# Perspectives on Quantitative Trading

*It Pays to be Bayes*

Saurabh Singal

Nov 19, 2014

# What Works…

-

$$\varepsilon_{mix} := \left[ \nu_1 \cdot \left( \varepsilon_1^{\frac{1}{3}} \right) + \nu_2 \cdot \left( \varepsilon_2^{\frac{1}{3}} \right) \right]^3$$ *where* $\nu_1, \nu_2$ *are the volume fractions of materials 1 and 2 ,*

*and ..* $\varepsilon_1, \varepsilon_2$ *are the full complex ( two component) dielectric constants of the full density materials 1 and 2 .*

*Note :* $\varepsilon_1 := \varepsilon'_1 - j \cdot \varepsilon''_1$ $\quad \varepsilon_2 := \varepsilon'_2 - j \cdot \varepsilon''_2$

*Note that* $\nu_1 + \nu_2 := 1.$ *and the density of the mixture is* $\rho_{mix} := \rho_1 \cdot \nu_1 + \rho_2 \cdot \nu_2$

*and that* $$\frac{\rho_{mix}}{\rho_1} := \nu_1 + \left( \frac{\rho_2}{\rho_1} \right) \cdot \left( \nu_1 - 1 \right)$$

$$\varepsilon_{mix} := \left[ \nu_1 \cdot \left( \varepsilon_1^{\frac{1}{3}} \right) + \left( 1 - \nu_1 \right) \cdot \left( \varepsilon_2^{\frac{1}{3}} \right) \right]^3$$

# Qui Numerare Incipit Errare Incipit

- **He who begins to count, begins to err …**
- Title of an article by the great Oskar Morgenstern, founder of Game Theory along with von Neumann.
- Niels Bohr said: *It's tough to make predictions, especially about the future.*
- Don’t predict the past, it can be embarrassing if you are wrong!
- We can study only the past, not the future…

# Statistical Paradoxes

- The pop-behavioral finance & pop-psychology books have had such a good run that most of us know all about various kinds of biases.

- Anchoring, Frame Dependence, Mental Accounting, Availability Bias, Representativeness, Overconfidence, Self-Attribution Bias, Illusion of Control, Confirmation Bias, Hindsight Bias, Recency Bias, the Winner's Curse, Endowment effect, Status Quo Bias, Loss Aversion, Regret Avoidance, Lottery Effect,…

- However, there are some very important **statistical biases** related to sample selection that are less commonly talked about.

- Theses biases can induce spurious correlations.

# Statistical Paradoxes and Fallacies

- Statistical Paradoxes
  - Simpson's Paradox
  - Berkson's Paradox
  - Stage Migration or the Will Rogers's Effect
  - Base Rate Neglect
  - Conservatism
- The last two are related to the Bayes' Rule.

# **SMOKING INCREASES LONGEVITY:** Simpson's Paradox

- **Simpson's paradox**: a reversal of the direction of an association when data from several groups are combined to form a single group.
  - For example, When the data are examined as one group, the association between X and Y is positive, but when the data are split into groups based on some other characteristic, say W, the association between X and Y is negative.
  - This is a type of **Omitted-Variable Bias (OVB)** : occurs when a model incorrectly leaves out one or more important causal factors. The model compensates for the missing factor by over- or underestimating the effect of one of the other factors.

# SMOKING INCREASES LONGETIVITY:
## Simpson's Paradox-2

| | Smoker ? | |
|---|---|---|
| | Yes | No |
| Dead | 107 | 132 |
| Alive | 174 | 175 |
| Total | 281 | 307 |
| % Dying | 38.10% | 43.00% |

