PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 13, Fasc. 2 (1992), pp. 293–310

EXTENSIONS OF CHEBYCHEV'S INEQUALITY WITH APPLICATIONS

BY

PETER J. BICKEL (BERKELEY, CALIFORNIA)
AND ABBA M. KRIEGER (PHILADELPHIA, PENNSYLVANIA)

Abstract. Chebychev's inequality provides a bound on $P[|X-\mu| \geqslant k\sigma]$, where X has an arbitrary cdf F with $\sigma^2 < \infty$. We extend this result by placing further restrictions on F. We first assume that X is n times divisible so that X can be viewed as an average of n i.i.d. random variables.

Camp-Meidell's inequality provides a tighter bound than Chebychev's by assuming that X is absolutely continuous with unimodal density function. We also extend this inequality by placing additional smoothness assumptions on the density of X.

1. Introduction. Statistical inference in regression analysis often assumes that the error terms are normally distributed. The bootstrap (cf. [6]) is a nonparametric procedure in which the data are used as an empirical distribution of the errors. The theoretical results on the bootstrap involve proving that the bootstrap distribution of a statistic as $n \to \infty$ converges to the true distribution (see [3] and [7]). The conditions under which the bootstrap distribution converges to the distribution is that $p/n \to 0$ for the estimation of one regression coefficient and $p^2/n \to 0$ for the simultaneous estimation of all regression coefficients. In contrast, the validity of the normal approximation (for one coefficient) depends on the stronger uniform convergence of the diagonal of the hat matrix to 0 (see [13]). A control versus treatments experiment is described in [4] to illustrate the asymptotic properties of these approaches.

Since the theoretical results are asymptotic, we did some empirical work to study the bootstrap distribution of a statistic and the distribution of a statistic assuming normality for finite sample sizes. By varying p and n in the experiment described in [4] and using the rates of convergence as a guide, we created three situations: (i) neither the bootstrap nor the normal distribution works, (ii) only the bootstrap works, (iii) both work. We also used three different distributions for the error terms: (i) $U(-3, 3)$, (ii) T with 3 d.f., and (iii) exponential. To our surprise, the empirical findings from the computer simulations showed that the normal distribution worked well in all nine cases and outperformed the bootstrap.

One reason that the normal confidence intervals appear to be doing well could be due to the fact that the quantiles of the distribution under the actual distribution are reasonably close to the quantiles of the statistics under the normal distribution. If we consider $\bar{X}$ in the extreme case of $n = 2$, then we know that $P[\sqrt{2}|X_1+X_2| \leqslant 1.96] = .95$ under normality. The probability that $\sqrt{2}|X_1+X_2| = 1.96$ is close to .95 for the uniform, exponential and T-distribution with 3 d.f. is evident from Table 1.1.

Table 1.1. $P_F[\sqrt{2}|X_1+X_2| \leqslant c]$

F (1)	$c = 1.645$	$c = 1.96$	$c = 2.575$
Uniform	.8921	.9600	.9974
Exponential	.9296	.9511	.9764
T with 3 d.f.	.9295	.9559	.9800

(1) A linear transformation is performed so that $\mu = 0$ and $\sigma^2 = 1$.

The results of the experiment raise the following questions which are the focus of this paper. Is the coverage probability .95 because of the choices of the distributions? Would the coverage probability be approximately the same as the coverage probability assuming normality for different levels (e.g. would ± 1 standard error correspond to approximately 68%)? Sharpe [17] shows that using 1.96 standard errors is robust, over a wide class of distributions, in obtaining 95% tolerance intervals. Using the requisite number of standard errors from the normal distribution for other levels of confidence, substantively different from 95%, could lead to coverage probabilities that are quite different from the actual ones.

One approach for obtaining insight into the coverage probabilities of confidence intervals for μ assuming normality is to consider

$$P_F[|\bar{X}-\mu| \geqslant c\sigma/\sqrt{n}], \tag{1.1}$$

where X_i are i.i.d. with cdf F. One bound on (1.1) follows from Chebychev's inequality

$$P_F[|\bar{X}-\mu| \geqslant c\sigma/\sqrt{n}] \leqslant c^{-2}. \tag{1.2}$$

This bound clearly is not tight except if $n = 1$ or $n = 2$ and $c = \sqrt{2}$, since Chebychev's inequality places no restriction on the distribution of $\bar{X}$, but $\bar{X}$ must be n times divisible. In the next section, the random variables that provide the bounds for the more general problem of independent but not necessarily identical random variables are characterized and the bound for $n = 2$ is derived. Motivated by the solution for $n = 2$, we consider $P_F[|\bar{X} - \mu| \geqslant c\sigma/\sqrt{n}]$ for a specific F in Section 3. We show that as $n \to \infty$ this F is a local optimum to (1.1). Interestingly, the bound given by F as $n \to \infty$ is close to Chebychev's bound.

The bound can be made smaller by restricting the nature of the distribution. If, for example, X_i is unimodal and symmetric, then Camp-Meidell's inequality implies that

$$P_F[|\bar{X} - \mu| \geqslant c\sigma/\sqrt{n}] \leqslant \begin{cases} 1 - c/\sqrt{3} & \text{if } c \leqslant 2/\sqrt{3}, \\ (4/9)c^{-2} & \text{if } c > 2/\sqrt{3}. \end{cases}$$

A discussion of Camp-Meidell's inequality and an extension to distributions where the density has a bounded derivative are given in the last section.

An alternative approach that appears in the literature dates back to the work of Bernstein [2]. Since $P[X \geqslant c] = P[g(X) \geqslant g(c)]$ for any strictly monotonic function g, additional bounds are obtained by appropriate choices of g. This approach is used in [1], [9] and [11]. For an excellent review of this literature see [14].

2. Results on bounds. In this section, we first characterize the behavior of F that bounds (1.1). We then determine the member from this class for $n = 2$. We know that Chebychev's bound applies for $n = 1$, so if $c = 2$, the bound is .227. We also know that as $n \to \infty$, the bound must be greater than the value obtained by our special case in Section 3 which is .209 for $c = 2$. The bound for $n = 2$ at $c = 2$ is .2 as will be demonstrated in this section. This shows that half of the distance between the bound as $n \to \infty$ and $n = 1$ is obtained at $n = 2$. Samuels [16] uses the same approach except for non-negative random variables.

