May 23 Lec—Covariance, Correlation Coefficient, Multinomial

A quick recap of the dot product

Dot product properties

(Let V be a vector space over $\mathbb{R}$. Then the inner product, also called the *dot product*, $\cdot : V \times V \to \mathbb{R}$ is a function such that:)

1. $\vec{u} \cdot (\vec{v}_1 + \vec{v}_2) = \vec{u} \cdot \vec{v}_1 + \vec{u} \cdot \vec{v}_2$
2. $\vec{u} \cdot (\alpha \vec{v}) = \alpha \vec{u} \cdot \vec{v}$
3. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
4. If $V = \mathbb{R}^n$ for some $n \in \mathbb{N}$, then $\vec{u} \cdot \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos \alpha$ where α is the angle between $\vec{u}$ and $\vec{v}$.

Definition:

Let X, Y be two random variables. The covariance between X and Y is defined as

$$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Remark

If $\mu_X = \mathbb{E}(X)$, $\mu_Y = \mathbb{E}(Y)$, then

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

Properties of covariance

1. $Cov(X, X) = V(X)$
2. $Cov(X, Y) = Cov(Y, X)$
3. $Cov(X, Y_1 + Y_2) = Cov(X, Y_1) + Cov(X, Y_2)$
4. $Cov(X, \alpha Y) = \alpha Cov(X, Y)$

These properties establish Cov as a dot product on the vector space of random variables.

Example

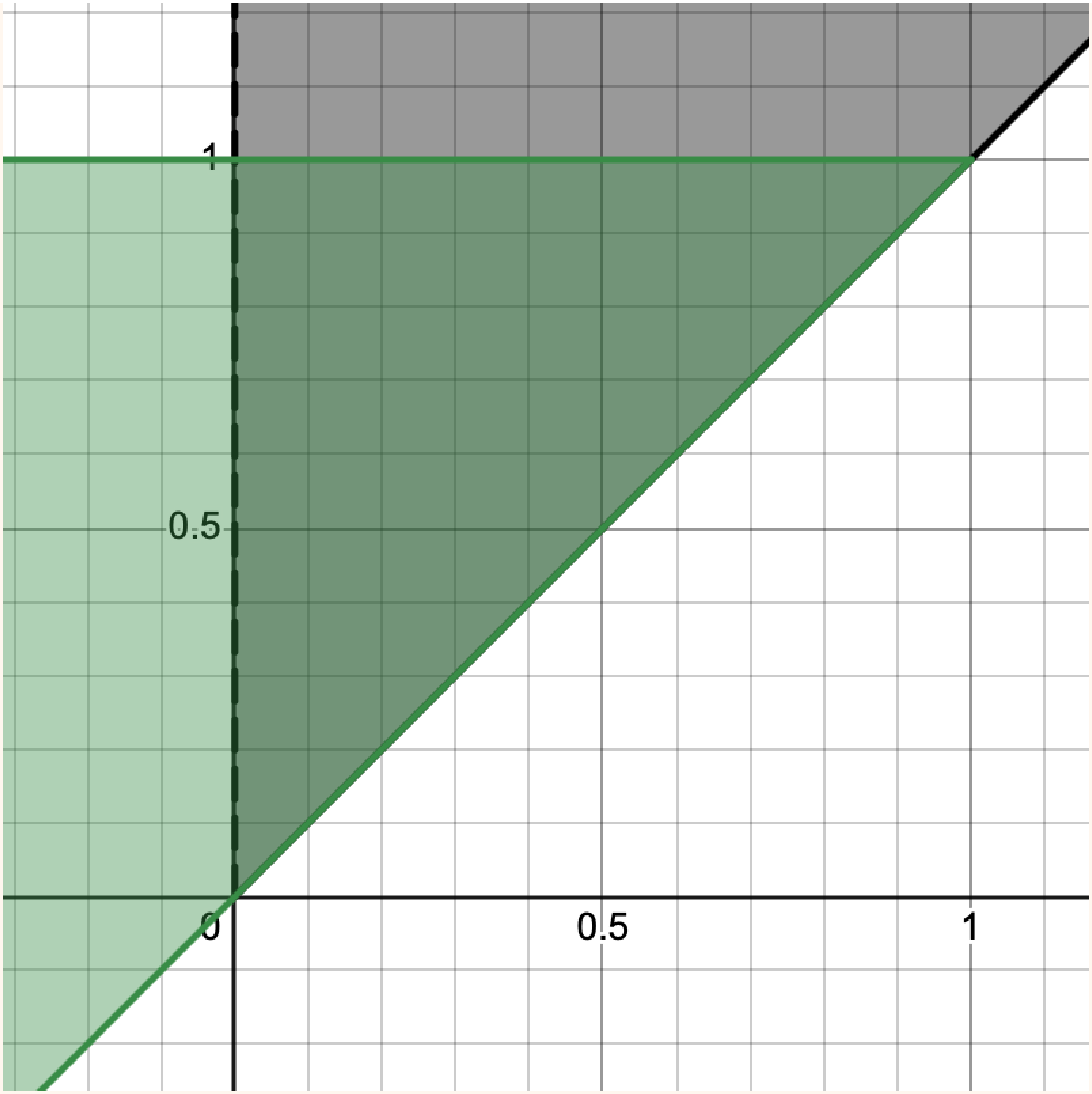
y**-value** x**-value**	-1	0	1	$p_Y(y)$
0	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{2}$
1	$\frac{1}{4}$	0	$\frac{1}{4}$	$\frac{1}{2}$
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$	

