Welcome to STAT 547C, *Topics in Probability*

Instructor: Alexandre Bouchard

Fall 2018

Plan for today:

- Logistics.
- Why you should care about probability.
- The vocabulary of probability

Logistics

Contact & other logistic issues

- Web site: always check first!
  http://www.stat.ubc.ca/~bouchard/teaching.html
  - Textbook and other readings
  - Hints and updates for assignments
  - Office hours
  - Contact:
    1. Piazza
    2. bouchard@stat.ubc.ca

Homeworks (40%)

- Main assignments (~3-4, 25%)
- ‘Exercises/Participation’ (15%): doing the short exercises, scribing, interacting in class, coming at office hours

Exams

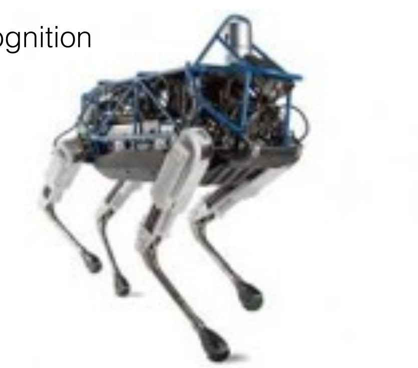
- In class midterm (20%)
- Finals:
  - Last day of class: 'Essay: what I have learned in the course' (10%)
  - Take home final/project (30%)

Why this topic is important

Automated theorem proving
Adaptive websites
Affective computing
Bioinformatics
Brain–machine interfaces
Cheminformatics
Classifying DNA sequences
Computational anatomy
Computer Networks
Computer vision, including object recognition
Detecting credit-card fraud
General game playing
Information retrieval

Machine perception
Medical diagnosis
Economics
Insurance
Natural language processing
Natural language understanding
Optimization and metaheuristic
Online advertising
Recommender systems
Robot locomotion
Search engines
Sentiment analysis (or opinion mining)
Sequence mining

Internet fraud detection
Linguistics
Marketing
Machine learning control
Software engineering
Speech and handwriting recognition
Financial market analysis
Structural health monitoring
Syntactic pattern recognition
Time series forecasting
User behavior analytics
Translation



Machine Learning

Probability

Optimization

Why this topic is important

- Fundamental tool in **statistics**, computer science, physics, econometrics, ... and increasingly, biology, linguistics, sociology, ...
  - Creating models
  - Inverting them (Bayesian statistics/conditioning)
  - Computational power of randomness
- Also a branch of pure math in its own right
- Replacing logic as the philosophical foundations of science and cognition?
  ('Dawning of the age of stochasticity', D. Mumford)

Probability in action: Diverse examples

Engineering, technology, logistics

The Search for Malaysia Airlines Flight 370

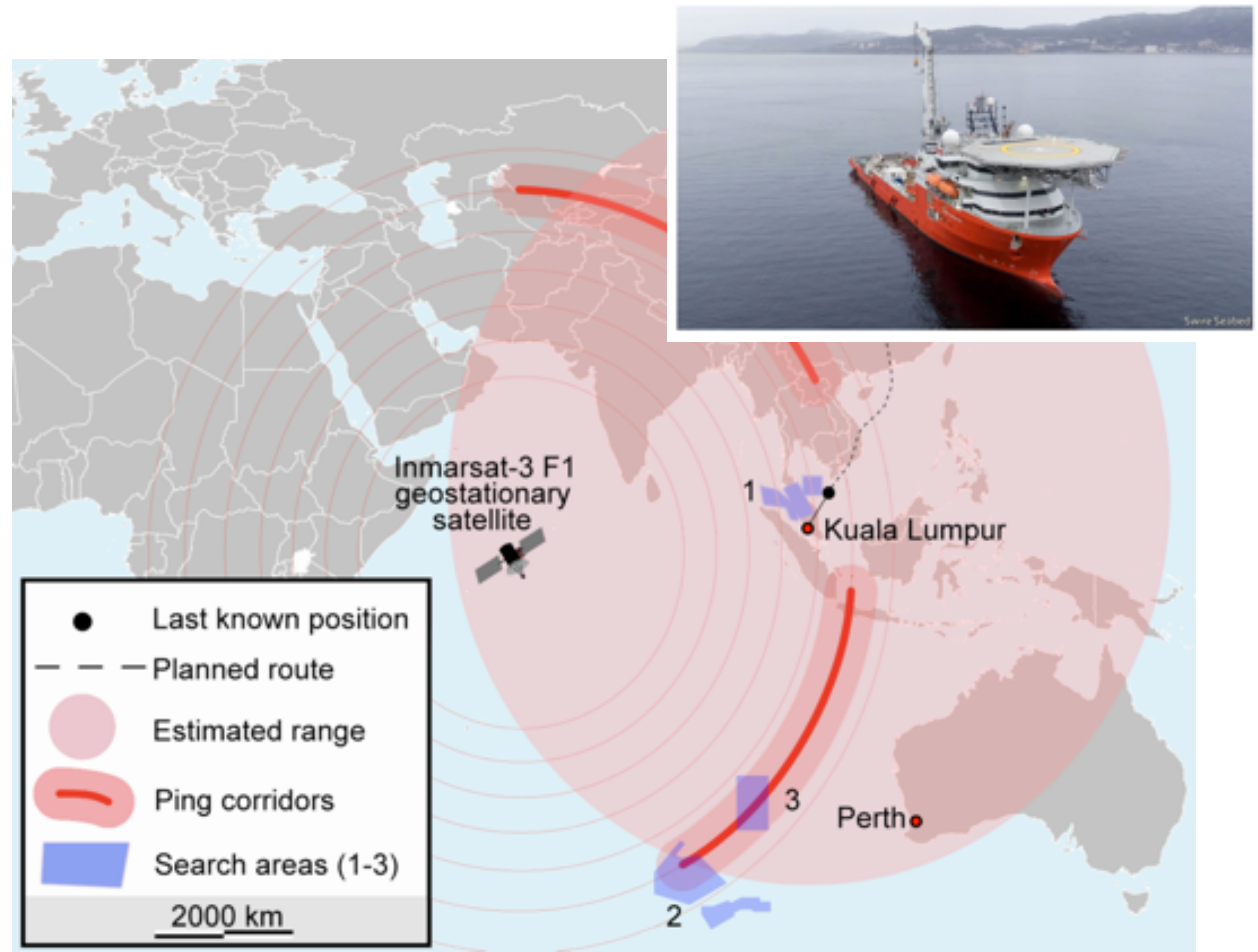
Ex. 1

The search for MH370
A fantastical ship has set out to seek Malaysian Airlines flight 370
A swarm of submarine drones will scour the depths for the plane

Goal: finding the location of the crash

Question: how to prioritize search

How to reconcile several sources of partial info:
- Last known position
- Fuel range
- Last satellite ping



http://tinyurl.com/lhzrufa

Ex. 3

Rational behavior and uncertainty

General question: how to **act** when

- we are facing uncertainty
- errors have different costs

Examples:

- fraud detection
- medical diagnosis
- spam classifiers

Key tool: *expected value*

Sciences

Ex. 4

Ecology: Estimating animal population sizes

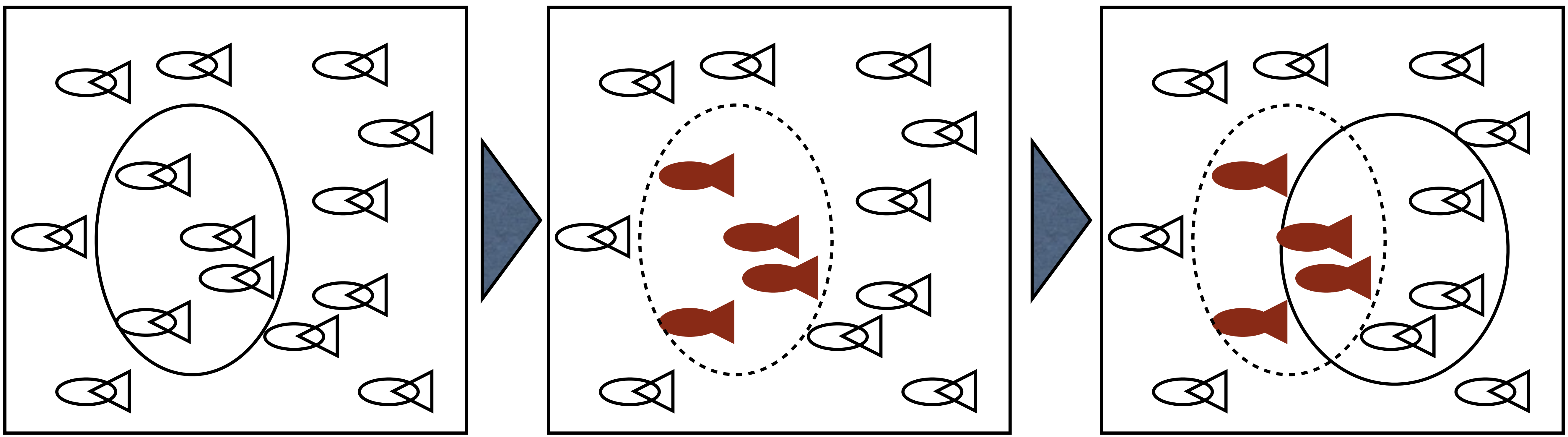
Example: finding the number of Sockeye salmon in the Pacific Ocean (!)

