BAYESIAN CHANGE-POINT DETECTION FOR PROCESS MONITORING WITH FAULT DETECTION

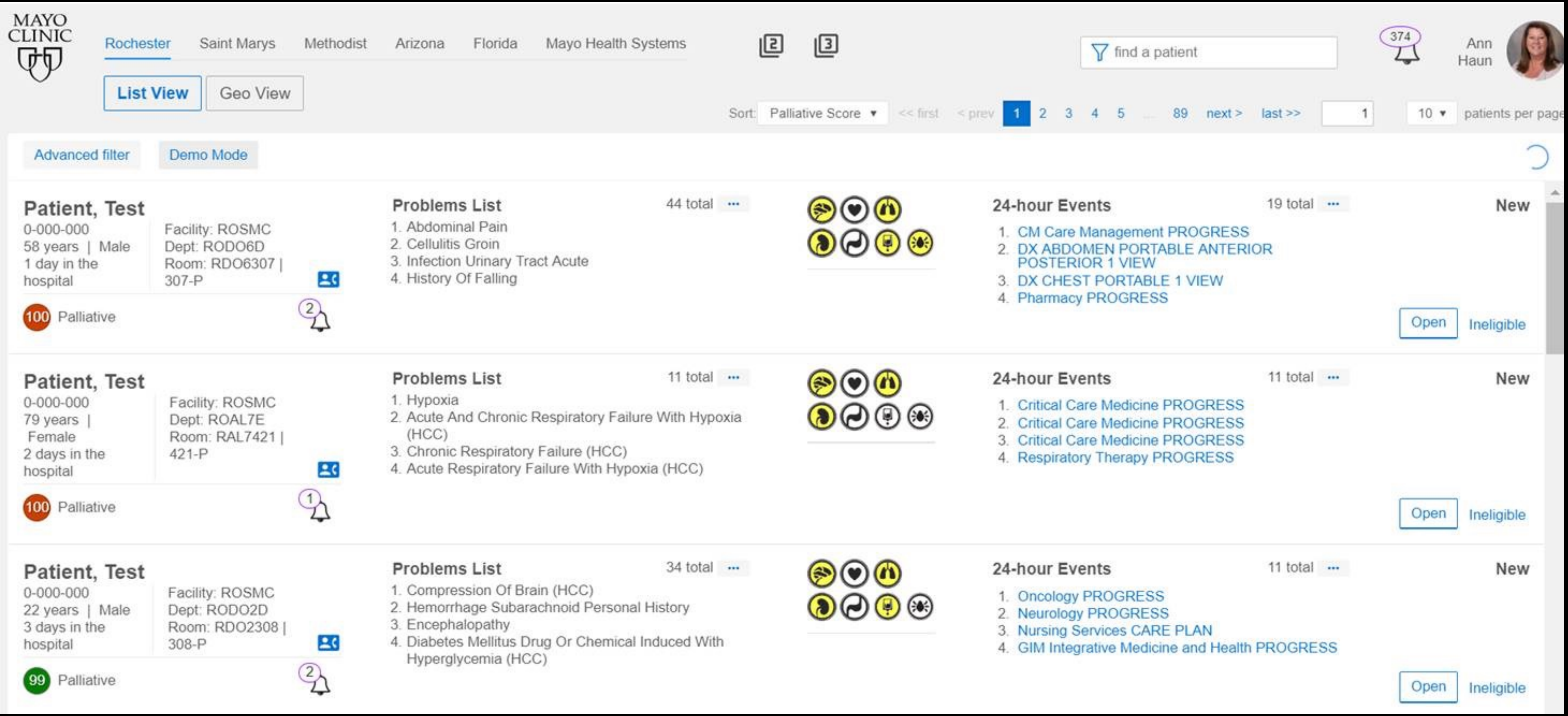
KERN CENTER FOR THE SCIENCE OF HEALTH CARE DELIVERY

Alexander C. Murph

April 21st, 2023

 |

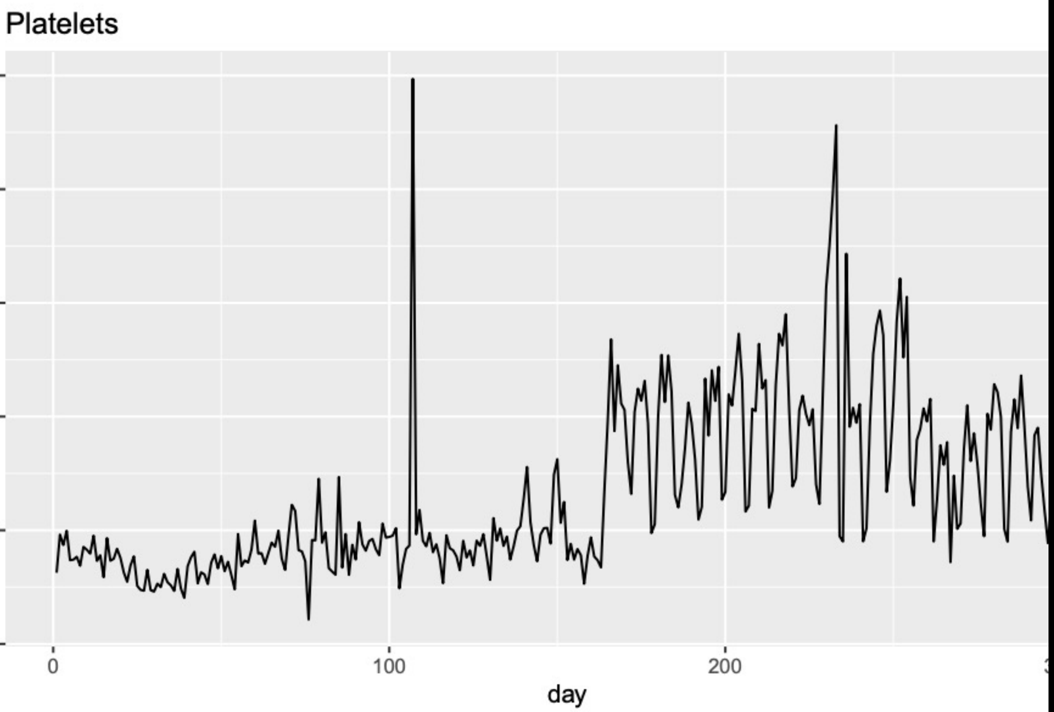
CONTROL TOWER INTERFACE

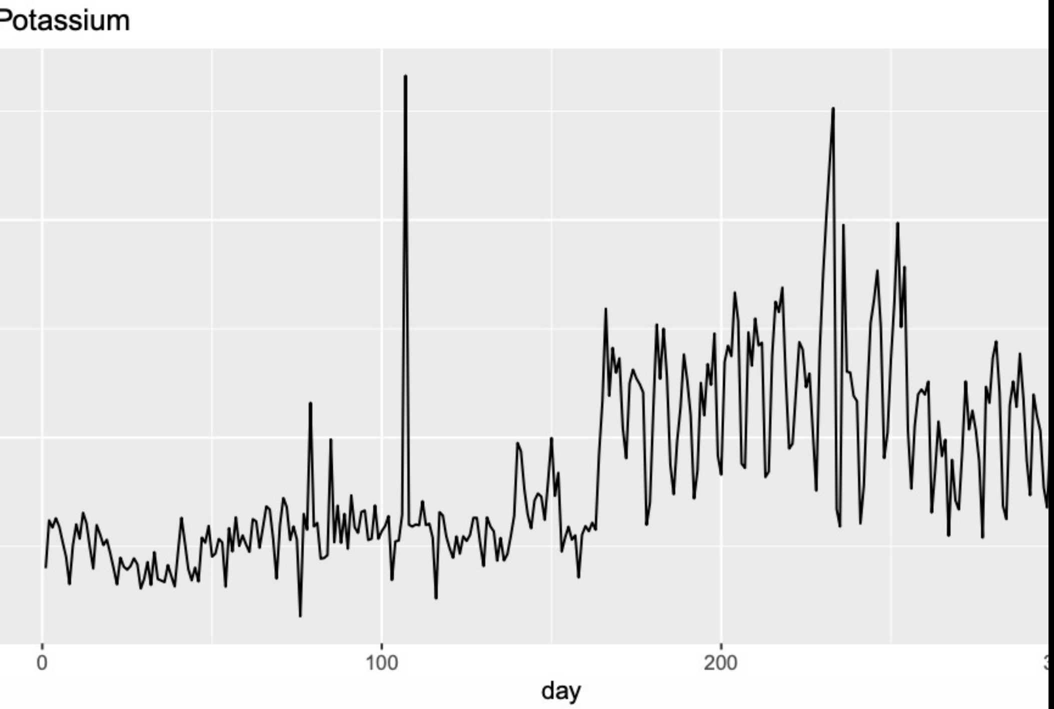


<u>Number in the circle</u>:

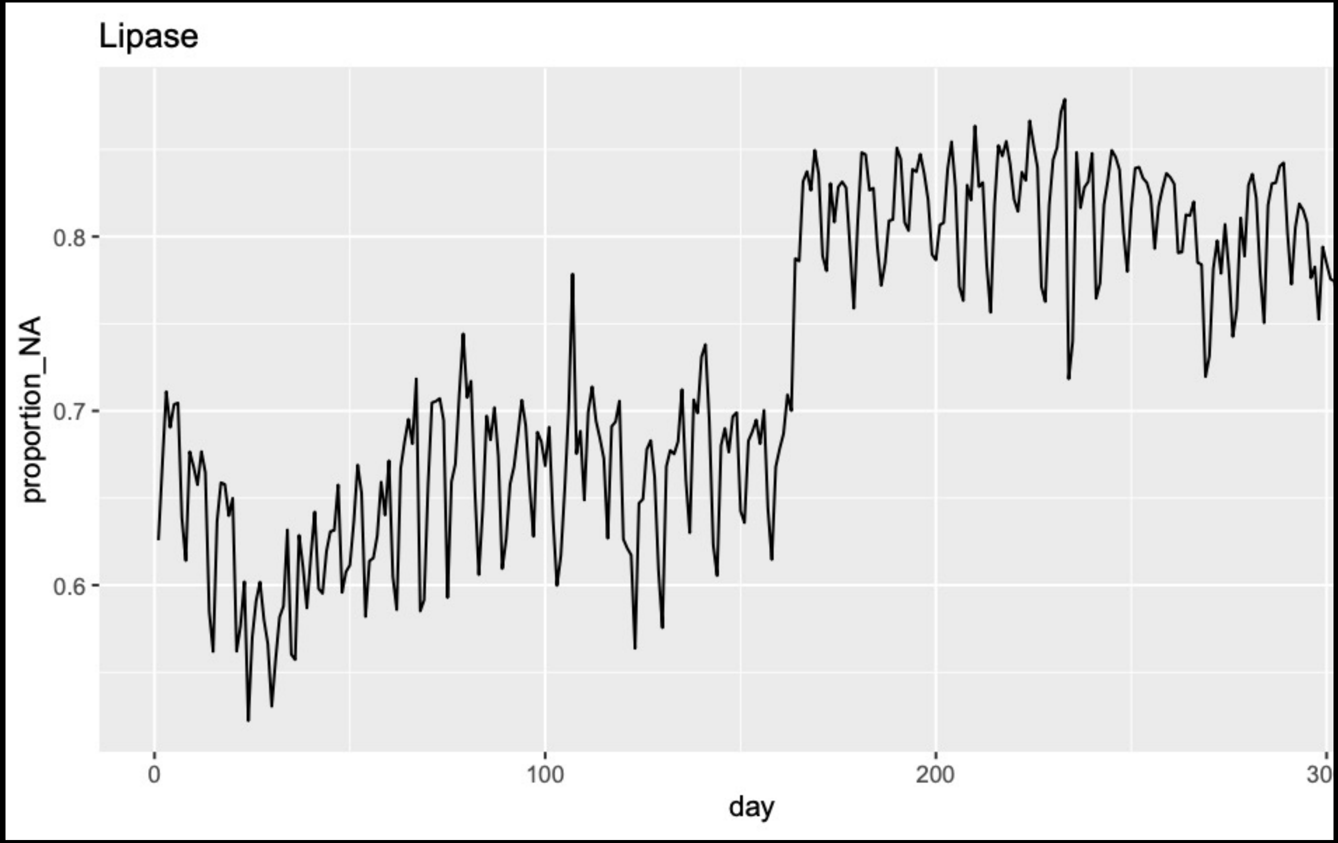
"The probability of getting an inpatient palliative care consultation in the next 7 days."

Thank you to Patrick Wilson for sharing this slide!





WELL, WHAT COULD GO WRONG?



A (NOT SO) ELEGANT PROBLEM

(Sub) problems:

- Missing Data
- Mixed Data (discrete, binary, continuous)
- Censored Values

- HUGE data (n in millions, p ~= 250)

(Sub) solutions:

- Bayesian Latent Variables
- Gaussian Graphical Models (GGMs)

(Sub-Sub) problems:

- Lack of information for some priors
- Learning parameters for a GGM (especially the conditional independence graph structure)

(Sub-Sub) solutions:

- Bayesian Hierarchical Model
- Double Reversible Jump Metropolis Hastings, Conjugate priors.

A (NOT SO) ELEGANT PROBLEM

(Sub) problems:

- Missing Data
- Mixed Data (discrete, binary, continuous)
- Censored Values
- HUGE data (n in millions, p ~= 250)

(Sub) solutions:

- Bayesian Latent Variables
- Gaussian Graphical Models (GGMs)

(Sub-Sub) problems:

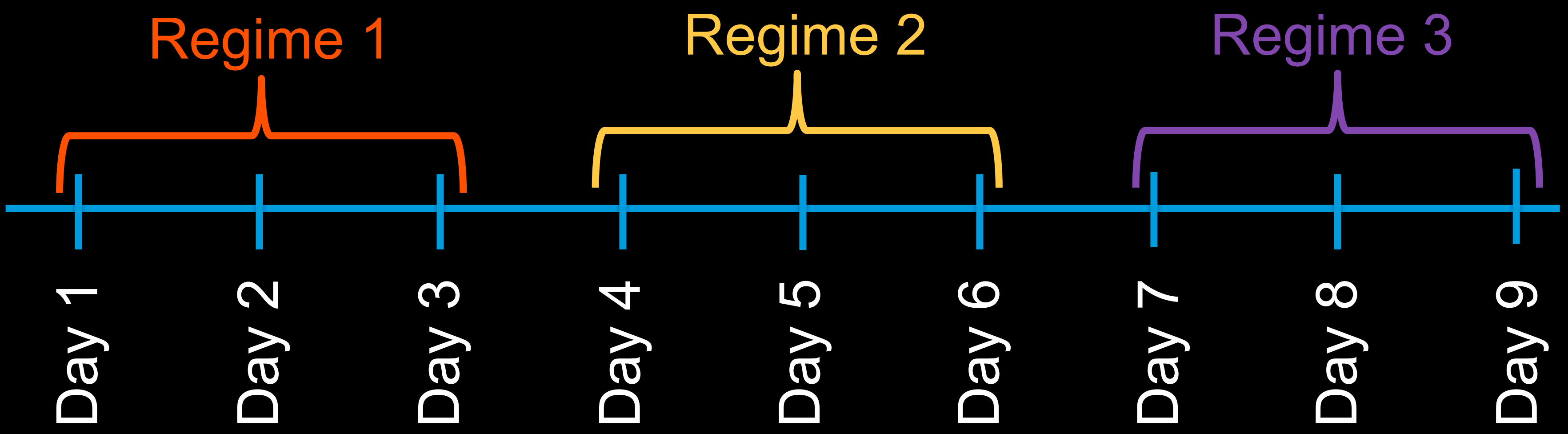
- Lack of information for some priors
- Learning parameters for a GGM (especially the conditional independence graph structure)

(Sub-Sub) solutions:

- Bayesian Hierarchical Model
- Double Reversible Jump Metropolis Hastings, Conjugate priors.

HOW DO WE FIND A CHANGE POINT?

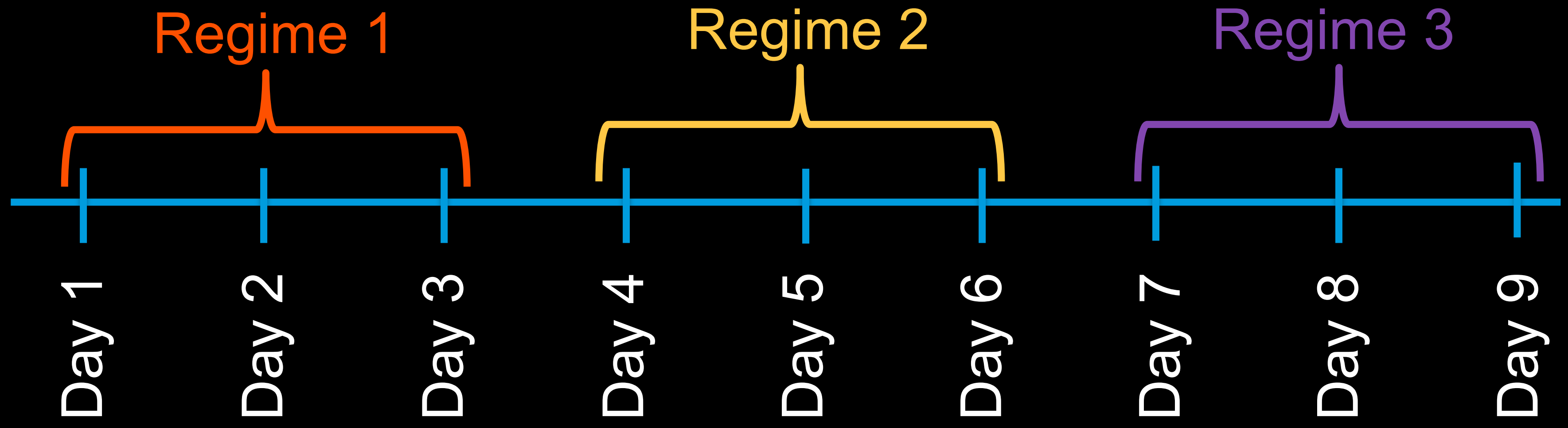
LET’S CLASSIFY OBSERVATIONS OVER TIME INTO “REGIMES”



Data we observed on ______

SUPPOSE WE OBSERVED 9 DAYS OF DATA…

We can *encode* each day’s regime assignment as a vector of length 9:



Data we observed on _____

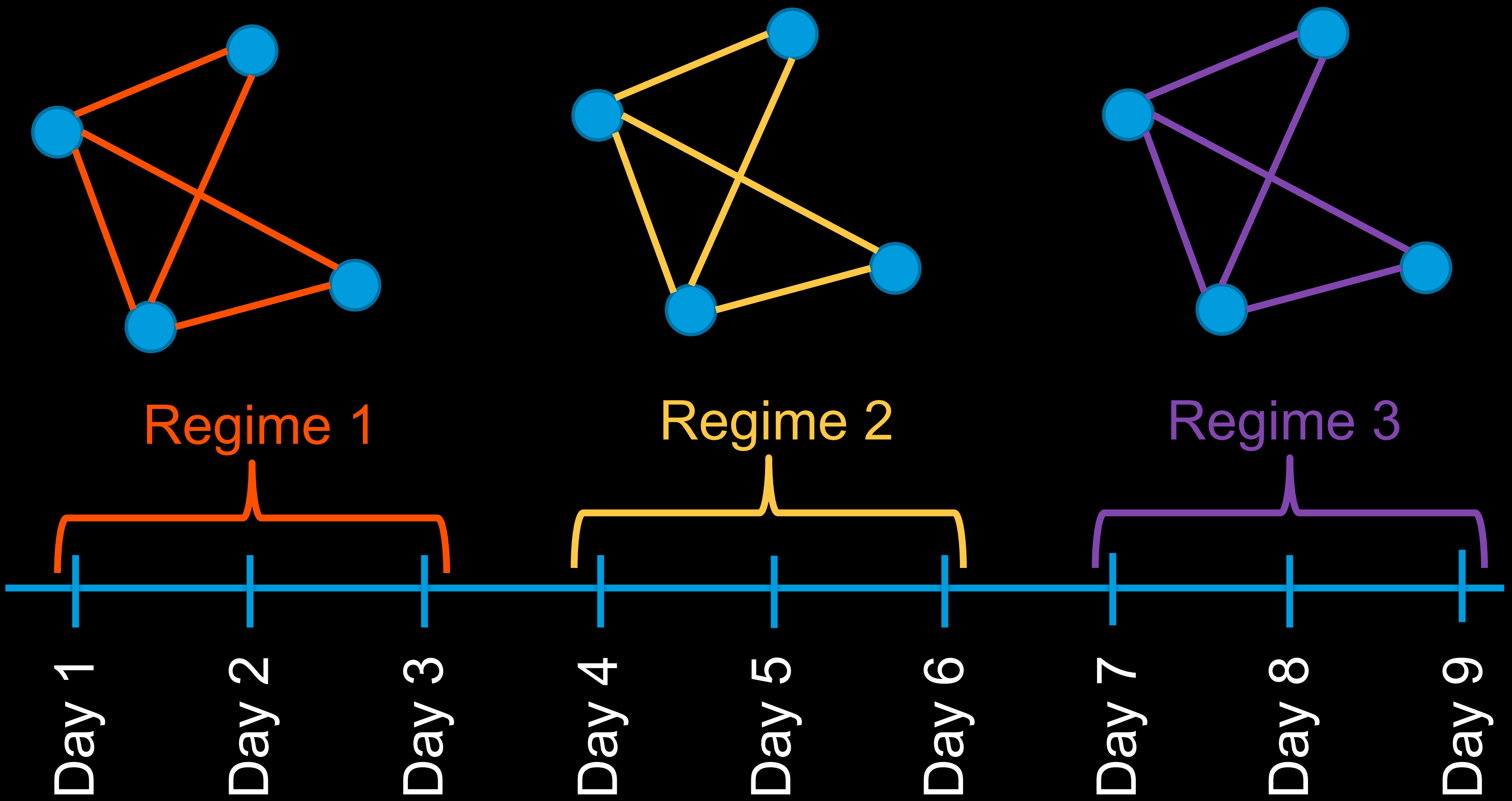
becomes

$$\phi = (1, 1, 1, 2, 2, 2, 3, 3, 3)$$

 |

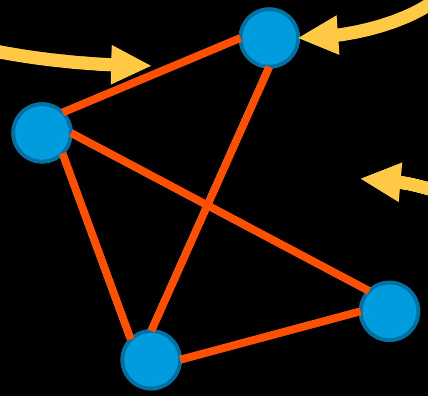
WHAT'S THE LIKELIHOOD?

WE MUST DEFINE THE DISTRIBUTION OF THE DATA *GIVEN* A REGIME VECTOR



Weighted edges correspond to values in the *precision matrix* of a multivariate normal model.

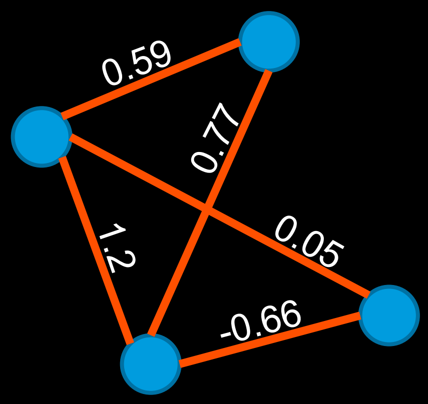
The nodes correspond to the variables in our dataset.



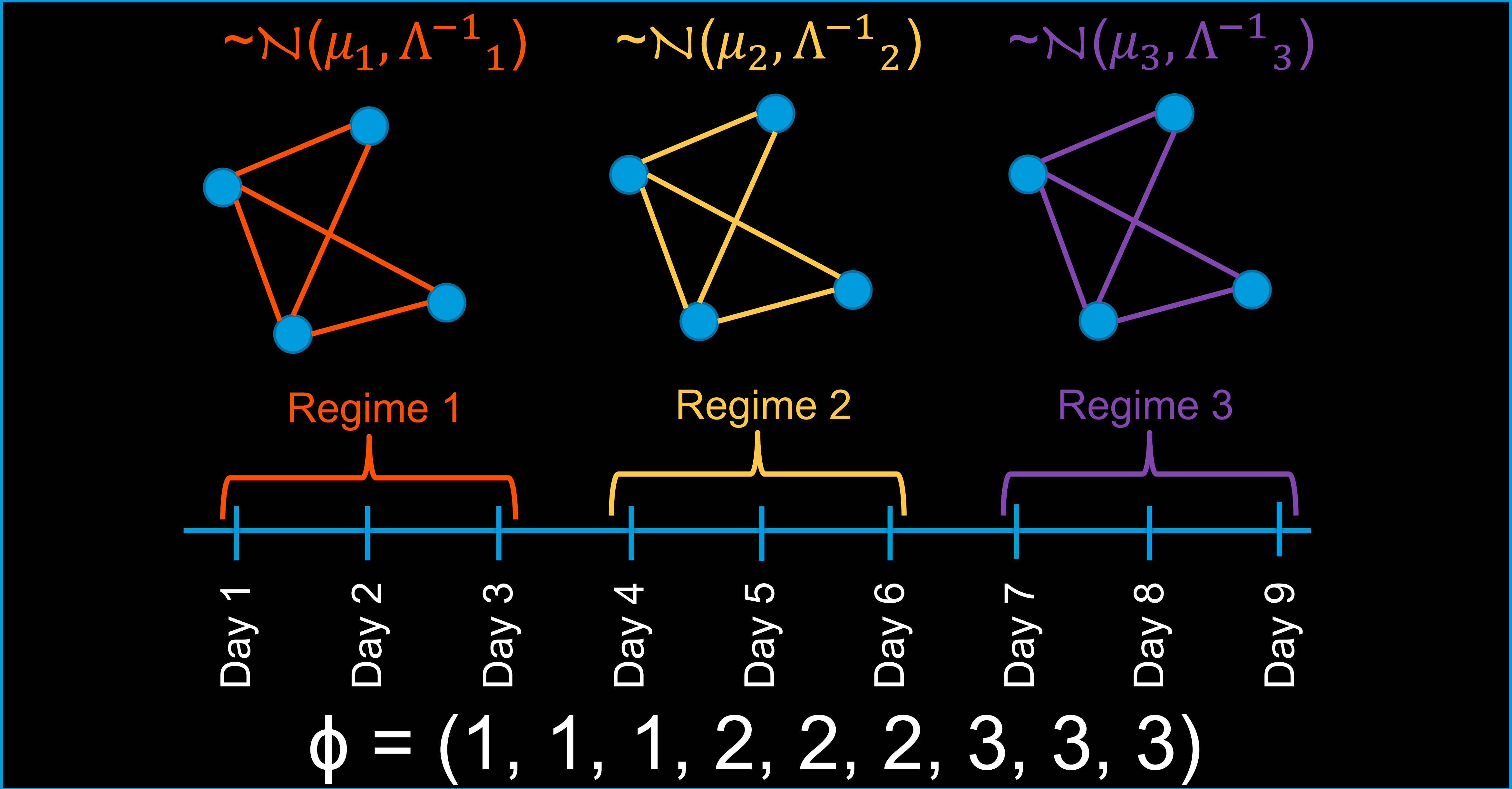
The *absence* of an edge corresponds to a zero in the precision matrix

$$\Lambda, \text{ where } \quad Y \sim \mathcal{N}_4(\mu, \Lambda^{-1})$$

FOR INSTANCE,



$$\Lambda = \begin{bmatrix} 1 & 0.59 & 0.05 & 1.2 \\ 0.59 & 1 & 0 & 0.77 \\ 0.05 & 0 & 1 & -0.66 \\ 1.2 & 0.77 & -0.66 & 1 \end{bmatrix}$$



$\sim\mathcal{N}(\mu_1, {\Lambda^{-1}}_1)$
$\sim\mathcal{N}(\mu_2, {\Lambda^{-1}}_2)$
$\sim\mathcal{N}(\mu_3, {\Lambda^{-1}}_3)$
Regime 1
Regime 2
Regime 3
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Day 7
Day 8
Day 9
$\phi = (1, 1, 1, 2, 2, 2, 3, 3, 3)$

INFERENCE ON *THREE* PARAMETERS OF INTEREST

- The set-up on the previous slides admits a likelihood for our data (a product of normal mixture models).
- To get good posterior samples on ϕ, μ, Λ, G, we require:
    - Priors on each parameter;
    - Sampling procedures for proposal values.

