WMS: 9.4

L. Sufficiency:

1. Sufficiency Criterion

   a. How much of information do we have to consider,

   b. and how much can we toss away as not giving information about the quantity of interest?

   c. Express generic data as $X_1, \cdots, X_n = \boldsymbol{X}$, with observed values $x_1, \cdots, x_n = \boldsymbol{x}$.

2. Sufficiency Example:

   a. $\boldsymbol{X} \sim \text{Bin}(m, \theta)$ an ind. sample.

   b. $\hat{\theta} = \sum_{i=1}^{n} X_i/(mn)$ is an unbiased, consistent estimator of θ.

   c. Is there any other part of the data, other than that summarized by $\hat{\theta}$, that gives information about θ?

   d. The separate p.m.f.s for the variables are $\binom{m}{x_i}\pi^{x_i}(1-\pi)^{m-x_i}$.

   e. Hence the joint p.m.f. is $p_{\boldsymbol{X}}(\boldsymbol{x};\pi) = \prod_{i=1}^{n} \binom{m}{x_i}\pi^{x_i}(1-\pi)^{m-x_i}$.

      i. Collect exponents

$$p_{\boldsymbol{X}}(\boldsymbol{x};\pi) = \pi^{\sum_{i=1}^{n} x_i}(1-\pi)^{mn-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \binom{m}{x_i}$$

      ii. Substitute in statistic value

$$p_{\boldsymbol{X}}(\boldsymbol{x};\pi) = \pi^{mn\hat{\theta}}(1-\pi)^{mn-mn\hat{\theta}} \prod_{i=1}^{n} \binom{m}{x_i}$$

iii. Calculate marginal probability from distribution of sum of binomials:

$$p(\hat{\theta};\pi) = \binom{mn}{mn\hat{\theta}} \pi^{mn\hat{\theta}}(1-\pi)^{mn-mn\hat{\theta}};$$

f. Hence $p_{\boldsymbol{X}|\hat{\theta}}(\boldsymbol{x}|\hat{\theta};\pi) = \prod_{i=1}^{n} \binom{m}{x_i} / \binom{mn}{\sum_{i=1}^{n} x_i}$.

g. Hence the additional information given by the X_i beyond their total tells nothing about π.

3. Sufficiency Definition:

a. $T(\boldsymbol{X})$ is *sufficient* for θ if the distń of $\boldsymbol{X}$ conditional on T doesn't depend on θ.

b. *factorization theorem*: T is sufficient if and only if full p.m.f. can be factored as

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = g(t(\boldsymbol{x});\theta)u(\boldsymbol{x}).$$

i. T sufficient $\Rightarrow$ p.m.f. $p_{\boldsymbol{X}}(\boldsymbol{x};\theta)$ is $p_T(t;\theta)p_{\boldsymbol{X}|T}(\boldsymbol{x}|t(\boldsymbol{x}))$.

- the latter factor independent of θ

ii. p.m.f. factors as described $\Rightarrow$ $p_{\boldsymbol{X}|T}(\boldsymbol{x}|t;\pi) = g(t;\theta)u(\boldsymbol{x}) / \sum_{\boldsymbol{z}|t(\boldsymbol{z})=t} g(t;\theta)u(\boldsymbol{z}) = u(\boldsymbol{x}) / \sum_{\boldsymbol{z}|t(\boldsymbol{z})=t} u(\boldsymbol{z})$.

- The conditional p.m.f. does not depend on θ.

c. The ideas and theorems above also hold for densities.

d. Entire data set $\boldsymbol{X}$ is sufficient.

i. For independent data, so is ordered data set.

ii. Generally want more concise summary.

4. Example: *********** ***********

a. Consider $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$.

b. The joint p.d.f. is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{\exp(-(x_i - \mu)^2/(2\sigma^2))}{\sigma\sqrt{2\pi}}$$

i. Simplify exponentials:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp(-(\sum_{i=1}^{n}(x_i - \mu)^2)/(2\sigma^2))}{\sigma^n (2\pi)^{n/2}}$$

ii. Expand squares:

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\exp\left(\frac{-\sum_{i=1}^{n} x_i^2 + 2\mu\sum_{i=1}^{n} x_i - n\mu^2}{2\sigma^2}\right)}{(\sigma^n (2\pi)^{n/2})}$$

iii. Simplify to obtain density $\frac{\exp((2\mu\sum_{i=1}^{n} x_i - n\mu^2)/(2\sigma^2)) \times \exp((-\sum_{i=}}{\sigma^n (2\pi)^{n/2}}$

c. If σ is known without looking at the data, sum of observations is sufficient.

i. Factorization shows that $\sum_{i=1}^{n} X_i$ is sufficient for μ.

ii. So is $\hat{\mu} = T/n$.

iii. $\hat{\mu}$ is a good estimator but T is not.

iv. Factorization shows that $(\sum_{i=1}^{n} X, \sum_{i-1}^{n} X_i^2)$ is sufficient for (μ, σ^2) .

v. So is $\bar{X}, s^2 = \sum_{i-1}^{n}(X_i - \bar{X})^2/(n-1)$

5. Poisson Example

a. $X, Y \sim$ P(θ)

b. Consider summary $\hat{\mu} = \frac{1}{3}X + \frac{2}{3}Y$

i. $\hat{\mu} = \frac{2}{3} \Rightarrow X = 2$ and $Y = 0$ or $X = 0$ and $Y = 1$

ii. $\mathrm{P}\left[X = 2|\hat{\mu} = \frac{2}{3}\right] =$

$$\frac{\exp(-\mu)\mu^2/2!\exp(-\mu)}{\exp(-\mu)\mu^2/2!\exp(-\mu) + \exp(-\mu)\exp(-\mu)\mu^1/1!} = \frac{\mu^2}{\mu^2 + 2\mu},$$

iii. depends on μ : $\hat{\mu}$ not sufficient

c. Consider summary $\hat{\mu} = \frac{1}{2}X + \frac{1}{2}Y$

i. $\mathrm{P}\left[X = x|\hat{\mu} = u\right] =$

$$\frac{\exp(-\mu)\mu^x/x!\exp(-\mu)\mu^{2u-x}/(2u-x)!}{\exp(-2\mu)\mu^{2u}/(2u)!} = \frac{2u!}{x!(2u-x)!},$$

ii. does not depend on μ : sufficient

6. Example where sufficient statistic doesn't tell whole story:

a. A collection of cars is inspected for defective wheels

b. Estimate the proportion π of wheels which are defective.

c. Under the binomial model, the sample proportion is sufficient for inference on π .

Am

d. Table 2 contains two scenarios:

Scenario 1: # of wheels defective	# of times observed	Scenario 2: # of wheels defective	# of times observed
0	5	0	44
1	19	1	0
2	36	2	0
3	27	3	0
4	13	4	56
Total	100	Total	100

i. Both scenarios give the same estimate of π

ii. the second case gives strong evidence that the binomial model is wrong.

iii. Hence the sufficient statistic tells about the parameters in the model; remainder tells about the suitability of the model itself.