Chapter 5
Discrete Random Variables and Probability Models:

Chapter Outcomes:

- Random variables and probability functions
- Discrete uniform distribution
- Hypergeometric distribution
- Binomial distribution
- Negative binomial distribution
- Geometric distribution
- Poisson distribution
- Combining other models with the Poisson process

Random Variables and Probability Functions (Section 5.1)

- Probability models are used to describe outcomes associated with random processes (e.g. flipping a coin, rolling a die, etc.).
- So far, we have used sets in sample spaces to describe such outcomes.
- Now, we introduce numerical-valued variables *X*, *Y*, … to describe outcomes.

- These numerical-valued variables are called **random variables** and represent outcomes in an experiment or random process.

- Every random variable (r.v.) is associated with a **set of possible values** denoted by ***A*** *(called the range).*

- For example, suppose an experiment consists of tossing a fair coin 3 times. Then, we could let the r.v. X = number of heads that occur.

- Here, X has a range of $A = \{0,1,2,3\}$.

More formally:

- A r.v. is a function that assigns a real number to each point in a sample space *S.*
- Random variables are denoted by capital letters such as X, Y, $\ldots$.
- Their possible values/realizations are denoted by lowercase letters such as x, y, … .
- This provides a nice shorthand notation for outcomes (e.g. "X=2" would represent that "2 heads occurred").

- Random variables are defined for every outcome of the random experiment (i.e. for every outcome $a \in S$).
- The outcome x occurs (i.e. "X=x") for a corresponding set of outcomes a in the sample space S.
- <u>Example</u>: Consider again our coin tossing example, where a fair coin is tossed 3 times.

We can list the 8 possible outcomes:

$$S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}$$

Let’s define the r.v. X to represent the number of heads obtained.

What are the possible values for the r.v. X?

X = 0, 1, 2, 3.

A Function vs. the Value of a Function

- A <u>function</u> is used to map each value in a domain into a unique point.
- For example, $f(x) = x^2$ maps the point $x = 2$ in the domain into the point $f(2) = 2^2 = 4$ in the range.
- In our setting, the rule of mapping a point in the sample space (domain) into a real number in the range of a r.v. is often expressed in words rather than by a formula.
- Note that a **function** is captured by $f(x)$ and the **value of the function** by $f(2) = 4$.

Since "$X = x$" represents an event of some kind, our interest lies in finding its probability, which we represent by:

$$P(X = x)$$

These probabilities can be solved for using appropriately defined functions.

To do so, we first need to look at classifying random variables.

Random variables can be classified into one of **two types**, and this is determined by looking at the possible values that X can take on.

X

Discrete r.v.

- Range consists of a *finite* or *countably infinite* set of values.
- The random variable can take on integer values or, in general, values in a countable set (e.g. 0,1,2,....).

Continuous r.v.

- Range consists of an *infinite* (uncountable) set of values.
- The random variable takes values on some interval of real numbers (e.g. (0,1) or $(0,\infty)$ or $(-\infty,\infty)$).

Examples of Random Variables

Discrete

- Number of students present in class on a given day
- Number of heads obtained when flipping 3 coins
- Number of coin flips needed to observe 5 consecutive tails
- Number of defective items in a batch of products
- Number of free throws made by the Toronto Raptors in a given game
- Number of accidents per day at the corner of Phillip and University Ave.

Continuous

- Time it takes to get to school on on a given day
- Price of a stock following a day of trading
- Monthly rainfall amount at a certain geographical location
- Volume of coke in a can marketed as 355 mL
- Height of a randomly chosen maple tree in the Waterloo Region
- Lengths or areas of manufactured components

We will start off by looking at <u>discrete</u> random variables.

Definition: The **probability function** (p.f.) of a random variable X, usually denoted by $f(x)$, is given by

$$f(x) = P(X = x) \quad \forall\, x \in A.$$

The set of pairs $\{(x, f(x)) : x \in A\}$ is called the **probability distribution of *X***.

All probability functions must satisfy the following:

1. $f(x) \geq 0 \;\; \forall\, x \in A$

2. $\sum_{x \in A} f(x) = 1$

3. $f(x) \leq 1 \;\; \forall\, x \in A$

Example: A fair coin is tossed 3 times. Let X represent the number of heads observed.

