*Journal of Theoretical Probability, Vol. 9, No. 2, 1996*

# Divergence Rates for the Number of Rare Numbers

**Eric S. Key**[1]

*Received September 26, 1994; revised March 24, 1995*

Suppose that $X_1, X_2,...$ is a sequence of i.i.d. random variables taking value in $Z^+$. Consider the random sequence $A(X) \equiv (X_1, X_2,...)$. Let $Y_n$ be the number of integers which appear exactly once in the first $n$ terms of $A(X)$. We investigate the limit behavior of $Y_n/E[Y_n]$ and establish conditions under which we have almost sure convergence to 1. We also find conditions under which we determine the rate of growth of $E[Y_n]$. These results extend earlier work by the author.

**KEY WORDS:** Rare numbers; almost sure convergence; subadditive.

## 1. INTRODUCTION

The topic of rare numbers was first introduced in Key[(2)] in response to a problem in density estimation formulated by Cuevas and Walters.[(1)] In the latter work, the authors addressed the problem of estimting generalized densities (mixtures of discrete and continuous distributions), and they needed to get a handle on the number of observed values which appear only once in the data and which are from the discrete part of the distribution.

To this end, in the former work, the following situation was analyzed. Suppose that $X_1, X_2,...$ is an i.i.d. sequence of positive integer valued random variables, with $p_k \equiv \Pr(X_i = k) > 0$ and $p_k$ a nonincreasing sequence. For a fixed positive integer $n$, the integers which appear once in $(X_1,..., X_n)$ were called *rare numbers*. We are interested in the behavior of the number of rare numbers, denoted by $Y_n$, as $n \to \infty$.

[1] Department of Mathematical Sciences, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201. E-mail: ericskey@csd.uwm.edu.



0894-9840/96/0400-0413$09.50/0 

It will be useful to express $Y_n$ as follows. For each positive integer $k$ let $I_{k,n}$ denote the indicator of the event that the integer $k$ appears exactly once in the $n$-tuple $(X_1,..., X_n)$. We may then write

$$Y_n = \sum_{k=1}^{\infty} I_{k,n}$$

from which we see that

$$E[Y_n] = n \sum_{k=1}^{\infty} (1-p_k)^{n-1} p_k \tag{1.1}$$

It is clear that $Y_n \leqslant n$ and that $|Y_{n+1} - Y_n| \leqslant 1$. In fact, if we put $Y_{m,n}$ = the number of positive integers which appear once in $(X_{m+1},..., X_n)$, it was pointed out in Key[2] that $Y_{m,n}$ is a subadditive process. It follows from Kingman's subadditive ergodic theorem and Eq. (1.1) that $Y_n/n$ convergerges almost surely to 0.

It is the point of this note to show how to improve upon the following two results in Key[2]:

**Theorem 1.** If $\delta \in [0, 1/2)$ and $\sum_{k=1}^{\infty} p_k^{1-\delta} < \infty$ then $\lim_{n\to\infty} n^{-(1-\delta)} Y_n = 0$ a.s. If $\delta \in [1/2, 2/3)$, $\varepsilon \in (2\delta - 1, 1-\delta)$ and $\sum_{k=1}^{\infty} p_k^{1-\delta-\varepsilon} < \infty$ then $\lim_{n\to\infty} n^{-(1-\delta)} Y_n = 0$ a.s.

**Theorem 2.** Suppose that there is some positive integer $K$ such that for all $k > K$, $0 < p_k < p_{k-1}$. Put $r \equiv \limsup_{k\to\infty} p_{k+1}/p_k$ and $s \equiv \liminf_{k\to\infty} p_{k+1}/p_k$. Then

$$\limsup_{n\to\infty} E[Y_n] \leqslant (\log(1/r))^{-1} + e^{-1} \tag{1.2}$$

$$\liminf_{n\to\infty} E[Y_n] \geqslant \frac{e^{-1/s} - e^{-s} + 1}{-\log s} \tag{1.3}$$

The key to our improvements is

**Theorem 5.** Suppose that $p \leqslant n$ is a positive integer. Then there is a polynomial $Q_p$ of degree $p-1$ so that for all $n \geqslant p$

$$E[Y_n^p] \leqslant E[Y_n]^p + (1+o(1))\, Q_p(E[Y_n])$$

First we use Theorem 5 to show in Theorem 6 that the second part of Theorem 1 may be improved to $\delta \in [1/2, 1)$. We then use it to establish

Theorem 7 on the convergence of $Y_n/E[Y_n]$. The first corollary of Theorem 7 is

**Corollary 8.** Suppose that $A \geqslant 1$ and

$$\sum_{n=1}^{\infty} \frac{1}{E[Y_{n^A}]} < \infty$$

$$\lim_{n \to \infty} \frac{n^{1-(1/A)}}{E[Y_n]} = 0$$

Then

$$\lim_{n \to \infty} \frac{Y_n}{E[Y_n]} = 1 \qquad \text{a.s.}$$

We then derive a series of results aimed at determining the rate of divergence of $E[Y_n]$. For example,

**Corollary 13.** Suppose that for some $C > 0$ and $\beta > 1$ we have

$$p_k = (C + o(1)) \frac{\log^{\alpha}(k)}{k^{\beta}}$$

For $n \geqslant 2$ put $g(n) = n^{1/\beta} \log^{\alpha/\beta}(n)$. Then

$$\lim_{n \to \infty} \frac{E[Y_n]}{g(n)} = \frac{C^{1/\beta}}{\beta^{(\alpha+\beta)/\beta}} \Gamma\left(\frac{\beta - 1}{\beta}\right)$$

Finally, we improve Theorem 2 by obtaining a complete description of the limit points of $E[Y_n]$ if $p_{k+1}/p_k$ converges to an element of (0, 1).

## 2. TWO LAWS OF LARGE NUMBERS FOR $Y_n$

We know that in general that if $q \geqslant 1$ and $R$ is a positive random variable then $E[R]^q \leqslant E[R^q]$. For $Y_n$ something like the reverse inequality holds. We begin with some simple observations.

First, it is easy to show that:

**Proposition 3.** For distinct positive integers $x_1, \dots, x_a$, and $a \leqslant n$,

$$E[I_{x_1,n} \cdots I_{x_a,n}] \leqslant \frac{n!}{(n-a)!} \prod_{j=1}^{a} p_{x_j}(1 - p_{x_j})^{n-a}$$

Next, a technical lemma to help get a handle on the rate of growth of $E[Y_n]$:

**Lemma 4.** For any $b > a \geqslant 0$, if

$$\sum_{k=1}^{\infty} p_k^{b-a} < \infty$$

then

$$\lim_{n\to\infty} n^a \sum_{k=1}^{\infty} (1-p_k)^n p_k^b = 0$$

*Proof.* For $n > 0$ and $u \in [0, 1]$ note that $(1-u)^n u^a$ is bounded by above by $a^a(n+a)^{-a}$, so Eq. (2.1) follows from the dominated convergence theorem. □

**Theorem 5.** Suppose that $p \leqslant n$ is a positive integer. Then there is a polynomial $Q_p$ of degree $p-1$ so that for all $n \geqslant p$

$$E[Y_n^p] \leqslant E[Y_n]^p + (1+o(1))\, Q_p(E[Y_n]) \tag{2.1}$$

*Proof.* To prove this theorem we need some additional notation: $N^j$ will stand for the set of ordered $j$-tuples of distinct positive integers.

Since

$$Y_n^p = \sum_{k_1=1}^{\infty} \cdots \sum_{k_p=1}^{\infty} I_{k_1,n} \cdots I_{k_p,n}$$

and the $I_{k,n}$ take only the values 0 and 1, there are non-negative integers $A_{p,j}$, independent of $n$, with $A_{p,p} = 1$ so that

$$Y_n^p = \sum_{j=1}^{p} A_{p,j} \left( \sum_{(k_1,\dots,k_j)\in N^j} I_{k_1} \cdots I_{k_j} \right) \tag{2.2}$$

From Proposition 3 we see that

$$\begin{aligned} E&\left[ \sum_{(k_1,\dots,k_j)\in N^j} I_{k_1} \cdots I_{k_j} \right] \\ &\leqslant \frac{n!}{(n-j)!} \sum_{(k_1,\dots,k_j)\in N^j} \prod_{i=1}^{j} p_{k_i}(1-p_{k_i})^{n-j} \\ &\leqslant \frac{n!}{(n-j)!\,(n-j+1)^j} \left( \sum_{k=1}^{\infty} (n-j+1)(1-p_k)^{n-j} p_k \right)^j \\ &= (1+O(n^{-1}))(E[Y_n]+o(1))^j \end{aligned}$$

Note that this last expression may be broken down into a sum of four terms:

- $E[Y_n]^j$.
- A polynomial of degree $j-1$ in $E[Y_n]$ times $o(1)$.
- $O(n^{-1})\,E[Y_n]^j$.
- A polynomial of degree $j-1$ in $E[Y_n]$ times $o(1)$ times $O(n^{-1})$.

