Risk, Sufficiency, completeness, and Ancillarity

- X —— RV. $X \sim P_\theta$ $\theta \in \mathbb{R}^p$

Model: a set of distributions for X: $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$.

- Statistic: a function of the data X.

ex $\delta(X) = \bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$

- Inference: Estimation hypothesis testing.

- estimation $\delta(X) \longrightarrow g(\theta)$

ex. $X \sim Bin(100, \theta)$ flip a coin 100 times

$\theta \in [0,1] = \Omega$ $P_\theta = Bin(100, \theta)$

$\mathcal{P} = \{P_\theta : \theta \in \Omega\}$

An estimator of θ is $\delta(X) = \frac{X}{100}$

- good or bad? decision theory.

Loss function: $L(\theta, d) \geq 0$ estimate $g(\theta)$ by a value d.

Risk function: $R(\theta, \delta) = E_\theta L(\theta, \delta(X))$

ex. $X \sim Bin(100, \theta)$ $\quad \delta(X) = \frac{X}{100}$ $\quad g(\theta) = \theta$

$L(\theta, d) = (\theta - d)^2$.

$R(\theta, \delta) = E_\theta\left(\theta - \frac{X}{100}\right)^2 = \frac{\theta(1-\theta)}{100}$ $\quad \theta \in [0,1]$.

- compare different estimators:

$\delta_0(X) = \frac{X}{100}$. $\quad \delta_1(X) = \frac{X+3}{106}$ $\quad \delta_2(X) = \frac{X+3}{106}$

$R(\theta, \delta_0) = \frac{\theta(1-\theta)}{100}$ $\quad R(\theta, \delta_1) = \frac{9 + 100\theta(1-\theta)}{100^2}$. $\quad R(\theta, \delta_2) = \frac{(9-8\theta)(1+8\theta)}{106^2}$

when θ is near $\frac{1}{2}$. δ_2 is prefered.

minimax estimator.

Sufficient statistics

ex. X, Y — independent $\quad f_\theta(x) = \theta e^{-\theta x}$ $\quad x > 0$.

U independent of X and Y $\quad U \sim Unif(0,1)$

$T = X + Y$. $\quad \begin{cases} \tilde{X} = UT \\ \tilde{Y} = (1-U)T \end{cases}$

density of T:

$$P(T \le t \mid Y = y) = P(X + Y \le t \mid Y = y) = F_X(t - y)$$

$$F_T(t) = P(T \le t) = E F_X(t - Y) = \int_0^t (1 - e^{-\theta(t-y)}) \theta e^{-\theta y} dy.$$

$$= 1 - e^{-\theta t} - t\theta e^{-\theta t}.$$

$$P_T(t) = t\theta^2 e^{-\theta t} \quad t \ge 0$$

joint density of (T, U):

$$P_\theta(t, u) = \begin{cases} t\theta^2 e^{-\theta t} & t \ge 0 \quad u \in (0,1) \\ 0 & \text{o.w.} \end{cases}$$

$$P\left(\begin{pmatrix} \tilde{X} \\ \tilde{Y} \end{pmatrix} \in B\right) = \iint 1_B(tu, t(1-u)) P_\theta(t, u) \, du \, dt$$

$$= \iint 1_B(x, y) \frac{1}{x+y} P_\theta(x+y, \frac{x}{x+y}) \, dy \, dx.$$

joint density $(\tilde{X}, \tilde{Y})$:

$$\frac{P_\theta(x+y, \frac{x}{x+y})}{x+y} = \theta^2 e^{-\theta(x+y)} \quad x \ge 0, \ y \ge 0$$

$\longrightarrow$ same as the joint density of (X, Y).

T provides as much information about θ as the pair (X, Y).

$T = X + Y \longrightarrow$ sufficient statistic.

conditional distribution Q_t for X and Y given $T=t$

$$Q_t(B) = P\left[\begin{pmatrix} X \\ Y \end{pmatrix} \in B \,\middle|\, T=t\right] = P\left[\begin{pmatrix} Ut \\ (1-U)t \end{pmatrix} \in B\right]$$

does not depend on θ.

Ref: $X \to P = \{P_\theta : \theta \in \Omega\}$.

