Statistical view of Spline

Spline

- In computational field, it was populous to interpolate points.
- The property of convex function $x = \lambda x_0 + (1-\lambda)\, x_1$ is fundamental of it.
- Spline is the method of such an interpolation that it connect the points through piecewise polynomial curves.
- (Thm) Let $(\mathbf{c}_i)_{i=1}^{n}$ be a set of control points for a spline curve f of degree d, with nondecreasing knots $(t_i)_{i=1}^{n+d+1}$,

$$f(t) = \sum_{i=d+1}^{n} p_{i.d}(t) B_{i,0}(t)$$

where $p_{i.d}$ is given recursively by

$$p_{i,d-r+1}(t) = \frac{t_{i+1} - t}{t_{i+r} - t_i} p_{i-1,d-r}(t) + \frac{t - t_i}{t_{i+r} - t_i} p_{i,d-r}(t)$$

for $i = d - r + 1, ..., n$, and $r = d, d-1, ..., 1$, while $p_{i,0}(t) = \mathbf{c}_i$ for $i = 1, ..., n$. This representation is equivalent to

$$f(t) = \sum_{i=1}^{n} \mathbf{c}_i B_{i,d}(t)$$

where $B_{i,d}$ is given recursively by

$$B_{i,d}(t) = \frac{t - t_i}{t_{i+1} - t_i} B_{i,d-1}(t) + \frac{t_{i+1+d} - t}{t_{i+1+d} - t_{i+1}} B_{i+1,d-1}(t)$$

where $B_{i,0}$ is

$$B_{i,0}(t) = \begin{cases} 1 & t \in [t_i, t_{i+1}) \\ 0 & \text{otherwise.} \end{cases}$$

- Our interest is to apply a spline method to approximate nonlinear function.
- Note that situation we face is a little different with the interpolation in that there are much data than control points.

Cubic spline

- Basis expansions : the input space can be expanded through a function $h_m : \mathbb{R}^p \to \mathbb{R}$. We call it mth transformation $m = 1, ..., M$.

$$f(X) = \sum_{m=1}^{M} \beta_m h_m(X),$$

which is called as “a linear basis expansion in X”.

- Let ξ_1 and ξ_2 be knots in X, then cubic spline is constructed with following transformation functions.

$$
\begin{aligned}
h_1(X) &= 1 \\
h_2(X) &= X \\
h_3(X) &= X^2 \\
h_4(X) &= X^3 \\
h_5(X) &= (X-\xi_1)_+^3 \\
h_6(X) &= (X-\xi_2)_+^3\,,
\end{aligned}
$$

  Total degree of freedom is 6 in that 3 seperated spaces that have cubic polynomial functions respectively, and each knot has three restriction, continuous on function, first derivative and second derivative, respectively. Thus,

$$
4 \times 3 - 2 \times 3 = 6.
$$

- This truncated formual is equal to original cubic spline with knots ξ_1 and ξ_2.

Natural cubic spline

- The behaviour of polynomials fit to data tends to be erratic near the boundaries, and extrapolation can be dangerous.
- By adding constraints that the function is linear beyond the boundary knots, the erratic behaviour would be alleviated.
- Two constraints each in both boundary regions.
- Basis functions with K knots are follows.

$$
\begin{aligned}
N_1(X) &= 1 \\
N_2(X) &= X \\
N_{k+2}(X) &= d_k(X) - d_{K-1}(X) \\
d_k(X) &= \frac{(X-\xi_k)_+^3 - (X-\xi_K)_+^3}{\xi_K - \xi_k}.
\end{aligned}
$$

B-spline

- (Natural) cubic spline is quite reasonable, but could make numerical unstability due to overflow,$100^3 = 1,000,000$.

- In this sense, B-spline can be a replacement in order to obtain efficient and stable calculation.

- It has recursive formula as degree d increases.

$$f(X) = \sum_{k=1}^{K} \beta_k B_{k,d}(X)$$

where $B_{k,d}(X)$ is a basis function of B-spline that has recursive formula,

$$B_{k,d}(X) = \frac{t - t_k}{t_{k+1} - t_k} B_{k,d-1}(X) + \frac{t_{k+1+d} - t}{t_{k+1+d} - t_{k+1}} B_{k+1,d-1}(X)$$

with $B_{k,0}$

$$B_{k,0}(t) = \begin{cases} 1 & t \in [t_k, t_{k+1}) \\ 0 & \text{otherwise.} \end{cases}$$

Varying coefficients model

Consider linear model $Y = X\beta + \epsilon$. If there is an another variable t so that it makes an interaction effect with X, then the linear model can be represented as

$$\begin{aligned} Y &= X\beta(t) + \epsilon. \\ y_i &= x_i\beta(t) + \epsilon_i \end{aligned}$$

We call this model varying coefficient model. Let $Y \in \mathbb{R}^{n\times 1}$, $X \in \mathbb{R}^{n\times 1}$, $t \in \mathbb{R}^{n\times 1}$, and $\epsilon \sim \mathrm{N}\left(0, \tau^{-1}\right)$.

