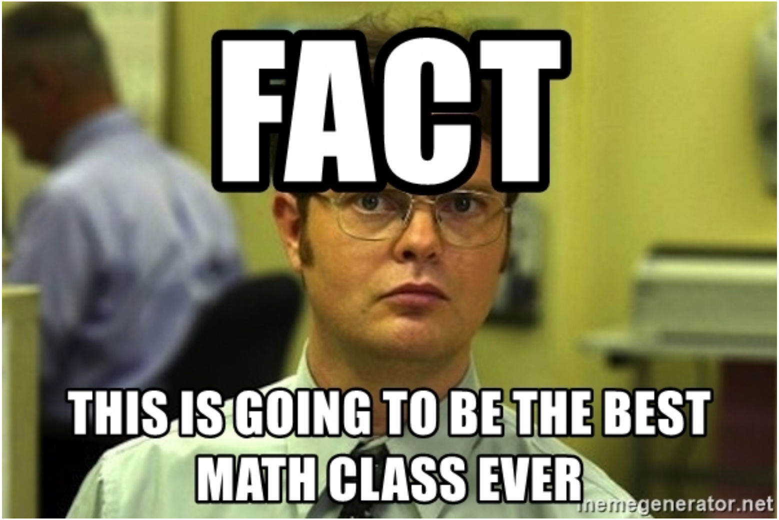
STAT 230 SECTION 2 LECTURE 10



Today's Agenda

Last time:

- Law of total probability
- Bayes Rule

Today (Lec 10, 05/25):

- Another example for conditional probabilities
- Discrete random variables (Chapter 5!!!)

Example

In an insurance portfolio 10% of the policy holders are in Class A1 (high risk), 40% are in Class A2 (medium risk), and 50% are in Class A3 (low risk). The probability there is a claim on a Class A1 policy in a given year is 0.10; similar probabilities for Classes A2 and A3 are 0.05 and 0.02. Find the probability that if a claim is made, it is made on a Class A1 policy.

5. DISCRETE RANDOM VARIABLES

Definition

A **random variable** is a function that maps from the sample space $S$to the set of real numbers $\mathbb{R}$. In other words, we say X is a random variable if

$$X : S \to \mathbb{R}.$$

Often "random variable" is abbreviated with "RV".

Definition

The values that a random variable takes is called the **range** of the random variable. We often denote the range of a random variable X by $X(S)$.

Why do we need random variables if we can already assign probability to events directly?

By assigning numerical values to events, we can manipulate them using familiar mathematical tools such as (linear) algebra and calculus.

As we will see later in the course, we've just enabled ourselves to model A LOT MORE situations with random variables.

Definition

- We say that a random variable is **discrete** if its range is a discrete subset of $\mathbb{R}$ (i.e., a finite or a countably infinite set).
- A random variable is **continuous** if its range is an interval that is a subset of $\mathbb{R}$ (e.g. $[0, 1]$, $(0, \infty)$, $\mathbb{R}$).

Remark:

- one can define rvs whose range is a mixture of discrete and continuous parts $\Rightarrow$ not in STAT 230
- In this chapter, we focus on discrete random variables.

Examples

Experiment	X	Range
Flip a coin	# of Heads	$\{0, 1\}$
Roll two 6 sided die	sum of die rolls	$\{2, ..., 12\}$
Monitor call center traffic	# of calls to center	$\{0, 1, 2, ...\}$
Measure distance to black hole	distance measured	$(10kpc, 300kpc)$
Measure time to solve the midterm	time measured	$(0min, 80min)$

Definition

The **probability (mass) function** of a discrete random variable X is the function

$$f_X(x) = P(X = x).$$

We use the short hand $P(X = x)$ to denote $P(\{\omega \in S : X(\omega) = x\})$. The definition f is valid for all x, but its value is zero when x is outside the range of random variable X.

Example

Suppose that X is the sum of the outcomes of a two fair six sided dice roll. Calculate the probability function of X

Properties of the pf f

A probability function has the properties that:

a)

$$0 \leq f_X(x) \leq \qquad \text{for all } x$$

b)

$$\sum_{x \in X(S)} f_X(x) = 1$$

Example

Find the value k which makes the function f given by

$$f(0) = 0.1, \quad f(1) = k, \quad f(2) = 3k, \quad f(3) = 0.3$$

and 0 elsewhere a valid probability function.

Example

A fair coin is tossed three times with each toss being independent. Let $X =$ number of heads occurring. Compute the probability function of X.

Now we have a way to assign probability to specific values of a random variable.

However, we often encounter situations like “at least” or “at most”. What do we do then?

Definition

The **cumulative distribution function** (cdf) of a random variable X is

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}.$$

Again we use the short hand notation

$$P(X \leq x) = P(\{\omega \in S : X(\omega) \leq x\}).$$

If X is discrete with probability function f_X, then

$$F_X(x) = P(X \leq x) = \sum_{y \colon y \leq x} f_X(y)$$

Properties of the cdf

Let F_X be a cdf. Then,

a) $0 \leq F_X(x) \leq 1$

b) $F_X(x) \leq F_X(y)$ for $x < y$

c) $\lim_{x \to -\infty} F_X(x) = 0$, and $\lim_{x \to \infty} F_X(x) = 1$.

Example

Suppose a fair six sided die is rolled, and let X denote the number of dots on the top face of the die. Compute and graph $F_X(x)$.