Stat210A: Theoretical Statistics **Lecture Date: September 11, 2014**

Basu's Theorem, Rao-Blackwell, and Unbiasedness

Lecturer: Michael I. Jordan *Scribe: Arturo Fernandez*

1 From Last Time

Recall that last lecture we dsicussed the concepts of sufficiency, minimal sufficiency, completeness and ancillarity.

Definition 1 (Ancillarity)**.** A statistic V is ancillary if its distribution does not depend on θ.

However, we note that several ancillary points put together could tell you something about θ. More importantly, we now have the machinery to state Basu's Thereom, which states an important relationship between complete, sufficient statistics and ancillary statistics.

2 Basu and Rao-Blackwell

Theorem 2 (Basu's Theorem, Thm 3.21 in Keener)**.** *(Keener, 2010, p.50) Let T be a complete, sufficient statistic for a family of distributions $\mathcal{P}$ (indexed by θ) and let V be ancillary. Then, T is independent of V for all θ.*

Proof. Define $p_A = \mathbb{P}_\theta(V \in A)$ and $q_A(T) = \mathbb{P}_\theta(V \in A \,|\, T)$. Now note that by smoothing (or the tower law)

$$p_A = \mathbb{P}_\theta(V \in A) = \mathbb{E}_\theta\left[\mathbb{1}_{\{V\in A\}}\right] = \mathbb{E}_\theta\left[\mathbb{E}[\mathbb{1}_{\{V\in A\}} \,|\, T]\right] = \mathbb{E}_\theta[\mathbb{P}_\theta(V \in A \,|\, T)] = \mathbb{E}_\theta[q_A(T)] \tag{1}$$

Thus, by the completeness of T, we see that $q_A(T) = p_A$ (a.e. $\mathcal{P}$). We can now prove theorem by the following

$$\begin{aligned}
\mathbb{P}_\theta(T \in B, V \in A) &= \mathbb{E}_\theta\left[\mathbb{1}_{\{T\in B\}}\mathbb{1}_{\{V\in A\}}\right] \\
&= \mathbb{E}_\theta\left[\mathbb{E}[\mathbb{1}_{\{T\in B\}}\mathbb{1}_{\{V\in A\}} \,|\, T]\right] && \text{(smoothing)} \\
&= \mathbb{E}_\theta\left[\mathbb{1}_{\{T\in B\}}\mathbb{E}[\mathbb{1}_{\{V\in A\}} \,|\, T]\right] && (\mathbb{E}[f(X)Z|X] = f(X)\mathbb{E}[Z|X]) \\
&= \mathbb{E}_\theta\left[\mathbb{1}_{\{T\in B\}} q_A(T)\right] \\
&= \mathbb{E}_\theta\left[\mathbb{1}_{\{T\in B\}} p_A\right] && \text{(by(1))} \\
&= p_A\mathbb{E}_\theta\left[\mathbb{1}_{\{T\in B\}}\right] && \text{(linearity)} \\
&= \mathbb{P}_\theta(V \in A)\mathbb{P}_\theta(T \in B)
\end{aligned}$$

□

To see how this helps us we will go through a well-known fact that the sample mean and sample variance of a guassian sample are independent.

Example 3. Let $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ and the family under consideration is $\mathcal{P}_\sigma = \{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}^n\}$ (i.e. $\mathcal{P}_\sigma$ is the family of all gaussian distributions with fixed standard deviation σ.) Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. We'll show that $\bar{X}$ is independent of s^2 by Basu's Theorem. That is, recall that by Example 2.3 (Keener, 2010, p.26) that $T(x) = \sum_{i=1}^n x_i$ is sufficient for μ (and in general $\sum_{i=1}^n T_j(X_i)$ are sufficient for exponential families). Moreover, by Theorem 4.3.1 of (Lehmann and Romano, 2005, p.142), $T(x) = \sum_{i=1}^n x_i$ is also complete. We note that scaling the statistic doesn't change theses properties. Thus to show independence, all that is left is to show is that s^2 is ancillary. We will do this by calculating the distribution and show that is doesn't depend on μ.

