DATA 252/DATA 551
Modeling and Simulation

Lecture 2: Common Discrete Distributions

Also: primer on R; ~~review of loops~~

January 27, 2020

R, RStudio, and RStudio Cloud

Common Discrete Distributions

Getting started

- We'll use RStudio, an integrated development environment (IDE) for R.
- Instead of running RStudio on your local computer, we can also use RStudio Cloud; it runs RStudio in a browser and does not require any downloading or installation.
- RStudio Cloud is also how I will be sharing sample R code in this class from now on.
- If you want to use R in the long run, you can download R and RStudio after class (note: RStudio requires R to be separately installed). Ask me or a classmate if you have any questions regarding how to set things up.

RStudio Cloud

Go to `https://rstudio.cloud/project/870063`. Choose your sign up option and sign in.

- This is where you can find all of the shared code for this class; the link has been posted on Moodle.
- "Your Workspace" is only accessible by you.
- When you open this link, you'll see a project opened in your workspace. This is a *temporary* copy of the shared project copied to your workspace. You can run or edit the code, but nothing is saved unless you choose to "save a permanent copy." When you do so, you do not change anything on the cloud.
- To get material from RStudio Cloud to your local RStudio, or vice versa, simply copy-paste any code.

Running R code

Two options going forward:

- If you have RStudio: Open RStudio;
- If you do not have RStudio: Click "Your Workspace" on RStudio Cloud and create a new project for this class.

We next demonstrate some basics in R

- Layout of RStudio: An R script is a text file where you can write code; console is where the code is run;
- Running code in console and **running code in R script**;
- Running code by highlighting or line-by-line;
- Basic syntax.

Exercise

Next, we will:
- Briefly go over algorithms in HW1
- Work in groups to finish the exercise questions on the R syntax handout

R, RStudio, and RStudio Cloud

Common Discrete Distributions

A little math...

We will encounter the following math operations:

- exp() and log()
- factorial: $x! = x(x-1)(x-2)\dots 1$
- combination: $\binom{N}{k} = \frac{N!}{(N-k)!k!}$

Exercise: use any programming language, calculate the following

- $\frac{\log(5^{0.7})+0.7^4}{5+0.3^4}$ = 0.2728992
- $\frac{\exp(5)\times(13.5)^4}{10!}$ = 1.358452
- $\binom{10}{x}0.3^x 0.7^{10-x}$ for $x = 1, 2, 3$

$\frac{10!}{(10-x)!\,x!}$

x=1 : 0.1210608
x=2 : 0.2334744
x=3 : 0.2668279

In R : 10!

factorial(10)

dbinom(c(1,2,3), size=10, prob=0.3)

Review

A *random variable* X is discrete if its range is countable (not necessarily finite). Example: $\{0,1\}$ $\{1,2,\cdots,+\infty\}$

The distribution of a discrete random variable can be conveniently specified by its probability mass function (pmf):

$f(x) = P(X=x)$

Example: If a random variable X has range $\{1, 2, \ldots\}$ with pmf

$$f(x) = \frac{\exp^{-x} x^{x-1}}{x!},$$

then: $\cancel{P(X=0) = f(0) = 1}$ $P(X=3) = f(3)$

We next introduce a few *common discrete distributions*, focusing on: 1) the type of scenario that can be modeled by each distribution, 2) its pmf and how to calculate probabilities, and 3) simulating (i.e., drawing random observations) from the distribution.

Discrete uniform

Scenario: X can be any value on a finite set $\{x_1, \ldots, x_k\}$, with equal probability.

Parameter: k is called a **parameter**, which is a constant that you need to specify when setting up a statistical model. k here represents the number of possible values X can take.

Probability mass function: $f(x) = P(X = x) = \frac{1}{k} \quad x \in \{x_1, \cdots, x_k\}$

Simulating from discrete uniform: ex: $X \sim$ discrete uniform $\{1, \cdots, 6\}$

in R:

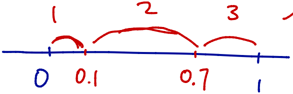
Sample (c(1:6), size = 100, replace = T)

sampling $\{1, \cdots, 6\}$ w/ replacement, equal probability.

Sampling from a finite set of numbers

In general, suppose we want to sample $X \in \{1, 2, 3\}$ according to probabilities $\{0.1, 0.6, 0.3\}$.

- In R, use function `sample`; equivalent functions likely exist in other languages.
- A more universal approach only requires a uniform pseudo-random number generator, which draws a random number between 0 and 1. In R, this is done by `runif(1)`.
- Let **a** be a randomly drawn number between 0 and 1, we can simulate X as follows:
  - Set $X = 1$ if $0 < a \leq 0.1$;
  - Set $X = 2$ if $0.1 < a \leq 0.7$;
  - Set $X = 3$ if $0.7 < a < 1$.



Exercise: Sample a sequence of 1000 $x's$ according to the above distribution; calculate the mean of the 1000 simulated values. You answer should be close to 2.2 (calculated as $1 \times 0.1 + 2 \times 0.6 + 3 \times 0.3$).

do at home

Binomial

Scenario: X represents the number of successes in n independent trials, where each trial has the same success probability p.

Parameter: $n =$ number of trials, $p =$ success probability

Pmf: $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$

$P(X=2) = f(2) \overset{\text{in R}}{=\!=}$ dbinom(2, size=n, prob= p)

Simulating from binomial:

in R: rbinom(100, size=n, prob= p)

make 100 random draws

Example

- Flip a fair coin 10 times; the number of head you get can be modeled as: binomial $n = 10$, $p = 0.5$

- What is the probability that 7 out of 10 people will recover from a certain disease if the probability of recover is 0.8?

\# of people recovered ~ binomial $n = 10$, $p = 0.8$

$P(X=7)$ = dbinom(7, size=10, prob=0.8) = 0.2013

- In a hospital of 15 new born babies, what is the chance that 10 or more babies were boys (assuming the probability of boy is 0.5)?

\# boys ~ binomial $n = 15$, $p = 0.5$

$P(X=10, 11, \cdots, 15)$ = sum(dbinom(c(10:15), size=15, prob=0.5)) = 0.1509