ST2132 Probability Part.

1. $X \sim Bin(n,p) \Leftrightarrow \sum_{i=1}^{n} Y_i$, $Y_i \sim Bernoulli(p)$.

$\therefore X = X_1 + \cdots + X_n$.

$\therefore E(X) = E(X_1) + \cdots + E(X_n) = n \cdot p$.

$V(X) = V(X_1) + \cdots + V(X_n) = n \cdot p \cdot q$.

For $k \in \{0,1,\cdots,n\}$ $I_k(X) = 1$ if $X = k$.

$E(I_k(X)) = \sum_{i=1}^{n} I_k(X) \cdot P(X=k)$

$= \sum_{i=1}^{n} E(X_i) = np$.

2. $I_A(X) = \begin{cases} 1, & X \in A \\ 0, & X \notin A \end{cases}$

$E(I_A(X)) = \sum_{i=1}^{n} I_A(X_i) \cdot p_i$

$= 1 \cdot P(X \in A) + 0 \cdot P(X \notin A) = P(X = x \in A)$.

3. $E(Z) = aE(X) + bE(Y) + c$

4. $\mu = E(X)$. $E\{X(X-\mu)\} = E(X^2 - \mu X) = E(X^2) - \mu E(X)$

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

$= E(X^2) - \mu^2 = var(X)$.

5. $var(X) = E((X-\mu)^2) = E(X^2) - [E(X)]^2$. $E(X)$

6. $SD(X) = \sqrt{var(X)}$ spare of X around its center.

Realisation of X:

$$\begin{cases} \bar{x} = \frac{1}{r}\sum_{i=1}^{r} x_i \\ \overline{x^2} = \frac{1}{r}\sum_{i=1}^{r} x_i^2 \end{cases}$$

realization

$$\underbrace{v = \frac{1}{r}\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$= \frac{1}{r}\sum_{i=1}^{n} (x_i^2 - 2\bar{x} \cdot x_i + \bar{x}^2)$

$= \frac{1}{r}\sum_{i=1}^{n} x_i^2 - 2\bar{x}\frac{1}{r}\sum_{i=1}^{n} x_i + r\bar{x}^2$

$= \overline{x^2} - 2\bar{x} \cdot \bar{x} + \bar{x}^2$

$= \boxed{\overline{x^2} - \bar{x}^2}$

denominator: $r-1$

(For correction).

7. LLN: $r \to \infty$, $\frac{1}{r}\sum_{i=1}^{r} h(X_i) \to E\{h(X)\}$

$$\begin{cases} \bar{x} \to E(X) \\ \overline{x^2} \to E(X^2) \\ v \to var(X) \end{cases}$$

Geometrically: $\bar{x}$ / SD / $\sqrt{v}$ are roughly the center and spread of X.

8. Draw without replacement $\Rightarrow X_1, X_2$.

Independent: ✗

$X \sim Exp(1) \Rightarrow f(x) = e^{-x}$, $x > 0$.

$E(\sqrt{X}) = \int_0^\infty \sqrt{x} \cdot e^{-x} dx = -\int_0^\infty \sqrt{x}\, de^{-x} \cdots$

9. Joint distribution:

$E\{h(X,Y)\} = \sum_{i=1}^{I}\sum_{j=1}^{J} h(x_i, y_j) \cdot f(x_i, y_j)$.

(a) $h(x,y) = x + y$.

$\therefore E(h(X,Y)) = \sum_{i=1}^{I}\sum_{j=1}^{J} (x+y) \cdot f(x_i, y_j)$

$= \sum_i \sum_j x \cdot f(x_i, y_j) + \sum_i \sum_j y_j f(x_i, y_j)$

$= \sum_i x_i \sum_j f(x_i, y_j) + \sum_j y_j \sum_i f(x_i, y_j)$

$= \sum_i x_i f(x_i) + \sum_j y_j f(y_j)$

$= E(X) + E(Y)$.

(b) $h(x,y) = x \Rightarrow$ By (a)

$h(x,y) = x = x + 0 \Rightarrow y = 0$

$\therefore E(h(x,y)) = E(X) + \underbrace{E(Y)}_{0} = E(X)$.

10. Continuous Joint Distribution.

① $f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$.

② $f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$

③ $P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x,y)\,dx\,dy$

11. Conditional

① $f_Y(y_i|x_i) = \frac{f(x_i, y_i)}{f_X(x_i)}$

② $f_X(x_i|y_i) = \frac{f(x_i, y_i)}{f_Y(y_i)}$

Discrete:

$E(Y|x_i) = \sum_{j=1}^{J} y_j f_Y(y_j|x_i)$.

$var[Y|x_i] = \sum_{j=1}^{J} (y_j - E[Y|x_i]) \cdot f_Y(y_j|x_i)$

$\boxed{= E((Y - E(Y|x_i)))}$

Continuous:

$E(Y|x_i) = \int_{-\infty}^{\infty} y f_Y(y|x)\,dy$.

$var[Y|x_i] = \int_{-\infty}^{\infty} (y - E(Y|x_i))^2 f_Y(y|x_i)\,dy$.

$var[Y|x_i] = E[Y^2|x_i] - E[Y|x_i]^2$.

12. Covariance / Independent.

D1· Random vector

$$X := \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \quad E(X) := \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_k) \end{bmatrix}$$

$$var(X) = \begin{bmatrix} var(X_1) & cov(X_1,X_2) & \cdots & cov(X_1,X_k) \\ cov(X_2,X_1) & var(X_2) & \cdots & \vdots \\ \vdots & \cdots & \cdots & var(X_k) \end{bmatrix}$$

$X_{k\times k}$

$$\text{For } a_{ij} = \begin{cases} var(X_i), & \text{if } i=j \\ cov(X_i,X_j), & \text{if } i\neq j \end{cases}$$

$$\begin{cases} E(AX+b) = A\,E(X)+b \\ var(AX+b) = A\,var(X)\,A^T \end{cases}$$

$$X_1 \cdots X_n \Rightarrow \text{IID} \Rightarrow var(X) = \begin{bmatrix} \sigma^2 & & \\ & \sigma^2 & \\ & & \ddots \ \sigma^2 \end{bmatrix}$$

(μ, σ^2)

$$E(X) = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix}$$

$$S_n = \sum_{i=1}^{n} X_i \Rightarrow \begin{cases} E(S_n) = n\,E(X_i) = n\mu \\ var(S_n) = n\,var(X_i) = n\sigma^2 \\ SD(S_n) = \sqrt{n}\,\sigma \end{cases}$$

$X_i \sim \text{Bernoulli}(p)$. $S_n = \sum_{i=1}^{n} X_i \sim \text{Bin}(n,p)$.

Mutinomial distribution: $\Rightarrow \sum_{i=1}^{k} \text{Bin}(n_i, p_i)$.

n runs → k outcomes: $E_1 \cdots E_k$.
$p_1 \cdots p_k$.

