Università
Ca'Foscari
Venezia

Do parametrisations matter?

Nathan Huet
Isadora Antoniano-Villalobos, Claudia Collarin, Ilaria Prosdocimi
Ca' Foscari University of Venice

JdS 2026
June 3, 2026

What is a parametrisation?

A parametrisation is a choice of parameters used to characterise a distribution.

Example: the Gamma distribution

- with shape and scale parameters $(\alpha, \theta) \in (\mathbb{R}_+^*)^2$

$$f^\Gamma(x; \alpha, \theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\frac{x}{\theta})$$

- with shape and rate parameters $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$

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

- with mean and dispersion parameters $(\mu, \nu) \in (\mathbb{R}_+^*)^2$

$$f^\Gamma(x; \mu, \nu) = \frac{\nu^{\nu}}{\Gamma(\nu)\mu^{\nu}} x^{\nu-1} \exp(-\frac{\nu x}{\mu})$$

What is a parametrisation?

A parametrisation is a choice of parameters used to characterise a distribution.

Example: the Gamma distribution

- with shape and scale parameters $(\alpha, \theta) \in (\mathbb{R}_+^*)^2$

$$f^\Gamma(x; \alpha, \theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\frac{x}{\theta})$$

- with shape and rate parameters $(\alpha, \beta) \in (\mathbb{R}_+^*)^2$

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

- with mean and dispersion parameters $(\mu, \nu) \in (\mathbb{R}_+^*)^2$

$$f^\Gamma(x; \mu, \nu) = \frac{\nu^{\nu}}{\Gamma(\nu)\mu^{\nu}} x^{\nu-1} \exp(-\frac{\nu x}{\mu})$$

Which one to use?

Orthogonal parametrisation

$l(\boldsymbol{\theta})$ a log-likelihood depending on

$$\boldsymbol{\theta} := (\psi_1, \dots, \psi_m, \phi_1, \dots, \phi_p) := (\boldsymbol{\psi}, \boldsymbol{\phi}).$$

$\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are orthogonal if $\forall j \in \{1, \dots, m\}, k \in \{1, \dots, p\}$,

$$i_{\psi_j \phi_k} := \mathbb{E}\Big[\frac{\partial l(\boldsymbol{\theta})}{\partial \psi_j} \frac{\partial l(\boldsymbol{\theta})}{\partial \phi_k}\Big] = 0.$$

In general, impossible to achieve ($m + p$ unknown quantities for mp restrictions), but...

How to obtain an orthogonal parametrisation?

(Cox and Reid, 1987)

if $\boldsymbol{\theta} = (\psi, \phi_1, \ldots, \phi_p) := (\psi, \boldsymbol{\phi})$, with ψ a parameter of interest, $\phi_1, \ldots, \phi_p$ nuisance parameters

easier to construct a parametrisation

$$(\psi, \boldsymbol{\lambda}) \mapsto \boldsymbol{\theta}(\psi, \boldsymbol{\lambda}) = (\psi, \boldsymbol{\phi}(\psi, \boldsymbol{\lambda})) \text{ such that } l(\boldsymbol{\theta}(\psi, \boldsymbol{\lambda})) = \tilde{l}(\psi, \boldsymbol{\lambda}),$$

with $\tilde{l}$ the reparametrised log-likelihood,

which is orthogonal, $\forall j \in \{1, \ldots, p\}$,

$$\mathbb{E}\left[\frac{\partial^2 \tilde{l}(\psi, \boldsymbol{\lambda})}{\partial \psi \partial \lambda_j}\right] = 0.$$

reduced to solve the PDE system

$$\sum_{i=1}^{p} i_{\phi_i \phi_j} \frac{\partial \phi_i(\psi, \boldsymbol{\lambda})}{\partial \psi} = -i_{\psi \phi_j},$$

with $i_{..}$ a Fisher information.

Example : Gamma distribution

- Gamma distribution in a classic parametrisation (α, θ):

$$f^{\Gamma}(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} \exp(-x/\theta),$$

for $x \geq 0, \alpha > 0$ and $\theta > 0$.

