Conditional probabilities continued

“Let’s Make a Deal” show with Monty Hall aired on NBC/ABC 1963-1986

Monty Hall problem

- **Let's Make a Deal.** In Monty Hall's game show *Let's Make a Deal* there are three closed doors. Behind a random one of these doors is a car.
  Behind the other two are goats. The contestant does not know where the car is, but Monty Hall does.
- After the contestant picks a door Monty always opens one of the remaining doors, one he knows does not hide the car. If the contestant has already chosen the door with the car behind, Monty is equally likely to open either of the two remaining doors.
- After Monty has shown a goat behind the door that he opens, the contestant is always given the option to switch doors.
- What is the probability of winning the car under switching and non-switching strategies?

Monty Hall problem.
What gives you better chances to win a car?

A. Better to switch doors
B. Better not to switch doors
C. Switching does not matter
D. It depends on the phase of the moon
E. I don't know

Get your i-clickers

Don’t’ feel bad if you guessed wrong

- When first presented with the Monty Hall problem an overwhelming majority of people assume that each door has an equal probability and conclude that switching does not matter
- Out of 228 subjects in one study, only 13% chose to switch
- Paul Erdős, one of the most prolific mathematicians in history, remained unconvinced until he was shown a computer simulation confirming the predicted result
- Pigeons repeatedly exposed to the problem show that they rapidly learn always to switch, unlike humans

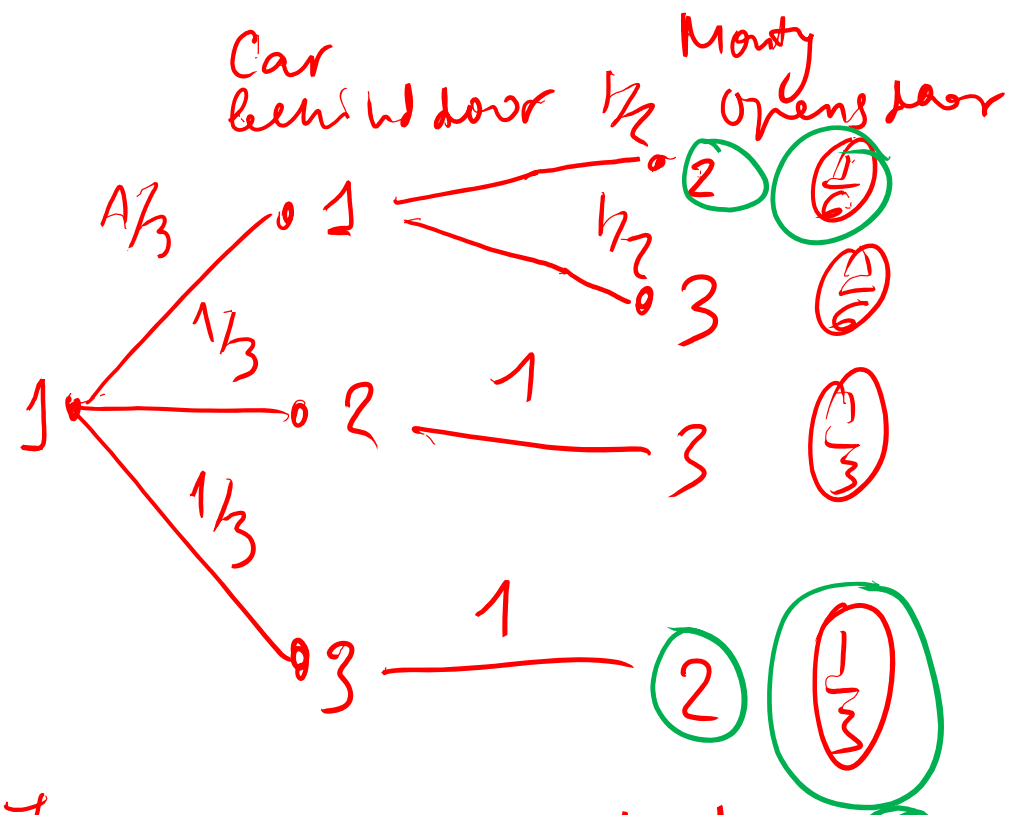
Solution #1 (intuitive)

