distribution on 20 degrees of freedom lies between 0.025 and 0.094, so that the interval for ϕ is from 664×0.025 to 664×0.094, that is, the interval (17, 62). It follows that the two methods do not give notably different answers.

2.9 The role of sufficiency

Definition of sufficiency

When we considered the normal variance with known mean, we found that the posterior distribution depended on the data only through the single number S. It often turns out that the data can be reduced in a similar way to one or two numbers, and as long as we know them we can forget the rest of the data. It is this notion that underlies the formal definition of sufficiency.

Suppose observations $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ are made with a view to gaining knowledge about a parameter θ, and that

$$t = t(\boldsymbol{x})$$

is a function of the observations. We call such a function a *statistic*. We often suppose that t is real-valued, but it is sometimes vector-valued. Using the formulae in Section 1.4 on 'Several random variables' and the fact that once we know $\boldsymbol{x}$ we automatically know the value of t, we see that for any statistic t,

$$p(\boldsymbol{x} \mid \theta) = p(\boldsymbol{x}, t \mid \theta) = p(t \mid \theta)\, p(\boldsymbol{x} \mid t, \theta).$$

However, it sometimes happens that

$$p(\boldsymbol{x} \mid t, \theta)$$

does not depend on θ, so that

$$p(\boldsymbol{x} \mid \theta) = p(t \mid \theta)\, p(\boldsymbol{x} \mid t).$$

If this happens, we say that t is a *sufficient statistic* for θ given $\boldsymbol{x}$, often abbreviated by saying that t is sufficient for θ. It is occasionally useful to have a further definition as follows: a statistic $u = u(\boldsymbol{x})$ whose density $p(u \mid \theta) = p(u)$ does not depend on θ is said to be *ancillary* for θ.

Neyman's factorization theorem

The following theorem is frequently used in finding sufficient statistics:

Theorem. A statistic t is sufficient for θ given $\boldsymbol{x}$ if and only if there are functions f and g such that

$$p(\boldsymbol{x} \mid \theta) = f(t, \theta)\, g(\boldsymbol{x})$$

where $t = t(\boldsymbol{x})$.

Proof. If t is sufficient for θ given $\boldsymbol{x}$ we may take

$$f(t, \theta) = p(t \mid \theta) \quad \text{and} \quad g(\boldsymbol{x}) = p(\boldsymbol{x} \mid t).$$

Conversely, if the condition holds, integrate or sum both sides of the equation over all values of $\boldsymbol{x}$ such that $t(\boldsymbol{x}) = t$. Then the left-hand side will, from the basic properties of a density, be the density of t at that particular value, and so we get

$$p(t \mid \theta) = f(t, \theta)\, G(t)$$

where $G(t)$ is obtained by summing or integrating $g(\boldsymbol{x})$ over all these values of $\boldsymbol{x}$. We then have

$$f(t, \theta) = p(t \mid \theta)/G(t).$$

Considering now any one value of $\boldsymbol{x}$ such that $t(\boldsymbol{x}) = t$ and substituting in the equation in the statement of the theorem, we obtain

$$p(\boldsymbol{x} \mid \theta) = p(t \mid \theta)\, g(\boldsymbol{x})/G(t).$$

Since, whether t is sufficient or not,

$$p(\boldsymbol{x} \mid t, \theta) = p(\boldsymbol{x}, t \mid \theta)/p(t \mid \theta) = p(\boldsymbol{x} \mid \theta)/p(t \mid \theta)$$

we see that

$$p(\boldsymbol{x} \mid t, \theta) = g(\boldsymbol{x})/G(t).$$

Since the right-hand side does not depend on θ, it follows that t is indeed sufficient, and the theorem is proved. ■

Sufficiency principle

Theorem. A statistic t is sufficient for θ given $\boldsymbol{x}$ if and only if

$$l(\theta \mid \boldsymbol{x}) \propto l(\theta \mid t)$$

whenever $t = t(\boldsymbol{x})$ (where the constant of proportionality does not, of course, depend on θ).

Proof. If t is sufficient for θ given $\boldsymbol{x}$ then

$$l(\theta \mid \boldsymbol{x}) \propto p(\boldsymbol{x} \mid \theta) = p(t \mid \theta) p(\boldsymbol{x} \mid t) \propto p(t \mid \theta) \propto l(\theta \mid t).$$

Conversely, if the condition holds then

$$p(\boldsymbol{x} \mid \theta) \propto l(\theta \mid \boldsymbol{x}) \propto l(\theta \mid t) \propto p(t \mid \theta)$$

so that, for some function $g(\boldsymbol{x})$,

$$p(\boldsymbol{x} \mid \theta) = p(t \mid \theta) g(\boldsymbol{x}).$$

The theorem now follows from the factorization theorem. ■

Corollary 1. For any prior distribution, the posterior distribution of θ given $\boldsymbol{x}$ is the same as the posterior distribution of θ given a sufficient statistic t.

Proof. From Bayes' theorem $p(\theta \mid \boldsymbol{x})$ is proportional to $p(\theta \mid t)$; they must then be equal as they both integrate or sum to unity. ■

Corollary 2. If a statistic $t = t(\boldsymbol{x})$ is such that $l(\theta \mid \boldsymbol{x}) \propto l(\theta \mid \boldsymbol{x}')$ whenever $t(\boldsymbol{x}) = t(\boldsymbol{x}')$, then it is sufficient for θ given $\boldsymbol{x}$.