Very important problem for conservation, setting fishing quotas, etc.

Ex. 4

Insight: the capture-recapture trick

Population Capture and tag Recapture and count

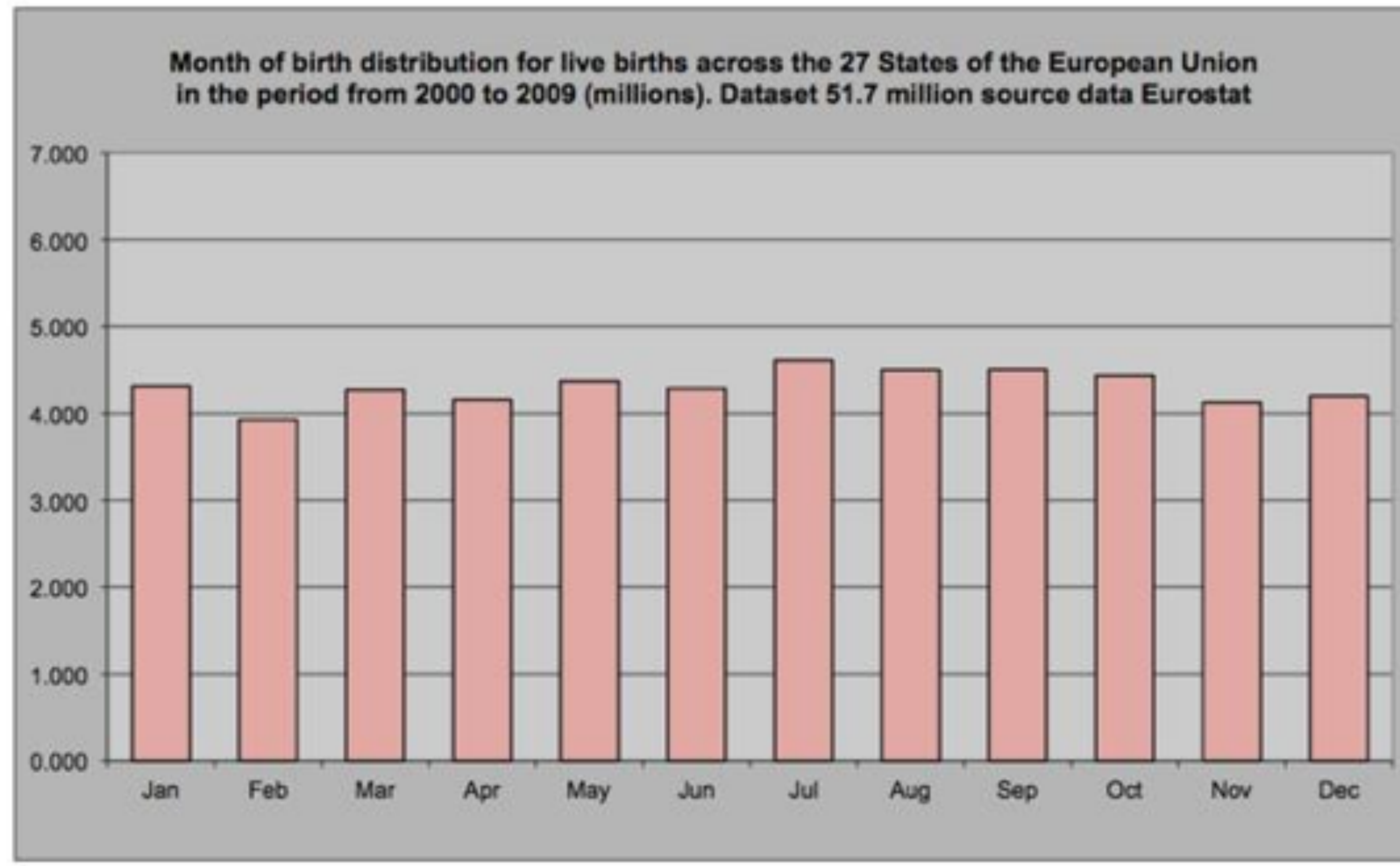


Examples:

Ex. 5

Assessing *significance*

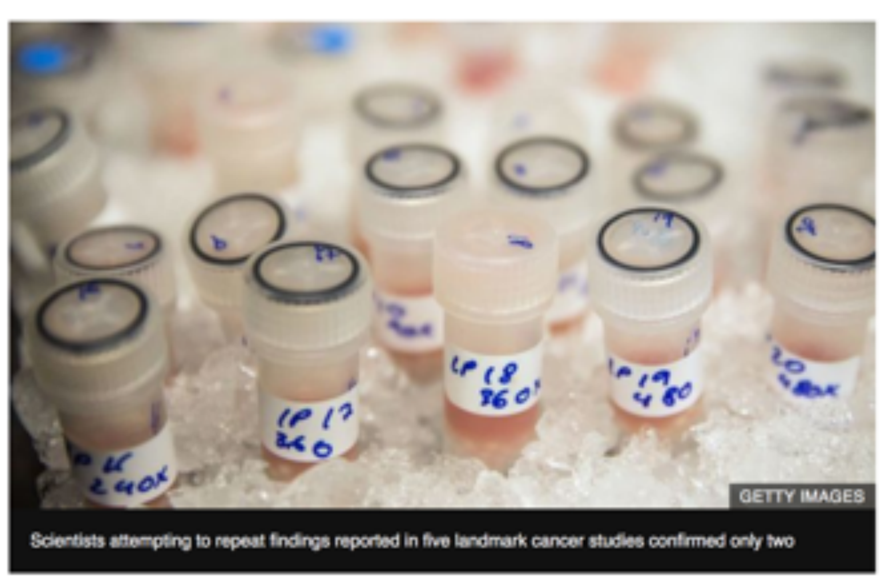
- *Histogram* of # of births organized by month:
- Question: is the # of births *uniform* across months?
- Note: even if the answer is yes, we would expect small differences across months.
    - How small?



Ex. 5

Assessing *significance*

- Tricky problem
- Mis-use of statistical tools has lead to the *reproducibility crisis*



Most scientists 'can't replicate studies by their peers'

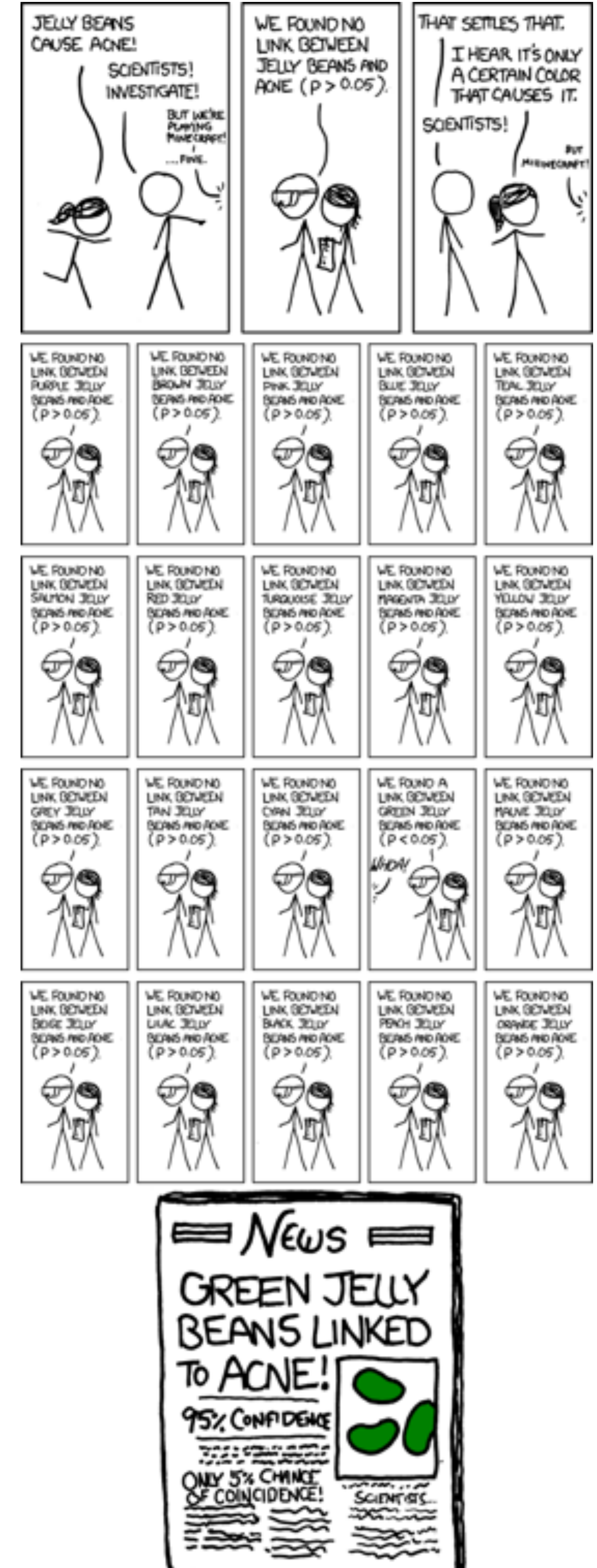
By Tom Feilden
Science correspondent, Today programme

