ATTAINABLE BOUNDS FOR EXPECTATIONS

E. SENETA and N. C. WEBER

Abstract

A simple technique for obtaining bounds in terms of means and variances for the expectations of certain functions of random variables in a given class is examined. The bounds given are sharp in the sense that they are attainable by at least one random variable in the class. This technique is applied to obtain bounds for moment generating functions, the coefficient of skewness and parameters associated with branching processes. In particular an improved lower bound for the Malthusian parameter in an age-dependent branching process is derived.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): primary 60 E 15; secondary 60 J 80.

1. Introduction

It has been known for some time that one can bound the exponential function above by a quadratic function over an interval of the form $(-\infty, B]$. This result was used by Bennett (1962) and by Brook (1966) to obtain an upper bound for the moment generating function of particular random variables in terms of their mean and variance. The bounds given by these authors are sharp in the sense that they are attainable by at least one random variable in the class considered.

We show first that any function with monotone second derivative can be bounded above or below over intervals of the form $(-\infty, B]$ or $[A, \infty)$ by a quadratic function. This simple result is used here to provide attainable bounds for the expectation of such functions of random variables in terms of the mean and variance.

The particular case due to Bennett, an upper bound for a moment generating function, has found more recent use by Serfling (1974) to derive bounds for the

411

tail probabilities of a sum of random variables which are sampled from a finite population without replacement. Our general result enables lower bounds for moment generating functions to be calculated in certain cases.

A second application results in moment inequalities such as attainable bounds for the coefficient of skewness for bounded random variables.

The bounding technique has substantial application in branching processes. It can be used to provide a convex upper bound for a probability generating function, which leads to an attainable upper bound for the extinction probability in the supercritical case of a Galton-Watson process. This upper bound for the extinction probability can be computed from a knowledge of only the mean and variance of the offspring distribution. This particular application was noted by Brook (1966) and rediscovered by Turnbull (1973). In the case of an age-dependent branching process, Brockwell (1969) uses the bound given by Brook to obtain a sharp upper bound for the Malthusian parameter; and also provides a sharp lower bound.

To find the upper bound, only the mean number of offspring A (> 1) and the mean and variance of the lifetime distribution need to be finite and known; in the case of his lower bound the value of the variance need not be finite or known. We show by an application of the general result of Section 2 that the lower bound given by Brockwell can be improved if there is an upper bound $B < \infty$ to the lifetime; the improved bound can be computed from a knowledge of B, the two means, and the variance of the lifetime. The degree of improvement is illustrated by numerical examples. In as much as estimation techniques can be used to yield approximations to the means, and to the variance of the lifetime distribution, such bounding techniques are of value in approximating the Malthusian parameter.

2. General results

The main theorem relies on the following two lemmas.

LEMMA 1. *Let $\tau(x)$ be a function from $(-\infty, B]$ to R, where B is finite. Suppose $\tau''(x)$, the second derivative of τ, is a strictly monotone function. Then for any $H < B$ there exist unique values u, v and w such that*

$$f(x) = ux^2 + vx + w - \tau(x)$$

satisfies (i) $f(B) = f(H) = 0$ and (ii) $f'(H) = 0$.

If τ'' is strictly increasing then the function f also satisfies

$$f(x) \geq 0, \qquad x \leq B$$

with equality only for $x = H, B$.

If τ'' is strictly decreasing then

$$f(x) \leqslant 0, \qquad x \leqslant B$$

with equality only for $x = H, B$.

PROOF. The system of equations corresponding to (i) and (ii):

$$\begin{pmatrix} B^2 & B & 1 \\ H^2 & H & 1 \\ 2H & 1 & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \tau(B) \\ \tau(H) \\ \tau'(H) \end{pmatrix}$$

is non singular and so the existence and uniqueness of u, v and w follow.

From the mean value theorem, since $f(B) = f(H) = 0$, there is at least one value γ, $H < \gamma < B$, such that $f'(\gamma) = 0$. Also, since $f'(\gamma) = f'(H) = 0$, there is at least one value α, $H < \alpha < \gamma$, such that $f''(\alpha) = 0$. Since τ'' is strictly monotone, f'' is strictly monotone and so α is the only value for which $f''(\alpha) = 0$.

