Probability and Statistics: Lecture-41

Monsoon-2020

by Dr. Pawan Kumar (IIIT, Hyderabad)
on November 18, 2020

» Asymptotic Properties of MLEs...

» Asymptotic Properties of MLEs...

Asymptotic Properties of MLEs

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ.

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ.

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

1. $\hat{\Theta}_{ML}$ is asymptotically consistent, i.e.,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of θ. Then, under some mild regularity conditions,

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

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of 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.,

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of 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., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of 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., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable.

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of 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., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely, the random variable

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of 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., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely, the random variable

$$\frac{\hat{\Theta}_{ML} - \theta}{\sqrt{\mathrm{Var}(\hat{\Theta}_{ML})}}$$

Asymptotic Properties of MLEs

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Let $\hat{\Theta}_{ML}$ denote the maximum likelihood estimator (MLE) of 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., $\lim_{n\to\infty} E[\hat{\Theta}_{ML}] = \theta$
3. As n becomes large, $\hat{\Theta}_{ML}$ is approximately a normal random variable. More precisely, the random variable

$$\frac{\hat{\Theta}_{ML} - \theta}{\sqrt{\mathrm{Var}(\hat{\Theta}_{ML})}}$$

converges in distribution to $N(0, 1)$.

» Solved Example 1 ...

Example

Show the following:

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable.

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable. Show that

$$\hat{\Theta}_2 = \hat{\Theta}_1 + W$$

is also an unbiased estimator for θ

$$E[\hat{\Theta}_2] = E[\hat{\Theta}_1] + E[W]$$
$$= E[\hat{\Theta}_1]$$

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable. Show that

$$\hat{\Theta}_2 = \hat{\Theta}_1 + W$$

   is also an unbiased estimator for θ

2. Let $\hat{\Theta}_1$ be an estimator for θ such that $E[\hat{\Theta}_1] = a\theta + b$, where $a \neq 0$.

$E[\hat{\theta}_2]$:

$\frac{E[\hat{\Theta}_1] - b}{a} = \frac{a\theta + \cancel{b} - \cancel{b}}{a}$

$= \underline{\underline{\theta}}$

Example

Show the following:

1. Let $\hat{\Theta}_1$ be an unbiased estimator for θ, and W is a zero mean random variable. Show that

$$\hat{\Theta}_2 = \hat{\Theta}_1 + W$$

   is also an unbiased estimator for θ

2. Let $\hat{\Theta}_1$ be an estimator for θ such that $E[\hat{\Theta}_1] = a\theta + b$, where $a \neq 0$. Show that

$$\hat{\Theta}_2 = \frac{\hat{\Theta}_1 - b}{a}$$

   is an unbiased estimator for θ

» Answer to previous problem...

» Solved Example...

Example

Let $X_1, X_2, \ldots, X_n$ be a random variable from a Uniform$(0, \theta)$ distribution, where θ is unknown.

Example

Let $X_1, X_2, \ldots, X_n$ be a random variable from a Uniform$(0, \theta)$ distribution, where θ is unknown. Consider the estimator

$$\hat{\Theta}_n = \max\{X_1, X_2, \cdots, X_n\}$$

Order statistics

$X_{(1)}$

Example

Let $X_1, X_2, \ldots, X_n$ be a random variable from a Uniform$(0, \theta)$ distribution, where θ is unknown. Consider the estimator

$$\hat{\Theta}_n = \max\{X_1, X_2, \cdots, X_n\}$$

1. Find the bias of $\hat{\Theta}_n$, $B(\hat{\Theta}_n)$

Example

Let $X_1, X_2, \ldots, X_n$ be a random variable from a Uniform$(0, \theta)$ distribution, where θ is unknown. Consider the estimator

$$\hat{\Theta}_n = \max\{X_1, X_2, \cdots, X_n\}$$

1. Find the bias of $\hat{\Theta}_n$, $B(\hat{\Theta}_n)$
2. Find the MSE of $\hat{\Theta}_n$, $\mathrm{MSE}(\hat{\Theta}_n)$

$B[\hat{\Theta}_n] = E[\hat{\Theta}_n] - \theta.$

Example

Let $X_1, X_2, \ldots, X_n$ be a random variable from a Uniform$(0, \theta)$ distribution, where θ is unknown. Consider the estimator

$$\hat{\Theta}_n = \max\{X_1, X_2, \cdots, X_n\}$$

1. Find the bias of $\hat{\Theta}_n$, $B(\hat{\Theta}_n)$
2. Find the MSE of $\hat{\Theta}_n$, MSE$(\hat{\Theta}_n)$
3. Is $\hat{\Theta}_n$ a consistent estimator of θ?

» Answer to previous problem...

$X \sim \text{Uniform}(0, \theta)$, then PDF and CDF

$$f_X(x) = \begin{cases} 1/\theta \, , & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/\theta & 0 \le x \le \theta \\ 1 & x > \theta \end{cases}$$

Since $\hat{\theta}_n = X_{(1)}$, PDF of $\hat{\theta}_n$ is

$$f_{\hat{\theta}_n}(y) = n \, f_X(y) \left[F_X(y) \right]^{n-1}$$

[Recall order stat.]

$$= \begin{cases} n \cdot \frac{1}{\theta} \cdot \left(\frac{y}{\theta}\right)^{n-1} & 0 \le y \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{n}{\theta^n} y^{n-1} & 0 \le y \le \theta \\ 0 & \text{otherwise.} \end{cases}$$

