

Modern Statistics

Xiangyu Chang

March 12, 2026

Abstract

To be undated.

1 Lecture 4: Random Vector and Expectation

In Lecture 3, we introduced random vectors (X, Y) and their joint distributions. A central question in statistics is: how does knowing one variable affect our beliefs about another? This lecture answers that question through **conditional distributions**. We then extend the framework to d -dimensional random vectors, introduce the multivariate normal and multinomial distributions, and finally define **expectation**—the first of many summary quantities that will play a fundamental role in estimation and inference in later lectures.

1.1 Recall: Joint and Marginal Distributions

We briefly recall the key concepts from Lecture 3. For a random vector (X, Y) :

- **Joint CDF:** $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$.
- **Joint PMF (discrete):** $f_{X,Y}(x, y) = P(X = x, Y = y)$.
- **Joint PDF (continuous):** $f_{X,Y}(x, y) \geq 0$, $\int_{\mathbb{R}^2} f_{X,Y}(x, y) \, dx \, dy = 1$, and $P((X, Y) \in A) = \iint_A f_{X,Y}(x, y) \, dx \, dy$.
- **Marginal distribution:** $f_X(x) = \sum_y f_{X,Y}(x, y)$ (discrete) or $f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dy$ (continuous).
- **Independence:** X and Y are independent if $P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$ for all Borel sets A, B , equivalently $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

With this foundation, we now study how to update our beliefs about Y when we observe X .

1.2 Conditional PMF and PDF

The **conditional distribution** of Y given $X = x$ describes how the distribution of Y changes when we know the value of X . It is the probabilistic analogue of “ Y given X .”

Definition 1.1 (Conditional PMF and PDF). For random variables X and Y with joint PMF or PDF $f_{X,Y}$ and marginal $f_X(x) > 0$:

- **Discrete:** $P(Y = y | X = x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f_{X,Y}(x,y)}{f_X(x)}$.
- **Continuous:** $f_{Y|X}(y | x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$.

In both cases, the **chain rule** holds:

$$f_{X,Y}(x, y) = f_{Y|X}(y | x) \cdot f_X(x).$$

When X and Y are independent, we have $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Substituting into the conditional formula yields

$$f_{Y|X}(y | x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y).$$

Thus, knowing X does not change the distribution of Y .

Example 1.2 (Conditional Uniform). Let $X \sim U[0, 1]$ and suppose that given $X = x$, we have $Y | X = x \sim U[x, 1]$. Find the marginal PDF of Y .

Step 1: The marginal of X and the conditional of Y given X are

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad f_{Y|X}(y | x) = \begin{cases} \frac{1}{1-x}, & x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Step 2: By the chain rule, the joint PDF is

$$f_{X,Y}(x, y) = f_{Y|X}(y | x) \cdot f_X(x) = \begin{cases} \frac{1}{1-x}, & 0 \leq x \leq y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Step 3: Marginalizing over x , for $y \in [0, 1]$:

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx = \int_0^y \frac{1}{1-x} \, dx = -\ln(1-y).$$

Remark 1.3 (Linear Model). A simple regression model posits $Y_i = \beta^\top x_i + \varepsilon_i$ for observations $\{(x_i, y_i)\}_{i=1}^n$, where ε_i are random errors. The conditional distribution $Y | X = x$ is central to this framework; we will return to it in the regression lectures.

1.3 Multivariate Random Vectors

We now extend the bivariate framework to d -dimensional random vectors $\mathbf{X} = (X_1, \dots, X_d)^\top \in \mathbb{R}^d$.

Definition 1.4 (Multivariate CDF and PDF). For $\mathbf{x} = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$:

- **Joint CDF:** $F_X(\mathbf{x}) = P(X_1 \leq x_1, \dots, X_d \leq x_d)$.
- **Joint PDF:** $f_X(\mathbf{x}) \geq 0$, $\int_{\mathbb{R}^d} f_X(\mathbf{x}) \, d\mathbf{x} = 1$, and $P(\mathbf{X} \in A) = \int_A f_X(\mathbf{x}) \, d\mathbf{x}$ for Borel sets $A \subseteq \mathbb{R}^d$.

Definition 1.5 (Marginal Distribution). The marginal PDF of (X_1, \dots, X_k) is obtained by integrating out the remaining variables:

$$f_{X_1, \dots, X_k}(x_1, \dots, x_k) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_X(\mathbf{x}) \, dx_{k+1} \cdots dx_d.$$

Definition 1.6 (Independence). Random variables X_1, \dots, X_d are **independent** if

$$f_X(\mathbf{x}) = \prod_{i=1}^d f_{X_i}(x_i).$$

1.3.1 Multivariate Normal Distribution

The multivariate normal distribution generalizes the univariate normal to \mathbb{R}^d .

