The Annals of Probability
1990, Vol. 18, No. 3, 1388–1402

ERROR BOUNDS FOR EXPONENTIAL APPROXIMATIONS OF GEOMETRIC CONVOLUTIONS[1]

BY MARK BROWN

The City College, CUNY

Define Y_0 to be a geometric convolution of X if Y_0 is the sum of N_0 i.i.d. random variables distributed as X, where N_0 is geometrically distributed and independent of X. It is known that if X is nonnegative with finite second moment, then as $p \to 0$, Y_0/EY_0 converges in distribution to an exponential distribution with mean 1. We derive an upper bound for $d(Y_0)$, the sup norm distance between Y_0 and an exponential with mean EY_0. This upper bound is $d(Y_0) \le cp$ for $0 < p \le \frac{1}{2}$, where $c = EX^2/(EX)^2$. It is asymptotically ($p \to 0$) tight. Also derived is a bound for $d(Y_0 + Z)$, where Z is independent of Y_0.

1. Introduction. If $\{X_i,\ i \ge 1\}$ is an i.i.d. sequence of nonnegative random variables and N_0 is geometrically distributed [$\Pr(N_0 = k) = q^k p$, $k = 0, 1, 2, \dots$] and independent of $\{X_i\}$, then $Y_0 = \sum_1^{N_0} X_i$ is called a geometric convolution of X. Closely related to Y_0 is the random variable $Y = \sum_1^N X_i$, where $N = N_0 + 1$. Y is also referred to as a geometric convolution.

Geometric convolutions arise naturally in many applied probability models. Gertsbakh (1984) discusses a rich variety of applications in reliability and queues and surveys research in the area, most of which was performed by Soviet authors. Feller (1971), Section XI.6, elegantly discusses terminating renewal processes, the time until termination being a geometric convolution. Several authors have studied random sampling, or "thinning," of renewal processes; this procedure results in new renewal processes with geometric convolution interarrival times. Jacobs (1986) investigates a geometric convolution in the context of combining random loads and waiting for the stress to exceed a given level. In a $G|G|1$ queue in equilibrium, the waiting time distribution is a geometric convolution and has been studied in this context by Szekli (1986) and Köllerström (1976). Finally, numerous applications arise in regenerative stochastic processes. Consider a regenerative process [Smith (1958)] where in each cycle an event A may or may not occur, independently of other cycles. The waiting time for A to occur is then of the form

$$W = \sum_1^{N_0} X_i + Z = Y_0 + Z, \tag{1.1}$$

Received April 1988; revised March 1989.
[1]Research supported by the Air Force Office of Scientific Research under Grant 84-0095 with City College, CUNY.
AMS 1980 *subject classifications*. Primary 60F05; secondary 60E15.
Key words and phrases. Geometric convolutions, exponential approximations, reliability theory, queuing theory, renewal theory, regenerative processes, NWU distributions.

where A occurs for the first time in cycle $N_0 + 1$, N_0 is geometrically distributed with parameter equal to the probability of A occurring during a specified cycle and Z is the waiting time from the beginning of cycle $N_0 + 1$ until A occurs.

Keilson (1966) recognized the prevalence of (1.1) and considered the case of small p. He showed that if the X_i are nonnegative with finite second moment, then W/EW converges in distribution to an exponential with mean 1 as $p \to 0$. Thus, the waiting time for a rare event (small p) to occur is approximately exponential. Solovyev (1971) considered a sequence of random variables of the form (1.1) in which the distribution of X varies with p and obtained conditions for asymptotic exponentiality as $p \to 0$. Solovyev also obtained error bounds for the exponential approximation.

In this paper we seek to bound the distance between a geometric convolution of nonnegative random variables and an exponential distribution with the same mean. This problem is cited by Gertsbakh (1984) as being "of great interest for engineering applications." Defining X, Y and Y_0 as above, $q = 1 - p$, $\mu = EX$, $\mu_2 = EX^2$, $\gamma = \mu_2/2\mu^2$ and $\bar{F}_{Y_0}(t) = \Pr(Y_0 > t)$, we derive

$$q \exp\left(-\frac{p}{q}\left(\frac{t}{\mu} + 2\gamma - 1\right)\right) \le \bar{F}_{Y_0}(t) \le \exp\left(-\frac{pt}{q\mu}\right) + \frac{\gamma p}{q}. \tag{1.2}$$

Defining $d(Y_0)$ as the sup norm distance between Y_0 and an exponential distribution with mean $EY_0 = q\mu/p$, it follows from (1.2) that for $0 < p \le \frac{1}{2}$,

$$d(Y_0) \le 2\gamma p = \mu_2 p/\mu^2. \tag{1.3}$$

For a given p, let $B(p, \mu, \mu_2)$ denote the best upper bound for $d(Y_0)$ among all distributions with common first two moments μ and μ_2. It is shown in Section 4 that $\lim_{p \to 0}[B(p, \mu, \mu_2)/p] = 2\gamma$. Thus the bound in (1.3) is asymptotically sharp as $p \to 0$.

Bounds are also obtained for $d(Y_0 + Z)$, where Z is a nonnegative random variable independent of Y_0, for $d(Y)$ and for $d(Y^*)$, where Y^* is the stationary renewal distribution corresponding to Y. It is easy to see that Y^* is also the stationary distribution corresponding to Y_0.

The current bounds offer improvement over those of Solovyev (1971) in that they are derived under less restrictive conditions, require less information about X and Z to compute and are in general tighter. This comparison is discussed in Section 4.

In the above X is assumed nonnegative with known first two finite moments. In Section 3 we restrict X further and obtain improved bounds. Perhaps the most interesting of these results is that if X is assumed NBUE (new better than used in expectation, defined in Section 2), then $d(Y_0)$ is exactly equal to p.

