Evaluating a point estimator (Chapter 7.7)

- Given sample $D = \{X_1, X_2, \dots, X_N\}$
- Given density/PMF: $\mathrm{f}(x, \theta)$
- Let $\hat{\theta}_D$ be any estimated value of θ, example maximum likelihood estimate.
- How do we measure quality of the estimate?
  - Square difference from actual parameter.
  - Square
  - $Error(\hat{\theta}_D) = (\hat{\theta}_D - \theta)^2$

This error is a function of a specific data sample D.

Often, we want the expected square error where expectation is over all possible Ds.

Expected square error of the mean estimate

A common estimated parameter is the mean of the distribution.

$$\theta = \mu = E_f(X), \qquad \hat{\theta} = (X_1 + X_2 + \cdots + X_N)/N \leftarrow \text{sample mean}$$

- Expected square error of the above estimate $E_f\left(\sum_i \underbrace{\frac{X_i}{N}}_{\hat{\theta}} - \theta\right)^2 = \sigma^2/N$
  where $\sigma^2 = E_f(X-\mu)^2$

$X_i \perp\!\!\!\perp X_j$

$$E_f\left(\frac{\sum_i x_i - N\theta}{N}\right)^2$$

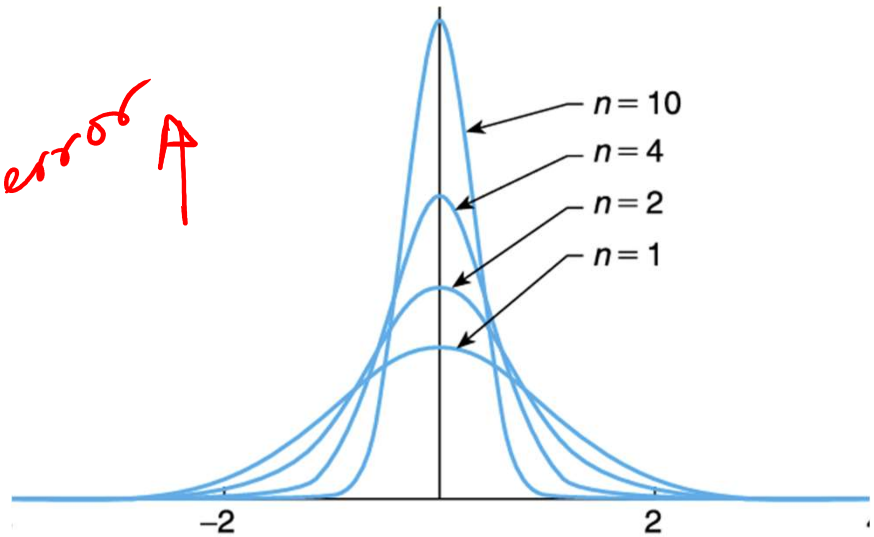
$$= \frac{1}{N^2} E_f\left[\sum_{i=1}^{N}(x_i-\theta)^2 + 2\sum_{i \neq j} E(x_i-\theta)(x_j-\theta)\right]$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \underbrace{E(x_i-\theta)^2} + 2 \sum_i E(x_i-\theta) \sum_{j \neq i} E(x_j-\theta)$$

$$= \frac{N\sigma^2}{N^2} + 2 \cdot 0 \qquad \left[\because E(x_i-\theta) = 0\right]$$

$$= \sigma^2/N$$

Square error A



Biased and Unbiased estimator

- The estimated parameter $\hat{\theta}_D$ is a random variable since it depends on D which is a random sample.

$\theta \equiv \lambda$ of an exponential distribution

- For example: |D| = 3.
- Two different samples and means.

$D_1 = \{1, 1.5, 0.5\}$

$\hat{\lambda}_{D_1} = \frac{3}{3} = 1$

$D_2 = \{1.2, 0.8, 1.8\}$

$\hat{\lambda}_{D_2} = \frac{3.8}{3} = 1.2\overline{6}$

$D_1, D_2 - - - D_\infty$

- An interesting question: what is the expected value $\hat{\theta}_D$ over different random samples D? How does that compare with true θ?
  - Unbiased: $E_D(\hat{\theta}_D) = \theta$
  - Biased: $E_D(\hat{\theta}_D) \neq \theta$

Example: two unbiased estimator

- Parameter $\theta = \mu$ of Gaussian distribution.

$D \sim f$

- Two different estimators:

$D = \{x_1, - - x_N\}$

  - Lame estimator: just take first element: $\hat{\theta}_D = X_1$

$E_D[\hat{\theta}_D] = E_f[x_1] = \mu$

  - MLE: $\hat{\theta}_D = \dfrac{x_1 + x_2 + - - x_N}{N}$

$E_f\left[\dfrac{x_1 + x_2 - - + x_N}{N}\right] = \dfrac{N \cdot \mu}{N} = \mu$

Example: a biased estimator

- A constant estimator.

$\hat{\theta}_D = 5.7$ foot.

- MLE of Variance parameter of Gaussian: $\hat{\sigma}_D = \frac{\sum_{i=1}^{N}(x_i - \hat{\mu}_D)^2}{N}$

$E_D[\hat{\sigma}_D] \neq \sigma^2$

$N = 1$: $\hat{\sigma}_{|D|=1} = 0$

- Proof in https://en.wikipedia.org/wiki/Bias_of_an_estimator#Sample_variance

Doubt.

An unbiased estimator of variance of Gaussian

$$S^2 = \frac{\sum_{i=1}^{N} (x_i - \hat{\mu}_D)^2}{N-1}$$

$$\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2$$

where $\overline{x} = \sum_{i=1}^{n} x_i/n$. It follows from this identity that

$$(n-1)S^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$$

Taking expectations of both sides of the preceding yields, upon using the fact that for any random variable W, $E[W^2] = \mathrm{Var}(W) + (E[W])^2$,

$$\begin{aligned}
(n-1)E[S^2] &= E\left[\sum_{i=1}^{n} X_i^2\right] - nE[\overline{X}^2] \\
&= nE[X_1^2] - nE[\overline{X}^2] \\
&= n\mathrm{Var}(X_1) + n(E[X_1])^2 - n\mathrm{Var}(\overline{X}) - n(E[\overline{X}])^2 \\
&= n\sigma^2 + n\mu^2 - n(\sigma^2/n) - n\mu^2 \\
&= (n-1)\sigma^2
\end{aligned}$$

or

$$E[S^2] = \sigma^2$$

Consistent estimator

- An estimator is consistent if the estimation error goes to zero as N (size of D) goes to infinity.

$$\hat{\theta}_D \rightarrow \theta \text{ } as \text{ } |D| \rightarrow \infty$$

Example of an unbiased estimator that is not consistent.

- Parameter $\theta = \mu$ of Gaussian distribution, Lame estimator: just take first element: $\hat{\theta}_D = X_1$

Example of an unbiased, consistent estimator:

- Parameter $\theta =$ mean of a distribution. $\hat{\theta} = (X_1 + X_2 + \cdots + X_N)/N$

Example of a biased, consistent estimator:

- Parameter $\theta = \sigma$ of Gaussian distribution, $\hat{\sigma}$ is sample variance.