Modern Statistics

Xiangyu Chang

March 9, 2026

Abstract

To be undated.

1 Lecture 3: Random Variable & Vector

In Lecture 2, we introduced random variables as measurable functions from a probability space to the real numbers, and we characterized their behavior using the cumulative distribution function (CDF). In this lecture, we go deeper: we define the probability density function (PDF) and probability mass function (PMF) for continuous and discrete random variables, survey important distributions that arise in practice, study how random variables transform under functions, and finally extend our framework to **random vectors**—pairs of random variables that will pave the way for conditional distributions and expectation in Lecture 4.

1.1 Recall: From Probability to Random Variables

Before proceeding, we briefly recall the key concepts from Lectures 1 and 2. The **sample space** Ω is the set of all possible outcomes of an experiment; **events** are subsets $A \subseteq \Omega$. A **probability measure** P assigns numbers to events and satisfies three axioms: non-negativity, normalization ($P(\Omega) = 1$), and countable additivity for pairwise disjoint events.

A **random variable** X is a measurable function $X : \Omega \to \mathbb{R}$. For any Borel set $A \subseteq \mathbb{R}$, we define

$$P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}).$$

The **cumulative distribution function (CDF)** of X is

$$F_X(x) = P(X \leqslant x) = P(\{\omega \in \Omega \mid X(\omega) \leqslant x\}),$$

which completely characterizes the distribution of X. With this foundation in place, we now introduce alternative ways to describe distributions—the PDF and PMF—and build a toolkit of commonly used distributions.

Example 1.1 (Example: Distribution of a Fair Die Roll)**.** Let X denote the outcome of a single roll of a fair six-sided die. We can illustrate the concepts above using this random variable:

1. PMF (Probability Mass Function):
X is discrete and takes values in $\{1, 2, 3, 4, 5, 6\}$, with equal probability for each value:

$$f_X(x) = P(X = x) = \begin{cases} \frac{1}{6}, & x \in \{1, 2, 3, 4, 5, 6\} \\ 0, & \text{otherwise} \end{cases}$$

2. CDF (Cumulative Distribution Function):

$$F_X(x) = P(X \leqslant x) = \begin{cases} 0, & x < 1 \\ \frac{1}{6}, & 1 \leqslant x < 2 \\ \frac{2}{6}, & 2 \leqslant x < 3 \\ \frac{3}{6}, & 3 \leqslant x < 4 \\ \frac{4}{6}, & 4 \leqslant x < 5 \\ \frac{5}{6}, & 5 \leqslant x < 6 \\ 1, & x \geqslant 6 \end{cases}$$

3. PDF (Probability Density Function):
Strictly speaking, discrete random variables do not have a classical PDF. Instead, their PMF describes the probability at each point and can be visualized as a "stick diagram."

4. Illustrative calculation:
To compute $P(2 < X \leqslant 5)$, that is, the probability the die shows 3, 4, or 5:

$$P(2 < X \leqslant 5) = P(X = 3) + P(X = 4) + P(X = 5) = 3 \times \frac{1}{6} = \frac{1}{2}$$

Alternatively, using the CDF:

$$P(2 < X \leqslant 5) = F_X(5) - F_X(2) = \frac{5}{6} - \frac{2}{6} = \frac{3}{6} = \frac{1}{2}$$

This example demonstrates how PMF and CDF can be used to analyze the distributional properties of a random variable.

1.2 Probability Density Function (PDF)

For continuous random variables, the CDF is often differentiable. In that case, we can work with a more convenient object: the probability density function.

Definition 1.2 (Continuous Random Variable and PDF)**.** A random variable X is **continuous** if there exists a nonnegative function $f_X : \mathbb{R} \to [0, \infty)$, called the **probability density function (PDF)**, such that

- $\int_{\mathbb{R}} f_X(x)\ \mathrm{d}x = 1$.
- For any interval (a, b), $P(a < X < b) = \int_a^b f_X(x)\ \mathrm{d}x$.

The second property follows from the fundamental theorem of calculus: $F_X(x) = \int_{-\infty}^x f_X(t)\ \mathrm{d}t$, so $P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\ \mathrm{d}x$.

