# 6. Sufficient, Complete, and Ancillary Statistics

## The Basic Statistical Model

Consider again the basic statistical model, in which we have a random experiment with an observable random variable $\boldsymbol{X}$ taking values in a set $S$. Once again, the experiment is typically to sample $n$ objects from a population and record one or more measurements for each item. In this case, the outcome variable has the form

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$$

where $X_i$ is the vector of measurements for the $i^{\text{th}}$ item. In general, we suppose that the distribution of $\boldsymbol{X}$ depends on a parameter $\theta$ taking values in a parameter space $\Theta$. The parameter $\theta$ may also be vector-valued. We will use subscripts in probability density functions, expected values, etc. to denote the dependence on $\theta$

## Sufficient Statistics

Let $U = h(\boldsymbol{X})$ be a statistic taking values in a set $T$. Intuitively, $U$ is sufficient for $\theta$ if $U$ contains all of the information about $\theta$ that is available in the entire data variable $\boldsymbol{X}$. Formally, $U$ is **sufficient** for $\theta$ if the conditional distribution of $\boldsymbol{X}$ given $U$ does not depend on $\theta$.

Sufficiency is related to the concept of **data reduction**. Suppose that $\boldsymbol{X}$ takes values in $\mathbb{R}^n$. If we can find a sufficient statistic $\boldsymbol{U}$ that takes values in $\mathbb{R}^j$, then we can reduce the original data vector $\boldsymbol{X}$ (whose dimension $n$ is usually large) to the vector of statistics $\boldsymbol{U}$ (whose dimension $j$ is usually much smaller) with no loss of information about the parameter $\theta$.

The following result gives a condition for sufficiency that is equivalent to this definition.

1. Let $U = h(\boldsymbol{X})$ be a statistic taking values in $T$, and let $f_\theta$ and $g_\theta$ denote the probability density functions of $\boldsymbol{X}$ and $U$ respectively. Show that $U$ is sufficient for $\theta$ if and only if the function

$$\frac{f_\theta(\boldsymbol{x})}{g_\theta(h(\boldsymbol{x}))}, \quad \boldsymbol{x} \in S$$

is independent of $\theta$. *Hint*: The joint distribution of $(\boldsymbol{X}, U)$ is concentrated on the set $\{(\boldsymbol{x}, u) : (\boldsymbol{x} \in S) \text{ and } (u = h(\boldsymbol{x}))\} \subseteq S \times T$.

### The Factorization Theorem

The definition precisely captures the intuitive notion of sufficiency given above, but can be difficult to apply.

We must know in advance a candidate statistic $U$, and then we must be able to compute the conditional distribution of $\boldsymbol{X}$ given $U$. The **factorization theorem** given in the next exercise frequently allows the identification of a sufficient statistic from the form of the probability density function of $\boldsymbol{X}$.

2. Let $f_\theta$ denote the probability density function of $\boldsymbol{X}$ and suppose that $U = h(\boldsymbol{X})$ is a statistic taking values in $T$. Show that $U$ is sufficient for $\theta$ if and only if there exists $G : T \times \Theta \to \mathbb{R}$ and $r : S \to \mathbb{R}$ such that

$$f_\theta(\boldsymbol{x}) = G(h(\boldsymbol{x}), \theta)\, r(\boldsymbol{x}), \quad \boldsymbol{x} \in S, \theta \in \Theta$$

Note that $r$ depends only on the data $\boldsymbol{x}$ but not on the parameter $\theta$.

3. Show that if $U$ and $V$ are equivalent statistics and $U$ is sufficient for $\theta$ then $V$ is sufficient for $\theta$.

### Special Distributions

We will determine sufficient statistics for several parametric families of distributions.

4. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the Bernoulli distribution with success parameter $p \in [0, 1]$. Thus, $X_i = 1$ if trial $i$ is a success, and $X_i = 0$ if trial $i$ is a failure. Let $Y = \sum_{i=1}^n X_i$ denote the number of successes, and recall that $Y$ has the binomial distribution with parameters $n$ and $p$. Show directly from the definition that $Y$ is sufficient for $p$. Specifically, show that the conditional distribution of $\boldsymbol{X}$ given $Y = k$ is the uniform distribution on the set of points

$$\left\{(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n : x_1 + x_2 + \cdots + x_n = k\right\}$$

.

