MATH 170B LECTURE NOTE 2: COVARIANCE AND CORRELATION

HANBAEK LYU

2. How can we quantify dependence between RVs?

2.1. **Covariance.** When two RVs X and Y are independent, we know that the pair (X,Y) is distributed according to the product distribution $\mathbb{P}((X,Y)=(x,y))=\mathbb{P}(X=x)\mathbb{P}(Y=y)$ and we can say a lot of things about their sum, difference, product, maximum, etc. For instance, the expectation of their product is the product of their expectations:

Exercise 2.1. Let X and Y be two independent RVs. Show that $\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$.

But what if they are not independent? Then their joint distribution $\mathbb{P}((X,Y)=(x,y))$ can be very much different from the product distribution $\mathbb{P}(X=x)\mathbb{P}(Y=y)$. Covariance is the quantity that measures the 'average disparity' between the true joint distribution $\mathbb{P}((X,Y)=(x,y))$ and the product distribution $\mathbb{P}(X=x)\mathbb{P}(Y=y)$.

Definition 2.2 (Covariance)**.** *Given two RVs X and Y, their covariance is denoted by* $\text{Cov}(X,Y)$ *and is defined by*

$$\text{Cov}(X,Y)=\mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y). \tag{1}$$

We say X and Y are correlated (resp., uncorrelated) if $\text{Cov}(X,Y)\neq 0$ *(resp.,* $\text{Cov}(X,Y)=0$*).*

Exercise 2.3. Show the following.

(i) $\text{Cov}(X,X)=\text{Var}(X)$.
(ii) $\text{Cov}(X,Y)=\mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))]$.

Exercise 2.4. Show that two RVs X and Y are uncorrelated if they are independent.

Example 2.5 (Uncorrelated but dependent)**.** Two random variables can be uncorrelated but still be dependent. Let (X,Y) be a uniformly sampled point from the unit circle in the 2-dimensional plane. Parameterize the unit circle by $S^1=\{(\cos\theta,\sin\theta)\,|\,0\le\theta<2\pi\}$. Then we can first sample a uniform angle $\Theta\sim\text{Uniform}([0,2\pi))$, and then define $(X,Y)=(\cos\Theta,\sin\Theta)$. Recall from your old memory that

$$\sin 2t = 2\cos t\sin t. \tag{2}$$

Now

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(\cos\Theta\sin\Theta) &(3)\\
&= \frac{1}{2}\mathbb{E}(\sin 2\Theta) &(4)\\
&= \frac{1}{2}\int_0^{2\pi}\sin 2t\,dt &(5)\\
&= \frac{1}{2}\left[-\frac{1}{2}\cos 2t\right]_0^{2\pi}=0. &(6)
\end{aligned}$$

On the other hand,

$$\mathbb{E}(X)=\mathbb{E}(\cos\Theta)=\int_0^{2\pi}\cos t\,dt=0 \tag{7}$$

and likewise $\mathbb{E}(Y)=0$. This shows $\text{Cov}(X,Y)=0$, so X and Y are uncorrelated. However, they satisfy the following deterministic relation

$$X^2+Y^2=1, \tag{8}$$

so clearly they cannot be independent.

So if uncorrelated RVs can be dependent, what does the covariance really measure? It turns out, $\text{Cov}(X,Y)$ measures the 'linear tendency' between X and Y.

Example 2.6 (Linear transform)**.** Let X be a RV, and define another RV Y by $Y = aX + b$ for some constants $a, b \in \mathbb{R}$. Let's compute their covariance using linearity of expectation.

$$\begin{aligned}
\text{Cov}(X,Y) &= \text{Cov}(X, aX+b) && (9)\\
&= \mathbb{E}(aX^2 + bX) - \mathbb{E}(X)\mathbb{E}(aX+b) && (10)\\
&= a\mathbb{E}(X^2) + b\mathbb{E}(X) - \mathbb{E}(X)(a\mathbb{E}(X)+b) && (11)\\
&= a[\mathbb{E}(X^2) - \mathbb{E}(X)^2] && (12)\\
&= a\text{Var}(X). && (13)
\end{aligned}$$

Thus, $\text{Cov}(X, aX+b) > 0$ if $a > 0$ and $\text{Cov}(X, aX+b) < 0$ if $a < 0$. In other words, if $\text{Cov}(X,Y) > 0$, then X and Y tend to be large at the same time; if $\text{Cov}(X,Y) > 0$, then Y tends to be small if X tends to be large.

