Spatiotemporal Modeling for HRF

Jungin Choi

Bias-Variance Trade-off

Case I: i.i.d errors

Let's assume Canonical + Derivative (DD) model for HRF estimation and assume that error terms follow iid distribution with unknown variance of σ^2.

For independent (Non-spatial Model), all timecourse for each voxel(coordinates) are fitted independently.

$$Y(x_1, x_2) = X\beta(x_1, x_2) + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

Then,

$$\hat{\sigma}^2 = \frac{1}{T-p}\hat{\epsilon}^{'}(x_1, x_2)\hat{\epsilon}(x_1, x_2) \;\; \text{where } \hat{\epsilon}(x_1, x_2) = Y(x_1, x_2) - X\hat{\beta}(x_1, x_2)$$
$$\hat{\beta}(x_1, x_2) = (X^{'}(x_1, x_2)X(x_1, x_2))^{-1}X^{'}(x_1, x_2)Y(x_1, x_2)$$

For indepdent model, we know fitted β is unbiased and sample variance for β is estimated using sample errors.

$$V\hat{a}r(\hat{\beta}) = (X^{'}X)^{-1}\hat{\sigma}^2 = (X^{'}X)^{-1}\frac{1}{T-p}\hat{\epsilon}^{'}\hat{\epsilon}$$

For Gaussian Kernel smoothing, we have biased but smaller variance of estimated β.

Gaussian kernel smoothing is used to capture spatial dependencies across neighboring voxels. The estimated smoothed time course data at time t at coordinates (x_1, x_2), denoted as $\tilde{Y}(t, x_1, x_2)$, is given by:

$$\tilde{Y}(t, x_1, x_2) = \frac{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)Y_i(t, x_{1i}, x_{2i})}{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)},$$

where the Gaussian kernel $K(x_1, x_2|\mu_1, \mu_2, \sigma)$ is defined as:

$$K(x_1, x_2|\mu_1, \mu_2, \sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(x_1-\mu_1)^2 + (x_2-\mu_2)^2}{2\sigma^2}\right).$$

Researchers typically use a fixed standard deviation for the Gaussian kernel such that the Full Width at Half Maximum (FWHM) of the kernel is between 4 mm and 8 mm—commonly set to 6 mm. Assuming the coordinates are in millimeters, we calculate:

$$\text{FWHM} = 2\sqrt{2\ln 2}\,\sigma = 6 \implies \sigma = \frac{3}{\sqrt{2\ln 2}} \approx 2.55.$$

Let $\tilde{Y}(x_1, x_2)$ represent the time course at voxel (x_1, x_2):

$$\tilde{Y}(x_1, x_2) = \begin{pmatrix} \tilde{Y}(1, x_1, x_2) \\ \vdots \\ \tilde{Y}(T, x_1, x_2) \end{pmatrix}.$$

For each voxel, the fitted β using Gaussian kernel smoothed time course data is expressed as:

$$\hat{\tilde{\beta}}(x_1, x_2) = (X^{'}\Sigma^{-1}X)^{-1}X^{'}\Sigma^{-1}\tilde{Y}(x_1, x_2).$$

The expected value of the fitted β is:

$$\mathbb{E}(\hat{\tilde{\beta}}(x_1, x_2)) = (X^{'}X)^{-1}X^{'}\mathbb{E}(\tilde{Y}(x_1, x_2)).$$

We know that $\mathbb{E}(Y(x_1, x_2)) = X\beta(x_1, x_2)$. Thus:

$$\mathbb{E}(\tilde{Y}(t, x_1, x_2)) = \mathbb{E}\left(\frac{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)Y(t, x_{1i}, x_{2i})}{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)}\right)$$

$$= \frac{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)\mathbb{E}(Y(t, x_{1i}, x_{2i}))}{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)}$$

$$\mathbb{E}(\tilde{Y}(x_1, x_2)) = \frac{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)X\beta(x_{1i}, x_{2i})}{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)}$$

$$\therefore \mathbb{E}(\tilde{\beta}(x_1, x_2)) = \frac{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)\beta(x_{1i}, x_{2i})}{\sum_i K(x_{1i}, x_{2i}|x_1, x_2, \sigma)}.$$

Thus, the fitted β using the Gaussian kernel smoothed time course data is biased, and its expectation is the Gaussian kernel smoothed true β.

