TWO-MOMENT APPROXIMATIONS FOR MAXIMA

by

Charles S. Crow, IV
IEOR Department
Columbia University
csc2103@columbia.edu

David Goldberg
CS Department
Columbia University
dag2020@columbia.edu

Ward Whitt
IEOR Department
Columbia University
ww2040@columbia.edu

June 22, 2005; Revision: July 3, 2006

Abstract

We introduce and investigate approximations for the probability distribution of the maximum of n i.i.d. nonnegative random variables, in terms of the number n and the first few moments of the underlying probability distribution, assuming the distribution is unbounded above but does not have a heavy tail. Since the mean of the underlying distribution can immediately be factored out, we focus on the effect of the squared coefficient of variation (SCV, c^2, variance divided by the square of the mean). Our starting point is the classical extreme-value theory for representative distributions with the given SCV - mixtures of exponentials for $c^2 \geq 1$, convolutions of exponentials for $c^2 \leq 1$ and gamma for all c^2. We develop approximations for the asymptotic parameters and evaluate their performance. We show that there is a minimum threshold n^*, depending on the underlying distribution, with $n \geq n^*$ required in order for the asymptotic extreme-value approximations to be effective. The threshold n^* tends to increase as c^2 increases above 1 or decreases below 1.

Key words: two-moment approximations, extreme-value theory, maximum of independent random variables, Gumbel distribution.

1. Introduction and Summary

In this paper we investigate simple approximations suitable for engineering applications of a classical probability model. In particular, we are interested in the probability distribution of the *maximum* $M_n \equiv \max\{Z_1, Z_2, \ldots, Z_n\}$, where $Z_1, \ldots Z_n$ are n independent and identically distributed (i.i.d.) nonnegative random variables, each distributed as a random variable Z with a cumulative distribution function (cdf) F.

The distributions of maxima have many applications, e.g., in extreme-value engineering and insurance; see Castillo (1988) and Embrechts et al (1997). However, we are primarily motivated by a queueing problem, in particular, approximating the probability distribution of the last departure time from a multi-server queue with a terminating arrival process (a finite number of customers), when the service-time distribution is only partially characterized; see Goldberg and Whitt (2006). That problem in turn arose in the study of congestion associated with various inspection schemes, such as inspecting shipping containers at ports of embarkation (leaving to come to the country); see Crow et al. (2006b).

How does the model above relate to the queueing problem? The last departure time can be expressed as the sum of the time of the last arrival and the remaining time required to serve all customers in the system at the time of the last arrival. Assuming that the total number of arrivals is relatively large, the queueing system will be approximately in steady state at the time of the last arrival. Then the remaining time until the last departure will be approximately independent of the time of the last arrival and itself be the maximum of the remaining completion times. As an approximation for the case of a large number of servers, we consider the case of infinitely many servers. When there are infinitely many servers, these remaining completion times are all residual service times. If in addition the arrival process is Poisson (even nonhomogeneous), then these residual service times (except for the very last arrival), turn out to be i.i.d. with a distribution that can be determined; see Goldberg and Whitt (2006). Hence the last departure time involves the maximum of a random number of i.i.d. random variables, but where the distribution of these random variables may only be partially characterized, e.g., by its first few moments.

Thus we are interested in the probability distribution of the maximum M_n. Given the cdf F, we can easily *numerically calculate* the cdf of M_n, because

$$P(M_n \leq x) = F(x)^n, \quad x \geq 0 \; . \tag{1.1}$$

We also can numerically calculate the moments via

$$E[M_n^k] = \int_0^\infty kx^{k-1}[1 - F(x)^n]\, dx \ ; \tag{1.2}$$

e.g., see p. 150 of Feller (1971); and we can calculate quantiles ($x_{(n,q)}$ such that $P(M_n \leq x_{(n,q)}) = q$) by performing binary search with the cdf in (1.1).

However, suppose that we have only a *partial characterization* of the cdf F, as mentioned in the queueing application above. In particular, suppose that we know only its first two moments - $m_k \equiv E[Z^k]$ for $k = 1, 2$ - or, equivalently, only its mean $m_1 \equiv E[Z]$ and its *squared coefficient of variation* $c^2 \equiv c_Z^2$ (SCV, variance divided by the square of the mean). What can we say about the distribution of M_n now?

Here, we start by making the elementary observation that M_n is proportional to the mean $m_1 = E[Z]$. If we multiply all the random variables Z_n by some constant, then M_n itself is multiplied by that same constant. Hence, without loss of generality, we can assume that $m_1 = 1$ and we make that assumption. We thus ask: Given that $m_1 = 1$, how does the distribution of M_n depend on n and the SCV c^2 of the cdf F? And to what extent do n and c^2 determine the distribution of M_n?

We are interested in developing approximations for two reasons: first, to gain insight into the way the distribution of M_n depends on the cdf F and, second, to do further analysis; e.g., apply calculus to do optimization and embed the model in larger queueing network models, as in Whitt (1983), Bitran and Dasu (1992), Buzacott and Shanthikumar (1993) and Suri et al. (1993). Given the motivating inspection application, we are especially interested in approximations for moderate values of n, e.g., $10 \leq n \leq 1000$.

The Need for Regularity Conditions. A basis for understanding lies in the classical extreme-value theory; see Castillo (1988), Embrechts et al. (1997), Galambos (1987), Kotz and Nadarajah (2000), Resnick (1987) and Thomas and Reiss (2001). Paralleling the normal approximation from the central limit theorem, the distribution of M_n will usually have a relatively simple asymptotic form that will be a good approximation when n is sufficiently large. However, there is not one possible asymptotic form, but three; see Section 3.2 of Embrechts (1997) or Proposition 0.3 of Resnick (1987). The particular asymptotic form and the specific approximation depends on different properties of the distribution of the random variable Z - the asymptotic behavior of the tail $F^c(x) \equiv 1 - F(x)$ as $x \to \infty$; see Chapter 3 of Embrechts et al. (1997) or Chapter 1 of Resnick (1987).

Thus, concerning approximations with limited two-moment partial information, **the main message from extreme-value theory is that we should be cautious: we do not have the right information.** But suppose that we want a rough answer even if we do not have the right information. To avoid gross approximation errors, we clearly need to assume more. Accordingly, first, we assume that Z is unbounded above and, second, we assume that we can rule out the possibility of a heavy tail; we assume that

$$F^c(x) \le Ke^{-\lambda x} \quad \text{for all} \quad x \ge 0 \ , \tag{1.3}$$

for some constant K, where $\lambda > 0$.

Under those extra assumptions, the extreme-value theory tells us that the relevant asymptotic form is the Gumbel distribution - defined in (2.2) below. Moreover, the extreme-value theory tells us that we need extra regularity conditions and shows the impact of these regularity conditions. From a practical perspective, it is natural to assume as regularity conditions that the tail probability has the asymptotic property

$$F^c(x) \sim \gamma(\lambda x)^{\alpha} e^{-\lambda x} \quad \text{as} \quad x \to \infty \ , \tag{1.4}$$

for $0 < \lambda < \infty$ and $0 < \gamma < \infty$, where $f(x) \sim g(x)$ as $x \to \infty$ means that $f(x)/g(x) \to 1$ as $x \to \infty$. In practice, it is usually difficult to distinguish between (1.4) and other asymptotic behavior consistent with (1.3).

We *assume* that regularity condition (1.4) holds, but without directly verifying it. Consequently, we do not know the asymptotic parameters in (1.4). It is important to emphasize that the condition (1.4) is critical. We can have the same SCV with a distribution having heavy-tail tail asymptotics of the form

$$F^c(x) \sim \gamma x^{-\alpha} \quad \text{as} \quad x \to \infty \quad \text{for} \quad \alpha > 2 \ , \tag{1.5}$$

which will produce vastly different (larger) maxima. From data, these two cases can be distinguished by estimating $\log(F^c(x))$. Under (1.4), $\log(F^c(x))$ is approximately a linear function of x for large x; under (1.5), $\log(F^c(x))$ is approximately a linear function of $\log(x)$. If we actually had data, then we could directly estimate the parameters in (1.4) and apply available extreme-value approximations, as described in Sections 3 and 7 below. Here we assume that condition (1.4) holds, but that we do not know the parameters.

Regularity condition (1.4) covers two distinct cases: (i) the cdf F has a pure exponential tail ($\alpha = 0$) and (ii) the cdf F does *not* have a pure-exponential tail ($\alpha \neq 0$). The case of a pure-exponential tail is illustrated by the familiar *hyperexponential* (H_k) distributions (mixtures of exponential distributions), which have $c^2 > 1$. For $c^2 < 1$, the case of a pure-exponential tail is illustrated by *hypoexponential* distributions (convolutions of exponential distributions) when the component exponential distributions have different means and by a limiting case, the shifted-exponential distribution (the distribution of a constant plus an exponential random variable).

The case of a non-exponential tail is illustrated by gamma distributions, which includes Erlang distributions (convolutions of exponential distributions when all the component exponential distributions have the same mean). Based on previous experience with asymptotic approximations, e.g., as in Abate and Whitt (1997), we anticipate that asymptotic extreme-value approximations will be more effective with a pure exponential tail, and that will be borne out here. We will see that the extreme-value approximations take a simpler form and are more accurate when F has a pure-exponential tail.

A Numerical Approach. To obtain numerical results, we suggest fitting a representative approximating distribution to the partial information and then computing the exact distribution of the maximum according to (1.1). We will show that approach is more reliable than the classical extreme-value approximations, even when we know the asymptotic parameters in (1.4), when the SCV c^2 is very large or small (e.g., $c^2 \geq 16$ or $c^2 \leq 1/16$) and n is not extremely large (e.g., $n \leq 1000$).

As is frequently done in queueing approximations, e.g., see Whitt (1982-1984), we suggest using the exponential distribution for $c^2 \approx 1$, the H_2 distribution with an appropriate choice of the third parameter for $c^2 > 1$ (matching the first three moments, if possible), and the shifted-exponential distribution or a convolution of exponential distributions for $c^2 < 1$. They all have pure-exponential tails. As an alternative, and for comparisons, we also consider the gamma distribution for all $c^2 > 0$, which does not have a pure-exponential tail. We give details in the remaining sections. As noted above, all four distributions are "exponential-like."

Simple Approximations and Insights. Nevertheless, here we are primarily interested in closed-form analytic approximations. We will show that it is possible to develop reasonable closed-form approximations for moderate values of n and c^2, such as $10 \leq n \leq 1000$ and $1/16 \leq c^2 \leq 16$. Our starting point is the classical extreme-value theory associated with (1.4) in the case of a pure exponential tail ($\alpha = 0$). Since the extreme-value approximations are not consistently good for all n and c^2, an important component of our approximation is an indication when the extreme-value approximations will be appropriate.

The extreme-value theory produces the following approximation for the q^{th} quantile of M_n:

$$x_{(n,q)} \approx \lambda^{-1} \left[\log(n\gamma) - \log\log(1/q)\right] \ , \tag{1.6}$$

where *log* is the natural logarithm (base e) and λ and γ are the asymptotic parameters in (1.4); see Section 3. Based on an examination of representative special cases, as crude approximations for the asymptotic parameters, assuming that the mean is $m_1 = 1$, we suggest

$$\lambda^{-1} \approx c^2 \quad \text{and} \quad \gamma \approx \frac{1}{c^2} \ . \tag{1.7}$$

Combining (1.6) and (1.7), and ignoring the $\log\log(1/q)$ term, we suggest the following **crude approximation**:

$$E[M_n] \approx x_{(n,q)} \approx c^2 \log\left(\frac{n}{c^2}\right) \ . \tag{1.8}$$

The approximation for $x_{(n,q)}$ follows directly from (1.6) and (1.7), if we assume that $\log\log(1/q) \approx 0$. That occurs for $q = 1/e \approx 0.368$. By ignoring this term involving $\log\log(1/q)$, we are regarding all central quantiles as being approximately the same. We relate to the mean by approximating the mean by the median or any other quantile. Clearly, if n is large enough, then we can ignore the $\log\log(1/q)$ term, but n may not be large enough, which partly explains why (1.8) is presented as a crude approximation.

