The Annals of Applied Probability
1993, Vol. 3, No. 3, 731–745

THE ASYMPTOTIC PROBABILITY OF A TIE FOR FIRST PLACE

BY BENNETT EISENBERG, GILBERT STENGLE AND GILBERT STRANG[1]

Lehigh University, Lehigh University and Massachusetts Institute of Technology

Suppose that the scores of n players are independent integer-valued random variables with probabilities p_j. We study the probability $P(T_n)$ that there is a tie for the highest score. The asymptotic behavior of this probability is surprising. Depending on the limit of p_{j+1}/p_j, we find different limits of different subsequences $P(T_{n(m)})$. These limits are evaluated for several families of discrete distributions.

This paper had a rather inauspicious beginning. A golf tournament on television ended in a tie. A playoff was required and the question arose: What is the probability of this outcome? That question established the context for the present paper: The model is *discrete*, there are *many players* and we are interested in the *number of winners* rather than the distribution of the scores.

By actual count, 19% of the tournaments on the professional golf tour since 1945 ended in ties. A month after the mentioned tournament, there was a five-way tie. This was an improbable event, but our paper will not go further with golf. Ultimately that question was replaced by a more mathematical one. The scores $j = 0, 1, 2, \dots$ will have probabilities p_j and the highest score wins. We examine the asymptotic probability of a tie and the asymptotic distribution of the number of winners as the number n of (independent) players approaches infinity.

We write S (or more precisely S_n) for the event that there is a single winner among the n players. This occurs with probability $P(S_n)$. The alternative outcome is T—a tie. The individual scores $X_1, \dots, X_n$ are independent random variables with $P(X_i = j) = p_j$. The cumulative probability is $F(j) = P(X_i \le j) = p_0 + \cdots + p_j$. Then the probability of S_n is the sum of probabilities that there is a single winner with score j:

$$P(S_n) = \sum_{j=1}^{\infty} np_jF(j-1)^{n-1}. \tag{1}$$

(If there is only one player and that player scores zero, then we say that there is no winner.) Our problem is the asymptotic behavior of $P(S_n)$ as $n \to \infty$.

Received June 1992; revised September 1992.
[1] Supported by NSF Grant DMS-90-06220.
AMS 1991 *subject classifications*. Primary 60G70; secondary 60F05.
Key words and phrases. Tie, asymptotic probability, record value, order statistics, logarithmic summability, geometric distribution.

The reader will recognize the competing influences on $P(S_n)$ as n increases. On the one hand, additional players may tie current winners. On the other hand, a new player may jump into the lead. This becomes less probable as the p_j decrease more rapidly: We are more inclined to expect a tie. What we did not expect, especially in the most innocent cases, was that different subsequences $P(S_{n(m)})$ approach different limits.

For clarity of presentation, the theorems are stated in terms of $P(S_n)$ rather than $P(T_n) = 1 - P(S_n)$. We will distinguish four separate cases (and the first is easy).

CASE 1. $p_j = 0$ for $j > M$. All scores are bounded by M. In this case, $P(S_n) \to 0$ as $n \to \infty$. Therefore, $P(T_n) \to 1$ (Theorem 1).

CASE 2. $p_{j+1}/p_j = q$ for a fixed number $0 < q < 1$. The scores then have the geometric distribution $p_j = q^j(1 - q)$ for $j \geq 0$. In this case Theorems 2, 3 and 4 imply the following:

(i) $\lim_{n \to \infty} P(S_n)$ does not exist.

(ii) If $n(m) = [cq^{-m}]$, the greatest integer in cq^{-m}, then $\lim_{m \to \infty} P(S_{n(m)})$ exists. The limit is a function of c that is approximately constant for q not too small.

(iii) $\lim_{N \to \infty}(1/\ln N)\Sigma_{n=1}^{N}(P(S_n))/n = (1 - q)/|\ln q|$. That is, the logarithmic means of $P(S_n)$ converge to $(1 - q)/|\ln q|$. We write $P(S_n) \to_{\log} L$ for the convergence of the logarithmic means of $P(S_n)$ to some value L.

CASE 3. $p_{j+1}/p_j \to 1$. In this case, $P(S_n) \to 1$ as $n \to \infty$ (Theorem 5).

CASE 4. $p_{j+1}/p_j \to 0$. In this case, $\liminf P(S_n) = 0$, but $\limsup P(S_n) = e^{-1}$ (Theorem 7). However, $P(S_n) \to_{\log} 0$ (Theorem 9).

One might expect the limit to be zero in Case 4, because it is so close to Case 1. The surprise in Theorem 6 is that, *whenever the scores are unbounded*, $\limsup P(S_n)$ *is at least* e^{-1}. We have the striking conclusion that $\lim P(S_n) = 0$ if and only if the scores are bounded. If the p_j's approach zero fast enough, however, then Theorems 8 and 9 show that $P(S_n)$ is logarithmically summable to zero. In this sense we can say that $P(S_n)$ converges to zero on the average.

Cases 1–4 are not exhaustive. The reader is invited to consider the behavior of $P(S_n)$ for other distributions. The results suggest that Case 1 and a variation of Case 3 are the only instances in which $\lim P(S_n)$ exists.

In the literature the winning score $W_n = \max(X_1, \dots, X_n)$ is called the *record value*. Many of the most important results about record values assume that the observations $X_1, \dots, X_n$ are continuous [Resnick (1987), Chapter 4 and Arnold and Balakrishnan (1988), Chapter 6.] The same is true in the related areas of rank tests and order statistics [Hájek (1967)]. The problem of ties for the record value and the related problem of ambiguous ranks of observations do not arise when the observations are continuous because the continuity implies that ties have probability zero. Nevertheless, observations in the real world are often intrinsically discrete—and a tie is frequently a significant event.

