Smoothing Spline Estimation of Variance Functions

Anna LIU, Tiejun TONG, and Yuedong WANG

This article considers spline smoothing of variance functions. We focus on selection of the smoothing parameters and develop three direct data-driven methods: unbiased risk (UBR), generalized approximate cross-validation (GACV), and generalized maximum likelihood (GML). In addition to guaranteed convergence, simulations show that these direct methods perform better than existing indirect UBR, generalized cross-validation (GCV), and GML methods. The direct UBR and GML methods perform better than the GACV method. An application to array-based comparative genomic hybridization data illustrates the usefulness of the proposed methods.

Key Words: Array-based comparative genomic hybridization; Generalized approximate cross-validation; Generalized maximum likelihood; Heteroscedasticity; Smoothing parameter; Unbiased risk.

1. INTRODUCTION

Modeling local variability in terms of variance functions is an important problem with a wide range of applications. For example, variance function estimation is necessary in finance, quality control, and immunoassay for measuring volatility or risk (Andersen and Lund 1997; Gallant and Tauchen 1997), experimental design (Box 1988), prediction (Carroll 1987; Yao and Tong 1994) and calibration (Raab 1981; Watters, Carroll, and Spiegelman 1987). Variance estimation is especially important for detecting genes with differential expression across experimental conditions based on microarray data (Huang and Pan 2002; Wang and Guo 2004). With a small number of replicated array experiments, the standard estimates of variances are unreliable. Various methods have been proposed to improve estimation of the variances which usually lead to more powerful tests (Huang and Pan 2002; Wang and Guo 2004; Cui et al. 2005). We apply our variance function estimation

---

Anna Liu is Assistant Professor, Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003 (E-mail: *anna@math.umass.edu*). Tiejun Tong is Postdoctoral Associate, Department of Epidemiology and Public Health, Yale University, New Haven, CT 06520 (E-mail: *tiejun.tong@yale.edu*). Yuedong Wang is Professor, Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106 (E-mail: *yuedong@pstat.ucsb.edu*). Address for correspondence: Yuedong Wang, Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106.

Journal of Computational and Graphical Statistics, Volume 16, Number 2, Pages 312–329
DOI: 10.1198/106186007X204528

methods to array-based comparative genomic hybridization (aCGH) data in Section 5. More applications of variance function estimation can be found in Carroll and Ruppert (1988).

Research on nonparametric estimation of variance functions has attracted a great deal of attention (Carroll 1982; Silverman 1985; Hall and Carroll 1989; Ruppert, Wand, Holst, and Hössjer 1997; Fan and Yao 1998; Yuan and Wahba 2004; Dai and Guo 2005). Most research concentrates on heteroscedastic regression. Both local polynomial smoothers and smoothing splines were used to model the variance function nonparametrically. Within the smoothing spline framework, Yuan and Wahba (2004) used the GACV method to select the smoothing parameter for estimating the variance function, while Dai and Guo (2005) treated squared pseudo-residuals (lag-one differences) as Gaussian data. The main goal of this article is to develop and compare several data-driven smoothing parameter selection methods for the smoothing spline estimation of variance functions.

To simplify exposition, we focus on the situation when direct observations on a variance function are available. Specifically, we have independent observations $\{(x_i, y_i),\ i = 1, \dots, n\}$, where

$$y_i = \exp(f(x_i))\chi^2_{i,k}/k, \tag{1.1}$$

$\chi^2_{i,k}$ are iid Chi-square random variables with k degrees of freedom. Our goal is to estimate the variance function f nonparametrically.

One typical situation leading to model (1.1) is the following heteroscedastic regression model with replicates

$$z_{ij} = \mu_i + \exp(f(x_i)/2)\epsilon_{ij}, \quad i = 1, \dots, n;\ j = 1, \dots, k+1, \tag{1.2}$$

where $\epsilon_{ij} \overset{\text{iid}}{\sim} N(0, 1)$. Then the sample variances $y_i = \sum_{j=1}^{k+1}(z_{ij} - \sum_{j=1}^{k+1} z_{ij}/(k+1))^2/k$ follow model (1.1). See Section 5 and Raab (1981) for real examples of model (1.2).

