$I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X_1)\right]$ is the Fisher information.

We can also start with

$$\frac{\partial}{\partial\theta}\log p_\theta(\underline{x}) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(x_i)$$ and show that

$$\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log p_\theta(\underline{X})\right\}^2\right] = n\,\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right\}^2\right]$$

$$= n\,I(\theta).$$

$$\therefore\ \mathrm{var}_\theta(T(\underline{X})) \geq \frac{\left\{\frac{d}{d\theta}\mathbb{E}_\theta(T(\underline{X}))\right\}^2}{n\,I(\theta)}$$

If T is unbiased for θ, then $\frac{d}{d\theta}\mathbb{E}_\theta(T(\underline{X})) = 1$

If T is unbiased for $a(\theta)$, then $\frac{d}{d\theta}\mathbb{E}_\theta(T(\underline{X})) = a'(\theta)$.

If $W = W(\underline{X})$ is an estimator. Then

$$\operatorname{var}_\theta(W) \geq \frac{\left\{\frac{d}{d\theta} \mathbb{E}_\theta(W)\right\}^2}{I_n(\theta)},$$

where $I_n(\theta) = \mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta} \log p_\theta(\underline{X})\right\}^2\right]$ is the Fisher information contained in the sample $\underset{\sim}{X}$.
$= (X_1, \ldots, X_n)$

$$\mathbb{E}_\theta(W) = \tau(\theta) \quad \forall\, \theta \in \Theta$$

So, another way of stating the Cramer-Rao lower bound:
The variance of any unbiased estimator of $\tau(\theta)$ must be greater than or equal to $\dfrac{\{\tau'(\theta)\}^2}{I_n(\theta)}$.

If $\tau(\theta) = \theta$. Then, for any W with $\mathbb{E}_\theta(W) = \theta \ \forall \theta \in \Theta$, it must hold that

$$\mathrm{var}_\theta(W) \geq \frac{1}{I_n(\theta)}.$$

$$I_n(\theta) = \mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log p_\theta(\underline{X})\right\}^2\right]$$

$$= \mathrm{var}_\theta\left[\frac{\partial}{\partial\theta}\log p_\theta(\underline{X})\right]$$

$X_1, \dots, X_n \overset{iid}{\sim} f_\theta$

$\Rightarrow p_\theta(\underline{x}) = \prod_{i=1}^n f_\theta(x_i)$

$$\therefore \frac{\partial}{\partial\theta}\log p_\theta(\underline{X}) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \log f_\theta(X_i)$$

$$= \sum_{i=1}^n \frac{\partial}{\partial\theta}\log f_\theta(X_i)$$

$$
\begin{aligned}
I_n(\theta) &= \mathrm{var}_\theta\left[\frac{\partial}{\partial\theta}\log p_\theta(\underline{X})\right] \\
&= \mathrm{var}_\theta\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right] \\
&= \sum_{i=1}^{n}\mathrm{var}_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X_i)\right] \\
&= n\,\mathrm{var}_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right] \qquad \left[\because \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right] = 0 \;\; \forall\theta \in \Theta\right] \\
&= n\,\underbrace{\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f_\theta(X_1)\right\}^2\right]}_{\text{Fisher information in a single observation}} \\
&= n\,I(\theta).
\end{aligned}
$$

$I_n(\theta) = n I(\theta)$

C.R. lower bound

for one obs.

$$\frac{[\tau'(\theta)]^2}{I(\theta)}$$

for n obs n.

$$\frac{[\tau'(\theta)]^2}{n I(\theta)}$$

With addition conditions

$$I_n(\theta) = \mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log p_\theta(\underline{X})\right\}^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log p_\theta(\underline{X})\right]$$

In this case,

$$I(\theta) = \mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f_\theta(X)\right\}^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right].$$

Eg. Let $X_1 \dots X_n \overset{iid}{\sim}$ Bernoulli(θ), $\theta \in [0,1]$.

$$f_\theta(x) = \begin{cases} \theta & \text{if } x = 1 \\ 1-\theta & \text{if } x = 0 \\ 0 & \text{o.w.} \end{cases} \quad \leftarrow \text{p.m.f of } X$$

$$= \begin{cases} \theta^x (1-\theta)^{1-x} & \text{if } x = \{0,1\} \\ 0 & \text{o.w.} \end{cases}$$

$$\log f_\theta(x) = x \log \theta + (1-x) \log(1-\theta)$$

$$\frac{\partial}{\partial \theta} \log f_\theta(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x(1-\theta) - \theta(1-x)}{\theta(1-\theta)}$$

$$= \frac{x - \theta x - \theta + \theta x}{\theta(1-\theta)} = \frac{x-\theta}{\theta(1-\theta)}.$$

$$\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f_\theta(X)\right\}^2\right] = \mathbb{E}_\theta\left[\left\{\frac{X-\theta}{\theta(1-\theta)}\right\}^2\right]$$

$$= \frac{1}{\theta^2(1-\theta)^2}\,\mathbb{E}_\theta\left[(X-\theta)^2\right]$$

$$= \frac{1}{\theta^2(1-\theta)^2}\,\mathrm{var}_\theta(X) = \frac{\theta(1-\theta)}{\theta^2(1-\theta)^2}$$

$$= \frac{1}{\theta(1-\theta)} = I(\theta)$$

C.R. lower bound:

$$\mathrm{var}_\theta(W) \geq \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$$ for any unbiased estimator of θ

Check that $\mathrm{var}_\theta(\bar{X}) = \frac{\theta(1-\theta)}{n}$

So, $\bar{X}$ achieves the min. possible variance.

Suppose that with a single observation X, we can find $g(X)$ such that

$$\mathrm{var}_\theta(g(X)) = \frac{1}{I(\theta)}$$

← The C.R. lower bound with a single obs.

Then, $$\mathrm{var}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i)\right) = \frac{1}{nI(\theta)}$$

← The C.R. lower bound for the entire sample.

Thus, $T = T(\underline{X}) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$ is the UMVUE of θ.

