Selective Inference for Poisson
Charles Zheng (and Jonathan Taylor)

1 Discrete Knots

Data S is a poisson process N with points in $\mathcal{X}$ and intensity

$$\frac{d\lambda(\beta)}{d\mu} = \exp\left(\sum_{j=1}^{p} \beta_j h_j(x)\right)$$

Take the change-point model where $h_x(x) = I\{x \in [u_j, 1]\}$.

2 Continuous Knots

2.1 Nonrandomized problem

Let $\vartheta(x)$ be the step function

$$\vartheta(x) = \sum_{i=1}^{K} w_{\theta_i} I\{x \in [\theta_i, 1]\}.$$

Here $K \in \{0, 1, \ldots\}$ and $\theta_i \in [0, 1]$.

Data S is a poisson process N with points in $\mathcal{X}$ and intensity

$$\frac{d\lambda(\beta)}{d\mu} = \exp(\vartheta(x))$$

Define the total-variation norm

$$||\vartheta||_{TV} = \min_{f,g} ||f||_\infty + ||g||_\infty \text{ subject to } f, g \text{ increasing; } \vartheta = f - g$$

Suppose we observe $x_1, \ldots, x_N \in [0, 1]$ where $x_i \sim F$.

Despite the fact that ϑ has an infinite-dimensional parameter space, it is possible to find the global minimum of the following problem

$$\text{minimize }_\vartheta \Lambda(\vartheta) - \sum_{i=1}^{n} \vartheta(x_i) + \lambda||\vartheta||_{TV}$$

because defining $h_t = I\{x \in [t, 1]\}$, the minimizing ϑ is given by

$$\vartheta(x) = \sum_{i=1}^{p} w_{x_i} h_{x_i}(x)$$

where w_{x_i} are determined by

$$\text{minimize }_w \Lambda(w) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{x_j} h_{x_j}(x_i) + \lambda ||w||_1$$

2.2 Randomized problem

Since the parameter space of ϑ is over continuous functions, the randomization parameter ω must parameterize a random functional of $[0,1]$.

2.2.1 Scheme 1

One example: $\omega = (p, w)$ where $p \in [0,1]^m$ and $w \in \mathbb{R}^m$. Then define the random functional $\Omega(\vartheta) = \sum_{i=1}^m w_i \vartheta(\omega_i)$.

In any case, the random optimization becomes

$$\text{minimize }_\vartheta \Lambda(\vartheta) - \sum_{i=1}^{n} \vartheta(x_i) + \lambda ||\vartheta||_{TV} - \Omega(\vartheta) + \frac{\epsilon}{2} ||\vartheta||_{TV}^2$$

2.2.2 Scheme 2

Another example which is a special case of the above. Sample $\omega_i \sim Q$ for $i = 1, \ldots, m$. Consider the optimization

$$\text{minimize }_\vartheta \Lambda(\vartheta) + \Lambda_\omega - \sum_{i=1}^{n} \vartheta(x_i) + \sum_{i=1}^{n} \vartheta(\omega_i) + \lambda ||\vartheta||_{TV}$$

where

$$\tilde{\Lambda}(\vartheta) = \log \left[\int_{\mathcal{X}} \left[\exp(\vartheta(x)) - 1\right] \mu(dx) \right]$$

This is just the original continuous optimization problem except augmenting the observed data with sampled points ω.

2.3 Support of Selective Distribution

We would like to condition on the event where the nonzero knots from x are $\hat{E}^x = \{z \in \{x_i\} : w_z \neq 0\} = E^x$ and the nonzero knots from ω are $\hat{E}^\omega = \{z \in \{\omega_i\} : w_z \neq 0\} = E^\omega$ and where $z_{\hat{E}} = s_E = s_{E^x \cup E^\omega}$.

Let $z_1^x, \ldots, z_k^x$ denote the nonzero knots from $\{x\}$ and $z_1^\omega, \ldots, z_\ell^\omega$ the nonzero knots from $\{\omega\}$. If we test the hypothesis that $H_{0,j|E} : w_{z_j} = 0$, then we will also condition on the sufficient statistics

$$t_i^x = \sum_{i=1}^{n} h_{z_i^x}(x_i) = \sum_{\ell=1}^{n} I(x_\ell < z_i^x)$$

$$t_i^\omega = \sum_{\ell=1}^{n} I(x_\ell < z_i^\omega)$$

Furthermore, we can condition on n, the observed number of points.

Let us find the distribution of augmented data $\tilde{S} = \{x\} \cup \{\omega\}$ conditional on all these conditions. A first step is to identify the support.

The support of $\tilde{S}$ consists of datasets with n observed points and m randomly added points. The condition of having knots $\{z^x\}, \{z^\omega\}$, means that all instances of $\tilde{S}$ in the support must contain the points $\{z^x\} \subset \{x\}$ and $\{z^\omega\} \subset \{\omega\}$. Let $z_{(1)}, \ldots, z_{(k+\ell)}$ denote the sorted elements of $\{z\}$, and $[i]$ defined where $z_{([j])} = z_j^x$. The condition on t_i means that the number of points in $(z_{(i)}, z_{(i+1)})$ is fixed, for all $i \notin \{[j]-1, [j]\}$, and that the number of points in $(z_{[j]-1}, z_{[j]+1})$ is fixed. This defines the support of the selective distribution of $\tilde{S}$.

2.4 Minimal λ

In order to compute or sample from the selective distribution, it will be useful to be able to compute on arbitrary intervals (a, b) the function $\lambda^*_{(a,b)}(y_1, \ldots, y_\ell)$ with argument $(y_1, \ldots, y_\ell) \in [0, 1]^\ell$ defined to be the minimal λ such that

$$\text{minimize}_\vartheta \Lambda(I_{(a,b)}\vartheta) - \sum_{i=1}^{\ell} \vartheta(y_i) + \lambda ||I_{(a,b)}\vartheta||_{TV}$$

has a nonconstant solution. The nonconstant solution takes the form $\vartheta(x) = \beta_0 I_{[0,a]\cup[b,1]}(x) + \beta_1 I_{(a,y_i)}(x) + \beta_2 I_{[y_i,1]}(x)$. Hence $\lambda^*_{(a,b)}(y_1, \ldots, y_\ell)$ is given by the same as the minimal λ where

$$\text{minimize}_{\beta_1,\beta_2,i} \log[(e^{\beta_1}-1)\mu((a, y_i))+(e^{\beta_2}-1)\mu([y_i, b))]-(i\beta_1+(\ell-i)\beta_2)+\lambda|\beta_1-\beta_2|$$

has a solution with $\beta_1 \neq \beta_2$.

Q: Shouldn't we be able to decompose the likelihood along the split??