Lecture 9: Sufficiency and the Rao-Blackwell Theorem

MATH 667-01
Statistical Inference
University of Louisville

October 15, 2019
Last modified: 10/15/2019

- We will define sufficiency and prove the Neyman-Fisher Factorization Theorem[1].
- We also discuss and prove the Rao-Blackwell Theorem[2].
- The proof of the Rao-Blackwell Theorem uses iterated expectation formulas[3].

[1] CB: Sections 6.1 and 6.2, HMC: Section 7.2
[2] CB: Section 7.3. HMC: Section 7.3
[3] CB: Section 4.4, HMC: Section 2.3

Sufficiency

- Now we examine data summarization and data reduction when making inferences about a fixed but unknown parameter θ based on a sample $X_1, \ldots, X_n$.
- When the sample size n is large, simply being given a list of the observed sample values $x_1, \ldots, x_n$ is not very useful.
- Instead, it is useful to provide a statistic $T(X_1, \ldots, X_n)$ and use the observed value $T(x_1, \ldots, x_n)$ to summarize the information about θ in the observed sample.
- Let $\mathcal{X}$ denote the sample space of $X_1, \ldots, X_n$. Then $\mathcal{T} = \{t : t = T(\boldsymbol{x}) \text{ for some } \boldsymbol{x} \in \mathcal{X}\}$ is the image of $\mathcal{X}$ under T.
- So $T(\boldsymbol{x})$ partitions $\mathcal{X}$ into sets $A_t = \{\boldsymbol{x} : T(\boldsymbol{x}) = t\}$ for $t \in \mathcal{T}$.

- The goal of the *sufficiency principle* is to summarize data while not losing information about θ.
- *Definition L9.1*:[4] A statistic $T(\boldsymbol{X})$ is a *sufficient statistic for* θ if the conditional distribution of the sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ given the value of $T(\boldsymbol{X})$ does not depend on θ.
- That is, $T(\boldsymbol{X})$ is sufficient for θ if the pdf/pmf $f_{\boldsymbol{X}|T(\boldsymbol{X})=T(\boldsymbol{x})}(\boldsymbol{x};\theta)$ is the same for all θ.

[4] CB: Definition 6.2.1 on p.272, HMC: page 420

- *Theorem L9.1*:[5] If $p(\boldsymbol{x};\theta)$ is the joint pdf/pmf of $\boldsymbol{X}$, and $q(t;\theta)$ is the pdf/pmf of $T(\boldsymbol{X})$, then $T(\boldsymbol{X})$ is a sufficient statistic for θ if, and only if, for every $\boldsymbol{x}$ in the sample space the ratio $p(\boldsymbol{x};\theta)/q(T(\boldsymbol{x});\theta)$ is constant as a function of θ.
- *Proof of Theorem L9.1*:

$$\begin{aligned} P_\theta(\boldsymbol{X}=\boldsymbol{x}|T(\boldsymbol{X})=T(\boldsymbol{x})) &= \frac{P_\theta\left(\boldsymbol{X}=\boldsymbol{x} \text{ and } T(\boldsymbol{X})=T(\boldsymbol{x})\right)}{P_\theta\left(T(\boldsymbol{X})=T(\boldsymbol{x})\right)} \\ &= \frac{P_\theta\left(\boldsymbol{X}=\boldsymbol{x}\right)}{P_\theta\left(T(\boldsymbol{X})=T(\boldsymbol{x})\right)} \\ &= \frac{p(\boldsymbol{x};\theta)}{q(T(\boldsymbol{x});\theta)}. \end{aligned}$$

So, $T(X)$ is sufficient if and only if the probability above is constant as a function of θ.

