ON THE ASYMPTOTIC DISTRIBUTION OF MAX AND MEX

Thomas S. Ferguson

A form of the asymptotic joint distribution of the maximum and of the minimal excludent of a sample of size n from a discrete distribution with exponential tails is derived. A strong law is given followed by an application to an inconsistent Bayes procedure.

Keywords. asymptotic distributions, exponential tails, inconsistent Bayes.

1. INTRODUCTION AND SUMMARY

The objective of this paper is to develop a form for the asymptotic joint distribution of the maximum (max) and the minimal excludent (mex) of a sample of size n from a discrete distribution on the non-negative integers with exponential tails.

Throughout this paper, p_j for $j = 0, 1, 2, \ldots$ denotes the frequency function of a random variable X defined on the non-negative integers, and $X_1, X_2, \ldots, X_n$ represents a sample from this distribution. Also, M_n denotes the maximum,

$$M_n = \max(X_1, \ldots, X_n)$$

and L_n denotes the minimal excludent,

$$\begin{aligned} L_n &= \operatorname{mex}(X_1, \ldots, X_n) \\ &= \min\{k \geq 0 : X_i \neq k \quad \text{for} \quad i = 1, \ldots, n\}, \end{aligned}$$

the smallest non-negative integer not hit by one of the X_i. The term mex is due to J. H. Conway (1976).

It seems as if the particular form given below for the asymptotic distribution of the max has not been worked out before. This is surprising in view of the relatively complete theory available for the continuous case. An excellent survey of this area may be found in J. Galambos (1978). The lack of attention given to the discrete case is due partly to the fact that the distribution of the maximum of a sample of size n from a distribution on the non-negative integers cannot, in general, be normalized so as to converge to a non-degenerate distribution. This was pointed out in Galambos (1978) page 120, Exercise 8, where "in general" means that the distribution is such that $P(X \geq k+1)/P(X \geq k)$ does not converge to one as $k \to \infty$. When the limit exists and is strictly between zero and one,

$$a = \lim_{k \to \infty} \frac{P(X \geq k+1)}{P(X \geq k)} \qquad 0 < a < 1, \tag{1.1}$$

C. W. Anderson (1970) has derived upper and lower bounds on the distribution of a centered version of the max. A corresponding treatment of the Poisson distribution, in which case $a = 0$, may be found in A. C. Kimber (1983). A form of the asymptotic distribution of max has been described in a special case by Heyde (1971), and in the general case by Iglehart (1977) and by Diaconis and Freedman (personal communication).

We shall see that when (1.1) holds, precise statements can be made regarding the approximate form of the distribution function of a centered maximum valid for all sufficiently large n. This is done in Theorem 1 where it is shown that by taking the limit as $n \to \infty$ over an appropriate subsequence of the integers, the joint distribution of a centered max and mex converges to a member of a family of distributions (2.5), indexed by a parameter, θ. For each large n, one may find the appropriate θ to use to obtain an approximate distribution.

In passing, an interesting semi-independence property is observed for the limiting distribution which implies a strong relation between the limiting marginal distributions of max and mex.

In section 3, it is shown that $L_n/\log(n)$ and $M_n/\log(n)$ converge almost surely, and an application is given to an example due essentially to D. Freedman (1963) of an inconsistent Bayes procedure.

Section 4 contains some examples illustrating the problems involved in choosing the centering constant and θ for a given n. Some tables clarifying the shape of the distribution and suggesting different choices of θ for the marginal distributions are also given.

2. THE MAIN THEOREM

We say that a frequency function, p_i, defined on the non-negative integers has *exponential tails at decay rate* a, $0 < a < 1$, if

$$\frac{\sum_{k+1}^{\infty} p_i}{\sum_{k}^{\infty} p_i} \to a \qquad \text{as} \qquad k \to \infty. \tag{2.1}$$

Lemma 1. *A frequency function,* p_i *for* $i = 0, 1, 2, \ldots$, *has exponential tails at decay rate* a, $0 < a < 1$, *if and only if*

$$\frac{p_{k+1}}{p_k} \to a \qquad \textit{as} \qquad k \to \infty. \tag{2.2}$$

Proof. (*only if*)

$$\frac{p_{k+1}}{p_k} = \frac{\sum_{k+1}^{\infty} p_i - \sum_{k+2}^{\infty} p_i}{\sum_{k}^{\infty} p_i - \sum_{k+1}^{\infty} p_i} \to \frac{a - a^2}{1 - a} = a.$$

