Lecture 30: Covariance, Variance of Sums
Friday, April 20

1 Expectation of Products of Independent Random Variables

We shall assume (unless otherwise mentioned) that the random variables we deal with have expectations. We shall use the notation

$$\mu_X = E[X],$$

where X is a random variable.

Proposition. *Let X and Y be independent random variables on the same probability space. Let g and h be functions of one variable such that the random variables $g(X)$ and $h(Y)$ are defined. Then*

$$E\left[g(X)h(Y)\right] = E\left[g(X)\right] E\left[h(Y)\right].$$

provided that either both X and Y are both discrete or that both X and Y have densities.

Proof. We first assume that X and Y are both discrete. Let p_X and p_Y be their probability mass functions. Since they are independent, their joint probability mass function is given by $p(x,y) = p_X(x)p_Y(y)$. Thus

$$\begin{aligned}
E\left[g(X)h(Y)\right] &= \sum_y \sum_x g(x)h(y)p(x,y) \\
&= \sum_y \sum_x g(x)h(y)p_X(x)p_Y(y) \\
&= \sum_y h(y)p_Y(y) \sum_x g(x)p_X(x) \\
&= \sum_y h(y)p_Y(y)E\left[g(X)\right] \\
&= E\left[g(X)\right] E\left[h(Y)\right]
\end{aligned}$$

Next we assume that X and Y both have densities. Let f_X and f_Y be their densities. Since they are independent, their joint density is given by $f(x,y) = f_X(x)f_Y(y)$. Thus

$$\begin{aligned}
E\left[g(X)h(Y)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} h(y)f_Y(y)\left(\int_{-\infty}^{\infty} g(x)f_X(x)\,dx\right)\,dy \\
&= E\left[g(X)\right]\int_{-\infty}^{\infty} h(y)f_Y(y)\,dy \\
&= E\left[g(X)\right]E\left[h(Y)\right]
\end{aligned}$$

□

Corollary. *If X and Y are independent random variables with expectations, then*

$$E\left[XY\right] = E\left[X\right]E\left[Y\right].$$

2 Covariance

Definition. Let X and Y be random variables defined on the same probability space. Assume that both X and Y have expectations and variances. The **covariance** of X and Y is defined by

$$\mathrm{Cov}\,(X,Y) = E\left[(X-\mu_X)\,(Y-\mu_Y)\right].$$

Proposition (Some Properties of Covariance)**.** *Let X, Y, X_j, $j = 1, \ldots m$, and Y_k, $k = 1 \ldots n$, be random variables with expectations and variances, and assume that they are all defined on the same probability space.*

1. $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$.
2. *If X and Y are independent. then* $\mathrm{Cov}(X,Y) = 0$.
3. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.
4. $\mathrm{Cov}(aX,Y) = a\mathrm{Cov}(X,Y)$. *for $a \in \mathbf{R}$.*
5. $\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$.
6. $\mathrm{Cov}(\sum_{j=1}^{m} X_j, \sum_{k=1}^{n} Y_k) = \sum_{j=1}^{m}\sum_{k=1}^{n} \mathrm{Cov}\,(X_j, Y_k)$.

Proof of 1.

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E\left[(X-\mu_X)\,(Y-\mu_Y)\right] \\
&= E\left[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y\right] \\
&= E\left[XY\right] - \mu_X E\left[Y\right] - \mu_X E\left[Y\right] + \mu_X\mu_Y \\
&= E\left[XY\right] - E\left[X\right]E\left[Y\right]
\end{aligned}$$

□

Proof of 2. Assertion *2* follows from *1* because $E[XY] = E[X]E[Y]$ for independent random variables X and Y. □

Remark. The converse is not true, as the following example shows.

Example 1. Let X be a random variable which satisfies

$$P\{X=-1\}=P\{X=0\}=P\{X=1\}=\frac{1}{3},$$

and let $Y=1-X^2$. Then $E[X]=0$ and $XY=0$. Thus,

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

but X and Y are not independent. Indeed, Y is a function of X.