However, from Eq. (1.1) and Lemma 4 we see that

$$O(n^{-1})\,E[Y_n]^j = o(1)\,E[Y_n]^{j-1}$$

so $E[\sum_{(k_1,\ldots,k_j)\in N^j} I_{k_1}\cdots I_{k_j}] \leqslant E[Y_n]^j +$ (a polynomial in $E[Y_n]$ of degree $j$) $o(1)$. Therefore the theorem follows by taking the expectation of both sides of Eq. (2.2). □

The following extends Theorem 1 up to $\delta < 1$:

**Theorem 6.** If $\delta \in [1/2, 1)$ and for some $\varepsilon > 0$ we have $\sum_{k=1}^{\infty} p_k^{1-\delta-\varepsilon} < \infty$ then $\lim_{n\to\infty} n^{-(1-\delta)} Y_n = 0$ a.s.

*Proof.* By hypothesis and Lemma 4 we have $E[Y_n] = o(n^{1-\delta-\varepsilon})$. Let $p$ be the smallest integer so that $(2p-1)\varepsilon > \delta$. From Theorem 5 it is sufficient to show that

$$\lim_{n\to\infty} \frac{Y_n^p - E[Y_n^p]}{n^{(1-\delta)p}} = 0 \qquad \text{a.s.} \tag{2.3}$$

Again using Theorem 5 we see that

$$\operatorname{Var}(Y_n^p) = n^{(1-\delta-\varepsilon)(2p-1)}\,o(1)$$

so Eq. (2.3) follows from the Borel-Cantelli lemma and Chebychev's inequality. □

We would like to determine a normalization of $Y_n$ which gives an almost sure limit which is finite and positive. To this end, we let $k_{n,A}$ denote the largest integer whose $A$th power does not exceed $n$, and if $x > 0$ is not an integer, we take $Y_x$ to be $Y_q$ where $q$ is the integer part of $x$.

**Theorem 7.** Let $p$ be a positive integer and let $A \geqslant 1$ be given. Suppose that $\phi(x) > 0$ for all $x > 0$. Let $l_n \equiv k_{n,A}^A$. Suppose that

$$\sum_{n=1}^{\infty} \frac{1}{\phi(n^A)^{2p}\,E[Y_{n^A}]} < \infty \tag{2.4}$$

$$E[Y_n] \geqslant C_1 > 0 \tag{2.5}$$

If $A=1$ or if

$$\lim_{n\to\infty} \frac{n^{1-(1/A)}}{\phi(n)\, E[Y_n]} = 0 \tag{2.6}$$

$$\frac{\phi(l_n)\, E[Y_{l_n}]}{\phi(n)\, E[Y_n]} \leqslant C_2 < \infty \tag{2.7}$$

then

$$\lim_{n\to\infty} \frac{1}{\phi(n)} \left( \frac{Y_n}{E[Y_n]} - 1 \right) = 0 \qquad \text{a.s.}$$

*Proof.* It follows from Chebychev's inequality and Theorem 5 that since $E[Y_n]$ is bounded away from 0 that there is a constant $K$ so that

$$\Pr\left[ \left| \frac{1}{\phi(n)} \left( \frac{Y_n}{E[Y_n]} - 1 \right) \right| > \varepsilon \right] \leqslant \frac{E[(Y_n - E[Y_n])^{2p}]}{(\varepsilon\phi(n)\, E[Y_n])^{2p}} \leqslant \frac{K}{(\varepsilon\phi(n))^{2p}\, E[Y_n]}$$

Hence it follows from the Borel-Cantelli lemma and (7) that

$$\lim_{n\to\infty} \frac{1}{\phi(n^A)} \left( \frac{Y_{n^A}}{E[Y_{n^A}]} - 1 \right) = 0 \qquad \text{a.s.}$$

To finish off the proof we need to "fill in the gaps" in case $A > 1$. First note that

$$\lim_{n\to\infty} \frac{1}{\phi(l_n)} \left( \frac{Y_{l_n}}{E[Y_{l_n}]} - 1 \right) = 0 \qquad \text{a.s.}$$

since $l_n$ increases to infinity and only takes values in the $A$th powers of positive integers.