- With Prob=1/3 you guessed the car door right:
  you loose the car if you switch
  you win the car if you don't switch
- With Prob=2/3 you got it wrong and picked a goat door. Then Monty opens another goat door. What is left is the car door.
  you win the car if you switch
  you loose the car if you don't switch

Solution #2.
Tree & conditional probabilities

Solution #2.
Tree & conditional probabilities

Contestant chooses door 1



If Monty opened door 2

$$\text{Prob}\left(\text{win by switching}\right) = \frac{\frac{1}{3}}{\frac{1}{3}+\frac{1}{6}} = \frac{2}{3}$$

$$\text{Prob}\left(\text{win by staying}\right) = \frac{1/6}{1/3+1/6} = \frac{1}{3}$$

Matlab program skeleton

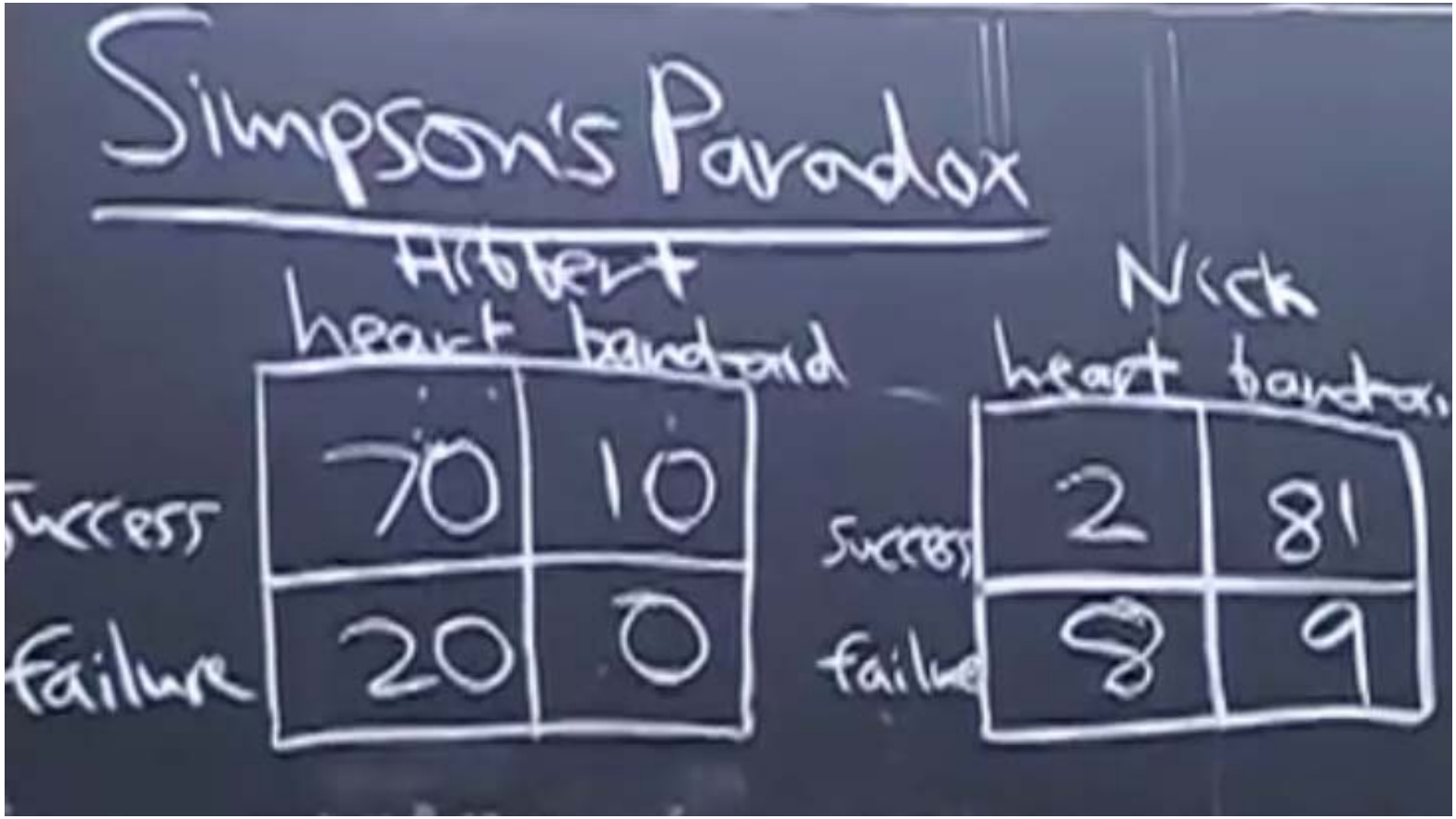
```
%set Stats large...
cars=0; carn=0; %counts successes of switch (cars) and not (carn) strategies. Initialize=0
for i = 1:Stats
a = randperm(3); %Monty places two goats and the car at random
%a(1) –the door for goat #1, a(2) - the door for goat #2, a(3) – the door for car
i= floor(3.*rand)+1; %you randomly select the door!
% SWITCH STRATEGY
if(i == a(1)) cars=cars+1; %a(2)-opened, switch to a(3), car!
elseif (i == a(2)) cars = cars + 1 ;%a(1) opened, switch to a(3), car!
else cars = cars + 0; %a(1)/a(2) opened, switch to a(2)/a(1), no car!
end
% NO SWITCH STRATEGY
if(i == a(1)) carn = carn + ??
elseif (i==a(2)) carn = carn + ??
else  carn = carn +  ??
end;
end;
disp(cars); disp(carn);
```

