17. Examples of sufficient statistics

17.1 Couple of notes about sufficient statistics

(i) Sufficiency is a property of the family of distributions, not of the particular parameter. "Sufficient for θ" means also sufficient for any function of θ. See examples of this both in 17.2 and 17.3.

(ii) If T is sufficient for θ (in some fanily of distributions), then any 1-1 function of T is sufficient. (LM Question 5.6.4, P. 405). Sufficient statstics are unique only up to 1-1 functions.

(iii) Another way to think of this is that T partitions the sample space. All 1-1 functions of each other will give the same partition.

17.2: Example from Midterm-1

(i) $X_1, ..., X_n$ i.i.d. $\mathcal{P}\mathrm{o}(\theta)$; we want to estimate $\theta^2 + \theta$.

(ii) For this family of distributions $T = \sum_{i=1}^n X_i$ is sufficient. Note it does not matter what function of θ we are estimating – sufficiency tells us the best estimators must be based on T.

(iii) $S = (1/n) \sum_{i=1}^n X_i^2$ is unbiased estimator of $\theta^2 + \theta$, but it is not a function of T.

(iv) $\overline{X_n} = T/n$, $\mathrm{E}(\overline{X_n}) = \theta$ and $\mathrm{E}(\overline{X_n}^2) = \theta^2 + \theta/n$.

(v) The MoM estimator $W = \overline{X_n}^2 + \overline{X_n}$ in midterm is a function of T but not unbiased.

(vi) However $W^* = \overline{X_n}^2 + \overline{X_n} - \overline{X_n}/n$ is a function of T, and is unbiased for $\theta^2 + \theta$.

(vii) So theory tells us W^* will have smaller variance that S.

17.3 A sample from a Normal distribution

(i) $X_1, ..., X_n$ are i.i.d. $N(\mu, \sigma^2)$; $f_X(x; \mu, \sigma^2) \;=\; (1/\sqrt{2\pi\sigma^2}) \exp(-(x-\mu)^2/(2\sigma^2))$

(ii)

$$\begin{aligned} L_n(\mu, \sigma^2) &= \prod_{i=1}^n f_X(x_i; \mu, \sigma^2) \;=\; (1/\sqrt{2\pi\sigma^2})^n \exp(-\sum_{i=1}^n (x_i - \mu)^2/(2\sigma^2)) \\ \ell_n(\mu, \sigma^2) &= \text{const.} \; - (n/2)\log(\sigma^2) - (1/2\sigma^2) \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

(iii) Note $\sum_{i=1}^n (x_i - \mu)^2 \;=\; S^2 + n(\overline{x_n} - \mu)^2$ where $s^2 \;=\; \sum_{i=1}^n (x_i - \overline{x_n})^2$ so

$$\ell_n(\mu, \sigma^2) \;=\; \text{const.} \; - (n/2)\log(\sigma^2) - (1/2\sigma^2)(s^2 + n(\overline{x_n} - \mu)^2)$$

(iv) So by the factorization criterion $(\overline{x_n}, S^2)$ is sufficient for (μ, σ^2), where $S^2 \;=\; \sum i = 1^n (X_i - \overline{X_n})^2$ so

(v) Messy algebra shows the MLE of μ is $\overline{X_n}$ and of σ^2 is S^2/n (see LM. Example 5.2.4; Pp.353-4).

(vi) Note these are the same as the MoM estimators. $\overline{X_n}$ is unbiased for μ, but S^2/n is biased for σ^2 (but asymptotically unbiased).

(vii) $\overline{X_n}$ is sufficient for μ if σ^2 is known.
S^2 is NOT sufficient for σ^2 if μ is known.
Instead it would be $\sum_{i=1}^n (X_i - \mu)^2$ – see the Homework Exercise 5.2.14 (LM. P.357).

Friday Feb 19: Nick Basch to teach

1. Cramer-Rao Lower Bound (LM 5.5)

We would like unbiased estimators with small variance.

How small can the variance be?

Turns out there is a formula, based on the log-likelihood.

This formula is known as the Cramer-Rao Lower Bound (CRLB): LM P.394

2. Minimum variance unbiased estimators

Estimators that have variance equal to the CRLB have minimum variance.

They are called Minimum Variance Unbiased Estimators (MVUE)

Sometimes they are also called *efficient* estimators (LM. P.396)

3. Why use maximum likelihood estimators?

We have seen some reasons already – they are always functions of the sufficient statistics, and the Rao-Blackwell Theorem tells us they will be better than estimators that are not.

Here are more reasons (for large sample size n):

As sample size $n \to \infty$, and subject to some conditions on the pdf/pmf that are beyond what we need to worry about:

(i) MLEs are approximately unbiased

(ii) MLEs achieve the CRLB

That is MLE's are *asymptotically* unbiased and efficient.

This will mean (subject to same conditions) MLE's are consistent (LM. P.409)