Let Ω be the class of cdf's corresponding to symmetric random variables X with $\mu = 0$ and $\sigma^2 = 1$. We consider the following extension to our previous problem. Let $X_1, \ldots, X_n$ be independent but not necessarily identically distributed random variables with $X_i \sim F_i \in \Omega$ for all i. Find

$$\sup_{F_1, \ldots, F_n} P[|X_1 + \ldots + X_n| \geqslant c/\sqrt{n}]. \tag{2.1}$$

By symmetry, we clearly only need to consider $P[\sum_{i=1}^{n} X_i \geqslant c\sqrt{n}]$. Let $F_1^*, \ldots, F_n^*$ be the cdf's that achieve the supremum in (2.1).

The next two lemmas serve to characterize F_i^* as the cdf corresponding to the random variable with point masses at $\pm a$ and $\pm b$ for $0 \leqslant a \leqslant 1 < b$. The characterization also follows from a general result in [10].

LEMMA 2.1. *Let Ω_0 be the subclass of cdf's such that $F \in \Omega_0$ if $F \in \Omega$ and F cannot be written as $\varepsilon F_1 + (1-\varepsilon)F_2$ for $F_1, F_2 \in \Omega$ and $0 < \varepsilon < 1$. Then $F_i^* \in \Omega_0$ for all i.*

Proof. Assume that

$$X_i \sim F = \varepsilon F_1 + (1-\varepsilon)F_2, \quad \text{where } F_1, F_2 \in \Omega \text{ and } 0 < \varepsilon < 1.$$

Then

$$(2.2)\quad P\left[(X_1+\ldots+X_n) \geqslant c\sqrt{n}\right] = \int_{x_i} P\left[(X_1+\ldots+X_n) \geqslant c\sqrt{n} \mid X_i = x_i\right] dF(x_i)$$
$$= \varepsilon \int_{x_i} P\left[(X_1+\ldots+X_n) \geqslant c\sqrt{n} \mid X_i = x_i\right] dF_1(x_i)$$
$$+(1-\varepsilon) \int_{x_i} P\left[(X_1+\ldots+X_n) \geqslant c\sqrt{n} \mid X_i = x_i\right] dF_2(x_i).$$

Since the last expression in (2.2) is linear in ε, the probability is maximized at $\varepsilon = 0$ or 1.

LEMMA 2.2. *A random variable $X \sim F \in \Omega_0$ iff X is a discrete random variable, $X(a, b)$, with point masses of $(b^2-1)/2(b^2-a^2)$ at $\pm a$ and $(1-a^2)/2(b^2-a^2)$ at $\pm b$ for $0 \leqslant a \leqslant 1 < b$.*

Proof. If $F \in \Omega$ but F is not as in the statement above, there exists $a_0 \in [0, 1]$ such that $F(1) > F(a_0) > F(0)$ and/or there exists $b_0 \in (1, \infty)$ such that $1 > F(b_0) > F(1)$. We consider the case for a_0 since the argument for b_0 follows in a similar manner. Let $p_1 = F(a_0) - F(0)$ and $p_2 = F(1) - F(a_0)$. Define

$$F_1(x) = \begin{cases} F(0) & \text{if } 0 \leqslant x \leqslant a_0, \\ (p_1+p_2)\left[F(x)-F(a_0)\right]/p_2 + F(0) & \text{if } a_0 < x \leqslant 1, \\ F(x) & \text{if } x > 1 \end{cases}$$

and

$$F_2(x) = \begin{cases} F(0) + (p_1+p_2)\left[F(x)-F(0)\right]/p_1 & \text{if } 0 \leqslant x \leqslant a_0, \\ F(1) & \text{if } a_0 < x \leqslant 1, \\ F(x) & \text{if } x > 1. \end{cases}$$

Then $F(x) = \varepsilon F_1(x) + (1-\varepsilon)F_2(x)$, where $\varepsilon = p_2/(p_1+p_2)$.

We now turn to the special case $n = 2$. Based on Lemmas 2.1 and 2.2, we need to maximize $P\left[X_1 + X_2 \geqslant k \equiv c\sqrt{2}\right]$, where $X_i = \pm a_i$ with probability $(b_i^2-1)/2(b_i^2-a_i^2)$ and $X_i = \pm b_i$ with probability $(1-a_i^2)/2(b_i^2-a_i^2)$ for

$0 \leqslant a_i \leqslant 1 < b_i$ and $i = 1, 2$. We prove a lemma which reduces the number of cases and defer the tedious task of going through all of the cases to a working paper (Bickel and Krieger (1991)).

LEMMA 2.3. *If X_2 is a random variable as described above, then in order to maximize $P[X_1+X_2 \geqslant k]$*

(i) *a_1 must satisfy $P[a_1+X_2 = k]+P[-a_1+X_2 = k] > 0$ provided that $0 < a_1 < 1$;*

(ii) *b_1 must satisfy $P[b_1+X_2 = k]+P[-b_1+X_2 = k] > 0$.*

Proof. Let

$$\phi(a_1, b_1) \equiv P[X_1+X_2 \geqslant k] = \frac{b_1^2-1}{b_1^2-a_1^2}\alpha(a_1)+\frac{1-a_1^2}{b_1^2-a_1^2}\beta(b_1),$$

where

$$\alpha(a_1) = \tfrac{1}{2}\{P[a_1+X_2 \geqslant k]+P[-a_1+X_2 \geqslant k]\}$$

and

$$\beta(b_1) = \tfrac{1}{2}\{P[b_1+X_2 \geqslant k]+P[-b_1+X_2 \geqslant k]\}$$

(i.e., $\alpha(a_1)$ and $\beta(b_1)$ are the conditional probabilities that $X_1+X_2 \geqslant k$ given $X_1 = \pm a_1$ and $X_1 = \pm b_1$, respectively).

(i) Assume (i) does not hold. That implies $\partial\alpha(a_1)/\partial a_1 = 0$. Hence by taking derivatives it follows that $\partial\phi(a_1, b_1)/\partial a_1 = 0$ if $\alpha(a_1) = \beta(b_1)$. But this implies $\phi(a_1, b_1) = \alpha(a_1)$ so that a_1 can either be decreased or increased until $P[a_1+X_2 = k]+P[-a_1+X_2 = k] > 0$ or until $a_1 = 0$ or $a_1 = 1$.

(ii) The argument for (ii) is similar to the argument for (i).

Clearly, the same conditions as (i) and (ii) in Lemma 2.3 for a_1 and b_1 apply to a_2 and b_2.