Suppose τ'' is strictly increasing. Then f'' is a strictly decreasing function so

$$f''(x) > 0 \text{ if } x < \alpha \quad \text{and} \quad f''(x) < 0 \text{ if } x > \alpha.$$

Thus f has a local minimum at $x = H$ and a local maximum at $x = \gamma$, and these are the only turning points for f. Hence

$$f(x) \geqslant 0, \qquad x \leqslant B$$

with equality only for $x = H, B$.

Similarly, if τ'' is strictly decreasing then f'' is strictly increasing and

$$f(x) \leqslant 0, \qquad x \leqslant B$$

with equality only for $x = H, B$.

One can use a similar argument to prove the following.

LEMMA 2. *Let $\tau(x)$ be a function from $[B, \infty)$ to R, where B is finite. Suppose $\tau''(x)$ is a strictly monotone function. Then for any $H > B$ there exist unique values u, v and w such that*

$$f(x) = ux^2 + vx + w - \tau(x)$$

satisfies (i) $f(B) = f(H) = 0$ and (ii) $f'(H) = 0$.

If τ'' is strictly increasing then the function f also satisfies

$$f(x) \leqslant 0, \qquad x \geqslant B$$

with equality only for $x = B, H$.

If τ'' is strictly decreasing then

$$f(x) \geqslant 0, \qquad x \geqslant B$$

with equality only for $x = B, H$.

THEOREM 1. *Let X be a random variable with mean 0 and variance σ^2 $(0 < \sigma^2 < \infty)$.*

(i) *Suppose X has support contained in $(-\infty, B]$, where $0 < B < \infty$, and $\tau(x)$ is a function with strictly monotone second derivative on $(-\infty, B]$. Then*

(a) *if $\tau''(x)$ is strictly increasing*

(1) $$E\tau(X) \leqslant \left[B^2\tau(-\sigma^2/B) + \sigma^2\tau(B)\right]/(\sigma^2 + B^2);$$

(b) *if $\tau''(x)$ is strictly decreasing*

(2) $$E\tau(X) \geqslant \left[B^2\tau(-\sigma^2/B) + \sigma^2\tau(B)\right]/(\sigma^2 + B^2),$$

with equality in (1) and (2) only if X has a two point distribution taking values B and $-\sigma^2/B$.

(ii) *Suppose X has support contained in $[-A, \infty)$, where $0 < A < \infty$, and $\tau(x)$ is a function with strictly monotone second derivative on $[-A, \infty)$. Then*

(a) *if τ'' is strictly increasing*

(3) $$E\tau(X) \geqslant \left[A^2\tau(\sigma^2/A) + \sigma^2\tau(-A)\right]/(\sigma^2 + A^2);$$

(b) *if τ'' is strictly decreasing*

(4) $$E\tau(X) \leqslant \left[A^2\tau(\sigma^2/A) + \sigma^2\tau(-A)\right]/(\sigma^2 + A^2)$$

with equality in (3) and (4) only if X has a two point distribution taking values $-A$ and σ^2/A.

PROOF. (i) Choose $H < B$ so that the random variable which only takes the values H and B has mean 0 and variance σ^2. If we denote this two point random variable by Y then $H = -\sigma^2/B$ and

$$P(Y = B) = \frac{\sigma^2}{\sigma^2 + B^2} \quad \text{and} \quad P(Y = -\sigma^2/B) = \frac{B^2}{\sigma^2 + B^2}.$$

The result now follows immediately from Lemma 1. For example, if τ'' is strictly increasing then $Ef(X) \geqslant 0$, that is, $E\tau(X) \leqslant u\sigma^2 + w$. In particular, this inequality is an equality for the random variable Y, so

$$E\tau(X) \leqslant u\sigma^2 + w = E\tau(Y) = \left[B^2\tau(-\sigma^2/B) + \sigma^2\tau(B)\right]/(\sigma^2 + B^2).$$

Part (ii) is proved similarly using Lemma 2 with $H = \sigma^2/A$ and $B = -A$.

