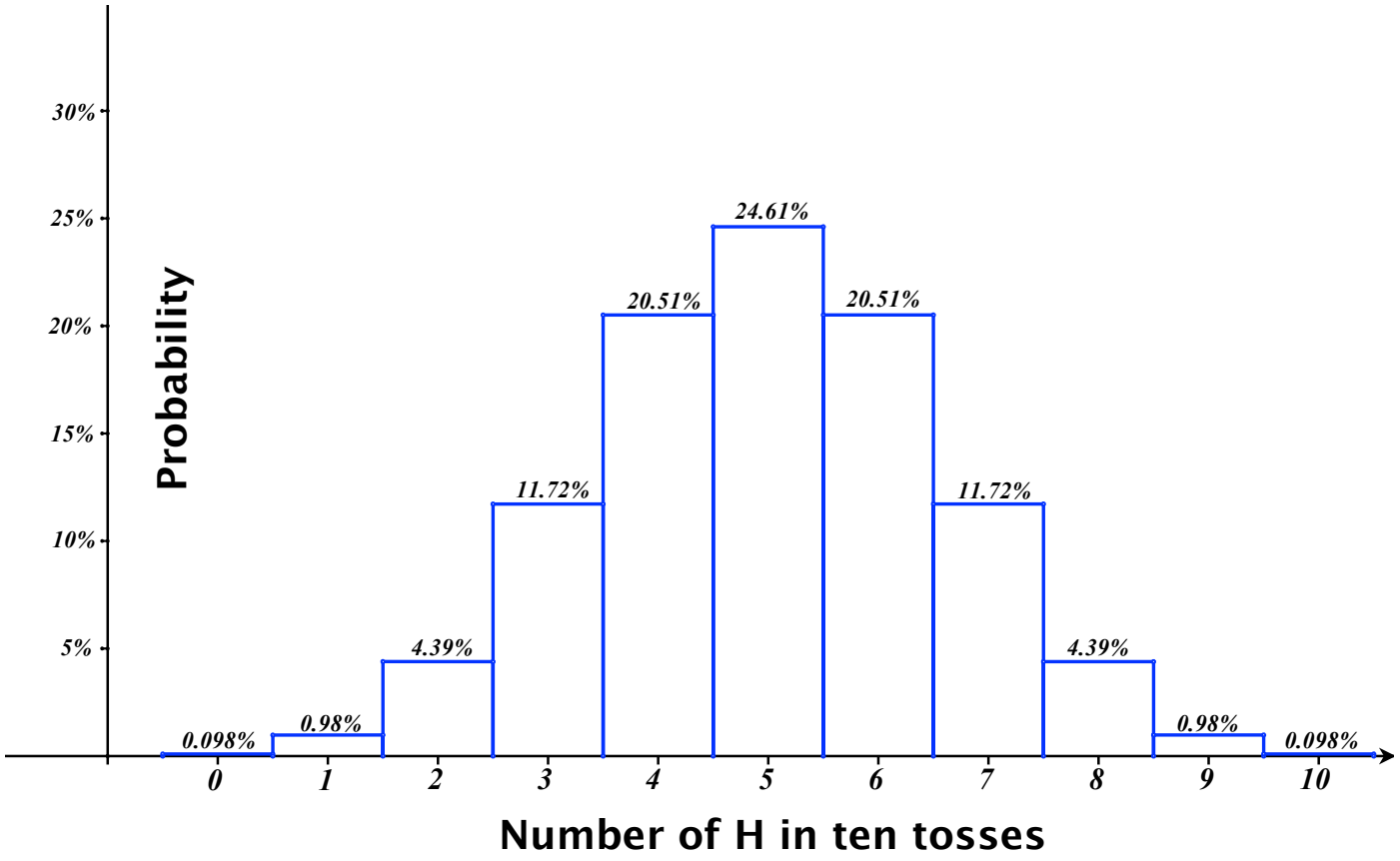
Coin tosses

If a fair coin is tossed 10 times, what will we see?



