Math 180B, Winter 2021

Notes on covariance and the bivariate normal distribution

1. Covariance. If X and Y are random variables with finite variances, then their *covariance* is the quantity

$$\mathrm{Cov}(X,Y) := \mathbf{E}[(X-\mu_X)(Y-\mu_Y)], \tag{1.1}$$

where $\mu_X = \mathbf{E}[X]$ and $\mu_Y = \mathbf{E}[Y]$. The covariance is a measure of the extent to which X and Y are linearly related. Because $(X-\mu_X)(Y-\mu_Y) = XY - \nu_X Y - \mu_Y X + \mu_X\mu_Y$, the covariance can also be expressed as

$$\mathrm{Cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\cdot\mathbf{E}[Y] = \mathbf{E}[XY] - \mu_X\mu_Y. \tag{1.2}$$

Observe that if X and Y are independent, then $\mathbf{E}[XY] = \mu_X\mu_Y$. Therefore

$$X \perp\!\!\!\perp Y \qquad \Longrightarrow \qquad \mathrm{Cov}(X,Y) = 0. \tag{1.3}$$

The converse implication fails as a general statement, by an example discussed in class [X uniformly distributed on $(-1,1)$, $Y = X^2$]. But see Corollary 3 below.

2. Variance. Part of the importance of covariance is the way in which it completes the addition formula for the variance of a sum of random variables:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y). \tag{2.1}$$

More generally, because we evidently have

$$\mathrm{Cov}(\alpha X, \beta Y) = \alpha\beta\mathrm{Cov}(X,Y) \tag{2.2}$$

for real numbers α and β, it is also true that

$$\mathrm{Var}[\alpha X + \beta Y] = \alpha^2\mathrm{Var}[X] + \beta^2\mathrm{Var}[Y] + 2\alpha\beta\mathrm{Cov}(X,Y). \tag{2.3}$$

3. Correlation. The correlation of two random variables X and Y is their standardized covariance:

$$\begin{aligned}\rho = \rho(X,Y) = \mathrm{corr}(X,Y) :&= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} \\ &= \mathbf{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right)\right].\end{aligned} \tag{3.1}$$

Here, for example, σ_X^2 is the variance of X.

4. Example. Suppose $Y = \alpha X + \beta$, where α and β are real constants. Then $\mu_Y = \alpha\mu_X + \beta$, $\text{Var}[Y] = \alpha^2\text{Var}[X]$, and $\text{Cov}(X, Y) = \alpha\text{Var}[X]$. Consequently,

$$\text{corr}(X,Y) = \begin{cases} 1, & \text{if } \alpha > 0; \\ 0, & \text{if } \alpha = 0; \\ -1, & \text{if } \alpha < 0 \end{cases} \tag{4.1}$$

when Y is a (non-random) straight-line function of X.

5. Theorem. [Cauchy-Schwarz Inequality]

$$\left|\text{Cov}(X,Y)\right| \le \sigma_X \sigma_Y \tag{5.1}$$

and

$$\left|\text{corr}(X,Y)\right| \le 1. \tag{5.2}$$

Assuming that $\sigma_X\sigma_Y > 0$, *equality holds in either of (5.1) or (5.2) only if*

$$\mathbf{P}[(X-\mu_X)/\sigma_X = \text{sign}(\text{corr}(X,Y)) \cdot (Y-\mu_Y)/\sigma_Y] = 1.$$

Proof. The inequalities (5.1) and (5.2) are equivalent, so it is enough to demonstrate (5.2), which (in view of the third equality in (3.1)) can be written as

$$\left|\text{Cov}(\hat{X},\hat{Y})\right| \le 1, \tag{5.3}$$

where $\hat{X} = (X-\mu_X)/\sigma_X$ and $\hat{Y} = (Y-\mu_Y)/\sigma_Y$. To see (5.3) we consider the function

$$g(t) := \mathbf{E}\left[(\hat{X} - t\hat{Y})^2\right], \qquad t \in \mathbf{R}.$$

Being the expectation of a square, $g(t) \ge 0$ for all real t. On the other hand, g is a quadratic:

$$g(t) = \mathbf{E}[\hat{X}^2] - 2t\mathbf{E}[\hat{X}\hat{Y}] + t^2\mathbf{E}[\hat{Y}^2] = 1 - 2t\text{Cov}(\hat{X},\hat{Y}) + t^2.$$

The only way a quadratic function can take only non-negative values is if its discriminant is non-positive. Thus we must have

$$[-2\text{Cov}(\hat{X},\hat{Y})]^2 - 4\cdot 1\cdot 1 \le 0.$$

That is, $4[\mathrm{Cov}(\hat{X},\hat{Y})]^2 - 4 \le 0$, or what is the same

$$[\mathrm{Cov}(\hat{X},\hat{Y})]^2 \le 1, \tag{5.4}$$

which is equivalent to (5.3).

