Probability I & II
Jiayi Li

Elements of probability
Conditional probability and independence
Exercise 1
Random Variables
Cumulative distribution functions
Probability mass functions
Probability density functions
Expectation
Variation
Exercise 2
Joint and marginal distributions

# Probability I & II

Jiayi Li

Department of Statistics, UCLA

# Elements of probability

In order to define a probability on a set we need a few basic elements,

- Sample space $\Omega$: The set of all the outcomes of a random experiment. Here, each outcome $\omega \in \Omega$ can be thought of as a complete description of the state of the real world at the end of the experiment.
- Set of events (or event space) $\mathcal{F}$: A set whose elements $A \in \mathcal{F}$ (called events) are subsets of $\Omega$.
- Probability measure: A function $P : \mathcal{F} \to \mathbb{R}$ that satisfies the following properties (Axioms of Probability),
  1. $P(A) \geq 0$, for all $A \in \mathcal{F}$
  2. $P(\Omega) = 1$
  3. If $A_1, A_2, \cdots$ are disjoint events (i.e. $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then

$$P(\cup_i A_i) = \sum_i P(A_i)$$

# Elements of probability

Example: Consider the event of tossing a six-sided die. The sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. An event space $\mathcal{F}$ can be the set of all subsets of $\Omega$. One valid probability measure is to assign the probability of each set in the event space to be $\frac{i}{6}$ where $i$ is the number of elements of that set; for example

$$P(\{1, 2, 3, 4\}) = \frac{4}{6}$$

$$P(\{1, 2, 3\}) = \frac{3}{6}$$

Elements of probability
Conditional probability and independence
Exercise 1
Random Variables
Cumulative distribution functions
Probability mass functions
Probability density functions
Expectation
Variation
Exercise 2
Joint and marginal distributions

# Elements of probability

Properties of probability:

- If $A \subseteq B \rightarrow P(A) \leq P(B)$
- $P(A \cap B) \leq \min(P(A), P(B))$
- (Union Bound) $P(A \cap B) \leq P(A) + P(B)$
- $P(\Omega \backslash A) = 1 - P(A)$
- (Law of Total Probability) If $A_1, \cdots, A_k$ are a set of disjoint events such that $\cup_{i=1}^{k} A_i = \Omega$, then $\sum_{i=1}^{k} P(A_k) = 1$

# Conditional probability and independence

Let $B$ be an event with non-zero probability. The conditional probability of any event $A$ given $B$ is defined as,

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

In other words, $P(A|B)$ is the probability measure of the event $A$ after observing the occurrence of event $B$. Two events are called independent if and only if $P(A \cap B) = P(A)P(B)$ (or equivalently, $P(A|B) = P(A)$). Therefore, independence is equivalent to saying that observing $B$ does not have any effect on the probability of $A$.
By law of total probability, if $\cup_{i=1}^{k} B_i = \Omega$, then

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

# Exercise 1

1.1 Two dice are rolled, find the probability that the sum is (a) 1 (b) 4 (c) 13

1.2 A die is rolled and a coin is tossed, find the probability that the die shows an odd number and the coin shows a head.

1.3 You draw a card at random from a standard deck of 52 cards. Find each of the following conditional probabilities.
a. The card is a heart, given that it is black.
b. The card is black, given that it is a heart.
c. The card is an ace, given that it is black.
d. The card is a queen, given that it is a face card.

1.4 A tennis player $A$ has probability of $\frac{2}{3}$ of winning a set against player $B$. A match is won by the player who first wins three sets. Find the probability that $A$ wins the match.

1.5 Employment data at a large company reveal that 72% of the workers are married, that 44% are college graduates, and that half of the grads are married. What is the probability that a randomly chosen worker

a. is neither married nor a college graduate?

b. is married but not a college graduate?

c. is married or a college graduate?

1.6 Suppose that 23% of adults smoke cigarettes. It is known that 57% of smokers and 13% of nonsmokers develop a certain lung condition by age 60. a. Explain how these statistics indicate that lung condition and smoking are not independent. b. What is the probability that a randomly selected 60-year-old has this lung condition?

1.7 Two fair dice are rolled and the sum of the two numbers is observed. What is the probability that a sum of 2 appears before a sum of 6?

# Random Variables

Consider an experiment in which we flip 10 coins, and we want to know the number of coins that come up heads. Here, the elements of the sample space $\Omega$ are 10-length sequences of heads and tails. For example, we have $w_0 = \langle H, H, T, H, T, H, H, T, T, T \rangle \in \Omega$.

In practice, we usually do not care about the probability of obtaining any particular sequence of heads and tails. Instead we usually care about real-valued functions of outcomes, such as the number of heads that appear among our 10 tosses, or the length of the longest run of tails. These functions, under some technical conditions, are known as random variables.