Our methodology is a combination of reliability and renewal theory geared to exploit the fact that Y_0 is NWU (new worse than used) as pointed out by

Daley and Trengove (1977). The technique of studying random variables through their aging properties was developed by several authors, most notably by Barlow, Marshall and Proschan, and is lucidly presented in the text of Barlow and Proschan (1975).

2. Bounds for general X.

A distribution F on $[0,\infty)$ is defined to be NWU if

$$\bar{F}(t+x) \geq \bar{F}(t)\bar{F}(x) \quad \text{for all } t, x \geq 0. \tag{2.1}$$

Similarly, NBU (new better than used) is defined by reversing the inequality in (2.1). Thus F is NWU (NBU) if its survival distribution at age 0 is stochastically smaller (larger) than its survival distribution at age x for all $x > 0$ with $\bar{F}(x) > 0$.

Let $\{X_i,\ i \geq 0\}$ be an i.i.d. sequence of nonnegative random variables; let N_0 be independent of this sequence with $\Pr(N_0 = k) = q^k p,\ k = 0, 1, \dots$ and define $S_n = \Sigma_1^n X_i$ and $Y_0 = S_{N_0}$. The following simple but useful result is discussed in Stoyan (1983), page 96. Stoyan credits Köllerström (1976) with deriving the NWU result for the stationary waiting time distribution in a $G|G|1$ queue and Daley and Trengove (1977) with noticing that it is a consequence of the waiting time being representable as a geometric convolution.

LEMMA 2.1. *Y_0 is NWU.*

PROOF. For $t > 0$, define $M_t = \min\{k\colon S_k > t\}$. Since M_t is independent of N_0, it follows from the lack of memory property of the geometric distribution that $(N_0 - M_t | N_0 \geq M_t) \sim N_0$. Furthermore since M_t is a stopping time, $\{X_{M_t+i},\ i \geq 1\} \sim \{X_i,\ i \geq 1\}$. Thus

$$\left(\sum_{M_t+1}^{N_0} X_i \Big| N_0 \geq M_t \right) \sim \sum_{1}^{N_0} X_i = Y_0. \tag{2.2}$$

Note that the events $\{Y_0 > t\}$ and $\{N_0 \geq M_t\}$ are equivalent. Thus their indicator functions are equal, thus $I_{Y_0 > t} = I_{N_0 \geq M_t}$. Now

$$\begin{aligned} (Y_0 - t) I_{Y_0 > t} &= \left[\left\{ \left(\sum_{1}^{M_t} X_i \right) - t \right\} + \sum_{M_t+1}^{N_0} X_i \right] I_{Y_0 > t} \\ &\geq \left(\sum_{M_t+1}^{N_0} X_i \right) I_{N_0 \geq M_t}. \end{aligned} \tag{2.3}$$

Next note that for $x \geq 0$,

$$\begin{aligned}\Pr(Y_0 > t + x) &= \Pr\big((Y_0 - t)I_{Y_0 > t} > x\big) \\ &\geq \Pr\left(\left(\sum_{M_t+1}^{N_0} X_i\right) I_{N_0 \geq M_t} > x\right) \quad \text{(by 2.3)} \\ &= \Pr\left(\sum_{M_t+1}^{N_0} X_i > x,\ N_0 \geq M_t\right) \\ &= \Pr\left(\sum_{M_t+1}^{N_0} X_i > x | N_0 \geq M_t\right) \Pr(N_0 \geq M_t) \\ &= \Pr(Y_0 > x)\Pr(N_0 > M_t) \quad \text{(by 2.2)} \\ &= \Pr(Y_0 > x)\Pr(Y_0 > t).\end{aligned}$$

Thus Y_0 is NWU. □

In this paper the renewal function corresponding to a distribution F on $[0, \infty)$ will be defined by $M(t) = \sum_{k=0}^{\infty} F^{(k)}(t)$ as in Feller (1971), Chapter XI. Some authors use $\sum_{k=1}^{\infty} F^{(k)}(t) = M(t) - 1$ for the renewal function.

Lemma 2.2 is a slight generalization of a known result [Barlow and Proschan (1975), page 162], the generalization allowing for an atom at zero which will be required for Y_0.

LEMMA 2.2. *Assume that W is NWU with an atom of size p at zero, but with no other atoms. Let F be the cdf of W and $M = \sum_0^{\infty} F^{(k)}$ the renewal function. Then*

$$\overline{F}(t) \geq q \exp(pq^{-1})\exp(-(M(t) - 1)) \geq \exp(-(M(t) - 1)).$$

PROOF. Let $\{N_1(t) - N_1(0),\ t > 0\}$ be a nonhomogeneous Poisson process with $E(N_1(t) - N_1(0)) = -\ln(\overline{F}(t)/q)$. This process has its first event epoch T_1 distributed as $W|W > 0$, its next interarrival time $T_2 - T_1$ distributed as $W - T_1|W > T_1$ and in general its kth interarrival time $T_k - T_{k-1}$ distributed as $W - T_{k-1}|W > T_{k-1}$. Since W is NWU,

(2.4) $$T_k - T_{k-1}|T_1 \cdots T_{k-1} \geq_{\text{st}} W \quad \text{for } k \geq 2.$$

Next, consider $\{N_2(t) - N_2(0),\ t \geq 0\}$, where N_2 is a renewal process with interarrival time W. The first event epoch of $N_2(t) - N_2(0)$ occurs at $S_1 \sim W|W > 0 \sim T_1$. Subsequent interarrival times are distributed as W. It follows from $S_1 \sim T_1$ and (2.4) that

(2.5) $$T_k \geq_{\text{st}} S_k \quad \text{for } k \geq 1.$$

Thus from (2.5),

(2.6) $$N_1(t) - N_1(0) \leq_{\text{st}} N_2(t) - N_2(0).$$

Taking expectations in (2.6),

$$-\ln(\bar{F}(t)/q) \le M(t) - q^{-1}. \tag{2.7}$$

Finally, the result follows from (2.7) and the observation that $q\exp(pq^{-1}) \ge q(1+qp^{-1}) = 1$. □

COROLLARY 2.1. *Suppose that $X \ge 0$ has $F(0) < 1$ and $\mu_2 = EX^2 < \infty$. Then*

$$\bar{F}_{Y_0}(t) \ge q\exp(-pq^{-1}(t\mu^{-1} + 2\gamma - 1)) = q\exp(-pq^{-1}(2\gamma - 1))\exp(-t/EY_0),$$

where $\gamma = \mu_2/2\mu^2$.