Definition 1.7 (Standard Multivariate Normal). Let $Z = (Z_1, \dots, Z_k)^\top$ where $Z_1, \dots, Z_k \sim N(0, 1)$ are independent. The density of Z is

$$f(z) = \prod_{i=1}^k f_{Z_i}(z_i) = \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^k z_j^2 \right\} = \frac{1}{(2\pi)^{k/2}} \exp \left\{ -\frac{1}{2} \mathbf{z}^\top \mathbf{z} \right\}.$$

We write $Z \sim N(\mathbf{0}, I)$, where I is the $k \times k$ identity matrix.

Definition 1.8 (General Multivariate Normal). A random vector \mathbf{X} has a **multivariate normal distribution** $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ if its density is

$$f(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

where $\boldsymbol{\mu} \in \mathbb{R}^k$ is the mean vector, Σ is a $k \times k$ symmetric positive definite covariance matrix, and $|\Sigma|$ denotes its determinant.

Lemma 1.9. If $\mathbf{X} \sim N(\mathbf{0}, I)$, then $\mathbf{Z} = \boldsymbol{\mu} + \Sigma^{1/2} \mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$.

Proof. The transformation $\mathbf{g}^{-1}(\mathbf{z}) = \Sigma^{-1/2}(\mathbf{z} - \boldsymbol{\mu})$ has Jacobian $\nabla \mathbf{g}^{-1}(\mathbf{z}) = \Sigma^{-1/2}$. By the multivariate change-of-variable formula,

$$f_Z(\mathbf{z}) = f_X(\mathbf{g}^{-1}(\mathbf{z})) \left| \det \Sigma^{-1/2} \right| = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right\}.$$

■

Definition 1.10 (Matrix Square Root). The **square root** $\Sigma^{1/2}$ of a symmetric positive definite matrix Σ satisfies:

- $\Sigma^{1/2}$ is symmetric.
- $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$.
- $\Sigma^{1/2}\Sigma^{-1/2} = \Sigma^{-1/2}\Sigma^{1/2} = I$, where $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$.
- If $\Sigma = UDU^\top$ is an eigendecomposition, then $\Sigma^{1/2} = UD^{1/2}U^\top$ and $\Sigma^{-1/2} = UD^{-1/2}U^\top$, where $D^{1/2} = \text{diag}(\sqrt{D_{11}}, \dots, \sqrt{D_{kk}})$.

1.3.2 Multinomial Distribution

Consider tossing a die with k faces n times. Let $p = (p_1, \dots, p_k)$ with $p_j \geq 0$ and $\sum_{j=1}^k p_j = 1$, and let $X = (X_1, \dots, X_k)^\top$ where X_j counts how many times face j appears, so $\sum_{j=1}^k X_j = n$.

Definition 1.11 (Multinomial Distribution). We say $X \sim \text{Multinomial}(n, p)$ if

$$P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1!n_2!\dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k},$$

where $n_1 + \dots + n_k = n$.

Lemma 1.12. If $X \sim \text{Multinomial}(n, p)$ with $X = (X_1, \dots, X_k)^\top$, then the marginal distribution of X_j is $X_j \sim \text{Bin}(n, p_j)$.

1.4 Expectation

We now introduce the **expected value** (or **mean**) of a random variable—a single number that summarizes the “center” of its distribution. Expectation will be the cornerstone of estimation, variance, and many limit theorems in later lectures.

Definition 1.13 (Expected Value). The **expected value** (or **mean**, or **first moment**) of a random variable X is

$$\mathbb{E}[X] = \int x \, dF_X(x) = \begin{cases} \sum_x x f_X(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f_X(x) \, dx & \text{if } X \text{ is continuous} \end{cases},$$

provided the sum or integral exists (i.e., $\mathbb{E}[|X|] < \infty$).

Example 1.14 (Coin Toss). Flip a fair coin twice. Let X be the number of heads. Then $X \in \{0, 1, 2\}$ with $f_X(0) = 1/4$, $f_X(1) = 1/2$, $f_X(2) = 1/4$. Thus

$$\mathbb{E}[X] = \sum_x x f_X(x) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1.$$

Definition 1.15 (Rule of the Lazy Statistician). Let $Y = r(X)$ for a (measurable) function r . Then

$$\mathbb{E}[Y] = \mathbb{E}[r(X)] = \int r(x) \, dF_X(x) = \int r(x) f_X(x) \, dx \quad (\text{continuous case}).$$

We do *not* need to find the distribution of Y ; we integrate with respect to the distribution of X .

Example 1.16 (Indicator and Probability). Let I_A be the indicator of event A : $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise. Then

$$\mathbb{E}[I_A(X)] = 0 \cdot P(X \notin A) + 1 \cdot P(X \in A) = P(X \in A).$$

Thus expectation of an indicator recovers the probability of the corresponding event.

References