6.2 Covariance and correlation

Unit 6: Joint Distributions and Conditional Expectation

Adapted from Blitzstein-Hwang Chapters 7 and 9.

Just as the mean and variance provided single-number summaries of the distribution of a single r.v., covariance is a single-number summary of the joint distribution of two r.v.s. Roughly speaking, covariance measures a tendency of two r.v.s to go up or down together, relative to their expected values: positive covariance between X and Y indicates that when X goes up, Y also tends to go up, and negative covariance indicates that when X goes up, Y tends to go down. Here is the precise definition.

Definition 6.2.1 (Covariance).

The *covariance* between r.v.s X and Y is

$$\mathrm{Cov}(X,Y) = E((X - EX)(Y - EY)).$$

Multiplying this out and using linearity, we have an equivalent expression:

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

Let's think about the definition intuitively. If X and Y tend to move in the same direction, then $X - EX$ and $Y - EY$ will tend to be either both positive or both negative, so $(X - EX)(Y - EY)$ will be positive on average, giving a positive covariance. If X and Y tend to move in opposite directions, then $X - EX$ and $Y - EY$ will tend to have opposite signs, giving a negative covariance.

If X and Y are independent, then their covariance is zero. We say that r.v.s with zero covariance are *uncorrelated*.

Theorem 6.2.2.

If X and Y are independent, then they are uncorrelated.

The converse of this theorem is false: just because X and Y are uncorrelated does not mean they are independent. For example, let $X \sim \mathcal{N}(0,1)$, and let $Y = X^2$. Then $E(XY) = E(X^3) = 0$ because the odd moments of the standard Normal distribution are equal to 0 by symmetry. Thus X and Y are uncorrelated,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 = 0,$$

but they are certainly not independent: Y is a function of X, so knowing X gives us perfect information about Y. Covariance is a measure of *linear* association, so r.v.s can be dependent in nonlinear ways and still have zero covariance, as this example demonstrates. The bottom right plot of Figure 6.2.3 shows draws from the joint distribution of X and Y in this example. The other three plots illustrate positive correlation, negative correlation, and independence.

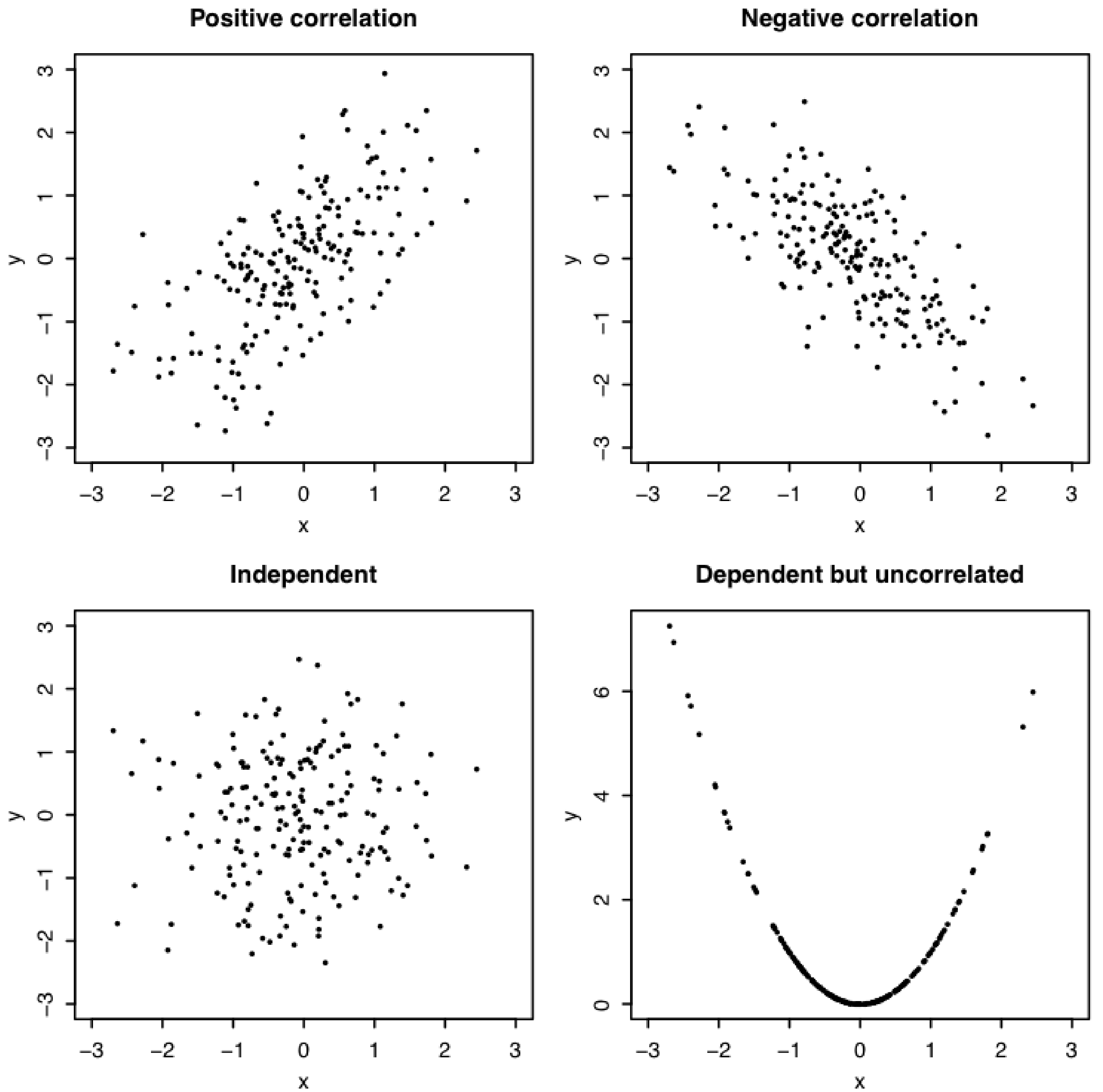

Figure 6.2.3: Draws from the joint distribution of (X, Y) under various dependence structures.
Top left: X and Y are positively correlated.
Top right: X and Y are negatively correlated.
Bottom left: X and Y are independent, hence uncorrelated.
Bottom right: Y is a deterministic function of X, but X and Y are uncorrelated.