From extreme-value theory, the role of $\log(n)$ in (1.6), and thus in (1.8), is well known, but we are unaware of any previous statements about the approximate role of the SCV c^2. However, we caution that our numerical experiments show that approximation (1.8) is *not* sufficiently accurate for practical purposes when $c^2 < 1$.

As a **simple rough approximation** for the q^{th} quantile $x_{(n,q)}$, suitably for practical purposes, we propose

$$x_{(n,q)} \approx \phi(c^2)\left[\log\left(n\psi(c^2)\right) - \log\log(1/q)\right], \quad n \ge n^* \equiv n^*(c^2, q) \ , \tag{1.9}$$

or, equivalently, the extreme-value approximation (1.6) with

$$\lambda^{-1} \approx \phi(c^2) \quad \text{and} \quad \gamma \approx \psi(c^2) \ , \tag{1.10}$$

where

$$\phi(c^2) \equiv \begin{cases} c^2 \ , & c^2 \ge 1 \ , \\ c \equiv \sqrt{c^2} \ , & c^2 \le 1 \ , \end{cases} \tag{1.11}$$

$$\psi(c^2) \equiv \begin{cases} \frac{c^2+1}{2(c^2)^2} \approx \frac{1}{c^2} \ , & c^2 \ge 1 \ , \\ e^{\{(1-\sqrt{c^2})/\sqrt{c^2}\}} \approx \frac{1}{c} \ , & c^2 \le 1 \ , \end{cases} \tag{1.12}$$

and

$$n^*(c^2, q) \equiv \begin{cases} \frac{c^2}{q} \ , & c^2 \ge 1 \ , \\ \frac{1}{q} \ , & c^2 \le 1 \ . \end{cases} \tag{1.13}$$

The first case of (1.9) with $c^2 \ge 1$ is based on the H_2 distribution (and an appropriate choice of the third parameter); the second case with $c^2 \le 1$ is based on the shifted-exponential distribution. They are both consistent with extreme-value theory; i.e., they are asymptotically correct as $n \to \infty$ (for an appropriate choice of the third parameter in the case of H_2).

The two separate cases of (1.9) agree at the boundary $c^2 = 1$, coinciding with the well-known exponential asymptotic extreme-value approximation in (2.3) and (3.6) below. The final term involving $-\log\log(1/q)$ tends to be relatively negligible when q is near 0.5 (the median); indeed it equals 0 when $q = e^{-1} \approx 0.368$. For $q = 0.25$, 0.5, 0.75 and 0.9, $-\log\log(1/q) = -0.327$, 0.367, 1.25 and 2.25, respectively.

The approximation for $c^2 \le 1$ in (1.9) can also be re-expressed as

$$x_{(n,q)} \approx 1 - \sqrt{c^2} + \sqrt{c^2}\log(n) - \sqrt{c^2}\log\log(1/q), \quad c^2 \le 1 \ . \tag{1.14}$$

This alternative form in (1.14) says that the quantile $x_{(n,q)}$ can be expressed as a convex combination of corresponding extreme-value approximations for the quantile when the underlying distribution is exponential and deterministic, using the weight $c \equiv \sqrt{c^2}$ on the exponential term.

Experience shows that the value of n required for asymptotic extreme-value approximations to be useful depends on the underlying cdf F. We find that the SCV c^2 can also be used to provide a simple rough indication of the range of n for which the simple rough approximation in (1.9) is reasonable. When $c^2 = 1$ and F is exponential, (1.9) is remarkably accurate even for small n (e.g., for all $n \ge 5$). We find that is decidedly not the case when c^2 is much greater than 1. In order to invoke approximation (1.9), and in order to apply other asymptotic extreme-value approximations more generally, n should be larger as the SCV c^2 increases above 1. We give a suggested range for n in (1.13).

For the shifted-exponential distribution, we do not need $n^*(c^2, q)$ to increase as c^2 decreases below 1, but we do for the gamma distribution; see Section 7. Since the the cdf F has mean 1, it is reasonable to require that any estimate of $x_{(n,q)}$ be at least 1. Thus we would refine (1.9) to be the maximum of 1 and the calculated value.

Organization of the rest of this paper. In Sections 2 and 3 we review the simple case of the exponential distribution and the asymptotic result for the case in which the cdf F has a pure exponential tail.

Starting with distributions having $c^2 \geq 1$, in Section 4 we consider the special case of an H_2 distribution. The H_2 distribution has three parameters, so there is an additional degree of freedom. For the H_2 distribution, we ask how the distribution of M_n depends on this additional parameter. We will show that the simple rough approximation above is reasonable, but in pathological cases it can break down completely.

Turning to distributions with $c^2 \leq 1$, in Sections 5, 6 and 7 we consider shifted-exponential distributions, convolutions of exponential distributions and gamma distributions. We consider reverse engineering in Section 8; we estimate F given the distribution of the maximum as a function of n. Finally, we draw conclusions in Section 9. Additional supporting tables and plots appear in an Internet supplement, Crow et al. (2006a).

2. The Exponential Distribution

The distribution of the maximum M_n when the cdf F is exponential is known to be the sum of n exponentials with means k^{-1}, $1 \leq k \leq n$, so that the mean is exactly the harmonic number $H_n \equiv \sum_{k=1}^{n} k^{-1}$. The distribution also has a simple accurate approximation. As we show in the next section, for the exponential distribution (with mean $m_1 = 1$), we have the limit

$$M_n - \log(n) \Rightarrow W \quad \text{as} \quad n \to \infty \ , \tag{2.1}$$

where $\Rightarrow$ denotes convergence in distribution and W is a random variable with the *Gumbel distribution*, i.e.,

$$G(x) \equiv P(W \leq x) \equiv \exp\{-e^{-x}\}, \quad \text{for all} \quad x \in \mathbf{R} \ , \tag{2.2}$$

which has mean and variance $E[W] = \zeta \approx 0.5772$, the Euler constant, and $Var(W) = \pi^2/6 \approx 1.6449$; see Johnson and Kotz (1970). The Gumbel distribution has q^{th} quantile $x_{(q)} = -\log\log(1/q)$, where $G(x_{(q)}) \equiv q$. The mode of the Gumbel distribution is at 0, which is the $(1/e)^{\text{th}} = 0.37^{\text{th}}$ quantile.

Hence, for the exponential distribution (with mean $m_1 = 1$), we have the approximations

$$\begin{aligned} M_n &\approx \log(n) + W \\ E[M_n] &\approx \log(n) + 0.5772 \\ Var(M_n) &\approx 1.6449 \\ x_{(n,q)} &\approx \tilde{x}_{(n,q)} \equiv \log(n) - \log\log(1/q)\ , \end{aligned} \tag{2.3}$$

where $P(M_n \leq x_{(n,q)}) \equiv q$.

The approximations in (2.3) based on the Gumbel distribution are remarkably accurate when F is exponential, even for small n, e.g., $n = 5$. In Table 1 we compare exact values to approximations for three quantiles of the cdf of the maximum: $q = 0.25$, $q = 0.50$ and $q = 0.75$. The results are spectacular, provided that q and n are not both too small. The final column gives the crude approximation in (1.8), which here only ignores the $\log\log(1/q)$ term. in (2.3).

	$q = 0.25$		$q = 0.50$		$q = 0.75$		
n	exact	approx.	exact	approx.	exact	approx.	crude (1.8)
5	1.42	1.28	2.04	1.98	2.88	2.86	1.60
10	2.04	1.98	2.70	2.67	3.56	3.55	2.30
100	4.29	4.28	4.97	4.97	5.85	5.85	4.61
1000	6.58	6.58	7.27	7.27	8.15	8.15	6.90

Table 1: A comparison of exact values with asymptotic approximations from (2.3) for three quantiles of the cdf of the maximum of n i.i.d. exponential random variables with mean 1 for four values of n.

The exponential case illustrates *important basic phenomena*: The mean $E[M_n]$ grows with n like $\log(n)$, while the variance is asymptotically constant, so the distribution concentrates about the mean (relatively) when n is sufficiently large. Similarly, any q^{th} quantile also grows like $\log(n)$, but the difference of two quantiles, $x_{(n,q_2)} - x_{(n,q_1)}$ for $0 < q_1 < q_2 < 1$, is asymptotically constant. The predictability for large n resulting from the asymptotically constant spread is remarkable. However, $\log(n)$ increases slowly with n, so that we do not see that relative concentration if n is not very large. In practice (for moderate n), the mean and standard deviation are usually of the same order.

3. Asymptotics for a Pure Exponential Tail

In this section we assume that the cdf F has a pure-exponential tail; in particular, we assume that (1.4) holds with $\alpha = 0$. (Here we make no assumption about the mean of F.) The pure-exponential-tail assumption is a "common case," and is satisfied by all H_k distributions. To make this discussion self-contained, we review the classical result and its proof.

Theorem 3.1. (*pure exponential tail*) *If condition* (1.4) *holds with* $\alpha = 0$, *then*

$$M_n - \frac{\log(n\gamma)}{\lambda} \Rightarrow \frac{W}{\lambda} \ , \tag{3.1}$$

where W *is a random variable with the Gumbel distribution in* (2.2), *while* λ *and* γ *are the asymptotic parameters in* (1.4).

To prove the classic result, we exploit the basic lemma:

Lemma 3.1. *If* c_n *are real numbers such that* $c_n \to c$ *as* $n \to \infty$, *then*

$$\left(1 + \frac{c_n}{n}\right)^n \to e^c \quad as \quad n \to \infty \ . \tag{3.2}$$

Proof of Theorem 3.1. Apply Lemma 3.1 to get

$$\begin{aligned} P\left(M_n \le \frac{\log(n\gamma) + x}{\lambda}\right) &= \left(1 - F^c\left(\frac{\log(n\gamma) + x}{\lambda}\right)\right)^n \\ &\sim \left(1 - \frac{e^{-x}}{n}\right)^n \to \exp\{-e^{-x}\} \quad \text{as} \quad n \to \infty \ , \end{aligned} \tag{3.3}$$

because

$$F^c\left(\frac{\log(n\gamma) + x}{\lambda}\right) \sim \gamma e^{-\lambda[(\log(n\gamma)+x)/\lambda]} = \frac{e^{-x}}{n} \quad \text{as} \quad n \to \infty \tag{3.4}$$

for each fixed x, by virtue of assumption (1.4), so that

$$nF^c\left(\frac{\log(n\gamma) + x}{\lambda}\right) \to e^{-x} \quad \text{as} \quad n \to \infty \ . \quad \blacksquare \tag{3.5}$$

Hence we have the approximations:

$$\begin{aligned} M_n &\approx \lambda^{-1}(\log(n\gamma) + W) \\ E[M_n] &\approx \lambda^{-1}(\log(n) + \log(\gamma) + 0.5772) \\ Var(M_n) &\approx 1.6449\lambda^{-2} \\ x_{(n,q)} &\approx \tilde{x}_{(n,q)} \equiv \lambda^{-1}\left[\log(n\gamma) - \log\log(1/q)\right] \ , \end{aligned} \tag{3.6}$$

4. The H_2 Distribution

In this section we suppose that Z has an H_2 distribution, with tail probability

$$F^c(x) \equiv P(Z > x) = pe^{-\lambda x} + (1-p)e^{-\eta x}, \quad x \ge 0, \tag{4.1}$$

where $0 < \lambda < \eta$ and $0 < p < 1$, which implies a pure exponential tail asymptotically, just as in (1.4) with $\alpha = 0$ and $\gamma = p < 1$. Hence we use the approximations in (3.6) with $\gamma = p$. Again we initially make no assumption about the mean. We propose using the H_2 distribution as an approximation when we are given only the first two or three moments of Z in the case $c^2 > 1$. We indicate how to do the fitting below.

