Week 3: Discrete Random Variables and Probability Distributions

MATH697

Sahir Bhatnagar

September 19, 2017

McGill University

Introduction

Introduction

In Chapter 2, we discussed the probability model as the central object of study in the theory of probability. This required defining a probability measure P on a class of subsets of the sample space Ω. For example, for an experiment with possible sample outcomes denoted by the *sample space* Ω, an *event* E was defined as any collection of sample outcomes, that is, any subset of the set Ω.

In Chapter 2, we discussed the probability model as the central object of study in the theory of probability. This required defining a probability measure P on a class of subsets of the sample space Ω. For example, for an experiment with possible sample outcomes denoted by the *sample space* Ω, an *event* E was defined as any collection of sample outcomes, that is, any subset of the set Ω.

$$\begin{array}{llll} \text{EXPERIMENT} & \longrightarrow \text{SAMPLE OUTCOMES} & \longrightarrow \text{EVENTS} & \longrightarrow \text{PROBABILITIES} \\ & \Omega = \{s_1, s_2, \ldots\} & \longrightarrow E \subseteq S & \longrightarrow P(E) \end{array}$$

In Chapter 2, we discussed the probability model as the central object of study in the theory of probability. This required defining a probability measure P on a class of subsets of the sample space Ω. For example, for an experiment with possible sample outcomes denoted by the *sample space* Ω, an *event* E was defined as any collection of sample outcomes, that is, any subset of the set Ω.

$$
\begin{array}{llll}
\text{EXPERIMENT} & \longrightarrow \text{SAMPLE OUTCOMES} & \longrightarrow \text{EVENTS} & \longrightarrow \text{PROBABILITIES} \\
 & \Omega = \{s_1, s_2, \ldots\} & \longrightarrow E \subseteq S & \longrightarrow P(E)
\end{array}
$$

In this framework, it is necessary to consider each experiment with its associated sample space separately - the nature of sample space Ω is typically different for different experiments.

Example 1 (Rainy Days)

Count the number of days in February which have zero precipitation.

$$\Omega = \{0, 1, 2, \ldots, 28\}$$

Let E_i = i days have zero precipitation. Then $E_0, \ldots, E_{28}$ partition Ω.

Example

Example 2 (Football Match)

Count the number of goals in a football match.

$$\Omega = \{0, 1, 2, 3, \ldots\}$$

Let E_i = i goals in the match. $E_0, E_1, E_2, \ldots$ partition Ω

Example

Example 2 (Football Match)

Count the number of goals in a football match.

$$\Omega = \{0, 1, 2, 3, \ldots\}$$

Let E_i = i goals in the match. $E_0, E_1, E_2, \ldots$ partition Ω

In both of these examples, we need a formula to specify each

$$P(E_i) = p_i$$

Example

Example 3 (Operating Temperature)

Measure the operating temperature of an experimental process.

$$\Omega = \{x : x > T_{min}\}$$

Example 3 (Operating Temperature)

Measure the operating temperature of an experimental process.

$$\Omega = \{x : x > T_{min}\}$$

Here it is difficult to express

$$P(\text{Measurement is } x)$$

Example 3 (Operating Temperature)

Measure the operating temperature of an experimental process.

$$\Omega = \{x : x > T_{min}\}$$

Here it is difficult to express

$$P(\text{Measurement is } x)$$

but possible to think about

$$P(\text{Measurement is } \leq x) = F(x)$$

Example 3 (Operating Temperature)

Measure the operating temperature of an experimental process.

$$\Omega = \{x : x > T_{min}\}$$

Here it is difficult to express

$$P(\text{Measurement is } x)$$

but possible to think about

$$P(\text{Measurement is } \leq x) = F(x)$$

We seek a formula for $F(x)$ which is a simpler way of presenting a particular probability assignment.

