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:

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

- $ni(\theta)=I(\theta)$: 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 $X_1,\cdots,X_n\overset{iid}{\sim}f(x;\theta)$. $\Rightarrow V[T(X)]\ge\frac{[\frac{\partial}{\partial\theta}E[T(X)]]^2}{nE[(\frac{\partial}{\partial\theta}\log f(x;\theta))^2]}=\frac{[\tau'(\theta)]^2}{ni(\theta)}=\{\tau'(\theta)\}^2I(\theta)^{-1}$

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. $P(X_1=x_1,\cdots,X_n=x_n|T(X_1,\cdots,X_n))$ 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

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

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 $E[\phi(T)]=\tau(\theta)$, $V[\phi(T)]\le V[W]$

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

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

Lehman-Scheffe Thm: $X_1,\dots,X_n$ 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

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

原味果汁好喝

canonical form: Let $\phi=(\phi_1,\cdots,\phi_k)=A(\vec{\theta})=\{A_1(\theta),\cdots,A_k(\theta)\}\Rightarrow f(x;\theta)=\exp\{\sum_{j=1}^{k}\phi_iB_i(x)+C(x)+D(\phi)\}$ 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 or 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.

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

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:

$E_\theta(g_1(X))=\frac{1}{n}\sum_{i=1}^{n}g_1(x_i),\cdots,E_\theta(g_k(X))=\frac{1}{n}\sum_{i=1}^{n}g_k(x_i)$ if $g_i(x)=x^i\Rightarrow$ 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 $\hat{\theta}_1,\hat{\theta}_2=0$: $\frac{\partial}{\partial\theta_1}H(\theta_1,\theta_2)\Big|_{\theta_1=\hat{\theta}_1,\theta_2=\hat{\theta}_2}=0=\frac{\partial}{\partial\theta_2}H(\theta_1,\theta_2)\Big|\dots$

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

EM: $E\{l_{comp}(\theta; y_{obs}, y_{miss}) | \theta^{(r)}, y_{obs}\} = \int [l \cdots] k(y_{miss} | y_{obs}, \theta)\, dy_{miss}$ M step: maximize previous value w.r.t. $\vec{\theta}$, set $\vec{\theta}^{(r+1)}$ to be the maximizer.
invariance property if $\hat{\theta}$ is MLE of $\hat{\theta}$, $\hat{\eta} = g(\hat{\theta})$ is MLE of η, when $\eta = g(\theta)$, $\theta = h(\eta)$, g & θ are 1-1 functions

MLE asymptotics: ① $W = \frac{1}{\sqrt{n}} l'(\theta; \vec{x}) \xrightarrow{D} N(0, i(\theta))$ ② $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, i(\theta)^{-1})$ ③ $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2) \Rightarrow \sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2 [g'(\theta)]^2)$ ④ $\tau(\theta)$ is a cont. $\tau(\hat{\theta}) \sim N(\tau(\theta), \frac{[\tau'(\theta)]^2}{I(\theta)})$ # Type I error: Reject H_0 when it's true. := in rejection region under H_0. # Type II error: do not reject H_0 when it's false. fall outside CR under H_1. # α: prob(type I error) # Power $\eta(\theta) = 1 - \beta = 1 - P(\text{Type II error}) = P(X \in C | H_1)$ P value $= P(T(X) \geq t(x) | H_0)$

Neyman-Pearson. $H_0: \theta = \theta_0$, $H_1: \theta = \theta_1$, $\lambda(x) = \frac{L(\theta_0; x)}{L(\theta_1; x)}$. test $C = \{\lambda(\vec{x}) \leq k\}$ For a given α, we could compare the power $\eta(\theta) \Rightarrow$ find a UMP test s.t. $\eta(\theta) \geq \eta(\theta^*)$ NP $\rightarrow$ UMP. # MLRT: not UMP. $\lambda(\vec{x}) = \frac{\max_{\theta \in \omega} L(\theta; \vec{x})}{\max_{\theta \in \Omega} L(\theta; \vec{x})}$, test ...

