

SECTION 2 | QUESTIONS WITH SOLUTIONS

The Palpatine Poll

Bernoulli, binomial, and a probabilistic model that turns a secret into a number

HOW TO USE THIS HANDOUT

Every topic starts with the idea in plain words, followed by questions with complete, step-by-step solutions. The boxes marked **Additional Example** are extra, easy examples that help the idea sink in. Try each question yourself before you read its solution.

THE THREE RULES FROM SECTION 1 (REVIEW)

NOT

$$P(E^C) = 1 - P(E)$$

OR

exclusive: $P(E) + P(F)$
 otherwise: $P(E) + P(F) - P(E \text{ and } F)$

AND

independent: $P(E) \cdot P(F)$
 otherwise: $P(E) \cdot P(F | E)$

THE BIG IDEAS OF THIS SECTION

BARS (PMF)

One bar per value, as tall as its chance. The bars add up to 1.

BERNOULLI

$X \sim \text{Ber}(p)$: one trial, 1 if it happens, 0 if not.
 $P(X = 1) = p$.

BINOMIAL

$X \sim \text{Bin}(n, p)$: the number of successes in n independent trials, each with chance p . Average: $n \times p$.

PROBABILISTIC MODEL

A story with random variables that makes data. Run it to make fake data. Count to answer questions. Multiply along the story for the chance of one outcome.

Notation. $X \sim \text{Ber}(p)$ is read “ X is Bernoulli with chance p ”, and $X \sim \text{Bin}(n, p)$ is read “ X is binomial with n trials and chance p ”. The letter H will always mean the number of yeses to the harmless questions, and $P(H = k)$ is the chance of exactly k harmless yeses.

Topic 1 Random variables and their bars

IN PLAIN WORDS

A **random variable** is a number that comes out of something random, such as the number a die shows. We name it with a capital letter like X .

The most important thing to know about a random variable is its **bars** (officially the probability mass function, or PMF):

- there is one bar for every value the variable can take,
- each bar is as tall as the chance of that value,
- all the bars together add up to exactly 1. If they do not, something is wrong.

QUESTION 1

What are the bars of a fair die?

SOLUTION

Let X be the number the die shows.

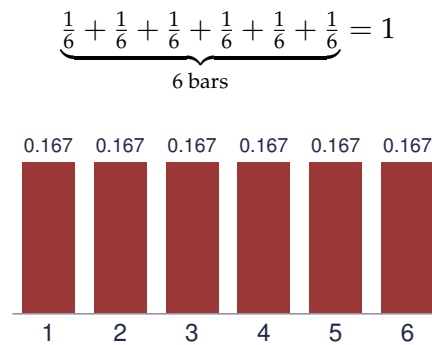
Step 1: list the values

X can be 1, 2, 3, 4, 5, 6, so there are six bars.

Step 2: find the height of each bar

“Fair” means every face is equally likely. Call that common chance c . The six chances must add up to 1:

$$6c = 1 \implies c = \frac{1}{6} \approx 0.167$$

Step 3: check

The bars of one fair die. Value shown (under each bar) and its chance (above).

Note. Adding bars answers questions. For example, the chance the die shows 5 or 6 is $P(X = 5) + P(X = 6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$. We may add because one roll cannot show 5 and 6 together (the events are exclusive).

ANSWER Six bars, at the values 1, 2, 3, 4, 5, 6, each of height $\frac{1}{6} \approx 0.167$. Together they add up to 1.

ADDITIONAL EXAMPLE 1

easy practice

- (a) Flip one fair coin. Let $X = 1$ if it lands heads and $X = 0$ if it lands tails. What are the bars of X ?
- (b) A friend says some random variable has bars of height 0.5, 0.3 and 0.4. Can that be right?

SOLUTION**Part (a)**

X has two values, 0 and 1. The coin is fair, so the two chances are equal and add to 1: each is $\frac{1}{2}$. So $P(X = 0) = 0.5$ and $P(X = 1) = 0.5$.



The bars of one coin flip (0 = tails, 1 = heads).

Part (b)

Add the bars: $0.5 + 0.3 + 0.4 = 1.2$. The total must be exactly 1, and 1.2 is more than 1, so these cannot be the bars of any random variable. If the friend had said 0.5, 0.3, 0.2, the total would be 1.0 and that would be fine.

ANSWER (a) Two bars of height 0.5 at the values 0 and 1. (b) No, because the bars add up to 1.2, not 1.

Topic 2 Bernoulli and binomial random variables**IN PLAIN WORDS**

Bernoulli: one trial. A trial is one try of something that either happens or does not. We write 1 if it happens and 0 if it does not. The number p is the chance that it happens. So $X \sim \text{Ber}(p)$ has just two bars: $P(X = 1) = p$ and $P(X = 0) = 1 - p$.

Binomial: count the successes in many trials. Repeat the same trial n times. Two conditions must hold: the trials are *independent* (one does not change another) and each has the *same* chance p . Let X count how many succeeded. Then $X \sim \text{Bin}(n, p)$, and X can be $0, 1, \dots, n$. Its average is $n \times p$.

The bars of a binomial.

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

In words: $\binom{n}{k}$ is the number of ways to choose *which* k of the n trials are the successes (for example $\binom{3}{2} = 3$, because with three trials the two successes can be trials 1&2, 1&3 or 2&3). Each particular way has chance $p^k(1 - p)^{n-k}$ by the AND rule. Different ways cannot happen together, so by the OR rule we add them, which is the same as multiplying by $\binom{n}{k}$.

QUESTION 2

one example of each

A bus is on time with chance 0.8 on any given day, and different days do not affect each other. Give one Bernoulli and one binomial random variable from this story, and write their bars.

SOLUTION**Step 1: the Bernoulli variable**

Look at *one* bus on *one* day: it is on time (success) or not. Let $X = 1$ if it is on time and $X = 0$ if not. One trial with two outcomes gives $X \sim \text{Ber}(0.8)$:

$$P(X = 1) = 0.8, \quad P(X = 0) = 1 - 0.8 = 0.2$$

Step 2: the binomial variable

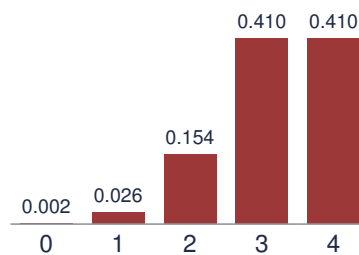
Now watch four days. Let Y be the number of days (out of 4) on which the bus is on time. These are 4 independent trials with the same chance 0.8, and Y counts the successes, so $Y \sim \text{Bin}(4, 0.8)$ and Y can be 0, 1, 2, 3, 4.

Step 3: the bars of Y

Use $P(Y = k) = \binom{4}{k} 0.8^k 0.2^{4-k}$.

k	$\binom{4}{k}$	0.8^k	0.2^{4-k}	$P(Y = k)$
0	1	1	0.0016	$1 \cdot 1 \cdot 0.0016 = 0.0016$
1	4	0.8	0.008	$4 \cdot 0.8 \cdot 0.008 = 0.0256$
2	6	0.64	0.04	$6 \cdot 0.64 \cdot 0.04 = 0.1536$
3	4	0.512	0.2	$4 \cdot 0.512 \cdot 0.2 = 0.4096$
4	1	0.4096	1	$1 \cdot 0.4096 \cdot 1 = 0.4096$

Check: $0.0016 + 0.0256 + 0.1536 + 0.4096 + 0.4096 = 1.0000$.



The bars of $Y \sim \text{Bin}(4, 0.8)$: number of on-time days out of four.

Step 4: the average

$$n \times p = 4 \times 0.8 = 3.2$$

Note. The most likely values are 3 and 4 (chance 0.4096 each). That makes sense: the bus is usually on time, so most weeks have three or four on-time days. Notice that the average, 3.2, is not itself a possible value. An average does not have to be one.

ANSWER $X \sim \text{Ber}(0.8)$ with bars 0.2 (at 0) and 0.8 (at 1). $Y \sim \text{Bin}(4, 0.8)$ with bars 0.0016, 0.0256, 0.1536, 0.4096, 0.4096 for $k = 0, 1, 2, 3, 4$, and average 3.2.

ADDITIONAL EXAMPLE 2*easy practice*

Flip a fair coin three times. Let Y be the number of heads. Find the bars of Y by listing every outcome, and then check them with the binomial formula.

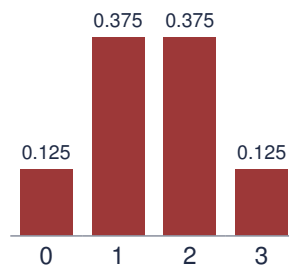
SOLUTION**Step 1: list all outcomes**

Each flip is H or T, so there are $2 \cdot 2 \cdot 2 = 8$ outcomes. Each particular sequence has chance $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$ (AND rule, the flips are independent).

Heads k	Outcomes with k heads	How many	$P(Y = k)$
0	TTT	1	$\frac{1}{8} = 0.125$
1	HTT, THT, TTH	3	$\frac{3}{8} = 0.375$
2	HHT, HTH, THH	3	$\frac{3}{8} = 0.375$
3	HHH	1	$\frac{1}{8} = 0.125$

Step 2: check with the formula

$Y \sim \text{Bin}(3, \frac{1}{2})$, so $P(Y = k) = \binom{3}{k}(\frac{1}{2})^3 = \binom{3}{k}/8$. Since $\binom{3}{0} = 1$, $\binom{3}{1} = 3$, $\binom{3}{2} = 3$, $\binom{3}{3} = 1$, we get $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$, the same as the list.



The bars of the number of heads in three fair flips.

Note. The bars add to $\frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$, and the average is $3 \times \frac{1}{2} = 1.5$. The “number of ways” column is exactly the number $\binom{n}{k}$: that is why the formula works.

