Chapter 3
Discrete Random Variables and Probability Distributions

Part 1: Discrete Random Variables

Section 2.9 Random Variables (section fits better here)
Section 3.1 Probability Distributions and Probability Mass Functions
Section 3.2 Cumulative Distribution Functions

Random Variables

- Consider tossing a coin two times. We can think of the following *ordered* sample space: $S = \{(T,T),(T,H),(H,T),(H,H)\}$
  NOTE: for a fair coin, each of these are equally likely.

- The outcome of a random experiment need not be a number, but we are often interested in some (numerical) measurement of the outcome.

- For example, the <u>*Number of Heads*</u> obtained is numeric in nature can be 0, 1, or 2 and is a **random variable**.

Definition (Random Variable)

A <u>random variable</u> is a *function* that assigns a real number to each outcome in the sample space of a random experiment.

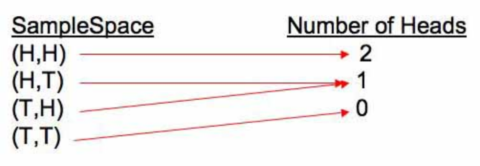
Random Variables

Definition (Random Variable)

A random variable is a *function* that assigns a real number to each outcome in the sample space of a random experiment.

Example (Random Variable)

For a fair coin flipped twice, the probability of each of the possible values for *Number of Heads* can be tabulated as shown:



Number of Heads	0	1	2
Probability	1/4	2/4	1/4

Let $X \equiv \#$ of heads observed. X **is a random variable**.

Discrete Random Variables

Definition (Discrete Random Variable)

A discrete random variable is a variable which can only take-on a countable number of values (finite or countably infinite)

Example (Discrete Random Variable)

- Flipping a coin twice, the random variable *Number of Heads* $\in \{0, 1, 2\}$ is a discrete random variable.
- Number of flaws found on a randomly chosen part $\in \{0, 1, 2, \ldots\}$.
- Proportion of defects among 100 tested parts $\in \{0/100, 1/100, \ldots, 100/100\}$.
- Weight measured to the nearest pound.*

  *Because the possible values are discrete and countable, this random variable is discrete, but it might be a more convenient, simple approximation to assume that the measurements are values on a continuous random variable as 'weight' is theoretically continuous.

Continuous Random Variables

Definition (Continuous Random Variable)

A **continuous random variable** is a random variable with an interval (either finite or infinite) of real numbers for its range.

Example (Continuous Random Variable)

- Time of a reaction.
- Electrical current.
- Weight.

Discrete Random Variables

We often omit the discussion of the underlying sample space for a random experiment and directly describe the distribution of a particular random variable.

Example (Production of prosthetic legs)

Consider the experiment in which prosthetic legs are being assembled **until** a defect is produced. Stating the sample space...

$$S = \{d, gd, ggd, gggd, \ldots\}$$

Let X be the trial number at which the experiment terminates (i.e. the sample at which the first defect is found).

The possible values for the random variable X are in the set $\{1, 2, 3, \ldots\}$

We may skip a formal description of the sample space S and move right into the random variable of interest X.

Probability Distributions and Probability Mass Functions

Definition (Probability Distribution)

A **probability distribution** of a random variable X is a description of the probabilities associated with the possible values of X.

Example (Number of heads)

Let $X \equiv \#$ of heads observed when a coin is flipped twice.

Number of Heads	0	1	2
Probability	1/4	2/4	1/4

Probability distributions for discrete random variables are often given as a table or as a function of X...

Example (Probability defined by function $f(x)$)

Table:

x	1	2	3	4
$\mathrm{P}(X = x) = f(x)$	0.1	0.2	0.3	0.4

Function of X: $f(x) = \frac{1}{10}x$ for $x \in \{1, 2, 3, 4\}$

Probability Distributions and Probability Mass Functions

Example (Transmitted bits, example 3.3 p.44)

There is a chance that a bit transmitted through a digital transmission channel is received in error.

Let X equal the number of bits in error in the next four bits transmitted. The possible values for X are $\{0, 1, 2, 3, 4\}$.

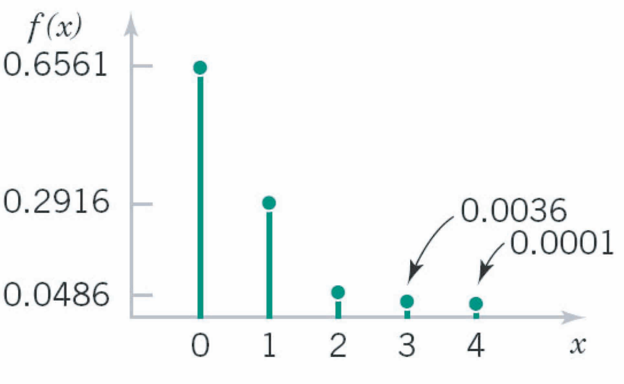
Suppose that the probabilities are...

x	$P(X = x)$
0	0.6561
1	0.2916
2	0.0486
3	0.0036
4	0.0001

Probability Distributions and Probability Mass Functions

Example (Transmitted bits, example 3.3 p.44, cont.)

