European Geosciences Union General Assembly 2011 Vienna, Austria, 3-8 April 2011 Session HS7.5: Hydroclimatic Stochastics

A general Monte Carlo method for the construction of confidence intervals for a function of probability distributions

D. Koutsoyiannis, S. Kozanis and H. Tyralis, Department of Water Resources and Environmental Engineering, National Technical University of Athens (itia.ntua.gr/1125)

1. Abstract

We derive an algorithm which calculates an exact confidence interval for a distributional parameter of location or scale family, based on a two-sided hypothesis test on the parameter of interest, using some pivotal quantities. We use this algorithm to calculate approximate confidence intervals for the parameter or a function of the parameter of one-parameter distributions. We show that these approximate intervals are asymptotically exact. We modify the algorithm and use it to obtain approximate confidence intervals for a parameter or a function of parameters for multi-parameter distributions. We compare the results of the method with those obtained by known methods of the literature for the normal, the gamma and the Weibull distribution and find them satisfactory. We conclude that the proposed method can yield approximate confidence intervals, based on Monte Carlo simulations, in a generic way, irrespectively of the distribution function, as well as of the type of the parameters or the function of parameters.

2. Computation of an approximate confidence interval

- Suppose that we seek an approximate $1-\alpha$ confidence interval for θ, of the one parameter distribution $f(x\,|\,\theta)$. We have proved that an asymptotically exact confidence interval is

$$\Big[\hat\theta + \frac{\hat\theta - \upsilon(\hat\theta)}{(d\upsilon/d\theta)_{|\theta=\hat\theta}},\ \hat\theta + \frac{\hat\theta - \lambda(\hat\theta)}{(d\lambda/d\theta)_{|\theta=\hat\theta}}\Big] \quad \text{where} \quad \lambda(\theta) = F^{-1}(\alpha/2|\theta) \text{ and } \upsilon(\theta) = F^{-1}(1 - \alpha/2|\theta)$$

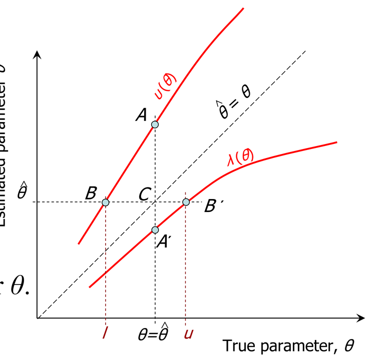
  and $\hat\theta$ is the maximum likelihood estimate.
- Suppose that we seek an approximate $1-\alpha$ confidence interval for $\beta = h(\boldsymbol\theta)$, of the multi-parameter distribution $f(x\,|\,\boldsymbol\theta)$. We have proved that an asymptotically exact confidence interval is again the above interval where we substitute for $(d\lambda/d\theta)_{|\theta=\hat\theta}$, $(d\upsilon/d\theta)_{|\theta=\hat\theta}$ according to the following

$$\frac{d\lambda}{d\beta} = \frac{q_{12}+q_{13}}{q_{22}+q_{23}}, \quad \frac{d\upsilon}{d\beta} = \frac{q_{31}+q_{32}}{q_{21}+q_{22}} \quad \boldsymbol{q} := \frac{d\boldsymbol\gamma}{d\boldsymbol\theta}\mathrm{Var}(\boldsymbol T)\Big(\frac{d\boldsymbol\gamma}{d\boldsymbol\theta}\Big)^{\mathrm T} \quad \frac{d\boldsymbol\gamma}{d\boldsymbol\theta} = \begin{bmatrix} \frac{d\lambda}{d\boldsymbol\theta} \\ \frac{d\beta}{d\boldsymbol\theta} \\ \frac{d\upsilon}{d\boldsymbol\theta} \end{bmatrix} = \begin{bmatrix} \frac{\partial\lambda}{\partial\theta_1} & \frac{\partial\lambda}{\partial\theta_2} & \cdots & \frac{\partial\lambda}{\partial\theta_k} \\ \frac{\partial\beta}{\partial\theta_1} & \frac{\partial\beta}{\partial\theta_2} & \cdots & \frac{\partial\beta}{\partial\theta_k} \\ \frac{\partial\upsilon}{\partial\theta_1} & \frac{\partial\upsilon}{\partial\theta_2} & \cdots & \frac{\partial\upsilon}{\partial\theta_k} \end{bmatrix}$$

- In the following we will use the maximum likelihood estimators of the parameters of interest.
- All derivatives are calculated using stochastic simulation
- The method extends, unifies and generalizes approximate confidence intervals by B. Ripley (Stochastic simulation, John Wiley & Sons 1987)

See details concerning our method in Koutsoyiannis D., Kozanis S. and Tyralis, H. (2011) An algorithm to construct Monte Carlo confidence intervals for an arbitrary function of probability distribution parameters (in preparation).

