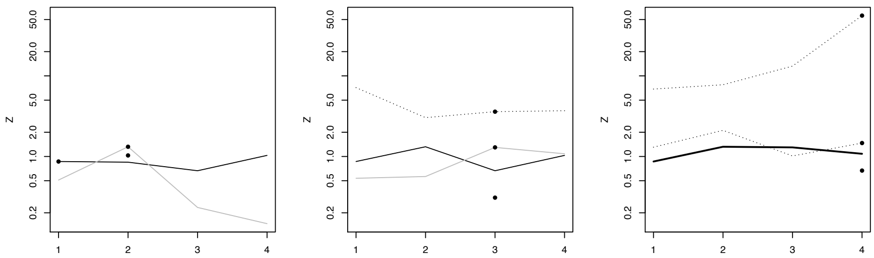

Figure 5.5 Illustration of Algorithm 5.4 with $D = 4$. For clarity the individual $R_d^* S_{-d}^*$ are joined by lines: those accepted are in grey, and those rejected are dotted. In each panel the successive values of R^* are shown as blobs; those without lines passing through them satisfy $R_d^* \le Z_d^*$. Left-hand panel: initial Z^* for $d = 1$ (solid) and simulated $R_2^* S_{-2}^*$. Middle panel: updated Z^* after $d = 2$ (solid) and simulated $R_3^* S_{-3}^*$. Right-hand panel: final Z^* for $d = 3$ (heavy) and simulated $R_4^* S_{-4}^*$.

– generate $S_{-d}^* \sim h_{-d}$,
– if $R_d^* S_{-d,d'}^* < Z_{d'}^*$ for each $d' = 1, \ldots, d-1$, set $Z^* = \max(Z^*, R_d^* S_{-d}^*)$,
– set $R_d^* = (1/R_d^* + E^*)^{-1}$.

Return Z^*. ●

Figure 5.5 illustrates this for $D = 4$ and hitting scenario $(1 \mid 2 \mid 3, 4)$.

5.3.4 Angular distributions

The angular distribution ν is the distribution of the angular variable W, which lies in the simplex $\mathcal{S}_{D-1} = \{(w_1, \ldots, w_D) : \sum_d w_d = 1, w_1, \ldots, w_D \ge 0\}$. Before we discuss this in more detail, we describe how to obtain an angular density $\dot{\nu}(w)$, if it exists, in terms of the corresponding exponent function or Poisson process measure μ.

Computation of angular density

Let $\mathcal{A}_{z_d} = \{x \in \mathcal{E} : x_d > z_d\}$, so that $\mathcal{A}_z = \bigcup_{d=1}^D \mathcal{A}_{z_d}$. Then the inclusion-exclusion formula gives

$$V(z) = \mu(\mathcal{A}_z) = \sum_{d=1}^{D} (-1)^{d+1} \sum_{\mathcal{C} \in \mathcal{D}_d} \mu(\mathcal{A}_{\mathcal{C}}), \tag{5.26}$$

where $\mathcal{D}_d$ denotes the set of distinct subsets of $\{1, \ldots, D\}$ of size d, and $\mathcal{A}_{\mathcal{C}} = \bigcap_{c \in \mathcal{C}} \mathcal{A}_{z_c}$. If the measure μ has an intensity function $\dot{\mu}$, then we can write

$$\mu(\mathcal{A}_{\mathcal{C}}) = \int_{z_{c_1}}^{\infty} \cdots \int_{z_{c_{|\mathcal{C}|}}}^{\infty} \int_0^{\infty} \cdots \int_0^{\infty} \dot{\mu}(x_1, \ldots, x_D)\, \mathrm{d}x_1 \cdots \mathrm{d}x_D,$$

where the first $|\mathcal{C}|$ integrals are over those x_c for which $c \in \mathcal{C}$. On differentiating (5.26) with respect to $z_1, \ldots, z_D$, all terms but one disappear, leaving

$$\begin{aligned} \frac{\partial^D V(z_1, \ldots, z_D)}{\partial z_1 \cdots \partial z_D} &= (-1)^{D+1} \frac{\partial^D}{\partial z_1 \cdots \partial z_D} \int_{z_1}^{\infty} \cdots \int_{z_D}^{\infty} \dot{\mu}(x_1, \ldots, x_D)\, \mathrm{d}x_1 \cdots \mathrm{d}x_D \\ &= -\dot{\mu}(z_1, \ldots, z_D). \end{aligned}$$