ANSWER $P(Y = 0), P(Y = 1), P(Y = 2), P(Y = 3) = 0.125, 0.375, 0.375, 0.125$. Average 1.5.

Topic 3 A probabilistic model is a story that makes data

IN PLAIN WORDS

A **probabilistic model** is a story told with random variables: a list of random steps, in order, that ends in some data. A model can be used in three ways.

- **Run it.** Follow the story with real random choices and you get one piece of fake data. Run it again to get another.
- **Count.** Run it many times and count how often something happens. The fraction answers questions such as “how likely is this?”.
- **Multiply along the story.** To get the exact chance of one particular outcome, multiply the chances of the steps that lead to it.

QUESTION 3

run, count, multiply

A story: flip a fair coin. If it lands heads, roll a fair die. If it lands tails, do nothing.

(a) What is the chance of “heads, and then a 6”?

(b) If the story is run 12,000 times, about how many runs end with heads and a 6?

SOLUTION

Let E be the event “the coin lands heads” and F the event “the die shows 6”. The die is only rolled after heads, so the die step depends on the coin step.

Part (a): multiply along the story

Chance of the first step: $P(E) = \frac{1}{2}$. Given heads, the die is rolled, so $P(F | E) = \frac{1}{6}$. By the AND rule:

$$P(E \text{ and } F) = P(E) \cdot P(F | E) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12} \approx 0.0833$$

Part (b): run it and count

Out of 12,000 runs we expect the fraction $\frac{1}{12}$ to end in “heads and 6”:

$$12,000 \times \frac{1}{12} = 1,000$$

Note. This is the “run and count” idea. A computer that really runs the story 12,000 times will not get exactly 1,000 every time. It might get 985 or 1,012. This small wobble is called *sampling noise*. It is not a mistake, and it gets smaller (relative to the total) when we run more times.

ANSWER (a) $\frac{1}{12} \approx 0.083$. (b) About 1,000 runs.

ADDITIONAL EXAMPLE 3*easy practice*

The bus from Topic 2 is on time with chance 0.8 each day, independently. Tell the story for Monday and Tuesday and answer: (a) both days on time, (b) exactly one day on time, (c) neither day on time. Then say how many of 1,000 simulated weeks you expect in each case.

SOLUTION**The story**

Step 1, Monday: on time with chance 0.8, late with chance 0.2. Step 2, Tuesday: the same, independent of Monday.

Multiply along the story

$$\begin{aligned}
 \text{(a) both on time: } & 0.8 \times 0.8 = 0.64 \\
 \text{(b) exactly one on time: } & \underbrace{0.8 \times 0.2}_{\text{Mon on time, Tue late}} + \underbrace{0.2 \times 0.8}_{\text{Mon late, Tue on time}} = 0.16 + 0.16 = 0.32 \\
 \text{(c) neither on time: } & 0.2 \times 0.2 = 0.04
 \end{aligned}$$

In (b) we add because the two ways cannot happen together (OR rule, exclusive).

Check and count

$0.64 + 0.32 + 0.04 = 1.00$, so every possible week is covered. Out of 1,000 simulated weeks we expect about 640, 320 and 40 weeks respectively.

ANSWER (a) 0.64 (b) 0.32 (c) 0.04. About 640, 320 and 40 weeks out of 1,000.

Topic 4 The game: Bernoulli, binomial, or neither?**HOW TO DECIDE**

Ask three questions about the random variable X .

1. What is X : a yes/no answer, a count, or something else?
2. If it is a count: how many trials n are there? Are they independent, and do they all have the same chance p ?
3. Are we counting *successes*?

One yes/no trial gives **Bernoulli**, $\text{Ber}(p)$. The number of successes in n independent trials with the same p gives **binomial**, $\text{Bin}(n, p)$. Any other random variable (not a yes/no trial and not a count of successes) is **neither**.

The family is decided by *the question X asks*, not by the object. You will see in Question 10 that one die can give all three answers, depending on what X is.

How the game runs. Each round is a short story with one random number X . Everybody thinks quietly first, then types an answer in the chat, and the room argues before the answer is revealed. There are five rounds (a team's five games, tomorrow's weather, a birthday, a chatbot, a die) and a sixth round on the harmless questions in Topic 7.

Two review examples

QUESTION 4*review example*

A footballer takes a penalty kick. Let X describe whether the kick goes in. Is X Bernoulli, binomial, or neither?

SOLUTION**Apply the three questions**

(1) X is a yes/no answer: the kick goes in or it does not. (2) There is exactly one trial. (3) We record 1 if it goes in and 0 if not.

One trial with two outcomes is a Bernoulli variable. If p is the chance the kick goes in, then

$$X \sim \text{Ber}(p), \quad P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Note. Someone might say “binomial”. That is also correct, because $\text{Bin}(1, p)$ is exactly the same thing as $\text{Ber}(p)$: a count of successes in one trial is just 0 or 1.

ANSWER Bernoulli, $X \sim \text{Ber}(p)$ (equivalently $\text{Bin}(1, p)$).

QUESTION 5*review example*

Flip a fair coin 10 times. Let X be the number of heads. Which family is X in, and what are $P(X = 5)$ and the average of X ?

SOLUTION**Apply the three questions**

(1) X is a count of heads. (2) There are $n = 10$ flips, each independent of the others, each with the same chance $p = \frac{1}{2}$ of heads. (3) We count successes (heads). So $X \sim \text{Bin}(10, \frac{1}{2})$.

The chance of exactly 5 heads

$$P(X = 5) = \binom{10}{5} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{\binom{10}{5}}{2^{10}} = \frac{252}{1024} \approx 0.246$$

Here $\binom{10}{5} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 252$ is the number of ways to choose which 5 of the 10 flips are heads.

The average

$$n \times p = 10 \times \frac{1}{2} = 5$$

Note. For comparison, getting no heads at all has chance $\binom{10}{0}/1024 = \frac{1}{1024} \approx 0.001$. The bars pile up around 5, the average.

ANSWER Binomial, $X \sim \text{Bin}(10, \frac{1}{2})$. $P(X = 5) = \frac{252}{1024} \approx 0.246$, and the average is 5.

The game, rounds 1 to 5

QUESTION 6*game round 1, warm-up*

Your favourite sports team plays 5 games. Its record is WIN, WIN, LOSS, LOSS, WIN. Let X be how many wins the team gets. Is X Bernoulli, binomial, or neither? What is X for this record? (For the last part only, suppose the team wins each game with chance $p = 0.6$; this number is just for practice.)

SOLUTION**Apply the three questions**

(1) X is a count of wins. (2) There are $n = 5$ games. We model them as independent, each with the same chance p of a win. (3) We count successes (wins). So

$$X \sim \text{Bin}(5, p).$$

The value of X for this record

The record WIN, WIN, LOSS, LOSS, WIN has three wins, so $X = 3$.

Practice with $p = 0.6$: the chance of exactly this record

Each game is independent, so we multiply along the story (AND rule). A win has chance 0.6 and a loss has chance $1 - 0.6 = 0.4$:

$$P(\text{WWLLW}) = 0.6 \cdot 0.6 \cdot 0.4 \cdot 0.4 \cdot 0.6 = 0.6^3 \cdot 0.4^2 = 0.216 \cdot 0.16 = 0.03456$$

Practice with $p = 0.6$: the chance of $X = 3$

Other records also have exactly three wins. We list them all:

WWWLL, WWLWL, **WWLLW**, WLWWL, WLWLW, WLLWW, LWWWL, LWLWL, LWLWW, LLWWW

That is $\binom{5}{3} = 10$ records. Each one has the same chance, $0.6^3 \cdot 0.4^2 = 0.03456$, because multiplication does not care about the order. The records cannot happen together, so we add (OR rule):

$$P(X = 3) = 10 \times 0.03456 = 0.3456$$

The same answer comes from the formula $\binom{5}{3} 0.6^3 0.4^2 = 10 \cdot 0.216 \cdot 0.16 = 0.3456$. The average number of wins is $5 \times 0.6 = 3$.

Note. Why not Bernoulli? Each single game is a yes/no trial, but X counts over five games, and a Bernoulli variable only has the values 0 and 1. Is “independent, same p ” exactly true for real games? Not quite: opponents differ, and teams have good and bad runs. A model is a story we choose; the binomial is the simplest good story. This makes a nice discussion for the room.

ANSWER Binomial, $X \sim \text{Bin}(5, p)$. For this record $X = 3$. With $p = 0.6$: the exact record WWLLW has chance 0.03456 and $P(X = 3) = 0.3456$.

QUESTION 7*game round 2, warm-up*

The weather forecast for tomorrow says there is a 30% chance of rain (this number is just for practice). Let $X = 1$ if it rains tomorrow and $X = 0$ if not. Is X Bernoulli, binomial, or neither? Then think about two other random variables: (a) the number of rainy days in the next 7 days, and (b) the amount of rain tomorrow in millimetres.

SOLUTION**Apply the three questions to X**

(1) X is a yes/no answer: it rains or it does not. (2) There is one trial (one day). (3) $X = 1$ means rain. So

$$X \sim \text{Ber}(0.3), \quad P(X = 1) = 0.3, \quad P(X = 0) = 0.7.$$

(a) The number of rainy days in 7 days

This is a count of successes, so it is binomial if the 7 days are independent and have the same chance $p = 0.3$. Then it is $\text{Bin}(7, 0.3)$ with average $7 \times 0.3 = 2.1$ rainy days. Real weather is not independent from day to day (rainy days come in streaks), so this is a modeling choice, just like the team's games.

(b) The amount of rain in millimetres

This is not a yes/no trial and not a count of successes, so it is **neither**.

ANSWER X is Bernoulli, $X \sim \text{Ber}(0.3)$. (a) Binomial $\text{Bin}(7, 0.3)$ if we model the days as independent with the same chance. (b) Neither.

