Predicting Spatiotemporal Counts of Opioid-related Fatal Overdoses via Zero-Inflated Gaussian Processes

Kyle Heuton*, Shikhar Shrestha[1], Thomas J. Stopka[1], Jennifer Pustz[1], Li-Ping Liu*, Michael C. Hughes*

* Department of Computer Science, Tufts University, Medford, MA

1 Dept. of Public Health and Community Medicine, Tufts University School of Medicine

Challenges and Project Goals

Public Health Challenges:

- There are over 2,000 opioid-related fatal overdoses each year in Massachusetts (more than traffic fatalities)
- There are limited resources for intervention, so model must help target these resources

Modeling Challenges:

- Death data is count-valued and very sparse with many exact zeroes

How we overcome these challenges:

- We used the Zero-Inflated Gaussian Process (ZIGP) model to flexibly capture the spatiotemporal trends in the data
- We introduce a Zero-Inflated Poisson likelihood better suited to model count data
- We develop a metric, %-Best Possible Reach which captures the limited-resource nature of the problem

Model: *Zero-Inflated Gaussian Process with Poisson Likelihood*

Model. For real-valued observation y with latent signal f and non-sparsity level g:

$$\mathbf{f}|\mathbf{g} \sim \mathcal{N}(\mathbf{0}_{1:N}, \mathbf{K}^F_{1:N,1:N} \odot \Phi(\mathbf{g})\Phi(\mathbf{g})^T)$$

$$\mathbf{g} \sim \mathcal{N}(\mathbf{0}_{1:N}, \mathbf{K}^G_{1:N,1:N})$$

where ϕ is the probit-link function.

The Zero-Inflated Gaussian Process [1], uses a Normal likelihood:

$$\tilde{y}_n | f_n \sim \mathcal{N}(f_n, \sigma^2), \text{ for } n \in 1, \dots N$$

This is inappropriate count data, so we propose a Zero-Inflated Poisson Likelihood

$$p(y_n = k | f_n, g_n) = \begin{cases} \Phi(g_n)\text{PoiPMF}(0|r(f_n)) + (1 - \Phi(g_n)) & k = 0 \\ \Phi(g_n)\text{PoiPMF}(k|r(f_n)) & k \in 1, 2, \dots \end{cases}$$

For scalability, we use the inducing point augmentation [2], introducing M random variables $\boldsymbol{h}^F$ and another M variables $\boldsymbol{h}^G$ representing latent function and non-sparsity level outputs.

Following the standard variational approach, we use approximate distributions over the inducing point locations:

$$q(\mathbf{h}^G) = \mathcal{N}(m^G, S^G), \; q(\mathbf{h}^F) = \mathcal{N}(m^F, S^F),$$

The evidence lower bound optimization objective becomes:

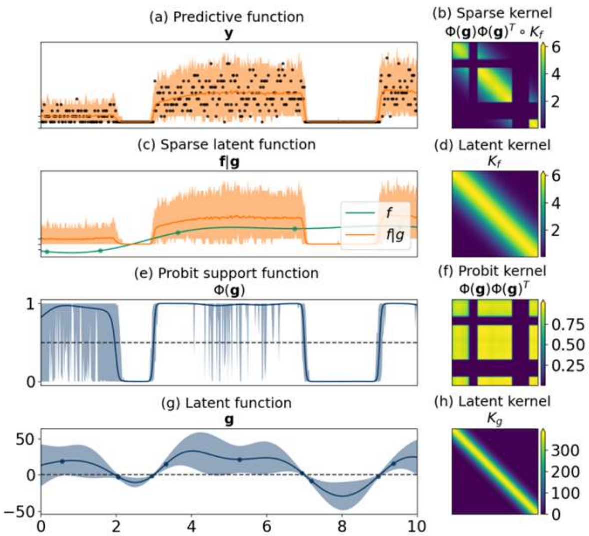
$$\mathcal{L}(\theta, \nu) = \mathbb{E}_{q_{\theta,\nu}(\mathbf{f},\mathbf{g})} \left[\sum_{n=1}^N \log p(y_n | f_n, g_n) \right] + \text{KL}(q_\nu(\mathbf{h}^G, \mathbf{h}^F) || p_\theta(\mathbf{h}^G, \mathbf{h}^F))$$