(*if*) Given $0 < \epsilon < a$, find an integer K such that for all $k \geq K$,

$$a - \epsilon \leq \frac{p_{k+1}}{p_k} < a + \epsilon.$$

Then, for all $k \geq K$ and $i \geq 0$,

$$(a - \epsilon)^i \leq \frac{p_{k+i}}{p_k} \leq (a + \epsilon)^i.$$

Take the sum on $i \geq 0$, then the limit as $k \to \infty$, and conclude from the arbitrariness of $\epsilon > 0$, that

$$\frac{1}{p_k} \sum_{k}^{\infty} p_i \to \frac{1}{1 - a} \qquad \text{as} \qquad k \to \infty. \tag{2.3}$$

Hence, the conclusion follows from

$$\frac{\sum_{k+1}^{\infty} p_i}{\sum_{k}^{\infty} p_i} = 1 - \frac{p_k}{\sum_{k}^{\infty} p_i} \to 1 - (1 - a) = a.$$

The equivalence of (2.1) and (2.2) is also valid for $a = 0$, and although (2.2) implies (2.1) for $a = 1$, the converse is not true for $a = 1$.

Theorem 1. *Let* $X_1, X_2, \ldots$ *be i.i.d. with frequency function,* $p_i > 0$ *for* $i = 0, 1, 2, \ldots$ *having exponential tails at decay rate* a, $0 < a < 1$, *and let* $M_n = \max(X_1, \ldots, X_n)$ *and* $L_n = \operatorname{mex}(X_1, \ldots, X_n)$. *Let* $\{n_k\}$ *be a sequence of integers such that*

$$n_k p_k \to \theta \qquad \textit{as} \qquad k \to \infty \tag{2.4}$$

for some θ with $0 < \theta < \infty$. Then, as $k \to \infty$,

$$P(L_{n_k} - k = l, M_{n_k} - k = m) \to f(l, m|a, \theta) \qquad \textit{for} \qquad l, m = 0, \pm 1, \pm 2, \ldots$$

where

$$f(l, m|a, \theta) = \prod_{i=-\infty}^{l-1} (1 - e^{-\theta a^i}) \cdot \begin{cases} e^{-\theta a^l}(1 - e^{-\theta a^m})e^{-\theta a^{m+1}/(1-a)} & \textit{for } m > l \\ e^{-\theta a^l/(1-a)} & \textit{for } m = l - 1 \\ 0 & \textit{otherwise.} \end{cases} \tag{2.5}$$

In particular, the limiting marginal distribution of M_n is

$$P(M_{n_k} - k = m) \to (1 - e^{-\theta a^m})e^{-\theta a^{m+1}/(1-a)} \quad \textit{for} \quad m = 0, \pm 1, \pm 2, \ldots \tag{2.6}$$

and the limiting marginal distribution of L_n is

$$P(L_{n_k} - k = l) \to \left[\prod_{i=-\infty}^{l-1} (1 - e^{-\theta a^i}) \right] e^{-\theta a^l} \quad \textit{for} \quad l = 0, \pm 1, \pm 2, \ldots. \tag{2.7}$$

Remark 1. The assumption that all p_i be positive is obviously unneeded for the asymptotic distribution of the max. It is just as clearly needed to obtain the asymptotic distribution of the mex in the form given in the theorem. However, since the assumption of exponential tails implies that $p_i > 0$ for all i sufficiently large, the result of Theorem 1 would hold more generally if we defined L_n to be the minimal excludent of $(X_1, \ldots, X_n)$ among i such that $p_i > 0$.

Remark 2. The asymptotic distribution of the max, (2.6), is the discretized version of the limiting distribution of the max for continuous variables in the "general case", namely, $F(x) = e^{-e^{-x}}$ up to change of location and scale.

Remark 3. Note that for $l > m$, the joint probability (2.5) is a product of the marginal probabilities (2.6) and (2.7). This means that the asymptotic joint distribution can be obtained by choosing l and m independently according to the asymptotic marginal distributions, (2.6) and (2.7), and then changing m to $l - 1$ if $m \leq l$. Strangely, this alteration does not change the marginal distribution of m. We can also obtain the same result by changing, instead, l to $m + 1$ if $l \geq m$. This remark implies that there is a strong relation between the marginal distributions of the limiting variables. This is expressed in the following lemma.

Lemma 2. *Let L and M be integer valued random variables, and let $f_l = P(L = l)$ and $g_m = P(M = m)$, and suppose that the joint distribution of L and M satisfies*

$$P(L = l, M = m) = \begin{cases} f_l g_m & \text{for } l < m \\ C_l & \text{for } m = l - 1 \\ 0 & \text{otherwise} \end{cases}$$

for some numbers C_l for $l = 0, \pm 1, \pm 2, \ldots$, then

$$P(L = l | L > l - 1) = P(M = l - 1 | M \leq l) \tag{2.8}$$

for all l for which both sides exist.

