MICHIGAN STATE UNIVERSITY

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

Discrete random variables & their probability distributions – Part 3 of 3

September 19, 2017

Nicole Mason
Michigan State University
Fall 2017

MICHIGAN STATE UNIVERSITY

GAME PLAN

- Review
- Graded in-class exercise
- Quick aside on random sampling
- Continue common discrete probability distributions:
  - Geometric
  - Negative binomial
  - Poisson
- Tchebysheff's Theorem

1

Review

- We discussed 2 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** (*Y* is # of Ss in n independent Bernoulli trials)

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

Any questions from last class?

GRADED IN-CLASS EXERCISE

Brief 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*.

Some common discrete probability distributions & their properties

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

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 (1st success on 1st trial) p
  - E_2: FS (1st success on 2nd trial) qp
  - E_3: FFS (1st success on 3rd trial) q^2p
  - …
  - E_k: $\underbrace{\text{FFFF…F}}_{k-1}$S (1st success on kth trial) $q^{k-1}p$

6

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**

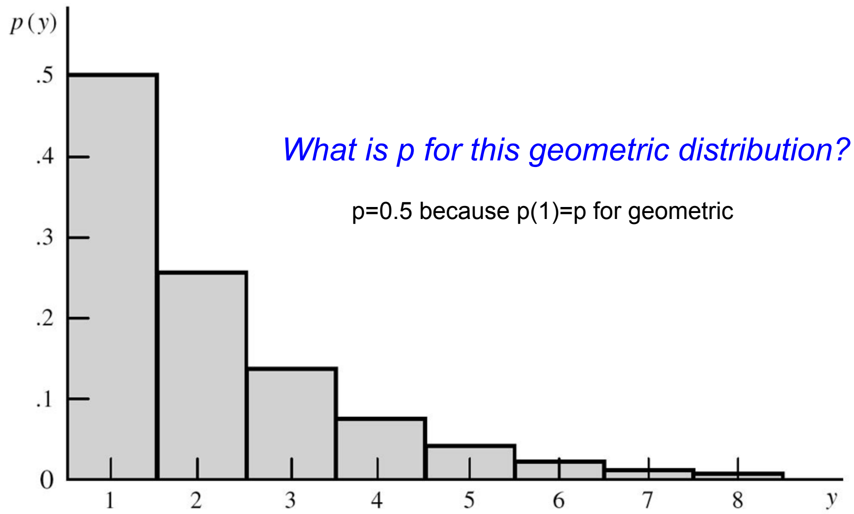
$$\boxed{\begin{aligned} \mu &= E(Y) = \frac{1}{p} \\ \sigma^2 &= V(Y) = \frac{1-p}{p^2} \end{aligned}}$$

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

7

MICHIGAN STATE UNIVERSITY

3. Geometric distribution - graphically



What is p for this geometric distribution?

p=0.5 because p(1)=p for geometric

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 for the first time in the 2^{nd} hour.
b. Find the probability that a given tractor engine will survive at least 2 hours.
c. Let *Y* be the number of the one-hour interval in which the first malfunction occurs. Find the mean and variance of *Y*.

9

Some common discrete probability distributions & their properties

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

4. Negative binomial distribution

- As usual: series of independent & identical Bernoulli trials, where *p* is the probability of "success" of each trial

- **Geometric**: *Y* is # of the trial of the 1st success
- **Negative binomial**: *Y* is # of the trial of the r^{th} success (for r = 2, 3, …)

4. Negative binomial distribution

MICHIGAN STATE UNIVERSITY

- *p(y)* is probability that the r^th success occurs on trial *y*
- Let A = { the first (y -1) trials contain (r -1) successes }
  B = { trial y results in a success }
- Using set notation and events A and B, *what is the event, "the r^th success occurs on trial y"?* $A \cap B$
- *What is this probability if A and B are independent?*

$$p(y) = P(A \cap B) = P(A) \times P(B)$$

- *What is P(B)?* p
- Note that *A* is a binomial experiment. *What is P(A)?*

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

- *Putting these together, p(y)=P(A) × P(B) is:*

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

12

MICHIGAN STATE UNIVERSITY

4. Negative binomial distribution

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

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

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

$$\mu = E(Y) = \frac{r}{p}$$

$$\sigma^2 = V(Y) = \frac{r(1-p)}{p^2}$$

13

4. Negative binomial distribution - example

A study on the economics of oil exploration indicates that an exploratory well drilled in a particular region should strike oil with probability 0.2.

a. Find the probability that the 3rd oil strike comes on the 5th well drilled.
b. Find the mean and variance of the number of wells that must be drilled for the 3rd oil strike to occur.

