7.3 Methods of Evaluating Estimators

Definition 7.3.1

The mean square error (MSE) of an estimator W of a parameter θ is the function $E[(W-\theta)^2]$

We have.

$$E[(W-\theta)^2] = E[(W - E[W] + E[W] - \theta)^2]$$

$$= E[(W - E[W])^2] + E[(E[W]-\theta)^2] + 2E[(W-E[W])(E[W]-\theta)]$$

$$= Var[W] + [Bias\, W]^2$$

Definition 7.3.2

The bias of a point estimator W of a parameter θ is defined as $Bias\, W = E[W] - \theta$. If $E[W] = \theta$, $\forall \theta$, W is an unbiased estimator.

For unbiased estimators $MSE = Var[W]$

Example

$$X_1, \dots, X_n \quad iid \quad N(\mu, \sigma^2)$$

$$\bar{X}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}, \qquad MSE(\bar{X}) = \frac{\sigma^2}{n}$$

$$\hat{\sigma}^2 = \frac{n-1}{n} S^2 = \frac{\sigma^2}{n}\left(\frac{(n-1)S^2}{\sigma^2}\right) \Rightarrow E[\hat{\sigma}^2] = \frac{\sigma^2}{n}(n-1), \quad Var[\hat{\sigma}^2] = \frac{\sigma^4 2(n-1)}{n^2}$$

$$MSE[\hat{\sigma}^2] = \frac{2(n-1)\sigma^4}{n^2} + \Big(\underbrace{\frac{1}{n}\sigma^2}_{Bias\, \hat{\sigma}^2}\Big)^2 = \frac{2n-1}{n^2}\sigma^4 < Var[S^2] = \frac{2\sigma^4}{n-1}$$

General problem with MSE is that it depends on θ

Example 7.3.5

MSE of a binomial Bayes estimator.

$X_1, \dots, X_n$ Bernoulli(p), $\hat{p} = \bar{X} = \frac{Y}{n}$ is unbiased ($Y = \sum_{i=1}^{n} X_i$)

and $MSE(p) = Var(\hat{p}) = \frac{p(1-p)}{n}$.

Let $\hat{p}_B$ be the Bayes estimator, $\hat{p}_B = \frac{Y+\alpha}{\alpha+\beta+n}$

$$E[(\hat{p}_B - p)^2] = Var[\hat{p}_B] + (\text{Bias } \hat{p}_B)^2$$

$$= Var\left[\frac{Y+\alpha}{\alpha+\beta+n}\right] + \left(E\left[\frac{Y+\alpha}{\alpha+\beta+n}\right] - p\right)^2$$

$$= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2$$

$$= \frac{1}{(\alpha+\beta+n)^2}\left\{np(1-p) + (\alpha - p(\alpha+\beta))^2\right\}$$

Without knowledge of p it is reasonable to try to make it a constant.

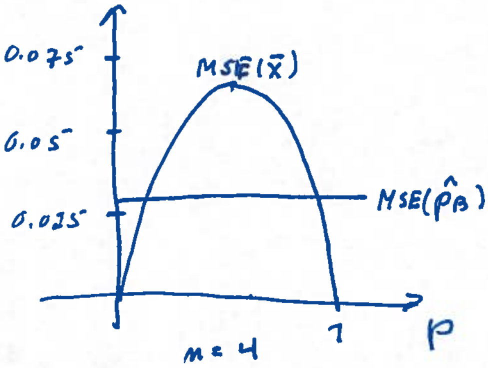
$$\frac{d\,MSE}{dp} = 0 \iff n - 2np - 2(\alpha - p(\alpha+\beta))(\alpha+\beta) = 0$$

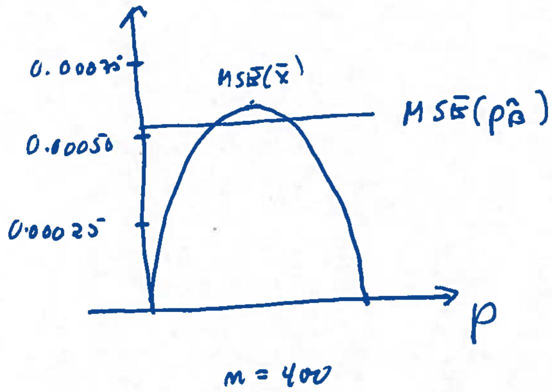
$$\iff n - 2\alpha^2 - 2\alpha\beta + 2p((\alpha+\beta)^2 - n) = 0$$

Satisfied if $\left.\begin{matrix} 2\alpha(\alpha+\beta) = n \\ (\alpha+\beta)^2 = n \end{matrix}\right\}$ satisfied if $\alpha = \beta = \sqrt{\frac{n}{4}}$

This gives $\hat{p}_B = \frac{Y + \sqrt{\frac{n}{4}}}{n + \sqrt{n}}$ and $E[(\hat{p}_B - p)^2]$

$$= \frac{np(1-p)}{(n+\sqrt{n})^2} + \frac{\left(\frac{\sqrt{n}}{2} - p\sqrt{n}\right)^2}{(n+\sqrt{n})^2} = \frac{np - np^2 + \frac{n}{4} - np + p^2 n}{(n+\sqrt{n})^2} = \frac{n}{4(n+\sqrt{n})^2}.$$





7.3.2 Best unbiased estimators

Problem. There is no one „best MSE".

