Introduction to Probability

Samrachana Adhikari
Suresh Manandhar

December 13, 2019

Probability

Tool to quantify outcomes of an experiment whose exact values cannot be predicted with certainty.

1. Probability in this course typically represents the relative frequency of outcomes after infinitely many repetitions
2. We study probability because it is a tool that lets us make an inference from a sample to a population
3. Probability is used to understand what patterns in nature are "real" and which are due to chance

Motivating Example 1

- Flip a coin for 5 times and record number of heads
- Repeat for 100 times
- Example: {2Heads, 3 Heads, 5 Heads, 3 Heads, 2 Heads, ... }

1. How can I summarize the number of heads in 5 coin flips?
2. If I were to repeat this experiment, what is the expected number of heads?
3. What is the probability of getting 2 heads in a row when flipping a coin for 5 times?

Motivating Example 1

- Flip a coin for 5 times and record number of heads
- Repeat for 100 times
- Example: {2Heads, 3 Heads, 5 Heads, 3 Heads, 2 Heads, ... }

1. How can I summarize the number of heads in 5 coin flips?
2. If I were to repeat this experiment, what is the expected number of heads?
3. What is the probability of getting 2 heads in a row when flipping a coin for 5 times?

Motivating Example 2

- From a 52-cards deck, draw one card.
- Without putting it back, draw a second card.

1. What is the probability of getting a Queen in the first draw?
2. What is the probability of getting a Diamond in the second draw?

Motivating Example 3

What is the probability of getting an even numbered face when rolling a dice?

Motivating Example 4: Prevalence of Diabetes among Men

- Prevalence of Diabetes in Nepal is 25%.
- Number of men with diabetes in Banepa is 100,000.
- Total # of Men is 250,000 and total population of the city is 500,000.

What is the probability that a random person with diabetes in Banepa is a man?

What is In This Lecture?

1. **Concepts in Probability**
   - Probabilistic space
   - Laws of probability
   - Laws of conditional probability
   - Total probability theorem
   - Independence of events
2. **Random variable**
   - Discrete random variable
   - Bayes rule for discrete random variables
   - Probability mass function
   - Expectation
   - Variance
3. **Probability Distribution**
   - Continuous distribution
   - Density functions and likelihood
   - Parameter estimation

Probabilistic Model

1. **The sample space** Ω: the set of all outcomes from an experiment
2. **The probability law**: which assigns a number $P(A)$ to a set A of possible outcome

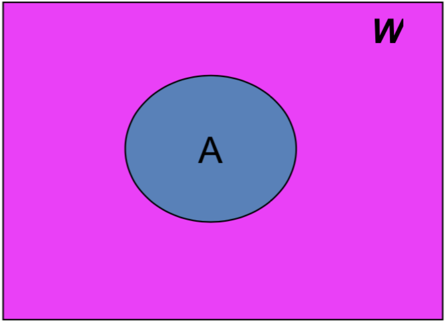
- A is called an **Event**. An event is a subset of the sample space Ω, or a collection of possible outcomes
- $P(A)$ is called the **probability of** A
  - $P(A)$ encodes our knowledge or belief about the "likelihood" of A
  - Interpretation in terms of frequency

Probabilistic Model

1. **The sample space** Ω: the set of all outcomes from an experiment
2. **The probability law**: which assigns a number $P(A)$ to a set A of possible outcome

- A is called an **Event**. An event is a subset of the sample space Ω, or a collection of possible outcomes
- $P(A)$ is called the **probability of** A
  - $P(A)$ encodes our knowledge or belief about the "likelihood" of A
  - Interpretation in terms of frequency

Probabilistic Model



$$P(A) = \frac{|A|}{|W|}$$

assuming the events in W are finite, discrete and equally likely.

