6.2 Covariance and correlation

- Covariance tells us if 2 r.v.s go up or down together. If covariance is positive then if X go up then Y also go up and vice versa.

Definition 6.2.1 (Covariance)

The covariance between X and Y is:

$$\mathrm{Cov}(X, Y) = E\big((X - EX)(Y - EY)\big)$$

Multiplying this out and using linearity, we have:

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

Interpret:
- If X and Y move in the same direction, then $(X - EX)$ and $(Y - EY)$ will be both positive or negative.
- If X and Y move in the opposite direction, then $(X - EX)$ and $(Y - EY)$ will have opposite signs.

If X and Y are independent, then their covariance are zero. We called X and Y uncorrelated

Theorem 6.2.2 (Uncorrelated)

If X and Y are independent, then they are uncorrelated.

The reverse is wrong:

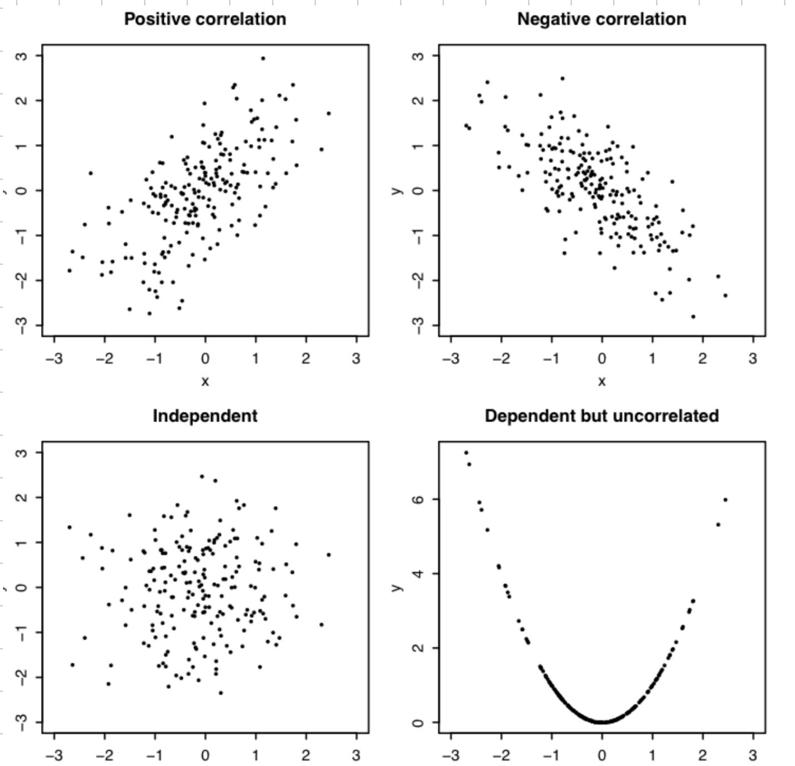
- Just because X and Y are uncorrelated, doesn't mean they are independent. For example: Let $X \sim N(0,1)$ and $Y = X^2$, then:

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E(XY) - E(X).E(Y)\\ &= E(X^3) - E(X).E(Y)\\ &= 0 - 0 = 0\end{aligned}$$

So X and Y are uncorrelated but they are not independent.

Note on Covariance:

Covariance is a measure of linear association, so r.v.s can be dependent in a nonlinear way (i.e quadratic, logarithmic) and still have zero covariance (uncorrelated). Here are some visuals of covariance:



Properties of Covariance:

1. $\text{Cov}(X, X) = \text{Var}(X)$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. $\text{Cov}(X, c) = 0$, for any constant c
4. $\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y)$ for any constant a
5. $\text{Cov}(X+Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
6. $\text{Cov}(X+Y, Z+W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$
7. $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y)$

   Generally, for n r.v.s $X_1, \dots, X_n$:

   $$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots \text{Var}(X_n) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

Correlation:

Correlation is a unit less version of covariance.