From the above example, it is clear that why the x- and y-coordinates of a uniformly sampled point from the unit circle are uncorrelated – they have no linear relation!

Exercise 2.7 (Covariance is symmetric and bilinear)**.** Let X and Y be RVs and fix constants $a, b \in \mathbb{R}$. Show the following.

(i) $\text{Cov}(aX+b, Y) = a\text{Cov}(X,Y)$.
(ii) $\text{Cov}(X+Z, Y) = \text{Cov}(X,Y) + \text{Cov}(Z,Y)$.
(iii) $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.

Next, let's say four RVs X, Y, Z, and W are given. Suppose that $\text{Cov}(X,Y) > \text{Cov}(Z,W) > 0$. Can we say that 'the positive linear relation' between X and Y is stronger than that between Z and W? Not quite.

Example 2.8. Suppose X is a RV. Let $Y = 2X$, $Z = 2X$, and $W = 4X$. Then

$$\text{Cov}(X,Y) = \text{Cov}(X, 2X) = 2\,\text{Var}(X), \tag{14}$$

and

$$\text{Cov}(Z,W) = \text{Cov}(2X, 4X) = 8\,\text{Var}(X). \tag{15}$$

But $Y = 2X$ and $W = 2Z$, so the linear relation between the two pairs should be same.

So to compare the magnitude of covariance, we first need to properly normalize covariance so that the effect of fluctuation (variance) of each coordinate is not counted: then only the correlation between the two coordinates will contribute. This is captured by the following quantity.

Definition 2.9 (Correlation coefficient)**.** *Given two RVs X and Y, their correlation coefficient $\rho(X,Y)$ is defined by*

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}\,Y}}. \tag{16}$$

Example 2.10. Suppose X is a RV and fix constants $a, b \in \mathbb{R}$. Then

$$\rho(X, aX+b) = \frac{a\text{Cov}(X,X)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(aX+b)}} = \frac{a\,\text{Var}(X)}{\sqrt{\text{Var}(X)}\sqrt{a^2\,\text{Var}(X)}} = \frac{a}{|a|} = \text{sign}(a). \tag{17}$$

Exercise 2.11 (Cauchy-Schwarz inequality)**.** Let X, Y are RVs. Suppose $\mathbb{E}(Y^2) > 0$. We will show that the 'inner product' of X and Y is at most the product of their 'magnitudes'

(i) For any $t \in \mathbb{R}$, show that

$$\mathbb{E}[(X - tY)^2] = t^2\mathbb{E}(Y^2) - 2t\mathbb{E}(XY) + \mathbb{E}(X^2) \tag{18}$$

$$= \mathbb{E}(Y^2)\left(t - \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}\right)^2 + \frac{\mathbb{E}(X^2)\mathbb{E}(Y^2) - \mathbb{E}(XY)^2}{\mathbb{E}(Y^2)}. \tag{19}$$

Conclude that

$$0 \le \mathbb{E}\left[\left(X - \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y\right)^2\right] = \frac{\mathbb{E}(X^2)\mathbb{E}(Y^2) - \mathbb{E}(XY)^2}{\mathbb{E}(Y^2)}. \tag{20}$$

(ii) Show that a RV Z satisfies $\mathbb{E}(Z^2) = 0$ if and only if $\mathbb{P}(Z = 0) = 1$.
(iii) Show that

$$\mathbb{E}(XY) \le \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}, \tag{21}$$

where the equality holds if and only if

$$\mathbb{P}\left(X = \frac{\mathbb{E}(XY)}{\mathbb{E}(Y^2)}Y\right) = 1. \tag{22}$$

Exercise 2.12. Let X, Y are RVs such that $\text{Var}(Y) > 0$. Let $\bar{X} = X - \mathbb{E}(X)$ and $\bar{Y} = Y - \mathbb{E}(X)$.
(i) Use (21) to show that

$$0 \le \mathbb{E}\left[\left(\bar{X} - \frac{\text{Cov}(X,Y)}{\text{Var}(Y)}\bar{Y}\right)^2\right] = \text{Var}(X)\left(1 - \rho(X,Y)^2\right). \tag{23}$$

(ii) Show that $|\rho(X,Y)| \le 1$.
(iii) Show that $|\rho(X,Y)| = 1$ if and only if $\bar{X} = a\bar{Y}$ for some constant $a \neq 0$.