QUESTION 8*game round 3, a tricky one*

You have 30 classmates. Let X be the number of them who share your birthday. Which family is X in? What are $P(X = 0)$ and $P(X \geq 1)$?

SOLUTION**Assumptions (say them out loud)**

Each classmate's birthday is one of 365 days, all equally likely, and different classmates do not affect each other. (We ignore leap days and twins.)

Apply the three questions

(1) X is a count. (2) There are $n = 30$ trials, one per classmate. Each classmate "succeeds" (shares your birthday) with the same chance $p = \frac{1}{365}$, independently. (3) We count successes. So

$$X \sim \text{Bin}\left(30, \frac{1}{365}\right).$$

The average

$$n \times p = \frac{30}{365} \approx 0.082$$

In a typical class, fewer than one person shares your birthday.

The chance that nobody shares it

Each classmate fails to share your birthday with chance $1 - \frac{1}{365} = \frac{364}{365}$. All 30 must fail, independently (AND rule):

$$P(X = 0) = \left(\frac{364}{365}\right)^{30} \approx 0.921$$

The chance that at least one person shares it (NOT rule)

$$P(X \geq 1) = 1 - P(X = 0) \approx 1 - 0.921 = 0.079$$

The Poisson alternative

When n is large and p is small, a Poisson variable with average $\lambda = n \times p = \frac{30}{365} \approx 0.082$ is a very close approximation. It gives $P(X = 0) \approx e^{-0.082} \approx 0.921$, the same number to three decimals. So “Poisson” is also a good answer; both are right.

Note. This is *not* the “birthday paradox”. Here *your* birthday is fixed and each classmate is compared only with you. The paradox asks about any two people in the class sharing a birthday, which is a different question.

ANSWER Binomial, $X \sim \text{Bin}(30, \frac{1}{365})$ (Poisson with $\lambda = \frac{30}{365}$ is a close approximation). $P(X = 0) \approx 0.921$ and $P(X \geq 1) \approx 0.079$.

QUESTION 9

game round 4, a tricky one

You ask a chatbot a question. Let X describe whether its reply starts with “Great question!”. Is X Bernoulli, binomial, or neither?

SOLUTION**Apply the three questions**

(1) X is a yes/no answer about one reply: it starts with “Great question!” or it does not. (2) There is one trial (one reply). (3) Write $X = 1$ for yes and $X = 0$ for no.

So $X \sim \text{Ber}(p)$, where p is the chance that the chatbot starts a reply that way.

Note. It does not matter how fancy the chatbot is. The family depends on the *question about the reply*, and one yes/no question about one reply is one trial. If we instead asked the same question about 100 independent replies and counted the yeses, that count would be $\text{Bin}(100, p)$.

ANSWER Bernoulli, $X \sim \text{Ber}(p)$.

QUESTION 10*game round 5, a tricky one*

Roll one fair die and let X be the number it shows. Is X Bernoulli, binomial, or neither? Then show how changing what X means changes the family.

SOLUTION**Apply the three questions**

(1) X is a number from 1 to 6. It is not a yes/no answer, and it is not a count of successes. (2) There is no “number of successes out of n trials” to count. (3) So neither description fits.

Why the tempting answers are wrong

- “Bernoulli, because it is one roll.” A Bernoulli variable has only *two* values, 0 and 1. The die has six.
- “Binomial, because there are six faces.” A binomial variable counts successes over several trials. Here there is one roll, and we are reading off a value, not counting successes.

X is simply a random variable with six equally tall bars (each $\frac{1}{6}$, as in Question 1). It is **neither**.

Change what X means, and the family changes

What X means	Family
The number the die shows	Neither (six bars)
$X = 1$ if the die shows a 6, otherwise 0	$\text{Ber}(\frac{1}{6})$
The number of 6's in 10 rolls	$\text{Bin}(10, \frac{1}{6})$, average $\frac{10}{6} \approx 1.67$

ANSWER Neither. The die is the same object in all three rows; only the question X asks has changed.

ADDITIONAL EXAMPLE 4*easy practice*

A student guesses answers on a quiz where every question has 4 options and exactly one is right. Classify each random variable (Bernoulli, binomial, or neither) and give its average if it has one in the formulas of this section.

- X is 1 if the student guesses *one* question correctly, 0 otherwise.
- Y is the number of correct guesses out of 10 questions.
- Z is the day of the week on which the quiz takes place (1, ..., 7).

SOLUTION

(a)

One yes/no trial. A guess is right with chance $\frac{1}{4}$, so $X \sim \text{Ber}(\frac{1}{4})$.

(b)

There are $n = 10$ trials. Each guess is independent and right with the same chance $p = \frac{1}{4}$, and we count the right ones. So $Y \sim \text{Bin}(10, \frac{1}{4})$ with average $10 \times \frac{1}{4} = 2.5$.

(c)

Z takes seven values. It is not a yes/no answer and not a count of successes, so it is **neither**.

ANSWER (a) Bernoulli $\text{Ber}(\frac{1}{4})$. (b) Binomial $\text{Bin}(10, \frac{1}{4})$, average 2.5. (c) Neither.

Topic 5 The hook: how many people support Palpatine?

THE SETTING

Emperor Palpatine is the evil Emperor who rules the galaxy in the Star Wars films. (You do not need to know the films.) He punishes anybody who tells a pollster that they do not like him. So even people who do *not* support him say “yes, I support him”, out of fear.

Our question: how many people really support Palpatine? Nobody will answer it honestly, so asking directly will not work. Let us see why, and then find a cleverer way.

QUESTION 11

What goes wrong if we simply ask everybody, “Do you support Palpatine?”

SOLUTION

Step 1: who says “yes”?

Two kinds of people say yes: (i) honest supporters, and (ii) frightened non-supporters who say yes to stay safe. A direct poll counts both as “support”, and nothing in a “yes” tells them apart.

Step 2: a numerical illustration

(The numbers here are made up for illustration.) Suppose the true support is 0.09, so 0.91 of people do not support him.

- If *all* non-supporters are scared into saying yes: the yes-fraction is $0.09 + 0.91 \times 1 = 1.00$. The poll says “everyone supports him”.
- If 80% of non-supporters say yes: the yes-fraction is $0.09 + 0.91 \times 0.8 = 0.09 + 0.728 = 0.818$.

In both cases the poll is far above the true 0.09.

Step 3: the lesson

A direct poll counts every frightened “yes” as support, so it *overestimates* the support. We need a poll in which a “yes” can never be held against anybody.

ANSWER A direct poll mixes honest supporters with frightened non-supporters, so the share of “yes” is much too high and does not tell us the real support.

ADDITIONAL EXAMPLE 5*easy practice*

A teacher has 30 students, and 6 of them did not do their homework. The teacher asks, “Raise your hand if you did not do your homework.” Only half of those 6 students raise a hand. What does the poll say, what is the truth, and what is different from the Palpatine poll?

SOLUTION**The truth**

$\frac{6}{30} = 0.20$, which is 1 student in 5.

What the poll says

Half of 6 is 3 hands, so the poll says $\frac{3}{30} = 0.10$, which is 1 student in 10.

What is different

The poll *undercounts*: shame makes people say “no”. In the Palpatine poll fear makes people say “yes”, so it *overcounts*. The direction is different, but in both cases asking a hidden-feeling question directly gives the wrong number.

ANSWER Poll: 0.10. Truth: 0.20. Shame makes a direct poll undercount; fear (Palpatine) makes it overcount.

Topic 6 The trick: hide the question in a list**THE TRICK**

Put the secret question inside a list of harmless questions, and let each person report only *how many* of the questions they answered “yes” to, never *which* ones.

1. Do you like surprises?
2. Do you like the beach more than the mountains?
3. Do you like paper books more than e-books?
4. Would you rather have more time than more money?
5. Do you like holidays at home more than abroad?
- 6. Do you support Palpatine?**

The first five are harmless, and each is answered “yes” about half the time (50/50). The sixth is the secret. The person reports one number, for example “3”.

QUESTION 12

A person reports “3”. Do you know whether they support Palpatine?

SOLUTION

Step 1: list every way to get a 3

- Three harmless yeses, and “no” to the Palpatine question: $3 + 0 = 3$.
- Two harmless yeses, and “yes” to the Palpatine question: $2 + 1 = 3$.

Step 2: conclude

Both stories give the same number, so the number 3 cannot tell us which one happened. We do *not* know. That is the whole trick: nobody can point to a single person and say what they secretly answered, and that is exactly what protects them.

Note. Optional check with numbers. Suppose the true support is 0.09. Using $P(H = 2) = P(H = 3) = \frac{10}{32} = 0.3125$ (we compute these bars in Question 15):

chance of saying 3 *and* supporting = $0.3125 \times 0.09 = 0.028125$;

chance of saying 3 *and* not supporting = $0.3125 \times 0.91 = 0.284375$.

Given that someone said 3, the chance they support him is $\frac{0.028125}{0.028125 + 0.284375} = \frac{0.028125}{0.3125} = 0.09$, which is the same as the support in the whole crowd. So hearing “3” teaches us nothing new about that person. The crowd as a whole, however, still carries the information, and Topic 8 shows how to read it.

ANSWER No. A 3 can mean three harmless yeses (not a supporter) or two harmless yeses plus the secret yes (a supporter).

QUESTION 13

this trick is real

In Vietnam in 2023, researchers hid a question about cheating with ChatGPT inside a list like this one. When students were asked directly, 9.6% said yes. The list method gave 23.7%. What does this tell us? (Source: Nguyen and Goto, *Unmasking academic cheating behavior in the artificial intelligence era*, Education and Information Technologies 29, 2024.)

SOLUTION

Step 1: read the two numbers in everyday language

9.6% is about 1 student in 10. 23.7% is about 1 student in 4.

