4 Moments

In the following $(\Omega, \mathcal{A}, \mathbf{P})$ will denote our canonical probability space and when we refer to a random variable, we mean a measurable application defined on Ω.

Definition 4.1. Let X and Y be real random variables.

(i) If $X \in \mathcal{L}^1(\mathbf{P})$, then X is called **integrable** and we call

$$\mathbf{E}[X] := \int X d\mathbf{P}$$

the **expectation** or **mean** of X. If $\mathbf{E}[X] = 0$, then X is called **centered**. More generally, we also write $\mathbf{E}[X] = \int X d\mathbf{P}$ if only X^- or X^+ is integrable.

(ii) If $n \in \mathbb{N}$ and $X \in \mathcal{L}^n(\mathbf{P})$, then the quantities

$$m_k := \mathbf{E}[X^k], \text{ for any } k = 1, \ldots, n,$$

are called the k**th moments** of X.

(iii) If $X \in \mathcal{L}^2(\mathbf{P})$, then X is called **square integrable** and

$$\mathbf{Var}[X] := \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

is the **variance** of X. The number

$$\sigma := \sqrt{\mathbf{Var}[X]}$$

is called the **standard deviation** of X.

(iv) If $X, Y \in \mathcal{L}^2(\mathbf{P})$, then we define the **covariance** of X and Y by

$$\mathbf{Cov}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

X and Y are called **uncorrelated** if $\mathbf{Cov}[X, Y] = 0$ and **correlated** otherwise.

Remark 4.2. Let X and Y be real random variables.

(i) The definition in (ii) is sensible since $\mathcal{L}^n(\mathbf{P}) \subset \mathcal{L}^k(\mathbf{P})$ for all $k = 1, \ldots, n$.

(ii) The standard deviation of X makes sense in definition (iii) since

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] \geq 0.$$

(iii) If $X, Y \in \mathcal{L}^2(\mathbf{P})$, then $XY \in \mathcal{L}^1(\mathbf{P})$ since $|XY| \leq X^2 + Y^2$. Hence the definition in (iv) makes sense and we have

$$\mathbf{Cov}[X, Y] = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].$$

In particular, $\mathbf{Var}[X] = \mathbf{Cov}[X, X]$.

Now, we collect the most important rules of expectations. All of these properties are direct consequences of the corresponding properties of the integral.

Property 4.3 (Rules for expectations)**.** Let $X, Y, X_1, X_2, \dots$ are integrable real random variables.

(i) [Linearity] Let $a, b \in \mathbb{R}$. Then $aX + bY$ is integrable and

$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y].$$

(ii) If $X \geq 0$ a.s., then

$$\mathbf{E}[X] = 0 \quad \Leftrightarrow \quad X = 0 \text{ a.s.}$$

(iii) [Monotonicity] If $X \leq Y$ a.s., then $\mathbf{E}[X] \leq \mathbf{E}[Y]$, with equality iff $X = Y$ a.s.

(iv) [Triangle inequality] $|\mathbf{E}[X]| \leq \mathbf{E}[|X|]$.

(v) If $X_n \geq 0$ a.s. for all $n \in \mathbb{N}$, then

$$\mathbf{E}\left[\sum_{n=1}^{+\infty} X_n\right] = \sum_{n=1}^{+\infty} \mathbf{E}[X_n].$$

(vi) If $X_n \uparrow X$, then

$$\mathbf{E}[X] = \lim_{n \to +\infty} \mathbf{E}[X_n].$$

Proof.

(v) It follows from the monotone convergence theorem.

(vi) Again, it follows from applying the monotone convergence theorem to $X_n - X_1$. □

Property 4.4. Let X be a real random variable and let $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable. Then $h \circ X \in \mathcal{L}^1(\mathbf{P})$ iff $h \in \mathcal{L}^1(\mathbf{P}_X)$, and in this case:

$$\mathbf{E}[h \circ X] = \int h(x)\mathbf{P}_X(dx). \tag{4.1}$$

Moreover, if $h \geq 0$ then equation 4.1 also holds.

Proof. This follows from the image measure property of integrals. □

Remark 4.5. Let X and Y are indentically distributed real random variables. Then, by virtue of the property above,

(i) $X, Y \in \mathcal{L}^1(\mathbf{P}) \quad \Rightarrow \quad \mathbf{E}[X] = \mathbf{E}[Y]$,

(ii) $X, Y \in \mathcal{L}^2(\mathbf{P}) \quad \Rightarrow \quad \mathbf{Var}[X] = \mathbf{Var}[Y]$.