View Larger Image
Image Description

Covariance has the following key properties.

1. $\text{Cov}(X, X) = \text{Var}(X)$.
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. $\text{Cov}(X, c) = 0$ for any constant c.
4. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$ for any constant a.
5. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.
6. $\text{Cov}(X + Y, Z + W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$.
7. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$. For n r.v.s $X_1, \ldots, X_n$,

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) + 2\sum_{i<j} \text{Cov}(X_i, X_j).$$

The first five properties follow readily from the definition and basic properties of expectation. Property 6 follows from Property 2 and Property 5, by expanding

$$\begin{aligned}\text{Cov}(X + Y, Z + W) &= \text{Cov}(X, Z + W) + \text{Cov}(Y, Z + W)\\ &= \text{Cov}(Z + W, X) + \text{Cov}(Z + W, Y)\\ &= \text{Cov}(Z, X) + \text{Cov}(W, X) + \text{Cov}(Z, Y) + \text{Cov}(W, Y)\\ &= \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W).\end{aligned}$$

Property 7 follows from writing the variance of an r.v. as its covariance with itself (by Property 1) and then using Property 6 repeatedly. We have now fulfilled our promise from Chapter 4 that for independent r.v.s, the variance of the sum is the sum of the variances. By Theorem 6.2.2, independent r.v.s are uncorrelated, so in that case all the covariance terms drop out of the expression in Property 7.

Since covariance depends on the units in which X and Y are measured---if we decide to measure X in centimeters rather than meters, the covariance is multiplied a hundredfold---it is easier to interpret a unitless version of covariance called correlation.

Definition 6.2.4 (Correlation).

The *correlation* between r.v.s X and Y is

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

(This is undefined in the degenerate cases $\text{Var}(X) = 0$ or $\text{Var}(Y) = 0$.)

Notice that shifting and scaling X and Y has no effect on their correlation. Shifting does not affect $\mathrm{Cov}(X, Y)$, $\mathrm{Var}(X)$, or $\mathrm{Var}(Y)$, so the correlation is unchanged. As for scaling, the fact that we divide by the standard deviations of X and Y ensures that the scale factor cancels out:

$$\mathrm{Corr}(cX, Y) = \frac{\mathrm{Cov}(cX, Y)}{\sqrt{\mathrm{Var}(cX)\mathrm{Var}(Y)}} = \frac{c\mathrm{Cov}(X, Y)}{\sqrt{c^2\mathrm{Var}(X)\mathrm{Var}(Y)}} = \mathrm{Corr}(X, Y).$$

Correlation is convenient to interpret because it does not depend on the units of measurement and is always between -1 and 1.

> **Theorem 6.2.5 (Correlation bounds).**
>
> For any r.v.s X and Y,
>
> $$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

Proof

Without loss of generality we can assume X and Y have variance 1, since scaling does not change the correlation. Let $\rho = \mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y)$. Using the fact that variance is nonnegative, along with Property 7 of covariance, we have

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) = 2 + 2\rho \geq 0,$$

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y) = 2 - 2\rho \geq 0.$$

Thus, $-1 \leq \rho \leq 1$.

Covariance properties are a helpful tool for finding *variances*, especially when the r.v. of interest is a sum of dependent random variables. The next example uses properties of covariance to derive the variance of the Hypergeometric distribution.

Example 6.2.6 (Hypergeometric variance).

Let $X \sim \mathrm{HGeom}(w, b, n)$. Find $\mathrm{Var}(X)$.

Solution

Interpret X as the number of white balls in a sample of size n from an urn with w white and b black balls. We can represent X as the sum of indicator r.v.s, $X = I_1 + \cdots + I_n$, where I_j is the indicator of the jth ball in the sample being white. Each I_j has mean $p = w/(w + b)$ and variance $p(1 - p)$, but because the I_j are dependent, we cannot simply add their variances. Instead, we apply properties of covariance:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}\left(\sum_{j=1}^{n} I_j\right)\\ &= \mathrm{Var}(I_1) + \cdots + \mathrm{Var}(I_n) + 2\sum_{i<j}\mathrm{Cov}(I_i, I_j)\\ &= np(1-p) + 2\binom{n}{2}\mathrm{Cov}(I_1, I_2).\end{aligned}$$

In the last step, we used the fact that all $\binom{n}{2}$ pairs of indicators have the same covariance by symmetry. Now we just need to find $\mathrm{Cov}(I_1, I_2)$. By the fundamental bridge,

$$\begin{aligned}\mathrm{Cov}(I_1, I_2) &= E(I_1 I_2) - E(I_1)E(I_2)\\ &= P(\text{1st and 2nd balls both white}) - P(\text{1st ball white})P(\text{2nd ball white})\\ &= \frac{w}{w+b}\cdot\frac{w-1}{w+b-1} - p^2.\end{aligned}$$

Plugging this into the above formula and simplifying, we can obtain

$$\mathrm{Var}(X) = \frac{N-n}{N-1}np(1-p),$$

Next, we introduce the Multinomial and the Multivariate Normal distributions. The Multinomial is the most famous discrete multivariate distribution, and the Multivariate Normal is the most famous continuous multivariate distribution.

Learn About Verified Certificates

© All Rights Reserved