#1

$(P_\theta \ \theta \in \Theta)$ $X_1, \cdots X_n \overset{iid}{\sim} \mathbb{P}_\theta$

① sufficiency & exponential family

$\downarrow$

not losing info

- factorization
- minimal sufficiency
- ancillary statistics
- completeness
- Rao - Blackwell thm

② decision theory

- loss $\ell(\hat{\theta}, \theta)$
- risk $E\,\ell(\hat{\theta}, \theta)$
- Bayes & minimax optimality
- admissibility
- James - Stein estimator (application in adaptive nonparametric estimation)
- Neyman - Pearson lemma
- minimax lower bound via Le Cam two-point method

③ estimation under constraint

- unbiasedness (UMVUE Lehman - Scheffe)
- invariance (location family, Pitman estimator)

④ likelihood & asymptotic theory

- consistency of MLE
- Fisher info & score
- LAN & DQM

Cramer Rao lower bound

Hodges estimator

Convolution theorem & local asymptotic minimaxity

Bernstein von-Mises theorem

books: ① E. Lehmann & G. Casella Theory of point estimation (pt. 1~2, 3) (4)

② E. Lehmann & J. Romano Testing statistical hypothesis (not too much)

③ I. Johnstone

Gaussian estimation:
Sequence and wavelet models

Very important book. Gaussian sequence model

④ A. van der Vaart Asymptotic Statistics important. (pt. 4)

Statistical model / experiment $(P_\theta : \theta \in \Theta)$

data / observations $X_1 \cdots X_n \overset{iid}{\sim} P_\theta$

Statistic: $T = T(X_1 \cdots X_n)$

Def 1: T is sufficient iff the conditional distribution of $X \mid T$ does not depend on θ $\forall \theta \in \Theta$

Learning X given T does not give further info on θ

Alice Bob

$X_1 \cdots X_n$ $T = T(X_1 \cdots X_n)$

(sufficient)

Bob Strategy: Sample $\tilde{X}_1\ \tilde{X}_2 \cdots \tilde{X}_n$ from the conditional distribution of $X \mid T$

can do b.c. definition

$$(\tilde{X}_1\ \tilde{X}_2, \cdots \tilde{X}_n) \overset{d}{=} (X_1 \cdots X_n)$$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1)$ $\quad T(X) = \bar{X}$

$\boxed{X_i - \bar{X} \mid \bar{X} \overset{d}{=} X_i - \bar{X}}$

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \Big| \bar{X} \sim N\left(\begin{pmatrix} \bar{X} \\ \bar{X} \\ \vdots \\ \bar{X} \end{pmatrix}, \underbrace{I_n - \frac{1}{n} 1_n 1_n^T}_{} \right) \qquad 1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$E(\bar{X} \mid \bar{X}) = \bar{X}$$

$$\downarrow$$

$$\frac{1}{n} \sum_{i=1}^{n} X_i$$

from

$$\begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & & \\ -\frac{1}{n} & 1-\frac{1}{n} & & \\ & & \ddots & \\ & & & 1-\frac{1}{n} \end{pmatrix}$$

Bob can sample $\begin{pmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_n \end{pmatrix}$

$$E(\tilde{X}_1) = E(E(\tilde{X}_1 \mid \bar{X})) = E\bar{X} = \theta$$

$$E(\tilde{X}_1^2) = E(E(\tilde{X}_1^2 \mid \bar{X})) = E(1 - \frac{1}{n} + \bar{X}^2)$$

$$= 1 - \frac{1}{n} + E(\bar{X})^2 + Var(\bar{X})$$

$$= 1 - \frac{1}{n} + \theta^2 + \frac{n}{n^2}$$

$$= 1 + \theta^2$$

$$Var(\tilde{X}_1) = E(\tilde{X}_1^2) - E(\tilde{X}_1)^2 = 1 + \theta^2 - \theta^2 = 1$$

$$E(\tilde{X}_1 \tilde{X}_2) = E(E(\tilde{X}_1 \tilde{X}_2 \mid \bar{X})) = E(-\frac{1}{n} + \bar{X}^2) = E(-\frac{1}{n} + \frac{1}{n} + \theta^2)$$

$$= \theta^2$$

$$Cov(\tilde{X}_1, \tilde{X}_2) = E(\tilde{X}_1 \tilde{X}_2) - E(\tilde{X}_1)E(\tilde{X}_2) = \theta^2 - \theta^2 = 0$$

$$\begin{pmatrix} \tilde{X}_1 \\ \vdots \\ \tilde{X}_n \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \vdots \\ \theta \end{pmatrix}, I_n \right)$$ so same distr.

e.g. $X_1 \cdots X_n \overset{iid}{\sim}$ Bernoulli(θ) $\quad T(x) = \sum_{i=1}^n X_i$

$$\mathbb{P}(X=x \mid T=t) = \frac{\mathbb{P}(X=x, T=t)}{\mathbb{P}(T=t)}$$

$$\mathbb{P}(X=x, T=t) = \begin{cases} \mathbb{P}(X=x) & \sum_{i=1}^n x_i = t \\ 0 & \sum_{i=1}^n x_i \neq t \end{cases}$$

$$= \mathbb{1}\{\sum_{i=1}^n x_i = t\} \, \mathbb{P}(X=x)$$

$$= \mathbb{1}\{\sum_{i=1}^n x_i = t\} \prod_{i=1}^n \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \mathbb{1}\{\sum_{i=1}^n x_i = t\} \, \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i}$$

$$= \mathbb{1}\{\sum_{i=1}^n x_i = t\} \, \cancel{\theta^t (1-\theta)^{n-t}}$$

$$\mathbb{P}(T=t) = \binom{n}{t} \cancel{\theta^t (1-\theta)^{n-t}}$$ cancelled out

$$\mathbb{P}(X=x \mid T=t) = \mathbb{1}\{\sum_{i=1}^n x_i = t\} \frac{1}{\binom{n}{t}}$$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} P_\theta$ $\quad T = \underbrace{(X_{(1)} \leq X_{(2)} \leq \cdots X_{(n)})}_{\text{order statistic}}$

don't know which is which

order statistic is sufficient!

$\underbrace{X_1 \cdots X_n \mid X_{(1)} \cdots X_{(n)}}_{\text{uniform } \frac{1}{n!}}$ # permutation is $n!$

uniform $\frac{1}{n!}$ (not depend on θ)

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \text{unif}(0, \theta)$ $\quad T = \max_{1 \le i \le n} X_i = X_{(n)}$ is sufficient

$X_{(1)} \cdots X_{(n-1)} \mid X_{(n)} = t$

are order-statistics from $n-1$ iid samples from unif $(0, t)$

(hw)

$X \mid T \quad \mathbb{P}(X \in A \mid T = t) = f_A(t)$

$\mathbb{P}(X \in A) = \int f_A(t)\, d\mathbb{P}_T(t)$

distri. of T

Shall we always use sufficient statistic?

Information-theoretic perspective yes

computational perspective sometimes no

Sampling $\tilde{X} \sim X \mid T$ can be NP hard

(Montanari 2015, Bresler, Shah, & ... 2014)

#2 Lecture 2

$(\mathbb{P}_\theta \quad \theta \in \Theta) \qquad X_1 \cdots X_n \overset{iid}{\sim} \mathbb{P}_\theta$

(Bayesian definition)

$\underline{\theta \to X \to T \text{ always!}}$

T is sufficient iff $\theta \to T \to X$ is Markov chain

$(\underbrace{\theta}_{\text{has a distri.}} \perp\!\!\!\perp X \mid T)$

Theorem (factorization) Suppose $(P_\theta : \theta \in \Theta)$ is continuous or discrete (has pdf or pmf)

then T is sufficient $\iff \mathbb{P}(x \mid \theta) = g_\theta(T(x))\, h(x)$ for some g_θ and h $\quad \forall \theta \in \Theta$

pf) (discrete)

$(\Leftarrow)$ Assume $\mathbb{P}(x|\theta) = g_\theta(T(x))h(x)$

$$\mathbb{P}(X=x \mid T=t) = \frac{\mathbb{P}(X=x,\, T(x)=t)}{\mathbb{P}(T(x)=t)} = \frac{1_{T(x)=t}\, \mathbb{P}(X=x)}{\sum_{T(x)=t} \mathbb{P}(x|\theta)} = \frac{1_{T(x)=t}\, \cancel{g_\theta(t)}\, h(x)}{\sum_{T(x)=t} \cancel{g_\theta(t)}\, h(x)}$$

$(\Rightarrow)$ $\quad X|T$ does not depend on θ

$$P(x|\theta) = \mathbb{P}_\theta(X=x) = \underbrace{\mathbb{P}_\theta(X=x \mid T(x)=T(x))}_{\text{does not depend on } \theta}\, \mathbb{P}_\theta(T(x)=T(x))$$

$$= h(x)\, g_\theta(T(x))$$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1)$

$$P(x|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i-\theta)^2\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum x_i^2 - \frac{1}{2}n\theta^2 + \theta\sum_{i=1}^n x_i\right)$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum x_i^2\right) \exp\left(-\frac{1}{2}n\theta^2 + \theta n \bar{X}\right) \Rightarrow \bar{X} \text{ is sufficient}$$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \text{unif}(0, \theta)$

$$p(x|\theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta} 1_{\{0<x_i<\theta\}} \right)$$

$$= \theta^{-n} \prod_{i=1}^{n} 1_{\{0<x_i<\theta\}}$$

$$= \theta^{-n} 1_{\{0<\min_i x_i, \max_i x_i < \theta\}}$$

$$= \left(1_{\{0<\min x_i\}}\right)\left(1_{\{\max x_i<\theta\}} \theta^{-n}\right)$$

$\Rightarrow \max_{1\le i\le n} x_i$ is sufficient

exponential family

$$p(x|\theta) = \exp\Big(\sum_{j=1}^{d} \underbrace{\eta_j(\theta)}_{\text{natural parameter}} \underbrace{T_j(x)}_{\text{sufficient statistic}} - \underbrace{B(\theta)}_{\text{normalizing factor "log partition function"}} \Big) \underbrace{h(x)}_{\text{"base measure"}}$$

pdf or pmf

$$B(\theta) = \log \int e^{\sum_{j=1}^{d} \eta_j(\theta) T_j(x)} h(x)\, d\mu(x)$$

e.g. exponential distribution $\text{Exp}(\theta)$

$$p(x|\theta) = \theta e^{-x\theta} 1_{\{x\ge 0\}}$$

$$= \exp(-\theta x + \log\theta)\, 1_{\{x\ge0\}}$$

e.g. $N(\mu, \sigma^2)$ $\quad \theta = (\mu, \sigma^2)$

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(x-\mu)^2/\sigma^2 \right)$$

$$= \exp\left(-\frac{1}{2\sigma^2}(x^2+\mu^2-2x\mu) - \frac{1}{2}\log(2\pi\sigma^2) \right)$$

$$= \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2) \right) \cdot 1$$

$$X_1 \cdots X_n \overset{iid}{\sim} p(x|\theta) = e^{\sum_{j=1}^{d} \eta_j(\theta) T_j(x) - B(\theta)} h(x)$$

$$\Rightarrow p(x_1 \cdots x_n|\theta) = \exp\Big(\underbrace{\sum_{j=1}^{d} \eta_j(\theta)}_{\text{does not change}} \big(\underbrace{\sum_{i=1}^{n} T_j(x_i)}_{\text{becomes a sum}} \big) - nB(\theta) \Big) \prod_{i=1}^{n} h(x_i)$$

sufficient statistic is $T = \left(\sum_{i=1}^{n} T_1(x_i), \ldots, \sum_{i=1}^{n} T_d(x_i) \right)$ (always d-dimensional vector!)

Canonical form $p(x|\eta) = \exp\left(\sum_{j=1}^{d} \eta_j T_j(x) - A(\eta)\right) h(x)$

for example, $\eta_1 = -\frac{1}{2\sigma^2}$ $\eta_2 = \frac{\mu}{\sigma^2}$

$A(\eta) = \log \int e^{\sum_{j=1}^{d} \eta_j T_j(x)} h(x)\, d\mu(x)$

Is "d" our best?

Def. an exponential family $(P_\eta \;\; \eta \in H)$ (of canonical form) is

minimal if its dimension cannot be reduced

(sufficient stats are linearly indpt natural parameters are linearly indpt)

e.g. (non-minimal examples)

linearly dependent

$p(x|\eta) = \exp(\eta_1 T(x) + \eta_2 (3T(x)+2) - A(\eta))$

$= \exp((\eta_1 + 3\eta_2) T(x) + 2\eta_2 - A(\eta))$

$p(x|\eta) = \exp(\eta T_1(x) + (4-5\eta) T_2(x) - A(\eta))$

$= \exp(\eta(T_1(x) - 5T_2(x)) - A(\eta)) \exp(4T_2(x))$

Two types of minimal exponential family

① full rank : H contains an open d-dimensional rectangle

② curved : $\eta_1 \cdots \eta_d$ are related in non-linear ways

e.g. $N(\mu, \sigma^2)$ $p(x|\mu,\sigma^2) = \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)$

$\begin{cases} T_1(x) = -x^2 \\ T_2(x) = x \end{cases}$ $\begin{cases} \eta_1 = \frac{1}{2\sigma^2} \\ \eta_2 = \frac{\mu}{\sigma^2} \end{cases}$

* $\mu = \sigma^2$ $N(\sigma^2, \sigma^2)$ $\eta_2 = 1$ non-minimal

* $\mu = \sqrt{\sigma^2}$ $\begin{cases} y_1 = \frac{1}{2\sigma^2} \\ y_2 = \frac{1}{\sqrt{\sigma^2}} \end{cases}$ $\eta_2^2 = 2\eta_1$ (non-linear) minimal & curved

* μ & σ^2 do not have further constraints minimal & full rank

$H = (0,\infty) \times \mathbb{R}$

minimal sufficiency

$X_1 \cdots X_n \sim$ iid $N(0,1)$

$T_1 = (X_1 \cdots X_n)$

$T_2 = (X_1+X_2, X_3+X_4, \cdots X_{n-1}+X_n)$

$T_3 = (\sum_{i \le k} X_i, \sum_{i > k} X_i)$

$T_4 = \sum_{i=1}^{n} X_i$ (will see min suff. statistic)

Def. S is minimal sufficient iff it is sufficient & $\forall$ sufficient T, S is a function of T

Q. How to find minimal sufficient statistic?

① sub-family method

② discussed later exponential family

Lemma: Suppose $\Theta_0 \subseteq \Theta$. Suppose S is minimal sufficient for $(P_\theta : \theta \in \Theta_0)$ and sufficient for $(P_\theta : \theta \in \Theta)$

it is minimal sufficient for $(P_\theta : \theta \in \Theta)$

pf) definition

(ex) rule-out unif $(0, \theta)$

Thm for $(P_\theta : \theta \in \{\theta_0 \cdots \theta_d\})$ (common support)

$$T(x) = \left(\frac{P_{\theta_1}(x)}{P_{\theta_0}(x)}, \frac{P_{\theta_2}(x)}{P_{\theta_0}(x)}, \cdots, \frac{P_{\theta_d}(x)}{P_{\theta_0}(x)} \right)$$ is minimal sufficient

($d=1$: likelihood)

pf)
$$\begin{cases} P_{\theta_0}(x) = P_{\theta_0}(x) \\ P_{\theta_j}(x) = T_j(x) P_{\theta_0}(x) \quad j = 1 \cdots d \end{cases}$$

$$g_{\theta_j}(T(x)) = \begin{cases} 1 & j=0 \\ T_j(x) & j = 1 \cdots k \end{cases} \qquad h(x) = P_{\theta_0}(x)$$

$\Rightarrow$ T is sufficient ?

$\Rightarrow$ Say T' sufficient: $\dfrac{P(x|\theta_1)}{P(x|\theta_0)} = \dfrac{g_{\theta_1}(T'(x))}{g_{\theta_0}(T'(x))}$

$\Rightarrow$ likelihood ratio is function of T'

$\Rightarrow$ T is a function of T'

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \text{Bern}(\theta)$ $\theta \in [0,1]$

$\sum_{i=1}^{n} X_i$ is sufficient.

Lecture 3

<u>Thm</u> minimal exponential family $\exp(\langle \eta, T(x)\rangle - A(\eta))\, h(x)$

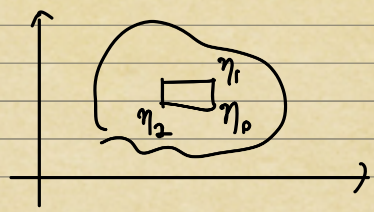
$\eta \in H \subseteq \mathbb{R}^d$ then $T(x) = (T_1(x) \cdots T_d(x))$ is minimal sufficient.

proof) Since exp. family is minimal, can find $\eta_0\ \eta_1 \cdots \eta_d \in H$

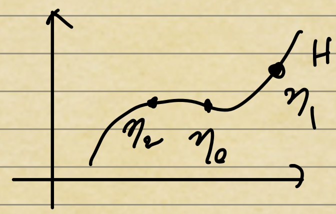
s.t. $\begin{pmatrix} (\eta_1 - \eta_0)^T \\ (\eta_2 - \eta_0)^T \\ \vdots \\ (\eta_d - \eta_0)^T \end{pmatrix}$ $d \times d$ matrix has full rank

Illustration of $d=2$

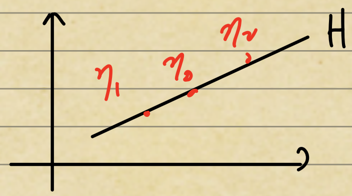
① full rank exp family $d=2$



② curved exp family



③ non-minimal exp family



a subfamily $\{\eta_0\ \eta_1 \cdots \eta_d\} \subseteq H$

minimal sufficient statistic $\dfrac{p(x|\eta_j)}{p(x|\eta_0)}$ $j = 1 \ldots d$

$$\frac{p(x|\eta_j)}{p(x|\eta_0)} = \frac{\exp(\langle \eta_j, T(x)\rangle - A(\eta_j))}{\exp(\langle \eta_0, T(x)\rangle - A(\eta_0))} = \exp(\langle \eta_j - \eta_0, T(x)\rangle - A(\eta_j) + A(\eta_0))$$

equiv. to $\langle \eta_i - \eta_0, T(x) \rangle \quad i = 1 \ldots d$

$$\Leftrightarrow \begin{pmatrix} \langle \eta_1 - \eta_0, T(x) \rangle \\ \langle \eta_2 - \eta_0, T(x) \rangle \\ \vdots \\ \langle \eta_d - \eta_0, T(x) \rangle \end{pmatrix} = \underbrace{\begin{pmatrix} (\eta_1 - \eta_0)^T \\ (\eta_2 - \eta_0)^T \\ \vdots \\ (\eta_d - \eta_0)^T \end{pmatrix}}_{\text{invertible}} T(x)$$

$\Leftrightarrow \quad T(x)$, minimal sufficient!

② completeness method (remove all ancillary information)

complete, sufficient $\Rightarrow$ minimal sufficient

e.g. $X_1, X_2 \overset{iid}{\sim} N(\theta, 1)$

$T = (X_1, X_2)$ is sufficient but not minimal

equivalent to $(\underline{\underline{X_1 - X_2}}, X_1 + X_2)$

$N(0, 2)$

"ancillary statistic"

Def. $A = A(X)$ is ancillary iff its distribution does not depend on $\theta \in \Theta$

is first-order ancillary iff its expectation ($E_\theta A(X)$) does not depend on $\theta \in \Theta$

Def. $T = T(X)$ is complete iff $E_\theta f(T(X)) = 0 \quad \forall \theta \in \Theta$ implies $f(T(X)) = 0$ a.s.