Step 2: the gap

$$23.7\% - 9.6\% = 14.1 \text{ percentage points}$$

Step 3: interpret

In a list, a “yes” hides among the harmless yeses, so people can answer honestly without anyone being able to point at them. If we trust the list number as closer to the truth (which is the idea of the method), then the direct question caught only

$$\frac{9.6}{23.7} \approx 0.41$$

of the students who had really done it: roughly 4 out of every 10 admitted it when asked directly, and about 6 out of 10 did not.

Step 4: the bigger picture

The trick works for anything people hide, out of *fear* (Palpatine) or out of *shame* (cheating).

ANSWER Direct questioning gave about 1 in 10 and the list gave about 1 in 4. People hide the truth when asked directly, and the list lets many of them answer honestly.

ADDITIONAL EXAMPLE 6*easy practice*

A short survey hides one secret question in a list of four: *Do you like tea?* *Do you like cricket?* *Do you walk to work?* and the secret *Did you skip homework?* A person reports how many of the four are “yes”. For each reported number 0, 1, 2, 3, 4, can we tell what they answered to the secret question?

SOLUTION

Let the harmless yeses be H (between 0 and 3) and the secret answer S (1 for yes, 0 for no). The reported number is $H + S$.

Reported	Ways it can happen	Do we know S ?
0	$H = 0$ and $S = 0$ only	Yes: $S = 0$
1	$H = 1, S = 0$ or $H = 0, S = 1$	No
2	$H = 2, S = 0$ or $H = 1, S = 1$	No
3	$H = 3, S = 0$ or $H = 2, S = 1$	No
4	$H = 3$ and $S = 1$ only	Yes: $S = 1$

Note. Only the two ends, 0 and 4, give a person away. Everything in the middle is protected. Remember this: Topic 9 asks exactly which answers expose a person in our Palpatine poll.

ANSWER Reported 1, 2, 3: we cannot tell. Reported 0 means “no” and reported 4 means “yes” to the secret question.

Topic 7 Warm-up: the five harmless questions**IN PLAIN WORDS**

Forget Palpatine for a moment. One person answers just the five harmless questions, each a 50/50 yes. How many of the five do they say yes to? Let

H = the number of yeses to the harmless questions.

This is the first half of the poll story, so it is worth understanding completely. In section you explore it with a *binomial calculator* (you can change n and p and read the bars) and with a *Galton board* (balls falling through pegs). Both appear below.

QUESTION 14*game round 6*

What kind of random variable is H ?

SOLUTION**Apply the three questions**

(1) H is a count of yeses. (2) There are $n = 5$ questions. Each is answered separately, so we treat them as independent, and each has the same chance $p = \frac{1}{2}$ of a yes. (3) We count successes (yeses).

So

$$H \sim \text{Bin}\left(5, \frac{1}{2}\right).$$

Note. Keep this one in mind: it is the heart of the poll. Some people prefer to say “five separate coins”. That describes exactly the same thing: five 50/50 yes/no questions, counted, form $\text{Bin}(5, \frac{1}{2})$.

ANSWER Binomial, $H \sim \text{Bin}(5, \frac{1}{2})$.

QUESTION 15*the binomial calculator*

Find all the bars of H . Which number of yeses is the most likely? How likely is it to get all five yes, and how likely is it to get 0 or 5?

SOLUTION**Step 1: the possible values**

$$H \in \{0, 1, 2, 3, 4, 5\}.$$

Step 2: the formula

$$P(H = k) = \binom{5}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{5-k} = \binom{5}{k} \left(\frac{1}{2}\right)^5 = \frac{\binom{5}{k}}{32}$$

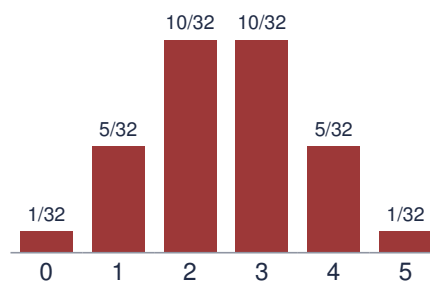
Every particular pattern of five answers has the same chance $\left(\frac{1}{2}\right)^5 = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{32}$, so each bar is just the number of patterns divided by 32.

Step 3: count the patterns

$\binom{5}{k}$ is the number of ways to choose which k of the 5 questions get a yes. For $k = 1$ there are 5 choices (any one question). For $k = 2$ there are $\frac{5 \cdot 4}{2 \cdot 1} = 10$ pairs. For $k = 3$ it is the same as choosing the two “no” questions: 10. And so on.

k yeses	$\binom{5}{k}$ patterns	$P(H = k)$	As a decimal
0	1	$\frac{1}{32}$	0.03125
1	5	$\frac{5}{32}$	0.15625
2	10	$\frac{10}{32}$	0.3125
3	10	$\frac{10}{32}$	0.3125
4	5	$\frac{5}{32}$	0.15625
5	1	$\frac{1}{32}$	0.03125

Check: $1 + 5 + 10 + 10 + 5 + 1 = 32$, so the bars add up to $\frac{32}{32} = 1$.



The bars of $H \sim \text{Bin}(5, \frac{1}{2})$: number of harmless yeses.

Step 4: read off the answers

- Most likely: 2 and 3 yeses, each with chance $\frac{10}{32} = 0.3125$.
- All five yes: $P(H = 5) = \frac{1}{32} \approx 0.031$, about 1 in 32.

- 0 or 5: $\frac{1}{32} + \frac{1}{32} = \frac{2}{32} = \frac{1}{16} = 0.0625$ (OR rule: the two events cannot both happen).

Note. On the calculator you can click a bar to read its exact chance. You can also change n and p : n is the number of questions and p is the chance of a yes. We will move p to 0.9 in Topic 9 and see all the bars slide to the right.

ANSWER $P(H = k) = \frac{1}{32}, \frac{5}{32}, \frac{10}{32}, \frac{10}{32}, \frac{5}{32}, \frac{1}{32}$ for $k = 0, \dots, 5$. Most likely: 2 or 3. All five yes has chance $\frac{1}{32} \approx 0.031$, and 0 or 5 has chance 0.0625.

QUESTION 16

the other two calculator views

The calculator can show three kinds of bars: $P(H = k)$ (exactly k yeses), $P(H \leq k)$ (k or fewer yeses) and $P(H \geq k)$ (k or more yeses). For $H \sim \text{Bin}(5, \frac{1}{2})$:

- (a) Find $P(H \leq 2)$. (b) Find $P(H \geq 4)$.
 (c) Check that $P(H \geq 3) = 1 - P(H \leq 2)$. (d) What is the chance of between 1 and 4 yeses, including both ends?

SOLUTION

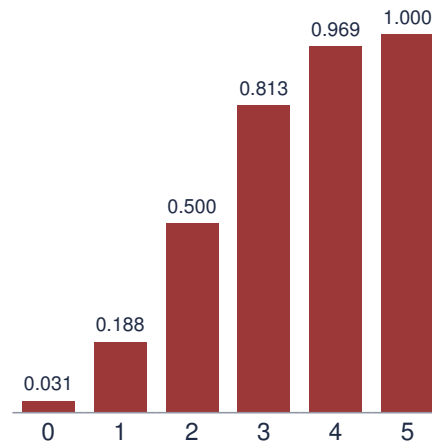
Step 1: what the symbols mean

$P(H \leq k)$ adds the bars from 0 up to k . $P(H \geq k)$ adds the bars from k up to 5. The bars $P(H = k)$ are from Question 15.

Step 2: the full table

k	$P(H = k)$	$P(H \leq k)$	$P(H \geq k)$
0	$\frac{1}{32} = 0.03125$	$\frac{1}{32} = 0.03125$	$\frac{32}{32} = 1$
1	$\frac{5}{32} = 0.15625$	$\frac{6}{32} = 0.1875$	$\frac{31}{32} = 0.96875$
2	$\frac{10}{32} = 0.3125$	$\frac{16}{32} = 0.5$	$\frac{26}{32} = 0.8125$
3	$\frac{10}{32} = 0.3125$	$\frac{26}{32} = 0.8125$	$\frac{16}{32} = 0.5$
4	$\frac{5}{32} = 0.15625$	$\frac{31}{32} = 0.96875$	$\frac{6}{32} = 0.1875$
5	$\frac{1}{32} = 0.03125$	$\frac{32}{32} = 1$	$\frac{1}{32} = 0.03125$

Each $P(H \leq k)$ is a running total of the previous column: $\frac{1}{32}$, then $\frac{1+5}{32} = \frac{6}{32}$, then $\frac{1+5+10}{32} = \frac{16}{32}$, and so on. Each $P(H \geq k)$ is a running total from the bottom.



The bars of $P(H \leq k)$: they climb up to 1.

Step 3: the four answers

- (a) $P(H \leq 2) = \frac{1+5+10}{32} = \frac{16}{32} = 0.5$. Half of the people say two yeses or fewer.
- (b) $P(H \geq 4) = \frac{5+1}{32} = \frac{6}{32} = 0.1875$.
- (c) $P(H \geq 3) = \frac{10+5+1}{32} = \frac{16}{32} = 0.5$, and $1 - P(H \leq 2) = 1 - 0.5 = 0.5$. They agree, because “3 or more” is the opposite (complement) of “2 or fewer” (NOT rule). In general $P(H \geq k) = 1 - P(H \leq k-1)$.
- (d) Add the bars 1, 2, 3, 4: $\frac{5+10+10+5}{32} = \frac{30}{32} = 0.9375$. Another way: “not 0 and not 5”, so $1 - \frac{1}{32} - \frac{1}{32} = \frac{30}{32}$.

Note. The table is symmetric ($P(H \leq k)$ read downwards equals $P(H \geq k)$ read upwards) because $p = \frac{1}{2}$: yes and no are equally likely. This will not stay true when $p = 0.9$.