Let one explainatory variable exist. The varying coefficient can be estimated via B-spline model.

$$\begin{aligned} \beta(t) &= \sum_{k=1}^{K} \phi_k B_{k,d}(t) \\ &= \mathbf{B}\phi, \end{aligned}$$

where $B_{k,d}(t)$ is basis function of B-spline with ξ_1, ξ_2, ..., ξ_K as knots, and $\mathbf{B}$ is

$$\mathbf{B}\phi = \begin{bmatrix} B_{1,d}(t_1) & B_{2,d}(t_1) & \cdots & B_{K-1,d}(t_1) & B_{K,d}(t_1) \\ B_{1,d}(t_2) & B_{2,d}(t_2) & \cdots & B_{K-1,d}(t_2) & B_{K,d}(t_2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ B_{1,d}(t_{n-1}) & B_{2,d}(t_{n-1}) & \cdots & B_{K-1,d}(t_{n-1}) & B_{K,d}(t_{n-1}) \\ B_{1,d}(t_n) & B_{2,d}(t_n) & \cdots & B_{K-1,d}(t_n) & B_{K,d}(t_n) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{K-1} \\ \phi_K \end{bmatrix} = \begin{bmatrix} B_{\cdot,d}(t_1)^\top\phi \\ B_{\cdot,d}(t_2)^\top\phi \\ \vdots \\ B_{\cdot,d}(t_{n-1})^\top\phi \\ B_{\cdot,d}(t_n)^\top\phi \end{bmatrix}$$

Then, the model $E[Y \mid X] = X\beta(t) = X\mathbf{B}\phi$,

$$
\begin{aligned}
X\mathbf{B}\phi &= \begin{bmatrix} x_1 B_{\cdot,d}(t_1)^\top \phi \\ x_2 B_{\cdot,d}(t_2)^\top \phi \\ \vdots \\ x_{n-1} B_{\cdot,d}(t_{n-1})^\top \phi \\ x_n B_{\cdot,d}(t_n)^\top \phi \end{bmatrix} \\
&= \begin{bmatrix} x_1 B_{1,d}(t_1) & x_1 B_{2,d}(t_1) & \cdots & x_1 B_{K-1,d}(t_1) & x_1 B_{K,d}(t_1) \\ x_2 B_{1,d}(t_2) & x_2 B_{2,d}(t_2) & \cdots & x_2 B_{K-1,d}(t_2) & x_2 B_{K,d}(t_2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ x_{n-1} B_{1,d}(t_{n-1}) & x_{n-1} B_{2,d}(t_{n-1}) & \cdots & x_{n-1} B_{K-1,d}(t_{n-1}) & x_{n-1} B_{K,d}(t_{n-1}) \\ x_n B_{1,d}(t_n) & x_n B_{2,d}(t_n) & \cdots & x_n B_{K-1,d}(t_n) & x_n B_{K,d}(t_n) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{K-1} \\ \phi_K \end{bmatrix}.
\end{aligned}
$$

Let $X\mathbf{B} = W$.

Let's study simulation!

Specifying prior distributions

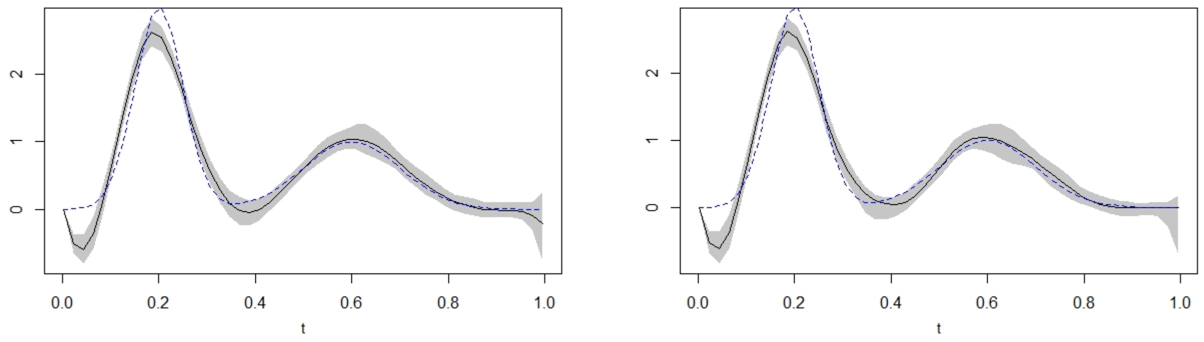
Following prior distributions are considered for each parameter.

$$
\begin{aligned}
&\phi_k \sim \mathrm{N}\left(0, \alpha_k^{-1}\right) \quad \forall k = 1, ..., K \\
&\alpha_k \sim \mathrm{Gamma}\left(a, b\right) \quad \forall k = 1, ..., K \\
&\tau \sim \mathrm{Gamma}\left(c, d\right).
\end{aligned}
$$

Setting

Varying coefficient model $\beta(t) = 3\exp\left(-200\left(t-0.2\right)^2\right) + \exp\left(-50\left(t-0.6\right)^2\right)$.

