

Definition. (Constructive approach) A stochastic process W in $C_+(S)$ with constant $\omega_0 > 0$ is a simple Pareto process if $W(s) = YV(s)$, for all $s \in S$, for some Y and $V = \{V(s)\}_{s \in S}$ satisfying:

- a) $V \in C_+(S)$ is a stochastic process satisfying $\sup_{s \in S} V(s) = \omega_0$ almost surely, $E[V(s)] > 0$ for all $s \in S$,
- b) Y is a standard Pareto random variable, $P(Y > y) = y^{-1}, y > 1$,
- c) Y and V are independent.

See Ferreira and de Haan (2014) for other variants of the definition.

Coefficient of asymptotic dependence. For random variables X_1 and X_2 having common marginal distribution F , let

$$\chi_{12} = \lim_{z \rightarrow z_+} P(X_1 > z | X_2 > z)$$

where z_+ is the (possibly infinite) right end-point.

For $s_1, s_2 \in S$ and $x > \omega_0$, then for $i = 1, 2$,

$$\begin{aligned} P(W(s_i) > x) &= P(YV(s_i) > x) \\ &= P\left(Y > \frac{x}{V(s_i)}\right) \\ &= E_{Y, V(s_i)} \left[\mathbb{1}\left(Y > \frac{x}{V(s_i)}\right) \right] \\ &= E_{V(s_i)} \left\{ E_{Y|V(s_i)} \left[\mathbb{1}\left(Y > \frac{x}{V(s_i)}\right) \middle| V(s_i) \right] \right\} \\ &= E_{V(s_i)} \left\{ P\left(Y > \frac{x}{V(s_i)} \middle| V(s_i)\right) \right\} \\ &= E_{V(s_i)} \left\{ \frac{V(s_i)}{x} \right\} = \frac{E[V(s_i)]}{x} \end{aligned}$$

and

$$\begin{aligned} P(W(s_1) > x, W(s_2) > x) &= P(YV(s_1) > x, YV(s_2) > x) \\ &= P\left(Y > \frac{x}{V(s_1)}, Y > \frac{x}{V(s_2)}\right) \\ &= P\left(Y > \frac{x}{V(s_1)} \vee \frac{x}{V(s_2)}\right) \\ &= P\left(Y > x \left(\frac{1}{V(s_1)} \vee \frac{1}{V(s_2)} \right)\right) \\ &= P\left(Y > x \left(\frac{1}{V(s_1) \wedge V(s_2)} \right)\right) \end{aligned}$$

$$= \frac{E[V(s_1) \wedge V(s_2)]}{x}$$

by using arguments similar in the first set of equations. Then for a simple Pareto process at points $s_1, s_2 \in S$, we have

$$\begin{aligned} \chi_{12}^W &= \lim_{x \rightarrow \infty} P(W(s_1) > x | W(s_2) > x) \\ &= \lim_{x \rightarrow \infty} \frac{x E[V(s_1) \wedge V(s_2)]}{x E[V(s_2)]} \\ &= \frac{E[V(s_1) \wedge V(s_2)]}{E[V(s_2)]} \end{aligned}$$

This is from Ferreira and de Haan (2014), but with some clarity (for the dummies of the universe) as to how they got to the expectations.

NOTE: This has some issues since the marginals of $W(s_1)$ and $W(s_2)$ need not be the same. Unless conditioning on them both being above a certain value (i.e. 1 or ω_0), they have different marginals (except the degenerate case when $W(s_1) = W(s_2)$ a.s. where they are both standard paretos. We did not explicitly account for this conditioning in the above calculation.

