As the **WORLDCOM** turns.

**BITTER:** Well, we lost it all! Kevin got laid off, after working there for over 8 years! We had just bought a house, and had a baby! But to mention: I had just quit my job at WorldCom to be a stay at home mom! We both lost our entire 401K's and Kevin also lost hundreds of thousands of dollars in vested stock options. Things were tough.

**SWEET:** This is where we WRESTLE! This is where we worked when we got married and had our first child! This is where our life together began. So even though we lost all of our money, we gained all of these things.

**KEVIN'S JOB DESCRIPTION:** Kevin started working at WorldCom when it was actually still 1991. It looked like just another gig on Wall St. He got laid off from WorldCom in April of 2002, which was 3 months after 9/11.

---

My Job Description:

I started working at WorldCom in September of 1997. Back then, I was the Director for their Corporate Training, which happened to be the fastest growing area in the world. I also did graphics for the internet site as well. It got my name after I was born in February of 2002!

---

“As an alternative to the traditional 30-year mortgage,
we also offer an interest-only mortgage, balloon mortgage, reverse mortgage, upside down mortgage, inside out mortgage, loop-de-loop mortgage, and the spinning double axle mortgage with a triple lutz.”

---

Gudrun January 2005

326 MEuro loss

72% due to forest losses

4 times larger than second largest
Maximum Likelihood (ML) inference (Coles p. 30-43)

Likelihood function = the function which shows how the “probability” (or likelihood) of getting the observed data depends on the parameters

\[ x_1, \ldots x_n \] observations of i.i.d. variables \( X_1, \ldots X_n \), density \( f(x) = f(x; \theta) \)

\[ \theta = (\theta_1, \ldots \theta_d) \] parameters

\[ L(\theta) = f(x_1; \theta)f(x_2; \theta) \ldots f(x_n; \theta) \] likelihood function

\[ \ell(\theta) = \log f(x_1; \theta) + \log f(x_2; \theta) + \ldots \log f(x_n; \theta) = \sum_{i=1}^{n} \log f(x_i; \theta) \]

ML estimates = the value \( \hat{\theta} = (\hat{\theta}_1, \ldots \hat{\theta}_d) \) which maximizes \( \ell(\theta) \)

- ML estimates often have to be found through numerical maximization
- sometimes a maximum doesn’t exist
- sometimes several local maxima (\( \rightarrow \) problem for numerical maximization)
- but typically no problems if the number of observations is “large”
**Example:** ML estimation of the parameters in the PoT model

\[ T = \text{length of observation period} \]
\[ N = \text{number of observed excesses (random variable!)} \]
\[ x_1, \ldots x_N \text{ observed excess sizes} \]
\[ \theta = (\lambda, \beta, \sigma) \text{ parameters} \]

The probability of observing \( N \) excesses is \( \frac{(\lambda T)^N}{N!} \exp\{-\lambda T\} \), plus independence and previous slide →

\[
L(\theta) = L(\lambda, \sigma, \gamma) = \frac{(\lambda T)^N}{N!} \exp\{-\lambda T\} \prod_{i=1}^{N} \frac{1}{\sigma} \left(1 + \frac{\gamma}{\sigma} x_i\right)_+^{-1/\gamma-1}
\]

\[
\ell(\lambda, \sigma, \gamma) = N \log(\lambda) + N \log(T) - \log(N!) - \lambda T
\]
\[- N \log(\sigma) - \sum_{i=1}^{N} \frac{1}{\gamma + 1} \log \left(1 + \frac{\gamma}{\sigma} x_i\right)_+
\]

\[
\frac{\partial}{\partial \lambda} \ell(\lambda, \sigma, \gamma) = \frac{N}{\lambda} - T = 0 \quad \text{so that} \quad \hat{\lambda} = \frac{N}{T}
\]

\( \hat{\sigma}, \hat{\gamma} \) obtained from numerical maximization of the second part of \( \ell(\lambda, \sigma, \gamma) \)
**ML inference: asymptotic properties**

\[ \mathcal{I}(\theta) = E_\theta(\left( - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta) \right)) \] expected Fisher information matrix, estimated by \( \mathcal{I}(\hat{\theta}) \) or by \( I(\hat{\theta}) \) where \( I(\theta) = \left( - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\theta) \right) \) is the observed Fisher information matrix. (In the expected Fisher information matrix, the observations are replaced by the corresponding random variables when the expectations are computed)

\[ \hat{\theta} = (\hat{\theta}_1, \ldots \hat{\theta}_d) \] asymptotically has a d-dimensional multivariate normal distribution with mean \( \theta \) and variance \( \mathcal{I}(\theta)^{-1} \)

In particular, the variance of \( \hat{\theta}_i \) may be estimated by \( (\mathcal{I}(\hat{\theta})^{-1})_{ii} \) (= the \( i \)-th diagonal element of \( \mathcal{I}(\hat{\theta})^{-1} \)), or by \( (I(\hat{\theta})^{-1})_{ii} \). The latter is often more accurate.

\[ k_{\alpha/2} = \text{the } \alpha/2\text{-th quantile from the top of the standard normal distribution} \]

