MSF100
Statistical Inference Principles - Lecture 3

Rebecka Jörnsten
Mathematical Statistics
University of Gothenburg/Chalmers University of Technology

January 23, 2012

1 Recap

Last lecture we talked about point estimates. We compared Method of Moment estimators and Maximum Likelihood estimators. MoMs tend to be easy to compute, but are usually not the best in terms of variance. MLEs are sometimes difficult to compute and may be biased for small n. However, as n increases, MLEs tend to be the better estimates.

Speaking of "best" estimator, we talked about Best Unbiased estimators last lecture. We focused on unbiased estimators (classical) for which variance is the performance measure of interest. We had the Cramer-Rao theorem that gave us a bound for the minimum achievable variance among unbiased estimators W of $\tau(\theta)$ as

$$Var_\theta[W] \geq \frac{\tau'(\theta)^2}{E_\theta[(\frac{d}{d\theta}\log f(\tilde{X}|\theta))^2]}.$$

If an estimator W achieves the CR bound, we know it is the UMVUE (be careful about checking that the assumptions for the CR theorem holds, i.e. that the order of integration and derivation can be changed). However, the CR bound is not tight. We may have a UMVUE W, but it may not achieve the bound and so we have not been able to verify that it is UMVUE.

However, a corollary to the CR theorem states that if W is an unbiased estimator for $\tau(\theta)$ and the pdf is such that

$$a(\theta)[W(\tilde{X}) - \tau(\theta)] = \frac{d}{d\theta}\log L(\theta)$$

for some $a(\theta)$, then W achieves the CR bound. That is, everything works out if $W - \tau(\theta)$ is proportional to the derivative of the loglikelihood function.

This doesn't solve the problem for pdf f where the CR conditions do not apply or when the estimator does not have the form above. What do we do then?

2 Rao-Blackwell, Linking sufficiency and unbiasedness

Things get a lot easier when we stay within the exponential family of distributions. Why?

Let's first start by revisiting the idea of sufficient statistics. We have the following result by Rao and Blackwell:

<u>Rao-Blackwell</u>: Let W be any unbiased estimator of $\tau(\theta)$ and T a sufficient statistic for θ. If we construct a new estimator $Q(T) = E[W\|T]$, then • $E_\theta[Q(T)] = \tau(\theta)$, so Q "inherits" the unbiasedness of W • and $Var_\theta[Q(T)] \leq Var_\theta W$ for all θ

Proof: Using the standard results for complete and conditional expectations and variance we have

$$\tau(\theta) = E_\theta[Q] = E_\theta[E(Q|T)] = E_\theta[Q(T)] \text{ and}$$
$$Var_\theta[W] = Var_\theta[E(Q|T)] + E_\theta[Var(Q|T)] \geq Var_\theta[Q(T)]$$

The results isn't really interesting unless we insist on T being sufficient. While conditioning on something reduces the variability or randomness, it doesn't guarantee we get a useful estimator as a result. Here's an example (from Casella/Berger).

- Take X_1, X_2 iid $N(\theta, 1)$. An unbiased estimator for θ is $\bar{X}$ with variance $V(\bar{X}) = 1/2$.
- Condition on the first sample, X_1, and we get

  $$Q = E_\theta[\bar{X}|X_1]$$

  where we have $E_\theta[Q(X_1)] = \theta$ (unbiased) and $Var_\theta[Q(X_1)] \leq Var_\theta[\bar{X}]$. (Note, we are working under the assumption that X_i are iid $N(\theta, 1)$ so $X_2|X_1 \sim N(\theta, 1)$ and $\bar{X} \sim N(\theta, 1/n)$. What would happen if the Xs were correlated?)
- However, $Q(X_1) = E_\theta[\bar{X}|X_1] = \frac{1}{2}E[X_2|X_1] + \frac{1}{2}X_1 = \frac{1}{2}\theta + \frac{X_1}{2}$ which is useless as an estimator since it depends on the unknown θ. (Try at home: What if X_i were $Be(\theta)$? and $n = 2$?)
- If T is sufficient (what we condition on), then we know $Q|T$ does *not* depend on θ except through T and everything works out. In the above example, X_1 is *not* sufficient.

Rao-Blackwell guarantees that we can always get smaller a variance for an unbiased estimator by conditioning it on a sufficient statistic to form a new estimator. Of course, we don't know that this gets us the best unbiased estimator, but we're almost done.

Consider a sufficient statistic, $T(\tilde{X})$ (this is pretty to easy to obtain as we've seen in lecture 1). Construct a new estimator $Q(T) = E_\theta(Q(\tilde{X})\|T(\tilde{X}))$ If $Q(T)$ achieves the CR bound, it is the best unbiased estimator.

If $Q(T)$ does not achieve this bound, we're stuck again. On the other hand, it turns out that if we can go ahead to find the best unbiased estimator if we can show that our estimator is *uncorrelated with all unbiased estimators of 0*. Consider an estimator $Q' = Q + aZ$, for some constant a. It is still unbiased for $\tau(\theta)$ if the expected value of Z is 0. Now

$$Var(Q') = Var(Q) + a^2Var(Z) + 2aCov(Q, Z)$$

Q is UMVUE if $a^2Var(Z) + 2aCov(Q, Z) \geq 0$. But this function can be negative, for some values of a if $Cov(Q, Z)$ is non-zero. If Q and Z are uncorrelated, $Var(Q') \geq Var(Q)$.

This looks difficult to work with. To see that Q is UMVUE we need to check its correlation with *all* unbiased estimators of 0...

