Binomial Probability Distribution

Math 122

Random Variables

- A variable x whose value is determined by the outcome of an experiment.
- P(x) = probability of a particular value of x
- Mean/Expected Value: $\mu = \sum xP(x)$
  - If the experiment is repeated many times and the values of x are averaged, the average should be near μ.
- Standard Deviation: σ

Range Rule of Thumb

- Usual values are between

$$\mu - 2\sigma \text{ and } \mu + 2\sigma$$

5% Rule

- If $P(x \leq N) \leq 5\%$, then N is unusually low

- If $P(x \geq N) \leq 5\%$, then N is unusually ~~low~~
  High

Special Distributions

- We want a few specific, common distributions so that we know what to do when we encounter them.
- Discrete
  - Binomial (counting successes in trials)
  - Poisson (counting events in an interval)
- Continuous
  - Uniform (simple)
  - Normal (pervasive bell curve)
  - t, F, χ^2

Binomial Distribution

Example

A short multiple choice quiz consists of 5 questions. Each question has 4 options, only one of which is correct. Bob guesses on every questions. Let x be the number of questions that Bob gets correct.

Find P(x=2).

$P(c) = 1/4$

$P(w) = 3/4$

Ways to get 2 correct

$P(CCWWW) = \frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} = \left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)^3$

$P(CWCWW) = \frac{1}{4} \times \frac{3}{4} \times \frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} = \left(\frac{1}{4}\right)^2 \times \left(\frac{3}{4}\right)^3$

CWWCW

CWWWC

WCCWW

WCWCW

WCWWC

WWCCW

WWCWC

WWWCC

all same

$P(\text{exactly } 2 \text{ correct}) = (\text{Add}) = 10 \times \left(\frac{1}{2}\right)^2 \times \left(\frac{3}{4}\right)^3$

\# successes \# failures

$$\underline{10} \times \underline{\left(\frac{1}{4}\right)}^{2} \times \left(\frac{3}{4}\right)^{2}$$

\# ways to get 2 correct

p(success)

p(failure)

Quiz Example

- 5 questions
- The questions are independent
- Each question is either correct or wrong
- The probability of being correct is the same for each question.
- x is the number of correct questions.

Binomial Distribution

- A fixed number of trials is repeated.
- The trials are independent.
- Each trial ends in success or failure.
- The probability of success is the same for each trial.
- The value of x is the number of successes.

Examples of Binomial Distributions

- The number of female children out of 10 randomly selected children
- The number of green peas in sets of 5 offspring peas.
- The number of correct responses when you guess on a multiple choice test.
- The number of Republicans among sets of 1000 random voters.

Binomial Distribution Notation

- n = number of trials
- p = probability of success
- q = probability of failure = 1-p
- P(x) = probability of getting exactly x successes in n trials.

A formula

#successes

#Fail

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

#ways to get X successes

P(success)

P(fail)

$n! =$ "n factorial" $= 1 \cdot 2 \cdot 3 \cdot 4 \cdots n$

Functions we will use

- $P(x = N)$ = binom**p**df(n, p, N)

- $P(x \leq N)$ = binom**c**df(n, p, N)

- “c” is for “cumulative”

What is the probability that a family with 5 children has exactly 3 boys?

$n = 5$ $\quad X = \#\text{Boys}$

$p = \frac{1}{2}$

$P(X=3) = \text{binompdf}(5, 1/2, 3) = 0.3125$

What is the probability that a family with 5 children has 3 or fewer boys?

$n = 5$ $\quad\quad$ $X = \#$ Boys

$p = \frac{1}{2}$

$P(X \leq 3) = \text{binomcdf}(5, 1/2, 3) = 0.8125$

What is the probability that a family with 5 children has at least 3 boys?

$n = 5$ $\quad X = \#$ Boys

$p = \frac{1}{2}$

$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \text{binomcdf}(5, 1/2, 2)$

complement

$P(X < 3)$

$= 0.5$

Lesser/Greater

What is the probability that x is…

- equal to N? pdf (N)
- less than or equal to N? cdf (N)
- less than N? cdf(N – 1)
- greater than or equal to N? 1 – cdf (N – 1)
- greater than N? 1 – cdf (N)

Look closely @ the online calculator.
There is a feature which helps w/ this.

