Welcome to STAT 302, *Intro to Probability*

Instructor: Alexandre Bouchard
Fall 2014

Plan for today:

- Logistics.
- Why you should care about probability.

To get this information & more

- Main website :
  http://www.stat.ubc.ca/~bouchard/courses/stat302-fa2014-15/
  - Piazza (contact link on webpage)
  - Click on 'Files' to get lecture slides, assignments, etc.

Administrative details

- Prerequisites: Math 200 or 226 (which may be taken concurrently)
- Exclusions: Stat 241/251, Math 302
- Textbook: *A first course in probability* Sheldon Ross
  - 9th edition recommended
  - Suggested, optional exercises posted on website

Assessment

- Assignments (4), 20%
  - We encourage you to discuss your work with other students...
  - However, you must write up your own solutions independently.
  - No extension possible!

Assessment

- Midterm, 25%
  - Date: Wed Oct 22 (check this week that this date is not a problem)
  - No make-up exam!

Assessment

- Webwork/clickers, 5%
  - Get your clicker this week! (I will start using clicker questions by Monday Sep 8)
  - More on this soon
- Final, 50%
  - After lectures are over
  - Date announced centrally by University
  - You must pass the final to pass the course

To get help

- Great TAs:

Sean Jewell	Vincent Huang
sean.jewell@stat.ubc.ca	finial79@hotmail.com

- My email: bouchard@stat.ubc.ca
- But remember: use piazza (unless it is a personal matter)
- Office hours: TBA, will create a Doodle on main website shortly

What STAT 302 is about:

- *Probability spaces*: arguably the best quantitative tool to model reality
- Properties of probability spaces
- Lots of examples

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

Ex. 1

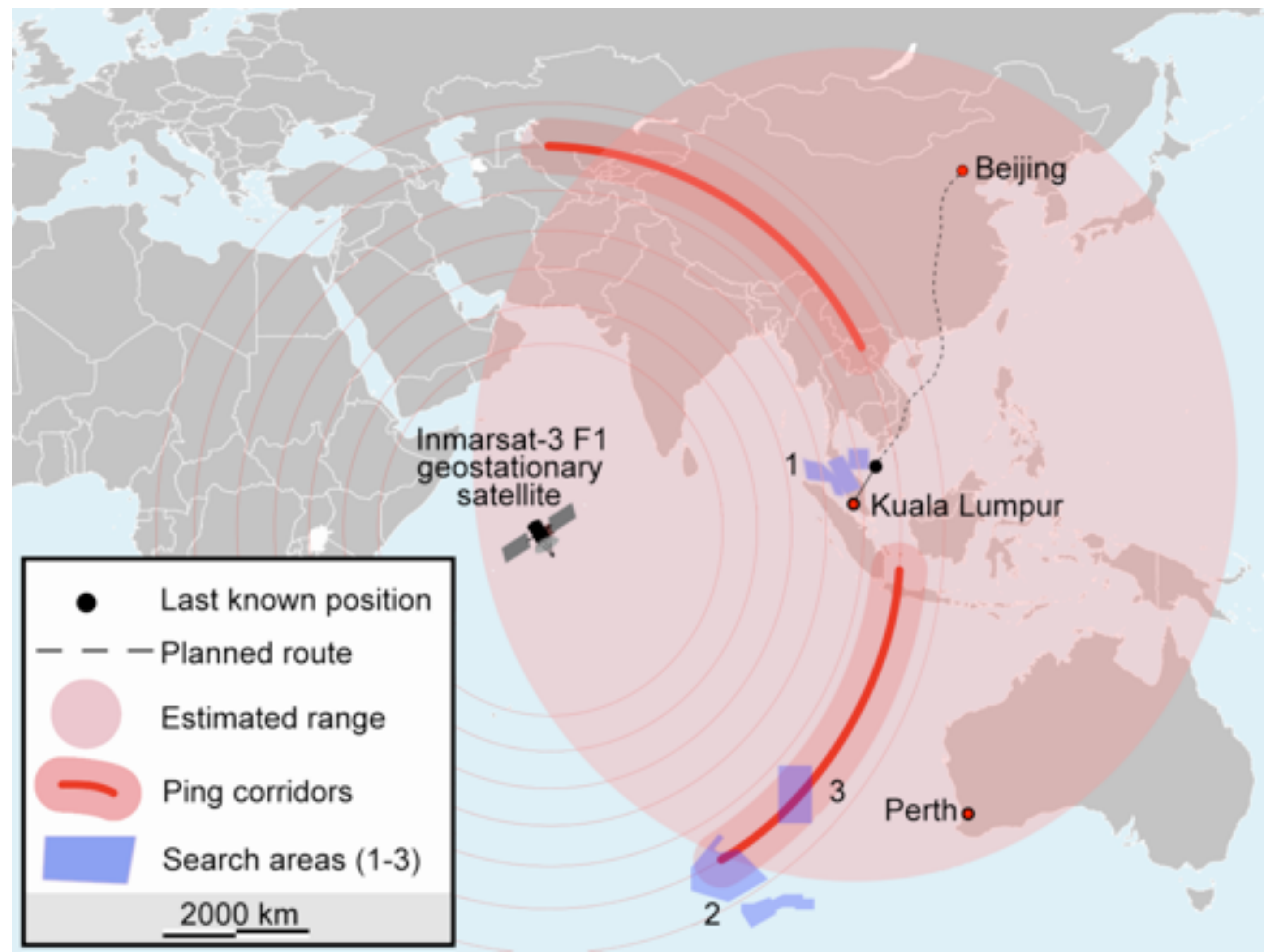
The Search for Malaysia Airlines Flight 370

