Lecture 9

1 Expectation

Definition 1.1 (Expectation – Non-Negative Random Variables)**.** *Let X be a non-negative random variable, i.e., $X(\omega) \geq 0$ for every $\omega \in \Omega$. Then the* expectation *of X is defined as*

$$\mathbb{E}\left(X\right) = \sum_{\omega \in \Omega} \mathbb{P}\left(\omega\right) X\left(\omega\right).$$

Note that it could be the case that $\mathbb{E}\left(X\right) = \infty$, which is still well-defined.

Example 1:

1. Let X be the outcome of the roll of a fair die. Then $\mathbb{E}\left(X\right) = 7/2$. Indeed, we can model this as follows:

$$\begin{aligned}
&\Omega = \left\{1, 2, 3, 4, 5, 6\right\}, \\
&\mathbb{P}\left(\omega\right) = \frac{1}{6} \quad \text{for every } \omega \in \Omega, \\
&X : \Omega \to \mathbb{R}, \quad X(\omega) = \omega, \\
&\mathbb{E}\left(X\right) = \sum_{\omega \in \Omega} \mathbb{P}\left(\omega\right) \cdot X(\omega) = \frac{1}{6}\left(1+2+3+4+5+6\right) = 7/2.
\end{aligned}$$

2. A coin with probability $1/3$ for heads is tossed twice. Let Y be the total number of heads in both tosses. Then $\mathbb{E}\left(Y\right) = 2/3$. Indeed, we can model this as follows:

$$\begin{aligned}
&\Omega = \left\{(1,1), (1,0), (0,1), (0,0)\right\}, \\
&\mathbb{P}\left((1,1)\right) = \frac{1}{9},\ \mathbb{P}\left((1,0)\right) = \frac{2}{9}, \\
&\mathbb{P}\left((0,1)\right) = \frac{2}{9},\ \mathbb{P}\left((0,0)\right) = \frac{4}{9}, \\
&Y : \Omega \to \mathbb{R}, \quad Y(\omega_1, \omega_2) = \omega_1 + \omega_2, \\
&\mathbb{E}\left(Y\right) = \sum_{(\omega_1,\omega_2) \in \Omega} \mathbb{P}\left((\omega_1, \omega_2)\right) \cdot Y\left((\omega_1, \omega_2)\right) = \frac{1}{9} \cdot 2 + \frac{2}{9} \cdot 1 + \frac{2}{9} \cdot 1 + \frac{4}{9} \cdot 0 = 2/3.
\end{aligned}$$

3. Let

$$\Omega = \mathbb{N} = \{1, 2, \ldots\},$$
$$\mathbb{P}(k) = \frac{1}{2^k} \quad \text{for every } k \in \Omega,$$
$$Z : \Omega \to \mathbb{R}, \quad Z(k) = 2^k.$$

Then
$$\mathbb{E}(Z) = \sum_{k \in \Omega} \mathbb{P}(k) \cdot Z(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot 2^k = \infty.$$

Definition 1.2 (Positive and Negative Parts)**.** *Let X be a random variable. Define two new random variables X^+ and X^-, called the* positive part *of X and the* negative part *of X, respectively, as follows:*

$$X^+ = \begin{cases} X(\omega) & \text{if } X(\omega) \geq 0 \\ 0 & \text{otherwise} \end{cases} \qquad X^- = \begin{cases} -X(\omega) & \text{if } X(\omega) < 0 \\ 0 & \text{otherwise} \end{cases}$$

Observe that, by definition, both X^+ and X^- are non-negative random variables, and it holds that $|X| = X^+ + X^-$ and $X = X^+ - X^-$.

Definition 1.3 (Expectation)**.** *Let X be a random variable.*

1. Finite expectation: *If $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$, then $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$, that is*
$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega),$$
and the series converges absolutely (See Theorem 1.4 below).

