Example Tossing 2 coins

$$\Omega = \{HH, HT, TH, TT\}$$

$$\mathcal{F} = \text{power set of } \Omega$$

$$P(A) = |A|/4, \qquad A \subseteq \Omega \Leftrightarrow A \in \mathcal{F}$$

X: number of H,

$$X(\omega) = \begin{cases} 2, & \omega = HH \\ 1, & \omega = HT \text{ or } TH \\ 0, & \omega = TT. \end{cases}$$

1.5

$X =$ result of coin flip

$Y =$ parity of dice

$X_1, X_2, \cdots, X_n$ independent, $Y = \max_{1 \le i \le n} X_i$, 1.7

$$P(Y \le x) = P(\max_i X_i \le x)$$
$$= P(X_1 \le x, \cdots, X_n \le x)$$
$$= \prod_{i=1}^{n} P(X_i \le x).$$

Recall $X_1 \sim \text{Poisson}(\lambda)$ if PDF of X_1 is $(\lambda > 0)$ 1.

$$p(x;\lambda) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \cdots$$

Suppose $X_1, \cdots, X_n$ independent, $X_i \sim \text{Poisson}(\lambda_i)$,

$$Y = \sum_{i=1}^{n} X_i$$

MGF of $X \sim \text{Poisson}(\lambda)$ is

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} e^{-\lambda} \cdot \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} \cdot e^{e^t\lambda} = \underline{e^{\lambda(e^t-1)}}$$

MGF of Y: $M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^{n} X_i})$

$$= E\left(\prod_{i=1}^{n} e^{tX_i}\right) = \prod_{i=1}^{n} E(e^{tX_i}) = e^{\left(\sum_{i=1}^{n}\lambda_i\right)(e^t-1)}$$

This shows $Y \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right)$

$(1-p)^{x-1}p.$ Geometric 1.18

$X_1, X_2, \cdots, X_n$ indep. $\quad X_i \sim \text{Gamma}(\alpha_i, \lambda)$ 1.22

Then $Y = \sum_{i=1}^{n} X_i \sim \text{Gamma}\left(\sum_{i=1}^{n} \alpha_i, \lambda\right).$

MGF of $X \sim \text{Gamma}(\alpha, \lambda)$

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \cdot \lambda^\alpha x^{\alpha-1} \cdot e^{-\lambda x} / \Gamma(\alpha) \, dx$$

$$= \lambda^\alpha / (\lambda - t)^\alpha, \qquad t < \lambda$$

$$= \left(1 - \frac{t}{\lambda}\right)^{-\alpha}.$$

$$M_Y(t) = E(e^{tY}) = \prod_{i=1}^{n} M_{X_i}(t) = \left(1 - \frac{t}{\lambda}\right)^{-\sum_{i=1}^{n} \alpha_i}$$

1.26

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)\, dt$$

1.27

$$X=\sum_{i=1}^{k} Z_i^2, \qquad Z_1,\cdots,Z_k \overset{IID}{\sim} N(0,1)$$

$$\frac{1}{2}-t=\frac{1}{2\sigma^2}$$

$$\sigma^2=\frac{1}{1-2t}$$

$$\begin{aligned}
E(e^{tX}) &= \left\{E(e^{tZ_1^2})\right\}^k \\
&= \left\{\int_{-\infty}^{\infty} e^{tz^2}\cdot\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz\right\}^k \\
&= \left\{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-z^2\cdot(\frac{1}{2}-t)}\,dz\right\}^k \\
&= \left\{\frac{1}{\sqrt{2\pi}}\Big/\frac{1}{\sqrt{2\pi\sigma^2}}\right\}^k \\
&= \sigma^k \\
&= (1-2t)^{-k}
\end{aligned}$$

$$\chi_k^2 = \mathrm{Gamma}\left(\frac{k}{2},\frac{1}{2}\right).$$

2.2

$X_1, \cdots, X_n \overset{IID}{\sim} N(\underset{\sim}{\mu}, 1)$.

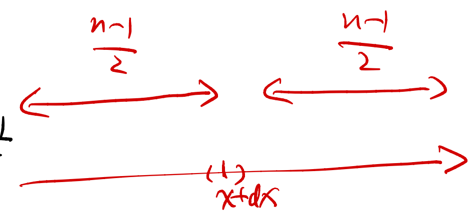
$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is unbiased because

$$E_\mu(\hat{\mu}) = E_\mu\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = E_\mu(X_1) = \mu.$$

$X_{(\frac{n+1}{2})}$ $\qquad X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ order statistics.

PDF: $p(X_{(\frac{n+1}{2})} - \mu = x)$

$$= \binom{n}{(n-1)/2} (\Phi(x))^{\frac{n-1}{2}} \phi(x) (1-\Phi(x))^{\frac{n-1}{2}}$$



is symmetric around 0. $\Rightarrow E(X_{(\frac{n+1}{2})} - \mu) = 0$

2.3

$$MSE(\hat\theta) = E_\theta(\hat\theta - \theta)^2$$

deterministic function of θ

$$= E_\theta\{(\hat\theta - E_\theta\hat\theta) + (\underline{E_\theta\hat\theta - \theta})\}^2$$

$$= \underbrace{E_\theta\{(\hat\theta - E_\theta\hat\theta)^2\}}_{Var(\hat\theta)} + \underbrace{(E_\theta\hat\theta - \theta)^2}_{bias(\hat\theta)^2} + \underbrace{2E_\theta\{(\hat\theta - E_\theta\hat\theta)(E_\theta\hat\theta - \theta)\}}$$

$$= 2(E_\theta\hat\theta - \theta)\cdot \underline{E_\theta(\hat\theta - E_\theta\hat\theta)}$$

$$= 0$$

2.7

Suppose $\hat{\theta} = T(X)$ is unbiased. Then

$$e^{-2\lambda} = E_\lambda(\hat{\theta}) = \sum_{x=0}^{\infty} T(x) \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} \qquad \text{for all } \lambda > 0.$$

By using Maclaurin series,

$$\sum_{x=0}^{\infty} T(x) \cdot \frac{\lambda^x}{x!} = e^{-\lambda} = \sum_{x=0}^{\infty} (-1)^x \cdot \frac{\lambda^x}{x!}.$$

This shows $T(x) = (-1)^x$, $x = 0, 1, 2, \cdots$.

Proof of factorization criterion. (in discrete case) 3.2

$\Leftarrow$ Suppose $p(x;\theta) = g(T(x), \theta) \cdot h(x)$.

$$P_\theta(X=x \mid T=t) = \frac{P_\theta(X=x, T=t)}{P_\theta(T=t)}$$

$$= \frac{P_\theta(X=x) \cdot 1\{T(x)=t\}}{\sum_x P_\theta(X=x, T=t)}$$

$$= \frac{\cancel{g(t,\theta)} \cdot h(x) \cdot 1\{T(x)=t\}}{\sum_x \cancel{g(t,\theta)} \cdot h(x) \cdot 1\{T(x)=t\}}$$

does not depend on θ. So $T(X)$ is sufficient.

$\Rightarrow$ Suppose T is sufficient.

w.p.1 given $X=x$

$$P_\theta(X=x) = P_\theta(X=x,\ \underline{T(X)=T(x)})$$

$$= \underbrace{P_\theta(X=x \mid T(X)=T(x))}_{h(x)} \cdot \underbrace{P_\theta(T(X)=T(x))}_{g(T(x),\theta)}$$

doesn't depend on θ by sufficiency

Bernoulli example:

$$p(x;\theta) = \underbrace{(1-\theta)^n \left(\frac{\theta}{1-\theta}\right)^{\Sigma x_i}}_{g(T(x),\theta)} \cdot \underbrace{1}_{h(x)}$$

So $T(X) = \sum_{i=1}^{n} X_i$ is sufficient.

3.2

Ex $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \text{Unif}[0, \theta]$, $\quad X = (X_1, \cdots, X_n)$

$$p(x;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \cdot \mathbb{1}\{0 \le x_i \le \theta\}$$

$$= \frac{1}{\theta^n} \cdot \mathbb{1}\{0 \le \min_i x_i \le \max_i x_i \le \theta\} \cdot \mathbb{1}$$

$$= \underbrace{\frac{1}{\theta^n} \cdot \mathbb{1}\{\max_i x_i \le \theta\}}_{g(T(x),\, \theta)} \cdot \underbrace{\mathbb{1}\{0 \le \min_i x_i \le \max_i x_i\}}_{h(x)}$$

So $T(X) = \max_i X_i$ is a sufficient statistic.

Proof for minimal sufficiency:

Sufficiency Let z_t representative element from $\{x: T(x)=t\}$, then

$$p(x;\theta) = \underbrace{p(z_{T(x)};\theta)}_{g(T(x),\theta)} \cdot \underbrace{\frac{p(x;\theta)}{p(z_{T(x)};\theta)}}_{h(x)}$$

b.c. $T(x)=T(z)$

$\Rightarrow \frac{p(x;\theta)}{p(y;\theta)}$ doesn't depend on θ.

so $T(x)$ is sufficient.

Minimality Let $S(x)$ be another suff. stat. so

$$p(x;\theta) = \tilde{g}(S(x),\theta) \cdot \tilde{h}(x).$$

Suppose $S(x)=S(y)$. Then $\frac{p(x;\theta)}{p(y;\theta)} = \frac{\tilde{h}(x)}{\tilde{h}(y)} \iff T(x)=T(y)$.

So $T(x)$ is minimal suff.

Ex $X_1, \cdots, X_n \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$

$$\frac{p(x;\theta)}{p(y;\theta)} = \frac{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\}}{\cancel{(2\pi\sigma^2)^{-n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i-\mu)^2\}}$$

$$= \exp\{-\frac{1}{2\sigma^2}(\Sigma x_i^2 - \Sigma y_i^2) - \frac{\mu}{\sigma^2}\cdot(\Sigma x_i - \Sigma y_i)\}$$

So $T(x) = (\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i)$ is minimal sufficient.

Recall minimal sufficient stat. is unique up to bijection.

$$\tilde{T}(x) = (\bar{X}, s^2) = \left(\frac{1}{n}\sum_{i=1}^n x_i, \frac{1}{n-1}\sum_{i=1}^n (x_i-\bar{x})^2\right)$$

$$= \left(\frac{1}{n}\Sigma x_i, \frac{1}{n-1}(\sum_i x_i^2 - n\bar{x}^2)\right)$$ is also min. suff.

3.5

$$\hat{\theta} = \int \tilde{\theta}(x) \underbrace{P_\theta(x|t)}_{\text{doesn't depend on } \theta \text{ by suff.}} dx \quad \text{is a function of } t.$$

Proof of Rao-Blackwell

$$E(\hat{\theta}) = E\{E(\tilde{\theta}|T)\} \overset{\text{law of total expectation}}{=} E(\tilde{\theta}). \quad \text{So } \operatorname{bias}(\hat{\theta}) = \operatorname{bias}(\tilde{\theta}).$$

$$\operatorname{Var}(\tilde{\theta}) \overset{\text{law of total variance}}{=} \underbrace{E\{\operatorname{Var}(\tilde{\theta}|T)\}}_{\geq 0} + \underbrace{\operatorname{Var}(E(\tilde{\theta}|T))}_{= \operatorname{Var}(\hat{\theta})}$$

So $\operatorname{Var}(\tilde{\theta}) \geq \operatorname{Var}(\hat{\theta})$, and $\operatorname{MSE}(\tilde{\theta}) \geq \operatorname{MSE}(\hat{\theta})$.

