Chapter 3: Random Variables and Their Distributions

Introduction to Probability, 2nd Edition
Blitzstein & Hwang

DRME
December 22, 2025

Chapter Overview

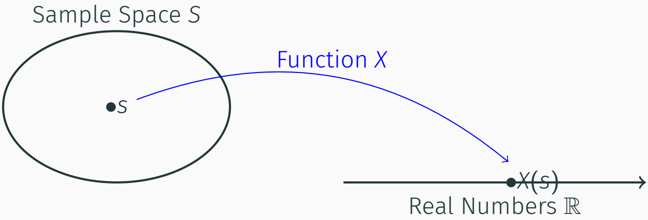
Key Idea

A **random variable** (RV) is a function that assigns a numerical value to each outcome of a random experiment.

Why Random Variables?

- Provide a **numerical language** for probability
- Allow us to use **mathematical tools** (calculus, algebra)
- Connect probability theory with **statistics**

Visualizing the Mapping



Wikipedia: Random Variable

Definitions and Types of Random Variables

Definition (Random Variable)

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A **random variable** X is a function $X : \Omega \to \mathbb{R}$ such that for every $x \in \mathbb{R}$:

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

Technical Condition: This ensures $P(X \leq x)$ is defined for all x.

Notation:

- Capital letters: X, Y, Z for random variables
- Lowercase letters: x, y, z for possible values
- Events: $\{X = x\}$, $\{X \leq x\}$, $\{a < X \leq b\}$

Types of Random Variables

Random variables are classified by their **range**:

Discrete Random Variables

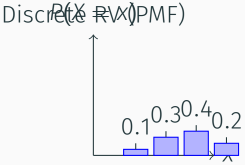
- Takes **countable** values
- Can list all possible values
- Examples:
  - Number of heads in 10 coin tosses
  - Number of customers arriving in an hour
  - Roll of a die (1, 2, 3, 4, 5, 6)

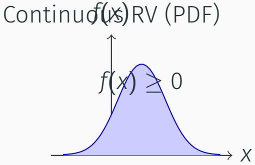
Continuous Random Variables

- Takes **uncountable** values
- Can take any value in an interval
- Examples:
  - Height of a randomly selected person
  - Time until next earthquake
  - Temperature in degrees Celsius

Mixed RVs: Some RVs are neither purely discrete nor purely continuous.

Visual Comparison





Key Difference: Discrete uses probability masses at points, continuous uses density over intervals.

Discrete Random Variables

Probability Mass Function (PMF)

Definition (PMF)

For a discrete random variable X, the **probability mass function** p_X is:

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

Properties of a Valid PMF:

1. **Nonnegativity**: $p_X(x) \geq 0$ for all x
2. **Total Probability**: $\sum_x p_X(x) = 1$ (sum over all possible values)
3. $P(X \in A) = \sum_{x \in A} p_X(x)$ for any set A

Example: Fair die roll

$$p_X(x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

Wikipedia: PMF

Example: Sum of Two Dice

Problem: Roll two fair dice. Let $X =$ sum of the two numbers.

Step 1: Sample space

$$S = \{(i, j) : 1 \leq i, j \leq 6\}, \quad |S| = 36$$

Step 2: Possible values of X

$$X \in \{2, 3, 4, \ldots, 12\}$$

Step 3: Count outcomes for each sum:

Sum x	2	3	4	5	6	7	8	9	10	11	12
Ways	1	2	3	4	5	6	5	4	3	2	1

Definition (CDF)

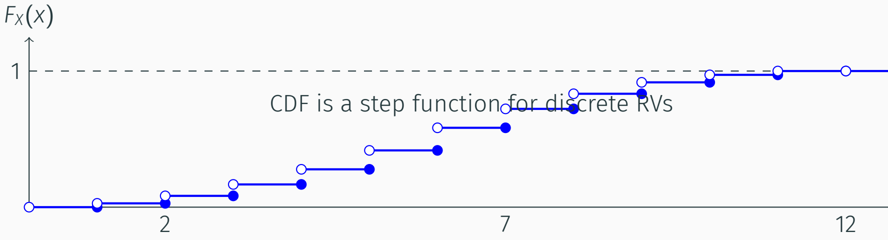
For any random variable X, the **cumulative distribution function** F_X is:

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

Properties of CDF:

1. $0 \leq F_X(x) \leq 1$ for all x
2. F_X is **non-decreasing**: $x < y \Rightarrow F_X(x) \leq F_X(y)$
3. $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$
4. F_X is **right-continuous**: $\lim_{h \to 0^+} F_X(x + h) = F_X(x)$

For discrete RVs: $F_X(x) = \sum_{t \leq x} p_X(t)$



Important: CDF jumps at each possible value; jump size = probability at that value.

Definition (Bernoulli)

A random variable X has a **Bernoulli** distribution with parameter p if:

$$P(X=1)=p, \quad P(X=0)=1-p, \quad 0 \leq p \leq 1$$

Notation: $X \sim \mathrm{Bern}(p)$

Interpretation:

- $X=1$: "Success" (with probability p)
- $X=0$: "Failure" (with probability $1-p$)

PMF:

$$p_X(x)=\begin{cases} p & \text{if } x=1 \\ 1-p & \text{if } x=0 \\ 0 & \text{otherwise} \end{cases} = p^x(1-p)^{1-x} \quad \text{for } x \in \{0,1\}$$

Wikipedia: Bernoulli Distribution

Definition (Binomial)

X counts the number of successes in n independent Bernoulli(p) trials:

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

Notation: $X \sim \text{Bin}(n, p)$

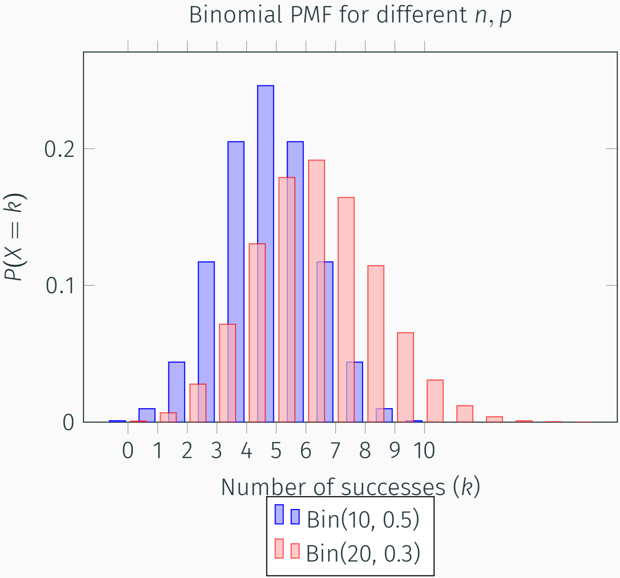
Derivation (Proof of PMF):

- Each trial: success (prob p) or failure (prob $1 - p$)
- Need exactly k successes in n trials
- Choose which k trials succeed: $\binom{n}{k}$ ways
- Probability of each such sequence: $p^k(1-p)^{n-k}$

Example: Number of heads in 10 coin flips: $X \sim \text{Bin}(10, 0.5)$

Wikipedia: Binomial Distribution

Visualizing Binomial Distribution



Observations:

Definition (Poisson)

X counts the number of events occurring in a fixed interval of time/space if:

- Events occur independently
- Average rate λ is constant

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

Notation: $X \sim \text{Pois}(\lambda)$

Applications:

- Number of phone calls received per hour
- Number of radioactive decays per second
- Number of typos per page

Relation to Binomial: Poisson approximates Binomial when n is large, p is small, $\lambda = np$.

