Multiple Random Variables

Univariate vs. Multivariate Models

- Discussed so far: *univariate models*.
- This lecture's focus: *multivariate models*.

Why Multivariate Models?

- Rare to observe just one random variable.
- E.g., with people: body weight, temperature, height, blood pressure, etc.
- Need models for multiple random variables.
- Focus: *Bivariate models* with two random variables.

Random Vectors

Definition: An n-**dimensional random vector** is a function from a sample space S into $\mathbb{R}^n$.

Example
Roll two dice. There are 36 outcomes.

- $X = |\text{difference of 2 dice}|$ (i.e., absolute value of the difference between the outcomes of the two dice)
- $Y = \text{sum of 2 dice}$

(X, Y) is a 2-dimensional random vector, specifically a *discrete* bivariate random vector.

Joint Probability Mass Function

Definition: For discrete bivariate random vector (X, Y), the function $f(x, y)$ defined as:

$$f(x, y) = P(X = x, Y = y)$$

is called the *joint pmf* of (X, Y).

Requirements:

1. $0 \leq f(x, y) \leq 1$
2. $\displaystyle\sum_{(x,y) \in \mathbb{R}^2} f(x, y) = 1$

That is, any nonnegative function from $\mathbb{R}^2$ into $\mathbb{R}$ that:

- is nonzero for at most a countable number of (x, y) pairs,
- sums to 1,

is the joint pmf for some bivariate discrete random vector (X, Y).

Examples

Example 1: Define:

$$f(x,y) = \begin{cases} f(0,0) = f(0,1) & = 1/6 \\ f(1,0) = f(1,1) & = 1/3 \\ \text{otherwise} & = 0 \end{cases}$$

Example 2 (Joint pmf for Two Dice)
Find $f_{X,Y}(x,y)$ for (X,Y) defined with two dice before.

$X\backslash Y$	2	3	4	5	6	7	8	9	10	11	12
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$
1	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0
2	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0
3	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0
4	0	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0	0
5	0	0	0	0	0	$\frac{1}{18}$	0	0	0	0	0

Using the Joint pmf

The joint pmf can be used to compute the probability of any event defined in terms of (X, Y). Given:

$$P((X, Y) \in A) = \sum_{(x,y)\in A} f(x, y)$$

Example: Consider:

$$A = \{(x, y) : x \leq 4 \text{ and } y = 7\}$$

Only relevant pairs: $(x, y) = (1, 7)$ and $(x, y) = (3, 7)$.

$$P(X \leq 4,\ Y = 7) = f(1, 7) + f(3, 7) = \frac{1}{18} + \frac{1}{18} = \frac{1}{9}$$

Marginal pmf

Even if considering a model for (X, Y), we might be interested in probabilities only for X.

$$f_X(x) = P(X = x)$$

This is called the *marginal pmf of X*.

Theorem (Calculating the Marginal pmf)
For random vector (X, Y) with joint pmf $f_{X,Y}(x, y)$:

$$f_X(x) = \sum_y f_{X,Y}(x, y)$$

and

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Example

Marginal pmf for the rv's in the two-dice example:

$X \backslash Y$	2	3	4	5	6	7	8	9	10	11	12	$f_X(x)$
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{6}{36}$
1	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{10}{36}$
2	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	$\frac{8}{36}$
3	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0	$\frac{6}{36}$
4	0	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0	0	$\frac{4}{36}$
5	0	0	0	0	0	$\frac{1}{18}$	0	0	0	0	0	$\frac{2}{36}$
$f_Y(y)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

Marginal pmf - Usage

The marginal pmf of X or Y:

- Matches the pmf of X or Y defined previously.
- Computes probabilities involving only X or only Y.
- Joint pmf is required for probabilities involving both X and Y.

Dice Probabilities: Marginal Distribution

Example
Calculate the quantities involved only X or Y. E.g.

$$P(X < 3) = f_X(0) + f_X(1) + f_X(2) = 1/6 + 5/18 + 2/9 = 2/3.$$

Importance of Joint Distribution

Notes:

- Marginal distributions of X and Y do not completely describe the joint distribution.
- There could be many joint distributions with the same marginal distributions.

Example
(Same Marginals, but Different Joint pmf)

$$\begin{aligned} (1) \quad & f(0,0) = 1/12, f(1,0) = 5/12, \\ & f(0,1) = f(1,1) = 3/12 \\ (2) \quad & g(0,0) = g(0,1) = 1/6, \\ & g(1,0) = g(1,1) = 1/3 \end{aligned}$$

Joint pmf Tables

$X \backslash Y$	0	1	$f_X(x)$
0	1/12	5/12	1/2
1	3/12	3/12	1/2
$f_Y(y)$	1/3	2/3	1

$U \backslash V$	0	1	$g_U(u)$
0	1/6	1/3	1/2
1	1/6	1/3	1/2
$g_V(v)$	1/3	2/3	1

(X, Y) and (U, V) have the same marginal distributions, but their joint pmfs are different!

Continuous Bivariate Random Vectors

To this point, we discussed discrete bivariate random vectors. We can also consider random vectors whose components are continuous random variables.

