Imperial College London

Maximum Likelihood Estimation CMEE MSc

Dr Tin-Yu Hui
tin-yu.hui11@imperial.ac.uk

10-14 Feb 2020

Schedule (amended)

- 10-14 Feb: Maximum Likelihood & Selected topics in Statistics

- 17 Feb (next Monday): Austin on gene drive
- 18-20 Feb: gene drive modelling & simulations
- 21 Feb: Bhavin

MSc Project

- Desk / computer based projects available
- Email us

Learning outcome

- Define random variables, probability distributions, expectations, and associated concepts
- Understand the principles of Maximum Likelihood Estimation
- Perform statistical modelling, hypothesis testing, and parameter estimation under the likelihood framework (by hand or with R)
- Develop your own likelihood models

- Appreciate Statistics, and start to believe that it is more than a subject ☺

Introductory lecture

- Probability vs Statistics

- Why do we need Statistics?

Example 0: German Tank Problem

- During WW2, the Allies wanted to know the number of tanks the German side had produced
- Two methods: conventional intelligence vs Statistics
- Statistical estimation made use of the serial numbers on those captured or destroyed tanks
- What is my best guess, if I spotted a tank with serial number #40?
  - Statistics says around 40

- If multiple tanks were spotted, then
- $\hat{N} = (largest\ serial\ number\ spotted) * \left(1 + \frac{1}{number\ of\ tanks\ captured}\right) - 1$
- "The largest serial number plus the average gap between observations"

- If we look at the real data:

Month	Statistical estimate	Intelligence estimate	German records
Jun 1940	169	1000	122
Jun 1941	244	1550	271
Aug 1942	327	1550	342

https://www.theguardian.com/world/2006/jul/20/secondworldwar.tvandradio

- Statistical models/estimations could be useful (Of course!)

Probability vs Statistics

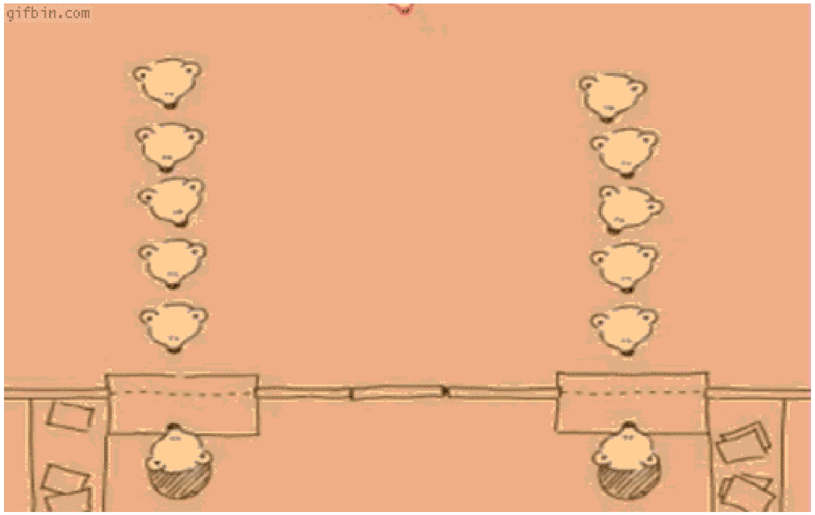
- A Probabilistic question:
  - Given a fair coin, what is the probability of tossing three heads in a row?

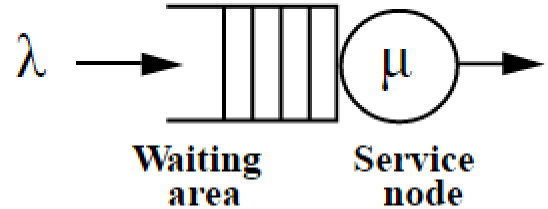
Calculate the chance of occurrence of a certain event, based on some (given) random mechanisms.

- A Statistical question:
  - I tossed three heads in a row, is the coin fair?

Given the observation, what inferences can we make about the underlying mechanism?

- Queuing system in fast food shop





@wikicommons

You are in: UK

Front Page
World
UK
England
Northern Ireland
Scotland
Wales
UK Politics
Business
Sci/Tech
Health
Education
Entertainment
Talking Point
In Depth
AudioVideo

COMMONWEALTH GAMES
BBC SPORT
BBC Weather

SERVICES
Daily E-mail
News Ticker
Mobiles/PDAs

Feedback
Help

Low Graphics

Monday, 13 April, 1998, 07:09 GMT 08:09 UK

Do diapers drive dads to drink?

