



Statistical Theory of Extremes

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



Part 2

Probabilistic Patterns of Univariate Statistical 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

- 2.1. If $Z = \max(X - a, Y - b)$ with (X, Y) independent with the reduced Gumbel distribution as well as z , show that $\text{Prob}\{X - a \leq Z \leq z\} = e^{-b} \Lambda(z)$.
- 2.2. Consider one of the sets of data given in the book (or elsewhere); verify using the probability paper if they can be fitted to the Gumbel, Weibull or Fréchet distributions; compare with the use of the choice statistic; estimate the parameters; take logarithms and exponentials of the same set of data and analyse the fit. Is there any apparent contradiction? Explain it.
- 2.3. Consider the Gumbel distribution $\Lambda((x - \lambda)/\delta)$. Obtain the estimators, confidence interval for λ , and tests corresponding to $\lambda_0 = \lambda_0$ vs. $\lambda > \lambda_0, \lambda < \lambda_0$ and $\lambda \neq \lambda_0$. Study the statistics used.

Passos Coelho and Gil (1963)

- 2.4. Consider the prediction problem for $\Lambda(x - \lambda)$ from a sample $(x_1, \dots, x_n)$ to a second sample (to be observed) $(Y_1, \dots, Y_n)$. Show that the least squares predictor of $W = \max(Y_1, \dots, Y_m)$ is $p(\hat{\lambda}) = \hat{\lambda} + k$ with $k = \gamma + \log \frac{m}{n} + \Gamma'(n)/\Gamma(n)$ with a mean square error of prediction of

$$E^2 = \frac{\pi^2}{6} + \Gamma'(n)/\Gamma(n) - (\Gamma'(n)/\Gamma(n))^2.$$

The shortest prediction interval for W , with level ω , is given by

$$c_\omega \leq W - \hat{\lambda} \leq d_\omega$$

Originally published in 'Statistical Analysis of Extremes', 1997, 2016

<http://dx.doi.org/10.21523/gcb1.1716>

© 2017 GATHA COGNITION® All rights reserved.

where c_ω and d_ω are given by the prediction level and the relation

$$\frac{c_\omega}{n+1} + \log\left(1 + \frac{m}{n} e^{-c_\omega}\right) = \frac{d_\omega}{n+1} + \log\left(1 + \frac{m}{n} e^{-d_\omega}\right)$$

or, approximately, $c_\omega + m e^{-c_\omega} = d_\omega + m e^{-d_\omega}$.

Tiago de Oliveira (1963)

- 2.5. Consider two populations P and P' whose distribution functions are $\Lambda(x - \lambda)$ and $\Lambda(x - \lambda')$, with the parameters λ and λ' estimable from samples $(x_1, \dots, x_n)$ and $(x'_1, \dots, x'_{n'})$.

By the likelihood ratio criterion, show that if $V_n = \frac{1}{n} \sum_1^n e^{-x_i}$ and $V'_{n'} = \frac{1}{n'} \sum_1^{n'} e^{-x'_i}$ we should discriminate for P or P' a new observation z by the rule

$$V_n^n \left(V'_{n'} - \frac{V'_{n'} - e^{-z}}{n'+1}\right)^{n'+1} \leq c(\lambda, \lambda') V_n^{n'} \left(V_n - \frac{V_n - e^{-z}}{n+1}\right)^{n+1},$$

c being computed to lead to equal chances of misclassification; c depends on $\lambda' - \lambda$. For large values of n and n' we have using the notation $\hat{\lambda} = -\log V_n$ and $\hat{\lambda}' = -\log V'_{n'}$ the approximation,
 $\hat{\lambda} - e^{\hat{\lambda}' - z} \leq \log c + \hat{\lambda}' - e^{\hat{\lambda}' - z}$.

