NMAI059 – Probability and Statistics 1

Mykhaylo Tyomkyn

Lecture 4 - Expectation of discrete random variables.

Given a discrete random variable, we want to measure its 'average value'. One approach is to use the cdf $F_X(x) = \mathbb{P}(X \le x)$ and consider x satisfying $F_X(x) = 1/2$. This value is known as the *median* of X. But while the median is a useful concept, it suffers from being not always a unique value (e.g., any value $x \in [0, 1)$ is a median of $X \sim Bern(1/2)$) and being difficult to handle computationally. A more commonly used concept in both theory and practice is the *expected value ('mean')* of X.

Definition 1 (Expected value of a discrete rv.) *Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ be a discrete random variable, and suppose that*

$$\sum_{x \in \operatorname{Im} X} |x| p_X(x) < \infty. \tag{1}$$

The expected value/expectation/mean *of X, is defined as*

$$\mathbb{E}(X) = \sum_{x \in \operatorname{Im} X} x p_X(x).$$

Remark 1 *The validity of this definition is based on the fact from Analysis that absolute convergence of a series implies convergence, and the limit does not depend on the summation order.*

Remark 2 $\mathbb{E}(X)$ *is a function of the distribution p_X ("doesn't care where X is coming from").*

If (1) does not hold, we say that X does not have an expectation. Note that if $\operatorname{Im} X$ is finite, condition (1) becomes vacuously true, so $\mathbb{E}(X)$ always exists.

Example 1 *Let $X \sim Bern(p)$. Then $\mathbb{E}(X) = p \cdot 1 + (1 - p) \cdot 0 = p$*

Example 2 *Let $X \sim Bin(n, p)$. Then*

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k p^k (1-p)^{n-k} \binom{n}{k} = \sum_{k=1}^{n} p^k (1-p)^{n-k} \frac{n! \cdot k}{k!(n-k)!} \\
&= \sum_{k=0}^{n-1} p^{k+1} (1-p)^{n-k-1} \frac{n! \cdot (k+1)}{(k+1)!(n-k-1)!} = np \sum_{k=0}^{n-1} p^k (1-p)^{n-k-1} \binom{n-1}{k} \\
&= np(p + (1-p))^{n-1} = np.
\end{aligned}$$

Example 3 *Let $X \sim Geo(p)$. Then*

$$\begin{aligned}\mathbb{E}(X) &= p + 2p(1-p) + 3p(1-p)^2 + \dots = p(1 + 2(1-p) + 3(1-p)^2 + \dots) \\ &= p(1 + (1-p) + (1-p)^2 + \dots)^2 = p\left(\frac{1}{1-(1-p)}\right)^2 \\ &= p \cdot \frac{1}{p^2} = \frac{1}{p}.\end{aligned}$$

We will shortly see more elegant ways of doing the last two examples.

Sometimes we will have that $\operatorname{Im} X$ is infinite but $X \geq 0$ a.s.[1], that is $\mathbb{P}(X \geq 0) = 1$. In this case either $\mathbb{E}(X)$ exists ('is finite') or we have $\sum xp_X(x) = \infty$, in which case we may informally write $\mathbb{E}(X) = +\infty$.

Example 4 *Let $p_X((-2)^k) = 2^{-k}$ for $k = 1, 2, \dots$. Let $Y = |X|$, so $p_Y(2^k) = 2^{-k}$ for $k = 1, 2, \dots$. Then*

$$\sum_{x \in \operatorname{Im} X} |x| p_k(x) = \sum_{k=1}^{\infty} 2^k \cdot 2^{-k} = \infty,$$

so X does not have an expectation. The same calculation shows that Y also does not have an expectation, but since $Y \geq 0$ a.s., we may write $\mathbb{E}(Y) = +\infty$.

Lemma 1 *Let Ω be countable and $X\colon (\Omega, 2^\Omega, \mathbb{P}) \to \mathbb{R}$ be a discrete random variable with finite expectation. We have*

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

Proof

$$\begin{aligned}\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) &= \sum_{x \in \operatorname{Im} X} \sum_{\omega \in \Omega\colon X(\omega)=x} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{x \in \operatorname{Im} X} \sum_{\omega \in \Omega\colon X(\omega)=x} x\mathbb{P}(\{\omega\}) \\ &= \sum_{x \in \operatorname{Im} X} x \sum_{\omega \in \Omega\colon X(\omega)=x} \mathbb{P}(\{\omega\}) = \sum_{x \in \operatorname{Im} X} x\mathbb{P}(X = x) \\ &= \sum_{x \in \operatorname{Im} X} xp_X(x) = \mathbb{E}(X).\end{aligned}$$

□

Example 5 *A fair coin is tossed twice, independently. Let $X : \{H, T\}^2 \to \{0, 1, 2\}$ count the number of 'heads'. Then*

$$\mathbb{E}(X) = \sum_{x \in \operatorname{Im} X} xp_X(x) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1, \text{ or alternatively}$$

$$\mathbb{E}(X) = 0 \cdot \mathbb{P}(TT) + 1 \cdot \mathbb{P}(HT) + 1 \cdot \mathbb{P}(TH) + 2 \cdot \mathbb{P}(HH) = 0 + \frac{1}{4} + \frac{1}{4} + \frac{1}{2} = 1.$$

[1] a.s. is short for "almost surely". "Event A holds a.s." means $\mathbb{P}(A) = 1$.

