The Best Way to Estimate Extreme Percentiles
Stuart Jarvis & Andrew Smith

Actuarial Mathematics Workshop, 19 March 2013, Leicester.
Value-at-Risk: Definitions

• Jorion (2007): “The worst loss over a target horizon such that there is a low, pre-specified probability that the actual loss will be larger”

• Legislative references: Insurance Solvency II Directive:

  The Solvency Capital Requirement ... shall correspond to the Value-at-Risk of the basic own funds of an insurance or reinsurance undertaking subject to a confidence level of 99.5 % over a one-year period.

• Regulation gives little guidance on how to deal with limited data and corresponding uncertainty in models or parameters

• Our presentation seeks to address this
Presentation Overview

• Clarifying the Problem
• Does the estimation method matter?
• Efficiency
• Robustness
• Conclusions
Clarifying the Problem

Find the best estimate of the 99.5%-ile

Possible definitions:
• Unbiased estimate
• Prediction interval
• Confidence interval

“best” could mean:
• Efficient
  • Low variability if distribution family correctly specified
• Robust
  • Approximate 99.5%-ile even if distribution is mis-specified
# Distributions and Estimation Methods

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Density Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss</td>
<td>( \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2}\right) )</td>
</tr>
<tr>
<td>Logistic</td>
<td>( \frac{e^x}{(1 + e^x)^2} )</td>
</tr>
<tr>
<td>Log Pareto(2)</td>
<td>( \frac{2e^x}{(1 + e^x)^3} )</td>
</tr>
<tr>
<td>Student T (df = 4)</td>
<td>( \frac{12}{(4 + x^2)^{3/2}} )</td>
</tr>
<tr>
<td>Gumbel</td>
<td>( \exp[-x - e^{-x}] )</td>
</tr>
<tr>
<td>Location-scale family</td>
<td>( \frac{1}{s} f\left(\frac{x - m}{s}\right) )</td>
</tr>
</tbody>
</table>

**Estimation methods**

- Method of Moments (MOM)
- Probability-weighted moments (PWM)
- Maximum likelihood (MLE)
- Bayesian

**Objectives**

- Substitution
- Unbiased estimate
- Prediction interval
- Confidence interval
Distributions: Densities (shifted and scaled to closest KS)
Cumulative Distribution Functions difference relative to logistic

Note: Largest difference within 4%. These are indistinguishable for small data sizes.
Lie Group Structure of Location-Scale Families

\[ X \sim \text{standard distribution} \]
Location \( m_X = 0 \)
Scale \( s_X = 1 \)

\[
\begin{pmatrix} 1 \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ m_Y & s_Y \end{pmatrix} \begin{pmatrix} 1 \\ X \end{pmatrix}
\]
Location \( m_Y \)
Scale \( s_Y \)

\[
\begin{pmatrix} 1 \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ m_Z & s_Z \end{pmatrix} \begin{pmatrix} 1 \\ X \end{pmatrix}
\]
Location \( m_Z \)
Scale \( s_Z \)

Displacements:
\[
\delta_Y = \log \left( \frac{s_z}{s_y} \right)
\]
\[
\delta_m = \frac{\delta_Y (m_z - m_y)}{s_z - s_y}
\]
Lie Group Structure (Continued)

- In general, affine group $G = \mathbb{R}^n \rtimes GL(\mathbb{R}^n)$ acts on $\mathbb{R}^n$
- For $n=1$, $\mathbb{R}^n \rtimes GL^+(\mathbb{R}^n) = \mathbb{R} \rtimes \mathbb{R}^+$
  - Group action is $(m,s),(n,t) = (m+ns, st)$. Identity is $(0,1)$.
- Lie algebra $g = \mathbb{R}^2$
  - Lie bracket $[(a, b), (c, d)] = (bc - ad, 0)$
- Exponential map $\exp: (a, b) \in g \mapsto \left( a \frac{\exp(b) - 1}{b}, \exp(b) \right) \in G$
- Estimation process is usually $G$-equivariant
  - e.g. estimate $(m, s) \in \mathbb{R}^2$ based on sample $\{X_i\} \in \Omega$
  - $Y = cX + d$ gives rise to estimates $(cm + d, cs)$
Comparing Estimation Methods

Does the Method Matter?
Does the Method Matter?
Logistic Parameter Estimates (sample n=20)
Another way to view the same data
Mean & Stdev displacement plot (Logistic dist)

Notes:
MLE smallest sampling error
Mean estimates unbiased, (unsurprisingly by symmetry)
PWM unbiased (fact)
Downward bias in MOM scale estimate (Jensen)
Displacement Plot for Location / Scale Parameters (Log Pareto 2 distribution, n=20)

Notes:
MLE smallest sampling error but largest bias
Location parameter bias different for MLE and MOM even though mean estimates are the same. PWM unbiased (theorem)

Classical result: MLE asymptotically \((n \rightarrow \infty)\) efficient and unbiased. Our experimental results support MLE efficiency also for small samples.
MLE vs MOM: Displacement Chart (n=20) Choice of Method Matters

- Logistic
- LogPareto
- Student T4
- Gumbel
Alternative Interval Definitions
Alternative Interval Constructions

Background: $k = k(X_1, \ldots, X_{20})$ is the upper limit of an interval. We consider four constructions