The $-(D+1)$-homogeneity of $\dot{\mu}$ yields

$$\dot{\mu}(z_1,\ldots,z_D) = \frac{1}{\left(\sum_d z_d\right)^{D+1}}\dot{\mu}\left(\frac{z_1}{\sum_d z_d},\ldots,\frac{z_D}{\sum_d z_d}\right) = \frac{1}{r^{D+1}}\dot{\mu}(w)$$

in terms of the pseudo-radius and angular variables,

$$r = z_1 + \cdots + z_D, \quad w = (z_1,\ldots,z_D)/r.$$

If the joint intensity of the Poisson process $\mathcal{P}$ exists, then equation (5.20) implies that it equals $Dr^{-2}\dot{\nu}(w)$. The Jacobian for the transformation from $(z_1,\ldots,z_D)$ to (r,w) is r^{D-1}, giving

$$Dr^{-2}\dot{\nu}(w) = r^{D-1}\times \dot{\mu}(z_1,\ldots,z_D)|_{z=rw} = r^{-2}\dot{\mu}(w),$$

so the angular density may be written as

$$\dot{\nu}(w) = \frac{1}{D}\dot{\mu}(w_1,\ldots,w_D) = -\frac{r^{D+1}}{D}\left.\frac{\partial^D V(z_1,\ldots,z_D)}{\partial z_1\cdots\partial z_D}\right|_{z=rw}. \tag{5.27}$$

In later chapters it will also be useful to be able to compute

$$\int_0^{u_1}\cdots\int_0^{u_C}\dot{\mu}(z_1,\ldots,z_C,z_{C+1},\ldots,z_D)\,\mathrm{d}z_1\cdots\mathrm{d}z_C,$$

and a little thought shows that this equals (Problem 5.16)

$$-\frac{\partial^{D-C}}{\partial z_{C+1}\cdots\partial z_D}V(u_1,\ldots,u_C,z_{C+1},\ldots,z_D). \tag{5.28}$$

Example 5.21 In Example 5.12, $D=2$, and taking the function V found there we obtain $\partial^2 V/\partial z_1\partial z_2 = -2/(z_1+z_2)^3$, yielding

$$\dot{\mu}(z_1,z_2) = \frac{2}{(z_1+z_2)^3}, \quad z_1,z_2>0, \quad \dot{\nu}(w) = -\frac{r^3}{2}\times\frac{-2}{r^3} = 1, \quad 0<w<1,$$

as expected. □

†This has to go later, or an example introducing this has to be earlier.

Example 5.22 (Hüsler–Reiss distribution) † Differentiation of (5.39) yields

$$-\frac{\partial^D V(z_1,\ldots,z_D)}{\partial z_1\cdots\partial z_D} = \frac{1}{z_D^2\prod_{d=1}^{D-1}(z_d\lambda_{d,D})}\phi_{D-1}(z^*_{-D};\Lambda_{-D}),$$

where $\phi_{D-1}(\cdot;\Omega)$ is the $(D-1)$-variate Gaussian density with variance matrix Ω, and

$$z^*_{d;-D} = \frac{\lambda_{d,D}}{2} + \frac{\log z_d - \log z_D}{\lambda_{d,D}}, \quad d=1,\ldots,D-1.$$

The angular density (5.27) therefore reduces to

$$\dot{\nu}(w) = \frac{1}{Dw_D^2\prod_{d=1}^{D-1}(w_d\lambda_{d,D})}\phi_{D-1}(w^*_{-D};\Lambda_{-D}), \quad \sum_{d=1}^{D} w_d = 1, \quad w_d>0,$$

where the dth element of w^*_{-D} is $\lambda_{d,D}/2 + \lambda_{d,D}^{-1}\log(w_d/w_D)$, and (5.28) is simply

$$\frac{1}{z_D^2\prod_{d=C+1}^{D-1}(z_d\lambda_{d,D})}\frac{\partial^{D-C-1}\Phi_{D-1}(z^*_{-D};\Lambda_{-D})}{\partial z_{C+1}\cdots\partial z_{D-1}},$$

with $z_1,\ldots,z_C$ replaced by $u_1,\ldots,u_C$. □

Construction of angular densities

The only constraint on the angular distribution ν is that every component of the angular variable W should have mean $1/D$, i.e., $\mathrm{E}(W) = D^{-1}1_D$. In principle this gives a lot of liberty in the choice of ν, but it is not immediately obvious how to construct models that satisfy this constraint. One general approach to doing so starts from a random variable $W^* = (W_1^*,\ldots,W_D^*)$ lying in the simplex $\mathcal{S}_{D-1}$ and having density $\dot\nu^*$ with arbitrary expectations

$$\mathrm{E}(W_d^*) = \int_{\mathcal{S}_{D-1}} w_d^*\dot\nu^*(w^*)\,\mathrm{d}w^* = m_d/D,\quad d=1,\ldots,D.$$

