Generalized exponential distribution in flood frequency analysis

Iwona Markiewicz
Witold G. Strupczewski
Krzysztof Kochanek

3rd STAHY International Workshop on Statistical Methods for Hydrology and Water Resources Management
October 1 & 2, 2012, Tunis, Tunisia
Contents

- Introduction
- Probability of correct selection for GE and IG distributions
  - Discrimination procedures
  - Simulation experiment
  - Results
- Fitting GE distribution and other models to data
  - Case study
  - Flood quantile estimates
  - Best fitting model by various discrimination procedures
- Conclusions
Many distributions have been suggested for fitting the flood extremes data.

**Introduction**

However…
Polish data

39 gauging stations
annual peak flows from 1921-2010, 90-year-long series
$C_v - Cs$ relation for some two-parameter distributions plotted with Polish data
Cv - Cs relation for some two-parameter distributions plotted with Polish data
Inverse Gaussian (IG) has been found to match the number of Polish data successfully.*

Comparison the probability of correct selection (PCS) for GE and IG distributions

<table>
<thead>
<tr>
<th></th>
<th><strong>Generalized exponential</strong></th>
<th><strong>Inverse Gaussian</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PDF</strong></td>
<td>( f(x) = \alpha \lambda \left(1 - e^{-\lambda x}\right)^{(1-\alpha)} e^{-\lambda x} ; \lambda, \alpha, x &gt; 0 )</td>
<td>( f(x) = \frac{\alpha}{\sqrt{\pi x^3}} \exp \left[ - \left( \frac{\alpha - \beta}{\alpha} \right)^2 / x \right] ; \alpha, \beta, x &gt; 0 )</td>
</tr>
<tr>
<td><strong>CDF</strong></td>
<td>( F(x) = \left(1 - e^{-\lambda x}\right)^\alpha )</td>
<td>( F(x; \alpha, \beta) = \frac{1}{2} \left[ 2 - \text{erfc} \left( \frac{-\alpha + x\beta / \alpha}{\sqrt{x}} \right) + \exp(4\beta)\text{erfc} \left( \frac{\alpha + x\beta / \alpha}{\sqrt{x}} \right) \right] )</td>
</tr>
<tr>
<td><strong>Quantile ( F )</strong></td>
<td>( x_F = -\frac{\ln \left(1 - F^{1/\alpha} \right)}{\lambda} )</td>
<td>No analytical form</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>( \mu = \frac{1}{\lambda} \left[ \psi(\alpha + 1) - \psi(1) \right] )</td>
<td>( \mu = \frac{\alpha^2}{\beta} )</td>
</tr>
<tr>
<td></td>
<td>( \psi(.) ) - digamma function</td>
<td></td>
</tr>
<tr>
<td><strong>Variation coefficient</strong></td>
<td>( C_V = \frac{\sigma}{\mu} = \sqrt{\frac{\psi'(1)-\psi'(\alpha + 1)}{\psi(\alpha + 1)-\psi(1)}} )</td>
<td>( C_V = \frac{1}{\sqrt{2\beta}} )</td>
</tr>
<tr>
<td></td>
<td>( \psi(.) ) - trigamma function</td>
<td></td>
</tr>
<tr>
<td><strong>Linear variation coefficient</strong></td>
<td>( LC_V = \frac{\lambda_2}{\lambda_1} = \frac{\psi(2\alpha + 1)-\psi(\alpha + 1)}{\psi(\alpha + 1)-\psi(1)} )</td>
<td>No analytical form</td>
</tr>
</tbody>
</table>
Discrimination procedures

(1) \(K\) procedure

\[
\max \left\{ f_i : \ln \left( L_f(x / \hat{\theta}) \right) \right\}
\]

\(L\) – likelihood function of \(f_i\)

Parameters are estimated by various methods:

- Method of (conventional) moments - MOM
- Method of linear moments - LMM
- Maximum likelihood method - MLM
Discrimination procedures

(2) QK procedure

$$\max \{ f_i : S_i \}$$

$$S_i = \int_0^\infty f_i (\lambda x_1, ..., \lambda x_N) \lambda^{N-1} d\lambda$$

Parameters are estimated by the maximum likelihood method
Discrimination procedures

(3) D procedure

\[ \min \left\{ f_i : D_{i}^{\text{max}} \right\} \]

\[ D_{i}^{\text{max}} = \max_{j=1,...,N} \left( \frac{F(x_{j:N}) - \hat{F}_{j:N}}{1 - F(x_{j:N})} \right) \]

- \( F(x_{j:N}) \) – theoretical probability of the \( j \)-th element of \( x_{1:N} \geq ... \geq x_{N:N} \)
- \( \hat{F}_{j:N} \) – empirical probability given by the Weibull formula: \( \hat{F}_{j:N} = j/(N + 1) \)

Sensitive within the area of probabilities approximating to one!
Discrimination procedures

(4) \( R \) procedure

\[
\begin{align*}
\min \left\{ f_i : R_1^1 \right\}, \min \left\{ f_i : R_2^2 \right\} \\
R_1^1 &= \hat{x}_{1\%(i)}^{MOM} - \hat{x}_{1\%(i)}^{MOM} \\
R_2^2 &= \hat{x}_{1\%(i)}^{MML} - \hat{x}_{1\%(i)}^{MML}
\end{align*}
\]

\( \hat{x}_{1\%(i)}^{MOM} \) — 1% quantile estimated by the method of moments

\( \hat{x}_{1\%(i)}^{MML} \) — 1% quantile estimated by the method of linear moments