- an orthogonal parametrisation $(\alpha, \omega) \mapsto (\alpha, \theta(\alpha, \omega))$:

$$i_{\theta\theta} \frac{\partial\theta(\alpha,\omega)}{\partial\alpha} = -i_{\alpha\theta} \iff \frac{\partial\theta(\alpha,\omega)}{\partial\alpha} = -\frac{\theta(\alpha,\omega)}{\alpha} \Longrightarrow \theta(\alpha,\omega) = \frac{C(\omega)}{\alpha}$$

- with $\theta(\alpha, \omega) = \frac{\omega}{\alpha}$, the density with orthogonal parameters :

$$f^{\Gamma}(x; \alpha, \omega) = \frac{\alpha^{\alpha}}{\Gamma(\alpha)\omega^{\alpha}} x^{\alpha-1} \exp(-\alpha x/\omega),$$

for $x \geq 0, \alpha > 0$ and $\omega > 0$.

Some new orthogonal parametriations

(Huet and Prosdocimi, 2026)

distribution	original density	reparametrisation
Gamma	$f^{\Gamma}(x;\alpha,\theta)=\frac{1}{\Gamma(\alpha)\theta^{\alpha}}x^{\alpha-1}\exp(-x/\theta)$	1. $(\alpha,\omega)\mapsto(\alpha,\theta(\alpha,\omega))$ with $\theta(\alpha,\omega)=\frac{\omega}{\alpha},\ \alpha,\omega>0$
Gumbel	$f^{Gumbel}(x;\mu,\sigma)=\frac{1}{\sigma}\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\exp\left(-\exp\left(-\left(\frac{x-\mu}{\sigma}\right)\right)\right)$	1. $(\nu,\sigma)\mapsto(\mu(\nu,\sigma),\sigma)$ with $\mu(\nu,\sigma)=(1-\gamma)\sigma+\nu,\nu\in\mathbb{R},\sigma>0$ 2. $(\mu,\rho)\mapsto(\mu,\sigma(\mu,\rho))$ with $\sigma(\mu,\rho)=\frac{1-\gamma}{\pi^2/6+\gamma^2-2\gamma+1}\mu+\rho,\nu\in\mathbb{R},\rho>0$
GP	$f^{GP}(x;\sigma,\xi)=\frac{1}{\sigma}\left(1+\frac{\xi x}{\sigma}\right)_+^{-1-1/\xi}$	1. (Chavez-Demoulin and Davison, 2005): $(\rho,\xi)\mapsto(\sigma(\rho,\xi),\xi)$ with $\sigma(\rho,\xi)=\rho/(\xi+1),\rho>0,\xi>-1$ 2. $(\sigma,\zeta)\mapsto(\sigma,\xi(\sigma,\zeta))$ with $\xi(\sigma,\zeta)=\zeta-\log(\sigma)/2,\zeta\in\mathbb{R},\rho>0$
GEV_2*	$f^{GEV_2}(x;\sigma,\xi)=\frac{1}{\sigma}\left(\frac{\xi x}{\sigma}\right)_+^{-1-1/\xi}\exp\left(-\left(\frac{\xi x}{\sigma}\right)_+^{-1/\xi}\right)$	1. $(\rho,\xi)\mapsto(\sigma(\rho,\xi),\xi)$ with $\sigma(\rho,\xi)=\rho\lvert\xi\rvert\exp\left((1-\gamma)\xi\right),\rho>0,\xi\neq 0$

* the GEV_2 is the classic GEV distribution when $\xi\neq 0$, and the upper- or lower-bound is fixed to zero

Advantages of such parametrisations?

Main advantages:

- Decoupling the inference of a parameter of interest:
  - asymptotically independent ML estimators;
  - asymptotic estimation error on one parameter does not deteriorate the estimation of the other;
  - ⇝ "Misspecification of one component of the model does not affect inference on the other parameter" (Heller et al., 2019; Stadlmann et al., 2023).
- Beneficial properties for Bayesian inference:
  - faster convergence of MCMC algorithms (Moins, 2023; Tanskanen, 2018);
  - facilitates the computation of Jeffreys priors.
- **That's it?**

A new way for comparing the parametrisations

Parameter set volume method

- **Input**: $\{Q_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ and $\{Q_{\boldsymbol{\theta}(\boldsymbol{\eta})} : \boldsymbol{\eta} \in H\}$
  P_0 a target distribution e.g., $P_0 = Q_{\boldsymbol{\theta}_0} = Q_{\boldsymbol{\theta}(\boldsymbol{\eta}_0)}$
- for a radius $r > 0$, consider the probability ball

