ONE SIDED LIMIT THEOREMS FOR THE RANGE OF A PARETO

ANDRÉ ADLER

Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois, 60616, USA.
E-mail: adler@iit.edu

Abstract

The limiting behaviour of $\sum_{k=1}^{n} a_k R_k / b_n$, where R_k is the range from our k^{th} sample of Pareto random variables is explored. Here we show that a weak laws of large numbers holds, while the corresponding strong law of large numbers fails. We use the weak law to show the lower and upper almost sure limits of our normalized weighted sum. The underlying density is $f(x) = x^{-2} I(x \geq 1)$ and the partial sum consists of the weights times the range from our samples. We also show how these weights must be of a certain form.

1. Introduction

Instead of studying the ratio of order statistics for random variables with infinite mean, here we will investigate the range. There are many papers that have examined the limiting behaviour of the ratio of these type of order statistics. Some of these papers have used this very same Pareto distribution, some of these papers have looked at exponentials, see [7], the uniform distribution, see [8], and a collection of various distributions, see [6]. The earliest papers on the ratios of Pareto's can be found in [1] and [2]. Here we are examining the limiting behaviour of the difference, between the largest and smallest order statistics. We look at sample sizes that grow at a particular rate, m_n.

We need to mention that the constant, C, used in the proofs denotes a generic real number that is not necessarily the same in each appearance. It is used either as an upper bound or as a lower bound in order to establish the convergence or divergence of our various series. We define $\lg x = \ln(\max\{e, x\})$, $\lg_2 x = \lg(\lg x)$ and $\lg_3 x = \lg(\lg(\lg x))$. Also, we pull slowly varying functions outside of integrals and sums quite freely using results from [5], pages 275-284.

2. Preliminary Results

The underlying distribution is the classic Pareto distribution, $f(x) = x^{-2}I(x \geq 1)$. Let $\{X_{k,j}, 1 \leq k \leq n, 1 \leq j \leq m_k\}$ be independent and identically Pareto distributed random variables. The order statistics are no longer independent and they are denoted by $\{X_{k(1)}, \ldots, X_{k(m_k)}\}$, where $X_{k(1)} \leq X_{k(2)} \leq \cdots \leq X_{k(m_k)}$. Next we observe the smallest and largest order statistics from this sample, $X_{k(1)}$ and $X_{k(m_k)}$. From these two random variables we obtain the range, $R_k = X_{k(m_k)} - X_{k(1)}$, $k = 1, 2, \ldots, n$. From [3], we have $f_{R_k}(r) \sim m_k r^{-2}$, as $r \to \infty$.

We will establish several things. First of all we will obtain a weak law of large numbers. And from our weak law we will obtain the almost sure lower bound on our normalized partial sums. We will also show that the almost sure upper limit is infinity, which proves that there isn't a corresponding strong law here, that this weak law is optimal. This further shows that in our Exact Strong Laws in [3], we had to set our coefficients, $a_k = L(k)/k$, for some slowly varying function, $L(x)$. The most common slowly varying function is our logarithm, which is used quite extensively in these Exact Strong Laws and here as well. In this paper we look at what happens if we ever so slightly increase our coefficients. In [3] we obtained strong laws for normalized partial sums of the form

$$\frac{\sum_{k=1}^{n} \frac{(\lg k)^{\alpha}}{k} R_k}{(\lg n)^{\alpha+\beta+2}}$$

for these same random variables. Where α can be any real number, as long as $\alpha + \beta + 2$ is positive. We will generalize $(\lg k)^{\alpha}$ to be any slowly varying function $L(x)$ and increase $1/k$ to k^{α}, as long as $\alpha > -1$. Note that whenever $\alpha \neq -1$, we cannot obtain an Exact Strong Law. (When $\alpha < -1$ these limits

will be zero and in this paper we show what happens when α exceeds minus one.)

3. Main Results

We start with our weak law and conclude with the almost sure limiting behaviour for these same normalized partial sums.

Theorem 1. *If $\alpha > -1$ and $\{X_{k,1}, \ldots, X_{k,m_k}\}$ is a sample from the Pareto distribution, with sample size $m_k \sim \gamma(\lg k)^\beta$, where the range is defined as $R_k = X_{k(m_k)} - X_{k(1)}$, then for all γ and β positive*

$$\frac{\sum_{k=1}^n L(k)k^\alpha R_k}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}} \xrightarrow{P} \frac{\gamma}{\alpha+1}.$$

