CS 70 Discrete Mathematics and Probability Theory
Summer 2019 Course Notes **Note 16**

Random Variables: Covariance

In this note, we will cover covariance, which is a measure of how related two random variables are. Before we can jump into covariance, we first discuss independent random variables, which have "no relation" to each other; they have no covariance.

1 Sum of Independent Random Variables

One of the most important and useful facts about variance is that if a random variable X is the sum of *independent* random variables $X = X_1 + \cdots + X_n$, then its variance is the sum of the variances of the individual r.v.'s. In particular, if the individual r.v.'s X_i are identically distributed (i.e., they have the same distribution), then $\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(X_i) = n \cdot \text{Var}(X_1)$. This means that the standard deviation is $\sigma(X) = \sqrt{n} \cdot \sigma(X_1)$. Note that by contrast, the expected value is $\mathbb{E}[X] = n \cdot \mathbb{E}[X_1]$. Intuitively this means that whereas the average value of X grows proportionally to n, the spread of the distribution grows proportionally to $\sqrt{n}$, which is much smaller than n. In other words the distribution of X tends to concentrate around its mean.

Let us now formalize these ideas. First, we have the following result which states that the expected value of the product of two independent random variables is equal to the product of their expected values.

Theorem 16.1. *For* independent *random variables* X, Y, *we have* $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Proof. We have

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_a \sum_b ab \times \mathbb{P}[X = a, Y = b] \\
&= \sum_a \sum_b ab \times \mathbb{P}[X = a] \times \mathbb{P}[Y = b] \\
&= \left(\sum_a a \times \mathbb{P}[X = a]\right) \times \left(\sum_b b \times \mathbb{P}[Y = b]\right) \\
&= \mathbb{E}[X] \times \mathbb{E}[Y],
\end{aligned}$$

as required. In the second line here we made crucial use of independence. □

We now use the above theorem to conclude the nice property that the variance of the sum of independent random variables is equal to the sum of their variances.

Theorem 16.2. *For* independent *random variables* X, Y, *we have*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Proof. From the alternative formula for variance in Theorem **??** and linearity of expectation, we have

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}[X+Y])^2 \\ &= \mathbb{E}\left[X^2\right] + \mathbb{E}\left[Y^2\right] + 2\,\mathbb{E}[XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\ &= (\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2) + (\mathbb{E}\left[Y^2\right] - \mathbb{E}[Y]^2) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]).\end{aligned}$$

Since X,Y are independent, Theorem 16.1 implies that the final term in this expression is zero. □

It is very important to remember that **neither** of the above two results is true in general when X,Y are not independent. As a simple example, note that even for a $\{0,1\}$-valued r.v. X with $\mathbb{P}[X=1]=p$, $\mathbb{E}\left[X^2\right]=p$ is not equal to $\mathbb{E}[X]^2 = p^2$ (because of course X and X are not independent!). This is in contrast to linearity of expectation, where we saw that the expectation of a sum of r.v.'s is the sum of the expectations of the individual r.v.'s, regardless of whether or not the r.v.'s are independent.

Example

Let us return to our motivating example of a sequence of n coin tosses. Let X_n denote the number of Heads in n tosses of a biased coin with Heads probability p (i.e., $X_n \sim \mathrm{Binomial}(n,p)$). As usual, we write $X_n = I_1 + I_2 + \cdots + I_n$, where $I_i = 1$ if the i-th toss is H, and $I_i = 0$ otherwise.

We already know $\mathbb{E}[X_n] = \sum_{i=1}^n \mathbb{E}[I_i] = np$. We can compute $\mathrm{Var}(I_i) = \mathbb{E}\left[I_i^2\right] - \mathbb{E}[I_i]^2 = p - p^2 = p(1-p)$. Since the I_i's are independent, by Theorem 16.2 we get $\mathrm{Var}(X_n) = \sum_{i=1}^n \mathrm{Var}(I_i) = np(1-p)$.

As an example, for a fair coin ($p = \frac{1}{2}$) the expected number of Heads in n tosses is $\frac{n}{2}$, and the standard deviation is $\sqrt{\frac{n}{4}} = \frac{\sqrt{n}}{2}$. Note that since the maximum number of Heads is n, the standard deviation is much less than this maximum number for large n. This is in contrast to the previous example of the uniformly distributed random variable (**??**), where the standard deviation $\sigma(X) = \sqrt{\frac{n^2-1}{12}} \approx \frac{n}{\sqrt{12}}$ (for large n) is of the same order as the largest value, n. In this sense, the spread of a binomially distributed r.v. is much smaller than that of a uniformly distributed r.v.