<table>
<thead>
<tr>
<th>Process Driven:</th>
<th>Substitution</th>
<th>Statistical criteria:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate $m$ and $s$</td>
<td>E($k$) = true 99.5%-ile</td>
</tr>
<tr>
<td></td>
<td>Calculate 99.5%-ile for fitted distribution</td>
<td>Prob{$X_{21} \leq k(X_1,X_2, \ldots,X_{20})$} = 99.5%</td>
</tr>
<tr>
<td></td>
<td>Unbiased Prediction interval</td>
<td>Prob{ true percentile $\leq k(X_1,X_2, \ldots,X_{20})$ } = 95%</td>
</tr>
</tbody>
</table>
Methods to Consider: Multiple of Stdev based on MOM

- Gauss
- Logistic
- LogPareto2
- Student T4
- Gumbel

Graph showing the distribution of unbiased, prediction, and confidence for various methods.
Prediction Intervals as Multiple of Stdev
Effect of Sample Size (capital for parameter risk)

- Gumbel
- Student T4
- LogPareto2
- Logistic
- Gauss
Efficiency of Prediction Interval Constructions:
(all based on n=20 and 99.5% prediction)
Effects of Model Mis-Specification

Robustness
Robustness for Method of Moments (assuming infinite sample)

1-in-49

1-in-8544

Multiple of sample stdev

1-cdf (log scale)

2 2.5 3 3.5 4

0.1

0.01

0.001

0.0001

20

© 2013 The Actuarial Profession • www.actuaries.org.uk
Robustness Comparison: Different Methods
Infinite Sample Size

<table>
<thead>
<tr>
<th>Method</th>
<th>Lower</th>
<th>Target</th>
<th>Upper</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>49</td>
<td>200</td>
<td>8544</td>
<td>Most robust</td>
</tr>
<tr>
<td>LSCALE</td>
<td>43</td>
<td>200</td>
<td>16303</td>
<td></td>
</tr>
<tr>
<td>MLE(logistic)</td>
<td>37</td>
<td>200</td>
<td>27993</td>
<td>Least robust</td>
</tr>
<tr>
<td>MLE(LogPareto2)</td>
<td>32</td>
<td>200</td>
<td>73573</td>
<td></td>
</tr>
<tr>
<td>MLE (Student T4)</td>
<td>32</td>
<td>200</td>
<td>56443</td>
<td></td>
</tr>
<tr>
<td>MLE (Gumbel)</td>
<td>#N/A</td>
<td>200</td>
<td>#N/A</td>
<td></td>
</tr>
</tbody>
</table>

Notes: In each case, Gauss produces the lowest %-ile and Gumbel the highest, so it turns out that the other distributions don’t affect the robustness criterion. Fitting MLE to Student T4 data, assuming (wrongly) a Gumbel, results in fitted scale parameters tending to infinity for large samples, hence the #N/A in the limit.
## Different Prediction Interval Methods

Value of the “Model Shopping” option

<table>
<thead>
<tr>
<th>Company A</th>
<th>Company B</th>
<th>Company C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always fits logistic using MOM</td>
<td>Uses whatever method produces the lowest answer: currently logistic &amp; MOM</td>
<td>Uses logistic &amp; MOM, and states an intention to continue. Has a track record of citing “new” research every year to support a different method</td>
</tr>
<tr>
<td>Prediction interval gives credit for method stability</td>
<td>Prediction interval uses larger multiple of stdev compared to company A, reflecting the option to switch methods</td>
<td>Auditors would like there to be an agreed “best” method to thwart opportunistic switching (and not because being “best” is important in itself).</td>
</tr>
</tbody>
</table>
Conclusions

- Estimating extreme percentiles from limited data is not a hopeless task but there are several possible definitions and the choice matters.
- Known asymptotic results relate to Fisher-efficiency, a criterion which favours maximum likelihood estimation. However, this is not the whole story and method of moments produces more efficient prediction intervals for small samples.
- Difficult in practice to get MLE to work. In the special case of location scale families, the problem is bounded and we fine-tuned the optimisation to ensure convergence.
- Robustness is difficult to address analytically but Monte Carlo experiments suggest MLE performs worse than method of moments, and both are much less robust than we would like.
- Although practitioners routinely apply hypothesis testing tools, there is much resistance to the frequentist requirement to randomise data sets in order to construct predictions.
- There is a regulatory risk of model shopping and the search for a “best” approach may address this more than optimality.
- Choice of ambiguity set is a social convention rather than a technical choice.
- Analysis of Bayes methods is numerically challenging and we haven’t figured out how to do this yet!
- Further work to test robustness of more advanced methods sensitive to distribution shape.
Acknowledgements and Disclaimers

Our contact details:  andrewdsmith8@deloitte.co.uk
                         stuart.jarvis@blackrock.com

Disclaimer  We are grateful for useful discussions with many friends
and colleagues, including Graeme Alexander, Gabriela
Baumgartner, Paul Coulthard, Antoon Pelsser, Andreas
Tsanakas and fellow members of the Extreme Events
working party. Any views, opinions or errors in this
presentation are ours alone.
The Best Way to Estimate Extreme Percentiles
Stuart Jarvis & Andrew Smith

Actuarial Mathematics Workshop, 19 March 2013, Leicester.