Neyman- Pearson lemma.

b) consider a test ϕ s.t.

$$\phi(x) = 1 \text{ if } f_{\theta_1}(\vec{x}) > k f_{\theta_0}(\vec{x})$$
$$= \gamma \text{ if } f_{\theta_1}(x) = k f_{\theta_0}(x)$$
$$= 0 \text{ otherwise.}$$

with k & γ satisfying $E_{\theta_0}[\phi(\vec{x})] = \alpha$.
is the MP test in Φ_α.

$$\lambda(\vec{x}) = \frac{f_{\theta_1}(\vec{x})}{f_{\theta_0}(\vec{x})}.$$

—— x —— x ——

Example :- $X_1, X_2, \ldots X_n \sim$ Bernoulli (p).

$H_0 : p = p_0$ & $H_1 : p = p_1$. $(p_0 < p_1)$

Given $\alpha \in [0,1]$, Design MP test in Φ_α.

$$P_{p_0}(\vec{x} = \vec{x}) = p_0^{\Sigma x_i}(1-p_0)^{n-\Sigma x_i}$$

$$P_{p_1}(\vec{x} = \vec{x}) = p_1^{\Sigma x_i}(1-p_1)^{n-\Sigma x_i}$$

$$\frac{P_{p_1}(\vec{x} = \vec{x})}{P_{p_0}(\vec{x} = \vec{x})} = \underbrace{\left(\frac{p_1}{p_0}\right)}_{>1}^{\Sigma x_i} \underbrace{\left(\frac{1-p_1}{1-p_0}\right)}_{<1}^{n-\Sigma x_i}$$

$\lambda(\vec{x})$ monotone increasing fn of $\sum_{i=1}^{n} x_i$

Take $\Sigma \vec{x} = N$

Check if $P_{p_0}(\vec{x}) \geq \alpha. \Rightarrow p_0^n \geq \alpha.$

stop & calculate γ, s.t.

$$\boxed{\gamma p_0^n = \alpha} \Rightarrow \gamma = \frac{\alpha}{p_0^n}, \quad k = \left(\frac{p_1}{p_0}\right)^n$$

$$\left(\begin{array}{l} \phi(\vec{x}) = \gamma \text{ if } \Sigma x_i = n \\ \qquad = 0 \quad \text{o.w.} \end{array}\right).$$

Prob. of H_1 being true is largest when $\sum x_i = n$.

$P_{p_0}(\vec{x}) < \alpha$.

then take $\vec{x}$ s.t. $\sum x_i = n-1$

Prob. of $p_0^n + \binom{n}{1} p_0^{n-1}(1-p_0) \geq \alpha$?

picking $\sum x_i = n$, or $\sum x_i = n-1$.

$$p_0^n + \gamma \binom{n}{1} p_0^{n-1}(1-p_0) = \alpha.$$

$$k = \frac{P_{p_1}(\sum x_i = n-1)}{P_{p_0}(\sum x_i = n-1)}$$

$$\frac{P_{p_1}(\sum X_i = n-u)}{P_{p_0}(\sum X_i = n-u)} = \left(\frac{p_1}{p_0}\right)^{n-u}\left(\frac{1-p_1}{1-p_0}\right)^u < \left(\frac{p_1}{p_0}\right)^{n-1}\left(\frac{1-p_1}{1-p_0}\right) = k$$

$$< \left(\frac{p_1}{p_0}\right)^{n}\left(\frac{1-p_1}{1-p_0}\right)^0$$

$$\boxed{\lambda(x) = \frac{f_{\theta_1}(\vec{x})}{f_{\theta_0}(\vec{x})} \quad \begin{array}{ll} > k & = 1 \\ = k & = \gamma \\ < k & = 0 \end{array}}$$

Example: $X_1, \ldots X_n \sim G(\mu, 1)$

$H_0 : \mu = \mu_0$ & $H_1 : \mu = \mu_1$ $(\mu_0 < \mu_1)$.

