LECTURE 5: MORE PROBABILITY DISTRIBUTIONS

AP STATS

Abstract. We review uniform distributions and cumulative distributions. We then define a "z-score" of a point in a distribution.

- If X is a discrete variable, then we call f a **probability mass function** (PMF)
- If X is a continuous variable, then we call f a **probability density function** (PDF)

Question 1. Suppose we have $X = \{1, 2, 5\}$ and suppose $p(x) = kx$ for some constant k. What is the value of k if we want p to be a probability mass function on X?

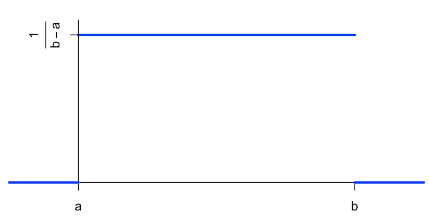
Solution. Rewriting X as $X = \{1\} \cup \{2\} \cup \{5\}$, from Additivity we get that

$$\begin{aligned} p(X) &= p(\{1,2,5\}) = p(\{1\} \cup \{2\} \cup \{5\}) \\ &= p(\{1\}) + p(\{2\}) + p(\{5\}) && \text{using additivity} \\ &= k(1) + k(2) + k(5) && \text{plugging } x \text{ into } p(x) \\ &= 8k \end{aligned}$$

Using the fact that $p(X) = 1$ from Property (2) Normalization, we conclude that $p(X) = 8k = 1 \Longrightarrow k = \dfrac{1}{8}$.

1. Uniform Distributions

A **uniform distribution** is a distribution where all of the data points are *uniformly distributed* (are equal). Graphically, a uniform distribution is a flat line:



That is, the probability function is a nonzero constant on X and is 0 outside of X. However, since coming to a flat line in life is rare, we use "roughly uniform" to describe distributions that are *almost* uniform.

Example 1. Suppose $X = [2, 2023]$ and let $f(x) = c$ for some constant c. If we want f to be the uniform probability density function on $X = [2, 2023]$, we use the fact that $f(X) = 1$.

Recalling that $f(X) =$area under f, and since $f(x) = c$ is a flat line at height c, the area under f is a box with height c and width $2023 - 2 = 2021$. Thus, we

$$c2021 = f(X) = 1 \Longrightarrow c = \frac{1}{2021}$$

Example 2. Let $X = \{1, 2, 3, 4, 5, 6\}$ be the random variable for a "fair die" (fair $\Longrightarrow$ uniform probability). So if $f(x) = c$ is the uniform probability distribution function on X, we get that $c = \dfrac{1}{6}$.

2. Cumulative Distribution Functions

Given a random variable X and a PMF/PDF f defined on X, we can consider *another distribution* called the **cumulative distribution function**, which sums all of the probabilities up to a given point.

Definition. The **cumulative distribution function** (CDF) is the function defined as

$$CDF(x) = \int_{t=-\infty}^{x} f(t), \quad f = PDF/PMF$$

The cumulative distribution function has the following properties:

- Domain = all real numbers $\mathbb{R}$
- Range of outputs = $[0, 1]$
- CDF is *monotonically increasing* (if $x \leq y \Longrightarrow CDF(x) \leq CDF(y)$)

Example 1. From **Question 1**, we have $X = \{1, 2, 5\}$ and $p(x) = \dfrac{1}{8}$ as our probability function.

Let's find the values of the CDF:

x	1	2	5
$p(x)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{5}{8}$
$CDF(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{8}{8}$

Note that we set the $CDF(x) = 0$ for any $x < 1$ in this example so that the CDF starts life at 0 and then concludes its final output to be $8/8 = 1$.

Example 2. If we set $X = [2, 2023]$ and let $f(x) = \dfrac{1}{2021}$ to be our (uniform) probability function, then we can evaluate the CDF at any given point between 2 and 2023. So if we were interested in $x = 2020$, then

$$CDF(x = 2020) = [\text{area from } x = 2 \text{ to } x = 2020] = \frac{2018}{2021}$$

3. Z-Scores

Definition. Given a quantitative variable, the **z-score** of a point x is the number

$$z_x = \frac{x - \mu}{\sigma}, \quad \mu = \text{mean}, \;\; \sigma = \text{standard dev}$$

The z-score tells you how many standard deviations you are away from the mean μ.

Remark. You can always compute the z-score of a point given μ and σ.

The only reason why we care about these z-scores is that they give us percentile interpretations when our distribution is *normal*.

Example 1. Suppose $\mu = 70$ years and $\sigma = 5$ years. Say that your tortoise lifespan was in the 2.3 standard deviations above the mean. How long was your tortoise lifespan?

We write $z_x = 2.3$ and then do the computations:

$$2.3 = z_x = \frac{x - 70}{5} \Longrightarrow x = 70 + 2.3(5) = 81.5 \text{ years}$$