Finite Expectations

Bryan S. Todd

19th July 2023

Abstract

When it takes a long time for a random process to terminate it is tempting to wonder whether waiting is futile. We propose a statistical test to determine from a few outlier waiting times whether or not the expected waiting time is actually finite.

Introduction

Imagine that we have a process inside a black box that runs each time we press "start". It takes a random amount of time to finish and the times are statistically independent. The process is guaranteed always to terminate eventually, but it can take an extremely long time to do so. After it has been running for a seemingly endless length of time it is tempting to wonder whether it is worth waiting. Past behaviour provides some insight into what to expect. We can measure statistics such as the average and median of the previous run times, but it would be helpful to know that the expected run time is actually finite. For example, a symmetric one-dimensional random walk always eventually returns to its starting point but, paradoxically, the expected time to do so is infinite.

Let the run time expressed in suitable units (say hours) be random variable X. Let its probability density function be p whose support is the non-negative reals. Thus

$$\int_0^\infty p(x)\,dx \;=\; 1 \tag{1}$$

We do not know what p is without looking inside the box, which we are forbidden to do. However, we would at least like to know whether the expected waiting time $\mathrm{E}_p[X]$ is finite.

$$\mathrm{E}_p[X] \;=\; \int_0^\infty x\,p(x)\,dx \tag{2}$$

Now if we can find a known probability density function q that is a strict upper bound on p

$$\forall x \bullet\; p(x) \le q(x) \tag{3}$$

and has *finite* expectation $\mathrm{E}_q[X]$

$$\mathrm{E}_q[X] \;=\; \int_0^\infty x\,q(x)\,dx \tag{4}$$

then $\mathrm{E}_p[X]$ is finite too.

Lomax Distribution

A suitable distribution is the Lomax [1]. It has two positive real parameters, a and b, and probability density function q where

$$q(x) = \frac{b}{a}\left(\frac{a}{x+a}\right)^{b+1} \tag{5}$$

This has expectation

$$\mathrm{E}_q[X] = \begin{cases} \infty & \textit{if } b < 1 \\ \textit{indeterminate} & \textit{if } b = 1 \\ \dfrac{a}{b-1} & \textit{if } b > 1 \end{cases} \tag{6}$$

Thus for $b > 1$ the expectation is finite. However, the limit of the expectation as b approaches unity from below is infinite.

$$\lim_{b \uparrow 1} \mathrm{E}_q[X] = \infty \tag{7}$$

Therefore the Lomax distribution that lies at the boundary between those with infinite expectation and those with finite expectation is that for which $b = 1$. We refer to this as the $Lomax_1$ distribution; it has only one parameter a. Therefore if we can show that our sample is drawn from a probability distribution that is strictly bounded above by Lomax_1 then the expectation of that unknown distribution is finite.

An Example

Suppose that L_0 is the list of six longest waiting times in decreasing order:

$$L_0 = [10, 5, 4, 3, 2, 1] \tag{8}$$

The null hypothesis is that these data are the largest six of a random sample of some unspecified size from the Lomax_1 distribution which has probability density function

$$q(x) = \frac{a}{(x+a)^2} \tag{9}$$

and cumulative distribution function

$$Q(x) = \int_0^x q(x)dx \tag{10}$$

$$= \frac{x}{x+a} \tag{11}$$

Normalization

Our sample has been conditioned on the values being greater than some chosen threshold. Therefore, let us normalize it by subtracting and discarding every instance of the smallest value s. It follows that the remaining values have an unconditional Lomax_1 distribution. Let X be drawn from a Lomax_1 distribution with parameter a' and conditioned on $X > s$. For any positive real x, we have from the definition of conditional probability

$$
\begin{aligned}
\Pr(X - s \le x \mid X > s) &= \Pr(X \le x + s \mid X > s) && (12) \\
&= \frac{\Pr(X \le x + s, X > s)}{\Pr(X > s)} && (13) \\
&= \frac{Q(x+s) - Q(s)}{1 - Q(s)} && (14) \\
&= \left(\frac{x+s}{x+s+a'} - \frac{s}{s+a'}\right) \Big/ \left(1 - \frac{s}{s+a'}\right) && (15) \\
&= \frac{(x+s)(s+a') \; - \; s(x+s+a')}{a'(x+s+a')} && (16) \\
&= \frac{a'x}{a'(x+s+a')} && (17) \\
&= \frac{x}{x+a} && (18)
\end{aligned}
$$

where $a = s + a'$. So the normalized random variable X–s is equivalently drawn from a Lomax_1 distribution with parameter $s + a'$. Thus after normalization our list L_0 becomes $[9, 4, 3, 2, 1]$ and the null hypothesis is that this is a random sample from a Lomax_1 distribution.

