Chapter 4

Variances and covariances

The expected value of a random variable gives a crude measure for the "center of location" of the distribution of that random variable. For instance, if the distribution is symmetric about a value μ then the expected value equals μ. To refine the picture of a distribution distributed about its "center of location" we need some measure of spread (or concentration) around that value. The simplest measure to calculate for many distributions is the **variance** (or, more precisely, the square root of the variance).

Definition. *The* **variance** *of a random variable X with expected value $\mathbb{E}X = \mu$ is defined as $\mathrm{var}(X) = \mathbb{E}\left((X-\mu)^2\right)$. The square root of the variance of a random variable is called its* **standard deviation**, *sometimes denoted by $\mathrm{sd}(X)$.*

The **covariance** *between random variables Y and Z, with expected values μ_Y and μ_Z, is defined as $\mathrm{cov}(Y,Z) = \mathbb{E}\left((Y-\mu_Y)(Z-\mu_Z)\right)$. The* **correlation** *between Y and Z is defined as*

$$\mathrm{corr}(Y,Z) = \frac{\mathrm{cov}(Y,Z)}{\sqrt{\mathrm{var}(Y)\mathrm{var}(Z)}}$$

The random variables Y and Z are said to be **uncorrelated** *if $\mathrm{corr}(Y,Z) = 0$.*

Notice that $\mathrm{cov}(X,X) = \mathrm{var}(X)$. Results about covariances contain results about variances as special cases.

> REMARK. Strictly speaking, the variance of a random variable is not well defined unless it has a finite expectation. Similarly, we should not talk about $\mathrm{corr}(Y,Z)$ unless both random variables have well defined variances for which $0 < \mathrm{var}(Y) < \infty$ and $0 < \mathrm{var}(Z) < \infty$.

Sometimes it is easier to subtract off the expected values at the end of the calculation, by means of the formulae $\mathrm{cov}(Y,Z) = \mathbb{E}(YZ) - (\mathbb{E}Y)(\mathbb{E}Z)$ and, as a particular case, $\mathrm{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$. Both formulae follow via an expansion of the product:

$$\begin{aligned}\mathrm{cov}(Y,Z) &= \mathbb{E}\left(YZ - \mu_Y Z - \mu_Z Y + \mu_Y\mu_Z\right)\\ &= \mathbb{E}(YZ) - \mu_Y\mathbb{E}Z - \mu_Z\mathbb{E}Y + \mu_Y\mu_Z\\ &= \mathbb{E}(YZ) - \mu_Y\mu_Z.\end{aligned}$$

A pair of random variables X and Y is said to be **independent** if *every event determined by X is independent of every event determined by Y*. For example, independence implies that events such as $\{X \le 5\}$ and $\{7 \le Y \le 18\}$ are independent, and so on. Independence of the random variables also implies independence of functions of those random variables. For example, $\sin(X)$ would be independent of e^Y, and so on. For the purposes of Stat241, you should not fret about the definition of independence: Just remember to explain why you regard some pieces of information as irrelevant when you calculate conditional probabilities and conditional expectations.

For example, suppose a random variable X can take values $x_1, x_2, \ldots$ and that X is independent of another random variable Y. Consider the expected value of a product $g(X)h(Y)$, for any functions g and h. Calculate by conditioning on the possible values taken by X:

$$\mathbb{E}g(X)h(Y) = \sum_i \mathbb{P}\{X = x_i\}\mathbb{E}(g(X)h(Y) \mid X = x_i).$$

Given that $X = x_i$, we know that $g(X) = g(x_i)$ but we get no help with understanding the behavior of $h(Y)$. Thus, independence implies

$$\mathbb{E}(g(X)h(Y) \mid X = x_i) = g(x_i)\mathbb{E}(h(Y) \mid X = x_i) = g(x_i)\mathbb{E}h(Y).$$

Deduce that

$$\mathbb{E}g(X)h(Y) = \sum_i \mathbb{P}\{X = x_i\}g(x_i)\mathbb{E}h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y).$$

Put another way,

$$\operatorname{cov}\big(g(X), h(Y)\big) = 0 \qquad \text{if } X \text{ and } Y \text{ are independent random variables;}$$

each function of X is uncorrelated with each function of Y. In particular, if X and Y are independent then they are uncorrelated. The converse is not usually true.

Example 1: An example of uncorrelated random variables that are dependent. (Do not look at this example until you have read how to calculate covariances between linear combinations of random variables.)

The variance of a random variable X is unchanged by an added constant: $\operatorname{var}(X + C) = \operatorname{var}(X)$ for every constant C, because $(X+C)-\mathbb{E}(X+C) = X-\mathbb{E}X$, the C's cancelling. It is a desirable property that the spread should not be affected by a change in location. However, it is also desirable that multiplication by a constant should change the spread: $\operatorname{var}(CX) = C^2\operatorname{var}(X)$ and $\operatorname{sd}(CX) = |C|\operatorname{sd}(X)$, because $(CX - \mathbb{E}(CX))^2 = C^2(X - \mathbb{E}X)^2$. In summary:

$\operatorname{var}(a + bX) = b^2\operatorname{var}(X)$ and $\operatorname{sd}(a + bX) = |b|\operatorname{sd}(X)$ for constants a and b.

