9: Independent Random Variables, Covariance, and Correlation

PSTAT 120A: Summer 2022

Ethan P. Marzban

July 14, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions
- Transformations of Random Variables
- Double Integrals
- Random Vectors and the basics of multivariate probability

Independence

Definition

Definition: Independence (of 2 Random Variables)

Given two random variables X and Y with marginal p.d.f.'s given by $f_X(x)$ and $f_Y(y)$, respectively, and joint p.d.f. $f_{X,Y}(x,y)$, we say that X and Y are **independent** (notated $X \perp Y$) if

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

In other words, two random variables are independent if their joint density factors as the product of their marginal densities.

- It turns out that an equivalent definition of independence is that the joint c.d.f. factors as the product of the marginal c.d.f.'s.

Definition

Definition: Independence (of n Random Variables)

Consider a collection of n random variables $X_1, \cdots, X_n$ with joint p.d.f. $f_{\vec{X}}(\vec{x})$ and marginal densities $f_{X_i}(x_i)$ for $i = 1, \cdots, n$.

(1) If $f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$, then $X_1, \cdots, X_n$ are independent.

(2) Conversely, if $X_1, \cdots, X_n$ are independent, then they are jointly continuous with joint density function $f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$.

Example

Consider a pair (X, Y) of discrete random variables with joint p.m.f. given by

		Y = 1	2	3	4
X	0	0.1	0.1	0.1	0
	1	0	0.2	0.1	0.1
	2	0.1	0.1	0	0.1

(a) Find the marginal p.m.f.'s $p_X(x)$ and $p_Y(y)$ of X and Y respectively.

(b) Compute $\mathbb{E}[XY]$.

(c) Are X and Y independent? Explain.

A Familiar Example

Suppose (X, Y) is a pair of random variables with joint density given by

$$f_{X,Y}(x,y) = \begin{cases} 2 \cdot e^{-(x+y)} & \text{if } x \leq y < \infty,\ 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Are X and Y independent? Explain.

Shortcut for Establishing Dependence

- There exists a shortcut for determining dependence: if the support of (X, Y) is nonrectangular, then X and Y will necessarily be dependent.
- Note that the logical inverse doesn't necessarily follow: just because a support is rectangular doesn't mean we can automatically conclude $X \perp Y$. To establish independence, you must use the definition.

Theorem

Given two random variables (X, Y) with joint p.d.f. $f_{X,Y}(x, y)$, if $X \perp Y$ then $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$

Proof.

- By independence, we have $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.
- Therefore, plugging into the LOTUS we find

$$\begin{aligned}
\mathbb{E}[XY] &= \iint_{\mathbb{R}^2} xy f_{X,Y}(x, y)\ \mathrm{d}A \\
&= \iint_{\mathbb{R}^2} xy \cdot f_X(x) f_Y(y)\ \mathrm{d}A \\
&= \iint_{\mathbb{R}^2} [x f_X(x)] \cdot [y f_Y(y)]\ \mathrm{d}A \\
&= \left(\int_{\mathbb{R}} x f_X(x)\ \mathrm{d}x \right) \cdot \left(\int_{\mathbb{R}} y f_Y(y)\ \mathrm{d}y \right) = \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}$$

□

Independence and Expectation

Theorem

Given n independent random variables $X_1, \cdots, X_n$, we have

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i]$$

Independence and Tranformations

Theorem

If $X_1, \cdots, X_{n+m}$ are independent random variables, and if $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$ are real-valued functions, then $g(X_1, \cdots, X_n) \perp h(X_{n+1}, \cdots, X_{n+m})$. In other words: functions of independent random variables are also independent.

- By the way, we won't talk much about multivariate transformations in this class. But, don't be scared by quantities like $g(X_1, \cdots, X_n)$; again, this is just a random variable!

Covariance and Correlation

Leadup

- Recall how our discussion on Variance started: we began with the (seemingly broad) question of "how can we measure the *spread* of a random variable?"
- With a pair of bivariate random variables (X, Y), we can ask ourselves another question: "how *related* are X and Y?"
- As a concrete example, consider taking a stick of length 1 and breaking it into two smaller pieces by picking a breakpoint uniformly along the length of the stick: let X denote the length of the shorter piece and Y denote the length of the longer piece. There is a clear "direct" relationship between X and Y: a one unit increase in X (i.e. making the shorter piece 1 unit longer) corresponds to a 1 unit decrease in Y (makes the longer piece shorten by 1 unit, since the length of the entire rod must remain constant).

