Lecture 8: Expectations

Rajat Mittal

IIT Kanpur

Many a times in probability theory, we are interested in a numerical value associated to the outcomes of the experiment. For example, number of heads in a sequence of tosses, pay out of a lottery, number of casualties after a Hurricane etc.. These functions which assign a numerical value to the outcomes of the experiment are called *random variables*.

In this lecture, we will look at random variables, their expectation and properties of expected value.

1 Random variable

Given the sample space Ω of an experiment, a random variable is a function $X : \Omega \to \mathbb{R}$. So a random variable assigns a real value $X(\omega)$ to every element ω of the sample space Ω. If the range of X is countable then X is called a discrete random variable. In this course, we will only be interested in discrete random variables.

Given a probability function Pr on Ω, the natural definition of probability of the random variable is,

$$Pr(X = x) = \sum_{\omega : X(\omega) = x} Pr(\omega).$$

This is called the *probability mass function* of a random variable.

Note 1. If X is a random variable then $g(X)$ is also a random variable, where g is any function from $\mathbb{R}$ to $\mathbb{R}$.

Let's look at some examples of random variables and their probability mass function.

– Suppose the experiment consists of tossing a fair coin 10 times. The sample space is all sequences of length 10 of H, T. Define the random variable to be the number of heads in the sequence. That is, $X(\omega)$ is the number of H's in ω.

  Exercise 1. Show that the probability of getting a sequence with k heads for a length 10 sequence is,

$$\binom{10}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10-k}.$$

  Then the probability mass function of the random variable for k between 0 and 10 is,

$$P(X = k) = \binom{10}{k}\left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{10-k}.$$

  Exercise 2. Generalize the probability mass function if the probability of head is p.

– For an experiment, we ask the birthday's of students in a class one by one. We stop as soon as we find two people with matching birthday. What is the probability mass function for the random variable X which counts the number of students queried?
  Let us find out the probability that we queried k people. Then the first $k-1$ birthdays are distinct and the last one matches at least one of the first $k-1$. First birthday will not match with anyone before that with probability 1. The next one will not match with probability 364/365 and so on. The last one will match with probability $(k-1)/365$.
  Hence,

$$Pr(X = k) = \frac{(k-1)}{365^k}(k-1)!\binom{365}{k-1}.$$

Exercise 3. Calculate these numbers using a calculator for k up to 23.

– In a set of 1000 balls, 150 balls have some defect. Say, we choose 50 balls and inspect, then let X be the random variable which denotes the number of defected balls found.
  The probability mass function of X is non-zero for $x = 1$ to 50. It is given by,

$$Pr(X = k) = \frac{\binom{150}{k}\binom{850}{50-k}}{\binom{1000}{50}}.$$

Like the case of probability function, two random variables are called independent if the product of their probability mass function gives the probability mass function of both.

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$$

Exercise 4. Let X be the random variable that assigns 1 if the number on the throw of a dice is even else it is -1. Let Y be the random variable that assigns 1 if the number on the throw of a dice is prime else it is -1. Show that X and Y are not independent.

2 Expectation

We introduced random variables because of our interest in numerical values associated with the set of outcomes. Many a times we are interested in the average, mean or expectation of this numerical value. Taking an example, in a sequence of 10 tosses, you get 1 Rs. every time it turns out to be head. What is your expected earning?

The expectation is easy to define if all outcomes are equally likely. In that case, if $\Omega = \{\omega_1, \cdots, \omega_n\}$, we would expect to get the average $\left(\frac{X(\omega_1)+X(\omega_2)+\cdots+X(\omega_n)}{n}\right)$.

Taking this idea further, the expected value of a random variable X is defined as,

$$E[X] := \sum_{x\in\mathbb{R}} Pr(X(\omega) = x)x.$$

Since random variables are discrete in our case, most of the time range of X will be much smaller. By range of X, we mean values attained by X with a non-zero probability. Suppose the range of X is R, then

$$E[X] := \sum_{x\in R} Pr(X(\omega) = x)x.$$

It is a common misinterpretation, probably because of name, when it is stated that X will attain value $E[X]$ with high probability. That is clearly not the case. It is easy to construct cases where $E[X]$ might not even be in R.

Exercise 5. Construct a random variable such that $E[X]$ is not in R.

The correct interpretation is, if we independently repeat the experiment multiple times, then with high probability the average value will be close to expectation. This intuition will be formalized in the later sections.

Exercise 6. In a probabilistic experiment, you get 100 Rs. every time an odd number shows up on a dice. You loose 100 Rs. every time an even number shows up. What is your expected earning.

Consider some more examples of expectation.

– Your friend is ready to give you 100 Rs. if on a throw of a dice, an odd prime turns up. What amount can you give him if the number is not an odd prime?

Exercise 7. What is the random variable in this case?