Definition 1.3 (Quantile)**.** For $q \in (0,1)$, the **quantile** (or **inverse CDF**) is defined as

$$F_X^{-1}(q) \triangleq \inf\{x : F_X(x) \geqslant q\}.$$

Special cases include:

- **Median**: $F_X^{-1}(1/2)$
- **First quartile**: $F_X^{-1}(1/4)$
- **Third quartile**: $F_X^{-1}(3/4)$

Example 1.4 (Uniform Distribution on the Unit Interval)**.** Let $X \sim U(0,1)$ have PDF and CDF

$$f_X(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad F_X(x) = \begin{cases} 0, & x \leqslant 0 \\ x, & 0 < x < 1 \\ 1, & x \geqslant 1 \end{cases}.$$

Then the median is $F_X^{-1}(1/2) = 1/2$, the first quartile is $1/4$, and the third quartile is $3/4$.

For discrete random variables, we use the probability mass function instead.

Definition 1.5 (Discrete Random Variable and PMF)**.** A random variable X is **discrete** if it takes values in a countable set $\{x_1, x_2, \ldots\}$. Its **probability mass function (PMF)** is defined by

$$f_X(x) \triangleq P(X = x).$$

Example 1.6 (Binomial PMF)**.** For $X \sim \text{Bin}(n, p)$, the PMF is $f_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ for $k = 0, 1, \ldots, n$.

1.3 Important Distributions

We now catalog several fundamental distributions that appear throughout statistics and machine learning. Mastering these will provide the building blocks for more complex models.

1. **Point mass (degenerate distribution)**: $P(X = a) = 1$. The CDF is $F_X(x) = \mathbf{1}_{x \geqslant a}$.
2. **Discrete uniform**: $X \in \{1, \ldots, n\}$ with $P(X = k) = 1/n$ for $k = 1, \ldots, n$.
3. **Bernoulli**(p): $X \in \{0, 1\}$ with $P(X = x) = p^x (1 - p)^{1-x}$.
4. **Binomial**(n, p):
$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n.$$
   - **Additivity**: If $X_1 \sim \text{Bin}(n_1, p)$ and $X_2 \sim \text{Bin}(n_2, p)$ are independent, then $X_1 + X_2 \sim \text{Bin}(n_1 + n_2, p)$.
   - **Sum of Bernoulli trials**: If $X_i \sim \text{Ber}(p)$ are independent, then $\sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$.

5. **Poisson**(λ):
$$f(x) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots.$$
Verification: $\sum_{k=0}^{\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$.

6. **Geometric**: $P(X = x) = p(1-p)^x$ for $x \geqslant 0$ (number of failures before first success).

7. **Continuous uniform** $U(a, b)$:
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{otherwise} \end{cases}.$$

8. **Normal (Gaussian)** $N(\mu, \sigma^2)$:
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \mu \in \mathbb{R}, \sigma^2 > 0.$$
Here μ is the mean and σ^2 is the variance.
   - **Standard normal**: $X \sim N(0, 1)$ has CDF $\Phi(x) = P(X \leqslant x)$.
   - **Standardization**: If $Z \sim N(\mu, \sigma^2)$, then $X = \frac{Z-\mu}{\sigma} \sim N(0, 1)$.

1.4 Transformation of Random Variables

When we apply a function g to a random variable X, we obtain a new random variable $Y = g(X)$. How do we find the distribution of Y? For monotonic g, the change-of-variable formula gives the answer.

1. **Discrete case**: $f_Y(y) = P(Y = y) = P(g(X) = y) = P(X \in g^{-1}(\{y\}))$. When g is injective, $f_Y(y) = f_X(g^{-1}(y))$.

2. **Continuous case (monotonic** g**)**:
$$F_Y(y) = P(Y \leqslant y) = P(g(X) \leqslant y) = P(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y)).$$
Differentiating,
$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} \right|.$$

Proof of the formula. Let $Y = g(X)$, where g is strictly monotonic and differentiable. We derive the PDF of Y via its CDF.