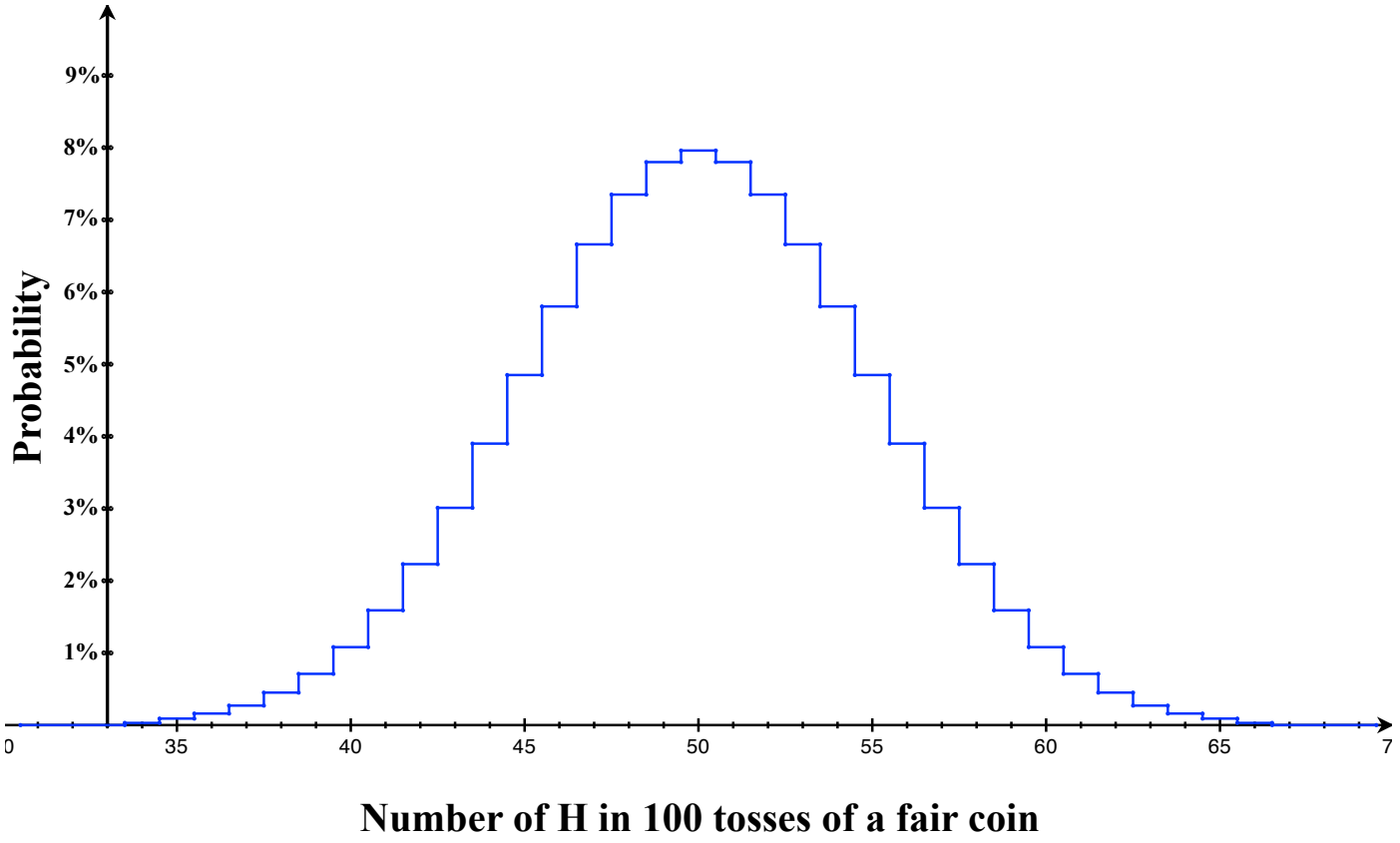
Probability histogram for the number of H in 10 tosses of a fair coin.

