CS 660: Mathematical Foundations of Analytics

Dr. Francis Parisi

Pace University

Spring 2019

IMS Chapter 1 – Probability and Distributions

Probability Distributions

Roadmap

- We'll start with a discussion of **Sets**
- This leads to **Functions on Sets**
- Set functions provide the basis for **Concepts in Probability**
- From here we develop a **Theory of Probability**
- We have the tools for a **Model for Random Experiments**
- Which allows us to make **Inferences** about a population from observed data

Probability Distributions

Basic Concepts in Probability

- A common interpretation of probability is the long run frequency of an event – this is a frequentist view
- Suppose we repeat a game many times and count the number of times we win; the number of wins divided by the number of games is the probability of winning

Definition

An **experiment** is a procedure that results in one of a number of possible outcomes. The set of all possible outcomes of an experiment is called the **sample space** for the experiment. Each subset of a sample space is called an **event.**

- We denote the sample space by $\mathcal{C}$ and an event by E, and $E \subseteq \mathcal{C}$
- If $\mathcal{C}$ is a sample space and $c \in \mathcal{C}$ then c is an **outcome**

Example (1.1.1)

In the toss of a fair coin, let tails be denoted by T and let heads be denoted by H. If we assume that the coin may be repeatedly tossed under the same conditions, then the toss of this coin is an example of a *random experiment* in which the outcome is one of the two symbols T and H; that is, the sample space is the collection of these two symbols. $\mathcal{C} = \{H, T\}$.

Example (1.1.2)

In the cast of one red die and one white die, let the outcome be the ordered pair (number of spots up on the red die, number of spots up on the white die). If we assume that these two dice may be repeatedly cast under the same conditions, then the cast of this pair of dice is a random experiment. The sample space consists of the 36 ordered pairs:
$\mathcal{C} = \{(1,1), \ldots, (1,6), (2,1), \ldots, (2,6), \ldots, (6,6)\}$.

Probability as Relative Frequency

Example (1.1.3)

Let $\mathcal{C}$ denote the sample space of Example 1.1.2 and let $\mathcal{B}$ be the collection of every ordered pair of $\mathcal{C}$ for which the sum of the pair is equal to seven.
That is, $\mathcal{B} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$.

Suppose the dice are cast $N = 400$ times and f, the frequency of a sum of seven, is $f = 60$. Then the *relative frequency* with which the outcome was in $\mathcal{B}$ is $f/N = \frac{60}{400} = 0.15$. Thus we might associate with $\mathcal{B}$ a number p that is close to 0.15, and p would be called the *probability of the event* $\mathcal{B}$.

Probability and Distributions

Sets

- A **Set** is a collection of objects
- We're concerned with sets of numbers and often will speak of the set of points
- Sets may be **countable** or **uncountable**
- A set is countable if it is finite or can be put in one-to-one correspondence with $\mathbb{N}$
- For example, the set of integers, $\mathbb{Z}$, is countable but the set of real numbers, $\mathbb{R}$, is not

Probability and Distributions

Sets

If $\mathcal{C}$ denotes the sample space of an experiment, then the events are subsets of $\mathcal{C}$

Some Important Definitions

Complement The complement of an event A is the set of all elements in $\mathcal{C}$ that are not in A. We denote this as A^c; $A^c = \{x \in \mathcal{C} : x \notin A\}$.

Subset If each element of a set A is also an element of set B, the set A is called a subset of the set B. This is indicated by writing $A \subset B$. If and also $B \subset A$, the two sets have the same elements, and this is indicated by writing $A = B$.

Union The set of all elements that belong to at least one of the sets A and B is called the union of A and B. The union of A and B is indicated by writing $A \cup B$.

Sets

Some Important Definitions

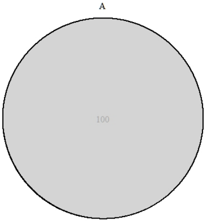
Intersection The set of all elements that belong to each of the sets A and B is called the intersection of A and B. The intersection of A and B is indicated by writing $A \cap B$.

Disjoint sets For two sets A and B, A and B are disjoint if $A \cap B = \emptyset$.

