

Ex. Gaussian

Lecture 6

X_1, \dots, X_n i.i.d. from $N(\theta, 1)$, a location family. $\mathcal{F}_\theta = \{N(\theta_0, 1), N(\theta_1, 1)\}$

Minimal sufficient statistic is

$$\begin{aligned} \frac{p_{\theta_1}(x)}{p_{\theta_0}(x)} &= \frac{e^{-\sum (x_i - \theta_1)^2 / 2}}{e^{-\sum (x_i - \theta_0)^2 / 2}} \\ &= e^{-\frac{1}{2} (-\sum x_i (\theta_0 + \theta_1) 2 + n(\theta_1^2 - \theta_0^2))} \\ &= \cancel{e^{-(\theta_1 + \theta_0)n \left(\frac{\sum x_i}{n} + n(\theta_1 - \theta_0) \right)}} \\ &= \cancel{e^{-\frac{1}{2} \dots}} \\ &= e^{-(\theta_1 + \theta_0)n \left(\frac{\sum x_i}{n} + \frac{n(\theta_1 - \theta_0)}{2} \right)} \end{aligned}$$

which is equivalent to $\frac{\sum x_i}{n} = \bar{x}$

We know that \bar{x} is sufficient for $\mathcal{F} = \{N(\theta, 1) : \theta \in \mathbb{R}\}$

By the previous lemma therefore \bar{x} is minimal sufficient for \mathcal{F} .

Ex 6.14 rest, read on your own. # ①

Minimal sufficient statistics for exponential family:

Corollary

Suppose X an exponential family distributed r.v. according (B). Then $T = (T_1, \dots, T_d)$ is minimal sufficient if

1) The family is full rank
or

2) The parameter space contains $s+1$ points $\eta^{(0)}, \dots, \eta^{(s)}$ which spans E_{η} .

Proof see detailed lecture notes.

The other extreme end of being a statistic is ancillarity.

Def A statistic V is called ancillary if its distribution does not depend on the parameter θ . It is called first order ancillary if $E(V(X))$ does not depend on the parameter θ .

Ex: If X_1, \dots, X_n are i.i.d. observations of a location family, then $X_i - X_j$ $i \neq j$ are ancillary statistics.

Def. A statistic T is called complete if

$$E_{\theta}(f(T)) = 0 \quad \forall \theta$$
$$\Rightarrow$$

$$f(t) = 0 \quad (\text{a.e. w.r.t } F) \quad \#$$

(3)

Barn's theorem

If T is complete minimal sufficient statistic for \mathcal{F} , and V is an ancillary statistic for \mathcal{F} , then T and V are independent.

Proof.

V is ancillary; therefore $P(V \in A)$ does not depend on θ , for any A .

We have that

$$E_{\theta}(P(V \in A | T)) = P(V \in A) \quad (1)$$

and

$$P(V \in A | T) = \tilde{f}(T) \quad \text{~~is~~}$$

and T is complete, so ~~is~~ (1), i.e.

~~is~~

$$(1') \quad E_{\theta}(\underbrace{P(V \in A | T) - P(V \in A)}_{\tilde{f}(T)}) = 0$$

\Rightarrow

$$\tilde{f}(T) = 0$$

(4)

i.e.

$$P(V \in A | T) = P(V \in A)$$

i.e. V, T are independent. #

Note that (1) holds because

$$\begin{aligned} E_{\theta}(P(V \in A | T)) &= E_{\theta}(E(1(V \in A) | T)) \\ &= E_{\theta}(1(V \in A)) = P(V \in A) \end{aligned}$$

Question: where did I use sufficiency in the proof?

Theorem 6.22

If X is distributed according to an exponential (or ^{on canonical form} ~~on canonical form~~) then

$$T = (T_1(X), \dots, T_s(X))$$

is complete.

