Estimates of Extremes in the Best of All Possible Worlds

R. R. P. van Nooijen[1] A. G. Kolechkina[1,2]

[1]Faculty of Civil Engineering and Geosciences
Delft University of Technology

[2]Aronwis

3rd StaHy International Workshop on Statistical Methods for Hydrology and Water Resources Management, October 1 & 2 - 2012

Outline

1. Prediction in Hydrology
    - Can it be done?
    - How to see what happens if we do
2. Indicators of uncertainty
    - Formalizing the problem
    - Numerical experiment
    - Bayesian
    - "Frequentist"

TUDelft

Outline

1. Prediction in Hydrology
   - Can it be done?
   - How to see what happens if we do
2. Indicators of uncertainty
   - Formalizing the problem
   - Numerical experiment
   - Bayesian
   - "Frequentist"

Reasons why it should not be done.

For more reasons read Klemeš

- Systems are not stationary.
- Time series are not generated from independent identically distributed random variables.
- Not enough data.

Reasons why it should not be done.

For more reasons read Klemeš

- Systems are not stationary.
- Time series are not generated from independent identically distributed random variables.
- Not enough data.

Reasons why it should not be done.

For more reasons read Klemeš

- Systems are not stationary.
- Time series are not generated from independent identically distributed random variables.
- Not enough data.

Other fields where predictions are made.

- Econometrics.
- Actuarial science.
- Medicine.

Other fields where predictions are made.

- Econometrics.
- Actuarial science.
- Medicine.

Other fields where predictions are made.

- Econometrics.
- Actuarial science.
- Medicine.

Reasons why perhaps it can be done.

- There is theory on extremes for non-stationary time series
- There is theory on extremes for non-iid time series
- There is theory on extremes for time series with correlation

TU Delft

Reasons why perhaps it can be done.

- There is theory on extremes for non-stationary time series
- There is theory on extremes for non-iid time series
- There is theory on extremes for time series with correlation

Reasons why perhaps it can be done.

- There is theory on extremes for non-stationary time series
- There is theory on extremes for non-iid time series
- There is theory on extremes for time series with correlation

Outline

Procedures

- Numerical experiments
- Frequentist approach to quantile as a statistic
- Bayesian approach to quantile as a statistic

Procedures

- Numerical experiments
- Frequentist approach to quantile as a statistic
- Bayesian approach to quantile as a statistic

Procedures

- Numerical experiments
- Frequentist approach to quantile as a statistic
- Bayesian approach to quantile as a statistic

Outline

TUDelft

Notation

- Random variable representing the event: X
- Random vector representing a set of n observations $\vec{Y}$

pdf f_X

cdf F_X

quantile function $Q_X = F_X^{-1}$

TU Delft

A question of Definition

- Suppose cdf, pdf, quantile function are parameter dependent
- For known parameter $\theta = \theta_0$
- $\Pr\left(X \leq Q_X\left(p_0, \theta_0\right) \mid \theta = \theta_0\right) = F_X\left(Q_X\left(p_0, \theta_0\right), \theta_0\right) = p_0$
- But if we *estimate* the parameter then in general
- $\Pr\left(X \leq Q_X\left(p_0, \hat{\theta}\left(\vec{y}\right)\right) \mid \theta = \theta_0\right) \neq p_0$

Gumbel

$$F_{\mathrm{G}}(x \mid \langle \xi, \zeta \rangle) = \exp\left(-\exp\left(-\frac{x-\xi}{\zeta}\right)\right)$$

Outline

1. Prediction in Hydrology
   - Can it be done?
   - How to see what happens if we do
2. Indicators of uncertainty
   - Formalizing the problem
   - Numerical experiment
   - Bayesian
   - "Frequentist"

Numerical experiment

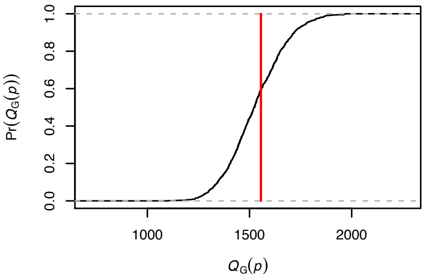


Figure : ecdf of quantile for $p = 0.99$ based on ML parameters, sample size $m = 20$

Outline

1. Prediction in Hydrology
   - Can it be done?
   - How to see what happens if we do
2. Indicators of uncertainty
   - Formalizing the problem
   - Numerical experiment
   - Bayesian
   - "Frequentist"

Quantile distribution

- $\Pr\left(Q_X\left(p_0 \mid \vec{\Theta}\right) \leq t\right)$ is the distribution of the quantile function
- For a given posterior distribution of $\vec{\Theta}$.