LEMMA 2.4. *If X_1 and X_2 are independent with cdf's $F_1, F_2 \in \Omega$, then $P[(X_1+X_2) \geqslant k]$ is maximized for $k \geqslant 2$ if either*

(i) *X_1 has point masses at 0 and $\pm(k-1)$; X_2 has point masses at ± 1;*

(ii) *X_1 and X_2 have point masses at 0 and $\pm k$;*

(iii) *X_1 has point masses at $\pm a$ and $\pm b$; X_2 has point masses at $\pm(k-b)$ and $\pm(k-a)$;*

(iv) *X_1 has point masses at $\pm a$ and $\pm k$; X_2 has point masses at 0 and $\pm(k-a)$;*

(v) *X_1 has point masses at $\pm a$ and $\pm(2k-a)$; X_2 has point masses at 0 and $\pm(k-a)$.*

Remarks. 1. The only solution that is a function of two variables is (iii). In the technical report we show that case (iii) can be reduced to a one-variable problem.

2. There are situations in which (iii), (iv) and (v) are optimized at an interior point. However, in all cases that we tried cases (i) and (ii) dominate.

3. Solution (ii) dominates solution (i) iff $k \geqslant 2.1939$ which is equivalent to $c \geqslant 1.55$. Thus case (ii), the only i.i.d. case, is not extremal in general for the problem we considered. However, we do not know whether case (ii) is extremal for the case of i.i.d. summands or not.

4. Similar results for two independent random variables, but where the random variables are nonnegative with given mean, appear in [5] for the case where X_1 and X_2 do not necessarily have the same distribution and in [12] for the case where X_1 and X_2 are identically distributed.

3. Special case. We want bounds on $P_F[|\bar{X}-\mu| \geqslant c\sigma/\sqrt{n}]$. Since we are implicitly assuming that σ is known, we let $\sigma = 1$ without loss of generality. Also, since we consider X_i in deviation form (i.e., $X_i-\mu$), we let $\mu = 0$. Hence the problem reduces to considering the functional

$$\phi_n(F; c) \equiv P[|\bar{X}| \geqslant c/\sqrt{n}] = P[|\sum_{i=1}^{n} X_i| \geqslant c\sqrt{n}], \tag{3.1}$$

where X_i are i.i.d. with cdf F.

Motivated by the solution for $n = 2$ in the previous section, we consider the discrete random variables that have probability mass of $p_n = 1/[2c^2 n]$ at $\pm c\sqrt{n}$ and the remaining mass of $1-2p_n$ at zero. Of course, $n \geqslant c^{-2}$. We assume $c \geqslant 1$ to avoid restrictions on n. These random variables are such that if the number of positive X_i differs from the number of negative X_i by at least one, then $|\sum_{i=1}^{n} X_i| \geqslant c\sqrt{n}$. For these random variables, with corresponding cdf denoted by G_n, it is easy to see that

$$\phi_n(G_n; c) = 1 - \sum_{j=0}^{\langle n/2 \rangle} \binom{n}{2j} (2p_n)^{2j} (1-2p_n)^{n-2j} \binom{2j}{j} \frac{1}{2^{2j}}, \tag{3.2}$$

where $\langle x \rangle$ denotes the greatest integer in x. When n is one, $\phi_n(G_n; c)$ coincides with Chebychev, and so $\phi_1(G_1; c) = c^{-2}$. Clearly, $\phi_n(G_n; c) \leqslant c^{-2}$ for all n. We show that $\phi_n(G_n; c)$ decreases monotonically in n for all c. We then find G_∞, the limiting distribution of G_n as $n \to \infty$, and hence $\phi_\infty(G_\infty; c)$. We show that G_∞ is a local maximum of $\phi_\infty(F; c)$. Finally, we show that $\phi_\infty(G_\infty; c)/\phi_1(G_1; c)$ increases in c to one as $c \to \infty$. The fact that there is an infinitely divisible distribution H such that

$$\lim_{c \to \infty} \phi_\infty(H, c)/c^{-2} = 1$$

is shown by Robbins in [15]. His H is different from G_∞ and is not tight since $\phi_\infty(H; c) \leqslant \phi_\infty(G_\infty; c)$.

The most difficult result to show is that $\phi_n(G_n; c)$ decreases in n. We prove that result by studying the behavior of

$$\beta(n; j, x) = \binom{n}{2j}\left(\frac{1}{xn}\right)^{2j}\left(1-\frac{1}{xn}\right)^{n-2j}, \tag{3.3}$$

where $x = c^2 \geqslant 1$. Clearly, if $\beta(n+1; j, x)/\beta(n; j, x) \geqslant 1$ for all n, j and x, we would be done. This is not quite the case but something close is true as is apparent from

LEMMA 3.1. (a) $\beta(n+1; j, x)/\beta(n; j, x)$ *has the following properties*:
(i) *it increases in* x *for* $j = 1, 2, \ldots, \langle n/2 \rangle$;
(ii) *it decreases in* x *for* $j = 0$;
(iii) *it is at least one for* $j = 0, 2, \ldots, \langle n/2 \rangle$ *for all* x;
(iv) *it is at least one for* $j = 1$ *and* $x \geqslant 2$.
(b) $\beta(n+1; 0, x)+\frac{1}{2}\beta(n+1; 1, x) \geqslant \beta(n; 0, x)+\frac{1}{2}\beta(n; 1, x)$ *for all* x.

The proof of Lemma 3.1 appears in the technical report.

We now turn to finding the limiting behavior of the sequence of random variables $\{X_n(c)\}$, where $P[X_n(c) = \pm c\sqrt{n}] = 1/[2c^2 n]$ and $P[X_n(c) = 0] = 1-1/[c^2 n]$ for a given value of c. The following lemma is standard:

LEMMA 3.2. *We have*

$$\sum_{i=1}^{n} X_{in}(c)/\sqrt{n} \xrightarrow{L} Y(c) \equiv c(Z_1(c)-Z_2(c)),$$

if $X_{1n}, \ldots, X_{nn}$ *are i.i.d. according to* $X_n(c)$ *and* $Z_1(c)$ *and* $Z_2(c)$ *are i.i.d. Poisson* $(\mu \equiv 1/(2c^2))$.

We now focus on proving certain results about $Y(c)$. It is easy to verify that

$$P[|Y(c)| \geqslant c] = 1-\sum_{i=0}^{\infty} e^{-2\mu}\mu^{2i}/(i!)^2, \tag{3.4}$$

where $\mu = 1/(2c^2)$. The following corollary to Lemma 3.2 is proved in the technical report.

COROLLARY 3.3. (i) $f(\mu) \equiv [1-\sum_{i=0}^{\infty} e^{-2\mu}\mu^{2i}/(i!)^2]/2\mu$ *decreases in* μ *and* (ii) $\lim_{\mu\to 0} f(\mu) = 1$.