Most nonparametric methods for heteroscedastic regression use either of the following as observations on the variance function: squared residuals after fitting a nonparametric model to the mean function (Silverman 1985; Hall and Carroll 1989; Ruppert, Wand, Holst, and Hössjer 1997; Fan and Yao 1998; Yuan and Wahba 2004), or squared pseudo-residuals after removing the mean by differences (Müller and Stadtmüller 1987; Fan and Yao 1998; Dai and Guo 2005). Under suitable conditions, these squared residuals and pseudo-residuals follow model (1.1) asymptotically (Fan and Yao 1998; Dai and Guo 2005) with $k = 1$. Therefore, results in this article also shed light on the selection of smoothing parameters for these methods where k is small.

We introduce smoothing spline models and an estimation procedure in Section 2. In Section 3, we review some existing indirect methods and propose three direct methods for estimating the smoothing parameters. We report results of extensive simulations in Section 4 and apply our methods to a real aCGH dataset in Section 5.

2. SMOOTHING SPLINE MODELS FOR VARIANCE FUNCTIONS

We assume model (1.1) in the remainder of this article. Considering the chi-square distribution as a special case of the Gamma distribution, we have $y_i \sim$ Gamma$(k/2,$

$2\exp(f(x_i))/k$ with log-likelihood

$$l_i(f_i) = -ky_i \exp(-f_i)/2 - kf_i/2 + c(y_i), \tag{2.1}$$

where $f_i = f(x_i)$ and $c(y_i) = \log(y_i^{k/2-1}(k/2)^{k/2}/\Gamma(k/2))$, which is independent of f_i. Spline smoothing methods for the exponential family in Wahba et al. (1995) can be employed to estimate the function f. Note, however, that instead of the canonical link (reciprocal) used in the previous publications, we use the logarithmic link in this article in order to free the positive constraint on the variance function. We also propose new direct methods for selecting the smoothing parameters.

For simplicity, we assume that the domain of the function f is [0, 1] and that f belongs to the reproducing kernel Hilbert space (RKHS)

$$W_m([0,1]) = \left\{ f : f, f', \ldots, f^{(m-1)} \text{absolutely continuous}, \int_0^1 \left(f^{(m)}\right)^2 dx < \infty \right\}. \tag{2.2}$$

Our methods apply to general smoothing spline and smoothing spline ANOVA models (Wahba et al. 1995). The smoothing spline estimate of f, f_λ, is the minimizer of the penalized likelihood

$$-\sum_{i=1}^{n} l_i(f_i) + \frac{n\lambda}{2} \int_0^1 \left(f^{(m)}\right)^2 dx, \tag{2.3}$$

where λ is a smoothing parameter controlling the trade-off between the goodness-of-fit and the smoothness of the function. Let B_r be Bernoulli polynomials, $\phi_r(x) = B_{r-1}(x)/(r-1)!$, $r = 1, \ldots, m$, and $R_1(s,t) = \phi_{m+1}(s)\phi_{m+1}(t) + (-1)^{m-1}\phi_{2m+1}(s-t)$. Then the solution to (2.3) is (Wahba et al. 1995; Gu 2002)

$$f_\lambda(x) = \sum_{i=1}^{m} d_i \phi_i(x) + \sum_{i=1}^{n} c_i R_1(x_i, x). \tag{2.4}$$

For any fixed λ, coefficients $\mathbf{c} = (c_1, \ldots, c_n)^T$ and $\mathbf{d} = (d_1, \ldots, d_m)^T$ can be solved by the Newton procedure (Wahba et al. 1995). Let $u_i = -dl_i(f_i)/df_i = -ky_i \exp(-f_i)/2 + k/2$ and $w_i = -d^2 l_i(f_i)/df_i^2 = ky_i \exp(-f_i)/2$. The Newton procedure updates $\mathbf{c}$ and $\mathbf{d}$ by iteratively reformulating the minimization problem (2.3) as

$$\sum_{i=1}^{n} w_{i-}(\tilde{y}_i - f_i)^2 + n\lambda \int_0^1 \left(f^{(m)}\right)^2 dx, \tag{2.5}$$

where $\tilde{y}_i = f_{i-} - u_{i-}/w_{i-}$ is the pseudo-data; the subscript *minus* denotes its corresponding quantities evaluated at f from the previous Newton iteration (Gu 1992; Wahba et al. 1995).