Better χ . From before, we can find the distribution function for $W_i \equiv W(s_i)$, as

$$F_{W_i}(w) = 1 - \frac{E(V_i)}{w}$$

where $V_i \equiv V(s_i)$. Using this fact, we can standardize W_i to be uniform and compute χ in this way

$$\begin{aligned} \chi_1 &= \lim_{u \rightarrow 1} P(F_{W_1}(W_1) > u | F_{W_2}(W_2) > u) \\ &= \lim_{u \rightarrow 1} P\left(1 - \frac{E(V_1)}{W_1} > u \mid 1 - \frac{E(V_2)}{W_2} > u\right) \\ &= \lim_{u \rightarrow 1} P\left(W_1 > \frac{E(V_1)}{1-u} \mid W_2 > \frac{E(V_2)}{1-u}\right) \\ &= \lim_{u \rightarrow 1} \frac{P\left(W_1 > \frac{E(V_1)}{1-u}, W_2 > \frac{E(V_2)}{1-u}\right)}{P\left(W_2 > \frac{E(V_2)}{1-u}\right)} \\ &= \lim_{u \rightarrow 1} \frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u}\right)}{1-u} \\ &= \lim_{u \rightarrow 1} \frac{1}{1-u} P\left(Y > \frac{E(V_1)}{(1-u)V_1}, Y > \frac{E(V_2)}{(1-u)V_2}\right) \\ &= \lim_{u \rightarrow 1} \frac{1}{1-u} P\left(Y > \frac{E(V_1)}{(1-u)V_1} \vee \frac{E(V_2)}{(1-u)V_2}\right) \\ &= \lim_{u \rightarrow 1} \frac{1}{1-u} P\left(Y > \frac{1}{1-u} \left(\frac{E(V_1)}{V_1} \vee \frac{E(V_2)}{V_2}\right)\right) \end{aligned}$$

$$\begin{aligned}
&= \lim_{u \rightarrow 1} \frac{1}{1-u} (1-u) E \left[\left(\frac{E(V_1)}{V_1} \vee \frac{E(V_2)}{V_2} \right)^{-1} \right] \\
&= E \left(\frac{V_1}{E(V_1)} \wedge \frac{V_2}{E(V_2)} \right)
\end{aligned}$$

which is nice and symmetric. Concerns: Are we correctly using the fact that Y is standard Pareto? If we don't explicitly state that $x > 1$:

$$P(Y > x) = \min(1, 1/x)$$

Another concern goes back to solving the distribution function for W_i . Since V_i is a random variable, we required conditioning on it using iterated expectations to get the desired result. However, it is possible that $P(V_i = 0) > 0$, and so dividing by V_i may pose an issue. This may be circumvented perhaps by conditioning on $V_i > 0$, but how would this affect the calculation?

For $x > 0$,

$$\begin{aligned}
P(W_i > x) &= P(YV_i > x) \\
&= P(YV_i > x | V_i = 0)P(V_i = 0) + P(YV_i > x | V_i > 0)P(V_i > 0) \\
&= P(Y > x/V_i | V_i = 0)P(V_i = 0) + P(YV_i > x | V_i > 0)P(V_i > 0) \\
&= P(Y > \infty)P(V_i = 0) + P(YV_i > x | V_i > 0)P(V_i > 0) \\
&= 0 \times P(V_i = 0) + P(YV_i > x | V_i > 0)P(V_i > 0) \\
&= P(YV_i > x | V_i > 0)P(V_i > 0) \\
&\stackrel{?}{=} P(V_i > 0) \times \frac{E(V_i | V_i > 0)}{x}
\end{aligned}$$

Proposition:

$$\frac{E(V_i)}{x} = P(V_i > 0) \times \frac{E(V_i | V_i > 0)}{x}$$

It is certainly possible in practice to observe a $V_i = 0$. For the joint distribution we may need to condition on $V_1 \wedge V_2 > 0$.

$$\begin{aligned}
\chi_2 &= \lim_{u \rightarrow 1} P(F_{W_1}(W_1) > u | F_{W_2}(W_2) > u) \\
&= \lim_{u \rightarrow 1} P \left(1 - \frac{E(V_1)}{W_1} > u \mid 1 - \frac{E(V_2)}{W_2} > u \right) \\
&= \lim_{u \rightarrow 1} P \left(W_1 > \frac{E(V_1)}{1-u} \mid W_2 > \frac{E(V_2)}{1-u} \right) \\
&= \lim_{u \rightarrow 1} \frac{P \left(W_1 > \frac{E(V_1)}{1-u}, W_2 > \frac{E(V_2)}{1-u} \right)}{P \left(W_2 > \frac{E(V_2)}{1-u} \right)}
\end{aligned}$$

