LECTURE 4: PROBABILITY DISTRIBUTIONS

AP STATS

Recall that we have the conversion formulas for percentiles and enumeration (ordering by numbers):

To go from Enumeration to Percentiles (given n, find p-percentile):

(1) Order the list. If k is the object we are interested in, mark the position of k as n (this is the ordinal)
(2) k is in the pth percentile where

$$\boxed{p = \frac{100}{N} \times n, \quad N = \text{size of } X}$$

To go from Percentiles to Ordinals (given p-percentile, find n in the formula above):

(1) Order the list.
(2) If k is in the nth position of the list, then

$$\boxed{n = \lceil \frac{p}{100} \times N \rceil, \quad N = \text{size of } X}$$

(3) Your answer is k.

See Lecture 3 for example computations.

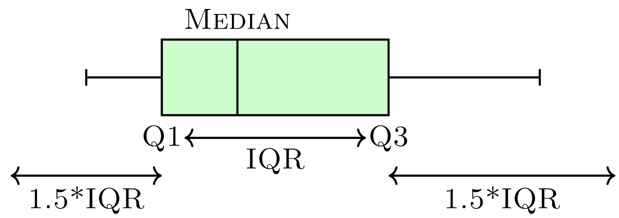
1. Five Point Summary

The five point summary gives a description of 5 points of the data set:

(1) min
(2) Q_1 25th percentile
(3) median (50th percentile)
(4) Q_3 75th percentile
(5) max

We usually use box and whisker plots to deescribe these pieces of data:

Box and Whisker Plot



2. Probability Distributions

Let's recall the definition of probability measures:

Definition. Given a random variable X, we can create a function f. We call the function f a **probability measure on** X if f satisfies the 3 properties:

(1) Non-negativity $f(A) \geq 0$ for all $A \subseteq X$ $\leftrightsquigarrow$ "you can never have negative probabilities"
(2) Normalization $f(X) = 1$ $\leftrightsquigarrow$ "the probability of obtaining the everything possible is 1"
(3) Additivity If $A, B \subseteq X$ and $A \cap B = \emptyset$, then $f(A \cup B) = f(A) + f(B)$ $\leftrightsquigarrow$ "prob(this or that) = prob(this) + prob(that)...as long as this and that can't both happen at the same time"

- 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 $X = \{1, 3, 5\}$ and suppose $p(x) = kx^2$ 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 $X = \{1\} \cup \{3\} \cup \{5\}$, from Additivity we get that

$$\begin{aligned} p(X) = p(\{1,3,5\}) &= p(\{1\} \cup \{3\} \cup \{5\}) \\ &= p(\{1\}) + p(\{3\}) + p(\{5\}) && \text{using additivity} \\ &= k(1)^2 + k(3)^2 + k(5)^2 && \text{plugging } x \text{ into } p(x) \\ &= k + 9k + 25k = 35k \end{aligned}$$

Using the fact that $p(X) = 1$ from Property (2) Normalization, we conclude that $p(X) = 35k = 1 \Longrightarrow k = \dfrac{1}{35}$.

2.1. Probabilities.

The purpose of creating the PMF/PDF is so that we can compute probabilities of obtaining values in a data set. That is, a PMF is constructed to satisfy the equation

$$Prob(a \leq X \leq b) = \sum_{t=a}^{b} f(t) = f(a) + f(a+1) + f(a+2) + \cdots + f(b-1) + f(b), \quad f = PMF$$

and similarly, a PDF is constructed to satisfy the equation

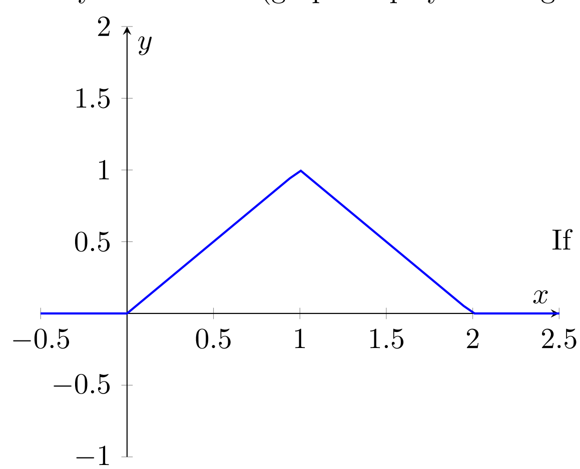
$$Prob(a \leq X \leq b) = \int_a^b f(t) = [\text{area of shape between } X = a \text{ and } X = b], \quad f = PDF$$