22 February 2017

Share

GETTY IMAGES

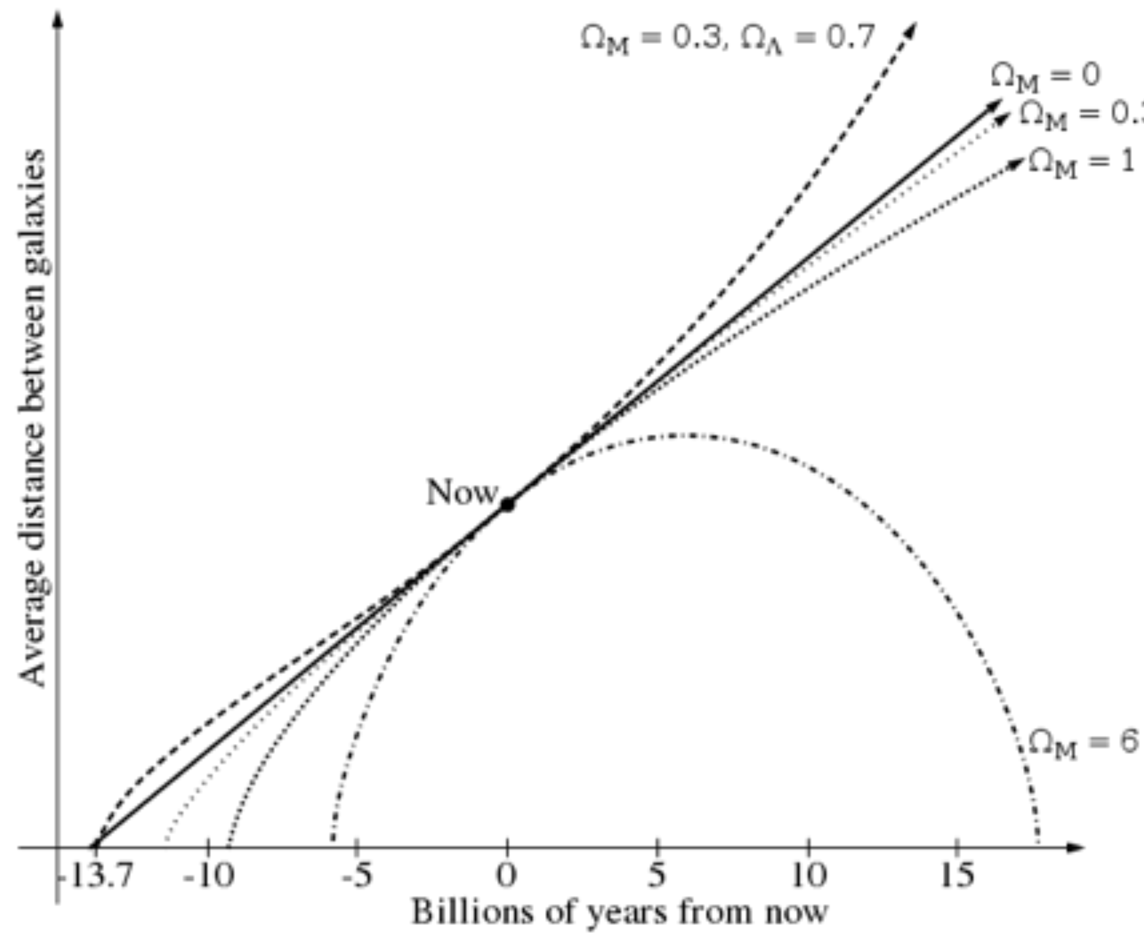
Scientists attempting to repeat findings reported in five landmark cancer studies confirmed only two

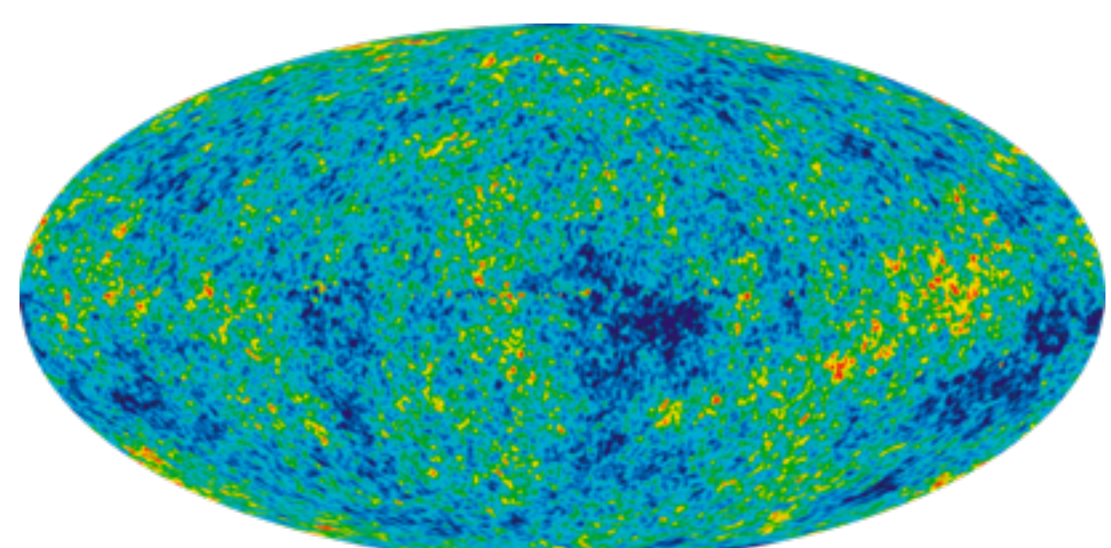


Ex. 7

Astrophysics: Estimating the age and fate of the Universe

- **Goals**: finding the Universe's
  - age
  - density (=> fate)
- **Data**: Cosmic Microwave Background (CMB): remnants of Big Bang
  - Detailed map from the Planck satellite
- Age, Physical constants => known *distribution* on CMP
- Invert using Bayes' rule

