Stat 543 1-19-05

A bit more about decision Thy

<u>Example</u> $X \sim N(\theta, 1)$ $\Theta = \mathcal{A} = \mathbb{R}$

$L(\theta, a) = (\theta - a)^2$

$\delta_c = cX \qquad c \in [0,1]$

$R(\theta, \delta_c) = (1-c)^2\theta^2 + c^2$

consider comparing these risk functions (comparing δ_c's)

From a minimax point of view

$$\text{"}\max_{\theta}\text{"}\ R(\theta,\delta_c) = \infty \quad \text{unless} \quad c=1$$

$$R(\theta,\delta_1) = 1$$

Since $1 < \infty$ δ_1 is "minimax" among $\{\delta_c \mid 0 \le c \le 1\}$

From a Bayes point of view .... one possible prior dsn for θ is $N(0,\gamma^2)$ (prior G is $N(0,\gamma^2)$)

The average of $R(\theta, \delta_c)$ against this (prior) dsn for θ is

$$R(G, \delta_c) = \int R(\theta, \delta_c)\, g(\theta)\, d\theta$$
$$= (1-c)^2 \gamma^2 + c^2$$
$$= (1+\gamma^2) c^2 - 2\gamma^2 c + \gamma^2$$

This is minimized over choices of c if

$$2(1+\gamma^2) c = 2\gamma^2$$
$$c = \frac{\gamma^2}{1+\gamma^2}$$

So for this prior dsn

$$\delta_{\frac{\gamma^2}{1+\gamma^2}}$$

is "Bayes" among $\{ \delta_c \mid 0 \le c \le 1 \}$

Sufficiency (Data Reduction w/o loss of Important Information)

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

we "automatically" reduce $X = (X_1, \ldots, X_n)$ to $(\bar{X}, s^2)$

<u>Example</u> $X_1, X_2, \ldots, X_5$ iid Ber(p)

we "automatically" reduce $X = (X_1, \ldots, X_5)$ to $\sum X_i$

Is this reasonable? How can it be justified? What is the corresponding rigamarole in other models?

<u>Def</u> We'll call a function of the observable X, $T(X)$, a statistic - usually T will take values in $\mathbb{R}^k$ and we can

then think of the statistic as being "k-dimensional"

Example iid $N(\mu, \sigma^2)$

$$T((X_1, \ldots, X_n)) = (\bar{X}, S^2)$$

Example iid Ber(p)

$$T((X_1, \ldots, X_5)) = \sum_{i=1}^{5} X_i$$

"Clearly" knowing $T(X)$ I know no more about θ than I do if I know X itself

Q: When do I really "lose nothing" in going from X to $T(X)$?

Example $X_1, X_2, \ldots, X_5$ iid Ber(p)

$T(X) = \sum X_i$

$$f(x|p) = I\begin{bmatrix}\text{all } x_i \\ \text{are } 0 \text{ or } 1\end{bmatrix} p^{\sum x_i} (1-p)^{5-\sum x_i}$$

$$\overset{\shortparallel}{P_p\left[X = x\right]}$$

Consider $P_p\left[X = x \mid T(X) = t\right]$

$$= \frac{P_p\left[X = x \text{ and } T(X) = t\right]}{P_p\left[T(X) = t\right]}$$

$$= \begin{cases} 0 & \text{if } \sum x_i \neq t \\ \dfrac{p^t(1-p)^{5-t}}{\binom{5}{t} p^t (1-p)^{5-t}} & \text{if } \sum x_i = t \end{cases}$$

$$= \begin{cases} 0 & \text{if } \sum x_i \neq t \\ \dfrac{1}{\binom{5}{t}} & \text{if } \sum x_i = t \end{cases}$$

this prescription doesn't depend on p !!!

in some sense, this says that if I know T(X) I don't really need to know X itself – this conditional dsn (that is uniform over x with $\sum x_i = t$) doesn't provide any information about p beyond what is in T(X)

In fact, if you give me $T(X)$ I can use the value of $T(X)$ plus some randomization that doesn't depend on p to invent X^* with the same dsn as X — how?

$$X^* \sim \text{uniform on } \{(x_1, \ldots, x_5) \mid \sum x_i = T(X)\}$$

This motivates

Def In the statistical model for X $\mathcal{P} = \{P_\theta\}$ we'll say that the statistic $T(X)$ is sufficient

for θ (or for the family/model $\mathcal{P}$)
provided for each t the conditional dsn of

$$X \mid T(X) = t$$

is the same for all $\theta \in \Theta$ (is free of dependence upon θ)

How to recognize a sufficient statistic?

Example $X_1, \ldots, X_5$ Bernoulli (p)

$T(X) = \sum X_i$ is sufficient for p

Theorem 1 (Fisher/Neyman/Halmos/Savage)

(Factorization Theorem) If $f(x|\theta)$ is either a pdf for X on $\mathbb{R}^k$ or a pmf for X

Then $T(X)$ is sufficient for θ $\iff$ There are nonnegative functions $g(t,\theta)$ and $h(x)$ such that

$$f(x|\theta) = g(T(x),\theta)\,h(x)$$

Interpretation: shape of the loglikelihood depends on X only through $T(X)$