# A Heavy-Tailed Algebra for Probabilistic Programming

**SUPPLEMENTARY MATERIAL**

## A Operations in the Generalized Gamma Algebra

In this section, we provide explanations, references, and new results for how operations on random variables affect their GGA tails. A summary of this, useful for referencing, appeared in Table 1.

### A.1 Ordering

A total ordering is imposed on the equivalence classes of $\mathcal{G}$ according to the heaviness of tails. In particular, we say that $(\nu_1, \sigma_1, \rho_1) \leq (\nu_2, \sigma_2, \rho_2)$ if $(x^{\nu_1} e^{-\sigma_1 x^{\rho_1}})/(x^{\nu_2} e^{-\sigma_2 x^{\rho_2}})$ is bounded as $x \to \infty$. As usual, we say $(\nu_1, \sigma_1, \rho_1) < (\nu_2, \sigma_2, \rho_2)$ if $(\nu_1, \sigma_1, \rho_1) \leq (\nu_2, \sigma_2, \rho_2)$ but $(\nu_1, \sigma_1, \rho_1) \not\equiv (\nu_2, \sigma_2, \rho_2)$.

### A.2 Addition

Tails of this form are closed under addition. Combining subexponentiality for $\rho < 1$ [1, Chapter X.1], with [2, Thm 3.1 & eqn. (8.3)], we obtain the following Proposition 1 for exactness of the proposed GGA addition operation.

**Proposition 1.** *Denoting the addition of random variables (additive convolution of densities) by $\oplus$,*

$$(\nu_1, \sigma_1, \rho_1) \oplus (\nu_2, \sigma_2, \rho_2) \equiv \begin{cases} \max\{(\nu_1, \sigma_1, \rho_1), (\nu_2, \sigma_2, \rho_2)\} & \text{if } \rho_1 \neq \rho_2 \text{ or } \rho_1, \rho_2 < 1 \\ (\nu_1 + \nu_2 + 1, \min\{\sigma_1, \sigma_2\}, 1) & \text{if } \rho_1 = \rho_2 = 1 \\ (\nu_1 + \nu_2 + 1 - \frac{\rho}{2}, (\sigma_1^{-\frac{1}{\rho - 1}} + \sigma_2^{-\frac{1}{\rho - 1}})^{1-\rho}, \rho) & \text{if } \rho = \rho_1 = \rho_2 > 1. \end{cases} \tag{3}$$

### A.3 Powers

For all exponents $\beta > 0$, by invoking a change of variables $x \mapsto x^\beta$, it is easy to show that $(\nu, \sigma, \rho)^\beta \equiv \left(\frac{\nu+1}{\beta} - 1, \sigma, \frac{\rho}{\beta}\right)$.

### A.4 Reciprocals

We *define* negative powers and reciprocals equivalently to positive powers in the case $\beta < 0$. This equivalence cannot be proven to hold in general since we cannot determine tail asymptotics of the reciprocal without knowledge of its behaviour around zero. Therefore, we implicitly assume that the behaviour around zero mimics the tail behaviour, that is, Equation (1) holds as $x \to 0^+$. Note that this can only hold provided $(\nu + 1)/\rho > 0$ and $\rho \neq 0$. To account for all other cases, including $\mathcal{R}_\nu$, we assume that the density of $X$ approaches some nonzero value near zero. In this case, Lemma 1 defines the reciprocal to be $\mathcal{R}_2$.

**Lemma 1.** *Assume that a random variable $X$ has a density $p$ that is continuous at zero and $p(0) > 0$. Then $X^{-1} \equiv \mathcal{R}_2$.*

*Proof.* From a change of variables, the density $q$ of $X^{-1}$ is given by $q(x) = |x|^{-2} p(x^{-1})$. By assumption, as $|x| \to \infty$, $q(x) \sim p(0)|x|^{-2}$. Therefore, $X^{-1} \equiv \mathcal{R}_2$. □

### A.5 Multiplication

For any $c \in \mathbb{R} \backslash \{0\}$, it can be readily seen from a change of variables $x \mapsto cx$ that $c(\nu, \sigma, \rho) = (\nu, \sigma/|c|^\rho, \rho)$. However, the case of multiplication convolution is not as straightforward. While additive convolutions of generalized Gamma random variables are relatively well-explored, to our knowledge, multiplicative convolution has not been examined at this level of generality. It turns out that the class $\mathcal{G}$ is also closed under multiplication (assuming independence of random variables), as we show in the following result. The proof requires some preliminary background on Mellin transforms and the Fox H function, which we cover in Appendix E.

**Proposition 2.** *Denoting the multiplication of independent random variables (multiplicative convolution) by* $\otimes$,

$$(\nu_1,\sigma_1,\rho_1)\otimes(\nu_2,\sigma_2,\rho_2)\equiv\begin{cases}\left(\frac{1}{\mu}\left(\frac{\nu_1}{|\rho_1|}+\frac{\nu_2}{|\rho_2|}+\frac{1}{2}\right),\sigma,-\frac{1}{\mu}\right) & \text{if } \rho_1,\rho_2<0\\ \left(\frac{1}{\mu}\left(\frac{\nu_1}{\rho_1}+\frac{\nu_2}{\rho_2}-\frac{1}{2}\right),\sigma,\frac{1}{\mu}\right) & \text{if } \rho_1,\rho_2>0\\ \mathcal{R}_{|\nu_1|} & \text{if } \rho_1\le 0,\rho_2>0\\ \mathcal{R}_{\min\{|\nu_1|,|\nu_2|\}} & \text{if } \rho_1=0,\rho_2=0\end{cases}$$

*where* $\mu=\frac{1}{|\rho_1|}+\frac{1}{|\rho_2|}=\frac{|\rho_1|+|\rho_2|}{|\rho_1\rho_2|}$ *and* $\sigma=\mu(\sigma_1|\rho_1|)^{\frac{1}{\mu|\rho_1|}}(\sigma_2|\rho_2|)^{\frac{1}{\mu|\rho_2|}}$.

