The Annals of Applied Probability
1995, Vol. 5, No. 3, 854–873

PRECISION CALCULATION OF DISTRIBUTIONS FOR TRIMMED SUMS[1]

BY SÁNDOR CSÖRGŐ AND GORDON SIMONS

University of Michigan and University of North Carolina, Chapel Hill

Recursive methods are described for computing the frequency and distribution functions of trimmed sums of independent and identically distributed nonnegative integer-valued random variables. Surprisingly, for fixed arguments, these can be evaluated with just a *finite* number of arithmetic operations (and whatever else it takes to evaluate the common frequency function of the original summands). These methods give rise to very accurate computational algorithms that permit a delicate numerical investigation, herein described, of Feller's weak law of large numbers and its trimmed version for repeated St. Petersburg games. The performance of Stigler's theorem for the asymptotic distribution of trimmed sums is also investigated on the same example.

1. Introduction. Trimmed sums of iid (independent and identically distributed) random variables appear in many contexts. Applied statisticians use them to improve estimators when the parent distribution has a heavy tail. [See, e.g., David (1981), pages 158–163, and for a list of relevant references, see Stigler (1973).] Probabilists, who have studied them extensively, have clearly documented the heavy influence, in some settings, of the largest observation(s). [Extensive reference lists can be found in a recent book edited by Hahn, Mason and Weiner (1991).]

We are concerned here with describing effective recursive methods for computing the frequency and distribution functions for trimmed sums of iid nonnegative integer-valued random variables. Not only do such methods exist, but, as we shall see, there exist methods that can be fully implemented with just a *finite* number of arithmetic operations. The model for this is provided by the convolution-based recursion for untrimmed sums

$$P\{S_n = s\} = \sum_{k=0}^{s} P\{X = k\}\, P\{S_{n-1} = s - k\}, \tag{1}$$

and by the simple sum

$$P\{S_n \le s\} = \sum_{k=0}^{s} P\{S_n = k\}, \qquad s = 0, 1, 2, \dots, \tag{2}$$

where $S_n := X_1 + \cdots + X_n$ is the sum of n iid random variables distributed as X. Both sums (1) and (2) contain just a finite number of terms.

Received April 1994; revised January 1995.

[1]Research supported in part by NSF Grant DMS-92-08067.

AMS 1991 *subject classifications*. Primary 60E99; secondary 60-04.

Key words and phrases. Trimmed sums on nonnegative integers, recursive algorithms for distributions, St. Petersburg game.

Turning to trimmed sums, let $S_n(m)$ denote the same sum but with the m largest summands excised, $m \le n$, that is, let

$$S_n(m) := X_{n,1} + X_{n,2} + \cdots + X_{n,n-m}, \qquad m = 0, 1, 2, \ldots, n,$$

where $X_{n,1} \le X_{n,2} \le \cdots \le X_{n,n}$ denote the order statistics for $X_1, X_2, \ldots, X_n$, so that $S_n(0) = S_n$. [Throughout, $S_0 := 0$ and $S_n(n) := 0$.] While, clearly, the distribution function $P\{S_n(m) \le s\}$ can be obtained by a finite sum, as in (2), the frequency function $P\{S_n(m) = s\}$, for $m \ge 1$, cannot be computed via a simple analogue of (1).

Consider the special case $m = 1$. A simple recursion in n, described in Theorem 3 below, *requiring nothing but finite sums*, links the functions $P\{S_n(1) = s, X_{n,n} = t\}$, $s, t = 0, 1, 2, \ldots$, $n \ge 1$. However, this approach leads to an *infinite* sum

$$P\{S_n(1) = s\} = \sum_{t=0}^{\infty} P\{S_n(1) = s,\ X_{n,n} = t\}. \tag{3}$$

Fortunately, this shortcoming can be finessed with an application of the following theorem.

THEOREM 1. *For integers $n \ge m \ge 0$ and $r \ge s \ge 0$,*

$$P\{S_n(m) = s,\ X_{n,n} \le r\} = \sum_{k=0}^{m} (-1)^k \binom{n}{k} [1-F(r)]^k P\{S_{n-k}(m-k) = s\}, \tag{4}$$

where $F(r) := P\{X \le r\}$ is the distribution function of X.

Combining (3) and (4), for the case $m = 1$, leads to the finite sum: for integers $n \ge 1$ and $r \ge s \ge 0$,

$$P\{S_n(1) = s\} = \sum_{t=0}^{r} P\{S_n(1) = s, X_{n,n} = t\} + n[1 - F(r)]P\{S_{n-1} = s\}. \tag{5}$$

The same trick works for a general $m \le n$. For instance, for $n \ge m = 2$ and $r \ge s \ge 0$,

$$\begin{aligned} &P\{S_n(2) = s\} \\ &\quad = \sum_{v=0}^{r} \sum_{u=0}^{v} P\{S_n(2) = s, X_{n,n-1} = u, X_{n,n} = v\} \\ &\qquad + n[1 - F(r)]P\{S_{n-1}(1) = s\} - \frac{n(n-1)}{2}[1 - F(r)]^2 P\{S_{n-2} = s\}. \end{aligned} \tag{6}$$

We readily concede the point to any critic who would argue, at this point, that it is possible, with proper care, to throw away an infinite number of small summands without introducing a substantial amount of error. This is true, but we would make three rejoinders:

1. Precise computations are more easily achieved when the issue of truncation does not arise (or is circumvented).

2. The exercise of "proper care" with a formula like (3) requires more memory, than with (5), to accomplish comparable accuracy. This can be a significant issue.
3. Recursive methods tend to propagate errors. Thus the "proper care" sufficient to handle the case $n = 10$, for instance, might not be adequate when the same calculations are extended to $n = 100$.

We have used the methods described herein in various settings. Based on considerable experience, we feel quite confident that they yield excellent results, when performed with double precision arithmetic, even when n assumes values in the low thousands. For example, we have obtained essentially identical results when using equation (5) with various values of r, the free choice of which is a potential means for checking accuracy. However, a careful error analysis has not been made.

The presence of mixed signs on the right of (6) can be overcome, in order to avoid potential losses of computational accuracy, by applying (5) to the next to last term in (6) to obtain

$$\begin{aligned} P\{S_n(2) = s\} &= \sum_{v=0}^{r} \sum_{u=0}^{v} P\{S_n(2) = s,\ X_{n,n-1} = u,\ X_{n,n} = v\} \\ &\quad + n[1 - F(r)] \sum_{t=0}^{r} P\{S_{n-1}(1) = s,\ X_{n-1,n-1} = t\} \\ &\quad + \frac{n(n-1)}{2}[1 - F(r)]^2 P\{S_{n-2} = s\}. \end{aligned} \tag{7}$$

Everything on the right side of (7), other than the factor $1 - F(r)$, can be evaluated *without subtracting terms of positive sign*. (A check of the recursion described in Theorem 3 below is required to verify this assertion.) Formula (7) is also a simple consequence, for the case $m = 2$, of Theorem 2 below, which, in a sense, inverts (4).

THEOREM 2. *For integers $n \geq m \geq 0$ and $r \geq s \geq 0$,*

$$P\{S_n(m) = s\} = \sum_{k=0}^{m} \binom{n}{k} [1 - F(r)]^k P\{S_{n-k}(m-k) = s,\ X_{n-k,n-k} \leq r\}. \tag{8}$$

Section 2 gives a proof of Theorems 1 and 2, and describes in Theorem 3 a recursion for general $m \geq 1$, which yields intermediate probabilities such as $P\{S_n(1) = s,\ X_{n,n} = t\}$ and $P\{S_n(2) = s,\ X_{n,n-1} = u,\ X_{n,n} = v\}$ in (5) and (7). Section 3 discusses some illustrative applications, which are of relevance to an ongoing study by the authors of the "St. Petersburg paradox."