ANSWER (a) 0.5 (b) 0.1875 (c) both sides equal 0.5 (d) $\frac{30}{32} = 0.9375$.

ADDITIONAL EXAMPLE 7

easy practice

For three fair coin flips (bars 0.125, 0.375, 0.375, 0.125 from Additional Example 2), let Y be the number of heads. Find (a) $P(Y \leq 1)$, (b) $P(Y \geq 2)$, (c) $P(Y \geq 1)$.

SOLUTION

(a)

$$P(Y \leq 1) = P(Y = 0) + P(Y = 1) = 0.125 + 0.375 = 0.5.$$

(b)

$$P(Y \geq 2) = P(Y = 2) + P(Y = 3) = 0.375 + 0.125 = 0.5.$$

(c)

“At least one head” is the opposite of “no heads”, so $P(Y \geq 1) = 1 - P(Y = 0) = 1 - 0.125 = 0.875$.

ANSWER (a) 0.5 (b) 0.5 (c) 0.875.

QUESTION 17*the Galton board*

A Galton board has 5 rows of pegs. A ball falls from the top, and at each row it bounces left or right, each with chance $\frac{1}{2}$, independently of the other rows. A bounce to the right counts as a “yes”. The ball finally lands in bin $0, 1, \dots, 5$.

- (a) What does the bin number count, and which family is it in?
- (b) How many different left/right paths lead to each bin?
- (c) What fraction of the balls should land in each bin?
- (d) If $N = 160$ balls fall, how many do we expect in each bin? Why will a real run not match exactly?

SOLUTION**(a) What the bin number counts**

Think of each ball as *one person* and each row of pegs as *one harmless question*. A bounce to the right is a yes. The bin number is the number of right bounces, that is, the number of yeses. The five bounces are independent and 50/50, so the bin number is

$$H \sim \text{Bin}\left(5, \frac{1}{2}\right),$$

the same H as in Question 14.

(b) Counting the paths

A path is a list of five bounces, for example R, L, R, R, L. There are $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32$ paths. To count how many lead to each bin we use one simple rule: a ball can reach a peg only from the peg above-left or the peg above-right, so *the number of paths to a spot is the sum of the two numbers above it*.

start						
			1			
after row 1		1		1		
after row 2		1	2	1		
after row 3		1	3	3	1	
after row 4	1	4	6	4	1	
after row 5	1	5	10	10	5	1
	bin 0	bin 1	bin 2	bin 3	bin 4	bin 5

Paths to each spot. Every number is the sum of the two numbers above it (this is Pascal's triangle).

For example, the 10 in the last row is $4 + 6$. The last row is 1, 5, 10, 10, 5, 1, which adds to $1 + 5 + 10 + 10 + 5 + 1 = 32 = 2^5$, so we have counted every path. These numbers are exactly $\binom{5}{k}$.

(c) The fraction of balls in each bin

Every path has the same chance $(\frac{1}{2})^5 = \frac{1}{32}$, so a bin with m paths gets the fraction $\frac{m}{32}$:

$$\frac{1}{32}, \frac{5}{32}, \frac{10}{32}, \frac{10}{32}, \frac{5}{32}, \frac{1}{32},$$

the bars of H from Question 15.

(d) The expected piles for $N = 160$ balls

Expected number in a bin = $N \times$ its fraction. Here $160 \times \frac{1}{32} = 5$, so the piles are

Bin	0	1	2	3	4	5
Expected balls	5	25	50	50	25	5

Check: $5 + 25 + 50 + 50 + 25 + 5 = 160$.

Why a real run wobbles

Each ball makes its own random bounces, so the number of balls in a bin is itself random. For a bin with chance q the usual wobble is about $\sqrt{Nq(1-q)}$:

$$\text{middle bins: } \sqrt{160 \cdot \frac{10}{32} \cdot \frac{22}{32}} \approx 5.9, \quad \text{end bins: } \sqrt{160 \cdot \frac{1}{32} \cdot \frac{31}{32}} \approx 2.2$$

So a middle pile of 44 or 56, or an end pile of 3 or 7, is completely normal. On the screen the *solid* bars are what the balls really did, and the *dashed outlines* mark the exact expected piles. The solid bars sit close to the outlines but not exactly on them.

Note. This works for any number of balls: expected balls in bin k is $N \times \binom{5}{k}/32$. With more balls the piles hug the outlines more tightly in proportion, because the wobble grows like $\sqrt{N}$ while the piles grow like N .

ANSWER Bin number = $H \sim \text{Bin}(5, \frac{1}{2})$. Paths: 1, 5, 10, 10, 5, 1 out of 32. Fractions $\frac{1}{32}, \frac{5}{32}, \frac{10}{32}, \frac{10}{32}, \frac{5}{32}, \frac{1}{32}$. For 160 balls: 5, 25, 50, 50, 25, 5 on average, with a real wobble of about ± 6 in the middle and ± 2 at the ends.

ADDITIONAL EXAMPLE 8*easy practice*

A smaller Galton board has only 3 rows of pegs, and 80 balls fall through it. Find the number of paths to each bin, the fraction of balls in each bin, and the expected number of balls in each bin.

SOLUTION**Paths**

Use the same “add the two numbers above” rule for three rows:

start: 1 row 1: 1 1 row 2: 1 2 1 row 3: 1 3 3 1

There are $2^3 = 8$ paths in all, and $1 + 3 + 3 + 1 = 8$.

Fractions and expected balls

Each path has chance $\frac{1}{8}$, so the fractions for bins 0, 1, 2, 3 are $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$. With 80 balls, $80 \div 8 = 10$ balls per path, so the expected piles are 10, 30, 30, 10. Check: $10 + 30 + 30 + 10 = 80$.

Note. These are the same chances as the three coin flips in Additional Example 2: a bounce to the right is a head, and the bin number is the number of heads.

ANSWER Paths 1, 3, 3, 1 out of 8; fractions $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$; expected balls 10, 30, 30, 10.

QUESTION 18*the average*

What is the average number of harmless yeses? Check your answer in two different ways.

SOLUTION**Method 1: the binomial average**

$$n \times p = 5 \times \frac{1}{2} = 2.5$$

Method 2: from the bars

Multiply each value by the height of its bar and add (each value counts as often as it happens):

$$\begin{aligned} \text{average} &= 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 2 \cdot \frac{10}{32} + 3 \cdot \frac{10}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{1}{32} \\ &= \frac{0 + 5 + 20 + 30 + 20 + 5}{32} = \frac{80}{32} = 2.5 \end{aligned}$$

Both methods agree.

Note. No single person can say 2.5 yeses, but over a large crowd the average is 2.5. On the Galton board it is the middle of the piles. We will need this number later, when we match the average of the model with the average of the real answers.

ANSWER 2.5 harmless yeses on average.

ADDITIONAL EXAMPLE 9*easy practice*

A shop has four customers in a row. Each one independently buys something with chance 0.3. Let B be the number of customers who buy. Find the bars and the average of B .

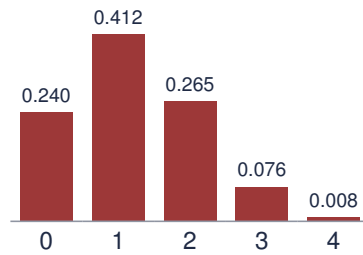
SOLUTION**Identify**

Four independent trials, the same chance 0.3, counting successes: $B \sim \text{Bin}(4, 0.3)$.

The bars

k	$\binom{4}{k}$	0.3^k	0.7^{4-k}	$P(B = k)$
0	1	1	0.2401	$1 \cdot 1 \cdot 0.2401 = 0.2401$
1	4	0.3	0.343	$4 \cdot 0.3 \cdot 0.343 = 0.4116$
2	6	0.09	0.49	$6 \cdot 0.09 \cdot 0.49 = 0.2646$
3	4	0.027	0.7	$4 \cdot 0.027 \cdot 0.7 = 0.0756$
4	1	0.0081	1	$1 \cdot 0.0081 \cdot 1 = 0.0081$

Check: $0.2401 + 0.4116 + 0.2646 + 0.0756 + 0.0081 = 1.0000$.

The bars of $B \sim \text{Bin}(4, 0.3)$.**The average**

$$4 \times 0.3 = 1.2$$

Note. When $p \neq \frac{1}{2}$ the bars lean to one side (here to the left, because $p < \frac{1}{2}$). With $p = \frac{1}{2}$, as for H , the bars are perfectly symmetric.

ANSWER Bars 0.2401, 0.4116, 0.2646, 0.0756, 0.0081 for $k = 0, \dots, 4$. Average 1.2.

Topic 8 Build the poll together**TODAY'S TASK: FIND THE HIDDEN SUPPORT**

1. Tell the story of one answer: what is random, and what is not.
2. Run the story. Do 1,000 of its answers look like the real ones?
3. Answer the question: how many people really support Palpatine?

In section you build the story step by step with a Code Agent: you say one sentence and it builds that step. Its screen shows three variables, H (the harmless yeses), S (whether the person supports Palpatine) and A (the answer they say), with an arrow from H and an arrow from S into A . It also has two sliders, called θ_1 (the chance of a yes to a harmless question, set to 0.5) and θ_2 (the chance that a person supports Palpatine). On the solution slides θ_2 is written p , and we do the same here: $\theta_2 = p$. The two sentences you say first are “model one person answering the poll, start with the 5 harmless questions” and “now their secret, we don’t know the chance yet”.

Spoiler warning: the solutions below reveal the hidden value, so try Question 19 to Question 24 on your own first.