Proof.

$$\begin{aligned} C_l &= f_l - \sum_{m>l} f_l g_m = f_l \sum_{m \leq l} g_m \\ &= g_{l-1} - \sum_{m<l-1} f_m g_{l-1} = g_{l-1} \sum_{m \geq l-1} f_m. \end{aligned}$$

The asymptotic distributions of the max and mex therefore satisfy (2.8). These equations exhibit a very strong relationship between the distributions of L and M, but it is not clear exactly how strong. For example, it is conjectured that for a given distribution of L such that $P(L \geq l) > 0$ for all l, there is at most one distribution of M satisfying (2.8). (There may be none.)

Remark 4. These limiting distributions have a logarithmic-periodic property in θ, namely,

$$f(l, m | a, a\theta) = f(l + 1, m + 1 | a, \theta). \tag{2.9}$$

This is reflected in the fact that from (2.4), $n_{k-1} p_k = n_{k-1} p_{k-1} (p_k / p_{k-1}) \to \theta a$. If we restrict, say, $a < \theta \leq 1$ in these formulas, the remaining distributions may be obtained from one of these by an integer change of location, the same change for both variables.

Proof of Theorem 1. The proof of Theorem 1 is broken into two parts. First it is shown that the result holds if the sample size is chosen according to a Poisson distribution with parameter n, rather than a fixed sample size n. Then, it is shown that these two sampling schemes are asymptotically equivalent.

Part 1. Let $X_1, X_2, \ldots$ be i.i.d. according to $p_i > 0$ for $i = 0, 1, 2, \ldots$, and let N be Poisson with parameter n independent of $X_1, X_2, \ldots$. For $j = 0, 1, 2, \ldots$, let N_j denote the number of X_i among $X_1, X_2, \ldots, X_N$ such that $X_i = j$. It is well known that the random variables $N_0, N_1, \ldots$ are independent (see, for example, Lemma 2 of Ferguson and Klass (1972), and furthermore that N_i has a Poisson distribution with parameter np_i for $i = 0, 1, 2, \ldots$.

Let $M_N = \max(X_1, \ldots, X_N)$ and $L_N = \text{mex}(X_1, \ldots, X_N)$ where we define M_0 and L_0 to be -1 and 0 for definiteness. Then, for $l < m$,

$$\begin{aligned} &P_n(L_N = l, M_N = m) \\ &\quad = P_n(N_0 > 0, \ldots, N_{l-1} > 0, N_l = 0, N_m > 0, N_{m+1} = 0, \ldots) \\ &\quad = \left[\prod_{i=0}^{l-1} (1 - e^{-np_i}) \right] e^{-np_l} (1 - e^{-np_m}) e^{-n \sum_{m+1}^{\infty} p_i}. \end{aligned}$$

We now let $n = n_k \to \infty$ in such a way that $n_k p_k \to \theta$ and use assumption (2.1) which entails (2.2) and (2.3), to conclude

$$P_{n_k}(L_N - k = l, M_N - k = m)$$
$$= \left[\prod_{i=-k}^{l-1}(1 - e^{-n_k p_{k+i}})\right] e^{-n_k p_{k+l}}(1 - e^{-np_{k+m}})e^{-n\sum_{m+1}^{\infty} p_{k+i}}$$
$$\to \left[\prod_{i=-k}^{l-1}(1 - e^{-\theta a^i})\right] e^{-\theta a^l}(1 - e^{-\theta a^m})e^{-\theta a^{m+1}/(1-a)}$$

as $k \to \infty$, term by term. To show that the infinite product converges, it is sufficient to show that for every $\epsilon > 0$ there exists an integer K such that for all sufficiently large k

$$\prod_{i=-k}^{-K}(1 - e^{-n_k p_{k+i}}) > 1 - \epsilon. \tag{2.10}$$

First find K_0 such that $k > K_0$ implies $n_k p_k > \theta_0$ where θ_0 is some number $0 < \theta_0 < \theta$. Find also K_1 such that $k > K_1$ implies $p_{k+1}/p_k > a_0$ where a_0 is some number $a_0 > a$. Let $K > K_0$ and $k > K + K_1$. Then,

$$\prod_{i=-k}^{-K}(1 - e^{-n_k p_{k+i}}) > 1 - \sum_{i=K}^{k} e^{-n_k p_{k-i}}$$
$$> 1 - \sum_{i=K}^{k} e^{-\theta_0 p_{k-i}/p_k}$$
$$> 1 - \sum_{i=K}^{k-K_1} e^{-\theta_0 a_0^{-i}} - K_1 e^{-\theta_0 \delta/p_k}$$

where $\delta = \min(p_0, p_1, \ldots, p_{K_1-1}) > 0$. Now (2.10) follows since the first sum with the upper limit replaced by $+\infty$ can be made greater than $-\epsilon/2$ by making K sufficiently large, and then the second term can be made greater than $-\epsilon/2$ by making k sufficiently large. Thus, the limiting distribution of (L_N, M_N) is as given in (2.5). for $l < m$. For $l = m + 1$, the result follows similarly.