Proof. By summing or integrating over all $\boldsymbol{x}$ such that $t(\boldsymbol{x}) = t$, it follows that

$$l(\theta \mid t) \propto p(t \mid \theta) = \sum p(x' \mid \theta) \propto \sum l(\theta \mid \boldsymbol{x}') \propto l(\theta \mid \boldsymbol{x}),$$

the summations being over all $\boldsymbol{x}'$ such that $t(\boldsymbol{x}') = t(\boldsymbol{x})$. The result now follows from the theorem. ■

Examples

Normal variance. In the case where the x_i are normal of known mean μ and unknown variance ϕ, we noted that

$$p(\boldsymbol{x} \mid \phi) \propto \phi^{-n/2} \exp(-\tfrac{1}{2}S/\phi)$$

where $S = \sum (x_i - \mu)^2$. It follows from the factorization theorem that S is sufficient for μ given $\boldsymbol{x}$. Moreover, we can verify the sufficiency principle as follows. If we had simply been given the value of S without being told the values of $x_1, x_2, \ldots, x_n$ separately we could have noted that, for each i,

$$(x_i - \mu)/\sqrt{\phi} \sim \mathrm{N}(0,1)$$

so that S/ϕ is a sum of squares of n independent N(0, 1) variables. Now a χ_n^2 distribution is often *defined* as being the distribution of the sum of squares of n random variables with an N(0, 1) distribution, and the density of χ_n^2 can be deduced from this. It follows that $S/\phi \sim \chi_n^2$ and hence if $y = S/\phi$ then

$$p(y) \propto y^{n/2-1} \exp(-\tfrac{1}{2}y).$$

Using the change-of-variable rule, it is then easily seen that

$$p(S \mid \phi) \propto S^{n/2-1} \phi^{-n/2} \exp(-\tfrac{1}{2}S/\phi).$$

We can thus verify the sufficiency principle in this particular case because

$$l(\phi \mid \boldsymbol{x}) \propto p(\boldsymbol{x} \mid \phi) \propto S^{-n/2+1} p(S \mid \phi) \propto p(S \mid \phi) \propto l(\phi \mid S).$$

Poisson case. Recall that the integer-valued random variable x is said to have a Poisson distribution of mean λ (denoted $x \sim \mathrm{P}(\lambda)$) if

$$p(x \mid \lambda) = (\lambda^x / x!) \exp(-\lambda) \qquad (x = 0, 1, 2, \ldots).$$

We shall consider the Poisson distribution in more detail later in the book. For the moment, all that matters is that it often serves as a model for the number of occurrences of a rare event, for example for the number of times the King's Arms on the riverbank at York is flooded in a year. Then if $x_1, x_2, \ldots, x_n$ have independent Poisson distributions with the same mean (so could, for example, represent the numbers of floods in several successive years), it is easily seen that

$$p(\boldsymbol{x} \mid \lambda) \propto \lambda^T \exp(-n\lambda)$$

where

$$T = \sum x_i.$$

It follows from the factorization theorem that T is sufficient for λ given $\boldsymbol{x}$. Moreover we can verify the sufficiency principle as follows. If we had simply been given the value of T without being given the values of the x_i separately, we could have noted that a sum of independent Poisson distributions has a Poisson distribution with mean the sum of the means (see question 7 on Chapter 1), so that

$$T \sim \mathrm{P}(n\lambda)$$

and hence

$$l(\lambda \mid T) \propto p(T \mid \lambda) = \{(n\lambda)^T / T!\} \exp(-n\lambda) \propto p(\boldsymbol{x} \mid \lambda) \propto l(\lambda \mid \boldsymbol{x})$$

in accordance with the sufficiency principle.

Order statistics and minimal sufficient statistics

It may be noted that it is easy to see that whenever $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ consists of independently and identically distributed observations whose distribution depends on a parameter θ, then the order statistic

$$\boldsymbol{x}_{(O)} = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$$

which consists of the values of the x_i arranged in increasing order, so that

$$x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$$

is sufficient for θ given $\boldsymbol{x}$.

This helps to underline the fact that there is, in general, no such thing as a *unique* sufficient statistic. Indeed, if t is sufficient for θ given $\boldsymbol{x}$, then so is (t, u) for *any* statistic $u(\boldsymbol{x})$. If t is a function of all other sufficient statistics that can be constructed, so that no further reduction is possible, then t is said to be *minimal sufficient*. Even a minimal sufficient statistic is not unique, since any one-to-one function of such a statistic is itself minimal sufficient.

It is not obvious that a minimal sufficient statistic always exists, but in fact it does. Although the result is more important than the proof, we shall now prove this. We define a statistic $u(\boldsymbol{x})$ which is a *set*, rather than a real number or a vector, by

$$u(\boldsymbol{x}) = \{\boldsymbol{x}';\ l(\theta \mid \boldsymbol{x}') \propto l(\theta \mid \boldsymbol{x})\}.$$

Then it follows from Corollary 2 to the sufficiency principle that u is sufficient. Further, if $v = v(\boldsymbol{x})$ is any other sufficient statistic, then by the same principle whenever $v(\boldsymbol{x}') = v(\boldsymbol{x})$ we have

$$l(\theta \mid \boldsymbol{x}') \propto l(\theta \mid v) \propto l(\theta \mid \boldsymbol{x})$$

and hence $u(\boldsymbol{x}') = u(\boldsymbol{x})$, so that u is a function of v. It follows that u is minimal sufficient. We can now conclude that the condition that

$$t(\boldsymbol{x}') = t(\boldsymbol{x})$$

if and only if

$$l(\theta \mid \boldsymbol{x}') \propto l(\theta \mid \boldsymbol{x})$$

is equivalent to the condition that t is minimal sufficient.