Evaluating the Performance of Estimators (Section 7.3)

Example: Suppose we observe $X_1, \ldots, X_n$ iid $N(\theta, \sigma_0^2)$, with σ_0^2 known, and wish to estimate θ.

Two possible estimators are:

$$\widehat{\theta} = \bar{X} \equiv \text{sample mean} \quad \text{and} \quad \widehat{\theta} = M \equiv \text{sample median}\,.$$

Which is better? How to measure performance?

Some possibilities:

1. Compare $E|\bar{X} - \theta|$ with $E|M - \theta|$.
2. Compare $E(\bar{X} - \theta)^2$ with $E(M - \theta)^2$.
3. Compare $EL(\theta, \bar{X})$ with $EL(\theta, M)$
   where $L(\cdot, \cdot)$ is an appropriate "loss function": the value $L(\theta, a)$ is some measure of the loss incurred when the true value is θ and our estimate is a.

absolute error loss:	$L(\theta, a) = \lvert a - \theta \rvert$
squared error loss:	$L(\theta, a) = (a - \theta)^2$
game show loss:	$L(\theta, a) = I(\lvert a - \theta \rvert > c)$
Stein's loss (for $\theta, a > 0$):	$L(\theta, a) = \dfrac{a}{\theta} - 1 - \log\left(\dfrac{a}{\theta}\right)$

Historically, estimators have been most frequently compared using Mean Squared Error: $\text{MSE}(\theta) = E_\theta(\widehat{\theta} - \theta)^2$.

This is because the MSE can often be calculated or approximated (for large samples), and has nice mathematical properties.

Admissible and Inadmissible Estimators

Let $W = W(\boldsymbol{X})$ be an estimator of $\tau = \tau(\theta)$.

Define $\text{MSE}_W(\theta) = E_\theta(W - \tau(\theta)^2$.

An estimator W is **inadmissible** (w.r.t. squared error loss) if there exists another estimator $V = V(\boldsymbol{X})$ such that

$$MSE_V(\theta) \leq MSE_W(\theta) \quad \forall \theta \in \Theta$$

with strict inequality for at least one value of θ. (An estimator is inadmissible if there is another estimator that "beats" it.) An estimator which is **not inadmissible** is called **admissible**.

(Draw some pictures.)

Note: An admissible estimator may actually be very bad. An inadmissible estimator can sometimes be pretty good.

Note: If we are using a loss function $L(\tau, a)$, we also define inadmissible and admissible estimators in the same way, replacing the MSE by the more general notion of a **risk function** $R(\theta, W) = E_\theta L\left(\tau(\theta), W(\boldsymbol{X})\right)$.

Examples: Again, suppose we observe $X_1, \ldots, X_n$ iid $N(\theta, \sigma_0^2)$, with σ_0^2 known, and wish to estimate θ.

Consider the estimator $W \equiv 0$ that always estimates θ by 0 regardless of the data $\boldsymbol{X}$.

This is a very bad estimator, but it is admissible because it is great when $\theta = 0$. No non-degenerate estimator V can possibly beat W since it would have to satisfy

$$\begin{array}{rrcl} & MSE_V(0) & \leq & MSE_W(0) \\ \Rightarrow & E_0(V-0)^2 & \leq & E_0(W-0)^2 \\ \Rightarrow & E_0V^2 & \leq & 0 \quad \Rightarrow \quad P_0(V=0)=1 \\ \Rightarrow & V & \equiv & 0\,. \end{array}$$

Now consider the estimator $M \equiv$ sample median.

We show later that the sample mean $\bar{X}$ has a uniformly smaller MSE than M so that M is inadmissible. (The two MSE functions are constant, i.e., flat.)

However, M is not a bad estimate of θ, and might be used if there were doubts about the normality assumption (perhaps the true distribution has thicker tails) or concern about outliers.