Parameter Determination

Let the list of n normalized values ($n = 5$ in our example) in decreasing order be x_0, x_1, ...x_{n-1}. We assume that $n > 1$, all the values are positive, and they are not all identical. Let us determine the maximum likelihood value for parameter a. This entails minimizing the total surprise S associated with the sample.

$$
\begin{aligned}
S &= -\sum_{i=0}^{n-1} \log q(x_i) && (19) \\
&= \sum_{i=0}^{n-1} (2\log(x_i + a) - \log a) && (20)
\end{aligned}
$$

Calculating the first and second derivatives we obtain

$$\frac{dS}{da} = \frac{1}{a}\sum_{i=0}^{n-1}\frac{a-x_i}{a+x_i} \tag{21}$$

$$\frac{d^2S}{da^2} = \frac{2}{a}\left(\sum_{i=0}^{n-1}\frac{x_i}{(x_i+a)^2}\right) - \frac{1}{a}\left(\frac{dS}{da}\right) \tag{22}$$

Minimum S is easily computed because

$$a = x_0 \implies \frac{dS}{da} > 0 \tag{23}$$

$$a = x_{n-1} \implies \frac{dS}{da} < 0 \tag{24}$$

The second derivative is strictly positive at the turning point and so the solution is the unique minimum. Therefore optimum a is found rapidly and accurately by binary search for zero first derivative. For our example $a = 2.9229$ to four decimal places and calculation took just 12 iterations. Parameter a is also the median of the distribution and close to that of the sample itself.

A Test Statistic

According to the null hypothesis, the cumulative Lomax probabilities $Q(x_i)$ for our sample should be drawn from the standard uniform distribution. If instead they appear to be clustered towards the mid-point ($\frac{1}{2}$) it implies that our sample is actually drawn from a distribution with a shorter tail.

A useful measure of deviation from the mid-point is the log-odds.

$$\log\frac{Q(x)}{1-Q(x)} = \log x - \log a \tag{25}$$

The calculations for our example are

Index (i)	Value (x_i)	Probability $Q(x_i)$	$\log x_i - \log a$
0	9	0.7549	1.1247
1	4	0.5778	0.3137
2	3	0.5065	0.0261
3	2	0.4063	-0.3794
4	1	0.2549	-1.0726

Since the direction of deviation from the mid-point is immaterial we are interested only in the absolute value of the log-odds. Our test statistic is thus v, suitably shifted and scaled:

$$v = \frac{1}{\sqrt{nc}}\sum_{i=0}^{n-1}\left(|\log x_i - \log a| - \log 4\right) \tag{26}$$

where the normalizing constant is

$$c = \frac{\pi^2}{3} - (\log 4)^2 \tag{27}$$

$$\approx 1.3681 \tag{28}$$

It follows that the v-statistic has zero mean under the null hypothesis in which $Q(x)$ has a standard uniform distribution. The expectation of each term in the sum is thus

$$\int_0^1 \left(\left| \log \frac{p}{1-p} \right| - \log 4 \right) dp = 2\int_{\frac{1}{2}}^1 \log \frac{p}{1-p}\, dp - \int_0^1 \log 4\, dp \tag{29}$$

$$= 2 \log 2 - \log 4 \tag{30}$$

$$= 0 \tag{31}$$

It also follows that each term in the sum has variance c.

$$\int_0^1 \left(\left| \log \frac{p}{1-p} \right| - \log 4 \right)^2 dp = \int_0^1 \left(\log \frac{p}{1-p} \right)^2 dp - 2\log 4 \int_0^1 \left| \log \frac{p}{1-p} \right| dp + \int_0^1 (\log 4)^2 dp \tag{32}$$

$$= \frac{\pi^2}{3} - (\log 4)^2 \tag{33}$$

$$= c \tag{34}$$

Therefore after division by $\sqrt{c}$ the term has unit variance.

Critical Values

For large n, the v-statistic has a standard normal distribution (central limit theorem). Critical values are provided in Table 1. They were obtained by randomly generating 10^6 samples from the Lomax_1 distribution. For our example, $v = -1.5352$ which is statistically significant at the $p < 0.05$ level and allows us to reject the null hypothesis. We can be reasonably confident that list L_0 relates to a process with finite expected duration.