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{I(\theta)}\right)$$

$$\hat{\theta}_{MLE} \overset{a}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

↑ CR lower bound.

informal statement

$$\sqrt{n}\left(g(\hat{\theta}_{MLE}) - g(\theta)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{[g'(\theta)]^2}{I(\theta)}\right)$$

$$g(\hat{\theta}_{MLE}) \overset{a}{\sim} N\left(g(\theta), \frac{[g'(\theta)]^2}{n I(\theta)}\right)$$

C.R lower bound for U.E.'s of $g(\theta)$

$T_n \xrightarrow{\mathcal{D}} T$ $\quad$ T is c w.p 1.

$T_n \xrightarrow{P} T$

If $T_n \xrightarrow{\mathcal{D}} c$, then $T_n \xrightarrow{P} c$

$$\sqrt{n}\,(\underbrace{T_n - \theta}) \xrightarrow{\mathcal{D}} N(0, \sigma^2(\theta))$$

$T_n - \theta \xrightarrow{\mathcal{D}} 0$ $\qquad$ $T_n \xrightarrow{P} \theta$

informal discussions.

The multiparameter case

Let $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_K)$ be a vector of parameters.

We are interested in $\underline{\tau}(\underline{\theta})$, which is a m-dimensional vector

We want to find a lower bound on the variance of an unbiased estimator of $\underline{\tau}(\underline{\theta})$.

Thus, the estimator $\underline{W} = \underline{W}(\underline{X})$ has to be an m-dim. vector.

$\mathrm{var}_\theta(\underline{W}) = V$ is an $m \times m$ matrix.

For random variables, $\mathrm{var}_\theta(T) \geq \mathrm{var}_\theta(S) \quad \forall \theta \in \Theta$

$$\Leftrightarrow \mathrm{var}_\theta(T) - \mathrm{var}_\theta(S) \geq 0 \quad \forall \theta \in \Theta.$$

So, we need to define "$\geq$" for matrices.

We are looking for a lower bound on a variance matrix, which is p.s.d.

<u>Def</u>: Let A & B be p.s.d. matrices. We say that A is greater than or equal to B if $A - B$ is p.s.d.

We denote it by writing $A \succeq B$.

A is said to be strictly greater than B if $A - B$ is p.d.

We denote this by writing $A \succ B$.

Lowener order between p.s.d. matrices.

$$A \succcurlyeq B \iff \underline{x}^T A \underline{x} \geqslant \underline{x}^T B \underline{x} \quad \forall \underline{x} \in \mathbb{R}^m.$$

$$V = \mathrm{var}_\theta(\underset{\sim}{w})$$

$$\underset{\sim}{a}^T V \underset{\sim}{a} = \underline{a}^T \mathrm{var}_\theta(\underline{w})\, \underline{a} = \mathrm{var}_\theta(\underline{a}^T \underline{w})$$

A lower bound for V would be such that

$$\underline{a}^T V \underline{a} \geqslant \underset{\sim}{a}^T L \underset{\sim}{a} \quad \forall \underline{a}$$

$$\Updownarrow$$

$$\mathrm{var}_\theta(\underline{a}^T \underset{\sim}{w}) \geqslant \underset{\sim}{a}^T L \underline{a} \quad \forall \underline{a}$$

Cramer-Rao lower bound: For any unbiased estimator

$\underline{W} = \underline{W}(\underline{X})$ of $\underline{\tau}(\underline{\theta}) \in \mathbb{R}^m$, we have

$$\underset{m \times m}{\mathrm{var}_{\theta}(\underline{W})} \succeq \left(\nabla_{\underset{\sim}{\theta}} \underline{\tau}(\underline{\theta})\right)^T \left\{ I_n(\underline{\theta}) \right\}^{-1} \left(\nabla_{\theta} \underline{\tau}(\underline{\theta}) \right)$$

$$\underset{K \times m}{\nabla_{\underline{\theta}} \underset{\sim}{\tau}(\underline{\theta})} = \left(\left(\frac{\partial}{\partial \theta_i} \tau_j(\underset{\sim}{\theta}) \right) \right)_{i,j}$$

$$\underset{K \times K}{I_n(\underset{\sim}{\theta})} = \mathbb{E}_{\underset{\sim}{\theta}} \left[\nabla_{\underset{\sim}{\theta}} \log p_{\underset{\sim}{\theta}}(\underline{X}) \, \nabla_{\underset{\sim}{\theta}} \log p_{\underline{\theta}}(\underline{X})^T \right]$$

$$= \left(\left(\mathbb{E}_{\underset{\sim}{\theta}} \left[\frac{\partial}{\partial \theta_i} \log p_{\underset{\sim}{\theta}}(\underline{X}) \frac{\partial}{\partial \theta_j} \log p_{\underset{\sim}{\theta}}(\underline{X}) \right] \right) \right)_{i,j}$$

Under additional conditions

$$I_n(\theta) = - \mathbb{E}_\theta \left[\nabla^2_{\underset{\sim}{\theta}} \log p_{\underset{\wedge}{\theta}}(\underline{X}) \right]$$

$$= - \left(\left(\mathbb{E}_\theta \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log p_{\underset{\wedge}{\theta}} (\underline{X}) \right] \right) \right)_{i,j}$$

For iid sample:

$$I_n(\theta) = n\, I(\theta) \text{ , where}$$

$$I(\underset{\sim}{\theta}) = \left(\left(\mathbb{E}_\theta \left[\frac{\partial}{\partial\theta_i} \log f_{\underline{\theta}}(X) \frac{\partial}{\partial\theta_j} \log f_{\underline{\theta}}(X) \right] \right) \right)_{i,j}$$

$$= - \left(\left(\mathbb{E}_\theta \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f_{\underline{\theta}}(X) \right] \right) \right)_{i,j}$$

under additional conditions.

- If T is an unbiased estimator of θ and it attains the Cramer-Rao lower bound, then it is the UMVUE of θ.

<u>Eg.</u> Let $X_1, \ldots, X_n \overset{iid}{\sim}$ Bernoulli(p), $p \in [0,1]$.

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

C.R. lower bound $\dfrac{p(1-p)}{n}$.

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ has $\mathbb{E}_p(\bar{X}) = p \quad \forall\, p \longleftarrow$ U.E.

$$\mathrm{Var}_p(\bar{X}) = \frac{p(1-p)}{n} \quad \forall\, p. \longleftarrow \text{attains the C.R. lower bound.}$$

$\therefore$ $\bar{X}$ is the UMVUE of p.

