The Rao-Blackwell Theorem and the UMVUE

Guy Lebanon

May 24, 2006

In a previous note, sufficient statistic was discussed as a data-reduction tool. Here, we explore an another application of sufficient statistic. Via the Rao-Blackwell theorem, it can be used to improve an estimator and reach the best unbiased estimator.

Theorem 1 (Rao-Blackwell)**.** *Let $\hat{\theta}$ be an estimator, and T a sufficient statistic, both for θ. Then the estimator $\hat{\theta}^*(X_1, \ldots, X_n) = E(\hat{\theta}(X_1, \ldots, X_n)|T(X_1, \ldots, X_n))$ is at least as good as θ in terms of the MSE:*

$$\forall\theta \qquad \mathit{MSE}(\hat{\theta}^*) \leq \mathit{MSE}(\hat{\theta}).$$

Proof. Since $\mathsf{Var}(X) \geq 0$, we have that $\mathsf{E}(X^2) \geq \mathsf{E}(X)^2$ and

$$\mathsf{E}((\hat{\theta} - \theta)^2|T) \geq (\mathsf{E}(\hat{\theta} - \theta)|T)^2 = (\mathsf{E}(\hat{\theta}|T) - \theta)^2 = (\hat{\theta}^* - \theta)^2.$$

Taking expectation of both sides we have

$$\mathsf{E}((\hat{\theta}^* - \theta)^2) \leq \mathsf{E}(\mathsf{E}((\hat{\theta} - \theta)^2|T)) = \mathsf{E}((\hat{\theta} - \theta)^2).$$

□

It can be shown that the inequality in the above theorem is strict if the two estimators are different $\hat{\theta}^* \neq \hat{\theta}$. The above procedure for improving an estimator is sometimes called a Rao-Blackwellization procedure. The Rao-Blackwell theorem may be extended to the multi-parameter case where θ is a vector, and is also correct when the MSE is replaced with the MAE: $\mathsf{E}(|\hat{\theta} - \theta|)$.

When we examine the form of the superior estimator $\hat{\theta}^*$, we note that it is a random variable that is a function of the sufficient statistic $T(X_1, \ldots, X_n)$ (it depends on the data $X_1, \ldots, X_n$ only through the value of $T(X_1, \ldots, X_n)$). In other words, if we have an estimator that is not a function of the sufficient statistic by itself, it can be improved upon by Rao-Blackwellization. We can thus restrict ourselves in our search of estimators to functions of sufficient statistics.

Example: Recall that for normal data and $\theta = \mu$, $\bar{X}$ is the minimal sufficient statistic. Therefore an estimator $\hat{\mu}$ that uses the sample median is not going to be as good as some function of $\bar{X}$ in terms of the MSE.

Example: Recall that $\bar{X}$ and S^2 are minimal sufficient statistics for (μ, σ^2) in normal data. An estimator $\hat{\sigma}$ that uses the range $\max(X_1, \ldots, X_n) - \min(X_1, \ldots, X_n)$ will be dominated by an estimator that uses $\bar{X}$ and S^2 .

The notion of complete statistics sharpen the Rao-Blackwell theorem in a way that enables characterizing the best estimator among all unbiased estimators.

Definition 1. *$\hat{\theta}$ is the uniformly minimum variance unbiased estimator (UMVUE) if it is an unbiased estimator of θ, and its variance is smaller than any other unbiased estimator (for all values of θ). By the bias-variance decomposition of the MSE, it is also the best estimator in terms of the MSE among the class of unbiased estimators.*

Definition 2. *A sufficient statistic T is complete (w.r.t θ) if*

$$E(g(T(X_1 \ldots, X_n)) = 0 \ \forall\theta \quad \Rightarrow \quad P(g(T) = 0) = 1 \ \forall\theta.$$

The above definition is sometimes referred to as a 'complete sufficient statistic does not admit any unbiased estimators of 0' (others than the zero function).

Theorem 2 (Lehmann-Scheffe). *An unbiased estimator that is a complete sufficient statistic is the unique UMVUE.*

The Lehmann-Sheffe theorem states that if you find an unbiased estimator that is a function of a complete sufficient statistic, it is the unique UMVUE. For such estimators, Rao-Blackwellization acts as an identity operator.

Example: Consider $\hat{\theta} = \bar{X}$ as an estimator of the parameter θ of a Bernoulli distribution. By the factorization theorem $L(\theta) = \prod \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} = g(\theta, \bar{X})$ it is a sufficient statistic. To show completeness, note that under the Bernoulli distribution, $\sum X_i$ is a binomial RV and

$$\mathsf{E}\left(g(\bar{X})\right) = \sum_{i=0}^{n} \frac{n!}{i!(n-i)!}\theta^i(1-\theta)^{n-i}g(i/n) = n!(1-\theta)^n \sum_{i=0}^{n} \frac{g(i/n)}{i!(n-i)!}\left(\frac{\theta}{1-\theta}\right)^i$$

is a polynomial in $\theta/(1-\theta)$ and for it to be identically zero (for all θ), all its coefficients have to be zero. This implies that $g(\bar{X}) = 0$ with probability 1 and therefore $\bar{X}$ is a complete sufficient statistic. Since it is also an unbiased estimator of θ, by the Lehmann-Sheffe theoremit is the unique UMVUE for θ.