CS 30: Discrete Math in CS (Winter 2020): Lecture 16

Date: 6th February, 2020 (X-hour)

Topic: Probability: Random Variables, Expectation

Disclaimer: These notes have not gone through scrutiny and in all probability contain errors. Please discuss in Piazza/email errors to deeparnab@dartmouth.edu

---

1. **Random Variable.**

   Given a random experiment with outcomes Ω, a *real valued random variable* X defined over this experiment is a mapping $X : \Omega \to \mathbb{R}$. An *integer valued random variable* X is a mapping from $X : \Omega \to \mathbb{Z}$.

   Examples:

   - We toss a fair coin. $X(\text{heads}) = 0$ and $X(\text{tails}) = 1$. This is a $\{0, 1\}$-random variable, or a Boolean random variable. Also called a *Bernoulli* random variable.
   - We roll a fair die. X takes the value on the face of the die.
   - We roll *two* fair dice. X takes the value of the sum. In this case, $X = Y + Z$ where Y, Z are two *identical* random variables of the kind from the previous bullet point.
   - We toss 1000 fair coins. Z takes the value of the number of heads we see.
   - Given any event $\mathcal{E}$, there is an associated random variable called the *indicator random variable* denoted as $\mathbf{1}_{\mathcal{E}}$, where $\mathbf{1}_{\mathcal{E}}(\omega) = 1$ if $\omega \in \mathcal{E}$, and 0 otherwise.

2. **Events associated with random variables.**

   Given a random variable X, we can associate many events and ask for their probabilities. For instance, we can ask $\mathbf{Pr}[X = x]$. More precisely, this is a shorthand for saying $\sum_{\omega \in \Omega : X(\omega) = x} \mathbf{Pr}[\omega]$.

   Similarly, $\mathbf{Pr}[X \geq k]$ is a shorthand for saying $\sum_{\omega \in \Omega : X(\omega) \geq k} \mathbf{Pr}[\omega]$.

3. **"Shape" of a Random Variable.**

   Since X is real valued (or integer valued), one can plot how the $\mathbf{Pr}[X = x]$ looks like with respect to X. The following plots show a couple of examples. The first set of figures (Figure 1) is related to dice. We roll N dice, each independent of one another, and we use X to denote the sum of the numbers seen. The plots show how $\mathbf{Pr}[X = x]$ changes with x, as x goes from 0 to $6N + 1$. As you can see, when $N = 1$, the probabilities are the same for each number, and equals $1/6$th. However, the distribution becomes less and less uniform as N grows.

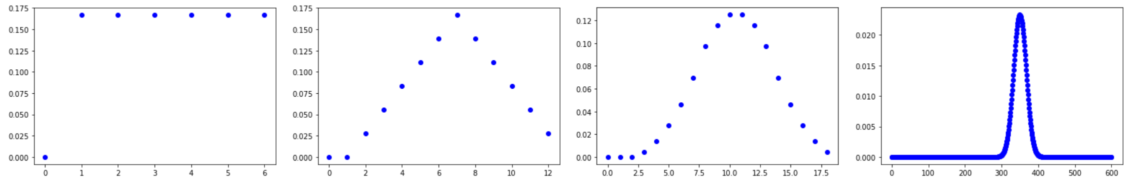

Figure 1: *The above graphs plot the probability of seeing a particular sum on the Y-axis against the possible sums on the X-axis. From left to right, the number of dice is* $1, 2, 3$ *and* 100*.*

The next set of figures (Figure 2) relate to coin tosses. We toss N coins and Z denotes the number of heads we see. The plots in blue (the ones to the left) are the plots of tosses of fair coins which turn up heads 50-50. The plots in green (the ones to the right) are for biased coins which come up heads with probability 0.3.

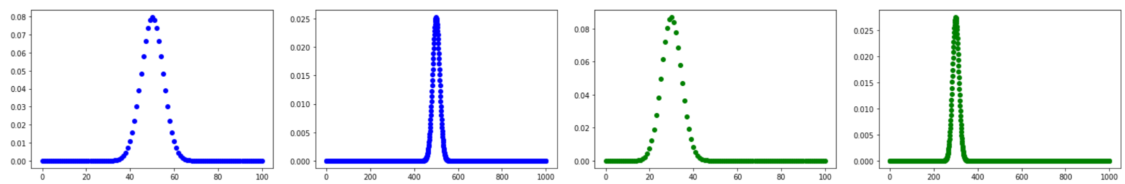

Figure 2: *The above graphs plot the probability of seeing a particular number of heads on the Y-axis against the reals on the X-axis. The first two figures (in blue) on the left are for fair coins, with* $N = 100$ *coins tossed and* $N = 1000$ *coins tossed. The two figures in the right (in green) are for biased coins which come heads with* 0.3 *probability. The number of coins are* $N = 100$ *and* $N = 1000$ *respectively.*

