

Lecture Notes: Independence, Covariance, and Correlation

From a Lecture Transcription

May 6, 2025

Abstract

These notes cover the concept of independence for events and random variables, explore equivalent conditions for the independence of random variables, and delve into measures of linear dependence such as covariance and correlation, culminating with the Cauchy-Schwarz inequality. Administrative details regarding an upcoming quiz are also included.

Contents

1	Independence of Events	2
1.1	Independence of Two Events	2
1.2	Independence of Multiple Events	3
2	Independence of Random Variables	4
2.1	Definition via Events	4
2.2	Equivalent Conditions for Independence (Lemma 2.12 in lecture)	4
2.3	Functions of Independent Random Variables	5
2.4	Examples of Independence/Dependence	6
3	Linear Independence (Uncorrelatedness)	7
3.1	Functional Dependence vs. Probabilistic Independence	7
3.2	Covariance and Correlation	8
3.3	Cauchy-Schwarz Inequality	8
3.4	Example: Uncorrelated but Dependent Variables	10

Administrative Announcements

- **Upcoming Quiz Details:**
 - **Duration:** 1.5 hours.
 - **Time:** 5:00 PM - 6:30 PM.
 - **Structure:** Approximately 3 multi-part questions. There is no choice of questions.
 - **Bonus:** There may be a small bonus component (around 5 points).
 - **Material Covered:** Chapter 2, up to and including the material from this lecture (specifically, excluding "Exponential Families" which will be covered next week but not on this quiz).
 - **Formula Sheet:** A formula sheet will be provided. This sheet will primarily contain theorems. Definitions are generally not included, as understanding and recalling definitions is considered part of the expected knowledge.
 - You may request specific additions to the formula sheet (e.g., Taylor series expansion).
 - **Definitions in Questions:** Questions may begin by asking for a relevant definition. This helps ensure a common understanding and can contribute to your score.
 - **Instructor Availability for Quiz:** The main instructor will not be physically present during the quiz (due to travel) but will be reachable by phone. A knowledgeable teaching assistant or student proctor will be present to oversee the quiz and answer logistical questions.
 - **Homework:** An exercise set covering the material from this lecture will be provided.
- **General Notes:**
 - Please feel free to ask questions about the material at any time.
 - The lecture acknowledges the challenge of a 2.5-hour session at this time of day.

1 Independence of Events

We begin our journey into the concept of independence, a cornerstone of probability theory. You've likely encountered this in a basic course, but we'll revisit and extend it.

1.1 Independence of Two Events

Definition 1.1 (Independence of Two Events). *Two events A and B are said to be **independent** if the occurrence of one event does not affect the probability of the occurrence of the other. Mathematically, this is expressed as:*

$$P(A \cap B) = P(A)P(B) \tag{1}$$

Remark 1.2 (Intuition from Conditional Probability). *The definition $P(A \cap B) = P(A)P(B)$ might seem a bit arbitrary at first. Its intuition becomes clearer when we consider conditional probability. Recall that the conditional probability of event A occurring given that event B has occurred is:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0 \quad (2)$$

If A and B are truly independent, then knowing B occurred should not change the probability of A . That is, we'd expect:

$$P(A|B) = P(A) \quad (3)$$

Substituting this into the conditional probability formula:

$$\begin{aligned} P(A) &= \frac{P(A \cap B)}{P(B)} \\ \implies P(A)P(B) &= P(A \cap B) \end{aligned}$$

This gives us the familiar definition of independence. Similarly, if A and B are independent, then $P(B|A) = P(B)$ (assuming $P(A) > 0$). The statement "the fact that B occurred does not affect the probability of A " is precisely what $P(A|B) = P(A)$ signifies.

1.2 Independence of Multiple Events

The concept of independence can be extended from two events to any finite collection of events.

Definition 1.3 (Independence of n Events (Definition 2.12 in lecture)). *Let A_1, A_2, \dots, A_n be a finite collection of events. These events are said to be **mutually independent** if for every subcollection of these events, say $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ (where $1 \leq i_1 < i_2 < \dots < i_k \leq n$ and $k \geq 2$), the probability of their intersection is the product of their individual probabilities. More formally, for any subset of indices $J \subseteq \{1, 2, \dots, n\}$ with $|J| \geq 2$,*

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \quad (4)$$

For the entire collection to be independent, this must hold for all $2^n - n - 1$ such subcollections (all subsets of size 2 or more). A common compact way to state the full requirement is that for any non-empty finite index set $J \subseteq \{1, \dots, n\}$,

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i). \quad (5)$$

This covers all cases, including single events (trivially) and the full set.

