Chapter 5: Discrete Probability Distribution

Engr 0021

Review

- Random variable
- Discrete: pmf, CDF
- Continuous: pdf, CDF
- Mean
- Variance

Learning Objectives

- Binomial distribution
- Hypergeometric distribution
- Poisson distribution and the Poisson process

Review of Discrete Random Variables

- A **discrete** random variable has possible values that form a finite set or can be listed in an infinite sequence ("countable").
- The probability distribution of a random variable X describes how total probability 1 is distributed among outcomes for X.

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

The Bernoulli Process

- The experiment consists of n repeated trials.
- Each trial results in success or failure.
- The probability of success, p, is constant from trial to trial.
- The repeated trials are independent.

Bernoulli Random Variable

Let X indicate the outcome of a single trial:

$$X = \begin{cases} 1, & \text{success} \\ 0, & \text{failure} \end{cases} \qquad \text{denoted } X \sim \text{Bernoulli}(p).$$

The pmf is

$$f(a) = P(X = a) = \begin{cases} p, & a = 1, \\ 1 - p, & a = 0, \end{cases} \qquad 0 \le p \le 1.$$

Motivating Binomial Examples

Consider:

- Flip a coin 10 times; $X =$ number of heads.
- A machine produces 1% defective parts; $X =$ number defective in next 25.
- In the next 20 births; $X =$ number of female births.

Binomial Experiment Conditions

An experiment is **binomial** if:

1. It consists of n trials (fixed in advance).
2. Trials are identical; each trial is success (S) or failure (F).
3. Trials are independent.
4. Probability of success is constant: p.

Example (Setup)

Assume the probability that a stolen car will be recovered is 60%. If 3 cars are stolen, what is the probability that exactly 1 of them will be recovered?

Example (Solution)

Let X = number of recovered cars. Then $X \sim \mathrm{Bin}(n = 3, p = 0.6)$.

$$P(X=1) = \binom{3}{1} 0.6 \times (0.4)^2$$

$$= .288$$

Binomial Distribution

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

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

Example (Binomial Probabilities)

If $P(\text{pass}) = 0.8$ and $n = 10$ students:

(a) $P(\text{all 10 pass})$

(b) $P(\text{exactly 3 pass})$

(c) $P(\text{at least 8 pass})$

$$P(X=10) = \binom{10}{10} 0.8^{10} 0.2^{0} = 0.8^{10}$$

$$P(X=10) = \binom{10}{3} 0.8^{3} 0.2^{7}$$

$$P(X \geq 8) = P(X=8) + P(X=9) + P(X=10)$$

$$\sum_{x=8}^{10} \binom{10}{x} 0.8^{x} 0.2^{10-x}$$

CDF for $X \sim \mathrm{Bin}(n, p)$

For an integer a with $0 \leq a \leq n$,

$$P(X \leq a) = \sum_{x=0}^{a} P(X = x).$$

Mean and Variance for Binomial

For $X \sim \mathrm{Bin}(n, p)$,

$$\mu = E[X] = np, \qquad \sigma^2 = \mathrm{Var}(X) = np(1-p) = npq, \quad q = 1 - p.$$

Example (Credit Card Purchases)

If 70% of purchases are made with a credit card and X is the number among 10 purchases made with a credit card:

(a) What is the distribution of X?

(b) Find $E[X]$.

(c) Find $\mathrm{Var}(X)$.

(d) Find $P(|X - \mu| \leq \sigma)$.

