Semiparametric Estimation for K-Sample Multivariate Extremes

de Carvalho, Miguel

Davison, Anthony

Institute of Mathematics, Analysis, and Applications; Ecole Polytechnique Fédérale de Lausanne

EPFL SB MATHAA STAT Bâtiment MA Station 8

CH-1015 Lausanne, Switzerland

E-mail: miguel.carvalho@epfl.ch   anthony.davison@epfl.ch

1   Introduction

Models for multivariate extremes are a mainstay in statistical modelling of risk. Introductions to the topic can be found in Kotz and Nadarajah (2000, §2–3), Coles (2001, §8), Beirlant et al. (2004, §8–9), and Castillo et al. (2005, §11). Mikosch (2005) overviews several issues arising in modelling multivariate extremes, and recently there has been a renewed interest in developing approaches able to cope with such difficulties (Boldi and Davison, 2007; Zhang et al., 2008; Ramos and Ledford, 2009; Stephenson, 2009; Cooley et al., 2010; Gudendorf and Segers, 2011). Recent advances have also been driven by an increasing need in several sciences to model joint tail phenomena, ranging from environmetrics (Dupuis, 2005), to oceanography (Stephenson, 2009).

An important concept in multivariate extreme value modelling is the so-called spectral measure, which controls the degree of dependence of the extremes of different variables. This measure must satisfy certain marginal moment constraints, so it is awkward both to devise suitable models and to find estimators that obey the constraints. A variety of parametric models have been proposed, for small to moderate numbers of dimensions (Tawn, 1988; Coles and Tawn, 1991; Kotz and Nadarajah, 2000), and others are currently under development. Boldi and Davison (2007) introduced a constrained mixture of Dirichlet distributions which is weakly dense in the class of all possible spectral measures, and Einmahl and Segers (2009) recently proposed a nonparametric estimator that imposes the marginal constraints using empirical likelihood (Owen, 1988, 2001).

Most available models are one-sample-based, but just as there are efficiency gains from using the threshold exceedance rather than the block maximum approach to extremal modelling (Coles, 2001), it seems less wasteful of data to attempt to combine the models for the individual subpopulations. The modelling discussed here allows us to estimate each of the spectral distribution functions H_κ using all K samples. Beyond the obvious efficiency gains, this borrowing of strength also allows improved estimation for subpopulations whose samples are too small to be individually informative about their tails.

The exposition below is largely based on Carvalho and Davison (2011), and more details on the statistical methodology and its applications can be found there.

2   Modelling K-Sample Multivariate Extremes

2.1   Multivariate Extremes and Point Processes

We begin with D-dimensional extreme value distributions specified in the classical one-sample framework. Let $Z_1, Z_2, \ldots$ be independent and identically distributed vectors of continuous random variables on $\mathbb{R}^D$ whose distribution function F lies in the joint domain of attraction of an extreme value distribution G.

Without loss of generality, suppose that F has unit Fréchet marginal distributions, i.e., $\exp(-1/z)$, for $z > 0$. Pickands' (1981) representation theorem asserts that the limiting distribution of the standardised maximum $M_n = n^{-1} \max\{Z_1, \ldots, Z_n\}$ may be written as $G(z) = \exp\{-V(z)\}$, where

$$V(z) = D \int_{\Delta_D} \max\{w_1/z_1, \ldots, w_D/z_D\} \, \mathrm{d}H(w).$$

Here H represents the so-called spectral measure defined on the unit simplex in $\mathbb{R}^D$, i.e., $\Delta_D = \{w \in \mathbb{R}_+^D : \sum_{i=1}^D w_i = 1, \; w = (w_1, \ldots, w_D)\}$. Further connections between the exponential measure V and the spectral distribution function H can be found in Coles and Tawn (1991) or Beirlant et al. (2004, §8.2). The distribution H determines the interaction between joint extremes, but must satisfy the normalization and moment conditions

(1) $\quad \displaystyle\int_{\Delta_D} \mathrm{d}H(w) = 1, \quad \int_{\Delta_D} w \mathrm{d}H(w) = D^{-1}\mathbf{1}_D,$

where $\mathbf{1}_D$ is the D-vector of ones.