The probability distribution shown graphically:



Notice that the sum of the probabilities of the possible random variable values is equal to 1.

Probability Mass Function (PMF)

Definition (Probability Mass Function (PMF))

For a discrete random variable X with possible values $x_1, x_2, x_3, \ldots, x_n$, a **probability mass function** $f(x_i)$ is a function such that

1. $f(x_i) \geq 0$
2. $\sum_{i=1}^{n} f(x_i) = 1$
3. $f(x_i) = P(X = x_i)$

Example (Probability Mass Function (PMF))

For the transmitted bit example,

$f(0) = 0.6561, f(1) = 0.2916, ..., f(4) = 0.0001$

$\sum_{i=1}^{n} f(x_i) = 0.6561 + 0.2916 + \cdots + 0.0001 = 1$

The probability distribution for a discrete random variable is described with a probability mass function (probability distributions for continuous random variables will use different terminology).

Probability Mass Function (PMF)

Example (Probability Mass Function (PMF))

Toss a coin 3 times.

- Let X be the number of heads tossed.
  Write down the probability mass function (PMF) for X:
  {Use a table...}

- Show the PMF graphically:

Probability Mass Function (PMF)

Example (Probability Mass Function (PMF))

A box contains 7 balls numbered 1,2,3,4,5,6,7. Three balls are drawn at random and *without replacement*.

- Let X be the number of 2's drawn in the experiment.

  Write down the probability mass function (PMF) for X:

  {Use your counting techniques}

Cumulative Distribution Function (CDF)

Sometimes it's useful to quickly calculate a **cumulative probability**, or $P(X \leq x)$, denoted as $F(x)$, which is the probability that X is less than or equal to some specific x.

Example (Widgets, PMF and CDF)

Let X equal the number of widgets that are defective when 3 widgets are randomly chosen and observed. The possible values for X are $\{0, 1, 2, 3\}$.

The probability mass function for X:

x	$P(X = x)$ or $f(x)$
0	0.550
1	0.250
2	0.175
3	0.025

Suppose we're interested in the probability of getting 2 or less errors (i.e. either 0, or 1, or 2). We wish to calculate $P(X \leq 2)$.

Cumulative Distribution Function (CDF)

Example (Widgets, PMF and CDF, cont.)

$$
\begin{aligned}
P(X \leq\ 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\
&= 0.550 \quad + \ 0.250 \quad + \ 0.175 \ = \ 0.975
\end{aligned}
$$

Below we see a table showing $P(X \leq\ x)$ for each possible x.

Cumulative Probabilities...

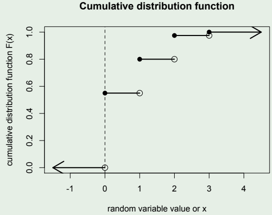
x	$P(X = x)$	$P(X \leq x) = F(x)$	
0	0.550	0.550	$P(X \leq 0) = F(0)$
1	0.250	0.800	$P(X \leq 1) = F(1)$
2	0.175	0.975	$P(X \leq 2) = F(2)$
3	0.025	1.000	$P(X \leq 3) = F(3)$

As x increases across the possible values for x, the cumulative probability increases, eventually getting 1, as you accumulate all the probability.

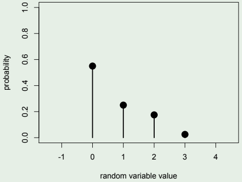
Cumulative Distribution Function (CDF)

Example (Widgets, PMF and CDF, cont.)

The cumulative probabilities are shown below as a function of x or $F(x) = P(X \leq x)$.



The above cumulative distribution function $F(x)$ is associated with the probability mass function $f(x)$ below:



Connecting the PMF and the CDF

- Connecting the PMF and the CDF
  - We can get the PMF (i.e. the probabilities for $P(X = x_i)$) from the CDF by determining the height of the jumps.
  - Specifically, because a CDF for a discrete random variable is a step-function with left-closed and right-open intervals, we have

    $$P(X = x_i) = F(x_i) - \lim_{x \uparrow x_i} F(x_i)$$

    and this expression calculates the difference between $F(x_i)$ and the limit as x increases to x_i.