The H_2 distribution highlights potential limitations of the extreme-value theory. With the extreme-value theory, we act as if we can ignore the second term in (4.1); i.e., we use the approximation

$$F^c(x) \approx pe^{-\lambda x}, \quad x \geq 0. \tag{4.2}$$

We can see that the approximation in (4.2), and thus the associated approximations in (3.6), will perform poorly if η is very close to λ or if p is extremely small when n is not large. To see the problem, suppose that $\lambda = \eta$. Then (1.4) holds with $\alpha = 0$ and $\gamma = 1$, but we use $\gamma = p$ instead. However, while it is good to be aware of these difficulties, we do not expect them to commonly arise.

We proceed by developing further approximations for the extreme-value approximations in terms of other parameters besides the initial three: λ, η and p. In particular, we want to use approximation (3.6) but with approximations for the parameters λ^{-1} and p.

Alternative Parameter Triples. For approximations, natural alternative parameters are the first three moments. Essentially equivalent are the mean m_1 (here allowed to be general) and the SCV $c^2 \equiv (m_2 - m_1^2)/m_1^2$ and the *skewness* $\beta \equiv m_3^2/m_2^3$. Like the SCV, the skewness is appealing because it is independent of scale: it remains unchanged if Z is multiplied by a constant. All moments are easy to compute, since moments of mixtures are mixtures of moments; Z has k^{th} moment

$$E[Z^k] = \frac{pk!}{\lambda^k} + \frac{(1-p)k!}{\eta^k} \ . \tag{4.3}$$

In terms of the first three moments, we can obtain expressions for the standard parameters:

$$\begin{aligned}
\lambda^{-1} &= m_1 \left(\frac{(u + 1.5v^2 + 3v) + \sqrt{(u + 1.5v^2 - 3v)^2 + 18v^3}}{6v} \right) , \\
\eta^{-1} &= m_1 \left(\frac{(u + 1.5v^2 + 3v) - \sqrt{(u + 1.5v^2 - 3v)^2 + 18v^3}}{6v} \right) , \\
p &= \frac{1 - (\eta^{-1}/m_1)}{(\lambda^{-1}/m_1) - (\eta^{-1}/m_1)} ,
\end{aligned} \tag{4.4}$$

where

$$u \equiv (m_3/m_1^3) - 1.5(m_2/m_1)^2 \quad \text{and} \quad v \equiv (m_2/m_1^2) - 2 \tag{4.5}$$

and we again use the convention that $\lambda^{-1} > \eta^{-1}$, e.g., see p. 136 of Whitt (1982). When we are fitting an H_2 distribution to the first three moments, we require $u \ge 0$ and $v \ge 0$. Having $v \ge 0$ is equivalent to $c^2 \ge 1$. If $c^2 = 1$, then the distribution must be exponential and there is no flexibility in the third moment. For $c^2 > 1$, we must have $m_3 \ge 1.5 m_2^2 m_1$, which is a restriction. If m_3 is initially too small, then we suggest choosing a value slightly above the lower bound $1.5 m_2^2 m_1$.

Given data, we can estimate the asymptotic parameters λ and γ in (1.4). Even if we know the asymptotic parameters λ and γ in (1.4) (with $\alpha = 0$), we might be interested in fitting an H_2 distribution, because the asymptotic extreme-value approximations developed in Section 3 are effective only for n above a threshold n^*. We might want an H_2 approximation to develop an approximation for smaller n. Given the mean m_1 and the asymptotic parameters λ and γ in (1.4) (with $\alpha = 0$), where $\lambda^{-1} > m_1$ and $\gamma < 1$, we suggest fitting the H_2 distribution by letting $p = \gamma$ and choosing η to make

$$m_1 = \frac{p}{\lambda} + \frac{1-p}{\eta} \; . \tag{4.6}$$

By the assumptions above, necessarily $\eta > \lambda$, so that the H_2 distribution has the given asymptotics. We can roughly judge the quality of the H_2 approximation by comparing the SCV of the H_2 distribution to the actual SCV.

As an appealing alternative third parameter to go with the mean and the SCV is the ratio of the dominant component mean to the total mean, i.e.,

$$r \equiv \frac{(p/\lambda)}{m_1} \equiv \frac{(p/\lambda)}{(p/\lambda) + ((1-p)/\eta)} \; . \tag{4.7}$$

As a function of m_1, c^2 and r, we can express the basic parameters λ, η, and p as

$$\begin{aligned}
\lambda m_1 &= \left(\frac{1}{c^2+1}\right)\left(\frac{1}{r}\right)\left(\frac{c^2+1}{2} + 2r - 1 - \frac{(c^2-1)}{2}\sqrt{1 + \frac{8r(1-r)}{(c^2-1)}}\right) \\
p/m_1 &= \lambda r \\
\eta m_1 &= \frac{(1-p)}{(1-r)} \; .
\end{aligned} \tag{4.8}$$

For an alternative expression in terms of r, see p. 169 of Whitt (1984).

Approximations for the parameters λ^{-1} and p. We develop an approximation for λ^{-1} by establishing upper and lower bounds for the square root. In particular, we use the following elementary lemma.

Lemma 4.1. *For all $x > 0$,*

$$1 + \frac{x}{2} - \frac{x^2}{8} \le \sqrt{1+x} \le 1 + \frac{x}{2} \ . \tag{4.9}$$

Theorem 4.1. *For all $m_1 > 0$, $c^2 > 1$ and $0 < r < 1$,*

$$m_1 \left(\frac{c^2+1}{2}\right)\left(\frac{1}{r}\right)\left(\frac{1}{1+\frac{2(1-r)^2}{(c^2-1)}}\right) \le \lambda^{-1} \le m_1 \left(\frac{c^2+1}{2}\right)\left(\frac{1}{r}\right) \ , \tag{4.10}$$

so that

$$\lambda^{-1} \sim m_1 \left(\frac{c^2+1}{2}\right)\left(\frac{1}{r}\right) \quad as \quad c^2 \to \infty \quad or \quad r \to 1 \ . \tag{4.11}$$

and

$$p \equiv \gamma = m_1 \lambda r \sim \frac{2r^2}{c^2+1} \quad as \quad c^2 \to \infty \quad or \quad r \to 1 \ . \tag{4.12}$$

Now, to simplify, we again assume that $m_1 = 1$. The limit in (4.11) yields the aesthetically appealing **product approximation** for λ^{-1}:

$$\lambda^{-1} \approx \left(\frac{c^2+1}{2}\right)\left(\frac{1}{r}\right) \quad \text{with} \quad p \equiv \gamma \approx \frac{2r^2}{c^2+1} \ . \tag{4.13}$$

Combining (4.13) and (3.6), we obtain the associated H_2 maximum-quantile approximation

$$x_{(n,q)} \approx \left(\frac{c^2+1}{2}\right)\left(\frac{1}{r}\right)\left(\log\left(\frac{2r^2 n}{c^2+1}\right) - \log\log\left(1/q\right)\right) \tag{4.14}$$

There are several observations to make about the product approximation for λ^{-1} in (4.13). First, Theorem 4.1 implies that the approximation in (4.13) is an upper bound for λ^{-1}, so it is conservative (overestimates λ^{-1}). Second, the bounds in Theorem 4.1 imply that the maximum percentage error for approximation (4.13) is

$$\text{max percentage error} \equiv 100 \times \frac{(UB - LB)}{LB} = \frac{200(1-r)^2}{(c^2-1)}\% \ . \tag{4.15}$$

For example, if $r = 1/2$ (balanced means), then the maximum percentage error in (4.13) is at most $[50/(c^2-1)]\%$; if in addition $c^2 = 6$, then it is at most 10%. Numerics indicates the error typically is much less. For example, if $c^2 = 6$ the maximum error in (4.13) over all r is at most 5%. However the product approximation for λ^{-1} in (4.13) is not good for c^2 close to 1.

Approximation as a function of the SCV alone. We obtain an approximation as a function of the SCV c^2 alone (with $m_1 = 1$) by just substituting a representative value of r. For that purpose, it is common to assume $r = 1/2$, corresponding to *balanced means.* When $r = 1/2$, formula (4.8) (with $m_1 = 1$) simplifies to

$$\lambda = 1 - \sqrt{(c^2-1)/(c^2+1)} \ . \tag{4.16}$$

However, it is actually not so natural to have r fixed at $1/2$ for all c^2. For most values of c^2, $r = 1/2$ is entirely reasonable, but we contend that it is more natural to have $r \uparrow 1$ as $c^2 \downarrow 1$. To see that, we can apply (4.8) to see what happens to the parameters as $c^2 \downarrow 1$. By (4.8), $\lambda \to 1$, $\eta \to 1$ and $p \to r$ as $c^2 \downarrow 1$ for any fixed r with $0 < r < 1$ (and $m_1 = 1$). The limiting H_2 distribution then has two exponential components, each with mean 1, one with probability r and the other with probability $1 - r$. That limit is equivalent to the correct exponential distribution, but it is pathological from the perspective of the regularity condition in (1.4). From the approach to the limit as $c^2 \downarrow 1$, we get $F^c(x) \sim re^{-\lambda x}$ as $x \to \infty$ instead of $F^c(x) \sim e^{-\lambda x}$ for the exponential.

Thus we here propose a different choice for r, making r a non-constant function of c^2, which approaches a simple exponential as $c^2 \downarrow 1$. In particular, we propose

$$r \equiv r(c^2) \equiv \frac{c^2 + 1}{2c^2} \ . \tag{4.17}$$

With (4.17), we have $r(c^2) \approx 1/2$ for most values of c^2, but we have $r(c^2) \uparrow 1$ as $c^2 \downarrow 1$, so that the pathology above is avoided. It is also reasonable to have the function of c^2 be monotone and continuous. Using (4.17), the product approximation in (4.13) and (4.14) become

$$\lambda^{-1} \approx c^2 \quad \text{with} \quad p \equiv \gamma \approx \frac{c^2 + 1}{2(c^2)^2} \tag{4.18}$$

and (1.9).

For this particular choice of r, the product approximation is consistently good by virtue of (4.10). Since

$$\frac{2(1-r)^2}{c^2 - 1} = \frac{c^2 - 1}{2(c^2)^2} \le \frac{1}{8} \ , \tag{4.19}$$

with the maximum occurring at $c^2 = 2$, the maximum relative error in λ^{-1} for this r in the product approximation is $1/8$ or 12.5%. Significant errors for H_2 in the approximation based on c^2 alone will thus occur only because the actual value of r differs substantially from (4.17).