This says that the ratio of the bound provided by $Y(c)$ to the Chebychev bound increases in c and goes to one as $c \to \infty$.

In Table 3.1, $\phi_n(G_n; c)$ (see (3.2)) is presented. Note that the last row, which is labelled $n \to \infty$, is the asymptotic result as given in (3.4). The three properties of $\phi_n(G_n; c)$, namely, $\phi_n(G_n; c)$ decreases in n, $\phi_\infty(G_n; c)/\phi_1(G_1; c)$ increases in c, and this ratio goes to one as $c \to \infty$ are evident from the table.

Table 3.1. Bounds for the special case

n	$c = 1.5$	$c = 2.0$	$c = 2.5$	$c = 3.0$	$c = 5.0$
1	.4444	.2500	.1600	.1111	.0400
2	.3704	.2266	.1504	.1065	.0394
5	.3422	.2155	.1454	.1040	.0391
10	.3342	.2122	.1439	.1032	.0389
15	.3317	.2111	.1434	.1029	.0389
20	.3304	.2106	.1431	.1028	.0389
25	.3297	.2102	.1430	.1027	.0389
50	.3282	.2096	.1427	.1026	.0389
100	.3275	.2093	.1425	.1025	.0388
$n \to \infty$	.3268	.2090	.1424	.1024	.0388
Ratio [1]	1.3601	1.1963	1.1236	1.0851	1.0302

[1] Ratio is the value at $n = 1$ divided by the value as $n \to \infty$.

We now turn to showing that G_∞ is a local optimal to ϕ_∞. If $F^*_{i,n}$, $1 \leqslant i \leqslant n$, achieves

$$\sup_{F_1,\ldots,F_n} P\left[|X_1+\ldots+X_n| \geqslant c\sqrt{n}\right],$$

then $\{(X_1+\ldots+X_n)/\sqrt{n}\}$ is necessarily tight. Further, since

$$\max_{1 \leqslant i \leqslant n} P_{F_i^*}\left[|X_i| \geqslant \varepsilon\sqrt{n}\right] \leqslant [n\varepsilon^2]^{-1} \to 0 \quad \text{as } n \to \infty,$$

any limit law of $(X_1+\ldots+X_n)/\sqrt{n}$ must be infinitely divisible, symmetric about 0, with variance $\leqslant 1$. If L is infinitely divisible, symmetric about 0, and $\mathrm{E}(L^2) < 1$, we can only increase $P[|L| \geqslant c]$ by scaling L up to 1. We conclude that

$$\underline{\lim}_{n} \sup_{F_1,\ldots,F_n} P\left[|X_1+\ldots+X_n| \geqslant c\sqrt{n}\right] = \sup_{\Omega_\infty} P_F[|L| \geqslant c],$$

where $\Omega_\infty = \{F \text{ infinitely divisible, symmetric about 0, } \mathrm{E}_F(L^2) = 1\}$.

We conjecture that

$$P_{G_\infty}[|L| \geqslant c] = \max_{\Omega_\infty} P_F[|L| \geqslant c].$$

We can only show the following local maximum result.

It is well known that

$$\Omega_\infty = \left\{F\colon \mathrm{E}_F(e^{itX}) = \exp \int_{-\infty}^{\infty} \frac{(\cos tx - 1)}{x^2} dM(x),\ M \in \Lambda\right\},$$

where $\Lambda = \{\text{probability measures on } (-\infty, \infty), \text{ symmetric about } 0\}$. Parametrize Ω_∞ by Λ and write P_M. Then G_∞ corresponds to $M_0 = \frac{1}{2}(\delta_c + \delta_{-c})$, where δ_c is point mass at c.

THEOREM 3.4. *With the above notation, if* $M_\varepsilon \equiv (1-\varepsilon)M_0 + \varepsilon M$, *where* $M \in \Lambda$ *is arbitrary, and* $c \geqslant \frac{1}{2}(2+\sqrt{2})^{1/2} = .924$, *then*

$$\overline{\lim_{\varepsilon \to 0}}\, \varepsilon^{-1}(P_{M_\varepsilon}[|L| \geqslant c] - P_{M_0}[|L| \geqslant c]) < 0$$

unless $M = M_0$.

The proof proceeds by a series of lemmas.

LEMMA 3.5. *Suppose* $\Lambda_0 = \{M \in \Lambda \colon \int_{-\infty}^{\infty} x^{-2}\, dM(x) < \infty\}$. *Then, if* $t > 0$,

$$\frac{\partial}{\partial \varepsilon} P_{M_\varepsilon}[L \geqslant c]\big|_{\varepsilon=0} = \int \{c^2 (P[L_0 \geqslant c] - P[L_0 + D_c \geqslant c]) - x^{-2}(P[L_0 \geqslant c] - P[L_0 + D_x \geqslant c])\}\, dM(x),$$

where $L_0 \sim F_0$, $D_a \sim \frac{1}{2}(\delta_a + \delta_{-a})$ *and is independent of* L_0.

Proof. Let the probability measure M^* correspond to M via

$$dM^*(x) = \frac{1}{A(M)} \frac{dM(x)}{x^2}, \quad \text{where } A(M) \equiv \int_{-\infty}^{\infty} x^{-2} dM(x).$$

Then it is well known that $X \sim M$ if

$$X = \sum_{i=1}^{N} U_i, \tag{3.5}$$

where U_i are i.i.d. with common distribution M^* and $N \sim \text{Poisson}(A(M))$. Further,

$$P_{M_\varepsilon}[X \geqslant c] = P[X_\varepsilon + Y_\varepsilon \geqslant c], \tag{3.6}$$

where $X_\varepsilon \sim F_{(1-\varepsilon)M_0}$ and $Y_\varepsilon \sim F_{\varepsilon M}$. Combining (3.5) and (3.6), we obtain

$$P_{M_\varepsilon}[X \geqslant c] = P\Big[\sum_{i=1}^{N_{1\varepsilon}} U_i + \sum_{i=1}^{N_{2\varepsilon}} V_i \geqslant c\Big],$$

where U_i are i.i.d. $M_0^* = M_0 = D_c$ and V_i are i.i.d. M^*, $N_{1\varepsilon} \sim \text{Poisson}(A_0(1-\varepsilon))$, $N_{2\varepsilon} \sim \text{Poisson}(A\varepsilon)$ all independent and $A_0 \equiv A(M_0) = c^{-2}$, $A \equiv A(M)$. Then

$$P_{M_\varepsilon}[X \geqslant c] = P[X_\varepsilon \geqslant c]e^{-A\varepsilon} + P[X_\varepsilon + V_1 \geqslant c]A\varepsilon e^{-A\varepsilon} + O(\varepsilon^2). \tag{3.7}$$

