

# On the Generalized Pareto Distribution Behaviour of Distributions' Tails

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.

Much of the following discussion is wrong. More precisely, it's based on an incorrect interpretation of the "tail function" defined in Pickands III (1975). The correct interpretation is as the conditional distribution of  $X$  given that  $X > u$ , rather than just the survival function of  $X$ . See Section 3 for a more correct analysis. Eventually, I will probably delete the old stuff.

## 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  $d\mathbb{F}/d\mathbb{G}(x) = \exp(x\mu - \mu^2/2)$ .

$$P(w) = p(W > w) \quad (1)$$

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

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

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

I missed this the first time, but we can't proceed without knowing the sign of  $\mu$ . 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

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

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

$$= p\left(\frac{X - \mu}{1} > -\frac{\mu}{2} + \frac{1}{\mu} \log(w)\right) \quad (7)$$

$$= 1 - \Phi\left(-\frac{\mu}{2} + \frac{1}{\mu} \log(w)\right) \quad (8)$$

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 easy to see 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  $\mu < 2/\Phi^{-1}(3/4)$  ( $c < 0$ ). In particular, if the latter condition holds, then  $W \leq a/|c|$ . Solving for  $c$  (and  $a$ ) in terms of  $\mu$  would make this truncation bound more interpretable.

## 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.

# 3 Normal Tail Behaviour

Caution: There is a lot of algebra and Taylor's theorem in this section. I'm pretty confident that I made at least one mistake somewhere, so don't take the derivation too literally. Nevertheless, I think that the process I used has some merit, and might be worth applying to more complicated problems. For now, I think it's best to look at the strategy and not take the specific details too seriously.

I need a relatively easy example to start. The exponential distribution is a bit too trivial, so let's try the next simplest: the standard normal. I am going to derive the tail function for this distribution, then find the matching Pareto tail function. I'm particularly interested in the sign of the parameter we've been calling  $k$  and Pickands III calls  $c$ . I think this is called the tail index, from a very cursory inspection of de Haan and Ferreira (2006). I'm going to use this terminology from now on.

Note that the tail index I derive will depend on  $u$ , the tail threshold. There is a strategy given by Pickands III for obtaining a  $u$ -independent tail index, as well as a simpler method we discussed but aren't completely certain about. As agreed, Dr. Ruth is going to focus on the Pickands III method and Dr. Nickchi is going to investigate our method. To start though, I need the tail function.

Writing down the tail function for the standard normal is actually pretty

simple:

$$P_u(x) := \frac{\bar{\Phi}(u+x)}{\bar{\Phi}(u)} \quad (9)$$

Note that the tail function can also be obtained using the truncated normal distribution. Inverting to get the tail quantile function, we get

$$Q_u(p) = \bar{\Phi}^{-1}[p\bar{\Phi}(u)] - u \quad (10)$$

$$= \Phi^{-1}[p\Phi(u) + (1-p)] - u \quad (11)$$

Note that this is the quantile function for the amount by which our random variable exceeds  $u$ , not the conditional distribution of a standard normal given that it exceeds  $u$ <sup>1</sup>. Applied math people tend to really like the error function, which has been extremely well studied. Let's try translating our problem from normal CDFs to error functions. In fact, it's even nicer if we work with survival functions and the complementary error function. Some relevant identities are .

$$Q_u(p) = \bar{\Phi}^{-1}[p\bar{\Phi}(u)] - u \quad (12)$$

$$= \sqrt{2}\operatorname{erfc}^{-1}\left[p \cdot \operatorname{erfc}\left(\frac{u}{\sqrt{2}}\right)\right] - u \quad (13)$$

Some relevant identities for the above derivation are:

$$\Phi(x) = [\operatorname{erf}(x/\sqrt{2}) + 1]/2 \quad (14)$$

$$\operatorname{erfc} := 1 - \operatorname{erf} \quad (15)$$

$$\bar{\Phi}(x) = \operatorname{erfc}(x/\sqrt{2})/2 \quad (16)$$

$$\bar{\Phi}^{-1}(p) = \sqrt{2}\operatorname{erfc}^{-1}(2p) \quad (17)$$

From Pickands III, the tail index is  $c_u = \log[(Q_u(1/4) - Q_u(1/2))/Q_u(1/2)]/\log(2)$ . We are particularly interested in the sign of  $c_u$ . This sign is unaffected by division by  $\log(2)$ , so an equivalent question is whether  $Q_u(1/4) - Q_u(1/2) > Q_u(1/2)$ .

Let's write  $T_u(p) = \sqrt{2}\operatorname{erfc}^{-1}[p \cdot \operatorname{erfc}(u/\sqrt{2})]$ , so that  $Q_u(p) = T_u(p) - u$ . Notice also that  $T_u(1) = u$ , so we actually have  $Q_u(p) = T_u(p) - T_u(1)$ . I'm not certain that this will be helpful, but it certainly lends itself more nicely to Taylor expansion.

