Random Samples.

(1) Basic Concepts:

① **Def**: $X_1, X_2 \cdots X_n$ is random samples from population of size n $f(x|\theta)$ if $X_k \sim f(x|\theta)$, i.i.d. $\forall\, 1 \le k \le n$.

Remark: This model sometimes called sampling from infinite population. When population is finite:

$$\nearrow \text{With replacement} \longrightarrow \{X_k\}_1^n \text{ is also random samples}$$

$$\searrow \text{Without replacement} \longrightarrow \{X_k\}_1^n \text{ isn't random samples!}$$

since it's not indept. but identical dist!

Def: Statistic T is a vector/real valued function of $(X_1, X_2 \cdots X_n)$, the random samples. $T = T(X_1 \cdots X_n)$

Its domain is on sample space.

e.g. $T(X_1, X_2 \cdots X_n) = \bar{X} = \frac{\sum_1^n X_k}{n}$, or $S^2 = \sum_1^n \frac{(X_k - \bar{X})^2}{n-1}$

Prop. i) $E(\bar{X}) = \mu$, $E(S^2) = \sigma^2$. Unbiased Estimator.

ii) $\min_{a \in \mathbb{R}} \sum_1^n (x_i - a)^2 = \sum_1^n (x_i - \bar{x})^2$, $(n-1)S^2 = \sum_1^n X_k^2 - n\bar{X}^2$

② Dist. of Statistic:

- Usually, the dist of statistic is difficult to generate. But if we sample from scale-location family. it's easy to derive.

Thm. If $X_k \sim f(x|\theta) = h(x)c(\theta)e^{\sum_1^k w_i(\theta)t_i(x)}$. exponential family.

Def $T_i(X_1 \cdots X_n) = \sum_{j=1}^n t_i(X_j)$, If $\{(w_1(\theta) \dots w_k(\theta)) | \theta \in \Theta\}$

contain an open set. Then. $T = (T_1(\vec{X}) \dots T_k(\vec{X}))$

has dist $= H(u_1 \cdots u_k)[c(\theta)]^n e^{\sum_1^k w_i(\theta)u_i}$

Pf: For $\vec{X}$: $f(x_1 \cdots x_n|\theta) = [\prod_1^n h(x_i)] c^n(\theta) e^{\sum_1^k w_i(\theta)[\sum_{k=1}^n t_i(x_k)]}$

Let $\sum_1^n t_i(x_j) = u_i$. $1 \le i \le k$. By reversing. (IFT)

$\Rightarrow \prod_1^n h(x_i) = h(u_1 \cdots u_k)$ □

Remark: It can't be applied such as $N(\theta, \theta^2)$, (θ, θ^2) is closed!

(2) Sampling from

Normal Dist:

① About $\bar{X}$ and S^2:

· Suppose $X_1, X_2 \cdots X_n \sim N(\mu, \sigma^2)$. i.i.d. We have:

i) $\bar{X}$ indept with S^2.

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

Pf: WLOG. Let $\mu = 0$. $\sigma = 1$.

i) Check pmf: By $f(x_1 \cdots x_n) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum_1^n x_i^2}{2}}$

Let $y_1 = \bar{x}$. $y_2 = x_2 - \bar{x}$ --- $y_n = x_n - \bar{x}$.

(It's reasonable. $\bar{x}$ is ancilary statistic for

σ^2. Sufficient for μ. $X_k - \bar{X}$ is about σ^2)

$\Rightarrow \left|\frac{\partial(x_1 \cdots x_n)}{\partial(\eta_1 \cdots \eta_n)}\right| = \frac{1}{n}$ $\therefore f(\vec{\eta}) = \left[\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{n\eta_1^2}{2}}\right]\left[\frac{n^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{\sum_2^n \eta_i^2 + (\sum_2^n \eta_k)^2}{2}}\right]$

i.e. $(X_2-\bar{X}, \cdots X_n-\bar{X})$ indept with $\bar{X}$.

