Covariance & Correlation

Tim Dosen Pengajar Probabilitas

Learning Goals

Understand the meaning of covariance and correlation.

Be able to compute the covariance and correlation of two random variables.

Covariance

- Covariance is a measure of how much two random variables vary together.
- For example, height and weight of giraffes have positive covariance because when one is big the other tends also to be big.
- Suppose X and Y are random variables with means μ_X and μ_Y. The covariance of X and Y is defined as

$$Cov(X,Y) = E\big((X-\mu_X)(Y-\mu_Y)\big)$$

Review of Expectation & Variance of Random Variable

Expectation

- For discrete R.V. of X, $E(X) = \sum x_i p(X = x_i)$
- For continues R.V. of X, $E(X) = \int x\, f_x(x)dx$
- For discrete R.V. of X with $g(x)$, $E(X) = \sum g(x)p(X = x_i)$
- For continues R.V. of X with $g(x)$, $E(X) = \int g(x)\, f_x(x)dx$
- $E(b) = b$, b constant
- $E(aX) = aE(X)$, a constant
- $E(aX + b) = aE(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- $E[g(x) + h(x)] = E[g(x)] + E[h(x)]$

Review of Expectation & Variance of Random Variable

Variance

- $Var(X) = E\left[\left(X - E(X)\right)^2\right] = E(X^2) - E(X)^2$
- $Var(X|Y) = E[X^2|Y] - [E(X|Y)]^2$
- For discrete R.V. of X, $Var(X) = E\left[\left(X - E(X)\right)^2\right] = \sum(x_i - \mu)^2 p(X = x_i)$
- For continues R.V. of X, $Var(X) = E\left[\left(X - E(X)\right)^2\right] = \int (x - \mu)^2 f_x(x) dx$
- If $Var(X)$ exists and $Y = a + bX$, then $Var(Y) = b^2 Var(X)$
- $Var(X) = Cov(X, X)$
- $Var(X) = E[Var(X|Y)] - Var[E(X|Y)]$

Properties of Covariance

1. $Cov(aX + b, cY + d) = ac\ Cov(X, Y)$ for constants a, b, c, d
2. $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$
3. $Cov(X, X) = Var(X)$
4. $Cov(X, Y) = E(XY) - \mu_X \mu_Y$
5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$ for any X and Y
6. If X and Y are independent then $Cov(X, Y) = 0$

   Warning! The converse is false: zero covariance does not always imply independence

Properties of Covariance

Notes.

- Property 4 is like the similar property for variance. Indeed, if X = Y it is exactly that property: $Var(X) = E(X^2) - \mu_X^2$.
- By Property 5, the formula in Property 6 reduces to the earlier formula $Var(X + Y) = Var(X) + Var(Y)$ when X and Y are independent.

Sums and integrals for computing covariance

- Since covariance is defined as an expected value we compute it in the usual way as a sum or integral.
- **Discrete case**: If X and Y have joint PMF $p(x_i, y_j)$ then,

$$Cov(X,Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)(x_i - \mu_X)(y_j - \mu_Y)$$

$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)x_i y_j\right) - \mu_X\mu_Y$$

Sums and integrals for computing covariance

- Since covariance is defined as an expected value we compute it in the usual way as a sum or integral.
- **Continuous case**: If X and Y have joint PDF $f(x, y)$ over range $[a, b] \times [c, d]$ then,

$$Cov(X, Y) = \int_c^d \int_a^b (x - \mu_x)(y - \mu_y) f(x, y)\, dx\, dy$$

$$= \left(\int_c^d \int_a^b xy\, f(x, y)\, dx\, dy \right) - \mu_x\, \mu_y$$

Proofs of the properties of covariance

1 and 2 follow from similar properties for expected value

3. $Cov(X,X) = E\big((X-\mu_X)(X-\mu_X)\big) = E\big((X-\mu_X)^2\big) = Var(X)$
4. Recall that $E(X-\mu_X) = 0$. So,

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\
&= E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y) \\
&= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y.
\end{aligned}$$

Proofs of the properties of covariance

5. Using properties 3 and 4, we get

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

6. If X and Y are independent, then $f(x,y) = f_X(x) f_Y(y)$. Therefore

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \int \int (x-\mu_X)(y-\mu_Y) f_X(x) f_Y(y)\, dx\, dy \\
&= \int (x-\mu_X) f_X(x)\, dx \int (y-\mu_Y) f_Y(y)\, dy \\
&= E(X-\mu_X) E(Y-\mu_Y) \\
&= 0.
\end{aligned}$$

Correlation

- The units of covariance $Cov(X, Y)$ are **units of X times units of Y** . This makes it hard to compare covariances: if we change scales then the covariance changes as well.
- Correlation is a way to remove the scale from the covariance.
- The correlation coefficient between X and Y is defined by

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

Properties of Correlation

1. ρ is the covariance of the standardization of X and Y
2. ρ is dimensionless (it's a ratio!)
3. $-1 \leq \rho \leq 1$. Furthermore,
   - $\rho = +1$ if and only if $Y = aX + b$ with $a > 0$
   - $\rho = -1$ if and only if $Y = aX + b$ with $a < 0$
   - Property 3 shows that ρ measures the linear relationship between variables. If the correlation is positive then when X is large, Y will tend to large as well. If the correlation is negative then when X is large, Y will tend to be small.

Exercise 1

Flip a fair coin 3 times. Let X be the number of heads in the first 2 flips and let Y be the number of heads on the last 2 flips (so there is overlap on the middle flip). Compute $Cov(X, Y)$.

Exercise 2

(*Zero covariance does not imply independence*). Let X be a random variable that takes values −2, −1, 0, 1, 2, each with probability 1/5. Let $Y = X^2$. Show that $Cov(X, Y) = 0$ but X and Y are not independent.

Exercise 3

Toss a fair coin 2n + 1 times. Let X be the number of heads on the first n + 1 tosses and Y the number on the last n + 1 tosses.

Compute $Cov(X, Y)$ and $Corr(X, Y)$.

Thank You