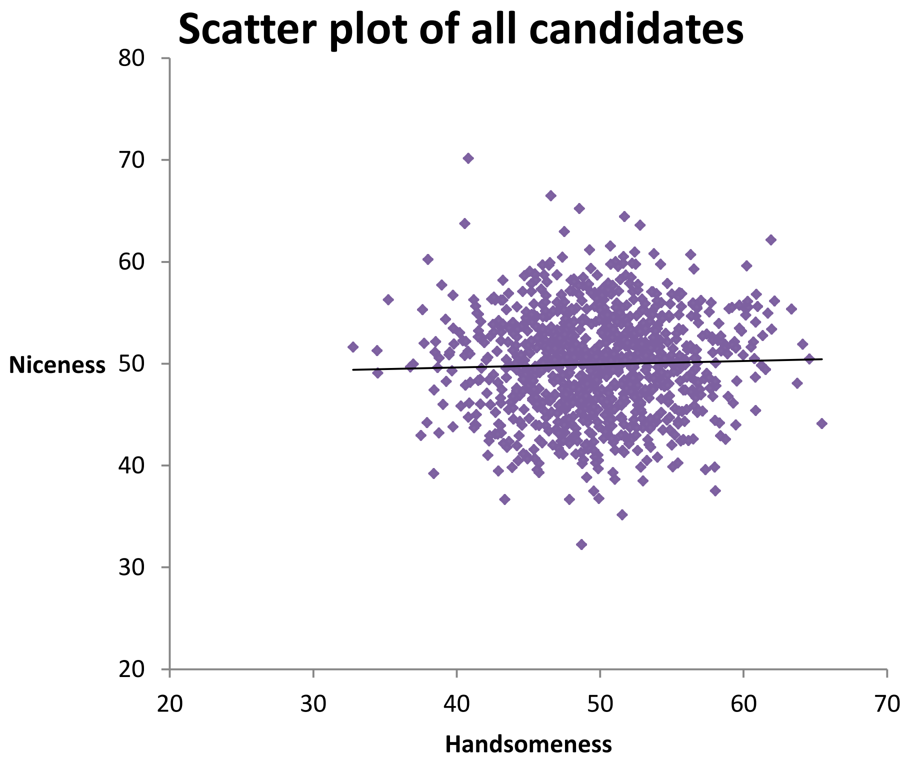
| | Age Group | | | | | |
|---|---|---|---|---|---|---|
| | 45-54 | | 55-64 | | 65-74 | |
| | Smoker ? | | Smoker ? | | Smoker ? | |
| | Yes | No | Yes | No | Yes | No |
| Dead | 27 | 12 | 51 | 40 | 29 | 101 |
| Alive | 103 | 66 | 64 | 81 | 7 | 28 |
| Total | 130 | 78 | 105 | 121 | 36 | 129 |
| % Dying | 20.80% | 15.40% | 48.60% | 33.10% | 80.60% | 78.30% |

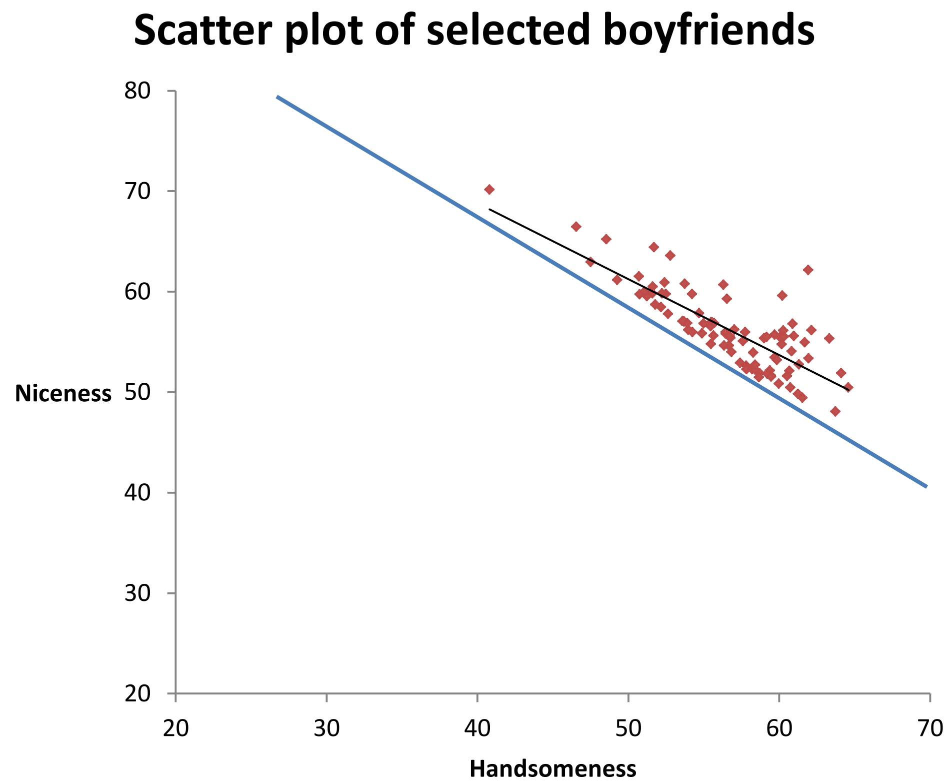
Source: http://www.stat.osu.edu/~biostat/newsletters/volume2_2/article_vol2_2.html

# Why Are Handsome Men Such Jerks?
## Berkson's Paradox

- Title of a Slate article by the Wisconsin professor, Jordan Ellenberg
- "you will date a man if his niceness plus his handsomeness exceeds some threshold. … So, among the men that you date, the nicer ones are less handsome on average (and vice-versa), even if these traits are uncorrelated in the general population."
- This kind of selection bias is called **Berkson's Paradox**.
- Beware of this when using sum of two ranks to select a universe.

# Candidates Ranked on Handsomeness and Niceness




The black lines are regression lines. Data on left show no correlation.
Negatively sloped regression line in right chart shows negative correlation. Only points above blue line can be selected, leading to sampling bias.

