Estimating Case Fatality Rate via Convolutional Modeling

Jeremy Goldwasser

Joint work with Ryan Tibshirani, Addison Hu, Daniel McDonald, & Alyssa Bilinski

April 28, 2023

Berkeley
UNIVERSITY OF CALIFORNIA

Outline

1. Case Fatality Rate
2. Likelihood of Convolutional Model
3. Deconvolution Methods

Case Fatality Rate

The CFR is the …

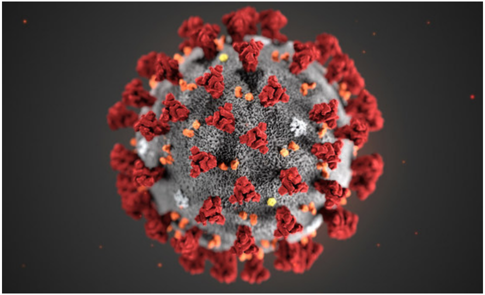
- Probability of dying from a disease

Berkeley
UNIVERSITY OF CALIFORNIA

Case Fatality Rate

The CFR is the …

- Probability of dying from a disease
- Probability of dying from a disease *at a given point in time*
  - *Omicron vs Delta vs …*



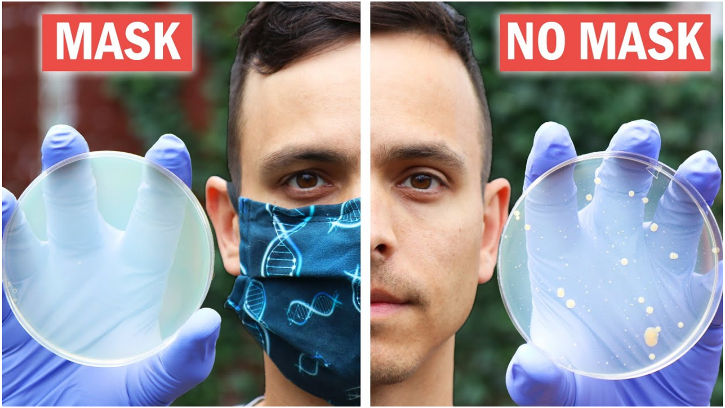
Berkeley
UNIVERSITY OF CALIFORNIA

Stakeholders

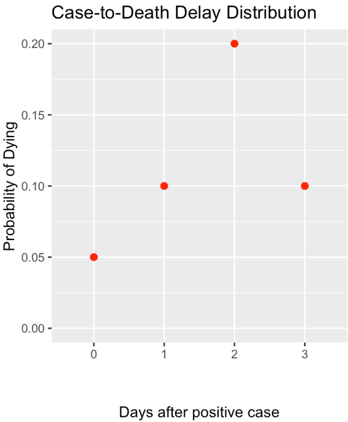
Berkeley
UNIVERSITY OF CALIFORNIA

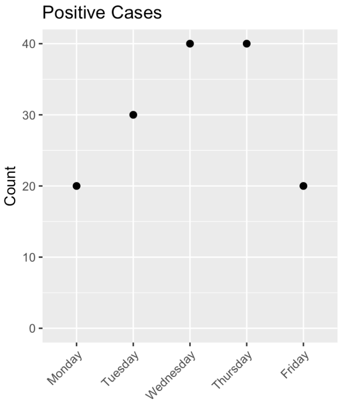
Stakeholders

Stakeholders



Convolving Cases to Deaths

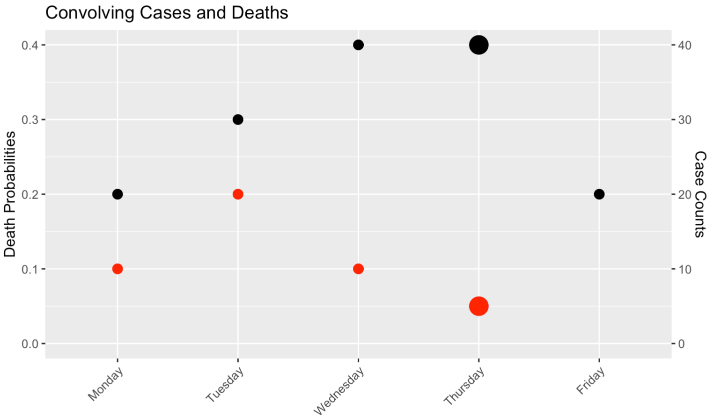




How many people will we expect to die on Thursday?