Approximation (4.13) dramatically shows that very small r will seriously invalidate the more elementary approximations based on the first two moments in (1.8) and (1.9). By using the extremal distributions and stochastic ordering for H_2 distributions in Whitt (1984), or by applying (4.8), we can conclude what is possible as we allow r to vary, for given first two moments. We see that the range is just as predicted by the product approximation in (4.13).

Theorem 4.2. *For $m_1 = 1$ and $c^2 > 1$ given, all values of λ^{-1} are possible with*

$$\lambda^{-1} \ge \left(\frac{c^2 + 1}{2} \right) \ . \tag{4.20}$$

The minimal value of λ^{-1} is obtained in the degenerate case with $\eta^{-1} = 0$ and

$$p \equiv \gamma = \frac{2}{c^2 + 1} \ . \tag{4.21}$$

From Theorem 4.2, we see that in order for the approximation in (1.9) based on the SCV to be useful, even for the H_2 distribution, we depend on extra regularity conditions. When the cdf F is H_2, we need to assume that r is not too small. If we can assume that r does not differ much from the new approximation in (4.17), then we can use the simple rough two-parameter approximations in (4.18) and (1.9).

Minimum thresholds for n. So far, we have focused on approximating the key parameter λ^{-1}. Now we want to consider another issue. We observe that we can still have problems with the approximation of the cdf of M_n if p is too small. Because of the mixing property of H_2 distributions, we can regard the n H_2 random variables as a random number, N, of exponential random variables with mean λ^{-1} and another random number number, $n - N$, of exponential random variables with mean η^{-1}, where N has a binomial distribution with parameters n and p. Because we exploit approximation (4.2), the approximation is driven by the N exponentials with mean λ^{-1}. Since $np = E[N]$, the mean number of exponential random variables with mean λ^{-1}, the approximation becomes problematic when np is small.

From (4.13) or (4.18), we know that p will be small when c^2 is large, so that we should anticipate difficulties in that case. Let n^* be the *threshold value* of n for which the extreme-value approximation will be effective. From above, we anticipate that we should have $n^* \geq 1/p$. From our numerical experiments, we see that n needs to be larger when q is smaller. Thus, as a rough criterion we propose $n \geq 1/(pq)$. Using (4.13), we can find an approximate expression for the threshold value, namely,

$$n^* \approx \frac{1}{pq} \approx \frac{(c^2+1)}{2r^2 q} \approx \frac{c^2}{q} \ . \tag{4.22}$$

To provide further insight in the case of small p (or high c^2), we suggest exploiting this mixing property further. We can express the cdf of the maximum exactly as a binomial mixture, and then for large n and small p we can perform a Poisson approximation, getting

$$\begin{aligned} P(M_n \leq x) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} (1 - e^{-\lambda x})^k (1 - e^{-\eta x})^{n-k} \\ &\approx \sum_{k=0}^{\infty} \frac{e^{-np}(np)^k}{k!} (1 - e^{-\lambda x})^k (1 - e^{-\eta x})^{n-k} \\ &= e^{-np} (1 - e^{-\eta x})^n e^{\{np(1-e^{-\lambda x})/(1-e^{-\eta x})\}} \end{aligned}$$

$$= \left[e^{-p}(1-e^{-\eta x})e^{\{p(1-e^{-\lambda x})/(1-e^{-\eta x})\}}\right]^n \ . \tag{4.23}$$

Assuming that $\lambda << \eta$ and that the relevant x is relatively large, we can further approximate, starting in the second line of (4.23). For small np and λ, we get the simple approximation

$$P(M_n \le x) \approx e^{-np}(1-e^{-\eta x})^n + \sum_{k=1}^{\infty} \frac{e^{-np}(np)^k}{k!}(1-e^{-\lambda x})^k \ . \tag{4.24}$$

From (4.24), we can get an associated approximation for the mean. First, from Section 2,

$$E[M_n] \approx e^{-np}\left(\frac{\log(n)+\zeta}{\eta}\right) + \sum_{k=1}^{\infty} \frac{e^{-np}(np)^k}{k!}\left(\frac{\log(k)+\zeta}{\lambda}\right) \ , \tag{4.25}$$

where $\zeta \approx 0.5772$ is again the Euler constant. Then, exploiting the asymptotic approximations $\log(n) \sim H_n - \zeta$, where $H_n \equiv \sum_{k=1}^{n} k^{-1}$ are the *harmonic numbers* (not to be confused with hyperexponential distribution), and the relation

$$\sum_{n=1}^{\infty} \frac{x^n H_n}{n!} = e^x[\log(x) + E_1(x) + \zeta] \ , \tag{4.26}$$

where $E_1(x)$ is the *exponential integral*, i.e.,

$$E_1(x) \equiv \int_x^{\infty} t^{-1}e^{-t}\, dt \ , \tag{4.27}$$

see Chapter 5 of Abramowitz and Stegun (1972) and (25) of Weisstein et al. (2005), we get

$$E[M_n] \approx e^{-np}\left(\frac{\log(n)+\zeta}{\eta}\right) + \frac{\log(np) + E_1(np) + \zeta}{\lambda} \ . \tag{4.28}$$

For reference, note that $E_1(1) \approx 0.22$, $E_1(x) \to 0$ as $x \to \infty$ and

$$\log(x) + E_1(x) + \zeta \to 0 \quad \text{as} \quad x \to 0 \ . \tag{4.29}$$

Formula (4.28) shows what happens as n increases for very small p. Initially, $e^{-np} \approx 1$ so that $E[M_n]$ behaves like $(\log(n)+\zeta)/\eta$ and eventually, as n increases, $e^{-np} \approx 0$, so that $E[M_n]$ behaves like $(\log(np)+\zeta)/\lambda$.

A revealing double limit. We now introduce an appealing asymptotic regime that provides additional insight into the threshold n^*. We let *both* $n \to \infty$ and $c^2 \to \infty$, but at a coordinated rate. To properly formulate the double limit, we consider cdf's F_n indexed by n. Let

$$M_{n,k} \equiv \max\{Z_{n,1}, \cdots, Z_{n,k}\}, \quad n \ge 1 \quad \text{and} \quad k \ge 1 \ , \tag{4.30}$$

where $\{Z_{n,k} : k \ge 1\}$ is a sequence of i.i.d. random variables distributed as F_n for each n. Let $\lfloor x \rfloor$ be the greatest integer less than x.

Theorem 4.3. *Consider a sequence of H_2 cdf's $\{F_n : n \geq 1\}$. Suppose that the cdf's F_n have common mean 1 and r, $0 < r < 1$, for all n. Let c_n^2 be the SCV of F_n and assume that $c_n^2 \to \infty$ as $n \to \infty$ such that*

$$\frac{2nr^2}{c_n^2 + 1} \to \beta \quad \textit{or, equivalently, by (4.12),} \quad np_n \to \beta \ . \tag{4.31}$$

If $\beta > 0$ in (4.31), then for any constant $\xi > 0$

$$P(\lambda_n M_{n,\lfloor \xi n \rfloor} \leq \log(\beta\xi) + x) \to e^{-e^{-x}} \quad \textit{as} \quad n \to \infty \quad \textit{for} \quad \log(\beta\xi) + x > 0 \ . \tag{4.32}$$

If instead $\beta = 0$ in (4.31), then

$$\lambda_n M_{n,\lfloor \xi n \rfloor} \Rightarrow 0 \quad \textit{as} \quad n \to \infty \ . \tag{4.33}$$

Proof. For (4.32), it suffices to apply Lemma 3.1 after noting that

$$P\left(M_{n,\lfloor \xi n \rfloor} \leq \lambda_n^{-1}(\log(\beta\xi) + x)\right) = \left(1 - \frac{\xi np_n}{\xi n} e^{-(\log(\beta\xi)+x)} - (1 - p_n)e^{-\eta_n \lambda_n^{-1}(\log(\beta\xi)+x))}\right)^{\lfloor \xi n \rfloor} , \tag{4.34}$$

provided that $\log(\beta\xi) + x > 0$ (to make the argument of the probability on the left positive). Assuming that $\log(\beta\xi)+x > 0$, the last term on the right in (4.34) is asymptotically negligible, because $\lambda_n^{-1} = r/p_n = O(n)$ and $\eta_n = (1-p_n)/(1-r) = (1-O(n^{-1}))/(1-r) \to \eta = 1/(1-r) > 0$ as $n \to \infty$. On the other hand, if $\beta = 0$, then for any $x > 0$

$$P\left(M_{n,\lfloor \xi n \rfloor} \leq \lambda_n^{-1}x)\right) = \left(1 - \frac{\xi np_n}{\xi n} e^{-x} - (1 - p_n)e^{-\eta_n \lambda_n^{-1} x}\right)^{\lfloor \xi n \rfloor} \to e^0 = 1 \quad \text{as} \quad n \to \infty \ , \tag{4.35}$$

again by Lemma 3.1. That implies (4.33). ■

Theorem 4.3 has interesting implications. Recall that the conditions have $1/p_n$ or, equivalently, c_n^2 increase as n increases. The limit (4.33) says that if $n << 1/p$, then the maximum is negligible compared to λ^{-1}, the mean of the dominant exponential. In that case, clearly the extreme-value approximation is not appropriate.

With condition (4.31) and $\xi = 1$, the condition $\log(\beta) + x > 0$ in (4.32) translates into $\log(np) + x > 0$. A relevant value of x is the q^{th} quantile of the Gumbel distribution, which is $-\log\log(1/q)$. In order for the extreme-value theory to yield a good approximation for $x_{(n,q)}$, this double limit suggests we should have

$$\log(np) - \log\log(1/q) > 0 \quad \text{or, equivalently,} \quad n > \frac{\log(1/q)}{p} \ . \tag{4.36}$$

In other words, the double limit suggests an approximate value for the threshold for n:

$$n^* \approx \frac{\log(1/q)}{p} \approx c^2 \log(1/q) \ , \qquad (4.37)$$

which is roughly consistent with (1.13).

Numerics for the H_2 distribution. Numerical comparisons between approximations and exact values for the H_2 distribution with $c^2 = 4.0$ are given in Tables 2-4, considering four values of n: $n = 10$, 20, 100 and 1000. We let the individual H_2 random variables have mean 1, $c^2 = 4.0$ and we consider three values of r: $r = 0.25$, $r = 0.50$ and $r = 0.75$. In these three cases, the (λ, p, η) triples are, respectively $(0.1303, 0.0326, 1.2899)$, $(0.2254, 0.1127, 1.7746)$ and $(0.3101, 0.2326, 3.0697)$.

We again treat the quantiles $q = 0.25$, $q = 0.50$ and $q = 0.75$, but start with the median $q = 0.50$ in Table 2. We give five values: (i) the exact values from (1.1) plus binary search, (ii) the asymptotic values from (3.6) and (4.8), (iii) the product approximation in (4.14), (iv) the simple rough approximation in (1.9) and the crude approximation in (1.8).

The balanced-means case $r = 0.5$ plays a special role, because the exact values in that case serve as a natural approximation based on the first two moments alone. The exact value with $r = 0.5$ is our proposed **numerical approximation** for all cdf's with SCV $c^2 = 4$ (assuming regularity condition (1.4)). The product approximation with $r = 0.5$ is a second two-moment approximation, serving as an alternative to (1.9). The exact value as a function of r is our proposed numerical approximation given the first three moments. (We then use (4.4) and (4.7) to get the conventional H_2 parameters.)

