Chapter 7: Joint Distributions

MATH/STAT 394: Probability I

Arman Jahangiri

University of Washington Department of Mathematics

Summer 2026

Learning Goals

By the end of this lecture, students should be able to:

- define joint, marginal, and conditional distributions;
- understand independence of random variables;
- apply the two-dimensional LOTUS formula;
- define covariance and correlation;
- distinguish independence from zero correlation;
- define pairwise and mutual independence;
- define the Multinomial distribution;
- understand the Multivariate Normal distribution.

Joint Distributions

Why Joint Distributions?

Previously, we studied one random variable at a time.

But many probabilistic questions involve several random variables simultaneously.

Examples:

- height and weight;
- temperature and rainfall;
- several coin tosses;
- sums of random variables;
- sample means.

Main idea

Joint distributions describe how multiple random variables behave together.

Joint PMF

Definition

For discrete random variables X, Y, the joint PMF is

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y).$$

Properties:

$$p_{X,Y}(x,y) \geq 0, \quad \textit{and} \quad \sum_{x}\sum_{y} p_{X,Y}(x,y) = 1.$$

Similarly, for n discrete random variables $X_1, \ldots, X_n$, the joint PMF is

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

Note $\{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\}$

Marginal PMFs

The marginal PMFs are obtained by summing out the other variable.

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

and

$$p_Y(y) = \sum_x p_{X,Y}(x, y).$$

Interpretation

Marginals describe each variable individually.

Example: Two Fair Dice and Marginals

Let X be the outcome of the first die and Y the outcome of the second die.

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) = \frac{1}{36}, \qquad x, y \in \{1,2,3,4,5,6\}.$$

$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y)$	$y = 1$	$y = 2$	$y = 3$	$y = 4$	$y = 5$	$y = 6$	$p_X(x) = \mathbb{P}(X = x)$
$x = 1$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 2$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 3$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 4$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 5$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$x = 6$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$
$p_Y(y) = \mathbb{P}(Y = y)$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	

For example,

$$p_X(3) = \sum_{y=1}^{6} p_{X,Y}(3,y) = \frac{1}{36} + \cdots + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}, \quad \textit{and} \quad p_Y(4) = \sum_{x=1}^{6} p_{X,Y}(x,4) = \frac{6}{36} = \frac{1}{6}.$$

The marginal PMFs appear on the "margins" of the joint PMF table.

Proof of Marginal PMFs

We prove the formula for $p_X(x)$.
Fix a value x in the support of X. Then

$$p_X(x) = \mathbb{P}(X = x).$$

The event $\{X = x\}$ can be written as the union of all events

$$\{X = x,\ Y = y\}$$

over all possible values y of Y:

$$\{X = x\} = \bigcup_y \{X = x,\ Y = y\}.$$

Reason:

$$\{X = x\} = (\{X = x\} \cap \mathcal{S}) = \{X = x\} \cap (\bigcup_y \{Y = y\}) = \bigcup_y \{X = x\} \cap \{Y = y\} = \bigcup_y \{X = x, Y = y\}$$

Thus, applying the probability function to the above equation yields the result. Indeed, these events are pairwise disjoint, because Y cannot take two different values at the same time. Therefore, by the third axiom of probability (countable additivity),

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x,\ Y = y).$$

Hence

Conditional PMFs

Definition

If

$$\mathbb{P}(X = x) > 0,$$

then

$$p_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

Conditional distributions describe one variable given information about another.

Definition

For continuous random variables X, Y, the joint PDF is a function

$$f_{X,Y}(x, y)$$

such that

$$\mathbb{P}((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\, dx\, dy.$$

Properties:

$$f_{X,Y}(x, y) \geq 0, \qquad \iint_{\mathbb{R}^2} f_{X,Y}(x, y)\, dx\, dy = 1.$$

Similarly, for n continuous random variables

$$X_1, \ldots, X_n,$$

the joint PDF is a function

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

such that

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = \int_A \cdots \int f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Marginal PDFs

Marginal densities are obtained by integrating out the other variable.

