STAT40810 — Stochastic Models

Brendan Murphy

Week 2

Count Data

Count Data

- Suppose we have data that consist of an independent number of counts in the range $\{0, 1, 2, \ldots\}$.
- The standard model for such a situation is to model the data as Poisson, however alternatives exist:
  - Negative binomial
  - PoissonGamma
  - ConwayMaxwellPoisson (CMP or COM-Poisson) distribution
  - $\cdots$

Count Data:

- Data were collected on the number of houses in 1200 small (equal sized) areas in Japan.
- The number of houses per region were as follows:

Number	Frequency
0	584
1	398
2	168
3	35
4	9
5	4
7	1
9	1

- We want to find a suitable model for the number of houses per region.

Count Data: Poisson Model

- The Poisson model arises in the following situation.
- Suppose that $X_1, X_2, \dots, X_n$ are independent count observations.
- If $X_i \sim \text{Poisson}(\lambda)$ then

$$f(x) = \mathbb{P}\{X_i = x\} = \frac{\lambda^x \exp(-\lambda)}{x!}, \text{ where } \lambda > 0.$$

- Under this model,

$$\mathbb{E}(X) = \lambda \text{ and } \mathbb{V}\text{ar}(X) = \lambda.$$

- So, the model cannot account for situations where

$$\mathbb{E}(X) \neq \mathbb{V}\text{ar}(X).$$

Count Data: Poisson-Gamma Model

- The Poisson-Gamma model arises in the following situation.
- Suppose that $X_1, X_2, \ldots, X_n$ are independent count observations.
- Assume that

  $$X_i | \lambda_i \sim \text{Poisson}(\lambda_i)$$

  and

  $$\lambda_i \sim \text{Gamma}(\alpha, \beta).$$

  That is, each subject has their own rate parameter in the Poisson and these are gamma distributed.
- Then,

  $$X_i \sim \text{Poisson-Gamma}(\alpha, \beta).$$

- And,

  $$f(x) = \frac{\Gamma(x+\alpha)\beta^x}{\Gamma(\alpha)(1+\beta)^{x+\alpha}\Gamma(x+1)}, \text{ where } \alpha, \beta > 0.$$

- Further,

  $$\mathbb{E}(X) = \alpha\beta \text{ and } \mathbb{V}\text{ar}(X) = \alpha\beta + \alpha\beta^2.$$

Poisson-Gamma: Code

- The following code gives the log probability mass function of the Poisson-Gamma model.

```
dPoissonGamma <- function(x,alpha,beta,log=FALSE)
{
if (log)
{
lognumer <- lgamma(x+alpha)+x*log(beta)
logdenom <- lgamma(alpha)+(alpha+x)*log(1+beta)+lgamma(x+1)
res <- lognumer-logdenom
}else
{
numer <- gamma(x+alpha)*beta^x
denom <- gamma(alpha)*(1+beta)^(alpha+x)*gamma(x+1)
res <- numer/denom
}
res
}
```

- Setting `log=TRUE` gives the log probability mass function.

Modeling Question

(a) Find a method of moments estimate of the Poisson-Gamma model parameters.

(b) Fit the Poisson-Gamma model to the data using maximum likelihood.

(c) Fit the Poisson model to the data using maximum likelihood.

(d) Explain whether the Poisson or Poisson-Gamma model provide a better model for the data.

(e) Propose a method for assessing the fit of the models to the data and compare the method using your method.