Proof. Let $a_k = L(k)k^\alpha$ and $b_n = L(n)(\lg n)^{\beta+1}n^{\alpha+1}$. We will use the Weak Law from page 356 of [4]. Let $\epsilon > 0$

$$\sum_{k=1}^n P\{a_k R_k/b_n| > \epsilon\} < C\sum_{k=1}^n \int_{\epsilon b_n/a_k}^\infty \frac{m_k dr}{r^2} < C\sum_{k=1}^n \frac{m_k a_k}{b_n}$$

$$< C\frac{\sum_{k=1}^n (\lg k)^\beta L(k)k^\alpha}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}} < \frac{CL(n)(\lg n)^\beta n^{\alpha+1}}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}}$$

$$= \frac{C}{\lg n} \to 0.$$

As for the variance term, we have

$$\sum_{k=1}^n V\left[\frac{a_k R_k}{b_n}I\left(\left|\frac{a_k R_k}{b_n}\right| < 1\right)\right] < C\sum_{k=1}^n \left(\frac{a_k^2}{b_n^2}\right)\int_0^{b_n/a_k} m_k dr < C\sum_{k=1}^n \frac{m_k a_k}{b_n} \to 0$$

once again.

Next, we must compute the expectation from that theorem

$$\sum_{k=1}^n E\left[\frac{a_k R_k}{b_n}I\left(\left|\frac{a_k R_k}{b_n}\right| < 1\right)\right] \sim b_n^{-1}\sum_{k=1}^n a_k \int_1^{b_n/a_k} \frac{m_k dr}{r}$$

$$= b_n^{-1}\sum_{k=1}^n m_k a_k \lg(b_n/a_k) = b_n^{-1}\sum_{k=1}^n m_k a_k \lg(b_n) - b_n^{-1}\sum_{k=1}^n m_k a_k \lg(a_k).$$

It's interesting that both of these terms are equally important

$$b_n^{-1} \sum_{k=1}^{n} m_k a_k \lg(b_n)$$

$$\sim \frac{\gamma \sum_{k=1}^{n} (\lg k)^\beta L(k) k^\alpha \left[\lg(L(n)) + (\beta + 1)\lg_2 n + (\alpha + 1)\lg n\right]}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}}$$

$$\sim \frac{\gamma(\alpha + 1) \sum_{k=1}^{n} (\lg k)^\beta L(k) k^\alpha \lg n}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}}$$

$$\sim \frac{\gamma(\alpha + 1)\left[(\lg n)^\beta L(n)\left(\frac{n^{\alpha+1}}{\alpha+1}\right)\right]\lg n}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}} = \gamma$$

and the other term is

$$b_n^{-1} \sum_{k=1}^{n} m_k a_k \lg(a_k) \sim \frac{\gamma \sum_{k=1}^{n} (\lg k)^\beta L(k) k^\alpha \left[\lg(L(k)) + \alpha \lg k\right]}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}}$$

$$\sim \frac{\alpha\gamma \sum_{k=1}^{n} (\lg k)^{\beta+1} L(k) k^\alpha}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}}$$

$$\sim \frac{\alpha\gamma \left[(\lg n)^{\beta+1} L(n)\left(\frac{n^{\alpha+1}}{\alpha+1}\right)\right]}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}} = \frac{\alpha\gamma}{\alpha+1}.$$

Combining these two terms, we see that our limit is

$$\gamma - \frac{\alpha\gamma}{\alpha+1} = \frac{\gamma}{\alpha+1}$$

which concludes this proof.   □

Now, one may hope or expect that there is a corresponding strong law. That is not the case at all.

Theorem 2. *If $\alpha > -1$ and $\{X_{k,1}, \ldots, X_{k,m_k}\}$ is a sample from the Pareto distribution, with sample size $m_k \sim \gamma(\lg k)^\beta$, where the range is defined as $R_k = X_{k(m_k)} - X_{k(1)}$, then for all γ and β positive*

$$\liminf_{n\to\infty} \frac{\sum_{k=1}^{n} L(k) k^\alpha R_k}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}} = \frac{\gamma}{\alpha+1} \quad \text{almost surely}$$

and

$$\limsup_{n\to\infty} \frac{\sum_{k=1}^{n} L(k) k^\alpha R_k}{L(n)(\lg n)^{\beta+1} n^{\alpha+1}} = \infty \quad \text{almost surely.}$$

Proof. In this proof we set $a_n = L(n)n^\alpha$, $b_n = L(n)(\lg n)^{\beta+1}n^{\alpha+1}$, $c_n = b_n/a_n = n(\lg n)^{\beta+1}$ and also $d_n = c_n/(\lg_2 n)^2 = n(\lg n)^{\beta+1}/(\lg_2 n)^2$.