SPARSITY ENCODED USING THE G-WISHART

Prior Distributions on the Parameters

$$\Lambda \sim \text{G-Wishart}(\delta, I)$$

(δ, degrees of freedom)

$$\mu | \Lambda \sim \mathcal{N}(\mu_0, \Lambda^{-1})$$

$$(\mu, \Lambda) \sim f_{\mu|\Lambda} f_{\Lambda}$$

$$p(\Lambda := \Sigma^{-1} | G) = \frac{1}{I_G(\delta, D)} (\det \Lambda)^{(\delta-2)/2} \exp\left\{ -\frac{1}{2} \langle \Lambda, D \rangle \right\}$$

★ is a normalizing constant gotten by integrating over a space of precision matrices with fixed zero structure

TO EVALUATE

Laplace Approximation:

(Dobra and Lenkowski 2011)

$$h_{\delta,D}(\Lambda) = -\frac{1}{2}[\mathrm{tr}(\Lambda^T D) - (\delta - 2)\log(\det K)]$$

$$I_G(\hat{\delta}, D) = \exp\left\{h_{\delta,D}(\hat{\Lambda})\right\}(2\pi)^{|V|/2}[\det H_{\delta,D}(\hat{\Lambda})]^{-1/2}$$

(V is the set of free elements in Λ, H is the hessian of h)

When we must calculate the normalizing constant, there exists a fast approximation!

TO SAMPLE

Algorithm 1 Exact sampling from the precision matrix.

Input: A graph $G = (V, E)$ with precision matrix K and $\Sigma = K^{-1}$
Output: An exact sample from the precision matrix.

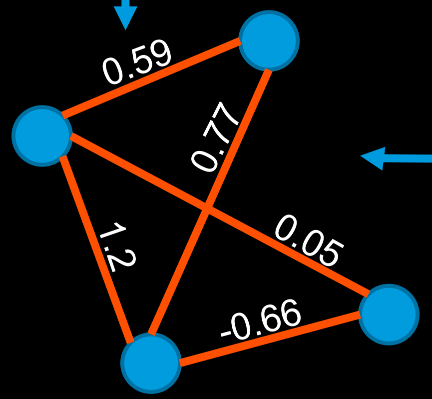
1: Set $\Omega = \Sigma$
2: **repeat**
3:    **for** $i = 1, \ldots, p$ **do**
4:       Let $N_i \subset V$ be the neighbor set of node i in G. Form Ω_{N_i} and $\Sigma_{N_i,i}$ and solve $\hat{\beta}_i^* = \Omega_{N_i}^{-1}\Sigma_{N_i,i}$.
5:       Form $\hat{\beta}_i \in R^{p-1}$ by padding the elements of $\hat{\beta}_i^*$ to the appropriate locations and zeros in those locations not connected to i in G.
6:       Update $\Omega_{i,-i}$ and $\Omega_{-i,i}$ with $\Omega_{-i,-i}\hat{\beta}_i$.
7:    **end for**
8: **until** convergence
9: **return** $K = \Omega^{-1}$

(algorithm Mohammadi 2019)

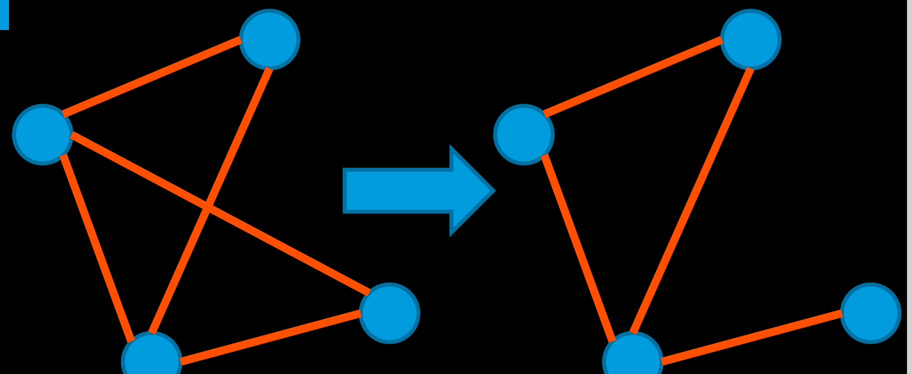
There exists algorithms to sample from a G-Wishart *without* having to calculate the normalizing constant!

Learning these edge weights is a relatively easy problem

Learning when edges are present/absent is **not** easy.



We would like to be able to…



DOUBLE REVERSIBLE JUMP METROPOLIS-HASTINGS MONTE CARLO

LEARNING THE *NUMBER* OF PARAMETERS, WHILE ALSO LEARNING THE PARAMETERS THEMSELVES.

This process combines two neat ideas:

1. Put a sampling distribution on the *Cholesky decomposition* Λ=Ψ`Ψ of the precision matrix Λ;
2. Learn an ancillary grid of values along with the observed graph and precision matrix. Then, propose a new graph and matrix by *swapping.*

$$(\Lambda^d, G^d, \hat{\Lambda}^d, \hat{G}^d)$$

$$(\Lambda^{d+1}, G^{d+1}, \hat{\Lambda}^{d+1}, \hat{G}^{d+1})$$

SO WHERE ARE THE CHANGEPOINTS?

- (1, 1, 1, 2, 2, 2, 3, 3, 3)
- (1, 1, 2, 2, 2, 2, 2, 3, 3)
- (1, 1, 2, 2, 2, 2, 2, 2, 2)
- (1, 1, 2, 2, 3, 3, 3, 3, 3)
- (1, 1, 1, 1, 2, 2, 2, 2, 2)
- (1, 1, 1, 1, 2, 2, 2, 2, 3)
- (1, 1, 1, 1, 2, 2, 2, 2, 2)
- (1, 1, 1, 1, 1, 1, 1, 1, 1)

$\vdots$

SO WHERE ARE THE CHANGEPOINTS?

