Expectation

Chapter 4 of Probability and Statistics

Yiping Cheng
ypcheng@bjtu.edu.cn
10-Jul-2017

School of Electronic and Information Engineering
Beijing Jiaotong University

Chapter Contents

The Expectation of a Random Variable

Introducing Expectation

Summaries of the distribution, such as the average value, or expected value, can be useful for giving people an idea of where we expect X to be without trying to describe the entire distribution.

Example

Fair Price for a Stock. An investor is considering whether or not to invest \$18 per share in a stock for one year. If the expected gain in this investment is bigger than the supposed bank interest $18 \times 0.05 = 0.9$, then the investment is worthwhile. Suppose the gain is a random variable X that can assume only the four different values -2,0,1, and 4, and that $P(X = -2) = 0.1$, $P(X = 0) = 0.4$, $P(X = 1) = 0.3$, and $P(X = 4) = 0.2$. Then the weighted average of the values is $-2 \times 0.1 + 0 \times 0.4 + 1 \times 0.3 + 4 \times 0.2 = 0.9$. This expected gain in this investment for one year is equal to the supposed bank interest, hence the stock price of \$18 seems fair.

Definition of Expectation

Definition: Let X be an RV with c.d.f. F. The expectation, or mean, of X, denoted by $E(X)$, is a number defined as

$$E(X) = \int_{-\infty}^{\infty} x dF(x)$$

where the integral is the Lebesgue-Stieltjes integral.

Remark

- This definition is for all RVs. However, since the Lebesgue-Stieltjes integral is too difficult to understand, we shall later give equivalent computation formulas for discrete and continuous RVs.
- Although $E(X)$ is called the expectation of X, it depends only on the distribution of X.
- This integral may converge, diverge to $+\infty$, diverge to $-\infty$, or diverge to nowhere, in which case we say $E(X)$ does not exist.

Expectation of Discrete RVs

Theorem: Let X be a discrete RV with p.f. f. The expectation of X can be computed as

$$E(X) = \sum_{\text{All } x} xf(x) = \underbrace{\sum_{\text{Positive } x} xf(x)}_{E^+(X)} + \underbrace{\sum_{\text{Negative } x} xf(x)}_{E^-(X)}.$$

- If both $E^+(X)$ and $E^-(X)$ are finite, then $E(X)$ is finite.
- If one of $E^+(X)$ and $E^-(X)$ is finite but the other is infinite, then $E(X)$ is infinite, either ∞ or $-\infty$.
- If both $E^+(X)$ and $E^-(X)$ are infinite, then $E(X)$ does not exist.

Expectation of Discrete RVs - Examples

Bernoulli RV. Let X have the Bernoulli distribution with parameter p, that is, assume that X takes only the two values 0 and 1 with $P(X = 1) = p$. Then the expectation of X is

$$E(X) = 0 \times (1 - p) + 1 \times p = p.$$

Infinite Expectation. Let X be an RV whose p.f. is

$$f(x) = \begin{cases} \frac{1}{x(x+1)} & \text{if } x = 1, 2, 3, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

It can be verified that this is indeed a p.f. Obviously $E^-(X) = 0$, so the expectation of X exists and is

$$E(X) = \sum_{x=1}^{\infty} x \frac{1}{x(x+1)} = \infty.$$

We say that the expectation of X is *infinite* in this case.

Expectation of Discrete RVs - Examples (Continued)

Nonexistent Expectation. Let X be an RV which has p.f.

$$f(x) = \begin{cases} \frac{1}{2|x|(|x|+1)} & \text{if } x = \pm1, \pm2, \pm3, \dots, \\ 0 & \text{otherwise.} \end{cases}$$

It can be verified that this is indeed a p.f. The two sums are

$$\begin{aligned} E^-(X) &= \sum_{x=-1}^{-\infty} x \frac{1}{2|x|(|x|+1)} = -\infty, \\ E^+(X) &= \sum_{x=1}^{\infty} x \frac{1}{2x(x+1)} = \infty. \end{aligned}$$

Hence $E(X)$ does not exist.