PROOF. CASE (i). We first consider the case in which X has no atoms. Then Y_0 has an atom of size p at zero and no other atoms. Define N_{Y_0} and N_X as renewal processes with interarrival times Y_0 and X, respectively, and $M_{Y_0} = EN_{Y_0}$, $M_X = EN_X$.

Note that N_{Y_0} has $1 + G_q$ renewals at zero and G_q renewals at each renewal epoch of N_X in $(0, \infty)$, where $\Pr(G_q = k) = p^k q$, $k = 0, 1, \dots$.

Thus,

$$M_{Y_0}(t) = q^{-1} + pq^{-1}(M_X(t) - 1) = 1 + pq^{-1}M_X(t). \tag{2.8}$$

Alternatively, (2.8) can be easily proved using Laplace transforms. Next, we note Lorden's (1970) upper bound for the renewal function,

$$M_X(t) \le \frac{t}{\mu} + 2\gamma. \tag{2.9}$$

The result now follows from Lemma 2.2, (2.8), (2.9) and $EY_0 = p^{-1}q\mu$.

CASE (ii). Now consider the general case. Define e_n to be uniformly distributed on $(0, \varepsilon_n)$, $n = 1, 2, \dots,$ with $\lim \varepsilon_n = 0$, e_n independent of X and $X_n = X + e_n$. Since X_n converges to X in quadratic mean, $\mu_n = EX_n \to \mu$ and $\gamma_n = EX_n^2/2\mu_n^2 \to \gamma$. Define $Y_{0,n} = \sum_1^{G_p} X_{n,i}$, the analogue of Y_0 with X replaced by X_n. Then by choosing $X_{n,i} = X_i + e_{n,i}$, we have $E(Y_{0,n} - Y_0)^2 = E(\sum_1^{G_p} e_{n,i})^2 \to 0$ as $n \to \infty$. Thus $Y_{0,n}$ converges in quadratic mean and thus in distribution to Y_0. By Case (i),

$$\bar{F}_{Y_{0,n}}(t) \ge q\exp(-pq^{-1}(t\mu_n^{-1} + 2\gamma_n - 1)). \tag{2.10}$$

It follows by letting $n \to \infty$ in (2.10) that the desired bound for $\bar{F}_{Y_0}(t)$ holds at all continuity points of Y_0. But since $\bar{F}_{Y_0}$ is right continuous and the bound is a continuous function of t, it follows that the inequality must hold for all $t \ge 0$.

A nonnegative random variable X with distribution F is defined to be NWUE (new worse than used in expectation) if $0 < \mu = EX < \infty$ and

$$E(X - t \mid X > t) \ge \mu \quad \text{for all } t \ge 0 \text{ with } \bar{F}(t) > 0. \tag{2.11}$$

Note that (2.11), with $\mu > 0$, implies that $\bar{F}(t) > 0$ for all t. X is defined to be NBUE if $0 < \mu < \infty$ and the inequality in (2.11) is reversed. NBUE distributions can have finite support.

Define X^* to be a random variable distributed as the stationary renewal distribution corresponding to X; X^* has cdf $G(x) = \mu^{-1}\int_0^x \bar{F}(t)\,dt$. Let $h^*(x)$ be the failure rate function of X^* defined by $h^*(x) = \bar{F}(x)/\mu\bar{G}(x)$ for x with $\bar{F}(x) > 0$. Noting that $h^*(x) = [E(X - x|X > x)]^{-1}$ it follows that NWUE is equivalent to each of the following:

$$h^*(x) \le \mu^{-1} \quad \text{for all } x \ge 0, \tag{2.12}$$

$$X \le_{\text{st}} X^*. \tag{2.13}$$

Moreover, (2.12) implies

$$X^* \ge_{\text{st}} \mu\varepsilon, \tag{2.14}$$

where $\mu\varepsilon$ is exponentially distributed with mean μ.

Similarly X NBUE is equivalent to the reverse inequality holding in (2.12) for all x with $\bar{F}(x) > 0$. It is also equivalent to the reverse inequality holding in (2.13).

For two probability distributions F_1, F_2 define $\mathscr{D}^*(F_1, F_2) = \sup_{B\in\beta}|F_1(B) - F_2(B)|$, where β is the collection of Borel sets in R. Let λ be a measure dominating both F_1 and F_2 and set $f_i = dF_i/d\lambda$, $i = 1,2$. Define $A = \{t\colon f_1(t) > f_2(t)\}$. Note that

$$\begin{aligned} F_1(B) - F_2(B) &= \int_B (f_1 - f_2)\,d\lambda \le \int_{B\cap A} (f_1 - f_2)\,d\lambda \\ &\le \int_A (f_1 - f_2)\,d\lambda. \end{aligned} \tag{2.15}$$

Similarly,

$$F_2(B) - F_1(B) \le \int_{\bar{A}} (f_2 - f_1)\,d\lambda. \tag{2.16}$$

Since $\int_A (f_1 - f_2)\,d\lambda = \int_{\bar{A}} (f_2 - f_1)\,d\lambda$, it follows from (2.15) and (2.16) that

$$\mathscr{D}^*(F_1, F_2) = \int_A (f_1 - f_2)\,d\lambda = F_1(A) - F_2(A). \tag{2.17}$$

Define $c\varepsilon$ to be an exponentially distributed random variable with mean c. For a random variable X on $[0,\infty)$ with $\mu_2 = EX^2 < \infty$, define $\rho_X = |(\mu_2/2\mu^2) - 1| = |\gamma_X - 1|$.