Further,

$$P[X_0 \geqslant c] = P[X_\varepsilon + X'_\varepsilon \geqslant c], \quad \text{where } X'_\varepsilon = \sum_{i=1}^{N'_{1\varepsilon}} U_i,$$

where U'_i are independent of U_i and i.i.d. M_0^* and $N'_{1\varepsilon} \sim \text{Poisson}(A_0\varepsilon)$ is independent of everything. So,

$$P_{M_0}[X \geqslant c] = e^{-\varepsilon A_0} P[X_\varepsilon \geqslant c] + \varepsilon A_0 e^{-\varepsilon A_0} P[X_\varepsilon + U'_1 \geqslant c] + O(\varepsilon^2). \tag{3.8}$$

Combining (3.7) and (3.8), we get

$$\begin{aligned} P_{M_\varepsilon}[X \geqslant c] - P_{M_0}[X \geqslant c] = (e^{\varepsilon(A_0-A)}-1)P[X_0 \geqslant c] \\ + A\varepsilon P[X_0+V_1 \geqslant c] \\ - A_0\varepsilon P[X_0+U_1' \geqslant c] + O(\varepsilon^2). \end{aligned}$$

Hence

$$\frac{\partial}{\partial\varepsilon} P_{M_\varepsilon}[X \geqslant c]\Big|_{\varepsilon=0} = A_0(P[X_0 \geqslant c] - P[X_0+U_1' \geqslant c]) - A(P[X_0 \geqslant c] - P[X_0+V_1 \geqslant c]). \tag{3.9}$$

Write

$$\begin{aligned} A(P[X_0 \geqslant c] - P[X_0+V_1 \geqslant c]) &= A\int (P[X_0 \geqslant c] - P[X_0+D_x \geqslant c])dM^*(x) \\ &= \int x^{-2}(P[X_0 \geqslant c] - P[X_0+D_x \geqslant c])dM(x). \end{aligned} \tag{3.10}$$

The lemma follows from (3.9) and (3.10).

By Lemma 3.5, the theorem follows for $M \in A_0$ from

LEMMA 3.6. *For all* $c \geqslant 1/\sqrt{2}$ *and all* x

$$P[X_0 \geqslant c] - P[X_0+D_x \geqslant c] \geqslant \frac{x^2}{c^2}(P[X_0 \geqslant c] - P[X_0+D_c \geqslant c]). \tag{3.11}$$

The proof of this lemma uses

LEMMA 3.7. *If* $Y = X - X'$, *where* X *and* X' *are independent* Poisson(λ) *and* $\lambda \leqslant 2-\sqrt{2}$, *then* $P[Y=j+1] - P[Y=j]$ *is increasing in* j *for* $j \geqslant 0$. *Hence* $P[Y=j]$ *is decreasing for* $j \geqslant 0$, *viz.* Y *is unimodal.*

Proof. If $j \geqslant 0$, then

$$P[Y=j] = e^{-2\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{2k+j}}{k!(k+j)!}.$$

Hence

$$\begin{aligned} &P[Y=j+2] - 2P[Y=j+1] + P[Y=j] \\ &\quad = e^{-2\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{2k+j}}{k!(k+j)!}\left(\frac{\lambda^2}{(k+j+2)(k+j+1)} - \frac{2\lambda}{k+j+1} + 1\right). \end{aligned}$$

The quadratic in parentheses has roots

$$k+j+2 \pm \sqrt{k+j+2} \geqslant 2-\sqrt{2}$$

for all $k, j \geqslant 0$. The result follows. Note that Y is unimodal under the weaker condition $\lambda \geqslant 1$.

Proof of Lemma 3.6. Using the notation of Lemma 3.7 it is enough to show that, for all $u \geqslant 0$,

$$P[Y \geqslant 1] - P[Y + D_u \geqslant 1] \geqslant u^2 (P[Y \geqslant 1] - P[Y + D_1 \geqslant 1]) \tag{3.12}$$

if $\lambda = 1/2c^2 \leqslant 1$ (or $c \geqslant 1/\sqrt{2}$).

Suppose $0 < u < 1$. Then (3.12) reduces to

$$0 \geqslant (u^2/2)(P[Y = 1] - P[Y = 0]) \tag{3.13}$$

which holds by Lemma 3.7. In general, if $j \leqslant u < j+1$, then

$$P[Y + D_u \geqslant 1] = P[Y + D_j \geqslant 1],$$

so that if (3.12) holds for $u = j \geqslant 1$, it holds for all u. Now, if $j \geqslant 1$, then

$$\begin{aligned} P[Y \geqslant 1] - \tfrac{1}{2}(P[Y \geqslant 1-j] + P[Y \geqslant 1+j]) \\ &= -\tfrac{1}{2}(P[Y = 0] + P[1-j \leqslant Y \leqslant -1] - P[1 \leqslant Y \leqslant j]) \\ &= -\tfrac{1}{2}(P[Y = 0] - P[Y = j]). \end{aligned}$$

So (3.12) becomes

$$P[Y = 0] - P[Y = j] \leqslant j^2 (P[Y = 0] - P[Y = 1]). \tag{3.14}$$

But, by Lemma 3.7,

$$\begin{aligned} P[Y = 0] - P[Y = j] &= \sum_{l=0}^{j-1} (P[Y = l] - P[Y = l-1]) \\ &\leqslant j(P[Y = 0] - P[Y = 1]) \leqslant j^2 (P[Y = 0] - P[Y = 1]) \end{aligned}$$

and (3.14) and the lemma follow.

To complete the proof of the theorem we need

LEMMA 3.8. *If* $\int_{-\infty}^{\infty} x^{-2}\, dM(x) = \infty$, *then* F_M *is continuous.*

Proof (due to P. W. Millar). Argue by contradiction. Without loss of generality suppose F_M has a mass at 0. If it does not, consider $F_M * F_M = F_{2M}$. Let $\{Y_t\colon t \geqslant 0\}$ be the Lévy process having $Y_0 = 0$, $Y_t = Y_1 \sim F_M$. By Lévy's inequality,

$$P[\sup_{0 < t \leqslant 1} Y_t > 0] \leqslant 2P[Y_1 > 0] < 1$$

by the symmetry of Y_t and $P[Y_1 = 0] > 0$. But $P[\sup_{0 < t \leqslant 1} Y_t \geqslant 0] = 1$ since $Y_0 = 0$. Hence

$$P[\sup_{0 < t < 1} Y_t = 0] > 0. \tag{3.15}$$

The events $E_1 \equiv \{Y_t < 0$ for t arbitrarily close to $0\}$ and $E_2 \equiv \{Y_t = 0$ for all $0 \leqslant t \leqslant t_0$, some $t_0 > 0\}$ have, by the Blumenthal 0-1 law, probability 0 or 1.