Wikipedia: Poisson Distribution

Continuous Random Variables

Probability Density Function (PDF)

Definition (PDF)

For a continuous random variable X, the **probability density function** f_X satisfies:

1. $f_X(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$
3. $P(a \leq X \leq b) = \int_a^b f_X(x)dx$

Important Notes:

- $f_X(x)$ is **NOT** a probability
- $P(X = x) = 0$ for any specific x
- Only intervals have positive probability
- $f_X(x)$ can be > 1 (but area under curve = 1)

Wikipedia: PDF

For continuous RVs, PDF is the derivative of CDF:

Theorem

If X is continuous with CDF F_X and PDF f_X, then:

$$f_X(x) = \frac{d}{dx} F_X(x)$$

at points where F_X is differentiable.

Proof: By definition of derivative and fundamental theorem of calculus:

$$f_X(x) = \lim_{h \to 0} \frac{F_X(x+h) - F_X(x)}{h} = \lim_{h \to 0} \frac{P(x < X \leq x+h)}{h}$$

Conversely: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$

Definition (Uniform)

X is **uniform** on $[a, b]$ if it's equally likely to be anywhere in $[a, b]$:

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

Notation: $X \sim \text{Unif}(a, b)$

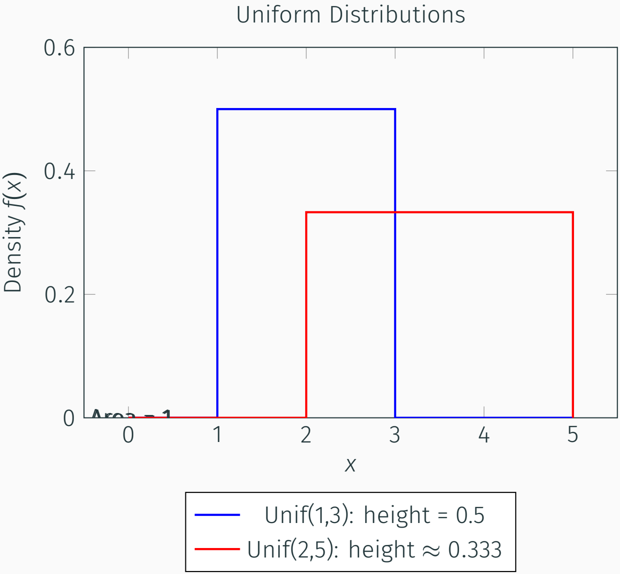
CDF:

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

Applications: Random number generation, round-off errors

Wikipedia: Uniform Distribution

Visualizing Uniform Distribution



Key Property: Density is constant over $[a, b]$, zero elsewhere.

Exponential Distribution

Definition (Exponential)

Models time until next event in a Poisson process with rate λ:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Notation: $X \sim \text{Exp}(\lambda)$

CDF:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Applications:

- Time between phone calls
- Lifetime of radioactive atoms
- Time until failure of electronic components

Wikipedia: Exponential Distribution

Theorem (Memoryless Property)

If $X \sim Exp(\lambda)$, then for all $s, t \geq 0$:

$$P(X > s + t \mid X > s) = P(X > t)$$

Proof:

$$\begin{aligned} P(X > s + t \mid X > s) &= \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t) \end{aligned}$$

Interpretation: "The future is independent of the past"

- If a lightbulb has lasted 100 hours, probability it lasts another 50 hours is same as new bulb lasting 50 hours
- Only continuous distribution with this property

Definition (Normal)

The **normal** distribution with mean μ and variance σ^2:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

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

Special Case: Standard Normal $Z \sim N(0, 1)$:

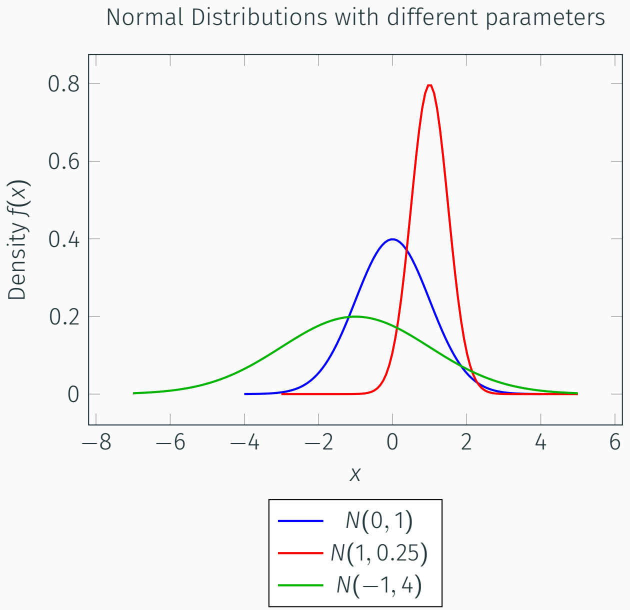
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = P(Z \leq z)$$

Why important?

- Central Limit Theorem: Sums of RVs approach normal
- Models measurement errors, biological traits, etc.

Wikipedia: Normal Distribution

Visualizing Normal Distribution



Expectation and Variance

Expected Value (Mean)

Definition (Expectation)

The **expected value** or **mean** of a random variable X is:

$$E[X] = \begin{cases} \sum_x x \cdot p_X(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{(continuous)} \end{cases}$$

Interpretation:

- Long-run average if experiment repeated many times
- ”Center of mass” of the distribution
- Denoted by μ: $\mu = E[X]$

Example: Fair die roll

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$$

Theorem (Linearity of Expectation)

For any random variables X, Y and constants a, b:

$$E[aX + bY] = aE[X] + bE[Y]$$

Important: This holds even if X and Y are dependent!

Other properties:

1. $E[c] = c$ for constant c
2. $E[X + Y] = E[X] + E[Y]$
3. If $X \geq 0$, then $E[X] \geq 0$
4. If $X \leq Y$, then $E[X] \leq E[Y]$

Caution: $E[XY] \neq E[X]E[Y]$ in general (unless independent).

Law of the Unconscious Statistician (LOTUS)

LOTUS allows calculating $E[g(X)]$ without finding the distribution of $g(X)$ first.

Theorem (LOTUS)

For any function g:

$$E[g(X)] = \begin{cases} \sum_x g(x)p_X(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx & \text{(continuous)} \end{cases}$$

Example: If $X \sim \text{Unif}(0, 1)$, then

$$E[X^2] = \int_0^1 x^2 \cdot 1dx = \frac{1}{3}$$

Without LOTUS, we would need to find PDF of $Y = X^2$ first.

Example: Expected Number of Fixed Points

Problem: n people put their hats in a box, hats randomly redistributed. Let $X =$ number of people who get their own hat back. Find $E[X]$.

Step 1: Define indicator variables

$$I_i = \begin{cases} 1 & \text{if person } i \text{ gets own hat} \\ 0 & \text{otherwise} \end{cases}$$

Then $X = I_1 + I_2 + \cdots + I_n$

Step 2: Find $E[I_i]$

$$E[I_i] = 1 \cdot P(I_i = 1) + 0 \cdot P(I_i = 0) = P(\text{person } i \text{ gets own hat}) = \frac{1}{n}$$

Step 3: Use linearity

$$E[X] = E\left[\sum_{i=1}^{n} I_i\right] = \sum_{i=1}^{n} E[I_i] = \sum_{i=1}^{n} \frac{1}{n} = 1$$

Surprising result: Expected number of fixed points is 1, regardless of n!

Variance and Standard Deviation

Definition (Variance)

The **variance** of X measures spread around the mean:

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

where $\mu = E[X]$

Definition (Standard Deviation)

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$

Properties:

- $\mathrm{Var}(X) \geq 0$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X, Y independent

Variance of Common Distributions

Bernoulli(p):

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

Binomial(n, p):

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

Poisson(λ):

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

Uniform(a, b):

$$E[X] = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Exponential(λ):

$$E[X] = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Normal(μ, σ^2):

$$E[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2$$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)

For any random variable X with finite mean μ and variance σ^2:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \textit{for any } k > 0$$

Interpretation:

- At most $1/k^2$ of probability is more than k standard deviations from mean
- For $k = 2$: At most 25% of probability is beyond 2 SDs
- For $k = 3$: At most 11% of probability is beyond 3 SDs

Example: If test scores have mean 70, SD 10, then at most 25% of students scored below 50 or above 90.