Next, note that $k_{n,A}$ is of the same order of magnitude as $n^{1/A}$, so

$$n - l_n \leqslant (k_n + 1)^A - k_n^A = O(n^{1-(1/A)}) \tag{2.8}$$

Since for any $n$ we have $|Y_n - Y_{n+1}| \leqslant 1$ it follows from Eqs. (2.6) and (2.7)

$$\limsup_{n\to\infty} \frac{1}{\phi(n)} \left| \frac{Y_n}{E[Y_n]} - 1 \right| \leqslant \limsup_{n\to\infty} C_2 \frac{1}{\phi(l_n)} \left| \frac{Y_{l_n} - E[Y_{l_n}]}{E[Y_{l_n}]} \right| = 0 \qquad \text{a.s.}$$

which proves the theorem. □

**Corollary 8.** Suppose that $A \geqslant 1$ and

$$\sum_{n=1}^{\infty} \frac{1}{E[Y_{n^A}]} < \infty \tag{2.9}$$

$$\lim_{n \to \infty} \frac{n^{1-(1/A)}}{E[Y_n]} = 0 \tag{2.10}$$

Then

$$\lim_{n \to \infty} \frac{Y_n}{E[Y_n]} = 1 \qquad \text{a.s.}$$

*Proof.* Set $\phi(x) \equiv 1$ in the theorem. Then Eq. (2.4) reduces to Eq. (2.9) and Eq. (2.6) reduces to Eq. (2.10).

Next, note that Eq. (2.10) implies that $E[Y_n]$ is divergent. Hence Eq. (2.5) is satisfied. Thus we have the case $A = 1$. If $A > 1$, since $|Y_n - Y_{n+1}| \leqslant 1$, we have

$$\left| \frac{E[Y_{l_n}]}{E[Y_n]} - 1 \right| \leqslant \frac{n - l_n}{E[Y_n]}$$

Together with Eqs. (2.8) and (2.10) this shows us that Eq. (2.7) holds as well. □

**Remark 2.1.** The hypotheses of this corollary cannot be satisfied if

$$\limsup_{n \to \infty} \frac{E[Y_n]}{\sqrt{n}} < \infty$$

while they are always satisfied if for some $f(n) > 0$,

$$\sum_{n=1}^{\infty} \frac{1}{n f(n^2)} < \infty$$

$$\limsup_{n \to \infty} \frac{\sqrt{n}\, f(n)}{E[Y_n]} < \infty$$

For example, $f(n) = \log(n+1)^2$.

In the next section we shall determine some condition on the $p_k$ under which the hypotheses of Theorem 7 are satisfied.

## 3. THE BEHAVIOR OF $E[Y_n]$

In order to better appreciate Theorem 7 we have to have a better grip on the behavior of $E[Y_n]$ as $n \to \infty$. Our goal is to describe the behavior of $E[Y_n]$ in terms of the behavior of the $p_k$. We already see from Lemma 4 that for $a \in (0, 1]$, if $p_k^a$ is summable then $E[Y_n] = o(1)\, n^a$. We have also seen in Theorem 2 that in some circumstances $E[Y_n]$ may be bounded. These results are not of much use to us in that what Theorem 7 indicates to us is that we want $E[Y_n]$ to be divergent. However, Lemma 4 does give us a clue about how quickly we might expect $E[Y_n]$ to diverge. We will now proceed to develop some results in that direction.

First off let us settle when $E[Y_n]$ is a bounded sequence. Here is a companion to to Theorem 2:

**Theorem 9.** If there is a strictly increasing sequence of positive integers, $k(j)$, such that $p_{k(j+1)}/p_{k(j)} < 1$ and

$$\lim_{j \to \infty} p_{k(j+1)}/p_{k(j)} = 1$$

then $\lim_n E[Y_n] = \infty$.

**Remark 3.1.** For example, if the $p_k$ are strictly decreasing and $\limsup_k p_{k+1}/p_k = 1$ then such a sequence $k(j)$ exists. What we have to avoid is a sequence $p_k$ such as $p_{2k-1} = p_{2k} = 2^{-k-1}$ for which the conclusion of the theorem is false.

*Proof.* Let $a_j = p_{k(j)}$. Choose $\varepsilon \in (0, 1)$. Then there is a positive integer $M$ so that if $j > M$ then $a_j/a_{j-1} \geqslant (1-\varepsilon)$. Then

$$\begin{aligned} E[Y_n] &\geqslant \sum_{j=M}^{\infty} n(1-a_j)^{n-1} a_j \\ &\geqslant \sum_{j=M}^{\infty} ((1-a_j)^n - (1-a_{j-1})^n) \left(\frac{a_{j-1}}{a_j} - 1\right)^{-1} \\ &\geqslant \frac{1-\varepsilon}{\varepsilon}(1-(1-a_{M-1})^n) \end{aligned}$$

Hence $\liminf_n E[Y_n] \geqslant (1-\varepsilon)/\varepsilon$, which implies the conclusion of the proposition. □

Of course, to apply Theorem 7 we need some idea of the rate of divergence of $E[Y_n]$.