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx.$$

Conditional PDFs

Definition

If

$$f_X(x) > 0,$$

then

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

Example: Uniform on the Unit Square

Suppose

$$(X, Y) \sim \text{Unif}([0,1]^2).$$

Then:

$$f_{X,Y}(x,y) = 1, \qquad 0 < x < 1,\ 0 < y < 1.$$

Marginals:

$$f_X(x) = 1, \qquad f_Y(y) = 1.$$

Thus:

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

Example: Uniform on the Unit Disk

Suppose (X, Y) is uniformly distributed on the unit disk

$$D = \{(x, y) : x^2 + y^2 \leq 1\}.$$

Since the area of the unit disk is π, the joint density is

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{\pi}, & x^2 + y^2 \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Indeed,

$$\iint_{\mathbb{R}^2} f_{X,Y}(x, y)\, dx\, dy = \iint_D \frac{1}{\pi}\, dx\, dy = \frac{1}{\pi} \operatorname{Area}(D) = 1.$$

For a fixed value of $X = x$, the possible values of Y satisfy

$$-\sqrt{1 - x^2} \leq Y \leq \sqrt{1 - x^2}.$$

Important

Knowing X restricts the possible values of Y, so X and Y are dependent.

Independence

Independence

Definition

Random variables X, Y are independent if for all sets A, B,

$$\mathbb{P}(X \in A,\ Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Intuitively: knowing one variable gives no information about the other.

Equivalent Definitions of Independence

Theorem

The following are equivalent for random variables X and Y:

1. X and Y are independent.
2. $$F_{X,Y}(x,y) = F_X(x)F_Y(y).$$
3. If discrete:
$$p_{X,Y}(x,y) = p_X(x)p_Y(y).$$
4. If continuous:
$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$
5. $$f_{Y|X}(y|x) = f_Y(y).$$
6. $$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$
7. If MGFs exist:
$$M_{X,Y}(s,t) = M_X(s)M_Y(t).$$

Which Criterion Should We Use?

Different characterizations are useful in different situations.

- PMF/PDF factorization: usually easiest computationally;
- conditional distributions: most conceptual;
- expectation factorization: useful theoretically;
- MGF factorization: useful for sums and CLT.

Independence of Several Random Variables

Definition

Random variables

$$X_1, \ldots, X_n$$

are mutually independent if for every subset

$$i_1, \ldots, i_k,$$

$$\mathbb{P}(X_{i_1} \in A_1, \ldots, X_{i_k} \in A_k) = \prod_{j=1}^{k} \mathbb{P}(X_{i_j} \in A_j).$$

Pairwise Independence

Definition

Random variables

$$X_1, \dots, X_n$$

are pairwise independent if every pair

$$(X_i, X_j)$$

is independent.

$$\text{mutual independence} \implies \text{pairwise independence.}$$

But the converse is false.

Pairwise Independent But Not Mutually Independent

Let
$$X, Y$$
be independent Bernoulli(1/2) random variables.
Define
$$Z = X \oplus Y,$$
where $\oplus$ denotes XOR.

Then:

- every pair is independent;
- but
  $$Z$$
  is completely determined by
  $$X, Y.$$

Hence:
$$X, Y, Z$$
are not mutually independent.

2D LOTUS

2D LOTUS

Theorem

For a function $g(X, Y)$,

$$\mathbb{E}[g(X, Y)] = \begin{cases} \sum_x \sum_y g(x, y) p_{X,Y}(x, y), & \text{if } X, Y \text{ are discrete,} \\ \iint g(x, y) f_{X,Y}(x, y)\, dx\, dy, & \text{if } X, Y \text{ are continuous.} \end{cases}$$

Meaning

To find the expected value of a function of two random variables, average $g(x, y)$ using the joint distribution of (X, Y).

Linearity of Expectation Revisited

Using 2D LOTUS:

$$E(X+Y) = E(X) + E(Y).$$

More generally:

$$E(aX+bY+c) = aE(X) + bE(Y) + c.$$

Importantly: independence is NOT required.

Covariance and Correlation

Covariance

Definition

The covariance of X, Y is
$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Equivalent formula:
$$\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Proposition: $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

Interpretation of Covariance

- Positive covariance: variables tend to move together;
- Negative covariance: one tends to increase while the other decreases;
- Zero covariance: no **linear** relationship (but there may be other forms of nonlinear relationships).

Important

Covariance measures linear dependence only.

