E2:202 Random Processes **Aug 2020**

Tutorial 4: Expectations, Some Cool Examples, and Some Inequalities

Lecturer: Parimal Parag *TA: Arvind* *Scribes: Krishna Chaythanya KV*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

4.1 Great Expectations

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a given probability space.

Recall that for an $\mathcal{F}$-measurable RV X, its expectation is given by

$$\mathbb{E}[X] = \int_{x\in\mathbb{R}} x dF_X(x).$$

In particular, if X is a simple RV, that takes values in a finite set $\mathfrak{X} \subseteq \mathbb{R}$, its expectation is given by

$$\mathbb{E}[X] = \sum_{x\in\mathfrak{X}} x\mathbb{P}_X(x).$$

Remark. 1. If $X \sim \text{Ber}(p)$, $\mathbb{E}[X] = p = \mathbb{P}_X(1)$.

2. If $X \sim \text{Bin}(n,p)$, then $X = \sum_{i=1}^n Y_i$, where $Y_i \overset{iid}{\sim} \text{Ber}(p)$. Hence, using the linearity of expectations, $\mathbb{E}[X] = n\mathbb{E}[Y] = np$.

Exercise 4.1. 1. Find the expectation of the $\text{Exp}(\lambda)$ distribution, with $\lambda > 0$.

2. Prove that for $X \sim \text{Geo}(p)$, for $p \in (0,1)$, $\mathbb{E}[X] = \frac{1}{p}$. Hence, find the quantity $H(X) \triangleq \mathbb{E}[-\log \mathbb{P}_X(x)]$, where $X \sim \text{Geo}(p)$.
Remark: The quantity $H(X)$ is called the *entropy* of the RV X.

We shall now consider some corner cases where the expectation of a RV may be infinite or may not even exist.

Example. Recall that $\mathbb{E}[X]$ is undefined when $\mathbb{E}[X^+] = \mathbb{E}[X^-] = \infty$.
Consider the RV $X : \Omega \to \mathbb{Z} \setminus \{0\}$ with pmf $P_X(\cdot)$ given by

$$P_X(x) = \frac{3}{\pi^2 x^2}, \text{ if } x \in \mathbb{Z} \setminus \{0\}.$$

Note: The sequence $\left\{\frac{1}{n^2}\right\}_{n\geq 1}$ is summable, and $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. However the sequence $\left\{\frac{1}{n}\right\}_{n\geq 1}$ is <u>not</u> summable, i.e., $\sum_{n\geq 1} \frac{1}{n} = \infty$. **Can you provide an argument to show that this sum is ∞? (Compare this sum with the integral $\int_1^{\infty} \frac{1}{x} dx$)**

Now, $\mathbb{E}[X^+] = \sum_{x\geq 1} \frac{x\cdot 3\cdot}{1\cdot\pi^2\cdot x^2} = \frac{3}{\pi^2}\sum_{x\geq 1}\frac{1}{x} = \infty$, and $\mathbb{E}[X^-] = \infty$ similarly. Hence, $\mathbb{E}[X]$ does not exist!

Exercise 4.2. Modify the argument above to construct a RV X whose expectation exists, and besides, $\mathbb{E}[X] = \infty$.

We now look at some interesting problems inspired by classical questions in theoretical computer science.

Example (The secretary problem)**.** Suppose an agency is looking to hire secretaries. Assume that $n > 1$ candidates apply. There exists an absolute ordering on the candidates in terms of their skill levels, but the agency is oblivious to this absolute order.

The hiring proceeds as follows:

- The candidates arrive in a random order (all permutations of the n candidates are equally likely)
- At time $i \in \mathbb{N}$, the agency hires the best candidate among the i candidates it has seen up until then.

What is the expected number of candidates hired?

Solution: The trick in such questions is to use the linearity of expectations.

Let X_i be the indicator random variable that denotes whether the i^{th} candidate is hired

$$X_i = \begin{cases} 1, & \text{if candidate } i \text{ is hired,} \\ 0, & \text{o.w.} \end{cases}$$

Clearly, we are interested in $\mathbb{E}[X]$ where $X = \sum_{i=1}^{n} X_i$.

We first note that $\mathbb{P}(X_1 = 1) = \mathbb{E}[X_i] = 1/i, i \in \mathbb{N}$ (To check this, consider all the permutations of i objects that are distant). Hence, $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \frac{1}{i} \approxeq \log n + \mathcal{O}(1)$.

Hence, on average, the agency hires $\log n$ candidates.

Example (Balls-in-Bins or the "Coupon Collector" problem)**.** Consider an experiment where we randomly toss identical balls into b bins numbered $1, 2, \ldots, b$. The tosses are independent, and on each toss, the ball is equally likely to fall in any bin. Hence the probability that a tossed ball lands in a given bin is $1/b$.

1. How many balls on average must we toss until a given bin contains a ball? [**Exercise**]
2. How many balls must we toss until every bin contains at least one ball? [**Solution below**]

Solution: Let us split the tosses into stages: the i^{th} stage consits of tosses after the $(i-1)^{\text{th}}$ "hit" until the i^{th} "hit". Here, a "hit" is a toss in which the ball falls into an empty bin.

Let N_i be the RV corresponding to the number of tosses in the i^{th} stage. Then,

$$\mathbb{E}[N_i] = \frac{b}{b-i+1} \quad \textbf{(Why? Show this.)}$$

We wish to compute $\mathbb{E}[N]$, where $N \triangleq \sum_{i=1}^{b} N_i$. **Can you use the linearity of expectation to do this**?

