MICHIGAN STATE UNIVERSITY

AFRE 802 Statistical Methods for Agricultural, Food, & Resource Economists

Discrete random variables & their probability distributions
(Part 2 of 3)
(WMS Ch. 3.3-3.5)

September 14, 2017

Nicole Mason
Michigan State University
Fall 2017

MICHIGAN STATE UNIVERSITY

GAME PLAN

1. Hand back graded in-class exercise & Ch. 2 HW
2. Review and questions from last class
3. Graded in-class exercise
4. Some common discrete probability distributions:
   i. Bernoulli
   ii. Binomial
   iii. Geometric

   Next Tuesday: Negative Binomial & Poisson

1

Review

- **Random variable** =
  - A variable that takes on numerical values and has an outcome that is determined by a random experiment
- **Random experiment** =
  - Experiment whose actual outcome can't be predicted with certainty but whose possible outcomes can be described prior to the experiment
- **Discrete** vs. **continuous** RVs
- **Notation**:
  - *Y*: the random variable
  - *y*: particular values of *Y*
  - *P(Y=y) or p(y)*: the probability distribution of *Y*
    - **Probability distribution** describes the probability that the RV takes on a particular value for all possible values of the RV
  - *EX) P(Y=5) or p(5)*: the probability that *Y* takes on the value of 5

Review (cont'd)

Expected value of *Y*: *E(Y)=μ*

- Measure of central tendency; population mean
- Weighted avg., where each value of *y* is weighted by its probability, *p(y)*:

$$\boxed{\text{For discrete RV: } E(Y) = \sum_i y_i p(y_i)}$$

Expected value of a function of *Y*, E[*g(Y)*]

$$\boxed{\text{For discrete RV: } E\left[g(Y)\right] = \sum_i g(y_i) p(y_i)}$$

3

MICHIGAN STATE UNIVERSITY

Review (cont'd)

Variance of Y: *V(Y)=σ²* (standard deviation is *σ*)

- Measure of variability/spread around the mean

$$\sigma^2 = V(Y) = E\left[(Y-\mu)^2\right] = E(Y^2) - \mu^2$$
$$= E(Y^2) - \left[E(Y)\right]^2$$
$$= \left[\sum_i y_i^2 p(y_i)\right] - \left[\sum_i y_i p(y_i)\right]^2$$

4

Review

MICHIGAN STATE UNIVERSITY

Useful rules for expected values & variances

FOR EXPECTED VALUES

For any constants b and c:

$(i)\ E(c) = c$

$(ii)\ E(bX) = bE(X)$

$(iii)\ E(bX+c) = bE(X)+c$

$(iv)\ E\left[g_1(X)+g_2(X)+...+g_k(X)\right] = E\left[g_1(X)\right]+E\left[g_2(X)\right]+...+E\left[g_k(X)\right]$

$(v)\ E\left[g(X)\right] = \sum_i g(x_i)p(x_i)$

FOR VARIANCES

For any constants b and c:

$(i)\ V(c) = 0$

$(ii)\ V(bX) = b^2V(X)$

$(iii)\ V(bX+c) = b^2V(X)$

5

Graded in-class exercise

Some common discrete probability distributions & their properties

1. Bernoulli
2. Binomial
3. Geometric
4. Negative Binomial
5. Poisson

What applications did you find for Bernoulli, binomial, and geometric?

**** Please look up negative binomial and Poisson applications for next class.*

1. Bernoulli distribution

- **Only two possible outcomes: *y*=1 or 0**
- Let *p* be the probability that *y*=1
  - *P(Y=1) = p(1) = p*
- What is the probability of *y*=0? Call this *q*.
  - *P(Y=0) = p(0) = 1-p = q*
- Probability distribution of a Bernoulli RV:

$$\boxed{p(y) = p^y (1-p)^{1-y} = p^y q^{1-y}}$$

- *Find E(Y) and V(Y).*
  - *E(Y) = p, V(Y) = pq = p(1-p)*

2. Binomial distribution

- **<u>*n* independent & identical Bernoulli trials</u>, so each trial has two possible outcomes: 1 or 0**
- 1="success" (S), 0="failure" (F)
- In a <u>single</u> trial
  - Probability of 1 (or S) = *p*
  - Probability of 0 (or F) = 1-*p* = *q*
  - Same as Bernouilli RV
- A **<u>binomial RV</u>**, *Y*: number of Ss (1s) observed in the *n* independent & identical Bernoulli trials

