Probability and Random Process (SWE3026)

Bayesian Inference

JinYeong Bak
jy.bak@skku.edu
College of Computing, SKKU

H. Pishro-Nik, "Introduction to probability, statistics, and random processes", available at https://www.probabilitycourse.com, Kappa Research LLC, 2014.

Statistical Inference

Frequentist (classical) inference

The unknown quantity which we want to estimate from data is assumed to be a fixed (deterministic, non-random) quantity,

It is to be estimated by the observed data

Bayesian inference

The unknown quantity is assumed to be a random variable,

We have some initial guess about the distribution of the quantity and we update the distribution using Bayes Rule

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.

Example

Let θ be the true portion of voters in your town who plan to vote for Party A.

$$\widehat{\boldsymbol{\theta}} = \frac{\mathbf{6}}{\mathbf{20}} = \mathbf{0.3}$$

But, the size of samples is too small.

Example

You look at that data and find out that, in the previous election, 40% of the people in your town voted for Party A.

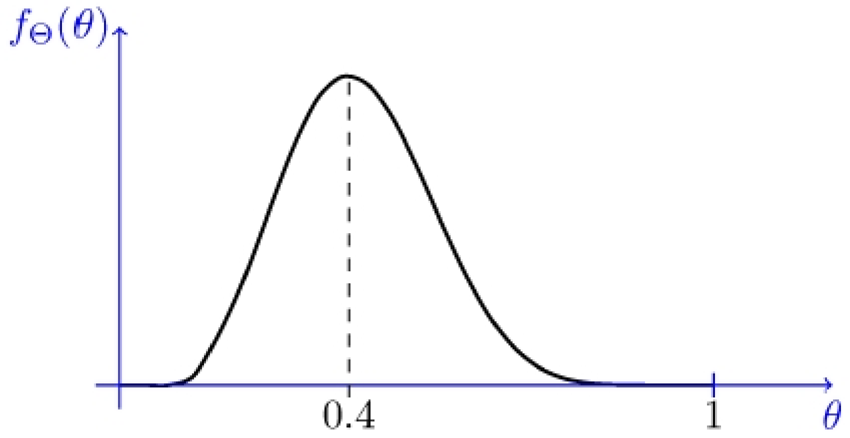
How to update θ with this knowledge?

Example

Assumption: A change of the portion of votes for Party A from one election to another is not usually very drastic

So you might model the portion of votes for Party A in the next election as a random variable $\boldsymbol{\Theta}$ with a pdf $\boldsymbol{f_{\Theta}(\theta)}$

Such a distribution shows your <u>prior belief</u> about $\boldsymbol{\Theta}$ in the absence of any additional data



Example

Prior distribution: $\boldsymbol{f_{\Theta}(\theta)}$

Likelihood function: $\boldsymbol{P(D|\theta)}$ **where** $\boldsymbol{D}$ **is some data**

Updated distribution for $\boldsymbol{\Theta}$ **(posterior distribution)**

$$\boldsymbol{f_{\Theta|D}(\theta|D) = \frac{P(D|\theta) f_{\Theta}(\theta)}{P(D)}}$$

Bayesian Statistical Inference

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

The unknown variable is modeled as a R.V. X, with <u>prior distribution</u> $f_X(x)$

After observing the value of the R.V. Y, we find the <u>posterior distribution</u> of X

This is the conditional PDF (or PMF) of X given $Y = y$,

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

Using the posterior distribution, we can find point or interval estimates of X

Statistical Inference

Frequentist (classical) inference

The unknown quantity which we want to estimate from data is assumed to be a fixed (deterministic, non-random) quantity,

It is to be estimated by the observed data

Bayesian inference

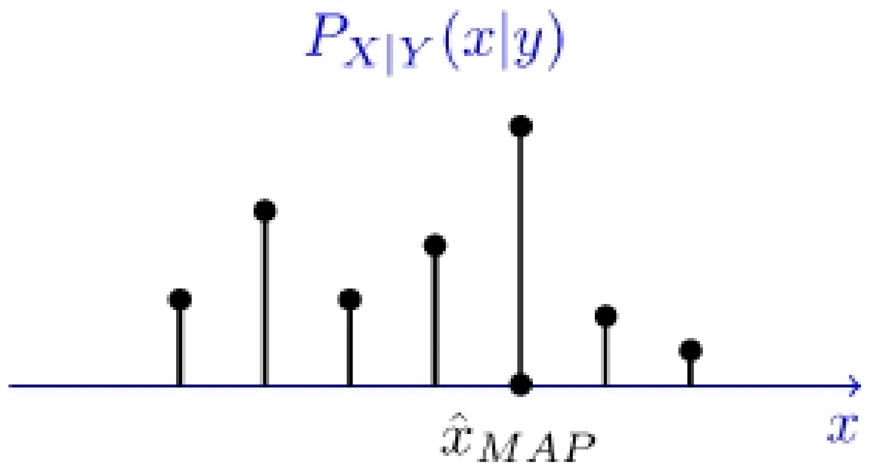
The unknown quantity is assumed to be a random variable,