Introduction

- This chapter is concerned with the definitions of random variables, distribution functions $F(x)$, probability/density functions $f(x)$, and the development of the concepts necessary for carrying out calculations for a probability model using these entities

Introduction

- This chapter is concerned with the definitions of random variables, distribution functions $F(x)$, probability/density functions $f(x)$, and the development of the concepts necessary for carrying out calculations for a probability model using these entities
- The concept of a random variable allows us to pass from the experimental outcomes themselves to a numerical function of the outcomes. There are two fundamentally different types of random variables :
    (i) discrete random variables
    (ii) continuous random variables

Introduction

- This chapter is concerned with the definitions of random variables, distribution functions $F(x)$, probability/density functions $f(x)$, and the development of the concepts necessary for carrying out calculations for a probability model using these entities
- The concept of a random variable allows us to pass from the experimental outcomes themselves to a numerical function of the outcomes. There are two fundamentally different types of random variables :
  (i) discrete random variables
  (ii) continuous random variables
- In this chapter, we examine the basic properties and discuss the most important examples of discrete variables. Chapter 4 focuses on continuous random variables.

Random Variables

A general notation useful for all such examples can be obtained by considering a sample space that is **equivalent** to Ω for a general experiment, but whose form is more familiar.

Definition 4 (Random Variable (RV))

A random variable X on Ω is a function from the sample space Ω to the set $\mathbb{R}$ of all real numbers denoted by

$$X : \Omega \to \mathbb{R}$$

Let R_X denote the range of X.

Definition

A general notation useful for all such examples can be obtained by considering a sample space that is **equivalent** to Ω for a general experiment, but whose form is more familiar.

Definition 4 (Random Variable (RV))

A random variable X on Ω is a function from the sample space Ω to the set $\mathbb{R}$ of all real numbers denoted by

$$X : \Omega \to \mathbb{R}$$

Let R_X denote the range of X.

X is called a discrete random variable if R_X is a countable set.

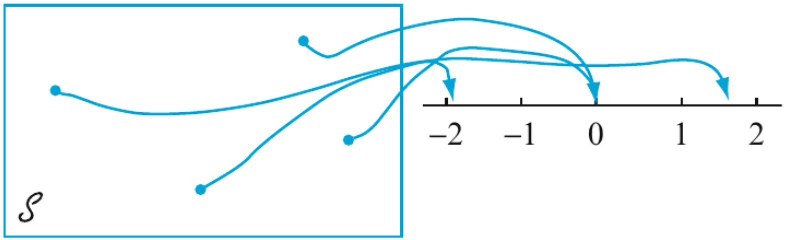


Figure 3.1 A random variable

Figure 1

Notation

- Random variables are customarily denoted by uppercase letters, such as X and Y, lowercase letters to represent some particular value of the corresponding random variable.

Notation

- Random variables are customarily denoted by uppercase letters, such as X and Y, lowercase letters to represent some particular value of the corresponding random variable.
- The notation $X(s) = x$ means that x is the value associated with the outcome s by the rv X.

Example 5 (Coin Toss)

Suppose a coin is tossed three times. Let X be the number of heads observed. The sample space is

$$\Omega = \left\{ \underbrace{HHT}_{3}, \underbrace{HHT}_{2}, \underbrace{HTH}_{2}, \underbrace{HTT}_{1}, \underbrace{THH}_{2}, \underbrace{THT}_{1}, \underbrace{TTH}_{1}, \underbrace{TTT}_{0} \right\}$$

That is, we have $X(HHH) = 3$, $X(HHT) = 2$, $X(HTH) = 2$, and so on. Hence $R_X = \{0, 1, 2, 3\}$

Example

Example 6 (A Very Simple Random Variable)

Let the random variable $X : \{\text{rain}, \text{snow}, \text{clear}\} \to \mathbb{R}$ by $X(rain) = 3$, $X(snow) = 6$, and $X(clear) = -2.7$.

Example 6 (A Very Simple Random Variable)

Let the random variable $X : \{\text{rain}, \text{snow}, \text{clear}\} \to \mathbb{R}$ by $X(rain) = 3$, $X(snow) = 6$, and $X(clear) = -2.7$.

We now present several further examples. The point is, we can define random variables any way we like, as long as they are functions from the sample space to $\mathbb{R}$.

Example

Example 7 (A Very Simple Random Variable 2)

For the case $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$, we might define a second random variable Y by saying that $Y = 0$ if it rains, $Y = -1/2$ if it snows, and $Y = 7/8$ if it is clear. That is $Y(rain) = 0$, $Y(snow) = 1/2$, and $Y(rain) = 7/8$.