REMARK. Try not to confuse properties of expected values with properties of variances: for constants a and b we have $\operatorname{var}(a + bX) = b^2\operatorname{var}(X)$ but $\mathbb{E}(a + bX) = a + b\mathbb{E}X$. Measures of location (expected value) and spread (standard deviation) should react differently to linear transformations of the variable. As another example: if a given piece of "information" implies that a random variable X must take the constant value C then $\mathbb{E}(X \mid \text{information}) = C$, but $\operatorname{var}(X \mid \text{information}) = 0$.

It is also a common blunder to confuse the formula for the variance of a difference with the formula $\mathbb{E}(Y - Z) = \mathbb{E}Y - \mathbb{E}Z$. If you ever find yourself wanting to assert that $\operatorname{var}(Y - Z)$ is equal to $\operatorname{var}(Y) - \operatorname{var}(Z)$, think again. What would happen if $\operatorname{var}(Z)$ were larger than $\operatorname{var}(Y)$? Variances can't be negative.

Covariances become useful when we consider variances of sums of random variables. For example, $\operatorname{var}(Y + Z) = \operatorname{var}(Y) + 2\operatorname{cov}(Y, Z) + \operatorname{var}(Z)$, a result that follows by taking expectations of both sides of the expansion

$$(X + Y - \mu_X - \mu_Y)^2 = (X - \mu_X)^2 + 2\,(X - \mu_X)\,(Y - \mu_Y) + (Y - \mu_Y)^2.$$

More generally, for constants a, b, c, d, and random variables U, V, Y, Z,

$$\begin{aligned}&\operatorname{cov}(aU + bV, cY + dZ)\\ &\qquad = ac\operatorname{cov}(U, Y) + bc\operatorname{cov}(V, Y) + ad\operatorname{cov}(U, Z) + bd\operatorname{cov}(V, Z).\end{aligned}$$

It is easier to see the pattern if we work with the centered random variables $U' = U - \mu_U$, $\ldots$, $Z' = Z - \mu_Z$. For then the left-hand side becomes

$$\begin{aligned}\mathbb{E}\big((aU' + bV')(cY' + dZ')\big) &= \mathbb{E}(ac\,U'Y' + bc\,V'Y' + ad\,U'Z' + bd\,V'Z')\\ &= ac\,\mathbb{E}(U'Y') + bc\,\mathbb{E}(V'Y') + ad\,\mathbb{E}(U'Z') + bd\,\mathbb{E}(V'Z').\end{aligned}$$

The expected values in the last line correspond to the four covariances.

If Y and Z are uncorrelated, the covariance term drops out from the expression for the variance of their sum, leaving $\text{var}(Y+Z) = \text{var}(Y) + \text{var}(Z)$. Similarly, if $X_1, \ldots, X_n$ are random variables for which $\text{cov}(X_i, X_j) = 0$ for each $i \neq j$ then

$$\text{var}(X_1 + \ldots + X_n) = \text{var}(X_1) + \ldots + \text{var}(X_n) \qquad \text{for } \textit{pairwise uncorrelated} \text{ rv's.}$$

You should check the last assertion by expanding out the quadratic in the variables $X_i - \mathbb{E}X_i$, observing how all the cross-product terms disappear because of the zero covariances.

There is an enormous body of probability literature that deals with approximations to distributions, and bounds for probabilities, expressible in terms of expected values and variances. One of the oldest and simplest examples, the Tchebychev inequality, is still useful, even though it is rather crude by modern standards.

Example 2: The Tchebychev inequality: $\mathbb{P}\{|X - \mu| \geq \epsilon\} \leq \text{var}(X)/\epsilon^2$, where $\mu = \mathbb{E}X$ and $\epsilon > 0$.

The Tchebychev bound explains an important property of sample means: their distributions concentrate increasingly around their expectations as the sample size increases.

Example 3: Concentration of sample mean about expected value.

The concentration phenomenon will also hold for averages of dependent random variables, if the variance is small.

Example 4: Comparison of spread in sample averages for sampling with and without replacement: the Decennial Census.

As with expectations, variances and covariances can also be calculated conditionally on various pieces of information. The conditioning formula in the final Example has the interpretation of a decomposition of "variability" into distinct sources, a precursor to the statistical technique know as the "analysis of variance".

Example 5: An example to show how variances can sometimes be decomposed into components attributable to difference sources. (Can be skipped.)

Things to remember

- $\mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y)$ if X and Y are independent random variables
- the initial definitions of variance and covariance, and their expanded forms $\text{cov}(Y, Z) = \mathbb{E}(YZ) - (\mathbb{E}Y)(\mathbb{E}Z)$ and $\text{var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
- $\text{var}(a + bX) = b^2\text{var}(X)$ and $\text{sd}(a + bX) = |b|\text{sd}(X)$ for constants a and b.
- For constants a, b, c, d, and random variables U, V, Y, Z,

$$\begin{aligned}&\text{cov}(aU + bV, cY + dZ)\\ &\qquad = ac\,\text{cov}(U, Y) + bc\,\text{cov}(V, Y) + ad\,\text{cov}(U, Z) + bd\,\text{cov}(V, Z).\end{aligned}$$

- Sampling without replacement gives smaller variances than sampling with replacement.