Sufficient Statistics and minimum variance of estimators

March 13, 2020

1 Estimating support of pdf's of random variables

We have seen that if U is a minimal, sufficient statistic for the estimation of θ, and we find a function g such that $\mathbb{E}[g(U)]=\theta$, then $\hat{\theta}=g(U)$ is the minimum variance, unbiased estimator of θ. In other words, it is the best estimator. When we say minimal, we mean we have just one statistic to estimate one parameter.

We also saw that if $f(X|\theta)$ is an exponential family of the form

$$f(X|\theta)=a(\theta)b(X)e^{-c(\theta)d(X)} \text{ if } a\leq x\leq b,$$

and $f(X|\theta)=0$ otherwise, where the bounds (IMPORTANT) a and b don't depend on θ, then $\sum d(x_i)$ is a minimal sufficient statistic. Now we will consider some cases where the bounds of the support of the pdf depend on θ.

Example 1.1. *Let $X_1,X_2,...,X_n$ be a random sample with pdf:*

$$f_X(x|\theta)=\frac{2x}{\theta^2} \text{ if } 0\leq x\leq\theta,$$

and $f(X|\theta)=0$ otherwise. Show that $X_{(n)}=\max\{X_1,X_2,...,X_n\}$ is a sufficient statistic, and then use it to find the MVUE.

We now verify that $X_{(n)}$ is sufficient using the Likelihood Factorization Criterion.

$$L(x_1,...,x_n|\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n}x_i \text{ if } 0\leq x_i\leq\theta \text{ for all } i$$

Letting $g(\theta,u)=\frac{1}{\theta^n}$ if $\theta>0$ and 0 otherwise, and $h(x_1,x_2,...,x_n)=\prod_{i=1}^n x_i$. Then the observation that $\theta\geq x_i$ for all i is equivalent to $\theta\geq x_{(n)}$, where $x_{(n)}=\max\{x_1,x_2,...,x_n\}$, gives

$$L(\theta)=g(\theta,x_{(n)})h(x_1,x_2,...,x_n)$$

so we have verified that $X_{(n)}$ is sufficient.

We would now like to find a function of $X_{(n)}$ that is an unbiased estimator. Our first try will be $X_{(n)}$, itself. If we are lucky its expectation is θ, or at least something close to θ.

In order to compute the expectation of $X_{(n)}$ we need to compute it's pdf. You should be able to derive

$$f_{X_{(n)}}=n\left(\frac{x^2}{\theta^2}\right)^{n-1}\left(\frac{2x}{\theta^2}\right)=n(\frac{x}{\theta})^{2n}$$

and then get

$$E[X_{(n)}]=\int_0^\theta n\left(\frac{x}{\theta}\right)^{2n}d\theta=\frac{2n}{2n+1}\theta.$$

So we see $X_{(n)}$ is not an unbiased estimator of θ but $\frac{2n+1}{2n}X_{(n)}$ is an unbiased estimator of θ. Since $\frac{2n+1}{2n}X_{(n)}$ is a function of the minimal, sufficient statistic, by the Rao-Blackwell Thm $\frac{2n+1}{2n}X_{(n)}$ is the MVUE.

We'll now briefly look at 2 more examples of random variables whose pdf's are supported on intervals that depend on unknown parameters.

EXAMPLE 1.2. *Let* $X_1, X_2, ..., X_n$ *be a random sample with pdf:*

$$f_X(x|\theta) = \frac{1}{2\theta} \text{ if } -\theta \leq x \leq \theta,$$

and $f(X|\theta) = 0$ *otherwise. In other words each random variable is uniformly distributed on the interval* $[-\theta, \theta]$.

In this case, similar to above, it can be shown that $\hat{\theta} = \max\{-X_{(1)}, X_{(n)}\}$ *is sufficient for the estimation of* θ. *The distribution of* $\hat{\theta}$ *can also be computed, by considering the CDF* $\mathbb{P}(\hat{\theta} < t)$ *and writing this event in terms of the random sample* $X_1, X_2, ..., X_n$, *then using independence as we do for computing the distribution of* $X_{(1)}$ *or* $X_{(n)}$.

