COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

STAT 4224/5224

Bayesian Statistics

Dobrin Marchev

Probability Review

Let H denote the sample space and let the subset $E \subset H$ be an event.

A **probability measure** P is defined on H such that for each event E, $P(E)$ satisfies the following axioms:

- Axiom 1: $P(E) \geq 0$, for each E.
- Axiom 2: $P(H) = 1$.
- Axiom 3:

If E_1, E_2, … are disjoint events, then:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

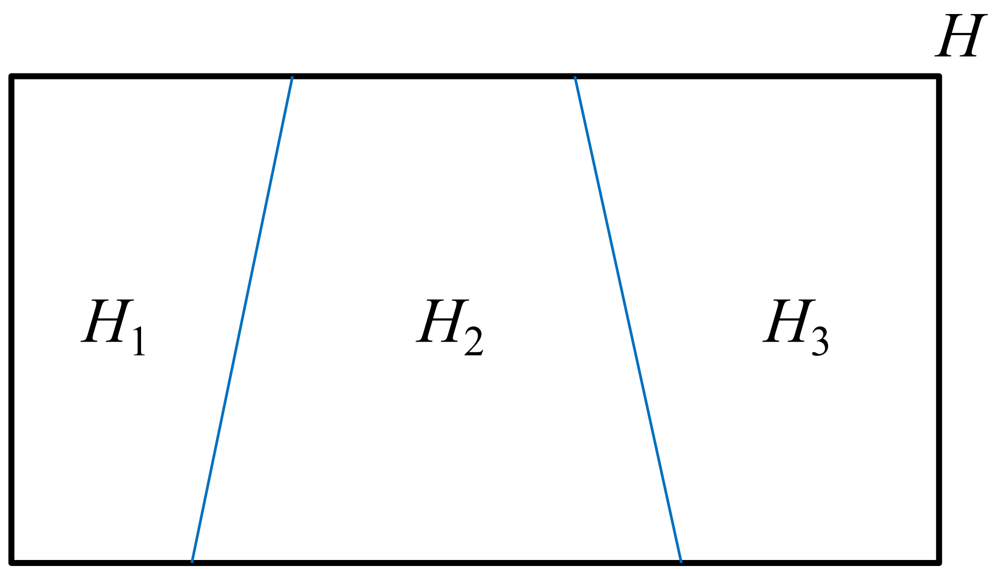
Basic Probability Rules

1. $P(E) \geq 0$, for each E.
2. $P(H) = 1$.
3. If A and B are disjoint, then $P(A \cup B) = P(A) + P(B)$
4. $P(\bar{A}) = 1 - P(A)$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Partitions

Definition: A collection of events $\{H_1, \ldots, H_k\}$ is a *partition* of the sample space H if:

1. The events are disjoint, that is, $H_i \cap H_j =$ for $i \neq j$
2. The union of all events is H, that is, $\bigcup_{i=1}^{k} H_i = H$



Partitions and Probability

Suppose $\{H_1, \ldots, H_k\}$ is a partition of the sample space H and E is any event. Then the following can be proved.

- **Multiplication Law:**

$$P\left(E \cap H_j\right) = P\left(H_j\right)P\left(E \middle| H_j\right)$$

- **Total Probability Law:**

$$P(E) = \sum_{i=1}^{k} P(E \cap H_i) = \sum_{i=1}^{k} P(H_i)P(E|H_i)$$

- **Bayes' Theorem:**

$$P\left(H_j \middle| E\right) = \frac{P\left(H_j\right)P\left(E \middle| H_j\right)}{P(E)} = \frac{P\left(H_j\right)P\left(E \middle| H_j\right)}{\sum_{i=1}^{k} P(H_i)P(E|H_i)}$$

Example 1

An insurance company insured 2000 scooter drivers, 4000 car drivers, and 6000 truck drivers. The probability of an accident involving a scooter driver, car driver, and a truck is 0.01, 0.03, and 0.015, respectively. One of the insured persons had an accident. What is the probability that he/she is a scooter driver?