Berkeley
UNIVERSITY OF CALIFORNIA

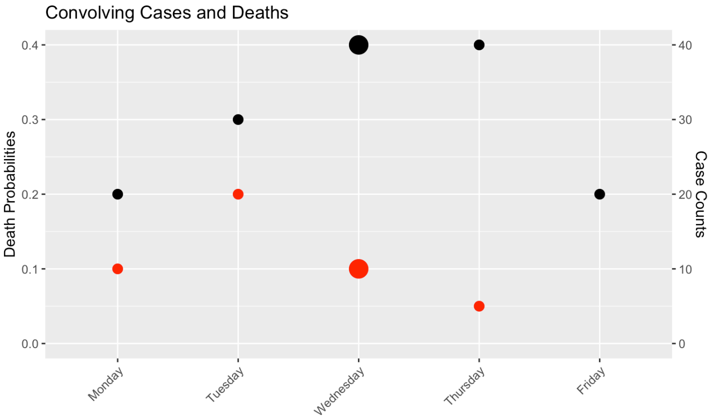
Convolving Cases to Deaths



$$E[Y_{Thurs}] = 40 * 0.05 + \ldots$$

Berkeley
UNIVERSITY OF CALIFORNIA

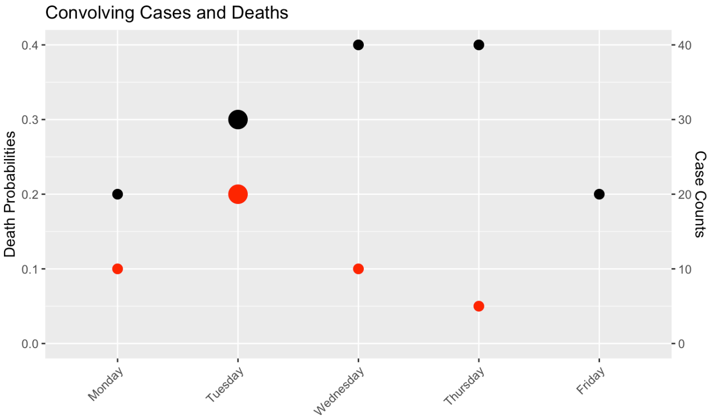
Convolving Cases to Deaths



$$E[Y_{Thurs}] = 40 * 0.05 + 40 * 0.1 + \dots$$

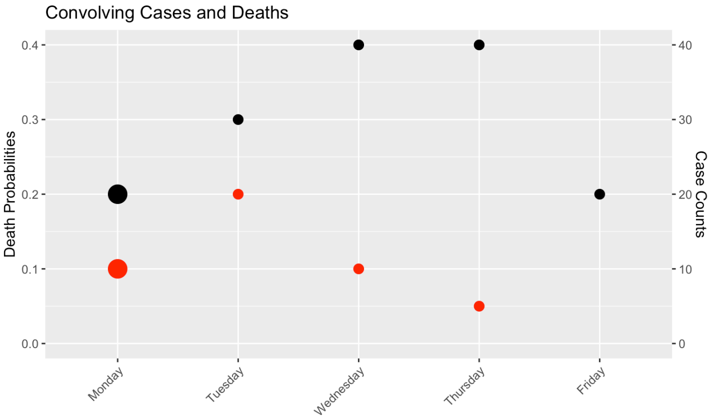
Berkeley
UNIVERSITY OF CALIFORNIA

Convolving Cases to Deaths



$$E[Y_{Thurs}] = 40 * 0.05 + 40 * 0.1 + 30 * 0.2 + \dots$$

Convolving Cases to Deaths



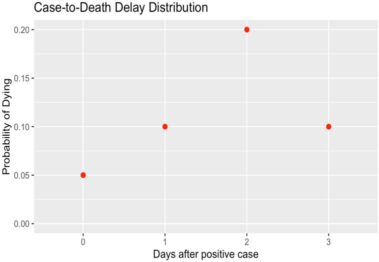
$$E[Y_{Thurs}] = 40 * 0.05 + 40 * 0.1 + 30 * 0.2 + 20 * 0.1 = 14$$

