Count Data Models

Derya Uysal

LMU, Munich, Germany,
derya.uysal@econ.lmu.de,
www.deryauysal.com

1 Basics

Poisson Distribution:

$$
\begin{aligned}
f(x;\lambda) = P(X = x) &= \frac{\lambda^x \exp(-\lambda)}{x!} \qquad \text{for} \\
E[X] = V[X] &= \lambda
\end{aligned}
$$

Gamma Distribution:

$$
\begin{aligned}
f(x;\alpha,\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x\beta), \quad x \in (0,\infty), \quad \alpha,\beta > 0 \\
E[X] &= \alpha/\beta \\
V[X] &= \alpha/\beta^2
\end{aligned}
$$

Negative Binomial Distribution

X denotes the trial at which the r^{th} success occurs:

$$
\begin{aligned}
f(x;\lambda) = P(X = x) &= \binom{r+x-1}{x} p^r (1-p)^x \qquad \text{for} \\
E[X] &= r\frac{1-p}{p} \\
V[X] &= r\frac{1-p}{p^2}
\end{aligned}
$$

2 Hierachical Model (Mixture Distribution)

$$
\begin{aligned}
Y|\nu &\sim Po(\lambda\nu) \\
\nu &\sim gamma(\alpha,\beta)
\end{aligned}
$$

Marginal pmf of Y

$$
\begin{aligned}
f_Y(y) &= \int_0^\infty f_{Y,\nu}(y,\nu)d\nu \\
&= \int_0^\infty f_Y(y|\nu)f_\nu(\nu)d\nu \\
&= \int_0^\infty \frac{(\lambda\nu)^y \exp(-\lambda\nu)}{y!}\frac{\beta^\alpha}{\Gamma(\alpha)}\nu^{\alpha-1}\exp(-\nu\beta)d\nu \\
&= \int_0^\infty \lambda^y \nu^{y+\alpha-1}\frac{\beta^\alpha}{y!\Gamma(\alpha)}\exp(-\nu(\beta+\lambda))d\nu \\
&= \frac{\beta^\alpha\lambda^y}{y!\Gamma(\alpha)}\int_0^\infty \underbrace{\nu^{y+\alpha-1}\exp(-\nu(\beta+\lambda))}_{\text{Kernel of a gamma pdf with }(y+\alpha,\beta+\lambda)} d\nu \\
&= \frac{\beta^\alpha\lambda^y}{y!\Gamma(\alpha)}\frac{\Gamma(y+\alpha)}{(\beta+\lambda)^{y+\alpha}}\underbrace{\int_0^\infty \frac{(\beta+\lambda)^{y+\alpha}}{\Gamma(y+\alpha)}\nu^{y+\alpha-1}\exp(-\nu(\beta+\lambda))d\nu}_{=1(gammapdf)} \\
&= \frac{\beta^\alpha\lambda^y}{y!\Gamma(\alpha)}\frac{\Gamma(y+\alpha)}{(\beta+\lambda)^{y+\alpha}} \\
&= \frac{(y+\alpha-1)!}{y!(\alpha-1)!}\left(\frac{\beta}{\beta+\lambda}\right)^\alpha\left(\frac{\lambda}{\beta+\lambda}\right)^y \\
&= \underbrace{\binom{y+\alpha-1}{y}\left(\frac{\beta}{\beta+\lambda}\right)^\alpha\left(1-\frac{\beta}{\beta+\lambda}\right)^y}_{\text{Negative bionomial with }r=\alpha,p=\beta/(\beta+\lambda)}
\end{aligned}
$$

with mean $E[Y] = r\frac{1-p}{p} = \alpha\frac{\lambda/(\beta+\lambda)}{\beta/\beta+\lambda} = \frac{\alpha}{\beta}\lambda$ and variance $V[Y] = \frac{\alpha\lambda(\beta+\lambda)}{\beta^2}$.

Unconditional Mean and Variance of Y (without deriving the marginal pmf of Y)

The unconditional mean and variance of Y can be derived using law of iterated expectations.

1. Unconditional Mean:

$$
\begin{aligned}
E[Y] &= E[E[Y|\nu]] = E[\lambda\nu] \\
&= \lambda E[\nu] = \lambda\frac{\alpha}{\beta}
\end{aligned} \tag{1}
$$

2. Unconditional Variance:

$$V[Y] = V[E[Y|\nu]] + E[V[Y|\nu]]$$

$$
\begin{aligned}
&= V[\lambda\nu] + E[\lambda\nu] \\
&= \lambda^2 V[\nu] + \lambda E[\nu] \\
&= \lambda^2 \frac{\alpha}{\beta^2} + \lambda \frac{\alpha}{\beta} \\
&= \frac{\alpha\lambda(\beta + \lambda)}{\beta^2} \qquad (2)
\end{aligned}
$$

3 NEGBIN I

(constant dispersion model)
Assume $\alpha = \beta \equiv \frac{\lambda}{\sigma^2}$, then $E[\nu] = 1$ and $V[\nu] = \frac{\sigma^2}{\lambda}$. Given equations 1 and 2, unconditional mean and variance of the count data Y:

$$
\begin{aligned}
E[Y] &= \lambda \\
V[Y] &= \lambda(1 + \sigma^2) \\
\frac{V[Y]}{E[Y]} &= (1 + \sigma^2) \text{ (constant dispersion)}
\end{aligned}
$$

4 NEGBIN II

(mean dispersion model)
Assume $\alpha = \beta \equiv \frac{1}{\sigma^2}$, then $E[\nu] = 1$ and $V[\nu] = \sigma^2$. Given equations 1 and 2, unconditional mean and variance of the count data Y:

$$
\begin{aligned}
E[Y] &= \lambda \\
V[Y] &= \lambda(1 + \lambda\sigma^2) \\
\frac{V[Y]}{E[Y]} &= (1 + \lambda\sigma^2) \text{ (mean dispersion)}
\end{aligned}
$$