Define the probability function (p.f.), $f(x)$ as follows:

Example: The random variable X has p.f. given by

x	0	1	2	3	4
$f(x)$	$0.1c$	$0.2c$	$0.5c$	c	$0.2c$

a) Determine the value of c.

b) Calculate $P(X > 2)$.

The p.f. is the most common way of describing a probability model, but another way is to use the **cumulative distribution function**.

Definition: The cumulative distribution function (c.d.f.) of X, usually denoted by $F(x)$, is given by

$$F(x) = P(X \le x) \;\forall\; x \in \mathbb{R}.$$

In general:

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

Properties of the c.d.f.:

1. $F(x)$ is a *non-decreasing* function of $x \; \forall \; x \in \mathbb{R}$. For example, $P(X \leq 8)$ cannot be less than $P(X \leq 7)$.

2. $0 \leq F(x) \leq 1 \; \forall \; x \in \mathbb{R}$.

3. $\lim\limits_{x \to -\infty} F(x) = 0$ and $\lim\limits_{x \to \infty} F(x) = 1$.

We can also obtain $f(x)$ from $F(x)$.

For instance, if X takes on integer values, then for values x and $x-1$ such that $x \in A$ and $x - 1 \in A$,

$$f(x) = F(x) - F(x-1).$$

In other words,

$$P(X = x) = P(X \leq x) - P(X \leq x - 1).$$

When a r.v. has been defined, it is generally simpler to find its p.f. first, but sometimes it is simpler to find its c.d.f. first.

We notice that $f(x)$ represents the size of the **jump** in $F(x)$ at the point x.

Plots of $\boldsymbol{f(x)}$ and $\boldsymbol{F(x)}$:

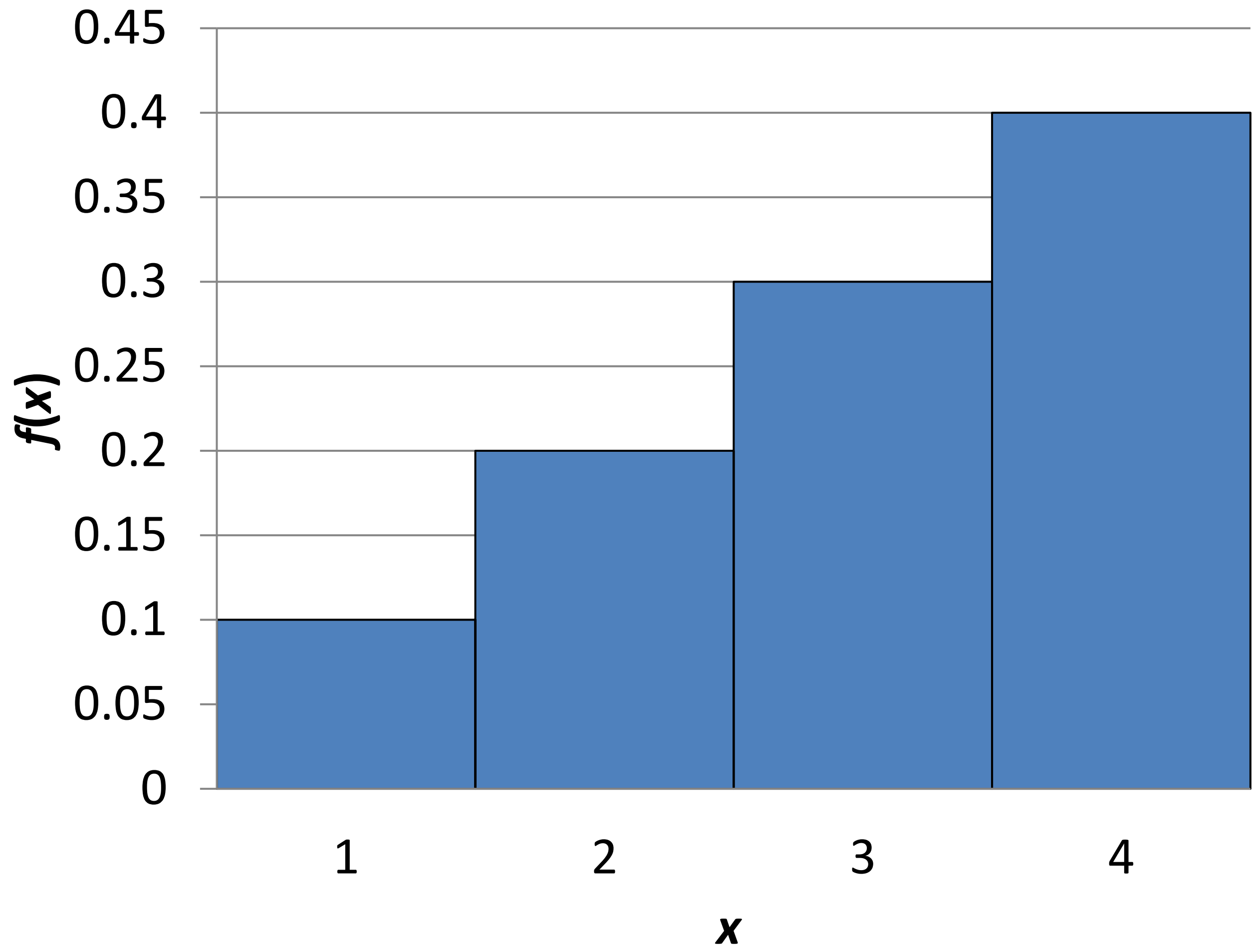
For discrete random variables, we find that the c.d.f. $F(x)$ is represented as a **step function.**