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \sim \text{Mutinomial}(n, \underset{\sim}{p})$$

$$\therefore P(X_1 = x_1 \cdots X_k = x_k) = \left(\frac{n!}{x_1! \cdots x_k!}\right) p_1^{x_1} \cdots p_k^{x_k}$$

$\underset{\sim}{p} = [p_1 \cdots p_k]^T$

X_i or X_j happens.

$$\left.\begin{array}{l} X_i \sim \text{Bin}(n, p_i) \\ X_j \sim \text{Bin}(n, p_j) \end{array}\right\} \Rightarrow X_i + X_j \sim \text{Bin}(n, p_i + p_j)$$

MSE: $E\{(Y-c)^2\}$

$$= E\{Y^2 - 2cY + c^2\}$$
$$= E(Y^2) - 2cE(Y) + c^2$$
$$= E(Y^2) - [E(Y)]^2 + [E(Y)]^2 - 2cE(Y) + c^2$$
$$= var(Y) + \{E(Y) - c\}^2 \geq var(Y)$$

$\Rightarrow c = E(Y)$. $\min MSE(Y) = var(Y)$.

Correlated RV. X, Y.

MSE: $E[(Y-c)^2 \mid X] = var(Y \mid X) + \{E(Y \mid X) - c\}^2$

$\Rightarrow E((Y-c)^2 \mid X)\ \min = var(Y \mid X)$ when $c = E(Y \mid X)$

Predict Y from X Given X if $c = E(Y)$.

$$MSE = var[Y \mid X] + \{E(Y \mid X) - E(Y)\}^2$$

Main Idea: MSE.

$$var(Y-c) = E\{(Y-c)^2\} - \{E(Y-c)\}^2 = var(Y)$$

RV conditional E/var

① $E[X_2 \mid x_1] \xrightarrow{X_1} E[X_2 \mid X_1]$ (formulation)

② $E(E[X_2 \mid X_1]) = E(X_2)$

③ $var(E[X_2 \mid X_1]) + E(var[X_2 \mid X_1]) = var(X_2)$

Mean MSE (for relasations)

$$\frac{1}{n}\sum_{i=1}^{n} var[Y \mid x_i] \xrightarrow{n\to\infty} E[var[Y \mid X]]$$

If using $E(Y)$ instead of $E[Y \mid X]$.

$\Rightarrow MSE = var(Y)$

$$var(Y) - E[var[Y \mid X]] = var(E[Y \mid X])$$

① prediction is based on realization of X instead of X itself.

② $E(Y \mid X) \Rightarrow$ RV of X insted of Y

③ Never think about $E(X=X)$ kinda thing

Sample variance: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

① $E(S^2) = \sigma^2$

② $var(S^2) = \frac{2\sigma^4}{n-1}$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\begin{cases} E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \\ var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2n-1 \end{cases}$$

ST2132 Probability Part(II) · Distribution ·

Normal: $\phi(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$.

$N(\mu, \sigma^2)$

let $X \sim N(\mu, \sigma^2)$ $Y \sim N(\nu, \tau^2)$.

① $aX + b \sim N(a\mu + b, a^2\sigma^2)$.

② $X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2 + 2cov(X, Y))$.

IID: $cov = 0$

CLT: $S_n = \sum_{i=1}^{n} X_i$, $n \to \infty$, $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$.

μ, σ → X_i Not S_n.

$E(S_n) = n\mu$.

$var(S_n) = n\sigma^2$.

$\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1) \Leftrightarrow S_n \sim N(n\mu, n\sigma^2)$.

standardized.

Example: $X_i \sim Bernoulli(p)$.

$S_n = \sum_{i=1}^{n} X_i \sim Bin(n, p)$.

$\therefore var(S_n) = np(1-p) \Rightarrow \sigma = \sqrt{np(1-p)}$.

$E(S_n) = np$.

$\therefore n \to \infty$. $\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0,1)$.

IID Normal RVs. $X_1, X_2 \cdots X_n \sim N(\mu, \sigma^2)$.

① $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow \begin{cases} E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \\ var(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n} var(X_i) = \frac{\sigma^2}{n} \end{cases}$

② $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$. $\Rightarrow$ ③ $\frac{(\bar{X}-\mu)^2}{\sigma^2/n} = \frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi_1^2$.

④ $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$

⑤ $\bar{X}$ and S^2 independent

$\Rightarrow$ ⑥ $\frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}{\sqrt{\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi_{n-1}^2} / (n-1)}} \sim t_{n-1}$

denominator.

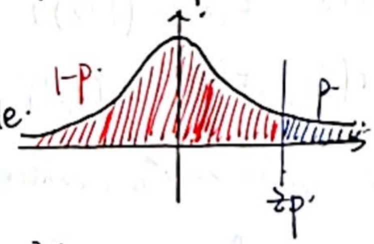
⑦ $\underbrace{\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2}_{\chi_n^2} = \underbrace{\boxed{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2}}_{\chi_{n-1}^2} + \underbrace{n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2}_{\chi_1^2}$ justify. (by tut1).

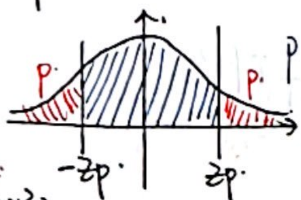
upper-tail quantile.

let $Z \sim N(0,1)$, $0 < p < 1$. let z_p s.t.

$p = Pr(Z > z_p)$.

z_p is $(1-p)$-quantile of Z.





$P(-z_p < Z < z_p) = 1 - 2p$

Normal MOM: let $X_1 \cdots X_n \sim IID\ N(\mu, \sigma^2)$

$\mu_1 = \mu$, $\mu_2 = \sigma^2 + \mu^2$.

① MOM estimator:

(i) $\hat{\mu} = \hat{\mu}_1 = \bar{X}$

(ii) $\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

② (i) $E(\hat{\mu}) = E(\bar{X}) = \mu$ → unbiased

(ii) $E(\hat{\sigma}^2) = E(\frac{n-1}{n} S^2) = \frac{n-1}{n} E(S^2) = \frac{(n-1)\sigma^2}{n}$

biased!

