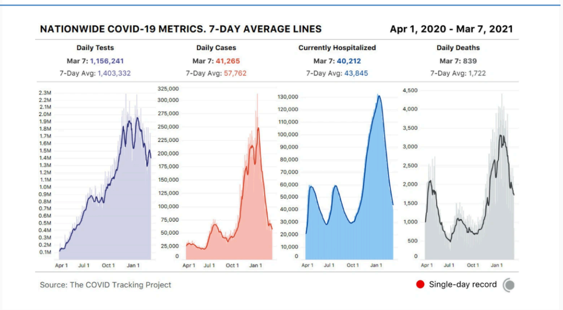
Mathematical Statistics II

STA2212H S LEC9101

Week 10

March2 24 2021

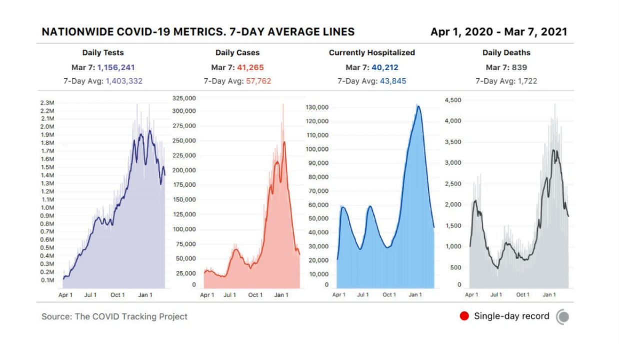
Start recording!

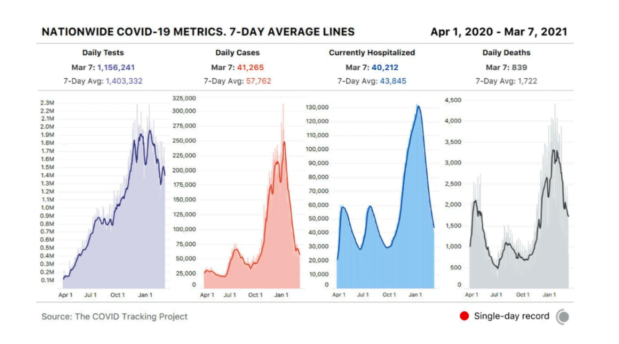


link

Why the Pandemic Experts Failed

The Atlantic, Mar 15, Meyer & Madrigal





"*Data* might seem like an overly technical obsession, an oddly nerdy scapegoat on which to hang the deaths of half a million Americans. But data are how our leaders apprehend reality. In some sense, data are the federal government's reality. As a gap opened between the data that the leaders imagined should exist and the data that actually did exist, it swallowed the country's pandemic response."

"Before March 2020, the country had no shortage of pandemic preparation plans. Many stressed the importance of data-driven decision making. Yet these plans largely assumed that detailed and reliable data would simply ... exist. They were less concerned with how those data would actually be made."

In Canada

Replying to @CTVNews and @CitImmCanada

Last year, Canada hit a record high in the number of deaths recorded in a single year, reporting over 300,000 deaths in 2020.

According to the Public Health Agency of Canada, COVID-19 accounted for approximately one death out of every 20 deaths in Canada last year.

Calling Bullshit @callin_bull · Mar 19

This is a train wreck of a data graphic from the Public Health Agency of Canada.

It implies a minimum of a 5% infection fatality rate, and that only if everyone in Canada had been infected.

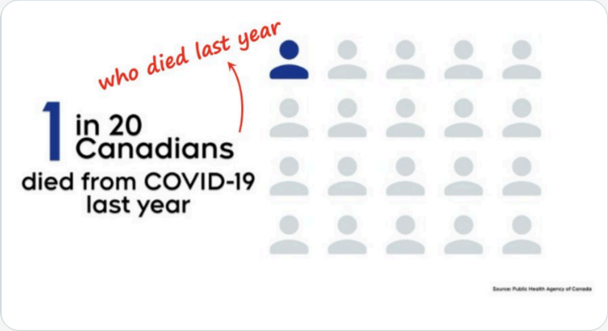
Can you spot went wrong?

Answer in the next post.

14 23 96

Calling Bullshit @callin_bull · Mar 19

What it should say is that one in twenty deaths in Canada last year were due to Covid, or that one in twenty Canadians *who died last year* died of Covid.



2 10 62

link

Recap: shrinkage estimation

- Bayesian hierarchical models lead to shrinkage estimators

$$x_i \mid \theta_i \sim N(\theta_i, \sigma^2); \theta_i \mid \mu \sim N(\mu, \tau^2); \mu \sim N(\mu_0, b^2)$$

- shrinks towards mean of hyper-prior $\tilde{\theta}_{HB,i} = w_i x_i + (1 - w_i)\widetilde{\mathrm{E}}(\mu \mid x)$ Mar 17 Slide 3

Recap: shrinkage estimation

- Bayesian hierarchical models lead to shrinkage estimators

$x_i \mid \theta_i \sim N(\theta_i, \sigma^2); \theta_i \mid \mu \sim N(\mu, \tau^2); \mu \sim N(\mu_0, b^2)$

- shrinks towards mean of hyper-prior $\tilde{\theta}_{HB,i} = w_i x_i + (1 - w_i)\widetilde{\mathrm{E}}(\mu \mid x)$

Mar 17 Slide 3

- empirical Bayes replaces hyper-prior with estimates $\hat{\mu}, \hat{\tau}^2$

using $m(\boldsymbol{x}; \mu, \tau^2) = \int L(\theta; \boldsymbol{x})\pi(\theta \mid \mu, \tau^2) d\theta$

- leads to shrinkage estimators $\tilde{\theta}_{EB,i} = w_i x_i + (1 - w_i)\hat{\mu}$
- in some models it may be possible to estimate prior nonparametrically

e.g. Poisson; EH §6.1

- SM, Example 11.30 "Shakespeare's vocabulary data"; also EH §6.2

"How many words did Shakespeare know?" (Efron & Thisted 1976)

"Did Shakespeare write a newly discovered poem?" (Thisted & Efron, 1987)

Recap: decision theory of estimation

- Loss function
- Risk function – average loss over model

$f(x \mid \theta)$

Recap: decision theory of estimation

- Loss function
- Risk function – average loss over model $f(x \mid \theta)$

- Bayes risk – average Risk function over prior $\pi(\theta)$
- Bayes estimator – minimizes Bayes risk

Recap: decision theory of estimation

- Loss function
- Risk function – average loss over model $f(x \mid \theta)$

- Bayes risk – average Risk function over prior $\pi(\theta)$
- Bayes estimator – minimizes Bayes risk

- minimax estimator – minimizes maximum risk

Recap: decision theory of estimation

- Loss function
- Risk function – average loss over model $f(x \mid \theta)$

- Bayes risk – average Risk function over prior $\pi(\theta)$
- Bayes estimator – minimizes Bayes risk

- minimax estimator – minimizes maximum risk

- admissible estimator – no other estimator has smaller Risk function for all $\theta \in \Theta$
- Bayes' estimators with proper priors are admissible

Recap: decision theory of estimation

- Loss function
- Risk function – average loss over model $f(x \mid \theta)$

- Bayes risk – average Risk function over prior $\pi(\theta)$
- Bayes estimator – minimizes Bayes risk

- minimax estimator – minimizes maximum risk

- admissible estimator – no other estimator has smaller Risk function for all $\theta \in \Theta$
- Bayes' estimators with proper priors are admissible

- Stein's paradox – when estimating many parameters, shrinkage estimators beat non-shrinkage (mle) estimators

no prior needed!

Today

1. Calendar, HW, Final HW
2. Friday
3. Stein's paradox – baseball example
4. multivariate distributions (Ch 14 AoS, SM 6.3)

- Mar 29 3.00 – 4.00 pm EDT
  Data Science ARES
  John Aston, University of Cambridge
  "Functional Data in Constrained Spaces"
- Professor of Statistics in Public Life;
  formerly Chief Scientific Adviser
  to the Home Office, UK Government.

Example J-S Parad

Table 7.1 *Eighteen baseball players;* MLE *is batting average in first 90 at bats;* TRUTH *is average in remainder of 1970 season; James–Stein estimator* JS *is based on arcsin transformation of MLEs. Sum of squared errors for predicting* TRUTH*:* MLE *.0425,* JS *.0218.*

Player	MLE	JS	TRUTH	x
1	.345	.283	.298	11.96
2	.333	.279	.346	11.74
3	.322	.276	.222	11.51
4	.311	.272	.276	11.29
5	.289	.265	.263	10.83
6	.289	.264	.273	10.83
7	.278	.261	.303	10.60
8	.255	.253	.270	10.13
9	.244	.249	.230	9.88
10	.233	.245	.264	9.64
11	.233	.245	.264	9.64
12	.222	.242	.210	9.40
13	.222	.241	.256	9.39
14	.222	.241	.269	9.39
15	.211	.238	.316	9.14
16	.211	.238	.226	9.14
17	.200	.234	.285	8.88
18	.145	.212	.200	7.50

Example J-S Paradox

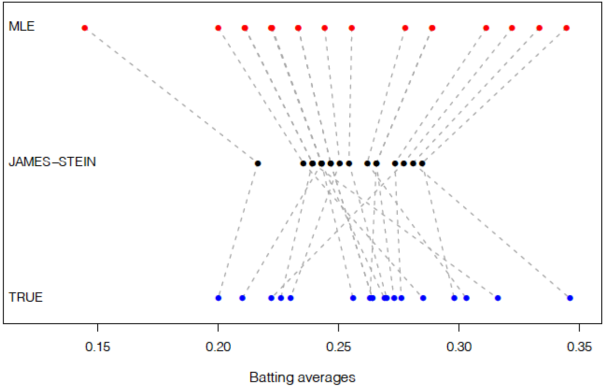


Figure 7.1 Eighteen baseball players; top line MLE, middle James–Stein, bottom true values. Only 13 points are visible, since there are ties.

vector random variables

expected value, variance, precision

i.i.d. samples

unbiased estimators

AoS Thm 14.1

confidence intervals for correlations AoS §14.2

Multivariate distributions

AoS Ch14; SM §6.3

multivariate normal distribution

AoS Thm 14.2

sampling and likelihood function

Hotelling's T^2 statistic

partial and marginal correlation

graphical models

Multivariate distributions

SM §6.3

264 6 · *Stochastic Models*

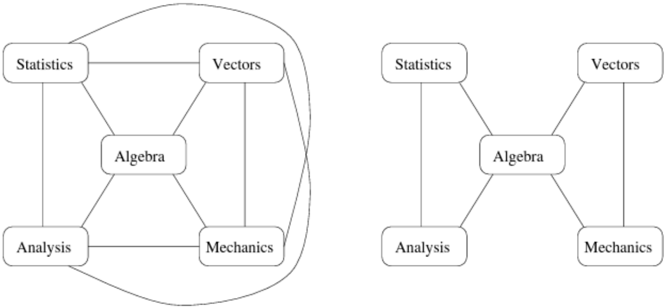


Figure 6.10 Graphs for the full model (left) and a reduced model (right) for the maths marks data. The interpretation of the reduced model is that given the result for algebra, results for vectors and mechanics are independent of those for analysis and statistics.

Multinomial distribution