Counting problems and probability mass functions

Problem statement

Suppose you flipped a coin 5 times and counted the number of heads. There are then 6 possible results:

$$0, 1, 2, 3, 4, 5.$$

What is the probability of each outcome?

In this video, we will look at a “first principles” approach using combinatorial principles. Later we will see that using the binomial distribution would be a more efficient use of time.

Key idea

For now, we will focus on the probability of flipping **exactly two heads**.

Consider the possible arrangements of heads and tails:

$$\begin{array}{ccccc} H & H & H & H & H \\ H & H & H & H & T \\ H & H & H & T & H \\ H & H & H & T & T \\ & & \vdots & & \\ T & T & T & T & H \\ T & T & T & T & T \end{array}$$

How many of these arrangements are there?

How many of these arrangements contain exactly two heads?

Binomial coefficients

Binomial coefficients are important for many discrete probability distributions such as the binomial distribution and the negative binomial distribution. The formula is:

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

In some texts, this might be written as ${}^nC_k$ or $C(n, k)$ or C^n_k or $C_{n,k}$.

It is pronounced "n choose k". This is because it counts the number of ways of choosing k (unordered) objects from an n-element set.

Counting

This means we have:

- $\binom{5}{0}$ ways to flip 0 out of 5 heads
- $\binom{5}{1}$ ways to flip 1 out of 5 heads
- $\binom{5}{2}$ ways to flip 2 out of 5 heads
- $\binom{5}{3}$ ways to flip 3 out of 5 heads
- $\binom{5}{4}$ ways to flip 4 out of 5 heads
- $\binom{5}{5}$ ways to flip 5 out of 5 heads

Number of heads	0	1	2	3	4	5
Number of ways	1	5	10	10	5	1

Solution

Thus we can obtain the probability mass function.

Number of heads	0	1	2	3	4	5
Number of ways	1	5	10	10	5	1

Recall that there are $2^5 = 32$ possible outcomes when flipping 5 coins.

Let X denote the number of heads obtained after flipping a coin five times. The probability mass function for X is as below:

x	0	1	2	3	4	5
$P(X = x)$	$1/32$	$5/32$	$10/32$	$10/32$	$5/32$	$1/32$

Important:

- This method worked because the probability of flipping heads is the same as the probability of flipping tails. If the probabilities were not equal, more careful treatment would be needed.
- This method is for illustrative purposes—we will see a more practical/useful approach when we consider the binomial distribution.

The Bernoulli distribution

The Bernoulli distribution

The term *probability distribution* is used to describe a generic template that can be assigned to a random variable.

One of the simplest distributions is the Bernoulli distribution. For a fixed probability p, we write

$$X \sim \mathrm{Bern}(p)$$

to indicate that X follows a Bernoulli distribution. In this case, X models a single attempt, where the probability of a succesful attempt is p.

- The sample space is $\Omega_X = \{0, 1\}$.
- The probability mass function is

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

- $E(X) = p$
- $\mathrm{Var}(X) = p(1 - p)$

Interpretation

The Bernoulli distribution is a model for a **single** trial where the random variable counts the number of "successes" (zero successes or one success).

Examples:

- Let X denote the number of heads after flipping **one** coin. Then $X \sim \text{Bern}(0.5)$.
- If you have a 0.0001 probability of winning the lottery, and you buy just **one** ticket, then this follows a Bern(0.0001) distribution.
- Suppose you roll a die **once**, and rolling a 6 counts as a "success". Let Y equal 1 if you roll a six, and let Y equal 0 otherwise. Then $Y \sim \text{Bern}(\frac{1}{6})$.

The Bernoulli distribution is not typically useful in practice, but it serves as a useful starting point for the underlying mathematics.

Success and failure

Important: the words 'success' and 'failure' should not always be taken literally. In other words, a success is not always a "good outcome" or a "positive outcome".

In the context of probability distributions, the word *success* is a convenience term: it refers only to a particular outcome meeting specified conditions.

- If a computer has a 1% probability of crashing on a single day, then a "success" could be defined as the computer crashing. Think of it as the crash "successfully" occurring. If the computer does not crash, this is called a "failure".

You can reframe a "success" as a "yes instance"

- Did the computer crash?
  - Yes $\Rightarrow$ "success"
  - No $\Rightarrow$ "failure"

For a different problem, you might do the opposite, and call a day without a crash a "success". It all depends on context.

The binomial distribution

The binomial distribution

For a given number of attempts n and probability p, we write

$$X \sim \mathrm{Bin}(n, p)$$

to indicate that X follows a binomial distribution. In this case, X models the number of successful attempts out of n attempts overall, where the probability of success is p.

- The sample space is $\Omega_X = \{0, 1, 2, \ldots, n\}$.
- The probability mass function is

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

- $E(X) = np$
- $\mathrm{Var}(X) = np(1 - p)$

Interpretation

A binomial distribution models a situation where:

- You know how many attempts will be made (the parameter n).
- Each attempt has the same probability of success (the parameter p).