3. Determination of l and u from an inversion of a hypothesis test



From this figure we have that $\frac{\upsilon(\hat\theta)-\hat\theta}{\hat\theta - l} = \frac{CA}{CD} \approx \big(\frac{d\upsilon}{d\theta}\big)_{|\theta=\hat\theta}$

Solving for l and u, we find $l \approx \hat\theta + \frac{\hat\theta - \upsilon(\hat\theta)}{(d\upsilon/d\theta)_{|\theta=\hat\theta}}$ and $u \approx \hat\theta + \frac{\hat\theta - \lambda(\hat\theta)}{(d\lambda/d\theta)_{|\theta=\hat\theta}}$.

According to the following, $[l,u]$ is an exact $1-\alpha$ confidence interval for θ.

We construct a test H_0: $\theta = \hat\theta$ vs H_1: $\theta \neq \hat\theta$ with acceptance region:

$$A(\hat\theta) = \{\boldsymbol x: F^{-1}(\alpha/2|\hat\theta) \le B(\boldsymbol x) \le F^{-1}(1-\alpha/2|\hat\theta)\} \tag{1}$$

where $B(\boldsymbol X)$ is an estimator of the parameter θ, which is a size α test because $\beta(\hat\theta) = 1 - P(F^{-1}(\alpha/2|\hat\theta) \le B(\boldsymbol X) \le F^{-1}(1-\alpha/2|\hat\theta)|\theta = \hat\theta) = 1 - [F(F^{-1}(1-\alpha/2|\hat\theta)|\hat\theta) - F(F^{-1}(\alpha/2|\hat\theta)|\hat\theta)] = 1 - (1 - \alpha/2 - \alpha/2) = \alpha$. From this test we obtain the following $1-\alpha$ confidence interval for θ:

$$C(\boldsymbol x) = \{\hat\theta: F^{-1}(\alpha/2|\hat\theta) \le B(\boldsymbol x) \le F^{-1}(1-\alpha/2|\hat\theta)\} \tag{2}$$

In our case we assume that we have an observation $\boldsymbol y = (y_1,\dots,y_n)$. We obtain the following $1-\alpha$ confidence interval for θ:

$$C(\boldsymbol y) = \{\theta: F^{-1}(\alpha/2|\theta) \le B(\boldsymbol y) \le F^{-1}(1-\alpha/2|\theta)\} \tag{3}$$

Now we define l and u as the solutions of the equations:

$$\upsilon(l) = B(\boldsymbol y) \text{ and } \lambda(u) = B(\boldsymbol y) \tag{4}$$

From the above equation we obtain that:

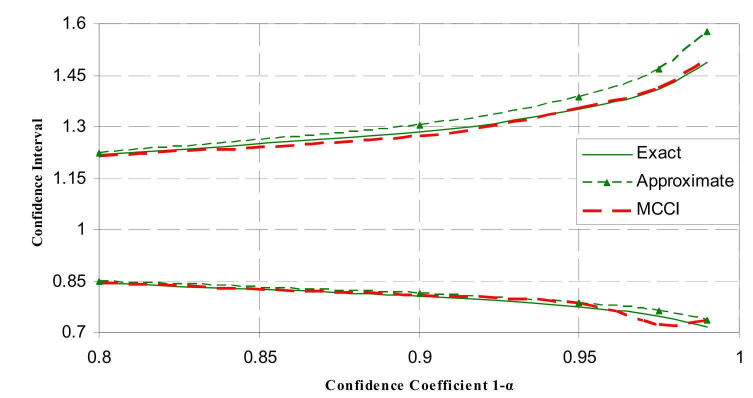
$$F^{-1}(\alpha/2|u) = B(\boldsymbol y) \text{ and } F^{-1}(1-\alpha/2|l) = B(\boldsymbol y) \tag{5}$$

We assume that $C(\boldsymbol y) = [\theta_1,\theta_2]$ where θ_1,θ_2 are the solutions of the equations

$$F^{-1}(\alpha/2|\theta_2) = B(\boldsymbol y) \text{ and } F^{-1}(1-\alpha/2|\theta_1) = B(\boldsymbol y) \tag{6}$$

Now it is obvious that $[l,u]$ is an approximate $1-\alpha$ confidence interval estimate for θ.