# Stage Migration

## Will Rogers Effect

- Will Rogers remarked: *When the Okies left Oklahoma and moved to California, they raised the average intelligence level in both states.*
- Improved techniques in detection of illness leads to some people being classified as "unhealthy" at an early stage.
- Stage Migration is the name given to this movement of people from the set of healthy people to the set of unhealthy people.
- This leads to the increase in average life span of both the healthy and unhealthy groups.
- Both lifespans are statistically lengthened, even if the patients did not live longer.
- Example : $A=\{1,2,3\}$ and $B=\{4,5,6\}$. If we move $4$ from set $B$ to set $A$, the mean of both sets will increase.

# How a Quant Strategy Develops

- Observation
- Common place events
- Unusual Events
- How improbable are these unusual events and what can be learned from them?
- Simplicity - Hemingway spent time forgetting tough words.
- Not  too simple - Brain surgery is impossible with a pen knife.
- We have already met the Omitted-Variable Bias.

# October 2014

- In October 2014,
  - the world's equity markets had a string of big negative days.
  - Many equity indices were in "correction" territory or down more than 10%. And talks of bear markets were rampant.
  - And then the equity markets also had a string of positive days.

- We could ask some questions:
  - What is the probability of a 1-day move>1%?
  - Or, drill down : in a bull market, what is the probability of a string of large 1-day moves
  - Or invert the question:

# Probability Inversions

- Is it the case that bear markets have more large up days than bull markets?
- Given that we saw a large number of big positive days, what is the probability we are in a bear market?
- Examine Not $P(\text{many big positive days } | \text{bull market})$ but
  $P(\text{we are in a bear market } | \text{many big positive days})$ ?

**Need to use Bayes' Theorem**

# Flaw of Averages
## November 2014

- Title of book by Sam Savage,  son of the great statistician James Savage
- James Savage wrote the famous book *How to Gamble if You Must.*
- How a statistician could drown in a river that was "on average" only three feet deep.
- In November 2014, The equity markets had a mind numbing dull sequence of days:

| | |
|---|---|
| 6-Nov-14 | 2031.21 |
| 7-Nov-14 | 2031.92 |
| 10-Nov-14 | 2038.26 |
| 11-Nov-14 | 2039.68 |
| 12-Nov-14 | 2038.25 |
| 13-Nov-14 | 2039.33 |
| 14-Nov-14 | 2039.82 |

- Average Volatility is moderate now. Use Gaussian Mixture model- very low volatility, very high high volatility and some medium volatility.

# Biases related to Bayes' Rule

$$P(X|D) = \frac{P(X \cap D)}{P(D)} = \frac{P(D|X) \times P(X)}{P(D)}$$

Which yields

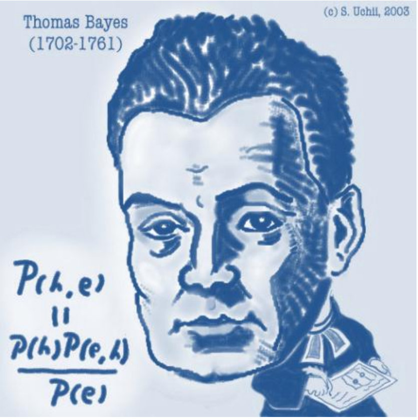
$$P(X|D) \propto P(D|X) \times P(X)$$

Where

$P(X|D)$ is the posterior probability

$P(D|X)$ is the likelihood

$P(X)$ is the prior probability




- Two biases arise which are related to Bayes' Rule
  - Base Rate Fallacy (ignoring the prior or underweighting it significantly)
  - Conservatism (slowness in revising our beliefs in the light of new information)

# Example 1: Boy-Girl paradox…

- A new neighbor moves in next door. You learn that he has two children.
- You see that one of the children is a boy.
- What is the probability that the other child is a girl?

# Boy-Girl Paradox: using Bayes' Theorem

Cases:
1. Both are girls (GG) : P(GG) = 1/4,
2. Both are boys (BB) : P(BB) = 1/4,
3. One boy, one girl- two case, (GB) and (BG)- with probability= (1/4)*2= 1/2

- Now we add the additional assumption that "at least one is a boy" : call this B.