Proof In TSH. (read on

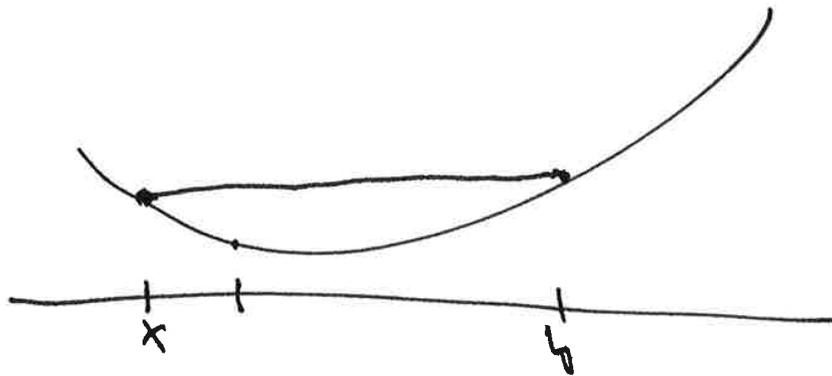
5) your own & if interested). #
book 2.

Ex. X obs of; $\text{Bin}(n, p)$, $0 < p < 1$
is one-parameter exponential family
 $T(x) = x$. Complete.

Ex. X_1, \dots, X_n i.i.d. $N(\xi, \sigma^2)$

Full rank and $T = (\bar{X}, s^2)$ is complete
for $\{N(\xi, \sigma^2) : \xi \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\}$
#

Convex loss functions



Def. φ convex function on $I = (a, b)$
if

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda \varphi(x) + (1-\lambda)\varphi(y)$$

for all $\lambda, 0 < \lambda < 1$, all $x, y \in (a, b)$.

φ strictly convex if strict inequality
for all λ, x, y

Theorem.

(i) If φ differentiable then φ convex
iff φ' nondecreasing. φ strictly convex
iff φ' strictly increasing

(ii) If φ twice differentiable then
 φ convex iff $\varphi'' \geq 0$, strictly
convex iff ~~$\varphi'' > 0$~~ $\varphi'' > 0$.
 #

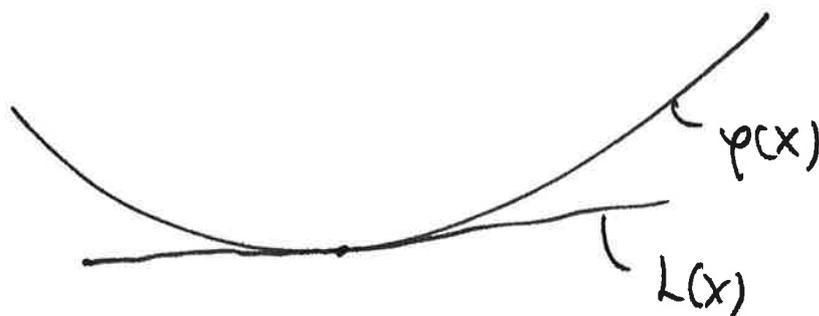
Supporting line theorem

φ convex function on $I = (a, b)$ and
 $t \in I$ fixed. Then there is a straight
 line

$$L(x) = c(x-t) + \varphi(t)$$

going through $(t, \varphi(t))$ such that

$$L(x) \leq \varphi(x) \quad \forall x \in I$$



⑧

#

Jensen's inequality

φ convex on open interval I , X r.v.
with $P(X \in I) = 1$, and $E(X) < \infty$.

Then

$$\varphi(E(X)) \leq E(\varphi(X))$$

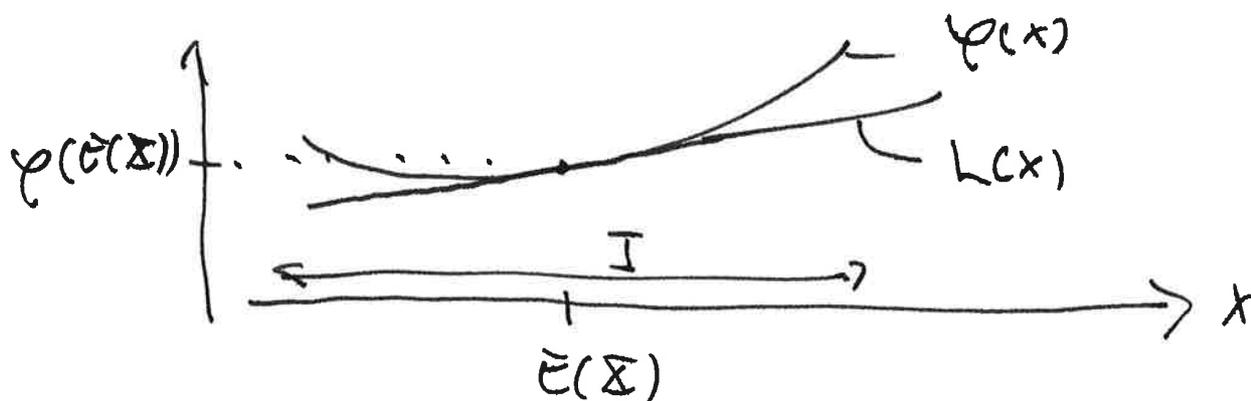
If φ strictly convex then strict inequality holds unless $X \stackrel{\text{a.s.}}{=} c$.