Equality iff $\operatorname{Var}(\tilde{\theta}|T) = 0$ w.p. 1 $\Leftrightarrow$ $\tilde{\theta}$ is func. of T.

$$\tilde{\theta} = 1\{X_1 = 0\} \qquad E_\theta(\tilde{\theta}) = P_\theta(X_1 = 0) = e^{-\lambda}. \qquad T = \sum_{i=1}^{n} X_i$$

3.6

$$\hat{\theta}(t) = E(\tilde{\theta} \mid T = t) = P(X_1 = 0 \mid \sum_{i=1}^{n} X_i = t)$$

$$= P(X_1 = 0, \sum_{i=2}^{n} X_i = t \mid \sum_{i=1}^{n} X_i = t)$$

$$= \frac{P(X_1 = 0) P(\sum_{i=2}^{n} X_i = t)}{P(\sum_{i=1}^{n} X_i = t)}$$

$$= \frac{\cancel{e^{-\lambda} \cdot e^{-(n-1)\lambda}} \cdot \frac{\{(n-1)\cancel{\lambda}\}^t}{\cancel{t!}}}{\cancel{e^{-n\lambda}} \cdot \frac{(n\cancel{\lambda})^t}{\cancel{t!}}}$$

$$= \left(\frac{n-1}{n}\right)^t.$$

So $\hat{\theta} = (1 - \frac{1}{n})^{\sum X_i}$. Sanity check: $\hat{\theta} = \{\underbrace{(1 - \frac{1}{n})^n}\}^{\bar{X}} \longrightarrow e^{-\lambda} = \theta$.

$X_1, \cdots, X_n \overset{\text{IID}}{\sim} \text{Unif}[0, \theta]$. $\quad \theta > 0$ unknown.

3.6

$T = \max_{1 \le i \le n} X_i$ is sufficient.

$X_1 \sim \text{Unif}[0, \theta]$. $\quad E_\theta(X_1) = \frac{\theta}{2}$. $\quad$ So $\tilde{\theta} = 2X_1$ is unbiased.

$$\hat{\theta}(t) = E(\tilde{\theta} \mid T = t) = 2E(X_1 \mid \max_{1 \le i \le n} X_i = t)$$

$$= 2\left\{ \frac{1}{n} E(X_1 \mid X_1 = \max_{1 \le i \le n} X_i = t) \right.$$

$$\left. + \frac{n-1}{n} E(X_1 \mid X_1 < t, \max_{2 \le i \le n} X_i = t) \right\}$$

$$= 2\left\{ \frac{t}{n} + \frac{n-1}{n} E(X_1 \mid X_1 < t) \right\}$$

$$= 2\left\{ \frac{t}{n} + \frac{n-1}{n} \cdot \frac{t}{2} \right\}$$

$$= \frac{n+1}{n} \cdot t$$

So $\hat{\theta} = \frac{n+1}{n} \cdot \max_{1 \le i \le n} X_i$ is unbiased.

1. $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \text{Bernoulli}(\theta)$.

4.2

$$L(\theta) = p(X;\theta) = \theta^{\sum_{i=1}^n X_i} \cdot (1-\theta)^{n-\sum_{i=1}^n X_i}$$

$$\ell(\theta) = \log L(\theta) = \left(\sum_{i=1}^n X_i\right) \cdot \log\theta + \left(n - \sum_{i=1}^n X_i\right) \cdot \log(1-\theta)$$

$$\ell'(\theta) = \frac{1}{\theta} \cdot \sum_{i=1}^n X_i - \frac{1}{1-\theta}\left(n - \sum_{i=1}^n X_i\right).$$

$\hat{\theta}$ solves $\ell'(\theta) = 0$, so $(1-\hat{\theta}) \cdot \sum_{i=1}^n X_i = \hat{\theta} \cdot \left(n - \sum_{i=1}^n X_i\right)$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}. \quad \text{unbiased.}$$

4.2

2. $X_1, \cdots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, $\quad \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

$$L(\theta) = p(X;\theta) = (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2\right\}.$$

$$\ell(\theta) = \log L(\theta) = -\frac{n}{2}\cdot\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\cdot\sum_{i=1}^n (X_i-\mu)^2.$$

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\cdot\sum_{i=1}^n (X_i-\mu). \quad \Longrightarrow \quad \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}.$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{\cancel{2}}\cdot\frac{1}{\cancel{\sigma^2}} + \frac{1}{\cancel{2}\sigma^{\cancel{4}\,2}}\cdot\sum_{i=1}^n (X_i-\mu)^2. \quad \Longrightarrow \quad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^n (X_i-\bar{X})^2$$

$$= S_{xx}/n$$

$\hat{\mu}_{MLE}$ is unbiased (for μ).

Later will show $\quad \frac{S_{xx}}{\sigma^2} \sim \chi^2_{n-1}. \Rightarrow E(\hat{\sigma}^2_{MLE}) = \frac{n-1}{n}\cdot\sigma^2.$ biased.

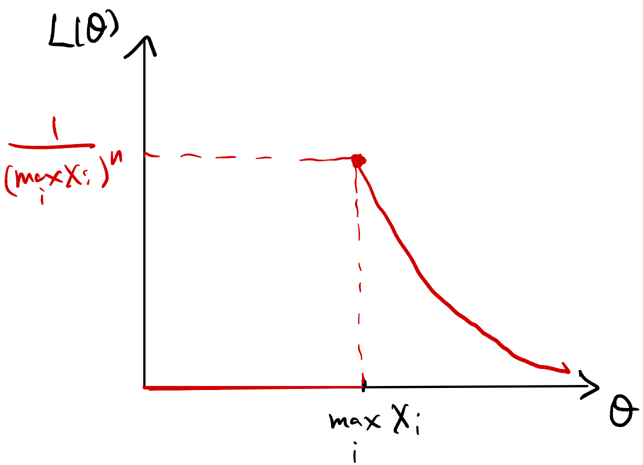
4.2

3. $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \text{Unif}[0, \theta], \quad \theta > 0$

$$L(\theta) = p(X;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \cdot I_{\{0 \le X_i \le \theta\}}$$

$$= \frac{1}{\theta^n} \cdot 1\{0 \le \min_i X_i \le \max_i X_i \le \theta\}.$$

$$= \frac{1}{\theta^n} \cdot 1\{\max_i X_i \le \theta\} \quad \text{(with probability 1).}$$



So $\hat{\theta}_{MLE} = \max_i X_i$.

$$E(\hat{\theta}_{MLE}) = \frac{n}{n+1} \cdot \theta.$$

So $\hat{\theta}_{MLE}$ biased in finite n, but asymptotically unbiased.

$$L(\theta) = P(X;\theta) \underset{\substack{T \\ \text{factorization}}}{=} g(T(X),\theta)\, h(X).$$

4.3

$$L(k) = P(\text{new colour 1st}) \cdot P(\text{new colour 2nd})\, P(\text{3rd} = \text{1st})\, P(\text{new 4th})$$

$$= 1 \cdot \frac{k-1}{k} \cdot \frac{1}{k} \cdot \frac{k-2}{k}$$

$$= \frac{(k-1)(k-2)}{k^3}.$$

4.4

$$E(N_j) = E\left(\sum_{i=1}^{S} 1\{X_i(1)=j\}\right) = \sum_{i=1}^{S} P(X_i(1)=j) = \sum_{i=1}^{S} e^{-\lambda_i} \cdot \frac{\lambda_i^j}{j!}.$$

4.5

$$E(M(t)) = E\Big(\sum_{i=1}^{S} I\{\text{doesn't appear in first } n. \text{ appear} \geq 1 \text{ in next } nt\}\Big)$$

$$= E\left(\sum_{i=1}^{S} I\{X_i(1)=0,\ X_i(\cancel{n}(1+t)) \geq 1\}\right) = E\Big(\sum 1_{\{\underbrace{X_i(1)=0}\ \underbrace{X_i(1:1+t)}\geq 1\}}\Big)$$

$$= \sum_{i=1}^{S} e^{-\lambda_i \cancel{t}} \cdot (1-e^{-\lambda_i t}).$$

$$= \sum_{i=1}^{S} e^{-\lambda_i \cancel{t}} \cdot \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j!} \cdot (\lambda_i t)^j.$$

$$= \sum_{j=1}^{\infty} (-1)^{j-1} \cdot t^j \cdot \boxed{\sum_{i=1}^{S} e^{-\lambda_i} \cdot \frac{\lambda_i^j}{j!}} = E(N_j)$$

$$\hat{M}(1) = N_1 - N_2 + N_3 - \cdots$$

So an unbiased estimator of $E(M(t))$ is $\sum_{j=1}^{\infty} (-1)^{j-1} \cdot t^j \cdot N_j$.

Suppose $X \sim P(x; \theta, \eta)$, $(\theta, \eta) \in \Theta \times H$. 5.1

θ is called primary parameter, η is nuissance parameter.

Def A set $S(X) \subseteq \Theta$ is a $(1-\alpha)$- confidence set / region of θ if

$$P_{\theta,\eta}(\theta \in S(X)) = 1-\alpha \quad \text{for all } (\theta,\eta) \in \Theta \times H.$$

Rem $=$ is often difficult. Some authors just require $\geq$.

5.2

$X_1, \cdots, X_n \overset{\text{IID}}{\sim} N(\theta, 1)$.

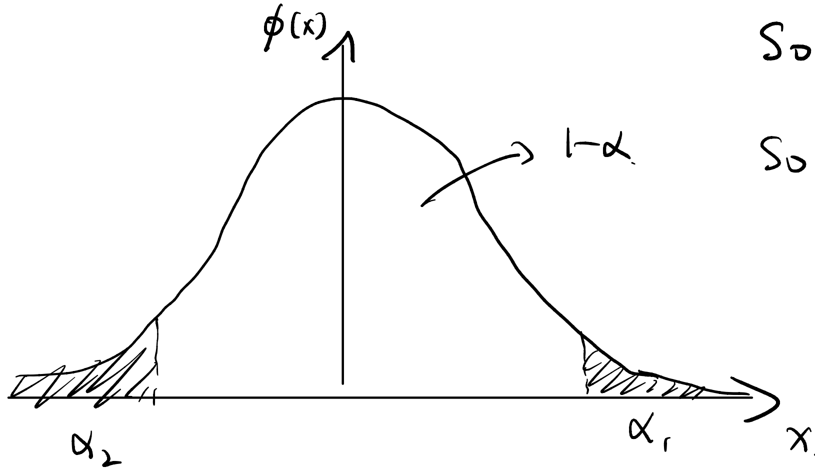
1. Find pivot

We know $\bar{X} \sim N(\theta, \frac{1}{n})$ or equivalently $\sqrt{n}(\bar{X}-\theta) \sim N(0,1)$.