2. Theory. Here, we assume the notation appearing in the Introduction. We begin with the proofs of Theorems 1 and 2. Then we describe in Theorem 3 an essential recursion, for general m, which, with the recursion in either Theorem 1 or 2, leads to a general scheme, in the spirit of (6) or (7), for computing $P\{S(m) = s\}$, for any integer $s \geq 0$, with just a finite number of arithmetic operations.

PROOF OF THEOREM 1. There is really nothing to prove when $m = 0$; (4) reduces to $P\{S_n = s, \ X_{n,n} \leq r\} = P\{S_n = s\}$, which is obvious when $r \geq s$.

For fixed integers $n \geq m \geq 1$ and $r \geq s \geq 0$, let $A_i = A_i(n,m)$ denote the event $A_i(n,m) := \{S_n(m) = s, \ X_i > r\}$, $i = 1, \dots, n$. Then, by inclusion and exclusion,

$$
\begin{aligned}
&P\{S_n(m) = s, X_{n,n} > r\} \\
&\quad = P\left\{\bigcup_{i=1}^{n} A_i\right\} = \sum_{k=1}^{n}(-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P\left\{\bigcap_{j=1}^{k} A_{i_j}\right\} \\
&\quad = \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k} P\left\{\bigcap_{i=1}^{k} A_i\right\} \\
&\quad = \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k} P\{X_i > r, \ 1 \leq i \leq k, \text{ and the sum of the} \\
&\qquad\qquad n - m \text{ smallest among } X_{k+1}, \dots, X_n \text{ equals } s\} \\
&\quad = \sum_{k=1}^{m}(-1)^{k-1}\binom{n}{k}[1 - F(r)]^k P\{S_{n-k}(m-k) = s\},
\end{aligned}
$$

where the assumption that $r \geq s$ is essential for the fourth equality. Since the probability $P\{S_n(m) = s, \ X_{n,n} > r\} = P\{S_n(m) = s\} - P\{S_n(m) = s, \ X_{n,n} \leq r\}$, we see that the desired probability $P\{S_n(m) = s, X_{n,n} \leq r\}$ is

$$
\begin{aligned}
P\{S_n(m) = s\} &+ \sum_{k=1}^{m}(-1)^{k}\binom{n}{k}[1 - F(r)]^k P\{S_{n-k}(m-k) = s\} \\
&= \sum_{k=0}^{m}(-1)^{k}\binom{n}{k}[1 - F(r)]^k P\{S_{n-k}(m-k) = s\},
\end{aligned}
$$

proving the theorem. □

PROOF OF THEOREM 2. While this theorem can be viewed as a corollary of Theorem 1, with (8) following from (4) by a direct combinatorial calculation, it is more instructive to present a straightforward probabilistic argument. The statement is trivial if $n = m$. Fix $n > m \geq 0$ and $r \geq s \geq 0$, and let $K_n(r)$ denote the number of $X_i > r$, $1 \leq i \leq n$. Then

$$P\{S_n(m) = s\} = \sum_{k=0}^{m} P\{S_n(m) = s, K_n(r) = k\}.$$

Now, for $0 \leq k \leq m$, introduce the events $B_k := \{X_i > r \text{ for } n - k < i \leq n\}$ and $C_k :=$ {the $n - m$ smallest among $X_1, \dots, X_{n-k}$ sum to s, $X_i \leq r$ for $1 \leq$

$i \le n-k\}$. Since there are $\binom{n}{k}$ ways of choosing exactly k of the X_i's to exceed r, $1 \le i \le n$, and thereby making $K_n(r) = k$, we have

$$\begin{aligned}
&P\{S_n(m) = s, K_n(r) = k\}\\
&\quad = \binom{n}{k} P\{B_k \cap C_k\} = \binom{n}{k} P\{B_k\}P\{C_k\}\\
&\quad = \binom{n}{k} [1 - F(r)]^k P\{S_{n-k}(m-k) = s,\ X_{n-k,n-k} \le r\}.
\end{aligned}$$

The two equations together complete the proof of (8). □

To proceed, we need some additional notation. For fixed $n \ge m \ge 1$, let

$$X_n(m) := (X_{n,n-m+1}, \dots, X_{n,n}).$$

When $X_n(m) = t$, then $t = (t_1, \dots, t_m)$, where $0 \le t_1 \le \cdots \le t_m$ are some integers. Given such a t, let $\underline{t}$ denote the smallest (the first) component of t, let $\{t\}$ denote the set of integer values appearing in t, without repetitions, and, for integers j and k, with $0 \le j \le \underline{t}$ and $k \in \{t\}$, let $t[j,k]$ denote the vector formed by augmenting t from the left with the integer j and deleting one of the k's appearing in t. Thus, $t[j,k]$ remains an m-dimensional vector with the same properties as t. Finally, for a given $t = (t_1, \dots, t_m)$ as above, set $\{t\}_* := \{0, 1, 2, \dots, \underline{t} - 1\}$ and let us agree that $\{t\}_* = \varnothing$ if $\underline{t} = 0$.

Clearly, $X_n \in \{t\}_* \cup \{t\}$ and $X_{n,n-m} \in \{0, 1, 2, \dots, \underline{t}\} = \{t\}_* \cup \{\underline{t}\}$ when $X_n(m) = t$. Furthermore, to handle an incoming new observation X_n at time n, we claim that for $n - 1 \ge m$, if $X_n(m) = t$, then

$$(9)\qquad \begin{aligned}
&(S_{n-1}(m), X_{n-1}(m))\\
&\quad = \begin{cases} (S_n(m) - X_n, t), & \text{if } X_n \in \{t\}_*,\\ (S_n(m) - X_{n,n-m}, t[X_{n,n-m}, X_n]), & \text{if } X_n \in \{t\}. \end{cases}
\end{aligned}$$

The first of these is obvious because a value of $X_n \in \{t\}_* = \{0, 1, 2, \dots, \underline{t} - 1\}$ cannot be in $\{t\}$, and hence $X_{n-1}(m) = X_n(m) = t$, and, for the same reason, the difference in the trimmed sums $S_n(m) - S_{n-1}(m)$ must be X_n. However, when $X_n \in \{t\}$, the new observation X_n is trimmed at time n, or, optionally, can be trimmed if it and some previous random variable are both equal to $\underline{t}$. If X_n is trimmed, the smallest member of $X_{n-1}(m)$, namely, $X_{n-1,(n-1)-m+1} = X_{n,n-m}$, must be deleted from the trimmed set at time n, because X_n is trimmed instead of it, and so $X_{n,n-m}$ is a term in $S_n(m)$, and hence $X_{n-1}(m) = t[X_{n,n-m}, X_n]$, and the difference $S_n(m) - S_{n-1}(m) = X_{n,n-m}$. Notice, for the optional case, that $S_n(m) - S_{n-1}(m)$ is $X_{n,n-m} = \underline{t} = X_n$ and $t[X_{n,n-m}, X_n] = t[\underline{t}, \underline{t}] = \{t\}$, so that either form of (9) can be used.

Understanding an empty sum as zero, using the indices j and k to represent the values of $X_{n,n-m}$ and X_n, respectively, and noting the independence of the vector $(S_{n-1}(m), X_{n-1}(m))$ and X_n, we are led to the following recursion.