Proof.

Let

$$L(x) = c(x - E(X)) + \varphi(E(X))$$

be a line such that $L(x) \leq \varphi(x)$ and going through $(E(X), \varphi(E(X)))$



Then

$$L(X(\omega)) = L(X)(\omega) \leq \varphi(X)(\omega) = \varphi(X(\omega)) \quad (9)$$

so that

$$\cancel{E(L(X)) \leq E(\varphi(X))}$$

$$E(\varphi(X)) \geq E(L(X)) =$$

$$= E(c(X - E(X)) + \varphi(E(X)))$$

$$= c(E(X) - E(X)) + \varphi(E(X))$$

$$= \varphi(E(X))$$

If φ strictly convex then

$$\varphi(X) > L(X)$$

\Rightarrow

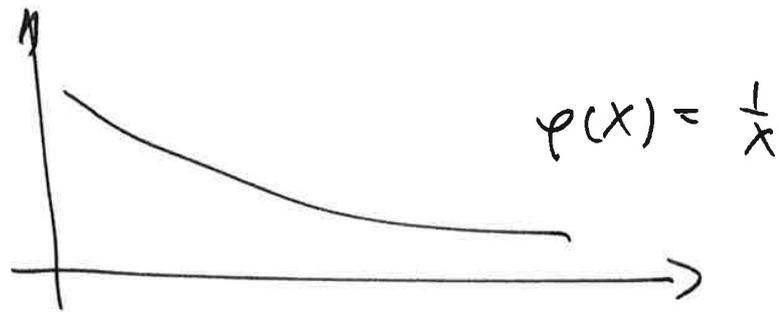
$$\cancel{E(\varphi(X)) >}$$

$$E(\varphi(X)) > \varphi(E(X))$$

If X constant, then $X = E(X), \dots$

Ex. If X positive r.v. with $E(X) < \infty$

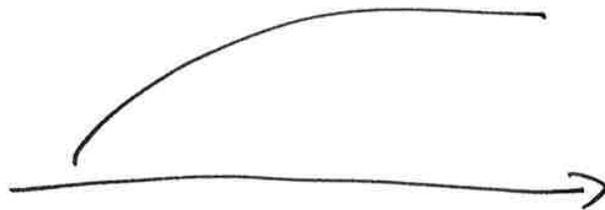
$$\varphi(E(X)) = \frac{1}{E(X)} < E\left(\frac{1}{X}\right) = E(\varphi(X))$$



and

$$E(\log X) < \log E(X)$$

$\varphi(x) = -\log x$
concave



$$\begin{aligned} \varphi(E(X)) &= -\log E(X) \\ &< E(-\log X) = E(\varphi(X)) \end{aligned}$$

\Leftrightarrow

$$\log E(X) > E(\log X)$$

#

Ex. (Kullback-Leibler distance)

Way to measure distance between two density functions f, g .

$$E_f \left(\log \frac{f(x)}{g(x)} \right) = \int \log \frac{f(y)}{g(y)} f(y) dy$$

Then

$$E_f \left(\log \frac{f(x)}{g(x)} \right) = - E_f \left(\log \frac{g(x)}{f(x)} \right)$$

$$\geq - \log E_f \left(\frac{g(x)}{f(x)} \right) =$$

$$= - \log \int \frac{g(y)}{f(y)} f(y) dy = - \log \int g(y) dy$$

$$= 0$$

so distance ≥ 0 and $= 0$ if

$$f = g \quad (\text{a.k.})$$

Rao-Blackwell's theorem

X a r.v. with d.f. $F_\theta \in \mathcal{F} = \{F_\theta : \theta \in \mathcal{R}\}$ (1)

T sufficient for \mathcal{F} . Let δ be an estimator of $g(\theta)$. Assume loss function $L(\theta, d)$ strictly convex function in d .