A pseudo-polar transformation is helpful in understanding H: if we map the joint extremal data $e_1, \ldots, e_D$, which have unit Fréchet marginal distributions, into the pseudo-angular coordinates $w_1 = e_1/r, \ldots, w_D = e_D/r$, with pseudo-radius $r = e_1 + \cdots + e_D$, then the directional part $w_1, \ldots, w_D$ has measure H. The limiting cases of independence and dependence in the two-dimensional case illustrate this: if extreme values of the two variables tend to occur individually then the mass of H is concentrated close to the limits 0 and 1, whereas if the extremes tend to occur together then H places most of its mass near $w = 1/2$. This pseudo-polar representation also arises in a point process characterization. As $n \to \infty$, the counting process of the rescaled standard Fréchet observations, $N_n = \{n^{-1}Z_i : i = 1, \ldots, n\}$, converges in distribution to a Poisson process whose intensity measure on $\mathbb{R}_+^D$ factorizes as

$$\Lambda(dz) = \frac{\mathrm{d}r}{r^2} \times D \, \mathrm{d}H(w), \quad r > 0, w \in \Delta_D.$$

These pseudo-polar representations generalise to the K-sample case. Suppose that we have independent sets of observations $\{w_{\kappa 1}, \ldots, w_{\kappa n_\kappa}\}$ from K unknown spectral distributions $(H_0, \ldots, H_{K-1})$. The measures must satisfy the normalization and moment constraints

(2) $\quad \begin{cases} \int_{\Delta_D} \mathrm{d}H_0(w) = 1, & \int_{\Delta_D} w \mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D, \\ \int_{\Delta_D} \mathrm{d}H_1(w) = 1, & \int_{\Delta_D} w \mathrm{d}H_1(w) = D^{-1}\mathbf{1}_D, \\ \quad \vdots & \quad\quad \vdots \\ \int_{\Delta_D} \mathrm{d}H_{K-1}(w) = 1, & \int_{\Delta_D} w \mathrm{d}H_{K-1}(w) = D^{-1}\mathbf{1}_D. \end{cases}$

A similar point process representation can be given for the K-sample case by replacing the univariate point process with a multivariate point process. For each sample κ, we now suppose that $Z_{1\kappa}, Z_{2\kappa}, \ldots$ are sequences of independent and identically distributed vectors on $\mathbb{R}^D$, with a distribution function F_κ in the joint domain of attraction of G. The counting process for the K-sample problem may be denoted by $\mathcal{N}_\mathbf{n} = (N_{n_0}, \ldots, N_{n_{K-1}})$, where $\mathbf{n} = (n_0, \ldots, n_{K-1})$ and

$$\mathcal{N}_\kappa = \{n_\kappa^{-1}Z_{i\kappa} : i = 1, \ldots, n_\kappa\}, \quad \kappa = 0, \ldots, K-1.$$

The process $\mathcal{N}_{\mathbf{n}}$ converges to a multivariate Poisson process on $\mathbb{R}_{+}^{KD}$ with intensity process $\mathbf{\Lambda} = (\Lambda_0, \ldots, \Lambda_{K-1})$, as $\min\{n_0, \ldots, n_{K-1}\} \to \infty$, where each component of the intensity process factorizes along the radial and directional parts, respective to each sample, viz.:

$$\Lambda_\kappa(dz) = \frac{\mathrm{d}r_\kappa}{r_\kappa^2} \times D \, \mathrm{d}H_\kappa(w), \quad \kappa = 0, \ldots, K-1, \qquad r_\kappa > 0, w \in \Delta_D.$$

The next section introduces the spectral density ratio model for modelling K-sample multivariate extremes. Just as the point process characterization for one-sample multivariate extremes avoids the wastefulness of data implied by block maximum strategies, our model enables more effective use of the available data in the K-sample framework.