Definition
Joint Density Function: A function $f(x, y)$ from $\mathbb{R}^2$ into $\mathbb{R}$ is called a **joint probability density function** or **joint pdf** of the continuous bivariate random vector (X, Y) if:

$$P((X, Y) \in A) = \iint_A f(x, y) dxdy.$$

Properties of Joint pdf

Notes:

1. A valid joint pdf $f(x, y)$ must satisfy:
    - $f(x, y) \geq 0$ for all x, y
    - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$
2. The marginal pdfs of X and Y (respectively):

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

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Example: Calculating Joint Probabilities - 1

Define the joint pdf of (X, Y) by:

$$f(x, y) = \begin{cases} 6xy^2, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Questions:

1. Show it's a valid joint pdf.
2. Find $f_X(x)$ and $f_Y(y)$.
3. Calculate $P(X + Y \geq 1)$.

Solution:

Example: Calculating Joint Probabilities - 2

Joint pdf for a continuous random vector:

$$f(x, y) = \begin{cases} e^{-y}, & \text{if } 0 < x < y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Find $P(X + Y \geq 1)$.
See the last slide for the **Solution**.

Joint CDF of (X,Y)

Definition
The *joint cdf* of (X, Y) is:

$$F(x, y) = P(X \leq x, Y \leq y)$$

For discrete case:

$$F(x, y) = \sum_{t \leq y} \sum_{s \leq x} f(s, t).$$

For continuous case:

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t) ds dt$$

From the Fundamental Theorem of Calculus (for two variables):

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y).$$

Conditional Distributions and Independence

When observing two random variables, (X, Y), the values are usually related. Knowing the value of X can give us information about Y. Conditional probabilities can be computed using the joint distribution of (X, Y).

Conditional Distribution: Discrete Case

Definition
For a discrete bivariate random vector (X, Y) with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$:

- Conditional pmf of X given $Y = y$: $f(x|y) = \dfrac{f(x, y)}{f_Y(y)}$ provided that $f_Y(y) \neq 0$
- Conditional pmf of Y given $X = x$: $f(y|x) = \dfrac{f(x, y)}{f_X(x)}$ provided that $f_X(x) \neq 0$

Properties of Conditional pmf

$f(x|y)$ is a valid pmf (provided $f_Y(y) \neq 0$) since:

- $f(x|y) \geq 0$ for all x.
- $\sum_x f(x|y) = 1$.

Similarly for $f(y|x)$.

Example (Calculating Conditional pmfs)

Define the joint pmf of (X, Y):

- $f(10, 0) = f(20, 0) = 2/18$
- $f(10, 1) = f(30, 1) = 3/18$
- $f(20, 1) = 4/18$, $f(30, 2) = 4/18$

1. Obtain conditional distribution of X given $Y = y$.
2. Find $P(X > 10|Y = 1)$.

Solution:

X \ Y	0	1	2	$f_X(x)$
10	$2/18$	$3/18$	0	$5/18$
20	$2/18$	$4/18$	0	$6/18$
30	0	$3/18$	$4/18$	$7/18$
$f_Y(y)$	$4/18$	$10/18$	$4/18$	1

Conditional Distribution: Continuous Case

Definition
For a continuous bivariate random vector with joint pdfs $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$:

- Conditional pdf of X given $Y = y$ (provided that $f_Y(y) \neq 0$):

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

- Conditional pdf of Y given $X = x$ (provided that $f_X(x) \neq 0$):

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

Indicator Functions

Recall

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

We will use the indicator function to specify support of a distribution or a r.v.

For example:

$$f(x; \lambda) = \lambda e^{-\lambda x} I(x > 0)$$

Example: Calculating Conditional pdfs

Let the continuous random vector (X, Y) have joint pdf:

$$f_{X,Y}(x, y) = e^{-y} \cdot I(0 < x < y < \infty)$$

Questions:

1. Marginal pdfs of X and Y.
2. Find $f(y|x)$ for any x such that $f_X(x) > 0$.

Solution:

Independence

Definition
Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. The random variables X and Y are *independent* if:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all } (x, y) \in \mathbb{R}^2$$

If X and Y are independent, then:

$$f(x|y) = f_X(x) \quad \text{and} \quad f(y|x) = f_Y(y)$$

Example: Checking Independence - 1

Consider the discrete bivariate random vector (X, Y) with:

$$f(10,1) = f(20,1) = f(20,2) = 1/10$$

$$f(10,2) = f(10,3) = 1/5$$

$$f(20,3) = 3/10$$

Are X and Y independent?

Solution:

Checking Independence without Marginals

Question: Can we determine independence from a joint pdf or pmf function without the marginals?

Theorem
Let (X, Y) have joint pdf or pmf $f(x, y)$. X and Y are independent iff functions $g(x)$ and $h(y)$ exist such that:

$$f(x, y) = g(x)h(y) \quad \text{for all } x, y \in \mathbb{R}$$

Note: If true, then $f_X(x) = cg(x)$ and $f_Y(y) = dh(y)$ where c and d make them valid pdfs or pmfs.