Wikipedia: Chebyshev's Inequality

Joint Distributions

Definition (Joint PMF)

For discrete RVs X and Y, the **joint PMF** is:

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

Definition (Joint PDF)

For continuous RVs X and Y, the **joint PDF** $f_{X,Y}$ satisfies:

$$P((X, Y) \in A) = \iint_A f_{X,Y}(x, y)dxdy$$

Marginal Distributions:

- Discrete: $p_X(x) = \sum_y p_{X,Y}(x, y)$
- Continuous: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$

Example: Joint Distribution Table

Problem: Roll two dice. Let $X =$ number on first die, $Y =$ sum of both dice.

Joint PMF (partial table):

X/Y	2	3	4	5	6	7	etc	Total
1	1/36	1/36	1/36	1/36	1/36	1/36	$\cdots$	1/6
2	0	1/36	1/36	1/36	1/36	1/36	$\cdots$	1/6
3	0	0	1/36	1/36	1/36	1/36	$\cdots$	1/6
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Total	1/36	2/36	3/36	4/36	5/36	6/36	$\cdots$	1

Marginal of X: Sum each row = 1/6 (uniform on 1..6)

Marginal of Y: Sum each column (gives our earlier PMF for sum of dice)

Independence of Random Variables

Definition (Independence)

Random variables X and Y are **independent** if:

- Discrete: $p_{X,Y}(x, y) = p_X(x)p_Y(y)$ for all x, y
- Continuous: $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all x, y

Equivalent conditions:

1. $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ for all x, y
2. $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for all sets A, B
3. $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for all functions g, h

Consequences of independence:

- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- $E[XY] = E[X]E[Y]$

Example: Checking Independence

Problem: From our dice example, are X (first die) and Y (sum) independent?

Check: Is $p_{X,Y}(1,2) = p_X(1)p_Y(2)$?

$$\begin{aligned} p_{X,Y}(1,2) &= P(X=1, Y=2) = P(\text{first die=1, sum=2}) \\ &= P((1,1)) = 1/36 \\ p_X(1) &= 1/6 \\ p_Y(2) &= 1/36 \\ p_X(1)p_Y(2) &= (1/6)(1/36) = 1/216 \neq 1/36 \end{aligned}$$

Conclusion: X and Y are **dependent**.

Covariance

Definition (Covariance)

The **covariance** of X and Y measures their linear relationship:

$$\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$$

Properties:

- $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X, Y + Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$
- If X, Y independent, then $\mathrm{Cov}(X, Y) = 0$ (but converse is false!)

Variance of sum:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$$

Definition (Correlation)

The **correlation coefficient** normalizes covariance:

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Properties:

1. $-1 \leq \rho(X, Y) \leq 1$
2. $\rho(X, Y) = 1$ iff $Y = aX + b$ with $a > 0$
3. $\rho(X, Y) = -1$ iff $Y = aX + b$ with $a < 0$
4. $\rho(X, Y) = 0$: uncorrelated (but not necessarily independent)

Interpretation:

- $\rho > 0$: Positive linear relationship
- $\rho < 0$: Negative linear relationship
- $\rho = 0$: No linear relationship

Wikipedia: Correlation

Transformations of Random Variables

Transformations of RVs: Discrete Case

Problem: Given X with PMF p_X, find distribution of $Y = g(X)$.