2. Use quantiles of pivot distribution.

$$P_\theta\left(-z_{\alpha_2} \le \sqrt{n}(\bar{X}-\theta) \le z_{\alpha_1}\right) = 1-\alpha.$$

for all $0 \le \alpha_1, \alpha_2 \le 0.5$, $\alpha_1 + \alpha_2 = \alpha$.



3. Rearrange terms.

So $P_\theta\left(\bar{X} - \frac{z_{\alpha_1}}{\sqrt{n}} \le \theta \le \bar{X} + \frac{z_{\alpha_2}}{\sqrt{n}}\right) = 1-\alpha$

So a $(1-\alpha)$-CI of θ is

$$\left[\bar{X} - \frac{z_{\alpha_1}}{\sqrt{n}},\ \bar{X} + \frac{z_{\alpha_2}}{\sqrt{n}}\right].$$

$\bar{X} \pm 1.96\ SE$

SE = standard error $\left(\frac{1}{\sqrt{n}}\right)$

$h : \mathbb{R} \to \mathbb{R}$ is monotone $\uparrow$ $\quad (\Theta \subseteq \mathbb{R})$. $\qquad$ 5.3

$$P_\theta(h(L) \le h(\theta) \le h(U))$$

$$\ge P_\theta(L \le \theta \le U)$$

$$= 1-\alpha, \quad \text{for all } \theta \in \Theta.$$

$X_1, \dots, X_n \overset{IID}{\sim} \text{Bernoulli}(p)$. $\quad \hat{p} = \bar{X}$. $\quad$ 5.5

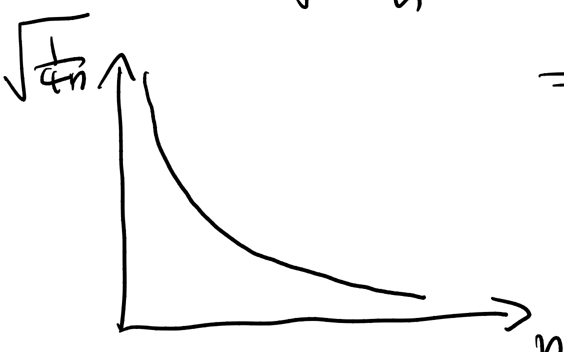
$$\sqrt{n}(\bar{X} - p) \xrightarrow{d} N(0, p(1-p)).$$

1. Wald / symmetric CI based on $\sqrt{n}\dfrac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{d} N(0,1)$.

$$\hat{p} \pm z_{\alpha/2} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \sqrt{\frac{1}{4n}} = 0.02$$

$$\Rightarrow \quad n = 25^2 = 625.$$



5.5

2. $$\sqrt{n} \cdot \frac{\hat{p} - p}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

$$P\left(-z_{\alpha/2} \le \sqrt{n} \cdot \frac{\hat{p} - p}{\sqrt{p(1-p)}} \le z_{\alpha/2}\right) \longrightarrow \alpha \quad \text{when } n \to \infty$$

$\Longrightarrow$ closed-form CI for p.

3. Exact: $$P\left(Q_{\frac{\alpha}{2}}(n,p) \le n\bar{X} \le Q_{1-\frac{\alpha}{2}}(n,p)\right) \ge 1-\alpha$$

Plot: $$P_p(p \in [L, U]) = \sum_{x=L}^{U} \binom{n}{p} p^x (1-p)^{1-x}$$

$$\pi(\theta|X) \propto \pi(\theta)\, \underline{p(X|\theta)} \underset{\text{factorization thm.}}{=} \pi(\theta) \cdot g(T(X), \theta) \cdot \underline{h(X)}$$ 6.2

$$\pi(\theta=1 \mid X=1) \propto \pi(\theta=1) \cdot \underline{p(X=1|\theta=1)} = 2\% \times 98\%$$ 6.3

$$\pi(\theta=0 \mid X=1) \propto \pi(\theta=0) \cdot p(X=1|\theta=0) = 98\% \times 1\%$$

So $\pi(\theta=1 \mid X=1) = \frac{2}{3}$.

6.4

$\theta \sim \text{Beta}(\alpha, \beta)$ $\quad \pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}, \quad \theta \in [0,1].$

$X|\theta \sim \text{Binom}(n, \theta)$ $\quad L(\theta) = p(X|\theta) = \binom{n}{X} \cdot \theta^X \cdot (1-\theta)^{n-X}.$

So $\quad \pi(\theta|X) \propto \pi(\theta)\, p(X|\theta) \propto \theta^{\alpha+X-1} \cdot (1-\theta)^{\beta+n-X-1}.$

So $\quad \theta | X \sim \text{Beta}(\alpha + X, \quad \beta + n - X)$

If $\alpha = \beta = 1$ (so $\theta \sim \text{Unif}[0,1]$ a priori), then

$$\theta | X \sim \text{Beta}(X+1, n-X+1)$$

$$E(\theta | X) = \frac{X+1}{n+2}.$$

6.6

$A > B.$

$$E(A) = \frac{1}{3} \times 1000$$

$$E(B) = P(\text{blue}) \times 1000$$

$$P(\text{blue}) < \frac{1}{3}$$

$C > D$

$$E(C) = \frac{2}{3} \times 1000$$

$$E(D) = (1 - P(\text{blue})) \times 1000$$

$$P(\text{blue}) > \frac{1}{3}.$$

6.7

1. $$R(\tilde{\theta}) = \int (\tilde{\theta}-\theta)^2 \pi(\theta|x)\, d\theta$$

$$R'(\tilde{\theta}) = \int 2(\tilde{\theta}-\theta)\, \pi(\theta|x)\, d\theta$$

So $\hat{\theta}$ satisfies $\int \hat{\theta}(x) \pi(\theta|x)\, d\theta = \int \theta\, \pi(\theta|x)\, d\theta$.

$$\hat{\theta}(x) = E_\pi(\theta|x).$$

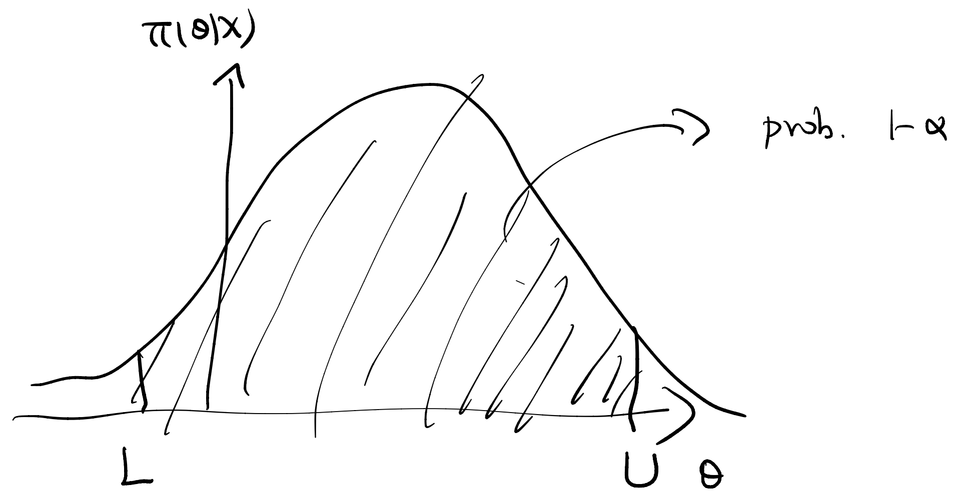
2. $$R(\tilde{\theta}) = \int |\tilde{\theta}-\theta|\, \pi(\theta|x)\, d\theta$$

$$= \int_{-\infty}^{\tilde{\theta}} (\tilde{\theta}-\theta)\cdot \pi(\theta|x)\, d\theta + \int_{\tilde{\theta}}^{\infty} (\theta-\tilde{\theta})\, \pi(\theta|x)\, d\theta.$$

$$R'(\tilde{\theta}) = \int_{-\infty}^{\tilde{\theta}} \pi(\theta|x)\, d\theta + \int_{\tilde{\theta}}^{\infty} -1\cdot \pi(\theta|x)\, d\theta.$$

So $\hat{\theta}$ solves $\int_{-\infty}^{\hat{\theta}} \pi(\theta|x)\, d\theta = \int_{\hat{\theta}}^{\infty} \pi(\theta|x)\, d\theta$. $P_\pi(\theta<\hat{\theta}|x) = P_\pi(\theta>\hat{\theta}|x)$.

6.8



$$\pi(\mu) \propto e^{-\frac{n_0}{2}(\mu-\mu_0)^2}$$

6.9

$$p(X|\mu) \propto e^{-\frac{n}{2}(\bar{x}-\mu)^2}$$ [$\bar{X}$ is suff. $\bar{x} \sim N(\mu, \frac{1}{n})$]

$$\pi(\mu|X) \propto \pi(\mu) \cdot p(X|\mu) = e^{-\frac{n_0}{2}(\mu-\mu_0)^2 - \frac{n}{2}(\bar{x}-\mu)^2}$$

$$\propto e^{-\frac{n_0+n}{2}\mu^2 + (n_0\mu_0 + n\bar{x})\mu}$$

$$\propto e^{-\frac{n_0+n}{2}\left(\mu - \frac{n_0\mu_0+n\bar{x}}{n_0+n}\right)^2}$$

$$\Rightarrow \quad \mu|X \sim N\left(\frac{n_0\mu_0+n\bar{x}}{n_0+n}, \frac{1}{n_0+n}\right).$$

$$\bar{x} \sim N(\mu, \tfrac{1}{n}).$$

7.1

1. $X \sim \text{Bin}(n, p)$. $H_0: p = \frac{1}{2}$ (unbiased) $H_1: p \neq \frac{1}{2}$ (biased)

2. $X \sim \text{Multi}(n, (p_1, p_2, p_3, p_4))$, $H_0: p_1 = \frac{9}{16}$, $p_2 = p_3 = \frac{3}{16}$, $p_4 = \frac{1}{16}$.

3. $X_1, \cdots, X_n \overset{\text{IID}}{\sim} P$, $Y_1, \cdots, Y_m \overset{\text{IID}}{\sim} P'$.

$H_0: P = P'$ $H_1: P \neq P'$.

Type I

$$L_1(T, \theta) = \begin{cases} 1, & \text{if } \theta \in \Theta_0 \text{ and } T=1 \\ 0, & \text{otherwise} \end{cases}$$

7.3

Type II

$$L_2(T, \theta) = \begin{cases} 1, & \text{if } \theta \in \Theta_1 \text{ and } T=0 \\ 0, & \text{otherwise.} \end{cases}$$

Type I error rate $= E_\theta(L_1(T,\theta))$

II $= E_\theta(L_2(T,\theta))$.