Covariance

Definition: Covariance

The **covariance** of two random variables X and Y is defined as

$$\text{Cov}(X, Y) := \mathbb{E}\left\{[X - \mathbb{E}(X)] \cdot [Y - \mathbb{E}(Y)]\right\}$$

By expanding out the RHS and simplifying, one can show that covariance is equivalent to

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

Our Familiar Example, Again!

Suppose (X, Y) is a pair of random variables with joint density given by

$$f_{X,Y}(x, y) = \begin{cases} 2 \cdot e^{-(x+y)} & \text{if } x \leq y < \infty,\ 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

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

Independence and Covariance

- Now, recall that when $X \perp Y$ we have that $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$. This leads to the following interesting observation:

Theorem

If random variables X and Y are independent, then i.e. $\mathrm{Cov}(X, Y) = 0$.

- Let me stress something very important: **THE CONVERSE IS NOT (IN GENERAL) TRUE!** There are several examples of random variables (X, Y) that have zero covariance but are dependent.
- Additionally: we can levarage this fact in some situations to enable us to bypass any need for computation. What I mean is the following: if given a joint p.d.f. $f_{X,Y}(x, y)$ that factors as $f_X(x) \cdot f_Y(Y)$, we can immediately conclude that $X \perp Y$ and therefore $\mathrm{Cov}(X, Y) = 0$. Perhaps something to keep in mind when you're doing your next homework assignment...

Properties of Covariance

Theorem: Bilinearity of Covariance

$$\text{Cov}\left(\sum_{i=1}^{n} a_i X_i\,,\ \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \text{Cov}(X_i, Y_j)$$

- For example,

$$\text{Cov}(aX+bY,\ cZ+dW) = ac\text{Cov}(X,Z)+ad\text{Cov}(X,W)+bc\text{Cov}(Y,Z)+bd\text{Cov}(Y,W)$$

Theorem: Symmetry of Covariance

$$\text{Cov}(X,Y) = \text{Cov}(Y,X)$$

Theorem: Self-Covariance

$$\text{Cov}(X,X) = \text{Var}(X)$$

Variance of Linear Combinations

Theorem

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

- Here, the sum on the rightmost end is a double sum over indices i and j such that the i index is strictly less than the j index. For example:

$$\begin{aligned}\mathrm{Var}(a_1X_1 + a_2X_2 + a_3X_3) = {} & a_1^2\mathrm{Var}(X_1) + a_2^2\mathrm{Var}(X_2) + a_3^2\mathrm{Var}(X_3) \\ & + 2a_1a_2\mathrm{Cov}(X_1, X_2) + 2a_1a_3\mathrm{Cov}(X_1, X_3) + 2a_2a_3\mathrm{Cov}(X_2, X_3)\end{aligned}$$

- Believe it or not, I find the proof of this theorem to be helpful in remembering its statement!

Variance of Linear Combinations

Proof.

- The first fact we use is that $\text{Var}(X) = \text{Cov}(X, X)$. Therefore,

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \text{Cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{n} a_j X_j\right)$$

[Note that in these sorts of double-sum computations it is very important to not reuse the same index multiple times, lest you get a bit confused and forget which indices are actually alike!]

- Now we use Bilinearity:

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \text{Cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{n} a_j X_j\right)\\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{Cov}(X_i, X_j)\end{aligned}$$

□

Proof.

- Next, we break the double sum up into two sums, using the following division: we consider the case where $i = j$ separate from where $i \neq j$:

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i,\ \sum_{j=1}^{n} a_j X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=j} a_i a_j \mathrm{Cov}(X_i, X_j) + \sum_{i\neq j} a_i a_j \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} a_i^2 \mathrm{Cov}(X_i, X_i) + \sum_{i\neq j} a_i a_j \mathrm{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + \sum_{i\neq j} a_i a_j \mathrm{Cov}(X_i, X_j)
\end{aligned}$$

□

Variance of Linear Combinations

Proof.

- Finally, we consider the rightmost sum: by the symmetry property of the covariance operator, we will have quite a few duplicated terms [for instance, $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$]. Therefore, we can consider summing only along the indices for which $i < j$, and then multiply by 2:

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2 \sum_{i<j} a_i a_j \text{Cov}(X_i, X_j)\end{aligned}$$

□

Example

Suppose (X, Y) is a pair of random variables with joint density given by

$$f_{X,Y}(x,y) = \begin{cases} 2 \cdot e^{-(x+y)} & \text{if } x \leq y < \infty,\ 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute $\text{Var}(X - Y)$.

Example

Let $X_1, \cdots, X_n$ be a sequence of random variables with the following covariance structure:

$$\text{Cov}(X_i, X_j) = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } |i - j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute $\text{Var}\left(\sum_{i=1}^{n} X_i\right)$

Independence and Variance

- Finally, let's tie together independence and variance.