The p.f. $f(x)$ is frequently represented by a **histogram.**

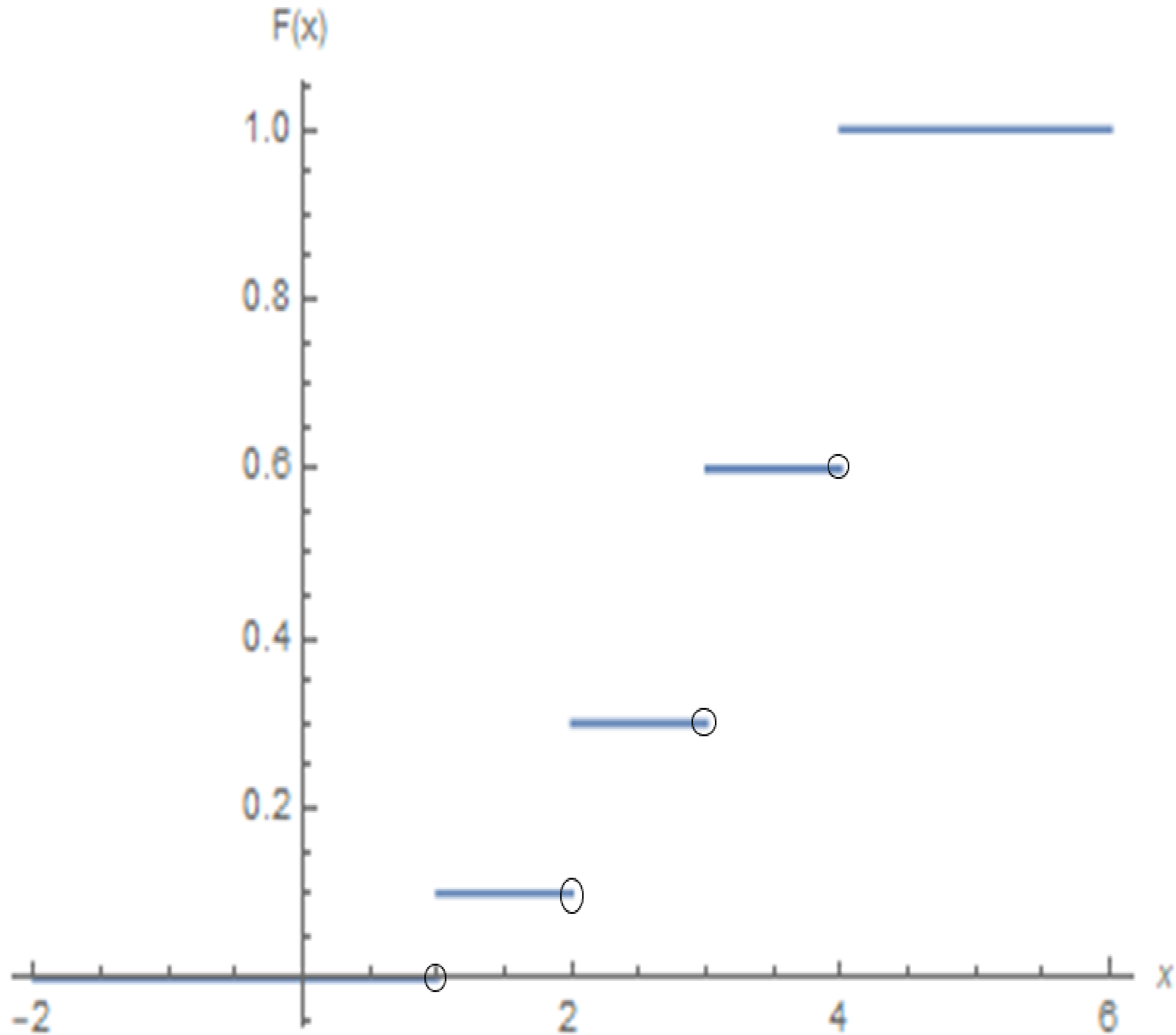
Example: Consider the r.v. X such that

$\boldsymbol{x}$	$\boldsymbol{f(x)}$
1	0.1
2	0.2
3	0.3
4	0.4

Plot $f(x)$ and $F(x)$.



0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
f(x)
1
2
3
4
x



F(x)
1.0
0.8
0.6
0.4
0.2
-2
2
4
6
x

Example: Recall the earlier example in which we toss a fair coin 3 times and X represents the number of heads observed.

a) Provide a plot of $f(x)$.
b) Determine $F(x)$ and provide a plot of $F(x)$.
c) Calculate $P(X \leq 2.4)$.

Special Distributions

In the rest of the chapter, our aim is to identify common types of processes or problems and to develop **probability distributions** that represent them.

Many processes or problems tend to have the same structure.

For example:

1. Toss a fair coin 10 times and let X record the "number of heads obtained".
2. Plant 20 seeds and let X record the "number of seeds that germinate".
3. Pick 12 items at random from a factory line and let X record the "number of items with no defects".

*All these problems are essentially the same, so why not develop a **general model** to accommodate them all?*

Discrete Uniform Distribution (Section 5.2)

Physical setup:

- If the r.v. X takes on values a, $a+1$, $a+2$, ..., b where each value is equally likely, then X is said to have a **discrete uniform distribution** on the set $\{a, a+1, ..., b\}$.
- We might sometimes write $X \sim DU(a, b)$.

