Stat210A: Theoretical Statistics **Lecture Date: September 16, 2014**

Lecture 6: Unbiasedness, UMVU, Normal one-sample distribution theory

Lecturer: Michael I. Jordan *Scribe: Ahmed El Alaoui*

1 Uniform Minimum Variance Unbiased estimators (UMVU) cont.

Recall the result establishing the existence of UMVU estimators (thm 4.4 in Keener (2010)): Let $g(\theta)$ be a U-estimable parameter. If T is a complete sufficient statistic, then there exist an essentially unique estimator of $g(\theta)$ based on T that is UMVU. Furthermore, the proof is constructive, where the UMVU can be derived from any unbiased estimator by a Rao-Blackwellization procedure with respect to the statistic T.

Example 1: Let $X_i \sim \text{Unif}(0, \theta)$ i.i.d. random variables for $i = 1, \cdots, n$. Then $T = \max_i X_i$ is a sufficient and complete statistic for this family of distributions. But T, considered as an estimator of θ, is biased towards the left since $T \le \theta$ a.s. Let's construct an unbiased estimator based on T. Let $\eta(T)$ be a statistic which is function of T. The unbiasedness of $\eta(T)$ amounts to

$$\int_0^\theta \eta(t)\frac{nt^{n-1}}{\theta^n}dt = \theta,$$

which leads to

$$\int_0^\theta \eta(t)nt^{n-1}dt = \theta^{n+1}.$$

By differentiating w.r.t. θ and rearranging terms, we get

$$\eta(t) = \frac{n+1}{n}t.$$

We saw in previous lectures another unbiased estimator (not based on T): $\delta = 2\bar{X}$, where $\bar{X}$ is the sample average. Which, if either, estimator is best? A calculation of the variances of these estimator shows that

$$\text{Var}(\eta(T)) = \frac{3}{n+2}\text{Var}(\delta).$$

Hence, the variance of $\eta(T)$ is much smaller than that of δ. The ratio is in fact vanishing as the sample size grows. Given that both estimators are unbiased, this means that their risks under the quadratic loss compare in the same way.

Example 2: Let $X_i \sim \text{Bern}(\theta)$ i.i.d. Bernoulli random variables of parameter θ for $i = 1, \cdots, n$, and let $g(\theta) = \theta^2$ be the quantity to be estimated. Squaring an estimator of θ (e.g. $\bar{X}$) will typically be a biassed estimator. Let's try $\delta = X_1X_2$. By independence of X_1 and X_2 we indeed have $E(\delta) = \theta^2$. To obtain a better (even UMVU) estimator, we Rao-Blackwellize δ: $T = \sum_{i=1}^n X_i$ is a complete sufficient statistic. Then,

$$\begin{aligned} E(\delta|T=t) &= E(X_1X_2|T=t) = P(X_1X_2 = 1|T=t) \\ &= \frac{P(X_1=1, X_2=1, T=t)}{P(T=t)} = \frac{\theta^2\binom{n-2}{t-2}\theta^{t-2}(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} = \frac{t(t-1)}{n(n-1)}. \end{aligned}$$

By theorem 4.4 in Keener (2010), $\frac{T(T-1)}{n(n-1)}$ is UMVU.

2 Bias-Variance tradeoff:

Let's consider the setting described Example 1. We showed that $\frac{n+1}{n}T$ is UMVU. Now, consider the estimator $\delta(X) = aT$ for some $a \in \mathbf{R}$. The quadratic loss of this estimator is

$$R(\theta, \delta) = E_\theta(aT - \theta)^2 = a^2 E(T^2) - 2a\theta E(T) + \theta^2.$$

We have $E(T) = \frac{n}{n+1}\theta$ and $E(T^2) = \frac{n}{n+2}\theta^2$. Then minimizing the above quadratic in a yields $a = \frac{n+2}{n+1}$, hence $\delta^*(X) = \frac{n+2}{n+1}T$ is optimal for the risk function R in the class of estimators that are proportional to T. This shows in particular that δ^* has a smaller risk than the UMVU, since the latter also belongs to this class. This together with the fact that δ^* is biased, suggests that allowing some bias can significantly reduce the overall risk, hinting to the possibility that considering only unbiased estimators is limiting. The next example will demonstrate the failure of unbiasedness as a general inferential principle.