To find the Bias of $\hat{\theta}_n$

$$E[\hat{\theta}_n] = \int_0^\theta y \cdot \frac{n \, y^{n-1}}{\theta^n} \, dy$$

$$= \frac{n}{\theta^n} \int_0^\theta y^n \, dy = \frac{n}{\theta^n} \left[\frac{y^{n+1}}{n+1} \right]_0^\theta$$

$$= \frac{n}{n+1} \theta$$

Thus Bias

$$B(\hat{\theta}_n) = E[\hat{\theta}_n] - \theta$$

$$= \frac{n}{n+1}\theta - \theta = \frac{-\theta}{n+1}$$

(b) $MSE[\hat{\theta}_n] = Var(\hat{\theta}_n) + B(\hat{\theta}_n)^2$

$$= \underline{Var(\hat{\theta}_n)} + \frac{\theta^2}{(n+1)^2}$$

need this

$$\cancel{Var} E(\hat{\theta}_n^2) = \int_0^\theta y^2 \frac{n y^{n-1}}{\theta^n} dy = \frac{n}{n+2}\theta^2$$

$$Var(\hat{\theta}_n) = \underline{E[\hat{\theta}_n^2]} - \underline{(E[\hat{\theta}_n])^2}$$

$$= \frac{n}{(n+2)(n+1)^2}\theta^2 \quad \text{(check)}$$

$$\Rightarrow MSE[\hat{\theta}_n] = \frac{n}{(n+2)(n+1)^2}\theta^2 + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+2)(n+1)}.$$

(c) $\lim_{n \to \infty} MSE(\hat{\theta}_n)$

$= \lim_{n \to \infty} \frac{2\theta^2}{(n+2)(n+1)} = 0$

$\Rightarrow \hat{\theta}_n$ is a consistent estimator of θ.

» Solved Example...

Example

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a Geometric(θ) distribution, where θ is unknown.

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

Example

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a Geometric(θ) distribution, where θ is unknown. Find the maximum likelihood estimator (MLE) of θ based on this random sample.

Solution: $X_i \sim \text{Geometric}(\theta)$, then

$$P_{X_i}(x_i;\theta) = (1-\theta)^{x_i-1}\,\theta$$

Likelihood fn

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

$$= P_{X_1}(x_1;\theta)\, P_{X_2}(x_2;\theta) \cdots P_{X_n}(x_n;\theta)$$

$$= (1-\theta)^{\sum_{i=1}^{n} x_i - n}\,\theta^n$$

Better to use log likelihood to maximize

$$\ln L(x_1, x_2, \dots, x_n;\theta)$$

$$= \left(\sum x_i - n\right)\ln(1-\theta) + n\ln\theta$$

Maximize:

$$\frac{d}{d\theta}\ln L(x_1, \dots x_n;\theta)$$

$$= \left(\sum x_i - n\right)\frac{-1}{1-\theta} + \frac{n}{\theta} = 0$$

Solve for θ

$$\Rightarrow \hat{\theta}_{ML} = \frac{n}{\sum_{i=1}^{n} x_i}$$

» Solved Example...

Example

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a Uniform$(0, \theta)$ distribution, where θ is unknown.

Exercise

Example

Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a Uniform$(0, \theta)$ distribution, where θ is unknown. Find the maximum likelihood estimator (MLE) of θ based on this random sample.

» Answer to previous problem...

» Interval Estimation and Confidence Level...

Interval Estimation and Confidence Level

1. Let $X_1, X_2, \ldots, X_n$ be random sample from a distribution with a parameter θ to be estimated

Interval Estimation and Confidence Level

1. Let $X_1, X_2, \ldots, X_n$ be random sample from a distribution with a parameter θ to be estimated
2. Suppose we observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and obtained point estimate $\hat{\theta}$ of θ

Interval Estimation and Confidence Level

1. Let $X_1, X_2, \ldots, X_n$ be random sample from a distribution with a parameter θ to be estimated
2. Suppose we observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and obtained point estimate $\hat{\theta}$ of θ
3. Without additional information, we don't know whether $\hat{\theta}$ is close to θ

Interval Estimation and Confidence Level

1. Let $X_1, X_2, \ldots, X_n$ be random sample from a distribution with a parameter θ to be estimated
2. Suppose we observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and obtained point estimate $\hat{\theta}$ of θ
3. Without additional information, we don't know whether $\hat{\theta}$ is close to θ
4. In an interval estimation, instead of just one value $\hat{\theta}$, we produce an interval $[\hat{\theta}_\ell, \hat{\theta}_h]$ that is likely to include true value of θ

Interval Estimation and Confidence Level

1. Let $X_1, X_2, \ldots, X_n$ be random sample from a distribution with a parameter θ to be estimated
2. Suppose we observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, and obtained point estimate $\hat{\theta}$ of θ
3. Without additional information, we don't know whether $\hat{\theta}$ is close to θ
4. In an interval estimation, instead of just one value $\hat{\theta}$, we produce an interval $[\hat{\theta}_\ell, \hat{\theta}_h]$ that is likely to include true value of θ
5. The confidence level is the probability that the interval that we construct includes the real value of θ

» Interval Estimation and Confidence Level...

Interval Estimation and Confidence Level

1. Let $X_1, X_2, \dots, X_n$ be random sample from a distribution with a parameter θ to be estimated
2. Suppose we observed $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$, and obtained point estimate $\hat{\theta}$ of θ
3. Without additional information, we don't know whether $\hat{\theta}$ is close to θ
4. In an interval estimation, instead of just one value $\hat{\theta}$, we produce an interval $[\hat{\theta}_\ell, \hat{\theta}_h]$ that is likely to include true value of θ
5. The confidence level is the probability that the interval that we construct includes the real value of θ
6. The smaller the interval, the higher the precision with which we can estimate θ, and higher the confidence level

» Interval Estimation with Confidence Level...