4. Confidence interval for the scale parameter of an exponential distribution



The $1-\alpha$ exact confidence interval is obtained by the following equations.

$$F(\bar x|n,l/n) = 1-\alpha/2,\ F(\bar x|n,u/n) = \alpha/2$$

where $f(x|\theta,k) = \frac{e^{-x/\theta}x^{k-1}}{\Gamma(k)\theta^k}$, $x \ge 0$ is the density of the gamma distribution, $\theta > 0$ is the scale parameter and $k > 0$ is the shape parameter

The $1-\alpha$ approximate confidence interval is obtained by the following equation (*).

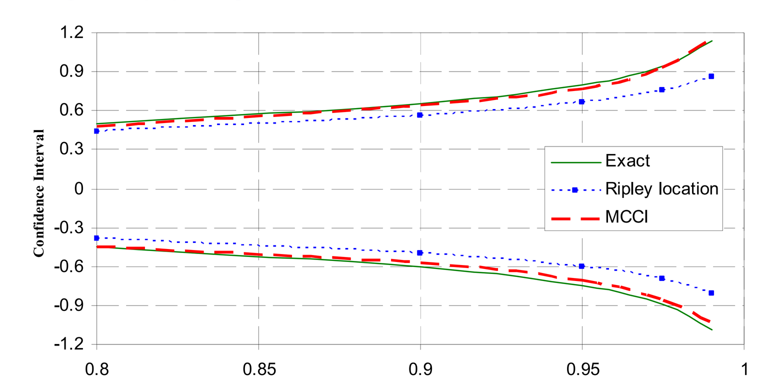
$$[L(\boldsymbol X),U(\boldsymbol X)] = \Big[\frac{\bar X}{1+\Phi^{-1}(1-\alpha/2)/\sqrt n}, \frac{\bar X}{1-\Phi^{-1}(1-\alpha/2)/\sqrt n}\Big]$$

MCCI in the following will denote the $1-\alpha$ confidence interval obtained by our method.

Here we have a simulated sample with 50 elements from an exponential distribution with $\sigma = 1$. For this sample $\hat\sigma = 1.002$.

(*) Papoulis A, Pillai U (2002) Probability, Random Variables and Stochastic Processes. McGraw-Hill

5. Confidence interval for the location parameter of a normal distribution

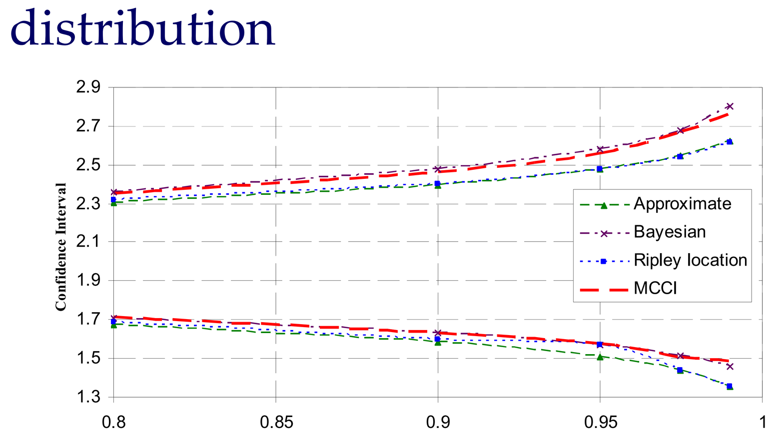


The $1-\alpha$ exact confidence interval is obtained by the following equation.

$$[l(\boldsymbol x),u(\boldsymbol x)] = [\bar x - t_{n-1}(1-\alpha/2)\frac{s}{\sqrt n}, \bar x + t_{n-1}(1-\alpha/2)\frac{s}{\sqrt n}]$$

The $1-\alpha$ interval obtained by Ripley's first method designated as Ripley location.
Here we have a simulated sample with 10 elements from a normal distribution with $\mu = 0$ and $\sigma = 1$. For this sample $\hat\mu = 0.026$ and $\hat\sigma = 1.023$.