3. METHODS FOR SELECTING SMOOTHING PARAMETERS

A good choice of the smoothing parameter λ is crucial to the performance of the smoothing spline estimate f_λ. One measure of the discrepancy between the spline estimate f_λ and the true function f is the Kullback-Leibler distance (Wahba et al. 1995)

$$\begin{aligned}\mathrm{KL}(f, f_\lambda) &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}_{f_i}\left(l_i(f_i) - l_i(f_{\lambda i})\right) \\ &= \frac{k}{2n}\sum_{i=1}^{n}(\exp(f_i - f_{\lambda i}) + f_{\lambda i}) - \frac{k}{2n}\sum_{i=1}^{n}(1 + f_i). \qquad (3.1)\end{aligned}$$

Ignoring the last term in (3.1), which is independent of λ, and also the multiplying constant $k/2$, we consider the comparative Kullback-Leibler criterion

$$\mathrm{CKL}(f, f_\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\exp\left(f_i - f_{\lambda i}\right) + f_{\lambda i}\right). \qquad (3.2)$$

One approach is to find λ such that $\mathrm{CKL}(f, f_\lambda)$ is minimized. However, similar to other loss functions, $\mathrm{CKL}(f, f_\lambda)$ cannot be minimized directly since f is unknown. Therefore, we need to find estimates or proxies of the CKL criterion. We review some existing indirect methods in Section 3.1 and propose three direct data-driven methods in Sections 3.2-3.4.

3.1 INDIRECT METHODS

Note that (2.5) is the penalized weighted least squares criterion for the working variables $\tilde{y}_i$ and working weights w_{i-}. An indirect (or iterative) method chooses a smoothing parameter at each iteration for the reformulated problem (2.5) and hopes it will converge (Gu 1992). To estimate λ at each iteration, we may use the UBR, GCV, or GML methods which respectively minimize the following UBR, GCV, and GML scores

$$\begin{aligned}U_{-}(\lambda) &= \frac{1}{n}\|(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{W}_{-}^{1/2}\tilde{\mathbf{y}}\|^2 + 2\frac{\hat{\sigma}^2}{n} tr\mathbf{A}(\lambda), \qquad (3.3)\\ V_{-}(\lambda) &= \frac{1/n\|(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{W}_{-}^{1/2}\tilde{\mathbf{y}}\|^2}{[(1/n)tr(\mathbf{I} - \mathbf{A}(\lambda))]^2}, \qquad (3.4)\\ M_{-}(\lambda) &= \frac{\tilde{\mathbf{y}}^T\mathbf{W}_{-}^{1/2}(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{W}_{-}^{1/2}\tilde{\mathbf{y}}}{[\det^{+}(\mathbf{W}_{-}^{1/2}(\mathbf{I} - \mathbf{A}(\lambda))\mathbf{W}_{-}^{1/2})]^{\frac{1}{n-m}}}, \qquad (3.5)\end{aligned}$$

where $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^T$, $\mathbf{W}_{-} = \mathrm{diag}(w_{1-}, \ldots, w_{n-})$, $f_{\lambda i} = f_\lambda(x_i)$, $\mathbf{A}(\lambda)$ satisfies $(w_{1-}^{1/2} f_{\lambda 1}, \ldots, w_{n-}^{1/2} f_{\lambda n})^T = \mathbf{A}(\lambda)(w_{1-}^{1/2}\tilde{y}_1, \ldots, w_{n-}^{1/2}\tilde{y}_n)^T$, $\hat{\sigma}^2 = \sum_{i=1}^{n} u_{i-}^2/nw_{i-}$, and $\det^{+}$ is the product of the nonzero eigenvalues (Wahba et al. 1995; Gu 2002).

The target criteria of these indirect methods change throughout the iterations. Therefore, one conceptual problem of the indirect methods is that their overall target criteria are not explicitly known. Asymptotically, $U_{-}(\lambda)$ provides a proxy of $\mathrm{KL}(f, f_\lambda)$ (Wang, Wahba, Chappell, and Gu 1995). No justifications have been provided for the indirect GCV and GML methods. One practical problem with the indirect methods is that they do not guarantee convergence. For binomial and Poisson data, extensive simulations indicate that convergence is achieved for almost all situations (Wang et al. 1995). The performance of these indirect methods for Gamma data has not been studied. For binomial and Poisson data, some direct methods have been developed and have been found to work better than the indirect methods (Xiang and Wahba 1996; Gu and Xiang 2001; Yuan 2005). However, direct methods for Gamma data have not been developed.