*Proof.* The $\rho_1\le 0,\rho_2>0$ and $\rho_1=\rho_2=0$ cases follow from Breiman's lemma [5, Lemma B.5.1]. Our argument proceeds similar to [2]. Assume that $\rho_1,\rho_2>0$ and let $0<\epsilon<1$ be such that $0<a_-<a_+<1$, where

$$a_+=\frac{(1+\epsilon)\rho_2}{\rho_1+\rho_2},\qquad a_-=1-\frac{(1+\epsilon)\rho_1}{\rho_1+\rho_2}.$$

Then for $\rho=\frac{\rho_1\rho_2}{\rho_1+\rho_2}$, if $X\equiv(\nu_1,\sigma_1,\rho_1)$ and $Y\equiv(\nu_2,\sigma_2,\rho_2)$, then

$$\begin{aligned}\mathbb{P}(XY>x,X\notin[x^{a_-},x^{a_+}])&\le\mathbb{P}(X>x^{a_+})+\mathbb{P}(Y>x^{1-a_-})\\&\sim c_1x^{\nu_1a_+}e^{-\sigma_1x^{\rho_1a_+}}+c_2x^{\nu_2(1-a_-)}e^{-\sigma_2x^{\rho_2(1-a_-)}}\\&\le\left(c_1x^{\nu_1a_+}+c_2x^{\nu_2(1-a_-)}\right)e^{-\min\{\sigma_1,\sigma_2\}x^{(1+\epsilon)\rho}}=o(x^\nu e^{-\sigma x^\rho}),\end{aligned}$$

for any $\nu,\sigma>0$. Hence, it will suffice to show the claimed tail asymptotics for the generalized Gamma distribution. In this case, since $a_->0$ and $a_+<1$, the tail of the distribution for the product of $X,Y$ depends only on the tail of the distributions for $X$ and $Y$.

Therefore, assume without loss of generality that $p_X(x)=c_Xx^{\nu_1}e^{-\sigma_1x^{\rho_1}}$ and $p_Y(x)=c_Yx^{\nu_2}e^{-\sigma_2x^{\rho_2}}$. Then

$$\mathcal{M}_s[p_{XY}]=c_Xc_Y\frac{\sigma_1^{-\nu_1/\rho_1}}{\rho_1}\frac{\sigma_2^{-\nu_2/\rho_2}}{\rho_2}\left(\sigma_1^{1/\rho_1}\sigma_2^{1/\rho_2}\right)^{-s}\Gamma\left(\frac{\nu_1}{\rho_1}+\frac{s}{\rho_1}\right)\Gamma\left(\frac{\nu_2}{\rho_2}+\frac{s}{\rho_2}\right).$$

Consequently,

$$p_{XY}(z)=c_Xc_Y\frac{\sigma_1^{-\nu_1/\rho_1}}{\rho_1}\frac{\sigma_2^{-\nu_2/\rho_2}}{\rho_2}H_{0,2}^{2,0}\left[\sigma_1^{1/\rho_1}\sigma_2^{1/\rho_2}z\,\Big|_{(\frac{\nu_1}{\rho_1},\frac{1}{\rho_1}),(\frac{\nu_2}{\rho_2},\frac{1}{\rho_2})}^{-}\right]$$

Computing the corresponding $\beta,\delta,\mu$ for the asymptotic expansion, we find that

$$\mu=\frac{1}{\rho_1}+\frac{1}{\rho_2},\qquad\delta=\frac{\nu_1}{\rho_1}+\frac{\nu_2}{\rho_2}-1,\qquad\beta=\rho_1^{-1/\rho_1}\rho_2^{-1/\rho_2}.$$

Consequently, for some $c>0$,

$$p_{XY}(z)\sim cz^{\frac{1}{\mu}(\frac{1}{2}+\delta)}\exp\left(-\mu\beta^{-\frac{1}{\mu}}(\sigma_1^{1/\rho_1}\sigma_2^{1/\rho_2})^{\frac{1}{\mu}}z^{\frac{1}{\mu}}\right),$$

which completes the $\rho_1,\rho_2>0$ case. The final case follows by composing the multiplication and reciprocal operations. Note that

$$\begin{aligned}(\nu_1,\sigma_1,-\rho_1)^{-1}\otimes(\nu_2,\sigma_2,-\rho_2)^{-1}&\equiv(-\nu_1-2,\sigma_1,\rho_1)\otimes(-\nu_2-2,\sigma_2,\rho_2)\\&\equiv\left(\frac{1}{\mu}\left(\frac{-\nu_1-2}{\rho_1}+\frac{-\nu_2-2}{\rho_2}-\frac{1}{2}\right),\sigma,\frac{1}{\mu}\right)\\&\equiv\left(\frac{1}{\mu}\left(\frac{-\nu_1}{\rho_1}+\frac{-\nu_2}{\rho_2}-2\mu-\frac{1}{2}\right),\sigma,\frac{1}{\mu}\right)\\&\equiv\left(\frac{1}{\mu}\left(\frac{-\nu_1}{\rho_1}+\frac{-\nu_2}{\rho_2}-\frac{1}{2}\right)-2,\sigma,\frac{1}{\mu}\right),\end{aligned}$$

and therefore

$$(\nu_1,\sigma_1,-\rho_1)\otimes(\nu_2,\sigma_2,-\rho_2)\equiv\left(\frac{1}{\mu}\left(\frac{\nu_1}{\rho_1}+\frac{\nu_2}{\rho_2}+\frac{1}{2}\right),\sigma,-\frac{1}{\mu}\right).$$