The left is the resulf of variational inference, and the right is of ordinary Gibbs sampler.



Variational inference

Steps of variational inference are introduced at https://jinwonsohn.github.io/statistics/bayesian/2019/03/22/Variational-Inference-(3).html

Gibbs sampler

Construction of Gibbs sampler is as follows.

$$Y = W\phi + \epsilon$$

$$\phi \sim \mathrm{N}\left(0, \alpha^{-1} I\right) \in \mathbb{R}^{p \times 1}$$

where $\alpha = (\alpha_1, ..., \alpha_p)$ has prior as

$$\alpha_k \sim \mathrm{Gamma}\left(a, b\right), \forall k = 1, ..., p$$

$$\epsilon \sim \mathrm{N}\left(0, \tau^{-1}\right) \in \mathbb{R}^{N \times 1}$$

where τ has prior as

$$\tau \sim \mathrm{Gamma}\left(c, d\right).$$

The joint likelihood is

$$\begin{aligned} p(\phi, \alpha, \tau \mid Y) &\propto p(Y \mid \phi, \alpha, \tau) p(\tau) p(\phi \mid \alpha) p(\alpha) \\ &= \mathrm{N}\left(W\phi, \tau^{-1} I\right) \mathrm{Gamma}\left(\tau \mid c, d\right) \mathrm{N}\left(\phi \mid 0, \alpha^{-1} I\right) \prod_{k=1}^{p} \mathrm{Gamma}\left(\alpha_k \mid a, b\right) \end{aligned}$$

Then, Gibbs sampler is

1. For ϕ,

$$p(\phi \mid -) \propto \mathrm{N}\left(W\phi, \tau^{-1}I\right) \times \mathrm{N}\left(\phi \mid 0, \alpha^{-1}I\right)$$

$$\begin{aligned}
&= \det\left(2\pi\tau^{-1}I\right)^{-1/2} \exp\left(-\frac{\tau}{2}\left(Y - W\phi\right)^\top\left(Y - W\phi\right)\right) \times \det\left(2\pi\alpha^{-1}I\right) \exp\left(-\frac{\alpha}{2}\phi^\top\phi\right) \\
&\propto \exp\left(-\frac{\tau}{2}\left[\phi^\top W^\top W\phi - 2\phi^\top W^\top Y\right] - \frac{\alpha}{2}\phi^\top\phi\right) \\
&= \exp\left(-\frac{1}{2}\left[\tau\phi^\top W^\top W\phi - 2\tau\phi^\top W^\top Y + \alpha\phi^\top\phi\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\phi^\top\left(\tau W^\top W + \alpha I\right)\phi - 2\phi^\top\tau W^\top Y\right]\right) \\
&\sim \mathrm{N}\left(\tau\left(\tau W^\top W + \alpha I\right)^{-1} W^\top Y, \left(\tau W^\top W + \alpha I\right)^{-1}\right).
\end{aligned}$$

2. For τ,

$$p(\tau \mid -) \propto \mathrm{N}\left(W\phi, \tau^{-1}I\right) \mathrm{Gamma}\left(\tau \mid c, d\right)$$

$$\begin{aligned}
&\propto \det\left(2\pi\tau^{-1}I\right)^{-1/2} \exp\left(-\frac{\tau}{2}\left(Y - W\phi\right)^\top\left(Y - W\phi\right)\right) \times \tau^{c-1}\exp\left(-d\tau\right) \\
&\propto \tau^{N/2+c-1} \exp\left(-\tau\left(\frac{1}{2}\left(Y - W\phi\right)^\top\left(Y - W\phi\right) + d\right)\right) \\
&\sim \mathrm{Gamma}\left(c + \frac{N}{2}, d + \frac{1}{2}\left(Y - W\phi\right)^\top\left(Y - W\phi\right)\right).
\end{aligned}$$

3. For α_k,

$$p(\alpha_k \mid -) \propto \mathrm{N}\left(\phi_k \mid 0, \alpha_k^{-1}\right) \mathrm{Gamma}\left(\alpha_k \mid a, b\right)$$

$$\begin{aligned}
&\propto \det\left(2\pi\alpha_k^{-1}\right)^{-1/2} \exp\left(-\frac{\alpha_k}{2}\phi_k^2\right) \times \alpha_k^{a-1}\exp\left(-b\alpha_k\right) \\
&\alpha_k^{1/2+a-1} \exp\left(-\alpha_k\left(\frac{\phi_k^2}{2} + b\right)\right) \\
&\sim \mathrm{Gamma}\left(a + \frac{1}{2}, \frac{\phi_k^2}{2} + b\right), \quad \forall k = 1, ..., p
\end{aligned}$$