Proof of 3. Assertion *3* follows immediately from the definition of covariance. $\square$

Proof of 4.

$$\begin{aligned}\mathrm{Cov}(aX,Y) &= E\left[aXY\right]-E[aX]E[Y]\\ &= aE\left[Y\right]\\ &= a\mathrm{Cov}(X,Y)\end{aligned}$$

$\square$

Proof of 5.

$$\begin{aligned}\mathrm{Cov}(X_1+X_2,Y) &= E\left[\left(X_1+X_2\right)Y\right]-E[X_1+X_2]E[Y]\\ &= E\left[X_1Y+X_2Y\right]-\left(E\left[X_1\right]+E\left[X_2\right]\right)E[Y]\\ &= E\left[X_1Y\right]+E\left[X_2Y\right]-E\left[X_1\right]E[Y]-E\left[X_2\right]E[Y]\\ &= \mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)\end{aligned}$$

$\square$

Proof of 6. It follows from *5*, by induction on m, that for any $m \in \mathbf{N}$

$$\operatorname{Cov}\left(\sum_{j=1}^{m} X_j, Y\right) = \sum_{j=1}^{m} \operatorname{Cov}(X_j, Y)$$

Using this and *1*, we see that for any $n \in \mathbf{N}$

$$\begin{aligned}
\operatorname{Cov}\left(X, \sum_{k=1}^{n} Y_k\right) &= \operatorname{Cov}\left(\sum_{k=1}^{n} Y_k, X\right) \\
&= \sum_{k=1}^{n} \operatorname{Cov}(Y_k, X) \\
&= \sum_{k=1}^{n} \operatorname{Cov}(X, Y_k)
\end{aligned}$$

Thus,

$$\begin{aligned}
\operatorname{Cov}\left(\sum_{j=1}^{m} X_j, \sum_{k=1}^{n} Y_k\right) &= \sum_{j=1}^{m} \operatorname{Cov}\left(X_j, \sum_{k=1}^{n} Y_k\right) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n} \operatorname{Cov}(X_j, Y_k)
\end{aligned}$$

□

3 Variance of a Sum

Proposition. *Let $X_1, \ldots, X_n$ be rnadom variables defined on the same probability space and all having expectations and variances. Then*

$$\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{Var}(X_k) + 2\sum_{j<k}\sum \operatorname{Cov}(X_j, X_k)$$

If $X_1, \ldots, X_n$ are pairwise independent,

$$\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{Var}(X_k)$$

Proof.

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) &= \operatorname{Cov}\left(\sum_{j=1}^{n} X_j, \sum_{k=1}^{n} X_k\right) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} \operatorname{Cov}(X_j, X_k) \\
&= \sum_{j=k}\sum \operatorname{Cov}(X_j, X_k) + \sum_{j\neq k}\sum \operatorname{Cov}(X_j, X_k)
\end{aligned}
$$

Each pair, α, β, of indices with $\alpha \neq \beta$ occurs twice in the sum: once as (α, β) and once as (β, α). Since the terms $\operatorname{Cov}(X_\alpha, X_\beta)$ and $\operatorname{Cov}(X_\beta, X_\alpha)$ are equal, we have

$$
\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{Var}(X_k) + 2\sum_{j<k}\sum \operatorname{Cov}(X_j, X_k)
$$

If X_j and X_k are independent, where (necessarily) $j \neq k$, then $\operatorname{Cov}(X_j, X_k) = 0$, and the second formula follows. □

Example 2. Let Y be a Bernoulli random variable with $P\{Y = 1\} = p$. Then $Y^2 = X$. Thus $E[Y] = p$ and $\operatorname{Var}(Y^2) - (E[Y])^2 = p - p^2 = pq$, where $q = 1 - p$.

Example 3. Let X be a binomial random variable with parameters (n, p). Then there exist independent Bernoulli random variables, all with parameter p, such that

$$
X = X_1 + \cdots + X_n.
$$

then

$$
\begin{aligned}
E[X] &= E[X_1] + \cdots + E[X_n] \\
&= np
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{Var}[X] &= \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n) \\
&= npq
\end{aligned}
$$

Remark. Using independent Bernoulli random variables gives us another way to see that if X and Y are independent binomial random variables with respective parameters (n, p) and (m, p), then the sum $X + Y$ is binomial with parameters $(n + m, p)$.