Example: Checking Independence - 2

Consider X and Y with joint pdf:

$$f_{X,Y}(x,y) = \frac{1}{384}x^2y^4e^{-y-x/2} \quad \text{for } x > 0,\ y > 0$$

Are X and Y independent?

Solution:

Notes on Independence

1. If the support of joint pdf of X, Y isn't rectangular, X and Y aren't independent.

2. If X and Y are independent with marginals $f_X(x)$ and $f_Y(y)$, then:

$$f(x, y) = f_X(x) f_Y(y)$$

Theorem
If $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ are independent, then $Z = X + Y$ has $N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ distribution.

Example

Let X and Y be independent Exponential(1) random variables.

(a) Joint pdf of (X, Y).

(b) Find $P(X \geq 4, Y < 3)$.

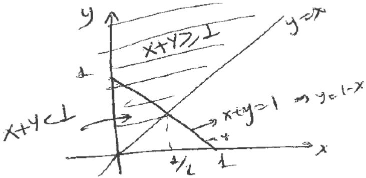
Solution:

$$f(x, y) = e^{-(x+y)} \quad \text{for } x > 0,\ y > 0$$

$$P(X \geq 4, Y < 3) = e^{-4}(1 - e^{-3}) \approx 0.017$$

(III) Solution:

$P(X+Y \geq 1) = ?$



$P(X+Y \geq 1) = 1 - P(X+Y<1)$

$= 1 - \int_0^{1/2}\int_x^{1-x} e^{-y}\,dy\,dx$

$= 1 + \int_0^{1/2}\left(e^{-y}\Big|_{y=x}^{y=1-x}\right)dx$

$= 1 + \int_0^{1/2}\left(e^{x-1} - e^{-x}\right)dx = 1 + \left(e^{x-1} + e^{-x}\right)\Big|_0^{1/2}$

$= 1 + e^{-1/2} + e^{-1/2} - e^{-1} - e^{0} = 2e^{-1/2} - e^{-1} = \frac{2-\sqrt{e}}{e}$

≈ 0.845

Multivariate Distributions

Multivariate PMFs and Densities

Introduction:

- Individual pmfs f_X and densities f_X don't describe correlations (i.e., joint trends) between variables.
- Need multivariate distributions to describe these relationships.

Multivariate PMFs

Multivariate Probability Mass Functions

Recall: For a single discrete random variable X, the distribution lists all values of X with their probabilities.

For discrete random variables U and V:

$$f_{U,V}(i, j) = P(U = i \text{ and } V = j) \tag{1}$$

Example: Yellow and Blue Marbles

Consider a bag with two yellow marbles, three blue ones, and four green ones. Four marbles are chosen at random, without replacement. Let Y and B denote the number of yellow and blue marbles chosen:

$$f_{Y,B}(i,j) = \frac{\binom{2}{i}\binom{3}{j}\binom{4}{4-i-j}}{\binom{9}{4}} \tag{2}$$

Distribution Table for Y and B

$i \setminus j$	0	1	2	3
0	0.0079	0.0952	0.1429	0.0317
1	0.0635	0.2857	0.1905	0.1587
2	0.0476	0.0952	0.0238	0

This table represents the distribution of the pair (Y, B).

Marginal PMFs

The univariate pmfs, termed *marginal pmfs*, can be derived from the multivariate pmf:

$$f_U(i) = \sum_j f_{U,V}(i,j) \tag{3}$$

and similarly for V.

Expected Value

For any function $g()$ of two discrete random variables U and V, the expected value of $g(U, V)$ is:

$$E[g(U,V)] = \sum_i \sum_j g(i,j) f_{U,V}(i,j) \tag{4}$$

Example: Consider the earlier marble example. If we wish to find the expected value of the product of the numbers of yellow and blue marbles:

$$E(YB) = \sum_{i=0}^{2} \sum_{j=0}^{3} ij\ f_{Y,B}(i,j) = 0.255 \tag{5}$$

Multivariate PDFs

Motivation and Definition

Extending our previous definition of cdf for a single variable, we define the two-dimensional cdf for a pair of random variables X and Y (discrete or continuous) as:

$$F_{X,Y}(x, y) = P(X \leq x \text{ and } Y \leq y) \tag{6}$$

If X and Y were discrete, we would evaluate that cdf via a double sum of their bivariate pmf. You may have guessed by now that the analog for continuous random variables would be a double integral, and it is. The integrand is the bivariate density:

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x, y) dxdy \tag{7}$$

1

[1]Some rv's are neither discrete nor continuous, there are some pairs of continuous random variables whose cdfs do not have the requisite derivatives. We will not pursue such cases here.

Bivariate Density

As in the univariate case, a bivariate density shows which regions of the $X - Y$ plane occur more frequently, and which occur less frequently.

By analogy, for any region A in the $X - Y$ plane,

$$P[(X, Y) \in A] = \iint_A f_{X,Y}(x, y)\ dx\ dy \tag{8}$$

Probabilities involving X and Y are found by taking the double integral of $f_{X,Y}$ over that region.