Eg. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$, $\mu \in \mathbb{R}$.

$I(\mu) = 1$

C.R. lower bound $\frac{1}{n}$.

$\mathbb{E}_\mu(\bar{X}) = \mu \quad \forall \mu \in \mathbb{R} \longleftarrow$ U.E.

$\mathrm{Var}_\mu(\bar{X}) = \frac{1}{n} \quad \forall \mu \in \mathbb{R} \longleftarrow$ attains C.R. lower bound.

$\therefore \bar{X}$ is the UMVUE of μ.

Eg. $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, 1)$, $\mu \in \mathbb{R}$.

We want to estimate μ^2.

$I(\mu) = 1$

C.R. lower bound $\dfrac{\left\{\frac{d}{d\mu}\mu^2\right\}^2}{n} = \dfrac{4\mu^2}{n}$

$$\mathbb{E}_\mu(\bar{X}^2) = \underbrace{\operatorname{var}_\mu(\bar{X})}_{\frac{1}{n}} + \{\underbrace{\mathbb{E}_\mu(\bar{X})}_{\mu}\}^2$$

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$$\mathbb{E}(X^2) = V(X) + [\mathbb{E}(X)]^2$$

$$\therefore \mathbb{E}_\mu\left(\bar{X}^2 - \frac{1}{n}\right) = \mu^2$$

$T = \bar{X}^2 - \frac{1}{n}$ is an U.E. of μ^2.

$$\operatorname{var}_\mu(T) = \operatorname{var}_\mu(\bar{X}^2)$$

$$= \mathbb{E}_\mu(\bar{X}^4) - \{\mathbb{E}_\mu(\bar{X}^2)\}^2$$

$$\bar{X} \sim N\left(\mu, \frac{1}{n}\right)$$

$$Z - \mu \sim N(0, \sigma^2)$$

For $Z \sim N(\mu, \sigma^2)$,

$$\mathbb{E}(Z^4) = \mathbb{E}(Z - \mu + \mu)^4$$

$$= \mathbb{E}((Z-\mu)^4) + 4\,\cancel{\mathbb{E}(Z-\mu)^3}\,\mu + 6\,\mathbb{E}(Z-\mu)^2\mu^2 + 4\,\cancel{\mathbb{E}(Z-\mu)}\,\mu^3 + \mu^4$$

$$= 3\sigma^4 + 6\sigma^2\mu^2 + \mu^4$$

[if $X \sim N(0, \sigma^2)$, then $\mathbb{E}(X^4) = 3\sigma^4$]

$$\mathbb{E}_\mu(\bar{X}^4) = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4 \qquad \bar{X} \sim N\left(\mu, \frac{1}{n}\right)$$

$$\sigma^2 = \frac{1}{n}$$

$$= \frac{3}{n^2} + \frac{6\mu^2}{n} + \mu^4$$

$$\mathrm{var}_\mu(\bar{X}^2) = \frac{3}{n^2} + \frac{6\mu^2}{n} + \mu^4 - \left(\mu^2 + \frac{1}{n}\right)^2$$

$$= \frac{3}{n^2} + \frac{6\mu^2}{n} + \cancel{\mu^4} - \cancel{\mu^4} - \frac{2\mu^2}{n} - \frac{1}{n^2}$$

$$= \frac{2}{n^2} + \frac{4\mu^2}{n} \quad \text{C.R. lower bound}$$

> 0

$> $ C.R. lower bound.

$T = \bar{X}^2 - \frac{1}{n}$ is the UMVUE of μ^2.

But it does not attain the C.R. lower bound.

Eg. $X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$, $\theta > 0$

C.R. lower bound $\dfrac{\theta^2}{n}$.

$\hat{\theta}_{MLE} = X_{(n)}$ has pdf $f_{(n)}(y) = \dfrac{n y^{n-1}}{\theta^n}$, $0 < y < \theta$

$\mathbb{E}_\theta(X_{(n)}) = \dfrac{n}{n+1}\theta \leftarrow$ not unbiased.

$\text{var}_\theta(X_{(n)}) = \dfrac{n \theta^2}{(n+1)^2(n+2)} = O\left(\dfrac{n}{n^3}\right) = O\left(\dfrac{1}{n^2}\right) << \dfrac{1}{n}$

If we use $T = \dfrac{n+1}{n} X_{(n)}$

$$\mathbb{E}_\theta(T) = \frac{n+1}{n}\mathbb{E}_\theta(X_{(n)}) = \frac{n+1}{n} \cdot \frac{n}{n+1}\theta = \theta \quad \forall \theta$$

$$\text{var}_\theta(T) = \left(\frac{n+1}{n}\right)^2 \text{var}_\theta(X_{(n)}) = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}.$$

C.R. lower bound does not apply in this case.

Characterization of UMVUE.

Let T be an u.E. of $\theta \Rightarrow \mathbb{E}_\theta(T) = \theta \quad \forall \theta \in \Theta$

Let Z be an estimator s.t. $\mathbb{E}_\theta(Z) = 0 \quad \forall \theta \in \Theta$

Z is an unbiased estimator of 0.

$\Rightarrow T+Z$ is an unbiased estimator of θ.

For any $\lambda \in \mathbb{R}$, $T+\lambda Z$ is an unbiased estimator of θ.

$$\mathrm{var}_\theta(T+\lambda Z) = \mathrm{var}_\theta(T) + \lambda^2 \mathrm{var}_\theta(Z) + 2\lambda \,\mathrm{cov}_\theta(T, Z)$$

<u>Result</u>: Let $\mathcal{Z}$ denote the class of all unbiased estimators of 0. Then, T is the UMVUE of $\mathbb{E}_\theta(T)$ if and only if $\mathrm{cov}_\theta(T, Z) = 0 \quad \forall \theta \in \Theta \quad \forall Z \in \mathcal{Z}$.

<u>Pf:</u> <u>If part:</u> We want to show that if $\mathrm{Cov}_\theta(T,Z)=0\ \forall\theta\ \forall Z\in\mathcal{Z}$

Then T is the UMVUE of $\mathbb{E}_\theta(T)$.

