April 8

Let $X_1, \dots, X_n$ are IID observations from a $N(\mu, \sigma^2)$ population.

$\boxed{\text{Case I}}$: Let μ is "known" and we want to estimate σ^2.

The MLE of σ^2.

The likelihood function is

$L(\sigma^2)$

$$= \prod_{i=1}^{n} f(x_i)$$

$$\prod^{n} \frac{1}{} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$= \prod_{i=1} \frac{}{\sqrt{2\pi}\,\sigma} e$$

$$\vdots$$

The MLE is

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

$\boxed{\text{Case II}}$: when μ is unknown.

The likelihood function is

$$L(\mu, \sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2}$$

$$\sum^{n} (x_i - \mu)^2$$

$$= c\,\sigma^{-n}\, e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\cdots}$$

The log likelihood function is

$$l(\mu, \sigma^2)$$

$$= \log c - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\underbrace{\sum_{i=1}^{n}(x_i-\mu)^2}$$

~~Note that~~

do calculation yourselves

The MLE's are

$$\hat{\mu} = \overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

$$\begin{cases} \mu_2 = \ldots \\ \hat{\sigma}_2^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \end{cases}$$

<u>MSE</u>

$$MSE(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta}-\theta)^2\right]$$

<u>Th</u>

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2$$

where $Bias(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$

Proof

$$\text{MSE}(\hat{\theta})$$

$$= E\left[\left(\left(\hat{\theta} - E(\hat{\theta})\right) + \left(E(\hat{\theta}) - \theta\right)\right)^2\right]$$

$$= E\left[\left(\hat{\theta} - E(\hat{\theta})\right)^2 + \left(E(\hat{\theta}) - \theta\right)^2 + 2\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right]$$

$$= \text{Var}(\hat{\theta}) + E\left[\left(E(\hat{\theta}) - \theta\right)^2\right] + 2E\left[\left(\hat{\theta} - E(\hat{\theta})\right)\left(E(\hat{\theta}) - \theta\right)\right]$$

$$= Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2$$

$$+ 2\left(E(\hat{\theta}) - \theta\right) E\left[\hat{\theta} - E(\hat{\theta})\right]$$

$$= Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2$$

$X_1, \dots, X_n$ are IID $U(0, \theta)$ RVs. $\theta > 0$.

the MLE is

$$X_{(n)} = \max\{X_1, \dots, X_n\}$$

(n)

$= \max(X_i)$

CDF of $X_{(n)}$

$$F_{X_{(n)}}(y) = \mathbb{P}(X_{(n)} \le y)$$

$$= \mathbb{P}(X_1 \le y, \dots, X_n \le y)$$

$$= \mathbb{P}(X_1 \le y) \dots \mathbb{P}(X_n \le y)$$

$$= \left(\frac{y}{\theta}\right)^n \cdot \mathbb{1}(y \le \theta)$$

$$T_1 = 2\bar{X}_n$$

$$MSE(T_1) = Var(T_1)$$

$$MSE(T_1) = Var(T_1)$$

$$= \frac{\theta^2}{3n}$$

$$T_2 = X_{(n)} = \max(X_i)$$

$$MSE(T_2) = \frac{2\theta^2}{(n+1)(n+2)}$$

Clearly,

$$\frac{2\theta^2}{(n+1)(n+2)} \le \frac{\theta^2}{3n}$$

$$\Rightarrow MSE(T_2) \le MSE(T_1)$$

How to combine two unbiased estimators?

Let T_1 & T_2 are two unbiased estimators for θ.

Let $var(T_1) = \sigma_1^2$,

$var(T_2) = \sigma_2^2$.

Define

$$T = \lambda T_1 + (1-\lambda) T_2$$

for some $\lambda \in \mathbb{R}$.

Then

$$E(T) = E(\lambda T_1 + (1-\lambda) T_2)$$

$$= \lambda E(T_1) + (1-\lambda) E(T_2)$$

$$= \lambda E(T_1) \ldots$$

$$= \lambda\theta + (1-\lambda)\theta$$

$$\Rightarrow \boxed{E(T) = \theta}$$

Next

$\operatorname{var}(T)$

$$= \operatorname{var}(\lambda T_1 + (1-\lambda) T_2)$$

$$= \lambda^2 \operatorname{var}(T_1) + (1-\lambda)^2 \operatorname{var}(T_2) + 2\lambda(1-\lambda) \operatorname{cov}(T_1, T_2)$$

let T_1 and T_2 are uncorrelated, that is

$\text{cor}(T_1, T_2) = 0$.

