1 Sufficiency

1.1 Definition

- so far, we have evaluated estimators using the concepts of *Bias* and *Efficiency*
- Another concept we may use to evaluate estimators is *Sufficiency*

Definition:
Let $X_1, ..., X_n$ be a random sample from a probability distribution with unknown parameter θ. Then the statistic $U = g(X_1, ..., X_n)$ is said to be *Sufficient* for θ if the conditional distribution of $X_1, ..., X_n$ given U does not depend on θ

- Sufficiency is meant to encapsulate the idea of information when we summarize the data (sample) with a statistic (estimator)
- In other words, if a statistic is sufficient, then we are essentially saying, that the statistic tells us everything that we want to know about θ that the raw data could tell us.

1.2 Example

Let $X_1, ..., X_n$ be i.i.d. $Bern(p)$ where p is unknown. Let $Y = \sum_{i=1}^n X_i$. Is Y a sufficient statistic?
Solution:
Y is sufficient if the conditional distribution of $X_1, ..., X_n$ given Y is free of p. in other words we need to find $P(X_1 = x_1, ..., X_n = x_n | Y = y)$

$$P(X_1 = x_1, ..., X_n = x_n | Y = y) = \frac{P(X_1 = x_1, ..., X_n = x_n, Y = y)}{P(Y = y)}$$

Note, if $y \neq \sum_{i=1}^n x_i$ then $P(X_1 = x_1, ..., X_n = x_n, Y = y)$ must be 0. If $y = \sum_{i=1}^n x_i$, the event $X_1 = x_1, ..., X_n = x_n, Y = y$ is the event that y of n i.i.d. Bernoullis are 1, and $n - y$ Bernoullis are 0. So,

$$\begin{aligned}
\frac{P(X_1 = x_1, ..., X_n = x_n, Y = y)}{P(Y = y)} &= \begin{cases} \frac{p^y(1-p)^{n-y}}{P(Y=y)} & \text{if} y = \sum_{i=1}^n x_i \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} \frac{p^y(1-p)^{n-y}}{\binom{n}{y} p^y(1-p)^{n-y}} & \text{if} y = \sum_{i=1}^n x_i \\ 0 & \text{else} \end{cases} \\
&= \begin{cases} \frac{1}{\binom{n}{y}} & \text{if} y = \sum_{i=1}^n x_i \\ 0 & \text{else} \end{cases}
\end{aligned}$$

Because the distribution of the data given Y is free of p, we conclude that Y is sufficient for p.

1.3 Factorization Theorem

- While this example appears fairly straight forward, there are many situations where checking to see if a statistic is sufficient is algebraically difficult
- Additionally, the definition only tells us how to confirm that a statistic is sufficient. that is, it cannot help us *find* a sufficient statistic
- Luckily there is a theorem that can help us with both of these issues, but first we will introduce some new notation

Now we can state our the *Factorization Theorem* that will make finding sufficient statistics easier: Theorem:
Let U be a statistic of the random sample $X_1, ..., X_n$ whose distribution depends on the unknown parameter θ. U is a sufficient statistic for θ iff $L(x_1, ..., x_n|\theta)$ can be factored into two nonnegative functions,

$$L(x_1, ..., x_n|\theta) = g(u, \theta) \times h(x_1, ..., x_n)$$

Where $g(u, \theta$ is a function of u and θ, and $h(x_1, ..., x_n)$ is not a function of θ

1.4 Example

Let $X_1, ..., X_n$ be i.i.d. $Bern(p)$ where p is unknown. Let $Y = \sum_{i=1}^n X_i$. Show that Y is a sufficient statistic using the factorization theorem. *Solution:*

$$\begin{aligned} L(x_1, ..., x_n|p) &= f(x_1, ...x_n|p)(= P(X_1 = x_1, ..., X_n = x_n)) \\ &= f(x_1|p)...f(x_n|p)(= P(X_1 = x_1)...P(X_n = x_n)) \\ &= (p^{x_1}(1-p)^{1-x_1})...(p^{x_1}(1-p)^{1-x_1}) \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

So, here $g(u, \theta) = g(y, p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$ and $h(x_1, ..., x_n) = 1$. Thus, by the factorization theorem $Y = \sum_{i=1}^n X_i$ is a sufficient statistic.

1.5 Exercises

1. Let $X_1, ..., X_n$ be i.i.d. $Exp(\delta)$. Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for δ
2. Let $X_1, ..., X_n$ be i.i.d. $Exp(\delta)$. Show that $\bar{X}$ is a sufficient statistic for δ
3. Let $X_1, ..., X_n$ be i.i.d. $Poisson(\lambda)$. Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for λ

1.6 Solutions

1. Let $X_1, ..., X_n$ be i.i.d. $Exp(\delta)$. Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for δ

$$\begin{aligned} L(x_1, ..., x_n|\delta) &= f(x_1, ..., x_n|\delta) \\ &= f(x_1|\delta)...f(x_n|\delta) \\ &= \frac{e^{-\frac{x_1}{\delta}}}{\delta} I(0 < x_1)...\frac{e^{-\frac{x_n}{\delta}}}{\delta} I(0 < x_n) \\ &= \frac{1}{\delta^n} e^{-\frac{\sum_{i=1}^n x_i}{\delta}} I(0 < x_1)...I(0 < x_n) \end{aligned}$$

   Here we have $g(\sum_{i=1}^n x_i, \delta) = \frac{1}{\delta^n} e^{-\frac{\sum_{i=1}^n x_i}{\delta}}$ and $h(x_1, ..., x_n) = I(0 < x_1)...I(0 < x_n)$. Thus, by the factorization theorem we see that $\sum_{i=1}^n X_i$ is sufficient for δ.