Expected Values & Marginal Densities

For any function $g(X, Y)$:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\; dx\; dy \tag{9}$$

Marginal densities:

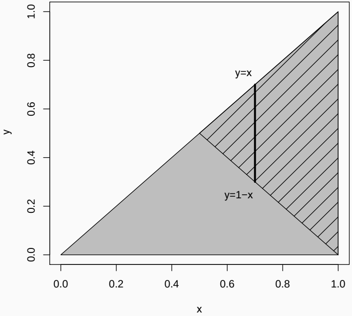
$$f_X(x) = \int_y f_{X,Y}(x,y)\; dy \tag{10}$$

Example (A Distribution with Triangular Support) Suppose (X, Y) has the density:

$$f_{X,Y}(x,y) = \begin{cases} 8xy, & \text{if } 0 < y < x < 1 \\ 0, & \text{otherwise.} \end{cases}, \tag{11}$$

i.e. the density is 0 outside the region $0 < y < x < 1$. Find $P(X + Y > 1)$.

Probability Calculation



To find $P(X+Y>1)$:

$$P(X+Y>1) = \int_{1/2}^{1}\int_{1-x}^{x} 8xy\ dy\ dx = \frac{5}{6} \tag{12}$$

Expected Value of $\sqrt{X+Y}$: Following the formula:

$$E[\sqrt{X+Y}] = \int_{0}^{1}\int_{0}^{x} \sqrt{x+y}\ 8xy\ dy\ dx \tag{13}$$

Marginal Densities and Other Calculations

$$f_X(x) = \int_0^x 8xy \; dy = 4x^3 \text{ for } 0 < x < 1$$

$$f_Y(y) = \int_y^1 8xy \; dx = 4y(1 - y^2) \text{ for } 0 < y < 1$$

$$E(X^2) = \int_0^1 x^2 \cdot 4x^3 \; dx = \frac{2}{3}$$

$$Var(X) = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = 0.027$$

$$E(Y^2) = \int_0^1 y^2 \cdot (4y - 4y^3) \; dy = \frac{1}{3}$$

$$Var(Y) = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = 0.049$$

$$\rho(X, Y) = \frac{0.018}{\sqrt{0.027 \cdot 0.049}} = 0.49$$

Example: Train Rendezvous

Train lines A and B intersect at a certain transfer point, with the schedule stating that trains from both lines will arrive there at 3:00 p.m. However, they are often late, by amounts X and Y (respectively), measured in hours, for the two trains. The bivariate density is

$$f_{X,Y}(x, y) = 2 - x - y,\ 0 < x, y < 1 \tag{14}$$

Two friends agree to meet at the transfer point, one taking line A and the other B. Let W denote the time in minutes the person arriving on line B must wait for the friend. Find $P(W > 6 \text{ minutes})$.

$$P(W > 6) = P(60(X - Y) > 6) = P(X - Y > 0.1)$$
$$= \int_{0.1}^{1} \int_{0}^{x-0.1} (2 - x - y)\ dy\ dx \approx 0.405$$

Probability Mass Functions and Densities Factor in the Independent Case

The joint pmf/density is the product of the marginal ones.
That is, if X and Y are independent:

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

Proof for Discrete Case

$$\begin{aligned} f_{X,Y}(i,j) &= P(X = i \text{ and } Y = j) \\ &= P(X = i)P(Y = j) \\ &= f_X(i)f_Y(j) \end{aligned}$$

Proof for Continuous Case

Recall that

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) \tag{15}$$

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} P(X \leq x \text{ and } Y \leq y) \\ &= \frac{\partial^2}{\partial x \partial y} \left[P(X \leq x) \cdot P(Y \leq y) \right] = \frac{\partial^2}{\partial x \partial y} (F_X(x) F_Y(y)) \\ &= f_X(x) f_Y(y) \end{aligned}$$

Moments, Conditional Expectation, Multivariate Distributions, and Inequalities

Moments and Moment Generating Functions

Moments and Moment Generating Functions

The various moments of a distribution are an important class of expectations.

Definition

For each non-negative integer k, **the k^{th} moment of** X is $E\left[X^k\right]$.

The k^{th} **central moment of** X is $E\left[(X-\mu)^k\right]$, where $\mu = E[X]$.

Notes:

The first moment of X is the mean $\mu = E[X]$.

The second moment of X is $E[X^2]$.

Variance

Aside from the mean, $E[X]$, of a random variable, perhaps the most important moment is the second central moment, known as the *variance*.

Definition
The *variance* of a random variable X is its second central moment

$$\text{Var}(X) = E\left[(X - E[X])^2\right] = E\left[(X - \mu)^2\right]$$

The positive square root of $\text{Var}(X)$ is the *standard deviation* of X.

Question: What is the first central moment of X?
Answer: $E[X - \mu] = 0$

Notes on Variance

- Variance and standard deviation are measures of spread.
- Var(X) is always ≥ 0, and so is the SD.
- When is Var(X) $= 0$ (or $SD(X) = 0$)?
- The standard deviation is easier to interpret since its measurement unit is the same as X.