Simpson's paradox

Is it possible for one doctor to be higher success rate than another in all operations but have lower overall success rate?

Dr. Hibbert Dr. Nick

Using Simpsons cartoon to illustrate Simpson paradox is from Prof. Joe Blitzstein teaching Harvard STAT 110: http://projects.iq.harvard.edu/stat110/youtube

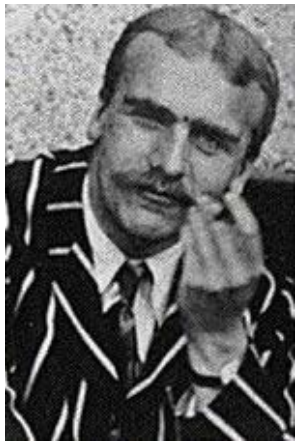


Dr. Hibbert: success rate =80%

Dr. Nick: success rate =83%

Simpson's paradox might explain altruism

- Darwinian evolution has a problem with altruism
- "Selfish gene" does not care about others
- J. B. S. Haldane, (1892-1964) British geneticist, evolutionary biologist



- When asked if he would give his life to save a drowning brother answered: "No, but I would to save two brothers or eight cousins"
- Altruism in some insect colonies like ants is because they are all genetically identical.

Altruism in bacteria

- Bacteria live in communities in close proximity to each other
- Individual bugs spend resources to produce extracellular molecules, excrete them outside of the cell to share with others
- Examples: extracellular enzymes, biofilm components, antimicrobial and anti-immune agents

