# Consistency of estimators

**Theorem**

*Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots,$ be a sequence of point estimators of $\theta^*$. If*

$$\lim_{n \to \infty} MSE(\hat{\Theta}_n) = 0$$

*then $\hat{\Theta}_n$ is a consistent estimator of $\theta^*$*

$$\begin{aligned}
P(|\hat{\Theta}_n - \theta^*| \geq \epsilon) &= P(|\hat{\Theta}_n - \theta^*|^2 \geq \epsilon^2) \\
&\leq \frac{E[\hat{\Theta}_n - \theta^*]^2}{\epsilon^2} \quad \text{Markov Inequality} \\
&= \frac{MSE(\hat{\Theta}_n)}{\epsilon^2} \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}$$

# Point Estimators for Mean and Variance

- We know by now that the sample mean $(\hat{\mu}_n)$ is an unbiased estimator for the mean and its MSE is $\frac{\sigma^2}{n}$. It is also consistent.
- What about sample variance ? How can it be defined ?
- Since $\sigma^2 = E[(X-\mu)^2]$, we can define sample variance estimator as $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.
- Problem with this estimator is that it needs the true mean which will not be available!
- What if we replace true mean by sample mean in the above formula?

# Point Estimators for Mean and Variance

- Let $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2$ denote the sample variance (with divisor $n$).
- HW: Is $S^2$ an unbiased estimator of the true variance $\sigma^2$? If not, compute the bias.
- **Step 1:** Expand using total variance identity:

$$\mathbb{E}[(X_i - \hat{\mu}_n)^2] = \mathbb{E}[(X_i - \mu + \mu - \hat{\mu}_n)^2]$$
$$= \mathbb{E}[(X_i - \mu)^2] + \mathbb{E}[(\hat{\mu}_n - \mu)^2] - 2\mathbb{E}[(X_i - \mu)(\hat{\mu}_n - \mu)]$$

# Point Estimators for Mean and Variance

- Let $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2$ denote the sample variance (with divisor $n$).

- HW: Is $S^2$ an unbiased estimator of the true variance $\sigma^2$? If not, compute the bias.

- **Step 1:** Expand using total variance identity:

$$\mathbb{E}[(X_i - \hat{\mu}_n)^2] = \mathbb{E}[(X_i - \mu + \mu - \hat{\mu}_n)^2]$$
$$= \mathbb{E}[(X_i - \mu)^2] + \mathbb{E}[(\hat{\mu}_n - \mu)^2] - 2\mathbb{E}[(X_i - \mu)(\hat{\mu}_n - \mu)]$$

- Use known results:

$$\mathbb{E}[(X_i - \mu)^2] = \sigma^2, \quad \mathbb{E}[(\hat{\mu}_n - \mu)^2] = \frac{\sigma^2}{n}$$
$$\mathbb{E}[(X_i - \mu)(\hat{\mu}_n - \mu)] = \frac{\sigma^2}{n}$$

- So:

$$\mathbb{E}[(X_i - \hat{\mu}_n)^2] = \sigma^2 + \frac{\sigma^2}{n} - 2 \cdot \frac{\sigma^2}{n} = \sigma^2 \left(1 - \frac{1}{n}\right)$$

- Thus,

$$\mathbb{E}[S^2] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[(X_i - \hat{\mu}_n)^2] = \left(1 - \frac{1}{n}\right) \sigma^2 = \frac{n-1}{n} \sigma^2$$

- So, $S^2$ is a **biased** estimator of $\sigma^2$.

$$\mathbb{E}[S^2] = \frac{n-1}{n} \sigma^2$$

$$\Rightarrow \text{Bias } B(S^2) = -\frac{1}{n} \sigma^2$$

# Point Estimators for Mean and Variance

- To obtain an unbiased estimator, define:

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

- Then $\mathbb{E}[\bar{S}^2] = \sigma^2$

> The sample variance defined by $\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2$ is an unbiased estimator of the variance.

- Is $\sqrt{\bar{S}^2}$ unbiased for $\sigma$?
  - Hint: Unbiasedness is not preserved under nonlinear transforms like square roots.

# Example: Estimating the Parameter of $\mathcal{U}(0, \theta)$

- Let $X_1, \ldots, X_n$ be i.i.d. samples from $\mathcal{U}(0, \theta)$. Consider the estimator:
$$\hat{\theta}_n = \max\{X_1, \ldots, X_n\}.$$
- **Bias:** $\hat{\theta}_n$ is a biased estimator of $\theta$.
$$\mathbb{E}[\hat{\theta}_n] = \frac{n}{n+1}\theta.$$
- **Unbiased Estimator:** Define
$$\tilde{\theta}_n = \frac{n+1}{n} \cdot \hat{\theta}_n.$$
Then $\mathbb{E}[\tilde{\theta}_n] = \theta$, hence it is unbiased.
- **Consistency:** Since $\hat{\theta}_n \xrightarrow{a.s.} \theta$, both estimators are consistent.

# Example: Consistent but Biased Estimator

- Let $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Consider

$$\hat{\mu}'_n = \frac{1}{n} \sum_{i=1}^{n} X_i + \frac{1}{n}.$$

- Then $\hat{\mu}'_n$ is biased:

$$\mathbb{E}[\hat{\mu}'_n] = \mu + \frac{1}{n} \Rightarrow \text{Bias} = \frac{1}{n}.$$

- But:

$$\hat{\mu}'_n \xrightarrow{p} \mu \quad \text{as } n \to \infty.$$

- Hence, **consistent but biased** estimator.