QUESTION 19*task 1: the story of one answer*

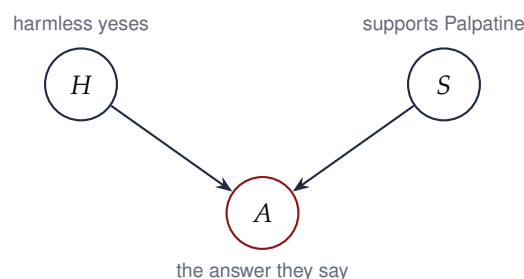
Tell the story of what one person says in the poll (five harmless questions plus the Palpatine question). Which parts are random and which are not?

SOLUTION**Notation (defined first)**

H is the number of harmless yeses (as before). S is the secret answer: $S = 1$ if the person supports Palpatine and $S = 0$ if not. T is the total number of yeses, and A is the number the person actually reports. The chance of a harmless yes is $\theta_1 = \frac{1}{2}$, and the chance of supporting Palpatine is $\theta_2 = p$.

The story, in words first

1. The person answers the five harmless questions, each 50/50. The count of yeses is $H \sim \text{Bin}(5, \frac{1}{2})$, exactly as in Topic 7.
2. The person has a secret: they support Palpatine with chance p . So $S \sim \text{Ber}(p)$. The number p is the *hidden support*: a fixed number that we do not know and want to find. (The “random” part is that we pick a person from the crowd, so whether that person supports him has chance p .) For this practice poll the hidden value is $p = 0.09$.
3. They say how many are yes in all: $T = H + S$, which can be $0, 1, \dots, 6$. Both H and S feed into the answer, which is why the screen draws an arrow from each of them into A .
4. One more rule, which the room adds last, after comparing with the real answers: **nobody says** 0. If $T = 0$ the person says 1 instead. In short, $A = T$ unless $T = 0$, and then $A = 1$.



The story as a picture: A is made from H and S .

What is random and what is not

Random	Not random (fixed)
H , the harmless yeses	the number of harmless questions (5)
S , the secret answer	the chance of a harmless yes, $\theta_1 = \frac{1}{2}$
$T = H + S$ and so A	the rule “a 0 becomes a 1”
	$p = \theta_2$ itself: fixed but unknown, the thing we want

ANSWER $H \sim \text{Bin}(5, \frac{1}{2})$, $S \sim \text{Ber}(p)$, $T = H + S$, and the reported number is $A = T$ unless $T = 0$, in which case $A = 1$.

QUESTION 20*one person at a time*

Four people answer. Work out what each one says.

Asha: harmless answers yes, no, yes, yes, no; secret: no.

Bilal: harmless answers no, no, yes, yes, no; secret: yes.

Chen: harmless answers no, no, no, no, no; secret: no.

Dana: harmless answers yes, yes, yes, yes, yes; secret: yes.

SOLUTION

For each person: count H , set S (yes = 1, no = 0), add to get $T = H + S$, then apply the rule “a 0 becomes a 1” to get A .

Person	H	S	$T = H + S$	Reports A
Asha	3	0	3	3
Bilal	2	1	3	3
Chen	0	0	0	1
Dana	5	1	6	6

Note. Asha and Bilal both report 3, but their secrets are different. That is the protection from Question 12. Chen’s total is 0, so the rule makes him say 1. Dana’s 6 means yes to everything, including the secret. Dana’s 6 and Chen’s honest 0 are the two answers that give a secret away (see Topic 9).

ANSWER Asha reports 3, Bilal reports 3, Chen reports 1, Dana reports 6.

QUESTION 21*task 2: run the story*

Using $p = 0.09$, find the chance of each total $T = 0, 1, \dots, 6$ (before applying the rule that a 0 becomes a 1). Then say how many people in 1,000 we expect for each total.

SOLUTION**Step 1: the ways a total can happen**

$T = k$ happens in exactly two ways: either $H = k$ and $S = 0$, or $H = k - 1$ and $S = 1$. These two ways cannot happen together (OR rule, we add), and in each way H and S are independent (AND rule, we multiply):

$$P(T = k) = P(H = k) (1 - p) + P(H = k - 1) p$$

Step 2: the numbers

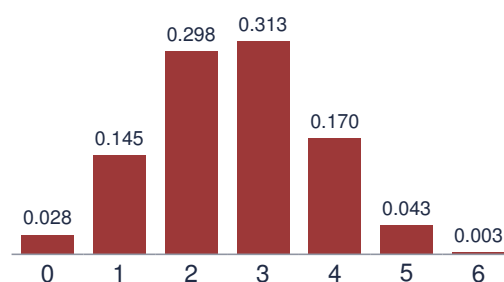
With $p = 0.09$ we have $1 - p = 0.91$. From Question 15, the bars of H for $k = 0, \dots, 5$ are

0.03125, 0.15625, 0.3125, 0.3125, 0.15625, 0.03125,

and $P(H = -1) = P(H = 6) = 0$ (these values are impossible).

k	secret "no": $P(H = k) \times 0.91$	secret "yes": $P(H = k - 1) \times 0.09$	$P(T = k)$	per 1,000
0	$0.03125 \times 0.91 = 0.0284375$	0	0.0284375	28
1	$0.15625 \times 0.91 = 0.1421875$	$0.03125 \times 0.09 = 0.0028125$	0.1450000	145
2	$0.3125 \times 0.91 = 0.2843750$	$0.15625 \times 0.09 = 0.0140625$	0.2984375	298
3	$0.3125 \times 0.91 = 0.2843750$	$0.3125 \times 0.09 = 0.0281250$	0.3125000	313
4	$0.15625 \times 0.91 = 0.1421875$	$0.3125 \times 0.09 = 0.0281250$	0.1703125	170
5	$0.03125 \times 0.91 = 0.0284375$	$0.15625 \times 0.09 = 0.0140625$	0.0425000	43
6	0	$0.03125 \times 0.09 = 0.0028125$	0.0028125	3

Check: the column $P(T = k)$ adds to $0.0284375 + 0.145 + 0.2984375 + 0.3125 + 0.1703125 + 0.0425 + 0.0028125 = 1$.



The model's bars for the total $T = H + S$ (before the "nobody says 0" rule).

Note. Running the story on a computer 1,000 times gives bars close to the "per 1,000" column. Each bar wobbles by roughly ± 15 because of sampling noise. (Topic 10 runs the story in Python.)

ANSWER $P(T = k) = 0.0284, 0.1450, 0.2984, 0.3125, 0.1703, 0.0425, 0.0028$ for $k = 0, \dots, 6$, which is about 28, 145, 298, 313, 170, 43, 3 people per 1,000.

QUESTION 22*task 2: compare with the real answers*

The real poll gave 1,000 answers: about 17% said 1, 30% said 2, 33% said 3, 17% said 4, 4% said 5 and fewer than 1% said 6 (rounded percentages). The average real answer is 2.62. Put the model's bars from Question 21 next to the real bars and read both charts bar by bar from the left. What is different?

SOLUTION**Step 1: put the two charts side by side**

We compare the bars at 0, 1, 2, ..., 6 one at a time. The model column is $P(T = k)$ from Question 21 written as a percentage.

Answer	Model so far (%)	Real answers (%)
0	2.8	no bar
1	14.5	17
2	29.8	30
3	31.3	33
4	17.0	17
5	4.3	4
6	0.3	< 1

Step 2: spot the differences, bar by bar

- At 0: the real chart has **no bar at all**. Nobody says 0. Our model expects about 2.8%, which is 28 people per 1,000.
- At 1: the real bar is taller (17% against 14.5%). Notice that $14.5\% + 2.8\% = 17.3\%$. It looks as if the people who would have said 0 said 1 instead.
- At 2, 3, 4, 5, 6: the bars agree to within about two percentage points.

Step 3: is the missing 0 just chance?

If zeros were allowed we would expect 28 of them, give or take about 5 (the usual wobble of a count like this is $\sqrt{1000 \times 0.0284 \times 0.9716} \approx 5.3$). Seeing *none at all* is more than five such wobbles below 28, so it is a real feature of the data and not sampling noise. The other differences (33% against 31.3% at 3, for example) are about one wobble, which is normal.

Note. The percentages on the real chart are rounded, so they add up to a little more than 100%. That is rounding and nothing more.

ANSWER The real answers contain no 0, while the model expects about 28 per 1,000. The real bar at 1 is also taller. The model is missing a rule.

QUESTION 23*task 2: fix the model*

Repair the model so that nobody says 0, find the new bars of the reported number A , and compare them with the real answers.

SOLUTION**Step 1: the rule**

Anyone whose total is $T = 0$ says 1 instead. So the people who used to be at 0 move to 1.

Step 2: the new bars

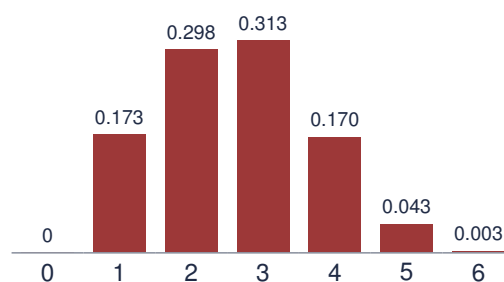
$$P(A = 0) = 0$$

$$P(A = 1) = P(T = 1) + P(T = 0) = 0.1450000 + 0.0284375 = 0.1734375$$

$$P(A = k) = P(T = k) \quad \text{for } k = 2, 3, 4, 5, 6$$

k	$P(A = k)$	per 1,000 people
0	0	0
1	0.1734375	173
2	0.2984375	298
3	0.3125000	313
4	0.1703125	170
5	0.0425000	43
6	0.0028125	3

Check: $0.1734375 + 0.2984375 + 0.3125 + 0.1703125 + 0.0425 + 0.0028125 = 1$.



The repaired model: the bar at 0 is gone and the bar at 1 has grown.

Step 3: compare with the real answers and with one run of the model

On the screen, one run of 1,000 people from the repaired model (with $p = 0.09$) gives the second column below, and the real answers are in the third column (all rounded percentages).