$\square$

### A.6 Product of Densities

We can also consider a product of densities operation acting on two random variables $X, Y$, denoted $X\&Y$, by $p_{X\&Y}(x) = cp_X(x)p_Y(x)$, where $c > 0$ is an appropriate normalizing constant and $p_X, p_Y, p_{X\&Y}$ are the densities of $X$, $Y$, and $X\&Y$, respectively. In terms of the equivalence classes:

$$(\nu_1, \sigma_1, \rho_1)\&(\nu_2, \sigma_2, \rho_2) \equiv \begin{cases} (\nu_1 + \nu_2, \sigma_1, \rho_1) & \text{if } \rho_1 < \rho_2 \\ (\nu_1 + \nu_2, \sigma_1 + \sigma_2, \rho) & \text{if } \rho = \rho_1 = \rho_2 \\ (\nu_1 + \nu_2, \sigma_2, \rho_2) & \text{otherwise,} \end{cases}$$

which follows directly by taking the product of the generalized Gamma tails in eq. (1). Note that this particular operation does not require either $p_X$ or $p_Y$ to be normalized — only the tail behaviour is needed. We may also use this to work out the tail behaviour of a posterior density, provided the tail behaviour of the likelihood in the parameters is known.

### A.7 Exponential and Logarithm

Tails of the generalized Gamma form are not closed under exponentiation or logarithms. Indeed, if both $X$ and $\exp X$ have generalized Gamma tails, then $X$ is exponentially distributed (and $\exp X$ has power law tails). As a workaround, we can consider an upper bound on the tail by projecting onto the nearest possible exponentially distributed / power law tail. If $\rho > 1$, then a change of variables shows the density of $\exp X$ satisfies

$$p_{\exp X}(x) \sim \frac{c}{x}(\log x)^\nu \exp\left(-\sigma(\log x)^\rho\right) \leq \frac{\tilde{c}}{x}\exp\left(-\sigma(\log x)\right) = cx^{-\sigma-1}, \text{ as } x \to \infty.$$

The inverse of this operation sends $\mathcal{R}_{\sigma+1}$ to $(0, \sigma, 1)$. With this in mind, we define the exponential and logarithmic operations according to the following: $\exp(\nu, \sigma, \rho) \equiv \mathcal{R}_{\sigma+1}$ if $\rho \geq 1$, otherwise $\mathcal{R}_1$; $\log(\nu, \sigma, \rho) \equiv (0, |\nu| - 1, 1)$ if $\nu < -1$ and $\rho \leq 0$, otherwise $\mathcal{L}$.

### A.8 Lipschitz Functions

There are many multivariate functions that cannot be readily represented in terms of the operations covered thus far. For these, it is important to specify the tail behaviour of pushforward measures under Lipschitz-continuous functions. Fortunately, this is covered by Theorem 2 below, presented in [31, Proposition 1.3].

**Theorem 2.** *For any Lipschitz continuous function $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\|f(x) - f(y)\| \leq L\|x - y\|$ for $x, y \in \mathbb{R}^d$, there is $f(X_1, \dots, X_d) \equiv L\max\{X_1, \dots, X_d\}$. More generally, for any Hölder continuous function $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\|f(x) - f(y)\| \leq L\|x - y\|^\alpha$ for $x, y \in \mathbb{R}^d$, there is $f(X_1, \dots, X_d) \equiv L\max\{X_1^\alpha, \dots, X_d^\alpha\}$.*

### A.9 Power Law Approximation

There are many cases where power laws arise not from a single operation of random variables, but cumulatively, through many successive operations. In these cases, $\rho$ becomes small while $\sigma$ becomes large, such that $\sigma = \mathcal{O}(\rho^{-1})$. To see how this regime induces a power law, note that as $x \to \infty$,

$$p_{|X|}(x) \sim cx^\nu e^{-\sigma x^\rho} = \tilde{c}x^\nu e^{-\sigma(x^\rho - 1)} = \tilde{c}x^\nu e^{-\sigma\rho\frac{x^\rho - 1}{\rho}} \approx \tilde{c}x^\nu e^{-\sigma\rho\log x} = \tilde{c}x^{\nu - \sigma\rho},$$

where we have used the approximation $\log x = \rho^{-2}(x^\rho - 1) + \mathcal{O}(\rho^2)$. Consequently, we can represent tails of this form by the Student $t$ distribution with $|\nu - \sigma\rho| - 1$ degrees of freedom. In practice, we find this approximation tends to *overestimate* the heaviness of the tail.

Alternatively, the generalized Gamma density (2) satisfies $\mathbb{E}X^r = \sigma^{-r/\rho}\Gamma(\frac{\nu+1+r}{\rho})/\Gamma(\frac{\nu+1}{\rho})$ for $r > 0$. Let $\alpha > 0$ be such that $\mathbb{E}X^\alpha = 2$. By Markov's inequality, the tail of $X$ satisfies $\mathbb{P}(X > x) \leq 2x^{-\alpha}$. Therefore, we can represent tails of this form by the Student $t$ distribution with $\alpha + 1$ degrees of freedom (generate $X \sim \text{StudentT}(\alpha)$). In practice, we find this approximation to be more accurate, and is hence used as our power law candidate distribution in Section 3.2.

### A.10 Posterior Distributions

Suppose that a random variable $X$ is dependent on a parameter $\theta$ and a latent random element $Z$ through a function $f$ by $X = f(Z;\theta)$. Letting $\pi$ denote a prior on $\theta$, since $p(\theta|x) \propto p_X(x|\theta)\pi(\theta)$, it will suffice to find the tail of $p_X(x|\theta)$ in $\theta$, as we can incorporate the tail of $\pi$ with the & operation. Assuming that $f$ is invertible with respect to both $Z$ and $\theta$ with respective inverses $f^{-1}(x;\theta)$ and $\Theta(x;z)$, a change of variables shows that

$$p_X(x|\theta) = p_Z(f^{-1}(x;\theta)) \left|\frac{\partial}{\partial x} f^{-1}(x;\theta)\right|.$$

