Journal of Statistics: Advances in Theory and Applications
Volume 13, Number 1, 2015, Pages 89-100

RESAMPLING USING ORDER STATISTICS

ANDRÉ ADLER

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

Abstract

This paper establishes strong laws for ratios and differences of order statistics from various distributions. Some of these results allow us to obtain estimators of the parameters from those distributions. This is an extension of the method of moments technique used in estimation theory and it has performed well in simulations.

1. Introduction

This paper establishes strong laws for resampling of order statistics from various distributions. From some of these strong laws, we can create an estimate of the parameters from these distributions. Let X_{nk} be independent and identically distributed random variables within each sample, where $k = 1, 2, \dots, m_n$ and $n = 1, 2, 3, \dots$. The sample size m_n, surprisingly, turns out to be quite unimportant in all of our theorems. The order statistics from this sample are $X_{n(1)} \leq X_{n(2)} \leq \cdots \leq X_{n(m_n)}$.

2010 Mathematics Subject Classification: 60F15.
Keywords and phrases: almost sure convergence, order statistics, strong laws of large numbers.
Received March 23, 2015

Next, we choose two order statistics from this sample of m_n random variables, say $X_{n(i)}$ and $X_{n(j)}$, and either take the ratio or the difference of these two. The final random variable in Sections 2 and 3 is

$$R_{nij} = \frac{X_{n(j)}}{X_{n(i)}}, \quad 1 \le i < j \le m_n,$$

while in the last section, we observe

$$S_{n12} = X_{n(2)} - X_{n(1)}.$$

We obtain strong laws for either R_{nij} or S_{n12} and in some cases we will use these theorems to estimate the parameter from the underlying distribution.

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.

2. Strong Laws from a Uniform Distribution

The underlying distribution in this section is the classic uniform distribution, $f(x) = (1/\theta_n) I(0 \le x \le \theta_n)$. It is important to note that the parameter, θ, can vary from sample to sample, hence the θ_n in our distribution. We establish strong laws for R_{nij}. Sometimes this expectation is finite, sometimes not. This depends solely on i and in all cases we obtain a strong law, even if $E(R_{nij}) = \infty$, see Adler [1]. In order to establish the density of R_{nij}, we need the joint density of $X_{n(i)}$ and $X_{n(j)}$

$$f(x_i, x_j) = \frac{m_n!}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,\theta_n^2}\left(\frac{x_i}{\theta_n}\right)^{i-1}\left(\frac{x_j}{\theta_n} - \frac{x_i}{\theta_n}\right)^{j-i-1}\left(1 - \frac{x_j}{\theta_n}\right)^{m_n - j}.$$

Next, let $w = x_i$ and $r = x_j / x_i$. The Jacobian is w and the joint density of w and r is

$$f(w,\, r) = \frac{m_n!\, w}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,\theta_n^2}\left(\frac{w}{\theta_n}\right)^{i-1}\left(\frac{rw}{\theta_n} - \frac{w}{\theta_n}\right)^{j-i-1}\left(1 - \frac{rw}{\theta_n}\right)^{m_n - j}.$$

Integrating out the dummy variable, w, and letting $u = rw/\theta_n$, the density of R_{nij} is

$$\frac{m_n!}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,\theta_n^2}\int_0^{\theta_n/r} w\left(\frac{w}{\theta_n}\right)^{i-1}\left(\frac{rw}{\theta_n} - \frac{w}{\theta_n}\right)^{j-i-1}\left(1 - \frac{rw}{\theta_n}\right)^{m_n - j} dw$$

$$= \frac{m_n!\,(r-1)^{j-i-1}}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,\theta_n^2}\int_0^{\theta_n/r} w\left(\frac{w}{\theta_n}\right)^{j-2}\left(1 - \frac{rw}{\theta_n}\right)^{m_n - j} dw$$

$$= \frac{m_n!\,(r-1)^{j-i-1}}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,\theta_n^j}\int_0^{\theta_n/r} w^{j-1}\left(1 - \frac{rw}{\theta_n}\right)^{m_n - j} dw$$

$$= \frac{m_n!\,(r-1)^{j-i-1}}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,\theta_n^j}\int_0^{1}\left(\frac{\theta_n u}{r}\right)^{j-1}(1-u)^{m_n - j}\left(\frac{\theta_n du}{r}\right)$$

$$= \frac{m_n!\,(r-1)^{j-i-1}}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,r^j}\int_0^{1} u^{j-1}(1-u)^{m_n - j} du$$

$$= \frac{m_n!\,(r-1)^{j-i-1}}{(i-1)!\,(j-i-1)!\,(m_n - j)!\,r^j}\left(\frac{(j-1)!\,(m_n - j)!}{m_n!}\right)$$

$$= \frac{(j-1)!\,(r-1)^{j-i-1}}{(i-1)!\,(j-i-1)!\,r^j},$$

which is free of both m_n and θ_n.