The final section of this paper studies the asymptotic probability that k players tie for first place. This yields some interesting limiting distributions for the number of winners.

THEOREM 1. *If* $p_j = 0$ *for* $j > M$, *then* $P(S_n) \to 0$.

PROOF. The sum in (1) is finite; it stops at $j = M$. We may suppose that $j = M$ is the highest possible score. Then $p_M > 0$ and $F(M-1) < 1$. Each term $np_jF(j-1)^{n-1}$ approaches zero as $n \to \infty$, so the sum for $P(S_n)$ approaches zero. □

When the scores are bounded, the probability of a tie approaches 1.

The geometric case. We turn to Case 2, the geometric distribution $p_j = q^j(1-q)$. The cumulative probabilities are $F(j-1) = 1 - q^j$. This arises for a sequence of independent trials with probability q of success on each trial. The score j is the number of consecutive successes until the first failure. The sum (1) becomes

$$P(S_n) = (1-q)\sum_{j=1}^{\infty} nq^j(1-q^j)^{n-1}. \tag{2}$$

We will find the limit of $P(S_n)$ when the number of players is $n \sim c/q^m$ as $m \to \infty$. More precisely, the integers n approach infinity along a sequence $n(m)$ such that $n(m)q^m \to c = q^{-x}$. The existence of the limit means that for large n, $P(S_n) \approx P(S_{[nq]})$. We find that the *limiting probability of a single winner* (or a tie) *depends on c and thus on x*. If $q \geq 0.3$, then the variation with c is very minute, but for smaller q the variation with c can be considerable.

It has come to our attention during the revision of this manuscript that the limiting behavior of the expression in (2) has been analyzed in an apparently different context by Lossers (1993), Mann (1992), and Brands, Steutel and Wilms (1992). They all consider a problem of Råde (1991), which at first glance does not look like ours. His problem is as follows:

> Suppose we have n identical coins for each of which heads occurs with probability p. Suppose we first toss all the coins, then toss those which show tails after the first toss, then toss those which show tails after the second toss, and so on until all the coins show heads. Let X be the number of coins involved in the last toss.
> (a) Find $P(X = i)$ for $i = 1, 2, \ldots, n$ and $E(X)$.
> (b) Let $p_n = P(X = 1)$. Analyze the behavior of p_n as $n \to \infty$.

This is a disguised version of our problem in Case 2. Each coin corresponds to a player whose score is the number of times that the coin shows tails. The score has a geometric distribution with $q = 1 - p$. X is just the number of players with the highest score and p_n is our $P(S_n)$. Mann's thesis has graphs and numerical estimates that complement our work. He also proves what we call Theorems 2, 3, 2k and 3k. Lossers (1992) emphasizes the "astonishing fact" that $\lim P(S_n)$ fails to exist. Brands, Steutel and Wilms also prove a version of Theorem 5.

We state our results in terms of $x = \ln c/|\ln q|$. The advantage of introducing x is that the limit of $P(S_{n(m)})$ is a periodic function of x with period 1. A subsequence with $n \sim cq^{-m} = q^{-x}q^{-m}$ yields the same limit as a subsequence with $n \sim (cq^{-1})q^{-m} = q^{-(x+1)}q^{-m}$. Theorem 2 establishes that limit, which is $L_q(x) = L_q(x+1)$. Following the proof we study the dependence of the limiting probability L on both q and x.

The following simple inequalities are used in many of the proofs.

LEMMA 1. (a) *If* $0 \le x \le 1$, *then* $(1-x)^n \le e^{-xn}$ *and* $xe^{-x} \le x$.
(b) *If* $1 \le x$, *then* $xe^{-x} \le 1/x$.

THEOREM 2. *If* $p_j = q^j(1-q)$ *and* $n(m)$ *is a sequence such that* $n(m)q^m \to q^{-x}$, *then*

$$\lim_{m\to\infty} P(S_{n(m)}) = (1-q)\sum_{t=-\infty}^{\infty} q^{t-x}\exp(-q^{t-x}) := L_q(x). \tag{3}$$

PROOF. Choose M so large that $q^{-x}/2 < n(m)q^m < 2q^{-x}$ for $m > M$. Let $j = m + t$ in (2). Then

$$\begin{aligned} P(S_{n(m)}) &= (1-q)\sum_{t=-m+1}^{\infty} n(m)q^{m+t}\left(1-q^{m+t}\right)^{n(m)-1} \\ &= (1-q)\sum_{t=-\infty}^{\infty} a_m(t), \end{aligned}$$

where $a_m(t) = 0$ for $t \le -m$ and $a_m(t) = n(m)q^{m+t}(1-q^{m+t})^{n(m)-1}$ for $t > -m$. When m exceeds M, the terms are bounded by $a_m(t) \le 2(1-q)^{-1}q^{t-x}\exp(-q^{t-x}/2)$. These bounds are summable. Thus by dominated convergence,

$$\lim_{m\to\infty} P(S_{n(m)}) = (1-q)\sum_{t=-\infty}^{\infty} \lim a_m(t) = (1-q)\sum_{t=-\infty}^{\infty} q^{t-x}\exp(-q^{t-x}). \quad \square$$

COROLLARY 1. *If* $n(m)q^m \to 1$, *then* $\lim_{m\to\infty} P(S_{n(m)}) > e^{-1} > 0.36$.