We calculated last time that $\mathbb{E}(XY) = 0$. Furthermore, $\mathbb{E}(X) = -\frac{1}{10}$ and $\mathbb{E}(Y) = \frac{1}{2}$. So $Cov(X, Y) = 0 - \left(-\frac{1}{6} \cdot \frac{1}{2}\right) = \frac{1}{12}$.

A continuous example

Again the function

$$f(x, y) = \begin{cases} 6x & 0 < x < y < 1 \\ 0 & \text{otherwise} \end{cases}$$



We have

$$\mathbb{E}(XY) = \iint_{\mathbb{R}^2} xyf(x,y)\,dA$$

$$= \int_0^1 \int_0^y xy \cdot 6x\,dx\,dy$$

$$\int_0^1 y \int_0^y 6x^2\,dx\,dy$$

$$\int_0^1 y \cdot 2y^3\,dy = \frac{2}{5}$$

We computed before that $X \sim Beta(2,2)$ and $Y \sim Beta(3,1)$, so $\mathbb{E}(X) = \frac{1}{2}, \mathbb{E}(Y) = \frac{3}{4}$. Therefore $Cov(X,Y) = \frac{1}{40}$.

Recall the famous Cauchy-Schwartz inequality:

$$|\vec{u} \cdot \vec{v}| = \|\vec{u}\|\|\vec{v}\|$$

with equality if and only if $\vec{u} = \pm\vec{v}$.

Proposition

$$|Cov(X, Y)| \leq \sqrt{V(X)} \cdot \sqrt{V(Y)}$$

Proof

Follows from the Cauchy-Schwartz inequality.

From here we have

$$-1 \leq \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \leq 1$$

Definition: Correlation coefficient

$\rho(x, y)$ is the correlation coefficient between X and Y.
It has the meaning of being a measure of the linear association between X and Y.

Properties of the correlation coefficient

1. $$-1 \leq \rho(X, Y) \leq 1$$
2. if $\rho(X, Y) = 1$, then $Y = \alpha X + b$ where $\alpha > 0$
3. if $\rho(X, Y) = -1$, then $Y = \alpha X + b$ where $\alpha < 0$.

Example

Suppose $V(X)=2$, $V(Y)=3$, $Cov(X,Y)=-\sqrt{6}$, $\mathbb{E}(X)=1$, $\mathbb{E}(Y)=3$. What is the relationship between X and Y?
We observe that $\rho(X,Y)=\frac{-\sqrt{6}}{\sqrt{6}}=-1$, so $Y=\alpha X+b$ where $\alpha<0$. Since $V(Y)=\alpha^2V(X)$, we deduce $\alpha^2=\frac{3}{2} \implies \alpha=-\sqrt{\frac{3}{2}}$, and since $\mathbb{E}(Y)=\alpha\mathbb{E}(X)+b$ we deduce $b=3+\sqrt{\frac{3}{2}}$.

Remark

Note that if $Cov(X,Y)$ had been $\sqrt{8}$ in the above example, then the scenario would have been impossible because $\sqrt{8}>\sqrt{2}\cdot\sqrt{3}$ which violates Cauchy-Schwartz.
Also, if $\rho(X,Y)\neq\pm 1$, we have insufficient information.

Definition

X and Y are said to be uncorrelated if $\rho(X,Y)=0$.

Remark

If X and Y are independent, then X and Y are uncorrelated. But the converse is not true, see the examples below.

Example

y-value\ x-value	-1	0	1	$p_Y(y)$
-1	q	p	q	$2q+p$
0	p	0	p	$2p$
1	q	p	q	$2q+p$
$p_X(x)$	$2q+p$	$2p$	$2q+p$	

If $p+q=\frac{1}{4}$ we get a joint probability distribution. $\mathbb{E}(XY)=0$, and you can verify that $\mathbb{E}(X)=\mathbb{E}(Y)=0$. But since the support is not of rectangular type, X and Y are not independent.