Example 1 Solution

An insurance company insured 2000 scooter drivers, 4000 car drivers, and 6000 truck drivers. The probability of an accident involving a scooter driver, car driver, and a truck is 0.01, 0.03, and 0.015, respectively. One of the insured persons had an accident. What is the probability that he/she is a scooter driver?

$H_1 = \{\text{scooter driver}\}$, $P(H_1) = \frac{2000}{12000} = \frac{1}{6}$

$H_2 = \{\text{car driver}\}$, $P(H_2) = \frac{4000}{12000} = \frac{1}{3}$

$H_3 = \{\text{truck driver}\}$, $P(H_3) = \frac{6000}{12000} = \frac{1}{2}$

$E = \{\text{accident}\}$, $\mathrm{P}(E|H_1) = 0.01$, $\mathrm{P}(E|H_2) = 0.03$, $\mathrm{P}(E|H_3) = 0.015$

$$P(E) = \sum_{i=1}^{3} P(H_i)P(E|H_i) = \frac{1}{6}(0.01) + \frac{1}{3}(0.03) + \frac{1}{2}(0.015) = 0.019167$$

$$P(H_1|E) = \frac{P(H_1)P(E|H_1)}{P(E)} = \frac{\frac{1}{6}(0.01)}{0.019167} = 0.0869565$$

Example 1 Analysis

Prior distribution of driver type is:

$$P(H_1) = \frac{1}{6}, P(H_2) = \frac{1}{3}, P(H_3) = \frac{6000}{12000} = \frac{1}{2}$$

Posterior distribution, given there is an accident is:

$$P(H_1|E) = 0.08696, P(H_2|E) = 0.52174, P(H_3|E) = 0.39130$$

Note that before the "experiment" E, the ratio of car to truck drivers was $\frac{P(H_2)}{P(H_3)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$. After we know driver had an accident, the ratio doubles:

$$\frac{P(H_2|E)}{P(H_3|E)} = \frac{0.5217}{0.3913} = \frac{4}{3}$$

$$= \frac{P(E|H_2)}{P(E|H_3)} \times \frac{P(H_2)}{P(H_3)} = \text{Bayes Factor} \times \text{Prior Ratio}$$

Exercise 1

A person uses his car 30% of the time, walks 30% of the time and rides the bus 40% of the time as he goes to work. He is late 10% of the time when he walks; he is late 3% of the time when he drives; and he is late 7% of the time he takes the bus.

a) What is the probability he took the bus if he was late?

b) What is the probability he walked if he is on time?

Answers:

a) 0.418

b) 0.289

Independence

Definition: Two events A and B are *independent* if

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

Definition: Two events A and B are *conditionally* independent given a third event C, if

$$P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$$

Note: By the multiplication law we *always* have that

$$P(A \cap B \mid C) = P(B \mid C)\, P(A \mid B \cap C)$$

Therefore, conditional independence means that

$$P(A \mid C) = P(A \mid B \cap C)$$

Example 2

A box contains two coins: a regular fair coin and a fake two-headed coin with P(H) = 1. A coin is chosen at random and tossed twice. Define the following events.

A = First coin toss results in an H

B = Second coin toss results in an H

C = The regular coin has been selected.

Find $P(A|C)$, $P(B|C)$, $P(A \cap B|C)$, $P(A)$, $P(B)$, and $P(A \cap B)$.

Note that A and B are *not* independent, but they are *conditionally* independent given C.

Example 2 Solution

A box contains two coins: a regular fair coin and a fake two-headed coin with P(H) = 1. A coin is chosen at random and tossed twice. Define the following events.

A = First coin toss results in an H

B = Second coin toss results in an H

C = The regular coin has been selected.