Again probability theory comes into play when independence enters the stage; that is, when we exit the realm of linear integration theory.

Theorem 4.6. Let X and Y be independent real random variables and let $h : \mathbb{R}^2 \to \mathbb{R}$ be measurable. If $h \geq 0$ then

$$\mathbf{E}[h(X,Y)] = \int h \, d\mathbf{P}_X \otimes \mathbf{P}_Y. \tag{4.2}$$

Moreover, if $h(X,Y) \in \mathcal{L}^1(\mathbf{P})$ then $h \in \mathcal{L}^1(\mathbf{P}_X \otimes \mathbf{P}_Y)$, and equation 4.2 holds.

Corollary 4.7. Let X and Y independent real random variables and let $f, g : \mathbb{R} \to \mathbb{R}$ be measurable functions. If $f, g \geq 0$ then

$$\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]. \tag{4.3}$$

Moreover, if $f(X), g(Y) \in \mathcal{L}^1(\mathbf{P})$ then $f(X)g(Y) \in \mathcal{L}^1(\mathbf{P})$ and equation 4.3 holds.

Proof. Let $h : \mathbb{R}^2 \to \mathbb{R}$, $h(x,y) = f(x)g(y)$. First, let us assume that $f, g \geq 0$ then, by theorem 4.6 and property 4.4,

$$\begin{aligned}
\mathbf{E}[f(X)g(Y)] &= \mathbf{E}[h(X,Y)] \\
&= \int h \, \mathbf{P}_X \otimes \mathbf{P}_Y \\
&= \int \left(\int f(x)g(y)\mathbf{P}_X(dx) \right) \mathbf{P}_Y(dy) \\
&= \int \mathbf{E}[f(X)]g(y)\mathbf{P}_Y(dy) \\
&= \mathbf{E}[f(X)]\mathbf{E}[g(Y)].
\end{aligned}$$

Now, let us assume that $f(X), g(Y) \in \mathcal{L}^1(\mathbf{P})$ then, by virtue of property 4.4, $f \in \mathcal{L}^1(\mathbf{P}_X)$ and $g \in \mathcal{L}^1(\mathbf{P}_Y)$, and thus

$$\begin{aligned}
\int |h| \, \mathbf{P}_X \otimes \mathbf{P}_Y &= \int \left(\int |f(x)||g(y)|\mathbf{P}_X(dx) \right) \mathbf{P}_Y(dy) \\
&= \int |f(x)|\mathbf{P}_X(dx) \int |g(y)|\mathbf{P}_Y(dy) \\
&< +\infty.
\end{aligned}$$

Hence, $h \in \mathcal{L}^1(\mathbf{P}_X \otimes \mathbf{P}_Y)$, and thus $h \in \mathcal{L}^1(\mathbf{P}_{X,Y})$, by the independence of X and Y. Finally, by the image measure property, $f(X)g(Y) = h(X,Y) \in \mathcal{L}^1(\mathbf{P})$, and the conclusion follows from theorem 4.6. □

Corollary 4.8. Let X and Y independent real random variables. If $X, Y \geq 0$ then

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]. \tag{4.4}$$

Moreover, if $X, Y \in \mathcal{L}^1(\mathbf{P})$ then $XY \in \mathcal{L}^1(\mathbf{P})$ and equation 4.4 holds.

Proof. First, let us suppose that $X, Y \geq 0$ then, considering $f, g : \mathbb{R} \to \mathbb{R}$, $f(x) = g(x) = |x|$, the conclusion follows, by corollary 4.7. Now, let us suppose that $X, Y \in \mathcal{L}^1(\mathbf{P})$ then, now considering $f(x) = g(x) = x$, the conclusion follows again from corollary 4.7. $\square$

And using an inductive argument it holds:

Corollary 4.9. Let $X_1, \dots, X_n$ be independent real random variables. If $X_1, \dots, X_n \geq 0$ then

$$\mathbf{E}[X_1 \cdots X_n] = \mathbf{E}[X_1] \cdots \mathbf{E}[X_n]. \tag{4.5}$$

Moreover, if $X_1, \dots, X_n \in \mathcal{L}^1(\mathbf{P})$ then $X_1 \cdots X_n \in \mathcal{L}^1(\mathbf{P})$ and equation 4.5 holds.

