Marginal and Joint Probabilities

Marginal and Joint Probabilities

- Grab a D10 {0, 1, 2, …, 9}
- Consider the following:
    a. Role a dice
    b. If the dice is >7, role again
    c. Record both rolls
    d. Record the total (sum) ← this is your outcome

- Create a table of the joint distribution of the values of each dice roll; Calculate the marginals as well
- Calculate the probability of every outcome
- What is/are the least likely outcome(s)?

Dice 1	Dice 2										
	NA | 0	1	2	3	4	5	6	7	8	9	TOTAL
0		0	0	0	0	0	0	0	0	0	
1		0	0	0	0	0	0	0	0	0	
2		0	0	0	0	0	0	0	0	0	
3		0	0	0	0	0	0	0	0	0	
4		0	0	0	0	0	0	0	0	0	
5		0	0	0	0	0	0	0	0	0	
6		0	0	0	0	0	0	0	0	0	
7		0	0	0	0	0	0	0	0	0	
8											
9											
TOTAL											

Marginal and Joint Probabilities

- Grab a D10 $\{0, 1, 2, \ldots, 9\}$
- Consider the following:
  a. Role a dice
  b. If the dice is >7, role again
  c. Record both rolls
  d. Record the total (sum) ← this is your outcome

- Perform the experiment 5 times and write down your results

How many 8s did we get (as a class)?
How many 3s did we get (as a class)?

Lab Recap

Q&A

After this week's classes, along with the required readings (SPEEGLE Chapter 1; CHIHARA Appendix A; SPEEGLE Chapter 2), you should be able to:

- Execute basic R commands for arithmetic, variable assignment, and installing packages
- Calculate the probability of complex events by applying the fundamental rules of probability
- Explain the concepts of marginal, joint, and conditional probability
- Solve for conditional probabilities like sensitivity, specificity, and positive predictive value in the context of diagnostic testing

PHP 2510 Principles of Biostatistics & Data Analysis

Week 3: Random Variables

This Week's Plan

1

Random Variables

Introduction

2

Discrete RVs

Common Distributions

Practice Problems

3

Continuous RVs

Common Distributions

Practice Problems

4

In Public Health

Examples

OUTCOMES

After this week's classes, along with the required readings (SPEEGLE Chapter 3; SPEEGLE Chapter 4; SPEEGLE Chapter 5, Sections 5.1-5.3), you should be able to:

- Define a random variable and distinguish between discrete and continuous types
- Calculate the expected value and variance using a probability mass function
- Solve for probabilities associated with Binomial, Poisson, and Normal random variables
- Appropriately select and justify probability distribution models for public health variables

Random Variables

What is a random variable?

- Formal definition
- Notation
- Discrete vs Continuous
- PMF/PDF
- CDF

Random Variables

What is an expectation? variance? standard deviation?

Random Variables

Common Random Variables (discrete)

- Bernoulli
- Binomial
- Poisson
- Geometric
- ...

Class Activity

Let's simulate the geometric* distribution together

- Grab a D10 die
- Pick a number (as a class)
- Roll until you get that number
- Write down the number of rolls it took you - 1 {the number of failures before the first success}

What is the event space?
What is the RV?
What does the mapping look like (& what is the support)?

* there are two common parametrizations of the geometric distribution. SPEEGLE and R uses the one that counts failures instead of total trials; CHIRARA counts total trials (1 is the smallest value of the RV)

Identify which (discrete) distribution would be most appropriate and why.

1. A nurse performs a wellness check. W = whether the patient has a fever
2. Officials are tracking the incidence of a measles outbreak. X = the number of new cases reported in Texas
3. A researcher is studying a population where 15% carry a specific genetic mutation. They randomly sample 40 individuals. Y = the number of individuals with the mutation; Z = the number of individuals without the mutation

Practice Problems

Assume an outbreak of a new disease follows a poisson distribution with some incidence = lambda * t (in years). An epidemiologist says that an outbreak happens at least once every 20 years with a 30% chance.

What is lambda?
What is the likelihood of one or more disease outbreaks within the next 3 years?

Practice Problems

~Use Theorem 3.8 in SPEEGLE~
~Let's prove that theorem together~

In an experiment to test whether participants have a certain auditory skills, scientists play sounds and participants say which of the sounds is being played. Participants get 1 pt for each correct answer; ¾ point for the 2 answers that are similar; 0 points for the 9 other answers

- If scientists play 36 sounds, what is the expected score of a participant that randomly guesses?
- Why might we need to calculate the expected value?

Theorem 3.8 does NOT require independence between X and Y

Dice 1	Dice 2									
	NA | 0	1	2	3	4	5	6	7	8	9	TOTAL
0	0.1	0	0	0	0	0	0	0	0	0	0.1
1	0.1	0	0	0	0	0	0	0	0	0	0.1
2	0.1	0	0	0	0	0	0	0	0	0	0.1
3	0.1	0	0	0	0	0	0	0	0	0	0.1
4	0.1	0	0	0	0	0	0	0	0	0	0.1
5	0.1	0	0	0	0	0	0	0	0	0	0.1
6	0.1	0	0	0	0	0	0	0	0	0	0.1
7	0.1	0	0	0	0	0	0	0	0	0	0.1
8	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.1
9	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.1
TOTAL	0.82	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	1

Practice Problems

CHALLENGE!!!

Find the expected number of *different* faces that appear in the first *n* rolls of an S-sided die

Then calculate the number exactly for rolling a 10–sided die rolled 6 times

HINT: again, use the previous result about expectations (which is true even without independence)