Stat 104: Quantitative Methods

Class 12: Discrete Probability Distributions

Review: E(X) and Var(X)

Economic Scenario	Profit ($ Millions) X		Probability P
Great	x_1 10	$P(X=x_1)$	0.20
Good	x_2 5	$P(X=x_2)$	0.40
OK	x_3 1	$P(X=x_3)$	0.25
Lousy	x_4 -4	$P(X=x_4)$	0.15

$E(X)=10(.2)+5(.4)+1(.25)+(-4)(.15)=3.65$

$E(X^2)=100(.2)+25(.4)+1(.25)+16(.15)=32.65$

$Var(X)= E(X^2)-E(X)^2=32.65-3.65^2=13.32$

Probability Distributions

- **Probability Distributions**
  - **Discrete** Probability Distributions
    - Binomial
    - Poisson
  - **Continuous** Probability Distributions
    - Uniform
    - Normal

Probability distributions and random variables go hand in hand. We talk for example about the binomial probability distribution or a binomial random variable-same thing.

The Binomial Distribution

- Many business experiments can by characterized by a binomial distribution (random variable).
- The key ideas are as follows:
  - We are doing something *n* times.
  - Each instance of a trial has only two outcomes, "success" or "failure".
  - The trials of the experiment are independent of each other.
  - The probability of a success, *p*, remains constant from trial to trial
  - We are interested in the **total number of successes** (out of *n* trials)

Binomial random variable is for modeling counts.

Example

- MSA Electronics is experimenting with the manufacture of a new transistor.
  - Every hour a random sample of 5 transistors is taken.
  - The probability of one transistor being defective is 0.15.
- What is the probability of finding more than 3 defective?

Example

- The probability that any particular customer who visits zappos.com makes a purchase is about 0.20.
- Assume that visitors make their purchase decisions independently-what is the probability that no more than 10 of the last 50 visitors make purchases?

The Binomial Assumptions

- You know it's a binomial situation if the answer to each question below is yes:
  - Are *n* independent trials taking place?
  - Is there success or failure on each trial?
  - Is the probability of success the same for each trial?
  - Is our interest in the total number of success in the *n* trials?

Example

- Assume the probability that a randomly selected student receives no financial aid is 25%.
- The college administration surveys students until they find one not on financial aid. Let the random variable *X* denote the number of students surveyed.
- **Is *X* a binomial random variable?**
- **A YES**
- **B NO** We don't have a fixed number of trials. Before you start, you don't know when you'll stop.

Example

- Unknown to the quality control inspector, the defect rate of new energy efficient green fluorescent light bulbs is 3%.
- The inspector selects 5 bulbs at random.
- Let the random variable *X* denote the number of defective bulbs in the inspector's sample.
- **Is *X* a binomial random variable?**
- **A YES**
- **B NO**

Notation: X~Bin(*n*,*p*)

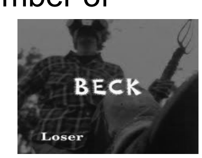
- We introduce the notation
  - X~Bin(n,p) or X~B(n,p)
- The "~" is read as "is distributed as"
- So this statement means, "X is distributed as a binomial random variable, with *n* trials and probability of success on each trial *p*".

The "end point probabilities"

- There is a slightly complicated formula to calculate binomial probabilities, but there are two special cases which are very easy to deal with.
- These two special cases are P(all failures), and P(all successes).

P(all failures)

- X~Bin(n, π) so X counts the number of **successes** in n trials.



- P(all failures) = P(X=0) ☹
- P(all failures)=
  - P(fail and fail and fail and….)
  - =P(fail)P(fail)P(fail)P(fail)….(why?) independent
  - =[P(fail)]^n (why?) probability of success/failure is same for all trials
  - =(1-p)^n

P(all successes)

- X~Bin(n, π) so X counts the number of **successes** in n trials.

- P(all successes) = P(X=n) ☺
- P(all succeses)=
  - P(win and win and win and….)
  - =P(win)P(win)P(win)P(win)….(why?)
  - =[P(win)]^n (why?)
  - =(p)^n

Example

Suppose that 30% of the citizens of Somerville are opposed to joining Cambridge to form the new hipster paradise Somebridge.