$\Rightarrow$ $T = T(X)$ is a sufficient statistic for P if for every t and θ, the conditional distribution of X under P_θ given $T=t$ does not depend on θ.

Suppose T is sufficient $\quad Q_t(B) = P_\theta(X \in B \mid T=t)$

$P_\theta(X \in B \mid T) = Q_T(B)$. $\quad P_\theta(X \in B) = E_T Q_T(B)$.

construct "fake" data $\tilde{X}$ from T taking $\tilde{X} \sim Q_t$ when $T=t$. then $\tilde{X} \mid T=t \sim Q_t$.

$P_\theta(\tilde{X} \in B) = E_\theta P_\theta(\tilde{X} \in B \mid T) = E_\theta Q_T(B)$.

X and $\tilde{X}$ have the same distribution.

Thm. $X \to P = \{P_\theta : \theta \in \Omega\}$ $\quad T = T(X)$ suff.

$\Rightarrow$ For any $\delta(X)$ of $g(\theta)$ there exists a randomized estimator based on T that has the same risk function as $\delta(X)$

Factorization Theorem.

- A sufficient statistic $T(X)$ conveys all of the information about θ from data X.

Def. A family of distributions $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ is dominated if there exists a measure μ with P_θ absolutely continuous w.r.t. μ, for all $\theta \in \Omega$.

Thm. $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ dominated by μ. A necessary and sufficient condition for T to be sufficient is that

$$p_\theta(x) = g_\theta(T(x))\, h(x) \quad \text{for a.e. } x \text{ under } \mu.$$

ex. $X_1 \dots X_n \overset{iid}{\sim} f_\theta(x) = (\theta+1) x^\theta \quad x \in (0,1) \quad \theta > -1$

$$p_\theta(x) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n (\theta+1) x_i^\theta = (\theta+1)^n \left(\prod_{i=1}^n x_i\right)^\theta \quad x \in (0,1)^n$$

$T = \prod_{i=1}^n X_i$ — suff.

ex $X_1 \dots X_n \overset{iid}{\sim} \text{Unif}(\theta, \theta+1)$ $\qquad f_\theta(x) = I_{(\theta, \theta+1)}(x)$

$$\prod_{i=1}^n 1_{(\theta, \theta+1)}(x_i) = 1_{(0, +\infty)}(\min x_i)\, 1_{(-\infty, \theta+1)}(\max x_i)$$

$T = (\min X_i, \max X_i)$ — suff.

Minimal sufficiency

- If T is suff. if $T = f(\tilde{T}) \Rightarrow \tilde{T}$ is also suff.

Def. T is minimal suff. if T is suff. and for every suff. $\tilde{T}$ there exists a function f s.t. $T = f(\tilde{T})$.

ex. $X_1 \cdots X_{2n} \overset{iid}{\sim} N(0,1)$

$\tilde{T} = \begin{pmatrix} \sum_{i=1}^{n} X_i \\ \sum_{i=n+1}^{2n} X_i \end{pmatrix}$ is suff. but not minimal suff.

$T = \sum_{i=1}^{2n} X_i$ — minimal suff. $T = f(\tilde{T})$. $f(t) = t_1 + t_2$.

- $P_\theta(x)$: viewed as a function of θ — likelihood

any suff. statistic must provide enough information to graph the shape of the likelihood

Thm: $P_\theta(x) = g_\theta(T(x))h(x)$. if $P_\theta(x) \propto P_\theta(y)$ implies $T(x) = T(y)$. then T is minimal sufficient

$P_\theta(x) = c(x,y) P_\theta(y)$

ex. $\mathcal{P}$ — s-parameter exponential family.

$$p_\theta(x) = e^{\eta(\theta)T(x) - B(\theta)} h(x) \qquad \theta \in \Omega$$

T — suff. Suppose $p_\theta(x) \propto_\theta p_\theta(y)$

$$e^{\eta(\theta)T(x)} \propto_\theta e^{\eta(\theta)T(y)} \qquad \eta(\theta)T(x) = \eta(\theta)T(y) + c(x,y)$$

if θ_0 and θ_1 are any two points in Ω.

