Lecture #17: Expectation of a Simple Random Variable

Recall that a simple random variable is one that takes on finitely many values.

Definition. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. A random variable $X : \Omega \to \mathbb{R}$ is called *simple* if it can be written as

$$X = \sum_{i=1}^{n} a_i 1_{A_i}$$

where $a_i \in \mathbb{R}$, $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots, n$. We define the expectation of X to be

$$\mathbb{E}(X) = \sum_{i=1}^{n} a_i \mathbf{P}\{A_i\}.$$

Example 17.1. Consider the probability space $(\Omega, \mathcal{B}_1, \mathbf{P})$ where $\Omega = [0, 1]$, $\mathcal{B}_1$ denotes the Borel sets of $[0, 1]$, and $\mathbf{P}$ is the uniform probability on Ω. Suppose that the random variable $X : \Omega \to \mathbb{R}$ is defined by

$$X(\omega) = \sum_{i=1}^{4} a_i 1_{A_i}(\omega)$$

where $a_1 = 4$, $a_2 = 2$, $a_3 = 1$, $a_4 = -1$, and

$$A_1 = [0, \tfrac{1}{2}), \quad A_2 = [\tfrac{1}{4}, \tfrac{3}{4}), \quad A_3 = (\tfrac{1}{2}, \tfrac{7}{8}], \quad A_4 = [\tfrac{7}{8}, 1].$$

Show that there exist finitely many real constants $c_1, \ldots, c_n$ and *disjoint* sets $C_1, \ldots, C_n \in \mathcal{B}_1$ such that

$$X = \sum_{i=1}^{n} c_i 1_{C_i}.$$

Solution. We find

$$X(\omega) = \begin{cases} 4, & \text{if } 0 \le \omega < 1/4, \\ 6, & \text{if } 1/4 \le \omega < 1/2, \\ 2, & \text{if } \omega = 1/2, \\ 3, & \text{if } 1/2 < \omega < 3/4, \\ 1, & \text{if } 3/4 \le \omega < 7/8, \\ 0, & \text{if } \omega = 7/8, \\ -1, & \text{if } 7/8 < \omega \le 1, \end{cases}$$

so that

$$X = \sum_{i=1}^{7} c_i 1_{C_i}$$

where $c_1 = 4$, $c_2 = 6$, $c_3 = 2$, $c_4 = 3$, $c_5 = 1$, $c_6 = 0$, $c_7 = -1$ and

$$C_1 = [0, \tfrac{1}{4}), \quad C_2 = [\tfrac{1}{4}, \tfrac{1}{2}), \quad C_3 = \{\tfrac{1}{2}\}, \quad C_4 = (\tfrac{1}{2}, \tfrac{3}{4}), \quad C_5 = [\tfrac{3}{4}, \tfrac{7}{8}), \quad C_6 = \{\tfrac{7}{8}\}, \quad C_7 = (\tfrac{7}{8}, 1].$$

Proposition 17.2. *If X and Y are simple random variables, then*

$$\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$$

for every α, $\beta \in \mathbb{R}$.

Proof. Suppose that X and Y are simple random variables with

$$X = \sum_{i=1}^{n} a_i 1_{A_i} \quad \text{and} \quad Y = \sum_{j=1}^{m} b_j 1_{B_j}$$

where $A_1, \ldots, A_n \in \mathcal{F}$ and $B_1, \ldots, B_m \in \mathcal{F}$ each partition Ω. Since

$$\alpha X = \alpha \sum_{i=1}^{n} a_i 1_{A_i} = \sum_{i=1}^{n} (\alpha a_i) 1_{A_i}$$

we conclude by definition that

$$\mathbb{E}(\alpha X) = \sum_{i=1}^{n} (\alpha a_i) \mathbf{P}\{A_i\} = \alpha \sum_{i=1}^{n} a_i \mathbf{P}\{A_i\} = \alpha\mathbb{E}(X).$$

The proof of the theorem will be completed by showing $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. Notice that

$$\{A_i \cap B_j : 1 \le i \le n, 1 \le j \le m\}$$

consists of pairwise disjoint events whose union is Ω and

$$X + Y = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j) 1_{A_i \cap B_j}.$$

Therefore, by definition,

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j) \mathbf{P}\{A_i \cap B_j\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mathbf{P}\{A_i \cap B_j\} + \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mathbf{P}\{A_i \cap B_j\} \\
&= \sum_{i=1}^{n} a_i \mathbf{P}\{A_i\} + \sum_{j=1}^{m} b_j \mathbf{P}\{B_j\}
\end{aligned}$$

and the proof is complete. □

Fact. If X and Y are simple random variables with $X \le Y$, then

$$\mathbb{E}(X) \le \mathbb{E}(Y).$$

Exercise 17.3. Prove the previous fact.

Having already defined $\mathbb{E}(X)$ for simple random variables, our goal now is to construct $\mathbb{E}(X)$ in general. To that end, suppose that X is a *positive random variable*. That is, $X(\omega) \geq 0$ for all $\omega \in \Omega$. (We will need to allow $X(\omega) \in [0, +\infty]$ for some consistency.)

Definition. If X is a positive random variable, define the *expectation* of X to be

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Y) : Y \text{ is simple and } 0 \leq Y \leq X\}.$$

That is, we approximate positive random variables by simple random variables. Of course, this leads to the question of whether or not this is possible.

Fact. For every random variable $X \geq 0$, there exists a sequence (X_n) of positive, simple random variables with $X_n \uparrow X$ (that is, X_n increases to X).

An example of such a sequence is given by

$$X_n(\omega) = \begin{cases} \frac{k}{2^n}, & \text{if } \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n} \text{ and } 0 \leq k \leq n2^n - 1, \\ n, & \text{if } X(\omega) \geq n. \end{cases}$$

(Draw a picture.)

Fact. If $X \geq 0$ and (X_n) is a sequence of simple random variables with $X_n \uparrow X$, then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$.

We will prove these facts next lecture.

Now suppose that X is any random variable. Write

$$X^+ = \max\{X, 0\} \quad \text{and} \quad X^- = -\min\{X, 0\}$$

for the *positive part* and the *negative part* of X, respectively.

Note that $X^+ \geq 0$ and $X^- \geq 0$ so that the positive part and negative part of X are both positive random variables and

$$X = X^+ - X^- \quad \text{and} \quad |X| = X^+ + X^-.$$

Definition. A random variable X is called *integrable* (or has *finite expectation*) if both $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ are finite. In this case we define $\mathbb{E}(X)$ to be

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

Statistics 851 (Fall 2013) October 18, 2013
Prof. Michael Kozdron

Lecture #18: Construction of Expectation

Recall that our goal is to define $\mathbb{E}(X)$ for all random variables $X : \Omega \to \mathbb{R}$. We outlined the construction last lecture. Here is the summary of our strategy.

Summary of Strategy for Constructing $\mathbb{E}(X)$ for General Random Variables

We will

(1) define $\mathbb{E}(X)$ for simple random variables,

(2) define $\mathbb{E}(X)$ for positive random variables,

(3) define $\mathbb{E}(X)$ for general random variables.

This strategy is sometimes called the "standard machine" and is the outline that we will follow to prove most results about expectation of random variables.