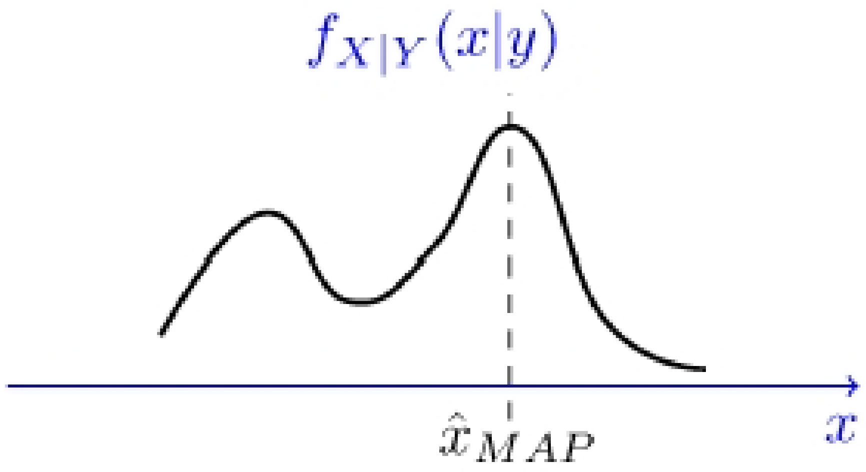
We have some initial guess about the distribution of the quantity and we update the distribution using Bayes Rule

Maximum A Posteriori (MAP) Estimation

Definition

A systematic way of parameter estimation that finds the parameter value that maximizes the posterior distribution

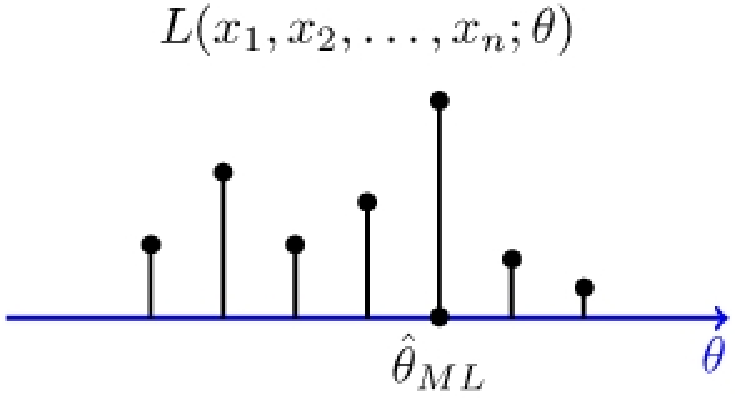


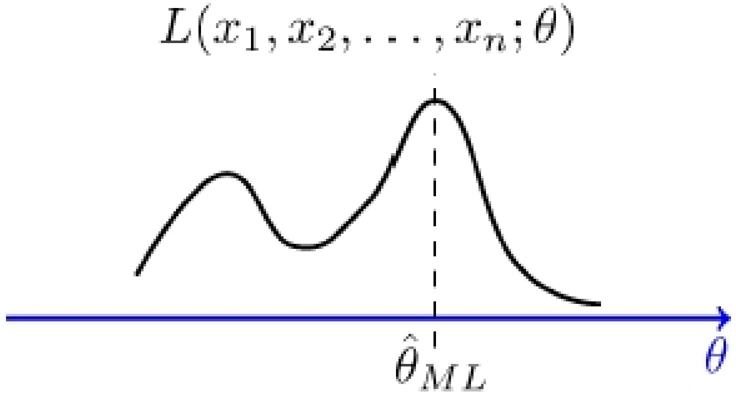


MLE vs MAP

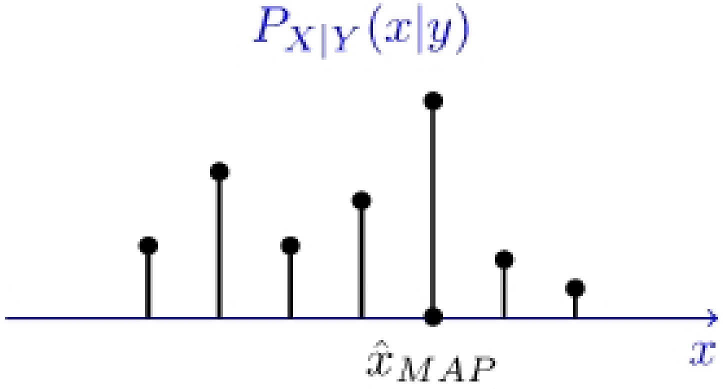
$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}$$