Theorem 1 (Basic properties of the expectation) *Let X be a discrete random variable with finite mean.*

1. *If $X \geq 0$ a.s. then $\mathbb{E}(X) \geq 0$.*
2. *If $\mathbb{E}(X) \geq 0$ then $\mathbb{P}(X \geq 0) > 0$.*
3. *If $X \geq 0$ a.s. and $\mathbb{E}(X) = 0$ then $X = 0$ a.s.*
4. *$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ for all $a, b \in \mathbb{R}$.*

Proof 1.-3. follow directly from the definitions. 4. is the consequence of the LOTUS theorem, see below. □

Theorem 2 (LOTUS: law of the unconscious statistician) *Let X be a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$ be a function. Then, provided $\mathbb{E}(g(X))$ exists,*

$$\mathbb{E}(g(X)) = \sum_{x \in \operatorname{Im} X} g(x) p_X(x)$$

Proof Let $Y = g(X)$. Then

$$\begin{aligned}
\mathbb{E}(g(X)) = \mathbb{E}(Y) &= \sum_{y \in \operatorname{Im} Y} y \mathbb{P}(Y = y) = \sum_{y \in \operatorname{Im} g(X)} y \mathbb{P}(g(X) = y) \\
&= \sum_{y \in \operatorname{Im} g(X)} \sum_{x \in \operatorname{Im} X :\, g(x) = y} g(x) \mathbb{P}(X = x) \\
&= \sum_{x \in \operatorname{Im} X} g(x) \mathbb{P}(X = x) = \sum_{x \in \operatorname{Im} X} g(x) p_X(x).
\end{aligned}$$

□

Applying LOTUS with $g(x) = ax + b$ readily gives statement 4. of Theorem 1:

$$\mathbb{E}(aX + b) = \sum_{x \in \operatorname{Im} X} (ax + b) p_X(x) = a \sum_{x \in \operatorname{Im} X} x p_X(x) + b \sum_{x \in \operatorname{Im} X} p_X(x) = a\mathbb{E}(X) + b.$$

Theorem 3 (Linearity of expectation) *Let $X, Y : \Omega \to \mathbb{R}$ be discrete random variables with finite expectation. Then*

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

In particular, $X + Y$ has finite expectation.

Proof We will prove it in the case when Ω is countable, and leave the general case as a harder exercise. When Ω is countable, by Lemma 1 we can write

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

□

Corollary 1 (Monotonicity of $\mathbb{E}$) *If $X \geq Y$ a.s. then $\mathbb{E}(X) \geq \mathbb{E}(Y)$.*

Proof By the previous theorems, since $X - Y \geq 0$ a.s., we obtain

$$0 \leq \mathbb{E}(X - Y) = \mathbb{E}(X) + \mathbb{E}(-Y) = \mathbb{E}(X) - \mathbb{E}(Y).$$

□

Example 6 *Let us revisit Example 2. We have $X : \{H, T\}^n \to \mathbb{R}$, $X \sim Bin(n, p)$, counting the 'heads'. Then $X = X_1 + \cdots + X_n$, where X_i is the number of heads at the i-th toss. So $X_i \sim Bern(p)$ for all i, and by linearity of expectation we obtain*

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

Note that we did not even use that the coin tosses are independent.

Just like conditional probabilities we may also define conditional expectations of random variables with respect to events.[2]

Definition 2 (Conditional expectation) *Let X be a discrete random variable, and let $A \in \mathcal{F}$, $\mathbb{P}(A) > 0$ be an event. The* conditional expectation of X given A *is (assuming convergence)*

$$\mathbb{E}(X \mid A) := \sum_{x \in \operatorname{Im} X} x \mathbb{P}(X = x \mid A).$$

Theorem 4 (Law of total expectation) *Let $A_1, A_2 \ldots$ be a countable partition of Ω with $\mathbb{P}(A_i) > 0$ for all i. Then (assuming convergence)*

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid A_i)\mathbb{P}(A_i).$$

Proof

$$\begin{aligned}\sum_i \mathbb{E}(X \mid A_i)\mathbb{P}(A_i) &= \sum_i \sum_{x \in \operatorname{Im} X} x \mathbb{P}(X = x \mid A_i)\mathbb{P}(A_i) = \sum_{x \in \operatorname{Im} X} x \sum_i \mathbb{P}(\{X = x\} \cap A_i) \\ &= \sum_{x \in \operatorname{Im} X} x \mathbb{P}(X = x) = \mathbb{E}(X).\end{aligned}$$

□

Example 7 *Revisiting Example 3, let $X \sim Geo(p)$ be measuring the time of the first 'heads' in a sequence of independent tosses. Let A be the event that the first toss is 'heads'. By the law of total expectation*

$$\mathbb{E}(X) = \mathbb{E}(X \mid A)\mathbb{P}(A) + \mathbb{E}(X \mid \overline{A})\mathbb{P}(\overline{A}) = 1 \cdot p + \mathbb{E}(X + 1)(1 - p) = p + (\mathbb{E}(X) + 1)(1 - p).$$

Resolving this (and assuming $\mathbb{E}(X)$ is finite) gives $\mathbb{E}(X) = 1/p$.

[2] A more general concept of conditional expectation with respect to another random variable will be out of scope for this course.