Berkeley
UNIVERSITY OF CALIFORNIA

Case Fatality Rate

Definition (Backward CFR)

$$\text{BCFR}(t) = \sum_{k=0}^{\infty} \mathbb{P}(\text{Die at } t \mid \text{Case at } t-k)$$



$$\text{BCFR}(t) = 0.05 + 0.1 + 0.2 + 0.1 = 0.45$$

Berkeley
UNIVERSITY OF CALIFORNIA

Case Fatality Rate

Definition (Forward CFR)

$$\text{FCFR}(t) = \sum_{k=0}^{\infty} \mathbb{P}(\text{Die at } t + k \mid \text{Case at } t)$$
$$= \mathbb{P}(\text{Die in future} \mid \text{Case at } t)$$

Conditions stagnant $\Rightarrow$ $BCFR(t) = FCFR(t)$.

Berkeley
UNIVERSITY OF CALIFORNIA

Lagged CFR

Let X_t denote cases and Y_t denote deaths at time t. For some ℓ,

$$\text{Lagged BCFR}(t) = \frac{Y_t}{X_{t-\ell}}$$

and

$$\text{Lagged FCFR}(t) = \frac{Y_{t+\ell}}{X_t}$$

Lagged CFR

Let X_t denote cases and Y_t denote deaths at time t. For some ℓ,

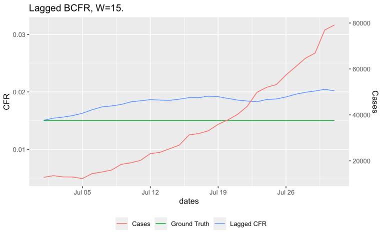
$$\text{Lagged BCFR}(t) = \frac{Y_t}{X_{t-\ell}}$$

and

$$\text{Lagged FCFR}(t) = \frac{Y_{t+\ell}}{X_t}$$

Model assumes all deaths after exactly ℓ days

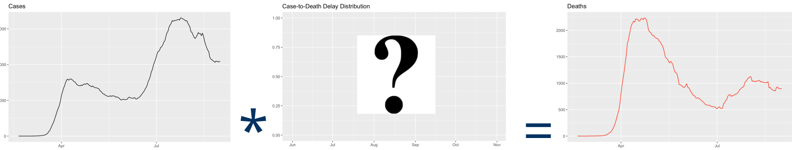
- *This isn't true!*
- *Induces bias: ↑ in surge, ↓ in downswing*



Berkeley
UNIVERSITY OF CALIFORNIA

Convolutional CFR

- By definition, $BCFR(t) = \sum_k \mathbb{P}(\text{Die at } t \mid \text{Case at } t-k) = \sum_k \beta_k$.
- *Can we estimate $\beta_k\ \forall k$?*
- Deconvolution problem: Given case & death counts, learn transfer function.



Convolutional Model

$$Y_t|X = \sum_k \left(\sum_{i=1}^{X_{t-k}} \mathbb{1}\{\text{Die at t } | \text{Case at t-k}\} \right)$$

is the sum of asymptotically independent normals by CLT. Therefore

Proposition (Distribution of $Y_t|X$)

$$Y_t|X \xrightarrow{d} \mathcal{N}(\mu_t, \sigma_t^2)$$

$$\textit{where } \mu_t = \sum_k X_{t-k}\beta_k \textit{ and } \sigma_t^2 = \sum_k X_{t-k}\beta_k(1-\beta_k)$$