LRT asymptotics Thm: under H_0, $n \to \infty$, $-2\log[\lambda(\vec{x})] \xrightarrow{D} \chi_1^2 \Rightarrow$ find with $P(-2\log(\lambda) > k^*) = 0.05$
when more than 1 constraints $-2\log(\lambda) \xrightarrow{D} \chi_\nu^2$ # The Score Test $\vec{u}(\vec{\theta}) = (\frac{\partial l}{\partial \theta_1}, \cdots \frac{\partial l}{\partial \theta_k})^T$, $H_0: \vec{\theta} = \vec{\theta}_0$ vs $H_1: \Omega - \{\vec{\theta}_0\}$, TS: $\vec{u}(\vec{\theta})^T I_{\theta_0}^{-1} \vec{u}(\vec{\theta}) \sim \chi_k^2$ # Wald Test: TS $(\hat{\vec{\theta}} - \vec{\theta}_0)^T I_{\hat{\theta}} (\hat{\vec{\theta}} - \vec{\theta}_0) \sim \chi_k^2$
Hypo testing & interval est. $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$, test $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$. $C = \{|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}| \geq z_{\frac{\alpha}{2}}\} \Rightarrow P(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\frac{\alpha}{2}}) = 1 - \alpha \Rightarrow P(\ \leq \mu \leq\ ) = 1 - \alpha$

pivotal quantity: $P(g_1 \leq g(X, \theta) \leq g_2) = 1 - \alpha$, $P(\theta_1(X) \leq \theta \leq \theta_2(X)) = 1 - \alpha$, $\hat{\theta} \sim N(\theta, I(\theta)^{-1})$, $\frac{\hat{\theta} - \theta}{1/\sqrt{I(\hat{\theta})}} \sim N(0,1)$, asymptotic $[\hat{\theta} - z_{\frac{\alpha}{2}} \frac{1}{\sqrt{I(\hat{\theta})}}, \hat{\theta} + z_{\frac{\alpha}{2}} \frac{1}{\sqrt{I(\hat{\theta})}}]$
Similarly $\tau(\hat{\theta}) \sim N(\tau(\theta), \frac{[\tau'(\theta)]^2}{I(\theta)})$, $\frac{\tau(\hat{\theta}) - \tau(\theta)}{\sqrt{\frac{(\tau'(\theta))^2}{I(\theta)}}} \sim N(0,1) \Rightarrow$ asymptotic $[\tau(\hat{\theta}) - z_{\frac{\alpha}{2}} \frac{\tau'(\hat{\theta})}{\sqrt{I(\hat{\theta})}}, \tau(\hat{\theta}) + z_{\frac{\alpha}{2}} \frac{\tau'(\hat{\theta})}{\sqrt{I(\hat{\theta})}}]$

