Experience vs Data

Markus Gesmann
Bayesian Mixer, 12 February 2016

My story for today

- How to asses risk with small data sets
- Three examples from insurance pricing
  - Using Bayes, Belief Networks and MCMC
  - R packages: gRain, RStan

The insurance data conundrum

- Insurance companies have many customers
- But most customers have very few claims
- How do you price risk?

How do you set the price?

- Three options:

1. Start with the costs of the insured
2. Start with the perceived value to the insured
3. Call your competitor and ask for the price

Example: Motor insurance

- We have clustered policyholders into ‘good’ and ‘bad’ drivers.

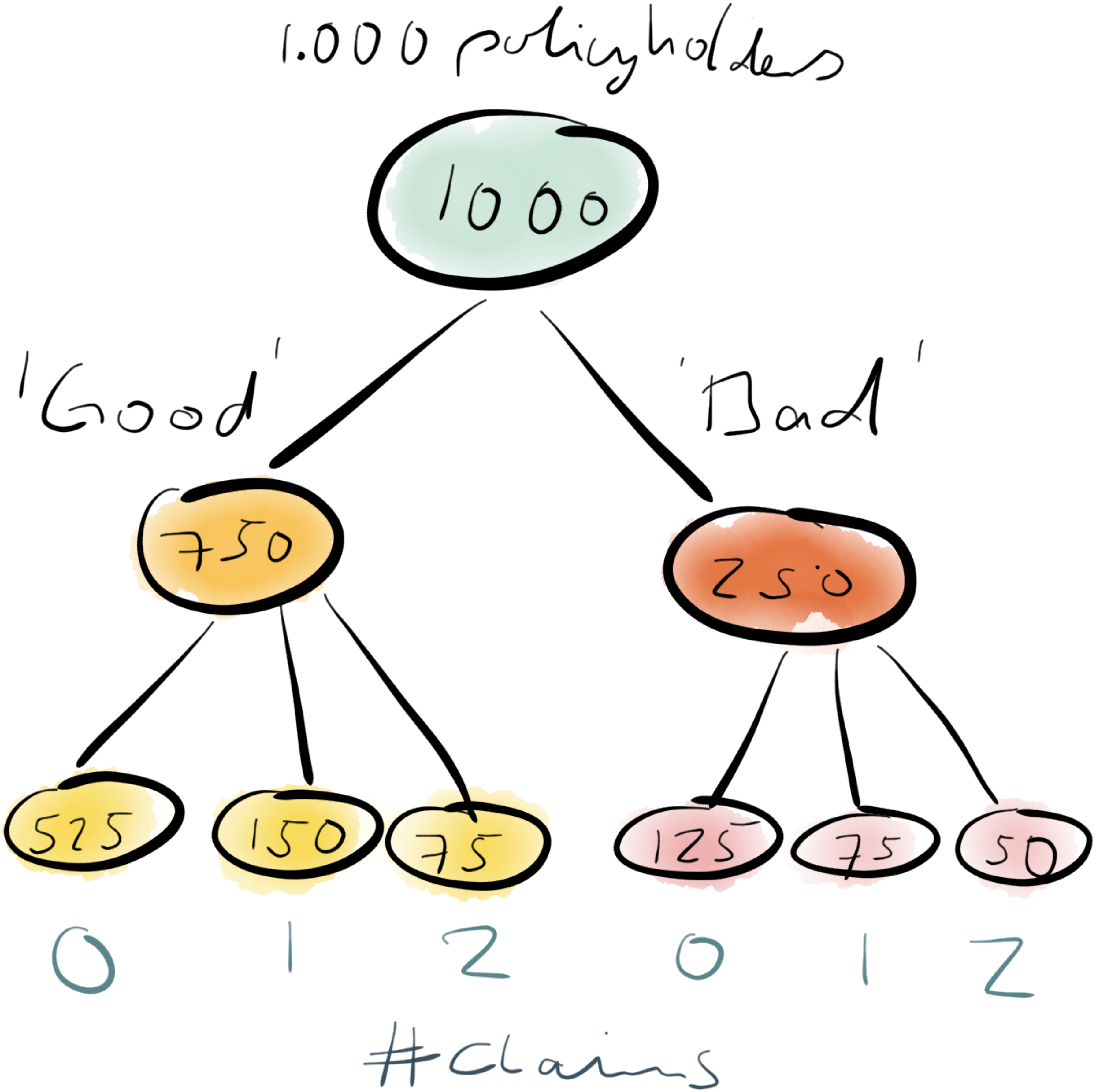
Average number of claims per year	Frequency for ‘Good’ drivers	Frequency for ‘Bad’ drivers
0	70%	50%
1	20%	30%
2	10%	20%

- Of our policyholders 75% are categorised as ‘good’, 25% as ‘bad’.