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



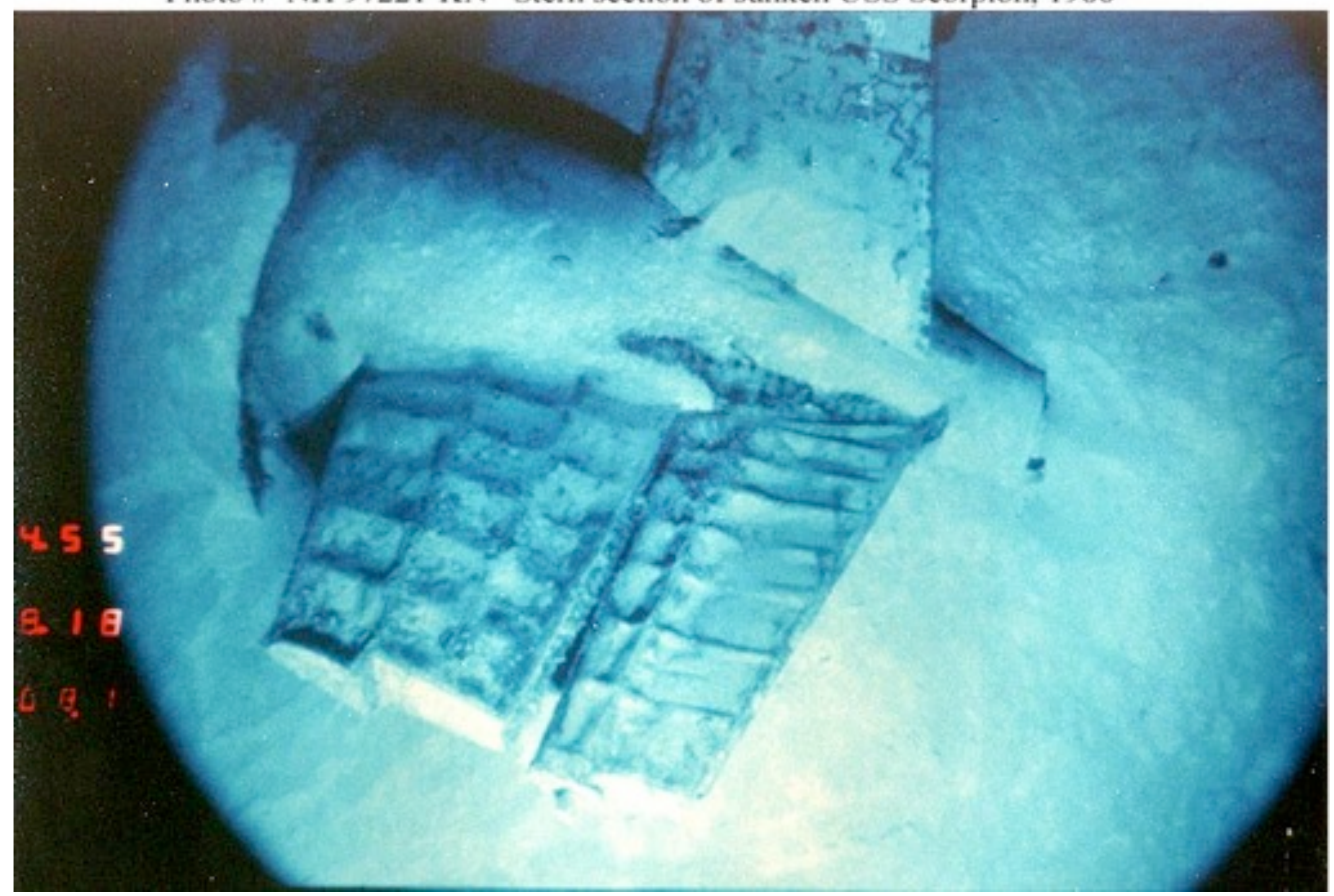
http://tinyurl.com/lhzrufa

Ex. 1

Bayesian Search

1966: Palomares B-52 crash

Photo # NH 97221-KN Stern section of sunken USS Scorpion, 1986



1968: USS Scorpion disappearance

Ex. 1

Conditioning

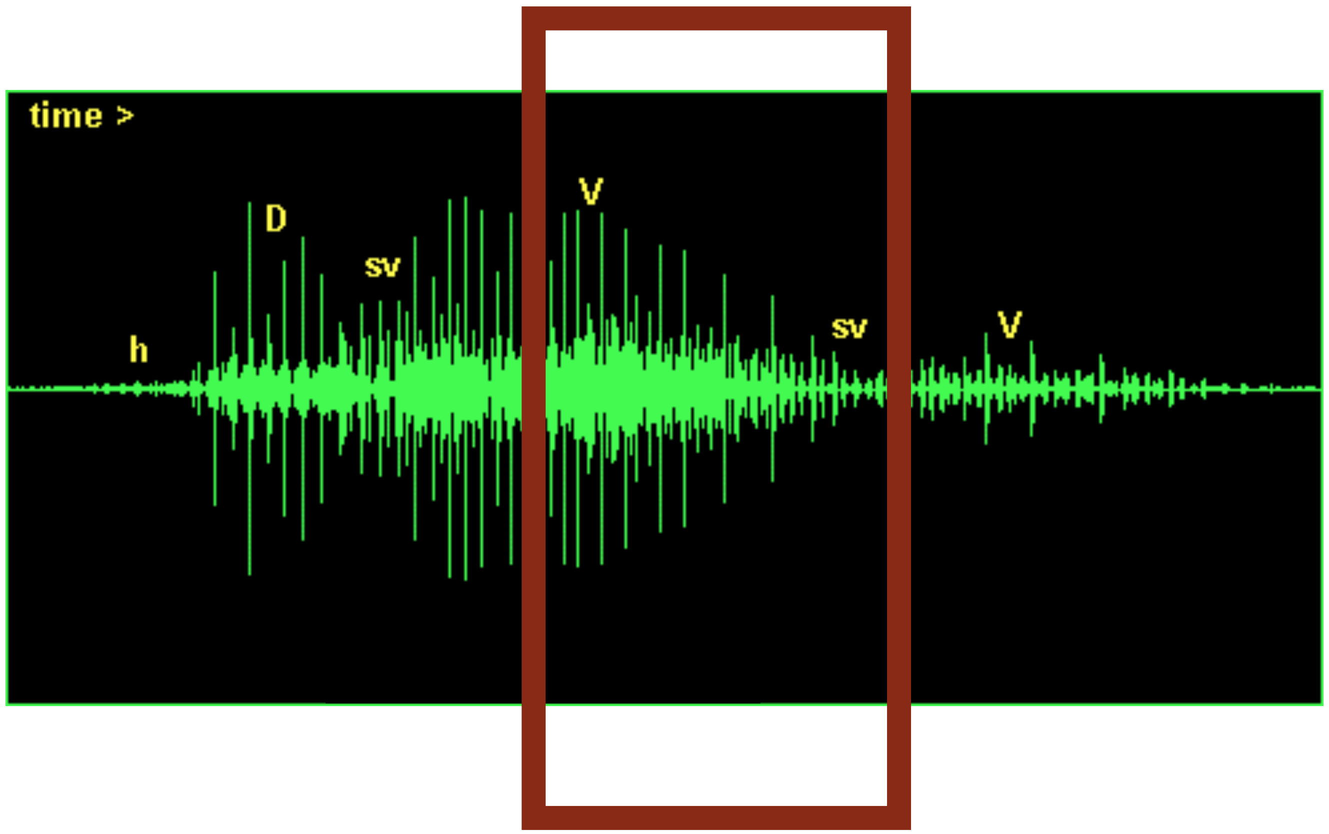
- Say you search in the square of highest success probability
  - You find nothing
  - What should you do next?
    - Note: even if the submarine is there, you might have missed it!
- Probability as a calculus of belief and uncertainty

Bayes theorem (Thomas Bayes), 1763

http://tinyurl.com/pcznhml

Ex. 2

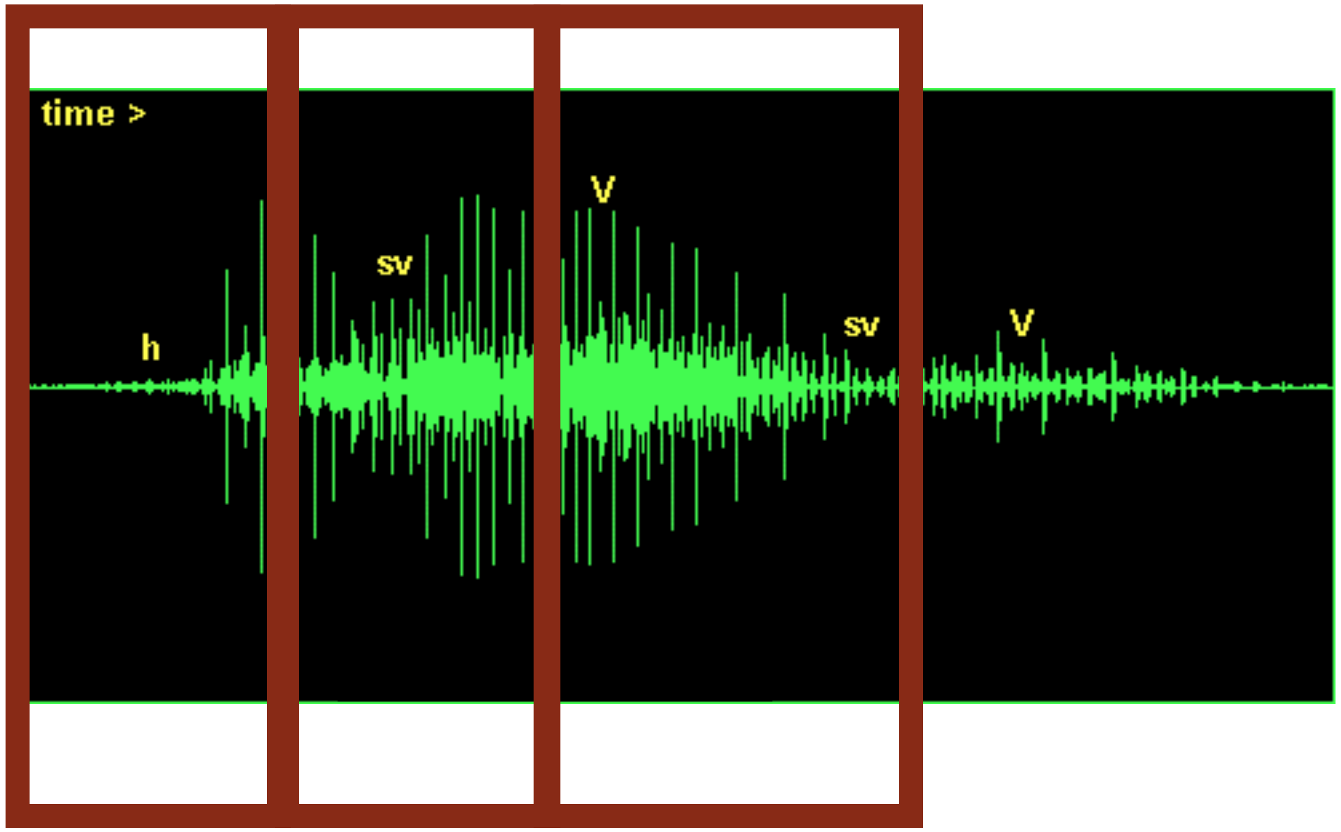
AI and machine learning: Speech recognition



???