χ^2-distribution: $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi_1^2$

$var(Z) = E(Z^2) - (E(Z))^2 = \boxed{E(Z^2) = 1}$

$\boxed{var(Z^2)} = E(Z^4) - E(Z^2)^2 = 3 - 1 = 2$

$V = \sum_{i=1}^{n} V_i \sim \chi_n^2$, where $V_i \sim \chi_1^2$ $\boxed{\text{IID RV}}$ ☆

$E(V) = \sum_{i=1}^{n} E(V_i) = \sum_{i=1}^{n} E(Z^2) = n$

$var(V) = \sum_{i=1}^{n} var(V) = 2n$

Gamma: $\alpha > 0$, $\lambda > 0$, $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, $x > 0$

Note that $\chi_1^2 \sim Gamma(\frac{1}{2}, \frac{1}{2}) \Leftarrow f(x) = \frac{1}{\sqrt{2\pi}} v^{-\frac{1}{2}} e^{-\frac{v}{2}}$

Theorem: $X_1 \sim Gamma(\alpha_1, \lambda)$ $X_2 \sim Gamma(\alpha_2, \lambda)$

$\therefore X_1 + X_2 \sim Gamma(\alpha_1 + \alpha_2, \lambda)$

$\Rightarrow \chi_n^2 \sim Gamma(\frac{n}{2}, \frac{1}{2})$

Theorem 2: $\int_0^{+\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = 1 \Rightarrow \frac{\Gamma(\alpha)}{\lambda^\alpha} = \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx$

$E(X) = \frac{\alpha}{\lambda}$ $\quad E(X^2) = \frac{(\alpha+1)\alpha}{\lambda^2}$

$var(X) = \frac{\alpha}{\lambda^2}$

t-distribution: $t_n = \frac{Z}{\sqrt{V/n}}$, $Z \sim N(0,1)$, $V \sim \chi_n^2$

F-distribution: $F_{m,n} = \frac{V/m}{W/n}$, $V \sim \chi_m^2$, $W \sim \chi_n^2$

$T \sim t_n \Rightarrow T^2 \sim \frac{Z^2}{V/n} = \frac{W/1}{V/n} \sim F_{1,n}$

ST2132 Population mean.

Population variable: RV. discrete.

discrete $\mu := \frac{1}{n}\sum_{i=1}^{n} v_i$ $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(v_i-\mu)^2$

parameter: unknow quantity.

Example:

① population with μ, σ^2 draw, $X = v$-value.

(i) X has the same distribution of v.

(ii) $E(X)=\mu$, $var(X)=\sigma^2$.

estimate: μ.

② Assume in ①, $v_1 \cdots v_n$ are distinct.

$E(X) = \sum_{i=1}^{N} v_i \frac{1}{N} = \mu$ $E(X^2) = \sum_{i=1}^{N} v_i^2 \frac{1}{N}$

Random Sample: $\boxed{X_1 \cdots X_n}$ RV. draw from population μ, σ^2 with replacement

① $X_1 \cdots X_n$ are RVs with μ and σ^2.

② $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \Rightarrow E(\bar{X}) = \mu$, $var(\bar{X}) = \frac{\sigma^2}{n}$.

③ let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mu$. unbias

reasonable estimate of μ since $E(\bar{X}) = \mu$.

Idea: estimate mean of population → estimate expection of RV. (Random draw)

Error let $\bar{x}$ be mean of $x_1, x_2 \cdots x_n$.

Note that error in $\bar{x}$: $\mu - \bar{x}$ cannot estimate.

use standard error $\boxed{SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}}$ to indicate size of error in $\bar{x}$ fixed.

Estimate: σ. Use sample variance.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

① $E(S^2) = \sigma^2$. $E\left(\sum_{i=1}^{n}(X_i-\mu)^2\right) = n\sigma^2$.

② $\sum_{i=1}^{n}(X_i - \bar{X})^2 \le \sum_{i=1}^{n}(X_i-\mu)^2$.

⇓

$E\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = (n-1)\sigma^2$.

③ $\hat{\sigma}^2 = \frac{n-1}{n} S^2 \Rightarrow E(\hat{\sigma}^2) = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2$.

※ SE is estimated as $\frac{s}{\sqrt{n}} = \frac{\sqrt{\frac{n}{n-1}\sigma^2}}{\sqrt{n}}$

※ $\mu \approx \bar{x} \pm \frac{s}{\sqrt{n}}$

Binary population: 1-0 pop with p.

$\mu = p$, $\sigma^2 = p(1-p)$, $X_1 \cdots X_n$ draws with replacement.

① $S_n = X_1 + \cdots + X_n \sim Bin(n, p)$

② $\bar{X} = S_n/n = \hat{p} \Rightarrow E(\hat{p}) = p$ $var(\hat{p}) = p(1-p)/n$

$SD(\hat{p}) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$

Estimate p: $p \approx \hat{p} \pm \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$

p is estimated as $\hat{p}$. → estimate / estimator. > Same symbol

Point of estimate: $\bar{x} \pm \frac{s}{\sqrt{n}}$.

interval estimation: works well in.

① $X_i \sim N(\mu, \sigma^2)$, $n \in N^*$.

② large number, other distribution.

[Refer Normal distribution for z quantile]

① Assume $X_1 \cdots X_n \sim IID\ N(\mu, \sigma^2)$ σ^2 known

$P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$.

$\Rightarrow P\left(\mu - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$.

$\Rightarrow P\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$.

Random interval: $\left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

realistion $\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$ $(1-\alpha)$ conf. interv.

※ We have $(1-\alpha)$ confidente that μ is between

NOT Probability. $\Rightarrow Pr(\square < \mu < \square) = 1-\alpha$ FALSE.

② Assume IID RV, $N(\mu, \sigma^2)$ σ^2 unknown

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1} \Rightarrow \left(\bar{x} - t_{\frac{\alpha}{2}, n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}, n-1}\frac{s}{\sqrt{n}}\right)$

~~③~~ If μ, σ^2 unknown, n large

$\because$ $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$ & large N: $S \approx \sigma$.

$\therefore$ $\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$

* Actually a bit less than $(1-\alpha)$, converge to $(1-\alpha)$ as $n\to\infty$.

* Special case: Bernoulli ⇒ $S = \sqrt{\hat{p}(1-\hat{p})}$.

Bias and Unbias.

let estimator X estimate w by realisation x.

is x unbiased? $\Leftrightarrow$ is X unbiased? $\Leftrightarrow$ $E(X)=w$?

$SE = SD(X) = \sigma$.　unknown: $n+1$.

$\therefore$ we have $x_i = w + e_i$ (e_i is error for x_i).

$\Rightarrow E(\bar{X}) = w + b$ (b is bias parameter).

$MSE := E\{(\bar{X}-w)^2\} = \frac{\sigma^2}{n} + b^2 = SE^2 + bias^2$.

$n \to \infty$, $\frac{\sigma^2}{n} \to 0 \Rightarrow MSE = \boxed{bias^2}$

no affect.

Generalisation: let $\hat{\theta}$ be estimator for IID RVs.

1. $SE = SD(\hat{\theta})$.
2. $Bias := E(\hat{\theta}) - \theta$　$\Rightarrow MSE = \boxed{SE^2 + Bias^2}$.
3. $MSE = E\{(\hat{\theta}-\theta)^2\}$

Formula sheet.

ST2132. MOM

typically continous regardless distribution.