**Proposition 10.** Suppose that $p\colon (0, \infty) \to (0, 1)$ is decreasing and $\sum_{k=1}^{\infty} p(k) = 1$. Put $p_k = p(k)$. Suppose that for some function $g\colon (0, \infty) \to (0, \infty)$ we have

$$p(g(x)) = \frac{c_g + o(1)}{x}$$

for some positive constant $c_g$. Then

$$\liminf_{n \to \infty} \frac{E[Y_n]}{n \sum_{k \geqslant g(n)} p_k} \geqslant \exp(-c_g) \tag{3.1}$$

If we can also find $h\colon (0, \infty) \to (0, \infty)$ so that

$$p(h(x)) = \frac{c_h + o(1)}{x}$$

for some positive constant $c_h$, and for every $x$, $g(x) < h(x)$, then

$$\liminf_{n \to \infty} \frac{E[Y_n]}{h(n) - g(n)} \geqslant \min(c_g e^{-c_g}, c_h e^{-c_h}) \tag{3.2}$$

*Proof.* For Eq. (3.1), observe that

$$E[Y_n] \geqslant n(1 - p(g(n)))^n \sum_{k \geqslant n} p_k$$

For Eq. (3.2), let $W_n$ denote the integers between $g(n)$ and $h(n)$. The key to the proof is that on any closed interval in $[0, 1]$, as a function of $u$, the expression $u(1-u)^{n-1}$ achieves its minimum at the endpoints of the interval. Therefore

$$\begin{aligned} E[Y_n] &\geqslant \sum_{k \in W_n} np_k(1 - p_k)^{n-1} \\ &\geqslant (h(n) - g(n)) \min(np(g(n))(1 - p(g(n))^n, bh(n))(1 - p(h(n))^n)) \end{aligned}$$

which implies Eq. (3.2). □

Equations (3.1) and (3.2) do not always yield the same estimates on the rate of growth of $E[Y_n]$. For example, it is easy to check that if

$$p(x) = (C + o(1)) \frac{\log^{\alpha}(x)}{x^{\beta}}$$

satisfies the hypotheses of Proposition 10 then we can take $g(x) = x^{1/\beta} \log^{\alpha/\beta}(x)$ and $h(x) = 2g(x)$. Equation (3.2) tells us that $E[Y_n]$ grows as least as fast as $g(n)$. However, Eq. (3.1) tells us that the growth rate is at least $g(n)$ if $\beta > 1$, while if $\beta = 1$, it is at least $\log(n)\, g(n)$.

We now turn our attention to describing a function $g$ in terms of $p_k$ so that $\lim_n E[Y_n]/g(n) \in (0, \infty)$.

**Lemma 11.** Suppose that $p: (0, \infty) \to (0, 1)$ is decreasing, $\sum_{k=1}^{\infty} p(k) = 1$, and that there is a divergent sequence $g(n)$ such that for all $x \in [a, b] \subset (0, \infty)$, $np(xg(n))$ converges uniformly to a function $f(x)$ which is integrable on $[a, b]$. Put $p_k = p(k)$ for any positive integer $k$. Then

$$\lim_{n \to \infty} \frac{1}{g(n)} \sum_{k \in [ag(n), bg(n)]} n(1-p_k)^n p_k = \int_a^b \exp(-f(x))\, f(x)\, dx$$

*Proof.* Since $0 < p_1 < 1$ we have

$$\exp(-np_k) \exp\left(-\frac{np_k^2}{2(1-p_1)}\right) \leqslant (1-p_k)^n \leqslant \exp(-np_k) \tag{3.3}$$

so for $k \in [ag(n), bg(n)]$ we have that there is a constant $C_{a,b}$ so that

$$\exp\left(-np\left(\left(\frac{k}{g(n)}\, g(n)\right)\right)\right) \exp(-C_{a,b}\, p_{ag(n)})$$
$$\leqslant (1-p_k)^n \leqslant \exp\left(-np\left(\left(\frac{k}{g(n)}\, g(n)\right)\right)\right)$$

Since $f(x)$ is Riemann integrable on $[a, b]$, so is $f(x) \exp(-f(x))$ and the lemma follows from the pinching theorem. □

