Variance: $\sigma^2$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

unbiased for $\sigma^2$

$$E(\sigma^2) = \sigma^2$$

Another estimator

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

unbiased for $\sigma^2$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{2}{n(n-1)} \left[ \sum_{i=1}^{n} (X_i - \mu) \sum_{j=1}^{n} (X_j - \mu) \right] + \frac{n}{n-1} (\bar{X} - \mu)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{2}{n(n-1)} n (\bar{X} - \mu)^2 + \frac{n}{n-1} (\bar{X} - \mu)^2$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2 - n (\bar{X} - \mu)^2$$

Example: Suppose r.v. $X$ has a Uniform distribution on $[0, \theta]$

density function

$$f(x) = \begin{cases} \frac{1}{\theta} & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

Observe $X, \ldots, X_n$

Want to estimate $\theta$

$$\hat{\theta} = \max(X, \ldots, X_n)$$

Probability problem: $Y = \max(X_1, \ldots, X_n)$

density function of $Y$

$$f_Y(y) = \sum_{k=0}^{y} \frac{n-k}{n} \frac{1}{\theta^n}$$

$$E(\hat{\theta}) = \int_{0}^{\theta} y \left( \sum_{k=0}^{y} \frac{n-k}{n} \frac{1}{\theta^n} \right) dy$$

$$= \int_{0}^{\theta} y \left( \frac{n}{n} \frac{1}{\theta^n} \right) dy$$

$$= \frac{n}{n+1} \left. \frac{y^2}{2} \right|_{0}^{\theta}$$

Bias of $\hat{\theta}$

$$\text{Bias of } \hat{\theta} = \frac{n}{n+1} \theta - \theta = \frac{-\theta}{n+1}$$

Unbiased estimator

$$\hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$E(\hat{\mu}) = \theta$$

Other unbiased estimator of $\mu$

- trimmed mean
- $X_i$

CONVERGENCE

An estimator is consistent if for any $\varepsilon > 0$

$$\lim_{n \to \infty} P\left( |\hat{\theta} - \theta| > \varepsilon \right) = 0$$

Example: $\hat{\mu}$ is a consistent estimator for $\mu$

by WLLN

- $\hat{\mu}$ is consistent for $\sigma^2$
- trimmed mean is consistent for $\mu$
- $X_i$ is not consistent for $\mu$

MINIMUM VARIANCE - VARIANCE

decreases as $n$ increases

An estimator of $\theta$, $\hat{\theta}$, is a function of r.v.'s $X_1, \ldots, X_n$

$\hat{\theta}$ has a probability distribution - called a sampling distribution
- the distribution of possible values of $\hat{\theta}$ from different samples
- Can assess variance of an estimator from its sampling distribution.
- A better estimator has smaller variance, more likely to be close to the true value $\theta$.

Examples
- Estimators of $\mu$: $\bar{X} = \frac{\sum X_i}{n}$ best unbiased
  $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

- Uniform $[0, \theta]$
  Another unbiased estimator of $\theta$: 
  $\hat{\theta}_{\text{UB}} = \bar{X}$
  (Exercise: show unbiased)
  $\text{Var}(\hat{\theta}_{\text{UB}}) = \frac{\sigma^2}{n}$
  $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

Suppose $\theta$ is a vector
$\theta = (\theta_1, \theta_2, ..., \theta_m)$

Method of moments estimates of $\theta_1, \theta_2, ..., \theta_m$ are obtained by equating the first $m$ sample moments to the corresponding first $m$ moments and solving for $\theta_1, \theta_2, ..., \theta_m$.

Example: Exponential ($\lambda$)
$f(x) = e^{-x}$ for $x > 0$

$E(X) = \frac{1}{\lambda}$

1st sample moment: $\bar{X}$

So method of moments estimator of $\lambda$ is 
$\hat{\lambda} = \frac{1}{\bar{X}}$

Example 2: MLE estimator for $\theta = \sigma^2$
1st sample moment: $\bar{X}$ or $\frac{1}{n} \sum X_i$
2nd sample moment: $\frac{1}{n} \sum X_i^2$

Method of Moments
- The $k^{th}$ moment of a r.v. $X$ is $E(X^k)$
- $X_1, ..., X_n$ random sample from a distribution with pdf $f(x|\theta)$, the $k^{th}$ sample moment is
  $\frac{1}{n} \sum X_i^k$

Distribution of $\bar{X}$ when $X_i \sim \text{Exp}(\lambda)$
$\bar{X} \sim \text{Gamma}(n, \lambda)$

Maximum Likelihood Estimation (MLE)
- Most recommended method because:
  - MLEs are consistent
  - When sample size ($n$) is large, the
MLE is approximately unbiased, and has variance that is nearly as small as any other estimator when $n$ is large, sampling distribution of MLE is approximately normal.

**Example**

Let $X_1, ..., X_n$ have joint pdf or pdf:

$$f(x_1, ..., x_n | \theta)$$

For $X_i$ i.i.d.

$$f(x_1, ..., x_n | \theta) = f(x_1 | \theta) f(x_2 | \theta) ... f(x_n | \theta)$$

When the $x_1, ..., x_n$ are observed sample values (data), the joint density function can be regarded as a function of $\theta$ called the likelihood function.

The MLE is the value of $\theta$ that maximizes the likelihood function.

Log-likelihood:

$$\ln f(x_1, ..., x_n | \mu, \sigma^2)$$

$$= -\frac{n}{2} \ln (2\pi \sigma^2) - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2$$