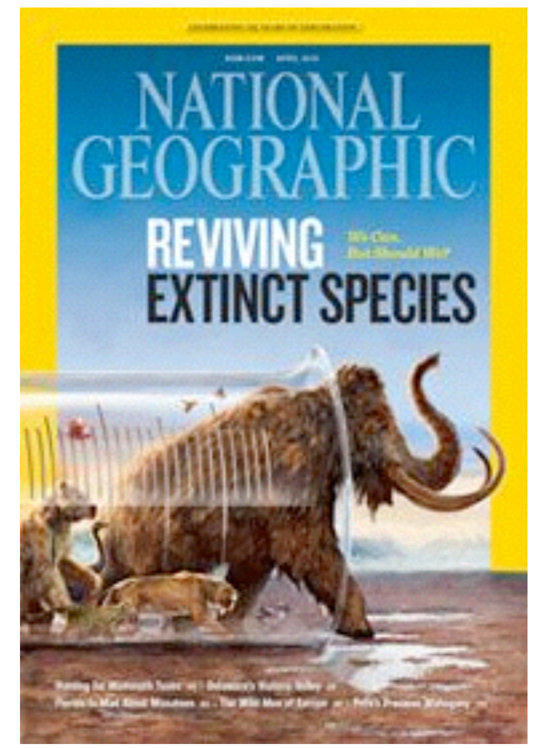




Ex. 7

Phylogenetics: Reconstruction of ancient species

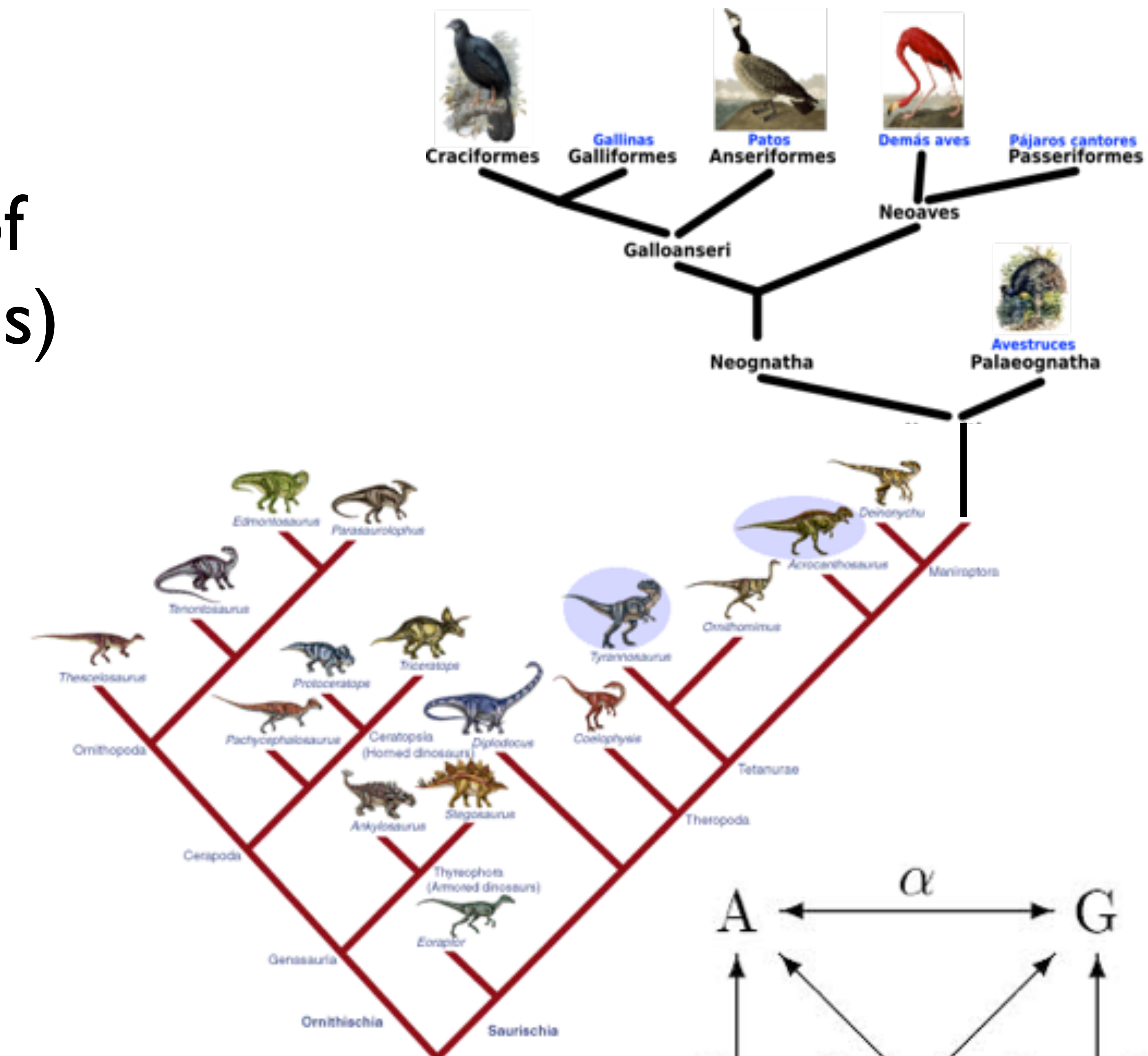
- Goals:
  - better understand ancient species
  - revive them?
- Fossil DNA degrades after few 1000s years
- Are dinosaurs' genomes completely lost?

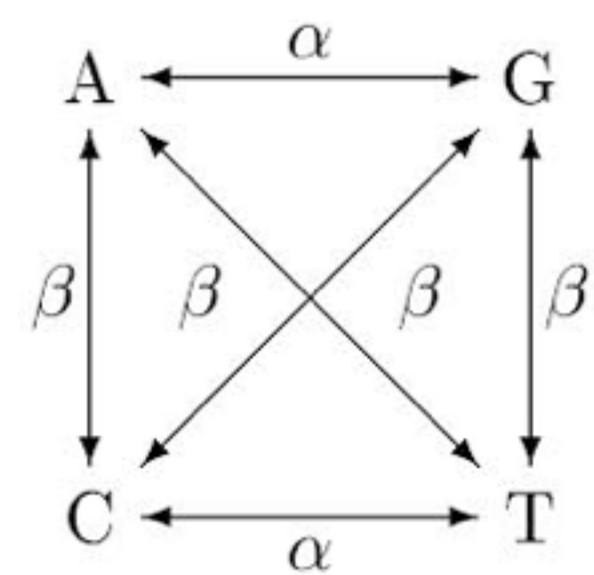


Ex. 7

Phylogenetic tree

- **Idea:** use the genomes from the descendants of dinosaurs (modern birds)
- We know how DNA change over time (probabilistically)
- Marginalization of unknown genomes
- Additional challenge: structure of tree is unknown





Core topics

- Applications of probability in statistics
- Formal treatment of the probability spaces and expectation and their properties (the language of probability)
- The ‘surprising challenges of composing r.v.s’
  - Asymptotics
  - Generating functions
- Conditioning
- Going beyond independence.e.g Markov chains

Additional topics from:

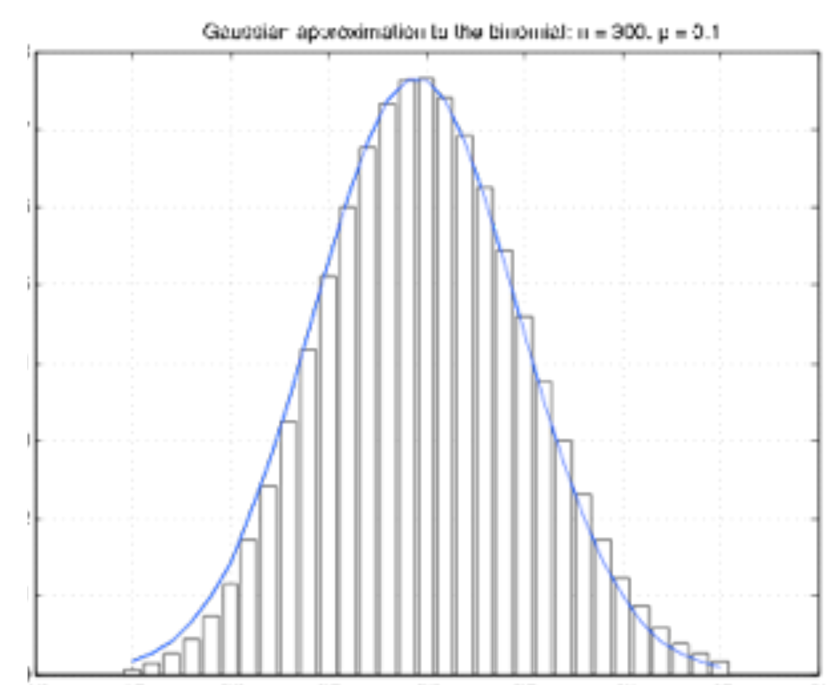
- Selected stochastic processes useful in the stats literatures (e.g. Continuous time Markov chains / piecewise deterministic Markov processes / Poisson processes ...)
- Martingales
- Bandits
- Optimal transports

First exciting landmark (coming up this month): *A Surprising Challenge*

- Sums of random variables
  - Omnipresent in statistics
  - Taking the sum of variables is easy, so taking the sum of *random* variables should also be easy, right?
  - Not quite... consider for example the problem of computing the probability that the sum of 1000 coins is greater that 500.
    - Would have been hard in the pre-computer era
    - Generalized versions of this problem still hard with computer

Limiting theory to the rescue

- Another surprise: sums of random variables can be approximated by something simple when large number of terms involved
- No matter what each X is!!! (almost)
- Also explains why we spend disproportionate amount of time on some specific types of random variables (normal, Poisson, ...)
    - In some sense, ultimate motivation for continuous random variables, as they often arise as limiting objects



300 coins

Building *models*

Two coins

- Let us ask Probability Theory:
  Flip 2 coins. What is the probability that the 2 coins both show heads?

