MAS2317/3317:
Introduction to Bayesian Statistics

Dr. Lee Fawcett

Case Study 2: Bayesian Modelling of Extreme Rainfall Data

Semester 2, 2014—2015

Motivation

Over the last 30 years or so, interest in the use of statistical methods for modelling **environmental extremes** has grown dramatically, for good reason.

Climate change has resulted in

- an increase in severity, and
- an increase in frequency,

of environmental phenomena resulting in huge economic loss, and loss of human life.

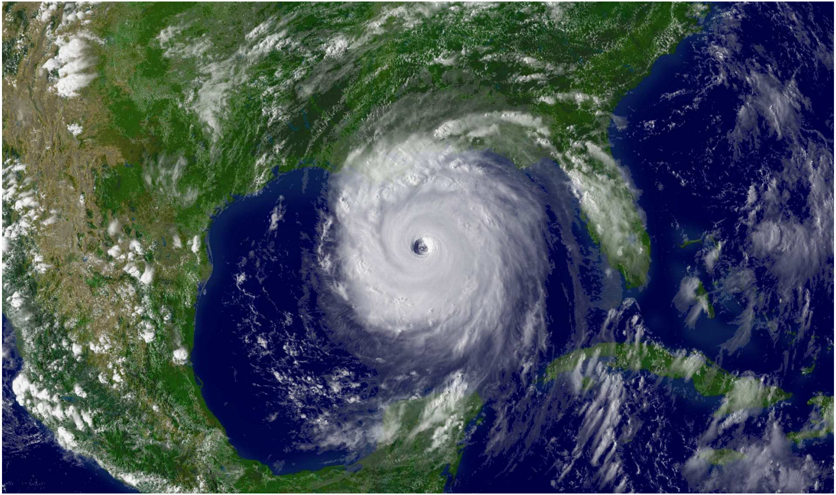
Motivation

For example, **Hurricane Katrina** (see Figure 1) hit southern states of the USA in September 2005,

- killing nearly **2000** people
- displacing well over **one million** people
- costing the US economy an estimated **$ 110 billion**

Billed as the “storm of the century” – just a few weeks later, **Hurricane Rita** battered Texas and Louisiana.

Sea–surge: Hurricane Katrina, 2005



Sea–surge: Hurricane Katrina, 2005

Sea–surge: Hurricane Katrina, 2005

Other environmental extremes

- Extreme **Extreme drought** in Sub–Saharan Africa $\longrightarrow$ famine, huge loss of life, civil war
- Extreme **cold spells** in Russia/China $\longrightarrow$ difficult to stockpile enough fuel
- Rapid shifts in climate can lead to **landslides** $\longrightarrow$ Venezuela, 2010

Closer to home

The **Great storm** of 1987

- Southern England
- 22 deaths
- £7.3 billion worth of damage
- Seemed to come as a surprise...

  `http://www.youtube.com/watch?v=uqs1YXfdtGE`
- Dubbed the UK's **Storm of the century** – two years later, the same type of storm hit the UK

Wind damage from UK storms

Closer to home: recent flooding

Over the past few years, there has been several extreme flooding events in several parts of the UK:

- North–west England, 2008 and 2009
- Central/South–west England 2007–2009
- Seem to be getting more **severe** and more **frequent**
- Loss of life, huge economic burden, including massive flood insurance premiums

Rainfall: Flooding in North–West England, 2009

- £100 million worth of damage
- A number of deaths
- Massive transport disruption

