CSC 1201 Probability and Statistics

Discrete Random Variables and Distributions

Denish Azamuke

Makerere University

February 24, 2025

Specific Discrete Distributions

In the following, we will study several important discrete distributions.

- Discrete Uniform Distribution
- Bernoulli Distribution
- Binomial Distribution
- Geometric and Negative Binomial Distributions
- Hypergeometric Distribution
- Poisson Distribution

These distributions are very useful in a wide range of scenarios, providing models for many real-world situations.

Discrete Uniform Distribution

Definition: Discrete Uniform Distribution

Let X be a random variable with possible values $x_1, x_2, \ldots, x_n$. If each of the n possible values in the range of X has equal probability, then the random variable is said to have a *discrete uniform distribution*. In other words,

$$P(X = x_i) = f(x_i) = \frac{1}{n}.$$

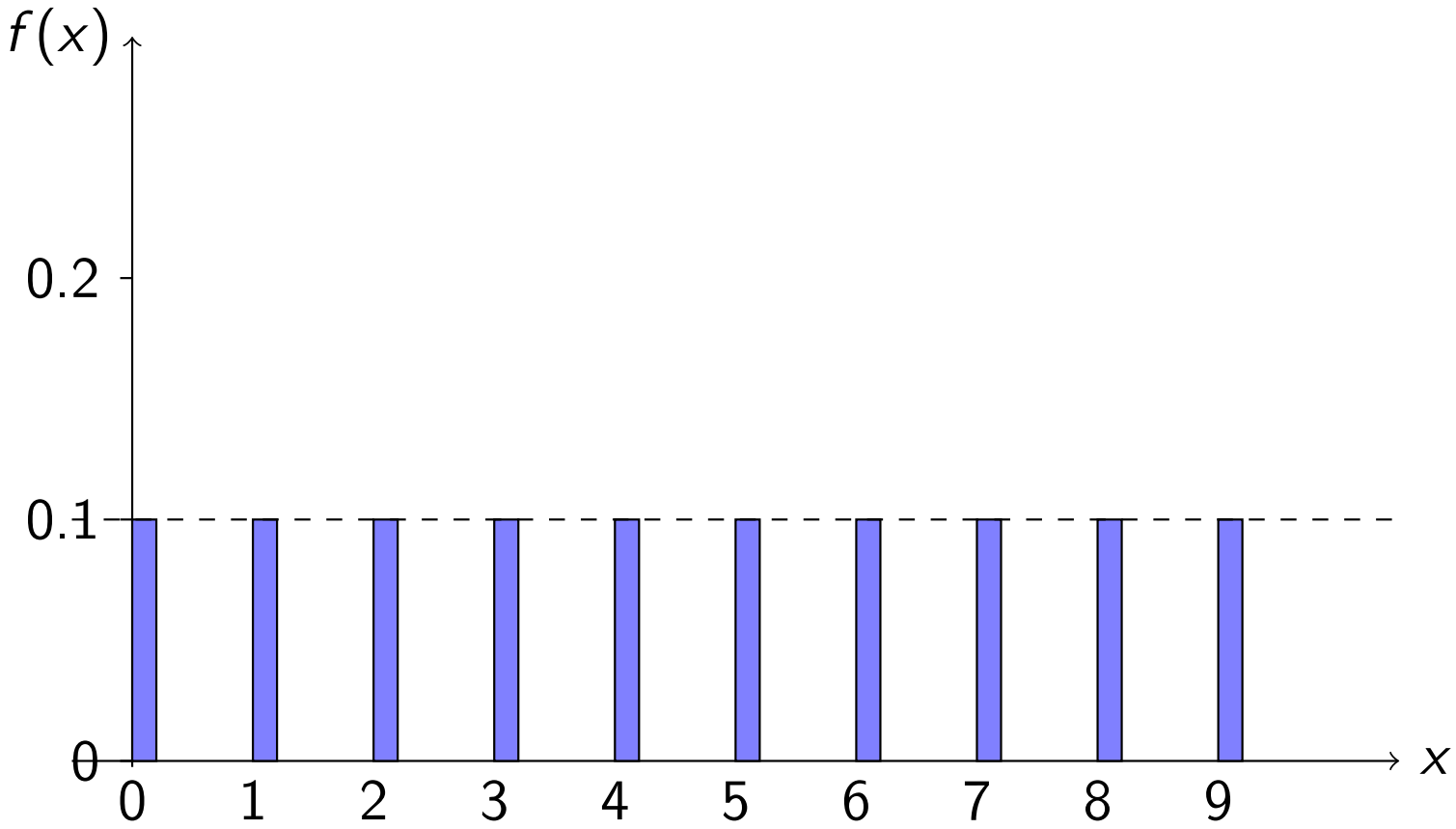
Example: The first digit of a part's serial number is equally likely to be any digit from 0 through 9. If we select one part at random, and let X be the first digit of its serial number, then X has a discrete uniform distribution with probability 0.1 for each value in $R = \{0, 1, 2, \ldots, 9\}$. In other words,

$$f(x) = 0.1$$

for each $x \in R$.