Let $Y_i = X_i - \mu$ and thus $Y_i \sim \mathcal{N}(0, \sigma^2)$. Furthermore

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \mu - \frac{1}{n}\sum_{j=1}^n (X_i - \mu))^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$$

so we see that distribution of s^2 does not depend on μ. By the Theorem of Basu, with $\bar{X}$ complete, sufficient and s^2 ancillary, we have that $\bar{X} \perp\!\!\!\perp s^2$

We briefly note that we see and use this phenomenon in many other places in statistics. In particular, when we have ratios of statistics. t-statistics and F-statistics come to mind and we'll often use the property of independence between the numerator and denominator. Closing up chapter 3, we will introduce the idea of *Rao-Blackwellisation*, which says that given an estimator and sufficient statistic T, we can make the estimator better (in the risk sense) by taking its conditional expectation with respect to T.

We also briefly noted the important contributions that David Blackwell provide to many fields, including game theory, convex analysis, statistics, mathematics, and economics. The following quote was pulled from his wikipedia page: '*I've worked in so many areas-I'm sort of a dilettante. Basically, I'm not interested in doing research and I never have been. I'm interested in understanding, which is quite a different thing. And often to understand something you have to work it out yourself because no one else has done it.*'

Theorem 4 (Rao-Blackwell, Thm 3.28 in Keener)**.** *(Keener, 2010, p.53) Let T be a sufficient statistics for $\mathcal{P}$ (again, indexed by θ), let delta be an estimator of $g(\theta)$, and define $\eta(T) = \mathbb{E}_\theta[\delta \mid T]$. Moreover, let $\mathcal{L}(\theta, \cdot)$ be convex in the second argument for all theta and suppose that $R(\theta, \delta) < \infty, \forall \theta \in \Omega$. Then*

$$R(\theta, \eta) \leq R(\theta, \delta) \tag{2}$$

Proof. We'll refer to Jensens inequality (Keener, 2010, p.52), but also allude to it's conditional analogue (Durrett, 2010, p.227), which is proved similarly. Note that

$$R(\theta, \eta(T)) = \mathbb{E}_\theta[\,\mathcal{L}(\theta, \mathbb{E}_\theta[\delta \mid T]\,)\,] \leq \mathbb{E}_\theta\,[\,\mathbb{E}_\theta[\mathcal{L}(\theta, \delta) \mid T\,]\,] = R(\theta, \delta)$$

where we have used Jensen's inequality and smoothing. □

Based on this theorem, one might as why we don't *rao-blackwellise* all of our estimators (if what we care about is risk). However, note that this first requires a sufficient statistic T and also the computational cost of evaluating that conditional expectation, which may or may not come cheap. On the other hand, this technique has been useful in many cases, including Monte Carlo Methods (Casella and Robert, 1996). That said, we are now done with Chapter 3 and our adventure into Theoretical Statistics continues with chapter 4.

3 Unbiasedness

Definition 5 (Unbiasedness)**.** An estimator δ is unbiased for $g(\theta)$ if

$$\mathbb{E}_\theta[\delta(X)] = g(\theta) \tag{3}$$

Furthermore, if $g(\theta)$ is known, and there exists an estimator $\delta(X)$ that is unbiased, then $g(\theta)$ is said to be *U-estimable.*

Moreover, it's not uncommon to when doing parameter estimation to start off by getting an unbiased estimator and then Rao-Blackwellising it. We proceed by showing an example in which we find the desired form of an unbiased estimator.

Example 6. Let $X\sim\text{Unif}(0,\theta)$. Then for $\delta(X)$ to be unbiased it must satisfy

$$\int_0^\theta \frac{1}{\theta}\delta(x)dx = g(\theta) \qquad \forall\theta > 0$$

similarly we can write this as

$$\int_0^\theta \delta(x)dx = \theta g(\theta) \qquad \forall\theta > 0$$

If we let $\theta \searrow 0$, then g is only estimable if $\theta g(\theta) \to 0$ as well. If the first derivative of g exists, then by the fundamental theorem of calculus note that we can differentiate both sides and arrive at

$$\delta(x) = g(x) + xg'(x)$$

Letting $g(\theta) = \theta$, it's clear that $\delta(x) = 2x$ is an unbiased estimator.

