Expectation of a function of a random vector

If $h : \mathbb{R}^2 \to \mathbb{R}$, and (X, Y) is a random vector, then the expectation $\mathbb{E}h(X, Y)$ can be computed directly from the formula

$$\mathbb{E}h(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} h(x, y)\mathbb{P}(X = x, Y = y)$$

when X and Y are discrete, or from the formula

$$\mathbb{E}h(X, Y) = \iint_{\mathbb{R}^2} h(x, y) f(x, y) \mathrm{d}x \mathrm{d}y,$$

when X and Y are continuous and f is the joint density of (X, Y).

Example

Compute $\mathbb{E}XY$ if (X, Y) is distributed uniformly over the unit disk.

Expectation of the product of r.v.s

Theorem

If X and Y are independent then

$$\mathbb{E}(XY) = \mathbb{E}X \cdot \mathbb{E}Y.$$

Variance of the sum of independent r.v.s

Theorem

If X and Y are independent (and $\mathbb{E}X^2 < \infty$, $\mathbb{E}Y^2 < \infty$), then

$$\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y.$$

Variance of binomial r.v.s

Let X denote the number of successes in n Bernoulli trials, so

$$X \sim \mathrm{bin}(n, p).$$

If

$$X_i = \begin{cases} 1, & \text{if a success at } i\text{-th trial}, \\ 0, & \text{if a failure at } i\text{-th trial}, \end{cases} \qquad i = 1, \ldots, n,$$

then $X = X_1 + \ldots + X_n$. Since Bernoulli trials are independent, so are r.v.s $X_1, \ldots, X_n$. Therefore

$$\operatorname{Var} X = \operatorname{Var}(X_1 + \ldots + X_n) = \operatorname{Var} X_1 + \ldots + \operatorname{Var} X_n.$$

Since

$$\operatorname{Var} X_i = p(1 - p),$$

we conclude that

$$\operatorname{Var} X = np(1 - p).$$

Covariance

Definition

Covariance between r.v.s X ***and*** Y *is*

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right].$$

Covariance is defined for r.v.s satisfying $\mathbb{E}X^2 < \infty$ and $\mathbb{E}Y^2 < \infty$.

Equivalent formula for covariance:

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

Definition

R.v.s X *and* Y *with* $\mathrm{Cov}(X, Y) = 0$, *are* ***uncorrelated****.*

Covariance - some properties

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathrm{Var}\, X$
- $\mathrm{Cov}(aX, bY) = ab\, \mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X, c) = \mathrm{Cov}(c, X) = 0$
- $$\begin{aligned}\mathrm{Cov}(X_1 + X_2, Y_1 + Y_2) = \mathrm{Cov}(X_1, Y_1) + \mathrm{Cov}(X_1, Y_2) + \\ + \mathrm{Cov}(X_2, Y_1) + \mathrm{Cov}(X_2, Y_2)\end{aligned}$$
- **Cauchy-Schwarz inequality**:

$$|\,\mathrm{Cov}(X, Y)| \leq \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}$$

Variance of $X + Y$ - general formula

Using some properties of covariance, we can easily derive an important formula for the variance of the sum of random variables in the general case (that is, *without assuming independence*).

Theorem

If X and Y are r.v.s (for which the variances exist) then

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

Correlation coefficient

Definition

*The **correlation** between r.v.s X and Y us*

$$\rho_{X,Y} := \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}}.$$

(This makes sense only if the covariance makes sense and is undefined in the degenerate cases $\operatorname{Var} X = 0$ *or* $\operatorname{Var} Y = 0$*.)*

Some properties:

- $-1 \leq \rho_{X,Y} \leq 1$ (correlation, unlike covariance, is bounded),
- $\rho_{X,Y} = 0$ iff X and Y are uncorrelated.

