Random Variables and Probability Distributions: A Deep, Intuitive, and Detailed Introduction

Saurabh

March 2, 2025

Contents

1 Why Probability? A Gentle Invitation **3**

2 Random Variables: The Core Concept **4**

2.1 What Is a Random Variable, Really? . . . 4

2.2 Discrete vs. Continuous at a Glance . . . 4

2.3 Why Do We Need Random Variables? . . . 4

3 Discrete Probability Distributions **5**

3.1 Probability Mass Function (PMF) . . . 5

3.2 Key Discrete Distributions in Real Life . . . 5

3.2.1 Bernoulli Distribution . . . 5

3.2.2 Binomial Distribution . . . 5

3.2.3 Poisson Distribution . . . 6

3.3 A Visual Example of a Discrete Distribution . . . 7

4 Practice Problems (Discrete) with Solutions **8**

5 Continuous Probability Distributions **10**

5.1 Probability Density Function (PDF) . . . 10

5.2 Key Continuous Distributions . . . 10

5.2.1 Uniform(a, b) . . . 10

5.2.2 Exponential(λ) . . . 10

5.2.3 Normal(μ, σ^2) . . . 11

5.3 A Continuous PDF in Action . . . 12

6 Practice Problems (Continuous) with Solutions 13

7 Wrapping Up and Further Adventures 15

1 What Are Random Variables?

1.1 Core Idea

A **random variable** is a function that assigns a real number to each outcome of a random process. Before the event happens, we are uncertain about the outcome; after the event, the random variable takes on one specific value.

Example 1: Rolling a Die
If you roll a fair 6-sided die, let X denote the number on the top face. Before rolling, X could be any of 1, 2, 3, 4, 5, or 6. After rolling, you observe one of these numbers.

Example 2: Measuring Height
When you measure a random person's height, let H represent that height (in centimeters). Here, H can be any real number within a reasonable range (e.g., 130 to 220 cm).

1.2 Why Do We Need Random Variables?

- **Numerical Grip on Uncertainty:** They transform unpredictable outcomes (like whether the bus is early or late) into numbers that we can analyze quantitatively.
- **Calculation Power:** Once outcomes are numerical, we can compute probabilities (e.g., $P(X > 3)$), averages, variances, and other statistics that summarize the behavior of the uncertain event.
- **Universality:** Random variables provide a common language for uncertainty across everyday events (coin tosses, dice rolls) and advanced fields (quantum mechanics, finance, machine learning, etc.).

1.3 Discrete vs. Continuous Random Variables

- **Discrete:** Takes on values from a *countable* set (e.g., $\{0, 1, 2, \dots\}$ or $\{1, 2, 3, 4, 5, 6\}$). Typically used when you count occurrences or specific outcomes.
- **Continuous:** Takes on any value in an *interval* (e.g., all real numbers between 130 and 220). Typically used for measurements like time, distance, weight, temperature.

2 Discrete Probability Distributions

When a random variable X is **discrete**, it can take on separate (often integer) values. We describe its distribution with a **probability mass function (PMF)**:

$$P(X = k) = \text{probability that } X \text{ equals } k,$$

and

$$\sum_k P(X = k) = 1.$$

2.1 Core Discrete Distributions and Their Stories

We'll focus on three widely used discrete distributions: **Bernoulli**, **Binomial**, and **Poisson**.

2.1.1 Bernoulli Distribution

- **Definition:** A Bernoulli(p) random variable models a single trial with exactly two outcomes (often called success/failure). Probability of success = p, probability of failure = $1 - p$.

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

- **Examples:**
  - A single coin flip (Heads = 1, Tails = 0).
  - Checking if one lightbulb is defective (defect = 1, working = 0).

2.1.2 Binomial Distribution (Detailed)

- **Definition:** If you repeat the Bernoulli(p) trial n *independent* times, and let X = the number of successes, then $X \sim \text{Binomial}(n, p)$. Its PMF is:

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

- **Real-Life Scenes:**
  - Number of heads in 10 flips of a biased coin.
  - Number of defective products in a sample of 50 if each item is defective with probability p.

Mean and Variance Proof (Step-by-Step and Detailed)

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

Strategy: Break X into n simpler random variables $X_1, X_2, \ldots, X_n$, where:

$$X_i = \begin{cases} 1, & \text{if trial } i \text{ is a success,} \\ 0, & \text{otherwise.} \end{cases}$$

Thus

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

We know each $X_i \sim \text{Bernoulli}(p)$.