Discrete Uniform Distribution

The probability mass function (pmf) of X is shown schematically below:



Discrete Uniform Distribution

Definition

Suppose X is a random variable with a discrete uniform distribution on the consecutive integers $a, a+1, \ldots, b$, where $a \leq b$. The mean of X is

$$E[X] = \frac{a+b}{2},$$

and the variance of X is

$$V(X) = \frac{(b-a+1)^2 - 1}{12}.$$

Discrete Uniform Distribution

Proof (Sketch)

$$P(X = i) = \frac{1}{b - a + 1} \quad \text{for} \quad i = a, a + 1, \ldots, b.$$

Then,

$$E[X] = \sum_{i=a}^{b} i\, P(X = i) = \sum_{i=a}^{b} i\, \frac{1}{b - a + 1} = \frac{a + b}{2}.$$

A similar summation yields

$$V(X) = \sum_{i=a}^{b} (i - E[X])^2\, P(X = i) = \frac{(b - a + 1)^2 - 1}{12}.$$

Dices

A regular die has six sides, showing 1, 2, 3, 4, 5, and 6 dots, respectively. If the die is fair, then the outcome of rolling it is given by the uniform random variable

$$X \sim \text{Unif}(1, 2, 3, 4, 5, 6).$$

Q1. What is $E[X]$?
Q2. What is $V(X)$?

Answers to the Dice Questions

Q1. What is $E[X]$?

For a fair die with faces $\{1, 2, 3, 4, 5, 6\}$:

$$X \sim \text{Uniform}\{1, 2, 3, 4, 5, 6\},$$

each outcome occurs with probability $\frac{1}{6}$. Since $\{1, 2, \ldots, 6\}$ corresponds to $a = 1$ and $b = 6$, the mean is

$$E[X] = \frac{a+b}{2} = \frac{1+6}{2} = 3.5.$$

Q2. What is $V(X)$?

We use the discrete uniform variance formula:

$$V(X) = \frac{(b-a+1)^2 - 1}{12}.$$

Plugging in $a = 1$ and $b = 6$ gives

$$V(X) = \frac{(6-1+1)^2 - 1}{12} = \frac{6^2 - 1}{12} = \frac{35}{12} \approx 2.9167.$$

Bernoulli Random Variables

A discrete random variable X is said to have a **Bernoulli distribution** with parameter p if its probability mass function is given by

$$f(x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}.$$

Definition: Bernoulli Random Variable

Let X be a random variable taking only two possible values $\{0, 1\}$. Let $p = P(X = 1)$. Then X is said to be a Bernoulli random variable with parameter p, denoted $X \sim \text{Ber}(p)$.

Bernoulli Random Variables

Properties

If $X \sim \text{Ber}(p)$, then:

$$E[X] = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = p,$$

$$E[X^2] = 0^2 \cdot P(X = 0) + 1^2 \cdot P(X = 1) = p,$$

$$V(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p).$$

Bernoulli Trials

A **Bernoulli trial** is a random experiment in which there are exactly two possible outcomes, typically called *failure* (F) and *success* (S). We define a random variable

$$X : \{F, S\} \rightarrow \{0, 1\},$$

such that

$$X(F) = 0, \quad X(S) = 1.$$

The **Bernoulli distribution** arises when the following three conditions hold:

1. Each trial of an experiment results in an outcome that may be classified as success or failure.
2. The probability of success $P(S) = p$ is the same for each trial.
3. The trials are independent; that is, the outcome of one trial has no effect on the outcome of any other trial.

Note: Independent Bernoulli random variables are important building blocks to construct more complicated random variables.

Bernoulli Random Variables

Example 1: What is the probability of getting a score of *not less than 5* on a throw of a six-sided die?

Answer: Although there are six possible outcomes $\{1, 2, 3, 4, 5, 6\}$, we can group them into two sets: $\{1, 2, 3, 4\}$ and $\{5, 6\}$.
Any outcome in $\{1, 2, 3, 4\}$ is deemed a *failure*, while any outcome in $\{5, 6\}$ is a *success*. Hence, this scenario can be viewed as a Bernoulli trial, with

$$P(X = 0) = P(\text{failure}) = \frac{4}{6}, \qquad P(X = 1) = P(\text{success}) = \frac{2}{6}.$$

Therefore, the probability of getting a score of at least 5 (i.e., "not less than 5") is

$$\frac{2}{6}.$$

Repeated Bernoulli Trials

Example 2: Suppose we send 3 packets through a communication channel. Each packet is independently received with probability $p = 0.9$. Denote by

$$X_i = \begin{cases} 1 & \text{if packet } i \text{ is received,} \\ 0 & \text{if packet } i \text{ is lost,} \end{cases} \qquad i = 1, 2, 3,$$

so that each $X_i \sim \mathrm{Ber}(0.9)$. Let $X = X_1 + X_2 + X_3$ be the total number of packets received.