Berkeley
UNIVERSITY OF CALIFORNIA

Convolutional Model

$$Y_t|X = \sum_k \left(\sum_{i=1}^{X_{t-k}} \mathbb{1}\{\text{Die at t } | \text{Case at t-k}\} \right)$$

is the sum of asymptotically independent normals by CLT. Therefore

Proposition (Distribution of $Y_t|X$)

$$Y_t|X \xrightarrow{d} \mathcal{N}(\mu_t, \sigma_t^2)$$

$$\textit{where } \mu_t = \sum_k X_{t-k}\beta_k \textit{ and } \sigma_t^2 = \sum_k X_{t-k}\beta_k(1-\beta_k)$$

Berkeley
UNIVERSITY OF CALIFORNIA

MLE of Convolutional Model

Assuming death counts on successive days are independent:

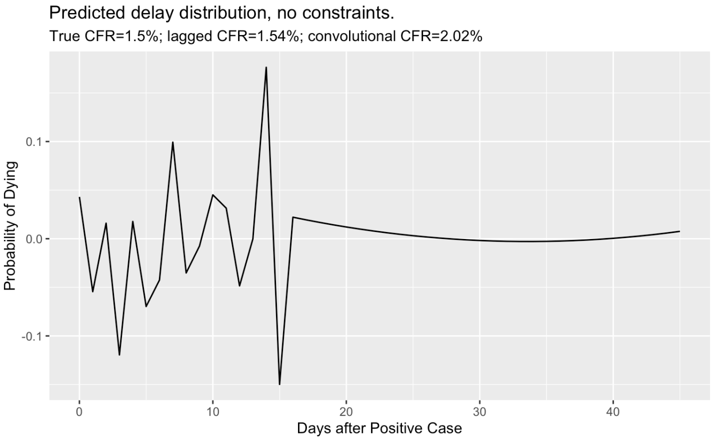
$$\begin{aligned}
\hat{\beta}^{MLE}(t) &= \operatorname*{argmax}_{\beta} \mathbb{L}(\beta|X, y) \\
&\approx \operatorname*{argmax}_{\beta} \sum_{s=1}^{n} \log \mathbb{P}(Y_s = y_s|X, \beta) \\
&= \operatorname*{argmax}_{\beta} \sum_{s=1}^{n} \log[\frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{1}{2\sigma_s^2}(y_s-\mu_s)^2}] \\
&= \operatorname*{argmin}_{\beta} \sum_{s=1}^{n} \frac{(y_s - \mu_s)^2}{\sigma_s^2} \\
&= \operatorname*{argmin}_{\beta} \sum_{s=1}^{n} \frac{(y_s - \sum_{k=1}^{d} X_{s-k}\beta_k)^2}{\sum_{k=1}^{d} X_{s-k}\beta_k(1-\beta_k)}
\end{aligned}$$

Berkeley
UNIVERSITY OF CALIFORNIA

Are we done?

Berkeley
UNIVERSITY OF CALIFORNIA

Are we done?



No

Why is the MLE so bad?

$$\hat{\beta}^{MLE}(t) \approx \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

Berkeley
UNIVERSITY OF CALIFORNIA

Why is the MLE so bad?

$$\hat{\beta}^{MLE}(t) \approx \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

Berkeley
UNIVERSITY OF CALIFORNIA

Why is the MLE so bad?

$$\hat{\beta}^{MLE}(t) \approx \underset{\beta \in \mathbb{R}^d}{\text{argmin}} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

Berkeley
UNIVERSITY OF CALIFORNIA

Why is the MLE so bad?

$$\hat{\beta}^{MLE}(t) \approx \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

Berkeley
UNIVERSITY OF CALIFORNIA

Why is the MLE so bad?

$$\hat{\beta}^{MLE}(t) \approx \underset{\beta \in \mathbb{R}^d}{\text{argmin}} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

Berkeley
UNIVERSITY OF CALIFORNIA

Why is the MLE so bad?

$$\hat{\beta}^{MLE}(t) \approx \underset{\beta \in \mathbb{R}^d}{\text{argmin}} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

Berkeley
UNIVERSITY OF CALIFORNIA

Why is the MLE so bad?

