Truncated Exponential Distribution with Mean Parameterization*

Kevin S. Van Horn
Adobe, Inc.

March 29, 2019

1 Problem statement

We wish to create a prior distribution for a quantity x that must lie between 0 and u, and we have in mind a desired mean μ for the distribution. As an example, the one-period autocorrelation coefficient $\rho = r^P$ for a quasiperiodic process of period $P > 0$ could be anywhere between 0 and 1, but we expect it to be close to 1 (the change in the periodic pattern from one period to the next is not large), so we want, say, $E[\rho] = 0.95$. Aside from this requirement, we would like the prior distribution to be as uninformative as possible.

Our initial thought might be to use a beta distribution for $y = x/u$,

$$\mathrm{Beta}\,(y \mid \alpha, \beta) \propto y^{\alpha-1}\,(1-y)^{\beta-1}\,,$$

which has a mean of α/β, so we would want $\mu/u = \alpha/\beta$ or $\beta = \alpha u/\mu$. But what should α be? We might consider maximizing the variance, which increases as α decreases, suggesting the improper prior obtained by setting $\alpha = 0$; but as Figure 1 shows, small values of α just push all of the probability mass to the extremes, which is not at all what we have in mind by an uninformative prior distribution.

The *entropy*

$$H(f) \triangleq -\int f(x)\log f(x)\mathrm{d}x$$

of a distribution with density function $f(x)$ measures how spread-out or diffuse it is; this suggests that the prior for x should be the maximum-entropy distribution having support $[0, u]$ and mean μ. This maximum-entropy distribution has the form

$$f\,(x \mid \theta, u) = \frac{1}{Z(\theta, u)}\exp\,(-\theta x)$$

$$Z(\theta, u) = \theta^{-1}\left(1 - e^{-\theta u}\right)$$

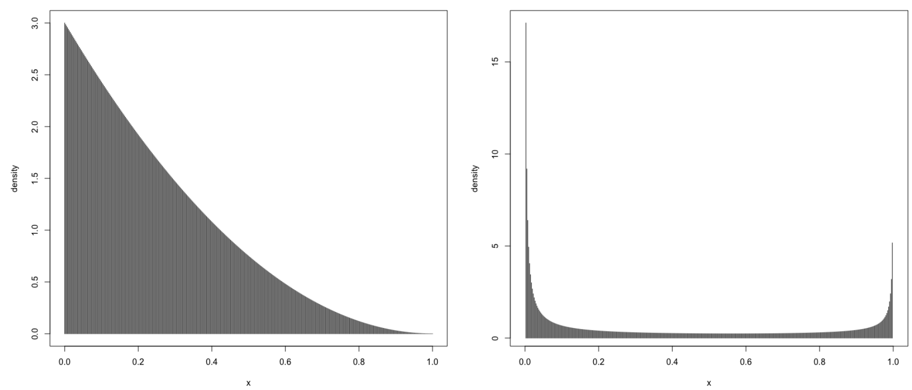

Figure 1: Beta distributions with mean $\alpha/\beta = 1/3$, and $\alpha = 1$ (left) or $\alpha = 0.1$ (right)

for some θ; that is, it is a truncated exponential distribution. Its mean is

$$E[x] = \theta^{-1}\left(1 - \frac{\theta u}{e^{\theta u} - 1}\right).$$

Figure 2 shows two examples. As expected, if we let $u \to \infty$ then the normalization constant $Z(\theta, u) \to \theta^{-1}$ and the mean $E[x] \to \theta^{-1}$, that is, we get the (untruncated) exponential distribution.

This document addresses two issues that arise in using this truncated exponential distribution in a Bayesian model:

1. How to robustly compute the log PDF.
2. How to robustly compute θ from μ using differentiable expressions.

The second issue arises when using modeling languages such as Stan or PyMC3, which use auto-differentiation to compute the gradient of the log joint PDF.