Because the X_i's are independent,

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \;=\; (0.9)^{x_1}(0.1)^{1-x_1} \;\times\; (0.9)^{x_2}(0.1)^{1-x_2} \;\times\; (0.9)^{x_3}(0.1)^{1-x_3}.$$

The table below enumerates all possible outcomes:

Repeated Bernoulli Trials

The table below enumerates all possible outcomes:

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Binomial Distribution

From the previous example (sending 3 packets where each is received with probability 0.9 independently), we had the following table:

x_1	x_2	x_3	$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$	$x_1 + x_2 + x_3$
0	0	0	0.001	0
0	0	1	0.009	1
0	1	0	0.009	1
0	1	1	0.081	2
1	0	0	0.009	1
1	0	1	0.081	2
1	1	0	0.081	2
1	1	1	0.729	3

Therefore, if $X = X_1 + X_2 + X_3$ is the total number of received packets, then:

Binomial Distribution

Therefore, if $X = X_1 + X_2 + X_3$ is the total number of received packets, then:

$$P(X = 0) = 0.001, \quad P(X = 1) = 0.027, \quad P(X = 2) = 0.243, \quad P(X = 3) = 0.729.$$

This is what is called a Binomial random variable!

A random variable is called a **binomial random variable** if it represents the total number of successes in n independent Bernoulli trials, each with the same probability of success p. In this example,

$$X \sim \text{Binomial}(n = 3,\ p = 0.9).$$

Binomial Distribution

Definition: Binomial Random Variable

Consider a random experiment consisting of $n \in \mathbb{N}$ independent Bernoulli trials, where each trial results in either *success* or *failure*. Assume the probability of success in each trial is p (with $0 \leq p \leq 1$), and it remains the same throughout all n trials.

The random variable X that counts the number of trials resulting in success is said to follow a **Binomial distribution** with parameters n and p. We write $X \sim \text{Bin}(n, p)$.

The probability mass function (pmf) of X is given by:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, \ldots, n.$$

We have seen $\binom{n}{k}$ before: it represents the number of ways to choose k items out of n.

Binomial Distribution

The name *binomial distribution* arises from its close resemblance to the **binomial expansion**:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}.$$

Indeed, this similarity shows that the distribution's pmf is valid, since

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+(1-p))^n = 1.$$

Computing the *mean* and *variance* of X *directly* from

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

is straightforward in principle but can be rather tedious. **There is a much easier way...**

Binomial Distribution

Note: Let

$$X = X_1 + X_2 + \cdots + X_n,$$

where each $X_i \sim \mathrm{Ber}(p)$ and the X_i's are independent Bernoulli trials.

- For independent random variables $X_1, \ldots, X_n$,

$$E\left[X_1 + X_2 + \cdots + X_n\right] = E[X_1] + E[X_2] + \cdots + E[X_n],$$
$$V\left(X_1 + X_2 + \cdots + X_n\right) = V(X_1) + V(X_2) + \cdots + V(X_n).$$

Since $E[X_i] = p$ and $V(X_i) = p(1-p)$ for each i,

$$E[X] = np \quad \text{and} \quad V(X) = np\,(1-p).$$

Hence X is a *Binomial* random variable with parameters n and p.

If $X \sim \mathrm{Bin}(n, p)$, **then**

$$E[X] = np, \quad V(X) = np(1-p).$$

Binomial Distribution Examples

Example 1: On a five-question multiple-choice test, there are five possible answers, exactly one of which is correct. If a student guesses randomly and independently on each question, what is the probability that she is correct on exactly two questions?

Example 2: Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the presence or absence of the pollutant in each sample is independent of all other samples.

a) Find the probability that in the next 18 samples, exactly 2 contain the pollutant.

b) Determine the probability that at least 4 samples contain the pollutant.

c) Find the probability that $3 \leq X < 7$, i.e., $X = 3, 4, 5, 6$, where X is the number of polluted samples.

Solution to Example 1

Problem Recap: A student takes a five-question multiple-choice test, each with 5 possible answers (and exactly 1 correct). The student guesses randomly and independently on each question. We want the probability that she is correct on exactly two questions.

