ECE 361 Probability for Engineers (Fall, 2016)
Lecture 9a

§4.2 Covariance and correlation

The covariance of two RVs, X, Y, is defined as:

$$\mathrm{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]. \quad (1)$$

Say X, Y are uncorrelated when $\mathrm{cov}(X,Y) = 0$. A large positive covariance indicates that RV X tends to be large when Y is large, and a large negative covariance indicates that RV X tends to be large when Y is small. See Fig. 4.11 for an illustration. Key properties include:

$$\begin{aligned} \mathrm{cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ \mathrm{cov}(X,X) &= \mathrm{var}(X) \\ \mathrm{cov}(X, aY+b) &= a\mathrm{cov}(X,Y) \\ \mathrm{cov}(X, Y+Z) &= \mathrm{cov}(X,Y) + \mathrm{cov}(X,Z). \end{aligned} \quad (2)$$

If RVs X, Y are independent then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, which means $\mathrm{cov}(X,Y) = 0$. Thus independence implies uncorrelatedness, but not vice versa, i.e., uncorrelated RVs may be dependent.

Example. The pair of RVs (X,Y) has PMF:

$$p_{X,Y}(1,0) = p_{X,Y}(0,1) = p_{X,Y}(-1,0) = p_{X,Y}(0,-1) = \frac{1}{4}, \quad (3)$$

as shown in Fig. 4.12. Clearly $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ by symmetry. Further, $XY = 0$ for all possible pairs (X,Y), and thus $\mathbb{E}[XY] = 0$. Hence X, Y are uncorrelated. But they are clearly not independent since knowing $X = -1$ (say) guarantees $Y = 0$.

The correlation coefficient is simply a normalization of the covariance:

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}. \quad (4)$$

One can show that $\rho \in [-1,1]$. Then $\rho < 0$ means $X - \mathbb{E}[X], Y - \mathbb{E}[Y]$ tend have opposite sign ($X < \mathbb{E}[X]$ suggests $Y > \mathbb{E}[Y]$) and $\rho > 0$ means $X - \mathbb{E}[X], Y - \mathbb{E}[Y]$ tend have the same sign ($X < \mathbb{E}[X]$ suggests $Y < \mathbb{E}[Y]$). In the extreme case of $\rho \in \{-1,1\}$ then Y is a linear function of X: there exists a c such that $Y - \mathbb{E}[Y] = c(X - \mathbb{E}[X])$.

Example. Let (Z_1, Z_2, Z_3) be independent standard normal RVs, and define the pair of RVs (X,Y) where $X = Z_1 + Z_2$ and $Y = Z_1 + Z_3$. Find the correlation of (X,Y).

First observe $X \sim N(0, \sqrt{2})$ and $Y \sim N(0, \sqrt{2})$ and that (X,Y) are not independent, as they have a shared dependence on Z_1. The covariance is:

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] \\ &= \mathbb{E}[(Z_1 + Z_2)(Z_1 + Z_3)] \\ &= \mathbb{E}[Z_1^2 + Z_1 Z_2 + Z_1 Z_3 + Z_2 Z_3] \\ &= \mathbb{E}[Z_1]^2 + \mathbb{E}[Z_1 Z_2] + \mathbb{E}[Z_1 Z_3] + \mathbb{E}[Z_2 Z_3] \\ &= \mathrm{Var}(Z_1) + \mathbb{E}[Z_1]\mathbb{E}[Z_2] + \mathbb{E}[Z_1]\mathbb{E}[Z_3] + \mathbb{E}[Z_2]\mathbb{E}[Z_3] \\ &= 1. \end{aligned} \quad (5)$$

The correlation is therefore

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Std}(X)\mathrm{Std}(Y)} = \frac{1}{\sqrt{2}\sqrt{2}} = \frac{1}{2}. \tag{6}$$

Example. Let X, Y be RVs with the same variance. Show that $(X-Y, X+Y)$ are uncorrelated.

