Generalized Fiducial Inference on Differentiable Manifolds

Alexander C. Murph
Department of Statistics & Operations Research,
University of NC at Chapel Hill
Visiting Graduate Student at the Mayo Clinic

Contents

Generalized Fiducial Inference:

1. Introduction to Generalized Fiducial Inference (GFI) (Murph *et al.*, 2022b)
2. GFI on Differentiable Manifolds (Murph *et al.*, 2022a)
3. A Geometric Perspective on Bayesian and Generalized Fiducial Inference (Liu *et al.*, 2022)

Applied Computational Bayesian:

1. Bayesian Change-point Detection for Process Monitoring with Fault Detection

Contents

Generalized Fiducial Inference:

1. Introduction to Generalized Fiducial Inference (GFI) (Murph *et al.*, 2022b)
2. GFI on Differentiable Manifolds (Murph *et al.*, 2022a)
3. A Geometric Perspective on Bayesian and Generalized Fiducial Inference (Liu *et al.*, 2022)

Applied Computational Bayesian:

1. Bayesian Change-point Detection for Process Monitoring with Fault Detection

Contents

Generalized Fiducial Inference:

1. Introduction to Generalized Fiducial Inference (GFI) (Murph *et al.*, 2022b)
2. GFI on Differentiable Manifolds (Murph *et al.*, 2022a)
3. A Geometric Perspective on Bayesian and Generalized Fiducial Inference (Liu *et al.*, 2022)

These projects are connected!

Applied Computational Bayesian:

1. Bayesian Change-point Detection for Process Monitoring with Fault Detection

Fiducial Inference

[fiducial statistics was R.A. Fisher's] "bold attempt to make the Bayesian omelette without breaking the Bayesian eggs" -J.Savage, 1961

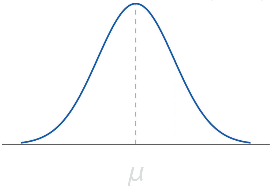
Fiducial Inference

FI Basic Idea:

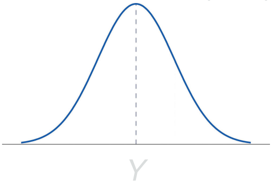
1. Define a structural equation. We will call this a *Data Generating Algorithm* (DGA).
   - Should express how the data were *generated*.
2. Fix the data and invert this equation
3. Get a distribution on the target parameter

Toy Example: a DGA for $Y \sim \mathcal{N}(\mu, 1)$ is $Y = \mu + Z$, $Z \sim \mathcal{N}(0, 1)$.

$$Y = \mu + Z \sim \mathcal{N}(\mu, 1)$$



$$\mu = Y - Z \sim \mathcal{N}(Y, 1)$$



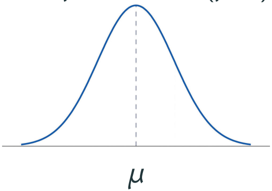
Fiducial Inference

FI Basic Idea:

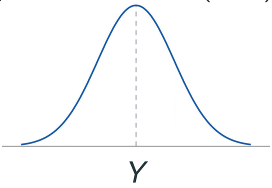
1. Define a structural equation. We will call this a *Data Generating Algorithm* (DGA).
   - Should express how the data were *generated*.
2. Fix the data and invert this equation
3. Get a distribution on the target parameter

Toy Example: a DGA for $Y \sim \mathcal{N}(\mu, 1)$ is $Y = \mu + Z$, $Z \sim \mathcal{N}(0, 1)$.

$Y = \mu + Z \sim \mathcal{N}(\mu, 1)$



$\mu = Y - Z \sim \mathcal{N}(Y, 1)$



Generalized Fiducial Inference (GFI)

This quickly becomes more difficult...

Example: a DGA for $Y \sim$ Gamma(α, β) is $Y = F^{-1}_{\alpha,\beta}(U)$, $U \sim$ Unif(0,1).

Theorem (Hannig *et al.*, 2016)

For continuous data,

$$r_n(\theta) = \frac{f(\mathbf{y}|\theta)J(\mathbf{y},\theta)}{\int f(\mathbf{y}|\theta')\, J(\mathbf{y},\theta')\, d\theta'} \tag{1}$$

where $J(\mathbf{y},\theta) = D\left(\nabla_\theta A(u,\theta)|_{u=A^{-1}(\mathbf{y},\theta)}\right)$, $A(u,\theta)$ the DGA s.t. $Y = A(u,\theta)$, and $D(M) = (\det M'M)^{\frac{1}{2}}$.

Notes on Notation:

- $A(u,\theta)$ is the DGA, e.g. $A(Z,\mu) = \mu + Z$ from before.
- $f(\mathbf{y}|\theta)$ is the distribution of your observed data.

Generalized Fiducial Inference (GFI)

This quickly becomes more difficult...

Example: a DGA for $Y \sim$ Gamma(α, β) is $Y = F^{-1}_{\alpha,\beta}(U)$, $U \sim$ Unif(0,1).

Theorem (Hannig *et al.*, 2016)

For continuous data,

$$r_n(\theta) = \frac{f(\mathbf{y}|\theta)J(\mathbf{y},\theta)}{\int f(\mathbf{y}|\theta')\, J(\mathbf{y},\theta')\, d\theta'} \tag{1}$$

where $J(\mathbf{y},\theta) = D\left(\nabla_\theta A(u,\theta)|_{u=A^{-1}(\mathbf{y},\theta)}\right)$, $A(u,\theta)$ the DGA s.t. $Y = A(u,\theta)$, and $D(M) = (\det M'M)^{\frac{1}{2}}$.

Notes on Notation:

- $A(u,\theta)$ is the DGA, e.g. $A(Z,\mu) = \mu + Z$ from before.
- $f(\mathbf{y}|\theta)$ is the distribution of your observed data.

Generalized Fiducial Inference (GFI)

This quickly becomes more difficult...

Example: a DGA for $Y \sim$ Gamma(α, β) is $Y = F^{-1}_{\alpha,\beta}(U)$, $U \sim$ Unif(0,1).

Theorem (Hannig *et al.*, 2016)

For continuous data,

$$r_n(\theta) = \frac{f(\mathbf{y}|\theta)J(\mathbf{y}, \theta)}{\int f(\mathbf{y}|\theta')\, J(\mathbf{y}, \theta')\, d\theta'} \tag{1}$$

where $J(\mathbf{y}, \theta) = D\left(\nabla_\theta A\,(u, \theta)|_{u=A^{-1}(\mathbf{y},\theta)}\right)$, $A(u, \theta)$ the DGA s.t. $Y = A(u, \theta)$, and $D(M) = (\det M'M)^{\frac{1}{2}}$.

Notes on Notation:

- $A(u, \theta)$ is the DGA, e.g. $A(Z, \mu) = \mu + Z$ from before.
- $f(\mathbf{y}|\theta)$ is the distribution of your observed data.

Generalized Fiducial Inference (GFI)

This quickly becomes more difficult...

Example: a DGA for $Y \sim$ Gamma(α, β) is $Y = F^{-1}_{\alpha,\beta}(U)$, $U \sim$ Unif(0,1).

Theorem (Hannig *et al.*, 2016)

For continuous data,

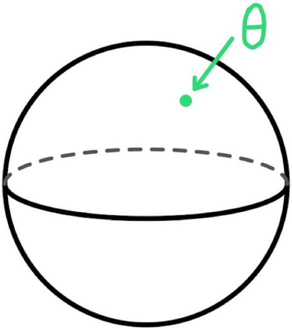
$$r_n(\theta) = \frac{f(\mathbf{y}|\theta)J(\mathbf{y},\theta)}{\int f(\mathbf{y}|\theta')\,J(\mathbf{y},\theta')\,d\theta'} \tag{1}$$

where $J(\mathbf{y},\theta) = D\left(\nabla_\theta A\left(u,\theta\right)\big|_{u=A^{-1}(\mathbf{y},\theta)}\right)$, $A(u,\theta)$ the DGA s.t. $Y = A(u,\theta)$, and $D(M) = (\det M'M)^{\frac{1}{2}}$.

Notes on Notation:

- $A(u,\theta)$ is the DGA, e.g. $A(Z,\mu) = \mu + Z$ from before.
- $f(\mathbf{y}|\theta)$ is the distribution of your observed data.

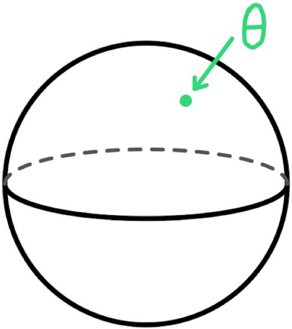
GFI on Constrained Parameter Spaces



$\leftrightarrow$ $g(\theta) = ||\theta|| - 1 = 0$

In general, we say $\theta \in \mathcal{M} \leftrightarrow g(\theta) = 0$ for $g : \mathbb{R}^d \to \mathbb{R}^t$.
Equivalently, $g^{-1}\{0\} = \mathcal{M}$.

GFI on Constrained Parameter Spaces



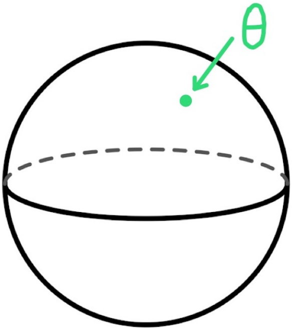
$\leftrightarrow$

$$g(\theta) = ||\theta|| - 1 = 0$$

In general, we say $\theta \in \mathcal{M} \leftrightarrow g(\theta) = 0$ for $g : \mathbb{R}^d \to \mathbb{R}^t$.

Equivalently, $g^{-1}\{0\} = \mathcal{M}$.

GFI on Constrained Parameter Spaces

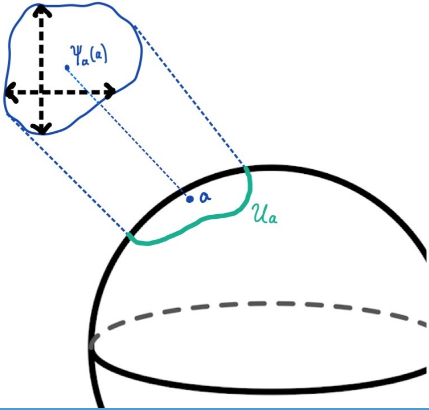


$\leftrightarrow$ $\quad g(\theta) = ||\theta|| - 1 = 0$

In general, we say $\theta \in \mathcal{M} \leftrightarrow g(\theta) = 0$ for $g : \mathbb{R}^d \to \mathbb{R}^t$.
Equivalently, $g^{-1}\{0\} = \mathcal{M}$.

GFI on Constrained Parameter Spaces

Assuming $\nabla_\theta g \neq 0$ everywhere, the level set $\mathcal{M} := g^{-1}\{0\}$ is a **Riemannian Manifold**.