Step 1: Simple Random Variables

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. Suppose that $X : \Omega \to \mathbb{R}$ is a simple random variable so that

$$X(\omega) = \sum_{j=1}^{m} a_j 1_{A_j}(\omega)$$

where $a_1, \ldots, a_m \in \mathbb{R}$ and $A_1, \ldots, A_m \in \mathcal{F}$. We define the expectation of X to be

$$\mathbb{E}(X) = \sum_{j=1}^{m} a_j \mathbf{P}\left\{A_j\right\}.$$

Step 2: Positive Random Variables

Suppose that X is a positive random variable. That is, $X(\omega) \geq 0$ for all $\omega \in \Omega$. (We will need to allow $X(\omega) \in [0, +\infty]$ for some consistency.) We are also assuming at this step that X is not a simple random variable.

Definition. If X is a positive random variable, define the *expectation* of X to be

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Y) : Y \text{ is simple and } 0 \leq Y \leq X\}. \tag{18.1}$$

Proposition 18.1. *For every random variable $X \geq 0$, there exists a sequence (X_n) of positive, simple random variables with $X_n \uparrow X$ (that is, X_n increases to X).*

Proof. Let $X \geq 0$ be given and define the sequence (X_n) by

$$X_n(\omega) = \begin{cases} \frac{k}{2^n}, & \text{if } \frac{k}{2^n} \leq X(\omega) < \frac{k+1}{2^n} \text{ and } 0 \leq k \leq n2^n - 1, \\ n, & \text{if } X(\omega) \geq n. \end{cases}$$

Then it follows that $X_n \leq X_{n+1}$ for every $n = 1, 2, 3, \dots$ and $X_n \to X$ which completes the proof. □

Proposition 18.2. *If $X \geq 0$ and (X_n) is a sequence of simple random variables with $X_n \uparrow X$, then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$. That is, if $X_n \leq X_{n+1}$ and*

$$\lim_{n\to\infty} X_n = X,$$

then $\mathbb{E}(X_{n+1}) \leq \mathbb{E}(X_n)$ and

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

Proof. Suppose that $X \geq 0$ is a random variable and let (X_n) be a sequence of simple random variables with $X_n \geq 0$ and $X_n \uparrow X$. Observe that since the X_n are increasing, we have $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$. Therefore, $\mathbb{E}(X_n)$ increases to some limit $a \in [0, \infty]$; that is,

$$\mathbb{E}(X_n) \uparrow a$$

for some $0 \leq a \leq \infty$. (If $\mathbb{E}(X_n)$ is an unbounded sequence, then $a = \infty$. However, if $\mathbb{E}(X_n)$ is a bounded sequence, then $a < \infty$ follows from the fact that increasing, bounded sequences have unique limits.) Therefore, it follows from (18.1), the definition of $\mathbb{E}(X)$, that $a \leq \mathbb{E}(X)$.

We will now show $a \geq \mathbb{E}(X)$. As a result of (18.1), we only need to show that if Y is a simple random variable with $0 \leq Y \leq X$, then $\mathbb{E}(Y) \leq a$. That is, by definition, $\mathbb{E}(X) = \sup\{\mathbb{E}(Y) : Y \text{ is simple with } 0 \leq Y \leq X\}$ and so if Y is an arbitrary simple random variable satisfying $0 \leq Y \leq X$ and $\mathbb{E}(Y) \leq a$, then the definition of supremum implies $\mathbb{E}(X) \leq a$. To this end, let Y be simple and write

$$Y = \sum_{k=1}^{m} a_k 1\{Y = a_k\}.$$

That is, take $A_k = \{\omega : Y(\omega) = a_k\}$. Let $0 < \epsilon \leq 1$ and define

$$Y_{n,\epsilon} = (1 - \epsilon) Y 1_{\{(1-\epsilon)Y \leq X_n\}}.$$

Note that $Y_{n,\epsilon} = (1 - \epsilon)a_k$ on the set

$$\{(1 - \epsilon)Y \leq X_n\} \cap A_k = A_{k,n,\epsilon}$$

and that $Y_{n,\epsilon} = 0$ on the set $\{(1 - \epsilon)Y > X_n\}$. Clearly $Y_{n,\epsilon} \leq X_n$ and so

$$\mathbb{E}(Y_{n,\epsilon}) = (1 - \epsilon) \sum_{k=1}^{m} a_k \mathbf{P}\{A_{k,n,\epsilon}\} \leq \mathbb{E}(X_n).$$

We will now show that $A_{k,n,\epsilon}$ increases to A_k. Since $X_n \le X_{n+1}$ and $X_n \uparrow X$ we conclude that

$$\{(1-\epsilon)Y \le X_n\} \subseteq \{(1-\epsilon)Y \le X_{n+1}\} \subseteq \{(1-\epsilon)Y \le X\}$$

and therefore

$$A_{k,n,\epsilon} \subseteq A_{k,n+1,\epsilon} \subseteq \{(1-\epsilon)Y \le X\} \cap A_k. \tag{18.2}$$

Since $Y \le X$ by assumption, we know that the event $\{(1-\epsilon)Y \le X\}$ is equal to Ω. Thus, (18.2) implies $A_{k,n,\epsilon} \subseteq A_k$ for all n and so

$$\bigcup_{n=1}^{\infty} A_{k,n,\epsilon} \subseteq A_k. \tag{18.3}$$

Conversely, let $\omega \in A_k = \{(1-\epsilon)Y \le X\} \cap A_k = \{\omega : (1-\epsilon)Y(\omega) \le X(\omega)\} \cap A_k$ so that $(1-\epsilon)a_k \le X(\omega)$. Since $Y \le X$, which is to say that

$$Y(\omega) \le X(\omega) = \lim_{n\to\infty} X_n(\omega)$$

for all ω, we know that if $\omega \in A_k$, then there is some N such that $Y(\omega) = a_k \le X_n(\omega)$ whenever $n > N$. We therefore conclude that if $\omega \in A_k$ and $n > N$, then $(1-\epsilon)a_k < X_n(\omega)$. (This requires that X is not identically 0.) That is, $\omega \in A_{k,n,\epsilon}$ which proves that

$$\bigcup_{n=1}^{\infty} A_{k,n,\epsilon} \supseteq A_k. \tag{18.4}$$

In other words, it follows from (18.3) and (18.4) that $A_{k,n,\epsilon}$ increases to A_k, so by the continuity of probability theorem (Theorem 10.2), we conclude that

$$\lim_{n\to\infty} \mathbf{P}\{A_{k,n,\epsilon}\} = \mathbf{P}\{A_k\}.$$

Hence,

$$\mathbb{E}(Y_{n,\epsilon}) = (1-\epsilon)\sum_{k=1}^{m} a_k \mathbf{P}\{A_{k,n,\epsilon}\} \to (1-\epsilon)\sum_{k=1}^{m} a_k \mathbf{P}\{A_k\} = (1-\epsilon)\mathbb{E}(Y) \le a.$$

That is,

$$E(Y_{n,\epsilon}) \le \mathbb{E}(X_n)$$

and

$$E(Y_{n,\epsilon}) \uparrow (1-\epsilon)\mathbb{E}(Y) \quad \text{and} \quad \mathbb{E}(X_n) \uparrow a$$

so that

$$(1-\epsilon)\mathbb{E}(Y) \le a$$

(which follows since everything is increasing). Since $0 < \epsilon \le 1$ is arbitrary,

$$\mathbb{E}(Y) \le a$$

which, as noted earlier in the proof, is sufficient to conclude that

$$\mathbb{E}(X) \le a.$$

Combined with our earlier result that $a \le \mathbb{E}(X)$ we conclude $\mathbb{E}(X) = a$ and the proof is complete. □

Step 3: General Random Variables

Now suppose that X is any random variable. Write

$$X^+ = \max\{X, 0\} \quad \text{and} \quad X^- = -\min\{X, 0\}$$

for the positive part and the negative part of X, respectively. Note that $X^+ \geq 0$ and $X^- \geq 0$ so that the positive part and negative part of X are both positive random variables and $X = X^+ - X^-$.