For comparison consider list L_1 in which the longest duration is increased by a factor of ten with respect to that of L_0.

$$L_1 = [100, 5, 4, 3, 2, 1] \tag{35}$$

In this case $v = -0.5706$ and we cannot reject the null hypothesis even at the $p < 0.1$ level. List L_1 is thus consistent with a process that has infinite expected duration.

More generally, suppose that our sample were actually drawn from the standard exponential distribution with probability density function

$$p(x) = e^{-x} \tag{36}$$

This has unit mean. Table 2 shows the probabilities of rejection of the null hypothesis if we draw samples of various sizes from this distribution. Compilation of the table entailed Monte

Table 1: *Lower and upper critical values of v statistic.*

n	$p \le 0.1$	$p \le 0.05$	$p \le 0.02$	$p \le 0.01$
4	-1.17, 1.33	-1.40, 1.83	-1.63, 2.44	-1.75, 2.87
5	-1.19, 1.33	-1.43, 1.81	-1.67, 2.39	-1.81, 2.80
6	-1.20, 1.33	-1.45, 1.80	-1.70, 2.37	-1.86, 2.77
7	-1.21, 1.33	-1.47, 1.80	-1.73, 2.35	-1.90, 2.74
8	-1.21, 1.32	-1.48, 1.78	-1.75, 2.33	-1.92, 2.72
9	-1.22, 1.32	-1.48, 1.78	-1.77, 2.31	-1.94, 2.70
10	-1.22, 1.32	-1.50, 1.77	-1.78, 2.31	-1.97, 2.68
12	-1.23, 1.32	-1.51, 1.76	-1.81, 2.29	-2.00, 2.65
15	-1.23, 1.32	-1.53, 1.75	-1.83, 2.26	-2.03, 2.61
25	-1.25, 1.31	-1.55, 1.73	-1.89, 2.22	-2.10, 2.55
50	-1.25, 1.30	-1.58, 1.70	-1.94, 2.17	-2.17, 2.48
100	-1.26, 1.30	-1.60, 1.69	-1.97, 2.14	-2.21, 2.43
200	-1.27, 1.29	-1.61, 1.67	-1.99, 2.12	-2.25, 2.41
400	-1.28, 1.29	-1.62, 1.67	-2.01, 2.10	-2.27, 2.38
>400	-1.28, 1.28	-1.64, 1.64	-2.05, 2.05	-2.33, 2.33

Carlo simulation with 10^4 iterations. The results show that a sample size of about 25 is needed in order to achieve equal odds of detecting that the expected duration of the process is finite at the $p < 0.05$ level of significance. A sample size of a little over 50 is sufficient to achieve equal odds at the $p < 0.01$ level of significance. Surprisingly there is still a reasonable chance (26%) of rejection at the $p < 0.05$ level of significance with just four values.

Since Lomax_1 lies on the boundary between those Lomax distributions with finite expectation and those with infinite, the test statistic also allows us to reject the null hypothesis that the sample was drawn from a distribution with finite expectation. For example consider L_2:

$$L_2 = [1000, 100, 4, 3, 2, 1] \tag{37}$$

For this sample $v = 1.8581$. This exceeds the upper critical level at the $p < 0.05$ level of significance. We can therefore conclude with reasonable confidence that this sample relates to a process with infinite expected duration.

More generally, suppose that our sample were actually from the standard log-Cauchy distribution with probability density function

$$p(x) = \frac{1}{\pi x(1 + (\log x)^2)} \tag{38}$$

This is a heavy-tailed distribution with infinite mean. Table 3 shows the probability of rejection of the null hypothesis if we draw samples of various sizes from this distribution. Here we compare the test statistic to the upper critical value and reject the null hypothesis in favour

Table 2: *Probability of rejecting null hypothesis that expectation is infinite at significance level p for sample of size n drawn from the exponential distribution.*

Size (n)	$p<0.10$	$p<0.05$	$p<0.02$	$p<0.01$
4	0.39	0.27	0.15	0.09
5	0.41	0.27	0.15	0.10
6	0.41	0.28	0.16	0.10
7	0.43	0.29	0.17	0.10
8	0.43	0.30	0.17	0.11
9	0.45	0.32	0.19	0.12
10	0.45	0.32	0.19	0.12
12	0.48	0.35	0.21	0.14
15	0.54	0.40	0.26	0.18
25	0.61	0.50	0.35	0.27
50	0.75	0.67	0.57	0.49
100	0.82	0.78	0.74	0.70
200	0.85	0.83	0.81	0.79
400	0.86	0.85	0.83	0.82

of the alternative that the distribution has infinite expectation. The results show that sample sizes as small as seven and nine achieve equal odds of detecting at the $p<0.05$ and $p<0.01$ levels of significance, respectively, that the process has infinite expected duration. There is still a reasonable chance (36%) of rejection at the $p<0.05$ level of significance with just four values.