Source: Loss Models: From Data to Decisions

Questions

- How many claims would you expect from 1,000 policyholders in a year?
- How many claims would you expect from a random policyholder in a year?


1.000 policyholders
1000
'Good'
'Bad'
750
250
525
150
75
125
75
50
0
1
2
0
1
2
#claims

Expected number of claims for one random policyholder

$$
\begin{aligned}
&75\%(0\cdot 70\% + 1\cdot 20\% + 2\cdot 10\%)+\\
&25\%(0\cdot 50\% + 1\cdot 30\% + 2\cdot 20\%)\\
=&0.475
\end{aligned}
$$

Customer asks for his renewal

- The customer is a policyholder of yours for the last two years.
- He had one claim over those two years.
- How many claims should we expect next year?

Thomas Bayes can help

$$P(H|D) = \frac{P(H)\,P(D|H)}{P(D)}$$

What is our hypothesis?
What is our data?

H = “Customer is a ‘good’ driver”

D = “1 claim in two years”
= {(no claim in year 1 & one claim in year 2),
(one claim in year 1 & no claim in year 2)}
= {(1,0), (0,1)}

Prior probability

$$P(H) = 75\%$$

Likelihood

$$p(D \mid H) = p(\{(1,0),(0,1)\} \mid H)$$

$$= p(\{0\} \mid H)\, p(\{1\} \mid H) +$$

$$p(\{1\} \mid H)\, p(\{0\} \mid H)$$

$$= 70\% \cdot 20\% + 20\% \cdot 70\%$$

$$= 28\%$$

Data: Sum over all hypothesises

$$p(D) = \sum_i p(D|H_i)\,p(H_i)$$

Data: Sum over all hypothesises

$$p(D) = p(D|H)p(H) + p(D|\overline{H})p(\overline{H})$$

$$= 28\% \cdot 75\% + (50\% \cdot 30\% + 30\% \cdot 50\%) 25\%$$

$$= 28.5\%$$

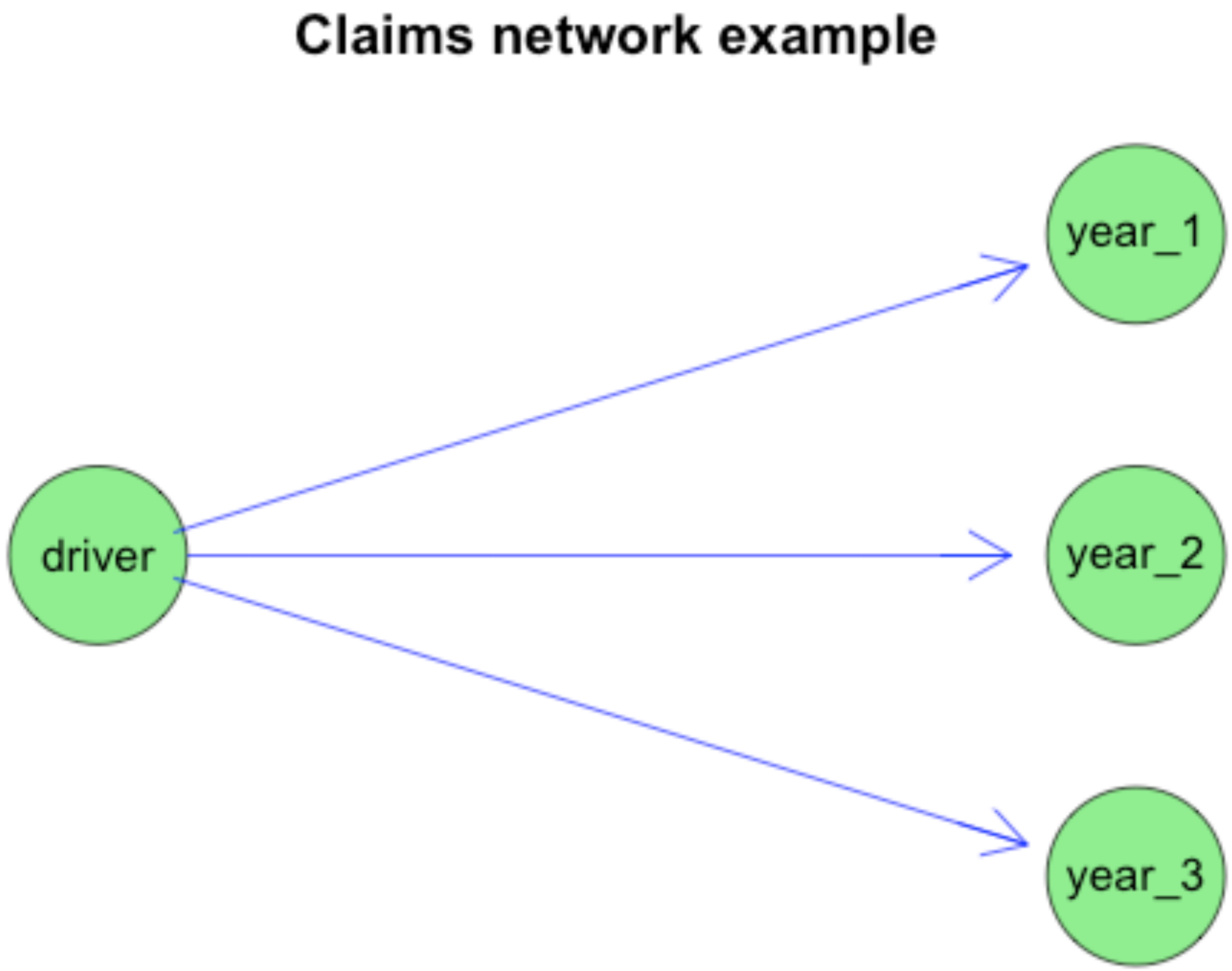
Probability that customer is a ‘good’ driver given that he had one claim in two years.