For simplicity, write $\mathbb{E}_\theta(T)=\tau(\theta)$.

Let S be another U.E. of $\tau(\theta)$. $\mathbb{E}_\theta(S)=\tau(\theta)\ \forall\theta$

$\therefore\ \mathbb{E}_\theta(T-S)=\tau(\theta)-\tau(\theta)=0\quad\forall\theta$

$\therefore\ T-S\in\mathcal{Z}$.

$\therefore\ \mathrm{Cov}_\theta(T,T-S)=0\quad\forall\theta$

$\iff \mathrm{Var}_\theta(T)-\mathrm{Cov}_\theta(T,S)=0$

$\iff \mathrm{Cov}_\theta(T,S)=\mathrm{Var}_\theta(T)\quad\forall\theta$

$$\mathrm{Var}_\theta(S)=\mathrm{Var}_\theta(T+S-T)$$
$$=\mathrm{Var}_\theta(T)+\underbrace{\mathrm{Var}_\theta(S-T)}_{\geq 0}+\underbrace{2\,\mathrm{Cov}_\theta(T,S-T)}_{0}\quad \because \mathrm{Cov}_\theta(T,T-S)=0$$

$\therefore\ \mathrm{Var}_\theta(S)\geq\mathrm{Var}_\theta(T)$.

Only if part We need to show that if T is the UMVUE of $\tau(\theta)$, then $Cov_\theta(T, Z) = 0$ $\forall \theta \in \Theta$
$\forall Z \in \mathcal{X}$

Take an arbitrary $Z \in \mathcal{X}$

For any $\lambda \in \mathbb{R}$, $\mathbb{E}_\theta(T + \lambda Z) = \mathbb{E}_\theta(T) = \tau(\theta)$ $\forall \theta \in \Theta$.

$\therefore T + \lambda Z$ is an U.E. of $\tau(\theta)$

Since T is the UMVUE, we must have

$$var_\theta(T + \lambda Z) \geq var_\theta(T)$$

$$\Leftrightarrow \cancel{var_\theta(T)} + \lambda^2 var_\theta(Z) + 2\lambda Cov_\theta(T, Z) \geq \cancel{var_\theta(T)}$$

$$\Leftrightarrow \lambda^2 var_\theta(Z) + 2\lambda Cov_\theta(T, Z) \geq 0$$

$$\Leftrightarrow \lambda Cov_\theta(T, Z) \geq -\frac{\lambda^2}{2} var_\theta(Z)$$

For $\lambda > 0$, $Cov_\theta(T, Z) \geq -\frac{\lambda}{2} var_\theta(Z)$ has to hold for every $\lambda > 0$.

Take $\lambda \to 0$, to get $\operatorname{cov}_\theta(T, Z) \geq 0$.

For $\lambda < 0$, $\operatorname{cov}_\theta(T, Z) \leq -\frac{\lambda}{2} \operatorname{var}_\theta(Z)$. true for any $\lambda < 0$

Take $\lambda \to 0$, to get $\operatorname{cov}_\theta(T, Z) \leq 0$

$\therefore \operatorname{Cov}_\theta(T, Z) = 0 \quad \forall \theta \in \Theta$

Let T be the UMVUE of $\tau(\theta)$

S be an U.E. of $\tau(\theta)$.

Then, $\operatorname{cov}_\theta(T, S-T) = 0 \quad \forall \theta$

$\Rightarrow \operatorname{Cov}_\theta(T, S) = \operatorname{var}_\theta(T) \quad \forall \theta$

$$\operatorname{Corr}_\theta(T, S) = \frac{\operatorname{cov}_\theta(T, S)}{\sqrt{\operatorname{var}_\theta(T)\operatorname{var}_\theta(S)}} = \frac{\operatorname{var}_\theta(T)}{\sqrt{\operatorname{var}_\theta(T)\operatorname{var}_\theta(S)}}$$

$$= \sqrt{\frac{\operatorname{var}_\theta(T)}{\operatorname{var}_\theta(S)}}$$

$$\therefore \quad \frac{\operatorname{var}_\theta(T)}{\operatorname{var}_\theta(S)} = \operatorname{corr}^2_\theta(T,S) = e(S|T) = e(S)$$

$\uparrow$ efficiency of the estimator S (w.r.t. the UMVUE T)

$$0 \le e(S) \le 1.$$

Efficiency is inversely proportional to the variance.

The more the variance, the less efficient the estimator.

⊛ MLEs are superefficient in the sense that their asymptotic variance matches the variance of the UMVUE. informal.

Our result: A statistic T is UMVUE for its expectation if it is uncorrelated with all unbiased estimators of 0.

Cor. Let T_1 be the UMVUE of $\tau_1(\theta)$ and T_2 be the UMVUE of $\tau_2(\theta)$. Then, $T_1 T_2$ is the UMVUE of $\mathbb{E}_\theta(T_1 T_2)$.

Pf: It is enough to show that $\text{Cov}_\theta(T_1 T_2, Z) = 0 \ \forall \theta \ \forall Z \in \mathcal{Z}$

$\iff \mathbb{E}_\theta(T_1 T_2 Z) = 0 \ \forall \theta \ \forall Z \in \mathcal{Z}$

$(\because \mathbb{E}_\theta(Z) = 0 \ \forall \theta)$

Since T_2 is the UMVUE of $\tau_2(\theta)$, so, $\mathbb{E}_\theta(T_2 Z) = 0 \ \forall \theta \in \Theta$

$\therefore T_2 Z \in \mathcal{Z}$. So, $\text{Cov}_\theta(T_1, T_2 Z) = 0 \quad \forall \theta \in \Theta$

$\iff \mathbb{E}_\theta(T_1 T_2 Z) = 0 \quad \forall \theta \in \Theta$

Cor: If T is the UMVUE of $\tau(\theta)$, then

(i) T^2 is the UMVUE of $\mathbb{E}_\theta(T^2)$

(ii) T^3 is the UMVUE of $\mathbb{E}_\theta(T^3)$

$\vdots$

(k) T^k is the UMVUE of $\mathbb{E}_\theta(T^k)$.

⊛ If T_1 is the UMVUE of $\tau_1(\theta)$ & T_2 is the UMVUE of $\tau_2(\theta)$. Then T_1+T_2 is the UMVUE of $\tau_1(\theta)+\tau_2(\theta)$.