Remark 1.4 (Pairwise vs. Mutual Independence). *Mutual independence is a stronger condition than pairwise independence.*

- **Mutual Independence implies Pairwise Independence:** *If a collection of events A_1, \dots, A_n is mutually independent, then any pair A_i, A_j (for $i \neq j$) is also independent.*
- **Pairwise Independence does NOT imply Mutual Independence:** *It is possible for events to be independent in pairs, yet not mutually independent as a whole collection. For instance, $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, and $P(B \cap C) = P(B)P(C)$ does not guarantee that $P(A \cap B \cap C) = P(A)P(B)P(C)$.*

Our definition above for n events is for mutual independence.

2 Independence of Random Variables

We now extend the idea of independence from events to random variables. A random variable X (mapping outcomes $\omega \in \Omega$ to real numbers) generates events. For example, for any set $\Gamma \subseteq \mathbb{R}$, the set of outcomes $\{\omega \in \Omega \mid X(\omega) \in \Gamma\}$ is an event. We denote this event concisely as $\{X \in \Gamma\}$.

2.1 Definition via Events

Definition 2.1 (Independence of Random Variables (Definition 2.1.11 in lecture)). *Let X_1, X_2, \dots, X_n be random variables. They are said to be **independent** if for every sequence of (Borel) sets $\Gamma_1, \Gamma_2, \dots, \Gamma_n \subseteq \mathbb{R}$, the events $\{X_1 \in \Gamma_1\}, \{X_2 \in \Gamma_2\}, \dots, \{X_n \in \Gamma_n\}$ are mutually independent. That is,*

$$P(X_1 \in \Gamma_1, X_2 \in \Gamma_2, \dots, X_n \in \Gamma_n) = \prod_{i=1}^n P(X_i \in \Gamma_i) \quad (6)$$

for all such sets Γ_i .

While this definition is fundamental, checking it for all possible sets Γ_i is impractical. Fortunately, there are equivalent, more workable conditions.

2.2 Equivalent Conditions for Independence (Lemma 2.12 in lecture)

Lemma 2.2 (Equivalent Conditions for Independence of Random Variables). *Random variables X_1, X_2, \dots, X_n are independent if and only if any one of the following conditions holds:*

1. **Product of Expectations of Functions:** For all (Borel-measurable) functions $h_1, h_2, \dots, h_n : \mathbb{R} \rightarrow \mathbb{R}$ such that the expectations exist:

$$E \left[\prod_{j=1}^n h_j(X_j) \right] = \prod_{j=1}^n E[h_j(X_j)] \quad (7)$$

(This is often stated as: the expectation of the product is the product of the expectations, provided the random variables are independent and the expectations are finite).

2. **Product of Cumulative Distribution Functions (CDFs):** For all $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{j=1}^n F_{X_j}(x_j) \quad (8)$$

where F_{X_1, \dots, X_n} is the joint CDF and F_{X_j} is the marginal CDF of X_j . ($F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$ and $F_{X_j}(x_j) = P(X_j \leq x_j)$).

3. **Product of Probability Density Functions (PDFs) (for continuous RVs):** If the random variables X_1, \dots, X_n are jointly continuous with joint PDF $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ and marginal PDFs $f_{X_j}(x_j)$, then they are independent if and only if:

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{j=1}^n f_{X_j}(x_j) \quad (9)$$

for all $(x_1, \dots, x_n) \in \mathbb{R}^n$ (almost everywhere).

4. **Product of Probability Mass Functions (PMFs) (for discrete RVs):** If the random variables X_1, \dots, X_n are discrete with joint PMF $P_{X_1, \dots, X_n}(x_1, \dots, x_n)$ and marginal PMFs $P_{X_j}(x_j)$, then they are independent if and only if:

$$P_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{j=1}^n P_{X_j}(x_j) \quad (10)$$

for all (x_1, \dots, x_n) in the support.

