Orthogonal parametrisations of Extreme-Value distributions

preprint arXiv:2602.16283

Nathan Huet, Isadora Antoniano-Villalobos, Claudia Collarin and Ilaria Prosdocimi

DAIS, Ca' Foscari University of Venice, Italy.

Context/Motivations

What is a parametrisation? a choice of parameters used to characterise a distribution.

example: the Gamma distribution has (at least) two common parametrisations with :

- a shape and a scale parameter (α, θ) with density
$$f^{\Gamma}(x; \alpha, \theta) = \theta^{-\alpha} x^{\alpha-1} \exp(-x/\theta)/\Gamma(\alpha);$$
- a mean and a dispersion parameter (μ, ν) with density
$$f^{\Gamma}(x; \mu, \nu) = (\nu/\mu)^{\nu} x^{\nu-1} \exp(-\nu x/\mu)/\Gamma(\nu).$$

$\rightsquigarrow$ do parametrisations matter?

depends on the context **but** orthogonal parametrisation are general advantageous parametrisations

Main drawback: loss of interpretability of the parameter (important if prior elicitation in Bayesian context)

Main advantages:

- decouple inference of a parameter of interest;
- improve computational performance;
- **that's it?**

How to obtain an orthogonal parametrisation? (Framework from [1]).

Given $l(\boldsymbol{\theta})$ a log-likelihood depending on $\boldsymbol{\theta} = (\psi, \phi_1, \ldots, \phi_p) := (\psi, \boldsymbol{\phi})$, with ψ parameter of interest, $\phi_1, \ldots, \phi_p$ nuisance parameters and $i_{..}$ Fisher information.

construct a reparametrisation $(\psi, \boldsymbol{\lambda}) \mapsto \boldsymbol{\theta}(\psi, \boldsymbol{\lambda}) = (\psi, \boldsymbol{\phi}(\psi, \boldsymbol{\lambda}))$ such that $l(\boldsymbol{\theta}(\psi, \boldsymbol{\lambda})) = \tilde{l}(\psi, \boldsymbol{\lambda})$, with $\tilde{l}$ the reparametrised log-likelihood,

which is orthogonal if, for all $j \in \{1, \ldots, p\}$,
$$\mathbb{E}\left[\frac{\partial^2 \tilde{l}(\psi, \boldsymbol{\lambda})}{\partial \psi \partial \lambda_j}\right] = 0,$$

which leads to the PDE system
$$\sum_{i=1}^{p} i_{\phi_i \phi_j} \frac{\partial \phi_i(\psi, \boldsymbol{\lambda})}{\partial \psi} = -i_{\psi \phi_j}.$$

Example (Gumbel distribution) with σ as parameter of interest
$$i_{\mu\mu} \frac{\partial \mu(\nu, \sigma)}{\partial \sigma} = -i_{\mu\sigma}.$$

Orthogonal parametrisations

distribution	original density	reparametrisation
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. [2]: $(\rho, \xi) \mapsto (\sigma(\rho, \xi), \xi)$ 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 $\sigma(\rho, \xi) = \rho\lvert\xi\rvert \exp((1-\gamma)\xi), \rho > 0, \xi \neq 0$

* the GEV_2 is the classic GEV distribution when $\xi \neq 0$, and the upper- or lower-bound is fixed to zero (implying $\mu = \sigma/\xi$)

Parameter set volume method

Comparing parametrisations : a proposal

Input a target probability distribution P_0; a divergence d between probability distributions; a radius $r > 0$; a parametric model $\{Q_{\boldsymbol{\theta}: \boldsymbol{\theta} \in \Theta}\}$.

Consider the probability ball $B(P_0, r) = \{Q : d(P_0, Q) \leq r\}$.

Compute a reparametrisation $\boldsymbol{\theta} : H \to \Theta$

$\rightsquigarrow$ alternative model $\{Q_{\boldsymbol{\theta}(\boldsymbol{\eta})} : \boldsymbol{\eta} \in H\}$

Compute $\Theta_0(r)$ and $H_0(r)$ the preimages of $B(P_0, r)$
$$\Theta_0(r) = \{\boldsymbol{\theta} \in \Theta : Q_{\boldsymbol{\theta}} \in B(P_0, r)\}, \quad \text{and}$$
$$H_0(r) = \{\boldsymbol{\eta} \in H : Q_{\boldsymbol{\theta}(\boldsymbol{\eta})} \in B(P_0, r)\}.$$

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

Interpretation in terms of volume:

if $V(\Theta_0(r))$ **larger than** $V(H_0(r))$, then in the $\boldsymbol{\theta}$-parametrisation:

- a neighbourhood in a distribution space corresponds to a broader region;

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

$\rightsquigarrow$ perturbations of the parameters may induce only small changes in the corresponding distribution;

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

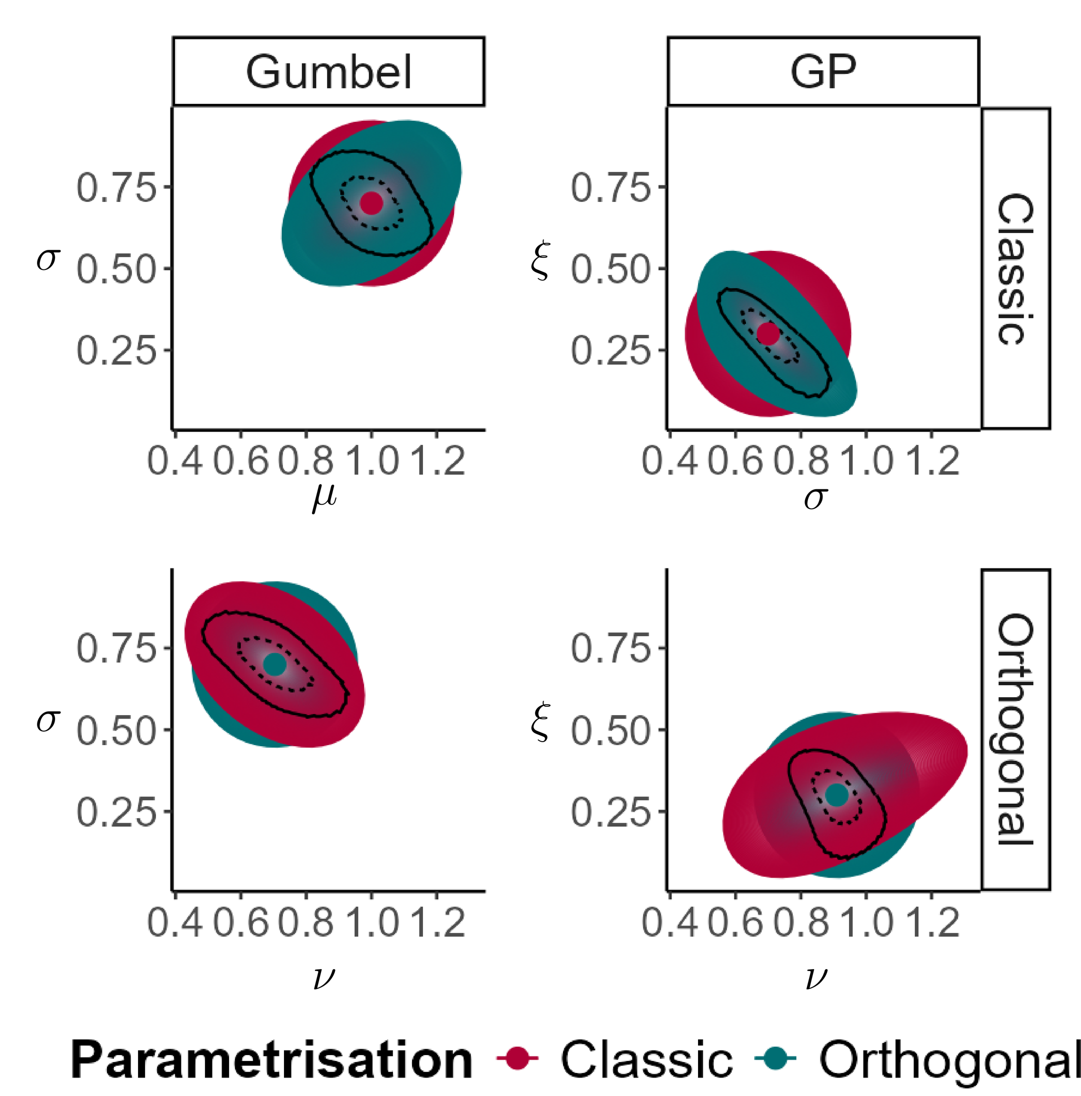

Fig. 1: Euclidean balls in the classical parameter space are mapped into the orthogonal space (and vice-versa), producing non-circular coloured shapes. Black solid (r_1) and dashed ($r_2 := r_1/2$) lines denote equidistant points in the distributional space endowed with the KL divergence.

	Solid (r_1)		Dashed (r_2)	
	Classic	Orthogonal	Classic	Orthogonal
Gumbel	0.11029	0.11025	0.05371	0.05380
GP	0.09341	0.14178	0.04641	0.07019

Table 1: Areas of $\Theta_0(r)$ and $H_0(r)$ drawn in Fig. 1.