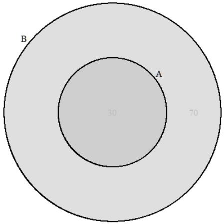
Null Set A set with no elements is the **null** or **empty** set denoted $\emptyset$ or $\{\}$

Probability and Distributions

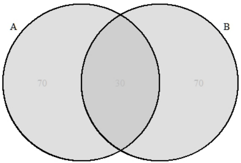
Venn Diagrams



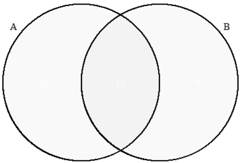
(a) A



(b) $A \subset B$



(c) $A \cup B$



(d) $A \cap B$

Probability and Distributions

For any three sets A, B, and C, the following laws hold:

Commutative Laws

$$A \cap B = B \cap A \quad \text{and} \quad A \cup B = B \cup A$$

Associative Laws

$$(A \cap B) \cap C = A \cap (B \cap C) \quad \text{and} \quad (A \cup B) \cup C = A \cup (B \cup C)$$

Distributive Laws

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

De Morgan's Laws

$$(A \cup B)^c = A^c \cap B^c \quad \text{and} \quad (A \cap B)^c = A^c \cup B^c$$

Probability and Distributions

Unions and Intersections of more than Two Sets

Suppose we have n sets $A_1, A_2, \ldots, A_n$, then

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i = \{x : x \in A_i \text{ for some } i = 1, 2, \ldots, n\}$$

and

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i = \{x : x \in A_i \text{ for all } i = 1, 2, \ldots, n\}$$

Probability and Distributions

Countable Unions and Intersections

For an infinite number of sets $A_1, A_2, \ldots$, we have

$$A_1 \cup A_2 \cup \cdots = \bigcup_{n=1}^{\infty} A_n$$

and

$$A_1 \cap A_2 \cap \cdots = \bigcap_{n=1}^{\infty} A_n$$

Probability and Distributions

Set Functions

- We often work with functions from real numbers to real numbers
- For our purpose we are concerned with functions from sets to real numbers
- These are set functions

Example

Let $\mathcal{C} = \mathbb{R}$, the set of real numbers. For some subset $A \in \mathcal{C}$ define the function $Q(A)$ as the number of elements in A corresponding to the positive integers. Then $Q(A)$ is a set function on the set A.

(i) If $A = \{x : 0 < x < 5\}$ then $Q(A) = 4$

(ii) If $A = \{x : -2 < x < 4\}$ then $Q(A) = 3$

(iii) If $A = \{x : -\infty < x < 6\}$ then $Q(A) = 5$

Probability and Distributions

Many set functions in probability are defined on integrals and sums

$\int_A f(x)dx$ integral over a one-dimensional set A

$\iint_A g(x,y)dxdy$ integral over a two-dimensional set A

$\sum_A f(x)$ sum over all $x \in A$

$\sum_A g(x,y)$ sum over all pairs $(x,y) \in A$

Probability and Distributions

Recall that for a geometric series, if $|a| < 1$

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a} \tag{1.1}$$

Example

For a set $A \subset \mathcal{C}$ the set of non-negative integers, define the set function

$$Q(A) = \sum_{n \in A} \left(\frac{2}{3}\right)^n$$

(i) It follows from (1.1) that $Q(\mathcal{C}) = 3$

(ii) Suppose we have $B = \{1, 3, 5, \ldots\}$ the set of odd positive integers

Probability and Distributions

Finding $Q(B)$ follows:

$$Q(B) = \sum_{n \in B} \left(\frac{2}{3}\right)^n = \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^{2n+1}$$
$$= \frac{2}{3} \sum_{n=0}^{\infty} \left[\left(\frac{2}{3}\right)^2\right]^n = \frac{2}{3}\left[\frac{1}{1-(4/9)}\right] = \frac{6}{5}$$

Similarly for an interval, let's define $Q(A) = \int_A e^{-x} dx \ = -e^{-x}|_A$

$Q[(1,3)] = \int_1^3 e^{-x} dx = -e^{-x}\Big|_1^3 = e^{-1} - e^{-3} \doteq 0.3181$

Definition

Let $\mathcal{C}$ be a sample space and $\mathcal{B}$ be the set of events. Let P be a real-valued function defined on $\mathcal{B}$. Then P is a **probability set function** if P satisfies the following conditions:

1. $P(A) \geq 0, \quad \forall\, A \in \mathcal{B}$
2. $P(\mathcal{C}) = 1$
3. If $\{A_n\}$ is a sequence of events in $\mathcal{B}$ and $A_m \cap A_n = \emptyset$ for all $m \neq n$, then
$$p\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Probability and Distributions

Some terminology

- A collection of pairwise disjoint events is a **mutually exclusive** collection, or a **disjoint union**
- If the union of events is the entire sample space then the collection is **exhaustive**
- A mutually exclusive and exhaustive collection is a partition of $\mathcal{C}$

Probability and Distributions

Some properties of P

- For each $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$
- $P(\emptyset) = 0$
- If A and B are events such that $A \subset B$ then $P(A) \leq P(B)$
- For each $A \in \mathcal{B}$, $0 \leq P(A) \leq 1$
- For $A, B \in \mathcal{C}$, $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probability and Distributions

Exercise

Let $\mathcal{C}$ denote the sample space of the outcomes from rolling two die (Example 1.1.2). P assigns a probability of $\frac{1}{36}$ to each outcome. If $C_1 = \{(1,1),(2,1),(3,1),(4,1),(5,1)\}$ and $C_2 = \{(1,2),(2,2),(3,3)\}$. Find the following probabilities:

(a) $P(C_1)$

(b) $P(C_2)$

(c) $P(C_1 \cap C_2)$

(d) $P(C_1 \cup C_2)$

Probability and Distributions

Exercise

Two coins are to be tossed and the outcome is the ordered pair (face on the first coin, face on the second coin). Thus the sample space may be represented as $\mathcal{C} = \{(H,H),(H,T),(T,H),(T,T)\}$. Let the probability set function assign a probability of 1/4 to each element of $\mathcal{C}$. Let $C_1 = \{(H,H),(H,T)\}$ and $C_2 = \{(H,H),(T,H)\}$. What is

- $P(C_1), P(C_2)$
- $P(C_1 \cup C_2)$
- $P(C_1 \cap C_2)$

Probability and Distributions

Suppose we have a finite sample space $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$

Let $p_1, p_2, \ldots, p_m$ be fractions such that $0 \leq p_i \leq 1$ and $\sum p_i = 1$

If we consider the roll of a fair die, then $p_1 = p_2 = \cdots = p_6 = \frac{1}{6}$
this is an example of the *equilikely case*

Probability and Distributions

Definition (Equilikely Case)

Let $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$ be a finite sample space. Let $p_i = 1/m$ for all i and for all subsets A of $\mathcal{C}$ define

$$P(A) = \sum_{x_i \in A} \frac{1}{m} = \frac{\#(A)}{m}$$

where $\#(A)$ denotes the number of elements in A. The P is a probability on $\mathcal{C}$ and is the **equilikely case.**

$\#(A)$ is also know as the cardinality of the set A

We see this often – flipping a coin, drawing a card, rolling dice

The Multiplication Rule Let $A = \{x_1, x_2, \ldots, x_m\}$ be a set with m elements and $B = \{y_1, y_2, \ldots, y_n\}$ be a set with n elements. Together there are mn ordered pairs, with the first from A and second from B

Consider a procedure consisting of a sequence of two tasks. To perform this procedure, one performs the first task followed by performing the second task. If there are m ways to perform the first task and n ways to perform the second task after the first task has been performed, then there are mn ways to perform the procedure.

Probability and Distributions

The Addition Rule When we have a procedure that can be done with either of two tasks but the two tasks that cannot be performed at the same time, the procedure is done when either of the two is done.

A procedure consists of two tasks that cannot be performed simultaneously. To perform this procedure, either of the two tasks is performed. If the first task can be performed in m ways and the second can be performed in n ways, then the number of ways of performing this procedure is $m + n$.

Definition

A **permutation** of a nonempty set S is an arrangement or ordered list of the elements of S. We denote a permutation of k elements from a set of n elements as

$$P_k^n = n(n-1)\cdots(n-(k-1)) = \frac{n!}{(n-k)!}$$

Example

Take the set $S = \{1, 2, 3, 4\}$.
There are four possible numbers for the first position, three for he second, two for the third and one for the final number
The total number of permutations is then by the multiplication rule $4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$.