We can use ADVI [3] to obtain gradients using minibatches from our data

Acknowledgments

The authors gratefully acknowledge support from the U.S. National Science Foundation under grant NSF IIS-1908617. KH is also supported by NSF award NRT-HDR 2021874

ZIGP with Zero-Inflated Poisson Likelihood



Demonstration on toy data. For real-valued observation y with latent signal f and non-sparsity level g, we plot the latent kernels, probit support kernels, and the predictive function. We observe that for $\phi(g)$ close to 0, the model can learn exact 0, and it learns a count-value otherwise.

%-Best Possible Reach

We wish to assess prediction accuracy by using a metric which reflects the need to spend limited intervention resources on the K highest need census tracts. This metric evaluates the number of deaths in the top K predicted tracts divided by the number of deaths in the actual K worst tracts.

A model that perfectly predicts the K most severe census tracts will have a %BPR of 100%, while a model that predicts no deaths will have a %BPR of 0%.

$$\%\text{BPR} = \frac{\text{\# deaths in the predicted K tracts}}{\text{\# deaths in best-possible K tracts}} \times 100\%$$

References

[1] P. Hegde, M. Heinonen, and S. Kaski. Variational zero-inflated Gaussian processes with sparse kernels. In Uncertainty in Artificial Intelligence, 2018

[2] M. Titsias. Variational learning of inducing variables in sparse Gaussian processes. In Artificial Intelligence and Statistics, pages 567–574, 2009

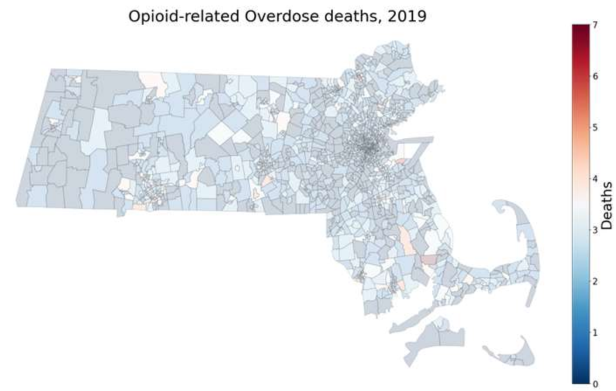
[3] A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei. Automatic Differentiation Variational Inference. Journal of Machine Learning Research, 18(14):1–45, 2017

Performance Metrics for 2019 in MA

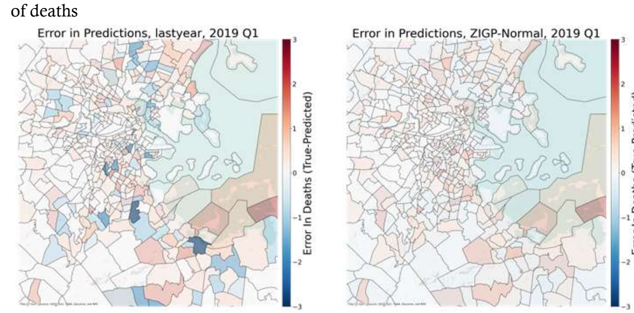
method	%BPR
allzero	16.0
lastyear	26.4
GLM+Poisson	25.2
RF+Poisson	25.1
ZIGP+Normal	34.5
ZIGP+ZIPoisson	31.9

The Zero-Inflated Gaussian Process models out-perform all baselines, but surprisingly the Normal likelihood works best, despite not being appropriate for count-valued data.

Results for Massachusetts



Total Opioid-related overdose deaths for Massachusetts in 2019 at the census tract level. We observe many exact zeros, and heterogeneity in the geographic distribution of deaths



Prediction errors for a naïve model using the deaths during this quarter last year as the prediction for 2019 Q1 in Metro Boston, left

Errors for the best-performing model, the ZIGP with Normal likelihood in 2019 Q1, right.

We observe that the ZIGP model lacks the dark blue hues indicating too-high predicted deaths and dark reds hues indicating too-low predictions