Expectation of Continuous RVs

Theorem: Let X be a continuous RV with p.d.f. f. The expectation of X can be computed as

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \underbrace{\int_{-\infty}^{0} xf(x)dx}_{E^-(X)} + \underbrace{\int_{0}^{\infty} xf(x)dx}_{E^+(X)}.$$

- If both $E^+(X)$ and $E^-(X)$ are finite, then $E(X)$ is finite.
- If one of $E^+(X)$ and $E^-(X)$ is finite but the other is infinite, then $E(X)$ is infinite, either ∞ or $-\infty$.
- If both $E^+(X)$ and $E^-(X)$ are infinite, then $E(X)$ does not exist.

Expectation of Continuous RVs - Examples

Expected Failure Time. An appliance has a maximum lifetime of one year. The time X until it fails is a random variable with a continuous distribution having p.d.f.

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then $E(X) = \int_0^1 x(2x)dx = \int_0^1 2x^2 dx = \frac{2}{3}$.

Failure after Warranty. A product has a warranty of one year. Let X be the time at which the product fails. Suppose that X has a continuous distribution with the p.d.f.

$$f(x) = \begin{cases} 0 & \text{for } x < 1, \\ \frac{2}{x^3} & \text{for } x \geq 1. \end{cases}$$

The expected time to failure is then

$$E(X) = \int_1^\infty x \frac{2}{x^3} dx = \int_1^\infty \frac{2}{x^2} dx = 2.$$

Expectation of Continuous RVs - Examples (Continued)

Infinite Expectation. Let X be an RV whose p.d.f. is

$$f(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & \text{if } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

It can be verified that this is indeed a p.d.f. Obviously $E^{-}(X) = 0$, so the expectation of X exists and is

$$E(X) = \int_0^\infty x \frac{2}{\pi(1+x^2)} = \infty.$$

We say that the expectation of X is *infinite* in this case.

Expectation of Continuous RVs - Examples (Continued 2)

Nonexistent Expectation. Let X be an RV which has p.d.f.

$$f(x) = \frac{1}{\pi(1+x^2)}.$$

It can be verified that this is indeed a p.d.f., and this distribution is called the *Cauchy Distribution*. The two sums are

$$\begin{aligned} E^-(X) &= \int_{-\infty}^{0} x\frac{1}{\pi(1+x^2)} = -\infty, \\ E^+(X) &= \int_{0}^{\infty} x\frac{1}{\pi(1+x^2)} = \infty. \end{aligned}$$

Hence $E(X)$ does not exist.

Expectation of a Function of an RV

Theorem: Let X be an RV with c.d.f. F. $Y = r(X)$, then

$$E(Y) = E(r(X)) = \int_{-\infty}^{\infty} r(x)dF(x)$$

where the integral is the Lebesgue-Stieltjes integral.

Remark

This theorem is, again, for all RVs. Again, since the Lebesgue-Stieltjes integral is too difficult to understand, we shall later give equivalent computation formulas for discrete and continuous RVs.

Expectation of a Function of a Discrete RV

Theorem: Let X be a discrete RV with p.f. f. $Y = r(X)$, then

$$E(Y) = E(r(X)) = \sum_{\text{All } x} r(x)f(x)$$

where the rule for infinite summation is according to slide 6.

Proof.

Let g be the p.f. of Y. Then

$$\begin{aligned} E(Y) &= \sum_y yg(y) = \sum_y yP(r(X) = y) \\ &= \sum_y y \sum_{x:r(x)=y} f(x) = \sum_y \sum_{x:r(x)=y} r(x)f(x) \\ &= \sum_x r(x)f(x). \end{aligned}$$

□

Expectation of a Function of a Discrete RV – Example

Suppose a random variable X has the following p.f.

x	-2	0	2
$f(x)$	0.1	0.4	0.5

Let $Y = X^2$. Find $E(Y)$.

Solution Method 1: $E(Y) = E(X^2) = (-2)^2 \times 0.1 + 0^2 \times 0.4 + 2^2 \times 0.5 = 0.24$

Solution Method 2: Find the p.f. $g(y)$ of Y.

Y	4	0
$g(y)$	0.6	0.4

$E(Y) = 4 \times 0.6 + 0 \times 0.4 = 0.24.$

Expectation of a Function of a Continuous RV

Theorem: Let X be a continuous RV with p.d.f. f. $Y = r(X)$, then

$$E(Y) = E(r(X)) = \int_{-\infty}^{\infty} r(x)f(x)dx$$

where the rule for infinite integration is according to slide 9.

