Section 6.3 - Binomial & Geometric Random Variables (pp. 386-410)

---

1. Binomial Settings and Binomial Random Variables

Definition: A **binomial setting** arises when we perform several independent trials of the same chance process and record the number of times that a particular outcome occurs. The four conditions for a binomial setting are: • Binary? The possible outcomes of each trial can be classified as a "success" or a "failure." • Independent? Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial. • Number? The number of trials *n* must be fixed in advance. • Success? On each trial, the probability *p* of success must be the same. **(BINS)**
Definition: The count X of successes in a binomial setting is a **binomial random variable**. The probability distribution of X is a **binomial distribution** with parameters *n* and *p*, where *n* is the number of trials of the chance process and *p* is the probability of success on any one trial. The possible values of X are whole numbers from 0 to *n*.

Example: Determine whether the random variables below have a binomial distribution. Justify your answer.

(a) Roll a fair die 10 times and let X = the number of 6's.

BINARY – YES: SUCCESS = 6, FAILURE = 6'
IND – YES, ~~REASONABLE TO ASSUME~~
NUMBER – FIXED YES (10)   SUCCESS – YES $P = 1/6$
∴ BINOMIAL

(b) Shoot a basketball 20 times from various distances on the court. Let Y = number of shots made.

BINARY – YES: SUCCESS = MAKE; FAILURE = MISS
IND – YES, REASONABLE TO ASSUME.
NUMBER – YES, 20.   SUCCESS: NO; $P(x)$ Δ'S.
∴ NOT BINOMIAL

(c) Observe the next 100 cars that go by. Let C = color.

BINARY – NO. MORE THAN 2 COLORS.   ∴ NOT BINOMIAL

Check Your Understanding - Complete CYU on p. 389

1
B: ACE, ACE' ✓
I: YES; REPLACE ✓
N: 10 DRAWS ✓
S: $P = 4/25$ ✓
∴ BINOMIAL

2
B: S > 6 ft, F < 6 ft ✓
I: NO; NO REPLACE X
N: 3 PEOPLE ✓
S: NO. X
∴ NOT BINOMIAL

3
B: S = 5, F = 5' ✓
I: YES, DIE ROLL ✓
N: 5 ROLLS ✓
S: P(SUCCESS) Δ'S X
∴ NOT BINOMIAL

2. Binomial Probabilities

Example. Consider rolling a 6-sided die 5 times. Suppose we wanted to know the probability of getting 2 6's. One way this can happen is to roll two 6's in a row and then 3 numbers other than a 6. Let's represent that as 6 6 6' 6' 6'. Since the rolls are *independent* of each other we can compute this probability:

$$P(6\,6\,6'\,6'\,6') = \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right)\left(\tfrac{5}{6}\right)\left(\tfrac{5}{6}\right)\left(\tfrac{5}{6}\right) = 0.0161$$

However, there are many other ways to roll 2 6's. We can compute the number of ways this can happen by using the formula for the combinations of rearranging 6 things taken 2 at a time. The formula for combinations of n things taken k at a time is:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$ This is known as the **binomial coefficient**.

$\frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot3\cdot2\cdot1} = 10$

So in our case, we are looking for $\binom{5}{2} = \frac{5!}{2!\,3!}$ which means to find the probability of 2 6's in 5 rolls is $10(0.0161) = \boxed{0.161}$

FORMULA SHEET

Binomial Probability

If X has the binomial distribution with n trials and probability p of success on each trial, the possible values of X are 0, 1, 2, ..., n. If k is any one of these values,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

K = # OF SUCCESSES
n − K = # OF FAILURES
p = P(SUCCESS)
1 − p = P(FAILURE).

Example (cont). For the scenario above, let X = the number of 6's in 5 rolls.

(a) Is X a binomial random variable?

B: YES: 6' OR 6
I: IND YES: DIE ROLL
N: YES: 5 ROLLS
S: P(6) = 1/6 ✓
∴ BINOMIAL SETTING

(b) Compute P(X = 3)

$P(X=3) = \binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = 0.032$ BINOMPDF(5, 1/6, 3)

(0 1 2 3 4 5) ★

(c) Compute P(X ≥ 3)

$P(X\geq 3) = P(X=3) + P(X=4) + P(X=5) = 0.035$ 1 − BINOMCDF(5, 1/6, 2)

0 1 2 3 4 5

(d) Compute P(X < 3)

$P(X<3) = 1 - P(X\geq 3) = 1 - 0.035 = 0.9645$ BINOMCDF(5, 1/6, 2)

0 1 2 3 4 5

Technology: NTA p. 52; Text p. 394.

