Lecture 1: Introduction

Instructor: **Aaron Schein**

Modern Methods in Applied Statistics

STAT 34800 (Spring 2023)

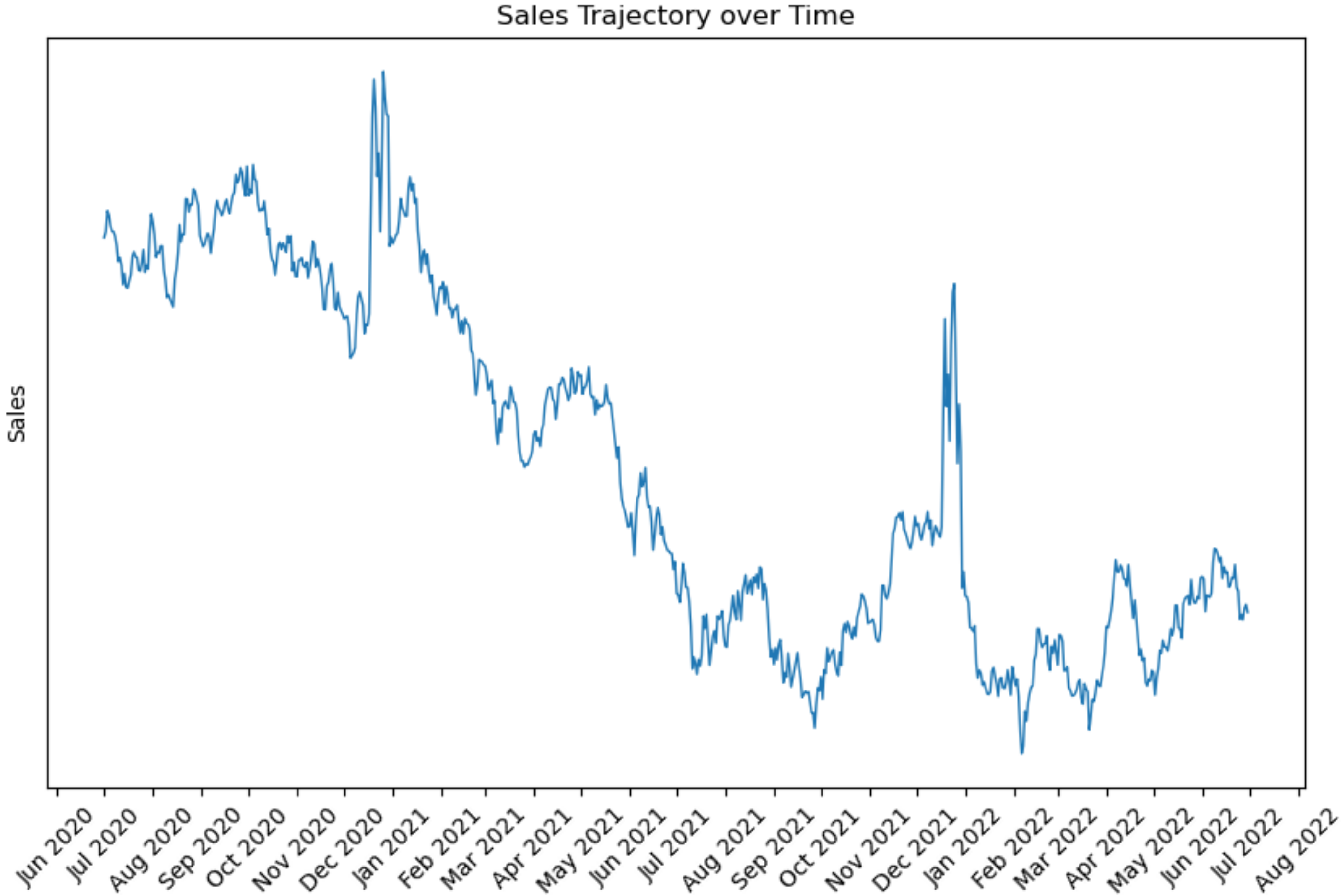
THE UNIVERSITY OF CHICAGO

Intro: Applied statistics

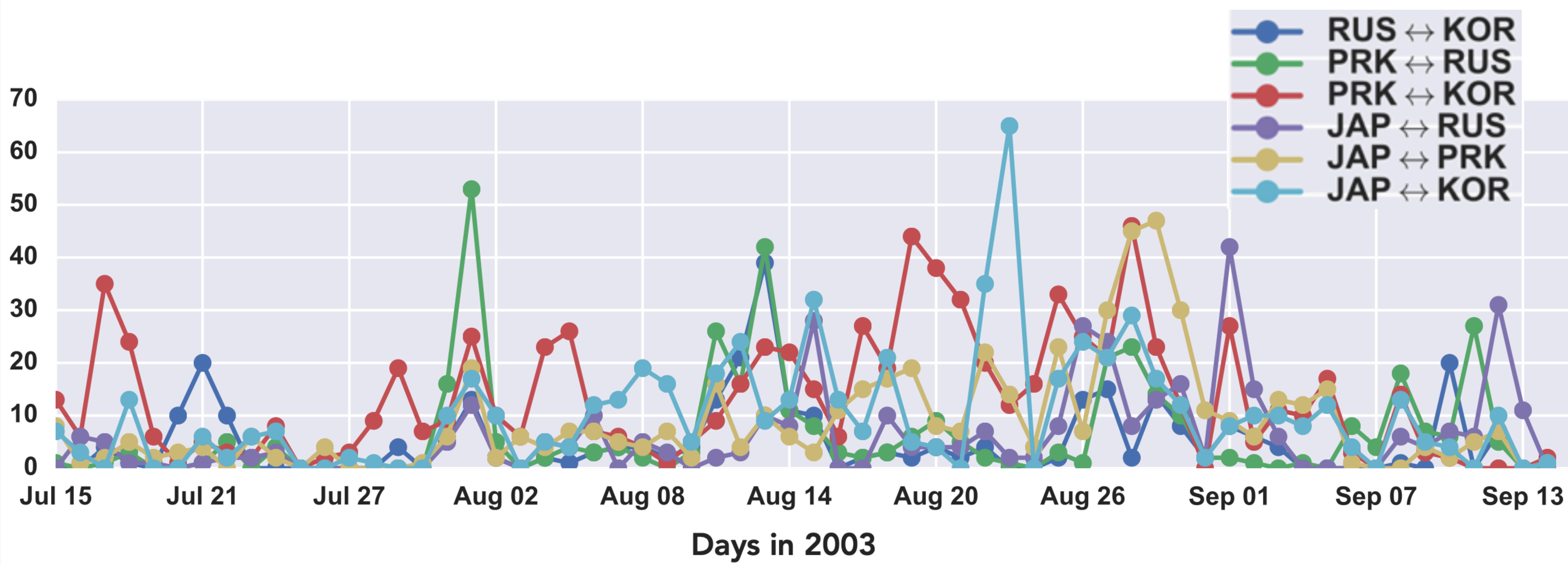
How to combine **domain knowledge** + **data**:

- make tailored predictions
- generate scientific hypotheses
- guide decisions

Intro: Structure in applied problems



Intro: Structure in applied problems



Intro in applied problems

Six-party talks

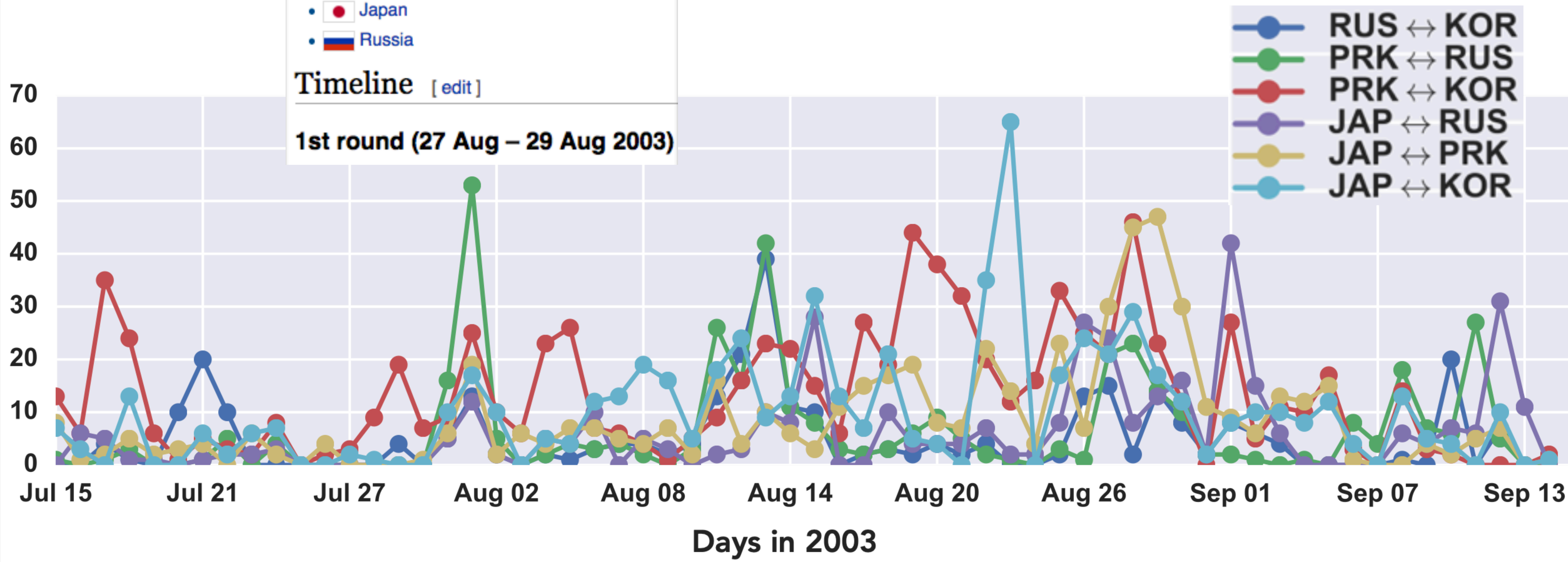
From Wikipedia, the free encyclopedia

The **six-party talks** aim to find a peaceful resolution to the security concerns as a result of the North Korean nuclear weapons program. There has been a series of meetings with six participating states:

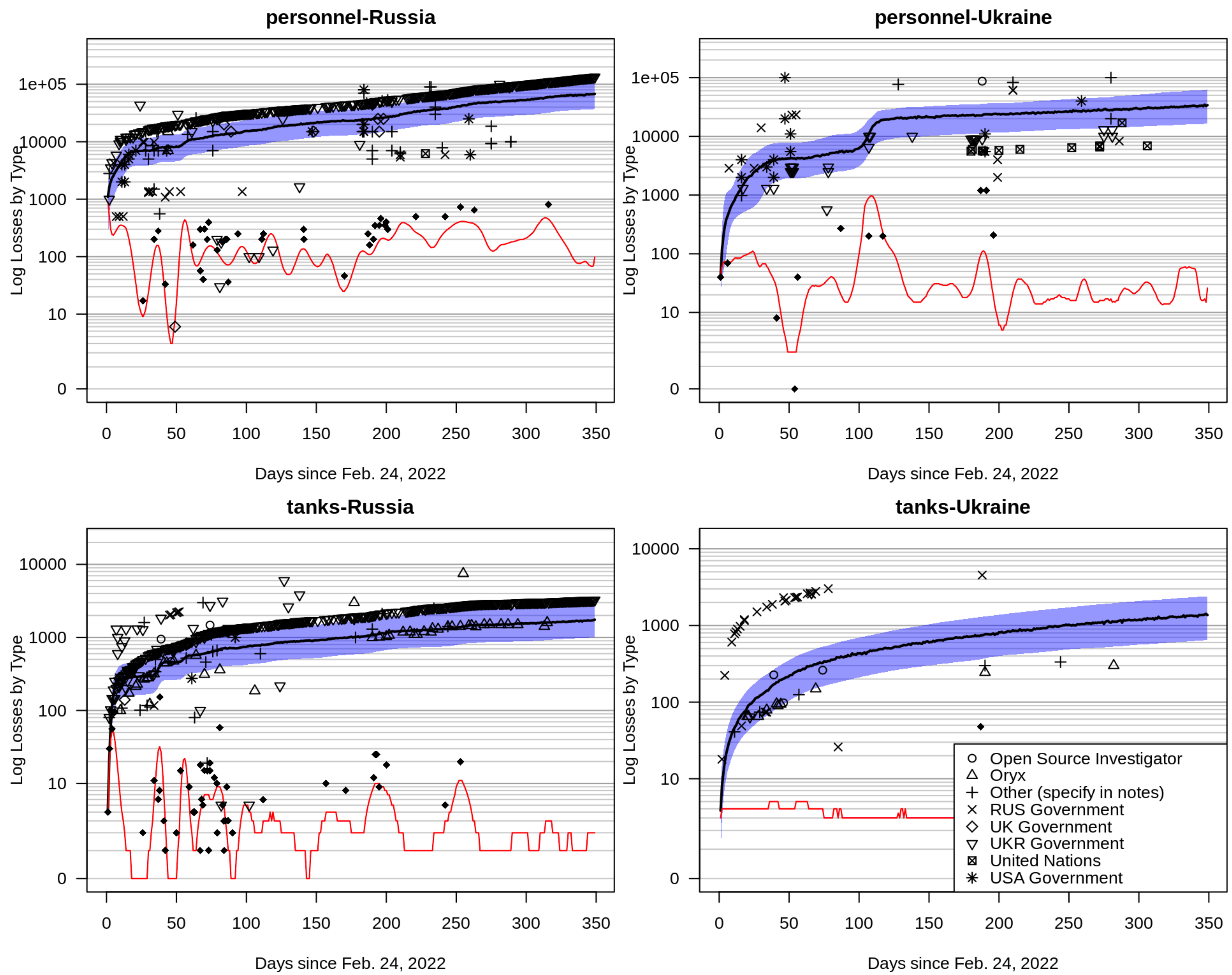
- South Korea
- North Korea
- United States of America
- China
- Japan
- Russia

Timeline [edit]

1st round (27 Aug – 29 Aug 2003)



Intro: Structure in applied problems



Intro: Probabilistic models

Modular framework for:

- Encoding structural knowledge
- Positing latent structure
- Integrating heterogeneous data
- Synthesizing into quantified uncertainty

Intro: Probabilistic models

joint distribution

$$P(Y, Z \mid \mathscr{H})$$

Y : **data**
what we know

Z : **latent variables** and **parameters**
what we don't (and seek to) know

$\mathscr{H}$: **assumptions** and **hyperparameters**
what we (think we) know

Intro: Probabilistic models

joint distribution **likelihood** **prior**

$$P(Y, Z \mid \mathscr{H}) = P(Y \mid Z, \mathscr{H})\; P(Z \mid \mathscr{H})$$

Y : **data**
what we know

Z : **latent variables** and **parameters**
what we don't (and seek to) know

$\mathscr{H}$: **assumptions** and **hyperparameters**
what we (think we) know

Intro: Probabilistic models

$$\overset{\textbf{joint distribution}}{P(Y, Z \mid \mathscr{H})} = \overset{\textbf{likelihood}}{P(Y \mid Z, \mathscr{H})}\ \overset{\textbf{prior}}{P(Z \mid \mathscr{H})}$$

Once you define a model…

$$= \overset{\textbf{posterior}}{P(Z \mid Y, \mathscr{H})}\ \overset{\textbf{evidence}}{P(Y \mid \mathscr{H})}$$

Bayes follows naturally…

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

Intro: Probabilistic models

Building models is a **negotiation…**

$$\overset{\text{posterior}}{P(Z \mid Y, \mathscr{H})} = \frac{\overset{\text{likelihood}}{P(Y \mid Z, \mathscr{H})}\,\overset{\text{prior}}{P(Z \mid \mathscr{H})}}{\underset{\text{evidence}}{P(Y \mid \mathscr{H})}}$$

The ``scientist'' wants to encode **rich structure**…

Intro: Probabilistic models

Building models is a **negotiation**...

$$\underbrace{P(Z \mid Y, \mathscr{H})}_{\text{posterior}} = \frac{\overbrace{P(Y \mid Z, \mathscr{H})}^{\text{likelihood}}\,\overbrace{P(Z \mid \mathscr{H})}^{\text{prior}}}{\underbrace{P(Y \mid \mathscr{H})}_{\text{evidence}}}$$

The ``scientist'' wants to encode **rich structure**...

...but the ``accountant'' is aware of the **cost**:

$$\text{evidence} = \int P(Y, Z \mid \mathscr{H})\, dZ$$

Intro: Probabilistic models

Building models is a **negotiation**...

$$\underbrace{P(Z \mid Y, \mathscr{H})}_{\text{posterior}} = \frac{\overbrace{P(Y \mid Z, \mathscr{H})}^{\text{likelihood}}\, \overbrace{P(Z \mid \mathscr{H})}^{\text{prior}}}{\underbrace{P(Y \mid \mathscr{H})}_{\text{evidence}}}$$

The ``scientist'' wants to encode **rich structure**...