Solution: Let X be the number of correct answers. Since each question has probability $p = \frac{1}{5} = 0.2$ of being correct and the 5 questions are independent,

$$X \sim \text{Binomial}(n = 5,\ p = 0.2).$$

Thus,

$$P(X = 2) \ = \ \binom{5}{2} (0.2)^2 (0.8)^3.$$

Numerically,

$$\binom{5}{2} = 10, \quad (0.2)^2 = 0.04, \quad (0.8)^3 = 0.512.$$

So

$$P(X = 2) \ = \ 10 \times 0.04 \times 0.512 \ = \ 0.2048 \ \approx \ 20.48\%.$$

Solutions to Example 2

Let X be the number of polluted samples out of $n = 18$. Each sample is polluted with probability $p = 0.1$ independently. Hence, $X \sim \text{Binomial}(18, 0.1)$.

a) **Probability that exactly 2 samples contain the pollutant:**

$$P(X = 2) = \binom{18}{2}(0.1)^2(0.9)^{16}.$$

b) **Probability that at least 4 samples contain the pollutant:**

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \big[P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)\big].$$

c) **Probability that $3 \leq X < 7$, i.e. $X \in \{3, 4, 5, 6\}$:**

$$P(3 \leq X < 7) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6).$$

Each term is computed using the pmf $\binom{18}{k}(0.1)^k(0.9)^{18-k}$.

Tip: Use a binomial table or software (e.g., R, Python, or a statistical calculator) to find exact values or approximate numerical answers.

Example 2: Summary of Computations

Setup: Suppose you collect 20 samples over time. Let $X \sim \text{Binomial}(n = 20,\ p = 0.1)$, where X is the number of contaminated samples (each sample has a 10% chance of containing the pollutant, independently of the others).

- $P(X = 2) = \binom{20}{2} (0.1)^2 (0.9)^{18} \approx 0.285.$
- $P(X \geq 2) = 1 - P(X \leq 1) = 1 - \left[P(X = 0) + P(X = 1)\right] \approx 0.608.$
- $P(X \leq 7) \approx 0.9996$ (from table or software).
- $P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 2) \approx 0.3207.$
- $P(3 < X < 7) = P(X = 4) + P(X = 5) + P(X = 6) \approx 0.1306.$
- $E[X] = n\,p = 20 \times 0.1 = 2.$
- $V(X) = n\,p\,(1 - p) = 20 \times 0.1 \times 0.9 = 1.8.$

Geometric Distribution

Let X be the number of times you must send a packet over a network until it is successfully received. Assume each transmission attempt is independent, and the probability of success (receipt) on any single attempt is $p = 0.9$.

Sample Space: $\{1, 2, 3, \dots\}$

Probability Computations:

$$P(X = 1) = p = 0.9.$$

$$P(X = 2) = (1 - p)\, p = (1 - 0.9) \times 0.9 = 0.09.$$

$$P(X = k) = (1 - p)^{k-1} p \quad \text{for } k = 1, 2, 3, \dots$$

Definition: A random variable X in a sequence of independent Bernoulli trials that equals the *number of trials up to and including the first success* is said to follow a **Geometric distribution** with parameter p. We write

$$X \sim \mathrm{Geom}(p).$$

Geometric Distribution

Definition: Geometric Random Variable

Consider a series of independent Bernoulli trials, each with a constant probability of success p. Let the random variable X denote the *number of trials up to and including the first success*. Then X is called a **geometric random variable** with parameter p, and its probability mass function is

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Mean and Variance

If $X \sim \text{Geom}(p)$, then

$$E[X] = \frac{1}{p}, \qquad V(X) = \frac{1-p}{p^2}.$$

Geometric Distribution: Example 1

Let X be the number of times you need to send a packet over a network to ensure it is received. Assume each transmission is independent, and the probability of a successful transmission is $p = 0.9$.

Sample Space: $\{1, 2, 3, \dots\}$

$$E[X] = \frac{1}{p} = \frac{1}{0.9} \approx 1.111, \qquad V(X) = \frac{1-p}{p^2} = \frac{0.1}{0.9^2} \approx 0.1234.$$

Geometric Distribution: Example 2

A gambler plays roulette at Monte Carlo, repeatedly betting on "Red" until his first win. The probability of "Red" (success) on any single spin is $p = \frac{18}{38} \approx 0.4737$. He has enough money for at most 5 bets.

(a) What is the probability that he wins at least once before running out of money?

Solution

Winning before exhausting his funds (i.e., in 5 or fewer spins):

$$P(\text{Win before 6th spin}) = P(X \leq 5) = \sum_{k=1}^{5} p\,(1-p)^{k-1}.$$

Numerically,

$$p = \frac{18}{38} \approx 0.4737, \quad 1 - p \approx 0.5263.$$

$$P(X \leq 5) = \sum_{k=1}^{5} 0.4737 \times (0.5263)^{k-1} \approx 0.4737 + 0.2496 + 0.1313 + 0.0692 + 0.0365 \approx 0.9603.$$

(For more precision, use calculator or software.)

Geometric Distribution: Example 2

A gambler plays roulette at Monte Carlo, repeatedly betting on "Red" until his first win. The probability of "Red" (success) on any single spin is $p = \frac{18}{38} \approx 0.4737$. He has enough money for at most 5 bets.

(b) What is the probability that he wins on the second bet?