2 Robustly computing the log PDF

The probability density function is

$$\begin{aligned} f(x \mid \theta, u) &= Z(\theta, u)^{-1} \exp(-\theta x) \\ &= \frac{\theta e^{-\theta x}}{1 - e^{-\theta u}} \end{aligned}$$

which, for $\theta < 0$, it is convenient to re-express as

$$f(x \mid \theta, u) = \frac{(-\theta) e^{\theta(u-x)}}{1 - e^{\theta u}}.$$

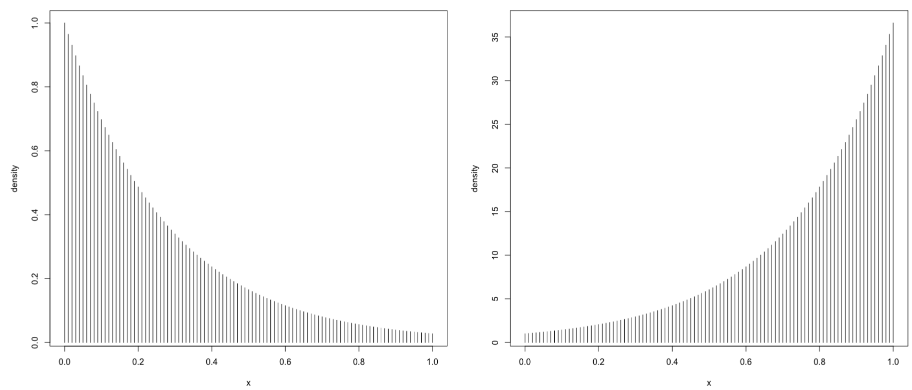

Figure 2: Truncated exponential distributions with upper bound 1. Left: $\theta = 3.6$ and mean 0.25. Right: $\theta = -3.6$ and mean 0.75.

For $\theta = 0$ we have

$$f(x \mid 0, u) = \frac{1}{u}.$$

We then get a log density of

$$\log f(x \mid \theta, u) = \begin{cases} \log\theta - \theta x - \psi(\theta u) & \text{if } \theta > 0 \\ -\log u & \text{if } \theta = 0 \\ \log(-\theta) + \theta(u - x) - \psi(-\theta u) & \text{if } \theta < 0. \end{cases}$$

$$\psi(y) \triangleq \log\left(1 - e^{-y}\right)$$

Note that

$$f(x \mid \theta, u) = f(u - x \mid -\theta, u)$$

and so we can restrict our attention to the case $\theta \geq 0$.

Robustly evaluating $\psi(y)$ presents some difficulties:

- If y is large then e^{-y} is very small, and in computing $\log(1 - e^{-y}) \approx e^{-y}$ we lose many bits of precision in the result. We can solve this by using

  $$\psi(y) = \text{log1p}\left(-e^{-y}\right)$$

  where $\text{log1p}(a)$ is a numerically stable implementation of $\log(1 + a)$ that uses

  $$\log(1 + a) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{a^n}{n}$$

  for $0 \leq a \ll 1$.

- If y is small then $e^{-y} \approx 1$, and the above approach works very badly. One option is to use
$$\psi(y) = \log\left(-\mathrm{expm1}\left(-y\right)\right)$$
where $\mathrm{expm1}(a)$ is a numerically stable implementation of $e^a - 1$ that uses
$$e^a - 1 = \sum_{n=1}^{\infty} \frac{a^n}{n!}$$
for $0 \le a \ll 1$.

The above is the method proposed by Mächler [1], using $y = \log 2$ as the cutoff for switching between the two approaches.