2.2. **Variance of sum of RVs.** Let X, Y be RVs. If they are not necessarily independent, what is the variance of their sum? Using linearity of expectation, we compute

$$\text{Var}(X+Y) = \mathbb{E}[(X+Y)^2] - \mathbb{E}(X+Y)^2 \tag{24}$$

$$= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}(X) + \mathbb{E}(Y))^2 \tag{25}$$

$$= [\mathbb{E}(X^2) - \mathbb{E}(X)^2] + [\mathbb{E}(Y^2) - \mathbb{E}(Y)^2] + 2[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] \tag{26}$$

$$= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y). \tag{27}$$

Note that $\text{Cov}(X,Y)$ shows up in this calculation. We can push this computation for sum of more than just two RVs.

Proposition 2.13. *For RVs $X_1, X_2 \cdots, X_n$, we have*

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2\sum_{1\le i,j\le n} \text{Cov}(X_i, X_j). \tag{28}$$

Proof. By linearity of expectation, we have

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] - \left(\mathbb{E}\left[\sum_{i=1}^n X_i\right]\right)^2 \tag{29}$$

$$= \mathbb{E}\left[\sum_{1\le i,j\le n} X_iX_j\right] - \sum_{1\le i,j\le n} \mathbb{E}(X_i)\mathbb{E}(X_j) \tag{30}$$

$$= \left[\sum_{1\le i,j\le n} \mathbb{E}(X_iX_j)\right] - \sum_{1\le i,j\le n} \mathbb{E}(X_i)\mathbb{E}(X_j) \tag{31}$$

$$= \sum_{1\le i,j\le n} \left[\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\right] \tag{32}$$

$$= \sum_{1\le i\le n} [\mathbb{E}(X_iX_i) - \mathbb{E}(X_i)\mathbb{E}(X_i)] + \sum_{1\le i\neq j\le n} \left[\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\right] \tag{33}$$

$$= \sum_{1\le i\le n} \mathrm{Var}(X_i^2) + 2\sum_{1\le i<j\le n} \left[\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)\right] \tag{34}$$

$$= \sum_{1\le i\le n} \mathrm{Var}(X_i^2) + 2\sum_{1\le i<j\le n} \mathrm{Cov}(X_i, X_j). \tag{35}$$

□

Exercise 2.14. Show that for independent RVs $X_1, X_2, \cdots, X_n$, we have

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i). \tag{36}$$

Example 2.15 (Number of fixed point in a random permutation)**.** Suppose n people came to a party and somehow the host mixed up their car keys and gave them back completely randomly at the end of the party. Let X_i be a RV, which takes value 1 if person i got the right key and 0 otherwise. Let $N_n = X_1 + X_2 + \cdots + X_n$ be the total number of people who got their own keys back. We will show that $\mathbb{E}(N_n) = \mathrm{Var}(N_n) = 1$.

First, we observe that each $X_i \sim \mathrm{Bernoulli}(1/n)$. So we know that $\mathbb{E}(X_i) = 1/n$ and $\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \mathbb{E}(X_i) - \mathbb{E}(X_i)^2 = n^{-1} - n^{-2} = (n-1)/n^2$. Clearly X_i's are not independent: If the first person got the key number 2, then the second person will never get the right key.