...but the ``accountant'' is aware of the **cost**:

$$\text{evidence} = \int P(Y, Z \mid \mathscr{H})\, dZ$$

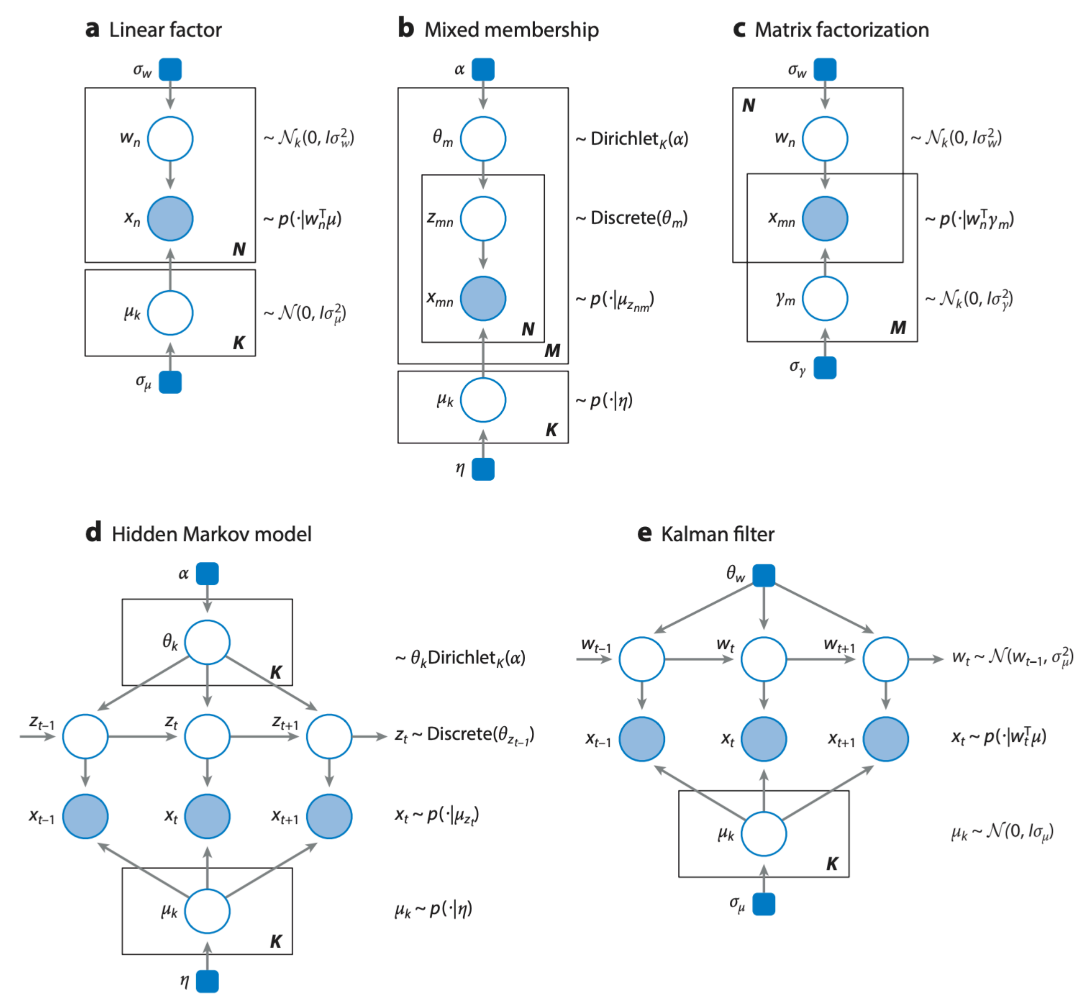
Ultimately, we rely on the ``mathematician'' and the ``engineer''
to provide principled and accurate approximations

Intro: Probabilistic models

This course is about the motifs, themes, techniques that help us build, fit, and check models

Emphasis on modularity, and ``object-oriented'' approach

Probabilistic graphical models:
a language for **modular** model-building



Blei (2014) "Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models"

Intro: Probabilistic models

``Box's Loop"

(This is really Blei's Loop)

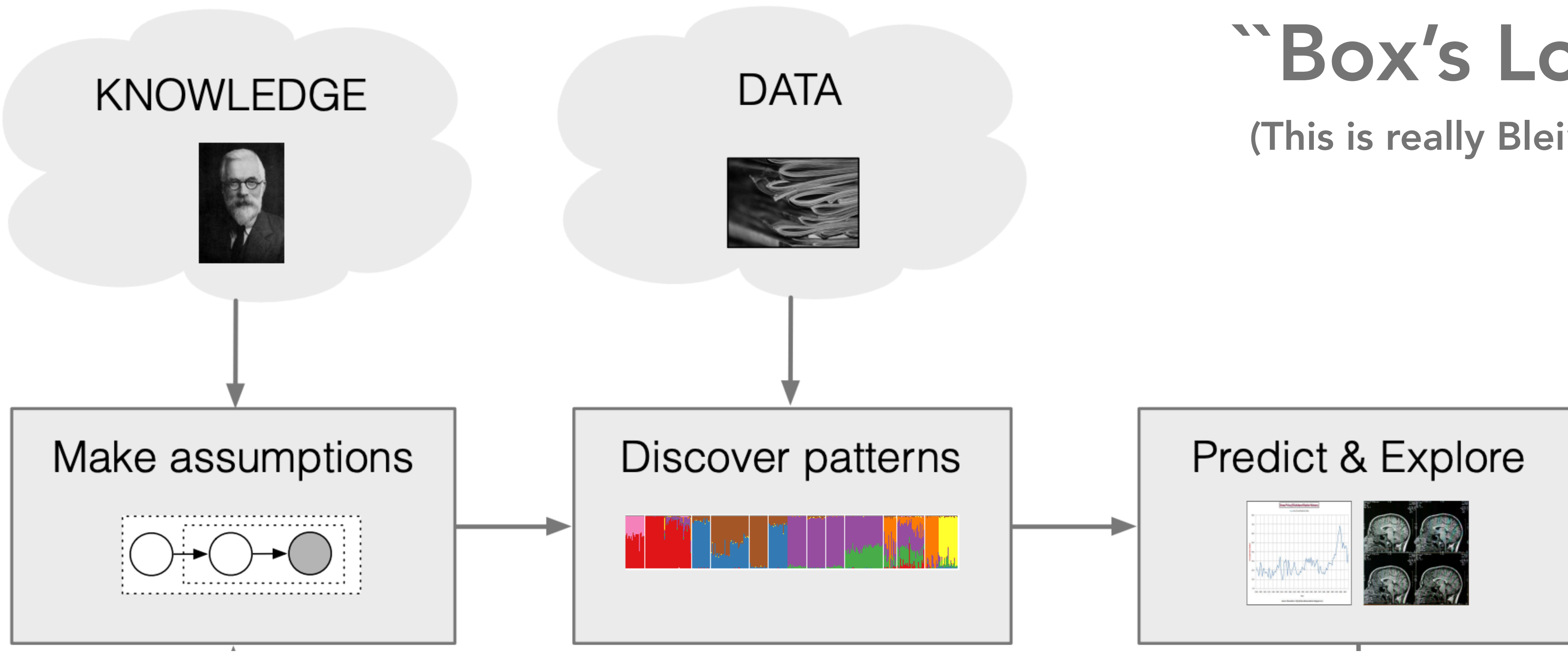


Image from David Blei's lecture notes

Intro: Probabilistic models

"Box's Loop"

(This is really Blei's Loop)

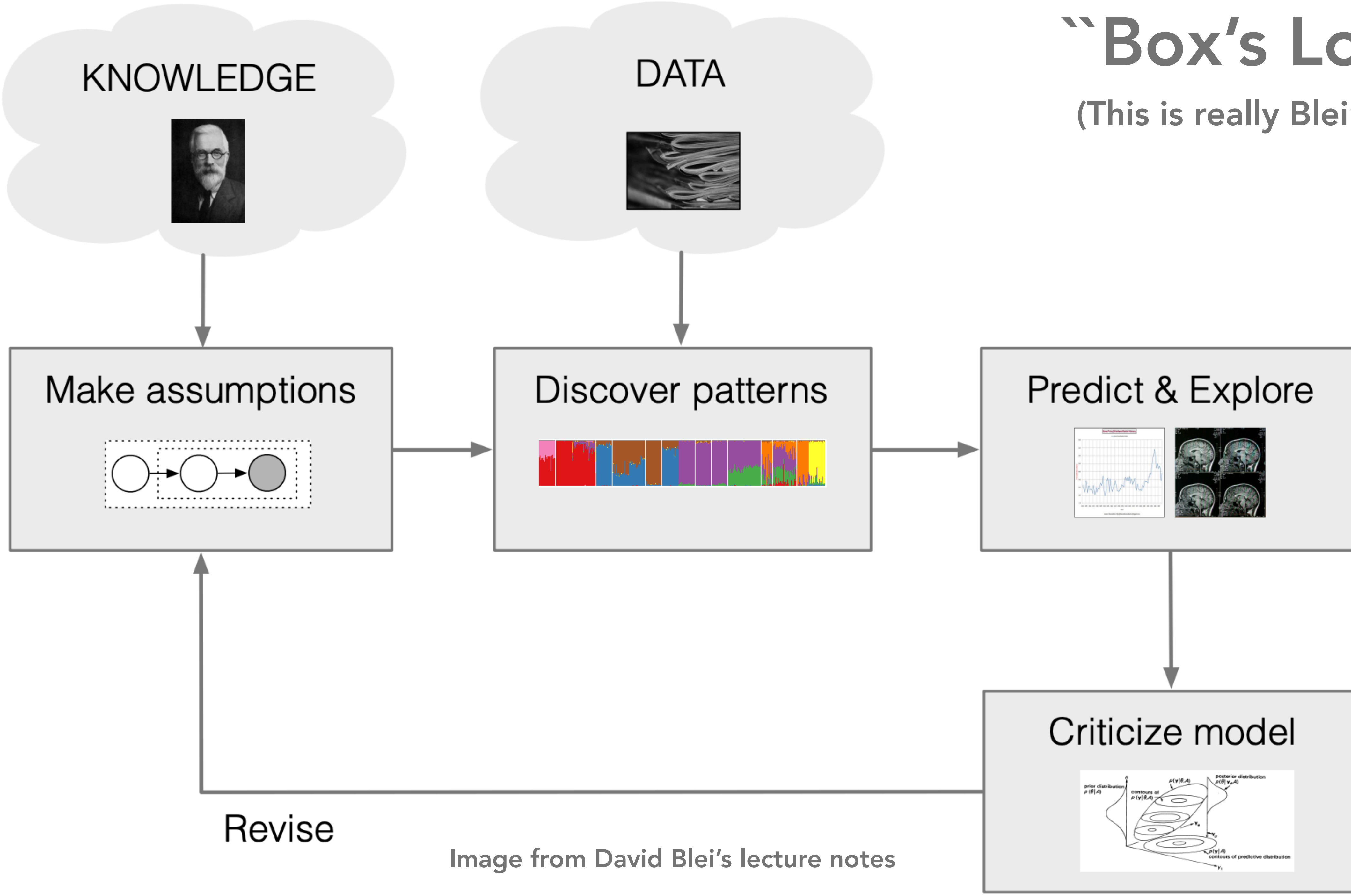
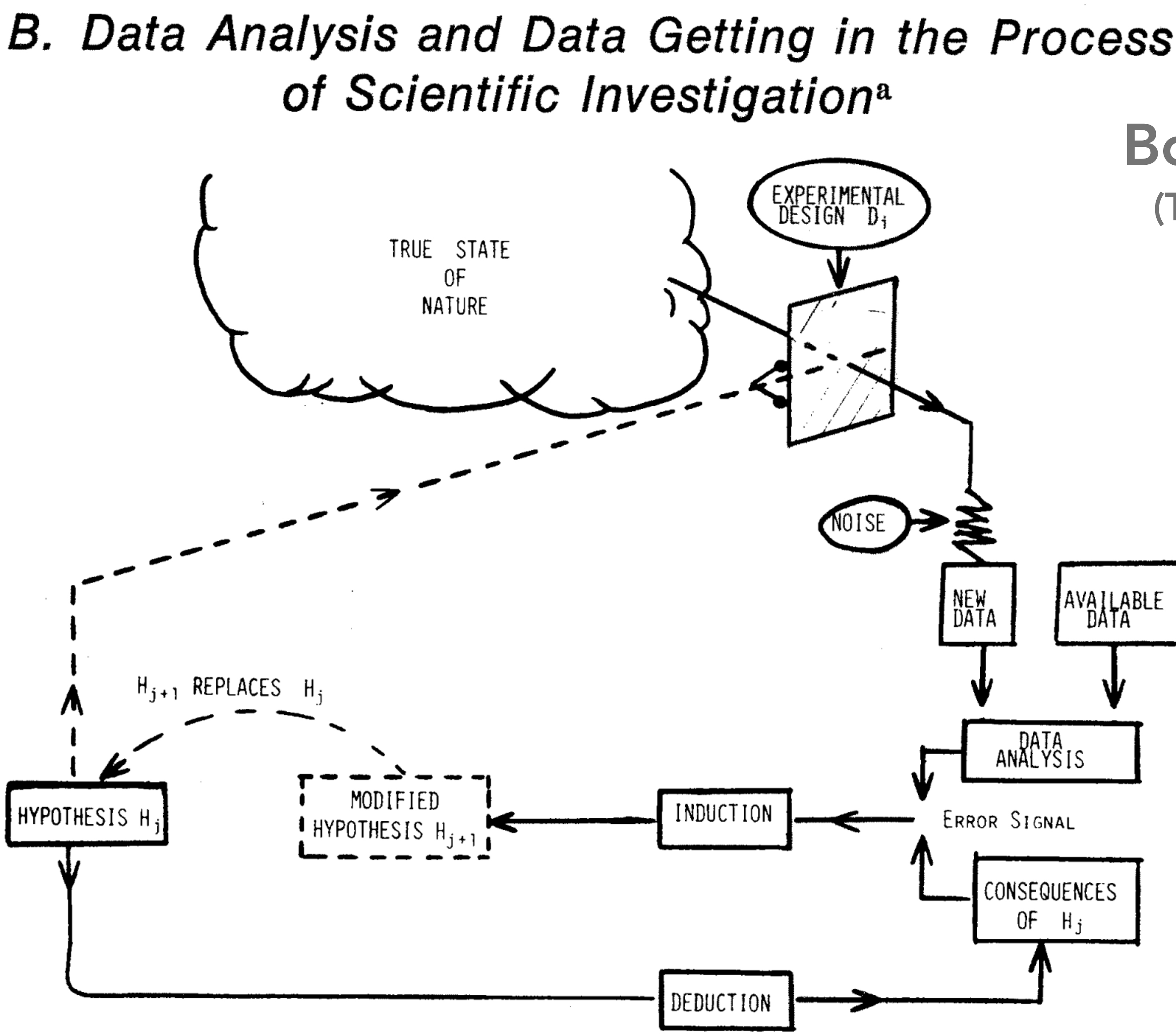


Image from David Blei's lecture notes

Intro: Probabilistic models

Box's Loop
(The original)



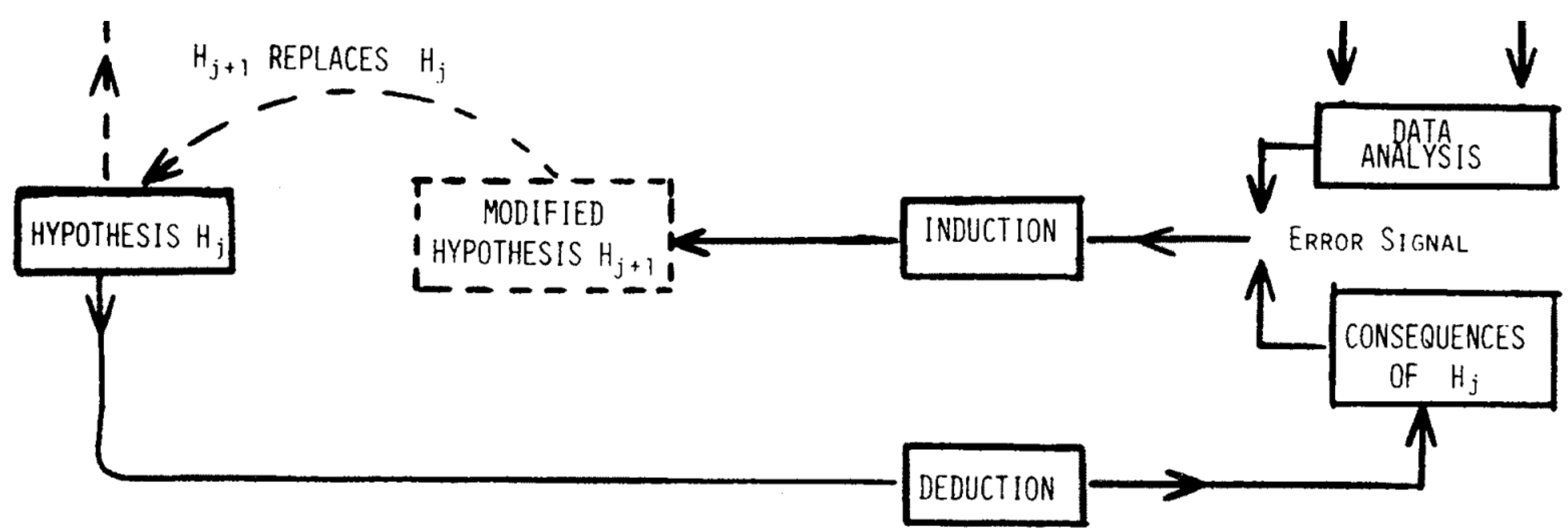
Box (1976) "Science and Statistics"

Intro: Probabilistic models

B. Data Analysis and Data Getting in the Process of Scientific Investigation[a]

Box's Loop
(The original)

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.



Box (1976) "Science and Statistics"