Observations:

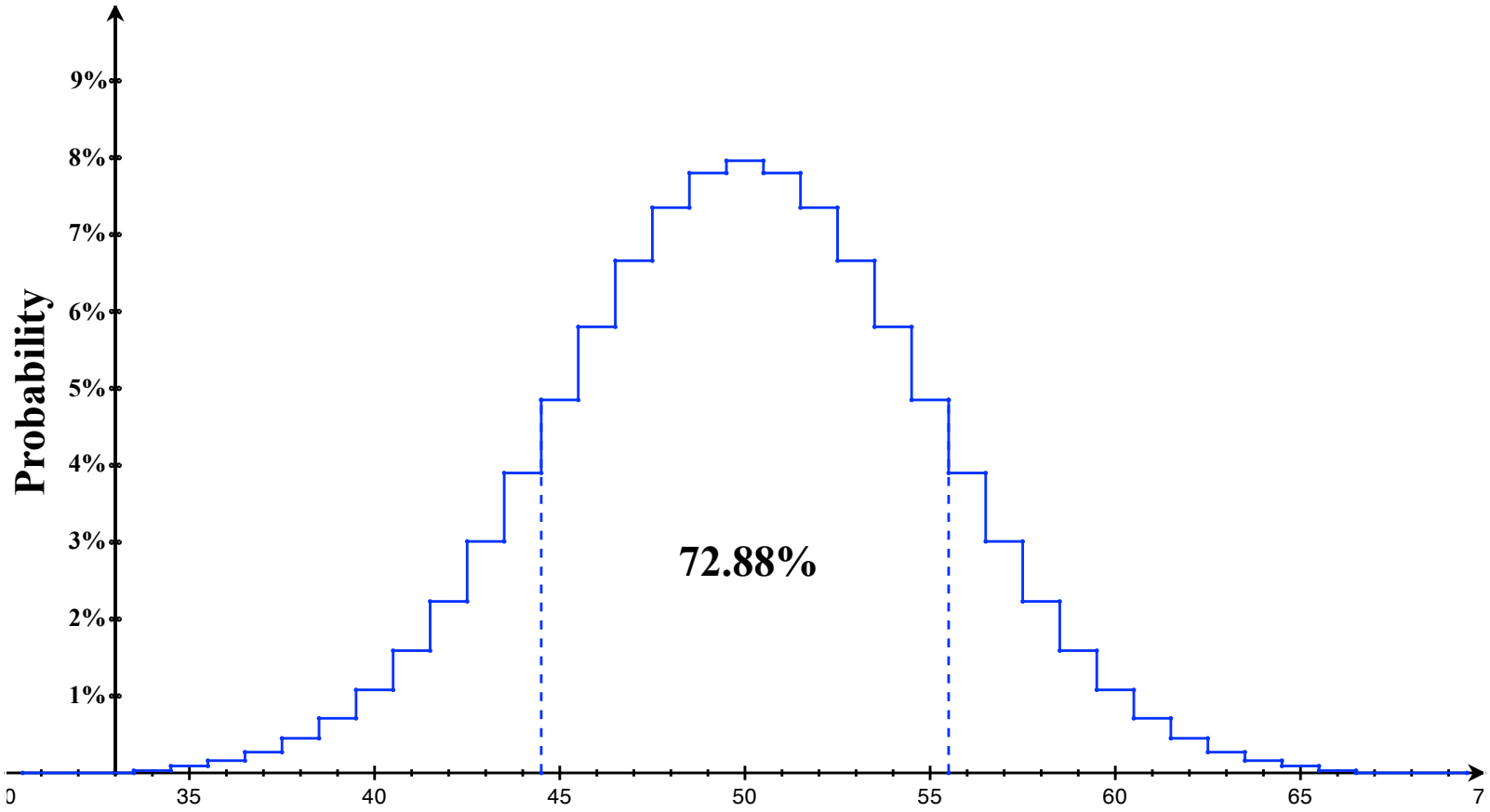
- The probability of seeing exactly 5 H in 10 tosses is just below 25%.
- The probability that the number of H is between 4 and 6 is close to 66%.
- The probability that the number of H is between 3 and 7 is about 89%.
- We could summarize these numbers by saying that we will ***probably*** see ***about*** 5 H in 10 tosses of a fair coin.