Answer	Exact model (%)	One model run (%)	Real answers (%)
1	17.3	19	17
2	29.8	29	30
3	31.3	32	33
4	17.0	16	17
5	4.3	5	4
6	0.3	< 1	< 1

Every bar of the model run is within two percentage points of the real bar. A single bar of 1,000 people wobbles by about 1.2 to 1.5 points (for the bar at 1: $\sqrt{0.173 \times 0.827 / 1000} \approx 0.012$), so the picture matches. (On the screen the dashed outlines on the model chart appear to mark the real bars, so you can see this bar by bar.)

ANSWER $P(A = k) = 0, 0.1734, 0.2984, 0.3125, 0.1703, 0.0425, 0.0028$ for $k = 0, 1, \dots, 6$. These match the real answers up to sampling noise.

QUESTION 24

task 3: answer the question

Find the hidden support p by matching the *average* of the model with the average of the real answers, which is 2.62.

SOLUTION

Step 1: the average of the model

Harmless part: average of $H = 5 \times \frac{1}{2} = 2.5$ (Question 18). Secret part: S is 1 with chance p and 0 otherwise, so its average is p . The average of a sum is the sum of the averages, so before the rule

$$\text{average of } T = 2.5 + p.$$

Step 2: add the effect of the “no zeros” rule

The rule raises a total of 0 to 1, adding 1 to that person’s answer. Total 0 happens when $H = 0$ and $S = 0$, with chance $\frac{1}{32}(1 - p)$. So the rule adds $\frac{1-p}{32}$ to the average:

$$\text{average of } A = 2.5 + p + \frac{1-p}{32}.$$

Step 3: sanity checks, out loud

Is p between 0 and 1? It must be, because it is a chance. Is the average between 2.5 and 3.5? Yes: the harmless part gives 2.5 and the secret part adds between 0 and 1.

Step 4: move the slider by hand

We want the model’s average to equal the data’s average, 2.62. Try different values of the slider $\theta_2 = p$:

p	model average $2.5 + p + \frac{1-p}{32}$	compared with 2.62	move the slider
0.05	2.5797	too low	up
0.09	2.6184	very close	stop
0.10	2.6281	very close (slightly high)	fine
0.12	2.6475	too high	down
0.20	2.7250	far too high	down

So the room calls “up” while the model is too low, “down” while it is too high, and lands near 0.09 to 0.10.

Step 5: solve exactly

$$\begin{aligned}
 2.5 + p + \frac{1-p}{32} &= 2.62 \\
 80 + 32p + 1 - p &= 32 \times 2.62 = 83.84 \quad (\text{multiply both sides by } 32) \\
 31p &= 83.84 - 81 = 2.84 \\
 p &= \frac{2.84}{31} \approx 0.0916
 \end{aligned}$$

Step 6: why the rule matters

If we had ignored the “no zeros” rule, we would have solved $2.5 + p = 2.62$ and got $p = 0.12$, which is about 0.03 too high. Modeling the quirk moves the answer a little, but in the right direction.

Step 7: how close is close?

One answer has a standard deviation of about 1.10, so the average of 1,000 people wobbles by about $\frac{1.10}{\sqrt{1000}} \approx 0.035$. On the screen, at $\theta_2 = 0.09$ (the true value) one run of 1,000 people shows “average 2.58, real 2.62”. That gap of 0.04 is only about one wobble, so it is noise, not a mistake. It also means p is only pinned down to about ± 0.04 , and a room answer of 0.10 when the truth is 0.09 is excellent.

ANSWER $p \approx 0.09$: about 1 person in 10 really supports Palpatine. Nobody told us their secret; the crowd did.

ADDITIONAL EXAMPLE 10

easy practice

A mini poll has only *one* harmless question (50/50) plus the secret question. A person reports $T = H_1 + S$, which is 0, 1 or 2. (a) If the hidden support is $p = 0.2$, find the bars of T and its average. (b) A pollster does not know p but sees that the average answer is 0.62. What is p ?

SOLUTION

Part (a): the bars

H_1 is 1 with chance 0.5 and 0 with chance 0.5. S is 1 with chance 0.2 and 0 with chance 0.8.

$$P(T = 0) = P(H_1 = 0) \cdot P(S = 0) = 0.5 \times 0.8 = 0.4$$

$$P(T = 1) = P(H_1 = 1)P(S = 0) + P(H_1 = 0)P(S = 1) = 0.5 \times 0.8 + 0.5 \times 0.2 = 0.4 + 0.1 = 0.5$$

$$P(T = 2) = P(H_1 = 1) \cdot P(S = 1) = 0.5 \times 0.2 = 0.1$$

Check: $0.4 + 0.5 + 0.1 = 1$. The average is $0 \cdot 0.4 + 1 \cdot 0.5 + 2 \cdot 0.1 = 0.7$, which equals $0.5 + 0.2$

(average of the harmless part plus p).

Part (b): read p off the data

The model says average = $0.5 + p$. The data say 0.62. So $0.5 + p = 0.62$ and $p = 0.12$.

Note. This is the whole method in miniature: write the model's average in terms of p , then choose p so that it matches the data's average. In the mini poll, an answer of 2 would reveal a "yes" and an answer of 0 a "no". Only 1 is safe. A longer list of harmless questions makes this much safer.

ANSWER (a) Bars 0.4, 0.5, 0.1; average 0.7. (b) $p = 0.12$.

Topic 9 Is anyone's secret out?

IN PLAIN WORDS

The list trick protects a person only when *two different stories* could have produced their number (as in Question 12). If just *one* story can produce it, the answer gives the person away. In our poll the reported total is $T = H + S$, so we look for totals that can come from only one pair (H, S) . There are two such totals: the biggest and the smallest.

QUESTION 25

is anyone's secret out? part 1

In the Palpatine poll (five 50/50 harmless questions, hidden support $p = 0.09$), what does an answer of 6 mean? How many people in 1,000 give themselves away with it?

SOLUTION

Step 1: which story gives a 6?

The harmless count H is at most 5 and the secret S is at most 1, so $T = 6$ happens only when $H = 5$ and $S = 1$: yes to everything, Palpatine too. There is no other way, so a 6 gives the secret away.

Step 2: the chance of a 6

We need all five harmless yeses and the secret yes (AND rule, independent):

$$P(\text{answer} = 6) = p \cdot \left(\frac{1}{2}\right)^5 = \frac{0.09}{32} = 0.0028125 \approx 0.003$$

Here $\left(\frac{1}{2}\right)^5$ means $\frac{1}{2}$ multiplied by itself five times: $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{32}$.

Step 3: in 1,000 people

$1,000 \times 0.0028125 \approx 2.8$, so about 3 people in 1,000 give themselves away.

Note. Only a supporter can ever say 6. That is why a 6 is so revealing.

ANSWER A 6 means yes to everything, Palpatine too. $P(\text{answer} = 6) = p \cdot \left(\frac{1}{2}\right)^5 = \frac{0.09}{32} \approx 0.003$, about 3 people in 1,000.

QUESTION 26*is anyone's secret out? part 2*

In the same poll, what does an honest total of 0 mean, why is it called the “dangerous answer”, and how many people in 1,000 would give themselves away with it? Then add up both kinds of giveaway.

SOLUTION**Step 1: which story gives a 0?**

$T = 0$ happens only when $H = 0$ and $S = 0$: no to everything, Palpatine too. That is the only way, so this answer also gives the secret away. It says out loud “I do not support Palpatine”, and in the story that is the dangerous thing to be known as, because Palpatine punishes people who say they do not like him.

Step 2: the chance of a 0

All five harmless answers are no and the secret answer is no (AND rule). A non-supporter has chance $1 - p = 0.91$:

$$P(T = 0) = (1 - p) \cdot \left(\frac{1}{2}\right)^5 = \frac{0.91}{32} = 0.0284375 \approx 0.028$$

Step 3: in 1,000 people

$1,000 \times 0.0284375 \approx 28$, so about 28 people in 1,000 would give themselves away.

Step 4: add up both kinds of giveaway

The answers 0 and 6 cannot both happen to the same person, so we add (OR rule):

$$\frac{0.91}{32} + \frac{0.09}{32} = \frac{0.91 + 0.09}{32} = \frac{1}{32} = 0.03125$$

That is about 31 people in 1,000. Notice that the support p has disappeared from the answer: with five 50/50 harmless questions, exactly $\frac{1}{32}$ of all people are at one of the two ends, whatever the support is.

The safe answers

1, 2, 3, 4, 5 can each be produced by two different stories (a secret yes or a secret no), so they protect the person.

Note. The slide writes $P(\text{Answer} = 0)$, but in the real poll nobody *says* 0. The 0 here is the person's honest total T , before the rule “a 0 becomes a 1” (Question 19).

Note. Our reading, not stated on the slides: this may be why the real answers contain no 0. A person whose honest total is 0 would rather say 1, because 1 can also come from a secret yes. That turns the missing bar from Question 22 into a story you can discuss with the room.

ANSWER An honest 0 means no to everything, Palpatine too. $P(T = 0) = (1 - p) \cdot \left(\frac{1}{2}\right)^5 = \frac{0.91}{32} \approx 0.028$, about 28 in 1,000. Together with the 6's, about $\frac{1}{32} \approx 31$ people in 1,000 are at an end.

QUESTION 27*what if the harmless questions are too easy?*

Suppose almost everybody says yes to each harmless question: 9 in 10 do. Where do the five harmless answers land now? How many answers give someone away, compared with the 50/50 case?

SOLUTION**Step 1: the new harmless count**

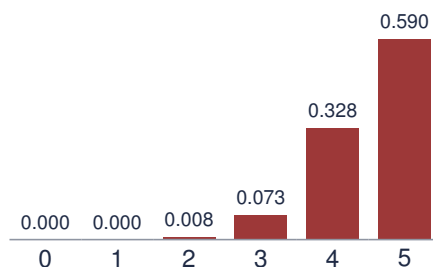
Call it H' . It is still five independent yes/no answers, but each has chance 0.9: $H' \sim \text{Bin}(5, 0.9)$. The average is $5 \times 0.9 = 4.5$. On the calculator, this is the same picture as before with the slider p moved from 0.5 to 0.9.

