

Bernoulli Experiments, Binomial Distribution

If a person randomly guesses the answers to 10 multiple choice questions, we can ask questions like

- ▶ what is the probability that they get none right?
- ▶ what is the probability that they get all ten right?
- ▶ what is the probability that they get at least three right?
- ▶ how many do they get right on average?

These and similar scenarios lead to *Bernoulli Experiments* and *the Binomial Distribution*. A *Bernoulli Experiment* involves repeated (in this case 10) independent trials of an experiment with 2 outcomes usually called “success” and “failure” (in this case getting a question right/wrong).

Bernoulli Experiment with n Trials

Here are the rules for a Bernoulli experiment.

1. The experiment is repeated a fixed number of times (n times).
2. Each trial has only two possible outcomes, “success” and “failure”. The possible outcomes are exactly the same for each trial.
3. The probability of success remains the same for each trial. We use p for the probability of success (on each trial) and $q = 1 - p$ for the probability of failure.
4. The trials are independent (the outcome of previous trials has no influence on the outcome of the next trial).
5. We are interested in the random variable X where $X =$ the number of successes. Note the possible values of X are $0, 1, 2, 3, \dots, n$.

Examples

Flip a coin 12 times, count the number of heads. Here $n = 12$. Each flip is a trial. It is reasonable to assume the trials are independent. Each trial has two outcomes heads (success) and tails (failure). The probability of success on each trial is $p = 1/2$ and the probability of failure is $q = 1 - 1/2 = 1/2$. We are interested in the variable X which counts the number of successes in 12 trials. This is an example of a Bernoulli Experiment with 12 trials.

Examples

A basketball player takes four independent free throws with a probability of 0.7 of getting a basket on each shot. The number of baskets made is recorded. Here each free throw is a trial and trials are assumed to be independent. Each trial has two outcomes basket (success) or no basket (failure). The probability of success is $p = 0.7$ and the probability of failure is $q = 1 - p = 0.3$. We are interested in the variable X which counts the number of successes in 4 trials. This is an example of a Bernoulli experiment with 4 trials.

Examples

A bag contains 6 red marbles and 4 blue marbles. Five marbles are drawn from the bag **without replacement** and the number of red marbles is observed. We might let a trial here consist of drawing a marble from the bag and let success be getting a red. However, this is **not** a Bernoulli experiment since the trials are not independent (the mix of reds and blues changes on each trial since we do not replace the marble) and the probability of success and failure vary from trial to trial.

Examples

A bag contains 6 red marbles and 4 blue marbles. A marble is drawn at random from the bag, its color is noted and then it is replaced. Five marbles are drawn from the urn in this way (**with** replacement) and the number of red marbles is observed. This **is** a Bernoulli experiment, where each time we draw a marble from the bag constitutes one trial. Trials are independent since we draw randomly from the bag and the probability of success (getting a red) is the same on each trial ($p = 6/10$) since we replace the marble after each draw. We are interested in the number of successes in five trials of this experiment.

Examples

1000 people are chosen at random from among likely Democrat primary voters in Indiana, and asked “do you plan to vote for Bernie Sanders in the upcoming primary on May 3?”. Their answers are recorded. Assuming people really are randomly selected, this is an independent repetition of the same trial 1000 times. If we declare voting for Bernie “success”, then p = the proportion of likely Democrat primary voters in Indiana who plan to vote for Bernie. If X = the number of successes, then a observation of X can be used to estimate (unknown) p . For example, if $X = 442$ in a particular running of this poll, we would concluded that p is likely around 44%.

Probability distribution of X

Our next goal is to calculate the probability distribution for the random variable X , where X counts the number of successes in a Bernoulli experiment with n trials. We will start with a small example for which a tree diagram can be drawn (we have already looked at a specific case of this example when we studied tree diagrams).