Bayesian $p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta)$ = likelihood × prior. Posterior mean of θ (mean of function $\tau(\theta)$: $\hat{\theta}_B = E[\theta|x] = \int_\Theta \theta p(\theta|x) d\theta$, $\hat{\tau}(\theta)_B = E[\tau(\theta)|x] = \int_\Theta \tau(\theta) p(\theta|x) d\theta$
conjugate. prior & posterior in the same class. Some common pairs (likelihood, posterior/prior) (N, N), (expo/Pois/Gamma/χ^2, Gamma), (Geom/Bino, Beta).
Bayesian testing: reject H_0 if $P(\theta \in \Theta_1 | x) > P(\theta \in \Theta_0 | x)$
Bayes factor & testing: $X_1, \cdots, X_n \stackrel{iid}{\sim} N(\theta, \sigma^2)$, $\theta \sim N(\mu, \tau^2)$, $H_0: \theta \leq \theta_0$ vs $H_1: \theta > \theta_0$ $[\theta|x] \sim N(\frac{\sigma^2 \mu + n\tau^2 \bar{x}}{\sigma^2 + n\tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2})$. Check $\int_{-\infty}^{\theta_0} p(\theta|x) d\theta$ vs $\int_{\theta_0}^{\infty} p(\theta|x) d\theta$
Take $\beta = 0$ as Bayes factors. Compare M_1 vs M_2 M_1: $y_i = \alpha + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$, $\theta_1 = \{\alpha, \sigma^2\}$, M_2 $y_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_i \sim N(0, \sigma^2)$. $\theta_2 = \{\alpha, \beta, \sigma^2\}$
posterior $\pi(M_i|x) = \frac{f(x|M_i)\pi(M_i)}{m(x)}$, $f(x|M_i) = \int_\theta f(x|M_i)\pi(\theta|M_i) d\theta_i$, $m(x) = \sum_{i=1}^2 f(x|M_i)\pi(M_i)$, $\frac{\pi(M_2|x)}{\pi(M_1|x)} = \frac{f(x|M_2)}{f(x|M_1)} \times \frac{\pi(M_2)}{\pi(M_1)} = BF(M_2; M_1) \times \frac{\pi(M_2)}{\pi(M_1)}$ (if $BF > 3.2$ reject H_0)
Bayesian Interval Estimation Find C s.t. $P_\pi(\theta|x)(C) = \int_C \pi(\theta|x) d\theta = 1 - \alpha$.
① Equal-tailed $\int_{-\infty}^{\theta_L} \pi(\theta|x) d\theta = \frac{\alpha}{2}$, $\int_{\theta_u}^{\infty} \pi(\theta|x) d\theta = \frac{\alpha}{2}$ ② Smallest length: minimize $\theta_u - \theta_L$ ③ HPD, $\theta_1 \in C$ & $\pi(\theta_2|x) > \pi(\theta_1|x) \Rightarrow \theta_2 \in C$
Bayesian sufficient: $T(\vec{X})$ is sufficient for θ iff $\theta|\vec{X}$ is the same as $\theta|T(\vec{X})$.
Risk function $R(\theta, \delta(\vec{x})) = \int L_s(\theta, \delta(\vec{x})) L(\theta, \vec{x}) d\vec{x}$. inadmissible δ_1, $\exists \delta_2$ s.t. $R(\theta, \delta_1) \geq R(\theta, \delta_2)\ \forall \theta$.
· minimax: minimize the $\max_\theta R(\theta, \delta)$.
Bayes Risk: $\int R(\theta, \delta) p(\theta) d\theta$ minimized $\Rightarrow \int_\chi p(x) [\int_\Theta L_s(\theta, \delta) p(\theta|x) d\theta] dx$ minimize posterior expected loss $\hat{\theta}_{min} = \bar{\theta}$
Simple loss functions: ① zero-one loss $L_s(\theta, \delta = \hat{\theta}) = \begin{cases} 0 & \text{when } |\hat{\theta} - \theta| < b \\ a & \text{when } |\hat{\theta} - \theta| \geq b \end{cases}$ $a, b > 0$ ② Absolute error loss: $L_s(\theta, \delta = \hat{\theta}) = a|\hat{\theta} - \theta|$, $\hat{\theta}_{min}$ = median ③ Squared-error loss $L_s(\theta, \delta = \hat{\theta}) = a(\hat{\theta} - \theta)^2$