Rainfall: Flooding in Central England, 2008

Rainfall: Flooding in Newcastle, June 2012

Rainfall: Flooding in Newcastle, June 2012

Rainfall: Flooding in Newcastle, June 2012

Rainfall: Flooding in Newcastle, June 2012

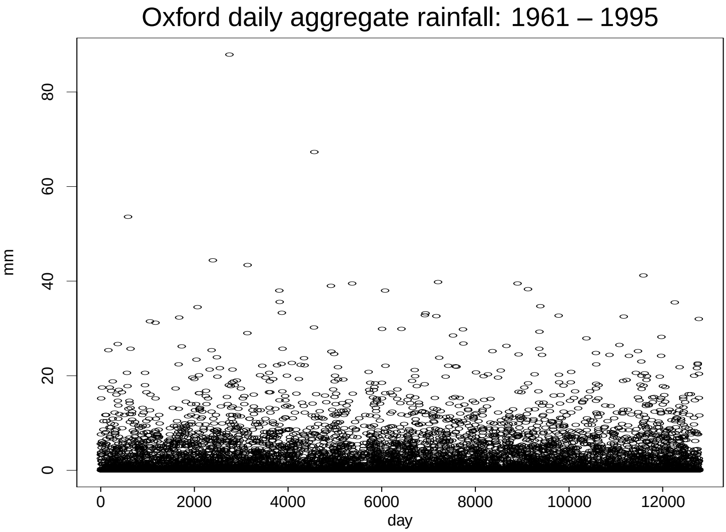
Rainfall: The Great North Sea Flood, 1953

Rainfall: The Great North Sea Flood, 2025?

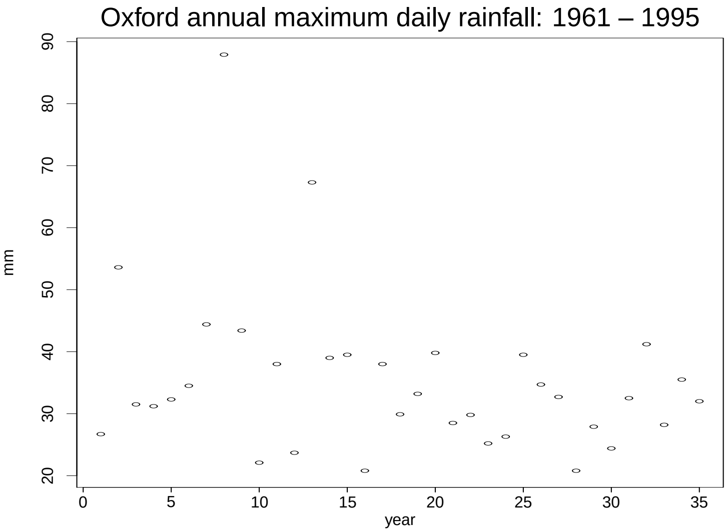
Statistical modelling of extreme rainfall data

- For the rest of this case study, we will focus on **extreme rainfall in the UK**
- In 2003, we were supplied with rainfall data for **204 sites** in the UK
  - daily rainfall accumulations
  - $1961 \rightarrow 1995$
  - Nearly 13,000 observations for each site!
  - However, not interested in most of them – e.g. zero values or indeed anything **non–extreme**!
- **Idea:** Extract annual maxima!

Statistical modelling of extreme rainfall data



Statistical modelling of extreme rainfall data



A statistical model for extremes

The **Generalised Extreme Value** distribution (GEV) – independently derived by von Mises (1954) and Jenkinson (1955).

Provides a limiting model for extremes of stationary series.

Has CDF

$$F_X(x|\mu,\sigma,\gamma) = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\},$$

where μ, σ and γ are **location**, **scale** and **shape** parameters.

- What data do we use for the "extremes", x?
- Can use the extracted annual maxima!