Pf: For any $Z \in \mathcal{Z}$. $\mathrm{Cov}_\theta(T_1+T_2, Z) = \mathrm{Cov}_\theta(T_1, Z) + \mathrm{Cov}_\theta(T_2, Z)$

$= 0 \quad \forall \theta$ ($\because T_1$ & T_2 are UMVUEs)

$\therefore$ T_1+T_2 is the UMVUE of $\mathbb{E}_\theta(T_1+T_2)$

$= \tau_1(\theta)+\tau_2(\theta)$.

⊛ If T is the UMVUE of $\mathbb{E}_\theta(T)$, then $a_0+a_1T+\cdots+a_kT^k$ is the UMVUE of $\mathbb{E}_\theta(a_0+a_1T+\cdots+a_kT^k)$.

Let T be an UMVUE of $\tau(\theta)$

For any unbiased estimator S of $\tau(\theta)$

$$\mathrm{Cov}_\theta(T,S) = \mathrm{var}_\theta(T).$$

$$\mathrm{var}_\theta(S-T) = \mathrm{var}_\theta(S) + \mathrm{var}_\theta(T) - 2\underbrace{\mathrm{Cov}_\theta(T,S)}_{\mathrm{var}_\theta(T)}$$

$$= \mathrm{var}_\theta(S) - \mathrm{var}_\theta(T)$$

If $\mathrm{var}_\theta(S) = \mathrm{var}_\theta(T)$ [That is S is also an UMVUE]

Then, $\mathrm{var}_\theta(S-T) = 0$

$\Rightarrow$ $S = T$ with prob. 1.

So, an UMVUE, if it exists, must be unique with prob 1.

- An estimator T is UMVUE if and only if T is uncorrelated with every unbiased estimator of 0.
- If T_1, T_2 are UMVUEs, then $T_1 T_2$ is a UMVUE.
- If T is an UMVUE, then $a_0 + a_1 T + \cdots + a_K T^K$ is the UMVUE of $a_0 + a_1 \mathbb{E}_\theta(T) + \cdots + a_K \mathbb{E}_\theta(T^K)$. Here, $a_0, a_1, \ldots, a_K$ are known constants.

<u>A different way of getting to the UMVUE</u>

Use sufficient statistics

Let T be an unbiased estimator of $\tau(\theta)$.

$$\mathbb{E}_\theta(T) = \tau(\theta) \quad \forall \theta \in \Theta$$

Theorem (Rao - Blackwell) Let T be an unbiased estimator of $\tau(\theta)$ and let S be a sufficient statistic for θ. Define $T^* = \mathbb{E}_\theta(T|S)$. Then,

(a) T^* is a ~~proper~~ estimator.

(b) $\mathbb{E}_\theta(T^*) = \tau(\theta) \quad \forall \theta \in \Theta \leftarrow T^*$ is an **U.E.** unbiased estimator

(c) $Var_\theta(T^*) \leq var_\theta(T) \quad \forall \theta \in \Theta \leftarrow T^*$ is uniformly better than T

Pf: S is sufficient for θ

$\Rightarrow$ The dist of $X_1 \ldots X_n$ given S is free of θ.

$\Rightarrow$ The dist. of $T(X_1 \ldots X_n)$ given S is free of θ

$\Rightarrow$ $\mathbb{E}_\theta(T|S)$ is free of θ.

$\therefore$ T^* does not involve θ. $\Rightarrow$ T^* can be used as an estimator.

$$\mathbb{E}_\theta(T^*) = \mathbb{E}_\theta(\mathbb{E}_\theta(T|S)) \qquad [\,\mathbb{E}(X) = \mathbb{E}\,\mathbb{E}(X|Y)\,]$$

$$= \mathbb{E}_\theta(T) = \tau(\theta) \quad \forall \theta \in \Theta$$

$\therefore$ T^* is an unbiased estimator of θ.

$$\operatorname{var}_\theta(T) = \mathbb{E}_\theta \operatorname{var}_\theta(T|S) + \operatorname{var}_\theta \underbrace{\mathbb{E}_\theta(T|S)}_{T^*}$$

$$= \underbrace{\mathbb{E}_\theta \underbrace{\operatorname{var}_\theta(T|S)}_{\geq 0 \text{ wp1.}}}_{\geq 0} + \operatorname{var}_\theta(T^*)$$

$$\Rightarrow \operatorname{var}_\theta(T) \geq \operatorname{var}_\theta(T^*). \quad \forall \theta \in \Theta$$

So, T^* has uniformly smaller variance than T.

"=" holds, i.e, $\mathrm{var}_\theta(T) = \mathrm{var}_\theta(T^*)$, iff

$$\mathbb{E}_\theta\, \mathrm{var}_\theta(T|S) = 0 \iff \mathrm{var}_\theta(T|S) = 0 \text{ wp. } 1$$

[$\because$ $\mathrm{var}_\theta(T|S)$ is a non-negative r.v.
and for a non-negative r.v. Z, $\mathbb{E}(Z)=0$ iff $Z=0$ wp 1]

$$\mathrm{var}_\theta(T|S) = 0 \text{ wp } 1.$$

$$\mathrm{var}_\theta(T|S) = \mathbb{E}_\theta\left[\left\{T - \mathbb{E}_\theta(T|S)\right\}^2 \middle| S\right]$$

$$\therefore \mathrm{var}_\theta(T|S) = 0 \text{ wp } 1 \iff T - \mathbb{E}_\theta(T|S) = 0 \text{ wp } 1$$

$$\iff T = \mathbb{E}_\theta(T|S) \text{ wp } 1$$

$$\iff T = T^* \text{ wp } 1.$$

The procedure where we start with an unb. estimator T and find $T^* = \mathbb{E}_\theta(T|S)$ for a sufficient statistic S is sometimes called Rao-Blackwallization.

- Rao-Blackwallization does not give us an improvement if and only if after taking the conditional expectation, we get back the same statistic.

Eg. Let X_1, X_2 be iid $N(\theta, 1)$, $\theta \in \mathbb{R}$.

