EXACT STRONG LAWS FOR THE RANGE

ANDRÉ ADLER

Department of Applied Mathematics, Illinois Institute of Technology, Chicago, Illinois, 60616, USA.

E-mail: adler@iit.edu

Abstract

In this paper we establish exact strong laws of large numbers for the range of a Pareto random variable. The underlying density is $f(x) = x^{-2}I(x \geq 1)$. Neither the first nor second moments of this random variable exist, which makes these theorems unusual. The results are of the form $\sum_{i=1}^{n} a_i R_i/b_n \to \gamma$, as $n \to \infty$, where R_i is the range from the ith sample.

1. Introduction

This paper establishes strong laws for the range from a sample of size m_n from a Pareto distribution. We examine two cases, one where the sample size is constant and another where it grows to infinity at a particular rate. Unusual strong laws have been established for the ratio of order statistics. This was started with the papers [2]-[5] and has exploded in the literature with [7]-[12] and many other papers. Here, we look at the range. So instead of looking at the ratio, we now examine the difference between our order statistics. And once again we see that Exact Strong Laws exist in this environment as well.

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 as an upper bound in order to establish the convergence of our

various series. Also, we define $\lg x = \ln(\max\{e, x\})$ and $\lg_2 x = \lg(\lg x)$, which is not a logarithm with a base of 2. Likewise $\lg_3 x = \lg(\lg(\lg x))$.

2. Preliminary Results

The underlying distribution is the classic Pareto, $f(x) = x^{-2} I(x \geq 1)$. This random variable does not possess a first and hence a second moment. So the usual estimator for the variance is a multiple of the range, which is the largest minus the smallest order statistics from the sample. We will look at two cases, one with a fixed sample size from each set of data, to one that increases in time.

Let $\{X_{ij}, 1 \leq i \leq n, 1 \leq j \leq m_i\}$ be independent and identically Pareto distributed random variables. The order statistics are no longer independent and they are denoted by $\{X_{i(1)}, \ldots, X_{i(m_i)}\}$, where $X_{i(1)} \leq X_{i(2)} \leq \cdots \leq X_{i(m_i)}$. Next we observe the smallest and largest order statistics from this sample, $X_{i(1)}$ and $X_{i(m_i)}$. From these two random variables we obtain the range, $R_i = X_{i(m_i)} - X_{i(1)}$, $i = 1, 2, \ldots n$.

We first obtain the joint density of $X_{i(1)}$ and $X_{i(m_i)}$, which is

$$f(x_1, x_{m_i}) = \frac{m_i!}{(m_i - 2)!} f(x_1)[F(x_{m_i}) - F(x_1)]^{m_i - 2} f(x_{m_i})$$

$$= m_i(m_i - 1) \frac{1}{x_1^2} \left[\left(1 - \frac{1}{x_{m_i}}\right) - \left(1 - \frac{1}{x_1}\right) \right]^{m_i - 2} \frac{1}{x_{m_i}^2} I(1 \leq x_1 \leq x_{m_i})$$

$$= \frac{m_i(m_i - 1)}{x_1^2 x_{m_i}^2} \left[\frac{1}{x_1} - \frac{1}{x_{m_i}} \right]^{m_i - 2} I(1 \leq x_1 \leq x_{m_i}).$$

Next, let $w = x_1$ and $r = x_{m_i} - x_1$. The Jacobian is one and the joint density of W and R_i is

$$f(w, r) = \frac{m_i(m_i - 1)}{w^2(r + w)^2} \left(\frac{r}{(r + w)w} \right)^{m_i - 2} I(w \geq 1) I(r \geq 0).$$

Integrating out the dummy variable, w, we see that the density of R_i is

$$f_{R_i}(r) = m_i(m_i - 1) r^{m_i - 2} \int_1^\infty \frac{dw}{(r + w)^{m_i} w^{m_i}}$$

where $r \geq 0$. The key to this paper is the messy integration which is explained in the following lemma.