$X \sim Bin(10, 0.7)$

$E(X) = 7$

$\sigma^2 = npq = 10(.7)(.3) = 2.1$

$\sigma = \sqrt{2.1} = 1.4$

$P(X - 7 \leq 1.4$ or $7 - X \leq 1.4) =$

$P(6 \leq X \leq 8) = \sum_{x=6}^{8} \binom{10}{x} 0.7^x 0.3^{10-x}$

Binomial Formula Examples

1. A die is rolled 10 times. What is the chance of getting exactly 2 aces (1's)?
2. A die is rolled until it first lands six. If this can be done using the binomial formula, find the chance of getting 2 aces. If not, why not?
3. Ten draws are made with replacement from a box:

   [1] [1] [1] [1] [1] [0]

   Before the last draw, one ticket labeled 1 is removed. True or False:

   $$P(\text{exactly 2 ones}) = \binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8$$

4. Four draws are made **without replacement** from the same box. True or False:

   $$P(\text{exactly 2 ones}) = \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$$

Binomial Formula: Example 1

Example 1.
A die is rolled 10 times. What is the chance of getting exactly 2 aces (1's)?

$$P(X=2) = \binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8$$

Binomial Formula: Example 2

Example 2. A die is rolled until it first lands six. If this can be done using the binomial formula, find the chance of getting 2 aces. If not, why not?

no, n is not fixed

Binomial Formula: Example 3

Example 3.
Ten draws are made with replacement from a box:

[1] [1] [1] [1] [1] [0]

Before the last draw, one ticket labeled 1 is removed.

True or False:

$$P(\text{exactly 2 ones}) = \binom{10}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^8$$

Binomial Formula: Example 4

Example 4.
Four draws are made **without replacement** from the same box.

True or False:

$$P(\text{exactly 2 ones}) = \binom{4}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^2$$

False

When Can We Use the Binomial Formula?

We can use the binomial model when:

1. Number of trials is fixed
2. Probability is constant
3. Trials are independent
4. Only two outcomes

Hypergeometric Distribution (When?)

Hypergeometric assumptions:

- Population size N is finite.
- Each item is success (S) or failure (F); there are k successes.
- Sample n items **without replacement** (all subsets equally likely).

Hypergeometric pmf

If $X \sim \mathrm{Hypergeom}(N, n, k)$, then

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

for

$$\max\{0,\ n - (N - k)\} \leq x \leq \min\{n,\ k\}.$$

Example (Lot Acceptance)

You ship a lot of $N = 24$ cameras. An inspector samples $n = 4$ and rejects the lot if any defective is found. You know $k = 3$ are defective. Find $P(\text{lot accepted})$.

$$P(\text{accepted}) = P(X=0) = \frac{\binom{3}{0}\binom{21}{4}}{\binom{24}{4}}$$

Mean and Variance (Hypergeometric)

For $X \sim \mathrm{Hypergeom}(N, n, k)$,

$$\mu = E[X] = n\frac{k}{N},$$

$$\sigma^2 = \mathrm{Var}(X) = \frac{N-n}{N-1}\, n\frac{k}{N}\left(1 - \frac{k}{N}\right).$$

What is the mean and variance of the previous example?

$$E[X] = n\frac{K}{N} = 4\frac{3}{24} = \frac{1}{2}$$

$$Var(X) = \frac{N-n}{N-1} n \frac{k}{N}\left(1-\frac{k}{N}\right) = \frac{24-4}{24-1} \times 4 \frac{3}{24}\left(1-\frac{3}{24}\right)$$

$$= \frac{35}{92}$$

Hypergeometric vs Binomial: Example

A box contains 4 red balls and 6 blue balls.

(a) Four balls are drawn **with replacement**.

(b) Four balls are drawn **without replacement**.

What is the probability of drawing exactly 2 red balls?

(a) $P(X=2) = \binom{4}{2}(.4)^2(.6)^2$

(b) $P(X=2) = \dfrac{\binom{4}{2}\binom{6}{2}}{\binom{10}{4}}$

Example Solution

Poisson Experiments (Counts)

A Poisson experiment records the number of events in a time interval or a region of space. Examples:

- Calls received per hour
- Days school is closed
- Accidents in a week
- Typing errors on a page

Poisson Process Properties

- Counts in disjoint intervals/regions are independent.
- Probability of one event in a very short interval is proportional to interval length/region size.
- Probability of more than one event in a very short interval is negligible.