6. Confidence interval for the quantile $\mu + 2\sigma$ of a normal distribution



The $1-\alpha$ approximate confidence interval estimate is given by the following equation.

$$[l(\boldsymbol x),u(\boldsymbol x)] = [\bar x + z_p s - \Phi^{-1}(1-\alpha/2)\frac{s}{\sqrt n}\sqrt{1+z_p^2/2}, \bar x + z_p s + \Phi^{-1}(1-\alpha/2)\frac{s}{\sqrt n}\sqrt{1+z_p^2/2}]\ ^{(*)}$$

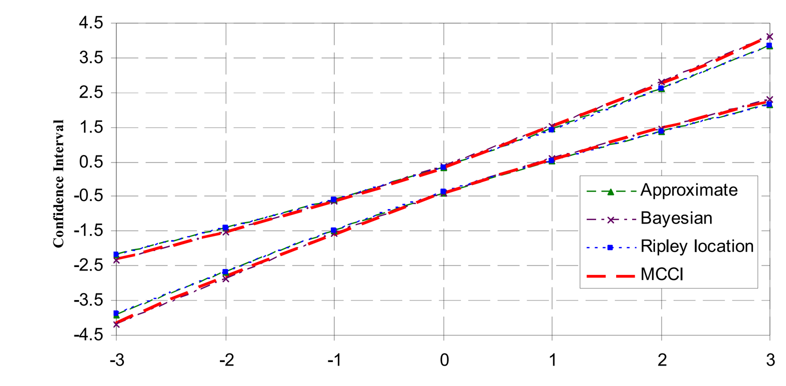
The $1-\alpha$ Bayesian confidence interval (**), if we choose a prior $P(\mu,\sigma) \propto 1/\sigma^2$, is obtained by a sampler based on the following mixture.

$$\sigma^2|\boldsymbol x \sim \text{inv-}\chi^2(n-1,s^2) \text{ and } \mu|\sigma^2,\boldsymbol x \sim \mathrm N(\bar x,\sigma^2/n)$$

Here we have a simulated sample with 50 elements from a normal distribution with $\mu = 0$ and $\sigma = 1$. For this sample $\hat\mu = -0.027$ and $\hat\sigma = 0.998$.

(*) Koutsoyiannis, D., (1997) Statistical Hydrology, Edition 4, National Technical University of Athens, Athens.
(**) Gelman A, Carlin J, Stern H, Rubin D (2004) Bayesian Data Analysis. Chapman & Hall/CRC.

7. Confidence intervals with confidence coefficient 0.99 of a normal distribution



The 0.99 approximate confidence interval estimate is given by the following equation.

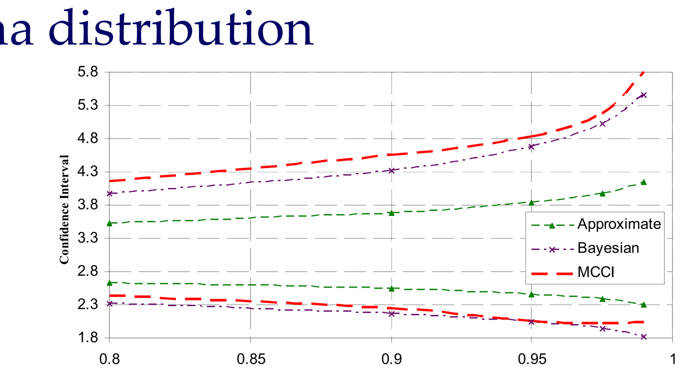
$$[l(\boldsymbol x),u(\boldsymbol x)] = [\bar x + z_p s - \Phi^{-1}(1-\alpha/2)\frac{s}{\sqrt n}\sqrt{1+z_p^2/2}, \bar x + z_p s + \Phi^{-1}(1-\alpha/2)\frac{s}{\sqrt n}\sqrt{1+z_p^2/2}]$$

The 0.99 Bayesian confidence interval, if we chose a prior $P(\mu,\sigma) \propto 1/\sigma^2$, is obtained by a sampler based on the following mixture.

$$\sigma^2|\boldsymbol x \sim \text{inv-}\chi^2(n-1,s^2) \text{ and } \mu|\sigma^2,\boldsymbol x \sim \mathrm N(\bar x,\sigma^2/n)$$

Here we have a simulated sample with 50 elements from a normal distribution with $\mu = 0$ and $\sigma = 1$. For this sample $\hat\mu = -0.027$ and $\hat\sigma = 0.998$.
In the figure we calculate the confidence intervals for values of z_p between −3 and 3.

