Week 13 Recap

Monday November 13 and Wednesday November 15: Expectation, variance, covariance for multivariate RVs. Multivariate normal RVs.

Expectation: Recall that for a multivariate RV $X=(X_1,\dots,X_n)$, $E(X_1+\cdots+X_n)=E(X_1)+\cdots+E(X_n)$. Also, if $Y=g(X_1,\dots,X_n)$ and X is a continuous multivariate RV with density $(x_1,\dots,x_n)\mapsto f_X(x_1,\dots,x_n)$ then

$$E(Y)=E(g(X_1,\dots,X_n))=\int\dots\int g(x_1,\dots,x_n)f_X(x_1,\dots,x_n)dx_1\dots dx_n.$$

Expectation of products: Let us apply the formula given above to a pair $Z=(X,Y)$ for which we want to compute $E(XY)$ or, more generally, $E(X^pY^q)$ where, say, p,q are non-negative integers. The formula gives

$$E(XY)=\iint xyf_Z(x,y)dxdy \text{ and } E(X^pY^q)=\iint x^py^qf_Z(x,y)dxdy.$$

If we know f_Z, we can try to compute these integrals. Something special and important happens if we assume that the RVs X,Y are independent. in this case, $f_Z(x,y)=f_X(x)f_Y(y)$ and

$$E(X^pY^q)=\iint x^py^qf_X(x)f_Y(y)dxdy=\left(\int x^pf_X(x)dx\right)\left(\int y^qf_Y(y)dy\right)=E(X^p)E(Y$$

We use this property to compute the moment generating function of the sum of two independent RVS, X,Y, namely,

$$M_{X+Y}(t)=E(e^{t(X+Y)})=E(e^{tX}e^{tY})=E(e^{tX})E(e^{tY})=M_X(t)M_Y(t)$$

Variance and covariance

Recall that the variance of X is defined, when it exists, by $\mathrm{Var}(X)=E(X^2)-E(X)^2=E\left((X-E(X))^2\right)$. This formula show that $\mathrm{Var}(aX+b)=a^2\mathrm{Var}(X)$. The variance of X gives some indication about how much X can deviate from its mean or average value $E(X)$.

In the case of a sum S of k random variables $X_1,\dots,X_k$,

$$\begin{aligned}\mathrm{Var}(S) &= E((S-E(S))^2)\\ &= \sum_1^k \mathrm{Var}(X_i) + 2\sum_{i<j} E((X_i - E(X_i))(X_j - E(X_j))).\end{aligned}$$

We now give a name to the quantities $E((X_i - E(X_i))(X_j - E(X_j)))$ appearing on the right-hand side of this equation.

Covariance Given a pair of random variables, X, Y (defined on the same probability space (Ω, P)), each with finite variance, their covariance is the quantity $\mathrm{Cov}(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y).$

For the sum $S = \sum_1^k X_i$, this gives $\mathrm{Var}(S) = \sum_1^k \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j).$

Covariance, variance, and independence

- If it exists, the covariance, $\mathrm{Cov}(X,Y)$, of two independent random variables X, Y is equal to 0 .
- For a sum $S = \sum_1^k X_i$ of mutually independent random variables, all with finite variance,

$$\mathrm{Var}(S) = \sum_1^k \mathrm{Var}(X_i).$$

Cauchy-Schwarz inequality, variance and covariance If X and Y each have finite variance then \[\left|\mbox{Cov} (X,Y) \right|\le \sqrt{\mbox{Var}(X)\mbox{Var}(Y)}.\] There is equality in this inequality if and only if there are reals a, b such that $Y = aX + b$.

Proof: Recall from linear algebra that for any two vectors $x = (x_i)_1^n, y = (y_i)_1^n$, we have

$$|x \cdot y| \le \sqrt{x \cdot x}\sqrt{y \cdot y} = \|x\|\|y\|.$$

Indeed, the dot product of the vector $tx + y$ with itself is non-negative, that is, $(tx+y)\cdot(tx+y) \ge 0$. This means that $\|x\|^2 t^2 + 2(x \cdot y)t + \|y\|^2 \ge 0.$ This a second degree polynomial in t and it has either one double real root or no real roots (because it does not change sign). The discriminant $4(x \cdot y)^2 - 4\|y\|^2\|x\|^2$ must be less or equal to zero. This gives exactly $|x \cdot y| \le \|x\|\|y\|$. This proof tells us what happens in the case of equality. Equality can only occur when the polynomial has a double root t_0 at which $(t_0 x + y)\cdot(t_0 x + y) = 0$. This implies $y = -t_0 x$ (the vector y is a multiple of x).

We can now repeat this argument with almost no changes for random variables X, Y as follows:

$$0 \le E((tX+Y)^2) = E(t^2X^2 + 2tXY + Y^2)$$
$$= t^2E(X^2) + 2tE(XY) + E(Y^2).$$

The right hand-side is a degree two polynomial in t which is aways non-negative. This means it has either one double real root or no real roots. Its discriminant $4E(XY)^2 - 4E(X^2)E(Y^2)$ must be less or equal to zero. This gives $|E(XY)| \le \sqrt{E(X^2)E(Y^2)}$ with equality if and only if there is t_0 such that $Y = -t_0X$. Now, it suffices to apply this inequality to the variables $X - E(X), Y - E(Y)$ to get

$$|\mathrm{Cov}(X,Y)| \le \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}.$$

Equality occurs if and only if there are reals a, b such that $Y = aX + b$. Here b must be equal to $E(Y) - aE(X)$.