The expectation of R_{nij} is finite iff $i \geq 2$, which produces a classic strong law.

Theorem 1. *If $X_{n(i)}$ and $X_{n(j)}$ are two order statistics from a uniform $(0,\, \theta_n)$ distribution with $j > i$ and sample size m_n, then for all $i \geq 2$*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \frac{X_{n(j)}}{X_{n(i)}}}{N} = \frac{j-1}{i-1} \text{ almost surely.}$$

Proof. All we need to do is establish the expected value of R_{nij}. Let $u = 1/r$

$$E(R_{nij}) = \frac{(j-1)!}{(i-1)!\,(j-i-1)!}\int_1^{\infty}(r-1)^{j-i-1}r^{-j+1}dr$$

$$= \frac{(j-1)!}{(i-1)!\,(j-i-1)!}\int_0^{1}(1-u)^{j-i-1}u^{i-2}du$$

$$= \left(\frac{(j-1)!}{(i-1)!\,(j-i-1)!}\right)\left(\frac{\Gamma(j-i)\Gamma(i-1)}{\Gamma(j-1)}\right)$$

$$= \frac{j-1}{i-1},$$

which concludes this proof. □

Next, we need to explore what happens when $i = 1$. In this situation, the expectation is infinite, but barely for all j.

Theorem 2. *If $X_{n(1)}$ and $X_{n(j)}$ are two order statistics from a uniform $(0, \theta_n)$ distribution with sample size m_n, then for all $j \geq 2$ and $\alpha > -2$*

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

Proof. This is an "exact strong laws". Our random variable R_{n1j} has the following density:

$$f(r) = (j-1)(r-1)^{j-2}r^{-j}.$$

Thus,

$$xP\{R_{n1j} > x\} = x(j-1)\int_x^{\infty}(r-1)^{j-2}r^{-j}dr$$

$$= x(j-1)\sum_{k=0}^{j-2}\binom{j-2}{k}(-1)^{j-2-k}\int_x^{\infty}r^{k-j}dr$$

$$= x(j-1)\sum_{k=0}^{j-2}\frac{\binom{j-2}{k}(-1)^{j-2-k}x^{k-j+1}}{j-k-1}$$

$$= (j-1)\sum_{k=0}^{j-2}\frac{\binom{j-2}{k}(-1)^{j-2-k}x^{k-j+2}}{j-k-1}$$

$$\sim j-1.$$

Using Example 2 from Adler [1], the conclusion is immediate. □

The conclusion of both theorems are free of the sample size, m_n, and also our parameter, θ_n, which is not the case for our next distribution. If the parameter does not disappear when we let $N \to \infty$, then we can use these results to obtain an estimator, based on these strong laws. It needs to be repeated that in all cases of our uniform distribution a strong law does exist and that the distribution of R_{n1j} is similar for all $j \geq 2$. All of those distributions have infinite mean, but barely.

3. Particular Beta Distribution

In this section, the underlying distribution is $f(x) = ax^{a-1}I(0 \leq x \leq 1)$, which belongs to the Beta family of densities. The uniform is a Beta "stretched" from $(0, \theta)$ to $(0, 1)$. The final result here is quite different from those in Section 2. The sample size once again disappears from the distribution of R_{nij}, but the parameter in this family doesn't. That allows us to estimate this unknown constant, a. Also, there are three different cases here. They all depend on the magnitude of ai. Once again, we start with the joint distribution of our order statistics, $X_{n(i)}$ and $X_{n(j)}$, where $1 \leq i < j \leq m_n$

$$f(x_i, x_j) = \frac{m_n!\,a^2}{(i-1)!\,(j-i-1)!\,(m_n-j)!}(x_i)^{ai-1}(x_j^a - x_i^a)^{j-i-1}$$

$$\times (x_j)^{a-1}(1-x_j^a)^{m_n-j}.$$

Next, let $w = x_i$ and $r = x_j / x_i$. The Jacobian is w and the joint density of w and r is

$$f(w, r) = \frac{m_n!\, a^2}{(i-1)!\,(j-i-1)!\,(m_n - j)!}(w)^{aj-1}(r^a - 1)^{j-i-1}(r)^{a-1}(1-(rw)^a)^{m_n - j}.$$

