NUMBER THEORY, BALLS IN BOXES, AND THE ASYMPTOTIC UNIQUENESS OF MAXIMAL DISCRETE ORDER STATISTICS

Jayadev S. Athreya
Department of Mathematics, Iowa State University, Ames, Iowa 50011, U.S.A.

Lukasz M. Fidkowski
Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138, U.S.A.

Received: 12/14/99, Accepted: 2/25/00, Published: 5/19/00

Abstract

We investigate the asymptotic uniqueness of the maximal order statistic of $X_1, X_2, \ldots, X_n$, i.i.d. positive integer random variables, by casting the problem in a balls in boxes setting. We give a necessary and sufficient condition on the distribution of the X_i's for the convergence of the probability of uniqueness as $n \to \infty$. We describe the connection to an interesting problem in number theory. The main techniques used are altering the sample to have random size, specifically, Poisson(n), and Karamata's Tauberian Theorem.

1. Introduction And Main Results

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables taking values on the positive integers, with $P(X = i) = p_i$. Denote by ρ_n the probability that the largest value in the sample is unique, which, in order statistics notation, reads

$$P(X_{(n)} > X_{(n-1)}).$$

What is the asymptotic behavior of ρ_n? The problem can be thought of in the following intuitive manner: Let n balls be thrown independently into an infinite number of boxes, numbered $1, 2, 3, \ldots$ according to the distribution $\{p_i\}_{i=1}^{\infty}$. Then ρ_n is the probability that the highest non-empty box has exactly one ball in it. Denote by $\rho_{n,t}$ the probability that t is the highest box with a ball and that it has exactly one ball in it. Clearly,

$$\rho_{n,t} = np_t(p_1 + p_2 + \ldots + p_{t-1})^{n-1}$$

so that

$$\rho_n = \sum_{t=1}^{\infty} \rho_{n,t} = \sum_{t=1}^{\infty} np_t(1 - \bar{p}_t)^{n-1}$$

where

$$\bar{p}_t = \sum_{j=t}^{\infty} p_j = P(X \geq t).$$

The question(s) we would like to answer is (are): What is $\lim_{n\to\infty} \rho_n$? For what distributions $\{p_i\}_{i=1}^{\infty}$ does the limit exist? The case $p_i = 2^{-i}$ was investigated by Shuguang Li [2]. He showed that $\lim_{n\to\infty} \rho_n$ does not exist in this case. Before we investigate the asymptotic behaviour of ρ_n we consider a slight modification which illustrates the techniques that we use. The modification is as follows: Instead of a sample of fixed size n, let our sample be of random size, specifically, Poisson(n). Then the number of balls in the boxes become independent Poisson random variables with mean $\{np_i\}_{i=1}^{\infty}$ respectively (see Ross [3]). So denoting by ρ'_n the probability that the highest box has only one ball, we have:

$$\rho'_n = \sum_{t=1}^{\infty} np_t e^{-n\bar{p}_t}.$$

Before stating our main theorem, we recall a key lemma: Karamata's Tauberian Theorem, as found, e.g., in Bingham et al. [1], pp. 37-8:

Lemma 1 *Let U be non-decreasing on the reals, with $U(x) = 0$ for $x < 0$, and such that $\hat{U}(s)$ (the Laplace transform) $< \infty$ for all large s. Let l be a slowly varying function (i.e., for any $t > 0, \lim_{x\to\infty} l(tx)/l(x) = 1$) and let $c \geq 0$ and $\rho \geq 0$ be constants. Then the following are equivalent:*

1. *$U(x) \sim cx^{\rho} l(1/x)/\Gamma(1+\rho)$ as $x \to 0+$;*
2. *$\hat{U}(s) \sim cs^{-\rho} l(s)$ as $s \to \infty$.*

If $c = 0$, the above result is to be interpreted to mean that $U(x) = o(x^{\rho} l(1/x)/\Gamma(1+\rho))$ as $x \to 0+$ is equivalent to $\hat{U}(s) = o(s^{-\rho} l(s))$ as $s \to \infty$.

We now prove a theorem that illustrates the technique to be used in the proof of our main theorem:

Theorem 1 *The following are equivalent:*

1. *$\lim_{n\to\infty} \rho'_n$ exists;*
2. *$\lim_{n\to\infty} \rho'_n = 1$;*
3. *$\lim_{t\to\infty} p_t/\bar{p}_t = 0$.*

In our main theorem we show the result holds with ρ'_n replaced by ρ_n.

