Probability Theory

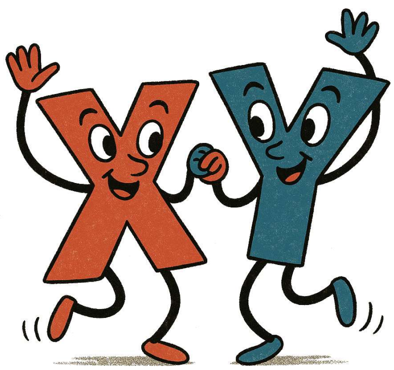
Textbook: *Introduction to Probability* by Blitzstein and Hwang

Previous Lecture

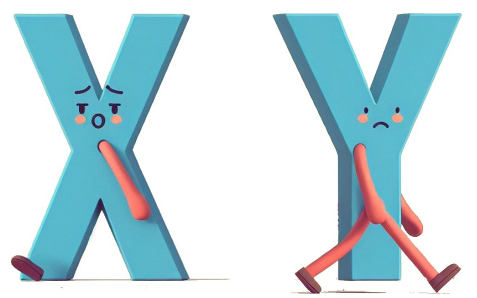
- Joint/Marginal/Conditional Distr
- Bayes' Rule and LOTP for two rvs
- Independence of Rvs
- 2D LOTUS

§7.3 - Covariance and Correlation

Just as mean and variance provided single-number summaries of the distr of a single rv, covariance $Cov(X, Y)$ is a single-number summary of the joint distr of X, Y.



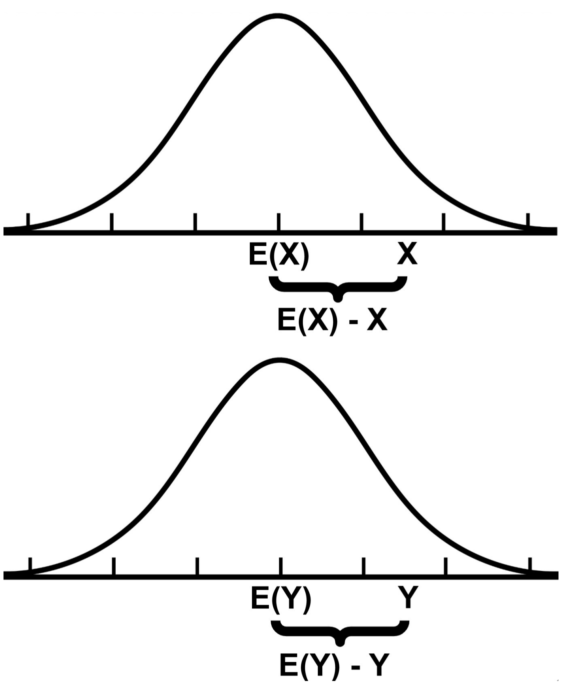
Positive Covariance



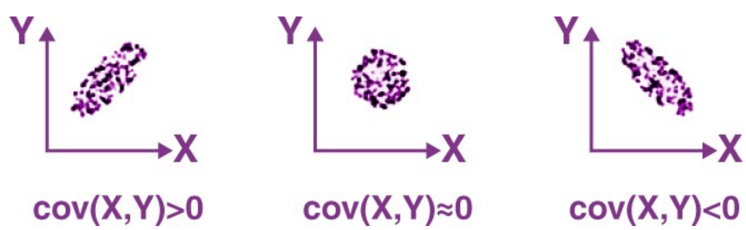
Negative Covariance

Covariance measures a tendency of X and Y to go up or down together, relative to their means $(E(X), E(Y))$

e.g. $Cov(X, Y) > 0$ means that when X goes up, so does Y (on avg).



$Cov(X, Y) > 0$: X increasing at same time Y increasing.



So, on avg, if they vary similarly, then $(X - E(X))(Y - E(Y))$ should be positive.
If they vary in opposite ways, $(X - E(X))(Y - E(Y))$ should be negative.

Def (**Covariance**). The covariance between X and Y is $Cov(X, Y) = E((X - E(X))(Y - E(Y)))$.

Multiplying this out and using linearity, we have an equivalent expression: $Cov(X, Y) = E(XY) - E(X)E(Y)$.

Pneumonic: If you consider covariance of X with itself, this is just variance. Substituting this in, we find:
$Var(X) = Cov(X,X) = E(XX) - E(X)E(X) = E(X^2) - E(X)^2$, as we expected.

If X and Y are indep, then their covariance is zero. We say that rvs w/zero covariance are *uncorrelated*.

Thm (**Indep and Correlation**): If X and Y are indep, then they are uncorrelated.

Proof (**cont**): We'll show this in the case where X and Y are cont, with PFs f_X and f_Y.