On setting all but one of the zs on the left-hand side of the following expression equal to infinity, we obtain

$$\mathrm{E}\left(\max_d \frac{W_d^*}{z_d m_d}\right) = \frac{D^{-1}}{z_d},\tag{5.29}$$

and comparison of this with (5.9) suggests that we rescale the W_d^*/m_d to form new variables W_d that lie in $\mathcal{S}_{D-1}$ and have the desired expectations. To achieve this we express the new variables W and the original variables W^* in terms of each other as

$$W_d = \frac{W_d^*/m_d}{\sum_{c=1}^D W_c^*/m_c},\quad W_d^* = \frac{m_dW_d}{\sum_{c=1}^D m_cW_c},\quad d=1,\ldots,D,$$

and write $\sum_{c=1}^D m_cW_c = m^{\mathrm{T}}W$. The Jacobian $|\partial w^*/\partial w|$ equals $(m^{\mathrm{T}}w)^{-D}\prod_{d=1}^D m_d$, so the left-hand side of (5.29) may be written as

$$\int_{\mathcal{S}_{D-1}} \max_d\left(\frac{w_d}{z_d}\right)(m^{\mathrm{T}}w)^{-(D+1)}\left(\prod_{d=1}^D m_d\right)\dot\nu^*\left(\frac{m_1w_1}{m^{\mathrm{T}}w},\ldots,\frac{m_Dw_D}{m^{\mathrm{T}}w}\right)\mathrm{d}w.$$

The joint density of the new variables W is therefore

$$\dot\nu(w) = (m^{\mathrm{T}}w)^{-(D+1)}\left(\prod_{d=1}^D m_d\right)\dot\nu^*\left(\frac{m_1w_1}{m^{\mathrm{T}}w},\ldots,\frac{m_Dw_D}{m^{\mathrm{T}}w}\right),\quad w\in\mathcal{S}_{D-1};$$

this is a version of $\dot\nu^*$ tilted to have the required mean vector.

Example 5.23 (Dirichlet model) The Dirichlet density is

$$\dot\nu^*(w^*;\alpha) = \frac{\Gamma(\alpha.)}{\prod_d\Gamma(\alpha_d)}\prod_d (w_d^*)^{\alpha_d-1},\quad (w_1^*,\ldots,w_D^*)\in\mathcal{S}_{D-1},$$

where $\alpha_1, \ldots, \alpha_D > 0$ and $\alpha_{\cdot} = \sum_d \alpha_d$; here $m_d = \alpha_d/\alpha_{\cdot}$. The resulting tilted density,

$$\dot{\nu}(w;\alpha) = \frac{\Gamma(\alpha_{\cdot}+1)}{\left(\sum_d \alpha_d w_d\right)^{D+1}} \prod_{d=1}^{D} \left\{ \frac{\alpha_d}{\Gamma(\alpha_d)} \left(\frac{\alpha_d w_d}{\sum_c \alpha_c w_c} \right)^{\alpha_d - 1} \right\}, \quad w \in \mathcal{S}_{D-1},$$

is symmetric around its mean $D^{-1}1_D$ only when all the α_d are equal. The corresponding Pickands function is shown in Figure 5.3 for various values of α_1 and α_2.

This model can be extended using Dirichlet densities $\dot{\nu}_1^*, \ldots, \dot{\nu}_K^*$ with respective parameter vectors $\alpha^1, \ldots, \alpha^K$ and mean vectors $\alpha^k/\alpha_{\cdot}^k$. Then if we write $\dot{\nu}^*$ as a mixture of these densities with weights $p_1, \ldots, p_K$, i.e., W^* has density

$$\dot{\nu}^*(w^*;\alpha,p) = \sum_{k=1}^{K} p_k \dot{\nu}_k^*(w^*;\alpha^k), \quad p_1, \ldots, p_K > 0, \sum_{k=1}^{K} p_k = 1;$$

in an obvious notation we have $\mathrm{E}(W_d^*) = m_d = \sum_k p_k \alpha_d^k/\alpha_{\cdot}^k$. This can be used to construct a tilted mixture that satisfies the mean constraint. □

Multivariate inference involves accounting for those features of the data that are key to the investigation at hand, and it is desirable that the key features can be directly related to particular model parameters. In many statistical settings these key features are first and second moments, but here the first moment is fixed, so dependence parameters are of primary interest.