EXAMPLE 1.3. *Let* $X_1, X_2, ..., X_n$ *be a random sample with pdf:*

$$f_X(x|\theta) = \frac{1}{2} \text{ if } \theta - 1 \leq x \leq \theta + 1,$$

and $f_X(x|\theta) = 0$ *otherwise. In other words each random variable is uniformly distributed on the interval* $[\theta - 1, \theta + 1]$.

In this case, we actually need to keep 2 statistics and the pair $(X_{(1)}, X_{(n)})$ *is sufficient for the estimation of* θ. *It turns out there is no way to compress the the pair* $(X_{(1)}, X_{(n)})$ *further, without losing information from the sample about* θ. *Note that compressing from the full sample to just* $(X_{(1)}, X_{(n)})$ *doesn't lose information about* θ. *In this case computing the MVUE can be much harder.*

2 Lower bounds on Variances of estimators

One of our goals in this Chapter is to find estimators with small variances. When we have a minimal sufficient this is accomplished by finding the MVUE, but it general we would like more ways to know if we have a good estimator. The following Theorem tells us the best possible situation, so if we manage to find an estimator with a variance close to this bound, we are satisfied.

THEOREM 2.1. *Let* $X_1, X_2, ..., X_n$ *be a random sample with pdf* $f_X(x|\theta)$. *Let* $\hat{\theta}$ *be an unbiased estimator of* θ, *then*

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n \frac{\partial^2}{\partial \theta^2} \mathbb{E}[\ln(f_X(X|\theta))]}$$

The term $n\mathbb{E}[\ln(f_X(X|\theta))$ should be thought of as taking the log-likelihood function for the sample, and evaluating it at the random variables X_i (instead of as usual x_i, the observed values of the random variables) and then taking the expectation. This is called the Fisher Information, but we will not discuss it more in this class. The main point of introducing this Theorem will be seen below.

3 Confidence intervals for MLE

We saw in Chapter 8, that if we have an estimator and we know its distribution, then we can make confidence intervals. Broadly speaking, we mainly considered 2 cases, one where we used properties of the sample distribution (for example sum of normal r.v.'s is normal, or formulas for order statistic) to get the distribution of the estimator exactly, and the other where we applied to central limit theorem to our estimator. More precisely, in the second situation, the empirical mean $\overline{Y}$ is approximately normal, so the normal table tells us the probability $\overline{Y}$ is a given number of standard deviations away from it's mean.

It turns out the MLE behaves similarly to $\overline{Y}$, and is also approximately normal. For large n, the variance of the normal converges to $\frac{1}{n\frac{\partial^2}{\partial\theta^2}\mathbb{E}[\ln(f_X(X|\theta))]}$. Which from the last section we saw is the smallest possible variance. So the MLE, at least in large samples, is basically the best estimator. We precisely record this in the following theorem:

THEOREM 3.1. *Let X_1,X_2,...,X_n be a random sample with pdf $f_X(x|\theta)$. Let $\hat{\theta}$ be the MLE estimator of θ. Then*

$$\frac{\hat{\theta}-\theta}{\left(n\frac{\partial^2}{\partial\theta^2}\mathbb{E}[\ln(f_X(X|\theta))]\right)^{-1/2}}$$

is approximately a standard normal random variable.

In fact we can get a stronger statement if we recall that for any function t the MLE for $t(\theta)$ is just $t(\hat{\theta})$, when $\hat{\theta}$ is the MLE for θ.

THEOREM 3.2. *Let X_1,X_2,...,X_n be a random sample with pdf $f_X(x|\theta)$. Let $\hat{\theta}$ be the MLE estimator of θ. Then*

$$\frac{t(\hat{\theta})-t(\theta)}{\left|\frac{\partial t(\theta)}{\partial\theta}\right|\left(n\frac{\partial^2}{\partial\theta^2}\mathbb{E}[\ln(f_X(X|\theta))]\right)^{-1/2}}$$

is approximately a standard normal random variable.

Example 9.18 in the book, shows you an example of computing $n\frac{\partial^2}{\partial\theta^2}\mathbb{E}[\ln(f_X(X|\theta))]$.