Intro: Structure in applied problems

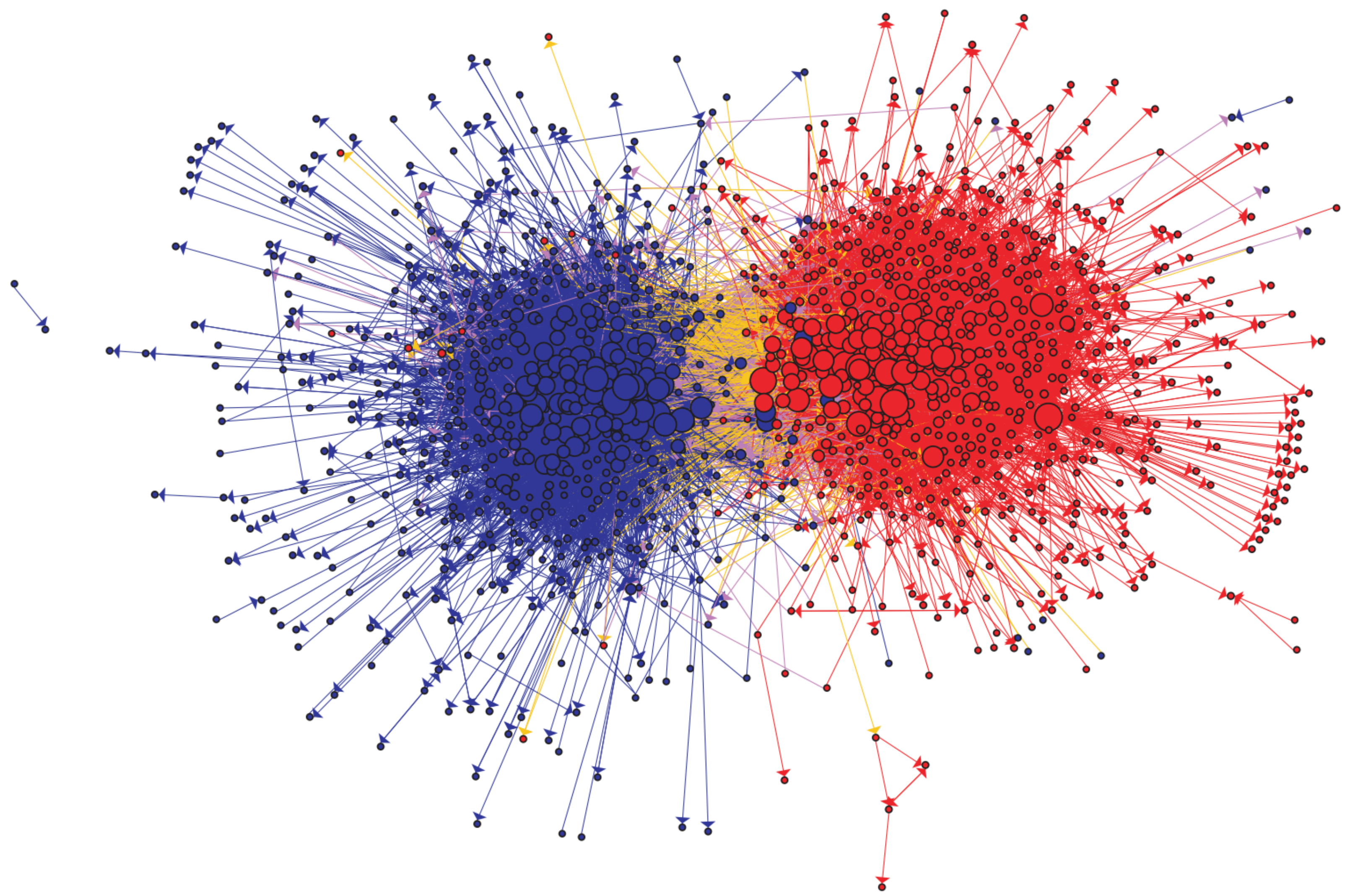
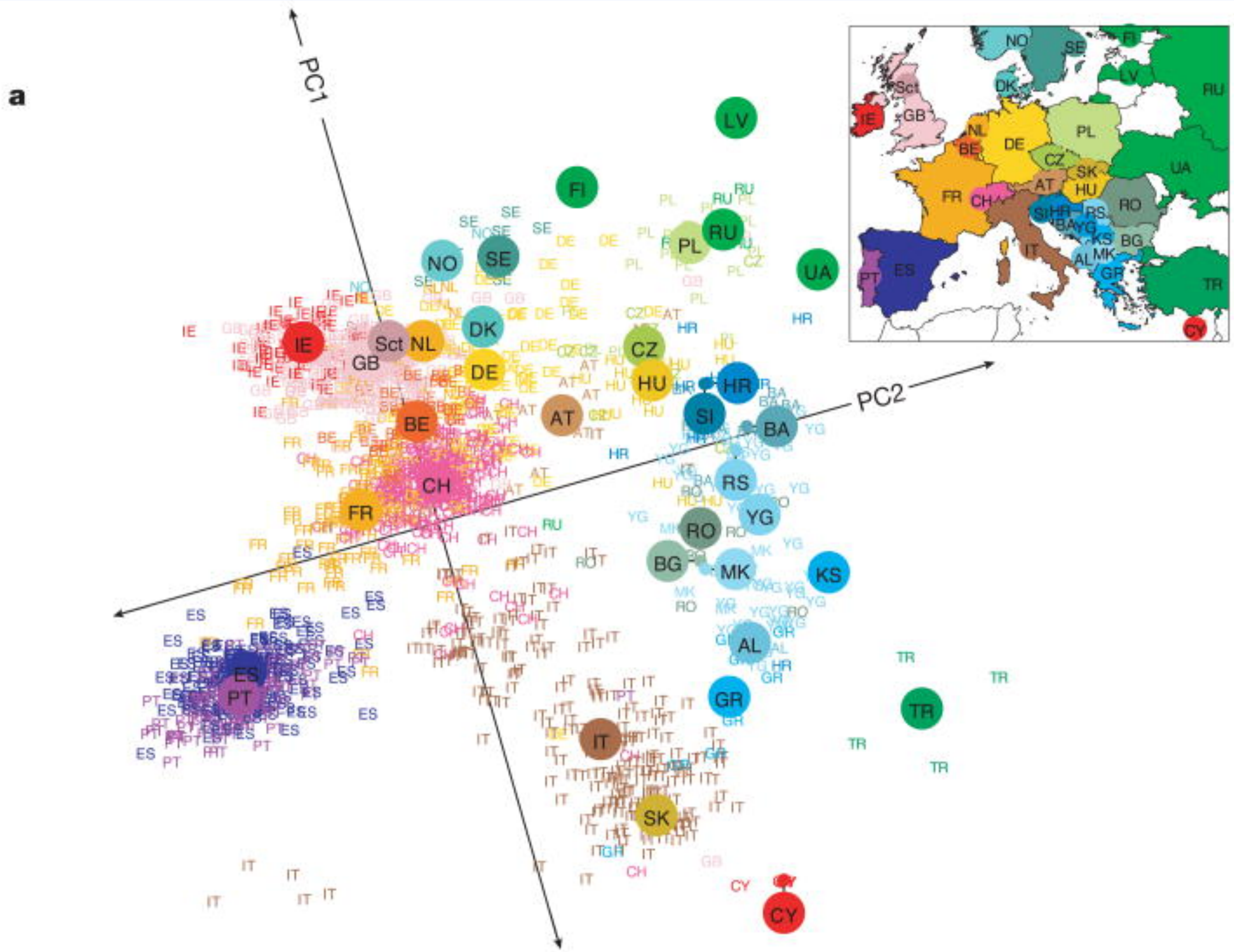


Figure 1: Community structure of political blogs (expanded set), shown using utilizing the GUESS visualization and analysis tool[2]. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

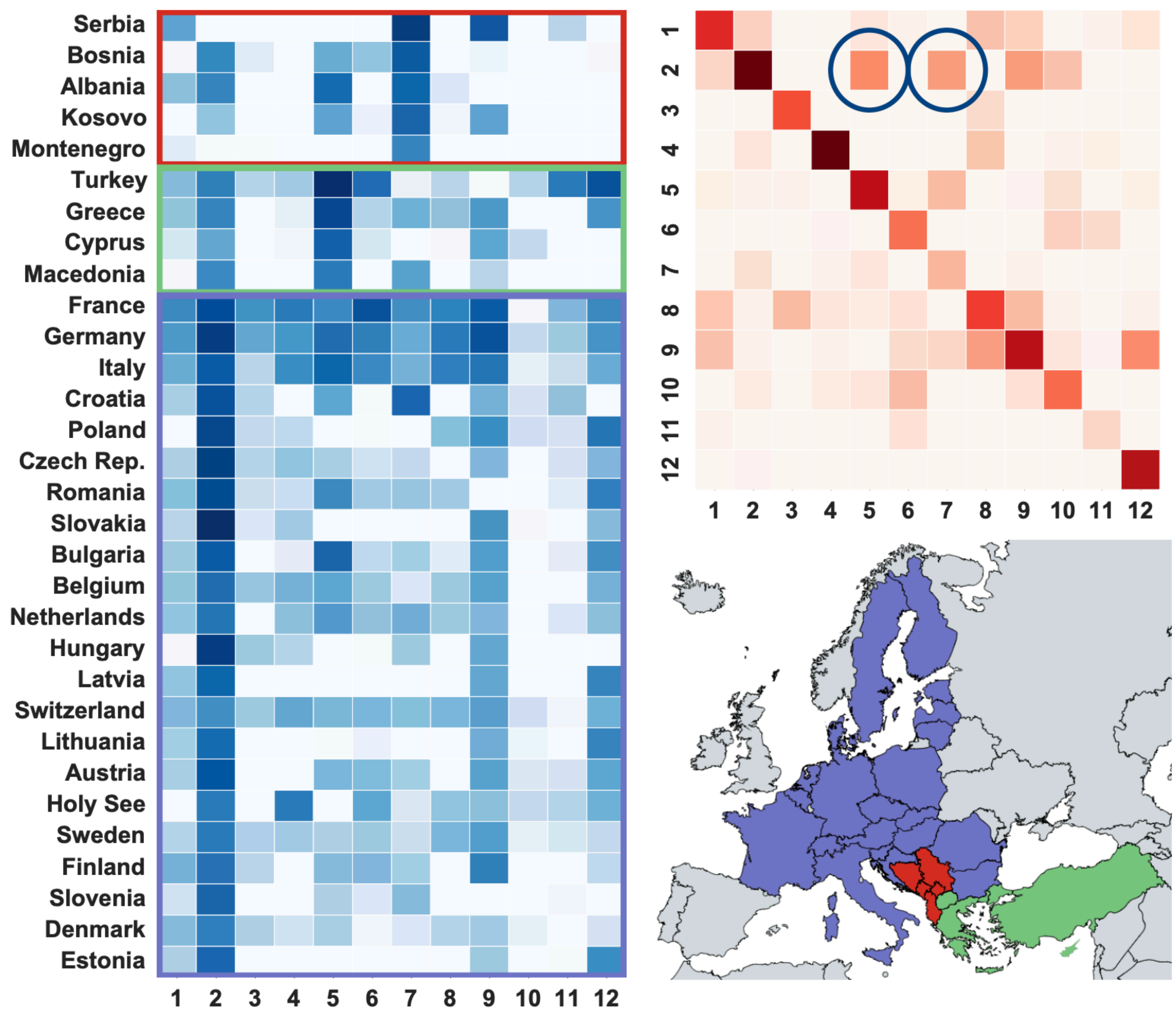
Adam and Glance (2005) "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog"

Intro: Structure in applied problems



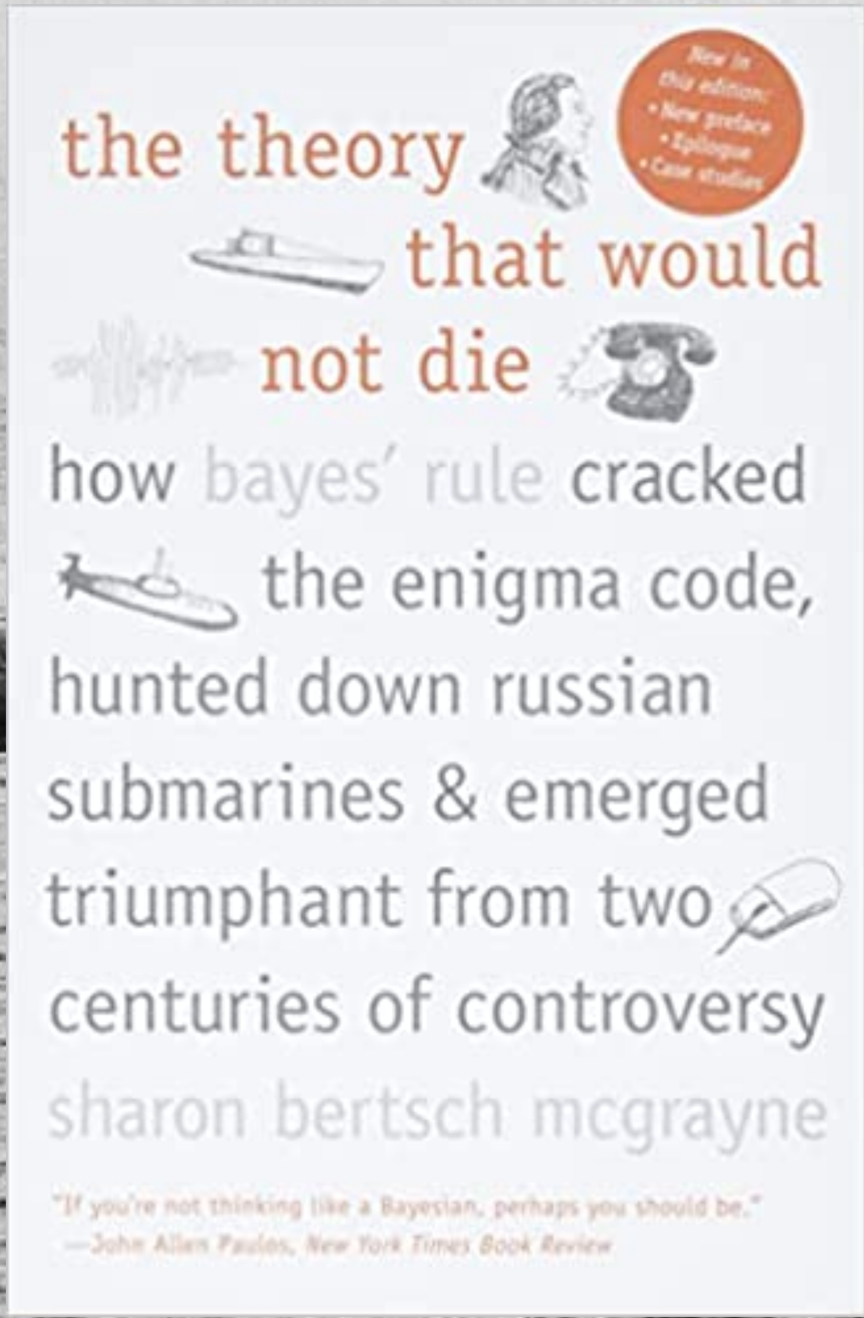
Novembre … Matthews, et al. (2008) "Genes mirror geography within Europe"

Intro: Structure in applied problems



Schein et al. (2016) "Poisson Tucker Decomposition for Learning the Structure of International Relations"

Story time: USS *Scorpion*

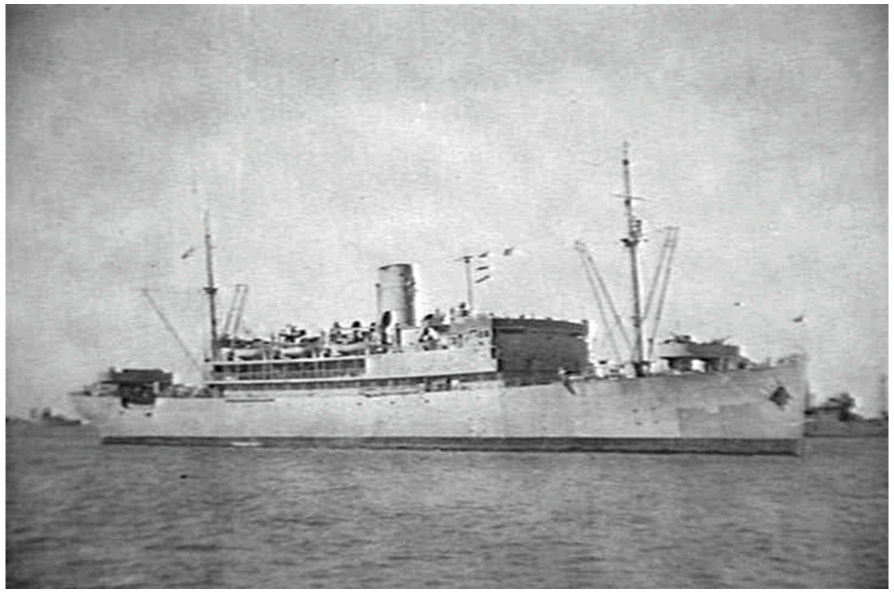


Story time: USS *Scorpion*

- May 1968 (Cold War)
- US Navy submarine went missing
- Carrying two nuclear missiles, 99 crew
- Urgent search-and-rescue operation by USS *Mizar*
- Led by statistician John Craven
- Pioneered ``Bayesian search theory"

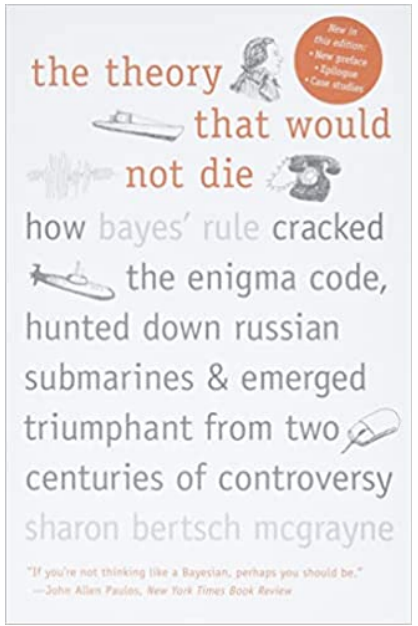
John Craven

USS *Mizar*



USS *Scorpion*

Source



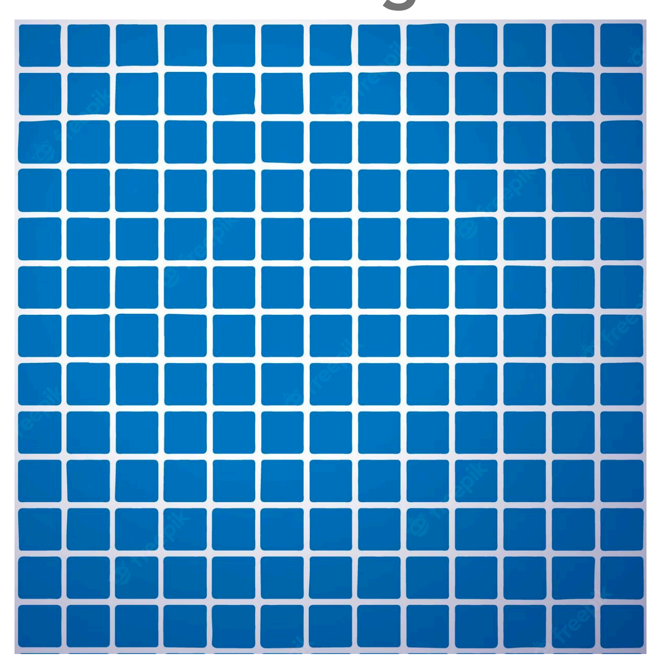
Story time: USS *Scorpion*

Bayesian search:

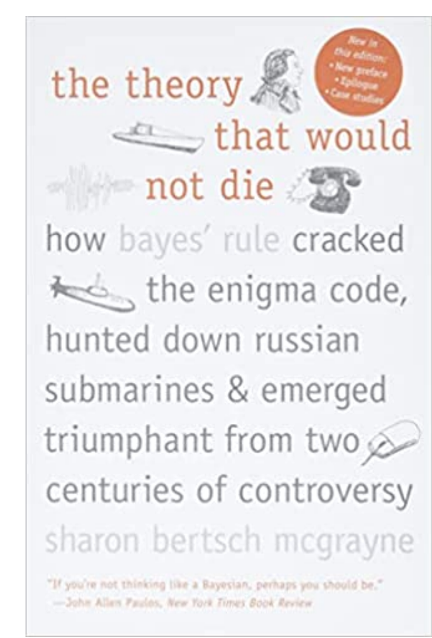
- Sonar "blips" narrowed search to 140 square mile area
- Somewhere near Azores in Atlantic
- Navy created **search grid** of 1 square mile cells
- Sonar, magnetometer, radar, camera, … data from cells
- Posterior updating of $P(\text{sub in cell } k \mid \text{data})$

Azores

Search grid



Source



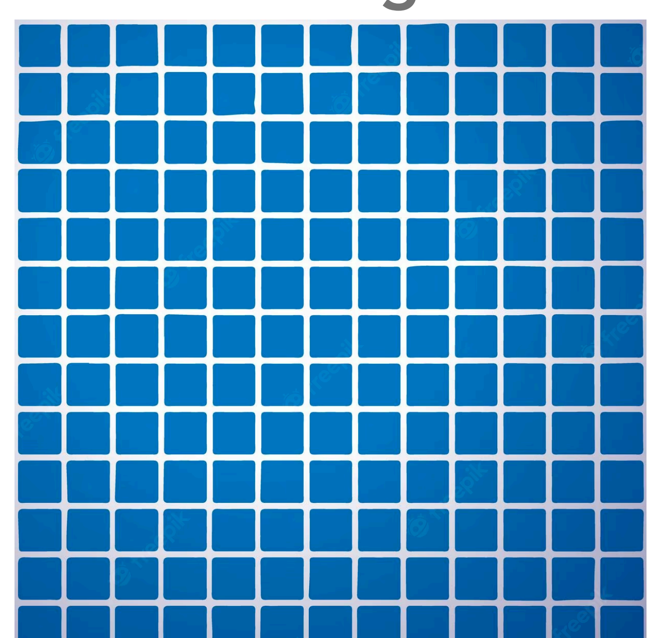
Story time: USS *Scorpion*

Bayesian **search:** Can this probability be interpreted as a frequency?