In words, no non-constant function of T is first-order ancillary

"ancillary" $\quad E_\theta f(T(X)) = c \Rightarrow f(T(X)) = c$

"complete" of $f(T(X)) - c$ $\quad E_\theta(f(T(X)) - c) = 0 \Rightarrow f(T(X)) - c = 0$

$\Downarrow$

already pared down to the essence of the data w.r.t. θ

no excess info,

ancillary "indicates" non-sufficiency

Theorem (Bahadur)

T is sufficient & complete $\Rightarrow$ T is minimal sufficient

pf) (sketch) assume minimal suff. statistic exists. $U = U(X)$

by def. $U = h(T)$

Want to show T is also a function of U.

define $g(u) = E_\theta(T \mid U = u)$ is a fn indep of θ by sufficiency of U

$E_\theta g(h(T)) = E_\theta g(U) = E_\theta(E_\theta(T|U)) = E_\theta T$

$\Rightarrow \quad E_\theta(g(h(T)) - T) = 0 \quad \forall \theta \in \Theta$

$\Rightarrow$ by completeness, $g(h(T)) = T$ a.s. $\Rightarrow$ $g(U) = T$ a.s. ∎

e.g. $X_1 \cdots X_n \overset{iid}{\sim} Ber(\theta)$ $\quad T = \sum_{i=1}^{n} X_i \sim Binom(n, \theta)$

Suppose $E_\theta f(T(X)) = 0$

$= \sum_{T=1}^{n} f(T) \binom{n}{T} \theta^T (1-\theta)^{n-T}$

$= \sum_{T=1}^{n} f(T) \binom{n}{T} \left(\frac{\theta}{1-\theta}\right)^T (1-\theta)^n = 0 \quad \forall \theta \in (0,1)$

$\Rightarrow \sum_{T=1}^{n} f(T) \binom{n}{T} \left(\frac{\theta}{1-\theta}\right)^T$

Set $\beta = \frac{\theta}{1-\theta}$ $\quad \sum_{T=1}^{n} f(T) \binom{n}{T} \beta^T = 0 \quad \forall \beta > 0$

$\Rightarrow$ at most n roots but ∞ roots $\Rightarrow$ coeff $\equiv 0$

$\Rightarrow f(T) = 0 \quad \forall T = 1 \cdots n$

e.g. $X_1 \dots X_n \overset{iid}{\sim} \mathrm{Unif}(0,\theta)$

$$T = \max_{1\le i\le n} X_i$$

$$\mathbb{P}(T\le t) = \prod_{i=1}^{n} \mathbb{P}(X_i \le t) = \left(\frac{t}{\theta}\right)^n \qquad t\in(0,\theta)$$

$$\mathbb{P}(t\mid\theta) = \frac{d}{dt}\mathbb{P}(T\le t) = \theta^{-n} n t^{n-1} \qquad t\in(0,\theta)$$

Suppose $E_\theta f(T(x)) = 0 \quad \forall\theta>0$

$$\Rightarrow \int_0^\theta f(t)\,\theta^{-n} n t^{n-1}\,dt = 0$$

$$\Rightarrow \int_0^\theta t^{n-1} f(t)\,dt = 0 \quad \forall\theta>0$$

$$\begin{cases} f^+ = \max(f,0) \\ f^- = \max(-f,0) \end{cases}$$

$$f = f^+ - f^-$$

$$\Rightarrow \int_0^\theta t^{n-1} f^+(t)\,dt = \int_0^\theta t^{n-1} f^-(t)\,dt \qquad \forall\theta>0$$

$$\Rightarrow \int_{\theta_1}^{\theta_2} t^{n-1} f^+(\theta)\,dt = \int_{\theta_1}^{\theta_2} t^{n-1} f^-(t)\,dt \qquad \forall\, 0<\theta_1<\theta_2$$

$$\Rightarrow \int_A t^{n-1} f^+(\theta)\,dt = \int_A t^{n-1} f^-(t)\,dt \quad \forall \text{ Borel } A$$

$$\Rightarrow t^{n-1} f^+(t) = t^{n-1} f^-(t)$$

$$\Rightarrow f = 0$$

$\Rightarrow T$ is complete

e.g. $X_1 \cdots X_n \overset{\text{iid}}{\sim} N(\theta, 1)$

$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i \sim N(\theta, 1)$

Suppose $E_\theta f(T(x)) = 0 \quad \forall \theta \in \mathbb{R}$

$$\int f(x) e^{-\frac{1}{2}x^2 + x\theta} = 0 \quad \forall \theta \in \mathbb{R}$$

$$\Rightarrow \int f^+ e^{-\frac{1}{2}x^2 + x\theta} dx = \int f^- e^{-\frac{1}{2}x^2 + x\theta} dx$$

$\Rightarrow$ take $\theta = 0$ $\quad \int f^+(x) e^{-\frac{1}{2}x^2} dx = \int f^-(x) e^{-\frac{1}{2}x^2} dx$

$$\Rightarrow \frac{\int f^+(x) e^{-\frac{1}{2}x^2} e^{\theta x} dx}{\int f^+ e^{-\frac{1}{2}x^2} dx} = \frac{\int f^-(x) e^{-\frac{1}{2}x^2} e^{\theta x} dx}{\int f^- e^{-\frac{1}{2}x^2} dx}$$

$E(e^{\theta x})$

MGF

$\Rightarrow f^+ = f^-$ a.e. $\Rightarrow f = 0$ a.e $\Rightarrow T$ is complete

e.g. for full rank, exponential family,

$e^{\sum_{j=1}^{d} \eta_j T_j(x) - A(\eta)} h(x) \quad h \in H \quad T = (T_1(x) \cdots T_d(x))$ is complete

Theorem (Basu) T is complete and sufficient

A is ancillary $\Rightarrow T \perp\!\!\!\perp A$

pf) Want to show $P_\theta(A \in B \mid T = t) = \mathbb{P}_\theta(A \in B) \quad \forall t$

Set $c = \mathbb{P}_\theta(A \in B)$ (does not depend on θ b/c A is ancillary)

$g(t) = \mathbb{P}_\theta(A \in B \mid T = t)$ (does not depend on θ b/c T is sufficient)

$T \in C$?

$E_\theta(g(T) - c) = E_\theta \mathbb{P}_\theta(A \in B \mid T) - \mathbb{P}_\theta(A \in B)$

$= \mathbb{P}_\theta(A \in B) - \mathbb{P}_\theta(A \in B) = 0 \quad \forall \theta \in \Theta$

$\Rightarrow$ By completeness, $g(t) = c$ a.s.

$X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1)$

$\bar{X} \perp\!\!\!\perp \sum_{i=1}^{n} (X_i - \bar{X})^2$

χ^2_{n-1}, ancillary

Lecture 4

Decision theory

$(\mathbb{P}_\theta : \theta \in \Theta)$ $\quad X_1 \dots X_n \overset{iid}{\sim} \mathbb{P}_\theta$

$\hat{\theta} = \hat{\theta}(X_1 \dots X_n)$

Loss function $L(\hat{\theta}, \theta)$ e.g. $\|\hat{\theta} - \theta\|^2$

risk $R(\hat{\theta}, \theta) = \mathbb{E}_\theta L(\hat{\theta}, \theta) = \int L(\hat{\theta}(x), \theta) P_\theta(x) dx$

<u>Theorem</u> (Rao-Blackwell) Assume $L(\hat{\theta}, \theta)$ is convex in $\hat{\theta}$

for any $\hat{\theta}$ and any sufficient T, define $\tilde{\theta} = E_\theta(\hat{\theta} | T)$

then $R(\tilde{\theta}, \theta) \leq R(\hat{\theta}, \theta)$

$\left(E((\hat{\theta}-\theta)^2 | T(x)) \geq (E(\hat{\theta} | T(x)) - \theta)^2 \right)$

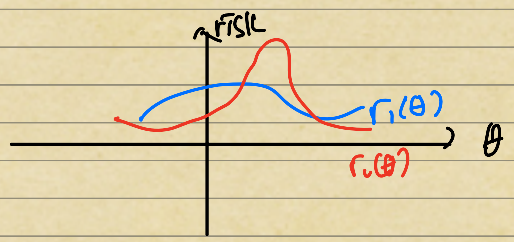
pf) $L(\tilde{\theta}, \theta) = L(E_\theta(\hat{\theta}|T), \theta)$

$\leq E_\theta(L(\hat{\theta}, \theta) | T)$ Jensen inequality

$\Rightarrow E_\theta L(\tilde{\theta}, \theta) \leq E_\theta E_\theta(L(\hat{\theta}, \theta)|T) \underset{\substack{\uparrow \\ \text{sufficiency}}}{=} E_\theta(L(\hat{\theta}, \theta))$

Two estimators $\hat{\theta}, \tilde{\theta} \rightsquigarrow r_2(\theta) = R(\tilde{\theta}, \theta)$

$r_1(\theta) = R(\hat{\theta}, \theta)$



average risk $\int R(\hat{\theta}, \theta) \pi(\theta) d\theta$ ← prior distribution

maximum risk $\sup_{\theta \in \Theta} R(\hat{\theta}, \theta)$

def : ① $\hat{\theta}$ is a Bayes estimator if $\hat{\theta} = \underset{\hat{\theta}}{\arg\min} \int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta$

$\iff \forall \tilde{\theta}, \int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta \le \int R(\tilde{\theta}, \theta)\pi(\theta)\, d\theta$

② $\hat{\theta}$ is a minimax estimator if

$$\hat{\theta} = \underset{\theta}{\arg\min} \sup_{\theta \in \Theta} R(\hat{\theta}, \theta)$$

$\iff \forall \tilde{\theta} \quad \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \le \sup_{\theta \in \Theta} R(\tilde{\theta}, \theta)$

Bayes estimator

$$\int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta = \iint L(\hat{\theta}(x), \theta)\, \underbrace{P_\theta(x)}_{p(x|\theta)}\, \pi(\theta)\, dx\, d\theta$$

↳ Joint distribution of (X, θ)

$$p(x|\theta)\pi(\theta) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\theta)\pi(\theta)\, d\theta} \int p(x|\theta)\pi(\theta)\, d\theta$$

$$= \pi(\theta|x)\, m(x)$$

$\pi(\theta|x)$ ← posterior, $m(x)$ ← marginal of X

$$\int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta = \int \boxed{\int L(\hat{\theta}(x), \theta)\, \pi(\theta|x)\, d\theta}\, m(x)\, dx$$

↳ a function of x

Claim: $\hat{\theta}_\pi(x) = \arg\min_a \int L(a, \theta)\pi(\theta|x)\, d\theta$ is Bayes

minimization over $\hat{\theta} \in \Theta$ vs. over all number

pf) WTS for any $\hat{\theta}$

$$\int R(\hat{\theta}_\pi, \theta)\pi(\theta)\, d\theta \le \int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta$$

$$\int R(\hat{\theta}_\pi, \theta)\pi(\theta)\, d\theta = \iint L(\hat{\theta}_\pi(x), \theta)\, \pi(\theta|x)\, d\theta\, m(x)\, dx$$

$$\le \iint L(\hat{\theta}(x), \theta)\, \pi(\theta|x)\, d\theta\, m(x)\, dx$$

$$= \int R(\hat{\theta}, \theta)\pi(\theta)\, d\theta$$

an important example

$\Theta \subseteq \mathbb{R}$ $\quad L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$

$$\hat{\theta}_\pi(X) = \underset{a}{\operatorname{argmin}} \int (a-\theta)^2 \pi(\theta | x) \, d\theta$$

$$= \underset{a}{\operatorname{argmin}} \, E_\theta((a-\theta)^2 | X) \Rightarrow E((\theta - E(\theta))^2 | X) + (E(\theta|X) - a)^2$$

$$= E(\theta | X)$$

$$\boxed{\begin{aligned} &\text{a r.v. } Y \in \mathbb{R} \quad \mu \in \mathbb{R} \quad E((Y-\mu)^2) = \mathrm{Var}(Y) + (EY - \mu)^2 \\ &\qquad E((a-\theta)^2 | X) = \mathrm{Var}(\theta | X) + (E(\theta|X) - a)^2 \end{aligned}}$$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \mathrm{Bern}(p)$ $\quad$ loss $(\hat{p} - p)^2$

prior $\pi = \mathrm{Beta}(\alpha, \beta)$ $\quad \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}$

$p | X_1 \cdots X_n \sim \mathrm{Beta}\left(\sum_{i=1}^n x_i + \alpha, \sum_{i=1}^n (1-x_i) + \beta\right)$

Bayes estimate $\quad \hat{p} = E(p | X_1 \cdots X_n) = \dfrac{\sum_{i=1}^n x_i + \alpha}{n + \alpha + \beta}$

$$R(\hat{p}, p) = E(\hat{p} - p)^2 = \mathrm{Var}\,\hat{p} + (E\hat{p} - p)^2$$

$$= \left(\frac{n}{n+\alpha+\beta}\right)^2 \frac{p(1-p)}{n} + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)^2 \left(\frac{\alpha}{\alpha+\beta} - p\right)^2$$

. minimax estimator $\quad \hat{\theta}_{minimax} = \underset{\hat{\theta}}{\operatorname{argmin}} \sup_{\theta \in \Theta} R(\hat{\theta}, \theta)$ "game theory"

Theorem If for some π, $\hat{\theta}$ satisfies $\sup_{\theta \in \Theta} R(\hat{\theta}, \theta) = \inf_{\tilde{\theta}} \int R(\tilde{\theta}, \theta) \pi(\theta) \, d\theta$

worst possible risk $\qquad$ best possible average risk

then $\hat{\theta}$ is minimax.

pf) $\quad \forall \tilde{\theta}, \quad \sup_{\theta \in \Theta} R(\tilde{\theta}, \theta) \geq \int R(\tilde{\theta}, \theta) \pi(\theta) \, d\theta$

$$\geq \inf_{\tilde{\theta}} \int R(\tilde{\theta}, \theta) \pi(\theta) \, d\theta$$

$$= \sup_{\theta \in \Theta} R(\hat{\theta}, \theta)$$

Corollary If $\hat{\theta} = \hat{\theta}_\pi$ for some π (Bayes estimator w.r.t. π) and $R(\hat{\theta}_\pi, \theta)$ is constant over $\theta \in \Theta$ then $\hat{\theta}$ is minimax.

function of θ ... and parameters of π make this risk constant

pf) $\sup_{\theta \in \Theta} R(\hat{\theta}, \theta) \underset{\text{constant}}{=} \int R(\hat{\theta}, \theta)\pi(\theta)d\theta$

$\underset{\hat{\theta} \text{ is Bayes}}{=} \inf_{\tilde{\theta}} \int R(\tilde{\theta}, \theta)\pi(\theta)d\theta$

by theorem $\hat{\theta}$ is minimax.

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ loss $(\hat{p}-p)^2$

$$\hat{p} = E(p \mid X_1 \cdots X_n) = \frac{\sum_{i=1}^n X_i + \alpha}{n+\alpha+\beta}.$$

$$R(\hat{p}, p) = E_p((\hat{p}-p)^2)$$

$$= \underbrace{\left(\frac{n}{n+\alpha+\beta}\right)^2 \frac{p(1-p)}{n}}_{\text{Variance}} + \underbrace{\left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 \left(\frac{\alpha}{\alpha+\beta} - p\right)^2}_{\text{bias}}$$

$$= \underbrace{\left[\left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 - \frac{1}{n}\left(\frac{n}{n+\alpha+\beta}\right)^2 \right]}_{=0} p^2$$

$$+ \underbrace{\left[\frac{1}{n}\left(\frac{n}{n+\alpha+\beta}\right)^2 - \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)^2 \frac{2\alpha}{\alpha+\beta} \right]}_{=0} p$$

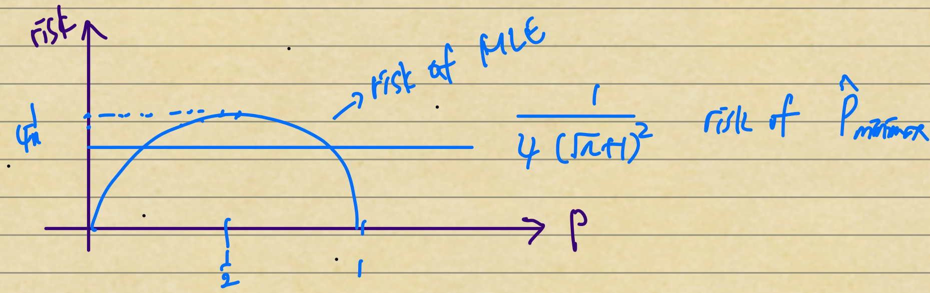
$$+ \frac{(\alpha+\beta)^2}{(n+\alpha+\beta)^2} \left(\frac{\alpha}{\alpha+\beta}\right)^2.$$

$$\left.\begin{aligned} (\alpha+\beta)^2 &= n \\ 2\alpha(\alpha+\beta) &= n \end{aligned}\right\} \Rightarrow \alpha = \beta = \frac{\sqrt{n}}{2}$$

$$\hat{p}_{\text{minimax}} = \frac{\sum_{i=1}^n X_i + \frac{\sqrt{n}}{2}}{n+\sqrt{n}} \qquad \hat{p}_{MLE} = \bar{X}$$

$$R(\hat{p}_{MLE}, p) = E_p(\hat{p}-p)^2 = \frac{p(1-p)}{n} \qquad \max_p \frac{p(1-p)}{n} = \frac{1}{4n}$$

$$R(\hat{p}_{\text{minimax}}, p) = \frac{1}{4(\sqrt{n}+1)^2} < \frac{1}{4n}$$



e.g. $X_1 \cdots X_n \overset{iid}{\sim} \text{Bernoulli}(p)$ loss $L(\hat{p}, p) = \frac{(\hat{p}-p)^2}{p(1-p)}$

$\pi(p) = 1$

$\hat{p} = \underset{a}{\arg\min} \int \frac{(a-p)^2}{p(1-p)} \pi(p|x)\, dp$

$= \underset{a}{\arg\min} \int (a-p)^2 \frac{\pi(p|x)}{p(1-p)}\, dp$

$\frac{\pi(p|x)}{p(1-p)} \propto p^{\sum X_i - 1} (1-p)^{\sum(1-X_i)-1} = \text{Beta}(\sum X_i, \sum(1-X_i))$

$\hat{p}(x) = \frac{\sum X_i}{\sum X_i + \sum(1-X_i)} = \bar{X} = \hat{p}_{MLE}$

$R(\hat{p}, p) = E_{\theta=p}\left(\frac{(\bar{X}-p)^2}{p(1-p)}\right) = \frac{1}{n}$ constant.

$\hat{p} = \bar{X}$ is minimax.

For squared error loss, Bayes estimator has to be biased

e.g. $X_1 \cdots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ $\mu \in \mathbb{R}$ loss $(\hat{\mu}-\mu)^2$

Q. is $\bar{X}$ minimax?

$R(\bar{X}, \mu) = E_x(\bar{X}-\mu)^2 = \frac{\sigma^2}{n}$ unbiased! So cannot see minimax

need new tool.

Lecture 5

$X \sim P_\theta \qquad (P_\theta : \theta \in \Theta)$

$L(\hat\theta, \theta) \qquad R(\hat\theta,\theta) = E_\theta L(\hat\theta,\theta) = \int L(\hat\theta(x),\theta)\,dP_\theta(x)$

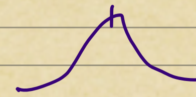
e.g. $X_1 \dots X_n \overset{iid}{\sim} N(\mu,\sigma^2)$ $\quad \mu\in\mathbb{R}$ $\quad$ loss $(\hat\mu-\mu)^2$

Q: Is $\bar X$ minimax?

$$R(\bar X,\mu) = E_\theta(\bar X-\mu)^2 = \frac{\sigma^2}{n}$$

Consider $\pi \sim N(0,\tau^2)$ $\Rightarrow$ maybe adjust τ later

unimodal



$$\pi(\mu|x) \propto \pi(\mu)\prod_{i=1}^n p(x_i|\mu) \propto e^{-\frac{\mu^2}{2\tau^2} - \frac{1}{2}\sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2}}$$

$$f(\mu) = \frac{\mu^2}{\tau^2} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{\sigma^2} \qquad f'(\mu) = \frac{2\mu}{\tau^2} + \frac{1}{\sigma^2}\sum_{i=1}^n 2(\mu - x_i) = 0$$

$$\Longleftrightarrow \frac{\mu}{\tau^2} + \frac{n\mu}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^n x_i$$

$$\Rightarrow E(\mu|x) = \frac{1}{\sigma^2}\sum_{i=1}^n x_i \Big/ \left(\frac{1}{\tau^2}+\frac{n}{\sigma^2}\right) = \boxed{\frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}}\ \bar X \neq \bar X$$

$\tau^2\to\infty$?

shrink factor

Actually it's unbiased so not Bayes for any prior

$$R(\hat\mu,\mu) = Var(\hat\mu) + E(\hat\mu-\mu)^2$$

$$= \left(\frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)^2\frac{\sigma^2}{n} + \left(\frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)^2\mu^2$$

$$\int R(\hat\mu,\mu)\,\pi(\mu)\,d\mu = \left(\frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)^2\frac{\sigma^2}{n} + \left(\frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)^2\tau^2$$

$$= \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} \rightarrow \frac{\sigma^2}{n}$$

To prove $\bar X$ is minimax $\quad \sup_{\mu\in\mathbb{R}} R(\bar X,\mu) = \frac{\sigma^2}{n}$

$\forall \hat{\mu} \quad \sup_{\mu \in \mathbb{R}} R(\hat{\mu}, \mu) \geq \int R(\hat{\mu}, \mu) \pi(\mu) d\mu$

~~To capture info. have to be spread out~~ $\quad N(0, \tau^2)$

$$\geq \inf_{\hat{\mu}} \int R(\hat{\mu}, \mu) \pi(\mu) d\mu$$

$$\overset{\text{Bayes}}{=} \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}$$

Letting $\tau \to \infty$. $\quad \sup_{\mu \in \mathbb{R}} R(\hat{\mu}, \mu) \geq \lim_{\tau \to \infty} \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}$

$$\Rightarrow \sup_{\mu \in \mathbb{R}} R(\hat{\mu}, \mu) \geq \frac{\sigma^2}{n} = R(\bar{X}, \mu)$$

$\Rightarrow \bar{X}$ is minimax

Theorem. $\exists (\pi_n)$ s.t. $\quad \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) = \lim_{n \to \infty} \inf_{\hat{\theta}} \int R(\hat{\theta}, \theta) \pi_n(\theta) d\theta$

we considered $\hat{\theta}_n$

then $\hat{\theta}$ is minimax

pf) $\sup_{\theta} R(\hat{\theta}, \theta) \leq \lim_{n \to \infty} \int R(\bar{\theta}, \theta) \pi_n(\theta) d\theta \leq R(\bar{\theta}, \theta) \ \forall \theta \leq \sup_{\theta} R(\bar{\theta}, \theta)$

Fact: $\sup_{\theta \in \Theta} R(\hat{\theta}, \theta) = \sup_{\pi} \int R(\hat{\theta}, \theta) \pi(\theta) d\theta$

Admissibility

"useless"

Def: $\hat{\theta}$ is inadmissible if $\exists \tilde{\theta}$ (estimator) s.t.

