UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Discussion 2
Fall 2023

1. **Numbered Balls**

   A bin contains balls numbered $1, \ldots, n$. You reach in and select $k \in \{1, \ldots, n\}$ balls at random, sampling *without* replacement, so you do not put the balls back into the bin after each draw. Let T be the sum of the numbers on the balls you picked.

   a. If $k = 1$, what is $\mathbb{E}(T)$?

   b. Find $\mathbb{E}(T)$ for any value of $k \in \{1, \ldots, n\}$.

   c. What is $\operatorname{var}(T)$ for any value of $k \in \{1, \ldots, n\}$? You may leave your answer in terms of summations.

   Solution:

   a. We can think of the $k = 1$ case as just picking one ball randomly from the bin. Each of the n balls is equally likely to be selected, so we have

   $$\mathbb{E}(T) = \sum_{i=1}^{n} \frac{i}{n} = \frac{n+1}{2}.$$

   b. Now let T_i be the value of the ith ball picked. We see that

   $$\mathbb{E}(T) = \sum_{i=1}^{k} \mathbb{E}(T_i) = k\,\mathbb{E}(T_i) = \frac{k(n+1)}{2},$$

   as $T_1, \ldots, T_k$ are identically distributed. The symmetry argument is as follows: Instead of thinking about drawing the k balls one by one, think about picking a random permutation of $\{1, \ldots, n\}$ and taking the first k numbers in the permutation as the balls which are drawn. A random permutation has the property that for each position, it is equally likely to be any of the numbers $1, \ldots, n$.

   c. The variance is slightly harder to calculate since the T_i are not independent — we need to find $\mathbb{E}(T^2)$:

   $$\begin{aligned}
   \mathbb{E}(T^2) &= \mathbb{E}\left(\left(\sum_{i=1}^{k} T_i\right)^2\right) \\
   &= k\,\mathbb{E}(T_1^2) + k(k-1)\,\mathbb{E}(T_1 T_2) \\
   &= \frac{k}{n} \sum_{i=1}^{n} j^2 + \frac{k(k-1)}{n(n-1)} \sum_{i \neq j} ij \\
   &= \frac{k(n+1)(2n+1)}{6} + \frac{k(k-1)}{n(n-1)} \sum_{i \neq j} ij.
   \end{aligned}$$

We note that

$$\sum_{i \neq j} ij = \sum_{i,j} ij - \sum_{i=1}^{n} i^2 = \left(\frac{n(n+1)}{2}\right)^2 - \frac{n(n+1)(2n+1)}{6},$$

so we have

$$k(k-1)\,\mathbb{E}(T_1 T_2) = \frac{k(k-1)}{n(n-1)}\left(\frac{n^2(n+1)^2}{4} - \frac{n(2n+1)(n+1)}{6}\right).$$

Lastly, we find

$$\begin{aligned}
\operatorname{var}(T) &= \mathbb{E}(T^2) - \mathbb{E}(T)^2 \\
&= \frac{k(n+1)(2n+1)}{6} + \frac{k(k-1)}{n(n-1)}\left(\frac{n^2(n+1)^2}{4} - \frac{n(n+1)(2n+1)}{6}\right) - \frac{k^2(n+1)^2}{4}.
\end{aligned}$$

Simplification is not necessary.

2. **Upperclassmen**

You meet two students in the library. Assume that each student is an upperclassman or underclassman with equal probability, and each student takes EECS 126 with probability $\frac{1}{10}$, independent of each other and independent of their class standing. What is the probability that both students are upperclassmen, given at least one of them is an upperclassman currently taking EECS 126?

Solution: We define the following events:

- A is the given event that at least one student is an upperclassman taking EECS 126.
- B is the event that both students are upperclassmen.
- U is the event that at least one student is an upperclassman.
- E is the event that at least one student is taking EECS 126.

We wish to find $\mathbb{P}(B \mid A)$, which we can rewrite by Bayes' rule as

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)}.$$

To find $\mathbb{P}(A \mid B)$, we notice that given B (both students are upperclassmen), the event A can be reduced to the event E that at least one student is taking EECS 126. The condition that at least one student is an upperclassman becomes redundant. Then, as taking EECS 126 is independent of being an upperclassman,

$$\mathbb{P}(A \mid B) = \mathbb{P}(E \mid B) = \mathbb{P}(E) = 1 - \mathbb{P}(E^c) = 1 - \left(\frac{9}{10}\right)^2.$$

The probability of B is straightforward by independence:

$$\mathbb{P}(B) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

Now, to find $\mathbb{P}(A)$, we observe that A^c is the event that none of the two students is an upperclassman taking EECS 126. By the given independences,

$$\begin{aligned}\mathbb{P}(A) = 1 - \mathbb{P}(A^c) &= 1 - (\mathbb{P}(\text{is not an upperclassman taking EECS 126}))^2 \\ &= 1 - \left(1 - \frac{1}{2} \cdot \frac{1}{10}\right)^2.\end{aligned}$$