Interval Estimation

* Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated

Interval Estimation

- Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated
- An interval estimator with confidence level $1 - \alpha$ consists of two estimators $\hat{\Theta}_l(X_1, X_2, \ldots, X_n)$ and $\hat{\Theta}_h(X_1, X_2, \cdots, X_n)$ such that

Interval Estimation

- Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated
- An interval estimator with confidence level $1-\alpha$ consists of two estimators $\hat{\Theta}_l(X_1, X_2, \ldots, X_n)$ and $\hat{\Theta}_h(X_1, X_2, \cdots, X_n)$ such that

$\hat{\Theta}_l \leq \theta \leq \hat{\Theta}_h$

$$P\left(\hat{\Theta}_l \leq \theta \text{ and } \hat{\Theta}_h \geq \theta\right) \geq 1-\alpha,$$

Interval Estimation

- Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated
- An interval estimator with confidence level $1-\alpha$ consists of two estimators $\hat{\Theta}_l(X_1, X_2, \ldots, X_n)$ and $\hat{\Theta}_h(X_1, X_2, \cdots, X_n)$ such that

$$P\left(\hat{\Theta}_l \leq \theta \text{ and } \hat{\Theta}_h \geq \theta\right) \geq 1-\alpha,$$

for every possible value of θ

Interval Estimation

* Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated
* An interval estimator with confidence level $1-\alpha$ consists of two estimators $\hat{\Theta}_l(X_1, X_2, \ldots, X_n)$ and $\hat{\Theta}_h(X_1, X_2, \cdots, X_n)$ such that

$$P\left(\hat{\Theta}_l \leq \theta \text{ and } \hat{\Theta}_h \geq \theta\right) \geq 1-\alpha,$$

  for every possible value of θ
* Equivalently, we say that $[\hat{\Theta}_l, \hat{\Theta}_h]$ is a $(1-\alpha)100\%$ confidence interval for θ

Interval Estimation

* Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated
* An interval estimator with confidence level $1-\alpha$ consists of two estimators $\hat{\Theta}_l(X_1, X_2, \ldots, X_n)$ and $\hat{\Theta}_h(X_1, X_2, \cdots, X_n)$ such that

$$P\left(\hat{\Theta}_l \leq \theta \text{ and } \hat{\Theta}_h \geq \theta\right) \geq 1-\alpha,$$

  for every possible value of θ
* Equivalently, we say that $[\hat{\Theta}_l, \hat{\Theta}_h]$ is a $(1-\alpha)100\%$ confidence interval for θ
* The randomness in these terms is due to $\hat{\Theta}_l$ and $\hat{\Theta}_h$, not θ

Interval Estimation

* Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ that is to be estimated
* An interval estimator with confidence level $1-\alpha$ consists of two estimators $\hat{\Theta}_l(X_1, X_2, \ldots, X_n)$ and $\hat{\Theta}_h(X_1, X_2, \cdots, X_n)$ such that

$$P\left(\hat{\Theta}_l \leq \theta \text{ and } \hat{\Theta}_h \geq \theta\right) \geq 1-\alpha,$$

  for every possible value of θ
* Equivalently, we say that $[\hat{\Theta}_l, \hat{\Theta}_h]$ is a $(1-\alpha)100\%$ confidence interval for θ
* The randomness in these terms is due to $\hat{\Theta}_l$ and $\hat{\Theta}_h$, not θ
* Here $\hat{\Theta}_l$ and $\hat{\Theta}_h$ are random variables because they are functions of $X_1, \ldots, X_n$

» Steps on Finding Interval Estimators...

1. Let X be a continuous random variable with CDF $F_X(x) = P(X \leq x)$

» Steps on Finding Interval Estimators...

1. Let X be a continuous random variable with CDF $F_X(x) = P(X \leq x)$
2. We are interested in finding two values x_l and x_h such that

» Steps on Finding Interval Estimators...

1. Let X be a continuous random variable with CDF $F_X(x) = P(X \leq x)$
2. We are interested in finding two values x_l and x_h such that

$$P\left(x_l \leq X \leq x_h\right) = 1 - \alpha$$

1. Let X be a continuous random variable with CDF $F_X(x) = P(X \leq x)$
2. We are interested in finding two values x_l and x_h such that

$$P\left(x_l \leq X \leq x_h\right) = 1 - \alpha$$

3. We can choose this as follows

$$P(X \leq x_l) = \frac{\alpha}{2} \quad \text{and} \quad P(X \geq x_h) = \frac{\alpha}{2}$$

» Steps on Finding Interval Estimators...

1. Let X be a continuous random variable with CDF $F_X(x) = P(X \leq x)$
2. We are interested in finding two values x_l and x_h such that

$$P\left(x_l \leq X \leq x_h\right) = 1 - \alpha$$

3. We can choose this as follows

$$P(X \leq x_l) = \frac{\alpha}{2} \quad \text{and} \quad P(X \geq x_h) = \frac{\alpha}{2}$$

$$= 1 - P(X \leq x_h) = 1 - \frac{\alpha}{2}$$

4. That is, we have from above

$$F_X(x_l) = \frac{\alpha}{2} \quad \text{and} \quad F_X(x_h) = 1 - \frac{\alpha}{2}$$

» Steps on Finding Interval Estimators...

