Chapter 4

4.1 Probability Distribution Function (PDF) for a Discrete Random Variable

A **discrete random variable** is a variable, X, that has a countable number of possible values. A **Probability Distribution Function (PDF)**, denoted as P(x), is a table, formula, or graph that gives the probability for each value of the random variable.

Key Properties

For any discrete PDF, two conditions must be met:

1. Each probability P(x) must be between 0 and 1, inclusive.

- $0 \leq P(x) \leq 1$

1. The sum of all probabilities must equal 1.

- $\sum P(x) = 1$

Example

Let X = the number of heads in two coin flips.

x (Value)	P(x) (Probability)
0	0.25
1	0.50
2	0.25

4.2 Mean or Expected Value and Standard Deviation

These metrics describe the center and spread of a probability distribution.

Mean or Expected Value (μ)

The **mean** or **expected value** is the long-term average value of a random variable. It's the weighted average of the possible values.

- **Formula:**

  $$\mu = E(X) = \sum x \cdot P(x)$$

  $$$$

Standard Deviation (σ)

The **standard deviation** measures the typical distance of the values from the mean. A smaller value indicates less variability.

- **Variance Formula (σ^2):**

  $$\sigma^2 = \sum (x - \mu)^2 \cdot P(x)$$

  $$$$

- **Standard Deviation Formula (σ):**

  $$\sigma = \sqrt{\sigma^2}$$

  $$$$

4.3 Binomial Distribution

Use the binomial distribution for experiments with a fixed number of independent trials, where each trial has only two possible outcomes.

Conditions

1. There is a fixed number of trials, n.
2. Each trial is independent.
3. There are only two outcomes: **success** (with probability p) and **failure** (with probability $q = 1 - p$).
4. The probability of success, p, is the same for each trial.

Formulas & Notation

- **Notation:** $X \sim B(n, p)$
- **PDF Formula:** Find the probability of exactly k successes in n trials.

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

- **Mean:** $\mu = np$
- **Standard Deviation:** $\sigma = \sqrt{np(1-p)}$

R Code

```
# P(X = k) → Probability of exactly k successes
dbinom(k, size=n, prob=p)
# P(X <= k) → Cumulative probability of k or fewer successes
pbinom(k, size=n, prob=p)
# Generate random values from a binomial distribution
rbinom(num_simulations, size=n, prob=p)
```

4.4 Geometric Distribution

Use the geometric distribution to model the number of trials needed to get the **first success** in a series of independent Bernoulli trials.

Conditions

1. Trials are independent.
2. There are two outcomes: success (p) and failure $(q = 1-p)$.
3. The probability of success, p , is the same for each trial.
4. The variable of interest, X, is the number of the trial on which the first success occurs.

Formulas & Notation

- **Notation:** $X \sim G(p)$
- **PDF Formula:** Find the probability that the first success occurs on trial k.
  $P(X=k)=(1-p)^{k-1}p$

- **Mean:** $\mu = \dfrac{1}{p}$
- **Standard Deviation:** $\sigma = \sqrt{\dfrac{1-2}{p^2}}$

R Code

Note: R's geometric functions count the number of *failures* before the first success. So, to find $P(X = k)$, you look for $k - 1$ failures.

```
# P(X = k) → First success on trial k
dgeom(k-1, prob=p)
# P(X <= k) → First success on or before trial k
pgeom(k-1, prob=p)
```

4.5 Hypergeometric Distribution

Use the hypergeometric distribution for experiments where you are sampling **without replacement** from a finite population with two groups (e.g., drawing cards from a deck).

Conditions

1. You are sampling from two groups.
2. You are concerned with a group of interest (successes).
3. You sample without replacement.
4. Each pick is not independent.

Formulas & Notation

- **Parameters:**
- N = Total population size.
- K = Number of successes in the population.
- n = Sample size.

- **Notation:** $X \sim H(N, K, n)$
- **PDF Formula:** Find the probability of getting exactly k successes in a sample of size n.

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

- **Mean:** $\mu = n\frac{K}{N}$

R Code

Note: In R, the arguments are m (successes in pop, K), n (failures in pop, N−K), and k (sample size, n).

```
# P(X = k) → Probability of exactly k successes

dhyper(k, m=K, n=N-K, k=n)

# P(X <= k) → Cumulative probability of k or fewer successes

phyper(k, m=K, n=N-K, k=n)
```

4.6 Poisson Distribution

Use the Poisson distribution to model the number of events occurring in a **fixed interval of time or space** when the events happen with a known average rate and independently of the time since the last event.

Conditions

1. Events occur at a known mean rate, λ.
2. Events are independent.
3. The random variable X is the number of occurrences in the interval.

Formulas & Notation

- **Notation:** $X \sim P(\lambda)$
- **PDF Formula:** Find the probability of exactly k events in an interval.

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

- **Mean:** $\mu = \lambda$
- **Standard Deviation:** $\sigma = \sqrt{\lambda}$

R Code

```
# P(X = k) → Probability of exactly k events
dpois(k, lambda=λ)
# P(X <= k) → Cumulative probability of k or fewer events
ppois(k, lambda=λ)
```