2. Binomial distribution

- EX) Flipping a weighted coin 5 times. Let *Y*=# of tails
- *What is n, the number of trials here?*
- *How should we define "success" (S) in this experiment?*
  - S is getting a tail
- *How many different ways could we get 1 S in 5 trials?*

$$C_1^5 = \binom{5}{1} = \frac{5!}{1!(5-1)!} = \frac{5!}{1!4!} = 5$$

- *0 Ss? 2 Ss? 3 Ss? 4 Ss? 5 Ss?*

$$\binom{5}{0} = 1, \ \binom{5}{2} = 10, \ \binom{5}{3} = 10, \ \binom{5}{4} = 5, \ \binom{5}{5} = 1$$

- *y Ss in n trials?*

$$\binom{n}{y}$$

MICHIGAN STATE UNIVERSITY

2. Binomial distribution

- EX) Flipping a weighted coin 5 times. Let *Y*=# of tails
- Previous slide: # of ways to get *y* Ss in n trials: $\binom{n}{y}$
- Suppose the coin is weighted such that p(T)=0.6=p, p(H)=0.4=q
- *One way to get Y=4 is TTHTT = SSFSS. What is the probability of SSFSS?* $0.6*0.6*0.4*0.6*0.6 = 0.6^4 0.4^1 = p^4 q^{n-4}$
- *What is P(Y=4)=p(4)?*

$$p(4) = \binom{5}{4} p^4 q^{n-4}$$

- *What is P(Y=y)=p(y)?*

$$p(y) = \binom{n}{y} p^y q^{n-y} = \binom{n}{y} p^y (1-p)^{n-y}$$

MICHIGAN STATE UNIVERSITY

2. Binomial distribution

- **Probability distribution of a binomial RV, *Y***
  (Y = number of "successes" in n independent & identical Bernoulli trials w/ probability of success, p, and probability of failure $q=1-p$)

$$\boxed{p(y) = \binom{n}{y} p^y q^{n-y} = \binom{n}{y} p^y (1-p)^{n-y} \quad \text{for } y = 0, 1, 2, \ldots, n}$$

- **Mean and variance of a binomial RV**

$$\boxed{\begin{aligned} \mu &= E(Y) = np \\ \sigma^2 &= V(Y) = npq \end{aligned}}$$

Proof is on p. 107 of WMS if you're interested

12

MICHIGAN STATE UNIVERSITY

2. Binomial distribution – example #1

Suppose a package of 5,000 bean seeds contains 5% that will not germinate. If a random sample of 5 seeds is tested, what is the probability of getting:

a. Zero seeds that don't germinate (i.e., no "defective" seeds)?
b. At least one seed that doesn't germinate (i.e., at least one defective seed)?

13

2. Binomial distribution – example #2 – using Table 1 (binomial probabilities)

Explain Table 1. *$P(Y \leq 2)$ for $n=5$ and $p=0.20$? $P(Y>2)$?*

Same package of 5,000 seeds w/ 5% defective. Now suppose you draw a random sample of 20 seeds. What is the probability of finding at least 4 defective seeds? Use Table 1.

Aside on random samples

Let N and n represent the numbers of elements in the population and sample, respectively. If the sampling is conducted in such a way that each of the $\binom{N}{n}$ samples has an equal probability of being selected, the sampling is said to be random, and the result is said to be a *random sample*.

MICHIGAN STATE UNIVERSITY

3. Geometric distribution

- Similar set-up as Binomial except that for a **geometric RV**, ***Y*: number of the trial on which the 1st success occurs**
- (Recall for binomial, *Y* was the # of successes)
- For a geometric RV, y = 1, 2, 3, …. (*Why no 0 or n?*)
- Sample space: *Probability?*
  - E_1: S (success on 1st trial) p
  - E_2: FS (success on 2nd trial) qp
  - E_3: FFS (success on 3rd trial) q^2p
  - …
  - E_k: $\underbrace{\text{FFFF…F}}_{k-1}$S (success on kth trial) $q^{k-1}p$

16

MICHIGAN STATE UNIVERSITY

3. Geometric distribution

- **Probability distribution of a geometric RV, *Y***
  (*Y* = number of the trial on which the 1st "success" occurs in a series of independent & identical Bernoulli trials w/ probability of success, *p*, and probability of failure *q=1-p*)