Some common discrete probability distributions & their properties

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

MICHIGAN STATE UNIVERSITY

5. Poisson distribution

- *Y* is the **# of times some event happens in a given interval** (of time, length, area, volume, etc.)

- **λ** is the average value of *Y*

16

MICHIGAN STATE UNIVERSITY

5. Poisson distribution

- **Probability distribution of a Poisson RV, *Y***
- (*Y* = number of times an event occurs in a given interval, where λ is the average value of *Y*)

$$\boxed{\begin{array}{l} p(y) = \dfrac{\lambda^y}{y!} e^{-\lambda} \\ \text{for } y = 0,\ 1,\ 2,\ \ldots \text{ and } \lambda > 0 \end{array}}$$

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

$$\boxed{\begin{array}{l} \mu = E(Y) = \lambda \\ \sigma^2 = V(Y) = \lambda \end{array}}$$

Note that **mean=variance=λ** for a Poisson RV

17

MICHIGAN STATE UNIVERSITY

5. Poisson distribution – example #1

A certain type of tree has seedlings randomly dispersed in a large area, with the mean density of seedling being approximately five per square yard.

a. What is the probability that a given 1-square yard area contains no seedlings?
b. If a forester randomly located ten 1-square-yard sampling regions in the area, find the probability that none of the regions will contain seedlings.
c. What is the mean and variance of the RV in part (b)?

18

MICHIGAN STATE UNIVERSITY

Poisson RVs: multiple intervals vs. a Poisson RV multiplied by a constant

E.g., previous example: seedlings per square yard, Y ~ Poisson with λ=5 → E(Y)=V(Y)=λ=5

- Multiple intervals – e.g., 10 square yards → assuming these are independent, then can think of this as a new Poisson RV, say X, with parameter ω=10λ=10*5=50 → E(X)=V(X)=ω=50

vs.

- A Poisson RV multiplied by a constant: e.g., each seedling can be sold for $10, and we want to know the mean and variance of seedling revenue per square yard. This is E(10Y)=10E(Y)=10*λ=50 and V(10Y)=100V(Y)=100*λ=500

19

MICHIGAN STATE UNIVERSITY

5. Poisson distribution

- Poisson distribution can be derived as the limit of a binomial distribution as the number of trials (n) → ∞
- Because of this relationship, Poisson probabilities can be used to approximate binomial probabilities when:
  - The # of trials **(*n*) is large**, and
  - The probability of success **(*p*) is small**, such that
  - ***λ=np* roughly < 7** (others say *λ=np* ≤ 20 or n≥100) – rules of thumb vary)
    - Recall that E(Y)=np for binomial, E(Y)=λ for Poisson

20

MICHIGAN STATE UNIVERSITY

5. Poisson distribution – example #2

Approximating binomial probabilities w/ Poisson probabilities

Suppose that Y ~ binomial with n=20 and p=0.1.

a. Find the exact value of P(Y<3) using the table of binomial probabilities (Appendix 3, Table 1).
b. Use Appendix 3, Table 3 to approximate this binomial probability using the corresponding Poisson probability.

21

Summary

MICHIGAN STATE UNIVERSITY

We have discussed 5 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)
4. **Negative binomial** (series of independent Bernoulli trials, *Y* is the # of the trial on which the rth S occurs)
5. **Poisson** (*Y* is the # of times an event occurs in a given interval, and λ is the average value of *Y*)

22

Summary

MICHIGAN STATE UNIVERSITY

Table 1 Discrete Distributions

Distribution	Probability Function	Mean	Variance
Bernoulli	$p(y) = p^y(1-p)^{1-y}$; $y = 0, 1$	p	$p(1-p)$
Binomial	$p(y) = \binom{n}{y} p^y(1-p)^{n-y}$; $y = 0, 1, \ldots, n$	np	$np(1-p)$
Geometric	$p(y) = p(1-p)^{y-1}$; $y = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$p(y) = \frac{\lambda^y e^{-\lambda}}{y!}$; $y = 0, 1, 2, \ldots$	λ	λ
Negative binomial	$p(y) = \binom{y-1}{r-1} p^r(1-p)^{y-r}$; $y = r, r+1, \ldots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$