REMARK. The above results are true (apart from equality being attained only for a two-point distribution) if the assumption that $\tau''(x)$ is strictly monotone is relaxed to $\tau''(x)$ monotone. The proofs of the corresponding lemmas are similar but a little more involved.

3. Applications

1. Moment generating functions.

Bennett, in his 1962 paper, was interested in improving Bernstein's inequality to obtain a sharper bound for the tail probability of a sum of independent, bounded random variables. In his paper Bennett developed an inequality similar to (1) for the specific function $\tau(x) = e^{tx}$, $t > 0$. That is, he developed an upper bound for the moment generating function of a bounded random variable X, $M_X(t) = Ee^{tX}$, $t > 0$. Brook (1966) developed an upper bound for Ee^{-tX} when $t > 0$, and $X \geqslant 0$ with finite variance. Theorem 1 gives the following generalisation of these two results.

COROLLARY 1. *Let X be a zero mean random variable with variance σ^2 $(0 < \sigma^2 < \infty)$.*
  (i) *If there is some finite value B such that $X \leqslant B$ a.s. then for $t > 0$*

$$(5) \qquad Ee^{tX} \leqslant \left(\sigma^2 e^{tB} + B^2 e^{-t\sigma^2/B}\right) / \left(\sigma^2 + B^2\right),$$

with equality only if X has a two point distribution taking values B and $-\sigma^2/B$.
  (ii) *If there is some finite value A such that $X \geqslant -A$ a.s. then for $t > 0$*

$$(6) \qquad Ee^{-tX} \leqslant \left(\sigma^2 e^{tA} + A^2 e^{-t\sigma^2/A}\right) / \left(\sigma^2 + A^2\right)$$

with equality only if X has a two point distribution on $-A$, σ^2/A.
  (iii) *Finally, if $|X| \leqslant B$ a.s. then for all real t*
(7)

$$\left(B^2 e^{|t|\sigma^2/B} + \sigma^2 e^{-|t|B}\right) / \left(\sigma^2 + B^2\right) \leqslant Ee^{tX} \leqslant \left(B^2 e^{-|t|\sigma^2/B} + \sigma^2 e^{|t|B}\right) / \left(\sigma^2 + B^2\right).$$

PROOF. If we set $\tau(x) = e^{tx}$, $t > 0$, then (5) follows immediately from (1). Similarly, by setting $\tau(x) = e^{-tx}$, $t > 0$, (6) follows from (4).
  Also e^{tx}, $t > 0$, has strictly increasing second derivative on $[-B, \infty)$ so from (3)

$$Ee^{tX} \geqslant \left(B^2 e^{t\sigma^2/B} + \sigma^2 e^{-tB}\right) / \left(\sigma^2 + B^2\right)$$

which together with (5) yields (7) for $t > 0$. If $t = 0$, (7) holds trivially; while if $t < 0$, (7) follows from (2) and (6).

  The equation (5) is Bennett's inequality; the relevant justification above clarifies his proof. If $X \geqslant 0$ with mean μ and variance σ^2, $0 < \sigma^2 < \infty$, then $-\mu \leqslant X - \mu$, so taking $A = \mu$ in (6) we obtain

$$Ee^{-t(X-\mu)} \leqslant \left(\sigma^2 e^{t\mu} + \mu^2 e^{-t\sigma^2/\mu}\right) / \left(\sigma^2 + \mu^2\right),$$

that is, for $t > 0$,

$$(8) \qquad Ee^{-tX} \leqslant \left(\sigma^2 + \mu^2 e^{-t(\sigma^2 + \mu^2)/\mu} \right) \Big/ \left(\sigma^2 + \mu^2 \right),$$

which is Brook's result.