$$\begin{aligned}
&= \lim_{u \rightarrow 1} \frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u}\right)}{1-u} \\
&= \lim_{u \rightarrow 1} \left[\frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u} \middle| V_1 \wedge V_2 = 0\right) P(V_1 \wedge V_2 = 0)}{1-u} \right. \\
&\quad \left. + \frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u} \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0)}{1-u} \right] \\
&= \lim_{u \rightarrow 1} \left[\frac{0 \times P(V_1 \wedge V_2 = 0)}{1-u} \right. \\
&\quad \left. + \frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u} \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0)}{1-u} \right] \\
&= \lim_{u \rightarrow 1} \frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u} \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0)}{1-u} \\
&= \lim_{u \rightarrow 1} \frac{1}{1-u} P\left(Y > \frac{E(V_1)}{(1-u)V_1}, Y > \frac{E(V_2)}{(1-u)V_2} \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0) \\
&= \lim_{u \rightarrow 1} \frac{1}{1-u} P\left(Y > \frac{E(V_1)}{(1-u)V_1} \vee \frac{E(V_2)}{(1-u)V_2} \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0) \\
&= \lim_{u \rightarrow 1} \frac{1}{1-u} P\left(Y > \frac{1}{1-u} \left(\frac{E(V_1)}{V_1} \vee \frac{E(V_2)}{V_2}\right) \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0) \\
&= \lim_{u \rightarrow 1} \frac{1}{1-u} (1-u) E\left[\left(\frac{E(V_1)}{V_1} \vee \frac{E(V_2)}{V_2}\right)^{-1} \middle| V_1 \wedge V_2 > 0\right] P(V_1 \wedge V_2 > 0) \\
&= E\left(\frac{V_1}{E(V_1)} \wedge \frac{V_2}{E(V_2)} \middle| V_1 \wedge V_2 > 0\right) P(V_1 \wedge V_2 > 0)
\end{aligned}$$

(Note the expectations within the larger expectation (i.e. $E(V_1)$ and $E(V_2)$) are not condition on $V_1 \wedge V_2 > 0$.) When conditioning $V_1 \wedge V_2 > 0$, there was no difference when computing estimates for χ_1 and χ_2 . This is perhaps due to the following.

Suppose we have samples X_1, \dots, X_n having p.d.f. $f(x) = p\delta_0(x) + (1-p)g(x)$ where $g(x)$ is defined on $(0, \infty)$. Then

$$\begin{aligned}
E(X) &= \int_0^\infty xp\delta_0(x)dx + \int_0^\infty x(1-p)g(x)dx \\
&= 0 + (1-p)E_g(X) \\
&= P(X > 0)E_g(X)
\end{aligned}$$

We could compute the sample mean in two ways:

$$\bar{X}_1 = \frac{1}{n} \sum_{i=1}^n x_i$$

and

$$\bar{X}_2 = \left[\frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i > 0) \right] \left[\frac{1}{|A|} \sum_{i \in A} x_i \right]$$

where $A = \{x : x > 0\}$ and $|A|$ is the size of A . The two multiplicands in \bar{X}_2 are estimates for $P(X > 0)$ and $E_g(X)$, respectively. Of course, \bar{X}_1 and \bar{X}_2 are equal, they just use the data differently:

$$\begin{aligned} \bar{X}_2 &= \left[\frac{1}{n} \sum_{i=1}^n \mathbf{1}(x_i > 0) \right] \left[\frac{1}{|A|} \sum_{i \in A} x_i \right] \\ &= \left[\frac{1}{|A|} \sum_{i=1}^n \mathbf{1}(x_i > 0) \right] \left[\frac{1}{n} \sum_{i \in A} x_i \right] \\ &= \left[\frac{1}{|A|} \sum_{i=1}^n \mathbf{1}(x_i > 0) \right] \left[\frac{1}{n} \sum_{i=1}^n x_i \right] \\ &= [1] \times [\bar{X}_1] = \bar{X}_1 \end{aligned}$$

I suspect we have a similar situation with χ_1 and χ_2 . That is, I believe $\chi_1 = \chi_2$.