Lemma. *As $r \to \infty$ we have*

$$\int_1^\infty \frac{dw}{(r+w)^m w^m} \sim \frac{1}{(m-1)r^m}.$$

Proof. Even though there are $2m$ terms in the partial fractions, we only need to be concerned with just three of them. Let $x = w/r$, then

$$\frac{1}{w^m(r+w)^m} = r^{-2m}\left(\frac{1}{x^m(x+1)^m}\right) = r^{-2m}\left(\sum_{i=1}^m \frac{a_i}{x^i} + \sum_{i=1}^m \frac{b_i}{(x+1)^i}\right).$$

Solving

$$\frac{1}{x^m(x+1)^m} = \sum_{i=1}^m \frac{a_i}{x^i} + \sum_{i=1}^m \frac{b_i}{(x+1)^i}$$

we have

$$1 = x^m(x+1)^m\left(\sum_{i=1}^m \frac{a_i}{x^i} + \sum_{i=1}^m \frac{b_i}{(x+1)^i}\right)$$

or

$$1 = a_1 x^{m-1}(x+1)^m + \cdots + a_m(x+1)^m + b_1 x^m(x+1)^{m-1} + \cdots + b_m x^m.$$

We have $2m$ equations with $2m$ terms, but all we need are two facts. The first is, by setting $x = 0$, we see that $a_m = 1$. The other important fact is that the largest term, x^{2m-1}, gives us the equation $a_1 + b_1 = 0$. Replacing w into the first term in each of the two series leads to

$$\frac{a_1}{x} + \frac{b_1}{x+1} = a_1 r\left(\frac{1}{w} - \frac{1}{w+r}\right).$$

The integration of these two terms produces

$$a_1 r \int_1^\infty \left(\frac{1}{w} - \frac{1}{w+r}\right)dw = a_1 r \lg(1+r).$$

All the other constants are, fortunately, unimportant.

Surprisingly, all the other terms involving the b_i's are of the same order.

If $i \geq 2$, we see that

$$\frac{b_i}{(x+1)^i} = \frac{b_i r^i}{(w+r)^i}$$

which, when integrated leads to

$$b_i r^i \int_1^\infty (w+r)^{-i} dw = \frac{b_i r^i}{(i-1)(1+r)^{i-1}} \sim \frac{b_i}{(i-1)} r.$$

The other terms, those involving the a_i's are of a much higher order. For $i \geq 2$, we have

$$\frac{a_i}{x^i} = \frac{a_i r^i}{w^i}$$

which, when integrated leads to

$$a_i r^i \int_1^\infty w^{-i} dw = \frac{a_i r^i}{i-1}.$$

So, the dominant term is the a_m term and since $a_m = 1$, we have

$$\int_1^\infty \frac{dw}{(r+w)^m w^m} \sim r^{-2m} \left(\frac{a_m r^m}{m-1} \right) = \frac{1}{(m-1)r^m}$$

which concludes this proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Using this Lemma, we have, for all $i \geq 2$

$$f_{R_i}(r) = m_i(m_i - 1)r^{m_i - 2} \int_1^\infty \frac{dw}{(r+w)^{m_i} w^{m_i}} \sim m_i r^{-2}$$

as $r \to \infty$. This fact allows us to establish an Exact Strong Law, see [1].

3. Main Results

We first look at what happens when the sample size is fixed.

Theorem 1. *If $b > 0$ and $\{X_{i1}, \ldots, X_{im}\}$ is a sample from the Pareto distribution, then the strong laws of large numbers for $R_i = X_{i(m)} - X_{i(1)}$ is*

$$\lim_{n \to \infty} \frac{\sum_{i=1}^n \frac{(\lg i)^{b-2}}{i} R_i}{(\lg n)^b} = \frac{m}{b} \quad \text{almost surely.}$$

Proof. Since $f_R(r) \sim mr^{-2}$, it follows that $rP\{R > r\} \sim m$. By applying Example 2 from [1] the conclusion follows. $\square$

Next we examine what happens as the sample size increases at a particular rate.

Theorem 2. *Let* $\{X_{i1}, \ldots, X_{im_i}\}$ *be a sample from the Pareto distribution where* $m_n \sim \gamma(\lg n)^\beta$, *then for* γ, β *and* $\alpha + \beta + 2$ *all positive*

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

Proof. Let $a_n = (\lg n)^\alpha/n$, $b_n = (\lg n)^{\alpha+\beta+2}$ and $c_n = b_n/a_n = n(\lg n)^{\beta+2}$. We use the partition

$$\frac{1}{b_n} \sum_{i=1}^n a_i R_i = \frac{1}{b_n} \sum_{i=1}^n a_i \big[R_i I(|R_i| \le c_i) - ER_i I(|R_i| \le c_i) \big]$$
$$+ \frac{1}{b_n} \sum_{i=1}^n a_i R_i I(|R_i| > c_i) + \frac{1}{b_n} \sum_{i=1}^n a_i ER_i I(|R_i| \le c_i).$$