2. Moment inequalities.

Let $|X| \leqslant B$ a.s., where X has zero mean and variance σ^2, $0 < \sigma^2 < \infty$. An application of (1) and (3) of the theorem with $\tau(x) = x^{2k+1}$ for integer $k \geqslant 1$ yields

$$\left(B^2 (\sigma^2/B)^{2k+1} + \sigma^2 (-B)^{2k+1} \right) \Big/ (\sigma^2 + B^2) \leqslant EX^{2k+1}$$

$$\leqslant \left(B^2 (-\sigma^2/B)^{2k+1} + \sigma^2 B^{2k+1} \right) \Big/ (\sigma^2 + B^2).$$

Hence

$$|EX^{2k+1}| \leqslant \left(B^2 (-\sigma^2/B)^{2k+1} + \sigma^2 B^{2k+1} \right) \Big/ (\sigma^2 + B^2)$$

$$= \sigma^2 B (B^{2k} + \sigma^{2k}) \left(1 - (\sigma/B)^{2k} \right) \Big/ (\sigma^2 + B^2)$$

and this bound is attained if X takes only the values $\{B, -\sigma^2/B\}$ or $\{-B, \sigma^2/B\}$. If we set $k = 1$, then we obtain a bound for the coefficient of skewness, $\gamma = EX^3/\sigma^3$:

$$|\gamma| \leqslant \left(\frac{B^2 - \sigma^2}{B\sigma} \right).$$

3. Branching processes.

In the case of a Galton-Watson process, approaches to bounding the various parameters generally rely on developing an effective approximation to the offspring probability generating function. For any non-negative, integer-valued random variable X with mean μ and variance σ^2, $0 < \sigma^2 < \infty$, we have by putting $s = e^{-t}$ in (8),

$$(9) \qquad P_X(s) = Es^X \leqslant \left(\sigma^2 + \mu^2 s^{(\sigma^2 + \mu^2)/\mu} \right) \Big/ (\sigma^2 + \mu^2), \qquad 0 \leqslant s \leqslant 1.$$

The bound in (9) is a convex function on $[0, 1]$, and is itself a probability generating function when $(\sigma^2 + \mu^2)/\mu$ is an integer. If $P_X(s)$ is the probability generating function of the offspring distribution and $\mu > 1$, the extinction probability, q, has as an upper bound the unique root in $s \in (0, 1)$ of

$$(10) \qquad s = \left(\sigma^2 + \mu^2 s^{(\sigma^2 + \mu^2)/\mu} \right) \Big/ (\sigma^2 + \mu^2).$$

This implicit bound on q, depending only on knowledge of the *offspring* mean and variance μ and σ^2 was produced by Brook (1966). Turnbull (1973), without reference to Brook, obtained a bound for $P_X(s)$ which is itself a probability generating function and also only relies on μ and σ^2. When $(\sigma^2 + \mu^2)/\mu$ is an

integer, this bound coincides with (9), but it is tighter than (9) when $(\sigma^2 + \mu^2)/\mu$ is not an integer, and hence gives a better bound on q than (10). Turnbull's bounding procedure relies heavily on the random variable X being integer-valued. Bounds such as (9) can also be used to obtain bounds for other parameters of the Galton-Watson process such as the mean time to extinction in the subcritical case.

In the case of an age-dependent branching process, let $M(t)$ denote the mean population size at time t of a branching process whch has one ancestor of age zero at time zero. Let T denote the lifetime of an individual in this process and let T have distribution function $G(t)$. Suppose each individual gives rise (at death) to an expected number A of *offspring*, $1 < A < \infty$. Then (Harris (1963), page 143) if $G(0) = 0$ and G is not a lattice distribution,

$$M(t) \sim \frac{(A-1)e^{ct}}{cA^2 \int te^{-ct}\, dG(t)} \quad \text{as } t \to \infty$$

where c, the Malthusian parameter for the population, is the unique positive value of p satisfying

$$Ee^{-pT} = A^{-1}.$$

Brockwell (1969) obtained upper and lower bounds for the parameter c based only on a knowledge of the *lifetime* mean, μ, and variance σ^2, $0 < \sigma^2 < \infty$. The upper bound for c given by Brockwell was obtained using Brook's bound for the moment generating function, bound (8). For if

$$A^{-1} = Ee^{-cT} \leqslant \left(\sigma^2 + \mu^2 e^{-c(\sigma^2 + \mu^2)/\mu} \right) / \left(\sigma^2 + \mu^2 \right)$$

then

$$c \leqslant \frac{1}{\mu(1 + \nu^2)} \log\left(\frac{A}{1 - (A-1)\nu^2} \right)$$

provided $\nu < (A-1)^{-1/2}$, where $\nu = \sigma/\mu$ is the coefficient of variation for T. Furthermore, this bound is the least upper bound for c over the class of distributions G which have mean μ and variance σ^2.