The result in the previous exercise is intuitively appealing: in a sequence of Bernoulli trials, all of the information about the probability of success $p$ is contained in the number of successes $Y$. The particular *order* of the successes and failures provides no additional information. Of course, the sufficiency of $Y$ follows more easily from the factorization theorem, but the conditional distribution provides additional insight.

5. Suppose that the distribution of $\boldsymbol{X}$ is a $k$-parameter exponential familiy with the natural statistic $U = h(\boldsymbol{X})$. Show that $U$ is sufficient for $\theta$. Because of this result, $U$ is referred to as the **natural sufficient statistic** for the exponential family.

6. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in (0, \infty)$.

   a. Show that $(Y, V)$ is sufficient for $\left(\mu, \sigma^2\right)$ where $Y = \sum_{i=1}^n X_i$ and $V = \sum_{i=1}^n X_i^2$.
   b. Show that $\left(M, S^2\right)$ is sufficient for $\left(\mu, \sigma^2\right)$ where $M$ is the sample mean of $\boldsymbol{X}$ and $S^2$ is the sample variance of $\boldsymbol{X}$. *Hint*: Use part (a) and equivalence.

7. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the Poisson distribution with

mean $a \in (0, \infty)$. Show that $Y = \sum_{i=1}^n X_i$ is sufficient for $a$.

8. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the gamma distribution with shape parameter $k \in (0, \infty)$ and scale parameter $b \in (0, \infty)$

   a. Show that $(Y, V)$ is sufficient for $(k, b)$ where $Y = \sum_{i=1}^n X_i$ and $V = \prod_{i=1}^n X_i$.
   b. Show that $(M, U)$ is sufficient for $(k, b)$ where $M$ is the sample (arithmetic) mean of $\boldsymbol{X}$ and $U$ is the sample geometric mean of $\boldsymbol{X}$. *Hint*: Use part (a) and equivalence.

9. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the beta distribution with left parameter $a \in (0, \infty)$ and right parameter $b \in (0, \infty)$. Show that $(U, V)$ is sufficient for $(a, b)$ where $U = \prod_{i=1}^n X_i$ and $V = \prod_{i=1}^n (1 - X_i)$.

10. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the Pareto distribution with shape parameter $a \in (0, \infty)$. Show that $U = \prod_{i=1}^n X_i$ is sufficient for $a$.

11. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the uniform distribution on the interval $[0, a]$ where $a \in (0, \infty)$ is the unknown parameter. Show that $X_{n,n} = \max\{X_1, X_2, \ldots, X_n\}$ (the $n^{\text{th}}$ order statistic) is sufficient for $a$.

**Minimal Sufficient Statistics**

The entire data variable $\boldsymbol{X}$ is trivially sufficient for $\theta$. However, as noted above, there usually exists a statistic $U$ that is sufficient for $\theta$ and has smaller dimension, so that we can achieve real data reduction. Naturally, we would like to find the statistic $U$ that has the smallest dimension possible. In many cases, this smallest dimension $j$ will be the same as the dimension $k$ of the parameter vector $\theta$. However, as we will see, this is not necessarily the case; $j$ can be smaller or larger than $k$.

Formally, suppose that a statistic $U$ is sufficient for $\theta$. Then $U$ is **minimally sufficient** if $U$ is a function of any other statistic $V$ that is sufficient for $\theta$. Once again, the definition precisely captures the notion of minimal sufficiency, but is hard to apply. The following exercise gives an equivalent condition.

12. Let $f_\theta$ denote the probability density function of $\boldsymbol{X}$ corresponding to the parameter value $\theta$ and suppose that $U = h(\boldsymbol{X})$ is a statistic taking values in $T$. Show that $U$ is minimally sufficient for $\theta$ if the following condition holds: for $\boldsymbol{x} \in S$ and $\boldsymbol{y} \in S$

$$\frac{f_\theta(\boldsymbol{x})}{f_\theta(\boldsymbol{y})} \text{ is independent of } \theta \text{ if and only if } h(\boldsymbol{x}) = h(\boldsymbol{y})$$

   *Hint*: If $V = g(\boldsymbol{X})$ is another sufficient statistic, use the factorization theorem and the condition above to show that $g(\boldsymbol{x}) = g(\boldsymbol{y})$ implies $h(\boldsymbol{x}) = h(\boldsymbol{y})$ for $\boldsymbol{x} \in S$ and $\boldsymbol{y} \in S$. Then conclude that $U$ is a function of $V$.

13. Show that if $U$ and $V$ are equivalent statistics and $U$ is minimally sufficient for $\theta$ then $V$ is minimally sufficient for $\theta$.

