

# Sufficient Statistics: Lecture XXI

Charles B. Moss

October 12, 2026

# Reading Questions

- One field in Industrial and Systems Engineering is referred to as Operations Research (at least it was when I took it). A classical problem in that field is how many lines to open up. This is based on the probability that  $k$  people will arrive needing a service.
- Poisson Distribution

$$f(x; \lambda) = \frac{\lambda^x \exp(-\lambda)}{x!}, \quad x = 0, 1, 2, \dots \quad (1)$$

What would this maximum likelihood problem look like?

- Likelihood function

$$L(\lambda; x_1, x_2, \dots, x_n) = \prod_{j=1}^n \exp(-\lambda) \frac{1}{x_j!} \lambda^{x_j} \quad (2)$$
$$\Rightarrow \ln(L) = -n\lambda - \sum_{j=1}^n \ln(x_j) + \ln(\lambda) \sum_{j=1}^n x_i$$

1 Data Reduction

2 Sufficiency Principle

# Data Reduction

- The typical mode of operation in statistics is to use information from a sample  $X_1, \dots, X_n$  to make inferences about an unknown parameter  $\theta$ .
  - Put slightly differently, the researcher summarizes the information in the sample (or the sample values) with a statistic.
  - Thus, any statistic  $T(X)$  summarizes the data, or reduces the information in the sample to a single number. We use only the information in the statistic instead of the entire sample.

- Put in a slightly more mathematical formulation, the statistic partitions the sample space into two sets
- Defining the sample space for the statistic

$$T = \{ t : t = T(x), x \in X \} \quad (3)$$

- Thus, a given value of a sample statistic  $T(x)$  implies that the sample comes from a space of sets  $A_t$  such that  $t \in T$ ,  $A_t = \{ x : T(x) = t \}$ .
- The second possibility (that is ruled out by observing a sample statistic of  $T(x)$ ) is  $A_t^C = \{ x : T(x) \neq t \}$ .
- Thus, instead of presenting the entire sample, we could report the value of the sample statistic.

# Sufficiency Principle

- Intuitively, a sufficient statistic for a parameter is a statistic that captures all the information about a given parameter contained in the sample.
- Sufficiency Principle: If  $T(X)$  is a sufficient statistic for  $\theta$ , then any inference about  $\theta$  should depend on the sample  $X$  only through the value of  $T(X)$ . That is, if  $x$  and  $y$  are two sample points such that  $T(x) = T(y)$ , then the inference about  $\theta$  should be the same whether  $X = x$  or  $X = y$ .
  - **Definition 6.2.1** (Cassela and Berger) A statistic  $T(X)$  is a sufficient statistic for  $\theta$  if the conditional distribution of the sample  $X$  given  $T(X)$  does not depend on  $\theta$ .

- **Definition 7.2.1** (Hogg, Craig, and McKean) Let  $X_1, X_2, \dots, X_n$  denote a random sample of size  $n$  from a distribution that has a pdf or pmf  $f(x; \theta)$ ,  $\theta \in \Theta$ . Let  $Y_1 = u_1(X_1, X_2, \dots, X_n)$  be a statistic whose pdf or pmf is  $f_{Y_1}(y_1; \theta)$ . Then  $Y_1$  is a sufficient statistic for  $\theta$  if and only if

$$\frac{f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)}{f_{Y_1}[u_1(x_1, x_2, \dots, x_n); \theta]} = H(x_1, x_2, \dots, x_n) \quad (4)$$

where  $H(x_1, x_2, \dots, x_n)$  does not depend on  $\theta \in \Theta$ .

- **Theorem 6.2.2** (Cassela and Berger) If  $p(x|\theta)$  is the joint pdf or pmf of  $X$  and  $q(t|\theta)$  is the pdf or pmf of  $T(X)$ , then  $T(X)$  is a sufficient statistic for  $\theta$  if, for every  $x$  in the sample space, the ratio of

$$\frac{p(x|\theta)}{q(t|\theta)} \quad (5)$$

is a constant as a function of  $\theta$ .