Theorem

If $X_1, \cdots, X_n$ are independent and if $a_1, \cdots, a_n \in \mathbb{R}$ are fixed constants, then

$$\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)$$

In other words, the only time we are able to pass a variance through a sum is when the random variables in the sum are independent.

Correlation

- Let's return to the notion of covariance for a moment.
- In general, there are no bounds on $\mathrm{Cov}(X, Y)$.
- A positive covariance means that X and Y are positively related (i.e. when X goes up, so does Y) where as a negative covariance means that X and Y are negatively related (i.e. when X goes up, Y goes down).
- The issue is the following: the *magnitude* of covariance doesn't give us a whole lot of information. That is, just because $\mathrm{Cov}(X, Y) > \mathrm{Cov}(Z, W) > 0$ doesn't mean that X and Y are "more strongly" related than Z and W. (The issue lies actually with standard deviations; random variables with large standard deviations tend to dominate covariances).
- For this reason, statisticians like to examine a standardized version of covariance:

Definition: Correlation

The **correlation** between two random variables X and Y is defined to be

$$\mathrm{Corr}(X, Y) := \frac{\mathrm{Cov}(X, Y)}{\mathrm{SD}(X) \cdot \mathrm{SD}(Y)}$$

- It turns out that correlations are always bound between -1 and 1, inclusive.

Example

Suppose (X, Y) is a pair of random variables with joint density given by

$$f_{X,Y}(x,y) = \begin{cases} 2 \cdot e^{-(x+y)} & \text{if } x \leq y < \infty,\ 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute $\text{Corr}(X, Y)$

Covariance Matrix

- By the way, that trick we used in the previous example of writing down a table consisting of the covariances between any two X_i and X_j is so useful, it has an associated mathematical quantity:

Definition: Covariance Matrix

Given a random vector $\vec{\boldsymbol{X}}$, we define the **covariance matrix** of $\vec{\boldsymbol{X}}$ to be the matrix $\boldsymbol{\Sigma}$ prescribed by

$$(\boldsymbol{\Sigma})_{ij} = \text{Cov}(X_i, X_j)$$

In other words, the $(i, j)^{\text{th}}$ element of $\boldsymbol{\Sigma}$ is $\text{Cov}(X_i, X_j)$.

- So, the diagonal entries of $\boldsymbol{\Sigma}$ represent the variances.
- Question: If $\boldsymbol{\Sigma}$ is diagonal, can we conclude the X_i's to be independent? No! We can only conclude them to be uncorrelated.

Example

Let $\vec{\boldsymbol{X}} = (X_1, X_2, X_3)$ denote a random vector with variance-covariance matrix given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} 10 & -4 & 3 \\ -4 & 5 & 2 \\ 3 & 2 & 5 \end{pmatrix}$$

- $\text{Var}(X_1 + X_3) = \text{Var}(X_1) + \text{Var}(X_2) + 2\text{Cov}(X_1, X_3) = (10) + (5) + (6) = 21$

Independent and Identically Distributed (I.I.D.)

- I'd like to leave off with one of the **MOST IMPORTANT** (and I'm not kidding!) acronyms in all of statistics:

Definition: I.I.D.

Suppose $X_1, \cdots, X_n$ are independent random variables that all follow the same distribution (from a marginal point of view). We then say that the n random variables are **independent and identically distributed**, or just **i.i.d.** for short.

- As an example, suppose we have

$$X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$$

What this means is that (1) the X_i's are all independent, and (2) each X_i follows the $\text{Exp}(\lambda)$ distribution. Consequently, the joint density is given by

$$f_{\vec{X}}(\vec{x}) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \cdot \mathbb{1}_{\{\text{all } x_i\text{'s greater than } 0\}}$$

A Quick Look Ahead

- I've offhandedly mentioned quantities like $\sum_{i=1}^{n} a_i X_i$ quite a bit during this lecture.
- A natural question might be: "...huh?"
- Perhaps think of it this way: the function $g : \mathbb{R}^n \to \mathbb{R}$ prescribed by $\vec{x} \mapsto \sum_{i=1}^{n} a_i x_i$ is, well, a function! A random vector $\vec{X}$ is a function from Ω to $\mathbb{R}^n$. Hence, $(g \circ X) : \Omega \to \mathbb{R}$, meaning $g(\vec{X}) = \sum_{i=1}^{n} a_i X_i$ is just a random variable!
- We've already seen how to compute its mean and variance; coming up, we'll talk about how to get more information about this random variable.