1. Let X be a continuous random variable with CDF $F_X(x) = P(X \leq x)$
2. We are interested in finding two values x_l and x_h such that

$$P(x_l \leq X \leq x_h) = 1 - \alpha$$

3. We can choose this as follows

$$P(X \leq x_l) = \frac{\alpha}{2} \quad \text{and} \quad P(X \geq x_h) = \frac{\alpha}{2}$$

4. That is, we have from above

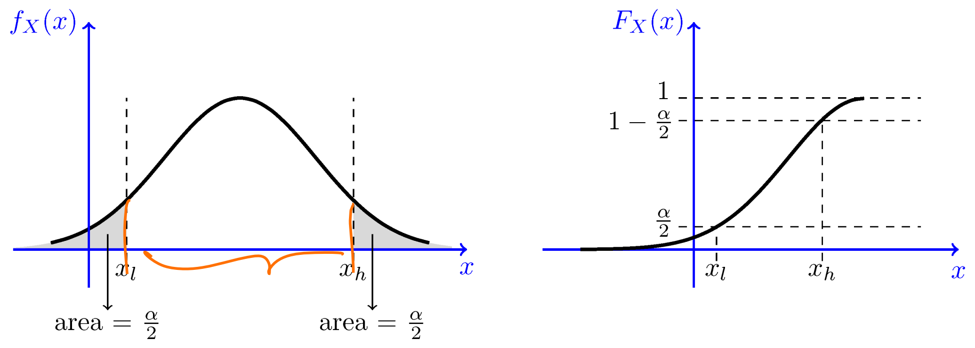
$$F_X(x_l) = \frac{\alpha}{2} \quad \text{and} \quad F_X(x_h) = 1 - \frac{\alpha}{2}$$

5. Rewriting these equations by using inverse, we have

$$x_l = F_X^{-1}\left(\frac{\alpha}{2}\right) \quad \text{and} \quad x_h = F_X^{-1}\left(1 - \frac{\alpha}{2}\right)$$

» Plot of confidence Interval...

» Plot of confidence Interval...



* $[x_l, x_h]$ is a $(1-\alpha)$ interval for X, that is, $P(x_l \leq X \leq x_h) = 1 - \alpha$

» Example of Interval Estimation...

» Example of Interval Estimation...

Example

Let $Z \sim N(0,1)$, find x_l and x_h such that

$$P(x_l \leq Z \leq x_h) = 0.95 \quad = 1-\alpha$$

$$\Rightarrow \alpha = 0.05$$

CDF of $Z = \Phi$

$$x_l = \Phi^{-1}\left(\frac{\alpha}{2}\right) = \Phi^{-1}(0.025) = -1.96$$

$$x_h = \Phi^{-1}\left(1-\frac{\alpha}{2}\right) = \Phi^{-1}(1-0.025) = 0.95$$

» Answer to previous problem...

» Answer to previous problem...

» Recall: Statistical Inference...

Statistical Inference: Compare frequentist and Bayesian

General setup for a statistical inference problem:

Statistical Inference: Compare frequentist and Bayesian

General setup for a statistical inference problem: There is an unknown quantity that we would like to estimate. We get some data.

Statistical Inference: Compare frequentist and Bayesian

General setup for a statistical inference problem: There is an unknown quantity that we would like to estimate. We get some data. From the data, we estimate the desired quantity.

Frequentist Approach

In that approach, the unknown quantity θ is assumed to be a fixed (non-random) quantity that is to be estimated by the observed data.

Statistical Inference: Compare frequentist and Bayesian

General setup for a statistical inference problem: There is an unknown quantity that we would like to estimate. We get some data. From the data, we estimate the desired quantity.

Frequentist Approach

In that approach, the unknown quantity θ is assumed to be a **fixed (non-random)** quantity that is to be estimated by the observed data.

Bayesian Approach

In the Bayesian framework, we treat the unknown quantity, Θ, as a random variable. More specifically, we assume that we have some **initial guess** about the distribution of Θ. This distribution is called the **prior distribution**. After observing some data, we update the distribution of Θ (based on the observed data).

» Motivating Example...

» Motivating Example...

Example (Motivating Example)

Example (Motivating Example)

Suppose that you would like to estimate the portion of voters in your town that plan to vote for Party A in an upcoming election.

» Motivating Example...

Example (Motivating Example)

Suppose that you would like to estimate the portion of voters in your town that plan to vote for Party A in an upcoming election. To do so, you take a random sample of size n from the likely voters in the town.

Example (Motivating Example)

Suppose that you would like to estimate the portion of voters in your town that plan to vote for Party A in an upcoming election. To do so, you take a random sample of size n from the likely voters in the town. Since you have a limited amount of time and resources, your sample is relatively small.

Example (Motivating Example)

Suppose that you would like to estimate the portion of voters in your town that plan to vote for Party A in an upcoming election. To do so, you take a random sample of size n from the likely voters in the town. Since you have a limited amount of time and resources, your sample is relatively small. Specifically, suppose that $n = 20$. After doing your sampling, you find out that 6 people in your sample say they will vote for Party A.

* Let θ be the true portion of voters in your town who plan to vote for Party A. You might want to estimate θ as

$$\hat{\theta} = 6/20 = 0.3$$

* Let θ be the true portion of voters in your town who plan to vote for Party A. You might want to estimate θ as

$$\hat{\theta} = 6/20 = 0.3$$

* In fact, in absence of any other data, that seems to be a reasonable estimate. However, you might feel that $n = 20$ is too small.

* Let θ be the true portion of voters in your town who plan to vote for Party A. You might want to estimate θ as

$$\hat{\theta} = 6/20 = 0.3$$

* In fact, in absence of any other data, that seems to be a reasonable estimate. However, you might feel that $n = 20$ is too small.
* Thus, your guess is that the error in your estimation might be too high.