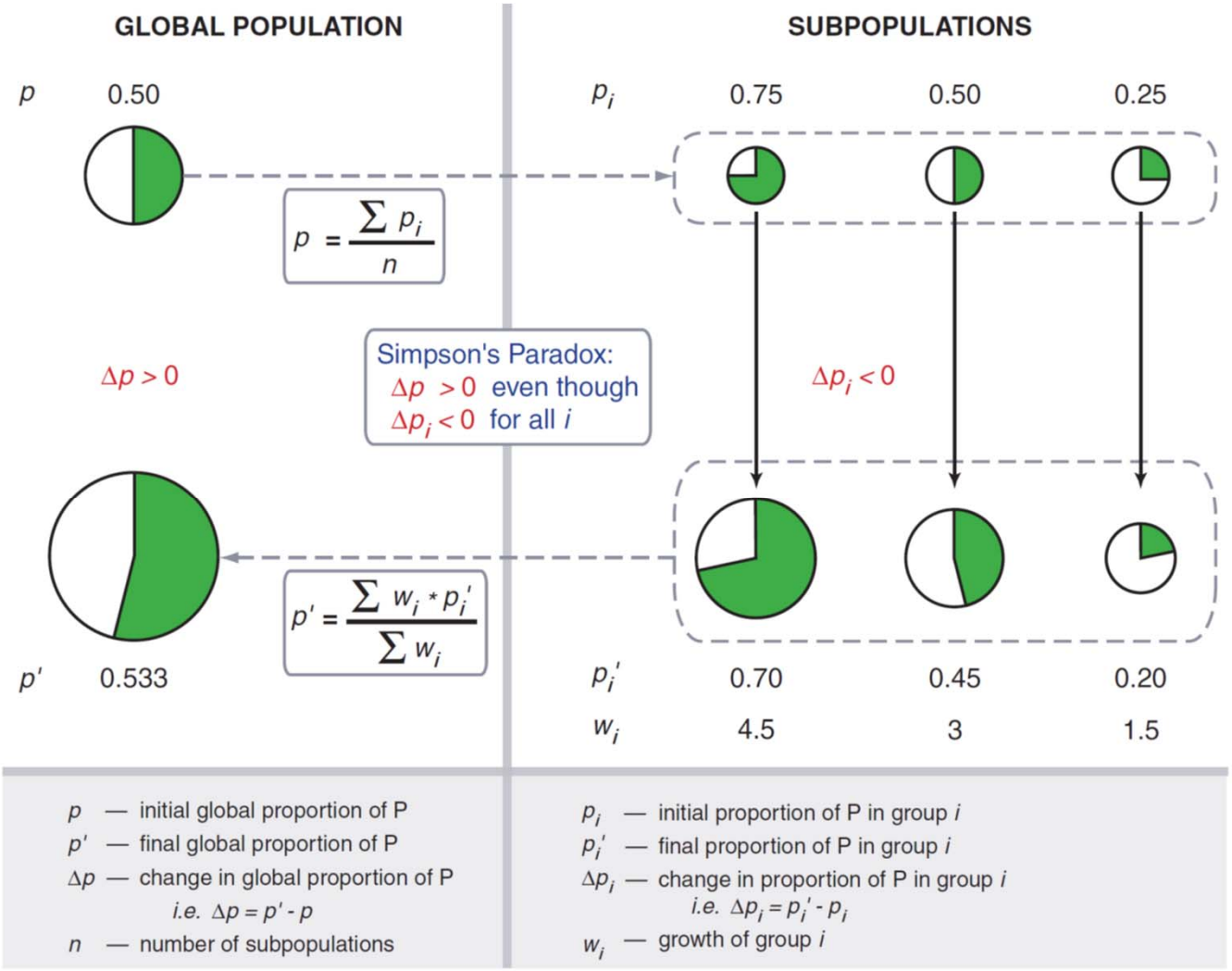
Chuang, Rivoire, and Leibler's answer

Simpson's Paradox in a Synthetic Microbial System

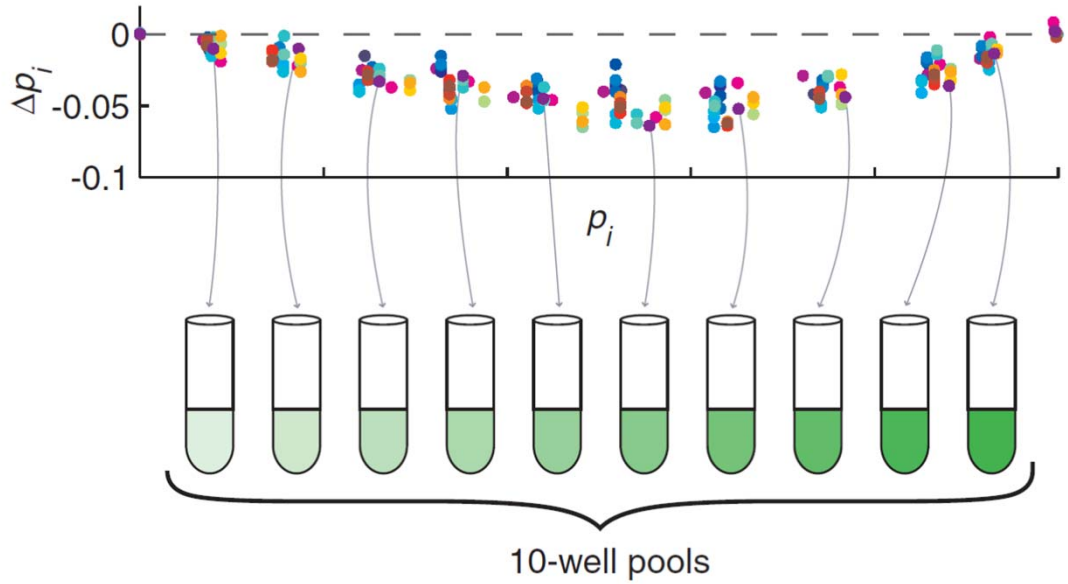
John S. Chuang,* Olivier Rivoire, Stanislas Leibler

The maintenance of "public" or "common good" producers is a major question in the evolution of cooperation. Because nonproducers benefit from the shared resource without bearing its cost of production, they may proliferate faster than producers. We established a synthetic microbial system consisting of two *Escherichia coli* strains of common-good producers and nonproducers. Depending on the population structure, which