BULLETIN OF THE
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
Volume 32, Number 2, June 2004

LIMIT THEOREMS FOR ARRAYS OF MINIMAL ORDER STATISTICS

BY

ANDRÉ ADLER

Abstract. Let $\{X_{nj}, 1 \leq j \leq m_n, n \geq 1\}$ be independently distributed Pareto random variables with parameter $p_n \to 0$. Let $X_{n(k)}$ be the k^{th} smallest order statistic from a sample of size $m_n \to \infty$. This paper establishes unusual limit theorems involving weighted sums for the sequence $\{X_{n(k)}, n \geq 1\}$ where our two parameters p_n and m_n are inversely related.

Consider independently distributed random variables $\{X_{nj}, 1 \leq j \leq m_n, n \geq 1\}$ with density $f_{X_{nj}}(x) = p_n x^{-p_n-1} I(x \geq 1)$ where $p_n > 0$. Let $X_{n(k)}$ be the k^{th} smallest order statistic from each row of our array. Thus the density of $X_{n(k)}$ is

$$f_{X_{n(k)}}(x) = \frac{p_n \cdot m_n!}{(k-1)!(m_n-k)!} x^{-p_n(m_n-k+1)-1}(1-x^{-p_n})^{k-1} I(x \geq 1).$$

In this paper we will explore limit theorems involving weighted sums of $\{X_{n(k)}, n \geq 1\}$. If we keep $\liminf_{n\to\infty} p_n > 0$, then it follows that $p_n m_n \to \infty$, so $X_{n(k)}$ has a finite first moment for all large n, thus a classical Strong law of Large Numbers will prevail. Hence we need to have $p_n \to 0$ whenever $m_n \to \infty$. In the past it was not difficult to obtain results when dealing with a small sample (see Adler [2]) or when dealing with the larger order statistics in either the finite or unbounded sample situation (see Adler [1]). This does

Received by the editors February 14, 2003 and in revised form May 30, 2003.
AMS 2000 Subject Classification: 60F15.

not seem to be the case when observing the smaller order statistics when the sample size approaches infinity. What proves to be even more unusual is that the strong law (Theorem 1) involving the smallest order statistic has total freedom in the selection of our sample size, which is not true when observing any other of our minimal order statistics.

As usual we define $\lg x = \log(\max\{e, x\})$ in order to avoid dividing by zero. Also we use the constant C to denote a generic real number that is not necessarily the same in each appearance.

In our first theorem the sample size, m_n, doesn't need to approach infinity. Our parameters, m_n and p_n, can be any real numbers as long as $m_n \in Z^+$ and $p_n = 1/m_n$.

Theorem 1. *If* $p_n m_n = 1$ *and* $\alpha > -2$, *then*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \frac{(\lg n)^\alpha}{n} X_{n(1)}}{(\lg N)^{\alpha+2}} = \frac{1}{\alpha+2} \quad \text{almost surely.}$$

Proof. Set $a_n = (\lg n)^\alpha/n$, $b_n = (\lg n)^{\alpha+2}$ and $c_n = b_n/a_n = n(\lg n)^2$. The density of $X_{n(1)}$ is

$$f_{X_{n(1)}}(x) = p_n \cdot m_n x^{-p_n m_n - 1} I(x \geq 1)$$

since $k = 1$. Next, if we set $p_n m_n = 1$ we have $f_{X_{n(1)}}(x) = x^{-2} I(x \geq 1)$. Thus

$$\sum_{n=1}^{\infty} P\{X_{n(1)} > c_n\} = \sum_{n=1}^{\infty} \int_{c_n}^{\infty} \frac{dx}{x^2} = \sum_{n=1}^{\infty} \frac{1}{c_n} = \sum_{n=1}^{\infty} \frac{1}{n(\lg n)^2} < \infty.$$

Similarly

$$\sum_{n=1}^{\infty} \frac{1}{c_n^2} E X_{n(1)}^2 I(1 \leq X_{n(1)} \leq c_n) = \sum_{n=1}^{\infty} \frac{1}{c_n^2} \int_1^{c_n} dx < \sum_{n=1}^{\infty} \frac{1}{c_n} < \infty.$$

Since

$$EX_{n(1)}I(1\le X_{n(1)}\le c_n)=\int_1^{c_n}\frac{dx}{x}=\log c_n\sim\log n$$

we have by the Khintchine-Kolmogorov Convergence Theorem (see Chow and Teicher [3])

$$\begin{aligned}\frac{\sum_{n=1}^N\frac{(\lg n)^\alpha}{n}X_{n(1)}}{(\lg N)^{\alpha+2}} &\sim \frac{\sum_{n=1}^N\frac{(\lg n)^\alpha}{n}EX_{n(1)}I(1\le X_{n(1)}\le c_n)}{(\lg N)^{\alpha+2}}\\ &\sim \frac{\sum_{n=1}^N\frac{(\lg n)^{\alpha+1}}{n}}{(\lg N)^{\alpha+2}}\to\frac{1}{\alpha+2}\end{aligned}$$

with probability one.