2.2 A Spectral Density Ratio Model for Multivariate Extremes

Later we focus our attention on the simplex Δ_D on which spectral measures are defined, but for now we let H_κ denote any distribution function. Our interest lies in the measures $(H_0, \ldots, H_{K-1})$, which are linked through a positive function g_κ with known functional form

$$(3) \quad \left\{ (H_0, \ldots, H_{K-1}) : \frac{\mathrm{d}H_\kappa(w)}{\mathrm{d}H_0(w)} = g(w, \gamma_\kappa), \text{ for some } g(w, \gamma_\kappa) > 0; \ \kappa = 0, \ldots, K-1 \right\}.$$

Here $\gamma = (\gamma_1, \ldots, \gamma_{K-1})^{\mathrm{T}}$ represents a q-vector of parameters, and we set $g(w, \gamma_0) = 1$ for identifiability. This specification is common to many models (Efron and Tibshirani, 1996; Qin and Zhang, 1997; Fokianos et al., 2001; Fokianos, 2004; Cheng et al., 2009). Under (3) the distributions H_κ are left unspecified but related through a known weight function. The measure H_0 acts as a reference from which the other $K-1$ measures are obtained, through a distortion controlled by g and γ_κ. Some examples are given below, where we take the two-sample case for ease of notation. Let $g(w; \alpha, \beta) = \exp\{\alpha + \beta c(w)\}$, where α is a scale parameter, β a is $p \times 1$ vector parameter, and $c(w)$ is a known distortion function, such as $c(w) = w$. We then obtain a model where the log-ratio of the densities is linear in the parameters, $\log\{\mathrm{d}H_1(w)/\mathrm{d}H_0(w)\} = \alpha + \beta w$. The logistic regression model can be derived by applying Bayes' theorem, and the multinomial logistic regression model arises for $p > 1$. A property of such models known as independence of irrelevant alternatives implies that the inference is independent of the baseline $\mathrm{d}H_0$ (Fokianos, 2004). This type of semiparametric approach provides gains in efficiency by estimating each density $\mathrm{d}H_\kappa$ on the basis of the entire sample (Gilbert et al., 1999; Fokianos, 2004; Kedem et al., 2009). Specification (3) can also be used to construct density estimates, by taking $g(w; \alpha, \beta) = \exp\{\alpha + \beta s(w)\}$, with $s(w)$ denoting a vector of sufficient statistics. Efron and Tibshirani (1996) used this representation to estimate a density $\mathrm{d}H_1/\mathrm{d}w$ on the basis of a carrier density $\mathrm{d}H_0/\mathrm{d}w$ obtained by nonparametric kernel procedures, the idea being that $\mathrm{d}H_0$ should control local adaptation to the data, while the exponential term $\exp\{\alpha + \beta s(w)\}$ should capture global features.

Specification (3) also turns out to be natural for modelling the K-sample multivariate extreme value framework discussed in §2.1, with the constraints (2) restated as

$$(4) \quad \begin{cases} \int_{\Delta_D} g(w, \gamma_0)\mathrm{d}H_0(w) = 1, & \int_{\Delta_D} w g(w, \gamma_0)\mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D, \\ \int_{\Delta_D} g(w, \gamma_1)\mathrm{d}H_0(w) = 1, & \int_{\Delta_D} w g(w, \gamma_1)\mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D, \\ \quad \vdots & \quad \vdots \\ \int_{\Delta_D} g(w, \gamma_{K-1})\mathrm{d}H_0(w) = 1, & \int_{\Delta_D} w g(w, \gamma_{K-1})\mathrm{d}H_0(w) = D^{-1}\mathbf{1}_D. \end{cases}$$

We suppress the dependence of α on β, and vice versa, but it should be noted in what follows that the normalization constraints in (4) force these parameters to be associated, and in particular if $p = 1$ they need to be perfectly correlated. Below we refer to the general semiparametric setting (3), subject to the normalization and marginal moment constraints (4), as the spectral density ratio model. We propose to fit it through empirical likelihood methods (Owen, 1988, 2001).

2.3 Estimation

Let $v = \{v_1, \ldots, v_n\}$ denote the combined sample $\{w_{01}, \ldots, w_{0n_0}, \ldots, w_{(K-1)1}, \ldots, w_{(K-1)n_{K-1}}\}$ from all K unknown spectral distributions H_κ. The likelihood of the K-sample multivariate extreme value problem under (3) is

$$\mathscr{L}(\gamma, H_0) = \prod_{k=0}^{K-1} \prod_{j=1}^{n_k} dH_k(w_{kj}) = \prod_{i=1}^{n} p_i \prod_{k=1}^{K-1} \prod_{j=1}^{n_k} g(w_{kj}, \gamma_k),$$

where $p_i = dH_0(v_i) = H_0(v_i^+) - H_0(v_i^-)$ denotes the size of the jump of the baseline spectral distribution function at the observed v_i.

