Stat 2911 Lecture Notes

Class 14, 2017

Uri Keich

Standard error of an estimator, Estimated standard error or MSE, Studying estimators: Poisson, HW problem estimators (Rice 8.4, 8.5)

Class Reps!

Initially, given a single sample $x_1, \dots, x_n$ from a dist. $F(\theta)$ we estimate θ by $\tilde{\theta} = \varphi(x_1, \dots, x_n)$, e.g., $\hat{\theta} = \bar{x}$.

To compare estimators we need to consider all possible samples, which we do by studying the <u>RV</u>

$$\Theta = \varphi(X_1, \dots, X_n),$$

where $X_1, \dots, X_n$, is now a random sample from $F(\theta)$, e.g. $\Theta = \frac{1}{n}\sum_1^n X_i$.

An estimator is <u>unbiased</u> if $E(\Theta) = \theta$.

The quality of an estimator can be measured by

$$MSE(\Theta) = E(\Theta - \theta)^2 = V(\Theta) + (E(\Theta) - \theta)^2.$$

<u>Examples</u>

1) $X_1, \dots, X_n$ are ind. Bernoulli(θ) RVs.

$\Theta = \Theta_n = \frac{1}{n}\sum_1^n X_i$ (strictly speaking, $\Theta = \Theta(n)$)

is an unbiased estimator:

$$E(\Theta) = \theta$$

$$MSE(\Theta_n) = V(\Theta_n) = \frac{\theta(1-\theta)}{n} \xrightarrow[n]{} 0$$

Def. The standard deviation of an L^2-RV $\underline{X}$ is

$$\sigma_{\underline{X}} = \sqrt{V(\underline{X})}.$$

The standard error of an estimator Θ is its standard deviation:

$$\sigma_{\Theta} = \sqrt{V(\Theta)}.$$

Example. For the Bernoulli(θ) estimator $\Theta = \overline{X}$,

$$\sigma_{\Theta} = \sqrt{\frac{\theta(1-\theta)}{n}}.$$

The std. error is often unknown, hence we introduce the estimated standard error, where the unknown θ is replaced with its estimate.

Example. Bernoulli(θ), with $\Theta = \overline{X}$

$$\hat{\sigma}_{\Theta} = \sqrt{\frac{\overline{x}(1-\overline{x})}{n}}$$

Unsurprisingly (why?) $\sigma_{\Theta} \xrightarrow[n]{} 0$ so in some sense $\Theta_n \to \theta$.
Can we make it more precise?

Recall: a sequence of RVs Y_n converges to the RV Y in probability, if $\forall \varepsilon > 0$,

$$P(|Y_n - Y| > \varepsilon) \xrightarrow[n]{} 0$$

Def. An estimator $\Theta = \Theta_n$ is consistent if $\Theta_n \xrightarrow[n]{} \theta$ in prob. ($\forall \varepsilon > 0$, $P(|\Theta_n - \theta| > \varepsilon) \xrightarrow[n]{} 0$)

Claim. If $MSE(\Theta_n) \xrightarrow[n]{} 0$ then Θ is a consistent estimator.

Proof.

$$P(|\Theta_n - \theta| \geq \varepsilon) = P((\Theta_n - \theta)^2 \geq \varepsilon^2)$$
$$\overset{\text{(Markov)}}{\leq} \frac{E(\Theta_n - \theta)^2}{\varepsilon^2}$$
$$= \frac{MSE(\Theta_n)}{\varepsilon^2} \xrightarrow[n]{} 0$$

Cor. $\Theta_n = \bar{X} = \frac{1}{n}\sum_{1}^{n} X_i$ is an unbiased and consistent estimator of θ for the Bernoulli(θ) dist.

Proof. We showed that $MSE(\Theta_n) \xrightarrow[n]{} 0$.

Does this ring a bell?

WLLN: $\frac{1}{n}\sum_{1}^{n} X_i \xrightarrow[n]{} \theta$

2) $X_1, X_2, \ldots, X_n$ are ind. Poisson(θ) RVs.