Correlation plots

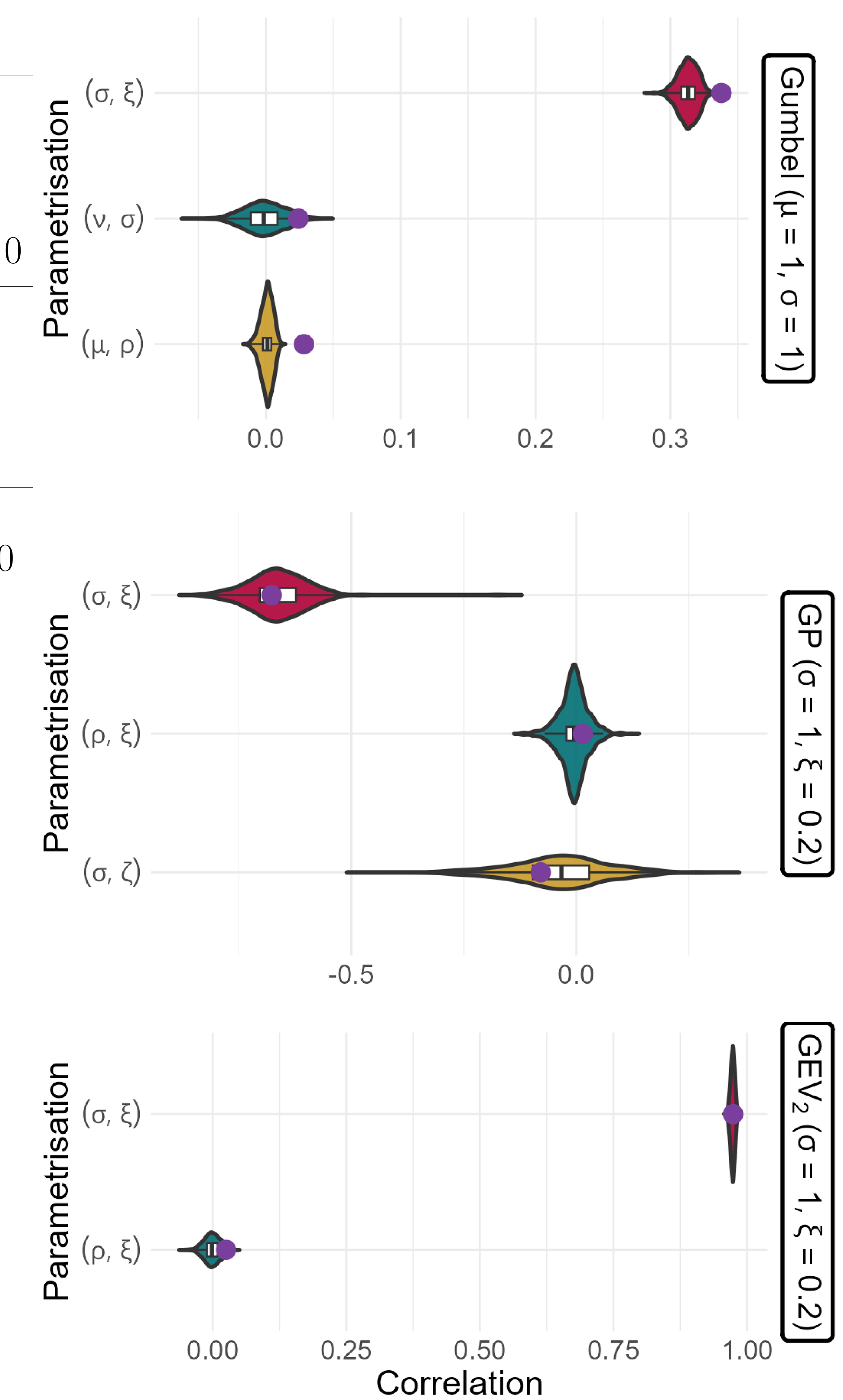

Fig. 2: Cross-correlation violin plots between parameter estimates, computed over $d = 1000$ independent replications of samples of size $n = 100$. Purple dots correspond to the correlations between the vectors of parameters obtained over all replications.

References

[1] David Roxbee Cox and Nancy Reid.
Parameter orthogonality and approximate conditional inference.
Journal of the Royal Statistical Society: Series B (Methodological), 49(1):1–18, 1987.

[2] Valérie Chavez-Demoulin and Anthony C Davison.
Generalized additive modelling of sample extremes.
Journal of the Royal Statistical Society Series C: Applied Statistics, 54(1):207–222, 2005.