Probability and Distributions

Combinations

Permutations are ordered but when order does not matter then we are interested in *combinations*

Example

Consider the set $A = \{1, 2, 3, 4\}$. How many subsets contain exactly 2 elements?

There are six two-element subsets

$$\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\},$$

We denote combinations by

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

this is the number of ways to choose k items from n items

If we expand the binomial series we find that $\binom{n}{k}$ represents the coefficient for the k^{th} term

$$
\begin{aligned}
(a+b)^n &= (a+b)(a+b)\cdots(a+b) \\
&= \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}
\end{aligned}
$$

Thus $\binom{n}{k}$ is known as the binomial coefficient

Exercise

We draw five cards from a standard deck of 52 cards. What is the probability of getting three kings and two queens? What is the probability of getting a full house?

- Conditional probability deals with the probability of an event from a subset of the sample space
- $P(B|A)$ is the probability of the event B given A has occurred (i.e., relative to the sample space A)

Probability and Distributions

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

where

1. $P(B|A) \geq 0$
2. $P(A|A) = 1$
3. $P\left(\bigcup_{n=1}^{\infty} B_n|A\right) = \sum_{n=1}^{\infty} P(B_n|A)$ provided the B_i are mutually exclusive

Probability and Distributions

Example

A hand of five cards is to be dealt at random without replacement from an ordinary deck of 52 playing cards. The conditional probability of an all spade hand (B), relative to the hypothesis that there are at least four spades in the hand (A) is, since $A \cap B = B$

$$\begin{aligned}
P(B|A) &= \frac{P(B)}{P(A)} \\
&= \frac{\binom{13}{5}/\binom{52}{5}}{\left[\binom{13}{4}\binom{39}{1} + \binom{13}{5}\right]/\binom{52}{5}} \\
&= \frac{\binom{13}{5}}{\left[\binom{13}{4}\binom{39}{1} + \binom{13}{5}\right]} = 0.0441
\end{aligned}$$

Probability and Distributions

From $P(B|A) = \frac{P(A \cap B)}{P(A)}$ we derive the multiplication rule for probabilities

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

$$\begin{aligned} P(A \cap B \cap C) &= P([A \cap B] \cap C) \\ &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

The multiplication rule lets us extend this to four or more events

Probability and Distributions

For k mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_k, P(A_i) > 0$, form a partition of $\mathcal{C}$. For another event $B, P(B) > 0$ which occurs with only one A_i

$$B = B \cap (A_1 \cup A_2 \cup \cdots A_k).$$

From the distributive law we get $B \cap A_i$, and they are mutually exclusive so

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k)$$

Probability and Distributions

From the multiplication rule $P(B \cap A_i) = P(A_i)P(B|A_i)$ we get

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k) \\ &= \sum_{i=1}^{k} P(A_i)P(B|A_i) \end{aligned}$$

This is the **law of total probability** which leads to Bayes' Theorem

Theorem

Let $A_1, A_2, \ldots, A_k$ be events such that $P(A_i) > 0$. Assume the A_i form a partition of the sample space $\mathcal{C}$. For any event B

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)} \tag{1.2}$$

Probability and Distributions

Example

Say it is known that bowl A_1 contains 3 red and 7 blue chips and bowl A_2 contains 8 red and 2 blue chips. All chips are identical in size and shape.

A die is cast and bowl A_1 is selected if 5 or 6 spots show on the side that is up; otherwise, bowl A_2 is selected. It seems reasonable to assign $P(A_1) = 2/6$ and $P(A_2) = 4/6$.

The selected bowl is handed to another person and one chip is taken at random. Say that this chip is red, an event which we denote by B. By considering the contents of the bowls, it is reasonable to assign the conditional probabilities $P(B|A_1) = 3/10$ and$P(B|A_2) = 8/10$.