$\hat{\Theta} = \bar{X}$ (MLE, or moment)

$\Rightarrow E(\Theta) = E(X_i) =$

$\Rightarrow \Theta$ is estimator

$MSE(\Theta_n) = Var(\Theta_n) = V(X_i)/n$

$= \theta/n \xrightarrow[n]{}$

$\Rightarrow \Theta_n$ is estimator

The std. error of Θ is

$$\sqrt{\theta/n}$$

and the estimated std. error is

$$\sqrt{\bar{x}/n}$$

Similarly, we can estimate the MSE as

$$\widehat{MSE}(\Theta_n) = \bar{x}/n$$

Hardy-Weinberg equilibrium

AA Aa aa

Model: $\underline{X} = (X_1, X_2, X_3) \sim$ multinomial $(n; (1-\theta)^2, 2\theta(1-\theta), \theta^2)$

Data: $n = 1029$ $x_1 = 342$ $x_2 = 500$ $x_3 = 187$

Goal: estimate θ, the frequency of the minor allele.

Moment estimator: $\tilde{\theta} = \sqrt{x_3/n} \underset{\text{case}}{\overset{\text{our}}{=}} \sqrt{187/1029} \simeq 0.4263$

MLE (also moment): $\hat{\theta} = \dfrac{x_2 + 2x_3}{2n} \underset{\text{case}}{\overset{\text{our}}{\simeq}} 0.4247$

Intuitively, $\hat{\theta}$ is better because it uses all available data.

To further study these estimators we disregard our specific sample and consider the multinomial random vector $\underline{X}$ above to define

$$\tilde{\Theta} = \tilde{\Theta}_n = \sqrt{X_3/n} \qquad \hat{\Theta} = \hat{\Theta}_n = \frac{X_2 + 2X_3}{2n}$$

Are these estimators consistent / unbiased?

$$E_\theta(\hat{\Theta}) = E\left(\frac{X_2 + 2X_3}{2n}\right)$$

$$= \theta \qquad \text{(we did this calculation)}$$

$\Rightarrow \hat{\Theta}$ is unbiased

$$E(\tilde{\Theta}) = E\left(\sqrt{X_3/n}\right)$$

$X_3 \sim$

$$= \sum_{k=0}^{n} \sqrt{k/n} \binom{n}{k} (\theta^2)^k (1-\theta^2)^{n-k} \quad = ?$$

Example: choose $\theta^2 = 1/2$ then $\theta = 1/\sqrt{2} \approx 0.7071$

n:	10	1000
$E(\tilde{\Theta})$:	0.6971	0.7070

This shows that $\tilde{\Theta}$ is but also suggests that the bias diminishes as the sample size, n, grows.

More important is the MSE.

$$MSE(\hat{\Theta}) \overset{?}{=} V(\hat{\Theta})$$
$$= V\left(\frac{X_2 + 2X_3}{2n}\right)$$
$$= \frac{1}{4n^2}\left[V(X_2) + 4V(X_3) + \underbrace{2Cov(X_2, 2\cdot X_3)}_{4Cov(X_2,X_3)}\right]$$

$V(X_2) = np_2(1-p_2)$ where $p_2 = 2\theta(1-\theta)$

$V(X_3) = np_3(1-p_3)$ " $p_3 = \theta^2$

$Cov(X_2, X_3) = -np_2p_3$ (tutorial problem)

algebra $\Rightarrow MSE(\hat{\Theta}_n) = \frac{1}{2n}\theta(1-\theta) \xrightarrow[n]{}$

$\Rightarrow \hat{\Theta}$ is estimator.