$$P(H|D) = \frac{P(H)\,P(D|H)}{P(D)}$$

$$= \frac{75\% \cdot 28\%}{28.5}$$

$$= 73.7\%$$

Alternatively think of this as a belief network



Define network in R

```
library(gRain)
library(Rgraphviz)
# Distribution of good and bad drivers
d <- cptable(~ driver, values=c(0.75, 0.25),
                       levels=c("good","bad"))
claims <- c("0", "1", "2")
cond.prop <- c(0.7, 0.2, 0.1, 0.5, 0.3, 0.2)
c1 <- cptable(~ year_1|driver, values=cond.prop, levels=claims)
c2 <- cptable(~ year_2|driver, values=cond.prop, levels=claims)
c3 <- cptable(~ year_3|driver, values=cond.prop, levels=claims)
plist <- compileCPT(list(d, c1, c2, c3))
pn <- grain(plist)
plot(pn[["dag"]], main="Claims network example",
     attrs = list(node = list(fillcolor = "lightgreen"),
                  edge = list(color = "blue"),
                  graph = list(rankdir = "LR")))
```

Set evidence

```
pn <- setEvidence(pn, nslist = list(year_1 = "0",
                                    year_2 = “1"))

querygrain(pn, nodes = "driver", type = “marginal")
$driver
driver
     good       bad
0.7368421 0.2631579
```

Predicting mid-air collisions

- The airline industry grew rapidly in the 1950s
- L.H. Longley-Cook was asked to price the risk for a mid-air collision of two planes
- All Longely-Cook knew was that there were no collisions in the previous 5 years

Source: Computational Actuarial Science with R

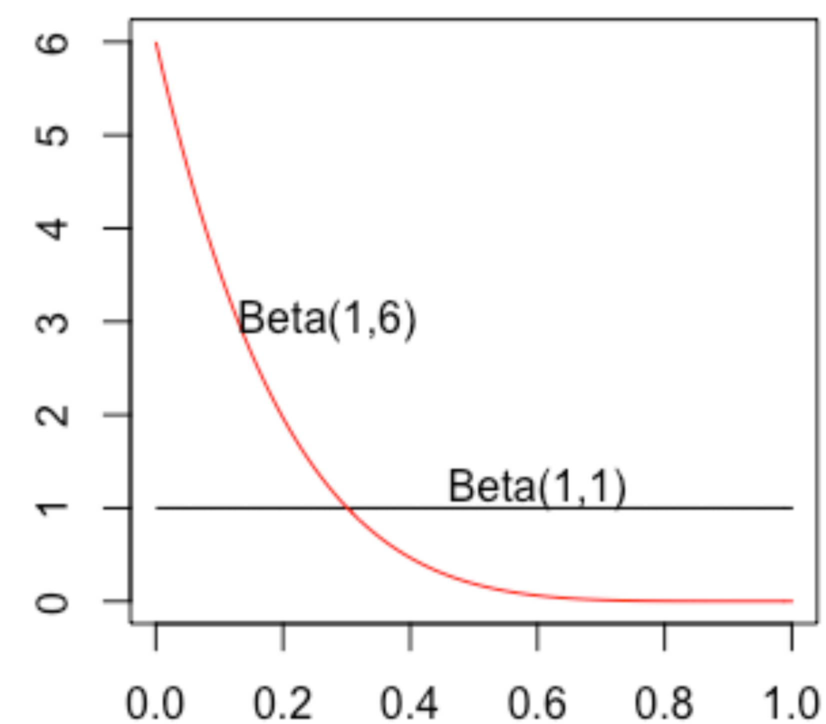
How do you think about this?

