When you go to school on Monday but 230 was cancelled

Today's Agenda

Last time:

- Cumulative distribution function
- Discrete uniform distribution

Today (Lec 12, 06/01):

- Hypergeometric Distribution
- Bernoulli Distribution
- Binomial Distribution

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 $X(S)$.

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. $[0, 1], (0, \infty), \mathbb{R}$).

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.

Any probability function f satisfied

a)

$$0 \leq f_X(x) \leq 1 \quad \text{for all } x$$

b)

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

Definition

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

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

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

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

If X takes values $0, 1, 2, \ldots$, we find

$$P(X = x) = P(X \leq x) - P(X \leq x - 1) = F(x) - F(x - 1),$$

for $x \in \{0, 1, 2, \ldots\}$.

Discrete Uniform Distribution

Definition

Suppose the range of X is $\{a, a+1, \ldots, b\}$, where $a, b \in \mathbb{Z}$, and suppose all values are equally likely. Then we say that X has a **discrete uniform distribution** on $\{a, a+1, \ldots, b\}$, shorthand: $X \sim U[a, b]$.

If $X \sim U[a, b]$, then its probability function is defined as

$$P(X = x) = \frac{1}{b - a + 1}, \quad \in \{a, a+1, \ldots, b\},$$

and if $x \notin [a, b]$, then $P(X = x) = 0$.

Examples of discrete uniform random variables

Experiment	X	Distribution
Roll a 6 sided die	# showing on die	$U[1, 6]$
Draw a number between 1 and 50	# Drawn	$U[1, 50]$

Hypergeometric distribution

Definition

- Consider a population that consists of N objects, of which r are considered "successes" and the remaining $N - r$ are considered "failures",
- Suppose that a subset of size n (with $n \leq N$) is drawn from the population without replacement.
- Let X=Number of successes obtained, then we say X follows a **hypergeometric distribution** with parameters (N, r, n).
- We sometimes write $X \sim hyp(N, r, n)$ or $X \sim HG(N, r, n)$.

Examples of hypergeometric random variables

Experiment	X	Distribution
Drawing 5 cards from a deck of cards	$\#$ of Ace's	$hyp(52, 4, 5)$
Lotto where 7 numbers are drawn from 50	$\#$ Matches	$hyp(50, 7, 7)$

Probability function

- Total number of arrangements: $\binom{N}{n}$.
- Number of ways to select x successes out of r successes: $\binom{r}{x}$.
- Number of ways to choose remaining $n-x$ failures from $N-r$ failures: $\binom{N-r}{n-x}$.
- Thus,

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},$$

  where

$$x \geq \max\{0, n-(N-r)\}$$

  (must have at least $n-(N-r)$ successes when there are more trials than failures)

$$x \leq \min\{r, n\}$$

  (cannot have more successes than there are (r) or more successes than trials (n)).

(Don't forget that we assign $f_X(x) = 0$ to values of x outside the defined range!)

Example

Consider drawing a 5 card hand at random from a standard 52 card deck.

1. What is the probability that the hand contains at least 3 Kings?
2. What is the probability that the hand contains 1 or fewer Aces?

Many real-life scenarios can be modelled with binary responses.

- “success” vs “failure”
- “old” vs “not old”
- “fried chicken” vs “not fried chicken”
- etc.

Definition

A **Bernoulli trial** with probability of success p is an experiment that results in either a success or failure, and the probability of success is p.

In other words, a Bernoulli trial models "either ... or". Moreover, we can build a distribution from it!

Definition

If a random variable X represents the number of successes in a Bernoulli trial with probability of success p, it follows the **Bernoulli distribution**, and we denote it as

$$X \sim Bernoulli(p),$$

where $X \in \{0, 1\}$.

Figure: Jacob Bernoulli, the dude who discovered Bernoulli distribution.

Examples of Bernoulli trials are

- A coin toss
- A single card draw
- etc.

Question: What do you think the probability function of a Bernoulli random variable X is? i.e: What is $P(X = x)$ if $X \sim Bernoulli(p)$?

...Alright, Bernoulli is cool, but one trial doesn't sound useful enough for anything.

Binomial Distribution

Definition

- Assume an experiment has two possible outcomes, "S" (success) and "F" (failure).
- Assume that in every try, $P(S) = p$ and $P(\mathit{failure}) = 1 - p$ for $0 < p < 1$.
- Repeat the experiment n times independently and record the number of successes obtained. Then X has a **binomial distribution**, denoted by $X \sim Bin(n, p)$.

Using previous techniques, we find

$$f(x) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

for $x \in \{0, 1, 2, \ldots, n\}$.

Examples of binomial random variables:

a) A Bernoulli(p) random variable is a Bin($1, p$) random variable.

b) Flip a fair coin independently 20 times, and let X denote the number of heads observed. Then

$$X \sim Bin(20, 0.5)$$

c) Consider drawing numbers from 1-9 **with replacement** to form a 5 number sequence, and let X denote the number of odd digits in the sequence. Then

$$X \sim Bin(5, 5/9)$$