2. Infinite expectation: *If $\mathbb{E}(X^+) = \infty$ and $\mathbb{E}(X^-) < \infty$, then $\mathbb{E}(X) := \infty$. Similarly, if $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) = \infty$, then $\mathbb{E}(X) := -\infty$.*

3. Undefined expectation: *If $\mathbb{E}(X^+) = \infty$ and $\mathbb{E}(X^-) = \infty$, then we say that the expectation of X is undefined (or, equivalently, that X has no expectation).*

Theorem 1.4. *$\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$ if and only if the sum $\sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega)$ converges absolutely. Moreover, in this case, $\sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega)$ converges to $\mathbb{E}(X^+) - \mathbb{E}(X^-)$.*

Proof. Assume first that $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$. Then

$$\begin{aligned} \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) &= \sum_{\omega \in \Omega} \left(X^+(\omega) - X^-(\omega)\right) \mathbb{P}(\omega) \\ &= \sum_{\omega \in \Omega} X^+(\omega) \mathbb{P}(\omega) - \sum_{\omega \in \Omega} X^-(\omega) \mathbb{P}(\omega) \\ &= \mathbb{E}\left(X^+\right) - \mathbb{E}\left(X^-\right), \end{aligned}$$

where the second equality holds since $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$ imply that $\sum_{\omega\in\Omega} X^+(\omega)\,\mathbb{P}(\omega)$ and $\sum_{\omega\in\Omega} X^-(\omega)\,\mathbb{P}(\omega)$ converge absolutely; in particular $\sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega)$ converges absolutely as it is the difference of two sums, each converging absolutely.

Next, assume that $\sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega)$ converges absolutely. Then

$$\mathbb{E}\left(X^+\right) + \mathbb{E}\left(X^-\right) = \sum_{\omega\in\Omega} X^+(\omega)\,\mathbb{P}(\omega) + \sum_{\omega\in\Omega} X^-(\omega)\,\mathbb{P}(\omega) = \sum_{\omega\in\Omega} |X(\omega)|\,\mathbb{P}(\omega) < \infty,$$

implying that $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$. It now follows by the same calculation as above that

$$\sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega) = \mathbb{E}\left(X^+\right) - \mathbb{E}\left(X^-\right).$$

□

The next claim, states that the expectation of a random variable depends solely on its distribution, rather than its definition as a function in a specific probability space.

Theorem 1.5. *Let $X : \Omega \to S$ be a non-negative random variable. Then*

$$\mathbb{E}(X) = \sum_{x\in S} x \cdot \mathbb{P}(X = x).$$

Proof. We have

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{\omega\in\Omega} X(\omega)\cdot\mathbb{P}(\omega) = \sum_{x\in S}\sum_{\omega\in\Omega : X(\omega)=x} X(\omega)\cdot\mathbb{P}(\omega) = \sum_{x\in S}\sum_{\omega\in\Omega : X(\omega)=x} x\cdot\mathbb{P}(\omega) \\
&= \sum_{x\in S} x \cdot \sum_{\omega\in\Omega : X(\omega)=x} \mathbb{P}(\omega) = \sum_{x\in S} x\cdot\mathbb{P}(\{\omega\in\Omega : X(\omega) = x\}) = \sum_{x\in S} x\cdot\mathbb{P}(X = x),
\end{aligned}$$

where the second equality holds since in a series with non-negative summands, changing the order of summation does not change the value of the series. □

Theorem 1.6. *Let $X : \Omega \to S$ be a random variable with finite expectation. Then*

$$\mathbb{E}(X) = \sum_{x\in S} x \cdot \mathbb{P}(X = x),$$

and the sum converges absolutely. The opposite direction is true as well, that is, if

$$\sum_{x\in S} x \cdot \mathbb{P}(X = x),$$

converges absolutely, then $\mathbb{E}(X)$ is finite and is equal to this sum.