Definition. A random variable X is called *integrable* (or has *finite expectation*) if both $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ are finite. In this case we define $\mathbb{E}(X)$ to be

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

Definition. If one of $\mathbb{E}(X^+)$ or $\mathbb{E}(X^-)$ is infinite, then we can still define $\mathbb{E}(X)$ by setting

$$\mathbb{E}(X) = \begin{cases} +\infty, & \text{if } \mathbb{E}(X^+) = +\infty \text{ and } \mathbb{E}(X^-) < \infty, \\ -\infty, & \text{if } \mathbb{E}(X^+) < \infty \text{ and } \mathbb{E}(X^-) = +\infty. \end{cases}$$

However, X is not integrable in this case.

Definition. If both $\mathbb{E}(X^+) = +\infty$ and $\mathbb{E}(X^-) = +\infty$, then $\mathbb{E}(X)$ does not exist.

Remark. We see that the standard machine is really not that hard to implement. In fact, it is usually enough to prove a result for simple random variables and then extend that result to positive random variables using Propositions 18.1 and 18.2. The result for general random variables usually then follows by definition.

Statistics 851 (Fall 2013) October 21, 2013
Prof. Michael Kozdron

Lecture #19: Expectation and Integration

Definition. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. If X is a simple random variable, say

$$X(\omega) = \sum_{i=1}^{n} a_i 1_{A_i}(\omega)$$

for $a_1, \ldots, a_n \in \mathbb{R}$ and $A_1, \ldots, A_n \in \mathcal{F}$, then we define

$$\mathbb{E}(X) = \sum_{i=1}^{n} a_i \mathbf{P}\{A_i\}.$$

If X is a positive random variable, then we define

$$\mathbb{E}(X) = \sup\{\mathbb{E}(Y) : Y \text{ is simple and } 0 \le Y \le X\}.$$

If X is any random variable, then we can write $X = X^+ - X^-$ where both X^+ and X^- are positive random variables. Provided that both $\mathbb{E}(X^+)$ and $\mathbb{E}(X^-)$ are finite, we define $\mathbb{E}(X)$ to be

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

and we say that X is *integrable* (or has *finite expectation*).

Remark. If $X : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B})$ is a random variable, then we sometimes write

$$\mathbb{E}(X) = \int_\Omega X \, \mathrm{d}\mathbf{P} = \int_\Omega X(\omega) \, \mathrm{d}\mathbf{P}\{\omega\} = \int_\Omega X(\omega) \mathbf{P}\{\mathrm{d}\omega\}.$$

That is, the expectation of a random variable is the *Lebesgue integral* of X. We will say more about this later.

Definition. Let $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ be the set of real-valued random variables on $(\Omega, \mathcal{F}, \mathbf{P})$ with finite expectation. That is,

$$\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P}) = \{X : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B}) : \; X \text{ is a random variable with } \mathbb{E}(X) < \infty\}.$$

We will often write $\mathcal{L}^1$ for $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ and suppress the dependence on the underlying probability space.

Theorem 19.1. *Suppose that $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, and let $X_1, X_2, \ldots$, X, and Y all be real-valued random variables on $(\Omega, \mathcal{F}, \mathbf{P})$.*

(a) $\mathcal{L}^1$ is a vector space and expectation is a linear map on $\mathcal{L}^1$. Furthermore, expectation is positive. That is, if X, $Y \in \mathcal{L}^1$ with $0 \le X \le Y$, then $0 \le \mathbb{E}(X) \le \mathbb{E}(Y)$.

(b) *$X \in \mathcal{L}^1$ if and only if $|X| \in \mathcal{L}^1$, in which case we have*

$$|\mathbb{E}(X)| \le \mathbb{E}(|X|).$$

(c) *If $X = Y$ almost surely (i.e., if $\mathbf{P}\{\omega : X(\omega) = Y(\omega)\} = 1$), then $\mathbb{E}(X) = \mathbb{E}(Y)$.*

(d) (Monotone Convergence Theorem) *If the random variables $X_n \ge 0$ for all n and $X_n \uparrow X$ (i.e., $X_n \to X$ and $X_n \le X_{n+1}$), then*

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}(X).$$

(We allow $\mathbb{E}(X) = +\infty$ if necessary.)

(e) (Fatou's Lemma) *If the random variables X_n all satisfy $X_n \ge Y$ almost surely for some $Y \in \mathcal{L}^1$ and for all n, then*

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \le \liminf_{n\to\infty} \mathbb{E}(X_n). \qquad (19.1)$$

In particular, (19.1) holds if $X_n \ge 0$ for all n.

(f) (Lebesgue's Dominated Convergence Theorem) *If the random variables $X_n \to X$, and if for some $Y \in \mathcal{L}^1$ we have $|X_n| \le Y$ almost surely for all n, then $X_n \in \mathcal{L}^1$, $X \in \mathcal{L}^1$, and*

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

Remark. This theorem contains **ALL** of the central results of Lebesgue integration theory.

Theorem 19.2. *Let X_n be a sequence of random variables.*

(a) *If $X_n \ge 0$ for all n, then*

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n) \qquad (19.2)$$

with both sides simultaneously being either finite or infinite.

(b) *If*

$$\sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty,$$

then

$$\sum_{n=1}^{\infty} X_n$$

converges almost surely to some random variable $Y \in \mathcal{L}^1$. In other words,

$$\sum_{n=1}^{\infty} X_n$$

is integrable with

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n).$$

Thus, (19.2) holds with both sides being finite.

Notation. For $1 \le p < \infty$, let $\mathcal{L}^p = \{\text{random variables } X : \Omega \to \mathbb{R} \text{ such that } |X|^p \in \mathcal{L}^1\}$.

Theorem 19.3 (Cauchy-Schwartz Inequality)**.** *If X, $Y \in \mathcal{L}^2$, then $XY \in \mathcal{L}^1$ and*

$$|\mathbb{E}(XY)| \le \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Theorem 19.4. *Let $X : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B})$ be a random variable.*

(a) (Markov's Inequality) *If $X \in \mathcal{L}^1$, then*

$$\mathbf{P}\left\{|X| \ge a\right\} \le \frac{\mathbb{E}(|X|)}{a}$$

for every $a > 0$.

(b) (Chebychev's Inequality) *If $X \in \mathcal{L}^2$, then*

$$\mathbf{P}\left\{|X| \ge a\right\} \le \frac{\mathbb{E}(X^2)}{a^2}$$

for every $a > 0$.

Lecture #20: Proofs of the Main Expectation Theorems

Our goal for today is to start proving all of the important results for expectation that were stated last lecture. Note that the proofs generally follow the so-called standard machine; that is, we first prove the result for simple random variables, then extend it to non-negative random variables, and finally extend it to general random variables. The key results for implementing this strategy are Proposition 18.1 and Proposition 18.2.

Theorem 20.1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. If $\mathcal{L}^1$ denotes the space of integrable random variables, namely*

$$\mathcal{L}^1 = \{\textit{random variables } X \textit{ such that } \mathbb{E}(X) < \infty\},$$

then $\mathcal{L}^1$ is a vector space and expectation is a linear operator on $\mathcal{L}^1$. Moreover, expectation is monotone in the sense that if X and Y are random variables with $0 \le X \le Y$ and $Y \in \mathcal{L}^1$, then $X \in \mathcal{L}^1$ and $0 \le \mathbb{E}(X) \le \mathbb{E}(Y)$.