Example

Example 8 (A Very Simple Random Variable 3)

If the sample space corresponds to flipping three different coins, then we could let X be the total number of heads showing, let Y be the total number of tails showing, let $Z = 0$ if there is exactly one head, and otherwise $Z = 17$.

Example

Example 9 (Constants as Random Variables)

As a special case, every constant value c is also a random variable, by saying that $c(s) = c$ for all $s \in \Omega$. Thus, 5 is a random variable, as is 3 or -21.6

Example

Example 10 (Indicator Functions)

If A is any event, then we can define the indicator function of A, written I_A, to be the random variable

$$I_A(s) = \begin{cases} 1 & s \in A \\ 0 & s \notin A \end{cases}$$

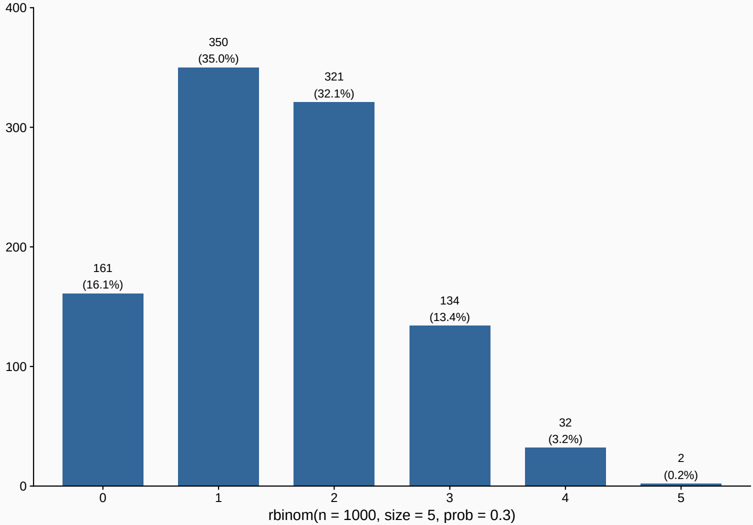
Distribution of X

Suppose X is a random variable. We know that different states s occur with different probabilities. It follows that $X(s)$ also takes different values with different probabilities. These probabilities are called the distribution of X; we consider them next.

Probability Distributions for Discrete Random Variables

Distribution of X

```r
sjPlot::sjp.frq(rbinom(n = 1000, size = 5, prob = .3))
```



Distribution of X

- Because random variables are defined to be functions of the outcome s, and because the outcome s is assumed to be random (i.e., to take on different values with different probabilities), it follows that the value of a random variable will itself be random (as the name implies).

Distribution of X

- Because random variables are defined to be functions of the outcome s, and because the outcome s is assumed to be random (i.e., to take on different values with different probabilities), it follows that the value of a random variable will itself be random (as the name implies).
- Specifically, if X is a random variable, then what is the probability that X will equal some particular value x? Well, $X = x$ precisely when the outcome s is chosen such that $X(s) = x$.

Example

Example 11 (A Very Simple Random Variable Revisited)

Let the random variable $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$ and X is defined by $X(rain) = 3$, $X(snow) = 6$, and $X(clear) = -2.7$. Suppose further that the probability measure P is such that

$$P(rain) = 0.4 \quad P(snow) = 0.15 \quad P(clear) = 0.45$$

Example

Example 11 (A Very Simple Random Variable Revisited)

Let the random variable $\Omega = \{\text{rain}, \text{snow}, \text{clear}\}$ and X is defined by $X(rain) = 3$, $X(snow) = 6$, and $X(clear) = -2.7$. Suppose further that the probability measure P is such that

$$P(rain) = 0.4 \quad P(snow) = 0.15 \quad P(clear) = 0.45$$

Then clearly, $X = 3$ only when it rains, $X = 6$ only when it snows, and $X = -2.7$ only when it is clear. Thus $P(X = 3) = P(rain) = 0.4$, $P(X = 6) = P(snow) = 0.15$, $P(X = -2.7) = P(clear) = 0.45$

Example 12 (A Very Simple Random Variable Revisited (cont))