Since $\sum(X_2-\bar{X}) = \bar{X}-X_1$ $\therefore (X_1-\bar{X}, X_2-\bar{X} \cdots X_n-\bar{X})$ indept with $\bar{X}$.

ii) <u>Lemma.</u> Since $\chi_p^2 \sim Gamma(2,p) = \frac{1}{\Gamma(\frac{p}{2})2^{\frac{p}{2}}} e^{-\frac{x}{2}} \cdot x^{\frac{p}{2}-1}$

p is the degree of freedom. $\longrightarrow$ For $X=\min(z_1, z_2)$ $X^2 \sim \chi_1^2$!

i) $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi_1^2$.

ii) X_k. indept. $X_k \sim \chi_{p_k}^2$. Then $\sum_1^n X_k \sim \chi^2_{\sum_1^n p_k}$.

$\Rightarrow$ Prove: $(n-1)S_n^2 \sim \chi_{n-1}^2$ by induction on n. ($n=1$ ✓)

Note that $(n-1)S_n^2 = (n-2)S_{n-1}^2 + \left(\frac{n-1}{n}\right)(X_n-\bar{X}_{n-1})^2$.

$X_n-\bar{X}_{n-1} \sim$ Normal dist. Check $Var(X_n-\bar{X}_{n-1}) = \frac{n}{n-1}$.

$\therefore \frac{n-1}{n}(X_n-\bar{X}_{n-1})^2 \sim \chi_1^2$. indept with $(n-2)S_{n-1}^2 \sim \chi_{n-2}^2$.

③ <u>The derived dist:</u>

<u>Student's t and Kronecker's F:</u>

- To do some estimation. Sometimes we need pivot statistic $\longrightarrow$ generate its dist!

i) <u>Estimate σ by S:</u>

- $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is the basis for inference about μ.

when σ is unknown. Let $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu/\sigma/\sqrt{n}}{\sqrt{\frac{S^2}{\sigma^2}}}$

$\sim \frac{Z}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \sim t_p$: $f_{t_p} = \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} \frac{1}{\sqrt{p\pi}} \frac{1}{(1+\frac{t^2}{p})^{\frac{p+1}{2}}}$. $t \in \mathbb{R}$. $p=n-1$ degree of free.

Remark: For t_p, it only has $p-1$ moments. By $t_p \sim \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$

Z indept with χ^2_{n-1}, since $\bar{X}$ indept with S^2.

$$\therefore E(t_p) = E(Z)\,E\left(\frac{1}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}\right) = 0$$

$$Var(t_p) = \frac{p}{p-2}, \quad p>2.$$

ii) Estimate ratio σ_X^2/σ_Y^2:

- The information of this ratio is contained in S_X^2/S_Y^2, where $X_k \sim N(\mu_X, \sigma_X^2)$, $Y_k \sim N(\mu_Y, \sigma_Y^2)$

$$S_X^2 = \frac{1}{n-1}\sum_1^n (X_k-\bar{X})^2, \quad S_Y^2 = \frac{1}{m-1}\sum_1^m (Y_k-\bar{Y})^2.$$

$\Rightarrow$ We can consider the statistic:

$$\frac{\sigma_X^2/\sigma_Y^2}{S_X^2/S_Y^2} = \frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2} \longrightarrow \text{derive } F_{m-1,n-1} \text{ dist.}$$

Note that $$\frac{S_Y^2/\sigma_Y^2}{S_X^2/\sigma_X^2} \sim \frac{\frac{\chi^2_{m-1}}{m-1}}{\frac{\chi^2_{n-1}}{n-1}} \sim F_{m-1,n-1}.$$

$$F_{p,q} \sim \frac{\Gamma(\frac{p+q}{2})}{\Gamma(\frac{p}{2})\Gamma(\frac{q}{2})}\left(\frac{p}{q}\right)^{\frac{p}{2}} \frac{x^{\frac{p}{2}-1}}{[1+\frac{p}{q}x]^{\frac{p+q}{2}}}$$