Proof. Suppose that $X \ge 0$ and $Y \ge 0$ are non-negative random variables, and let $\alpha \in [0, \infty)$. We know that there exist sequences X_n and Y_n of non-negative simple random variables such that (i) $X_n \uparrow X$ and $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$, and (ii) $Y_n \uparrow Y$ and $\mathbb{E}(Y_n) \uparrow \mathbb{E}(Y)$. Therefore, αX_n is a sequence of non-negative simple random variables with $(\alpha X_n) \uparrow (\alpha X)$ and $\mathbb{E}(\alpha X_n) \uparrow \mathbb{E}(\alpha X_n)$. Moreover, $X_n + Y_n$ is a sequence of non-negative simple random variables with $(X_n + Y_n) \uparrow (X + Y)$ and $\mathbb{E}(X_n + Y_n) \uparrow \mathbb{E}(X + Y)$. However, we know that expectation is linear on simple random variables so that

$$\mathbb{E}(\alpha X_n) = \alpha \mathbb{E}(X_n) \quad \text{and} \quad \mathbb{E}(X_n + Y_n) = \mathbb{E}(X_n) + \mathbb{E}(Y_n).$$

Thus, we find

$$\begin{array}{ccccccc} \mathbb{E}(\alpha X_n) & = & \alpha\mathbb{E}(X_n) & & \mathbb{E}(X_n+Y_n) & = & \mathbb{E}(X_n)+\mathbb{E}(Y_n) \\ \downarrow & & \downarrow & \text{and} & \downarrow & & \downarrow \\ \mathbb{E}(\alpha X) & & \alpha\mathbb{E}(X) & & \mathbb{E}(X+Y) & & \mathbb{E}(X)+\mathbb{E}(Y) \end{array}$$

so by uniqueness of limits, we conclude $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$ and $\mathbb{E}(X + Y) = \mathbb{E}(X + Y)$. Also note that if $0 \le X \le Y$, then the definition of expectation of non-negative random variables immediately implies that $0 \le \mathbb{E}(X) \le \mathbb{E}(Y)$ so that $Y \in \mathcal{L}^1$ implies $X \in \mathcal{L}^1$.

Suppose now that X and Y are general random variables and let $\alpha \in \mathbb{R}$. Since

$$(\alpha X) = (\alpha X)^+ - (\alpha X)^- \le (\alpha X)^+ + (\alpha X)^- \le |\alpha|(X^+ + X^-)$$

and

$$(X + Y) = (X + Y)^+ - (X + Y)^- \le (X + Y)^+ + (X + Y)^- \le X^+ + X^- + Y^+ + Y^-.$$

Hence, since X^+, X^-, Y^+, $Y^- \in \mathcal{L}^1$, we conclude that $\alpha X \in \mathcal{L}^1$ and $X + Y \in \mathcal{L}^1$. Finally, since $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$ by definition, and since we showed above that expectation is linear on non-negative random variables, we conclude that expectation is linear on general random variables. □

Theorem 20.2. *If $X : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B})$ is a random variable, then $X \in \mathcal{L}^1$ if and only if $|X| \in \mathcal{L}^1$.*

Proof. Suppose that $X \in \mathcal{L}^1$ so that $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$. Since $|X| = X^+ + X^-$ and since expectation is linear, we conclude that

$$\mathbb{E}(|X|) = \mathbb{E}(X^+ + X^-) = \mathbb{E}(X^+) + \mathbb{E}(X^-) < \infty$$

so that $|X| \in \mathcal{L}^1$.

On the other hand, suppose that $|X| \in \mathcal{L}^1$ so that $\mathbb{E}(X^+) + \mathbb{E}(X^-) < \infty$. However, since $X^+ \geq 0$ and $X^- \geq 0$, we know that $\mathbb{E}(X^+) \geq 0$ and $\mathbb{E}(X^-) \geq 0$. We now use the fact that if the sum of two non-negative numbers is finite, then each number must be finite to conclude that $\mathbb{E}(X^+) < \infty$ and $\mathbb{E}(X^-) < \infty$. Thus, by definition, $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-) < \infty$ so that $X \in \mathcal{L}^1$. □

Corollary. *If $X \in \mathcal{L}^1$, then $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$. In particular, if $\mathbb{E}(|X|) = 0$, then $\mathbb{E}(X) = 0$.*

Proof. Since $\mathbb{E}(X^+) \geq 0$ and $\mathbb{E}(X^-) \geq 0$, we have from the triangle inequality

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

Since $|\mathbb{E}(X)| \geq 0$, if $\mathbb{E}(|X|) = 0$, then $0 \leq |\mathbb{E}(X)| \leq \mathbb{E}(|X|) = 0$ implying $\mathbb{E}(X) = 0$. □

Theorem 20.3. *If $X, Y : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B})$ are integrable random variables with $X = Y$ almost surely, then $\mathbb{E}(X) = \mathbb{E}(Y)$.*

Proof. Suppose that $X = Y$ almost surely so that $\mathbf{P}\{\omega \in \Omega : X(\omega) = Y(\omega)\} = 1$. To begin, assume that $X \geq 0$ and $Y \geq 0$ and let $A = \{\omega : X(\omega) \neq Y(\omega)\}$ so that $\mathbf{P}\{A\} = 0$. Write

$$\mathbb{E}(Y) = \mathbb{E}(Y1_A + Y1_{A^c}) = \mathbb{E}(Y1_A) + \mathbb{E}(Y1_{A^c}) = \mathbb{E}(Y1_A) + \mathbb{E}(X1_{A^c}). \qquad (*)$$

We know that there exist sequences X_n and Y_n of non-negative simple random variables such that (i) $X_n \uparrow X$ and $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$, and (ii) $Y_n \uparrow Y$ and $\mathbb{E}(Y_n) \uparrow \mathbb{E}(Y)$. Thus, $X_n 1_A \uparrow X1_A$ and $\mathbb{E}(X_n 1_A) \uparrow \mathbb{E}(X1_A)$ and similarly $Y_n 1_A \uparrow Y1_A$ and $\mathbb{E}(Y_n 1_A) \uparrow \mathbb{E}(Y1_A)$. For each n, the random variable X_n takes on finitely many values and is therefore bounded by K, say, where K may depend on n. Thus, since $X_n \leq K$, we obtain $X_n 1_A \leq K$ and so

$$0 \leq \mathbb{E}(X_n 1_A) \leq \mathbb{E}(K1_A) = K\mathbf{P}\{A\} = 0.$$

This implies that $\mathbb{E}(X_n 1_A) = 0$ and so by uniqueness of limits, $\mathbb{E}(X1_A) = 0$. Similarly, $\mathbb{E}(Y1_A) = 0$. Therefore, by $(*)$, we obtain

$$\mathbb{E}(Y) = \mathbb{E}(Y1_A) + \mathbb{E}(X1_{A^c}) = 0 + \mathbb{E}(X1_{A^c}) = \mathbb{E}(X1_A) + \mathbb{E}(X1_{A^c}) = \mathbb{E}(X).$$

In general, note that $X = Y$ almost surely implies that $X^+ = Y^+$ almost surely and $X^- = Y^-$ almost surely. □

Lecture #21: Proofs of the Main Expectation Theorems (continued)

We will continue proving the important results for expectation that were stated in Lecture #19. Recall that a random variable is said to have finite mean or have finite expectation or be integrable if $\mathbb{E}(X) < \infty$. The vector space of all integrable random variable on a given probability space $(\Omega, \mathcal{F}, P)$ is denoted by $\mathcal{L}^1$ and if $1 \le p < \infty$, then

$$\mathcal{L}^p = \{\text{random variables } X : \Omega \to \mathbb{R} \text{ such that } |X|^p \in \mathcal{L}^1\}.$$

