RECAP

- A point estimator $\hat{\Theta}$ is a function of the random samples $\hat{\Theta} = h(X_1, \ldots X_n)$
- The Bias $B(\hat{\Theta})$ of an estimator $\hat{\Theta}$ is defined as

$$B(\hat{\Theta}) = E[\hat{\Theta}] - \theta^*$$

- The mean squared error of an estimator $\hat{\Theta}$ is defined as

$$MSE(\hat{\Theta}) = E[(\hat{\Theta} - \theta^*)^2]$$

.

- $MSE(\hat{\Theta}) = Var(\hat{\Theta}) + Bias(\hat{\Theta})^2$
- We say that $\hat{\Theta}_n$ is a **consistent estimator** of θ, if

$$\lim_{n \to \infty} P(|\hat{\Theta}_n - \theta^*| \geq \epsilon) = 0, \quad \text{for all } \epsilon > 0$$

Markov's Inequality: Statement

Markov's Inequality: Let X be a non-negative random variable, and let $a > 0$. Then:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Key Points:

- Applies to **non-negative** random variables.
- Provides an upper bound on the probability of large deviations.
- Useful in analyzing tail probabilities.

Proof of Markov's Inequality

Proof:
Let X be a positive continuous random variable. We start by writing the expectation $\mathbb{E}[X]$ as:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{0}^{\infty} x f_X(x)\, dx \quad (\text{since } X \geq 0).$$

For any $a > 0$, we can split the integral as follows:

$$\mathbb{E}[X] = \int_{0}^{a} x f_X(x)\, dx + \int_{a}^{\infty} x f_X(x)\, dx.$$

Thus,

$$\mathbb{E}[X] \geq \int_{a}^{\infty} x f_X(x)\, dx.$$

Proof of Markov's Inequality (cont'd)

Since $x \geq a$ for $x \in [a, \infty)$, we have

$$\int_a^\infty x f_X(x)\,dx \geq \int_a^\infty a f_X(x)\,dx = a \int_a^\infty f_X(x)\,dx.$$

Now, we recognize that $\int_a^\infty f_X(x)\,dx = \mathbb{P}(X \geq a)$, so:

$$\mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a).$$

Dividing both sides by a (for $a > 0$), we conclude:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

$\square$

Consistency of estimators

Theorem

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

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

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

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

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 $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2$.
- HW Exercise: Is S^2 an unbiased estimator ? If no, find $B(\bar{S}^2)$.
- You will see that $E[S^2] = \frac{(n-1)\sigma^2}{n}$ and therefore $B(S^2) = \frac{-\sigma^2}{n}$.
- Can you think of an unbiased estimator of the variance ?
- How about $\bar{S}^2 = \frac{nS^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu}_n)^2$?

> 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 $\sqrt{\bar{S}^2}$ and unbiased estimator for the standard deviation σ.

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 θ^*. 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 θ
- 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 θ 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 θ 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 θ^*. 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., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta^*$