**Theorem 12.** Under the hypotheses of Lemma 11 we have

$$\liminf_{n \to \infty} \frac{E[Y_n]}{g(n)} \geqslant \int_a^b \exp(-f(x))\, f(x)\, dx$$

$$\limsup_{n \to \infty} \frac{E[Y_n]}{g(n)} \leqslant \frac{a}{c} + \int_a^b \exp(-f(x))\, f(x)\, dx + \limsup_{n \to \infty} \frac{n}{g(n)} \sum_{k > bg(n)} p_k$$

*Proof.* We may write

$$E[Y_n] = n \sum_{k=1}^{\infty} (1-p_k)^n p_k + o(1)$$

and then divide the sum according to $k < ag(n)$, $k \in [ag(n), bg(n)]$ and $k > bg(n)$. For the terms with $k < ag(n)$ note that

$$n(1-p_k)^n p_k \leqslant \frac{n}{n+1}\left(1-\frac{1}{n+1}\right)^n$$

The lemma covers the middle values of $k$, and for the large values of $k$ overestimate $(1-p_k)^n$ by 1. □

**Corollary 13.** Suppose that for some $C>0$ and $\beta>1$ we have

$$p_k = (C+o(1))\frac{\log^\alpha(k)}{k^\beta}$$

For $n \geqslant 2$ put $g(n) = n^{1/\beta}\log^{\alpha/\beta}(n)$. Then

$$\lim_{n\to\infty}\frac{E[Y_n]}{g(n)} = \frac{C^{1/\beta}}{\beta^{(\alpha+\beta)/\beta}}\Gamma\left(\frac{\beta-1}{\beta}\right)$$

*Proof.* The hypotheses of the theorem are satisfied for any $0<a<b<\infty$ with $f(x)=C\beta^{-\alpha}x^{-\beta}$. It is easy to check that

$$\limsup_{n\to\infty}\frac{n}{g(n)}\sum_{k>bg(n)}(1-p_k)^n p_k$$

converges to 0 as $b\to\infty$, so letting $a\to 0$ and the $b\to\infty$ gives

$$\lim_{n\to\infty}\frac{E[Y_n]}{g(n)} = \int_0^\infty f(x)\exp(-f(x))\,dx$$

Making the substitution $u=x^{-\beta}$ allows us to compute the integral. □

To see what happens when $\beta=1$ we formulate the result slightly differently to avoid some technical problems about sums of series.

**Corollary 14.** Suppose that for some $C>0$ and for all $u>1$ we have

$$\sum_{k>u} p_k = (C+o(1))\log^\alpha(u)$$

Then

$$\lim_{n\to\infty}\frac{E[Y_n]}{n\log^\alpha(n)} = C$$

*Proof.* It is straightfoward to show that $g(n) \equiv n^{-1} \log^{\alpha-1}(n)$ satisfies the hypotheses of Theorem 12 with $f(x) = C'/x$ for some positive constant $C'$ and $[a, b] = [1, 2]$. Therefore

$$\begin{aligned}
\limsup_{n \to \infty} \frac{E[Y_n]}{n \log^{\alpha}(n)} &= \limsup_{n \to \infty} \frac{E[Y_n]}{g(n) \log(n)} \\
&\leqslant \limsup_{n \to \infty} (C + o(1)) \left( \frac{\log(2) + \log(n) + (\alpha - 1) \log \log(n)}{\log(n)} \right)^{\alpha} \\
&= C
\end{aligned}$$

On the other hand, if in Proposition 10 we take $b > 0$ and $g(n) = bn \log^{\alpha-1}(n)$ we have

$$\begin{aligned}
&\liminf_{n \to \infty} \frac{E[Y_n]}{n \log^{\alpha}(n)} \\
&\quad = \liminf_{n \to \infty} (C + o(1)) \left( \frac{\log(bn) + (\alpha - 1) \log \log(n)}{\log(n)} \right)^{\alpha} \frac{E[Y_n]}{g(n) \log(n)} \\
&\quad \geqslant C \exp(-C'/b)
\end{aligned}$$

Since we may take $b$ to be arbitrarily large we have

$$\liminf_{n \to \infty} \frac{E[Y_n]}{n \log^{\alpha}(n)} = C = \limsup_{n \to \infty} \frac{E[Y_n]}{n \log^{\alpha}(n)}$$

which is what we wanted to show. □

The next corollary gives us an idea of the rate of convergence of $Y_n / E[Y_n]$ to 1:

**Corollary 15.** Suppose that

$$p_k = (C + o(1)) \frac{\log^{\alpha}(k)}{k^{\beta}}$$

as $k \to \infty$ for some $\beta \geqslant 1$ and that $p$ is a positive integer. Put $x = (2p + 1)\beta/(2p\beta - 2p + 1)$ and $g(n) = n^{1/\beta} \log^{\alpha/\beta}(n)$. If $u(n)$ is increasing and divergent, and

$$\sum_{n=1}^{\infty} \frac{(\log n)^{(2p-1)\alpha/\beta}}{nu^{2p}(n^{x})} < \infty$$

then

$$\lim_{n \to \infty} \frac{n^{(2-\beta)/((2p+1)\beta)}(\log n)^{\alpha/\beta}}{u(n)} \left( \frac{Y_n - E[Y_n]}{g(n)} \right) = 0 \qquad \text{a.s.}$$

**Remark 3.2.** Notice that the factor

$$n^{(2-\beta)/((2p+1)\beta)}(\log n)^{\alpha/\beta}/u(n)$$

may be taken to be divergent if $\beta \in [1, 2)$ or if $\beta = 2$ and $\alpha > 0$. For example, if $\beta = 1$ then we must have $\alpha < -1$. If we set $p = 1$, we get $n^{1/3}(\log n)^{\alpha/\beta}/u(n)$ and we may take $u(n)$ to diverge as slowly as we like.

*Proof.* Take $A = x$ and $\phi(n) = n^{(\beta-2)/((2p+1)\beta)}(\log n)^{\alpha/\beta}u(n)$ in Theorem 7. Theorem 12 allows us to verify that the hypotheses of Theorem 7 are satisfied by these choices. □

Finally, let us look at a sharpening of Theorem 2, where it was shown that under certain conditions $E[Y_n]$ was bounded.

**Lemma 16.** Suppose that $\lim_{k \to \infty}(p_{k+1}/p_k) = p \in (0, 1)$. Let $k(n) = \min\{k: p_k \leqslant 1/n\}$. Let $L$ denote the limit points of $np_{k(n)}$. Then $(p, 1] \subset L \subset [p, 1]$.

*Proof.* It is relatively straightforward to establish the upper inclusion. Observe that

$$\frac{p_{k(n)}}{p_{k(n)-1}} \leqslant np_{k(n)} \leqslant 1$$

The lower inclusion is a little more unpleasant. Choose $x \in (p, 1]$, and define the positive integer sequence $n_j$ by $n_j = \max(1, [x/p_j])$, where $[y]$ stands for the integer part of $y$. Observe that for all large positive integers $j$ we have

$$p_j \leqslant \frac{p_j}{x} \leqslant \frac{1}{n_j} \leqslant \frac{p_j}{x - p_j} = \frac{p_{j-1}}{x - p_j} \frac{p_j}{p_{j-1}}$$

Therefore, for all range $j$, since $x > p$ and $p_j$ converges to 0, $p_{k(n_j)} = p_j$. Therefore,

$$\lim_{j \to \infty} n_j p_{k(n_j)} = x$$

which finishes the proof of the lemma. □

Eisenberg, Stengle, and Strang[2] analyze the behavior as $n \to \infty$ of

$$\Pr(S_n) \equiv n \sum_{k=2}^{\infty} p_k \left(1 - \sum_{j=k}^{\infty} p_j\right)^{n-1}$$

For geometric distributions, we will see that $E[Y_n]$ and $\Pr(S_n)$ have the same asymptotic behavior.

**Lemma 17.** Suppose $p \in (0, 1)$. Then for every $x \in (0, \infty)$

$$f_p(x) \equiv x \sum_{k=-\infty}^{\infty} \exp(-xp^k)\, p^k$$

defines a continuous function on $(0, \infty)$ with the property that $f_p(p^x)$ defines a nonconstant infinitely differentiable function of period 1 on $(-\infty, \infty)$.

*Proof.* This is a restatement of Theorem 3 of Eisenberg, Stengle, and Strang [1993]. □

**Theorem 18.** Suppose that $\lim_{k \to \infty}(p_{k+1}/p_k) = p \in (0, 1)$. Let $f_p$ be as in Lemma 17. Then the range of $f_p$ is equal to the set of limit points of $E[Y_n]$.