On the other hand, $g(\theta)$ need not always be U-estimable. Next, we provide an example of this case.

Example 7. Let $X \sim \text{Bin}(n, \theta)$ and $g(\theta) = \sin(\theta)$, then for δ to be unbiased it would have to satisfy

$$\sum_{k=0}^{n} \delta(k)\binom{n}{k}\theta^k(1-\theta)^{n-k} = \sin(\theta)$$

Note that the left hand side is a polynomial in θ with degree at most n. Yet, the sine function can not be represented by the sum of finite polynomials so it follows that $\sin(\theta)$ is not U-estimable.

However, there may exist multiples unbiased estimators so one might ask if there exists such a things as "best" unbiased estimator. If we consider risk under squared loss then we can formalize by recalling the bias-variance decomposition of risk and noting that for unbiased estimator this simplifies to just the variance of an estimator. Thus we will introduce the concept of a Uniform Minimum Variance Unbiased (UMVU) Estimator, which is uniform in the sense that it all applies for all θ in some index set, minimum variance in reference to risk under quadratic loss, and unbiased as the property that will define the class of estimators that we consider.

Definition 8 (UMVU)**.** An estimator δ is called UMVU if

$$\text{Var}_\theta(\delta) \leq \text{Var}_\theta(\delta') \qquad \forall\theta \in \Omega$$

for any competing δ', where all estimators under consideration are unbiased.

Although, this property is great, it's not clear that there always exist a UMVU estimator. However, the following theorem will lay out conditions for when it does.

Theorem 9 (UMVU Existence, Thm 4.4 in Keener)**.** *(Keener, 2010, p.62) Suppose $g(\theta)$ is U-estimable and let T be a complete, sufficient statistic. Then there exists an essentially unique (up to a set of measure zero) unbiased estimator based on T that is UMVU.*

Proof. Let δ be any unbiased estimator and let $\eta(T) = \mathbb{E}_\theta[\delta \,|\, T]$. By smoothing

$$g(\theta) = \mathbb{E}_\theta[\delta] = \mathbb{E}_\theta[\,\mathbb{E}_\theta[\delta \,|\, T]\,] = \mathbb{E}_\theta[\eta(T)] \tag{4}$$

so $\eta(T)$ is unbiased (i.e Rao-Blackwellisation leaves an unbiased estimator unbiased). Now let $\eta'(T)$ be a competing unbiased estimator. Since both are unbised, we have that

$$\mathbb{E}_\theta[\eta(T) - \eta'(T)] = 0, \qquad \forall \theta \in \Omega$$

By completeness, it follows that

$$\eta(T) = \eta'(T) \qquad \text{a.e. } \mathcal{P} \tag{5}$$

Now let δ' be a competing unbiased estimator. Then $\eta'(T) = \mathbb{E}_\theta[\delta' \,|\, T]$ is also unbiased by (3). Since risk is just variance, when under squared loss and considering unbiased estimators, we have by Rao-Blackwell (Theorem 4) and (5) that

$$\mathrm{Var}_\theta(\delta') = R(\theta, \delta') \geq R(\theta, \eta') = \mathrm{Var}_\theta(\eta') = \mathrm{Var}_\theta(\eta)$$

Thus, $\eta(T)$ is UMVU. □

In terms of what this means for actual parameter estimation, we note that this provides to avenues for finding a "good" estimator. One possibility is that we begin with an unbiased estimator and then condition it on a complete, sufficient statistic (i.e. Rao-Blackwellize it). On the other hand, we can start off with a complete, sufficient statistic and then de-bias it.

References

Casella, G. and Robert, C. P. (1996). Rao-blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94.

Durrett, R. (2010). *Probability: Theory and Examples*. Cambridge, New York, NY, 4 edition.

Keener, R. (2010). *Theoretical Statistics: Topics for a Core Course*. Springer, New York, NY.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer-Verlag, New York, NY, 3 edition.