LEMMA 2.3. *If X is either NWUE or NBUE, then $\mathscr{D}^*(X^*, \mu\varepsilon) \le \rho_X$.*

PROOF. Consider the NWUE case, the NBUE case being totally analogous. Define $A = \{t\colon \bar{F}(t) > \exp(-t\mu^{-1})\}$. By (2.17),

$$\mathscr{D}^*(X^*, \mu\varepsilon) = \mu^{-1}\int_A \left(\bar{F}(t) - \exp(-t\mu^{-1})\right)dt. \tag{2.18}$$

From (2.13), (2.14) and (2.18),

$$\mathscr{D}^*(X^*, \mu\varepsilon) \le \mu^{-1}\int_A \left(\overline{G}(t) - \exp(-t\mu^{-1})\right) dt \tag{2.19}$$
$$\le \mu^{-1}\int_0^\infty \left(\overline{G}(t) - \exp(-t\mu^{-1})\right) dt = \rho_X. \qquad \square$$

Define Y^* to be the stationary renewal distribution corresponding to Y. As previously mentioned Y^* is also the stationary renewal distribution corresponding to Y_0.

For $X_1 \sim F_1$, $X_2 \sim F_2$ define $\mathscr{D}(X_1, X_2) = \mathscr{D}(F_1, F_2) = \sup_t|F_1(t) - F_2(t)|$, the sup norm distance between F_1 and F_2.

COROLLARY 2.2. *If $X \ge 0$ with $F(0) < 1$ and $\mu_2 < \infty$, then $\mathscr{D}(Y^*, (EY_0)\varepsilon) \le pq^{-1}\gamma$.*

PROOF. By simple computation, $\rho_{Y_0} = pq^{-1}\gamma$. The result now follows from Lemmas 2.1 and 2.3. $\square$

Recall that $d(Y_0) = \mathscr{D}(Y_0, (EY_0)\varepsilon)$, the sup norm distance between Y_0 and an exponential distribution with the same mean.

THEOREM 2.1. *If $X \ge 0$ with $F(0) < 1$ and $\mu_2 < \infty$, then*

(i)
$$q\exp(-pq^{-1}(2\gamma - 1))\exp(-t/EY_0) \le \overline{F}_{Y_0}(t) \le \exp(-t/EY_0) + \gamma pq^{-1},$$

(ii)
$$d(Y_0) \le p\max(2\gamma, \gamma q^{-1}) = \begin{cases} 2\gamma p, & 0 < p \le \frac{1}{2}, \\ \gamma q^{-1}p, & p > \frac{1}{2}. \end{cases}$$

PROOF. The bound on the left of (i) is the conclusion of Corollary 2.1. The upper bound follows from Lemma 2.1, (2.13) and Corollary 2.2 by

$$\overline{F}_{Y_0}(t) - \exp(-t/EY_0) \le \overline{F}_{Y^*}(t) - \exp(-t/EY_0) \le \mathscr{D}(Y^*, (EY_0)\varepsilon) \tag{2.20}$$
$$\le \gamma pq^{-1}.$$

Finally (ii) follows from (i) noting that

$$\exp(-t/EY_0) - \overline{F}_{Y_0}(t) \le \exp(-t/EY_0)\left[1 - q\exp(-pq^{-1}(2\gamma - 1))\right] \tag{2.21}$$
$$\le 1 - q\left(1 - pq^{-1}(2\gamma - 1)\right) = 2\gamma p.$$

Thus, $d(Y_0) \le \max(2\gamma p, \gamma pq^{-1}) \le 2\gamma p$ for $p \le \frac{1}{2}$. $\square$

COROLLARY 2.3. *Under the conditions of Theorem 2.1,*

$$d(Y) \le pq^{-1}\max(1 + \gamma q^{-1}, 2\gamma - 1).$$

PROOF. Since $\bar{F}_Y(t) = q^{-1}\bar{F}_{Y_0}(t)$, we can multiply all three components of inequality (i) of Theorem 2.1, obtaining an upper and lower bound for $\bar{F}_Y(t)$. Subtracting $\exp(-t/EY)$ from both $\bar{F}_Y(t)$ and the upper bound, we obtain

$$\begin{aligned}\bar{F}_Y(t) - \exp(-t/EY) &\le \sup\big(q^{-1}\exp(-t/EY_0) - \exp(-t/EY)\big) + \gamma q^{-2}p \\ &= pq^{-1}(1 + \gamma q^{-1}).\end{aligned} \tag{2.22}$$

Subtracting $\bar{F}_Y(t)$ from $\exp(-t/EY)$ and from the lower bound for $\bar{F}_Y(t)$ yields

$$\exp(-t/EY) - \bar{F}_Y(t) \le \sup(\exp(-t/EY) - b\exp(-t/EY_0)), \tag{2.23}$$

where $b = \exp(-pq^{-1}(2\gamma - 1))$.

For $b \ge q$, the right side of (2.23) equals $pq^{p/q}\exp(2\gamma - 1)$ which is bounded above by p, while for $b < q$ the right side of (2.23) equals $1 - b$ which is bounded above by $pq^{-1}(2\gamma - 1)$. Thus $d(Y)$ is bounded above by the larger of $pq^{-1}(1 + \gamma q^{-1})$ and $pq^{-1}(2\gamma - 1)$, and the result is thus proved. □

We next seek bounds for $d(Y_0 + Z)$ with $Z \ge 0$ and independent of Y_0. A few simple preliminary results are first presented.