- Let's think of the years as a series of Bernoulli trials with unknown probability p

- Start with an uninformed prior, such as a Beta(α,β), with α=1, β=1 and mean $p_0 = \alpha/(\alpha + \beta)$=1/2

- Use the concept of conjugate prior to update:
  $\alpha' = \alpha + \sum x_i = 1$, $\beta' = \beta + n - \sum x_i = 6$

- Posterior predictive mean:
  $p' = 1/7 = 14.3\%$



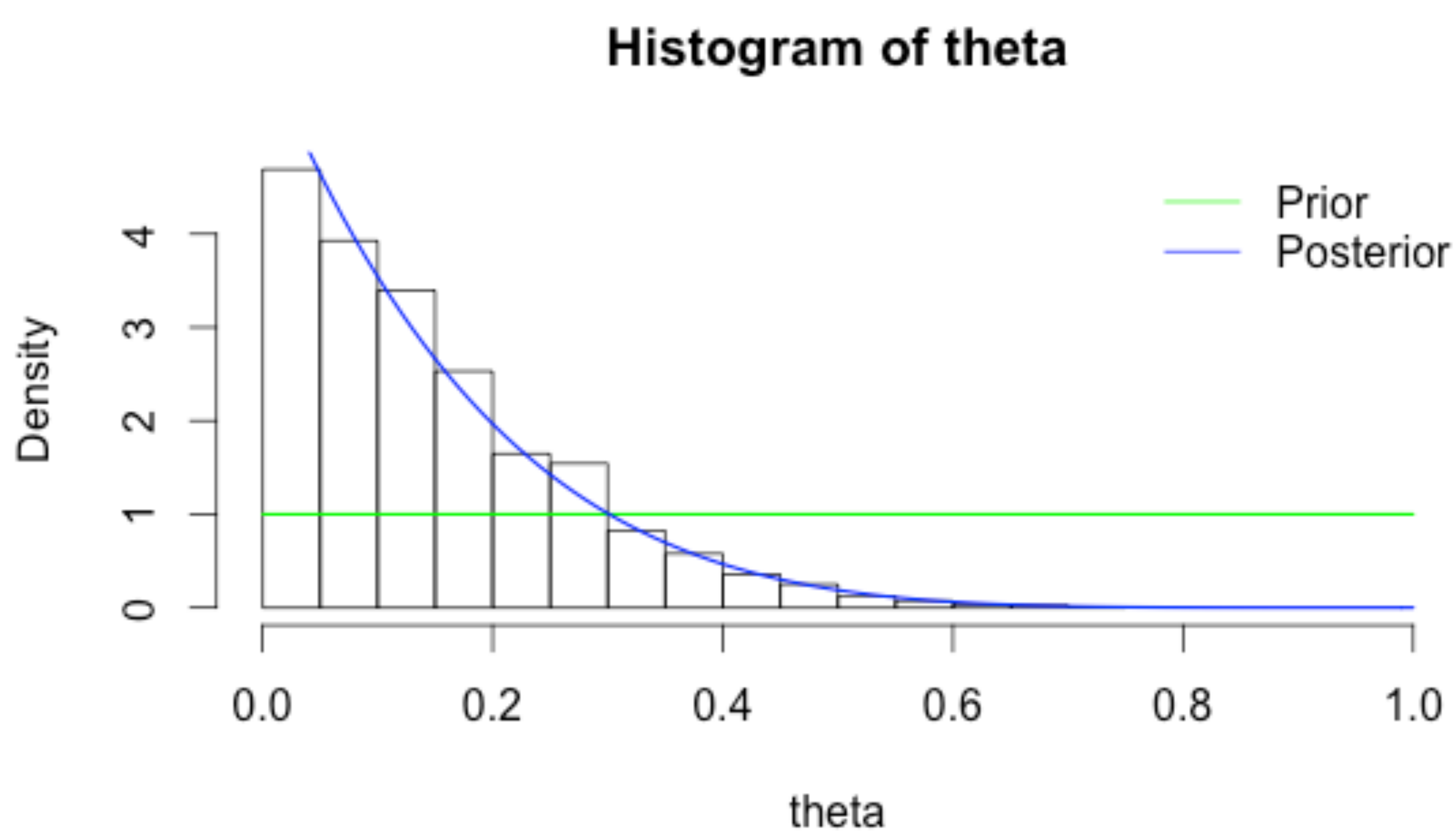
Or use R/Stan

```
library(rstan)
stanmodelcode <- "
data {
  int<lower=0> N;
  int<lower=0, upper=1> y[N];
  }
  parameters {
  real<lower=0, upper=1> theta;
  }
  model {
  theta ~ beta(1, 1);
  for (n in 1:N)
  y[n] ~ bernoulli(theta);
  }
"
fit <- stan(model_code=stanmodelcode, model_name="Longley-Cook",
            data = list(N = 5, y = rep(0,5)))
```

