Week 4: Random Variables and Distributions

Part 1: Discrete Random Variables

Random Variables

A **random variable** is a real-valued variable whose value is determined by an underlying random experiment.

Examples:

- Flip two coins, X = number of heads
- Roll a die, X = number shown
- Soccer game, X = number of goals scored

Range of a Random Variable: The set of possible values the random variable can take.

Exercises:

1. I toss a coin 100 times. Let X be the number of heads I observe. What's the range?
   - **Answer:** $[0, 100]$
2. I toss a coin until the first heads appears. Let Y be the total number of coin tosses. What's the range?
   - **Answer:** $[1, \infty)$

Discrete Random Variables

A random variable X is **discrete** if its range is countable (finite or countably infinite).

Example: $\{0, 1, 2, 3\}, \{1, 2, 3, \ldots\}$

Probability Mass Function (PMF)

Let X be a discrete random variable with range $R_X = \{x_1, x_2, x_3, \ldots\}$

The **PMF** is the function $P(X = x_k)$ for $k = 1, 2, 3, \ldots$

The PMF tells us the probability that X takes each specific value in its range.

Example: I toss a fair coin twice, X = number of heads. Find the range and PMF.

Solution:

- Range: $\{0, 1, 2\}$
- $P(X = 0) = 1/4$ (TT)
- $P(X = 1) = 1/2$ (HT or TH)
- $P(X = 2) = 1/4$ (HH)

Properties of PMF:

1. $0 \leq P(X = x) \leq 1$ for all x
2. $\sum P(X = x) = 1$ (sum over all possible values)

Independence of Random Variables

Two random variables X and Y are **independent** if:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \text{ for all } x, y$$

This extends to n random variables:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots \cdot P(X_n = x_n)$$

Example: Two coin flips are independent. Knowing the first flip doesn't change the probability of the second flip.

Special Discrete Distributions

1. Bernoulli Distribution

The simplest random variable: only two possible outcomes.

Definition: $X \sim \text{Bernoulli}(p)$

PMF:

$$P_X(x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$

Interpretation:

- $X = 1$ with probability p (success)
- $X = 0$ with probability $1 - p$ (failure)

Examples:

- Single coin flip: $X = 1$ if heads, $X = 0$ if tails, $p = 1/2$
- Pass/fail exam: $X = 1$ if pass, $X = 0$ if fail

---

2. Geometric Distribution

Question: How many trials until the first success?

Setup: I have a coin with $P(H) = p$. I toss until I observe the first heads. X = total number of tosses.

Definition: $X \sim \text{Geometric}(p)$

PMF:

$$P_X(k) = \begin{cases} p(1-p)^{k-1} & \text{for } k = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$

Intuition: We need $(k-1)$ failures, then one success.

Examples:

- $P(X = 1) = p$ (success on first trial)
- $P(X = 2) = (1-p) \cdot p$ (fail once, then succeed)
- $P(X = 3) = (1-p)^2 \cdot p$ (fail twice, then succeed)

Think of this as: repeating independent Bernoulli trials until observing the first success.

---

3. Binomial Distribution

Question: In n trials, how many successes?

Setup: I have a coin with $P(H) = p$. I toss n times. X = total number of heads.

Definition: $X \sim \text{Binomial}(n, p)$

PMF:

$$P_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and $0 < p < 1$

Derivation intuition:

- p^k: probability of k successes
- $(1-p)^{n-k}$: probability of $(n-k)$ failures
- $\binom{n}{k}$: number of ways to arrange k successes in n trials

Important: Binomial = sum of n independent Bernoulli random variables

Examples:

- 10 coin flips, count heads: Binomial$(10, 0.5)$
- 20 patients, count responses to treatment: Binomial$(20, p)$

Cumulative Distribution Function (CDF)

The **CDF** of a random variable X is:

$$F_X(x) = P(X \leq x), \quad \text{for all } x \in \mathbb{R}$$

The CDF works for ANY random variable (discrete, continuous, or mixed).

For discrete random variables:

$$F_X(x) = \sum_{x_k \leq x} P_X(x_k)$$

This follows from **mutual exclusivity**: X can only take one value at a time, so we add probabilities.

Example: Toss a coin twice, X = number of heads.

- $P(X = 0) = 1/4$
- $P(X = 1) = 1/2$
- $P(X = 2) = 1/4$

CDF:

- $F_X(x) = 0$, for $x < 0$
- $F_X(x) = 1/4$, for $0 \leq x < 1$
- $F_X(x) = 3/4$, for $1 \leq x < 2$
- $F_X(x) = 1$, for $x \geq 2$

Properties of CDF:

1. $F_X(x)$ is non-decreasing: if $y \geq x$, then $F_X(y) \geq F_X(x)$
2. $\lim_{x \to -\infty} F_X(x) = 0$
3. $\lim_{x \to \infty} F_X(x) = 1$
4. For discrete RVs, the CDF has "jumps" at values where $P(X = x) > 0$

Useful property:

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

Expectation

The **expected value** (or mean) of X is the long-run average value.