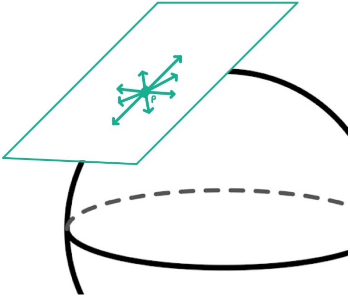


- Locally like $\mathbb{R}^{d-t}$ at every point a on the manifold;
- There exists an inner product on the tangent vectors at a;
- We can meaningfully define the derivative on a manifold.

GFI on Constrained Parameter Spaces

GFD Jacobian term

$$J(\mathbf{y}, \theta) = D\left(\nabla_\theta A(u, \theta)\big|_{u=A^{-1}(\mathbf{y},\theta)} \right).$$

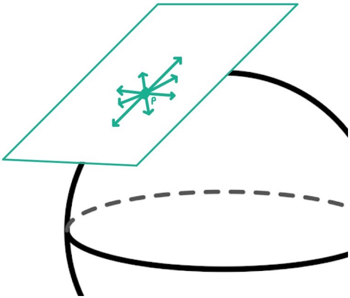


GFI on Constrained Parameter Spaces

GFD Jacobian term

$$J(\mathbf{y}, \theta) = D\left(\nabla_\theta A(u, \theta)|_{u=A^{-1}(\mathbf{y},\theta)}\right).$$

This is no longer a gradient on the *ambient* space and requires more thought.



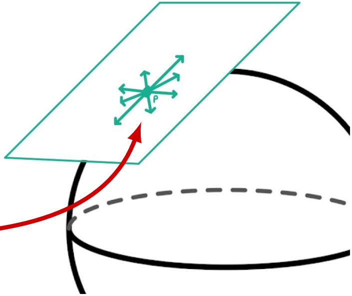
GFI on Constrained Parameter Spaces

GFD Jacobian term

$$J(\mathbf{y}, \theta) = D\left(\nabla_\theta A(u, \theta)|_{u=A^{-1}(\mathbf{y},\theta)} \right).$$

This is no longer a gradient on the *ambient* space and requires more thought.

We take the derivative along the tangent directions.



GFI on Constrained Parameter Spaces

First Idea

Project this gradient onto the manifold using

$$P_\theta := I_d - (\nabla_\theta g)'(\nabla_\theta g(\nabla_\theta g)')^{-1}\nabla_\theta g.$$

to get

$$J(\mathbf{y},\theta) = D\left(\nabla_\theta A\,(u,\theta)|_{u=A^{-1}(\mathbf{y},\theta)}\,P_\theta\right).$$

Recall that $D(M) = (\det M'M)^{\frac{1}{2}}$.

But then $J(\mathbf{y},\theta) = 0$ since $\nabla_\theta A\,(u,\theta)|_{u=A^{-1}(\mathbf{y},\theta)}\,P_\theta$ is not full rank!

GFI on Constrained Parameter Spaces

Constrained GFD (CGFD)

$$r_{n,\mathcal{M}}(\theta|\mathbf{y}) = \frac{f(y|\theta)D^*\left(\nabla_\theta A(w,\theta)|_{w=A^{-1}(\mathbf{y},\theta)}P_\theta\right)}{\int_{\mathcal{M}} f(y|\theta^*)D^*\left(\nabla_{\theta^*} A(w,\theta^*)|_{w=A^{-1}(\mathbf{y},\theta^*)}P_{\theta^*}\right)d\lambda(\theta^*)},$$

for $\theta \in \mathcal{M}$, where $D^*(M) = (\operatorname{pdet} M'M)^{\frac{1}{2}}$, the pdet operator denotes the pseudodeterminant, and the integral in the denominator is taken over the manifold $\mathcal{M}$.

- Note that the CGFD is unnecessary when a parameterization to an unconstrained space is known.
- E.g. the CGFD $r_{n,\mathcal{M}}(x,y,z)$ can be defined on the unit sphere, but you could equally define the GFD on this sphere using polar coordinates, $r_n(\varphi,\theta)$.

GFI on Constrained Parameter Spaces

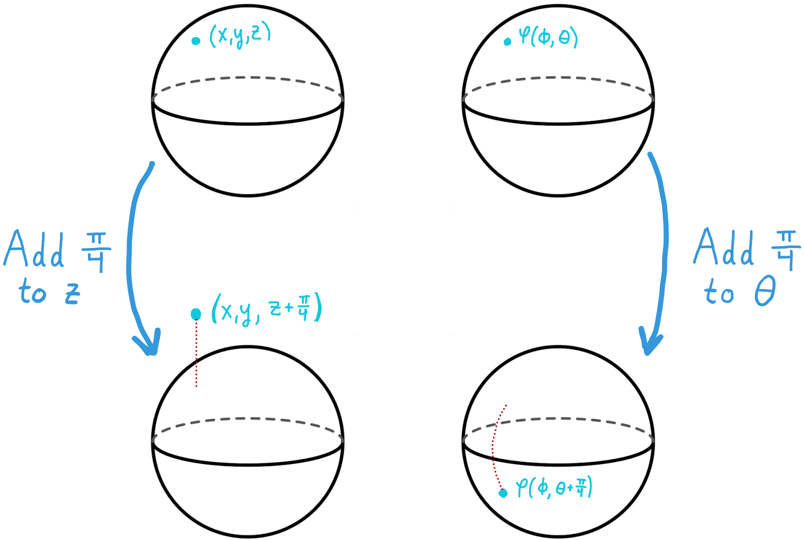
Constrained GFD (CGFD)

$$r_{n,\mathcal{M}}(\theta|\mathbf{y}) = \frac{f(y|\theta)D^{*}\left(\nabla_{\theta}A(w,\theta)|_{w=A^{-1}(\mathbf{y},\theta)}P_{\theta}\right)}{\int_{\mathcal{M}} f(y|\theta^{*})D^{*}\left(\nabla_{\theta^{*}}A(w,\theta^{*})|_{w=A^{-1}(\mathbf{y},\theta^{*})}P_{\theta^{*}}\right)d\lambda(\theta^{*})},$$

This is the Riemann–Lebesgue volume measure

fo et $M'M)^{\frac{1}{2}}$, the pdet operator denotes th integral in the denominator is taken over the manifold $\mathcal{M}$.

- Note that the CGFD is unnecessary when a parameterization to an unconstrained space is known.
- E.g. the CGFD $r_{n,\mathcal{M}}(x,y,z)$ can be defined on the unit sphere, but you could equally define the GFD on this sphere using polar coordinates, $r_n(\varphi,\theta)$.



$$\varphi(\varphi, \theta) = (\cos\theta \sin\varphi \ ; \ \sin\theta \sin\varphi \ ; \ \cos\varphi)'$$

Theoretical Qualities of CGFD

"Theorem Idea:" The CGFD is equivalent to the CFD defined using a parameterization.

Theorem (Invariant to parameterization)

- Fix a point $\theta \in \mathcal{M}$ and coordinate chart $(\psi_\theta, \mathcal{U}_\theta)$, such that $\theta \in \mathcal{U}_\theta$.
- Let $V \times W \subset \mathbb{R}^d$ be the open set around the point θ on which the implicit function $\varphi : V \to W$ exists and is diffeomorphic.

Then for any open set $M_\theta \subseteq \mathcal{U}_\theta \cap (V \times W)$,

$$\int_{M_\theta} f(y|\theta^*) D^* \left((\nabla_\theta A(w, \theta^*)) P_\theta \right) d\lambda(\theta^*) = \int_{\psi_\theta(M_\theta)} f(y|\psi_\theta^{-1}(u)) D \left(\nabla_u A(w, \psi_\theta^{-1}(u)) \right) du$$

Theoretical Qualities of CGFD

“Theorem Idea:" The CGFD is equivalent to the CFD defined using a parameterization.

Theorem (Invariant to parameterization)

- Fix a point $\theta \in \mathcal{M}$ and coordinate chart $(\psi_\theta, \mathcal{U}_\theta)$, such that $\theta \in \mathcal{U}_\theta$.
- Let $V \times W \subset \mathbb{R}^d$ be the open set around the point θ on which the implicit function $\varphi : V \to W$ exists and is diffeomorphic.

Then for any open set $M_\theta \subseteq \mathcal{U}_\theta \cap (V \times W)$,

$$\int_{M_\theta} f(y|\theta^*) D^* \left((\nabla_\theta A(w, \theta^*)) P_\theta \right) d\lambda(\theta^*) = \int_{\psi_\theta(M_\theta)} f(y|\psi_\theta^{-1}(u)) D \left(\nabla_u A(w, \psi_\theta^{-1}(u)) \right) du$$

Sphere Example

- $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m \sim \mathcal{N}_3(\mu, I)$ with DGA:

$$\mathbf{Y}_i = \mu + \mathbf{Z}_i, \quad i = 1, \ldots, m,$$

  where $Z_i \overset{i.i.d.}{\sim} \mathcal{N}_3(0, I)$.
- Consider constraint function $g(\mu) = ||\mu|| - 1$, the mean vector on the unit sphere, and $\mu := \begin{pmatrix} \mu_1 & \mu_2 & \mu_3 \end{pmatrix}'$.
- We can also parameterize this problem directly with $\mu := \begin{pmatrix} \cos\theta \sin\varphi \\ \sin\theta \sin\varphi \\ \cos\varphi \end{pmatrix}$ and define a fiducial density on θ, φ.

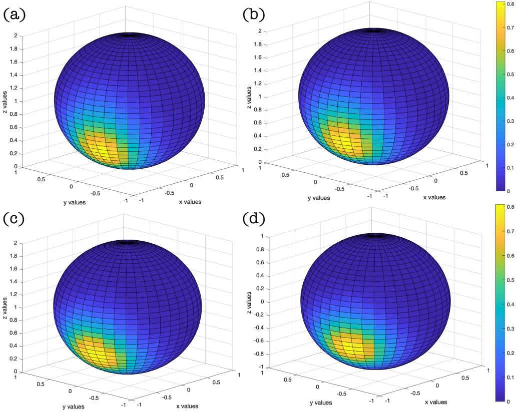
Sphere Example

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix};$$

explicit

$$\mu = \begin{pmatrix} \cos\theta \sin\varphi \\ \sin\theta \sin\varphi \\ \cos\varphi \end{pmatrix};$$

explicit



$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix};$$

sampled with HMC

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix};$$

sampled with MH

Theoretical Qualities of CGFD

Lemma (Invariance of g)

Let $g^{-1}(\{0\}) = h^{-1}(\{0\}) = \mathcal{M}$. Then, for all $\theta \in \mathcal{M}$,

$$r_{n,g^{-1}(\{0\})}(\theta) = r_{n,h^{-1}(\{0\})}(\theta).$$

- E.g. $g(x) = 1 - ||x||$ and $h(x) = c(1 - ||x||)$ for $c \in \mathbb{R} \backslash \{0\}$.
- *Proof idea.* Write P_θ in terms of the implicit function.

