

Notes: Covariance, Correlation, Bivariate Gaussians

CS 3130 / ECE 3530: Probability and Statistics for Engineers

Feb 27, 2024

Expectation of Joint Random Variables. When we have two random variables X, Y described jointly, we can take the expectation of functions of both random variables, $g(X, Y)$. This is defined how you think it would be.

For discrete:

$$E[g(X, Y)] = \sum_i \sum_j g(a_i, b_j) P(X = a_i, Y = b_j)$$

For continuous:

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

Linearity of expectation revisited. We've already stated expectation was linear; now we show why. Let $g(X, Y) = rX + sY$, where r, s are constants. Plugging this into the formulas above, we can see that $E[rX + sY] = rE[X] + sE[Y]$. Here we run through the discrete case (continuous case works exactly the same):

$$\begin{aligned} E[rX + sY] &= \sum_i \sum_j (ra_i + sb_j) P(X = a_i, Y = b_j) \\ &= r \sum_i \sum_j a_i P(X = a_i, Y = b_j) + s \sum_i \sum_j b_j P(X = a_i, Y = b_j) \\ &= r \sum_i a_i \left(\sum_j P(X = a_i, Y = b_j) \right) + s \sum_j b_j \left(\sum_i P(X = a_i, Y = b_j) \right) \\ &= r \sum_i a_i P(X = a_i) + s \sum_j b_j P(Y = b_j) \\ &= rE[X] + sE[Y] \end{aligned}$$

Covariance. The **covariance** of two random variables X, Y is defined as

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

Notice the similarity to the variance definition. In fact, $\text{Cov}(X, X) = \text{Var}(X)$. Covariance is a measure of how related X and Y are. If $\text{Cov}(X, Y)$ is positive, it means that “ X and Y tend to go in the same direction”. If $\text{Cov}(X, Y)$ is negative, it means that “ X and Y tend to go in opposite directions.” As an example, let $Y = X$. Now X and Y really go in the same direction! In this case $\text{Cov}(X, Y) = \text{Var}(X)$, which is always positive. Now consider the case that $Y = -X$. So, X and Y are really going in opposite directions. You can check that $\text{Cov}(X, Y) = -\text{Var}(X)$, which is always negative.

Just like variance, we have an alternate definition for covariance:

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

Exercise: Prove these two formulas for $\text{Cov}(X, Y)$ are equal.

So, $E[X + Y] = E[X] + E[Y]$ holds for expectation. Does it also hold for variance? In other words, does $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$?

$$\begin{aligned}\text{Var}(X + Y) &= E[(X + Y)^2] - E[X + Y]^2 \\ &= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\ &= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2 \\ &= (E[X^2] - E[X]^2) + (E[Y^2] - E[Y]^2) + 2(E[XY] - E[X]E[Y]) \\ &= \text{Var}(X) + \text{Var}(Y) + 2(E[XY] - E[X]E[Y]) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)\end{aligned}$$

So, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ if and only if $\text{Cov}(X, Y) = 0$.

Notation: Remember we had the notation $\sigma_X^2 = \text{Var}(X)$. We will also use the notation $\sigma_{X,Y} = \text{Cov}(X, Y)$.

Important Fact: If X and Y are independent, then $\text{Cov}(X, Y) = 0$ (see book for proof). This matches our intuition that independence means that X and Y are not related and that $\text{Cov}(X, Y)$ is a numerical measure of how related X and Y are.

Tricky Important Fact: If $\text{Cov}(X, Y) = 0$, this does *not* necessarily mean that X and Y are independent!

Correlation. One problem with covariance is that it scales with the random variables X and Y . That is, $\text{Cov}(rX, sY) = rs \text{Cov}(X, Y)$. (This follows directly from the linearity of expectation.) Therefore, if we change the units of X and Y , we will scale their covariance. This makes it really difficult to know how strongly two random variables are based on how large their covariance is. For example, let's think about X and Y variables that are given in meters. If we were to rewrite them in terms of centimeters, then each variable will scale by 100, and the covariance will scale by $100^2 = 10,000$. However, these are really just the same random variables, and their larger covariance does not mean they are more strongly related to each other.

To overcome this problem, the **correlation** is defined to remove these scale factors:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

Notice that scaling cancels out in the numerator and denominator, so $\rho(rX, sY) = \rho(X, Y)$. So, correlation is *invariant* to the units in which we write X and Y .

Bivariate Gaussian Distribution. One of the most important examples of a continuous joint distribution is the bivariate Gaussian distribution. Let's begin with understanding what it looks like when we combine two independent Gaussian random variables $X \sim N(\mu_x, \sigma_x)$ and $Y \sim N(\mu_y, \sigma_y)$. Because of independence, the joint pdf is given by

$$\begin{aligned} f(x, y) &= f(x)f(y) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{(x - \mu_x)^2}{2\sigma_x^2}\right) \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{(y - \mu_y)^2}{2\sigma_y^2}\right) \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{1}{2} \left[\frac{(x - \mu_x)^2}{2\sigma_x^2} + \frac{(y - \mu_y)^2}{2\sigma_y^2} \right] \right) \end{aligned}$$

Now, if we allow X and Y to be correlated with $\rho = \rho(X, Y)$, we get a more general form of the bivariate Gaussian pdf:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left[\frac{(x - \mu_x)^2}{2\sigma_x^2} + \frac{(y - \mu_y)^2}{2\sigma_y^2} - \frac{2\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} \right] \right)$$

See the R source code that we covered in class for some plots of what these joint pdf's look like.

Summary of important formulas:

Covariance:

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

Correlation:

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

Variance of Addition:

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