If a random sample of seven adult residents of Somerville were selected, what is the probability that

- None would be opposed to joining Cambridge?
- All would be opposed to joining Cambridge

Two more rules: At least one

- From the complement rule
  - P(at least one success)=1-P(all failures)
  - P(at least one failure)=1-P(all successes)

All successes = p^n
All failures = (1-p)^n
1-p^n = negating all successes = at least one failure
1-(1-p)^n = negating all failures = at least one success

Example

Suppose that 30% of the citizens of Somerville are opposed to joining Cambridge to form the new hipster paradise Somebridge.

If a random sample of seven adult residents of Somerville were selected, what is the probability that at least one person would be opposed to joining Cambridge?

1-(0.7)^7

Finding Other Binomial Probabilities

- We have to learn how to count
- As an example, suppose you have 3 interesting friends and each of them (without consulting the other two) decides daily to wear caps with probability 0.3
- You are about to meet the three of them for lunch.
- Define the random variable X to be the number of your friends wearing caps.

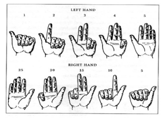
Wearing Caps

- We know that X can take on the values 0,1,2 and 3.
- From the endpoint probability rules we know the following:

$$P(X=0)=(0.7)^3$$
$$P(X=3)=(0.3)^3$$

- Now we need to find P(X=1) and P(X=2) [do we need to find both probabilities???]No. 1 - (sum of other 3)

Learning to Count



- We want to find P(X=1)
- This is the probability that 1 friend out of 3 is wearing caps.
- How can this happen? Lets list the scenarios

$$C\bar{C}\bar{C} \text{ or } \bar{C}C\bar{C} \text{ or } \bar{C}\bar{C}C$$

- Then

$$P(X=1) = P(C\bar{C}\bar{C} \text{ or } \bar{C}C\bar{C} \text{ or } \bar{C}\bar{C}C)$$

Scenario Probabilities are the Same

- We have

$$P(X=1) = P(C\bar{C}\bar{C} \text{ or } \bar{C}C\bar{C} \text{ or } \bar{C}\bar{C}C)$$
$$= P(C\bar{C}\bar{C}) + P(\bar{C}C\bar{C}) + P(\bar{C}\bar{C}C)$$

- Notice that the probabilities are all the same

1 person wears a cap and the others don't

$$P(C\bar{C}\bar{C}) = (0.3)(0.7)(0.7) = (0.3)(0.7)^2$$
$$P(\bar{C}C\bar{C}) = (0.7)(0.3)(0.7) = (0.3)(0.7)^2$$
$$P(\bar{C}\bar{C}C) = (0.7)(0.7)(0.3) = (0.3)(0.7)^2$$

Summarizing this Result

- We have found that

$$P(X=1) = 3(0.3)(0.7)^2$$

Number of scenarios

Probability of 1 "success" [cap]

Probability of 2 "failures" [no cap]

- The hard part is figuring out the number of scenarios; the probability of each scenario is easy.

Counting Rule for Combinations

- A combination is an outcome of an experiment where x objects are selected from a group of n objects

$$\boxed{\binom{n}{x} = \frac{n!}{x!(n-x)!}}$$

"How many ways can x successes appear in n trials"

where:

n! =n(n - 1)(n - 2) . . . (2)(1)

x! = x(x - 1)(x - 2) . . . (2)(1)

0! = 1 (by definition)

Binomial Distribution Formula

$$\boxed{P(X=x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

x = number of 'successes' in sample, (x = 0, 1, 2, ..., n)

p = probability of "success" per trial

q = probability of "failure" = (1 – p)

n = number of trials (sample size)

Example: Flip a coin four times, let x = # heads:

n = 4

p = 0.5

q = (1 - .5) = .5

x = 0, 1, 2, 3, 4

Deconstruct the Formula

$$\boxed{P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

Number of ways to have x successes in n trials

Probability of x successes

Probability of n-x failures

x+(n-x) = n. there are n total trials

Example

- You sell sandwiches. 70% of people choose chicken, the rest choose veggie.
- What is the probability of selling 7 chicken sandwiches to the next 10 customers?

$$P(X=7) = \binom{10}{7}(0.7)^7(0.3)^3$$
$$= \frac{10!}{7!(10-7)!}(0.7)^7(0.3)^3$$
$$= 0.27$$

Binomial Calculations in R

- The command `dbinom(n,p,x)` computes $P(X=x)$
- The command `pbinom(n,p,x)` computes $P(X \leq x)$

Example

- The Jordan Sports Equipment Company finds that 10% of the general adult population is left-handed. If 20 adults are randomly selected
- Find the probability that exactly 3 are left-handed.
- Find the probability that less than 6 are left-handed.