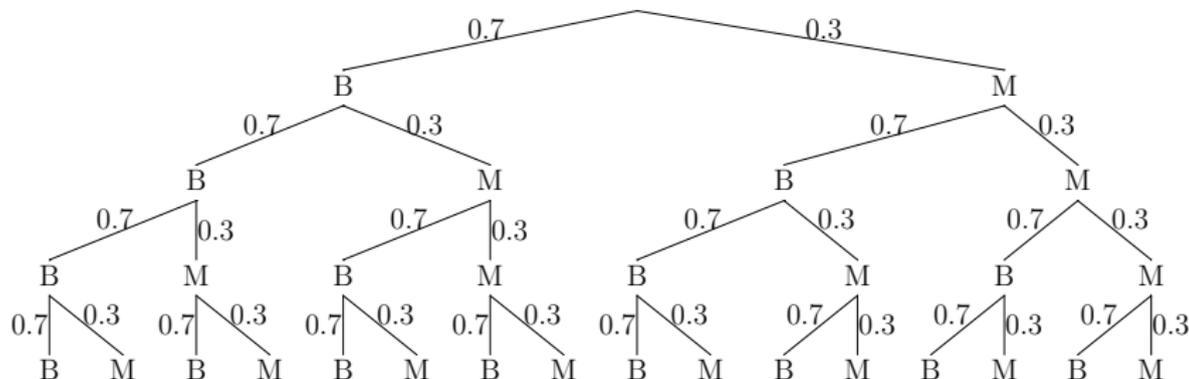
Example: A basketball player takes 4 independent free throws with a probability of 0.7 of getting a basket on each shot. Let $X =$ the number of baskets he gets.

Notice that this is indeed a Bernoulli experiment with $n = 4$ and $p = 0.7$.

Probability distribution of X

Use the tree diagram below to find the probability that he gets exactly 2 baskets or $\mathbf{P}(X = 2)$.

B = gets a basket, M = misses.



$$\mathbf{P}(X = 2) = \mathbf{C}(4, 2)(0.7)^2(0.3)^2 = 0.2646$$

Probability distribution of X

In general we have the following:

If X is the number of successes in a Bernoulli experiment with n independent trials, where the probability of success is p in each trial (and the probability of failure is then $q = 1 - p$), then

$$\mathbf{P}(X = k) = \mathbf{C}(n, k)p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}$$

for $k = 0, 1, 2, \dots, n$.

Probability distribution of X

We can see why this is true if we visualize a tree diagram for the n independent trials. The number of paths with exactly k success (out of n trials) is $\mathbf{C}(n, k)$ and the probability of every such path equals $p^k q^{n-k}$. The event that $X = k$ can result from any one of these outcomes (paths), hence the $\mathbf{P}(X = k)$ is the sum of the probabilities of all paths with exactly k successes which is $\mathbf{C}(n, k)p^k q^{n-k}$.

Example

A basketball player takes 4 independent free throws with a probability of 0.7 of getting a basket on each shot. Let $X =$ the number of baskets he gets. Write out the full probability distribution for X .

X	$\mathbf{P}(X)$
0	
1	
2	
3	
4	

Example

X	P(X)	
0	$C(4, 0)(0.7)^0(0.3)^4$	0.0081
1	$C(4, 1)(0.7)^1(0.3)^3$	0.0756
2	$C(4, 2)(0.7)^2(0.3)^2$	0.2646
3	$C(4, 3)(0.7)^3(0.3)^1$	0.4116
4	$C(4, 4)(0.7)^4(0.3)^0$	0.2401

Note $0.0081 +$
 $0.0756 + 0.2646 +$
 $0.4116 + 0.2401 = 1$

Binomial Random Variables

For a Bernoulli experiment with n trials, let X denote the number of successes in the n trials, where the probability of success in each trial is p . This distribution of random the variable X is called a *binomial distribution with parameters n and p* .