Cumulative Distribution Function (CDF)

Definition (CDF for a discrete random variable)

The cumulative distribution function of a discrete random variable X, denoted as $F(x)$, is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Definition (CDF for a discrete random variable)

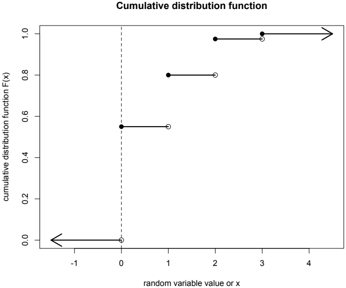
For a discrete random variable X, F(x) satisfies the following properties:

1. $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$
2. $0 \leq F(x) \leq 1$
3. If $x \leq y$, then $F(x) \leq F(y)$

- The CDF is defined on the real number line.
- The CDF is a non-decreasing function of X (i.e. increases or stays constant as $x \to \infty$).

Cumulative Distribution Function (CDF)

- For each probability mass function (PMF), there is an associated CDF.
- If you're given a CDF, you can come-up with the PMF and vice versa (know how to do this).
- Even if the random variable is discrete, the CDF is defined between the discrete values (i.e. you can state $P(X \leq x)$ for any $x \in \Re$).
- The CDF 'step function' for a discrete random variable is composed of left-closed and right-open intervals with steps occurring at the values which have positive probability (or 'mass').



Cumulative Distribution Function (CDF)

- The cumulative distribution function $F(x)$ for a discrete random variable is a step-function.

Example (Widgets, PMF and CDF, cont.)

In the widget example, the range of X is $\{0, 1, 2, 3\}$. There is no chance of a getting value outside of this set, e.g. $f(1.8) = P(X = 1.8) = 0$. But $F(1.8) = P(X \leq 1.8) \neq 0$. Specifically...

$$\begin{aligned} F(1.8) = P(X \leq 1.8) &= P(X \leq 1) \\ &= P(X = 0) + P(X = 1) = 0.800. \end{aligned}$$

So, if $f(x) = 0$, it does not necessarily mean $F(x) = 0$.

Here is $F(x)$ for the widget example:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.550 & \text{if } 0 \leq x < 1 \\ 0.800 & \text{if } 1 \leq x < 2 \\ 0.975 & \text{if } 2 \leq x < 3 \\ 1.0000 & \text{if } x \geq 3 \end{cases}$$

Cumulative Distribution Function (CDF)

Example (Monitoring a chemical process)

The output of a chemical process is continually monitored to ensure that the concentration remains within acceptable limits. Whenever the concentration drifts outside the limits, the process is shut down and recalibrated. Let X be the number of times in a given week that the process is recalibrated. The following table presents values of the cumulative distribution function $F(x)$ of X.

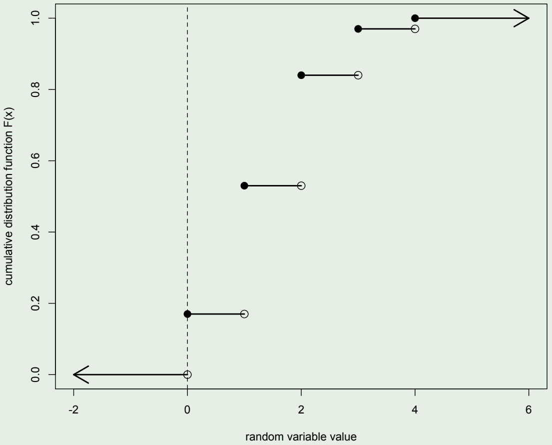
$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.17 & \text{if } 0 \leq x < 1 \\ 0.53 & \text{if } 1 \leq x < 2 \\ 0.84 & \text{if } 2 \leq x < 3 \\ 0.97 & \text{if } 3 \leq x < 4 \\ 1.0000 & \text{if } x \geq 4 \end{cases}$$

From the values in the far right column, I know that $X \in \{0, 1, 2, 3, 4\}$.

Cumulative Distribution Function (CDF)

Example (Monitoring a chemical process, cont.)

(1) Graph the cumulative distribution function.



Cumulative Distribution Function (CDF)

Example (Monitoring a chemical process, cont.)

(2) What is the probability that the process is recalibrated fewer than 2 times during a week?

(3) What is the probability that the process is recalibrated more than three times during a week?

Cumulative Distribution Function (CDF)

Example (Monitoring a chemical process, cont.)

(4) What is the probability mass function (PMF) for X?

(5) What is the most probable number of recalibrations in a week? (I'm not asking for an *expected value* here, just the one most likely).