Bernstein–von Mises Inheritance

“Theorem Idea:" Local Asymptotic Normality for the ambient GFD is inherited by the CGFD.

Theorem: Local Asymptotic Normality Inheritance

Recall the local parameter: $s = \sqrt{n}(\theta - \theta_0)$. Under technical assumptions,

$$\int_{\mathbb{R}^d} \left| r^*_{\mathbf{n}}(s) - \varphi^*_{0,I(\theta_0)}(s) \right| ds \overset{n}{\to} 0 \ \text{ implies that}$$

$$\int_{\mathcal{M}} \left| r^*_{\mathbf{n},\mathcal{M}}(s) - \varphi^*_{0,I(\theta_0),\mathcal{M}}(s) \right| d\lambda(s) \overset{n}{\to} 0,$$

where $q^*(s) := n^{-d/2} q(n^{-1/2} s + \theta_0)$ for any function q.

Bernstein–von Mises Inheritance

"Theorem Idea:" Local Asymptotic Normality for the ambient GFD is inherited by the CGFD.

Theorem: Local Asymptotic Normality Inheritance

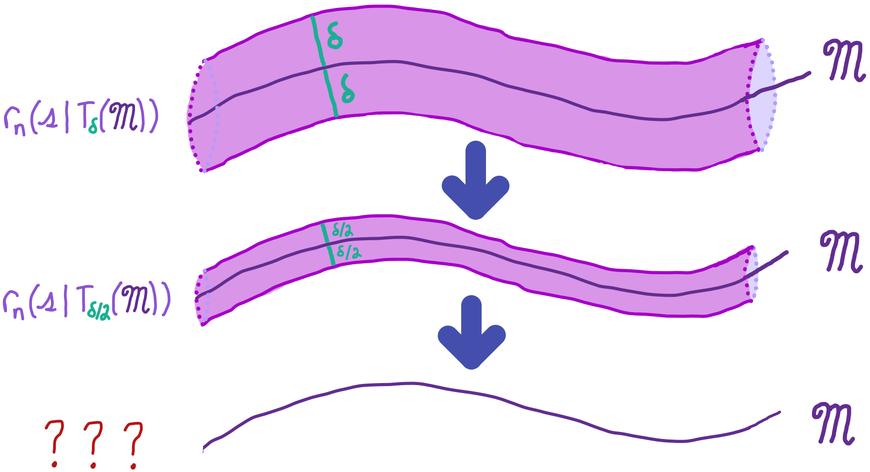
Recall the local parameter: $s = \sqrt{n}(\theta - \theta_0)$. Under technical assumptions,

$$\int_{\mathbb{R}^d} \left| r^*_{\mathbf{n}}(s) - \varphi^*_{0,I(\theta_0)}(s) \right| ds \overset{n}{\to} 0 \text{ implies that}$$

$$\int_{\mathcal{M}} \left| r^*_{\mathbf{n},\mathcal{M}}(s) - \varphi^*_{0,I(\theta_0),\mathcal{M}}(s) \right| d\lambda(s) \overset{n}{\to} 0,$$

where $q^*(s) := n^{-d/2} q(n^{-1/2}s + \theta_0)$ for any function q.

One Idea



$\Gamma_n(s \mid T_\delta(\mathcal{M}))$
δ
δ
$\mathcal{M}$
$\Gamma_n(s \mid T_{\delta/2}(\mathcal{M}))$
$\delta/2$
$\delta/2$
$\mathcal{M}$
???
$\mathcal{M}$

Bernstein–von Mises Inheritance

Theorem (Hwang, 1980)

Under the technical assumptions from Hwang (1980), the limiting density is

$$\frac{dP}{d\lambda}(u) = \frac{r_n(\theta|\mathbf{y}) \det\left(\nabla_\theta g (\nabla_\theta g)^T\right)^{-1/2}}{\int_{\mathcal{M}} r_n(\theta|\mathbf{y}) \det\left(\nabla_\theta g (\nabla_\theta g)^T\right)^{-1/2} d\lambda}.$$

We'll call this the *extrinsic* constrained GFD.

The extrinsic constrained GFD is **not** the same as the CGFD (which, in contrast, is *intrinsic*).

This result requires that the manifold be compact.

Bernstein–von Mises Inheritance

E.g.: Consider the equal means problem: two independent observations X_1, X_2 with DGA,

$$X_i = A(Z_i, \theta) := \mu + \Sigma^{1/2} Z_i, \tag{2}$$

where $\mu = (\mu_1, \mu_2)'$, $\Sigma^{1/2} = \texttt{diag}(\sigma_1, \sigma_2)$, $\theta := (\mu', \sigma_1, \sigma_2)'$, and $Z_i \overset{iid}{\sim} \mathcal{N}_2(0, I_2)$ for $i \in \{1, 2\}$.

1. For $g(\theta) := \mu_2 - \mu_1$, $\det\left(\nabla_\theta g (\nabla_\theta g)^T\right)^{-1/2} = 1$,
2. For $h(\theta) = \mu_2^3 - \mu_1^3$, $\det\left(\nabla_\theta h (\nabla_\theta h)^T\right)^{-1/2} = (9\mu_1^4 + 9\mu_2^4)^{-1/2}$.

Bernstein–von Mises Inheritance

The extrinsic constrained GFD is useful for understanding how ambient qualities are inherited by submanifold densities.

Proof outline:

- Prove that the extrinsic constrained GFD "inherits“ the Bernstein-von Mises property when satisfied by the ambient GFD;
- Show that the extrinsic constrained GFD and the intrinsic CGFD are indistinguishable in the local limit on a compact subset around the truth;
- Assume the intrinsic CGFD decays off of the compact subset of the manifold (satisfied for i.i.d. with bounded DGA Jacobian).

Bernstein–von Mises Inheritance

The extrinsic constrained GFD is useful for understanding how ambient qualities are inherited by submanifold densities.

Proof outline:

- Prove that the extrinsic constrained GFD "inherits“ the Bernstein-von Mises property when satisfied by the ambient GFD;
- Show that the extrinsic constrained GFD and the intrinsic CGFD are indistinguishable in the local limit on a compact subset around the truth;
- Assume the intrinsic CGFD decays off of the compact subset of the manifold (satisfied for i.i.d. with bounded DGA Jacobian).