First, for $c_1 < c_2$, a routine calculus argument proves

$$\mathscr{D}(c_1\varepsilon, c_2\varepsilon) = \left(1 - \frac{c_1}{c_2}\right)\left(\frac{c_1}{c_2}\right)^{c_1/c_2 - c_1} \le 1 - \frac{c_1}{c_2}. \tag{2.24}$$

Next, note that for any constant β,

$$\mathscr{D}(W_1 + \beta, W_2 + \beta) = \mathscr{D}(W_1, W_2). \tag{2.25}$$

It follows from (2.25) that for any random variable V independent of W_1 and W_2 that

$$\mathscr{D}(W_1 + V, W_2 + V) \le \mathscr{D}(W_1, W_2). \tag{2.26}$$

Next, let $Z \ge 0$ be independent of ε with Laplace transform $\mathscr{L}$.

LEMMA 2.4. $\mathscr{D}(c\varepsilon, c\varepsilon + Z) \le 1 - \mathscr{L}(c^{-1}) \le 1 - \exp(-c^{-1}EZ) \le c^{-1}EZ$.

PROOF. Let $F(t) = 1 - \exp(-c^{-1}t)$. For $t \ge y \ge 0$,

$$F(y) = (\bar{F}(t-y) - \bar{F}(t))/\bar{F}(t-y) \ge \bar{F}(t-y) - \bar{F}(t). \tag{2.27}$$

For $0 \le t \le y$,

$$F(y) = 1 - \bar{F}(y) \ge 1 - \bar{F}(t) = \bar{F}\big[(t-y)^+\big] - \bar{F}(t). \tag{2.28}$$

Thus $F(y) \ge \bar{F}[(t-y)^+] - \bar{F}(t)$ for $y, t \ge 0$. Consequently,

$$\Pr(c\varepsilon + Z > t) - \Pr(c\varepsilon > t) = E\big[\bar{F}(t-Z)^+ - \bar{F}(t)\big] \le EF(Z). \tag{2.29}$$

Now

$$\begin{aligned}EF(Z) &= E\big[1 - \exp(-c^{-1}Z)\big] \\ &= 1 - \mathscr{L}(c^{-1}) \le 1 - \exp(-c^{-1}EZ).\end{aligned} \tag{2.30}$$

□

Finally, we need a simple but useful result.

LEMMA 2.5. *Suppose that X and Y are both either stochastically larger or stochastically smaller than Z. Then*

$$\mathscr{D}(X,Y) \le \max(\mathscr{D}(X,Z), \mathscr{D}(Y,Z)).$$

PROOF. Immediate. □

The above inequalities now enable us to derive

THEOREM 2.2. *Let X be as in Theorem 2.1 and let $Z \ge 0$ be independent of X with $EZ < \infty$. Define $\delta_Z = EZ/\mu$. Then*:

(i) $d(Y_0 + Z) \le [2\gamma + \delta_Z q^{-1}]p$, *for* $0 < p \le \frac{1}{2}$.
(ii) $d(Y^*) \le \gamma p q^{-1}$.

PROOF. By the triangle inequality,

$$d(Y_0 + Z) \le \mathscr{D}(Y_0 + Z, (EY_0)\varepsilon + Z) + \mathscr{D}((EY_0)\varepsilon + Z, (E(Y_0 + Z))\varepsilon). \tag{2.31}$$

By (2.26) and Theorem 2.1,

$$\mathscr{D}(Y_0 + Z, (EY_0)\varepsilon + Z) \le d(Y_0) \le 2\gamma p, \quad \text{for } 0 < p \le \tfrac{1}{2}. \tag{2.32}$$

By Lemma 2.5,

$$\begin{aligned} &\mathscr{D}((EY_0)\varepsilon + Z, (E(Y_0 + Z))\varepsilon) \\ &\quad \le \max[\mathscr{D}((EY_0)\varepsilon + Z, (EY_0)\varepsilon), \mathscr{D}((E(Y_0 + Z))\varepsilon, (EY_0)\varepsilon)]. \end{aligned} \tag{2.33}$$

By Lemma 2.4,

$$\begin{aligned} &\mathscr{D}((EY_0)\varepsilon + Z, (EY_0)\varepsilon) \\ &\quad \le 1 - \exp(-EZ/EY_0) = 1 - \exp(-pq^{-1}\delta_Z) \le pq^{-1}\delta_Z. \end{aligned} \tag{2.34}$$

By (2.24),

$$\mathscr{D}((E(Y_0 + Z))\varepsilon, (EY_0)\varepsilon) \le 1 - \frac{EY_0}{E(Y_0 + Z)} \le \frac{EZ}{EY_0} = pq^{-1}\delta_Z. \tag{2.35}$$

Result (i) now follows from (2.31)–(2.35).
Finally by (2.13), (2.14), Lemma 2.5, Corollary 2.2 and (2.24),

$$\begin{aligned} d(Y^*) &\le \max[\mathscr{D}(Y^*, (EY_0)\varepsilon), \mathscr{D}((EY^*)\varepsilon, (EY_0)\varepsilon)] \\ &\le pq^{-1}\gamma. \end{aligned} \tag{2.36}$$

Thus, (ii) holds and the proof of Theorem 2.2 is complete. □

3. Improved bounds under additional assumptions. In this section we outline the improvements in the results under various aging assumptions on the distribution of X.