We restrict our attention to the tilting function $g(w, \gamma_\kappa) = \exp\{\alpha_\kappa + \beta_\kappa c(w)\}$, where $\gamma_\kappa = (\alpha_\kappa, \beta_\kappa)$; for identifiability we set $\alpha_0 = \beta_0 = 0$. The loglikelihood is thus

$$(5) \quad \ell(\gamma, H_0) = \sum_{i=1}^{n} \log(p_i) + \sum_{k=1}^{K-1} \sum_{j=1}^{n_k} \{\alpha_k + \beta_k c(w_{kj})\}.$$

Empirical likelihood estimation of the spectral density ratio model involves maximizing ℓ with respect to p_i, for a fixed γ, subject to the empirical versions of constraints (4), conveniently rewritten as

$$p_i \geq 0, \qquad v_i \in \Delta_D$$

$$\sum_{i=1}^{n} p_i = 1, \qquad \sum_{i=1}^{n} p_i \{v_i - D^{-1}\mathbf{1}_D\} = 0,$$

$$(6) \quad \sum_{i=1}^{n} p_i\{g(v_i, \gamma_1) - 1\} = 0, \qquad \sum_{i=1}^{n} p_i\{v_i g(v_i, \gamma_1) - D^{-1}\mathbf{1}_D\} = 0,$$

$$\vdots \qquad\qquad \vdots$$

$$\sum_{i=1}^{n} p_i\{g(v_i, \gamma_{K-1}) - 1\} = 0, \qquad \sum_{i=1}^{n} p_i\{v_i g(v_i, \gamma_{K-1}) - D^{-1}\mathbf{1}_D\} = 0.$$

Using an approach similar to that of Qin and Lawless (1994), it can be shown that if we use Lagrange multiplier procedures to profile p_i with the normalization and marginal moment constraints (6), then the jump size for the baseline spectral distribution function can be written as

$$(7) \quad p_i = \frac{1}{n_0} \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}}\mathcal{M}(v_i, \gamma)},$$

where $\rho_\kappa = n_\kappa/n_0$,

$$(8) \quad \mathcal{M}(v, \gamma) = \mathcal{G}(v, \gamma) \otimes (v - D^{-1}\mathbf{1}_D),$$

and $\mathcal{G}(v,\gamma) = (1, g_1(v,\gamma_1), \ldots, g_{K-1}(v,\gamma_{K-1}))^{\mathrm{T}}$. Here and below, $\delta = (\delta_0, \ldots, \delta_{K-1})^{\mathrm{T}}$ denotes the Lagrange multipliers corresponding to the marginal moment constraints, which are determined through the conditions

$$(9) \qquad \frac{1}{n_0} \sum_{i=1}^{n} \frac{v_i g(v_i, \gamma_\kappa) - D^{-1}\mathbf{1}_D}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}}\mathcal{M}(v_i, \gamma)} = 0, \quad \kappa = 0, \ldots, K-1.$$

Thus apart from a constant the profiled empirical loglikelihood for γ can be written as

$$(10) \quad \ell_{\mathrm{p}}(\gamma) = -\sum_{i=1}^{n} \log\left\{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}}\mathcal{M}(v_i, \gamma)\right\} + \sum_{k=0}^{K-1} \sum_{j=0}^{n_k} \{\alpha_k + \beta_k c(w_{kj})\},$$

and so the semiparametric empirical likelihood estimator $\widehat{\gamma} = \arg\max_\gamma \ell_{\mathrm{p}}(\gamma)$ of the spectral density ratio model can be obtained by combining (9) with the score equations