The **expected value** of X is

$$\mathbf{E}(X) = np$$

and the **standard deviation** of X is

$$\mathbf{\sigma}(X) = \sqrt{npq}$$

where $q = 1 - p$.

Examples

If a basketball player takes 8 independent free throws, with a probability of 0.7 of getting a basket on each shot, what is the probability that she gets exactly 6 baskets?

$$C(8, 6)(0.7)^6(0.3)^2 \approx 0.296$$

What is the expected number of baskets that she gets?

$np = 8(.7) = 5.6$ (**not** 6, or 5! — expected value doesn't have to be a value that can actually occur)

Examples

A student is given a multiple choice exam with 10 questions, each question with five possible answers. He guesses randomly for each question.

(a) What's P (he will get exactly 6 questions correct)?

$$n = 10, k = 6, p = 0.2, \text{ so } \mathbf{C(10, 6)(0.2)^6(0.8)^4 \approx 0.0055}$$

(b) What is the probability he will get **at least** 6?

$$\begin{aligned} &\mathbf{C(10, 6)(0.2)^6(0.8)^4 + C(10, 7)(0.2)^7(0.8)^3 +} \\ &\mathbf{C(10, 8)(0.2)^8(0.8)^2 + C(10, 9)(0.2)^9(0.8)^1 +} \\ &\mathbf{C(10, 10)(0.2)^{10}(0.8)^0 \approx 0.0063.} \end{aligned}$$

(c) What is the expected number of correct answers, and what's the standard deviation?

$$\mathbf{E(X) = np = 10 \cdot 0.2 = 2,}$$

$$\mathbf{\sigma(X) = \sqrt{npq} = \sqrt{10 \cdot 0.2 \cdot 0.8} = \sqrt{1.6} \approx 1.26}$$

A slightly off-topic example

Assume that the Mets and the Royals are in the world series, that the Mets have a $3/5$ chance of winning any given game, and that the games are independent experiments. What is the probability of a 7 game series?

Note 1: The world series is **not** a Bernoulli experiment! (number of games is not fixed in advance)

Note 2 A seven game series will occur only when each team wins 3 of the first 6 games.

A seven game series will occur whenever the Mets win exactly 3 of the first 6 games. The probability of this is $C(6, 3)(0.6)^3(0.4)^3 \approx 0.27$.

It is also true that a seven game series will occur whenever the Royals win exactly 3 of the first 6 games. The probability of this is $C(6, 3)(0.4)^3(0.6)^3 \approx 0.27$.

A quality control example

The Everlasting Lightbulb company produces light bulbs, which are packaged in boxes of 20 for shipment. Tests have shown that 4% of their light bulbs are defective.

(a) What is the probability that a box, ready for shipment, contains exactly 3 defective light bulbs?

$$\mathbf{C}(20, 3)(0.04)^3(0.96)^{17} \approx 0.036.$$

(b) What is the probability that the box contains 3 or more defective light bulbs?

$$1 - (\mathbf{C}(20, 2)(0.04)^2(0.96)^{18} + \mathbf{C}(20, 1)(0.04)^1(0.96)^{19} + \mathbf{C}(20, 0)(0.04)^0(0.96)^{20}) \approx 1 - 0.956 = 0.044.$$

We can also compute the expected number of defective bulbs, $\mathbf{E}(X) = 20 \cdot 0.04 = 0.8$, and the standard deviation, $\sigma(X) = \sqrt{20 \cdot 0.04 \cdot 0.96} \approx 0.876$.

Polling example

Suppose that the voting population in Utopia is 300 million and 60% of the voting population intend to vote for Melinda McNulty in the next election. We take a random sample of size 100 from the same voting population and ask each person chosen whether they will vote for Melinda McNulty in the next election or not. Let X be the number of YESes in our sample. The possible values of X (the number of successes) are $0, 1, 2, 3, \dots, 100$.