$\bar{X} = \frac{X_1 + X_2}{2}$. $\quad \mathbb{E}_\theta(\bar{X}) = \theta$, $\operatorname{var}_\theta(\bar{X}) = \frac{1}{2}$

$S = X_1 \qquad T^* = \mathbb{E}_\theta(\bar{X} \mid X_1)$

Then, $\mathbb{E}_\theta(T^*) = \theta \quad \forall \theta$

$\operatorname{var}_\theta(T^*) \le \operatorname{var}_\theta(T) \quad \forall \theta$

$$T^* = \mathbb{E}_\theta(\bar{X} \mid X_1) = \mathbb{E}_\theta\left(\frac{X_1 + X_2}{2} \,\middle|\, X_1\right)$$

$$= \mathbb{E}_\theta\left(\frac{X_1}{2} \,\middle|\, X_1\right) + \mathbb{E}_\theta\left(\frac{X_2}{2} \,\middle|\, X_1\right)$$

$$= \frac{X_1}{2} + \mathbb{E}_\theta\left(\frac{X_2}{2}\right) \quad [\because X_2 \text{ is indep of } X_1]$$

$$= \frac{X_1}{2} + \frac{\theta}{2}$$

T^* involves the unknown parameter θ.

$\therefore$ T^* is not a statistic. X_1 is not sufficient.

- Conditioning w.r.t. a sufficient statistic is important.

<u>Theorem.</u> (Lehmann - Scheffe). Let T be an unbiased estimator of $\tau(\theta)$ and let S be a complete sufficient statistic for θ. Then, $\mathbb{E}_\theta(T|S)$ is the UMVUE of $\tau(\theta)$.

Suppose that $T_1 = f_1(S)$ and $T_2 = f_2(S)$ are unbiased estimators of $\tau(\theta)$.

$$\mathbb{E}_\theta(T_1) = \tau(\theta) = \mathbb{E}_\theta(T_2) \quad \forall \theta \in \Theta$$

$$\therefore \mathbb{E}_\theta(T_1 - T_2) = 0 \quad \forall \theta \in \Theta$$

$$\Rightarrow \mathbb{E}_\theta(f_1(S) - f_2(S)) = 0 \quad \forall \theta \in \Theta$$

$$\Rightarrow \mathbb{P}_\theta(f_1(S) - f_2(S) = 0) = 1 \quad \forall \theta \in \Theta$$

$$\Leftrightarrow \mathbb{P}_\theta(f_1(S) = f_2(S)) = 1 \quad \forall \theta \in \Theta. \quad \therefore T_1 = T_2 \text{ wp } 1.$$

Therefore, we cannot have two unbiased estimators of $\tau(\theta)$ which are functions of the complete sufficient statistic S.

Lehmann-Scheffe Theorem tells us that if we start with a function of a complete sufficient statistic, which is unbiased for $\tau(\theta)$, then it is the UMVUE of $\tau(\theta)$.

Eg. Let $X_1, \ldots X_n \overset{iid}{\sim}$ Bernoulli (p), $0 \le p \le 1$.

Then, $\sum_{i=1}^{n} X_i$ is complete sufficient for p

Also, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ is unbiased for p.

$\therefore$ $\bar{X}$ is the UMVUE of p.

Similarly, for $X_1, \ldots, X_n \overset{iid}{\sim}$ Poisson(θ), $\theta > 0$,

$\bar{X}$ is the UMVUE of θ.

for $X_1, \ldots, X_n \overset{iid}{\sim}$ Normal$(\theta, 1)$, $\theta \in \mathbb{R}$

$\bar{X}$ is the UMVUE of θ.

<u>Eg.</u> $X_1, \ldots, X_n \overset{iid}{\sim}$ Uniform$(0, \theta)$, $\theta > 0$.

$$\hat{\theta}_{MLE} = X_{(n)} = \max\{X_1, \ldots, X_n\}$$

$X_{(n)}$ is complete sufficient.

$$\mathbb{E}_\theta(X_{(n)}) = \frac{n}{n+1}\theta \quad \forall \theta \quad \leftarrow \text{not unbiased.}$$

$$\mathbb{E}_\theta\Big(\underbrace{\frac{n+1}{n} X_{(n)}}_{T}\Big) = \theta \quad \forall \theta$$

$T = \frac{n+1}{n} X_{(n)}$ is a fn of a complete sufficient stat.

T is unbiased.

$\therefore$ T is the UMVUE of θ.

Tests of hypotheses

Let $X_1 - - X_n$ be a sample with jt. dist. (pmf/pdf)

$p_\theta(\underline{x})$, $\theta \in \Theta$

In tests of hypotheses, we are interested to check whether $\theta \in \Theta_0$ or $\theta \in \Theta_1$, where Θ_0, Θ_1 are non-overlapping subsets of Θ.

The assertions that $\theta \in \Theta_0$ or $\theta \in \Theta_1$ are known as hypotheses.

A hypothesis is any statement regarding the underlying model.

In our parametric setup, the underlying model is completely specified w.r.t. the unknown parameter. So, any statement about the underlying model will be a statement about the unknown parameter.

Eg. $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$. $p \in [0,1]$.

We may want to check whether $p \leq 0.05$
or $p > 0.05$

Tests of hypotheses

Def: A hypothesis is a statement/assertion/conjecture about the underlying statistical model.

Eg. Suppose $X_1 \ldots X_n \overset{iid}{\sim}$ Bernoulli(p), $0 \le p \le 1$.

p : the prob. of success (getting 1)

"I think this probability is 50%." $(p = \frac{1}{2})$

— a statement/assertion/conjecture about the underlying statistical model. A hypothesis.

"The probability is no more than 70%." $(p \le 0.7)$

— a statement/assertion/conjecture about the underlying statistical model. A hypothesis.

Eg. $X_1, \dots, X_n \overset{iid}{\sim} f$, f is some cont. dist.

"I think f is $N(0,1)$". — a hypothesis

|| Anderson - Darling test.

"f is symmetric about 0". — a hypothesis

"f is normal" — a hypothesis.

The aim in test of hypothesis is to device a technique to decide whether our hypothesis (assertion / conjecture) is true/valid based on the data we observe.

Based on the nature, hypotheses are categorized into two categories:

- Simple hypothesis: If the hypothesis completely specifies the statistical model.