A very important fact is that the linearity of expectation holds regardless of dependence (c.f. Exercise 1.8 in Note 0), so

$$\mathbb{E}[N_n] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n \frac{1}{n} = 1. \tag{37}$$

On the other hand, to compute the covariance, let's take a look at $\mathbb{E}(X_1X_2)$. Note that if the first person got her key, then the second person gets his key with probability $1/(n-1)$. So

$$\mathbb{E}(X_1X_2) = 1\cdot\mathbb{P}(X_1 = 1,\, X_2 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1\,|\,X_1 = 1) = \frac{1}{n}\cdot\frac{1}{n-1}. \tag{38}$$

Hence we can compute their covariance:

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{n-(n-1)}{n^2(n-1)} = \frac{1}{n^2(n-1)}. \tag{39}$$

Since there is nothing special about the pair (X_1, X_2), we get

$$\mathrm{Var}(N_n) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\le i,j\le n} \mathrm{Cov}(X_i, X_j) \tag{40}$$

$$= \sum_{i=1}^n \frac{n-1}{n^2} + 2\sum_{1\le i,j\le n} \frac{1}{n^2(n-1)} \tag{41}$$

$$= \frac{n-1}{n} + 2\binom{n}{2}\frac{1}{n^2(n-1)} \tag{42}$$

$$= \frac{n-1}{n} + 2\frac{n(n-1)}{2!}\frac{1}{n^2(n-1)} \tag{43}$$

$$= \frac{n-1}{n} + \frac{1}{n} = 1. \tag{44}$$

So in the above example, we have shown $\mathbb{E}(N) = \mathrm{Var}(N) = 1$. Does this ring a bell? If $X \sim \mathrm{Poisson}(1)$, then $\mathbb{E}(X) = \mathrm{Var}(N) = 1$ (c.f. Exercise 1.21 in Note 0). So is N_n somehow related to the Poisson RV with rate 1? In the following two exercises, we will show that N_n approximately follows Poisson(1) if n is large.

Exercise 2.16 (Derangements)**.** In reference to Example 2.15, let D_n be the total number of arrangements of n keys so that no one gets the correct key.

(i) Show that the total number of arrangements of n keys is $n!$.
(ii) Show that there are $(n-1)!$ arrangements where person 1 got the right key.
(iii) Show that there are $(n-2)!$ arrangements where person 1 and 2 got the right key.
(iv) Show that there are $\binom{n}{2}(n-2)!$ arrangements where at least two people got the right key.
(v) Show that there are $\binom{n}{k}(n-k)!$ arrangements where at least k people got the right key.
(vi) By using inclusion-exclusion, show that

$$D_n = n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \binom{n}{3}(n-3)! + \cdots + (-1)^n \binom{n}{n}(n-n)! \tag{45}$$

$$= n!\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right) \to \frac{n!}{e} \quad \text{as } n \to \infty. \tag{46}$$

Exercise 2.17. Let $N_n = X_1 + X_2 + \cdots + X_n$ be as in Example 2.15.

(i) Use Exercise 2.16 to show that for each $1 \le k \le n$,

$$\mathbb{P}(N_n = k) = \binom{n}{k}\frac{D_{n-k}}{n!} \tag{47}$$

$$= \frac{n!}{k!(n-k)!}\frac{(n-k)!}{n!}\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!}\right) \tag{48}$$

$$= \frac{1}{k!}\left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^{n-k}\frac{1}{(n-k)!}\right). \tag{49}$$

(ii) Conclude that

$$\lim_{n\to\infty} \mathbb{P}(N_n = k) = \frac{e^{-1}}{k!} = \mathbb{P}(\text{Poisson}(1) = k). \tag{50}$$

Remark 2.18. Recall that Poisson(1) can be obtained from Binomial(n, p) where $p = 1/n$, for large n (c.f. Example 1.20 in Note 0). In other words, the sum of n independent Bernoulli$(1/n)$ RVs is distributed approximately as Poisson(1). In the key arrangement problem in Example 2.15, note that the correlation coefficient between X_i and X_j is very small:

$$\rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_j)}\sqrt{\text{Var}(X_j)}} = \frac{n^2}{n^2(n-1)^2} = \frac{1}{(n-1)^2}. \tag{51}$$

So it's kind of make sense that X_i's are almost independent for large n, so $N_n \sim$ Poisson(1) approximately for large n.