Extremvärdesteori som ett hjälpmedel att sätta gränser

Blickar bakåt och framåt

Georg Lindgren

Lunds universitet, Matematisk statistik

22 mars, 2012

Cramérsällskapet

''New and growing'' ?

Emil Julius Gumbel, a German mathematician, pacifist and anti-Nazi campaigner developed new distributions in the 1950s. The Gumbel-distribution, the Generalized Extreme Value distribution and the Generalized Pareto Distribution (GPD) are just the tip of the ice-berg of an entire new and quickly growing branch of statistics. The first application was to answer environmental questions, quickly followed by the finance industry.
Extreme Value Theory can save your neck Valérie Chavez-Demoulin ETHZ & Armin Roehrl, Approximity GmbH, (2004)

Springerkatalogen har (minst) sex titlar: ''Statistical extremes''

Kom in i AMS Math Subject classification 2000, 60G70, 62G32

Tidskriften ''Extremes'' startade 1998

NATO-ASI – Statistical extremes and Applications 1983

Avslutning på klassisk extremvärdesteori, start på ny användbar Inferens

Exempel på gränser, Sverige, SO_2 timme

Naturvårdsverkets Miljökvalitetesnormer för svaveldioxid i luft:

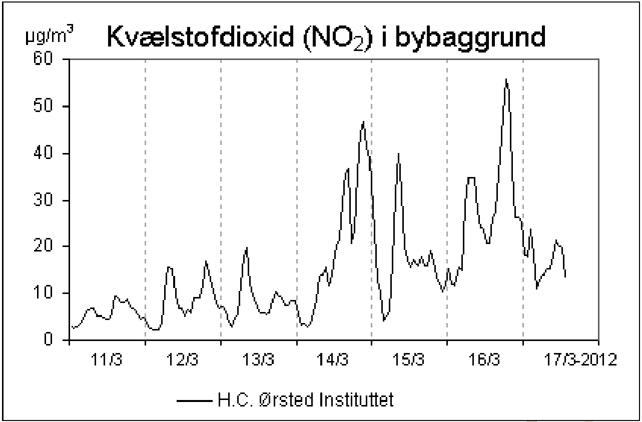
Medelvärdestid	Värde	Anmärkning
1 timme	200 μg/m3	Värdet får överskridas 175 gånger per kalenderår förutsatt att föroreningsnivån aldrig överskrider 350 mikrogram per kubikmeter luft under en timme mer än 24 gånger per kalenderår.

Exempel på gränser, Sverige, SO_2 dygn

Naturvårdsverkets Miljökvalitetesnormer för svaveldioxid i luft:

Medelvärdestid	Värde	Anmärkning
1 dygn	100 μg/m3	Värdet får överskridas 7 gånger per kalenderår förutsatt att föroreningsnivån aldrig överskrider 125 mikrogram per kubikmeter luft under en timme mer än 3 gånger per kalenderår.

Men hur ser det egentligen ut?



Klassisk extremvärdesteori

Vad kan man säga om MAXIMUM/MINIMUM av ett stort antal (oberoende eller beroende) stokastiska variabler?

- Oberoende (Dodd (1923), Fréchet (1927), Fisher & Tippett (1928),
  - Gumbel-fördelningen – "extremvärdesfördelningen", (Cramér: "after some calculation"), för normal(0,1)-variabler:

$$P(\max_{k=1,\dots,n} X_k \leq \sqrt{2\log n} - \dots + \frac{x}{\sqrt{2\log n}}) \to \exp(-e^{-x})$$

  - Weibull-fördelningen – "svagaste länken"
  - Gnedenko – "de tre typerna av extremvärdesfördelningar"
  - von Mises, Jenkinson (1955): Allt i ett, Den Generaliserade extremvärdesfördelningen, GEV, med läges- och skal-parameter plus en formparameter, k:

    $k < 0$, uppåt begränsad (omvänd Weibull); $k = 0$, Gumbel; $k > 0$, "tung-svansad"

Gumbel

”It seems that the rivers know the theory.
It only remains to convince the engineers of the validity of this analysis.”
Gumbel (1958)

Hydrologer och ingenjörer ("safety of structures") accepterade snabbt extremvärdesfördelningarna på 50-, 60-, 70-talen – Waloddi Weibull!

Beroende!

- Beroende variabler bör ge lägre extremvärden
- Intuitionen: I två-dimensionell normalfördelning (X, Y) med korrelation $\rho < 1$, om $X = u$ är mycket stort så är $Y \approx \rho u$
- Slutsats: Finns chans att beroendet inte spelar någon roll!
- Loynes, Leadbetter $\approx$ 1975
- Om $Cov(X_n, X_{n+k})/\log k \to 0$ när $k \to \infty$ så beter sig en beroende normalföljd som om den vore oberoende.

Ett typiskt exempel på 70-talsteori

Ylvisaker - Mittal (1975): Om $Cov(X_n, X_{n+k}) / \log k \to \gamma > 0$ så är maximum fördelat som "Gumbel" + $\sqrt{2\gamma}$ * "Normal"

Vad har detta med miljönormer att göra?