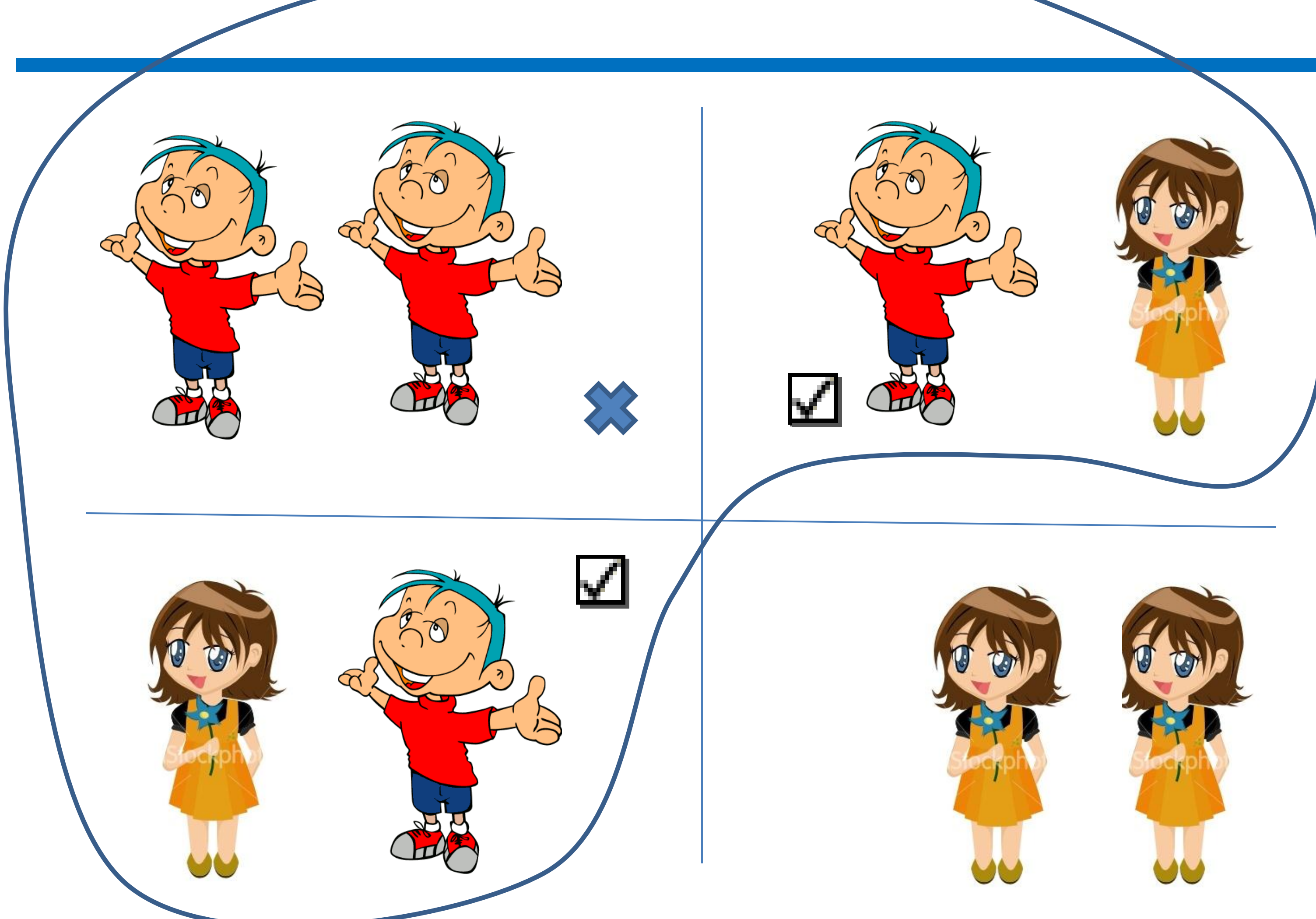
Using Bayes' Theorem, we find:
- $P(BB|B) = P(B|BB) \cdot P(BB) / P(B)$

- $P(B|BB)$ = probability of at least one boy given both are boys = 1
- $P(B)$ = probability of at least one being a boy, which includes cases (2) and (3)
  = 1/4 + 1/2 =3/4

Finally, **P(BB|B)** = ( 1 * 1/4 )/ (3/4 ) = **1/3**

**So the probability other child is a girl is 2/3**

# Boy Girl Paradox… Intuitive Solution



Out of 3 possible cases with at least one boy, there are two cases with a girl-

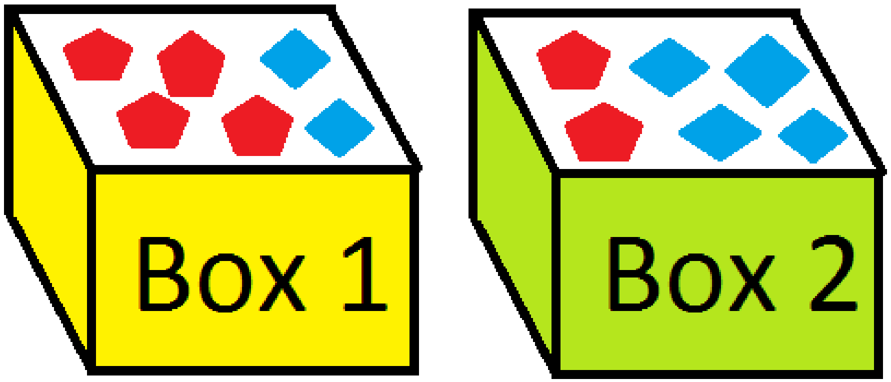
$$Prob(other\ child\ is\ a\ girl) = \frac{2}{3}$$

# Application of Bayes' Theorem in Finance

- **Black-Litterman Framework**
  - In the Markowitz mean-variance optimization framework – the stocks or assets have a multivariate-normal distribution.
  - If we want to revise our estimates of return on individual stocks the proper way is to go through the variance-covariance matrix due to the correlations – i.e., to compute the distribution conditional upon our views. This is the Black-Litterman Framework.
- **Bayes-Stein shrinkage**: We "shrink" stock beta towards unity. Very high values of beta are lowered somewhat, and very low values get increased.
- **Less-Technical Example**
  - Finance people and ordinary people are somewhat Bayesians without being conscious of it. For example, the research analysts revise their price targets based on the current or historical prices.
  - And any one who takes into account the long term drift in equities is accounting for the base rate. For the S&P 500, about 54% days are positive, as are about 60% months, 66% quarters and 75% years.