THEOREM 3. *For integers $n > m \geq 1$ and $t_m \geq \cdots \geq t_1 = \underline{t}$, $s \geq 0$, with $t = (t_1, \ldots, t_m)$,*

$$P\{S_n(m) = s, X_n(m) = t\}$$

$$(10) \qquad = \sum_{k=0}^{\underline{t}-1} P\{X = k\} P\{S_{n-1}(m) = s - k,\ X_{n-1}(m) = t\}$$

$$+ \sum_{k \in \{t\}} P\{X = k\} \sum_{j=0}^{\underline{t}} P\{S_{n-1}(m) = s - j,\ X_{n-1}(m) = t[j,k]\}.$$

PROOF. If $X_n(m) = t$, the observation X_n is confined to the set $\{t\}_* \cup \{t\} = \{0, 1, 2, \ldots, \underline{t} - 1\} \cup \{t\}$. The values $X_n = k \in \{t\}_* = \{0, 1, 2, \ldots, \underline{t} - 1\}$ give rise to the first sum on the right-hand side of (10):

$$P\{S_n(m) = s,\ X_n(m) = t,\ X_n = k\}$$
$$= P\{X = k\} P\{S_{n-1}(m) = s - k,\ X_{n-1}(m) = t\},$$

in accordance with the first case in (9). The values $X_n = k \in \{t\}$ give rise to the double sum on the right-hand side of (10) as follows:

$$P\{S_n(m) = s,\ X_n(m) = t,\ X_n = k\}$$

$$= \sum_{j=0}^{\underline{t}} P\{S_n(m) = s,\ X_n(m) = t,\ X_n = k,\ X_{n,n-m} = j\}$$

$$= \sum_{j=0}^{\underline{t}} P\{S_n(m) = s,\ X_{n-1}(m) = t[j,k],\ X_n = k\}$$

$$= \sum_{j=0}^{\underline{t}} P\{S_{n-1}(m) = s - j,\ X_{n-1}(m) = t[j,k],\ X_n = k\}$$

$$= P\{X = k\} \sum_{j=0}^{\underline{t}} P\{S_{n-1}(m) = s - j,\ X_{n-1}(m) = t[j,k]\},$$

in accordance with the second case in (9). This establishes (10). □

3. Applications to the St. Petersburg game. The context of the St. Petersburg paradox is a game, based on a sequence of fair coin tosses, in which Peter agrees to pay Paul $X = 2^k$ ducats, where k is the number of tosses required to produce the first head, so that $P\{X = 2^k\} = 2^{-k}$, $k = 1, 2, \ldots$. The simple fact that Paul's expected winnings, $E(X)$, is infinite provides the basis for the paradox. For as Nicolaus Bernoulli, who posed the problem in 1713, wrote in 1728 to his younger cousin Daniel, "... there ought not to exist any even halfway sensible person who would not sell the right of this gain for forty ducats." [The original numbers of ducats are doubled here and everywhere in our discussion to conform with a more convenient payoff scheme used by many subsequent writers. The translation from the Latin is taken from Martin-Löf

(1985); we like it better than the standard form in the English translation of Bernoulli (1738), where Daniel cites Nicolaus' letter. See Jorland (1987) and Dutka (1988) for recent historical accounts.]

3.1. *Laws of large numbers.* Despite the attention of many well known mathematicians, stretching over a quarter of a millennium, a significant mathematical treatment of the subject did not occur until Feller (1945) addressed the topic, arguing that the question of "Paul's fair price" only makes sense when one considers a sequence of independent St. Petersburg games, with payoffs $X_1, X_2, \ldots$ distributed as X, and asks what a "fair price" would be for playing n such games. Addressing this issue, he showed [see also Feller (1950)] that

$$\frac{S_n}{n \operatorname{Log} n} \to 1 \quad \text{in probability as } n \to \infty,$$

where $S_n := X_1 + \cdots + X_n$ and $\operatorname{Log} n$ denotes the base 2 logarithm of n. Subsequently, Chow and Robbins (1961) showed that the convergence in Feller's law cannot be upgraded to almost sure convergence. Indeed, it can easily be shown that $P\{X_n > c\, n \operatorname{Log} n \text{ infinitely often}\} = 1$ for every $c > 0$. On the other hand, the authors [Csörgő and Simons (1994)] have shown for every $m \geq 1$ that

$$\frac{S_n(m)}{n \operatorname{Log} n} \to 1 \quad \text{almost surely as } n \to \infty.$$

These facts suggest that $p_n(\varepsilon) := P\{S_n > (1+\varepsilon)\, n \operatorname{Log} n\}$ might go to zero quite slowly with n, and its trimmed analogue, $p_n(m, \varepsilon) := P\{S_n(m) > (1+\varepsilon)\, n \operatorname{Log} n\}$ might converge to zero more rapidly when $m \geq 1$. This conjecture is investigated numerically for $m = 1$ in Figures 1 and 2, which contain overlying plots of $p_n(\varepsilon)$ and $p_n(1, \varepsilon)$ for $\varepsilon = 0.25$ and $\varepsilon = 1$, respectively.

The plots in the untruncated case are based on the simple recursion appearing in (1) and (2), made somewhat easier to compute by the fact that the index k in (1) is restricted to integer powers of 2. Frankly, we were initially surprised that this simple recursion could be run out to n values well into the thousands, with high accuracy maintained, and without major difficulties.

The calculations for $m = 1$ and $n = 1, \ldots, N$ proceed as follows: Beginning with

$$P\{S_1(1) = s,\ X_{1,1} = 2^t\} = \frac{I_{\{0\}}(s)}{2^t}, \qquad t \geq 1, \quad \text{and} \quad P\{S_1(1) = s\} = I_{\{0\}}(s),$$

where $I_{\{0\}}(s) = 0$ or 1 as $s \geq 1$ or $s = 0$, one computes $P\{S_n(1) = s, X_{n,n} = 2^t\}$ and $P\{S_n(1) = s\}$ successively for $n = 2, \ldots, N$ with the recursions

$$P\{S_n(1) = s,\ X_{n,n} = 2^t\} = \sum_{k=0}^{t-1} \frac{1}{2^k} P\{S_{n-1}(1) = s - 2^k,\ X_{n-1,n-1} = 2^t\} + \frac{1}{2^t} \sum_{j=0}^{t} P\{S_{n-1}(1) = s - 2^j,\ X_{n-1,n-1} = 2^j\},$$

for the integers $0 \le s \le r$, $0 \le t \le \lfloor \operatorname{Log} r \rfloor$, and

$$P\{S_n(1) = s\} = \sum_{t=0}^{\lfloor \operatorname{Log} r \rfloor} P\{S_n(1) = s,\ X_{n,n} = 2^t\} + \frac{n}{2^{\lfloor \operatorname{Log} r \rfloor}} P\{S_{n-1} = s\},$$

for $0 \le s \le r$, where $\lfloor x \rfloor := \max\{k = 0, \pm 1, \pm 2, \dots :\ k \le x\}$ is the usual integer part and, below, $\lceil x \rceil := \min\{k = 0, \pm 1, \pm 2, \dots :\ k \ge x\}$ is the upper integer part of $x \in \mathbb{R}$. Then

$$p_n(1, \varepsilon) = 1 - P\{S_n(1) \le (1+\varepsilon)\, n \operatorname{Log} n\} = 1 - \sum_{s=0}^{r_\varepsilon(n)} P\{S_n(1) = s\},$$

where $r_\varepsilon(n) := \lfloor (1+\varepsilon)\, n \operatorname{Log} n \rfloor$.

In order to do all the required calculations for n up to N, *with just one set of recursions*, one must work with a single $r \ge r_\varepsilon(N)$. Here, $N = 4096 = 2^{12}$ for Figures 1 and 2, resulting in $r_\varepsilon(N) = 61{,}440$ and $r_\varepsilon(N) = 98{,}304$ for $\varepsilon = 0.25$ and $\varepsilon = 1$, respectively.