$$(11) \quad \begin{aligned} \frac{\partial \ell_{\mathrm{p}}}{\partial \alpha_\kappa} &= -\sum_{i=1}^{n} \frac{\rho_\kappa g(v_i, \gamma_\kappa) + \delta^{\mathrm{T}} M_{\alpha_\kappa}(v_i, \gamma)}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}}\mathcal{M}(v_i, \gamma)} + n_\kappa = 0, \\ \frac{\partial \ell_{\mathrm{p}}}{\partial \beta_\kappa} &= -\sum_{i=1}^{n} \frac{\rho_\kappa c(w_{\kappa j}) g(v_i, \gamma_\kappa) + \delta^{\mathrm{T}} M_{\beta_\kappa}(v_i, \gamma)}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \gamma_k) + \delta^{\mathrm{T}}\mathcal{M}(v_i, \gamma)} + \sum_{j=0}^{n_\kappa} c(w_{\kappa j}) = 0, \quad \kappa = 0, \ldots, K-1. \end{aligned}$$

Here, M_{α_κ} and M_{β_κ} represent the partial derivatives of the moment estimating function (8). The existence of the semiparametric likelihood estimates $\widehat{\gamma}$ is a corollary of Lemma 1 in Qin and Lawless (1994) which states that under suitable regularity conditions the semiparametric empirical likelihood estimator lies in the interior of the ball $||\gamma - \gamma_o|| \le n^{-1/3}$, and thus is centred on the true value γ_o, as $n \to \infty$.

On using the estimates obtained from the score equations (11), the maximum likelihood estimator of the size of the jump of the baseline spectral density function turns out to be

$$\widehat{p_i} = \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \widehat{\gamma_k}) + \widehat{\delta}^{\mathrm{T}}\mathcal{M}(v_i, \widehat{\gamma})},$$

so the maximum likelihood estimator of the baseline spectral distribution function is

$$(12) \quad \begin{aligned} \widehat{H}_0(w) &= \sum_{i=1}^{n} \widehat{p_i} I(v_i \le w) \\ &= \frac{1}{n_0} \sum_{i=1}^{n} \frac{1}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \widehat{\gamma_k}) + \widehat{\delta}^{\mathrm{T}}\mathcal{M}(v_i, \widehat{\gamma})} I(v_i \le w), \end{aligned}$$

and for $\kappa = 1, \ldots, K-1$ the other spectral spectral distribution functions are estimated as

$$(13) \quad \begin{aligned} \widehat{H}_\kappa(w) &= \sum_{i=1}^{n} \widehat{p_i} g(v_i, \widehat{\gamma_k}) I(v_i \le w) \\ &= \frac{1}{n_0} \sum_{i=1}^{n} \frac{g(v_i, \widehat{\gamma_k})}{1 + \sum_{k=1}^{K-1} \rho_k g(v_i, \widehat{\gamma_k}) + \widehat{\delta}^{\mathrm{T}}\mathcal{M}(v_i, \widehat{\gamma})} I(v_i \le w). \end{aligned}$$

3 Numerical Experiments

3.1 A Two-Sample Bivariate Spectral Density Ratio Model

To model the spectral density corresponding to each subpopulation, we consider two symmetric Beta distributions,

$$(14) \quad \begin{cases} \mathrm{d}H_0(w) & = \frac{1}{B(\phi_0)} w^{\phi_0-1}(1-w)^{\phi_0-1}\mathrm{d}w, \quad \phi_0 > 0, \\ \mathrm{d}H_1(w) & = \frac{1}{B(\phi_1)} w^{\phi_1-1}(1-w)^{\phi_1-1}\mathrm{d}w, \quad \phi_1 > 0, \end{cases}$$

where $B(\phi) = \int_0^1 \{u(1-u)\}^{\phi-1}\mathrm{d}u$. The mean for each spectral density equals $1/2$, so that constraints (4) are satisfied. Using the distortion function $c(w) = \log\{w(1-w)\}$ we can rewrite (14) as

$$\begin{cases} \mathrm{d}H_0(w) & = \exp\{a_0 + b_0 c(w)\}\mathrm{d}w, \\ \mathrm{d}H_1(w) & = \exp\{a_1 + b_1 c(w)\}\mathrm{d}w, \\ (a_1, b_1) & = (-\log B(\phi_1), \phi_1 - 1), \\ (a_0, b_0) & = (-\log B(\phi_0), \phi_0 - 1). \end{cases}$$

Hence, making use of (14), we obtain the following spectral density representation of the K-sample bivariate extreme value beta model,

$$(15) \quad \frac{\mathrm{d}H_1(w)}{\mathrm{d}H_0(w)} = \exp\left\{\alpha_1 + \beta_1 c(w)\right\},$$

where the tilt parameters are

$$(16) \quad \begin{cases} (\alpha_1, \beta_1) & = (\log\{B(\phi_0)/B(\phi_1)\}, \phi_1 - \phi_0), \\ (\alpha_0, \beta_0) & = (0, 0) \end{cases}$$

Observe that this model is closed, in the sense that tilting always yields a symmetric Beta distribution.