14. Suppose that the distribution of $\boldsymbol{X}$ is a $k$-parameter exponential family with natural sufficient statistic $U = h(\boldsymbol{X})$. Show that $U$ is a minimally sufficient for $\theta$.

15. Show that the sufficient statistics given above for the Bernoulli, Poisson, normal, gamma, and beta families are minimally sufficient for the given parameters.

16. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from the uniform distribution on the interval $[a, a + 1]$ where $a \in (0, \infty)$ is the unknown parameter. Show that $\left(X_{n,1}, X_{n,n}\right)$, the vector consisting of the first and last order statistics, is minimally sufficient for $a$. Note that we have a single parameter, but the minimally sufficient statistic is a vector of dimension 2.

### Properties of Sufficient Statistics

Sufficiency is related to several of the methods of constructing estimators that we have studied.

17. Suppose that $U$ is sufficient for $\theta$ and that there exists a maximum likelihood estimator of $\theta$. Show that there exists a MLE $V$ that is a function of $U$. *Hint*: Use the factorization theorem.

In particular, suppose that $V$ is the unique MLE of $\theta$ and that $V$ is sufficient for $\theta$. If $U$ is sufficient for $\theta$ then $V$ is a function of $U$ by the previous exercise. Hence it follows that $V$ is minimally sufficient for $\theta$.

18. Suppose that the statistic $U$ is sufficient for the parameter $\theta$ and that $V$ is a Bayes' estimator of $\theta$. Show that $V$ is a function of $U$. *Hint*: Use the factorization theorem.

The following exercise gives the **Rao-Blackwell theorem**, named for CR Rao and David Blackwell. The theorem shows how a sufficient statistic can be used to improve an unbiased estimator.

19. Suppose that $U$ is sufficient for $\theta$ and that $V$ is an unbiased estimator of a real parameter $\lambda = \lambda(\theta)$. Use sufficiency and properties of conditional expectation and conditional variance to show that

    a. $\mathbb{E}_\theta(V|U)$ is a valid statistic. That is, it does not depend on $\theta$, in spite of the formal dependence on $\theta$ in the expected value.
    b. $\mathbb{E}(V|U)$ is a function of $U$.
    c. $\mathbb{E}(V|U)$ is an unbiased estimator of $\lambda$.
    d. $\text{var}_\theta(\mathbb{E}(V|U)) \leq \text{var}_\theta(V)$ for any $\theta \in \Theta$ so $\mathbb{E}(V|U)$ is uniformly better than $V$.

## Complete Statistics

Suppose that $U = h(\boldsymbol{X})$ is a statistic taking values in a set $T$. Then $U$ is a **complete** statistic for $\theta$ if for any real-valued function $g$ on $T$

$$\mathbb{E}_\theta(g(U)) = 0 \text{ for all } \theta \in \Theta \Rightarrow \mathbb{P}_\theta(g(U) = 0) = 0 \text{ for all } \theta \in \Theta$$

To understand this rather strange looking condition, suppose that $g(U)$ is a statistic constructued from $U$ that is being used as an estimator of 0 (thought of as a function of $\theta$). The completeness condition means that the only such unbiased estimator is the statistic that is 0 with probability 1.

20. Show that if $U$ and $V$ are equivalent statistics and $U$ is complete for $\theta$ then $V$ is complete for $\theta$.

### Special Distributions

21. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the Bernoulli distribution with success parameter $p \in (0, 1)$. Show that the number or successes $Y = \sum_{i=1}^n X_i$ is complete for $p$. *Hint*: Note that $\mathbb{E}_p\left(g(Y)\right)$ can be written as a polynomial in $t = \frac{p}{1-p}$ If this polynomial is 0 for all $t$ in an open interval, then the coefficients must be 0.

22. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the Poisson distribution with parameter $a \in (0, \infty)$. Show that the sum of the sample values $Y = \sum_{i=1}^n X_i$ is complete for $a$. *Hint*: Note that $\mathbb{E}_a(g(Y))$ can be written as a power series in $a$. If this series is 0 for all $a$ in an open interval, then the coefficients must be 0.

23. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the exponential distribution with scale parameter $b \in (0, \infty)$. Show that the sum of the sample values $Y = \sum_{i=1}^n X_i$ is complete for $b$. *Hint*: Note that $\mathbb{E}_b(g(Y))$ is the Laplace transform of a certain function. If this transform is 0 for all $b$ in an open interval, then the function must be 0.

