November 7, 2021

EXPECTATION I

Abstract. We will define the expectation of a discrete random variable and prove basic properties.

1. Introduction

Let X be a discrete real-valued random variable taking values in the countable set $R \subseteq \mathbb{R}$ with pmf $f(x) = \mathbb{P}(X = x)$. Suppose that

$$\sum_{x \in R} |x| f(x) < \infty.$$

We define the **expectation** or **mean** of X via

$$\mathbb{E}X = \sum_{x \in \mathbb{R}} x f(x).$$

The justification for this definition is the following theorem.

Theorem 1 (Law of large numbers)**.** *Let $(X_i)_{i=1}^{\infty}$ be a i.i.d. sequence of (discrete) real-valued random variables, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume $\mathbb{E}X_1 = \mu \in \mathbb{R}$. Let the sample average be given by*

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

There exists an event $\Omega' \in \mathcal{F}$ with unit probability, such that for all $\omega \in \Omega'$ we have $S_n(\omega) \to \mu$ as $n \to \infty$.

Remark 1. *We have not defined expectation for continuous (and other) random variables yet, but Theorem 1 will hold as stated, given the appropriate definitions. The full proof of Theorem 1 is slightly beyond the scope of this course. However, we will be able to prove a slightly less general version that will be sufficient for many of our applications and examples.*

Remark 2. *We require the sum defining expectation to be absolutely convergent; this ensures that the order in which the summation takes place does not matter. In the case that X is a nonnegative random variable, it makes sense to define expectation to be infinite if the associated sum fails to converge.*

Remark 3. *Note that in the definition of expectation, it is computed using the only the pmf of X so expectation is invariant under choice of probability space; that is, if X and Y have the same law, then $\mathbb{E}X = \mathbb{E}Y$.*

Exercise 1.1. *Let $X \sim Bern(p)$, where $p \in [0, 1]$. Show that $\mathbb{E}X = p$.*

Exercise 1.2. *Let $X \sim Poi(\mu)$, where $\mu > 0$. Show that $\mathbb{E}X = \lambda$.*

2. Properties of Expectation

Theorem 2 (Change of variables, law of the unconscious statistician)**.** *Let X be a discrete random variable taking values on a countable set R with pdf f. If $g : R \to \mathbb{R}$, then*

$$\mathbb{E}g(X) = \sum_x g(x)f(x),$$

whenever the sum is absolutely convergent.

Proof. Note that $g(X)$ is a discrete real-valued random variable. If $B = \{g(x) \in R : x \in R\}$, then

$$\mathbb{E}g(X) = \sum_{y \in B} y\mathbb{P}(g(X) = y).$$

Let $R_y := \{x \in R : g(x) = y\}$. Note that the sets R_y partition the set R. Thus

$$\mathbb{P}(g(X) = y) = \mathbb{P}(R_y) = \sum_{x \in R_y} f(x)$$

and

$$\mathbb{E}g(X) = \sum_{y \in B} \sum_{x \in R_y} yf(x) = \sum_{y \in B} \sum_{x \in R_y} g(x)f(x) = \sum_{x \in R} g(x)f(x);$$

here the assumption of absolute convergence allows us to freely interchange the order of summation. □

Applying Theorem 2 with the the absolute value function, we have that the expectation of X is well-defined provided $\mathbb{E}|X| < \infty$.

If X and Y are random variables defined on same probability space, taking values on sets A and B respectively, we say that they are **jointly distributed** and if they are both discrete we say that they **joint distribution** is given by $j : A \times B \to [0, 1]$ with

$$j(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}\big(\{X = x\} \cap \{Y = y\}\big).$$

Notice that j is the pmf for the discrete random variable $Z = (X, Y)$.

Corollary 3 (Triangle inequality)**.** *Let $a, b \in \mathbb{R}$. If X and Y are jointly distributed discrete real-valued random variables, such that $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$, then $\mathbb{E}(|aX + bY| \leq |a|\mathbb{E}|X| + |b|\mathbb{E}|Y| < \infty$.*

Proof. Suppose X and Y take values on a countable sets $A, B \subseteq \mathbb{R}$, respectively. Consider the discrete random variable $Z = (X, Y)$ with joint pdf j and the function $g : A \times B \to \mathbb{R}$ given by

$$g(z) = g(x, y) = |ax + by|.$$

Note that

$$\begin{aligned}
\sum_{A\times B} g(x,y)j(x,y) &\leq |a| \sum_{A\times B} |x|j(x,y) + |b| \sum_{A\times B} |y|j(x,y) \\
&= |a| \sum_{x\in A} \sum_{y\in B} |x|j(x,y) + |b| \sum_{y\in B} \sum_{x\in A} |y|j(x,y) \\
&= |a| \sum_{x\in A} |x|\mathbb{P}(X = x) + |b| \sum_{y\in B} |y|\mathbb{P}(Y = y) \\
&= |a|\mathbb{E}|X| + |b|\mathbb{E}|Y|.
\end{aligned}$$

□

Corollary 4 (Linearity of expectation)**.** *Let $a, b \in \mathbb{R}$. If X and Y are jointly distributed discrete real-valued random variables, such that $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$, then $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$.*

Remark 4. *Note that in Corollary 4, we do not make any assumption on the joint distribution of X and Y; in particular, it is not necessary to assume they are independent.*

Exercise 2.1. *Prove Corollary 4.*

Exercise 2.2. *Let $X \sim Bin(n, p)$. Find $\mathbb{E}X$.*

Solution. We know that if $X_1, \ldots, X_n$ are i.id. Bernoulli random variables with parameter p, then $S = X_1 + \cdots X_n$ is a binomial random variable with parameter (n, p). Thus X has the same law as S. By Remark 3, we know that $\mathbb{E}S = \mathbb{E}X$ and Exercise 1.1 and Corollary 4, we have

$$\mathbb{E}S = \mathbb{E}(X_1 + \cdots + X_n) = np.$$

□

Remark 5. *Our solution to Exercise 2.2 is an example of a technique in probability theory sometimes referred to as coupling. We are easily able to compute $\mathbb{E}X$ by considering a probability space where it was easy to compute since in this case we envisioned X as an independent sum of Bernoulli's.*