# Monotone B-spline Smoothing for a Generalized Linear Model Response Supplementary Material

Wei Wang *

Center for Outcomes Research, Children's Hospital of Philadelphia

Dylan S. Small †

Department of Statistics, Wharton School, University of Pennsylvania

## APPENDIX A: B-spline basis functions

We provide a brief description of B-splines basis functions and associated

*Wei Wang is Senior Statistician, Center for Outcomes Research, Children's Hospital of Philadelphia, Philadelphia, PA, 19104 (email: wangw@email.chop.edu).

†Dylan S. Small is Professor, Department of Statistics, the Wharton School, University of Pennsylvania, Philadelphia, PA, 19104 (email: dsmall@wharton.upenn.edu).

properties. See de Boor (1978) for complete and detailed studies. Roughly speaking, B-spline basis functions are piecewise polynomials. The definition of B-spline basis functions relies on two concepts, the knots and the degree of the polynomial. The knots are a set of non-decreasing numbers, $x_1 \leq \ldots \leq x_p \leq x_{p+1} \leq \ldots$ . A recursive definition of the $p$-th B-spline basis function of degree $k$, $\psi_p^k(x)$, is

$$\psi_p^0(x) = \begin{cases} 1, \text{ if } x_p \leq x < x_{p+1}; \\ 0, \text{ otherwise.} \end{cases}$$

$$\psi_p^k(x) = \frac{x - x_p}{x_{p+k} - x_p}\psi_p^{k-1}(x) + \frac{x_{p+k+1} - x}{x_{p+k+1} - x_{p+1}}\psi_{p+1}^{k-1}(x).$$

As in de Boor (1978), the derivative of a B-spline basis function is

$$\frac{d\ \psi_p^k(x)}{d\ x} = k\left(\frac{\psi_p^{k-1}(x)}{x_{p+k} - x_p} - \frac{\psi_{p+1}^{k-1}(x)}{x_{p+k+1} - x_{p+1}}\right).$$

Let $\beta_p$'s be a series of coefficients and consider a linear combination of B-spline basis functions $\sum_p \psi_p^k(x)\beta_p$. Its derivative follows as $k\sum_p(\beta_{p+1} - \beta_p)\psi_{p+1}^{k-1}(x)/(x_{p+k+1} - x_{p+1})$. Since the $\psi_p^{k-1}$'s are non-negative, an increasing order of the coefficients ensures that the linear combination is an increasing function of $x$.

# APPENDIX B: Approximation of the log likelihood

Given the estimate at the previous estimation, to obtain (3), we approximate the log likelihood by Taylor expansion to the second order. Let $l_i(\boldsymbol{\beta})$ be the $i$-th individual's log likelihood function, i.e. $\Lambda(\boldsymbol{\beta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\beta})$. As assumed, the distribution of $y_i$ is in the exponential family (McCullagh and Nelder, 1989). It is clear $\mu_i$ depends on $\boldsymbol{\beta}$ through the linear predictor $\eta_i$. By standard calculation (O'Sullivan et al., 1986; Eilers and Marx, 1996), we get

$$\frac{\partial l_i(\boldsymbol{\beta})}{\partial \beta_k} = \frac{\partial l_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_k} = \frac{y_i - \mu_i}{v_i}\frac{\partial \mu_i}{\partial \eta_i}\phi_k(x_i),$$

$$\mathrm{E}\frac{\partial^2 l_i(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_{k'}} = -\frac{1}{v_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \phi_k(x_i)\frac{\partial \eta_i}{\partial \beta_{k'}} = -\frac{1}{v_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \phi_k(x_i)\phi_{k'}(x_i),$$

where $k',\ k = 1, \ldots, K+1$.

We obtain the Taylor expansion of $l_i(\boldsymbol{\beta})$ at the previous estimate $\tilde{\boldsymbol{\beta}}$ to the second order, replacing the second order derivative by its expected value, as

$$\Lambda(\boldsymbol{\beta}) \approx \Lambda(\tilde{\boldsymbol{\beta}}) + \left[\sum_{i=1}^{n} \frac{y_i - \mu_i}{v_i}\frac{\partial \mu_i}{\partial \eta_i}\boldsymbol{\phi}_i'\right](\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'\left[\sum_{i=1}^{n}\frac{1}{v_i}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \boldsymbol{\phi}_i\boldsymbol{\phi}_i'\right](\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),$$

where $\mu_i$, $v_i$ and $\partial \mu_i / \partial \eta_i$ are evaluated at $\tilde{\boldsymbol{\beta}}$. We rearrange the terms and leave out those independent of $\boldsymbol{\beta}$ to get

$$\Lambda(\boldsymbol{\beta}) \approx \left[\sum_{i=1}^{n} \frac{y_i - \mu_i}{v_i} \frac{\partial \mu_i}{\partial \eta_i} \boldsymbol{\phi}_i'\right] \boldsymbol{\beta} + \tilde{\boldsymbol{\beta}}' \left[\sum_{i=1}^{n} \frac{1}{v_i} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \boldsymbol{\phi}_i \boldsymbol{\phi}_i'\right] \boldsymbol{\beta} - \frac{1}{2} \boldsymbol{\beta}' \left[\sum_{i=1}^{n} \frac{1}{v_i} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 \boldsymbol{\phi}_i \boldsymbol{\phi}_i'\right] \boldsymbol{\beta}.$$

Expressed in a vector/matrix form, we have $\Lambda(\boldsymbol{\beta}) \approx \left[(\mathbf{y} - \tilde{\boldsymbol{\mu}})\tilde{\mathbf{D}} + \tilde{\boldsymbol{\beta}}'\mathbf{B}'\tilde{\mathbf{W}}\right]' \mathbf{B}\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}'\mathbf{B}'\tilde{\mathbf{W}}\mathbf{B}\boldsymbol{\beta}$. Hence, maximizing $\Lambda(\boldsymbol{\beta}) - (\lambda/2)\boldsymbol{\beta}'\mathbf{P}\boldsymbol{\beta}$ over $\boldsymbol{\beta}$ is approximated by the following iteration. Given the previous estimate $\tilde{\boldsymbol{\beta}}$ and its associated quantities, we maximize $\left[(\mathbf{y} - \tilde{\boldsymbol{\mu}})\tilde{\mathbf{D}} + \tilde{\boldsymbol{\beta}}'\mathbf{B}'\tilde{\mathbf{W}}\right]' \mathbf{B}\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{\beta}'\left(\mathbf{B}'\tilde{\mathbf{W}}\mathbf{B} + \lambda\mathbf{P}\right)\boldsymbol{\beta}$ to find the current estimate of $\boldsymbol{\beta}$.

# References

de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. New York: Chapman & Hall, 2nd edition.

O'Sullivan, F., Yandell, B. S., and Raynor, W. J. J. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of American Statistical Association*, 81(393):96–103.