3. Probability distribution:

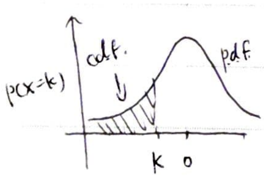
A probability distribution is a mathematical function that describes the possible outcomes of a random variable along with their corresponding probability values.

Terminology and Notion:

- Expected value $E(x)$ – mean of the variable X
- Parameters: Characteristics that are used to define the probability distribution.
- r.v. $X \sim$ Uniform (a,b) means that X is a random variable, distributed as Uniform with parameters a and b.
- $P(X=k) = 0.9$ means the probability of random variable X equals to k is 0.9
- Probability mass function (p.m.f) : the equation that discribes $P(X=k)$ for discrete distribution.
- Probability density function (p.d.f) : the equivalent to p.m.f for continuous distributions
- Cumulative distibution function (c.d.f) the equation that describes $P(X<=k)$

Continuous or discrete?

- Discrete distribution: the random variable can assume one of a countable number of values
- Continuous distribution: the random variable can assume one of an infinite number of values



Continuous Distribution:

- The c.d.f is the area under the curve of the p.d.f from 0 to k

Discrete Uniform Distribution

- Discribes an r.v. which can take one of a countable number of values each outcome equally likely.

$U(a,b)$ $\quad n = b-a+1$ is the number of distinct possible values that x can take

p.m.f : $P(X=k) = \frac{1}{n}$ $\quad$ Mean: $\mu = \frac{a+b}{2}$

example: $X \sim$ Uniform $(1,6)$

Bernoulli Distribution: $X \sim$ Bernoulli (0.7) $p=0.7$

- Discribes an r.v. which is the result of a single experiment and can take values "success = 1" or "failure = 0"
- Single parameter "p" which refers to the probability of success   p=赢的概率
- $q = 1-p$ is the probability of failure.
- p.m.f: $P(X=k) \begin{cases} 1-p & \text{if } k=0 \\ p & \text{if } k=1 \end{cases}$ Mean: $\mu = p$

Binomial Distribution $X \sim$ Binomial $(1, 0.7)$ 1次实验, 概率 0.7 $(n=1, p=0.7)$.

- Discribes the number of success from repeated independent experiments each of which can take values "Success = 1" or "failure = 0"
- p = the prob. of success in each experiment   n = the number of experiments   k= 赢的次数
- p.m.f $= \binom{n}{k} p^k (1-p)^{n-k}$ Mean: $\mu = np$

$\binom{n}{k}$ = nCk

→ permutation: $nPk = \frac{n!}{(n-k)!}$ Combination: $nCk = \frac{nPk}{k!} = \frac{n!}{(n-k!)k!}$

$\binom{n}{k}$

Poisson Distribution:

- Describes an r.v. which is the number of events that take place per unit time
- One para. "λ" refers to the rate of events.

p.m.f: $P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$ Mean: $\mu = \lambda$ X: the number of events will happen in a unit of time

Notes: $A \sim Poi(a)$, $B \sim Poi(b)$ and $C = A+B$, then $C \sim Poi(a+b)$

Poisson to binomial!

Binomial $(n, p) \approx$ Poisson(λ) $(\lambda = np)$

Exponential Distribution:

- How to find probabilities for time (T) between events of r.v.

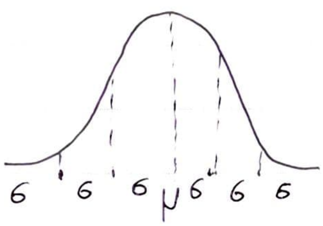
$X \sim$ Poisson (λ)

- If λ = event rate per unit time, mean time between events $= 1/\lambda$

P.d.f: $P(T=t) = \lambda e^{-t\lambda}$ c.d.f: $P(T \leq t) = 1 - e^{-t\lambda}$ Mean: $\mu = 1/\lambda$

Normal (Gaussian) Distribution:

- Symmetrical
- "Bell Shaped"
- Mean = Median = mode
- Location is deterimined by the mean μ
- Spread is determined by the standard deviation σ
- The variable has infinite theoretical range $+\infty$ to $-\infty$



等分6份

68% of obs. fall between Mean $\pm$ 1 SD

2.5% are $\geq$ Mean + 1.96 SD

5% are $\leq$ mean − 1.64 SD

CLT: Central Limit Theorm:

$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ 95% between

$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ ← ~~St~~ standard error

$\boxed{\text{The t-distribution}}$

small samples

para.: $\boxed{\text{Degrees of Freedom: Sample size} - 1}$

Standard Normal Distribution:

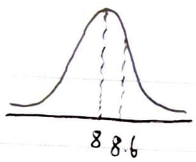
- Z distribution
- A normal distribution with Mean 0 and SD 1
- p.d.f. $f(x) = \frac{e^{-\frac{(x^2)}{2}}}{\sqrt{2\pi}}$ c.d.f: $F(x) = \int_{-\infty}^{x} \frac{e^{-\frac{(x^2)}{2}}}{\sqrt{2\pi}}$

Z scores:

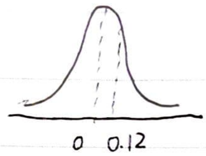
- Translate from X to the standardized normal

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

e.g. Suppose X is normal with mean 8.0 and standard deviation 5.0 Find $P(8 \le X \le 8.6)$



$\Rightarrow$



$$z_1 = \frac{x-\mu}{\sigma} = \frac{8-8}{5} = 0$$

$$z_2 = \frac{x-\mu}{\sigma} = \frac{8.6-8}{5} = 0.12$$

$$P(8 \le x \le 8.6) = P(0 \le z \le 0.12) = P(z \le 0.12) - P(z \le 0) = 0.55 - 0.5 = 0.05$$

↑ 查表得到

Estimation $\boxed{X \pm CV \frac{SD}{\sqrt{n}}}$

Find the X value for the known probability:

1. Find the z value for the known probability
2. Convert to x units using the formula: $\boxed{x = \mu + z\sigma}$

e.g. in z distribution, 95% was between ± 1.96, What about $Y \sim N(1,2)$ Normal, but not standard

Then, for our variable Y 95% will be between $1 \pm 1.96(2) \approx 1 \pm 3.92$ $(-2.92, 4.92)$

CI with normal distribution: z-distribution

90% CI means we want the middle 90% of the distribution

with Sample: size n. $\boxed{\mu \pm 1.96 \frac{\sigma}{\sqrt{n}}}$ (95% CI)

CI with t-distribution: t-distribution

$$\begin{cases} \alpha\text{-level} = 1 - \text{desired confidence level} \\ df = \text{sample size} - 1 \end{cases}$$

for: proportion $SE(\hat{p}) = \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$ use z-distribution

$\hat{p} \pm (CV) \cdot SE(\hat{p})$

↑ 查表, z distribution

always use z-distribution.

Always check $n\hat{p} > 10$ and $n(1-\hat{p}) > 10$