Advanced Probability: WEEK 3

1 Expectation

From the point of view of measure theory, random variables are maps from $\Omega$ to $\mathbb{R}$. Correspondingly, the expectation (or mean) of a random variable $X$ is the Lebesgue integral of the map $X$, that is, the "area under the curve $\omega \mapsto X(\omega)$", where the horizontal axis is measured with the probability measure $\mathbb{P}$.

1.1 Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X$ be a random variable defined on this probability space. The expectation of $X$, denoted as $\mathbb{E}(X)$, will be defined in three steps.

**Step 1.** Assume first that $X$ is a non-negative discrete random variable, i.e. that $X$ may be written as $X(\omega) = \sum_{i=1}^{\infty} x_i 1_{A_i}(\omega)$, where $x_i \geq 0$ are distinct and $A_i \in \mathcal{F}$ are disjoint (notice that $A_i = \{X = x_i\}$). The expectation of $X$ is then defined as

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \mathbb{P}(A_i),$$

which corresponds to the traditional definition of expectation in elementary probability courses. Notice here that since the sum is infinite, $\mathbb{E}(X)$ may take the value $+\infty$; but because of the assumption that $x_i \geq 0$, $\mathbb{E}(X)$ is always non-negative.

Notice also that in the particular case where $X = 1_A$, with $A \in \mathcal{F}$, one has $\mathbb{E}(X) = \mathbb{P}(A)$.

**Step 2.** Assume now that $X$ is a generic non-negative random variable (i.e. $X(\omega) \geq 0$, $\forall \omega \in \Omega$). Let us define the following sequence of discrete random variables:

$$X_n(\omega) = \sum_{i=1}^{\infty} \frac{i-1}{2^n} 1_{\left\{ \frac{i-1}{2^n} < X \leq \frac{i}{2^n} \right\}}(\omega).$$

Notice that $x_i = \frac{i-1}{2^n} \geq 0$ and that $\{\frac{i-1}{2^n} < X \leq \frac{i}{2^n}\} \in \mathcal{F}$, since $X$ is $\mathcal{F}$-measurable. So according to Step 1, one has for each $n$

$$\mathbb{E}(X_n) = \sum_{i=1}^{\infty} \frac{i-1}{2^n} \mathbb{P}\left(\left\{ \frac{i-1}{2^n} < X \leq \frac{i}{2^n}\right\}\right) \in [0, +\infty].$$

It should be observed that $(X_n, n \in \mathbb{N})$ is actually an increasing sequence of non-negative "staircases", that is,

$$0 \leq X_n(\omega) \leq X_{n+1}(\omega), \quad \forall n.$$

As the size of the steps is divided by two from $n$ to $n + 1$, the staircase gets refined. Likewise, one easily sees that $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$ for all $n$, so $(\mathbb{E}(X_n), n \in \mathbb{N})$ is an increasing sequence, that therefore converges (possibly to $+\infty$). One defines

$$\mathbb{E}(X) = \lim_{n \to \infty} \mathbb{E}(X_n) = \lim_{n \to \infty} \sum_{i=1}^{\infty} \frac{i-1}{2^n} \mathbb{P}\left(\left\{ \frac{i-1}{2^n} < X \leq \frac{i}{2^n}\right\}\right) \in [0, \infty].$$
Step 3. Finally, consider a generic random variable $X$. One defines its positive and negative parts:

$$X^+(\omega) = \max(0, X(\omega)), \quad X^- (\omega) = \max(0, -X(\omega))$$

Notice that both $X^+(\omega) \geq 0$ and $X^- (\omega) \geq 0$, and that

$$X^+(\omega) - X^- (\omega) = X(\omega), \quad X^+(\omega) + X^- (\omega) = |X(\omega)|.$$ 

In measure theory, one does not want to deal with ill-defined quantities such as $\infty - \infty$. One therefore defines $\mathbb{E}(X)$ only when $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-) < \infty$:

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-).$$

Two important particular cases. Let $X$ be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a Borel-measurable function such that $\mathbb{E}(|g(X)|) < \infty$ (this last condition is verified if for example $g$ is a bounded function).