By (3.15), $P[E_1]=1$ or $P[E_2]=1$. But, by symmetry, $P[E_1]=1$ implies $P[Y_t>0$ for t arbitrarily close to $0]=1$, which contradicts (3.15). Therefore, $P[E_2]=1$ and it follows by standard theory (see, e.g., [8], pp. 274–275) that $M\in\Lambda_0$, a contradiction.

We can now complete the proof of the theorem. By Lemma 3.8, if $M\notin\Lambda_0$, then F_M is continuous. Then, if $X\sim F_M$, $X_\varepsilon\sim F_{(1-\varepsilon)M_0}$, $Y_\varepsilon\sim F_{\varepsilon M}$ are independent, we have

$$(3.16)\qquad \begin{aligned} P[X\geqslant c] &= P[X_\varepsilon+Y_\varepsilon\geqslant c] \\ &= P[X_\varepsilon+Y_\varepsilon\geqslant c]-P[X_0+Y_\varepsilon\geqslant c]+P[X_0+Y_\varepsilon\geqslant c]. \end{aligned}$$

Now,

$$(3.17)\qquad \begin{aligned} &|P[X_\varepsilon+Y_\varepsilon\geqslant c]-P[X_0+Y_\varepsilon\geqslant c]| \\ &\quad = \Big|\sum_{k=-\infty}^{\infty}(P[X_\varepsilon=ck]-P[X_0=ck])P[Y_\varepsilon\geqslant c(1-k)]\Big| \\ &\quad \leqslant \sum_{k=-\infty}^{\infty}|P[X_\varepsilon=ck]-P[X_0=ck]|\to 0 \end{aligned}$$

as $\varepsilon\to 0$. But

$$(3.18)\qquad \begin{aligned} &P[X_0+Y_\varepsilon\geqslant c]-P[X_0\geqslant c] \\ &\quad = \sum_{k=-\infty}^{\infty}P[X_0=ck]\big(P[Y_\varepsilon\geqslant c(1-k)]-1\ (k\geqslant 1)\big). \end{aligned}$$

As $\varepsilon\to 0$, we have

$$P[Y_\varepsilon\geqslant c(1-k)]\to\begin{cases}0 & \text{if } k<1,\\ 1 & \text{if } k\geqslant 2,\end{cases}\qquad\text{and}\qquad P[Y_\varepsilon\geqslant 0]=\tfrac{1}{2}$$

since $F_{\varepsilon M}$ is continuous. By (3.16)–(3.18) we obtain

$$\lim_{\varepsilon\downarrow 0}(P_{M_\varepsilon}[X\geqslant c]-P_{M_0}[M\geqslant c]) = -\tfrac{1}{2}P[X_0=c]<0.$$

The theorem follows.

4. Unimodal densities. In this section, we first prove Camp-Meidell's inequality constructively so that we can extend this result in two directions. We consider $\sup_{F\in\Omega_1}\phi_n(F;c)$, where $F\in\Omega_1$ if F corresponds to a unimodal density with $F\in\Omega$. We then turn to the problem of maximizing $\phi_1(F;c)$ over all $F\in\Omega_2(s)$, where $F\in\Omega_2(s)$ if $F\in\Omega_1$ and $|F'(x)|\leqslant s$.

LEMMA 4.1. *Let X be a symmetric absolutely continuous random variable with $\mu=0$, $\sigma^2=1$ and unimodal density function $f(x)$. Then $P[|X-\mu|\geqslant c\sigma]$ is maximized when X is a mixture of point mass at zero and $U(-k,k)$.*

Proof. Without loss of generality we can assume that $\mu = 0$ and $\sigma^2 = 1$. By symmetry, we only need to consider $P[X \geqslant c]$. Let Y be a random variable with decreasing density over $(0, \infty)$. Let $g(y)$ be the constant function on $(0, k)$ with value equal to $f_Y(c)$, where $k = c + P[Y \geqslant c]/f_Y(c)$. Let X be a mixture of $U(0, k)$ with probability $f_Y(c)k$ and point mass at zero with probability $.5 - f_Y(c)k$. Since $f_Y(x) \geqslant f_Y(c)$ for $c \leqslant x \leqslant k$, we have $kf_Y(c) \leqslant P[Y \geqslant 0] = .5$. Also $\mathrm{E}(X^2) \leqslant .5$. Hence, we can increase k and reduce the probability that X equals zero either until $\mathrm{E}(X^2) = .5$ or $P[X = 0] = 0$. If the latter occurs first, we can then reduce the height of the uniform and increase k to obtain $U(0, k')$, where $k' = \sqrt{3}$.

COROLLARY 4.2. *If X is an absolutely continuous symmetric unimodal random variable, then*

$$P[|X-\mu| \geqslant c\sigma] \leqslant \begin{cases} 1 - c/\sqrt{3} & \text{if } c < 2/\sqrt{3}, \\ (4/9)c^{-2} & \text{if } c \geqslant 2/\sqrt{3}. \end{cases}$$

Proof. The results follow from Lemma 4.1 and the best choice of k.

Remark. If $c = 2$, the bound is 1/9 which is much closer to .05 than Chebychev's bound.

We can follow the same approach as in Section 3 to obtain a lower limit for bounds on ϕ_n in (3.1) for symmetric absolutely continuous unimodal random variables. Let X_i be a mixture of $U(-b\sqrt{n}, b\sqrt{n})$ with probability $3/(b^2 n)$ and point mass at zero with probability $1 - 3/(b^2 n)$. As $n \to \infty$, the number $\tilde{M}$ of X_i that are not zero has a Poisson distribution with $\mu = 3/b^2$. Hence

$$\text{(4.1)} \quad \lim_{n\to\infty} P\Big[\Big|\sum_{i=1}^{n} X_i\Big| \geqslant k\sqrt{n}\Big] = \lim_{n\to\infty} \sum_{m=1}^{\infty} P\Big[\Big|\sum_{i=1}^{m} Y_i\Big| \geqslant k\sqrt{n}\Big] e^{-\mu}\mu^m/m!$$
$$\geqslant \mu e^{-\mu}(b-k)/b + .5\mu^2 e^{-\mu}(2b-k)^2/4b^2$$
$$+ .5\big(1 - e^{-\mu}(1+\mu+\mu^2/2)\big)(2b-k)^2/4b^2,$$

where Y_i are i.i.d. $U(-b\sqrt{n}, b\sqrt{n})$. The first term on the right-hand side comes from $P[|Y_1| \geqslant k\sqrt{n}]$ and the second term comes from $P[|Y_1 + Y_2| \geqslant k\sqrt{n}]$, where $Y_1 + Y_2$ has a triangular density. Note that neither term depends on n. The last term comes from the inequality that, for any $j \geqslant 3$,

$$P\big[|Y_1 + \ldots + Y_j| \geqslant k\sqrt{n}\big]$$
$$\geqslant P\big[|Y_1 + Y_2| \geqslant k\sqrt{n}\big]\, P[Y_3 + \ldots + Y_j \text{ has the same sign as } Y_1 + Y_2].$$

The probability that the signs agree is .5.