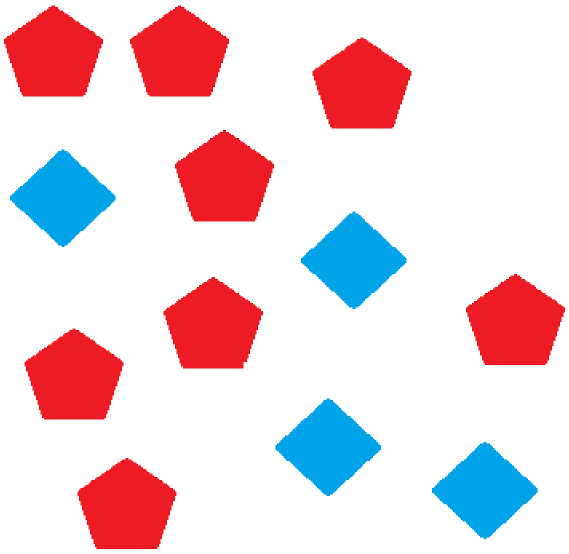
# Let’s Play a Game!

- There are two boxes containing red and blue checkers.
- **Box 1** has 100 red and 50 blue checkers.
- **Box 2** has 50 red and 100 blue checkers.



# Let's Play a Game: 2

- We select a box at random
- Assume there is 50% chance of either box being selected.
- Sampling with replacement, we pull 12 checkers out of this same box
- Of the 12 checkers,  8 are red and 4 are blue.
- What is the chance we picked them from the second box? (the one with 50 red and 100 blue checkers)



# Let’s Play a Game: Answer

- The correct answer is 1/17, which is close to 6%.

- It is difficult at first blush for most of us to guess this.
  - The prior probability of either box being selected is 50%.
  - When we see more red than blue checkers, we can assume there is a higher than 50% chance that they are drawn from box 1, which has more red checkers.
  - so the chance of box 2 having been selected is $< 50\%$.
  - It is difficult to guess it is so much lower.

- The human brain is very good at pattern recognition but poorly designed for computing probabilities, esp. conditional probabilities.

# Dangers of Chart Reading

- Human eye-brain system is good at the cognitive task of comparing position along a common scale.
- However, it is not so good at elementary perceptual tasks that involve making judgment about length, area, shading, direction, angle, volume, and curvature.
- Source: William S. Cleveland and Robert McGill, "*Graphical Perception: Theory, Experimentation and Application to the Development of Graphical Methods*."
- Use of charts can be dangerous as we can fall prey to illusions.