Examples:

- Let X denote the number of heads attained after flipping 10 coins. Then $X \sim \text{Bin}(10, 0.5)$.
- You receive 500 emails, and each has a 1% chance of being spam. If Y denotes the number of spam emails, then $Y \sim \text{Bin}(500, 0.01)$.
- If you roll a die ten times, and rolling a 6 counts as a "success", then the number of successful rolls follows a $\text{Bin}(10, \frac{1}{6})$ distribution.
- If you have 15 entries to a competition, and each entry has a 0.001 chance of winning, then the number of winning entries you have is $\text{Bin}(15, 0.001)$ distributed.

Interpretation

It is important that the number of attempts is known in advance.

- I flip a coin *until* I attain 3 heads. This does **not** follow a binomial distribution, because you don't know in advance how many times you will flip.
- How many car accidents will there be tomorrow? This does not follow a binomial distribution either. There are no 'attempts' being made in this case.

Revisiting an earlier example

Recall the problem from a previous video:

> *Suppose you flipped a coin 5 times and counted the number of heads. There are then 6 possible results:*
>
> $$0, 1, 2, 3, 4, 5.$$
>
> *What is the probability of each outcome?*

This is much easier with the binomial distribution!

Frame the problem:

- There are 5 attempts, so $n = 5$.
- Each heads has probability 0.5, so the probability p of 'success' is 0.5.

Thus, the number of heads flipped follows a distribution. If we let X denote the number of heads, then:

$$X \sim$$

Revisiting an earlier example

Recall: if $Y \sim \text{Bin}(n, p)$, then

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

Let $X \sim \text{Bin}(5, 0.5)$. The probability mass function for X in table form is:

x	0	1	2	3	4	5
$P(X = x)$						

Another example

It is also more versatile, since it does not assume success and failure are equally likely.

> *Suppose it is known that, on average, 2% of credit card transactions per day are fraudulent. On one day, 432 transactions are made.*

1. What is the expected number of fraudulent transactions in that day?
2. What is the probability that there are no fraudulent transactions?

Another example

Suppose it is known that, on average, 2% of credit card transactions per day are fraudulent. On one day, 432 transactions are made.

Many practical applications will be expressible in terms of inequalities.

- What is the probability that fewer than 10 transactions are fraudulent?
- What is the probability that at least 15 transactions are fraudulent?

We will see how these can be calculated in the video on cumulative probabilities.

The geometric distribution

The geometric distribution

For a given probability p, we write

$$X \sim \mathrm{Geo}(p)$$

to indicate that X follows a geometric distribution. In this case, X models the number of failed attempts before one success, where the probability of success is p.

- The sample space is $\Omega_X = \{0, 1, 2, \ldots\}$.
- The probability mass function is

$$P(X = k) = (1 - p)^k \cdot p$$

- $E(X) = \frac{1-p}{p}$
- $\mathrm{Var}(X) = \frac{1-p}{p^2}$

Important: the single success is *not* counted by X.

Interpretation

A geometric distribution models a situation where:

- Each attempt has the same probability of success (the parameter p).
- You will stop as soon as the first success occurs.

Examples:

- Let X denote the number of heads attained before flipping tails once. Then $X \sim \text{Geo}(0.5)$.
- You contact potential customers until you make your first sale. Each contact has a 10% chance of making a purchase. If Y denotes the number of customers contacted before the first sale is made, then $Y \sim \text{Geo}(0.1)$.
- Products are being manufactured one-by-one, with a 5% chance of having a defect. As soon as a defective item is found, production will be stopped to determine the cause. The number of functional items produced before one defective item is found follows a Geo(0.05) distribution.

Example

Products are being manufactured one-by-one, with a 5% chance of having a defect. As soon as a defective item is found, production will be stopped to determine the cause.

1. What is the probability that the first item produced is defective?
2. What is the probability that the third item produced is defective?

Comments

Often, we want to speak in terms of *at least k* failures.

> *Products are being manufactured one-by-one, with a 5% chance of having a defect. As soon as a defective item is found, production will be stopped to determine the cause. What is the probability that at least 3 functional items are produced before a defective item is found?*

If X denotes the number of functional products, then we would want to find $P(X \geqslant 3)$. So we would need to find:

$$P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + \dots$$

...an infinite sum!

It is easier to reframe using complements:

$$P(X \geqslant 3) = 1 - P(X < 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

The negative binomial distribution

The negative binomial distribution

For a given number of successes r and probability p, we write

$$X \sim \text{NB}(r, p)$$

to indicate that X follows a negative binomial distribution. In this case, X models the number of failed attempts before r successes, where the probability of success is p.

- The sample space is $\Omega_X = \{0, 1, 2, \ldots\}$.
- The probability mass function is

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

- $E(X) = \frac{r(1-p)}{p}$
- $\text{Var}(X) = \frac{r(1-p)}{p^2}$

Important: the r successes are *not* counted by X.

Interpretation

A negative binomial distribution models a situation where:

- Each attempt has the same probability of success (the parameter p).
- You will stop after the first r successes occur.

