

## Is it for real?

Using statistics to make decisions

## True art or a fake?

An online quiz by Mikhail Simkin  
[http://reverent.org/true\\_art\\_or\\_fake\\_art.html](http://reverent.org/true_art_or_fake_art.html)

## What should the pass mark for the quiz be?

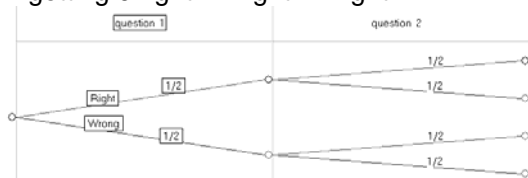
- Imagine the quiz is the first phase in choosing an art expert to appear on TV. What do you think the pass mark should be? We don't want someone who guessed them all and got lucky to end up on TV!

## Galton's grid

Francis Galton was a cousin of Charles Darwin. He worked on heredity, statistics and the classification of fingerprints. His grid consists of pins; small beads fall and go to left or right at a pin with equal probability.

## Probability

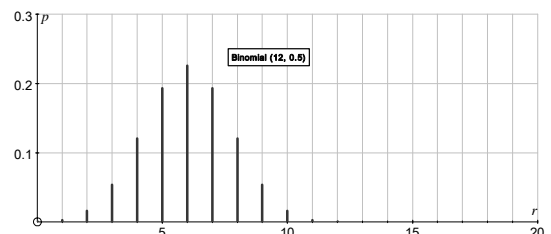
If there were just 2 questions in the quiz, what would be the probability of a guesser getting 0 right? 1 right? 2 right?



What if there were 3 questions?

4 questions?

## Probabilities for a guesser



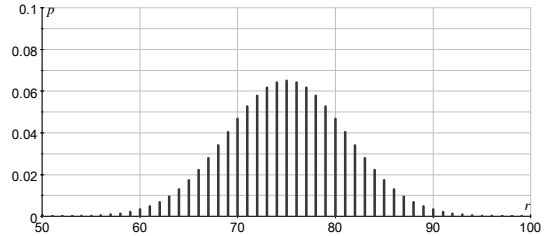
## 150 tosses of a fair coin

HtttHtttHtHtHtHHttHttHHHtHHttttHHtH  
 tttHHtttHHtHttttHHHHHtHHHHHttHH  
 HHHtHHtHttHttHtttHttttHHttHttttHHH  
 HHHtHHtttHHHHHttHHttttHHHtttHH  
 tHHtHHtttHHHHHttt

75 heads, 75 tails.

Real or fake?

## Probabilities of different numbers of heads on 150 tosses of a fair coin



## Tossing a coin

- A tosses a fair coin until he gets the sequence HTT.
- B tosses a fair coin until she gets the sequence HTH.

Which of the following is true?

1. On average, A takes longer than B.
2. On average, B takes longer than A.
3. On average, they take the same time.

## One way to think about it

Write out the ways you can get the streak in exactly 3, 4, 5, ... tosses:

No. of tosses	3	4	5	6
Ends HTT	HTT	HHTT THTT		
Ends HTH	HTH	HHTH THTH		

## Penney's game

- Invented by mathematician Walter Penney in 1969.
- Brought to public attention by Derren Brown in 2009.
- Two players take it in turns to toss a coin.
- Player A picks a sequence of 3 outcomes (e.g. HTT). A wins if this sequence comes up.
- No matter what sequence A chooses to look for, player B can pick a different sequence that is more likely to come up first.

## Winning Penney's game

- Whatever sequence of three tosses A chooses, B can win by choosing
  - The opposite of A's middle choice as his first choice
  - A's first choice as his second choice
  - B's middle choice as his third choice
  - So A picks HTH, B picks HHT

### HTH: A wins; HHT: B wins

- HHH
- HHT
- HTH
- HTT
- THH
- THT
- TTH
- TTT

### Streaks of heads

- Tossing  $n$  coins.
- $U_n$  = number of ways of getting a sequence without 5 (or more) heads in a row
- $U_1 = 2, U_2 = 4, U_3 = 8, U_4 = 16, U_5 = 31$
- For more than 5 coins, if there are no more than 4 heads in a row, the sequence must end with one of these:  
T or TH or THH or THHH or THHHH and the bit before the ending must have no more than 4 heads in a row

- $U_6 = U_5 + U_4 + U_3 + U_2 + U_1$
- $U_7 =$
- $U_n =$
- [coins.xls](#)

### 150 tosses of a fair coin

HtttHtttHtHtHtHHttHttHHHtHHttttHttH  
tttHHtttHHtHttttHHHHHtHHHHHtHH  
HHHtHHtHttHttHtttHttttHHttHttttHHH  
HHHtHHtttHHHHHtHttHttttHHHtttHH  
tHHtHHtttHHHHHttt

75 heads, 75 tails.  
Real or fake?

### Pay per click

- If an internet user clicks on an advert, the advertiser pays the website hosting the advert.
- The website makes money.
- The advertiser pays in proportion to people viewing the advert.
- Not all internet advertising is pay per click.

### Click Fraud

- Imagine you are advertising on the internet.
- If your competitor repeatedly clicks on your advert, it will cost you money.
- If the website hosting the advert repeatedly clicks on it, they will earn money for nothing.

### Click fraud in court

- Google and Yahoo were taken to court in 2004 by an advertiser who thought he had been overcharged; they settled out of court.
- In June 2009 Microsoft started a civil court case against individuals suspected of click fraud.

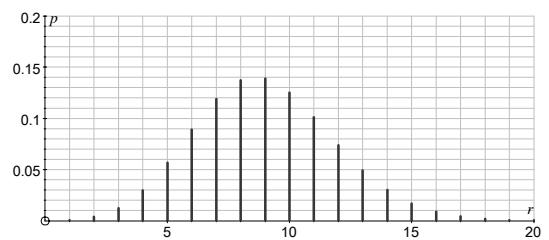
### Detecting click fraud

- Search engines have systems for detecting click fraud. They keep these secret.
- Many adverts will not count you more than once a day no matter how often you click.

### Conversion rate

- The advert may want the customer to buy a product or ask for further details etc. The percentage of clickers who do this is the conversion rate.
- Suppose the typical conversion rate for a particular advert is 10%. One day there are 90 clicks. How low can the number of conversions go before I suspect there is a problem?

90 clicks, probability of conversion = 0.1



graph