Example 4. Suppose we want to estimate the mean height of ND undergraduate males, that is, the arithmetic mean μ of the heights of all the individuals in this population of $N \approx 4000$. The parameter μ is called the **population mean.** We are also interested in estimating the **population variance** σ^2, that is the arithmetic mean of the squares of the deviations of the individual heights from μ.

We select a random sample of size n of them. (Here random means that each of the $\binom{N}{n}$ possible samples of size n is equally likely.) We measure and record the heights of the individuals in the sample.

Now let X_1 be the height of the first person selected, X_2 the height of the second, and so on. We assume that the sampling is done in such a way that we can treat these random variables as independent. Each X_k has mean μ and variance σ^2. Let

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Now $\bar{X}$ is called the **sample mean**; it is the arithmetic mean of the heights of the individuals in the sample. Now,

$$E\left[X_k\right] = \mu$$

for each k, $k = 1, \ldots, n$. Thus,

$$\begin{aligned} E\left[\bar{X}\right] &= E\left[\frac{1}{n}\sum_{k=1}^{n} X_k\right] \\ &= \frac{1}{n}\sum_{k=1}^{n} E\left[X_k\right] \\ &= \frac{1}{n}\cdot n \cdot \mu \\ &= \mu \end{aligned}$$

Statisticians would summarize this calculation by saying that $\bar{X}$ is an unbiased estimator of the population mean μ. This is one of the (many) reasons that we would use the observed value of $\bar{X}$ as our estimate of the population mean μ.

We also have

$$\begin{aligned}
\operatorname{Var}\left(\bar{X}\right) &= \operatorname{Var}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n^2}\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) \\
&= \frac{1}{n^2}\sum_{k=1}^{n} X_k \operatorname{Var}\left(X_k\right) \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

where σ^2 is the **population variance** Note that the variance of the sample mean decreases as the sample size n increases.

To estimate the population variance, σ^2, we would use the **sample variance:**

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$$

We show that

$$E\left[S^2\right] = \sigma^2.$$

Now

$$\begin{aligned}
(n-1)S^2 &= \sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2 \\
&= \sum_{k=1}^{n}\left(X_k - \mu + \mu - \bar{X}\right)^2 \\
&= \sum_{k=1}^{n}\left((X_k - \mu) + \left(\mu - \bar{X}\right)\right)^2 \\
&= \sum_{k=1}^{n}\left((X_k - \mu)^2 - 2(X_k - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\right) \\
&= \sum_{k=1}^{n}(X_k - \mu)^2 - 2(\bar{X} - \mu)\sum_{k=1}^{n}(X_k - \mu) + n(\bar{X} - \mu)^2
\end{aligned}$$

Since

$$\begin{aligned}\sum_{k=1}^{n}(X_k-\mu) &= \sum_{k=1}^{n}X_k - n\mu \\ &= n\bar{X}-n\mu,\end{aligned}$$

we have

$$\begin{aligned}(n-1)S^2 &= \sum_{k=1}^{n}(X_k-\mu)^2 - 2n(\bar{X}-\mu)^2 + n(\bar{X}-\mu)^2 \\ &= \sum_{k=1}^{n}(X_k-\mu)^2 - n(\bar{X}-\mu)^2\end{aligned}$$

Therefore,

$$\begin{aligned}(n-1)E\left[S^2\right] &= \sum_{k=1}^{n}E\left[(X_k-\mu)^2\right] - nE\left[\left(\bar{X}-\mu\right)^2\right] \\ &= n\sigma^2 - n\mathrm{Var}\left(\bar{X}\right) \\ &= n\sigma^2 - n\frac{\sigma^2}{n} \\ &= (n-1)\sigma^2\end{aligned}$$

Thus the sample variance S^2 is an unbiased estimator of the population variance σ^2.

4 Correlation

Let X and Y be random variables defined on the same probability space, and assume that the expectations and variances of X and Y exist.

Definition. The **correlation** of X and Y is denoted $\rho(X,Y)$ and is defined by

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

provided that $\mathrm{Var}(X) \neq 0 \neq \mathrm{Var}(Y)$.