Remark: i) $E(F_{p,q}) = E\left(\frac{\chi^2_{p-1}}{p-1}\right) E\left(\frac{q-1}{\chi^2_{q-1}}\right) = \frac{q-1}{q-3}$

ii) It's easy to see $\frac{1}{X} \sim \frac{1}{\chi_p^2/p\,/\,\chi_q^2/q} = \frac{\chi_q^2/q}{\chi_p^2/p} \sim F_{q,p}$

iii) Kronecker's F can be derived by Beta dist:

Since $X \sim F_{n,m} \Rightarrow \frac{\frac{n}{m}X}{1+\frac{n}{m}X} \sim Beta(\frac{n}{2}, \frac{m}{2})$

Thm. For $X \sim t_p$

i) $X^2 \sim F_{1,p}$ ⟶ cor. $2F_{2,p} \to \chi_2^2$.

ii) If $f(x|p)$ is pmf of X. Then $p \to \infty$. $f(x|p) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Pf: i) $X \sim \dfrac{Z}{\sqrt{\frac{\chi_p^2}{p}}} \Rightarrow X^2 \sim \dfrac{\chi_1^2}{\frac{\chi_p^2}{p}} \sim F_{1,p}$

ii) By $\Gamma(n) \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$.

$$f(x|p) \sim \frac{\left(\frac{p+1/2}{e}\right)^{\frac{p+1}{2}}}{\left(\frac{p/2}{e}\right)^{\frac{p}{2}}} \cdot \frac{\sqrt{(p+1)\pi}}{\sqrt{p\pi}} \cdot \frac{\left(1+\frac{t^2}{p}\right)^{-\frac{p+1}{2}}}{\sqrt{p\pi}}$$

$$= \sqrt{\frac{p+1}{2e}} \left(1+\frac{1}{p}\right)^{\frac{p}{2}} \cdot \sqrt{\frac{p+1}{p}} \cdot \frac{\left(1+\frac{t^2}{p}\right)^{-\frac{p+1}{2}}}{\sqrt{p\pi}} \to \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}.$$

(3) Order Statistic:

Thm. $X_1, X_2 \cdots X_n$ random samples from discrete dist

$f_X(x_i) = p_i$. $x_1 < x_2 \cdots < \cdots$. $P_i = \sum_{k=1}^{i} p_k$. Then.

$$P(X_{(j)} \le x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1-P_i)^{n-k}.$$

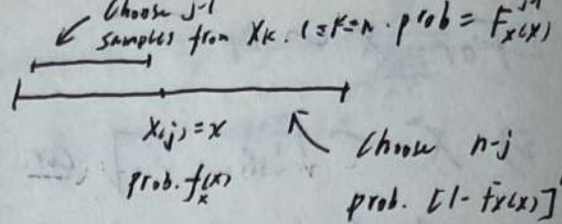
Remark: It means at least $k \ge j$ $X_{(i)}$. $1 \le i \le k$. fall into $[-\infty, x_i]$. easy to prove!

Thm. $X_1 \dots X_n$ random samples from conti. dist $F_X(x)$. $f_X(x)$

$$\text{Then } f_{X_{(j)}}(x) = \binom{n}{j-1}\binom{n-j+1}{n-j} f_X(x) F_X^{j-1}(x) [1-F_X(x)]^{n-j}$$

Pf: Note that $F_{X_{(j)}}(x) = \sum_{k \ge j}^{n} \binom{n}{k} F_X^k(x) [1-F_X(x)]^{n-k}$

Remark: Geometrical interpretation:

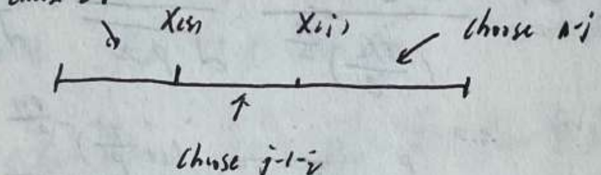


Cor. $$f_{X_{(i)},X_{(j)}}(u,v) = \binom{n}{i-1}\binom{n-i}{j-i-1}\binom{n-j+1}{n-j} f_X(u) f_X(v)\cdot F_X^{i-1}(u)$$

$$\cdot\,[F_X(v)-F_X(u)]^{j-i-1}[1-F_X(v)]^{n-j}\text{; where } j>i.$$

And in particular, $f_{X_{(1)},X_{(2)}\cdots X_{(n)}}(x_1\cdots x_n) = n!\prod_1^n f_X(x_i)$.