3.2 UNBIASED RISK METHOD

In this subsection we derive an unbiased estimate of $\mathrm{E}(\mathrm{CKL}(f, f_\lambda))$. Let $h_\lambda(i, z, \cdot)$ be the minimizer of (2.3) when the ith observation, y_i, is replaced by z. Let $g_{\lambda i}(z) = h_\lambda(i, z, x_i)$ and $v(t) = \int_0^t \exp(-g_{\lambda i}(z)) z^{k/2-1} dz$. Then for any fixed $\mathbf{y}_{-i} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)^T$, when $k \geq 3$, we have

$$\begin{aligned}
&\mathrm{E}(\exp(f_i - f_{\lambda i}))\\
&\quad= \frac{\exp(f_i)}{\Gamma(k/2)} \left(\frac{2\exp(f_i)}{k}\right)^{-\frac{k}{2}} \int_0^\infty \exp(-g_{\lambda i}(t)) t^{\frac{k}{2}-1} \exp\left(\frac{-kt}{2\exp(f_i)}\right) dt\\
&\quad= \frac{\exp(f_i)}{\Gamma(k/2)} \left(\frac{2\exp(f_i)}{k}\right)^{-\frac{k}{2}} \left\{ v(t) \exp\left(\frac{-kt}{2\exp(f_i)}\right)\Bigg|_0^\infty \right.\\
&\quad\quad \left. - \int_0^\infty v(t) d\left(\exp\left(\frac{-kt}{2\exp(f_i)}\right)\right)\right\}\\
&\quad= \frac{k}{2\Gamma(k/2)} \left(\frac{2\exp(f_i)}{k}\right)^{-\frac{k}{2}} \int_0^\infty v(t) \exp\left(\frac{-kt}{2\exp(f_i)}\right) dt\\
&\quad= \frac{k}{2}\mathrm{E}\left(v(y_i) y_i^{-(k/2-1)}\right),
\end{aligned}$$

where we used the facts that $\mathrm{E}(\exp(-f_{\lambda i}))$ exists and $v(t)\exp(-kt/2\exp(f_i))|_0^\infty = 0$ (Appendix A, p. 325). The derivation is similar to those in Berger (1980) and Wong (2006), where the above facts were assumed as conditions. Then, when $k \geq 3$, an unbiased estimator of $\mathrm{E}(\mathrm{CKL}(f, f_\lambda))$ is

$$\mathrm{UBR}(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{k}{2} v(y_i) y_i^{-(k/2-1)} + f_{\lambda i} \right). \tag{3.6}$$

The direct UBR estimate of λ is the minimizer of $\mathrm{UBR}(\lambda)$. Gaussian quadrature may be used to approximate $v(y_i) = \int_0^{y_i} \exp(-g_{\lambda i}(z)) z^{k/2-1} dz$. However, it requires calculating $h_\lambda(i, z, \cdot)$ for several different values of z, which can be computationally intensive when the sample size is large. Simulations indicate that $\exp(g_{\lambda i}(z))$ is approximately linear in z: $\exp(g_{\lambda i}(z)) \approx \beta_0 + \beta_1 z$. We have tried several methods to compute the coefficients β_0 and β_1, and found that the following two work well: (1) compute $g_{\lambda i}(y_i)$ and $g_{\lambda i}(y_i/2)$, and then compute β_0 and β_1 as the intercept and slope of the line joining two points $(y_i/2, \exp(g_{\lambda i}(y_i/2)))$ and $(y_i, \exp(g_{\lambda i}(y_i)))$; (2) from Taylor expansion $\exp(g_{\lambda i}(z)) \approx \exp(g_{\lambda i}(y_i)) + \exp(g_{\lambda i}(y_i))(\partial g_{\lambda i}/\partial z)|_{z=y_i}(z - y_i) \approx \exp(f_{\lambda i}) + \exp(f_{\lambda i}) d_{ii}(z - y_i)$, where $(\partial g_{\lambda i}/\partial z)|_{z=y_i} \approx d_{ii}$; this is based on a similar argument as that in Appendix B (p. 326), and d_{ii} is defined later in Section 3.3. Thus, $\beta_0 \approx \exp(f_{\lambda i})(1 - d_{ii} y_i)$ and $\beta_1 \approx \exp(f_{\lambda i}) d_{ii}$. Based on the linear approximation, $v(y_i)$ can be calculated through recursive formulas obtained by lengthy algebra (not shown). Extensive simulations indicate that the above linear approximations lead to similar UBR estimates of the smoothing parameters as those based on Gaussian quadrature; the two methods for computing the coefficients β_0 and β_1 also lead to similar UBR estimates. We use the linear approximation based on the Taylor expansion in our simulations since it requires the least amount of computation.