PROOF. In this case $x = 0$ and $L_q(0) > (1-q)\sum_{t=0}^{\infty} q^t e^{-1} = e^{-1}$. $\square$

Thus in the coin tossing problem, if $p = 0.99$ and the number of coins is an arbitrarily large power of 100, then the probability of a tie for the largest sequence of tails is less than 0.64—a somewhat surprising result considering that about 99.99% of such coins land heads by the second toss. If the number of coins is very large, but not a power of 100, then the conclusion is very different, as we shall see.

The form of (3) shows that $L_q(x)$ is periodic with period 1. The limiting probability $L_q(x)$ has the following Fourier expansion, with coefficients involving the gamma function.

THEOREM 3. *$L_q(x)$ is an infinitely differentiable periodic function with Fourier expansion*

$$L_q(x) = \frac{1-q}{|\ln q|} \sum_{n=-\infty}^{\infty} \Gamma\left(1 + \left\{\frac{2\pi in}{\ln q}\right\}\right) \exp(2\pi inx). \tag{4}$$

Because $L_q(x)$ is not constant, $\lim_{n\to\infty} P(S_n)$ does not exist.

PROOF. By Lemma 1 and the Weierstrass M-test, the infinite series (3) for $L_q(x)$ and the series of its termwise kth derivatives converge uniformly on $[0, 1]$ for all k. The nth Fourier coefficient of the sum in (3) is

$$c(n) = (1-q)\int_0^1 \sum_{t=-\infty}^{\infty} q^{t-x} \exp(-q^{t-x}) \exp(-2\pi inx)\, dx$$

$$= (1-q)\int_{-\infty}^{\infty} q^{-x} \exp(-q^{-x}) \exp(-2\pi inx)\, dx.$$

Change variables to $u = q^{-x}$. The Fourier coefficient $c(n)$ becomes

$$\frac{1-q}{|\ln q|}\int_0^{\infty} e^{-u} u\left(\frac{2\pi in}{\ln q}\right) du = \frac{1-q}{|\ln q|}\Gamma\left(1 + \left\{\frac{2\pi in}{\ln q}\right\}\right).$$

Then (4) is the Fourier series, it converges uniformly and L varies with x. □

To understand the nature of $L_q(x)$, first note that $c(0) = (1-q)/|\ln q| = \int_0^1 L_q(x)\, dx$ is the average value of $L_q(x)$. We call this the *average asymptotic probability of a single winner*. It is the approximate probability of a single winner if the number of players is a random variable distributed as $N = [q^{-m-X}]$, where m is a large integer and X is uniformly distributed on $[0, 1]$. Alternatively, the number of players is distributed as $N = [q^{-m}Y]$, where Y has density $f(y) = (y|\ln q|)^{-1}$ for $1 \le y \le q^{-1}$. A table of the average values of L_q is therefore useful; see Table 1.

The greater the probability q of success per trial, the greater the average asymptotic probability of a single winner when the score is the number of

TABLE 1
The average asymptotic probability of a single winner

q	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\int_0^1 L_q(x)\, dx$	0.39	0.50	0.58	0.65	0.72	0.78	0.84	0.90	0.95

consecutive successes. Intuitively, the highest score tends to be relatively isolated within the wider range of scores that occurs as q approaches 1.

For q not close to zero, $L_q(x)$ stays remarkably close to its average value $c(0)$. This follows from an examination of the magnitudes of the Fourier coefficients $c(n)$ for $n \neq 0$. Let $c(n) = c(0)a(n)$. Then $|c(n)| < |a(n)|$ for all n. Using a gamma function identity [Jahnke (1960)],

$$|a(n)|^2 = a(n)a(-n) = \Gamma(1+z)\Gamma(1-z) = \pi z/\sin \pi z \quad \text{with} \quad z = 2\pi i n/|\ln q|.$$

It follows that if $\theta = 1/|\ln q|$, then

$$(5) \qquad |a(n)|^2 = |a(-n)|^2 = 4\pi^2 n\theta/\{\exp(2\pi^2 n\theta) - \exp(-2\pi^2 n\theta)\}.$$

Because $2\pi^2 \approx 20$, we have for $n \geq 1$, $|a(n)|^2 \approx 40n\theta \exp(-20n\theta)$ for q not too close to zero. In particular, $q = 0.5$ gives $\theta \approx 1.45$ and $|a(1)| \approx 5 \times 10^{-6}$. Mann (1992) has more detailed estimates in this case. The coefficients decrease very rapidly with n, especially for larger q. Even for $q = 0.3$ and $\theta \approx 0.83$, we find $|a(1)| \approx 0.0016$ with a rapid decrease. This gives $|c(-1)| + |c(1)| \approx 0.0018$ with the sum of the other coefficients insignificant compared to this. On the other hand, $|a(n)|^2 \to 1$ as $\theta \to 0$, so that the nonconstant terms in the Fourier series are significant as q approaches zero. For $q = 0.2$, we find $|c(-1)| + |c(1)| \approx 0.03$, which is not insignificant.

Nevertheless, for q not too small, Theorem 3 has a useful corollary: For all practical purposes $P(S_n) \approx (1-q)/|\ln q|$ for large n.

COROLLARY 2. $|L_q(x) - (1-q)/|\ln q|| \leq 0.002$ for all x for $q \geq 0.3$.

