

Discrete Random Variables

Dr. Tom Ilvento
FREC 408

Where have we been?

- We started with an introduction to statistics
- and the descriptive nature of statistics for qualitative and quantitative variables
- And the basic ideas and tools of probability

Moving towards

- We want to shift towards inferential properties of statistics
- To do this we need some sense of probability distributions for discrete and continuous random variables

Example – organ transplants

- One problem in organ transplants is the body's rejection of the transplanted tissue
- If the antigens attached to the tissue cells match, the body will accept the transplanted tissue.
- The probability of a match for:
 - Siblings = .25
 - General population = .001
- Let's work out the probabilities of a match for three siblings – assume the probability of a match for one sibling is independent of the probability for another sibling

Sample space and probabilities of a Match for Three Siblings

Sibling 1	Sibling 2	Sibling 3	Probability
Yes	Yes	Yes	$(.25)(.25)(.25) = .0156$
Yes	Yes	No	$(.25)(.25)(.75) = .0469$
Yes	No	Yes	$(.25)(.75)(.25) = .0469$
Yes	No	No	$(.25)(.75)(.75) = .1406$
No	Yes	Yes	$(.75)(.25)(.25) = .0469$
No	Yes	No	$(.75)(.25)(.75) = .1406$
No	No	Yes	$(.75)(.75)(.25) = .1406$
No	No	No	$(.75)(.75)(.75) = .4219$

Or, I could rethink this as:

Number of Matches	Sample Points	p(x) Probabilities
0	No No No	$.75^3 = .422$
1	Yes No No; No Yes No; No No Yes	$(.25)(.75)(.75)*3 = .422$
2	Yes Yes No Yes No Yes No Yes Yes	$(.75)(.75)(.25)*3 = .141$
3	Yes Yes Yes	$.25^3 = .016$

X is a discrete random variable that measures the number of matches in three siblings

Random Variables

- **Random Variables** – variables that assume numerical values associated with random outcomes from an experiment (Def5.1 P238)
 - Discrete
 - Continuous
- Probability Distributions for Random Variables

This will lead us to

- Expected Values (mean) and Variance of Random Variables
- Special Distributions
 - Binomial Distribution
 - Normal Distribution
- Central Limit Theorem
- And into inferences
 - Single sample
 - Two samples

Example of a random variable

- Suppose we were recording the number of dentists that use nitrous oxide (laughing gas) in their practice
- We know that 60% of dentists use the gas.

Nitrous Oxide Example

- Let X = number of dentists in a random sample of **five** dentists that use laughing gas.
- X is a random variable that can take on the following values:
 - 0, 1, 2, 3, 4, 5

Nitrous Oxide Example

- We can list the values of our random variable X

X	0	1	2	3	4	5

Nitrous Oxide Example

- And then assign probabilities to each value of the random variable X

X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

Probability Distribution

- The last table shows the **probability distribution** for the discrete random variable X (Def5.4 P240)
- The table is also referred as probability distribution table
- Properties of probability distribution
 - Each probability is between 0 and 1
 - The sum of the probabilities for all values of x is equal to 1

How we assign probabilities

- $P(0) = (.4)(.4)(.4)(.4)(.4) = .4^5 = .01024$
- $P(1) = (.6)(.4)(.4)(.4)(.4) \times 5 = .0768$
 - 5 is the the number of combinations with only 1 doctor using the gas (use $C_{n,r}$ formula)
 - Yes No No No No
 - No Yes No No No
 - No No Yes No No
 - No No No Yes No
 - No No No No Yes
- $\binom{5}{1} = \frac{5!}{1!(5-1)!} = \frac{120}{24} = 5$
- **Note: we are assuming independence**

Nitrous Oxide Example

X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

What is the probability of 4 of 5 dentists selected randomly Using laughing gas?

$$p(4) = .2592$$

Nitrous Oxide Example

What is the probability of less than 2 of 5 using laughing gas?

$$p(< 2) = p(0) + p(1) =$$

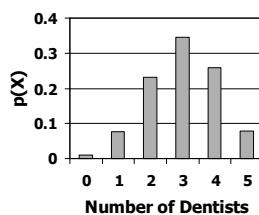
$$p(< 2) = .0102 + .0768 = .0870$$

If I randomly selected 5 dentists, how many would I expect to use laughing gas?

$$\text{Expectation} = 3$$

Nitrous Oxide Example

Probability Distribution of X



Types of Random Variables

- The variable in the dentist example is called a **discrete random variable** (Def5.2 P239)
 - Finite countable number of distinct possible values
 - We can assume that the values can be listed or counted
- Random Variables that fall along points on an interval, and can't be fully counted, are call **Continuous Random Variables** (Def5.3 P239)

How can you tell it is a discrete random variable?