Examples of Events

- Draw an ace from a pack of cards.
  - Sample space:= All possible draws
  - Event: $E = \{\mathrm{1Heart}, \mathrm{1Spade}, \mathrm{1Club}, \mathrm{1Diamond}\}$
  - **P(E) = ??**

- Draw two spades from a pack of cards.

$$\begin{aligned} E &= \{(1S, 2S), (1S, 3S), \ldots, (2S, 1S), \ldots\} \\ &\equiv \{(xS, yS) | x \in 1 \ldots 13, y \in 1 \ldots 13, x \neq y\} \end{aligned}$$

Examples of Events

- Draw an ace from a pack of cards.
  - Sample space:= All possible draws
  - Event: $E = \{1\text{Heart}, 1\text{Spade}, 1\text{Club}, 1\text{Diamond}\}$
  - **P(E) = ??**

- Draw two spades from a pack of cards.

$$
\begin{aligned}
E &= \{(1S, 2S), (1S, 3S), \ldots, (2S, 1S), \ldots\} \\
&\equiv \{(xS, yS) | x \in 1 \ldots 13, y \in 1 \ldots 13, x \neq y\}
\end{aligned}
$$

Example: Probability of getting an even number when rolling a dice

- **The sample space** Ω: $\{1,2,3,4,5,6\}$
- $A = \{2,4,6\}$ represents an event of getting an even number.

$$P(A) = \frac{\text{Number of elements in } A}{\text{Total number of elements in } \Omega} = \frac{3}{6} = 0.5.$$

Note that in this experiment, each outcome is equally likely. Hence, we call the dice 'fair'.

Probability Axioms: Things to Keep in Mind

1. (**Non-negativity**). For every event A

$$P(A) \geq 0$$

2. (**Additivity**). If A and B are two **disjoint events**, then the probability of their union satisfies

$$P(A \cup B) = P(A) + P(B).$$

More generally, if $A_1, A_2, \ldots$ is a sequence of disjoint events, then

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots.$$

3. (**Normalization**). The probability of the entire sample space Ω is 1:

$$P(\Omega) = 1.$$

Probability Axioms: Things to Keep in Mind

1. (**Non-negativity**). For every event A

$$P(A) \geq 0$$

2. (**Additivity**). If A and B are two **disjoint events**, then the probability of their union satisfies

$$P(A \cup B) = P(A) + P(B).$$

   More generally, if $A_1, A_2, \ldots$ is a sequence of disjoint events, then

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots.$$

3. (**Normalization**). The probability of the entire sample space Ω is 1:

$$P(\Omega) = 1.$$

Probability Axioms: Things to Keep in Mind

1. (**Non-negativity**). For every event A

$$P(A) \geq 0$$

2. (**Additivity**). If A and B are two **disjoint events**, then the probability of their union satisfies

$$P(A \cup B) = P(A) + P(B).$$

   More generally, if $A_1, A_2, \ldots$ is a sequence of disjoint events, then

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots.$$

3. (**Normalization**). The probability of the entire sample space Ω is 1:

$$P(\Omega) = 1.$$

Conditional Probability: Formal Definition

Definition.

For two events A and B where $P(B) > 0$, the conditional probability of A **given** B is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Remark.

- For discrete events, conditional probability can be estimated as

$$P(A|B) = \frac{c(A,B)}{c(B)} = \frac{\text{\# of both } A\&B \text{ in the sample}}{\text{\# of event} B}$$

Conditional Probability: Formal Definition

Definition.

For two events A and B where $P(B) > 0$, the conditional probability of A **given** B is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Remark.

- For discrete events, conditional probability can be estimated as

$$P(A|B) = \frac{c(A,B)}{c(B)} = \frac{\text{\# of both } A\&B \text{ in the sample}}{\text{\# of event} B}$$

Independence

When A is independent of B,

$$P(A|B) = P(A). \tag{1}$$

By the definition of conditional probability,

$$P(A \cap B) = P(B)P(A|B). \tag{2}$$

Combining (1)-(2), A is independent of B if and only if

$$P(A \cap B) = P(A)P(B).$$

- A is independent of B if and only if B is independent of A.
- We usually simply say "A and B are independent".