We remarked earlier that we can think of the covariance as an inner product. This analogy works most precisely if we think of

$$\mathrm{Cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

as the inner product of the inner product of the random variables $X - \mu_X$ and $Y - \mu_Y$, both of which have mean 0. Then the variance is anologous to the square of the length of the vector, and the correlation is analogous to the cosine of the angle between the two vectors.

If we have vectors $\vec{x}$ and $\vec{y}$ with $\vec{x} \neq \vec{0}$, then there is a scalar a and there is a vector $\vec{z}$ such that

$$\begin{aligned} \vec{y} &= a\vec{x} + \vec{z} \\ \vec{z} \cdot \vec{x} &= 0 \end{aligned}$$

Now $|a\vec{x}| = |\cos\theta||\vec{y}|$ and $|\vec{z}| = \sin\theta|\vec{y}|$, where θ is the angle between $\vec{x}$ and $\vec{y}$.

Thus $\cos\theta$ is a measure of the strength of the component of $\vec{y}$ in the direction of $\vec{x}$.

Returning to the random variables X and Y, and assuming that $\mathrm{Var}(X) \neq 0$, we have that $\vec{x}$ corresponds to $X - \mu_X$ and $\vec{y}$ corresponds to $Y - \mu_Y$. Also $\cos\theta$ corresponds to $\rho(X, Y)$. A multiple of $X - \mu_X$ has the form

$$\begin{aligned} a(X - \mu_X) &= aX - a\mu_X \\ &= aX + b \end{aligned}$$

of a linear function of X. Thus $\rho(X, Y)$ measures the extent to which Y is a linear function of X. (Since $\rho(X, Y) = \rho(Y, X)$, this is also the extent to which X is a linear function of y.)

Terminology. If $\rho(X, Y) > 0$, we say that X and Y are *positively correlated.* If $\rho(X, Y) < 0$, we say that X and Y are *negatively correlated.* If $\rho(X, Y) = 0$, we say that X and Y are *uncorrelated.*

Remark. $\rho(X, Y)$ and $\mathrm{Cov}(X, Y)$ have the same sign.

Example 5. Let A be an event with $P(A) > 0$. Let I_A be the indicator of A; that is, $I_A = 1$ if A occurs, and $I_A = 0$ otherwise. Similarly let B be an event (in the same probability space) with $P(B) > 0$, and let I_B be the indicator of B. We shall show that

$$Cov(I_A, I_B) = P(AB) - P(A)P(B).$$

Solution.

$$\begin{aligned} E[I_A] &= P(A) \\ E[I_B] &= P(B) \\ E[I_{AB}] &= P(AB) \end{aligned}$$

Thus

$$\begin{aligned} \mathrm{Cov}(I_A, I_B) &= E[I_A I_B] - E[I_A]E[I_B] \\ &= P(AB) - P(A)P(B) \end{aligned}$$

□

Remark. Now $P(AB) = P(A|B)P(B)$, so

$$\begin{aligned} \mathrm{Cov}(I_A I_B) &= P(A|B)P(B) - P(A)P(B) \\ &= (P(A|B) - P(A))\,P(B) \end{aligned}$$

Thus I_A and I_B are positively correlated if and only if $P(A|B) > P(A)$; negatively correlated if and only if $P(A|B) < P(A)$; and uncorrelated if and only if $P(A|B) = P(A)$, i.e, if and only if A and B are independent.

Note that the argument above is symmetric in A and B.

Example 6. Let's specialize the example above. Consider rolling a fair die twice. Let A be the event that the resulting sum is even and let B be the event that the sum is a perfect square, i.e., 4 or 9.

$$\begin{aligned} \mathrm{Cov}(I_A, I_B) &= E[I_A I_B] - E[I_A]E[I_B] \\ &= \frac{1}{12} - \frac{1}{2}\frac{7}{36} \\ &= -\frac{1}{72} \end{aligned}$$

Thus I_A and I_B are negatively correlated. This reflects the fact that if the sum is a perfect square, it is more likely to be 9 than 4.