Then

$\text{var}(T) =: \gamma(\lambda)$

$= \lambda^2 \sigma_1^2 + (1-\lambda)^2 \sigma_2^2$

we want to obtain λ such that $\text{Var}(T)$ is minimum.

$\frac{d}{d\lambda} \gamma(\lambda) =$

Maximum Likelihood Estimation (MLE) is not just an abstract calculus exercise; it is the fundamental engine driving modern scientific inference. When researchers observe real-world data, they use MLE to reverse-engineer the underlying parameters of the physical or biological systems that generated that data. Here are two concrete examples.

Healthcare: Survival Analysis in Clinical Trials

The Problem:

An oncologist is leading a clinical trial for a revolutionary new targeted therapy for leukemia. The team tracks a sample of $n = 50$ patients to measure the time until remission ends (the "survival time," T). Biological survival times are strictly positive and typically right-skewed, suppose they follow an Exponential distribution, $T \sim \text{Exp}(\lambda)$, where λ represents the underlying hazard rate (the rate at which patients relapse).

The MLE Application:
The doctors observe specific relapse times for the patients: $t_1, t_2, \ldots, t_{50}$. To understand the drug's true efficacy, they must estimate the unknown hazard rate, λ. They construct the likelihood function, which represents the joint probability of seeing exactly these 50 survival times:

$$L(\lambda) = \prod_{i=1}^{50} \lambda e^{-\lambda t_i}$$

Using calculus to maximize the log-likelihood, the Maximum Likelihood Estimator is

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{1}{\bar{t}}$$

Conclusion:
By mathematically proving that this specific $\hat{\lambda}$ makes the observed clinical data more probable than any other possible parameter, the oncologist can confidently model the entire survival curve. This allows the hospital to give fu-

ture patients highly accurate prognoses, such as calculating the exact probability that a patient will remain in remission for more than 5 years, $P(T > 5) = e^{-5\hat{\lambda}}$.

Pollution Control: Groundwater Arsenic Modeling

The Problem:

Environmental engineers are investigating a region where agricultural runoff is suspected of contaminating the groundwater with toxic arsenic. The engineers take $n = 100$ water samples across different wells. Because environmental toxins build up multiplicatively, contaminant concentrations (X) almost always follow a Lognormal distribution rather than a standard Normal distribution. The engineers need to find the true underlying parameters (μ and σ^2) of this regional contamination.

The MLE Application:

If X is lognormally distributed, then the natural logarithm of the data follows a normal distribution: $Y = \ln(X) \sim N(\mu, \sigma^2)$. The engineers transform their 100 raw arsenic readings into log-space $(y_1, y_2, \ldots, y_{100})$. They then construct the Gaussian likelihood function for these transformed values. Maximizing this likelihood yields the MLEs for both the geometric mean and the variance of the contamination:

$$\hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{and} \quad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2$$

The Result:

With the MLE parameters securely identified, the regulators possess the exact mathematical equation governing the region's groundwater. They can now calculate the precise probability that a randomly drilled well will exceed the EPA's legal toxic threshold of 10 parts per billion (ppb). If this MLE-derived probability is deemed too high, the EPA has the statistical and legal justification required to declare the region a Superfund site and release millions of dollars in federal cleanup funds.

MTL108
Estimation Theory

Rahul Singh

Mean Square Error (MSE)

Definition 1 (MSE)**.** For an estimator $\hat{\theta}$ for a parameter θ, the MSE is defined as

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2].$$

Lemma 1. *For an estimator $\hat{\theta}$ for a parameter θ,*

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + \left(Bias(\hat{\theta})\right)^2,$$

where $\left(Bias(\hat{\theta})\right) = \mathbb{E}(\hat{\theta}) - \theta$. Moreover, if an estimator is unbiased, its MSE is simply equal to its variance.