Note that $z = f^{-1}(x;\Theta(x;z))$ and so $\Theta^{-1}(\theta;x) = f^{-1}(x;\theta)$, where $\Theta^{-1}(x;\theta)$ is the inverse of $z \mapsto \Theta(x;z)$ at $\theta$. Therefore, the density of $\Theta(x;Z)$ is

$$p_\Theta(\theta,x) = p_Z(f^{-1}(x;\theta)) \left|\frac{\partial}{\partial \theta} f^{-1}(x;\theta)\right|.$$

Consequently,

$$p_X(x|\theta) = p_\Theta(\theta,x)R(x,\theta),$$

where $R(x,\theta) = |\frac{\partial}{\partial x} f^{-1}(x;\theta)|/|\frac{\partial}{\partial \theta} f^{-1}(x;\theta)|$. Since the inverse of a composition of operations is a composition of inverses, the tail of $p_\Theta$ is relatively straightforward to determine by tracing back through the computation graph and sequentially applying inverse operations, i.e., $\oplus$ (addition) becomes $\ominus$ (subtraction), etc. For example, if $X = \mu + Z$, then $f(z,\mu) = \mu + z$, $f^{-1}(x;\mu) = x - \mu$, and $R(x,\mu) = 1$. Therefore, $\mu|X = x \equiv (x - z) \,\&\, \pi$. Similarly, if $X = \sigma Z$, then $f(z,\sigma) = \sigma z$, $f^{-1}(x,\sigma) = x/\sigma$, and $R(x,\sigma) = \sigma^{-1}/(x\sigma^{-2}) \equiv \sigma$. Therefore, $\mu|X = x \equiv (x/Z) \,\&\, (1,1,0) \,\&\, \pi$. If $X = Z/\sigma$, then $f(Z,\sigma) = z/\sigma$, $f^{-1}(x,\sigma) = \sigma x$, and $R(x,\sigma) = \sigma/x \equiv \sigma$.

## B List of Univariate Distributions

To demonstrate the scope of our algebra and facilitate implementation in a general PPL, Table 5 lists many families of one-dimensional densities and their corresponding tail class.

Table 5: List of univariate distributions

| Name | Support | Density $p(x)$ | Class |
|---|---|---|---|
| Benktander Type II | $(0,\infty)$ | $e^{\frac{a}{b}(1-x^b)}x^{b-2}(ax^b - b + 1)$ | $(2b-2, \frac{a}{b}, b)$ |
| Beta prime | $(0,\infty)$ | $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1+x)^{-\alpha-\beta}$ | $\mathcal{R}_{\beta+1}$ |
| Burr | $(0,\infty)$ | $ckx^{c-1}(1+x^c)^{-k-1}$ | $\mathcal{R}_{ck+1}$ |
| Cauchy | $(-\infty,\infty)$ | $(\pi\gamma)^{-1}\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]^{-1}$ | $\mathcal{R}_2$ |
| Chi | $(0,\infty)$ | $\frac{1}{2^{k/2-1}\Gamma(k/2)}x^{k-1}e^{-x^2/2}$ | $(k-1, \frac{1}{2}, 2)$ |
| Chi-squared | $(0,\infty)$ | $\frac{1}{2^{k/2}\Gamma(k/2)}x^{\frac{k}{2}-1}e^{-x/2}$ | $(\frac{k}{2}-1, \frac{1}{2}, 1)$ |
| Dagum | $(0,\infty)$ | $\frac{ap}{x}\left(\frac{x}{b}\right)^{ap}\left(\left(\frac{x}{b}\right)^a+1\right)^{-p-1}$ | $\mathcal{R}_{a+1}$ |
| Davis | $(0,\infty)$ | $\propto (x-\mu)^{-1-n}/\left(e^{\frac{b}{x-\mu}}-1\right)$ | $(-1-n, b, -1)$ |
| Exponential | $(0,\infty)$ | $\lambda e^{-\lambda x}$ | $(0,\lambda,1)$ |
| $F$ | $(0,\infty)$ | $\propto x^{d_1/2-1}(d_1 x + d_2)^{-(d_1+d_2)/2}$ | $\mathcal{R}_{d_2/2+1}$ |
| Fisher $z$ | $(-\infty,\infty)$ | $\propto \frac{e^{d_1 x}}{(d_1 e^{2x}+d_2)^{(d_1+d_2)/2}}$ | $(0, d_2, 1)$ |