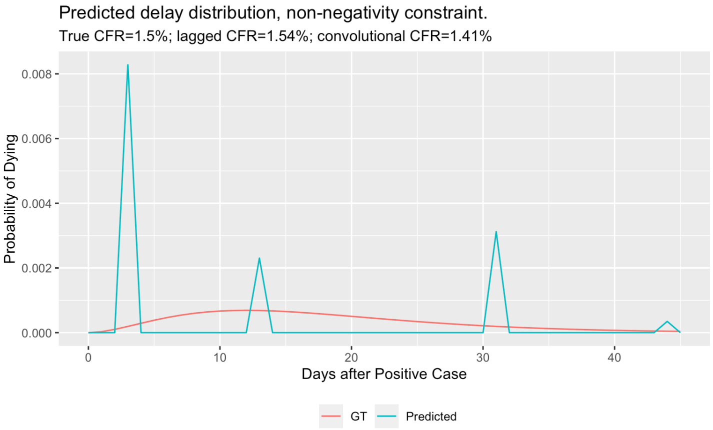
$$\hat{\beta}^{MLE}(t) \approx \underset{\beta \in \mathbb{R}^d}{\operatorname{argmin}} \|W(\beta)(Y - X\beta)\|_2^2$$

1. Small n
2. Large d
3. High σ^2

What can we do about this? Shape-constrained regression!