$$\boxed{p(y) = q^{y-1}p \qquad \text{for } y = 1,\ 2,\ 3,\ \ldots}$$

- **Mean and variance of a geometric RV**

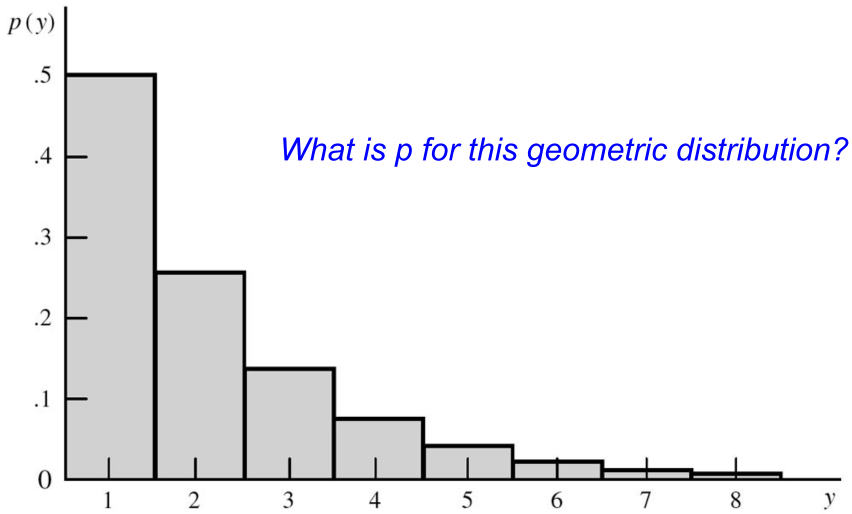
$$\boxed{\begin{array}{l} \mu = E(Y) = \dfrac{1}{p} \\[2ex] \sigma^2 = V(Y) = \dfrac{1-p}{p^2} \end{array}}$$

Proofs are on p. 116-117 and Exercise 3.85 in WMS if you're interested

17

MICHIGAN STATE UNIVERSITY

3. Geometric distribution - graphically



What is p for this geometric distribution?

MICHIGAN STATE UNIVERSITY

3. Geometric distribution - example

Suppose that the probability of tractor engine malfunction during any one-hour period is p=0.02.

a. Find the probability that a given tractor engine will malfunction in the 2nd hour.
b. Find the probability that a given tractor engine will survive at least 2 hours.
c. Let *Y* be the number of one-hour intervals until the first malfunction. Find the mean and variance of *Y*.

Summary

MICHIGAN STATE UNIVERSITY

- We discussed 3 specific, common discrete probability distributions:
  1. **Bernoulli** (1 trial, only 2 outcomes: S (Y=1) or F (Y=0); *p* is probability of S, *q=1-p* is probability of F)
  2. **Binomial** (n independent Bernoulli trials, *Y* is # of Ss)
  3. **Geometric** (series of independent Bernoulli trials, *Y* is the # of the trial on which the 1st S occurs)

Distribution	Probability Function	Mean	Variance
Bernoulli	$p(y) = p^y(1-p)^{1-y}$	p	$p(1-p)$
Binomial	$p(y) = \binom{n}{y} p^y (1-p)^{n-y}$; $y = 0, 1, \dots, n$	np	$np(1-p)$
Geometric	$p(y) = p(1-p)^{y-1}$; $y = 1, 2, \dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

Homework:

- WMS Ch. 3 (part 2 of 3)
  - Binomial distribution: 3.38. 3.44, 3.51, 3.60 – use Table 1 as applicable
  - Geometric distribution: 3.67, 3.68, 3.72, 3.75
- Ch. 3 HW will be due the class after we finish Ch. 3 (probably due next Thursday)

Next class:

- Discrete random variables (Part 3 of 3) – wrap up specific distributions (negative binomial & Poisson) & Tchebysheff's Inequality

Reading for next class:

- WMS Ch. 3: read 3.6, 3,8, 3.11-3.12

Application to look into for next class:

- Find an example of how the negative binomial or Poisson distribution is applied in your field

21

In-class exercise #1 - binomial

3.45 A fire-detection device utilizes three temperature-sensitive cells acting independently of each other in such a manner that any one or more may activate the alarm. Each cell possesses a probability of $p = .8$ of activating the alarm when the temperature reaches 100° Celsius or more. Let Y equal the number of cells activating the alarm when the temperature reaches 100°.

a Find the probability distribution for Y.

b Find the probability that the alarm will function when the temperature reaches 100°.

c Find the mean and variance of Y

In-class exercise #2 – binomial (using Table 1)

3.52 The taste test for PTC (phenylthiocarbamide) is a favorite exercise in beginning human genetics classes. It has been established that a single gene determines whether or not an individual is a "taster." If 70% of Americans are "tasters" and 20 Americans are randomly selected, what is the probability that

a at least 17 are "tasters"?

b fewer than 15 are "tasters"?

c Find the mean and variance of Y

In-class exercise #3 – geometric

3.73 A certified public accountant (CPA) has found that nine of ten company audits contain substantial errors. If the CPA audits a series of company accounts, what is the probability that the first account containing substantial errors

a is the third one to be audited?

b will occur on or after the third audited account?

c Find the mean and variance of Y

In-class exercise #4 – geometric

3.81 How many times would you expect to toss a balanced coin in order to obtain the first head?