06 - Conjugate models - 01

Master in Foundations of Data Science
Bayesian Statistics and Probabilistic Programming
Fall 2018-2019

Josep Fortiana

Wednesday, October 24, 2018

Universitat de Barcelona

06 - Conjugate models - 01

The negative binomial distribution

The Poisson-Gamma model

06 - Conjugate models - 01

The negative binomial distribution

The Poisson-Gamma model

First definition

A sequence of independent binary 0/1 experiments, whose indicators are equally distributed as a $\mathrm{Ber}(p)$, $p \in (0,1)$.

The number X of realizations needed to obtain a number $r \in \mathbb{N}$ of successes (1's), is a r.v. following the *negative binomial distribution* with *size* r and probability p.

Alternative: $Y = X - r =$ number of failures (0's) before obtaining a number r of successes.

Probability mass function

$$\mathrm{P}(x) = \binom{x-1}{r-1} \cdot (1-p)^{x-r} \cdot p^r, \qquad x = 1, 2, \ldots,$$

$$\mathrm{P}(y) = \binom{y+r-1}{r-1} \cdot (1-p)^{y} \cdot p^r, \qquad y = 0, 1, 2, \ldots,$$

The second one is more usual (see e.g. the `dnbinom` function in `R`).

Alternative (and the reason for the name)

$$\mathrm{P}(y) = \binom{-r}{y} \cdot p^r \cdot (-q)^y,$$

where $q = 1 - p$. Indeed:

$$\binom{-r}{y} = \frac{(-r) \cdot (-r-1) \cdot \cdots \cdot (-r-y+1)}{y!}$$

$$= (-1)^y \cdot \binom{y+r-1}{r-1}.$$

General definition

For an integer r,

$$\binom{y+r-1}{r-1} = \frac{(y+r-1)!}{(r-1)!\cdot y!} = \frac{\Gamma(y+r)}{\Gamma(r)\cdot y!}.$$

The right hand is valid for real $r>0$. Thus the pmf:

$$\mathrm{P}(y) = \frac{\Gamma(y+r)}{\Gamma(r)\cdot y!}\cdot(1-p)^y\cdot p^r, \qquad y=0,1,2,\dots,$$

defines the $\mathrm{NegBin}(r,p)$, for $r>0$ and probability p.

Relation to the geometric distribution

For $r = 1$, the $\mathsf{NegBin}(1, p)$ is the $\mathsf{Geom}(p)$.

For integer r, a r.v. distributed as $\mathsf{NegBin}(r, p)$ can be considered as the sum of r i.i.d. copies of a $\mathsf{Geom}(p)$.

Expectation, variance of a negative binomial

For $Y \sim \mathrm{NegBin}(r, p)$,

$$\mathrm{E}(Y) = \mu \equiv r \cdot \frac{1-p}{p}$$

$$\mathrm{var}(Y) = \sigma^2 \equiv r \cdot \frac{1-p}{p^2} = \frac{\mu}{p} = \mu + \frac{\mu^2}{r}.$$

06 - Conjugate models - 01

The negative binomial distribution

The Poisson-Gamma model

Likelihood

The likelihood is:

$$(y|\lambda) \sim \text{Poisson}(\lambda),$$

with pmf:

$$f(y|\lambda) = e^{-\lambda}\,\frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots, \quad \lambda > 0,$$

Conjugate prior pdf

λ's prior is:

$$\mathrm{Gamma}(\alpha,\beta),\quad \alpha,\beta>0,$$

with pdf:

$$h(\lambda)=\frac{\beta^\alpha}{\Gamma(\alpha)}\,\lambda^{\alpha-1}\,e^{-\beta\lambda},\quad \alpha,\beta,\lambda>0.$$

Joint pdf

The joint "density" of (y, λ) is:

$$f(y, \lambda) = f(y|\lambda) \cdot h(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{\lambda^{\alpha+y-1}}{y!}\, e^{-(\beta+1)\lambda},$$

for $\alpha, \beta, \lambda > 0$.