In Table 2 we also display the exact values of np based on (4.8) and the approximations based on (4.13). (These are not repeated in Tables 3 and 4 because the values do not change with q.) We have indicated that we need np to be suitably large, not just n. We highlight in bold those problematic cases in which $np < 1/q = 2.0$. We anticipate that in these cases n is not yet large enough for the extreme-value approximations to perform well, and that is confirmed. First we see that we obtain good rough approximations for np, which is only used to estimate whether the approximations should be effective. Next we see that the extreme-value approximations in (3.6) are again spectacular if n is large enough, in particular, for $np > 1/q = 2.0$. From either the exact or the approximate values of np in Table 2, we are able to accurately predict when the approximations will perform well. With that guide, the extreme-value approximations perform well: we first use the approximation of p in (4.13) or

	exact (1.1)			asymp. (3.6)			product (4.14)			simple (1.9)	crude (1.8)
$n\backslash r$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75		
10	2.4	2.8	4.0	**-5.8**	**2.2**	3.9	**-10.2**	**1.8**	3.9	**3.3**	3.7
20	3.3	5.3	6.2	**-0.5**	5.2	6.1	**-3.3**	5.3	6.2	6.0	6.4
100	11.9	12.4	11.3	11.9	12.4	11.3	12.8	13.3	11.6	12.5	12.9
1000	29.5	22.6	18.8	29.5	22.6	18.8	35.9	24.9	19.3	21.7	22.1
	exact np						approx. np				
$n\backslash r$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	$n\psi(c^2)$	in (1.12)
10	**0.3**	**1.1**	2.3				**0.2**	**1.0**	2.3	**1.6**	
20	**0.7**	2.3	4.7				**0.5**	2.0	4.5	3.1	
100	3.3	11.3	23.3				2.5	10.0	22.5	15.6	
1000	32.6	112.7	232.6				25.0	100.0	225.0	156.2	

Table 2: A comparison of exact values with approximations for the $q = 0.50$ quantile of the cdf of the maximum of n i.i.d. H_2 random variables with mean 1 and $SCV = 4$ for four values of n and three values of r. Also displayed are exact values and approximations for np, indicating when the asymptotics should be used. The problematic values with $np < 1/q = 2.0$ are highlighted.

	exact (1.1)			asymp. (3.6)			product (4.14)			simple (1.9)	crude (1.8)
$n\backslash r$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75		
10	3.7	6.1	6.8	**1.0**	**6.1**	6.7	**-1.4**	**6.2**	6.9	6.8	3.7
20	6.5	9.2	9.0	**6.3**	9.1	9.0	**5.5**	9.7	9.2	9.5	6.4
100	18.6	16.3	14.2	18.6	16.3	14.2	21.6	17.7	14.5	16.0	12.9
1000	36.3	26.5	21.6	36.3	26.5	21.6	44.6	29.2	22.2	25.1	22.1

Table 3: A comparison of exact values with approximations for the $q = 0.75$ quantile of the cdf of the maximum of n i.i.d. H_2 random variables with mean 1 and $SCV = 4$ for four values of n and three values of r. The problematic values with $np < 1/q = 1.33$ are highlighted.

(4.18) to indicate if np is large enough and, if so, we use the approximation of $x_{(n,q)}$ in (4.14) or (1.9) to predict the quantile itself. From Table 2 we see that the simple rough approximation in (1.9) based on the SCV alone is reasonable, considering that the exact values for the three values of r vary considerably.

Turning to Tables 3 and 4, we see that the results are better for higher quantiles than for lower quantiles. In the present setting, we want to be estimating quantiles that are at least several times the mean of F, which here is 1. In Table 3 with $q = 0.75$, the approximations perform even better than in Table 2, but in Table 4 with $q = 0.25$, they perform worse. We clearly need to require n to be larger when we decrease q. Experience with results such as these led us to propose the rough guideline that $np \geq 1/q$. That is reflected in (1.9).

Given the excellent performance of the extreme-value asymptotic approximations when np

	exact (1.1)			asymp. (3.6)			product (4.14)			simple (1.9)	crude (1.8)
$n \backslash r$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75		
10	1.7	1.6	1.9	**-11.1**	**-0.9**	**1.7**	**-17.1**	**-1.6**	**1.6**	**0.5**	3.7
20	2.4	2.8	4.0	**-5.8**	**2.2**	3.9	**-10.2**	**1.8**	3.9	**3.3**	6.4
100	6.7	9.3	9.1	**6.6**	9.3	9.1	**5.9**	9.9	9.3	9.7	12.9
1000	24.2	19.5	16.5	24.2	19.5	16.5	28.9	21.4	17.0	18.9	22.1

Table 4: A comparison of exact values with approximations for the 0.25 quantile of the cdf of the maximum of n i.i.d. H_2 random variables with mean 1 and $SCV = 4$ for four values of n and three values of r. The problematic values with $np < 1/q = 4.0$ are highlighted.

is not too small, the performance of the three-parameter product approximation in (4.14) and the corresponding simple two-parameter approximation obtained by setting $r = 0.5$ or the alternative in (1.9) can be judged by evaluating the approximations for λ^{-1}. Theorem 4.1 and (4.15) provide bounds on the error, and show that performance improves as r and c^2 increase. The error bound in (4.15) shows that there is no trouble at all in estimating λ^{-1} with (4.13) if $r \geq 0.9$ or if $c^2 \geq 20$.

We conclude this section by displaying results for an H_2 distribution with much larger SCV, in particular, for $c^2 = 16$ (again with mean 1). To provide a basis for comparison with Table 2, we display the results for $q = 0.5$ in Table 5. As in Table 2, we consider four values of n: $n = 10$, 20, 100 and 1000, and three values of r: $r = 0.25$, $r = 0.50$ and $r = 0.75$. In these three cases, the (λ, p, η) triples are, respectively $(0.0315, 0.0079, 1.3228)$, $(0.0607, 0.0303, 1.9393)$ and $(0.0889, 0.0667, 3.7332)$.

As before, the extreme-value approximations are spectacular when n is large enough, but now there are more cases in which $n < n^* = 1/(pq)$. We again highlight the values for which $np < 1/q = 2.0$. There are more of these problematic values now because the higher SCV leads to a smaller value of p. We thus confirm that n needs to be larger for the extreme-value approximations to perform well when the SCV increases above 1. We remark that our criterion seems to be conservative, because the approximations are good when $n = 20$, $r = 0.75$ and $np = 1.3$, even though our criterion flags that case as potentially difficult. Observe that the results for $c^2 = 16$ in Table 5 are quite different from those for $c^2 = 4$ in Table 2, and approximation (1.9) predicts the behavior relatively well. Observe that the threshold $n^* \approx 1/pq$ is important for the exact values as well as the approximations; the exact values jump up in the region of n^*. For example, compare $n = 100$ and $n = 1000$ for $r = 0.25$. That can be explained by (4.28).

	exact (1.1)			asymp. (3.6)			product (4.14)			simple (1.9)	crude (1.8)
$n\backslash r$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75		
10	2.1	1.7	1.3	**-69.0**	**-13.6**	**-0.4**	**-76.3**	**-14.6**	**-0.5**	**-11.8**	-7.5
20	2.7	2.5	7.5	**-47.0**	**-2.2**	**7.4**	**-52.7**	**-2.8**	**7.3**	**-0.7**	3.6
100	6.0	24.4	25.5	**4.1**	24.3	25.5	**2.0**	24.6	25.6	25.1	29.3
1000	77.1	62.3	51.4	77.1	62.3	51.3	80.3	63.7	51.7	61.9	66.2
	exact np						approx. np				
$n\backslash r$	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	$n\psi(c^2)$	in (1.12)
10	**0.1**	**0.3**	**0.7**				**0.1**	**0.3**	**0.7**	**0.3**	
20	**0.2**	**0.6**	**1.3**				**0.1**	**0.6**	**1.3**	**0.7**	
100	**0.8**	3.0	6.7				**0.7**	2.9	6.6	3.3	
1000	7.9	30.3	66.7				7.4	29.4	66.2	33.2	

Table 5: A comparison of exact values with approximations for the $q = 0.50$ quantile of the cdf of the maximum of n i.i.d. H_2 random variables with mean 1 and $SCV = 16$ for four values of n and three values of r. Also displayed are exact values and approximations for np, indicating when the asymptotics should be used. The problematic values with $np > 1/q = 2.0$ are highlighted.

5. The Shifted-Exponential Distribution

In order to have a representative class of "exponential-like" distributions with $0 < c^2 < 1$, in this section we consider the shifted-exponential distribution. That is, we suppose that $Z \stackrel{\mathrm{d}}{=} d + X$, where d is a constant with $0 < d < 1$ and X is an exponential random variable with mean λ^{-1}. As before, we assume that $m_1 \equiv E[Z] = 1$, so that we have

$$1 = d + \lambda^{-1} \quad \text{and} \quad c^2 = \lambda^{-2} \ . \tag{5.1}$$

We thus let $\lambda^{-1} = c \equiv \sqrt{c^2}$.

The shape of the shifted-exponential density tends not to be too realistic, but it has a pure-exponential tail and is easy to work with. We can derive the extreme-value asymptotics either from Section 2 or from Section 3. We get the same result from both approaches. Noting that

$$\gamma = e^{\lambda d} = e^{\lambda - 1} = e^{(1-c)/c} > 1, \tag{5.2}$$

we get

$$\begin{aligned} M_n &\approx \lambda^{-1}\left[\log(n\gamma) + W\right] \\ &\approx c\left[\log(ne^{(1-c)/c}) + W\right] \\ &\approx 1 - c + c\left(\log(n) + W\right) \ . \end{aligned} \tag{5.3}$$

We thus have the approximations

$$\begin{aligned} M_n &\approx 1 - c + c(\log(n) + W) \\ E[M_n] &\approx 1 - c + c\log(n) + 0.5772c \\ Var(M_n) &\approx 1.6449c^2 \\ x_{(n,q)} &\approx \tilde{x}_{(n,q)} \equiv 1 - c + c\log(n) - c\log\log(1/q) \ . \end{aligned} \tag{5.4}$$

where $c \equiv \sqrt{c^2}$. Since the extreme-value asymptotics reduces to the exponential case, we know that the extreme-value asymptotics perform extremely well for the shifted-exponential distribution, just as in Table 1. In practice, then, the only approximation remaining is the approximation of the given distribution with $0 < c^2 < 1$ by the shifted-exponential distribution.

We use the extreme-value approximation in (5.4) as our simple rough approximation for $0 < c^2 < 1$ in (1.9). As noted before, the approximation coincides with the "naive" approximation, involving a *convex combination* of the exponential extreme-value approximation, denoted by $x^M_{(n,q)}$ and the deterministic extreme-value approximation, denoted by $x^D_{(n,q)} \equiv 1$; i.e.,

$$x_{(n,q)} = (1-c)x^D_{(n,q)} + cx^M_{(n,q)} \ , \tag{5.5}$$

with a weight c placed on the exponential term.

6. Convolutions of Exponential Distributions

Continuing to focus on distributions with $c^2 < 1$ (and $m_1 = 1$), we now consider convolutions of exponential distributions, i.e., the distribution of a sum of n independent random variables: $Z \equiv Z_1 + \cdots + Z_n$, where Z_i is exponential with mean $1/\lambda_i$, but with the restriction that all the component exponential distributions have different means. With that condition, the convolution has a pure-exponential tail. Moreover, the tail of the cdf F has a relatively simple expression:

$$F^c(x) \equiv F^c_Z(x) = \sum_{i=1}^{n} C_{i,n} e^{-\lambda_i x}, \quad x \geq 0 \ , \tag{6.1}$$

where

$$C_{i,n} = \prod_{j, j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} \ ; \tag{6.2}$$

see Section 5.2.4 of Ross (2003).