MAYO CLINIC

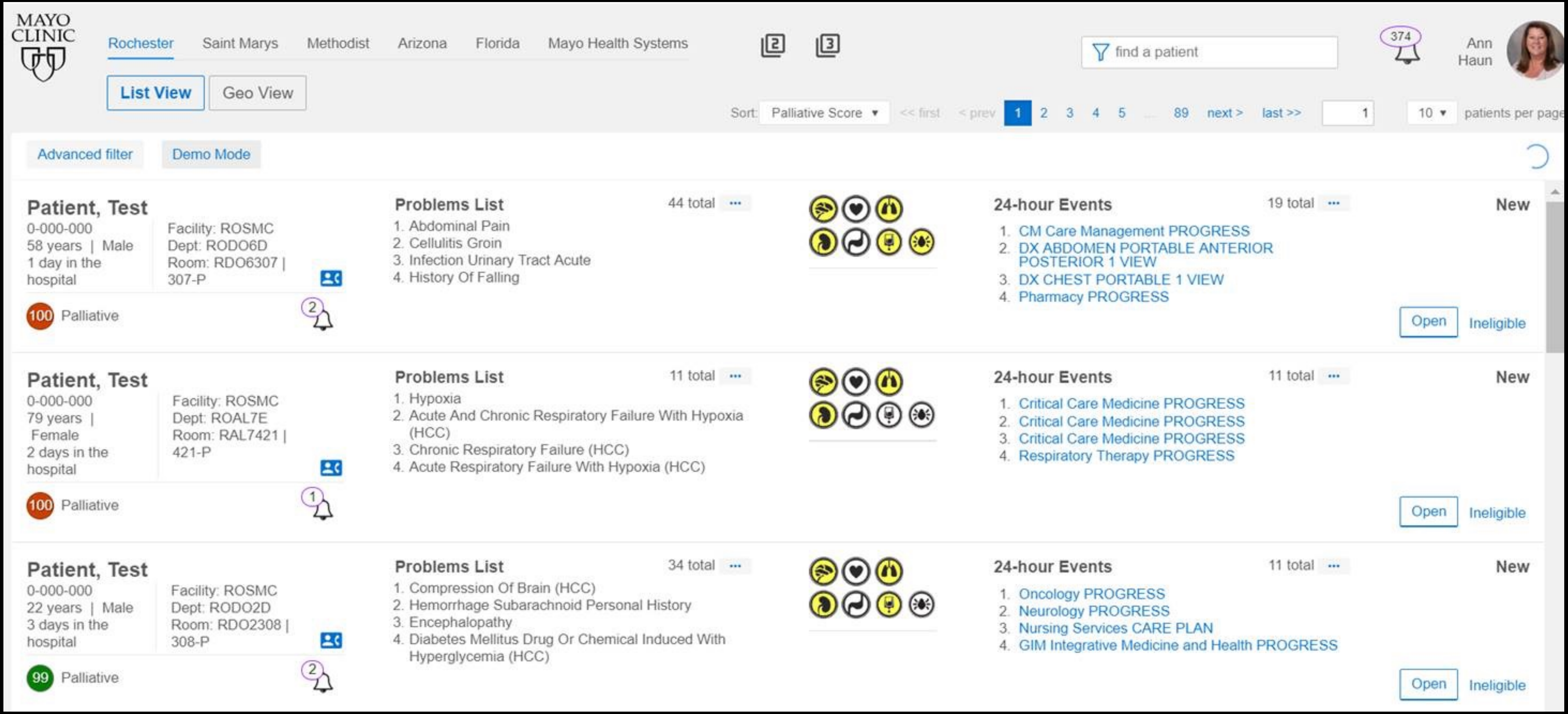
BAYESIAN CHANGE-POINT DETECTION FOR PROCESS MONITORING WITH FAULT DETECTION

KERN CENTER FOR THE SCIENCE OF HEALTH CARE DELIVERY

Alexander C. Murph

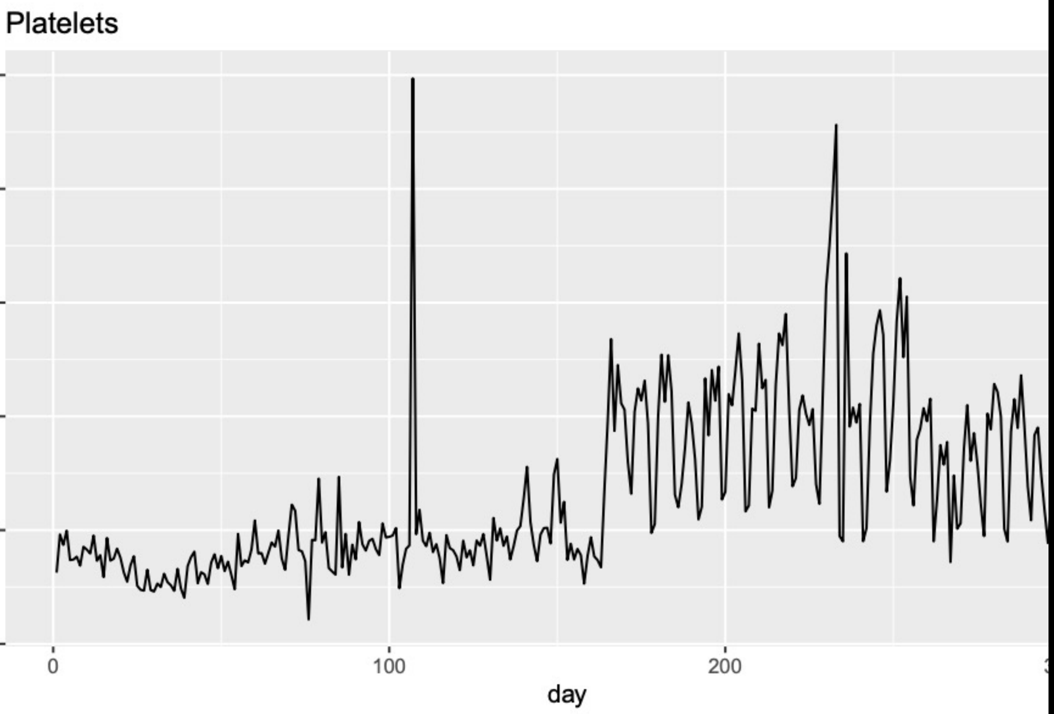
April 21st, 2023

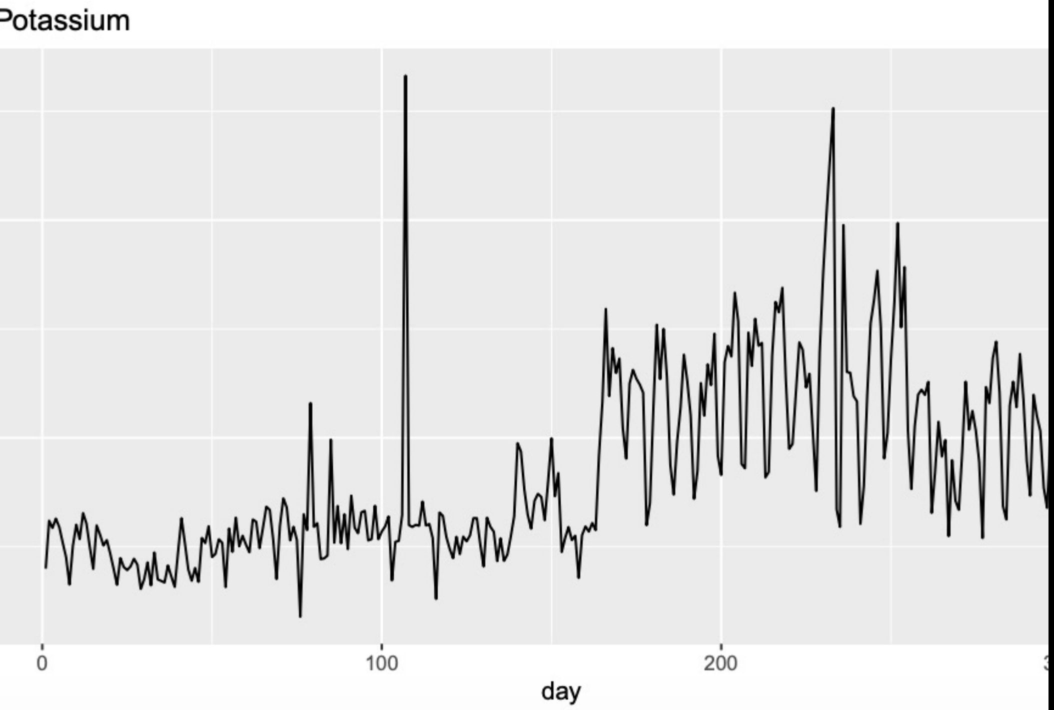
CONTROL TOWER INTERFACE



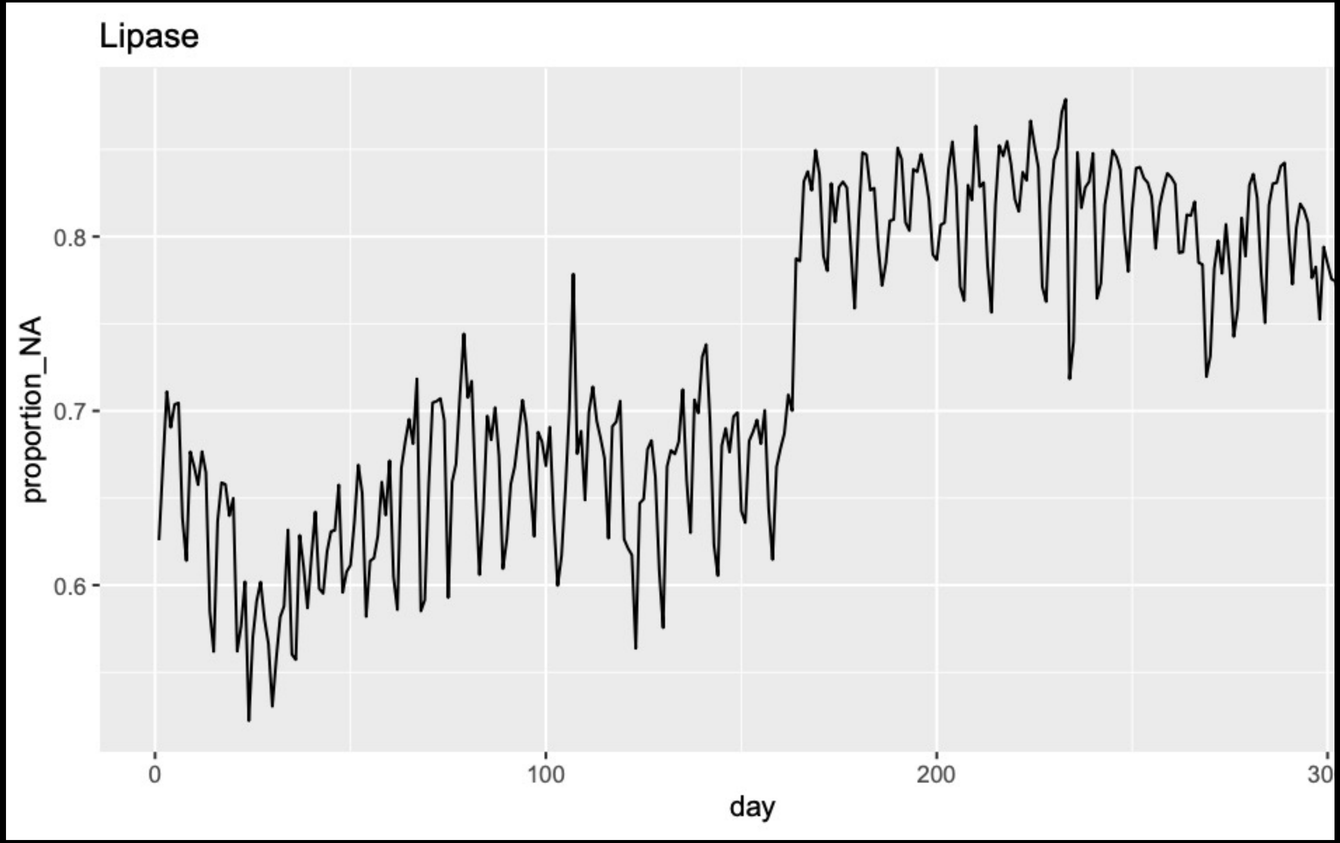
Number in the circle:
"The probability of getting an inpatient palliative care consultation in the next 7 days."

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-Sub) problems:

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

(Sub) solutions:

- Bayesian Latent Variables

- Gaussian Graphical Models (GGMs)

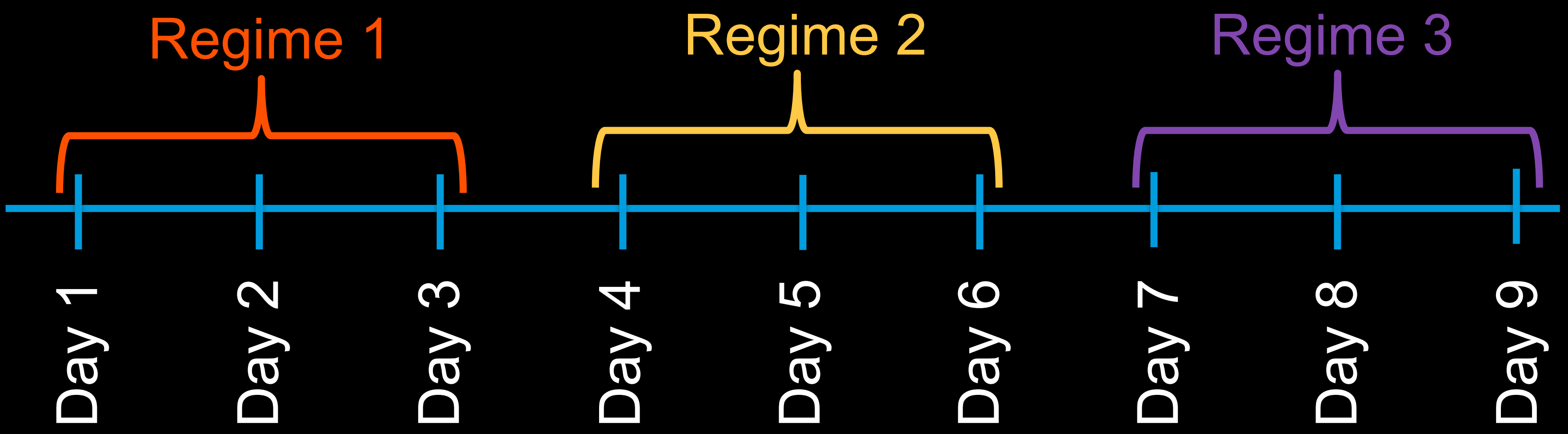
(Sub-Sub) solutions:

- Bayesian Hierarchical Model

- Double Reversible Jump Metropolis Hastings, Conjugate priors.