- 2.6. Develop sequential probability ratio tests for $\Lambda(x - \lambda)$. Show that to test $\lambda = 0$ vs. $\lambda = \lambda_1$, as the logarithm of the sequential probability ratio is $Z_n = n[\lambda_1 + (1 - e^{\lambda_1})V_n]$ ($V_n = \frac{1}{n} \sum_1^n e^{-x_i}$), then α and β being the approximate values of the errors of first and second kinds, the test rule is to accept $\lambda = 0$ if $V_n \leq b_n(\lambda_1)$, to continue sampling if $b_n(\lambda_1) < V_n < a_n(\lambda_1)$, and to reject $\lambda = 0$ if $a_n(\lambda_1) \leq V_n$, if $\lambda_1 < 0$, and with the inequalities reversed if $\lambda_1 > 0$; we have denoted by $a_n(\lambda) = (\log A - n\lambda)/(1 - e^\lambda)$, $b_n(\lambda) = (\log B - n\lambda)/(1 - e^\lambda)$ where $A = \frac{1-\beta}{\alpha}$ and $B = \frac{1-\beta}{\alpha}$. Also obtain the operating characteristic and the average sample number of the sequential test.

Mexia (1967/68)

- 2.7. Form a two-sided sequential test of $\lambda = 0$ vs. $\lambda = \bar{\lambda} < 0$ and $\lambda = \bar{\bar{\lambda}} > 0$ by “intersection” of two one-sided tests as before with $\lambda_1 = \bar{\lambda}$ and $\lambda_1 = \bar{\bar{\lambda}}$, “intersection” meaning accept $\lambda = 0$ if both one-sided tests accept it, reject it if any of the one-sided tests rejected it, and continue sampling otherwise. Obtain the special feature of the graph of the

sequential probability ratio test and evaluate approximately the operating characteristic and the average sample number.

Mexia (1967/68)

- 2.8. Develop a sequential estimation procedure for the parameter λ for $\Lambda(x - \lambda)$.

Mexia (1967/68)

- 2.9. Consider the last six exercises and solve the analogous questions for the exponential distribution, which is equivalent to taking $Y = e^{-X}$; as X is $\Lambda(x - \lambda)$ then Y has the exponential distribution with zero location and $\delta = e^{-\lambda}$. The reverse method could be followed: to deduce from the results for the exponential distribution the results for $\Lambda(x - \lambda)$.
- 2.10. Obtain the locally most powerful one-sided tests for $\lambda = \lambda_0$ vs. $\lambda = \lambda_0$ (or $\lambda < \lambda_0$) with fixed δ ($\delta = 1$, for simplicity), for $\delta = \delta_0$ vs. $\delta = \delta_0$ (or $\delta < \delta_0$) with fixed λ ($\lambda = 0$, for simplicity); obtain the corresponding locally most powerful unbiased two-sided tests in both, cases and study asymptotic behaviour.
- 2.11. For the Gumbel distribution consider a pair of estimators (λ^*, δ^*) of (λ, δ) and assume that $V(\lambda^*) \sim A \delta^2/n$, $V(\delta^*) \sim B \delta^2/n$, $C(\lambda^*, \delta^*) \sim C \delta^2/n$.

a) Compute the asymptotic efficiency for the estimation of any quantile and obtain its minimum (worst asymptotic efficiencies);

b) Choose a pair of central quantiles (of probabilities $0 < p < q < 1$); evaluate the worst asymptotic efficiency; seek the best pair (p, q) , i.e., the one giving the largest value to the smallest efficiency.

- 2.12. For the Gumbel distribution consider an estimator (λ^*, δ^*) of (λ, δ) , asymptotically standard binormal, with mean values λ and δ , and $V(\lambda^*) \sim A \delta^2/n$, $V(\delta^*) \sim B \delta^2/n$ and $C(\lambda^*, \delta^*) \sim C \delta^2/n$. Compute the asymptotic variance of the estimator (or predictor) $\lambda^* + \chi \delta^*$ of the quantile $\lambda + \chi \delta$ and obtain its asymptotic efficiency with respect to the estimator (or predictor) $\hat{\lambda} + \chi \hat{\delta}$ based on the maximum likelihood. In particular, compute this efficiency when (λ^*, δ^*) are based on two quantiles, and analyse the choice of the quantiles with different criteria (optimization for $\chi = a$, best asymptotic efficiency for $\chi \rightarrow \pm\infty$, etc.).
- 2.13. Supposing that the samples $(x_1, \dots, x_n)$ and $(x'_1, \dots, x'_n)$ have a Gumbel distribution with parameters (λ, δ) and (λ', δ') , determine the likelihood ratio test for $\lambda = \lambda'$, $\delta = \delta'$, and obtain its asymptotic behaviour. Using the approximate relation $(\bar{x} - \hat{\lambda})/\hat{\delta} \approx \gamma$ simplify the expression and determine its asymptotic behaviour (use the δ -method).

