CS 30: Discrete Math in CS (Winter 2019): Lecture 24

Date: 20th February, 2019 (Wednesday)

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 random variables of the kind from the previous bullet point.
   - 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. **Expectation of a Random Variable.**

   Theorem 1. The expectation of a random variable X is defined to be

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

   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 toss a fair coin.* $X(\textit{heads}) = 0$ *and* $X(\textit{tails}) = 1$. *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?

- *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* $\mathbf{Exp}[Z]$*?*

4. **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.*

5. **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 2. 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, we get:

Theorem 3. 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 X be the number of heads we see. What is $\mathbf{Exp}[X]$?*

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. 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 are sent apart (in expectation)?*

> **Remark:** *A graph is randomly split into two. 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). 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, $\mathbf{Exp}[Z] = \sum_{i=1}^{n} \sum_{j>n} \mathbf{Exp}[X_{ij}] = \frac{1}{2} \cdot \frac{n(n-1)}{2}$.