Gamma LRT

This document describes the steps to obtain the tail function for the importance sampling weights in Gamma distribution example. We use an Exponential distribution as the proposal dist with pdf $g(x)$, and our target distribution is Gamma with pdf $f(x)$.

Steps are as follows (I think!) This looks really good Payman. I just added a couple notes in blue.

Step 1. Set the target distribution, $f(x)$ as a Gamma distribution with α as shape and β as scale:

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{\beta x} \tag{1}$$

Step 2. Set the proposal distribution, $g(x)$, as an Exponential dist with scale parameter of one:

$$g(x) = e^{-x} \tag{2}$$

More generally, we can use any scale parameter for the exponential. For the purposes of this discussion, we just need the scale parameter to be fixed. I think it's pretty reasonable to keep $\lambda = 1$ for now to avoid cluttering our formulas, although eventually we'll need to do the general case.

Step 3. The pdf of weights is $w(x; \alpha, \beta) = \frac{f(x;\alpha,\beta)}{g(x)}$. This actually isn't quite the pdf of the weights. Instead, the Xs are iid from g, and the weights are obtained by evaluating w at the simulated Xs. The distribution on the weights is then induced by the distribution on X and the functional relationship between X and W. If the words "pushforward measure" help clarify things, that's exactly what's happening here. If not, W is a transformation of X. In principle, we can get the pdf of W by computing the Jacobian of the (inverse) transformation (i.e. the derivative of the inverse of the likelihood ratio function). Personally, I prefer to work directly with the CDF or survival function; for one, I think it's easier, and for two, we need the survival function for step 4 anyway.

Step 4. Use the pdf in step 3 and compute its tail function as described in Pickands paper:

$$P(x) = \frac{F(X \geq u + x)}{F(X \geq u)} \tag{3}$$

where $F(.)$ is the CDF of weights as defined in step 3.

Step 5. Find Q_3, and median as described in the paper. Compute the c and a.

1 William's Derivation

I found a few mistakes. See the git log for changes. I still think my new answer is wrong, but at least it's a bit more reasonable. I'll make another pass at some point and see if I notice more mistakes.

I'm going to mostly stick to the above notation, but I would like to use the general exponential distribution until I find myself needing to set $\lambda = 1$. I will also suppress dependence on parameters in the various densities to avoid extra typing. Thus, $g(x) = \exp(-x/\lambda)/\lambda$.

At a high-level, my plan is to compute the survival function of the weights, as well as their density and the derivative of their density. I can then use Equation (1.1.30) on page 15 of de Haan and Ferreira (2006) to (hopefully) compute the tail index. There are enough steps here that it's going to take a few days' work. One step at a time.

To start, we note that the weight, W, can be written as a random variable in the following way.

$$W(X) = \frac{f(X)}{g(X)} \tag{4}$$

$$= \frac{\lambda}{\beta^\alpha \Gamma(\alpha)} X^{\alpha-1} \exp\left[-X\left(\frac{1}{\beta} + \frac{1}{\lambda}\right)\right] \tag{5}$$

$$=: cX^{\alpha-1} \exp(-X/\gamma) \tag{6}$$

where $c = \lambda/\beta^\alpha \Gamma(\alpha)$, and $\gamma = (1/\beta + 1/\lambda)^{-1}$. I doubt that it will every be relevant, but I think it's cute that γ is a harmonic mean.

We can now work out the survival function of W. We write $\mathscr{W}$ for the inconveniently named Lambert-W function, which is the inverse function of $x \mapsto xe^x$. I.e. $\mathscr{W}$ satisfies $\mathscr{W}(x)\exp[\mathscr{W}(x)] = x$.

$$\mathbb{P}(W > w) = \mathbb{P}\left(X^{\alpha-1}e^{-X/\gamma} > \frac{w}{c}\right) \tag{7}$$

$$= \mathbb{P}\left(X > -(\alpha-1)\gamma \cdot \mathscr{W}\left[-\frac{(w/c)^{1/(\alpha-1)}}{(\alpha-1)\gamma}\right]\right) \tag{8}$$

$$= \exp\left(\frac{(\alpha-1)\gamma}{\lambda} \cdot \mathscr{W}\left[-\frac{(w/c)^{1/(\alpha-1)}}{(\alpha-1)\gamma}\right]\right) \tag{9}$$

where we pass from (7) to (8) with the help of Wolfram Alpha.