More coin tosses

If a fair coin is tossed 100 times, what will we see?



Probability histogram for the number of H in 100 tosses of a fair coin.



Probability histogram for the number of H in 100 tosses of a fair coin.

There is a probability of 72.88% that we will see between 45 and 55 H in 100 tosses.

Observations:

- The probability of seeing exactly 50 H in 100 tosses is 7.96%.
- The probability that the number of H in 100 tosses is between 49 and 51 is 23.56%.
- The probability that the number of H in 100 tosses is between 48 and 52 is 38.26%.
- The probability that the number of H in 100 tosses is between 47 and 53 is 51.58%.
- The probability that the number of H in 100 tosses is between 46 and 54 is 63.18%.
- The probability that the number of H in 100 tosses is between 45 and 55 is 72.88%.

Analogously to the 10-toss scenario, we can say that in 100 tosses of a fair coin, we will ***probably*** see ***about*** 50 H.

Question: *In which of the two scenarios is our prediction more accurate?*

Answer: *It depends on how we are measuring the accuracy.*

- In terms of the ***number*** of H, the prediction for 10 tosses gives a narrower range of possible values with a higher probability.
- In terms of the ***proportion*** of H, the prediction for 100 tosses gives a narrower range of possible ***percentages*** with higher probability.
  - The probability is 66% that ***percentage*** of H in 10 tosses will be between 40% and 60%.
  - The probability that the ***percentage*** of H in 100 tosses is between 46% and 54% is 63.18%.
  - The probability is 72% that ***percentage*** of H in 100 tosses will be between 45% and 55%.

The 3R7B box: ***Ten tickets are drawn at random with replacement from a box that contains three red tickets and seven blue tickets...***

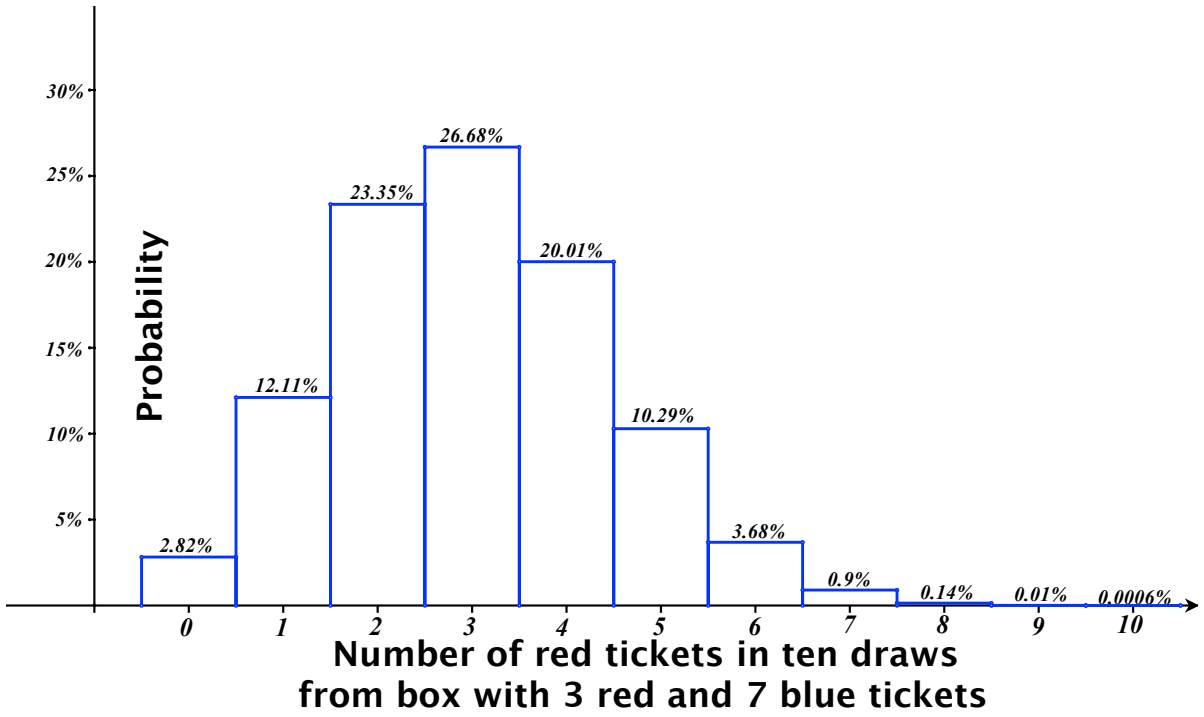
Questions:

1. How many red tickets do you ***expect*** to see when you draw 10 tickets from the 3R7B box?
2. How ***accurate*** is our answer to the first question ***likely*** to be?
3. What does '***expect***' mean in this context?

Answers:

1. We expect ***about*** 3 red tickets in 10 draws... *Why?*

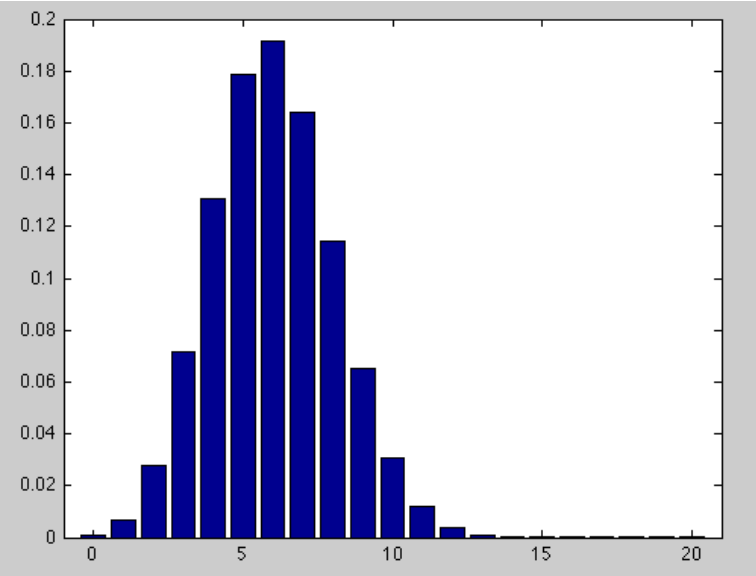
2. To answer question 2., we need to study the probabilities of all the possible outcomes:



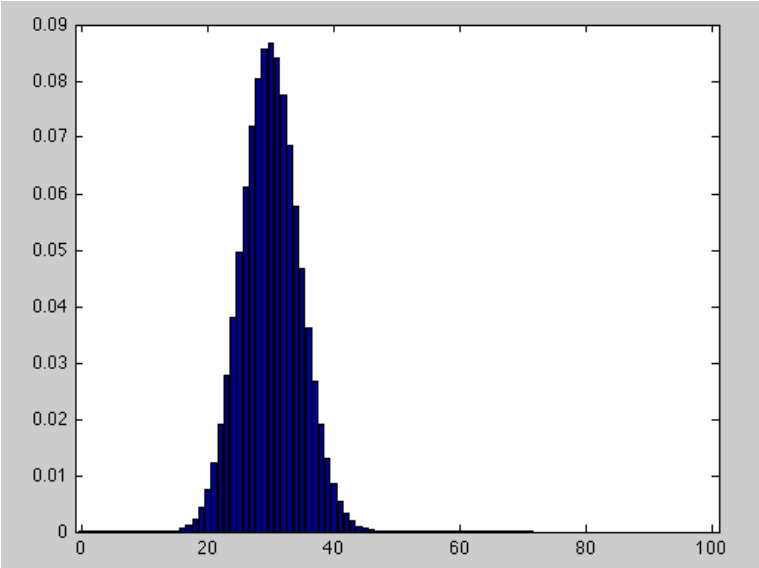
- The probability of exactly 3 red tickets in 10 draws is 26.68%.
- The probability that the number of red tickets in 10 draws is between 2 and 4 is about 70%.

The small number of draws is a little misleading (as it was in the case of coin tosses):

As the number of draws gets bigger, the ***observed number*** of red tickets is more and more likely to be farther and farther from the ***expected number*** of reds tickets.