Remark 1. If $c = 2$ and $b = 3$, as in Camp-Meidell's inequality, then the right-hand side of (4.1) is approximately .099, which is close to .111.

Remark 2. We can of course get a tighter result on (4.1) by considering the probability relating to the convolution of three uniforms. If $c=2$ and $b=3$, however, then $P(\tilde{N} \geqslant 3) = .0048$, so the $n=3, 4, \ldots$ terms are negligible.

Remark 3. Let $b = 1.5c$ as in Camp-Meidell's inequality. As $c \to \infty$, the right-hand side of (4.1) behaves as

$$\mu e^{-\mu}(b-c)/b = \tfrac{4}{9}c^2 e^{-4/3c^2}.$$

But

$$\frac{\tfrac{4}{9}c^2 e^{-4/3c^2}}{\tfrac{4}{9}c^2} \to 1 \quad \text{as } c \to \infty.$$

So for large c the ratio of the bound for infinitely divisible random variables and Camp-Meidell (which is sharp for $n=1$) goes to one.

Let $f \in \Omega$ if f is a symmetric (about zero without loss of generality) unimodal density function with $\int_{-\infty}^{\infty} x^2 f(x)dx = 1$ as above. Consider

$$\gamma_f(c) = \int_c^{\infty} f(x)dx. \tag{4.2}$$

Camp-Meidell shows that $\sup_{f \in \Omega} \gamma_f(c) = P[X_0 \geqslant c]$, where X_0 is of the form: mixture of $U(-k, k)$ and point mass at zero with respective probabilities $3/k^2$ and $1-3/k^2$. Strictly speaking, X_0 is not an absolutely continuous r.v. but there exists $f_\varepsilon \in \Omega$ such that

$$\lim_{\varepsilon \to 0} \int_c^{\infty} f_\varepsilon(x)dx = P[X_0 \geqslant c].$$

The $\sup_{f \in \Omega} \gamma_f(c)$ can be reduced by restricting Ω to Γ, a subset of Ω. One logical course would be to place a restriction on $f'(x)$. To this end, we define $\Gamma(s)$, where $f \in \Gamma(s)$ if $f \in \Omega$ and $-f'(x) \leqslant s$ for $x > 0$, where $s > 0$. Note that the restriction of f' makes sense since f is decreasing for $x > 0$. Of course, we need only consider $f(x)$ for $x > 0$ by symmetry.

We state the following known result since it is used throughout the ensuing discussion.

LEMMA 4.3. *If* $\int_t^u f(x)dx = \int_t^u g(x)dx$ *and there exists* $x_0 \in [t, u]$ *such that* $f(x) \geqslant g(x)$ *for* $t \leqslant x \leqslant x_0$ *and* $f(x) \leqslant g(x)$ *for* $x_0 \leqslant x \leqslant u$, *then*

$$\int_t^u h(x)f(x)dx \leqslant \int_t^u h(x)g(x)dx, \tag{4.3}$$

where h *is any non-decreasing function.*

Lemma 4.3 implies that s must be at least 1/6. This follows by considering

$$f_0(x) = \begin{cases} s(b-x) & \text{for } 0 \leqslant x \leqslant b, \\ 0 & \text{for } x > b; \end{cases}$$

then b must be $s^{-1/2}$, which implies $\mathrm{E}(X^2) = 1/(6s)$.

The proposition of interest uses the following subset of functions $H(s)$ contained in $\Gamma(s)$, i.e., $h(x;\, a, b, d) \in H(s)$, where

$$h(x;\, a, b, d) = \begin{cases} s(a-x)+s(d-b) & \text{for } 0 \leqslant x \leqslant a, \\ s(d-b) & \text{for } a \leqslant x \leqslant b, \\ s(d-x) & \text{for } b \leqslant x \leqslant d, \\ 0 & \text{for } x > d. \end{cases} \tag{4.4}$$

Obviously, there are constraints on a, b and d since

$$\int_0^\infty h(x;\, a, b, d)dx = 1/2 \quad \text{and} \quad \int_0^\infty x^2 h(x;\, a, b, d)dx = 1/2.$$

Straightforward calculus gives

$$s(d^2+a^2-b^2) = 1 \tag{4.5}$$

and

$$s(d^4+a^4-b^4) = 6. \tag{4.6}$$

Letting $x = a^2$, $y = b^2$ and $z = d^2$ it follows from (4.5) and (4.6) that

$$y = z - k/(zs-1) \tag{4.7}$$

and

$$x = (1/s) - k/(zs-1), \tag{4.8}$$

where $k = 3 - 1/(2s)$.

We are now prepared to prove the main result.

LEMMA 4.4. *$\sup_{f\in\Gamma(s)} \gamma_f(c)$ occurs if $f \in H(s)$.*

Proof. Let g be any element in $\Gamma(s)$. There exists an $a \in [0, c]$ such that

$$\int_0^c h_1(x;\, a)dx = \int_0^c g(x)dx,$$

where

$$h_1(x;\, a) = \begin{cases} s(a-x)+g(c) & \text{for } 0 \leqslant x \leqslant a, \\ g(c) & \text{for } a \leqslant x \leqslant c. \end{cases}$$

This is so since $h_1(x;\, 0) \leqslant g(x)$ for all $x \in [0, c]$ and $h_1(x;\, c) \geqslant g(x)$ for all $x \in [0, c]$. Similarly, there exists a $b \in [c, \infty]$ such that

$$\int_c^\infty h_2(x; b)dx = \int_c^\infty g(x)dx,$$

where

$$h_2(x; b) = \begin{cases} g(c) & \text{for } c \leqslant x \leqslant b, \\ s(b-x)+g(c) & \text{for } b \leqslant c \leqslant b+g(c)/s, \\ 0 & \text{for } x > b+g(c)/s. \end{cases}$$