Independence

When A is independent of B,

$$P(A|B) = P(A). \tag{1}$$

By the definition of conditional probability,

$$P(A \cap B) = P(B)P(A|B). \tag{2}$$

Combining (1)-(2), A is independent of B if and only if

$$P(A \cap B) = P(A)P(B).$$

- A is independent of B if and only if B is independent of A.
- We usually simply say "A and B are independent".

Discussion Question

Given a regular 52-card deck. Consider:

- We draw one card from the deck.
- Let A be the event that you get an Ace.
- Let H be the event that you get a Heart.

1. In experiment 1, show that A and H are independent.

Discussion Question

- We draw one card, then without putting it back, we draw a second card.
- Let Q be the event you get a Queen in the first draw.
- Let D be the event you get a diamond in the second draw.

1. Compute $P(Q)$ and $P(D)$.
2. Compute $P(Q \cap D)$.
3. Compute $P(D|Q)$

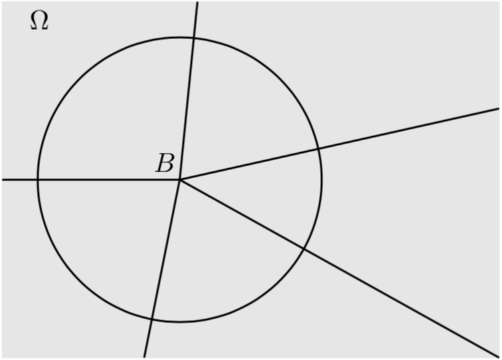
Total Probability Theorem

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample and assume that $P(A_i) > 0$ for all i. Then for any event B,

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \ldots + P(A_n)P(B|A_n) \\ &= \sum_{i=1}^{n} P(A_i)P(B|A_i). \end{aligned} \tag{3}$$

Recall that $A_1, A_2, \ldots, A_n$ is a partition of B if A_i are disjoint and $B = A_1 \cup A_2 \ldots \cup A_n$.

Law of Total Probability



Bayes Rule

Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $P(A_i) > 0$, for all i. Then for any event B such that $P(B) > 0$,

$$
\begin{aligned}
P(A_i|B) &= \frac{P(A_i)P(B|A_i)}{P(B)} \\
&= \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)} \\
&\propto P(A_i)P(B|A_i).
\end{aligned}
$$

Implication of Bayes Rule

- Prevalence of Diabetes in Nepal is 25%. Let D be an event that a person has diabetes.

$$P(D) = 0.25$$

- Number of men with diabetes is 100,000 in Banepa. Total # of Men in the area is 250,000, while the total population is 500,000. Let M be the event that a random person in the city is a man, i.e.

$$P(M) = \frac{250000}{500000} = 0.5$$

and

$$P(D|M) = \frac{75000}{250000} = 0.4$$

Example

What is the probability that a random person with diabetes in Banepa is a man? $P(M|D) = ??$

Using Bayes rule:

$$\begin{aligned} P(M|D) &= \frac{P(M \cap D)}{P(D)} \\ &= \frac{P(D|M)P(M)}{P(D)} \\ &= ?? \end{aligned}$$

Random Variables

A random variable is a **real-valued function** defined on the sample space, usually denoted by a capital letter, such as $X, Y, Z, \ldots$

- **Range** of a random variable: the set of values that it can take.
- A random variable is **discrete** if its range is either finite of countably infinite.
- Discrete random variables are different from **continuous random variables**.

Examples of Discrete Random Variables

- Coin flips

$$X = \begin{cases} 1 & \text{if Head} \\ 0 & \text{otherwise} \end{cases}$$

Range = $\{0,1\}$. (Finite)

- Calls in a call center

$$X = \text{Number of daily calls.}$$

Range = Set of all non-negative integers. (Countably infinite)

- Other examples of discrete random variables??