was varied by forming groups with different initial compositions, an apparently paradoxical situation could be attained in which nonproducers grew faster within each group, yet producers increased overall. We show that a simple way to generate the variance required for this effect is through stochastic fluctuations via population bottlenecks. The synthetic approach described here thus provides a way to study generic mechanisms of natural selection.

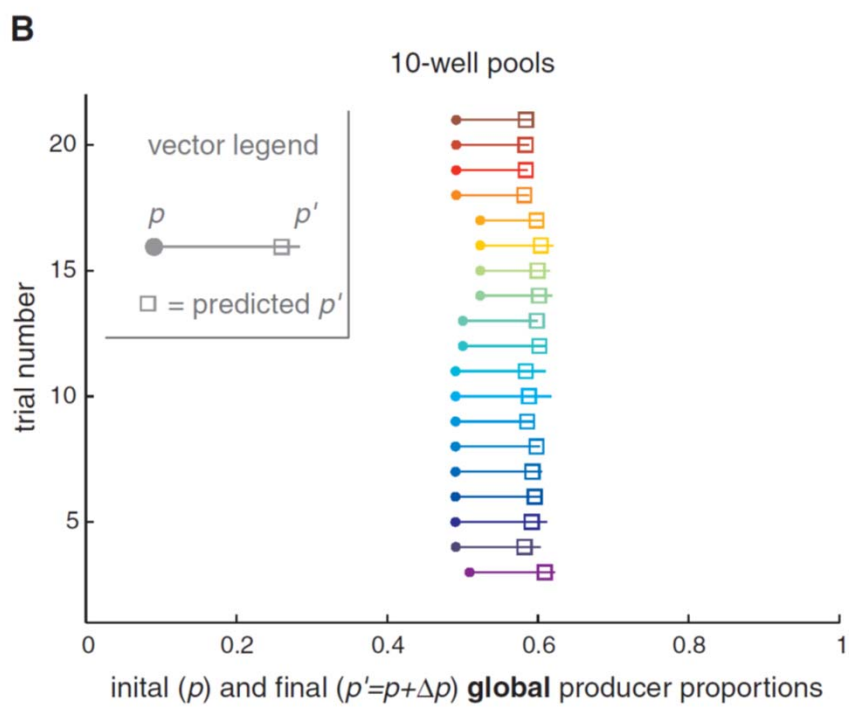
- The common good was a membrane-permeable Rhl autoinducer molecule rewired to activate antibiotic (chloramphenicol; Cm) resistance gene expression.