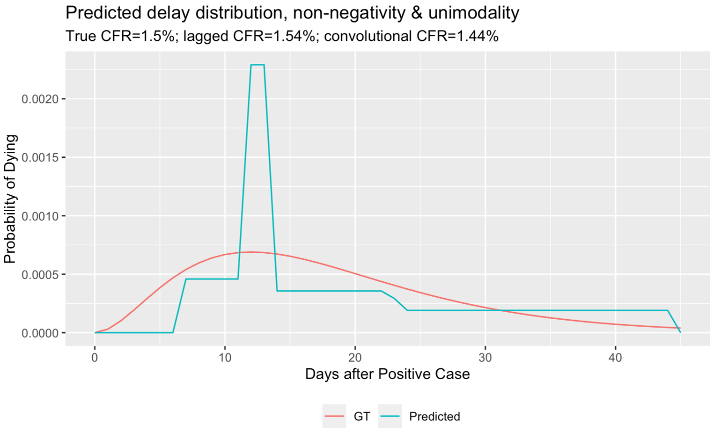
Nonnegativity

Nonnegativity



Better…

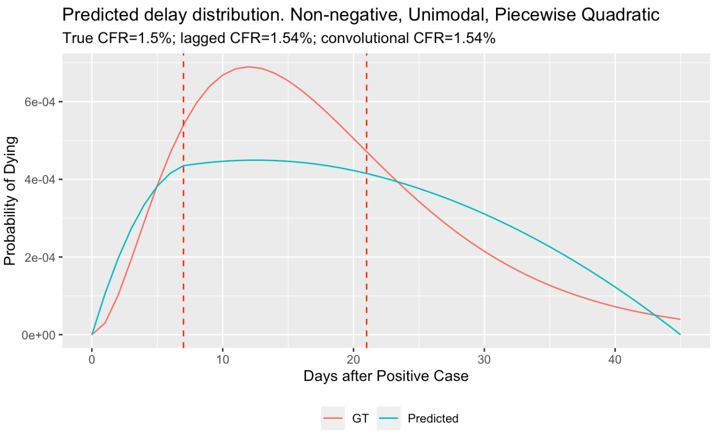
Unimodality



Better…

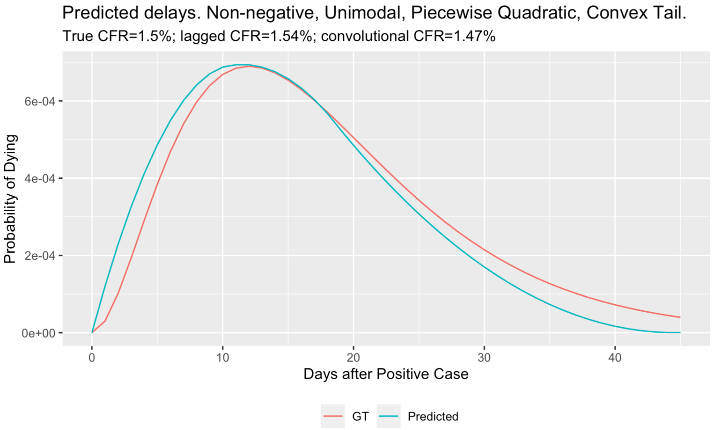
Berkeley
UNIVERSITY OF CALIFORNIA

Piecewise Quadratic



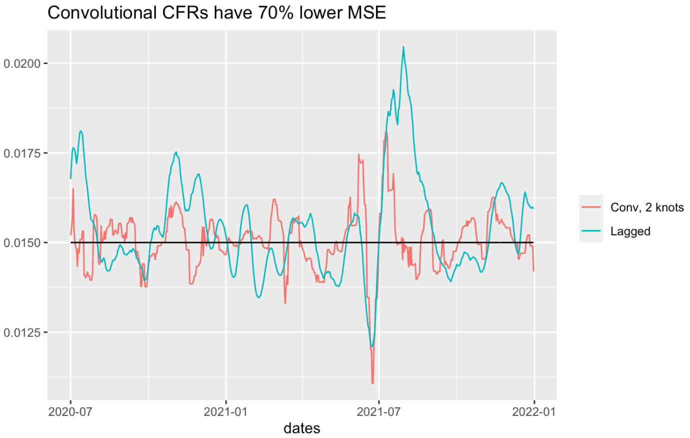
Better…

Convex Tail



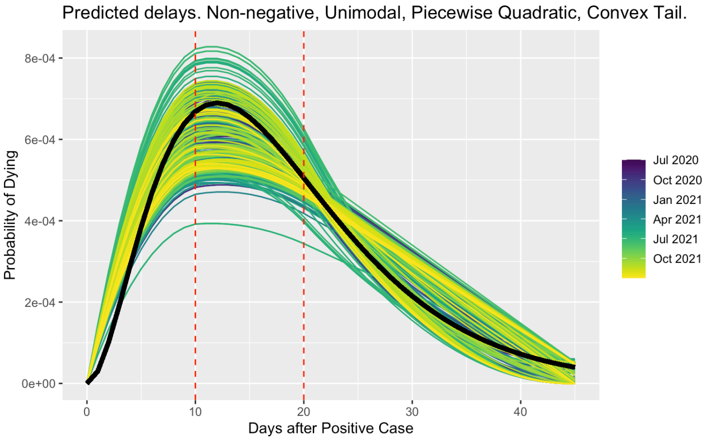
Looks good!

CFRs, All Time



Berkeley
UNIVERSITY OF CALIFORNIA

Delay Distributions, All Time



Aside: Trend Filtering

Can fix knots ahead of time... *or learn them adaptively*

- Trend Filtering enables us to do this!
- Smoothness hyperparameter λ controls number of knots. Choose whichever produces two.

Definition (2nd-order Trend Filtering)

$$\hat{\beta}^{TF} = \underset{\beta}{\mathbf{argmin}} \|Y - X\beta\|_2^2 + \lambda \|D^{(3)}\beta\|_1$$

Advantages

- Accurate CFRs from realistic delay distributions
- Nonparametric regression enables flexible modeling
- Convex loss, if we omit or fix weights

Advantages

- Accurate CFRs from realistic delay distributions
- Nonparametric regression enables flexible modeling
- Convex loss, if we omit or fix weights

Berkeley
UNIVERSITY OF CALIFORNIA

Advantages

- Accurate CFRs from realistic delay distributions
- Nonparametric regression enables flexible modeling
- Convex loss, if we omit or fix weights