Thus the values in Table 1 are essentially the limiting probabilities of having a single winner. Although $\lim P(S_n)$ does not exist, the lim sup and lim inf differ by less than 0.004 for $q \geq 0.3$. Numerical computation of the exact probabilities of having a single winner shows that n does not have to be very large for $P(S_n)$ to be close to the values in Table 1 for $q \geq 0.3$.

The inequality in this corollary is not valid as $q \to 0$ because Corollary 1 shows that $L_q(0) > 0.36$, whereas $(1-q)/|\ln q| \to 0$ as $q \to 0$. Because $(1-q)/|\ln q|$ is the average value of L_q, it follows that L_q must have sharp peaks at 0 and 1 for q near 0. In particular, for $q = 0.01$, the average probability of a single winner is approximately 0.21, whereas numerical computation shows that $L_{0.01}(0)$ is approximately 0.38.

It is worth noting that in the unrelated problem of fitting splines with knots at the geometric progression q^m to $f(x) = \log_q(x)$ for $q > 1$, Newman and Schoenberg (1975) found asymptotic behavior that is remarkably similar to $L_q(x)$. Series similar to that in (3) appear there along with the gamma function. They found that the splines $S_n(x)$ do not converge to $f(x)$, but that their logarithmic means $(\ln n)^{-1}\{\sum_{k=1}^{n} S_k(x)/k\}$ converge uniformly for $x > 0$. This type of convergence also holds for $P(S_n)$. In fact, $P(S_n)$ is logarithmically summable to the average asymptotic probability of a single winner.

THEOREM 4. *If the scores have a geometric distribution with parameter* q, *then*

$$\lim_{N\to\infty}\frac{1}{\ln N}\sum_{n=1}^{N}\frac{P(S_n)}{n}=\frac{1-q}{|\ln q|}=\int_0^1 L_q(x)\,dx.$$

PROOF. From (2) we have that

$$\sum_{n=1}^{N}P(S_n)/n=(1-q)\sum_{j=1}^{\infty}\sum_{n=1}^{N}q^j(1-q^j)^{n-1}.$$

It is easily seen that $\sum_{n=1}^{N}q^j(1-q^j)^{n-1}=1-(1-q^j)^N$ is a decreasing function of j. It follows that $\sum_{j=1}^{\infty}\sum_{n=1}^{N}q^j(1-q^j)^{n-1}$ differs from $\int_0^\infty\sum_{n=0}^{N-1}q^x(1-q^x)^n\,dx$ by at most 1. Setting $u=q^x$, this integral becomes

$$\frac{1}{|\ln q|}\int_0^1\sum_{n=0}^{N-1}(1-u)^n\,du=|\ln q|^{-1}\left(1+\frac{1}{2}+\cdots+\frac{1}{N}\right).$$

The theorem follows because $(1+1/2+\cdots+1/N)\sim\ln N$. □

The use of logarithmic means when limits do not exist is analogous to the use of Cesaro averages for probabilities in periodic Markov chains when the limits of those probabilities do not exist. The logarithmic mean is more natural here because $P(S_n)$ is approximately "multiplicatively periodic" as opposed to being approximately "additively periodic." The graph of $P(S_n)$ versus n tends to repeat with periods growing by a factor of $1/q$.

The extreme cases where p_{j+1}/p_j approaches zero or one. For the distribution $p_j=q^j(1-q)$, the ratio p_{j+1}/p_j has a constant value $0<q<1$. Table 1 indicates the different asymptotic behavior of $P(S_n)$ in case q is near 0 or 1. Indeed, $(1-q)/|\ln q|\to 0$ as $q\to 0$ and $(1-q)/|\ln q|\to 1$ as $q\to 1$. This suggests that $P(S_n)\to 0$ whenever $p_{j+1}/p_j\to 0$ and $P(S_n)\to 1$ whenever $p_{j+1}/p_j\to 1$. In this section we show that the second of these conjectures is true, but the first is false.

LEMMA 2. $\sum_{j=0}^{\infty}np_{j+1}F(j)^{n-1}\le 1\le\sum_{j=0}^{\infty}np_jF(j)^{n-1}$.

PROOF. Writing p_j as $F(j)-F(j-1)$, the inequalities become

$$\sum_{j=0}^{\infty}n[F(j+1)-F(j)]F(j)^{n-1}\le 1\le\sum_{j=0}^{\infty}n[F(j)-F(j-1)]F(j)^{n-1}.$$

The left-hand side is a lower Riemann sum and the right-hand side is an upper Riemann sum for $\int_0^1 nx^{n-1}\,dx=1$. The lemma follows. □

THEOREM 5. *If* $p_{j+1}/p_j \to 1$, *then* $P(S_n) \to 1$.

PROOF. From (1) and Lemma 2, a tie has probability

$$P(T_n) = 1 - P(S_n) \le \sum_{j=0}^{\infty} n(p_j - p_{j+1})F(j)^{n-1}$$
$$= \sum_{j=0}^{\infty} np_{j+1}\left(\frac{p_j}{p_{j+1}} - 1\right)F(j)^{n-1}.$$

Given $\varepsilon > 0$, it follows from the assumption $p_{j+1}/p_j \to 1$ and from Lemma 2 that for M large enough, $\sum_{j=M}^{\infty} np_{j+1}(p_j/p_{j+1} - 1)F(j)^{n-1} \le \varepsilon/2$ for all n. Because each term in the series goes to zero, one can then choose n large enough that the sum of the first M terms of the series is less than $\varepsilon/2$. Thus $P(T_n) < \varepsilon$ for n large enough and this proves the theorem. □

The hypothesis of Theorem 5 is stronger than is absolutely necessary. For example, there is the following corollary:

COROLLARY 3. *Let* $0 = j(0), j(1), \dots,$ *be an increasing sequence of integers and let* $r_i = p_{j(i)} + \cdots + p_{j(i+1)-1}$ *for* $i = 0, 1, \dots$. *If* $r_{i+1}/r_i \to 1$, *then* $P(S_n) \to 1$.