Brockwell's lower bound for c is $\mu^{-1}\log A$. This lower bound can be improved using Theorem 1 if one has knowledge of an upper bound for the lifetime, T. Suppose $T \in (0, B]$, for some finite value B. Then $T - \mu \leqslant B - \mu$, so from (2)

$$Ee^{-c(T-\mu)} \geqslant \left[(B-\mu)^2 e^{c\sigma^2/(B-\mu)} + \sigma^2 e^{-c(B-\mu)} \right] / \left(\sigma^2 + (B-\mu)^2 \right)$$

so

$$Ee^{-cT} \geqslant \left[(B-\mu)^2 e^{-c(\mu - \sigma^2/(B-\mu))} + \sigma^2 e^{-cB} \right] / \left(\sigma^2 + (B-\mu)^2 \right).$$

Consider $f(\alpha) = [(B - \mu)^2 e^{-\alpha(\mu - \sigma^2/(B-\mu))} + \sigma^2 e^{-\alpha B}]/(\sigma^2 + (B - \mu)^2)$. Because T is a non lattice random variable, $\sigma^2 < \mu(B - \mu)$ so $f(\alpha)$ is a strictly decreasing, continuous function in α, $f(0) = 1$ and $f(\infty) = 0$. Thus an implicit lower bound for c is the unique value α, say, given by

$$(11) \qquad A^{-1} = \left[(B - \mu)^2 e^{-\alpha(\mu - \sigma^2/(B-\mu))} + \sigma^2 e^{-\alpha B}\right] / \left(\sigma^2 + (B - \mu)^2\right).$$

An explicit lower bound for c, which is appropriate when B is large and $A > 1 + \sigma^2/(B - \mu)^2$, is α^* where

$$A^{-1} = \frac{(B - \mu)^2}{\sigma^2 + (B - \mu)^2} \exp\left[-\alpha^*\left(\mu - \frac{\sigma^2}{(B - \mu)}\right)\right].$$

If we put $B = (K + 1)\mu$, then

$$\alpha^* = \mu^{-1}\left(1 - \frac{\nu^2}{K}\right)^{-1} \log\left(\frac{A}{1 + (\nu/K)^2}\right)$$

and $\alpha^* < \alpha \leqslant c$. Furthermore, as $B \to \infty$, $\alpha^* \to \mu^{-1} \log A$ which is the lower bound given by Brockwell (1969). Moreover, α^* is an improvement of Brockwell's bound, that is $\alpha^* > \mu^{-1} \log A$, provided

$$(12) \qquad\qquad\qquad \log A > \frac{K}{\nu^2} \log\left(1 + \left(\frac{\nu}{K}\right)^2\right).$$

Since $\log(1 + x^2) < x^2$, $x \neq 0$, α^* is certainly an improvement of Brockwell's bound if $\log A \geqslant K^{-1}$.

Also, from Theorem 1, the lower bound given by the root of the equation (11) is sharp in the sense that it is the greatest lower bound when calculated over all possible nonlattice random variables, T, with mean μ, variance σ^2, taking values in $(0, B]$. In fact, the lower bound is attained by the distribution

$$P\left(T = \mu - \frac{\sigma^2}{(B - \mu)}\right) = \frac{(B - \mu)^2}{\sigma^2 + (B - \mu)^2},$$

$$P(T = B) = \frac{\sigma^2}{\sigma^2 + (B - \mu)^2}$$

which unfortunately is lattice, although it has no mass at the origin.

To illustrate the improvement in the lower bound for c when B is known, the bounds α and α^* are compared with Brockwell's bound for various values of the coefficient of variation ν and bounds B.