① $R(\tilde{\theta}, \theta) \leq R(\hat{\theta}, \theta) \quad \forall \theta \in \Theta$

② $R(\tilde{\theta}, \theta_0) < R(\hat{\theta}, \theta_0)$ for some θ_0 (number)

$\hat{\theta}$ is admissible if it is not inadmissible

Theorem. If $\hat{\theta}$ is Bayes, then it is admissible

pf) suppose $\hat{\theta}$ is inadmissible. $\exists \tilde{\theta}$ s.t.

① $R(\tilde{\theta},\theta) \le R(\hat{\theta},\theta) \quad \forall \theta \in \Theta$

② $R(\tilde{\theta},\theta_0) < R(\hat{\theta},\theta_0)$ for some $\theta_0 \in \Theta$

$\Rightarrow$ $\exists$ an open set $\Theta_0 \ni \theta_0$ and $\varepsilon > 0$

s.t. $R(\tilde{\theta},\theta) < R(\hat{\theta},\theta) - \varepsilon \quad \forall \theta \in \Theta_0$ (need continuity of R in θ)

$$\int R(\tilde{\theta},\theta)\pi(\theta)\,d\theta = \int_{\Theta_0} R(\tilde{\theta},\theta)\pi(\theta)\,d\theta + \int_{\Theta_0^c} R(\tilde{\theta},\theta)\pi(\theta)\,d\theta$$

$$< \int_{\Theta_0} R(\hat{\theta},\theta)\pi(\theta)\,d\theta + \int_{\Theta_0^c} R(\hat{\theta},\theta)\pi(\theta)\,d\theta$$ (π has to be non zero integrated on Θ_0 too!)

$$= \int R(\hat{\theta},\theta)\pi(\theta)\,d\theta$$

$\Rightarrow$ $\hat{\theta}$ is not Bayes

(roughly converse)

Complete class theorem (Brown, 1986)

$\hat{\theta}$ is admissible $\rightarrow$ $\exists \{\pi_n\}_n$ s.t. $\hat{\theta}_{\pi_n} \to \hat{\theta}$

$\boxed{\text{admissible} \approx \text{Bayes or close to Bayes !!}}$

e.g. $X_1 \dots X_n \overset{iid}{\sim} N(\theta, 1) \quad \theta \in \mathbb{R} \quad (\hat{\theta}-\theta)^2$

Q. Is $\bar{X}$ admissible? Yes   Wald (1939)

proof. (Blyth's method) Suppose $\hat{\theta} = \bar{X}$ is not admissible

$\exists \tilde{\theta}$ s.t. ① $R(\tilde{\theta},\theta) \le \frac{1}{n} \quad \forall \theta \in \mathbb{R}$

② $R(\tilde{\theta},\theta_0) < \frac{1}{n} \quad \exists \theta_0 \in \mathbb{R}$

$\exists\, a<b \quad \varepsilon > 0$ s.t. $(a,b) \ni \theta_0$ & $R(\tilde{\theta},\theta) < \frac{1}{n} - \varepsilon \quad \forall \theta \in (a,b)$

Consider $\pi_m = N(0,m)$ $\int R(\hat{\theta}_m,\theta)\pi_m(\theta)\,d\theta = \frac{1}{n+\frac{1}{m}}$

$\frac{1}{n} - \int R(\hat{\theta}_m,\theta)\pi_m(\theta)\,d\theta = \frac{1}{n} - \frac{1}{n+\frac{1}{m}} = \frac{1}{n}\frac{\frac{1}{m}}{n+\frac{1}{m}} \asymp \frac{1}{m}$ (n kept fixed, order)

$$\frac{1}{n} - \int R(\tilde{\theta}, \theta)\pi_m(\theta)d\theta$$

$$= \frac{1}{n} - \int_a^b R(\tilde{\theta}, \theta)\pi_m(\theta)d\theta - \int_{(a,b)^c} R(\tilde{\theta}, \theta)\pi_m(\theta)d\theta$$

$$= \underbrace{\int_{(a,b)} \left(\frac{1}{n} - R(\tilde{\theta}, \theta)\right)\pi_m(\theta)d\theta}_{>0} + \underbrace{\int_{(a,b)^c} \left(\frac{1}{n} - R(\tilde{\theta}, \theta)\right)\pi_m(\theta)d\theta}_{\geq 0}$$

$$\geq \varepsilon \int_{(a,b)} \pi_m(\theta)d\theta = \varepsilon\, \mathbb{P}(a < N(0,m) < b)$$

$$= \varepsilon\, \mathbb{P}\left(\frac{a}{\sqrt{m}} < N(0,1) < \frac{b}{\sqrt{m}}\right)$$

$$= \varepsilon \int_{\frac{a}{\sqrt{m}}}^{\frac{b}{\sqrt{m}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \gtrsim \quad \frac{1}{\sqrt{m}}$$

$\exists m$ sufficiently large s.t.

$$\cancel{\frac{1}{n}} - \int R(\tilde{\theta}, \theta)\pi_m(\theta)d\theta > \cancel{\frac{1}{n}} - \int R(\hat{\theta}_m, \theta)\pi_m(\theta)d\theta$$

$$\int R(\tilde{\theta}_m, \theta)\pi_m(\theta)d\theta < \int R(\hat{\theta}_m, \theta)\pi_m(\theta)d\theta \quad \text{impossible}$$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} N(\theta, I_2)$ $\quad \|\hat{\theta} - \theta\|^2$

$\theta \in \mathbb{R}^2$

Q. Is $\bar{X}$ admissible?

yes (proved by Stein)

Constructed complicated prior distribution

e.g. $X_1 \cdots X_n \sim N(0, I_3)$ $\quad \|\hat{\theta} - \theta\|^2$

$\theta \in \mathbb{R}^3$

Q. is $\bar{X}$ admissible. No.

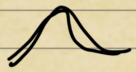
James-Stein estimator

$$\hat{\theta}_{JS} = \left(1 - \frac{p-2}{n\|X\|^2}\right)\bar{X}$$

Stein paradox

Theorem: $E_\theta \|\hat{\theta}_{JS} - \theta\|^2 < E_\theta \|\bar{X} - \theta\|^2 \quad \forall \theta \in \mathbb{R}^p$ for $p \geq 3$

(Brown, 1971)

 $\hat{\theta}_{JS}$ is minimax, even if it's non-constant

Lecture 6

$X_1 \cdots X_n \overset{iid}{\sim} N(\theta, I_p)$ $\quad \theta \in \mathbb{R}^p$ $\quad \|\hat{\theta} - \theta\|^2$

$p \geq 3$ $\quad \bar{X}$ inadmissible

$$\hat{\theta}_{JS} = \left(1 - \frac{p-2}{n\|\bar{X}\|^2}\right) \bar{X}$$

Thm. $E_\theta \|\hat{\theta}_{JS} - \theta\|^2 < \frac{p}{n} = E_\theta \|\bar{X} - \theta\|^2$ $\quad \forall \theta \in \mathbb{R}^p$ $\quad p \geq 3$

An empirical Bayes view (Efron & Morris)

empirical Bayes framework (Robbins) $\quad$ Estimate hyper parameter using data

$P(x|\theta)$ $\quad \pi(\theta|\tau^2)$

$\quad \hookrightarrow$ hyper parameter

can estimate τ^2 using $m(x|\tau^2) = \int P(x|\theta)\pi(\theta|\tau^2)\,d\theta$

particularly useful when $\theta \in \mathbb{R}^p$ for large p

and $\theta_1 \cdots \theta_p | \tau^2$ are iid.

Compound decision theory $\frac{1}{p} E_\theta \|\hat{\theta} - \theta\|^2 = E_\theta \frac{1}{p} \sum_{j=1}^{p} (\hat{\theta}_j - \theta_j)^2$

$\theta \sim \frac{1}{p} \sum_{j=1}^{p} \delta_{\theta_j}$

derivation of $\hat{\theta}_{JS}$ from empirical Bayes

$X_1 \cdots X_n | \theta \overset{iid}{\sim} N(\theta, I_p)$ $\quad \theta|\tau^2 \sim N(0, \tau^2 I_p)$

$$\hat{\theta} = E(\theta | X_1 \cdots X_n) = \frac{n}{n + \frac{1}{\tau^2}} \bar{X} = \left(1 - \frac{\frac{1}{n}}{\frac{1}{n} + \tau^2}\right) \bar{X}$$

calculated before for minimax

$m(x|\tau^2) = \int p(x|\theta)\pi(\theta|\tau^2)\,d\theta$

$X_i = \theta + z_i$ $\quad z_i \overset{iid}{\sim} N(0, I_p)$ $\quad \theta = \tau W$ $\quad W \sim N(0, I_p)$

$\Rightarrow X_i = \tau W + z_i$

$\Rightarrow \bar{X} = \tau W + \bar{z} \sim N(0, (\tau^2 + \frac{1}{n}) I_p)$

$$\frac{\|\bar{X}\|^2}{\tau^2 + \frac{1}{n}} \sim \chi_p^2 \Rightarrow \frac{\tau^2 + \frac{1}{n}}{\|\bar{X}\|^2} \sim \text{inv-}\chi_p^2$$

$$\mathbb{E}\, \frac{\tau^2+\frac{1}{n}}{\|x\|^2} = \frac{1}{p-2} \iff \mathbb{E}\left(\frac{p-2}{\|\bar{x}\|^2}\right) = \frac{1}{\tau^2+\frac{1}{n}}$$

method of moment $\frac{p-2}{\|\bar{x}\|^2}$ is an unbiased estimator of $\frac{1}{\tau^2+\frac{1}{n}}$

$$\hat{\theta}_{JS} = \left(1 - \frac{p-2}{n\|x\|^2}\right)\bar{x}$$

Theorem $p \geq 3$ $E_\theta \|\hat{\theta}_{JS} - \theta\|^2 < \frac{p}{n} = E_\theta \|\bar{x}-\theta\|^2$ $\forall \theta \in \mathbb{R}^p$

Lemma (Stein's identity) $Z \sim N(0,1)$ Some measure-theory condition on g as well

$$\mathbb{E}\, Z g(Z) = E\, g'(Z)$$

$\forall g \Rightarrow Z$ has to be normal

prove CLT by Stein's method

pf) $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}(-x) = -x\phi(x)$

$$\begin{aligned}
\mathbb{E}\, Zg(Z) &= \int x g(x)\phi(x)\,dx \\
&= \int -\phi'(x) g(x)\,dx \\
&= -\underbrace{\phi(x)g(x)\Big]_{-\infty}^{\infty}}_{0} + \int \phi(x) g'(x)\,dx \\
&= E(g'(Z))
\end{aligned}$$

$Z \sim N(0, I_p)$ $g: \mathbb{R}^p \to \mathbb{R}^p$

$$E(\langle Z, g(Z)\rangle) = E\langle \nabla, g(Z)\rangle = \sum_{j=1}^{p} E \frac{\partial}{\partial z_j} g(Z)$$

pf of theorem) $E_\theta \|\hat{\theta}_{JS} - \theta\|^2$

$$= E_\theta \left\|\left(1-\frac{p-2}{n\|\bar{x}\|^2}\right)\bar{x} - \theta\right\|^2$$

$$= E_\theta \left\|\bar{x} - \theta - \frac{p-2}{n\|\bar{x}\|^2}\bar{x}\right\|^2$$

$\bar{x} \sim N(\theta, \frac{1}{n} I_p)$

$$\begin{aligned}
\iff \bar{x} &= \theta + \tfrac{1}{\sqrt{n}} Z \quad Z \sim N(0, I_p) \\
&= \tfrac{1}{\sqrt{n}}(\sqrt{n}\,\theta + Z) \\
&= \tfrac{1}{\sqrt{n}}(\mu + Z) \quad \mu = \sqrt{n}\,\theta
\end{aligned}$$

$$= E\left\|\frac{1}{\sqrt{n}} Z - \frac{p-2}{\|\mu + Z\|^2}\frac{1}{\sqrt{n}}(\mu+Z)\right\|^2$$

$$= \frac{1}{n} E\left\|Z - \frac{p-2}{\|\mu+Z\|^2}(\mu+Z)\right\|^2$$ reduce to sample size 1 problem

$$= \frac{1}{n}\left(E\|Z\|^2 + E\frac{(p-2)^2}{\|\mu+Z\|^2} - 2(p-2)E\left\langle Z, \frac{\mu+Z}{\|\mu+Z\|^2}\right\rangle\right)$$

analysis of $E\left\langle Z, \frac{\mu+Z}{\|\mu+Z\|^2}\right\rangle$ $g(Z) = \frac{\mu+Z}{\|\mu+Z\|^2} = \begin{pmatrix} g_1(Z) \\ \vdots \\ g_p(Z)\end{pmatrix}$

$$g_j(Z) = \frac{\mu_j + Z_j}{\|\mu+Z\|^2}$$

$$\frac{\partial}{\partial z_j} g_j(z) = \frac{\|\mu+z\|^2 - 2(\mu_j+z_j)^2}{\|\mu+z\|^4}$$

$$E\left\langle z, \frac{\mu+z}{\|\mu+z\|^2}\right\rangle = \sum_{j=1}^{p} E \frac{\partial}{\partial z_j} g_j(z)$$

$$= E \sum_{j=1}^{p} \frac{\|\mu+z\|^2 - 2(\mu_j+z_j)^2}{\|\mu+z\|^4}$$

$$= E\left(\frac{(p-2)}{\|\mu+z\|^2}\right)$$

$$E_\theta \|\hat\theta_{JS} - \theta\|^2 = \frac{1}{n}\left(p + E\frac{(p-2)^2}{\|\mu+z\|^2} - 2(p-2)\, E\frac{p-2}{\|\mu+z\|^2}\right)$$

$$= \frac{1}{n}\left(p - (p-2)^2 E\frac{1}{\|\mu+z\|^2}\right) \qquad z \sim N(0, I_p),\ \mu = \sqrt{n}\,\theta$$

$$< \frac{p}{n} \qquad \blacksquare$$

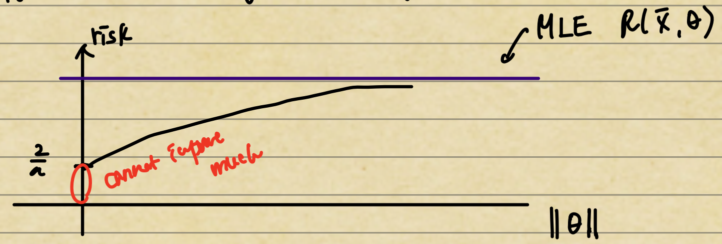
$$R(\hat\theta_{JS}, \theta) = \frac{1}{n}\left(p - (p-2)^2 E\frac{1}{\|\sqrt{n}\,\theta + z\|^2}\right)$$

$$R(\hat\theta_{JS}, 0) = \frac{1}{n}\left(p - (p-2)^2 E\frac{1}{\|z\|^2}\right)$$

$$= \frac{2}{n}$$

By symmetry, $R(\hat\theta_{JS}, \theta)$ depends on θ through $\|\theta\|$

as $\|\theta\| \to \infty$ $\quad R(\hat\theta_{JS}, \theta) \to \frac{p}{n}$



$$\sup_{\theta\in\mathbb{R}^p} R(\hat\theta_{JS}, \theta) = \sup_{\theta\in\mathbb{R}^p} R(\bar X, \theta)$$

$\left(1 - \frac{\frac{1}{n}}{\frac{1}{n}+\tau^2}\right)\bar X$ is Bayes and is admissible

Shrinkage estimation $\quad X_1, \dots, X_n \overset{iid}{\sim} N(\theta, I_p) \quad \hat\theta_c = c\bar X$

$$R(\hat\theta_c, \theta) = E_\theta \|c\bar X - c\theta\|^2 + \|c\theta - \theta\|^2$$

$$= c^2 \frac{p}{n} + (c-1)^2 \|\theta\|^2 = f(c)$$

$$f'(c) = 2c\frac{p}{n} + 2(c-1)\|\theta\|^2 = 0$$

$$\Rightarrow \quad c^* = \frac{\|\theta\|^2}{\frac{p}{n} + \|\theta\|^2}$$

$$\hat{\theta}_{c^*} = \frac{\|\theta\|^2}{\frac{p}{n} + \|\theta\|^2}\bar{X} = \left(1 - \frac{p}{p + n\|\theta\|^2}\right)\bar{X}$$

$$R(\hat{\theta}_{c^*}, \theta) = \left(\frac{b}{a+b}\right)^2 a + \left(\frac{a}{a+b}\right)^2 b \qquad a = \frac{p}{n}$$

$$b = \|\theta\|^2$$

"oracle estimator"

$$= \frac{b^2 a + a^2 b}{(a+b)^2} = \frac{ab}{a+b} \leq \min(a, b)$$

$$= \min\left(\frac{p}{n}, \|\theta\|^2\right)$$

MLE risk

$$R(\hat{\theta}_{JS}, \theta) = \frac{1}{n}\left(p - (p-2)^2 E\frac{1}{\|Z+\mu\|^2}\right) \qquad Z \sim N(0, I_p)$$

$$\mu = \sqrt{n}\,\theta$$

non central parameter

$$\|Z+\mu\|^2 \sim \chi^2_{p, \|\mu\|^2} \overset{d}{=} \chi^2_{p+2N} \qquad N \sim \text{poisson}\left(\frac{\|\mu\|^2}{2}\right)$$

$$E\frac{1}{\|Z+\mu\|^2} = E\frac{1}{\chi^2_{p+2N}} = E\left(E\left(\frac{1}{\chi^2_{p+2N}} \middle| N\right)\right)$$

$$= E\frac{1}{p+2N-2} \geq \frac{1}{E(p+2N-2)}$$

$$= \frac{1}{p + \|\mu\|^2 - 2}$$

$$R(\hat{\theta}_{JS}, \theta) \leq \frac{1}{n}\left(p - (p-2)^2\frac{1}{p+\|\mu\|^2-2}\right)$$

$$= \frac{1}{n}\,\frac{\cancel{p^2} + p\|\mu\|^2 - 2p - \cancel{p^2} + 4p - 4}{p + \|\mu\|^2 - 2}$$

$$= \frac{1}{n}\,\frac{p\|\mu\|^2 + 2(p-2)}{p + \|\mu\|^2 - 2}$$

$$= \frac{1}{n}\left(2 + \frac{(p-2)\|\mu\|^2}{p-2+\|\mu\|^2}\right) \qquad \mu = \sqrt{n}\,\theta$$

$$= \frac{2}{n} + \frac{\frac{(p-2)}{n}\|\theta\|^2}{\frac{p-2}{n} + \|\theta\|^2} \leq \frac{2}{n} + \frac{\frac{p}{n}\|\theta\|^2}{\frac{p}{n} + \|\theta\|^2}$$

$$= \frac{2}{n} + R(\hat{\theta}_{c^*}, \theta)$$

Theorem (oracle inequality)

P23 $\quad R(\hat{\theta}_{JS}, \theta) \leq \inf_c R(\hat{\theta}_c, \theta) + \frac{2}{n} \rightarrow$ dimension-free

comparable

Lecture 7

$X_1 \cdots X_n \overset{iid}{\sim} f \in S_\alpha(R)$ $\quad \|\hat f - f\|^2 = \int_0^1 (\hat f(x) - f(x))^2\, dx$

$\subseteq L^2[0,1]$

Fourier analysis $\quad f(x) = a_0 + \sum_{j=1}^{\infty} \left(a_j \cos(2\pi j x) + b_j \sin(2\pi j x) \right)$