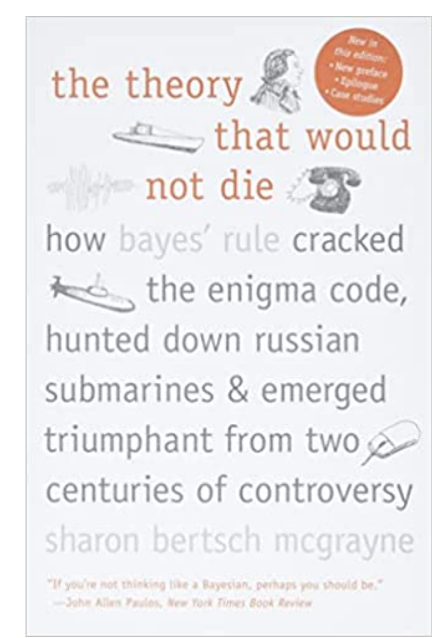
$$P(\text{sub in cell } k \mid \text{data})$$

Azores

Search grid



Source



Story time: USS *Scorpion*

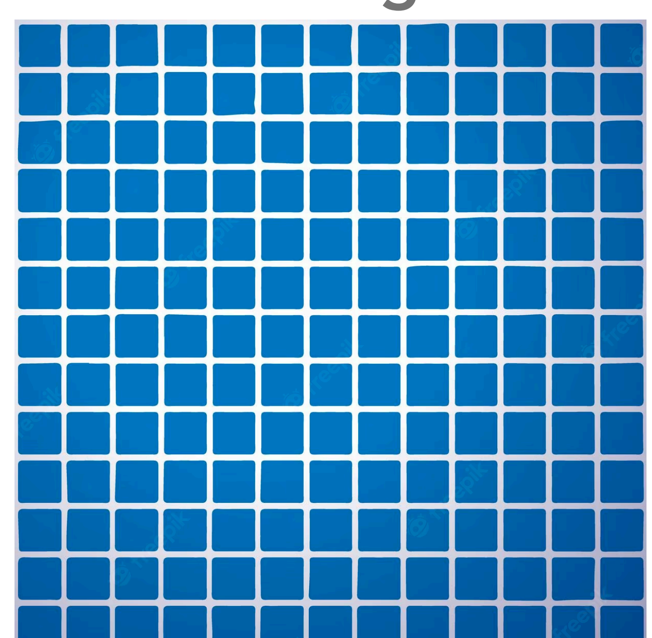
Bayesian search: Can this probability be interpreted as a frequency?

$$P(\text{sub in cell } k \mid \text{data})$$

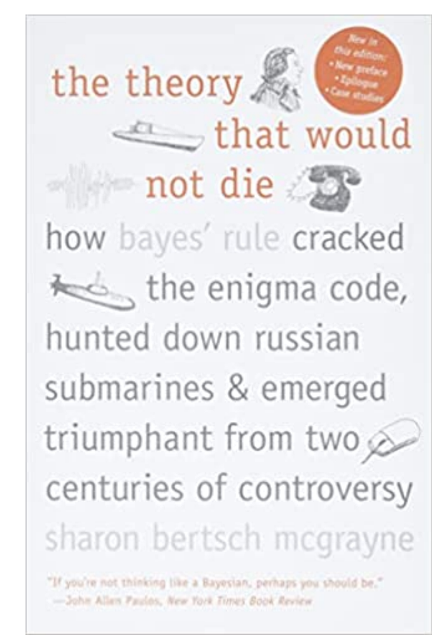
``Subjectivist probability'': degree of belief

Azores

Search grid



Source



Story time: USS *Scorpion*

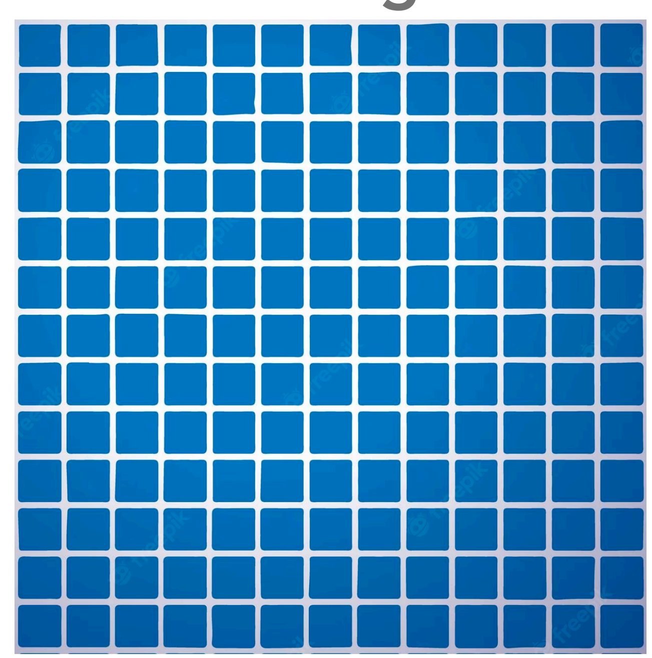
Bayesian **search:** Pioneered many facets of modern-day Bayes methods

- Structured priors (e.g., possible sub paths)
- Monte Carlo / MCMC
- Posterior updating
- Measurement error models
- Bayesian decision theory
- …

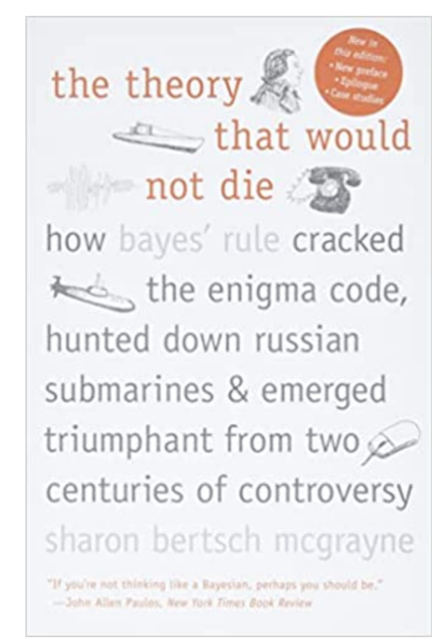
(we will explore all of these)

Azores

Search grid



Source



Story time: USS *Scorpion*

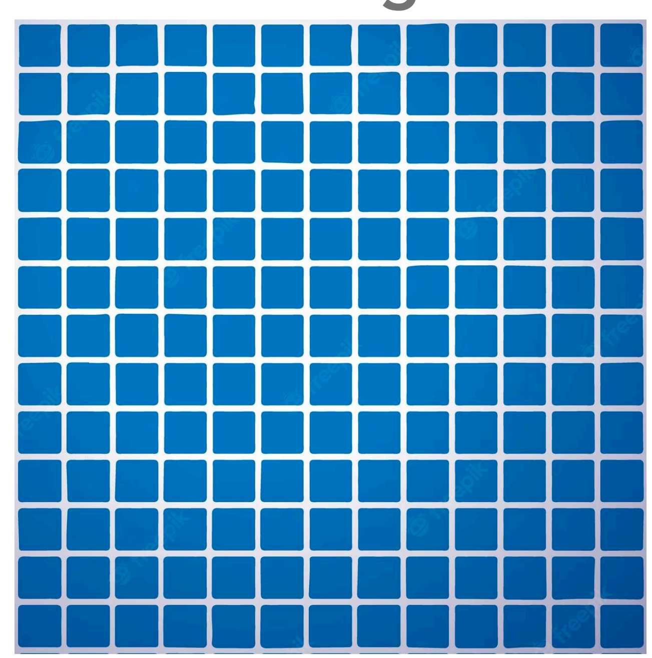
<u>Bayesian</u> search: *A natural methodology…*

integrates:
- heterogeneous data (e.g., sonar, magnetometer, radar, …)
- noisy, missing, sparse data
- expert opinion, background knowledge

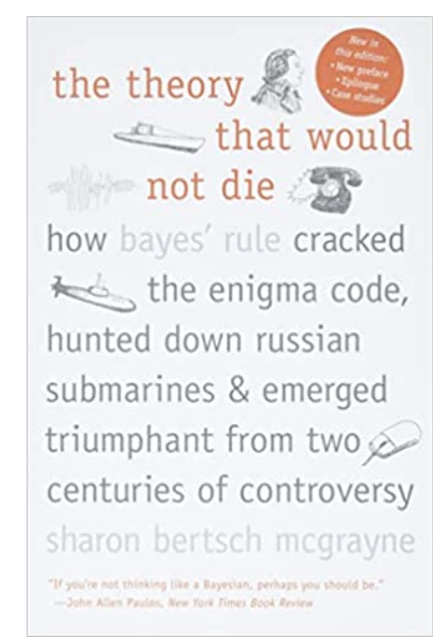
yields:
- principled decision rules
- precise uncertainty quantification

Azores

Search grid



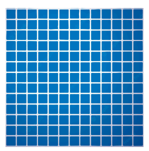
Source



This course: schedule

First half(-ish) of this course:

"The hunt for the USS Scorpion"



P(sub in cell k | data)

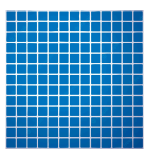
<u>Topic list (roughly by week)</u>

1. Foundations
2. Conjugate priors, Markov chains
3. Graphical models, HMMs
4. MCMC and auxiliary variables
5. Bayesian regression
6. Model checking and evaluation

This course: schedule

First half(-ish) of this course:

"The hunt for the USS Scorpion"



P(sub in cell k | data)

Topic list (roughly by week)

1. Foundations
2. Conjugate priors, Markov chains
3. Graphical models, HMMs
4. MCMC and auxiliary variables
5. Bayesian regression
6. Model checking and evaluation

Second half(ish):

More standard (unless we *love* the Scorpion)

Survey some popular models, algorithm and applications

1. Finite mixture models / EM
2. Infinite mixture models / MCMC
3. Admixture models / Gibbs
4. Matrix factorization / MAP
5. Tensor factorization / VI
6. Other topics (some flexibility)

This course: goals and expectations

Goals:

1. Internalize conceptual and mathematical foundations
2. Develop modeling intuition, *creativity* and taste
3. Roadmap to the larger space of methods
4. Exposure to python™

Expectations:

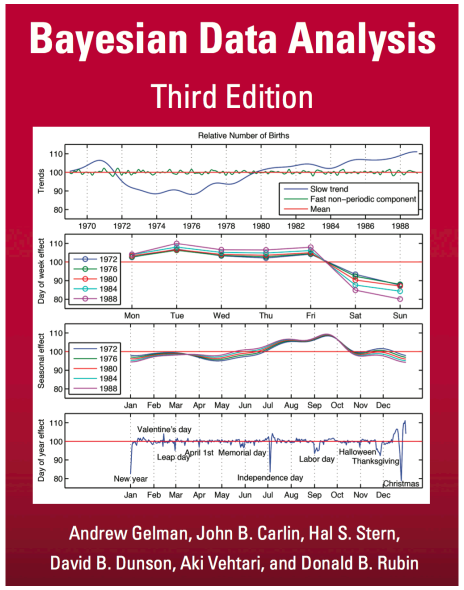
1. *Readings*: Posted after lecture
2. *Attendance*: Come to class (if able); I will often be at the board

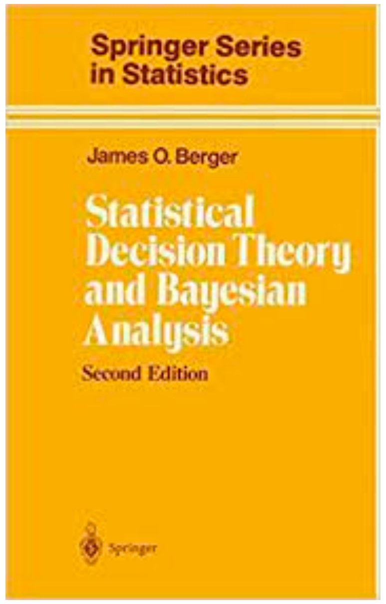
Grades:

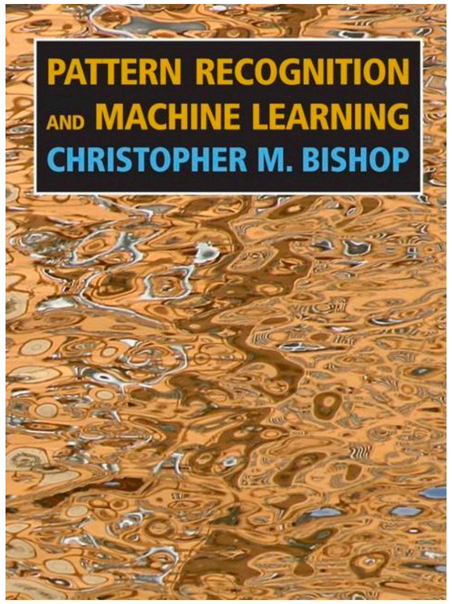
1. *Homeworks (65%)*: ~1/week (released Friday), lowest dropped
2. *Midterm / Final (both 15%)*
3. *Participation (5%)*: scribing, in-class input, out-class contributions

This course: resources

I will draw on a variety of books / papers, but often these:







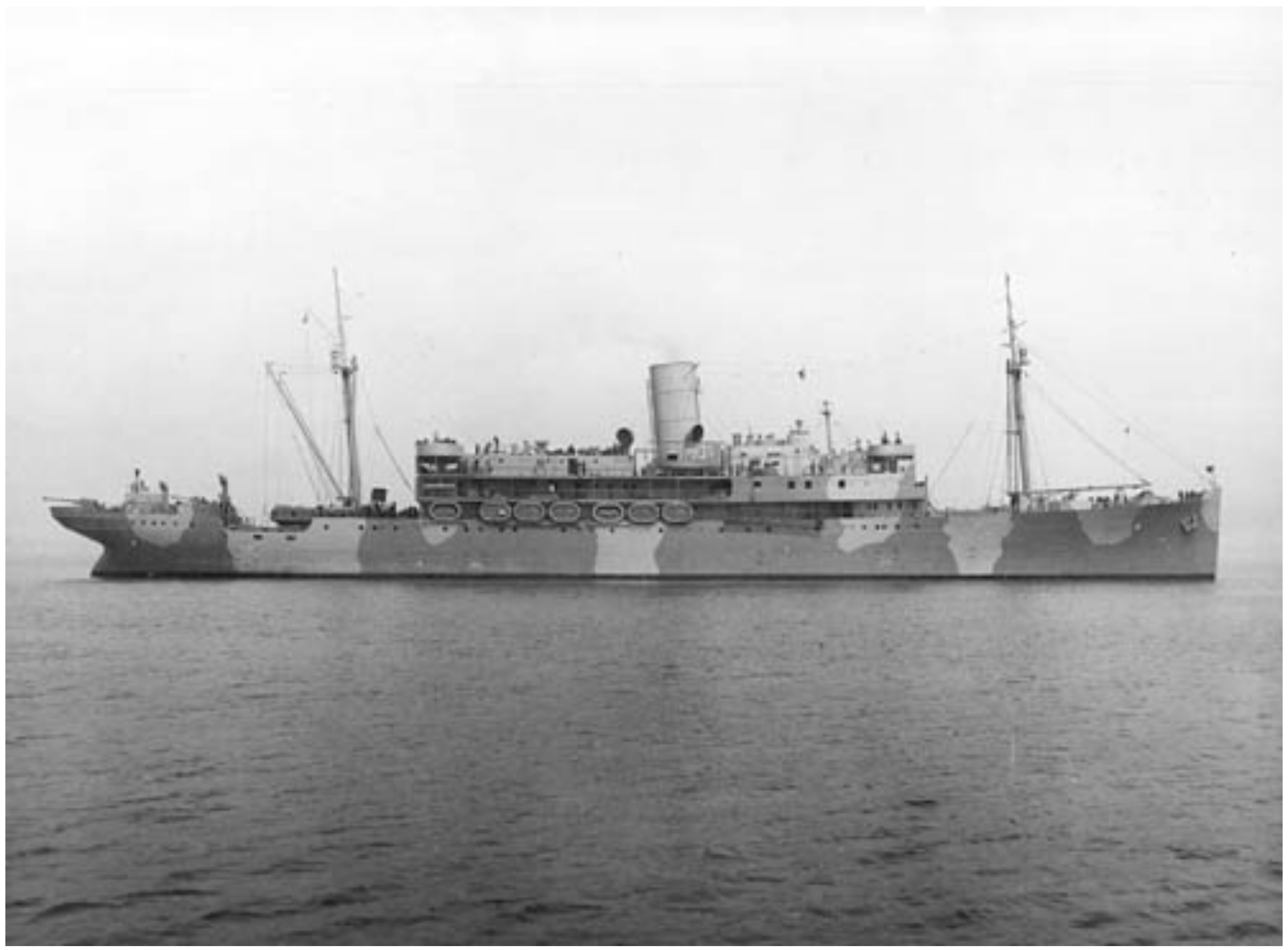
Will post papers (and homework) to Canvas:

https://canvas.uchicago.edu/courses/48880

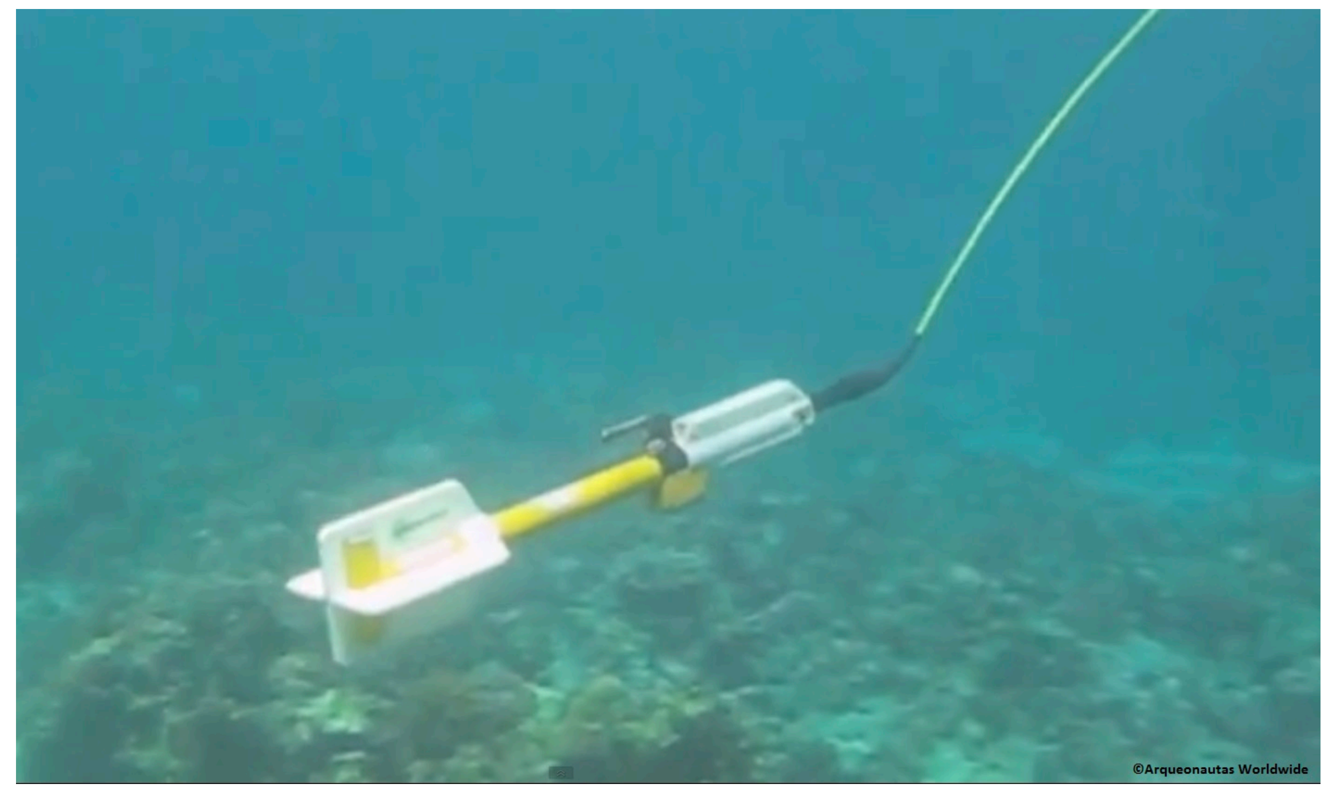
Also see Matthew Stephens' resources from 2021 version:

https://tinyurl.com/yc8c4ebc

Magnetometer data



USS *Mizar*



Magnetometer

- *Mizar* dragged a magnetometer around the search area
- A magnetometer measures magnetic field
- Produces big numbers when close to large metal objects (like subs)