Correlation: The ratio

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

is called the correlation coefficient of X and Y. It is always between -1 and 1 included (see above). It provides a measure of the correlation between X and Y.

Correlation equals to \$1\$ or \$-1\$: If X, Y have finite variance and correlation coefficient equal to 1 then there is $a > 0$ and $b \in \mathbb{R}$ such that $Y = aX + b$. If instead the correlation coefficient of X, Y equals \$-1\$ then there is $a < 0$ and $b \in \mathbb{R}$ such that $Y = aX + b$.

Example: Consider a multinomial random variable $X = (X_1, \dots, X_k)$ with parameters n and $p_1, \dots, p_k$, $\sum_1^k p_i = 1$. Compute the covariances $\mathrm{Cov}(X_i, X_j)$ and the associated correlation coefficients, $1 \le i, j \le k$.

Recall that the mass distribution function of this random vector is

$$p_{x_1,\dots,x_k} = p_1^{x_1} \dots p_k^{x_k} \binom{n}{x_1, \dots, x_k}.$$

To facilitate computations, we can think of a sequence of independent identical experiments with k possible outcomes (type 1,..., type k) occurring with probability $p_1, \dots, p_k$. Each X_i counts how many outcomes of type i occurs in the sequence so that we see that each X_i (the i-th marginal) is binomial with parameter n, p_i. It follows that

$$E(X_i) = np_i \quad \text{and} \quad \mathrm{Var}(X_i) = np_i(1-p_i).$$

Let's compute $E(X_1X_2)$ which is

$$E(X_1X_2) = \sum_{x_1,\dots,x_k} x_1x_2p_1^{x_1}p_2^{x_2}p_3^{x_3}\dots p_k^{x_k}\binom{n}{x_1,\dots,x_k}.$$

Consider the function

$$(p,q) \mapsto \sum_{x_1,\dots,x_k} p^{x_1}q^{x_2}p_3^{x_3}\dots p_k^{x_k}\binom{n}{x_1,\dots,x_k}.$$

By the multinomial theorem it is equal to

$$(p,q) \mapsto (p+q+p_3+\cdots+p_k)^n = \sum_{x_1,\dots,x_k} p^{x_1}q^{x_2}p_3^{x_3}\dots p_k^{x_k}\binom{n}{x_1,\dots,x_k}$$

and its mixed partial derivative of order 2,

$$\frac{\partial^2}{\partial q\partial p}(p+q+p_3+\cdots+p_k)^n = n(n-1)(p+q+p_3+\cdots+p_k)^{n-2}$$

is also equal to

$$\sum_{x_1,\dots,x_k} x_1x_2p^{x_1-1}q^{x_2-1}p_3^{x_3}\dots p_k^{x_k}\binom{n}{x_1,\dots,x_k}.$$

This gives that

$$\begin{aligned}E(X_1X_2) &= \sum_{x_1,\dots,x_k} x_1x_2p_1^{x_1}p_2^{x_2}p_3^{x_3}\dots p_k^{x_k}\binom{n}{x_1,\dots,x_k}\\ &= p_1p_2\sum_{x_1,\dots,x_k} x_1x_2p_1^{x_1-1}p_2^{x_2-1}p_3^{x_3}\dots p_k^{x_k}\binom{n}{x_1,\dots,x_k}\\ &= p_1p_2n(n-1)(p_1+p_1+p_3+\cdots+p_k)^{n-2}\end{aligned}$$
$$= n(n-1)p_1p_2.$$

The last equality is because $p_1+\cdots+p_k = 1$, by definition.

The same argument gives

$$E(X_iX_j) = n(n-1)p_ip_j, \quad 1 \le i \le j \le k.$$

So, we can now compute

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\ &= n(n-1)p_i p_j - n^2 p_i p_j = -np_i p_j.\end{aligned}$$

The correlation coefficient between X_i and X_j is

$$\begin{aligned}\mathrm{Corr}(X_i, X_j) &= -\frac{np_i p_j}{\sqrt{np_i(1-p_i)}\sqrt{np_j(1-p_j)}} \\ &= -\frac{\sqrt{p_i p_j}}{\sqrt{(1-p_i)(1-p_j)}}.\end{aligned}$$

Consider what happens when $k = 2$ and $i = 1, j = 2$. Because $p_1 + p_2 = 1$, $p_i = 1 - p_j$ in this case and the formula above tells us that the correlation $\mathrm{Corr}(X_1, X_2)$ is equal -1. To explain why this is the case, observe that when $k = 2$ there are only two possible outcomes for each experiment. Because X_1 counts how many time the experiment ends with a type 1 outcome, and X_2 counts how many time the experiment ends with a type 2 outcome, and there are n experiments performed, $X_1 + X_2 = n$, that is, $X_2 = n - X_1$. Now suppose that there are three possible types of outcomes ($k = 3$) but the third type occurs only with very small probability compared to the first two types (that is $p_3 \lll p_1$ and $p_3 \lll p_2$). In this case, it is almost true that $p_1 = 1 - p_2$ because $p_1 = 1 - p_2 - p_3$ and $p_3 \lll p_2$ is very small. So the correlation coefficient between X_1 and X_2 is close to -1.