On the Generalized Pareto Distribution Behaviour of Distributions' Tails

March 13, 2024

This work is based on the paper by Pickands III (1975). It is cited heavily by Vehtari et al. (2022).

The first part of the paper is surprisingly readable. I've gotten as far as Section 3, skipping large chunks as necessary. At some point, it will be helpful to summarise what is said there. For now though, I just want to document some calculations in hopes that doing so reduces my error rate.

1 Exponential Tail

This is just a proof-of-concept example to illustrate how the calculations are done. Let $X \sim \text{Exp}(\theta)$, with survival function $\bar{F}(x) = \exp(-\theta x)$. Since $X \geq 0$, $\bar{F}$ is a tail function. Write $P(x) = \bar{F}(x)$ for the tail function of X.

In order to identify the GPD associated with P, we require the value of P^{-1} evaluated at $1/2$ and $1/4$. It is straightforward to show that $P^{-1}(r) = -\log(r)/\theta$. Thus, the relevant quantiles are $x_1 := P^{-1}(1/2) = \log(2)/\theta$ and $x_2 := P^{-1}(1/4) = \log(4)/\theta$.

We now recover the GPD parameters under the usual parameterization using equalities (3.2) and (3.3) from Pickands III (1975). First, note that $x_2 - x_1 = \log(2)/\theta$, so $(x_2 - x_1)/x_1 = 1$ and $c = 0$. Although we only really care about c, we remark here that $a = 1/\theta$.

Having established that the GPD associated with the tail of X has $c = 0$, we recover simply that the tails of the exponential distribution behave like the exponential distribution, which is a special case of the GPD when $c = 0$.

2 Normal Likelihood Ratio

We now move on to a less trivial, albeit still not terribly exciting, example. Let $\mathbb{F} \sim \text{N}(0, 1)$ and $\mathbb{G} \sim \text{N}(\mu, 1)$. Let $W = d\mathbb{F}/d\mathbb{G}$ be the importance ratio corresponding to using $\mathbb{G}$ to approximate sampling from $\mathbb{F}$. Note that the distribution of W is induced by its dependence on X, and hence on the Monte Carlo distribution of X, $\mathbb{G}$.

The tail function of W, $P(w)$, is obtained as follows. Recall that the likelihood ratio $d\mathbb{F}/d\mathbb{G}(x) = \exp(x\mu - \mu^2/2)$.

$$P(w) = p(W > w) \tag{1}$$

$$= p\left(\frac{d\mathbb{F}}{d\mathbb{G}}(X) > w\right) \tag{2}$$

$$= p\left(\exp(X\mu - \frac{\mu^2}{2}) > w\right) \tag{3}$$

$$= p\left(X\mu > \frac{\mu^2}{2} + \log(w)\right) \tag{4}$$

I missed this the first time, but we can't proceed without knowing the sign of μ. To make my life easier for now, we'll handle these cases separately. They're not really all that different, but this should help me keep the formulas straight.

2.1   Case I: $\mu > 0$

Proceeding now from (4), we get

$$P(w) = p\left(X\mu > \frac{\mu^2}{2} + \log(w)\right) \tag{5}$$

$$= p\left(X > \frac{\mu}{2} + \frac{1}{\mu}\log(w)\right) \tag{6}$$

$$= 1 - \Phi\left(\frac{\mu}{2} + \frac{1}{\mu}\log(w)\right) \tag{7}$$

This is about as much simplification as I've been able to get here. I tried doing something with the error function, but it didn't make things easier for me.

Next, we need the quantiles of W. I.e. We need to solve the equations $P(w_1) = 1/2$ and $P(w_2) = 1/4$. Fortunately, the first one is easy, since $\Phi^{-1}(1/2) = 0$. Plugging this in and simplifying, we get $w_1 = \exp(-\mu^2/2)$. Getting w_2 is a bit more involved, but not too bad. We eventually end up with $w_2 = \exp[-\mu^2/2 + \mu\Phi^{-1}(3/4)] =: w_1\eta$.

Before directly applying the formula for c, we observe that $(w_2 - w_1)/w_1 = \eta - 1 = \exp[\mu\Phi^{-1}(3/4)] - 1$. Taking the log of this expression makes me feel like I should be using a Taylor series. Unfortunately, the nice series is for $\log(1 + x)$ not for $\log(-1 + x)$, but nevertheless we can expand the latter function around $x = 2$ to get $c \approx (\log(2))^{-1}(\mu\Phi^{-1}(3/4) - 2)$. Not the most informative expression I've ever derived, but at least it tells us the sign of c. We get ordinary Pareto distribution behaviour when $\mu > 2/\Phi^{-1}(3/4) \approx 2.97$ (i.e. $c > 0$), and truncated Pareto behaviour when $\mu < 2/\Phi^{-1}(3/4)$ ($c < 0$).

2.2   Case II: $\mu < 0$

I'm not going to go through this tonight, but I expect there to be a lot of symmetry with Case I.

2.3 Discussion

I think the take-away here is that, in the limit, we get very well-behaved tails for the normal likelihood ratio, provided that the proposal mean is close enough to the target mean. As for Pareto smoothing, it does suggest that we should be getting negative values for k in the normal setting. Since we aren't, maybe we aren't yet in the asymptotic regime. I'm going to need to think more carefully about equation (2.1) of Pickands III (1975). Actually, a possibility I hadn't yet considered is that these calculations might not even be relevant for W. It's possible that W isn't in the "basin of attraction" of a GPD limiting tail distribution. I'm also going to look through the paper again for sufficient conditions which guarantee the GPD type limit behaviour that we've assumed above.

References

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1).

Vehtari, A., Simpson, D., Gelman, A., Yao, Y., and Gabry, J. (2022). Pareto smoothed importance sampling. *arXiv*.