- Countable
- Usually is described as "the number of..."
- Tends to be whole numbers
 - Number of students applying to a university
 - Number of errors on a test
 - Number of bacteria per cubic centimeter of water
 - Number of heart beats of a patient

To describe a discrete random variable

- Specify all the possible values it can assume
- Assign corresponding probabilities to each value

Specifying a discrete random variable – The number of Heads when tossing two coins

- Let X=number of heads observed
 - Heads Heads
 - Heads Tails
 - Tails Heads
 - Tails Tails
- So X takes on the following values
 - 0 TT
 - 1 HT or TH
 - 2 HH

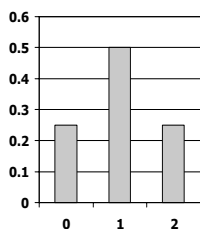
Specifying a discrete random variable – Tossing two coins

- We can a priori assign probabilities to X

Number of Heads	Sample Points	p(x)
0	TT	$.5 * .5 = .25$
1	HT TH	$2(.5 * .5) = .50$
2	HH	$.5 * .5 = .25$

Specifying a discrete random variable – Tossing two coins

- This completely defines the discrete random variable X
- Connecting probabilities to the values results in the probability distribution



The probabilities are distributed over the values

Probability Distribution: p(x)

- Can be shown by a
 - Graph
 - Table
 - Formula – will come later
- Specifies the probability associated with each value
- **Requirements:**
 - $p(x) \geq 0$ for all values of x
 - $\sum p(x) = 1$

Probability Distribution: $p(x)$

- **Notation**
 - $p(x=0) = .25$
 - $p(x=1) = .50$
 - $p(x=2) = .25$
 - $p(x < 2) = .75$

Probability Distributions of Discrete Random Variables

- Sometimes the probabilities are known a priori
- Sometimes they are observed in experiments, a posteriori
- And sometimes we use a formula which seems to apply to a class of discrete random variables, such as the Binomial Distribution

Example - chromosomes

- Males have XY chromosomes
- Females have XX chromosomes
- Assume each are equally likely to be contributed
- **If a couple has three children, what is the probability of at least one male?**

Probability of at least one male in three births

- **Here's our approach**
 - Show the sample points in three births by showing all the possible outcomes in three births
 - Let $X =$ number of male offspring
 - Assign probabilities and depict the probability distribution
 - Sum the probabilities for at least one male

Probability of at least one male in three births

Birth 1	Birth 2	Birth 3
F	M	M
F	M	F
F	F	M
F	F	F
M	M	M
M	M	F
M	F	M
M	F	F

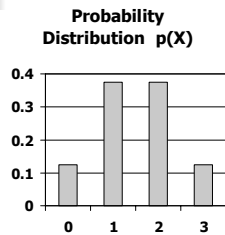
There are:
 $2 \times 2 \times 2 = 8$
 Combinations of males and females in three births

Probability of at least one male in three births

Number of Males	Sample Points	$p(X)$
0	FFF	1/8
1	FMF FFM MFF	3/8
2	FMM MMF MFM	3/8
3	MMM	1/8

Let $X =$ number of male offspring and find its probability distribution

Probability of at least one male in three births



This shows the probability of 0, 1, 2, or 3 males in three births

Probability of one male in three births

- Probability of at least one male in three births
- = $P(1 \text{ male}) + P(2 \text{ males}) + P(3 \text{ males})$
- = $3/8 + 3/8 + 1/8$
- = $7/8 = .875$
- We could have solved this using the compliment –
 - $1 - P(\text{no males}) = 1 - 1/8 = 7/8$

The relationship between observed relative frequency distribution and the population

- If we were to observe a large number of observations of a discrete random variable x (e.g., from a large number of experiments)
- The relative frequency distribution of x
- **Would begin to resemble the probability distribution for the population**

Mean and Variance of a Discrete Variable

- We can think of our observed probability distribution for x as having a mean and variance
- That are identical to the population mean and variance
- For the population
 - μ represents the mean
 - σ^2 represents the variance

Expected Values of Discrete Random Values (Def5.5 P241)

- An expected value is another term for the mean when dealing with a probability distribution
- The expected value of a discrete random variable is

$$E(x) = \sum_{i=1}^n x_i \cdot p(x_i) = \mu$$

The sum of each value times the probability of that value

The Expectation of a Discrete Variable

- I take each value of the discrete variable
- Multiply by the probability associated with the value
- Sum the calculations

$$E(x) = \sum_{i=1}^n x_i \cdot p(x_i) = \mu$$

Expectation of the Male Baby Example

Number of Males	Sample Points	p(X)
0	FFF	1/8 = .125
1	FMF FFM MFF	3/8 = .375
2	FMM MMF MFM	3/8 = .375
3	MMM	1/8 = .125

Expectation of the Male Baby Example

$$E(x) = 0(.125) + 1(.375) + 2(.375) + 3(.125)$$

$$E(x) = 0 + .375 + .750 + .375$$

$$E(x) = 1.5$$

The Variance of a Discrete Variable

(Def5.6 P241)