Correlation

Definition

The correlation coefficient is
$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Properties:
$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

Proof: Chapter 10a (inequalities) - needs Cauchy-Schwarz inequality.

Theorem

For random variables X, Y,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y).$$

If X, Y are independent:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Independence Implies Zero Correlation

Theorem

If X, Y are independent, then
$$\mathrm{Cov}(X, Y) = 0.$$
Hence:
$$\mathrm{Corr}(X, Y) = 0.$$

This follows because:
$$E(XY) = E(X)E(Y).$$

Zero Correlation Does Not Imply Independence

Let

$$X \sim N(0,1),$$

and define

$$Y = X^2.$$

Then:

$$E(X) = 0,$$

and

$$E(XY) = E(X^3) = 0.$$

Hence:

$$\mathrm{Cov}(X,Y) = 0.$$

But:

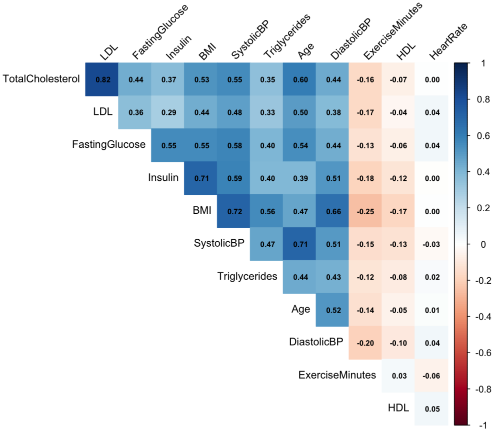
$$Y$$

is completely determined by

$$X.$$

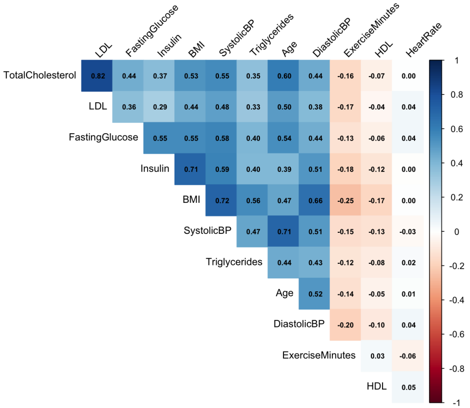
Thus X, Y are dependent.

Now, suppose that instead of 2 variables, we have n variables $X_1, \dots, X_n$. If we consider all the correlations, i.e., $\{\mathrm{Corr}(X_i, X_j) \ : \ 1 \le i \ne j \le n\}$, then there is $\binom{n}{2}$ distinct correlations, so we can define a matrix whose ij entry is $\mathrm{Corr}(X_i, X_j)$. This matrix is symmetric since $\mathrm{Corr}(X_i, X_j) = \mathrm{Corr}(X_j, X_i)$. There is an interesting way to visualize this matrix, because all entries are between -1 and 1, we can use a heatmap.



Correlation Heatmap

- A **heatmap** provides a visual summary of pairwise correlations between variables.
- Each cell shows the correlation coefficient
$$r(X, Y)$$
between two variables.
- Interpretation:
  - Blue: positive correlation.
  - Red: negative correlation.
  - White/light colors: weak or near-zero correlation.
- The closer $|\mathrm{Corr}(X, Y)|$ is to 1, the stronger the linear relationship.



Multinomial Distribution

Multinomial Distribution (generalization of Binomial Distribution)

In the Binomial setting, each trial has two possible outcomes:

$$o_1 : \text{Success} \qquad \text{or} \qquad o_2 : \text{Failure},$$

with probabilities

$$p_1 = p, \quad p_2 = 1 - p, \qquad \textit{where} \qquad p_1 + p_2 = 1$$

After n independent trials,

$$X_1 := X \qquad \textit{and} \qquad X_2 := n - X$$

count the number of trials resulting in category o_1: Success and o_2: Failure, respectively. Note that $X_1 + X_2 = X + n - X = n$

In the multinomial setting, each trial can result in one of k (instead of 2) possible categories

$$o_1, \ldots, o_k,$$

with probabilities

$$p_1, \ldots, p_k, \qquad \textit{where} \qquad p_1 + \cdots + p_k = 1$$

After n independent trials,

$$X_1, \ldots X_k$$

count the number of trials resulting in category $o_1, \ldots, o_k$. Note that $X_1 + \cdots + X_k = n$, since each trial results in exactly one category.