Advantages

- Accurate CFRs from realistic delay distributions
- Nonparametric regression enables flexible modeling
- Convex loss, if we omit or fix weights

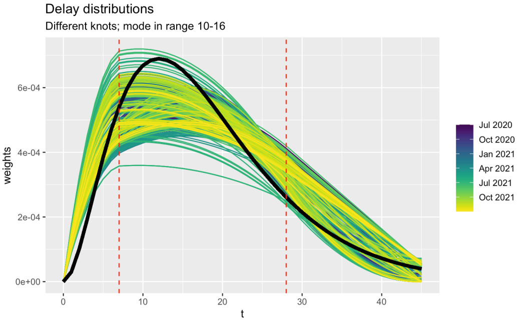
$$\hat{\beta} = \underset{\beta \in \Theta}{\text{argmin}} \| W(\beta)(Y - X\beta) \|_2^2$$

$$\approx \underset{\beta \in \Theta}{\text{argmin}} \| Y - X\beta \|_2^2 \qquad \text{or}$$

$$\approx \underset{\beta \in \Theta}{\text{argmin}} \| W(Y - X\beta) \|_2^2 \quad \text{with } W_{ii} = \frac{1}{Y_i}, W_{ij} = 0$$

Drawbacks

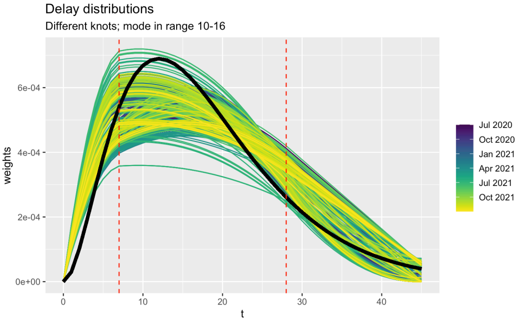
Need to specify hyperparameters for mode and knots.



Berkeley
UNIVERSITY OF CALIFORNIA

Drawbacks

Need to specify hyperparameters for mode and knots.

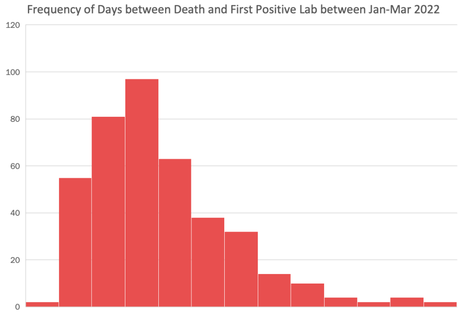


Can we get good shapes without hyperparameters?

Berkeley
UNIVERSITY OF CALIFORNIA

Parametric Model

Idea: Let β be a scaled PMF from probability family.
Our task will be to learn its parameters.

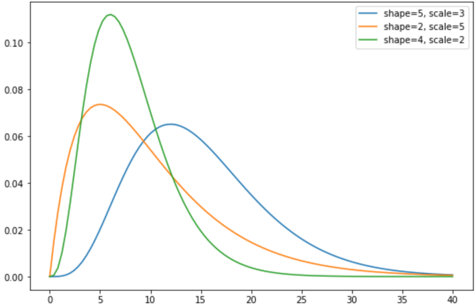


Berkeley
UNIVERSITY OF CALIFORNIA

Parametric Model

Idea: Let β be a scaled PMF from probability family.
Our task will be to learn its parameters.

The gamma distribution is a good candidate!