Parameter Space: contains every possible value that parameter could be.

using Θ to refer to the family of distribution

Bernoulli(p): $(0,1) = \{p \in R, 0 < p < 1\}$.

Possion(λ): $R+ = \{\lambda \in R: \lambda > 0\}$

Exponential(λ) $R+ = \{\lambda \in R, \lambda > 0\}$.

Normal(μ, σ^2). $R \times R+ = \{\mu \in R, \sigma^2 \in R+: \sigma^2 > 0\}$ R. We interested in μ. largest

k-th moment of RV X is $\mu_k := E(X^k)$.

	μ_1	μ_2
Bernoulli(p).	p	$\mu_2 = var + \mu_1^2 = p$.
Normal(μ, σ^2)	μ	$\mu^2 + \sigma^2$
Bin(n,p)	np	$n(1-p)p + n^2p^2$ * Not np.
Exp(λ).	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2} + (\frac{1}{\lambda})^2 = \frac{2}{\lambda^2}$
Possion(λ)	λ	$\lambda + \lambda^2$.

Estimating moments: $X_1 \cdots X_n$, IID RV.

estimator $\widehat{\mu_k} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ same symbol different meaning.

$\downarrow$ estimate

$\mu_k \Rightarrow \widehat{\mu_k} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$.

estimate.

$\therefore E(\widehat{\mu_k}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i^k) = \frac{1}{n}\sum_{i=1}^{n} E(X^k) = E(X^k) = \mu_k$.

$\therefore$ unbiased estimator!

$SE = \frac{\sigma}{\sqrt{n}}$.

Poisson MOM:

① let $X \sim$ Poisson(λ), $\mu_1 = \lambda$

② MOM estimator of λ is $\hat{\lambda} = \widehat{\mu_1} = \bar{X}$.

③ $var(X) = \lambda \Rightarrow SE = SD(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{\lambda}}{\sqrt{n}}$.

Bernoulli MOM.

① let $X \sim$ Bernoulli(p), $\mu_1 = p$.

② MOM estimate of p is $\hat{p} = \widehat{\mu_1} = \bar{X}$.

③ $var(X) = p(1-p) \Rightarrow SE = SD(\bar{X}) = \frac{\sqrt{p(1-p)}}{\sqrt{n}}$

Geometric distribution.

let $x_1 \cdots x_{20}$ from IID Geometric RV. $X_1 \cdots X_{20}$

$X \sim$ Geometric(p). $P(X=x) = (1-p)^{x-1} \cdot p$.

$E(X) = \frac{1}{p} \Rightarrow \mu_1 = \frac{1}{p}$

$\therefore$ ① MOM estimator: $\hat{p} = \frac{1}{\widehat{\mu_1}} = \frac{1}{\bar{X}}$

② estimate of p: $= \frac{1}{\bar{x}} = 0.71$.

Bootstrap: replace para by estimates create IID.

let $X_1^* \cdots X_{20}^* \sim$ Geo(0.71).

$\Rightarrow SD(1/\bar{X}) \approx SD(1/\bar{X}^*)$

$\bar{X}$ replaced by $\bar{X}^*$.
p replaced by 0.71.

let $\bar{x}_1^*, \cdots \bar{x}_r^*$ be realisation of $\bar{X}^*$.

$\therefore SD(1/\bar{X}^*) \approx SD\left(\frac{1}{\bar{x}_1^*} \cdots \frac{1}{\bar{x}_r^*}\right)$.

SE in 0.71, with r

Bias $= E(1/\bar{X}) - p \approx E(1/\bar{X}^*) - 0.71$

$\Rightarrow$ Monte Carlo approximation of $1/\bar{X}^*$.

let $\bar{x}_1^* \cdots \bar{x}_r^*$ be r simulations of $\bar{X}^*$.

$\therefore E(1/\bar{X}^*) \approx \frac{1}{r}\sum_{i=1}^{n} \frac{1}{\bar{x}_i^*}$ $\therefore$ get bias.

$\therefore$ We have bias-corrected estimation.

$p \approx (0.71 - \text{bias}) \pm SE$

Gamma Distribution with Bootstrap.

Let $X_1 \cdots X_n \sim$ Gamma(α, λ) RV., $\alpha > 0$, $\lambda > 0$.

Note that $\mu_1 = \frac{\alpha}{\lambda}$, $\mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$.

$\therefore$ We have $\alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$ $\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$.

$\therefore$ MOM estimator

(i) $\hat{\alpha} = \frac{\widehat{\mu_1}^2}{\widehat{\mu_2} - \widehat{\mu_1}^2} = \frac{\widehat{\mu_1}^2}{\hat{\sigma}^2} = \frac{\bar{X}^2}{\hat{\sigma}^2}$

(ii) $\hat{\lambda} = \frac{\widehat{\mu_1}}{\widehat{\mu_2} - \widehat{\mu_1}^2} = \frac{\bar{X}}{\hat{\sigma}^2}$

Use $\bar{x}, \hat{\sigma}^2$ from data $\Rightarrow \hat{\alpha} \approx 0.38$ $\hat{\lambda} \approx 1.68$.

Bootstrap: let $X_1^* \cdots X_n^* \sim$ IID Gamma(0.38, 1.68)

$\therefore SD(\frac{\bar{X}^2}{\hat{\sigma}^2}) \approx SD(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}})$. $SD(\frac{\bar{X}}{\hat{\sigma}^2}) = SD(\frac{\bar{X}^*}{\hat{\sigma}^{*2}})$.

Monte Carlo: let $(\bar{x}_1^*, \hat{\sigma}_1^{*2}) \cdots (\bar{x}_r^*, \hat{\sigma}_r^{*2})$ r sim..

Approximation:

$$SD\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^2}\right) = SD\left(\frac{\bar{x}_1^{*2}}{\hat{\sigma}_1^{*2}}, \dots \frac{\bar{x}_r^{*2}}{\hat{\sigma}_r^{*2}}\right).$$

$$SD\left(\frac{\bar{X}^*}{\hat{\sigma}^{*2}}\right) = SD\left(\frac{\bar{x}_1^*}{\hat{\sigma}_1^{*2}} \dots \frac{\bar{x}_r^*}{\hat{\sigma}_r^{*2}}\right).$$

Bias:

① $E\left(\frac{\bar{X}^2}{\hat{\sigma}^2}\right) - \alpha \approx E\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}}\right) - 0.38$

② $E\left(\frac{\bar{X}}{\hat{\sigma}^2}\right) - \lambda = E\left(\frac{\bar{X}^*}{\hat{\sigma}^{*2}}\right) - 1.68$

Monte Carlo:

$$E\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}}\right) = \frac{1}{r}\sum_{i=1}^{r} \frac{\bar{x}_i^{*2}}{\hat{\sigma}^{*2}}.$$

$$E\left(\frac{\bar{X}^*}{\hat{\sigma}^{*2}}\right) = \frac{1}{r}\sum_{i=1}^{r} \frac{\bar{x}_i}{\hat{\sigma}^{*2}}.$$

∴ Bias correction estimation:

$$\alpha \approx (0.38 - \text{bias}) \pm SD\left(\frac{\bar{X}^{*2}}{\hat{\sigma}^{*2}}\right).$$

$$\lambda \approx (1.68 - \text{bias}) \pm SD\left(\frac{\bar{X}^*}{\hat{\sigma}^{*2}}\right).$$

MOM is consistent. when $n \to \infty$

① $SE \to 0$.