Nonparametric density estimation $= \sum_j \theta_j \phi_j(x) \qquad \int \phi_i \phi_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

$f'(x) = \sum \left[a_j (2\pi j) \sin(2\pi j x) + 2\pi j\, b_j \cos(2\pi j x) \right]$ $\quad$ fact: $\cos x \perp\!\!\!\perp \sin x$ $[0, 2\pi]$

$\|f\|^2 = \frac{1}{2} \sum_j (a_j^2 + b_j^2) \qquad \|f'\|^2 = \frac{1}{2} \sum_j (2\pi j)^2 (a_j^2 + b_j^2)$

$\|f''\|^2 = \frac{1}{2} \sum_j (2\pi j)^4 (a_j^2 + b_j^2)$ $\quad$ need to decay faster

smoothness power

$\|f^{(\alpha)}\| = \sum_j (2\pi j)^{2\alpha} (a_j^2 + b_j^2)$

$S_\alpha(R) = \left\{ f = \sum \theta_j \phi_j : f \geq 0, \int f = 1, \sum_j j^{2\alpha} \theta_j^2 \leq R^2 \right\} \subseteq L^2[0,1]$

radius

Sobolev ball

$\theta_j = \int f \phi_j \qquad \hat\theta_j = \frac{1}{n} \sum_{i=1}^{n} \phi_j(X_i) \overset{\text{CLT approx}}{\sim} N\left(\theta_j, \frac{Var(\phi_j(X))}{n}\right)$

$= E \phi_j(X)$

Gaussian sequence model

$X_j = \theta_j + \frac{1}{\sqrt{n}} Z_j \qquad j = 1 \cdots \infty \qquad Z_j \overset{iid}{\sim} N(0,1)$

$\sim N(\theta_j, \frac{1}{n}) \qquad \Theta(R) = \{ \theta : \sum_j j^{2\alpha} \theta_j^2 \leq R^2 \}$

$\|\hat\theta - \theta\|^2 = \sum_{j=1}^{\infty} (\hat\theta_j - \theta_j)^2$

Le Cam's asymptotic equivalence $\quad X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1) \quad \theta \in \mathbb{R}$ $\quad$ equivalence

$n \to \infty \qquad \bar X \sim N(\theta, \frac{1}{n}) \quad \theta \in \mathbb{R}$

Brown & Low (1996): connection btw
- non parametric regression
- white noise model $\quad dY(t) = f(t)\,dt + \frac{1}{\sqrt{n}} dB(t)$

$\quad ||$

Gaussian sequence model

Nussman (1996) Connection btw $\begin{cases} \text{density estimation} \\ \text{white noise model} \end{cases}$

$$\hat{\theta}_j = \begin{cases} X_j & j \le k \\ 0 & j > k \end{cases}$$

idea: $\Theta(R) = \{ \theta : \sum_j j^{2\alpha} \theta_j^2 \le R^2 \}$ so if j large $\theta_j \Downarrow$

$$E \|\hat{\theta} - \theta\|^2 = \underbrace{\sum_{j=1}^{k} E(X_j - \theta_j)^2}_{\text{variance}} + \underbrace{\sum_{j>k} \theta_j^2}_{\text{bias}}$$

Variance : $\sum_{j=1}^{k} E(X_j - \theta)^2 = \frac{k}{n}$

bias: $\sum_{j>k} \theta_j^2 = \sum_{j>k} j^{-2\alpha} j^{2\alpha} \theta_j^2$

$\le k^{-2\alpha} \sum_{j>k} j^{2\alpha} \theta_j^2$

$\le k^{-2\alpha} R^2$

"Approximation error"

$$E \|\hat{\theta} - \theta\|^2 \le \frac{k}{n} + k^{-2\alpha} R^2$$

$$\le \underbrace{\underline{C(R,\alpha)}}_{\text{fixed w.r.t. } n}\ \underline{n^{-\frac{2\alpha}{2\alpha+1}}}_{\text{optimal rate}}$$

nonparametric rate

$\frac{k}{n} = k^{-2\alpha} R^2$

$k \asymp n^{\frac{1}{2\alpha+1}}$

small α: rough function $\Rightarrow$ need more terms

Pinsker constant

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta(R)} E_\theta \|\hat{\theta} - \theta\|^2 = (1 + o(1))\, C_{\alpha,R}\, n^{-\frac{2\alpha}{2\alpha+1}}$$ minimax

$o(1) \to 0$ as $n \to \infty$

problem: α unknown. Need to adapt to α!

$X \sim N(\theta, \frac{1}{n} I_p)$ $\quad \hat{\theta}_{JS} = \left(1 - \frac{p-2}{n\|X\|^2}\right) X$

$$E_\theta \|\hat{\theta}_{JS} - \theta\|^2 \le \underline{\min_c E_\theta \|c\bar{X} - \theta\|^2} + \frac{2}{n} \qquad \forall \theta \in \mathbb{R}^p$$

$$\le \min\left(\frac{p}{n}, \|\theta\|^2\right) + \frac{2}{n}$$

$c=1 \qquad c=0$

$\{1 \cdots n\} = B_1 \cup B_2 \cup \cdots \cup B_m$

$B_1 = \{1, 2, 3\} \quad \cdots \quad B_\ell = \{3^{\ell-1} - 1, \cdots, 3^\ell\}$

$|B_\ell| = 3^\ell - 3^{\ell-1} = \frac{2}{3} 3^\ell$

$$\hat{\theta} : \begin{pmatrix} \hat{\theta}_{B_1} \\ \hat{\theta}_{B_2} \\ \vdots \\ \hat{\theta}_{B_m} \\ 0 \end{pmatrix} \qquad \hat{\theta}_{B_\ell} = \left(1 - \frac{|B_\ell| - 2}{n \|X_\ell\|^2}\right) X_\ell$$

$B_m = \{ \cdots 3^m \}$ $\quad m = \log_3 n$

$3^m \to n$

Block-wise James Stein estimator ("BJS")

$$E\,\|\hat\theta-\theta\|^2 = \sum_{\ell=1}^{m} E\|\hat\theta_{B_\ell}-\theta_{B_\ell}\|^2 + \sum_{j>m}\theta_j^2 \lesssim \underbrace{\sum_{\ell=1}^{m}\min\left(\frac{|B_\ell|}{n},\ \|\theta_{B_\ell}\|^2\right)}_{①} + \underbrace{\frac{2m}{n}}_{②} + \underbrace{\sum_{j>n}\theta_j^2}_{③}$$

② $\displaystyle \frac{2m}{n} = \frac{2\log_3 n}{n} = o\left(n^{-\frac{2\alpha}{2\alpha+1}}\right)$ " $\frac{m}{n}$ very small"

③ $\displaystyle \sum_{j>n}\theta_j^2 = \sum_{j>n} j^{-2\alpha}\cdot j^{2\alpha}\,\theta_j^2 \lesssim n^{-2\alpha}\sum_{j>n} j^{2\alpha}\theta_j^2 \leq R^2 n^{-2\alpha} = o\left(n^{-\frac{2\alpha}{2\alpha+1}}\right)$

threshold before

① $\displaystyle \sum_{\ell=1}^{m}\min\left(\frac{|B_\ell|}{n},\ \|\theta_{B_\ell}\|^2\right)$

Choose $L = \min\{\ell\in\mathbb{N} \quad 3^\ell \geq n^{\frac{1}{2\alpha+1}}\}$

$3^{L-1} \leq n^{\frac{1}{2\alpha+1}} \leq 3^L$

$\displaystyle \sum_{j>k}\theta_j^2 \lesssim k^{-2\alpha}$

$$\leq \sum_{\ell=1}^{L}\frac{|B_\ell|}{n} + \sum_{\ell>L}\|\theta_{B_\ell}\|^2$$

$$\leq \sum_{\ell=1}^{L}\frac{2}{3n}\cdot 3^\ell + \sum_{j>3^L}\theta_j^2$$

$$\lesssim \frac{3^L}{n} + \sum_{j>n^{\frac{1}{2\alpha+1}}}\theta_j^2$$

$$\lesssim \frac{n^{\frac{1}{2\alpha+1}}}{n} + \left(n^{\frac{1}{2\alpha+1}}\right)^{-2\alpha} \lesssim n^{-\frac{2\alpha}{2\alpha+1}}$$

Lecture 8

$X_j = \theta_j + \frac{1}{\sqrt{n}} z_j \qquad z_j \overset{iid}{\sim} N(0,1) \qquad \Theta_d(R) = \{\theta : \sum j^{2d}\theta_j^2 \le R^2\}$

$$\sup_{\theta \in \Theta_d(R)} E_\theta \|\hat\theta - \theta\|^2 \lesssim C\, n^{-\frac{2d}{2d+1}}$$ achieve minimax rate

$$\inf_{\hat\theta} \sup_{\theta \in \Theta_d(R)} E_\theta \|\hat\theta - \theta\|^2 \gtrsim C\, n^{-\frac{2d}{2d+1}}$$

Hypothesis test

two-point test (simple vs. simple)

$H_0 : X \sim P \qquad H_1 : X \sim Q$

testing function $\phi : \mathcal{X} \mapsto \{0,1\}$

Type I-error $P\phi = E_{X\sim P}\,\phi(x)$

Type II-error $Q(1-\phi)$

testing error $P\phi + Q(1-\phi)$

Optimal testing error $\inf_{\phi} (P\phi + Q(1-\phi))$

def. total variation distance

why not max q(x)?

$$TV(P,Q) = \sup_B |P(B) - Q(B)|$$

Theorem $TV(P,Q) = P(p(x) > q(x)) - Q(p(x) > q(x))$ where $\begin{cases} p = \frac{dP}{dP+dQ} \\ q = \frac{dQ}{dP+dQ} \end{cases}$

$$= \frac{1}{2}\int |p-q|(x)\, d(P+Q)(x)$$

$$= 1 - \underbrace{\int \min(p,q)}_{\text{total variation affinity}}$$

pf) $A = \{p(x) > q(x)\}$

$TV(P,Q) \ge P(A) - Q(A)$

$\forall B \quad |P(B) - Q(B)| = \left|\int_B (p-q)\right|$

$$= \left|\int_{B\cap A}(p-q) + \int_{B\cap A^c}(p-q)\right|$$

$$= \left|\int_{B\cap A}(p-q) - \int_{B\cap A^c}(q-p)\right|$$

$$\le \max\left(\int_{B\cap A}(p-q),\ \int_{B\cap A^c}(q-p)\right)$$

$$\le \max\left(\int_A (p-q),\ \int_{A^c}(q-p)\right)$$

$= \max\left(P(A)-Q(A),\ Q(A^c)-P(A^c)\right)$

$= P(A)-Q(A)$

Taking sup, $TV(P,Q) \le P(A)-Q(A)$

$\Rightarrow TV(P,Q) = P(A)-Q(A)$

$$\frac{1}{2}\int |p-q| = \frac{1}{2}\int_{p>q}(p-q) + \frac{1}{2}\int_{p\le q}(q-p)$$

$$= \frac{1}{2}\left(P(A)-Q(A)\right) + \frac{1}{2}\left(Q(A^c)-P(A^c)\right)$$

$$= P(A) - Q(A) = TV(P,Q)$$

$$\int \min(p,q) = \int_{p>q} \min(p,q) + \int_{p\le q} \min(p,q)$$

$$= \int_{p>q} q + \int_{p\le q} p$$

$$= Q(A) + P(A^c)$$

$$= Q(A) + 1 - P(A)$$

$$= 1-(P(A)-Q(A))$$

$$= 1 - TV(P,Q)$$

Theorem (Neyman-Pearson lemma)

$$\inf_{\phi}\left(P\phi + Q(1-\phi)\right) = 1-TV(P,Q) = \int \min(p,q)$$

likelihood ratio test

the optimal ϕ is $\phi(x) = 1\{p(x) < q(x)\}$

pf) $\forall \phi$ $\quad P\phi + Q(1-\phi) = \int p\phi + \int q(1-\phi)$

$$= \int p\phi + q(1-\phi)$$

$$\ge \int \min(p,q)$$

$$\Rightarrow \inf_{\phi}\left(P\phi + Q(1-\phi)\right) \ge \int \min(p,q)$$

$$\inf_{\phi}\left(P\phi + Q(1-\phi)\right) \le P(p<q) + Q(p\ge q)$$

$$= 1 - TV(P,Q)$$

$$= \int \min(p,q)$$

■

Le Cam two-point method

$$\inf_{\hat\theta}\sup_{\theta\in\Theta} E_\theta(\hat\theta-\theta)^2 \ge \inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat\theta-\theta)^2 \quad (\theta_1,\theta_2\in\Theta)$$

$$\ge \underbrace{\frac{(\theta_1-\theta_2)^2}{4}}_{\text{far away}}\ \underbrace{\int \min(P_{\theta_1},P_{\theta_2})}_{\text{but not too far}}$$

$$\int \min(P_{\theta_1},P_{\theta_2}) = \inf_{\phi} P_{\theta_1}\phi + P_{\theta_2}(1-\phi)$$

pf) $\inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat\theta-\theta)^2$

$$\ge \inf_{\hat\theta} \frac12\left(E_{\theta_1}(\hat\theta-\theta_1)^2\right)+\frac12 E_{\theta_2}(\hat\theta-\theta_2)^2)$$

$$= \inf_{\hat\theta}\frac12\int\left((\hat\theta-\theta_1)^2P_{\theta_1}+(\hat\theta-\theta_2)^2P_{\theta_2}\right)$$

$$\ge \inf_{\hat\theta}\frac12\int\left[(\hat\theta-\theta_1)^2+(\hat\theta-\theta_2)^2\right]\min(P_{\theta_1},P_{\theta_2})$$

$$\ge \frac12\int\frac{(\theta_1-\theta_2)^2}{2}\min(P_{\theta_1},P_{\theta_2})$$

$$= \frac{(\theta_1-\theta_2)^2}{4}\int\min(P_{\theta_1},P_{\theta_2})$$

$(x+y)^2 \le 2x^2+2y^2$

$(\theta_1-\theta_2)^2 = (\theta_1-\hat\theta+\hat\theta-\theta_2)^2$

$\le 2(\hat\theta-\theta_1)^2+2(\hat\theta-\theta_2)^2$

e.g. $X_1\cdots X_n \overset{iid}{\sim} N(\theta,1) \quad \theta\in\mathbb{R}$

good for 1d

$\bar X$ CLT $\frac{1}{\sqrt n}$ right order

$\bar X$ CLT

two tails apart as far as possible

$$\inf_{\hat\theta}\sup_{\theta\in\mathbb{R}} E_\theta(\hat\theta-\theta)^2 \ge \inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat\theta-\theta)^2$$

Chose $\theta_1=0$ $\theta_2=\frac{1}{\sqrt n}$, $P_\theta=N(\theta,1)$

endpt

$$\inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat\theta-\theta)^2 \ge \frac{1}{4n}\inf_{\phi}\left(P_0^n\phi+P_{\frac{1}{\sqrt n}}^n(1-\phi)\right) \qquad \left(\frac{(\theta_1-\theta_2)^2}{4}\int\min(P_{\theta_1},P_{\theta_2})\right)$$

$$\inf_{\phi}\left(P_0^n\phi+P_{\frac{1}{\sqrt n}}^n(1-\phi)\right) \ge P_0^n\left(P_0^n(x)<P_{\frac{1}{\sqrt n}}^n(x)\right)$$

$$= P_0^n\left(\prod_{i=1}^n e^{-\frac12(X_i-\frac{1}{\sqrt n})^2}\Big/ e^{-\frac12 X_i^2}>1\right)$$

$$= P_0^n\left(\sum_{i=1}^n\left[(X_i-\tfrac{1}{\sqrt n})^2-X_i^2\right]<0\right)$$

$$= P_0^n\left(\sum_{i=1}^n\left(-\frac{2}{\sqrt n}X_i+\frac1n\right)<0\right)$$

$$= P_0^n\left(\frac{1}{\sqrt n}\sum X_i>\frac12\right)$$

$\bar X > \frac{1}{2\sqrt n}$ $\quad (0+\frac{1}{\sqrt n})\frac12$

$$= \mathbb{P}\left(N(0,1)>\frac12\right)$$

$$\Rightarrow \inf_{\hat\theta}\sup_{\theta\in\mathbb{R}} E(\hat\theta-\theta)^2 \ge \frac{\mathbb{P}(N(0,1)>\frac12)}{4}\frac1n$$

constant : not optimal

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$ $\quad \theta \in [0,1]$

LAE (local asymptotic equivariality)

$$\inf_{\hat\theta} \sup_{\theta\in[0,1]} E_\theta(\hat\theta-\theta)^2 \asymp n^{-?}$$

upper bound $\hat\theta = \max_{1\le i\le n} X_i$ (MLE)

$$F(t) = P(\hat\theta \le t) = \prod_{i=1}^n P(X_i \le t) = \left(\frac{t}{\theta}\right)^n$$

$$f(t) = n\left(\frac{t}{\theta}\right)^{n-1}\frac{1}{\theta} = \theta^{-n} n t^{n-1} \quad t\in(0,\theta)$$

$$E_\theta(\hat\theta-\theta)^2 = \int_0^\theta (t-\theta)^2\, \theta^{-n} n t^{n-1}\,dt$$

$$= \theta^{-n} n \int_0^\theta (t^2+\theta^2-2t\theta)\, t^{n-1}\,dt$$

$$= \theta^{-n} n \left(\frac{t^{n+2}}{n+2}\Big]_0^\theta + \theta^2 \frac{t^n}{n}\Big]_0^\theta - 2\theta\frac{t^{n+1}}{n+1}\Big]_0^\theta \right)$$

$$= \theta^{-n} n \left(\frac{\theta^{n+2}}{n+2} + \frac{\theta^{n+2}}{n} - 2\frac{\theta^{n+2}}{n+1}\right)$$

$$= \theta^2 n \left(\frac{1}{n+2} - \frac{1}{n+1} + \frac{1}{n} - \frac{1}{n+1}\right)$$

$$= \theta^2 n \left(\frac{-1}{(n+2)(n+1)} + \frac{1}{n(n+1)}\right)$$

$$= \frac{\theta^2 n}{n+1}\left(\frac{1}{n} - \frac{1}{n+2}\right)$$

$$= \frac{\theta^2 n}{n+1}\,\frac{2}{n(n+2)}$$

$$= \frac{2\theta^2}{(n+1)(n+2)}$$

$$\sup_{\theta\in(0,1)} E_\theta(\hat\theta-\theta)^2 = \frac{2}{(n+1)(n+2)} = O(n^{-2})$$

choose 1

only for LAE

↓

fisher X

↓

Unifom non-smooth ⇒ things are very different

$X_1 \cdots X_n \overset{iid}{\sim} P_\theta$

$\sqrt{n}(\underbrace{\hat\theta-\theta}_{\frac{1}{\sqrt n}}) \sim N(0, \frac{1}{I_\theta})$ $\Rightarrow \frac{1}{n}$

Lower bound: $\inf_{\hat\theta}\sup_{\theta\in(0,1)} E_\theta(\hat\theta-\theta)^2 \ge \inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat\theta-\theta)^2$

close to 1 ← $\theta_1 = 1 \quad \theta_2 = 1-\frac{1}{n}$

$P_\theta \sim \text{Unif}(0,\theta)$

$$\inf_{\hat\theta}\sup_{\theta\in\{\theta_1,\theta_2\}} E_\theta(\hat\theta-\theta)^2 \ge \frac{1}{4n^2}\int P_1^n \overset{\text{minimum}}{\wedge} P_{1-\frac{1}{n}}^n$$

inequality: $\int p\wedge q \ge \frac{1}{2}\left(\underbrace{\int\sqrt{pq}}\right)^2$ "Hellinger affinity" easier to work

$$\int P_1^n \wedge P_{1-\frac1n}^n \ge \frac12\left(\int\sqrt{P_1^n P_{1-\frac1n}^n}\right)^2$$

$$= \frac12\left(\int\sqrt{\prod_{i=1}^n P_1(x_i)\prod_{i=1}^n P_{1-\frac1n}(x_i)}\right)^2$$

$$= \frac{1}{2}\left(\int \sqrt{P_1(x_1) P_{1-\frac{1}{n}}(x_1)}\right)^{2n}$$

$$= \frac{1}{2}\left(\int_0^{\theta_2} \sqrt{\frac{1}{\theta_1}\frac{1}{\theta_2}}\right)^{2n}$$

$$= \frac{1}{2}\left(\sqrt{\frac{\theta_2}{\theta_1}}\right)^{2n} = \frac{1}{2}\left(1-\frac{1}{n}\right)^{n} \to \frac{1}{2}e^{-1}$$

$$\geq \frac{1}{8} \quad \forall n \geq 2$$

$$\Rightarrow \inf_{\hat{\theta}} \sup_{\theta \in [0,1]} E_\theta\left((\hat{\theta}-\theta)^2\right) \geq \frac{1}{32n^2}$$