$$\lambda(\vec{x}) = \frac{f_{\mu_1}(\bar{x})}{f_{\mu_0}(\bar{x})} = \prod_{i=1}^{n} \frac{f_{\mu_1}(x_i)}{f_{\mu_0}(x_i)}$$

$$= e\left[-\sum_{i=1}^{n}\left[(x_i-\mu_1)^2 - (x_i-\mu_0)^2\right]/2\right]$$

$$(x_i-\mu_1)^2 - (x_i-\mu_0)^2 = \cancel{x_i^2} - 2x_i\mu_1 + \mu_1^2 - \cancel{x_i^2} + 2x_i\mu_0 - \mu_0^2$$

$$= -2x_i(\mu_1-\mu_0) + (\mu_1^2 - \mu_0^2)$$

$$\Rightarrow e^{-n(\mu_1^2-\mu_0^2)} \, e^{2(\mu_1-\mu_0)\sum x_i}$$

$$> 0$$

$\lambda(\vec{x})$ is monotone increasing fn of $\sum x_i$.

Page No.
Date

Test $\phi(\vec{x}) = 1$ if $\sum x_i \geq z$.
$= 0$ o.w.

$E_{\mu_0}[\phi(\vec{x})] = \alpha$. $\quad \sim G(n\mu_0, n^2)$

i.e. $P_{\mu_0}(\phi(x) = 1) = P_{\mu_0}(\sum x_i \geq z)$ → convert in error func.

$= P_{\mu_0}\left(\frac{\sum x_i - n\mu_0}{n} \geq \frac{z - n\mu_0}{n}\right) = Erfc\left(\frac{z - n\mu_0}{n}\right)$

Find z. s.t.
$Erfc\left(\frac{z - n\mu_0}{n}\right) = \alpha$.

Result: MP test is unique if ϕ & ϕ' are both MP, then $P(\phi(\vec{x}) \neq \phi'(\vec{x})) = 0$.

1/10/17 *

- Interval Estimation

$X_1, \ldots, X_n \sim f_\theta$, $\quad \theta \in \Theta \subseteq \mathbb{R}$

l ——— u.

Find $L : \mathcal{X} \to \mathbb{R}$
$U : \mathcal{X} \to \mathbb{R}$.

$S(\vec{x}) = [L(\vec{x}), U(\vec{x})]$.
↳ A random interval.

"Good" event → $\{\theta \in S(x)\}$.
"Bad" event → $\{\theta \notin S(x)\}$

Problem: Find $\underbrace{[L(x), U(x)]}_{S(x)}$ s.t. $P_\theta(\theta \in S(x)) \geq 1 - \alpha$, $\forall \theta \in \Theta$.

* $L(X)$ Lower Confidence bound for θ at confidence level $1-\alpha$.

* A family of random sets $\{ S(\vec{x}) : \vec{x} \in X \}$ of $\Theta \subseteq \mathbb{R}$ is said to constitute a family of confidence sets at conf. level $(1-\alpha)$ if

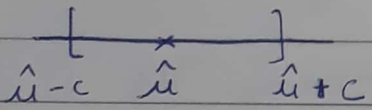
$\cancel{P_\theta(\theta \in S}$

$$P_\theta(\theta \in S(\vec{x})) \geq 1-\alpha, \ \forall \ \theta \in \Theta$$

* Example $X_1, \dots X_n \sim G(\mu, 1)$ $\mu \in \mathbb{R}$.

α is specified.

UMVUE $\vdash$ Sample Mean. $\frac{1}{n}\sum_{i=1}^{n} x_i$ (Point estimation)



$P_\mu(\hat{\mu} - c \leq \mu \leq \hat{\mu} + c)$.

$= P_\mu(\hat{\mu} \leq \mu + c, \ \hat{\mu} \geq \mu - c)$ (as $\hat{\mu}$ is our R.V., we need condn on this)

$= P_\mu(\mu - c \leq \hat{\mu} \leq \mu + c)$

$$P_\mu\left(-c \leq \frac{1}{n}\sum_{k=1}^{n} x_k - \mu \leq c\right)$$

$G(0, 1/n)$

2/11/17.