Berkeley
UNIVERSITY OF CALIFORNIA

Parametric Model

Let $f_\theta \in \mathbb{R}^d$ be a PMF parameterized by θ.

e.g. Gamma has two nonnegative parameters, shape and rate

Parametric Model

Let $f_\theta \in \mathbb{R}^d$ be a PMF parameterized by θ.

e.g. Gamma has two nonnegative parameters, shape and rate

Definition (Parametric Model)

Find $\hat{\beta} = \hat{c} f_{\hat{\theta}}$, where

$$\hat{c}, \hat{\theta} = \underset{c, \theta \in \Theta}{\operatorname{argmin}} \| W(c, \theta)(Y - cXf_\theta) \|_2^2$$

Berkeley
UNIVERSITY OF CALIFORNIA

Drawbacks

1. **Distributions may not be expressive enough.**

Drawbacks

1. **Distributions may not be expressive enough.**
2. **Loss is nonconvex!**

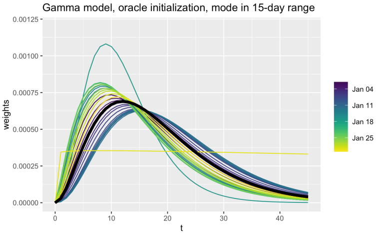
$\Rightarrow$ **Heavy dependence on initialization.**

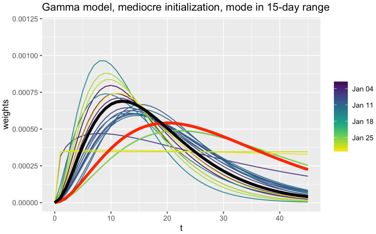
For any distribution whose tail decays exponentially,
the loss $\mathcal{L}(c, \theta) = \|Y - cXf_\theta\|_2^2$ resembles $g(\theta) = (1 - e^{-\theta})^2$.

$g'' \not> 0$ on whole domain $\Rightarrow g, \mathcal{L}$ not convex.

Berkeley
UNIVERSITY OF CALIFORNIA

Initialization



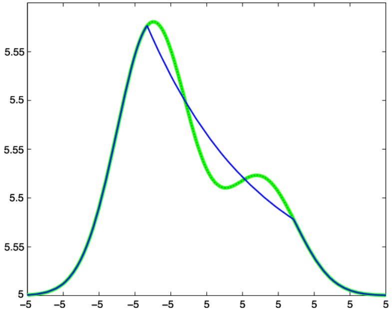


Berkeley
UNIVERSITY OF CALIFORNIA

Log-Concave Motivation

Class of log-convave functions best of both worlds

- Very expressive
- No mode hyperparameter



Unimodal, exponentially decaying tails

Log-Concave Weights

β is log-concave iff $\log(\beta)$ is concave.
If we reformulate our problem in terms of $u := \log(\beta) \in \mathbb{R}^d$, this will be a linear inequality constraint.

Definition (Log-Concave Weights)

Find $\hat{\beta} = e^{\hat{u}}$, where

$$\hat{u} = \underset{\substack{D^{(2)}u \preceq 0 \\ u = N\alpha}}{\operatorname{argmin}} \|W(u)(Y - Xe^u)\|_2^2$$

Log-Concave Weights

β is log-concave iff $\log(\beta)$ is concave.
If we reformulate our problem in terms of $u := \log(\beta) \in \mathbb{R}^d$, this will be a linear inequality constraint.

Definition (Log-Concave Weights)

Find $\hat{\beta} = e^{\hat{u}}$, where

$$\hat{u} = \underset{\substack{D^{(2)}u \preceq 0 \\ u = N\alpha}}{\operatorname{argmin}} \|W(u)(Y - Xe^u)\|_2^2$$

PROBLEM: Exponential renders nonconvex

- Get caught in local minimum

Conclusion

- Deconvolve relation between cases & deaths $\longrightarrow$ better interpretations & predictions of CFR
- Found MLE of deconvolution is approximately WLS
- Explored parametric & nonparametric estimators

Conclusion

- Deconvolve relation between cases & deaths $\longrightarrow$ better interpretations & predictions of CFR
- Found MLE of deconvolution is approximately WLS
- Explored parametric & nonparametric estimators

Berkeley
UNIVERSITY OF CALIFORNIA

Conclusion

- Deconvolve relation between cases & deaths $\longrightarrow$ better interpretations & predictions of CFR
- Found MLE of deconvolution is approximately WLS
- Explored parametric & nonparametric estimators

Thank You!