A. 1/2

B. 1/3

C. 1/4

D. I cannot tell you

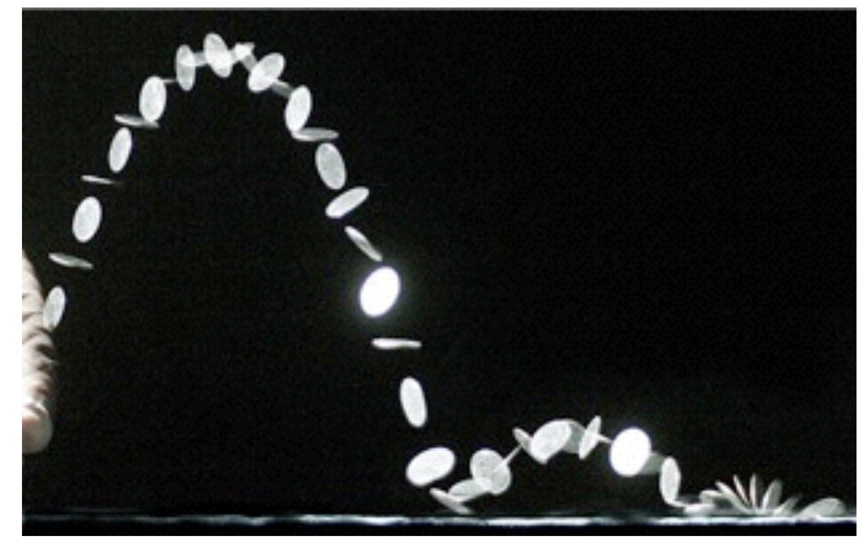
Probability that 2 coins show heads

- 1/2? Either they do, or they don't.
- 1/3? Either both heads, both tails, or heads-tails.
- 1/4? Imagine one coin is painted red, one is painted blue. There are then 4 possibilities:
  - red: heads, blue: heads
  - red: heads, blue: tails
  - red: tails, blue: heads
  - red: tails, blue: tails

- These correspond to 3 different models
- None is 'true'
- But the third is more useful (accurate at doing predictions)

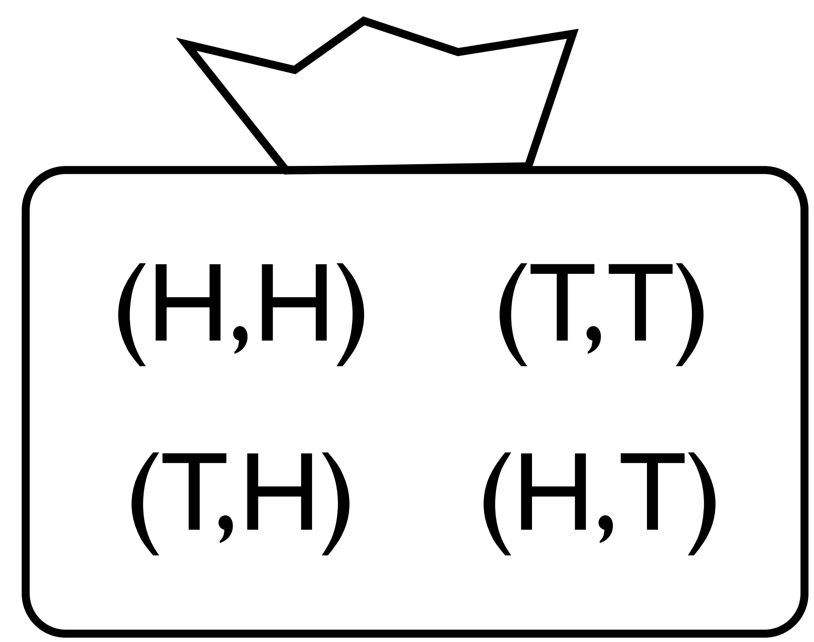
Models are approximations

- Reality: a complex dynamical system



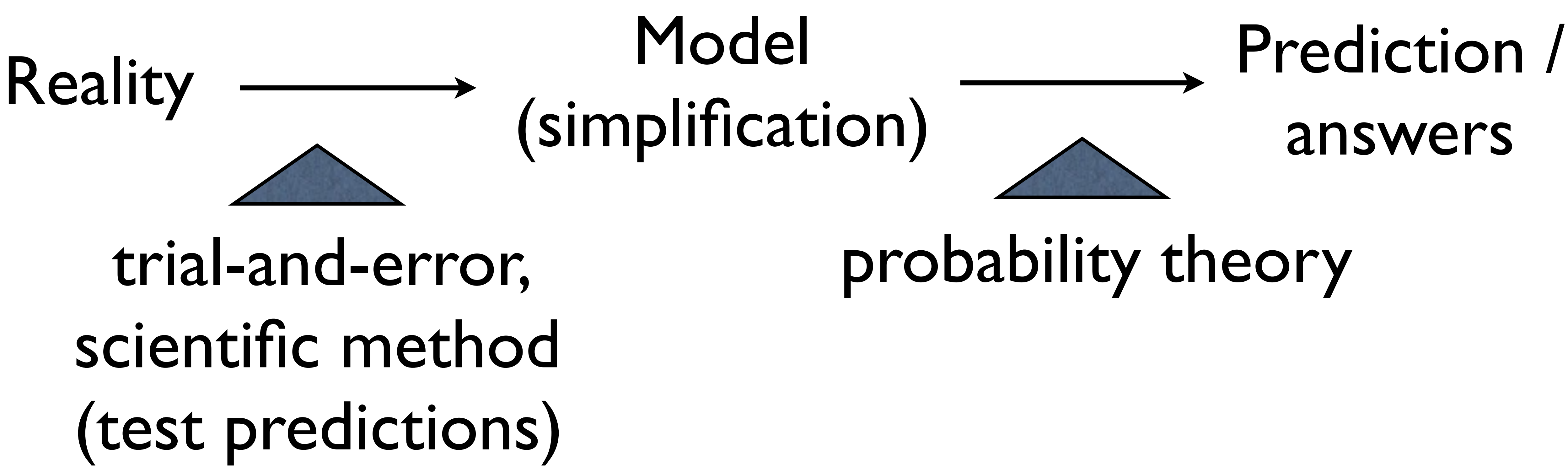
Stroboscopic image of a coin flip by Andrew Davidhazy

- Model: a 'bag' with 4 'objects' in it



Which model to use?

Probability theory alone does cannot answer this question.