[5] CB: Theorem 6.2.2 on p.274, HMC: Definition 7.2.1 on p.421

- *Example L9.1*: Let $X_1, \ldots, X_n$ be iid Poisson(λ) random variables. Show that $\sum_{i=1}^{n} X_i$ is sufficient for λ.
- *Answer to Example L9.1*:

$$P\left((X_1, \ldots, X_n) = (x_1, \ldots, x_n) \middle| \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i\right) =$$

$$\frac{P\left((X_1, \ldots, X_n) = (x_1, \ldots, x_n)\right)}{P\left(\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i\right)} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda} / \left(\prod_{i=1}^{n} x_i!\right)}{(n\lambda)^{\sum_{i=1}^{n} x_i} e^{-n\lambda} / \left(\sum_{i=1}^{n} x_i\right)!}$$

since $\sum_{i=1}^{n} X_i \sim$ Poisson$(n\lambda)$. Simplifying this expression, we obtain $n^{-\sum_{i=1}^{n} x_i} \left(\sum_{i=1}^{n} x_i\right)! / \left(\prod_{i=1}^{n} x_i!\right)$ which does not depend on λ.

- We can use *Theorem L9.1* to verify that a statistic is sufficient for θ, but it is better to have a way of finding sufficient statistics without having a candidate in mind.
- This can be done with the following result known as the **Neyman-Fisher Factorization Theorem**.
- *Theorem L9.2*:[6] Let $f(\boldsymbol{x};\theta)$ denote the joint pdf/pmf of a sample $\boldsymbol{X}$. A statistic $T(\boldsymbol{X})$ is a sufficient statistic for θ if and only if there exist functions $g(t;\theta)$ and $h(\boldsymbol{x})$ such that, for all sample points $\boldsymbol{x}$ and all parameter points θ, $f(\boldsymbol{x};\theta) = g(T(\boldsymbol{x});\theta)h(\boldsymbol{x})$.

[6] CB: Theorem 6.2.6 on p.276, HMC: Theorem 7.2.1 on p.422

- *Sketch of proof of Theorem L9.2 for the discrete case*:
- Suppose $T(\boldsymbol{X})$ is a sufficient statistic. Then

$$\begin{aligned} f(\boldsymbol{x};\theta) &= P_\theta(\boldsymbol{X}=\boldsymbol{x}) \\ &= P_\theta\left(\boldsymbol{X}=\boldsymbol{x} \text{ and } T(\boldsymbol{X})=T(\boldsymbol{x})\right) \\ &= P_\theta\left(T(\boldsymbol{X})=T(\boldsymbol{x})\right) P_\theta\left(\boldsymbol{X}=\boldsymbol{x} | T(\boldsymbol{X})=T(\boldsymbol{x})\right) \\ &= g(T(\boldsymbol{x});\theta)h(\boldsymbol{x}). \end{aligned}$$

- Suppose that $f(\boldsymbol{x};\theta)=g(T(\boldsymbol{x});\theta)h(\boldsymbol{x})$. Then

$$\begin{aligned} \frac{f(\boldsymbol{x};\theta)}{q(T(\boldsymbol{x});\theta)} &= \frac{g(T(\boldsymbol{x});\theta)h(\boldsymbol{x})}{q(T(\boldsymbol{x});\theta)} \\ &= \frac{g(T(\boldsymbol{x});\theta)h(\boldsymbol{x})}{\sum_{\boldsymbol{y}\in A_{T(\boldsymbol{x})}} g(T(\boldsymbol{y});\theta)h(\boldsymbol{y})} \\ &= \frac{g(T(\boldsymbol{x});\theta)h(\boldsymbol{x})}{g(T(\boldsymbol{x});\theta)\sum_{\boldsymbol{y}\in A_{T(\boldsymbol{x})}} h(\boldsymbol{y})} = \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y}\in A_{T(\boldsymbol{x})}} h(\boldsymbol{y})} \end{aligned}$$

does not depend on θ.