Proposition

$V(X+Y) = Cov((X+Y),(X+Y)) = V(X) + V(Y) + 2Cov(X,Y)$
more generally,

$$V\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i^2 V(X_i) + 2\sum_{i<j} \alpha_i \alpha_j Cov(X_i, X_j)$$

Example

Suppose X, Y, W are independent, $X \sim N(0,2)$, $Y \sim \chi^2(p=3)$, $W \sim Exponential(3)$.
Let us calculate $V(2X+3Y-W)$:

$$= 4V(X) + 9V(Y) + V(W) + \cancel{12Cov(X,Y)} \cancel{-4Cov(X,CW)} \cancel{-6Cov(Y,W)}$$
$$= 4 \cdot 2 + 9 \cdot 6 + 9$$

Big long formula for taking covariance

$$Cov\left(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{m} \beta_i Y_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_i Cov(X_i, Y_i)$$

Multinomial distribution

In a binomial distribution, the number of repetitions (N) is fixed, each trial is independent, and there are two outcomes (failure and success) with constant probability.
Let X be the number of successes and Y the number of failures.
$X \sim B(N,p)$ and $Y \sim B(N, 1-p)$. Furthermore, $Y = N - p$, so $\rho(X,Y) = -1$.
Hence $Cov(X,Y) = -\sqrt{V(X)} \cdot \sqrt{V(Y)} = -Np(1-p)$
But what if we have more than 2 outcomes?

Example

A fair die is such that 3 faces are labeled a, 2 are labeled b, and 1 is labeled c. Roll the die 10 times. Let X_1 be the number of A, X_2 the number of B, X_3 the number of C.

Taken by themselves, $X_1 \sim B\left(10, \frac{1}{2}\right), X_2 \sim B\left(10, \frac{1}{3}\right), X_3 \sim B\left(10, \frac{1}{6}\right)$.

We also have $X_1 + X_2 + X_3 = 10$.

We know that events like $X_1 = 1, X_2 = 3, X_3 = 7$ is impossible because the sum is not 10.

Calculating $P(X_1 = 1, X_2 = 3, X_3 = 6)$:

$$= \begin{bmatrix} 10 \\ 1, 3, 6 \end{bmatrix} \left(\frac{1}{2}\right)^1 \left(\frac{1}{3}\right)^3 \left(\frac{1}{6}\right)^6$$

Definition: Multinomial distribution

Assume that a random experiment has outcomes $w_1, w_2, \ldots, w_k$ with constant probability $p_1, p_2, \ldots, p_k$. Repeat the experiment N times, and let X_i be the number of times w_i occurred. Let the trials be independent. Then $X_1, X_2, \ldots, X_k$ are said to have a (joint) multinomial distribution with parameters $N, p_1, p_2, \ldots, p_k$.

Properties of the multinomial distribution

1. $X_i \sim B(N, p_i)$ for $i = 1, \ldots, k$
2. $$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{N!}{x_1! x_2! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$$

Example

Continue the die example above ($N = 10, p_1 = \frac{1}{2}, p_2 = \frac{1}{3}, p_3 = \frac{1}{6}$). Given $X_3 = 5$, what is the (conditional) distribution of X_1?

$$P(X_1 = x_1 | X_3 = 5) = \frac{P(X_1 = x_1, X_3 = 5)}{P(X_3 = 5)}$$

$$= \frac{P(X_1 = x_2, X_2 = 5 - x_1, X_3 = 5)}{P(X_3 = 5)}$$

$$\frac{\frac{\cancel{10!}}{x_1!(5-x_1)!\cancel{5!}}\left(\frac{1}{2}\right)^{x_1}\left(\frac{1}{3}\right)^{5-x_1}\left(\frac{1}{6}\right)^5}{\frac{\cancel{10!}}{5!\cancel{5!}}\left(\frac{5}{6}\right)^5\cancel{\left(\frac{1}{6}\right)^5}}$$

$$= \frac{5!}{x_1!(5-x_1)!}\left(\frac{3}{5}\right)^{x_1}\left(\frac{2}{5}\right)^{5-x_1}$$

so $X_1 \sim B\left(5, \frac{3}{5}\right)$.

More generally, if X_1, X_2, X_3 have a multinomial distribution with parameters N, p_1, p_2, p_3, then given $X_3 = n$ (where $n < N$) we have $X_1 \sim B\left(N - n, \frac{p_1}{1-p_3}\right)$ and $X_2 \sim B\left(N - n, \frac{p_2}{1-p_3}\right)$, or equivalently $X_1 \sim B\left(N - n, \frac{p_1}{p_1+p_2}\right)$ and $X_2 \sim B\left(N - n, \frac{p_2}{p_1+p_2}\right)$

Exercise

Prove the above.

Proposition

If $X_1, X_2, \ldots, X_k$ have a (joint) multinomial distribution with parameter $N, p_1, p_2, \ldots, p_k$, then:

1. $V(x_i) = Np_i(1 - p_i)$
2. $Cov(X_i, Xj) = -Np_ip_j$ if $i \neq j$

Proof

Left till next time.