Biometrika (2013), –, -, *pp*. 1–3

Printed in Great Britain

Supplementary material for "A sum characterization of hidden regular variation with likelihood inference via expectation–maximization"

BY GRANT B. WELLER

Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, U.S.A.

grantw@andrew.cmu.edu

AND DANIEL COOLEY

Department of Statistics, Colorado State University, Fort Collins, Colorado 80523-1877, U.S.A.

cooleyd@stat.colostate.edu

1. PROOF OF TAIL EQUIVALENCE ON $\mathfrak{C}$

With Y and V defined in § 3, we adapt Lemma 3·12 of Jessen & Mikosch (2006) to show tail equivalence on the full cone $\mathfrak{C}$. Consider a relatively compact rectangle $A \in \mathfrak{C}$; that is, A is bounded away from 0. This class of sets A generates vague convergence in $\mathfrak{C}$ (Resnick, 2007, Lemma 6.1); thus it is sufficient to show

$$\lim_{t\to\infty} t \text{ pr}\left\{\frac{Y+V}{b(t)} \in A\right\} = \lim_{t\to\infty} t \text{ pr}\left\{\frac{Z}{b(t)} \in A\right\} = \nu(A).$$

Without loss of generality let $A = [a, b] = \{x \in \mathfrak{C} : a \leq x \leq b\}$. For small $\epsilon > 0$, define $a^{-\epsilon} = (\max\{0, a_1 - \epsilon\}, \ldots, \max\{0, a_d - \epsilon\})^{\mathrm{T}}$, and define $b^{-\epsilon}$ analogously. Define the rectangles $A^{-\epsilon} = [a^{-\epsilon}, b]$ and $A^{\epsilon} = [a, b^{-\epsilon}]$. For small ϵ, the rectangles A^{ϵ} and $A^{-\epsilon}$ are relatively compact in $\mathfrak{C}$, and $A^{\epsilon} \subset A \subset A^{-\epsilon}$. Note that $\nu(\partial A) = 0$ (Jessen & Mikosch, 2006); there is no mass on the edges of A.

For small $\epsilon > 0$ and fixed $t > 0$,

$$\begin{aligned}\text{pr}\left\{\frac{Y+V}{b(t)} \in A\right\} &= \text{pr}\left\{\frac{Y+V}{b(t)} \in A, \frac{\|V\|}{b(t)} > \epsilon\right\} + \text{pr}\left\{\frac{Y+V}{b(t)} \in A, \frac{\|V\|}{b(t)} \leq \epsilon\right\} \\ &\leq \text{pr}\left\{\|V\| > b(t)\epsilon\right\} + \text{pr}\left\{\frac{Y}{b(t)} \in A^{-\epsilon}\right\}.\end{aligned}$$

Thus

$$\begin{aligned}\limsup_{t\to\infty} t \text{ pr}\left\{\frac{Y+V}{b(t)} \in A\right\} &\leq \limsup_{t\to\infty} t \text{ pr}\left\{\|V\| > b(t)\epsilon\right\} + \limsup_{t\to\infty} t \text{ pr}\left\{\frac{Y}{b(t)} \in A^{-\epsilon}\right\} \\ &= \nu(A^{-\epsilon}) \searrow \nu(A), \ \epsilon \to 0,\end{aligned}$$

since $\alpha^* > \alpha$ by assumption. For the lower bound, recognize

$$\text{pr}\left\{\frac{Y+V}{b(t)} \in A\right\} \geq \text{pr}\left\{\frac{Y}{b(t)} \in A^{\epsilon}, \frac{\|V\|}{b(t)} \leq \epsilon\right\} \geq \text{pr}\left\{\frac{Y}{b(t)} \in A^{\epsilon}\right\} - \text{pr}\left\{\|V\| > b(t)\epsilon\right\},$$

and so

$$\liminf_{t\to\infty} t \operatorname{pr}\left\{\frac{Y+V}{b(t)} \in A\right\} \geq \liminf_{t\to\infty} t \operatorname{pr}\left\{\frac{Y}{b(t)} \in A^{\epsilon}\right\} - \liminf_{t\to\infty} t \operatorname{pr}\{\|V\| > b(t)\epsilon\}$$
$$= \nu(A^{\epsilon}) \nearrow \nu(A),\ \epsilon \to 0.$$

Collecting the upper and lower bounds, and using the fact that A is a ν-continuity set, we see that

$$t \operatorname{pr}\left\{\frac{Y+V}{b(t)} \in \cdot\right\} \longrightarrow \nu(\cdot),\, t \to \infty,$$

vaguely in $M_+(\mathfrak{C})$.