Remark 2.3 (Further Conditions - For Information). The lemma can be extended to include other conditions, though we won't delve into their proofs here:

- **Moment Generating Functions (MGFs):** If the joint MGF $M_{X_1, \dots, X_n}(t_1, \dots, t_n)$ and marginal MGFs $M_{X_j}(t_j)$ exist in a neighborhood of the origin, then independence is equivalent to:

$$M_{X_1, \dots, X_n}(t_1, \dots, t_n) = \prod_{j=1}^n M_{X_j}(t_j) \quad (11)$$

- **Characteristic Functions (CFs):** Characteristic functions $\phi_{X_j}(t_j) = E[e^{it_j X_j}]$ always exist. Independence is equivalent to:

$$\phi_{X_1, \dots, X_n}(t_1, \dots, t_n) = \prod_{j=1}^n \phi_{X_j}(t_j) \quad (12)$$

These are powerful tools, especially the characteristic function, which always uniquely determines a distribution.

2.3 Functions of Independent Random Variables

A very useful consequence of independence is that functions of independent random variables are also independent.

Corollary 2.4 (Functions of Independent Random Variables (Corollary 2.13 in lecture)). Let X_1, X_2, \dots, X_n be independent random variables. Let g_1, g_2, \dots, g_n be (Borel-measurable) functions, where each $g_j : \mathbb{R} \rightarrow \mathbb{R}$. Define new random variables $Y_j = g_j(X_j)$ for $j = 1, \dots, n$. Then the random variables Y_1, Y_2, \dots, Y_n are also independent.

Proof. We want to show that Y_1, \dots, Y_n are independent. We can use condition (1) from Lemma 2.2. Let h_1, \dots, h_n be arbitrary (Borel-measurable) functions such that the expectations exist. Consider:

$$\begin{aligned} E \left[\prod_{j=1}^n h_j(Y_j) \right] &= E \left[\prod_{j=1}^n h_j(g_j(X_j)) \right] \\ &= E \left[\prod_{j=1}^n (h_j \circ g_j)(X_j) \right] \end{aligned}$$

Let $k_j = h_j \circ g_j$. Then k_j is a function of X_j . Since X_1, \dots, X_n are independent, by Lemma 2.2(1):

$$\begin{aligned} E \left[\prod_{j=1}^n k_j(X_j) \right] &= \prod_{j=1}^n E[k_j(X_j)] \\ &= \prod_{j=1}^n E[(h_j \circ g_j)(X_j)] \\ &= \prod_{j=1}^n E[h_j(g_j(X_j))] \\ &= \prod_{j=1}^n E[h_j(Y_j)] \end{aligned}$$

Since $E \left[\prod_{j=1}^n h_j(Y_j) \right] = \prod_{j=1}^n E[h_j(Y_j)]$ for arbitrary functions h_j , the random variables Y_1, \dots, Y_n are independent by Lemma 2.2(1). ■

2.4 Examples of Independence/Dependence

Example 2.5 (Independent Bernoulli Variables from Joint PMF (Example 2.14 in lecture)). *Let X_1, \dots, X_n be random variables, each taking values in $\{0, 1\}$. Suppose their joint probability mass function (PMF) is given by:*

$$P(X_1 = x_1, \dots, X_n = x_n) = p^{\sum_{i=1}^n x_i} (1-p)^{n - \sum_{i=1}^n x_i} \quad (13)$$

for $x_i \in \{0, 1\}$ and $0 < p < 1$. We want to show that X_1, \dots, X_n are independent.

We can rewrite the joint PMF as:

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= p^{x_1 + x_2 + \dots + x_n} (1-p)^{(1-x_1) + (1-x_2) + \dots + (1-x_n)} \\ &= (p^{x_1} (1-p)^{1-x_1}) (p^{x_2} (1-p)^{1-x_2}) \dots (p^{x_n} (1-p)^{1-x_n}) \\ &= \prod_{j=1}^n (p^{x_j} (1-p)^{1-x_j}) \end{aligned}$$

Now, consider the marginal PMF for any X_j . For $X_j = x_j$ (where $x_j \in \{0, 1\}$), its PMF is $P(X_j = x_j) = p^{x_j} (1-p)^{1-x_j}$. This is the PMF of a Bernoulli random variable with success probability p , i.e., $X_j \sim \text{Bernoulli}(p)$.

Since the joint PMF is the product of the marginal PMFs:

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{j=1}^n P(X_j = x_j) \quad (14)$$

By Lemma 2.2(4), the random variables X_1, \dots, X_n are independent. Moreover, they are identically distributed as $\text{Bernoulli}(p)$.