Without loss of generality, label the component random variables so that $\lambda_1 < \lambda_2 < \cdots < \lambda_n$. Then, for the extreme value theory in Section 3, we are interested in λ_1 and $\gamma = C_{1,n}$.

Since λ_1 is the smallest of all the λ_i, we see from (6.2) that $\gamma > 1$. However, the weights on other terms may be negative.

We see that the extreme-value asymptotics will produce problematic results when λ_2 is close to λ_1. As $\lambda_2 \downarrow \lambda_1$, $C_{1,n} \uparrow \infty$. For reasonable results, we assume that λ_2 is not too close to λ_1.

We also observe that the shifted exponential distribution considered in the previous section is in fact a limiting case of a convolution of exponentials. By the law of large numbers, the sum of a large number of independent exponential random variables will be approximately constant with a mean equal to the sum of the means. In particular, suppose that one exponential random variable has mean $1-d$, while the sum of the remaining $n-1$ independent exponential random variables is fixed at $d < 1$, and we let $n \to \infty$ while ensuring that each individual exponential among the $n-1$ is asymptotically negligible. Then the sum of all n exponentials approaches the shifted exponential distribution.

We now consider a convolution of exponentials with a given SCV $c^2 < 1$. We can achieve any SCV value between $1/n$ and 1 with a convolution of n exponentials; e.g., see Aldous and Shepp (1987). If we restrict attention to $n = 2$, there are only two parameters, which we can match to the mean and the SCV. We get the equations

$$\frac{1}{\lambda_1} + \frac{1}{\lambda_2} = 1 \quad \text{and} \quad \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} = c^2 \ , \tag{6.3}$$

with the constraints: $\lambda_1 < \lambda_2$ and $0.5 < c^2 < 1$. That yields a quadratic equation for λ_1^{-1} in terms of c^2.

We present numerical results for three values of c^2 - 0.8, 0.6 and 0.51 - in Tables 6-8. In the first two cases the extreme-value asymptotics are spectacular for all values of n. The simple rough approximation in (1.9) is spectacular for $c^2 = 0.8$ and quite good for $c^2 = 0.6$, with a gap appearing for very large n; e.g., there is about a 6% error for $n = 10^6$.

We include the case $c^2 = 0.51$ to deliberately include a difficult case, in which λ_2^{-1} is close to λ_1^{-1}. In this case the extreme-value approximation is not initially so good, but eventually becomes good. In fact, for small n, the simple rough approximation performs better than the extreme-value asymptotics. Overall, the simple rough approximation performs well in this case too.

To obtain values of c^2 less than 0.5, we need to consider more exponential terms. To illustrate, in Table 9 we give results for the sum of 4 independent exponential random variables with means 0.4, 0.3, 0.2 and 0.1. Here the SCV is $c^2 = 0.3$ and the mean is again 1. As before,

	$q = 0.50$			$q = 0.75$			
n	exact	asymp. (3.6)	(1.9)	exact	asymp. (3.6)	(1.9)	(1.8)
10	2.520	2.489	2.49	3.282	3.269	3.28	2.0
20	3.119	3.104	3.11	3.891	3.884	3.90	2.6
100	4.535	4.532	4.55	5.313	5.312	5.34	3.9
1000	6.575	6.575	6.61	7.355	7.355	7.40	5.7
$(10)^5$	10.661	10.661	10.73	11.441	11.441	11.52	9.4
$(10)^6$	12.704	12.704	12.79	13.485	13.484	13.48	11.2

Table 6: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. random variables, each distributed as the convolution of two exponential distributions, having overall mean 1 and $SCV = 0.8$ for six values of n. The distribution parameters are $\lambda_1^{-1} = 0.8873$, $\lambda_2^{-1} = 0.1127$, $C_{1,2} = 1.1455$ and $C_{2,2} = -0.1455$.

	$q = 0.50$			$q = 0.75$			
n	exact	asymp. (3.6)	(1.9)	exact	asymp. (3.6)	(1.9)	(1.8)
10	2.30	2.28	2.29	2.93	2.92	2.97	1.7
20	2.79	2.78	2.83	3.42	3.42	3.51	2.1
100	3.95	3.95	4.08	4.58	4.58	4.78	3.1
1000	5.61	5.61	5.86	6.25	6.25	6.54	4.4
$(10)^5$	8.94	8.94	9.43	9.58	9.58	10.11	7.2
$(10)^6$	10.61	10.61	11.21	11.25	11.25	11.89	8.6

Table 7: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. random variables, each distributed as the convolution of two exponential distributions, having overall mean 1 and $SCV = 0.6$ for six values of n. The distribution parameters are $\lambda_1^{-1} = 0.723$, $\lambda_2^{-1} = 0.277$, $C_{1,2} = 1.618$ and $C_{2,2} = -0.618$.

	$q = 0.50$			$q = 0.75$			
n	exact	asymp. (3.6)	(1.9)	exact	asymp. (3.6)	(1.9)	(1.8)
10	2.20	2.32	2.19	2.73	2.82	2.82	1.5
20	2.62	2.72	2.69	3.15	3.22	3.32	1.9
100	3.58	3.63	3.84	4.09	4.14	4.46	2.7
1000	4.92	4.95	5.48	5.43	5.43	6.11	3.9
$(10)^5$	7.57	7.58	8.77	8.07	8.08	9.40	6.2
$(10)^6$	8.89	8.89	10.41	9.39	9.39	11.04	7.4

Table 8: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. random variables, each distributed as the convolution of two exponential distributions, having overall mean 1 and $SCV = 0.51$ for six values of n. The distribution parameters are $\lambda_1^{-1} = 0.5707$, $\lambda_2^{-1} = 0.4293$, $C_{1,2} = 4.0355$ and $C_{2,2} = -3.0355$.

	$q = 0.50$			$q = 0.75$			
n	exact	asymp. (3.6)	(1.9)	exact	asymp. (3.6)	(1.9)	(1.8)
10	1.91	2.01	1.91	2.29	2.36	2.40	1.05
20	2.21	2.29	2.29	2.58	2.64	2.78	1.26
100	2.89	2.94	3.18	3.25	3.29	3.66	1.74
1000	3.84	3.86	4.44	4.19	4.21	4.92	2.43
$(10)^5$	5.69	5.70	6.96	6.05	6.05	7.44	3.82
$(10)^6$	6.62	6.62	8.22	6.97	6.97	8.70	4.51

Table 9: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. random variables, each distributed as the convolution of four exponential distributions with individual means 0.4, 0.3, 0.2 and 0.1, having overall mean 1 and $SCV = 0.3$, for six values of n. The remaining asymptotic parameter is $p = C_{1,4} = 10.667$.

the asymptotic extreme-value results are excellent, although there is about 5% error for small n. The simple rough approximation is good for smaller n, but begins to deviate for larger n. Overall, the simple rough approximation in (1.9) is reasonable though.

7. The Gamma Distribution

In this section we suppose that Z has a gamma distribution, with probability density function (pdf)

$$f(x) \equiv \frac{\lambda^\nu x^{\nu-1} e^{-\lambda x}}{\Gamma(\nu)} \;, \tag{7.1}$$

with the two parameters $\nu > 0$ and $\lambda > 0$, where Γ is the *gamma function*, with $\Gamma(k) = (k-1)!$ for k a positive integer. When $\nu = 1$, the gamma distribution reduces to the exponential distribution, but we will not consider that special case. The first two moments have a simple form: $E[Z] = \nu/\lambda$ and $c_Z^2 = 1/\nu$.

For the gamma pdf in (7.1), the associated tail probability has asymptotics

$$F^c(x) \sim \frac{\lambda^\nu x^{\nu-1} e^{-\lambda x}}{\lambda \Gamma(\nu)} \quad \text{as} \quad x \to \infty \;; \tag{7.2}$$

see p. 186 of Abate and Whitt (1997), which refers to p. 17 of Erdélyi (1956). Further analysis shows that the next term on the right in (7.2) in an asymptotic expansion has the form $Cx^{\nu-2}e^{-\lambda x}$; e.g, that is easy to see when ν is an integer greater than or equal to 2 (an Erlang distribution). In contrast, for cdf's with a pure exponential tail, the next term is typically of the form $Ce^{-\eta x}$, where $\eta > \lambda$; e.g., that is the case for any finite mixture of exponentials. Consequently, the relative error in (7.2) decays linearly instead of exponentially (as in the pure-exponential-tail case).

Even though the cdf F does not have a pure exponential tail, an appropriately scaled version of the maximum converges in distribution to a Gumbel random variable W; see p. 72 of Resnick (1987). In particular, by essentially the same argument as used to prove Theorem 3.1, we obtain

Theorem 7.1. *If*

$$F^c(x) \sim \gamma(\lambda x)^{\nu-1}e^{-\lambda x} \quad as \quad x \to \infty , \tag{7.3}$$

then

$$M_n - \frac{\log(n) + (\nu - 1)\log\log(n) - \log(\gamma)}{\lambda} \Rightarrow \frac{W}{\lambda} \quad as \quad n \to \infty . \tag{7.4}$$

However, when performing the calculation, we see that we have eliminated a complicated factor that is only asymptotically equal to 1; i.e., following the details of the proof suggests that the natural approximation stemming from Theorem 7.1 is likely to be less accurate than the corresponding approximations with the H_2 distribution.

Henceforth we focus on the gamma distribution, for which $\gamma = 1/\Gamma(\nu)$. Since $m_1 = \nu/\lambda$ and $c^2 = 1/\nu$ for the gamma distribution, $\lambda^{-1} = m_1 c^2$ and $\gamma = 1/\Gamma(1/c^2)$. We now assume $m_1 = 1$ as before. Thus, for the gamma case in terms of the single parameter c^2, we obtain the **asymptotic approximations**

$$\begin{aligned}
M_n &\approx c^2\left[\log(n) + \left(\frac{1}{c^2} - 1\right)\log\log(n) - \log(\Gamma(1/c^2)) + W\right] \\
E[M_n] &\approx c^2\left[\log(n) + \left(\frac{1}{c^2} - 1\right)\log\log(n) - \log(\Gamma(1/c^2)) + 0.5772\right] \\
Var(M_n) &\approx 1.6449c^2 \\
x_{(n,q)} &\approx c^2\left[\log(n) + \left(\frac{1}{c^2} - 1\right)\log\log(n) - \log(\Gamma(1/c^2)) - \log\log(1/q)\right] .
\end{aligned} \tag{7.5}$$

From (7.5), we can continue and eliminate asymptotically negligible terms, yielding

$$M_n \approx c^2[\log(n) + W] \approx c^2\log(n) , \tag{7.6}$$

as in (1.8), but unless n is large, those deleted terms are actually not negligible, as we show in Table 10, where we present a breakdown of the contributions to the approximation for the mean $E[M_n]$ in (7.5). In particular, the *loglog* terms are the same order as the *log* terms.