Lecture 9

$X_j \sim N(\theta_j, \frac{1}{n}) \quad j \in \mathbb{N} \qquad \Theta_\alpha(R) = \{ \theta : \sum_{j=1}^{\infty} j^{2\alpha} \theta_j^2 \le R^2 \}$

$$\inf_{\hat\theta} \sup_{\theta \in \Theta_\alpha(R)} E_\theta \|\hat\theta - \theta\|^2 \gtrsim C n^{-\frac{2\alpha}{2\alpha+1}}, \qquad n^{-\frac{2\alpha}{2\alpha+1}} = \frac{k}{n} \qquad k = n^{\frac{1}{2\alpha+1}}$$

Le Cam two-point

$$\inf_{\hat\theta} \sup_{\theta \in \{\theta_1, \theta_0\}} E_\theta (\hat\theta - \theta)^2 \ge \frac{(\theta_0 - \theta_1)^2}{4} \int P_{\theta_0} \wedge P_{\theta_1}$$

$\Theta_0 = \{ \theta : \theta_j \in \{0, \frac{1}{\sqrt{n}}\} \ j = 1 \cdots k, \ \theta_j = 0 \ j > k \}$

$|\Theta_0| = 2^k \qquad$ want $\Theta_0 \subseteq \Theta_\alpha(R)$

$\forall \theta \in \Theta_0 \quad \sum j^{2\alpha} \theta_j^2 \underset{\text{want}}{\le} R^2$

$$\sum_{j=1}^{k} j^{2\alpha} \theta_j^2 \le \left(\sum_{j=1}^{k} j^{2\alpha} \right) \frac{1}{n} \le \frac{k^{2\alpha+1}}{n}$$

$\Rightarrow$ Choose $k \asymp n^{\frac{1}{2\alpha+1}}$

$$\begin{aligned}
\inf_{\hat\theta} \sup_{\theta \in \Theta_\alpha(R)} E_\theta \|\hat\theta - \theta\|^2 &\ge \inf_{\hat\theta} \sup_{\theta \in \Theta_0} E_\theta \|\hat\theta - \theta\|^2 \\
&\ge \inf_{\hat\theta} \operatorname{ave}_{\theta \in \Theta_0} E_\theta \|\hat\theta - \theta\|^2 \\
&\ge \inf_{\hat\theta} \operatorname{ave}_{\theta \in \Theta_0} E_\theta \sum_{j=1}^{k} (\hat\theta_j - \theta_j)^2 \\
&= \inf_{\hat\theta} \sum_{j=1}^{k} \operatorname{ave}_{\theta \in \Theta_0} E_\theta (\hat\theta_j - \theta_j)^2 \\
&= \inf_{\hat\theta} \sum_{j=1}^{k} \operatorname{ave}_{\theta_{-j}} \left[\frac{1}{2} E_{\theta_{-j}, \theta_j = 0} (\hat\theta_j - 0)^2 + \frac{1}{2} E_{\theta_{-j}, \theta_j = \frac{1}{\sqrt{n}}} (\hat\theta_j - \tfrac{1}{\sqrt{n}})^2 \right] \\
&\ge \sum_{j=1}^{k} \operatorname{ave}_{\theta_{-j}} \inf_{\hat\theta_j} \left[\frac{1}{2} E_{\theta_{-j}, \theta_j = 0} (\hat\theta_j - 0)^2 + \frac{1}{2} E_{\theta_{-j}, \theta_j = \frac{1}{\sqrt{n}}} (\hat\theta_j - \tfrac{1}{\sqrt{n}})^2 \right] \\
&\ge \sum_{j=1}^{k} \operatorname{ave}_{\theta_{-j}} \frac{1}{4n} \int P_{(\theta_{-j}, \theta_j = 0)} \wedge P_{(\theta_{-j}, \theta_j = \frac{1}{\sqrt{n}})}
\end{aligned}$$

$H_0 : \theta_{-j}$ fixed $\quad \theta_j = 0 \quad \theta_l = 0 \ \forall l \ge k$

$H_1 : \theta_{-j}$ fixed $\quad \theta_j = \frac{1}{\sqrt{n}} \quad \theta_l = 0 \ \forall l \ge k$

$\Leftrightarrow$ likelihood ratio $H_0: X_j \sim N(0, \frac{1}{n})$ $\Leftrightarrow$ $X_1 \cdots X_n \overset{iid}{\sim} N(0,1)$

$H_1: X_j \sim N(\frac{1}{\sqrt{n}}, \frac{1}{n})$ $\qquad X_1 \cdots X_n \overset{iid}{\sim} N(\frac{1}{\sqrt{n}}, 1)$

$$\geq \sum_{j=1}^{k} \underset{\theta_{-j}}{\text{ave}} \frac{1}{4n} \mathbb{P}_{\theta_j, \theta_j = 0}\left(\frac{d\mathbb{P}_{\theta_j, \theta_j = \frac{1}{\sqrt{n}}}}{d\mathbb{P}_{\theta_{-j}, \theta_j = 0}}(x) \geq 1 \right)$$

$$= \sum_{j=1}^{k} \underset{\theta_{-j}}{\text{ave}} \frac{1}{4n} \mathbb{P}_{X_j \sim N(0, \frac{1}{n})}\left(\frac{e^{-\frac{1}{2}(x_j - \frac{1}{\sqrt{n}})^2}}{e^{-\frac{1}{2} x_j^2}} \geq 1 \right)$$

$$= \sum_{j=1}^{k} \underset{\theta_{-j}}{\text{ave}} \frac{1}{4n} \mathbb{P}\left(N(0,1) > \frac{1}{2}\right)$$

$$= \frac{k}{4n} \mathbb{P}\left(N(0,1) > \frac{1}{2}\right) \times n^{-\frac{2\alpha}{2\alpha+1}}$$

Using this method, can get minimax for $X_1 \cdots X_n \sim N(\theta, I_p)$: alternatively, could've used Bayesian approach before (squared loss)

unbiased estimators

$E_\theta \hat{\theta} = \theta$

$E(\hat{\theta} - \theta)^2 = Var(\hat{\theta})$

UMVUE : $\hat{\delta}(x)$ is UMVUE for $g(\theta)$ with respect to $\theta \in \Theta$

if $E_\theta \hat{\delta}(x) = g(\theta)$ $\forall \theta \in \Theta$ and for any other unbiased $\tilde{\delta}$

$Var(\hat{\delta}) \leq Var(\tilde{\delta})$ $\forall \theta \in \Theta$

Thm (Lehmann Scheffe)

If $\delta(x) = h(T(x))$ for T that is both sufficient and complete, and unbiased

then ① $h(T(x))$ is the only unbiased function of $T(x)$

② $h(T(x))$ is the unique UMVUE

pf) $\tilde{h}(T(x))$ is also unbiased. $E_\theta \tilde{h}(T(x)) = g(\theta) = E_\theta h(T(x))$ $\forall \theta \in \Theta$

$E_\theta\left(\tilde{h}(T(x)) - h(T(x))\right) = 0 \;\; \forall \theta \xRightarrow{\text{complete}} \tilde{h}(T(x)) = h(T(x))$ a.s.

Consider an unbiased $\delta(x)$ $\quad E_\theta \delta(x) = g(\theta)$

$h(T(x)) = E(\delta(x) | T(x))$ is also unbiased $\Rightarrow$ only fn of $T(x)$ that is unbiased

Rao-Blackwell $\quad \underbrace{E\left(\left(E(\delta(x)|T(x)) - g(\theta)\right)^2\right)}_{= Var(E(\delta(x)|T(x)))} \leq \underbrace{E\left((\delta(x) - g(\theta))^2\right)}_{= Var(\delta(x))}$

$X_1 \cdots X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$

① UMVUE for θ $\quad \delta = \bar{X}$

② UMVUE for θ^2 $\quad E(X_1 X_2) = \theta^2$

$$\delta(X) = E(X_1 X_2 \mid \Sigma X_i)$$

$$E(X_1 X_2 \mid \Sigma X_i = t) = \mathbb{P}(X_1 X_2 = 1 \mid \Sigma X_i = t)$$

Binom $(n-2, \theta)$

$$= \frac{\mathbb{P}(X_1 = 1 \;\; X_2 = 1 \;\; \Sigma X_i = t)}{\mathbb{P}(\Sigma X_i = t)} = \frac{\mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 1)\,\mathbb{P}(\sum_{i=3}^{n} X_i = t-2)}{\mathbb{P}(\Sigma X_i = t)}$$

$$= \frac{\theta^2 \binom{n-2}{t-2} \theta^{t-2} (1-\theta)^{n-t}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}}$$

$$= \frac{\frac{(n-2)!}{(t-2)!\,(n-t)!}}{\frac{n!}{t!\,(n-t)!}}$$

$$= \frac{t(t-1)}{n(n-1)}$$

$$\delta(X) = \frac{\sum_{i=1}^{n} X_i \left(\sum_{i=1}^{n} X_i - 1\right)}{n(n-1)}$$

Lecture 10

UMVUE

$(P_\theta : \theta \in \Theta)$ find $\delta(X)$ s.t. $E_\theta \delta(X) = g(\theta)$ $\forall \theta \in \Theta$

and has the smallest variance among all unbiased estimators $\forall \theta \in \Theta$

Two methods ① find $h(T(X))$ for sufficient & complete statistic $T(X)$

s.t. $E_\theta(h(T(X))) = g(\theta)$ $\forall \theta \in \Theta$

② find $\delta(X)$ s.t. $E(\delta(X)) = g(\theta)$

UMVUE is $E(\delta(X) \mid T(X))$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} \mathrm{Unif}.(0, \theta)$

$E(X_1) = \frac{\theta}{2} \Rightarrow E(2X_1) = \theta$

$E(2X_1 \mid X_{(n)}) = 2\left(\frac{1}{n} X_{(n)} + (1-\frac{1}{n}) \frac{X_{(n)}}{2}\right) = (\frac{2}{n} + 1 - \frac{1}{n}) X_{(n)} = (1+\frac{1}{n}) X_{(n)}$

sample $y_1 \cdots y_{n-1}$ iid $\sim \mathrm{Unif}(0, X_{(n)}]$

assign $(y_1 \cdots X_{(n)})$ randomly to $X_1 \cdots X_n$

$X_1 \mid X_{(n)} \sim \frac{1}{n} \delta_{X_{(n)}} + (1-\frac{1}{n}) \mathrm{Unif}(0, X_{(n)})$

e.g. $X_1 \cdots X_n \sim N(\mu, \sigma^2)$ $\mu \in \mathbb{R}$ $\sigma^2 > 0$

Complete suff $(\sum_{i=1}^n X_i \quad \sum_{i=1}^n X_i^2) \Leftrightarrow (\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$ (⊥ by Basu)

UMVUE for μ: $\bar{X}$

for σ^2: $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

for σ: $E|X_1 - \bar{X}|$ $\quad X_1 - \bar{X} \sim N(0, \frac{n-1}{n}\sigma^2)$

$= \sqrt{\frac{n-1}{n}}\, \sigma\, \underbrace{E(|Z|)}_{\sqrt{\frac{2}{\pi}}} \Leftarrow X_1 - \bar{X} = \frac{n-1}{n} \sigma Z$

Why not Γ ?

$X_i = \mu + \sigma Z_i$

$$\sqrt{\sum (X_i - \bar{X})^2} = \sqrt{\sigma^2 \sum (Z_i - \bar{Z})^2} = \sqrt{\sigma^2 \chi^2_{n-1}}$$

UMVUE is $\hat{\sigma} = \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \Big/ E\sqrt{\chi^2_{n-1}}$ "chi distribution"

for μ^2: $\mathbb{E}\bar{X}^2 = \mu^2 + \frac{\sigma^2}{n}$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\quad E\left(\frac{1}{n(n-1)} \sum (X_i - \bar{X})^2\right) = \frac{\sigma^2}{n}$

UMVUE is $\bar{X}^2 - \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2$

e.g. $X_1 \cdots X_n \overset{iid}{\sim}$ poisson(θ) estimate $e^{-\theta}$

$\sum_{i=1}^{n} X_i \sim$ poisson$(n\theta)$ $\quad \mathbb{P}(\sum X_i = k) = \frac{(n\theta)^k}{k!} e^{-n\theta}$

complete sufficient

$$E\left(h\left(\sum_{i=1}^{n} X_i\right)\right) = e^{-\theta} \quad \forall \theta > 0$$

$$\sum_{k=0}^{\infty} h(k) \frac{e^{-n\theta}}{k!} (n\theta)^k = e^{-\theta} \iff \sum_{k=0}^{\infty} h(k) n^k \theta^k / k! = e^{(n-1)\theta}$$

Taylor expansion should converge

Taylor expansion

$$e^{(n-1)\theta} = \sum_{k=0}^{\infty} \frac{(n-1)^k \theta^k}{k!}$$

$$\Rightarrow \frac{h(k) n^k}{k!} = \frac{(n-1)^k}{k!}$$

$$h(k) = \left(\frac{n-1}{n}\right)^k$$

$$\Rightarrow \text{UMVUE is } \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i}$$

$X \sim N(\theta, I_p)$ X is complete, sufficient

$$E(\|\hat\theta - \theta\|^2) = g(\theta)$$

$$E_\theta \|\hat\theta - X + X - \theta\|^2$$

$$= E_\theta(\|\hat\theta - X\|^2 + \|X - \theta\|^2 + 2\langle X-\theta, \hat\theta - X\rangle)$$

$$= E_\theta\left(\underline{\|\hat\theta - X\|^2} + \underline{p} + 2\langle X-\theta, \hat\theta\rangle \underline{- 2\langle X-\theta, X\rangle}\right)$$

$$E(-2\langle X-\theta, X\rangle) = E(-2\langle Z, \theta+Z\rangle) = -2E(\|Z\|^2) = -2p$$

$X = \theta + Z$ $Z \sim N(0, I_p)$

$$E(\langle X-\theta, \hat\theta\rangle) \qquad E(\langle Z, f(Z)\rangle) = E(\nabla \cdot f(Z))$$

$$= E(\langle Z, \hat\theta(\theta+Z)\rangle)$$

$$= E \sum_{i=1}^{p} \frac{\partial}{\partial z_i} \hat\theta(\theta+Z)$$

$$= E\left(\sum_{i=1}^{p} \hat\theta_i(X)\right)$$

$$\Rightarrow \text{SURE}(\hat\theta) = \|\hat\theta - X\|^2 - p + 2\overbrace{\sum_{i=1}^{p} \frac{\partial}{\partial x_i}\hat\theta_i(X)}^{\text{degrees of freedom of } \hat\theta}$$

$\searrow$ Stein's unbiased risk estimate

$$E_\theta \, \text{SURE}(\hat\theta) = E_\theta \|\hat\theta - \theta\|^2$$

Applications

① $\hat\theta = cX$ $\hat c = \underset{c}{\operatorname{argmin}}\, \text{SURE}(cX)$ function of X

② Linear model $y \sim N(\theta, I_n)$

$$\text{SURE}(\hat\theta) = \|y - \hat\theta\|^2 - n + 2\sum_{i=1}^{n} \frac{\partial}{\partial y_i}\hat\theta_i$$

$X \in \mathbb{R}^{n\times p}$ $\hat\beta = \underset{\beta}{\operatorname{argmin}} \|y - X\beta\|^2$ $\hat\theta = X\hat\beta = \underbrace{X(X^TX)^{+}X^T}_{\text{projection}} y$

$= Hy$

$$\sum_{i=1}^{n} \frac{\partial}{\partial y_i} \hat{\theta}_i = \sum_{i=1}^{n} H_{ii} = Tr(H) = p$$

or $rank(X)$ if not full rank

$Tr\left(X(X^TX)^{-1}X^T\right)$

$= Tr\left((X^TX)^{-1}X^TX\right)$

$= Tr(I) = p$

$$SURE(X\hat{\beta}) = \underbrace{\|y - X\hat{\beta}\|^2}_{\text{encourage large model}} + \underbrace{2p}_{\text{penalize large model}} - n$$

= "AIC"

Akaike

$= E(\|X\hat{\beta} - \theta\|^2)$

Lecture 11

Best Equivariant Estimator

- Location model

$(P_\theta \ \ \theta \in \mathbb{R}^p) \quad X \sim P_\theta \iff X-\theta \sim P_0$

- Assuming density exists, we have

$$P_\theta(X \le x) = P_\theta(X-\theta \le x-\theta)$$

$$= P_0(X \le x-\theta)$$

density ft of $X \sim P_0$

$$P_\theta(x) = \frac{d}{dx} P_\theta(X \le x) = \frac{d}{dx} P_0(X \le x-\theta) = f(x-\theta)$$

Equivariant Estimator

$$\hat{\theta}(X_1+c, X_2+c, \cdots, X_n+c) = \hat{\theta}(X_1, X_2, \cdots, X_n) + c$$

Location invariant loss

$$L(\hat{\theta}+c, \theta+c) = L(\hat{\theta}, \theta)$$

Thm location model & Equivariant Estimator & Location invariant loss

$\Rightarrow$ risk is Constant

pf) $E_\theta[L(\hat{\theta}, \theta)]$

$= E_\theta(\rho(\hat{\theta}-\theta))$

$= E_\theta(\rho(\hat{\theta}((X_1-\theta, \cdots, X_n-\theta)))$

$= E_0(\rho(\hat{\theta}(X_1, \cdots X_n)) = \text{constant}$

Q. Does this mean $L(\hat{\theta}, \theta)$ doesn't depend on θ?

Not really

MREE: minimal risk equivariant estimator

$\delta(x) = \delta_0(x) - u(x)$

$u(x+c) = \delta_0(x+c) - \delta(x+c)$

$= \delta_0(x) - \delta(x)$

$= u(x)$

$\Rightarrow u(x_1+c, \dots, x_n+c) = u(x_1, \dots, x_n)$

$c = -x_n$

$\Rightarrow u(x_1, \dots, x_n) = \underbrace{u(x_1-x_n, \dots, x_{n-1}-x_n, 0)}$

$Y = (x_1-x_n, \dots, x_{n-1}-x_n)$

$u(x) = v(Y)$

$\delta(x) = \delta_0(x) - v(Y)$

$E_\theta(\rho(\delta(x)-\theta)) = E_0(\rho(\delta(x)))$

$= E_0\, \rho(\delta_0(x) - v(Y))$

$= E_0\left(E_0(\rho(\delta_0(x) - v(Y)) \mid Y)\right)$

$\delta(x) = \delta_0(x) - v(Y) = \delta_0(x) - \underset{a}{\operatorname{argmin}}\ E_0(\rho(\delta_0(x) - a) \mid Y)$ seems to be unique

If $\rho(t) = t^2$, $v(Y) = E_0(\delta_0(x) \mid Y)$

(ex) $X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1)$

$\delta_0(x) = \bar{X}$: sufficient, complete

$v(Y) = E_0(\bar{X} \mid Y)$ ancillary

$= E_0(\bar{X})$

$= 0$

$\delta(X) = \bar{X}$

(ex) $X_1 \cdots X_n \overset{iid}{\sim} E(\theta, b)$

density $\frac{1}{b} e^{-\frac{x-\theta}{b}} 1_{\{x > \theta\}}$

$\delta_0(X) = X_{(1)}$

$E_\delta(X_{(1)} \mid Y) = E_0(X_{(1)})$

$X_i = \theta + bZ_i$ where $Z_i \sim e^{-x} 1_{\{x>0\}}$

$X_{(1)} = \theta + bZ_{(1)}$

$\mathbb{P}(Z_{(1)} \leq t) = 1 - \mathbb{P}(Z_{(1)} > t)$

$= 1 - \prod P(Z_i > t)$

$= 1 - e^{-nt}$

$nt = y \quad ndt = dy$

$\Rightarrow$ density $Z_{(1)}$ is ne^{-nt} with $\int_0^\infty t n e^{-nt} dt = \int_0^\infty y e^{-y} \frac{1}{n} dy$

$= \frac{1}{n} \left. -ye^{-y} \right]_0^\infty - \left. e^{-y} \right]_0^\infty$