* Let θ be the true portion of voters in your town who plan to vote for Party A. You might want to estimate θ as
$$\hat{\theta} = 6/20 = 0.3$$
* In fact, in absence of any other data, that seems to be a reasonable estimate. However, you might feel that $n = 20$ is too small.
* Thus, your guess is that the error in your estimation might be too high.
* While thinking about this problem, you remember that the data from the previous election is available to you.

* Let θ be the true portion of voters in your town who plan to vote for Party A. You might want to estimate θ as

$$\hat{\theta} = 6/20 = 0.3$$

* In fact, in absence of any other data, that seems to be a reasonable estimate. However, you might feel that $n = 20$ is too small.
* Thus, your guess is that the error in your estimation might be too high.
* While thinking about this problem, you remember that the data from the previous election is available to you.
* You look at that data and find out that, in the previous election, 40% of the people in your town voted for Party A.

* Let θ be the true portion of voters in your town who plan to vote for Party A. You might want to estimate θ as

$$\hat{\theta} = 6/20 = 0.3$$

* In fact, in absence of any other data, that seems to be a reasonable estimate. However, you might feel that $n = 20$ is too small.
* Thus, your guess is that the error in your estimation might be too high.
* While thinking about this problem, you remember that the data from the previous election is available to you.
* You look at that data and find out that, in the previous election, 40% of the people in your town voted for Party A.
* How can you use this data to possibly improve your estimate of θ?

* Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic.

* Although the portion of votes for Party *A* changes from one election to another, the change is not usually very drastic.
* Therefore, given that in the previous election 40% of the voters voted for Party *A*, you might want to model the portion of votes for Party *A* in the next election as a random variable Θ with a probability density function, $f_\Theta(\theta)$, that is mostly concentrated around $\theta = 0.4$.

* Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic.
* Therefore, given that in the previous election 40% of the voters voted for Party A, you might want to model the portion of votes for Party A in the next election as a random variable Θ with a probability density function, $f_\Theta(\theta)$, that is mostly concentrated around $\theta = 0.4$.
* For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

* Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic.
* Therefore, given that in the previous election 40% of the voters voted for Party A, you might want to model the portion of votes for Party A in the next election as a random variable Θ with a probability density function, $f_\Theta(\theta)$, that is mostly concentrated around $\theta = 0.4$.
* For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

* That is, before taking your random sample of size $n = 20$, this is your guess about the distribution of Θ.

* Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic.
* Therefore, given that in the previous election 40% of the voters voted for Party A, you might want to model the portion of votes for Party A in the next election as a random variable Θ with a probability density function, $f_\Theta(\theta)$, that is mostly concentrated around $\theta = 0.4$.
* For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

* That is, before taking your random sample of size $n = 20$, this is your guess about the distribution of Θ.
* Therefore, you initially have the prior distribution $f_\Theta(\theta)$. Then you collect some data, shown by D.

1 2

* Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic.
* Therefore, given that in the previous election 40% of the voters voted for Party A, you might want to model the portion of votes for Party A in the next election as a random variable Θ with a probability density function, $f_\Theta(\theta)$, that is mostly concentrated around $\theta = 0.4$.
* For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

* That is, before taking your random sample of size $n = 20$, this is your guess about the distribution of Θ.
* Therefore, you initially have the prior distribution $f_\Theta(\theta)$. Then you collect some data, shown by D.
* More specifically, here your data is a random sample of size n=20 voters, 6 of whom are voting for Party A.

- Although the portion of votes for Party A changes from one election to another, the change is not usually very drastic.
- Therefore, given that in the previous election 40% of the voters voted for Party A, you might want to model the portion of votes for Party A in the next election as a random variable Θ with a probability density function, $f_\Theta(\theta)$, that is mostly concentrated around $\theta = 0.4$.
- For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

- That is, before taking your random sample of size $n = 20$, this is your guess about the distribution of Θ.
- Therefore, you initially have the prior distribution $f_\Theta(\theta)$. Then you collect some data, shown by D.
- More specifically, here your data is a random sample of size n=20 voters, 6 of whom are voting for Party A.
- you can then proceed to find an updated distribution for Θ, called the posterior distribution, using Bayes' rule:

prior

$$f_{\Theta|D}(\theta|D) = \frac{P(D|\theta)\, f_\Theta(\theta)}{P(D)}. \tag{1}$$

- We can now use the posterior density, $f_{\Theta|D}(\theta|D)$ to further draw inferences about Θ

» Answer to previous problem...

» Bayesian Inference...

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y
2. The unknown variable is modelled as a random variable X, with prior distribution $f_X(x)$, if X is continuous, $P_X(x)$, if X is discrete

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y
2. The unknown variable is modelled as a random variable X, with prior distribution $f_X(x)$, if X is continuous, $P_X(x)$, if X is discrete
3. After observing the value of the random variable Y, we find the posterior distribution of X.

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y
2. The unknown variable is modelled as a random variable X, with prior distribution $f_X(x)$, if X is continuous, $P_X(x)$, if X is discrete
3. After observing the value of the random variable Y, we find the posterior distribution of X. This is the conditional PDF (or PMF) of X given $Y = y$, $f_{X|Y}(x|y)$ or $P_{X|Y}(x|y)$

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y
2. The unknown variable is modelled as a random variable X, with prior distribution $f_X(x)$, if X is continuous, $P_X(x)$, if X is discrete
3. After observing the value of the random variable Y, we find the posterior distribution of X. This is the conditional PDF (or PMF) of X given $Y = y$, $f_{X|Y}(x|y)$ or $P_{X|Y}(x|y)$
4. The posterior distribution is usually found using Bayes' formula.