Partial Proof

Suppose that $r(x)$ is strictly increasing or strictly decreasing. Let g be the p.d.f. of Y. Then

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} yg(y)dy = \int_{-\infty}^{\infty} r(x)g(r(x))|dr(x)| \\ &= \int_{-\infty}^{\infty} r(x)g(r(x))|r'(x)|dx = \int_{-\infty}^{\infty} r(x)f(x)dx. \end{aligned}$$

□

Expectation of a Function of a Continuous RV – Example

Suppose a random variable X has the following p.d.f.

$$f(x) = \begin{cases} \frac{1}{2} & -1 \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y = X^2$. Find $E(Y)$.

Solution Method 1: $E(Y) = E(X^2) = \int_{-1}^{1} x^2 \frac{1}{2} dx = \frac{1}{3}$.

Solution Method 2: Find the p.d.f. $g(y)$ of Y: (see section 3.8)

$$g(y) = \begin{cases} \frac{1}{2y^{1/2}} & 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$E(Y) = \int_{-\infty}^{\infty} y g(y) dy = \int_{0}^{1} y \frac{1}{2y^{1/2}} dy = \frac{1}{3}.$$

Expectation of a Function of Several RVs

Theorem: Suppose that $X_1, ..., X_n$ are RVs and $Y = r(X_1, \ldots, X_n)$. Then if $X_1, ..., X_n$ have a discrete joint distribution with p.f. $f(x_1, ..., x_n)$, we have

$$E(Y) = \sum ... \sum r(x_1, ..., x_n) f(x_1, ..., x_n).$$

Similarly, if $X_1, ..., X_n$ have a continuous joint distribution with p.d.f. $f(x_1, ..., x_n)$, we have

$$E(Y) = \int ... \int r(x_1, ..., x_n) f(x_1, \ldots, x_n) dx_1 ... dx_n.$$

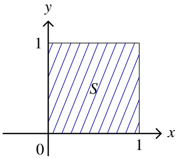
Expectation of a Function of Several RVs – Example

Suppose that a point (X, Y) is chosen at random from the square

$$S = \{(x, y) : 0 \le x \le 1 \text{ and } 0 \le y \le 1\}.$$

Determine the expected value of $X^2 + Y^2$.

Solution: We have $f(x, y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$



$$\begin{aligned} E(X^2 + Y^2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x^2 + y^2) f(x, y) dx dy \\ &= \int_0^1 \int_0^1 (x^2 + y^2) dx dy = \frac{2}{3}. \end{aligned}$$

Property 1 of Expectation – Locality

Theorem: Let a be a constant and X be an RV.

1. If $P(X = a) = 1$, then $E(X) = a$.
2. If $P(X \geq a) = 1$, then $E(X) \geq a$.
3. If $P(X \leq a) = 1$, then $E(X) \leq a$.
4. If $E(X) = a$ and either $P(X \geq a) = 1$ or $P(X \leq a) = 1$, then $P(X = a) = 1$.

Note: It follows from (2,4) that if $P(X \geq a) = 1$ and $P(X > a) > 0$, then $E(X) > a$. Therefore $P(X > a) = 1$ implies $E(X) > a$.

Corollary: If $E(X^2) = 0$, then $P(X = 0) = 1$.

Proof.

Obviously $P(X^2 \geq 0) = 1$. Now since $E(X^2) = 0$, by (4) of the above theorem, $P(X^2 = 0) = 1$, which means $P(X = 0) = 1$. □

Property 2 of Expectation – Linearity

Theorem [Linearity 1]: If $Y = aX + b$, where a and b are finite constants, then

$$E(Y) = aE(X) + b.$$

Note: In general, $E[g(X)] \neq g[E(X)]$, for example, $E[X^2] \neq [E(X)]^2$.

Theorem [Linearity 2]: If $X_1, \ldots, X_n$ are n RVs such that each expectation $E(X_i)$ is finite ($i = 1, \ldots, n$), then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n).$$

Note: It is true even when $X_1, \ldots, X_n$ are dependent.