Probability Mass Function (PMF)

A random variable is characterized by its PMF.

- Let X be a random variable
- Let x be a value in its range (not a point Ω!)

PMF

PMF is a function p_X that is defined as

$$p_X(x) = P(X = x),$$

where for short,

$$P(X = x) = P(\{X = x\}).$$

For a subset S in the *range* of X,
$P(X \in S) = P(X \text{ takes a value within a set } S)$.

Key Properties of PMF

- $\sum_x p_X(x) = 1$.
- $P(X \in S) = \sum_{x \in S} p_X(x)$.

Expectation

For a random variable X with a PMF, the **expected value**, or the **expectation**, or the **mean** of X is defined as

$$E[X] = \sum_{x} x p_X(x).$$

The expectation can be viewed as

- *Weight average* of x, with $p_X(x)$ being the weights.
- The gravitational center.

Properties of mean and variance

- $E[X] = \sum_x x p_X(x)$
- $E[g(X)] = \sum_x g(x) p_X(x)$
- $\mathrm{var}(X) = E[(x - E[X])^2] = E[X^2] - (E[X])^2$:
- Since $\mathrm{var}(X) \geq 0$, it follows that for any X, $(E[X])^2 \leq E[X^2]$.
- $E[aX + b] = aE[X] + b$
- $\mathrm{var}(aX + b) = a^2 \mathrm{var}(X)$;

Example

Random Variable with Bernoulli Distribution

- A coin is tossed; let X denote the random variable indicating whether a head was flipped. $X = \begin{cases} 1 & \text{if head is flipped} \\ 0 & \text{otherwise} \end{cases}$
- Let p denote the probability of coming heads.

- X has a Bernoulli distribution with probability p; denoted as $X \sim \mathrm{Bernoulli}(p)$
- PMF of X can be written as

$$Pr(X = 1) = p_x(1) = p$$

$$Pr(X = 0) = p_x(0) = 1 - p.$$

Example

Random Variable with Bernoulli Distribution

- A coin is tossed; let X denote the random variable indicating whether a head was flipped. $X = \begin{cases} 1 & \text{if head is flipped} \\ 0 & \text{otherwise} \end{cases}$
- Let p denote the probability of coming heads.

- X has a Bernoulli distribution with probability p; denoted as $X \sim \mathrm{Bernoulli}(p)$
- PMF of X can be written as

$$Pr(X = 1) = p_x(1) = p$$

$$Pr(X = 0) = p_x(0) = 1 - p.$$

Example

Random Variable with Binomial Distribution

- Suppose a coin is tossed for n times.
- Let X denote the number of heads in the n tosses.
- Let p be the probability of head in each coin toss.

- $X \sim \mathrm{Binomial}(n, p)$
- PMF of X is

$$Pr(X = k) = p_x(k) = \binom{n}{k} p^k (1 - p)^k.$$

- $E(X) = np$

Example

Random Variable with Binomial Distribution

- Suppose a coin is tossed for n times.
- Let X denote the number of heads in the n tosses.
- Let p be the probability of head in each coin toss.

- $X \sim \mathrm{Binomial}(n, p)$
- PMF of X is

$$Pr(X = k) = p_x(k) = \binom{n}{k} p^k (1 - p)^k.$$

- $E(X) = np$

Example

Random Variable with Geometric Distribution

- Suppose a coin is tossed until a head is flipped.
- Let X denote the number of trials required until the first success.
- Let p be the probability of head in each toss.

- X has a Geometric distribution
- PMF of X is:

$$Pr(X = k) = p_x(k) = (1 - p)^{k-1}p, \text{ with } k = 1, 2, \ldots$$

The first $k - 1$ trials are failures, followed by success in n^{th} trail.