Case 1. $A = 2$

Coefficient of Variation, ν	$B/\mu = K + 1$	Lower Bounds for $c\mu$			Upper Bound for $c\mu$
		$\log A$	$\alpha^*\mu$	$\alpha\mu$	
0 · 2	5	0 · 6931	0 · 6976	0 · 6977	0 · 7058
	10	0 · 6931	0 · 6957	0 · 6957	0 · 7058
	50	0 · 6931	0 · 6936	0 · 6936	0 · 7058
0 · 8	5	0 · 6931	0 · 7784	0 · 7803	1 · 0457
	10	0 · 6931	0 · 7377	0 · 7377	1 · 0457
	50	0 · 6931	0 · 7020	0 · 7020	1 · 0457
1 · 0	5	0 · 6931	0 · 8433	0 · 8456	∞
	10	0 · 6931	0 · 7659	0 · 7660	
	50	0 · 6931	0 · 7071	0 · 7071	
1 · 5	5	0 · 6931	1 · 2835	1 · 2845	∞
	10	0 · 6931	0 · 8876	0 · 8876	
	50	0 · 6931	0 · 7255	0 · 7255	

Case 2. $A = 1 \cdot 1$

Coefficient of Variation, ν	$B/\mu = K + 1$	Lower Bounds for $c\mu$			Upper Bound for $c\mu$
		$\log A$	$\alpha^*\mu$	$\alpha\mu$	
0 · 2	5	0 · 0953	0 · 0937	0 · 0954	0 · 0955
	10	0 · 0953	0 · 0952	0 · 0954	0 · 0955
	50	0 · 0953	0 · 0953	0 · 0953	0 · 0955
0 · 8	5	0 · 0953	0 · 0667	0 · 0980	0 · 0985
	10	0 · 0953	0 · 0941	0 · 0976	0 · 0985
	50	0 · 0953	0 · 0963	0 · 0963	0 · 0985
1 · 0	5	0 · 0953	0 · 0462	0 · 0997	0 · 1004
	10	0 · 0953	0 · 0934	0 · 0990	0 · 1004
	50	0 · 0953	0 · 0968	0 · 0968	0 · 1004
1 · 5	5	0 · 0953	N/A	0 · 1066	0 · 1078
	10	0 · 0953	0 · 0905	0 · 1045	0 · 1078
	50	0 · 0953	0 · 0989	0 · 0989	0 · 1078

When $A = 1 \cdot 1$, the condition for α^* to give an improvement for Brockwell's bound, $\log A \geqslant K^{-1}$, implies an improvement when $B \geqslant 11 \cdot 5\mu$. The above example suggests, at $B/\mu = 10$, that the conservative condition $\log A \geqslant K^{-1}$ is

close to the precise condition (12). In fact putting $\beta = (\nu/K)^2$ in (12) we see that

$$(K/\nu^2)\log\left(1 + (\nu/K)^2\right) = K^{-1}\{\beta^{-2}\log(1 + \beta^2)\} \to K^{-1} \quad \text{as } \beta \to 0,$$

so for small ν, or large K, the condition (12) can be written $\log A > K^{-1}$.

References

G. Bennett (1962), 'Probability inequalities for the sum of independent random variables,' *J. Amer. Statist. Assoc.* **57**, 33–45.

P. J. Brockwell (1969), 'Bounds for the asymptotic growth rate of an age-dependent branching process,' *J. Austral. Math. Soc.* **10**, 231–235.

D. Brook (1966), 'Bounds for moment generating functions and for extinction probabilities,' *J. Appl. Probability* **3**, 171–178.

T. E. Harris (1963), *The theory of branching processes* (Springer, Berlin).

R. J. Serfling (1974), 'Probability inequalities for the sum in sampling without replacement.' *Ann. Statist.* **2**, 39–48.

B. W. Turnbull (1973), 'Inequalities for branching processes,' *Ann. Probability* **1**, 457–474.

Department of Mathematical Statistics
The University of Sydney
N.S.W. 2006
Australia