Step 2: the bars of H'

Use $P(H' = k) = \binom{5}{k} 0.9^k 0.1^{5-k}$:

k	$P(H' = k)$
0	$1 \cdot 1 \cdot 0.00001 = 0.00001$
1	$5 \cdot 0.9 \cdot 0.0001 = 0.00045$
2	$10 \cdot 0.81 \cdot 0.001 = 0.0081$
3	$10 \cdot 0.729 \cdot 0.01 = 0.0729$
4	$5 \cdot 0.6561 \cdot 0.1 = 0.32805$
5	$1 \cdot 0.59049 \cdot 1 = 0.59049$

Check: $0.00001 + 0.00045 + 0.0081 + 0.0729 + 0.32805 + 0.59049 = 1.00000$.



The bars of $H' \sim \text{Bin}(5, 0.9)$: the piles crowd into 4 and 5.

Clicking the last bar on the calculator reads $P(H' = 5) = 0.590$, which is 59%. Most people say 4 or 5: $P(H' \geq 4) = 0.32805 + 0.59049 = 0.91854$, about 92% of people.

Step 3: the chance of giving yourself away

A 6 needs all five harmless yeses and the secret yes:

$$P(T = 6) = p \cdot P(H' = 5) = 0.09 \times 0.59049 \approx 0.053$$

An honest 0 needs all five harmless noes and the secret no:

$$P(T = 0) = (1 - p) \cdot P(H' = 0) = 0.91 \times 0.00001 \approx 0.00001$$

So about 5 answers in 100 give someone away, and almost all of them are 6's. With 50/50 questions it was about 3 in 1,000 for a 6 (and 28 in 1,000 for a 0).

Step 4: look at it from the supporters' side

A supporter says 6 exactly when all five harmless answers are yes, so 59% of supporters (59 of every 100) would give themselves away, compared with $\frac{1}{32} \approx 3\%$ with 50/50 questions.

Step 5: the lesson

When the harmless questions are too easy, the harmless count piles up at 4 and 5, so a supporter's secret yes often pushes the total to 6, and only supporters can reach 6. Pick harmless questions near 50/50.

ANSWER $H' \sim \text{Bin}(5, 0.9)$ piles up at 4 and 5 (92% of people). About 5 in 100 answers give someone away (mostly 6's), and 59 in every 100 supporters are exposed.

ADDITIONAL EXAMPLE 11*easy practice*

A survey uses a short list of only 3 harmless questions plus a secret question, and the true support is $p = 0.2$. Two answers give a person away: 4 (yes to everything) and 0 (no to everything). Find the chance of each, and their sum, when (a) each harmless question is 50/50, and (b) each harmless question is answered yes by 9 in 10 people.

SOLUTION**Setup**

The answer 4 needs three harmless yeses and the secret yes, so its chance is $P(H = 3) \times p$. The answer 0 needs three harmless noes and the secret no, so its chance is $P(H = 0) \times (1 - p)$, where $1 - p = 0.8$.

(a) 50/50 questions

$P(H = 3) = P(H = 0) = \left(\frac{1}{2}\right)^3 = \frac{1}{8} = 0.125$. So $P(T = 4) = 0.125 \times 0.2 = 0.025$ and $P(T = 0) = 0.125 \times 0.8 = 0.1$. Together: $0.025 + 0.1 = 0.125$, which is $\frac{1}{8}$, about 12 people in 100.

(b) Easy questions

$P(H = 3) = 0.9^3 = 0.9 \times 0.9 \times 0.9 = 0.729$ and $P(H = 0) = 0.1^3 = 0.001$. So $P(T = 4) = 0.729 \times 0.2 = 0.1458$ and $P(T = 0) = 0.001 \times 0.8 = 0.0008$. Together: 0.1466, about 15 people in 100.

Note. A short list is already far less safe than our five-question list (compare $\frac{1}{32} \approx 0.03$ in Question 26), and easy questions make it worse. Notice that with 50/50 questions the total is again $\frac{1}{8} = \left(\frac{1}{2}\right)^3$, whatever p is.

ANSWER (a) $0.025 + 0.1 = 0.125$ (b) $0.1458 + 0.0008 = 0.1466$. Easy harmless questions expose more people, mostly supporters.

EXTENSION QUESTIONS

QUESTION 28

extension

“How would you protect them?” Give ideas to make it harder to expose a person. In particular, what changes if we use 10 harmless questions instead of 5?

SOLUTION

Idea 1: choose harmless questions near 50/50

Question 27 showed that easy questions expose 59% of supporters. In general we want yes-to-everything and no-to-everything to be rare for the harmless part.

Idea 2: use more harmless questions

With 10 harmless questions, $H \sim \text{Bin}(10, \frac{1}{2})$, the two giveaway totals are 0 (all noes) and 11 (ten harmless yeses and the secret yes):

$$P(T = 11) = p \cdot \left(\frac{1}{2}\right)^{10} = \frac{0.09}{1024} \approx 0.000088, \quad P(T = 0) = (1 - p) \cdot \left(\frac{1}{2}\right)^{10} = \frac{0.91}{1024} \approx 0.00089$$

Together they make $\frac{1}{1024} \approx 0.00098$, about 1 person in 1,000. With five questions it was $\frac{1}{32}$, so the list is now 32 times safer.

The price of more safety

More harmless questions add more random noise on top of the secret. One answer then has a standard deviation of about 1.60 instead of 1.10, so the average of 1,000 people wobbles by about 0.051 instead of 0.035. Our estimate of p is therefore a little less sharp (about ± 0.05 instead of ± 0.04).

What about the model?

Only the numbers change: the harmless average is $10 \times \frac{1}{2} = 5$, so the model's average becomes $5 + p + \frac{1-p}{1024}$. (A total of 0 now has chance $\frac{1-p}{1024}$, so the “no zeros” rule barely matters.)

ANSWER Pick 50/50 harmless questions and use more of them. Ten questions cut the number of people at an end from about 31 in 1,000 to about 1 in 1,000, at the cost of a slightly noisier estimate of p .

QUESTION 29

extension

What if some supporters still lie and leave the Palpatine question out of their count? What does the poll then estimate?

SOLUTION**Step 1: what their answers look like**

A lying supporter reports only their harmless yeses, H . This looks exactly like the answer of a non-supporter, so the data cannot tell them apart.

Step 2: what the model finds

Say a fraction f of supporters lie like this. Then only a fraction $1 - f$ of supporters add their secret yes, and the slider finds

$$p_{\text{found}} = p(1 - f).$$

Step 3: an example

If $p = 0.09$ and one third of supporters lie ($f = \frac{1}{3}$), the poll finds

$$0.09 \times (1 - \frac{1}{3}) = 0.09 \times \frac{2}{3} = 0.06.$$

Note. The list trick protects people, but it cannot force honesty. It estimates the share of people who *include* their secret yes, so it gives a lower bound on the true support when some supporters lie.

ANSWER The poll then finds $p(1 - f)$, which is less than the true support p . With $p = 0.09$ and $f = \frac{1}{3}$ it finds 0.06.

QUESTION 30

extension

How could a computer find the support by itself, without a slider?

SOLUTION**Idea 1: automate the slider**

Let the computer try many values of p between 0 and 1. For each value it computes the model's average $2.5 + p + \frac{1-p}{32}$, compares it with the data's average, and keeps the p whose model average is closest. (Topic 10 does exactly this in Python.)

Idea 2: here we can even solve it directly

Set the model average equal to the data average, written $\bar{a}$:

$$2.5 + p + \frac{1-p}{32} = \bar{a} \implies 80 + 32p + 1 - p = 32\bar{a} \implies p = \frac{32\bar{a} - 81}{31}.$$

With $\bar{a} = 2.62$ this gives $p = \frac{83.84 - 81}{31} \approx 0.0916$, as before.

Idea 3: the bigger picture

For bigger models there is no simple formula, so computers improve their guess step by step, always moving p in the direction that makes the model fit the data better. That is what *learning* means, and it is exactly where the course is going. Today we did it by hand.

ANSWER Try values of p (or solve $p = \frac{32\bar{a}-81}{31}$) and keep the one whose model matches the data. Automating that search is the idea behind learning from data.

WHAT TO REMEMBER

- A random variable is known by its bars: one per value, as tall as its chance, adding to 1. The calculator can also show running totals, $P(H \leq k)$ and $P(H \geq k)$.
- $\text{Ber}(p)$ is one yes/no trial. $\text{Bin}(n, p)$ counts successes in n independent trials with the same p (average $n \times p$). Anything else is neither, and the question X asks decides the family.
- A Galton board is a binomial in action: the bin is the number of right bounces, and the piles follow Pascal's triangle (1, 5, 10, 10, 5, 1 out of 32).
- A probabilistic model is a story that makes data: run it, count, and multiply along the story.
- Asking directly fails when people hide the truth. Hiding the question in a list of harmless ones, and reporting only the *count* of yeses, protects each person.
- The poll story: $H \sim \text{Bin}(5, \frac{1}{2})$, a secret $S \sim \text{Ber}(p)$ with $p = \theta_2$, total $H + S$, and nobody says 0. Its average is $2.5 + p + \frac{1-p}{32}$. Matching it to the real average 2.62 finds $p \approx 0.09$.
- Two answers give a person away: 6 (about 3 in 1,000, a supporter) and an honest 0 (about 28 in 1,000, the dangerous answer). Together $\frac{1}{32}$. Choose harmless questions near 50/50, and use more of them for more safety.