Examples:

- A child plays a claw machine with a 5% chance of winning a prize. The child will stop only after winning 3 prizes. Then the number of failed attempts follows a $\mathrm{NB}(3, 0.05)$ distribution.
- You contact potential customers until you make 5 sales. Each contact has a 10% chance of making a purchase. If Y denotes the number of failed attempts at making a sale, then $Y \sim \mathrm{NB}(5, 0.1)$.
- Products are being manufactured one-by-one, with a 5% chance of having a defect. To try and detect the cause of the problem, the manager wants to make products until 10 defective items are found. The number of functional products produced before the 10 defective products are found follows a $\mathrm{NB}(10, 0.05)$ distribution.

Interpretation

Note carefully that the r successes are not included in the count. Suppose X was a random variable following a negative binomial distribution. If a sequence of attempts was:

$$F\ F\ F\ S\ F\ F\ F\ F\ S\ F\ F\ F\ F\ F\ F\ S,$$

then X takes the value 13 (because there are 13 failures).

- What is the probability that 7 **attempts** will be needed for 3 successes to be had?
  - Taking 7 attempts minus 3 successes means there would be 4 failures. Thus we require $P(X = 4)$.
- What is the probability that 7 **failures** will occur before 3 successes?
  - Since failures are specifically mentioned, we would require $P(X = 7)$.

Example

A child plays a claw machine with a 5% chance of winning a prize. The child will stop only after winning 3 prizes.

1. If each attempt costs \$1, what is the expected cost for the child?
2. What is the probability that the child spends exactly \$25?

Cumulative probabilities

Cumulative probabilities

Cumulative probabilities occur when expressing a probability in terms of an inequality. Key phrases that could occur include:

- *at least*
- *at most*
- *no more than*
- *fewer than*
- *up to*
- *between*

There are other phrases that could occur, so understanding inequalities will be important when solving word problems.

Useful rules

Some useful rules include:

- $P(X \leqslant a) = P(X < a) + P(X = a)$
- $P(X < a) = P(X \leqslant a) - P(X = a)$
- $P(X \geqslant a) = 1 - P(X < a)$
- $P(a \leqslant X \leqslant b) = P(X \leqslant b) - P(X < a)$
- $P(a < X \leqslant b) = P(X \leqslant b) - P(X \leqslant a)$

This is not an exhaustive list—there are too many possible combinations to write them all.

Don't try to memorise them. Instead, try to understand the logic.

- Adding probabilities amounts to including the event.
- Subtracting probabilities amounts to excluding the event.

$$P(a \leqslant X \leqslant b) = P(X \leqslant b) - P(X < a)$$

Example

Suppose it is known that, on average, 2% of credit card transactions per day are fraudulent. On one day, 432 transactions are made.

1. What is the probability that fewer than 2 transactions are fraudulent?
2. What is the probability that at least 2 transactions are fraudulent?
3. What is the probability that between 2 and 4 transactions are fraudulent?

Example

Products are being manufactured one-by-one, with a 5% chance of having a defect. As soon as a defective item is found, production will be stopped to determine the cause.

1. What is the probability that at least 3 functional items are made?
2. What is the probability that the defective item is one of the first 5 items?

Example

A child plays a claw machine with a 5% chance of winning a prize. The child will stop only after winning 3 *prizes.*

1. If each attempt costs \$1, what is the probability that the child spends more than \$5?
2. If each attempt costs \$1, what is the probability that the child spends less than \$5?

Using R

Many probability distributions are built into R. Probabilities of the form $P(X \leqslant a)$ can be calculated using the `pxxxxx` functions. For example,

- `pbinom` for the binomial distribution
- `pgeom` for the geometric distribution
- `pnbinom` for the negative binomial distribution

There are many built-in distributions—to see a list, type `?distributions`.

Parameters for the distribution need to be specified. It is best to check the documentation to see the syntax.

For probabilities of the form $P(X > a)$, the optional argument `lower.tail` can be set to `FALSE`.

For any other form, you must re-express it to be in terms of probabilities of the form $P(X \leqslant a)$ and $P(X > a)$.

Example: using R

Suppose it is known that, on average, 2% of credit card transactions per day are fraudulent. On one day, 432 transactions are made.

1. What is the probability that there are fewer than 10 fraudulent transactions?
2. What is the probability that there are at least 15 fraudulent transactions?
3. What is the probability that the number of fraudulent transactions is between 10 and 20 inclusive?

Example: using R

Products are being manufactured one-by-one, with a 5% chance of having a defect. As soon as a defective item is found, production will be stopped to determine the cause.

1. What is the probability that fewer than 8 functional items are made?
2. What is the probability the defective item occurs after the first 13 items are made?

Example: using R

A child plays a claw machine with a 5% chance of winning a prize. The child will stop only after winning 3 prizes. Each attempt costs \$1.

1. Suppose the child has \$30 to spend. What is the probability the child wins the three prizes?
2. What is the probability that the child spends less than \$10?