Outline

Upper bound with given probability of exceedance.

- A function $b_{\text{ub}}(p, \vec{y})$
- such that
- we have $\Pr\left(X \leq b_{\text{ub}}\left(p, \vec{Y}\right)\right) = p$

Upper bound with given probability of exceedance.

- A function $b_{\text{ub}}(p, \vec{y})$
- such that
- we have $\Pr\left(X \leq b_{\text{ub}}\left(p, \vec{Y}\right)\right) = p$

Upper bound with given probability of exceedance.

- A function $b_{\mathrm{ub}}(p,\vec{y})$
- such that
- we have $\Pr\left(X \leq b_{\mathrm{ub}}\left(p,\vec{Y}\right)\right) = p$

Upper bound on 1:N event size with a probability of exceedance.

- A function $b_{\text{ub}}(\alpha, p, \vec{y})$
- such that
- we have $\Pr\left(Q_X(p) \leq b_{\text{ub}}\left(\alpha, p, \vec{Y}\right)\right) = \alpha$

Upper bound on 1:N event size with a probability of exceedance.

- A function $b_{\mathrm{ub}}(\alpha, p, \vec{y})$
- such that
- we have $\Pr\left(Q_X(p) \leq b_{\mathrm{ub}}\left(\alpha, p, \vec{Y}\right)\right) = \alpha$

Upper bound on 1:N event size with a probability of exceedance.

- A function $b_{\mathrm{ub}}(\alpha, p, \vec{y})$
- such that
- we have $\Pr\left(Q_X(p) \leq b_{\mathrm{ub}}\left(\alpha, p, \vec{Y}\right)\right) = \alpha$

Upper bound on 1:100 event, 10% probability of exceedance, $m = 20$

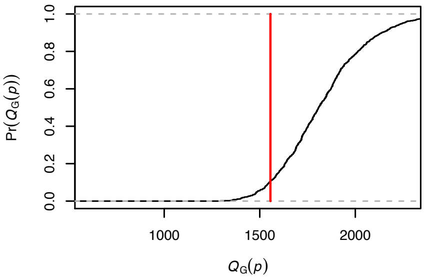
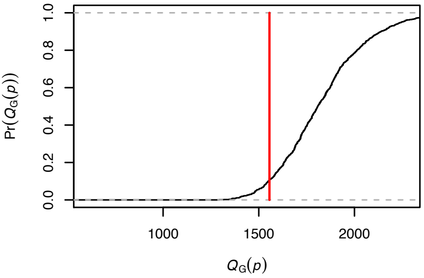
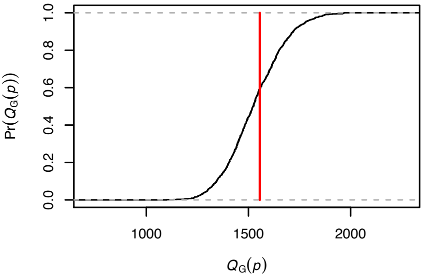


Figure : ecdf of upper bound for quantile estimates for $m = 20$

Comparison


0.0
0.2
0.4
0.6
0.8
1.0
Pr($Q_G(p)$)
1000
1500
2000
$Q_G(p)$


0.0
0.2
0.4
0.6
0.8
1.0
Pr($Q_G(p)$)
1000
1500
2000
$Q_G(p)$

Upper bound on 1:100 event, 10% probability of exceedance, $m = 100$.

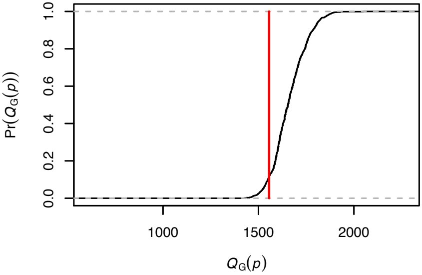


Figure : ecdf of upper bound for quantile estimates for $m = 100$

Summary

- There are ways to display uncertainty.
- For small samples the prediction uncertainty can be large.
- Both "frequentist" and "Bayesian" techniques can be applied.

For Further Reading I

Carlin, B.P. and Luis, T.A.
Bayes and Empirical Bayes Methods for Data Analysis.
Chapman and Hall/CRC, 2000.

Klemeš, V.
Tall tales about tails of hydrological distributions I.
Journal of Hydrologic Engineering 5:227-231, 2000.

Klemeš, V.
Tall tales about tails of hydrological distributions II.
Journal of Hydrologic Engineering 5:232–239, 2000.

Lawless, J.F. and Fredette, M.
Frequentist prediction intervals and predictive distributions.
Biometrika. 92:529–542, 2005.