Lecture 30: Basic probability theory

Nitin Saxena *

IIT Kanpur

0.1 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,

$$E[X+Y] = E[X] + E[Y]\,.$$

This property is known as *linearity of expectation.*

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

$$\begin{aligned}
E[X+Y] &= \sum_{x,y} (x+y) P(X=x, Y=y) \\
&= \sum_{x,y} xP(X=x, Y=y) + yP(X=x, Y=y) \\
&= \sum_{x} x \sum_{y} P(X=x, Y=y) + \sum_{y} y \sum_{x} P(X=x, Y=y) \\
&= \sum_{x} xP(X=x) + \sum_{y} yP(Y=y) \\
&= E[X] + E[Y]\,.
\end{aligned}$$

□

Exercise 1. Extend the linearity of expectation to more than two random variables using induction.

Because this property works irrespective of the dependence between random variables, it has many applications. Let us 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 expect to buy before you collect all the stickers?

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

Let S_1 be the count at which we get the first sticker (clearly $S_1 = 1$). Let S_2 be the extra number of chewing-gums for getting the 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,

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

Exercise 2. Show that $E[S_k] = \frac{n}{n-(k-1)}$.

$E[S_k] = \sum_{r\geq 0} r \cdot P(S_k = r) = \sum_{r\geq 0} r \cdot ((k-1)/n)^{r-1} \left(1 - \frac{k-1}{n}\right) = \left(1 - \frac{k-1}{n}\right)^{-1} = \frac{n}{n-k+1}$.

This implies that the expected number of chewing-gums needed to collect all n stickers is $E[T] = \sum_{k\in[n]} \frac{n}{n-k+1} = n \cdot \sum_{k\in[n]} k^{-1} \approx n \ln n$. (It is between $n\ln(n+1)$ and $n\ln n + n$.)

* Edited from Rajat Mittal's notes.

1 Markov, Chernoff bounds

We interpreted expectation in the previous section as: if the random variable is measured a 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 (Chebyshev-Markov, 1867). *Given a positive random variable X and $a > 0$,*

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

Note 1. If the random variable is not positive then, $P(|X| \geq a) \leq \frac{E[|X|]}{a}$, by applying Markov's inequality to $|X|$.

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

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

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*– it is a stronger result in the case of independent experiments. Suppose an experiment succeeds with probability p. 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 (close to 1 depending inverse-*exponentially* upon n).

Theorem 2 (Chernoff, 1952). *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 measured n times (the experiment is independently repeated n times). Define $S = \sum_{i=1}^n X_i$, then (for any $\delta \in (0,1)$)*

$$P(\,S < (1-\delta) \cdot n \cdot E[X]\,) \leq e^{-nE[X]\delta^2/2}\,.$$

Note 2. 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

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

We can apply Markov's inequality for e^{-tS} (note: it is positive valued),

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

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

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

Exercise 3. Show that for mutually independent random variables the expectation operator $E[\cdot]$ is *multiplicative*.

$\square$

References

1. D. Stirzaker. Elementary Probability. *Cambridge University Press*, 2003.