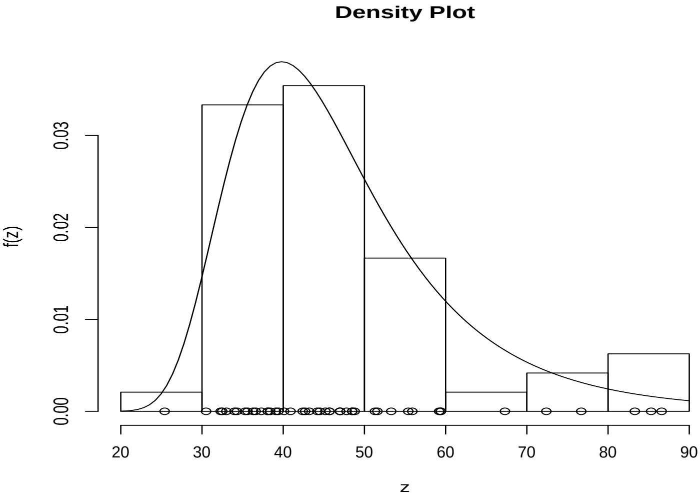
A statistical model for extremes

Estimating the GEV parameters:

- Usual approach: maximise the likelihood w.r.t. each of μ, σ and γ
- No closed–form solutions for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\gamma}$
- Use R – Newton–Raphson type procedure
- This gives

$$\hat{\mu} = 40.8(1.58) \qquad \hat{\sigma} = 9.7(1.19) \qquad \hat{\gamma} = 0.1(0.11)$$

A statistical model for extremes



Practical use of the GEV

So we have a statistical model for extremes which seems to fit our annual maximum daily rainfall data quite well.

So what?

One practical application of such a model is to aid the **design of flood defences**. For example:

- Suppose we wish to protect a town (Oxford?) against a flooding event we would expect to occur once every hundred years
- We only have **35 years** worth of data
- In effect – trying to estimate a flooding event which is more extreme than has ever occurred before

Practical use of the GEV

- This requires **extrapolation** beyond the range of our data
- There is both a **theoretical** and **practical** basis for using the GEV here
- We can estimate such quantities by calculating **high quantiles** from our fitted distribution.

Practical use of the GEV

For our Oxford rainfall extremes, solve the following for $\hat{z}_{100}$:

$$\exp\left\{-\left[1+0.1\left(\frac{\hat{z}_{100}-40.8}{9.7}\right)\right]^{-1/0.1}\right\}=0.99,$$

where $\hat{z}_{100}$ is known as the **100 year return level**.

- A flood defence would need to be tall enough to withstand a daily rainfall total of at least $\hat{z}_{100}$ mm
- Storm systems might last longer than one day
- Calculation of the height of the flood defence would have to take into account the accumulation of successive daily rainfall totals $\hat{z}_{100}$ mm
- The height of the flood defence would be a function of $\hat{z}_{100}$ and the duration of the storm event

Practical use of the GEV

Generically, we have

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\gamma}} \left\{ \left[-\ln\left(1 - \frac{1}{r}\right) \right]^{-\hat{\gamma}} \right\}.$$

- Can obtain **standard errors** via likelihood theory to account for uncertainty in our estimates
- Can then form **confidence intervals**

r (years)	**10**	**50**	**200**	**1000**
$\hat{z}_r$	65.54	87.92	98.64	140.34
	(4.53)	(11.48)	(16.22)	(41.83)

Practical use of the GEV

A Bayesian perspective

Drawback of frequentist approach/beauty of Bayesian approach

- GEV parameter estimates, and resulting estimated return levels, have large standard errors
  - Reduced the sample from about 13,000 observations to just 35
  - 95% CI for $\hat{z}_{1000}$: (58, 222)mm
  - Engineers don't like this:

    "Design your flood defence so that it will withstand a daily rainfall total of somewhere between 58 and 222 mm"
- A Bayesian analysis allows us to incorporate expert information
  - By this point in the course you should know that this is the right thing to do!
  - But it can also reduce estimation uncertainty!

Bring in the expert!

Duncan Reede: independent consulting hydrologist

- PhD (Newcastle, 1977) in Applied Science
- Over 30 years experience
- Can he give us **prior distributions** for μ, σ and γ?
  - Probably not...
  - Very difficult to express your prior opinion about likely values of the “shape” parameter γ...
  - ... perhaps easier for μ?

Bring in the expert!

Idea: Re–express our GEV in terms of parameters Dr. Reede will feel comfortable with – perhaps **return levels**!