While these calculations were carried out with good accuracy, a substantial memory burden was encountered that required the storage of approximately $\lfloor (1+\varepsilon) N (\operatorname{Log} N)^2 \rfloor$ double precision numbers: about 0.7 million when $\varepsilon = 0.25$ and about 1.2 million when $\varepsilon = 1$.

The horizontal axes in Figures 1 and 2 are expressed in units of $\operatorname{Log} n$, rather than n, in order to draw attention to empirical evidence indicating a link between the values of $p_n(\varepsilon)$ and $p_n(1, \varepsilon)$ and the location of n between

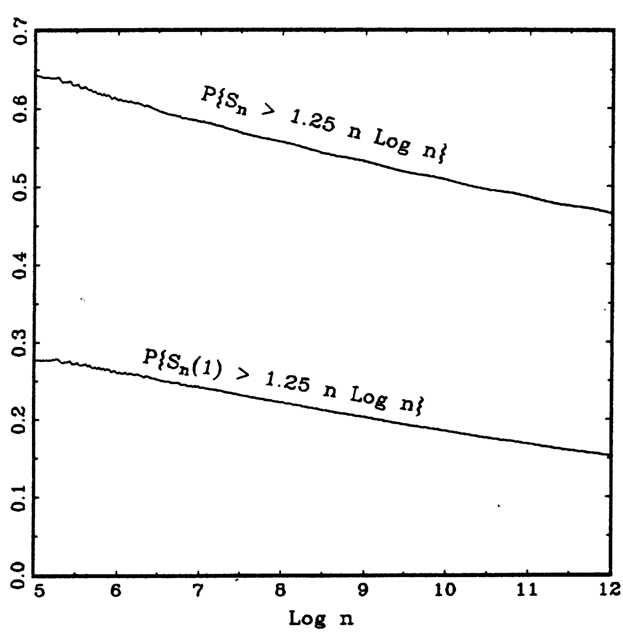

FIG. 1.

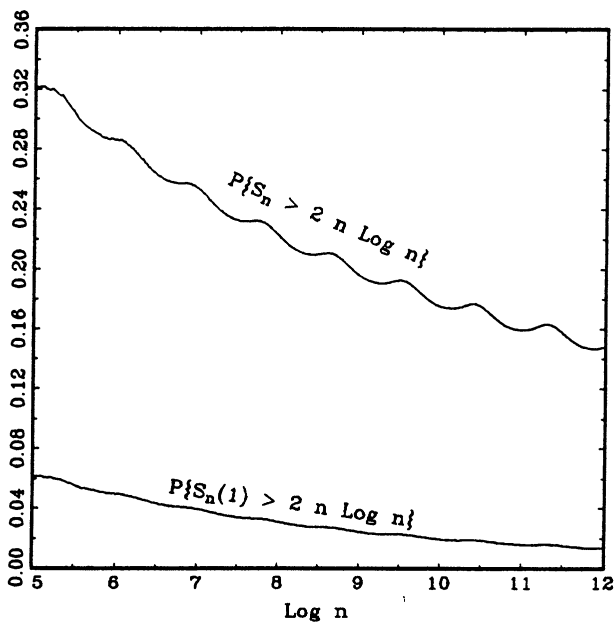

FIG. 2.

consecutive integer powers of 2. Theoretical support for this link is provided by the fact that the distribution functions of $(S_n - n\operatorname{Log} n)/n$ and $(S_n(1) - n\operatorname{Log} n)/n$ are asymptotically approximated, as $n \to \infty$, by the distribution functions of certain infinitely divisible random variables and their "trimmed" analogues, respectively, chosen on the basis of the value $\gamma_n := n/2^{\lceil \operatorname{Log} n\rceil}$, $1/2 < \gamma_n \le 1$. This is described in a forthcoming book by the authors.

Evidence of slow convergence to zero of both $p_n(\varepsilon)$ and $p_n(1,\varepsilon)$, and especially of $p_n(\varepsilon)$, is apparent in Figures 1 and 2. By methods outside the scope of the present paper, we can prove for every fixed $\varepsilon > 0$ that

$$1 \le \liminf_{n\to\infty}\, [\varepsilon \operatorname{Log} n]\, p_n(\varepsilon) \le \limsup_{n\to\infty}\, [\varepsilon \operatorname{Log} n]\, p_n(\varepsilon) \le 2. \tag{11}$$

Also, we have sufficient grounds to conjecture, but for the time being cannot prove, that for every fixed $\varepsilon > 0$,

$$\begin{aligned}\frac{1}{(m+1)!} &\le \liminf_{n\to\infty}[\varepsilon \operatorname{Log} n]^{m+1} p_n(m,\varepsilon)\\ &\le \limsup_{n\to\infty}[\varepsilon \operatorname{Log} n]^{m+1} p_n(m,\varepsilon) \le \frac{2^{m+1}}{(m+1)!}.\end{aligned} \tag{12}$$

Notice that (12) reduces to (11) when $m = 0$.

Figure 3 below provides numerical evidence, for four different values of ε, supporting the truth of (11), and it strongly indicates that the influence of the asymptotics arises quickly when ε is relatively large, and more slowly when

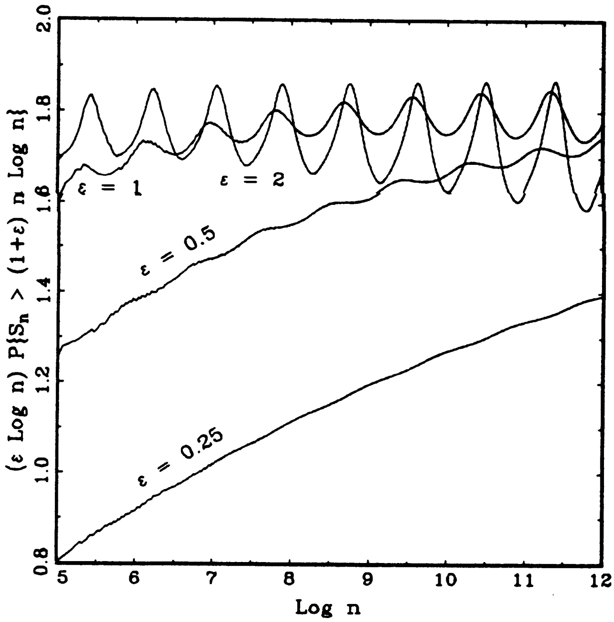

FIG. 3.

it is small. Moreover these graphs suggest that neither bound in (11) is tight. Working with the two values of ε for which we have data, we see evidence in Figure 4 that supports our conjecture in (12) for the particular case $m = 1$.

3.2. *Asymptotic distributions: Stigler's theorem.* One of the basic difficulties with the St. Petersburg game is that there is no limit theorem for the asymptotic distribution of Paul's cumulative gain S_n in the usual sense. This is because $P\{X > x\} = L(x)/x$ for any $x \geq 2$, where the oscillating function $L(x) = x/2^{\lfloor \operatorname{Log} x \rfloor} \in [1, 2)$ is not slowly varying at infinity, and so the St. Petersburg distribution is not in the domain of attraction of any (stable) distribution [cf. Gnedenko and Kolmogorov (1954), page 175, or Corollaries 1 and 3 in Csörgő, Haeusler and Mason (1988a)]. One is, therefore, tempted to trim the largest gains to see how Stigler's (1973) theorem works in this situation, that is, to look at the trimmed sums $S_n\{\beta\} := S_n(n - \lfloor \beta n \rfloor) = \sum_{j=1}^{\lfloor \beta n \rfloor} X_{j,n}$ for a fixed number $\beta \in (0, 1)$. As is heuristically clear, it turns out below that β has to be restricted to $[1/2, 1)$.