Further, using the asymptotic result introduced in the previous problem, **can you put down a similar asymptotic result in** b?

Exercise 4.3 (Simple Exercise)**.** I have a packet that I wish to send to a receiver over a noisy medium that drops packets with probability $\epsilon \in (0, 1)$. I try sending this packet at time instants $t = 1, 2, \ldots$, and after each attempt to send the packet, I receive instantaneous feedback from the receiver about whether it received the packet. If the packet has not been received at time t, I send the packet again at time $t + 1$. What is the expected number of times I need to send a single packet for correct reception?

4.2 Some Inequalities

4.2.1 Markov Inequality

In class, we saw the *Markov inequality*, which states the following.

Theorem 4.4 (Markov Inequality)**.** If $X : \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable RV, then for any monotonically non-decreasing function $f : \mathbb{R} \to \mathbb{R}_+$,

$$\mathbb{P}\{X \geq \epsilon\} \leq \frac{\mathbb{E}[f(X)]}{f(\epsilon)}, \text{ for } \epsilon \in (0, \infty).$$

Corollary 4.5. Iif X is a non-negative RV, then,

$$\mathbb{P}\{X \geq x\} \leq \frac{\mathbb{E}[X]}{x} \ \forall x > 0.$$

We shall now use this to show an important property of non-negative RVs.

Lemma. For $X \geq 0$, w.p. 1, if $\mathbb{E}[X] = 0$, then $X = 0$ w.p. 1.

Proof. We intend to show that $\mathbb{P}(\{\omega : X(\omega) = 0\}) = 1$ or equivalently that $\mathbb{P}(\{\omega : X(\omega) > 0\}) = 0$. We know that for any $n \in \mathbb{N}$,

$$\mathbb{P}\left(\left\{\omega : X(\omega) \geq \frac{1}{n}\right\}\right) \leq \frac{\mathbb{E}[X]}{1/n} = 0.$$

Now, using the *continuity of probability*, we get that

$$\mathbb{P}(\{\omega : X(\omega) > 0\}) = \lim_{n\to\infty} \mathbb{P}\left(\left\{\omega : X(\omega) \geq \frac{1}{n}\right\}\right) = 0.$$

□

Exercise 4.6. Let $h : \mathbb{R} \to [0, \alpha]$ be a non-negative bounded function. Show that for $0 \leq a < \alpha$,

$$\mathbb{P}(h(X) \geq a) \geq \frac{\mathbb{E}[h(x)] - a}{\alpha - a}.$$

4.2.2 A Longer Look at the Chernoff Bound

In class, we saw how we can obtain the Chernoff bound as a special case of Markov's inequality. In particular, for any $t \in \mathbb{R}$ and $\lambda > 0$,

$$\begin{aligned}\mathbb{P}(X \geq t) &= \mathbb{P}(\lambda X \geq \lambda t) \\ &= \mathbb{P}\left(e^{\lambda X} \geq e^{\lambda t}\right) \quad \textbf{(Why?)} \\ &\leq \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda t}}. \longrightarrow (*)\end{aligned}$$

Remark 1. The quantity $\mathbb{E}\left[e^{\lambda X}\right]$, as a function of λ for a RV X, is called the moment-generating function (MGF) of X, $M_X(\lambda)$.

Note that, in the inequality $(*)$, the RHS is a function of λ, but the LHS is independent of λ. That is, the inequality is valid for every $\lambda > 0$. Thus to get the "tightest" bound, we can optimize over λ (> 0) to obtain

$$\mathbb{P}\left(X \geq t\right) \leq \inf_{\lambda > 0} e^{-\lambda t} M_X\left(\lambda\right). \longrightarrow (\#)$$

Let us now compute the RHS for the case of a Bernoulli RV.

Example 4.7. Let $X \sim \text{Ber}\left(p\right), p < 1/2$. Further, for $\lambda > 0$, we define

$$\psi_X\left(\lambda\right) \triangleq \log \mathbb{E}\left[e^{\lambda X}\right] = \log M_X\left(\lambda\right).$$

By a simple calculation, we see that $M_X\left(\lambda\right) = pe^{\lambda} + 1 - p$. Hence,

$$\psi_X\left(\lambda\right) = \log\left(pe^{\lambda} + 1 - p\right).$$

Our Chernoff bound then becomes

$$\begin{aligned} \mathbb{P}\left(X \geq t\right) &\leq \min_{\lambda > 0} e^{\left(\psi_X(\lambda) - \lambda t\right)} \quad t \in (0, 1) \\ &= e^{\min\limits_{\lambda > 0}\left(\psi_X(\lambda) - \lambda t\right)} \quad (\textbf{Why?}) \\ &= e^{\min\limits_{\lambda > 0}\left(\log\left(pe^{\lambda} + 1 - p\right) - \lambda t\right)} \\ &= e^{-D(t || p)}, \end{aligned}$$

where,

$$D\left(t \mid\mid p\right) \triangleq (1 - p) \log\left(\frac{1 - t}{1 - p}\right) + t \log \frac{t}{p}$$

is called the binary relative entropy or the binary *Kullback-Leibler Divergence*. Further, it is easy to see that

$$\mathbb{P}\left(X \geq t\right) = \begin{cases} 0, & \text{for } t > 1, \\ 1, & \text{for } t < 0. \end{cases}$$

Exercise 4.8. Compute the RHS of $(\#)$ for $X \sim \text{Poi}\left(\lambda\right)$.