Non-parametric
Empirical dist: $\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I_{(x_i \leq x)}$, $F(x) = P(X \leq x)$. $E[\hat{F}(x)] = \frac{1}{n} \sum_{i=1}^n F(x) = \frac{n_x}{n}$ ← # of observed less than or equal to x. $V(\hat{F}) = V(\frac{n_x}{n}) = \frac{1}{n^2} V(n_x) = \frac{1}{n^2} np(1-p) = \frac{1}{n} F(x)(1 - F(x))$
Bootstrap dist: $\hat{Bias}_B = E_{\hat{F}}\{\theta(\hat{F}^*)\} - \theta(\hat{F})$, $\hat{V}_B\{\theta(\hat{F}^*)\} = E_{\hat{F}}\{\theta(\hat{F}^*)^2\} - [E_{\hat{F}}\{\theta(\hat{F}^*)\}]^2$. After resampling: estimate $\hat{\theta}_b = \theta(\hat{F}_b^*) \Rightarrow$
$\hat{B}_B \approx \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* - \hat{\theta}$, $\hat{V}_B\{\theta(\hat{F}^*)\} \approx \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*)^2$, $\hat{\sigma}_B^2 = \sqrt{\frac{1}{B-1}(\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2}$ Bootstrap interval est. $[\hat{\theta} \pm z_{\frac{\alpha}{2}} \hat{\sigma}_B(\hat{\theta})]$, for small sample, use t-dist quant. instead of normal.
BS test: Draw B samples of size $N = n + m$ with replacement from $z = \{y_1, \cdots, y_n, x_1, \cdots, x_m\}$, the first n obs. $:= \vec{y}^*$, remaining m obs $:= \vec{x}^*$. Evaluate TS $t(\vec{y}^*, \vec{x}^*) = \bar{y}^* - \bar{x}^*$.
Approx. p-value $= \#\{t(\ ) \geq t_{obs}\}/B$.
Inverse CDF sample generating: ① $Y = -\beta \sum_{j=1}^a \log(U_j) \sim gamma(a, \beta)$ ② $\beta = 2$, $Y = -2\sum_{j=1}^\nu \log(U_j) \sim \chi_{2\nu}^2$. ③ $Y = \frac{\sum_{j=1}^a \log(U_j)}{\sum_{j=1}^{a+b} \log(U_j)} \sim Beta(a, b)$
χ_1^2? $N(0,1)$? Box-Muller Algorithm: $U_1, U_2 \sim Unif(0,1)$, set $R = \sqrt{-2\log(U_1)}$, $\theta = 2\pi U_2$, $X = R\cos(\theta)$, $Y = R\sin(\theta)$, then $X, Y \stackrel{iid}{\sim} N(0,1)$. χ_1^2? sample X^2, Y^2.
Discrete dist. generating: $y_1 < \cdots < y_k$. $P[F_Y(y_i) < U \leq F_Y(y_{i+1})] = P(Y = y_{i+1})$. stepwise function.
Accept/Reject Algorithm: $Y \sim f_Y(y)$, $V \sim f_V(v)$, densities have common support & $M = \sup_y \frac{f_Y(y)}{f_V(y)} < \infty$. Want to sample from Y, able to sample from V.
① Generate $U \sim U(0,1)$, $V \sim f_V$ independently ② if $U < \frac{1}{M} \frac{f_Y(V)}{f_V(V)}$, set $Y = V$, o.w. return to step ① NB: envelope $= M f_V(v) \geq f_Y(v)$
Metropolis Hastings, $Y \sim f_Y(y)$, $Y^* \sim f_V(v)$ have the same support. Then to generate $Y \sim f_Y$.
① set $Z_0 = c$, any starting value, could draw a Y^* from $f_V(v)$
② For $i = 1, \cdots$:

(2.1) Generate $Y_i^* \sim f_V$ and $U_i \sim U(0,1)$ and calculate:

$$\rho_i = \min\{\underbrace{\frac{f_Y(Y_i^*)}{f_Y(Z_{i-1})} \times \frac{f_V(Z_{i-1})}{f_V(Y_i^*)}}_{r}, 1\}$$

(2.2) Set

$$Z_i = \begin{cases} Y_i^* & \text{if } U_i \leq \rho_i \\ Z_{i-1} & \text{if } U_i > \rho_i \end{cases}$$

as $i \to \infty$, $Z_i \xrightarrow{D} Y$
· intuition, $r > 1$, accept Y^*
$r \leq 1$, accept Y^* at rate r.
· common symmetric proposal dist. $f_V(Y_i^* | Z_{i-1}) = U(Z_{i-1} - \delta, Z_{i-1} + \delta)$
$f_V(Y_i^* | Z_{i-1}) = N(Z_{i-1}, \sigma)$
σ & δ are customized size of "jump".

Example of Decision Theory δ_3: $L_s(\theta, \delta_3) \geq L_s(\theta, \delta_j)$ $j = 1, 2$. $\forall \theta \Rightarrow \delta_3$ is inadmissible.

	θ_1	θ_2	θ_3
δ_1	-10	-5	-3
δ_2	-5	-5	-1
δ_3	1	0	-1

$\pi(\theta_1) = 0.2$, $\pi(\theta_2) = 0.3$, $\pi(\theta_3) = 0.5$
minimize $\sum_{i=1}^3 L_s(\theta_i, \delta) p(\theta_i)$, pick the minimum.
(no data) $\min_\delta \max_\theta R(\theta, \delta) = \min_\delta \max_\theta L_s(\theta, \delta)$