$$[\eta(\theta_0) - \eta(\theta_1)]\, T(x) = [\eta(\theta_0) - \eta(\theta_1)] \cdot T(y)$$

$$[\eta(\theta_0) - \eta(\theta_1)] \cdot [T(x) - T(y)] = 0$$

$T(x) - T(y)$ is orthogonal to every vector in

$$\eta(\Omega) \ominus \eta(\Omega) \triangleq \{\eta(\theta_0) - \eta(\theta_1) : \theta_0 \in \Omega, \theta_1 \in \Omega\}$$

it must lie in the orthogonal complement of the linear span of $\eta(\Omega) \ominus \eta(\Omega)$ if $\eta(\Omega) \ominus \eta(\Omega) = \mathbb{R}^s$ $T(x) = T(y)$

ex. $X_1 \cdots X_n$ iid $f_\theta(x) = \frac{1}{2} e^{-|x-\theta|}$

joint $p_\theta(x) = \frac{1}{2^n} e^{-\sum_{i=1}^n |x_i - \theta|}$

$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ — order statistic — suff.

$p_\theta(x) \propto p_\theta(y)$ $\qquad \sum_{i=1}^n |x_i - \theta| = \sum_{i=1}^n |y_i - \theta| + c$.

piecewise linear functions of θ with a slope that increase by two at each order statistic. The difference can only be constant in θ if x and y have the same order statistic.

Completeness

Ref: A statistic T is <u>complete</u> for a family $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ if $E_\theta f(T) = c$ for all θ implies $f(T) = c$ (a.e. $\mathcal{P}$).

ex. $X_1 \cdots X_n$ i.i.d Unif$(0, \theta)$

joint density $I(\min X_i > 0)\, I(\max X_i < \theta)\, \frac{1}{\theta^n}$

$T = \max\{X_1 \cdots X_n\}$ sufficient.

$$P_\theta(T \le t) = P_\theta(X_1 \le t, \cdots X_n \le t) = P_\theta(X_1 \le t) \cdots P_\theta(X_n \le t)$$
$$= \left(\frac{t}{\theta}\right)^n \qquad t \in (0, \theta)$$

T has density $\frac{n t^{n-1}}{\theta^n}$ $\quad t \in (0, \theta)$

If $E_\theta f(T) = c$ for all $\theta > 0$

$$E_\theta(f(T) - c) = \frac{n}{\theta^n} \int_0^\theta (f(t) - c)\, t^{n-1}\, dt = 0.$$

$f(T) = c$. $\Rightarrow T$ is complete.

Thm: If T is complete and sufficient, then T is minimal sufficient

pf. Let $\tilde{T}$ be a minimal suff. statistics. $\tilde{T} = f(T)$

Define $g(\tilde{T}) = E_\theta(T|\tilde{T})$ — independent of θ.

$E_\theta g(\tilde{T}) = E_\theta(T)$ $\quad E_\theta(T - g(\tilde{T})) = 0$ for all θ.

$T - g(\tilde{T}) = T - g(f(T)) \Rightarrow T = g(\tilde{T})$ a.e. P.

<u>Def.</u> An exponential family with densities $p_\theta(x) = e^{\eta(\theta)\,T(x) - B(\theta)} h(x)$ $\theta \in \Omega$ is said to be of full rank if the interior of $\eta(\Omega)$ is not empty. and if $T_1, \dots T_s$ do not satisfy a linear constraint $v \cdot T = c$. (a.e. μ).

---

if $\Omega \subset \mathbb{R}^s$ and η is continuous and one-to-one, and the interior of Ω is nonempty, then the interior of $\eta(\Omega)$ cannot be empty.

$\eta(\Omega) \ominus \eta(\Omega)$ will be all of $\mathbb{R}^s$. $\Rightarrow T$ will be minimal suff.

<u>Thm</u> In an exponential family of full rank, T is complete.