② $\text{bias} \to 0$. ⇒ hit the target.

unlike Gauss Model (inconsistent)

ST2132. MLE. Assume better than MOM: asymptotically unbiased / normality distribute.

let $x_1, x_2 \cdots x_n$ be realisation of IID RVs. $X_1 \cdots X_n$ with mass / density $f(\cdot)$ specified by $\theta \in \Theta \subset \mathbb{R}^k$. General idea: $l(\hat{\theta}) \geq l(\theta)$ for any choice of $\hat{\theta}$

① Likelihood function: $L: \Theta \to \mathbb{R}$:

$$L(\theta) = P(X_1 = x_1, X_2 = x_2 \cdots X_n = x_n) = \prod_{i=1}^{n} f(x_i).$$

② log likelihood: $l(\theta) = \log L(\theta)$

③ MLE of θ: maximiser of $l(\theta)$ and $L(\theta)$.
estimate $\to$ estimator: $\hat{\theta}$

④ ML estimator: $\hat{\theta}$ for $\prod_{i=1}^{n} f(X_i)$.

Ex(a) Poisson MLE. $\therefore f(x_i) = \frac{\lambda^x e^{-\lambda}}{x!}$

$\mathbb{R}^+ \to \mathbb{R}^+$

$$\therefore L(\lambda) = \frac{\lambda^{\sum_{i}^{n} x_i} e^{-n\lambda}}{x_1! \cdots x_n!} \qquad l(\lambda) = \sum_{i=1}^{n} x_i \log\lambda - n\lambda - \sum_{i=1}^{n} \log x_i!$$

$L(\bar{x}) \geq L(\lambda) \Rightarrow \rho_{max}$

$$\therefore l'(\lambda) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

$\therefore$ ML estimator: $\hat{\lambda} = \bar{X}$

(b) Normal MLE: $f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $x_i \in \mathbb{R}$

$\mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$

$$\therefore L(\mu, \sigma) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \cdot e^{-\sum_{i=1}^{n}\left(\frac{(x_i-\mu)^2}{2\sigma^2}\right)}.$$

$$\therefore l(\mu, \sigma) = -\frac{n}{2}\log 2\pi - n\log\sigma - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}.$$

$$\frac{\partial l(\mu,\sigma)}{\partial\mu} = \sum \frac{1}{2\sigma^2} \cdot 2(x_i - \mu) = \sum \frac{x_i - \mu}{\sigma^2} = 0 \Rightarrow \mu = \bar{x}$$

$$\frac{\partial l(\mu,\sigma)}{\partial\sigma} = \frac{1}{\sigma}\cdot(-n) + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3} = 0 \Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \Rightarrow \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\therefore \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \end{cases}$$

(c) Gamma estimation $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$.

Track: $x \Rightarrow \log +$ take log first then $\sum n$.

$$\log f(x_i) = \alpha\log\lambda - \log\Gamma(\alpha) + (\alpha-1)\log x - \lambda x$$

$$\therefore l(\alpha,\lambda) = \sum_{i=1}^{n} \log f(x_i) = n\alpha\log\lambda - n\log\Gamma(\alpha) - \lambda\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}(\alpha-1)\log x_i$$

$$\therefore \frac{\partial l(\alpha,\lambda)}{\partial\lambda} = \frac{n\alpha}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \boxed{\lambda = \frac{\alpha}{\bar{x}}}$$

$$\frac{\partial l(\alpha,\lambda)}{\partial\alpha} = n\log\lambda - \frac{n}{\Gamma(\alpha)}\cdot\frac{d\Gamma(\alpha)}{d\alpha} + \sum_{i=1}^{n}\log x_i = 0$$

$$n\log\lambda - n\psi(\alpha) + n\bar{y} = 0$$

$$\begin{cases} \psi(\alpha) = \frac{d\log\Gamma(\alpha)}{d\alpha} \\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n}\log x_i \end{cases} \Rightarrow \log\frac{\alpha}{\bar{x}} - \psi(\alpha) + \bar{y} = 0$$

$$\therefore \text{ML estimator: } \begin{cases} \hat{\lambda} = \frac{\hat{\alpha}}{\bar{X}} \\ \hat{\alpha}: \log\left(\frac{\hat{\alpha}}{\bar{X}}\right) - \psi(\hat{\alpha}) + \bar{Y} = 0 \end{cases}$$

Bias and SE for estimator: Bootstrap and Monte Carlo.

Do not write $G \cdot (\alpha 441, 1.96)$

let $X_1^* \cdots X_n^*$ IID Gamma(α^*, λ^*)

where $\alpha^* = \boxed{}$ $\lambda^* = \boxed{}$ get from estimate

$$\therefore SD(\hat{\alpha}) = SD(\hat{\alpha}^*) \cdot \quad E(\hat{\alpha}) - \alpha \approx E(\hat{\alpha}^*) - \alpha^*.$$

$$SD(\hat{\lambda}) = SD(\hat{\lambda}^*) \quad E(\hat{\lambda}) - \lambda \approx E(\hat{\lambda}^*) - \lambda^*.$$

Monte Carlo: simulate $(\hat{\alpha}_1^*, \hat{\lambda}_1^*) \cdots (\hat{\alpha}_n^*, \hat{\lambda}_n^*)$.

$$\therefore \begin{cases} SD(\hat{\alpha}^*) \approx SD(\hat{\alpha}_1^*, \hat{\alpha}_2^* \cdots \hat{\alpha}_n^*) \\ E(\hat{\alpha}^*) \approx \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i^* \end{cases}$$

[$\hat{\lambda}$ similarly]

$\therefore \Rightarrow$ get $\boxed{\text{bias-correction}}$ $(\alpha^* - \text{bias}) \pm SE$

ML estimates have smaller SE/bias than MOM

$\begin{cases} \text{MLE: No formulae} \Rightarrow \text{have to use } \hat{\lambda}^*, \hat{\mu}^* \\ \text{MOM: estimator explicitly donate by } \bar{X}^*, S^* \end{cases}$

Simplify L and l:

① In L, any constant factor can be left out.

② In l: any additive constant can be left out.

(d) Multinormal data. $(X_1 \cdots X_k) \sim \text{Multi}(n, p)$.

