L5. Commonly used Discrete Distribution

- Review r.v., p.m.f., p.d.f., c.d.f. and expected value

- Population variance and standard deviation

- Commonly use discret distributions
  - Bernoulli
  - Binomial
  - Poisson

1. Review:

Random variable X : a function that numerically records the outcomes of a random experiment

p.m.f. (Discrete) and p.d.f. (continous) : functions that define the values and probability distribution of the values of X.

- p.m.f. : $f(x) = P(X = x), \quad x \in S$

- p.d.f. : $f(x) = F'(x), \quad x \in S$

c.d.f.: another function that obtains the information of the distribution of X

$$F(x) = \mathbb{P}(X \leq x)$$

Expected value: the weighted average of X that shows the "center" of the population

μ

Discrete: $E(X) = \sum_{x \in S} x \cdot f(x)$

Continuous: $E(X) = \int_{x \in S} x \cdot f(x)\, dx$

In real life, working on estimating the distribution of the whole population is rather difficult, so most statistical methods focus on some key characteristics of the distribution: $\boxed{\text{parameters}}$

Two most commonly used parameters:

- mean (expected value) (center)
- variance (spread-out)

For ex: a data, by nature, is close to a Normal distribution behavior, then knowing the mean and variance can fully define this Normal distribution.

2. Population variance σ^2 or $Var(X)$

What is variance? : it measures the variability in the outcomes. The larger it is, the more spread-out the outcome values are.

How is it calculated? : stand deviation measures the "average" distance of the values from the mean; variance is the squared version of that.

$\boxed{\text{Def.}}$ The variance of a r.v. X (or its distribution) is given by

$$\sigma_X^2 = Var(X) = \underset{\text{"average"}}{E}\Big[\underbrace{(X-\mu_X)}_{\text{distance from the mean}}{}^2\Big]$$

$\boxed{\text{Thm}}$ $\sigma_X^2 = E[(X-\mu_X)^2] = E(X^2) - \mu_X^2$

pf. $E[(X-\mu_X)^2]$

$= E[X^2 - 2\mu_X \cdot X + \mu_X^2]$

$= E(X^2) - 2\mu_X \cdot \underbrace{E(X)}_{=\mu_X} + \mu_X^2$

remark: for any r.v. X,

μ_X is a fixed constant

(no matter unknown or known)

$= E(X^2) - 2\mu_X^2 + \mu_X^2$

$= E(X^2) - \mu_X^2$

$\boxed{\text{Ex 1}}$ Bernoulli (p)

$Var(X) = E[(X-\mu_X)^2]$

$= E[(X-p)^2]$

$= \Sigma (x-p)^2 \cdot f(x)$

$$= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p$$

$$= p^2(1-p) + (1-2p+p^2) \cdot p$$

$$= p^2 - p^3 + p - 2p^2 + p^3$$

$$= p - p^2$$

$$= p(1-p)$$

$$= E(X^2) - \mu_X^2$$

$$= \Sigma x^2 \cdot f(x) - p^2$$

$$= 0^2 \cdot (1-p) + 1^2 \cdot p - p^2$$

$$= p - p^2$$

$$= p(1-p)$$

$X \sim \text{Bernoulli}(p)$

p.m.f. : $f(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}$

$\mu = p, \quad \sigma^2 = p(1-p)$

Take away: when we have a binary data, we can assume it's from a Bernoulli distribution. if we can estimate p, we get all the facts about the distribution.

$\boxed{\text{Ex 2}}$ Uniform distribution on $[a, b]$

$$\sigma^2 = E[(X-\mu_x)^2]$$

$$= \int_a^b \left(x - \frac{a+b}{2}\right)^2 \cdot \frac{1}{b-a}\, dx$$

$$= \left.\frac{\left(x-\frac{a+b}{2}\right)^3}{3(b-a)}\right|_a^b$$

$$= \frac{\left(b-\frac{a+b}{2}\right)^3}{3(b-a)} - \frac{\left(a-\frac{a+b}{2}\right)^3}{3(b-a)}$$

$$= \frac{\left(\frac{b-a}{2}\right)^3}{3(b-a)} - \frac{\left(\frac{a-b}{2}\right)^3}{3(b-a)}$$

$$= \frac{(b-a)^2}{24} + \frac{(b-a)^2}{24} = \frac{(b-a)^2}{12}$$

Try the other way?