Scenario: Magnetometer data

Scenario: Magnetometer data

- We are the statisticians onboard the *Mizar*
- The *Mizar* sweeps cell k and obtains magnetometer data…

```
[0.12149101 1.74388112 2.45639549 ... 0.3103083  0.10229448 4.65486283]
```

- Do the data suggest the *Scorpion* is in cell k?
- Should we deploy the deep-sea divers?
- We need a *model*…

Building a model: likelihood

likelihood *how (we believe) the data would look if…*

$P(Y \mid Z = 1)$ *the Scorpion is there*

$P(Y \mid Z = 0)$ *the Scorpion is not there*

Building a model: likelihood

likelihood *how (we believe) the data would look if…*

$P(Y \mid Z = 1)$ *the Scorpion is there*

$P(Y \mid Z = 0)$ *the Scorpion is not there*

likelihood odds *has a ``human scale''*

$$\frac{P(Y \mid Z = 1)}{P(Y \mid Z = 0)}$$

e.g., 5 means *data favor Z=1 by 5-to-1*

Building a model: likelihood

likelihood *how (we believe) the data would look if…*

$P(Y \mid Z = 1)$ *the Scorpion is there*

$P(Y \mid Z = 0)$ *the Scorpion is not there*

But how do we build this? Especially $P(Y \mid Z = 1)$…?

likelihood odds *has a "human scale"*

$$\frac{P(Y \mid Z = 1)}{P(Y \mid Z = 0)}$$

e.g., 5 means *data favor Z=1 by 5-to-1*

Building a model: likelihood

likelihood *how (we believe) the data would look if…*

$P(Y \mid Z = 1)$ *the Scorpion is there*

$P(Y \mid Z = 0)$ *the Scorpion is not there*

But how do we build this? Especially $P(Y \mid Z = 1)$…?

Use whatever we know!

likelihood odds *has a "human scale"*

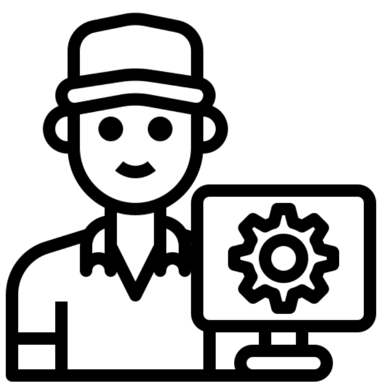
$$\frac{P(Y \mid Z = 1)}{P(Y \mid Z = 0)}$$

e.g., 5 means *data favor Z=1 by 5-to-1*

Scenario: Magnetometer data

- We chat with the technician onboard to learn more

- First, she informs us that the magnetometer always **scrambles the order** of data for operational security



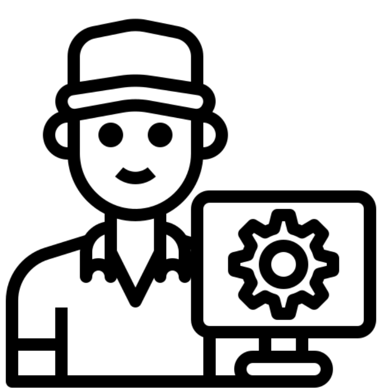
Technician

Scenario: Magnetometer data

- We chat with the technician onboard to learn more

- First, she informs us that the magnetometer always **scrambles the order** of data for operational security

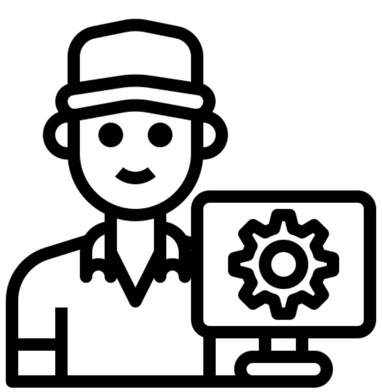
- *We* don't know where the *Mizar* was within the cell when it took each reading



Technician

Scenario: Magnetometer data

- We chat with the technician onboard to learn more
- First, she informs us that the magnetometer always **scrambles the order** of data for operational security
- *We* don't know where the *Mizar* was within the cell when it took each reading



Technician

- To us, the data are **exchangeable**…

Exchangeable data

A sequence of random variables is *infinitely exchangeable* if for any N, their joint probability is invariant to permutation:

$$P(y_1, \ldots, y_N) = P(y_{\Delta(1)}, \ldots, y_{\Delta(N)})$$

Exchangeable data

A sequence of random variables is *infinitely exchangeable* if for any *N*, their joint probability is invariant to permutation:

$$P(y_1, \ldots, y_N) = P(y_{\Delta(1)}, \ldots, y_{\Delta(N)})$$

De Finetti's Theorem: The sequence is infinitely exchangeable ***IFF***:

$$P(y_1, \ldots, y_N) = \int \left[\prod_{i=1}^{N} P(y_i \mid \theta) \right] dP(\theta)$$

Exchangeable data

A sequence of random variables is *infinitely exchangeable* if for any *N*, their joint probability is invariant to permutation:

$$P(y_1, \ldots, y_N) = P(y_{\Delta(1)}, \ldots, y_{\Delta(N)})$$

De Finetti's Theorem: The sequence is infinitely exchangeable ***IFF*:**

$$P(y_1, \ldots, y_N) = \int \left[\prod_{i=1}^{N} P(y_i \mid \theta) \right] dP(\theta)$$

This is very paraphrased (see the readings), but the moral to many Bayesians has been: **conditionally independent** models for exchangeable data.

Scenario: Magnetometer data

- Inspired by de Finetti, we will build a likelihood that *factorizes:*

$$\underbrace{P(y_1, \ldots, y_N \mid Z)}^{\textbf{likelihood}} = \underbrace{\prod_{i=1}^{N} P(y_i \mid Z)}_{\textbf{conditionally independent}}$$

Scenario: Magnetometer data

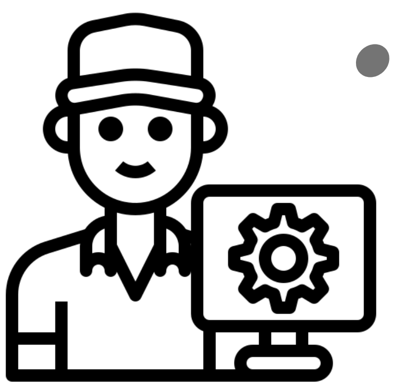
- Inspired by de Finetti, we will build a likelihood that *factorizes:*

likelihood

$$P(y_1, \ldots, y_N \mid Z) = \prod_{i=1}^{N} P(y_i \mid Z)$$

conditionally independent

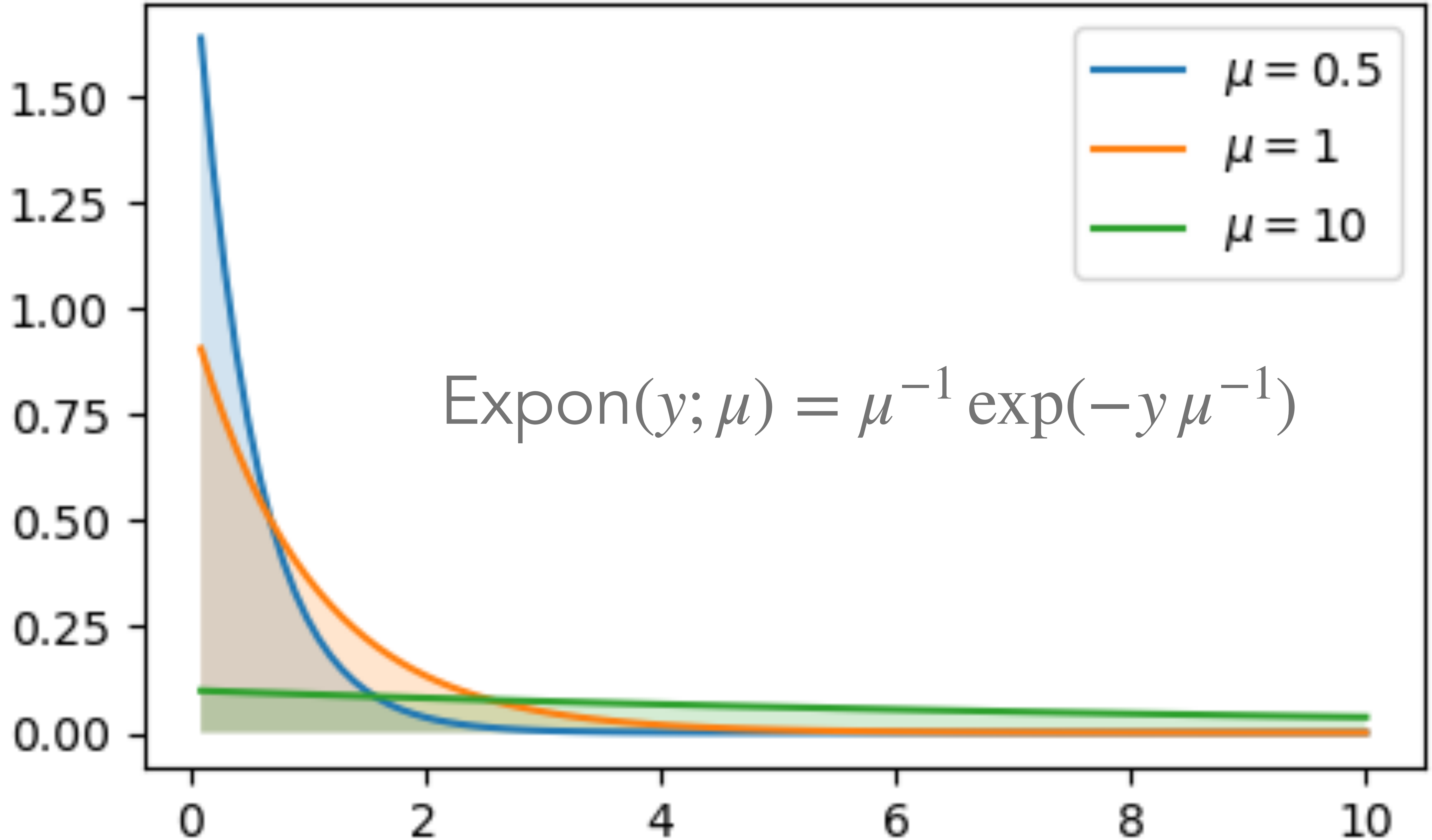
- The technician informs us of another useful fact…



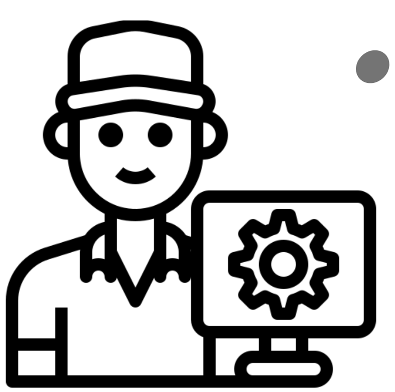
Technician

- In tests, where distance to a metal object is known:

$$y_i \sim \mathrm{Expon}(\alpha_0 + \alpha_1\,\mathrm{distance}^{-3})$$



$$\text{Expon}(y; \mu) = \mu^{-1} \exp(-y\,\mu^{-1})$$



Technician

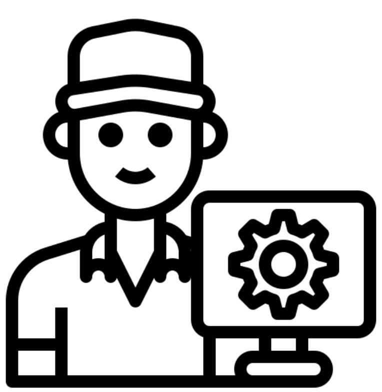
- In tests, where distance to a metal object is known:

$$y_i \sim \text{Expon}(\alpha_0 + \alpha_1\,\text{distance}^{-3})$$

Scenario: Magnetometer data

- With the technician's help, we know (where $D \equiv$ distance)

$$P(y_i \mid Z = 1, D) = \mathrm{Expon}(y_i;\, \alpha_0 + \alpha_1 D^{-3})$$



Technician

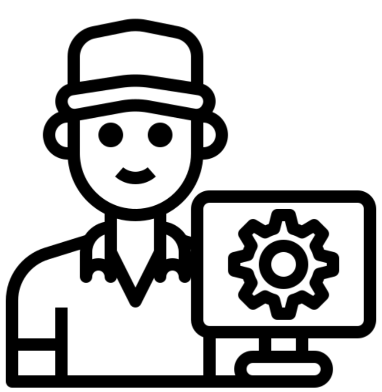
Scenario: Magnetometer data

- With the technician's help, we know (where $D \equiv$ distance)

$$P(y_i \mid Z = 1, D) = \mathrm{Expon}(y_i;\ \alpha_0 + \alpha_1 D^{-3})$$

- And by extension, if there is no metal object nearby

$$P(y_i \mid Z = 0) = \mathrm{Expon}(y_i;\ \alpha_0)$$



Technician

Scenario: Magnetometer data

- With the technician's help, we know (where $D \equiv$ distance)

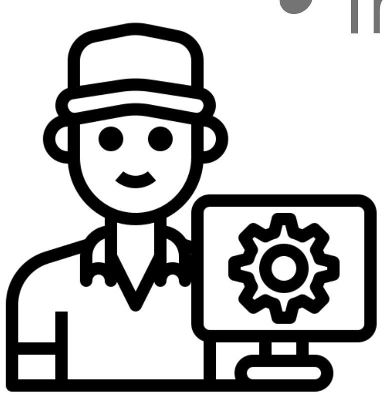
$$P(y_i \mid Z = 1, D) = \text{Expon}(y_i;\, \alpha_0 + \alpha_1 D^{-3})$$

- And by extension, if there is no metal object nearby

$$P(y_i \mid Z = 0) = \text{Expon}(y_i;\, \alpha_0)$$

- In practice, we don't know *D*—it is a **nuisance variable**
    - What we want:

$$P(y_i \mid Z = 1) = \mathbb{E}_D[P(y_i \mid Z = 1, D)]$$



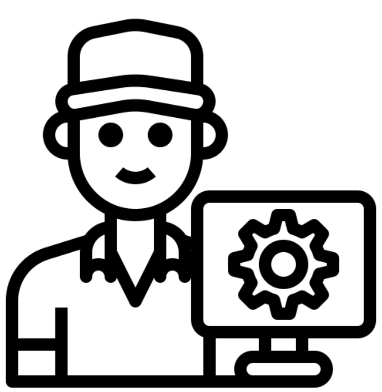
Technician

Scenario: Magnetometer data

$$\underbrace{P(y_i \mid Z = 1)}_{\text{marginal likelihood}} = \mathbb{E}_D[P(y_i \mid Z = 1, D)]$$

$$= \int \underbrace{P(y_i \mid Z = 1, D)}_{\text{complete data likelihood}} \; \underbrace{P(D \mid Z = 1)}_{\text{prior (for } D\text{)}} \, dD$$

- The likelihood we want *marginalizes out* the nuisance variable



Technician

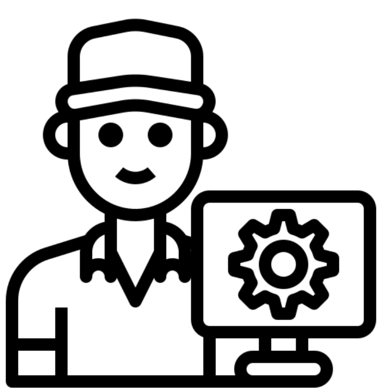
Scenario: Magnetometer data

$$P(y_i \mid Z = 1) = \mathbb{E}_D[P(y_i \mid Z = 1, D)]$$

marginal likelihood

$$= \int \underbrace{P(y_i \mid Z = 1, D)}_{\textbf{complete data likelihood}} \; \underbrace{P(D \mid Z = 1)}_{\textbf{prior (for } D\textbf{)}} \, dD$$

- The likelihood we want *marginalizes out* the nuisance variable
  - This involves specifying a prior distribution for *D*
    - If the *Scorpion* were in cell *k*, how far might it have been from the *Mizar's* magnetometer at any given time?