Remark: *A few points are noteworthy*

- *Note the shapes become "narrower" as the number of coins/dice grow.*
- *Note that the shape of fair coin is similar to the shape of biased coins with just a shift.*
- *Note that the* 100 *dice shape looks quite similar to the shape with* 1000 *coins.*

All of these happen for a very important reason (which we will not cover, unfortunately). The reason, informally, states that if we take many, many independent copies of the same random variable (dice, coin, whatever), and add them all up, their shape (or "distribution" more formally) all tend to look the same (like a bell curve). This unifying shape is called the "normal distribution" or the "Gaussian distribution".

4. **Expectation of a Random Variable.**

The expectation of a random variable X is defined to be

$$\mathbf{Exp}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{Pr}[\omega]$$

Here is another simpler, and possibly more useful, formula to calculate expectation.

Theorem 1. For any random variable X, we have

$$\mathbf{Exp}[X] = \sum_{k \in \mathbb{R}} k \cdot \mathbf{Pr}[X = k]$$

Proof.

$$\begin{aligned}
\mathbf{Exp}[X] &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbf{Pr}[\omega] &&= \sum_{k \in \mathbb{R}} \left(\sum_{\omega \in \Omega : X(\omega) = k} X(\omega) \cdot \mathbf{Pr}[\omega] \right) \qquad (1) \\
&= \sum_{k \in \mathbb{R}} \left(\sum_{\omega \in \Omega : X(\omega) = k} k \cdot \mathbf{Pr}[\omega] \right) &&= \sum_{k \in \mathbb{R}} k \cdot \left(\sum_{\omega \in \Omega : X(\omega) = k} \mathbf{Pr}[\omega] \right) \\
&= \sum_{k \in \mathbb{R}} k \cdot \mathbf{Pr}[X = k]
\end{aligned}$$

The main idea is to partition Ω based on various valued $X(\omega)$ takes, and for each of those, $X(\omega)$ can be pulled out of the summation. □

Remark: *The expectation is therefore often thought of as an inner-product (aka dot-product) of two vectors. These vectors have $|\Omega|$ dimensions. One vector is $(X(\omega) : \omega \in \Omega)$, and the other is $(\mathbf{Pr}[\omega] : \omega \in \Omega)$. This dot-product view is often useful (although, sadly, we may not see its ramifications in this course).*

Examples: We now use the above formula to calculate expectations of a bunch of random variables.

- *We toss a fair coin. $X(heads) = 0$ and $X(tails) = 1$. This is a $\{0,1\}$-random variable, or a Boolean random variable. Also called a* Bernoulli *random variable.*

  $$\mathbf{Exp}[X] = 0 \cdot \mathbf{Pr}[X = 0] + 1 \cdot \mathbf{Pr}[X = 1] = 1/2$$

  Indeed, if the coin were not fair, and the probability that tails would come with probability p, then $\mathbf{Exp}[X] = p$.

- *We roll a fair die. X takes the value on the face of the die.*

  $$\mathbf{Exp}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

- *We roll two fair dice. X takes the value of the sum. In this case, $X = Y + Z$ where Y, Z are random variables of the kind from the previous bullet point.*

  This is requires a little work. The range of X is $\{2,3,4,5,6,7,8,9,10,11,12\}$. We can calculate the probabilities for each (remember, it is not uniform), and then do the calculation. ✍

Exercise: *Please do the calculation.*

We get the answer 7. Did you?

- *We toss a fair coin* 100 *times.* Z *is the number of heads.*

  This is a lot more work. First, we observe the $\mathrm{range}(Z) = \{0, 1, 2, \ldots, 100\}$. Then, we try to figure out $\mathbf{Pr}[Z = k]$. This is $\frac{1}{2^{100}} \cdot \binom{100}{k}$. (Do you see how? There are 2^{100} possible outcomes, each equally likely coz the coins are fair, and $\binom{100}{k}$ have exactly k heads.). Therefore,

$$\mathbf{Exp}[Z] = \sum_{k=0}^{100} k \cdot \binom{100}{k} \cdot \frac{1}{2^{100}}$$

  Phew!