Proof of N-P

7.4

Let $R^* = \{x : T^*(x) = 1\}$ be the rejection region of LRT.

$$\beta(T^*;\theta) = P_\theta(T^* = 1) = \int_{R^*} p(x;\theta)\,dx.$$

Optimization: maximize $\int_R p(x;\theta_1)\,dx$ s.t. $\int_R p(x;\theta_0) \le \alpha$.

Take any test $T(X) \in \{0,1\}$ with size $\le \alpha$. Let $R = \{x : T(x) = 1\}$.

$$\beta(T^*;\theta_1) - \beta(T;\theta_1) = \int_{R^*} p(x;\theta_1)\,dx - \int_R p(x;\theta_1)\,dx$$

$$= \int_{R^*\setminus R} \underbrace{\frac{p(x;\theta_1)}{p(x;\theta_0)}}_{\Lambda(x) > c} \cdot p(x;\theta_0)\,dx - \int_{R\setminus R^*} \underbrace{\frac{p(x;\theta_1)}{p(x;\theta_0)}}_{\Lambda(x) \le c} \cdot p(x;\theta_0)\,dx$$

$$\ge c \cdot \left\{ \int_{R^*\setminus R} p(x;\theta_0)\,dx - \int_{R\setminus R^*} p(x;\theta_0)\,dx \right\}$$

$$= c \cdot \left\{ \int_{R^*} - \int_R \right\} p(x;\theta_0)\,dx = c \cdot (\alpha - \text{size of } T) \ge 0$$

$X_1, \cdots, X_n \overset{IID}{\sim} N(\mu, \sigma_0^2)$, σ_0^2 is known. $\mu_1 > \mu_0$ given

$H_0: \mu = \mu_0$ vs. $H_1: \mu = \mu_1$.

7.7

$$\Lambda(X) = \frac{L(\mu_1)}{L(\mu_0)} = \frac{p(X;\mu_1)}{p(X;\mu_0)} = \frac{\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i-\mu_1)^2\right\}}{\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i-\mu_0)^2\right\}}$$

$$= \exp\left\{\frac{(\mu_1-\mu_0)\cdot n\overline{X}}{\sigma_0^2} + \frac{n(\mu_0^2-\mu_1^2)}{2\sigma_0^2}\right\}$$

So $\Lambda(X)$ is a monotone increasing function in $\overline{X}$.

So $T = \mathbb{1}\{\Lambda(x) > c\} = \mathbb{1}\{\overline{X} > b\}$

Size of $T = P_{\mu_0}(T=1) = P_{\mu_0}(\overline{X} > b)$

$$= P_{\mu_0}\Bigg(\underbrace{\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma_0}}_{\sim N(0,1) \text{ under } \mu_0} > \frac{\sqrt{n}(b-\mu_0)}{\sigma_0}\Bigg) = 1 - \Phi\left(\frac{\sqrt{n}(b-\mu_0)}{\sigma_0}\right) = \alpha$$

8.1

$$\beta(\mu) = P_\mu(T(X)=1) = P_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma_0} > z_\alpha\right)$$

$$= P_\mu\Bigg(\underbrace{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma_0}}_{\sim N(0,1)} > z_\alpha + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\Bigg)$$

$$= 1 - \Phi\left(z_\alpha + \frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\right).$$

At $\mu=\mu_0$. $\beta(\mu_0) = 1 - \Phi(z_\alpha) = \alpha$.

UMP $\quad$ z-test rejects $\mu=\mu_0$ if $z=\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma_0}>z_\alpha$. $\quad T=1\{z>z_\alpha\}$.

8.2

① Note that this test doesn't depend on μ_1.

By the N-P Lemma, the same test is most powerful against any $\mu_1>\mu_0$.

(*)

Meaning $\quad \beta(\tilde{T};\mu_1)\le\beta(T;\mu_1) \quad$ if $\quad \beta(\tilde{T};\mu_0)\le\beta(T;\mu_0)$.

② Power function of the z-test $\beta(T;\mu)=1-\Phi\left(z_\alpha+\frac{\sqrt{n}(\mu-\mu_0)}{\sigma_0}\right)$.

is increasing in μ. So $\sup_{\mu\le\mu_0}\beta(T;\mu)=\beta(T;\mu_0)=\alpha$.

③ For any test $\tilde{T}$ with size $\le\alpha$, $\quad \beta(\tilde{T};\mu_0)\le\alpha$.

Then apply (*).

A model has <u>monotone likelihood ratio</u> if there exists $S: \mathcal{X} \to \mathbb{R}$ s.t. 8.2

$$\frac{p(x;\theta_1)}{p(x;\theta_0)} \text{ is an increasing func. of } S(x), \quad \forall\, \theta_0 < \theta_1.$$

One example: (one-parameter) exponentialy family

$$p(x;\theta) = e^{h(\theta)S(x) - k(\theta)} \cdot p(x;0), \qquad \theta \in \Theta \subseteq \mathbb{R}.$$

h is increasing in θ. $S(X)$ is a sufficient statistic.

$X_1, \cdots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma_0^2)$, σ_0^2 is known. **8.6**

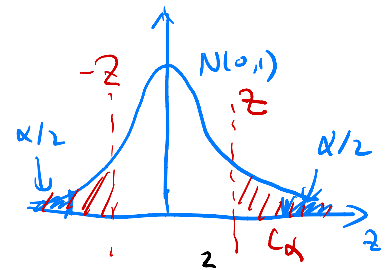
$H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$ μ_0 given.

$$\Lambda(X) = \frac{\sup_{\mu \in \Theta_0 \cup \Theta_1} p(X;\mu)}{\sup_{\mu \in \Theta_0} p(X;\mu)} = \frac{\cancel{(2\pi\sigma_0^2)^{-n/2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i - \bar{X})^2\right\}}{\cancel{(2\pi\sigma_0^2)^{-n/2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i - \mu_0)^2\right\}}$$

So $$2\log\Lambda(X) = \frac{1}{\sigma_0^2}\left\{\sum_i (X_i - \mu_0)^2 - \sum_i (X_i - \bar{X})^2\right\} \underset{\substack{\uparrow \\ \text{algebra}}}{=} \frac{n}{\sigma_0^2}(\bar{X} - \mu_0)^2 = Z^2$$

So LRT rejects H_0 if

$$2\log\Lambda(X) = Z^2 > c_\alpha.$$



where

$$P_{\mu_0}(Z^2 > c_\alpha) = P_{\mu_0}\Big(\underbrace{\frac{n}{\sigma_0^2}(\bar{X} - \mu_0)^2}_{} > c_\alpha\Big) = \alpha.$$ So $c_\alpha = \chi_1^2(\alpha)$.

Under H_0 $\bar{X} \sim N\left(\mu_0, \frac{\sigma_0^2}{n}\right)$. $Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma_0} \sim N(0,1)$

8.7

1. $\Theta_0 = \{\theta_0\}$, $df = 0$.

2. $\Theta_0 = \{\theta \in \mathbb{R}^p : \theta_1 = \cdots = \theta_{p_0} = 0\}$. $df = p - p_0$

3. $A \in \mathbb{R}^{p_0 \times p}$, $b \in \mathbb{R}^{p_0}$. $\Theta_0 = \{\theta \in \mathbb{R}^p : A\theta = b\}$ affine subspace

$df = p - p_0$.

4. $\Theta_0 = \{\theta \in \mathbb{R}^p : \theta_i = f_i(\phi),\ i = 1, \cdots, p \text{ for } \phi \in \mathbb{R}^{p-p_0}\}$.

Under regularity conditions, $df = p - p_0$.

9.1

$T(X) = \mathbb{1}\{\Lambda(X) > c_\alpha\}$, c_α is chosen s.t. $P(\underset{\sim}{\Lambda}(X) \geq c_\alpha) = \alpha$

assume continuous.

So $c_\alpha = \inf\{c : \sup_{\theta\in\Theta_0} P_\theta(\Lambda(X) > c) \leq \alpha\}$

This tells me

$$T(x) = 1 \iff \Lambda(x) > c_\alpha \iff \underbrace{\sup_{\theta\in\Theta_0} P_\theta(\Lambda(\tilde{X}) > \Lambda(x) \mid x)}_{\text{p-value}} \leq \alpha$$

One-sided normal location: $P = P(\tilde{Z} > Z) = 1 - \Phi(Z) = \Phi(-Z)$
$\tilde{Z} \sim N(0,1)$

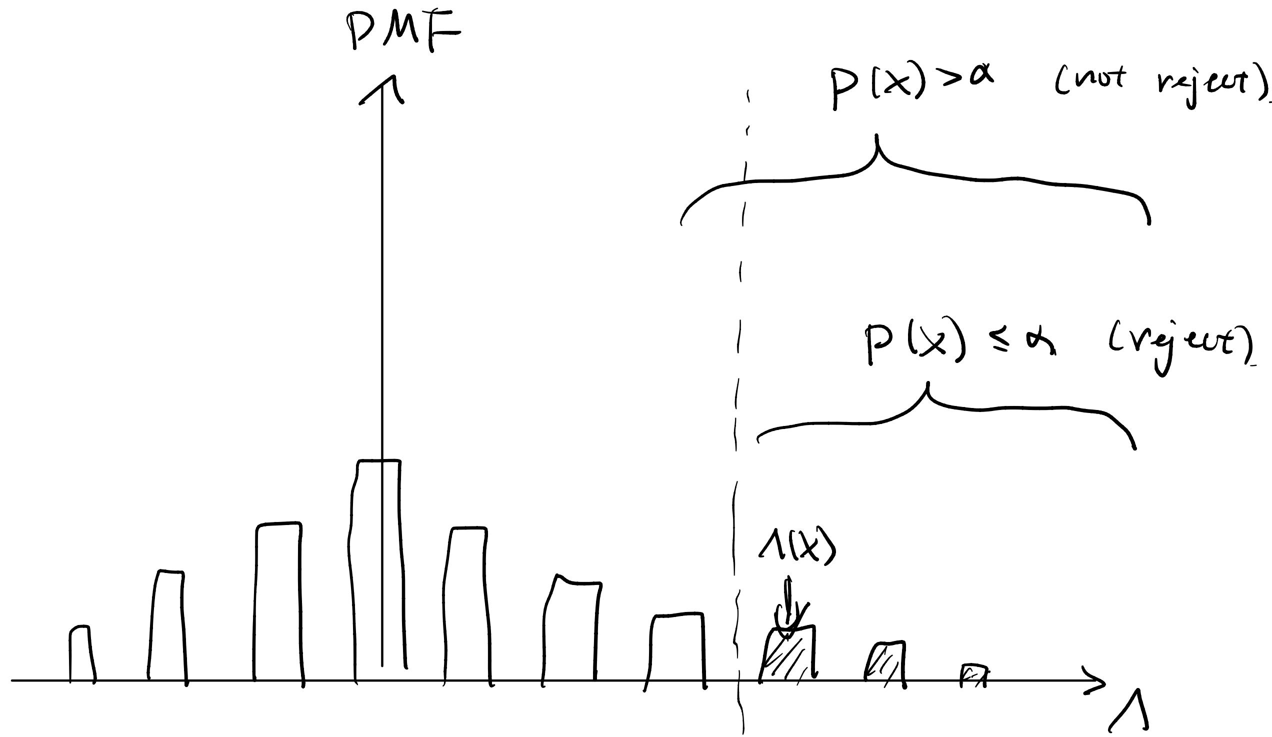
Two-side ~ —— ~ : $P = P(|\tilde{Z}| > |Z|) = 2\,\Phi(-|Z|)$.
$\tilde{Z} \sim N(0,1)$