For example:

- If X is the number obtained from rolling a fair die, then X has a discrete uniform distribution over the set $\{1, 2, ..., 6\}$ (i.e. $a=1$ and $b=6$).

Probability function:

- If X ranges between values a and b, this means that we have $b–a+1$ possible values for X, which are all equally likely.
- Hence, each of these values has a probability of $1/(b–a+1)$ of being observed, leading to

$$f(x) = \begin{cases} \frac{1}{b-a+1} & \text{for } x = a, a+1, \ldots, b \\ 0 & \text{otherwise} \end{cases}$$

- Note that

$$\sum_{x=a}^{b} f(x) = 1$$

Example: Suppose a fair die is thrown once and let X be the number obtained.

Determine the c.d.f. of X, $F(x)$.

Example: Let Y be a discrete uniform r.v. with a=0 and b=6.

Calculate the probability that the roots of the equation $g(z) = 0$ are real when:

$$g(z) = 0.75z^2 + 3Yz + 5Y + 8.$$

Hypergeometric Distribution (Section 5.3)

Physical setup:

- We have a population of ***N* objects** that can be classified into exactly one of **two** distinct types, namely "success" (S) and "failure" (F).
- Of the ***N*** objects, let ***r*** represent the number of *success-type* objects and let ***N-r*** be the number of *failure-type* objects. We pick ***n*** objects (***n* ≤ *N***) at random from the population **without replacement**.
- If *X* represents the number of success-type objects obtained, then *X* has a **hypergeometric distribution**.
- We might sometimes write $X \sim HG(N, r, n)$.