  — $X_1, \ldots X_n \overset{iid}{\sim} Ber(p)$, Hypothesis: $p = \frac{1}{2}$

  Under the hyp., $X_1 \ldots X_n \overset{iid}{\sim} Ber(\frac{1}{2})$ ← completely known.

  — $X_1, \ldots X_n \overset{iid}{\sim} f$, hypothesis: $f = N(0,1)$

  Under the hyp., $X_1 \ldots X_n \overset{iid}{\sim} N(0,1)$ ← complete specification.

· <u>Composite hypothesis</u>: If the hypothesis does not completely specify the statistical model,

- $X_1, \ldots, X_n \overset{iid}{\sim} \text{Ber}(p)$, hyp: $p \leq 0.7$

Here, under hyp., we can have $X_1 - X_n \overset{iid}{\sim} \text{Ber}(0.4)$

also $X_1 - - X_n \overset{iid}{\sim} \text{Ber}(0.6)$

. . . . infinitely many other possibilities.

- $X_1 \ldots X_n \overset{iid}{\sim} f$. hyp. f is normal.

↑ not complete specification

f can be $N(0,1)$ or $N(5,1)$ or $N(0,10)$ . . .

- $X_1 \ldots X_n \overset{iid}{\sim} f$, hyp. f is symmetric about 0

f can be Cauchy$(0,1)$

Laplace$(0,1)$ . . .

In our context of parametric inference.

$X_1 \cdots X_n$ have jt dist. (pdf/pmf) $p_\theta(\underline{x})$, $\theta \in \Theta$

parameter space.

$p_\theta(\cdot)$ is known to us except for θ

So, any statement/assertion/conjecture about the statistical model $p_\theta(\cdot)$ will be a statement/assertion/conjecture about the unknown parameter θ.

In particular, any such assertion will be of the form $\theta \in \tilde{\Theta}$, where $\tilde{\Theta} \subset \Theta$

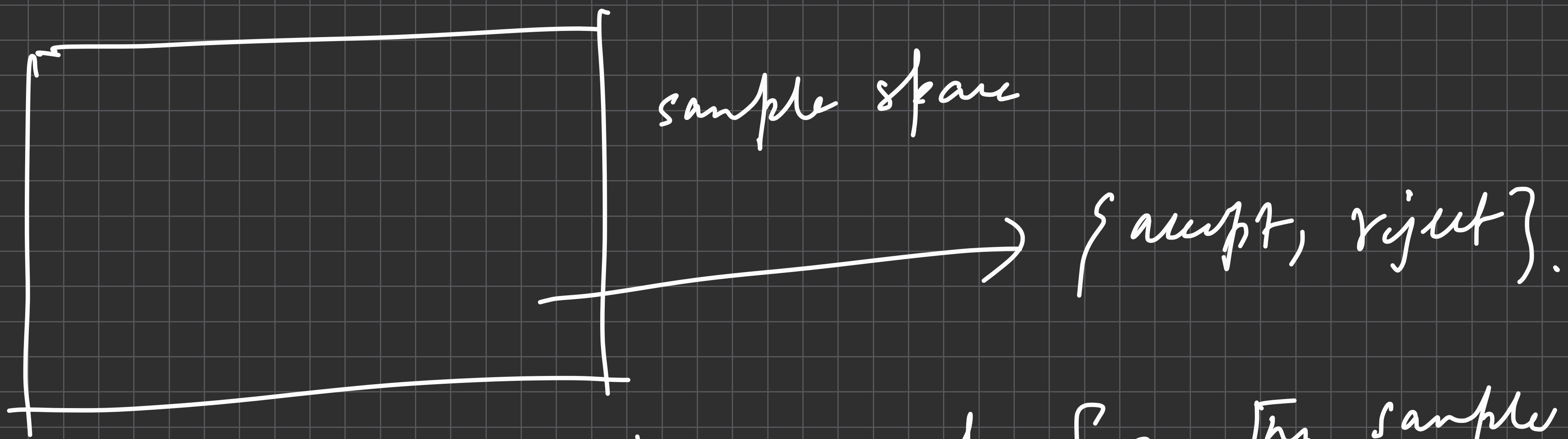
- A simple hyp. if $\tilde{\Theta}$ is a singleton set
- A composite hyp. if $\tilde{\Theta}$ has cardinality more than 1

Notation: We use H / $\mathcal{H}$ to denote a hypothesis.

In fact, in a hypothesis testing problem, there will be two hypothesis H_0 and H_1 which will be tested against each other. (to be seen soon).

In a test of hyp. problem, we either accept or reject a hypothesis.

Def: A test function is a rule that tells us whether to accept or reject a hypothesis H based on the data $X_1 \dots X_n$.

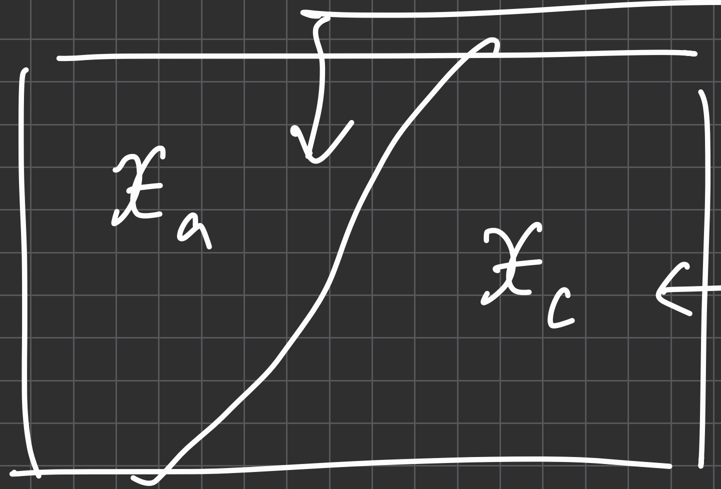


A test function is a map from the sample space $\mathcal{X}$ to {accept, reject}.

