Sufficiency and Unbiased Estimation

BS2 Statistical Inference, Lecture 3
Michaelmas Term 2004

Steffen Lauritzen, University of Oxford; October 19, 2004

Sufficiency

A statistic $T = t(X)$ is said to be *sufficient* for the parameter θ if $P_\theta\{X = x \,|\, T = t\}$ does not depend on θ.

Intuitively, a sufficient statistic is capturing all information in data x which is relevant for θ.

Clearly, it can be of interest to find a statistic which does so as compactly as possible. Such a statistic is called minimal sufficient.

Formally a statistic is said to be *minimal sufficient* if it is sufficient and it can be calculated from any other sufficient statistic.

Neyman's factorization theorem

Sufficient statistics are most easily recognized through the following fundamental result:

A statistic $T = t(X)$ is sufficient for θ if and only if the family of densities can be factorized as

$$f(x;\theta) = h(x)k\{t(x);\theta\}, \quad x \in \mathcal{X}, \theta \in \Theta, \tag{1}$$

i.e. into a function which does not depend on θ and one which only depends on x through $t(x)$.

This is true in wide generality (essentially when densities exist), but we only prove it here in the case where $\mathcal{X}$ is discrete.

Proof of the factorization theorem

Assume T sufficient and let $h(x) = P_\theta\{X = x \,|\, T = t(x)\}$ be independent of θ. Let $k(t;\theta) = P_\theta(T = t)$. Then

$$f(x;\theta) = P_\theta\{X = x \,|\, T = t(x)\} P_\theta\{T = t(x)\} = h(x)k\{t(x);\theta\}.$$

Conversely assume (1). Then

$$\begin{aligned} P_\theta\{X = x \,|\, T = t\} &= \frac{h(x)k\{t(x);\theta\}}{\sum_{y:t(y)=t} h(y)k\{t(y);\theta\}} 1_{\{t(x)=t\}}(x) \\ &= \frac{h(x)k\{t;\theta\}}{k\{t;\theta\}\sum_{y:t(y)=t} h(y)} 1_{\{t(x)=t\}}(x) \\ &= \frac{h(x)}{\sum_{y:t(y)=t} h(y)} 1_{\{t(x)=t\}}(x), \end{aligned}$$

which is independent of θ.

Example

Let $X = (X_1, \ldots, X_n)$ be independent and Poisson distributed with mean θ so that

$$f(x;\theta) = \prod_i \frac{\theta^{x_i}}{x_i!} e^{-\theta} = \frac{\theta^{\sum_i x_i}}{\prod_i x_i!} e^{-n\theta}.$$

This factorizes as in (1) with $t(x) = \sum_i x_i$ if we let $h(x) = 1/\prod_i x_i!$ and $k(t;\theta) = \theta^t e^{-n\theta}$.

Thus $t(x) = \sum_i x_i$ is sufficient (and indeed minimal sufficient).

Minimal sufficiency of the likelihood function

As previously mentioned, likelihood is probably the most fundamental concept in statistic.

In fact, the likelihood function itself is always minimal sufficient, thus implicitly the universal representation of available information in the data about a parameter θ. To establish this we first show

Let $U = u(X)$ be any statistic from which the likelihood function $L_x(\theta)$ can be reconstructed up to a factor which is constant in θ. Then U is sufficient. In particular L_x is itself sufficient.

To see this, we argue as follows: If L_x can be reconstructed from $u = u(x)$ up to a constant factor, we have

$$L_x(\theta) = c(x)d\{u(x);\theta\}.$$

But then we have for the density that

$$f(x;\theta) = e(x)L_x(\theta) = e(x)c(x)d\{u(x);\theta\}.$$

Using the factorization theorem with $h(x) = e(x)c(x)$ and $k = d$ shows that U is sufficient.

The likelihood function is minimal sufficient.

For if $T = t(X)$ is sufficient, the factorization theorem yields

$$L_x(\theta) = h(x)k\{t(x);\theta)\}$$

so the likelihood function can be calculated (up to a constant factor) from the value of t and L_x must therefore be minimal sufficient.

The Rao–Blackwell theorem

Sufficiency is important for obtaining minimum variance for unbiased estimators:

If $U = u(X)$ *is an unbiased estimator of a function* $g(\theta)$ *and* $T = t(X)$ *is sufficient for* θ *then* $U^* = u^*(X)$ *where* $u^*(x) = \mathbf{E}_\theta\{U \mid T = t(x)\}$ *is also unbiased for* $g(\theta)$ *and*

$$\mathbf{V}(U^*) \leq \mathbf{V}(U).$$

The process of modifying U to the improved estimator U^* by taking conditional expectation w.r.t. a sufficient statistic T, is known as *Rao–Blackwellization* and can be important in many contexts.