- *Given any event* $\mathcal{E}$*, there is an associated random variable called the* indicator random variable *denoted as* $\mathbf{1}_{\mathcal{E}}$*, where* $\mathbf{1}_{\mathcal{E}}(\omega) = 1$ *if* $\omega \in \mathcal{E}$*, and* 0 *otherwise.*

$$\mathbf{Exp}[\mathbf{1}_{\mathcal{E}}] = 0 \cdot \mathbf{Pr}[\neg\mathcal{E}] + 1 \cdot \mathbf{Pr}[\mathcal{E}] = \mathbf{Pr}[\mathcal{E}]$$

  This is quite important. Why? Because it turns a probability calculation (the RHS) into an expectation calculation. As we show below, calculating expectations is often easier than calculating probabilities.

✍

Exercise: *Suppose you have a fair coin. Construct the following random variable* Z *whose range is* $\mathbb{N}$*. You keep tossing the fair coin till you get a heads.* Z *is the number of times you have tossed the coin. What is* $\mathbf{Exp}[Z]$*? To do this, figure out what is* $\mathbf{Pr}[Z = k]$*. Then write the expectation as a sum. Then see if you can simplify the sum.*

5. **Multiplication by a scalar.** If X is a random variable, and c is a "scalar" (a constant), then $Z = c \cdot X$ is another random variable. $\mathbf{Exp}[c \cdot X] = c \cdot \mathbf{Exp}[X]$. ✍

Exercise: *Prove this.*

6. **Expectation of a function of a random variable.** Let X be a random variable, and let $f : \mathbb{R} \to \mathbb{R}$ be any function. One can then define a random variable $Z := f(X)$, defined as $Z(\omega) = f(X(\omega)$. The following easily follows as in the proof of Theorem 1.

Theorem 2. $\mathbf{Exp}[f(X)] = \sum_{k \in \mathbb{R}} f(k) \cdot \mathbf{Pr}[X = k]$.

Proof.

$$\begin{aligned}
\mathbf{Exp}[f(X)] = \mathbf{Exp}[Z] &= \sum_{\omega \in \Omega} Z(\omega) \cdot \mathbf{Pr}[\omega] = \sum_{\omega \in \Omega} f(X(\omega)) \cdot \mathbf{Pr}[\omega] \\
&= \sum_{k \in \mathbb{R}} \left(\sum_{\omega \in \Omega : X(\omega) = k} f(X(\omega)) \cdot \mathbf{Pr}[\omega] \right) = \sum_{k \in \mathbb{R}} \left(\sum_{\omega \in \Omega : X(\omega) = k} f(k) \cdot \mathbf{Pr}[\omega] \right) \\
&= \sum_{k \in \mathbb{R}} f(k) \cdot \left(\sum_{\omega \in \Omega : X(\omega) = k} \mathbf{Pr}[\omega] \right) \\
&= \sum_{k \in \mathbb{R}} f(k) \cdot \mathbf{Pr}[X = k]
\end{aligned}$$

□

Example.

- *We roll a fair die. X takes the value on the face of the die.*

$$\mathbf{Exp}[X^2] = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

and

$$\mathbf{Exp}\left[\frac{1}{X}\right] = \frac{1}{1} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \cdots + \frac{1}{6} \cdot \frac{1}{6} = \frac{49}{120}$$

✍

Exercise: *Which is bigger –* $\mathbf{Exp}[X^2]$ *or* $(\mathbf{Exp}[X])^2$*?* $\mathbf{Exp}\left[\frac{1}{X}\right]$ *or* $\frac{1}{\mathbf{Exp}[X]}$*?*

7. **Linearity of Expectation.** This is one of the most powerful equations in all of probability. Literally. It states the following. It literally has a four line proof.

Theorem 3. For any two random variables X and Y, let $Z := X + Y$. Then,

$$\mathbf{Exp}[Z] = \mathbf{Exp}[X] + \mathbf{Exp}[Y]$$

Proof.

$$\begin{aligned}
\mathbf{Exp}[Z] &= \sum_{\omega\in\Omega} Z(\omega)\mathbf{Pr}[\omega] && \text{Definition of Expectation} \\
&= \sum_{\omega\in\Omega} (X(\omega) + Y(\omega))\,\mathbf{Pr}[\omega] && \text{Definition of } Z \\
&= \sum_{\omega\in\Omega} X(\omega)\mathbf{Pr}[\omega] + \sum_{\omega\in\Omega} Y(\omega) \cdot \mathbf{Pr}[\omega] && \text{Distributivity} \\
&= \mathbf{Exp}[X] + \mathbf{Exp}[Y] && \text{Definition of Expectation}
\end{aligned}$$

□

As a corollary, by applying the above again and again $k - 1$ times, we get:

Theorem 4. For any k random variables $X_1, X_2, \ldots, X_k$,

$$\mathbf{Exp}\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} \mathbf{Exp}[X_i]$$

Examples of applications.