Recap:

Confidence intervals:

$X_1, ---- X_n \sim f_\theta(\cdot)$ i.i.d. $\theta \in \mathbb{R}$.

Find $L(\vec{x})$ and $U(\vec{x})$, equivalently,

$$S(\vec{x}) = [L(\vec{x}), U(\vec{x})] \quad \text{s.t.}$$

$$P_\theta(\theta \in S(\vec{x})) \geq 1-\alpha \quad \forall \theta$$

and given value of α.

$X_1, X_2, X_3 --- X_n \sim$ Geometric (p).

$$P_p(X_i = k) = p^{k-1}(1-p).$$

Unbiased estimator for p.

$$E_p[1 - \mathbb{1}\{X_1 = 1\}] = 1 - P(X_1 = 1)$$
$$= 1 - (1-p)$$
$$= p.$$

$$P(X_1 = 1 \mid \Sigma X_i = k)$$

$$= \frac{P(\Sigma X_i = k \mid X_1 = 1) P(X_1 = 1)}{P(\Sigma X_i = k)}$$

$$= \frac{P(\sum_{i=2}^{n} X_i = k-1) P(X_1 = 1)}{P(\Sigma X_i = k)}$$

Page No.
Date

2/11/17
Recap:

Example $x_1, x_2, \ldots, x_n \sim G(\mu, 1)$.

$$u(\vec{x}) = \frac{1}{n} \sum x_i + c.$$

$$L(\vec{x}) = \frac{1}{n} \sum x_i - c \qquad c > 0.$$

Result: by choosing appropriate value of c, you can achieve any level of confidence α.

$\frac{1}{n} \sum x_i$ converges to μ. the prob. of $\frac{\sum x_i}{n} \to n$ increases towards 1 as we $\uparrow n$. weak law of large no.s

* Uniformly most accurate (UMA)

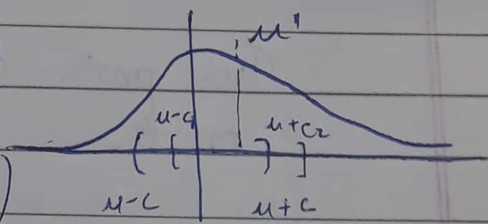
- If pdf symmetric, then interval can be symmetric around the unbiased estimator.

The collection of $S(\vec{x})$ is said to be UMA if
$P_\theta(\theta' \in S(\vec{x}) b) \leq P_\theta(\theta' \in S'(\vec{x}))$

$\forall\ \theta' \neq \theta$ & interal collection $S'(\vec{x})$.

$P(\mu' \in [\frac{1}{n} \sum x_i - c_1, \frac{1}{n} \sum x_i + c_2])$

$\geq P(\mu' \in [\frac{1}{n} \sum x_i - c, \frac{1}{n} \sum x_i + c])$



where c, c_1, c_2 are chosen to meet the required α.

$P(\mu' - c_2 \leq \frac{1}{n} \sum x_i \leq \mu' + c_1)$ Here in Gaussian, max when symmetric.

Pivot: $T : \mathcal{X} \times \Theta \to \mathbb{R}$ ~~Random~~

random variable is called pivot if the distribution of T is independent of θ.

Sufficient statistic's distribution can depend on θ. Also, T can be a func of parameters as well, while Suff. sta is indep. of θ.

Example : $X_1 - - - X_n \sim G(\mu, 1)$

$$T(\vec{x}, \mu) = \frac{1}{n} \Sigma x_i - \mu \sim G\left(0, \frac{1}{n}\right).$$

Find c_1 & c_2 s.t.

$$P_\theta (c_1 \leq T(\vec{x}, \theta) \leq c_2) \geq 1 - \alpha.$$

$$= P_\theta \left(c_1 \leq \frac{1}{n} \Sigma x_i - \mu \leq c_2 \right)$$

$$= P_\theta \left(\mu + c_1 \leq \frac{1}{n} \Sigma x_i \leq \mu + c_2 \right).$$

The distrib. of T is indep. of θ.