Let $F(t) = P_{\theta_0}(\Lambda(X) \le t)$. $f(t) = F'(t)$. density. 9.1

Then $p = 1 - F(\Lambda) = \int_{\Lambda}^{\infty} f(t)\, dt$.

$$
\begin{aligned}
\mathbb{P}(p \le u) &= \mathbb{P}(1 - F(\Lambda) \le u) \\
&= \mathbb{P}(F(\Lambda) \ge 1-u) \\
&= \mathbb{P}(\Lambda \ge F^{-1}(1-u)) \\
&= 1 - F(F^{-1}(1-u)) \\
&= 1 - (1-u) \\
&= u. \qquad \text{for all } u \in [0,1].
\end{aligned}
$$

9.1



Duality

Proof $T(\theta_0, x)$ has size α for $H_0: \theta = \theta_0$, $\forall\, \theta_0 \in \Theta$.

$$\iff P_{\theta_0}(T(\theta_0, X) = 1) = \alpha, \quad \forall\, \theta_0 \in \Theta.$$

$$\iff P_{\theta_0}(\theta_0 \notin I(X)) = \alpha, \quad \forall\, \theta_0 \in \Theta.$$

$$\iff P_{\theta_0}(\theta_0 \in I(X)) = 1 - \alpha, \quad \forall\, \theta_0 \in \Theta.$$

$\iff$ $I(X)$ is $(1-\alpha)$-Confidence set. for θ.

---

p-value: $P(\theta_0, X)$ is p-value for $H_0: \theta = \theta_0$.

Then $T(\theta, x) = 1\{p(\theta, x) \leq \alpha\}$

$I(x) = \{\theta : p(\theta, x) > \alpha\}$. is a $(1-\alpha)$-CI because $P(\theta, x)$ is a pivotal quantity.

Ex (two-sided normal) $X_1, \cdots, X_n \overset{\text{IID}}{\sim} N(\mu, \sigma_0^2)$, $\sigma_0^2 > 0$ known.

$$CI(1-\alpha) = \left[\bar{X} - \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}, \ \bar{X} + \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}\right].$$

9.3

Test: $T(\mu_0, x) = 1\left\{|\bar{x} - \mu_0| > \frac{z_{\alpha/2}\sigma_0}{\sqrt{n}}\right\}.$

p-value: $$p = P(|\tilde{Z}| > |Z| \mid Z)$$

$$= 2\,\Phi\left(-\left|\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma_0}\right|\right).$$

$p = (p_1, \cdots, p_k)$. $\sum_{i=1}^{k} p_i = 1$. $\Theta_0 = \{p_0\}$. $p_0 = (p_{01}, \cdots, p_{0k})$ given.

$L(p) \propto p_1^{N_1} \cdots p_k^{N_k}$. $\ell(p) = \log L(p) = \sum_i N_i \log p_i + \text{const}$.

9.5

$$2 \log \Lambda = 2 \left(\sup_{p \in \Theta_0 \cup \Theta_1} \ell(p) - \sup_{p \in \Theta_0} \ell(p) \right).$$

$$= 2 \left(\sup_{p_i \geq 0, \sum p_i = 1} \sum_i N_i \log p_i - \sum_{i=1}^{k} N_i \log p_{0i} \right).$$

Lagrange multiplier: $\frac{\partial}{\partial p_i} \left(\sum_i N_i \log p_i + \lambda \left(\sum_i p_i - 1 \right) \right) = \frac{N_i}{p_i} + \lambda$.

Set this to zero, $\hat{p}_i \propto N_i$. so $\hat{p}_i = \frac{N_i}{n}$.

$$2 \log \Lambda = 2 \sum_i N_i \log \frac{N_i}{n p_{0i}} = 2 \sum_i O_i \cdot \log \left(\frac{O_i}{E_i} \right)$$

where $O_i = N_i$, $E_i = n p_{0i}$

"observed" "expected".

Pearson's statistic. Let $\delta_i = O_i - E_i$, assume $\frac{\delta_i}{E_i} \simeq 0$.

9.5

$$2\log\Lambda = 2\sum_i O_i \log\left(\frac{O_i}{E_i}\right)$$

$$= 2\sum_i (E_i + \delta_i)\log\left(1 + \frac{\delta_i}{E_i}\right)$$

$$\simeq 2\sum_i (E_i + \delta_i)\cdot\left(\frac{\delta_i}{E_i} - \frac{1}{2}\frac{\delta_i^2}{E_i^2}\right)$$

$$\simeq 2\sum_i \left(\delta_i + \frac{\delta_i^2}{E_i} - \frac{1}{2}\cdot\frac{\delta_i^2}{E_i}\right)$$

$$= \underbrace{2\sum_i \delta_i}_{=0} + \sum_i \frac{\delta_i^2}{E_i} = \sum_i \frac{(O_i - E_i)^2}{E_i}.$$

By Wilks' theorem, $2\log\Lambda \xrightarrow[H_0]{d} \chi^2_{\dim(\Theta)-\dim(\Theta_0)} = \chi^2_{(k-1)-0} = \chi^2_{k-1}$

Reject H_0 if $2\log\Lambda > \chi^2_{k-1}(\alpha)$ or $\sum_i \frac{(O_i - E_i)^2}{E_i} > \chi^2_{k-1}(\alpha)$.

10.1

$$2\log\Lambda = 2\left\{\sup_{p_i\geq 0,\ \sum p_i=1} \ell(p) - \sup_{p=p(\theta),\ 0\leq\theta\leq 1} \ell(p)\right\}$$

$$\left(\ell(p) = \sum_i N_i \log p_i + \text{const.}\right)$$

$$= 2\left\{\ell(\hat{p}) - \ell(p(\hat{\theta}))\right\}$$

$$\hat{p}_i = \frac{N_i}{n}.$$

$$= 2\sum_i N_i \log\frac{\hat{p}_i}{p_i(\hat{\theta})} = 2\sum_i N_i \log\frac{\overset{O_i}{N_i}}{\underset{E_i}{n p_i(\hat{\theta})}} = 2\sum_i O_i \log\frac{O_i}{E_i}.$$

$$\simeq \sum_i \frac{(O_i-E_i)^2}{E_i}.$$

$\hat{\theta}$ maximizes $\ell(p(\theta)) = N_{AA}\cdot 2\log\theta + N_{Aa}\log\{2\theta(1-\theta)\} + N_{aa}\cdot 2\log(1-\theta)$.

$$\Longrightarrow \quad \hat{\theta} = \frac{2N_{AA}+N_{Aa}}{2n}.$$

By Wilks' theorem

$$2\log\Lambda \xrightarrow[H_0]{d} \chi^2_{\underbrace{\dim(\Theta_0\cup\Theta_1)}_{=2} - \underbrace{\dim(\Theta_0)}_{=1}}$$

$$P_{x+} = \sum_{y=1}^{c} P_{xy}, \qquad P_{+y} = \sum_{x=1}^{r} P_{xy}.$$

10.3

"marginal probabilities".

$$\ell(p) = \sum_{x=1}^{r} \sum_{y=1}^{c} N_{xy} \log p_{xy} + \text{const}.$$

MLE under H_0: $\hat{P}_{xy} = \hat{P}_{x+} \cdot \hat{P}_{+y} = \frac{N_{x+}}{n} \cdot \frac{N_{+y}}{n}$ (Lagrange multiplier)

MLE under $H_0 \cup H_1$: $\hat{P}_{xy} = \frac{N_{xy}}{n}$.

So
$$2 \log \Lambda = 2 \sum_{x=1}^{r} \sum_{y=1}^{c} N_{xy} \log \frac{N_{xy}/n}{N_{x+}N_{+y}/n^2}$$

$$= 2 \sum_{x=1}^{r} \sum_{y=1}^{c} N_{xy} \log \frac{\overset{= O_{xy}}{N_{xy}}}{\underset{= E_{xy}}{N_{x+}N_{+y}/n}}$$

$$= 2 \sum_x \sum_y O_{xy} \log \frac{O_{xy}}{E_{xy}} \xrightarrow[H_0]{d} \chi^2_{(rc-1)-(r-1)-(c-1)} = \chi^2_{(r-1)(c-1)}$$

Wilks' thm.

$$\approx \sum_x \sum_y \frac{(O_{xy} - E_{xy})^2}{E_{xy}}$$

10.4

$$\begin{array}{cccc} & p_{1|1} & p_{2|1} & p_{3|1} \\ & p_{1|2} & p_{2|2} & p_{3|2} \\ & p_{1|3} & p_{2|3} & p_{3|3} \\ & \| & \| & \| \\ H_0: & p_1 & p_2 & p_3 \end{array}$$

10.5

$$P(X_1 = x_1, Y_1 = y_1, \dots, X_n = x_n, Y_n = y_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i, Y_i = y_i)$$

$$= \prod_{i=1}^{n} P(X_i = x_i) P(Y_i = y_i)$$

$$= \prod_{i=1}^{n} P(X_i = x_i) \cdot \prod_{i=1}^{n} P(Y_i = y_i)$$

$$= \prod_{i=1}^{n} P(X_i = x_i) \cdot \prod_{i=1}^{n} P(Y_{\pi(i)} = y_{\pi(i)})$$

$$= \dots$$

$$= P(X_1 = x_1, Y_{\pi(1)} = y_{\pi(1)}, \dots)$$

permutation

$$\text{p-value} = \frac{1}{n!} \sum_{\pi} I\{2\log\Lambda(X, Y_\pi) \geq 2\log\Lambda(X,Y)\}.$$

10.5

- Monte-Carlo approximation.

sample B random permutations.

$$\hat{p} = \frac{1 + \sum_{i=1}^{B} I\{2\log\Lambda(X, Y_{\pi_i}) \geq 2\log\Lambda(X,Y)\}}{1+B}.$$

$$\mu = E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}, \qquad \Sigma = Cov(X) = \begin{pmatrix} Var(X_1) & \cdots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \cdots & Var(X_n) \end{pmatrix}$$

$$a^T X \sim N(a^T\mu, \ a^T \Sigma a)$$

11.1

$$Z \sim N(\mu, \sigma^2), \quad Z \in \mathbb{R}.$$

$$E(e^Z) = \int_{-\infty}^{\infty} e^z \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

$$= \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-(\mu+\sigma^2))^2}{2\sigma^2}}}_{\text{PDF of } N(\mu+\sigma^2, \sigma^2)} dz \cdot e^{\frac{(\mu+\sigma^2)^2-\mu^2}{2\sigma^2}}$$

$$= e^{\mu+\sigma^2/2}.$$

So $M_X(a) = E(e^{a^T X}) = e^{a^T\mu + a^T \Sigma a/2}$.