Definition 6.2.4 (Correlation)

The correlation between r.v.s X and Y is:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\ \mathrm{Var}(Y)}}$$

(This is undefined in the degenerate case $\mathrm{Var}(X) = 0$ or $\mathrm{Var}(Y) = 0$)

Shifting and scaling don't affect correlation:

- Shifting doesn't affect $\mathrm{Cov}(X, Y)$, $\mathrm{Var}(X)$ or $\mathrm{Var}(Y)$
- For Scaling, the denominator of the equation cancels out any scaling:

$$\mathrm{Corr}(cX, Y) = \frac{\mathrm{Cov}(cX, Y)}{\sqrt{\mathrm{Var}(cX)\ \mathrm{Var}(Y)}} = \frac{c\,\mathrm{Cov}(X, Y)}{\sqrt{c^2\,\mathrm{Var}(X)\ \mathrm{Var}(Y)}} = \mathrm{Corr}(X, Y)$$

Correlation is easier to interpret because its unitless.

Theorem 6.2.5 (Correlation bounds)

For any r.v.s X and Y:

$$-1 \leqslant \mathrm{Corr}(X, Y) \leqslant 1$$

Proof:

- We can assume X and Y have variance 1 (since scaling doesn't change the correlation) and $\rho = \mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$
- Using the fact that variance $\geqslant 0$ and property 7 of Covariance:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y) = 2 + 2\rho \geqslant 0$$

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y) = 2 - 2\rho \geqslant 0$$

Thus, $-1 \leqslant \rho \leqslant 1$

Covariance properties are useful tools for finding variances.

Example 6.2.6 (Hypergeometric variance)

Let $X \sim HGeom(w, b, n)$. Find $Var(X)$

- Interpret X as the number of white balls in sample size n that is draw from an urn with w white balls and b black balls.
- X can be represented as the sum of indicator r.v.s :

$$X = I_1 + \dots + I_n \quad \text{where} \begin{cases} \cdot\ I_j \sim Bern(p) \text{ is } j^{th} \text{ ball being white} \\ \cdot\ \text{Mean}: E(I_j) = p = \dfrac{w}{w+b} \\ \cdot\ \text{Variance}: Var(I_j) = p(1-p) \end{cases}$$

- Since I_j are dependent (p decrease as j increase), we cannot simply add all the variances of I_j to find $Var(X)$. Instead, we apply properties of covariance:

$$Var(X) = Var\left(\sum_{j=1}^{n} I_j\right) \qquad \text{<property 7 of covariance>}$$

$$= Var(I_1) + \dots + Var(I_n) + 2\sum_{i<j} Cov(I_i, I_j)$$

$$= n\,p(1-p) + \underbrace{2\binom{n}{2} Cov(I_1, I_2)}$$

Explanation:

Using the fact that all $\binom{n}{2}$ pairs of indicators have the same covariance by symmetry.

Proof: $Cov(I_i, I_j) = E(I_i \cdot I_j) - E(I_i \cdot I_j)$

$$= \underset{\text{1st draw}}{p_i}\ \underset{\text{2nd draw}}{p_j} - \underset{\text{individual draw}}{p} \cdot p$$

$$= \frac{w}{w+b} \cdot \left(\frac{w-1}{w+b-1}\right) - \left(\frac{w}{w+b}\right)^2$$

$$= \frac{w}{N} \cdot \left(\frac{w-1}{N-1}\right) - \left(\frac{w}{N}\right)^2$$

We can see the result don't depends on i^{th} and j^{th}, as long as $i \neq j$. So with any i and j, $Cov(I_i, I_j)$ are the same

So we just need to find $Cov(I_1, I_2)$ since i, j don't matter.

By fundamental bridge:

$$Cov(I_1, I_2) = E(I_1 I_2) - E(I_1) E(I_2)$$

$= P(1^{st}$ and 2^{nd} balls both white$) - P(1^{st}$ ball white$) P(2^{nd}$ ball white$)$

$$= \frac{w}{w+b} \cdot \frac{w-1}{w+b-1} - p^2$$

Plugging this to above formula, we arrive at variance of $X \sim HGeom$

Variance of HyperGeometric $X \sim HGeom(w, b, n)$

$$Var(X) = \frac{N-n}{N-1} np(1-p)$$