The first term vanishes almost surely by the Khintchine-Kolmogorov Convergence Theorem, see page 113 of [6], and Kronecker's lemma since

$$\sum_{n=1}^\infty \frac{1}{c_n^2} ER_n^2 I(|R_n| \le c_n) < C \sum_{n=1}^\infty \frac{1}{c_n^2} \int_1^{c_n} m_n dr < C \sum_{n=1}^\infty \frac{m_n}{c_n}$$
$$< C \sum_{n=1}^\infty \frac{1}{n(\lg n)^2} < \infty.$$

The second term vanishes, with probability one, by the Borel-Cantelli lemma since

$$\sum_{n=1}^\infty P\{|R_n| > c_n\} < C \sum_{n=1}^\infty \int_{c_n}^\infty \frac{m_n}{r^2} dr = C \sum_{n=1}^\infty \frac{m_n}{c_n} < \infty.$$

Thus, our almost sure limit follows from the last term in our partition

$$\frac{\sum_{i=1}^n a_i ER_i I(|R_i| \le c_i)}{b_n} \sim \frac{\sum_{i=1}^n \frac{(\lg i)^\alpha}{i} \int_1^{c_i} \frac{m_i}{r} dr}{(\lg n)^{\alpha+\beta+2}}$$

$$= \frac{\sum_{i=1}^{n} \frac{(\lg i)^{\alpha}}{i} m_i \lg c_i}{(\lg n)^{\alpha+\beta+2}}$$

$$\sim \frac{\gamma \sum_{i=1}^{n} \frac{(\lg i)^{\alpha+\beta+1}}{i}}{(\lg n)^{\alpha+\beta+2}}$$

$$\rightarrow \frac{\gamma}{\alpha+\beta+2}.$$

which concludes this proof. □

Next we explore what happens when $\alpha + \beta + 2 = 0$.

Theorem 3. *Let $\{X_{i1}, \ldots, X_{im_i}\}$ be a sample from the Pareto distribution where $m_n \sim \gamma(\lg n)^{\beta}$, then for γ and β both positive*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}} R_i}{\lg {}_2 n} = \gamma \quad \text{almost surely.}$$

Proof. Let $a_n = 1/(n(\lg n)^{\beta+2})$, $b_n = \lg {}_2 n$ and $c_n = b_n/a_n = n(\lg n)^{\beta+2}\lg {}_2 n$. Again, we use the partition

$$\frac{1}{b_n} \sum_{i=1}^{n} a_i R_i = \frac{1}{b_n} \sum_{i=1}^{n} a_i \big[R_i I(|R_i| \le c_i) - E R_i I(|R_i| \le c_i) \big]$$

$$+ \frac{1}{b_n} \sum_{i=1}^{n} a_i R_i I(|R_i| > c_i) + \frac{1}{b_n} \sum_{i=1}^{n} a_i E R_i I(|R_i| \le c_i).$$

The first term vanishes almost surely by the Khintchine-Kolmogorov Convergence Theorem and Kronecker's lemma since

$$\sum_{n=1}^{\infty} \frac{1}{c_n^2} E R_n^2 I(|R_n| \le c_n) < C \sum_{n=1}^{\infty} \frac{1}{c_n^2} \int_{1}^{c_n} m_n dr < C \sum_{n=1}^{\infty} \frac{m_n}{c_n}$$

$$< C \sum_{n=1}^{\infty} \frac{1}{n(\lg n)^2 \lg {}_2 n} < \infty.$$

The second term vanishes, with probability one, by the Borel-Cantelli lemma since

$$\sum_{n=1}^{\infty} P\{|R_n| > c_n\} < C \sum_{n=1}^{\infty} \int_{c_n}^{\infty} \frac{m_n}{r^2} dr = C \sum_{n=1}^{\infty} \frac{m_n}{c_n} < \infty.$$

Thus, our almost sure limit follows from the last term in our partition

$$\frac{\sum_{i=1}^{n} a_i ER_i I(|R_i| \le c_i)}{b_n} \sim \frac{\sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}} \int_1^{c_i} \frac{m_i}{r} dr}{\lg_2 n}$$

$$= \frac{\sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}} m_i \lg c_i}{\lg_2 n}$$

$$\sim \frac{\gamma \sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}} (\lg i)^{\beta} \lg i}{\lg_2 n}$$

$$= \frac{\gamma \sum_{i=1}^{n} \frac{1}{i \lg i}}{\lg_2 n}$$

$$\to \gamma.$$

which concludes this proof.   □

We can continue to get smaller and smaller coefficients and norming sequences. We conclude with one such result.

