

06: SUFFICIENCY

Larsen & Marx 5.6

Prof Amanda Luby

1 Sufficient Estimators

So far, we've seen a few desirable properties for estimators: that they should be unbiased, that they should have minimum variance, and that they should converge to the parameter value with unlimited data. All of these properties are easy to motivate: they impose conditions on the probabilistic behavior of $\hat{\theta}$ that "make good sense". The next property we're going to introduce is not so intuitive, but has really important theoretical implications.

Assume we draw $X_1, \dots, X_n \sim f_x$. Imagine two statisticians:

Whether or not an estimator is sufficient refers to the amount of "information" it contains about the unknown parameter. Estimates are calculated using values obtained from random samples (drawn from either p_x or f_x). If everything that we can possibly know from the data about θ is encapsulated in the estimate $\hat{\theta}$, then the corresponding estimator $\hat{\theta}$ is said to be sufficient.

Example of an estimator that is not sufficient: Let $Y_1, \dots, Y_n \sim f_y$, where $f_y = \frac{2y}{\theta^2}$ for $0 \leq y \leq \theta$. The MoM estimator for this distribution is $\hat{\theta}_{MoM} = \frac{3}{2}\bar{Y}$. Consider two random samples of size 3: $\{3, 4, 5\}$ and $\{1, 3, 8\}$.

Sufficiency

Let W_1, \dots, W_n be a random sample from $f_w(w; \theta)$. The estimator $\hat{\theta} = h(W_1, \dots, W_n)$ is said to be *sufficient* for θ if $P(W_1, \dots, W_n | \hat{\theta} = t)$ does not depend on θ .

Factorization Criterion

Let W_1, \dots, W_n be a random sample from $f_w(w; \theta)$. The estimator $\hat{\theta} = h(W_1, \dots, W_n)$ is sufficient for θ if and only if the likelihood function, $L(\theta)$, factors into the product of the pdf for $\hat{\theta}$ and a function of the sample that does not involve θ :

Example: Let $X_1, \dots, X_n \sim \text{Pois}(\lambda)$. Show that $\hat{\lambda} = \sum X_i$ is a sufficient statistic for λ .

Factorization Criterion Round 2

Let W_1, \dots, W_n be a random sample from $f_w(w; \theta)$. The estimator $\hat{\theta} = h(W_1, \dots, W_n)$ is sufficient for θ if and only if the likelihood function, $L(\theta)$, factors into:

Example: Let $X_1, \dots, X_n \sim \text{Pois}(\lambda)$. Show that $\hat{\lambda} = \sum X_i$ is a sufficient statistic for λ .

Proof:

Example: Suppose Y_1, \dots, Y_n are drawn from $f_y(y; \theta) = \frac{2y}{\theta^2}$ where $0 \leq y \leq \theta$. The MLE for θ is $\hat{\theta} = Y_{max}$. Is Y_{max} sufficient for θ ?

More notation:

2 Jointly Sufficient Statistics

When a parameter θ is multidimensional, sufficient statistics will typically need to be multidimensional as well. Sometimes, no one-dimensional statistic is sufficient even when θ is one-dimensional. In either case, we need to extend the concept of sufficient statistic to deal with cases in which more than one statistic is needed in order to be sufficient.

Jointly Sufficient Statistics

Suppose that for each θ and each possible value of (t_1, \dots, t_k) of (T_1, \dots, T_k) , where each $T_i = h_i(X_1, \dots, X_n)$, the conditional joint distribution of (X_1, \dots, X_n) given $(T_1, \dots, T_k) = (t_1, \dots, t_k)$ does not depend on θ . Then (T_1, \dots, T_k) are called *jointly sufficient statistics* for θ

Factorization Theorem for Jointly Sufficient Statistics

Let r_1, \dots, r_k be functions. The statistics $T_i = r_i(X_1, \dots, X_n)$ are jointly sufficient for θ if and only if the joint pdf $f(x_1, \dots, x_n | \theta)$ can be factored into:

Example: Jointly sufficient statistics for the parameters of a normal distribution

3 Rao-Blackwell Theorem

Mean Squared Error

The following theorem says that if we want an estimator with small MSE we can confine our search to estimators which are functions of sufficient statistics.

Rao-Blackwell Theorem

Let $\hat{\theta}$ be an estimator of θ with $E(\hat{\theta}^2) < \infty$. Suppose that T is a sufficient estimator of θ , and let $\theta^* = E(\hat{\theta}|T)$. Then, for all θ ,

Example: Let $X_1, \dots, X_n \sim \text{Pois}(\lambda)$. We know that $\hat{\lambda} = \sum X_i$ is a sufficient statistic for λ . Let's "Rao-Blackwellize" the unbiased (but bad) estimator $\tilde{\lambda} = X_1$: