Poisson and Negative Binomial Distribution

Chao Cheng

February 15, 2024

Contents

1 Introduction

This notes is about poisson and negative binomial distribution, and their application in modeling count data.

2 Poisson Distribution

2.1 Basic Information

The p.m.f of a random variable X that follows Poisson distribution is

$$P\left(X = x\right) = \frac{\mathrm{e}^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, \cdots.$$

The expectation and variance of X is

$$\mathrm{E}X = \mathrm{Var}X = \lambda.$$

2.2 Some Additoinal Properties

1. Summation of independent poisson distributed variables is still poisson distributed. That is to say,

   if

   $$X_1 \sim Poisson(\lambda_1), \quad X_2 \sim Poisson(\lambda_2), \quad X_1 \perp X_2,$$

   then

   $$X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2).$$

2. Poisson distribution can be related to Gamma distribution, which is also mentioned in "gamma_and_beta.pdf" notes.

3. A poisson distributed variable can be seen as counts of customer to a store or occurence of some events. And the **time interval between each count** follows exponential distribution.

3 Negative Binomial Distribution

3.1 Basic Information

Denote X the number of failing attempts just before reaching a total of r times of success and p the probability of success. If each attempt is independent of each other, then the distribution of X is called a negative binomial distribution. And the p.m.f of X is

$$P\left(X = x\right) = C_{x+r-1}^{x} p^r \left(1 - p\right)^x = \frac{(x + r - 1)!}{(r-1)!x!} p^r \left(1 - p\right)^x, \quad x = 0, 1, \cdots.$$

In this notation, X is the number of failing attempts. In some cases, one will model the distribution of number of total attempts X', which is just $X' = X + r$.

Previously, r is the number of required success, which is naturally an positive integer. But one can actually generalize r to positive real number using Gamma function. Then the p.m.f of X is written as

$$P\left(X = x\right) = \frac{\Gamma(x + r)}{\Gamma(r)\Gamma(x + 1)} p^r \left(1 - p\right)^x, \quad x = 0, 1, \cdots,$$

which is also called **Polya** distribution.

The expectation of a negative binomial/Polya distribution is

$$\mathrm{E}X = \frac{r\left(1 - p\right)}{p}.$$

The variance of a negative binomial/Polya distribution is

$$\mathrm{Var}X = \frac{r\left(1 - p\right)}{p^2}.$$

3.2 Some Additional Properties

1. Summation of independent negative binomial distributed variables. That is to say, if $X_1 \sim NB(r_1, p)$, $X_2 \sim NB(r_2, p)$ and $X_1 \perp X_2$, then

$$X_1 + X_2 \sim NB(r_1 + r_2, p).$$

2. Relationship to Poisson and Gamma distribution.

   - Related to poisson when $r \to \infty$.
   - Related to poisson and gamma distribution.

3. Dispersion parameter and other parametrization of negative binomial distribution.

   - In some literature, expectation and variance are used to describe negative binomial distribution:
   $$\mathrm{E}X = \mu$$
   $$\mathrm{Var}X = \sigma^2$$
   $$r = \frac{\mu^2}{\sigma^2 - \mu}$$
   $$p = \frac{\mu}{\sigma^2}$$

   - In some literature, $\alpha = \frac{1}{r}$ is called the dispersion parameter while others call r the dispersion parameter. It also has other names such as "shape parameter", "clustering coefficient", "heterogeneity" and "aggregation parameter". Note that the expectation and the variance of a negative binomial distributed variable satisfies
   $$\sigma^2 = \mu + \alpha\mu^2 = \mu + \frac{\mu^2}{r}.$$

   - Denote β the **failure odds**, then
   $$\beta = \frac{1 - p}{p}$$
   $$P(X = x) = \frac{\Gamma(x + r)}{\Gamma(r)\Gamma(x + 1)} \left(\frac{\beta}{1 + \beta}\right)^r \left(\frac{1}{1 + \beta}\right)^x, \quad x = 0, 1, \cdots,$$
   $$\mathrm{E}X = r\beta$$
   $$\mathrm{Var}X = r\beta(1 + \beta).$$

4. Geometric distribution is a special case of negative binomial distribution.

4 Modeling Counts Data

4.1 Basic Modeling

4.2 Zero Inflation

References