# Three Optical Illusions

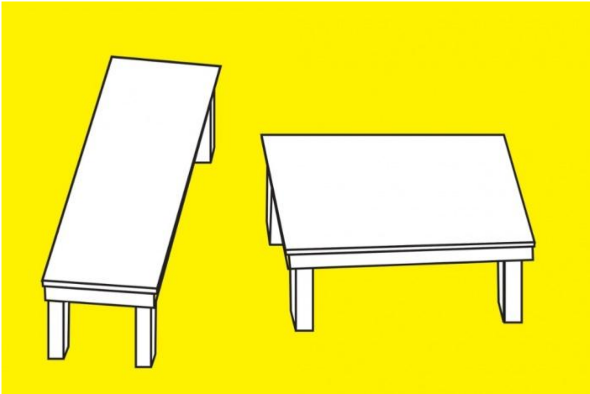


*Figure 1.*

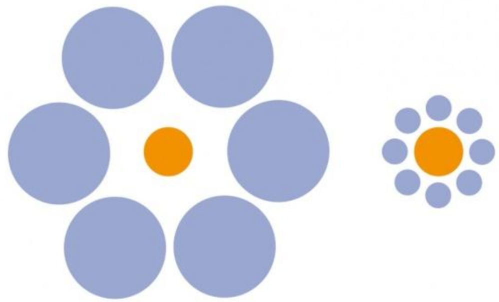


*Figure 2.*

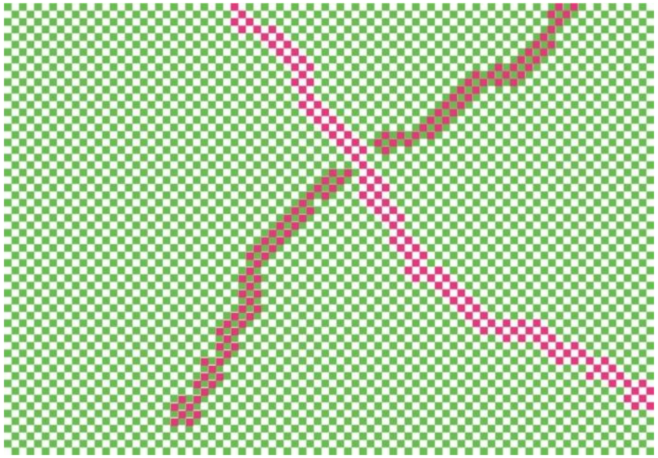


*Figure 3.*

Figure 1: The two tables have same length.
Figure 2: The two orange circles have the same radius.
Figure 3. The two lines of pixels which appear pink and red are of the exact same shade.
*Source: Eye Benders by Clive Gifford*.

# Seemingly Useful But…

- Are seemingly useful data sometimes *useless*?

- Oskar Morgenstern wrote a book ***On the Accuracy of Economic Observations***. He discussed how officials decide economic policy based on Economic numbers based on estimates have that very large standard errors. Point estimates are reported and not intervals, and also without the estimates of uncertainty.

- Recent example: June 25, 2014 news release stated (Bureau of Economic Analysis, 2014):"Real gross domestic product . . . decreased at an annual rate of 2.9 % in the first quarter of 2014 according to the "third" estimate released by the Bureau of Economic Analysis. . . . The GDP estimate released today is based on more complete source data than were available for the "second" estimate issued last month. In the second estimate, real GDP was estimated to have decreased 1.0%."

# Bayes' Theorem Revisited

$$P(X|D) = \frac{P(X \cap D)}{P(D)} = \frac{P(D|X) \times P(X)}{P(D)}$$

- Sometimes we can use other more visual & intuitive methods to examine such probabilities as in examples 1 and 2.

- But sometimes, the inverse probabilities cannot be derived without Bayes' Theorem as example 3 shows.

# Seemingly Useless But…

- Are seemingly useless data sometimes *useful*?
- We have all read of the Monty Hall paradox about 3 doors, 2 goats and a car. The proper strategy is to switch. Your chances of winning goes up from 1/2 to  2/3.
- At first blush, it might appear that the knowledge of one of the doors with a goat is useless. But appreciation of what this knowledge signifies is a key factor.

# Monty Hall using Bayes' Theorem

- Let ***A, B, C*** be the events that prize is behind $door_A$ , $door_{B,}$ $door_c$ respectively.
- Suppose the Player picks $door_A$ .
- Let ***O*** be the event that the show host reveals $door_B$ .
- We want to know whether $P(A|O) < P(C|O)$
- ***Bayes' Theorem*** *tells us that*
  $P(A|O) = \frac{P(A) \times P(O|A)}{P(O)}$ *and* $P(C|O) = \frac{P(C) \times P(O|C)}{P(O)}$
- $P(O|A) = \frac{1}{2}$, (if $door_A$ has prize, either $door_B$ or $door_c$ can be opened)
- $P(O|B) = 0$, (if $door_B$ has the prize, it won't be opened.)
- $P(O|C) = 1$, (if $door_c$ has the prize, $door_B$ will be opened.)

# Monty Hall using Bayes Theorem-2

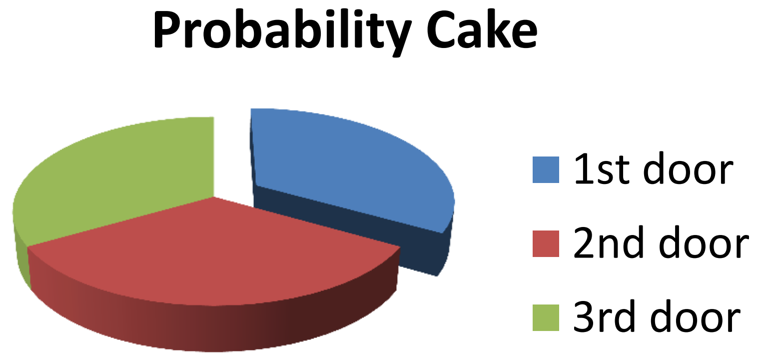
$$P(O) = P(O \cap A) + P(O \cap B) + P(O \cap C)$$
$$= P(A)P(O|A) + P(B)P(O|B) + P(C)P(O|C) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{2}$$

$$P(A|O) = \frac{P(A) \times P(O|A)}{P(O)} = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{2}} = \mathbf{\frac{1}{3}}$$
$$P(C|O) = \frac{P(C) \times P(O|C)}{P(O)} = \frac{\frac{1}{3} \times 1}{\frac{1}{2}} = \mathbf{\frac{2}{3}}$$