Ex. 2

AI and machine learning: Speech recognition



How are ???

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

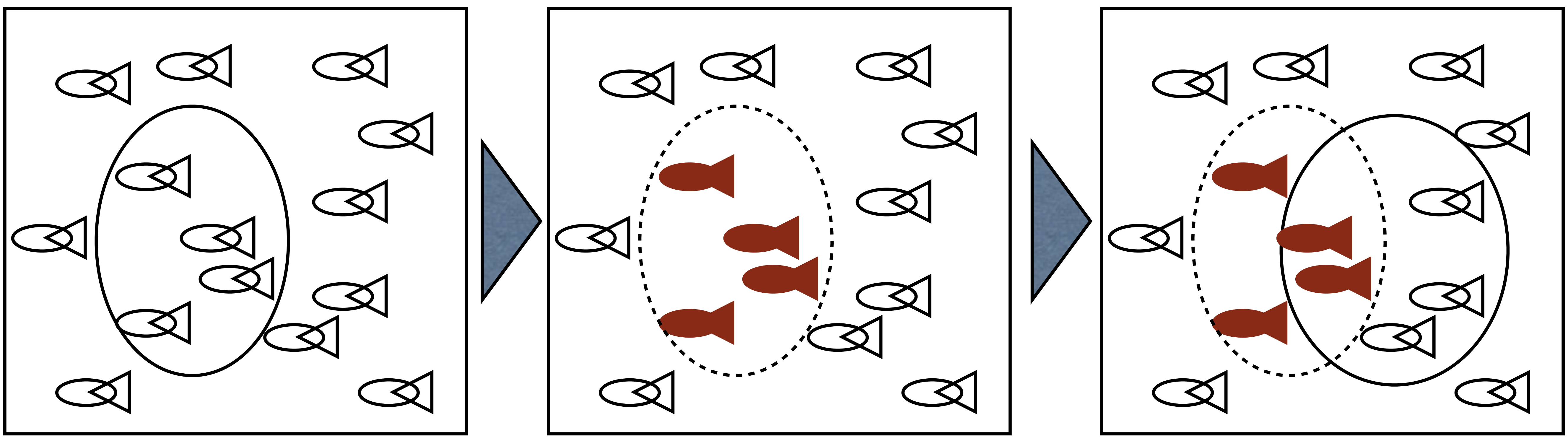
Example: finding the number of Sockeye salmon in the Pacific Ocean (!)

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

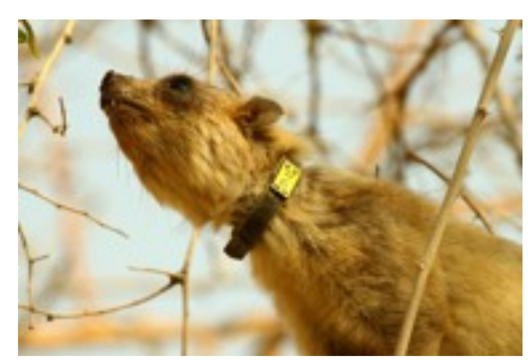
Ex. 4

Insight: the capture-recapture trick

Population | Capture and tag | Recapture and count



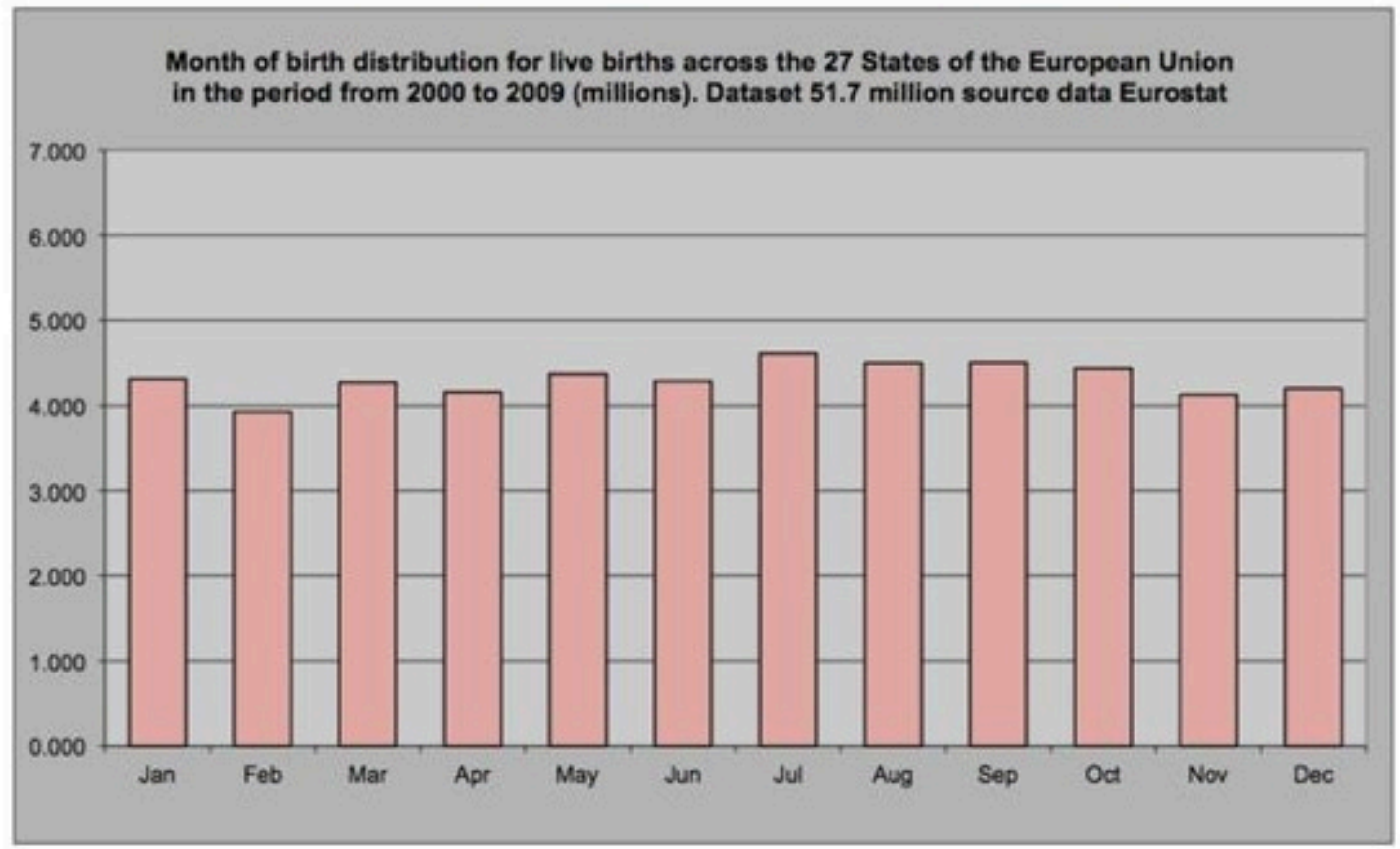
Examples:



Ex. 5

Assessing *significance*

- *Histogram* of births organized by month:

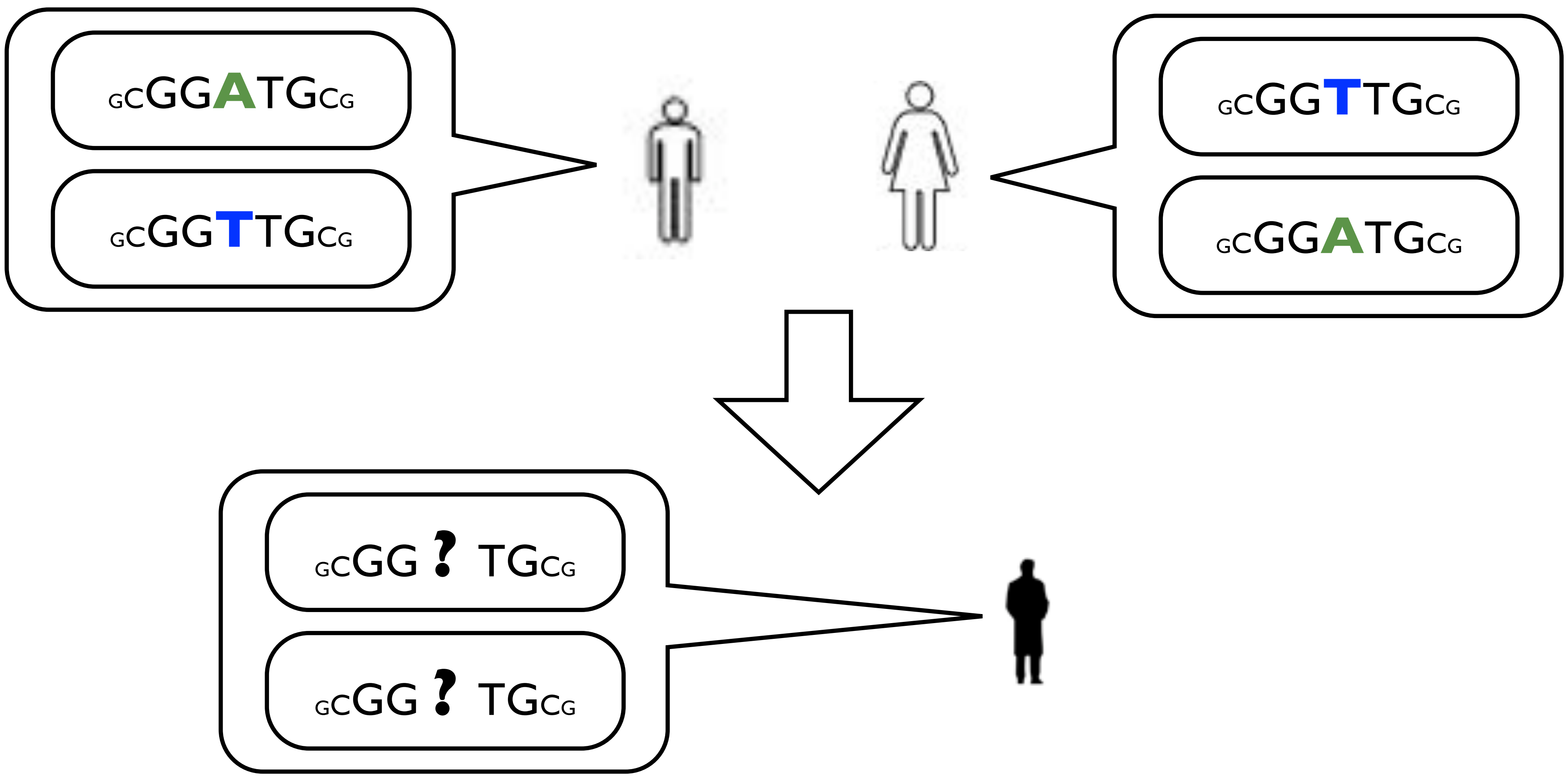


- Question: in *general*, are there months where there are more births than others? (are births *uniform* across months?)
- Note: even if the answer is no, we would expect small differences across months.
  - How small?

Ex. 6

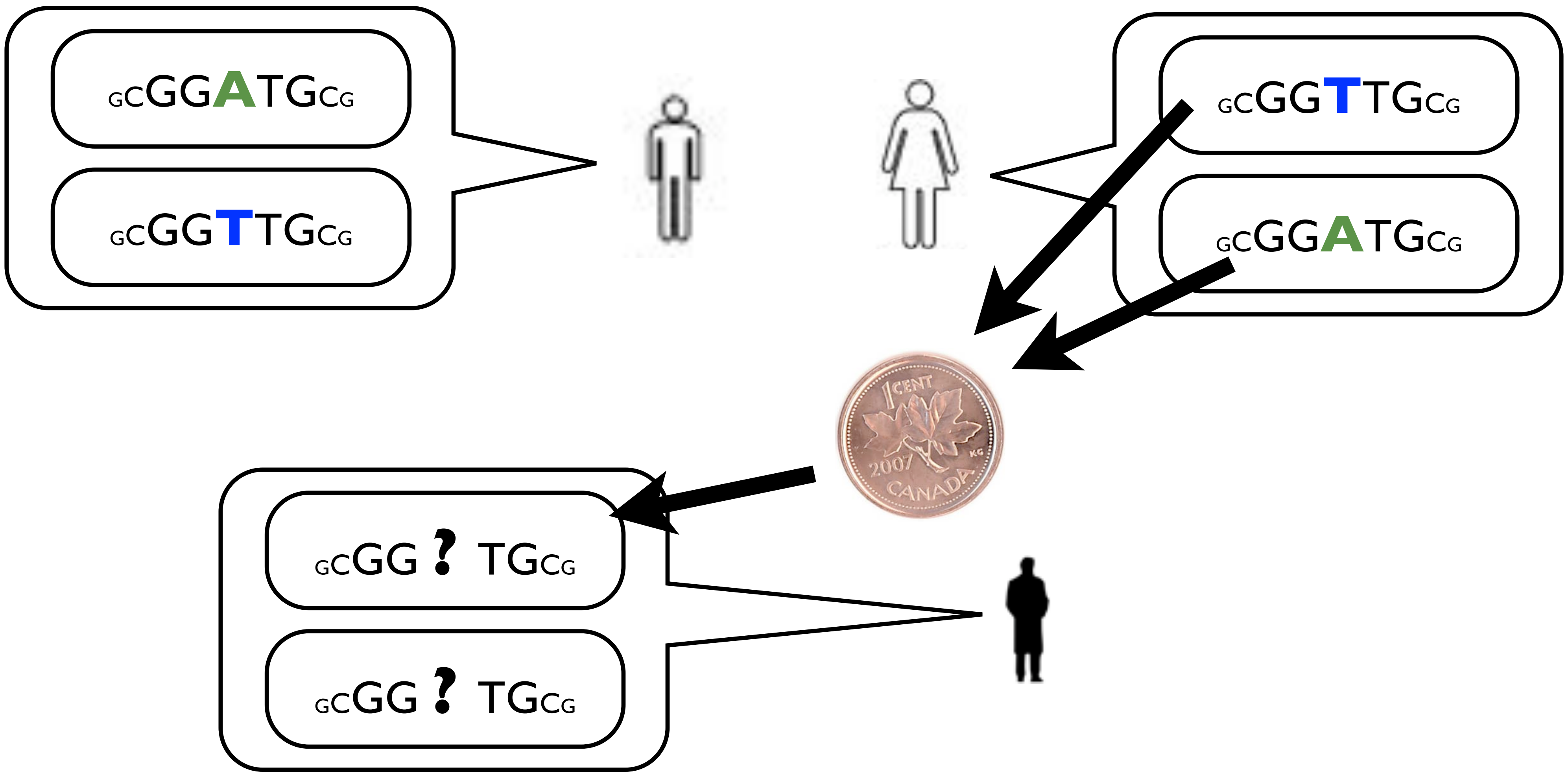
Genetics: inheritable diseases

Randomness in inheritance :