Solution

Winning on the second bet:

Recall

$$P(X = k) = p\,(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Which implies

$$P(X = 2) = (1-p)\,p = 0.5263 \times 0.4737 \approx 0.2496.$$

Properties of the Geometric Distribution

CDF: For a geometric random variable X with parameter p,

$$F(k) = P(X \leq k) = \sum_{i=1}^{k} p\,(1-p)^{i-1} = 1 - (1-p)^{k}, \quad k = 1, 2, \ldots$$

Memoryless Property: The geometric distribution has "no memory" of the past. Formally,

$$P(X = k + m \mid X > k) = P(X = m).$$

This means that, given you have already waited k trials without success, the probability you still need to wait m additional trials is the same as if you were starting fresh.

Properties of the Geometric Distribution

Derivation:

$$P(X = k + m \wedge X > k) = P(X = k + m),$$

since "$X = k + m$" already implies "$X > k$." Meanwhile,

$$P(X = k + m) = p\,(1 - p)^{k+m-1}.$$

and

$$P(X > k) = (1 - p)^{k}.$$

Hence,

$$P(X = k + m \mid X > k) = \frac{p\,(1 - p)^{k+m-1}}{(1 - p)^{k}} = p\,(1 - p)^{m-1} = P(X = m).$$

Memoryless Property

$$P\big(X = k + m \mid X > k\big) \;=\; P(X = m).$$

The geometric distribution is the only discrete distribution satisfying this relation.

This implies that objects or components with no "wear" under normal conditions have their time-to-failure well modeled by a geometric distribution.

Examples:

- The number of days until an LED fails (assuming no gradual degradation).
- The number of years a resistor lasts, given appropriate heat dissipation.

Negative Binomial Distribution

We can generalize the idea behind the geometric distribution (waiting for the *first* success) to instead wait until we achieve r successes.

Example: Packets are sent through a communication channel, each packet being transmitted independently with success probability p. Let X be the number of transmissions required to receive r successful packets.

- For $k < r$, $P(X = k) = 0$, since you cannot have r successes in fewer than r trials.
- For $k = r$, all r transmissions must succeed:

$$P(X = r) = p \times p \times \cdots \times p \; = \; p^r.$$

- For $k \geq r$,

$$P(X = k) \; = \; \binom{k-1}{r-1} (1-p)^{k-r} \, p^r,$$

where the first r successes can occur in any arrangement among the k trials, but the k-th trial must be the r-th success.

Negative Binomial Distribution

Definition: Negative Binomial Random Variable

Consider a series of independent Bernoulli trials with constant success probability p. Let the random variable X be the total number of trials needed to achieve r successes. Then X is called a **negative binomial random variable** with parameters r and p, written $X \sim \mathrm{NB}(r, p)$. Its pmf is:

$$P(X = k) = \begin{cases} 0, & k < r, \\ \binom{k-1}{r-1} (1-p)^{k-r} p^r, & k \geq r,\ k \in \mathbb{N}. \end{cases}$$

Geometric Distribution as a Special Case:
When $r = 1$, the negative binomial reduces to the geometric distribution.

Why "negative binomial"?
Despite the name, there is nothing negative in the definition. The terminology reflects how it contrasts with the binomial distribution:

- In the *binomial* model, we fix the number of trials n and count how many successes occur.
- In the *negative binomial* model, we fix the number of successes r and count how many trials are needed.

Properties of the Negative Binomial Distribution

Interpretation: A negative binomial random variable can be viewed as the sum of r independent geometric random variables, each with success probability p. Symbolically,

$$X = G_1 + G_2 + \cdots + G_r$$

where $G_i \sim \text{Geom}(p)$ and the G_i's are i.i.d.

Mean and Variance: If $X \sim \text{NegBin}(r, p)$, then

$$E[X] = \frac{r}{p}, \quad V(X) = \frac{r(1-p)}{p^2}.$$

This sum-of-geometric-random-variables perspective provides an easy way to compute $E[X]$ and $V(X)$ using known properties of the geometric distribution.

Negative Binomial Distribution: Examples

Example 1: What is the probability that the fifth head occurs on the 10^{th} flip of a fair coin?

Solution: Let X be the total number of flips needed to observe the 5^{th} head. Then $X \sim \text{NegBin}(r = 5,\, p = \frac{1}{2})$. We want $P(X = 10)$:

$$P(X = 10) = \binom{10-1}{5-1}\left(\tfrac{1}{2}\right)^{5}\left(\tfrac{1}{2}\right)^{10-5} = \binom{9}{4}\left(\tfrac{1}{2}\right)^{10} = 126 \times \frac{1}{1024} = \frac{63}{512}.$$

Example 2: A survey-taker needs 3 completed surveys before finishing. Each phone call independently has a 9% chance of reaching an adult who will complete the survey. What is the probability that the 3^{rd} completed survey occurs on the 10^{th} call?

Solution: Let Y be the total number of calls needed to get 3 completed surveys. Then $Y \sim \text{NegBin}(r = 3,\, p = 0.09)$. We want $P(Y = 10)$:

$$P(Y = 10) = \binom{10-1}{3-1}(0.09)^{3}(0.91)^{7} = \binom{9}{2}(0.09)^{3}(0.91)^{7}.$$

(Exact or approximate numerical values can be found using a calculator or software.)