The class of estimators is too large.

Example. $X \sim Po(\lambda)$

$E[\bar{X}] = \lambda$ $\quad \hat{\lambda} = \bar{X}$

$E[S^2] = \lambda$ $\quad \hat{\lambda} = S^2$ } unbiased and so are $a\bar{X} + (1-a)S^2$

~~for all (a,1)~~

Definition 7.3.7

An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E[W^*] = \tau(\theta)$, $\forall \theta$ and for any other estimator W with $E[W] = \tau(\theta)$, $Var[W^*] \leq Var[W]$. W^* is also called a uniform minimum variance unbiased estimator (UMVUE).

Problem. The class of unbiased estimators is also very large. Try to construct a lower bound for the variance such that if it is obtained, we have found the best unbiased estimator.

Let $\underline{X} = X_1, \dots X_n$ be a sample with likelihood

$$L(\theta|\underline{x}) = f(\underline{x}|\theta) \quad \text{that is differentiable.}$$

Define $S(\underline{X}|\theta) = \frac{d}{d\theta} \log L(\theta|\underline{X})$ and $I(\theta) = Var[S(\underline{X}|\theta)]$

$S(\underline{X}|\theta)$ is called the score statistic and $I(\theta)$ is called the Fisher information.

Example $f(x|\theta) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$, $-\infty < X < \infty$

$$\ell(\theta|x) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(x-\theta)^2$$

$$S(X|\theta) = \frac{X-\theta}{\sigma^2}, \quad Var[S(X|\theta)] = \frac{1}{\sigma^2}$$

We have

$$E[S(\underline{X}|\theta)] = 0$$

Proof. $$\frac{d}{d\theta}\log L(\theta|\underline{x}) = \frac{d}{d\theta}\log f(\underline{x}|\theta) = \frac{\frac{d}{d\theta} f(\underline{x}|\theta)}{f(\underline{x}|\theta)}$$

Hence $$E[S(\underline{X}|\theta)] = \int_{\underline{x}} \frac{d}{d\theta}\log f(\underline{x}|\theta)\, f(\underline{x}|\theta)\, d\underline{x} = \int_{\underline{x}} \frac{d}{d\theta} f(\underline{x}|\theta)\, d\underline{x}$$

$$\overset{\text{if allowed}}{=} \frac{d}{d\theta}\int_{\underline{x}} f(\underline{x}|\theta)\, d\underline{x} = \frac{d}{d\theta} 1 = 0$$

Theorem 7.3.9. Cramer-Rao inequality

Let $X_1, \dots, X_n$ be a sample with pdf $f(\underline{x}|\theta)$ and let $W(\underline{X})$ be an estimator satisfying.

$$\frac{d}{d\theta} E[W(\underline{X})] = \int_{\underline{x}} \frac{d}{d\theta}[W(\underline{x}) f(\underline{x}|\theta)]\, d\underline{x}$$

and let $Var\, W(\underline{X}) < \infty$

Then
$$Var[W(\underline{X})] \geq \frac{\left(\frac{d}{d\theta} E[W(\underline{X})]\right)^2}{E\left[\left(\frac{d}{d\theta}\log f(\underline{x}|\theta)\right)^2\right]} = \frac{\left(\frac{d}{d\theta} E[W(\underline{X})]\right)^2}{I(\theta)} = \frac{(\tau'(\theta))^2}{I(\theta)}$$

where $E[W(\underline{X})] = \tau(\theta)$, with equality if and only if

$S(\underline{X}|\theta) = a(\theta)[W(\underline{X}) - \tau(\theta)]$

Proof. Let $E[W(\underline{X})] = \tau(\theta)$

Then $\tau(\theta) = E[W(\underline{X})] = \int_{\underline{x}} W(\underline{x}) f(\underline{x}|\theta)\, d\underline{x}$

and $\tau'(\theta) = \int_{\underline{x}} W(\underline{x}) \left[\frac{d}{d\theta} f(\underline{x}|\theta)\right] d\underline{x} = \int_{\underline{x}} W(\underline{x}) \left[\frac{d}{d\theta} \log f(\underline{x}|\theta)\right] f(\underline{x}|\theta)\, d\underline{x}$

$= E[W(\underline{X}) \cdot S(\underline{X}|\theta)] = Cov[W(\underline{X}), S(\underline{X}|\theta)]$

Cauchy-Schwartz: $\frac{|Cov(X,Y)|}{SD(X)\cdot SD(Y)} \leq 1 \Leftrightarrow Cov(X,Y)^2 \leq Var[X]\cdot Var[Y]$

Thereby $(\tau'(\theta))^2 \leq Var[W(\underline{X})] \cdot Var[S(\underline{X}|\theta)]$

$\Rightarrow Var[W(\underline{X})] \geq \frac{(\tau'(\theta))^2}{I(\theta)}$

Further. $|Cov(X,Y)| = SD(Y)\cdot SD(X) \Leftrightarrow Y = aX + b$. Hence to have equality we must have $S(\underline{X}|\theta) = a(\theta) W(\underline{X}) + b(\theta)$ for arbitrary functions $a(\theta)$ and $b(\theta)$. Further $E[S(\underline{X}|\theta)] = 0$

$\Rightarrow -a(\theta)\tau(\theta) = b(\theta) \Rightarrow S(\underline{X}|\theta) = a(\theta)[W(\underline{X}) - \tau(\theta)]$