Suppose now that $\mathrm{corr}(X,Y) = 1$. In this case we have $g(t) = (1-t)^2$, so $g(1) = 0$. But $g(1) = \mathbf{E}\left[(\hat{X}-\hat{Y})^2\right]$, and the only way this (the expectation of a non-negative random variable) can be 0 is if that random variable is itself 0 with probability 1. This shows that if $\mathrm{corr}(X,Y) = 1$ then $\mathbf{P}[\hat{X} = \hat{Y}] = 1$, as required by the final sentence of the Theorem. Similar considerations apply when $\mathrm{corr}(X,Y) = -1$. □

6. Bivariate Normal Distribution. A pair of random variables X and Y is said to have the *bivariate normal distribution* provided

$$\alpha X + \beta Y \tag{6.1}$$

has the univariate normal distribution for each pair (α,β) of real numbers. This state of affairs will be indicated by the notation

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2,$$

or by

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}\right) \tag{6.2}$$

if I wish to indicate the means (μ_X,μ_Y), the variances (σ_X^2,σ_Y^2), and the covariance σ_{XY} of X and Y. Observe that if (6.2) holds then $X \sim \mathcal{N}_1$ (that is X has a univariate normal distribution—take $\alpha = 1$ and $\beta = 0$) and also $Y \sim \mathcal{N}_1$. In what follows I will write

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{6.3}$$

for the correlation of X and Y. The column vector

$$\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}$$

appearing in (6.2) is the *mean vector* for the pair $(X,Y)^t$, while the 2×2 matrix

$$\begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \sigma_Y^2 \end{pmatrix}$$

is its *variance-covariance matrix*. The *standard* bivariate normal distribution is the special case in which the means are both 0 and the variances are both 1.

7. Representation. The following construction of a standard bivariate normal pair, in terms of iid univariate normals, is useful for various calculations. Let X and Z be independent standard normal random variables. Then for real constants α and β, the random variables αX and βY are also independent, and so their sum $\alpha X + \beta Y$ is also normally distributed, as discussed in class. It follows from the discussion in **6.** that X and Z have a bivariate normal distribution. More precisely,

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \tag{7.1}$$

the standard bivariate normal distribution.

Now fix $\rho \in [-1, 1]$, and consider the random variable Y defined by

$$Y = \rho X + \sqrt{1-\rho^2}\, Z. \tag{7.2}$$

It is easy to check that $\mathbf{E}[Y] = 0$, $\mathrm{Var}[Y] = 1$, and $\mathrm{Cov}(X, Y) = \rho$. Also, if α and β are any two real numbers, then

$$\alpha X + \beta Y = (\alpha + \beta\rho)X + \beta\sqrt{1-\rho^2}\, Z$$

is normally distributed, because X and Z have a bivariate normal distribution as remarked above. It follows that X and Y themselves have a bivariate normal distribution. More precisely,

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \tag{7.3}$$

8. Corollary 1. *With X and Y as in (7.3), the conditional distribution of Y, given that X takes the value x, is normal, with mean ρx and variance $1-\rho^2$. In symbols*

$$Y \mid X = x \sim \mathcal{N}_1(\rho x, 1-\rho^2). \tag{8.1}$$

Proof. Look at (7.2): If I tell you that X equals x, this has no effect on the (independent) random variable Z, which still has the standard normal distribution. Thus, under the condition $X = x$, the random variable Y is Z scaled by a factor of $\sqrt{1-\rho^2}$ (which changes its variance to $1-\rho^2$) and then translated by ρx (which changes its mean to ρx).

Neither the scaling nor the translation alter the fact that the random variable is normally distributed. ∎

9. Corollary 2. *Suppose that*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \sigma_Y^2 \end{pmatrix}\right). \tag{9.1}$$

Then the conditional distribution of Y, given that X takes the value x is normal:

$$Y \mid X = x \sim \mathcal{N}(\mu_Y + \frac{\sigma_Y\rho}{\sigma_X}(x-\mu_X), (1-\rho^2)\sigma_Y^2). \tag{9.2}$$

Proof. Just apply Corollary 1 to the standardized random variables $\hat{X} = (X-\mu_X)/\sigma_X$ and $\hat{Y} = (Y-\mu_Y)/\sigma_Y$. ∎

10. Corollary 3. *Suppose that*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho \\ \sigma_X\sigma_Y\rho & \sigma_Y^2 \end{pmatrix}\right).$$

Then X and Y are independent if and only if $\rho = 0$.

Proof. We need only consider the "if" part of this assertion—the "only if" part holds for any two random variables, bivariate normal or not. If $\rho = 0$, then according to (9.2) the conditional density of Y given the value of X, namely $f_{Y|X}(y|x)$, does not depend on x, and so is a function of y alone, call it $g(y)$. This means that the joint density $f(x, y)$ of X and Y factors:

$$f(x,y) = f_X(x) \cdot f_{Y|X}(y|x) = f_X(x) \cdot g(y). \tag{10.1}$$

Now integrate out x:

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{-\infty}^{\infty} f_X(x) \cdot g(y)\,dx = g(y). \tag{10.2}$$

Using (10.2) in (10.1) we find that

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

which proves to the asserted independence. ∎

11. References. The material discussed above can be found in sections 7.3 and 7.8 of *A First Course in Probability* by Sheldon Ross, and also in sections 6.4 and 6.4 of *PROBABILITY* by Jim Pitman. The latter text is available in pdf form for UCSD students at

`http://link.springer.com/book/10.1007%2F978-1-4612-4374-8`