Revision Questions

Question 1: Customers at Fred's Cafe win \$100 if their receipts show a star on each of the five consecutive days (Mon–Fri) in a given week. The cash register prints a star with probability 10% on any given receipt, independently. If Mark dines at Fred's Cafe once a day for four consecutive weeks, what is the probability that he wins at least \$100?

Question 2: Thirty percent of all electrical fuses manufactured by a certain company fail to meet municipal building standards. In a random sample of 10 fuses, what is the probability that exactly 3 fail to meet those standards?

Question 3: A player in a video game faces a series of opponents, each with an independent 80% probability of being defeated ($p = 0.8$). The player continues to face new opponents until defeated for the first time.

1. What is the probability mass function (pmf) of the number of opponents contested in a single game?
2. What is the probability that the player defeats at least two opponents in a game?
3. What is the expected number of opponents contested in a game?
4. What is the probability that a player contests four or more opponents in a game?
5. What is the expected number of game plays until a player contests four or more opponents in at least one game?

Revision Questions

Question 4: Heart failure can be attributed to either **natural occurrences** (87%) or **outside factors** (13%). - Outside factors relate to induced substances or foreign objects. - Natural occurrences can be caused by arterial blockage, disease, or infection.
Assume each patient's cause of heart failure is independent of the others.

1. What is the probability that the first patient with heart failure to enter the emergency room has a condition due to outside factors?
2. What is the probability that the *third* patient with heart failure to enter the emergency room is the *first* one due to outside factors?
3. What is the mean (expected) number of heart failure patients whose condition is due to natural causes entering the emergency room *before* the first patient with heart failure from outside factors?

Solution to Some RQs

Question 1: Customers at Fred's Cafe win \$100 if their receipts show a star on each of 5 consecutive days (Mon–Fri). Mark visits once a day for 4 consecutive weeks (so 4 "tries" to get 5 stars in a row). Each receipt has a star with probability 0.1, independently of other days.

Let $A =$ "5 stars in a single Mon–Fri week." Then

$$P(A) \;=\; 0.1^5 \;=\; 0.00001.$$

The probability that Mark does *not* get 5 stars in a row in a given week is $1 - 0.1^5$. Hence, the probability he never wins over 4 weeks is

$$\left(1 - 0.1^5\right)^4,$$

so the probability he wins at least once is

$$1 - \left(1 - 0.1^5\right)^4 \;\approx\; 0.00004 \;=\; 0.004\%.$$

Solutions to Some RQs

Question 2: Suppose 30% of electrical fuses fail to meet standards. In a random sample of 10 fuses, what is the probability that exactly 3 fail?

Model the number of failing fuses by $X \sim \text{Binomial}(n = 10, p = 0.3)$. Then

$$P(X = 3) = \binom{10}{3}(0.3)^3(0.7)^7 = 120 \times 0.027 \times 0.0823543$$

$$\approx 0.267 = 26.7\%.$$

Hypergeometric Distribution

Consider a bag containing N balls, of which K are green and $N-K$ are red. We draw one ball at random (each ball equally likely). Define the random variable

$$X = \begin{cases} 1, & \text{if the drawn ball is green,} \\ 0, & \text{if the drawn ball is red.} \end{cases}$$

Clearly, X is a Bernoulli random variable with success probability

$$P(X = 1) = \frac{K}{N}.$$

When multiple draws occur without replacement, the probability of drawing a certain number of green balls changes after each draw, giving rise to the **Hypergeometric Distribution**. Specifically, if we draw n balls (without replacement) from the bag and let Y be the number of green balls in that sample, then Y follows a Hypergeometric distribution.

Hypergeometric Distribution

Scenario: If we *replace* each ball drawn from the bag (and re-shuffle), then each draw is independent, and the number of green balls drawn in n trials follows a **Binomial** distribution:

$$X \sim \text{Binomial}\left(n, \frac{K}{N}\right).$$

But what if we do not replace the balls?

- The outcome of each draw depends on the outcomes of previous draws.
- In this scenario, we get what is known as a **Hypergeometric** random variable.

When drawing n balls *without* replacement from a bag of N total balls (with K of them being green), the number of green balls in the drawn sample no longer follows a binomial distribution, but a **Hypergeometric** distribution.

Hypergeometric Distribution

Definition: Hypergeometric Random Variable

Consider a set of N objects, of which K are labeled "good" and the remaining $N - K$ are labeled "bad." Suppose we draw a sample of n objects *without replacement*. Let X be the number of good objects in the sample.

Then X is called a **Hypergeometric** random variable, and its probability mass function (pmf) is given by

$$P(X = i) = \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}} \quad \text{for } \max\{0, n-(N-K)\} \leq i \leq \min\{K, n\},$$

and $P(X = i) = 0$ otherwise.