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y
2. The unknown variable is modelled as a random variable X, with prior distribution $f_X(x)$, if X is continuous, $P_X(x)$, if X is discrete
3. After observing the value of the random variable Y, we find the posterior distribution of X. This is the conditional PDF (or PMF) of X given $Y = y$, $f_{X|Y}(x|y)$ or $P_{X|Y}(x|y)$
4. The posterior distribution is usually found using Bayes' formula. Using the posterior distribution, we can then find point or interval estimates of X

Bayesian Inference: main ideas

1. The goal is to draw inferences about an unknown variable X by observing a related random variable Y
2. The unknown variable is modelled as a random variable X, with prior distribution $f_X(x)$, if X is continuous, $P_X(x)$, if X is discrete
3. After observing the value of the random variable Y, we find the posterior distribution of X. This is the conditional PDF (or PMF) of X given $Y = y$, $f_{X|Y}(x|y)$ or $P_{X|Y}(x|y)$
4. The posterior distribution is usually found using Bayes' formula. Using the posterior distribution, we can then find point or interval estimates of X
5. Note that in the above setting, X or Y (or possibly both) could be random vectors

» Prior and Posterior...

Prior and Posterior

1. Using our notation for PMF and CDF, we have

Prior and Posterior

1. Using our notation for PMF and CDF, we have

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

Prior and Posterior

1. Using our notation for PMF and CDF, we have

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

2. If X is continuous RV and Y is discrete,

Prior and Posterior

1. Using our notation for PMF and CDF, we have

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

2. If X is continuous RV and Y is discrete,

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)f_X(x)}{P_Y(y)}$$

Prior and Posterior

1. Using our notation for PMF and CDF, we have

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

2. If X is continuous RV and Y is discrete,

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)f_X(x)}{P_Y(y)}$$

3. To find the denominator $P_Y(y)$ or $f_Y(y)$, we often use the law of total probability

Prior and Posterior

1. Using our notation for PMF and CDF, we have

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)P_X(x)}{P_Y(y)}$$

2. If X is continuous RV and Y is discrete,

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)f_X(x)}{P_Y(y)}$$

P(y|x)P(x)

3. To find the denominator $P_Y(y)$ or $f_Y(y)$, we often use the law of total probability
4. Here $f_{X|Y}(x|y)$ is called posterior distribution

» Solved Example...

Example

Let $X \sim \text{Uniform}(0, 1)$.

Example

Let $X \sim \text{Uniform}(0, 1)$. Suppose that we know

Example

Let $X \sim \text{Uniform}(0, 1)$. Suppose that we know

$$Y \mid X = x \sim \text{Geometric}(x).$$

» Solved Example...

① Baye's rule $f_{X|Y}(x|2) = \dfrac{P_{Y|X}(2|x)\, f_X(x)}{P_Y(2)}$

Since $Y|X \sim \text{Geometric}(x)$

$P_{Y|X}(y|x) = x(1-x)^{y-1}$ for $y = 1, 2, \ldots \Rightarrow P_{Y|X}(2|x) = x(1-x)$

Example

Let $X \sim \text{Uniform}(0, 1)$. Suppose that we know

$$Y \mid X = x \sim \text{Geometric}(x).$$

Find the posterior density of X given $Y = 2$, $f_{X|Y}(x|2)$

$P_Y(2) = \int_{-\infty}^{\infty} P_{Y|X}(2|x)\, f_X(x)\, dx = \int_0^1 x(1-x) \cdot 1\, dx = \frac{1}{6}$

$\Rightarrow f_{X|Y}(x|2) = \dfrac{x(1-x)\cdot 1}{1/6} = 6x(1-x)$ $\quad 0 \le x \le 1$

» Maximum Apriori Estimation (MAP)...

Definiton of MAP

The posterior distribution, $f_{X|Y}(x|y)$ (or $P_{X|Y}(x|y)$), contains all the knowledge about the unknown quantity X.

Definiton of MAP

The posterior distribution, $f_{X|Y}(x|y)$ (or $P_{X|Y}(x|y)$), contains all the knowledge about the unknown quantity X. Therefore, we can use the posterior distribution to find point or interval estimates of X.

Definiton of MAP

The posterior distribution, $f_{X|Y}(x|y)$ (or $P_{X|Y}(x|y)$), contains all the knowledge about the unknown quantity X. Therefore, we can use the posterior distribution to find point or interval estimates of X. One way to obtain *a* point estimate is to choose the value of x that maximizes the posterior PDF (or PMF).

Definiton of MAP

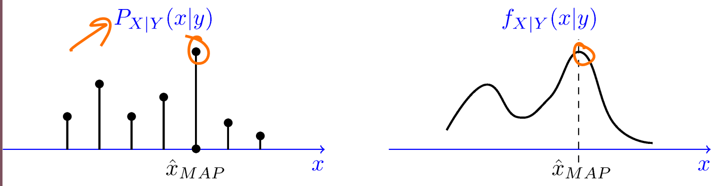
The posterior distribution, $f_{X|Y}(x|y)$ (or $P_{X|Y}(x|y)$), contains all the knowledge about the unknown quantity X. Therefore, we can use the posterior distribution to find point or interval estimates of X. One way to obtain a point estimate is to choose the value of x that maximizes the posterior PDF (or PMF). This is called the maximum a posteriori (MAP) estimation.

» Maximum Apriori Estimation (MAP)...

$$\frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)}$$ does not depend on X

Definiton of MAP

The posterior distribution, $f_{X|Y}(x|y)$ (or $P_{X|Y}(x|y)$), contains all the knowledge about the unknown quantity X. Therefore, we can use the posterior distribution to find point or interval estimates of X. One way to obtain a point estimate is to choose the value of x that maximizes the posterior PDF (or PMF). This is called the maximum a posteriori (MAP) estimation.