| | | | |
|---|---|---|---|
| Frechet | $(0,\infty)$ | $\frac{\alpha}{\lambda}\left(\frac{x-m}{\lambda}\right)^{-1-\alpha} e^{-\left(\frac{x-m}{\lambda}\right)^{-\alpha}}$ | $(-1-\alpha, \lambda^\alpha, -\alpha)$ |
| Gamma | $(0,\infty)$ | $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ | $(\alpha-1, \beta, 1)$ |
| Gamma/Gompertz | $(0,\infty)$ | $bse^{bx}\beta^s/(\beta-1+e^{bx})^{s+1}$ | $(0, bs, 1)$ |
| Generalized hyperbolic | $(-\infty,\infty)$ | $\propto e^{\beta(x-\mu)} \frac{K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2})}{(\delta^2+(x-\mu)^2)^{1/4-\lambda/2}}$ | $(\lambda-1, \alpha-\beta, 1)$ |
| Generalized normal | $(-\infty,\infty)$ | $\frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left(\frac{\lvert x-\mu\rvert}{\alpha}\right)^\beta\right)$ | $(0, \alpha^{-\beta}, \beta)$ |
| Geometric stable | $(-\infty,\infty)$ | no closed form | $\mathcal{R}_{\alpha+1}$ |
| Gompertz | $(0,\infty)$ | $\sigma\eta\exp(\eta+\sigma x-\eta e^{\sigma x})$ | $\mathcal{L}$ |
| Gumbel | $(0,\infty)$ | $\beta^{-1}e^{-(\beta^{-1}(x-\mu)+e^{-\beta^{-1}(x-\mu)})}$ | $(0, \frac{1}{\beta}, 1)$ |
| Gumbel Type II | $(0,\infty)$ | $\alpha\beta x^{-\alpha-1}e^{-\beta x^{-\alpha}}$ | $(-\alpha-1, \beta, -\alpha)$ |
| Holtsmark | $(-\infty,\infty)$ | no closed form | $\mathcal{R}_{5/2}$ |
| Hyperbolic secant | $(-\infty,\infty)$ | $\frac{1}{2}\mathrm{sech}\left(\frac{\pi x}{2}\right)$ | $(0, \frac{\pi}{2}, 1)$ |
| Inverse chi-squared | $(0,\infty)$ | $\frac{2^{-k/2}}{\Gamma(k/2)} x^{-k/2-1} e^{-1/(2x)}$ | $(-\frac{k}{2}-1, \frac{1}{2}, -1)$ |
| Inverse gamma | $(0,\infty)$ | $\frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}$ | $(-\alpha-1, \beta, -1)$ |
| Levy | $(0,\infty)$ | $\sqrt{\frac{c}{2\pi}}(x-\mu)^{-3/2}e^{-\frac{c}{2(x-\mu)}}$ | $(-\frac{3}{2}, \frac{c}{2}, -1)$ |
| Laplace | $(-\infty,\infty)$ | $\frac{1}{2\lambda}\exp\left(-\frac{\lvert x-\mu\rvert}{\lambda}\right)$ | $(0, \frac{1}{\lambda}, 1)$ |
| Logistic | $(-\infty,\infty)$ | $\frac{e^{-(x-\mu)/\lambda}}{\lambda(1+e^{-(x-\mu)/\lambda})^2}$ | $(0, \frac{1}{\lambda}, 1)$ |
| Log-Cauchy | $(0,\infty)$ | $\frac{\sigma}{x\pi}((\log x-\mu)^2+\sigma^2)^{-1}$ | $\mathcal{R}_1$ |
| Log-Laplace | $(0,\infty)$ | $\frac{1}{2\lambda x}\exp\left(-\frac{\lvert\log x-\mu\rvert}{\lambda}\right)$ | $\mathcal{R}_{1/\lambda+1}$ |
| Log-logistic | $(0,\infty)$ | $\frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1}\left(1+\left(\frac{x}{\alpha}\right)^\beta\right)^{-2}$ | $\mathcal{R}_{\beta+1}$ |
| Log-$t$ | $(0,\infty)$ | $\propto x^{-1}(1+\frac{1}{\nu}(\log x-\mu)^2)^{-\frac{\nu+1}{2}}$ | $\mathcal{R}_1$ |
| Lomax | $(0,\infty)$ | $\frac{\alpha}{\lambda}\left(1+\frac{x}{\lambda}\right)^{-\alpha-1}$ | $\mathcal{R}_{\alpha+1}$ |
| Maxwell-Boltzmann | $(0,\infty)$ | $\sqrt{\frac{2}{\pi}}\frac{x^2e^{-x^2/(2\sigma^2)}}{\sigma^3}$ | $(2, \frac{1}{2\sigma^2}, 2)$ |
| normal | $(-\infty,\infty)$ | $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ | $(0, \frac{1}{2\sigma^2}, 2)$ |
| Pareto | $(x_0,\infty)$ | $\alpha x_0^\alpha x^{-\alpha-1}$ | $\mathcal{R}_{\alpha+1}$ |
| Rayleigh | $(0,\infty)$ | $\frac{x}{\sigma^2}e^{-x^2/(2\sigma^2)}$ | $(1, \frac{1}{2\sigma^2}, 2)$ |
| Rice | $(0,\infty)$ | $\frac{x}{\sigma^2}\exp\left(-\frac{(x^2+\nu^2)}{2\sigma^2}\right)I_0\left(\frac{x\nu}{\sigma^2}\right)$ | $(\frac{1}{2}, \frac{1}{2\sigma^2}, 2)$ |
| Skew normal | $(-\infty,\infty)$ | no closed form | $(0, \frac{1}{2\sigma^2}, 2)$ |
| Slash | $(-\infty,\infty)$ | $\frac{1-e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}x^2}$ | $(-2, \frac{1}{2}, 2)$ |
| Stable | $(-\infty,\infty)$ | no closed form | $\mathcal{R}_{\alpha+1}$ |

| | | | |
|---|---|---|---|
| Student's $t$- | $(-\infty,\infty)$ | $\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ | $\mathcal{R}_{\nu+1}$ |
| Tracy-Widom | $(-\infty,\infty)$ | no closed form | $(-\frac{3\beta}{4}-1,\frac{2\beta}{3},\frac{3}{2})$ |
| Voigt | $(-\infty,\infty)$ | no closed form | $\mathcal{R}_2$ |
| Weibull | $(0,\infty)$ | $\frac{\rho}{\lambda}\left(\frac{x}{\lambda}\right)^{\rho-1}e^{-(x/\lambda)^{\rho}}$ | $(\rho-1,\lambda^{-\rho},\rho)$ |

The following densities are not supported by our algebra: Benini distribution; Benktander Type I distribution; Johnson's $S_U$-distribution; and the log-normal distribution. All of these densities exhibit log-normal tails.

## C Theoretical Examples

To verify that our GGA yields accurate predictions of tail behaviour, we work out some explicit GGA computations on several standard distributions using operations in Table 1. By doing so, we recover some common probability identities.

**Example 3** (Chi-squared random variables)**.** Let $X_1,\dots,X_k$ be $k$ independent standard normal random variables. The variable $Z=\sum_{i=1}^k X_i^2$ is *chi-squared distributed* with $k$ degrees of freedom. Using the GGA, we can accurately determine the tail behaviour of this random variable directly from its construction. Recall that each $X_i \equiv (0,1/2,2)$, and by the power operation, $X_i^2 \equiv (-1/2,1/2,1)$. Applying the addition operation $k$ times reveals that $Z \equiv (k/2-1,1/2,1)$ and implies that the density of $Z$ is asymptotically $cx^{k/2-1}e^{-x/2}$ as $x\to\infty$. In fact, it is known that the density of $Z$ is exactly $p_Z(x)=c_k x^{k/2-1}e^{-x/2}$, where $c_k=2^{-k/2}/\Gamma(k/2)$.

