30 + 25 + 44 + 15 + 19 + 13 + 72 = 218

Generating Random Variables

MC integration

iid sample $X_1, \cdots, X_n$ from $f(X;\theta)$, approx $\mu = E(h(X))$ by sample average $\hat{\mu}_{MC} = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \to \int h(x)f(x)dx = \mu$ (based on LLN)

$\hat{\sigma}_{MC}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(h(X_i) - \hat{\mu}_{MC})^2 \longrightarrow \sigma^2$

Prob inverse transform

X has a continuous cdf $F_X(x) = Y$, then Y is uniformly distributed on $(0,1)$. $P(Y \le y) = y$, $0<y<1$.

$X = F_X^{-1}(Y)$

Revision.

Thm 1: Z, $U = Z^2 \sim \chi_1^2$

Thm 2: $U_1, \cdots, U_n$ iid & $U_i \sim \chi_1^2$ then $\sum_{i=1}^{n} U_i \sim \chi_n^2$

Thm 3: Z, $U \sim \chi_n^2$. $Z \perp U \Rightarrow T = Z/\sqrt{U/n} \sim t_n$

Thm 4: $U \sim \chi_m^2$, $V \sim \chi_n^2$, $U \perp V \Rightarrow W = \frac{U/m}{V/n} \sim F(m,n)$

Thm 5: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2) \Rightarrow$ 1. $\bar{X} \sim N(\mu, \sigma^2/n)$ 2. $\bar{X}$ & S^2 are independent 3. $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

Thm 6: $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

unbiasedness: $\hat{\theta} = T(X_1, \cdots, X_n)$. $E[T(X)] = \theta$, $bias(\hat{\theta}) = E[T(X)] - \theta$

MSE $MSE[\hat{\theta}] = E[(\hat{\theta}-\theta)^2] = V(\hat{\theta}) + Bias(\hat{\theta})^2$

consistency: $\hat{\theta}$ is weakly consistent if $P(|\hat{\theta}-\theta| > \varepsilon) \longrightarrow 0$ as $n \to \infty$ $\forall \varepsilon > 0$.

By Chebyshev's ineq. $P(|\hat{\theta}-\theta|>\varepsilon) \le \frac{1}{\varepsilon^2}[V(\hat{\theta}) + bias(\hat{\theta})^2]$. Thus $V(\hat{\theta}) \to 0$ & $bias(\hat{\theta}) \to 0 \Rightarrow$ consistency.

MVUE (minimum variance unbiased estimator) for $\tau(\theta)$. If unbiased $E[T^*] = \tau(\theta)$ & any other T with $E[T] = \tau(\theta)$ we have $V(T^*) \le V(T)$ $\forall \theta$.

(Also, it is the most efficient one!)

Cramér-Rao Ineq. [lower bound] r.v. $X_1, \cdots, X_n$ w.r.t. $f_X(x;\theta)$ where θ is a scalar para. let $T = t(X_1, \cdots, X_n)$ be an unbiased est. for $\tau(\theta)$, then under certain regularity conditions:

$$Var(T) \ge \frac{\{\tau'(\theta)\}^2}{ni(\theta)} = \{\tau'(\theta)\}^2 I(\theta)^{-1}$$

- $ni(\theta) = I(\theta)$: expected Fisher Information.
- $I(\theta) = E[(\frac{\partial l(\theta)}{\partial \theta})^2] = -E[\frac{\partial^2 l(\theta)}{\partial \theta^2}]$

Cramer-Rao Inequality Extended. $V[T(X)] \ge \frac{[\frac{\partial}{\partial\theta}E[T(X)]]^2}{E[(\frac{\partial}{\partial\theta}\log f(x|\theta))^2]} = \frac{[\frac{\partial}{\partial\theta}E[T(X)]]^2}{I(\theta)}$

- if unbiased $E[T(X)] = \tau(\theta) \Rightarrow V[T(X)] \ge [\tau'(\theta)]^2/I(\theta)$
- iid $\Rightarrow V[T(X)] \ge [\tau'(\theta)]^2/ni(\theta)$

Regularity conditions: ① $\frac{\partial}{\partial\theta}\ln\{f(x|\theta)\}$ exists $\forall x$ & θ. ② Interchange of integration & differentiation is permissible. ③ $i(\theta) = E[[\frac{\partial}{\partial\theta}\ln\{f(x|\theta)\}]^2]$. X w.r.t. $f(x;\theta) < \infty$ $\forall \theta \in \Theta$.