- *Example L9.2*: Let $X_1, \ldots, X_n$ be iid random variables from a Normal$(\mu, 1)$ distribution. Find a sufficient estimator for μ.
- *Answer to Example L9.2*: Let $\boldsymbol{x} = (x_1, \ldots, x_n)$. The joint pdf of $X_1, \ldots, X_n$ is

$$
\begin{aligned}
f(\boldsymbol{x}; \mu) &= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \\
&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \exp\left(n\bar{x}\mu - \frac{n}{2}\mu^2\right) \\
&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right) e^{-\frac{n}{2}(\bar{x}-\mu)^2} \\
&= h(\boldsymbol{x}) g(\bar{x}; \mu)
\end{aligned}
$$

where $h(\boldsymbol{x}) = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$ does not depend on μ and $g(t; \mu) = e^{-\frac{n}{2}(t-\mu)^2}$. Thus, $\bar{X}$ is sufficient for μ.

- *Example L9.3*: Let $X_1, \ldots, X_n$ be iid random variables from a Uniform(α, ω) distribution where α and ω are real numbers with $\alpha \leq \omega$. Show that $(X_{(1)}, X_{(n)})$ is sufficient for (α, ω) where $X_{(1)} = \min_{i=1,\ldots,n} X_i$ and $X_{(n)} = \max_{i=1,\ldots,n} X_i$.

- *Answer to Example L9.3*: Let $\boldsymbol{x} = (x_1, \ldots, x_n)$. The joint pdf of $X_1, \ldots, X_n$ is

$$
\begin{aligned}
f(\boldsymbol{x}; \alpha, \omega) &= \frac{1}{(\omega - \alpha)^n} \prod_{i=1}^{n} I_{(\alpha,\omega)}(x_i) \\
&= \frac{1}{(\omega - \alpha)^n} I_{(\alpha,\infty)}(x_{(1)}) I_{(-\infty,\omega)}(x_{(n)}) \\
&= g(x_{(1)}, x_{(n)}; \alpha, \omega) h(\boldsymbol{x})
\end{aligned}
$$

where $g(t_1, t_2; \alpha, \omega) = \frac{1}{(\omega-\alpha)^n} I_{(\alpha,\infty)}(t_1)\, I_{(-\infty,\omega)}(t_2)$ and $h(\boldsymbol{x}) = 1$ does not depend on (α, ω). Thus, $(X_{(1)}, X_{(n)})$ is sufficient for (α, ω).

Rao-Blackwell Theorem

- Sufficient statistics are related to unbiased estimators through a well-known result known as the Rao-Blackwell Theorem.
- *Theorem L9.3*:[7] Let W be any unbiased estimator of $\tau(\theta)$, and let T be a sufficient statistic for θ. Define $\phi(t) = \mathsf{E}[W|T = t]$. Then
  (1) $\mathsf{E}_\theta[\phi(T)] = \tau(\theta)$ and
  (2) $\mathsf{Var}_\theta[\phi(T)] \leq \mathsf{Var}_\theta[W]$ for all θ;

  that is, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$.
- Consequently, conditioning any unbiased estimator on a sufficient statistic will uniformly "improve" the estimator, so the Rao-Blackwell Theorem shows that we only need to consider statistics which are functions of sufficient statistics when searching for a UMVUE.

[7] CB: Theorem 7.3.17 on p.342, HMC: Theorem 7.3.1 on p.427

- *Theorem L9.4*:[8] If X and Y are any two random variables, then

$$\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X|Y]],$$

provided that the expectations exist.

- *Theorem L9.5*:[9] For any two random variables X and Y,

$$\mathsf{Var}[X] = \mathsf{E}[\mathsf{Var}[X|Y]] + \mathsf{Var}[\mathsf{E}[X|Y]]$$

provided that the expectations exist.