\( (\hat{\theta}_i - k_{\alpha/2} \sqrt{(I(\hat{\theta})^{-1})_{i,i}}, \hat{\theta}_i + k_{\alpha/2} \sqrt{(I(\hat{\theta})^{-1})_{i,i}}) \) asymptotic 100(1- \( \alpha \))% confidence interval for \( \theta_i \)
ML inference: the delta method

\[ \eta = g(\theta) = g(\theta_1, \ldots, \theta_d) \]  function of the parameters

\[ \hat{\eta} = g(\hat{\theta}) = g(\hat{\theta}_1, \ldots, \hat{\theta}_d) \]  estimate of the function of the parameters

\[ \nabla(\theta) = (\frac{\partial}{\partial \theta_1} g(\theta), \ldots, \frac{\partial}{\partial \theta_d} g(\theta)) \]  gradient, \[ \nabla(\hat{\theta}) \] estimate of gradient

\[ \hat{\eta} \] asymptotically normal with mean \( \eta \) and variance \[ \nabla(\theta) \mathcal{I}(\theta)^{-1} \nabla(\theta)^t \]
(which e.g. can be estimated by \[ \nabla(\hat{\theta}) I(\hat{\theta})^{-1} \nabla(\hat{\theta})^t \].

From this one can construct confidence intervals for \( \eta \) in the same way as the confidence intervals for \( \theta \) on the previous page.
ML inference: Likelihood Ratio (LR) tests

\[ \theta = (\theta_1, \theta_2) \]  partition of \( \theta \) into two vectors \( \theta_1 \) and \( \theta_2 \) of dimensions \( d-p \) and \( p \). \( \hat{\theta}_2^* \) maximizes \( l(\theta_1, \theta_2) \) over \( \theta_2 \), for \( \theta_1 \) “kept fixed” (so function of \( \theta_1 \) )

\[ 2(\ell(\hat{\theta}) - \ell(\theta_1, \hat{\theta}_2^*)) \]  asymptotically has a \( \chi^2 \) distribution with \( d-p \) degrees of freedom if \( \theta_1 \) is the true value  \( \rightarrow \) LR test:

Reject \( H_0 : \theta_1 = \theta_1^0 \) at the significance level 100 \( \alpha \)% if

\[ 2(\log \ell(\hat{\theta}) - \ell(\theta_1^0, \hat{\theta}_2^*)) > \chi^2_\alpha(d-p) \] , where \( \chi^2_\alpha(d-p) \) is the \( \alpha \)-th quantile from the top of the \( \chi^2 \) distribution with \( d-p \) degrees of freedom.
ML inference: profile likelihood confidence intervals
(often more accurate than delta method intervals, plots from Coles)

Profile likelihood confidence intervals for the shape parameter in the Block Maxima model. The delta method probably would give similar interval in the left case, but not in the right.
The real problem!
The problems

How much reinsurance should LFAB buy?

Should LFAB worry about windstorm losses getting worse?

How should LFAB adjust if its forest insurance portfolio grows?

and:

Can detailed modeling give better risk estimates?

Are windstorms becoming more frequent?
**1994 PoT analysis of 1982-1993 LFAB data** *(the basic method, more sophisticated analysis of 1982-2005 data in later paper)*

<table>
<thead>
<tr>
<th>Risk (MSEK)</th>
<th>next year</th>
<th>next 5 years</th>
<th>next 15 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>66</td>
<td>215</td>
<td>473</td>
</tr>
<tr>
<td>1%</td>
<td>366</td>
<td>1149</td>
<td>2497</td>
</tr>
</tbody>
</table>

\[ X_i \sim \text{GP}(y; \sigma_t, \gamma) \]
\[ \sigma_t = \exp(\alpha + \beta t) \]
\[ \hat{\alpha} = 15.1 \]
\[ \hat{\beta} = 0.013 \pm 0.013 \]

no evidence of trend in extremes

conditional probability that a loss in excess of the reinsurance level 850 MSEK exceeds \( x \)

Gudrun: 2912 MSEK, 12 years later

Windstorms of 1902 and 1969 probably comparable to Gudrun
Choice of threshold/number of order statistics in PoT, model diagnostics

Threshold choice compromise between low bias (= good fit of model), which requires high threshold/few order statistics, and low variance, which requires low threshold/many order statistics

- mean excess plots (high variability for heavy tails)
- median excess plots
- plots of parameter estimates as function of threshold/number of order statistics
- qq- and pp-plots

Automatic threshold selection procedures exist, but perhaps not all that reliable (“optimal” threshold depends on the underlying distribution which is unknown and has to be estimated).
Quantiles of GPD

Windstorm losses 1982-1993
excesses of 0.9 MSEK

Storm loss, MSEK

Windstorm losses 1982-1993

Median excess of level

Level

Parameter stability plot, $\hat{\gamma}$

Parameter stability plot, $\hat{\delta} - \hat{\gamma}u$

Windstorm losses 1982-2005

parameter stability plots

Windstorm loss

Quantiles of GPD

Parameter stability plots
Some conclusions

• risk cannot be summarized into one number
• extreme value statistics provide the simplest methods (but other methods may sometimes be needed)
• didn’t find clear trends
• meteorological data didn’t help
• don’t trust computer simulation models unless statistically validated
• companies should develop systematic techniques for thinking about “not yet seen” catastrophes

• put contractual limits to aggregate exposure
A step in another direction:
catastrophe risks

BIG --- ”happens only once”

• can’t adjust and improve as experience is gained
• methods based on means, variances, central limit theory have little meaning
• difficult to keep in mind that catastrophes can (and will!) occur

a gamble --- find the odds of a gamble!