We now need to evaluate  $Q_u(1/4) - Q_u(1/2) = T_u(1/4) - T_u(1/2)$ . I'm trying out some more precise notation for derivatives. The ambiguity that I want to eliminate is that  $d/df(g(x)) \neq (d/df)(g(x))$ . This is all fine if you're willing to use a lot of brackets, but people often like to just write  $\nabla f(g(x))$ . The prime mark addresses this issue to some extent, but I prefer an operator to adding

---

<sup>1</sup>The former can be obtained from the latter by working with  $X - u$ .

an extra superscript to the function. To this end, I propose that  $\nabla$  be a low “binding power” derivative operator and  $\partial$  be a high “binding power” derivative operator. By this I mean that  $\nabla f(g(x)) = \partial f(g(x)) * \partial g(x)$ . I’m still testing to see if I really like this convention and whether it sticks. So far though, I think it’s pretty nice, particularly when doing multivariate problems. Proceeding now with my adopted notation, we have

$$T_u(1/4) - T_u(1/2) \approx -\partial T_u(1/2)/4 \quad (18)$$

$$= \frac{\sqrt{2\pi}}{8} \operatorname{erfc}\left(\frac{u}{\sqrt{2}}\right) \exp\left(\operatorname{erfc}^{-1}\left[p \cdot \operatorname{erfc}\left(\frac{u}{\sqrt{2}}\right)\right]^2\right) \Big|_{p=1/2} \quad (19)$$

$$=: \frac{\sqrt{2\pi}}{8} \eta \exp\left[\operatorname{erfc}^{-1}\left(\frac{\eta}{2}\right)^2\right] \quad (20)$$

where  $\eta := \operatorname{erfc}(u/\sqrt{2})$  (note that this quantity is constant in  $p$ ). A similar argument shows that

$$T_u(1/2) - T_u(1) \approx \frac{\sqrt{2\pi}}{4} \eta \exp\left(\frac{u^2}{2}\right) \quad (21)$$

The ratio which shows up inside the logarithm for our tail index formula is thus

$$\frac{T_u(1/4) - T_u(1/2)}{T_u(1/2) - T_u(1)} \approx \frac{1}{2} \exp\left[\operatorname{erfc}^{-1}\left(\frac{\eta}{2}\right)^2 - \frac{u^2}{2}\right] \quad (22)$$

$$=: \mathcal{R}_u \quad (23)$$

Our approximation to the tail index itself is thus

$$c_u \approx \log(\mathcal{R}_u) / \log(2) \quad (24)$$

$$= \frac{\operatorname{erfc}^{-1}\left(\frac{\eta}{2}\right)^2 - \frac{u^2}{2}}{\log(2)} - 1 \quad (25)$$

Let’s do one more Taylor expansion to approximate the numerator of the first term. Recall that  $u^2/2 = \operatorname{erfc}^{-1}(p\eta)^2|_{p=1}$ . Therefore,

$$\operatorname{erfc}^{-1}\left(\frac{\eta}{2}\right)^2 - \frac{u^2}{2} \approx \nabla_p \operatorname{erfc}^{-1}(p\eta)^2|_{p=1} \cdot \left(-\frac{1}{2}\right) \quad (26)$$

$$= \sqrt{\frac{\pi}{8}} u \exp\left(\frac{u^2}{2}\right) \quad (27)$$

Putting this all together now, we get

$$c_u \approx \sqrt{\frac{\pi}{8}} \cdot \frac{u \exp(u^2/2)}{\log(2)} - 1 \quad (28)$$

### 3.1 Discussion

I did most of the above derivation by hand, so there’s a pretty good chance that I messed something up somewhere. Having the tail index go to  $\infty$  as  $u \rightarrow \infty$  doesn’t really align with my expectations; I would have preferred that the first term go to zero, so the asymptotic tail index is just  $-1$ . It’s entirely possible that I just missed a minus sign somewhere along the way. Going back over this with Maple would help, I just don’t have time for that today.

It still remains to do the optimization outlined by Pickands III (1975) to find a tail index which doesn’t depend on  $u$ . I’m going to have to do that another day. I don’t really want to go much further until I’ve verified that there aren’t any typos in the above derivation. Of course, there is also the issue of all the Taylor approximations. It’s not totally clear to me that all of the remainder terms are negligible. Sounds like a good thing to think about while double-checking the derivation.

There is a more serious issue with the above math as well. Ultimately, we’re going to write code to compute tail indices. Since we’re going to evaluate everything numerically anyway, we might as well just stick with the original expression for the tail function in terms of normal CDFs. There are methods for finding roots of arbitrary functions which can just solve for  $x_1$  and  $x_2$  exactly<sup>2</sup>. It will certainly be easier to apply the optimization-flavoured method of Pickands III (1975) on the tail index formula in (28), but this leaves lots of questions related to the validity of the approximations. Sounds like a good direction to explore if we want to write a math-stat paper. Less fruitful if we want to write code to improve importance sampling.

I guess I also need to look back through Pickands III (1975) more closely to see what examples he does. Tails of the standard normal distribution have been extremely well-studied, so somebody has almost certainly done the above derivation before, or at least something similar. I need to have a better look through the literature.

## References

- de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*. Springer.
- Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1).

---

<sup>2</sup>Rather, as exactly as we can expect on a computer. In statistics, we tend not to worry about this discrepancy.

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