If $X \sim N(\mu, \Sigma)$, then $\quad (\Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{1}{2}} = \Sigma)$

$$\underbrace{\Sigma^{-1/2}}_{=A}(X - \underbrace{\mu}_{b=\Sigma^{-1/2}\mu}) \sim N(0, \underbrace{I}_{A\Sigma A^T}) \qquad \underline{11.1}$$

Let $Z_1, \cdots, Z_n \overset{\text{IID}}{\sim} N(0,1)$. $\quad Z = (Z_1, \cdots, Z_n) \sim N(0, I)$

By uniqueness of MVN, PDF of $N(0, I_n)$ is

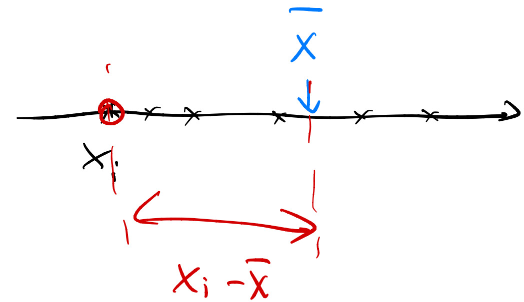
$$(2\pi)^{-n/2} \cdot e^{-z^T z/2} = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2}.$$

11.2

$$Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma_0}$$ Z-test

$\sigma_0 \to$ known

$H_0: \mu = \mu_0$.



Prop $Z \sim N(0, I_n)$, $U \in \mathbb{R}^{n\times n}$, $UU^T = U^TU = I_n$

then $UZ \sim N(0, I_n)$

Pf

11.2

$$U = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\times 1}} & \frac{-1}{\sqrt{2\times 1}} & 0 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3\times 2}} & \frac{1}{\sqrt{3\times 2}} & \frac{-2}{\sqrt{3\times 2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & \frac{-(n-1)}{\sqrt{n(n-1)}} \end{pmatrix}$$

Not difficult to verify $U^T U = U U^T = I_n$.

$\left[\mathbb{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \right]$

$X \sim N(\mu \underline{\mathbb{1}}, \sigma^2 I)$

$\mu \cdot \mathbb{1} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix}$

$$Y = UX = \begin{pmatrix} \frac{1}{\sqrt{n}} \sum x_i \\ * \end{pmatrix} = \begin{pmatrix} \sqrt{n}\,\bar{x} \\ * \end{pmatrix} \sim N\Big(\underbrace{\begin{pmatrix} \sqrt{n}\mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{U \cdot \mu \mathbb{1}}, \sigma^2 I \Big)$$

$$Y = \begin{pmatrix} \sqrt{n}\bar{X} \\ * \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \sigma^2 I \right)$$

$Y = UX$

11.2

$$Y_1 = \sqrt{n}\bar{X}$$

$$\begin{aligned} Y_2^2 + \cdots + Y_n^2 &= \underbrace{Y^T Y} - Y_1^2 \\ &= X^T X - n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{x}^2 \\ &= \sum_{i=1}^n (X_i - \bar{x})^2 = S_{XX}. \end{aligned}$$

$$Y_1 = \sqrt{n}\bar{X} \perp\!\!\!\perp \underbrace{(Y_2, \cdots, Y_n)}_{S_{XX} \text{ is a function of}} \overset{\text{IID}}{\sim} N(0, \sigma^2)$$

$$\frac{S_{XX}}{\sigma^2} = \sum_{i=2}^n \frac{Y_i^2}{\sigma^2} = \sum_{i=2}^n \left(\frac{Y_i}{\sigma}\right)^2 \sim \chi^2_{n-1}.$$

11.2

Student's t: $H_0: \mu = \mu_0$

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{S_{xx}/(n-1)}}$$

[Compare with $Z = \dfrac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0}$]

$$= \frac{\sqrt{n}(\bar{x} - \mu_0)/\sigma}{\sqrt{(S_{xx}/\sigma^2)/(n-1)}}$$

$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

$\frac{S_{xx}}{\sigma^2} \sim \chi^2_{n-1}$

$$\overset{\text{under } H_0}{\sim} \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}}$$

independence.

$$= t_{n-1}.$$

upper α-quantile of t_{n-1}.

One-sided t-test: $H_0: \mu \le \mu_0$ vs. $H_1: \mu > \mu_0$. Reject if $T > t_{n-1}(\alpha)$

Two-sided t-test: $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$. Reject if $|T| > t_{n-1}(\frac{\alpha}{2})$.

$X_1, \cdots, X_n \overset{\text{IID}}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ unknown. 12.1

$$\bar{X} = \frac{1}{n}\sum X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right) \perp\!\!\!\perp (n-1)\hat{\sigma}^2 = S_{xx} = \sum (X_i - \bar{X})^2 \sim \sigma^2 \chi^2_{n-1}$$

Two-sided CI: $\bar{X} \pm t_{n-1}\left(\frac{\alpha}{2}\right) \cdot \boxed{\hat{\sigma}/\sqrt{n}}$

standard error

12.2

$$Y = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_1 \\ \vdots \\ \mu_k \\ \vdots \\ \mu_k \end{pmatrix}, \sigma^2 I_n \right)$$

$H_0: \mu_1 = \cdots = \mu_k$. vs. $H_1: \mu_1, \cdots, \mu_k \in \mathbb{R}$. (but H_0 is not true)

General problem

$$Y \sim N(\mu(\theta), \sigma^2 I_n)$$

MLE:

$$L(\theta, \sigma^2) = (2\pi)^{-n/2} \cdot |\sigma^2 I_n|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2}(Y-\mu(\theta))^T (\sigma^2 I_n)^{-1} (Y-\mu(\theta))\right\}$$

$$= (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{-\frac{1}{2\sigma^2} \| Y - \mu(\theta) \|^2\right\}.$$

So

$$\hat{\theta} = \operatorname{argmin} \| Y - \mu(\theta) \|^2,$$

12.2

$$L(\theta,\sigma^2) = (2\pi\sigma^2)^{-n/2} \cdot \exp\left\{-\frac{1}{2\sigma^2}\|Y-\mu(\theta)\|^2\right\}.$$

$$\ell(\theta,\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\|Y-\mu(\theta)\|^2 + \text{const}.$$

$$\frac{\partial \ell(\theta,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2}\|Y-\mu(\theta)\|^2.$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\|Y-\mu(\hat{\theta})\|^2 \Rightarrow L(\hat{\theta},\hat{\sigma}^2) \propto (\hat{\sigma}^2)^{-n/2} \propto \|Y-\mu(\hat{\theta})\|^{-n}.$$

$$\Lambda = \frac{\sup_{\mu_1,\cdots,\mu_k\in\mathbb{R},\sigma^2>0} L(\mu_1,\cdots,\mu_k,\sigma^2)}{\sup_{\mu\in\mathbb{R},\sigma^2>0} L(\mu,\sigma^2)}$$

$$\rightarrow \hat{\mu}_i = \bar{Y}_{i+} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}.$$

$$\rightarrow \hat{\mu} = \bar{Y}_{++} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}.$$

$$\propto \left(\frac{\sum_{i,j}(Y_{ij}-\bar{Y}_{i+})^2}{\sum_{i,j}(Y_{ij}-\bar{Y}_{++})^2}\right)^{-n/2} = \left(\frac{SSE}{SST}\right)^{-n/2} = \left(\frac{SST}{SSE}\right)^{n/2}.$$

Reject H_0 if $\frac{SST}{SSE}$ large

Variance decomposition

12.2

$$SST = \sum_{ij} (Y_{ij} - \bar{Y}_{++})^2$$

$$= \sum_{ij} (Y_{ij} - \bar{Y}_{i+} + \bar{Y}_{i+} - \bar{Y}_{++})^2$$

$$= \underbrace{\sum_{ij} (Y_{ij} - \bar{Y}_{i+})^2}_{SS\cancel{X}E \text{ "within group"}} + \underbrace{\sum_{ij} (\bar{Y}_{i+} - \bar{Y}_{++})^2}_{SS\cancel{E}A \text{ "between group"}}$$

$$+ 2 \sum_{i=1}^{k} \sum_{j=1}^{n_i} \underbrace{(Y_{ij} - \bar{Y}_{i+})}_{=0} (\bar{Y}_{i+} - \bar{Y}_{++}) = 0.$$

$$= SSA + SSE.$$

So we reject H_0 if $\frac{SSA}{SSE}$ is large.

It follows from t-test:

1. $\frac{SSE}{\sigma^2} \sim \chi^2_{\sum_{i=1}^{k}(n_i - 1)} = \chi^2_{n-k}$.
2. $\frac{SSA}{\sigma^2} \sim \chi^2_{k-1}$. [later]
3. $SSA \perp\!\!\!\perp SSE$.

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \in \mathbb{R}^n \qquad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \in \mathbb{R}^p. \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix} \in \mathbb{R}^n.$$

12.5

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Normal linear model: $\quad Y = X\beta + \varepsilon, \quad \varepsilon \perp\!\!\!\perp X, \quad \varepsilon \sim N(0, \sigma^2 I_n)$

ANOVA

$$X = \begin{pmatrix} \vdots & & 0 \\ & \vdots & \\ 0 & & \ddots \end{pmatrix} = \begin{pmatrix} 1_{n_1} & & 0 \\ & 1_{n_2} & \\ 0 & & \ddots \\ & & & 1_{n_k} \end{pmatrix}. \qquad \beta = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}.$$

12.5

$$X^{\text{full}} = \begin{pmatrix} 1_{n_1} & & & \\ & 1_{n_2} & & 0 \\ & 0 & \ddots & \\ & & & 1_{n_k} \end{pmatrix}, \quad X^{\text{null}} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1_n$$

$$X^{\text{saturated}} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = I_n. \qquad Y = X\beta + \varepsilon$$

$$\vec{\mu}^{\text{saturated}} = Y = \text{projection of } Y \text{ onto colspan}(X^{\text{saturated}}) = \mathbb{R}^n$$

$$\vec{\mu}^{\text{full}} = \begin{pmatrix} \bar{Y}_{1+} \\ \vdots \\ \bar{Y}_{1+} \\ \vdots \\ \bar{Y}_{k+} \\ \vdots \\ \bar{Y}_{k+} \end{pmatrix} \begin{matrix} \}n_1 \\ \\ \\ \\ \}n_k \end{matrix} = \text{projection of } Y \text{ onto colspan}(X^{\text{full}})$$

$$\vec{\mu}^{\text{null}} = \begin{pmatrix} \bar{Y}_{++} \\ \vdots \\ \bar{Y}_{++} \end{pmatrix} = \bar{Y}_{++} \cdot 1_n = \text{projection of } Y \text{ onto colspan}(X^{\text{null}})$$

$$SSA = \| \vec{\mu}^{\text{full}} - \vec{\mu}^{\text{null}} \|^2, \quad SSE = \| \vec{\mu}^{\text{sat.}} - \vec{\mu}^{\text{full}} \|^2.$$

$$SST = SSA + SSE = \| \vec{\mu}^{\text{sat.}} - \vec{\mu}^{\text{null}} \|^2.$$

12.5

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \qquad \beta = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \qquad \text{[abuse notation.]}$$

$$Y_i = X_i^T \beta + \varepsilon_i = \alpha + \beta x_i + \varepsilon_i.$$

1. $Y = X\beta + \varepsilon, \quad \varepsilon \perp\!\!\!\perp X, \quad \varepsilon \sim N(0, \sigma^2 I_n)$ 13.1