Proof of the theorem 4.6. We shall proceed in four steps, and in each of them equation 4.2 follows:

Step 1. Let us assume that h is an indicator function. Let $h = \mathbf{1}_A$ for some $A \in \mathcal{B}(\mathbb{R}^2)$; then

$$\begin{aligned}
\int \mathbf{1}_A(X, Y) d\mathbf{P} &= \int \mathbf{1}_{(X,Y)^{-1}(A)} d\mathbf{P} \\
&= \mathbf{P}((X, Y)^{-1}(A)) \\
&= \mathbf{P}_{X,Y}(A) \\
&= \mathbf{P}_X \otimes \mathbf{P}_Y(A) \\
&= \int \mathbf{1}_A \, d\mathbf{P}_X \otimes \mathbf{P}_Y.
\end{aligned}$$

Step 2. Let us assume that h is a simple and positive function. Let $a_1 \mathbf{1}_{A_1} + \cdots + a_n \mathbf{1}_{A_n}$ be a normal representation of h; then, by step 1,

$$\begin{aligned}
\mathbf{E}[h(X, Y)] &= a_1 \mathbf{E}[\mathbf{1}_{A_1}(X, Y)] + \cdots + a_n \mathbf{E}[\mathbf{1}_{A_n}(X, Y)] \\
&= a_1 \int \mathbf{1}_{A_1} \, d\mathbf{P}_X \otimes \mathbf{P}_Y + \cdots + a_n \int \mathbf{1}_{A_n} \, d\mathbf{P}_X \otimes \mathbf{P}_Y \\
&= \int h \, d\mathbf{P}_X \otimes \mathbf{P}_Y.
\end{aligned}$$

Step 3. Let us assume that $h \geq 0$. Let (h_n) be a sequence of simple and positive functions such that $h_n \uparrow h$; then, by step 2 and the monotone convergence theorem,

$$\begin{aligned}
\mathbf{E}[h(X, Y)] &= \lim_{n \to +\infty} \mathbf{E}[h_n(X, Y)] \\
&= \lim_{n \to +\infty} \int h_n \, d\mathbf{P}_X \otimes \mathbf{P}_Y \\
&= \int h \, d\mathbf{P}_X \otimes \mathbf{P}_Y.
\end{aligned}$$

Step 4. Let us assume that $h(X,Y) \in \mathcal{L}^1(\mathbf{P})$. Then, by the image measure property, $h \in \mathcal{L}^1(\mathbf{P}_X \otimes \mathbf{P}_Y)$ and, by step 3,

$$\begin{aligned}
\mathbf{E}[h(X,Y)] &= \mathbf{E}[h(X,Y)^+] - \mathbf{E}[h(X,Y)^-] \\
&= \mathbf{E}[h^+(X,Y)] - \mathbf{E}[h^-(X,Y)] \\
&= \int h^+ \, d\mathbf{P}_X \otimes \mathbf{P}_Y - \int h^- \, d\mathbf{P}_X \otimes \mathbf{P}_Y \\
&= \int h \, d\mathbf{P}_X \otimes \mathbf{P}_Y.
\end{aligned}$$

□

In the following, an important identity that simplifies the calculation of the expected value of the sum of a random number of random quantities.

Theorem 4.10 (Wald's equation)**.** Let $N, X_1, X_2, \ldots$ be independent and integrable real random variables. If N takes nonnegative integer values and $X_1, X_2, \ldots$ are identically distributed then $X_1 + \cdots + X_N \in \mathcal{L}^1(\mathbf{P})$ and

$$\mathbf{E}[X_1 + \cdots + X_N] = \mathbf{E}[N]\mathbf{E}[X_1].$$

Proof. Let $S_N := X_1 + \cdots + X_N$ and let $S_n := X_1 + \cdots + X_n$, for every $n \in \mathbb{N}$. Then

$$S_N = \sum_{n=1}^{+\infty} S_n \mathbf{1}_{(N=n)}.$$

Hence, by corollary 4.8, because $|S_n|$ and $\mathbf{1}_{(N=n)}$ are independent,

$$\begin{aligned}
\mathbf{E}[|S_N|] &= \sum_{n=1}^{+\infty} \mathbf{E}[|S_n|\mathbf{1}_{(N=n)}] \\
&= \sum_{n=1}^{+\infty} \mathbf{E}[|S_n|]\mathbf{E}[\mathbf{1}_{(N=n)}] \\
&\leq \sum_{n=1}^{+\infty} n\mathbf{E}[|X_1|]\mathbf{P}(N=n) \\
&\leq \mathbf{E}[|X_1|]\mathbf{E}[N].
\end{aligned}$$