PROOF. Consider a new scoring system, where the score is $X^* = i$ if the original score X takes values from $j(i)$ to $j(i+1) - 1$. Then $P(X^* = i) = r_i$. Let S_n^* be the event that there is a single winner in the new scoring system. Because a tie in the old scoring system implies a tie in the new scoring system, we must have $P(S_n^*) \le P(S_n)$. By Theorem 5, $P(S_n^*) \to 1$. Hence $P(S_n) \to 1$. □

Corollary 3 strictly extends Theorem 5 in the case where $p_{2j} = c/j^2$ and $p_{2j+1} = 2c/j^2$, with $p_0 = 0$ and c as the normalizing constant. If $j(i) = 2i$, then $r_i = 3c/i^2$ for $i \ne 0$. Corollary 3 applies, but not Theorem 5.

THEOREM 6. $P(S_n) \to 0$ *as* $n \to \infty$ *if and only if the scores are bounded. If the scores are unbounded, then*

$$\limsup_{n\to\infty} P(S_n) \ge e^{-1} > 0.36.$$

PROOF. With bounded scores, Theorem 1 gives $P(S_n) \to 0$. If the scores are unbounded, we may choose a subsequence $n(m)$ so that $n(m)(1 - F(m)) \to 1$. Then

$$P(S_{n(m)}) \ge \sum_{j=m+1}^{\infty} n(m)p_j F(j-1)^{n(m)-1} \ge n(m)(1 - F(m))F(m)^{n(m)-1}.$$

This converges to e^{-1} under the assumption on $n(m)$. Thus $\limsup_{n\to\infty} P(S_n) \ge e^{-1}$. □

We next consider the extreme case where $p_{j+1}/p_j \to 0$. One might conjecture that a tie is highly probable because the scores would be concentrated around fairly low values. Although Theorem 6 shows that $\lim_{n\to\infty} P(T_n)$ cannot equal 1 in this case, Theorem 7 shows that at least $\limsup_{n\to\infty} P(T_n) = 1$.

LEMMA 3. *If $p_{j+1}/p_j \to 0$, there exists a sequence $n(m)$ such that $n(m)p_m \to \infty$ and $n(m)p_{m+1} \to 0$.*

PROOF. Let $n(m)$ be the greatest integer in $(p_m p_{m+1})^{-1/2}$. □

Recall that $W_n = \max(X_1, \dots, X_n)$ is the winning score with n players.

LEMMA 4. *If $p_{j+1}/p_j \to 0$ and $n(m)p_m \to \infty$ and $n(m)p_{m+1} \to 0$, then $P(W_{n(m)} = m) \to 1$.*

PROOF. We show that a score above m is improbable, and then that a lower score is not likely to win. Choose J large enough so that $p_{j+1} < p_j/2$ for $j > J$. Then $m > J$ implies

$$P\big(W_{n(m)} > m\big) \le \sum_{i=1}^{n(m)} P(X_i > m) = n(m) \sum_{j=m+1}^{\infty} p_j \le 2n(m)p_{m+1} \to 0.$$

On the other hand, all the scores are below m with probability

$$P\big(W_{n(m)} < m\big) \le (1-p_m)^{n(m)} \le \exp(-n(m)p_m) \to 0.$$

Therefore the probability that the winning score is m approaches 1. □

LEMMA 5. *For any $\delta > 0$, there is an $\varepsilon_0 > 0$ such that if $\varepsilon < \varepsilon_0$ and $x > 0$, then*

$$xe^{-x} + \varepsilon x e^{-\varepsilon x} \le e^{-1} + \delta.$$

PROOF. First note that $ye^{-y} \le e^{-1}$ for all y and $ye^{-y} \to 0$ as $y \to \infty$. Choose $t \ge 1$ so that $te^{-t} < \delta$. Let $\varepsilon_0 = \delta/t$. If $x \ge t$ and $\varepsilon \le \varepsilon_0$, then $xe^{-x} + \varepsilon x e^{-\varepsilon x} \le te^{-t} + \varepsilon x e^{-\varepsilon x} \le \delta + e^{-1}$. If $x \le t$, then $xe^{-x} + \varepsilon x e^{-\varepsilon x} \le e^{-1} + \varepsilon_0 t \le e^{-1} + \delta$. □

THEOREM 7. *If $\lim_{j\to\infty} p_{j+1}/p_j = 0$, then $\liminf_{n\to\infty} P(S_n) = 0$ and $\limsup_{n\to\infty} P(S_n) = e^{-1}$.*

PROOF. Let $c = \sup(1-p_j)^{-1}$. Choose $n(m)$ as in Lemma 3. Then

$$\begin{aligned} P\big(S_{n(m)}\big) &\le n(m)p_m(1-p_m)^{n(m)-1} + P(W_n \ne m) \\ &\le cn(m)p_m \exp(-n(m)p_m) + P(W_n \ne m). \end{aligned}$$

This approaches zero by the assumption on $n(m)$ and Lemma 4. This proves the first part.