- **Cooley et al. (2010)**
- **Ballini etc.**
- **inversion**

Angular distribution

The angular distribution can be a complex object, since ν could have densities and/or point masses on any combination of the $2^D - 1$ faces of $\mathcal{S}_{D-1}$. With $D = 3$, for example, there may be point masses on the corners $\{1,0,0\}$, $\{0,1,0\}$ and $\{0,0,1\}$, densities on the edges $\{(0,w_2,w_3) : w_2 + w_3 = 1\}$, $\{(w_1,0,w_3) : w_1 + w_3 = 1\}$ and $\{(w_1,w_2,0) : w_1 + w_2 = 1\}$, and on $S_2 = \{(w_1,w_2,w_3) : w_1 + w_2 + w_3 = 1\}$, in addition to point masses on S_2.† The only restrictions are that the total probability equal unity, and that $\mathrm{E}(W_1) = \mathrm{E}(W_2) = \mathrm{E}(W_3) = 1/3$.

†Make a figure for this.

A point mass p_1 at $\{1,0,0\}$ corresponds to extremes of Z_1 alone appearing with probability p_1, a total mass p_{23} on the face $\{(0,w_2,w_3) : w_2 + w_3 = 1\}$ corresponds to a probability p_{23} of large values of (Z_2, Z_3) occurring without large values of Z_1, with their relative sizes determined by the distribution on this face, and so forth. Since any of the faces of S_{D-1} is a lower-dimensional simplex, we can obtain the masses and/or densities from the exponent function by using (5.27) with the appropriate derivatives and then replacing the unwanted z_ds by zero.

Example 5.24 (Extremal-t model) The exponent function of the bivariate extremal-t model may be expressed as

$$V(z_1, z_2) = \frac{1}{z_1} T_{\alpha+1}\left\{a(z_2/z_1)\right\} + \frac{1}{z_2} T_{\alpha+1}\left\{a(z_1/z_2)\right\}, \quad z_1, z_2 > 0, \quad (5.30)$$

where $T_\nu(\cdot)$ denotes the Student t distribution function with ν degrees of freedom and $a(x) = (x^{1/\alpha} - c)/b$ for $x > 0$, $|c| < 1$ and $b^2 = (1-c^2)/(\alpha+1)$. If $\alpha \to \infty$ and $c \to 1$ in such a way that $2\alpha(1-c) \to \lambda^2 > 0$, then $a(x) \to \lambda/2 + \lambda^{-1}\log x$, and (5.30) converges to the Hüsler–Reiss exponent function (**??**).

Differentiation of (5.30) with respect to z_1 yields

$$-\frac{\partial V}{\partial z_1} = \frac{1}{z_1^2} T_{\alpha+1}\left\{a(x)\right\} + \frac{z_2}{z_1^2} a'(x) T'_{\alpha+1}\left\{a(x)\right\} - \frac{1}{z_2^2} a'(1/x) T'_{\alpha+1}\left\{a(1/x)\right\},$$

where $x = z_2/z_1$ and prime denotes differentiation. On setting $z_1 = rw$ and $z_2 = r(1-w)$, we obtain

$$-r^2 \frac{\partial V\{rw, r(1-w)\}}{\partial z_1} \quad \to \quad T_{\alpha+1}\left\{-(\alpha+1)^{1/2}\frac{c}{\sqrt{1-c^2}}\right\}, \quad w \to 1,$$

corresponding to $z_2 \to 0$. The same expression results from consideration of $-r^2 \partial V/\partial z_2$ when $w \to 0$, corresponding to $z_1 \to 0$, so for any finite α this model places equal point masses at $w = 0, 1$ that decrease to zero as $\alpha \to \infty$.

It is straightforward to see that $\theta(x_1, x_2) = 2T_{\alpha+1}[\{(1-c)(1+\alpha)/(1+c)\}^{1/2}]$, which has upper bound $2T_{\alpha+1}\{(1+\alpha)^{1/2}\}$; thus independence can only be attained as $\alpha \to \infty$.† □

†This para elsewhere?

The discussion above concerns limiting distributions whose statistical relevance may be questioned. Atoms of probability and observations with some component exactly equal to zero in the limiting angular distribution would lead to data in which certain values could appear repeatedly or in which extremes for one variable coincide with zeros for other variables, both of which seem unlikely in applications. Although the second might be a useful summary of certain types of events, distributions having a density function only on $\mathcal{S}_{D-1}$ will usually provide better finite-sample models.

5.3.5 Computation of exponent functions

We now describe a general approach to the computation of exponent functions. Let X be a D-dimensional continuous random variable with identical marginal distributions, and define events $\mathcal{B}_d = \{X_d > y_d\}$ and $\mathcal{C}_d = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_d$