Since we have convergence in probability from Theorem 1, we can claim that

$$\liminf_{n\to\infty} \frac{\sum_{k=1}^n L(k)k^\alpha R_k}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}} \leq \frac{\gamma}{\alpha+1} \quad \text{almost surely.}$$

Hence we need to prove that

$$\liminf_{n\to\infty} \frac{\sum_{k=1}^n L(k)k^\alpha R_k}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}} \geq \frac{\gamma}{\alpha+1} \quad \text{almost surely.}$$

This is where the sequence d_n comes into play. Clearly

$$b_n^{-1}\sum_{k=1}^n a_k R_k \geq b_n^{-1}\sum_{k=1}^n a_k R_k I(0 \leq R_k \leq d_k)$$

$$= b_n^{-1}\sum_{k=1}^n a_k\left[R_k I(0 \leq R_k \leq d_k) - E\left(R_k I(0 \leq R_k \leq d_k)\right)\right]$$

$$+ b_n^{-1}\sum_{k=1}^n a_k E\left(R_k I(0 \leq R_k \leq d_k)\right).$$

The first term vanishes almost surely by the Khintchine-Kolmogorov Convergence Theorem, see page 113 of [4], and Kronecker's lemma since

$$\sum_{n=1}^\infty \frac{1}{c_n^2} E R_n^2 I(0 \leq R_n \leq d_n) < C\sum_{n=1}^\infty \frac{1}{c_n^2}\int_0^{d_n} m_n dr$$

$$= C\sum_{n=1}^\infty \frac{m_n d_n}{c_n^2} < C\sum_{n=1}^\infty \frac{(\lg n)^\beta n(\lg n)^{\beta+1}/(\lg_2 n)^2}{\left(n(\lg n)^{\beta+1}\right)^2}$$

$$= C\sum_{n=1}^\infty \frac{1}{n\lg n(\lg_2 n)^2} < \infty.$$

And the limit of the second term is

$$b_n^{-1}\sum_{k=1}^n a_k E\left(R_k I(0 \leq R_k \leq d_k)\right)$$

$$\sim b_n^{-1}\sum_{k=1}^n a_k \int_1^{d_k} \frac{m_k dr}{r} = b_n^{-1}\sum_{k=1}^n a_k m_k \lg(d_k)$$

$$\sim \frac{\sum_{k=1}^{n} L(k)k^{\alpha}\gamma(\lg k)^{\beta}\left[\lg k + (\beta+1)\lg_2 k - 2\lg_3 k\right]}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}}$$

$$\sim \frac{\gamma\sum_{k=1}^{n} L(k)k^{\alpha}(\lg k)^{\beta+1}}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}}$$

$$\sim \frac{\gamma L(n)\left(\frac{n^{\alpha+1}}{\alpha+1}\right)(\lg n)^{\beta+1}}{L(n)(\lg n)^{\beta+1}n^{\alpha+1}}$$

$$= \frac{\gamma}{\alpha+1}.$$

Thus showing that the almost sure lower limit is indeed $\gamma/(\alpha+1)$.

The upper limit is easier. Here, we use C in the opposite direction, since we want this series to diverge. Let M be any positive real number, then

$$\sum_{n=1}^{\infty} P\left\{\frac{a_n R_n}{b_n} > M\right\} > C\sum_{n=1}^{\infty}\int_{Mc_n}^{\infty}\frac{m_n dr}{r^2}$$

$$> C\sum_{n=1}^{\infty}\frac{m_n}{c_n}$$

$$> C\sum_{n=1}^{\infty}\frac{(\lg n)^{\beta}}{n(\lg n)^{\beta+1}}$$

$$= C\sum_{n=1}^{\infty}\frac{1}{n\lg n} = \infty.$$

Thus

$$\limsup_{n\to\infty}\frac{\sum_{k=1}^{n} a_k R_k}{b_n} \geq \limsup_{n\to\infty}\frac{a_n R_n}{b_n} = \infty \quad \text{almost surely}$$

which concludes this proof.   □

This theorem combined with [3] shows that in order to obtain a strong law we must have coefficients of the form $a_n = L(n)/n$ for some slowly varying function, $L(x)$. And for each $L(x)$ one must be careful in how one selects the norming sequence in order for the Exact Strong Law to hold.

References

1. A. Adler, Limit theorems for arrays of ratios of order statistics, *Bull. Inst. Math. Acad. Sinica*, **33** (2005), 327-344.

2. A. Adler, Unusual strong laws for arrays of ratios of order statistics, *Brazilian Journal of Probability and Statistics*, **25** (2011), 34-43.

3. A. Adler, Exact strong laws for the range, *Bull. Inst. Math. Acad. Sinica*, **15(1)** (2020), 53-61.

4. Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, third edition, Springer-Verlag, New York, 1997.

5. W. Feller, *An Introduction to Probability Theory and Its Applications*, second edition, Wiley, New York, 1971.

6. P. Matula, P. Kurasinski and A. Adler, On exact strong laws of large numbers for ratios of random variables and their applications, *Communications in Statistics, Theory and Methods*, (2019) doi.org/10.1080/03610926.2019.1586935

7. Y. Miao, R. Wang and A. Adler, Limit theorems for order statistics from exponentials, *Statist. Probab. Lett.* **110** (2016), 51-57.

8. S. F. Xu and Y. Miao, Some limit theorems for ratio of order statistics from uniform random variables, *J. Inequal. Appl.* 2017:295, 1-18. doi.org/10.1186/s13660-017-1569-7