- (1, 1, 1, 2, 2, 2, 3, 3, 3)
- (1, 1, 2, 2, 2, 2, 2, 3, 3)
- (1, 1, 2, 2, 2, 2, 2, 2, 2)
- (1, 1, 2, 2, 3, 3, 3, 3, 3)
- (1, 1, 1, 1, 2, 2, 2, 2, 2)
- (1, 1, 1, 1, 2, 2, 2, 2, 3)
- (1, 1, 1, 1, 2, 2, 2, 2, 2)
- (1, 1, 1, 1, 1, 1, 1, 1, 1)

⋮

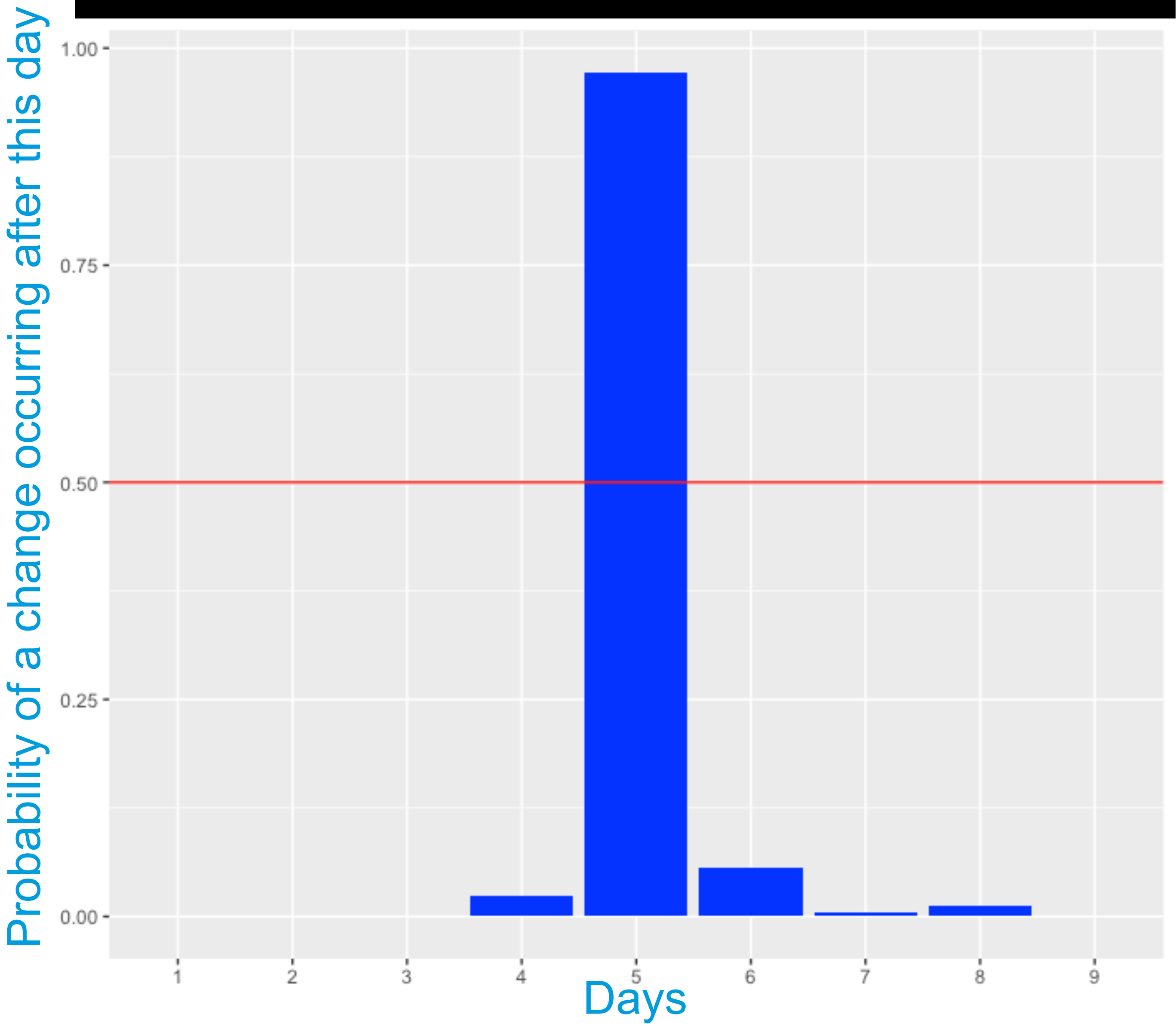
- (0, 0, 1, 0, 0, 1, 0, 0)
- (0, 1, 0, 0, 0, 0, 1, 0)
- (0, 1, 0, 0, 0, 0, 0, 0)
- (0, 1, 0, 1, 0, 0, 0, 0)
- (0, 0, 0, 1, 0, 0, 0, 0)
- (0, 0, 0, 1, 0, 0, 0, 1)
- (0, 0, 0, 1, 0, 0, 0, 0)
- (0, 0, 0, 0, 0, 0, 0, 0)

⋮

 |

- (0, 0, 1, 0, 0, 1, 0, 0)
- (0, 1, 0, 0, 0, 0, 1, 0)
- (0, 1, 0, 0, 0, 0, 0, 0)
- (0, 1, 0, 1, 0, 0, 0, 0)
- (0, 0, 0, 1, 0, 0, 0, 0)
- (0, 0, 0, 1, 0, 0, 0, 1)
- (0, 0, 0, 1, 0, 0, 0, 0)
- (0, 0, 0, 0, 0, 0, 0, 0)

⋮



COMPETING METHODS

Hotelling T^2 (HT2) Sliding Window

Given four days of data, days 1,2, & 3 are compared against day 4.

$$T*2 = (\bar{y}_1 - \bar{y}_2)' \left(\frac{S_1}{n_1} + \frac{S_2}{n_2} \right)^{-1} (\bar{y}_1 - \bar{y}_2) \sim T^2_{p,v}$$

$\bar{y}_1$ Mean value for days 1, 2, & 3

$\bar{y}_2$ Mean value for day 4

S_1 Sample variance for days 1, 2, & 3

S_2 Sample variance for day 4

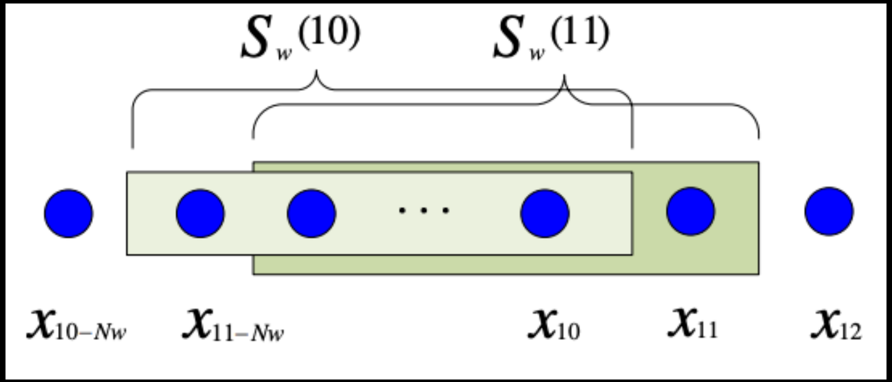
$T^2_{p,v}$ Hotelling T^2 statistic, with pooled degrees of freedom v and number of dimensions p

n_1 Number of observations for days 1, 2, & 3

n_2 Number of observations for day 4

Real Time Contrasts (RTC) with a Gradient Boosting Machine (GBM) Classifier

- Given a classifier $c(x)$ learned to a hold-out sample, does this classifier have significant predictive power?
- We simulate a distribution of AUC values on $c(x)$'s ability to distinguish day 4 from days 1, 2, & 3.
- If 0.5 is NOT in the center 95% confidence region, a changepoint is flagged.