Find P(*A*|*C*), P(*B*|*C*), P(*A*∩*B*|*C*), P(*A*), P(*B*), and P(*A*∩*B*).

Obviously, $P(A|C) = P(B|C) = \frac{1}{2}$. Also, $P(A \cap B|C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

For P(*A*) we need the law of total probability:

$$P(A) = P(C)P(A|C) + P(\bar{C})P(A|\bar{C}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 1 = \frac{3}{4}$$

Similarly, $P(B) \ = \frac{3}{4}$. And for P(*A*∩*B*) we apply again the law of total probability:

$$\begin{aligned} P(A \cap B) &= P(C)P(A \cap B|C) + P(\bar{C})P(A \cap B|\bar{C}) \\ &= P(C)P(A|C)P(B|C) + P(\bar{C})P(A|\bar{C})P(B|\bar{C}) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 1 \cdot 1 = \frac{5}{8} \end{aligned}$$

Exercise 2

Consider rolling a die and let:

$A = \{1, 2\}$,

$B = \{2, 4, 6\}$,

$C = \{1, 4\}$

Show that A and B are independent, but are *not* conditionally independent given C.

Answers:

$$P(A \cap B) = \frac{1}{6}, P(A \cap B|C) = 0$$

Random Variables

Let X be a random variable (r.v.) and $\mathcal{X}$ be the set of all possible values of X. We say that X is *discrete* if the set $\mathcal{X}$ is finite or countable, meaning you can write it as $\mathcal{X} = \{x_1, x_2, \ldots\}$. If $\mathcal{X}$ is uncountably infinite (like an interval), then X is called *continuous*.

Examples:

- Counts are discrete r.v. For example, number of children in a family, or number of years of education.
- Measurements are continuous r.v. For example, height, weight.

Discrete Random Variables

Let X be a discrete random variable. For each $x \in \mathcal{X}$, we write $p(x)$ for $P(X = x)$. The function $p(x)$, $x \in \mathcal{X}$ is called the *probability mass function* (pmf) or *probability density function* (pdf) of X. It satisfies the following properties:

- $0 \le p(x) \le 1$ for all $x \in \mathcal{X}$.
- $\sum_{x \in \mathfrak{X}} p(x) = 1$

Any other property of X can be derived from the pdf. For example, $P(X \in A) = \sum_{x \in A} p(x)$ and $E(X) = \sum_{x \in \mathfrak{X}} x p(x)$.
Often the pdf has parameter(s) θ.

Example 3: Binomial Distribution

Let $\mathcal{X} = \{0, 1, 2, \dots, n\}$, where n is a natural number. Then X is said to have *Binomial distribution* with parameter $\theta \in (0, 1)$ if

$$p(x; \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

Interpretation: X can be thought of as the number of successes in a sequence of n independent trials, each having probability of success θ.

Notation: $X \sim \text{Binom}(n, \theta)$.

In R:

`dbinom(x, size, prob)` gives you $\binom{\text{size}}{x}\text{prob}^x(1-\text{prob})^{\text{size}-x}$

For example, if $n = 8$, $\theta = 0.9$, and $x = 4$, then

$$p(4; 0.9) = \binom{8}{4} 0.9^4 (0.1)^4 = 0.0045927 = \text{dbinom}(4, 8, 0.9)$$

Exercise 3

The probability that Ari will score above a 90 on any mathematics test is 4/5 and he takes the tests independently. What is the probability that he will score above a 90 on exactly three of the four tests this semester?

Answer: 0.4096

Example 4: Poisson Distribution

Let $\mathcal{X} = \{0, 1, 2, \ldots\}$. Then X is said to have a *Poisson distribution* with parameter $\theta > 0$ if

$$p(x; \theta) = \frac{\theta^x e^{-\theta}}{x!}$$

Interpretation: X can be thought of as the number of times an event is happening within a given time or space interval, where θ is the average intensity per unit interval.

Notation: $X \sim \text{Poisson}(\theta)$.