Ex. 6

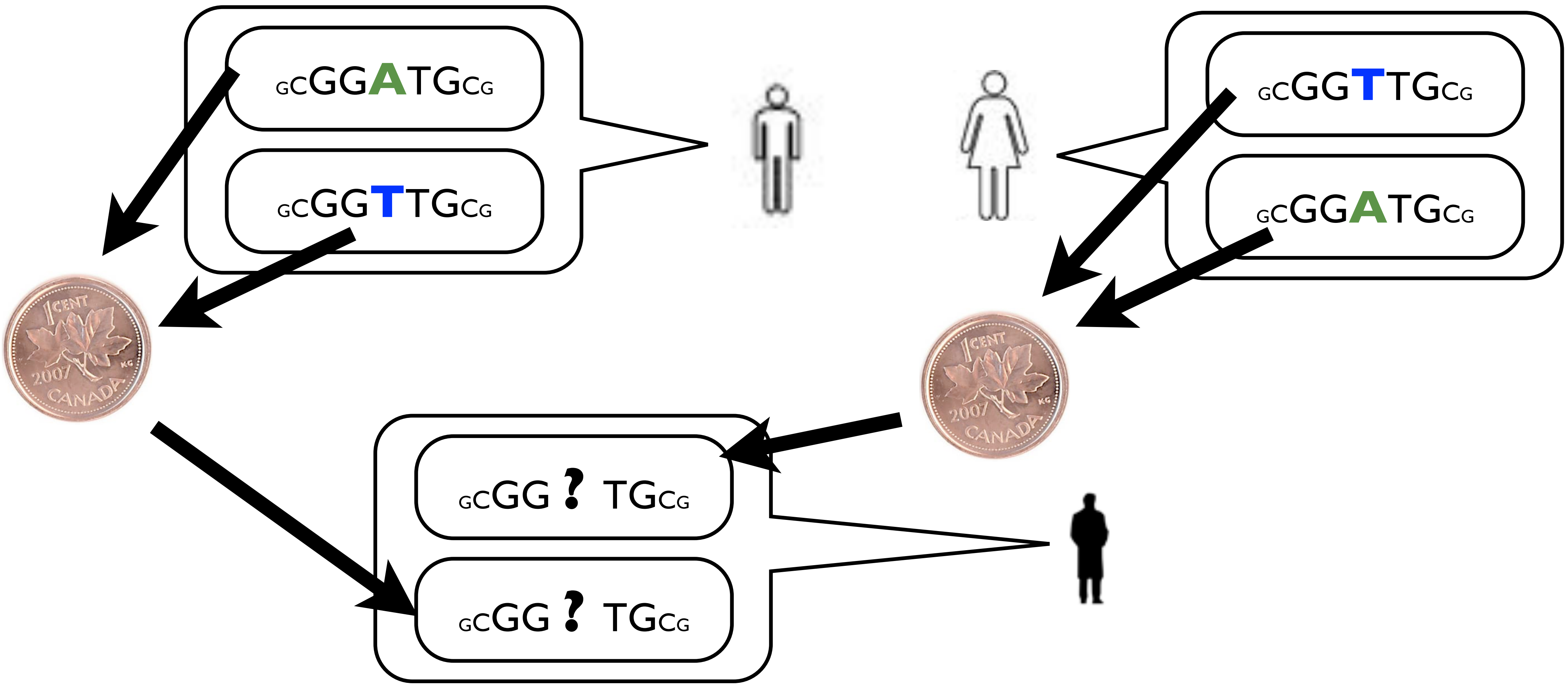
Genotype inheritance



1) Flip a fair coin to decide if you inherit mom's T or A

Ex. 6

Genotype inheritance

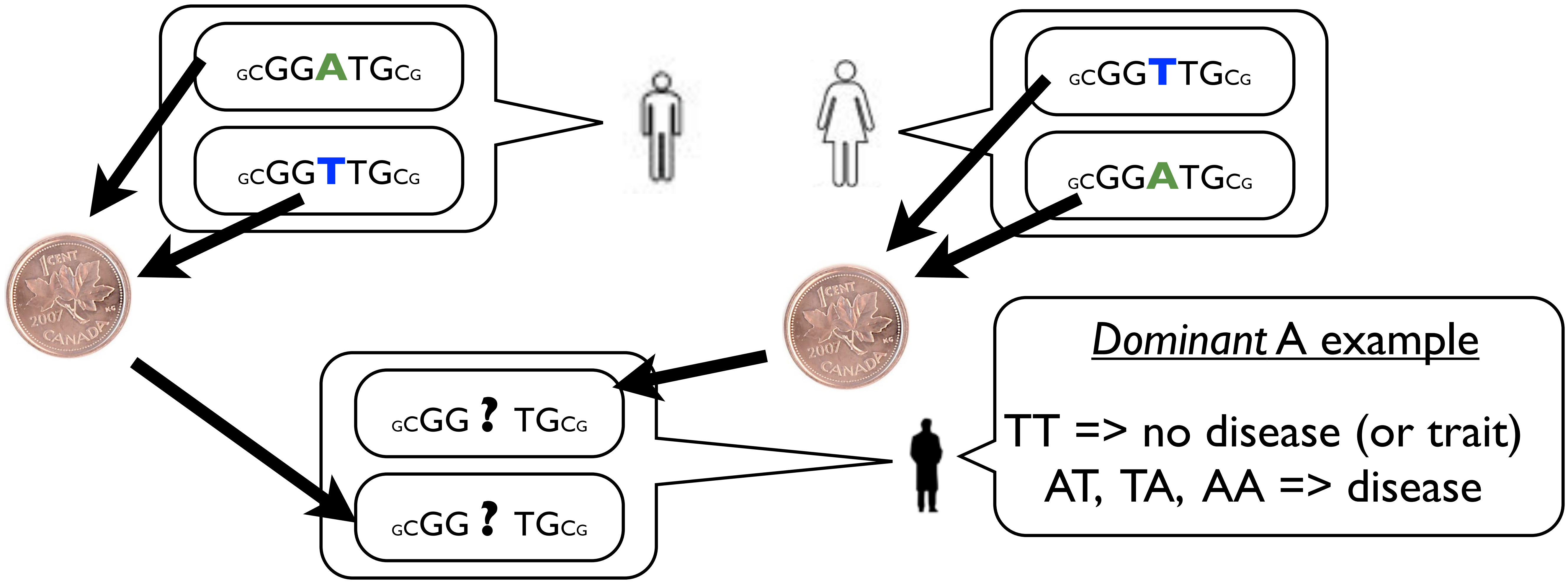


1) Flip a fair coin to decide if you inherit mom's T or A

2) Flip another fair coin to decide if you inherit dad's T or A

Ex. 6

Genotype inheritance



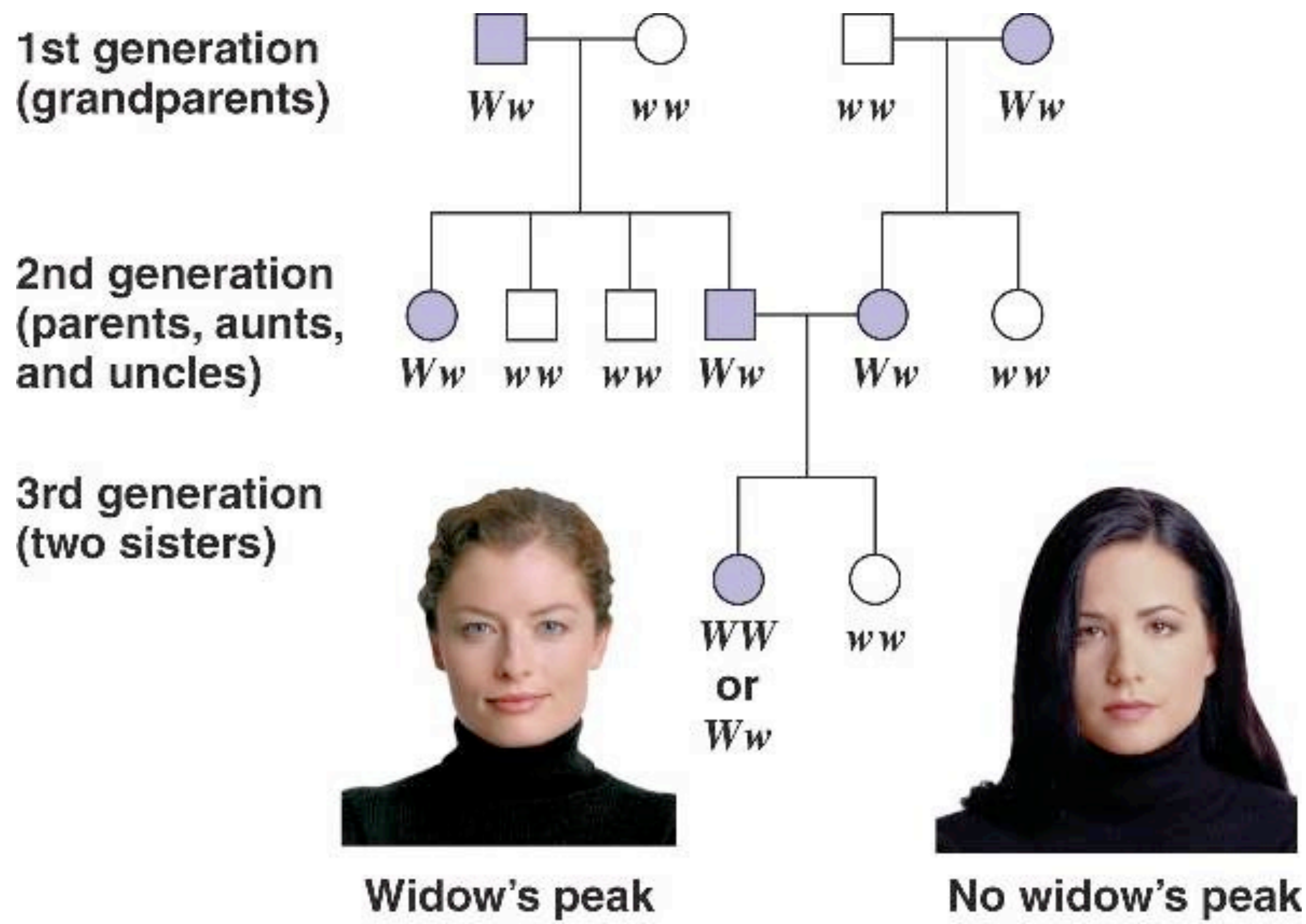