Simplifying our final expression, we get

$$\mathbb{P}(B \mid A) = \frac{\left(1 - \left(\frac{9}{10}\right)^2\right) \cdot \frac{1}{4}}{1 - \left(\frac{19}{20}\right)^2} = \frac{\frac{19}{400}}{\frac{39}{400}} = \frac{19}{39}.$$

3. **Law of the Unconscious Statistician**

a. Prove the *Law of the Unconscious Statistician* (LOTUS): Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X\colon \Omega \to \mathbb{Z}$ and $F\colon \mathbb{Z} \to \mathbb{Z}$ be random variables. Note that the composition $Y = F(X)\colon \Omega \to \mathbb{Z}$ is another random variable. If $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$, and $\mathbb{E}_{\mathcal{L}_X}$ is expectation with respect to the *law* of X on $\mathbb{Z}$, then

$$\mathbb{E}(F(X)) = \mathbb{E}_{\mathcal{L}_X}(F).$$

You should assume that Ω is **discrete** for the sake of simplicity, although LOTUS holds more generally.

b. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the space of all sequences of independent fair coin tosses. Formulate N, the minimum number of tosses needed until we see heads, as a random variable on Ω.

c. Find $\mathbb{E}(N^2)$.
Hint: By the linearity of expectation, $\mathbb{E}(N^2) = \mathbb{E}(N(N-1)) + \mathbb{E}(N)$. You may use the Law of the Unconscious Statistician from part a, and the following identity:

$$\sum_{k=1}^{\infty} k(k-1)x^{k-2} = \frac{d}{dx}\sum_{k=1}^{\infty} kx^{k-1}.$$

Solution:

a. By the definition of expectation, the left-hand side is equal to

$$\begin{aligned}\mathbb{E}(F(X)) &= \sum_{y\in\mathbb{Z}} y\,\mathbb{P}(F(X) = y)\\ &= \sum_{y\in\mathbb{Z}} y\,\mathbb{P}(X \in \{x : F(x) = y\}).\end{aligned}$$

The $\mathbb{Z}$ above refers to the second $\mathbb{Z}$ in $\Omega \xrightarrow{X} \mathbb{Z} \xrightarrow{f} \mathbb{Z}$. Now, the law $\mathcal{L}_X$ of X is a probability measure on (the first) $\mathbb{Z}$, such that

$$\mathcal{L}_X(B) = \mathbb{P}(X \in B) \quad \text{for } B \subset \mathbb{Z}.$$

So, the above is precisely equal to

$$\begin{aligned}&= \sum_{y\in\mathbb{Z}} y\mathcal{L}_X(\{x : F(x) = y\})\\ &= \sum_{y\in\mathbb{Z}} y\mathcal{L}_X(F = y) = \mathbb{E}_{\mathcal{L}_X}(F).\end{aligned}$$

b. We write each outcome ω as $(\omega_1, \omega_2, \omega_3, \ldots)$, where $\omega_n \in \{H, T\}$ is the result of the nth toss. Then, we define N by

$$N(\omega) = \min\{n \geq 1 : \omega_n = H\}.$$

c. Per the hint, $\mathbb{E}(N^2) = \mathbb{E}(N(N-1)) + \mathbb{E}(N)$. We observe that N is a Geometric random variable with parameter $p = \frac{1}{2}$, which has expected value $\mathbb{E}(N) = 2$. Now, by the Law of the Unconscious Statistician,

$$\begin{aligned}
\mathbb{E}(N(N-1)) &= \sum_{k=1}^{\infty} k(k-1)\,\mathbb{P}(N=k) \\
&= \sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1}
\end{aligned}$$

Pulling out a factor of $p(1-p)$, we can apply the final hint.

$$\begin{aligned}
&= p(1-p)\sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} \\
&= -p(1-p)\frac{d}{dp}\sum_{k=1}^{\infty} k(1-p)^{k-1}
\end{aligned}$$

Note that $\mathbb{E}(N) = \frac{1}{p}$ is, by definition, equal to $p\sum_{k=1}^{\infty} k(1-p)^{k-1}$. Then,

$$\begin{aligned}
&= -p(1-p)\frac{d}{dp}\frac{1}{p^2} \\
&= \frac{2(1-p)}{p^2}.
\end{aligned}$$

For $p = \frac{1}{2}$, we find that $\mathbb{E}(N^2) = 4 + 2 = 6$.

Alternatively, we can observe the following recurrence relation. We remark that the approach above of finding $\mathbb{E}(N(N-1))$ is also applicable when N is *Poisson*, but not the following approach.

$$\begin{aligned}
\mathbb{E}(N^2) = \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} &= 1^2 \cdot p + (1-p)\sum_{k=1}^{\infty}(k+1)^2 p(1-p)^{k-1} \\
&= p + (1-p)\,\mathbb{E}((N+1)^2) \\
&= p + (1-p)(\mathbb{E}(N^2) + 2\,\mathbb{E}(N) + 1).
\end{aligned}$$