In R:

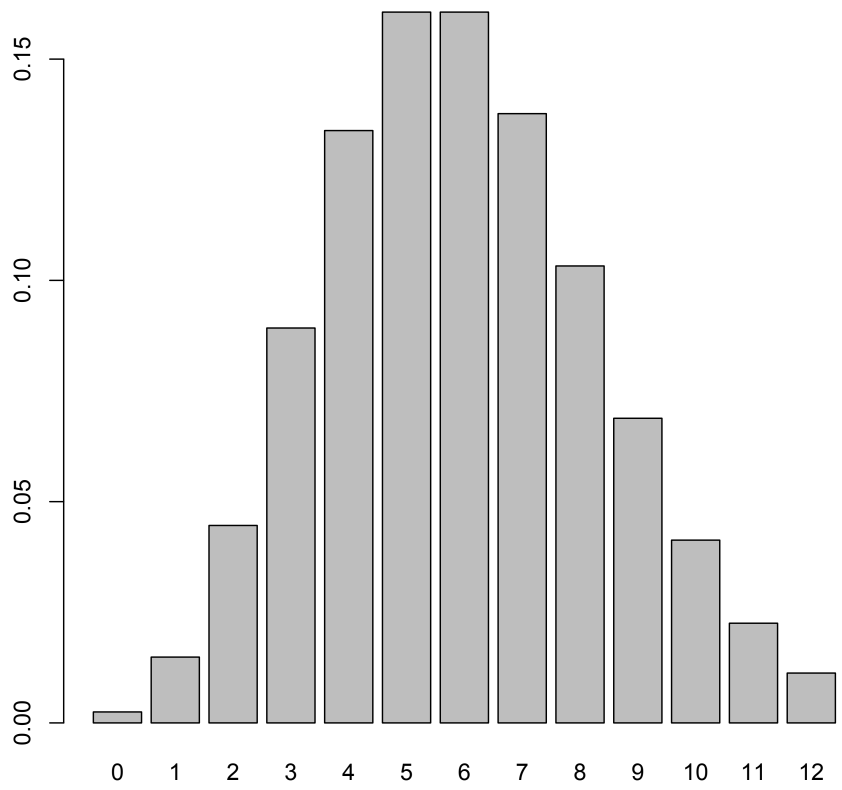
`dpois(x, lambda)` gives you $\dfrac{\text{lambda}^x e^{-\text{lambda}}}{x!}$

For example, if $\theta = 6$, and $x = 11$, then

$$p(11; 6) = \frac{6^{11} e^{-6}}{11!} = 0.02252896 = \text{dpois}(11\ 6)$$

Example 4: Poisson Distribution

The graph of the Poisson pdf when $\theta = 6$ is shown below:



Note: Poisson distribution converges to the normal when $\theta \to \infty$

Exercise 4

Passengers drop by a busy store at an average rate of $\theta = 4$ per minute. If the number of passengers dropping by the store obeys a Poisson distribution, what is the probability that 16 passengers drop by the store in a particular 4-minute period?

Answer: 0.09921753

Continuous Random Variables

Let X be a continuous random variable with a probability density function $f(x)$, $x \in \mathcal{X}$. It satisfies the following properties:

- $0 \leq f(x)$ for all $x \in \mathcal{X}$.
- $\int f(x) = 1$

Often, for such variables we use the *cumulative distribution function* $F(x)$, defined as

$$F(x) = \mathrm{P}(X \leq x),\ x \in \mathcal{X}$$

with the following properties:

- $\mathrm{F}(\infty) = 1$
- $\mathrm{F}(-\infty) = 0$
- $F(a) \leq F(b)$ if $a < b$
- $\mathrm{P}(a < X \leq b) = F(b) - F(a)$
- $F(a) = \int_{-\infty}^{a} f(x)dx$

Example 5: Normal Distribution

Let $\mathcal{X}$ = (-∞, ∞). Then X is said to have a *Normal distribution* with parameter vector $\boldsymbol{\theta} = (\mu, \sigma^2)$ if

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Interpretation: The normal distribution is symmetric with most values clustering around a central region, and values tapering off as they go further away from the center. The mean is μ and the variance is σ^2.