An alternate formula for variance is:

$$\text{Var}(X) = E\left[X^2\right] - (E[X])^2 = E\left[X^2\right] - \mu^2$$

A Property of Variance

If X is a random variable with finite variance, then, for any (finite) constants a and b,

$$\mathrm{Var}\left(aX+b\right) = a^2\mathrm{Var}(X).$$

Proof:

$$\begin{aligned}
\mathrm{Var}\left(aX+b\right) &= E\left[(aX+b)^2\right] - \left(E\left[aX+b\right]\right)^2 \\
&= E\left[a^2X^2+2abX+b^2\right] - (aE[X]+b)^2 \\
&= a^2E\left[X^2\right] + 2abE[X] + b^2 - (a^2(E[X])^2 + 2abE[X] + b^2) \\
&= a^2E\left[X^2\right] - a^2(E[X])^2 \\
&= a^2(E\left[X^2\right] - (E[X])^2) \\
&= a^2\mathrm{Var}(X).
\end{aligned}$$

Example: Variance of Exponential Distribution

Let X have an Exponential(λ) distribution. We previously found that $E[X] = \frac{1}{\lambda}$. Find Var(X).

Solution:

$$E\left[X^2\right] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$$
$$= ... = \frac{2}{\lambda^2} \quad \text{(integration by parts twice)}$$

So,

$$\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Moment Generating Functions (MGFs)

Definition
Let X be a random variable with cdf F_X. The *moment generating function* (MGF) of X is

$$M_X(t) = E\left[e^{tX}\right],$$

provided the expectation exists for t near 0. If not, the MGF doesn't exist.

Discrete Case:

$$M_X(t) = \sum_{x \in \mathcal{X}} e^{tx} f_X(x) = \sum_{x \in \mathcal{X}} e^{tx} P(X = x)$$

Continuous Case:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \int_{\mathcal{X}} e^{tx} f_X(x) dx$$

MGF to Moments

Note: MGFs can be used to find moments of X.

Theorem
If X has MGF $M_X(t)$, then the k-th moment of X can be found as

$$E\left[X^k\right] = M_X^{(k)}(0)$$

where $M_X^{(k)}(0) = \left.\frac{d^k}{dt^k} M_X(t)\right|_{t=0}$.

Proof:

MGF of Exponential Distribution

Example: Let X have Exponential(λ) distribution,

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{for } 0 < x < \infty,\ \lambda > 0.$$

Find the MGF of X.

Solution:

MGF of Binomial Distribution

Example: Find the MGF of a Binomial(n, p) random variable X.

Solution:

$$\begin{aligned} M_X(t) = E\left[e^{tX}\right] &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x (1-p)^{n-x} \\ &= \left(pe^t + 1 - p\right)^n \end{aligned}$$

Notes on MGFs

The main use of the MGF is not in its ability to generate moments, but rather, an MGF uniquely determines a distribution.

Theorem
If X and Y have MGFs $M_X(t)$ and $M_Y(t)$ respectively, and if $M_X(t) = M_Y(t)$ for all t in an interval around 0, then X and Y have the same distribution.

MGFs for Linear Transformation of Random Variables

Theorem
For any constants a and b, the MGF of the random variable $aX + b$ is given by $M_{aX+b}(t) = e^{bt} M_X(at)$.

Proof:

Multivariate Distributions

Extension to Multivariate Random Vectors

Recall that *joint pmf* of a discrete bivariate random vector, (X, Y), is

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

and *joint pdf* of a continuous bivariate random vector, (X, Y) is

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

Extension to Multivariate Random Vectors

Let $(X_1, X_2, \ldots, X_n)$ be a discrete multivariate random vector. The function

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

is called the *joint probability mass function* or *joint pmf* of the multivariate random vector.

Extension to Multivariate Random Vectors (cont.)

Let $(X_1, X_2, \ldots, X_n)$ be a continuous multivariate random vector. The function

$$f(x_1, x_2, \ldots, x_n)$$

is called the *joint probability density function* or *joint pdf* of the multivariate random vector.

Notes:

- The joint pmf and pdf provide a framework to study the behavior of multiple random variables simultaneously.
- These distributions capture the relationships and dependencies between variables.
- They extend naturally from bivariate to multivariate scenarios.

Generalization of Independence

- **Joint Distribution is the Product of the Marginals:**
  Given independent random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with each variable X_i having pmf or pdf $f_{X_i}(x_i)$ for $i = 1, \ldots, n$. The joint pmf or pdf of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is
  $$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \ldots f_{X_n}(x_n).$$
- **Checking Independence given the Joint Distribution:**
  Given random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with joint pmf or pdf $f(\mathbf{x}_1, \ldots, \mathbf{x}_n)$. $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent if and only if there are functions $g_i(\mathbf{x}_i)$, with each being a function of only $\mathbf{x}_i$, $i = 1, \ldots, n$, such that:
  $$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) = g_1(\mathbf{x}_1) \cdots g_n(\mathbf{x}_n).$$
- **Functions of Indep RVs are Indep.** Let random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent. For functions $g_i(\mathbf{x}_i)$, where each is a function of only $\mathbf{x}_i$, $i = 1, \ldots, n$, the random variables $U_i = g_i(\mathbf{X}_i)$, $i = 1, \ldots, n$, are also independent.