3.1. *NBUE.* Suppose that X is NBUE distributed. Then $Y^* =_{st} Y_0 + X^* \le_{st} Y_0 + X =_{st} Y$; thus Y is NBUE. Note that Y^* is the stationary renewal distribution corresponding to both Y_0 and Y. It thus follows from (2.12) that

$$\frac{p}{\mu} \le h_{Y^*}(t) \le \frac{p}{q\mu} \quad \text{for all } t \ge 0. \tag{3.1.1}$$

Consequently,

$$\exp\left(-\frac{px}{q\mu}\right) \le \frac{\bar{F}_{Y^*}(t+x)}{\bar{F}_{Y^*}(t)} \le \exp\left(-\frac{px}{\mu}\right) \quad \text{for all } t, x \ge 0. \tag{3.1.2}$$

Thus for p small, Y^* has an approximate lack of memory in that the residual age distribution varies with t by at most p in sup norm.

Note that from (3.1.1), $(EY_0)\varepsilon \le_{st} Y^* \le_{st} (EY)\varepsilon$ and from (2.13), $Y_0 \le_{st} Y^* \le_{st} Y$. Since $\mathscr{D}(Y, Y_0) \le p$ and $\mathscr{D}((EY_0)\varepsilon, (EY)\varepsilon) \le p$, it follows that

$$\max\big(d(Y^*, Y_0), d(Y^*, Y), d(Y^*, (EY_0)\varepsilon)\big) \le p. \tag{3.1.3}$$

Furthermore by Lemma 2.3,

$$d(Y^*, (EY)\varepsilon) \le \rho_Y = p\rho_X. \tag{3.1.4}$$

From (3.1.3) and (3.1.4), using the methodology of Section 2, it is straightforward to derive

$$d(Y_0) \le p, \tag{3.1.5}$$

$$d(Y) \le p, \tag{3.1.6}$$

$$d(Y^*) \le p\rho_X, \tag{3.1.7}$$

$$d(Y_0 + Z) \le p\big(1 + q^{-1}\delta_Z\big), \tag{3.1.8}$$

$$\bar{F}_Y(t) \ge \exp(-t/EY) - p\rho_X, \tag{3.1.9}$$

$$\exp(-t/EY_0) \le \bar{F}_Y(t) \le q^{-1}\exp(-t/EY). \tag{3.1.10}$$

Note that $\Pr(Y_0 = 0) = p/(1 - q\Pr(X = 0)) \ge p$. It follows that for any nonnegative X with finite mean,

$$d(Y_0) \ge p. \tag{3.1.11}$$

Thus (3.1.5) and (3.1.11) show that for F NBUE,

$$d(Y_0) = p. \tag{3.1.12}$$

Finally, we mention that when $\rho_X < p/2$ we can improve on (3.1.6) by using Daley's (1988) bound for NBUE distributions applied to Y. This yields

$$d(Y) \le \sqrt{2\rho_Y} = \sqrt{2p\rho_X}. \tag{3.1.13}$$

3.2. *NBU.* If X is NBU, then an argument similar to the NBUE case of Section 3.1 shows that Y is also NBU. Now Y_0 is NWU, Y is NBU and

$Y_0 - t|Y_0 > t =_{st} Y - t|Y > t$ for all $t \geq 0$. Thus

$$Y_0 \leq_{st} Y - t|Y > t \leq_{st} Y, \tag{3.2.1}$$

$$q\overline{F}_Y(x) \leq \frac{\overline{F}_Y(t+x)}{\overline{F}_Y(t)} \leq \overline{F}_Y(x) \quad \text{for all } x, t \geq 0. \tag{3.2.2}$$

The residual age distributions thus cannot vary by more than p in sup norm. Since Y is NBU, we can derive an analogue of Corollary 2.1 for Y:

$$\overline{F}_Y(t) \leq \exp(p)\exp(-t/EY). \tag{3.2.3}$$

Combine (3.1.9), (3.1.10) and (3.2.3) to obtain

$$\begin{aligned} &\max(\exp(-t/EY_0), \exp(-t/EY) - p\rho_X) \\ &\quad \leq \overline{F}_Y(t) \leq \exp(p)\exp(-t/EY). \end{aligned} \tag{3.2.4}$$

3.3. *NWUE.* Assume that X is NWUE distributed. It follows from the argument of Section 2.1 that Y is also NWUE. Thus by Lemma 2.3,

$$\mathscr{D}(Y^*, (EY)\varepsilon) \leq p\rho_X. \tag{3.3.1}$$

Since $Y \leq_{st} Y^*$ [(2.13)], it follows from (3.3.1) that

$$\overline{F}_Y(t) - \exp(-t/EY) \leq \overline{F}_{Y^*}(t) - \exp(-t/EY) \leq p\rho_X. \tag{3.3.2}$$

Finally using (3.3.1), Lemma 2.5, (2.13), (2.14) and (2.24),

$$d(Y^*) \leq \max[\mathscr{D}(Y^*, (EY)\varepsilon), \mathscr{D}((EY^*)\varepsilon, (EY)\varepsilon)] \leq p\rho_X. \tag{3.3.3}$$

3.4. *NWU.* Assume that X is NWU. We can easily derive the analogue of Corollary 2.1:

$$\overline{F}_Y(t) \geq \exp(-p[t\mu^{-1} + 2\gamma - 1]) = \exp(-p(2\gamma - 1))\exp(-t/EY). \tag{3.4.1}$$

Then from (3.4.1),

$$\exp(-t/EY) - \overline{F}_Y(t) \leq 1 - \exp(-p(2\gamma - 1)) \leq (2\gamma - 1)p. \tag{3.4.2}$$

Then (3.3.2) and (3.4.2) yield

$$d(Y) \leq (2\gamma - 1)p. \tag{3.4.3}$$

3.5. *IMRL.* A random variable $[0, \infty)$ is defined to be IMRL (increasing mean residual life) distributed if $EX < \infty$ and $E(X - t|X > t)$ is increasing in $t \geq 0$. For X IMRL, Brown (1980), page 231, derived the upper bound on the renewal function:

$$M(t) \leq t\mu^{-1} + \gamma. \tag{3.5.1}$$

We can therefore improve upon (3.4.1) and its consequences, obtaining

$$d(Y) \leq p\rho_X. \tag{3.5.2}$$

3.6. *DFR.* A random variable on $[0, \infty)$ is defined to be DFR (decreasing failure rate) distributed if $X - t|X > t$ is stochastically increasing in $t \geq 0$.

