Statistical Theory of Extremes

Homepage: http://www.gathacognition.com/book/gcb14
http://dx.doi.org/10.21523/gcb1

Part 3
Multivariate Extremes

Exercises

J. Tiago de Oliveira

Academia das Ciências de Lisboa (Lisbon Academy of Sciences), Lisbon, Portugal.

Editor(s)	**Published Online**
J.C. Tiago de Oliveira	23 June 2017

3.1. Obtain the inequality $\frac{\max(1, e^w)}{1 + e^w} \leq k(w) \leq 1$ or $\max(u, 1-u) \leq A(u) \leq 1$ from the Boole-Fréchet inequality and a passage to the limit.

3.2. Show that the sets $\{k(w)\}$ or $\{A(u)\}$ are convex, closed and symmetrical (corresponding to exchangeability $k(w) = k(-w)$ or $A(u) = A(1-u)$.

3.3. Obtain the conditions that $k(w)$ or $A(u)$ must satisfy in the absolutely continuous case.

3.4. The equidistribution median curve for the case with reduced Gumbel margins is between the curves $e^{-x} + e^{-y} = \log 2$ and $\max(e^{-x}, e^{-y}) = \log 2$; are these curves convex?: study the corresponding situation for the equisurvival median curve (case of $A(u)$).

3.5. For bivariate distributions attracted for maxima to $\Lambda(x, y)$, show that if there is negative dependence (i.e., $F(x, y) \leq F(x, +\infty)\, F(+\infty, y)$) then $\Lambda(x, y) = \Lambda(x)\, \Lambda(y)$ (*asymptotic attraction to independence*).

3.6. We can expect that for bivariate samples $(\max_1^n X_i, \max_1^n Y_i)$ are asymptotically independent. Obtain conditions for this, but show that in the singular case where $\text{Prob}\{X + Y = a\} = 1$this is not the case.

Originally published in 'Statistical Analysis of Extremes', 1997, 2016
http://dx.doi.org/10.21523/gcb1.1721

3.7. Recall that $\min_1^n Y_i = -\max_1^n (-Y_i)$, and use Sibuya's necessary and sufficient condition or Geffroy's sufficient condition for independence in maxima.

3.8. In the characterization of $A(u)$, show that if $A(0) = A(1) = 1$ and $A(u)$ convex in $[0,1]$ the conditions

a) $0 \leq A'(0), A' \leq 1$,

b) $\max(u, 1-u) \leq A(u) \leq 1$ and

c) $A(u)/(1-u)$ non-decreasing and $A(u)/u$ non-increasing are equivalent. Show also that $A'(0) \leq A'(u) \leq A'(1), A(u) \geq 1 + A'(0)u$, and $A(u) \geq 1 - A'(1)\,(1-u)$.

3.9. Obtain the relation between the function $P(u, v)$ defined by Sibuya and the structure function $\overline{S}(\xi, \eta)$. Convert the results of Sibuya on $P(u, v)$ concerning asymptotic independence and asymptotic complete dependence (diagonal case) to $\overline{S}(\xi, \eta)$.

3.10. The condition of $k''(w)$, for the bivariate distribution function $\Lambda(x, y)$, implies that $\varphi(w) = (1 + e^{w})k(w)$ and $\Psi(w) = (1 + e^{-w})k(w)$ are such that $\varphi''(w) \geq \varphi'(w)$ and so $\varphi'(w) \geq 0$ and $\Psi''(w) + \Psi'(w) \geq 0$ and so $\Psi'(w) \leq 0$, which eliminates the two monotonicity conditions.

3.11. Show that if (X, Y) has the distribution function $\Lambda(x, y)$ then the pair (V, W) with $e^{-V} = e^{-X} + e^{-Y}, W = Y - X$ has the distribution function

$$\text{Prob}\,\{V \leq v, W \leq w\} = \Lambda(v) + \left(\frac{k'(w)}{k(w)} + \frac{1-e^{-w}}{1+e^{-w}}\right) \Lambda(v - \log k(w))$$

$$+ \int_{-\infty}^{w} \frac{e^{-\rho}}{(1+e^{-\rho})^2} (e^{-v}k(\rho) - 1)\, \Lambda(v - \log k(\rho))\, d\rho.$$

3.12. For the distribution function $D(w)$, show that $D^a(w + b)$, if $a \geq 1$, is also a $D(w)$- function for convenient b.

3.13. Show that if (X, Y) is a bivariate pair with reduced Gumbel margins, and so distribution function $\Lambda(x, y)$ then

$$\text{Prob}\{X \leq x, Y - X \leq w\} = D(w)\, \Lambda(x, x + w).$$

3.14. Translate maxima results to minima results and vice-versa; use, in particular, the relation between the distribution function and the survival function.

3.15. Show that $\Lambda(w+\gamma)$ and $\frac{w+\theta}{2\,\theta}$ for $0 \leq \theta \leq 2, |w| < \theta$ are $D(w)$ -functions.

3.16. Suppose that (X, Y) is an independent random pair with Gumbel margins whose parameters are (λ_X, δ) and (λ_Y, δ) (the same dispersion parameter). Compute $\text{Prob}\{Y > X\}$. Solve the corresponding question for minima with exponential margins and compare the results.

