Day 5: Probability Distributions — PMF, PDF, and CDF

The Mathematical Framework for Describing Uncertainty Patterns

1. The Fundamental Question

> *How do we systematically describe and quantify the complete behavior of random variables, from individual outcomes to cumulative probabilities?*

Probability distributions provide the complete mathematical description of how random variables behave, enabling precise modeling of uncertainty in AI systems through PMFs, PDFs, and CDFs.

2. Mathematical Foundations

2.1. Probability Mass Function (PMF)

For discrete random variables:

$$p_X(x) = P(X = x), \quad \text{with} \sum_{x \in \mathcal{X}} p_X(x) = 1 \text{ and } 0 \leq p_X(x) \leq 1$$

2.2. Probability Density Function (PDF)

For continuous random variables:

$$f_X(x) \geq 0, \quad \int_{-\infty}^{\infty} f_X(x)dx = 1, \quad P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

2.3. Cumulative Distribution Function (CDF)

Universal definition for all random variables:

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

with properties: $F_X(-\infty) = 0$, $F_X(\infty) = 1$, and $F_X(x)$ is non-decreasing.

Key Relationships

- **Discrete:** $F_X(x) = \sum_{t \leq x} p_X(t)$
- **Continuous:** $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ and $f_X(x) = \frac{d}{dx}F_X(x)$
- **Probability Intervals:** $P(a < X \leq b) = F_X(b) - F_X(a)$
- **Survival Function:** $S_X(x) = 1 - F_X(x) = P(X > x)$

3. Comprehensive Examples and Analysis

3.1. Discrete Distribution: Bernoulli Random Variable

Bernoulli Distribution Specification

- $X \sim \text{Bernoulli}(p)$ with $p = 0.6$
- Support: $\{0, 1\}$ where 1 represents "success"
- Applications: Binary classification, coin flips, A/B testing

Mathematical Analysis:

$$\text{PMF: } p_X(0) = 0.4, \quad p_X(1) = 0.6$$

$$\text{CDF: } F_X(x) = \begin{cases} 0 & x < 0 \\ 0.4 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

$$\text{Expectation: } E[X] = 0 \cdot 0.4 + 1 \cdot 0.6 = 0.6$$

$$\text{Variance: } \text{Var}(X) = p(1-p) = 0.6 \times 0.4 = 0.24$$

3.2. Continuous Distribution: Standard Normal

Normal Distribution Properties

- $Y \sim \mathcal{N}(0, 1)$: Standard normal distribution
- PDF: $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$
- CDF: $\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$
- Applications: Error modeling, natural phenomena, central limit theorem

Probability Calculations:

$$P(-1 \leq Y \leq 1) = \Phi(1) - \Phi(-1) \approx 0.8413 - 0.1587 = 0.6827$$
$$P(-2 \leq Y \leq 2) = \Phi(2) - \Phi(-2) \approx 0.9772 - 0.0228 = 0.9544$$
$$P(-3 \leq Y \leq 3) = \Phi(3) - \Phi(-3) \approx 0.9987 - 0.0013 = 0.9974$$

4. Why Distribution Functions are Fundamental to AI

Critical Roles in Machine Learning

- **Likelihood Calculation:** PDFs/PMFs for maximum likelihood estimation
- **Uncertainty Quantification:** CDFs for confidence intervals and risk assessment
- **Generative Modeling:** Learning data distributions for sample generation
- **Hypothesis Testing:** Distribution functions for statistical significance
- **Bayesian Inference:** Prior and posterior distribution specification

5. Real-World AI Applications

5.1. Natural Language Processing

- **Word Distributions:** PMFs for word frequencies in documents
- **Topic Modeling:** Dirichlet distributions over topic mixtures
- **Language Models:** Categorical distributions for next-word prediction
- **Performance:** Modern LLMs achieve perplexity scores below 20 on benchmark datasets

5.2. Computer Vision

- **Pixel Values:** Continuous distributions for color intensities
- **Feature Distributions:** Multivariate normals for learned representations
- **Anomaly Detection:** Using tail probabilities from CDFs
- **Generative Models:** GANs and VAEs learning data distributions

6. Implementation in Modern AI Frameworks

6.1. PyTorch Distributions Module

```
import torch
import torch.distributions as dist

# Discrete distribution (Binomial)
n, p = 10, 0.3
binomial = dist.Binomial(n, p)
pmf_values = torch.exp(binomial.log_prob(torch.arange(0, n+1)))
print(f"PMF at k=5: {pmf_values[5]:.4f}")

# Continuous distribution (Normal)
normal = dist.Normal(0.0, 1.0)
x = torch.linspace(-3, 3, 100)
pdf_values = torch.exp(normal.log_prob(x))
cdf_values = normal.cdf(x)

# Compute probabilities
prob_interval = normal.cdf(1.0) - normal.cdf(-1.0)
print(f"P(-1 < X < 1) = {prob_interval:.4f}")
```


7. Practical Exercises for Mastery

Hands-On Distribution Analysis

1. **Bernoulli Analysis:** Derive PMF, CDF, mean, and variance for Bernoulli(p)
2. **Normal Calculations:** Compute $P(\mu - \sigma < X < \mu + \sigma)$ for $X \sim \mathcal{N}(\mu, \sigma^2)$
3. **Empirical CDF:** Implement empirical CDF from sample data
4. **Distribution Fitting:** Fit distributions to real datasets and assess goodness-of-fit
5. **Quantile Functions:** Implement and analyze quantile functions for various distributions

8. Common Pitfalls and Best Practices

Distribution Modeling Guidelines

- **Support Matching:** Ensure distribution support matches data range
- **Tail Behavior:** Choose distributions with appropriate tail thickness
- **Discrete vs Continuous:** Never use continuous PDF for discrete data
- **Parameter Interpretation:** Understand what each parameter controls
- **Computational Stability:** Use log-probabilities for numerical stability

9. Key Insight: The Complete Uncertainty Picture

Probability distribution functions provide the complete mathematical framework for understanding and working with uncertainty in AI systems:

- **PMFs:** Capture discrete uncertainty with exact point probabilities
- **PDFs:** Describe continuous uncertainty through density functions
- **CDFs:** Provide cumulative perspective and enable probability calculations
- **Complementarity:** Each function offers different insights into the same underlying uncertainty

The power of this framework lies in its ability to transform vague notions of "uncertainty" into precise, computable mathematical objects that can be estimated from data, manipulated through mathematical operations, and optimized in machine learning objectives.

Next: Expectation and Variance — Mean and Spread of Random Variables

Tomorrow we'll explore how to summarize and characterize probability distributions through their moments—the essential statistical quantities that capture central tendency, dispersion, and shape properties of random variables.