For the second part, it follows from Theorem 6 that it is enough to show that for any $\delta > 0$, if n is large enough, then $P(S_n) \le e^{-1} + 5\delta$. Using Lemma 5, choose ε so small that the following three inequalities hold:

(i) $xe^{-x} + \varepsilon_1 xe^{-\varepsilon_1 x} \le e^{-1} + \delta$ for all $\varepsilon_1 < \varepsilon$ and $x > 0$.
(ii) $(1+\varepsilon)(e^{-1} + \delta) \le e^{-1} + 2\delta$.
(iii) $\varepsilon(1+\varepsilon)/(1-\varepsilon) \le \delta$.

Then choose J so that $(1 - p_J)^{-1} < 1 + \varepsilon$ and $p_{j+1}/p_j < \varepsilon$ for $j \ge J$. For n large enough, $np_J > 1$. Thus there is a unique $j(n) \ge J$ so that $np_{j(n)} \ge 1$ and $np_{j(n)+1} < 1$. From (1),

$$P(S_n) \le \sum_{j=1}^{\infty} np_j(1-p_j)^{n-1}.$$

Break this sum into four parts: from $j = 1$ to J, from $j = J + 1$ to $j(n) - 1$, from $j = j(n)$ to $j(n) + 1$ and from $j = j(n) + 2$ to infinity. We will show that only the third sum with two terms is significant.

For n large enough the first finite sum can be made less than δ because each term approaches zero as $n \to \infty$. Next notice that $np_{j(n)+2} \le \varepsilon$. Thus from the fact that $p_{j+1}/p_j < \varepsilon$ for $j \ge J$, the fourth sum is at most $\varepsilon/(1-\varepsilon) < \delta$. For the second sum note that the assumption on $j(n)$ shows that $np_{j(n)-1} \ge 1/\varepsilon$. From Lemma 1(a) and (b) and the assumption on J, $np_{j(n)-1}(1 - p_{j(n)-1})^{n(j)-1} \le (1+\varepsilon)\varepsilon$. Looking at the sum from right to left it follows that the second sum is at most $\varepsilon(\varepsilon + 1)/(\varepsilon - 1)$, which is at most δ by (iii). Furthermore, it follows from Lemma 1(a) and (i) that the third sum can be made less than $(1+\varepsilon)(e^{-1} + \delta)$, which is less than $e^{-1} + 2\delta$ by (ii). Putting this all together we have that for n large enough, $P(S_n) \le e^{-1} + 5\delta$. □

The following theorem is a useful variation of Theorem 7.

THEOREM 7′. *Assume that* $p_{j+1}/p_j \to 0$.

(i) *If* $n(m)p_m \to 1$, *then* $P(S_{n(m)}) \to e^{-1}$.
(ii) *If* $n(m)p_m \to \infty$ *and* $n(m)p_{m+1} \to 0$, *then* $P(S_{n(m)}) \to 0$.

PROOF. If $p_{j+1}/p_j \to 0$ and $n(m)p_m \to 1$, then $n(m)(1 - F(m-1)) \to 1$. The first part then follows from the proof of Theorem 6. The second part follows from the proof of Theorem 7. □

Because the p_j decrease rapidly in Theorem 7′, the sequence $n(m)$ with $n(m)p_m \to 1$ grows very quickly.

EXAMPLE. Poisson distribution with $\lambda = 1$. The distribution $p_j = e^{-1}/j!$ has $p_{j+1}/p_j \to 0$. Hence $n(m) \sim em!$ leads to $P(S_{n(m)}) \to e^{-1}$. On the other hand, $n(m) \sim em!(m+1)^{1/2}$ yields $P(S_{n(m)}) \to 0$. It is interesting to note that in the geometric case $P(S_n)$ is approximately "multiplicatively periodic" with "period" q^{-1}; that is, $P(S_n) \approx P(S_{[nq^{-1}]})$. For q very small the period is very large but constant. In the Poisson case even the apparent "multiplicative period" of $P(S_n)$ is approaching infinity. That is, in the sequences above $n(m+1)/n(m)$ approach infinity.

Although $\limsup P(S_n) \geq e^{-1}$ for any unbounded distribution, the next result shows that if the distribution is close enough to being bounded, then $P(S_n)$ goes to zero on the average. In Theorem 8 recall that $P(S_1) = P(X_1 \neq 0)$.

THEOREM 8. *For any distribution of scores*:

(i) $\sum_{n=1}^{N}(P(S_n))/n \leq E(W_N)$ *for all* N.
(ii) *If* $E(W_n)/\ln n \to 0$, *then* $P(S_n) \to_{\log} 0$.
(iii) *If* $\liminf p_j/\{1 - F(j-1)\} > 0$, *then*

$$P(S_n) \to_{\log} 0 \quad \textit{if and only if } E(W_n)/\ln n \to 0.$$

PROOF. Assume that the scores are unbounded. Then from (1),

$$\sum_{n=1}^{N} \frac{P(S_n)}{n} = \sum_{j=1}^{\infty}\sum_{n=1}^{N} p_j F(j-1)^{n-1} = \sum_{j=1}^{\infty} \frac{p_j}{(1 - F(j-1))}\left(1 - F(j-1)^N\right)$$
$$\leq \sum_{j=1}^{\infty} 1 - F^N(j-1) = \sum_{j=1}^{\infty} P(W_N \geq j) = E(W_N).$$

If the maximum possible score is M, then replace the ∞ in the summations by M. Parts (i) and (ii) follow immediately.