Theorem 4. *Let $\{X_{i1}, \ldots, X_{im_i}\}$ be a sample from the Pareto distribution where $m_n \sim \gamma(\lg n)^{\beta}$, then for γ and β both positive*

$$\lim_{n \to \infty} \frac{\sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2} \lg_2 i} R_i}{\lg_3 n} = \gamma \quad \text{almost surely.}$$

Proof. Let $a_n = 1/(n(\lg n)^{\beta+2} \lg_2 n)$, $b_n = \lg_3 n$ and

$c_n = b_n/a_n = n(\lg n)^{\beta+2} \lg_2 n \lg_3 n$. Once again, we use the partition

$$\frac{1}{b_n} \sum_{i=1}^{n} a_i R_i = \frac{1}{b_n} \sum_{i=1}^{n} a_i \big[R_i I(|R_i| \le c_i) - ER_i I(|R_i| \le c_i) \big]$$

$$+ \frac{1}{b_n} \sum_{i=1}^{n} a_i R_i I(|R_i| > c_i) + \frac{1}{b_n} \sum_{i=1}^{n} a_i ER_i I(|R_i| \le c_i).$$

The first term vanishes almost surely by the Khintchine-Kolmogorov Convergence Theorem and Kronecker's lemma since

$$\sum_{n=1}^{\infty} \frac{1}{c_n^2} ER_n^2 I(|R_n| \le c_n) < C \sum_{n=1}^{\infty} \frac{1}{c_n^2} \int_1^{c_n} m_n dr < C \sum_{n=1}^{\infty} \frac{m_n}{c_n}$$

$$< C \sum_{n=1}^{\infty} \frac{1}{n(\lg n)^2 \lg_2 n \lg_3 n} < \infty.$$

The second term vanishes, with probability one, by the Borel-Cantelli lemma since

$$\sum_{n=1}^{\infty} P\{|R_n| > c_n\} < C \sum_{n=1}^{\infty} \int_{c_n}^{\infty} \frac{m_n}{r^2} dr = C \sum_{n=1}^{\infty} \frac{m_n}{c_n} < \infty.$$

Thus, our almost sure limit follows from the last term in our partition

$$
\begin{aligned}
\frac{\sum_{i=1}^{n} a_i E R_i I(|R_i| \le c_i)}{b_n} &\sim \frac{\sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}\lg_2 i} \int_1^{c_i} \frac{m_i}{r} dr}{\lg_3 n} \\
&= \frac{\sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}\lg_2 i} m_i \lg c_i}{\lg_3 n} \\
&\sim \frac{\gamma \sum_{i=1}^{n} \frac{1}{i(\lg i)^{\beta+2}\lg_2 i}(\lg i)^{\beta}\lg i}{\lg_3 n} \\
&= \frac{\gamma \sum_{i=1}^{n} \frac{1}{i \lg i \lg_2 i}}{\lg_3 n} \\
&\rightarrow \gamma.
\end{aligned}
$$

which concludes this proof.   □

References

1. A. Adler, Exact strong laws, *Bull. Inst. Math. Acad. Sinica*, **28** (2000), 141-166.

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

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

4. A. Adler, Strong laws for ratios of order statistics from exponentials, *Bull. Inst. Math. Acad. Sinica*, **10** (2015), 101-111.

5. A. Adler, Laws of large numbers for ratios of uniform random variables, *Open Mathematics*, **13** (2015), 571-576.

6. Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, 3rd ed., Springer-Verlag, New York, 1997.

7. P. Matula, P. Kurasinski and A. Adler, Exact strong laws of large numbers for ratios of the smallest order statistics, *Statistics and Probability Letters*, **152** (2019), 69-73.

8. Y. Miao, H. Ma, S. F. Xu and A. Adler, Complete convergence for arrays of ratios of order statistics, *Open Mathematics*, **17** (2019), 439-451.

9. Y. Miao, R. Wang and A. Adler, Limit theorems for order statistics from exponential, *Statistics and Probability Letters*, **110** (2016), 51-57.

10. Y. Miao, Y. Sun, R. Wang and M. Dong, Various limit theorems for ratios from the uniform distribution, *Open Mathematics*, **14** (2016), 393-403.

11. S. F. Xu and Y. Miao, Some limit theorems for ratios of order statistics from uniform random variables, *J. Inequal. Appl.*, (2017).

12. Y. Zhang and X. Ding, Limit properties for ratios of order statistics from exponentials, *J. Inequal. Appl.*, (2017).