<u>Def</u> A statistic V is called <u>ancillary</u> if its distribution does not depend on θ. — no information about θ

Thm (Basu) If T is complete and suff. for $\mathcal{P}=\{P_\theta, \theta\in\Omega\}$ and if V is ancillary, then T and V are independent under P_θ for any $\theta\in\Omega$

pf: $q_A(t)=P_\theta(V\in A \mid T=t)$ $\quad q_A(T)=P_\theta(V\in A\mid T)$

$p_A=P_\theta(V\in A)$ By sufficiency and ancillarity, neither p_A nor $q_A(t)$ depend on θ.

$$p_A=P_\theta(V\in A)=E_\theta\, q_A(T)$$

By completeness, $q_A(T)=p_A$.

$$P_\theta(T\in B, V\in A)=E_\theta\, 1_B(T)1_A(V)=E_\theta E_\theta(1_B(T)1_A(V)\mid T)$$

$$=E_\theta\, 1_B(T)\,E_\theta(1_A(V)\mid T)=E_\theta 1_B(T)q_A(T)$$

$$=E_\theta 1_B(T)\,p_A = P_\theta(T\in B)\,P_\theta(V\in A).$$

---

ex. $X_1\cdots X_n \overset{iid}{\sim} N(\mu,\sigma^2)$ $\quad P=P_\sigma=\{N(\mu,\sigma^2)^n : \mu\in\mathbb{R}\}$

$\bar X=\frac{1}{n}(X_1+\cdots+X_n)$

Joint density: $\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\exp\left[\frac{n\mu}{\sigma^2}\bar X-\frac{n\mu^2}{2\sigma^2}-\frac{1}{2\sigma^2}\sum_{i=1}^n X_i^2\right]$

full rank exponential family.

$\bar X$ is <u>complete</u>.

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ — sample variance. ancillary.

$Y_i = X_i - \mu \sim N(0, \sigma^2)$. $\bar{Y} = \bar{X} - \mu$ $X_i - \bar{X} = Y_i - \bar{Y}$.

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$. joint distribution of $Y_1 \cdots Y_n$ depends on σ but not μ. $\Rightarrow$ $\bar{X}$ and S^2 are independent.

Convex Loss and the Rao-Blackwell Theorem.

Def. $f : C \to \mathbb{R}$ $C \subset \mathbb{R}^p$ convex

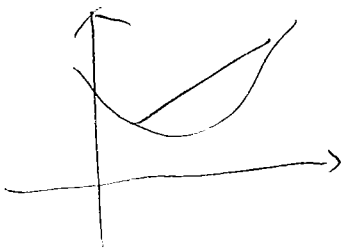
f convex if $f(rx + (1-r)y) \le r f(x) + (1-r) f(y)$.

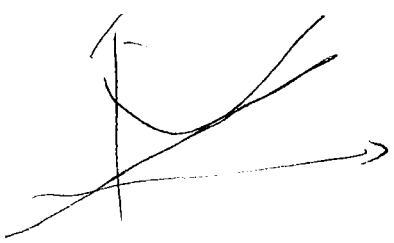
$r \in (0,1)$

strictly convex

$p=1$: $f'' \ge 0 \Rightarrow f$ convex

Thm. f — convex on an open interval C and if t is an arbitrary point in C. then there exists a constant $c = c_t$ s.t. $f(t) + c(x-t) \le f(x)$ $\forall x \in C$.





Thm: (Jensen's inequality) C — an open interval

f — convex on C. $P(X \in C) = 1$ $EX < \infty$.

$\Rightarrow f(EX) \le Ef(X)$.

pf. with $t = EX$. $f(EX) + c(X - EX) \le f(x)$ $\forall x \in C$

$f(EX) + c(X - EX) \le f(X)$ (a.e. P).

Ex. $\frac{1}{x}$ $-\log x$ strictly convex $(0, +\infty)$

$X > 0$. $\frac{1}{EX} \le E\frac{1}{X}$. $\log EX \ge E\log X$

Thm (Rao-Blackwell). T — suff. for $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$.

δ — an estimator of $g(\theta)$ define $\eta(T) = E[\delta(X) | T]$.

If $\theta \in \Omega$, $R(\theta, \delta) < \infty$, $L(\theta, \cdot)$ is convex. then

$R(\theta, \eta) \le R(\theta, \delta)$.

If $L(\theta, \cdot)$ is strictly convex, the inequality will be strict unless $\delta(X) = \eta(T)$ (a.e. P_θ).

Pf. Jensen's Inequality:

$$L(\theta, \eta(T)) \leq E_\theta\left(L(\theta, \delta(X)) \mid T \right).$$

Taking expectations, $R(\theta, \eta) \leq R(\theta, \delta)$.

- If estimators are judged by their risk, it should be a function of T, but not X.