- 2.14. Suppose that $X_1, \dots, X_n$ have a Gumbel distribution with parameters $\lambda + \beta i$ ($i = 1, \dots, n$) and δ . Obtain the least squares estimators of the parameters (λ, β, δ) of this linear regression with additive Gumbel errors ($X_i = \lambda + \beta i + \delta Z_i$, where the Z_i are reduced Gumbel random variables) and its asymptotic behaviour, and formulate a test of $\beta = 0$ (no regression). Generalize, substituting the term βi by βc_i , with the c_i known.
- 2.15. Determine the equations defining the values (k_1, k_2) that minimize the mean length of the (β, ω) and α -average tolerance intervals for Gumbel distribution.

Themido (1985)

- 2.16. In the search for one-sided and two-sided tolerance intervals for Gumbel distribution obtain the first-order development of the coefficients k_i (i.e., the a such that $k_i(n) = k_i + a(n^{-1/2}) + O(n^{-1/2})$), using the δ -method.

Themido (1985)

- 2.17. Generate 50 plus 20 random numbers for the Gumbel (Fréchet) distribution for maxima and study, in the way used for the Weibull distribution of minima, possible predictors of maxima of the next 20 observations based on the initial sample of 50 observations.
- 2.18. Using the method of moments or of the two quantiles or of Downton or of blocks, obtain the test of $(\lambda, \delta) = (\lambda_0, \delta_0)$ vs. $(\lambda, \delta) \neq (\lambda_0, \delta_0)$.
- 2.19. Consider the following situation: the random variables X_{ij} ($i = 1, \dots, m$, $j = 1, \dots, p$) have a Gumbel distribution with location parameters λ_{ij} and unit ($\delta_{ij} = 1$) dispersion parameter. For the sample of $n = m p$, test the periodicity of the λ_{ij} , i.e., $\lambda_{ij} = \lambda_j$, using the likelihood ratio criterion. Show that the statistic obtained is translation invariant in the hypothesis tested (i.e., for the transformation $x_{ij} \rightarrow x_{ij} + a_j$), and obtain its asymptotic behaviour, obtain the corresponding formulation after transformation to exponential margins.
- 2.20. Let $\{X_k\}$ be i.i.d. with distribution function $\Lambda((x - \lambda)/\delta)$. Now consider the dependent sequence $M_1 = x_1, M_k = \max(M_{k-1}, x_k) = \max(x_1, \dots, x_k)$ for $k > 1$.

a) Show that the sequence $\{M_1, \dots, M_n\}$ is increasing, i.e., $M_1 < M_n$, with probability $1 - 1/n$;

b) Show that

$$M(M_k) = \lambda + \delta((\gamma + \log k), V(M_k) = \frac{\pi^2}{6} \delta^2 \text{ and for } p \geq 1$$

$$C(M_k, M_{k+p}) =$$

$$\left[\frac{\pi^2}{6} + \iint_{x \leq y + \log p/k} (x - \gamma) (y + \log p/k - x) d\Lambda(x) d\Lambda(y) \right] \delta^2$$

$$= \left[\iint_{x \leq y + \log p/k} (x - \gamma)(y + \log p/k) + \frac{k}{p+k} \log \frac{p+k}{k} \right] \delta^2;$$