Notice above that if the covariance is zero, we have: $E(XY) = E(X)E(Y)$. So this is what we must show.

Since X and Y are indep, their joint PF is the product of the marginal PFs. So, by 2D LOTUS:

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty} yf_Y(y)\left(\int_{-\infty}^{\infty} xf_X(x)dx\right)dy \qquad \text{(pull things out of the } \int dx \text{ that don't depend on } x\text{)}$$

$$= \left(\int_{-\infty}^{\infty} xf_X(x)dx\right)\left(\int_{-\infty}^{\infty} yf_Y(y)dy\right) \qquad \text{(pull things out of the } \int dy \text{ that don't depend on } y\text{)}$$

$$= E(X)E(Y). \qquad \blacksquare$$

The proof in the discrete case is the same, with PMFs instead of PDFs.

❗ The converse of this theorem is false: just because X and Y are uncorrelated does not mean they are indep.

Covariance Properties:

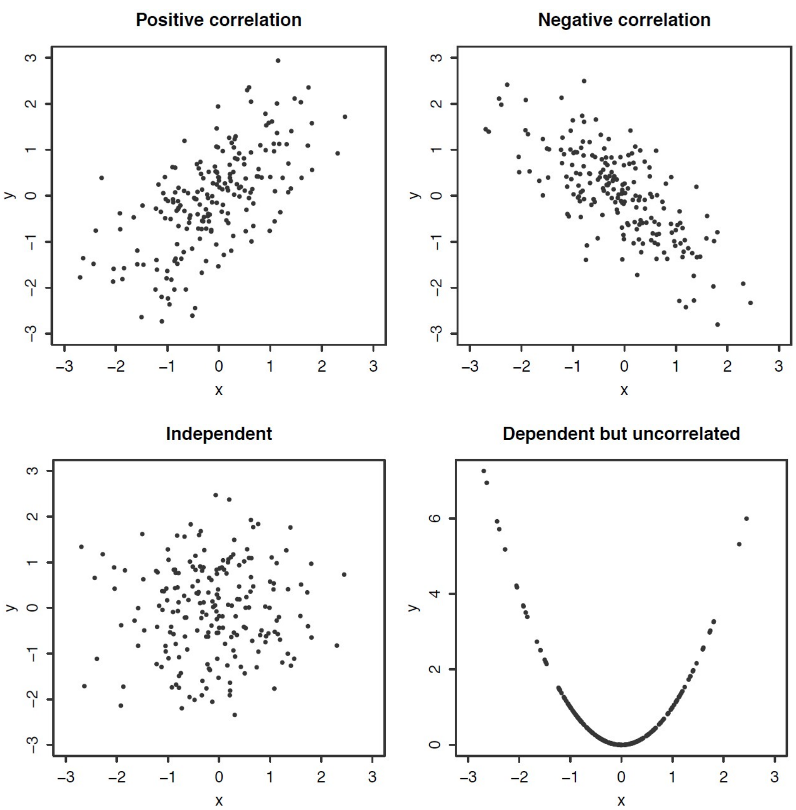
1. $Cov(X,X) = Var(X)$. (self variation)
2. $Cov(X, Y) = Cov(Y,X)$. (symmetry)
3. $Cov(X, c) = 0$ for any constant c. (annihilation)
4. $Cov(aX, Y) = aCov(X, Y)$ for any constant a. (homogeneity)

5. $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$. (additivity)

6. $Cov(X+Y,Z+W) = Cov(X,Z) + Cov(X,W) + Cov(Y,Z) + Cov(Y,W)$. (additivity)

7. $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$. (variation of a sum)

For n rvs $X_1,\ldots,X_n$ we have: $Var(X_1+\ldots+X_n) = Var(X_1)+\ldots+Var(X_n) + 2\sum_{i<j} Cov(X_i,X_j)$.



If X and Y are indep, then properties of covariance give $Var(X-Y) = Var(X) + Var(-Y) = Var(X) + Var(Y)$.
Common mistake: "$Var(X-Y) = Var(X) - Var(Y)$." This is an error since $Var(X) - Var(Y)$ could be negative!
For general X and Y, we have $Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$.

Ex (**Discrete Covariance**): Let X and Y have joint distr:

	Y		
X	0	1	2
1	0.05	0.10	0.05
2	0.15	0.05	0.05
3	0.20	0.15	0.20

a. Find $E(XY)$.