1.1 PDF

Next, we need the PDF of W, and later its derivative. I'm going to find it easier to do this in stages. Let $\bar{\varphi}(w) = \mathbb{P}(W > w)$ and η be the argument to $\mathscr{W}$ in

(9). The (negative) PDF of w is obtained by differentiating $\bar\varphi$. Note that the derivative of $\mathscr{W}$ can be obtained implicitly from its[1] defining equation

$$\nabla_w\bar\varphi(w) = \frac{(\alpha-1)\gamma}{\lambda}\bar\varphi(w)\nabla_w\mathscr{W}(\eta) \tag{10}$$

$$\nabla_w\mathscr{W}(\eta) = \frac{\mathscr{W}(\eta)}{\eta[1+\mathscr{W}(\eta)]}\nabla_w\eta \tag{11}$$

$$\nabla_w\eta = -\frac{1}{c\gamma(\alpha-1)^2}\left(\frac{w}{c}\right)^{\frac{1}{\alpha-1}-1} = \frac{\eta}{w(\alpha-1)} \tag{12}$$

Putting this all together, and after a surprising amount of cancellation, I get

$$f_W(w) := -\nabla_w\bar\varphi(w) \tag{13}$$

$$= -\bar\varphi(w)\frac{\gamma}{\lambda w}\cdot\frac{\mathscr{W}(\eta)}{1+\mathscr{W}(\eta)} \tag{14}$$

1.2 Derivative of 1 over Hazard Rate

We return now to Theorem 1.1.8 of de Haan and Ferreira (2006), specifically expression $(1.1.29)$[2]. We first compute the survival function of W over its density, wherein we get some nice cancellation.

$$\frac{\bar\varphi(w)}{f_W(w)} = -\frac{\lambda w}{\gamma}\cdot\frac{1+\mathscr{W}(\eta)}{\mathscr{W}(\eta)} \tag{15}$$

where $\mathscr{W}$ and η are defined in Section 1.1. We now need to differentiate the expression in 15. Fortunately, we can recycle identities (8) and (9). We first present some identities which help with simplification, then proceed with com-

---

[1] Usually, I try not to use "it" to refer to anything in academic writing (Tom drilled this into me during my master's thesis revisions). Here however, I think it's a pretty good fit and doing anything else would be very wordy.

[2] Note that, in survival analysis, the hazard rate of a distribution is defined as the density over the survival function. The quantity being differentiated in (1.1.29) before taking the limit is thus the reciprocal of the hazard rate of W

puting the derivative.

$$\nabla_w \mathscr{W}(\eta) = \partial \mathscr{W}(\eta) \cdot \partial \eta \tag{16}$$

$$= \frac{1}{w(\alpha - 1)} \cdot \frac{\mathscr{W}(\eta)}{1 + \mathscr{W}(\eta)} \tag{17}$$

$$\nabla_w \frac{1 + \mathscr{W}(\eta)}{\mathscr{W}(\eta)} = \nabla_w \frac{1}{\mathscr{W}(\eta)} \tag{18}$$

$$= \frac{-\nabla_w \mathscr{W}(\eta)}{\mathscr{W}(\eta)^2} \tag{19}$$

$$= -\frac{1}{w(\alpha - 1)\mathscr{W}(\eta)(1 + \mathscr{W}(\eta))} \tag{20}$$

$$\nabla_w \frac{\bar{\varphi}(w)}{f_W(w)} = -\frac{\lambda}{\gamma} \left[\frac{1 + \mathscr{W}(\eta)}{\mathscr{W}(\eta)} + w \nabla_w \frac{1 + \mathscr{W}(\eta)}{\mathscr{W}(\eta)} \right] \tag{21}$$

$$= \frac{\lambda}{\gamma} \left[\frac{1}{(\alpha - 1)\mathscr{W}(\eta)(1 + \mathscr{W}(\eta))} - \frac{1 + \mathscr{W}(\eta)}{\mathscr{W}(\eta)} \right] \tag{22}$$

All that remains is to compute the limit of (22) as $w \uparrow \infty$. Note that $\eta \to -\infty$ as $w \to \infty$, so $\mathscr{W}(\eta) \to -\infty$ as well. Re-write (22) as $(\lambda/\gamma)[A-B]$. We now explore the asymptotic behaviour of A and B. This is actually pretty straightforward: $A \to 0$ and $B \to 1$. Thus, the tail index, $\nabla_w[\bar{\varphi}(w)/f_W(w)]$, goes to $-\lambda/\gamma$ as $w \uparrow \infty$. Expanding this ratio, we get

$$-\frac{\lambda}{\gamma} = -\frac{\lambda}{(1/\beta + 1/\lambda)^{-1}} \tag{23}$$

$$= -\left(1 + \frac{\lambda}{\beta}\right) \tag{24}$$

1.3 Discussion

This final conclusion, while pleasantly simple, cannot possibly be correct. Surely, the tail index must depend on both parameters of the target distribution. Frankly, an incorrect answer isn't terribly surprising here. It's a long derivation with lots of room for mistakes. If I have time between now and our next meeting, I'll spend some time reviewing my work. If not, I would like to go over it again with you Payman, both to explain my reasoning and to double-check my work.