$n = 100$, $p = 0.6$, $q = 1 - p = 0.4$. Calculate the following:

- (a) What is $\mathbf{P}(X = 60)$? (b) What is $\mathbf{P}(X \leq 20)$?
- (c) What is $\mathbf{P}(X > 70)$? (d) What is $\mathbf{P}(X < 50)$?
- (e) What is $\mathbf{P}(50 \leq X \leq 60)$?

Polling example

$$\mathbf{P}(X = 60) \approx 0.0812$$

$$\mathbf{P}(X \leq 20) \approx 3.42 \times 10^{-16}$$

$$\mathbf{P}(X > 70) = 1 - \mathbf{P}(X \leq 70) \approx 0.0147$$

$$\mathbf{P}(X < 50) = \mathbf{P}(X \leq 49) \approx 0.0167$$

$$\begin{aligned} \mathbf{P}(50 \leq X \leq 60) = \\ \left(\mathbf{P}(X = 50) + \mathbf{P}(X = 51) + \cdots + \mathbf{P}(X = 60) \right) \approx 0.521 \end{aligned}$$

OR

$$\begin{aligned} \mathbf{P}(50 \leq X \leq 60) = \mathbf{P}(X \leq 60) - \mathbf{P}(X \leq 49) \approx \\ 0.537 - 0.0167 \approx 0.521. \end{aligned}$$

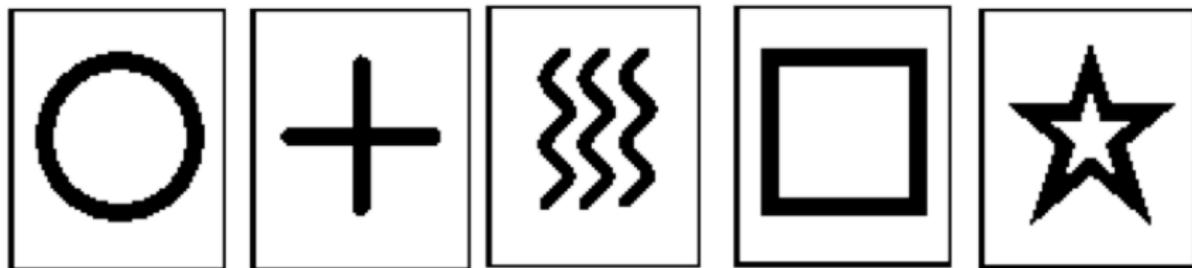
ESP — Zener Test

Zener Cards

One controversial test for ESP involves using a deck of Zener cards. This deck consists of 5 copies of the 5 cards shown below. The tester (sender) shuffles the 25 cards thoroughly, looks at the one on top of the deck and “sends” the information to the “receiver” (the person being tested for E.S.P.), without letting them see the card of course. This process is repeated many times. Many of the early results about the test were controversial because of flaws in how the test was conducted and miscalculation of probabilities (see [Zener Cards Skepticism](#)).

ESP — Zener Test

In this online version: [Zener Test](#), you have to guess 25 cards, each of which is selected randomly by the computer prior to your guess. The actual card will be shown after you click on your chosen symbol. To show evidence of ESP, you need to score at least 10 correct guesses (hits). If the selection of the card is random, this is a Bernoulli Experiment with 25 trials and a probability of $p = 0.2$ of success (correct guess) in each.



ESP — Zener Test

(a) What is the expected number of correct answers in the above test if the person taking the test is guessing?

$$\mathbf{E}(X) = 25 \cdot (0.2) = 2.$$

(b) What is $\mathbf{P}(X \geq 10)$?

$$0.0173$$

(c) Approximately how many students in a class of 100 would you expect to get a score of 10 or greater on this test by randomly guessing?

$$100 \cdot \mathbf{P}(X \geq 10) = 100 \cdot 0.0173 = 1.73, \text{ so close to } 2.$$

Old Exam questions

Recall the notation

$$\mathbf{C}(n, k) = \binom{n}{k}$$

1 An Olympic pistol shooter has a $\frac{2}{3}$ chance of hitting the target at each shot. Find the probability that he will hit exactly 10 targets in a game of 15 shots.

