Probability → Distributions → AI

The Foundation of AI: Understanding Probability

01

Probability Basics

The fundamental language of chance

02

Distributions

Blueprints of randomness

03

Discrete vs Continuous

Counting versus measuring data

04

Famous Distributions

Patterns that appear everywhere

05

AI Applications

How machines learn from probability

Part 1: Probability, the Language of Chance

Probability is simply a number between zero and one. Zero means impossible—it will never happen. One means certain—guaranteed to happen. Every other event falls somewhere between.

Coin Flip

$$P(\text{heads}) = 0.5 = 50\%$$

Die Roll

$$P(\text{rolling 3}) = \frac{1}{6} \approx 0.167$$

Golden Rule

$$\sum_{i=1}^{n} P(x_i) = 1$$

We write probabilities with P. For example, $P(\text{coin} = \text{heads})$ means "the probability that the coin comes up heads."

The golden rule: If you list every possible outcome and add up their probabilities, they must equal exactly one. This guarantees we've accounted for everything.

Part 2: Random Variables

A random variable is a placeholder for a number determined by some random event. It's simpler than it sounds!

Die Roll

$X \in \{1, 2, 3, 4, 5, 6\}$

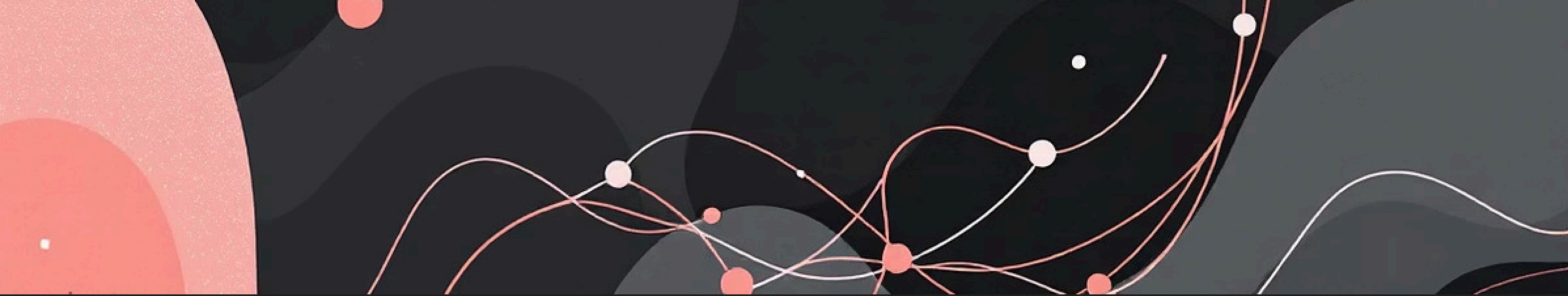
Coin Flip

$$X = \begin{cases} 0 & \text{if tails} \\ 1 & \text{if heads} \end{cases}$$

Height Measurement

$X \in (0, \infty)$ centimeters

Random variables connect probability to real numbers from random processes.



Part 3: Probability Distributions

A probability distribution is the complete map of all possible values of a random variable and their likelihoods. Think of it as the ultimate cheat sheet for randomness.

Discrete Distributions

Probability Mass Function (PMF)

$$P(X = x) = p_x$$

Tells you probability for each specific point

Continuous Distributions

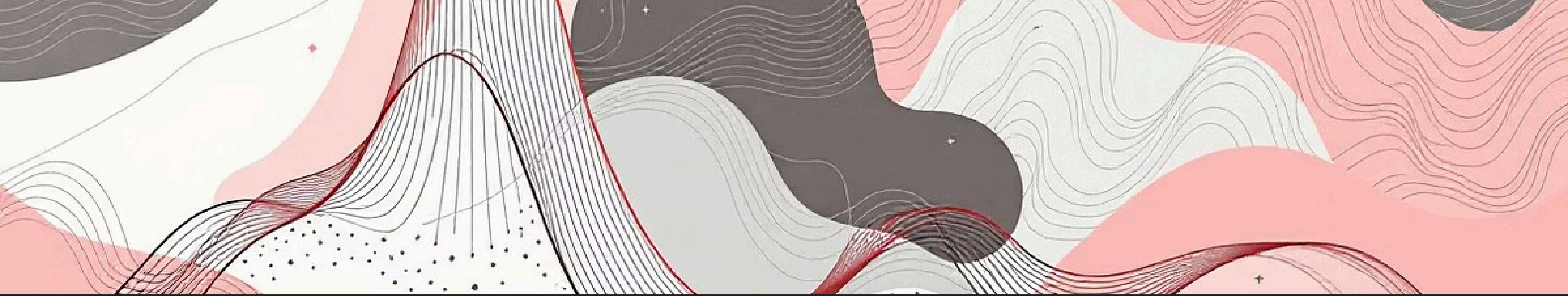
Probability Density Function (PDF)

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

Area under curve gives probability over intervals

In both cases, the golden rule applies: total probability must equal one.

$$\sum_{\text{all } x} P(X = x) = 1 \quad \text{or} \quad \int_{-\infty}^{\infty} f(x)dx = 1$$



Part 4: Famous Distributions in the Real World

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The famous bell curve for heights, test scores, blood pressure. Defined by mean μ and variance σ^2.

68-95-99.7 Rule: 68% within 1σ, 95% within 2σ, 99.7% within 3σ

Binomial Distribution

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

Models repeated yes/no experiments, like flipping a coin 10 times and counting heads.

Poisson Distribution

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

Events in fixed time/space: customers arriving, emails received, server requests.

Uniform Distribution

$$f(x) = \frac{1}{b-a} \text{ for } a \leq x \leq b$$

Every outcome equally likely, like a fair die roll.

Part 5: How AI Uses Probability Distributions

At its core, AI is always estimating probabilities. When you give it data, it builds a model of the underlying distribution.

Classification

$$P(\text{cat}|x) + P(\text{dog}|x) + P(\text{horse}|x) = 1$$

AI estimates probability an image belongs to each class

Regression

$$y = f(x) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

Assumes outputs have Gaussian noise around predictions

Anomaly Detection

$$P(x) < \text{threshold} \rightarrow \text{anomaly}$$

Flags data points far from learned distribution

Generative AI

$$x_{\text{new}} \sim P_{\text{data}}(x)$$

Learns entire data distribution to generate new samples

Loss functions connect directly to distributions: MSE assumes Gaussian noise, cross-entropy uses categorical distributions.

Conclusion: Seeing the World Through Probability

1 **Probability**

Language of chance: numbers between 0 and 1

2 **Random Variables**

Placeholders connecting randomness to numbers

3 **Distributions**

Complete maps of outcomes and likelihoods

4 **AI Foundation**

How machines learn, predict, and generate

> The world around us often looks messy and chaotic. But when you understand probability distributions, you begin to see the hidden shape in that randomness. That's exactly what AI does—it finds the blueprint behind the noise.

Your Turn: What patterns in your life might follow these distributions? Traffic wait times? Daily messages? Game scores? Once you see the world through probability, you see it the way AI does.