2 Second Attempt

There are a few mistakes in the above analysis. First, a typo. In the definition of W, the exponent should contain the difference between $1/\beta$ and $1/\lambda$ rather than their sum (see Equation (5)). Fortunately, we just set this sum equal to $1/\gamma$ in the rest of our analysis, so we simply re-refine $\gamma = (\beta^{-1} - \lambda^{-1})^{-1}$ and the rest of our calculations hold (insofar as the rest of the calculations are themselves correct).
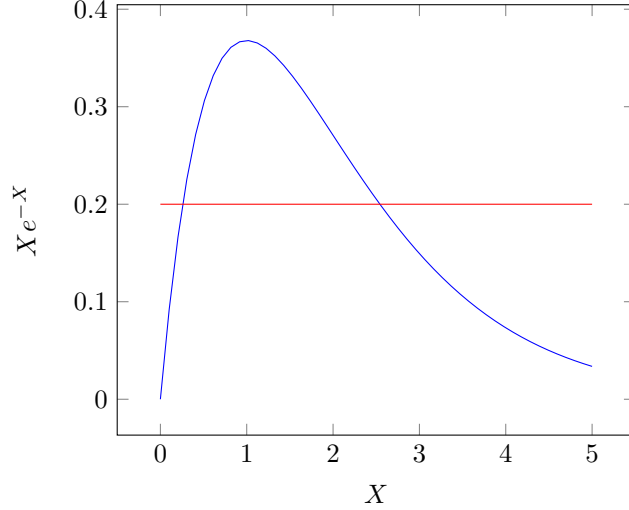
Figure 1: Graph of $y = xe^{-x}$, with a horizontal line at $y = 0.2$.

The second error is more egregious. In moving from equation (7) to (8), we used the Lambert-W function, $\mathscr{W}$, to invert a function of the form $x \mapsto axe^{-bx}$, with $a, b > 0$. However, in part of its domain, $\mathscr{W}$ is multi-valued. I find it easiest to understand this behaviour geometrically. Consider the simplest function of this form, xe^{-x}. Figure 1 shows this function, as well as a horizontal line at $Y = 0.2$. Note that this line intersects the curve at two points, so there are two solutions to the equation $xe^{-x} = 0.2$. In fact, this will be true for any positive value of y below the maximum. We will prove this fact later, but for now take it as motivation for investigating the multi-valuedness of $\mathscr{W}$.

The ultimate reason that we care about the multi-valuedness of $\mathscr{W}$ is that the set $\{W(X) > w\}$ corresponds to an interval in X. Thus, the survival function of W is computed as a difference in the survival function of X at two points. We now proceed with fixing our derivation of the tail index for W.

To start, note that W is defined as the likelihood ration between the $\text{Gamma}(\alpha, \beta)$ and $\text{Exp}(\lambda)$ distributions, evaluated at a random variable with the matching exponential distribution.

$$W(X) = \frac{\lambda}{\beta^\alpha \Gamma(\alpha)} X^{\alpha-1} \exp\left[-X\left(\frac{1}{\beta} - \frac{1}{\lambda}\right)\right] \tag{25}$$

$$=: cX^{\alpha-1}\exp(-X/\gamma) \tag{26}$$

Next, we need the survival function of W.

$$\mathbb{P}(W > w) = \mathbb{P}\left\{X^{\alpha-1}\exp(-X/\gamma) > \frac{w}{c}\right\} \tag{27}$$

We now present some properties of the function $h(x) = x^a e^{-bx}$, which we will then use to simplify Expression (27).

5

2.1 On the function $h(x) = x^a \cdot e^{-bx}$

To start, $\lim_{x\to\infty} h(x) = 0$ and $h(0) = 0$ (we will not concern ourselves with negative arguments of h). We can maximize h by differentiating and solving for a root. This is easier if we instead work with $\log h$.

$$\nabla \log h(x) = \nabla \left[a \log(x) - bx \right] \tag{28}$$

$$= \frac{a}{x} - b \tag{29}$$

This expression has a unique root at $x_* = a/b$. The second derivative test trivially confirms that x_* is a maximizer. Evaluating $h(x_*) = (a/eb)^a =: y_*$ gives us an upper bound for h.