Example: Random Variable with Geometric Distribution

$$
\begin{aligned}
E(x) &= \sum_{x} x p(x) \\
&= \sum_{k=1}^{\infty} k(1-p)^{k-1} p \\
&= \frac{1}{p}
\end{aligned}
$$

Independence of Random Variables

Two random variables X and Y are independent if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \qquad \text{for all } x, y.$$

Since $p_{X,Y}(x, y) = p_Y(y)p_{X|Y}(x|y), \qquad \text{for all } x, y$
we conclude that when X and Y are independent,
$p_{X|Y}(x|y) = p_X(x), \qquad \text{for any } y \text{ at which } p_Y(y) > 0.$

Independence: Implication in the 2-Dimensional Table

When X and Y are independent, then in the table of joint PMF, the value in each cell is the product of the sum of the corresponding row times the sum of the corresponding column!

	$X = 1$	$X = 2$	$X = 3$
$Y = 1$	$1/3$	$1/6$	$1/6$
$Y = 2$	$1/6$	$1/12$	$1/12$

Conditional independence

The notion can be further extended to the case where we condition X and Y on an event A:

Definition. We say X and Y are independent given an event A, $P(A) > 0$, if $P(X = x, Y = y|A) = P(X = x|A)P(Y = y|A)$, or for short, we write $p_{X,Y|A}(x) = p_{X|A}(x)p_{Y|A}(y)$.

Once again, this is equivalent to that
$p_{X|Y,A}(x) = p_{X|A}(x), \qquad$ for all x and y with $p_{X|A}(y) > 0$

Remark

Independence does not necessarily imply conditional independence, and vice versa!

Properties of Independent Random Variables

If X and Y are independent, then

- $E[XY] = E[X] \cdot E[Y]$
- $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$.

Remark.

- The above property is usually not true when X and Y are *not independent*
- As X and itself are usually *not* independent, it is not true that $E[X^2] = E[X] \cdot E[X]$.

More than 2 Random Variable

Extensions to more than 2 random variables are straightforward:

- Random variables X_1, X_2, ..., X_n are independent if the joint PMF satisfies
  $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1)p_{X_2}(x_2)\ldots p_{X_n}(x_n)$.
- $E[X_1 \cdot X_2 \ldots X_n] = E[X_1] \cdot E[X_2] \ldots E[X_n]$.
- $\mathrm{var}(X_1 + X_2 + \ldots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \ldots \mathrm{var}(X_n)$.

Example: Variance of the Binomial

If X is Binomial random variable with parameter p, $0 < p < 1$, then

$$X = X_1 + X_2 + \ldots X_n,$$

where X_i are independent of each other, and each X_i is Bernoulli random variable with parameter p. We know that

$$E[X_i] = p, \qquad \mathrm{var}(X_i) = p(1-p).$$

It follows that

$$E[X] = E[X_1] + \ldots + E[X_n] = np,$$

and

$$\mathrm{var}(X) = \mathrm{var}(X_1) + \ldots + \mathrm{var}(X_n) = np(1-p).$$

Mean and Variance

	Mean	Variance
Bernoulli(p)	p	$p(1-p)$
Binomial(n,p)	np	$np(1-p)$
Geometric(p)	$1/p$	$(1-p)/p^2$
Poisson(λ)	λ	λ

Sample mean

Suppose X_i, $i = 1, 2, \ldots, n$, are independent Bernoulli random variable with parameter p, $0 < p < 1$.

Question. If p is not known to us. What is a good way to estimate p?