- **We double the chance of winning if we switch.**

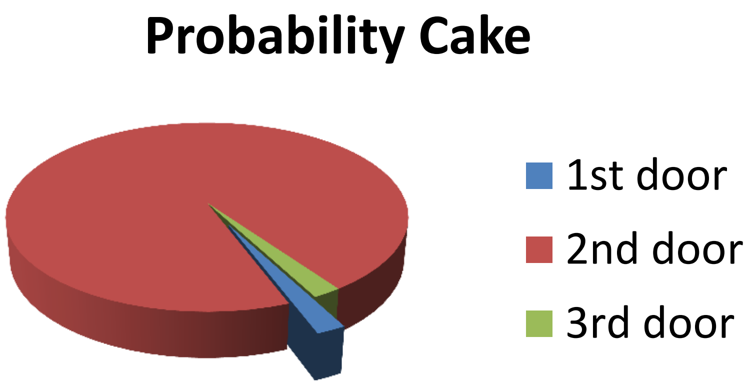
# Monty Hall: Dollar Under Cake

- Imagine again the same problem but represent it as a cake with three equal slices, under one of which is a silver dollar.
- If you picked up one slice, there is 2/3rd chance the dollar is in the remaining part of the cake.
- Given a chance to switch, you should, even if you know some part of the REMAINING CAKE, say the red slice, does not have the dollar.



# Monty Hall: Dollar Under Cake-2

- Imagine if there are 100 doors, the player chooses door 1.
- Now the chance of this door containing the prize is 1%
- The chance of the "rest of the cake" containing the dollar is 99%
- Shouldn't the player switch if given the chance?



# Example 2: Bayesian Reasoning

## Base Rate Fallacy

### Testing for HIV

Here **positive** test MEANS test says “**there is an infection**”

- About 0.01 % men are affected with HIV.
- If an ***infected*** man is tested, there is a 99.9% chance the test result is positive.
- If a man is ***not infected***, there is a 99.99% chance he will test negative.

**What is the chance a man who tests positive is infected with the HIV virus?**

# Example 2: Bayesian Reasoning-2

$$Prob\,(Infected\,|Test\,Positive) = \frac{Prob\,(Infected\ AND\ Test\ Positve)}{Prob\,(Test\ Positive)}$$

$$= \frac{Prob\,(Test+ve\,|\,Infected) \times Prob\,(Infected)}{Prob\,(Test+ve\,|\,Infected) \times Prob\,(Infected)\ +\ Prob\,(Test+ve\,|\,NOT\ Infected) \times Prob\,(NOT\ Infected)}$$

$$= \frac{99.9\% \times 0.01\,\%}{99.9\% \times 0.01\,\% + 0.01\,\% \times 99.9\%} \approx \frac{0.01\%}{0.01\% + 0.01\%} = \frac{1}{2}$$

**Probability(Infected, given test is positive)**

**= approx 50%.**

*This is far lower than what most people guess.*

*It is easy to forget that the incidence rate of this disease is very low.*

*Both the **test accuracy**, and the **incidence rate** will influence the probability of a positive test result.*

# Example 2: Bayesian Reasoning (without tears)

Assume 10,000 men.
Of these 10,000 men, 1 is infected. He will test positive.
Of the 9,999 men, 1 tests positive (9999 X 0.0001 = 0.9999)

10,000

1 (has HIV) | 9,999 (no HIV)

1 +ve | 0 - ve | 1 +ve | 9,998 - ve

Two men test positive, of which only one has HIV.
So if he tests positive, there is only a 50% chance a man has HIV

**Prob ( infected| positive test )** vs. **Prob ( positive test| infected )**

**Two are different (Prosecutor's fallacy)**

# Example 3: Drunk Driver

*When Bayes' Rule is essential, though problem seems easy*

- Sober wife to Drunk Husband: Don't drive!
- DH: Why not?
- SW: Because it is unsafe. 25% accidents caused by drunk drivers…
- DH: It is the safer alternative, as 75% accidents must be caused by sober drivers!
- Shall we hire a blind chimpanzee to chauffeur us?
- Since less than 25% of drivers are drunk, of course drunk drivers are causing more than their proportionate share of accidents

# Example 3: Drunk driver

The currect probability to examine is of course

Prob(Accident | Driver is Drunk) and compare it with

Prob(Accident | Driver is Sober)

- *There is no direct way to compute the probability a drunk driver has an accident and **without Bayes' theorem we cannot proceed.***

- Suppose we know 5% of drivers are drunk.

$$P(A|D) = P(D|A)*P(A)/P(D),$$

$$P(A|S) = P(S|A)*P(A)/P(S)$$

where P(D)= probability a driver is drunk, P(S) = prob. driver is sober, P(A)= probability a drive results in an accident.