3.2 Numerical Results

We now report computational experience with the method described above. We consider two scenarios: in A the parameters are $(\phi_0^A, \phi_1^A) = (0.10, 0.30)$, so that $(\alpha_1^A, \beta_1^A) = (1.18, 0.20)$; in B the parameters are $(\phi_0^B, \phi_1^B) = (0.90, 2.00)$, so that $(\alpha_1^B, \beta_1^B) = (2.00, 1.10)$. Given the computational cost of obtaining full optimisation estimates, in practice we have found it best to use a two-step strategy wherein one maximizes the unconstrained outer objective function with respect to γ, and minimizes the inner dual problem with respect to the nuisance parameter δ. For large n the inner dual problem not only has much lower dimensionality than the corresponding primal problem, but it also has the advantage of being subject to a set of linear constraints that can be removed by using a pseudo-logarithmic function (Owen, 2001, p. 62). In Figures 1 and 2 we plot instances of our numerical experiments, and to assess the possibility of borrowing strength from a larger number of extremes on one of the samples, in both cases we used an unbalanced design and simulated data with $n_0 = 50$ and $n_1 = 100$.

Baseline Sample (k=0)

Sample 2 (k=1)

Baseline Sample (k=0)

Sample 2 (k=1)
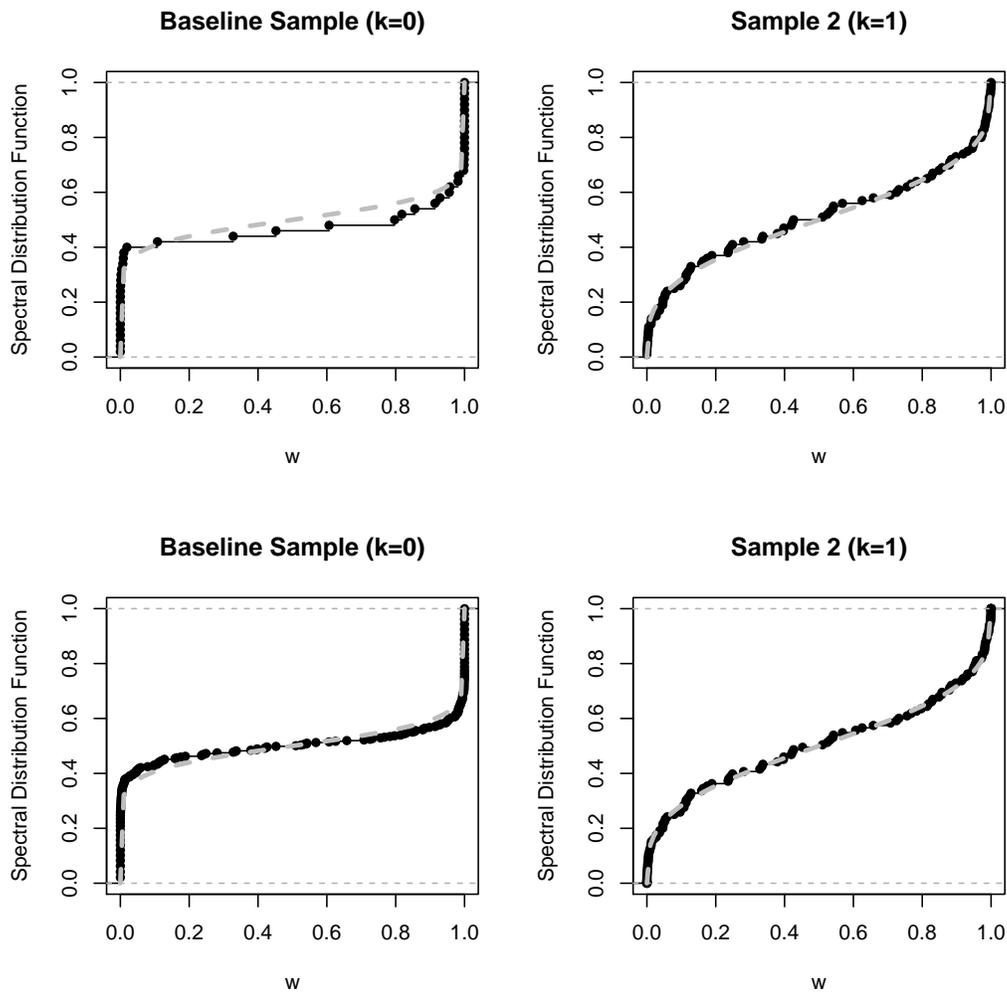
Figure 1: Scenario A—spectral distribution functions estimated by empirical likelihood (above) and using the spectral density ratio model (below); the dashed lines represent the true distribution functions.
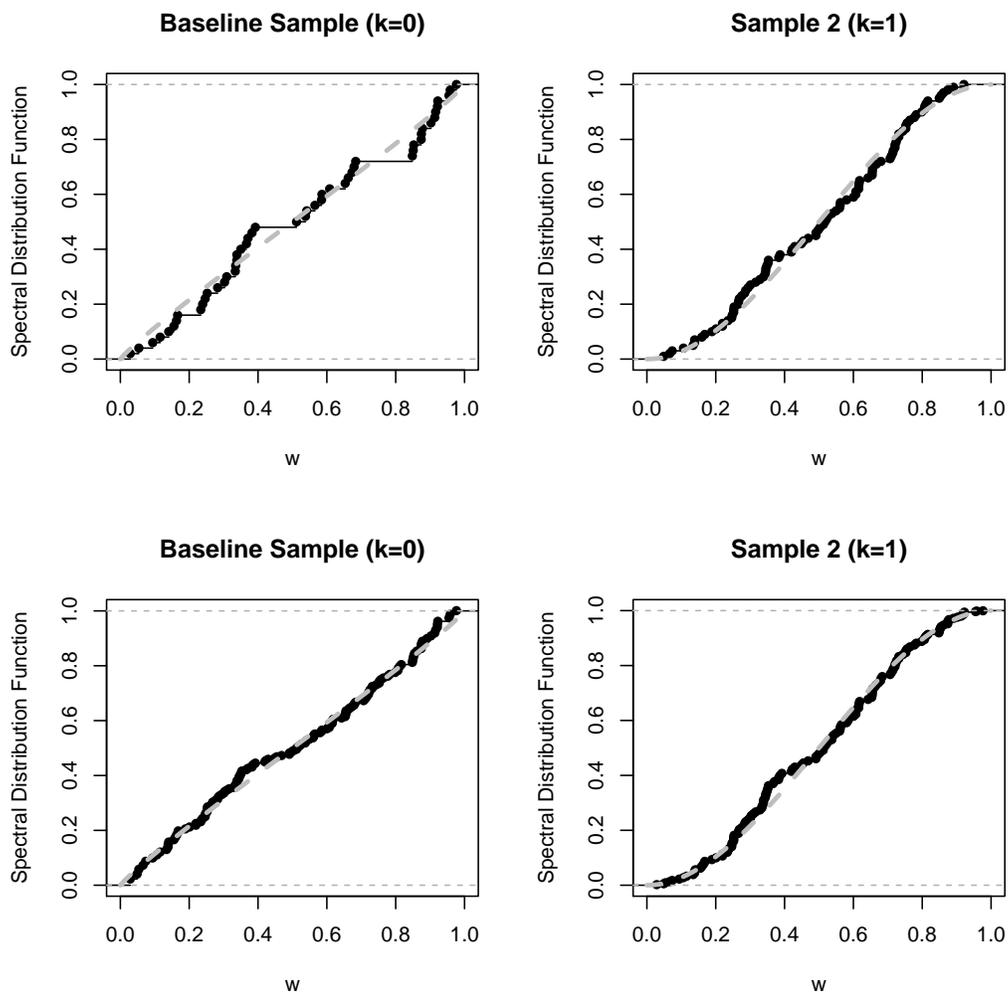
Figure 2: Scenario B—spectral distribution functions estimated by empirical likelihood (above) and using the spectral density ratio model (below); the dashed lines represent the true distribution functions.