Part 2. To show the two sampling schemes are equivalent, we put all random variables on the same probability space and show that

$$P_n(L_n = L_N, M_n = M_N) \to 1 \qquad \text{as} \qquad n \to \infty.$$

Let $T_1, T_2, \ldots$ be i.i.d. Poisson random variables with parameter 1, independent of the sequence $X_1, X_2, \ldots$. Let $S_n = \sum_1^n T_i$ and let $L_1' = \text{mex}(X_1, \ldots, X_{S_n})$ and $M_n' =$

$\max(X_1, \ldots, X_{S_n})$. Then the distribution of (L_N, M_N) of part 1 is identical to the distribution of (L_n', M_n'). Moreover,

$$P(L_n = L_n', M_n = M_n') = \sum_{k=0}^{\infty} P(S_n = k) P(L_n = L_k, M_n = M_k)$$
$$\geq \sum_{|k-n|<\alpha\sqrt{n}} P(S_n = k) P(A(\alpha, n)),$$

where α is an arbitrary positive number, and $A(\alpha, n)$ represents the event that no new observations occur among the X_i with i in the interval from $n - \alpha\sqrt{n}$ to $n + \alpha\sqrt{n}$. The probability that a new observation occurs at time $n+1$ is

$$\sum_{j=0}^{\infty} p_j (1-p_j)^n < \sum_{j=0}^{\infty} p_j e^{-np_j}$$
$$\leq \frac{1}{n} \sum_{j=0}^{k} p_j + \sum_{j=k+1}^{\infty} p_j$$
$$\leq \frac{1}{n} + \sum_{j=k+1}^{\infty} p_j$$

where $k = k(n)$ is the largest integer such that $\exp\{-np_j\} \leq 1/n$ for all $j \leq k$ (i.e. $p_j \geq \log(n)/n$ for all $j \leq k$ and $p_{k+1} < \log(n)/n$). Since assumption (2.3) implies $\sum_{k+1}^{\infty} p_j < c\, p_{k+1}$ for some constant $c > 0$ and all k sufficiently large, we can further bound this probability for n sufficiently large by

$$\sum_{j=0}^{\infty} p_j (1-p_j)^n \leq \frac{1}{n} + c\, p_{k+1} \leq \frac{1}{n}(1 + c\log(n)).$$

Hence, the probability that a new observation occurs at X_i for some i in the interval, $n - \alpha\sqrt{n}$ to $n + \alpha\sqrt{n}$, is bounded by

$$\frac{2\alpha\sqrt{n}}{n - \alpha\sqrt{n}}(1 + c\log(n - \alpha\sqrt{n})).$$

By choosing α so that $P(|S_n - n| < \alpha\sqrt{n}) > 1 - (\epsilon/2)$ for all n sufficiently large, and then choosing n sufficiently large so that $P(A(\alpha, n)) > 1 - (\epsilon/2)$, we obtain

$$P(L_n = L_n', M_n = M_n') > (1 - \frac{\epsilon}{2})(1 - \frac{\epsilon}{2}) > 1 - \epsilon.$$

3. A STRONG LAW AND AN APPLICATION

For the application to an inconsistent Bayes estimate, the following strong law is needed. More sensitive strong laws for the max in the geometric case have been found by Diaconis and Freedman (personal communication).

Theorem 2. *Under the assumptions of Theorem 1,*

$$\frac{L_n}{\log(n)} \log(1/a) \to 1 \qquad \text{a.s.}$$
$$\frac{M_n}{\log(n)} \log(1/a) \to 1 \qquad \text{a.s.}$$

Proof. Since $L_n \leq M_n + 1$, it is sufficient to show

$$P(L_n \leq (1-\epsilon)\log(n)/\log(1/a) \quad \text{i.o.}) = 0 \qquad \text{and}$$
$$P(M_n \leq (1+\epsilon)\log(n)/\log(1/a) \quad \text{i.o.}) = 0$$

for all $\epsilon > 0$. First note that for every $a_0 < a$, there exists a constant $c > 0$ such that $p_j > ca_0^j$ for all j. Hence,

$$\begin{aligned} P(L_n \leq K) &= P(\text{at least one} \quad j \leq K \quad \text{is not hit by} \quad X_1, \ldots, X_n) \\ &\leq \sum_{j=o}^{K} (1-p_j)^n \leq \sum_{j=0}^{K} e^{-np_j} \\ &\leq \sum_{j=0}^{K} e^{-nca_0^j} = \sum_{j=0}^{K} e^{-nca_0^{K-j}}. \end{aligned}$$

Next note that $a_0^{-j} \geq 1 + bj$ where $b = (1-a_0)/a_0$, so that

$$P(L_n \leq K) \leq e^{-nca_0^K}/(1 - e^{-ncba_0^K}).$$

For $K = K_n = (1-\epsilon)\log(n)/\log(1/a_0)$, the sum $\sum P(L_n \leq K_n)$ converges, so that $P(L_n \leq K_n \quad \text{i.o.}) = 0$ by Borel-Cantelli. The arbitrariness of $a_0 < a$ and $\epsilon > 0$ implies the same result for $a = a_0$ as well.