Approximations for the gamma function. To further simplify the asymptotic formulas in (7.5), we can approximate $\log(\Gamma(\nu))$. There is a large body of literature on the gamma function including approximations, many related to Stirling's formula; e.g., see Chapter 6 of Abramowitz and Stegun (1972). We observe that

$$\log(\Gamma(\nu)) \sim -\log(\nu) \quad \text{as} \quad \nu \downarrow 0 \quad \text{and} \quad \log(\Gamma(\nu)) \sim \nu \log(\nu) \quad \text{as} \quad \nu \uparrow \infty . \tag{7.7}$$

We propose the following simple rational approximation for $\log(\Gamma(\nu))$, which is exact at $\nu = 1$, $\nu = 2$, as $\nu \to 0$ and as $\nu \to \infty$:

$$\log(\Gamma(\nu)) \approx \phi(\nu) \equiv \left(\frac{\nu^2 - 4}{\nu + 4}\right) \log(\nu) . \tag{7.8}$$

The ratio $\phi(\nu)/\log(\Gamma(\nu))$ falls between 0.99 and 1.04 on the interval $(0, 1)$, assuming its largest value 1.04 as $\nu \uparrow 1$, where both terms are 0. The ratio rises up to a maximum of 1.26 at around $\nu = 23$ and then declines to 1 as ν increases further. Approximation 6.1.41 of Abramowtiz and Stegun (1972) performs well for $c^2 < 1$. However, we do not need approximations for numerics, because the gamma function can be calculated.

For $c^2 < 1$, we combine (7.5) and (7.7) to get an asymptotic approximation for small c^2 and large n, where

$$\beta \equiv c^2 \log(n) \geq 1 . \tag{7.9}$$

Let F have a gamma distribution with mean $m_1 = 1$ and SCV c^2. Note that

$$\begin{aligned} x_{(n,q)} &\sim c^2 \left[\log(n) + \left(\frac{1}{c^2} - 1\right) \log\log(n) - \log(\Gamma(1/c^2)) - \log\log(1/q)\right] \quad \text{as} \quad n \to \infty \\ &\approx c^2 \left[\log(n) + \left(\frac{1}{c^2}\right) \log\log(n) - \frac{1}{c^2} \log(1/c^2) - \log\log(1/q)\right] \quad \text{for small} \quad c^2 \\ &\approx \beta + \log(\beta) - c^2 \log\log(1/q) \quad \text{under (7.9)} . \end{aligned} \tag{7.10}$$

The final line in (7.10) provides an alternative to the second row in approximation (1.9), but notice that it requires that $\log(n) \geq \frac{1}{c^2}$. Because we use the asymptotic form of the gamma

ν	n	c^2	$\log(n)$	$(\nu - 1)\log\log(n)$	$-\log\Gamma(\nu)$	$E[W]$	$E[M_n]$
	10^2	0.50	4.6	1.5	0.0	0.6	3.2
2	10^4	0.50	9.2	2.2	0.0	0.6	5.8
	10^6	0.50	13.8	2.6	0.0	0.6	8.3
	10^2	0.25	4.6	4.6	-1.8	0.6	2.0
4	10^4	0.25	9.2	6.7	-1.8	0.6	3.7
	10^6	0.25	13.8	7.9	-1.8	0.6	5.1

Table 10: A breakdown of the contributions to the approximation of the mean $E[M_n]$ in (7.5) for the case of gamma random variables with mean 1, for three values of n and two values of the shape parameter $\nu = 1/c^2$.

	$q = 0.50$				$q = 0.75$				
n	exact	asymp.	(1.9)	H_2 ex.	exact	asymp.	(1.9)	H_2 ex.	(1.8)
10	4.0	3.0	**3.3**	**2.8**	6.5	6.5	6.8	6.1	3.7
20	5.9	5.0	6.0	5.3	8.6	8.5	9.5	9.2	6.4
100	10.9	10.2	12.4	12.4	13.9	13.7	16.0	16.3	12.9
1000	18.8	18.1	21.7	22.6	21.9	21.7	25.2	26.5	22.1
$(10)^5$	35.5	35.0	40.1	43.0	38.8	38.6	43.6	46.9	40.5
$(10)^8$	61.6	61.3	67.7	73.7	65.0	64.8	71.2	77.6	68.1

Table 11: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. gamma random variables with mean 1 and $SCV = 4$ for six values of n. The approximations are the asymptotic approximation in (7.5), the associated simple rough approximation in (1.9), the exact values for H_2 with $r = 0.5$, and the crude approximation in (1.8). The problematic values with $np > 1/q = 2.0$ in the H_2 framework with $r = 0.5$ are highlighted.

function as $c^2 \downarrow 0$, we anticipate that approximation (7.10) will perform best when c^2 is quite small. But notice that n must grow rapidly as c^2 decreases in order to have $\beta = c^2 \log(n) > 1$, which is needed to have $x_{(n,q)} \geq 1 = m_1$.

Numerics for the gamma distribution. We now report numerical comparisons between approximations and exact values for the gamma distribution. First, paralleling Tables 2-4, we give results for $c^2 = 4$ in Table 11. As we anticipated, the asymptotic extreme-value approximations are not as accurate here for this case of a cdf without a pure-exponential tail. There is an error of about $5 - 15\%$ when $n = 100$ and $n = 1000$ for $q = 0.5$. We display results for larger n to show that we have not made a mistake; convergence is evident, even though the error remains at about 1% for $r = 0.25$ at $n = 10^8$. The simple rough approximation in (1.9) consistently overestimates the exact value, but it too is reasonable, performing as well as the exact values for H_2 with $r = 0.5$. For the simple rough approximation in (1.9) and the H_2 approximation with $r = 0.5$ we can estimate when n is large enough by seeing if $np \geq 1/q = 2.0$. The problematic values with $np < 1/q = 2.0$ are highlighted in Table 11.

In Tables 12 and 13 we turn to the case of $c^2 < 1$, considering $c^2 = 1/4 = 0.25$ and $c^2 = 1/16 = 0.0625$. Here the asymptotic values do not approximate the exact values very well, having a relatively greater error than in Table 11. The situation is extremely bad for $c^2 = 1/16$. Even for $c^2 = 0.25$, the asymptotic extreme-value approximation still have about 10% error when $n = 10^5$ and 2.5% error when $n = 10^{12}$. The simple rough approximation in (1.9) based on the shifted-exponential distribution performs much better than the full asymptotic approximation itself for smaller n, e.g., for $n \leq 100$, but performs much worse when n gets

	$q = 0.50$				$q = 0.75$				
n	exact	asymp.	(7.10)	(1.9)	exact	asymp.	(7.10)	(1.9)	(1.8)
10	1.83	0.84	0.12	1.83	2.15	1.06	0.33	2.27	0.92
20	2.08	1.22	0.55	2.18	2.39	1.44	0.77	2.62	1.10
100	2.64	1.94	1.38	2.99	2.93	2.16	1.60	3.43	1.50
1000	3.38	2.82	2.36	4.14	3.66	3.04	2.58	4.58	2.07
$(10)^5$	4.77	4.35	4.03	6.44	5.03	4.57	4.25	6.88	3.22
$(10)^8$	6.75	6.43	6.22	9.89	6.99	6.65	6.44	10.33	4.95
$(10)^{10}$	8.03	7.75	7.60	12.20	8.27	7.97	7.82	12.64	6.10
$(10)^{12}$	9.28	9.04	8.93	14.50	9.52	9.26	9.15	14.94	7.25

Table 12: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. gamma random variables with mean 1 and $SCV = 0.25$ for eight values of n. The approximations are the asymptotic approximation in (7.5), the associated simple approximation (7.10), the simple rough approximation based on the shifted-exponential in (1.9) and the crude approximation in (1.8).

very large. When n is very large, the simple approximation in (7.10) closely agrees with the asymptotic approximation in (7.5). Thus one might use (1.9) for $n \leq 1000$.

8. Reverse Engineering

We now show how we can obtain a rough estimate of the first two moments of the underlying cdf F and the full cdf F itself, given knowledge of the distribution of the maximum for a few values of n. To do so, we make the assumption that the cdf has a pure-exponential tail. We thus apply the extreme-value approximations in (3.6). Given $x_{(n,q)}$ for known q and at least two values of n, we can estimate the parameters λ^{-1} and γ, assuming the approximation for $x_{(n,q)}$ in (3.6) holds as an equality.

Then given λ^{-1} and γ, we estimate the first two moments of the cdf F using our simple rough approximation in (1.9). We denote the estimates of the mean m_1 and the SCV c^2 by $\hat{m}_1$ and $\hat{c}^2$. Having estimated the mean and SCV, we fit an exponential distribution if $\hat{c}^2 \approx 1$, an H_2 distribution if $\hat{c}^2 > 1$, and a shifted exponential distribution or a convolution of exponentials if $\hat{c}^2 < 1$.

Since we no longer can assume the mean is 1, we need to determine the mean now. In this reverse direction, we first determine the SCV c^2 from the asymptotic constant γ. Exploiting the relation

$$\gamma \approx \psi(c^2) , \tag{8.1}$$

where ψ is given in (1.12), we let

$$\hat{c}^2 = \psi^{-1}(\gamma) , \tag{8.2}$$

	$q = 0.50$				$q = 0.75$				
n	exact	asymp.	(7.10)	(1.9)	exact	asymp.	(7.10)	(1.9)	(1.8)
10	1.40	-0.79	-1.77	1.42	1.53	-0.74	-1.72	1.64	0.32
20	1.50	-0.50	-1.47	1.59	1.62	-0.45	-1.41	1.81	0.36
100	1.72	-0.001	-0.93	1.99	1.83	0.05	-0.88	2.21	0.46
1000	1.99	0.52	-0.39	2.57	2.09	0.58	-0.33	2.79	0.61
$(10)^5$	2.48	1.29	0.41	3.72	2.56	1.34	0.47	3.94	0.89
$(10)^8$	3.12	2.16	1.32	5.45	3.19	2.22	1.37	5.67	1.32
$(10)^{10}$	3.51	2.66	1.83	6.60	3.59	2.71	1.88	6.82	1.61
$(10)^{12}$	3.90	3.12	2.30	7.75	3.97	3.17	2.35	7.97	1.90

Table 13: A comparison of exact values with approximations for two quantiles ($q = 0.50$ and 0.75) of the cdf of the maximum of n i.i.d. gamma random variables with mean 1 and $SCV = 1/16 = 0.0625$ for five values of n. The approximations are the asymptotic approximation in (7.5), the associated simple approximation (7.10), the simple rough approximation based on the shifted-exponential in (1.9) and the crude approximation in (1.8).

using the fact that ψ is a strictly-decreasing continuous function of c^2 with $\psi(1) = 1$. We plot the function ψ in log-log scale in Figure 1. As noted in (1.12), $\gamma \approx 1/c^2$ when $c^2 \geq 1$ and $\gamma \approx 1/c$ when $c^2 \leq 1$.

Now, given $\hat{c}^2$, we estimate the mean using (1.9). In particular, we let

$$\hat{m}_1 \hat{c}^2 = \lambda^{-1} \quad \text{and} \quad \hat{m}_1 = \frac{1}{\hat{c}^2 \lambda} \quad \text{if} \quad \hat{c}^2 > 1 \tag{8.3}$$

and

$$\hat{m}_1 \hat{c} = \lambda^{-1} \quad \text{and} \quad \hat{m}_1 = \frac{1}{\hat{c} \lambda} \quad \text{if} \quad \hat{c}^2 \leq 1 \; . \tag{8.4}$$

Given $\hat{m}_1$ and $\hat{c}^2 > 1$, we estimate a full H_2 cdf by applying (4.17) and letting the third parameter be

$$\hat{r} = \frac{\hat{c}^2 + 1}{2\hat{c}^2} \; . \tag{8.5}$$

Given $\hat{m}_1$ and $\hat{c}^2 \leq 1$, we estimate a full shifted-exponential cdf by directly using λ and letting

$$\hat{d} = \hat{m}_1 - \lambda^{-1} \; . \tag{8.6}$$

Alternatively, if $0.5 < c^2 < 1$, we can fit a convolution of two exponentials with estimated means $\hat{\lambda}_1^{-1}$ and $\hat{\lambda}_2^{-1}$ by solving the pair of equations

$$\frac{1}{\hat{\lambda}_1} + \frac{1}{\hat{\lambda}_2} = \hat{m}_1 \quad \text{and} \quad \frac{1}{\hat{\lambda}_1^2} + \frac{1}{\hat{\lambda}_2^2} = \hat{m}_1^2 \hat{c}^2 \; , \tag{8.7}$$

with the constraints: $\lambda_1 < \lambda_2$ and $0.5 < c^2 < 1$.