3.3 GENERALIZED APPROXIMATE CROSS-VALIDATION METHOD

Let $f_\lambda^{(-i)}$ be the minimizer of (2.3) without the ith observation and $f_{\lambda i}^{(-i)} = f_\lambda^{(-i)}(x_i)$. Replacing $\exp(f_i - f_{\lambda i})$ by $y_i \exp(-f_{\lambda i}^{(-i)})$ and ignoring the multiplying constant $1/n$, we obtain a cross-validation estimate of $\mathrm{CKL}(f, f_\lambda)$

$$\mathrm{CV}(\lambda) = \sum_{i=1}^{n} \left(y_i \exp\left(-f_{\lambda i}^{(-i)}\right) + f_{\lambda i} \right). \tag{3.7}$$

It is usually expensive to compute $\mathrm{CV}(\lambda)$ for large n. We now introduce an approximation of $\mathrm{CV}(\lambda)$. Let $\mathbf{f} = (f_1, \ldots, f_n)^T$, $\mathbf{T}_{n\times m} = \{\phi_v(x_i)\}_{i=1}^{n}{}_{v=1}^{m}$ and $\boldsymbol{\Sigma} = \{R_1(x_i, x_j)\}_{i,j=1}^{n}$. Let $\mathbf{T} = (\mathbf{Q}_1\ \mathbf{Q}_2)(\mathbf{R}^T\ \mathbf{0}^T)^T$ be the QR decomposition of $\mathbf{T}$ and $\boldsymbol{\Omega} = \mathbf{Q}_2(\mathbf{Q}_2^T \boldsymbol{\Sigma} \mathbf{Q}_2)^\dagger \mathbf{Q}_2^T$ where $\dagger$ is the Moore-Penrose generalized inverse. Then $\mathbf{f} = \mathbf{Td} + \boldsymbol{\Sigma}\mathbf{c}$, $\int_0^1 (f^{(m)})^2 dx = \mathbf{f}^T \boldsymbol{\Omega} \mathbf{f}$, and the penalized likelihood (2.3) can be rewritten as (Xiang and Wahba 1996)

$$J = -\sum_{i=1}^{n} l_i(f_i) + \frac{n\lambda}{2} \mathbf{f}^T \boldsymbol{\Omega} \mathbf{f}. \tag{3.8}$$

Let $\mathbf{W}_\lambda = \mathrm{diag}(ky_1 \exp(-f_{\lambda 1})/2, \ldots, ky_n \exp(-f_{\lambda n})/2)$, $\mathbf{V} = \mathrm{diag}(k \exp(-f_{\lambda 1})/2, \ldots, k \exp(-f_{\lambda n})/2)$, and $\mathbf{D} = (\mathbf{W}_\lambda + n\lambda\boldsymbol{\Omega})^{-1}\mathbf{V}$. An approximation of $\mathrm{CV}(\lambda)$ is (Appendix B, p. 326)

$$\mathrm{ACV}(\lambda) = L(\lambda) + \sum_{i=1}^{n} \frac{d_{ii} \exp(-f_{\lambda i}) y_i (y_i - \exp(f_{\lambda i}))}{1 - d_{ii} \exp(f_{\lambda i})},$$

where $L(\lambda) = \sum_{i=1}^{n} (y_i \exp(-f_{\lambda i}) + f_{\lambda i})$ and d_{ii} is the ith diagnal element of $\mathbf{D}$. Replacing $\exp(f_{\lambda i}) d_{ii}$ by $\mathrm{tr}(\mathbf{W}_0^{1/2} \mathbf{D} \mathbf{W}_0^{1/2})/n$ where $\mathbf{W}_0 = \mathrm{diag}(\exp(f_{\lambda 1}), \ldots, \exp(f_{\lambda n}))$, we obtain a generalized ACV

$$\mathrm{GACV}_1(\lambda) = L(\lambda) + \frac{\mathrm{tr}\left(\mathbf{W}_0^{1/2} \mathbf{D} \mathbf{W}_0^{1/2}\right)}{n - \mathrm{tr}\left(\mathbf{W}_0^{1/2} \mathbf{D} \mathbf{W}_0^{1/2}\right)} \sum_{i=1}^{n} y_i (y_i - \exp(f_{\lambda i})) \exp(-2 f_{\lambda i}).$$