Definition: For discrete X with range $R_X = \{x_1, x_2, x_3, \ldots\}$

$$E[X] = \sum_{x_k \in R_X} x_k \cdot P(X = x_k)$$

Think of this as a **weighted average**: each value is weighted by its probability.

Examples

Example 1: $X \sim \text{Bernoulli}(p)$. Find $E[X]$.

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

Example 2: $X \sim \text{Geometric}(p)$. Find $E[X]$.

$$E[X] = \sum_{k=1}^{\infty} k \cdot p \cdot (1 - p)^{k-1}$$

Using geometric series calculus: $E[X] = \frac{1}{p}$

Intuition: If p is small (success is rare), we need many trials on average ($1/p$ is large). If p is large, we need few trials.

Properties of Expectation

1. Linearity of Expectation:

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

This is **always true**, even if the random variables are dependent!

2. Scaling:

$$E[aX + b] = a \cdot E[X] + b$$

Intuition:

- Scaling X by a scales the mean by a
- Shifting X by b shifts the mean by b

Example 3: $X \sim \text{Binomial}(n, p)$. Find $E[X]$.

Key insight: $X = X_1 + X_2 + \cdots + X_n$ where each $X_i \sim \text{Bernoulli}(p)$

By linearity:

$$E[X] = E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] = p + p + \cdots + p = n \cdot p$$

Variance

The **variance** measures spread or variability around the mean.

Let $\mu_X = E[X]$

Definition:

$$\text{Var}(X) = E[(X - \mu_X)^2]$$

Computational formula (easier to use):

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Standard deviation: $\text{SD}(X) = \sqrt{\text{Var}(X)}$

Why square the deviations? So positive and negative deviations don't cancel out.

Properties of Variance

1. Scaling:

$$\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$$

Note:

- Adding a constant b doesn't change variance (just shifts the distribution)
- Multiplying by a scales the variance by a^2

2. Variance of sums (independent case):

If $X_1, X_2, \ldots, X_n$ are **independent**:

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n)$$

Warning: This only works for independent random variables!

Examples

Example 1: $X \sim \text{Bernoulli}(p)$. Find $\text{Var}(X)$.

- $E[X] = p$
- $E[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$
- $\text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$

Example 2: $X \sim \text{Binomial}(n, p)$. Find $\text{Var}(X)$.

Key insight: $X = X_1 + X_2 + \cdots + X_n$ where each $X_i \sim \text{Bernoulli}(p)$ independently.

By independence:

$$\text{Var}(X) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n) = p(1-p) + p(1-p) + \cdots + p(1-p) = n \cdot p \cdot (1-p)$$

Part 2: Continuous Random Variables

Key Differences from Discrete

For continuous random variables:

- Range is uncountable (e.g., all real numbers in an interval)
- $P(X = x) = 0$ **for any specific value** x
- We can only talk about probability of intervals: $P(a < X < b)$

Why is $P(X = x) = 0$?

Asking for the probability that someone's height is exactly 170 cm (not 169.9999 or 170.0001, but exactly 170.00000...) doesn't make sense. There are infinitely many possible values.

BUT: We can ask: $P(169.9 < X < 170.1)$ — this has a meaningful probability!

Probability Density Function (PDF)

Since PMF doesn't work, we use the **PDF** instead.

The PDF $f(x)$ is the **probability density** at point x.

Key idea: Probability = area under the PDF curve

$$P(a < X < b) = \int_a^b f(x)\,dx$$

Properties of PDF:

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (total area under curve is 1)

Important: $f(x)$ can be greater than 1! It's a density, not a probability.

Cumulative Distribution Function (Continuous)

Definition: Same as for discrete!

$$F_X(x) = P(X \leq x)$$

But for continuous:

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt$$

Relationship between PDF and CDF:

$$f(x) = \frac{dF_X(x)}{dx} = F'_X(x)$$

(The PDF is the derivative of the CDF)

Conversely:

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt$$

(The CDF is the integral of the PDF)

Key difference from discrete: The CDF is a **continuous function** (no jumps).

Expectation and Variance (Continuous)

Same concepts, different formulas:

Replace summation $(\sum)$ with integration $(\int)$

Replace PMF with PDF

Expectation:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

Variance:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x)\,dx = E[X^2] - (E[X])^2$$

Properties: Same as discrete!