Proof. We are interested in the behavior of

$$\rho'_n = \sum_{t=1}^{\infty} np_t e^{-n\bar{p}_t} = nE(e^{-n\bar{p}_X}) = n\phi(n)$$

where $\phi(n)$ is the Laplace-Stieltjes transform of the cumulative distribution function of the random variable $\bar{p}_X$, with X an random variable with distribution $\{p_i\}$. For $\lim_{n\to\infty} \rho'_n$ to exist and be positive, we would like to have

$$\phi(n) \sim l/n$$

as $n \to \infty$, and with $l > 0$ as our limit. Now since ϕ is monotone (because it is the moment generating function of $\bar{p}_X$), this is equivalent to $\phi(s) \sim l/s$ as $s \to \infty$ continuously. To get an equivalent condition for this, we use our lemma, i.e., Karamata's Tauberian Theorem with $l(x) = l$, $c = \rho = 1$, and $U(x) = F_{\bar{p}_X}$, where $F_{\bar{p}_X}$ is the cumulative distribution function of $\bar{p}_X$. This gives us the following condition in terms of the original distribution function near zero:

$$F_{\bar{p}_X}(y) \sim ly$$

as $y \to 0^+$. So we are interested in how

$$f(y) = F_{\bar{p}_X}(y)/y = P(\bar{p}_X \leq y)/y$$

varies as $y \to 0$. Let y approach 0 along the sequence $\{\bar{p}_t\}$. Clearly, $P(\bar{p}_X \leq \bar{p}_t) = P(X \geq t) = \bar{p}_t$. So $f(y) = 1$ along this sequence. Thus, if a positive limit were to exist it would have to be 1. Note that this calculation also eliminates the case of the limit being 0, since if we took $c = 0$, we could use our theorem to tell us that $\phi(s) = o(1/n)$ as $n \to \infty$ is equivalent to $f(y) = o(1)$ as $y \to 0^+$, which cannot happen since $f(y) = 1$ along the sequence $\{\bar{p}_t\}$.

Now suppose the limit existed, and thus was 1. Look at $y \in [\bar{p}_t, \bar{p}_{t-1})$. Then $P(\bar{p}_X \leq y) = \bar{p}_t$. So

$$1 \geq f(y) \geq \bar{p}_t/\bar{p}_{t-1}$$

and

$$\lim_{y \to \bar{p}_{t-1}^-} f(y) = \bar{p}_t/\bar{p}_{t-1}.$$

Thus for $f(y) \to 1$ as $y \to 0$, $\bar{p}_t/\bar{p}_{t-1}$ must tend to 1. Now, $\bar{p}_t/\bar{p}_{t-1} = 1 - p_{t-1}/\bar{p}_{t-1}$. So we must have $p_{t-1}/\bar{p}_{t-1} \to 0$. We finish by noting that if $p_{t-1}/\bar{p}_{t-1} \to 0$, $f(y)$ is squeezed and must go to 1. □

Theorem 2 *Theorem 1 holds when we replace ρ'_n with ρ_n.*

Proof. To note that the result holds for ρ_n, we proceed in a similar vein. First we write:

$$\rho_n = \sum_{t=1}^{\infty} np_t(1-\bar{p}_t)^{n-1} = \sum_{t=1}^{\infty} np_t e^{(n-1)\ln(1-\bar{p}_t)} = n\psi(n-1),$$

with $\psi(n)$ representing the Laplace-Stieltjes transform of the cumulative distribution function of the random variable $-\ln(1-\bar{p}_X)$. So for ρ_n we would like, as above with ρ'_n and $\phi(n)$,

$$\psi(n) \sim l/n$$

as $n \to \infty$. As with $\phi(n)$, $\psi(n)$ is a moment generating function and thus monotone, so going to ∞ discretely is the same as going continuously.So we can again use the Tauberian theorem, to give us that this is equivalent to:

$$F(y) \sim ly$$

as $y \to 0^+$ with F the cumulative distribution function of $-\ln(1-\bar{p}_X)$. Now let us inspect F. We have

$$F(y) = P(-\ln(1-\bar{p}_X) \leq y) = P(1-\bar{p}_X \geq e^{-y})$$
$$= P(1-e^{-y} \geq \bar{p}_X) = F_{\bar{p}_X}(1-e^{-y}).$$