Step 1: Compute the Mean

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

By the linearity of expectation (which *always* holds, even without independence):

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

Since each X_i is Bernoulli(p), we have $E[X_i] = p$. Therefore,

$$E[X] = n \times p.$$

Step 2: Compute the Variance

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

Key point: Now we use *independence* (which is crucial for variances to add easily). The variance of a sum of *independent* random variables is the sum of their variances:

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

Each X_i is Bernoulli(p), so

$$\text{Var}(X_i) = p(1 - p).$$

Hence

$$\text{Var}(X) = n \times p(1 - p).$$

Conclusion: For Binomial(n, p),

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

2.1.3 Poisson Distribution (Detailed)

- **Definition:** A Poisson(λ) random variable X models the number of events in a fixed interval, assuming events happen *independently* at a constant average rate λ. The PMF:

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

- **Examples:**
  - Number of calls arriving at a call center per hour if calls come at an average rate λ.
  - Count of rare accidents or occurrences in a large population over a given time.

Mean and Variance Proof (Step-by-Step and Detailed) Let $X \sim \text{Poisson}(\lambda)$.

Mean:

$$E[X] = \sum_{k=0}^{\infty} k \, \frac{\lambda^k e^{-\lambda}}{k!}.$$

We can manipulate the sum in a classical way:

$$\sum_{k=0}^{\infty} k \, \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \quad \text{(reindex } m = k - 1\text{)},$$

which equals λe^{λ}. Multiplying by $e^{-\lambda}$ yields

$$E[X] = \lambda.$$

Variance: A standard approach is to find $E[X^2]$. Alternatively, one can use the moment generating function of a Poisson, but we'll do a direct sum approach:

$$E[X^2] = \sum_{k=0}^{\infty} k^2 \, \frac{\lambda^k e^{-\lambda}}{k!}.$$

You can rewrite $k^2 = k(k-1) + k$. Then you split the sum into two simpler sums, each related to the same exponential expansions used for the mean. Eventually, you get:

$$E[X^2] = \lambda^2 + \lambda.$$

Therefore,

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

Hence for Poisson(λ),

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

2.2 Discrete Example PMF Plot

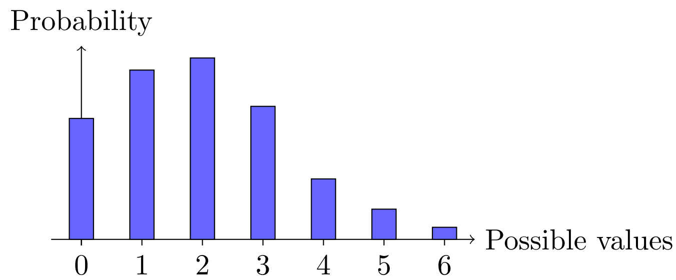

Figure 1: Illustration of a PMF: each bar is the probability mass for that integer outcome. Summation = 1.

3 Practice Problems (Discrete) with Solutions

Note: Some problems are straightforward, others a bit more challenging. Try them sequentially.

Problem 1: Biased Coin (Binomial)

Statement: You have a coin that lands heads with probability $p = 0.3$. Flip it 5 times. Let X = number of heads.

(a) Find $P(X = 2)$.

(b) Probability of at least 1 head?

(c) Compute $E[X]$ and $\mathrm{Var}(X)$.

Solution:

$$X \sim \text{Binomial}(5, 0.3).$$

$$P(X = 2) = \binom{5}{2}(0.3)^2(0.7)^3 \approx 0.3087.$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0.7)^5 \approx 0.83193.$$

$$E[X] = 5 \times 0.3 = 1.5, \quad \mathrm{Var}(X) = 5 \times 0.3 \times 0.7 = 1.05.$$

Problem 2: Poisson Emails

Statement: Emails arrive at rate $\lambda = 4$ per hour, independently. Let Y be the number of emails in 1 hour, so $Y \sim \text{Poisson}(4)$.

(a) Find $P(Y = 2)$.

(b) Probability of at least 1 email?

(c) Mean and variance of Y.

Solution:

$$P(Y = 2) = \frac{4^2 e^{-4}}{2!} = 8e^{-4} \approx 0.1465.$$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - e^{-4} \approx 0.9817.$$

$$E[Y] = \lambda = 4, \quad \mathrm{Var}(Y) = 4.$$

Problem 3: Quality Control (Binomial)

Statement: A factory has a 2% defect rate. You sample 20 items, let X = number of defectives.