Shantikumar (1988) recently proved that geometric convolutions of DFR are DFR. Thus if X is DFR, then so are Y_0, Y and Y^*. Using the DFR property of Y and Y_0, it follows from Brown (1983), page 422, that

$$\max(d(Y), d(Y^*)) \le \frac{p\rho_X}{p\rho_X + 1}, \tag{3.6.1}$$

$$d(Y_0) \le \frac{\rho_{Y_0}}{\rho_{Y_0} + 1} = \frac{p\gamma}{p\gamma + q}. \tag{3.6.2}$$

A geometric convolution of DFR random variables arises naturally in the study of time to first failure for repairable systems [Brown (1984), page 611].

3.7. *IFR*. Assume that X is IFR (increasing failure rate). Then it follows from Brown (1987) that

$$M(t) \ge \frac{t}{\mu} + \frac{\sigma^2}{\mu^2}, \tag{3.7.1}$$

where M is the renewal function corresponding to X. Then (3.7.1) and an analogue of Corollary 2.1 yield

$$\bar{F}_Y(t) \le \exp(-p(M(t) - 1)) \le \exp(2\rho_X p)\exp(-t/EY). \tag{3.7.2}$$

Since for X NBUE, $0 \le \rho_X \le \frac{1}{2}$, we see that (3.7.2) improves upon (3.2.3).

3.8. $\Pr(X = 0) = \beta \in (0, 1)$, *known*. If $\beta = \Pr(X = 0)$ is known, with $0 < \beta < 1$, then an improvement in the bound for $d(Y_0)$ can be achieved. Define X' to be distributed as the conditional distribution of X given $X > 0$. Then

$$\sum_1^{G_p} X_i =_{\text{st}} \sum_1^{G_{p^*}} X_i', \quad \text{where } p^* = p/(1 - \beta q), \tag{3.8.1}$$

$$\gamma_{X'} = (1 - \beta)\gamma_X. \tag{3.8.2}$$

From Theorem (2.1), for $0 < p^* \le \frac{1}{2}$,

$$d(Y_0) \le 2\gamma_{X'}p^* = (2\gamma_X p)(1 - \beta)/(1 - \beta q). \tag{3.8.3}$$

4. Comments and additions.

4.1. Recall that $Y_0 = \sum_1^{N_0} X_i$. If we replace each X_i by cX_i with $c > 0$, then $Y_0 \to cY_0$ and $d(cY_0) = d(Y_0)$. Thus $d(Y_0)$ depends on (μ, μ_2) only through $\gamma = \mu_2/2\mu^2$. Therefore, for a fixed p, if we define $B(p, \mu, \mu_2)$ to be the best bound for $d(Y_0)$ among all distributions (of X) with fixed (μ, μ_2) and define $B(p, \gamma)$ to be the best bound for $d(Y_0)$ among all distributions with fixed $\gamma = \mu_2/2\mu^2$, then $B(p, \mu, \mu_2) = B(p, \gamma)$. We now argue that $\lim_{p \to 0}[B(p, \gamma)/p] = 2\gamma$ and thus that (1.3) is asymptotically (as $p \to 0$) sharp.

Consider $X_\alpha \sim \text{Bin}(1, \alpha)$, i.e., $\Pr(X_\alpha = 1) = \alpha$, $\Pr(X_\alpha = 0) = 1 - \alpha$. Then $\gamma_\alpha = EX_\alpha^2/2(EX_\alpha)^2 = (2\alpha)^{-1}$. Thus as α ranges from 1 to 0, γ_α ranges from $\frac{1}{2}$

to ∞. Thus all possible values of γ are assumed by the Bin$(1, \alpha)$ family. Now, let $Y_0(\alpha)$ be a sum of G_p X_α's. Then

$$\Pr(Y_0(\alpha) = 0) = \frac{p}{p + q\alpha} = \left(\frac{2}{q + 2\gamma p}\right)\gamma p. \tag{4.1.1}$$

It follows from (4.1.1) and (1.3) that for $0 < p \le \frac{1}{2}$,

$$2\gamma p \ge B(p, \gamma) \ge 2\gamma p/(q + 2\gamma p). \tag{4.1.2}$$

From (4.1.2) it immediately follows that $\lim_{p \to 0}(B(p, \gamma)/p) = 2\gamma$.

I conjecture that $B(p, \gamma) = 2\gamma p/(q + 2\gamma p)$ and thus that the Bin$(1, (2\gamma)^{-1})$ distribution maximizes $d(Y_0)$ among all distributions with fixed γ.

4.2. Solovyev's (1971) bounds are more difficult to compute than the current bounds and are derived under more stringent conditions. The two methodologies are entirely different. It is hard to make a comparison to cover all possible cases, but it appears that the bounds of this paper are, in general, tighter.