Example

- Let X=number in group of 20 who is left handed.
- Then X~Binomial (n=20,p=0.1)

Find P(X=3)

- What is the probability that three people in your sample are left handed?

$$P(X=3) = \binom{20}{3}(0.1)^3(0.9)^{17} = 0.19$$

```
> dbinom(3,20,.1)
  [1] 0.1901199
```

Find P(X ≤5)

- R

```
> pbinom(5,20,.1)
[1] 0.9887469
```

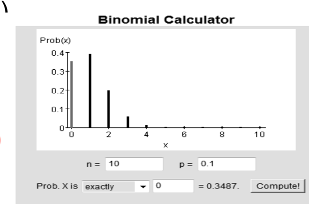
Binomial Example

- An oil exploration firm is to drill ten wells, each in a different location. Each well has a probability of 0.1 of producing oil. It will cost the firm $60,000 to drill each well.
- A successful well will bring in oil worth $1 million.
- Define X to be the number of good wells.
- X~Bin(10,.1)

Example: Probability of Losing Money

- What is the probability of losing money?
- Profit=1000000X-60000(10) (why?)
- P(profit<0)
- =P(1,000,000X-600,000<0)
- =P(X<.6)
- =P(X=0) (why?)

Discrete random variable.
Only takes values 0,1,2…10



Example

- Suppose you own a catering company. You hire temporary employees to act as servers from the local college.
- Not being the most reliable employees, there is an 80% chance that any one server will actually show up for a scheduled event.
- For a wedding scheduled on a given Saturday you need at least 5 servers.
- Suppose you schedule 5 employees, what is the probability that all 5 come to work?

```
> dbinom(5,5,.8)
[1] 0.32768
```

Example

- Suppose you schedule 7 employees, what is the probability that at least 5 come to work?

```
> 1-pbinom(4,7,.8)
[1] 0.851968
```

Binomial Distribution Characteristics

- Mean

$$\mu = E(X) = np$$

- Variance and Standard Deviation

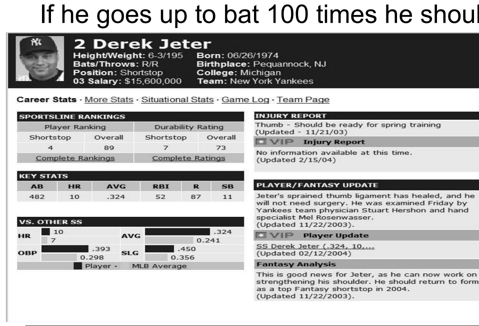
$$\sigma^2 = npq$$

$$\sigma = \sqrt{npq}$$

Where n = sample size
p = probability of success
q = (1 – p) = probability of failure

Example: Derek Jeter is a career .324 hitter.

If he goes up to bat 100 times he should get about 32.4 hits



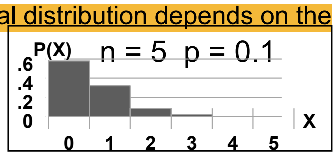
The variability of hits is 100(.324)(.636)= 22.5

Std Dev = sqrt(22.5)=4.74

Binomial Distribution

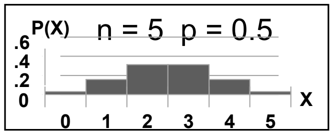
- The shape of the binomial distribution depends on the values of p and n
  - Here, n = 5 and p = .1

    Skewed to the right



  - Here, n = 5 and p = .5

    Symmetric



The Poisson Distribution

- Another widely-used theoretical discrete probability distribution is the Poisson; named for S. D. Poisson, a brilliant 19th century French mathematician.
- The Poisson distribution is used to model the number of events within a given time interval.