(a) Find $P(X = 0)$ and $P(X = 1)$.

(b) If more than 2 defects are found, the entire batch is rejected. Probability of rejection?

(c) Show from the Binomial formulas that $E[X] = 0.4$ and $\mathrm{Var}(X) = 0.392$.

Solution:

$$X \sim \text{Binomial}(20, 0.02).$$

$$P(X = 0) = (0.98)^{20}, \quad P(X = 1) = 20 \times 0.02 \times (0.98)^{19}.$$

Rejection means $X > 2$. So $P(X > 2) = 1 - P(X \leq 2)$.

$$E[X] = 20 \times 0.02 = 0.4, \quad \mathrm{Var}(X) = 20 \times 0.02 \times 0.98 = 0.392.$$

---

4 Continuous Probability Distributions (Detailed)

Now we move to **continuous** random variables, which take values in a continuum (like a line segment or the entire real line).

4.1 The Probability Density Function (PDF)

A continuous random variable X is described by a **probability density function** $f_X(x)$. Probabilities come from integrating this density over intervals:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx,$$

and

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1.$$

No single point has positive probability because an exact point has zero "width" (area).

4.1.1 Analogy: Spreading "1 Unit of Paint or Sand"

Think of painting the x-axis: The *thickness* at each point is $f_X(x)$. The total paint used is 1 liter (for total probability = 1). The probability that X lands in $[a, b]$ is the *amount of paint* over that interval, i.e., the area under the curve from a to b.

4.2 Core Continuous Distributions and Their Stories

We'll discuss **Uniform**, **Exponential**, and **Normal**.

4.2.1 Uniform Distribution (Detailed)

- **Definition:** A Uniform(a, b) random variable says every point in $[a, b]$ is equally likely.

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

- **Real-Life Scenes:**
  - A random time between 9:00 and 9:30 AM if you have no preference for any minute or second.
  - Selecting a random point on a line segment of length $b - a$.
- **Mean and Variance:**

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

- **Interpretation:**

$$P(a \leq X \leq c) = \int_a^c \frac{1}{b-a}\, dx = \frac{c-a}{b-a} \quad \text{for } a \leq c \leq b.$$

4.2.2 Exponential Distribution (Detailed)

- **Definition:** An Exponential(λ) random variable models the waiting time for the *first event* in a Poisson process with rate λ. The PDF:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

- **Real-Life Scenes:**
  - Time until the next phone call in a call center (if calls arrive at average rate λ).
  - Time until a radioactive decay event, given a constant decay rate.

- **Memoryless Property:**

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

  This is unique among continuous distributions (except for some discrete analogs like geometric).

- **Mean and Variance:**

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

4.2.3 Normal (Gaussian) Distribution (Detailed)

- **Definition:** The Normal (or Gaussian) distribution with mean μ and variance σ^2 has PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Bigl(-\frac{(x-\mu)^2}{2\sigma^2}\Bigr), \quad x \in \mathbb{R}.$$

- **Real-Life Scenes:**
  - Heights or weights in a large population, approximated by a bell curve.
  - Measurement errors in experiments, which often cluster around 0 with some spread σ.

- **Central Limit Theorem Connection:** The Normal distribution emerges as the sum/average of many small independent random effects, making it ubiquitous in nature and data.

- **Interpretation:**

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Bigl(-\frac{(x-\mu)^2}{2\sigma^2}\Bigr) dx.$$

  Typically evaluated via standard normal tables or software, since the integral doesn't have a simple closed form in elementary functions.

4.3 Continuous Example PDF Plot

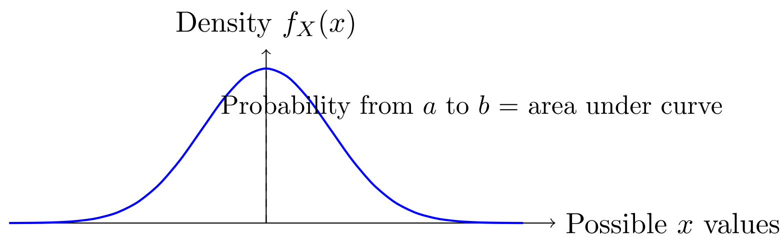

Figure 2: A continuous PDF (approximately a normal curve). The total area = 1. Single points have zero probability.