This is so since $h_2(x; c) \leqslant g(x)$ for $x \geqslant c$ and $h_2(x, \infty) \geqslant g(x)$ for $x \geqslant c$. We can now consider

$$h_0(x, a, b) = \begin{cases} h_1(x; a) & \text{if } 0 \leqslant x \leqslant c, \\ h_2(x; b) & \text{if } c < x. \end{cases} \tag{4.9}$$

It follows that

$$\int_0^\infty h_0(x; a, b)dx = \int_0^\infty g(x)dx = 1/2,$$

$$\int_c^\infty h_0(x; a, b)dx = \int_c^\infty g(x)dx \quad \text{and} \quad \int_0^\infty x^2 h_0(x; a, b)dx \leqslant \int_0^\infty x^2 g(x)dx.$$

The inequality holds by Lemma 4.3 and the facts that there exists $x_1 \in [u, c]$ such that $h_1(x; a) \geqslant g(x)$ for $0 \leqslant x \leqslant x_1$ and $h_1(x; a) \leqslant g(x)$ for $x_1 \leqslant x \leqslant c$ and there exists $x_2 \in [c, \infty]$ such that $h_2(x, b) \geqslant g(x)$ for $c \leqslant x \leqslant x_2$ and $h_2(x, b) \leqslant g(x)$ for $x_2 \leqslant x$. We can now alter $h_0(x; a, b)$ in (4.9) by increasing b and reducing a so that the area constraint is maintained until $\int_0^\infty x^2 h_0(x; a, b)dx = 1/2$. This adjustment increases $\int_c^\infty h_0(x; a, b)dx$. The only concern is whether $\int_0^\infty x^2 h_0(x; 0, b)dx > 1/2$ so that no $a \geqslant 0$ exists satisfying the second moment constraint.

Consider $h(x; 0, b, d)$ as defined by equation (4.4). If $h(0; 0, b, d) \geqslant g(c)$, then

$$\int_0^\infty x^2 h_0(x; b, d)dx \geqslant \int_0^\infty x^2 h(x; 0, b, d) = 1/2$$

by Lemma 4.3. If $h(0; 0, b, d) < g(c)$, then $h(x; 0, b, d) \leqslant g(x)$ for all $0 \leqslant x \leqslant c$, and so

$$\int_c^\infty h(x; 0, b, d)dx > \int_c^\infty g(x)dx.$$

Hence g cannot be optimal.

The solution to the calculus problem of finding $a_0 \leqslant c \leqslant b_0 \leqslant d_0$ such that

$$\int_c^\infty h(x; a_0, b_0, d_0)dx \geqslant \int_c^\infty h(x; a, b, d)dx \quad \text{for all } a \leqslant c \leqslant b \leqslant d$$

appears in the working paper.

We illustrate the results for $c = 2$, that is we want $P[|X-\mu| \geqslant 2\sigma]$. Chebychev's bound is .25 and Camp-Meidell's bound is 1/9. The bound here depends on s (see Table 4.1). We observe that if the underlying density has slope no steeper than the maximum slope of the density of the normal, then the bound of .067 is close to .05.

Table 4.1. Bounds on $P[|X| \geqslant 2]$

s	a	b	d	$P[|X| \geqslant 2]$
.2	1.864	2.262	2.577	.05284
.2419707 (1)	1.699	2.446	2.689	.06664
.3	1.522	2.587	2.777	.07747
.4	1.312	2.710	2.850	.08733
.6	1.065	2.816	2.909	.09608
.8	.920	2.868	2.938	.10011
1.0	.822	2.896	2.951	.10243
1.5	.670	2.936	2.973	.10543
2	.579	2.949	2.977	.10689
5	.365	2.972	2.983	.10945
∞	0	3.0	3.0	.11111

(1) .2419707 corresponds to $(1/\sqrt{2\pi})e^{-1/2}$ which is the maximum slope for a standard normal density.

REFERENCES

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

[2] S. Bernstein, *Sur une modification de l'inégalité de Tchebichef*, Ann. Sci. Instit. Sav. Ukraine, Sect. Math., I (1924).

[3] P. J. Bickel and D. A. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist. 9 (1981), pp. 1196–1217.

[4] — *Bootstrapping Regression Models with Many Parameters (In a festschrift for Erich L. Lehmann)*, Wadsworth, Belmont, California, 1983.

[5] Z. W. Birnbaum, J. Raymond and H. S. Zuckerman, *A generalization of Tchebychev's inequality to two dimensions*, Ann. Math. Statist. 18 (1947), pp. 70–79.

[6] B. Efron, *Bootstrap methods: Another look at the Jackknife*, Ann. Statist. 7 (1979), pp. 1–26.

[7] D. A. Freedman, *Bootstrapping regression models*, ibidem 9 (1981), pp. 1218–1228.

[8] I. I. Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes*, W. B. Saunders, Philadelphia 1965.

[9] H. J. Godwin, *On generalizations of Tchebycheff's inequality*, J. Amer. Statist. Assoc. 50 (1955), pp. 923–945.

[10] W. Hoeffding, *The extrema of the expected value of a function of independent random variables*, Ann. Math. Statist. 26 (1955), pp. 268–275.

[11] — *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. 58 (1963), pp. 113–129.

[12] — and S. S. Shrikhande, *Bounds for the distribution of a sum of independent identically distributed random variables*, Ann. Math. Statist. 26 (1955), pp. 439–449.

[13] P. Huber, *Robust regression: Asymptotic conjectures and Monte Carlo*, Ann. Statist. 1 (1973), pp. 799–821.

[14] S. Karlin and W. J. Studden, *Tchebycheff-Systems with Applications in Analysis and Statistics*, Interscience, New York 1966.

[15] H. Robbins, *Some remarks on the inequality of Tchebycheff*, in: *Studies and Essays (presented to R. Courant on his 60th birthday)*, Interscience, New York 1948.

[16] S. M. Samuels, *On a Chebychev-type inequality for sums of independent random variables*, Ann. Math. Statist. 37 (1966), pp. 248–259.

[17] K. Sharpe, *Robustness of normal tolerance intervals*, Biometrika 57 (1970), pp. 71–78.

Department of Statistics
University of California
Berkeley, USA

Department of Statistics
The Wharton School
of the University of Pennsylvania
3000 Steinberg Hall–Dietrich Hall
Philadelphia, PA 19104-6302, USA

Received on 2.4.1992