Men who buy nappies are likely to put some cans of beer in their shopping trolley to shore up their macho image, according to a supermarket chain.

Research by Tesco reveals a link between diaper and beer purchases by male customers.

Records show that fathers buying nappies are much more likely than other men to pick up a six-pack or two.

Tesco Clubcard manager Nick Green puts the trend down to the fathers wanting to prove they are still "real men".

He said: "Despite advances towards sexual equality, many men still feel embarrassed when they have to buy nappies.

"They fear people will think them hen-pecked husbands ordered by their wives to buy the nappies.

See also:
- 04 Apr 98 | UK
  Dads 'to be paid' for baby love

Internet links:
- Tesco
- The Campaign for Real Ale
- Families Need Fathers
- UK Homebrew
- Pubworld

The BBC is not responsible for the content of external internet sites

Top UK stories now:
- Postcode lottery in GP services
- Leanne killer jailed for life
- Tories attack Brixton drugs scheme
- Straw defends arms sales change
- IVF mix-up heads for court
- Police shoot man on the M6
- New challenge excites Venables
- Judge urges life sentence shake-up

Links to more UK stories are at the foot of the page.

- Beer and nappies (and dad?) association…

- Wright-Fisher model

- If the current allele frequency is p, then the allele counts in the next generation due to drift will be binomially distributed with size $2N$ and prob p.

- For two populations with migration rate m, the mean population differentiation between them is $F_{ST} \approx \frac{1}{1+4Nm}$

EVOLUTION IN MENDELIAN POPULATIONS

SEWALL WRIGHT
University of Chicago, Chicago, Illinois

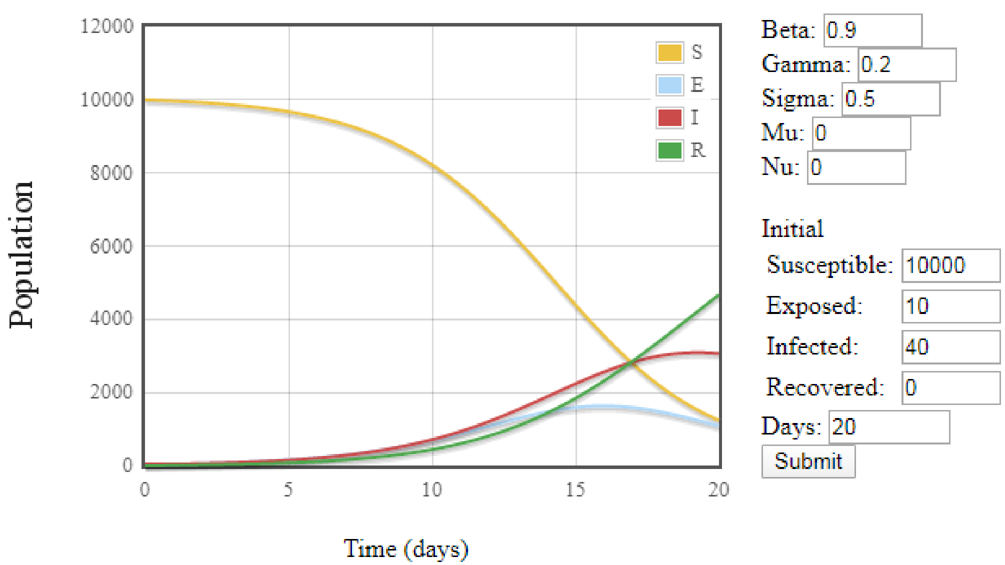
Received January 20, 1930

TABLE OF CONTENTS

- If I observed a certain amount of change in allele frequencies, what's my best guess of N?

- The data suggested $F_{ST} = 0.2$. Can I rule out the no-migration scenario?