Proof. Since X^+ and X^- are non-negative random variables, it follows by Theorem 1.5 that

$$\begin{aligned}
\mathbb{E}\left(X^+\right) &= \sum_{x\in S} x\cdot\mathbb{P}\left(X^+ = x\right) = \sum_{x\in S : x\geq 0} x\cdot\mathbb{P}(X = x) \qquad \text{and} \\
\mathbb{E}\left(X^-\right) &= \sum_{x\in S} x\cdot\mathbb{P}\left(X^- = x\right) = \sum_{x\in S : x< 0} (-x)\cdot\mathbb{P}(X = x).
\end{aligned}$$

Assume first that $\mathbb{E}(X) < \infty$. It follows by Definition 1.3 that $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$, implying that $\sum_{x \in S : x \geq 0} x \cdot \mathbb{P}(X = x)$ and $\sum_{x \in S : x < 0} (-x) \cdot \mathbb{P}(X = x)$ converge absolutely. Then, by Definition 1.3 we have

$$\begin{aligned}\mathbb{E}(X) = \mathbb{E}\left(X^+\right) - \mathbb{E}\left(X^-\right) &= x \cdot \sum_{x \in S : x \geq 0} \mathbb{P}(X = x) - \sum_{x \in S : x \leq 0} (-x) \cdot \mathbb{P}(X = x) \\ &= \sum_{x \in S} x \cdot \mathbb{P}(X = x), \end{aligned} \tag{1}$$

and the last sum converges absolutely as it is the difference of two sums, each converging absolutely. For the opposite direction, assume that $\sum_{x \in S} x \cdot \mathbb{P}(X = x)$ converges absolutely. Then

$$\sum_{x \in S} |x| \cdot \mathbb{P}(X = x) = \sum_{x \in S : x \geq 0} x \cdot \mathbb{P}(X = x) + \sum_{x \in S : x < 0} (-x) \cdot \mathbb{P}(X = x) = \mathbb{E}\left(X^+\right) + \mathbb{E}\left(X^-\right).$$

Hence $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$. Repeating the calculation Equation (1) completes the proof. $\square$

We can use Theorems 1.5 and 1.6 to calculate the expectation of a random variable X, using only its distribution. In the following subsection we do so for some of the common distributions defined in Lectures 7 and 8.

1.1 Expectation of Common Distributions

1.1.1 Uniform Distribution

Recall the definition of the Uniform distribution: $X \sim \mathrm{U}(S)$ if

$$\mathbb{P}(X = s) = \begin{cases} \frac{1}{|S|} & s \in S \\ 0 & s \notin S \end{cases}$$

Then we have the following claim.

Claim 1.7. *If $S = \{a, a+1, \ldots, b\}$, for some $a, b \in \mathbb{N}$, then $\mathbb{E}(X) = \frac{a+b}{2}$.*

Proof. It holds that

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=a}^{b} k \cdot \mathbb{P}(X = k) \\ &= \sum_{k=a}^{b} \frac{k}{b-a+1} \\ &= \frac{1}{b-a+1} \sum_{k=a}^{b} k \\ &= \frac{1}{b-a+1} \cdot \frac{(b+a) \cdot (b-a+1)}{2} \\ &= \frac{a+b}{2}.\end{aligned}$$

□

1.1.2 Bernoulli Distribution

Recall the definition of the Bernoulli distribution: $X \sim \text{Ber}(p)$ if

$$X \sim \begin{cases} 1 & p \\ 0 & 1-p \end{cases}$$

Claim 1.8. $\mathbb{E}(X) = p$.

Proof.

$$\mathbb{E}(X) = p \cdot 1 + (1-p) \cdot 0 = p.$$

□

In particular, for any event A, we have that

$$\mathbb{E}(1_A) = \mathbb{P}(A).$$

1.1.3 Binomial Distribution

Recall the definition of the Binomial distribution: $X \sim \text{Bin}(n, p)$ if

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

for every integer $k \in \{0, 1, \ldots, n\}$.

Claim 1.9. $\mathbb{E}(X) = np$.

The proof is left as an exercise.

1.1.4 Geometric Distribution

Recall the definition of the Geometric distribution: $X \sim \text{Geom}(p)$ if

$$\mathbb{P}(X = k) = (1-p)^{k-1} \cdot p,$$

for every integer $k \geq 1$.

Claim 1.10. $\mathbb{E}(X) = \frac{1}{p}$.

Proof.