Bias, Variance, and MSE (for an estimator W of $\tau(\theta)$)

$$\begin{aligned} \text{Bias}_W(\theta) &= E_\theta(W - \tau(\theta)) \\ \text{Var}_W(\theta) &= E_\theta(W - E_\theta W)^2 \equiv \text{Var}_\theta(W) \end{aligned}$$

$$\begin{aligned} \textbf{Fact:} \quad \text{MSE}_W(\theta) &= \text{Bias}_W^2(\theta) + \text{Var}_W(\theta) \\ \text{MSE} &= \text{Bias}^2 + \text{Var} \quad \text{(in brief)} \end{aligned}$$

Proof: For any rv Y with finite second moment, we know

$$EY^2 = (EY)^2 + \text{Var}(Y)\,.$$

Taking $Y = W - \tau$ leads to

$$\text{MSE} = \text{Bias}^2 + \text{Var}$$

since $\text{Var}(W - \tau) = \text{Var}(W)$.

Terminology: An estimator W with $\text{Bias}_W(\theta) \equiv 0$, that is,

$$E_\theta W = \tau(\theta) \quad \forall\, \theta \in \Theta$$

is said to be **unbiased**. If not, it is **biased**.

For an unbiased estimator, MSE = Var.

Example: Coin-tossing

$X_1, \ldots, X_n$ iid Bernoulli(θ)

$$\hat{\theta}_{MLE} = \frac{T}{n}, \quad T = \sum X_i.$$

$$\hat{\theta}_{Bayes} = (1-p)\, a + p\left(\frac{T}{n}\right)$$

$\frac{\alpha+\beta}{\alpha+\beta+n}$ ↑ prior mean $\frac{\alpha}{\alpha+\beta}$ ↑ $\frac{n}{\alpha+\beta+n}$

$$MSE_{MLE}(\theta) = Bias^2 + Variance$$

$$= 0 + \frac{\theta(1-\theta)}{n}$$

$$MSE_{Bayes}(\theta) = Bias^2 + Variance$$

↓

$$\left[(1-p)\, a + \underbrace{p\theta - \theta}_{-(1-p)\theta}\right]^2 + p^2 \frac{\theta(1-\theta)}{n}$$

$$(1-p)^2 (a-\theta)^2 + p^2 \frac{\theta(1-\theta)}{n}$$

Which is better? (according to MSE)

Answer: Neither dominates the other.

$$\mathrm{MSE}_{\mathrm{Bayes}}(a) = \frac{p^2\theta(1-\theta)}{n} < \frac{\theta(1-\theta)}{n} = \mathrm{MSE}_{\mathrm{MLE}}(a)\,,$$

$$\mathrm{MSE}_{\mathrm{Bayes}}(0) = (1-p)^2a^2 > 0 = \mathrm{MSE}_{\mathrm{MLE}}(0)\,,$$

and similarly, $\mathrm{MSE}_{\mathrm{Bayes}}(1) > \mathrm{MSE}_{\mathrm{MLE}}(1)$.

Thus, the Bayes estimate is superior in the neighborhood of $\theta = a$, and the MLE is superior near $\theta = 0$ and 1.

Note: both $\mathrm{MSE}_{\mathrm{Bayes}}(\theta)$ and $\mathrm{MSE}_{\mathrm{MLE}}(\theta)$ are parabolas (quadratic functions of θ).

Note: Regarding (in)admissibility, the above remarks prove nothing. But it can be shown that both the Bayes estimate and MLE are admissible here. Typically, Bayes estimates (with proper priors) are admissible.