- The use of epidemiological models to predict virus outbreak



$$\begin{aligned} \frac{dS}{dt} &= \mu(N-S) - \beta\frac{SI}{N} - \nu S \\ \frac{dE}{dt} &= \beta\frac{SI}{N} - (\mu+\sigma)E \\ \frac{dI}{dt} &= \sigma E - (\mu+\gamma)I \\ \frac{dR}{dt} &= \gamma I - \mu R + \nu S \\ N &= S+E+I+R \end{aligned}$$

- Estimating R_0 the basic reproductive number from infection data

Table 1: Best-case, central and worst-case estimates of 2019-nCoV human-to-human R_0 compatible with either 4000 (top half of table) or 1000 (bottom half of table) total cases by 18/01/2020. Values of R_0 >1 represent self-sustaining human-to-human and are highlighted in red. Baseline estimates highlighted in bold.

Number of cases caused by zoonotic exposure	Assumed total number of cases by 18/01/2020	Best-case R_0	Central (median) R_0	Worst-case R_0
40	**4000**	**2.1**	**2.6**	**3.5**
80	4000	1.8	2.2	2.7
120	4000	1.7	2.0	2.4
160	4000	1.6	1.8	2.2
200	4000	1.5	1.7	2.0
40	1000	1.4	1.9	2.7
80	1000	1.2	1.5	2.0
120	1000	1.1	1.3	1.7
160	1000	1.0	1.2	1.5
200	1000	0.9	1.1	1.3

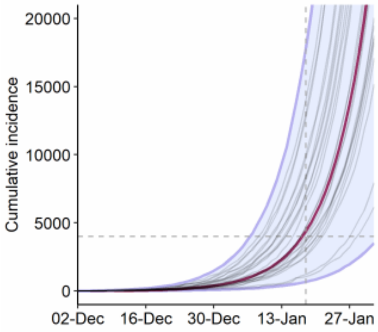


Figure 1: Illustration of estimation method for central estimate of R_0=2.6. Red curve represents median cumulative case numbers over time, calculated from 5000 simulated trajectories of the epidemic, assuming zoonotic exposure of 40 cases in December 2019 and the generation time and variability in infectiousness of SARS. The grey region indicates the 95 percentile range of trajectories – individual simulated epidemics (a random subset of which are shown as light grey curves) are highly variable, reflecting the random nature of disease transmission. Dotted lines indicate January 18th (vertical) and 4000 cumulative cases (horizontal).

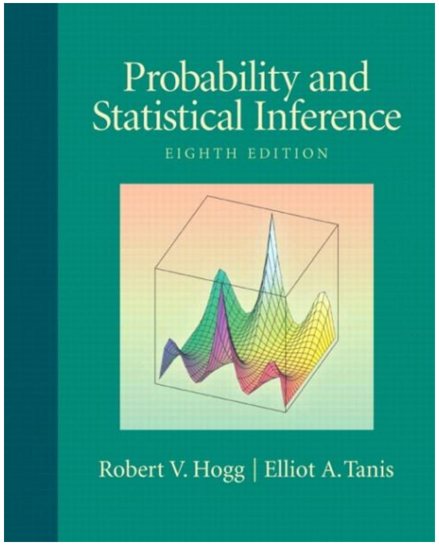
Imai et al. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/news--wuhan-coronavirus/

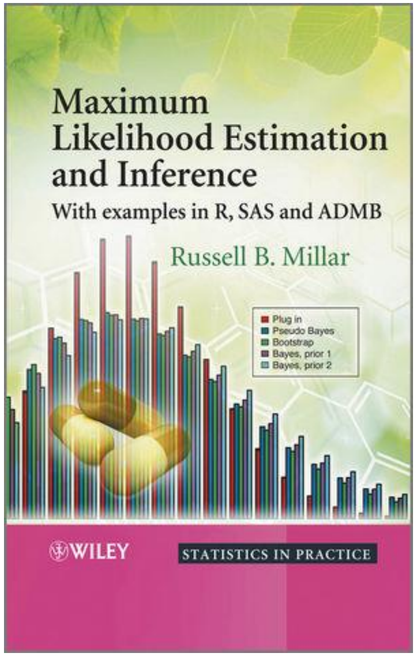
Statistical inference

- Point estimation
  - one-numbered estimate ("best guess")
- Interval estimation
  - 95% confidence interval
- Hypothesis testing
  - H_0 vs H_1
  - model selection

Suggested readings

- Hogg & Tanis, *Probability and Statistical Inference.*
- Millar, *Maximum Likelihood Estimation and Inference.*
- Crawley, *The R Book.*





"In war-time, truth is so precious that she should always be attended by a bodyguard of lies. " – Winston Churchill