Example 2.6 (Proving Dependence (Original lecture example)). *Let $X \sim U(0, 1)$ (uniform distribution on the interval $(0, 1)$). Define a new random variable $Y = I(X < 1/2)$, where $I(\cdot)$ is the indicator function. So, $Y = 1$ if $X < 1/2$ and $Y = 0$ if $X \geq 1/2$. We want to show that X and Y are dependent.*

We will use Lemma 2.2(1) and check if $E[XY] = E[X]E[Y]$. First, let's find the individual expectations:

- $E[X]$: Since $X \sim U(0, 1)$, its PDF is $f_X(x) = 1$ for $0 < x < 1$, and 0 otherwise.

$$E[X] = \int_0^1 x \cdot 1 \, dx = \left[\frac{x^2}{2} \right]_0^1 = \frac{1}{2} \quad (15)$$

- $E[Y]$: Since Y is an indicator variable, $E[Y] = P(Y = 1) = P(X < 1/2)$.

$$P(X < 1/2) = \int_0^{1/2} 1 \, dx = [x]_0^{1/2} = \frac{1}{2} \quad (16)$$

So, $E[Y] = 1/2$.

Therefore, $E[X]E[Y] = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$.

Next, let's find $E[XY]$:

$$\begin{aligned} E[XY] &= E[X \cdot I(X < 1/2)] \\ &= \int_{-\infty}^{\infty} x \cdot I(x < 1/2) f_X(x) \, dx \\ &= \int_0^1 x \cdot I(x < 1/2) \cdot 1 \, dx \end{aligned}$$

The term $I(x < 1/2)$ is 1 if $x < 1/2$ and 0 otherwise. So the integral becomes:

$$\begin{aligned} E[XY] &= \int_0^{1/2} x \cdot 1 \cdot 1 \, dx + \int_{1/2}^1 x \cdot 0 \cdot 1 \, dx \\ &= \int_0^{1/2} x \, dx = \left[\frac{x^2}{2} \right]_0^{1/2} = \frac{(1/2)^2}{2} - 0 = \frac{1}{8} \end{aligned}$$

We have $E[XY] = 1/8$ and $E[X]E[Y] = 1/4$. Since $E[XY] = 1/8 \neq E[X]E[Y] = 1/4$, the random variables X and Y are **dependent** by Lemma 2.2(1). This makes intuitive sense: the value of Y is completely determined by X .

3 Linear Independence (Uncorrelatedness)

While general independence is a strong condition, sometimes we are interested in a weaker form of non-association, specifically linear association.

3.1 Functional Dependence vs. Probabilistic Independence

Consider a random variable X and another random variable $Y = g(X)$ for some function g . Intuitively, X and Y are functionally dependent. When are they probabilistically independent?

Proposition 3.1. *Let X be a random variable and $Y = g(X)$ for some function g . Then X and Y are independent if and only if $g(X)$ is a constant almost surely (i.e., Y is a constant random variable).*

Proof. Suppose X and $Y = g(X)$ are independent. By Lemma 2.2(1), for any suitable functions h_1, h_2 , we have $E[h_1(X)h_2(Y)] = E[h_1(X)]E[h_2(Y)]$. The argument from the lecture was as follows: If X and $Y = g(X)$ are independent, then taking $h_1(u) = g(u)$ and $h_2(v) = v$ would imply

$E[g(X) \cdot Y] = E[g(X)]E[Y]$. Substituting $Y = g(X)$, we get $E[g(X) \cdot g(X)] = E[g(X)]E[g(X)]$. So, $E[(g(X))^2] = (E[g(X)])^2$. This implies $E[(g(X))^2] - (E[g(X)])^2 = 0$. This means $\text{Var}(g(X)) = 0$. A random variable with zero variance must be equal to its mean (a constant) almost surely. Thus, $g(X) = E[g(X)] = c$ for some constant c , almost surely. So Y is a constant.