This prob. can compute & will not depend on on θ.

6/11/17.

Recap : confidence Interval:-

$X_1, \dots, X_n$ iid $f_\theta(\cdot)$ $\theta \in \mathbb{R}$.

Find $L : \mathcal{X} \to \mathbb{R}$ & $U : \mathcal{X} \to \mathbb{R}$ s.t.

$P[\theta \in [L(\vec{x}), U(\vec{x})]) \geq 1-\alpha$ $\forall \theta \in \Theta$, for any given $\alpha \in [0,1]$.

$(L[\vec{x}], U[\vec{x}])$ is called α-confidence interval.

Method of Pivots :

$T(\vec{x}, \theta)$ is called Pivot if the distribution of $T(\vec{x}, \theta)$ does not depend on θ.

$X_1, \dots, X_n \sim G(\mu, 1)$.

$$\frac{1}{n}\sum_{i=1}^{n} x_i - \mu = T(\vec{x}, \mu) \sim G\left(0, \frac{1}{n}\right).$$

$$P(-c \leq \underbrace{\bar{x} - \mu}_{} \leq c) = P(\bar{x} - c \leq \mu \leq \bar{x} + c)$$

(ind. of μ, as ind. of μ)

$$g(c) = 1-\alpha.$$

$X_1, X_2, \dots X_n \sim$ Uniform $[0, \theta)$.

Find α confidence interval for θ.

$E[X_i] = \theta/2$.

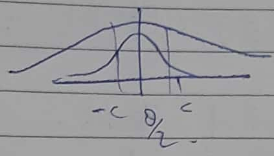
$$P\left(-c \leq \frac{1}{n}\sum_{i=1}^{n} x_i - \frac{\theta}{2} \leq c\right)$$

this might depend on θ, as variance not independent of θ. var $\to \frac{\theta^3}{3} - \frac{\theta^2}{4}$. not independent

$\therefore$ to keep prob. same, c_0 should also change.

trying to build $\left[\frac{1}{n}\sum x_i - c, \frac{1}{n}\sum x_i + c\right]$.

$\frac{\theta}{2} - c \le \bar{x} \le \frac{\theta}{2} + c$. Prob. of $\bar{x}$ lying in $\left(\frac{\theta}{2} - c, \frac{\theta}{2} + c\right)$ is changing with θ.



* $Y = \max\{X_1, \ldots, X_n\}$.

If multiply by λ, only variance changes.
If $ax+b$, shrink, expand, and shift.

$X \sim U[0,\theta]$. $\quad ax. \Rightarrow P(aX \le u) = P(X \le u/a)$.

$T(\bar{X},\theta) = \max\{X_1/\theta, X_2/\theta, \ldots, X_n/\theta\} = \frac{Y}{\theta}$

$P(\max\{Y_1, \ldots, Y_n\} \le x\}$.

$$= 0 \quad \text{if} \quad x < 0.$$
$$= x^n \quad \text{if} \quad x \in [0,1].$$
$$= 1 \quad \text{o.w.}$$

$$f_T(x) = 0 \quad \text{if} \quad x \notin [0,1].$$
$$= nx^{n-1} \quad \text{o.w.}$$

$$P(c \le T(\bar{X},\theta) \le 1) = \int_c^1 nx^{n-1}\,dx = 1 - c^n$$

To ~~gener~~ guarantee α- confidence level, it is enough to ensure $\alpha = c^n \Rightarrow c = \alpha^{1/n}$.

$\left[\alpha^{1/n} \le Y/\theta \le 1 \right)$

$$= \left(\frac{1}{\alpha^{1/n}} \ge \frac{\theta}{Y} \ge 1 \right) = \left(\frac{Y}{\alpha^{1/n}} \ge \theta \ge Y \right].$$

confidence interval for θ :- $\left[Y, \frac{Y}{\alpha^{1/n}} \right]$

θ can't be less than max. of X.

