4.2 Covariance and correlation

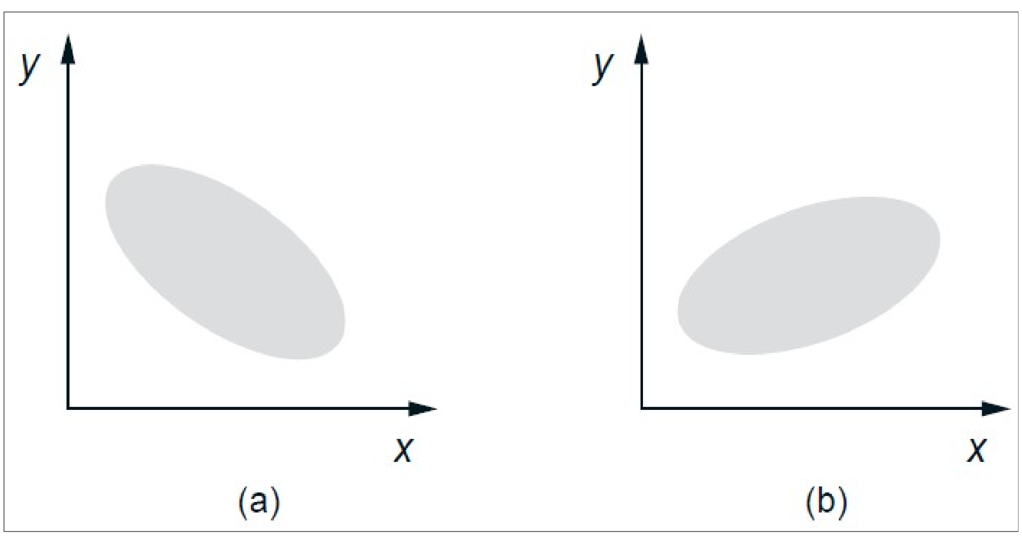
Definition

The **covariance** of two random variables X and Y is
$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$.
X and Y are **uncorrelated** if $\text{cov}(X, Y) = 0$.

Alternate form

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

Covariance visualized



Positively and negatively correlated patterns.

Other results

For scalars a and b:

- $\mathrm{cov}(X, X) = \mathrm{var}(X)$.
- $\mathrm{cov}(X, aY + b) = a\mathrm{cov}(X, Y)$.
- $\mathrm{cov}(X, Y + Z) = \mathrm{cov}(X, Y) + \mathrm{cov}(X, Z)$.
- $\mathrm{cov}(X, Y) = 0$ if X and Y are indepedent.

Example

Let X be a discrete uniform on $\{-1, 0, 1\}$. Let $Y = X^2$.

- Determine $\mathrm{cov}(X, Y)$.
- Are X and Y independent?

Example (cont)

Correlation

The **correlation coefficient** of random variables X and Y is

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}},$$

when $\mathrm{var}(X) > 0$ and $\mathrm{var}(Y) > 0$.

Interpretation

The correlation coefficient is a scaled version of covariance with $\rho \in [-1, 1]$.
If $\rho > 0$, then the values of $X - E(X)$ and $Y - E(Y)$ tend to have the same sign (deviate from the mean in the same direction).
If $\rho < 0$, then the deviations tend to have opposite signs.

Example 4.14

Suppose that $X - E(X) = -[Y - E(Y)]$. Determine $\operatorname{cor}(X, Y)$.

Example 4.14 (cont)

Variance of the Sum of Random Variables

Variance of a sum

Let $X_1, X_2, \ldots, X_n$ be a sequence of random variables with finite variance. Then

$$\begin{aligned}\operatorname{var}\left(\sum_{i=1}^{n}\right) &= \sum_{i=1}^{n} \operatorname{var}(X_i) + \sum_{\{(i,j)|i\neq j\}} \operatorname{cov}(X_i, X_j) \\ &= \sum_{i=1}^{n} \operatorname{var}(X_i) + 2\sum_{i<j} \operatorname{cov}(X_i, X_j).\end{aligned}$$

Variance of a sum (cont)

Variance of a sum (cont)

Example 4.15

Suppose that n people throw their hats in a box and then each picks one hat at random (without replacement). Let X_i be the random variable that person i picks their own hat.
Determine

$$\operatorname{var}\left(\sum_{i=1}^{n} X_i\right).$$

Example 4.15 (cont)