**Example 4** (Products of random variables)**.** To demonstrate the multiplication operation in our algebra, we consider the product of two exponential, Gaussian, and reciprocal Gaussian random variables. Traditionally, asymptotics for the distribution of the product of two random variables would be found analytically. For example, consider the following Lemma 3.

**Lemma 3.** *Let $X_1,X_2\sim$ Exp$(\lambda)$ and $Z_1,Z_2\sim\mathcal{N}(0,1)$ be independent. As $x\to\infty$, the densities of $X_1X_2$, $Z_1Z_2$ and $Z=1/Z_1\cdot 1/Z_2$ satisfy*

$$p_{X_1X_2}(x)\sim\frac{\lambda^{3/2}\sqrt{\pi}}{x^{1/4}}e^{-2\lambda\sqrt{x}},\quad p_{Z_1Z_2}(x)\sim\frac{1}{\sqrt{2\pi x}}e^{-x},\quad p_Z(x)\sim\frac{1}{\sqrt{2\pi}|z|^{3/2}}e^{-1/|z|}.$$

With ease, our algebra correctly determines that $X_1X_2\equiv(-\frac{1}{4},2\lambda,\frac{1}{2})$, $Z_1Z_2\equiv(-\frac{1}{2},1,1)$ and $Z\equiv(-\frac{3}{2},1,-1)$. We now demonstrate how one would ascertain these asymptotics manually.

*Proof of Lemma 3.* The proof relies on the following integral definition [56, pg. 183] and asymptotic relation as $z\to\infty$ [56, pg. 202] of the modified Bessel function $K_\nu(z)$ for $z>0$ and $\nu\geq 0$,

$$K_\nu(z)=\frac{1}{2}\left(\frac{z}{2}\right)^\nu\int_0^\infty u^{-\nu-1}\exp\left(-u-\frac{z^2}{4u}\right)\mathrm{d}u\sim\sqrt{\frac{\pi}{2z}}e^{-z}. \tag{4}$$

We also make use of the known density for the product of two independent continuous random variables: if $X$ and $Y$ have densities $p_X$ and $p_Y$ respectively, then $Z=XY$ has density

$$p_Z(z)=\int_{\mathbb{R}}p_X(x)p_Y(z/x)|x|^{-1}\mathrm{d}x.$$

- **Density of $X_1X_2$:** Recalling that the density of $X\sim\mathrm{Exp}(\lambda)$ is $p_X(x)=\lambda e^{-\lambda x}$ for $x\geq 0$, for $Z=XY$ where $X\sim\mathrm{Exp}(\lambda_1)$ and $Y\sim\mathrm{Exp}(\lambda_2)$ are independent,

$$p_Z(z)=\int_0^\infty x^{-1}\lambda_1 e^{-\lambda_1 x}\lambda_2 e^{-\lambda_2 z/x}\mathrm{d}x=\lambda_1\lambda_2\int_0^\infty x^{-1}e^{-\lambda_1 x-\lambda_2 z/x}\mathrm{d}x.$$

Since $2K_0(2\sqrt{z}) = \int_0^\infty u^{-1}\exp(-u - \frac{z}{u})\mathrm{d}u$, let $u = \lambda_1 v$, so that $\mathrm{d}u = \lambda_1 \mathrm{d}v$,

$$2K_0(2\sqrt{\lambda_1\lambda_2 z}) = \int_0^\infty u^{-1}\exp\left(-\lambda_1 v - \lambda_2\frac{z}{v}\right)\mathrm{d}v.$$

Therefore, letting $\lambda = \sqrt{\lambda_1\lambda_2}$,

$$p_Z(z) = 2\lambda^2 K_0(2\lambda\sqrt{z}) \sim \sqrt{\pi}\lambda^{3/2}z^{-1/4}e^{-2\lambda z^{1/2}}.$$

- **Density of $Z_1Z_2$:** Recalling that the density of $X \sim \mathcal{N}(0,1)$ is $p_X(x) = (2\pi)^{-1/2}\exp(-\frac{1}{2}x^2)$, for $Z = XY$ where $X, Y \sim \mathcal{N}(0,1)$ are independent,

$$\begin{aligned} p_Z(z) &= \frac{1}{2\pi}\int_{\mathbb{R}} |x|^{-1}\, e^{-\frac{1}{2}x^2}e^{-\frac{1}{2}z^2/x^2}\mathrm{d}x \\ &= \frac{1}{\pi}\int_0^\infty x^{-1}e^{-\frac{1}{2}x^2-\frac{1}{2}z^2/x^2}\mathrm{d}x \\ &= \frac{1}{\pi}\int_0^\infty x^{-1}e^{-\frac{1}{2}x^2-\frac{1}{2}z^2/x^2}\mathrm{d}x. \end{aligned}$$

Let $u = \frac{1}{2}x^2$ so that $\mathrm{d}u = x\mathrm{d}x$ and

$$K_\nu(z) = z^\nu\int_0^\infty x^{-2\nu-1}\exp\left(-\frac{1}{2}x^2 - \frac{z^2}{2x^2}\right)\mathrm{d}x.$$