You cannot be certain to win this game if you bet any positive amount. The bet will be profitable to you if the expected profit for you is greater than zero (at least when the experiment is repeated multiple times).
If you bet x Rs., then the expected earning should be greater than zero, $1/3 \times 100 + 2/3 \times (-x) \geq 0$. So you can agree to pay any amount less than 50. This example suggests that in a fair bet, the expected profit/loss should be zero.

Exercise 8. Suppose the expected value of a random variable X is zero. What is the expected value of $-X$?

- You toss a coin till you get head. What is the expected number of tosses?
  The random variable is the number of tosses.

  Exercise 9. Show that $Pr(X = k) = (1/2)^k$.

  The expected number of tosses is,
  $$E[X] = \sum_k k\,(1/2)^k\,.$$
  Let $S = \sum_k k\,(1/2)^k$, then $(1/2)\,S = \sum_k k\,(1/2)^{k+1}$. Subtracting and then further manipulation tell us that $S = 2$.
- Your friend asks you to bet on the rise/fall of stock market. Both of you put 100 Rs. in a pot and guess. If the guesses are same then both get 100 Rs., otherwise the one with the correct guess gets all the money. Assuming that stock market rises/falls with equal probability, this bet is fair. Now your friend wants to include her brother's share also. So she will put 200 Rs. and also the guess for her and her brother. Should you take the bet?
  Suppose you friend puts opposite guesses for her and her brother all the time. If you guess is correct then you will get 150 Rs. and if you incorrectly guess then you get 0 Rs.. The expected value of your pot earning is 75 Rs., which is less than the amount you put in the pot. So it is not an advisable bet.

One point of caution is that expectation need not be defined all the time. Let X be a random variable such that $Pr(X = k) = \frac{6}{\pi^2 k^2}$. It can be shown that $\sum_k Pr(X = k) = 1$ but $\sum_k Pr(X = k)k$ is not convergent.
If X is a random variable then so is $g(X)$ where g is any function from $\mathbb{R}$ to $\mathbb{R}$. Then,
$$E[g(X)] = \sum_{x \in \mathbb{R}} g(x) Pr(X = x).$$

Note 2. We need to assume that $\sum_x |g(x)| Pr(X = x)$ converges.

2.1 Linearity of expectation

One of the most important property of expectation is that it is linear. This means that given two random variables X and Y,
$$E[X + Y] = E[X] + E[Y].$$

Note 3. Here say $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$, then random variable $X + Y : \Omega \to \mathbb{R}$ is defined as $X + Y(\omega) = X(\omega) + Y(\omega),\ \forall \omega \in \Omega$.
This property is known as *linearity of expectation.*

Note 4. Nothing is assumed about the relationship between X and Y. They might be dependent or independent random variables.

Proof. The expectation $E[X+Y]$ is given using the probability mass function $Pr(X=x,Y=y)$.

$$\begin{aligned}
E[X+Y] &= \sum_{x,y}(x+y)Pr(X=x,Y=y) \\
&= \sum_{x,y} xPr(X=x,Y=y) + yPr(X=x,Y=y) \\
&= \sum_{x} x \sum_{y} Pr(X=x,Y=y) + \sum_{y} y \sum_{x} Pr(X=x,Y=y) \\
&= \sum_{x} xPr(X=x) + \sum_{y} yPr(Y=y) \\
&= E[X]+E[Y]
\end{aligned} \tag{1}$$

□

The linearity of expectation can be extended to more than two events using induction.

Exercise 10. What will be the statement?

Because this property works irrespective of the dependence between random variables, it has many applications. Let's look at one example.

Suppose you want to collect stickers which accompany your favorite chewing-gum. Every time you buy one, one sticker comes, out of n, with equal probability. How many chewing-gums should you buy to collect all stickers in expectation?

Let T be the random variable which counts the number of packets to be bought to collect n stickers.

Let S_1 be the random variable that we get first sticker (clearly $S_1 = 1$). Let S_2 be the extra number of chewing-gums for getting second different sticker, similarly define S_k. We need to calculate $E[T] = E[S_1 + S_2 + \cdots + S_n]$.

By linearity of expectation, we only need to worry about $E[S_k]$. The probability that $S_k = r$ is,

$$Pr(S_k = r) = ((k-1)/n)^{r-1}\left(1 - \frac{k-1}{n}\right).$$

Exercise 11. Show that $E[S_k] = \frac{n}{n-(k-1)}$.

This implies that the expected number of days needed to collect all stickers is $E[T] = \sum_k \frac{n}{n-(k-1)}$.

3 Chernoff bound

We gave the interpretation of expectation in the previous section. The statement was, if the random variable is repeated large number of times, then the average is close to the expected value with high probability. This statement will be formalized in this section.

First, we will prove *Markov's inequality.* It follows from the definition of expectation.

Theorem 1. *Given a positive random variable X and $a > 0$,*

$$Pr(X \geq a) \leq \frac{E[X]}{a}.$$

Note 5. If the random variable is not positive then, $Pr(|X| \geq a) \leq \frac{E[|X|]}{a}$, by applying Markov's inequality to $|X|$.