Let $R(\theta, \delta) = E L(\theta, \delta(X)) < \infty$

and define the estimator

$$\eta(T) = E_{\theta}(\delta(X) | T) = E_{\theta}(\delta(X) | T(X)).$$

Then $R(\theta, \eta) < R(\theta, \delta)$ unless $\delta(X) = \eta(T)$ almost surely.
 note that by sufficiency of T , $\eta(T)$ does not depend on θ , i.e. it is a statistic.

Proof.

Recall that we have a Jensen's inequality

$$\varphi(E(X)) \leq E(\varphi(X))$$

if φ convex; if φ strictly convex then there is strict inequality unless $X = c$ a.s.

The proof of this is via "the supporting line theorem" (supporting hyperplane theorem)

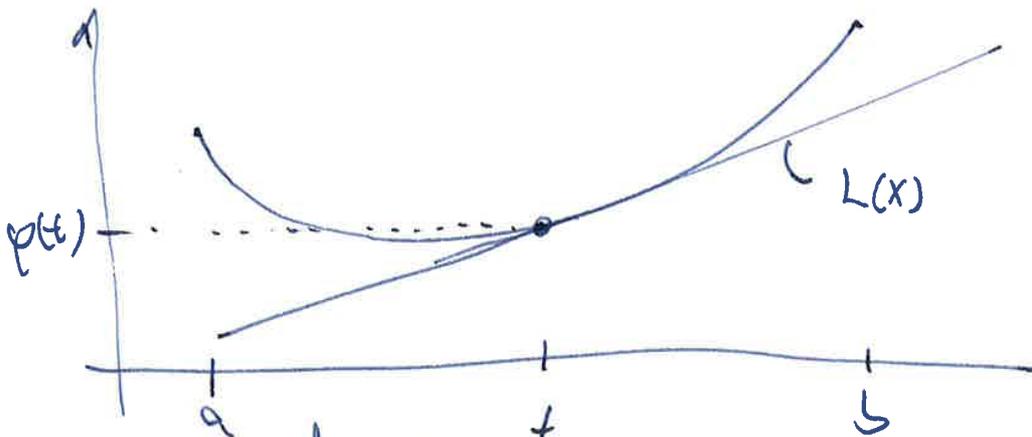
φ convex on $I = (a, b)$ and $t \in I$ fixed. Then there is a straight line

$$L(x) = c(x-t) + \varphi(t)$$

that goes through $(t, \varphi(t))$ and that

$$L(x) \leq \varphi(x)$$

(2)



One can ^{from this} similarly prove a conditional Jensen's inequality look in Willinger's book

$$\varphi(E(X|Y)) \leq E(\varphi(X)|Y)$$

(2)

with strict inequality unless

(1) $E(X|Y) = c$ almost surely

"Probability with martingales"
martingales"

~~we see first take E of (1)~~
 ~~$E(E(X|Y)) = E(X) = c$~~

Now we apply ⁽²⁾ that to φ , with \downarrow see lecture note

$$\varphi(d) = L(\theta, d)$$

that

$$L(\theta, \varphi(T)) = L(\theta, E_{\theta}(\varphi(X)|T)) \leq E_{\theta}(L(\theta, \varphi(X)) | T)$$

note given on page 16.

(14)

Now take E_{θ} of both side

(3)

$$E_{\theta}(L(\theta, Y(T))) \leq$$

$$\leq E_{\theta} E_{\theta}(L(\theta, \sigma(X)) | T)$$

$$= E_{\theta}(L(\theta, \sigma))$$

(15')

Now with $\varphi(d) = L(\theta, d)$, so φ strictly convex

$$\underline{L(\theta, \eta(T))} = L(\theta, E_{\theta}(\delta(\Delta) | T))$$

$$= \varphi(E_{\theta}(\delta(\Delta) | T))$$

$$\star \underline{E_{\theta}(\varphi(\delta(\Delta)) | T)} \quad \varphi \text{ strictly convex}$$

$$= \underline{E_{\theta}(L(\theta, \delta(\Delta)) | T)}$$

Now take E_{θ} of both sides to get

$$\underline{R(\theta, \eta(T))} = E_{\theta}(L(\theta, \eta(T)))$$

$$< E(E_{\theta}(L(\theta, \delta(\Delta)) | T))$$

$$= E_{\theta}(L(\theta, \delta(\Delta))) = \underline{R(\theta, \delta)}$$