$Y_i = X_i^T \beta + \varepsilon_i, \quad \varepsilon_1, \cdots, \varepsilon_n \overset{IID}{\sim} N(0, \sigma^2), \quad \varepsilon = (\varepsilon_1, \cdots, \varepsilon_n) \perp\!\!\!\perp X.$

2. $Y = X\beta + \varepsilon, \quad E(\varepsilon | X) = 0, \quad Var(\varepsilon | X) = \sigma^2 I_n.$

$$L(\beta, \sigma^2) = \underbrace{(2\pi)^{-\frac{n}{2}} \cdot |\sigma^2 I_n|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (Y - X\beta)^T (\sigma^2 I_n)^{-1} (Y - X\beta) \right\}}_{P_{\beta,\sigma^2}(Y | X)} \cdot \underbrace{P_X(X)}_{\text{const.}}$$

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \|Y - X\beta\|^2 + \text{const.}$$ 13.2

$$\Rightarrow \text{MLE } \hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \|Y - X\beta\|^2 = \underbrace{\underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2}_{\text{Least squares.}}$$

$$\frac{\partial \|Y - X\beta\|^2}{\partial \beta_k} = 2\sum_{i=1}^{n} (Y_i - X_i^T\beta)\cdot X_{ik} \qquad k=1,\cdots,p.$$

13.2

So MLE $\hat{\beta}$ solves $\sum_{i=1}^{n} (Y_i - X_i^T\beta)\, X_{ik} = 0$, $k=1,\cdots,p$.

$$\iff (X^TX)\beta = X^TY \xrightarrow{X \text{ has full col. rank}} \hat{\beta} = (X^TX)^{-1}X^TY$$

$$\iff X^T(Y - X\beta) = 0$$

"normal equation".

<u>Ex</u> Simple linear regression: $Y_i = \alpha + \beta X_i + \varepsilon_i$, $\varepsilon_i \overset{\text{IID}}{\sim} N(0,\sigma^2)$

$(\hat{\alpha}, \hat{\beta})$ solves $\min. \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$

So $\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0 \iff \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$.

$\sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta}X_i)\cdot X_i = 0$. So $\sum_{i=1}^{n} \{(Y_i - \bar{Y}) - \hat{\beta}(X_i - \bar{X})\} X_i = 0$.

$$\Rightarrow \quad \hat{\beta} = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y}) X_i}{\sum_{i=1}^{n} (X_i - \bar{X}) X_i} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \left[= \frac{S_{XY}}{S_{XX}} \right].$$

<u>Pf</u> (Gauss-Markov) <u>13.2</u>

$\hat{\beta}$ is <u>linear</u> in Y. ✓

$\hat{\beta}$ is <u>unbiased</u> because

$$\begin{aligned} E(\hat{\beta} \mid X) &= E((X^T X)^{-1} X^T Y \mid X) \\ &= (X^T X)^{-1} X^T E(Y \mid X) \\ &= (X^T X)^{-1} X^T X \beta \\ &= \beta. \quad ✓ \end{aligned}$$

Now consider $\tilde{\beta} = \overset{\text{fixed } n \cdot n \text{ matrix}}{C} Y$.

$\tilde{\beta}$ is unbiased $\Leftrightarrow$ $\beta = E(\tilde{\beta} \mid X) = C \cdot E(Y \mid X) = CX\beta \quad \forall \beta \in \mathbb{R}^p$.

$$\Rightarrow \quad CX = I_p$$

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \quad p \le n \quad \text{usually } p \text{ much smaller. skinny matrix.}$$

13.2

$$A = C - (X^T X)^{-1} X^T, \text{ so } AX = CX - \underbrace{(X^T X)^{-1} X^T X}_{} = I_p - I_p = 0$$

$$\begin{aligned} Var(\tilde{\beta} \mid X) &= Var(CY \mid X) = C \cdot (\sigma^2 I_n) \cdot C^T \\ &= \sigma^2 C C^T = \sigma^2 (A + (X^T X)^{-1} X^T)(A + (X^T X)^{-1} X^T)^T \\ &= \sigma^2 A A^T + \sigma^2 \cdot (X^T X)^{-1} \cancel{X^T X} \cancel{(X^T X)^{-1}} + \sigma^2 \underbrace{AX}_{=0} (X^T X)^{-1} + \sigma^2 (X^T X)^{-1} \underbrace{X^T A^T}_{=0} \\ &= \sigma^2 (X^T X)^{-1} + \sigma^2 A A^T . \end{aligned}$$

$$\begin{aligned} Var(\hat{\beta} \mid X) &= Var((X^T X)^{-1} X^T Y \mid X) \\ &= (X^T X)^{-1} X^T \cdot \sigma^2 I_n \cdot X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} . \end{aligned}$$

$$\mathrm{Var}(a^T\tilde{\beta}|X) = \sigma^2 \cdot a^T(X^TX)^{-1}a + \sigma^2 \cdot a^T A A^T a$$ 13.2

$$= \mathrm{Var}(a^T\hat{\beta}|X) + \sigma^2 \cdot \|A^Ta\|^2$$

$$\geq \mathrm{Var}(a^T\hat{\beta}|X).$$

13.4

$\mu_i = x_i^T \beta$

$P = X(X^TX)^{-1}X^T$ is symmetric, idempotent.

$$P^2 = (X(X^TX)^{-1}\cancel{X^T})\cdot(\cancel{X(X^TX)^{-1}}X^T) = P.$$

13.4

$\Rightarrow$ $\forall$ $v \in \text{colspan}(P)$, so $v = Pw$, $w \in \mathbb{R}^n$

$$Pv = P(Pw) = P^2 w = Pw = v$$

Now for any $v \in \text{colspan}(P)^\perp$, then $P^T v = 0$ by definition.

So $Pv = P^T v = 0$. So $P = UU^T$.

$\Leftarrow$ If $P = UU^T$. symmetric ✓

$$P^2 = U\underbrace{U^T U}_{= I_{\text{rank}(P)}} U^T = UU^T = P.$$ idempotent ✓.

$(I-P)^T = I - P^T = I - P$ 13.5

$(I-P)(I-P) = I - 2P + P^2 = I - 2P + P = I - P.$

---

1. $Y = X\beta + \varepsilon.$ $\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \|Y - X\beta\|^2.$

$\hat{\mu} = \underset{\mu \in \text{colspan}(X)}{\arg\min} \|Y - \mu\|^2 = PY.$

2. t-test. $Y_1, \cdots, Y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ $X = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$

$\hat{\mu}_{MLE} = \bar{Y}.$ $P = X(X^TX)^{-1}X^T = \begin{pmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & & \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}.$ $PY = \begin{pmatrix} \bar{Y} \\ \vdots \\ \bar{Y} \end{pmatrix}$

$$S_{YY} = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \| Y - PY \|^2 = \left\| \begin{pmatrix} Y_1 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{pmatrix} \right\|^2$$

$$= \| (I-P) Y \|^2 .$$

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \| Y - X\beta \|^2$$

14.1

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \| Y - X\beta \|^2$$

$$\Rightarrow \quad \hat{\sigma}^2_{MLE} = \frac{1}{n} \| Y - X\hat{\beta} \|^2 . \qquad \hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= \frac{1}{n} \| Y - PY \|^2$$

$$= \frac{1}{n} \| (I - P) Y \|^2 .$$

Lemma proof

1. $$\begin{pmatrix} P\varepsilon \\ (I-P)\varepsilon \end{pmatrix} = \begin{pmatrix} P \\ I-P \end{pmatrix} \varepsilon$$

$$\sim N\left(0, \begin{pmatrix} P \\ I-P \end{pmatrix} \sigma^2 I \begin{pmatrix} P \\ I-P \end{pmatrix}^T\right)$$

$$= N\left(0, \sigma^2 \begin{pmatrix} P^2 & P(I-P) \\ (I-P)P & (I-P)^2 \end{pmatrix}\right)$$

$$= N\left(0, \sigma^2 \begin{pmatrix} P & 0 \\ 0 & I-P \end{pmatrix}\right).$$

2. $P = UU^T$, $U \in \mathbb{R}^{n \times \operatorname{rank}(P)}$ orthnormal basis for colspan(P)

$$\frac{\|P\varepsilon\|^2}{\sigma^2} = \frac{\varepsilon^T P^T P \varepsilon}{\sigma^2} = \frac{\varepsilon^T P \varepsilon}{\sigma^2} = \frac{\|U^T \varepsilon\|^2}{\sigma^2}$$

$$= \sum_{i=1}^{\operatorname{rank}(P)} \frac{(U^T\varepsilon)_i^2}{\sigma^2} \sim \chi^2_{\operatorname{rank}(P)}.$$

$[U^T\varepsilon \sim N(0, U^T\sigma^2 I_n U) = N(0, \sigma^2 I)]$

14.3

Theorem proof

$$\hat\beta = (X^TX)^{-1}X^TY$$

$$= (X^TX)^{-1}X^T(X\beta+\varepsilon) \qquad \varepsilon \mid X \sim N(0, \sigma^2 I)$$

$$= \beta + (X^TX)^{-1}X^T\varepsilon$$

$$= \beta + (X^TX)^{-1}X^TP\varepsilon \qquad [PX = X \text{ so } X^TP = X^TP^T = X^T.]$$

$$\sim N(\beta, (X^TX)^{-1}X^T\cdot\sigma^2 I\cdot X(X^TX)^{-1}) = N(\beta, \sigma^2(X^TX)^{-1})$$

$$\hat\sigma^2_{MLE} = \frac{1}{n}\|Y - X\hat\beta\|^2$$

$$= \frac{1}{n}\|(I-P)Y\|^2$$

$$= \frac{1}{n}\|(I-P)(X\beta+\varepsilon)\|^2$$

$$= \frac{1}{n}\|(I-P)\varepsilon\|^2. \qquad \frac{n\hat\sigma^2_{MLE}}{\sigma^2} \sim \chi^2_{\text{rank}(I-P)} = \chi^2_{n-p}.$$

So $\hat\beta$ is a function of $P\varepsilon$, $\hat\sigma^2_{MLE}$ is a func. of $(I-P)\varepsilon$.

$\Rightarrow \hat\beta \perp\!\!\!\perp \hat\sigma^2_{MLE}$.

Student's t

$Y_1, \cdots, Y_n \overset{IID}{\sim} N(\mu, \sigma^2)$, $\quad X = I_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. $\quad \beta = \mu$.

