

March 1, 1996, Friday, Lecture 13 :

Summary of preceding lecture:

Properties of estimators ; after efficiency and a lower bound on an estimator's variance ( $\frac{1}{nT(\theta)}$ ), I introduced the notion of sufficiency of an estimator, if an estimator is sufficient for a parameter  $\theta$  we can compute just that estimate and throw away all the other data.

The rigorous definition of a sufficient statistic is that it is suff. iff the conditional distribution (density or frequency) of the vector  $\underline{X}$  does not depend on  $\theta$  for any value of  $T = t$ .

I showed the example of the binomial, I am now going to give a necessary and sufficient condition for sufficiency:

**Theorem 2.6.1** *A nec. and suff. condition for  $T(\underline{X}) \equiv T(X_1, X_2, \dots, X_n)$  to be sufficient for a parameter is that the joint dist. (density or frequency) factors into two parts, one that depends on  $\hat{\theta}$  and on  $\underline{x}$  only through  $T(\underline{x})$  the other that does not depend on  $\theta$  :*

$$f(x_1, x_2, \dots, x_n | \theta) = g[T(x_1, x_2, \dots, x_n), \theta] h(x_1, x_2, \dots, x_n)$$

or

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

Proof: the condition is sufficient, i.e. if we have the condition we will have sufficiency.

First partition

$$\begin{aligned} P(T = t) &= \sum_{T(\underline{x})=t} P(\underline{X} = \underline{x}) \\ &= g(t, \theta) \sum_{T(\underline{x})=t} h(\underline{x}) = g(t, \theta) H(\underline{x}) \\ P(\underline{X} = \underline{x} | T = t) &= \frac{P(\underline{X} = \underline{x}, T = t)}{P(T = t)} \\ &= \frac{h(\underline{x}) g(t, \theta)}{H(\underline{x}) g(t, \theta)} \end{aligned}$$

Cancellation giving the result.

The other direction, i.e. sufficiency implies the condition:  $T$  is sufficient for  $\theta$  means we can write:  $P(\underline{X} = \underline{x} | T = t)$  as a function of  $\underline{x}$ , call it  $h$ :  $P(\underline{X} = \underline{x} | T = t) = h(\underline{x})$ , we then have:

$$P(\underline{X} = \underline{x} | \theta) = P(\underline{X} = \underline{x} | T = t) P(t = t | \theta) = h(\underline{x}) g(t, \theta)$$

## Exponential Families

$$f(x | \theta) = \exp[c(\theta) K(x) + d(\theta) + S(x)]$$

Joint density of an iid sample from this distribution will be :

$$\begin{aligned} f(\underline{x} | \theta) &= \prod \exp[c(\theta) K(x_i) + d(\theta) + S(x_i)] \\ &= \exp[c(\theta) \sum K(x_i) + nd(\theta)] \exp[\sum S(x_i)] \end{aligned}$$

So that  $T(\underline{x}) = \sum K(x_i)$  is a sufficient statistic.

### Bernouilli Example

$P(X = x) = \theta^x (1 - \theta)^{1-x} = \exp[x \log(\frac{\theta}{1-\theta}) + \log(1 - \theta)]$   $K(x) = x, T = \sum X_i$  is the sufficient statistic.

### Normal Example

Corollary:

If  $T$  is sufficient for  $\theta$  the mle is a function of  $T$ .

Proof:

The mle is built by maximising  $f(\underline{x} | \theta)$  which can be factored as:  $g(T, \theta)h(\underline{x})$  the dependence on  $\theta$  is only through  $T$ . To maximise this we only need to look at  $g(T, \theta)$ .

The following quantifies how much better it can be to use a sufficient statistic as a basis for an estimator, it always provides a method for improving an estimator.

**Theorem 2.6.2** *Let  $\hat{\theta}$  be any finite-varianced estimator of  $\theta$ . Suppose that we have a sufficient statistic for  $\theta$  we call  $T$ . Now taking as a new estimate  $\tilde{\theta} = E(\hat{\theta} | T)$  we will have a better estimator because it has smaller MSE:*

$$E(\tilde{\theta} - \theta)^2 \leq E(\hat{\theta} - \theta)^2$$

*The equality is strict unless  $\hat{\theta} = \tilde{\theta}$ .*

Example of Rao-Blackwellisation:

$X_1, X_2, \dots, X_n \sim \mathcal{N}(\theta, \sigma^2)$  we want to estimate  $\theta$ , using the silly estimate:  $g(\underline{X}) = X_1$ , and we know a sufficient statistic:  $X_1 + X_2 + \dots + X_n$ . Then the Rao-Blackwellisation would give us:

$$E[X_1 | X_1 + X_2 + \dots + X_n] = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Because  $E(X | X + Y) + E(Y | X + Y) = 2E(X | X + Y) = E(X + Y | X + Y) = X + Y$ . So just the one step of conditioning on a sufficient statistic took us a long way.

Extension to other loss functions than the MSE, any convex  $W(\tilde{\theta}, \theta)$  is such that Rao-Blackwellisation makes things better.

# Chapter 3

## Testing Hypotheses

Statistical Hypotheses testing is a formal means of choosing between two distributions on the basis of a particular statistic or random variable generated from one of them.

Comparisons of two normals with known variances and unknown means.

### 3.1 Neyman-Pearson Paradigm

- Null hypothesis  $\mathcal{H}_0$
- Alternate hypothesis  $\mathcal{H}_A$  or  $\mathcal{H}_1$

There are two type of hypotheses, simple ones where the hypothesis completely specifies the distribution. (In the case of the comparisons of the normals above, both hypotheses were simple).

Here is an example when they are both composite:

- $X_i \sim \text{Poisson}$  with unknown parameter
- $X_i$  is not Poisson

Here is an example of one of each kind: ESP experiment : guess the suit of 52 cards with replacement.

- 

$$\mathcal{H}_0 T \sim \mathcal{B}(1/4, n = 52)$$

- 

$$\mathcal{H}_1 T \sim \mathcal{B}(p, n = 52), p > 1/4$$

Intuitive tradeoff : the burnt toast and teh smoke detector.