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{k=1}^{\infty}(1-p)^{k-1}\cdot p\cdot k \\
&= \sum_{k=1}^{\infty}(1-p)^{k-1}\cdot p\cdot[(k-1)+1] \\
&= \sum_{k=1}^{\infty}(1-p)^{k-1}\cdot p\cdot(k-1)+\sum_{k=1}^{\infty}(1-p)^{k-1}\cdot p \\
&= \sum_{m=0}^{\infty}(1-p)^{m}\cdot p\cdot m+1 \\
&= (1-p)\cdot\sum_{m=1}^{\infty}(1-p)^{m-1}\cdot p\cdot m+1 \\
&= (1-p)\cdot\mathbb{E}(X)+1,
\end{aligned}
$$

where the fourth equality holds by the substitution $m = k - 1$ and since the second sum is the sum of probabilities of the random variable X and thus equals 1. It follows that

$$\mathbb{E}(X) = (1-p)\cdot\mathbb{E}(X)+1 = \mathbb{E}(X) - p\mathbb{E}(X)+1$$

entailing $\mathbb{E}(X) = \frac{1}{p}$ as claimed. □

1.1.5 Negative-Binomial Distribution

Recall the definition of the Negative-Binomial distribution: $X \sim \mathrm{NB}(r, p)$ if

$$\mathbb{P}(X = n) = \binom{n-1}{r-1}\cdot p^r\cdot(1-p)^{n-r},$$

for every integer $n \geq r$.

Claim 1.11. $\mathbb{E}(X) = \frac{r}{p}$.

Proof. Observe that

$$\binom{a}{b} = \frac{a!}{b!(a-b)!} = \frac{a}{b}\cdot\frac{(a-1)!}{(b-1)![(a-1)-(b-1)]!} = \frac{a}{b}\cdot\binom{a-1}{b-1}$$

and thus

$$\binom{a-1}{b-1} = \frac{b}{a}\cdot\binom{a}{b}. \tag{2}$$

Hence

$$\mathbb{E}(X) = \sum_{n=r}^{\infty} n\cdot\binom{n-1}{r-1}\cdot p^r\cdot(1-p)^{n-r} = \frac{r}{p}\cdot\sum_{n=r}^{\infty}\binom{n}{r}\cdot p^{r+1}\cdot(1-p)^{n-r},$$

where the second equality holds by (2). Therefore, it suffices to prove that

$$\sum_{n=r}^{\infty} \binom{n}{r} \cdot p^{r+1} \cdot (1-p)^{n-r} = 1.$$

Indeed, using the substitution $m = n+1$, implies that

$$\sum_{n=r}^{\infty} \binom{n}{r} \cdot p^{r+1} \cdot (1-p)^{n-r} = \sum_{m=r+1}^{\infty} \binom{m-1}{r} \cdot p^{r+1} \cdot (1-p)^{m-(r+1)} = 1,$$

where the last equality holds since its left hand side is the sum of probabilities of a random variable $Y \sim \mathrm{NB}\left(r+1, p\right)$ over the support of its distribution. $\square$

1.1.6 Hypergeometric Distribution

Recall the definition of the Hypergeometric distribution: $X \sim \mathrm{Hyp}\left(N, D, n\right)$ if

$$\mathbb{P}\left(X = k\right) = \frac{\binom{D}{k} \cdot \binom{N-D}{n-k}}{\binom{N}{n}},$$

for every integer $0 \leq k \leq n$.

Claim 1.12. $\mathbb{E}\left(X\right) = n \cdot \frac{D}{N}$.

The proof will be presented in the practical session.

1.1.7 Poisson Distribution

Recall the definition of the Poisson distribution: $X \sim \mathrm{Poi}\left(\lambda\right)$ if

$$\mathbb{P}\left(X = k\right) = e^{-\lambda} \cdot \frac{\lambda^k}{k!},$$

for every integer $k \geq 0$.

Claim 1.13. $\mathbb{E}\left(X\right) = \lambda$.

The proof will be presented in the practical session.

1.2 Linearity of Expectation

Theorem 1.14 (Linearity of Expectation)**.** *Let X and Y be two random variables with a defined expectation, and let $c \in \mathbb{R}$. Then*

1. Homogeneity: *The expectation of cX is defined and it holds that $\mathbb{E}\left(cX\right) = c\mathbb{E}\left(X\right)$.*

2. Linearity: *If it does not holds that* $\mathbb{E}(X) = \infty$ *and* $\mathbb{E}(Y) = -\infty$, *or* $\mathbb{E}(X) = -\infty$ *and* $\mathbb{E}(Y) = \infty$, *then the expectation of* $X + Y$ *is defined, and it further holds that* $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Proof.