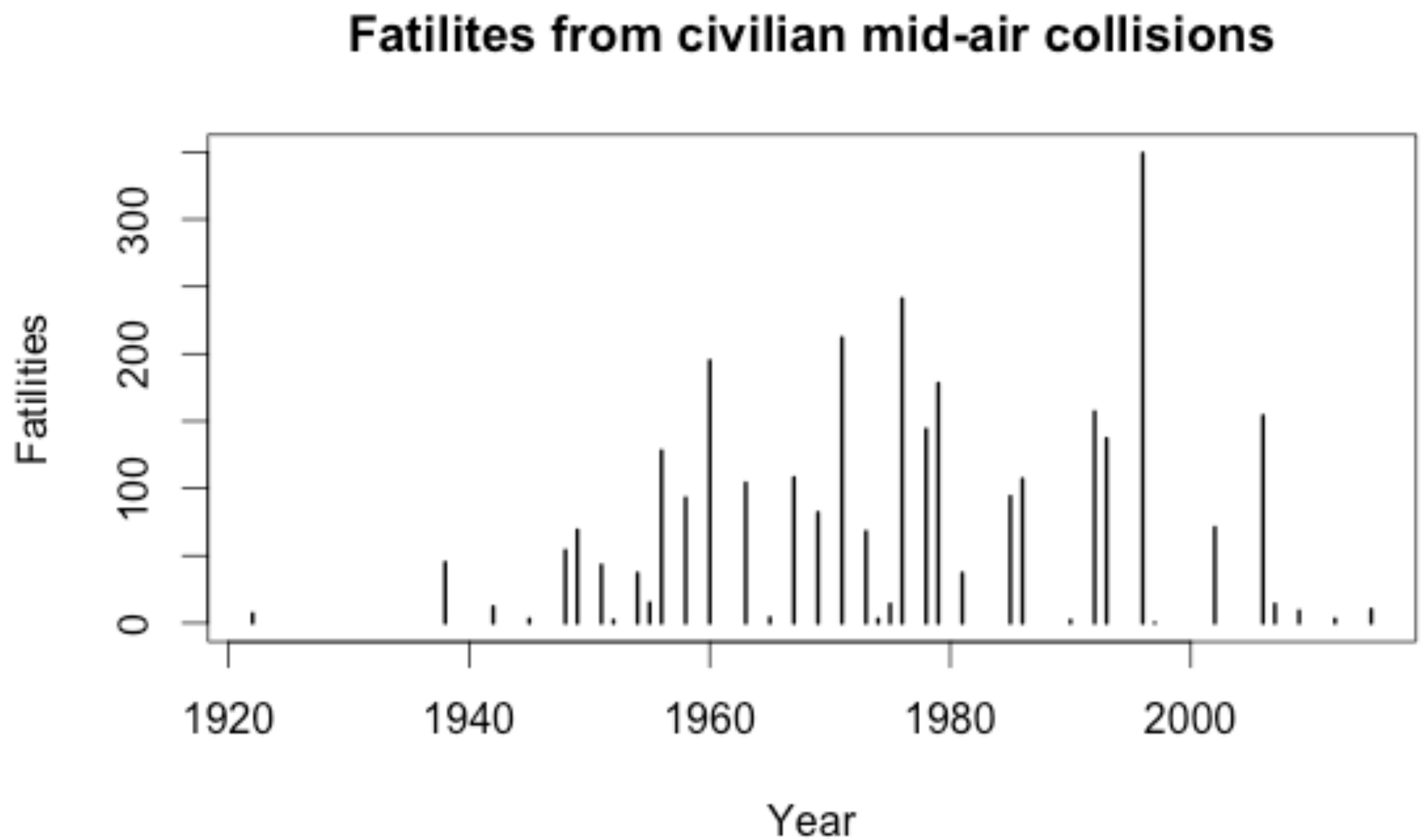
Review model output

```
print(fit, probs=c(0.25, 0.5, 0.75, 0.9))
Inference for Stan model: Longley-Cook.
4 chains, each with iter=2000; warmup=1000; thin=1;
post-warmup draws per chain=1000, total post-warmup draws=4000.

       mean se_mean   sd   25%   50%   75%   90% n_eff Rhat
theta  0.14    0.00 0.12  0.05  0.11  0.21  0.31  1374    1
lp__  -3.39    0.02 0.72 -3.56 -3.10 -2.93 -2.88  1043    1
```