Multinomial PMF

Definition

If $(X_1, \ldots, X_k)$ counts the numbers of occurrences of the k categories in n independent trials, then

$$\boxed{(X_1, \ldots, X_k) \sim \text{Mult}(n, p_1, \ldots, p_k).}$$

$$\boxed{\mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},}$$

subject to:

$$x_1 + \cdots + x_k = n.$$

Note. The Multinomial distribution is a multivariate distribution $k - \textit{dimensional}$, while the Binomial is a univariable distribution $1 - \textit{dimensional}$. A multivariate random variable is also called a *random vector*. **Note**. It can be shown using the Multinomial Theorem (an extension of the Binomial Theorem) that the Multinomial PMF sums up to 1. Click here to see the Multinomial Theorem in Wikipedia.

Moments of the Multinomial

For

$$(X_1, \dots, X_k) \sim \mathrm{Mult}(n, p_1, \dots, p_k),$$

$$E(X_i) = np_i,$$

$$\mathrm{Var}(X_i) = np_i(1 - p_i),$$

and for

$$i \neq j,$$

$$\boxed{\mathrm{Cov}(X_i, X_j) = -np_i p_j.}$$

Multivariate Normal Distribution

Bivariate Normal Distribution

Big Picture

The bivariate normal distribution is the two-dimensional version of the normal distribution.
It describes two continuous random variables X and Y whose joint density has a bell-shaped surface over the xy-plane.

A bivariate normal distribution is determined by five parameters:

$$\mu_X, \quad \mu_Y, \quad \sigma_X^2, \quad \sigma_Y^2, \quad \rho.$$

Here,

$$\rho = \mathrm{Corr}(X, Y)$$

controls the strength and direction of the linear relationship between X and Y.

▸ Open interactive bivariate normal visualization Click this link for interactive bivariate normal visualization.

Bivariate Normal Distribution: Density

Definition

We say

$$(X, Y) \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$

if the joint density of (X, Y) is

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}.$$

where $\sigma_X > 0$, $\sigma_Y > 0$, and $-1 < \rho < 1$.

If

$$(X, Y) \sim N_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right),$$

then:

- The means are

$$\mathbb{E}(X) = \mu_X, \qquad \mathbb{E}(Y) = \mu_Y.$$

- The variances are

$$\mathrm{Var}(X) = \sigma_X^2, \qquad \mathrm{Var}(Y) = \sigma_Y^2.$$

- The covariance is

$$\mathrm{Cov}(X, Y) = \rho\sigma_X\sigma_Y.$$

- Therefore,

$$\mathrm{Corr}(X, Y) = \rho.$$

Marginal Distributions

A beautiful property of the bivariate normal distribution is that the marginal distributions are normal.

Marginals

If

$$(X, Y) \sim N_2\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right),$$

then

$$\boxed{X \sim N(\mu_X, \sigma_X^2)}$$

and

$$\boxed{Y \sim N(\mu_Y, \sigma_Y^2).}$$

So the joint distribution is bivariate normal, but each marginal variable individually is ordinary normal.

Conditional Distributions

Another beautiful property is that the conditional distributions are also normal.

Conditionals

If

$$(X, Y)$$

is bivariate normal, then

$$\boxed{X \mid Y = y \sim N\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y),\ \sigma_X^2(1 - \rho^2)\right)}$$

and

$$\boxed{Y \mid X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X),\ \sigma_Y^2(1 - \rho^2)\right).}$$

The conditional mean is linear in the observed value.

Independence and Zero Correlation

For general random variables,

$$\mathrm{Corr}(X, Y) = 0$$

does not necessarily imply independence.

But for the bivariate normal distribution, it does.

Important Gaussian Property

If (X, Y) is bivariate normal, then

$$\boxed{X \text{ and } Y \quad \textit{independent} \quad \iff \quad \rho = 0.}$$

When $\rho = 0$, the joint density factors:

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

So in the bivariate normal case,

$$\text{uncorrelated} \quad \iff \quad \text{independent}.$$