A Poisson–Gamma Mixture Is Negative-Binomially Distributed

We can view the negative binomial distribution as a Poisson distribution with a gamma prior on the rate parameter. I work through this derivation in detail.

PUBLISHED
16 September 2019

Consider a Poisson model for count data,

$$y \sim \text{Poisson}(\theta), \qquad \theta \geq 0.$$

The parameter θ can be interpreted as the *rate of arrivals*, and importantly, $\mathbb{E}[y] = \text{Var}(y) = \theta$. An unfortunate property of this Poisson model is that it cannot model *overdispersed* data or data in which the variance is greater than the mean. This is because Poisson regression has one free parameter. However, if we place a gamma prior on θ,

$$\begin{aligned} y &\sim \text{Poisson}(\theta) \\ \theta &\sim \text{gamma}\Big(r, \frac{p}{1-p}\Big), \end{aligned}$$

and then marginalize out θ, we get a negative binomial (NB) distribution, which has the useful property that its variance can be greater than its mean. The derivation is

$$\begin{aligned} p(y) &= \int_0^\infty p(y \mid \theta) p(\theta) \mathrm{d}\theta \\ &= \int_0^\infty \Big(\frac{\theta^y e^{-\theta}}{y!}\Big)\Big(\frac{1}{\Gamma(r)\big(\frac{p}{1-p}\big)^r} \theta^{r-1} e^{-\theta(1-p)/p}\Big) \mathrm{d}\theta \\ &= \frac{(1-p)^r p^{-r}}{y!\Gamma(r)} \int_0^\infty \theta^{r+y-1} e^{-\theta/p} \mathrm{d}\theta \\ &\stackrel{\star}{=} \frac{(1-p)^r p^{-r}}{y!\Gamma(r)} p^{r+y} \Gamma(r+y) \\ &= \frac{\Gamma(r+y)}{\Gamma(r) y!} p^y (1-p)^r \\ &\stackrel{\dagger}{=} \frac{(r+y-1)!}{(r-1)! y!} p^y (1-p)^r \\ &= \binom{y+r-1}{y} p^y (1-p)^r \\ &= \text{NB}(r, p). \end{aligned}$$

Step $\star$ holds because of the following equality,

$$\int_0^\infty x^b e^{-ax} \mathrm{d}x = \frac{\Gamma(b+1)}{a^{b+1}}.$$

Wikipedia claims that this is part of the usefulness of the gamma function: integrals of expressions of the form $f(x) e^{-g(x)}$, which model exponential decay, can be sometimes solved in closed form using the above equation.

Step $\dagger$ uses the fact that $\Gamma(x) = (x-1)!$.