2. Let $X_1, ..., X_n$ be i.i.d. $Exp(\delta)$. Show that $\bar{X}$ is a sufficient statistic for δ

$$\begin{aligned} L(x_1, ..., x_n|\delta) &= f(x_1, ..., x_n|\delta) \\ &= f(x_1|\delta)...f(x_n|\delta) \\ &= \frac{e^{-\frac{x_1}{\delta}}}{\delta} I(0 < x_1)...\frac{e^{-\frac{x_n}{\delta}}}{\delta} I(0 < x_n) \\ &= \frac{1}{\delta^n} e^{-\frac{\sum_{i=1}^n x_i}{\delta}} I(0 < x_1)...I(0 < x_n) \\ &= \frac{1}{\delta^n} e^{-\frac{n\bar{x}}{\delta}} I(0 < x_{(1)}) \end{aligned}$$

   Here we have $g(\bar{x}, \delta) = \frac{1}{\delta^n} e^{-\frac{n\bar{x}}{\delta}}$ and $h(x_1, ..., x_n) = I(0 < x_{(1)})$. Thus, by the factorization theorem we see that $\bar{X}$ is sufficient for δ.

3. Let $X_1, ..., X_n$ be i.i.d. $Poisson(\lambda)$. Show that $\sum_{i=1}^n X_i$ is a sufficient statistic for λ

$$\begin{aligned} L(x_1, ..., x_n|\lambda) &= f(x_1, ...x_n|\lambda)(= P(X_1 = x_1, ..., X_n = x_n)) \\ &= f(x_1|\lambda)...f(x_n|\lambda)(= P(X_1 = x_1)...P(X_n = x_n)) \\ &= \frac{\lambda^{x_1}}{x_1!} e^{-\lambda}...\frac{\lambda^{x_n}}{x_n!} e^{-\lambda} \\ &= \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda} \prod_{i=1}^n \frac{1}{x_i!} \end{aligned}$$

   Here we have $g(\sum_{i=1}^n x_i, \lambda) = \lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}$ and $h(x_1, ..., x_n) = \prod_{i=1}^n \frac{1}{x_i!}$. Thus, by the factorization theorem we see that $\sum_{i=1}^n x_i$ is sufficient for λ.

2 Rao-Blackwell Theorem

2.1 The Theorem

Theorem: Let $\hat{\theta}$ be an unbiased estimator for θ such that $V[\hat{\theta}] < \infty$. If U is a sufficient statistic for θ, then $\hat{\theta}* = E[\hat{\theta}|U]$ has the following properties:

1. $\hat{\theta}^*$ is a statistic
2. $E[\hat{\theta}^*] = \theta$
3. $V[\hat{\theta}^*] \leq V[\hat{\theta}]$

Proof.

1. $\hat{\theta}^*$ is a statistic
By definition we have that
$$f(\hat{\theta}|U) = \frac{f(\hat{\theta}, U)}{f(U)}$$
We can think of the joint distribution of $\hat{\theta}$ and U as
$$f(\hat{\theta}, U)) = f(\hat{\theta}, U|X_1 = x_1, ..., X_n = x_x)f(X_1 = x_1, ..., X_n = x_x)$$
This implies that
$$\begin{aligned} f(\hat{\theta}|U) &= \frac{f(\hat{\theta}, U)}{f(U)} \\ &= \frac{f(\hat{\theta}, U|X_1 = x_1, ..., X_n = x_x)f(X_1 = x_1, ..., X_n = x_x)}{f(U)} \end{aligned}$$
We see that $f(\hat{\theta}, U|X_1 = x_1, ..., X_n = x_x)$ is a non-negative function that only depends of the values of $x_1, ..., x_n$, and $\frac{f(X_1=x_1,...,X_n=x_x)}{f(U)}$ cannot depend on θ since U is a sufficient statistic.

2. $E[\hat{\theta}^*] = \theta$
Note $E[\hat{\theta}^*] = E[E[\hat{\theta}|U]]$. Recall $E[X] = E[E[X|Y]]$ for any random variables X and Y. So,
$$\begin{aligned} E[\hat{\theta}^*] &= E[E[\hat{\theta}|U]] \\ &= E[\hat{\theta}] \\ &= \theta \leftarrow \text{Since } \hat{\theta} \text{ is unbiased} \end{aligned}$$

3. $V[\hat{\theta}^*] \leq V[\hat{\theta}]$
Recall, $V[X] = V[E[X|Y]] + E[V[X|Y]]$ for any X, Y. So,
$$\begin{aligned} V[\hat{\theta}] &= V[E[\hat{\theta}|U]] + E[V[\hat{\theta}|U]] \\ &= V[\hat{\theta}^*] + E[V[\hat{\theta}|U]] \\ &\geq V[\hat{\theta}^*] \leftarrow \text{Since } V[\hat{\theta}|U] \geq 0 \end{aligned}$$

$\square$

This means, that if we have an unbiased estimator and a sufficient statistic, we can actually make a *better* estimator that is also unbiased, but has at worst the same variance.