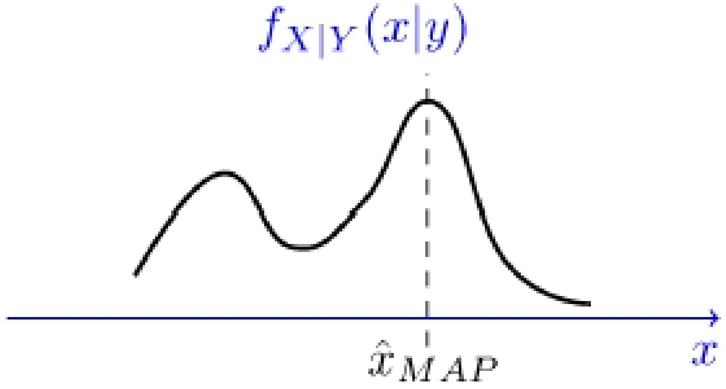
MLE





MAP





Hypothesis

Definition

H_0: null hypothesis, initially assumed to be true

H_1: alternative hypothesis, contradictory to H_0

Example

Let’s consider a radar system that uses radio waves to detect aircraft.

H_0: No aircraft is present

H_1: An aircraft is present

Hypothesis Testing for the Mean

Definition

H_0: null hypothesis, initially assumed to be true

H_1: alternative hypothesis, contradictory to H_0

Example

We have n random samples from a distribution and let's make inference about the mean of the distribution μ

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

Bayesian Hypothesis Testing

H_0: null hypothesis, initially assumed to be true

H_1: alternative hypothesis, contradictory to H_0

Assumption: we know prior probabilities of H_0 and H_1, $P(H_0)$ and $P(H_1)$

Data: we observe the random variable Y and its distribution under the hypotheses

$$f_Y(y|H_0) \text{ and } f_Y(y|H_1)$$

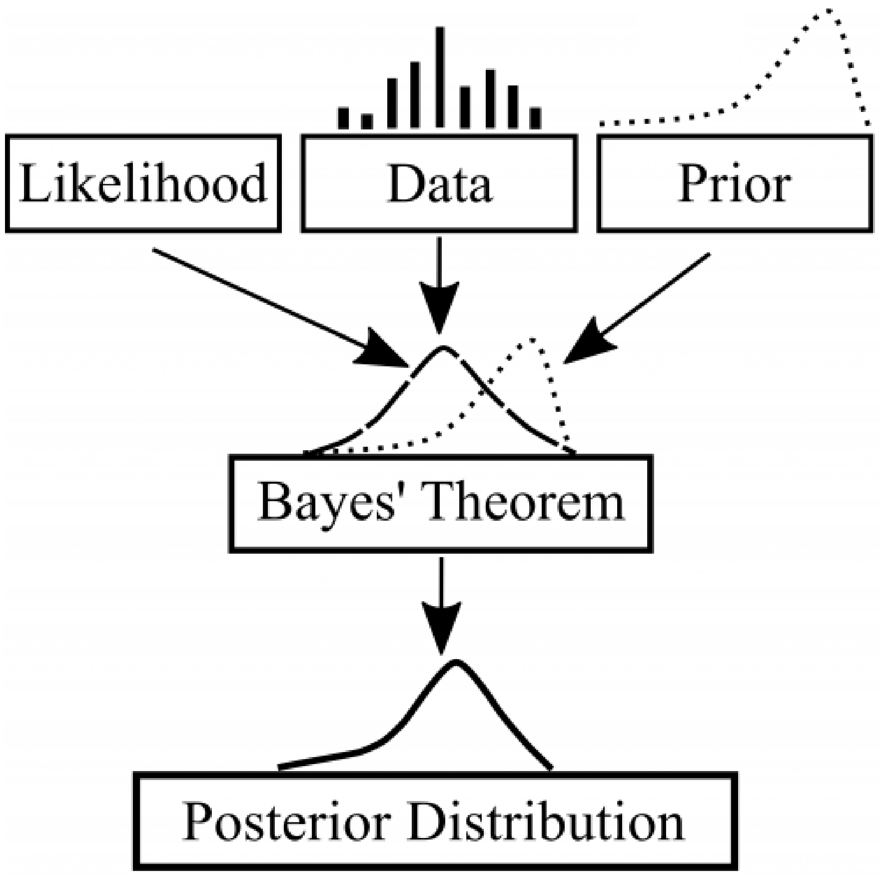
Then, compute the posterior probabilities of H_0 and H_1

$$P(H_0|Y=y) = \frac{f_Y(y|H_0)P(H_0)}{f_Y(y)}$$

$$P(H_1|Y=y) = \frac{f_Y(y|H_1)P(H_1)}{f_Y(y)}$$

And, accept the hypothesis with the higher posterior probability

Bayesian Modeling



https://medium.com/analytics-vidhya/hyperparameter-search-bayesian-optimization-14be6fbb0e09

Bayesian Modeling