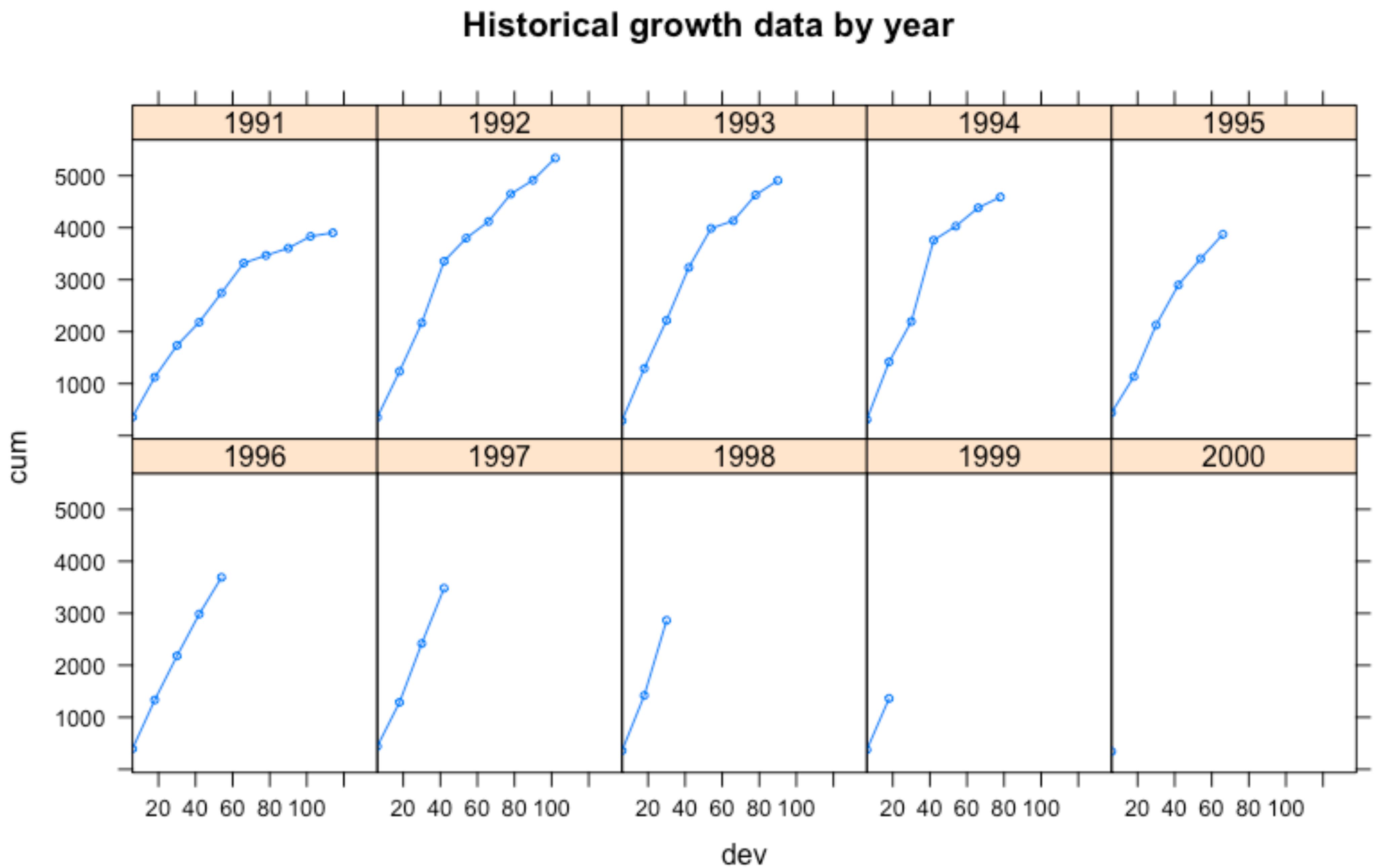


Since 1955 there were 11 incidents with more than 100 fatalities



Source: http://en.wikipedia.org/wiki/Mid-air_collision

Growth curves

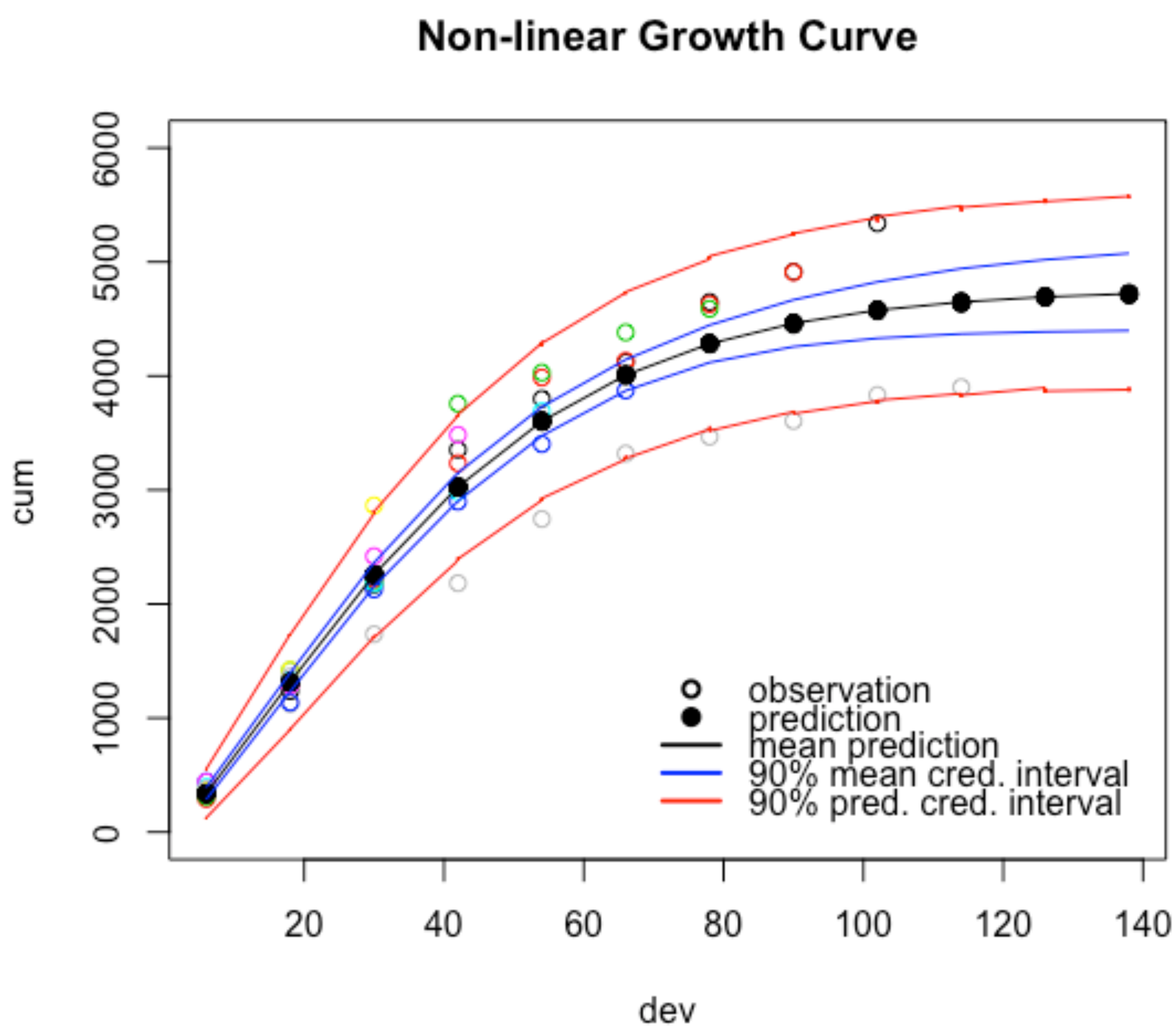


Pooled model

$$G(dev|\omega, \theta) = 1 - \exp\left(-\left(\frac{dev}{\theta}\right)^{\omega}\right)$$

$$\begin{aligned} CL_{AY,dev} &\sim \mathcal{N}(\mu_{dev}, \sigma_{dev}^2) \\ \mu_{dev} &= Ult \cdot G(dev|\omega, \theta) \\ \sigma_{dev} &= \sigma\sqrt{\mu_{dev}} \end{aligned}$$