HOW DO WE FIND A CHANGE POINT?

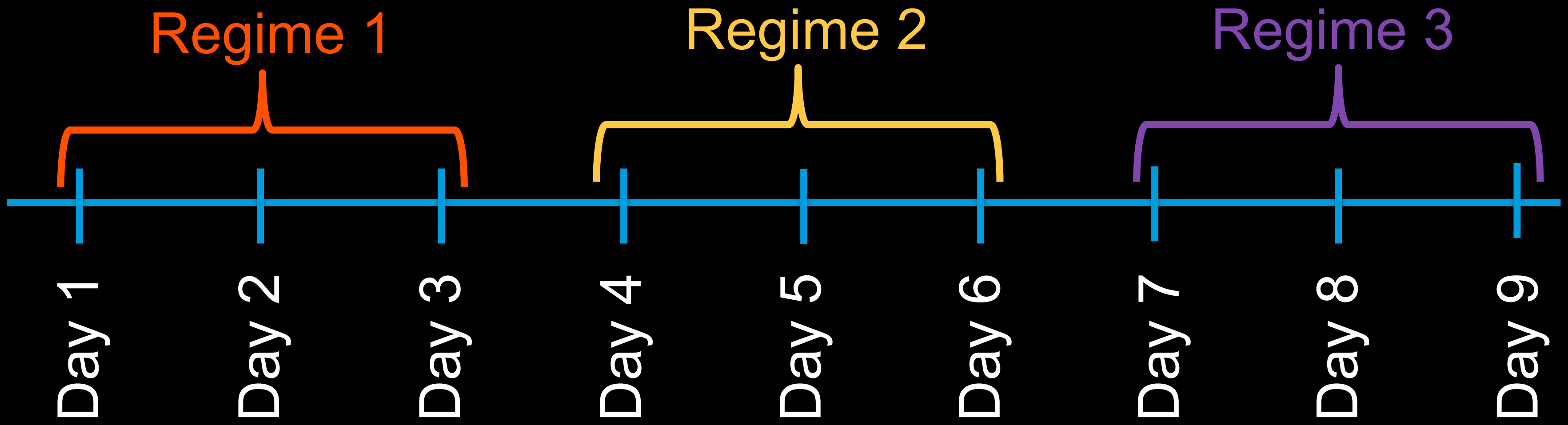
LET'S CLASSIFY OBSERVATIONS OVER TIME INTO "REGIMES"



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:



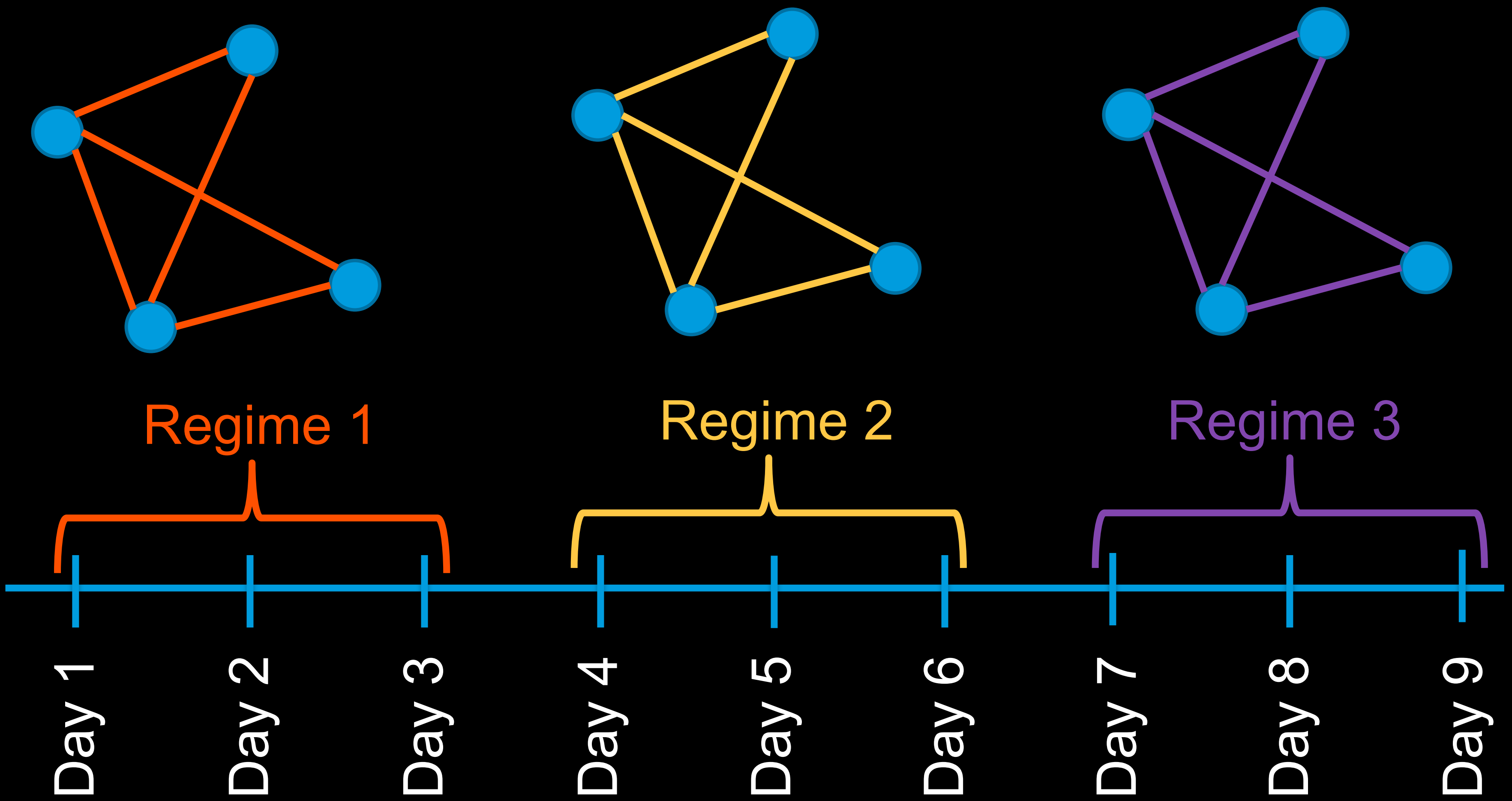
Data we observed on ______

becomes

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

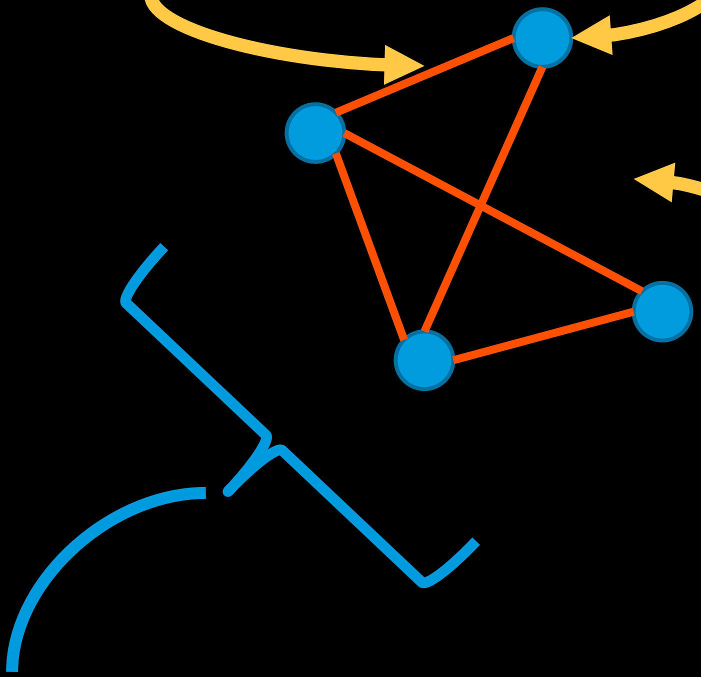
WHAT'S THE LIKELIHOOD?

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.

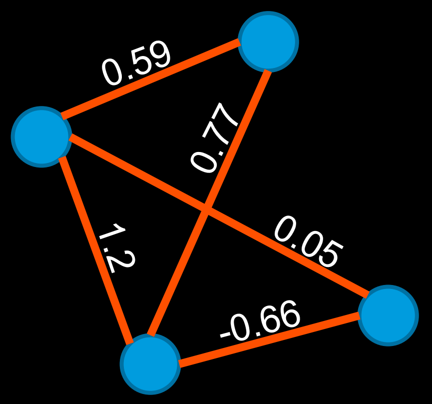
The nodes correspond to the variables in our dataset.



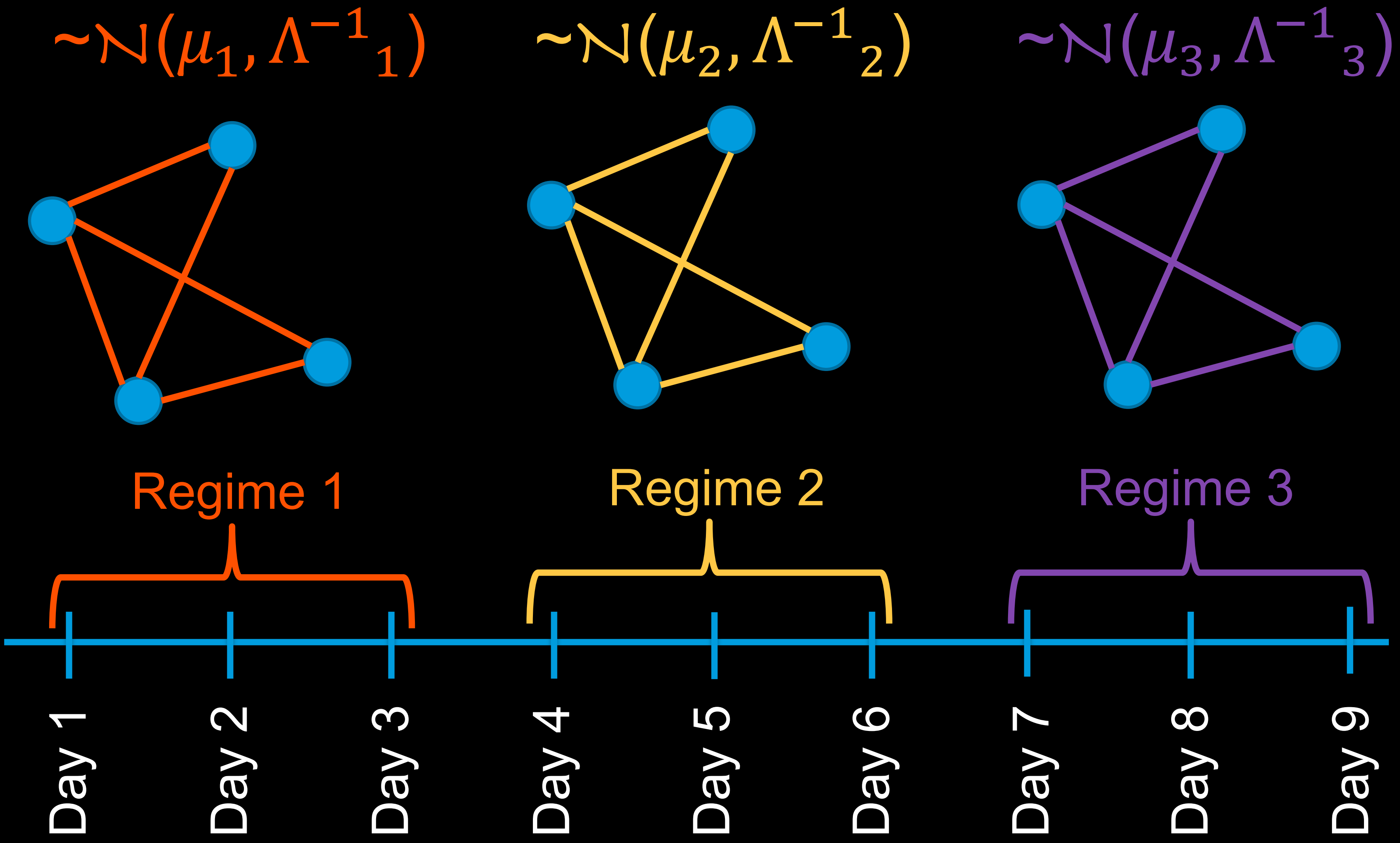
The *absence* of an edge corresponds to a zero in the precision matrix