Solution: For each y, find all x such that $g(x) = y$:

$$p_Y(y) = P(Y = y) = P(g(X) = y) = \sum_{x:g(x)=y} p_X(x)$$

Example: Let $X \sim \text{Bern}(p)$, $Y = 2X - 1$

$$p_Y(y) = \begin{cases} p & \text{if } y = 1 \quad (x = 1) \\ 1 - p & \text{if } y = -1 \quad (x = 0) \\ 0 & \text{otherwise} \end{cases}$$

Linear transformation: If $Y = aX + b$, then:

$$E[Y] = aE[X] + b, \quad \text{Var}(Y) = a^2\text{Var}(X)$$

Theorem (Change of Variables)

Let X be continuous with PDF f_X, and $Y = g(X)$ where g is strictly monotone and differentiable. Then:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Proof idea: Use CDF method:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \quad \text{(if } g \text{ increasing)} \\ &= F_X(g^{-1}(y)) \end{aligned}$$

Differentiate with chain rule to get PDF.

For non-monotone g: Break into pieces where g is monotone.

The **Uniform distribution** serves as a "universal" starting point for simulation:

1. **Generating RVs**: If $U \sim \text{Unif}(0,1)$, then $X = F^{-1}(U)$ has CDF F
2. **Normalizing RVs**: If X has continuous CDF F, then $F(X) \sim \text{Unif}(0,1)$

Example: To generate $X \sim \text{Exp}(\lambda)$:

- Let $U \sim \text{Unif}(0,1)$
- Set $X = F^{-1}(U) = -\frac{\ln(1-U)}{\lambda}$
- Then $X \sim \text{Exp}(\lambda)$

Wikipedia: Probability Integral Transform

Problem: If $Z \sim N(0,1)$, find distribution of $Y = Z^2$.

Solution using CDF method: For $y \geq 0$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(Z^2 \leq y) \\ &= P(-\sqrt{y} \leq Z \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\Phi(\sqrt{y}) - 1 \end{aligned}$$

Differentiate:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = 2\phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0$$

This is the **Chi-square with 1 degree of freedom**: $Y \sim \chi^2_1$

Problem: Given independent X and Y, find distribution of $Z = X + Y$.

Discrete case:

$$p_Z(z) = \sum_x p_X(x) p_Y(z - x)$$

Continuous case:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

Example: Sum of independent Poissons is Poisson If $X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$, independent, then:

$$X + Y \sim \text{Pois}(\lambda + \mu)$$

Proof: Use convolution or moment generating functions.

Applications and Real-World Examples

Problem: A factory produces lightbulbs. Probability a bulb is defective is 0.01. In a batch of 100 bulbs, what's the probability of at most 2 defective bulbs?

Solution: Let $X =$ number of defective bulbs.

$$X \sim \mathrm{Bin}(n = 100, p = 0.01)$$

Exact binomial calculation:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{100}{0}(0.01)^0(0.99)^{100} + \binom{100}{1}(0.01)^1(0.99)^{99} \\ &\quad + \binom{100}{2}(0.01)^2(0.99)^{98} \\ &\approx 0.3660 + 0.3697 + 0.1849 = 0.9206 \end{aligned}$$

Problem: A factory produces lightbulbs. Probability a bulb is defective is 0.01. In a batch of 100 bulbs, what's the probability of at most 2 defective bulbs?

Solution: Let $X =$ number of defective bulbs.

$$X \sim \text{Bin}(n = 100, p = 0.01)$$

Poisson approximation: $\lambda = np = 1$

$$P(X \leq 2) \approx e^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!}\right) = e^{-1}(1 + 1 + 0.5) \approx 0.9197$$

Good approximation!

Problem: An insurance company insures 10,000 people. Probability of claim in a year is 0.005 per person, average claim amount is \$10,000. What premium should they charge to be 95% sure of covering claims?

Solution: Let $N =$ number of claims $\sim \text{Bin}(10000, 0.005)$ Let $X_i =$ claim amounts, $E[X_i] = 10,000$, $\text{Var}(X_i) = \sigma^2$

Total claims: $S = \sum_{i=1}^{N} X_i$

Compound Poisson approximation: $N \approx \text{Pois}(50)$ since $\lambda = 10000 \times 0.005 = 50$

Using Central Limit Theorem, S is approximately normal. Set premium P such that $P(S \leq P) = 0.95$. Need to consider both frequency (N) and severity (X_i)*distributions*.