Fraction of altruists in
each of individual
test tubes dropped



Yet the overall fraction of
altruists in
all test tubes combined
increased



Credit: XKCD comics

QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KLINGONS DIFFERENT
WHY ARE THERE SQUIRRELS
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS SEX SO IMPORTANT
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG
WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY IS YKK ON ALL ZIPPERS
WHY DO IGUANAS DIE
WHY ARE THERE DINOSAUR GHOSTS
WHY IS THERE LAVA
WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND
WHY ARE THERE FEMALE MR MIMES
WHY IS GPS FREE
WHY IS THERE HELL IF GOD FORGIVES
WHY IS LIFE SO BORING

Random variables
Discrete Probability Distributions

Random Variables

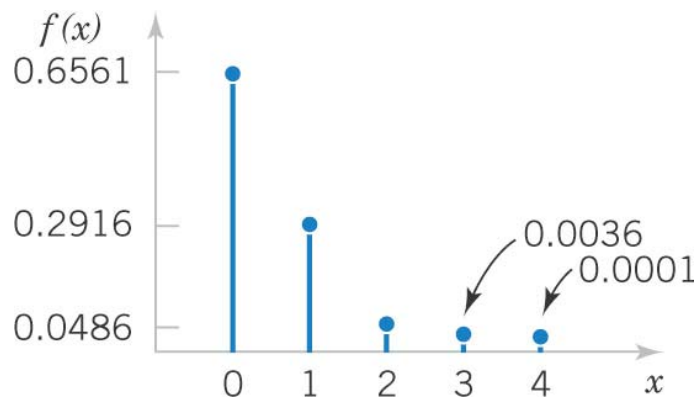
- A variable that associates a number with the outcome of a random experiment is called a random variable.
- Notation: random variable is denoted by an uppercase letter, such as X. After the experiment is conducted, the measured value is denoted by a lowercase letter, such a x.
- The probability of a random variable to have specific value is denoted as $P(X=x)$, e.g. $P(X=4)$

Continuous & Discrete Random Variables

- A discrete random variable is usually integer number
  - N - the number of p53 proteins in a cell
  - D - the number of nucleotides different between two sequences
- A continuous random variable is a real number
  - C=N/V – the concentration of p53 protein in a cell of volume V
  - Percentage (D/L)*100% of different nucleotides in protein sequences of different lengths L (depending on the set of L's may be discrete but dense)

Probability Mass Function (PMF)

- I want to compare all 4-mers in a pair of human genomes
- *X* – random variable: the number of nucleotide differences or Single Nucleotide Polymorphisms (SNPs) in a given 4-mer
- Probability Mass Function: f(x) or P(X=x) – the probability that the # of SNPs is exactly equal to x



Probability Mass Function for the # of mismatches in 4-mers

P(*X* =0) =	0.6561
P(*X* =1) =	0.2916
P(*X* =2) =	0.0486
P(*X* =3) =	0.0036
P(*X* =4) =	0.0001
$\sum_x$ P(X=x)=	1.0000

Cumulative Distribution Function (CDF)

P(X≤x)	P(X>x)		
0.6561	0.3439		
0.9477	0.0523		
0.9963	0.0037		
0.9999	0.0001		
1.0000	0.0000		

CDF: $P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 0.9999$

Complementary cumulative distribution function (tail distribution)

CCDF: $P(X > 3) = 1 - P(X \leq 3) = 0.0001$

Mean or Expected Value of X

The mean or expected value of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x \cdot P(X = x) = \sum_x x \cdot f(x)$$

- The notation: *E[X]* sometimes *EX*, where *E* stands for "expected value".