It should be pointed out that if $p_nm_n>1$, then $EX_{n(1)}<\infty$ and so classical strong laws exist and nothing unusual happens. What proves to be quite unusual is that if we look at some other minimal order statistics we are forced to have our sample size grow at a very slow rate.

Theorem 2. *If* $m_n=[\![\lg n]\!]$, $p_n(m_n-k+1)=1$ *and* $\alpha>-k-1$, *then*

$$\lim_{N\to\infty}\frac{\sum_{n=1}^N\frac{(\lg n)^\alpha}{n}X_{n(k)}}{(\lg N)^{\alpha+k+1}}=\frac{\gamma_k}{\alpha+k+1}\quad\text{almost surely}$$

where

$$\gamma_k=\frac{1}{(k-1)!}\left[1+\sum_{j=1}^{k-1}\frac{\binom{k-1}{j}(-1)^{j+1}}{je^j}-\sum_{j=1}^{k-1}\frac{1}{j}\right].$$

Proof. Set $a_n=(\lg n)^\alpha/n$, $b_n=(\lg n)^{\alpha+k+1}$ and $c_n=b_n/a_n=n(\lg n)^{k+1}$. The density of $X_{n(k)}$ is

$$\begin{aligned}f_{X_{n(k)}}(x) &= \frac{p_n\cdot m_n\cdots(m_n-k+1)}{(k-1)!}x^{-p_n(m_n-k+1)-1}(1-x^{-p_n})^{k-1}I(x\ge 1)\\ &= \frac{m_n\cdots(m_n-k+2)}{(k-1)!}x^{-2}(1-x^{-p_n})^{k-1}I(x\ge 1).\end{aligned}$$

Thus

$$\sum_{n=1}^{\infty} P\{X_{n(k)} > c_n\} < \sum_{n=1}^{\infty} m_n^{k-1} \int_{c_n}^{\infty} \frac{dx}{x^2} = \sum_{n=1}^{\infty} \frac{m_n^{k-1}}{c_n}$$
$$< C\sum_{n=1}^{\infty} \frac{(\lg n)^{k-1}}{n(\lg n)^{k+1}} = C\sum_{n=1}^{\infty} \frac{1}{n(\lg n)^2} < \infty.$$

Likewise

$$\sum_{n=1}^{\infty} \frac{1}{c_n^2} EX_{n(k)}^2 I(1 \le X_{n(k)} \le c_n) < \sum_{n=1}^{\infty} \frac{m_n^{k-1}}{c_n^2} \int_1^{c_n} dx < \sum_{n=1}^{\infty} \frac{m_n^{k-1}}{c_n} < \infty.$$

Our truncated expectation is a bit more difficult in this case. It is important to note that in this situation

$$c_n^{p_n} = [n(\lg n)^{\alpha+1}]^{\frac{1}{[\![\lg n]\!]-k+1}} \to e.$$

Hence

$$\begin{aligned}
& EX_{n(k)} I(1 \le X_{n(k)} \le c_n) \\
\sim\ & \frac{m_n^{k-1}}{(k-1)!} \int_1^{c_n} x^{-1}(1-x^{-p_n})^{k-1} dx \\
=\ & \frac{m_n^{k-1}}{(k-1)!} \Big[\int_1^{c_n} \frac{dx}{x} + \sum_{j=1}^{k-1} \binom{k-1}{j} (-1)^j \int_1^{c_n} x^{-p_n j-1} dx\Big] \\
=\ & \frac{m_n^{k-1}}{(k-1)!} \Big[\log c_n + \sum_{j=1}^{k-1} \frac{\binom{k-1}{j}(-1)^{j+1}}{p_n j c_n^{p_n j}} + \sum_{j=1}^{k-1} \frac{\binom{k-1}{j}(-1)^j}{p_n j}\Big] \\
\sim\ & \frac{m_n^{k}}{(k-1)!} \Big[\log c_n^{p_n} + \sum_{j=1}^{k-1} \frac{\binom{k-1}{j}(-1)^{j+1}}{j c_n^{p_n j}} - \sum_{j=1}^{k-1} \frac{1}{j}\Big] \\
\sim\ & m_n^k \gamma_k
\end{aligned}$$

by applying equation 0.155#4 from page 4 of Gradshteyn and Ryzhik [4].

Therefore

$$\frac{\sum_{n=1}^{N}\frac{(\lg n)^{\alpha}}{n}EX_{n(k)}I(1\le X_{n(k)}\le c_n)}{(\lg N)^{\alpha+k+1}}\sim\frac{\gamma_k\sum_{n=1}^{N}\frac{(\lg n)^{\alpha+k}}{n}}{(\lg N)^{\alpha+k+1}}\to\frac{\gamma_k}{\alpha+k+1}$$

as $N\to\infty$. Thus completing this proof.