Problem: Model daily returns of a stock. Returns often modeled as:

$$R_t = \mu + \sigma Z_t, \quad Z_t \sim N(0,1)$$

where μ is expected daily return, σ is volatility.

Probability of large loss: What's $P(R_t < -0.05)$ (5% loss)?

Assume $\mu = 0.001$, $\sigma = 0.02$ (2% daily volatility):

$$P(R_t < -0.05) = P\left(Z_t < \frac{-0.05 - 0.001}{0.02}\right)$$
$$= P(Z_t < -2.55) = \Phi(-2.55) \approx 0.0054$$

About 0.5% chance of 5% daily loss.

VaR (Value at Risk): 1% VaR = loss level with 1% probability Find x such that $P(R_t < -x) = 0.01$:

$$-x = \mu + \sigma\Phi^{-1}(0.01) = 0.001 + 0.02(-2.33) \approx -0.0456$$

So 1% VaR ≈ 4.56% loss.

Summary and Key Concepts

The Four Fundamental Objects

To master random variables, understand the interplay between:

1. **Distributions**: The blueprint (PMF/PDF/CDF/Story)
2. **Random Variables**: The specific instance (Function)
3. **Events**: Subsets of what can happen (e.g., $X > 5$)
4. **Numbers**: Quantifying uncertainty (Probabilities/Expectations)

Analogy for Understanding Distributions

Think of a Distribution as a Blueprint for a house. The blueprint tells you exactly where the rooms and walls go (the shape of the PDF/PMF). However, the blueprint is not the house itself. The Random Variable is the actual House built from that blueprint. Different houses (Random Variables) can be built from the exact same blueprint (Distribution), but they are distinct objects located on different plots of land (different experiments).

1. **PMF**: $p_X(x) = P(X = x)$
2. **PDF**: $f_X(x)$ satisfies $P(a \le X \le b) = \int_a^b f_X(x)dx$
3. **CDF**: $F_X(x) = P(X \le x)$
4. **Expectation**:
$$E[X] = \begin{cases} \sum_x x p_X(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} x f_X(x)dx & \text{(continuous)} \end{cases}$$
5. **Variance**: $\text{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$
6. **Linearity**: $E[aX + bY] = aE[X] + bE[Y]$
7. **Independence**: X, Y independent $\iff f_{X,Y}(x,y) = f_X(x)f_Y(y)$

8. **LOTUS**: $E[g(X)] = \int g(x) f_X(x) dx$ (continuous case)
9. **Covariance**: $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
10. **Correlation**: $\rho(X, Y) = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$
11. **Sum of variances**: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
12. **Change of variables**: $f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$
13. **Convolution**: $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$
14. **Chebyshev**: $P(|X - \mu| \geq k\sigma) \leq 1/k^2$
15. **Memoryless property**: Exponential: $P(X > s + t | X > s) = P(X > t)$
16. **Universality of Uniform**: $F(X) \sim \text{Unif}(0, 1)$ for continuous X

Common Distributions Summary

Distribution	PMF/PDF	Mean	Variance
Bernoulli(p)	$p^x(1-p)^{1-x}$	p	$p(1-p)$
Binomial(n, p)	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Poisson(λ)	$\frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ
Uniform(a, b)	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential(λ)	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal(μ, σ^2)	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

Problem-Solving Strategies

1. **Identify the type**: Discrete vs Continuous
2. **Choose appropriate distribution**: Based on problem context
3. **Use indicator variables**: For counting problems
4. **Leverage linearity of expectation**: Often easier than finding full distribution
5. **Consider transformations**: CDF method for $Y = g(X)$
6. **Check independence**: Simplifies calculations
7. **Use approximations**: Binomial $\rightarrow$ Poisson, Sums $\rightarrow$ Normal
8. **Simulate**: When analytical solution is difficult

Questions?

Complete R code for simulations available at
`http://stat110.net`