Modeling V in the bivariate case

The primary result we use is Theorem 3.2 from Ferreira and de Haan (2014). Let $C(S)$ be the space of continuous real functions on S , equipped with the supremum norm, where S is a compact subset of \mathbb{R}^d . Let X be from $C(S)$. Then the conditions of their Theorem 3.1 imply

$$\lim_{t \rightarrow \infty} P \left(T_t X \in A \mid \sup_{s \in S} T_t X(s) > 1 \right) = P(W \in A)$$

with $A \in \mathcal{B}(C_1^+(S))$, $P(\partial A) = 0$, W some simple Pareto process, and

$$T_t X = \left(1 + \xi \frac{X - u_t}{\sigma_t} \right)_+^{1/\xi}.$$

We assume that t is large enough that the theorem kicks in (implying $u = u_t$ and $\sigma = \sigma_t$). Being interested in the bivariate case, we have observations at only two fixed locations $s_1, s_2 \in S$. The particular values of s_1 and s_2 are irrelevant since we are comparing climate simulations to observations which have no quantitative meaning as far as their position in S is concerned; we only require that the labels for the observations be appropriately distinguished.

We further assume that the parameters ξ and σ are indexed by $s \in S$, so that our transformation is

$$T_t X(s) = \left(1 + \xi(s) \frac{X(s) - u_t(s)}{\sigma_t(s)} \right)_+^{1/\xi(s)}.$$

The first stage of our analysis involves estimating $\xi(s)$ and $\sigma(s)$ marginally, which is accomplished by selecting a high threshold $u(s)$ and fitting the generalized Pareto distribution to the excesses $X(s) - u(s)$. The posterior means for $\xi(s)$ and $\sigma(s)$ are then used for the transformation from $X(s)$ to $W(s)$.

Every observation of $X(s)$, say $X_1(s), \dots, X_{n(s)}(s)$ is transformed with

$$W_i(s) = T_t X_i(s) = \left(1 + \hat{\xi}(s) \frac{X_i(s) - u_t(s)}{\hat{\sigma}(s)} \right)_+^{1/\hat{\xi}(s)}, \quad i = 1, \dots, n(s) \quad (1)$$

forming the vector $\mathbf{W}(s) = (W_1(s), \dots, W_{n(s)}(s))^\top$. After performing this transformation, two components are combined to form a joint vector $(W(s_1), W(s_2))$ having realizations $\mathbf{W}_{12} = (\mathbf{W}(s_1), \mathbf{W}(s_2))$, an $n(s) \times 2$ matrix. Note that when we perform the bivariate analysis, we guarantee that $n(s) = n(s_1) = n(s_2)$.

By Theorem 3.2, \mathbf{W}_{12} has rows that are realizations of a simple Pareto process. By the constructive definition of a simple Pareto process, we can write \mathbf{W}_{12} as

$$\mathbf{W}_{12} = \begin{pmatrix} Y_1 V_1(s_1) & Y_1 V_1(s_2) \\ Y_2 V_2(s_1) & Y_2 V_2(s_2) \\ \vdots & \vdots \\ Y_n V_n(s_1) & Y_n V_n(s_2) \end{pmatrix}$$

where Y_i is a standard Pareto, $V_i(s_j) \geq 0$ ($j = 1, 2$), and $V_i(s_1) \vee V_i(s_2) = 1$, for $i = 1, \dots, n = n(s)$. This is easily obtained by

$$Y_i = W_i(s_1) \vee W_i(s_2), \quad \text{and} \quad V_i(s_j) = W_i(s_j) / Y_i \quad (j = 1, 2), \quad \text{for } i = 1, \dots, n.$$

The points $(V_i(s_1), V_i(s_2))$ fall along the curve of the non-negative unit sphere with supremum norm $\{(v_1, v_2) : \|(v_1, v_2)\|_\infty = 1, v_1 \geq 0, v_2 \geq 0\}$ which is thus one dimensional. An alternative representation is to specify $(V_i(s_1), V_i(s_2))$ in terms of a scaled angle

$$\phi_i = \frac{2}{\pi} \arctan \left(\frac{V_i(s_2)}{V_i(s_1)} \right) \in [0, 1].$$

We scale ϕ_i to be in $[0, 1]$ so we can model the angle using a mixture of beta distributions and (possibly) point masses at zero and one. The theorem holds when $\sup_{s \in S} T_t X(s) > 1$ which is equivalent to $Y_i > 1$. Therefore, we need only those ϕ_i for which $Y_i > 1$, ignoring the rest. Beyond this point we assume that ϕ_i have been relabeled to include only the appropriate angles which are indexed $i = 1, \dots, k \leq n$.