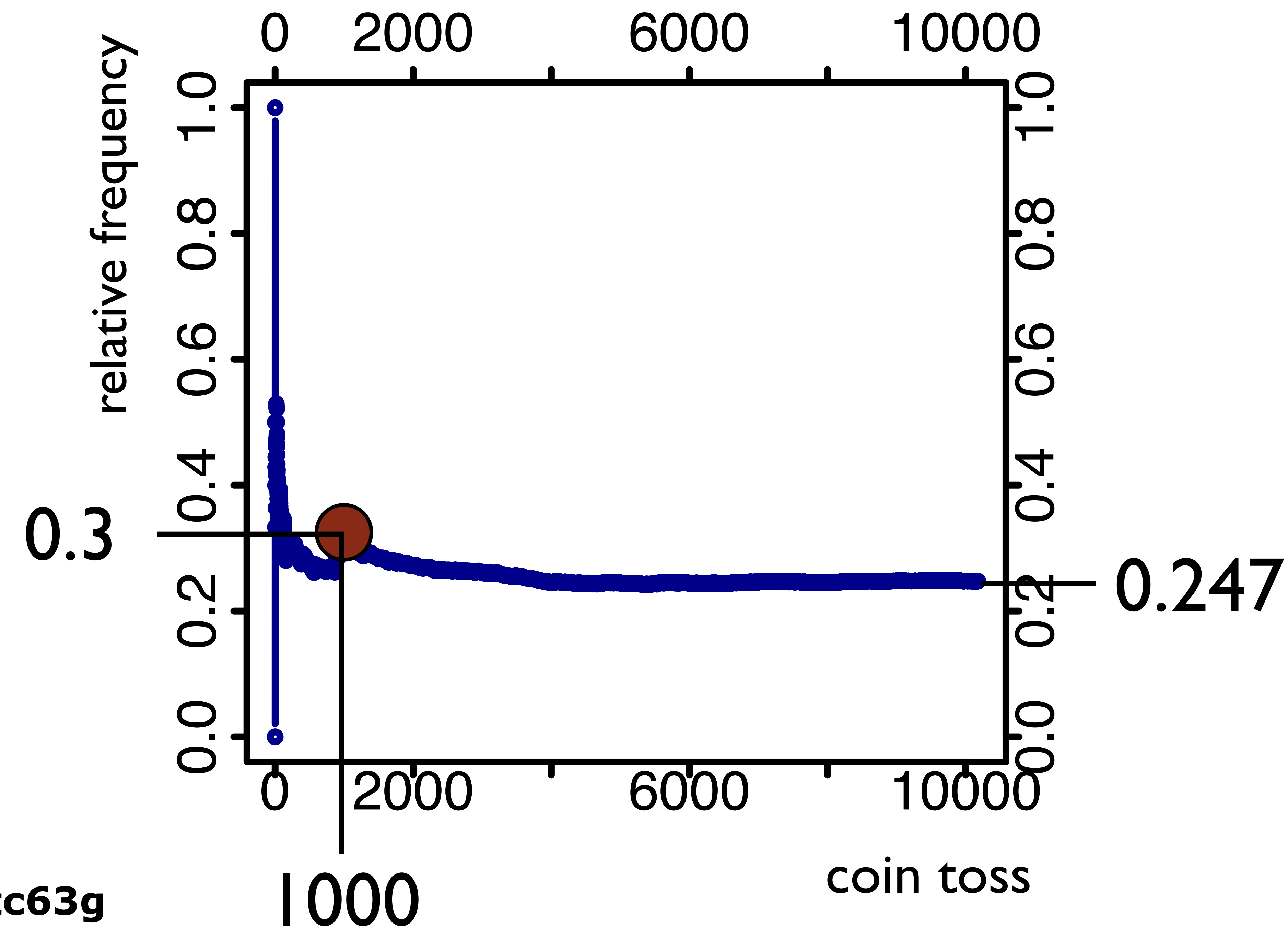


But...

- Probability theory still useful:
    - Given a model, it makes certain predictions
    - We can then test those predictions
- Example: *law of large numbers*
    - Relates *probability* and *frequency* in repeated experiments

Reality check

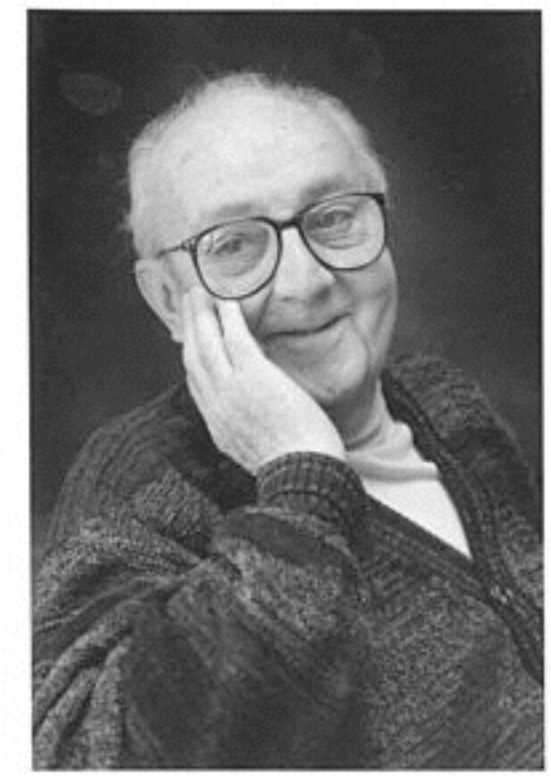
- Dataset of 10,000 actual tosses of two coins available on the web (!)



http://tinyurl.com/ljtc63g

All models are wrong

- Given the large number of throws, 0.247 may seem a bit far from 0.25 (we will formalize this idea later)
  - is there something fishy?
  - after reading the fine prints of the data, realized all throws started with same face in hand of thrower
  - see P. Diaconis' paper, http://tinyurl.com/yked5fk
- Essentially, all models are wrong, but some are useful. -- G. Box



Probability models
Measure-theoretic vocabulary

Why bother with measure theory

- Many concepts more natural to define and easier to prove
  - independence
  - exchange of integral / limits / derivatives
- Unified treatments of discrete/continuous, univariate/multivariate
- Read the literature fearlessly

The 3 ingredients (axioms of probability)

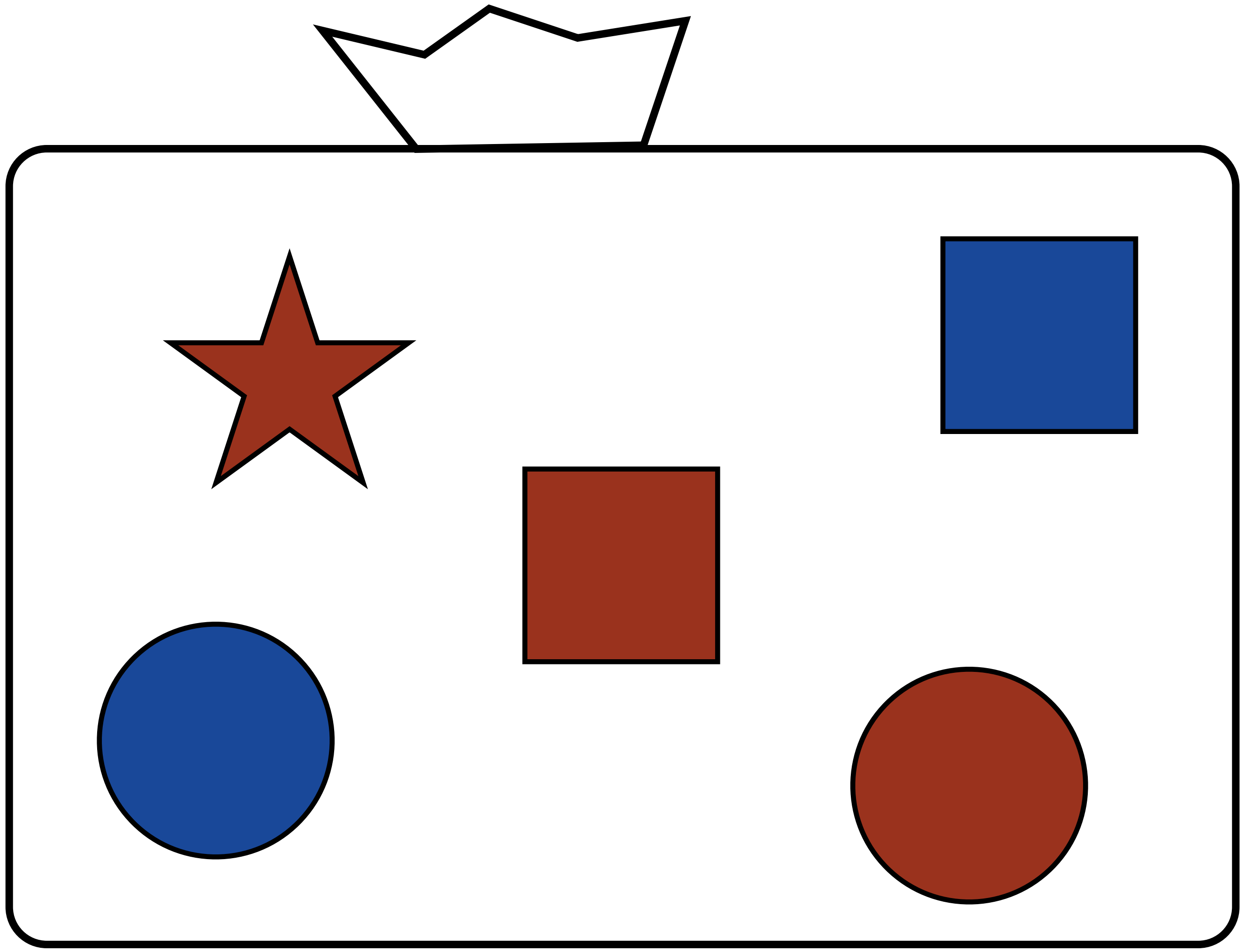
1. a set Ω, called the sample space,
2. a closed collection of events, $\mathcal{F} \subset 2^{\Omega}$, called a σ-algebra,
3. a probability measure (synonym: probability distribution), $\mathbb{P} : \mathcal{F} \to [0, 1]$.