REFERENCES (RÉFERENCES)

[1] BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004) *Statistics of Extremes: Theory and Applications.* New York: Wiley.

[2] BOLDI, M.-O. and DAVISON, A. C. (2007) A mixture model for multivariate extremes. *J. R. Statist. Soc. B*, **69**, 217–229.

[3] CASTILLO, E., HADI, A., ALEGRÍA, J. and BALAKRISHNAN, N. (2005) Extreme value and related models with applications in engineering and science. Wiley: New York.

[4] CHENG, J., QIN, J. and ZHANG, B. (2009) Semiparametric estimation and inference for distributional and general treatment effects. *J. R. Statist. Soc. B*, **71**, 881–904.

[5] COLES, S. G. (2001) *An Introduction to Statistical Modelling of Extreme Values.* London: Springer.

[6] COLES, S. G. and TAWN, J. A. (1991) Modelling extreme multivariate events. *J. R. Statist. Soc. B*, **53**, 377–392.

[7] COOLEY, D., DAVIS, R. A. and NAVEAU, P. (2010) The pairwise Beta distribution: A flexible parametric multivariate model for extremes. *J. Multivariate Anal.*, **101**, 2103–2117.

[8] DE CARVALHO, M. and DAVISON, A. C. (2011) Spectral density ratio models for multivariate extremes. Mimeo.

[9] DUPUIS, D. J. (2005) Ozone concentrations: a robust analysis of multivariate extremes. *Technometrics*, **47**, 191–201.

[10] EFRON, B. and TIBSHIRANI, R. (1996) Using specially designed exponential families for density estimation. *Ann. Statist.*, **24**, 2431–2461.

[11] EINMAHL, J. H. J. and SEGERS, J. (2009) Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.*, **37**, 2953–2989.

[12] FOKIANOS, K. (2004) Merging information for semiparametric density estimation. *J. R. Statist. Soc. B*, **66**, 941–958.

[13] FOKIANOS, K., KEDEM, B., QIN, J. and SHORT, D. A. (2001) A semiparametric approach to the one-way layout. *Technometrics*, **43**, 56–65.

[14] GILBERT, P. B., LELE, S. R. and VARDI, Y. (1999) Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, **10**, 839–866.

[15] GUDENDORF, G. and SEGERS, J. (2011) Nonparametric estimation of an extreme-value copula in arbitrary dimensions. *J. Multivariate Anal.*, **102**, 37–47.

[16] KEDEM, B., KIM, E.-Y., VOULGARAKI, A. and GRAUBARD, B. I. (2009) Two-dimensional semiparametric density ratio modeling of testicular germ cell data. *Stat. Med.*, **28**, 2147–2159.

[17] KOTZ, S. and NADARAJAH, S. (2000) *Extreme Value Distributions: Theory and Applications.* London: Imperial College Press.

[18] MIKOSCH, T. (2005) How to model multivariate extremes if one must? *Statist. Neerl.*, **59**, 324–338.

[19] OWEN, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

[20] OWEN, A. (2001) *Empirical Likelihood.* Boca Raton: Chapman and Hall.

[21] PICKANDS, J. (1981) Multivariate extreme value distributions. *Proc. 43rd Sess. Int. Statist. Inst.*, 859–878.

[22] QIN, J. and LAWLESS, J. F. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300–325.

[23] QIN, J. and ZHANG, B. (1997) A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, **84**, 609–618.

[24] RAMOS, A. and LEDFORD, A. (2009) A new class of models for bivariate joint tails. *J. R. Statist. Soc. B*, **71**, 219–241.

[25] STEPHENSON, A. (2009) High-dimensional parametric modelling of multivariate extreme events. *Aust. N. Z. J. Stat.*, **51**, 77–88.

[26] TAWN, J. A. (1988) Bivariate extreme value theory: models and estimation. *Biometrika*, **75**, 397–415.

[27] ZHANG, D., WELLS, M. T. and PENG, L. (2008) Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *J. Multivariate Anal.*, **99**, 577–588.