Example:

Of the 120 applicants competing for a job, suppose only 80 of them are qualified. In total, 5 applicants are selected at random for an interview. If X represents the number of qualified applicants that are interviewed, then X has a hypergeometric distribution where:

$$N = 120$$

r = 80 "successes"

N-r = 40 "failures"

$$n = 5$$

And what about the possible values of the r.v. X?

$$A = \{0, 1, 2, 3, 4, 5\}$$

Probability function:

- Using our counting techniques, in which we choose not to consider order, we have:

- Total number of arrangements = $\binom{N}{n}$
- Number of ways to select *x* success-type objects from the *r* available successes = $\binom{r}{x}$
- Number of ways to choose the remaining *n*-*x* objects from the *N* – *r* failures = $\binom{N-r}{n-x}$

- Hence, we obtain the p.f. as follows:

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

where $x \geq \max\{0, n-(N-r)\}$ and $x \leq \min\{r, n\}$.

- Note that $\sum_{x \in A} f(x) = \sum_{x=0}^{\infty} f(x) = 1$.

Back to the applicants example:

Find the probability that two of the five selected will be qualified for the job.

Example:

In the game of *Texas Hold’Em*, each player is dealt two private cards, and five community cards are dealt face up on the table. Each player is to make the best 5-card hand they can with their two private cards and the five community cards.

What is the probability that a particular player will have a spade flush? (i.e. 5 or more spades in this case)?

You Try:

A manufacturer of auto parts just shipped 25 auto parts to a dealer. Later on, it was discovered that 5 of those parts were defective. By the time the company manager contacted the dealer, 4 auto parts from that shipment had been sold.

What is the probability that 3 of the 4 parts that were sold were good parts?

Solution:

Let X = # of good parts in the sample of 4.

Here, X has a Hypergeometric Distribution with N = 25, r = 20, and n = 4. We want P(X = 3).

Here, f(x) = P(X = x) = $\frac{\binom{20}{x}\binom{5}{4-x}}{\binom{25}{4}}$, for x = 0, 1, 2, 3, 4; 0, otherwise.

P(X = 3) = $\frac{\binom{20}{3}\binom{5}{1}}{\binom{25}{4}}$ = 0.450593

Binomial Distribution (Section 5.4)

Physical setup:

- Suppose an experiment has **two** possible outcomes, namely "success" (S) and "failure" (F).
- For a single experiment, let P(success) = p and P(failure) = $1 - p$ where $0 < p < 1$.
- Repeat the experiment ***n*** **independent** times.
- Let X count the number of successes observed.
- The r.v. X is said to have a **binomial distribution**.

We write $X \sim Bin(n, p)$.

Note: The ***n* individual** experiments are called "trials" or "Bernoulli trials" and the process is called a Bernoulli process (or binomial process).

Underlying assumptions:

T: ***Two outcomes***

I: ***Independent trials***

M: ***Multiple trials***

S: ***Same probability of success in each trial***

Example:

A fair coin is tossed 12 times. Let X represent the number of heads obtained. Then, $X \sim Bin(n, p)$ with:

$$n = 12$$

$$p = 0.5$$

What are the possible values of the r.v. X?

$A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Probability function:

- If we are to have *x* successes (S's) (and *n–x* failures (F's)), then the total number of different arrangements of *x* S's and (*n–x)* F's is $\binom{n}{x}$.
- For each of these arrangements, since the trials are independent, the probability is:

$$p^x(1-p)^{n-x}$$

- Putting this altogether, we ultimately obtain:

$f(x) = \binom{n}{x}p^x(1-p)^{n-x}$ for $x = 0, 1, 2, \dots, n.$

- Note that $\sum_{x \in A} f(x) = \sum_{x=0}^{n} f(x) = 1.$

Example:

Seventy-five percent of the students at a college with a large student population use Instagram. A sample of five students from this college is selected. What is probability that at least 3 students use Instagram?