Replacing $\exp(f_{\lambda i}) d_{ii}$ in the denominator by $\mathrm{tr}(\mathbf{W}_0^{1/2} \mathbf{D} \mathbf{W}_0^{1/2})/n$ and d_{ii} in the numerator by $\mathrm{tr}(\mathbf{D})/n$, we obtain another generalized ACV

$$\mathrm{GACV}_2(\lambda) = L(\lambda) + \frac{\mathrm{tr}(\mathbf{D})}{n - \mathrm{tr}\left(\mathbf{W}_0^{1/2} \mathbf{D} \mathbf{W}_0^{1/2}\right)} \sum_{i=1}^{n} y_i (y_i - \exp(f_{\lambda i})) \exp(-f_{\lambda i}).$$

$\mathrm{GACV}_2(\lambda)$ is the same as the GACV in Yuan and Wahba (2004) (note that there is a typographical error in their formula). The direct GACV1 and GACV2 estimates of λ are minimizers of $\mathrm{GACV}_1(\lambda)$ and $\mathrm{GACV}_2(\lambda)$, respectively.

3.4 GENERALIZED MAXIMUM LIKELIHOOD METHOD

Let $F(x) = \sum_{i=1}^{M} \theta_i \phi_i(x) + b^{\frac{1}{2}} Z(x)$ be the prior for the function f, where $\theta_i \overset{\text{iid}}{\sim} N(0, a)$, $b = 2/kn\lambda$, $Z(x)$ is a Gaussian process independent of θ_i's with $\mathrm{E}(Z(x)) = 0$ and $\mathrm{E}(Z(s)Z(t)) = R_1(s, t)$. We assume that observations $\mathbf{y} = (y_1, \ldots, y_n)^T$ are generated according to model (1.1) conditional on $f = F$. As $a \to \infty$, Gu (1992) showed that the posterior mean $\mathrm{E}(F(x)|\mathbf{y})$ approximately equals the spline estimate $f_\lambda(x)$.

Let u_{ic} and w_{ic} be u_i and w_i, respectively, evaluated at the convergence point of the Newton procedure. Let $\mathbf{u}_c = (u_{1c}, \ldots, u_{nc})^T$, $\mathbf{W}_c = \mathrm{diag}(w_{1c}, \ldots, w_{nc})$, $\mathbf{f}_\lambda = (f_{\lambda 1}, \ldots, f_{\lambda n})^T$, $\mathbf{y}_c = \mathbf{f}_\lambda - \mathbf{W}_c^{-1}\mathbf{u}_c$, $\boldsymbol{\Sigma}_c = \mathbf{W}_c^{1/2}\boldsymbol{\Sigma}\mathbf{W}_c^{1/2}$, and $\mathbf{T}_c = \mathbf{W}_c^{1/2}\mathbf{T}$. Let $(\mathbf{Q}_{1c}\ \mathbf{Q}_{2c})(\mathbf{R}_c^T\ \mathbf{0}^T)^T$ be the QR decomposition of $\mathbf{T}_c$, and $\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ be the spectral decomposition of $\mathbf{Q}_{2c}^T\boldsymbol{\Sigma}_c\mathbf{Q}_{2c}$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_{\nu n}, \nu = 1, \ldots, n-m)$. Let $\mathbf{z} = (z_1, \ldots, z_{n-m})^T = \mathbf{U}^T\mathbf{Q}_{2c}^T\mathbf{W}_c^{1/2}\mathbf{y}_c$. Then, using Equation (16) in Liu, Meiring, and Wang (2005) and ignoring a constant, we have an approximation of the negative log-marginal likelihood of $\mathbf{y}$,

$$\begin{aligned} \mathrm{GML}(\lambda) = &- \sum_{i=1}^{n} l_i(f_{\lambda i}) - \frac{k}{4}\mathbf{u}_c{}^T\mathbf{W}_c^{-1}\mathbf{u}_c \\ &+ \frac{1}{2}\sum_{\nu=1}^{n-m}\left(\ln(\lambda_{\nu n}/n\lambda + 1) + \frac{kz_\nu^2/2}{\lambda_{\nu n}/n\lambda + 1}\right) + \ln|\mathbf{R}_c|. \end{aligned} \quad (3.9)$$