For example, Solovyev's bound, specialized to $d(Y_0)$ takes the form

$$d(Y_0) \le C(\alpha_r/r - 2)p, \tag{4.2.1}$$

where $2 < r \le 3$, $\alpha_r = (EX^r/(EX)^r)^{1/r-1}$ and $C \le 24$. For $r = 3$, Solovyev reports that C can be reduced to approximately 12. The bound (4.2.1) requires existence of EX^r for some $r > 2$. Since $2 = (1/(r-1))r + ((r-2)/(r-1))1$, it follows from the log-convexity of moments [Marshall and Olkin (1979), page 74] that

$$\alpha_r \ge 2\gamma \quad \text{for } r \ge 2. \tag{4.2.2}$$

Thus α_r ranges from 2γ to ∞, depending on EX^r. For $r = 3$, the right side of (4.2.1) with $C = 12$ is $12\alpha_3 p \ge 24\gamma p$. Thus Solovyev's bound for $d(Y_0)$, employing $r = 3$, is at least 12 times as large as the bound (1.3) of this paper. For $2 < r < 3$, Solovyev's bound (using $C = 24$) is at least $(24/r - 2)$ times as large as (1.3).

4.3. The bounds for $F_{Y_0}(t)$ derived in Sections 2 and 3 immediately yield bounds for the renewal function of a terminating renewal process [see Feller (1971), Section XI.6].

4.4. A simple argument is now presented to show that under very general conditions geometric convolutions are asymptotically exponential as $p \to 0$.

Consider a random sequence $\{X_i,\ i \ge 1\}$ which obeys the strong law of large numbers for $\mu > 0$, that is,

$$\Pr\left(\lim_{n \to \infty} \overline{X}_n = \mu\right) = 1, \tag{4.4.1}$$

where $\overline{X}_n = (1/n)\sum_1^n X_i$. Define G_p to be a random variable which is geometrically distributed with parameter p. Consider $Y_0(p) = \sum_1^{G_p} X_i$. Now

$$pY_0(p) = (pG_p)\overline{X}_{G_p}. \tag{4.4.2}$$

It follows from (3.1.12) that $d(pG_p) = p$ and thus pG_p converge in distribution to an exponential with mean 1. By (4.4.1), $\overline{X}_{G_p} \to \mu$ a.s. Thus $pY_0(p)$ converges in distribution to an exponential with mean μ.

In the i.i.d. case, $0 < EX < \infty$ suffices for exponential convergence of Y_0. It is not necessary that X be nonnegative, or that G_p be independent of $\{X_i,\ i \geq 1\}$ or that $EX^2 < \infty$.

It is also seen that a large variety of dependent sequences lead to exponential convergence of geometric convolutions, for example, stationary ergodic sequences with $0 < \mu < \infty$. An interesting problem is to obtain error bounds for $d(Y_0)$ [also $d(Y)$ and $d(Y_0 + Z)$] for various classes of dependent sequences $\{X_i,\ i \geq 1\}$.

If we relax (4.4.1) to convergence in probability but impose the condition that G_p be independent of $\{X_i\}$, then again $pY_0(p)$ is asymptotically exponential.

Acknowledgment. It is a pleasure to acknowledge stimulating correspondence with Professor D. J. Daley on various topics related to this paper.

REFERENCES

BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rhinehart and Winston, New York.

BROWN, M. (1980). Bounds, inequalities, and monotonicity properties for some specialized renewal processes. *Ann. Probab.* **8** 227–240.

BROWN, M. (1983). Approximating IMRL distributions by exponential distributions, with applications to first passage times. *Ann. Probab.* **11** 419–427.

BROWN, M. (1984). On the reliability of repairable systems. *Oper. Res.* **32** 607–615.

BROWN, M. (1987). Inequalities for distributions with increasing failure rate. In *Contributions to the Theory and Applications of Statistics, A Volume in Honor of Herbert Solomon* (A. E. Gelfand, ed.) 3–17. Academic, Orlando, Fla.

BROWN, M. and GE, G. (1984). Exponential approximations for two classes of aging distributions. *Ann. Probab.* **12** 869–875.

DALEY, D. J. (1984). Inequalities for a geometric random sum. Unpublished.

DALEY, D. J. (1988). Tight bounds on the exponential approximations of some aging distributions. *Ann. Probab.* **16** 414–423.

DALEY, D. J. and TRENGOVE, C. D. (1977). Bounds for mean waiting times in single server queues: A survey. Dept. Statistics (IAS), Australian National Univ.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

GERTSBAKH, I. B. (1984). Asymptotic methods in reliability theory: A review. *Adv. in Appl. Probab.* **16** 147–175.

JACOBS, P. A. (1986). First passage times for combinations of random loads. *SIAM J. Appl. Math.* **46** 643–656.

KEILSON, J. (1966). A limit theorem for passage times in ergodic regenerative processes. *Ann. Math. Statist.* **37** 886–890.

KÖLLERSTRÖM, J. (1976). Stochastic bounds for the single server queue. *Math. Proc. Cambridge Philos. Soc.* **80** 521–525.

LORDEN, G. (1970). On excess over the boundary. *Ann. Math. Statist.* **41** 520–527.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

SHANTIKUMAR, J. G. (1988). DFR property of first passage times and its preservation under geometric compounding. *Ann. Probab.* **16** 397–406.

SMITH, W. L. (1958). Renewal theory and its ramifications (with discussion). *J. Roy. Statist. Soc. Ser. B* **20** 243–302.

SOLOVYEV, A. D. (1971). Asymptotic behaviour of the time of first occurrence of a rare event. *Engrg. Cybernetics* **9** 1038–1048.

STOYAN, D. (1983). *Comparison Methods for Queues and Other Stochastic Models* (D. J. Daley, ed., with revisions). Wiley, New York.

SZEKLI, R. (1986). On the concavity of the waiting time distribution in some $GI/G/1$ queues. *J. Appl. Prob.* **23** 555–561.

DEPARTMENT OF MATHEMATICS
THE CITY COLLEGE, CUNY
138TH STREET AND CONVENT AVENUE
NEW YORK, NEW YORK 10031