Next, we attempt to invert h. We will proceed formally, ignoring any problems which can arise from raising negative numbers to arbitrary powers, and verify that the inverse we end up with is, well-defined. Later, we will verify that we have, in fact, found all inverses of h for the parameter values we consider. Our derivation of the inverse will ultimately involve the Lambert-W function; the rest of the details revolve around setting up its application.

$$x^a e^{-bx} = y \tag{30}$$

$$x e^{-bx/a} = y^{1/a} \tag{31}$$

$$-\frac{bx}{a} e^{-bx/a} = -\frac{b}{a} \cdot y^{1/a} \tag{32}$$

$$-\frac{bx}{a} = \mathscr{W}\left(-\frac{b}{a} \cdot y^{1/a} \right) \tag{33}$$

$$x = -\frac{a}{b} \cdot \mathscr{W}\left(-\frac{b}{a} \cdot y^{1/a} \right) = h^{-1}(y) \tag{34}$$

A few notes. First, our final expression for h^{-1} is perfectly well-defined. The only possible problem is that we could have lost one or more roots along the way. Second, in moving from (32) to (33), we have applied the $\mathscr{W}$ function without checking whether it is well-defined. Doing so now, we note that $\mathscr{W}(x)$ is not defined for $x < -e^{-1}$, has two real values for $-e^{-1} \leq x < 0$ (although these roots coincide for $x = -e^{-1}$), and one real value for $x \geq 0$. In the context of our h function, these conditions correspond to $y > (a/eb)^a = y_*$, $y_* \geq y > 0$, and $y \leq 0$ respectively. Note that solutions for $y < 0$ are negative in x, so are disregarded. These algebraic possibilities correspond exactly with the geometry in Figure 1. Above y_* or below 0 there are no inverses, below y_* but above 0 there are two inverses, and at 0 or y_* there are unique inverses.

While the above analysis is nice, it does still remain to check that we haven't missed any inverses of h. To this end, observe that the function h is strictly monotonic both below and above x_* (increasing and decreasing respectively). This can be checked easily by inspection of Equation (29). Thus, there is at most one solution of the equation $h(x) = c$ on each of the intervals $(-\infty, x_*]$ and $[x_*, \infty)$. We have already identified all such solutions.

6

In summary, we have completely characterized the inverses of the function $h(x) = x^a e^{-bx}$ for $x \geq 0$ and $a, b > 0$. These are given by Equation (34), possibly with two distinct values of $\mathscr{W}$. In practice, when writing code to evaluate $h^{-1}(y)$, we can check whether $y > y_*$ or $y < 0$ (no solution), $0 < y < y_*$ (two solutions), or $y \in \{0, y_*\}$ (one solution), where $y_* = (a/eb)^a$.

It would be nice to use proper names for the two branches of $\mathscr{W}$. I think one is called the "principal" branch, but I will need to confirm. Also do this in the following section.

2.2 Computing the Tail Index of W

We now return to the problem of computing $\mathbb{P}(W > w)$. Recalling Equation (27) and applying our inverse formula given in Equation 34, this probability is

$$\mathbb{P}\left\{X^{\alpha-1}\exp(-X/\gamma) > \frac{w}{c}\right\} \tag{35}$$

$$= \mathbb{P}\left\{-\gamma(\alpha-1)\mathscr{W}_u\left[-\frac{1}{\gamma(\alpha-1)}\left(\frac{w}{c}\right)^{1/(\alpha-1)}\right] < X \right. \tag{36}$$

$$\left. < -\gamma(\alpha-1)\mathscr{W}_l\left[-\frac{1}{\gamma(\alpha-1)}\left(\frac{w}{c}\right)^{1/(\alpha-1)}\right]\right\} \tag{37}$$

$$=: \mathbb{P}\left\{X_l(w) < X < X_u(w)\right\} \tag{38}$$

where $\mathscr{W}_l \leq \mathscr{W}_u$ are two (possibly equal) real-valued branches of the Lambert-W function. Finally, since $X \sim \text{Exp}(\lambda)$, the survival function of W is $\mathbb{P}(W > w) = \exp(-X_u(w)/\gamma) - \exp(-X_l(w)/\gamma)$.

Having established the survival function of W, we now need to compute its density (i.e. the negative derivative of the survival function; this is going to get messy). We then need to take their ratio, differentiate, and take the limit as $w \to w_*$. At some point, we may need to start working with numerical derivatives, but I would like to do as much analytically as is reasonably possible.

References

de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction.* Springer.