$E(XY) = \Sigma_x\Sigma_y xyf(x,y)$

$= \Sigma_x[(x\cdot 0)f(x,0) + (x\cdot 1)f(x,1) + (x\cdot 2)f(x,2)]$

$= (1\cdot 1)f(1,1) + (1\cdot 2)f(1,2) + (2\cdot 1)f(2,1) + (2\cdot 2)f(2,2) + (3\cdot 1)f(3,1) + (3\cdot 2)f(3,2)$

$= 1 \cdot (0.1) + 2 \cdot (0.05) + 2 \cdot (0.05) + 4 \cdot (0.05) + 3 \cdot (0.15) + 6 \cdot (0.2) = 2.15.$

b. $E(X) = 2.35$ and $E(Y) = 0.9$. Find of the covariance of X and Y.

$Cov(X, Y) = E(X, Y) - E(X)E(Y) = 2.15 - 2.35 \cdot 0.9 = 0.035$. (so they are correlated) □

Correlation

❗ Covariance depends on the units in which X and Y are measured.
The resulting numbers are therefore nonstandard, in that they depend upon the units used.
Instead, it's often more convenient to use correlation, where we **divide out the units**.

Def (**Correlation**): The correlation between X and Y is $\rho := Corr(X, Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$,

(This is undefined in the degenerate cases $Var(X) = 0$ or $Var(Y) = 0$.).

Ex (**Cont Correlation**): Let cont X and Y have joint PF: $f(x,y) = \begin{cases} 2(1-y) \text{ for } 0 \leq y \leq 1 \text{ and } 0 \leq x \leq y, \\ 0 \text{ otherwise.} \end{cases}$

a. Find the marginal distr of Y.

$$2(1-y)\int_0^y dx = 2(1-y)[x]_0^y = 2y(1-y).$$

$$f_Y(y) = \begin{cases} 2y(1-y) \text{ for } 0 \leq y \leq 1, \\ 0 \text{ otherwise.} \end{cases}$$

b. Find the expectation of Y.

$$E(Y) = \int_0^1 y f_Y(y) dy = \int_0^1 (2y^2 - 2y^3) dy = \left[\tfrac{2}{3}y^3 - \tfrac{1}{2}y^4\right]_0^1 = \tfrac{2}{3} - \tfrac{1}{2} - 0 = \tfrac{1}{6}.$$

c. Given $E(X) = \frac{1}{2}$, find the covariance of X and Y.

$$E(XY) = \int_0^1 \int_0^y xy(2(1-y))dxdy = \int_0^1 2(y-y^2)\int_0^y xdxdy = \int_0^1 (y-y^2)[x^2]_0^y dy = \int_0^1 (y^3 - y^4)dy$$

$$= \left[\tfrac{1}{4}y^4 - \tfrac{1}{5}y^5\right]_0^1 = \tfrac{1}{4} - \tfrac{1}{5} - 0 = \tfrac{1}{20}.$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = \tfrac{1}{20} - \tfrac{1}{2}\tfrac{1}{6} = -\tfrac{1}{30}.$$

d. The variance of X is $\frac{1}{4}$ and the variance of Y is $\frac{91}{720}$. Find the correlation ρ of X and Y.

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

$$= \frac{-\frac{1}{30}}{\sqrt{\frac{1}{4}\frac{91}{720}}} \approx -0.1875.$$

e. Find the variance of $X+Y$.

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) = \frac{1}{4} + \frac{91}{720} + 2\left(-\frac{1}{30}\right) = \frac{223}{720}. \qquad \square$$

Note: Scaling a rv does not affect the correlation:

$$Corr(cX,Y) = \frac{Cov(cX,Y)}{\sqrt{Var(cX)Var(Y)}} = \frac{cCov(X,Y)}{\sqrt{c^2 Var(X)Var(Y)}} = \frac{cCov(X,Y)}{c\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = Corr(X,Y). \qquad \blacksquare$$

Along w/eliminating the units, it turns out correlation also always exists between -1 and 1.

Thm (**Correlation Bounds**). For any X and Y, we have $-1 \leq Corr(X,Y) \leq 1$.

Proof. Without loss of generality we can assume X and Y have variance 1, since scaling does not change the correlation!

Let $\rho := Corr(X,Y) = Cov(X,Y)$. Since variance is nonnegative, along w/property 7 of covariance, we have:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) = 2 + 2\rho,$$

$$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y) = 2 - 2\rho.$$

$$0 \leq 2 + 2\rho \quad \Rightarrow \quad -2 \leq 2\rho \quad \Rightarrow \quad -1 \leq \rho.$$

$$0 \leq 2 - 2\rho \quad \Rightarrow \quad -2 \leq -2\rho \quad \Rightarrow \quad 1 \geq \rho.$$

Thus, $-1 \leq \rho \leq 1$. $\blacksquare$

Activity 16

Harvard Video: youtube.com/watch?v=IujCYxtpszU&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=22

What did we learn?

- Covariance: $E(XY) - E(X)E(Y)$
- Covariance Properties
- Correlation: $\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$