- Poissonprocessen!
- Extrema värden uppträder ungefär som en Poissonprocess
- Finns det ett samband mellan extrema värden och inte riktigt extrema värden?
- Tillåtet: högst 7 dygnsvärden över 100 μg/m3 per år men högst 3 värden över 125 μg/m3

SO_2 i Köpenhamn

För SO_2 i Köpenhamn 1990 var median för timmedelvärdet 20 μg/m3.
Övre 98% kvantilen kan uppskattas till 60 μg/m3.

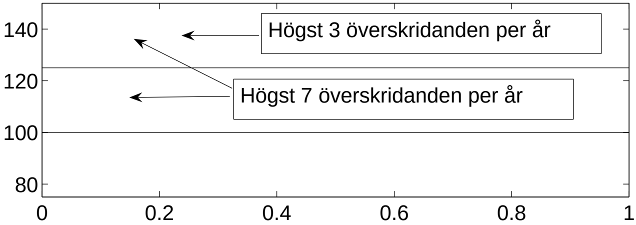
Antag log-normalfördelning ger $E(\log SO_2) = 3$, $D(\log SO_2) = 0.54$.

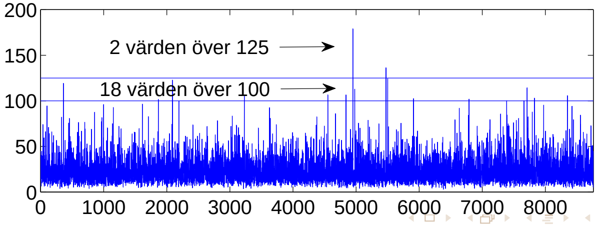
Klassisk extremvärdesteori ger för normerade log årsmaximum
$a_{365*24} = b_{365*24} = 4.26$.

Sannolikheten att årsmaximum < 100 (125) är 43% (87%).

"Fullständig Poissonkonvergens" säger att man kan förstora upp svansen i data och få en Poisonprocess i planet!

Extremvärden uppträder som en Poissonprocess i planet





Fullständig Poissonkonvergens ger nya möjligheter

- Beräkna sannoliheten att kraven uppfylls under ett visst år givet en fördelning.
- Komplikation: beroende mellan timvärden – identifiera kluster av höga värden
- Pekar på svårigheterna att skriva normer.
- Är 8 dagar i rad med värden över 100 (varav 4 över 125) lika allvarligt som 8(4) isolerade överskridanden

Inferens blir allmänt tillgängligt

- Plott på Weibullpapper – MK-metoden
- Momentmetoder – Pickands III – 2:a från vänster
- Likelihood-metoder
- Likelihood-metoder med kovariater
- Standardmetoder för att skatta läge, skala, och formparametrar med kovariater, t.ex. trend, säsong och geografiskt läge

POT, kovariater och bra dataprogram gör extremvärdesanalys användbar

Citat ur Coles:

> *Extreme value theory has emerged as one of the most important statistical disciplines for the applied sciences over the last 50 years. ... At the time of writing, in the past twelve months alone, applications of extreme value modeling have been published in the fields of alloy strength prediction, ocean wave modeling, memory cell failure, wind engineering, management strategy, biomedical data processing, thermodynamics of earthquakes, assessment of meteorological change, nonlinear beam vibration, food science. (2000-2001).*

Intresse från studenter – inte bara från extremister

- Riskanalys - både teknisk och finansiell
- LU kurs i Statistisk analys av extremvärden lockar studenter

kurs	2009/10	2010/11	2011/12
Analys av extremvärden	32	33	55
Tidsserieanalys	28	46	53
Finansiell statistik	23	20	33

Utmaningar

- Multivariata extremvärden
- Extrema "episoder" - extremvärdesteori för former, bilder, händelser
- Spatiala extremvärden – nivåkurvor
- Exakt numerisk beräkning – Alan Genz (Washington State): "My primary area of research interest is the numerical computation of multiple integrals".
  Numeriska rutiner för exakt beräkning av sannolikheter i flerdimensionella normafördelningar: Dimension > 100:

  $$P(X_k \leq x_k; k = 1, \ldots, 100)$$

  med $(X_1, \ldots, X_{100}) \in N(m, \Sigma)$

Katz & Naveau, 2010

Today, weather and climate extremes remain a fertile area of application, with virtually limitless data and a wide variety of extremal behavior . . . requires treatment of spatial fields of observations which evolve over time. . . . with external (especially the sun) and internal drivers and dynamical constraints with variations over a wide range of spatial and temporal scales. Further, the specter of global climate change suggests that non-stationarity need be considered in any treatment of extremes. As such, weather and climate extremes pose a number of problems which could serve as a catalyst for further developments in extreme value theory. Despite much effort, this theory does not yet even provide a fully satisfactory spatio-temporal model for extremes. Still climate scientists, in collaboration with statisticians, need to address critical questions about weather and climate extremes.

References

- S. Coles (2001): *An Introduction to Statistical Modeling of Extreme Values*, Springer
- R.W.Katz & P.Naveau (2010): Editorial: special issue on extremes in weather and climate. *Extremes*, **13**, 107–108.
- A.Stephenson & E.Gilleland (2006): Software for the analysis of extreme events: The current state and future directions. *Extremes*, **8**, 87–109.
- WAFO-group, (2011). *WAFO – a Matlab Toolbox for Analysis of Random Waves and Loads; Tutorial for WAFO Version 2.5*, URL http://www.maths.lth.se/matstat/wafo