5 Practice Problems (Continuous) with Solutions

Try these to solidify your understanding of PDFs, integrals, and key continuous distributions.

Problem 1: Uniform(0,10)

Statement: Let $X \sim \text{Uniform}(0, 10)$.

(a) Find $P(2 \le X \le 6)$.

(b) Compute $E[X]$ and $\text{Var}(X)$.

Solution:

$$P(2 \le X \le 6) = \int_2^6 \frac{1}{10}\,dx = \frac{6-2}{10} = 0.4.$$

For a uniform distribution on $[0, 10]$,

$$E[X] = \frac{0+10}{2} = 5, \quad \text{Var}(X) = \frac{(10-0)^2}{12} = \frac{100}{12} = 8.3333.$$

Problem 2: Exponential with $\lambda = 2$

Statement: Let $T \sim \text{Exponential}(2)$.

(a) Find $P(T \le 1)$.

(b) Find $P(T > 1.5)$.

(c) Verify $\text{Var}(T) = 1/\lambda^2$.

Solution:

$$P(T \le 1) = \int_0^1 2e^{-2t}\,dt = 1 - e^{-2} \approx 0.8647.$$

$$P(T > 1.5) = 1 - P(T \le 1.5) = e^{-3} \approx 0.0498.$$

Exponential($\lambda = 2$) has mean $= 1/2$ and variance $= 1/(2^2) = 1/4 = 0.25$.

Problem 3: Piecewise PDF

Statement: A random variable X has PDF:

$$f_X(x) = \begin{cases} \frac{x}{8}, & 0 \le x \le 4, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Show it integrates to 1.

(b) Find $P(1 \le X \le 3)$.

(c) Compute $E[X]$.

Solution:

$$\int_0^4 \frac{x}{8}\,dx = \left[\frac{x^2}{16}\right]_0^4 = \frac{16}{16} = 1.$$

$$P(1 \le X \le 3) = \int_1^3 \frac{x}{8}\,dx = \left[\frac{x^2}{16}\right]_1^3 = \frac{9}{16} - \frac{1}{16} = 0.5.$$

$$E[X] = \int_0^4 x \cdot \frac{x}{8}\,dx = \int_0^4 \frac{x^2}{8}\,dx = \left[\frac{x^3}{24}\right]_0^4 = \frac{64}{24} = \frac{8}{3} \approx 2.6667.$$

Problem 4: Poisson to Exponential (Memoryless)

Statement: Suppose emails arrive by a Poisson process with rate $\lambda = 5$ per hour. Let Z = time until the *first* email.

(a) Argue that $Z \sim \text{Exponential}(5)$.

(b) Find $P(Z > 0.2)$ (time in hours).

(c) Show the memoryless property: $P(Z > s + t \mid Z > s) = P(Z > t)$.

Solution:

(a) In a Poisson(λ) process, the waiting time for the first event is Exponential(λ). So $Z \sim \text{Exponential}(5)$.

(b)

$$P(Z > 0.2) = e^{-5\times 0.2} = e^{-1} \approx 0.3679.$$

(c) Exponential memorylessness:

$$P(Z > s + t \mid Z > s) = \frac{P(Z > s + t)}{P(Z > s)} = \frac{e^{-5(s+t)}}{e^{-5s}} = e^{-5t} = P(Z > t).$$

6 Summary and Next Steps

We've explored:

- **Random Variables**: A unifying concept to represent uncertain outcomes with numbers.
- **Discrete Distributions** (Bernoulli, Binomial, Poisson) and the *detailed* proofs of mean and variance for Binomial/Poisson.
- **Continuous Distributions** (Uniform, Exponential, Normal) with the "paint or sand" analogy for probability density.
- **Practice Problems**: Ranging from straightforward to more conceptual, each with solutions.

Where to go from here?

- **Joint distributions** and **conditional probability** for multiple random variables.
- The **Central Limit Theorem** (why sums/averages of many small independent variables approximate the normal distribution).
- **Statistical Inference**: Hypothesis testing, confidence intervals, regression, etc.
- **Advanced Probability Topics**: Markov chains, Brownian motion, advanced stochastic processes in finance, physics, or data science.

Keep the analogies in mind:

- Discrete = "placing lumps of probability mass into labeled boxes."
- Continuous = "spreading a continuous layer of probability paint across a line."

These mental images help keep the concepts *easy to grasp* and *fun to remember.*

— End of Document —