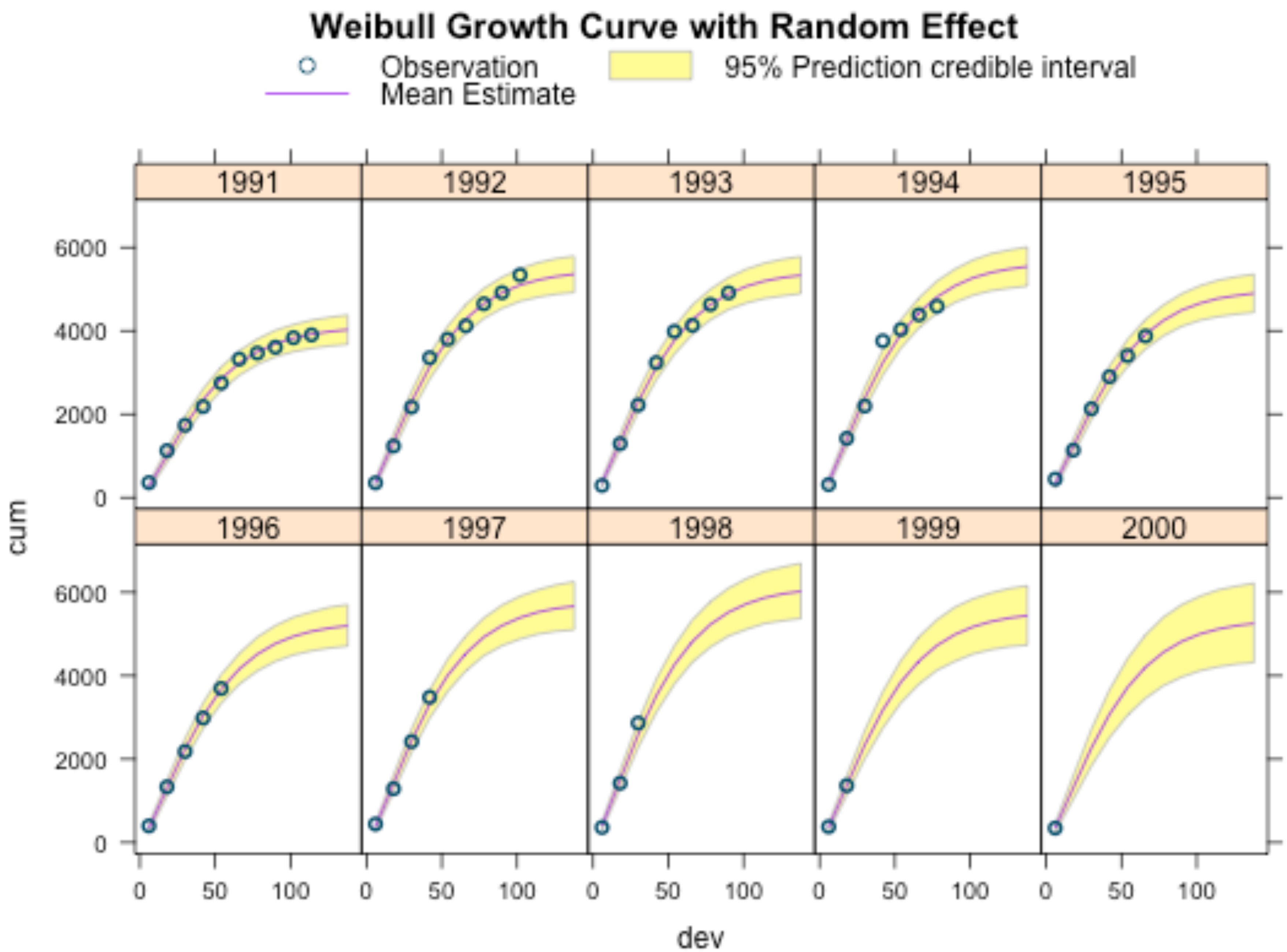
Pooled model



Hierarchical model

$$
\begin{aligned}
CL_{AY,dev} &\sim \mathcal{N}(\mu_{AY,dev}, \sigma^2_{dev}) \\
\mu_{AY,dev} &= Ult_{AY} \cdot G(dev|\omega, \theta) \\
\sigma_{dev} &= \sigma\sqrt{\mu_{dev}} \\
Ult_{AY} &\sim \mathcal{N}(\mu_{ult}, \sigma^2_{ult}) \\
G(dev|\omega, \theta) &= 1 - \exp\left(-\left(\frac{dev}{\theta}\right)^{\omega}\right)
\end{aligned}
$$

Hierarchical model



Conclusions

- More data is often better
- More thinking time is even better
- Bayesian concepts can turbo charge ‘little’ data/ beliefs by borrowing insight from other ‘bigger’ data

R in Insurance Conference

- Cass Business School, London
- 11 July 2016
- Keynote speakers:
  - Mario Wüthrich (RiskLab, ETH Zürich)
  - Dan Murphy (Trinostic, San Francisco)
- www.rininsurance.com

Cass Business School
CITY UNIVERSITY LONDON

Sponsors

2016 Gold Sponsors

2016 Silver Sponsors

References

Klugman, S. A., Panjer, H. H. & Willmot, G. E. (2004), Loss Models: From Data to Decisions, Wiley Series in Probability and Statistics.

Daniel Kahneman. (2011). Thinking, Fast and Slow. New York : Farrar, Straus and Giroux.

Søren Højsgaard (2012). Graphical Independence Networks with the gRain Package for R. Journal of Statistical Software, 46(10), 1-26. URL http://www.jstatsoft.org/v46/i10/

Computational Actuarial Science with R, Edited by Arthur Charpentier, Chapman and Hall/CRC Reference - 656 Pages

Stan Development Team. 2016. RStan: the R interface to Stan, Version 2.9.0. http://mc-stan.org/rstan.html.

The End

Contact:
@MarkusGesmann
markus.gesmann@gmail.com
www.magesblog.com