- **Example 6.2.4** (Cassela and Berger) Normal sufficient statistic: Let  $X_1, \dots, X_n$  be independently and identically distributed  $N(\mu, \sigma^2)$  where the variance is known. The sample mean  $T(X) = \bar{X} = 1/n \sum_{i=1}^n X_i$  is the sufficient statistic for  $\mu$ . Starting with the joint distribution function

$$\begin{aligned} f(x|\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right]. \end{aligned} \tag{6}$$

- Next, we add and subtract  $\bar{x}$  yielding

$$f(x|\mu\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\sum_{i=1}^n \frac{(x_i - \bar{x} + \bar{x} - \mu)^2}{2\sigma^2} \right] \quad (7)$$

Note that

$$\begin{aligned} [(x_i - \bar{x}) + (\bar{x} - \mu)]^2 &= \\ (x_i - \bar{x})^2 + 2(\bar{x} - \mu)(x_i - \bar{x}) + (\bar{x} - \mu)^2 \\ \Rightarrow \sum_{i=1}^N [(x_i - \bar{x}) + (\bar{x} - \mu)]^2 &= \\ \sum_{i=1}^N (x_i - \bar{x})^2 + 2(\bar{x} - \mu) \sum_{i=1}^N (x_i - \bar{x}) + n(\bar{x} - \mu)^2 \end{aligned} \quad (8)$$

- Hence,

$$f(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right] \quad (9)$$

- Note that we can rewrite the exponential term as

$$\exp \left[ -\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{2\sigma^2} \right] \exp \left[ -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right] \quad (10)$$

- Given that the distribution of the sample mean is

$$q(T(X) | \theta) = \frac{1}{\left(2\pi \frac{\sigma^2}{n}\right)^{1/2}} \exp \left[ -\frac{n(\bar{x} - \mu)^2}{2\sigma^2} \right] \quad (11)$$

- The ratio of the information in the sample to the information in the statistic becomes

$$\begin{aligned}\frac{f(x|\theta)}{q(T(x)|\theta)} &= \frac{\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\sum_{i=1}^n \frac{(x_i - \bar{x} + \bar{x} - \mu)^2}{2\sigma^2}\right]}{\frac{1}{\left(2\pi\frac{\sigma^2}{n}\right)^{1/2}} \exp\left[-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}\right]} \\ &= \frac{1}{n^{1/2} (2\pi\sigma^2)^{\frac{n-1}{2}}} \exp\left[-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2}\right]\end{aligned}\tag{12}$$

which does not depend on  $\mu$ .

- **Theorem 6.2.6** (Cassela and Berger) (Factorization Theorem) Let  $f(x|\theta)$  denote the joint pdf or pmf of a sample  $X$ . A statistic  $T(X)$  is a sufficient statistic for  $\theta$  if and only if there exists functions  $g(t|\theta)$  and  $h(x)$  such that, for all sample points  $x$  and all parameter points  $\theta$

$$f(x|\theta) = g(T(x)|\theta) h(x). \quad (13)$$

- **Definition 6.2.11** (Cassela and Berger) A sufficient statistic  $T(X)$  is called a minimal sufficient statistic if, for any other sufficient statistic  $T'(X)$ ,  $T(X)$  is a function of  $T'(X)$ .

# Another Simple Example - the Exponential Distribution

- Let's start with the exponential distribution

$$\begin{aligned} f(x_k|\lambda) &= \lambda \exp(-\lambda x_k) \\ L(\{x_k\}_{k=1}^N|\lambda) &= \prod_{k=1}^N \lambda \exp(-\lambda x_k) \\ &= \lambda^N \exp\left(-\lambda \sum_{k=1}^N x_k\right) \end{aligned} \tag{14}$$

Based on this result, let us consider the statistic

$$T(x) = \sum_{k=1}^N x_k.$$

- Following our approach to normal distribution, we add and subtract  $T$  to the exponential term

$$\begin{aligned} L(\{x_k\}_{k=1}^N | \lambda) &= \lambda^N \exp \left( -\lambda \left( \sum_{k=1}^N x_k - T + T \right) \right) \\ &= \lambda^N \exp \left( -\lambda \left( \sum_{k=1}^N x_k - T \right) \right) \exp(-\lambda T) \end{aligned} \quad (15)$$

- Since  $\sum_{k=1}^N x_k = T$ , the first exponential term simply becomes one (i.e.,  $\exp(-\lambda \times 0) = 1$ ), so we can rewrite Equation 15 as

$$L(\{x_k\}_{k=1}^N | \lambda) = 1 \times (\lambda^N \exp(-\lambda T)) \quad (16)$$

- A couple of ways to look at this. First, the second part of Equation 16 is the probability density function for  $T(x)$  given  $\lambda$ .

$$g(T|\lambda) = \lambda^N \exp(-\lambda T) \quad (17)$$

The first part is the remaining distribution. This would imply a factorization.

- Following the division approach we used earlier, we can demonstrate that

$$\begin{aligned} \frac{L(\{x_k\}_{k=1}^N|\lambda)}{g(T|\lambda)} &= \frac{1 \times (\lambda^N \exp(-\lambda T))}{\lambda^N \exp(-\lambda T)} \\ &= 1 \end{aligned} \quad (18)$$

# Reading

- Moss 7.6 - 7.7
- Salsburg, D. *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* Chapter 1 – The Lady Tasting Tea