$= \frac{1}{n} \Rightarrow$ expected value

$E_0(X_{(1)}) = b \frac{1}{n} = \frac{b}{n}$

$\delta(X) = X_{(1)} - \frac{b}{n}$ is MREE

pitman estimator

$\delta_0(X) = X_n$

$\delta(X) = X_n - E_0(X_n \mid Y)$

$$Z = \begin{pmatrix} Y \\ X_n \end{pmatrix} = \begin{pmatrix} X_1 - X_n \\ \vdots \\ X_{n-1} - X_n \\ X_n \end{pmatrix}$$

$$\begin{cases} x_1 = z_1 + z_n \\ \vdots \\ x_{n-1} = z_{n-1} + z_n \\ x_n = z_n \end{cases} \qquad \left|\frac{dx}{dz}\right| = \left|\begin{pmatrix} 1 & \cdots & 1 \\ & \ddots & \vdots \\ 0 & & 1 \end{pmatrix}\right| = 1$$

$$P_{z,0}(z_1 \cdots z_n) = P_{x,0}(x_1 \cdots x_n)\left|\frac{dx}{dz}\right|$$

$$= P_{x,0}(x_1 \cdots x_n)$$

$$= \prod_{i=1}^{n} f(x_i)$$

$$= f(Y_1 + x_n) f(Y_2 + x_n) \cdots f(Y_{n-1} + x_n) f(x_n)$$

$$E_0(x_n \mid Y) = \int x_n P_{x,0}(x_n \mid Y)\, dx_n$$

$$= \frac{\int x_n P_{x,0}(Y, x_n)\, dx_n}{\int P_{x,0}(Y, x_n)\, dx_n}$$

$$= \frac{\int t f(Y_1 + t) \cdots f(Y_{n-1} + t) f(t)\, dt}{\int f(Y_1 + t) \cdots f(Y_{n-1} + t) f(t)\, dt}$$

$$\delta(x) = x_n - E_0(x_n \mid Y)$$

$$= x_n - \frac{\int t f(Y_1 + t) \cdots f(Y_{n-1} + t) f(t)\, dt}{\int f(Y_1 + t) \cdots f(Y_{n-1} + t) f(t)\, dt}$$

$$= \frac{\int (x_n - t) f(Y_1 + t) \cdots f(Y_{n-1} + t) f(t)\, dt}{\int f(Y_1 + t) \cdots f(Y_{n-1} + t) f(t)\, dt}$$

$x_n - t = u$

$$= \frac{\int u f(x_1 - u) \cdots f(x_n - u)\, du}{\int f(x_1 - u) \cdots f(x_n - u)\, du}$$

$$\delta(x) = \frac{\int u \prod_{i=1}^{n} f(x_i - u)\, du}{\int \prod_{i=1}^{n} f(x_i - u)\, du}$$

"posterior mean under uniform prior even if it is improper"

e.g. $X_1 \cdots X_n \sim \text{Unif}\left(\theta - \frac{b}{2}, \theta + \frac{b}{2}\right)$

density is $\frac{1}{b} \mathbb{1}_{\{\theta - \frac{b}{2} < x_i < \theta + \frac{b}{2}\}}$

$\theta = 0$

$$f(x_i - u) = \frac{1}{b} \mathbb{1}_{\{u - \frac{b}{2} < x_{(i)} < u + \frac{b}{2}\}}$$

$$\prod_{i=1}^{n} f(x_i - u) = \frac{1}{b^n} \mathbb{1}_{\{u - \frac{b}{2} < x_{(1)}\}} \mathbb{1}_{\{x_{(n)} < u + \frac{b}{2}\}}$$

$$= \frac{1}{b^n} \mathbb{1}_{\{x_{(n)} - \frac{b}{2} < u < x_{(1)} + \frac{b}{2}\}}$$

$$\delta(x) = \frac{\int_{x_{(n)} - \frac{b}{2}}^{x_{(1)} + \frac{b}{2}} u\, du}{\int_{x_{(n)} - \frac{b}{2}}^{x_{(1)} + \frac{b}{2}} du}$$

$$= \frac{\frac{u^2}{2} \Big]_{x_{(n)} - \frac{b}{2}}^{x_{(1)} + \frac{b}{2}}}{\left(x_{(1)} + \frac{b}{2}\right) - \left(x_{(n)} - \frac{b}{2}\right)}$$

$$= \frac{1}{2}\left(x_{(1)} + x_{(n)}\right) \quad \text{is MREE}$$

Lecture 12

Location family $X \sim P_\theta \iff X - \theta \sim P_0$

Lemma. for $\rho(t) = t^2$ $\qquad E_0\,\delta(X-\theta) \overset{\shortparallel}{} = E_0(\delta(X))$

① $\delta(X)$ equivariant then its bias b is constant

and $\delta(X) - b$ has smaller risk

② MREE is unbiased

③ If UMVUE exists and is equivariant, then MREE = UMVUE

Location scale model

$X \sim P_{\theta,\tau} \iff \frac{X-\theta}{\tau} \sim P_{0,1}$ $\quad \frac{1}{\tau} f(\frac{x-\theta}{\tau})$ $\quad$ (ex) Cauchy distribution, τ not variance

$\hat\theta(aX_1+b, aX_2+b, \cdots, aX_n+b) = a\hat\theta(X)+b$ $\qquad$ location-scale equivariance

$(a>0 \quad b\in\mathbb{R})$

$L(a\hat\theta+b, a\theta+b, a\tau) = L(\hat\theta, \theta, \tau)$

Location-scale invariance of loss

$L(\hat\theta, \theta, \tau) = L(\hat\theta-\theta, 0, \tau)$

$= L(\frac{\hat\theta-\theta}{\tau}, 0, 1)$

$= \rho(\frac{\hat\theta-\theta}{\tau})$

fact. Risk is constant

$E_{\theta,\tau}\,\rho\left(\frac{\hat\theta(X_1,\cdots,X_n)-\theta}{\tau}\right) = E_{\theta,\tau}\,\rho\left(\hat\theta\left(\frac{X_1-\theta}{\tau}, \frac{X_2-\theta}{\tau}, \cdots, \frac{X_n-\theta}{\tau}\right)\right)$

$= E_{0,1}\,\rho(\hat\theta(X))$

$\delta(X)$ is location-scale equivariant

$\delta_0(aX_1+b, \cdots, aX_n+b) = a\delta_0(X)+b$ $\qquad \frac{\delta(X)-\delta_0(X)}{\delta_1(X)}$ only depends on ancillary information

$\delta_1(aX_1+b, \cdots, aX_n+b) = a\delta_1(X)$

$$\frac{\delta(aX_1+b,\dots,aX_n+b)-\delta_0(aX_1+b,\dots,aX_n+b)}{\delta_1(aX_1+b,\dots,aX_n+b)}=\frac{\delta(x)-\delta_0(x)}{\delta_1(x)}=u(x)$$

$$u(aX+b)=u(X)$$

$$u(X_1\cdots X_n)=u(X_1-X_n,\dots,X_{n-1}-X_n,0)$$

$$=u\Big(\underline{\frac{X_1-X_n}{|X_1-X_n|},\frac{X_2-X_n}{|X_1-X_n|},\dots,\frac{X_{n-1}-X_n}{|X_1-X_n|}},0\Big)$$

$$Z=\Big(\frac{X_1-X_n}{|X_1-X_n|},\frac{X_2-X_n}{|X_1-X_n|},\dots,\frac{X_{n-1}-X_n}{|X_1-X_n|}\Big)$$ is ancillary of (θ,τ)

$$\Rightarrow \delta(x)=\delta_0(x)-W(Z)\delta_1(x)$$

$$E_{0,1}\,\rho(\delta(x))=E_{0,1}\,\rho(\delta_0(x)-W(Z)\delta_1(x))$$

$$=E_{0,1}\,E_{0,1}\big(\rho(\delta_0(x)-W(Z)\delta_1(x))\mid Z\big)$$

$$W(Z)=\underset{a}{\arg\min}\;E_{0,1}\big(\rho(\delta_0(x)-a\delta_1(x))\mid Z\big)$$

When $\rho(t)=t^2$ $\quad W(Z)=\underset{a}{\arg\min}\;E_{0,1}\big((\delta_0(x)-a\delta_1(x))^2\mid Z\big)$

$$=\underset{a}{\arg\min}\;E_{0,1}\big[a^2E_{0,1}(\delta_1^2(x)\mid Z)-2a\,E_{0,1}(\delta_0(x)\delta_1(x)\mid Z)\big]$$

$$W(Z)=\frac{E_{0,1}(\delta_0(x)\delta_1(x)\mid Z)}{E_{0,1}(\delta_1^2(x)\mid Z)}$$ ... Similar to $\min_\beta \frac{1}{n}\sum(y_i-\beta x_i)^2 \Rightarrow \hat\beta=\dfrac{\frac{1}{n}\sum x_iy_i}{\frac{1}{n}\sum x_i^2}$

$$\Rightarrow \delta(x)=\delta_0(x)-\frac{E_{0,1}(\delta_0(x)\delta_1(x)\mid Z)}{E_{0,1}(\delta_1^2(x)\mid Z)}\delta_1(x)$$

e.g. $\quad X_1\cdots X_n$ iid $\quad \frac{1}{b}e^{-\frac{x-\theta}{b}}I_{\{x>\theta\}}$

$$\Big(\underbrace{X_{(1)}}_{\text{suff for }\theta},\ \sum_{i=1}^n X_i-X_{(1)}\Big)$$ complete, suff for (θ,b) & Z is ancillary for (θ,b)

$\downarrow$ $\qquad\qquad$ $\downarrow$

$\delta_0(X)$ $\qquad\qquad$ $\delta_1(X)$

$$\delta(X) = \delta_0(X) - \frac{E_{0,1}(\delta_0(X)\delta_1(X) \mid Z)}{E_{0,1}(\delta_1^2(X) \mid Z)}\,\delta_1(X)$$

$E_{0,1}(\delta_0(X))\,E_{0,1}(\delta_1(X))$

$$= \delta_0(X) - \frac{E_{0,1}(\delta_0(X)\delta_1(X))}{E_{0,1}(\delta_1^2(X))}\,\delta_1(X)$$

Under $P_{0,1}$ $\quad X_1 \cdots X_n \overset{iid}{\sim} e^{-x} 1_{\{x>0\}}$

$X_{(1)} \sim n e^{-nx} 1_{\{x>0\}}$

$E_{0,1} X_{(1)} = \frac{1}{n}$

$$f_{X_{(1)}\cdots X_{(n)}}(x_1 \cdots x_n) = n!\, f(x_1 \cdots x_n)\, 1_{\{0 \le x_1 \le x_2 \le \cdots \le x_n\}}$$

$$= n!\, e^{-\sum x_i}\, 1_{\{0 \le x_1 \le x_2 \le \cdots \le x_n\}}$$

$Y_1 = n X_{(1)}$

$Y_2 = (n-1)(X_{(2)} - X_{(1)})$

$Y_3 = (n-2)(X_{(3)} - X_{(2)})$

$\vdots$

$Y_n = 1 \cdot (X_{(n)} - X_{(n-1)})$

$$f_{Y_1 \cdots Y_n}(y_1 \cdots y_n) = e^{-\sum y_i}\, 1_{\{y_1 \cdots y_n \ge 0\}}$$

$$= \prod_{i=1}^{n} e^{-y_i}\, 1_{\{y_i \ge 0\}}$$

$\left|\frac{dY}{dX}\right| = n!$

$$\delta_1(X) = \sum_{i=1}^{n} (X_i - X_{(1)}) = \sum X_i - \sum X_{(1)} = \sum_{i=2}^{n} y_i$$

Memoryless property

$P(X > t+s \mid X > s) = P(X > t) \quad \forall t, s \ge 0$

$E_{0,1}(\delta_1(X)) = n-1$

$$E_{0,1}(\delta_1^2(X)) = Var_{0,1}(\delta_1(X)) + E_{0,1}(\delta_1(X))$$

$$= (n-1) + (n-1)^2$$

$$= n(n-1)$$

$$W(z) = \frac{\frac{1}{n}(n-1)}{n(n-1)} = \frac{1}{n^2}$$

$\delta(X) = X_{(1)} - \frac{1}{n^2}\sum_{i=1}^{n}(X_i - X_{(1)})$ is MRE

Lemma. If $\delta(x)$ is the best location-equivariant estimate for θ when τ is known and
$\begin{cases} \delta(x) \text{ does not depend on } \tau \\ \delta(x) \text{ is also location-scale equivariant} \end{cases}$

then $\delta'(x)$ is the best location-scale equivariant estimate for θ with unknown τ

pf) for fixed τ. constant $\Rightarrow$ true for $\forall \tau$

(ex) Normal distri case

Scale parameter

$\begin{cases} \text{Scale family} \quad X \sim P_\tau \Leftrightarrow \frac{X}{\tau} \sim P_1 \\ \text{location-scale family} \quad X \sim P_{\theta,\tau} \Leftrightarrow \frac{X-\theta}{\tau} \sim P_{0,1} \end{cases}$

goal. estimate τ^r $\quad \delta(aX_1 \cdots aX_n) = a^r \delta(x)$ $\quad$ scale

$\delta(aX_1+b \cdots aX_n+b) = a^r \delta(X)$ $\quad$ location-scale

Loss $\quad L(a^r \delta(x), a\tau) = L(\delta(x), \tau)$

$= L\left(\frac{\delta(x)}{\tau^r}, 1\right) = \gamma\left(\frac{\delta(x)}{\tau^r}\right)$

e.g. $\gamma(t) = (t-1)^2$ close to 1

$\gamma(t) = |t-1|$

$L(\delta, \tau) = \frac{(\delta - \tau^r)^2}{\tau^{2r}}$

fact: risk function is constant under both cases

$E_{\theta,\tau} \frac{(\delta - \tau^r)^2}{\tau^{2r}}$ $\qquad$ $E_\tau \frac{(\delta - \tau^r)^2}{\tau^{2r}}$

$= E_{0,1} \frac{(\delta(\tau(X+\theta)) - \tau^r)^2}{\tau^{2r}}$ $\quad$ ($= \tau^r \delta(x)$) $\qquad$ $= E_1 \frac{(\delta(\tau X) - \tau^r)^2}{\tau^{2r}}$

$= E_{0,1} (\delta(X) - 1)^2$ $\qquad$ $= E_1 (\delta(X) - 1)^2$

should ignore θ essentially b.c. we estimate τ^r

Lecture 13

1) $X \sim P_\tau \iff X/\tau \sim P_1$ scale family 2) $\delta(aX) = a^r \delta(X)$

$X \sim P_{\theta,\tau} \iff \frac{X-\theta}{\tau} \sim P_{0,1}$ location-scale family $\delta(aX+b) = a^r \delta(X)$

3) loss $\gamma\left(\frac{\delta(X)}{\tau^r}\right)$ e.g. $\gamma(t) = (t-1)^2$

① Scale family $\delta_0(aX) = a^r \delta_0(X)$ $\frac{\delta(X)}{\delta_0(X)} = W(X)$ $W(aX) = W(X)$

$$W(X_1 \cdots X_n) = W\Big(\underbrace{\frac{X_1}{|X_n|}, \frac{X_2}{|X_n|}, \cdots \frac{X_n}{|X_n|}}_{Z}\Big) = W(Z)$$

$\delta(X) = \delta_0(X) W(Z)$ $E_1 \gamma(\delta_0(X) W(Z))$

$$W(Z) = \arg\min_a E_1\left(\gamma(a\delta_0(X)) \mid Z\right)$$

covariate response

when $\gamma(t) = (t-1)^2$ $W(Z) = \arg\min_a E_1\left((\delta(X)a - 1)^2 \mid Z\right)$

$$= \frac{E_1(\delta(X) \mid Z)}{E_1(\delta^2(X) \mid Z)}$$

e.g. $X_1 \cdots X_n$ iid $\sim N(0, \sigma^2)$

$\delta(X) = \sum_{i=1}^n X_i^2 \overset{\sigma^2=1}{\sim} \chi^2_n$ $W(Z) = \frac{n}{n^2+2n} = \frac{1}{n+2}$ $\delta(X) = \frac{1}{n+2} \sum X_i^2$

for scale parameter we usually get biased estimate

② Location scale family

$\delta(aX+b) = a^r \delta(X)$ $W(X) = \frac{\delta(X)}{\delta_0(X)}$ $W(aX+b) = W(X)$

$$W(X_1 \cdots X_n) = W(X_1 - X_n, X_2 - X_n, \cdots, X_{n-1} - X_n, 0) = W\Big(\underbrace{\frac{X_1 - X_n}{|X_1 - X_n|}, \frac{X_2 - X_n}{|X_1 - X_n|}, \cdots, \frac{X_{n-1} - X_n}{|X_1 - X_n|}}_{Z}, 0\Big)$$

$= U(Z)$

$\delta(X) = \delta_0(X) U(Z)$

a function of pairwise difference

$u(z) = \underset{a}{\text{argmin}}\ E_{0,1}[r(\delta_0(X)a) \mid z)$

when $r(t) = (t-1)^2 \Rightarrow u(z) = \dfrac{E_{0,1}(\delta(X) \mid z)}{E_{0,1}(\delta^2(X) \mid z)}$

e.g. $X_1 \cdots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ $\quad \delta_0(X) = \sum_{i=1}^{n} (X_i - \bar{X})^2$

$\theta = 0\ \sigma^2 = 1 \qquad \sim \chi^2_{n-1}$

$$u(z) = \frac{n-1}{(n-1)^2 + 2(n-1)} = \frac{1}{n+1}$$

$$\Rightarrow \delta(X) = \frac{1}{n+1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

we'll get this when we consider $c_- \sim$ (?)

Likelihood method and asymptotic theory

MLE $X_1 \cdots X_n \overset{iid}{\sim} P_{\theta^*}$

$\hat{\theta} = \underset{\theta}{\text{argmax}} \sum_{i=1}^{n} \log P_\theta(X_i)$

WTS Consistency $P_{\theta^*}(\|\hat{\theta} - \theta^*\|^2 > \varepsilon) \xrightarrow{n \to \infty} 0$? $\forall \varepsilon > 0$

$$\max_\theta \frac{1}{n} \sum_{i=1}^{n} \log P_\theta(X_i) \iff \min_\theta \frac{1}{n} \sum_{i=1}^{n} \log \frac{1}{P_\theta(X_i)}$$

$$\iff \min_\theta \frac{1}{n} \sum_{i=1}^{n} \log \frac{P_{\theta^*}(X_i)}{P_\theta(X_i)}$$

$\frac{1}{n} \sum_{i=1}^{n} \log \frac{P_{\theta^*}(X_i)}{P_\theta(X_i)} \xrightarrow{\text{LLN}} \int P_{\theta^*} \log \frac{P_{\theta^*}}{P_\theta} = D(P_{\theta^*} \| P_\theta)$ & $\text{argmin} \frac{1}{n} \sum_{i=1}^{n} \log \frac{P_{\theta^*}(X_i)}{P_\theta(X_i)} = \hat{\theta}$

$\text{argmin}\ D(P_{\theta^*} \| P_\theta) = \theta^*$

Def. $D(P \| Q) = \int p \log \frac{p}{q}$

relative entropy, Kullback-Leibler divergence

$D(P \| Q) \geq 0$ $\quad D(P \| Q) = \int p \log \frac{p}{q} = \int q \left(\frac{p}{q} \log \frac{p}{q}\right)$

$= \int q\, f\left(\frac{p}{q}\right) \quad f(t) = t \log t$

$= \underset{X \sim q}{E} f\left(\frac{p(X)}{q(X)}\right) \quad \begin{cases} f \text{ is convex} \\ f(1) = 0 \end{cases}$

$\geq f\left(\underset{X \sim q}{E} \left(\frac{p(X)}{q(X)}\right)\right)$

$= 0$

Stronger result: $D(P\|Q) \geq \frac{1}{2}\underbrace{\int(\sqrt{p}-\sqrt{q})^2}_{\text{squared Hellinger distance}}$

< Coding theory >

Identifiability $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$

Ass ① a quantitative version $\forall \varepsilon > 0 \ \exists \delta > 0$ s.t. $\|\theta_1 - \theta_2\| > \varepsilon \Rightarrow D(P_{\theta_1}\|P_{\theta_2}) > \delta$

< e.g. $P_\theta = P \ \forall \theta$ >

$$P_{\theta^*}(\|\hat\theta - \theta^*\| > \varepsilon) \leq P_{\theta^*}(D(P_{\theta^*}\|P_{\hat\theta}) > \delta)$$

$$= P_{\theta^*}\left(\int P_{\theta^*}\log P_{\theta^*} - P_{\theta^*}\log P_{\hat\theta} > \delta\right)$$

$$\underset{\frac{1}{n}\sum \log P_{\theta^*}(X_i)}{\approx} \quad \overset{\text{by def}}{(\geq)} \quad \underset{\frac{1}{n}\sum \log P_{\hat\theta}(X_i)}{\approx}$$

not iid

VC dimension
covering number
Dudley integral
Radem complexity

empirical process: Van Der Vaart

Ass ② uniform law of large number

$$\sup_\theta \left| \frac{1}{n}\sum_{i=1}^n P_\theta(X_i) - \int P_{\theta^*}\log P_\theta \right| \xrightarrow{P_{\theta^*}} 0$$

e.g. $X_1 \cdots X_n \overset{\text{iid}}{\sim} F$ cdf

$$\sup_t |\hat F(t) - F(t)| \xrightarrow{P} 0 \quad \text{Glivenko-Cantelli}$$

$\Downarrow$

$$P_{\theta^*}\left(\sup_\theta \left|\frac{1}{n}\sum_{i=1}^n \log P_\theta(X_i) - \int P_{\theta^*}\log P_\theta\right| > \frac{\delta}{2}\right) \to 0$$

$$D(P_{\theta^*}\|P_{\hat\theta}) = \int P_{\theta^*}\log P_{\theta^*} - \int P_{\theta^*}\log P_{\hat\theta}$$

$$\leq \underbrace{\frac{1}{n}\sum \log P_{\theta^*}(X_i) - \frac{1}{n}\sum\log P_{\hat\theta}(X_i)}_{\leq 0} + \left|\frac{1}{n}\sum\log P_{\theta^*}(X_i) - \int P_{\theta^*}\log P_{\theta^*}\right| + \left|\frac{1}{n}\sum \log P_{\hat\theta}(X_i) - \int P_{\theta^*}\log P_{\hat\theta}\right|$$

$$\leq 2\sup_\theta \left|\frac{1}{n}\sum\log P_\theta(X_i) - \int P_{\theta^*}\log P_\theta\right|$$

$$\Rightarrow \quad \mathbb{P}_{\theta^*}(D(P_{\theta^*}\|P_{\hat\theta}) > \delta) \leq \mathbb{P}_{\theta^*}\left(\sup_\theta \left|\frac{1}{n}\sum\log P_\theta(X_i) - \int P_{\theta^*}\log P_\theta\right| > \delta/2\right) \to 0$$

Lecture 14

$$X_1 \dots X_n \overset{iid}{\sim} P_\theta \qquad \hat\theta = \arg\max_\theta \frac{1}{n}\sum_{i=1}^n \log P_\theta(X_i)$$

$$P_{\theta^*}(\|\hat\theta - \theta^*\| > \delta) \to 0 \quad \text{as} \quad n \to \infty \quad \forall \delta > 0$$

$$\Updownarrow$$

$$\|\hat\theta - \theta^*\| = o_p(1)$$

Convergence in probability: does not penalize how much deviation we had

Score function $S_\theta(x) = \frac{d}{d\theta} \log P_\theta(x)$ gradient

$E_\theta S_\theta(X) = 0$

↳ true theta

1) intuition MLE check

2) KL $\theta = \theta^*$ minimize $\left(E_{\theta^*} \log \frac{P_{\theta^*}}{P_\theta} \geq 0\right)$

$$\frac{d}{d\theta'} E_\theta \left(\log \frac{P_\theta}{P_{\theta'}}\right)\Big|_{\theta'=\theta} = 0$$

$$\Rightarrow \frac{d}{d\theta'} \int P_\theta \log \frac{P_\theta}{P_{\theta'}}\Big|_{\theta'=\theta} = 0$$

3) direct computation

. Fisher information matrix $I_\theta = Var_\theta(S_\theta(X)) = E_\theta(S_\theta(X)^2)$

(Intuition: $D_{KL}(f(x;\theta) \| f(x;\theta+\delta)) \approx \frac{1}{2} I(\theta)\delta^2$)

Have to specify, with respect to what parameter?