- 2.21. Obtain the least squares estimators of (λ, δ) in the previous exercise, if $M_1 < M_n$.
- 2.22. Obtain maximum likelihood estimators of δ for the Fréchet and Weibull distributions, supposing $\lambda = 0$ and α known. Compare these results with the estimators of λ for the Gumbel distribution with $\delta = 1$. Reduce all the results to the study of the dispersion parameter for exponential distribution.
- 2.23. Compute β_1 and β_2 for the Fréchet distribution.
- 2.24. Draw the graphs of the densities of $\alpha^{-1} \Phi'_\alpha(1 + y/\alpha)$ and verify that for $\alpha > 6$ is very close to $\Lambda'(y)$.
- 2.25. Compute β_1 and β_2 for the Weibull distribution of minima.
- 2.26. Show that if Z has the distribution function $W_\alpha(z)$ then $\text{Prob}\{\alpha(1 - Z) = y\}$ converges to $\Lambda(y)$. Draw the corresponding graphs.
- 2.27. Compute tables, similar to the one of [Chapter 5](#), for the shortest interval for the same probabilities and different values of the shape parameter (α) for the Fréchet and Weibull distributions.
- 2.28. Obtain prediction of maxima of the next m observations for a Weibull distribution of maxima.
- 2.29. Obtain tests and estimators using two and three central quantiles for the Gumbel and Fréchet distributions of maxima and the Weibull distribution of minima.

Tiago de Oliveira and Littauer (1976), Kubat and Epstein (1980), Neves (1986)

- 2.30. Determine the expression of the efficiency for the estimation of quantiles when using the method of moments.
- 2.31. For a Fréchet distribution (for maxima) with location parameter known (zero for convenience) and shape parameter known, obtain maximum likelihood and least squares (and approximations), homogeneous estimators (i.e., such that $\delta^*(\alpha x_i) = \alpha \delta^*(x_i), \alpha > 0$) of the quantiles, homogeneous predictors, and their asymptotic distributions, as well as the behaviour of overpassing probability.

Neves (1987)

- 2.32. For a Weibull distribution (for minima) with location parameter known (zero for convenience) and shape parameter known, obtain maximum likelihood and least squares (and approximations), homogeneous estimators (i.e., such that $\delta^*(\alpha x_i) = \alpha \delta^*(x_i), \alpha > 0$) of the quantiles, homogeneous predictors, and their asymptotic distributions, as well as the behaviour of underpassing probability.
- 2.33. Compute the variance-covariance matrix of maximum likelihood estimators for the Fréchet distribution (for maxima) with $\alpha > 0$.
- 2.34. Compute the variance-covariance matrix of maximum likelihood estimators for the Weibull distribution (for minima) with $\alpha > 2$.
- 2.35. Seek the probability $p(0 < p < 1)$ such that $\lambda^* = X'_1 - a(Q_n(P) - X'_1), \delta^* = b(Q_n(P) - X'_1)$, where $Q_n(P)$ is the sample quantile (where $[np] > 0$) and X'_1 is the sample minimum and $a, b > 0$, are the best estimators of (λ, δ) for the Fréchet distribution with $\alpha = \alpha_0$ known, 'best' being taken either in the meaning of minimizing the generalized asymptotic variance or of minimizing the asymptotic variance of some quantile; note that the estimators are quasi-linearly invariant. Compute the efficiency of (λ^*, δ^*) .
- 2.36. Do the same exercise for the Weibull distribution (for minima).
- 2.37. Seek the probabilities $p, q (0 < p < q < 1)$ such that for Fréchet distribution $\lambda^* = X'_1 - a_1(Q_n(p) - X'_1) - a_2(Q_n(q) - X'_1), \delta^* = b_1(Q_n(p) - X'_1) + b_2(Q_n(q) - X'_1), \alpha^* = g^{-1}(\frac{Q_n(q) - X'_1}{Q_n(p) - X'_1})$, where $X'_1, Q_n(p), Q_n(q)$ have the usual meaning, and we suppose $0 < [np] < [nq]$, and $a_i, b_i > 0$, are such that $(\lambda^*, \delta^*, \alpha^*)$ minimizes either the generalized asymptotic variance or the asymptotic variance of some quantile; note that $(\lambda^*, \delta^*, \alpha^*)$ are quasi-linearly invariant. Compute the efficiency of $(\lambda^*, \delta^*, \alpha^*) \cdot g(\alpha)$ can be defined by $\frac{Q_n(q) - X'_1}{Q_n(p) - X'_1} \xrightarrow{P} g(\alpha) = \frac{\lambda_q}{\lambda_p}$.
- 2.38. Do the same exercise for the Weibull distribution (for minima).
- 2.39. Consider the Weibull distribution for minima with $\alpha = 1/2$, i.e.,
- $$W_{1/2}(x|\lambda, \delta) = 0 \quad \text{if } x \leq \lambda$$
- $$W_{1/2}(x|\lambda, \delta) = 1 - \exp(-(\frac{x-\lambda}{\delta})^{1/2}) \quad \text{if } x \geq \lambda$$