$\therefore L(p) = p_1^{x_1} \cdots p_k^{x_k}$

$l(p) = x_1 \log p_1 + \cdots + x_k \log p_k$. > after simplify.

Use Lagrange multiplier maximize: $\hat{p}_i = \frac{X_i}{n}$.

HWE example. $(X_1, X_2, X_3)^T \sim \text{Multi}(3, p)$.

$p_1 = (1-\theta)^2 \quad p_2 = 2\theta(1-\theta) \quad p_3 = \theta^2$.

$L(\theta) = (1-\theta)^{2X_1} \cdot [\theta(1-\theta)]^{X_2} \cdot \theta^{2X_3}$. (simplify) remove 2.

$\therefore l(\theta) = (X_2 + 2X_3)\log\theta + (2X_1 + X_2)\log(1-\theta)$.

$\hat{\theta} = \frac{X_2 + 2X_3}{2n}$ $\therefore \text{var}(\hat{\theta}) = \frac{\theta(1-\theta)}{2n}$.

Asymptotic normality: $n \to \infty$

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, I(\theta)^{-1}).$$

$$\Downarrow$$

$$\hat{\theta} \to N\left(\theta, \frac{I^{-1}(\theta)}{n}\right).$$

$\boxed{E(\hat{\theta}_n) = \theta}$, $\text{var}(\hat{\theta}_n) = \frac{I^{-1}(\theta)}{n}$

unbiased. estimate.

$(1-\alpha)$ CI: $\hat{\theta}_n \pm z_{\frac{\alpha}{2}} \cdot SE = \hat{\theta}_n \pm z_{\frac{\alpha}{2}} \sqrt{\frac{I(\theta)^{-1}}{n}}$

Condition: ① $f(\cdot)$ is differentiable in θ.

② each elem. specify a unique distribution.

③ All distributions: same set of possible values.

Fact. $\hat{\theta} \sim$ ML estimator of θ.

for strictly increasing / decreasing h: $\Rightarrow$

we have. $h(\hat{\theta})$ is ML estimator of $h(\theta)$.

$n \to \infty$. $h(\hat{\theta}) \to$ normal.

Fisher Information: $I(\theta) = -E\left[\frac{d^2 \log f(x)}{d\theta^2}\right]$

① Measures info. about θ in one sample X.

② $I(\theta)$ is symmetric. (i, j) entry: $-E\left[\frac{\partial^2 \log f(x)}{\partial\theta_i \partial\theta_j}\right]$

Cramer-Rao inequality:

$$\text{var}(\hat{\theta}_n) \geq \boxed{\frac{I(\theta)^{-1}}{n}}.$$

cramer rao lower bound.
(CRLB).

Property: If an unbiased estimator's var. is CRLB $\Rightarrow$ it is efficient.

function of. $X_1 \cdots X_n$.

ML estimator is asymptotically $\begin{cases} \text{efficient!} \\ \text{unbiased} \\ \text{normal.} \end{cases}$

Best large sample estimator.

SE estimation / CI construction are fine with out Monte Carlo.

Fisher information example:

① Bernoulli(p). Discrete ML estimator: $\hat{p}$.

$f(x) = p^x(1-p)^{1-x} \Rightarrow \log f(x) = x\log p + (1-x)\log(1-p)$.

$\therefore \frac{d\log f(x)}{dp} = \frac{X}{p} - \frac{1-X}{1-p}$

$\frac{d^2\log f(x)}{dp^2} = -\frac{X}{p^2} - \frac{1-X}{(1-p)^2}$

$\therefore I(p) = -E\left(-\frac{X}{p^2} - \frac{1-X}{(1-p)^2}\right) = \frac{1}{p(1-p)}$

When $n\to\infty$ the distribution $\to N\left(p, \frac{p(1-p)}{n}\right)$ ← rough.

Never be continuous.

$E(\hat{p}) = p$. $var(\hat{p}) = \frac{p(1-p)}{n}$ exact.

② $X \sim$ Geometric(p). $n\to\infty$ Distribution $\to N\left(p, \frac{p^2(1-p)}{n}\right)$.

can not use CLT as $\hat{p} = 1/\bar{X}$.

$\log f(X) = (X-1)\log(1-p) + \log p$. $\boxed{E(\hat{p})} > E(p)$.

$\therefore \frac{d\log f(x)}{dp} = -\frac{X-1}{1-p} + \frac{1}{p}$ By ($E(\frac{1}{X}) \geq \frac{1}{E(X)}$, X positive)

$\boxed{var(\hat{p})} \approx \frac{p^2 p(1-p)}{n}$

$\frac{d^2\log f(x)}{dp^2} = -\frac{X-1}{(1-p)^2} - \frac{1}{p^2}$ rough.

$\therefore I(p) = -E\left(-\frac{X-1}{(1-p)^2} - \frac{1}{p^2}\right) = \frac{1}{p}\left(\frac{1}{1-p} + \frac{1}{p}\right)$

$= \frac{1}{p^2(1-p)}$

③ $X \sim N(\mu, \sigma^2)$. $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $x\in\mathbb{R}$.

$\log f(X) = -\frac{1}{2}\log 2\pi - \log\sigma - \frac{(X-\mu)^2}{2\sigma^2}$

$\therefore \frac{\partial\log f(x)}{\partial\sigma} = -\frac{1}{\sigma} + \frac{(X-\mu)^2}{\sigma^3}$ $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

$\frac{\partial^2\log f(X)}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}$ $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2} \sim N\left(\sigma, \frac{\sigma^2}{2n}\right)$.

$\therefore$ Approximate $(1-\alpha)$ CI.

$\frac{\partial\log f(x)}{\partial\mu} = \frac{X-\mu}{\sigma^2}$ $\frac{\partial^2\log f(x)}{\partial\mu^2} = \frac{-1}{\sigma^2}$ μ: $\left(\bar{X} - z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\right)$

σ: $\left(\hat{\sigma} - z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{2n}}, \hat{\sigma} + z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{2n}}\right)$.

$\therefore \frac{\partial^2\log f(X)}{\partial\mu\partial\sigma} = \frac{-2(X-\mu)}{\sigma^3}$

Here s is not used

$\because n \to$ large

$s \to \sigma$

$$\therefore E\left(-\frac{d^2\log f(x)}{d\theta^2}\right) = \begin{bmatrix} E\left(-\frac{\partial^2 f(x)}{\partial\mu^2}\right) & E\left(-\frac{\partial^2 f(x)}{\partial\mu\partial\sigma}\right) \\ E\left(-\frac{\partial^2 f(x)}{\partial\mu\partial\sigma}\right) & E\left(-\frac{\partial^2 f(x)}{\partial\sigma^2}\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

Note that $E\left((X-\mu)^2\right) = E\left(X^2 - 2\mu X + \mu^2\right) = \sigma^2 + \mu^2 - 2\mu^2 + \mu^2 = \sigma^2$.