For variance, we also using kernel smoothed residuals to estimate variance.

$$\hat{Var}(\hat{\tilde{\beta}}) = (X^{'}X)^{-1}\hat{\tilde{\sigma}}^2$$
$$\text{where } \hat{\tilde{\sigma}}^2 = \frac{1}{T-p}(\hat{\tilde{\epsilon}})^{'}(\hat{\tilde{\epsilon}})$$

Also, we know that residuals from gaussian kernel smoothed timecourse data are gaussian kernel smoothed residuals from independent model. i.e,

$$
\begin{aligned}
\tilde{\hat{\epsilon}}(x_1,x_2) &= \tilde{Y}(x_1,x_2) - X\tilde{\hat{\beta}}(x_1,x_2) \\
&= \frac{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)Y(x_{1i},x_{2i})}{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)} \\
&- X(x^{'}X)^{-1}X^{'}\frac{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)Y(x_{1i},x_{2i})}{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)} \\
&= \frac{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)(Y(x_{1i},x_{2i}) - X\hat{\beta}(x_{1i},x_{2i}))}{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)} \\
&= \frac{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)\hat{\epsilon}(x_{1i},x_{2i})}{\sum_i K(x_{1i},x_{2i}|x_1,x_2,\sigma)}
\end{aligned}
$$

Then, we can conclude that expectation of sample variance for gaussian kernel smoothed beta is smaller than for independent model which shows bias-variance trade off.

$$
\begin{aligned}
\mathbb{E}[\hat{Var}(\hat{\beta}) - \hat{Var}(\tilde{\hat{\beta}})] &= \mathbb{E}[(X^{'}X)^{-1}(\hat{\sigma}^2 - \tilde{\hat{\sigma}}^2)] \\
&= (X^{'}X)^{-1}\frac{1}{T-p}\mathbb{E}[\sum_{t=1}^{T}[(\hat{\epsilon}(t))^2 - (\tilde{\hat{\epsilon}}(t))^2]] \\
&= (X^{'}X)^{-1}\frac{1}{T-p}\sum_{t=1}^{T}[(1 - \frac{\sum_i K_i^2}{(\sum_i K_i)^2})\sigma^2(X(X^{'}X)^{-1}X^{'})_{tt}] \\
&\geq 0
\end{aligned}
$$

Case II: AR(p) errors

Let's assume error terms are following AR(1) model for independent(non-spatial) model. We will estimate ρ by iterative Yule-Walker methods so we assume we have estimator for ρ which is consistent and asymptotically efficient.

Then, set

$$
\begin{aligned}
Y^* &= DY, \;\; X^* = DX \\
&\text{where } D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & \cdots & 0 \\ \vdots & & & \cdots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix} \\
DY &= DX\beta + D\epsilon \\
Y^* &= X^*\beta + u \\
&\text{where } u \sim N(0,\sigma^2)
\end{aligned}
$$

For independently fitted beta, with σ unknown, it's also unbiased.

$$
\begin{aligned}
\mathbb{E}[\hat{\beta}] &= \mathbb{E}[(X^{*^{'}}X^*)^{-1}X^{*^{'}}Y^*] \\
&= \beta
\end{aligned}
$$

$$
\begin{aligned}
\hat{Var}[\hat{\beta}] &= (X^{*'} X^{*})^{-1}\hat{\sigma}^2 \\
&= (X^{'} D^{'} DX)^{-1}\hat{\sigma}^2 \\
&\text{where } \hat{\sigma}^2 = \frac{1}{T-p}\hat{u}^{'}\hat{u} \\
&\text{where } \hat{u} = DY - DX\hat{\beta}
\end{aligned}
$$

For gaussian kernel smoothing beta we have biased beta but reduced variance of estimated sample variance.

$$
\begin{aligned}
\mathbb{E}[\hat{Var}(\hat{\beta}) - \hat{Var}(\tilde{\hat{\beta}})] &= (X^{*'} X^{*})^{-1}[\mathbb{E}(\hat{\sigma}^2 - \frac{\sum_i [K_i^2]}{[\sum_i K_i]^2}\tilde{\hat{\sigma}}^2)] \\
&\geq 0
\end{aligned}
$$