For $\beta \in (1/2, 1)$ define

$$\sigma^2(\beta) := \frac{3}{2^{\lfloor \operatorname{Log}(1-\beta) \rfloor}} - (1 - \lfloor \operatorname{Log}(1 - \beta) \rfloor)^2 - 2,$$

$$r(\beta) := -\frac{\beta + (1 - \beta) \operatorname{Log}(1 - \beta)}{\sqrt{3 - (1 - \beta)\{[1 - \operatorname{Log}(1 - \beta)]^2 + 2\}}},$$

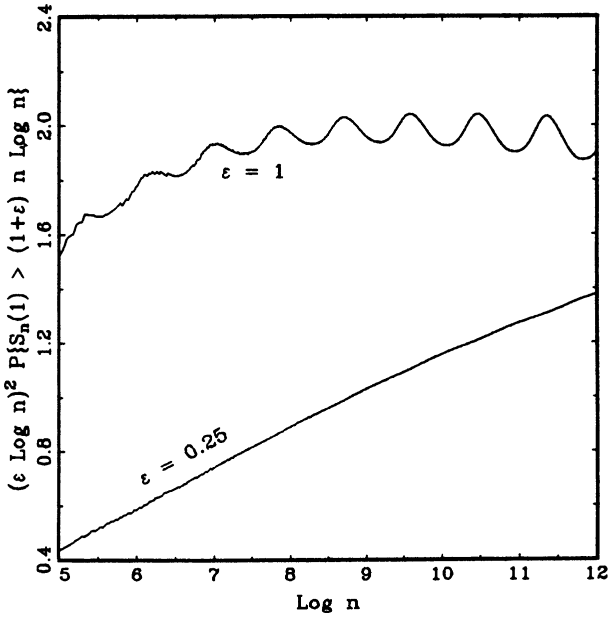

FIG. 4.

and set also

$$s^2(\beta) := \sigma^2(\beta) - 2\,\frac{\sigma(\beta)\,r(\beta)}{\sqrt{1-\beta}} + \frac{\beta}{1-\beta} \quad \text{and} \quad v^2(\beta) := \sigma^2(\beta)\,\frac{\beta - r^2(\beta)}{1-\beta}.$$

For some $q > 0$, let $\phi_{q^2}(\cdot)$ denote the $\mathrm{N}(0, q^2)$ density function and let $\Phi(\cdot)$ be the $\mathrm{N}(0,1)$ distribution function. If $\delta(t) := 1 + \langle \mathrm{Log}\, t\rangle - 2^{\langle \mathrm{Log}\, t\rangle}$, $t > 0$, where $\langle y\rangle = y - \lfloor y\rfloor$ is the fractional part of $y \in \mathbb{R}$, and $\to_{\mathscr{D}}$ denotes convergence in distribution as $n \to \infty$, Stigler's theorem for Paul's ordered winnings can be stated as follows:

CASE 1. *If $\beta \in (1/2, 1)$ and $-\mathrm{Log}(1-\beta)$ is not an integer, then*

$$\sqrt{n}\left[\frac{1}{n}\sum_{j=1}^{\lfloor \beta n\rfloor} X_{j,n} - \left\{\mathrm{Log}\,\frac{1}{1-\beta} + \delta(1-\beta)\right\}\right] \to_{\mathscr{D}} \sigma(\beta) Z,$$

where Z is a standard normal random variable.

CASE 2. *If $\beta \in (1/2, 1)$ and $-\mathrm{Log}(1-\beta)$ is an integer, then*

$$\sqrt{n}\left[\frac{1}{n}\sum_{j=1}^{\lfloor \beta n\rfloor} X_{j,n} - \mathrm{Log}\,\frac{1}{1-\beta}\right] \to_{\mathscr{D}} W_\beta := \sigma(\beta) Z + \frac{\max(0, -Z_*)}{\sqrt{1-\beta}},$$

where (Z, Z_*) is a bivariate normal vector with zero mean vector, variances $E(Z^2) = 1$ and $E(Z_*^2) = \beta$ and covariance $r(\beta)$. The density function $g_\beta(\cdot)$ of

W_β is given by

$$g_\beta(x) = \phi_{\sigma^2(\beta)}(x)\,\Phi\left(\frac{r(\beta)}{v(\beta)\sqrt{1-\beta}}\,x\right) + \phi_{s^2(\beta)}(x)\,\Phi\left(\frac{\beta-\sigma(\beta)\,r(\beta)\sqrt{1-\beta}}{(1-\beta)\,s(\beta)\,v(\beta)}\,x\right), \qquad x \in \mathbb{R}.$$

CASE 3. *For $\beta = 1/2$ we have*

$$\sqrt{n}\left[\frac{1}{n}\sum_{j=1}^{\lfloor n/2\rfloor} X_{j,n} - 1\right] \to_{\mathscr{D}} W_{1/2} := \max(0, Z),$$

where Z is standard normal, so that the distribution function $G_{1/2}(\cdot)$ of $W_{1/2}$ is

$$G_{1/2}(x) = \begin{cases} 0, & x < 0, \\ \Phi(x), & x \geq 0. \end{cases}$$

Here Case 3 follows directly from a half-sided version of Stigler's (1973) theorem, or from Theorem 1 in Csörgő, Haeusler and Mason (1991) as applied to $-X$, while the statements in Cases 1 and 2 were derived from the general "quantile" variant of Stigler's theorem in Theorem 5 of Csörgő, Haeusler and Mason (1988b) [cf. Csörgő and Dodunekova (1991)]. [Since $\sigma(1/2) = 0$, Case 3 may be viewed as a special case of Case 2, the latter formally extended for $\beta = 1/2$. Of course, $W_{1/2}$ does not have a Lebesgue density.] By an elementary combinatorial argument one can verify that

$$P\left\{\sum_{j=1}^{\lfloor\beta n\rfloor} X_{j,n} = 2\lfloor\beta n\rfloor\right\} = \frac{1}{2^n}\sum_{k=\lfloor\beta n\rfloor}^{n}\binom{n}{k} = 1 - \frac{1}{2^n}\sum_{k=0}^{\lfloor\beta n\rfloor-1}\binom{n}{k} \quad \text{for all fixed } \beta \in (0,1),$$

and $n \geq 1/\beta$. Then, letting $n \to \infty$ and using the symmetry of Pascal's triangle, we see that $P\{\sum_{j=1}^{\lfloor\beta n\rfloor} X_{j,n} = 2\lfloor\beta n\rfloor\} \to 0$ if $\beta \in (1/2, 1)$,

$$P\left\{\sum_{j=1}^{\lfloor n/2\rfloor} X_{j,n} = 2\left\lfloor\frac{n}{2}\right\rfloor\right\} = \frac{1}{2} + \frac{1}{2^{n+1}}\sum_{k=\lfloor n/2\rfloor}^{n-\lfloor n/2\rfloor}\binom{n}{k} \to \frac{1}{2} \quad \text{if } \beta = \frac{1}{2},$$

and $P\{\sum_{j=1}^{\lfloor\beta n\rfloor} X_{j,n} = 2\lfloor\beta n\rfloor\} \to 1$, if $\beta \in (0, 1/2)$. The first of these relations explains the lack of an atom in the limiting distributions in Cases 1 and 2, the second justifies the presence of it in Case 3, while the third shows that for $\beta \in (0, 1/2)$ the trimmed mean becomes asymptotically degenerate. In fact, an *ad absurdum* argument based on a joint application of both theorems in Csörgő, Haeusler and Mason (1991) to $-X$ again, also shows that when $\beta \in (0, 1/2)$, the trimmed sum $S_n\{\beta\} = \sum_{j=1}^{\lfloor\beta n\rfloor} X_{j,n}$ has a degenerate asymptotic distribution for arbitrary deterministic centering and norming sequences.

