$$E[\underbrace{L(\underline{W(X)}_{\text{estimator}}-\theta)}_{\text{loss function}}] = \int L(w(x)-\theta)\, f_X(x|\theta)\,dx$$
$$\equiv R(\theta, W(\cdot))$$

Defn: Risk is the expected loss from using the statistic $W(X)$ to estimate θ.

Mean square error

Defn: The mean square error (MSE) of an estimator W for θ is the function of θ given by
$E_\theta[(W(X)-\theta)^2]$.

eg Let
- $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $i=1,\dots,n$
- $W(X) = \bar{X}$ $\qquad \theta = \mu$
- $E[(W(X)-\theta)^2] = E[(\bar{X}-\mu)^2] = Var(\bar{X}) = \frac{\sigma^2}{n}$

Remarks:

1. Expectation is taken with respect to X.
2. E_θ reminds us what density we use to evaluate the expectation. (ie "true" θ)
3. Expanding MSE:

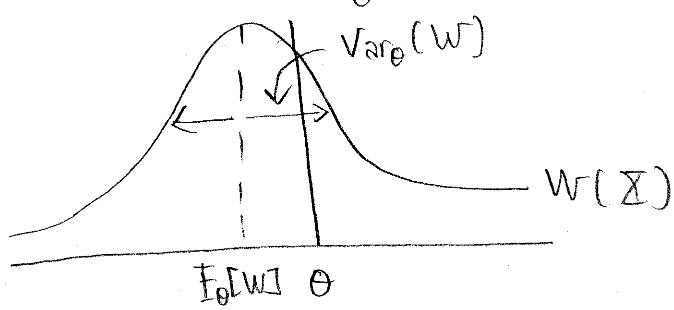
$$E_\theta[(W(X)-\theta)^2] = E_\theta[(W - E_\theta[W] + \underbrace{E_\theta[W]}_{\text{not random}} - \theta)^2]$$
$$= E_\theta[(W - E_\theta[W])^2] + E_\theta[(E_\theta[W]-\theta)^2]$$
$$+ 2E_\theta[(W-E_\theta[W])(E_\theta[W]-\theta)]$$

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

$$= 2(E_\theta[W]-\theta)E_\theta[(W-E_\theta[W])]$$
$$= 2(E_\theta[W]-\theta)(\underbrace{E_\theta[W]-E_\theta[W]}_{=0})$$
$$=0$$

$$\Rightarrow E_\theta[(W(\underline{X})-\theta)^2] = E_\theta[(W(\underline{X})-E_\theta[W(\underline{X})])^2] + (E_\theta[W(\underline{X})]-\theta)^2$$
$$= Var_\theta(W(\underline{X})) + (\text{bias})^2$$

Defn 7.3.2: The bias of an estimator $W(\underline{X})$ for θ is $E_\theta[W(\underline{X})]-\theta$.



- Bigger bias $\Rightarrow$ need smaller variance to compensate.

- If we restrict attention to unbiased estimators, we have MSE = Var, so all we need to do is focus on the variance.

eg 7.3.4: Let $X_1, \dots, X_n$ be iid $N(\mu, \sigma^2)$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$
- standard (unbiased) estimator
  - $E[S^2]=\sigma^2$

$$\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$
- MLE for σ^2

$$E[\bar{s}^2] = E\left[\frac{n-1}{n} s^2\right] = \frac{n-1}{n} E[s^2] = \frac{n-1}{n}\sigma^2$$

- biased towards zero

$$MSE(s^2) = Var(s^2) + [E[s^2] - \sigma^2]^2$$
$$= Var(s^2) = \frac{2\sigma^4}{n-1}$$

$$MSE(\bar{s}^2) = Var(\bar{s}^2) + (E[\bar{s}^2] - \sigma^2)^2$$
$$= Var\left(\frac{n-1}{n} s^2\right) + \left(-\frac{1}{n}\sigma^2\right)^2$$
$$= \left(\frac{n-1}{n}\right)^2 \cdot \frac{2\sigma^4}{n-1} + \frac{\sigma^4}{n}$$
$$= \frac{(n-1)2\sigma^4 + n\sigma^4}{n} = \frac{\sigma^4}{n^2}(2n-1)$$

$$MSE(s^2) - MSE(\bar{s}^2) = \sigma^4\left(\frac{2}{n-1} - \frac{2n-1}{n^2}\right)$$
$$= \sigma^4\left(\frac{2n^2 - (2n-1)(n-1)}{n^2(n-1)}\right)$$
$$= \sigma^4\left(\frac{2n^2 - 2n^2 + 2n + n - 1}{n^2(n-1)}\right)$$
$$= \sigma^4\left(\frac{3n-1}{n^2(n-1)}\right) > 0 \quad \forall n$$

Thus, $\bar{s}^2$ always has a smaller MSE than s^2.

"Two frequentists shoot at a deer. One hits a tree ten feet to the right of the deer and the other hits a tree ten feet to the left. They both high five and leave."

