One parameter models in Stan - part 2

Data Analytics

Jerzy Baranowski

Poisson distribution

What it is?

- discrete probability distribution that expresses the probability of a given **number of events** occurring in a **fixed interval** of time or space if these events occur with a known **constant mean rate** and independently of the time since the last event.
- Probability mass function

$$\pi(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Mean and variance equal to λ

Poisson distribution

Why it is useful?

- Many real issues are behaving similar to Poisson distributed variables
  - Number of accidents
  - Number of customers
  - Number of packets lost during transmission
  - etc.

How likely are you going to die on a plane?

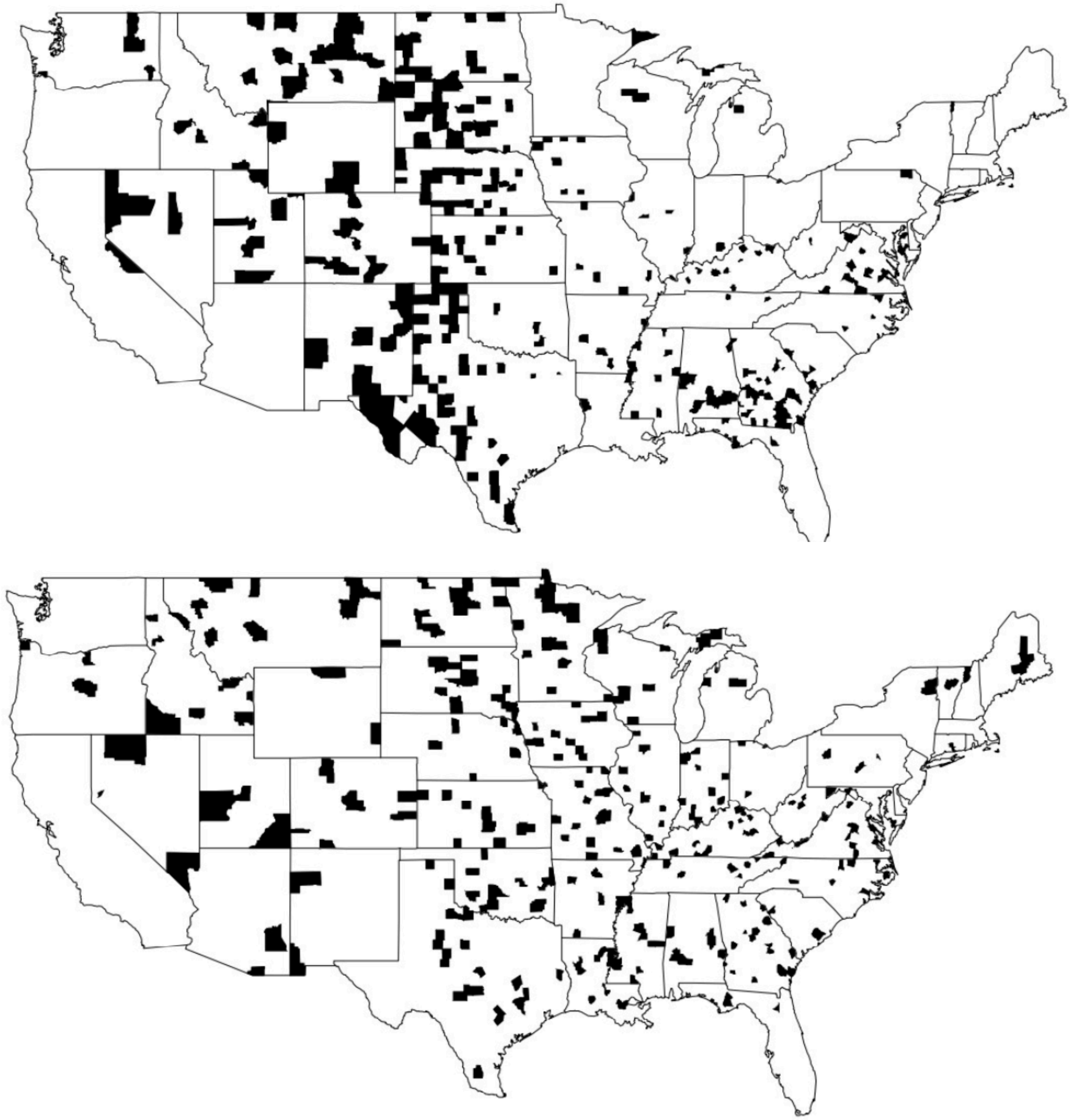
- In classic 80’s comedy anything could happen, but things came to a happy ending (even for autopilot)
- Let us on the other hand consider less happy coincidences look at the data of actual fatal accidents

Wikipedia

Kidney Cancer

Just data is not enough

- We consider a data for kidney/ ureter cancer deaths in individual counties in US in the 1980s
- Problem with population sizes
- Model approach is justified



The counties of the United States with the lowest and highest 10% age-standardized death rates for cancer of kidney/ureter

BDA3 Gelman et al 2014

Cancer modelling

- Poisson distribution for counts of rare occurrences

$$y_i \propto \text{Poisson}(10 n_i \theta_i)$$

- Informative prior distribution

$$\theta_i \propto \text{Gamma}(20{,}430000)$$

What’s else?

There are one-parameter models we have not discussed

- Bernouli distribution - single binary trial
- Exponential distribution - good for modeling waiting times
- More complicated distributions (for example normal) with some of parameters fixed