Remark: There are multiple approaches to deriving this pmf (e.g., direct combinatorial arguments, conditional probability, etc.).

Derivation of the Hypergeometric PMF

Setup: We have a total of N objects, of which K are "good" and $N-K$ are "bad." We draw n objects *without replacement*, and let X be the number of good objects in our sample.

Goal: Compute $P(X = i)$.

1. **Count the favorable outcomes:** To get exactly i good objects in our sample of n:

$$\binom{K}{i} \times \binom{N-K}{n-i}$$

   The first factor picks which i good objects appear, the second factor picks the remaining $(n-i)$ from the bad ones.

2. **Count the total outcomes:** There are $\binom{N}{n}$ ways to choose *any* set of n objects from N.

3. **Probability:** Since each sample is equally likely,

$$P(X = i) = \frac{\binom{K}{i}\,\binom{N-K}{n-i}}{\binom{N}{n}},$$

   valid for $\max\{0,\, n-(N-K)\} \le i \le \min\{K, n\}$.

Hence, X follows a **Hypergeometric distribution**: $X \sim \text{Hypergeom}(N, K, n)$.

Properties of the Hypergeometric Distribution

If $X \sim \text{Hypergeom}(N, K, n)$, then:

$$E[X] = n\,p, \quad \text{where } p = \frac{K}{N},$$

$$V(X) = n\,p\,(1-p)\,\frac{N-n}{N-1}.$$

Comments:

- The expectation $E[X] = np$ resembles that of a Binomial(n, p).
- The variance looks similar to $np(1-p)$, but includes the factor $\frac{N-n}{N-1}$. This is often called a *finite population correction factor*, reflecting that draws *without replacement* reduce variability compared to independent draws.

Hypergeometric Distribution: Example 1

A batch of parts contains 100 from a local supplier and 200 from an out-of-state supplier (so 300 total). If four parts are selected *randomly and without replacement*, answer the following:

(a) What is the probability that all four selected parts are from the local supplier?

Setup: Let X be the number of local (out of 100) parts in the sample of 4 drawn *without replacement* from the total $N = 300$. Then $X \sim \text{Hypergeom}(N = 300,\ K = 100,\ n = 4)$.

Solution: $P(\text{all 4 are local}) = P(X = 4)$

$$= \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = \frac{\binom{100}{4}}{\binom{300}{4}}.$$

Hypergeometric Distribution: Example 1

A batch of parts contains 100 from a local supplier and 200 from an out-of-state supplier (so 300 total). If four parts are selected *randomly and without replacement*, answer the following:

(b) What is the probability that two or more parts in the sample are from the local supplier?

Solution: Probability that two or more are local is

$$P(X \geq 2) = 1 - \big[P(X = 0) + P(X = 1)\big].$$

Use the pmf:

$$P(X = 0) = \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}}, \quad P(X = 1) = \frac{\binom{100}{1}\binom{200}{3}}{\binom{300}{4}}.$$

Hypergeometric Distribution: Example 1

A batch of parts contains 100 from a local supplier and 200 from an out-of-state supplier (so 300 total). If four parts are selected *randomly and without replacement*, answer the following:

(c) What is the probability that at least one part in the sample is from the local supplier?

Solution: Probability that at least one part is from the local supplier is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}}.$$

Note: Numerical values can be found by evaluating these binomial coefficients and ratios.

Revision Question

A company that packages breakfast cereals regularly assesses the quality of their product. Unfortunately, the assessment procedure is somewhat destructive, because they must open both the box and the cereal bag.

They know that, in a batch of 1000 boxes, there are 32 with a defect detectable by the assessment procedure. Suppose they select 100 boxes *at random* for testing.

Question: What is the probability that they will detect *at least one* defective box?

Hypergeometric Distribution: Example 2

A company packages breakfast cereals and must occasionally test for defects. Because the assessment procedure is destructive, they test a sample of boxes rather than the entire batch.

Setup:

- Total boxes in the batch: $N = 1000$
- Defective boxes (detectable by the procedure): $K = 32$
- Boxes selected for testing: $n = 100$

Define X to be the number of defective boxes in the tested sample. Since the boxes are sampled *without replacement*,

$$X \sim \text{Hypergeom}\big(N = 1000,\ K = 32,\ n = 100\big).$$

Probability of detecting ≥ 1 defective box:

$$P(X > 0) = 1 - P(X = 0) = 1 - \frac{\binom{32}{0}\binom{1000-32}{100}}{\binom{1000}{100}} \approx 0.9675.$$

Caution!

Binomial Approximation (if we mistakenly assume independence):

$$X \sim \text{Binomial}\big(n = 100,\ p = \tfrac{32}{1000}\big).$$

Then

$$P(X > 0) = 1 - (1 - \tfrac{32}{1000})^{100} \approx 0.9613.$$

Thus, ignoring the "without replacement" aspect slightly *underestimates* the probability of detecting at least one defective box.