Integrating out the dummy variable, w, and letting $u = rw$, then $x = u^a$, the density of R_{nij} is

$$\frac{m_n!\, a^2}{(i-1)!\,(j-i-1)!\,(m_n - j)!} r^{a-1}(r^a - 1)^{j-i-1}\int_0^{1/r} w^{aj-1}(1-(rw)^a)^{m_n-j}\,dw$$

$$= \frac{m_n!\, a^2}{(i-1)!\,(j-i-1)!\,(m_n - j)!} r^{a-aj-1}(r^a - 1)^{j-i-1}\int_0^{1} u^{aj-1}(1-u^a)^{m_n-j}\,du$$

$$= \frac{m_n!\, a}{(i-1)!\,(j-i-1)!\,(m_n - j)!} r^{a-aj-1}(r^a - 1)^{j-i-1}\int_0^{1} x^{j-1}(1-x)^{m_n-j}\,dx$$

$$= \frac{m_n!\, a}{(i-1)!\,(j-i-1)!\,(m_n - j)!} r^{a-aj-1}(r^a - 1)^{j-i-1}\left(\frac{(j-1)!\,(m_n - j)!}{m_n!}\right)$$

$$= \frac{(j-1)!\, a}{(i-1)!\,(j-i-1)!} r^{a-aj-1}(r^a - 1)^{j-i-1}.$$

Next, we observe the value of ai. If $ai < 1$, then no strong law can be established. In that situation, we will just need to increase which smaller order statistic we select, $X_{n(i)}$, so that $ai \geq 1$. If $ai > 1$, then a classic strong law exists.

Theorem 3. *Let $X_{n(i)}$ and $X_{n(j)}$ be two order statistics from our Beta distribution with $j > i$ and sample size m_n. If $ai > 1$, then*

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \frac{X_{n(j)}}{X_{n(i)}}}{N} = \frac{(j-1)!\,\Gamma(i - \frac{1}{a})}{(i-1)!\,\Gamma(j - \frac{1}{a})} \text{ almost surely.}$$

Proof. Once again, all we need to do is establish the expected value of R_{nij}. Let $u = r^{a}$, then $x = 1/u$

$$E(R_{nij}) = \frac{(j-1)!\,a}{(i-1)!\,(j-i-1)!}\int_1^{\infty} r^{a-aj}(r^{a}-1)^{j-i-1}dr$$

$$= \frac{(j-1)!}{(i-1)!\,(j-i-1)!}\int_1^{\infty} u^{\frac{1}{a}-j}(u-1)^{j-i-1}du$$

$$= \frac{(j-1)!}{(i-1)!\,(j-i-1)!}\int_0^{1} x^{i-\frac{1}{a}-1}(1-x)^{j-i-1}dx$$

$$= \left(\frac{(j-1)!}{(i-1)!\,(j-i-1)!}\right)\left(\frac{\Gamma(i-\frac{1}{a})\Gamma(j-i)}{\Gamma(j-\frac{1}{a})}\right)$$

$$= \frac{(j-1)!\,\Gamma(i-\frac{1}{a})}{(i-1)!\,\Gamma(j-\frac{1}{a})},$$

which is finite since $aj > ai > 1$, which implies that both $j - \frac{1}{a}$ and $i - \frac{1}{a}$ are positive, concluding this proof. □

Theorem 3 allows us to estimate the parameter a by resampling the ratio of various order statistics from our data. In Theorem 4, we obtain an "exact strong laws" when $ai = 1$. And even though $E(R_{nij}) = \infty$ a strong law still exists.