Example: Estimating the Variance

$X_1, \ldots, X_n$ iid $N(\underbrace{\mu, \sigma^2}_{\theta})$

$\tau(\theta) = \sigma^2$

$$W = c \sum_{i=1}^{n} (X_i - \bar{X})^2 = c\, SS$$

$c = \frac{1}{n-1}$ usual unbiased estimator

$c = \frac{1}{n}$ is MLE

Which is better?

$$\frac{SS}{\sigma^2} \sim \chi^2_{n-1} \rightarrow \begin{array}{l} \text{mean} = n-1 \\ \text{var} = 2(n-1) \end{array}$$

Thus $E\, cSS = c\sigma^2(n-1)$

$\text{Var}\, cSS = c^2\sigma^4 \cdot 2(n-1)$

$$MSE_{cSS}(\mu, \sigma^2) = E(cSS - \sigma^2)^2$$

$= \text{Bias}^2 + \text{Variance}$

$= (c\sigma^2(n-1) - \sigma^2)^2 + c^2\sigma^4 \cdot 2(n-1)$

$= \sigma^4\left((c(n-1)-1)^2 + c^2 2(n-1) \right)$

$$MSE_{cSS}(\mu, \sigma^2) = \sigma^4 \psi(c)$$

with $\psi(c) = (c(n-1)-1)^2 + 2(n-1)c^2$

$$\psi'(c) = 2(c(n-1)-1)(n-1) + 4(n-1)c$$
$$= 2(n-1)[c(n+1)-1]$$

$$\psi''(c) = 2(n-1)(n+1) > 0$$
for $n \geq 2$

$\psi'(c) = 0$ when $c = \frac{1}{n+1}$.

Both $c = \frac{1}{n-1}$ and $c = \frac{1}{n}$ give inadmissible estimators

$c = \frac{1}{n-1}$ is the best c with

Stein's loss function: (page 351)

$$L(\sigma^2; a) = \frac{a}{\sigma^2} - \log\frac{a}{\sigma^2} - 1$$

Other Plausible loss functions

$$L(\sigma^2; a) = \frac{\sigma^2}{a} - \log\frac{\sigma^2}{a} - 1$$

$$= \frac{\sigma^2}{a} + \frac{a}{\sigma^2} - 2$$

$$= (\log a - \log \sigma^2)^2$$

What is the best c for these?

Comments: (state earlier)

Estimator with best c might not be admissible!

MSE inappropriate (or dubious, anyway) for estimation of σ^2.

The Rao-Blackwell Theorem

If $T = T(X)$ is a sufficient statistic for θ,

$$E_\theta S(X) = \tau(\theta) \text{ for all } \theta, \text{ and}$$

$$E_\theta (S(X) - \tau(\theta))^2 < \infty \text{ for all } \theta,$$

then $S^*(X) = E(S(X) \mid T(X))$ satisfies

$$E_\theta S^*(X) = \tau(\theta) \text{ for all } \theta, \text{ and}$$

$$E_\theta (S^*(X) - \tau(\theta))^2 \leq E_\theta (S(X) - \tau(\theta))^2 \text{ for all } \theta.$$

Notes:

① $E(S(X)|T(X))$ does <u>not</u> depend on θ because $T(X)$ is sufficient so that $\mathcal{L}(X|T)$ (and thus $\mathcal{L}(S|T)$) does <u>not</u> depend on θ.

② Equality of MSE's for a particular θ can occur iff $P_\theta(S(X) = S^*(X)) = 1$.

③ $S^*(X)$ is a function of $T(X)$ (i.e., $\exists\, \psi$ such that $S^*(X) = \psi(T(X))$).

Proof:

Recall: For any rv's X, Y with $EX^2 < \infty$, we have

$$EX = E(E(X \mid Y))$$

$$\mathrm{Var}(X) = E(\mathrm{Var}(X \mid Y)) + \mathrm{Var}(E(X \mid Y))$$

Now apply these facts:

$$\begin{aligned} E_\theta[S^*(\boldsymbol{X})] &= E_\theta[E_\theta(S(\boldsymbol{X}) \mid T(\boldsymbol{X}))] \\ &= E_\theta\, S(\boldsymbol{X}) = \tau(\theta) \end{aligned}$$

$$\begin{aligned} E_\theta(S-\tau)^2 &= \mathrm{Var}_\theta(S) \\ &= \underbrace{E[\mathrm{Var}(S \mid T)]}_{\geq 0} + \mathrm{Var}[\underbrace{E(S \mid T)}_{S^*}] \\ &\geq \mathrm{Var}(S^*) = E_\theta(S^* - \tau(\theta))^2 \end{aligned}$$