Apart from mean used for any function of X: *E[f(X)]*

- The mean = the weighted average of all possible values of *X*. It represents its "center of mass"
- The mean may, or may not, be an allowed value of X
- It is also called the arithmetic mean (to distinguish from e.g. the geometric mean discussed later)
- Mean may be infinite if X any integer and $P(X=x)>1/x^2$

Variance $V(X)$: square of a typical deviation from the mean $\mu = E(X)$

$V(X) = \sigma^2$, where σ is called standard deviation

$$\sigma^2 = V(X) = E((X-\mu)^2) =$$

$$= E(X^2 - 2\mu X + \mu^2) = E(X^2) -$$

$$- 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 =$$

$$= E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$

Variance of a Random Variable

If X is a discrete random variable with probability mass function $f(x)$,

$$E\left[h(X)\right] = \sum_x h(x) \cdot P(X = x) = \sum_x h(x) f(x) \qquad (3\text{-}4)$$

If $h(x) = (X - \mu)^2$, then its expectation, $V(x)$, is the variance of X.

$\sigma = \sqrt{V(x)}$, is called standard deviation of X

$\sigma^2 = V(X) = \sum_x (x - \mu)^2 f(x)$ is the definitional formula

$$= \sum_x \left(x^2 - 2\mu x + \mu^2\right) f(x)$$

$$= \sum_x x^2 f(x) - 2\mu \sum x f(x) + \mu^2 \sum_x f(x)$$

$$= \sum_x x^2 f(x) - 2\mu^2 + \mu^2$$

$= \sum_x x^2 f(x) - \mu^2$ is the computational formula

Variance can be infinite
if X can be any integer
and $P(X=x) \geq 1/x^3$

Skewness of a random variable

- Want to quantify how asymmetric is the distribution around the mean?
- Need any odd moment: $E[(X-\mu)^{2n+1}]$
- Cannot do it with the first moment: $E[X-\mu]=0$
- Normalized 3-rd moment is skewness: $\gamma_1=E[(X-\mu)^3/\sigma^3]$
- Skewness can be infinite if X takes unbounded integer values and $P(X=x) \geq 1/x^4$

Geometric mean of a random variable

- Useful for very broad distributions (many orders of magnitude)?
- Mean may be dominated by very unlikely but very large events. Think of a lottery
- Exponent of the mean of *log X*: *Geometric mean=exp(E[log X])*
- Geometric mean usually is not infinite

Summary: Parameters of a Probability Distribution

- The mean, $\mu=E[X]$, is a measure of the center of mass of a random variable
- The variance, $V(X)=E[(X-\mu)^2]$, is a measure of the dispersion of a random variable around its mean
- The standard deviation, $\sigma=sqrt[V(X)]$, is another measure of the dispersion around mean
- The skewness, $\gamma_1=E[(X-\mu)^3/\sigma^3]$, measure of asymmetry around mean
- The geometric mean, $\exp(E[\log X])$, is useful for very broad distributions
- All can be infinite! Practically it means they increase with sample size
- <u>Different distributions can have identical parameters</u>

Credit: XKCD comics

QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE MALE AND FEMALE BIKES
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KLINGONS DIFFERENT
WHY ARE THERE SQUIRRELS
WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL
WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY
WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD
WHY IS SEX SO IMPORTANT
WHY ARE THERE SLAVES IN THE BIBLE
WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT
WHY AREN'T MY ARMS GROWING
WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE
WHY ARE THERE ANTS IN MY LAPTOP
WHY IS EARTH TILTED
WHY IS SPACE BLACK
WHY IS OUTER SPACE SO COLD
WHY ARE THERE PYRAMIDS ON THE MOON
WHY IS NASA SHUTTING DOWN
WHY ARE THERE GHOSTS
WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS
WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE
WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA
WHY AREN'T THERE GUNS IN HARRY POTTER
WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG
WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY IS YKK ON ALL ZIPPERS
WHY DO IGUANAS DIE
WHY ARE THERE DINOSAUR GHOSTS
WHY IS THERE LAVA
WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND
WHY ARE THERE FEMALE MR MIMES
WHY IS GPS FREE
WHY IS THERE HELL IF GOD FORGIVES
WHY IS LIFE SO BORING