Also, $P(X = 17) = 0$, and in fact $P(X = x) = P(\varnothing) = 0$ for all $x \notin \{3, 6, -2.7\}$. We can also compute that

$$P(X \in \{3, 6\}) = P(X = 3) + P(X = 6) = 0.4 + 0.15 = 0.55$$

Example 12 (A Very Simple Random Variable Revisited (cont))

Also, $P(X = 17) = 0$, and in fact $P(X = x) = P(\emptyset) = 0$ for all $x \notin \{3, 6, -2.7\}$. We can also compute that

$$P(X \in \{3, 6\}) = P(X = 3) + P(X = 6) = 0.4 + 0.15 = 0.55$$

while

$$P(X < 5) = P(X = 3) + P(X = -2.7) = 0.4 + 0.45 = 0.85$$

Distribution of X

We see from this example that, if B is any subset of the real numbers, then

$$P(X \in B) = P(\{s \in \Omega : X(s) \in B\})$$

Furthermore, to understand X well requires knowing the probabilities $P(X \in B)$ for different subsets B. That is the motivation for the following definition.

We see from this example that, if B is any subset of the real numbers, then

$$P(X \in B) = P(\{s \in \Omega : X(s) \in B\})$$

Furthermore, to understand X well requires knowing the probabilities $P(X \in B)$ for different subsets B. That is the motivation for the following definition.

Definition 13 (Distribution of X)

If X is a random variable, then the distribution of X is the collection of probabilities $P(X \in B)$ for all subsets B of the real numbers.

For a set B we must find the elements in $s \in S$ such that $X(s) \in B$. These elements are given by the set $\{s \in S : X(s) \in B\}$. Then we evaluate the probability $P(\{s \in S : X(s) \in B\})$. We must do this for every subset $B \in \mathbb{R}$.

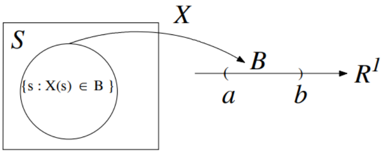


Figure 2.2.1: If $B = (a, b) \subset R^1$, then $\{s \in S : X(s) \in B\}$ is the set of elements such that $a < X(s) < b$.

Figure 2

Definition 14 (Discrete Distribution)

A random variable X is discrete if

$$\sum_{x \in \mathbb{R}} P(X = x) = 1$$

Definitions of Discrete Distributions (Alternative)

Definition 15 (Discrete Distribution)

A random variable X is discrete if there is a finite or countable sequence $x_1, x_2, \ldots$ of distinct real numbers, and a corresponding sequence $p_1, p_2, \ldots$ of nonnegative real numbers such that

$$P(X = x_i) = p_i \quad \forall\ i \quad \text{and} \quad \sum_i p_i = 1$$

This second definition also suggests how to keep track of discrete distributions. It prompts the following definition

Definition 16 (Probability Mass Function)

For a discrete random variable X, its probability mass function (PMF) is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$f_X(x) = P(X = x)$$

Probability Mass Function

Definition 16 (Probability Mass Function)

For a discrete random variable X, its probability mass function (PMF) is the function $p_X : \mathbb{R} \to [0, 1]$ defined by

$$f_X(x) = P(X = x)$$

Hence if $x_1, x_2, \ldots$ are the distinct values such that $P(X = x_i) = p_i$ for all i with $\sum_i p_i = 1$, then

$$f_X(x) = \begin{cases} p_i & x = x_i \\ 0 & otherwise \end{cases}$$

All the information about the distribution of X is contained in its probability function, but only if we know that X is a discrete RV