Equality can occur only when $E_\theta\, \mathrm{Var}(S \mid T) = 0$. But

$$\begin{aligned} E_\theta \mathrm{Var}(S \mid T) &= E_\theta\{E[(S - E(S \mid T))^2 \mid T]\} \\ &= E_\theta\{(S - E(S \mid T))^2\} = E_\theta\{(S - S^*)^2\} \\ &= 0 \quad \text{iff } P_\theta(S = S^*) = 1. \end{aligned}$$

Arguing more loosely, $E_\theta\, \mathrm{Var}(S \mid T) = 0 \;\Rightarrow\; \mathrm{Var}(S \mid T) = 0 \;\Rightarrow$ S is a function of $T \;\Rightarrow\; S^* \equiv E(S \mid T) = S$.

Example: $X_1, X_2, \ldots, X_n$ iid Bernoulli(p)
$T = \sum X_i$ is suff. stat. for p

$\mathcal{L}(X|T)$ puts equal prob. of $\frac{1}{\binom{n}{T}}$ on all strings with T 1's and $n-T$ 0's.

Generate from $\mathcal{L}(X|T=t)$ by placing t 1's and $n-t$ 0's in an urn, and randomly drawing (without replacement) until the urn is empty.

① Estimation of p.

$EX_1 = p$ for all p so that $S = X_1$ is an unbiased estimator of p.

X_1 is not a function of T, so it can be improved by conditioning (as in the Rao-Blackwell Theorem).

$$S^* = E(S|T) = E(X_1|T) = P(X_1 = 1|T) = \frac{T}{n}$$

$E(X_1|T)$: indicator rv, 0-1 valued

$S^* = T/n$ is the usual estimator (the sample proportion)!

Clearly

$$E\,S^* = p \quad \forall p$$

$$\mathrm{Var}(S^*) = p(1-p)/n < \mathrm{Var}\,S = p(1-p)$$

verifying the conclusions of the R-B Thm. $\forall p$,

(2) Estimation of p^2.

Note that $E\,\underbrace{X_1X_2}_{\text{indicator}} = P(X_1X_2=1)$

$$= P(X_1=X_2=1) = p^2.$$

Thus $S = X_1X_2$ is an unbiased estimator of p^2.

It is <u>not</u> a function of T, so it can be improved by conditioning.

$$S^* = E(X_1X_2 \mid T) = P(X_1X_2=1 \mid T)$$

$$= P(X_1=X_2=1 \mid T) = \frac{T}{n} \cdot \frac{T-1}{n-1}$$

By R-B Thm, S^* is an unbiased estimate of p^2 with smaller variance than S. This can be verified by straightforward calculations.

For comparison, what is the MLE of p^2?

The MLE of p is T/n, so the invariance principle for MLE's says the MLE of p^2 is $\left(\frac{T}{n}\right)^2$.

Clearly $\frac{T}{n} \cdot \frac{T-1}{n-1}$ and $\left(\frac{T}{n}\right)^2$ are very close when n is large.

Which is better? Neither dominates.

$\left(\frac{T}{n}\right)^2$ is biased, but the bias is negligible for large n.

Estimation of p^3, etc.

$S = X_1 X_2 X_3$ is unbiased for p^3

$$S^* = E(S|T) = P(X_1 = X_2 = X_3 = 1 | T)$$

$$= \frac{T}{n} \cdot \frac{T-1}{n-1} \cdot \frac{T-2}{n-2}$$

is the Rao-Blackwell improvement on S.