Generalized Extreme Value Distribution

An alternative to using Lomax$_1$ is to model our data with the generalized extreme value (GEV) distribution. This describes the maxima of a set of independent and identically distributed random variables. We use the standardized variable $s=(x-\mu)/\sigma$ where μ is the location and σ the scale. The cumulative GEV distribution function is

$$F(s) = \begin{cases} e^{-e^{-s}} & \textit{if} \quad c=0 \\ e^{-(1-cs)^{1/c}} & \textit{if} \quad c\neq 0 \quad \textit{and} \quad cs<1 \\ 0 & \textit{if} \quad c<0 \quad \textit{and} \quad cs\geq 1 \\ 1 & \textit{if} \quad c>0 \quad \textit{and} \quad cs\geq 1 \end{cases} \tag{39}$$

where c is the shape parameter. A variety of methods are available for rejecting the null hypotheses that $c=0$, $c\leq 0$, and $c\geq 0$ [2]. However, we are interested here in testing whether $c\leq -1$. This is because the mean of the GEV distribution is finite precisely when $c>-1$.

Table 3: *Probability of rejecting null hypothesis that expectation is finite at significance level p for sample of size n drawn from the log-Cauchy distribution.*

Size (n)	$p<0.10$	$p<0.05$	$p<0.02$	$p<0.01$
4	0.41	0.37	0.32	0.30
5	0.46	0.42	0.37	0.35
6	0.52	0.47	0.42	0.40
7	0.55	0.51	0.46	0.42
8	0.60	0.56	0.51	0.47
9	0.63	0.59	0.54	0.51
10	0.66	0.61	0.56	0.53
12	0.71	0.67	0.62	0.59
15	0.77	0.73	0.69	0.66
25	0.88	0.86	0.83	0.80
50	0.97	0.97	0.95	0.94
100	1.00	1.00	1.00	1.00
200	1.00	1.00	1.00	1.00
400	1.00	1.00	1.00	1.00

A maximum-likelihood estimator of the three parameters (μ, σ, and c) for any given sample has been implemented in the Python programming language by the SciPy group [3]. We therefore compared the estimate of the shape parameter c with the v-statistic for discriminating between the exponential and log-Cauchy distributions. For each of three different sample sizes (5, 10, and 25) we generated 10^4 samples from the exponential distribution and an identical number from the log-Cauchy distribution. From these samples we derived the receiver operating characteristic (ROC) curve for the v-statistic. We repeated the procedure for the c-statistic. Figures 1, 2, and 3 plot the corresponding ROC curves. The v-statistic appears to be a significantly better discriminant than the c-statistic for small samples. For a sample of size 25 there appears to be no significant difference.

Tail Index Estimation

The tail index estimator due to Hill [4] is another well established method of assessing the tail of the distribution from which a sample has been drawn. The index α is estimated by

$$\alpha = \left(\left(\frac{1}{n-1} \sum_{i=0}^{n-2} \log x_i \right) - \log x_{n-1} \right)^{-1} \tag{40}$$

The mean of the underlying distribution is finite precisely when $\alpha > 1$. The ROC curves for the α-statistic were generated similarly to those of the other two statistics and are included in Figures 1, 2, and 3. For the particular discrimination task the α-statistic performs less well than the υ-statistic with all three sample sizes.

Conclusion

The apparent effectiveness of the υ-statistic and the relative ease with which it can be reliably calculated suggest that it is a useful method for determining whether a sample is from a distribution with finite or infinite expectation.

B.S. Todd 19/07/2023

Figure 1: *ROC curves ($n = 5$) for υ-statistic, c-statistic, and α-statistic showing ability to discriminate exponential distribution from log-Cauchy.*

