Econ-240A (1st half - Fall 2008)
Note on Completeness 2008-10-19 (v2)

Owen Ozier

The idea, expressed in CB Theorem 6.2.25, that a particular sufficient statistic is complete only if (among other conditions from the exponential family representation) $\{(w_1(\theta), ..., w_k(\theta)) : \theta \in \Theta\}$ contains an open set in $\mathbb{R}^k$, has gone more or less without explanation; let me try to give an intuitive one based on some examples.

Suppose that we have random variables drawn from the Exponential(θ) distribution (in the sense of CB equation 3.3.11). It is straightforward to see that this family of distributions is an exponential family (the CB equation 3.4.1 meaning, now):

$$\frac{1}{\theta}e^{-x/\theta}\mathbb{1}(x \geq 0) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right) \text{ where: } \begin{array}{rcl} h(x) & = & \mathbb{1}(x \geq 0) \\ c(\theta) & = & \frac{1}{\theta} \\ k & = & 1 \\ w_1(\theta) & = & \frac{1}{\theta} \\ t_1(x) & = & x \end{array}$$

Ordinarily, we consider the parameter space $\theta \in \Theta = (0, \infty)$. Since $w_1(\theta) = 1/\theta$ is a continuous function, then because Θ contains an open set in $\mathbb{R}^k = \mathbb{R}^1 = \mathbb{R}$ (in fact, it is one), so does $\{w_1(\theta) : \theta \in \Theta\}$. Thus, by Theorem 6.2.25, the statistic $T(\mathbf{X}) = \sum_{i=1}^n t_1(X_i) = \sum_{i=1}^n X_i$ is complete (and by Theorem 6.2.10, sufficient).

What if Θ didn't include an open set (and thus $w(\Theta)$ also didn't)?

Suppose $\Theta = \{1, 2\}$. Then in this case, we may be able to construct a function $g(\cdot)$, based only on $T(\mathbf{X}) = \sum_{i=1}^n X_i$, that shows that T is not complete. Referring back to the definition of completeness (6.2.21), paraphrased here:

$$\text{If } E[g(T)] = 0\ \forall\theta \implies P(g(T) = 0) = 1\ \forall\theta, \text{ then } T(X) \text{is a complete statistic.}$$

Note that $E[T] = n\theta$ and $E[T^2] = n^2\theta^2 + n\theta^2$ (you can work out why this is true). Thus, $E[T^2/(n^2+n)] = \theta^2$, and $E[T/n] = \theta$.

To construct a $g(\cdot)$ that shows that T is not complete, we need $E[g(T)] = 0\ \forall\theta$, but now, we know that θ could only take on two values: 1 and 2, so we only need an expression that is zero for those two values. If we had a function whose expectation were $(\theta - 1)(\theta - 2)$, it would satisfy this condition. Multiplying out, we get: $(\theta-1)(\theta-2) = \theta^2 - 3\theta + 2$. Based on the notes above, it is easy to construct a function with this expectation:

$$E\left[g(T) = \frac{T^2}{(n^2+n)} - \frac{3T}{n} + 2\right] = \theta^2 - 3\theta + 2 = 0\ \forall\theta \in \Theta = \{1, 2\}$$

However, $P[g(T) = 0] \neq 1$ (easy to verify by example), so in this case, $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is not complete.

Why the k-dimensional open set in $\mathbb{R}^k$?

The discussion below Theorem 6.2.25 in the textbook discusses a case of this, the set of normals with parameter values $\mathcal{N}(\theta, \theta^2)$. The problem this introduces is that the set of $\{w_i(\theta)\}$ now only contains values on a parabola in $\mathbb{R}^2$, so it does not contain any two-dimensional open neighborhoods. Practically, the reason this makes the usual statistics $(\bar{X}, S^2)$ no longer complete, is that it gives us two ways to create functions with expectations related to theta, based on either $\bar{X}$ or S^2. These two methods may have the same expectation, but different realized values in any given sample. As above, you can construct functions of $\bar{X}$ and S^2 that have expectation zero, but that are not zero with probability 1, so the statistics are no longer complete.

Summary:

In essence, if $\{(w_1(\theta), ..., w_k(\theta)) : \theta \in \Theta\}$ does not contain an open set in $\mathbb{R}^k$ (it contained only singletons in the first example, and was of too low a dimension in the second example), then it will be possible to construct a function $g(T(\mathbf{X}))$ that has expectation zero but is not always zero. This would, in turn, make it possible to make many different unbiased estimators of θ based only on $T(\mathbf{X})$, so we could not be sure, without further work, whether any of them were actually the UMVU estimator of θ.