Proof. The result will be proved by contradiction. Assume that the converse holds, $Pr(X \geq a) > \frac{E[X]}{a}$.

$$
\begin{aligned}
E[X] &= \sum_x Pr(X = x)x \\
&\geq \sum_{x<a} Pr(X = x).0 + \sum_{x \geq a} Pr(X = x).a \\
&= a \sum_{x \geq a} Pr(X = x) \\
&> E[x]
\end{aligned}
\tag{2}
$$

Where the last inequality follows from assumption. So the assumption is false and hence Markov inequality is proved. □

Using Markov's inequality we can prove *Chernoff bound*. Suppose an experiment succeeds with probability p and otherwise fails. The expected value of success is p. If we repeat the experiment n times then the expectation is np by linearity of expectation. Chernoff bound shows that if we repeat the experiment many times (say n), then the number of successes will be close to np with very high probability (depending upon n).

Theorem 2. *Chernoff bound: Let X be a random variable which takes value 1 with probability p and 0 otherwise. Let $X_1, X_2, \cdots, X_n$ correspond to random variable X repeated n times (the experiment is repeated n times). Define $S = \sum_{i=1}^n X_i$, then*

$$Pr(S < (1-\delta)nE[X]) \leq e^{\frac{-nE[X]\delta^2}{2}}.$$

Note 6. We have taken a very special form of random variable X, but it can be generalized.

Proof. This proof is taken from John Canny's lecture notes, http://www.cs.berkeley.edu/ jfc/cs174/lecs/lec10/lec10.pdf.

The proof of Chernoff bound follows by looking at the random variable e^{-tS}, where t is a parameter and will be optimized later. Define $u := E[S] = nE[X]$, so

$$Pr(S < (1-\delta)u) = Pr(e^{-tS} > e^{-t(1-\delta)u}).$$

We can apply Markov's inequality for e^{-tS},

$$Pr(S < (1-\delta)u) \leq \frac{E[e^{-tS}]}{e^{-t(1-\delta)u}}.$$

But e^{-tS} is the product of e^{-tX_i}, where X_i are independent. So,

$$Pr(S < (1-\delta)u) \leq \frac{\Pi_{i=1}^n E[e^{-tX_i}]}{e^{-t(1-\delta)u}}. \tag{3}$$

Exercise 12. Show that $E[e^{-tX_i}] = 1 - p(1-e^{-t}) \leq e^{p(e^{-t}-1)}$.

Above exercise implies that $\Pi_{i=1}^n E[e^{-tX_i}] \leq e^{u(e^{-t}-1)}$. From Eq. 3, we get

$$Pr(S < (1-\delta)u) \leq e^{u(e^{-t}+t(1-\delta)-1)}$$

Exercise 13. Show that the bound on right is minimized for $t = \ln \frac{1}{1-\delta}$.

Putting the best t, we get

$$Pr(S < (1-\delta)u) \leq \left(\frac{e^{-\delta u}}{(1-\delta)^{u(1-\delta)}}\right).$$

Using the Taylor expansion of $\ln(1-\delta)$,

$$Pr(S < (1-\delta)u) \leq e^{\frac{-u\delta^2}{2}}.$$

Hence proved.

□

4 Assignment

Exercise 14. In a group of 23 people, we ask birthday of everyone. Define the random variable X to be the number of pairs whose birthdays match. What is $Pr(X \geq 1)$?

Exercise 15. Suppose you pick two cards, from a deck with cards numbered from 1 to 1000. What is the expected value of the greater number?

Exercise 16. Let X and Y be two independent random variables. Prove that,

$$E[XY] = E[X]E[Y].$$

Exercise 17. Let X be a random variable with $Pr(X = 1) = p$ and $Pr(X = 0) = 1-p$. Find $E[X_1X_2\cdots X_n]$ where $X_1, X_2, \cdots X_n$ are identical and independent copies of X.

Exercise 18. Let X be a random variable with $Pr(X = 1) = p$ and $Pr(X = -1) = 1-p$. Find $E[X^n]$.

Exercise 19. Prove Chebyshev's inequality.

$$P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}$$

Where $Var(X) = E[(X - E[X])^2]$.

Exercise 20. For the sticker collection problem in the linearity of expectation. Let us look at a different solution. Say T_i be the random variable which counts the packets needed to collect i^{th} sticker. Then $E[T_i] = n$, and $E[T] = \sum_i E[T_i] = n^2$. Is this argument correct, if not, what is wrong with this argument?

References

1. H. Tijms Understanding Probability. *Cambridge University Press*, 2012.
2. D. Stirzaker. Elementary Probability. *Cambridge University Press*, 2003.
3. D. Kahneman. Thinking, Fast and Slow. *Farrar, Straus and Giroux*, 2011.