The standard error of $\hat{\Theta}$ is: $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})} = \sqrt{\theta(1-\theta)/2n}$

The actual values of $\sigma_{\hat{\Theta}}$ and $MSE(\hat{\Theta})$ depend on the unknown θ. Instead we can:

(i) plug in $\hat{\theta} = \frac{X_2 + 2X_3}{2n} \approx 0.4247 \Rightarrow \widehat{MSE(\hat{\Theta})} \approx 1.187 \times 10^{-4}$

(ii) use $\theta(1-\theta) \leq 1/4$ to get

$$MSE(\hat{\Theta}) = \frac{1}{2n}\theta(1-\theta) \leq \frac{1}{8n}$$

Pluggin in $n = 1029$ yields

$MSE(\hat{\Theta}) \leq 1.215 \times 10^{-4}$ (greater than the estimated MSE)

$$MSE(\tilde{\Theta}) = E[(\tilde{\Theta}-\theta)^2] = E[(\sqrt{X_3/n}-\theta)^2]$$

$X_3 \sim \text{Binom}(n,\theta^2)$

$$= \sum_{k=0}^{n} (\sqrt{k/n}-\theta)^2 \binom{n}{k} (\theta^2)^k (1-\theta^2)^{n-k}$$

Like in the case of $E(\tilde{\Theta})$, we can compute this expression numerically for a given value of θ.

Before though, note the MSE can be simplified a little:

$$MSE(\tilde{\Theta}) = E(\tilde{\Theta}^2) - 2\theta E(\tilde{\Theta}) + \theta^2$$

Recall that $\tilde{\Theta} = \sqrt{X_3/n}$, so $\tilde{\Theta}^2 = X_3/n$.

$$\Rightarrow E(\tilde{\Theta}^2) = E(X_3)/n = \theta^2$$

$$\Rightarrow \quad MSE(\tilde{\Theta}) = 2\theta^2 - 2\theta E(\tilde{\Theta})$$

$$= -2\theta\underbrace{[E(\tilde{\Theta})-\theta]}_{\text{bias}}$$

(this shows the bias is ≤ 0; specific sample?)

We can't compute the MSE but we can estimate it:

plugged-in θ		$E(\tilde{\Theta})$	$MSE(\tilde{\Theta})$
$\tilde{\theta} = \sqrt{x_3/n}$	$\simeq 0.4263$	$\simeq 0.4261$	$\simeq 1.996\times10^{-4}$
$\hat{\theta} = \frac{x_2+2x_3}{2n}$	$\simeq 0.4247$	$\simeq 0.4245$	$\simeq 1.998\times10^{-4}$

$$\Rightarrow \frac{MSE(\tilde{\Theta})}{MSE(\hat{\Theta})} \underset{\text{data}}{\overset{\text{our}}{\sim}} \begin{cases} 1.6769 & \theta = \tilde{\theta} = \sqrt{x_3/n} \\ 1.6814 & \theta = \hat{\theta} = \frac{x_2+2x_3}{2n} \end{cases}$$

which estimator is better? What if our θ-estimation is off?

Suppose we can show that for any $n \in \mathbb{N}$ and $\theta \in [0,1]$:

$$\frac{MSE(\tilde{\Theta}_n)}{MSE(\hat{\Theta}_n)} \geq 3/2 \qquad (1)$$

In this case,

$$MSE(\hat{\Theta}_{2/3 n}) = \frac{\theta(1-\theta)}{2(2/3 n)}$$

$$= \frac{3}{2} \frac{\theta(1-\theta)}{2n}$$

$$= \frac{3}{2} MSE(\hat{\Theta}_n)$$

$$\leq MSE(\tilde{\Theta}_n)$$

<u>Cor.</u> Using 2/3 of the sample size, $\hat{\Theta}$ has, on average, a smaller error than $\tilde{\Theta}$! It can reduce the error without increasing the sample size.

That is, $\hat{\Theta}$ is a more *efficient* estimator.

Unfortunately, we can't exactly show (1) and, in fact, it is false when $\theta \approx 1$: it isn't hard to show that $\tilde{\Theta} \approx \hat{\Theta}$ in that case.

We can however show that (1) holds for any $\theta \in (0, 1/2)$ and any "sufficiently large" n.