Proof of the Rao–Blackwell theorem

Firstly, $u^*(x)$ is indeed a function of x since the conditional expectation $\mathbf{E}_\theta\{U \mid T = t(x)\}$ is independent of θ because T is sufficient.

Secondly, U^* is unbiased since iterating expectations yields

$$\mathbf{E}(U^*) = \mathbf{E}\{\mathbf{E}(U \mid T)\} = \mathbf{E}(U) = g(\theta).$$

Finally $[\mathbf{E}\{U - g(\theta) \mid T\}]^2 \leq \mathbf{E}[\{U - g(\theta)\}^2 \mid T]$ so

$$\begin{aligned}
\mathbf{V}(U^*) &= \mathbf{E}[\{\mathbf{E}(U \mid T) - g(\theta)\}^2] \\
&= \mathbf{E}[\{\mathbf{E}(U - g(\theta) \mid T)\}^2] \\
&\leq \mathbf{E}(\mathbf{E}[\{U - g(\theta)\}^2 \mid T]) = \mathbf{V}(U).
\end{aligned}$$

Example

Consider the Poisson case above. Here X_1 is clearly an unbiased estimator of θ, but far from the best possible.

Now Rao-Blackwellize this to

$$\hat{\theta} = \mathbf{E}\{X_1 \,|\, T = X_1 + \cdots + X_n\} = T/n = \bar{X},$$

where we have used that the conditional distribution of $(X_1, \cdots, X_n)$ given $T = t$ is multinomial with parameters $p = (1/n, \cdots, 1/n)$ and t.

Uniqueness of the MVUE

A MVUE is essentially unique.

For if T and T' are both MVUE for $g(\theta)$, consider

$$V\{T+\lambda(T'-T)\} = \mathbf{V}(T)+\lambda^2\mathbf{V}(T'-T)+2\lambda\,\mathrm{Cov}(T,T'-T). \tag{2}$$

If $\mathbf{V}(T'-T)$= 0 we clearly have $T'=T$ so assume that $\gamma = \mathbf{V}(T'-T) > 0$, let $\rho = -\,\mathrm{Cov}(T, T'-T)$ and $\lambda = \rho/\gamma$. Then

$$\mathbf{V}\{T+\lambda(T'-T)\} = \mathbf{V}(T)+\rho^2/\gamma-2\rho^2/\gamma = \mathbf{V}(T)-\rho^2/\gamma.$$

Since T was MVUE, we get $\rho = \mathrm{Cov}(T, T'-T) = 0$. Inserting this into (2) and letting $\lambda = 1$ yields $\mathbf{V}(T') = \mathbf{V}(T) + \mathbf{V}(T'-T)$. As T' is MVUE we have $\mathbf{V}(T'-T) = 0$ and thus $T = T'$.

Minimal sufficiency and MVUE

The Rao–Blackwell theorem and the essential uniqueness of the MVUE implies that

A MVUE must essentially be a function of any minimal sufficient statistic.

To see this, assume U is MVUE and let T be minimal sufficient. Then Rao–Blackwellize U to $U^* = \mathbf{E}\{U \mid T\}$.

We then have $\mathbf{V}(U^*) \leq \mathbf{V}(U)$, but as U was already MVUE, U^* is also MVUE. The essential uniqueness of a MVUE implies $U = U^*$.

Completeness

A statistic T is said to be *complete* w.r.t. θ if for all functions h

$$\mathbf{E}_\theta\{h(T)\} = 0 \text{ for all } \theta \implies h(t) = 0 \text{ a.s.}$$

It is *boundedly complete* if the same holds when only bounded functions h are considered.

The wording is not quite accurate. It would be more precise to say the *family* of densities of T $\mathcal{F}_T = \{f_T(t;\theta), \theta \in \Theta\}$ is complete, but the shorter usage has become common in statistical literature.

Completeness and sufficiency

Any estimator of the form $U = h(T)$ of a complete and sufficient statistic T is the unique unbiased estimator based on T of its expectation.

For if h_1 and h_2 were two such estimators, we would have

$$\mathbf{E}_\theta\{h_1(T) - h_2(T)\} = 0 \text{ for all } \theta,$$

and hence $h_1 = h_2$.

In fact, *if T is complete and sufficient, it is also minimal sufficient.* Note this makes the assumption of minimality in Lemma 2.7 of Garthwaite et al. (2002) redundant.

Hence, *if T is complete and sufficient, $U = h(T)$ is the MVUE of its expectation.*