where:

1. the σ-algebra satisfies:
   (a) $\Omega \in \mathcal{F}$,
   (b) $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots, \Longrightarrow \cap_{i=1}^{\infty} A_i \in \mathcal{F}$,
   (c) $A \in \mathcal{F} \Longrightarrow A^{\complement} \in \mathcal{F}$.
2. and the probability measure satistifies:
   (a) $\mathbb{P}(\Omega) = 1$,
   (b) if $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots$ are disjoint ($i \neq j \Longrightarrow A_i \cap A_j = \emptyset$), then

$$\mathbb{P}(\cup A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Example: bag of distinct objects of same size



- Proportion of red shapes?
- Probability of drawing a red shape?
- *Outcome:* an individual object in the bag
  ex.: s = 

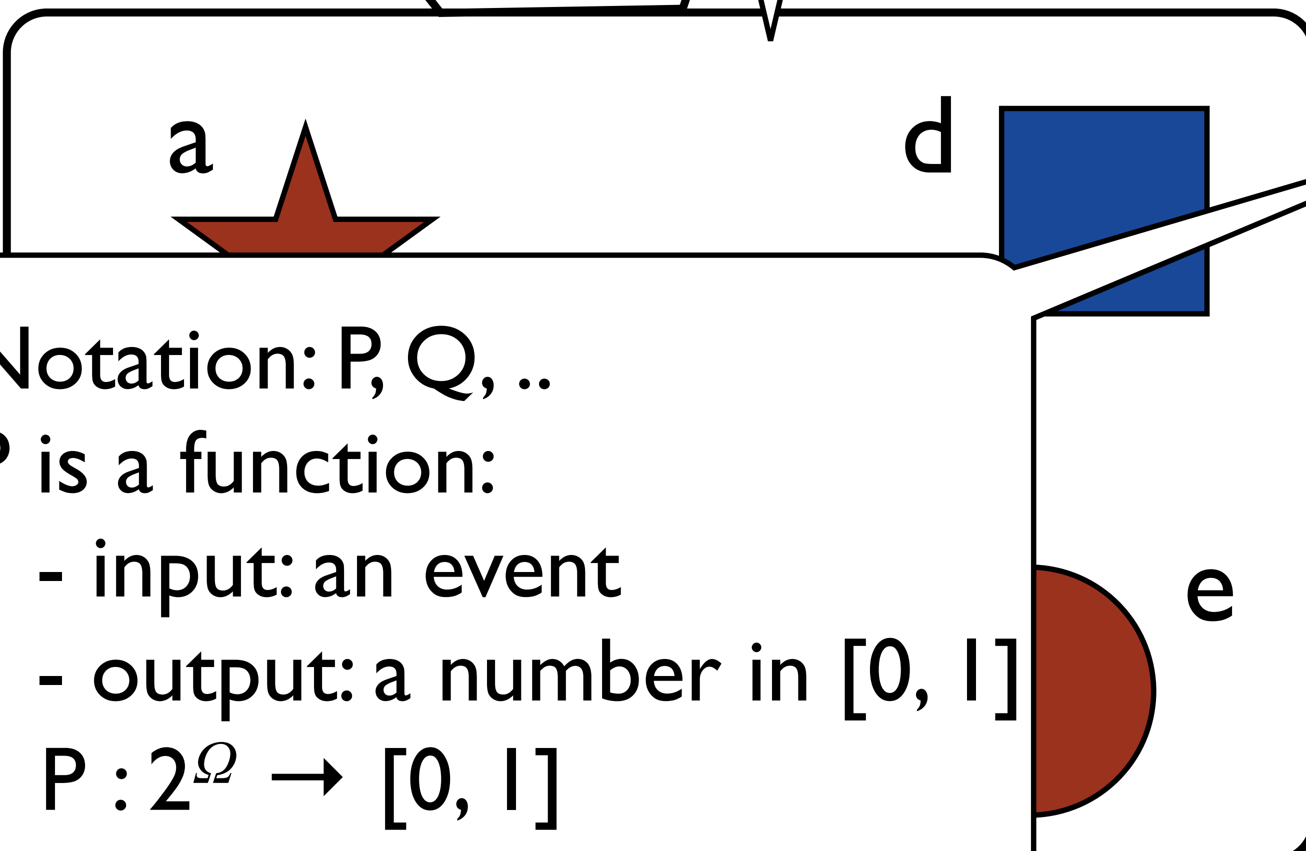
Example of a probability <u>where outcomes are equally likely</u>:

$$\frac{\text{\# of outcomes of interest}}{\text{\# of outcomes}}$$

Notation

Sample space
Typical notation: Ω

This is a *set of outcomes*
Nickname: *event*
Typical notation: red, *E*, blue, *F*, ..
E = {a, c, e}



Notation: P, Q, ..
P is a function:
- input: an event
- output: a number in [0, 1]

$P : 2^{\Omega} \rightarrow [0, 1]$

Example: P(*E*) = 3/5

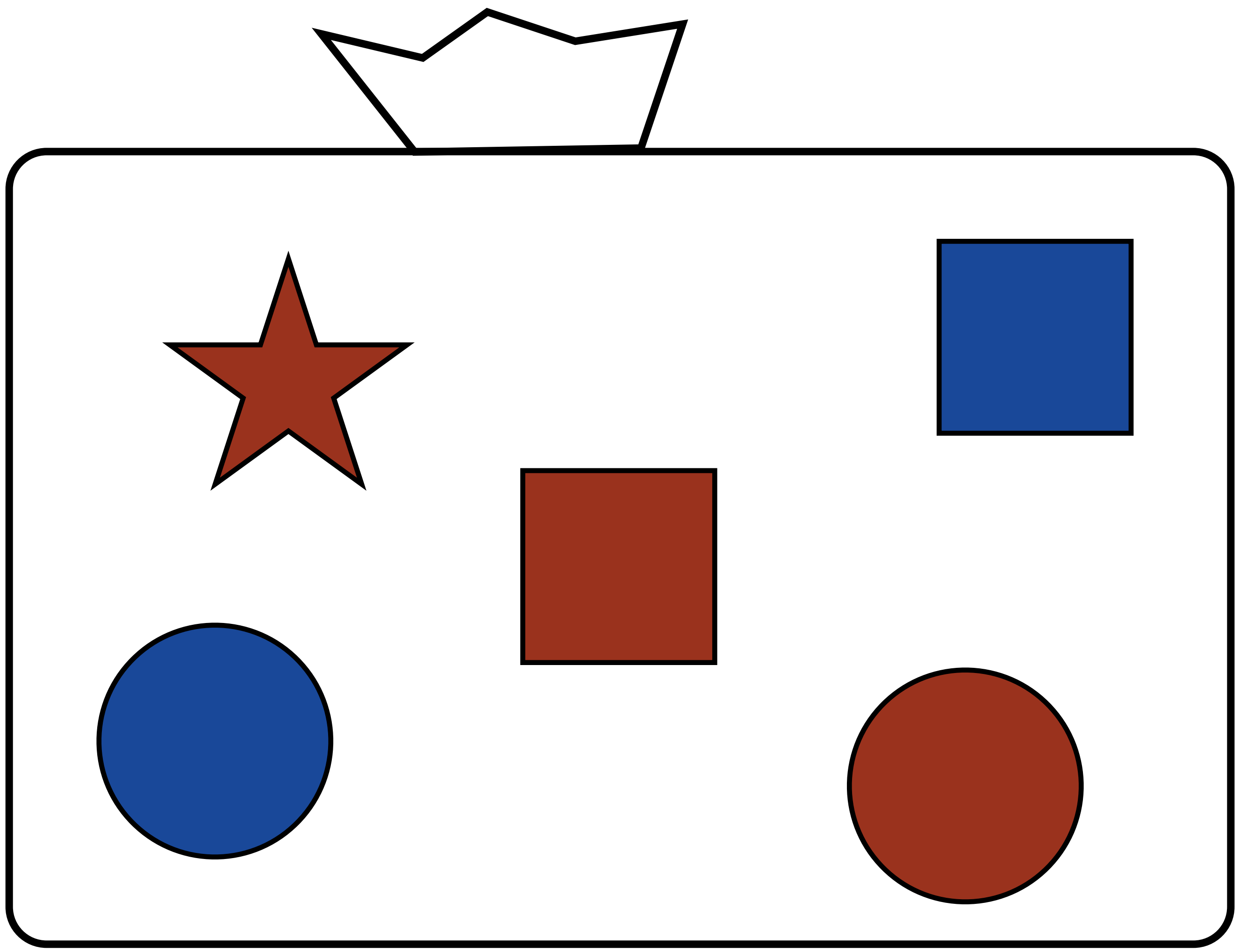
- Proportion of red shapes?
- Probability of drawing a red shape?
- *Outcome:* an individual object in the bag