3.17. Compute $k(.)$ for the use of max-technique; what is the corresponding technique if $S(x, y)$ is used? Interpret the statistical implications for small samples of the fact that the index of dependence is $\leq 1/4$; what does this imply for the discrimination between models?

3.18. Show that for general bivariate distribution of maxima with reduced Gumbel margins $\text{Prob}\{X \leq Y\} = \frac{1}{2} - \frac{k'(0)}{k(0)}$ so $|\frac{k'(0)}{k(0)}| \leq 1/2$ and that for minima with standard exponential margins we have, correspondingly, $\text{Prob}\{X' \leq Y'\} = \frac{1}{2} - \frac{A'(1/2)}{4\,A(1/2)}$ and so $|\frac{A'(1/2)}{A(1/2)}| \leq 2$.

3.19. Let $G_1(x)$ and $G_2(y)$ be distribution function such that $G_1^k(\lambda_k + \delta_k\, x) \to \Lambda(x)$ and $G_2^k(\lambda_k^{'} + \delta_k^{'} y) \to \Lambda(y)$. As stated, we know by Fréchet inequalities, that any bivariate distribution function $F(x, y)$ such that $F(x, +\infty) = G_1(x), F(+\infty, y) = G_2(y)$ satisfies the inequalities $\underline{D}(x, y) = \max(0, G_1(x) + G_2(y) - 1) \leq F(x, y) \leq$ $\min\big(G_1(x), G_2(y)\big) = \overline{D}(x, y)$. Show that the family of distribution functions $\{\theta\, \underline{D}(x, y) + (1-\theta)\, \overline{D}(x, y)\}$ is attracted, with the same attraction coefficients for the margins, to $\exp\{-\,[\theta\,(e^{-x} + e^{-y}) + (1-\theta)\max(e^{-x}, e^{-y})]\}$, which is the Gumbel model.

3.20. Show that the bivariate model of maxima with Weibull margins and shape parameter $\alpha = 1 (\xi = -e^{-x}, \eta = -e^{-y})$ is $\Psi_1(\xi, \eta) = \exp\{(\xi + \eta) k(\exp\frac{\xi}{\eta})\}$ and the corresponding logistic is $\Psi_{1L}(\xi, \eta) = \exp\{((-\,\xi)^{1/(1-\theta)} + (-\eta)^{1/(1-\theta)}\,)^{1-\theta}\}$.

3.21. Show that if we have a reduced sample $\{(x_i, y_i), i = 1, 2, \ldots, n\}$, and we expect to observe x, the MSE linear predictor of $y (y^* = \alpha + \beta\, x + \sum_1^n \varphi_i\; x_i + \sum_1^n \Psi_i\, y_i)$ is given by the usual linear regression $L_y(x) = \gamma + \rho(x - \gamma)$, the sample having an intervention only through the computation of $\rho^* = r$, as usual.

3.22. For bivariate extremes with reduced Gumbel margins, the linear regression line of y in x is $L_y(x) = \gamma + \rho(x - \gamma)$. Show that all regression lines are in the area: if $x \leq \gamma$ then $x \leq L_y(x) \leq \gamma$ and if $x \geq \gamma$ then $\gamma \leq L_y(x) \leq x$.

3.23. Show that, as all regression lines are in the area described above, statistical choice of bivariate model through linear regression is difficult even for moderate sample sizes.

3.24. Draw the graphs of $A(u|\theta)$ for different models (and values of $\theta = 0., .5$ and 1.0). They are all in the triangle bounded by $\max(u, 1-u) \leq A(u|\theta) \leq 1$. If they are not close to independence or to the diagonal case it is difficult to distinguish them.

3.25. Using the explicit expression of the correlation coefficients, show that: for the logistic model: $\rho(\theta) = \theta(2-\theta), v(\theta) = 2^{-2^{1-\theta}} - 1$;

for the mixed model:

$\rho(\theta) = \frac{6}{\pi^2}\,(\arccos(1-\theta/2))^2 = \frac{24}{\pi^2}(\operatorname{arctg}\sqrt{\frac{\theta}{4-\theta}})^2, v(\theta) = 2^{\theta/2} - 1$;

for the Gumbel model: $\rho(\theta) = \frac{12}{\pi^2}\int_0^\theta \frac{\log(2-t)}{1-t}dt, v(\theta) = 2^\theta - 1$;

for the biextremal model: $\rho(\theta) = -\frac{6}{\pi^2}\int_0^\theta \frac{\log t}{1-t}dt, v(\theta) = 2^\theta - 1$.

3.26. Analyse the data of Table 1, contained in Gumbel and Goldstein (1964):

Test the fit of the Gumbel distribution to each set of oldest ages of death (men, women) and, after that, test for (the expected) independence.

Add to Table 1 the data of Table 2, given in Fransén and Tiago de Oliveira (1984).

Do the same tests as before and compare the estimators. Also compare the estimators of the margins for Table 1 and Tables 1 and 2. Use the previous result to estimate the probability that women have greater ages of death than men (take δ^*as the average of the δ's for Tables 1 and 2) and test if women "live longer" than men.