Theorem 4. *Let $X_{n(i)}$ and $X_{n(j)}$ be two order statistics from our Beta distribution with $j > i$ and sample size m_n. If $ai = 1$, then for all $\alpha > -2$*

$$\lim_{N\to\infty}\frac{\sum_{n=1}^{N}\frac{(\log n)^{\alpha}X_{n(j)}}{nX_{n(i)}}}{(\lg N)^{\alpha+2}} = \frac{\binom{j-1}{i}}{\alpha+2} \text{ almost surely.}$$

Proof. Once again, we quote Example 2 from Adler [1]

$$xP\{R_{nij} > x\} = \frac{ax(j-1)!}{(i-1)!\,(j-i-1)!}\int_x^\infty r^{a-aj-1}(r^a - 1)^{j-i-1}dr$$

$$= \frac{ax(j-1)!}{(i-1)!\,(j-i-1)!}\sum_{k=0}^{j-i-1}\binom{j-i-1}{k}(-1)^{j-i-1-k}\int_x^\infty r^{a-aj-1+ak}dr$$

$$= \frac{ax(j-1)!}{(i-1)!\,(j-i-1)!}\left[\sum_{k=0}^{j-i-2}\binom{j-i-1}{k}(-1)^{j-i-1-k}\int_x^\infty r^{a-aj-1+ak}dr\right.$$

$$\left. + \int_x^\infty r^{a-aj-1+a(j-i-1)}dr\right].$$

The last term is the important one. Using $ai = 1$,

$$\frac{ax(j-1)!}{(i-1)!\,(j-i-1)!}\int_x^\infty r^{a-aj-1+a(j-i-1)}dr = \frac{ax(j-1)!}{(i-1)!\,(j-i-1)!}\int_x^\infty r^{a-aj-1+aj-ai-a}dr$$

$$= \frac{aix(j-1)!}{i!\,(j-i-1)!}\int_x^\infty r^{-ai-1}dr$$

$$= \frac{x(j-1)!}{i!\,(j-i-1)!}\int_x^\infty r^{-2}dr$$

$$= \binom{j-1}{i}.$$

As for the other terms in our series

$$\frac{ax(j-1)!}{(i-1)!\,(j-i-1)!}\sum_{k=0}^{j-i-2}\binom{j-i-1}{k}(-1)^{j-i-1-k}\int_x^\infty r^{a-aj-1+ak}dr$$

$$< Cx\sum_{k=0}^{j-i-2}\int_x^\infty r^{a-aj-1+ak}dr$$

$$< Cx\sum_{k=0}^{j-i-2}x^{a(k-j+1)}$$

$$< Cx \cdot x^{-ai-a} = Cx^{-a} = o(1).$$

So, we have

$$xP\{R_{nij} > x\} \sim \binom{j-1}{i},$$

which allows us to obtain an "exact strong laws" via Example 2 from Adler [1]. □

We can use either Theorem 3 or Theorem 4 to estimate the parameter a. But it is better to use Theorem 3 since the partial sums in Theorem 4 converges very slowly. Using Theorem 3 and adjacent order statistics, $i = j - 1,$ we have

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \frac{X_{n(j)}}{X_{n(j-1)}}}{N} = \frac{j-1}{j-1-\frac{1}{a}} \text{ almost surely.}$$

Set $L_{Nj} = N^{-1}\sum_{n=1}^{N} X_{n(j)} / X_{n(j-1)}$. We have L_{Nj} converging to $(j-1)/(j-1-1/a)$. Solving for a, we have as a natural estimator of our parameter $\frac{L_{Nj}}{(j-1)(L_{Nj}-1)}$, for all $j = 2, \ldots, m_n$. Next, we sum over all j, which leads to

$$\hat{a}_N = \sum_{j=2}^{m} \frac{L_{Nj}}{(j-1)(L_{Nj}-1)},$$

as a viable estimator of a. This is the type of estimator that was proved successful in Adler and Skountrianos [2]. In that paper, we resampled over and over again and then averaged over all those samples to obtain a very powerful estimator of the parameter. Likewise, we can average over more pairs of ratios of order statistics by looking at non-adjacent pairings to develop a more complex estimator of a. Also, we can use Theorem 4 to estimate the parameter a. The limit in Theorem 4 is $\binom{j-1}{i} / (\alpha + 2),$ but with $ai = 1$ we can solve for $a = 1/i$. However, the partial sums in these theorems converge very slowly, hence it takes too many observations in order to get a good estimator. That is why, it is better to use a classic strong law instead on an "exact strong laws" when running these simulations.