Probability and Distributions

Thus the conditional probability of bowl A_1, given that a red chip is drawn, is

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$
$$= \frac{(\frac{2}{6})(\frac{3}{10})}{(\frac{2}{6})(\frac{3}{10}) + (\frac{4}{6})(\frac{8}{10})} = \frac{3}{19}$$

Similarly we have $P(A_2|B) = 16/19$

Probability and Distributions

In the previous example, $P(A_1)$ and $P(A_2)$ are **prior probabilities**: probabilities known because of the state of the world *before* the 'new' information

$P(A_1|B)$ and $P(A_2|B)$ are **posterior probabilities**: updated probabilities of the events given some new knowledge

Exercise

Three plants, $A_1, A_2,$ and $A_3,$ produce respectively, 10%, 50%, and 40% of a company's output. Only 1% of the products from plant A_1 are defective, 3% of the products from plant A_2 are defective, and 4% of the products from plantA_3 are defective. One item is selected at random from all the products and observed to be defective, say event B. What is the probability the defective item came from plant A_1?

Solution

We have as prior probabilities, $P(A_1) = 0.1, P(A_2) = 0.5$ and $P(A_3) = 0.4$.

The conditional probabilities of getting a defect are $P(B|A_1) = 0.01, P(B|A_2) = 0.03$ and $P(B|A_3) = 0.04$.

The probability is then

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{(0.10)(0.01)}{(0.1)(0.01) + (0.5)(0.03) + (0.4)(0.04)} = \frac{1}{32}$$

- For two events A and B, if the occurrence of A does not change the probability of B then A and B are said to be independent
- We have $P(B|A) = P(B)$ and the multiplication rule becomes $P(A \cap B) = P(A)P(B|A) = P(A)P(B)$

Definition

Let A and B be two events. We say that A and B are independent if $P(A \cap B) = P(A)P(B)$.

Probability and Distributions

For three events A_1, A_2, A_3, we say they are **mutually independent** if and only if they are pairwise independent

That is, if $P(A_1 \cap A_2) = P(A_1)P(A_2)$, $P(A_1 \cap A_3) = P(A_1)P(A_3)$, $P(A_2 \cap A_3) = P(A_2)P(A_3)$, then
$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$

This extends to k mutually independent events

Probability and Distributions

Exercise

A coin is flipped independently several times. Let the event A_i represent a head (H) on the i^{th} toss; thus A_i^c represents a tail (T). For a fair coin $P(A_i) = P(A_i^c) = 1/2$. Find the probability of the ordered sequence HHTH.

Solution

Since the events are independent we have

$$P(A_1 \cap A_2 \cap A_3^c \cap A_4) = P(A_1)P(A_2)P(A_3^C)P(A_4) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

Probability and Distributions

Exercise

A computer system is built so that if component K_1 fails, it is bypassed and K_2 is used. If K_2 fails, then K_3 is used. Suppose that the probability that K_1 fails is 0.01, that K_2 fails is 0.03, and that K_3 fails is0.08. Assume that the failures are mutually independent events. Find the probability of failure of the system.

Solution

The events are independent so

$$\begin{aligned} P(K_1 \cap K_2 \cap K_3) &= P(K_1)P(K_2)P(K_3) \\ &= (0.01)(0.03)(0.08) = 0.000024 \end{aligned}$$

Random variables represent the elements of a sample space as real numbers or more precisely

Definition

Consider a random experiment with a sample space $\mathcal{C}$. A function X, which assigns to each element $c \in \mathcal{C}$ one and only one number $X(c) = x$, is called a **random variable.** The **space** or **range** of X is the set of real numbers $\mathcal{D} = \{x : x = X(c), c \in \mathcal{C}\}$.

Probability and Distributions

In the definition, $\mathcal{D}$ is a countable set, or it is an interval of real numbers

When $\mathcal{D}$ is a countable set we say the random variables are **discrete**

When $\mathcal{D}$ is an interval of real numbers we say the random variables are **continuous**

Probability and Distributions

- For some random variable X $\mathcal{D}$ is its range or the sample space of interest
- We say X *induces* the sample space $\mathcal{D}$
- and, X induces a probability called the **distribution** of X

Probability and Distributions

- First let's consider the case where X is a discrete random variable with a finite space $\mathcal{D} = \{d_1, \ldots, d_m\}$, we are interested in subsets of $\mathcal{D}$
- Define the function $p_X(d_i)$ on $\mathcal{D}$ by

$$p_X(d_i) = P\left[\{c : X(c) = d_i\}\right], \quad \text{for} \quad i = 1, 2, \ldots, m$$