Visual of the PMF

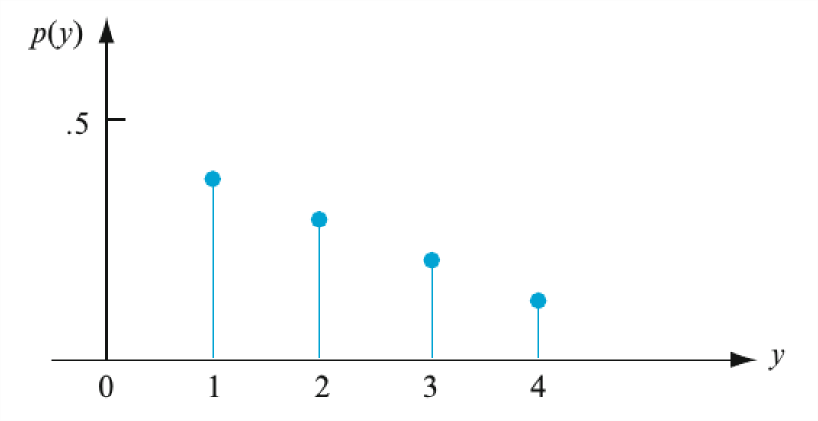


Figure 3

Theorem 17 (Law of total probability, discrete random variable version)

Let X be a random variable, and let A be some event. Then

$$P(A) = \sum_{x \in \mathbb{R}} P(X = x) P(A | X = x)$$

This follows from the probability axioms where f_X must exhibit the following properties

(i) $f_X(x_i) \geq 0 \ \forall \ i$

(ii) $\sum_i f_X(x_i) = 1$

Discrete Cumulative Distribution Function

The **cumulative distribution function** or **CDF**, F_X, is defined by

$$F_X(x) = P[X \leq x] \qquad for\; x \in \mathbb{R}$$

Visual of the CDF

$$F(y) = \begin{cases} 0 & y < 1 \\ .05 & 1 \le y < 2 \\ .15 & 2 \le y < 4 \\ .50 & 4 \le y < 8 \\ .90 & 8 \le y < 16 \\ 1 & 16 \le y \end{cases}$$

A graph of this cdf is shown in Figure 3.5.

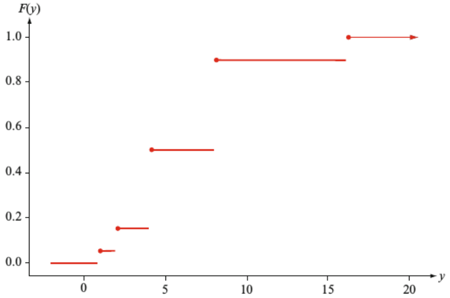


Figure 4

The CDF F_X must exhibit the following properties:

(i) $\lim_{x \longrightarrow -\infty} F_X(x) = 0$

Properties of the CDF

The CDF F_X must exhibit the following properties:

(i) $\lim\limits_{x \longrightarrow -\infty} F_X(x) = 0$

(ii) $\lim\limits_{x \longrightarrow \infty} F_X(x) = 1$

Properties of the CDF

The CDF F_X must exhibit the following properties:

(i) $\lim_{x \longrightarrow -\infty} F_X(x) = 0$

(ii) $\lim_{x \longrightarrow \infty} F_X(x) = 1$

(iii) $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$ [i.e. F_X is continuous from the right]

Properties of the CDF

The CDF F_X must exhibit the following properties:

(i) $\lim_{x \longrightarrow -\infty} F_X(x) = 0$

(ii) $\lim_{x \longrightarrow \infty} F_X(x) = 1$

(iii) $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$ [i.e. F_X is continuous from the right]

(iv) $a < b \Longrightarrow F_X(a) \leq F_X(b)$ [i.e. F_X is non-decreasing]

Properties of the CDF

The CDF F_X must exhibit the following properties:

(i) $\lim_{x \longrightarrow -\infty} F_X(x) = 0$

(ii) $\lim_{x \longrightarrow \infty} F_X(x) = 1$

(iii) $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$ [i.e. F_X is continuous from the right]

(iv) $a < b \Longrightarrow F_X(a) \leq F_X(b)$ [i.e. F_X is non-decreasing]

(v) $P[a < X \leq b] = F_X(b) - F_X(a)$

PMF and CDF

The functions f_X and/or F_X can be used to describe the probability distribution of random variable X.