Alternative formula (may not be always true)

$$I_\theta = -E_\theta\left(\frac{d^2}{d\theta^2} \log P_\theta(X)\right)$$

$$\frac{d^2}{d\theta^2}\log P_\theta = \frac{d}{d\theta}\left(\frac{\frac{d}{d\theta}P_\theta}{P_\theta}\right) = \frac{\frac{d^2}{d\theta^2}P_\theta \, P_\theta - (\frac{d}{d\theta}P_\theta)^2}{P_\theta^2}$$

$$E_\theta\left(\frac{d^2}{d\theta^2}\log P_\theta(X)\right)$$

$$= \int \cancel{\frac{d^2}{d\theta^2}P_\theta} - \frac{(\frac{d}{d\theta}P_\theta)^2}{P_\theta}$$

$$= -\int \left(\frac{\frac{d}{d\theta}P_\theta}{P_\theta}\right)^2 P_\theta$$

$$= -E_\theta(\log P_\theta(X))$$

$$\frac{1}{n}\sum \log p_\theta(x_i) \approx \frac{1}{n}\sum \log p_{\theta^*}(x_i) + (\theta-\theta^*)\frac{1}{n}\frac{d}{d\theta}\log p_\theta(x_i)\Big|_{\theta=\theta^*}$$

$$+\frac{1}{2}(\theta-\theta^*)^2\left(\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log p_\theta(x_i)\Big|_{\theta=\theta^*}\right) \xrightarrow{\text{LLN}} -I_{\theta^*}$$

$$\approx \frac{1}{n}\sum \log p_{\theta^*}(x_i) + \underline{(\theta-\theta^*)\frac{1}{n}S_{\theta^*}(x_i) - \frac{1}{2}(\theta-\theta^*)^2 I_{\theta^*} = \tilde{L}_n(\theta)}$$

concave

$$\tilde{\theta} = \arg\max_\theta \tilde{L}_n(\theta)$$

$$\frac{1}{n}S_{\theta^*}(x_i) - (\tilde{\theta}-\theta^*) I_{\theta^*} = 0$$

$$\tilde{\theta}-\theta^* = \frac{1}{n}\sum S_{\theta^*}(x_i)/I_{\theta^*}$$

$$\sqrt{n}(\tilde{\theta}-\theta^*) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{\theta^*}(x_i)\Big/ I_{\theta^*} \overset{\text{CLT}}{\sim} N\left(0, \mathrm{Var}\left(\frac{S_\theta(x)}{I_\theta^*}\right)\right)$$

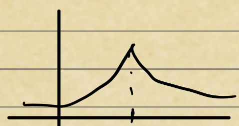
$$= N(0, I_{\theta^*}^{-1})$$

Le Cam: proved this w.o. ∃ second derivative

$\sqrt{n}(\hat{\theta}-\theta^*) \sim N(0, I_{\theta^*}^{-1})$ More info you have, more accurate it becomes

cvg rate : $\frac{1}{\sqrt{n}}$ (asymptotic normality)   In decision theory, minimax $\frac{1}{n}$ "parametric rate"

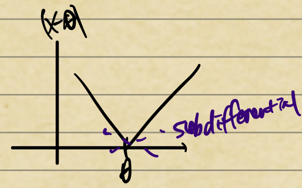
e.g. $X_1 \cdots X_n \overset{iid}{\sim} \frac{1}{2}e^{-|x-\theta|}$



$$\hat{\theta} = \arg\min_\theta \sum_{i=1}^{n}|X_i-\theta|$$

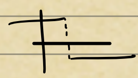
$$= \text{median}(X_1 \cdots X_n)$$

Likelihood $= -|x-\theta| - \log 2$



$$S_\theta(x) = -1_{\{x<\theta\}} + 1_{\{x\geq\theta\}}$$

$$I_\theta = E_\theta(S_\theta(x))^2 = 1 \qquad \frac{d^2}{d\theta^2}\log p_\theta(x)$$



$$= -\frac{d}{d\theta}S(x) = \begin{cases} 0 & \theta \neq x \\ \text{"}\infty\text{"} & \theta = x \end{cases}$$

alternative formulation does not work

$\|\hat{\theta}-\theta\|$

MLE uniform cvg $\frac{1}{n}$ } uniform is special, ∄ score, fisher

minimax for squared loss $\frac{1}{n^2}$

$\|\hat{\theta}-\theta\|^2$

Convergence rate is exponential, when we only have two choices ⇐ check

hw4 P5

Idea by pollard (first LLN then second derivative) integration given some "smoothness"

$$\frac{1}{n_1}\sum_{i=1}^{n}\log P_\theta(X_i) \to \int P_{\theta^*}\log P_\theta = E_{\theta^*}(-|X-\theta|-\log 2)$$

$$\frac{d}{d\theta}\int P_{\theta^*}\log P_\theta = E_{\theta^*}\left(-1_{\{X<\theta\}}+1_{\{X>\theta\}}\right) = -P_{\theta^*}(X<\theta)+P_{\theta^*}(X>\theta)$$

$$= 1-2P_{\theta^*}(X<\theta)$$

$$\frac{d^2}{d\theta^2}\int P_{\theta^*}\log P_\theta = -2\,P_{\theta^*}(\theta^*) = -1 \Rightarrow -\frac{d^2}{d\theta^2}\int P_{\theta^*}\log P_\theta\Big|_{\theta=\theta^*} = I_{\theta^*}$$

$$\begin{cases} L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\log P_\theta(X_i) \\ L(\theta) = E_{\theta^*}\log P_\theta(X) = \int P_{\theta^*}\log P_\theta \end{cases}$$

Empirical process

$$\nu_n f = \frac{\frac{1}{n}\Sigma\left(f(X_i)-E(f(X))\right)}{\frac{1}{\sqrt{n}}}$$

$$\sqrt{n}\,\frac{1}{n}\sum_{i=1}^{n}\left(\log P_\theta(X_i) - E_{\theta^*}\log P_\theta(X)\right) = \nu_n \log P_\theta = \sqrt{n}\left(L_n(\theta)-L(\theta)\right)$$

first-order expansion $\log P_\theta(X)$

$$\log P_{\theta^*+t}(X) = \log P_{\theta^*}(X) + t\,S_{\theta^*}(X) + \underbrace{|t|\,r(X,t)}_{\text{reminder}}$$

Second-order expansion $L(\theta)$

(regular Taylor expansion)

$$L(\theta^*+t) = L(\theta^*) - \frac{1}{2}I_{\theta^*}t^2 + o(t^2) + \cdots$$

$$\frac{t^2}{2}\frac{d^2}{d\theta^2}\log P_\theta\Big|_{\theta=\theta^*+t}$$

$$L_n(\theta^*+t) = L(\theta^*+t) + \frac{1}{\sqrt{n}}\nu_n \log P_{\theta^*+t}$$

$$= L(\theta^*) - \frac{1}{2}I_{\theta^*}t^2 + o(t^2) + \frac{1}{\sqrt{n}}\nu_n\left(\log P_{\theta^*} + t\,S_{\theta^*} + |t|\,r(\cdot,t)\right)$$

$$= \left(L(\theta^*) + \frac{1}{\sqrt{n}}\nu_n\log P_{\theta^*}\right) + t\,\frac{1}{\sqrt{n}}\nu_n S_{\theta^*}(X) - \frac{1}{2}I_{\theta^*}t^2 + o(t^2) + \frac{|t|}{\sqrt{n}}\nu_n r(\cdot,t)$$

$$= L_n(\theta^*) + t\,\frac{1}{n}\sum_{i=1}^{n}S_{\theta^*}(X_i) - \frac{1}{2}I_{\theta^*}t^2 + o(t^2) + \frac{|t|}{\sqrt{n}}\nu_n r(\cdot,t)$$

Stochastic differentiability $\sup_{t\in U_n}\frac{|\nu_n r(\cdot,t)|}{1+|\sqrt{n}\,t|} \xrightarrow{P_{\theta^*}} 0$ for every shrinking $U_n \to 0$

Lec 15

$X_1, \dots, X_n \overset{iid}{\sim} P_{\theta^*}$ $\quad \hat\theta = \underset{\theta}{\arg\max}\, L_n(\theta) \quad L_n(\theta) = \frac{1}{n}\sum \log p_\theta(X_i)$

Assumptions: ① $\|\hat\theta - \theta\| = o_p(1)$ $\qquad$ score $S_\theta(x) = \frac{d}{d\theta}\log p_\theta(x)$

② $\log p_{\theta^*+t}(x) = \log p_{\theta^*}(x) + t\, S_{\theta^*}(x) + |t|\, r(x,t)$

$$\sup_{t\in U_n} \frac{|\nu_n r(\cdot,t)|}{1+|\sqrt{n}\, t|} = O_p(1)$$

for any shrinking nbhd $U_n \ni 0$

③ $L(\theta) = E_{\theta^*} L_n(\theta) = \int p_{\theta^*}\log p_\theta$

$L(\theta^*+t) = L(\theta^*) - \frac{1}{2} I_{\theta^*} t^2 + o(t^2)$

consistent $\hat\theta$ need θ^* to be MLE $\qquad I_{\theta^*} = E_{\theta^*}(S_{\theta^*}(X))^2$

Step 1: derive a quadratic expansion for $L_n(\theta)$

$L_n(\theta^*+t) = L(\theta^*+t) + \frac{1}{\sqrt{n}}\nu_n \log p_{\theta^*+t}$

$= L(\theta^*) - \frac{1}{2} I_{\theta^*} t^2 + o(t^2) + \frac{1}{\sqrt{n}}\nu_n\left(\log p_{\theta^*}(x) + t\, S_{\theta^*}(x) + |t|\, r(x,t)\right)$

(ν in terms of θ^*) $= \left(L(\theta^*) + \frac{1}{\sqrt{n}}\nu_n \log p_{\theta^*}\right) + \frac{t}{\sqrt{n}}\nu_n S_{\theta^*} - \frac{1}{2} I_{\theta^*} t^2 + o(t^2) + \frac{|t|}{\sqrt{n}}\nu_n r(\cdot,t)$

$= L_n(\theta^*) + t\,\frac{1}{n}\sum_{i=1}^n S_{\theta^*}(X_i) - \frac{1}{2} t^2 I_{\theta^*} + o(t^2) + \frac{|t|}{\sqrt{n}}\nu_n r(\cdot,t)$

Step 2 show $\hat\theta$ is $\sqrt{n}$-consistent $\qquad g_n(x) = O_p(f_n(x)) \Rightarrow P\left(\left|\frac{g_n(x)}{f_n(x)}\right| > \varepsilon\right) \to 0$

$\hat t = \hat\theta - \theta^*$

$L_n(\theta^*+\hat t) = L_n(\theta^*) + \hat t\,\frac{1}{n}\sum_{i=1}^n S_{\theta^*}(X_i) - \frac{1}{2}\hat t^2 I_{\theta^*} + o_p(\hat t^2) + \frac{|\hat t|}{\sqrt{n}}\nu_n r(\cdot,\hat t)$

$\hat t = o_p(1)$ (consistency) $\Rightarrow \dfrac{|\nu_n r(\cdot,\hat t)|}{1+|\sqrt{n}\,\hat t|} < \sup_{t\in U_n}\dfrac{|\nu_n r(\cdot,t)|}{1+|\sqrt{n}\, t|} = O_p(1)$

$\Rightarrow \dfrac{|\hat t|}{\sqrt{n}}|\nu_n r(\cdot,\hat t)| \le \dfrac{|\hat t|}{\sqrt{n}}\left(O_p(1) + O_p(\sqrt{n}\,|\hat t|)\right) = O_p\left(\dfrac{|\hat t|}{\sqrt{n}}\right) + O_p(\hat t^2)$

by definition of $\hat\theta$ $\quad L_n(\theta^*+\hat t) \ge L_n(\theta^*)$

$\Rightarrow 0 \le \hat t\,\frac{1}{n}\sum_{i=1}^n S_{\theta^*}(X_i) - \frac{1}{2}\hat t^2 I_{\theta^*} + O_p(\hat t^2) + \frac{|\hat t|}{\sqrt{n}}\nu_n r(\cdot,\hat t)$

$\Rightarrow \frac{1}{2}\underbrace{\hat t^2}_{|\hat t|} I_{\theta^*} \le \cancel{|\hat t|}\left|\frac{1}{n}\sum_{i=1}^n S_\theta(X_i)\right| + \underbrace{|O_p(\hat t^2)|}_{O_p(|\hat t|)} + \underbrace{\frac{|\hat t|}{\sqrt{n}}}_{\frac{1}{\sqrt{n}}}|\nu_n r(\cdot,\hat t)|$

$(\frac{1}{2} I_{\theta^*} - o_p(1)) |\hat{t}| \leq |\frac{1}{n}\sum S_\theta(x_i)| + o_p(\frac{1}{\sqrt{n}}) \Rightarrow |\hat{t}| = O_p(\frac{1}{\sqrt{n}})$

little o

$\frac{I_\theta(x)}{n}$

small o

std $\sim \frac{1}{\sqrt{n}}$

$\exists M \quad \mathbb{P}(|\frac{X_n}{a_n}| > M) < \varepsilon$

$= \mathbb{P}(|\sqrt{n} X_n| > M) < \varepsilon$

$\mathbb{P}(n X_n \leq t) \to F(t)$

$\Rightarrow \mathbb{P}(\sqrt{n} X_n > \frac{k}{\sqrt{n}}) \to 1 - F(k)$ $\quad k$ fixed $\quad \exists n$ large

$X \sim N(0, \frac{1}{n})$

$\sqrt{n} X \sim N(0,1)$

$\sqrt{n} X \sim O_p(1)$

$X \sim O_p(\frac{1}{\sqrt{n}})$

Step 3 show asymptotic normality of $\hat{\theta}$ for any

$|t| = o_p(\frac{1}{\sqrt{n}})$

$$L_n(\theta^* + t) = L_n(\theta^*) + t \underbrace{\frac{1}{n}\sum_{i=1}^n S_{\theta^*}(x_i)}_{a} - \frac{1}{2} t^2 \underbrace{I_{\theta^*}}_{b} + \underbrace{o_p(t^2) + \frac{|t|}{\sqrt{n}} \mathrm{var}(\cdot\,t)}_{o_p(\frac{1}{n})}$$

$$at - \frac{1}{2}t^2 b = -\frac{1}{2} b (t^2 - 2\frac{a}{b} t) + \frac{a^2}{b^2} - \frac{a^2}{b^2}$$

$$= -\frac{1}{2} b (t - \frac{a}{b})^2 + \frac{1}{2}\frac{a^2}{b}$$

$$L_n(\theta^* + t) = L_n(\theta^*) - \frac{1}{2} I_{\theta^*} (t - \frac{1}{n}\sum_{i=1}^n S_{\theta^*}(x_i) / I_{\theta^*})^2 + \frac{1}{2} \frac{(\frac{1}{n}\sum_{i=1}^n S_\theta(x_i))^2}{I_{\theta^*}} + o_p(\frac{1}{n})$$

$\tilde{t} : \frac{1}{n}\sum_{i=1}^n S_\theta(x_i) / I_{\theta^*} = O_p(\frac{1}{\sqrt{n}})$ $\qquad |\hat{t}| = O_p(\frac{1}{\sqrt{n}})$

$L_n(\theta^* + \hat{t}) \geq L_n(\theta^* + \tilde{t})$ $\qquad$ what $\hat{t}$ maximizes?

$$\Rightarrow -\frac{1}{2} I_{\theta^*} (\hat{t} - \frac{1}{n}\sum_{i=1}^n \frac{S_\theta(x_i)}{I_{\theta^*}})^2 + o_p(\frac{1}{n}) \geq o_p(\frac{1}{n})$$

$$\Rightarrow \frac{1}{2} I_{\theta^*} (\hat{t} - \frac{1}{n}\sum_{i=1}^n \frac{S_\theta(x_i)}{I_{\theta^*}})^2 = o_p(\frac{1}{n})$$

$$\Rightarrow \hat{t} = \frac{1}{n}\sum_{i=1}^n \frac{S_{\theta^*}(x_i)}{I_{\theta^*}} + o_p(\frac{1}{\sqrt{n}})$$

$$\Rightarrow \sqrt{n}(\hat{\theta} - \theta^*) = \frac{1}{\sqrt{n}}\sum_{i=1}^n S_{\theta^*}(x_i) / I_{\theta^*} + o_p(1)$$

$$\rightsquigarrow N(0, I_{\theta^*}^{-1})$$

$X_1 \cdots X_n \overset{iid}{\sim} P_{\theta^*}$ $\quad \hat\theta = \arg\max_\theta L_n(\theta)$ $\quad L_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \log P_\theta(X_i)$

$\log P_{\theta^*+t}(x) = \log P_{\theta^*}(x) + t S_{\theta^*}(x) + o(t)$

$L(\theta^*+t) = L(\theta^*) - \frac{1}{2} t^2 I_{\theta^*} + o(t^2)$ ← Le Cam (1986) said that is not needed

$\forall t = O_p\left(\frac{1}{\sqrt{n}}\right)$

variance of score

$L_n(\theta^*+t) = L_n(\theta^*) + t\frac{1}{n}\sum S_{\theta^*}(X_i) - \frac{1}{2}t^2 I_{\theta^*} + O_p\left(\frac{1}{n}\right)$

$t = \frac{h}{\sqrt{n}}$

$\log \prod_{i=1}^{n} \frac{P_{\theta^*+\frac{h}{\sqrt{n}}}}{P_{\theta^*}}(X_i) = h\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_{\theta^*}(X_i) - \frac{1}{2}h^2 I_{\theta^*} + O_p(1)$ for any h not depending on n

1:01:10

↖ LAN (local asymptotic normality)