To show the result for M_n, note that for any non-decreasing sequence K_n, $P(M_n \geq K_n \quad \text{i.o.}) = P(X_n \geq K_n \quad \text{i.o.})$. Let $a_1 > a$ and find a constant c such that $p_j \leq ca_0^j$ for all j. Then

$$P(X_n \geq K_n) = \sum_{j=K_n}^{\infty} p_j \leq ca_1^{K_n}/(1-a_1).$$

For $K_n = (1+\epsilon)\log(n)/\log(1/a_1)$, we have $\sum_n P(X_n \geq K_n) < \infty$ so that $P(X_n \geq K_n \quad \text{i.o.}) = 0$ and the theorem follows.

As an application, we consider an example of an inconsistent Bayes procedure in which the mex of the sample plays a role. This example is a modification of an example of Freedman (1963) in which the phenomenon was first observed. According to a theorem of Doob (1948), Bayes procedures are consistent at all points of the parameter space except for a set of points of probability zero under the prior distribution. Therefore, no really strong example like those for inconsistency for maximum likelihood estimates (see for example, Ferguson (1982) can be found. In the example below, there are countably many distributions, indexed by $\theta = 0, 1, 2, \ldots$, with a single limit point, $\theta = \infty$, which is in the support of the prior; yet if $\theta = \infty$ is the true value, the posterior distribution given the sample converges to $\theta = 0$. For other examples of inconsistency of Bayes procedures, see Diaconis and Freedman (1986).

Let the parameter space be $\Theta = \{0, 1, 2, \ldots, \infty\}$. For $\theta = \infty$, let the distribution of X be geometric with parameter .5,

$$p_\infty(x) = (.5)^{x+1} \qquad \text{for} \qquad x = 0, 1, 2, \ldots.$$

For $\theta = 1, 2, \ldots$, let the distribution of X be the same with the modification that the mass at θ is removed and placed at 0,

$$p_\theta(x) = \begin{cases} .5 + (.5)^{\theta+1} & \text{for } x = 0 \\ 0 & \text{for } x = \theta \\ (.5)^{x+1} & \text{for } x = 1, 2, \ldots \text{ and } x \neq \theta. \end{cases}$$

For $\theta = 0$, any finite modification of p_∞ giving positive weight to all $x = 0, 1, 2, \ldots$ would work, but let us take for definiteness the modification in which the mass on 0 and 1 are interchanged,

$$p_0(x) = \begin{cases} (.5)^2 & \text{for } x = 0 \\ (.5) & \text{for } x = 1 \\ (.5)^{x+1} & \text{for } x = 2, 3, \ldots. \end{cases}$$

The prior distribution, $q(\theta)$, will give positive weight to each of $\theta = 0, 1, 2, \ldots$, and though $\theta = \infty$ gets zero weight, it is in the support of the prior. Yet if $\theta = \infty$ is the true value, the posterior distribution does not give more and more weight to those distributions close to $\theta = \infty$ as $n \to \infty$ as one would hope. Rather, the posterior distribution gives more and more weight to $\theta = 0$.

Theorem 3. *Let $X_1, X_2, \ldots$ be a sample from $p_\theta(x)$, and suppose the prior probabilities, $q(\theta)$, are positive for $\theta = 0, 1, 2, \ldots$, and $q(\infty) = 0$. Then, if $\theta = \infty$ is the true value, and if $q(\theta) \to 0$ sufficiently fast as $\theta \to \infty$ (how fast is noted in the proof), the* *posterior probability of $\theta = 0$ given $X_1, \ldots, X_n$ tends almost surely to one as $n \to \infty$, i.e.*

$$q(0|X_1, \ldots, X_n) \to 1 \qquad \text{a.s..}$$

Proof. The posterior probabilities are proportional to

$$q(\theta) \prod_1^n p_\theta(X_i) \qquad \theta = 0, 1, 2, \ldots.$$

Any observation $X_i > 1$ does not change the relative weights except to put this weight for $\theta = X_i$ to zero if it is not already zero. So these weights are proportional to

$$\begin{array}{ll} q(0)(.25)^{N_0}(.5)^{N_1} & \text{for} \quad \theta = 0 \\ q(\theta)(.5 + .5^{\theta+1})^{N_0}(.25)^{N_1} I(N_\theta = 0) & \text{for} \quad \theta = 1, 2, \ldots \end{array}$$

where N_j is the number of X_i among $X_1, \ldots, X_n$ that are equal to j, and I represents the indicator function. Let $L_n = \text{mex}(X_1, \ldots, X_n)$. Then, the sum of the weights on $1, 2, 3, \ldots$ is bounded above by

$$\sum_{\theta=L_n}^{\infty} q(\theta)(.5 + .5^{\theta+1})^{N_0}(.25)^{N_1} \leq (.5)^{N_0+2N_1}(1 + .5^{L_n})^{N_0} \sum_{\theta=L_n}^{\infty} q(\theta).$$