$X \sim \text{Uniform}[a.b]$

$f(x) = \frac{1}{b-a}, \quad x \in [a.b]$

$\mu = \frac{a+b}{2}, \quad \sigma^2 = \frac{(b-a)^2}{12}$

Properties of Variance

(1) $\text{Var}(h(X)) = E[(h(X))^2 - (E[h(X)])^2]$

$= E[(h(X))^2] - [E(h(X))]^2$

(2) $\text{Var}(c) = 0$ for a constant c

proof: $\text{Var}(c) = E[c^2] - [E(c)]^2$

$= c^2 - c^2 = 0$

(3) $\text{Var}(aX+b) = a^2 \text{Var}(X)$

proof?

Do we have $\text{Var}(h_1(x) + h_2(x)) = \text{Var}(h_1(x)) + \text{Var}(h_2(x))$?

[Optimal] The Moment-Generating Function (m.g.f.)

- A lot of times it's hard to calculate the mean and variance of a distribution, or the higher moments: $E(X^3)$, $E(X^4)$, etc.

- m.g.f. helps with the calculation

$\boxed{\text{Def}}$: the moment-generating function of X is defined as

$$M(t) = E(e^{tX})$$

$\boxed{\text{Thm}}$: $M^{(r)}(0) = E(X^r)$

particularly, $\mu = M'(0)$

$$\sigma^2 = M''(0) - [M'(0)]^2$$

$\boxed{\text{Ex}}$ $X \sim \text{Bernoulli}(p)$

$$M(t) = E(e^{tX})$$

$$= \sum_x e^{tx} \cdot f(x)$$

$$= e^{t \cdot 0} \cdot (1-p) + e^{t \cdot 1}(p)$$

$$= (1-p) + p \cdot e^{t}$$

$$M'(t) = p \cdot e^{t}$$

$$M''(t) = p \cdot e^{t}$$

$$M'(0) = p \cdot e^{0} = p = \mu$$

$$M''(0) = p \cdot e^{0} = p = E(X^2)$$

$$M''(0) - (M'(0))^2 = p - p^2 = p(1-p) = \sigma^2$$

3. Commonly used discrete distribution

When given a data, most likely we don't know the "exact" distribution. However, based on the behavior and nature of the data, we can make reasonable assumptions of the distribution the data is drawn from. Based on the assumption, we can then choose appropriate analysis methods.

a. Bernoulli (p) $\longleftrightarrow$ Binary outcomes

$\underline{\underline{\quad}}$ parameter p

A random experiment has two mutually exclusive outcomes.

Let r.v. X denote the outcome of the experiment, then $X \sim$ Bernoulli (p).

$$f(x) = \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \end{cases}$$

$\mu = p, \ \sigma^2 = p(1-p)$

Application: this is the mostly used distribution in all classification models.

b. Binomial (n, p) or $b(n, p)$ parameters (n, p)

- Perform the Bernoulli(p) experiment n times.
- The trials are independent
- Let r.v. X equals the number of successes (1s) in the n trials