23

MICHIGAN STATE UNIVERSITY

Tchebysheff's Theorem (Chebyshev's Inequality)

- Recall the "**empirical rule**": useful for probability distributions that are roughly bell-shaped → can determine approx. probability of being in $\mu \pm k\sigma$
- But many distributions are NOT bell-shaped
- **Tchebysheff's Theorem**: can use for any probability distribution to determine the lower bound for probability of being in $\mu \pm k\sigma$

24

MICHIGAN STATE UNIVERSITY

Tchebysheff's Theorem (cont'd)

For any RV, *Y,* with with mean μ & variance σ^2:

$$P\left[(\mu - k\sigma) < Y < (\mu + k\sigma)\right] \geq 1 - \frac{1}{k^2}$$

or

$$P\left(|Y - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

for any constant $k > 0$

The probability of being less than k standard deviations from the mean is at least $1-1/k^2$

The probability of being at least k standard deviations from the mean is no more than $1/k^2$

25

MICHIGAN STATE UNIVERSITY

Tchebysheff's Theorem - example

The number of customers per day at a sales counter, *Y*, has been observed for a long period of time and found to have mean 20 and standard deviation 2. The probability distribution of *Y* is not known. What can be said about the probability that Y will be greater than 16 but less than 24? (Hint: find k by determining # of standard deviations 16 and 24 are from their means, then use the formula on the previous slide.)

26

MICHIGAN STATE UNIVERSITY

Tchebysheff's Theorem (cont'd)

k	$P[(\mu-k\sigma)<Y<(\mu+k\sigma)] \geq 1-1/k^2$	$P[Y\leq(\mu-k\sigma) \text{ OR } Y\geq(\mu+k\sigma)] \leq 1/k^2$
1	0	1
2	0.750	0.250
3	0.889	0.111
4	0.938	0.063
5	0.960	0.040
6	0.972	0.028
7	0.980	0.020
8	0.984	0.016
9	0.988	0.012
10	0.990	0.010
	Etc.	

- *Which of these is* ***upper bound*** *(max.) vs.* ***lower bound*** *(min.) of a probability?*
- *Lower bound (min.) probability of being less than 2 standard deviations from the mean for any distribution?*
- *Upper bound (max.) probability of being 3 or more standard deviations from the mean for any distribution?*

27

MICHIGAN STATE UNIVERSITY

Homework:

- WMS Ch. 3 (part 3 of 3)
  - Geometric distribution: 3.67, 3.68, 3.72, 3.75
  - Negative binomial distribution: 3.93, 3.94
  - Poisson distribution: 3.121, 3.126, 3.134
  - Tchebysheff's theorem: 3.167, 3.171
- If we finish Ch. 3 today, HW will be due on Thurs. (Sep. 22); otherwise, it will be due next Tues. (Sep. 26)

Next class:

- Integration (to prepare us for continuous random variables)

Reading for next class:

- Chiang & Wainwright Ch. 14 (sections 14.1-14.4) – posted to D2L

28

MICHIGAN STATE UNIVERSITY

In-class exercises #1 & 2 – negative binomial

3.90 The employees of a firm that manufactures insulation are being tested for indications of asbestos in their lungs. The firm is requested to send three employees who have positive indications of asbestos on to a medical center for further testing. If 40% of the employees have positive indications of asbestos in their lungs, find the probability that ten employees must be tested in order to find three positives.

3.91 Refer to Exercise 3.90. If each test costs \$20, find the expected value and variance of the total cost of conducting the tests necessary to locate the three positives.

29

In-class exercises #3 – Poisson

3.122 Customers arrive at a checkout counter in a department store according to a Poisson distribution at an average of seven per hour. During a given hour, what are the probabilities that

a no more than three customers arrive?

b at least two customers arrive?

c exactly five customers arrive?

MICHIGAN STATE UNIVERSITY

In-class exercises #4 – Tchebysheff's Theorem

3.170 The U.S. mint produces dimes with an average diameter of .5 inch and standard deviation .01. Using Tchebysheff's theorem, find a lower bound for the number of coins in a lot of 400 coins that are expected to have a diameter between .48 and .52.