In particular, for any $z \in \mathbb{R}$,

$$K_0(|z|) = \int_0^\infty x^{-1}\exp\left(-\frac{1}{2}x^2 - \frac{z^2}{2x^2}\right)\mathrm{d}x, \tag{5}$$

and so

$$p_Z(z) = \frac{1}{\pi}K_0(|z|) \sim \frac{1}{\sqrt{2\pi|z|}}e^{-|z|}.$$

- **Density of $Z$:** Finally, by a change of variables, we note that the density of $X^{-1}$ where $X \sim \mathcal{N}(0,1)$ is $p_{X^{-1}}(x) = (2\pi)^{-1/2}x^{-2}\exp(-\frac{1}{2x^2})$. Therefore, the density of $Z = 1/(XY)$ where $X, Y \sim \mathcal{N}(0,1)$ are independent is given by

$$\begin{aligned} p_Z(z) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}x^2}e^{-\frac{1}{2x^2}}\frac{x^2}{\sqrt{2\pi}z^2}e^{-\frac{x^2}{2z^2}}\frac{1}{|x|}\mathrm{d}x \\ &= \frac{1}{2\pi z^2}\int_{\mathbb{R}} e^{-\frac{1}{2x^2}-\frac{x^2}{2z^2}}\frac{1}{|x|}\mathrm{d}x \\ &= \frac{1}{\pi z^2}\int_0^\infty e^{-\frac{1}{2x^2}-\frac{x^2}{2z^2}}\frac{1}{x}\mathrm{d}x \\ &= \frac{1}{\pi z^2}K_0(|z|^{-1}) \sim \sqrt{\frac{1}{2\pi}}|z|^{-3/2}e^{-|z|^{-1}}, \end{aligned}$$

where we have once again used (5).

□

**Example 5** (Reciprocal distributions)**.** Perhaps the most significant challenge with our tail algebra is correctly identifying the tail behaviour of reciprocal distributions. Here, we test the efficacy of our formulation with known reciprocal distributions.

- *Reciprocal normal:* $X \sim \mathcal{N}(0,1) \equiv (0, 1/2, 2)$, and $X^{-1} \equiv (-2, 1/2, -2)$.
- *Inverse exponential:* $X \sim \mathrm{Exp}(\lambda) \equiv (0, \lambda, 1)$, and $X^{-1} \equiv (-2, \lambda, -1)$.

- *Inverse $t$-distribution:* $X \equiv \mathcal{R}_\nu$, and $X^{-1} \equiv \mathcal{R}_2$.
- *Inverse Cauchy:* $X \equiv \mathcal{R}_2$, it is known $X^{-1}$ has the same distribution and our theory predicts $X^{-1} \equiv \mathcal{R}_2$.

**Example 6** (Cauchy distribution)**.** A simple special case of the Student $T$ distribution is the Cauchy distribution, which arises as the ratio of two standard normal random variables. For $X \sim \mathcal{N}(0,1)$, $X \equiv (0, 1/2, 2)$ and $X^{-1} \equiv (-2, 1/2, -2)$. Hence, the multiplication operation correctly predicts that the ratio of two standard normal random variables is in $\mathcal{R}_2$.

**Example 7** (Student $T$ distribution)**.** Let $X$ be a standard normal random variable, and $V$ a chi-squared random variable with $\nu$ degrees of freedom. The random variable $T = X/\sqrt{V/\nu}$ is *$t$-distributed* with $\nu$ degrees of freedom. Since $V \equiv (\nu/2 - 1, 1/2, 1)$, multiplying by the constant $1/\nu$ reveals $V/\nu \equiv (\nu/2 - 1, 1/(2\nu), 1)$. Applying the square root operation, $\sqrt{V/\nu} \equiv (\nu - 1, 1/(2\nu), 2)$. To compute the division operation, we first take the reciprocal to find $(V/\nu)^{-1/2} \equiv (-\nu - 1, 1/(2\nu), -2)$. Finally, since $\rho = -2 < 1$ for this random variable, the multiplication operation with $X \equiv (0, 1/2, 2)$ yields $T \equiv \mathcal{R}_{\nu+1}$. Thus, the density of $T$ is asymptotically $cx^{-\nu-1}$ as $x \to \infty$. In fact, it is known that the density of $T$ satisfies $p_T(x) = c_\nu(1 + x^2/\nu)^{-(\nu+1)/2}$ where $c_\nu = \Gamma(\frac{\nu+1}{2})/\Gamma(\frac{\nu}{2})(\nu\pi)^{-1/2}$, which exhibits the predicted tail behaviour.

**Example 8** (Log-normal distribution)**.** Although the log-normal distribution does not lie in $\mathcal{G}$, the existence of log-normal tails arising from the multiplicative central limit theorem is suggested by our algebra. Let $X_1, X_2, \dots$ be independent standard normal random variables, and let $Z_k = X_1 \cdots X_{2^k}$ for each $k = 1, 2, \dots$. By the multiplicative central limit theorem, letting $\tau = \exp(\mathbb{E}\log|X_i|) \approx 1.13$,

$$\left(\frac{X_1 \cdots X_n}{\tau}\right)^{1/\sqrt{n}} \xrightarrow{\mathcal{D}} Z \qquad \text{as } n \to \infty,$$

where $Z$ is a log-normal random variable with density

$$p_Z(x) = \frac{1}{x\sqrt{2\pi}} \exp(-\tfrac{1}{2}(\log x)^2).$$

Therefore, the same is true for $V_k = (Z_k/\tau)^{2^{-k/2}}$. Using our algebra, we will attempt to reproduce the tail of this density. Letting $\tilde{Z}_k = X_{2^k} \cdots X_{2^{k+1}}$, we see that $Z_{k+1} = Z_k \tilde{Z}_k$, and $Z_k, \tilde{Z}_k$ are iid. Let $Z_k \equiv (\nu_k, \sigma_k, \rho_k)$, by induction using the multiplication operation, we find that

$$\begin{aligned}
\nu_{k+1} &= \frac{1}{\mu}\left(\frac{2\nu_k}{\rho_k} - \frac{1}{2}\right) = \nu_k - \frac{\rho_k}{4} \\
\sigma_{k+1} &= \mu\,(\sigma_k\rho_k)^{\frac{2}{\mu\rho_k}} = \frac{2}{\rho_k}(\sigma_k\rho_k) = 2\sigma_k \\
\rho_{k+1} &= \frac{1}{\mu} = \frac{\rho_k}{2}.
\end{aligned}$$