It should be noted that Theorem 2 also applies to the smallest order statistic, but of course one should use Theorem 1 in that case. But, if you wish to use Theorem 2, note that $\gamma_1 = 1$. The difficulty that arises with $k \ge 2$ is with the conflict between having the first two series converging and having our normalized first moment converging to a nonzero constant. We were able to get around this problem by allowing our sample size to grow at a logarithmic rate. In our next result we show that if we slightly increase that rate of increase our problem still persists, but we can still find an Exact Strong Law involving these minimal order statistics. In our next theorem we show what happens when we observe the second smallest order statistic and slightly increase our sample size. Similar theorems can be achieved for other, $k \ge 3$, order statistics.

Theorem 3. *If* $m_n = [\![(\lg n)^m]\!]$, *where* $m > 2$, $p_n(m_n - 1) = 1$ *and* $\alpha > -2m + 1$, *then*

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}\frac{(\lg n)^{\alpha}}{n}X_{n(2)}}{(\lg N)^{\alpha+2m-1}}=\frac{1}{2(\alpha+2m-1)}\quad \textit{almost surely}.$$

Proof. Set $a_n = (\lg n)^{\alpha}/n$, $b_n = (\lg n)^{\alpha+2m-1}$ and $c_n = b_n/a_n = n(\lg n)^{2m-1}$. The density of $X_{n(2)}$ is $f_{X_{n(2)}}(x) = m_n x^{-2}(1 - x^{-p_n})I(x \ge 1)$.

Here

$$\begin{aligned}\sum_{n=1}^{\infty}P\{X_{n(2)}>c_n\}&<\sum_{n=1}^{\infty}m_n\int_{c_n}^{\infty}\frac{dx}{x^2}=\sum_{n=1}^{\infty}\frac{m_n}{c_n}\\ &< C\sum_{n=1}^{\infty}\frac{(\lg n)^m}{n(\lg n)^{2m-1}}=C\sum_{n=1}^{\infty}\frac{1}{n(\lg n)^{m-1}}<\infty\end{aligned}$$

and

$$\sum_{n=1}^{\infty} \frac{1}{c_n^2} EX_{n(2)}^2 I(1 \le X_{n(2)} \le c_n) < \sum_{n=1}^{\infty} \frac{m_n}{c_n^2} \int_1^{c_n} dx < \sum_{n=1}^{\infty} \frac{m_n}{c_n} < \infty.$$

In this case we need to note that

$$c_n^{p_n} = [n(\lg n)^{2m-1}]^{\frac{1}{[\![(\lg n)^m]\!]-1}} \to 1$$

thus

$$\begin{aligned} EX_{n(2)} I(1 \le X_{n(2)} \le c_n) &= m_n \int_1^{c_n} x^{-1}(1 - x^{-p_n})dx \\ &= m_n \int_1^{c_n} (x^{-1} - x^{-p_n-1})dx \\ &= \frac{m_n}{p_n}(\log c_n^{p_n} - 1 + c_n^{-p_n}) \\ &\sim \frac{m_n^2}{2}(c_n^{p_n} - 1)^2 \\ &\sim \frac{1}{2}(\lg n)^{2m-2} \end{aligned}$$

since

$$\frac{\log x - 1 + x^{-1}}{(x-1)^2} \to \frac{1}{2} \quad \text{as} \quad x \to 1$$

and

$$\lg n(c_n^{p_n} - 1) \to 1 \quad \text{as} \quad n \to \infty$$

or should we say as $c_n^{p_n} \to 1$.

Therefore

$$\frac{\sum_{n=1}^{N} \frac{(\lg n)^{\alpha}}{n} EX_{n(2)} I(1 \le X_{n(2)} \le c_n)}{(\lg N)^{\alpha+2m-1}} \sim \frac{\sum_{n=1}^{N} \frac{(\lg n)^{\alpha+2m-2}}{n}}{2(\lg N)^{\alpha+2m-1}} \to \frac{1}{2(\alpha+2m-1)}$$

as $N \to \infty$. Thus completing this proof.

Acknowledgment.

I would like to thank the referee for his/her careful reading and kind remarks.

References

1. A. Adler, *Limit theorems for arrays of maximal order statistics*, Probab. Math. Statist., **22** (2002), 211-220.

2. A. Adler, *Exact laws for sums of order statistics from the Pareto distribution*, Bull. Inst. Acad. Sinica, **31** (2003), 181-193.

3. Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, 3^{rd} ed., Springer, New York, 1997.

4. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, 4^{th} ed., Academic Press, New York, 1980.

Department of Mathematics, Illinois Institute of Technology, Chicago, Illinois, 60616, U.S.A.

E-mail: adler@iit.edu