Conversely, if $Y = g(X) = c$ (a constant), then Y is independent of any random variable X . To see this, for any Borel sets $A, B \subseteq \mathbb{R}$: $P(X \in A, Y \in B) = P(X \in A, c \in B)$. If $c \in B$, then $P(X \in A, c \in B) = P(X \in A)$. Also, $P(Y \in B) = P(c \in B) = 1$. So $P(X \in A) = P(X \in A) \cdot 1 = P(X \in A)P(Y \in B)$. If $c \notin B$, then $P(X \in A, c \in B) = P(\emptyset) = 0$. Also, $P(Y \in B) = P(c \in B) = 0$. So $0 = P(X \in A) \cdot 0 = P(X \in A)P(Y \in B)$. In both cases, $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$, so independence holds. ■

This means that if Y is a non-trivial function of X , then X and Y are dependent.

3.2 Covariance and Correlation

Definition 3.2 (Covariance). The **covariance** between two random variables X and Y is defined as:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \quad (17)$$

An alternative, often computationally simpler, formula is $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$.

Covariance measures the degree to which X and Y tend to vary together.

- $\text{Cov}(X, Y) > 0$: X and Y tend to increase or decrease together.
- $\text{Cov}(X, Y) < 0$: When X increases, Y tends to decrease, and vice-versa.
- $\text{Cov}(X, Y) = 0$: X and Y are said to be **uncorrelated**. This implies no linear relationship.

Remark 3.3. If X and Y are independent, then $E[XY] = E[X]E[Y]$ (assuming expectations exist). Thus, if X and Y are independent, $\text{Cov}(X, Y) = E[X]E[Y] - E[X]E[Y] = 0$. So, **independent random variables are always uncorrelated**. However, the converse is not true: **uncorrelated random variables are not necessarily independent**. (We will see an example soon).

The magnitude of covariance depends on the scale of X and Y . To get a scale-free measure, we use the correlation coefficient.

Definition 3.4 (Correlation Coefficient). The **correlation coefficient** (or *Pearson correlation coefficient*) between X and Y is:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \quad (18)$$

provided $\text{Var}(X) > 0$ and $\text{Var}(Y) > 0$. If $\text{Var}(X) = 0$ or $\text{Var}(Y) = 0$, ρ is typically undefined or taken as 0.

3.3 Cauchy-Schwarz Inequality

The correlation coefficient $\rho(X, Y)$ is bounded between -1 and 1. This is a consequence of the Cauchy-Schwarz inequality.

Theorem 3.5 (Cauchy-Schwarz Inequality (Theorem 2.17 in lecture)). *Let X and Y be random variables with finite second moments (i.e., $E[X^2] < \infty$ and $E[Y^2] < \infty$). Then:*

$$(E[XY])^2 \leq E[X^2]E[Y^2] \quad (19)$$

Equality holds if and only if $Y = cX$ for some constant c almost surely, or $X = 0$ a.s., or $Y = 0$ a.s. (i.e., X and Y are linearly dependent in this sense, or one is trivial). More precisely, equality holds if $P(Y = cX) = 1$ for some constant c , or $P(X = 0) = 1$, or $P(Y = 0) = 1$.

Proof. Case 1: $E[Y^2] = 0$. This implies $Y = 0$ almost surely (since $Y^2 \geq 0$). Then $E[XY] = E[X \cdot 0] = 0$. The inequality becomes $0^2 \leq E[X^2] \cdot 0$, which is $0 \leq 0$. So it holds, and equality holds. In this case $Y = 0 \cdot X$, so $c = 0$.

Case 2: $E[Y^2] > 0$. Consider the function $h(a) = E[(X - aY)^2]$ for any real number a . Since $(X - aY)^2 \geq 0$, we must have $h(a) = E[(X - aY)^2] \geq 0$ for all a . Expanding $h(a)$:

$$\begin{aligned} h(a) &= E[X^2 - 2aXY + a^2Y^2] \\ &= E[X^2] - 2aE[XY] + a^2E[Y^2] \end{aligned}$$

This is a quadratic function of a , of the form $Aa^2 + Ba + C$ where $A = E[Y^2]$, $B = -2E[XY]$, and $C = E[X^2]$. Since $A = E[Y^2] > 0$, this parabola opens upwards. Since $h(a) \geq 0$ for all a , the parabola must have at most one real root. This means its discriminant must be less than or equal to zero: $B^2 - 4AC \leq 0$.

$$\begin{aligned} (-2E[XY])^2 - 4(E[Y^2])(E[X^2]) &\leq 0 \\ 4(E[XY])^2 - 4E[X^2]E[Y^2] &\leq 0 \\ (E[XY])^2 - E[X^2]E[Y^2] &\leq 0 \\ \implies (E[XY])^2 &\leq E[X^2]E[Y^2] \end{aligned}$$

This proves the inequality.