Since $\rho_0 = 2$, $\sigma_0 = 1/2$, and $\nu_0 = 0$, we find that $\rho_k = 2^{1-k}$ and $\sigma_k = 2^{k-1}$. Furthermore, $\nu_{k+1} = \nu_k - 2^{-k-1}$ and so $\nu_k = -1 + 2^{-k}$. Therefore

$$\begin{aligned}
Z_k &\equiv (-1 + 2^{-k}, 2^{k-1}, 2^{1-k}), \qquad \text{and,} \\
V_k &\equiv (-1 + 2^{-k/2}, 2^{k-1}\tau^{-2^{1-k}}, 2^{1-k/2}),
\end{aligned}$$

and letting $\epsilon_k = 2^{-k/2}$, the tail behaviour of the density of $V_k$ satisfies

$$\begin{aligned}
p_k(x) &\sim c_k x^{-1+\epsilon_k} \exp\left(-\frac{\epsilon_k^{-2}}{2\tau^{-2\epsilon_k^2}} x^{2\epsilon_k}\right) \\
&\sim c_k x^{-1+\epsilon_k} \exp\left(-\frac{1}{2\tau^{-2\epsilon_k^2}}\left(\frac{x^{\epsilon_k} - 1}{\epsilon_k}\right)^2\right) \approx c_k x^{-1} \exp\left(-\frac{1}{2}(\log x)^2\right),
\end{aligned}$$

as $x \to \infty$, where the approximation improves as $k$ gets larger. The quality of this approximation is demonstrated in Figure 6.

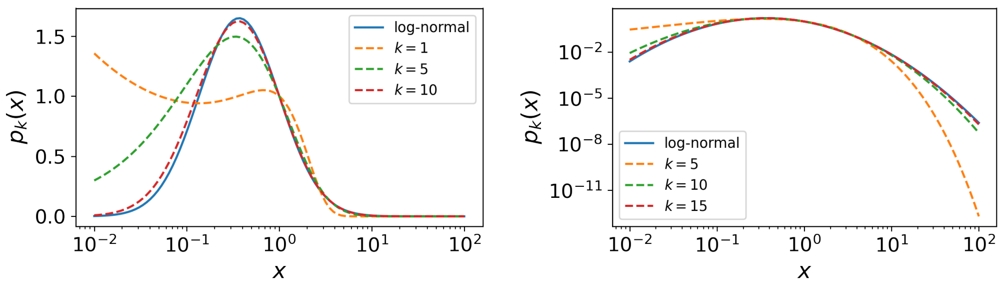


Figure 6: Estimation of the log-normal density (blue) by the representative density chosen by the GGA applied to $V_k$ for $k = 1, 5, 10$ (orange, green, red, respectively), as presented on a log-linear scale (left) and a log-log scale (right).

## D Additional Details for Experiments

The targets in Table 2 and Table 4 are analyzed using the GGA in Appendix C. Note that Inverse Gamma ("IG") corresponds to the inverse exponential. We selected closed form targets so that the Pareto tail index $\alpha$ is known analytically and the quality of theoretical predictions as well as empirical results can be rigorously evaluated. All experiments are repeated on i7-8700K CPU and GTX 1080 GPU hardware for 100 trials. 1000 samples from the model (as well as the approximation in VI) were used to compute each gradient estimate. Losses were trained until convergence, which all occurred in under $10^4$ iterations at a $0.05$ learning rate and the Adam [27] optimizer.

## E Mellin Transforms

Recall that the Mellin transform of a function $f$ on $(0, \infty)$ is given by

$$\mathcal{M}_s[f] = \int_0^\infty x^{s-1} f(x) \mathrm{d}x.$$

Letting $p_{XY}$ denote the density of the product of independent random variables $X, Y$ with respective densities $p_X$ and $p_Y$, $\mathcal{M}_s[p_{XY}] = \mathcal{M}_s[p_X]\mathcal{M}_s[p_Y]$. There is

$$\mathcal{M}_s[cx^\nu e^{-\sigma x^\rho}] = \frac{c\sigma^{-\nu/\rho}}{\rho} \sigma^{-s/\rho} \Gamma\left(\frac{\nu}{\rho} + \frac{s}{\rho}\right).$$

To facilitate the proof of Proposition 2, we define the Fox $H$-function

$$H_{p,q}^{m,n}\left[z \,\middle|\, \begin{smallmatrix}(a_1,A_1),\dots,(a_p,A_p)\\(b_1,B_1),\dots,(b_q,B_q)\end{smallmatrix}\right]$$

as the inverse Mellin transform of

$$\Theta(s) = z^{-s} \frac{\prod_{j=1}^m \Gamma(b_j + B_j s) \cdots \prod_{j=1}^n \Gamma(1 - a_j - A_j s)}{\prod_{j=m+1}^q \Gamma(1 - b_j - B_j s) \prod_{j=n+1}^p \Gamma(a_j + A_j s)}.$$

An important property of the Fox $H$-function is its asymptotic behaviour as $z \to \infty$. From [34, Theorem 1.3],

$$H_{p,q}^{q,0}\left[z \,\middle|\, \begin{smallmatrix}(a_1,A_1),\dots,(a_p,A_p)\\(b_1,B_1),\dots,(b_q,B_q)\end{smallmatrix}\right] \sim c x^{(\delta+\frac{1}{2})/\mu} \exp(-\mu \beta^{-1/\mu} x^{1/\mu}), \qquad \text{as } x \to \infty,$$

for some constant $c > 0$, where $\beta = \prod_{j=1}^p (A_j)^{-A_j} \prod_{j=1}^q B_j^{B_j}$, $\mu = \sum_{j=1}^q B_j - \sum_{j=1}^p A_j$, and $\delta = \sum_{j=1}^q b_j - \sum_{j=1}^p a_j + \frac{p-q}{2}$.