Pf: geometrical:



$\Rightarrow$ **Prop.** (Conditional Case)

$$f_{X_{(i)}|X_{(j)}}(u,v) = \binom{j-1}{i-1}\binom{j-i}{j-i-1}\frac{f_X(u)}{F_X(v)}\left[\frac{F_X(u)}{F_X(v)}\right]^{i-1}\left[1-\frac{F_X(u)}{F_X(v)}\right]^{j-i}$$

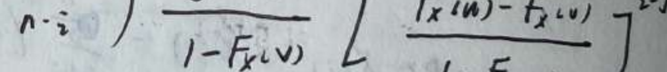
where $j>i$, $u\le v$.

$$f_{X_{(i)}|X_{(j)}}(u,v) = \binom{n-j}{i-j-1}\binom{n-i+1}{n-i}\frac{f_X(u)}{1-F_X(v)}\left[\frac{F_X(u)-F_X(v)}{1-F_X(v)}\right]^{i-j-1}$$

$$\cdot\left[1-\frac{F_X(u)-F_X(v)}{1-F_X(v)}\right]^{n-i}\text{.}\quad j<i,\ u\ge v.$$

Pf:

$\text{prob}\,|X_{(j)} = \dfrac{\text{prob}}{F_X(v)}$



$\text{prob}\,|X_{(j)} = \dfrac{\text{prob}}{1-F_X(v)}$

(4) Delta Method:

· It can be apply to find dist of function of r.v.'s.

Mostly. We consider one-moment expansion:

Denote $g_i'(\theta_1 \cdots \theta_k) = \frac{\partial g}{\partial t_i}\Big|_{t_1=\theta_1 \cdots t_k=\theta_k}$

$\Rightarrow g(t_1, t_2 \cdots t_k) = g(\theta_1, \theta_2, \cdots \theta_k) + \sum_1^k g_i'(\vec{\theta})(t_i - \theta_i) + o$

Suppose $\vec{T} = (T_1 \cdots T_k)$. T_i is r.v. with mean θ_i.

$\Rightarrow E_\theta(g(T)) \simeq g(\vec{\theta}) + \sum_1^k g_i'(\vec{\theta}) E(T_i - \theta_i) = g(\vec{\theta})$.

$\therefore Var_\theta(g(T)) \simeq Var((g(\vec{T}) - g(\vec{\theta}))^2) = Var((\sum_1^k g_i'(\vec{\theta})(T_i - \theta_i))^2)$

$= \sum_1^k g_i'(\vec{\theta})^2 Var_\theta(T_i) + 2\sum_{i<j}^k g_i'(\theta) g_j'(\theta) Cov_\theta(T_i, T_j)$

$\Rightarrow$ Thm. For $g: \mathbb{R}^k \to \mathbb{R}$.$^{(*)}$ If $\sqrt{n}(\hat{\theta}_n - \vec{\theta}) \to N(0, \Sigma)$

(*) Extend to $\mathbb{R}^k$. replace α by matrix

(Denote $\hat{\theta}_n \sim AN(\vec{\theta}, \Sigma/n)$) Then, we have:

$\sqrt{n}(g(\hat{\theta}_n) - g(\vec{\theta})) \longrightarrow N(0, \alpha^T \Sigma \alpha)$. i.e.

$g(\hat{\theta}_n) \sim AN(g(\theta), \frac{\alpha^T \Sigma \alpha}{n})$. $\alpha = (\frac{\partial g}{\partial \hat{\theta}_n})\Big|_{\hat{\theta}_{n1}=\theta_1 \cdots \hat{\theta}_{nk}=\theta_k}$

Pf: Lemma. (Cramer Wold Device)

$(X_{1n}, X_{2n} \cdots X_{kn}) \xrightarrow{d} (X_1, X_2 \cdots X_k)$ $(n \to \infty)$

$\Leftrightarrow \forall (a_i) \subseteq \mathbb{R}$. $\sum_{i=1}^k a_i X_{in} \xrightarrow{d} \sum_1^k a_i X_i$ $(n \to \infty)$

Pf: By Characteristic Func. !