The pattern is now clear for p^4, etc.

Suppose $T = T(X)$ is a **complete** and sufficient statistic for θ. Then ...

(1) For any parameter $\tau(\theta)$, there is at most one unbiased estimator which is a function of T.

(2) If $S = S(X)$ is unbiased for $\tau(\theta)$ (and $\text{Var}(S) < \infty$ for all θ), then

$$S^*(X) = S^* = E(S \mid T)$$

is the UMVUE for $\tau(\theta)$.

Definition: $S = S(X)$ is the UMVUE (uniformly minimum variance unbiased estimator) for $\tau(\theta)$ if

$$E_\theta S = \tau(\theta) \quad \text{for all } \theta,$$

and

$$\text{Var}_\theta(S) \leq \text{Var}_\theta(S') \quad \text{for all } \theta$$

for any other unbiased estimator S'.

Note: For unbiased estimators, MSE = Variance.

Terminology: UMVUE = "best unbiased estimator"

(3) An unbiased estimator (with finite variance) which is a function of T is the UMVUE.

Proof of ①:

Suppose $S_1(X) = \psi_1(T(X))$ $[S_1 = \psi_1(T)]$

$S_2(X) = \psi_2(T(X))$ $[S_2 = \psi_2(T)]$

and $ES_1 = ES_2 = \tau(\theta)$ $\forall \theta$.

Then $E(S_1 - S_2) = 0$ $\forall \theta$

or $E_\theta\, g(T) = 0$ $\forall \theta$ where

$g(t) = \psi_1(t) - \psi_2(t)$.

By completeness

$P_\theta(g(T) = 0) = P_\theta(S_1 = S_2) = 1$ $\forall \theta$.

Thus $S_1 = S_2$ (with prob 1 for all θ).

Proof of ②: S^* is unbiased and a fn. of T.

Suppose W is <u>any</u> unbiased estimator for $\tau(\theta)$.

Then R-B Thm says that $W^* = E(W \mid T)$ is an unbiased estimator of $\tau(\theta)$ and

$\mathrm{Var}_\theta\, W^* \le \mathrm{Var}_\theta\, W$ for all θ. (†)

But W^* is unbiased and a function of T, so ① implies $W^* = S^*$.

Thus (†) implies

$$\mathrm{Var}_\theta S^* \le \mathrm{Var}_\theta W \text{ for all } \theta,$$

and S^* is the UMVUE.

Proof of ③:

Suppose $E_\theta S(X) = \tau(\theta)\ \forall \theta$

and $S(X) = \psi(T(X))$.

Then $S^* = E(S|T)$ is the UMVUE by ②.

But S is a function of T so that

$E(S|T) = S$.

Thus S is the UMVUE.

Example: Observe $X_1, \ldots, X_n$ iid Bernoulli(p).

• Find the UMVUE of p.

$T = \sum_i X_i$ is a CSS. $\quad E(T/n) = p$.

Since T/n is an unbiased estimator of p which is a function of the CSS T, it is the UMVUE.

• Find the best unbiased estimator of p^2.

$$E\left(\frac{T(T-1)}{n(n-1)}\right) = p^2 \quad \left(\begin{array}{l}\text{Found indirectly using Rao-Blackwell.}\\ \text{For a direct argument, see below.}\end{array}\right)$$

Since $\left(\frac{T(T-1)}{n(n-1)}\right)$ is an unbiased estimator of p^2 which is a function of the CSS T, it is the UMVUE.

Checking unbiasedness:

$$\begin{aligned} ET(T-1) &= E(T^2) - ET = \text{Var}(T) + (ET)^2 - ET \\ &= np(1-p) + (np)^2 - np = n(n-1)p^2 \end{aligned}$$

Comment: "Estimate a parameter by its UMVUE" is another approach to estimation, but **not** a very good one. Often, no unbiased estimator exists, or the only one that exists is bad.