Theorem 21.1 (Cauchy-Schwartz Inequality)**.** *If X, $Y \in \mathcal{L}^2$, then $XY \in \mathcal{L}^1$ and*

$$|\mathbb{E}(XY)| \le \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

Proof. Since $0 \le (X+Y)^2 = X^2 + Y^2 + 2XY$ and $0 \le (X-Y)^2 = X^2 + Y^2 - 2XY$, we conclude that $2|XY| \le X^2 + Y^2$ implying $2\mathbb{E}(|XY|) \le \mathbb{E}(X^2) + \mathbb{E}(Y^2)$. Thus, if $X, Y \in \mathcal{L}^2$, we conclude that $XY \in \mathcal{L}^1$. For every $x \in \mathbb{R}$, note that

$$0 \le \mathbb{E}((xX+Y)^2) = x^2\mathbb{E}(X^2) + 2x\mathbb{E}(XY) + \mathbb{E}(Y^2).$$

Since $x^2\mathbb{E}(X^2) + 2x\mathbb{E}(XY) + \mathbb{E}(Y^2)$ is a non-negative quadratic in x, its discriminant is necessarily non-positive; that is,

$$4[\mathbb{E}(XY)]^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2) \le 0,$$

or, equivalently,

$$|\mathbb{E}(XY)| \le \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

as required. □

Theorem 21.2. *Let $X : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B})$ be a random variable.*

(a) (Markov's Inequality) *If $X \in \mathcal{L}^1$, then*

$$\mathbf{P}\{|X| \ge a\} \le \frac{\mathbb{E}(|X|)}{a}$$

for every $a > 0$.

(b) (Chebychev's Inequality) *If $X \in \mathcal{L}^2$, then*

$$\mathbf{P}\{|X| \ge a\} \le \frac{\mathbb{E}(X^2)}{a^2}$$

for every $a > 0$.

Proof. Recall that $X \in \mathcal{L}^1$ if and only if $|X| \in \mathcal{L}^1$. Since $|X| \geq 0$, we can write

$$|X| \geq a 1_{\{|X| \geq a\}}.$$

Taking expectations of the previous expression implies

$$\mathbb{E}(|X|) \geq a\mathbb{E}(1_{\{|X|\geq a\}}) = \mathbf{P}\left\{|X| \geq a\right\}$$

and Markov's inequality follows. Similarly, $X^2 \geq a^2 1_{\{X^2 \geq a^2\}}$ so that if $X \in \mathcal{L}^2$, taking expectations implies

$$\mathbb{E}(X^2) \geq a^2 \mathbf{P}\left\{X^2 \geq a^2\right\}.$$

Hence,

$$\mathbf{P}\left\{|X| \geq a\right\} \leq \mathbf{P}\left\{X^2 \geq a\right\} \leq \frac{\mathbb{E}(X^2)}{a^2}$$

yielding Chebychev's inequality. □

Definition. If $X \in \mathcal{L}^2$ we say that X has finite variance and define the *variance* of X to be

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

Thus, Chebychev's inequality sometimes takes the form

$$\mathbf{P}\left\{|X - \mathbb{E}(X)| \geq a\right\} \leq \frac{\text{Var}(X)}{a^2}.$$

Computing Expectations

Having proved a number of the main expectation theorems, we will now take a small detour to discuss computing expectations. We will also show over the course of the next several lectures how the formulas for the expectations of discrete and continuous random variables encountered in introductory probability follow from the general theory developed. The first examples we will discuss, however, are the calculations of expectations directly from the definition and theory.

Example 21.3. Consider $([0,1], \mathcal{B}_1, \mathbf{P})$ where $\mathcal{B}_1$ are the Borel sets of $[0,1]$ and $\mathbf{P}$ is the uniform probability. Let $X : \Omega \to \mathbb{R}$ be given by $X(\omega) = \omega$ so that X is a uniform random variable. We will now compute $\mathbb{E}(X)$ directly by definition. Note that X is not simple since the range of X, namely $[0,1]$, is uncountable. However, X is positive. This means that as a consequence of Propositions 18.1 and 18.2, if X_n is a sequence of positive, simple random variables with $X_n \uparrow X$, then $\mathbb{E}(X_n) \uparrow \mathbb{E}(X)$. Thus, let

$$X_1(\omega) = \begin{cases} 0, & 0 \leq \omega < 1/2, \\ 1/2, & 1/2 \leq \omega \leq 1 \end{cases}$$

$$X_2(\omega) = \begin{cases} 0, & 0 \leq \omega \leq 1/4, \\ 1/4, & 1/4 \leq \omega < 1/2, \\ 1/2, & 1/2 \leq \omega < 3/4, \\ 3/4, & 3/4 \leq \omega \leq 1, \end{cases}$$

and in general,

$$X_n(\omega) = \sum_{j=1}^{2^n-2} \frac{j-1}{2^n} 1\left\{\frac{j-1}{2^n} \le \omega < \frac{j}{2^n}\right\} + \frac{2^n-1}{2^n} 1\left\{\frac{2^n-2}{2^n} \le \omega \le \frac{2^n-1}{2^n}\right\}.$$

Thus,

$$\begin{aligned}
\mathbb{E}(X_n) &= \sum_{j=1}^{2^n-2} \frac{j-1}{2^n} \mathbf{P}\left\{\left[\frac{j-1}{2^n}, \frac{j}{2^n}\right)\right\} + \frac{2^n-1}{2^n} \mathbf{P}\left\{\left[\frac{2^n-2}{2^n}, \frac{2^n-1}{2^n}\right]\right\} \\
&= \sum_{j=1}^{2^n-1} \frac{j-1}{2^n}\left(\frac{j}{2^n} - \frac{j-1}{2^n}\right) \\
&= \frac{1}{2^{2n}} \sum_{j=1}^{2^n-1} (j-1) \\
&= \frac{1}{2^{2n}} \frac{(2^n-2)(2^n-1)}{2} \\
&= \frac{1}{2} - \frac{1}{2^n} - \frac{1}{2^{n+1}} + \frac{1}{2^{2n}}
\end{aligned}$$

implying

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty}\left(\frac{1}{2} - \frac{1}{2^n} - \frac{1}{2^{n+1}} + \frac{1}{2^{2n}}\right) = \frac{1}{2}.$$

Example 21.4. Consider $([0,1], \mathcal{B}_1, \mathbf{P})$ where $\mathcal{B}_1$ are the Borel sets of $[0,1]$ and $\mathbf{P}$ is the uniform probability. Let $\mathbb{Q}_1 = [0,1] \cap \mathbb{Q}$ and consider the random variable $Y : \Omega \to \mathbb{R}$ given by

$$Y(\omega) = \begin{cases} \omega, & \text{if } \omega \in [0,1] \setminus \mathbb{Q}_1, \\ 0, & \text{if } \omega \in \mathbb{Q}_1. \end{cases}$$

Note that Y is not a simple random variable, although it is non-negative. In order to compute $\mathbb{E}(Y)$ it is easiest to use Theorem 20.3. That is, let $X(\omega) = \omega$ for $\omega \in [0,1]$ and note that

$$\{\omega : X(\omega) \neq Y(\omega)\} = \mathbb{Q}_1 \setminus \{0\}.$$

(The technical point here is that $X(0) = Y(0) = 0$. However, if $\omega \in \mathbb{Q}_1$ with $\omega \neq 0$, then $X(\omega) \neq Y(\omega)$.) Since $\mathbf{P}\{\mathbb{Q}_1 \setminus \{0\}\} = 0$, we see that X and Y differ on a set of probability 0 so that $X = Y$ almost surely. Since X is a uniform random variable on $[0,1]$, we know from the previous example that $\mathbb{E}(X) = 1/2$. Therefore, using Theorem 20.3 we conclude $\mathbb{E}(Y) = \mathbb{E}(X) = 1/2$.