The most general situation is when we partition the sample space into two parts $\mathcal{X}_a$ and $\mathcal{X}_c$ (called the acceptance region and the rejection/critial region) so that if $(x_1 \ldots x_n) \in \mathcal{X}_a$ — accept H

$(x_1 \ldots x_n) \in \mathcal{X}_c$ — reject H

accept if $(x_1 \dots x_n)$ falls here

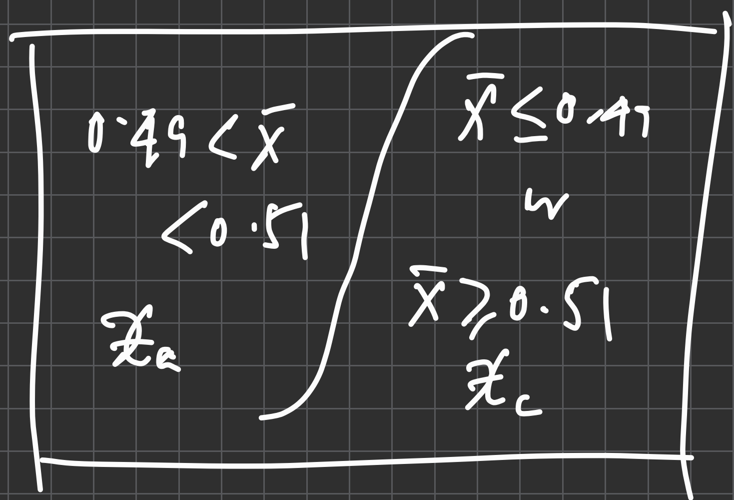


reject if $(x_1 \dots x_n)$ falls here

Eg. $X_1 \dots X_n \overset{iid}{\sim} \mathrm{Ber}(p)$. $H: p = \frac{1}{2}$.

Accept H if $0.49 < \bar{X} < 0.51$

Reject H if $\bar{X} \leq 0.49$ or $\bar{X} \geq 0.51$



$\mathcal{X}$ all possible values $\bar{X} \in [0,1]$

Alternatively, we could also do the following.

Toss a coin. If head comes, we accept H

If tail comes, we reject H.

Here, our decision (to accept/reject H) depends on a random trial, which has nothing to do with the observation.

If such random trial is used in the construction of a test, then the test is called a randomized test.

Usually, a randomized test will partition the sample space $\mathcal{X}$ into three parts —

- $\mathcal{X}_a \longleftarrow$ acceptance region
- $\mathcal{X}_c \longleftarrow$ rejection / critical region.
- $\mathcal{X}_r \longleftarrow$ randomized region

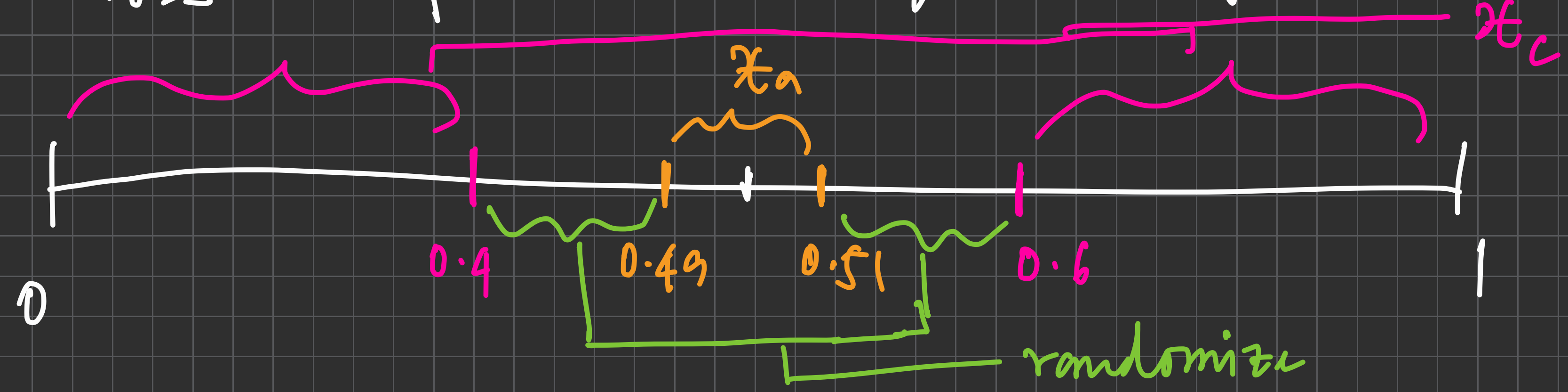
Eg. $X_1, \ldots, X_n \overset{iid}{\sim} Ber(p)$, $H: p = \frac{1}{2}$.

Accept H if $0.49 < \bar{X} < 0.51$,

Reject H if $\bar{X} < 0.4$ or $\bar{X} > 0.6$

If $0.4 \leq \bar{X} \leq 0.49$ or $0.51 \leq \bar{X} \leq 0.6$ then

toss a fair coin, Reject H if head appears.



A non-randomized / deterministic test is a particular case of a randomized test where the randomization region $\mathcal{X}_r$ is ϕ. (null).

An alternative way of representing a randomized test is via a critical function:

$$\varphi(\underline{x}) = \begin{cases} 1 & \text{if } \underline{x} \in \mathcal{X}_c \\ 0 & \text{if } \underline{x} \in \mathcal{X}_a \\ \gamma & \text{if } \underline{x} \in \mathcal{X}_r \end{cases}$$

$\gamma \in [0,1]$ is the probability with which we reject H if the observation falls in the randomization region.

$\varphi(\cdot)$ gives us the prob. of rejecting H (depending on the observed values).

If $\underline{x} \in \mathcal{X}_a$ — then $\varphi(\underline{x}) = 0 \leftarrow$ we reject with prob 0 / we accept

$\underline{x} \in \mathcal{X}_c$ — then $\varphi(\underline{x}) = 1 \leftarrow$ we reject with prob 1 / we reject

$\underline{x} \in \mathcal{X}_r$ — then $\varphi(\underline{x}) = \gamma \leftarrow$ we reject with prob. γ.

γ does not depend on $\underline{x}$

[toss a coin with prob of head equal to γ. Reject if head appears

In a testing problem, we will have one of the following four scenarios:

		Reality	
		H is true	H is false
Our decision	accept H	✓	Error of type II
	Reject H	Error of type I	✓