A: Continuous, Has Missing Values, Unimodal, Covariance Change
Proportion of Simulations Flagged
1.00
0.75
0.50
0.25
0.00
10
14
20
Day Before Possible Regime Change
Method Avg. FP %
T2 Scan Stat 0.0592
RTM GBM 0.0485
Mixture Model 0.0592
MVN Model 0.0577
B: Continuous, No Missing Values, Unimodal, MeanChange
Proportion of Simulations Flagged
1.00
0.75
0.50
0.25
0.00
10
14
20
Day Before Possible Regime Change
Method Avg. FP %
T2 Scan Stat 0.0508
RTM GBM 0.0515
Mixture Model 0.0577
MVN Model 0.0546
C: Continuous, No Missing Values, Bimodal, Mean Change
Proportion of Simulations Flagged
1.00
0.75
0.50
0.25
0.00
10
14
20
Day Before Possible Regime Change
Method Avg. FP %
T2 Scan Stat 0.0385
RTM GBM 0.0592
Mixture Model 0.0538
MVN Model 0.0631
D: Missingness Amount Change
Proportion of Simulations Flagged
1.00
0.75
0.50
0.25
0.00
10
14
20
Day Before Possible Regime Change
Method Avg. FP %
T2 Scan Stat 0.0615
RTM GBM 0.0546
Mixture Model 0.0608
MVN Model 0.0592

FAULT DETECTION ANALYSIS

Once a changepoint is found, how do we determine what *caused* it?

Let

Dist'n Before: X

Dist'n After: Y

Hellinger Distance: H(X,Y)

and

Marginal Dist'n **w/o** Variable i: X \ i

Marginal Dist'n **of** Variable i: X : i

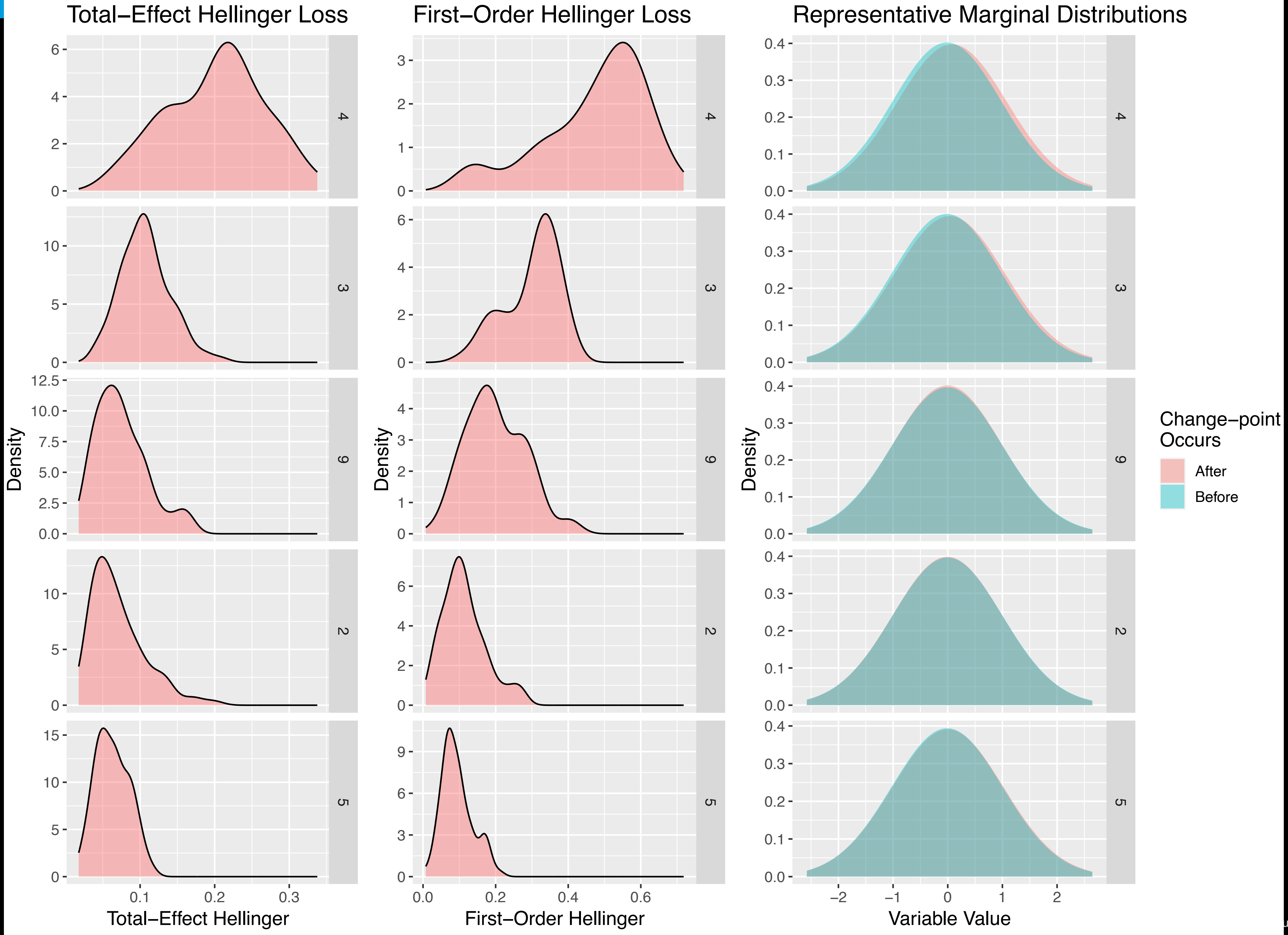
then we Define:

Total-Effect Loss of i: 1 – H(X \ i, Y \ i)/H(X,Y)