(a) $1 - \binom{15}{10} \left(\frac{2}{3}\right)^{10} \left(\frac{1}{3}\right)^5$ (b) $\binom{15}{10} \left(\frac{2}{3}\right)^{10} \left(\frac{1}{3}\right)^5$ (c) $\left(\frac{2}{3}\right)^{10}$

(d) $\binom{15}{10} \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^{10}$ (e) $1 - \left(\frac{1}{3}\right)^5$

Old Exam questions

$$(a) 1 - \binom{15}{10} \left(\frac{2}{3}\right)^{10} \left(\frac{1}{3}\right)^5 \quad (b) \binom{15}{10} \left(\frac{2}{3}\right)^{10} \left(\frac{1}{3}\right)^5 \quad (c) \left(\frac{2}{3}\right)^{10}$$

$$(d) \binom{15}{10} \left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^{10} \quad (e) 1 - \left(\frac{1}{3}\right)^5$$

(b) is “hits exactly 10 targets out of 15 shots” (so is correct answer)

(d) is “hits exactly 5 targets out of 15 shots”

(a) is “does not hit exactly 10 targets out of 15 shots”

(c) is “hits 10 targets in his first 10 shots”

(e) is “does not miss all 5 of his first 5 shots”

Old Exam questions

A random variable X is the number of successes in a Bernoulli experiment with n trials, each with a probability of success p and a probability of failure q . The probability distribution table of X is shown below:

k	$\mathbf{P}(X = k)$
0	$\frac{1}{81}$
1	$\frac{8}{81}$
2	$\frac{24}{81}$
3	$\frac{32}{81}$
4	$\frac{16}{81}$

Old Exam questions

Which of the following values of n , p , q give rise to this probability distribution?

(a) $n = 4$, $p = \frac{2}{3}$, $q = \frac{1}{3}$ (b) $n = 4$, $p = \frac{1}{3}$, $q = \frac{2}{3}$

(c) $n = 4$, $p = \frac{1}{6}$, $q = \frac{5}{6}$ (d) $n = 5$, $p = \frac{1}{3}$, $q = \frac{2}{3}$

(e) $n = 5$, $p = \frac{2}{3}$, $q = \frac{1}{3}$

The listed probabilities add up to 1 so they are a probability distribution and therefore $n = 4$.

$\mathbf{P}(X = 0) = \mathbf{C}(4, 0)p^0q^4 = q^4$ so $q = \frac{1}{3}$. Hence (a) is the right answer.

Old Exam questions

Peter is taking a quiz with 6 multiple choice questions. Each question has five options for the answer. Peter, who hasn't studied for the quiz, randomly guesses at each answer. Which of the following gives the probability that Peter gets 2 questions or fewer correct?

(a) $1 - \left[\binom{6}{0} (.2)^0 (.8)^6 + \binom{6}{1} (.2)^1 (.8)^5 + \binom{6}{2} (.2)^2 (.8)^4 \right]$

(b) $\binom{6}{2} (.2)^2 (.8)^4$

(c) $\binom{6}{0} (.2)^0 (.8)^6 + \binom{6}{1} (.2)^1 (.8)^5 + \binom{6}{2} (.2)^2 (.8)^4$

(d) $\binom{6}{3} (.2)^3 (.8)^3$

(e) $1 - \left[\binom{6}{0} (.2)^0 (.8)^6 + \binom{6}{1} (.2)^1 (.8)^5 \right]$

Old Exam questions

(b) is “gets exactly 2 right”

(d) is “gets exactly 3 right”

(c) is “gets 0,1 or 2 right” (so this is the correct answer)

(a) is “does not get 0, 1 or 2 right”

(e) is “does not get 0 or 1 right”