Notation: $X \sim \mathrm{N}(\mu, \sigma^2)$.

In R:

`dnorm(x, mean, sd)` gives you the pdf $\frac{1}{\sqrt{2\pi \mathrm{sd}^2}} e^{-\frac{1}{2}\frac{(x-\mathrm{mean})^2}{\mathrm{sd}^2}}$

`pnorm(q, mean, sd)` gives you the cdf $\int_{-\infty}^{q} \frac{1}{\sqrt{2\pi \mathrm{sd}^2}} e^{-\frac{1}{2}\frac{(x-\mathrm{mean})^2}{\mathrm{sd}^2}} dx$

Example 5: Normal Distribution

For normally distributed variables, you can find probabilities in R or by using the *z*-table.

For example, let $X \sim$ N(-3.4, 4.5) and we are interested in finding P($X \leq$ - 4.5).

In R: `pnorm(-4.5, -3.4, sqrt(4.5)) = 0.302039`

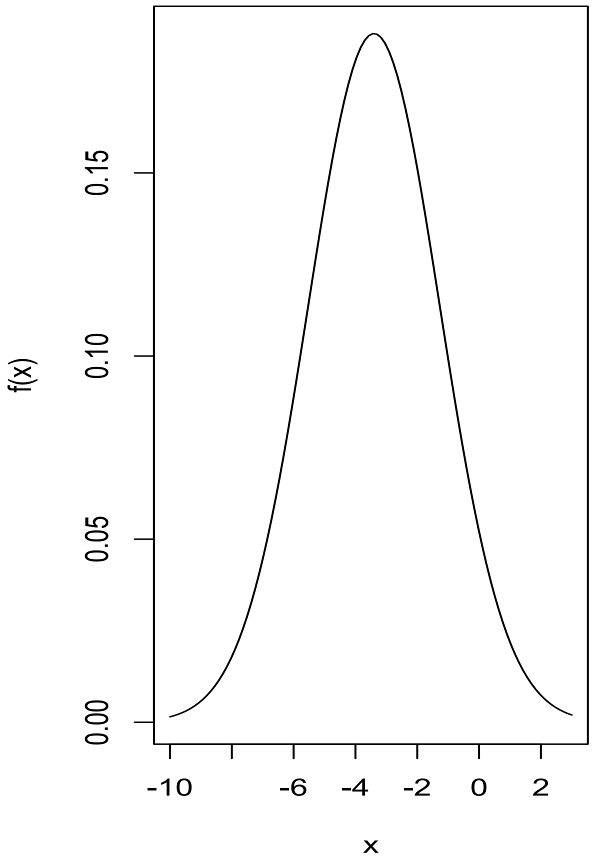
Using the *z*-table we first standardize:

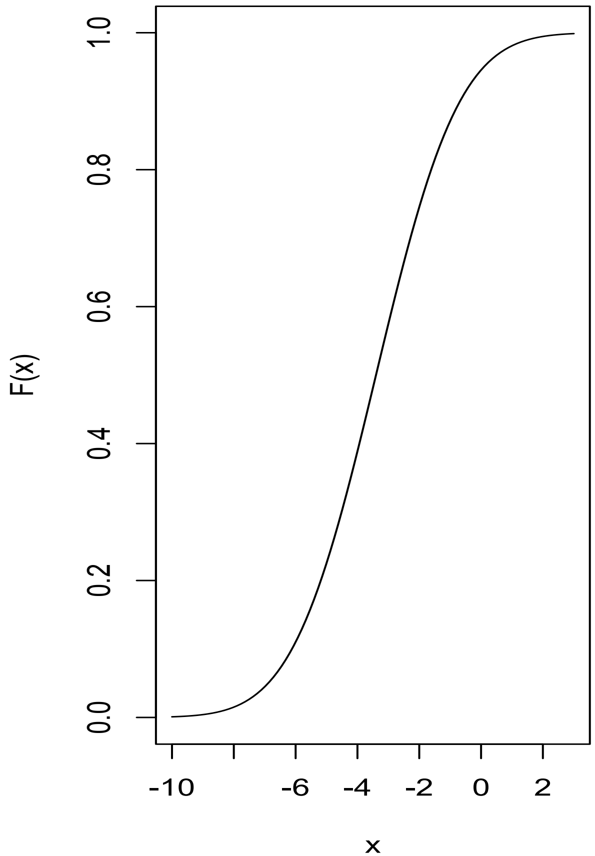
$$P(X \ \leq \ -4.5) = P\left(Z \leq \frac{-4.5 - (-3.4)}{\sqrt{4.5}}\right) = P(Z \leq -0.518545)$$

Then look-up -0.52 in the margins of the table and find the approximate answer inside: .3015

Example 5: Normal Distribution

The graphs of the pdf and cdf of $X \sim$ N(-3.4, 4.5) are shown below:





Exercise 5

If X is a normally distributed variable with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find

a) $P(X < 40)$

b) $P(X > 21)$

c) $P(30 < X < 35)$

Answers:

a) 0.9937903

b) 0.9877755

c) 0.3943502

Characteristics of Random Variables

- The *mean* or *expected value* of X is defined as follows:

  $\mathrm{E}(X) = \sum_{x \in \mathcal{X}} xp(x)$, when X is discrete

  $\mathrm{E}(X) = \int xf(x)dx$, when X is continuous

- The *mode* of X is defined as the value x for which $p(x)$ or $f(x)$ achieves its largest value:

  $$\text{mode} = \operatorname{argmax}_{x \in \mathcal{X}} p(x) \text{ or mode} = \operatorname*{argmax}_{x \in \mathcal{X}} f(x)$$

- The *median* of X is defined as the value m such that $\mathrm{P}(X \leq m) = 0.5$
- The *variance* of X is defined as follows:

  $\mathrm{Var}(X) = \sum_{x \in \mathcal{X}} [x - \mathrm{E}(X)]^2 p(x)$, when X is discrete

  $\mathrm{Var}(X) = \int [x - \mathrm{E}(X)]^2 f(x)dx$, when X is continuous

- The *standard deviation* is defined as $\sqrt{\mathrm{Var}(X)}$

Example 6

- If $X \sim \text{Binom}(n, \theta)$, then

$$\text{E}(X) = n\theta$$
$$\text{Var}(X) = n\theta(1-\theta)$$
$$\text{Mode } = \lfloor (n+1)\theta \rfloor \text{ or } \lceil (n+1)\theta \rceil - 1$$
$$\text{Median } = \lfloor n\theta \rfloor \text{ or } \lceil n\theta \rceil$$

- If $X \sim \text{Poisson}(\theta)$, then

$$\text{E}(X) = \text{Var}(X) = \theta$$
$$\text{Mode } = \lceil \theta \rceil - 1 \text{ or } \lfloor \theta \rfloor$$
$$\text{Median} \approx \left\lfloor \theta - \frac{1}{3} - \frac{1}{50\theta} \right\rfloor$$

- If $X \sim \text{N}(\mu, \sigma^2)$, then

$$\text{E}(X) = \text{Median} = \text{Mode} = \mu$$
$$\text{Var}(X) = \sigma^2$$

Exercise 6

A fair coin is flipped twice. Let X denote the number of heads and Y denote the number of tails. Find:

a) $\mathrm{Var}(X)$

b) $\mathrm{Var}(Y)$

c) $\mathrm{Var}(X + Y)$

Answers:

a) ½

b) ½