Equality holds if and only if the discriminant is zero, meaning $h(a)$ has exactly one real root, say a^* . This root is $a^* = \frac{-B}{2A} = \frac{2E[XY]}{2E[Y^2]} = \frac{E[XY]}{E[Y^2]}$. At this root, $h(a^*) = E[(X - a^*Y)^2] = 0$. Since $(X - a^*Y)^2$ is a non-negative random variable, if its expectation is zero, then $X - a^*Y = 0$ almost surely. So, $X = a^*Y$ almost surely. If $a^* \neq 0$, let $c = 1/a^*$. Then $Y = cX$ a.s. If $a^* = 0$, then $E[XY] = 0$. If $E[Y^2] > 0$, $X = 0 \cdot Y = 0$ a.s. Thus, equality holds if and only if one random variable is a scalar multiple of the other almost surely (or one is zero a.s.). ■

Corollary 3.6 (Bounds for Correlation Coefficient). *For any random variables X, Y with $\text{Var}(X) > 0$, $\text{Var}(Y) > 0$:*

$$-1 \leq \rho(X, Y) \leq 1 \quad (20)$$

Furthermore, $\rho(X, Y) = \pm 1$ if and only if $Y = aX + b$ for some constants $a \neq 0, b$, almost surely (i.e., X and Y have a perfect linear relationship).

Proof. Apply the Cauchy-Schwarz inequality to the centered random variables $X' = X - E[X]$ and $Y' = Y - E[Y]$. We have $E[X'] = 0$ and $E[Y'] = 0$. Then $E[X'^2] = \text{Var}(X)$ and $E[Y'^2] = \text{Var}(Y)$. And $E[X'Y'] = E[(X - E[X])(Y - E[Y])] = \text{Cov}(X, Y)$. The Cauchy-Schwarz inequality states $(E[X'Y'])^2 \leq E[X'^2]E[Y'^2]$. So, $(\text{Cov}(X, Y))^2 \leq \text{Var}(X)\text{Var}(Y)$. Taking the square root (and noting $\sqrt{\text{Var}(X)\text{Var}(Y)} \geq 0$):

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)} \quad (21)$$

Dividing by $\sqrt{\text{Var}(X)\text{Var}(Y)}$ (which is positive):

$$\frac{|\text{Cov}(X, Y)|}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \leq 1 \quad (22)$$

This means $|\rho(X, Y)| \leq 1$, which is equivalent to $-1 \leq \rho(X, Y) \leq 1$. Equality $\rho(X, Y) = \pm 1$ holds if and only if equality holds in the Cauchy-Schwarz inequality for X' and Y' , i.e., $Y' = cX'$ a.s. for some constant c . $Y - E[Y] = c(X - E[X])$. $Y = cX - cE[X] + E[Y]$. This is of the form $Y = aX + b$ where $a = c$ and $b = E[Y] - cE[X]$. If $c = 0$, then $Y' = 0$ a.s., so $Y = E[Y]$ a.s., meaning $\text{Var}(Y) = 0$, which contradicts our assumption that $\text{Var}(Y) > 0$. So $c \neq 0$, hence $a \neq 0$. ■

3.4 Example: Uncorrelated but Dependent Variables

Example 3.7 (Uncorrelated but Dependent (Example 2.18 in lecture)). Let $X \sim U(-1, 1)$. Define:

- $Y = X \cdot I(|X| < 1/2)$
- $Z = X \cdot I(|X| \geq 1/2)$

Here $I(\cdot)$ is the indicator function. Essentially, $Y = X$ if $-1/2 < X < 1/2$ and $Y = 0$ otherwise. And $Z = X$ if $X \in [-1, -1/2] \cup [1/2, 1]$ and $Z = 0$ otherwise. Notice that for any given outcome of X , it is never the case that both Y and Z are non-zero simultaneously (unless $X = 0$, in which case $Y = Z = 0$). If $X \neq 0$ and $|X| < 1/2$, then $Y = X$ and $Z = 0$. If $X \neq 0$ and $|X| \geq 1/2$, then $Y = 0$ and $Z = X$. If $X = 0$, then $Y = 0$ and $Z = 0$. In all cases, $Y \cdot Z = 0$. Therefore, $E[Y \cdot Z] = E[0] = 0$.