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

FOR INSTANCE,



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



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

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.

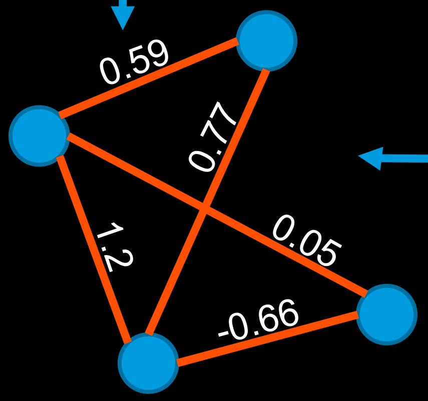
1: Set $\Omega = \Sigma$
2: **repeat**
3: **for** $i = 1, \ldots, p$ **do**
4: Let $N_i \subset V$ be the neighbor set of node i in G. Form Ω_{N_i} and $\Sigma_{N_i,i}$ and solve $\hat{\beta}_i^* = \Omega_{N_i}^{-1}\Sigma_{N_i,i}$.
5: Form $\hat{\beta}_i \in R^{p-1}$ by padding the elements of $\hat{\beta}_i^*$ 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)

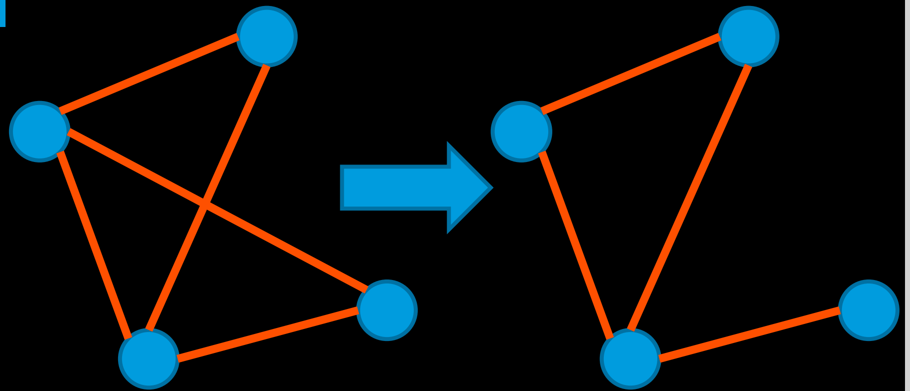
There exists algorithms to sample from a G-Wishart *without* having to calculate the normalizing constant!

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)

⋮

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)

⋮

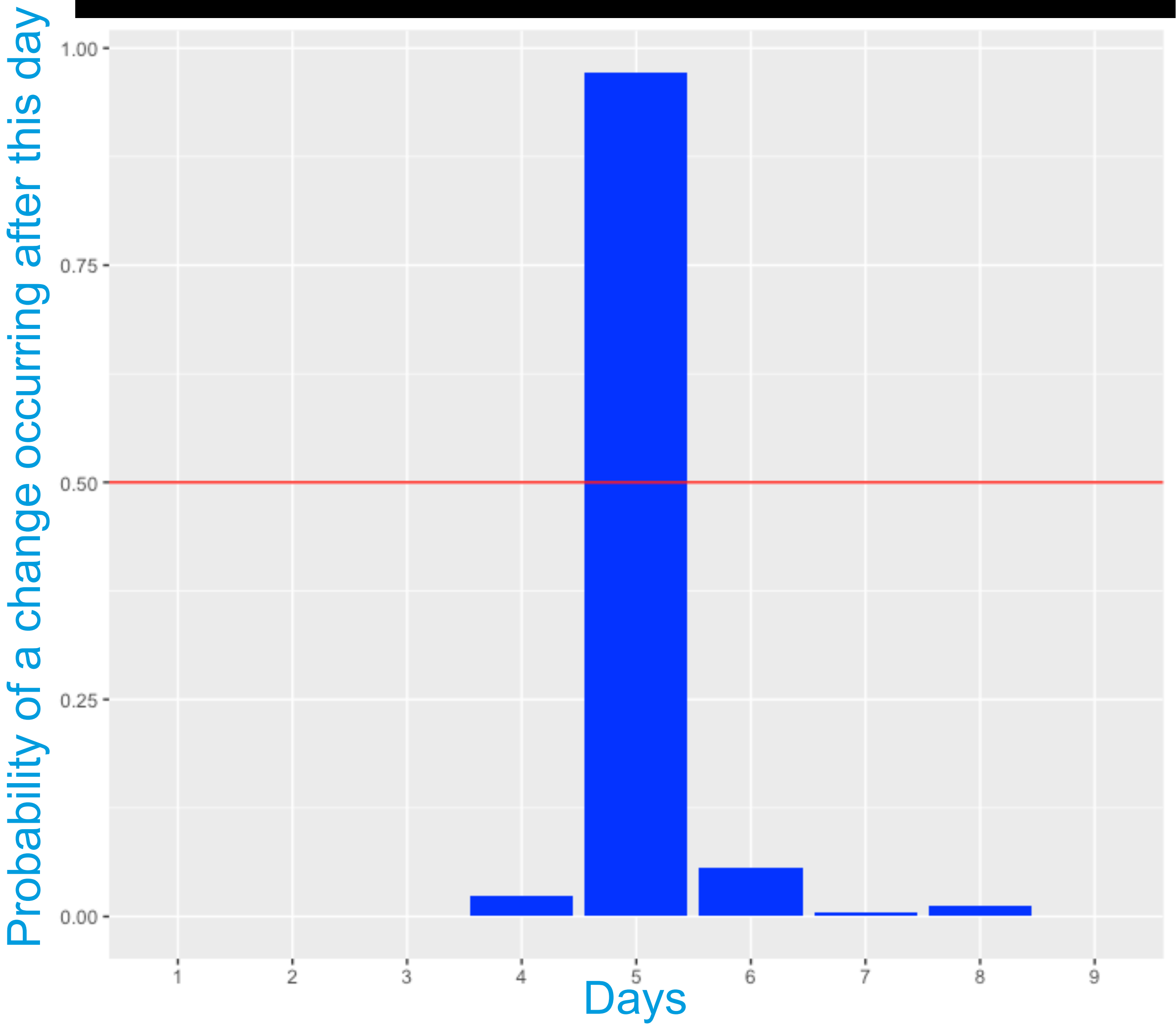
$\Rightarrow$

- (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

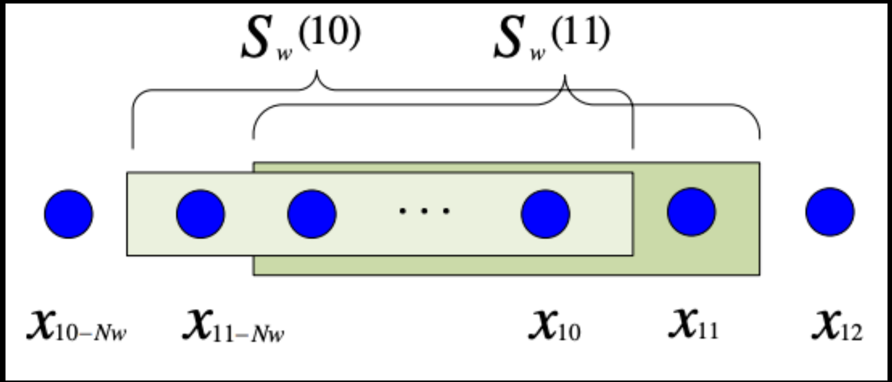
$T^2_{p,v}$ Hotelling T^2 statistic, with pooled degrees of freedom v and number of dimensions p

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
B: Continuous, No Missing Values, Unimodal, MeanChange
C: Continuous, No Missing Values, Bimodal, Mean Change
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 %
A: T2 Scan Stat 0.0592; RTM GBM 0.0485; Mixture Model 0.0592; MVN Model 0.0577
B: T2 Scan Stat 0.0508; RTM GBM 0.0515; Mixture Model 0.0577; MVN Model 0.0546
C: T2 Scan Stat 0.0385; RTM GBM 0.0592; Mixture Model 0.0538; MVN Model 0.0631
D: 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