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \frac{1}{n} \sum_{i=1}^{n} Y_i = \bar{Y}.$$

14.3

$$\hat{\sigma}^2_{MLE} = \frac{1}{n} \| Y - X\hat{\beta} \|^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = S_{YY}/n.$$

$\Rightarrow$

1. $\bar{Y} \perp\!\!\!\perp S_{YY}$.

2. $\bar{Y} \sim N(\mu, \sigma^2/n)$.

3. $\dfrac{S_{YY}}{\sigma^2} \sim \chi^2_{n-1}$.

$$\hat{\beta}_j \sim N\left(\beta_j,\ \sigma^2 (X^T X)^{-1}_{jj}\right)$$

$$\hat{\sigma}^2 = \frac{1}{n-p} \| Y - \hat{\mu} \|^2.$$

$$(\hat{\beta}_j - \beta_{0j}) \Big/ \sqrt{\sigma^2 (X^T X)^{-1}_{jj}} \sim N(0,1) \quad \text{under } H_0$$

and $\dfrac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p}$ are independent.

$$t_j = \frac{(\hat{\beta}_j - \beta_{0j}) \Big/ \sqrt{\cancel{\sigma^2} (X^T X)^{-1}_{jj}}}{\sqrt{\dfrac{\cancel{(n-p)}\hat{\sigma}^2}{\cancel{\sigma^2}} \Big/ \cancel{(n-p)}}} = \frac{\hat{\beta}_j - \beta_{0j}}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}} \overset{\text{under } H_0}{\sim} t_{n-p}$$

$(1-\alpha)$-CI for β_j : $\quad \hat{\beta}_j \pm t_{n-p}(\alpha/2) \cdot \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}$.

$$\hat{\beta} \overset{H_0}{\sim} N(\beta_0, \sigma^2 (X^T X)^{-1}) \implies (X^T X)^{\frac{1}{2}} (\hat{\beta} - \beta_0) \sim N(0, \sigma^2 I_p).$$

$$\|X(\hat{\beta} - \beta_0)\|^2 = (\hat{\beta} - \beta_0)^T X^T X (\hat{\beta} - \beta_0) \sim \sigma^2 \chi^2_p. \quad \perp\!\!\!\perp \hat{\sigma}^2.$$

14.4

$$\|\hat{\mu} - \mu_0\|^2$$

$$\frac{\{\|X(\hat{\beta} - \beta_0)\|^2 / \cancel{\sigma^2}\} / p}{\{\cancel{(n-p)}\hat{\sigma}^2 / \cancel{\sigma^2}\} / \cancel{(n-p)}} \sim F_{p, n-p} \quad \text{under } H_0.$$

$(1-\alpha)$ - confidence set of β : $\left\{ \beta \in \mathbb{R}^p : \dfrac{\|X(\hat{\beta} - \beta)\|^2}{p \hat{\sigma}^2} \le F_{p, n-p}(\alpha) \right\}$.

[is an ellipsoid.]

Two-sample t-test

$$Y = \begin{pmatrix} A_1 \\ \vdots \\ A_n \\ B_1 \\ \vdots \\ B_m \end{pmatrix}, \quad X = \begin{pmatrix} 1_n & 0_n \\ 1_m & 1_m \end{pmatrix} \quad \beta = \begin{pmatrix} \mu_1 \\ \mu_2 - \mu_1 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} n+m & m \\ m & m \end{pmatrix}, \quad (X^T X)^{-1} = \frac{1}{nm} \begin{pmatrix} m & -m \\ -m & n+m \end{pmatrix}.$$

$$X^T Y = \begin{pmatrix} \Sigma A_i + \Sigma B_i \\ \Sigma B_i \end{pmatrix} = \begin{pmatrix} n\bar{A} + m\bar{B} \\ m\bar{B} \end{pmatrix}$$

$$\hat{\beta}_2 = \left\{ (X^T X)^{-1} (X^T Y) \right\}_2 = \frac{1}{nm} \cdot \left\{ -m(n\bar{A} + m\bar{B}) + (n+m) \cdot m\bar{B} \right\}$$

$$= \frac{1}{nm} \left\{ nm\bar{B} - nm\bar{A} \right\} = \bar{B} - \bar{A}.$$

$$\hat{\beta}_2 \sim N\left(\beta_2,\ \sigma^2 (X^T X)^{-1}_{22}\right) = N\left(\mu_2 - \mu_1,\ \sigma^2 \cdot \frac{n+m}{nm}\right)$$

14.5

$$\hat{\beta}_2 = \bar{B} - \bar{A}$$

$$= N\left(\mu_2 - \mu_1,\ \sigma^2 \cdot \left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

To test $H_0: \mu_1 = \mu_2 \iff \beta_2 = 0.$

The two-sided t-test rejects H_0 if

$$\frac{|\hat{\beta}_2 - 0|}{\sqrt{\hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{1}{m}\right)}} > t_{n-p}(\alpha/2) \iff \frac{(\bar{B} - \bar{A})^2}{\hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{1}{m}\right)} > \left(t_{n-p}(\alpha/2)\right)^2.$$

One-way ANOVA

$$SS_{between} = n\left(\bar{A} - \frac{n\bar{A}+m\bar{B}}{n+m}\right)^2 + m\left(\bar{B} - \frac{n\bar{A}+m\bar{B}}{n+m}\right)^2$$

(SSA)

$$= \frac{mn}{m+n}(\bar{A}-\bar{B})^2$$

14.5

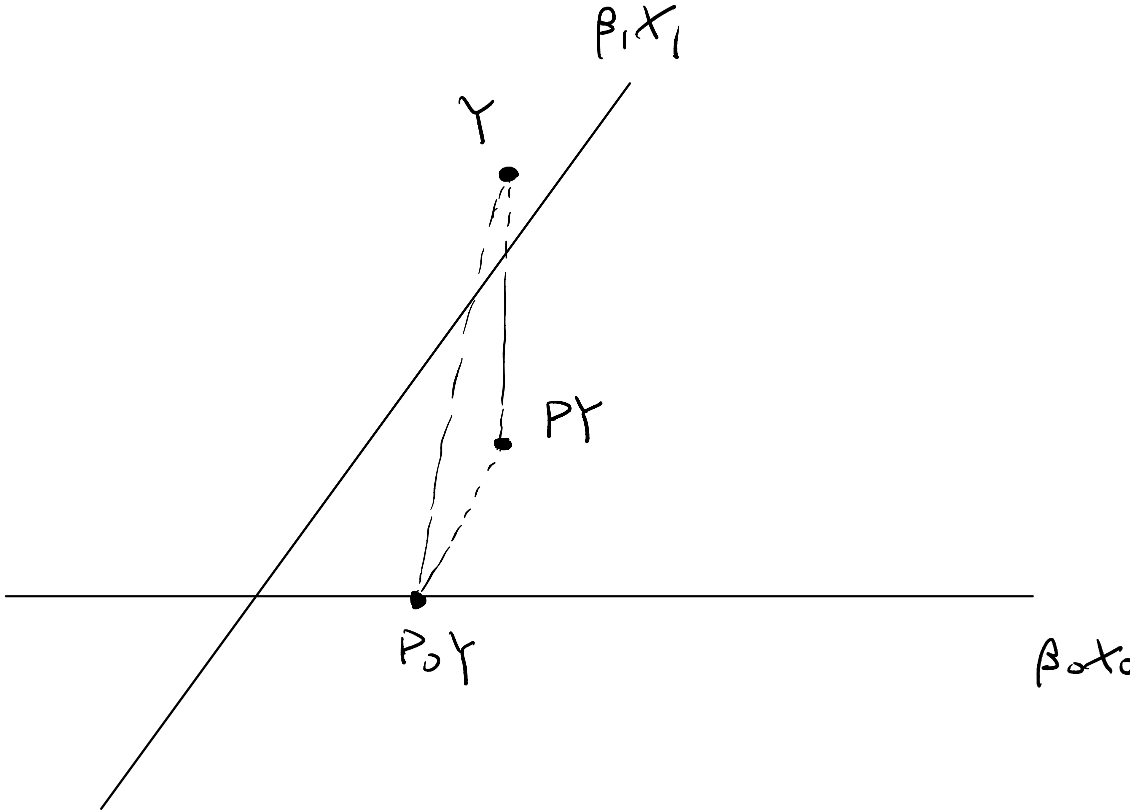
$$SS_{within} = (m+n-2)\hat{\sigma}^2$$

(SSE)

The ANOVA test rejects H_0 if $\dfrac{\frac{mn}{m+n}(\bar{A}-\bar{B})^2/1}{\cancel{(m+n-2)}\hat{\sigma}^2/\cancel{(m+n-2)}} > F_{1,m+n-2}(\alpha)$

If $T = \dfrac{N(0,1)}{\sqrt{\chi_p^2/p}} \sim t_p$, then $T^2 = \dfrac{\chi_1^2/1}{\chi_p^2/p} \sim F_{1,p}$

15.1



$p_0 = 1$

$p = 2$

$n = 3$

$$\varepsilon = (I - P)\varepsilon + (P - P_0)\varepsilon + P_0\varepsilon \qquad \underline{15.1}$$

<u>Thm</u>

$$\begin{pmatrix} (I-P)\varepsilon \\ (P-P_0)\varepsilon \\ P_0\varepsilon \end{pmatrix} \sim N\left(0, \sigma^2 \begin{pmatrix} I-P & 0 & 0 \\ 0 & P-P_0 & 0 \\ 0 & 0 & P_0 \end{pmatrix}\right)$$

$$\|(I-P)\varepsilon\|^2 / \sigma^2 \sim \chi^2_{n-p}$$

$$\|(P-P_0)\varepsilon\|^2 / \sigma^2 \sim \chi^2_{p-p_0}$$

$$\|P_0\varepsilon\|^2 / \sigma^2 \sim \chi^2_{p_0}$$

H_0: sub-model $\Longleftrightarrow H_0: \beta_1 = 0$ $\quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ $\quad \beta_0 \in \mathbb{R}^{p_0}$, $\beta_1 \in \mathbb{R}^{p_1}$

<u>Pf</u> Under H_0, $\beta_1 = 0$ so $Y = X\beta + \varepsilon = X_0 \beta_0 + \varepsilon$ <u>15.2</u>

$$(P - P_0) Y = (P - P_0)(X_0\beta_0 + \varepsilon) = (P - P_0)\varepsilon.$$

$$(I - P) Y = (I - P)(X_0 \beta_0 + \varepsilon) = (I - P)\varepsilon$$

$$\frac{\|(P-P_0)Y\|^2/(p-p_0)}{\|(I-P)Y\|^2/(n-p)} = \frac{\|(P-P_0)\varepsilon\|^2/(p-p_0)}{\|(I-P)\varepsilon\|^2/(n-p)} = \frac{\cancel{\sigma^2}\chi^2_{p-p_0}/(p-p_0)}{\cancel{\sigma^2}\chi^2_{n-p}/(n-p)}$$

$$\sim F_{p-p_0,\, n-p}.$$