You Try:

At *Skip The Dishes*, home delivery providing high-quality service to customers is the top priority of the management. The company guarantees a refund of all charges if your meal is not delivered within 40 minutes of placing your order. Through past history, it is known that 2% of the deliveries do not arrive within the 40 minutes. Suppose *Skip The Dishes* receives 10 orders during a specific hour.

a) Find the probability that exactly 1 of these orders is not delivered within 40 minutes.

b) Find the probability that at most 1 of these orders is not delivered within 40 minutes.

Skip The Dishes You Try Solution:

At *Skip The Dishes*, home delivery providing high-quality service to customers is the top priority of the management. The company guarantees a refund of all charges if your meal is not delivered within 40 minutes of placing your order. Through past history, it is known that 2% of the deliveries do not arrive within the 40 minutes. Suppose *Skip The Dishes* receives 10 orders during a specific hour.

a) Find the probability that exactly 1 of these orders is not delivered within 40 minutes.

Solution: Let X represent the number of orders that are not delivered within 40 minutes.

X~Binomial(n = 10, p = 0.02). P(X = x) = $\binom{10}{x}(0.02)^x(0.98)^{10-x}$**, x = 0, 1, …., 10 ; 0, oth.**

We want P(X = 1) = $\binom{10}{1}(0.02)^1(0.98)^9$ **= 0.16675**

b) Find the probability that at most 1 of these orders is not delivered within 40 minutes.

Solution: Let X represent the number of orders that are not delivered within 40 minutes.

X~Binomial(n = 10, p = 0.02). P(X = x) = $\binom{10}{x}(0.02)^x(0.98)^{10-x}$**, x = 0, 1, …., 10 ; 0, oth.**

We want P(X ≤ 1) = P(X = 0) + P(X = 1)

P(X ≤ 1) = $\binom{10}{0}(0.02)^0(0.98)^{10} + \binom{10}{1}(0.02)^1(0.98)^9$ **= 0.98382**

Binomial Distribution vs. Hypergeometric Distribution

Similarities:

1. Both have 2 types of outcomes; success and failure.
2. The experiment is repeated *n* times in each.
3. The r.v. *X* records the number of successes.

The Main Difference:

- The binomial distribution requires *n* INDEPENDENT trials, where the probability of success is the *same* in each trial (TIMS).
- In the hypergeometric setting, the *n* draws are made from a fixed number of objects (*N*) WITHOUT replacement. Hence, the trials are NOT independent.

Example: Suppose we have 50 bottles of drinks placed in a big ice container such that the labels are not visible. It is known that 20 are energy drinks and 30 are soda pop. 5 cans are randomly selected. Find the probability that 3 are energy drinks.

Let’s do the previous example assuming that selections are done <u>with</u> replacement (i.e. the draws are independent, which is actually the incorrect assumption).

What does the previous example tell us?

It seems to be suggesting that the binomial distribution can be used as an **approximation** for the hypergeometric distribution.

Why is this so?

When N is large and n, the number of objects being drawn, is relatively small, then the chance of selecting the same object (even if it is placed back in the population before the next draw) would be remote.

When $n/N \leq 0.05$ (or $N/n \geq 20$), the approximation generally works well.

Example: In a recent shipment of 5,000 tires to the *ABC Tire* company, 1,000 of them are slightly marked. Ten tires from the shipment are randomly chosen and purchased by a consumer.

a) What is the **exact** probability that 3 of the 10 tires purchased are slightly marked?

b) Using a suitable approximation, what is the **approximate** probability that 3 of the 10 tires purchased are slightly marked?

How well did the approximation work?

Negative Binomial Distribution (Section 5.5)

Physical setup:

- The setup is very similar to the binomial distribution, in the sense we have two outcomes (success or failure) possible on a single trial (or experiment), independent trials, and the probability of success is the same in each trial (namely, p).
- We continue to repeat the experiment until k successes are obtained.
- If X represents the number of failures obtained before the k^{th} success, then X has a **negative binomial distribution**.
- We write $X \sim NB(k, p)$.

Example:

Draw cards from a standard deck of 52 cards **with replacement** until you get 3 Aces. Let X represent the number of non-Aces that appear before the 3rd Ace is obtained. Then, $X \sim NB(k, p)$ with:

$$k = 3$$

$$p = 4/52 = 1/13$$

What are the possible values of the r.v. X?