Corollary: Assume that $E(X_i)$ is finite for $i = 1, \ldots, n$. For all constants $a_1, \ldots, a_n$ and b,

$$E(a_1X_1 + \cdots + a_nX_n + b) = a_1E(X_1) + \cdots + a_nE(X_n) + b.$$

Property 3 of Expectation – Multiplicativity on Independence

Theorem: If $X_1, \ldots, X_n$ are n independent RVs such that each expectation $E(X_i)$ is finite ($i = 1, ..., n$), then

$$E[X_1 X_2 \cdots X_n] = E(X_1) \times E(X_2) \times \cdots \times E(X_n).$$

Partial Proof

Suppose that $X_1, \ldots, X_n$ have a continuous joint distribution with joint p.d.f. f. Let f_i be the p.d.f. of X_i ($i = 1, \ldots, n$). Then, it follows from the independence assumption that

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i).$$

Therefore, $E(\prod_{i=1}^{n} X_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\prod_{i=1}^{n} x_i) f(x_1, \ldots, x_n) dx_1 \cdots dx_n$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\prod_{i=1}^{n} x_i f(x_i)) dx_1 \cdots dx_n = \prod_{i=1}^{n} \underbrace{\int_{-\infty}^{\infty} x_i f(x_i) dx_i}_{=E(X_i)}. \quad \square$$

Example of Property 3 of Expectation

Let X_1, X_2, and X_3 be independent RVs such that $E(X_i) = 0$ and $E(X_i^2) = 1$ for $i = 1, 2, 3$. Determine $E(X_1^2(X_2 - 4X_3)^2)$.

Solution: Since X_1, X_2, and X_3 are independent, it follows that the two RVs X_1^2 and $(X_2 - 4X_3)^2$ are also independent. Therefore,

$$\begin{aligned} E(X_1^2(X_2 - 4X_3)^2) &= E(X_1^2)E[(X_2 - 4X_3)^2] \\ &= E(X_2^2 - 8X_2X_3 + 16X_3^2) \\ &= E(X_2^2) - 8E(X_2X_3) + 16E(X_3^2) \\ &= 1 - 8E(X_2)E(X_3) + 16 \\ &= 17. \end{aligned}$$

Variance

Definition of Variance and Standard Deviation

It is useful to give some measure of how spread out the distribution of X is. The variance of X is one such measure.

Definition: Let X be a random variable with finite mean $\mu = E(X)$. The variance of X is defined as follows:

$$\mathrm{Var}(X) = E[(X-\mu)^2].$$

If X has infinite mean or if the mean of X does not exist, we say that $\mathrm{Var}(X)$ does not exist. The standard deviation σ_X of X is

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

if $\mathrm{Var}(X)$ exists.

Possible Combinations of Mean and Variance

Mean	Variance
finite	finite
finite	∞
$\pm\infty$	nonexistent
nonexistent	nonexistent

Remark

When do we say “expectation" and when do we say “mean"? What is their difference? Mathematically, they mean the same. However, in a practical setting the term “mean" is usually reserved for the main RV of concern, and therefore it is more appropriate to use “expectation" for the derived RVs which are functions of the main RV. In addition, sometimes we use “mean" to differentiate it from the variance which is also an expectation.

Alternative Method for Calculating the Variance

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

Proof.

$\mathrm{Var}(X) = E[(X-\mu)^2] = E(X^2-2\mu X+\mu^2) = E(X^2)-2\mu E(X)+\mu^2 = E(X^2)-\mu^2.$ □ □

Example

Suppose a random variable X has the following p.f.

x	-2	0	2
$f(x)$	0.1	0.4	0.5

$$\begin{aligned} E(X) &= -2\times 0.1 + 0\times 0.4 + 2\times 0.5 = -0.8 \\ E(X^2) &= (-2)^2\times 0.1 + 0^2\times 0.4 + 2^2\times 0.5 = 2.4 \\ \mathrm{Var}(X) &= E(X^2) - [E(X)]^2 = 2.4 - (-0.8)^2 = 1.76 \end{aligned}$$

Properties of Variance – 1

Theorem: For all RV X with finite mean, $\mathrm{Var}(X) \geq 0$. If X is bounded, then $\mathrm{Var}(X)$ must exist and be finite.

Theorem: $\mathrm{Var}(X) = 0$ if and only if there exists a constant c such that $P(X = c) = 1$.

Proof.