Marginal pmf of x (Prior predictive pmf)

To integrate with respect to λ we split $f(y,\lambda)$:

$$\frac{\beta^{\alpha}}{\Gamma(\alpha)}\cdot\frac{1}{y!}\cdot\frac{\Gamma(\alpha+y)}{(\beta+1)^{(\alpha+y)}}\times\frac{(\beta+1)^{(\alpha+y)}}{\Gamma(\alpha+y)}\cdot\lambda^{\alpha+y-1}\,e^{-(\beta+1)\lambda},$$

The second factor is a $\mathsf{Gamma}(\alpha+y,\beta+1)$ pdf, which integrates to 1.

Marginal pmf of x

The result of integrating out λ is the marginal pmf of y:

$$f(y) = \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\,y!}\frac{\beta^\alpha}{(\beta+1)^{\alpha+y}}, \quad y = 0, 1, \ldots, \quad \alpha, \beta > 0$$

Noting that:

$$\frac{\beta^\alpha}{(\beta+1)^{\alpha+y}} = \left(\frac{1}{\beta+1}\right)^y \cdot \left(\frac{\beta}{\beta+1}\right)^\alpha,$$

Marginal pmf of y

We identify $f(y)$ as the pdf of a $\mathsf{NegBin}(r, p)$, a negative binomial r.v. , with parameters

$$r = \alpha, \quad \text{and} \quad p = \frac{\beta}{\beta + 1}.$$

As always, this is the *prior predictive* distribution for the observed y.

Posterior pdf

We apply Bayes' formula, dividing $f(y, \lambda)$ by the marginal pmf, and we obtain the posterior pdf of λ, given y, the first factor above:

$$h(\lambda|y) = \frac{(\beta+1)^{\alpha+y}}{\Gamma(\alpha+y)} \lambda^{\alpha+y-1} e^{-(\beta+1)\lambda}, \quad \lambda > 0,$$

which is a $\mathsf{Gamma}(\alpha+y, \beta+1)$.

Case of an n-sample: posterior pdf

For $y_1, \ldots, y_n$ i.i.d.$\sim$ Poisson(λ), the sum:

$$y = \sum_{i=1}^{n} y_i \sim \mathrm{Poisson}(n\,\lambda)$$

Thus, for a prior $\lambda \sim \mathrm{Gamma}(\alpha, \beta)$ and n observed Poisson(λ) data, the posterior pdf is $\mathrm{Gamma}(\alpha + y, \beta + n)$.

Case of an n-sample: prior predictive pmf

Similarly, for n observed $\mathrm{Poisson}(\lambda)$ data, the prior predictive (marginal) distribution of the total count number $y = \sum_{i=1}^{n} y_i$ is a $\mathrm{NegBin}(r, p)$, a negative binomial r.v. , with parameters:

$$r = \alpha, \quad \text{and} \quad p = \frac{\beta}{\beta + n}.$$

An n-sample with different exposures

In many applications we find data of the form:

$$y_i \sim \mathrm{Poisson}(\lambda_i), \quad \text{where } \lambda_i = x_i \cdot \theta, \quad 1 \leq i \leq n.$$

The values x_i are known positive values of an explanatory variable, usually called *exposure*, and θ is the common *rate* parameter.

Posterior pdf for an n-sample with different exposures

For a prior $\theta \sim \mathrm{Gamma}(\alpha, \beta)$,

observations $\boldsymbol{y} = (y_1, \ldots, y_n)$, corresponding to

known exposures $\boldsymbol{x} = (x_1, \ldots, x_n)$,

the posterior pdf is:

$$\theta | \boldsymbol{y} \sim \mathrm{Gamma}(\alpha + y, \beta + x),$$

where $y = \sum_{i=1}^{n} y_i$ and $x = \sum_{i=1}^{n} x_i$.