Expectation for Multivariate RVs

Expectation of Functions of Multiple RVs

Let (X, Y) be a bivariate random vector.

Question: How do we compute the expected value of a function $g(X, Y)$ of (X, Y)?

- Expectations of functions of random vectors are computed similarly to univariate random variables.
- For a real-valued function $g(x, y)$ defined for all possible values (x, y) of a discrete random vector (X, Y), its expected value $E[g(X, Y)]$ is:

$$E[g(X, Y)] = \sum_{(x,y)\in\mathbb{R}^2} g(x, y)f(x, y) = \sum_{(x,y)\in A} g(x, y)f(x, y)$$

where A is the support of $f(x, y)$.

Example: Two Dice Rolls

Recall the two dice roll example:

$$X = \text{sum of 2 dice}$$

$$Y = |\text{difference of 2 dice}|$$

Find $E[XY]$.

Solution:

For $g(x, y) = xy$, we have

$$\begin{aligned} E[g(X,Y)] &= E[XY] = \sum_{(x,y)\in\mathbb{R}^2} xyf(x,y) \\ &= (2)(0)\frac{1}{36} + \ldots + (7)(5)\frac{1}{18} \\ &= 13\frac{11}{18} \approx 13.61. \end{aligned}$$

Properties of Expectation

Notes:

- The marginal pmf of X or Y can be used to compute expectations that involve only X or Y.
- To compute an expectation that involves both X and Y, use the joint pmf of X and Y.

For continuous bivariate random vectors, the expected value is:

$$E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y)f(x,y)dxdy$$

The expectation continues to have properties listed in previous lectures when the random variable X is replaced by the random vector (X,Y).

Theorems on Expectation - I

(1) For random vector (X, Y):

$$E\left[ag(X,Y) + bh(X,Y) + c\right] = aE[g(X,Y)] + bE[h(X,Y)] + c$$

In particular, $E[X+Y] = E[X] + E[Y]$.

(2) If X and Y are **independent**:

$$\text{Var}\left(ag(X) + bh(Y) + c\right) = a^2\text{Var}(g(X)) + b^2\text{Var}(h(Y))$$

In particular,

$$\text{Var}\left(aX + bY + c\right) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

and

$$\text{Var}\left(X + Y\right) = \text{Var}(X) + \text{Var}(Y)$$

Theorems on Expectation - II

(3) Let X and Y be **independent** random variables. Let $g(x)$ be a function of x only and $h(y)$ be a function of y only. Then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

In particular,

$$E[XY] = E[X]E[Y].$$

Generalized Theorem for Multiple RVs:

Let $X_1, \ldots, X_n$ be independent random variables with functions $g_1, \ldots, g_n$ such that $g_i(x_i)$ is only a function of x_i, for $i = 1, \ldots, n$. Then,

$$E\left[g_1(X_1) \cdots g_n(X_n)\right] = E\left[g_1(X_1)\right] \cdots E\left[g_n(X_n)\right].$$

In particular,

$$E[X_1 X_2 \cdots X_n] = E[X_1]E[X_2] \cdots E[X_n].$$

Theorems on Expectation - III

(4) Let X and Y and be **independent** random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by

$$M_Z(t) = M_X(t)M_Y(t).$$

Generalized Theorem for Multiple RVs Let $X_1, \ldots, X_n$ be **independent** random variables with MGFs $M_{X_1}(t), \ldots, M_{X_n}(t)$. If $Z = X_1 + \cdots + X_n$, then:

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t).$$

In particular, if all $X_1, \ldots, X_n$ have the same distribution with MGF $M_X(t)$, then:

$$M_Z(t) = (M_X(t))^n.$$

Example: Sum of Two Independent Normals

Given $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent. Find the distribution of $Z = X + Y$.
Note that $M_U(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$

Solution (using Moment Generating Functions):

The MGF of X is: $M_X(t) = \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right)$ and
the MGF of Y is: $M_Y(t) = \exp\left(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right)$

Since X and Y are independent, the MGF of $Z = X + Y$ is the product of the MGFs of X and Y: $M_Z(t) = M_X(t) \cdot M_Y(t)$. So,

$$M_Z(t) = \exp\left((\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right)$$

Since the resulting MGF is of the form of a normal distribution, we can infer that $Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Extension: Multivariate Normals

Consider n independent normal random variables: $X_1, X_2, \ldots, X_n$ such that $X_i \sim N(\mu_i, \sigma_i^2)$ and are independent for $i = 1, 2, \ldots, n$.