1. The case where $c = 0$ is trivial (we assume that $0 \cdot \infty = 0$). We prove the claim for $c > 0$, while the case where $c < 0$ can be proven analogously. Since $(cX)^+ = c(X^+)$ it follows that

$$\mathbb{E}\left((cX)^+\right) = \sum_{\omega\in\Omega} (cX)^+(\omega)\,\mathbb{P}(\omega) = c \cdot \sum_{\omega\in\Omega} X^+(\omega)\,\mathbb{P}(\omega) = c\mathbb{E}\left(X^+\right).$$

Similarly, $\mathbb{E}((cX)^-) = c\mathbb{E}(X^-)$. Therefore $\mathbb{E}(cX)$ is defined and it holds that

$$\mathbb{E}(cX) = \mathbb{E}\left((cX)^+\right) - \mathbb{E}\left((cX)^-\right) = c\left(\mathbb{E}\left(X^+\right) - \mathbb{E}\left(X^-\right)\right) = c\mathbb{E}(X).$$

2. We will only prove the statement in the case case where both $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ are finite. It holds that

$$\mathbb{E}(X+Y) = \sum_{\omega\in\Omega} (X+Y)(\omega)\,\mathbb{P}(\omega) = \sum_{\omega\in\Omega} X(\omega)\,\mathbb{P}(\omega) + \sum_{\omega\in\Omega} Y(\omega)\,\mathbb{P}(\omega) = \mathbb{E}(X) + \mathbb{E}(Y),$$

where the series $\sum_{\omega\in\Omega}(X+Y)(\omega)\,\mathbb{P}(\omega)$ corresponding to $\mathbb{E}(X+Y)$ absolutely converges since it is the sum of two absolutely converging series.

□

Claim 1.15. *[Monotonicity of Expectation] Let* X *and* Y *be two random variables satisfying* $\mathbb{P}(X \leq Y) = 1$. *If* $\mathbb{E}(X)$ *and* $\mathbb{E}(Y)$ *are defined, then* $\mathbb{E}(X) \leq \mathbb{E}(Y)$, *and equality is attained if and only if* $\mathbb{P}(X = Y) = 1$.

Proof. Observe first that the assumption $\mathbb{P}(X \leq Y) = 1$ implies that, for every $\omega \in \Omega$, it holds that $X(\omega) \leq Y(\omega)$ or $\mathbb{P}(\omega) = 0$. Indeed, let $A = \{\omega \in \Omega : X(\omega) > Y(\omega)\}$, then $\mathbb{P}(X \leq Y) = 1$ implies that

$$0 = \mathbb{P}(X > Y) = \mathbb{P}(\{\omega \in \Omega : X(\omega) > Y(\omega)\}) = \mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\omega).$$

Therefore, $X(\omega) \leq Y(\omega)$ or $\mathbb{P}(\omega) = 0$ for every $\omega \in \Omega$.

It is enough to show that $\mathbb{E}(Y^-) \leq \mathbb{E}(X^-)$ and $\mathbb{E}(Y^+) \geq \mathbb{E}(X^+)$. We will only prove the latter, as the former can be proven analogously.

$$\mathbb{E}\left(X^+\right) = \sum_{\omega\in\Omega} X^+(\omega)\mathbb{P}(\omega) \leq \sum_{\omega\in\Omega} Y^+(\omega)\mathbb{P}(\omega) = \mathbb{E}\left(Y^+\right),$$

where the inequality holds since, as previously noted, $X(\omega) \leq Y(\omega)$ or $\mathbb{P}(\omega) = 0$ for every $\omega \in \Omega$. Finally, observe that equality can be attained if and only if for every $\omega \in \Omega$ it holds that $X(\omega) = Y(\omega)$ or $\mathbb{P}(\omega) = 0$, i.e., if and only if $\mathbb{P}(X = Y) = 1$. □