- If $X$ is a discrete random variable with values in a countable set $C$, then

$$\mathbb{E}(g(X)) = \sum_{x \in C} g(x) \mathbb{P}({X = x}).$$

- If $X$ is a continuous random variable with pdf $p_X$, then

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) p_X(x) \, dx.$$ 

Terminology. - If $\mathbb{E}(|X|) < \infty$, then $X$ is said to be an integrable random variable. 
- If $\mathbb{E}(X^2) < \infty$, then $X$ is said to be a square-integrable random variable.
- If there exists $c > 0$ such that $|X(\omega)| \leq c, \forall \omega \in \Omega$, then $X$ is said to be a bounded random variable.
- If $\mathbb{E}(X) = 0$, then $X$ is said to be a centered random variable.

One has the following series of implications:

- $X$ is bounded $\Rightarrow$ $X$ is square-integrable $\Rightarrow$ $X$ is integrable,
- $X$ is integrable and $Y$ is bounded $\Rightarrow$ $XY$ is integrable,
- $X,Y$ are both square-integrable $\Rightarrow$ $XY$ is integrable.

Negligible and almost sure sets. An event $A \in \mathcal{F}$ is said to be negligible if $\mathbb{P}(A) = 0$. On the contrary, an event $B \in \mathcal{F}$ is said to be almost sure (a.s.) if $\mathbb{P}(B) = 1$. For example, if $\mathbb{P}({X \geq c}) = 1$, one says that “$X \geq c$ almost surely”.

1.2 Basic properties of the expectation

Linearity. If $c \in \mathbb{R}$ is a constant and $X, Y$ are integrable, then both

$$\mathbb{E}(cX) = c\mathbb{E}(X) \quad \text{and} \quad \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Positivity. If $X$ is integrable and $X \geq 0$ a.s., then $\mathbb{E}(X) \geq 0$.

Strict positivity. If $X$ is integrable, $X \geq 0$ a.s. and $\mathbb{E}(X) = 0$, then $X = 0$ a.s.

Monotonicity. If $X, Y$ are integrable and $X \geq Y$ a.s., then $\mathbb{E}(X) \geq \mathbb{E}(Y)$. 

2
1.3 Inequalities

Cauchy-Schwarz’s inequality. If $X$, $Y$ are square-integrable random variables, then the product $XY$ is integrable and
\[ E(|XY|) \leq \sqrt{E(X^2)} \sqrt{E(Y^2)}. \]
In particular, considering $Y = 1$ shows that if $X$ is square-integrable, then it is also integrable.

Jensen’s inequality. If $X$ is a random variable and $\psi : \mathbb{R} \to \mathbb{R}$ is Borel-measurable, convex and such that $E(|\psi(X)|) < \infty$, then
\[ \psi(E(X)) \leq E(\psi(X)). \]
In particular, $|E(X)| \leq E(|X|)$.

Also, if $X$ is such that $P(\{X = a\}) = P(\{X = b\}) = 1/2$, then the above inequality says that
\[ \psi\left(\frac{a+b}{2}\right) \leq \frac{\psi(a) + \psi(b)}{2}, \]
which is pretty much the definition of the convexity of $\psi$!

Chebychev’s inequality. If $X$ is a random variable and $\varphi : \mathbb{R} \to \mathbb{R}_+$ is Borel-measurable, increasing on $\mathbb{R}_+$ and such that $E(\varphi(X)) < \infty$, then for any $a > 0$, one has
\[ P(\{X \geq a\}) \leq \frac{E(\varphi(X))}{\varphi(a)}. \]
In particular, if $X$ is square-integrable, then taking $\varphi(x) = x^2$ gives
\[ P(\{X \geq a\}) \leq \frac{E(X^2)}{a^2}. \]

Proof. Using the assumptions made, we obtain
\[ P(\{X \geq a\}) = \mathbb{E}(1_{\{X \geq a\}}) \leq \mathbb{E}\left(\frac{\varphi(X)}{\varphi(a)} 1_{\{X \geq a\}}\right) \leq \frac{E(\varphi(X))}{\varphi(a)}. \]

1.4 Variance, covariance and independence

Definition 1.1. Let $X, Y$ be two square-integrable random variables. The variance of $X$ is defined as
\[ \text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0 \]
and the covariance of $X$ and $Y$ is defined as
\[ \text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \]

Terminology. If $\text{Cov}(X, Y) = 0$, then $X$ and $Y$ are said to be uncorrelated.

Facts. Let $c \in \mathbb{R}$ be a constant and $X, Y$ be square-integrable random variables.

a) $\text{Var}(cX) = c^2 \text{Var}(X)$.
b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.

In addition, if $X, Y$ are independent, then
c) $\text{Cov}(X, Y) = 0$, i.e. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (but the reciprocal statement is wrong).
d) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. 

3