The direct GML estimate of λ is the minimizer of $\mathrm{GML}(\lambda)$. Liu, Meiring, and Wang (2005) used the approximated marginal likelihood to construct GML tests for generalized linear models. This approximation has not been explored as a tool for selecting the smoothing parameter for non-Gaussian data, therefore, its performance is unknown.

4. SIMULATIONS

One simple and commonly used approach is to transform y_i in model (1.1) using the logarithm scale and then fitting the transformed data using a penalized least square smoothing spline. We refer to this simple approach as the transformation method. In the following simulation, we use the GCV method to select the smoothing parameter for the transformation approach.

We generate data from model (1.1) with $x_i = i/n$. We use a factorial design with seven functions

$$\begin{aligned} f_j(x) &= 2\sin(2\pi jx) + 3, \quad j = 1, 2, 3, \\ f_4(x) &= 3 - (5x - 2.5)^2, \\ f_5(x) &= \mathrm{logit}((-1.6x + .9)I_{\{x \le .5\}} + (1.6x - .7)I_{\{x > .5\}}), \\ f_6(x) &= \mathrm{logit}((3.5x/3)I_{\{x \le .6\}} + .7I_{\{x > .6\}}), \\ f_7(x) &= 0.218 - 4.312x, \end{aligned}$$

four different sample sizes $n = 100, 200, 300, 400$, and four different degrees of freedom $k = 1, 2, 3, 4$. Functions f_1-f_7 are similar to those in Gu and Xiang (2001). We fit model

Table 1. Number of replications out of 100 that failed to converge for the indirect UBR method.

	$k=1$						$k=2$							$k=3$				$k=4$
	f_1	$f_{2,3}$	f_4	f_5	f_6	f_7	f_1	f_2	f_3	f_4	f_5	f_6	f_7	$f_{1,2,4,7}$	f_3	f_5	f_6	f_{1-7}
$n=100$	98	100	97	95	93	95	19	37	34	11	8	10	13	0	1	0	0	0
$n=200$	97	100	94	97	90	91	8	15	12	7	6	5	6	0	0	1	0	0
$n=300$	92	100	87	93	85	85	4	6	4	2	3	2	2	0	0	1	1	0
$n=400$	88	100	88	97	86	84	2	2	1	0	3	2	2	0	0	0	0	0

(1.1) using eight methods: direct UBR, direct GML, GACV1, GACV2, indirect UBR, indirect GCV, indirect GML and transformation. For each simulation setting, we repeat the simulation 100 times.

Because UBR, GACV1, GACV2, and GML are all continuous functions of the smoothing parameter, we used the quasi-Newton method to optimize them. Specifically, the R function `optim` with the option `L-BFGS-B` is used. This option performs the quasi-Newton optimization and allows the user to specify lower and upper bounds for the smoothing parameter. We used -10 and 3 as the bounds for $\log_{10}(n\lambda)$. The R code is available from the authors.

Contrary to our experience with the binomial and Poisson data, the indirect methods sometimes fail to converge. Table 1 lists the number of replications in which the indirect UBR method failed to converge. The nonconvergence problem of this method is quite severe for small k. When $k=1$, the nonconvergence problem remains even when the sample size is as large as $n=400$. The indirect GML and GCV failed to converge occasionally with