Under the additional assumption, the preceding expressions show that $\sum_{n=1}^{N}(P(S_n))/n \geq c_1 E(W_N) - c_2$ for positive constants c_1 and c_2. The converse of (ii) then follows. □

COROLLARY 4. *In the geometric case* $\sum_{n=1}^{N}(P(S_n))/n = (1-q)E(W_N)$ *for all* N. *In particular*, $E(W_n)/\ln n \to |\ln q|^{-1}$.

PROOF. In the geometric case $p_j/\{1 - F(j-1)\} = 1 - q$ is constant. The first part then follows from the proof of part (i) of Theorem 8. The second part follows from Theorem 4 and the first part. □

THEOREM 9. *If* $\{1 - F(j-1)\}q^{-j} \to 0$ *for all* $q > 0$, *then* $E(W_n)/\ln n \to 0$ *and* $P(S_n) \to_{\log} 0$.

PROOF. Under the assumption for any q one can choose J so that for $j > J$, $1 - F(j-1) \leq q^j$. Hence

$$E(W_N) = \sum_{j=1}^{\infty} 1 - F(j-1)^N \leq J + \sum_{j=J+1}^{\infty} 1 - (1 - q^j)^N \leq J + E(W_N^*),$$

where W_N^* is the winning score when the individual scores are geometric with parameter q. It follows from Corollary 4 that $\limsup E(W_n)/\ln n \leq 1/|\ln q|$.

But q is arbitrary. Letting $q \to 0$ we see that $E(W_n)/\ln n \to 0$. The result then follows from part (ii) of Theorem 8. □

COROLLARY 5. *If $p_{j+1}/p_j \to 0$, then $E(W_n)/\ln n \to 0$ and $P(S_n) \to_{\log} 0$.*

PROOF. It is easily seen that $\{1 - F(j-1)\}q^{-j} \to 0$ for all q in this case. □

The probability of k players tied for first place. Let N_n be the number of winners. Equation (1) gives the probability that $N_n = 1$. A single winner was the event S_n. The same reasoning leads directly to the probability that there are k winners. For each j, we look for k ($k \geq 1$) players to have score j and $n - k$ players to have a lower score:

$$P(N_n = k) = \sum_{j=1}^{\infty} \binom{n}{k} p_j^k F(j-1)^{n-k}. \tag{6}$$

(If all n players score zero, then we say that there are no winners.) The problem is still the asymptotic behavior of this sum. We consider the same four classes of distributions.

CASE 1. Bounded scores (still easy). As in Theorem 1, let M be the highest possible score. The sum in (6) stops at $j = M$ and $F(M-1) < 1$. Each term in (6) approaches zero as $n \to \infty$.

CASE 2. The geometric distribution. The probability (6) of k winners becomes

$$P(N_n = k) = (1-q)^k \sum_{j=1}^{\infty} \binom{n}{k} q^{jk}(1-q^j)^{n-k}.$$

The proofs of Theorems 2 and 3 now lead to extensions of the theorems.

THEOREM 2k. *If $p_j = q^j(1-q)$ and $n(m)$ is a sequence such that $n(m)q^m \to q^{-x}$, then*

$$\lim_{m\to\infty} P(N_{n(m)} = k) = \frac{(1-q)^k}{k!} \sum_{t=-\infty}^{\infty} q^{k(t-x)} \exp(-q^{t-x}) := L_{q,k}(x). \tag{7}$$

THEOREM 3k. *$L_{q,k}(x)$ is an infinitely differentiable periodic function with Fourier expansion*

$$L_{q,k}(x) = \frac{(1-q)^k}{k!\,|\ln q|} \sum_{n=-\infty}^{\infty} \Gamma\left(k + \frac{2\pi i n}{\ln q}\right) \exp(2\pi i n x). \tag{8}$$

The average value is

$$\int_0^1 L_{q,k}(x)\,dx = \frac{(1-q)^k}{k|\ln q|}.$$

Once again $L_{q,k}$ stays close to its average value for q not too close to zero. Let

$$l_q(k) = (1-q)^k/(k|\ln q|). \tag{9}$$

Then $\Sigma_{k=1}^{\infty} l_q(k) = 1$. The values $l_q(k)$ form a probability distribution, which we call the $\mathfrak{L}_q$ distribution. $l_q(k)$ represents the *average asymptotic probability* of k players being tied for first place when the scores have a geometric distribution with parameter q. Table 2 is a short table of these distributions.

The apparent monotonicity of $l_q(k)$ as a function of q for fixed k is easily verified for $k = 1$. It does not hold, however, for $k \geq 2$. For example, Mann (1992) shows that $l_q(2)$ is maximized for $q \approx 0.28$, $l_q(3)$ is maximized for $q \approx 0.15$ and $l_q(4)$ is maximized for $q \approx 0.10$. Evidently the smaller q is, the greater the chance of a large number of players being tied for first place.

If N_∞ is a random variable with the $\mathfrak{L}_q$ distribution, then simple calculations show that

$$E(N_\infty) = (1-q)/(q|\ln q|) \quad \text{and} \quad E(N_\infty^2) = (1-q)/(q^2|\ln q|). \tag{10}$$

It is easily verified that $E(N_\infty)$ increases as q decreases. Mann (1992) and Griffin (1993) show that for finite n, $E(N_n) = P(S_n)/q$. Because $P(S_n) \to_{\log} (1-q)/|\ln q|$, we can thus say that $E(N_n) \to_{\log} E(N_\infty)$.

CASE 3. $p_{j+1}/p_j \to 1$. Theorem 5 shows that $P(S_n) \to 1$. Thus $P(N_n = 1) \to 1$ and $P(N_n = k) \to 0$ for $k \neq 1$.