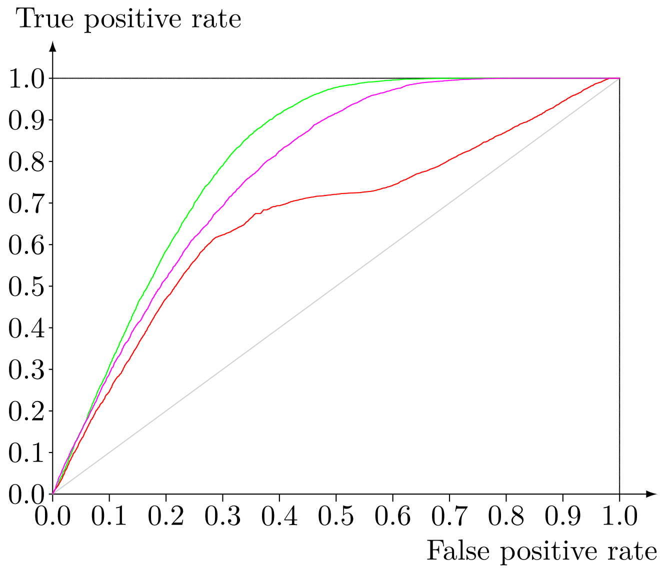


Figure 2: *ROC curves (n = 10) for v-statistic, c-statistic, and α-statistic showing ability to discriminate exponential distribution from log-Cauchy.*

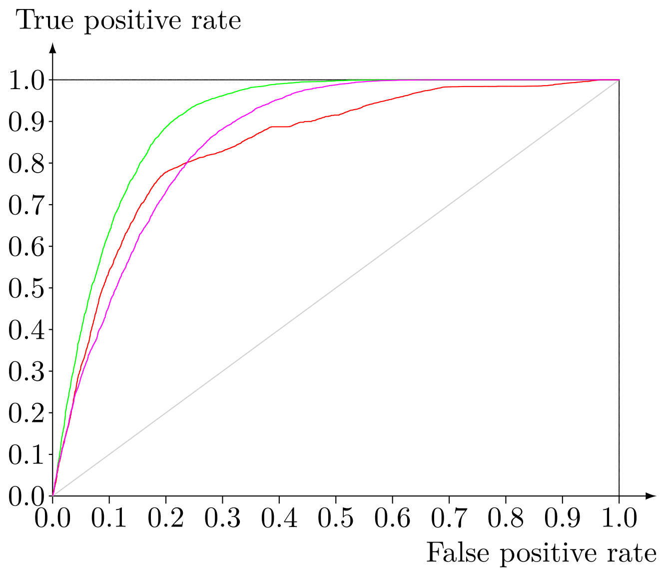
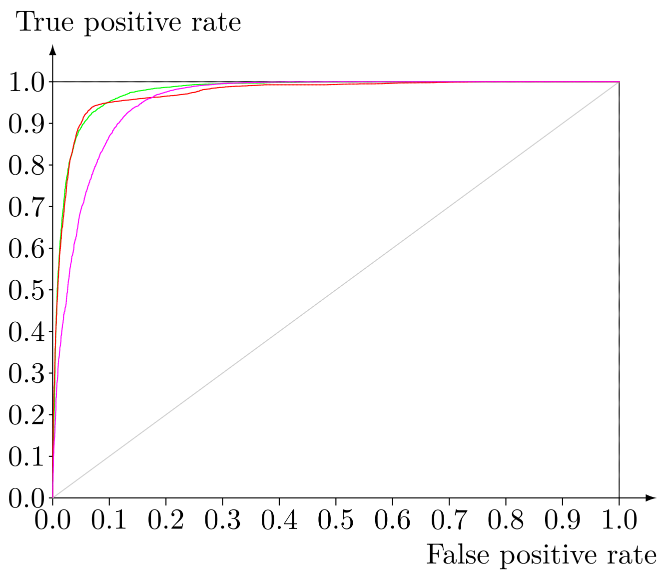


Figure 3: *ROC curves (n = 25) for v-statistic, c-statistic, and α-statistic showing ability to discriminate exponential distribution from log-Cauchy.*



References

[1] **Lomax KS.** Business failures: another example of the analysis of failure data. (1954) *Journal of the American Statistical Association* **49** *268* 847–52.
https://www.jstor.org/stable/2281544

[2] **Hosking JRM.** Testing whether the shape parameter is zero in the generalized extreme-value distribution. (1984) *Biometrika* **71** *ii* 367–74.
https://doi.org/10.1093/biomet/71.2.367

[3] **Virtanen P et al.** SciPy 1.0: fundamental algorithms for scientific computing in Python. (2020) *Nature Methods* **17** *iii* 261–72.
https://www.nature.com/articles/s41592-019-0686-2

[4] **Hill BM.** A simple general approach to inference about the tail of a distribution. (1975) *The Annals of Statistics* **3** *v* 1163–74.
https://projecteuclid.org/journals/annals-of-statistics/volume-3/issue-5/A-Simple-General-Approach-to-Inference-About-the-Tail-of/10.1214/aos/1176343247.full