$$B(P_0, r) = \{Q : d(P_0, Q) \leq r\}$$

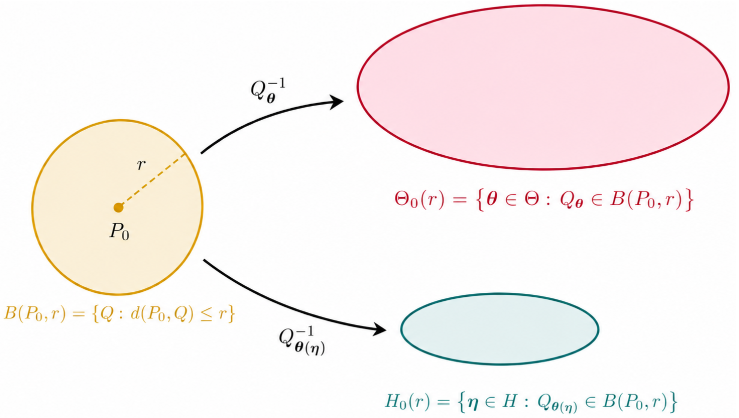
- **Compute** $\Theta_0(r)$ and $H_0(r)$ the preimages of $B(P_0, r)$

$$\Theta_0(r) = \{\boldsymbol{\theta} \in \Theta : Q_{\boldsymbol{\theta}} \in B(P_0, r)\},$$
$$H_0(r) = \{\boldsymbol{\eta} \in H : Q_{\boldsymbol{\theta}(\boldsymbol{\eta})} \in B(P_0, r)\}.$$

- **Compare** $\Theta_0(r)$ and $H_0(r)$.

How to interpret the results?

in terms of volume



$V(\Theta_0(r)) > V(H_0(r))$, then in the $\boldsymbol{\theta}$-parametrisation:

$\rightsquigarrow$ the mapping from parameters to distributions is less sensitive;

$\Longrightarrow$ distribution estimation may be more robust to parameter estimation error, but the parameters are less identifiable.

Gumbel, Gamma and GP distributions

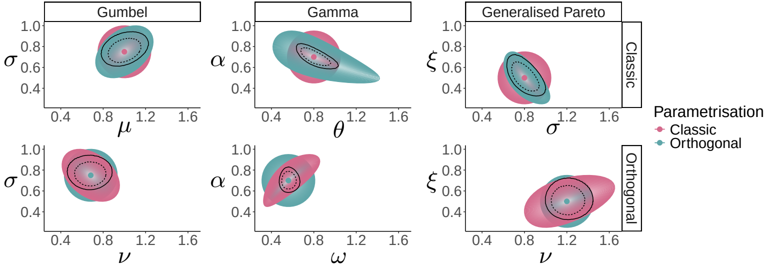


Figure: Euclidean balls in the classical parameter space are mapped into the orthogonal space (and vice-versa). Black solid (r_1) and dashed ($r_2 := r_1/2$) sets are the Θ_0 and H_0 sets.

Table: Areas of $\Theta_0(r)$ and $H_0(r)$.

	Solid (r_1)		**Dashed (r_2)**	
	Classic	**Orthogonal**	**Classic**	**Orthogonal**
Gumbel	0.11029	0.11025	0.05371	0.05380
Gamma	0.05308	0.03725	0.02640	0.01843
Generalised Pareto	0.09341	0.14178	0.04641	0.07019

What comes next?

- Explore other potential advantages of such parametrisations (do you have any ideas to share?);
- Compare our parametrisations with existing methods, both in regression and Bayesian frameworks.

  $\rightsquigarrow$ Our goal is to provide a comprehensive study of (orthogonal) parametrisations;
- In a non-stationary extreme-value context, use orthogonal parametrisations to investigate the temporal evolution of each parameter.

References

- V. Chavez-Demoulin and A. C. Davison, *Generalized additive modelling of sample extremes*, J. R. Stat. Soc., C: Appl. Stat, 2005;
- D. R. Cox and N. Reid, *Parameter orthogonality and approximate conditional inference*, J. R. Stat. Soc., B: Stat. Methodol., 1987;
- G. Z. Heller, D.-L. Couturier, and S. R. Heritier, *Beyond mean modelling: Bias due to misspecification of dispersion in poisson-inverse gaussian regression*, Biom. J., 2019;
- N. Huet and I. Prosdocimi, *Orthogonal parametrisations of Extreme-Value distributions*, Stat. Probab. Lett., 2026;
- T. Moins, *Bayesian computational methods for estimating extreme quantiles from environmental data*, PhD thesis, 2023;
- S. Stadlmann, G. Heller, T. Kneib, and L. Koens, *Parameter orthogonality transformations in distributional regression model*, preprint, 2023;
- V. Tanskanen, *Aspects of reparametrization in Gaussian process regression with the Weibull model*, PhD thesis, 2018.

Thank you for your attention!