Proof. We have

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta}-\theta)^2] \\ &= \mathbb{E}\left[\left((\hat{\theta}-\mathbb{E}(\hat{\theta})) + (\mathbb{E}(\hat{\theta})-\theta)\right)^2\right] \\ &= \mathbb{E}\Big[(\hat{\theta}-\mathbb{E}(\hat{\theta}))^2 + (\mathbb{E}(\hat{\theta})-\theta)^2 + 2(\hat{\theta}-\mathbb{E}(\hat{\theta}))(\end{aligned}$$

Using the linearity of expectation,

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta}-\mathbb{E}(\hat{\theta}))^2\right] + \mathbb{E}\left[(\mathbb{E}(\hat{\theta})-\theta)^2\right] + 2\mathbb{E}\Big[(\hat{\theta} \tag{1}$$

Observe that $(\mathbb{E}(\hat{\theta})-\theta)$ is **not** a random variable, so

$$\mathbb{E}\left[(\mathbb{E}(\hat{\theta})-\theta)^2\right] = (\mathbb{E}(\hat{\theta})-\theta)^2 = \left(\mathrm{Bias}(\hat{\theta})\right)^2. \tag{2}$$

Next, using the linearity of expectation, we get

$$\begin{aligned}\mathbb{E}\left[(\hat{\theta}-\mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta})-\theta)\right] =& (\mathbb{E}(\hat{\theta})-\theta)\,\mathbb{E}\left[(\hat{\theta}-\mathbb{E}(\hat{\theta}))\right] \\ =& (\mathbb{E}(\hat{\theta})-\theta)\ \Big[\mathbb{E}(\hat{\theta})-\mathbb{E}(\mathbb{E}(\hat{\theta})) \\ =& (\mathbb{E}(\hat{\theta})-\theta)\ \left[\mathbb{E}(\hat{\theta})-\mathbb{E}(\hat{\theta})\right] =\end{aligned} \tag{3}$$

Combining (1), (2) and (3), we obtain

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \left(\text{Bias}(\hat{\theta})\right)^2.$$

This proves the first part. Next, if an estimator is unbiased, then $\text{Bias}(\hat{\theta}) = 0$, consequently

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}).$$

$\square$

Remark 1. While unbiasedness is a desirable property, it is not the only criterion for evaluating an estimator. Sometimes, we might prefer an estimator with a tiny amount of bias if it significantly reduces the variance. The MSE is a comprehensive metric that evaluates an estimator by combining both its variance and its bias.

Combining Independent Unbiased Estimators

Let T_1 and T_2 denote independent unbiased estimators of θ, having known variances σ_1^2 and σ_2^2. That is,

$$\mathbb{E}[T_i] = \theta, \quad \text{Var}(T_i) = \sigma_i^2 \quad \text{for } i = 1, 2$$

Any combined estimator of the form

$$T = \lambda T_1 + (1-\lambda)T_2$$

is also an unbiased estimator for θ. Because, using linearity of expectation

$$\begin{aligned}\mathbb{E}(T) =& \mathbb{E}[\lambda T_1 + (1-\lambda)T_2] \\ =& \lambda \mathbb{E}[T_1] + (1-\lambda)\mathbb{E}[T_2] \\ =& \lambda\theta + (1-\lambda)\theta = \theta.\end{aligned}$$

To determine the value of λ that results in T having the smallest possible MSE (which, since T is unbiased, is just its variance), we compute

$$r(\lambda) = \mathrm{Var}(T) = \lambda^2 \mathrm{Var}(T_1) + (1-\lambda)^2 \mathrm{Var}(T_2) = \lambda^2\sigma_1^2 + (1$$

To minimize this variance, we differentiate with respect to λ

$$\frac{d}{d\lambda} r(\lambda) = 2\lambda\sigma_1^2 - 2(1-\lambda)\sigma_2^2.$$

Setting this to 0 and solving for λ (let's call it $\hat{\lambda}$), we have

$$2\hat{\lambda}\sigma_1^2 = 2(1-\hat{\lambda})\sigma_2^2 \implies \hat{\lambda}(\sigma_1^2 + \sigma_2^2) = \sigma_2^2.$$

Consequently,

$$\hat{\lambda} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}.$$

Thus, the optimal weight given to an estimator is inversely proportional to its variance.