2 Covariance and Correlation

The expression $\mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$ in the proof of Theorem 16.2 is a measure of association between X,Y, and is called the *covariance*:

Definition 16.1 (Covariance)**.** *The* covariance *of random variables X and Y, denoted* $\mathrm{Cov}(X,Y)$*, is defined as*

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

where $\mu_X = \mathbb{E}[X]$ *and* $\mu_Y = \mathbb{E}[Y]$.

Remarks. We note some important facts about covariance.

1. If X,Y are independent, then $\mathrm{Cov}(X,Y) = 0$. However, the converse is **not** true.

2. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

3. Covariance is *bilinear*; i.e., for any collection of random variables $\{X_1,\ldots,X_n\},\{Y_1,\ldots,Y_m\}$ and fixed constants $\{a_1\ldots,a_n\},\{b_1,\ldots,b_m\}$,

$$\mathrm{Cov}(\textstyle\sum_{i=1}^n a_iX_i,\sum_{j=1}^m b_jY_j)=\sum_{i=1}^n\sum_{j=1}^m a_ib_j\mathrm{Cov}(X_i,Y_j).$$

For general random variables X and Y,

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

While the sign of $\mathrm{Cov}(X,Y)$ is informative of how X and Y are associated, its magnitude is difficult to interpret. A statistic that is easier to interpret is *correlation*:

Definition 16.2 (Correlation)**.** *Suppose X and Y are random variables with $\sigma(X)>0$ and $\sigma(Y)>0$. Then, the correlation of X and Y is defined as*

$$\mathrm{Corr}(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Correlation is more useful than covariance because the former always ranges between -1 and $+1$, as the following theorem shows:

Theorem 16.3. *For any pair of random variables X and Y with $\sigma(X)>0$ and $\sigma(Y)>0$,*

$$-1\le\mathrm{Corr}(X,Y)\le+1.$$

Proof. Let $\mathbb{E}[X]=\mu_X$ and $\mathbb{E}[Y]=\mu_Y$, and define $\widetilde{X}=(X-\mu_X)/\sigma(X)$ and $\widetilde{Y}=(Y-\mu_Y)/\sigma(Y)$. Then, $\mathbb{E}[\widetilde{X}^2]=\mathbb{E}[\widetilde{Y}^2]=1$, so

$$\begin{aligned}0&\le\mathbb{E}[(\widetilde{X}-\widetilde{Y})^2]=\mathbb{E}[\widetilde{X}^2]+\mathbb{E}[\widetilde{Y}^2]-2\mathbb{E}[\widetilde{X}\widetilde{Y}]=2-2\mathbb{E}[\widetilde{X}\widetilde{Y}]\\0&\le\mathbb{E}[(\widetilde{X}+\widetilde{Y})^2]=\mathbb{E}[\widetilde{X}^2]+\mathbb{E}[\widetilde{Y}^2]+2\mathbb{E}[\widetilde{X}\widetilde{Y}]=2+2\mathbb{E}[\widetilde{X}\widetilde{Y}],\end{aligned}$$

which implies $-1\le\mathbb{E}[\widetilde{X}\widetilde{Y}]\le+1$. Now, noting that $\mathbb{E}[\widetilde{X}]=\mathbb{E}[\widetilde{Y}]=0$, we obtain $\mathrm{Corr}(X,Y)=\mathrm{Cov}(\widetilde{X},\widetilde{Y})=\mathbb{E}[\widetilde{X}\widetilde{Y}]$. Hence, $-1\le\mathrm{Corr}(X,Y)\le+1$. □

Note that the above proof shows that $\mathrm{Corr}(X,Y)=+1$ if and only if $\mathbb{E}[(\widetilde{X}-\widetilde{Y})^2]=0$, which implies $\widetilde{X}=\widetilde{Y}$ with probability 1. Similarly, $\mathrm{Corr}(X,Y)=-1$ if and only if $\mathbb{E}[(\widetilde{X}+\widetilde{Y})^2]=0$, which implies $\widetilde{X}=-\widetilde{Y}$ with probability 1. In terms of the original random variables X,Y, this means the following: if $\mathrm{Corr}(X,Y)=\pm1$, then there exist constants a and b such that, with probability 1,

$$Y=aX+b,$$

where $a>0$ if $\mathrm{Corr}(X,Y)=+1$ and $a<0$ if $\mathrm{Corr}(X,Y)=-1$.