Cor. For $n=1$. $g(\hat{\theta}_n) \sim AN(g(\theta), g'(\theta)^2 \sigma^2/n)$

For $\hat{\theta}_n \sim AN(\theta, \frac{\sigma^2}{n})$.

Thm. (Second - Order Moment for one dimension)

For $\theta_n \sim AN(\theta, \frac{\sigma^2}{n})$. If $g'(\theta)=0$. $g''(\theta)\neq 0$

Then $n(g(\theta_n)-g(\theta)) \xrightarrow{d} \sigma^2 g''(\theta)\chi_1^2/2$.

Pf: $n(g(\theta_n)-g(\theta)) = n\frac{g''(\theta)}{2}(\theta_n-\theta)^2$

$= \frac{\sigma^2 g''(\theta)}{2}\left[\frac{\sqrt{n}}{\sigma}(\theta_n-\theta)\right]^2 \to \frac{\sigma^2 g''(\theta)}{2}\chi_1^2$.

e.g. $X_k \sim X$. $1\le k\le n$. i.i.d. $V_4 = E(|X-\mu|^4)<\infty$.

Then $S^2 \sim AN(\sigma^2. \frac{(V_4-\sigma^4)}{n})$

Pf: $S^2 = \frac{1}{n-1}\sum_1^n (X_k-\bar{X})^2 \sim \frac{1}{n}\sum_1^n (X_k-\bar{X})^2 \quad (n\to\infty)$

Note that $\frac{1}{n}\sum(X_k-\mu)^2 \sim AN(\sigma^2. V_4-\sigma^4)$

By $E(X_k-\mu)^2=\sigma^2$. $Var((X_k-\mu)^2) = V_4-\sigma^4$.

$\therefore S^2 = \frac{\sum_1^n (X_k-\mu)^2 - n\sigma^2}{n} + \frac{n\sigma^2 + \sum((X_k-\bar{X})^2-(X_k-\mu)^2)}{n}$

$= \frac{\sum_1^n((X_k-\mu)^2-\sigma^2)}{n} + \sigma^2 + \frac{(\mu-\bar{X})}{n}\sum(2X_k-\mu-\bar{X})$

$\therefore \sqrt{n}(S^2-\sigma^2) = \frac{1}{n}\sum_1^n((X_k-\mu)^2-\sigma^2) + (\bar{X}-\mu)^2$

$\longrightarrow N(0, V_4-\sigma_4) \quad (n\to\infty)$

since $\bar{X}-\mu \xrightarrow{P} 0$. $(n\to\infty)$

(6) Generate a Random Samples:

We will transform Uni-dist to desired dist following:

① Direct Method:

i) For Y conti. r.v. Besides, F_Y is bijective d.f.

Then $F_Y^{-1}(u) \sim Y$. where u is uniform dist.

e.g. $F_Y^{-1}(u) = -\lambda \log(1-u) \sim Exp(\lambda)$.

($-\lambda \log u \sim Exp(\lambda)$, too. Note that $-\lambda\log(1-u)\uparrow$

$-\lambda \log u \downarrow$. Caution to reverse "$\le$" !)

$\Rightarrow$ Since $\chi_2^2 \sim Exp(2)$.

$\therefore \chi_{2\nu}^2 = \sum_1^{\nu} -2\log(u_i)$.

Moreover. $Y = \sum_1^{\alpha} -\beta\log(u_i) \sim Gamma(\alpha, \beta)$.

ii) For Y is discrete r.v. $p(Y=y_{i+1}) = F_Y(y_{i+1}) - F_Y(y_i)$

$= p(F_Y(y_i) < u \le F_Y(y_{i+1}))$

$\therefore$ To generate Y: If u falls into $(F_Y(y_i), F_Y(y_{i+1})]$

Then set $Y = y_{i+1}$.

iii) For Y r.v. F_Y is difficult to figure out.