Hence the posterior probability of $\theta = 0$ is at least

$$q(0|X_1, \ldots, X_n) \geq \left[1 + 2^{N_0 - N_1}(1 + .5^{L_n})^{N_0} \sum_{\theta \geq L_n} q(\theta)/q(0) \right]^{-1}.$$

From the strong law of large numbers, $N_0/n \to .5$ a.s. and $N_1/n \to .25$ a.s. From Theorem 2, we see that $L_n/\log_2(n) \to 1$ a.s. so that $q(0|X_1, \ldots, X_n) \to 1$ a.s. provided

$$2^{n/4} \sum_{\theta \geq L_n} q(\theta) \to 0 \qquad \text{a.s..}$$

Obviously, this can be achieved by choosing the tail of the prior to go to zero sufficiently rapidly; for example, $\sum_{\theta > n} q(\theta) = 2^{-4^n}$ works.

4. EXAMPLES AND TABLES

In the applications of the formula of Theorem 1, some care must be exercised in choosing the values of k and θ given n, as these values are not uniquely determined.

First consider sampling from the geometric distribution, $p_k = (1-a)a^k$ for $k = 0, 1, 2, \ldots$. For $a < \theta \leq 1$, the equation, $np_k = \theta$ is satisfied uniquely by

$$k = \lceil (\log(n) + \log(1-a))/\log(1/a) \rceil$$
$$\theta = np_k.$$

For example, if $n = 10^9$ and $a = .5$, then $k = 29$ and $\theta = .93\cdots$. In the geometric case, the non-unicity of the choice of k and θ is not important. In fact, if we choose $k = 28$ and $\theta = 1.86\cdots$, we would get the same approximation from equation (2.5) as for $k = 29$ and $\theta = .93\cdots$.

However, if the distribution of X is negative binomial of the form $p_k = (1 - a)^2 a^k (k+1)$ for $k = 0, 1, 2, \ldots$, then $np_k \to 1$ implies that

$$k \sim [\log(n) + \log\log(n) + 2\log(1-a) - \log\log(1/a)] \,/\, \log(1/a),$$

in the sense that for large n the difference is small. For $n = 10^9$ and $a = .5$ again, we find

$$\begin{aligned} &\text{for} \quad k = 33, \quad \theta = .989\cdots, \\ &\text{for} \quad k = 34, \quad \theta = .509\cdots. \end{aligned}$$

Both values of θ are in $(a, 1)$. Either pair (k, θ) may be used in(2.5) but they give slightly different approximations.

For the logarithmic series distribution,

$$p_k = a^{k+1}/((k+1)\log(1/(1-a))) \qquad \text{for} \quad k = 0, 1, 2, \ldots,$$

we have the opposite behavior with

$$k \sim [\log(n) - \log\log(n) + \log(1/a) - \log\log(1/a)] \,/\, \log(1/a).$$

For $n = 6 \times 10^8$ and $a = .5$, we find $np_k = \theta$ solved by

$$\begin{aligned} &\text{for} \quad k = 24, \quad \theta = 1.032\cdots, \\ &\text{for} \quad k = 25, \quad \theta = 0.496\cdots. \end{aligned}$$

For this choice of n, neither value of θ is in $(a, 1)$. The choice of which pair should be used depends on how the approximation is to be used. The general problem requires further investigation.

Formulas (2.5), (2.6) and (2.7) for the asymptotic distribution of max and mex are rather complex and do not lend much insight. To help gain understanding, some tables of these distributions are presented. In Table 1, the asymptotic joint distribution of max and mex is displayed for $a = .5$ and for two values of θ, $\theta = 1$ and $\theta = \sqrt{.5}$. This shows how the mass shifts toward larger values with decreasing θ. It also illustrates one surprising aspect of this distribution that occurs only when $a = .5$, namely, the sum of the subdiagonal elements in both tables is equal to .5. Specifically, let $I(\theta) = P(L = M + 1|\theta, a = .5)$. Then,

$$
\begin{aligned}
I(\theta) &= \sum_{l=-\infty}^{\infty} \left[\prod_{i=-\infty}^{l} (1 - e^{-\theta 2^{-i}}) \right] e^{-\theta 2^{-l}} \\
&= \sum_{l=-\infty}^{\infty} \left[\prod_{i=-\infty}^{l-1} (1 - e^{-\theta 2^{-i}}) \right] (e^{-\theta 2^{-l}} - e^{-\theta 2^{-l+1}}) \\
&= \sum_{l=-\infty}^{\infty} \left[\prod_{i=-\infty}^{l-1} (1 - e^{-\theta 2^{-i}}) \right] e^{-\theta 2^{-l}} - I(\theta).
\end{aligned}
$$