Covariance/correlation - some examples

Example

Compute Cov(X, Y) *and* $\rho_{X,Y}$ *when*

- (X, Y) *has the joint distribution given by the table*

$X \backslash Y$	0	1
0	$1/4 + \varepsilon$	$1/4 - \varepsilon$
1	$1/4 - \varepsilon$	$1/4 + \varepsilon$

- (X, Y) *is uniformly distributed over the unit disk,*
- $(X, Y) \sim U([0, 1]^2)$.

Independence vs uncorrelatedness

Recall:

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

On the other hand, if X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y$.

Corollary

If X and Y are independent (and $\mathbb{E}X^2 < \infty$, $\mathbb{E}Y^2 < \infty$), then X and Y are uncorrelated (so $\mathsf{Cov}(X, Y) = 0$, equivalently $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$).

Corollary

If X and Y are uncorrelated, then $\mathsf{Var}(X + Y) = \mathsf{Var}\, X + \mathsf{Var}\, Y$.

Independence vs uncorrelatedness - cont'd

So

if X and Y are independent (and their covariance is well defined), then they are uncorrelated.

The converse is false:

if X and Y are uncorrelated then they are *not* necessarily independent.

Covariance matrix

Definition

*If (X, Y) is a random vector with $\mathbb{E}X^2 < \infty$ and $\mathbb{E}Y^2 < \infty$, then the **covariance matrix** of (X, Y) is*

$$\boldsymbol{C}_{(X,Y)} := \begin{bmatrix} \operatorname{Cov}(X, X) & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(Y, X) & \operatorname{Cov}(Y, Y) \end{bmatrix} = \begin{bmatrix} \operatorname{Var} X & \operatorname{Cov}(X, Y) \\ \operatorname{Cov}(X, Y) & \operatorname{Var} Y \end{bmatrix}.$$

A straightforward generalization to n-dimensions:

Definition

*If $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a random vector with $\mathbb{E}X_i^2 < \infty$ for all $i = 1, \ldots, n$ then the **covariance matrix** of vector $\boldsymbol{X}$ is the $n \times n$ matrix*

$$\boldsymbol{C}_{\boldsymbol{X}} = \left[\operatorname{Cov}(X_i, X_j)\right]_{i,j=1}^{n}.$$

Covariance matrix - some properties

- covariance matrices are symmetric and non-negative definite,
- the covariance matrix of random vector $\mathbf{AX} + \mathbf{b}$ is

$$\mathbf{A}\mathbf{C}_{\mathbf{X}}\mathbf{A}^T.$$

($\mathbf{A}$ is a matrix of numbers, $\mathbf{b}$ is a vector of numbers).

Multivariate normal distribution

The Multivariate Normal is a continuous multivariate distribution that generalizes the Normal distribution into higher dimensions. Recall that the (one dimensional) normal distribution $\mathcal{N}(\mu, \sigma^2)$ depends on two parameters:

- a real number μ,
- a positive number σ.

The equivalents of these parameters for the multivariate normal distribution are

- a vector of reals $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{R}^n$,
- a symmetric and non-negative definite (aka: positive semi-definite) matrix $\mathbf{C} \in \mathbb{R}^{n \times n}$.

Multivariate normal distribution - cont'd

Definition

*A random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ has a **multivariate normal distribution**, if its joint density function $f : \mathbb{R}^n \to \mathbb{R}$ is given by the formula*

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{C}}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{m})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{m})}{2}\right), \ \boldsymbol{x} \in \mathbb{R}^n.$$

Let us focus on the two-dimensional case ($n = 2$). Plugging

$$\mathbf{m} = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \in \mathbb{R}^2, \quad \mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{pmatrix},$$

into the general formula we get

$$f(x, y) = \frac{1}{2\pi\sqrt{\det \mathbf{C}}} \times$$
$$\exp\left\{-\frac{1}{2\det \mathbf{C}}\left[c_{22}(x - m_1)^2 - 2c_{12}(x - m_1)(y - m_2) + c_{11}(y - m_2)^2\right]\right\}$$

Bivariate normal distribution - cont'd

What is the probabilistic meaning of $\mathbf{m}$ and $\mathbf{C}$? If $(X, Y) \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$ then

- $m_1 = \mathbb{E}X$,
- $m_2 = \mathbb{E}Y$,
- $c_{11} = \operatorname{Var} X = \sigma_X^2$,
- $c_{22} = \operatorname{Var} Y = \sigma_Y^2$,
- $c_{12} = \operatorname{Cov}(X, Y) = \rho_{X,Y}\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}$.