More formally, a random variable $X$ is a function $X : \Omega \to \mathbb{R}$. Typically, we will denote random variables using upper case letters $X(\omega)$ or more simply $X$ (where the dependence on the random outcome $\omega$ is implied). We will denote the value that a random variable may take on using lower case letters $x$.

# Random Variables

In our experiment above, suppose that $X(\omega)$ is the number of heads which occur in the sequence of tosses $\omega$. Given that only 10 coins are tossed, $X(\omega)$ can take only a finite number of values, so it is known as a discrete random variable. Here, the probability of the set associated with a random variable $X$ taking on some specific value $k$ is

$$P(X = k) := P(\{\omega : X(\omega) = k\})$$

Suppose that $X(\omega)$ is a random variable indicating the amount of time it takes for a radioactive particle to decay. In this case, $X(\omega)$ takes on a infinite number of possible values, so it is called a continuous random variable. We denote the probability that $X$ takes on a value between two real constants $a$ and $b$ (where $a < b$) as

$$P(a \leq X \leq b) := P(\{\omega : a \leq X(\omega) \leq b\})$$

# Cumulative distribution functions

In order to specify the probability measures used when dealing with random variables, it is often convenient to specify alternative functions (CDFs, PDFs, and PMFs).

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0,1]$ which specifies a probability measure as

$$F_X(x) := P(X \leq x)$$

CDF has the following properties:

- $0 \leq F_X(x) \leq 1$
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to \infty} F_X(x) = 1$
- $x \leq y \to F_X(x) \leq F_X(y)$

# Probability mass functions

When a random variable $X$ takes on a finite set of possible values (i.e., $X$ is a discrete random variable), a simpler way to represent the probability measure associated with a random variable is to directly specify the probability of each value that the random variable can assume. In particular, a probability mass function (PMF) is a function $p_X : \Omega \to \mathbb{R}$ such that

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

In the case of discrete random variable, we use the notation $\text{Val}(X)$ for the set of possible values that the random variable X may assume. For example, if $X(\omega)$ is a random variable indicating the number of heads out of ten tosses of coin, then $\text{Val}(X) = \{0, 1, 2, \cdots, 10\}$.

PMF has the following properties:

- $0 \leq p_X(x) \leq 1$
- $\sum_{x \in \text{Val}(X)} p_X(x) = 1$
- $\sum_{x \in A} p_X(x) = P(X \in A)$

# Probability density functions

For some continuous random variables, the cumulative distribution function $F_X(x)$ is differentiable everywhere. In these cases, we define the Probability Density Function or PDF as the derivative of the CDF, i.e.,

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Note here, that the PDF for a continuous random variable may not always exist (i.e., if $F_X(x)$ is not differentiable everywhere). According to the properties of differentiation, for very small $\Delta x$,

$$P(x \leq X \leq x + \Delta x) \approx f_X(x)\Delta x$$

# Probability density functions

Both CDFs and PDFs can be used for calculating the probabilities of different events.

The value of PDF at any given point $x$ is not the probability of that event, i.e., $f_X(x) \neq P(X = x)$.

PDF has the following properties:

- $f_X(x) \leq 0$
- $\int_{-\infty}^{\infty} f_X(x) = 1$
- $\int_{x \in A} f_X(x)dx = P(X \in A)$

# Expectation

Suppose that $X$ is a discrete random variable with PMF $p_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. In this case, $g(X)$ can be considered a random variable, and we define the expectation or expected value of $g(X)$ as

$$\mathbb{E}[g(X)] := \sum_{x \in \mathrm{Val}(x)} g(x) p_X(x)$$

If $X$ is a continuous random variable with PDF $f_X(x)$, then the expected value of $g(X)$ is defined as,

$$\mathbb{E}[g(X)] := \int_{\infty}^{\infty} g(x) f_X(x) dx$$

# Expectation

Properties of expectation

- $\mathbb{E}(a) = a$ for any constant $a \in \mathbb{R}$
- $\mathbb{E}(af(X)) = a\mathbb{E}(f(X))$ for any constant $a \in \mathbb{R}$
- $\mathbb{E}(f(X) + g(X)) = \mathbb{E}(f(X)) + \mathbb{E}(g(X))$

# Variation

The variance of a random variable $X$ is a measure of how concentrated the distribution of a random variable $X$ is around its mean. Formally, the variance of a random variable $X$ is defined as

$$\mathbb{V}(X) := \mathbb{E}(X - \mathbb{E}(X))^2$$

Using the properties of expectation, we have

$$\mathbb{V}(X) := \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

- $\mathbb{V}(a) = 0$ for any constant $a \in \mathbb{R}$
- $\mathbb{V}(af(X)) = a^2\mathbb{V}(f(X))$ for any constant $a \in \mathbb{R}$