$P(E) = |E| \,/\, |S|$

Example of a probability <u>where outcomes are equally likely</u>:

$$\frac{\text{\# of outcomes of interest}}{\text{\# of outcomes}}$$

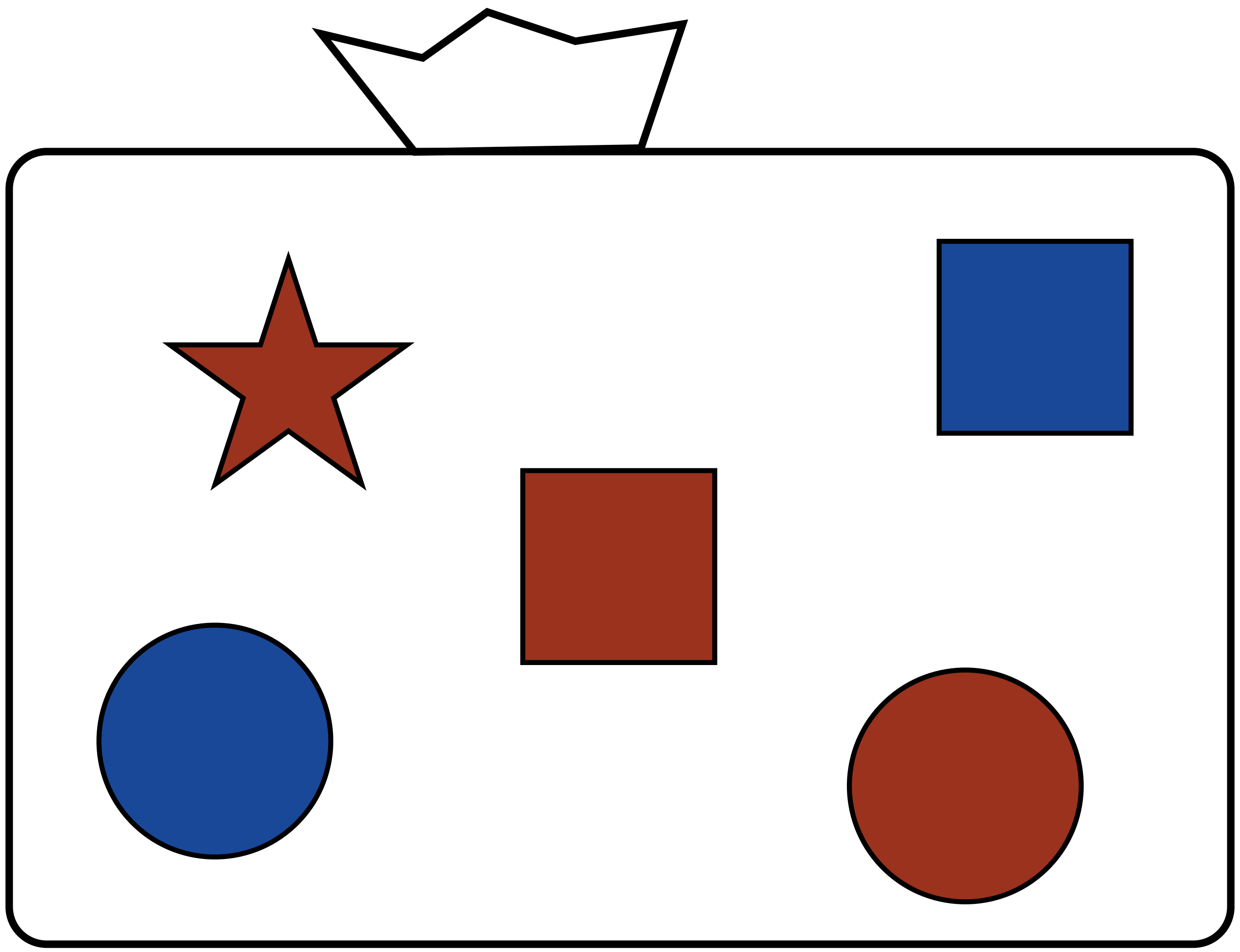
Basic properties



$$P(E^c) = 1 - P(E)$$

- We know P(*E*)
  - Example: P(square) = 2/5
- What is P(E^c) ?
  - Example: P(not square)
- E^c means:
  - the *complement* of *E*
  - the outcomes not in *E*
  - = $S \setminus E$ (minus, for sets)

Basic properties



- We know P(*E*), P(*F*)
  - Example: P(blue) = 2/5
    P(star) = 1/5
- What is P(*E* ∪ *F*) ?
  - Example: P(blue or star)
- Try now on:
  - Example: P(red) = 3/5
    P(square) = 2/5

$$P(E \cup F) = P(E) + P(F)$$ if *E* and *F* are *disjoint*, i.e. $E \cap F = \varnothing$

Axioms of probability

These are called the **axioms of probability** ***

Assume:

a) $0 \leq P(E) \leq 1$

b) $P(\Omega) = 1$

c) $P(E \cup F \cup \ldots) = P(E) + P(F) + \ldots$ if $E, F, \ldots$ are all disjoint

Discrete, equally weighted

Definition (1):

$$P(E) = |E| \,/\, |S|$$

Properties (2):

a) $0 \leq P(E) \leq 1$

b) $P(\Omega) = 1$

c) $P(E \cup F) = P(E) + P(F)$ if E and F are disjoint

Not equally weighted (or not discrete)

More interesting examples: random walk; statistical models

Some *type checking*

1. a set Ω, called the sample space,
2. a closed collection of events, $\mathcal{F} \subset 2^{\Omega}$, called a σ-algebra,
3. a probability measure (synonym: probability distribution), $\boxed{\mathbb{P} : \mathcal{F} \to [0, 1].}$

where:

1. the σ-algebra satisfies:
   (a) $\Omega \in \mathcal{F}$,
   (b) $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots, \Longrightarrow \cap_{i=1}^{\infty} A_i \in \mathcal{F}$,
   (c) $A \in \mathcal{F} \Longrightarrow A^{\complement} \in \mathcal{F}$.
2. and the probability measure satistifies:
   (a) $\mathbb{P}(\Omega) = 1$,
   (b) if $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots$ are disjoint ($i \neq j \Longrightarrow A_i \cap A_j = \emptyset$), then

$$\boxed{\mathbb{P}(\cup A_i)} = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Axioms for measure spaces

1. a set Ω, called the sample space,
2. a closed collection of events, $\mathcal{F} \subset 2^{\Omega}$, called a σ-algebra,
3. a probability measure (synonym: probability distribution), $\mathbb{P} : \mathcal{F} \to [0, 1]$.

Change upper bound to ∞

where:

1. the σ-algebra satisfies:
   (a) $\Omega \in \mathcal{F}$,
   (b) $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots, \Longrightarrow \cap_{i=1}^{\infty} A_i \in \mathcal{F}$,
   (c) $A \in \mathcal{F} \Longrightarrow A^{\complement} \in \mathcal{F}$.
2. and the probability measure satistifies:
   (a) $\mathbb{P}(\Omega) = 1$, Change boundary condition to $\mathbb{P}(\emptyset) = 0$
   (b) if $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots$ are disjoint ($i \neq j \Longrightarrow A_i \cap A_j = \emptyset$), then

$$\mathbb{P}(\cup A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Next: how this set of axioms is exquisitely fine tuned

1. a set Ω, called the sample space,
2. a closed collection of events, $\mathcal{F} \subset 2^{\Omega}$, called a σ-algebra,
3. a probability measure (synonym: probability distribution), $\mathbb{P} : \mathcal{F} \to [0, 1]$.

where:

1. the σ-algebra satisfies:
   (a) $\Omega \in \mathcal{F}$,
   (b) $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots, \Longrightarrow \cap_{i=1}^{\infty} A_i \in \mathcal{F}$,
   (c) $A \in \mathcal{F} \Longrightarrow A^{\complement} \in \mathcal{F}$.
2. and the probability measure satistifies:
   (a) $\mathbb{P}(\Omega) = 1$,
   (b) if $A_1 \in \mathcal{F}, A_2 \in \mathcal{F}, A_3 \in \mathcal{F}, \ldots$ are disjoint ($i \neq j \Longrightarrow A_i \cap A_j = \emptyset$), then
   $$\mathbb{P}(\cup A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$