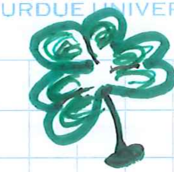
Lect. 24
Wed. 3/17

St. Patrick's Day!



Sufficient Statistics

A statistic is a "measurable" function of the observed data in a random experiment...

$Y \in \Gamma$

$T : \Gamma \longrightarrow \Delta$ $\quad$ $T(Y)$ is a random varible in its own right.

It's a statistic.

Goal: Is to estimate θ, a parameter indexing the family of distributions describing model

$$P_\theta : \theta \in \Lambda, \quad f_\theta(y) \quad \theta \in \Lambda$$

Usually we think of θ in the non-Bayesian way.

Q: Is possible that some $T(Y)$ carries info needed to est. θ?

Π

Bernoulli Trials

$Y_1\ Y_2 \cdots Y_m$ 0,1 rand. vars.
iid

$$P_\theta(y_i) = \begin{cases} \theta & y_i = 1 \\ 1-\theta & y_i = 0 \end{cases}$$

$$= \underbrace{\theta^{y_i}(1-\theta)^{1-y_i}}$$

$$P_\theta(y) = \prod_{i=1}^{m} P_\theta(y_i) = \theta^k (1-\theta)^{m-k}$$

$(y_1\, y_2 \cdots y_m)^T$

where
$$k = \sum_{i=1}^{m} y_i$$
$= \#$ of 1s in y
$\triangleq T(y)$

Let K be the corresp. r.v.
It is well Known that

$$P_\theta(k) = \binom{m}{k} \theta^k (1-\theta)^{m-k}$$

Consider the joint pmf of Y and K:

$$p_\theta(y,k) = P_\theta(\{Y=y, K=k\})$$

$$= \begin{cases} p_\theta(y) & \text{if } k = \#\text{ ones in } y \\ 0 & \text{otherwise} \end{cases}$$

With this the conditional pmf is

$$p_\theta(y|k) = \frac{p_\theta(y,k)}{p_\theta(k)} = \frac{p_\theta(y)}{p_\theta(k)} \quad \text{if } k = \#\text{ ones in } y.$$

$$= \frac{\theta^k(1-\theta)^{m-k}}{\binom{m}{k}\theta^k(1-\theta)^{m-k}} = \frac{1}{\binom{m}{k}} \quad \text{if } k = \#\text{ ones in } y.$$

$\Rightarrow$ Condition prob. of y given k is "unif. dist." on all of the $\binom{m}{k}$ seqs. y which have k ones.

See that the cond. dist. of Y given $K=k$ is indep. of θ.

$\Rightarrow$ The pmf of Y is a scaled version of that for K..

$$P_\theta(y) = \underbrace{P_\theta(y|k)}_{\binom{m}{k}^{-1} \text{ and indep of } \theta} P_\theta(k)$$

$\Rightarrow$ The dependence of the original random sample Y on the param θ is carried by the statistic $T(Y)$.

(Classical Suff. Stat.)

Def: A statistic $T(Y)$ is sufficient for the family P_θ $\theta \in \Lambda$ of dists. for Y if the cond. dist. of Y given $T(Y)$ does not dep. on θ for any $\theta \in \Lambda$.

Bayesian Sufficiency

Θ a rand. vector.

$T(Y)$ is said to be sufficient for est. Θ if

$$p_{\Theta|T}(\theta|t) = p_{\Theta|Y}(\theta|y) \qquad t = t(y)$$

Classical → comes from Fisher 1922

Bayes → Kolmogorov 1942

$$p_\Theta(y) = p_{Y|\Theta}(y|\theta).$$

Another "Look" At Sufficiency (Van Trees Book)

- A likelihood ratio in detection was our 1st example of a suff. statistic.
- The notion of s.s. was a byproduct of standard procedure for solving a H.T. ** Didn't Save Any Work**

$\Longrightarrow$ Most important ... Find s.s. before lots of work.

Say $Y = (Y_1, Y_1, \dots Y_n)^T$ a vector of observations.
$T(Y)$ be a one-dim. statistic.

Imagine a coord transformation

$$X = \tilde{T}(Y) = (T(Y), \underbrace{X_2, \dots X_n}_{Z})^T \qquad \tilde{T} \text{ is invertible}$$

Must be able to write the likelihood ratio

$$L(y) = \tilde{L}(t,z) = \frac{p_{T,Z|H_1}(t,z|H_1)}{p_{T,Z|H_0}(t,z|H_0)}$$

$$\tilde{L}(t,z) = \frac{p_{Z|T,H_1}(z|t,H_1)\, p_{T|H_1}(t|H_1)}{p_{Z|T,H_0}(z|t,H_0)\, p_{T|H_0}(t|H_0)}$$

If $T(Y)$ is a suff. stat., then $L(y)$ must reduce to some $\tilde{L}(t)$.

Must be able to cancel terms

ie

$$p_{Z|T,H_1}(z|t,H_1) = p_{Z|T,H_0}(z|t,H_0).$$

Choosing a scalar suff. stat. is equiv. to picking a coord syst. in which one coord contains all info needed to make dec.