Case 1: g **is strictly increasing.** Then g^{-1} is also strictly increasing, and
$$g(X) \leqslant y \Longleftrightarrow X \leqslant g^{-1}(y).$$
Thus
$$F_Y(y) = P(Y \leqslant y) = P(g(X) \leqslant y) = P(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y)).$$

Differentiating with respect to y and applying the chain rule,

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = f_X(g^{-1}(y)) \cdot \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}.$$

Since g^{-1} is increasing, $\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} > 0$, so

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

Case 2: g is strictly decreasing. Then g^{-1} is also strictly decreasing, and

$$g(X) \leqslant y \iff X \geqslant g^{-1}(y).$$

Thus

$$F_Y(y) = P(g(X) \leqslant y) = P(X \geqslant g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

Differentiating,

$$f_Y(y) = -\frac{\mathrm{d}}{\mathrm{d}y} F_X(g^{-1}(y)) = -f_X(g^{-1}(y)) \cdot \frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y}.$$

Since g^{-1} is decreasing, $\frac{\mathrm{d}g^{-1}(y)}{\mathrm{d}y} < 0$, so the right-hand side is positive and

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

In both cases we obtain

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

■

3. **Standardization of normal**: If $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$, then $P(X \leqslant x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

4. **Sum of independent normals**: If $X_i \sim N(\mu_i, \sigma_i^2)$ are independent, then $\sum_{i=1}^k X_i \sim N\left(\sum_{i=1}^k \mu_i, \sum_{i=1}^k \sigma_i^2\right)$.

Example 1.7 (Linear Transformation of Standard Normal)**.** If $X \sim N(0,1)$ and $Z = \mu + \sigma X$, then $g^{-1}(z) = (z-\mu)/\sigma$ and $|\mathrm{d}g^{-1}/\,\mathrm{d}z| = 1/\sigma$. Thus

$$f_Z(z) = f_X\left(\frac{z-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right),$$

so $Z \sim N(\mu, \sigma^2)$.

1.5 Random Vectors and Bivariate Distributions

So far we have studied single random variables. In many applications, we observe multiple quantities simultaneously—for example, height and weight of a person, or the returns of two stocks. We now extend our framework to **random vectors** (X, Y), which will be essential for conditional distributions and expectation in the next lecture.

Definition 1.8 (Joint PMF)**.** For discrete random variables X and Y, the **joint PMF** is

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

Definition 1.9 (Marginal Distribution)**.** The **marginal PMF** of X is obtained by summing over all possible values of Y:

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_j p_{ij} = p_{i\cdot}.$$

Definition 1.10 (Joint CDF and PDF)**.** For continuous (X, Y), the **joint CDF** is

$$F_{X,Y}(x, y) = P(X \leqslant x, Y \leqslant y).$$

If $F_{X,Y}$ is differentiable, the **joint PDF** $f_{X,Y}$ satisfies:

- $f_{X,Y}(x, y) \geqslant 0$ and $\int_{\mathbb{R}^2} f_{X,Y}(x, y)\ \mathrm{d}x\ \mathrm{d}y = 1$.
- $P((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\ \mathrm{d}x\ \mathrm{d}y$ for any Borel set $A \subseteq \mathbb{R}^2$.

Definition 1.11 (Independence of Random Variables)**.** Random variables X and Y are **independent** if

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B) \quad \text{for all Borel sets } A, B.$$

Equivalently, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ (for both PMF and PDF).

Transformation of random vectors: If $\boldsymbol{Z} = (X, Y)^\top$ and $\boldsymbol{W} = g(\boldsymbol{Z})$ is an invertible differentiable map, then

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = f_{\boldsymbol{Z}}(g^{-1}(\boldsymbol{w})) \cdot \left| \det \frac{\partial g^{-1}}{\partial \boldsymbol{w}} \right|.$$

Example 1.12 (Probability Integral Transform)**.** Let X be a continuous random variable with CDF F_X, and define $Y = F_X(X)$. Then $Y \sim U(0, 1)$.

Proof: For $y \in [0, 1]$,

$$P(Y \leqslant y) = P(F_X(X) \leqslant y) = P(X \leqslant F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y.$$

Hence Y has the CDF of the uniform distribution on $[0, 1]$.

References