The issues of computing for $S_n(m)$ for $m > 1$ are similar to those for $S_n(1)$. Roughly speaking, increasing m by 1 increases the memory requirement by a factor of $(m + \operatorname{Log} r)/(m+1)$ for a given n. For $n = 100$ we were able to go up to $m = 4$ without substantial memory problems; larger m's were possible for a smaller n.

Let us write, in obvious notation, the Stigler approximations for the trimmed St. Petersburg sums in all three Cases 1–3 as $P\{(S_n\{\beta\} - n\,c\{\beta\})/\sqrt{n} \le x\} \approx G_\beta(x)$, for all $x \in \mathbb{R}$ except for $x = 0$ for the case $\beta = 1/2$ in Case 3. With selected values of n and β, Figures 5–14 illustrate some typical findings for the accuracy of the resulting approximations for the distribution function

$$F_{n,\beta}(s) := P\{S_n\{\beta\} \le s\} = P\{S_n(m) \le s\} \approx H_{n,\beta}(s)$$
$$:= G_\beta((s - n\,c\{\beta\})/\sqrt{n}),$$

for all $s \in \mathbb{R}$ except for $s = n\,c\{1/2\} = n$ in Case 3, where $m = n - \lfloor \beta n \rfloor$. Figures 5–10 are for the normal approximation from Case 1, Figures 11 and 12 for the Case 2 approximation and Figures 13 and 14 illustrate the case $\beta = 1/2$ in Case 3. Quite naturally, the general picture is that all three types of approximation improve on the right tail if β is kept fixed and n grows and if n is kept fixed and β decreases.

Figures 5 and 6 are typical examples for large β and $-\operatorname{Log}(1-\beta)$ not an integer. The approximation is excellent for middle values of s and, by necessity,

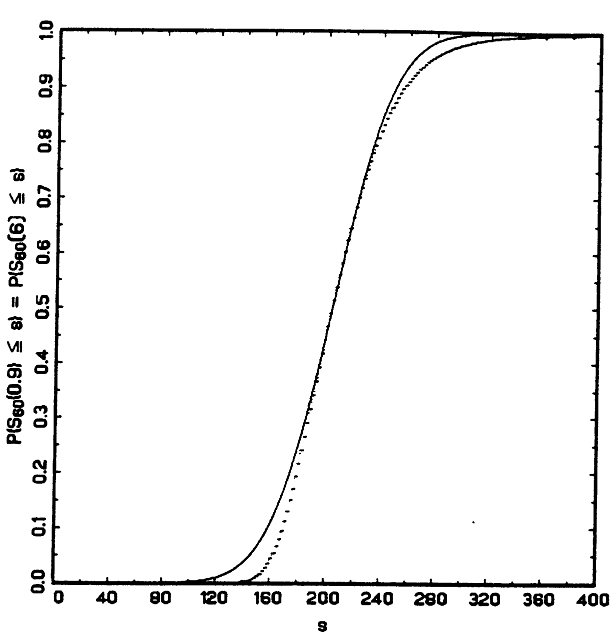

FIG. 5.

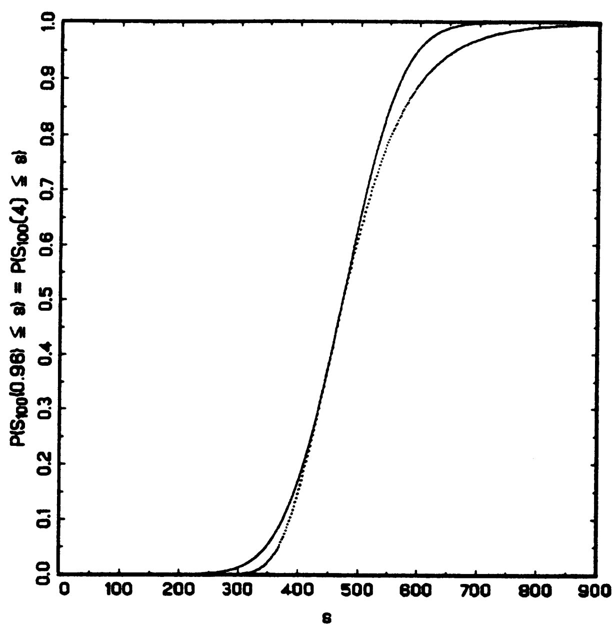

FIG. 6.

far out in the tails. For intermediate values of s, the Stigler approximation of the distribution function is too large. The overall effect *suggests* that a better approximate distribution for the distribution of $S_n\{\beta\}$ should be that of a random variable which is stochastically a little bit larger than the approximating normal here. Indeed, the proof of Stigler's theorem in Csörgő, Haeusler and Mason (1988b) shows that the normal approximation neglects a nonnegative term which asymptotically vanishes when β is a continuity point of the underlying quantile function (and does not when β is a discontinuity point, resulting in a Case 2 approximation). Figures 7 and 8 clearly show that while the normal approximation improves on the right tail as the trimming becomes heavier, it deteriorates on the left tail; in a more pronounced fashion, of course, when n is smaller.

Figures 9 and 10 illustrate the deleterious effects of having $-\text{Log}(1-\beta)$ close to an integer without using the additional term: Compare Figure 9 with Figure 12 and Figure 10 with Figure 13. Stigler (1973) himself has already discussed this possibility.

Figures 11 and 12 illustrate two examples for which $-\text{Log}(1-\beta)$ is an integer with $1/2 < \beta < 1$. Both approximations look very good. The advantage of having β farther from 1 can be seen by comparing these figures ($n = 40$ for both).

Figures 13 and 14 illustrate two examples of $\beta = 1/2$. It can be observed for both that the (limiting) Stigler approximation is exact for $s < n$. While the advantage of having n twice as large in Figure 13 is apparent, the quality

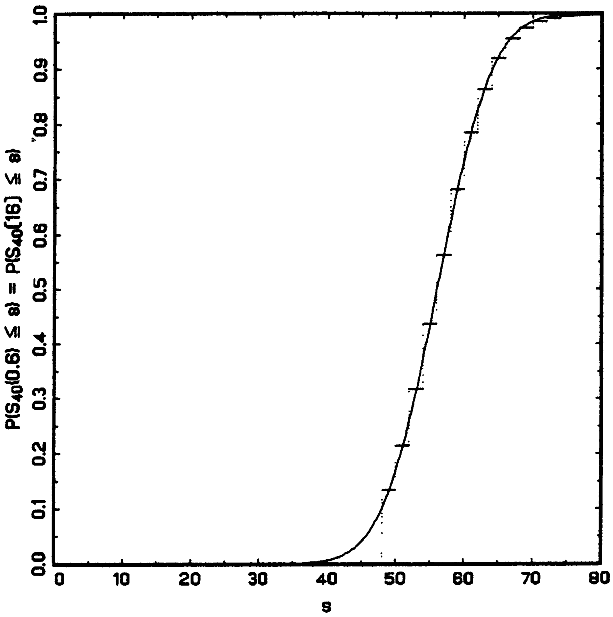
P{S40(0.6) ≤ s} = P{S40(16) ≤ s}
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
10
20
30
40
50
60
70
80
s

FIG. 7.