The sum is given by:

$$S = X_1 + X_2 + \ldots + X_n$$

Using the properties of moment generating functions, we can derive:

$$M_S(t) = \prod_{i=1}^{n} \exp\left(\mu_i t + \frac{1}{2}\sigma_i^2 t^2\right) = \exp\left((\sum_{i=1}^{n} \mu_i)t + \frac{1}{2}(\sum_{i=1}^{n} \sigma_i^2)t^2\right)$$

Thus, the distribution of S is:

$$S \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Corollary: MGF of Linear Combinations

Given $X_1, \ldots, X_n$ as independent random variables with MGFs $M_{X_1}(t), \ldots, M_{X_n}(t)$ and constants $a_1, \ldots, a_n$, $b_1, \ldots, b_n$:

$$Z = (a_1 X_1 + b_1) + \cdots + (a_n X_n + b_n)$$

has MGF:

$$M_Z(t) = e^{t\left(\sum_{i=1}^{n} b_i\right)} M_{X_1}(a_1 t) \cdots M_{X_n}(a_n t).$$

Proof:

Recall that MGF of the random variable $aX + b$ is given by $M_{aX+b}(t) = e^{bt} M_X(at)$.

Example: Sum of two Independent Poisson variables

Given

- $X \sim \text{Poisson}(\lambda)$
- $Y \sim \text{Poisson}(\theta)$

are independent. Then, $X + Y \sim \text{Poisson}(\lambda + \theta)$.

Solution (MGF Method):
Note that the MGF of Poisson(λ) distribution is $e^{\lambda(e^t-1)}$. Let $Z = X + Y$, then

$$M_Z(t) = M_X(t)M_Y(t) = e^{\lambda(e^t-1)}e^{\theta(e^t-1)} = e^{(\lambda+\theta)(e^t-1)}$$

which is the MGF of Poisson($\lambda + \theta$) distribution. Hence, $X + Y \sim \text{Poisson}(\lambda + \theta)$.

MGF of a Sum of Poisson Variables

Suppose $X_1, \ldots, X_n$ are independent with $X_i \sim \text{Poisson}(\lambda_i)$. For $Z = X_1 + \cdots + X_n$, show that

$$Z \sim \text{Poisson}\left(\lambda_1 + \cdots + \lambda_n\right) \equiv \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Solution:

Using the expressions from above:

$$M_Z(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{\sum_{i=1}^{n} \lambda_i(e^t-1)} = e^{(\sum_{i=1}^{n} \lambda_i)(e^t-1)}$$

This is the MGF of a Poisson $\left(\sum_{i=1}^{n} \lambda_i\right)$ distribution. Thus,

$$Z \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Covariance and Correlation

Covariance and Correlation

We've discussed the absence or presence of a relationship between two random variables, such as independence. However, if there is a relationship, it may vary in strength. In this section, we discuss the covariance and correlation, which measure the strength of the relationship between two random variables.

Illustrative Example:

Consider two different experiments:

- X is the weight of water and Y is its volume. Data points will closely fall on a straight line.
- X is a human's weight and Y is height. We expect an upward trend in the plot, but data points do not necessarily fall on a straight line (instead will be scattered around a line).

Covariance and Correlation

For simplicity, denote

$$E[X] = \mu_X,\ E[Y] = \mu_Y,\ \text{Var}(X) = \sigma_X^2, \text{ and } \text{Var}(Y) = \sigma_Y^2.$$

Covariance of X and Y:

$$\text{Cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

Correlation of X and Y:

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Notes:

1. $-\infty < \text{Cov}(X, Y) < \infty$ and $-1 \leq \rho_{XY} \leq 1$.
2. Positive covariance implies small (large) values of X observed with small (large) values of Y.
3. Negative covariance implies small (large) values of X observed with large (small) values of Y.
4. $\rho_{XY} = -1$ or 1 implies perfect linear relationship.

Theorem: Expression for Covariance

For any random variables X and Y:

$$\text{Cov}(X, Y) = E[XY] - \mu_X \mu_Y$$

Proof:

- Recall the definition of covariance:

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

- Expanding inside the expectation:

$$= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]$$

- $= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y$ (expectation is linear)
- Now, $E[Y] = \mu_Y$ and $E[X] = \mu_X$, so:

$$= E[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y$$

- Simplifying: $\text{Cov}(X, Y) = E[XY] - \mu_X \mu_Y$

Example: Bivariate Random Variables

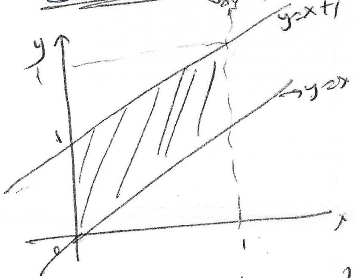
Given the joint pdf:

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

1. Find marginal pdfs.
2. Find means and variances.
3. Find covariance and correlation.

Soln:

$f_{XY}(x,y) = 1$ for $0 < x < 1$, $x < y < x+1$



(1) $f_X(x) = \int_x^{x+1} f_{XY}(x,y)\,dy = \int_x^{x+1} dy = 1$ $\quad 0<x<1$

i.e. $X \sim \mathrm{UNIF}(0,1)$

for $0 < y < 1$: $f_Y(y) = \int_0^y dx = y$

for $1 \le y < 2$: $f_Y(y) = \int_{y-1}^1 dx = 1-(y-1) = 2-y$

$$\Rightarrow f_Y(y) = \begin{cases} y & 0 \le y < 1 \\ 2-y & 1 \le y < 2 \end{cases}$$