Thus, $S_N \in \mathcal{L}^1(\mathbf{P})$, and the same computation without absolute values yields the remaining part of the claim. □

Property 4.11 (Rules for variance and covariance)**.** Let $X, Y, X_1, \ldots, X_n$ be square integrable real random variables and $\alpha \in \mathbb{R}$. Then:

(i) $\mathbf{Var}[X] = 0 \quad \Leftrightarrow \quad X = \mathbf{E}[X]$ a.s.

(ii) The map $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto \mathbf{E}[(X-x)^2]$, is minimal at $\mathbf{E}[X]$ with $f(\mathbf{E}[X]) = \mathbf{Var}[X]$.

(iii) $\mathbf{Var}[\alpha X] = \alpha^2 \mathbf{Var}[X]$.

(iv) The map $\mathbf{Cov} : \mathcal{L}^2(\mathbf{P}) \times \mathcal{L}^2(\mathbf{P}) \to \mathbb{R}$, $\mathbf{Cov}[X,Y]$, is a positive semidefinite symmetric bilinear form and X is almost surely constant if $\mathbf{Cov}[X,X] = 0$.

(v) If $X_1 + \cdots + X_n$ are uncorrelated, then

$$\mathbf{Var}[X_1 + \cdots + X_n] = \mathbf{Var}[X_1] + \cdots + \mathbf{Var}[X_n].$$

(vi) [Cauchy-Schwarz inequality]

$$\mathbf{Cov}[X,Y]^2 \leq \mathbf{Var}[X]\mathbf{Var}[Y].$$

Equality holds iff there are $a, b, c \in \mathbb{R}$ with $|a| + |b| + |c| > 0$ and $aX + bY + c = 0$ a.s.

Proof.

(ii) Since $f(x) = \mathbf{Var}[X] + (\mathbf{E}[X] - x)^2$, the conclusion follows.

(vi) We shall prove only the case where $\mathbf{Var}[Y] > 0$. Let $\theta := -\mathbf{Cov}[X,Y]/\mathbf{Var}[Y]$. Then

$$\begin{aligned} 0 &\leq \mathbf{Var}[X + \theta Y]\mathbf{Var}[Y] \\ &= (\mathbf{Var}[X] + 2\theta\mathbf{Cov}[X,Y] + \theta^2\mathbf{Var}[Y])\mathbf{Var}[Y] \\ &= \mathbf{Var}[X]\mathbf{Var}[Y] - \mathbf{Cov}[X,Y]^2 \end{aligned}$$

with equality if and only if $X + \theta Y$ is a.s. constant. Now let $a = 1$, $b = \theta$ and $c = -\mathbf{E}[X] - b\mathbf{E}[Y]$.

□

Example 4.12. Let $X : \Omega \to \mathbb{R}$ be a random variable.

(i) Let $p \in [0,1]$ and let $X \sim \mathrm{Ber}_p$. Then $\mathbf{E}[X] = p$ and $\mathbf{Var}[X] = p(1-p)$.

(ii) Let $n \in \mathbb{N}$ and $p \in [0,1]$, and let $X \sim \mathrm{B}_{n,p}$. Then $\mathbf{E}[X] = np$ and $\mathbf{Var}[X] = np(1-p)$.

(iii) Let $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, and let $X \sim \mathcal{N}_{\mu,\sigma^2}$. Then $\mathbf{E}[X] = \mu$ and $\mathbf{Var}[X] = \sigma^2$.

(iv) Let $\theta > 0$ and let $X \sim \exp_\theta$. Then $\mathbf{E}[X] = \frac{1}{\theta}$ and $\mathbf{Var}[X] = \frac{1}{\theta^2}$.

In the following, an important identity that simplifies the calculation of the variance value of the sum of a random number of random quantities.