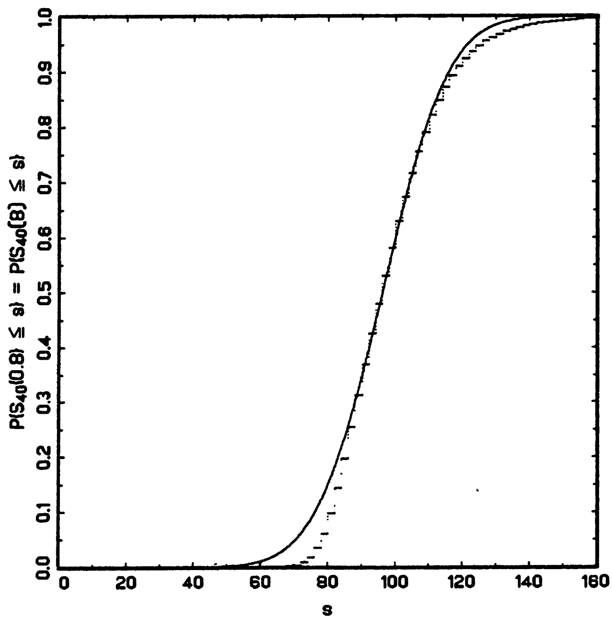
P{S40(0.8) ≤ s} = P{S40(8) ≤ s}
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
20
40
60
80
100
120
140
160
s

FIG. 8.

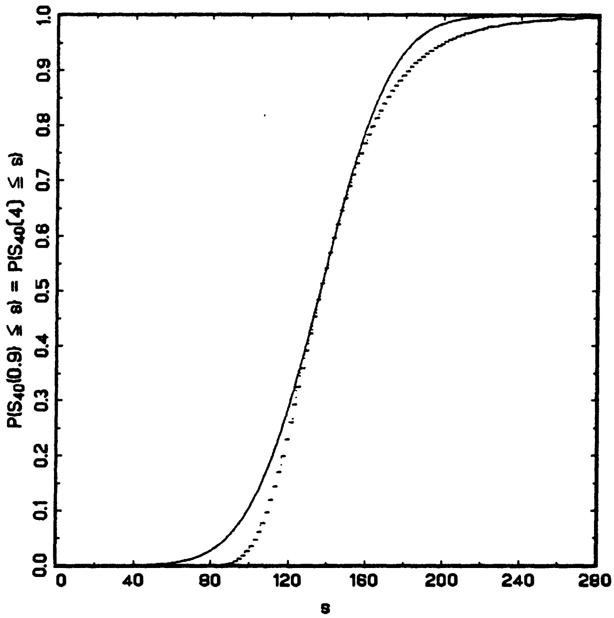
P{S40(0.9) ≤ s} = P{S40(4) ≤ s}
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
40
80
120
160
200
240
280
s

FIG. 9.

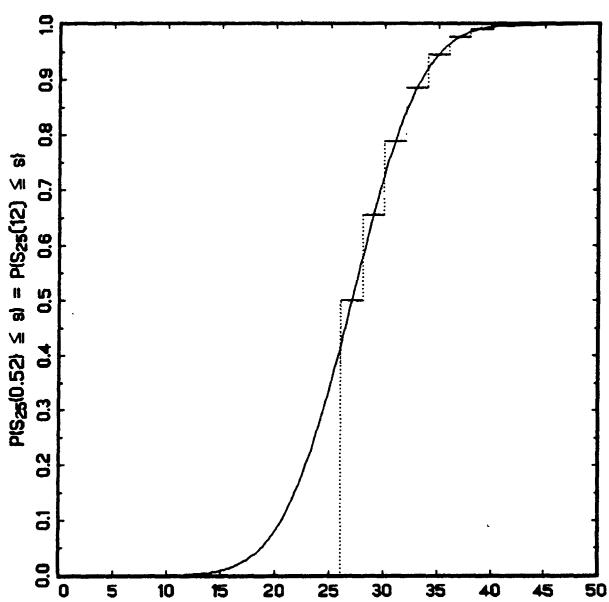
P{S25(0.52) ≤ s} = P{S25(12) ≤ s}
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
5
10
15
20
25
30
35
40
45
50
s

FIG. 10.

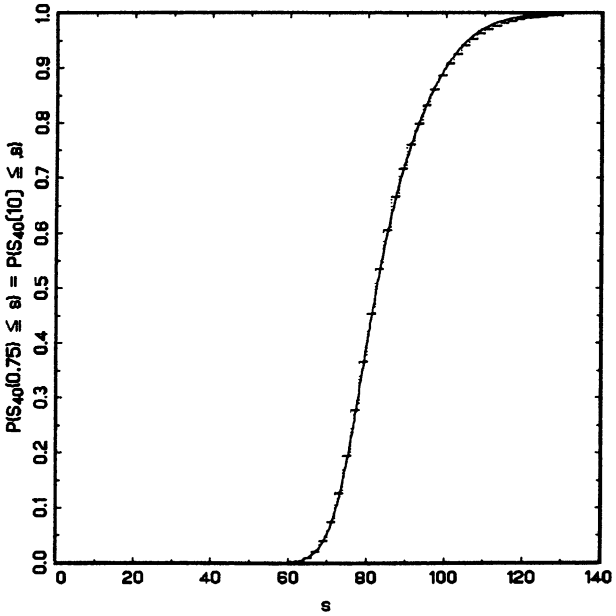
P(S40(0.75) ≤ s) = P(S40(10) ≤ s)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
20
40
60
80
100
120
140
s

FIG. 11.

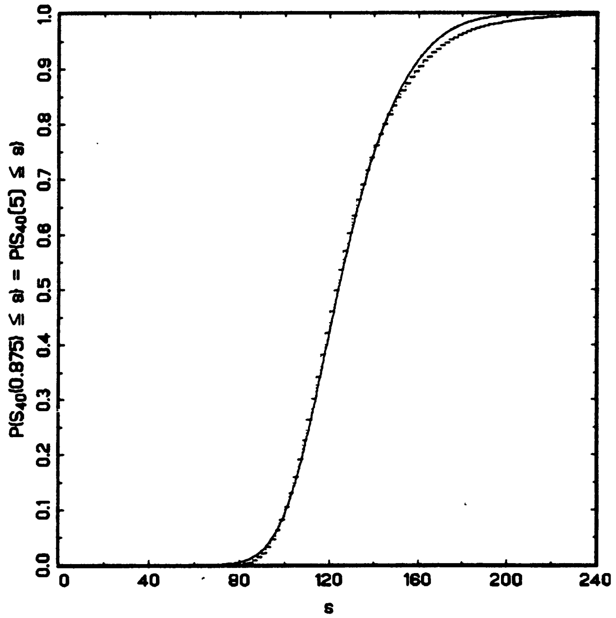
P(S40(0.875) ≤ s) = P(S40(5) ≤ s)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
40
80
120
160
200
240
s

FIG. 12.

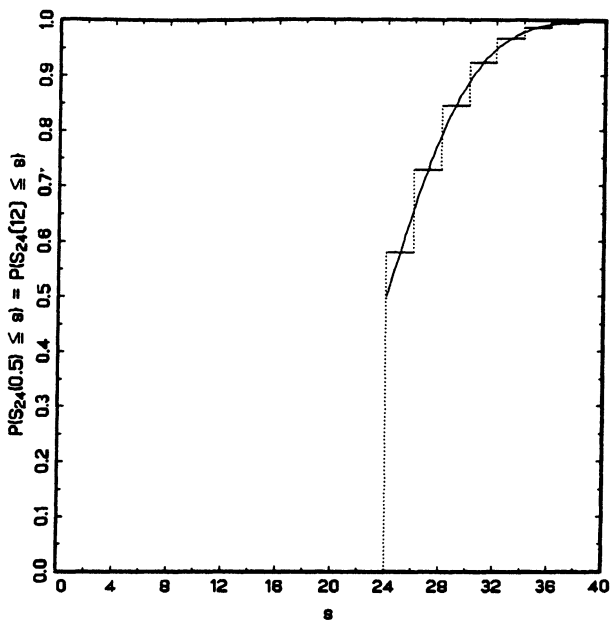
P{S24(0.5) ≤ s} = P{S24(12) ≤ s}
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
4
8
12
16
20
24
28
32
36
40
s

FIG. 13.