- The induced probability distribution, $P_X(\cdot)$, of X is

$$P_X(D) = \sum_{d_i \in D} p_X(d_i), D \subset \mathcal{D}$$

$p_X(d_i)$ is the **probability mass function** (**pmf**) of X

Probability and Distributions

- Now let's consider the case where X is a continuous random variable
- For continuous random variables he events of interest are intervals
- Given a nonnegative function $f_X(x)$ and any interval of real numbers $(a, b) \in \mathcal{D}$, the induced probability distribution of X, $P_X(\cdot)$ is defined as

$$P\left[(a,b)\right] = P\left[\{c \in \mathcal{C} : a < X(c) < b\}\right] = \int_a^b f_X(x)dx \quad (1.3)$$

- What (1.3) says is, the probability that X is between a and b is the area under the curve $y = f_X(x)$ between a and b

f_X is the **probability density function** (**pdf**) of X
$f_X(x) \geq 0$ and $P_X(\mathcal{D}) = \int_{\mathcal{D}} f_X(x)dx = 1$

Probability and Distributions

Example (1.5.2)

For an example of a continuous random variable, consider the following simple experiment: choose a real number at random from the interval $(0,1)$. Let X be the number chosen. In this case the space of X is$\mathcal{D} = (0,1)$. Because the number is chosen at random, it is reasonable to assign

$$P_X\left[(a,b)\right] = b - a, \quad \text{for} \quad 0 < a < b < 1$$

The pdf of X is

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases} \tag{1.4}$$

Example (1.5.2 cont'd)

What if we want to find the probability that $X < 1/8$ or $X > 7/8$?

$$P\left[\left\{X < \frac{1}{8}\right\} \cup \left\{X > \frac{7}{8}\right\}\right] = \int_0^{\frac{1}{8}} dx + \int_{\frac{7}{8}}^{1} dx = \frac{1}{4}$$

Probability and Distributions

We have talked about the pmf of a discrete random variable, and the pdf of a discrete random variable

Now we introduce the **cumulative distribution function** (cdf) of a random variable

Definition (Cumulative Distribution Function)

Let X be a random variable. Then its **cumulative distribution function** (**cdf**) is defined by $F_X(x)$, where

$$F_X(x) = P_X\left[(-\infty, x]\right] = P\left[\{c \in \mathcal{C} : X(c) \leq s\}\right]$$

Theorem (1.5.1)

Let X be a random variable with cumulative distribution function $F(x)$. Then

(a) *For all a and b, if $a < b$, then $F(a) \leq F(b)$ (F is nondecreasing)*

(b) $\lim_{x \to -\infty} F(x) = 0$ *(the lower limit of F is 0)*

(c) $\lim_{x \to \infty} F(x) = 1$ *(the upper limit of F is 1)*

(d) $\lim_{x \downarrow x_0} F(x) = F(x_0)$ *(F is right continuous).*

Probability and Distributions

We will use the following theorem when evaluating probabilities using cdfs

Theorem (1.5.2)

Let X *be a random variable with the cdf* F_X*. Then for* $a < b$,
$P\left[a < X \leq b\right] = F_X(b) - F_X(a)$

Probability and Distributions

Example (1.5.5)

Let X be the lifetime in years of a mechanical part. Assume that X has the cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & 0 \leq x \end{cases}$$

The derivative of F_X gives us the pdf of X

$$f_X(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

The probability that a part lasts between one and three years is

$$P\left[1 < X < 3\right] = F_X(3) - F_X(1) = \int_1^3 e^{-x} dx = 0.318$$

Probability and Distributions

By the law of total probability we have

$$\sum_{x \in \mathcal{D}} p_X(x) = 1 \quad \text{and} \quad \int_{\mathcal{D}} f_X(x)dx = 1$$

Exercise

Suppose X has the pdf

$$f_X(x) = \begin{cases} cx^3 & 0 < x < 2 \\ 0 & \text{otherwise,} \end{cases}$$

for some constant c. Find c.

In our discussion of random variables we identified two types: the first of these is the **Discrete Random Variable**

Definition

A random variable is a **Discrete Random Variable** if its space is either finite or countable

Probability and Distributions

Example

We flip a fair coin repeatedly. What is the probability that the first head appears on an odd flip?