(a) *We roll two fair dice. X takes the value of the sum. In this case, $X = Y + Z$ where Y, Z are random variables of the kind from the previous bullet point.*
Tailor-made application. $\mathbf{Exp}[Y] = \mathbf{Exp}[Z] = 3.5$, the expected value of a single roll of a die. Thus, $\mathbf{Exp}[X] = \mathbf{Exp}[Y + Z] = 7$ by linearity of expectation.

(b) *We have a biased coin which lands heads with probability p. We toss it 100 times. Let Z be the number of heads we see. What is $\mathbf{Exp}[Z]$? Note that earlier we had the question for $p = 0.5$.*

> **Remark:** *Try doing this the "first-principle" way. That is, for each $0 \le k \le 100$, figure out the probability $\mathbf{Pr}[X = k]$ (that is, the probability we get exactly k heads), and then sum $\sum_{k=0}^{100} k \cdot \mathbf{Pr}[X = k]$. Please try it; feel the sweat needed to do this. It will make you appreciate the next three lines more!*

Define new random variables; define X_i to take the value 1 if the ith toss is heads, and 0 otherwise. Note, $X = X_1 + X_2 + \cdots + X_{100}$. Note, $\mathbf{Exp}[X_i] = p$ (it is a Bernoulli random variable). Thus, linearity of expectation gives $\mathbf{Exp}[X] = 100p$.

(c) *n people checked in their hats, but on their way out, were handed back hats randomly. What is the expected number of people who get their correct hats?*

Define X_i to be 1 if the ith person gets his or her back correctly. What is $\mathbf{Exp}[X_i]$? It is $1/n$; it is the probability that $\sigma(i) = i$ for a random ordering σ. This question was there in the UGP. Let $Z = \sum_{i=1}^{n} X_i$. Note, Z is the number of people who get their correct hats. By linearity of expectation, $\mathbf{Exp}[Z] = 1$.

(d) *In a party of n people there are some pairs of people who are friends, and some pairs who are not. In all there are m pairs of friends. The host randomly divides the party by taking each person and sending them left or right at the toss of a fair coin. How many friends, in expectation, are sundered apart?*

> **Remark:** *In terms of graphs (which we will see soon) the question is: a graph with m edges is randomly partitioned. How many edges, in expectation, have endpoints in different parts?*

For each pair of friends (u, v), define X_{uv} which takes the value 1 if u and v are split, and takes the value 0 if u and v are not split. The probability u and v are split is $1/2$ (either u is sent left, v is sent right, or vice-versa – do you see this?). Thus, $\mathbf{Exp}[X_{uv}] = 1/2$. Define $Z = \sum_{(u,v):\text{ friends}} X_{uv}$; Z is the number of friends sent apart. $\mathbf{Exp}[Z] = \sum_{(u,v):\text{ friends}} \mathbf{Exp}[X_{uv}] = m/2$. In expectation, half the friendships are sundered apart.

(e) *In an ordering σ of $(1, 2, \ldots, n)$, an inversion is a pair $i < j$ such that $\sigma(i) > \sigma(j)$. How many inversions, in expectation, are there in a random permutation?*

Let σ be a random permutation. Define the *indicator random variable* X_{ij} for $i < j$, which takes the value 1 if $\sigma(i) > \sigma(j)$, and 0 otherwise. Note that $\mathbf{Pr}[X_{ij} = 1] = \frac{1}{2}$; there are equally many orderings with $\sigma(i) > \sigma(j)$ as $\sigma(i) < \sigma(j)$. Now note that $Z = \sum_{i=1}^{n} \sum_{j>i} X_{ij}$ is the number of inversions in σ. Thus, $\mathbf{Exp}[Z] = \sum_{i=1}^{n} \sum_{j>n} \mathbf{Exp}[X_{ij}] = \frac{1}{2} \cdot \frac{n(n-1)}{2}$.