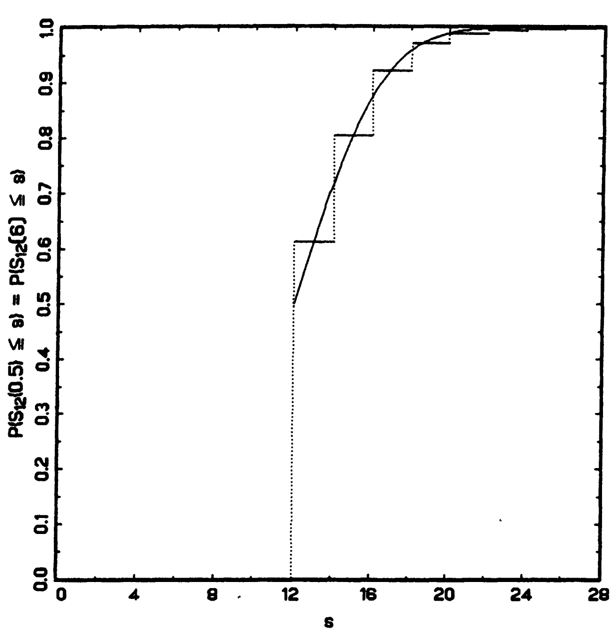
P{S12(0.5) ≤ s} = P{S12(6) ≤ s}
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
4
8
12
16
20
24
28
s

FIG. 14.

TABLE 1
Four measures of accuracy ($\Delta_L, \Delta_M, \Delta_R$ and Δ) for the examples in Figures 5–14

Figure	5	6	7	8	9	10	11	12	13	14
n m	60 6	100 4	40 16	40 8	40 4	25 12	40 10	40 5	24 12	12 6
β	0.9	0.96	0.6	0.8	0.9	0.52	0.75	0.875	0.5	0.5
Δ_L	0.0763	0.0371	0.1030	0.0901	0.0899	0.1587	0.0420	0.0274	0.0000	0.0000
Δ_M	0.0656	0.0516	0.0627	0.0754	0.0850	0.4207	0.0532	0.0365	0.5000	0.5000
Δ_R	0.0321	0.0670	0.0447	0.0295	0.0448	0.0537	0.0177	0.0217	0.0435	0.0713
Δ	0.0701	0.0653	0.0042	0.0757	0.0819	0.0009	0.0084	0.0187	0.0005	0.0067

of the approximation can still be seen in Figure 14 to be good for n as small as 12.

Figures 10–14 suggest that using a "continuity correction" is in order: The continuity correction in the present context calls for the replacement of a Stigler approximation of the form $H_{n,\beta}(s)$ by $H_{n,\beta}(s+1)$, that is, shifting the variable s one unit to the right. Numerical evidence supporting the use of this continuity correction is provided in Table 1.

Corresponding to Figures 5–10 and setting $D_{n,\beta}(s) := |F_{n,\beta}(s) - H_{n,\beta}(s)|$, the first three rows of Table 1 give $\Delta_L := \Delta_L(n,\beta) = \sup\{D_{n,\beta}(s)\colon H_{n,\beta}(s) \le 0.2\}$, $\Delta_M := \Delta_M(n,\beta) = \sup\{D_{n,\beta}(s)\colon 0.2 \le H_{n,\beta}(s) \le 0.8\}$ and $\Delta_R := \Delta_R(n,\beta) = \sup\{D_{n,\beta}(s)\colon H_{n,\beta}(s) \ge 0.8\}$ to numerically assess the maximal deviations on the left, middle and right sections of the distributions, where we use the 0.2 and 0.8 quantiles of the approximating distributions to avoid ambiguity. The last row contains the respective values of $\Delta := \Delta(n,\beta) = \sup\{|F_{n,\beta}(s) - H_{n,\beta}(s+1)|\colon s = n-m, n-m+2, n-m+4, \ldots\}$ to indicate the global effect of the continuity correction, respectable everywhere and rather dramatic for Figures 10–14. The size of $\Delta(n,\beta)$ appears to be a good (one-dimensional) summary of the intrinsic quality of Stigler's approximations for the distribution of heavily (or proportionally) trimmed St. Petersburg sums. Of course, with trimming like this, all of Paul's excitement of playing longer series of St. Petersburg games is gone!

REFERENCES

BERNOULLI, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae* **5** 175–193. Ad annos 1730 et 1731, published in 1738. [Reprinted in *Die Werke von Daniel Bernoulli* **2** 223–234. Birkhäuser, Basel, 1982. English transl. Sommer, L. (1954). Exposition of a new theory of the measurement of risk. *Econometrica* **22** 23–36; reprinted in *Precursors in Mathematical Economics: An Anthology* (W. J. Baumol and S. M. Goldfeld, eds.). *Series of Reprints of Scarce Works on Political Economy* **19** 15–26. London School of Economics and Political Science, London, 1968. Also reprinted in (1964) *Mathematics and Psychology* (G. A. Miller, ed.). *Perspectives in Psychology* 36–52, Wiley, New York, and in (1968) *Utility Theory: A Book of Readings* (A. N. Page, ed.) 199–214. Wiley, New York.]

CHOW, Y. S. and ROBBINS, H. (1961). On sums of independent random variables with infinite means and "fair" games. *Proc. Natl. Acad. Sci. U.S.A.* **47** 330–335.

CSÖRGŐ, S. and DODUNEKOVA, R. (1991). Limit theorems for the Petersburg game. In *Sums, Trimmed Sums and Extremes* (M. G. Hahn, D. M. Mason and D. C. Weiner, eds.) *Progress in Probability* **23** 285–315. Birkhäuser, Boston.

CSÖRGŐ, S., HAEUSLER, E. and MASON, D. M. (1988a). A probabilistic approach to the asymptotic distribution of sums of independent, identically distributed random variables. *Adv. in Appl. Math.* **9** 259–333.

CSÖRGŐ, S., HAEUSLER, E. and MASON, D. M. (1988b). The asymptotic distribution of trimmed sums. *Ann. Probab.* **16** 672–699.

CSÖRGŐ, S., HAEUSLER, E. and MASON, D. M. (1991). The asymptotic distribution of extreme sums. *Ann. Probab.* **19** 783–811.

CSÖRGŐ, S. and SIMONS, G. (1995). A strong law of large numbers for trimmed sums, with applications to generalized St. Petersburg games. *Statist. Probab. Lett.* **24**. To appear.

DAVID, H. A. (1981). *Order Statistics,* 2nd ed. Wiley, New York.

DUTKA, J. (1988). On the St. Petersburg paradox. *Arch. Hist. Exact Sci.* **39** 13–39.

FELLER, W. (1945). Note on the law of large numbers and "fair" games. *Ann. Math. Statist.* **16** 301–304.

FELLER, W. (1950). *An Introduction to Probability Theory and Its Applications* **1**. Wiley, New York. [Second and third editions: 1957, 1968; Chapter 10, Section 4, in all editions.]

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, MA.

HAHN, M. G., MASON, D. M. and WEINER D. C., eds. (1991). *Sums, Trimmed Sums and Extremes. Progress in Probability* **23**. Birkhäuser, Boston.

JORLAND, G. (1987). The Saint Petersburg paradox 1713–1937. In *The Probabilistic Revolution. Ideas in History* (L. Krüger, L. D. Daston and M. Heidelberger, eds.) **1** 157–190. MIT Press.

MARTIN-LÖF, A. (1985). A limit theorem which clarifies the "Petersburg paradox." *J. Appl. Probab.* **22** 634–643.

STIGLER, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1** 472–477.

DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN
1444 MASON HALL
ANN ARBOR, MICHIGAN 48109–1027

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
322 PHILLIPS HALL
CHAPEL HILL, NORTH CAROLINA 27599–3260