Sufficiency

MIT 18.443

Dr. Kempthorne

Spring 2015

Outline

Sufficient Statistics

Definition: Sufficiency

- $X_1, X_2, \ldots, X_n$ iid with distribution P_θ with density/pmf $f(x \mid \theta)$.
- $T(X_1, \ldots, X_n)$ is a *statistic* (a well-defined function of the data – computed wihout knowledge of θ).
- The statistic $T(X_1, \ldots, X_n)$ is sufficient for θ if the conditional distribution of $X_1, \ldots, X_n$ given $T = t$ does not depend on θ for any value of t.

Power of Sufficient Statistics

- If $T(X_1, \ldots, X_n)$ is sufficient for θ, then statistical inference about θ can focus exclusively on T and its conditional distribution given θ : $\quad T \sim f_T(t \mid \theta)$.
- Data reduction of the original sample $(X_1, \ldots, X_n)$ to $T(X_1, \ldots, X_n)$ maintains all the information in the sample about θ.

Outline

Sufficiency: Example

Example 8.8.1A Bernoulli Trials Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the outcome of n i.i.d *Bernoulli*(θ) random variables

- The pmf function of X is:

$$\begin{aligned} p(\mathbf{X} \mid \theta) &= P(X_1 = x_1 \mid \theta) \times \cdots \times P(X_n = x_n \mid \theta) \\ &= \theta^{x_1}(1-\theta)^{1-x_1} \times \theta^{x_2}(1-\theta)^{1-x_2} \times \cdots \theta^{x_n}(1-\theta)^{1-x_n} \\ &= \theta^{\sum x_i}(1-\theta)^{(n-\sum x_i)} \end{aligned}$$

- Consider $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ whose distribution has pmf:

$$p(t \mid \theta) = \binom{n}{t} \theta^t (1-\theta)^{n-t}, 0 \leq t \leq n.$$

- The distribution of $\mathbf{X}$ given $T(\mathbf{X}) = t$ is uniform over the n-tuples X: $T(\mathbf{X}) = t$.

Thus, $T(\mathbf{X})$ is sufficient for θ.

Sufficient Statistic for Bernoulli Trials

Consequences of Sufficiency

- The distribution of $\mathbf{X}$ given θ (not conditioned on T) can be simulated by generating $T \sim Binomial(n, \theta)$, and then choosing $\mathbf{X}$ randomly according to the uniform distribution over all tuples $\{\mathbf{x} = (x_1, \ldots, x_n) : T(\mathbf{x}) = t\}$
  Given $T(\mathbf{X}) = t$, the choice of tuple X does not require knowledge of θ.
- After knowing $T(\mathbf{X}) = t$, the additional information in $\mathbf{X}$ is the sequence/order information which does not depend on θ.
- To make statistical inferences concerning θ, we should only need the information of $T(\mathbf{X}) = t$, since the value of X given t reflects only the order information in X which is independent of θ.

Outline

1. Sufficiency
   - Definition
   - Example
   - Theorems

Sufficient Statistics

Proposition

- Suppose $\tilde{\theta} = q(X_1, \ldots, X_n)$ is any estimate of θ specified by a function $q(X_1, \ldots, X_n)$
(which is well-specified without knowledge of θ).
- There always exists an estimate $\tilde{\theta}^*$ which depends only on the sufficient statistic T which is as good as $\tilde{\theta}$.
$$\tilde{\theta}^*(X_1, \ldots, X_N) = q^*(t),$$
where $t = T(X_1, \ldots, X_n)$ and $q^*()$ is well-specified without knowledge of θ.

Proof: Application of statistical decision theory covered in 18.466

Sufficient Statistics: Theorems (Rice Section 8.8)

Factorization Theorem

A necessary and sufficient condition for $T(X_1, \ldots, X_n)$ to be sufficient for a parameter θ is that the joint probability density/pmf function factors in the form

$$f(x_1 \ldots, x_n \mid \theta) = g[T(x_1, \ldots, x_n), \theta]h(x_1, \ldots, x_n).$$

Corollary A

If T is sufficient for θ, then the maximum likelihood estimate is a function of T.

Rao-Blackwell Theorem

- Let $\hat{\theta}$ be an estimator of θ with $E[\hat{\theta}^2] < \infty$ for all θ.
- Suppose that T is sufficient for θ
- Define $\tilde{\theta} = E[\hat{\theta} \mid T]$.

Then for all θ,

$$E[(\tilde{\theta} - \theta)^2] \leq E[(\hat{\theta} - \theta)^2].$$

The inequality is strict unless $\tilde{\theta} \equiv \hat{\theta}$.

18.443 Statistics for Applications
Spring 2015