The results in the previous exercises generalize to exponential families, although the proof is complicated. Specifically, if the distribution of $\boldsymbol{X}$ is a $k$-parameter exponential family with the natural sufficient statistic $U = h(\boldsymbol{X})$ then $U$ is complete for $\theta$ (as well as minimally sufficient for $\theta$). This applies to random samples from the Bernoulli, Poisson, normal, gamma, and beta distributions discussed above.

The notion of completeness depends very much on the parameter space.

24. Suppose that $\boldsymbol{X} = (X_1, X_2, X_3)$ is a random sample of size 3 from the Bernoulli distribution with success parameter $p \in \left\{\frac{1}{3}, \frac{1}{2}\right\}$. Show that $Y = X_1 + X_2 + X_3$ is not complete for $p$.

### The Lehmann-Scheffé Theorem

The next exercise shows the importance of complete sufficient statistics; it is known as the **Lehmann-Scheffé theorem**, named for Erich Lehmann and Henry Scheffé.

25. Suppose that $U$ is sufficient and complete for $\theta$ and that $T = r(U)$ is an unbiased estimator of a real parameter $\lambda = \lambda(\theta)$. Show that $T$ is a uniformly minimum variance unbiased estimator of $\lambda$. The proof is

based on the following steps:

a. Suppose that $V$ is an unbiased estimator of $\lambda$. By the Rao-Blackwell theorem, $\mathbb{E}(V|U)$ is also an unbiased estimator of $\lambda$ and is uniformly better than $V$.
b. Since $\mathbb{E}(V|U)$ is a function of $U$, use completeness to conclude that $T = \mathbb{E}(V|U)$ with probability 1.

26. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the Bernoulli distribution with success parameter $p \in [0, 1]$. As usual, let $Y = \sum_{i=1}^n X_i$ denote the number of successes. Show that an UMVUE for $p(1 - p)$, the variance of the sample distribution, is

$$\frac{Y}{n-1}\left(1 - \frac{Y}{n}\right)$$

27. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the Poisson distribution with parameter $\mu$. Let $Y = \sum_{i=1}^n X_i$. Show that an UMVUE for $\mathbb{P}(X = 0) = e^{-\mu}$ is

$$\left(\frac{n-1}{n}\right)^Y$$

*Hint*: Use the probability generating function of $Y$.

## Ancillary Statistics

Suppose that $V = r(\boldsymbol{X})$ is a statistic taking values in a set $T$. If the distribution of $V$ does not depend on $\theta$, then $V$ is called an **ancillary** statistic for $\theta$. Thus, the notion of an ancillary statistic is complementary to the notion of a sufficient statistics (which contains all information about the parameter that is contained in the sample). Thus, the result in the following exercise, known as **Basu's Theorem** and named for Debabrata Basu, makes this point more precisely.

28. Suppose that $U$ is complete and sufficient for a parameter $\theta$ and that $V$ is an ancillary statistic. Show that $U$ and $V$ are independent. The following steps sketch the proof:

a. Let $g$ denote the probability density function of $V$ and let $v \mapsto g(v|U)$ denote the conditional probability density function of $V$ given $U$.
b. Use properties of conditional expected value to show that $\mathbb{E}(g(v|U)) = g(v)$ for $v \in T$.
c. Use completeness to conclude that $g(v|U) = g(v)$ with probability 1.

29. Show that if $U$ and $V$ are equivalent statistics and $U$ is ancillary for $\theta$ then $V$ is ancillary for $\theta$.

30. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample from a scale family with scale parameter $b \in (0, \infty)$. Show that $V$ is an ancillary statistic for $b$ if $V$ is a function of

$$\left(\frac{X_1}{X_n}, \frac{X_2}{X_n}, \ldots, \frac{X_{n-1}}{X_n}\right)$$

31. Suppose that $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is a random sample of size $n$ from the gamma distribution with shape parameter $k \in (0, \infty)$ and scale parameter $b \in (0, \infty)$. Let $M$ denote the sample arithmetic mean of $\boldsymbol{X}$ and let $U$ denote the sample geometric mean of $\boldsymbol{X}$. Show that $\frac{M}{U}$ is ancillary for $b$, and thus conclude that $M$ and $\frac{M}{U}$ are independent. *Hint*: Use the previous exercise.

Virtual Laboratories > 7. Point Estimation > 1 2 3 4 5 6
Contents | Applets | Data Sets | Biographies | External Resources | Keywords | Feedback | ©