- **binompdf**
- **binomcdf**

$\boxed{\text{2ND}}\boxed{\text{DISTR}}$ A! AND B!

Note: Do not use "calculator speak." In answers to binomial questions, include parameters n, p, and k. It is best to show binomial probability formula with numbers substituted in.

How to Find Binomial Probabilities

Step 1: **State the distribution and the values of interest.** Specify a binomial distribution with the number of trials n, success probability p, and the values of the variables clearly defined.
Step 2: **Perform calculations – show work!** Do one of the following: (i) Use the binomial formula to find the desired probability; or (ii) Use the `binompdf` or `binomcdf` command and label each of the inputs.
Step 3: **Answer the question *in context.***
(Example on p. 396)

Check Your Understanding - Complete CYU on p. 397. Use technology.

10 ITEM M/C TEST, 5 ANSWERS $P(\checkmark) = \frac{1}{5}$ $P(\text{INC}) = \frac{4}{5}$

① BINOM:
B - YES ✓ OR X
I - YES RANDOM
N - YES $n = 10$
S - YES $p = 0.2$
YES BINOM.

② $P(X=3) = \binom{10}{3}(0.2)(0.8)^7 = \boxed{.2013}$

③ PASSING: $P(X \geq 6) =$
$P(X=6) + P(X=7) + \dots + P(X=10) =$
$\binom{10}{6}(0.2)^6(0.8)^4 + \binom{10}{7}(0.2)^7(0.8)^3 + \dots$
$= \boxed{0.0064}$ .64% CHANCE OF PASSING
YES - SURPRISED.

---

3. Mean and Standard Deviation of a Binomial Distribution

Mean and Standard Deviation of a Binomial Random Variable

If a count X has the binomial distribution with number of trials *n* and ~~probability of success~~ *p*, the **mean** and **standard deviation** of X are

$$\mu_X = np$$

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

ON FORMULA SHEET

Note: these formulas are only for *binomial distributions.*

Example. The makers of a diet cola claim that its taste is indistinguishable from the full-calorie version of the same cola. To investigate, an AP Stats student prepared small samples of each type of cola in identical cups. Then she had volunteers taste each cola in a random order and try to identify which one was the regular cola. Overall, 23 of the 30 subjects made the correct identification. If we assume the volunteers really could not tell the difference, then each one was guessing with a ½ chance of being correct. Let X = the number of volunteers who correctly identify the colas.

(a) Explain why X is a binomial random variable.

B - YES: ✓ OR X
I - YES - GUESSES ARE IND.
N - YES 30 TRIALS ✓
S : P(SUCCESS) = 0.5 ✓

(b) Find the mean and standard deviation of X. Interpret each value in context.

$\mu_X = (30)(0.5) = 15$

$\sigma_X = \sqrt{30(0.5)(0.5)} = 2.74$

IF WE REPEAT EXP. MANY TIMES, WE EXPECT AVG # OF CORRECT GUESSES TO BE 15 AND THIS # WOULD VARY FROM THE MEAN BY ABOUT 2.74 ON AVG.

(c) Of the 30 volunteers, 23 made correct identifications. Does this give convincing evidence that the volunteers can taste the difference between the diet and regular colas?

$P(X \geq 23) = 1 - \text{BINOMCDF}(30, 0.5, 22) = 0.0026$

THIS MEANS THERE IS ONLY A SMALL CHANCE OF 23 OR MORE CORRECT GUESSES. $\therefore$ WE HAVE CONVINCING EVIDENCE THEY CAN TELL THE DIFFERENCE.

Check Your Understanding - Complete CYU on p. 400.

<u>QUIZ</u> $p = 0.2 \quad n = 10$

① $\mu_X = (10)(0.2) = 2 \Rightarrow$ ON AVG, IF MANY STUDENTS TAKE QUIZ, WE WOULD EXPECT STUDENTS TO GET 2 CORRECT.

② $\sigma_X = \sqrt{10(.2)(.8)} \approx 1.265$. IF MANY STUDENTS TAKE QUIZ, WE EXPECT INDIVIDUAL SCORES TO DIFFER FROM MEAN OF 2 CORRECT BY ABOUT 1.265

③ $\cancel{P(X > 2 + 2(1.265)) = P(X \geq 4.53)} = 1 - P(X \leq 4) = 1 - 0.9672 =$ ANSWERS 0.0328

4. Binomial Distributions in Statistical Sampling

A very common application of the binomial distribution in statistics is when we are counting the number of times a particular outcome occurs in a random sample from some population, for example, the number of defective CDs in a sample of size 10 from a population of 10,000. In cases like this, the sampling is almost always done *without replacement*. This means the trials are no longer *independent*.