[8] CB: Theorem 4.4.3 on p.164, HMC: Theorem 2.3.1(a) on p.114
[9] CB: Theorem 4.4.7 on p.167, HMC: page 114

Rao-Blackwell Theorem

- *Proof of Theorem L9.3*: Since T is sufficient, $W|T$ does not depend on θ and thus $\phi(T) = \mathsf{E}[W|T]$ is only a function of the sample and thus an estimator. Using the iterated formulas, we have

$$\mathsf{E}[\phi(T)] = \mathsf{E}[\mathsf{E}[W|T]] = \mathsf{E}[W] = \tau(\theta)$$

and

$$\begin{aligned}\mathsf{Var}[W] &= \mathsf{E}[\mathsf{Var}[W|T]] + \mathsf{Var}[\mathsf{E}[W|T]] \\ &= \mathsf{E}[\mathsf{Var}[W|T]] + \mathsf{Var}[\phi(T)] \\ &\geq \mathsf{Var}[\phi(T)]\end{aligned}$$

since $\mathsf{Var}[W|T] \geq 0$, and thus, $\mathsf{E}[\mathsf{Var}[W|T]] \geq 0$.

- *Example L9.4*: Let X_1 and X_2 be independent identically distributed (iid) Poisson(θ) random variables.
  (a) Find a sufficient statistic for θ.
  (b) Show that $W = \begin{cases} 1 & \text{if } X_1 = 0 \\ 0 & \text{otherwise} \end{cases}$
  is an unbiased estimator of $\tau(\theta) = e^{-\theta}$.
  (c) Compute $\mathsf{E}[W|X_1 + X_2 = y]$.
  (d) For the estimator W in part (b), find a uniformly better unbiased estimator of $e^{-\theta}$.

- *Answer to Example L9.4*: (a) The joint pmf of X_1 and X_2 is

$$\begin{aligned} f(x_1, x_2|\theta) &= f(x_1|\theta)f(x_2|\theta) = \frac{\theta^{x_1}e^{-\theta}}{x_1!}\frac{\theta^{x_2}e^{-\theta}}{x_2!} \\ &= \frac{\theta^{x_1+x_2}e^{-2\theta}}{x_1!x_2!} = g(x_1+x_2|\theta)h(x_1,x_2) \end{aligned}$$

where $g(t|\theta) = \theta^t e^{-2\theta}$ and $h(\boldsymbol{x}) = \frac{1}{x_1!x_2!}$. So, $X_1 + X_2$ is sufficient for θ.

- (b) $\mathsf{E}[W] = P(W=1) = P(X_1 = 0) = \frac{\theta^0 e^{\theta}}{0!} = e^{-\theta}$
- (c) Since $X_1 + X_2 \sim$ Poisson(2θ), we have

$$\begin{aligned} \mathsf{E}[W|X_1+X_2=y] &= P(W=1|X_1+X_2=y) \\ &= P(X_1=0|X_1+X_2=y) \\ &= \frac{P(X_1=0 \text{ and } X_1+X_2=y)}{P(X_1+X_2=y)} \\ &= \frac{P(X_1=0 \text{ and } X_2=y)}{P(X_1+X_2=y)} \end{aligned}$$

- *Answer to Example L9.4 continued*:

$$\begin{aligned}
\mathsf{E}[W|X_1+X_2=y] &= \frac{P(X_1=0 \text{ and } X_2=y)}{P(X_1+X_2=y)} \\
&= \frac{P(X_1=0)P(X_2=y)}{P(X_1+X_2=y)} \\
&= \frac{e^{-\theta}(\theta^y e^{-\theta}/y!)}{(2\theta)^y e^{-2\theta}/y!} \\
&= \frac{\theta^y}{(2\theta)^y} = \left(\frac{1}{2}\right)^y .
\end{aligned}$$

- (d) Since W is an unbiased estimator of $e^{-\theta}$ and X_1+X_2 is sufficient for θ (and consequently $e^{-\theta}$), the Rao-Blackwell Theorem implies that

$$\phi(X_1+X_2) = \mathsf{E}[W|X_1+X_2] = \left(\frac{1}{2}\right)^{X_1+X_2}$$

is a uniformly better unbiased estimator of $e^{-\theta}$.