8. Confidence interval for the scale parameter of a gamma distribution



The $1-\alpha$ approximate confidence interval is given by the following equation.

$$[L(\boldsymbol X),U(\boldsymbol X)] = \Big[\frac{2n\bar X}{\chi^2_{v+2n\hat k}(1-\alpha/2) - 2n\hat k R_n/c}, \frac{2n\bar X}{\chi^2_{v+2n\hat k}(\alpha/2) - 2n\hat k R_n/c}\Big]\ ^{(*)}$$

The $1-\alpha$ Bayesian confidence interval (**), if we chose a prior $P(k,\theta) \propto 1/\theta$, is obtained by the following Gibbs sampler.

$$\theta|k,\boldsymbol x \sim IG(nk,1/\sum_{i=1}^{n} x_i)$$

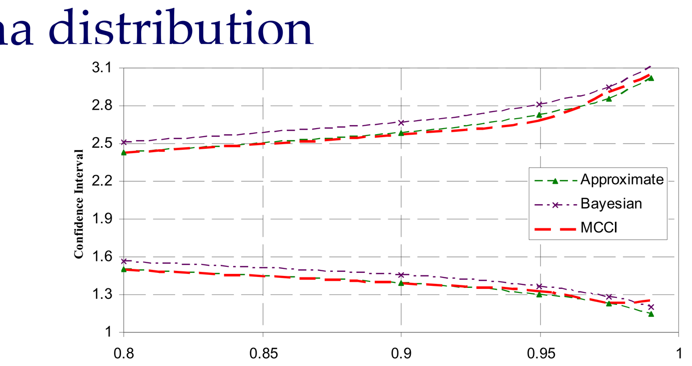
where $IG(a,b)$ denotes the inverse gamma distribution with parameters a and b, with density function defined by $f(x|a,b) = [\Gamma(a)b^a]^{-1}x^{-(a+1)}\exp(-1/bx)$, x, a, $b > 0$. And

$$P(k|\theta,\boldsymbol x) \propto [\Gamma(k)]^{-n}\theta^{-nk}\prod_{i=1}^{n} x_i^{k-1}$$

Here we have a simulated sample with 50 elements from a gamma distribution with $k = 2$ and $\theta = 3$. For this sample $\hat k = 1.979$ and $\hat\theta = 3.007$.

(*) Bhaumik D, Kapur K, Gibbons R (2009) Testing Parameters of a Gamma Distribution for Small Samples. Technometrics 51(3):326-334
(**) Son YS, Oh M (2006) Bayesian Estimation of the Two-Parameter Gamma Distribution. Communications in Statistics: Simulation and Computation 35(2):285-293

9. Confidence interval for the shape parameter of a gamma distribution



The $1-\alpha$ approximate confidence interval is given by the following inequality where we solve for k.

$$\frac{\mathrm{Var}(R_n)}{\mathrm E(R_n)}\chi^2_v(\alpha/2) < 2R_n < \frac{\mathrm{Var}(R_n)}{\mathrm E(R_n)}\chi^2_v(1-\alpha/2)\ ^{(*)}$$

The $1-\alpha$ Bayesian confidence interval (**), if we chose a prior $P(k,\theta) \propto 1/\theta$, is obtained by the following Gibbs sampler.

$$\theta|k,\boldsymbol x \sim IG(nk,1/\sum_{i=1}^{n} x_i)$$

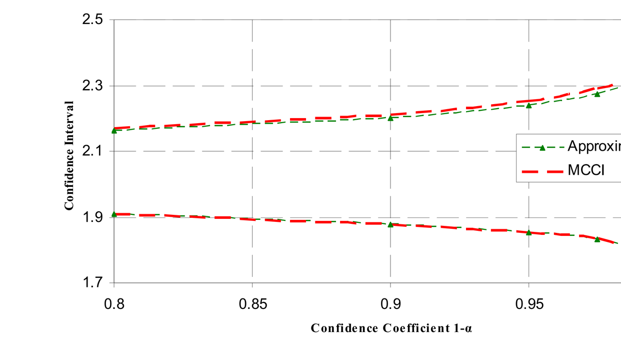
where $IG(a,b)$ denotes the inverse gamma distribution with parameters a and b, with density function defined by $f(x|a,b) = [\Gamma(a)b^a]^{-1}x^{-(a+1)}\exp(-1/bx)$, x, a, $b > 0$. And

$$P(k|\theta,\boldsymbol x) \propto [\Gamma(k)]^{-n}\theta^{-nk}\prod_{i=1}^{n} x_i^{k-1}$$

Here we have a simulated sample with 50 elements from a gamma distribution with $k = 2$ and $\theta = 3$. For this sample $\hat k = 1.979$ and $\hat\theta = 3.007$.