Technician

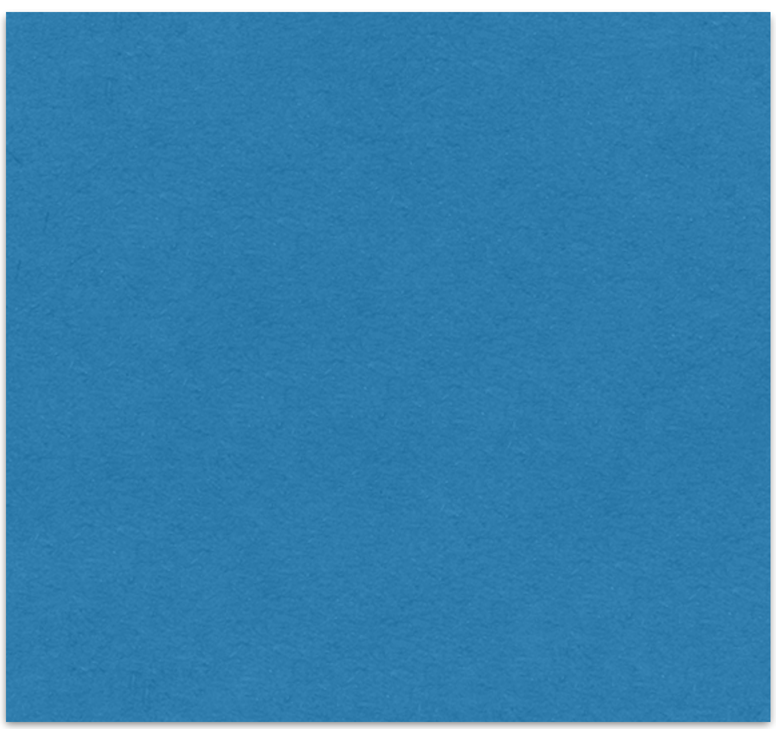
Scenario: Magnetometer data

$$P(D \mid Z = 1) = ?$$

prior (for *D*)

"Principle of insufficient reason"

—Laplace, Bernoulli, Keynes, …



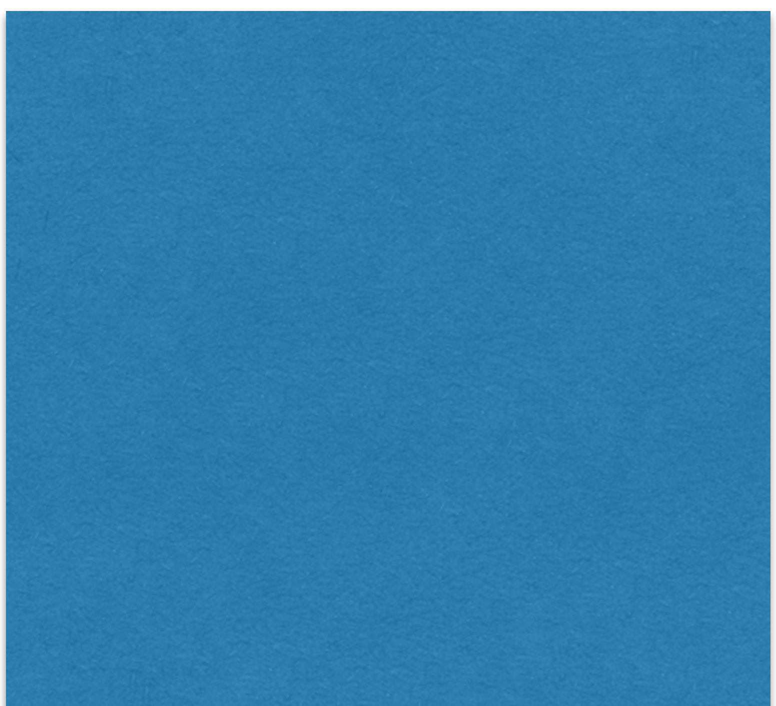
cell *k*

- Recall the *Mizar* scrambled its data
- All we know: *Mizar* was **somewhere** in cell
- Also don't know where *Scorpion* was
- **Principle:** *all locations equally probable*

Scenario: Magnetometer data

$$P(D \mid Z = 1) =$$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)



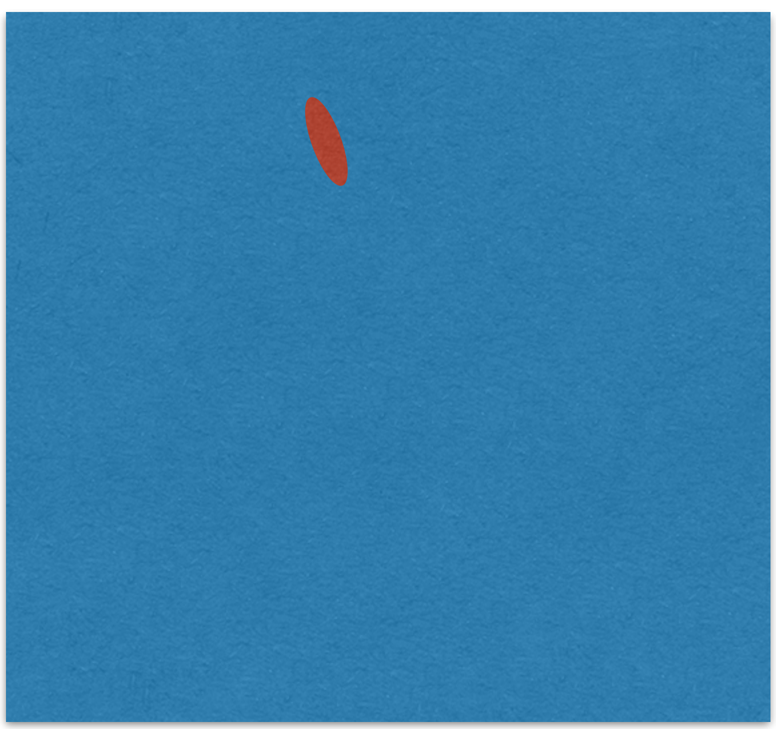
cell *k*

- Recall the *Mizar* scrambled its data
- All we know: *Mizar* was **somewhere** in cell
- Also don't know where *Scorpion* was
- **Principle:** *all locations equally probable*

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)



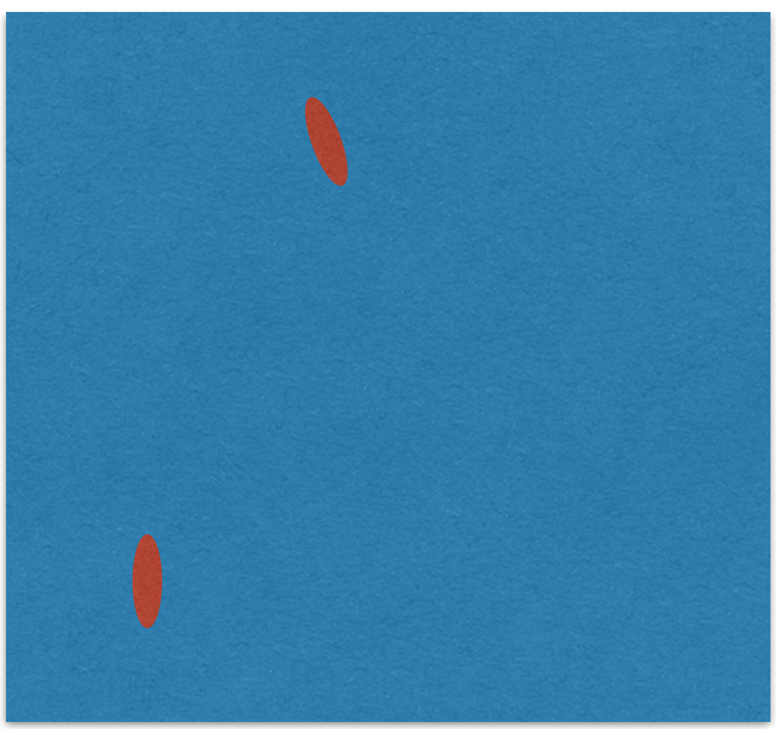
cell *k*

- Recall the *Mizar* scrambled its data
- All we know: *Mizar* was **somewhere** in cell
- Also don't know where *Scorpion* was
- **Principle:** *all locations equally probable*

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)



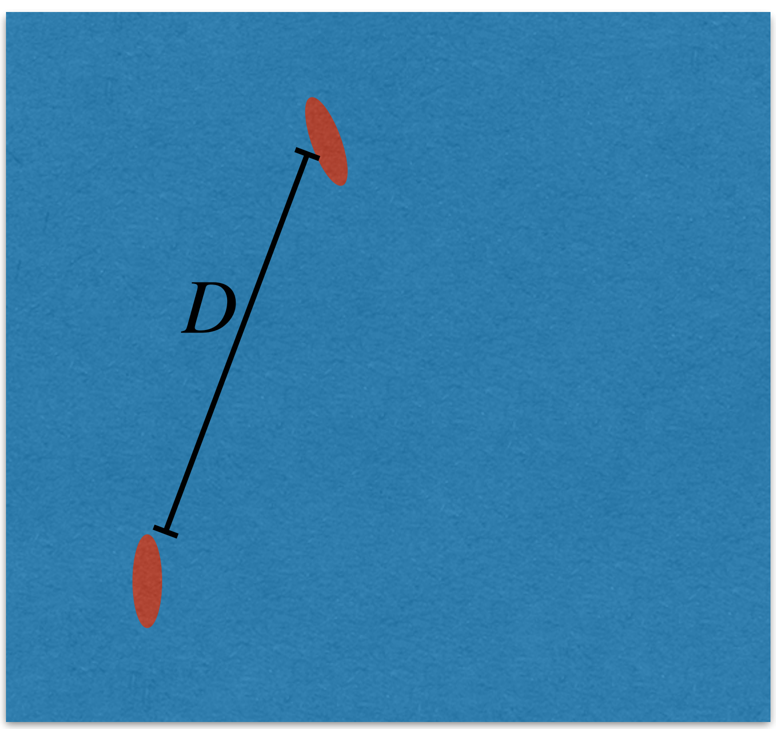
cell *k*

- Recall the *Mizar* scrambled its data
- All we know: *Mizar* was **somewhere** in cell
- Also don't know where *Scorpion* was
- **Principle:** *all locations equally probable*

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)



cell *k*

- Recall the *Mizar* scrambled its data
- All we know: *Mizar* was **somewhere** in cell
- Also don't know where *Scorpion* was
- **Principle:** *all locations equally probable*

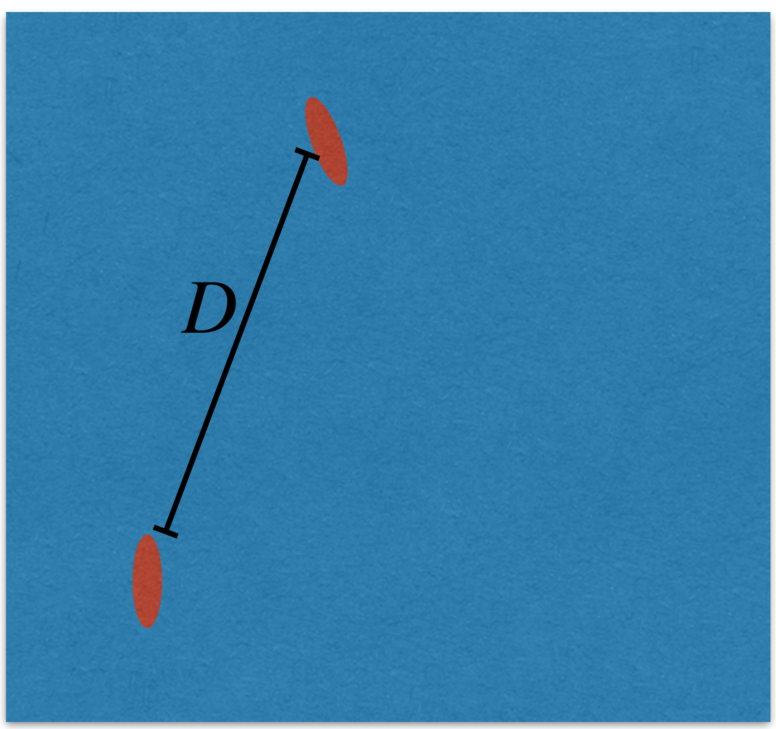
Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \mathrm{dist}(p_1, p_2)\right)\right]$$

- This is itself a marginal over nuisance variables



cell *k*

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** D between 2 points placed uniformly at random **)**

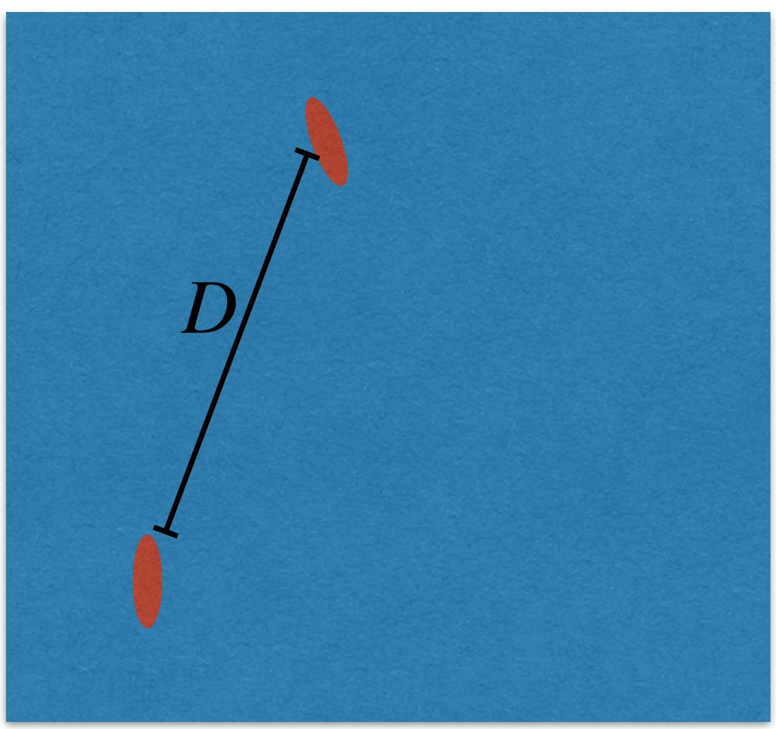
prior (for *D*)

$$= \mathbb{E}_{p_1, p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$

- This is itself a marginal over nuisance variables

$$= \int_{\square}\int_{\square} \delta\left(D = \text{dist}(p_1, p_2)\right) dp_1\, dp_2$$

- We could try to solve this analytically…



cell *k*

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \mathrm{dist}(p_1, p_2)\right)\right]$$

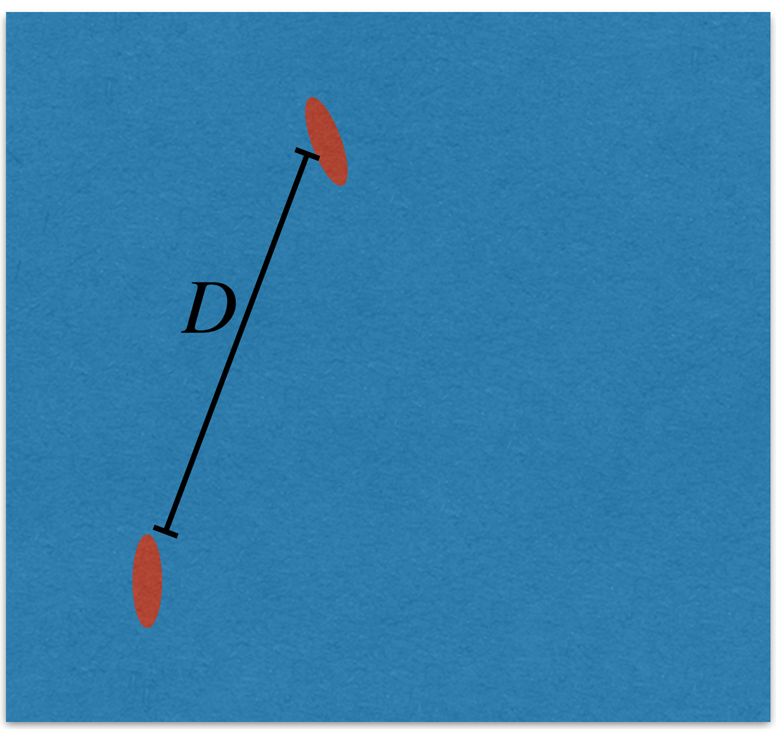
- This is itself a marginal over nuisance variables

$$= \int_{\square}\int_{\square} \delta\left(D = \mathrm{dist}(p_1, p_2)\right) dp_1\, dp_2$$

- We could try to solve this analytically…
- However, recall we ultimately want…

$$P(y \mid Z = 1) = \int P(y \mid Z = 1, D)\, P(D \mid Z = 1) dD$$

- Looks hard (*extra credit* if you solve it)



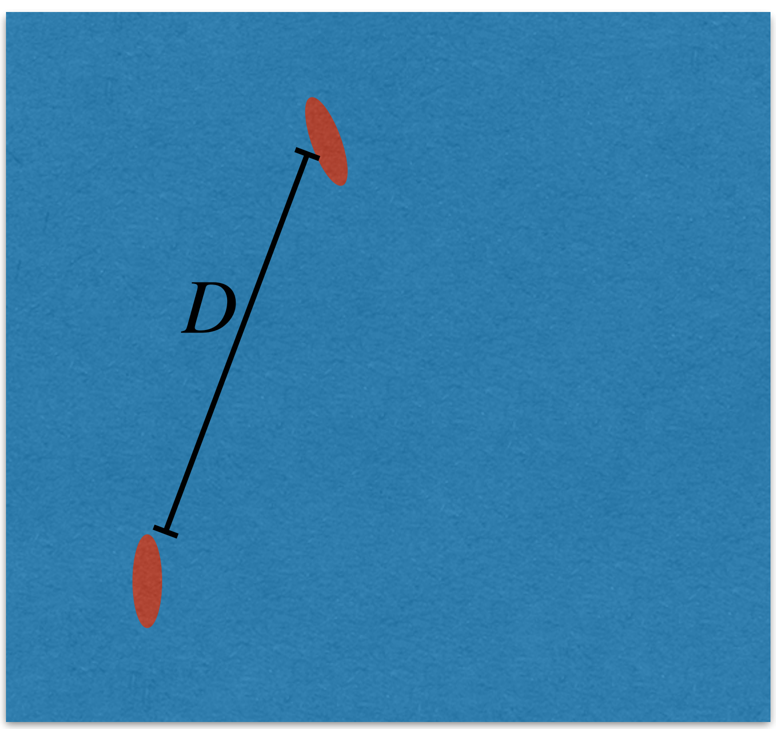
cell *k*

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** D between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \mathrm{dist}(p_1, p_2)\right)\right]$$



cell *k*

- **Monte Carlo** simulation:

for s in num_samples:

$$p_1^{(s)} \sim \mathrm{Uniform}(\square)$$

$$p_2^{(s)} \sim \mathrm{Uniform}(\square)$$

$$D_s = \mathrm{dist}(p_1^{(s)}, p_2^{(s)})$$

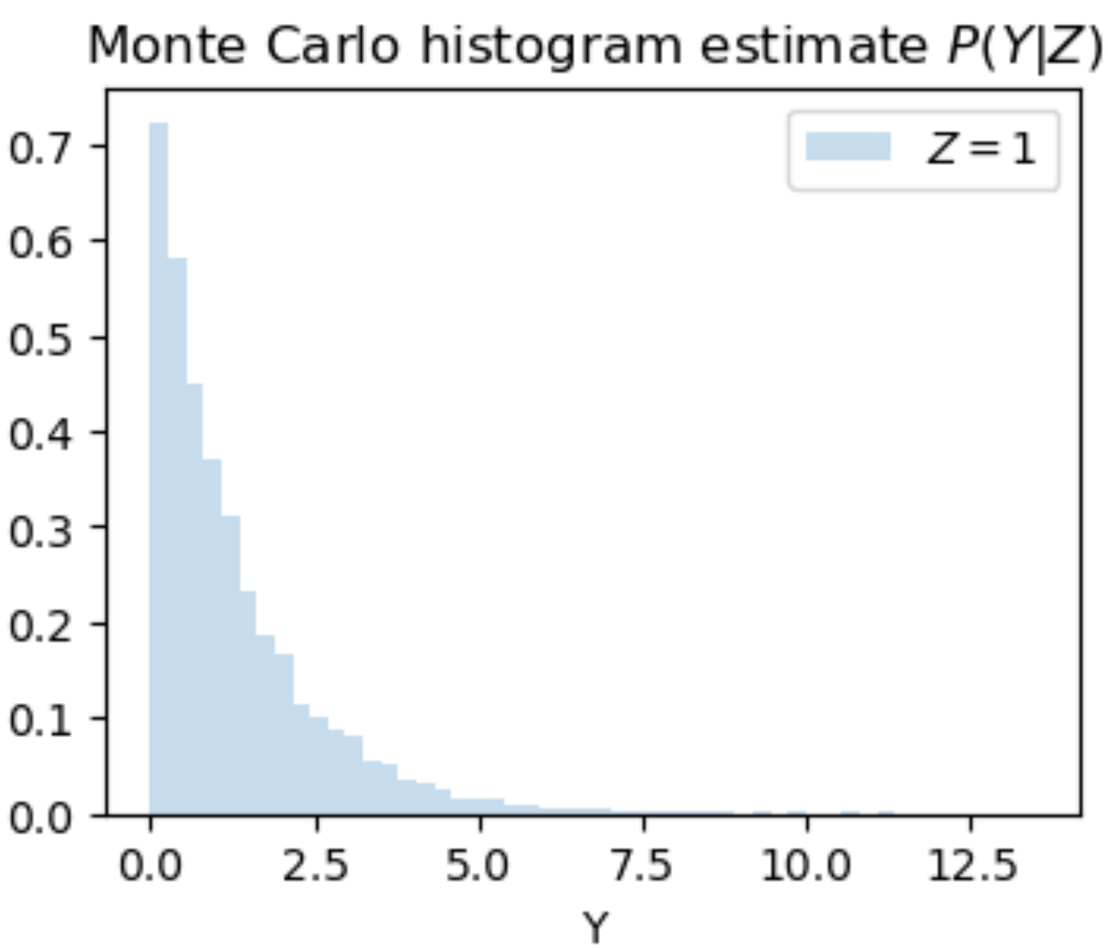
$$y^{(s)} \sim \mathrm{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