Show that if $\lambda = \lambda_0$ is known then $\hat{\delta} = (\frac{1}{n} \sum_{i=1}^n \sqrt{x_i - \lambda_0})^2$ is asymptotically normal. If λ and δ are unknown parameters, consider $\hat{\lambda} = \min(x_1, \dots, x_n)$ and $\hat{\delta} = (\frac{1}{n} \sum_{i=1}^n \sqrt{x_i - \hat{\lambda}})^2$. $\hat{\delta}$ is a modified estimator of δ where in the previous expression we substituted λ_0 by $\hat{\lambda}$.

which converges quickly to λ . Show that $(\hat{\lambda}, \hat{\delta})$ is an estimator of (λ, δ) with independent Weibull and normal margins. Use the δ -method.

Tiago de Oliveira (1982)

2.40. Consider $0 < p < q < r < 1$ such that $\beta = \frac{\log p}{\log q} = \frac{\log q}{\log r} > 1$, and define

the statistic $Q^* = \frac{X'_{[nr]} - X'_{[nq]}}{X'_{[nq]} - X'_{[np]}} \xrightarrow{P} \beta^\theta$, where θ is the shape parameter in the von Mises-Jenkinson formula; note that Q^* is positive iff $[np] < [nq] < [nr]$. Then $k^* = \frac{\log Q^*}{\log \beta}$ is an estimator of θ which is asymptotically normal. Use θ^* for statistical choice between the Weibull, Gumbel and Fréchet distributions, as done in the test, and compare it with $\hat{V}_n(x)$.

2.41. Consider the choice statistic

$$\hat{V}_n(x) = \frac{1}{n} \left\{ \sum_1^n (x_i - \bar{x})^2 / \hat{\delta}^2 - n \frac{\sum_1^n ((x_i - \bar{x}) / \hat{\delta})^2 e^{-(x_i - \bar{x}) / \hat{\delta}}}{\sum_1^n e^{-(x_i - \bar{x}) / \hat{\delta}}} \right\},$$

with

$$\hat{\delta} = \bar{x} - \frac{\sum_1^n x_i e^{-x_i / \hat{\delta}}}{\sum_1^n e^{-x_i / \hat{\delta}}}.$$

It was been shown that if $\theta = 0$ (Gumbel distribution) $\hat{V}_n(x) / \sqrt{n} \hat{\delta}$ is asymptotically standard normal. Consider then the following modified approximate statistics:

(a) as $\frac{1}{n} \sum_1^n e^{-(x_i - \bar{x}) / \delta} \approx M(e^{-(x - (\lambda + \gamma \delta)) / \delta}) = e^\gamma$, study

$$\hat{V}_n(x) = \frac{1}{n} \left\{ \sum_1^n \frac{(x_i - \bar{x})^2}{\hat{\delta}^2} - e^{-\gamma} \sum_1^n \frac{(x_i - \bar{x})^2}{\hat{\delta}^2} e^{-(x_i - \bar{x}) / \hat{\delta}} \right\}$$

and show its asymptotically normality;

b) as $\frac{1}{n} \sum_1^n \frac{(x_i - \bar{x})^2}{\hat{\delta}^2} e^{-(x_i - \bar{x}) / \hat{\delta}} \approx M\left\{\left(\frac{x - (\lambda + \gamma \delta)}{\delta}\right)^2 e^{-(x - (\lambda + \gamma \delta)) / \delta}\right\} = \frac{\pi^2}{6}$

study the statistic $\hat{V}_{n,2} = \frac{n}{2} (s^2 / \hat{\delta}^2 - \pi^2 / 6)$ and show its asymptotic normality; note that this is a comparison between the estimates $\hat{\delta}^2$ and $\delta^{*2} = \frac{6}{\pi^2} s^2$, the first approximation to $\hat{\delta}$.

c) Study the statistic $\hat{V}_{n,3}$, obtained by substituting $\hat{\delta}$ by δ^* in $\hat{V}_{n,1}$, and show its asymptotic normality;

d) Study the statistic $\hat{V}_{n,4}$ when (λ, δ) have been estimated by the method of moments, i.e., $\lambda^* = \bar{x} - \gamma \delta^*$, $s^2 = \frac{6}{\pi^2} \delta^{*2}$, and show its asymptotic normality;

e) Compare the efficiencies of these statistics with respect to $\hat{V}_n$.