(*) Bhaumik D, Kapur K, Gibbons R (2009) Testing Parameters of a Gamma Distribution for Small Samples. Technometrics 51(3):326-334
(**) Son YS, Oh M (2006) Bayesian Estimation of the Two-Parameter Gamma Distribution. Communications in Statistics: Simulation and Computation 35(2):285-293

10. Confidence intervals for the scale parameter of a Weibull distribution



The $1-\alpha$ approximate confidence interval is given by the following equation (*).

$$[l(\boldsymbol x),u(\boldsymbol x)] = \Big[\Big(\frac{2S(\hat b)}{c_1\chi^2_{2n}(1-\alpha/2) - 2n(c_1-1)}\Big)^{1/\hat b}, \Big(\frac{2S(\hat b)}{c_1\chi^2_{2n}(\alpha/2) - 2n(c_1-1)}\Big)^{1/\hat b}\Big]$$

where $S(b) := \sum_{i=1}^{n} x_i^b$ and $c_1 := \sqrt{1+0.607927\cdot 0.422642^2}$. We denote $\hat b$ the modified maximum likelihood estimate given by the minimization of the following function

$$L(b) := \frac{n-2}{b} - \Big(n\sum_{i=1}^{n} x_i^b \ln x_i\Big)\Big(\sum_{i=1}^{n} x_i^b\Big)^{-1} + \sum_{i=1}^{n}\ln x_i = 0$$

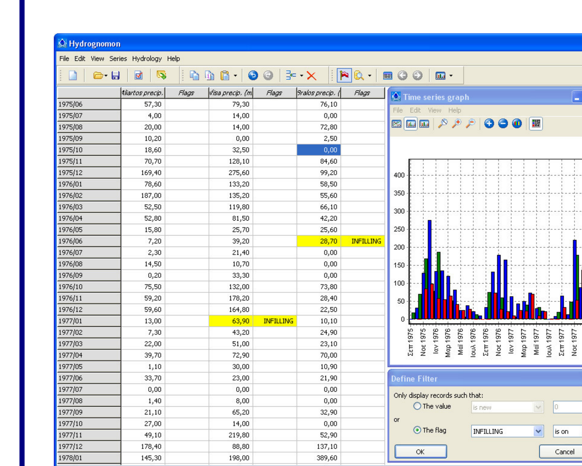
and $\hat a$ the modified maximum likelihood estimate given by the following equation.

$$\hat a = \Big[(1/n)\sum_{i=1}^{n} x_i^{\hat b}\Big]^{1/\hat b}$$

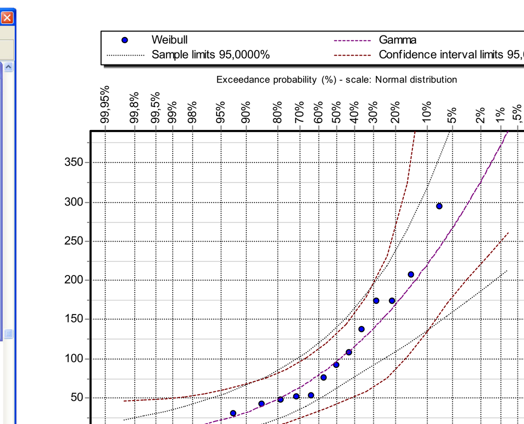
Here we have a simulated sample with 50 elements from a Weibull distribution with $a = 2$ and $b = 3$. For this sample $\hat a = 2.022$ and $\hat b = 3.097$.

(*) Yang Z, Xie M, Wong A (2007) A unified confidence interval for reliability-related quantities of two-parameter Weibull distribution. Journal of Statistical Computation and Simulation 77(5):365-378

11. Implementation of the method in software package 'Hydrognomon'



User interface



Monthly precipitation in Zographou campus. (November 1993-2006)
The sample is modelled by a gamma distribution.

12. Conclusions

- Here we propose a generalized numerical method for the computation of approximate confidence intervals of any distribution. The new algorithm unifies the advantages of two Monte Carlo methods by Ripley (1987).
- The most important characteristic of the method is its generic algorithm, that does not depend on the distribution function.
- Application of the algorithm in many cases and yields confidence intervals better than Ripley's or other approximate confidence intervals.
- We propose this algorithm for a first approximation of an exact confidence interval because it is easily applicable in every case and gives good results.
- Our method has already been applied to the software package 'Hydrognomon' (http://hydrognomon.org). 'Hydrognomon' implements this method to estimate confidence intervals of the parameters and quantiles of about 20 probability distributions.