However, what if we knew that for our $f(x|\theta)$ there are no unbiased estimators of 0, except for 0 itself, then we're OK since 0 is always uncorrelated with any other variable. This should remind you of *completeness* from lecture 1. Recall,

> Completeness: $f(t|\theta)$ is a complete family of pdf's for statistic $T(X)$. If $E_\theta[g(T)] = 0$ for all θ implies $g(T) = 0$ for all θ, then $T(X)$ is a complete statistic (and vice versa).

Now we're home.

> Let T be a complete sufficient statistic for θ and $Q(T)$ be an estimator based only on T (as we would get if we use $Q(T) = E_\theta[Q|T]$). Then $Q(T)$ is the unique best unbiased estimator of its expectation $E_\theta[Q]$.

What this means is that, for any (any you can come up with - fairly easy task) unbiased estimator for $\tau(\theta)$ and we have a complete sufficient statistic T for θ, then we can construct the best unbiased estimator as above. As we saw in lecture 1, the tricky part is really the completeness.... which turned out to not be tricky at all if we stayed within the exponential family of distributions. Note, we have not shown that this estimator achieves the CR bound though. It may not, in which case there are no unbiased estimators that do.

Example

- Let X_i iid with $f(x|\theta) = \theta x^{\theta-1},\ 0 < x < 1,\ \theta > 0$.
- We can write $f(x|\theta) = exp[(\theta - 1)\log(x) + \log(\theta)]$ and so f belongs to the exponential family of distributions with $t(x) = \log(x)$ and the complete sufficient statistic is $T(\tilde{X}) = \sum_i \log(x_i)$
- The likelihood is $L(\theta|\tilde{x}) = \prod_i \theta x_i^{\theta-1}$, which is easier to work with on a log-scale.
- The loglikelihood is $l(\theta|\tilde{x}) = \sum_i(\theta - 1)\log(x_i) + n\log(\theta)$.
- The derivative is $\frac{d}{d\theta}l(\theta|\tilde{x}) = \sum_i \log(x_i) + n/\theta$, which gives us a zero-crossing at $\hat{\theta} = \frac{-n}{\sum_i \log(x_i)}$
- This is indeed a maximizer of L since $\frac{d^2 l}{d\theta^2} = -n/\theta^2 < 0$
- The CR bound is $1/E_\theta[(\frac{d}{d\theta}l(\theta|\tilde{X}))^2]$, where here we have that

$$E_\theta[(\sum_i \log(X_i) + n/\theta)^2] = \frac{n^2}{\theta^2} + \frac{2n}{\theta}\sum_i E_\theta[\log(X_i)] + E_\theta[(\sum_i \log(X_i))^2]$$

$$E_\theta[\log(X)] = -1/\theta,\ E_\theta[(\log(X))^2] = 2/\theta^2$$

  and so

$$E_\theta[(\sum_i \log(X_i) + n/\theta)^2] =$$

$$= \frac{n^2}{\theta^2} - \frac{2n^2}{\theta^2} + \sum_i E[\log(X_i)^2] + 2\sum_{1\le i<j\le n} E[\log(X_i)\log(X_j)] = \frac{-n^2}{\theta^2} + \frac{2n}{\theta^2} + 2\frac{n(n-1)}{2}(\frac{-1}{\theta})(\frac{-1}{\theta}) =$$

$$= \frac{n}{\theta^2}$$

  and so the CR bound is θ^2/n
- But does the MLE achieve this? First of all, is it unbiased? Try to check this at home (you can do it by simulation if you want). We will go through the delta-method next week. What if we want to estimates $g(\theta) = 1/\theta$?
- What *do* we know about MLEs? We know that they're *asymptotically* unbiased, but may be quite bad for small n. As we shall see, we also know MLEs *asymptotically* achieve the CR bound, i.e. for large sample sizes n the MLE is UMVUE.

Example

- Let $X_i \sim Poi(\lambda)$ but we care about parameter $\theta = e^{-\lambda}$.
- What is the UMVUE for $\theta = e^{-\lambda}$?
- We have that $L(\theta|\tilde{x}) = f(\tilde{x}|\theta) = h(\tilde{x}) \prod_i \theta(-\log\theta)^{x_i}$
- The loglikelihood is $l(\theta|\tilde{x}) = n\log(\theta) + \log(-\log(\theta))\sum_i x_i + \ constant$
- At home, come up with the MLE for θ.
- This is an exponential family so the complete sufficient statistic is $T(\tilde{X}) = \sum_i X_i$
- Now we need an unbiased estimator for θ. Well, $\theta = e^{-\lambda}$ is the probability that the Poisson r.v. takes on value 0, i.e. $P(X = 0)$. An unbiased estimator for θ is thus $\sum_i 1\{X_i = 0\}/n$
- We could now try to figure out what $Q(T) = E_\theta[\sum_i 1\{X_i = 0\}|T = \sum_i X_i]$ is, but that looks pretty hard.
- The point is we don't need to go for the "best" estimator of θ at this stage since it is the conditioning that makes the trick. Let's instead consider the estimator $1\{X_1 = 0\}$, i.e. just checking the *first* sample value. It is an unbiased estimator for θ, just not a very efficient one since it uses only one sample value.
- We can improve upon it by conditioning on T: $E[X_1 = 0|\sum_i X_i = t] = P(X_1 = 0|\sum_i X_i = t) = \frac{P(X_1=0,\sum_{i=2}^n X_i=t)}{P(\sum_i X_i=t)}$
- The sum of K iid $Poi(\lambda)$ has distribution $Poi(K\lambda)$ (compound poisson) and so

$$E[X_1 = 0|\sum_i X_i = t] = \frac{e^{-\lambda}e^{-(n-1)\lambda}((n-1)\lambda)^t}{e^{-n\lambda}(n\lambda)^t} = (1 - \frac{1}{n})^t$$

- Our estimator is thus $\hat{\theta} = (1 - \frac{1}{n})^{\sum_i x_i}$.