④ Gamma(α, λ). $f(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)}X^{\alpha-1}e^{-\lambda X}$, $X>0$

$\log f(x) = \alpha\log\lambda - \log\Gamma(\alpha) + (\alpha-1)\log X - \lambda X$

$\therefore \frac{\partial\log f(x)}{\partial\alpha} = \log\lambda - \frac{d\log\Gamma(\alpha)}{d\alpha} + \log X$

$= \log\lambda - \psi(\alpha) + \log X$

$\frac{\partial^2\log f(x)}{\partial\alpha^2} = -\psi'(\alpha)$

$\frac{\partial\log f(x)}{\partial\lambda} = \frac{\alpha}{\lambda} - X$

$\frac{\partial^2\log f(x)}{\partial\lambda^2} = -\frac{\alpha}{\lambda^2}$

$\frac{\partial^2\log f(x)}{\partial\alpha\partial\lambda} = \frac{1}{\lambda}$

$$\therefore I(\alpha,\lambda) = \begin{bmatrix} \psi'(\alpha) & -\frac{1}{\lambda} \\ -\frac{1}{\lambda} & \frac{\alpha}{\lambda^2} \end{bmatrix}.$$

$$\Rightarrow \begin{bmatrix} \hat{\alpha} \\ \hat{\lambda} \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha \\ \lambda \end{bmatrix}, \frac{I(\theta)^{-1}}{n}\right)$$

$$var\begin{bmatrix} \hat{\alpha} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} var(\hat{\alpha}) & cov(\hat{\alpha},\hat{\lambda}) \\ cov(\hat{\lambda},\hat{\alpha}) & var(\hat{\lambda}) \end{bmatrix}$$

$\approx \frac{I^{-1}(\theta)}{n} \approx \cdots$

⑤ Bin(n, p).

$$f(x) = \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x}$$

$$\therefore \log f(x) = \log\binom{n}{x} + x\log p + (n-x)\log(1-p)$$

$$\therefore \frac{d\log f(x)}{dp} = \frac{x}{p} - \frac{n-x}{1-p}$$

$$\frac{d^2\log f(x)}{dp^2} = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}$$

$$\therefore -E\left(-\frac{x}{p^2} - \frac{n-x}{(1-p)^2}\right) = \frac{E(X)}{p^2} + \frac{n-E(X)}{(1-p)^2}$$

$$= \frac{n}{p} + \frac{n\cancel{(1-p)}}{(1-p)^{\cancel{2}}}$$

$$= \frac{n}{p(1-p)}$$

$$\therefore \underline{I(p)_{Bin} = nI(p)_{Bernoulli}}$$

ST2132 Hypotheses testing.

1. $X_1 \cdots X_n$ IID $N(\mu, \sigma^2)$ RVs. σ known.

$H_0: \mu=\mu_0 \quad H_1: \mu=\mu_1>\mu_0$.

Compare $\bar{x}$ and $\mu_0 \Rightarrow$ if $\bar{x}>\mu_0 \to$ reject H_0

test statistic: $\underbrace{Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}}$ $\xleftarrow{\text{realization}}$ $z=\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$.

$\therefore H_0: \mu=\mu_0 \Rightarrow Z \sim (0,1)$. $z\uparrow$ $\bar{x}-\mu_0\uparrow \Rightarrow$ more likely to reject H_0

$H_1: \mu=\mu_1>\mu_0 \Rightarrow Z \sim N(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}, 1)$.

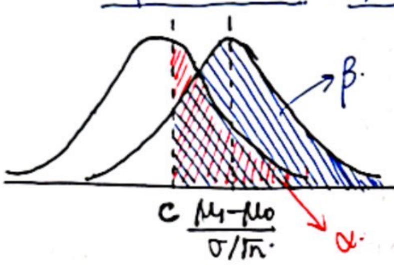
Let Rejection Region $[c, +\infty)$.

	$z<c$	$z\geq c$
H_0	Not reject H_0	Reject H_0 $\Rightarrow$ Type I error $\alpha=P_{H_0}(Z\geq c)$
H_1	Not reject H_0	Reject H_0

Type II error

under different Hypothese.

$1-\beta=P_{H_1}(Z<c) \Rightarrow \beta=P_{H_1}(Z\geq c)$. power of test



$c\uparrow \Rightarrow \beta\downarrow$ and $\alpha\downarrow$.

Impossible $\alpha=0, \beta=1$.

$\alpha=0 \Rightarrow c=\infty$

H_0 is never reject if it is false.

Under H_0: $\alpha=P_{H_0}(Z\geq c) \Leftrightarrow P_{H_0}(N(0,1)\geq c)$.

$\therefore c=z_\alpha$. $\therefore$ Rejection Region $[z_\alpha, +\infty)$.

Under H_1: $\beta=P_{H_1}(Z\geq c) \Rightarrow P(N(\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}, 1)\geq c)$.

$\therefore$ To adjust $P(\underbrace{Z-\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}}_{N(0,1)} \geq c-\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}})=\beta$.

$\therefore c-\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}=z_\beta \Rightarrow \underbrace{z_\alpha=\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}+z_\beta}$.

when $n\to\infty$ $P(N(0,1)\geq c-\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}}) \to P(N(0,1)\geq -\infty)=1$

2. $X_1 \cdots X_n \sim$ IID $N(\mu, \sigma^2)$.

$H_0: \mu=\mu_0 \quad H_1: \mu=\mu_1\neq\mu_0$

Rejection Region: $(-\infty, z_{\alpha/2}]$ and $[z_{\alpha/2}, +\infty)$.

Not reject H_0 when

$$-z_{\frac{\alpha}{2}} \leq \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}$$

$$\Rightarrow \bar{x}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow (1-\alpha)\text{CI}: \left(\bar{x}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right).$$

Fact: We can show a hypothesis is wrong but seldom can show it is true!

Sample Size: Find smallest n s.t. $\beta\geq 0.9$

By $z_\beta=\frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma}+z_\alpha$, we have.

$$n=\sigma^2\left(\frac{z_\alpha-z_\beta}{\mu_1-\mu_0}\right)^2.$$

Limitation of Reject Test: Highly depend on c

P value: under H_0, probability that test stats. (z) is equal or more extreme [i.e. more far from H_0]

P value $\downarrow \to$ Doubt $H_0\uparrow$.

z falls in Reject Region: $p<\alpha$.

z falls outside: $p>\alpha$.

$p=Pr_{H_0}(Z\geq z)$.

$P_{\text{two-tail}}=2P_{\text{one tail}}$

$P_{H_0}(|Z|\geq|z|)=2Pr_{H_0}(N(0,1)\geq|z|)$.