Here $\hat{x}_{MAP}$ is the value of X for which the posterior $f_{X|Y}(x|y)$ is maximized

» Finding the MAP Estimate...

MAP Estimate

We note that $f_Y(y)$ does not depend on the value of x.

MAP Estimate

We note that $f_Y(y)$ does not depend on the value of x. Hence, to find the MAP estimate of X given that we have observed $Y = y$,

MAP Estimate

We note that $f_Y(y)$ does not depend on the value of x. Hence, to find the MAP estimate of X given that we have observed $Y = y$, we find the value of x that maximizes

$$f_{Y|X}(y|x) f_X(x).$$

(forget $f_Y(y)$ in denominator)

MAP Estimate

We note that $f_Y(y)$ does not depend on the value of x. Hence, to find the MAP estimate of X given that we have observed $Y = y$, we find the value of x that maximizes

$$f_{Y|X}(y|x)f_X(x).$$

Whenever, X or Y is discrete,

» Finding the MAP Estimate...

MAP Estimate

We note that $f_Y(y)$ does not depend on the value of x. Hence, to find the MAP estimate of X given that we have observed $Y = y$, we find the value of x that maximizes

$$f_{Y|X}(y|x) f_X(x).$$

Whenever, X or Y is discrete, we replace PDF by its PMF.

» Example of MAP Estimate...

Example (Example of MAP Estimate)

Let X be a continuous random variable with the following PDF:

Example (Example of MAP Estimate)

Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

» Example of MAP Estimate...

Example (Example of MAP Estimate)

Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, let

Example (Example of MAP Estimate)

Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, let

$$Y|X = x \sim \text{Geometric}(x).$$

$P_{Y|X}(y|x) = x(1-x)^{y-1} \quad y = 1, 2, \ldots$

$\Rightarrow P_{Y|X}(3|x) = \underline{x(1-x)^2}.$

Need to find value $x \in [0,1]$ that maximize. $\underline{P_{Y|X}(y|x)}\,\underline{f_X(x)}$

Example (Example of MAP Estimate)

Let X be a continuous random variable with the following PDF:

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, let

$$Y|X = x \sim \text{Geometric}(x).$$

Find the MAP estimate of X given $Y = \underline{3}$.

$= x(1-x)^2 \cdot 2x = 2x^2(1-x)^2 \qquad 2x(1-x)^2 - 2(1-x)x^2 = 0$

Maximize: $\quad \frac{d}{dx}\left(x^2(1-x)^2 \right] =$

$\Rightarrow \hat{x}_{MAP} = \frac{1}{2}.$ (Check.

» Answer to previous problem...

» Comparison of MAP to ML Estimator...

» Comparison of MAP to ML Estimator...

Comparison of ML and MAP

1. Let $Y = y$ be observed value.

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes

$$f_{Y|X}(y|x).$$

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes

$$f_{Y|X}(y|x).$$

The ML estimate is shown by $\hat{x}_{ML}$.

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x).$$
The ML estimate is shown by $\hat{x}_{ML}$.
2. On the other hand,

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x).$$
The ML estimate is shown by $\hat{x}_{ML}$.
2. On the other hand, the MAP estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x) f_X(x)$$
prior

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x).$$
The ML estimate is shown by $\hat{x}_{ML}$.
2. On the other hand, the MAP estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x)f_X(x)$$
3. The two expressions are somewhat similar.

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x).$$
The ML estimate is shown by $\hat{x}_{ML}$.
2. On the other hand, the MAP estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x)f_X(x)$$
3. The two expressions are somewhat similar. The MAP has one extra term $f_X(x)$

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x).$$
The ML estimate is shown by $\hat{x}_{ML}$.
2. On the other hand, the MAP estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x)f_X(x)$$
3. The two expressions are somewhat similar. The MAP has one extra term $f_X(x)$
4. If X is uniformly distributed over a finite interval,

Comparison of ML and MAP

1. Let $Y = y$ be observed value. The maximum likelihood (ML) estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x).$$
The ML estimate is shown by $\hat{x}_{ML}$.
2. On the other hand, the MAP estimate of X is the value of x that maximizes
$$f_{Y|X}(y|x) f_X(x)$$
3. The two expressions are somewhat similar. The MAP has one extra term $f_X(x)$
4. If X is uniformly distributed over a finite interval, then ML and MAP estimate is same

» Solved Example...

Example

Suppose that the signal $X \sim N(0, \sigma_X^2)$ is transmitted over a communication channel.

» Solved Example...

Example

Suppose that the signal $X \sim N(0, \sigma_X^2)$ is transmitted over a communication channel. Assume that the received signal is given by

$$Y = X + W,$$

where $W \sim N(0, \sigma_W^2)$ is independent of X.

Example

Suppose that the signal $X \sim N(0, \sigma_X^2)$ is transmitted over a communication channel. Assume that the received signal is given by

$$Y = X + W,$$

where $W \sim N(0, \sigma_W^2)$ is independent of X.

1. Find the ML estimate of X, given $Y = y$ is observed

» Solved Example...

Sol^n: $f_X(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma_X} e^{-\frac{x^2}{2\sigma_X^2}}$

$\underbrace{Y|X = x \sim N(x, \sigma_W^2)}$ (Check) $\Rightarrow f_{Y|X}(y|x) = \dfrac{1}{\sqrt{2\pi}\,\sigma_W} e^{-\frac{(y-x)^2}{2\sigma_W^2}}$

Example

Suppose that the signal $X \sim N(0, \sigma_X^2)$ is transmitted over a communication channel. Assume that the received signal is given by

$$Y = X + W,$$

where $W \sim N(0, \sigma_W^2)$ is independent of X.