Parameter $\rho_{X,Y}$ appearing in the formula for c_{12} is the **correlation** between X and Y.

Bivariate normal distribution - cont'd

Plugging

- $m_1 = \mathbb{E}X$, $m_2 = \mathbb{E}Y$,
- $c_{11} = \sigma_X^2$, $c_{22} = \sigma_Y^2$,
- $c_{12} = \rho_{X,Y}\sigma_X\sigma_Y$.

into the formula for f, we get a third form of the density for two-dimensional normal distribution:

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_X\sigma_Y}\times$$
$$\exp\left\{-\frac{1}{2(1-\rho_{X,Y}^2)}\left[\frac{(x-\mathbb{E}X)^2}{\sigma_X^2} - 2\rho_{X,Y}\frac{(x-\mathbb{E}X)(y-\mathbb{E}Y)}{\sigma_X\sigma_Y} + \right.\right.$$
$$\left.\left. + \frac{(y-\mathbb{E}Y)^2}{\sigma_Y^2}\right]\right\}.$$

Bivariate normal distribution - cont'd

Notice that the previous formula for the join density of $(X, Y) \sim \mathcal{N}(\mathbf{m}, \mathbf{C})$ simplifies significantly when $\rho_{X,Y} = 0$:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left\{-\frac{1}{2}\left(\frac{(x-m_1)^2}{\sigma_X^2} + \frac{(y-m_2)^2}{\sigma_Y^2}\right)\right\}.$$

So $\rho_{X,Y}$ implies that

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-m_1)^2}{2\sigma_X^2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(-\frac{(y-m_2)^2}{2\sigma_Y^2}\right),$$

which means that the joint density of vector (X, Y) is the products of the marginal densities of r.v.s X and Y. This also proves normality of the marginal distributions:

$$X \sim \mathcal{N}\left(m_1, \sigma_X^2\right), \;\; Y \sim \mathcal{N}\left(m_2, \sigma_Y^2\right).$$

Bivariate normal vectors - cont'd

From the above considerations we get the following conclusion:

Corollary

If $(X, Y) \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$ and $\rho_{X,Y} = 0$ (equivalently, matrix $\boldsymbol{C}$ is diagonal), then r.v.s X and Y are independent.

It turns out that all the marignal distributions of any multivariate normal distribution are normal (not only in the case of independence):

Theorem

If $(X, Y) \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$ then the marginals of (X, Y) are normal, and

$$X \sim \mathcal{N}(m_1, c_{11}), \;\; Y \sim \mathcal{N}(m_2, c_{22}).$$

How to identify the parameters of a bivariate normal distribution?

Example

Let (X, Y) *has a two-dimensional normal distribution with the joint density:*

$$f_{(X,Y)}(x, y) = \frac{1}{8\pi} \exp\left\{ -\frac{1}{16} \left[4(x-2)^2 - 4(x-2)(y+1) + 2(y+1)^2 \right] \right\}.$$

Find $\boldsymbol{m}$ *and* $\boldsymbol{C}$.

Convolution

A **convolution** is a sum of *independent* r.v.s (more precisely, it is its distribution). We often add independent r.v.s because the sum is a useful summary of an experiment, and because sums lead to averages, which are also useful.

Example

Compute the convolution of independent r.v.s X and Y, if

- X, Y *are binomially distributed with parameters* p, *and* n_1 *(*X*) and* n_2 *(*Y*),*
- X, Y *have Poisson distributions with parameters* λ_1 *and* λ_2*,*
- X, Y *are uniformly distributed over* $[0, 1]$*.*