Approximate Test: consider σ unknown.

$X_1 \cdots X_n$ can be any distribution $n\to\infty$ $\bar{X}\to N$.

$Z=\frac{\bar{X}-\mu}{S/\sqrt{n}} \to$ Roughly RR: $\approx\alpha$. Roughly P-value.

Exact Test: consider σ unknown (T-test).

$X_1 \cdots X_n \sim N(\mu, \sigma^2)$, σ unknow.

$$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \longrightarrow T=\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

H_0: $\bar{X}\sim N(\mu_0, \frac{\sigma^2}{n})$.

Goodness-of-fit: likely-hood ratio test.

- L_0: $\boxed{\text{ML value}}$ for estimate model. (small) use estimates. $L(p)=\ell(p)_{max}$.. Estimate: $\dim(\Omega_0)$
- L_1: - - - - of general model (big) simply use the probability. (n is larger) $L(p)=\ell(p)_{max}$. General: $\dim(\Omega_1)$ (real).
- LR Statistic: $\underbrace{2\log(\frac{L_1}{L_0})}=2(\log L_1-\log L_0)$.
- $P\sim\chi^2_n$, n=difference of # para. between two model. $p(\chi^2_n>2\log\frac{L_1}{L_0})$
- $\dim(\Omega_0)$ nested $\dim(\Omega_0)$. [$\dim\Omega_0<\dim\Omega_1$] Argue by testing $H_0: p\in\Omega_0$. [estimate holds]

Multinomial Estimate. LR test.

let $(X_1, \dots X_k) \sim Multi(n, p)$

① $\Omega_0 = \{(p_1(\theta), \dots, p_k(\theta)) : \theta \in \Theta \subset R^h\}$.

② $\Omega_1 = \{(p_1, p_2, \dots p_k)\}$.

$\dim \Omega_0 = h < \dim \Omega_1 = k-1$

Object: test $H_0: p \in \Omega_0$.

$\Omega_1: l(p) = \sum_{i=1}^{k} X_i \log p_i$

General model: $\hat{p}_i = \frac{X_i}{n}$

$\therefore \log L_1 = l(\frac{X_i}{n}) = \sum_i^k X_i \log \frac{X_i}{n}$

$\Omega_0: l(\theta) = \sum_{i=1}^{k} X_i \cdot \log p_i(\theta)$

$\therefore \frac{\partial l}{\partial \theta} = 0 \to$ MLE: $\hat{\theta}$

$\therefore \log L_0 = l(\hat{\theta}) = \sum_i^k X_i \log p_i(\hat{\theta})$.

$\therefore G = 2(\log L_1 - \log L_0) = 2 \sum_{i=1}^{k} X_i \log \frac{X_i}{n p_i(\hat{\theta})}$

use data $(x_1, \dots x_n) \Rightarrow g$.

$P_{H_0}(G > g) \simeq P_r(\chi^2_{k-h-1} > g)$.

Test of independence. let $(X_{ij}; i \in [1, I], j \in [1, J]) \sim Multi(n, p)$

$\dim \Omega_1 = I \cdot J - 1$

Independency: $\exists$ positive number $\boxed{\sum_1^I q_i = \sum^J r_j = 1}$

s.t. $p_{ij} = q_i \cdot r_j$

$\Rightarrow \dim \Omega_0 = (I-1) + (J-1) = I + J - 2$.

Note that: $(IJ-1) - (I+J-2) = (I-1) \cdot (J-1)$.

$\therefore$ For H_0: $G \sim \chi^2_{(I-1)(J-1)}$.

let $\underbrace{X_{i+} = \sum_{j=1}^{J} X_{ij}}_{q}$ $\quad \underbrace{X_{+j} = \sum_{i=1}^{I} X_{ij}}_{r}$

$\therefore$ By definition:

$\Omega_1: l(p) = \sum X_{ij} p_{ij}$, $\hat{p}_{ij} = \frac{X_{ij}}{n}$

$\log L_1 = l(\hat{p}_{ij}) = \sum X_{ij} \log (\frac{X_{ij}}{n})$.

$\Omega_0: l(q, r) = \sum X_{ij} \log (p_{ij})$

$= \sum_{i,j} X_{ij} \cdot \log (q_i + r_j)$

$= \sum_{i,j} X_{ij} \log q_i + \sum_{i,j} X \log r_j$

$= \sum_i X_{i+} \log q_i + \sum_j X_{+j} \log r_j$

$\therefore \hat{q}_i = \frac{X_{i+}}{n} \quad \hat{r}_j = \frac{X_{+j}}{n}$

$\therefore \log L_0 = l(\hat{q}_i, \hat{r}_j)$

$= \sum_i X_{i+} \log \frac{X_{i+}}{n} + \sum X_{+j} \log \frac{X_{+j}}{n}$

$\therefore G = 2(\log L_1 - \log L_0) = 2 \sum X_{ij} \log \frac{X_{ij}}{X_{i+} X_{+j}/n}$

Pearson X^2 test statistic: $X^2 \sim \chi^2_{k-h-1}$ for large n.

Normal Data. (let $H_0: \mu = 0$).

σ	Ω_1	k_1	Ω_0	k_0
known	R	1	$\{0\}$	0
unknown	$R \times R_+$	2	$\{0\} \times R_+$	1

$\therefore$ For large n under H_0 $G \sim \chi^2_1$.

(i) $\Omega_1: l(\mu) = -\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{2\sigma^2}$ $\quad \hat{\mu} = \bar{X}$ (General)

$\therefore \log L_1 = -\frac{n \hat{\sigma}^2}{2\sigma^2}$. $\quad$ No MLE. $\quad$ 2-Moment

$\Omega_0: l(0) = -\frac{\sum_i X_i^2}{2\sigma^2} = \frac{-n \hat{\mu}_2}{2\sigma^2}$

$\therefore G = \frac{n \bar{X}^2}{\sigma^2} \sim \chi^2_1$.

(ii) unknown

$\Omega_1: l(\mu, \sigma) = -n \log \sigma - \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{2\sigma^2}$

$\therefore$ MLE: $\bar{X}, \hat{\sigma}$ (from Tut)

$\therefore \log L_1 = -\frac{n}{2} \log \hat{\sigma}^2 - \frac{n}{2}$

$\Omega_0: l(\sigma) = -n \log \sigma - \frac{\sum X_i^2}{2\sigma^2}$

$\frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum X_i^2}{\sigma^3} \Rightarrow \hat{\sigma} = \sqrt{\hat{\mu}_2}$

$\therefore \log L_0 = -\frac{n}{2} \log \hat{\mu}_2 - \frac{n}{2}$

$\therefore G = n \log (\frac{\hat{\mu}_2}{\hat{\sigma}^2}) \sim \chi^2_1$

Rejection Region: $[\chi^2_{k_1 - k_0}, \infty]$.