Probability and Statistics (Monsoon 2022)

Lecture-24

Pawan Kumar

IIIT, Hyderabad

November 16, 2022

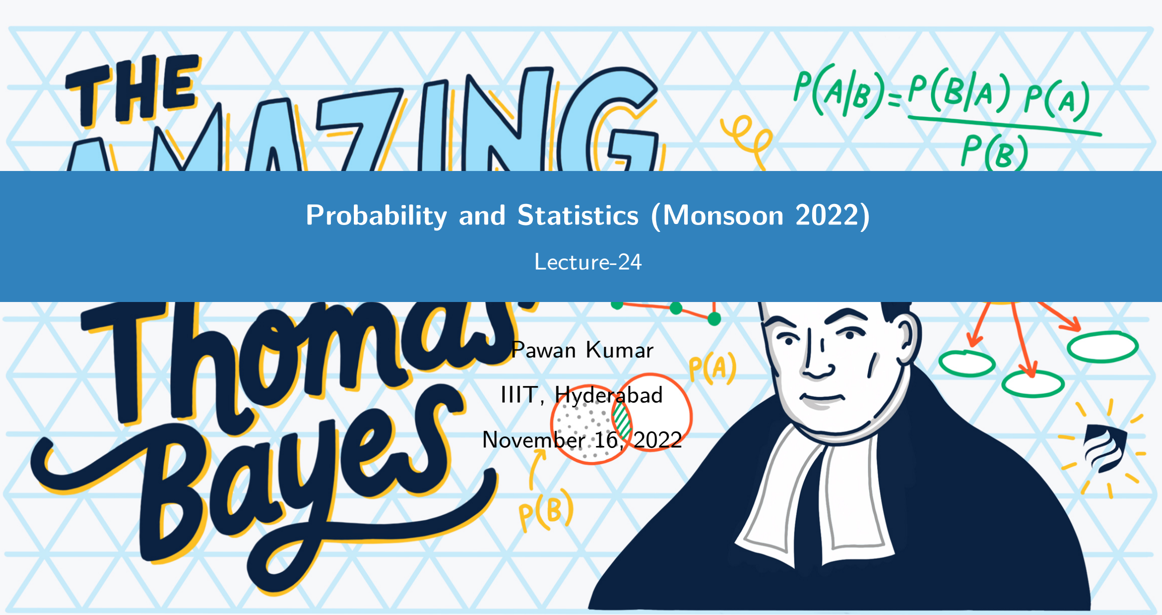

Table of Contents

Recall: Statistical Inference...

Recall: Statistical Inference...

Statistical Inference: Compare frequentist and Bayesian

General setup for a statistical inference problem:

Recall: Statistical Inference...

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.

Recall: Statistical Inference...

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.

Recall: Statistical Inference...

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).

Table of Contents

Motivating Example...

Motivating Example...

Example (Motivating 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.

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. Since you have a limited amount of time and resources, your sample is relatively small.

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. 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.

Some Observations...

Some Observations...

- 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

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

Some Observations...

- 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.

Some Observations...

- 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.

Some Observations...

- 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.

Some Observations...

- 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.

Some Observations...

- 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.

Some Observations...

- 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.
- 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 $\theta = 0.4$.

Some Observations...

Some Observations...

- For example, you might want to choose the density such that

$$E[\Theta] = 0.4$$

Some Observations...

- 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.

Some Observations...

- 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.

Some Observations...

- 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.

Some Observations...

- 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:

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

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

Bayesian Inference...

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...

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...

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...

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...

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...

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...

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

Example

Solved example Let $X \sim N(0, 1)$. Suppose that we know

$$Y \mid X = x \sim N(x, 1).$$

Show that the posterior density of X given $Y = y$, $f_{X|Y}(x \mid y)$ is given by

$$X \mid Y = y \sim N\left(\frac{y}{2}, \frac{1}{2}\right).$$

Table of Contents

Prior and Posterior...

Prior and Posterior...

Prior and Posterior

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

Prior and Posterior...

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...

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...

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...

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...

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
4. Here $f_{X|Y}(x|y)$ is called posterior distribution

Solved Example...

Solved Example...

Example

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

Solved Example...

Example

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

Solved Example...

Example

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

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

Solved Example...

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 $Y = 2$, $f_{X|Y}(x|2)$

Table of Contents

Maximum Apriori Estimation (MAP)...

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.

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. Therefore, we can use the posterior distribution to find point or interval estimates of X.

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. 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).

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. 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)...

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.

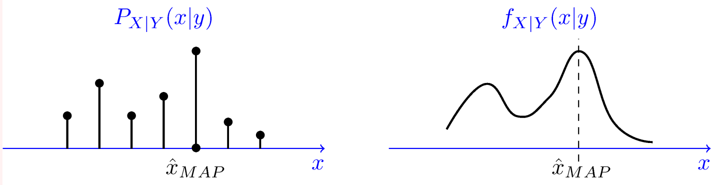

Figure: Here $\hat{x}_{MAP}$ is the value of X for which the posterior $f_{X|Y}(x|y)$ is maximized

Finding the MAP Estimate...

Finding the MAP Estimate...

MAP Estimate

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

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 $Y = y$,

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

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

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,

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 of MAP Estimate...

Example (Example of MAP Estimate)

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

Example of MAP Estimate...

Example (Example of MAP Estimate)

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

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

Example of MAP Estimate...

Example (Example of MAP Estimate)

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

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

Moreover, let

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

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

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

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

Find the MAP estimate of X given $Y = 3$.

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 MAP to ML Estimator...

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 MAP to ML Estimator...

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 MAP to ML Estimator...

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 MAP to ML Estimator...

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

Comparison of MAP to ML Estimator...

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 MAP to ML Estimator...

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 MAP to ML Estimator...

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 MAP to ML Estimator...

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...

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.

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.

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

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.

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

Table of Contents

Minimum Mean Squared Error (MMSE) Estimation...

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.

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.
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

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.
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

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

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.
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,

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.
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

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.
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 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

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

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

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 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

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

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

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

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

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

Mean Squared Error (MSE) ...

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.

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. The mean squared error (MSE) of this estimator is defined as

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. The mean squared error (MSE) of this estimator is defined as

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

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. The mean squared error (MSE) of this estimator is defined as

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

The MMSE estimator of X,

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. The mean squared error (MSE) of this estimator is defined as

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

The MMSE estimator of X,

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

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. The mean squared error (MSE) of this estimator is defined as

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

The MMSE estimator of X,

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

has the lowest MSE among all possible estimators.