Now let's find $E[Y]$ and $E[Z]$. Since $X \sim U(-1, 1)$, its PDF is $f_X(x) = 1/2$ for $-1 < x < 1$.

$$\begin{aligned} E[Y] &= E[X \cdot I(|X| < 1/2)] = \int_{-1}^1 x \cdot I(|x| < 1/2) f_X(x) dx \\ &= \int_{-1/2}^{1/2} x \cdot \frac{1}{2} dx = \frac{1}{2} \left[\frac{x^2}{2} \right]_{-1/2}^{1/2} = \frac{1}{4} ((1/2)^2 - (-1/2)^2) = 0 \end{aligned}$$

Similarly,

$$\begin{aligned} E[Z] &= E[X \cdot I(|X| \geq 1/2)] = \int_{-1}^1 x \cdot I(|x| \geq 1/2) f_X(x) dx \\ &= \int_{-1}^{-1/2} x \cdot \frac{1}{2} dx + \int_{1/2}^1 x \cdot \frac{1}{2} dx \\ &= \frac{1}{2} \left(\left[\frac{x^2}{2} \right]_{-1}^{-1/2} + \left[\frac{x^2}{2} \right]_{1/2}^1 \right) \\ &= \frac{1}{4} (((-1/2)^2 - (-1)^2) + (1^2 - (1/2)^2)) \\ &= \frac{1}{4} ((1/4 - 1) + (1 - 1/4)) = \frac{1}{4} (-3/4 + 3/4) = 0 \end{aligned}$$

(Alternatively, note that X is symmetric about 0, and the function $x \cdot I(|x| < 1/2)$ is odd, as is $x \cdot I(|x| \geq 1/2)$. So $E[Y] = 0$ and $E[Z] = 0$ by symmetry, since the integral of an odd function over a symmetric interval around 0 is 0.)

Now, the covariance:

$$\text{Cov}(Y, Z) = E[YZ] - E[Y]E[Z] = 0 - (0)(0) = 0 \quad (23)$$

So, Y and Z are uncorrelated.

Are they independent? No. If they were independent, then for example, $P(Y \neq 0, Z \neq 0)$ should equal $P(Y \neq 0)P(Z \neq 0)$. As we established, $Y \cdot Z = 0$, which means it's impossible for both Y and Z to be non-zero simultaneously. So, $P(Y \neq 0, Z \neq 0) = 0$.

However,

$$\begin{aligned} P(Y \neq 0) &= P(-1/2 < X < 1/2 \text{ and } X \neq 0) \\ &= P(-1/2 < X < 1/2) \quad (\text{since } P(X = 0) = 0 \text{ for a continuous RV}) \\ &= \int_{-1/2}^{1/2} \frac{1}{2} dx = \frac{1}{2}[x]_{-1/2}^{1/2} = \frac{1}{2}(1/2 - (-1/2)) = \frac{1}{2} \cdot 1 = 1/2 \end{aligned}$$

And

$$\begin{aligned} P(Z \neq 0) &= P(|X| \geq 1/2) \\ &= P(X \in [-1, -1/2] \cup [1/2, 1]) \\ &= \left(\int_{-1}^{-1/2} \frac{1}{2} dx \right) + \left(\int_{1/2}^1 \frac{1}{2} dx \right) \\ &= \frac{1}{2}((-1/2 - (-1)) + (1 - 1/2)) = \frac{1}{2}(1/2 + 1/2) = 1/2 \end{aligned}$$

So, $P(Y \neq 0)P(Z \neq 0) = (1/2)(1/2) = 1/4$. Since $P(Y \neq 0, Z \neq 0) = 0 \neq P(Y \neq 0)P(Z \neq 0) = 1/4$, the random variables Y and Z are **dependent**.

This example clearly illustrates that uncorrelatedness does not imply independence. Y and Z have a strong dependence (if one is non-zero, the other must be zero, given $X \neq 0$), yet their linear association (covariance) is zero.

Remark 3.8 (Reading Material Mentioned in Lecture). The lecture mentioned that further material on linear prediction (e.g., finding a, b to minimize $E[(X - (aY + b))^2]$, the mean squared error) would be provided as reading material. This topic often involves concepts like the best linear predictor and its connection to covariance and variance, which builds upon the ideas discussed here.