Stat 543 1-26-05

Recall $T(X)$ sufficient plus

log likelihoods for x and y with same shape $\Rightarrow T(x) = T(y)$

gives $T(X)$ minimal sufficient

Example $X_1, X_2, \ldots, X_n$ iid Beta (α, β)

We've seen that $T(X) = \left(\prod_{i=1}^{n} X_i, \prod_{i=1}^{n} (1-X_i)\right)$ is sufficient for (α, β) – now argue that it is minimal sufficient

Joint pdf on $[0,1]^n$ is

$$f(x \mid \alpha,\beta) = \frac{1}{(B(\alpha,\beta))^n} \left(\prod x_i\right)^{\alpha-1} \left(\prod (1-x_i)\right)^{\beta-1}$$

$f(y \mid \alpha,\beta) = k(y,x)\, f(x \mid \alpha,\beta)$ as a function of $(\alpha,\beta) \in (0,\infty)^2$ requires

$$\left(\prod y_i\right)^{\alpha-1} \left(\prod (1-y_i)\right)^{\beta-1} = k(y,x) \left(\prod x_i\right)^{\alpha-1} \left(\prod (1-x_i)\right)^{\beta-1}$$

Set $\beta = 1$ and not that this requires

$$\left(\prod y_i\right)^{\alpha-1} = k(y,x) \left(\prod x_i\right)^{\alpha-1}$$

$$\left(\frac{\prod y_i}{\prod x_i}\right)^{\alpha-1} = k(y,x) \quad \text{as a function of } \alpha$$

and this requires $\frac{\prod y_i}{\prod x_i} = 1$ i.e. $\prod y_i = \prod x_i$

a similar argument setting $\alpha = 1$ requires

$\prod(1-x_i) = \prod(1-y_i)$ i.e. $T(x) = T(y)$

and by the theorem T is minimal sufficient

Are there any "mechanical" means of identifying a minimal sufficient statistic? Yes

1) In general ... but the "statistic" is a bit hard to think about

2) In special convenient families of dsns called "exponential families"

General Result

<u>Thm</u> (Dynkin/Lehmann/Scheffé) If $f(x|\theta)$ $\forall \theta$ is either a pdf for X on $\mathbb{R}^k$ or is a pmf for X and $\exists\ \theta_0$ s.t.

if $f(x|\theta) > 0$ for any θ, $f(x|\theta_0) > 0$

then the (random function of θ)

$$R(x,\theta) = \frac{f(x|\theta)}{f(x|\theta_0)}$$

is a minimal sufficient statistic

Example $f(x|\theta)$ pmf

θ \ x	0	1	2	3	4
1	.1	.25	.2	.4	.05
2	.2	.2	.4	.1	.1
3	.1	.35	.2	.3	.05

$\frac{f(x|\theta)}{f(x|3)} = R(x,\theta)$

θ \ x	0	1	2	3	4
1	1	5/7	1	4/3	1
2	2	4/7	2	1/3	2
3	1	1	1	1	1

Note that the 3 functions of θ for $x=0,2,4$ are identical and are different from those of $x=1,3$ (which are not the same) — so anything equivalent to e.g.

$$T(0)=T(2)=T(4)=17$$
$$T(3)=35$$
$$T(1)=8$$

Example $X_1, X_2, \ldots, X_n$ iid $N(\mu,\sigma^2)$

$$R(x,\mu,\sigma^2)=\frac{f(x\mid\mu,\sigma^2)}{f(x\mid 0,1)}=\frac{1}{\sigma^n}\exp\left[-\frac{1}{2}\left[\left(\frac{1}{\sigma^2}-1\right)\sum x_i^2-\frac{2\mu}{\sigma^2}\sum x_i+\frac{n\mu^2}{\sigma^2}\right]\right]$$

Clearly if x and y have same $\sum x_i=\sum y_i$ and $\sum x_i^2=\sum y_i^2$, $R(x,\mu,\sigma^2)=R(y,\mu,\sigma^2)$ as functions of μ,σ^2

I want to argue that if $\left(\sum x_i, \sum x_i^2\right) \neq \left(\sum y_i, \sum y_i^2\right)$ that they are different functions

If $\sum x_i \neq \sum y_i$ and consider $R(x,1,1)$ and $R(y,1,1)$

$$R(x,1,1) = \exp\left(\sum x_i + n\right) \neq$$

$$R(y,1,1) = \exp\left(\sum y + n\right)$$

and $R(x,\mu,\sigma^2)$ and $R(y,\mu,\sigma^2)$ are different functions

If $\sum x_i^2 \neq \sum y_i^2$ consider $R(x,0,2)$ and $R(y,0,2)$

$$R(x,0,2) = \left(\frac{1}{\sqrt{2}}\right)^n \exp\left(-\frac{1}{2}\left[-\frac{3}{4}\sum x_i^2\right]\right) \neq$$

$$R(y,0,2) = \left(\frac{1}{\sqrt{2}}\right)^n \exp\left(-\frac{1}{2}\left(-\frac{3}{4}\sum y_i^2\right)\right)$$

and $R(x,\mu,\sigma^2)$ and $R(y,\mu,\sigma^2)$ are different functions i.e. $R(x, \mu, \sigma^2) = R(y, \mu, \sigma^2)$ as functions of (μ, σ^2) iff $\Sigma x_i = \Sigma y_i$ and $\Sigma x_i^2 = \Sigma y_i^2$

i.e.

$$T(X) = \left(\Sigma X_i , \Sigma X_i^2 \right)$$

is equivalent to the minimal sufficient statistic

$$R(X, \mu, \sigma^2)$$

Start reading Section 1.6 on exponential families — They have many nice properties, among them is the fact that identifying a minimal sufficient statistic is easy

---

Handout