There is another method for handling the case when θu is especially small. Note that for $0 < y \ll 1$,

$$\begin{aligned}
\psi(y) &= \log\left(1 - e^{-y}\right) \\
&\approx \log\left(1 - (1 - y)\right) \\
&= \log y
\end{aligned}$$

and so for $0 < \theta u \ll 1$,

$$\begin{aligned}
\log f\left(x \mid \theta, u\right) &\approx \log\theta - \theta x - \log\left(\theta u\right) \\
&= -\log u - \theta x
\end{aligned}$$

with the potentially large (in absolute value) terms $\log\theta$ cancelling out. So define
$$\zeta\left(y\right) \triangleq \psi(y) - \log y$$
and use

$$\begin{aligned}
\log f\left(x \mid \theta, u\right) &= \log\theta - \theta x - \psi\left(\theta u\right) + \log\left(\theta u\right) - \log\left(\theta u\right) \\
&= \log\theta - \theta x - \zeta\left(\theta u\right) - \log\left(\theta u\right) \\
&= -\log u - \theta x - \zeta\left(\theta u\right).
\end{aligned}$$

Using Mathematica, we find

$$\begin{aligned}
\zeta\left(y\right) = &-\frac{1}{2}y + \frac{1}{24}y^2 - \frac{1}{2880}y^4 + \frac{1}{181440}y^6 - \frac{1}{9676800}y^8 \\
&+ \frac{1}{479001600}y^{10} - \frac{691}{15692092416000}y^{12} + \frac{1}{1046139494400}y^{14} - \cdots
\end{aligned}$$

The relative error for various values of y and series truncation points is given below, where k is the maximum power of y used, and ϵ is the relative error:

y	k	ϵ
0.1	8	-4.2×10^{-18}
0.25	10	2.1×10^{-17}
0.4	12	-1.3×10^{-17}

In summary, we use the following to compute the log pdf:

$$\log f(x \mid \theta, u) = \begin{cases} \log f(u - x \mid -\theta, u) & \text{if } \theta < 0 \\ -\log u - \theta x - \zeta(\theta u) & \text{if } 0 \le \theta u \le 0.1 \\ \log \theta - \theta x - \psi(\theta u) & \text{if } \theta u > 0.1 \end{cases}$$

$$\begin{aligned} \zeta(y) &= \psi(y) - \log y \\ &\approx -\frac{1}{2}y + \frac{1}{24}y^2 - \frac{1}{2880}y^4 + \frac{1}{181440}y^6 - \frac{1}{9676800}y^8 \end{aligned}$$

$$\psi(y) = \begin{cases} \log(-\operatorname{expm1}(-y)) & \text{if } y \le \log 2 \\ \operatorname{log1p}(-e^{-y}) & \text{if } y > \log 2. \end{cases}$$

3 Mean for rate parameterization

The mean $\mu = E[x \mid \theta, u]$ is

$$\begin{aligned} \mu &= Z(\theta, u)^{-1} \int_0^u x \exp(-\theta x)\, \mathrm{d}x \\ &= Z(\theta, u)^{-1} \left[-\theta^{-2}\left(e^{-\theta x}(1 + \theta x)\right)\right]_0^u \qquad \{\text{using wolframalpha}\} \\ &= (\theta Z(\theta, u))^{-1} \left(-\theta^{-1}\right)\left(e^{-\theta u}(1 + \theta u) - 1\right) \\ &= \left(1 - e^{-\theta u}\right)^{-1}\left(\theta^{-1} - e^{-\theta u}\left(\theta^{-1} + u\right)\right) \\ &= \left(1 - e^{-\theta u}\right)^{-1}\left(\theta^{-1}\left(1 - e^{-\theta u}\right) - u e^{-\theta u}\right) \\ &= \theta^{-1} - \frac{u e^{-\theta u}}{1 - e^{-\theta u}} \\ &= \theta^{-1} - \frac{u}{e^{\theta u} - 1} \\ &= \theta^{-1}\left(1 - \frac{\theta u}{e^{\theta u} - 1}\right) \end{aligned}$$

All of this holds whether θ is positive or negative. Defining

$$\begin{aligned} \mu_r &= \mu / u \\ \theta_r &= \theta u \end{aligned}$$

we can rewrite the above as

$$\begin{aligned} \mu_r &= g(\theta_r) \\ g(y) &\triangleq \frac{1}{y}\left(1 - \frac{y}{e^y - 1}\right). \end{aligned}$$