Answer. Let $\bar{X}$ be the sample mean

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

then

$$E[\bar{X}] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = \frac{1}{n}(np) = p,$$

and

$$\mathrm{var}(\bar{X}) = \frac{1}{n^2}\mathrm{var}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}(np(1-p)) = \frac{p(1-p)}{n}.$$

TEA BREAK

Continuous Random Variable

Example of continuous random variable

- Height; Support $= [0, \infty)$
- Daily change in stock market; Support $= (-\infty, \infty)$
- Income; Support $= [0, \infty)$
- Prevalence of flu; support $= [0,1]$

Probability Density Function (pdf)

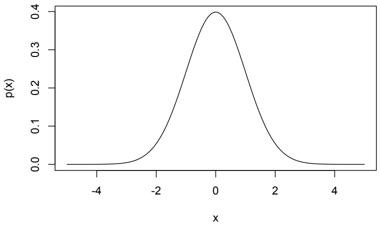
A random variable X is called **continuous** if there is a non-negative function $p(x)$, or **PDF**, such that:

$$P(X \in S) = \int_S p(x)dx.$$

Properties of a PDF:

- $p(x) \geq 0 \forall x$
- $\int_{-\infty}^{\infty} p(x)dx = 1$
- For any subset S of the real line

$$P(X \in S) = \int_S p(x)dx$$



Probability Density Function (pdf)

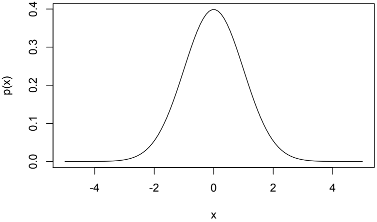
A random variable X is called **continuous** if there is a non-negative function $p(x)$, or **PDF**, such that:

$$P(X \in S) = \int_S p(x)dx.$$

Properties of a PDF:

- $p(x) \geq 0 \forall x$
- $\int_{-\infty}^{\infty} p(x)dx = 1$
- For any subset S of the real line

$$P(X \in S) = \int_S p(x)dx$$



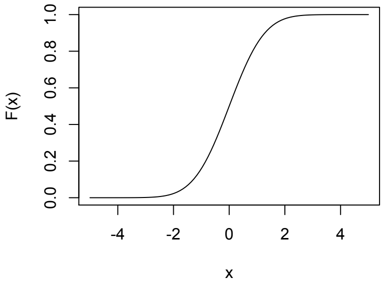
Cumulative Density Function (cdf)

- The **CDF** of a random variable X is defined by

$$F_X(x) = P(X \le x), \forall x.$$

The CDF has following properties:

- If $x \le y$, then $F_X(x) \le F_X(y)$
  Note: x and y are two possible values a random variable X could take and the distribution is defined wrt X.
- $F_X(x) \to 0$ as $x \to -\infty$ and $F_X(x) \to 1$ as $x \to \infty$



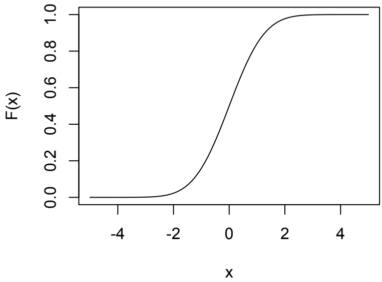
Cumulative Density Function (cdf)

- The **CDF** of a random variable X is defined by

$$F_X(x) = P(X \leq x), \forall x.$$

The CDF has following properties:

- If $x \leq y$, then $F_X(x) \leq F_X(y)$
  Note: x and y are two possible values a random variable X could take and the distribution is defined wrt X.
- $F_X(x) \to 0$ as $x \to -\infty$ and $F_X(x) \to 1$ as $x \to \infty$



CDF of a Discrete Random Variable

The probability that x lies in the interval $[a, b]$ is given by
$P(X \leq x) = F(a \leq X \leq b) = \sum_a^b p(X)dx$

CDF of a Continuous Random Variable

The probability that x lies in the interval $[a, b]$ is given by
$P(X \leq x) = F(a \leq X \leq b) = \int_a^b p(X)dx$

Expectation and Variance

Expectation of random variable X:

$$E(X) = \int_{\mathrm{Range}(x)} xp(x)dx.$$

Variance of random variable X:

$$Var(X) = \int_{\mathrm{Range}(x)} (x - E(X))^2 p(x)dx$$

Example: Normal Distribution

Random variable X has a Normal distribution denoted as $X \sim \text{Normal}(\mu, \sigma^2)$, $\mu = \text{Mean}$, $\sigma^2 = \text{Variance}$