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

*A has a **countably infinite** number of values.*

Probability function:

- In total, we have x failures and k successes, which means that the total number of trials is $x + k$.
- However, the very last trial **MUST** be a success.
- Hence, in the first $x+k-1$ trials, we must observe x failures and $k-1$ successes, **in any order**.
- Therefore, the total number of arrangements is

$$\binom{x+k-1}{x}$$

- Each arrangement will have probability

$$p^{k-1}(1-p)^x p = p^k(1-p)^x$$

- Putting this altogether, the p.f. of X is given by

$$f(x) = \binom{x+k-1}{x} p^k (1-p)^x \text{ for } x = 0, 1, 2, \dots .$$

- Note that

$$\sum_{x \in A} f(x) = \sum_{x=0}^{\infty} f(x) = 1.$$

Binomial Distribution vs. Negative Binomial Distribution

- These two distributions can be distinguished by looking at what is specified or known in advance.
- In the case of the binomial distribution, we **know** that there are *n* independent trials, but we **do not know** the number of successes that we will obtain.
- In the case of the negative binomial distribution, we **know** the number of successes *k* to be obtained, but we **do not know** the number of trials (or failures) needed to obtain the *k* successes.

Remark:

- In some textbooks, the negative binomial r.v. is **not** defined as the *number of failures* before the k^{th} success, but instead the *total number of trails* needed to get the k^{th} success.
- Let Y represent this r.v. instead. Clearly, we have

$$Y = X + k \text{ where } X \sim NB(k, p)$$

What is the p.f. of Y?

Example: A start-up company is looking for 5 investors. Each investor will independently agree to invest in the company with probability 20%. The founder asks investors one at a time until 5 "yes" responses are obtained. **Let *X* represent the total number of investors asked.**

a) What is the p.f. of X?

b) Use the p.f. in part a) to calculate $P(X=6)$.

Geometric Distribution (Section 5.6)

Physical setup:

- Again, we have independent Bernoulli trials, each having two possible outcomes (success vs. failure).
- The probability of success, p, is the same on each trial.
- If X represents the **number of failures** before the **FIRST** success occurs (i.e. a $NB(k, p)$ r.v. with k = 1), then X is said to have a **geometric distribution**.
- We write $X \sim Geo(p)$.

Some Illustrative Examples:

- Suppose that the probability you win a lottery prize in any given week is a constant *p*. Then, the number of weeks before you win a prize for the first time has a geometric distribution.

- Pascal is a professional basketball player. He is a 70% free throw shooter. That means his probability of making a free throw is 0.7. If *X* represents the number of free throws before Pascal makes his 1^{st} free throw shot, then *X* has a geometric distribution (with $p = 0.7$).

Probability function:

- There is only **one way** to arrange x failures followed by 1 success, leading to the p.f.

$$f(x) = (1-p)^x p \text{ for } x = 0, 1, 2, \ldots .$$

- Note that we can also obtain the above p.f. as a special case of the negative binomial distribution with $k = 1$.
- We remark that

$$\sum_{x \in A} f(x) = \sum_{x=0}^{\infty} f(x) = 1.$$

In summary:

- We take notice that the binomial, negative binomial, and geometric probability models all assume:

1. Two outcomes in each Bernoulli trial
2. Independent trials
3. The probability of success remains constant from trial to trial

Example:

Suppose that a company receives 60% of its orders over the internet.

a) What is the probability that the fifth order received is the first internet order?

b) What is the probability that the eighth order received is the fourth internet order?

c) What is the probability that more than 3 total orders are required to get the first internet order?

You Try:

Suppose that the probability is 0.75 that an applicant for a driver's license will pass the road test on any given attempt.

What is the probability that an applicant will finally pass the test on the fourth attempt?

Solution:

Let X represent the number of failures before the applicant finally passes the road test (i.e. first success).

X~Geom(p = 0.75)

with P(X = x) = $(0.25)^x(0.75)$, x = 0, 1, …… ; 0, otherwise

Note: If the applicant finally passes on the 4th attempt, then they must have failed on all 3 previous attemps.

So, we want P(X=3).

P(X = 3) = $(0.25)^3(0.75) = \frac{3}{256} =$ 0.01172