3.5 Discrete Uniform

Unit 3: Discrete Random Variables

Adapted from Blitzstein-Hwang Chapter 3.

A very simple story, closely connected to the naive definition of probability, describes picking a random number from some finite set of possibilities.

Story 3.5.1 (Discrete Uniform distribution).

Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the *Discrete Uniform distribution* with parameter C; we denote this by $X \sim \mathrm{DUnif}(C)$. The PMF of $X \sim \mathrm{DUnif}(C)$ is

$$P(X = x) = \frac{1}{|C|}$$

for $x \in C$ (and 0 otherwise), since a PMF must sum to 1. As with questions based on the naive definition of probability, questions based on a Discrete Uniform distribution reduce to counting problems. Specifically, for $X \sim \mathrm{DUnif}(C)$ and any $A \subseteq C$, we have

$$P(X \in A) = \frac{|A|}{|C|}.$$

Example 3.5.2 (Random slips of paper).

There are 100 slips of paper in a hat, each of which has one of the numbers $1, 2, \dots, 100$ written on it, with no number appearing more than once. Five of the slips are drawn, one at a time. *First consider random sampling with replacement (with equal probabilities).*

(a) What is the distribution of how many of the drawn slips have a value of at least 80 written on them?

(b) What is the distribution of the value of the j-th draw (for $1 \leq j \leq 5$)?

(c) What is the probability that the number 100 is drawn at least once?

Now consider random sampling without replacement (with all sets of five slips equally likely to be chosen).

(d) What is the distribution of how many of the drawn slips have a value of at least 80 written on them?

(e) What is the distribution of the value of the j-th draw (for $1 \leq j \leq 5$)?

(f) What is the probability that the number 100 is drawn at least once?

Solution

(a) By the story of the Binomial, the distribution is $\textbf{Bin}(5, 0.21)$.

(b) Let X_j be the value of the jth draw. By symmetry, $X_j \sim \textbf{DUnif}(1, 2, \ldots, 100)$.

(c) Taking complements,

$$P(X_j = 100 \textbf{ for at least one } j) = 1 - P(X_1 \neq 100, \ldots, X_5 \neq 100).$$

By the naive definition of probability, this is

$$1 - (99/100)^5 \approx 0.049.$$

This solution just uses new notation for concepts from Unit 1. It is useful to have this new notation since it is compact and flexible. In the above calculation, it is important to see why

$$P(X_1 \neq 100, \ldots, X_5 \neq 100) = P(X_1 \neq 100) \ldots P(X_5 \neq 100).$$

This follows from the naive definition in this case, but a more general way to think about such statements is through *independence* of r.v.s, a concept discussed in detail later in this unit.

(d) By the story of the Hypergeometric, the distribution is $\textbf{HGeom}(21, 79, 5)$.

(e) Let Y_j be the value of the jth draw. By symmetry, $Y_j \sim \textbf{DUnif}(1, 2, \ldots, 100)$. Here learning any Y_i gives information about the other values (so $Y_1, \ldots, Y_5$ are *not* independent, as defined in Definition 3.8.1), but symmetry still holds since, unconditionally, the jth slip drawn is equally likely to be any of the slips.

(f) The events $Y_1 = 100, \ldots, Y_5 = 100$ are disjoint since we are now sampling without replacement, so

$$P(Y_j = 100 \textbf{ for some } j) = P(Y_1 = 100) + \cdots + P(Y_5 = 100) = 0.05.$$