$S_\theta(x) = \frac{d}{d\theta}\log P_\theta(x) = \frac{\frac{d}{d\theta}P_\theta(x)}{P_\theta(x)}$

$\frac{d}{d\theta}\sqrt{P_\theta(x)} = \frac{1}{2}P_\theta^{-\frac{1}{2}}\frac{d}{d\theta}P_\theta = \frac{1}{2}\sqrt{P_\theta(x)}\,\frac{\frac{d}{d\theta}P_\theta(x)}{P_\theta(x)} = \frac{1}{2}\sqrt{P_\theta(x)}\,S_\theta(x)$

flattened unit vector

$\int\left(\frac{\sqrt{P_{\theta+t}(x)} - \sqrt{P_\theta}}{t} - \frac{1}{2}\sqrt{P_\theta(x)}\,S_\theta(x)\right)^2 \longrightarrow 0$ DQM at θ

differentiability in quadratic mean

Thm (Le Cam) mainly ② → ③ in above assumptions

Assume DQM at θ

① $E_\theta S_\theta(X) = 0$ (no need to exchange derivative & integration)

② LAN hold at θ with $I_\theta = E_\theta(S_\theta(X))^2$ "first order diff'bility implies second order expansion"

Lec 16

LAN: $\log \prod_{i=1}^{n} \frac{p_{\theta+\frac{h}{\sqrt{n}}}}{p_\theta}(x_i) = h\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_\theta(x_i)\right) - \frac{h^2}{2} I_\theta + o_{p_\theta}(1) \quad \forall h$

DQM $\int \left(\sqrt{p_{\theta+t}} - \sqrt{p_\theta} - \frac{1}{2}t\sqrt{p_\theta} S_\theta\right)^2 = o(t^2)$

(differentiable in quadratic mean)

Theorem (Le Cam): If DQM holds at θ

then ① $E_\theta S_\theta(x) = 0$

② LAN holds at θ, $I_\theta = E_\theta S_\theta(x)^2$

DQM $\Leftrightarrow$ ($t = \frac{h}{\sqrt{n}}$) $\int \Big(\underbrace{\sqrt{p_{\theta+\frac{h}{\sqrt{n}}}} - \sqrt{p_\theta}}_{f_n} - \underbrace{\frac{1}{2}\frac{h}{\sqrt{n}}\sqrt{p_\theta} S_\theta}_{g_n}\Big)^2 = o\left(\frac{1}{n}\right)$ $\qquad \int (f_n - g_n)^2 = o\left(\frac{1}{n}\right)$

$\Leftrightarrow \int \Big(\underbrace{\sqrt{n}\left(\sqrt{p_{\theta+\frac{h}{\sqrt{n}}}} - \sqrt{p_\theta}\right)}_{f_n} - \underbrace{\frac{1}{2}h\sqrt{p_\theta} S_\theta}_{f}\Big)^2 = o(1)$

$\Leftrightarrow \int p_\theta \left(\sqrt{\frac{p_{\theta+\frac{h}{\sqrt{n}}}}{p_\theta}} - 1 - \frac{1}{2}\frac{h}{\sqrt{n}} S_\theta\right)^2 = o\left(\frac{1}{n}\right)$

good control

$h E_\theta S_\theta(x) = h\int p_\theta S_\theta = 2\int \frac{1}{2} h\sqrt{p_\theta} S_\theta \sqrt{p_\theta}$

$= 2\int \left(\frac{1}{2}h\sqrt{p_\theta} S_\theta - \sqrt{n}\left(\sqrt{p_{\theta+\frac{h}{\sqrt{n}}}} - \sqrt{p_\theta}\right)\right)\sqrt{p_\theta} + 2\int \sqrt{n}\left(\sqrt{p_{\theta+\frac{h}{\sqrt{n}}}} - \sqrt{p_\theta}\right)$

① ②

$|①| \le 2\sqrt{\int \left(\frac{1}{2}h\sqrt{p_\theta} S_\theta - \sqrt{n}\left(\sqrt{p_{\theta+\frac{h}{\sqrt{n}}}} - \sqrt{p_\theta}\right)\right)^2}\sqrt{\int p_\theta}$

$\to 0$

$|②| = 2\sqrt{n}\left|\left(\int \sqrt{p_{\theta+\frac{h}{\sqrt{n}}} p_\theta} - 1\right)\right|$

$= \sqrt{n}\int \left(\sqrt{p_{\theta+\frac{h}{\sqrt{n}}}} - \sqrt{p_\theta}\right)^2 = \sqrt{n}\int f_n^2$

$= \sqrt{n}\int (f_n - g_n + g_n)^2 = \sqrt{n}\int g_n^2 + \sqrt{n}\int (f_n - g_n)^2 + 2\sqrt{n}\int (f_n - g_n) g_n$

$\sqrt{n}\int g_n^2 = \sqrt{n}\int \frac{1}{4}\frac{h^2}{n} p_\theta S_\theta^2 = \frac{h^2}{4\sqrt{n}} I_\theta \to 0$

$\sqrt{n}\int (f_n - g_n)^2 \to 0$

$\left|2\sqrt{n}\int (f_n - g_n) g_n\right| \le 2\sqrt{n}\underbrace{\sqrt{\int (f_n - g_n)^2}}_{o\left(\frac{1}{n}\right)}\underbrace{\sqrt{\int g_n^2}}_{\sqrt{\frac{I}{n}}} \to 0$

$|h E_\theta S_\theta(x)| \le \lim_{n\to\infty}(|①| + |②|) = 0$ $\qquad$ Done

$$2\log\prod_{i=1}^{n}\left(\sqrt{\frac{P_{\theta+\frac{h}{\sqrt{n}}}}{P_\theta}(X_i)}-1+1\right) = 2\sum_{i=1}^{n}\log(W_i+1) \qquad W_i = \sqrt{\frac{P_{\theta+\frac{h}{\sqrt{n}}}}{P_\theta}(X_i)}-1$$

$$\log(X+1) = X - \frac{X^2}{2}$$

$$= 2\sum W_i - \sum W_i^2 + o_p(1)$$ → reading assgnment

$$E_\theta \sum_i W_i = n E_\theta W_1$$

$$= n\left(\int\sqrt{P_\theta P_{\theta+\frac{h}{\sqrt{n}}}}-1\right)$$

$$= -\frac{n}{2}\left(\int\sqrt{P_{\theta+\frac{h}{\sqrt{n}}}}-\sqrt{P_\theta}\right)^2$$

$$= -\frac{1}{2}\int\left[\sqrt{n}\left(\sqrt{P_{\theta+\frac{h}{\sqrt{n}}}}-\sqrt{P_\theta}\right)\right]^2$$

$$= -\frac{1}{2}\int(f_n-f+f)^2$$

$$= -\frac{1}{2}\int f^2 - \frac{1}{2}\int(f_n-f)^2 - \int(f_n-f)f \longrightarrow \frac{1}{8}h^2 I_\theta$$

$$\left\{\begin{array}{l} -\frac{1}{2}\int f^2 = -\frac{1}{2}\int\frac{1}{4}h^2 P_\theta S_\theta^2 = -\frac{1}{8}h^2 I_\theta \\ -\frac{1}{2}\int(f_n-f)^2 \to 0 \\ \left|\int(f_n-f)f\right| \le \sqrt{\int(f_n-f)^2}\sqrt{\int f^2} \to 0 \end{array}\right.$$

$$E_\theta\sum_{i=1}^{n}\left(W_i - \frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X)\right) \longrightarrow -\frac{1}{8}h^2 I_\theta$$

$$\mathrm{Var}_\theta\left(\sum_{i=1}^{n}\left(W_i-\frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X)\right)\right) = n\,\mathrm{Var}_\theta\left(W_1-\frac{1}{2}\frac{h}{\sqrt{n}}S(X_1)\right) \le n E_\theta\left(W_1-\frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X_1)\right)^2$$

$$= n\int P_\theta\left(\sqrt{\frac{P_{\theta+\frac{h}{\sqrt{n}}}}{P_\theta}}-1-\frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X_1)\right)^2 \to 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(W_i-\frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X)\right) \xrightarrow{P_\theta} -\frac{1}{8}h^2 I_\theta \text{ (chebyshev)}$$

$$\sum W_i = \frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X) - \frac{1}{8}h^2 I_\theta + o_p(1)$$

$$\sum W_i^2 = \sum_{i=1}^{n}\left(W_i - \frac{1}{2}\frac{h}{\sqrt{n}}S(X_i) + \frac{1}{2}\frac{h}{\sqrt{n}}S(X)\right)^2$$

$$= \underbrace{\sum_{i=1}^{n}\left(W_i-\frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X_i)\right)^2}_{①} + \underbrace{\sum_{i=1}^{n}\left(\frac{1}{2}\frac{h}{\sqrt{n}}S_\theta(X_i)\right)^2}_{②} + \underbrace{2\sum_{i=1}^{n}\left(W_i-\frac{1}{2}\frac{h}{\sqrt{n}}S(X_i)\right)\frac{1}{2}\frac{h}{\sqrt{n}}S(X)}_{③}$$

① $= o_p(1)$

② $= \dfrac{1}{4}\dfrac{h^2}{n}\sum S_\theta(X_i)^2 \overset{LLN}{=} \dfrac{h^2}{4}I_\theta + o_p(1)$ — know that I_θ's finite

$|③| \le 2\sqrt{①}\sqrt{②} \to 0$

$\sum W_i^2 = \frac{1}{4} h^2 I_\theta + o_p(1)$

$2\sum W_i - \sum W_i^2 + o_p(1)$

$= h \frac{1}{\sqrt{n}} \sum S_\theta(X) - \frac{1}{2} h^2 I_\theta + o_p(1)$ $\quad$ <u>LAN</u>

Discussion

$\boxed{\begin{array}{l} L^2 \text{ cvg} \to \text{pointwise cvg } \times \\ L^2 = 0 \Rightarrow \text{uniqueness?} \end{array}}$

Cramer - Rao lower bound

$X \sim P_\theta$ $\quad$ score & Fisher info exist

$\hat\theta$ unbiased $\quad E_\theta \hat\theta = \theta$

then $\mathrm{Var}_\theta(\hat\theta) \ge I_\theta^{-1}$

$\Leftrightarrow \mathrm{Var}_\theta(\hat\theta) I_\theta \ge 1$

Cauchy-Schwarz

$= \int P_\theta (\hat\theta - \theta)^2 \int P_\theta \left(\frac{\frac{d}{d\theta} P_\theta}{P_\theta}\right)^2 \ge \left(\int \cancel{P_\theta} (\hat\theta - \theta) \frac{\frac{d}{d\theta} P_\theta}{\cancel{P_\theta}}\right)^2 = \left(\frac{d}{d\theta}\int \hat\theta P_\theta - \theta \frac{d}{d\theta}\int P_\theta\right)^2 = 1$

Corollary $\quad X_1 \cdots X_n \overset{iid}{\sim} P_\theta \quad \hat\theta$ is unbiased

$\Rightarrow \mathrm{Var}_\theta(\hat\theta) \ge \frac{1}{n I_\theta}$

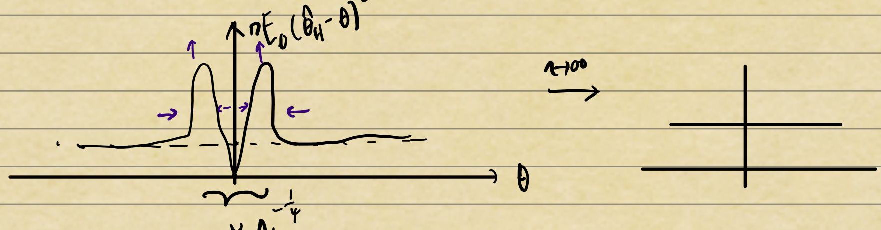
$\updownarrow$

MLE $\quad \sqrt{n}(\hat\theta - \theta) \rightsquigarrow N(0, I_\theta^{-1})$

Hodge's estimator $\quad X_1 \cdots X_n \overset{iid}{\sim} N(\theta, 1)$ $\quad$ "Is MLE optimal?"

$\hat\theta_H = \begin{cases} \bar X & |\bar X| \ge n^{-\frac14} \\ 0 & |\bar X| < n^{-\frac14} \end{cases}$ $\quad$ Q. Can you construct $\hat\theta$ whose asymptotic behavior is better?

$\sqrt{n}(\hat\theta_H - \theta) \rightsquigarrow \begin{cases} N(0,1) & \theta \ne 0 \\ 0 & \theta = 0 \end{cases}$ $\quad$ Super-efficiency (even better than James-Stein)



Lecture 17

$X_1 \dots X_n \overset{iid}{\sim} P_\theta$

$\hat\theta = \arg\max_\theta \sum_{i=1}^n \log p_\theta(X_i)$

$\sqrt{n}(\hat\theta - \theta) \rightsquigarrow N(0, I_\theta^{-1})$

Cramer-Rao lower bound

$X \sim P_\theta$ $\hat\theta$ is unbiased for θ

$Var(\hat\theta) \geq \frac{1}{I_\theta}$; iid $X_1 \dots X_n \Rightarrow Var(\hat\theta) \geq \frac{1}{nI_\theta}$ Fisher info for $X_1 \dots X_n$

Hodges estimator $X_1 \dots X_n \overset{iid}{\sim} N(\theta, 1)$ $\bar X \sim N(\theta, \frac{1}{n})$

$$\hat\theta = \begin{cases} \bar X & |\bar X| > n^{-\frac14} \\ 0 & |\bar X| \leq n^{-\frac14} \end{cases} = \bar X \mathbb{1}_{\{\bar X > n^{-\frac14}\}} \qquad \bar X = \theta + \frac{1}{\sqrt n} Z \qquad Z \sim N(0,1)$$

$$= (\theta + \tfrac{1}{\sqrt n}) \mathbb{1}_{\{|\theta + \frac{1}{\sqrt n} Z| > n^{-\frac14}\}}$$

$$= \theta - \theta \mathbb{1}_{\{|\theta + \frac{1}{\sqrt n} Z| \leq n^{-\frac14}\}} + \frac{1}{\sqrt n} Z \mathbb{1}_{\{|\theta + \frac{1}{\sqrt n} Z| > n^{-\frac14}\}}$$

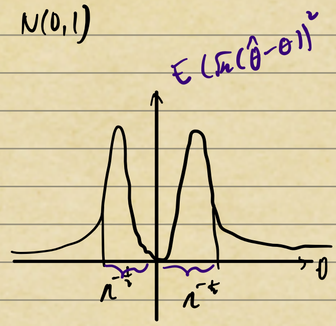
$$\sqrt n(\hat\theta - \theta) = -\sqrt n\, \theta\, \mathbb{1}_{\{|\sqrt n \theta + Z| \leq n^{\frac14}\}} + Z \mathbb{1}_{\{|\sqrt n \theta + Z| > n^{\frac14}\}}$$

① $\theta \neq 0$ does not depend on n, $\sqrt n(\hat\theta - \theta) \sim N(0,1)$

Superefficient point ② $\theta = 0$ $\sqrt n(\hat\theta - \theta) \rightsquigarrow N(0,0)$

③ $n^{-\frac12} << \theta << n^{-\frac14}$ $|\sqrt n(\hat\theta - \theta)| \to \infty$

minimax worse than $\frac{1}{n}$



"sufficient to consider $\frac{1}{\sqrt n}$ nbhd"

DQM (at θ) $\Rightarrow$ LAN $\log \prod_{i=1}^n \frac{p_{\theta + \frac{h}{\sqrt n}}}{p_\theta}(X_i) = h \frac{1}{\sqrt n} \sum S_\theta(X_i) - \frac12 h^2 I_\theta + o_p(1)$ $\forall h$

Local experiment $(P^n_{\theta + \frac{h}{\sqrt n}}\ h \in \mathbb{R})$ MLE : $\hat h = \frac{1}{\sqrt n} \sum I_\theta^{-1} S_\theta(X_i) + o_p(1) \rightsquigarrow N(0, I_\theta^{-1})$

$\}$ idea: optimality $\updownarrow$ optimality

$(N(h, I_\theta^{-1}) : h \in \mathbb{R})$

"sample 1"

why? $\log \frac{dN(h, I_\theta^{-1})}{dN(0, I_\theta^{-1})}(x) = \log \frac{e^{-\frac{I_\theta}{2}(x-h)^2}}{e^{-\frac{I_\theta}{2}x^2}}$

$= h I_\theta x - \frac{1}{2}h^2 I_\theta \qquad x \sim N(0, I_\theta)$

$\boxed{I_\theta x / I_\theta}$

$\Updownarrow$

$h \frac{1}{\sqrt{n}}\sum S_\theta(X_i) - \frac{1}{2}h^2 I_\theta + o_p(1) \quad \forall h$

parallel

$\hat{h} := \frac{1}{\sqrt{n}}\sum I_\theta^{-1} S_\theta(X_i) + o_p(1) \rightsquigarrow N(0, I_\theta^{-1})$

Appendix ① Hajek - LeCam Almost Everywhere Convolution Theorem

Theorem P_θ is DQM at θ, let T_n be a statistic of $(P^n_{\theta+\frac{h}{\sqrt{n}}} : h \in \mathbb{R})$ that satisfies

$\sqrt{n}\left(T_n - \left(\theta + \frac{h}{\sqrt{n}}\right)\right) \xrightarrow{P^n_{\theta+\frac{h}{\sqrt{n}}}} L_{\theta,h}$ every h (θ fixed)

then there exists a randomized statistic T from $(N(h, I_\theta^{-1}) : h \in \mathbb{R})$ such that

$T - h \overset{N(h, I_\theta^{-1})}{\sim} L_{\theta,h}$ for every h $\qquad T = T(X, U)$, $X \perp U$

Df. equivariance in law

$T - h \overset{N(h, I_\theta^{-1})}{\sim} L_\theta$ every h

proposition If T is a randomized statistic of $(N(h, I_\theta^{-1}) : h \in \mathbb{R})$ that is equivariant in law, then

we must have $L_\theta = N(0, I_\theta^{-1}) * M_\theta$ for some distribution M_θ

"Gaussian mixture"

(or $T - h \overset{d}{=} \underbrace{Z + V}_{\text{independent}} \quad Z \sim N(0, I_\theta^{-1})$)

Corollary (Convolution Theorem)

If P_θ is DQM at θ for a statistic of $(P^n_{\theta+\frac{h}{\sqrt{n}}} \; h \in \mathbb{R})$, T_n that satisfies $\sqrt{n}\left(T_n - \left(\theta + \frac{h}{\sqrt{n}}\right)\right) \xrightarrow{P^n_{\theta+\frac{h}{\sqrt{n}}}} L_\theta$ every h

↳ "perturbation is stable"

then $L_\theta = N(0, I_\theta^{-1}) * M_\theta$ for some M_θ

Lemma Suppose $(P^n_\theta : \theta \in \Theta)$ is DQM for $\forall \theta \in \Theta$ ($\Theta \subseteq \mathbb{R}$)

and T_n is a statistic satisfying

$\sqrt{n}(T_n - \theta) \xrightarrow{P_\theta^n} L_\theta$ every θ

then $\exists$ subsequence of $\{n\}$ s.t. for Lebesgue almost everywhere (θ, h)

along this subsequence $\sqrt{n}(T_n - (\theta + \frac{h}{\sqrt{n}})) \xrightarrow{P^n_{\theta+\frac{h}{\sqrt{n}}}} L_\theta$

Theorem (almost everywhere condition)

If $(P_\theta^n : \theta \in \Theta)$ is DQM for $\theta \in \Theta \subseteq \mathbb{R}$, and T_n is a statistic satisfying

$\sqrt{n}(T_n - \theta) \xrightarrow{P_\theta^n} L_\theta$ every θ

then for Lebesgue almost every θ

$L_\theta = N(0, I_\theta^{-1}) * M_\theta$ for some M_θ

countable enumeration? $\boxed{\text{superefficient set has measure } 0}$

Approach 2 Hajek-Le Cam local asymptotic minimaxity

If P_θ is DQM at θ (what if asymptotics doesn't exist?)

Theorem If $(P_\theta^n \; \theta \in \Theta)$ is DQM for $\theta \in \Theta$ any estimator T_n and a convex function $\ell(\cdot)$

$$\liminf_{\substack{n \to \infty \\ h_n \to \infty}} \sup_{h \in [-h_n, h_n]} \mathbb{E}^n_{\theta + \frac{h}{\sqrt{n}}} \left(\ell\left(\sqrt{n}\left(T_n - \left(\theta + \frac{h}{\sqrt{n}}\right)\right)\right)\right) \geq \int \ell \, dN(0, I_\theta^{-1})$$

Achieving the lower bound

① If DQM holds. Le Cam's one-step discretized estimator that satisfies

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n I_\theta^{-1} S_\theta(X_i) + o_p(1)$$

② DQM + empirical process conditions implies MLE