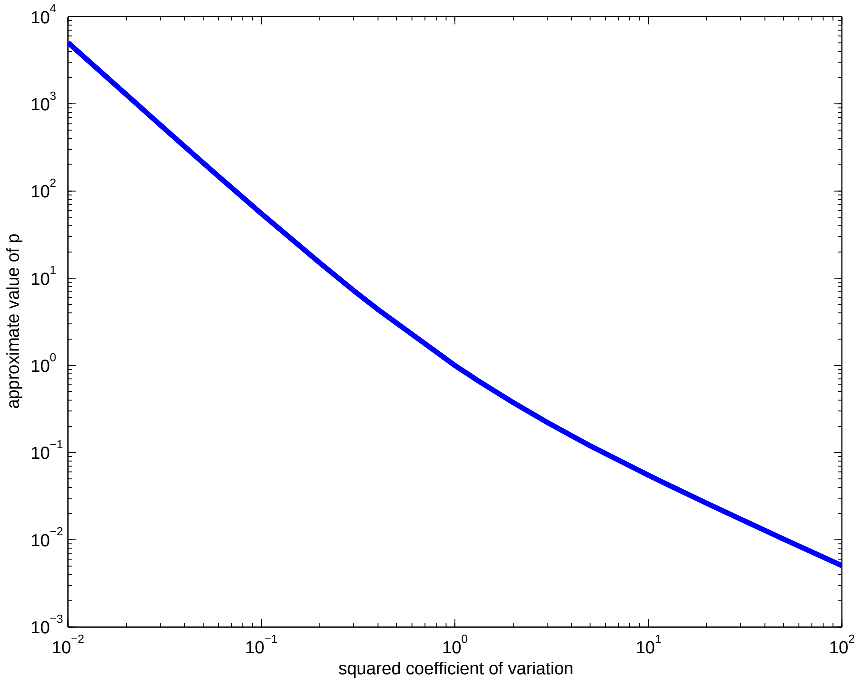

Figure 1: The approximation function ψ in (1.12), mapping c^2 into p for distributions with pure-exponential tail.

9. Conclusions

Unlike the normal-distribution approximation for sums of n i.i.d. random variables, based on the central limit theorem, the associated Gumbel-distribution approximation for the maximum of n i.i.d. random variables, based on the extreme-value theorems, depends on the underlying cdf F beyond its first two moments. First, we need to assume regularity conditions such as (1.4) assumed here and, second, we need to approximate the key parameters λ, α and γ appearing there. But, fortunately, unlike for sums of i.i.d. random variables, the exact distribution of the maximum of n i.i.d. random variables is easy to compute directly for any n and any cdf F, as indicated in the introduction. Thus, for numerical calculations, we propose fitting representative distributions to the first few moments and calculating the desired characteristics of the distribution of M_n for that representative distribution.

Nevertheless, we have focused on closed-form approximations, which have the advantage of directly providing insight. They also can be used for further analysis within other models. As in Abate and Whitt (1997), we find that the asymptotic approximations take a simpler form and perform better when the cdf F has a pure-exponential tail. The story in Sections 2-6 for cdf's with a pure-exponential tail is much better than for the gamma distribution in Section 7. When the cdf F has a pure-exponential tail, the extreme-value approximation in (3.6) usually performs spectacularly well unless n is too small.

There are difficulties when F does not have a pure-exponential tail, especially when c^2 is small. For $n \leq 1000$, the asymptotic extreme-value approximation performs poorly for $c^2 = 0.25$ and is totally useless for $c^2 = 0.0625$, as shown in Tables 12 and 13. In contrast, the simple rough approximation in (1.9) based on the shifted-exponential distribution performs well for $n \leq 1000$, even though it does not have the correct asymptotic behavior.

An important idea introduced here is that there is a threshold $n^* \equiv n^*(F, q)$, depending on the underlying cdf F and the target quantile q in $x_{(n,q)}$: It is necessary to have $n \geq n^*$ before the extreme-value approximations become useful. In addition to evaluating the performance of the extreme-value approximations and developing approximations for the key asymptotic parameters appearing there, we have developed approximations for the threshold n^*. For $c^2 \geq 1$, we suggest the threshold $n^* \equiv n^*(c^2, q) \approx c^2/q$. In particular, we observed that n needs to be larger as c^2 increases above 1 and as q decreases. Insight into this relation was given in Section 4. For the case of a pure-exponential tail, in Section 8 we also considered how to do reverse engineering to obtain an estimate of the underlying cdf F given known behavior of the maximum for two or more values of n.

The crude approximation in (1.8) shows the main tendency, but is not accurate when $c^2 < 1$. The proposed approximations in (1.9), (4.14), (4.32), (5.4), (7.5) and (7.10) can be viewed as refinements of (1.8). When F has a pure-exponential tail, there are two refinements: (i) finding a better approximation for the multiplier λ^{-1} than c^2 and (ii) finding an appropriate function of c^2 to include with n inside the logarithm.

For the practically important range $10 \leq n \leq 1000$, the simple rough approximation in (1.9) based on the first two moments of F seems to be satisfactory throughout when $c^2 \leq 1$. But it is important to note that the first two moments do not pin down the asymptotic parameters exceptionally well. The numerical results in this paper show the limitations of working with only that partial information.

When c^2 is not large and F is indeed only partially characterized by its first two moments, approximation (1.9) should do as well as exact calculations for the representative distributions, because the closed-form approximation formulas tend to be closer to the exact values for the representative distributions (on which they are based) than the exact values for the representative distributions are to the exact values for other distributions. That is illustrated here in the numerical examples, e.g., for the case $c^2 = 4.0$ by considering examples when F has a gamma distribution and an H_2 distribution with $r = 0.25$, $r = 0.5$ and $r = 0.75$.

The threshold n^* for n in order for the extreme-value approximations to be accurate increases in c^2 for $c^2 \geq 1$. The approximation for n^* in (1.13) seems to be relatively accurate. Tables 5 and 13 for $c^2 = 16$ and $c^2 = 1/16$ show that n has to be quite large before the extreme-value-based approximations are useful when c^2 is either very large or very small. For high values of c^2 and for moderate n, e.g., for $n \leq 1000$, it is better to use the exact distribution based on (1.1) for representative distributions than it is to use the extreme-value approximation, even given the asymptotic parameters in (1.4). In particular, we can use exact numerical values based on H_2 distributions, preferably based on the first three moments; see Table 5.

When c^2 is large, it becomes more important to have an additional parameter. Of course, the asymptotic parameters in (1.4) would be preferred, but in lieu of that, we suggest three-moment approximations using the H_2 distribution. Given the first three moments (assuming that $c^2 > 1$), we fit an H_2 distribution by applying (4.4). For understanding, we can then calculate the associated parameter r in (4.7). With r, the product approximation in (4.13) and (4.14) is highly accurate for H_2 when c^2 is large, as shown by (4.10) and (4.15).

When the SCV c^2 gets very large, we should also be concerned that the distribution may actually not satisfy the regularity condition (1.4), and instead have a heavy tail as in (1.5), which will yield much larger maxima, growing as $n^{1/\alpha}$ instead of $\log(n)$. In particular, instead of (1.6), we would then have

$$x_{(n,q)} \approx \left(\frac{\gamma n}{-\log(q)}\right)^{1/\alpha} ; \tag{9.1}$$

see Chapter 3 of Embrechts et al. (1997). With a heavy tail, there is less predictability: The spread, as measured by $x_{(n,q_2)} - x_{(n,q_1)}$, grows like $n^{1/\alpha}$, just like $x_{(n,q)}$. There is no relative concentration in the limit as $n \to \infty$ with a heavy tail.

References

[1] Abate, J., W. Whitt. 1997. Asymptotics for $M/G/1$ low-priority waiting-time tail probabilities. *Queueing Systems* 25, 173–233.

[2] Abramowitz, M., I. A. Stegun. 1972. *Handbook of Mathematical Functions*, National Bureau of Standards.

[3] Aldous, D. A., L. A. Shepp. 1987. The least variable phase type distribution is Erlang. *Stochastic Models* 3, 467–473.

[4] Bitran, G. R., S. Dasu. 1992. A review of open queueing network models of manufacturing systems. *Queueing Systems* 12, 95–134.

[5] Buzacott, J. A., J. G. Shanthikumar. 1993. *Stochastic Models of Manufacturing Systems*, Prentice-Hall.

[6] Castillo, E. 1988. *Extreme Value Theory in Engineering*, Academic Press.

[7] Crow, C. S., IV, D. Goldberg, W. Whitt. 2006a. Two-moment approximations for maxima: supplement. Available at http://www.columbia.edu/∼ww2040/recent.html

[8] Crow, C. S., IV, D. Goldberg, W. Whitt. 2006b. Congestion caused by inspection. In preparation.

[9] Embrechts, P., C. Klüppelberg, T. Mikosch. 1997. *Modelling Extremal Events*, Springer, New York.

[10] Erdélyi, A. 1956. *Asymptotic Expansions*, Dover.

[11] Feller, W. 1971. *An Introduction to Probability Theory and its Applications*, second edition, Wiley.

[12] Galambos, J. 1987. *Asymptotic Theory of Extreme Order Statistics*, second edition, Krieger, Malabar, Fl.

[13] Goldberg, D., W. Whitt. 2006. The last departure time from an $M_t/GI/\infty$ queue with a terminating arrival process. Columbia University. Submitted for publication. Available at http://www.columbia.edu/∼ww2040/recent.html

[14] Johnson, N. L., S. Kotz. 1970. *Continuous Univariate Distributions - I* Wiley, New York.

[15] Kotz, S., S. Nadarajah. 2000. *Extreme Value Distributions*, Imperial College Press.

[16] Resnick, S. I. 1987. *Extreme Values, Regular Variation and Point Processes*, Springer, New York.

[17] Ross, S. M. 2003. *Introduction to Probability Models*, eighth edition, Academic Press.

[18] Suri, R., J. L. Sanders, M. Kamath. 1993. Performance evaluation of production networks. In *Handbooks in Operations Research and Management Science, Vol. 4: Logistics of Production and Inventory*, S. C. Graves, A. H. G. Rinnooy Kan, and P. H. Zipkin (eds.), Elsevier, Amsterdam, 199–286.

[19] Thomas, M., R. D. Reiss. 2001. *Statistical Analysis of Extreme Values*, Birkhauser.

[20] Weisstein, E. W. 2005. Harmonic number. In *Math World – A Wolfram Web Resource*, Available at: http://mathworld.wolfram.com.

[21] Whitt, W. 1982. Approximating a point process by a renewal process, I: two basic methods. *Operations Research* 30, 125–147.

[22] Whitt, W. 1983. The queueing network analyzer. *Bell System Tech. J.* 62, 2779–2815.

[23] Whitt, W. 1984. On approximations for queues, III: mixtures of exponential distributions. *AT&T Bell Lab. Tech. J.* 63, 163-175.