If: Trivial.
Only if: Let $\mu = E(X)$. Since $E[(X - \mu)^2] = \mathrm{Var}(X) = 0$, by the corollary in slide 22, we have $P(X = \mu) = 1$. □ □

Properties of Variance – 2

Theorem: For constants a and b, let $Y = aX + b$. Then

$$\text{Var}(Y) = a^2\text{Var}(X).$$

Proof.

Let $\mu = E(X)$. Then $E(Y) = a\mu + b$ by theorem in slide 20. Therefore,

$$\begin{aligned}\text{Var}(Y) &= E[(aX + b - a\mu - b)^2] = E[(aX - a\mu)^2] \\ &= a^2E[(X - \mu)^2] = a^2\text{Var}(X).\end{aligned}$$

$\square$

Note: This implies $\text{Var}(b) = 0$ and $\text{Var}(X + b) = \text{Var}(X)$.

Properties of Variance – 3

Theorem: If X_1 and X_2 are independent RVs with finite means, then

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2).$$

Proof.

Let $\mu_1 = E(X_1)$ and $\mu_2 = E(X_2)$. Then $E(X_1 + X_2) = \mu_1 + \mu_2$. Therefore,

$$\begin{aligned}
\mathrm{Var}(X_1 + X_2) &= E[(X_1 + X_2 - \mu_1 - \mu_2)^2] \\
&= E[(X_1 - \mu_1)^2 + (X_2 - \mu_2)^2 + 2(X_1 - \mu_1)(X_2 - \mu_2)] \\
&= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2E[(X_1 - \mu_1)(X_2 - \mu_2)].
\end{aligned}$$

Since X_1 and X_2 are independent, $E[(X_1 - \mu_1)(X_2 - \mu_2)] = E(X_1 - \mu_1)E(X_2 - \mu_2) = 0$.
It follows that $\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2)$. □

Corollary: For constants $a_1, \dots, a_n$ and b, if $X_1, \dots, X_n$ are independent RVs with finite means, then

$$\mathrm{Var}(a_1X_1 + \cdots + a_nX_n + b) = a_1^2\mathrm{Var}(X_1) + \cdots + a_n^2\mathrm{Var}(X_n).$$

Covariance and Correlation

Definition of Covariance

The covariance and correlation measure how much two random variables depend on each other.

Definition: Let X and Y be random variables having finite means. Let $E(X) = \mu_X$ and $E(Y) = \mu_Y$. The covariance of X and Y is defined as

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

Remark

- The covariance only describes the LINEAR relationship between the two random variables.
- X and Y are positively correlated if $\text{Cov}(X, Y) > 0$; negatively correlated if $\text{Cov}(X, Y) < 0$; uncorrelated if $\text{Cov}(X, Y) = 0$.

Alternative Method for Calculating the Covariance

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

Proof.

Let $\mu_X = E(X)$, $\mu_Y = E(Y)$. Then

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\
&= E(XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y) \\
&= E(XY) - 2\mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y.
\end{aligned}$$

$\square$

Definition of Correlation

Correlation, or correlation coefficient, is the dimension-less version of covariance.

Definition [Correlation]: Let X and Y be random variables with finite variances. Then the correlation of X and Y is defined as follows:

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

Definition [Uncorrelated]: We say two random variables X and Y are uncorrelated, if

$$\rho(X, Y) = 0.$$

Inequalities Related to Correlation

Theorem [Schwarz Inequality]: For all RVs U and V,

$$[E(UV)]^2 \leq E(U^2)E(V^2).$$

Suppose $0 < E(U^2)E(V^2) < \infty$. Then

$$E(UV) = \sqrt{E(U^2)E(V^2)} \Leftrightarrow P(V = aU) = 1 \text{ for some } a > 0;$$
$$E(UV) = -\sqrt{E(U^2)E(V^2)} \Leftrightarrow P(V = aU) = 1 \text{ for some } a < 0.$$

Corollary: For all RVs X and Y with finite variance,

$$[\mathrm{Cov}(X, Y)]^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y),$$

and

$$-1 \leq \rho(X, Y) \leq 1.$$

Furthermore, $\rho(X, Y) = 1$ if and only if $P(Y = aX + b) = 1$ for some $a > 0$; $\rho(X, Y) = -1$ if and only if $P(Y = aX + b) = 1$ for some $a < 0$.