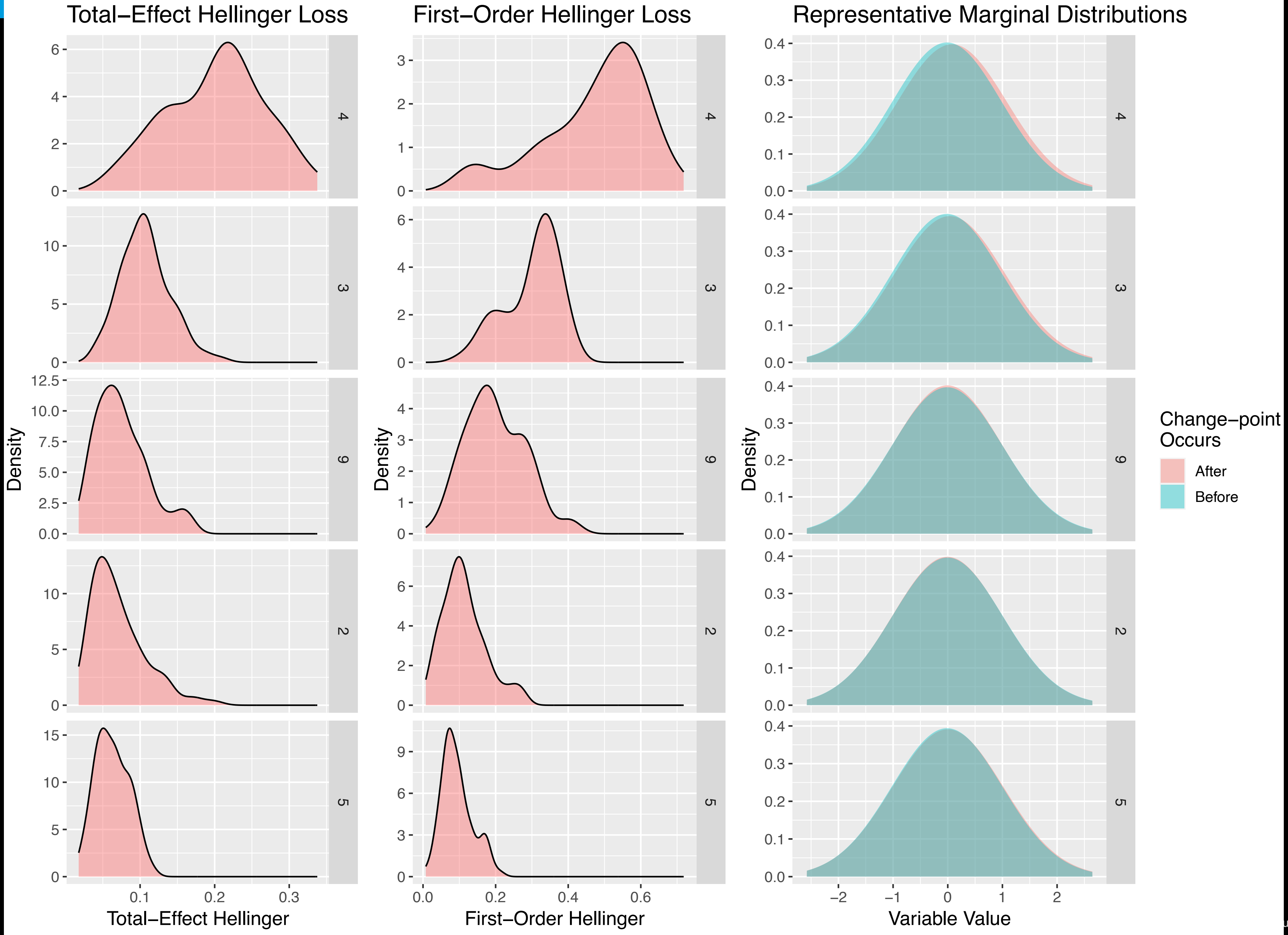
then we Define:

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

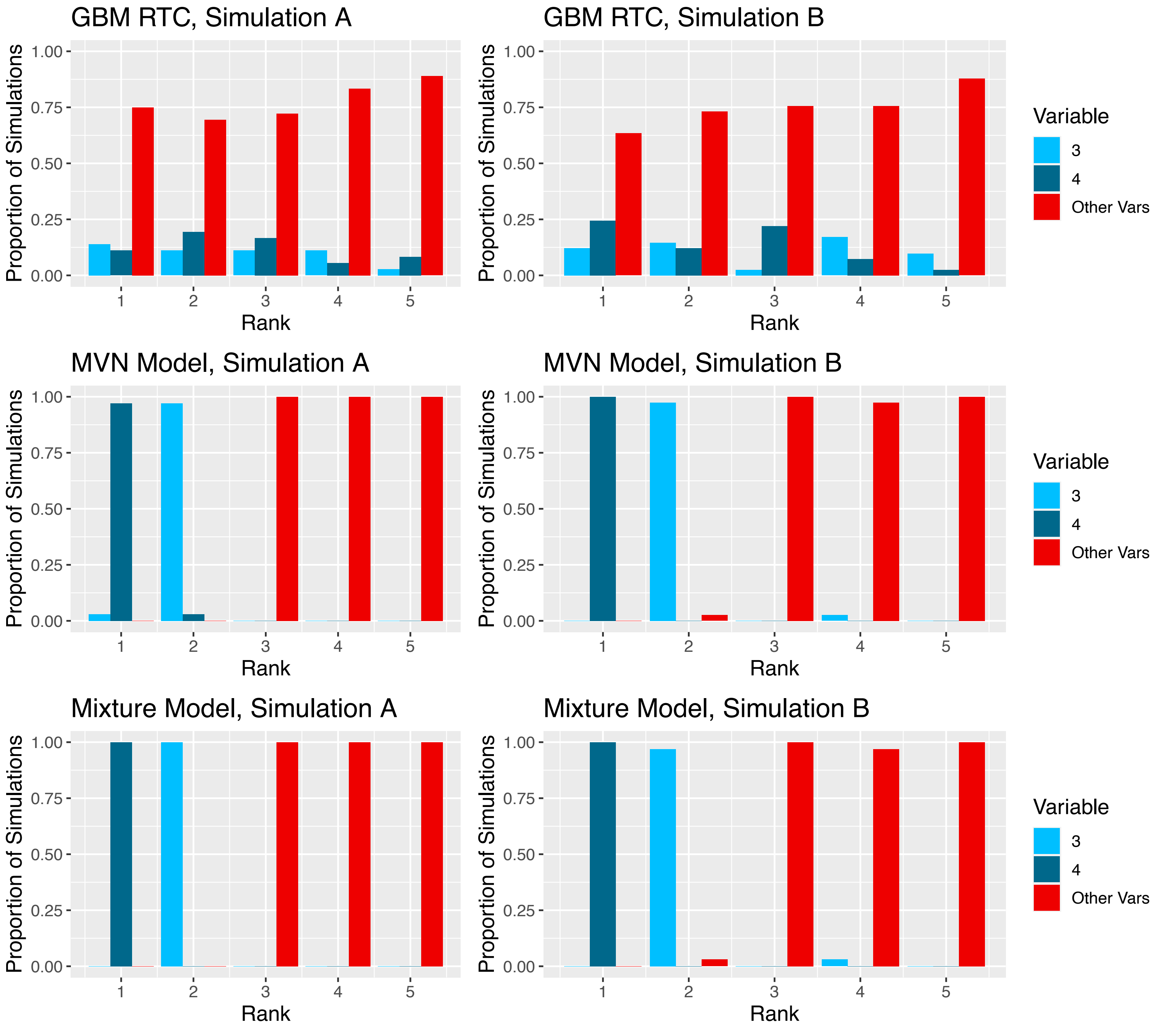
First-Order Loss of i: H(X : i, Y : i)/H(X,Y)

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
Rank
Variable
3
4
Other Vars

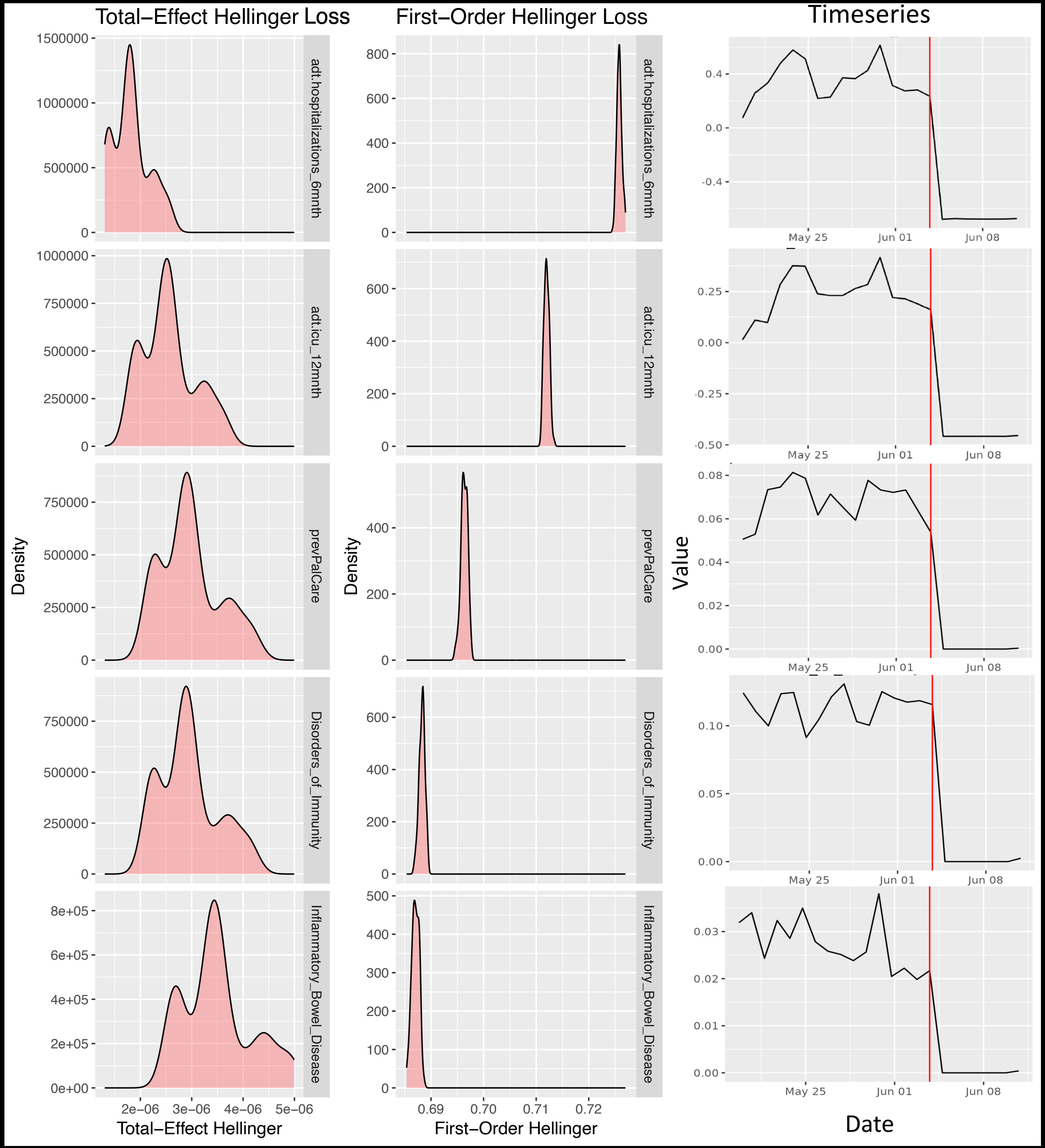
DATA APPLICATION

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

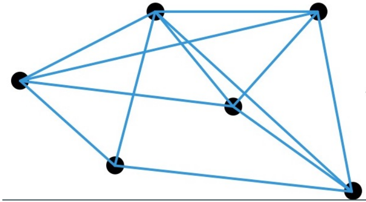


QUESTIONS?