CASE 4. $p_{j+1}/p_j \to 0$. As in Case 2, the limit of $P(N_n = k)$ depends on the choice of the subsequence. We have the following extension of Theorem 7.

TABLE 2
$\mathfrak{L}_q$ *distributions*

	k						
q	**1**	**2**	**3**	**4**	**5**	**6**	**7**
0.2	0.50	0.20	0.11	0.06	0.04	0.03	0.02
0.3	0.58	0.20	0.10	0.05	0.03	0.02	0.01
0.5	0.72	0.18	0.06	0.02	0.01		
0.65	0.81	0.14	0.03	0.01			
0.7	0.84	0.13	0.03				

THEOREM 7k. *Assume that* $p_{j+1}/p_j \to 0$.

(i) *If* $n(m)$ *is a sequence such that* $n(m)p_m \to \infty$ *and* $n(m)p_{m+1} \to 0$, *then* $P(N_{n(m)} = k) \to 0$ *for each* k. *In fact,* $N_{n(m)}$ *is asymptotically binomial with parameters* $n(m)$ *and* p_m.

(ii) *If* $n(m)$ *is a sequence such that* $n(m)p_m \to \lambda$ *for some* $\lambda > 0$, *then* $\lim_{m\to\infty} P(N_{n(m)} = k) = e^{-\lambda}\lambda^k/k!$ *for* $k \geq 1$.

PROOF. Let $n(m)p_m \to \infty$ and $n(m)p_{m+1} \to 0$. By Lemma 4, the probability that the winning score with $n(m)$ players is m approaches 1. The number of players with this score is binomial with parameters $n(m)$ and p_m. Because $n(m)p_m \to \infty$, $P(N_{n(m)} = k) \to 0$ for all k.

A subsequence such that $n(m)p_m \to \lambda \neq 0$ has $n(m)p_{m+1} \to 0$ and $n(m)p_{m-1} \to \infty$. It follows as in the proof of Lemma 4 that $P(W_{n(m)} > m) \to 0$ and $P(W_{n(m)} \leq m - 2) \to 0$. Thus $P(W_{n(m)} = m - 1 \text{ or } m) \to 1$; that is, the winning score will very likely be m or $m - 1$.

Let Y_{m-1} be the number of players scoring $m - 1$ and let Y_m be the number of players scoring m when there are $n(m)$ players. Then Y_{m-1} is binomial with parameters $n(m)$ and p_{m-1}, where $n(m)p_{m-1} \to \infty$. Thus

$$P(N_{n(m)} = k \cap W_{n(m)} = m - 1) \leq P(Y_{m-1} = k) \to 0 \quad \text{for all } k.$$

On the other hand, Y_m is asymptotically Poisson with parameter λ. Thus

$$P(N_{n(m)} = k \cap W_{n(m)} = m) = P(Y_m = k \cap W_{n(m)} = m)$$
$$= P(Y_m = k \cap W_{n(m)} \leq m) \quad \text{for } k \neq 0.$$

Because $P(Y_m = k) \to e^{-\lambda}\lambda^k/k!$ and $P(W_{n(m)} \leq m) \to 1$, it follows that

$$P(N_{n(m)=k} \cap W_{n(m)} = m) \to e^{-\lambda}\lambda^k/k! \quad \text{for } k \neq 0.$$

Combining these two results with the fact that $P(W_{n(m)} = m - 1 \text{ or } m) \to 1$, we have that $P(N_{n(m)} = k) \to e^{-\lambda}\lambda^k/k!$ for $k \neq 0$. □

For the original golf problem, it is interesting to pursue the (unsupported) assumption that players have equal skills and that the number of strokes below par has a geometric distribution. If 19% of large tournaments end in ties, then Table 2 ($q = 0.65$) suggests that less than 1% of tournaments should end in five-way ties, but about 5% should end in ties of three or more. Corollary 4 with $q = 0.65$ then suggests that the winning score (strokes below par) should average $2.3 \ln n$, where n is the number of players. For $n = 100$ this is 10.6.

REFERENCES

ARNOLD, B. C and BALAKRISHNAN, N. (1988). *Relations, Bounds and Approximations for Order Statistics*. Springer, New York.

BRANDS, J. J. A. M., STEUTEL, F. W. and WILMS, R. J. M. (1992). On the number of maxima in a discrete sample. Preprint.

GRIFFIN, P. (1993). Solution to part (a) of a problem of Lennart Råde. *Amer. Math. Monthly. To appear*.

HÁJEK, J. (1967). *Theory of Rank Tests*. Academic, New York.
JAHNKE, E., EMDE, F. and LOSCH, F. (1960). *Tables of Higher Functions*. McGraw-Hill, New York.
LOSSERS, O. P. (1993). Solution to part (b) of a problem of Lennart Råde. *Amer. Math. Monthly. To appear*.
MANN, B. (1992). How do you stop flipping a lot of coins? A.B. Thesis, Princeton Univ.
NEWMAN, D. J. and SCHOENBERG, I. J. (1975). Splines and the logarithmic function. *Pacific J. Math.* **61** 241–258.
RÅDE, L. (1991). Problem E3436. *Amer. Math. Monthly* **99** 366.
RESNICK, S. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

BENNETT EISENBERG
GILBERT STENGLE
DEPARTMENT OF MATHEMATICS
CHRISTMAS–SAUCON HALL
14 LEHIGH UNIVERSITY
BETHLEHEM, PENNSYLVANIA 18015

GILBERT STRANG
DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139