Example: Observe $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ with $\theta = (\mu, \sigma^2)$ unknown.

Here $T = (\bar{x}, s^2)$ is a CSS. (Recall the derivation: T is a 1-1 function of the natural SS for a 2pef.)

- **Estimation of $\tau(\mu, \sigma^2) = \mu$:**

$\bar{x}$ is unbiased ($E\bar{x} = \mu$) and a function of $T \Rightarrow \bar{x}$ is UMVUE.

MLE of θ is $\hat{\theta} = \left(\bar{x}, n^{-1}\sum_i (X_i - \bar{x})^2\right)$. So invariance principle says MLE of μ is $\tau(\hat{\theta}) = \bar{x}$.

MOM estimate is also $\bar{x}$ since $E\bar{x} = \mu$.

Note: For estimating μ, the MLE, MOM, UMVUE all agree on $\bar{X}$. But Bayes estimate is different.

What about the sample median M?

M is an unbiased estimator of μ. (Proof?) But it is **not** a function of the CSS T. Thus Rao-Blackwellizing M leads to the UMVUE (which we know is $\bar{x}$) which has a strictly smaller variance than M.

Thus: $E(M \,|\, T) = \bar{x}$ and $\text{Var}(M) > \text{Var}(\bar{x}) = \sigma^2/n$.

- **Estimating $\tau(\mu, \sigma^2) = \sigma^2$:**

Let $SS = \sum_i (X_i - \bar{x})^2$.

$s^2 = SS/(n-1)$ is an unbiased estimator of σ^2 and a function of the CSS T. Therefore s^2 is the UMVUE.

By the invariance principle, the MLE of σ^2 is SS/n. This is slightly biased.

• **Estimation of $\tau(\mu, \sigma^2) = \mu^2$:**

The MLE of μ^2 is $(\bar{x})^2$ by invariance of MLE's.

$\bar{x}^2$ is biased for μ^2:

$$E(\bar{x}^2) = \text{Var}(\bar{x}) + (E\bar{x})^2 = \frac{\sigma^2}{n} + \mu^2 > \mu^2 \,.$$

An unbiased estimate of μ^2 is $W \equiv \bar{x}^2 - \dfrac{s^2}{n}$:

$$E\left(\bar{x}^2 - \frac{s^2}{n}\right) = \left(\frac{\sigma^2}{n} + \mu^2\right) - \frac{\sigma^2}{n} = \mu^2.$$

Subtracting s^2/n removes (or corrects for) the bias in the MLE.

W is the UMVUE since it is unbiased and a function of T.

Which is better: $\bar{x}^2$ or W?

For $n > 3$, W has slightly smaller MSE than $\bar{x}^2$. (Verify?)

Thus $\bar{x}^2$ is inadmissible for $n > 3$ (but is a perfectly reasonable estimator).

But W is also inadmissible because it sometimes takes on "impossible" values.

$\mu^2 \geq 0$, but W can be negative!

$P(W < 0)$ is positive and will be sizeable when μ is small ($\approx 1/2$ when $\mu = 0$).

A better estimate is clearly $W_+ = \max(W, 0)$.

Whenever $W_+ \neq W$, we know W_+ is closer to the true value of μ^2. More formally

$$E(W - \mu^2)^2 - E(W_+ - \mu^2)^2 = E\left[(W - \mu^2)^2 - (W_+ - \mu^2)^2\right]$$
$$= E[\underbrace{\{(W - \mu^2)^2 - (0 - \mu^2)^2\} I(W < 0)}_{\text{always} \geq 0 \text{ and sometimes} > 0}] > 0$$

But W_+ is biased! Oh, well.

No unbiased estimator of μ^2 exists which does **not** take on negative values.

Fact: There are situations where there are **no** unbiased estimators (and hence, no UMVUE exists).

Example: Observe $X_1, \ldots, X_n$ iid Poisson(λ). There exists no unbiased estimator of $1/\lambda$.