4. The Normal Distribution

Now for something completely different. In this section, we use differences of our order statistics to estimate the standard deviation from a normal distribution. Here we stick to a fixed sample size of $m = 2$. Our random variables are X_{n1} and X_{n2}, which are independent and identically distributed $N(\mu_n, \sigma^2)$. The mean can vary from data set to data set, but not the variance. The statistic we pivot on will be $S_n = X_{n(2)} - X_{n(1)}$, which is the difference between the minimum and maximum of each data set and from this statistic we can estimate σ. This statistic is also known as the range. In order to obtain the distribution of S_n, we need the joint density of our two order statistics

$$f(x_1, x_2) = \frac{1}{\pi\sigma^2} \exp\left(\frac{-1}{2\sigma^2}[(x_1 - \mu_n)^2 + (x_2 - \mu_n)^2]\right).$$

Next, let $w = x_1$ and $s = x_2 - x_1$. The Jacobian is 1 and the joint density of w and s is

$$\begin{aligned} f(w, s) &= \frac{1}{\pi\sigma^2} \exp\left(\frac{-1}{2\sigma^2}[(w - \mu_n)^2 + (w + s - \mu_n)^2]\right) \\ &= \frac{1}{\pi\sigma^2} \exp\left(\frac{-1}{\sigma^2}[w^2 - 2\mu_n w + sw]\right) \exp\left(\frac{-1}{2\sigma^2}[s^2 - 2\mu_n s + 2\mu_n^2]\right). \end{aligned}$$

Integrating out the dummy variable, w and completing the square, the density of S_n is

$$f(s) = \frac{1}{\pi\sigma^2}\int_{-\infty}^{\infty}\exp\left(\frac{-1}{\sigma^2}[w^2 - 2\mu_n w + sw]\right)dw \cdot \exp\left(\frac{-1}{2\sigma^2}[s^2 - 2\mu_n s + 2\mu_n^2]\right)$$

$$= \frac{1}{\pi\sigma^2}\int_{-\infty}^{\infty}\exp\left(\frac{-1}{\sigma^2}\left[w + \frac{s - 2\mu_n}{2}\right]^2\right)dw$$

$$\cdot \exp\left(\frac{1}{\sigma^2}\left[\frac{s - 2\mu_n}{2}\right]^2\right)\exp\left(\frac{-1}{2\sigma^2}[s^2 - 2\mu_n s + 2\mu_n^2]\right)$$

$$= \frac{1}{\sqrt{2\pi}(\sigma/\sqrt{2})}\int_{-\infty}^{\infty}\exp\left(\frac{-2}{2\sigma^2}\left[w + \frac{s - 2\mu_n}{2}\right]^2\right)dw$$

$$\cdot \frac{1}{\sqrt{\pi}\sigma}\exp\left(\frac{1}{\sigma^2}\left[\frac{s - 2\mu_n}{2}\right]^2\right)\exp\left(\frac{-1}{2\sigma^2}[s^2 - 2\mu_n s + 2\mu_n^2]\right)$$

$$= \frac{1}{\sqrt{\pi}\sigma}\exp\left(\frac{-1}{2\sigma^2}\left[\frac{-(s - 2\mu_n)^2}{2} + s^2 - 2\mu_n s + 2\mu_n^2\right]\right)$$

$$= \frac{1}{\sqrt{\pi}\sigma}\exp\left(\frac{-s^2}{4\sigma^2}\right),$$

which is free of the parameter μ_n. This allows us to change the mean from sample to sample, if we choose to. The expected value of our random variable S_n is $\frac{2\sigma}{\sqrt{\pi}}$, which is easy to calculate. Thus our estimate of σ is

$$\hat{\sigma}_N = \frac{\sqrt{\pi}\sum_{n=1}^{N} S_n}{2N}.$$

And as we did in Adler and Skountrianos [2], we can now estimate the standard deviation in a normal distribution by resampling these pairs of order statistics. In that paper, we reshuffle our original data to create more and more pairs of order statistics to obtain a very powerful estimator of our parameter.

References

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

[2] A. Adler and G. Skountrianos, Resampling ratios of order statistics, Comm. Stats. 41 (2012), 1891-1894.

■