Corollary (iid case): regularity conditions hold, $T(X)$ is an unbiased estimator for $\tau(\theta)$ and we have $X_1, \cdots, X_n \overset{iid}{\sim} f(x;\theta)$. $\Rightarrow V[T(X)] \ge \frac{[\frac{\partial}{\partial\theta}E[T(X)]]^2}{nE[(\frac{\partial}{\partial\theta}\log f(x;\theta))^2]} = \frac{[\tau'(\theta)]^2}{ni(\theta)} = \{\tau'(\theta)\}^2 I(\theta)^{-1}$

Fisher Information

$I(\theta) = E[(\frac{\partial}{\partial\theta}\log f(\vec{x};\theta))^2] = -E[(\frac{\partial^2}{\partial\theta^2}\log f(\vec{x};\theta))]$

one data $i(\theta)$; iid data $ni(\theta) = I(\theta)$

 # Sufficiency Principle: $T(X_1,\cdots,X_n)$ is a sufficient statistic for θ, then any inference about θ should be on the sample $\vec{X}$ only through $T(X_1,\cdots,X_n)$

 # Sufficiency: A statistic $T(X_1,\cdots,X_n)$ is sufficient for θ if the conditional distribution of sample $X_1,\cdots,X_n$ given $T(X_1,\cdots,X_n)$ does not depend on θ.

i.e. $P(X_1=x_1,\cdots,X_n=x_n \mid T(X_1,\cdots,X_n))$ does not depend on θ.

 # the factorization thm: Sps $X_1,\cdots,X_n \sim f(x;\theta)$, then $T(\vec{X})$ is a sufficient statistic for θ iff $\exists$ two non-negative function K_1 & K_2 s.t. the Likelihood $L(\theta;\vec{x})$ can be written

$$f(\vec{x};\theta)=L(\theta;\vec{x})=K_1[t(\vec{x});\theta]\cdot K_2[\vec{x}]$$

If there exist multiple θs as $\vec{\theta}$, then $\exists$ $\vec{t}(\vec{x})$ multiple statistics which are sufficient.

 # Minimal Sufficient: A sufficient statistic $T(\vec{X})$ is called a minimal sufficient statistic if, for any other sufficient statistic $T'(\vec{X})$, $T(\vec{X})$ is a function of $T'(\vec{X})$. (Not easy to find one with such definition.)

 # Lemma: Let $f(\vec{x};\theta)$ be pdf or pmf of a sample $\vec{X}$. Sps $\exists$ a function $T(\vec{X})$ s.t. for every two sample points $\vec{x}$ & $\vec{y}$ the ratio

$$\frac{L(\theta;\vec{x})}{L(\theta;\vec{y})} \text{ is constant as function of } \theta \text{ iff } T(\vec{x})=T(\vec{y})$$

or a vector

Then $T(\vec{x})$ is a minimal sufficient statistic.

· Any 1-1 function of a minimal sufficient statistic is also minimal sufficient. (so not unique)

 # Rao-Blackwell Thm: W be any unbiased estimator of $\tau(\theta)$, T be a sufficient statistic for θ. Define $\phi(T)=E[W|T]$, then $E[\phi(T)]=\tau(\theta)$, $V[\phi(T)]\leq V[W]$

If we have unbiased estimator & condition it on a sufficient statistic, our new statistic $\phi(T)$ has the same or smaller variance!

↳ also unbiased

· Note: ① $E[X]=E[E[X|Y]]$; ② $V[X]=V[E(X|Y)]+E[V(X|Y)]$

· "Better estimator", key idea is sufficiency.

 # Complete Statistics: $f_T(t;\theta)$ be a family of pdfs or pmfs for a statistic $T(\vec{x})$. The family of prob. distributions is called complete if

$$E[h(T)]=\int h(t)f_T(t)\,dt=0 \text{ for all } \theta \implies P(h(T)=0)=1 \ \forall\theta.$$

 # Lehman-Scheffe Thm: $X_1,\cdots,X_n$ r.sample from dist. with pdf $f(x;\theta)$. If $T=T(\vec{X})$ is a complete & sufficient statistic, and $\phi(T)$ is an unbiased estimator of $\tau(\theta)$, then $\phi(T)$ is the unique MVUE of $\tau(\theta)$.

 # How to find MVUEs?

① a. Find or construct a sufficient & complete statistic T

b. Find unbiased estimator W for $\tau(\theta)$

c. Compute $\phi(T)=E[W|T]$, then $\phi(T)$ is MVUE.

② b' Find a func $h(T)$ s.t. $E[h(T)]=\tau(\theta)$

c' $h(T)$ is the MVUE.

