Lec25 Friday March 19

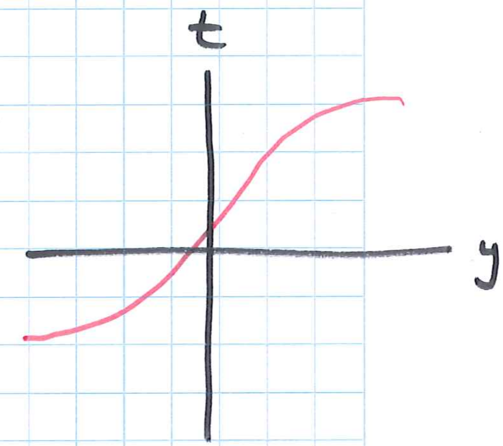
Sufficiency with "The Correct" Notation ...

Y, T jointly distributed with parm θ

$$f_{Y|T;\theta}(y|t) = \frac{f_{Y,T;\theta}(y,t)}{f_{T;\theta}(t)}$$

Then interested in case $T = T(Y)$

$$f_{Y,T;\theta}(y, T(y)) = f_{Y;\theta}(y)$$



Generally have

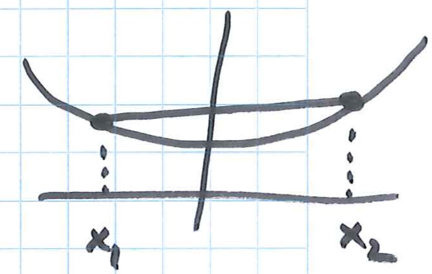
$$f_{Y,T;\theta}(y, T(y)) = f_{Y;\theta}(y) = f_{Y|T;\theta}(y|T(y))\, f_{T;\theta}(T(y))$$

T is sufficient if this does not depend on θ

Jensen's Inequality

A function $f(x)$ is convex over (a,b) if for every $x, y \in (a,b)$ and $0 \le \lambda \le 1$

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$



Strictly convex if equality holds only for $\lambda = 0$ or $\lambda = 1$.

Jensen: f is a convex function and X is a random variable then

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

Moreover, if f is strictly convex then equality implies $X = EX$ with prob. one.

Proof for Discrete Distributions

Go by induction on the # of mass points in dist.

Say $X \in \{x_1, x_2, x_3 \ldots\}$

$$p_k = P(X = x_k)$$

Suppose 2 mass points

$$Ef(X) = p_1 f(x_1) + p_2 f(x_2)$$

$$\forall |$$

$$f(EX) = f(p_1 x_1 + p_2 x_2)$$

$p_i \geq 0$
$p_1 + p_2 = 1.$

Assume J. I. true $\forall$ dists having $n-1$ mass pts.

Consider a case

$$P(X = x_k) = p_k \quad k = 1, 2 \ldots n \qquad \sum_{k=1}^{n} p_k = 1$$

Form C
APPROVED FOR USE IN
PURDUE UNIVERSITY

Write $p_i' = \dfrac{p_i}{1-p_n} \quad 1 \le i \le n-1$

$$\sum_{i=1}^{n} p_i f(x_i) = p_n f(x_n) + (1-p_n) \sum_{i=1}^{n-1} p_i' f(x_i)$$

$$\geq p_n f(x_n) + (1-p_n) f\left(\sum_{i=1}^{n-1} p_i' x_i \right).$$

By induction

$$\geq f\Big(p_n x_n + \underbrace{(1-p_n) \sum_{i=1}^{n-1} p_i' x_i}_{\sum_{i=1}^{n-1} p_i x_i} \Big)$$

This actually proves Jensen for any discrete r.v. with a dist. having finite support

Back to Est. in Classical Setup

Y rand. var. $\in \Gamma$ $\quad Y \sim f_\theta(y) \quad \theta \in \Lambda$

θ is merely unknown $\quad \Lambda = \mathbb{R}, \mathbb{R}^m$

Seek a "good" estimate $\hat{\theta}(y)$

↳ ie try to minimize some average error criterion.

Say squared error....

Since there is no prior for θ we can only look @ conditional risks ....

$$R_\theta(\hat{\theta}) = E_\theta\left\{\left(\hat{\theta}(Y) - \theta\right)^2\right\} \qquad \theta \in \Lambda$$

$E_{Y;\theta}$

Cannot generally expect to minimize $R_\theta(\hat{\theta})$ uniformly for $\theta \in \Lambda$.

<u>Example</u> Let's pick parameter value $\theta_0 \in \Lambda$ define an estimator

$$\hat{\theta}(y) = \theta_0 \qquad \forall y \in \Gamma$$

$$\Longrightarrow R_{\theta_0}(\hat{\theta}) = 0$$

but performance could be quite poor for $\theta \neq \theta_0$.

Must come up with a way to exclude estimators such as $\hat{\theta}$ above.

A reasonable restriction:

$$E_\theta\{\hat{\theta}(Y)\} = \theta \quad \forall\, \theta \in \Lambda$$

such is called <u>unbiased</u>.

and if this holds then

$$R_\theta(\hat{\theta}) = E_\theta\{(\hat{\theta}(Y)-\theta)^2\}$$

$$= E_\theta\{(\hat{\theta}(Y) - E_\theta\{\hat{\theta}(Y)\})^2\}$$

$$= Var_\theta\{\hat{\theta}(Y)\}.$$

ie MSE is Var.

An <u>unbiased</u> <u>estimator</u> minimizing the MSE for every $\theta \in \Lambda$ is called

<u>minimum</u> <u>variance</u> <u>unbiased</u> <u>estimator</u> (MVUE).

Def.

1) sufficient stat. ....

2) A function T of Y is said to be <u>minimal sufficient</u> if

* it is sufficient
* it is a function of every other suff. statistic.

$\Rightarrow$ represents the most compression of observations, which does not lose any info about θ.

$\Rightarrow$ These sometimes don't exist.

<u>NP Factorization Theorem</u> A stat. T is suff. for θ if and only if $\exists$ functions g_θ and h st.

$$f_\theta(y) = g_\theta\big(T(y)\big)\, h(y) \quad \forall y \in \Gamma, \ \forall \theta \in \Lambda$$