Tres debonair, no?

The Poisson Distribution

- The Poisson distribution models counts (like the binomial) but in a different situation.
- If events happen at a constant rate over time, the Poisson gives the probability of *x* number of events occurring in a time period *T.*
- If X=# counts per second (say), then

$$P(X = x) = \frac{\lambda^x}{e^\lambda x!}$$

Where λ=average number of counts per time period

Mean and Variance of the Poisson

- λ is called the rate parameter
- It can be shown that if *X* is a Poisson random variable with rate parameter λ, then
- $E(X) = \lambda$
- $Var(X) = \lambda$

The Poisson Conditions

- Three basic conditions need to be met in order for a random phenomenon to qualify as Poisson:

(1) We need to be assessing probability for the number of occurrences of some event *per unit time, space, or distance.*

(2) The average number of occurrences per unit of time, space, or distance is constant and proportionate to the size of the unit of time, space or distance involved. For example, if the average number of occurrences in a one-minute interval is 1.5, then the average number of occurrences in a two-minute interval will be exactly two times 1.5 (that is, 3.0); the three-minute average will be three times 1.5, etc.

(3) Individual occurrences of the event are *random* and *statistically independent.*

Example

- For a recent period of 100 years, there were 530 Atlantic hurricanes. Assume the Poisson distribution is a suitable model so the rate is 530/100 = 5.3 per year.
- If *P*(*x*) is the probability of *x* hurricanes in a randomly selected year, find *P*(2).

```
> dpois(2,5.3)
[1] 0.07010694
```

$$P(2) = \frac{\lambda^x}{e^\lambda x!} = \frac{5.3^2}{(2.71828)^{5.3}\, 2!} = 0.0701$$

Example

- Patients arrive at the emergency room of Mercy Hospital at the average rate of 6 per hour on weekend evenings.
- What is the probability of 4 arrivals in 30 minutes on a weekend evening?
- We have to adjust the rate for the unit of time we are interested in.
- 6 per hour = 3 per half hour.

Example

- We have λ=3

$$P(X=4) = \frac{3^4 (2.71828)^{-3}}{4!} = \quad .1680$$

```
> dpois(4,lambda=3)
[1] 0.1680314
```

Example

- If calls to your cell phone are a Poisson process with a constant rate λ **= 2 calls per hour**, what is the probability that, ***if you forget to turn your phone off*** in a **1.5 hour class**, your phone rings during that time?
- If **X = #** calls in **1.5 hours**, we want

P(X ≥ 1) = 1 – P(X = 0)

$$P(X=0) = \frac{(2*1.5)^0 e^{-2(1.5)}}{0!} \frac{(3)^0 e^{-3}}{0!} = e^{-3} = .05$$

P(X ≥ 1) = 1 – .05 = 95% chance

```
> 1-ppois(0,lambda=3)
[1] 0.9502129
```

Using R

- **ppois(4,lambda=2.72**) gives you the probability of getting 4 or less events in a Poisson distribution with mean λ = 2.72 ;
- **dpois(4,lambda=2.72)** gives you the probability of exactly 4.

Example: Whole Foods

- Arrivals at the checkout counter of your local Whole Foods Store average 3 per minute.
- Use the Poisson function to compute the probability of
  a) Exactly one arrival in the next minute observed
  b) No arrivals in the next minute observed
  c) No more than 2 arrivals in the next minute observed
  d) Exactly one arrival in the next 30 seconds.

R Output

a) **Exactly one arrival in the next minute observed**

```
> dpois(1,3)
[1] 0.1493612
```

b) **No arrivals in the next minute observed**

```
> dpois(0,3)
[1] 0.04978707
```

c) **No more than 2 arrivals in the next minute observed**

```
> ppois(2,3)
[1] 0.4231901
```

d) **Exactly one arrival in the next 30 seconds**

```
> dpois(1,1.5)
[1] 0.3346952
```

Note we made λ=1.5 since an average of 3 arrivals in 1 minute reduces to 1.5 arrivals per 30 seconds.

Things you should know

- ❑The Binomial Distribution
- ❑The endpoint probabilities
- ❑np and np(1-p) as the binomial mean and variance
- ❑The Poisson Distribution