The failure of UMVU. Let X be a Poisson-distributed r.v. with mean θ. Suppose that the distribution of X is truncated at 0, i.e. we observe X only if its value is greater than or equal to 1. Suppose that we want to estimate the mass of the probability distribution of X that is lost to this truncation process (i.e. estimate $P(X = 0) = e^{-\theta}$). If δ is any unbiased estimator of the former parameter, one has

$$\sum_{k=1}^{\infty} \delta(k) \frac{\theta^k e^{-\theta}}{k!(1 - e^{-\theta})} = e^{-\theta}$$

which can be rewritten as

$$\sum_{k=1}^{\infty} \delta(k) \frac{\theta^k}{k!} = 1 - e^{-\theta} = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{\theta^k}{k!}$$

Hence, $\delta(k) = (-1)^{k+1}$. This means that if we insist on unbiasedness, we have to estimate $e^{-\theta}$ —which is a positive quantity— by $(-1)^{X+1}$ which is equal to -1 whenever X is even, which happens about half of the time! Alternatively, if one gives up on unbiasedness, better estimators could be found.

3 Normal one-sample distribution theory

Consider a random variable $X \sim N(\mu, \sigma^2)$ and define $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$. Some properties of the normal distribution:

- $aX + b \sim N(a\mu + b, \sigma^2)$.
- The moment generating function (MGF) of Z is
$$M_Z(u) = e^{u^2/2}.$$
- Hence the MGF of X is
$$M_X(u) = e^{\mu u + \frac{u^2\sigma^2}{2}}.$$
- If $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2$ independent, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Given a random sample $X_1, \cdots, X_n$ from $N(\mu, \sigma^2)$, the joint distribution form to a full-rank exponential family of parameter $\theta = (\mu, \sigma^2)$ with a complete sufficient statistic $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$. Alternatively, one can work with the statistic $(\bar{X}, S^2)$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ which is also complete and sufficient since there is a one-to-one mapping that links it to T. Recall that by Basu's theorem (thm 3.21 in Keener (2010)), $\bar{X}$ and S^2 are independent. From the first bullet above, the distribution of $\bar{X}$ is $N(\mu, \frac{\sigma^2}{n})$. Deriving the distribution of S^2 is a bit more intricate. Consider the Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$, denoted $\Gamma(\alpha, \beta)$ of density

$$f_{\alpha,\beta}(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

for $x > 0$ and zero otherwise. Γ being the standard Gamma function defined as $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. For a r.v. X having this density we have $E(X) = \alpha\beta$ and $\text{Var}(X) = \alpha\beta^2$. The MGF of X is

$$\begin{aligned} M_X(u) = Ee^{uX} &= \int_0^\infty e^{ux} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx \\ &= \int_0^\infty x^{\alpha-1} \frac{e^{(u-1/\beta)x}}{\beta^\alpha \Gamma(\alpha)} dx \\ &= \frac{1}{(1-\beta u)^\alpha} \quad \text{for} \quad u < 1/\beta. \end{aligned}$$

Chi-square: The chi-square distribution is a special case of the Gamma distribution, but generally defined using the normal distribution: If $Z \sim N(0,1)$ then Z^2 is said to have a chi-square distribution with one degrees of freedom, denoted χ_1^2. To establish the link between χ_1^2 the Gamma distribution, let's compute the MGF of Z^2:

$$M_{Z^2}(u) = Ee^{uZ^2} = \int_0^\infty e^{ux^2} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{(1-2u)^{1/2}} \equiv \Gamma(1/2, 2).$$

From the formula above, it is immediate that if $X_i \sim \Gamma(\alpha_i, \beta)$ are independent r.v.'s then $\sum_i X_i \sim \Gamma(\sum_i \alpha_i, \beta)$. Hence, if $Z_i \sim N(0,1)$ are independent then

$$\sum_{i=1}^p Z_i^2 \sim \Gamma(p/2, 2) \equiv \chi_p^2.$$

χ_p^2 is called the chi-square distribution with p degrees of freedom. Now, let's back to the problem of deriving the distribution of S^2. Let $X_1, \cdots, X_n$ from $N(\mu, \sigma^2)$, $Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1)$ for every i and $\bar{Z} = \frac{\bar{X}-\mu}{\sigma} \sim N(0, 1/n)$. Then $\sqrt{n}\bar{Z} \sim N(0,1)$ and $n\bar{Z}^2 \sim \chi_1^2$.

Let $V = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2$. This implies $V + n\bar{Z}^2 \sim \chi_n^2$. Note that $n\bar{Z}^2$ is only a function of $\bar{X}$ and V is only a function of S^2, hence $n\bar{Z}^2$ and V are independent. Thus, the MGF of $V + n\bar{Z}^2$ can be decomposed:

$$M_{V+n\bar{Z}^2}(u) = M_{n\bar{Z}^2}(u) M_V(u) = \frac{1}{(1-2u)^{n/2}}.$$

Which yields $V \sim \chi_{n-1}^2$. This along with the fact $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ determines the joint distribution of the pair $(\bar{X}, S^2)$ by independence.

References

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