Hence, $2I(\theta) = \sum_{l=-\infty}^{\infty} P(L = l|\theta, a = .5) = 1$, so $I(\theta) = .5$ for all θ.

TABLE 1. The joint asymptotic distribution of max and mex when sampling from a distribution with exponential tails with decay rate, $a = .5$.

$\theta = 1$:

					max							
mex	-3	-2	-1	0	1	2	3	4	5	6	7	
-3		.0000	.0000	.0001	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0003
-2	.0003		.0021	.0043	.0044	.0032	.0019	.0010	.0005	.0003	.0001	.0183
-1		.0180		.0309	.0317	.0229	.0138	.0076	.0040	.0020	.0010	.1328
0			.1148		.0745	.0538	.0324	.0178	.0093	.0048	.0024	.3122
1				.1973		.0560	.0337	.0185	.0097	.0050	.0025	.3253
2					.1280		.0170	.0094	.0049	.0025	.0013	.1644
3						.0364		.0023	.0012	.0006	.0003	.0412
4							.0048		.0002	.0001	.0000	.0052
5								.0003		.0000	.0000	.0003
	.0003	.0180	.1170	.2325	.2387	.1723	.1037	.0569	.0298	.0153	.0077	1.0000

$\theta = \sqrt{.5}$:

					max							
mex	-3	-2	-1	0	1	2	3	4	5	6	7	
-3		.0002	.0006	.0009	.0007	.0005	.0003	.0001	.0001	.0000	.0000	.0035
-2	.0035		.0108	.0147	.0123	.0080	.0046	.0024	.0013	.0006	.0003	.0589
-1		.0554		.0570	.0477	.0310	.0177	.0094	.0049	.0025	.0012	.2279
0			.1725		.0732	.0475	.0271	.0145	.0075	.0038	.0019	.3499
1				.1774		.0343	.0196	.0104	.0054	.0027	.0014	.2526
2					.0752		.0070	.0037	.0019	.0010	.0005	.0898
3						.0145		.0007	.0003	.0002	.0001	.0159
4							.0013		.0000	.0000	.0000	.0014
5								.0001		.0000	.0000	.0001
	.0035	.0556	.1840	.2499	.2091	.1358	.0774	.0414	.0214	.0109	.0055	1.0000

This result implies that if the distribution has exponential tails with decay rate $a = .5$, then $P(L_n = M_n + 1)$ actually converges (to one-half) as $n \to \infty$. Computer results show that for $a \neq .5$, $P(L = M + 1|\theta, a)$ does depend on θ, so in the other cases we do not get convergence as $n \to \infty$.

An alternate proof that $I(\theta) = .5$ can be found using the following more general result of Ferguson and Melolidakis (1984): *Let $X_1, \ldots, X_n$ be i.i.d. geometric with parameter .5, and let K_n be the number of empty cells less than the largest occupied cell. Then K_n has a geometric distribution with parameter .5 for all n.* The event $K_n = 0$ is the same as the event $L_n = M_n + 1$. It would be of interest to find the asymptotic distribution of K_n for an arbitrary geometric distribution. However, the resulting distribution should have a logarithmic-periodic form as in Theorem 1, since $P(L = M + 1|\theta, a)$ depends on θ when $a \neq .5$.

Table 2 exhibits the asymptotic distribution of max for various values of a. The standard theory of the asymptotic distribution of max suggests using as a centering constant a value $k' = k'(n)$ satisfying $n \sum_{i=k'+1}^{\infty} p_i \to \theta$, where θ is close to 1. Such a value of k' should yield better approximations than the one using $np_k \to \theta$ with θ close to 1, since the resulting distribution should have mean closer to zero. For small values of a, this makes no difference, but for $a = .8$ this results in a $k' = k + 6$, and for $a = .9$, $k' = k + 21$. From (2.3), such a change is asymptotically equivalent to using k such that $np_k \to \theta$ where θ is close to $(1 - a)/a$. These are the values of θ used in Table 2.

TABLE 2. Asymptotic distribution of max.