"What sort of daily rainfall accumulation would you expect to see, at Oxford, in a storm that might occur once in ten years?"

"... 60mm–65mm? Range might be 50mm$\longrightarrow$80mm..."

Can use the *MATCH* tool to hep here:

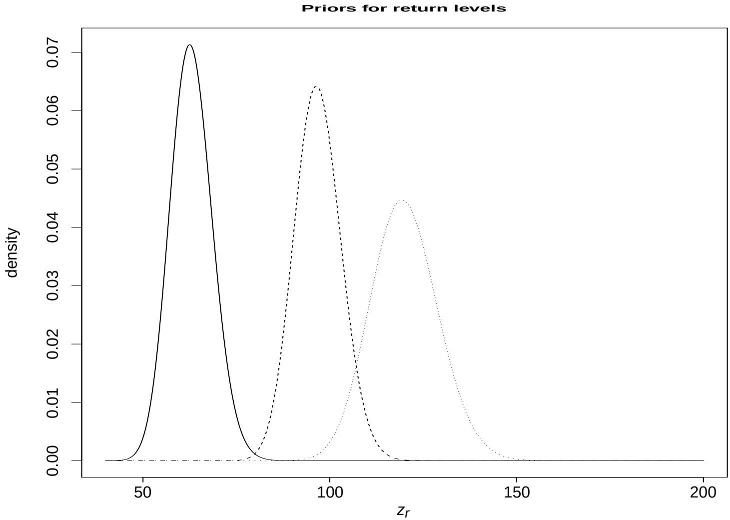
`http://optics.eee.nottingham.ac.uk/match/uncertainty.php`

Bring in the expert!

We get:

$$
\begin{aligned}
z_{10} &\sim Ga(126, 2) \qquad \text{and} \\
z_{50} &\sim Ga(242, 2.5) \\
z_{200} &\sim Ga(180, 1.5)
\end{aligned}
$$

Bring in the expert!



Converting to a prior for (μ, σ, γ)

We use a result from Distribution Theory (**Equation 3.7** from lecture notes) to “convert” the expert’s priors into a prior for (μ, σ, γ).

This gives an **improper**, **non–conjugate** prior for the GEV.

Obtaining the posterior distribution

- Non–conjugate prior for the GEV
- Cannot find the posterior analytically
- Use **Markov chain Monte Carlo** (MCMC – see *MAS3321: Bayesian Inference*)
- This gives a sample from the posteriors for μ, σ and γ
- Apply **Equation (3)** to obtain posterior distribution for return levels

Results

r (years)	**10**	**50**	**200**	**1000**
$E(\hat{z}_r \mid \boldsymbol{x})$	64.21 (2.14)	91.05 (6.31)	110.31 (8.05)	150.73 (14.79)
$\hat{z}_r$	*65.54 (4.53)*	*87.92 (11.48)*	*98.64 (16.22)*	*140.34 (41.83)*

r (years)	**10**	**50**	**200**	**1000**
Bayesian	(60.0,68.4)	(78.7,103.4)	(94.5,126.1)	(121.7,179.7)
Frequentist	*(56.7,74.4)*	*(65.4,110.4)*	*(66.8,130.5)*	*(58.3,222.4)*

Conclusions

- Incorporating the beliefs of an **expert hydrologist**
  - gives us a **more informed analysis** (somewhere "between the prior and the data")
  - dramatically **reduces our uncertainty** about estimates of return levels
- Engineers designing flood defences like this!

Conclusions

- Difficult to get an expert to quantify their uncertainty about things like “shape parameters”
  - Got round this by re–parameterising to something the expert would feel **more comfortable with**, and used *MATCH*
  - Then we can “convert back” to get our prior for (μ, σ, γ)
  - Then a **Bayesian analysis** follows
- The Statistician then feeds back their results to the Marine Engineers designing the flood defence system – they usually build to a height specified by the upper endpoint of a **95% Bayesian confidence interval** for $\hat{z}_r$!