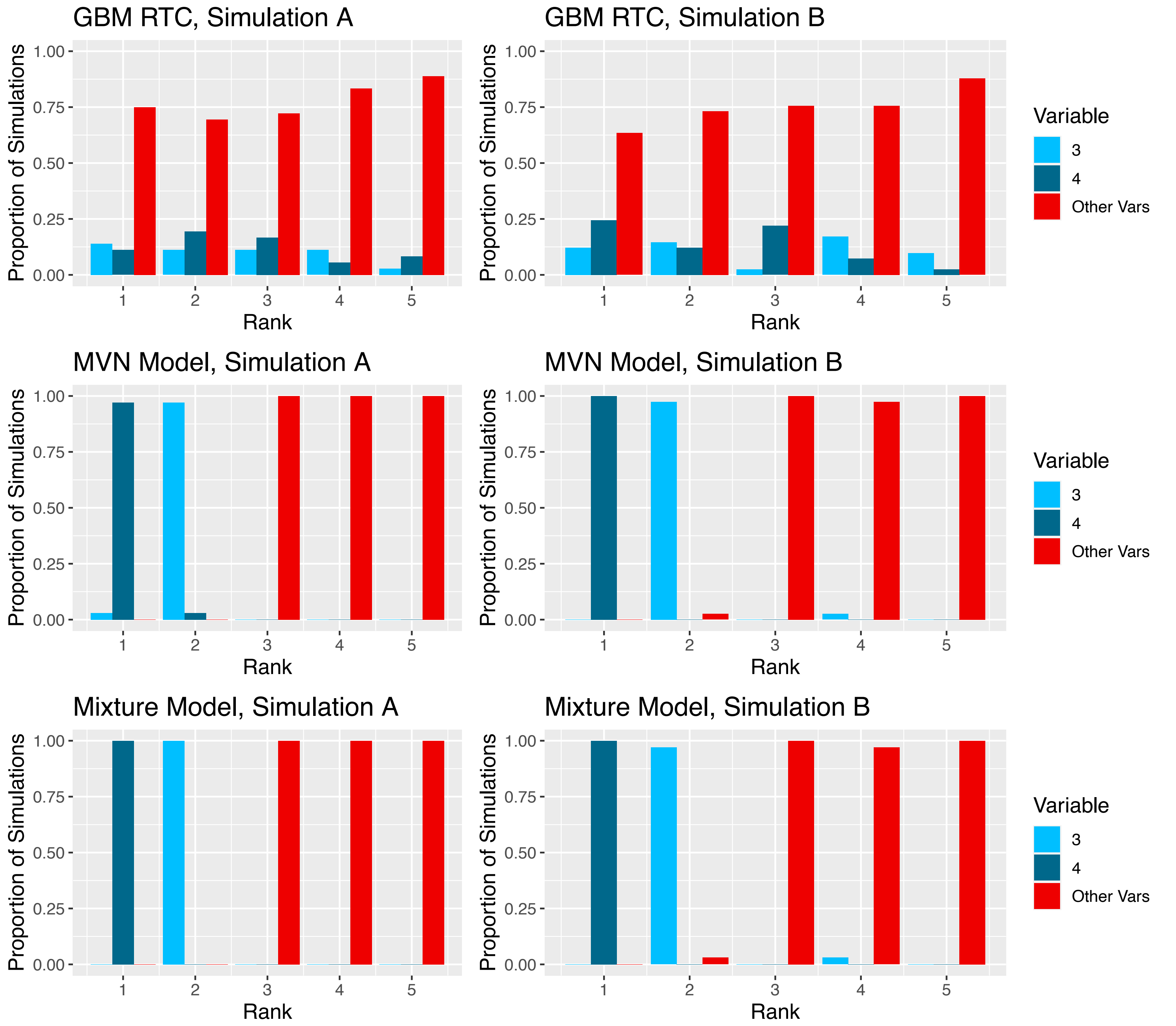
First-Order Loss of i: H(X : i, Y : i)/H(X,Y)

These metrics *decompose* the effects of each individual variable on the changepoint

Simulation



Total−Effect Hellinger Loss
First−Order Hellinger Loss
Representative Marginal Distributions
Density
Total−Effect Hellinger
First−Order Hellinger
Variable Value
Change−point Occurs
After
Before
4
3
6
2
5


GBM RTC, Simulation A
GBM RTC, Simulation B
MVN Model, Simulation A
MVN Model, Simulation B
Mixture Model, Simulation A
Mixture Model, Simulation B
Proportion of Simulations
0.00
0.25
0.50
0.75
1.00
Rank
1
2
3
4
5
Variable
3
4
Other Vars

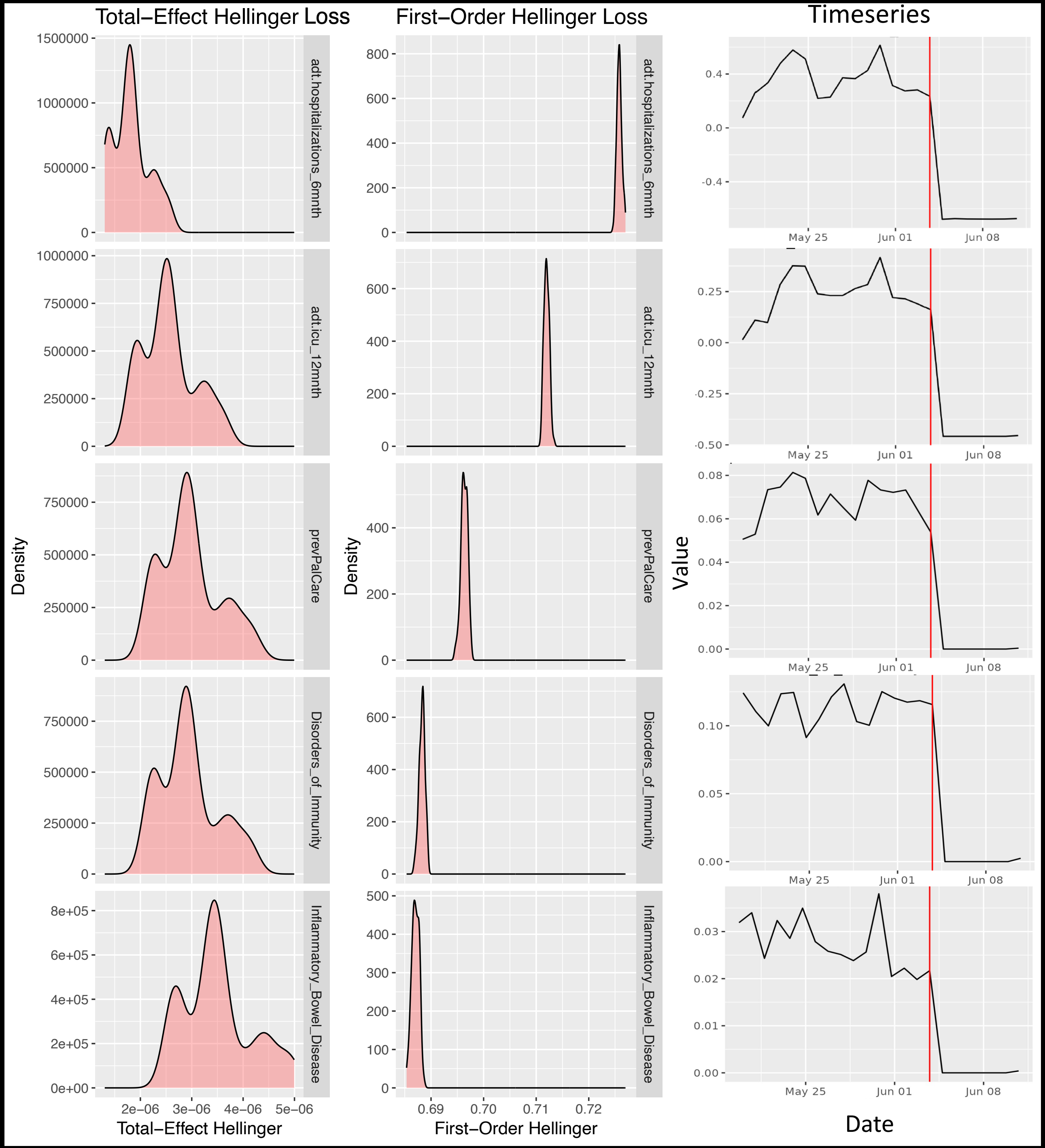
DATA APPLICATION

Let's investigate a month (mid-May to mid-June 2020) of the data from the Palliative Care Model

- Each observation in the data is a palliative care prediction that occurs for a patient at the Mayo Clinic in Rochester, MN.
- Variables include patient labs, the presence/absence of specific chronic diseases, basic demographics, and information about a patient's movement within the hospital.
- There are **123 variables**, 35 of which are continuous, with 27 variables having some amount missing values.
- During the month investigated, there were **298039 data observations**.

This is a *big* dataset with *many* data challenges; there are missing values, and various data types (censored & discrete variables).

Application



QUESTIONS?