- 2.42. Compare the $\text{Prob}\{ \text{accept } k = 0 | k \}$ for the statistic $\hat{V}_n(x)$ and for the statistic Q_n for the same significance level ($\alpha = .05, .02, .01, .$) and different values of k .
- 2.43. Analyse also the same question for the statistics given in exercise 2.41.
- 2.44. Compute Δ_n and $\tilde{P}'_n(0)$, or its orders, for the approximation of order n^{-2} for the solution of the statistical trilemma of the locally optimal approach.
- 2.45. Show that efficiency of largest observations vs. subsamples methods changes with the value of the shape parameter k of the von Mises-Jenkinson formula, which shows some lack of robustness.
- 2.46. For the techniques with partial information, seek the optimal quantiles (for quantile estimation and for block estimation) that minimize the asymptotic variance of the estimators (λ^*, δ^* or α^* if it exists) and compare with those with maxima (global) asymptotic variance.
- 2.47. Owing to the importance of λ in $\Phi_\alpha(\frac{x-\lambda}{\delta})$ and $W_\alpha(\frac{x-\lambda}{\delta})$, which can be estimated by $X'_1 = \min_{1 \leq i \leq n} X_i$, consider the (new) random variables $Y_i = X_i - X'_1$ ($\min_{1 \leq i \leq n} Y_i = 0$). Suppose that they have distributions $\Phi_\alpha(y/\delta)$ and $W_\alpha(y/\delta)$ (which they have approximately), and study the estimators (α^*, δ^*) of (α, δ) based on the positive Y_i (in number of $n - 1$ with probability one). Obtain sufficient conditions for the statistics (α^*, δ^*) to be estimators of (α, δ) . Recall that $\Phi_\alpha(y/\delta)$ and $W_\alpha(y/\delta)$ can be easily transformed to have a Gumbel distribution.
- 2.48. Develop the theory of “excesses over thresholds” for the limiting distributions of maxima both for the P.O.T. and A.E.O.T. methods.

References

- Kubat, P. and Epstein, B., 1980. Estimation of quantiles of location scale distribution based on 2 or 3 order statistics. *Trab. Estad. Inv. Oper.*, 22, 575-581.
- Mexia, J. T., 1967. Studies on the extreme double exponential distribution -II, Sequential estimation and testing for the location parameter of Gumbel distribution. *Rev. Fac. Ciencias de Lisboa, A*, 12, 5-14.
- Neves, M., 1986. Estimacao de quantis das distribuicoes de Gumbel e Frechet cots parametros de escala e de localizacao desconhecidos, baseados em duas estatisticas ordinais. *Centro de Estatistica e Aplicacoes*, 25/86, Lisboa.
- Neves, M., 1987. Predicao homogenea para a distribuicao de Frechet. *Jorn. Matem. Luso-Espanholas*, Braga.
- Passos Coelho, D. and Gil, T. P., 1963. Studies on extreme double exponential distribution, I - The location parameter. *Rev. Fac. Ciencias de Lisboa, A*, 9, 37-46.

- Themido, T., 1985. Intervalos de tolerancia equivariantes para a distribuicao de Gumbel, Universidade de Lisboa, Tese de Mestrado.
- Tiago de Oliveira, J. and Littauer, S. B., 1976. Mean square invariant foreca the Weibull distributions. *Naval Res. Logist. Quat.*, 23, 487-511.
- Tiago de Oliveira, J., 1963. Decision results for the parameters of the extreme value distribution (Gumbel) based on the mean and the standard deviation. *Trab. Estad. Inv. Oper.*, 14, 61-81, Madrid.
- Tiago de Oliveira, J., 1982. The δ -method for obtention of asymptotic distributions; applications. *Publ. Inst. Statist. Univ. Paris*, 27, 49-70.