Table 1. Oldest ages at death, Sweden

Year	Men	Women	Year	Men	Women
1905	100.88	102.54	1932	102.55	104.87
1906	101.17	106.13	1933	103.17	105.98
1907	104.65	103.46	1934	103.98	103.38
1908	105.12	102.12	1935	106.09	105.32
1909	102.57	101.69	1936	103.43	103.77
1910	101.70	102.92	1937	105.72	105.86
1911	100.49	102.78	1938	103.24	104.27
1912	100.90	106.15	1939	103.25	105.45
1913	103.06	105.02	1940	103.40	105.71
1914	102.63	103.56	1941	101.66	106.15
1915	102.69	106.52	1942	106.48	104.71
1916	100.82	101.50	1943	101.26	103.83
1917	102.52	104.01	1944	105.12	105.19
1918	100.08	105.01	1945	104.88	105.03
1919	101.67	104.52	1946	102.41	105.88
1920	101.41	103.94	1947	104.22	107.49
1921	101.76	103.14	1948	102.88	105.83
1922	102.57	104.33	1949	103.57	103.41
1923	101.63	102.32	1950	105.12	105.64
1924	103.47	103.56	1951	103.80	103.53
1925	105.48	103.86	1952	102.94	107.90
1926	104.01	105.87	1953	103.00	104.42
1927	105.83	103.31	1954	106.50	104.85
1928	105.00	104.37	1955	103.36	103.97
1929	102.78	102.72	1956	103.15	107.89
1930	102.61	105.01	1957	102.54	104.46
1931	105.55	104.40	1958	104.92	104.12

Table 2. Oldest ages at death, Sweden

Year	Men	Women	Year	Men	Women
1959	104.23	104.77	1965	105.28	104.31
1960	103.59	106.13	1966	104.93	104.98
1961	103.74	107.10	1967	105.27	105.83
1962	103.00	104.56	1968	105.92	106.35
1963	104.25	110.07	1969	101.81	107.58
1964	104.12	106.15	1970	104.02	105.42

3.27. Consider any river at two points with a sample of $k \geq 50$ yearly floods. Check if the margins are Weibull, Gumbel or Fréchet distributed. If they are not Gumbel distributed, make the necessary transformations to be so, and then:

- Estimate (λ, δ) for both points;
- Choose the bivariate model for the "estimated" reduced margins (exactly or approximately).

3.28. Obtain subroutines for statistical choice of bivariate maxima models with Gumbel margins or for bivariate minima with exponential margins.

3.29. Show that the "estimated" best linear predictor $\frac{y^*-\hat{\lambda}_y}{\hat{\delta}_y} = \gamma + r.\frac{x-\hat{\lambda}_x}{\hat{\delta}_x} - \gamma)$ is to be expected very close to $\frac{y^*-\bar{y}}{s_y} = r \cdot (\frac{x-\bar{x}}{s_x})$ as

$$\bar{x} - \hat{\lambda}_x - \gamma\,\hat{\delta}_x \approx 0, \bar{y} - \hat{\lambda}_y - \gamma\,\hat{\delta}_y \approx 0 \text{ and } \hat{\delta}_y/\hat{\delta}_x \approx s_y/s_x.$$

3.30. Study the estimator $\tilde{\theta}$ of θ in the biextremal case with Gumbel margins (moments, convergence, asymptotic distribution, etc.).

3.31. Show that from the sample distribution function of the $w_i, D^*(w) = \frac{1}{n}\sum_1^n H(w - w_i) \stackrel{P}{\rightarrow} D(w)$ we can obtain estimators $k^*(w)$ of $k(w)$ and $D^*(w)$ of $D(w)$, both being non-parametric but not intrinsic, which is a more restrictive condition. It is easy to prove this by showing that $A(u)$ is non-convex (using exponential margins).

3.32. Consider the trivariate model with Gumbel margins

$$\Lambda(x_1, x_2, x_3) = \exp\{-(e^{-x_1} + e^{-x_2} + e^{-x_3}) + \theta_3 \min(e^{-x_1}, e^{-x_2})$$

$$+\theta_2 \min(e^{-x_1}, e^{-x_3}) + \theta_1 \min(e^{-x_2}, e^{-x_3}) - \tau \min(e^{-x_1}, e^{-x_2}, e^{-x_3})\}.$$

Using the inequalities for the margins, show that

$0 \leq \theta_1, \theta_2, \theta_3 \leq 1$ and $\max(0, \theta_1 + \theta_2 + \theta_3 - 1) \leq \tau \leq (\theta_1 + \theta_2 + \theta_3)/2$ and so $\theta_1 + \theta_2 + \theta_3 \leq 2$.

References

Fransén, A. and Tiago De Oliveira, J., 1984. Statistical choice of univariate extreme models. Part II, *Statistical Extremes and Applications*, J. Tiago de Oliveira ed., 373-394, D. Reidel, Dordrecht.

Gumbel, E. J. and Goldstein, N., 1964. Analysis of empirical bivariate extremal distributions. *J. Amer. Statist. Assoc.*, 59 (307), 794-816.

*****