(2) $EX = \mu_X = \frac{1}{2}$ $\qquad \sigma_X^2 = \mathrm{Var}\,X = \frac{1}{12}$

$\mu_Y = EY = 1$ $\qquad \sigma_Y^2 = \mathrm{Var}\,Y = \frac{1}{6}$

(3) $EXY = \int_0^1 \int_x^{x+1} xy\,dy\,dx = \int_0^1 \frac{xy^2}{2}\Big|_x^{x+1} dx = \int_0^1 (x^2 + \frac{1}{2}x)\,dx = \frac{7}{12}$

so $\mathrm{Cov}(X,Y) = \frac{7}{12} - (\frac{1}{2})1 = \frac{1}{12}$

$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{1/12}{\sqrt{1/12}\cdot\sqrt{1/6}} = \frac{1}{\sqrt{2}} \approx 0.707$

Properties of Covariance and Correlation

Covariance and Independence: If X and Y are independent:

$$\text{Cov}(X, Y) = 0 \text{ and } \rho_{XY} = 0$$

However, zero covariance does not imply independence.
Covariance only measures linear relationship!

Variances and Covariances: For random variables X and Y and constants a, b:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

If X and Y are indep.: $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$

Correlation: For any random variables X and Y:

1. $-1 \leq \rho_{XY} \leq 1$
2. $|\rho_{XY}| = 1$ if and only if a linear relationship exists between X and Y.

Example: Covariance and Correlation

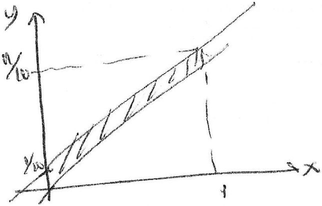
Given the joint pdf:

$$f_{X,Y}(x, y) = 10, \quad 0 < x < 1, x < y < x + 1/10$$

Compute $\text{Cov}(X, Y)$ and ρ_{XY}.

Sol'n:

(X,Y) have joint pdf $f_{X,Y}(x,y) = 10$ for $0<x<1$, $x<y<x+1/10$



$EX = \frac{1}{2}$ $\quad EY = E(X+Z) = \frac{1}{2} + \frac{1}{20} = \frac{11}{20}$

$\Rightarrow Cov(X,Y) = EXY - EXEY$ $\quad \hookrightarrow Y = X+Z$

$= E(X(X+Z)) - EX\,E(X+Z)$

$= EX^2 + EXZ - EX(EX + EZ)$

$= EX^2 + EXZ - (EX)^2 - EXEZ$

$= EX^2 + EXZ - (EX)^2$

$= Var(X) = 1/12$

$Var(Y) = Var(X+Z) = Var(X) + Var(Z) = \frac{1}{12} + \frac{1}{1200}$

Thus $\rho_{XY} = \dfrac{1/12}{\sqrt{1/12}\sqrt{\frac{1}{12}+\frac{1}{1200}}} = \sqrt{\dfrac{100}{101}} \approx 0.995$

Inequalities in Statistical Theory

Inequalities in Statistical Theory

Statistical theory is abundant with inequalities and identities.

Markov's Inequality:

Given a random variable X and a nonnegative function $g(x)$, for any $t > 0$:

$$P(g(X) \geq t) \leq \frac{E[g(X)]}{t}.$$

Chebyshev's Inequality (Special Case of Markov Ineq.):

Using Markov's Inequality with $g(x) = (x - \mu)^2/\sigma^2$:

$$P\left(|X - \mu| \geq t\sigma\right) \leq \frac{1}{t^2}$$

equivalently
$$P\left(|X - \mu| \leq t\sigma\right) \geq 1 - \frac{1}{t^2}$$

E.g. for $t = 3$, the probability that any random variable is within 3 standard deviations of its mean is at least 88.89%.

Appendix: Additional Inequalities

Cauchy-Schwarz Inequality

For any two random variables U and V:

$$|E[UV]| \leq \sqrt{E[U^2]E[V^2]},$$

where equality holds if $V = aU$ for some real number a.

Covariance Inequality:
Given random variables X and Y with variances σ_X^2 and σ_Y^2:

$$|\mathrm{Cov}(X, Y)| \leq \sigma_X \sigma_Y \quad \text{and} \quad |\mathrm{Corr}(X, Y)| = |\rho_{XY}| \leq 1,$$

with equality if $P(Y = aX + b) = 1$ for real numbers a and b.

Jensen's Inequality

Convex Functions: A function $g(x)$ is *convex* if, for all values x and y in its domain and any $t \in [0, 1]$:

$$g(tx + (1 - t)y) \leq tg(x) + (1 - t)g(y).$$

A function is *concave* if $-g(x)$ is convex.

Jensen's Inequality: For a random variable X with expected value $E[X]$, if $g(x)$ is a convex function, then:

$$E[g(X)] \geq g(E[X]).$$

Applications of Jensen's Inequality:

- For X with $\text{Var}(X)$, $\text{Var}(X) \geq 0$ by setting $g(x) = x^2$ in Jensen's inequality.
- For any random variable X, $E[\log X] \leq \log E[X]$.