Example 1. Let $X = \{1, 3, 5\}$. Assuming uniform probability $f(x) = \frac{1}{3}$ for each $x \in X$, compute the two following probabilities:

(1) $Prob(1 \leq X \leq 3) = 2/3$
(2) $Prob(1 \leq X < 3) = 1/3$
(3) $Prob(3 \leq X < 5) = 1/3$
(4) $Prob(1 \leq X < 5) = 2/3$
(5) $Prob(1 \leq X \leq 5) = 1$

Example 2. Suppose we have the our random variable X takes on values on the real line $\mathbb{R}$ and suppose we were given a PDF $p(x)$ determined by the formula (graph displayed on right)

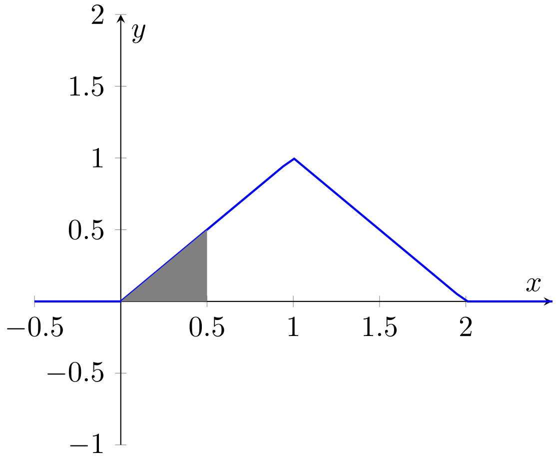
$$p(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1 \\ 2 - x & \text{if } 1 < x \leq 2 \\ 0 & \text{else} \end{cases}$$



If we were trying to compute $Prob(0 \leq X < \frac{1}{2})$. To compute this, since $p(x)$ is our PDF for the continuous random variable X,

$$\int_0^{\frac{1}{2}} p(x) = \text{area of shape between } x = 0, x = \frac{1}{2}$$

so we need to find the area of the shaded region



Since the shape given is a right triangle, we can compute the area:

$$\int_0^{1/2} p(x) = \frac{1}{2}(base)(height) = \frac{1}{2}(\frac{1}{2} \cdot \frac{1}{2}) = \frac{1}{8}.$$

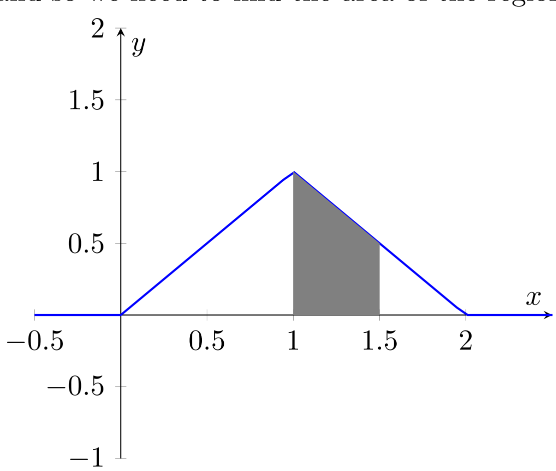
So we conclude that

$$Prob(0 \leq X < 1/2) = \boxed{\frac{1}{8}}$$

Example 3. Using the same probability distribution function above, let's compute $Prob(1 \leq X < 1.5)$:

$$Prob(1 \leq X < 1.5) = \int_1^{1.5} p(x) = \text{area between } x = 1 \text{ and } x = 1.5$$

and so we need to find the area of the region shaded below:



and since the shape is the sum of a rectangle and a triangle, we compute the area as:

$$\int_1^{1.5} p(x) = (rectanglearea) + (trianglearea)$$

$$
\begin{aligned}
&= \left(\underbrace{(1.5-1)}_{\text{base}} \times \underbrace{0.5}_{\text{height}} \right) + \frac{1}{2}(0.5 \times 0.5) \\
&= 0.25 + 0.125 \\
&= 0.375
\end{aligned}
$$

So $Prob(1 \leq X < 1.5) = 0.375$.

Remark. Note that **for continuous variables** that

$$P(a < X < b) = P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b)$$

since $P(X = \text{a fixed number}) = 0$.

For example, if X =weight of a student at VCHS, then $Prob(X = 120.0\ lbs) = 0$ since no student will be *exactly* 120 pounds- so instead we use a range of numbers (e.g. anyone who weights between $120.0 \leq X < 121.0$). Click HERE for a deeper discussion about this.