Comparing Estimators for a Uniform Distribution

Let $X_1, \ldots, X_n \sim U(0, \theta)$, where $U(0, \theta)$ is continuous uniform distribution with support $[0, \theta]$. We want to evaluate two estimators for the parameter θ.

Estimator 1: Since $\mathbb{E}[X_i] = \theta/2$, a natural unbiased estimator is based on the sample mean,

$$T_1 = 2\bar{X} = \frac{2}{n}\sum_{i=1}^{n} X_i$$

Since $\mathbb{E}[T_1] = \theta$, its MSE is just its variance. We know that $\text{Var}(X_i) = \theta^2/12$, so

$$\text{MSE}(T_1) = \text{Var}(T_1) = \frac{4}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{4}{n^2}\left(n\frac{\theta^2}{12}\right) =$$

Estimator 2: Maximum Likelihood Estimator The MLE for θ is $T_2 = \max(X_i) = X_{(n)}$. The CDF of T_2 is $F_2(x) = (x/\theta)^n$ for $x \leq \theta$, which gives the PDF

$$f_2(x) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n}, & if \quad 0 \leq x \leq \theta \\ 0, & otherwise. \end{cases}$$

Now we find its mean and variance; observe that

$$\mathbb{E}[T_2] = \int_0^\theta x \frac{nx^{n-1}}{\theta^n}\, dx = \frac{n}{\theta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta,$$

$$\mathbb{E}[T_2^2] = \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n}\, dx = \frac{n}{n+2}\theta^2.$$

So,

$$\mathrm{Var}(T_2) = \mathbb{E}[d_2^2] - (\mathbb{E}[d_2])^2 = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{\quad}{(n}$$

Now, The MSE of T_2 (since T_2 is biased),

$$\begin{aligned}
\mathrm{MSE}(T_2) &= (\mathbb{E}[T_2] - \theta)^2 + \mathrm{Var}(T_2) \\
&= \left(\frac{n}{n+1}\theta - \theta\right)^2 + \frac{n\theta^2}{(n+2)(n+1)^2} \\
&= \frac{\theta^2}{(n+1)^2} + \frac{n\theta^2}{(n+2)(n+1)^2} = \frac{(n+2)\theta^2 + n}{(n+2)(n+1}
\end{aligned}$$

Conclusion: Because $\dfrac{2\theta^2}{(n+1)(n+2)} \leq \dfrac{\theta^2}{3n}$ for all $n \geq 1$, the Maximum Likelihood Estimator (T_2) is fundamentally superior to the estimator T_1, even though T_2 is biased.

Asymptotic properties of MLE

The Rao-Cramer Lower Bound (RCLB)

Theorem 1 (without proof)**.** *For an unbiased estimator $\hat{\theta}$, under standard regularity conditions,*

$$Var(\hat{\theta}) \geq \frac{1}{nI(\theta)}$$

where n is the sample size, and $I(\theta)$ is the Fisher Information of a single observation, defined as

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta)\right]$$

An estimator that achieves this bound is called efficient.

The Rao-Cramer Lower Bound establishes the absolute minimum possible variance for any *unbiased* estimator of a parameter θ. It represents the theoretical limit of estimation precision.

Distribution of MLE under Standard Assumptions

Maximum Likelihood Estimators possess incredibly powerful asymptotic (large-sample) properties.

Theorem 2 (without proof)**.** *Under standard regularity conditions, as $n \to \infty$, the MLE $\hat{\theta}$ is*

1. ***Consistent:*** $\hat{\theta} \xrightarrow{p} \theta$.
2. ***Asymptotically Normal:*** *The distribution of the MLE approaches a normal distribution centered on the true parameter.*

3. ***Asymptotically Efficient:*** *The variance of the MLE reaches the Rao-Cramer Lower Bound.*

Formally, we write this asymptotic distribution as:

$$\hat{\theta}_{MLE} \sim N\left(\theta, \frac{1}{nI(\theta)}\right) \quad as\ n \to \infty$$