Stat 230: Probability

Lecture 8

Jeremy VanderDoes

University of Waterloo

Wednesday, May 17th

Review

Example

A groups of CS students are interviewed–1% of them don't know how to code. The interviewers learned 94% do poorly in coding exams if they don't know how to code. However, these tests are difficult so only 42% of students who know how to code do well in the exams.

(1) Find probability a student does well on the exam

(2) Given a student does well on the exam, what is the probability they know how to code?

(3) Given a student does not do well on the exam, what is the probability they know do not know how to code?

(4) What is the probability a student does not know how to code given they did well on the exam? How about the probability the student knows how to code given they did poorly on the exam?

Review

Last time we talked about:

(1) Law of total probability

(2) Bayes theorem

For today:

(1) Random variables

(2) Probability function

(3) Cummulative distribution function

Review

- Review question
- Office hours
- Quiz review

Mathematical Results

Some basic mathematical results:

(1) Geometric Series Formula:

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r}, \quad if \quad |r| < 1.$$

(2) Partial Geometric Series:

$$\sum_{i=0}^{k} r^i = \frac{1 - r^{k+1}}{1-r}$$

Mathematical Results

(3) Binomial Theorem:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$$

(4) Taylor Series for the Exponential Function:

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

Definitions

Definition

A **random variable** *is a function that maps the sample space S into 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 possible values for a random variable are called the **range** *of the random variable. We often denote the range of a random variable X by $X(S)$.*

Definitions

Definition

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

One could imagine defining random variables whose range is a mixture of discrete and continuous parts. For the sake of simplicity, we will not consider such cases in this class.

Definitions

Example

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	$(10^3 pc, 30^3 pc)$
Measure time to catch pokemon	time measured	$(0 sec, 15 sec)$

Definitions

Example

Consider the rand() function in C++ (or the random.random() function in python). If the random variable X denotes the output of the function, what kind of random variable is X?

Definitions

In "real life" we cannot actually observe continuous random variables, so in that sense every random variable derived from "real life" situations is discrete.

However, in many instances the range of a random variable is so large and dense that is makes more sense to model it as a continuum than as a discrete set.

Probability Function

Definition

The **probability 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 it is only different from zero when* x *is in the range of* X.

Probability Function

Example

Suppose that X is the sum of the outcomes of two fair six sided die rolls. Calculate the probability function of X.

Probability Function

A probability function has the properties:

(1) $0 \leq f_X(x) \leq 1$

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

Cumulative Distribution Function

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

When X is discrete:

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

Cumulative Distribution Function

The CDF satisfies:

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

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

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

Cumulative Distribution Function

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 $F_X(x)$.