Statistics 851 (Fall 2013) October 28, 2013
Prof. Michael Kozdron

Lecture #22: Computing Expectations of Discrete Random Variables

Recall from introductory probability classes that a random variable X is called *discrete* if the range of X is at most countable. The formula for the expectation of a discrete random variable X given in these classes is

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} j\mathbf{P}\{X = j\}.$$

We will now derive this formula as a consequence of the general theory developed during the past several lectures.

Suppose that the range of X is at most countable. Without loss of generality, we can assume that $\Omega = \{1, 2, \ldots, \}$, $\mathcal{F} = 2^{\Omega}$, and $X : \Omega \to \mathbb{R}$ is given by $X(\omega) = \omega$. Assume further that the law of X is given by $\mathbf{P}\{X = j\} = p_j$ where $p_j \in [0, 1]$ and

$$\sum_{j=1}^{\infty} p_j = 1.$$

Let $A_j = \{X = j\} = \{\omega : X(\omega) = j\}$ so that X can be written as

$$X(\omega) = \sum_{j=1}^{\infty} j 1_{A_j}.$$

Observe that if $p_j = 0$ for infinitely many j, then X is simple and can be written as

$$X(\omega) = \sum_{j=1}^{j} 1_{A_j}$$

for some $n < \infty$ which implies that

$$\mathbb{E}(X) = \sum_{j=1}^{j} \mathbf{P}\{A_j\} = \sum_{j=1}^{n} j\mathbf{P}\{X = j\}.$$

On the other hand, suppose that $p_j \neq 0$ for infinitely many j. In this case, X is not simple. We can approximate X by simple functions as follows. Let

$$A_{j,n} = \{\omega : X(\omega) = j,\, j \leq n\}$$

so that $A_{j,n} \subseteq A_{j,n+1}$ and

$$\bigcup_{n=1}^{\infty} A_{j,n} = A_j.$$

That is, $A_{j,n} \uparrow A_j$ and so by continuity of probability we conclude

$$\lim_{n\to\infty} \mathbf{P}\{A_{j,n}\} = \mathbf{P}\{A_j\}.$$

If we now set

$$X_n(\omega) = \sum_{j=1}^{n} X(\omega) 1_{A_{j,n}} = \sum_{j=1}^{n} j 1_{A_{j,n}}$$

so that

$$\mathbb{E}(X_n) = \sum_{j=1}^{n} j\mathbf{P}\{A_{j,n}\},$$

then (X_n) is a sequence of positive simple random variable with $X_n \uparrow X$. Proposition 18.2 then implies

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} \sum_{j=1}^{n} j\mathbf{P}\{A_{j,n}\} = \sum_{j=1}^{\infty} \mathbf{P}\{A_j\} = \sum_{j=1}^{\infty} j\mathbf{P}\{X = j\}$$

as required.

Computing Expectations of Continuous Random Variables

We will now turn to that other formula for computing expectation encountered in elementary probability courses, namely if X is a continuous random variable with density f_X, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, \mathrm{d}x.$$

It turns out that verifying this formula is somewhat more involved than the discrete formula. As such, we need to take a brief detour into some general function theory.

Some General Function Theory

Suppose that $f : \mathbf{X} \to \mathbf{Y}$ is a function. We are implicitly assuming that f is defined for all $x \in \mathbf{X}$. We call $\mathbf{X}$ the *domain* of f and call $\mathbf{Y}$ the *codomain* of f.

The *range* of f is the set

$$f(\mathbf{X}) = \{y \in \mathbf{Y} : f(x) = y \text{ for some } x \in \mathbf{X}\}.$$

Note that $f(\mathbf{X}) \subseteq \mathbf{Y}$. If $f(\mathbf{X}) = \mathbf{Y}$, then we say that f is *onto* $\mathbf{Y}$.

Let $B \subseteq \mathbf{Y}$. We define $f^{-1}(B)$ by

$$f^{-1}(B) = \{x \in \mathbf{X} : f(x) = y \text{ for some } y \in B\} = \{f \in B\} = \{x : f(x) \in B\}.$$

We call $\mathbf{X}$ a *topological space* if there is a notion of open subsets of $\mathbf{X}$. The Borel σ-algebra on $\mathbf{X}$, written $\mathcal{B}(\mathbf{X})$, is the σ-algebra generated by the open sets of $\mathbf{X}$.

Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces. A function $f : \mathbf{X} \to \mathbf{Y}$ is called *continuous* if for every open set $V \subseteq \mathbf{Y}$, the set $U = f^{-1}(V) \subseteq \mathbf{X}$ is open.

A function $f : \mathbf{X} \to \mathbf{Y}$ is called measurable if $f^{-1}(B) \in \mathcal{B}(\mathbf{X})$ for every $B \in \mathcal{B}(\mathbf{Y})$.

Since the open sets generate the Borel sets, this theorem follows easily.

Theorem 22.1. *Suppose that $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ and $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$ are topological measure spaces. The function $f : \mathbf{X} \to \mathbf{Y}$ is measurable if and only if $f^{-1}(O) \in \mathcal{B}(\mathbf{X})$ for every open set $O \in \mathcal{B}(\mathbf{Y})$.*

The next theorem tells us that continuous functions are necessarily measurable functions.

Theorem 22.2. *Suppose that $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ and $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}))$ are topological measure spaces. If $f : \mathbf{X} \to \mathbf{Y}$ is continuous, then f is measurable.*

Proof. By definition, $f : \mathbf{X} \to \mathbf{Y}$ is continuous if and only if $f^{-1}(O) \subseteq \mathbf{X}$ is an open set for every open set $O \subseteq \mathbf{Y}$. Since an open set is necessarily a Borel set, we conclude that $f^{-1}(O) \in \mathcal{B}(\mathbf{X})$ for every open set $O \in \mathcal{B}(\mathbf{Y})$. However, it now follows immediately from the previous theorem that $f : \mathbf{X} \to \mathbf{Y}$ is measurable. □

The following theorem shows that the composition of measurable functions is measurable.

Theorem 22.3. *Suppose that $(\mathbf{W}, \mathcal{F})$, $(\mathbf{X}, \mathcal{G})$, and $(\mathbf{Y}, \mathcal{H})$ are measurable spaces, and let $f : (\mathbf{W}, \mathcal{F}) \to (\mathbf{X}, \mathcal{G})$ and $g : (\mathbf{X}, \mathcal{G}) \to (\mathbf{Y}, \mathcal{H})$ be measurable. Then the function $h = g \circ f$ is a measurable function from $(\mathbf{W}, \mathcal{F})$ to $(\mathbf{Y}, \mathcal{H})$.*

Proof. Suppose that $H \in \mathcal{H}$. Since g is measurable, we have $g^{-1}(H) \in \mathcal{G}$. Since f is measurable, we have $f^{-1}(g^{-1}(H)) \in \mathcal{F}$. Since

$$h^{-1}(H) = (g \circ f)^{-1}(H) = f^{-1}(g^{-1}(H)) \in \mathcal{F}$$

the proof is complete. □

Lecture #23: Proofs of the Main Expectation Theorems (continued)

Theorem 23.1 (Monotone Convergence Theorem). *Suppose that $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, and let $X_1, X_2, \ldots,$ and X be real-valued random variables on $(\Omega, \mathcal{F}, \mathbf{P})$. If $X_n \ge 0$ for all n and $X_n \uparrow X$ (i.e., $X_n \to X$ and $X_n \le X_{n+1}$), then*

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}(X)$$

(allowing $\mathbb{E}(X) = +\infty$ if necessary).