Example

Example 18 (An electrical circuit comprises six fuses)

let X =number of fuses that fail within one month. Then

$$\mathbb{X} = \{0, 1, 2, 3, 4, 5, 6\}$$

To specify the probability distribution of X, can use the PMF f_X or the CDF F_X. For example,

x	0	1	2	3	4	5	6
$f_X(x)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$
$F_X(x)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{7}{16}$	$\frac{11}{16}$	$\frac{13}{16}$	$\frac{15}{16}$	$\frac{16}{16}$

as $F_X(0) = P[\, X \leq 0 \,] = P[\, X = 0 \,] = f_X(0)$,
$F_X(1) = P[\, X \leq 1 \,] = P[\, X = 0 \,] + P[\, X = 1 \,] = f_X(0) + f_X(1)$, and so on.

Example 19 (An electrical circuit comprises six fuses (cont))

Note also that, for example,

$$P[\,X \leq 2.5\,] \equiv P[\,X \leq 2\,]$$

as the random variable X only takes values $0, 1, 2, 3, 4, 5, 6$.

Example 20 (A computer is prone to crashes)

Suppose that $P[\text{Computer crashes on any given day}] = \theta$, for some $0 \leq \theta \leq 1$, independently of crashes on any other day. Let $X =$ number of days until the first crash. Then $\mathbb{X} = \{1, 2, 3, \ldots\}$. To specify the probability distribution of X, can use the PMF f_X or the cdf F_X. Now,

Example 20 (A computer is prone to crashes)

Suppose that $P[\text{Computer crashes on any given day}] = \theta$, for some $0 \leq \theta \leq 1$, independently of crashes on any other day. Let $X =$ number of days until the first crash. Then $\mathbb{X} = \{1, 2, 3, \ldots\}$. To specify the probability distribution of X, can use the PMF f_X or the cdf F_X. Now,

$$f_X(x) = P[\, X = x \,] = (1 - \theta)^{x-1}\theta$$

for $x = 1, 2, 3, \ldots$ (if the first crash occurs on day x, then we must have a sequence of $x - 1$ crash-free days, followed by a crash on day x).

Example

Example 20 (A computer is prone to crashes)

Suppose that $P[\text{Computer crashes on any given day}] = \theta$, for some $0 \leq \theta \leq 1$, independently of crashes on any other day. Let $X =$ number of days until the first crash. Then $\mathbb{X} = \{1, 2, 3, ...\}$. To specify the probability distribution of X, can use the PMF f_X or the cdf F_X. Now,

$$f_X(x) = P[\,X = x\,] = (1-\theta)^{x-1}\theta$$

for $x = 1, 2, 3, ...$ (if the first crash occurs on day x, then we must have a sequence of $x - 1$ crash-free days, followed by a crash on day x). Also

$$F_X(x) = P[\,X \leq x\,] = P[\,X = 1\,] + P[\,X = 2\,] + ... + P[\,X = x\,] = 1 - (1-\theta)^x$$

as the terms in the summation are a geometric progression with first term θ and common term $1 - \theta$.

Relationship Between f_X and F_X

The fundamental relationship between f_X and F_X is obtained by noting that if $x_1 \leq x_2 \leq \ldots \leq x_n \leq \ldots$,

Relationship Between f_X and F_X

The fundamental relationship between f_X and F_X is obtained by noting that if $x_1 \leq x_2 \leq \ldots \leq x_n \leq \ldots$, then

$$P[X \leq x_i] = P[X = x_1] + \ldots + P[X = x_i],$$

The fundamental relationship between f_X and F_X is obtained by noting that if $x_1 \leq x_2 \leq \ldots \leq x_n \leq \ldots$, then

$$P[X \leq x_i] = P[X = x_1] + \ldots + P[X = x_i],$$

so that

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i) \ ,$$

and

Relationship Between f_X and F_X

$$f_X(x_1) = F_X(x_1) \qquad\qquad f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \quad \text{for } i \geq 2$$

so $P[c_1 < X \leq c_2] = F_X(c_2) - F_X(c_1)$ for any real numbers $c_1 < c_2$.

Relationship Between f_X and F_X

$$f_X(x_1) = F_X(x_1) \qquad\qquad f_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

$$f_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \quad \text{for } i \geq 2$$

so $P[c_1 < X \leq c_2] = F_X(c_2) - F_X(c_1)$ for any real numbers $c_1 < c_2$.

Hence, in the discrete case, we can calculate F_X from f_X by **summation**, and calculate f_X from F_X by **differencing**.

-
-