Scenario: Magnetometer data

$$P(D \mid Z = 1) =$$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1, p_2}\left[\delta\left(D = \mathrm{dist}(p_1, p_2)\right)\right]$$



- **Monte Carlo** simulation:

for s in num_samples:

$$p_1^{(s)} \sim \mathrm{Uniform}(\square)$$

$$p_2^{(s)} \sim \mathrm{Uniform}(\square)$$

$$D_s = \mathrm{dist}(p_1^{(s)}, p_2^{(s)})$$

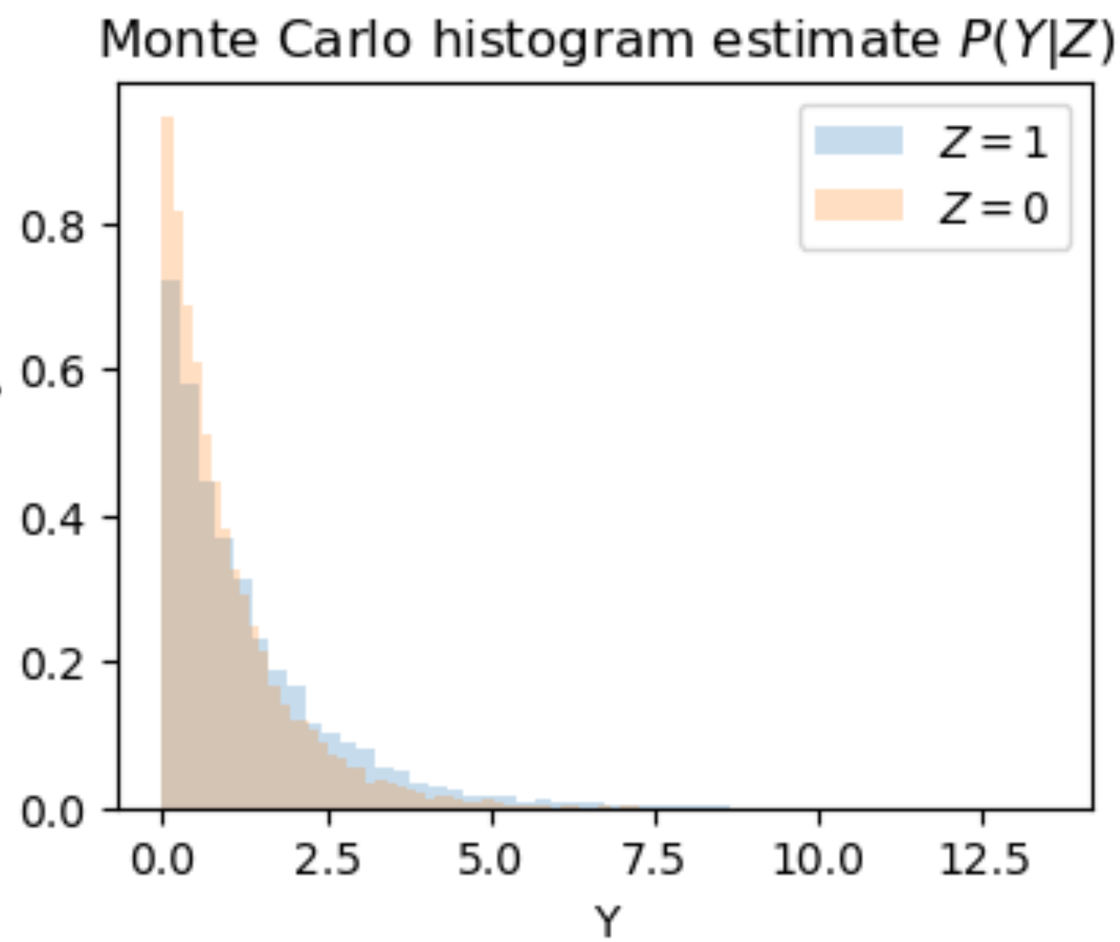
$$y^{(s)} \sim \mathrm{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** D between 2 points placed uniformly at random **)**

prior (for D)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \mathrm{dist}(p_1, p_2)\right)\right]$$



- **Monte Carlo** simulation:

for s in num_samples:

$$p_1^{(s)} \sim \mathrm{Uniform}(\square)$$

$$p_2^{(s)} \sim \mathrm{Uniform}(\square)$$

$$D_s = \mathrm{dist}(p_1^{(s)}, p_2^{(s)})$$

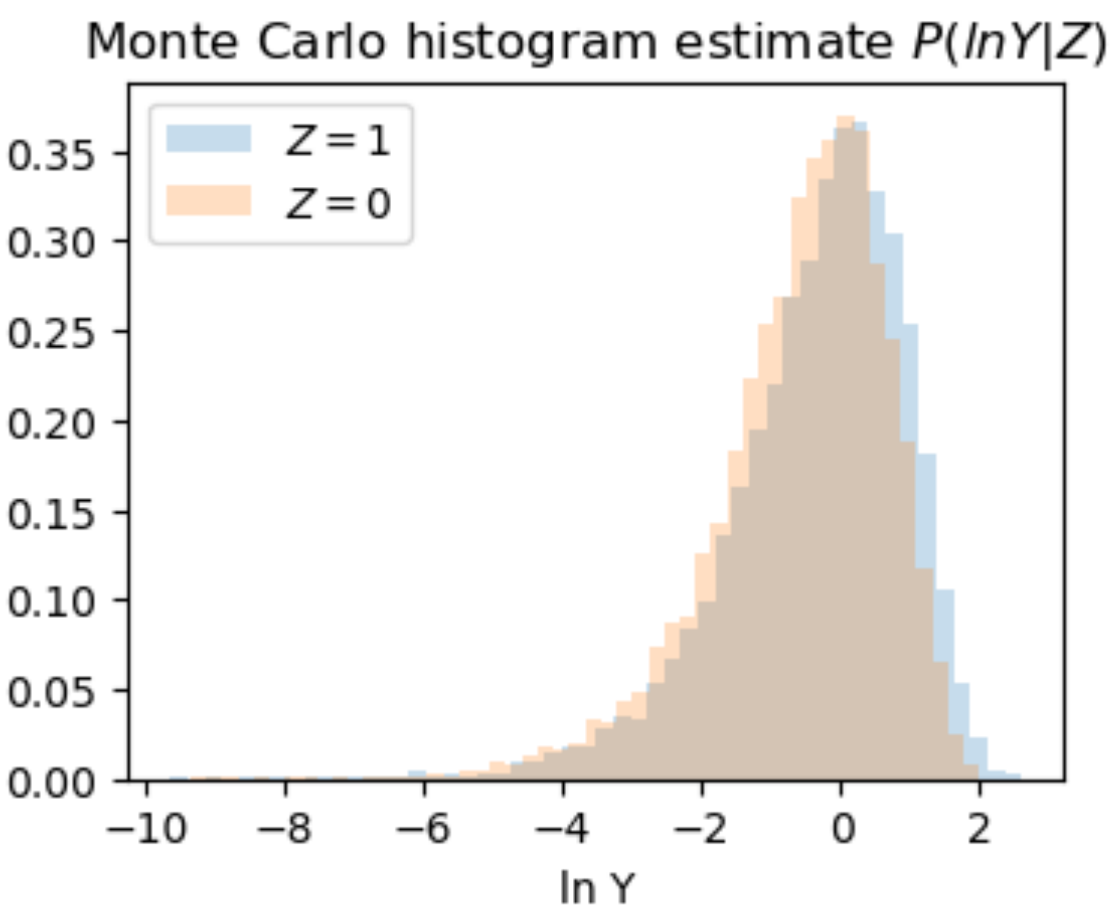
$$y^{(s)} \sim \mathrm{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$



- **Monte Carlo** simulation:

for s in num_samples:

$$p_1^{(s)} \sim \text{Uniform}(\square)$$

$$p_2^{(s)} \sim \text{Uniform}(\square)$$

$$D_s = \text{dist}(p_1^{(s)}, p_2^{(s)})$$

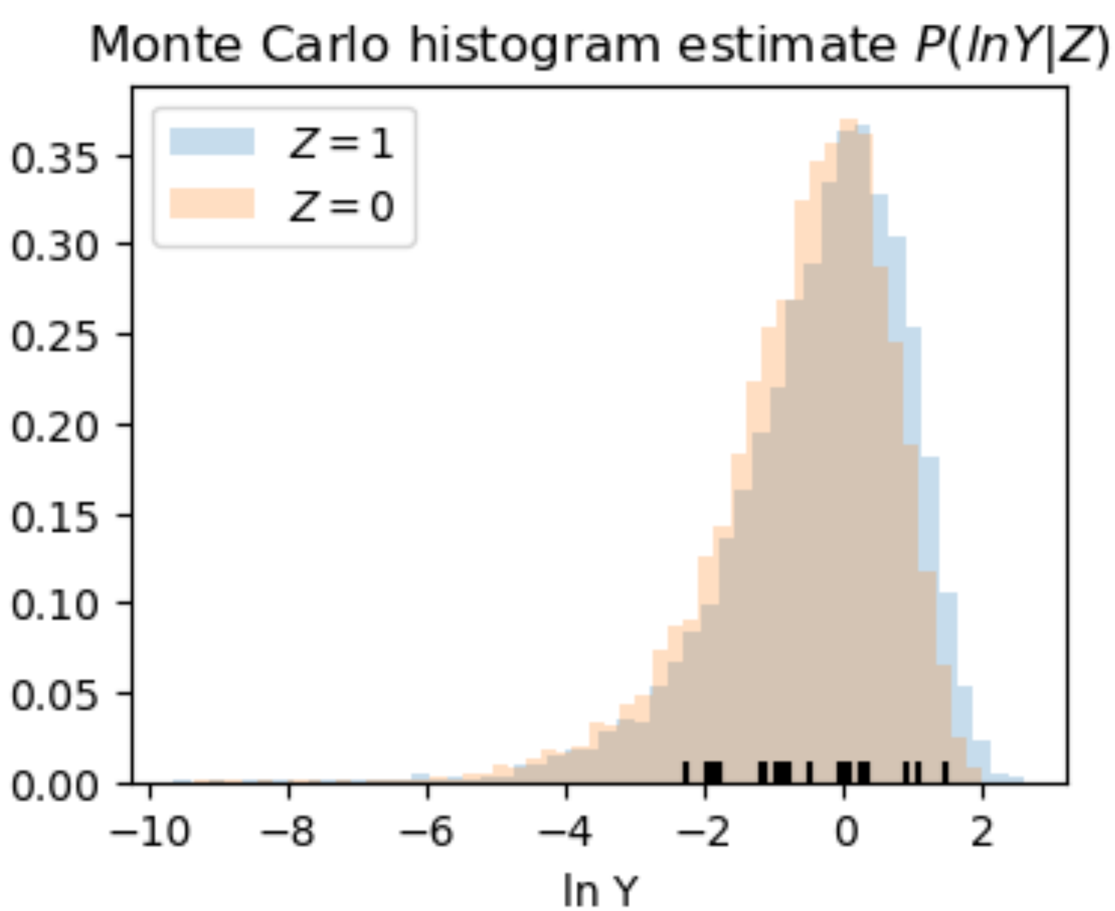
$$y^{(s)} \sim \text{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** D between 2 points placed uniformly at random **)**

prior (for D)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$



- **Monte Carlo** simulation:

for s in num_samples:

$$p_1^{(s)} \sim \text{Uniform}(\square)$$

$$p_2^{(s)} \sim \text{Uniform}(\square)$$

$$D_s = \text{dist}(p_1^{(s)}, p_2^{(s)})$$

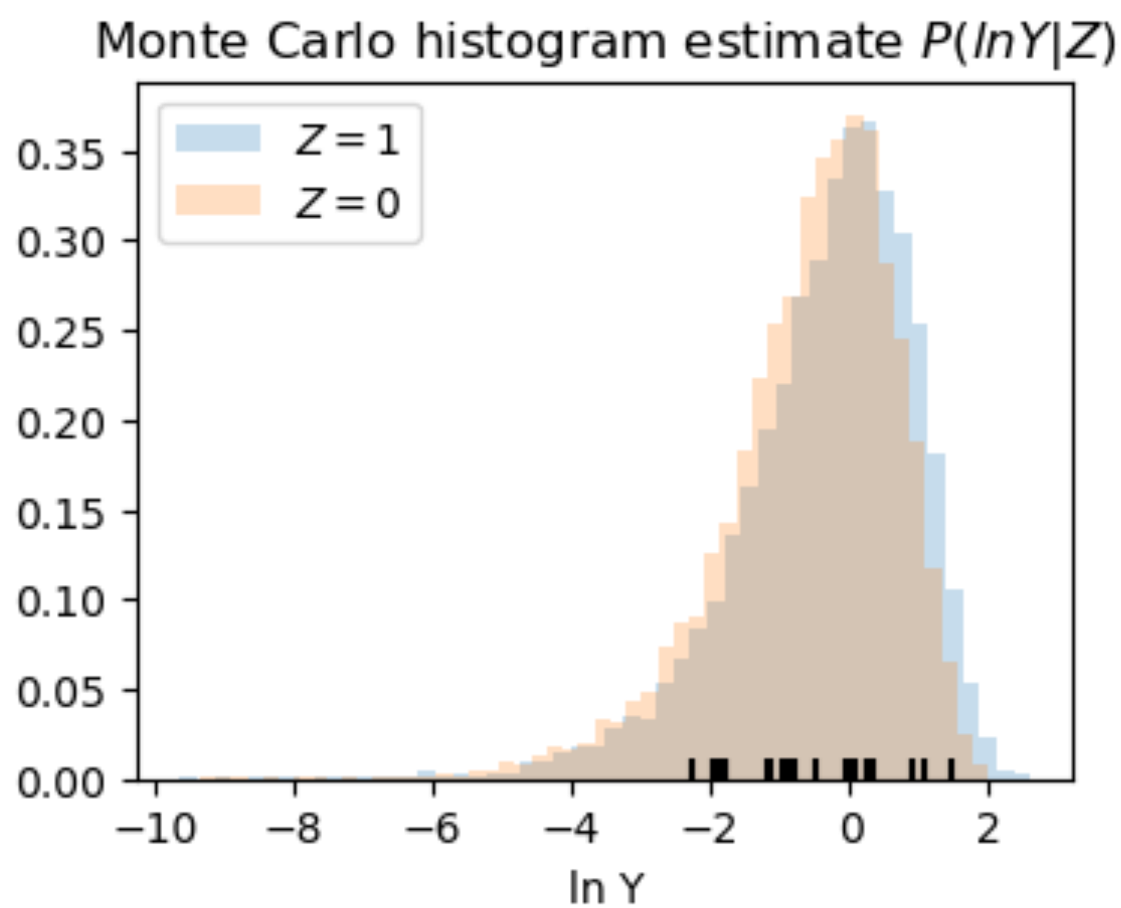
$$y^{(s)} \sim \text{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

Scenario: Magnetometer data

$$P(D \mid Z = 1) = \textbf{P(}\ D \text{ between 2 points placed uniformly at random } \textbf{)}$$

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$



- **Monte Carlo** simulation:

for s in num_samples:

$$p_1^{(s)} \sim \text{Uniform}(\square)$$

$$p_2^{(s)} \sim \text{Uniform}(\square)$$

$$D_s = \text{dist}(p_1^{(s)}, p_2^{(s)})$$

$$y^{(s)} \sim \text{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

- **Histogram density** estimate:

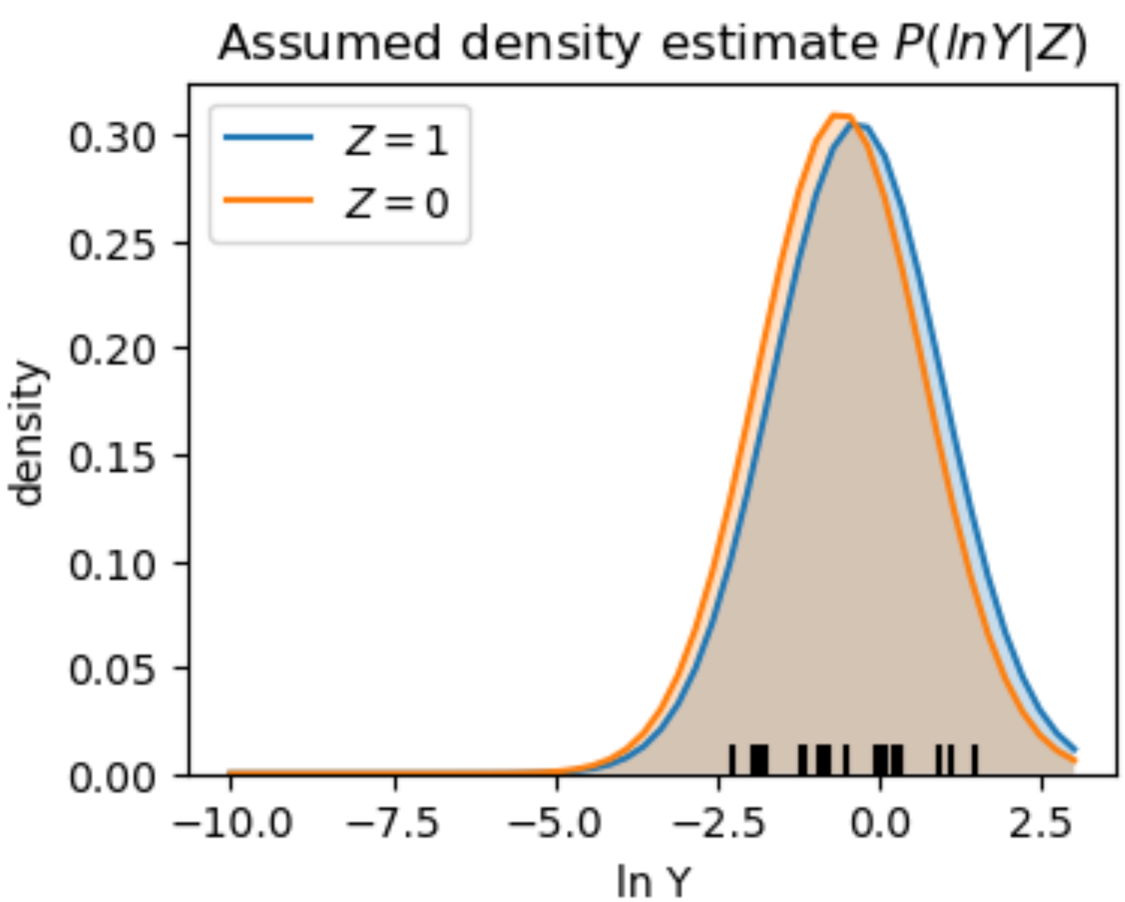
$$P(y \mid Z = 1) \approx \sum_{m=1}^{M} \delta(\ln y \in B_m) \frac{\text{num samples in bin } m}{\text{length of bin}}$$