# Maximum likelihood estimation

- We have seen point estimators for mean and variance. What if we want to estimate other parameter in general like shape, scale, rate?
- Let $X_1, \ldots, X_n$ be i.i.d samples from a distribution with a parameter $\theta^*$. Let $\mathcal{D} = \{X_1 = x_1, \ldots, X_n = x_n\}$.
- If $X_i$'s are discrete, then the likelihood function is defined

$$L(x_1, x_2, \ldots, x_n; \theta) = p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \theta)$$

- $L(x_1, \ldots, x_n; \theta) = f_{X_1, \ldots, X_n}(x_1, \ldots, x_n; \theta)$ ($X_i's$ continuous)
- When samples are i.i.d, this is just the product of the densities/pmf's with parameter $\theta$
- In such cases, it is easier to work with the log likelihood function given by $\ln L(x_1, x_2, \ldots, x_n; \theta)$
- Find the likelihood when $\mathcal{D}$ are samples from $exp(\theta)$, $\mathcal{N}(\theta, 1)$, $Binom(\theta, p)$, $Binom(n, \theta)$ etc.

# Maximum likelihood estimation

- $L(x_1, \ldots, x_n; \theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta)$
- You want to find the best $\theta$ that represents the data!

Given $\mathcal{D} = \{x_1, \ldots, x_n\}$, the estimate $\hat{\Theta}_{ML}$ is given by

$$\begin{aligned}\hat{\Theta}_{ML} &= \arg\max_{\theta} L(x_1, \ldots, x_n; \theta) \\ &= \arg\max_{\theta} logL(x_1, \ldots, x_n; \theta)\end{aligned}$$

- We can generalize this to setting where more than one parameters say $(\theta_1^*, \ldots, \theta_k^*)$ are unknown.
- Note that differentiating w.r.t $\theta$ and equating to zero may not help if the parameter we are estimating is known to be an integer.

# Properties of MLEs (without proof)

Let $X_1, \ldots, X_n$ be a i.i.d sample from a distribution with parameter $\theta^*$. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e.,
   $\lim_{n\to\infty} P(|\hat{\Theta}_{ML} - \theta^*| > \epsilon) = 0$
2. $\hat{\Theta}_{ML}$ is asymptotically unbiased, i.e.,
   $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta^*$

# Example: MLE for Exponential Distribution

- Let $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$ with unknown rate parameter $\lambda$
- PDF: $f_X(x; \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$
- Log-likelihood:

$$\ln L(\lambda) = \sum_{i=1}^{n} \ln(\lambda e^{-\lambda x_i}) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

- Differentiate and set to 0:

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{n}{\lambda} - \sum x_i = 0 \Rightarrow \hat{\lambda}_{ML} = \frac{n}{\sum x_i}$$

# Example: MLE for Bernoulli Distribution

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ where $p$ is unknown
- PMF: $P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}, \quad x_i \in \{0, 1\}$
- Likelihood:

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \quad \Rightarrow \quad \ln L(p) = \sum x_i \ln p + \sum (1-x_i) \ln(1-$$

- Maximize:

$$\frac{d}{dp} \ln L(p) = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0 \Rightarrow \hat{p}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

# Example: MLE for Uniform $[0, \theta]$

- Let $X_1, \dots, X_n \sim \text{Uniform}(0, \theta)$ where $\theta$ is unknown
- PDF: $f_X(x; \theta) = \frac{1}{\theta}$ for $0 \leq x \leq \theta$, else 0
- Likelihood:

$$L(\theta) = \begin{cases} \theta^{-n}, & \text{if } \theta \geq \max x_i \\ 0, & \text{otherwise} \end{cases}$$

- So to maximize, choose the smallest such $\theta$:

$$\hat{\theta}_{ML} = \max(x_1, \dots, x_n)$$

# Example: MLE for $\mathcal{N}(\mu, \sigma^2)$ (known $\sigma^2$)

- $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, only $\mu$ unknown
- Log-likelihood:

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

- Maximize:

$$\frac{d}{d\mu} \ln L = \frac{1}{\sigma^2} \sum (x_i - \mu) = 0 \Rightarrow \hat{\mu}_{ML} = \frac{1}{n} \sum x_i$$

## Example: MLE for $\mathcal{N}(\mu, \sigma^2)$ (both unknown)

- $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, both $\mu$ and $\sigma^2$ unknown
- Log-likelihood:

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

- MLEs:

$$\hat{\mu}_{ML} = \frac{1}{n} \sum x_i$$
$$\hat{\sigma}^2_{ML} = \frac{1}{n} \sum (x_i - \hat{\mu})^2$$

Note: this differs from the unbiased sample variance (which divides by $n - 1$).

# Example: MLE for Poisson Distribution

- $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$
- PMF: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
- Log-likelihood:

$$\ln L(\lambda) = \sum_{i=1}^{n} (x_i \ln \lambda - \lambda - \ln x_i!) = n\bar{x} \ln \lambda - n\lambda + C$$

- Maximize:

$$\frac{d}{d\lambda} \ln L = \frac{n\bar{x}}{\lambda} - n = 0 \Rightarrow \hat{\lambda}_{ML} = \bar{x}$$