\( \hat{x}_{1\%(i)}^{MML} \) — 1% quantile estimated by the maximum likelihood method
Probability of correct selection

GE, IG
\( \mu > 0 \)
\( C_V = 0.3 - 1.5 \)
\( N = 20, 50, 100 \)

MC = 10,000

K, QK procedures

\[ PCS = \frac{\text{number of CS}}{\text{number of MC}} \]
Cv - Cs relation for some two-parameter distributions plotted with Polish data
\( \mu = 0, \sigma = 0.41 \)
Conclusions

- PCS of GE versus IG and vice-versa is low for hydrological sample sizes.
- PCS even falls beyond 50% level for Cv value closed to 0.4, being useless and simple decision rule gives better recognition.
- The use of a discrimination procedure without the knowledge of its performance for the considered PDF may lead to erroneous conclusions (erroneous quantile estimates).
Case study

Kleczany on Ropa; tributary of Wisloka, of Vistula

Annual peak flows from 1921-2010
Case study

Annual peak flows for Kleczany on the Ropa River, 1921-2010

\[ \hat{\mu} = 125.78 \]
\[ \hat{C}_V = 0.7882, \hat{C}_S = 1.6732 \]
\[ \hat{LC}_V = 0.4045, \hat{LC}_S = 0.3266 \]
# Flood quantile estimates

<table>
<thead>
<tr>
<th>Method</th>
<th>Ga</th>
<th>We</th>
<th>GE</th>
<th>IG</th>
<th>LN</th>
<th>LL</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>460.3</td>
<td>448.2</td>
<td>465.7</td>
<td>498.2</td>
<td>497.7</td>
<td>473.4</td>
<td>467.2</td>
</tr>
<tr>
<td>LMM</td>
<td>451.4</td>
<td>429.1</td>
<td>461</td>
<td>531.7</td>
<td>544.2</td>
<td>606.7</td>
<td>689.4</td>
</tr>
<tr>
<td>MLM</td>
<td>419.1</td>
<td>415</td>
<td>424.9</td>
<td>530.3</td>
<td>539.7</td>
<td>716.7</td>
<td>584.9</td>
</tr>
</tbody>
</table>
# Discrimination procedures

## K procedure

<table>
<thead>
<tr>
<th></th>
<th>Ga</th>
<th>We</th>
<th>GE</th>
<th>IG</th>
<th>LN</th>
<th>LL</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>-515.97</td>
<td>-518.04</td>
<td>-515.25</td>
<td>-510.9</td>
<td>-512.09</td>
<td>-521.48</td>
<td>-751.64</td>
</tr>
<tr>
<td>LMM</td>
<td>-515.59</td>
<td>-517.57</td>
<td>-515.01</td>
<td>-510.49</td>
<td>-511.58</td>
<td>-514.65</td>
<td>-535.14</td>
</tr>
<tr>
<td>MLM</td>
<td>-514.9</td>
<td>-517.42</td>
<td>-514.15</td>
<td>-510.49</td>
<td>-511.57</td>
<td>-514.17</td>
<td>-524.72</td>
</tr>
</tbody>
</table>

## QK procedure

<table>
<thead>
<tr>
<th></th>
<th>Ga</th>
<th>We</th>
<th>GE</th>
<th>IG</th>
<th>LN</th>
<th>LL</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLM</td>
<td>-516.57</td>
<td>-519.08</td>
<td>-515.82</td>
<td>-512.15</td>
<td>-513.2</td>
<td>-515.76</td>
<td>-498.28</td>
</tr>
</tbody>
</table>

## D procedure

<table>
<thead>
<tr>
<th></th>
<th>Ga</th>
<th>We</th>
<th>GE</th>
<th>IG</th>
<th>LN</th>
<th>LL</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>1.459</td>
<td>2.257</td>
<td>1.213</td>
<td>0.72</td>
<td>0.862</td>
<td>1.562</td>
<td>1.816</td>
</tr>
<tr>
<td>LMM</td>
<td>1.819</td>
<td>3.867</td>
<td>1.36</td>
<td>0.475</td>
<td>0.535</td>
<td>0.625</td>
<td>0.666</td>
</tr>
<tr>
<td>MLM</td>
<td>4.078</td>
<td>6.061</td>
<td>3.126</td>
<td>0.484</td>
<td>0.555</td>
<td>0.444</td>
<td>1.352</td>
</tr>
</tbody>
</table>

## R procedure

<table>
<thead>
<tr>
<th></th>
<th>Ga</th>
<th>We</th>
<th>GE</th>
<th>IG</th>
<th>LN</th>
<th>LL</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>41.22</td>
<td>33.26</td>
<td>40.82</td>
<td>42</td>
<td>32.19</td>
<td>243.3</td>
<td>117.7</td>
</tr>
<tr>
<td>R2</td>
<td>32.34</td>
<td>14.1</td>
<td>36.08</td>
<td>4.496</td>
<td>1.4</td>
<td>110.1</td>
<td>104.5</td>
</tr>
</tbody>
</table>
Conclusions

✓ The choice of the best fitting model to the chosen annual peak flow series depends on the estimation method and the discrimination procedure. It is characteristic for hydrological size of samples.

✓ Generalized exponential distribution occupies one of the leading positions among distributions used in FFA.
Acknowledgements
This work was financed by the Polish Ministry of Science and Higher Education under the Grant IP 2010 024570 titled “Analysis of the efficiency of estimation methods in flood frequency modeling”.

THANK YOU