1) Flip a fair coin to decide if you inherit mom's T or A

2) Flip another fair coin to decide if you inherit dad's T or A

Ex. 6

Larger family trees

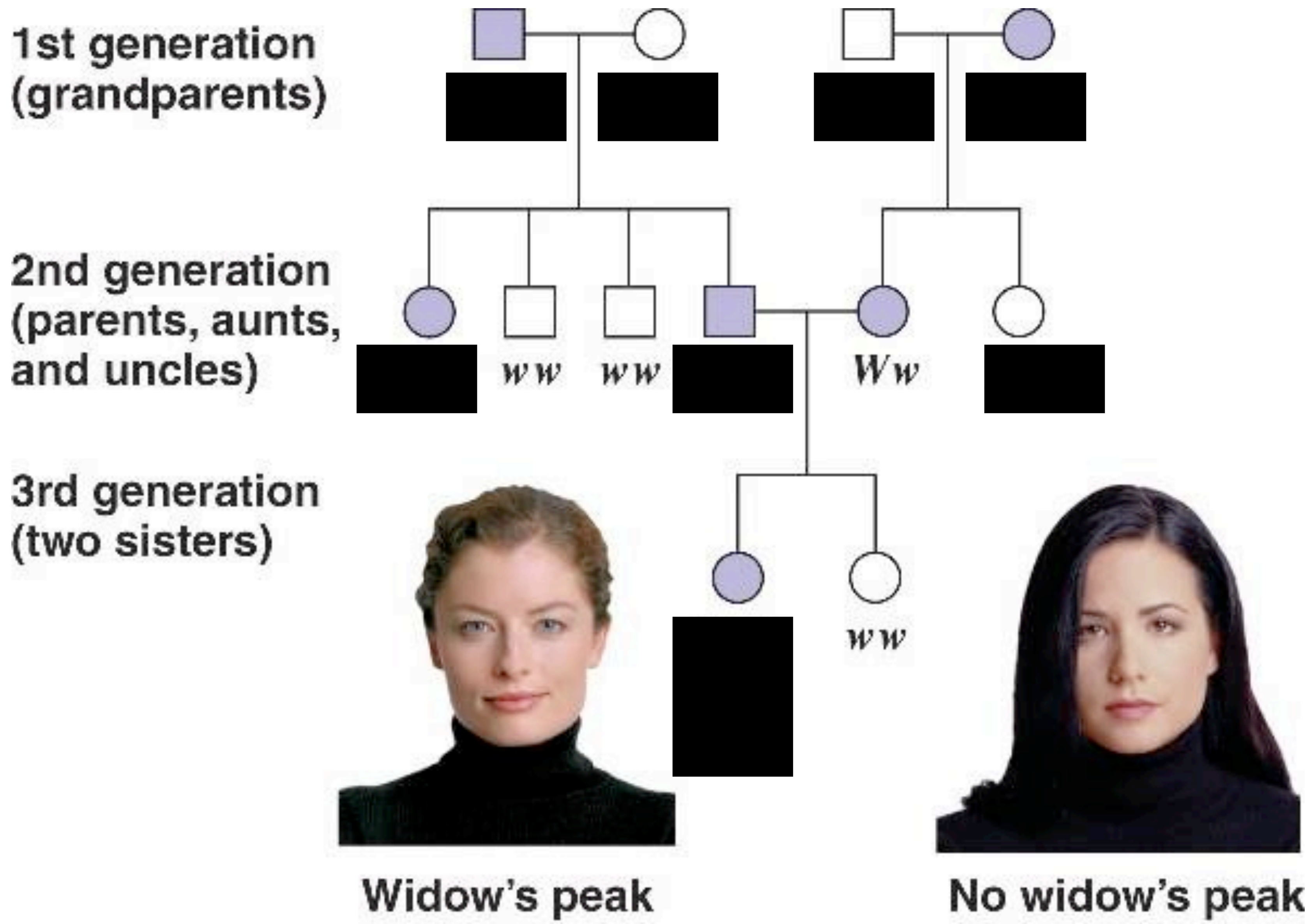
- A larger example where *W* is dominant over *w*
- Goals:
  - genetic counseling
  - finding genetic factor of diseases / traits
- Complication factor
  - incomplete data