Scenario: Magnetometer data

$P(D \mid Z=1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \text{dist}(p_1,p_2)\right)\right]$$



- **Monte Carlo** simulation:

  for s in num_samples:

$$p_1^{(s)} \sim \text{Uniform}(\square)$$
$$p_2^{(s)} \sim \text{Uniform}(\square)$$
$$D_s = \text{dist}(p_1^{(s)}, p_2^{(s)})$$
$$y^{(s)} \sim \text{Expon}(\alpha_0 + \alpha_1 D_s^{-3})$$

- **Assumed density** estimate:

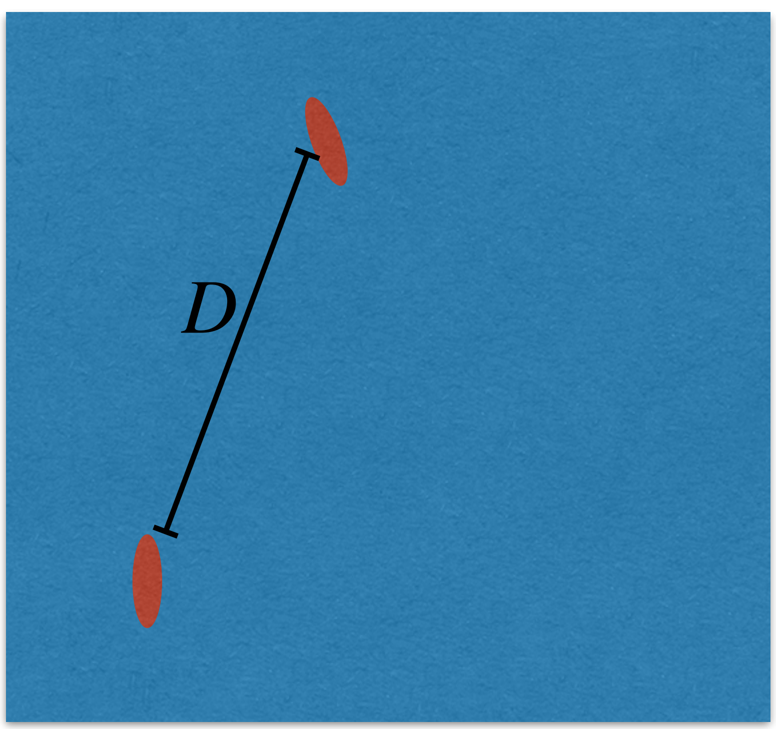
$$P(\ln y \mid Z) \approx \mathcal{N}(\ln y;\, \hat{\mu}_Z, \hat{\sigma}_Z)$$

Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** D between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1, p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$



cell *k*

Scenario: Magnetometer data

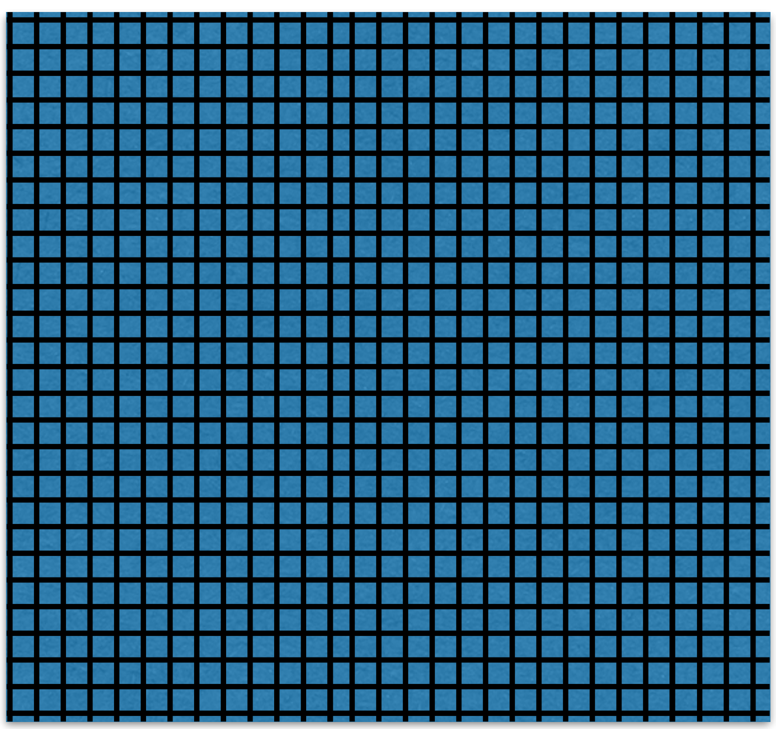
$$P(D \mid Z = 1) =$$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1,p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$

- **Discrete distribution** estimate: Grid up the cell, calculate distance between every pair of vertices

$$\approx \sum_k \pi_k \delta(D = d_k)$$ (arbitrarily exact for finer grids)



cell *k*

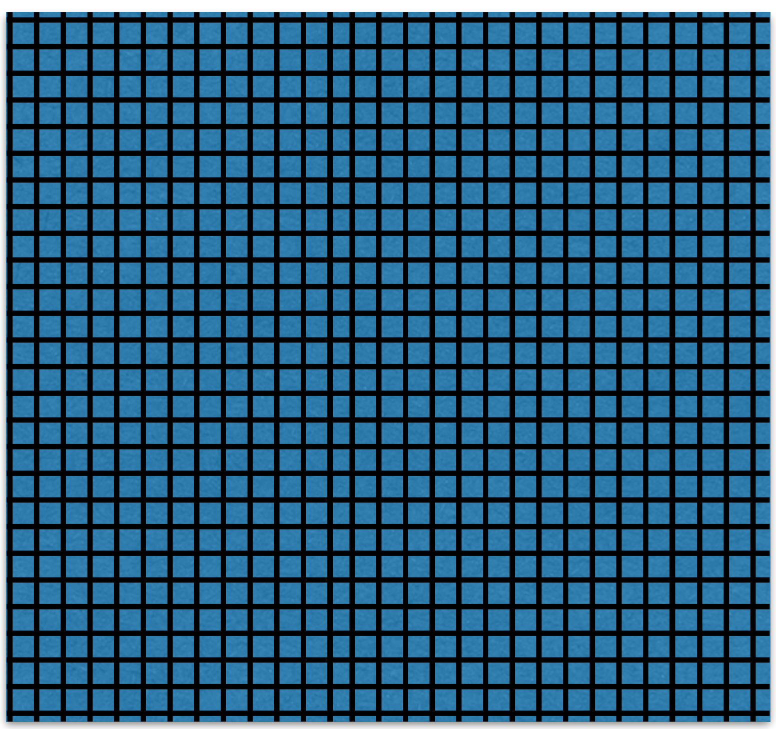
Scenario: Magnetometer data

$P(D \mid Z = 1) =$ **P(** *D* between 2 points placed uniformly at random **)**

prior (for *D*)

$$= \mathbb{E}_{p_1, p_2}\left[\delta\left(D = \text{dist}(p_1, p_2)\right)\right]$$

- **Discrete distribution** estimate: Grid up the cell, calculate distance between every pair of vertices



cell *k*

$$\approx \sum_k \pi_k \delta(D = d_k)$$ (arbitrarily exact for finer grids)

$$P(y \mid Z = 1) = \int P(y \mid Z = 1, D)\, P(D \mid Z = 1)\, dD$$

$$\approx \sum_k P(y \mid Z = 1, D = d_k)\, \pi_k$$

Scenario: Magnetometer data

Recap on how we built our **likelihood**:

$$P(y_1, \ldots, y_N \mid Z) = \prod_{i=1}^{N} P(y_i \mid Z) \qquad \text{(}\textbf{exchangeability}\text{)}$$

Scenario: Magnetometer data

Recap on how we built our **likelihood**:

$$P(y_1, \ldots, y_N \mid Z) = \prod_{i=1}^{N} P(y_i \mid Z) \quad \text{(exchangeability)}$$

$$P(y_i \mid Z = 0) = \mathrm{Expon}(y_i;\, \alpha_0) \quad \text{(background knowledge)}$$

Scenario: Magnetometer data

Recap on how we built our **likelihood**:

$$P(y_1, \ldots, y_N \mid Z) = \prod_{i=1}^{N} P(y_i \mid Z) \quad \textbf{(exchangeability)}$$

$$P(y_i \mid Z = 0) = \text{Expon}(y_i;\ \alpha_0) \quad \textbf{(background knowledge)}$$

$$P(y_i \mid Z = 1) = \sum_k \pi_k \, \text{Expon}(y_i; \alpha_0 + \alpha_1 d_k^{-3}) \quad \textbf{(marginalize)}$$

Scenario: Magnetometer data

Recap on how we built our **likelihood**:

$$P(y_1, \ldots, y_N \mid Z) = \prod_{i=1}^{N} P(y_i \mid Z) \quad \textbf{(exchangeability)}$$

$$P(y_i \mid Z = 0) = \mathrm{Expon}(y_i;\, \alpha_0) \quad \textbf{(background knowledge)}$$

$$P(y_i \mid Z = 1) = \sum_k \pi_k \, \mathrm{Expon}(y_i; \alpha_0 + \alpha_1 d_k^{-3}) \quad \textbf{(marginalize)}$$

We can now compute the **likelihood odds**:

$$\frac{\prod_{i=1}^{N} P(y_i \mid Z = 1)}{\prod_{i=1}^{N} P(y_i \mid Z = 0)} = 2.8$$

Scenario: Magnetometer data

Recap on how we built our **likelihood**:

$$P(y_1, \ldots, y_N \mid Z) = \prod_{i=1}^{N} P(y_i \mid Z) \quad \text{(\textbf{exchangeability})}$$

$$P(y_i \mid Z = 0) = \mathrm{Expon}(y_i;\ \alpha_0) \quad \text{(\textbf{background knowledge})}$$

$$P(y_i \mid Z = 1) = \sum_k \pi_k \,\mathrm{Expon}(y_i; \alpha_0 + \alpha_1 d_k^{-3}) \quad \text{(\textbf{marginalize})}$$

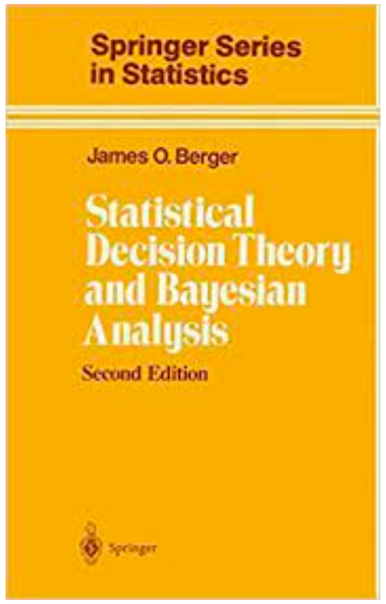
We can now compute the **likelihood odds**:

$$\frac{\prod_{i=1}^{N} P(y_i \mid Z = 1)}{\prod_{i=1}^{N} P(y_i \mid Z = 0)} = 2.8$$

is this big?

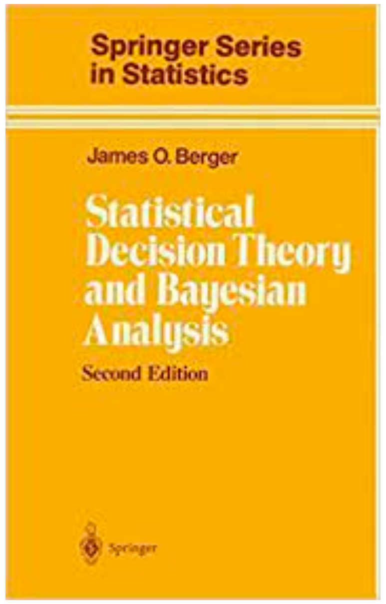
Lady drinking tea

1. A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.



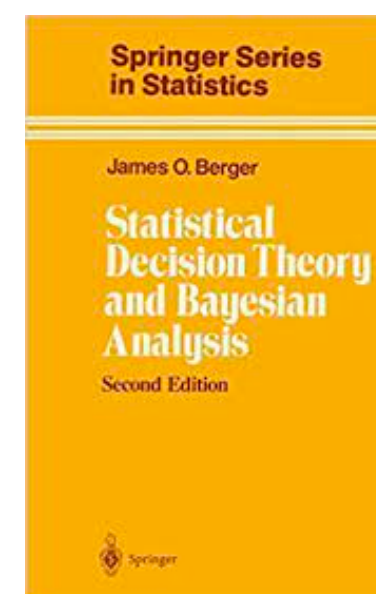
Lady drinking tea

1. A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score. In ten trials conducted to test this, he makes a correct determination each time.



Lady drinking tea

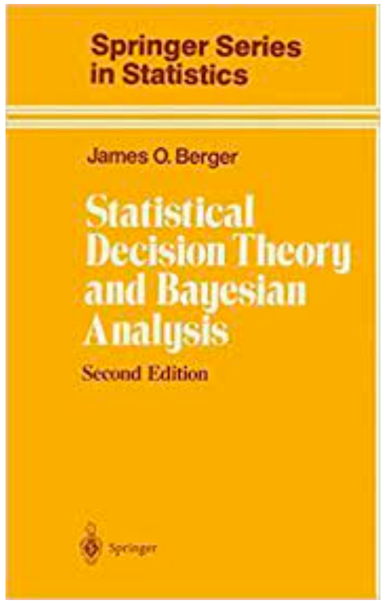
1. A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score. In ten trials conducted to test this, he makes a correct determination each time.
3. A drunken friend says he can predict the outcome of a flip of a fair coin. In ten trials conducted to test this, he is correct each time.



Lady drinking tea

1. A lady, who adds milk to her tea, claims to be able to tell whether the tea or the milk was poured into the cup first. In all of ten trials conducted to test this, she correctly determines which was poured first.
2. A music expert claims to be able to distinguish a page of Haydn score from a page of Mozart score. In ten trials conducted to test this, he makes a correct determination each time.
3. A drunken friend says he can predict the outcome of a flip of a fair coin. In ten trials conducted to test this, he is correct each time.

In all three situations, the unknown quantity θ is the probability of the person answering correctly. A classical significance test of the various claims would consider the null hypothesis (H_0) that $\theta = 0.5$ (i.e., the person is guessing). In all three situations this hypothesis would be rejected with a (one-tailed) significance level of 2^{-10}. Thus the above experiments give strong evidence that the various claims are valid.



Prior, posterior odds

$$P(Z = 1 \mid Y) = \frac{P(Y \mid Z = 1)\,\pi_1}{P(Y \mid Z = 1)\,\pi_1 + P(Y \mid Z = 0)\,\pi_0}$$

Prior, posterior odds

$$P(Z=1 \mid Y) = \frac{P(Y \mid Z=1)\,\pi_1}{P(Y \mid Z=1)\,\pi_1 + P(Y \mid Z=0)\,\pi_0}$$

$$= \frac{\frac{P(Y \mid Z=1)}{P(Y \mid Z=0)}\,\pi_1}{\frac{P(Y \mid Z=1)}{P(Y \mid Z=0)}\,\pi_1 + \pi_0}$$

Posterior = likelihood odds reweighted by the prior

Prior, posterior odds

$$P(Z = 1 \mid Y) \propto P(Y \mid Z = 1)\, \pi_1$$

$$P(Z = 0 \mid Y) \propto P(Y \mid Z = 0)\, \pi_0$$

Prior, posterior odds

$$P(Z = 1 \mid Y) \propto P(Y \mid Z = 1)\, \pi_1$$

$$P(Z = 0 \mid Y) \propto P(Y \mid Z = 0)\, \pi_0$$

$$\frac{P(Z = 1 \mid Y)}{P(Z = 0 \mid Y)} = \frac{P(Y \mid Z = 1)}{P(Y \mid Z = 0)}\, \frac{\pi_1}{\pi_0}$$

posterior odds = likelihood odds × prior odds

Scenario: Magnetometer data

- We obtained a **likelihood odds** = **2.8**
- Whether this is ``big" depends on how plausible Z=1 is
- What is P(Z=1)?
- This is a different kind of prior than the others we saw
- *uninformative* vs *informative* priors

Scenario: Magnetometer data

- We want an *informative* prior P(Z=1)
- We know the Captain likes to gamble
- We ask the Captain to give us the odds that the *Scorpion* is in cell *k* (he hasn't seen the magnetometer data)
- He says **20:1 against**

Captain

Scenario: Magnetometer data

- We want an *informative* prior P(Z=1)
- We know the Captain likes to gamble
- We ask the Captain to give us the odds that the *Scorpion* is in cell *k* (he hasn't seen the magnetometer data)
- He says **20:1 against**

Captain

$$\text{posterior odds} = \text{likelihood odds} \times \text{prior odds}$$

$$= 2.8 \times \frac{1}{20} = 0.14$$

(data + prior) is about 7-to-1 in favor of Z=0

Eliciting prior odds

- In this case, we used a **subjectivist** prior
- The same calculation is often used with **frequencies**
- e.g., in medicine: a patient presents with mild discomfort and worries they have small pox:

$$P(\text{small pox} \mid \text{discomfort}) \propto P(\text{discomfort} \mid \text{small pox})\, P(\text{small pox})$$

Eliciting prior odds

- In this case, we used a **subjectivist** prior
- The same calculation is often used with **frequencies**
- e.g., in medicine: a patient presents with mild discomfort and worries they have small pox:

$$P(\text{small pox} \mid \text{discomfort}) \propto P(\text{discomfort} \mid \text{small pox})\, P(\text{small pox})$$

Captain

- Back to subjectivist priors: what is the rationale for eliciting them with odds?
- Let's play a game… (blackboard)