We can use LLN: $\dfrac{\sum_1^n I\{Y \le x\}}{n} \longrightarrow p(Y \le x) = F_Y(x)$

② Indirect Method:

· For $Y \sim f_Y(x)$, $f_Y(x) \equiv 0$. $x \notin [a,b]$, $\sup_x f_Y(x) \le c < \infty$. For some c.

1°) Generate (U, V). indept Uniform dist. $U \sim U(0,1)$. $V \sim U(a,b)$

2°) If $U < \frac{1}{c} f_Y(V)$. set $Y = V$. otherwise return to 1°)

Pf: Firstly. $p(V \le y, U \le \frac{1}{c} f_Y(V)) = \int_a^y \int_0^{\frac{f_Y(v)}{c}} du\,dv = \dfrac{p(Y \le y)}{c}$

where $c \geq \sup_x f_Y(x)$.

Then let $\eta = b$, since $Y \in [a,b]$. w.p. 1.

$\therefore \frac{1}{c} = P(U \leq \frac{1}{c} f_Y(V))$

$\therefore P(Y \leq \eta) = \dfrac{P(V \leq \eta,\ U \leq \frac{1}{c} f_Y(V))}{P(U \leq \frac{1}{c} f_Y(V))} = P(V \leq \eta \mid U \leq \frac{1}{c} f_Y(V))$

$\therefore V \mid U \leq \frac{1}{c} f_Y(V) \sim Y$.

Remark: The optimal choice of c is $\sup_x f_Y(x)$.

Actually, N = number of (U,V) generate one Y

$\sim \exp(\frac{1}{c})$, since $\frac{1}{c} = P(U \leq \frac{1}{c} f_Y(V))$

③ The Accept/Reject

Algorithm:

- Note that in indirect method, it's wasteful in the area $U > \frac{1}{c} f_Y(V)$. Actually step 2°) is a testing to whether V looks like it's from density $f_Y(\eta)$

$\Rightarrow$ A generalization:

Suppose $V \sim f_V(\eta)$, has same support of $f_Y(\eta)$.

If $M = \sup_\eta \dfrac{f_Y(\eta)}{f_V(\eta)} < \infty$, we can compare $U \sim U(0,1)$

to $\dfrac{1}{M} \dfrac{f_Y(V)}{f_V(V)}$ to check how much V looks like Y.

Step.

1°) Generate $U \sim U[0,1]$. $V \sim f_V$, indept. (V is called candidate r.v.)

$supp f_V = supp f_Y$. $M = \sup_{\eta} \frac{f_Y(\eta)}{f_V(\eta)} < \infty$.

2°) If $U < \frac{1}{M} \frac{f_Y(V)}{f_V(V)}$. Set $Y = V$, otherwise return to 1°)

Pf: $$P(V \le \eta \mid U \le \frac{1}{M}\frac{f_Y(V)}{f_V(V)}) = \frac{P(V \le \eta, U \le \frac{1}{M}\frac{f_Y(V)}{f_V(V)})}{P(U \le \frac{1}{M}\frac{f_Y(V)}{f_V(V)})}$$

$$= \frac{\int_{-\infty}^{\eta}\int_0^{\frac{1}{M}\frac{f_Y(v)}{f_V(v)}} du\, f_V(v)\, dv}{\int_{\mathbb{R}}\int_0^{\frac{1}{M}\frac{f_Y(v)}{f_V(v)}} du\, f_V(v)\, dv} = \int_{-\infty}^{\eta} f_Y(v)\,dv = P(Y \le \eta)$$

Rmk: Suppose $Y, V \in [a,b]$. W.p.1. Set $\eta = b$.

$\Rightarrow P(U \le \frac{1}{M}\frac{f_Y(V)}{f_V(V)}) = \frac{1}{M}$. $\therefore N \sim \exp(\frac{1}{M})$

We can let M be small to make the algorithm more efficient.

④ MCMC: Gibbs Sampler and Metropolis Algorithm:

. Note that when Y has heavy tail, the method above can't be applied any more. So we introduce Markov Monte Carlo Method!