ST 518: Data Analytics II

Poisson Counts

Poisson Counts

Relationship to Multinomial

Poisson Counts

Another probability model that is used for count data is the Poisson distribution—this is used for counts obtained in some unit of time or space. Some examples:

- The number of hits on an internet site in one hour
- The number of fish caught in a 1km trawl in the ocean
- The number of patients at an emergency clinic in one 24-hour period

The Poisson Distribution

For $Y \sim \text{Poisson}(\lambda)$, the probability distribution, or probability mass function of Y is:

$$Pr(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

where $y \in \{0, 1, 2, \ldots\}$ and $\lambda > 0$.

Recall that $y! = y \times (y-1) \times (y-2) \times \cdots \times 1$ (with $0! = 1$).

The parameter, λ, is called the rate parameter because it indicates the average number of our count outcome in some unit of time or space.

Relationship to Multinomial

It turns out that if $Y_1, Y_2, \dots, Y_k$ are all independent Poisson counts, each possibly with it's own rate parameter, $\lambda_1, \lambda_2, \dots, \lambda_k$, then if we condition on the total counts (i.e., the sum of the Y_j's), we can think of the conditional counts,

$$Y_1, \dots, Y_k | M \sim \text{multinomial}(p_1, \dots, p_k),$$

where

$$M = Y_1 + Y_2 + \cdots + Y_k$$

and

$$p_j = \frac{\lambda_j}{\Lambda},$$

where Λ is the sum of the λ_j's.

It's not important for you to understand the details of this conditioning argument, only that by using it we have more flexibility in how we model categorical data in contingency tables.

An Example

Suppose that we have obtained the following data from a survey of likely voters:

Democrat	Independent	Republican
330	425	261

Using the multinomial distribution to model these counts, we obtain estimates of the category probabilities:

$$
\begin{aligned}
\hat{p}_1 &= 330/1016 = 0.32 \\
\hat{p}_2 &= 425/1016 = 0.42 \\
\hat{p}_3 &= 261/1016 = 0.26
\end{aligned}
$$

Example, Continued

Now, what if we consider the three counts on the previous slide to be independent Poisson variables:

$$\begin{aligned} Y_1 &\sim \text{Poisson}(\lambda_1) \\ Y_2 &\sim \text{Poisson}(\lambda_2) \\ Y_3 &\sim \text{Poisson}(\lambda_3) \end{aligned}$$

then the connection to the multinomial gives:

$$(Y_1, Y_2, Y_3)|M \sim \text{multinomial}(p_1, p_2, p_3)$$

where $M = Y_1 + Y_2 + Y_3$.

And, we just estimated these p_j's on the previous slide—so our inferences will be the same whether we consider the counts to be Poisson distributed and then condition on the total, or multinomial distributed.