- $E[aX + b] = a \cdot E[X] + b$
- $\text{Var}(aX + b) = a^2 \cdot \text{Var}(X)$

Special Continuous Distributions

1. Uniform Distribution

The simplest continuous distribution: All values equally likely.

Definition: $X \sim \text{Uniform}(a, b)$

PDF:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

CDF:

$$F_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x > b \end{cases}$$

Expected Value:

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

Variance:

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Intuition: The PDF is constant (flat) over $[a, b]$. It's a rectangle with height $\frac{1}{b-a}$ and width $(b - a)$, so area = 1. The mean is the midpoint of the interval.

Example: Pick a random number between 0 and 1 → Uniform$(0, 1)$

---

2. Normal Distribution

The most important distribution in statistics!

Why? Many natural phenomena follow normal distributions, and it appears in statistical inference (more on this next week).

Definition: $X \sim N(\mu, \sigma^2)$

PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad \text{for all } x \in \mathbb{R}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$

Parameters:

- μ = mean (center of distribution)
- σ^2 = variance (spread of distribution)
- σ = standard deviation

Expected Value:

$$E[X] = \mu$$

Variance:

$$\text{Var}(X) = \sigma^2$$

Sufficient statistics: μ and σ completely determine the shape.

Properties of Normal Distribution

1. Symmetry

- Bell-shaped and symmetric around the mean
- Mean = Median = Mode = μ

2. The tails never touch the x-axis

- Although very unlikely, extreme values are always possible
- "ANYTHING IS POSSIBLE!" (probability is never exactly 0)

3. Standard deviation marks inflection points

- The curve changes from "bending down" to "bending up" at $\mu \pm \sigma$

4. The 68-95-99.7 Rule

- Approximately 68% of data within $\mu \pm 1\sigma$
- Approximately 95% of data within $\mu \pm 2\sigma$
- Approximately 99.7% of data within $\mu \pm 3\sigma$

Standard Normal Distribution

Definition: $Z \sim N(0, 1)$

A normal distribution with mean 0 and standard deviation 1.

PDF:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

Why is this special?

Any normal distribution can be converted to the standard normal!

---

z-scores (Standardization)

Definition: For any value x from $N(\mu, \sigma)$, the z-score is:

$$z = \frac{x - \mu}{\sigma}$$

Interpretation: z is the number of standard deviations x is from the mean.

Examples:

- $z = 0$: x is at the mean
- $z = 1$: x is 1 standard deviation above the mean
- $z = -2$: x is 2 standard deviations below the mean

Why use z-scores?

1. Compare values from different distributions
2. Calculate probabilities using standard normal tables
3. Understand how unusual a value is

Example: Test Scores

Test scores: $N(100, 15^2)$

- $\mu = 100$ (mean score)
- $\sigma = 15$ (standard deviation)

Question: A student scores 115. What percentile are they in?

Step 1: Calculate z-score

$$z = \frac{115 - 100}{15} = \frac{15}{15} = 1$$

Step 2: Find the cumulative probability

$$P(Z \leq 1) \approx 0.84 \text{ (or use R: pnorm(1))}$$

Answer: 84th percentile (better than 84% of students)

The Φ Function (CDF of Standard Normal)

Notation: $\Phi(x) = P(Z \leq x)$ for $Z \sim N(0, 1)$

Properties:

1. $\lim_{x\to\infty} \Phi(x) = 1$
2. $\lim_{x\to-\infty} \Phi(x) = 0$
3. $\Phi(0) = 0.5$ (by symmetry)
4. $\Phi(-x) = 1 - \Phi(x)$ (by symmetry)

In practice: Use R functions!

- `pnorm(x)` gives $\Phi(x)$
- `pnorm(x, mean, sd)` for non-standard normal

Key Takeaways

Discrete vs. Continuous

Feature	**Discrete**	**Continuous**
Range	Countable	Uncountable (interval)
Probability function	PMF: $P(X = x)$	PDF: $f(x)$
$P(X = x)$	Can be > 0	Always $= 0$
Probabilities	Direct from PMF	Area under PDF
Summation vs. Integration	$\sum$	$\int$
CDF has jumps?	Yes	No

Important Distributions

Discrete:

- Bernoulli(p): Single trial
- Geometric(p): Trials until first success
- Binomial(n, p): Number of successes in n trials

Continuous:

- Uniform(a, b): All values equally likely
- Normal(μ, σ^2): The bell curve

Expectation and Variance

Always true:

- $E[aX + b] = a \cdot E[X] + b$ (linearity)
- $E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$ (linearity)
- $\mathrm{Var}(aX + b) = a^2 \cdot \mathrm{Var}(X)$ (scaling)

Only for independent variables:

- $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$