20 draws



100 draws

Observations:

- The most likely number of red tickets in all three examples is

  $$P(\text{red ticket in one draw}) \cdot (\text{number of draws}).$$

  This is the ***expected number*** of red tickets in each case.

- The probability that we see precisely the expected number of red tickets ***decreases*** as the number of draws increases. From about 27% (10 draws), to about 19% (20 draws), to about 8.5% (100 draws), to about 2.75% (1000 draws).

- The probability that the ***number*** of red tickets is close (e.g., within 2) of the expected number also decreases:

  $P(\text{between 1 and 5 red tickets in 10 draws}) = 92.45\%$

  $P(\text{between 4 and 8 red tickets in 20 draws}) = 77.96\%$

  $P(\text{between 28 and 32 red tickets in 100 draws}) = 41.43\%$

  $P(\text{between 298 and 302 red tickets in 1000 draws}) = 13.69\%$

As the number of draws from the box increases, the chance increases that the observed number of red tickets will deviate significantly from the expected number of red tickets:

- In 10,000 draws from the 3R7B box, the probability that the number of red tickets is ***more than*** 30 away from 3000 is about 50.57%
- In 1,000,000 draws from the 3R7B box, the probability that the number of red tickets is ***more than*** 300 away from 300,000 is about 51.2%.

Proportions, not numbers

It is more useful to compare the ***observed percentage*** of red tickets drawn to the ***expected percentage***.

⋆ The expected percentage of red tickets is the same as the percentage of red tickets in the box, namely 30%.

- In 10 draws from the 3R7B box, the probability that between 20% and 40% of the tickets are red is about 0.70.
- In 20 draws from the 3R7B box, the probability that between 20% and 40% of the tickets are red is about 0.78.
- In 100 draws from the 3R7B box, the probability that between 20% and 40% of the tickets are red is about 0.98.
- In 1000 draws from the 3R7B box, the probability that between 20% and 40% of the tickets are red is more than 0.99.

- In 1000 draws from the 3R7B box, the probability that between 28% and 32% of the tickets are red is about 0.85.
- In 10000 draws from the 3R7B box, the probability that between 28% and 32% of the tickets are red is more than 0.99.
- In 10000 draws from the 3R7B box, the probability that between 29% and 31% of the tickets are red is about 0.97.
- In 10000 draws from the 3R7B box, the probability that between 29.5% and 30.5% of the tickets are red is about 0.73.
- In 1,000,000 draws from the 3R7B box, the probability that between 29.5% and 30.5% of the tickets are red is more than 0.99.

Summarizing:

As the number of draws from the 3R7B box increases, the chance approaches 100% *that the observed percentage of red tickets is very close to the expected percentage.*

In other words,

> *If many tickets are drawn at random with replacement from the 3R7B box, then it is very likely that about 30% of the tickets will be red.*

This is called the ***Law of Averages*** (for the 3R7B box).

The "law of averages" does <u>not</u> say that...

- ...we will ***definitely*** see ***exactly*** 30% red tickets

  or

- ... we will ***probably*** see ***exactly*** 30% red tickets.

  or

- ... we will ***definitely*** see ***about*** 30% red tickets.

There is always a (small) chance that the observed percentage of red tickets will be far from the expected percentage, even for an extremely large number of draws.

The law of averages, in general:

If tickets are drawn from a box containing $\boxed{1}$*s and* $\boxed{0}$*s , then as the number of draws increases, the* ***probability*** *approaches 100% that the* ***observed percentage of*** $\boxed{1}$***s*** *is very close to the* ***expected percentage of*** $\boxed{1}$***s*** *(= the percentage of* $\boxed{1}$*s in the box).*

Comments:

- The law is true for draws with replacement and for draws without replacement. In fact, the results are even sharper when the draws are done without replacement.
- The difference between the observed *number* of $\boxed{1}$s and the expected *number* of $\boxed{1}$s is ***likely to get bigger*** as the number of draws grows (for draws with replacement).
- The law of averages does ***not*** say anything about what will happen on the ***next draw***.