*Proof.* Let $k(n)$ be as in Lemma 16. Observe that

$$E[Y_{n+1}] = n \sum_{k<k(n)} (1-p_k)^n p_k + n \sum_{k \geq k(n)} (1-p_k)^n p_k + o(1) \quad (3.4)$$

Now, suppose that $n \to \infty$ in such a way that $np_{k(n)} \to x \in [p, 1]$. Then

$$n \sum_{k \geq k(n)} (1 - p_k)^n p_k = np_{k(n)} \sum_{k \geq k(n)} \left(1 - p_{k(n)} \prod_{j=0}^{k-1} \frac{p_{j+k(n)+1}}{p_{j+k(n)}}\right)^n \prod_{k=0}^{k-1} \frac{p_{j+k(n)+1}}{p_{j+k(n)}}$$

so it follows from the dominated convergence theorem that along that same subsequence

$$n \sum_{k \geq k(n)} (1-p_k)^n p_k \to x \sum_{k=0}^{\infty} \exp(-xp^k)\, p^k$$

Next, we will show that the dominated convergence theorem implies that

$$\sum_{k<k(n)} (1-p_{k(n)-k})^n \frac{p_{k(n)-k}}{p_{k(n)}} \to x \sum_{k=1}^{\infty} \exp(-xp^{-k})\, p^{-k}$$

along any sequence where $np(k(n))$ converges to $x$. This is rather easy to see: For $k$ bounded away from 1 and all large $n$, since $k(n) \to \infty$, there are $q_i \in (0, 1)$ such that

$$(1 - p_{k(n)-k})^n \frac{p_{k(n)-k}}{p_{k(n)}} \leqslant (1 - q_1^{-k} p_{k(n)})^n q_2^{-k} \leqslant \exp(-q_1^{-k} p)\, q_2^{-k}$$

Since this last function of $k$ is summable, the dominated convergence theorem is applicable. Therefore we have shown that if $np_{k(n)} \to x$ along some subsequence, then $E[Y_n]$ converges to $f_p(x)$ along that same subsequence. Hence the range of $f$ is contained in the limit points of $E[Y_n]$. (Here we use that $f(p) = f(1)$, just in case $x = p$ and $p$ is not a limit point of $np_{k(n)}$.)

The reverse inclusion is easy: The preceding part of the proof shows that $E[Y_n]$ is bounded (as does Theorem 2). Along any convergent subsequence we can restrict to a subsequence along which $np_{k(n)} \to x \in [p, 1]$, so the convergent subsequence of $E[Y_n]$ must converge to $f(x)$. □

If Theorem 2 is applied when the hypotheses of Theorem 18 are satisfied, then the difference in the error bounds of Theorem 2 **increases** from $1/e$ to $3/e$ as $p$ increases from 0 to 1. Numerical evidence suggests that the diameter of the range of $f_p(\cdot)$ **decreases** from $1/\varepsilon$ to 0 as $p$ increases from 0 to 1. Furthermore, for any $p \in (0, 1)$, for any positive integers $M$ and $N$, for all $x \in [0, 1]$,

$$0 < f_p(p^x) - \sum_{k=-M}^{N} p^{k+x} \exp(-p^{(k+x)}) \leqslant \frac{p^{N+1}}{1-p} + \frac{\exp(-(1/p)^{M-1})}{\log(1/p)}$$

Choosing $p = 1/2$ as in Key,[2] for $x \in [0, 1]$ we can numerically estimate the diameter of the range of the partial sum

$$\sum_{k=-5}^{20} (1/2)^{k+x} \exp(-(1/2)^{(k+x)})$$

to be no more than $3 \times 10^{-5}$ and the error in replacing $f_{1/2}((\frac{1}{2})^x)$ with the partial sum for $x \in [0, 1]$) to be no more than $1.12 \times 10^{-6}$. The value of the partial sum at $x = 0$ is approximately 1.44270. All that Theorem 2 tells us is that the limit points of $E[Y_n]$ lie between 0.7629 and 1.8106. Corollary 2 of Eisenberg, Stengle, and Strang[3] shows $|f_p(p^x) - (1/\log(1/p))| \leqslant 0.0002$ for all $x$ and $p \geqslant 0.3$.

## ACKNOWLEDGMENT

I should like to thank Robin Pemantle for pointing out to me (the week before I got the page proofs of this paper) Eisenberg, Stengle, and Strang's proof that $f_p$ was not constant.

## REFERENCES


1. Cuevas, Antonio, and Walters, Gilbert G. (1992). A note on estimation of generalized densities. *Comm. Statist. Theory Methods* **21**, 1807–1821.
2. Eisenberg, B., Stengle, G., and Strang, G. (1993). The asymptotic probability of a tie for first place. *Ann. Appl. Prob.* **3**, 731–745.
3. Key, Eric S. (1992). Rare numbers. *J. Theoret. Prob.* **5**, 375–389.