* Inequalities & central limit thm:

⇒ Chebyshev's inequality:

$$P(|X - EX| > \varepsilon) \le \frac{E[X - EX|^2}{\varepsilon^2}$$

$X_1, \ldots, X_n \sim$ Bernoulli (p) $\quad p \in (0,1)$.

$$P\left(\left| \frac{1}{n}\sum X_i - p \right| > \varepsilon \right) \le \frac{Var(Y)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}$$

Y.

$$\le \frac{1}{4n\varepsilon^2} \quad \because p(1-p) \le \frac{1}{4}.$$

$$\Rightarrow P\left(\frac{1}{n}\sum X_i - c \le p \le \frac{1}{n}\sum X_i + c \right) \ge 1 - \alpha.$$

$$= P\left(\left| \frac{1}{n}\sum X_i - p \right| \le c \right) = 1 - P\left(\left| \frac{1}{n}\sum X_i - p \right| > c \right)$$

$$1 - P\left(\left| \frac{1}{n}\sum X_i - p \right| > c \right) \ge 1 - \alpha.$$

$$P\left(\left| \frac{1}{n}\sum X_i - p \right| > c \right) \le \alpha.$$

$\therefore$ It is enough to ensure $\frac{1}{4nc^2} \le \alpha \Rightarrow c \ge \frac{1}{\sqrt{4n\alpha}}$.

Page No.
Date

$$\left[\frac{1}{n}\sum x_i - \frac{1}{\sqrt{4n\alpha}},\ \frac{1}{n}\sum x_i + \frac{1}{\sqrt{4n\alpha}} \right]$$

* Central limit Thm:

$$\frac{\sum_{i=1}^{n} x_i - \overset{\text{mean}}{np}}{\underset{\sigma}{\sqrt{np(1-p)}}} \approx G(0,1)$$

σ standard deviation.

Regression:

$x_1, x_2 \ldots x_r \rightarrow \boxed{?} \rightarrow y_1, \ldots, y_r$

Think inside the box. What's in the box?

Linear regression:

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i x_i + \epsilon \rightarrow \text{Random error.}$$

estimate these.

zero ~~mean~~ mean r.v.

(If some dc value, add it to β_0, then estimate)

$$y_i = \alpha + \beta x_i + \epsilon_i \leftarrow \text{simple regression.}$$

Aim: Given - $x_1, x_2, \ldots, x_n$ - inputs

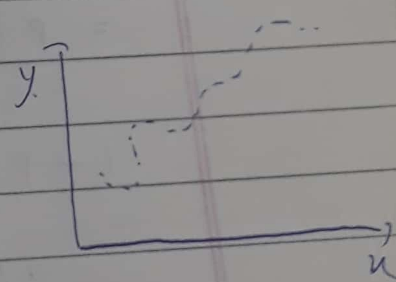
$y_1, y_2, \ldots, y_n$ - responses.

Estimate / find α & β, s.t. Error is minimum.

$$\mathcal{E}(A,B) = \sum_{i=1}^{n} (y_i - (A + Bx_i))^2$$

estimate for α

estimate for β.



Pick A, B s.t $\mathcal{E}(A,B)$ is minimized.

ch-9
Sheldon
Ross.

$Y \sim (a, \sigma^2)$

$nY \sim (an, n^2\sigma^2)$

Page No.

Date

$$\frac{\partial E}{\partial A} = -\sum_{i=1}^{n} \left(2(Y_i - (A + Bx_i)\right) = 0$$

$$\Rightarrow \frac{\partial E}{\partial B} = -\sum_{i=1}^{n} 2x_i(Y_i - A - Bx_i) = 0$$

$A = \bar{Y} - B\bar{x}$ $\qquad \bar{Y} = 1/n \sum Y_i$, $\bar{x} = 1/n \sum x_i$

$$B = \frac{\sum x_i Y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2} \sim \text{Gaussian}$$

E_i's are assumed to be iid Gaussian $(0, \sigma^2)$

$Y_i \sim G(\alpha + \beta x_i, \sigma^2)$ independent. (not identical, as mean different).

$E[B] = \sum(\alpha + \beta x_i) - n\bar{x}$

$= \beta$. $\qquad E[A] = \alpha$.