Now, as $y \to 0^+$, the variable $x = 1-e^{-y}$ has $x \sim y$. But we know about the behavior of $F_{\bar{p}_X}(x)$ as $x \to 0^+$, in particular that its limit is 1 if it exists at all, and that it exists iff $p_t/\bar{p}_t \to 0$. So our conditions are the same for ρ_n as for ρ'_n. □

2. The Number Theory Connection

How does this problem connect to number theory? Li's original paper [2] considered the following problem: Let a and n be integers with $(2a, n) = 1$. Denote the order of $a \pmod n$ by $e_n(a)$. Denote by $\lambda(n)$ the Carmichael-λ function, which is the maximal order of any element of the multiplicative group mod n. We would like to know when $\lambda(n)/e_n(a)$ is odd.

To do this, we first classify the prime divisors of n into the classes $p = 2^j + 1 \pmod{2^{j+1}}$ as j runs through positive integers. This gives us the highest power of 2 dividing $\varphi(p) = \lambda(p) = p-1$. It is known that the proportion of all primes in the jth class is asymptotically 2^{-j}. Now, the following can be proved easily:

Lemma 2 *For $\lambda(n)/e_n(a)$ to be odd, a must be a quadratic non-residue modulo at least one of the prime factors of n which has maximal j.*

So as n has more prime factors in the highest class, the more likely it is that $\lambda(n)/e_n(a)$ is odd. Asymptotic uniqueness would mean a small probability that $\lambda(n)/e_n(a)$ is odd. So to investigate the likelihood of $\frac{\lambda(n)}{e_n(a)}$ being odd, we model a "balls in boxes" problem as above, with the balls being the prime factors of n, the boxes being the j-classes, and $p_j = 2^{-j}$. In this paper, we have adopted the more general approach with general p_j and shown that $\lim \rho_j$ exists if and only if $\lim_{j\to\infty} \frac{p_j}{\sum_{i\geq j} p_i} = 0$, in which case, the limit is 1. So we can see that our condition is not satisfied in the special case under consideration by Li. It was shown by Li [2] that ρ_j oscillates, and while our result does not give this precision, it is a nice application of Tauberian theorems, and provides a general condition under which this limit does (not) exist.

Note that most "standard" distributions on the integers such as the geometric, Poisson etc. lead to the non-existence of the limit; a normalized zeta-function distribution would lead to the limit existing. Our calculations for the geometric ($p_j = 2^{-j}$), show oscillations in the fourth decimal place; other calculations have been done by Li [2].

Acknowledgements

We would like to thank the National Science Foundation and Michigan Tech University for supporting this research through award DMS-9619889. We would like to thank our colleagues Ashwani Sastry, Ben Wieland, and Donald Ying for useful discussions. We would like to thank Krishna B. Athreya for valuable discussions on the Poisson modification and Tauberian theorems. We would like to thank Carl Pomerance and Robert Vaughan for discussions on the number theory connection. Finally, we would like to thank our REU director, Anant Godbole, not only for discussions but also for the opportunity to participate in the Michigan Tech program.

Acknowledgement of Priority (added: 4/25/01)

It has recently been brought to our attention that the results we obtained had been obtained some years ago by Eisenberg et. al. in a series of papers from the mid-1990's (Stat and Prob. Letters 23 (1995), 203-209; Annals of Applied Probability 1993, Vol. 3, No. 3, 731-745; J. Math. Anal. and Applications, 198, 458-472 (1996), article 0092.). Even the proofs appear to be quite similar. However, neither we, nor our advisor A. Godbole, had knowledge of this previous work, and thus our work was completely independent. We also believe that the connection to number theory was not explored in the previous papers.

We would like to thank Bennet Eisenberg for bringing this matter to our attention recently.

References

[1] N.H. Bingham, C.M. Goldie, J.L. Teugels, *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1987.

[2] S. Li, *On Artin's Conjecture for Composite Moduli*, Ph.D. Thesis, University of Georgia, 1998.

[3] S. Ross, *A First Course in Stochastic Processes*, MacMillan, 1998.