Theorem 4.13 (Blackwell-Girshick equation)**.** If $N, X_1, X_2, \dots$ are independent and square integrable, N takes nonnegative integer values and $X_1, X_2, \dots$ are identically distributed then $X_1 + \cdots + X_N \in \mathcal{L}^2(\mathbf{P})$ and

$$\mathbf{Var}[X_1 + \cdots + X_N] = \mathbf{Var}[N]\mathbf{E}[X_1]^2 + \mathbf{E}[N]\mathbf{Var}[X_1].$$

Proof. Exercise. □

Exercise 4.1. Let $h : \mathbb{R} \to \mathbb{R}$ be measurable and let X and Y be identically distributed. Prove the following propositions.

(i) $h(X)$ is integrable iff $h(Y)$ is integrable.

(ii) In this case we have $\mathbf{E}[h(X)] = \mathbf{E}[h(Y)]$.

Exercise 4.2. Let X be integrable and with symmetric distribution. Prove the following propositions.

(i) $\mathbf{E}[X] = 0$.

(ii) If $h : \mathbb{R} \to \mathbb{R}$ is measurable and odd, then $h(X)$ has symmetric distribution.

(iii) If $X \sim \mathcal{N}_{0,1}$, then $\mathbf{E}[X^k] = 0$ for every odd $k \geq 1$.

Exercise 4.3. Prove that if $X \in \mathcal{L}^1(\mathbf{P})$ and has density f, then

$$\mathbf{E}[X] = \int x f(x) \lambda(dx).$$

Exercise 4.4. Assume that (X, Y) are uniformly distributed on a circle with radius a, then

$$f_{(X,Y)}(x, y) = \begin{cases} \frac{1}{\pi a^2} & \text{if } x^2 + y^2 \leq a^2, \\ 0 & \text{elsewhere.} \end{cases}$$

Find $\mathbf{E}[X]$.

Exercise 4.5. Suppose that X and Y are independent with probability densities:

$$f_X(x) = \begin{cases} \frac{8}{x^3} & \text{if } x > 2, \\ 0 & \text{elsewhere,} \end{cases}$$

and

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

Find $\mathbf{E}[XY]$.

Exercise 4.6. Let $X_1, X_2, \dots \geq 0$ be i.i.d. Prove the following propositions.

(i)

$$\limsup_{n\to+\infty} \frac{1}{n} X_n = \begin{cases} 0 \text{ a.s.} & \text{if } \mathbf{E}[X_1] < +\infty, \\ +\infty \text{ a.s.} & \text{if } \mathbf{E}[X_1] = +\infty. \end{cases}$$

(ii) For any $c \in]0,1[$:

$$\sum_{n=1}^{+\infty} e^{X_n} c^n \begin{cases} < +\infty \text{ a.s.} & \text{if } \mathbf{E}[X_1] < +\infty, \\ = +\infty \text{ a.s.} & \text{if } \mathbf{E}[X_1] = +\infty. \end{cases}$$

Exercise 4.7. Let $\Omega =]0,1[$, $\mathcal{A}$ be the class of Borel sets and $\mathbf{P}$ be the Lebesgue measure. If $X_n(\omega) = \sin(2\pi n\omega)$, $n = 1, 2, \ldots$, then prove that $X_1, X_2, \ldots$ are uncorrelated but not independent.

Exercise 4.8. Prove that if $\mathbf{P}[X \in [0,1]] = 1$, then $\mathbf{Var}[X] \leq 1/4$.

Exercise 4.9. By investing in a particular stock, a person can make a profit in one year of \$4,000 with probability 0.3 or take a loss of \$1,000 with probability 0.7.

(i) What is the person's expected gain?

(ii) What is the variance?

Exercise 4.10. Suppose that X represents the number of errors per 100 lines of software code and has the following probability distribution:

X	2	3	4	5	6
Probability	0.01	0.25	0.40	0.30	0.04

(i) Find the variance of X

(ii) Find the mean and variance of $3X - 2$.

Exercise 4.11. Let a six-sided die. Take the number on the die (call it T) and roll that number of six-sided dice to get the numbers $X_1, \ldots, X_T$, and add up their values. What is the expected value of this sum?

Exercise 4.12. Let a particle in the x axis with probability $2/3$ to move one meter to the right and $1/3$ to move one meter to the left. Take a number on $\mathbb{N}_0$ and call it T; suppose that $T \sim \text{Poi}_3$. Then starting at the origin, the particle performs T movements along the axis, say $X_1, \ldots, X_T$. What is the expected final position of this particle?