Poisson Distribution

If events occur at average rate λ per unit time (or space), then the number of events in interval length t is

$$X \sim \mathrm{Poisson}(\lambda t),$$

with pmf

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

Example (Poisson)

Let X be the number of creatures captured in a week. Suppose that X has a Poisson distribution with $\lambda = 4.5$, so on average traps will contain 4.5 creatures. Suppose $X \sim \text{Poisson}(\lambda t = 4.5)$.

(a) $P(X = 5)$

(b) $P(X \leq 5)$

(c) $P(X \geq 5)$

$$P(X=5) = e^{-4.5}\frac{(4.5)^5}{5!}$$

$$P(X\leq 5) = \sum_{x=0}^{5} e^{-4.5}\frac{(4.5)^x}{x!}$$

$$P(X\geq 5) = 1 - P(X\leq 4)$$

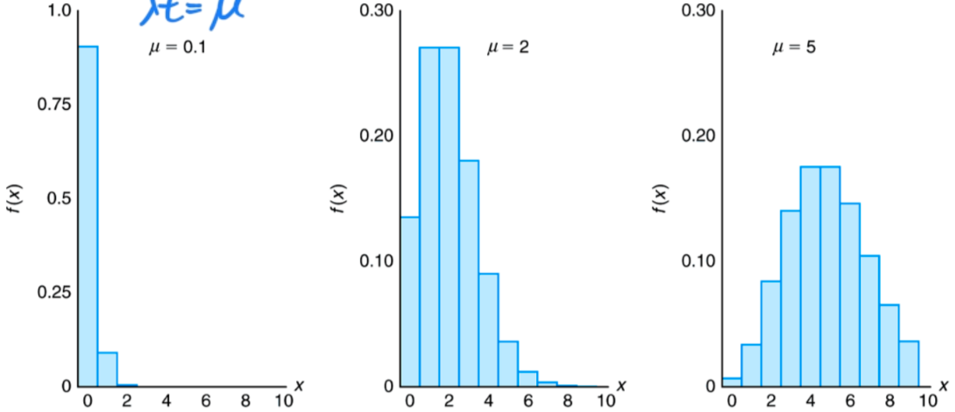
Mean and Variance (Poisson)

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

$$\mu = E[X] = \lambda t, \qquad \sigma^2 = \mathrm{Var}(X) = \lambda t.$$

Nature of the Poisson Probability function

$\lambda t = \mu$



Example (Arrivals at a Counter)

Suppose pulses arrive at a counter at an average rate of 6 per minute, so $\lambda = 6$.

(a) Find the probability that in a 0.5-minute interval at least one pulse is received.

(b) Find the mean number of pulses received in the 30-second interval.

Poisson Example

A factory produces screws, and 0.1% are defective.
A box contains 2000 screws.
What is the probability the box has exactly 3 defective screws?

Poisson Approximation: Example 3 (Solution)

Poisson Example

A store receives customers in bursts: when one arrives, their friends often arrive immediately after.
Is Poisson a good model for the number of arrivals per minute? Why or why not?

Chapter Summary: Discrete Distributions

- **Bernoulli Process:** Repeated independent trials with constant success probability p.
- **Binomial Distribution:**

$$X \sim \mathrm{Bin}(n, p), \quad P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E[X] = np, \qquad \mathrm{Var}(X) = np(1-p)$$

- **Hypergeometric Distribution:** Sampling **without replacement** from finite population.

$$X \sim \mathrm{Hypergeom}(N, n, k)$$

- **Poisson Distribution:** Counts of events over time/space.

$$X \sim \mathrm{Poisson}(\lambda t), \quad E[X] = \mathrm{Var}(X) = \lambda t$$

- **Key Idea:** Choose the distribution based on independence, replacement, and event rate.