Elements of probability

Conditional probability and independence

Exercise 1

Random Variables

Cumulative distribution functions

Probability mass functions

Probability density functions

Expectation

Variation

Exercise 2

Joint and marginal distributions

# Exercise 2

2.1 A worker can arrive to the workplace at any moment between 6 and 7 in the morning with the same likelihood.
Compute and plot the probability density function of the variable that measures the arrival time.
Compute and plot the distribution function.
Compute the probability of arriving before quarter past six and after half past six.
What is the expected arrival time?
2.2 Two balls are chosen randomly without replacement from an urn containing 8 white, 4 black, and 2 orange balls. Suppose that we win \$2 for each black ball selected and we lose \$1 for each white ball selected. We neither win nor we lose any money for selecting an orange ball. Let $X$ denote our winnings.
a. What is the expected value of $X$?
b. What is the variance of $X$?
c. Given that our winnings are negative, what is the probability that we lost exactly \$2?

# Exercise 2

2.3 Tay-Sachs disease is a rare fatal genetic disease occurring chiefly in children, especially of Jewish or Slavic extraction. Suppose that we limit ourselves to families which have (a) exactly three children, and (b) which have both parents carrying the Tay-Sachs disease. For such parents, each child has independent probability $\frac{1}{4}$ of getting the disease.

Write $X$ to be the random variable representing the number of children that will have the disease.

(a) Show that the probability distribution for $X$

(b) Find the expectation and variance of $X$

2.4 Find the probability that none of the three bulbs in a set of traffic lights will have to be replaced during the first 1200 hours of operation if the lifetime $X$ of a bulb (in thousands of hours) is a random variable with probability density function $f(x) = 6[0.25 - (x-1.5)^2]$ when $1 \le x \le 2$ and $f(x) = 0$ otherwise. You should assume that the lifetimes of different bulbs are independent.

# Exercise 2

2.5 Let $X \sim U(0,1)$. Find $\mathbb{E}(X)$, $\mathbb{V}(X)$.

2.6 $f_X(x) = \lambda e^{-\lambda x}$. Find $\mathbb{E}(X)$, $\mathbb{V}(X)$.

# Joint and marginal distributions

Suppose that we have two random variables $X$ and $Y$. One way to work with these two random variables is to consider each of them separately. If we do that we will only need $F_X(x)$ and $F_Y(y)$. But if we want to know about the values that $X$ and $Y$ assume simultaneously during outcomes of a random experiment, we require a more complicated structure known as the joint cumulative distribution function of $X$ and $Y$ , defined by

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

The joint CDF $F_{XY}(x, y)$ and the joint distribution functions $F_X(x)$ and $F_Y(y)$ of each variable separately are related by

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x, y)dy$$

$$F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y)dx$$

# Joint and marginal distributions

Here, we call $F_X(x)$ and $F_Y(y)$ the marginal cumulative distribution functions of $F_{XY}(x, y)$.

- $0 \leq F_{XY}(x, y) \leq 1$
- $\lim_{x,y\to\infty} F_{XY}(x, y) = 1$
- $\lim_{x,y\to-\infty} F_{XY}(x, y) = 0$

If $X$ and $Y$ are discrete random variables, then the joint probability mass function $p_{XY}[0, 1]$ is defined by

$$p_{XY}(x, y) = P(X = x, Y = y)$$

Here, $0 \leq P_{XY}(x, y) \leq 1$ for all $x, y$, and $\sum_{x\in \text{Val}(X)} \sum_{y\in \text{Val}(Y)} P_{XY}(x, y) = 1$

Elements of probability
Conditional probability and independence
Exercise 1
Random Variables
Cumulative distribution functions
Probability mass functions
Probability density functions
Expectation
Variation
Exercise 2
Joint and marginal distributions

# Joint and marginal distributions

$$p_X(x, y) = \sum_y p_{XY}(x, y)$$

and similarly for $p_Y(y)$. In this case, we refer to $p_X(x)$ as the marginal probability mass function of $X$.

Let $X$ and $Y$ be two continuous random variables with joint distribution function $F_{XY}$ . In the case that $F_{XY}(x, y)$ is everywhere differentiable in both $x$ and $y$, then we can define the joint probability density function,

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

Elements of probability
Conditional probability and independence
Exercise 1
Random Variables
Cumulative distribution functions
Probability mass functions
Probability density functions
Expectation
Variation
Exercise 2
Joint and marginal distributions

# Joint and marginal distributions

Like in the single-dimensional case, $f_{XY}(x, y) \neq P(X = x, Y = y)$, but rather

$$\int \int_{x \in A} f_{XY}(x, y) dxdy = P((X, Y) \in A)$$

Note that the values of the probability density function $f_{XY}(x, y)$ are always non-negative, but they may be greater than 1.

Analagous to the discrete case, we define

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy$$

as the marginal probability density function (or marginal density) of $X$, and similarly for $f_Y(y)$.