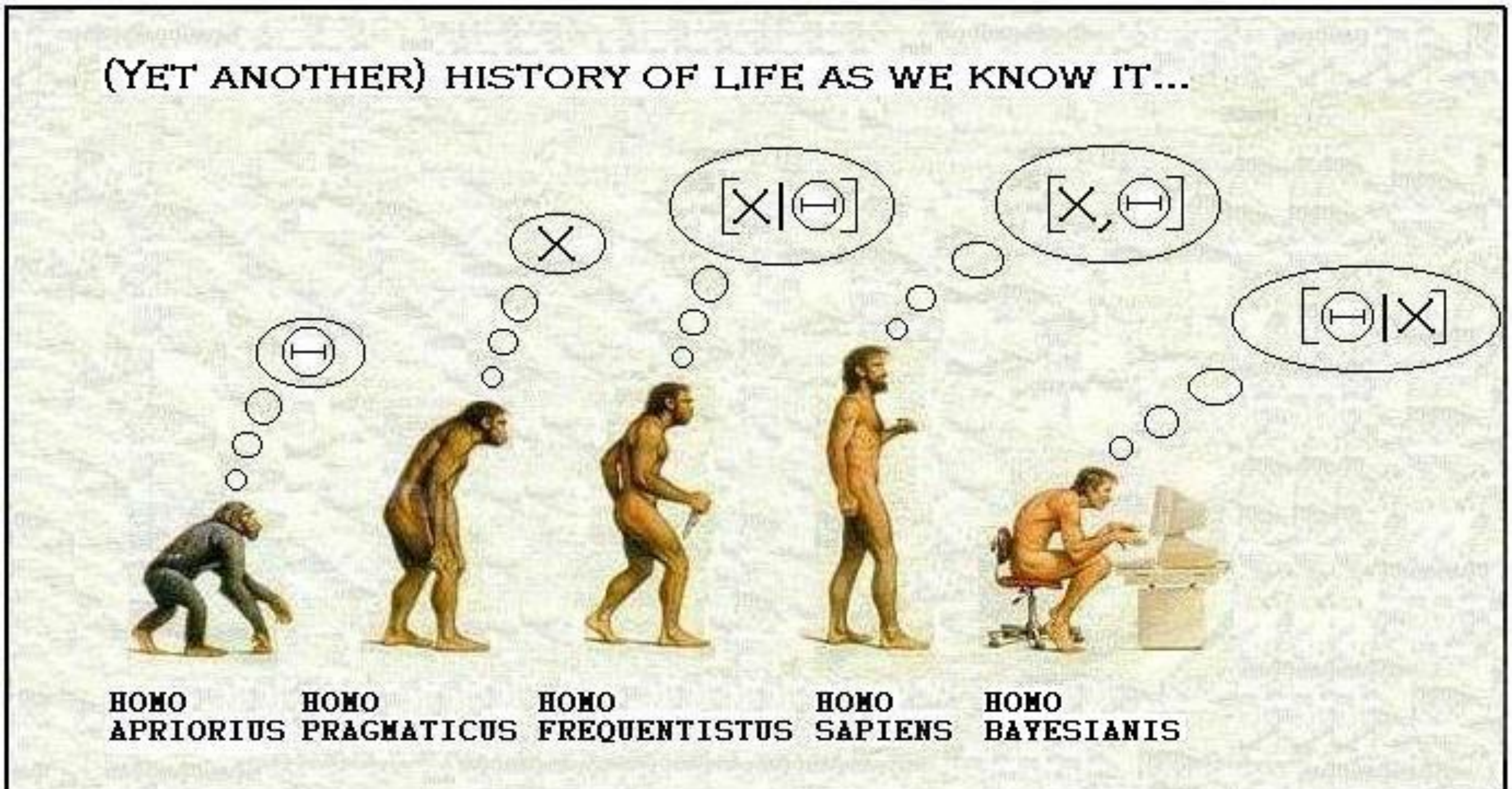
$$P(A|D)/P(A|S) = [P(D|A)/P(S|A)]*P(S)/P(D)$$

$$=(25\%/75\%)*(95\%/5\%)= 19/3$$

$$=6.33$$

- ✓ Thus even though 75% accidents are caused by sober drivers, a drunk driver has 6.33 times as much probability of having an accident as a sober driver.

# We are all Bayesians now!



# Explanation of Cartoon

***Θ*** - Will I be attacked?
***X*** - I heard a noise nearby

- **A priori** - Oh no, someone might attack me!!
- **Pragmatic** - what's that noise? Prepare for an attack!
- **Frequentist** - how often did I hear a noise before I was attacked?
- **Sapiens** - thinks about how often he hears noises and how often he was attacked, but does not have the intuitive or instinctive ability to combine the ideas, as we saw in this talk….
- **Bayesian** - answers the question, if I hear a noise, what is the chance of me being attacked?
- So be aware that our brains do not intuitively give Bayesian results.
- You may not need a computer but if you find yourself wrestling with a problem like the one above, at least reach for a pencil and paper.

# Summary

- We started with:

  He who begins to count, begins to err.
- One must be careful when applying statistics  as biases abound.
- One must question all assumptions underlying a model or analysis.
- One must always remember that:

**He who stops counting, does not stop erring!**