$X \sim b(n, p)$ s.t.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n$$

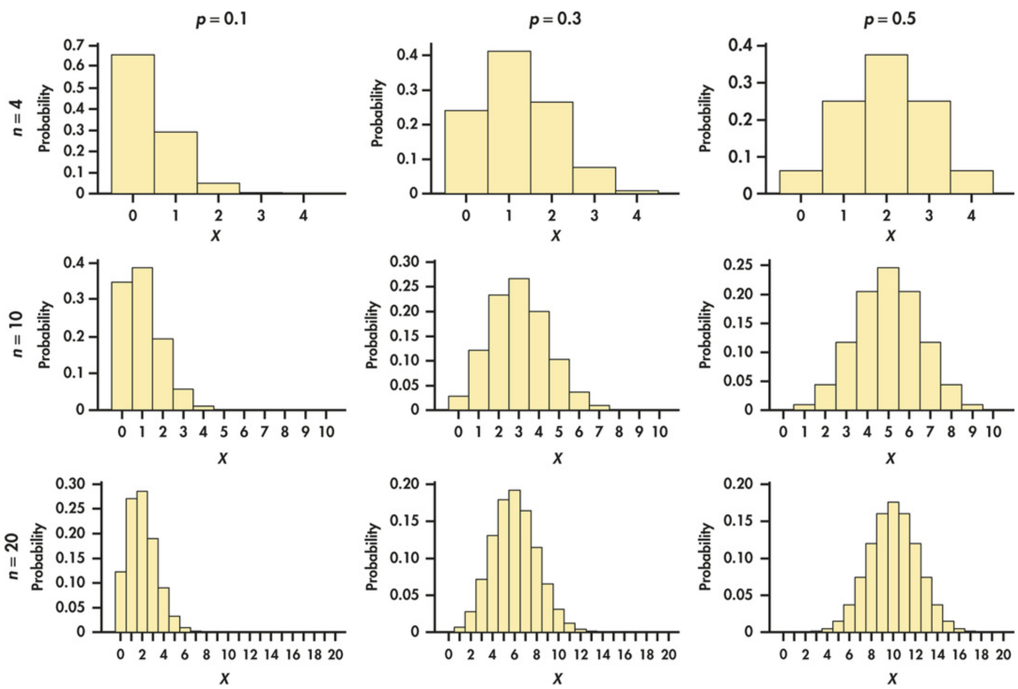
$$\mu = np$$

$$\sigma^2 = np(1-p)$$

How does binomial distribution look like?

when n is large, $\overset{\text{Appr}}{\sim}$ Normal $(np, np(1-p))$

(used in z test for p later)



$\boxed{\text{Ex 3}}$ Lottery! Match 3 Numbers (1–10)

I have 10 people bought tickets (one for each) this time, what's chance that nobody won?

For a single ticket:

$$P_{win} = \frac{1}{10^3} = 0.001$$

Let X be the number of winning tickets out of 10.

$$X \sim b(10, 0.001)$$

$$P(X=0) = \binom{10}{0} 0.001^{0} (1-0.001)^{10-0}$$

$$= 0.999^{10}$$

$$\simeq 0.99$$

what is $\mu = np = 10 \times 0.001 = 0.01$?

the expected number of people will win!

C. Poisson ($\underline{\underline{\lambda}}$) parameter λ

Let X be the number of occurance of an event in a given continuous interval, knowing that the average number of occurance during the same interval is λ. Then

$$X \sim \text{Poisson}(\lambda)$$

$\boxed{\text{Ex 4}}$ Earthquake!

It's known that the average number of earthquakes in the area is 5/year.

Let X be the number of earthquakes next year.

units match

$$X = 0, 1, 2, \cdots, \infty$$

[optional] How to get p.m.f. of this process?

- split the interval into many small intervals, so small that at most one occurance could happen

- count how many intervals that event happened out of n.

- $X \overset{\text{Appr}}{\sim} b(n, p)$

- $p \approx \lambda \cdot \frac{1}{n}$ (this is the trick)

$$\mathbb{P}(X = x) = \lim_{n \to \infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x e^{-\lambda}}{\lambda !}$$

$\boxed{\text{Def}}$ a r.v. $X \sim \text{Poisson}(\lambda)$ when

$$f(x) = \frac{\lambda^x e^{-\lambda}}{\lambda !}, \quad x = 0, 1, 2, \ldots$$

correction: $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$

$$\left.\begin{aligned} E(X) &= \lambda \\ Var(X) &= \lambda \end{aligned}\right\} \text{m.g.f.}$$

$\boxed{\text{Revisit Ex 4}}$

$X \sim \text{Poisson}(5)$

$$P(X=0) = \frac{5^0 e^{-5}}{0!} \approx 0.007$$

$\boxed{\text{Applications}}$

- Poisson Regression : predict the number of events occuring within an given interval of time.

- Bridge between discrete and continuous :

  Let T be the time between two occurances of the events,

  $T \sim \text{Exponential}(\lambda)$

key distribution used in survival analysis.