2. FINITE HIDDEN MEASURE SIMULATION STUDY

We apply the proposed methodology to simulated data of dimension $d = 2$ which exhibit asymptotic independence and hidden regular variation with finite hidden angular measure. We generate n independent realizations of $Z_s = Y_s + V_s$, where $Y_s = [RW, R(1-W)]^{\mathrm{T}}$ and $V_s = [R_0 W_0, R_0(1-W_0)]^{\mathrm{T}}$, with

$$1 - F_R(r) = 2/r,\ r > 2 \qquad W \sim Bernoulli(1/2)$$
$$1 - F_{R_0}(r) = r^{-1/\eta},\ r > 1 \qquad W_0 \sim H_0(\cdot),$$

all mutually independent, with H_0 the integrated measure density associated with the bivariate logistic dependence model (Gumbel, 1960). The angular density of this dependence model is

$$h_0(w;\beta) = \frac{1}{2}\left(\frac{1}{\beta} - 1\right)\{w(1-w)\}^{-1-1/\beta}\{w^{-1/\beta} + (1-w)^{-1/\beta}\}^{\beta-2},\, \beta \in (0,1).$$

As $\beta \to 1$, h_0 degenerates to point masses at $w = 0$ and $w = 1$, while the limiting case $\beta \to 0$ corresponds to a single point mass at $w = 1/2$. We set $\eta = 0{\cdot}75$ and assume it is known, and we aim to estimate β via the proposed expectation–maximization procedure.

While the full density of Z_s could be written as a convolution in this case, we aim to study the effects of misspecification of the model for non-extreme realizations of Y_s and V_s. Although the true radial component densities of Y_s and V_s follow Pareto distributions, here we let g_Y and g_V be densities associated with the same angular component models for Y_s and V_s, but Fréchet distributed radial components $\|Y\|$ and $\|V\|$ with scale parameters 2, 1 and shape parameters 1, $1/\eta$, respectively. These densities differ from true densities of Y_s and V_s for small $\|y\|$ and $\|v\|$ but rapidly converge to these as the magnitude grows, despite having differing supports.

To illustrate the performance of the proposed estimation scheme over different parameter values, threshold settings, and sample sizes, we perform the estimation scheme on 500 replications of n realizations of Z_s, with three different settings, shown in Table 1. The dependence parameters chosen signify moderate, weak, and strong hidden tail dependence, respectively, while the thresholds chosen correspond to the 0·9, 0·95, and 0·99 theoretical quantiles of the imposed Fréchet distributions of radial components. In each case, we choose an initial value of $m = 250$ for the number of Monte Carlo replications, and implement the scheme of Booth & Hobert (1999) for increasing m as described in § 5.3 with $\alpha = 0{\cdot}25$ and $r = 3$. We determine the algorithm has converged when the convergence criterion criterion (20) is met for three successive iterations, with $\delta_1 = 0{\cdot}001$ and $\delta_2 = 0{\cdot}002$. Initial values $\beta^{(0)} \sim U$, with U following a uniform distribution centered at β and of widths 0·5, 0·4, and 0·3 for settings 1, 2, and 3, respectively.

Table 1. *Sample sizes, true values (×100), selected thresholds, mean parameter estimates (×100), root mean square errors (×100),* 95% *confidence interval coverage rates, and median number of iterations for estimation procedure applied to 500 repetitions of simulated data*

Setting	n	β	r_Y^*	r_V^*	$\hat{\beta}$	RMSE	Coverage	median(k)
1	2500	50	19·0	5·4	52·0	3·1	94·4	20
2	5000	70	39·0	9·3	67·0	3·3	77·6	12
3	10000	25	199·0	31·5	28·8	6·2	70·2	18

Simulations were performed using R on the Lynx computing system at the National Center for Atmospheric Research.

Table 1 shows mean parameter estimates, their root mean square errors, coverage rates of 95% confidence intervals constructed via a normal approximation, and the median number of iterations needed to obtain convergence for each simulation scenario. In each case, the algorithm converged relatively quickly, with median number of iterations of 20, 12, and 18, respectively. We note the bias in the estimates of β from the estimation procedure due to the misspecification of the model, which is largest in setting 3, which corresponds to strong tail dependence in the V_s component. Further examination found that this bias is most severe when β is close to 0 or 1; that is, near the limiting degenerate cases. This bias was reduced by choosing a higher threshold; however, in small samples relatively low thresholds must be chosen to reduce uncertainty. Confidence intervals constructed via Louis (1982) were somewhat anticonservative in all cases, with coverage rates decreasing as the bias increases.

References

Booth, J. G. & Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *J. R. Stat. Soc.* B **61**, 265–285.

Gumbel, E. J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris* **9**, 171–173.

Jessen, A. H. & Mikosch, T. (2006). Regularly varying functions. *Publ. Inst. Math.* **80**, 171–192.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Stat. Soc.* B **44**, 226–233.

Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.