- The variance of a discrete random variable is given as the
- Expectation of the squared deviations about the population mean:

$$E[(x - \mu)^2] = \sum_{i=1}^n (x_i - \mu)^2 p(x_i) = \sigma^2$$

Sum of the squared deviation of each value from the mean times the probability of the value

The Variance of a Discrete Variable

- I take each value of the discrete variable
- Subtract the mean
- Square the result
- Multiply by the probability associated with the value
- Sum the calculations

$$E[(x - \mu)^2] = \sum_{i=1}^n (x_i - \mu)^2 p(x_i) = \sigma^2$$

Variance of the Male Baby Example

Number of Males	Sample Points	p(X)
0	FFF	1/8 = .125
1	FMF FFM MFF	3/8 = .375
2	FMM MMF MFM	3/8 = .375
3	MMM	1/8 = .125

Variance of the Male Baby Example

$$E(x-1.5)^2 = (0-1.5)^2(.125) + (1-1.5)^2(.375) + (2-1.5)^2(.375) + (3-1.5)^2(.125)$$

$$= .281 + .094 + .094 + .281 = .75$$

$$\sigma^2 = .75$$

Mean and Variance of Discrete Random Variable

- Remember, we said the variance is the mean squared deviation about the mean
- The **standard deviation** is the square root of the variance (Def 5.7 P241)
- In our example, $\sigma = .866$

Nitrous Oxide Example

X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

If I randomly selected 5 dentists, how many would I **expect** to use laughing gas?

Nitrous Oxide Example - Expectation

X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

$$E(X) = 0(.0102) + 1(.0768) + 2(.2304) + 3(.3456) + 4(.2592) + 5(.0778)$$

$$E(X) = 3.0002$$

Nitrous Oxide Example

X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

If I randomly selected 5 dentists, **what is the variance (σ^2)?**

Nitrous Oxide Example - Expectation

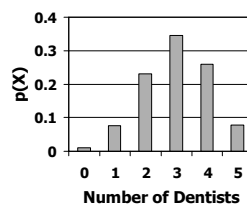
X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

$$E(X-\mu)^2 = (0-3)^2(.0102) + (1-3)^2(.0768) + (2-3)^2(.2304) + (3-3)^2(.3456) + (4-3)^2(.2592) + (5-3)^2(.0778)$$

$$E(X-\mu)^2 = 1.1998 = s^2 \quad s = 1.09535$$

Probability Distribution of the Discrete Variable X

Probability Distribution of X



- $E(X) = 3$
- $S^2 = 1.1998$
- $S = 1.09535$

So what?

- Knowing the mean and standard deviation of a probability distribution can be used with
 - Chebyshev's Rule
 - Empirical Rule
- We can find probabilities associated with particular values of x

Solitaire Example

- I have been asking my students to help in an experiment playing solitaire on the computer
- You need to get on any computer with solitaire and set the game
- We have been using the Vegas style solitaire game.
 - Under GAME picks OPTIONS
 - Pick draw three
 - Pick Vegas Style
 - Pick Keep Score

Solitaire Example

- Vegas Style plays a game where it costs you \$52 to play one hand.
- The draw three options turns the deck over three cards at a time and let's you go through the deck three times.
- You get \$5 for each card you place on the top (i.e., Ace, two, three... of a suit).
- If you get all 52 cards on top you would get back \$208 $(52 \times 5) - 52$

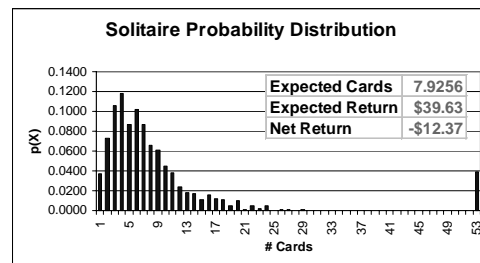
Solitaire Example

- There are 53 outcomes of this discrete random variable
 - Dollars won/lost - $-\$52$ to $+\$208$
 - Number of cards - 0 to 52
- In reality, there are some values that just don't happen

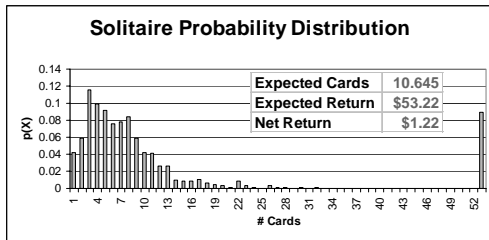
Some results of one class n = 219

Dollars	Cards	Freq	P(x)
-\$52	0	7	.0320
-\$47	1	11	.0502
-\$42	2	26	.1187
-\$37	3	24	.1096
-\$32	4	23	.1050
\$208	52	9	.0411

Student solitaire example n=820



Ilvento Solitaire results n = 735



Solitaire Example