Computing Covariance and Correlation - Example 1

Let X and Y be two RVs having the following joint probability function $f(x, y)$.

$f(x,y)$	$x=0$	$x=1$	$x=2$	
$y=0$	$\frac{1}{6}$	$\frac{2}{9}$	$\frac{1}{36}$	$\frac{5}{12}$
$y=1$	$\frac{1}{3}$	$\frac{1}{6}$	0	$\frac{1}{2}$
$y=2$	$\frac{1}{12}$	0	0	$\frac{1}{12}$
	$\frac{7}{12}$	$\frac{7}{18}$	$\frac{1}{36}$	1

(1) Find the **covariance** of X and Y.

(2) Find the **correlation coefficient** of X and Y.

$f(x,y)$	$x=0$	$x=1$	$x=2$	
$y=0$	$\frac{1}{6}$	$\frac{2}{9}$	$\frac{1}{36}$	$\frac{5}{12}$
$y=1$	$\frac{1}{3}$	$\frac{1}{6}$	0	$\frac{1}{2}$
$y=2$	$\frac{1}{12}$	0	0	$\frac{1}{12}$
	$\frac{7}{12}$	$\frac{7}{18}$	$\frac{1}{36}$	1

Solution: (1)

$$
\begin{aligned}
E(XY) &= \sum_x \sum_y (xy) f(x,y) \\
&= (0\times 0)\times \frac{1}{6} + (0\times 1)\times \frac{1}{3} + (0\times 2)\times \frac{1}{12} \\
&\quad +(1\times 0)\times \frac{2}{9} + (1\times 1)\times \frac{1}{6} + (1\times 2)\times 0 \\
&\quad +(2\times 0)\times \frac{1}{36} + (2\times 1)\times 0 + (2\times 2)\times 0 \\
&= \frac{1}{6}
\end{aligned}
$$

The (marginal) p.f. of X,

x	$x=0$	$x=1$	$x=2$
$f_X(x)$	$\frac{7}{12}$	$\frac{7}{18}$	$\frac{1}{36}$

$$
\begin{aligned}
E(X) &= \sum_x x f_X(x) \\
&= 0\times\frac{7}{12}+1\times\frac{7}{18}+2\times\frac{1}{36} \\
&= 4/9. \\
E(X^2) &= \sum_x x^2 f_X(x) \\
&= 0^2\times\frac{7}{12}+1^2\times\frac{7}{18}+2^2\times\frac{1}{36} \\
&= 1/2. \\
\sigma_X^2 &= E(X^2)-(EX)^2=\frac{1}{2}-\left(\frac{4}{9}\right)^2=\frac{49}{162}.
\end{aligned}
$$

The (marginal) p.f. of Y,

y	$y=0$	$y=1$	$y=2$
$f_Y(y)=P(Y=y)$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$

$$
\begin{aligned}
E(Y) &= \sum_y y f_Y(y) \\
&= 0\times\frac{5}{12}+1\times\frac{1}{2}+2\times\frac{1}{12} \\
&= 2/3. \\
E(Y^2) &= \sum_y y^2 f_Y(y) \\
&= 0^2\times\frac{5}{12}+1^2\times\frac{1}{2}+2^2\times\frac{1}{12} \\
&= 5/6. \\
\sigma_Y^2 &= E(Y^2)-(EY)^2=\frac{5}{6}-\left(\frac{2}{3}\right)^2=\frac{7}{18}.
\end{aligned}
$$

The covariance

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY)-E(X)E(Y)\\ &= \frac{1}{6}-\frac{4}{9}\times\frac{2}{3}\\ &= -\frac{7}{54}.\end{aligned}$$

(2) The correlation coefficient

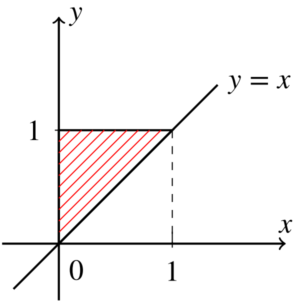
$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y} = \frac{-\frac{7}{54}}{\sqrt{\frac{49}{162}}\sqrt{\frac{7}{18}}} = -\frac{\sqrt{7}}{7}.$$

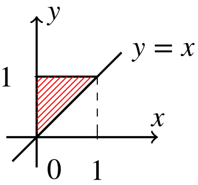
Computing Covariance and Correlation - Example 2

Let X and Y be two RVs having the joint density function

$$f(x, y) = \begin{cases} 2 & 0 < x \leq y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

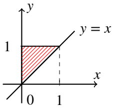
Find the correlation coefficient between X and Y.