Proof. For every n, let $Y_{n,k}$, $k = 1, 2, \ldots,$ be non-negative and simple with $Y_{n,k} \uparrow X_n$ and $\mathbb{E}(Y_{n,k}) \uparrow \mathbb{E}(X_n)$ as $k \to \infty$ which is possible by Propositions 18.1 and 18.2. Set

$$Z_k = \max_{n\le k} Y_{n,k},$$

and observe that $0 \le Z_k \le Z_{k+1}$ so that (Z_k) is an increasing sequence of non-negative simple random variables which necessarily has a limit

$$Z = \lim_{k\to\infty} Z_k.$$

By Proposition 18.2, $\mathbb{E}(Z_k) \uparrow \mathbb{E}(Z)$. We now observe that if $n \le k$, then

$$Y_{n,k} \le Z_k \le X_n \le X_k \le X. \tag{23.1}$$

We now deduce from (23.1) that $X_n \le Z \le X$ almost surely, and so letting $n \to \infty$ implies $X = Z$ almost surely. We also deduce from (23.1) that

$$\mathbb{E}(Y_{n,k}) \le \mathbb{E}(Z_k) \le \mathbb{E}(X_k)$$

for $n \le k$. Fix n and let $k \to \infty$ to obtain

$$\mathbb{E}(X_n) \le \mathbb{E}(Z) \le \lim_{k\to\infty} \mathbb{E}(X_k).$$

Now let $n \to \infty$ to obtain

$$\lim_{n\to\infty} \mathbb{E}(X_n) \le \mathbb{E}(Z) \le \lim_{k\to\infty} \mathbb{E}(X_k).$$

Thus,

$$\mathbb{E}(Z) = \lim_{n\to\infty} \mathbb{E}(X_n).$$

But $X = Z$ almost surely so that $\mathbb{E}(X) = \mathbb{E}(Z)$ and we conclude

$$\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n)$$

as required. □

Theorem 23.2 (Fatou's Lemma)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbf{P})$ *is a probability space, and let* $X_1, X_2, \ldots$ *be real-valued random variables on* $(\Omega, \mathcal{F}, \mathbf{P})$*. If* $X_n \geq Y$ *almost surely for some* $Y \in \mathcal{L}^1$ *and for all* n*, then*

$$\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n). \tag{23.2}$$

In particular, (23.2) holds if $X_n \geq 0$ *for all* n*.*

Proof. Without loss of generality, we can assume $X_n \geq 0$. Here is the reason. If $\tilde{X}_n = X_n - Y$, then $\tilde{X}_n \in \mathcal{L}^1$ with $\tilde{X}_n \geq 0$ and

$$\mathbb{E}\left(\liminf_{n\to\infty} \tilde{X}_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(\tilde{X}_n) \text{ if and only if } \mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$$

because

$$\liminf_{n\to\infty} \tilde{X}_n = \left(\liminf_{n\to\infty} X_n\right) - Y.$$

Hence, if $X_n \geq 0$, set

$$Y_n = \inf_{k\geq n} X_k$$

so that $Y_n \geq 0$ is a random variable and $Y_n \leq Y_{n+1}$. This implies that Y_n converges and

$$\lim_{n\to\infty} Y_n = \liminf_{n\to\infty} X_n.$$

Since $X_n \geq Y_n$, monotonicity of expectation implies $\mathbb{E}(X_n) \geq \mathbb{E}(Y_n)$. This yields

$$\liminf_{n\to\infty} \mathbb{E}(X_n) \leq \lim_{n\to\infty} \mathbb{E}(Y_n) = \mathbb{E}\left(\lim_{n\to\infty} Y_n\right) = \liminf_{n\to\infty} \mathbb{E}(X_n)$$

by the Monotone Convergence Theorem and the proof is complete. □

Theorem 23.3 (Lebesgue's Dominated Convergence Theorem)**.** *Suppose that* $(\Omega, \mathcal{F}, \mathbf{P})$ *is a probability space, and let* $X_1, X_2, \ldots$ *be real-valued random variables on* $(\Omega, \mathcal{F}, \mathbf{P})$*. If the random variables* $X_n \to X$*, and if for some* $Y \in \mathcal{L}^1$ *we have* $|X_n| \leq Y$ *almost surely for all* n*, then* $X_n \in \mathcal{L}^1$*,* $X \in \mathcal{L}^1$*, and*

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X).$$

Proof. Suppose that we define the random variables U and V by

$$U = \liminf_{n\to\infty} X_n \quad \text{and} \quad V = \limsup_{n\to\infty} X_n.$$

The assumption that $X_n \to X$ implies that $U = V = X$ almost surely so that $\mathbb{E}(U) = \mathbb{E}(V) = \mathbb{E}(X)$. And if we also assume that $|X_n| \leq Y$ almost surely for all n so that $X_n \in \mathcal{L}^1$, then we conclude $|X| \leq Y$. Since $Y \in \mathcal{L}^1$, we have $X \in \mathcal{L}^1$. Moreover, $X_n \geq -Y$ almost surely and $-Y \in \mathcal{L}^1$ so by Fatou's Lemma,

$$\mathbb{E}(U) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

However, we also know $-X_n \geq -Y$ almost surely and

$$-V = \liminf_{n\to\infty}(-X_n)$$

so Fatou's lemma also implies

$$-\mathbb{E}(V) = \mathbb{E}(-V) \geq \liminf_{n\to\infty} \mathbb{E}(-X_n) = -\limsup_{n\to\infty} \mathbb{E}(X_n).$$

Combining these two inequalities gives

$$\mathbb{E}(X) = \mathbb{E}(U) \leq \liminf_{n\to\infty} \mathbb{E}(X_n) \leq \limsup_{n\to\infty} \mathbb{E}(X_n) \leq \mathbb{E}(V) = \mathbb{E}(X)$$

so that $\mathbb{E}(X_n) \to \mathbb{E}(X)$ as required. □

Theorem 23.4. *Let X_n be a sequence of random variables.*

(a) If $X_n \geq 0$ for all n, then

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n) \tag{23.3}$$

with both sides simultaneously being either finite or infinite.

(b) If

$$\sum_{n=1}^{\infty} \mathbb{E}(|X_n|) < \infty,$$

then

$$\sum_{n=1}^{\infty} X_n$$

converges almost surely to some random variable $Y \in \mathcal{L}^1$. In other words,

$$\sum_{n=1}^{\infty} X_n$$

is integrable with

$$\mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}(X_n).$$

Thus, (23.3) holds with both sides being finite.