- pdf

$$p_X(x) = p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp - \frac{(x-\mu)^2}{2\sigma^2}$$

- $E(X) = \mu$
- $Var(X) = \sigma^2$

Common Continuous Probability Distributions

Distribution	Support / Range
Uniform distribution	(a, b)
Gaussian (or Normal) distribution	$(-\infty, \infty)$
Exponential distribution	$(0, +\infty)$
Beta distribution	$[0, 1]$
Dirichlet distribution	$[0, 1]^n$
Multivariate Gaussian distribution	$(-\infty, +\infty)^n$

Parameter and Estimation

Going back to Normal distribution with unknown mean and known variance

$X \sim \text{Normal}(\mu, \sigma^2)$

μ is unknown

Can we use observed data to estimate μ?

Maximum Likelihood Estimation: MLE

- Let random variables $X_1, \dots, X_n$ denote heights of n adults, assume independent and identical events.
- Assume $X \sim \mathrm{Normal}(\mu, \sigma^2)$ and $\sigma^2 = 1$.
- Let $x_1, x_2, \dots, x_n$ be observed heights of n adults.

Maximum Likelihood Estimation

Goal is to find μ that maximizes $p(x|\mu)$ read as likelihood of μ.

$$\begin{aligned}\hat{\mu} &= \mathrm{argmax}_{\mu} p(x_1, \dots, x_n|\mu) \\ &= \mathrm{argmax}_{\mu} \prod_{i=n}^{n} p(x_i|\mu).\end{aligned}$$

In general let, Θ be the parameter of interest for random variable $X_1, \ldots, X_n$ with density p.

Maximum Likelihood Estimation

Then MLE of Θ written as $\hat{\Theta}$ is computed as

$$\hat{\Theta} \;=\; \mathrm{argmax}_{\Theta} \prod_{i=n}^{n} p(x_i|\Theta).$$

Bayes for Continuous Random Variable

- From Bayes Theorem, we know:

$$p(\mu|x_1, \ldots, x_n) = \frac{p(x_1, \ldots, x_n|\mu)p(\mu)}{p(x_1, \ldots, x_n)} \propto p(x_1, \ldots, x_n|\mu)p(\mu).$$

- $p(\mu|x_1, \ldots, x_n)$ is called the **posterior distribution** of μ.
- $p(\mu)$ is called the **prior** of μ.

Bayes rule

Posterior $\propto$ Likelihood $\times$ Prior
REMARK: as long as we know how to sample from likelihood and prior, we can use iterative algorithm to get samples from the posterior distribution.

Application of Bayes Rule: MAP

Let's go back to the Normal distribution example.

- Let random variables $X_1, \ldots, X_n$ denote heights of n adults, assume independent and identical events.
- Assume $X \sim \mathrm{Normal}(\mu, \sigma^2)$ and $\sigma^2 = 1$.
- Let $x_1, x_2, \ldots, x_n$ be observed heights of n adults.
- Assume $p(\mu) \sim \mathrm{Normal}(\mu_0, \sigma_0^2)$.

MAP

MAP Estimate of μ

Goal is to find μ that maximizes $p(\mu|x_1, \dots, x_n)$.
Using Bayes rule:

$$p(\mu|X_1, \dots, X_n) \propto \prod_{i=1}^{n} p(x_i|\mu)p(\mu)$$

Can show:

$$\mu|X_1, \dots, X_n \sim \mathrm{Normal}(\mu_{post}, \sigma^2_{post})$$

Sampling from Posterior Distribution

$$\mu | X_1, \ldots, X_n \sim \mathrm{Normal}(\mu_{post}, \sigma^2_{post})$$

- $\mu_{post} = \frac{1}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum x_i}{\sigma^2} \right)$
- $\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2} \right)^{-1}$