Consider the covariance. Define $\tilde{X} = X - \mathbb{E}[X]$ and $\tilde{Y} - \mathbb{E}[Y]$. Then:

$$\begin{aligned}
\mathrm{Cov}(X-Y, X+Y) &= \mathbb{E}[(X - Y - (\mathbb{E}[X] - \mathbb{E}[Y]))(X + Y - (\mathbb{E}[X] + \mathbb{E}[Y]))] \\
&= \mathbb{E}[(\tilde{X} - \tilde{Y})(\tilde{X} + \tilde{Y})] \\
&= \mathbb{E}[\tilde{X}^2 - \tilde{X}\tilde{Y} + \tilde{X}\tilde{Y} - \tilde{Y}^2] \\
&= \mathbb{E}[\tilde{X}^2] - \mathbb{E}[\tilde{Y}^2]
\end{aligned} \tag{7}$$

As $\mathbb{E}[\tilde{X}] = \mathbb{E}[\tilde{Y}] = 0$ it follows that $\mathrm{Var}(\tilde{X}) = \mathbb{E}[\tilde{X}^2]$ and $\mathrm{Var}(\tilde{Y}) = \mathbb{E}[\tilde{Y}^2]$. As $\mathrm{Var}(X) = \mathrm{Var}(\tilde{X})$ and $\mathrm{Var}(Y) = \mathrm{Var}(\tilde{Y})$, and $\mathrm{Var}(X) = \mathrm{Var}(Y)$ by assumption, it follows that $\mathrm{Cov}(X-Y, X+Y) = 0$.

The natural extension of the result given earlier, $\mathrm{cov}(X, Y+Z) = \mathrm{cov}(X,Y) + \mathrm{cov}(X,Z)$, is that

$$\mathrm{cov}(X_1 + \cdots + X_k, Y_1 + \cdots + Y_l) = \sum_{i=1}^{k}\sum_{j=1}^{l} \mathrm{cov}(X_i, Y_j). \tag{8}$$

This result is used below.

Example. The multinomial distribution is a generalization of the binomial distribution. Recall $X \sim \mathrm{bin}(n,p)$ may be thought of as the number of successes in n independent Bernoulli trials, each of which succeeds with probability p or fails with probability $1-p$. Recall the binomial PMF

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \ x \in \{0, \ldots, n\}. \tag{9}$$

The multinomial generalizes the binomial from trials with two outcomes to trials with k outcomes. Let $(p_1, \ldots, p_k)$ be the outcome distribution, with p_j the probability that a random trial has outcome j for $j \in [k]$, and naturally $p_1 + \cdots + p_k = 1$. Now the outcome of n independent trials is reported as the vector of counts, meaning the number of trials that resulted in each possible outcome. Thus $(X_1, \ldots, X_k) \sim \mathrm{mult}(n, p_1, \ldots, p_k)$ has support $\mathcal{S}_{k,n} = \{(x_1, \ldots, x_k) \in \mathbb{Z}_+ : x_1 + \cdots + x_k = n\}$, with

$$\mathbb{P}((X_1, \ldots, X_k) = (x_1, \ldots, x_k)) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k}, \ (x_1, \ldots, x_k) \in \mathcal{S}_{k,n}. \tag{10}$$

Recall that $\binom{n}{x_1,\ldots,x_k} = \frac{n!}{x_1! \cdots x_k!}$ is the *multinomial coefficient*, the generalization of the *binomial coefficient*. Fix two distinct possible outcomes, say i and j, with $1 \leq i < j \leq k$, and let (X_i, X_j) be the number of trials resulting in those two outcomes. Find the correlation $\rho(X_i, X_j)$.

We first find the mean and variance of any one component, say X_i, of $(X_1, \ldots, X_k)$. Think of each outcome as a "bin", so there are k bins, and each of the n trials is a ball, so the experiment consists of throwing n balls into the k bins, where each ball is thrown in bin i with probability p_i. Observe that by aggregating together all bins aside from i we recover a binomial distribution with success probability p_i, and as such $\mathbb{E}[X_i] = np_i$ and $\mathrm{Var}(X_i) = np_i(1-p_i)$. Similarly, $\mathbb{E}[X_j] = np_j$ and $\mathrm{Var}(X_j) = np_j(1-p_j)$.

(TO BE CONTINUED NEXT LECTURE)

References

[1] *Introduction to Probability, 2nd Edition* by Dimitri P. Bertsekas and John N. Tsitsiklis, Athina Scientific Press, 2008.