Solution:

$$
\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \\
&= \int_0^1 \left[\int_0^y xy \times 2dx\right] dy \\
&= \int_0^1 \left[y \times x^2|_{x=0}^{x=y}\right] dy \\
&= \int_0^1 y^3 dy = \frac{y^4}{4}|_{y=0}^{y=1} = \frac{1}{4}.
\end{aligned}
$$

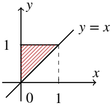


The (marginal) density of X is as follows,

- If $x \le 0$ or $x \ge 1$, $f_X(x) = 0$.
- If $0 < x < 1$, $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_x^1 2dy = 2(1-x)$.

$$f_X(x) = \begin{cases} 2(1-x) & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x \times 2(1-x)dx = \frac{1}{3}. \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_0^1 x^2 \times 2(1-x)dx = \frac{1}{6}. \\ \sigma_X^2 &= E(X^2) - [E(X)]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}. \end{aligned}$$



The (marginal) density of Y is as follows,

- If $y \le 0$ or $y \ge 1$, $f_Y(y) = 0$
- If $0 < y < 1$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^y 2dx = 2y$.

$$f_Y(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^1 y \times 2y dy = \frac{2y^3}{3}\Big|_{y=0}^{y=1} = \frac{2}{3}. \\ E(Y^2) &= \int_{-\infty}^{\infty} y^2 f_Y(y)dy = \int_0^1 y^2 \times 2y dy = \frac{y^4}{2}\Big|_{y=0}^{y=1} = \frac{1}{2}. \\ \sigma_Y^2 &= E(Y^2) - [E(Y)]^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}. \end{aligned}$$

The covariance

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\ &= \frac{1}{4} - \frac{1}{3} \times \frac{2}{3} \\ &= \frac{1}{36}.\end{aligned}$$

The correlation coefficient

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{1}{36}}{\sqrt{\frac{1}{18}}\sqrt{\frac{1}{18}}} = \frac{1}{2}.$$

Dependent but Uncorrelated RVs – Example

Suppose X has the following p.f.

x	-1	0	1
$f(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

and $Y = X^2$. Show that X and Y are dependent but uncorrelated.

Solution: The p.f. $g(y)$ of Y is

y	1	0
$g(y)$	$\frac{2}{3}$	$\frac{1}{3}$

$P(Y=1, X=1) = P(X^2=1, X=1) = P(X=1) = \frac{1}{3}$,
$P(Y=1) = \frac{2}{3}$, $P(X=1) = \frac{1}{3}$, $P(Y=1, X=1) \neq P(Y=1) \times P(X=1)$, so X and Y are dependent.
$E(X) = (-1) \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$
$E(XY) = E(X^3) = (-1)^3 \times \frac{1}{3} + 0^3 \times \frac{1}{3} + 1^3 \times \frac{1}{3} = 0$
$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$, so X and Y are uncorrelated.

Properties of Covariance

Let X and Y be random variables with finite variances, and a, b, and c be constants. Then

1. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$.
2. $\mathrm{Var}(aX+bY+c) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)$.
3. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.
4. $\mathrm{Cov}(aX,bY) = ab\mathrm{Cov}(X,Y)$.
5. $\mathrm{Cov}(X_1+X_2,Y) = \mathrm{Cov}(X_1,Y) + \mathrm{Cov}(X_2,Y)$.
6. $\mathrm{Cov}(X,c) = 0$.
7. If X and Y are independent, then $\mathrm{Cov}(X,Y) = 0$. But the converse is not true in general.
8. $\rho(X,Y) = 1$ if and only if $P(Y = aX+b) = 1$ for some $a > 0$;
   $\rho(X,Y) = -1$ if and only if $P(Y = aX+b) = 1$ for some $a < 0$.

Homework for Chapter 4

- P216: 8, 11
- P224: 3, 8
- P233: 4, 7
- P255: 12, 13, 18
- P272: 7, 8, 13