1. Find the ML estimate of X, given $Y = y$ is observed
2. Find the MAP estimate of X, given $Y = y$ is observed

(1) $f_{Y|X}(y|x) = \dfrac{1}{\sqrt{2\pi}\,\sigma_W} e^{-(y-x)^2/2\sigma_W^2}$

Maximize: $\min (y-x)^2 \Rightarrow \hat{x}_{ML} = y$

(2) $f_{Y|X}(y|x) f_X(x) = (C)\, e^{-\left[\frac{(y-x)^2}{2\sigma_W^2} + \frac{x^2}{2\sigma_X^2}\right]}$

Maximize $\Rightarrow$ minimize $\dfrac{(y-x)^2}{2\sigma_W^2} + \dfrac{x^2}{2\sigma_X^2}$

$\Rightarrow \boxed{\hat{x}_{MAP} = \sigma_X^2 / (\sigma_X^2 + \sigma_W^2)}$

» Minimum Mean Squared Error (MMSE) Estimation...

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.
2. To find a point estimate of X, we can just choose a summary statistic of the posterior such as its mean, median, or mode

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.
2. To find a point estimate of X, we can just choose a summary statistic of the posterior such as its mean, median, or mode
3. If we choose the mode (the value of x that maximizes $f_{X|Y}(x|y)$), we obtain the MAP estimate of X

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.
2. To find a point estimate of X, we can just choose a summary statistic of the posterior such as its mean, median, or mode
3. If we choose the mode (the value of x that maximizes $f_{X|Y}(x|y)$), we obtain the MAP estimate of X
4. Another possibility would be to choose the posterior mean, i.e., $\hat{x} = E[X|Y = y]$

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.
2. To find a point estimate of X, we can just choose a summary statistic of the posterior such as its mean, median, or mode
3. If we choose the mode (the value of x that maximizes $f_{X|Y}(x|y)$), we obtain the MAP estimate of X
4. Another possibility would be to choose the posterior mean, i.e., $\hat{x} = E[X|Y = y]$

The minimum mean squared error (MMSE) estimate of the random variable X,

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.
2. To find a point estimate of X, we can just choose a summary statistic of the posterior such as its mean, median, or mode
3. If we choose the mode (the value of x that maximizes $f_{X|Y}(x|y)$), we obtain the MAP estimate of X
4. Another possibility would be to choose the posterior mean, i.e., $\hat{x} = E[X|Y = y]$

The minimum mean squared error (MMSE) estimate of the random variable X, given that we have observed $Y = y$, is given by

MMSE

1. The posterior distribution, $f_{X|Y}(x|y)$, contains all the knowledge that we have about the unknown quantity X.
2. To find a point estimate of X, we can just choose a summary statistic of the posterior such as its mean, median, or mode
3. If we choose the mode (the value of x that maximizes $f_{X|Y}(x|y)$), we obtain the MAP estimate of X
4. Another possibility would be to choose the posterior mean, i.e., $\hat{x} = E[X|Y = y]$

The minimum mean squared error (MMSE) estimate of the random variable X, given that we have observed $Y = y$, is given by

$$\hat{x}_M = E[X|Y = y].$$

» Example of MMSE Computation...

Example

Let X be *a* continuous random variable with the following PDF

» Example of MMSE Computation...

Example

Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example

Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we are given that

» Example of MMSE Computation...

Example

Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we are given that

$$f_{Y|X}(y|x) = \begin{cases} 2xy - x + 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example

Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we are given that

$$f_{Y|X}(y|x) = \begin{cases} 2xy - x + 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the MMSE estimate of X, given $Y = y$ is observed.

Exercise

Example

Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Moreover, we are given that

$$f_{Y|X}(y|x) = \begin{cases} 2xy - x + 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the MMSE estimate of X, given $Y = y$ is observed.

» Answer to previous problem...

» Mean Squared Error (MSE) ...

MSE

Let $\hat{X} = g(Y)$ be an estimator of the random variable X, given that we have observed the random variable Y.

MSE

Let $\hat{X} = g(Y)$ be an estimator of the random variable X, given that we have observed the random variable Y. The mean squared error (MSE) of this estimator is defined as

MSE

Let $\hat{X} = g(Y)$ be an estimator of the random variable X, given that we have observed the random variable Y. The **mean squared error (MSE)** of this estimator is defined as

$$E[(X-X)^2] = E[(X-g(Y))^2]$$

MSE

Let $\hat{X} = g(Y)$ be an estimator of the random variable X, given that we have observed the random variable Y. The mean squared error (MSE) of this estimator is defined as

$$E[(X\text{-}X)^2] = E[(X\text{-}g(Y))^2]$$

The MMSE estimator of X,

MSE

Let $\hat{X} = g(Y)$ be an estimator of the random variable X, given that we have observed the random variable Y. The mean squared error (MSE) of this estimator is defined as

$$E[(X\text{-}X)^2] = E[(X\text{-}g(Y))^2]$$

The MMSE estimator of X,

$$\hat{X}_M = E[X|Y],$$

MSE

Let $\hat{X} = g(Y)$ be an estimator of the random variable X, given that we have observed the random variable Y. The mean squared error (MSE) of this estimator is defined as

$$E[(X\text{-}X)^2] = E[(X\text{-}g(Y))^2]$$

The MMSE estimator of X,

$$\hat{X}_M = E[X|Y],$$

has the lowest MSE among all possible estimators.