Ex. 6

Larger family trees

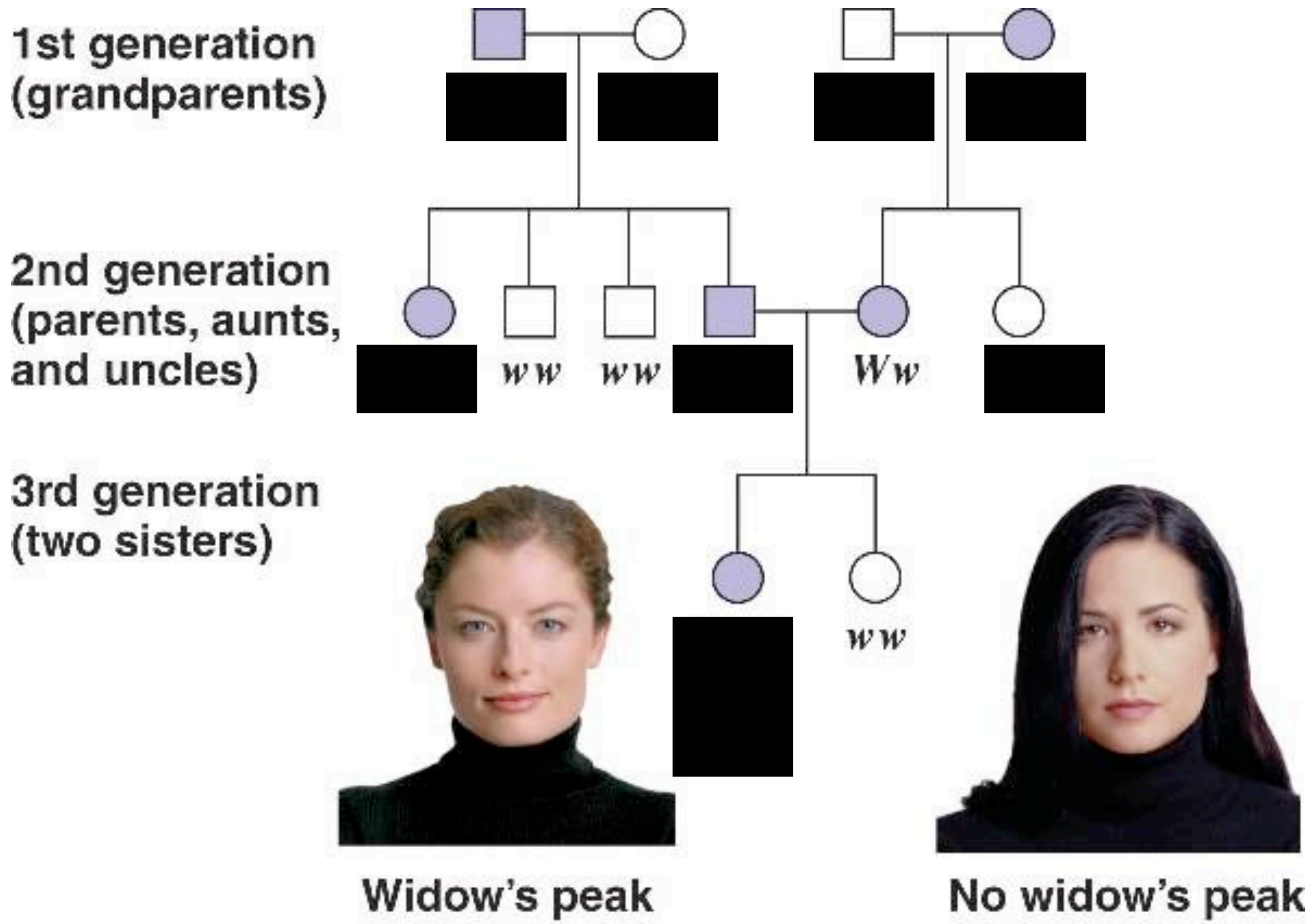
- A larger example where *W* is dominant over *w*
- Goals:
  - genetic counseling
  - finding genetic factor of diseases / traits
- Complication factor
  - incomplete data



Ex. 6

Larger family trees

- A larger example where *W* is dominant over *w*
- Goals:
  - genetic counseling
  - finding genetic factor of diseases / traits
- Complication factor
  - incomplete data

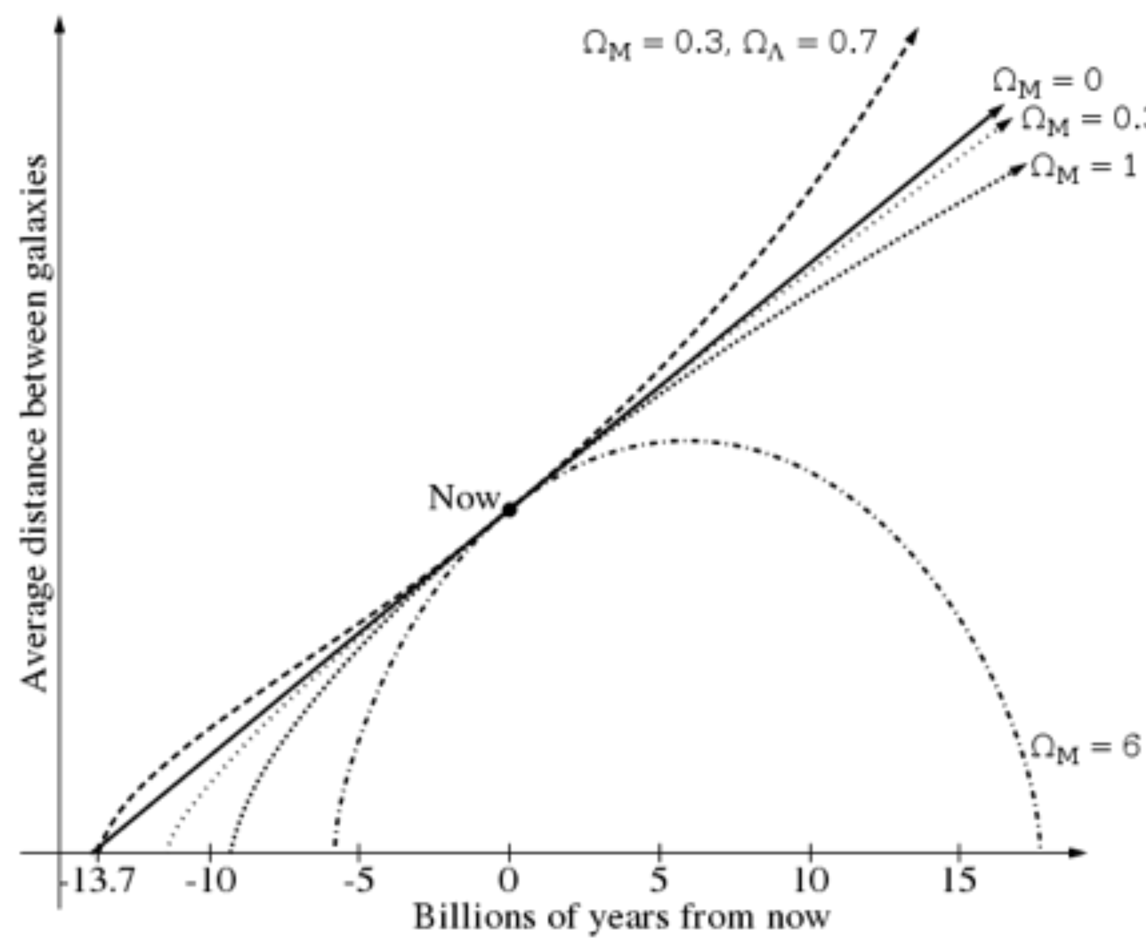


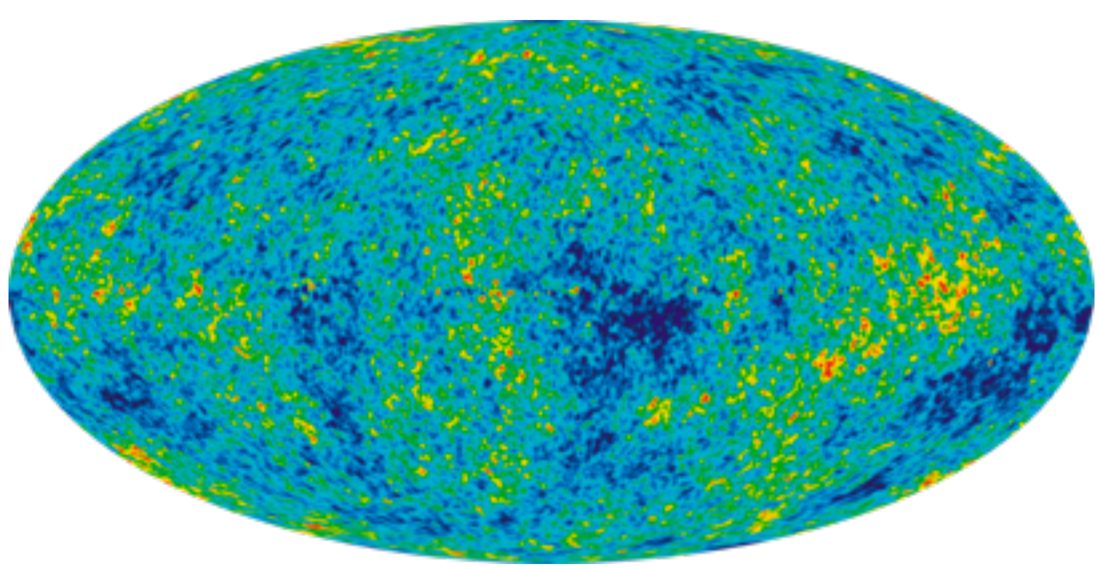
Technique: marginalization

Ex. 7

Astrophysics: Estimating the age and faith of the Universe

- **Goals**: finding the Universe's
  - age
  - density (=> faith)
- **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