Lesser/Greater

- **at most** means **less than or equal**
- **at least** means **greater than or equal**
- **no more than** means **less than or equal**
- **no less than** means **greater than or equal**
- **up to** means **less than or equal**

Guessing

$p=\frac{1}{2}$ n

- A true/false test has 100 questions. Each question has 2 options, of which one is correct. You guess on every question.
  - What is the probability that you get at least half correct?

    $P(X \geq 50) = 1 - P(X < 50) = 1 - P(X \leq 49) = 1 - \text{binomcdf}(100, \frac{1}{2}, 49)$

  - What is the probability you get no more than 30 correct?

    $P(X \leq 30) = \text{binomcdf}(100, \frac{1}{2}, 30)$

  - What is the probability that you get ~~60 or more~~ correct?

    more than 60

    $P(X > 60) = 1 - P(X \leq 60) = 1 - \text{binomcdf}(100, \frac{1}{2}, 60)$

  - What is the probability that you get exactly 50 correct?

    $P(X = 50) = \text{binompdf}(100, \frac{1}{2}, 50)$

  - What is the probability that you get 100 correct?

    $P(X = 100) = \text{binompdf}(100, \frac{1}{2}, 100)$

Another 5% Rule

- In sampling without replacement, the individuals are not independent, but…

- If the sample size is no more than 5% of the population, then we can treat the individual as independent.

On a college campus of 10,000 students, 2/3 of the students are female. If 10 students are chosen at random, what is the probability that no more than half of them are female?

$X=$ # females out of 10 random students

$p = 1/2 \qquad n=10$

X is binomial (sort of - sample is a small % of pop)

$P(X \le 5) = \text{binomcdf}(\underset{n}{\underset{\uparrow}{10}}, \underset{p}{\underset{\uparrow}{1/2}}, \underset{x}{\underset{\uparrow}{5}}) = 0.6230$

Binomial Mean and Standard Deviation

Know!

$$\mu = np$$

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

$$\sigma = \sqrt{np(1-p)}$$

Gregor Mendel

- Gregor Mendel estimated that the probability that a pea pod with green/yellow genes turns out to be green is 0.75.
- To test his estimate, he bred 580 pea pods. Of these, 428 were green.
- If Mendel was correct, would this be unusual (according to the Range Rule of Thumb)?

X= # green peas out of 580

$p = 0.75$

$n = 580$

$q = 0.25 \quad (1-p)$

$\mu = np = 0.75 \times 580 = 435$

$\sigma = \sqrt{npq} = \sqrt{580 \cdot .75 \cdot .25} \approx 10.43$

$\mu + 2\sigma = 455.9$

$\mu - 2\sigma = 414.14$

428 is not unusual

This is consistent w/ Gregor Mendel's assumption

X

Gregor Mendel

- Gregor Mendel estimated that the probability that a pea pod with green/yellow genes turns out to be green is 0.75.
- To test his estimate, he bred 580 pea pods. Of these, 428 were green.
- If Mendel was correct, would this be unusual (according to the 5% Rule)?

$X = \#$ green peas out of 580

$n = 580$ $\qquad \mu = np = 435$

$p = 0.75$

Considering 428

Is 428 unusual? ← check $\leq$ b/c $428 < \mu$

$P(x \leq 428) = \text{binomcdf}(580, 0.75, 428) = 0.265$

0.265 is not unusual

428 is consistent w/ Mendel's guess

(He may be correct)

Is 473 unusual?

$P(x \geq 473) = 1 - P(x < 473) = 1 - P(x \leq 472)$

Unusual

↓

$= 1.049\text{E-}4 = 1.049 \times 10^{-4} = 0.0001049$

Racial Discrimination

p

- 79.1% of the population of Hidalgo County, TX, is of Hispanic descent.

n

- Of 870 people selected for jury duty for a case of burglary against Rodrigo Partida, 339 or 39% were Hispanic.
- After conviction, Partida was granted a new trial because of the discrepancy of 39% compared to 79%.
- Statistically, would 339 of 870 be an unusually low number in this case?

Binomial

$n = 870$

$p = .791$

Is 339 unusually Low?

$P(x \leq 339) = \text{binomcdf}(870, .791, 339) = 0$

This is unusually low.

X

Summary

- Describe/identify binomial distributions
- Calculate probabilities with binomcdf and binompdf
- Mean and standard deviation for a binomial distributions
- Range Rule of Thumb
- 5% Rule (for usual values)