Proof. Suppose that

$$S_n = \sum_{k=1}^{n} |X_k| \quad \text{and} \quad T_n = \sum_{k=1}^{n} X_k.$$

Since S_n contains finitely many terms, we know by linearity of expectation that

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{k=1}^{n} |X_k|\right) = \sum_{k=1}^{n} \mathbb{E}(|X_k|).$$

Moreover, $0 \leq S_n \leq S_{n+1}$ so that S_n increases to some limit

$$S = \sum_{k=1}^{\infty} |X_k|.$$

Note that $S(\omega) \in [0, +\infty]$. By the Monotone Convergence Theorem,

$$\mathbb{E}(S) = \lim_{n\to\infty} \mathbb{E}(S_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}(|X_k|) = \sum_{k=1}^{\infty} \mathbb{E}(|X_k|).$$

If $X_n \geq 0$ for all n, then $S_n = T_n$ and (a) follows. If X_n are general random variables, with

$$\sum_{k=1}^{\infty} \mathbb{E}(|X_k|) < \infty,$$

then $\mathbb{E}(S) < \infty$. Now, for every $\epsilon > 0$, since $1_{S=\infty} \leq \epsilon S$, we conclude $\mathbf{P}\{S = \infty\} = \mathbb{E}(1_{S=\infty}) \leq \epsilon \mathbb{E}(S)$. Since $\epsilon > 0$ is arbitrary and $\mathbb{E}(S) < \infty$, we conclude that $\mathbf{P}\{S = \infty\} = 0$. Therefore,

$$\sum_{n=1}^{\infty} X_n$$

is absolutely convergent almost surely and its sum is the limit of the sequence T_n. Moreover, $|T_n| \leq S_n \leq S$ and $S \in \mathcal{L}^1$, so by the Dominated Convergence Theorem,

$$\mathbb{E}\left(\sum_{k=1}^{\infty} X_k\right) = \mathbb{E}\left(\lim_{n\to\infty} \sum_{k=1}^{n} X_k\right) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}(X_k) = \sum_{k=1}^{\infty} \mathbb{E}(X_k)$$

proving (b). $\square$

Lecture #24: The Expectation Rule

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable so that $X^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$. Suppose that $h : \mathbb{R} \to \mathbb{R}$ is a measurable function so that $h^{-1}(B) \in \mathcal{B}$ for every $B \in \mathcal{B}$.

Proposition 24.1. *The function $h \circ X : \Omega \to \mathbb{R}$ given by $h \circ X(\omega) = h(X(\omega))$ is a random variable.*

Proof. To prove that $h \circ X$ is a random variable, we must show that $(h \circ X)^{-1}(B) = X^{-1}(h^{-1}(B)) \in \mathcal{F}$ for every $B \in \mathcal{B}$. Thus, suppose that $B \in \mathcal{B}$. Since h is measurable, we know $h^{-1}(B) \in \mathcal{B}$. Let $A = h^{-1}(B)$. Since A is a Borel set, i.e., $A \in \mathcal{B}$, we know that $X^{-1}(A) \in \mathcal{F}$ since X is a random variable. That is,

$$(h \circ X)^{-1}(B) = X^{-1}(h^{-1}(B)) = X^{-1}(A) \in \mathcal{F}$$

so that $h \circ X : \Omega \to \mathbb{R}$ is a random variable as required. $\square$

We know that X induces a probability on $(\mathbb{R}, \mathcal{B})$ called the *law* of X and defined by

$$\mathbf{P}^X\{B\} = \mathbf{P}\left\{X^{-1}(B)\right\} = \mathbf{P}\{X \in B\} = \mathbf{P}\{\omega \in \Omega : X(\omega) \in B\}.$$

In other words, $(\mathbb{R}, \mathcal{B}, \mathbf{P}^X)$ is a probability space. This means that if $h : \mathbb{R} \to \mathbb{R}$ is a measurable function on $(\mathbb{R}, \mathcal{B}, \mathbf{P}^X)$, then we can also call h a random variable. As such, it is possible to discuss the expectation of h, namely

$$\mathbb{E}(h) = \int_{\mathbb{R}} h(x)\, \mathbf{P}^X\{\mathrm{d}x\}.$$

As we will now show, the expectation of h and the expectation of $h \circ X$ are intimately related.

Theorem 24.2 (Expectation Rule)**.** *Let $X : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B})$ be a random variable with law $\mathbf{P}^X$, and let $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ be measurable.*

(a) The random variable $h(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$ if and only if $h \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}, \mathbf{P}^X)$.

(b) If $h \geq 0$, or if either condition in (a) holds, then

$$\mathbb{E}(h(X)) = \int_{\Omega} h(X(\omega))\, \mathbf{P}\{\mathrm{d}\omega\} = \int_{\mathbb{R}} h(x)\, \mathbf{P}^X\{\mathrm{d}x\}.$$

Remark. The second integral in the theorem is actually a Lebesgue integral since it is the integral of a random variable, namely h, with respect to a probability, namely $\mathbf{P}^X$. As we will see shortly, we can always relate this to a Riemann-Stieltjes integral and in some cases to a Riemann integral.

Proof. Since $h \circ X : \Omega \to \mathbb{R}$ is a random variable by Proposition 24.1, we know

$$\mathbb{E}(h(X)) = \int_\Omega h(X(\omega))\, \mathbf{P}\{\mathrm{d}\omega\}$$

by definition. The content of the theorem is that this integral is equal to

$$\int_\mathbb{R} h(x)\, \mathbf{P}^X\{\mathrm{d}x\}.$$

In order to complete the proof, we will follow the standard machine. That is, let $h(x) = 1_B(x)$ for $B \in \mathcal{B}$ so that

$$\begin{aligned}\mathbb{E}(h(X)) = \mathbb{E}(1_B(X)) = \mathbf{P}\{X \in B\} = \mathbf{P}\left\{X^{-1}(B)\right\} = \mathbf{P}^X\{B\} &= \int_B \mathbf{P}^X\{\mathrm{d}x\} \\ &= \int_\mathbb{R} 1_B(x)\, \mathbf{P}^X\{\mathrm{d}x\} \\ &= \int_R h(x)\, \mathbf{P}^X\{\mathrm{d}x\}.\end{aligned}$$

Hence, if h is simple, say

$$h(x) = \sum_{i=1}^n b_i 1_{B_i}(x)$$

for some $b_1, \dots, b_n \in \mathbb{R}$ and $B_1, \dots, B_n \in \mathcal{B}$, then

$$\begin{aligned}\mathbb{E}(h(X)) = \mathbb{E}\left(\sum_{i=1}^n b_i 1_{B_i}(x)\right) &= \sum_{i=1}^n b_i \mathbf{P}^X\{B_i\} = \sum_{i=1}^n b_i \int_\mathbb{R} 1_{B_i}(x)\, \mathbf{P}^X\{\mathrm{d}x\} \\ &= \int_\mathbb{R} \left(\sum_{i=1}^n b_i 1_{B_i}(x)\right) \mathbf{P}^X\{\mathrm{d}x\} \\ &= \int_R h(x)\, \mathbf{P}^X\{\mathrm{d}x\}\end{aligned}$$

using linearity of expectation. If $h \geq 0$, let $h_n \geq 0$ be simple with $h_n \uparrow h$ so that

$$\begin{aligned}\mathbb{E}(h(X)) = \mathbb{E}\left[\lim_{n\to\infty} h_n(X)\right] = \lim_{n\to\infty} \mathbb{E}(h_n(X)) = \lim_{n\to\infty} &\int_R h_n(x)\, \mathbf{P}^X\{\mathrm{d}x\} \\ &= \int_R \left[\lim_{n\to\infty} h_n(x)\right] \mathbf{P}^X\{\mathrm{d}x\} \\ &= \int_R h(x)\, \mathbf{P}^X\{\mathrm{d}x\}\end{aligned}$$

using the Monotone Convergence Theorem twice. In particular, (b) follows for $h \geq 0$. If we apply the above procedure to $|h|$, then (a) follows. If h is not positive, then writing $h = h^+ - h^-$ implies the result by subtraction. $\square$