Best unbiased estimator

Defn 7.3.7 An estimator W^* is a best unbiased estimator of $\tau(\theta)$ and for any other unbiased estimator W' of $\tau(\theta)$, $Var_\theta(W^*) \leq Var_\theta(W')$ $\forall \theta \in \Theta$.

Remarks: 1] W^* is also called uniformly minimum variance unbiased estimator. (UMVU)

2] "uniform" is uniform in $\theta \in \Theta$ (parameter space.)

3] Not particularly constructive.

How low can $Var(W)$ get? Cramer-Rao Lower Bound. (CRLB)

Key: Use Cauchy-Schwarz inequality:

$$Var(Y)\,Var(Z) \geq [Cov(Y,Z)]^2$$

$$\Leftrightarrow \quad Var(Y) \geq \frac{[Cov(Y,Z)]^2}{Var(Z)}$$

Need to choose an appropriate Z.

Let $Z = \frac{\partial}{\partial \theta} \log f_X(x|\theta)$

Thm. 7.3.7 (CRLB) Let $X_1, \dots, X_n$ be iid with pdf $f_X(x|\theta)$. Let $W(X)$ be such that $E_\theta[W(X)]$ is differentiable in X and

$$\frac{d}{d\theta}\int\cdots\int h(x) f_X(x|\theta)\,dx_1\cdots dx_n = \int\cdots\int h(x)\frac{\partial}{\partial\theta} f_X(x|\theta)\,dx_1\cdots dx_n.$$

for any $h(x)$ such that $E_\theta|h(x)| < +\infty$. Then

$$\mathrm{Var}_\theta(W(X)) \geq \frac{\left(\frac{d}{d\theta}E_\theta[W(X)]\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(x|\theta)\right)^2\right]}$$

Remarks:

1] Key assumption is interchange of limits and integrals. Rules out certain pdfs:
  in: exponential family
  out: uniform with endpoints as fcns of parameters.

2] Terminology
  $E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(x|\theta)\right)^2\right] = I(\theta)$ is referred to as the Fisher Information.

3] If $W(X)$ is unbiased, then

$$E_\theta[W(X)] = \theta \Rightarrow \frac{d}{d\theta}E_\theta[W(X)] = 1$$

$$\Rightarrow \mathrm{Var}_\theta(W(X)) \geq \frac{1}{I(\theta)}$$

4] Simplifications: If iid data, then

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_{\mathbb{X}}(x|\theta)\right)^2\right] = nE_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_{\mathbb{X}_i}(x_i|\theta)\right)^2\right]$$

5] 2 forms for $I(\theta)$.

$$\underbrace{I(\theta)}_{\text{outer product form}} = \underbrace{-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_{\mathbb{X}}(x|\theta)\right]}_{\text{Hessian form}} \text{ for}$$

exponential form.

<u>Example</u>: $\mathbb{X}_1, \ldots, \mathbb{X}_n$ iid $N(\mu, \sigma^2)$

- $\bar{\mathbb{X}}$ for μ
- s^2 for σ^2

$$f_{\mathbb{X}_i}(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}$$

$$\log f_{\mathbb{X}_i}(x_i|\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\Rightarrow \frac{\partial}{\partial\mu}\log f_{\mathbb{X}_i}(x_i|\theta) = \frac{x_i-\mu}{\sigma^2}$$

Thus, $I(\mu) = nE_\mu\left[\left(\frac{\mathbb{X}_i-\mu}{\sigma^2}\right)^2\right] = \frac{n}{\sigma^4}E[(\mathbb{X}_i-\mu)^2] = \frac{n\sigma^2}{\sigma^2} = \frac{n}{\sigma^2}$

$$\Rightarrow Var(\bar{\mathbb{X}}) \geq \frac{1}{I(\mu)} = \frac{\sigma^2}{n}$$

But $Var(\bar{\mathbb{X}}) = \frac{\sigma^2}{n} \Rightarrow \bar{\mathbb{X}}$ is best unbiased estimator.

alternatively, $I(\mu) = -n\, E_{\mu}\left[\frac{\partial^2}{\partial\theta^2}\log f_{X_i}(x_i|\theta)\right]$

$$= -n\, E_{\mu}\left[-\frac{1}{\sigma^2}\right] = \frac{n}{\sigma^2}$$

For the variance,

$$\frac{\partial}{\partial\sigma^2}\log f_{X_i}(x_i|\theta) = -\frac{1}{2\sigma^2} + \frac{(x_i-\mu)^2}{2\sigma^4}$$

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log f = \frac{1}{2\sigma^4} - \frac{(x_i-\mu)^2}{\sigma^6}$$

$$I(\sigma^2) = -n\, E\left[\frac{1}{2\sigma^4} - \frac{(x_i-\mu)^2}{\sigma^6}\right]$$

$$= -\frac{n}{2\sigma^4} + \frac{n\sigma^2}{\sigma^6} = \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} = \frac{n}{2\sigma^4}$$

$\Rightarrow$ CRLB $= \frac{2\sigma^4}{n}$, but $Var(s^2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}$.

- Lehmann + Cassella – Theory of Point Estimation
  - The bible of exact distribution stuff.