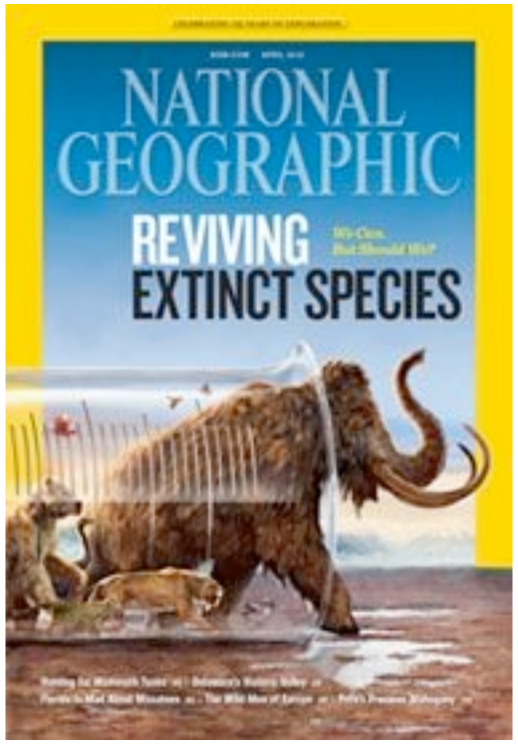




Ex. 8

Phylogenetics: Reconstruction of ancient species

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

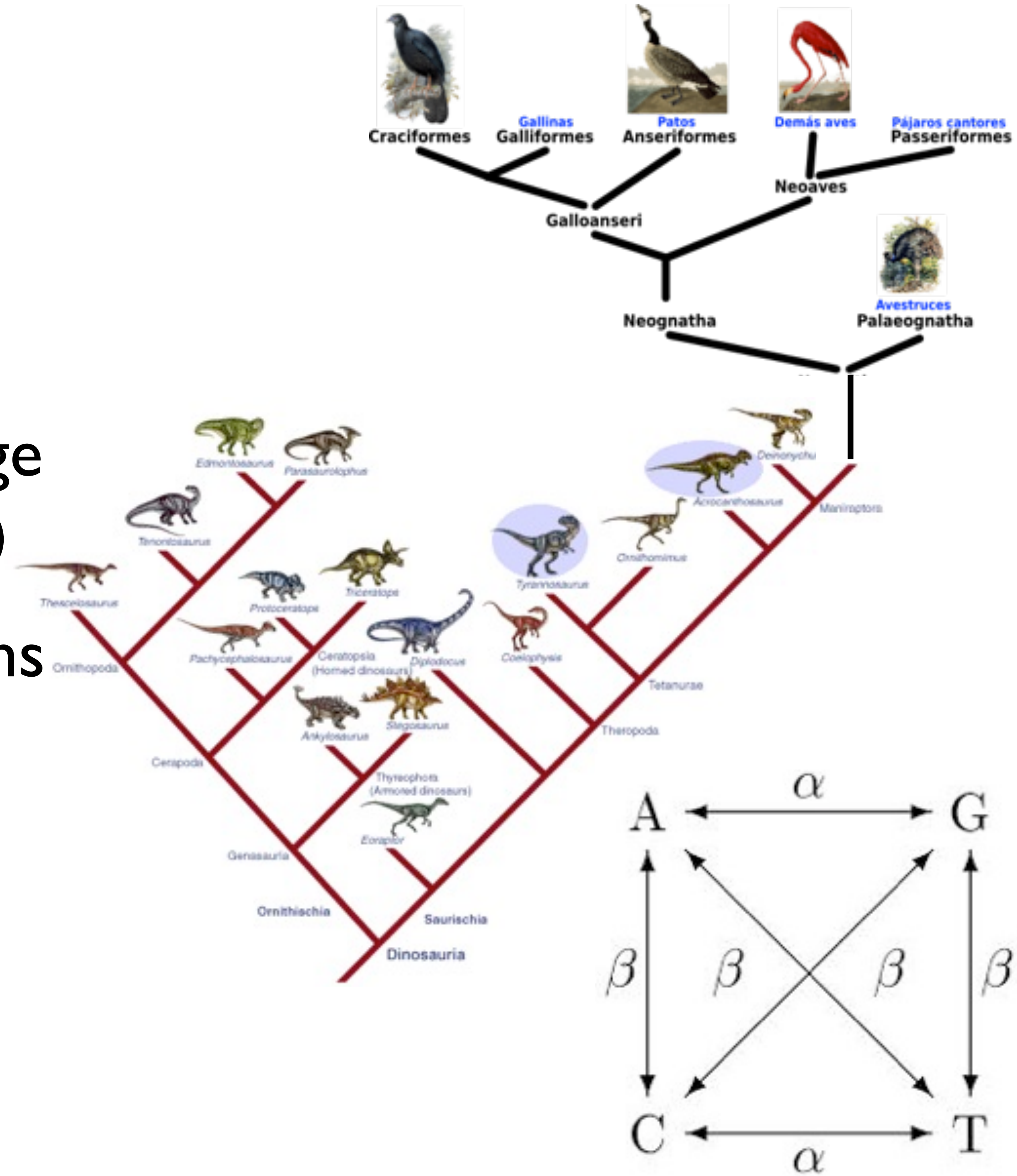


JURASSIC PARK

Ex. 8

Phylogenetic tree

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



Outline of the course

- Discrete probability models
- Conditioning and Bayes
- Expectation
- Continuous probability models
- Asymptotics

Random variables

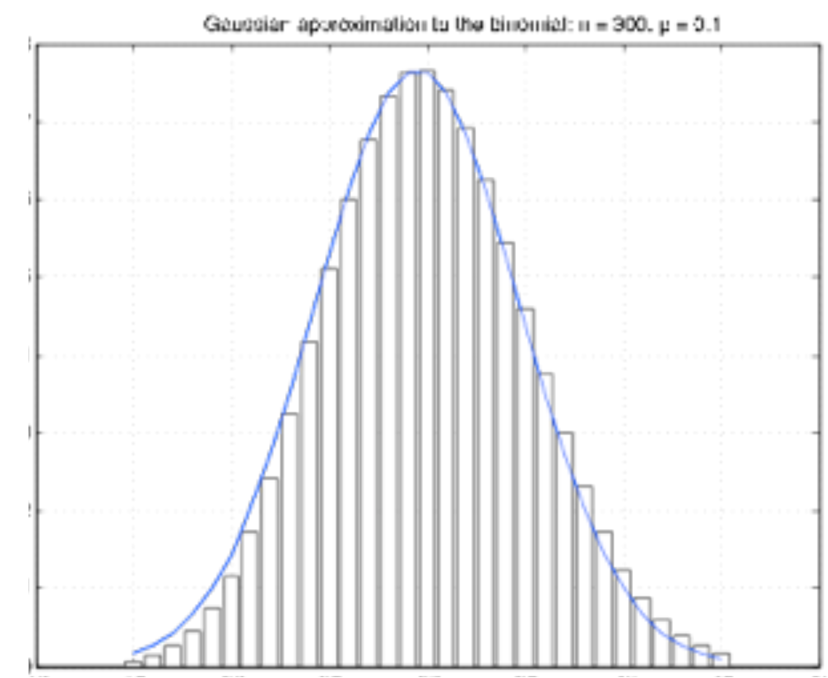
- Fundamental object of study
- Examples of a random variable X
  - The height of a UBC student picked at random
  - Gambling example ('Rademacher coin')

Surprising challenge

- Sums of random variables
  - Omnipresent
  - 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 100 coins is greater that 50.
    - Would have been hard in the pre-computer era
    - Generalized versions of this problem still hard with computer

Asymptotics 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 will spent disproportionate amount of time on some specific types of random variables (normal/gaussian, Poisson, ...)



300 coins