 # Exponential Families: a r.v. in k-parameter expo. family of dist. if its pdf is

$$f(x;\vec{\theta})=\exp\left(\sum_{j=1}^{k}A_j(\vec{\theta})B_j(\vec{x})+C(\vec{x})+D(\vec{\theta})\right) \text{ or } C^*(x)D^*(\vec{\theta})\exp\left(\sum_{j=1}^{k}A_j(\vec{\theta})B_j(x)\right)$$

原味果汁好喝

 # canonical form: Let $\phi=(\phi_1,\cdots,\phi_k)=A(\vec{\theta})=\{A_1(\theta),\cdots,A_k(\theta)\} \implies f(x;\theta)=\exp\left\{\sum_{j=1}^{k}\phi_i B_i(x)+C(x)+D(\phi)\right\}$ with $\theta=A^{-1}(\phi)$, $D(\phi)=D\{A^{-1}(\phi)\}$.

#expo.family sufficiency: Under regularity condition, a vector of k sufficient statistics $\vec{T}$ exists for a vector of parameters $\vec{\theta}$ iff the dist. of $\vec{X}$ belong to the k-para expo. family.

LEMMA: $\vec{T}$ is also minimal sufficient.

LEMMA: $\vec{T}$ is also complete.

$\cdot\ \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$

#Point Estimation

#MOM: $\mu_k = E_\theta(X^k)$, $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$, $k=1,\cdots,K$.

#Central moment $\mu_k' = E_\theta(\{X-E_\theta(X)\}^k)$ (raw), $\hat{\mu}_k' = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^k$ (sample), $k=2,\cdots,K$. $\mu_1=\hat{\mu}_1=\bar{x}$

↳generalization:

$$E_\theta(g_1(X)) = \frac{1}{n}\sum_{i=1}^{n} g_1(x_i), \cdots, E_\theta(g_k(X)) = \frac{1}{n}\sum_{i=1}^{n} g_k(x_i) \qquad \text{if } g_i(x)=x^i \Rightarrow \text{MOM}$$

#MLE $\hat{\theta} \to$ maximizes $L(\theta;x)=L(\theta;x_1,\cdots,x_n)=f(x_1,\cdots,x_n;\theta)$

check $l''<0$.

log-likelihood $l(\theta)=\log[L(\theta)]$, $|\theta|\to\infty$, $l(\theta)\to 0$,

Score Function is the gradient of $l(\theta)$: $U(\theta)=\frac{\partial l}{\partial \theta}=l'(\theta)$

#with right censoring $X_1,\cdots,X_m$ obs. $X_{m+1},\cdots,X_n$ censored

$$L(\theta)=\prod_{i=1}^{m} f_X(x_i;\theta)\prod_{i=(m+1)}^{n}\underbrace{(1-F_X(T;\theta))}_{P(X>T)}$$

#general checking: (3 hold)

① First order partial derivatives at $\hat{\theta}_1,\hat{\theta}_2=0$: $\frac{\partial}{\partial\theta_1}H(\theta_1,\theta_2)\Big|_{\theta_1=\hat{\theta}_1,\theta_2=\hat{\theta}_2}=0=\frac{\partial}{\partial\theta_2}H(\theta_1,\theta_2)\Big|\ \cdots$

② At least one of the second-order derivs is negative.

③ The determinant of the matrix of 2nd order partial derivs is positive.

$$\begin{vmatrix} \frac{\partial}{\partial\theta_1^2} & \frac{\partial}{\partial\theta_1\partial\theta_2} \\ \frac{\partial}{\partial\theta_2\partial\theta_1} & \frac{\partial}{\partial\theta_2^2} \end{vmatrix}_{\theta_1=\hat{\theta}_1,\theta_2=\hat{\theta}_2} > 0$$

#MLE computations

#Newton-Raphson (N-R) method (also applicable for multivariate)

① $U=l'(\theta|\vec{x})$, $H(\theta)=l''(\theta|\vec{x})$, θ_0, $\hat{\theta}$

② $U(\theta)=U(\theta_0)+(\theta-\theta_0)H(\theta_0)+\cdots$

③ At $\theta=\hat{\theta}$, $U(\hat{\theta})=0=U(\theta_0)+(\hat{\theta}-\theta_0)H(\theta_0)+\cdots$

④ $\hat{\theta}=\theta_0-H^{-1}(\theta_0)U(\theta_0)$

⑤ update: $\theta_1=\theta_0-H^{-1}(\theta_0)U(\theta_0)\ \cdots$

⑥ until θ_k converges to θ_{real}.

#Method of "scoring"

Hessian $H(\vec{\theta})$ is replaced by its expectation. $E[H(\vec{\theta})]=-I(\vec{\theta})$

so $\vec{\theta}_{t+1}=\vec{\theta}_t+I^{-1}(\vec{\theta}_t)U(\vec{\theta}_t)$