Univariate model

We propose the likelihood

$$f(\phi|\alpha, \beta, \mathbf{p}) = \prod_{i=1}^k \left[p_1 \delta_0(\phi_i) + p_2 \delta_1(\phi_i) + \sum_{j=1}^m p_{j+2} \delta_{(0,1)}(\phi_i) \times g(\phi_i|\alpha_j, \beta_j) \right]$$

with $\delta_A(y)$ is equal to one if $y \in A$ and zero otherwise, fixed $m \geq 1$, and g is the density of a beta distribution. Also written,

$$\begin{aligned} f(\phi|\alpha, \beta, \mathbf{p}) &= \prod_{i=1}^k \left[p_1^{\mathbb{1}(\phi_i=0)} p_2^{\mathbb{1}(\phi_i=1)} \left(\sum_{j=1}^m p_{j+2} g(\phi_i|\alpha_j, \beta_j) \right)^{\mathbb{1}(0<\phi_i<1)} \right] \\ &= p_1^{\sum_{i=1}^k \mathbb{1}(\phi_i=0)} p_2^{\sum_{i=1}^k \mathbb{1}(\phi_i=1)} \prod_{i=1}^k \left[\left(\sum_{j=1}^m p_{j+2} g(\phi_i|\alpha_j, \beta_j) \right)^{\mathbb{1}(0<\phi_i<1)} \right] \end{aligned}$$

Log-likelihood:

$$\begin{aligned} \log f(\phi|\alpha, \beta, \mathbf{p}) &= \log(p_1) \sum_{i=1}^k \mathbb{1}(\phi_i = 0) + \log(p_2) \sum_{i=1}^k \mathbb{1}(\phi_i = 1) + \\ &\quad \sum_{i=1}^k \log \left[\mathbb{1}(0 < \phi_i < 1) \left(\sum_{j=1}^m p_{j+2} g(\phi_i|\alpha_j, \beta_j) \right) \right] \end{aligned}$$

Hierarchical model

Asymptotic dependence

We calculated a measure of asymptotic dependence from the simple Pareto process as

$$\begin{aligned} \chi_1 &= \lim_{u \rightarrow 1} P(F_{W_1}(W_1) > u | F_{W_2}(W_2) > u) \\ &= E \left(\frac{V_1}{E(V_1)} \wedge \frac{V_2}{E(V_2)} \right), \end{aligned}$$

but we are really interested in the level of dependence between the original variables X_1 and X_2 . Is χ_1 measuring what we want? Is it different than

$$\chi_3 = \lim_{u \rightarrow 1} P(F_{X_1}(X_1) > u | F_{X_2}(X_2) > u)?$$

Let's explore this a bit. From (1), then

$$X_i(s) = u_t(s) + \frac{\hat{\sigma}(s)}{\hat{\xi}(s)} \left(W_i(s)^{\hat{\xi}(s)} - 1 \right) \quad (2)$$

when $W_i(s) > 0$, and

$$P(X_i(s) \leq y) = E(V_i) \left(1 + \frac{\hat{\xi}(s)}{\hat{\sigma}(s)} (y - u_t(s)) \right)^{-1/\hat{\xi}(s)}. \quad (3)$$

So, we have

$$\begin{aligned} \chi_3 &= \lim_{z \rightarrow 1} P(F_{X_1}(X_1) > z | F_{X_2}(X_2) > z) \\ &= \lim_{z \rightarrow 1} P(F_{X_1}(X_1) > z | F_{X_2}(X_2) > z) \end{aligned}$$

$$E(V_i) \left(1 + \frac{\hat{\xi}(s)}{\hat{\sigma}(s)} (X_1 - u_t(s)) \right)^{-1/\hat{\xi}(s)} > z$$

References

Ferreira, A. and de Haan, L. (2014), “The generalized Pareto process; with a view towards application and simulation,” *Bernoulli*, 20, 1717–1737.