Solution

The first head on an odd flip is the set $X = \{1, 3, 5, \ldots\}$ and the probability is

$$P\left[X \in \{1, 3, 5, \ldots\}\right] = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^{2x-1} = \frac{1}{2} \sum_{x=1}^{\infty} \left(\frac{1}{4}\right)^{x-1} = \frac{1/2}{1-(1/4)} = \frac{2}{3}$$

Probability and Distributions

Definition (Probability Mass Function (pmf))

Let X be a discrete random variable with space $\mathcal{D}$. The probability mass function (pmf) of X is given by

$$p_X(x) = P\left[X = x\right], \text{ for } \ x \in \mathcal{D}.$$

The second type of random variable we will look at is the **continuous** random variable

Definition (Continuous Random Variables)

We say a random variable is a continuous random variable if its cumulative distribution function $F_X(x)$ is a continuous function for all $x \in \mathbb{R}$.

Unlike discrete random variables for a continuous random variable X, there are no points of discrete mass, $P(X = x) = 0$

If a random variable has a **cumulative distribution function** $F_X(x)$, the from the Fundamental Theorem of Calculus we have the **probability density function**

$$\frac{d}{dx}F_X(x) = f_X(x)$$

Moreover,

$$P\left[a < X \leq b\right] = F_X(b) - F_X(a) = \int_a^b f_X(t)dt$$

Example

Let X be the random variable representing the time in seconds between incoming phone calls at a call center, and its pdf is given by

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-x/4} & 0 < x < \infty \\ 0 & \text{elsewhere} \end{cases}$$

What is the probability that the time between calls exceeds 4 seconds?

$$P\left[X > 4\right] = \int_4^\infty \frac{1}{4}e^{-x/4}dx = e^{-1} = 0.3679$$

Sometimes we are interested in a transformation of a random variable, say $Y = g(X)$ where we know the distribution of X and want to find the distribution of Y

Theorem

Let X be a continuous random variable with pdf $f_X(x)$ and support S_X. Let $Y = g(X)$, where $g(x)$ is a one-to-one differentiable function, on the support of X, S_X. Denote the inverse of g by $x = g^{-1}(y)$ and let $dx/dy = d[g^{-1}(y)]/dy$. Then the pdf of Y is given by

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|, \text{ for } y \in S_Y, \tag{1.5}$$

where the support of Y is the set $S_Y = \{y = g(x) : x \in S_X\}$.

Probability and Distributions

Transformation of a Random Variable

Example

Let $X \sim \exp\theta$ and $Y \sim g(X)$. Find the pdf of Y.

Solution

$Y = g(X) = e^X$ then $X = \log Y$.

$$
\begin{aligned}
f_Y(y) &= f_X(g^{-1(x)}) \left| \frac{d\log y}{dy} \right| \\
&= \theta e^{-\theta \log y} \left| \frac{1}{y} \right| \\
&= \frac{\theta e^{-\theta \log y}}{y}
\end{aligned}
$$

Now we turn our attention to the expectation operator and the expectation of a random variable

Definition (Expectation or Expected Value)

Let X be a random variable. If X is a continuous random variable with pdf $f(x)$ and its integral converges absolutely, then the **expectation** of X is

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx$$

If X is a discrete random variable then then the **expectation** of X is

$$\mathbb{E}X = \sum_{x} x p_X(x)$$

Exercise

Let X be a discrete random variable with the following pmf:

x	1	2	3	4
$p(x)$	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{10}$

Find $\mathbb{E}(X)$

Exercise

Let X be a continuous random variable with pdf

$$f(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1.6}$$

Find $\mathbb{E}(X)$

Probability and Distributions

Definition (Mean)

Let X be a random variable whose expectation exists. The **mean** (also the expected value of X) value μ of X is defined to be $\mu = \mathbb{E}(X)$.

Definition (Variance)

Let X be a random variable with finite mean μ and such that $\mathbb{E}[(X-\mu)^2]$ is finite. Then the **variance** of X is defined to be $\mathbb{E}[(X-\mu)^2]$. It is usually denoted by σ^2 or by $\mathrm{Var}(X)$.

References

Hogg, R. V., McKean, J. W. and Craig, A. T. (2018).
Introduction to Mathematical Statistics.
Pearson, Boston, eighth edition.

Ross, S. (2010).
A First Course in Probability.
Prentice Hall, Upper Saddle River, eigth edition.