Gaussian Graphical Models

- $X \sim \mathcal{N}_d\left(\mu, \Lambda\right)$ where $\Lambda := \Sigma^{-1}$;
- $G = (V, E)$ be an undirected graph, where V and E are the sets of nodes and edges, respectively. Nodes $i, j \in V$ are connected $\leftrightarrow (i, j) \in E$.

X is said to *satisfy the undirected Gaussian Graphical Model with graph* G if $\Lambda_{i,j} = 0$ for all $E_{i,j} = 0$.



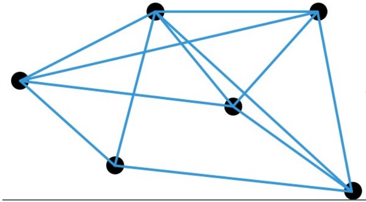
$$\Leftrightarrow \begin{pmatrix} 5.61 & 0.80 & 2.33 & 0.03 & 2.18 & 0.00 \\ 0.80 & 6.66 & 0.15 & -0.06 & 1.78 & 3.04 \\ 2.33 & 0.15 & 3.16 & 0.00 & 0.90 & -0.74 \\ 0.03 & -0.06 & 0.00 & 3.93 & 0.00 & 2.58 \\ 2.18 & 1.78 & 0.90 & 0.00 & 4.57 & 2.39 \\ 0.00 & 3.04 & -0.74 & 2.58 & 2.39 & 5.07 \end{pmatrix}$$ [1]

[1] example numbers from Kundu *et al.* (2019)

Gaussian Graphical Models

- $X \sim \mathcal{N}_d(\mu, \Lambda)$ where $\Lambda := \Sigma^{-1}$;
- $G = (V, E)$ be an undirected graph, where V and E are the sets of nodes and edges, respectively. Nodes $i, j \in V$ are connected $\leftrightarrow (i,j) \in E$.

X is said to *satisfy the undirected Gaussian Graphical Model with graph G* if $\Lambda_{i,j} = 0$ for all $E_{i,j} = 0$.



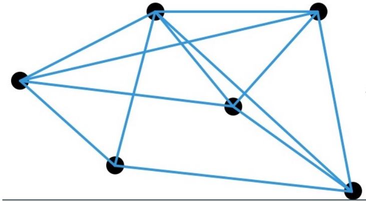
$$\Leftrightarrow \begin{pmatrix} 5.61 & 0.80 & 2.33 & 0.03 & 2.18 & 0.00 \\ 0.80 & 6.66 & 0.15 & -0.06 & 1.78 & 3.04 \\ 2.33 & 0.15 & 3.16 & 0.00 & 0.90 & -0.74 \\ 0.03 & -0.06 & 0.00 & 3.93 & 0.00 & 2.58 \\ 2.18 & 1.78 & 0.90 & 0.00 & 4.57 & 2.39 \\ 0.00 & 3.04 & -0.74 & 2.58 & 2.39 & 5.07 \end{pmatrix}$$ [1]

[1] example numbers from Kundu *et al.* (2019)

Gaussian Graphical Models

- $X \sim \mathcal{N}_d(\mu, \Lambda)$ where $\Lambda := \Sigma^{-1}$;
- $G = (V, E)$ be an undirected graph, where V and E are the sets of nodes and edges, respectively. Nodes $i, j \in V$ are connected $\leftrightarrow (i, j) \in E$.

X is said to *satisfy the undirected Gaussian Graphical Model with graph G* if $\Lambda_{i,j} = 0$ for all $E_{i,j} = 0$.



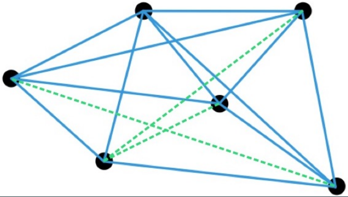
$$\Leftrightarrow \begin{pmatrix} 5.61 & 0.80 & 2.33 & 0.03 & 2.18 & 0.00 \\ 0.80 & 6.66 & 0.15 & -0.06 & 1.78 & 3.04 \\ 2.33 & 0.15 & 3.16 & 0.00 & 0.90 & -0.74 \\ 0.03 & -0.06 & 0.00 & 3.93 & 0.00 & 2.58 \\ 2.18 & 1.78 & 0.90 & 0.00 & 4.57 & 2.39 \\ 0.00 & 3.04 & -0.74 & 2.58 & 2.39 & 5.07 \end{pmatrix}$$ [1]

[1] example numbers from Kundu *et al.* (2019)

Gaussian Graphical Models

- $X \sim \mathcal{N}_d(\mu, \Lambda)$ where $\Lambda := \Sigma^{-1}$;
- $G = (V, E)$ be an undirected graph, where V and E are the sets of nodes and edges, respectively. Nodes $i, j \in V$ are connected $\leftrightarrow (i, j) \in E$.

X is said to *satisfy the undirected Gaussian Graphical Model with graph G* if $\Lambda_{i,j} = 0$ for all $E_{i,j} = 0$.



$$\Leftrightarrow \begin{pmatrix} 5.61 & 0.80 & 2.33 & 0.03 & 2.18 & \boxed{0.00} \\ 0.80 & 6.66 & 0.15 & -0.06 & 1.78 & 3.04 \\ 2.33 & 0.15 & 3.16 & \boxed{0.00} & 0.90 & -0.74 \\ 0.03 & -0.06 & \boxed{0.00} & 3.93 & \boxed{0.00} & 2.58 \\ 2.18 & 1.78 & 0.90 & \boxed{0.00} & 4.57 & 2.39 \\ \boxed{0.00} & 3.04 & -0.74 & 2.58 & 2.39 & 5.07 \end{pmatrix}$$ [2]

[2] example numbers from Kundu *et al.* (2019)

Gaussian Graphical Models

Our work thus far allows us to define a GFD on X with a fixed graph G. We can express

$$\Sigma^{-1}_{i,j} = 0 \leftrightarrow (i,j) \notin E,$$

as the zeros of the function

$$g_{i,j}(\Sigma) = \Sigma^{-1}_{i,j}.$$

Gaussian Graphical Models

Combining this with our previous work on the Multivariate Normal (Murph *et al.*, 2022), we can use the Cayley Transform to decompose

$$g_{i,j}(\Sigma) = \Sigma_{i,j}^{-1}.$$

into

$$g_{i,j}(A, \Lambda) = \Big((I - A)(I + A)^{-1}\Lambda^{-2}(I - A)^{-1}(I + A)\Big)_{i,j}$$

(The full DGA is: $\mathbf{Y}_i = \mu + (I_d + A)(I_d - A)^{-1}\Lambda\mathbf{Z}_i, \quad i = 1, \dots, m.$)

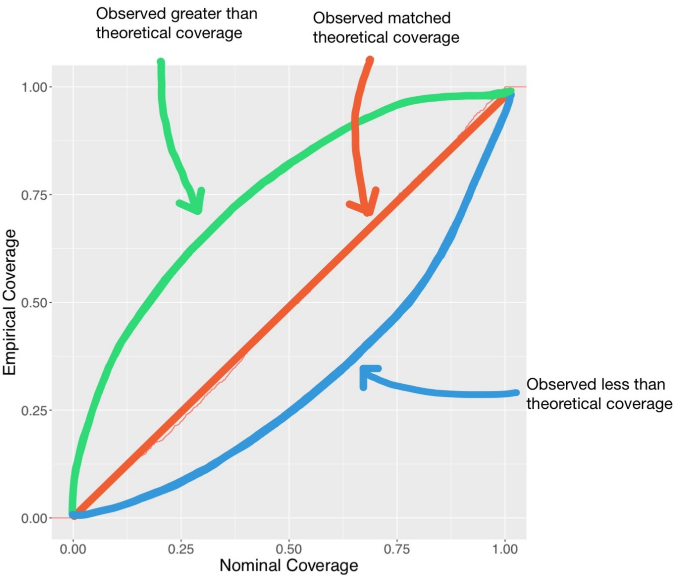
Gaussian Graphical Models

Ex: Let's sample GGMs from the Bayesian posterior and the CGFD for fixed graph structure G (assuming a zero mean).

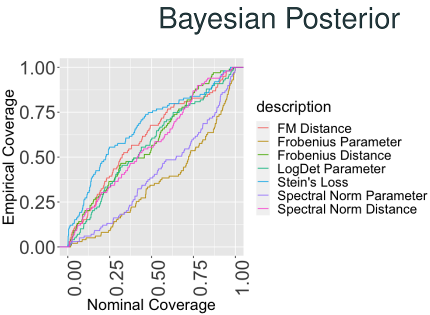
Bayesian: MVN model with G-Wishart prior; we can sample values directly from the posterior;

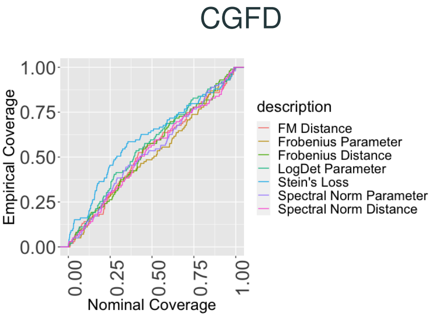
Generalized Fiducial: Propose values in a Metropolis-Hastings MCMC set-up using a G-Wishart with a change-of-variables to (A, Λ). Accepted samples are from the CGFD.

Gaussian Graphical Models



Gaussian Graphical Models





Thanks for your attention!

Questions?

References i

Kundu, S., Mallick, B. K. and Baladandayuthapani, V. (2019) Efficient Bayesian Regularization for Graphical Model Selection. *Bayesian Analysis*, **14**, 449 – 476.

Murph, A. C., Hannig, J. and Williams, J. P. (2022) Introduction to generalized fiducial inference. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*. Chapman & Hall.

References i

Hannig, J., Iyer, H., Lai, R. C. S. and Lee, T. C. M. (2016) Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, **111**, 1346–1361. URL https://doi.org/10.1080/01621459.2016.1165102.

Hwang, C.-R. (1980) Laplace's method revisited: Weak convergence of probability measures. *The Annals of Probability*, **8**, 1177 – 1182.

Liu, Y., Hannig, J. and Murph, A. C. (2022) A geometric perspective on bayesian and generalized fiducial inference. URL https://arxiv.org/abs/2210.05462.

Murph, A. C., Hannig, J. and Williams, J. P. (2022a) Generalized fiducial inference on differentiable manifolds.

References ii

— (2022b) Introduction to generalized fiducial inference. In *Handbook of Bayesian, Fiducial, and Frequentist Inference*. Chapman & Hall.