	$\theta = 9.000$	4.000	2.333	1.500	1.000	0.667	0.429	0.250	0.111
M	a=.1	.2	.3	.4	.5	.6	.7	.8	.9
-3	.0000	.0000	.0000	.0000	.0003	.0093	.0386	.0548	.0359
-2	.0000	.0000	.0000	.0019	.0180	.0524	.0757	.0678	.0373
-1	.0000	.0067	.0357	.0802	.1170	.1267	.1097	.0769	.0382
0	.3678	.3611	.3322	.2858	.2325	.1790	.1282	.0814	.0387
1	.5370	.4509	.3729	.3024	.2387	.1809	.1287	.0814	.0387
2	.0852	.1421	.1731	.1818	.1723	.1489	.1160	.0780	.0383
3	.0090	.0312	.0594	.0859	.1037	.1081	.0970	.0720	.0375
4	.0009	.0064	.0186	.0367	.0569	.0727	.0769	.0646	.0365
5	.0001	.0013	.0056	.0151	.0298	.0467	.0587	.0567	.0352
6	.0000	.0003	.0017	.0061	.0153	.0292	.0437	.0488	.0337
7	.0000	.0001	.0005	.0025	.0077	.0180	.0319	.0414	.0321

The asymptotic distribution of mex for various values of a is found in Table 3. Here the approximate choice of θ for a given value of a to center the mode of the distribution between zero and one is as the solution of $e^{-a\theta} + e^{-\theta} = 1$. These are the values of θ chosen for Table 3. Unlike for Table 2, the values of θ are increasing in a, so the appropriate choice of the centralizing k is *smaller* than that given by $np_k \to \theta$ with θ close to one. The choice of the same centralizing k for both max and mex, found in Theorem 1, represents a convenient compromise.

TABLE 3. Asymptotic distribution of mex.

	$\theta = 0.730$	0.773	0.824	0.886	0.962	1.061	1.198	1.406	1.802
L	a=.1	.2	.3	.4	.5	.6	.7	.8	.9
-3	.0000	.0000	.0000	.0000	.0005	.0074	.0302	.0609	.0685
-2	.0000	.0000	.0001	.0039	.0213	.0521	.0835	.0987	.0803
-1	.0007	.0210	.0641	.1086	.1428	.1604	.1587	.1361	.0894
0	.4816	.4519	.4105	.3658	.3192	.2699	.2173	.1600	.0945
1	.4813	.4516	.4102	.3660	.3191	.2698	.2173	.1600	.0945
2	.0362	.0732	.1068	.1351	.1550	.1639	.1587	.1353	.0892
3	.0003	.0023	.0080	.0195	.0374	.0606	.0840	.0961	.0792
4	.0000	.0000	.0002	.0011	.0045	.0136	.0320	.0570	.0661
5	.0000	.0000	.0000	.0000	.0003	.0018	.0087	.0280	.0516
6	.0000	.0000	.0000	.0000	.0000	.0002	.0017	.0113	.0376
7	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0033	.0255

Acknowledgement. The author gratefully acknowledges the help of Michael J. Klass on this problem.

References.

Anderson, C. W. (1970) Extreme value theory for a class of discrete distributions with applications to some stochastic processes, *J. Appl. Prob.* **7**, 99-113.

Conway, J. H. (1976) *On Numbers and Games*, Academic Press, New York.

P. Diaconis, P. and Freedman, D. A. (1986) On inconsistent Bayes estimates of location, *Ann. Statist.* **14**, 68-87.

Doob, J. L. (1948) Application of the theory of martingales, *Le Calcul des Probabilités et ses Applications.* Colloques Internationaux du Centre National de la Researche Scientifique, Paris, 23-28.

Ferguson, T. S. and Klass, M. J. (1972) A representation of independent increment processes without Gaussian components, *Ann. Math. Statist.* **43**, 1634-1643.

Ferguson, T. S. (1982) An inconsistent maximum likelihood estimate, *J. Amer. Stat-ist. Assoc.* **77**, 831-834.

Ferguson, T. S. and Melolidakis, C. (1984) Problem E 3061, *Amer. Math. Mo.* **91**, 580. Solution by D. E. Knuth (1987) **94**, 189.

Freedman, D. A. (1963) On the asymptotic behavior of Bayes estimates in the discrete case, *Ann. Math. Statist.* **34**, 1386-1403.

Galambos, J. (1978) *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, New York.

Heyde, C. C. (1971) On the growth of the maximum queue length in a stable queue, *Op. Res.* **19**, 447-452.

Iglehart, D. L. (1977) Regenerative simulation for extreme values, *Tech. Report #43*, Dept. of Op. Res., Stanford.

Kimber, A. C. (1983) A note on Poisson maxima, *Zeit. für Wahr.* **63**, 551-552.

Thomas S. Ferguson
Mathematics Department
UCLA
405 Hilgard Avenue
Los Angeles, CA 90024