4, Elsevier, Amsterdam, 1986)

Weber, R. R. 1980. On the marginal benefit of adding servers to $G/GI/m$ queues. *Management Science* 26 (9), 946-951.

Wolff, R. W. 1982. Poisson Arrivals See Time Averages. *Operations Research* 30, 223-231.

Wolff, R. W. 1989. *Stochastic Modeling and the Theory of Queues*, Prentice-Hall, Englewood Cliffs, New Jersey.

Wolff, R. W. and Wang, C.-L. 2002. On the convexity of loss probabilities. *J. Appl. Prob.* 39, 402-406.