Poisson Distribution

The **Poisson distribution** arises in a variety of contexts, especially where we want to model the number of events that occur over a fixed period of time or in a given region of space. It is particularly relevant for studying queues, as it can represent the arrival of requests to a server over time.

Illustrative Example: Imagine counting how many customers enter a store during one day. We make the following assumptions:

- Each customer arrives *independently* of the others.
- The expected (average) rate of customers per unit time is *constant* (does not change over the day).

Under these conditions, the random variable X (the number of customers arriving in one day) can often be well-modeled by a Poisson distribution with some parameter λ, which represents the mean number of arrivals per day.

Derivation of the Poisson Distribution

One common way to derive the Poisson distribution is as a *limit* of the Binomial distribution:

- **Partition the time interval (e.g., one day) into n small subintervals.**
  - Let each subinterval be so small that the probability of exactly one arrival in a subinterval is $p = \frac{\lambda}{n}$, and the probability of more than one arrival is negligible.
- **Model the number of arrivals X with a Binomial:**

$$X \sim \text{Binomial}\left(n,\, p = \tfrac{\lambda}{n}\right).$$

Hence,

$$P(X = k) = \binom{n}{k} \left(\tfrac{\lambda}{n}\right)^k \left(1 - \tfrac{\lambda}{n}\right)^{n-k}.$$

Derivation of the Poisson Distribution

- **Take the limit as $n \to \infty$ while $\lambda = n\,p$ stays fixed.** We use the fact that $\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}$ as $n \to \infty$.
  Detailed expansions show:

$$\lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\,\frac{\lambda^k}{k!}.$$

Thus, in the limit of many small "trial intervals," the number of arrivals in the entire period follows a **Poisson**(λ) distribution:

$$P(X = k) = e^{-\lambda}\,\frac{\lambda^k}{k!}.$$

Poisson Distribution

Definition: Poisson Random Variable

A random variable X is said to have a **Poisson distribution** with parameter $\lambda > 0$ if its pmf is given by

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Key Properties

If $X \sim \text{Poisson}(\lambda)$, then

$$E[X] = \lambda \quad \text{and} \quad V(X) = \lambda.$$

The fact that $E[X] = V(X)$ reflects the nature of Poisson processes, where λ interprets both the average rate of events and the distribution's variance.

Poisson Distribution: Key Points

Poisson process

- A Poisson process models events occurring *independently* and *randomly* over continuous time, at a constant average rate λ.
- Key properties:
  - *Independent increments*: the number of events in disjoint time intervals are independent.
  - *Stationary increments*: the distribution depends only on the length of the time interval, not on its position.
  - The number of events in a time interval of length t follows a Poisson(λ_t) distribution.

Poisson Distribution: Key Points

Poisson random variable, pmf, mean, and variance.

- A Poisson random variable X (with parameter λ) represents the count of events occurring in a fixed interval (time or space) under a Poisson process assumption.
- Its pmf is

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

- The expected value and variance are both λ:

$$E[X] = \lambda, \quad V(X) = \lambda.$$

Poisson Distribution: Exercises

Exercise 1: Customers arrive at a bank according to a Poisson process with a constant average rate of 8.6 customers per hour. Suppose we begin observing the bank at some point in time.

1. What is the expected number of customers arriving in the first 30 minutes?
2. What is the probability that 3 customers arrive in the first 30 minutes?
3. What is the probability that 5 or fewer customers arrive in the first 30 minutes?

Poisson Distribution: Exercises

Exercise 2: A 1 nanogram sample of plutonium-239 undergoes an average of 2.3 radioactive decays per second. The number of decays follows a Poisson distribution.

1. What is the probability that there are exactly 3 radioactive decays in a 2-second period?
2. What is the probability that there are at most 3 radioactive decays in that 2-second period?

Poisson Distribution: Solution

Exercise 1:

- Rate per hour: $\lambda_{\text{hour}} = 8.6$.
- Interval length: 30 mins $= 0.5$ hours.
- $\lambda_{30\text{ min}} = 8.6 \times 0.5 = 4.3$.
  1. $\mathbf{E}[\mathbf{X}] = 4.3$.
  2. $P(X = 3) = e^{-4.3}\,\dfrac{4.3^3}{3!}$.
  3. $P(X \leq 5) = \displaystyle\sum_{k=0}^{5} e^{-4.3}\,\frac{4.3^k}{k!}$.

Poisson Distribution: Solution

Exercise 2:

- Rate per second: $\lambda_{\text{sec}} = 2.3$.
- Interval length: 2 seconds.
- $\lambda_{2\ \text{sec}} = 2.3 \times 2 = 4.6$.
  1. $P(X = 3) = e^{-4.6}\,\dfrac{4.6^3}{3!}$.
  2. $P(X \leq 3) = \displaystyle\sum_{k=0}^{3} e^{-4.6}\,\frac{4.6^k}{k!}$.

Next Lecture