Then

$$\begin{aligned}
\lim_{y\to 0} g(y) &= \lim_{y\to 0} y^{-1}\left(1-\frac{y}{e^y-1}\right) \\
&= \lim_{y\to 0} y^{-1}\left(\frac{e^y-1-y}{e^y-1}\right) \\
&= \lim_{y\to 0} y^{-1}\left(\frac{\sum_{n=2}^{\infty} y^n/n!}{\sum_{n=1}^{\infty} y^n/n!}\right) \\
&= \lim_{y\to 0} y^{-1}\frac{y^2/2!}{y/1!} \\
&= \frac{1}{2}
\end{aligned}$$

and so $\mu_r = 1/2$ when $\theta_r = 0$. Furthermore,

$$\begin{aligned}
g\left(-y\right) &= -y^{-1} - \frac{1}{e^{-y}-1} \\
&= -y^{-1} - \frac{e^y}{1-e^y} \\
&= -y^{-1} + \frac{e^y}{e^y-1} \\
&= -y^{-1} + \frac{(e^y-1)+1}{e^y-1} \\
&= -y^{-1} + 1 + \frac{1}{e^y-1} \\
&= 1 - \left(y^{-1} - \frac{1}{e^y-1}\right) \\
&= 1 - g\left(y\right)
\end{aligned}$$

and so $\mu_r\left(-\theta_r\right) = 1 - \mu_r\left(\theta_r\right)$. Finally, for small values of y it may be better to compute $g(y)$ using

$$g(y) = y^{-1}\left(\frac{e^y-1-y}{e^y-1}\right) = y^{-1}\left(\frac{\sum_{n=2}^{\infty} y^n/n!}{\sum_{n=1}^{\infty} y^n/n!}\right).$$

In practice we need to obtain θ from μ, which we do by obtaining θ_r from μ_r, which in turn requires inverting g:

$$\begin{aligned}
\theta_r &= g^{-1}\left(\mu_r\right) \\
\theta &= \frac{1}{u} g^{-1}\left(\frac{\mu}{u}\right).
\end{aligned}$$

[TODO: Prove that g is strictly decreasing, to ensure that it has an inverse.] Furthermore, computation of the gradient via autodifferentiation requires that we express $g^{-1}\left(x\right)$ as a (piecewise) differentiable expression.

We compute $g^{-1}(x)$ as follows:

1. If $x < g(22) \approx 4.545 \times 10^{-2}$, then $g^{-1}(x) = x^{-1}$ to within a relative error of 10^{-6}.

    - The relative error was verified empirically on a grid of one million x values running from $10^{-6} \cdot g(22)$ to $g(22)$, spaced uniformly on a logarithmic scale.

2. If $g(22) \leq x \leq g(4.5) \approx 0.2110$, then $g^{-1}(x) = x^{-1} h_1(g(4.5) - x)$ to within a relative error of 10^{-6}, where

    - $h_1(x)$ is a degree-9 polynomial.
    - The coefficients were found by taking a grid of y values from 4.5 to 22 with a spacing of 10^{-6}, computing $x = g(y)$ for each of them, then doing a least-squares fit of xy to $g(4.5) - x$.

3. If $g(4.5) < x \leq 0.5$, then $g^{-1}(x) = h_2\,(0.5 - x)$ to within a relative error of 10^{-6}, where

    - $h_2(x)$ is a degree-9 polynomial.
    - The coefficients were found by taking a grid of y values from 0 to 4.5 with a spacing of 10^{-6}, computing $x = g(y)$ for each of them, then doing a least-squares fit of y to $0.5 - x$.

In cases 2 and 3 the relative error was verified empirically on a grid of values from the smallest x to the largest x, using 10 times as many points as were used for the least-squares fit.ß

References

[1] Martin Mächler, 2019. “Accurately Computing $\log(1 - \exp(-|a|))$ Assessed by the Rmpfr package.” A vignette of the R package Rmpfr, “R MPFR—Multiple Precision Floating-Point Reliable.” https://cran.R-project.org/package=Rmpfr.