Model for Counts

STAT-UB.0001 Statistics for Business Control

Ningshan Zhang

IOMS Department
nzhang@stern.nyu.edu

Jul 17, 2018

* Office Hours: Wed & Fri 10:00 - 11:30 AM, KMC 8-174

Review

Random Variable

- Probability distribution function (PDF)
- Expected value of a random variable
- Variance and standard deviation of a random variable
- Properties of expected value
  - Affine transformation
  - Sum

Example

A biased coin has a 25% of landing heads, and a 75% chance of landing tails. If you flip the coin 4 times, what is the chance of getting exactly 2 heads?

Random variable X = number of heads out of 4 flips. How to find $p(2)$?

(a) What are the sample points that lead to the observation $X = 2$? How many are there?

(b) What is the probability for each of the sample points in (a)?

(c) What is $p(2)$?

(d) What is $p(2)$ if you flip the coin 10 times?

Binomial Random Variable and Binomial Distribution

Binomial experiment:

- It consists of a fixed number n of statistically independent trials;
- each trial has the same probability of success p;
- we want to count the number of successes.

Let $X =$ the number of successes. Then X is a *binomial random variable* that has *binomial distribution*, written as $X \sim B(n, p)$.
The PDF:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

The Binomial Distribution

If $X \sim B(n, p)$, then it has the following properties:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \tag{PDF}$$

$$\mathbb{E}(X) = np, \tag{mean}$$

$$\mathrm{var}(X) = np(1-p). \tag{variance}$$

The Binomial Distribution

Characteristics of a Binomial Experiment:

- The experiment consists of n identical and independent trials.
- There are only two possible outcomes (success and failure) on each trial.
- The probability of success is the same on each trial.
- The random variable of interest is the number of successes out of n trials.

Example

The manager at NYU bookstore wants to figure out how many customers will visit the store between 4 - 5 pm. By examining past records, he finds that on average 20.3 customers will visit the store within that hour. Let X = number of customers visit the store between 4 - 5 pm. What is the PDF of X?

- Approximate X with the outcome of 60 independent trials: whether a customer visits the store in a given minute.
- Approximate X with the outcome of 3600 independent trials: whether $\cdots$ in a given second.
- Getting more and more granular, leads to the *Possion distribution*.

The Poisson Distribution

Let X = the number of events that occur in a fixed interval of time, space, etc. Assume that

- Events occur with a known constant rate.
- The events occur independently of the time since the last event.

Then X follows a *Poisson distribution*.

The Poisson Distribution

The PDF of Poisson distribution

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

- $k = 0, 1, 2, \ldots$. There is no upper limit on k (unlike binomial).
- λ is the (only) parameter of the distribution.
- $e \approx 2.72$ is a constant, the base of the natural logarithm.

Mean and variance of X:

$$\mathbb{E}(X) = \mathrm{var}(X) = \lambda.$$

The Poisson Distribution: Example

The manager at NYU bookstore wants to figure out how many customers will visit the store between 4 - 5 pm. By examining past records, he finds that on average 20.3 customers will visit the store within that hour. Let X = number of customers visit the store between 4 - 5 pm. What is the PDF of X?

Solution: X follows a Poisson distribution with $\lambda = 20.3$. Then,

- What is the probability of $X = 10$?

$$\mathbb{P}(X = 10) = \frac{\lambda^{10} e^{-\lambda}}{10!}$$

- What is the probability of $X = 20$?

$$\mathbb{P}(X = 20) = \frac{\lambda^{20} e^{-\lambda}}{20!}$$

Connection between Binomial and Poisson Distribution

In binomial distribution, let $n \to \infty$, $p \to 0$, but keep the mean as a constant: $np = \lambda$. Then,

$$B(n, p) \to \text{Poisson}(\lambda).$$

Proof: for any fixed number k and parameter λ,

$$\begin{aligned}
&\binom{n}{k} p^k (1-p)^{n-k} && \text{(PDF of } X \sim B(n, p)\text{)} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \underbrace{\left[\frac{n(n-1)\cdots(n-k+1)}{n^k}\right]}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&\to \frac{\lambda^k e^{-\lambda}}{k!} && \text{(PDF of } X \sim \text{Poisson}(\lambda)\text{)}
\end{aligned}$$

Empirical Rule

Empirical rule works for binomial distribution and Poisson distribution.

- $X \sim B(n, p)$, then $\mu = np$, $\sigma = \sqrt{np(1-p)}$.
- $X \sim \text{Poisson}(\lambda)$, then $\mu = \lambda$, $\sigma = \sqrt{\lambda}$.

Summary

- Binomial random variable and binomial distribution.
- Poisson random variable and Poisson distribution.
- Empirical rule applies to both.