However, if the sample is a small fraction of the population, the lack of independence does not have much effect on the probabilities we calculate.

10% Condition

When taking an SRS of size *n* from a population of size *N*, we can use a binomial distribution to model the count of success in the sample as long as

$$n \leq \frac{1}{10} N$$

[illegible]

Example. Suppose a drawer contains 8 AAA batteries but only 6 of them are good. You need to choose 4 for your graphing calculator. If you randomly select 4 batteries, what is the probability that all 4 of the batteries will work? Explain why the answer isn't $P(X = 4) = \binom{4}{4}(0.75)^4(0.25)^0$. (The actual probability is 0.2143)

SINCE SAMPLING W/O REPLACEMENT, TRIALS ARE <u>NOT</u> INDEPENDENT. WE CAN IGNORE THIS PROBLEM IF $n \leq \frac{1}{10} N$ BUT WE ARE SAMPLING 50% OF POP. (4/8) SO IT IS NOT REASONABLE TO IGNORE LACK OF INDEPENDENCE + USE BINOM. DIST. THIS IS WHY PROB'S ARE SO DIFFERENT.

5. Geometric Random Variables

Definition: A **geometric setting** arises when we perform several independent trials of the same chance process and record the number of trials until a particular outcome occurs. The four conditions for a binomial setting are:

- Binary? The possible outcomes of each trial can be classified as a "success" or a "failure."
- Independent? Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial.
- Trials? The goal is to count the number of trials until the first success occurs.
- Success? On each trial, the probability p of success must be the same.

(BITS)

Definition: The number of trials Y that it takes to get a success in a geometric setting is a **geometric random variable**. The probability distribution of Y is a **geometric distribution** with parameter p, the probability of success on any one trial. The possible values of Y are 1, 2, 3, ...

Example. The random variable of interest in this example is Y = number of attempts it takes to roll doubles one time. Each attempt is one trial of the chance process. Is this a geometric setting?

BINARY : YES - SUCCESS = DOUBLES; FAIL = NOT.

IND : YES - DIE ROLLS

TRIALS: YES, WE ARE COUNTING # OF TRIALS UNTIL DOUBLES

SUCCESS : $P(\text{SUCCESS}) = \frac{1}{6}$ ALWAYS

Geometric Probability Formula

NOT ON FORMULA SHEET

If Y has the geometric distribution with probability p of success on each trial, the possible values of Y are 1, 2, 3, ... If k is any one of these values,

$$P(Y = k) = (1-p)^{k-1}p$$

K TRIALS, 1 SUCCESS ⇒ K - 1 FAILURES.

Example. From the previous example, find

2 FAILS, 1 SUCCESS.

(a) the probability it takes 3 turns to roll doubles

$$\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) = 0.116$$

(b) the probability it takes more than 3 turns to roll doubles and interpret the value in context.

$$P(Y>3) = 1 - P(Y \le 3) = 1 - [P(Y=3) + P(Y=2) + P(Y=1)]$$

$$= 1 - \left[\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^1\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)^0\left(\frac{1}{6}\right)\right] = 0.5787$$

IF WE REPEATEDLY ROLL UNTIL WE GET DOUBLES, WE EXPECT TO HAVE TO ROLL MORE THAN 3 TIMES ≈ 58% OF THE TIME.

Technology: NTA p. 53; Text p. 406.

- **geompdf**

$\boxed{2ND}$ $\boxed{DISTR}$ F: GEOMPDF (P, K) $P(Y=3) \Rightarrow$ GEOMPDF$(\frac{1}{6}, 3)$ $\boxed{0.1157}$

- **geomcdf**

$\boxed{2ND}$ / $\boxed{DISTR}$ F: GEOMCDF (P, K)

$P(Y>3) = 1 - P(Y \leq 3) =$

$1 -$ GEOMCDF$(\frac{1}{6}, 3) = \boxed{0.5787}$

Mean (Expected Value) of a Geometric Random Variable
If *Y* is a geometric random variable with probability of success *p* on each trial, then its **mean** (expected value) is $E(Y) = \mu_Y = \frac{1}{p}$ That is, the expected number of trials required to get the first success is $1/p$.

~~**Check Your Understanding**: Complete CYU on p. 408.~~

1) 3 -

HW: Read pp. 386-410, 65, 66, 68-89 odd, 101-105.