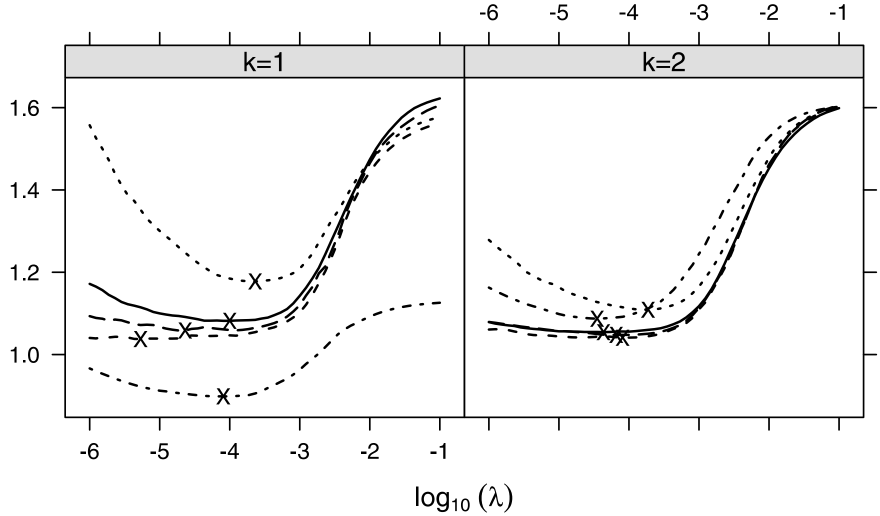

Figure 1. Plots of CKL (solid lines), UBR (long-dashed lines), GACV1 (short-dashed lines), GACV2 (dotted lines), and GML (mixed dotted-dashed lines) criteria as functions of $\log_{10}(\lambda)$ for a typical simulated dataset with $f(x)=f_5(x)$, $n=100$, and $k=1$ and $k=2$. Minima of these curves are marked by crosses.

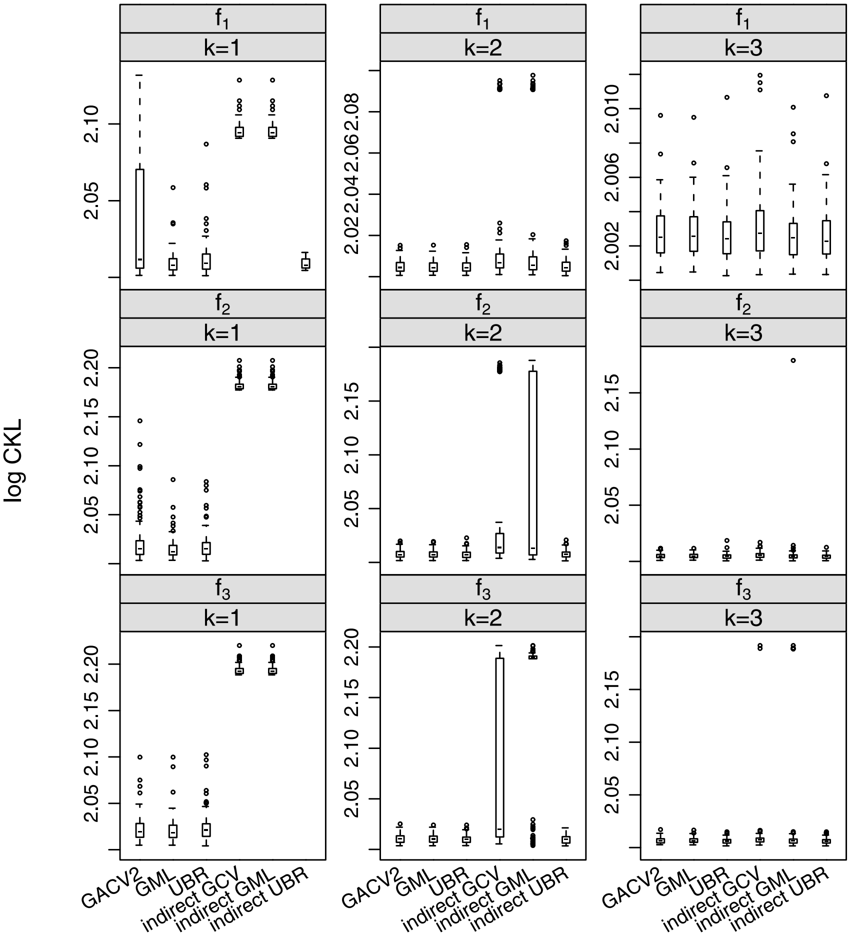

Figure 2. Boxplots of log CKL for f_1-f_3 with $n = 200$. Boxplots for indirect methods are based on converged replications.

frequency ranging from 1% to 4%. Convergence is achieved by all other methods.

Figure 1 contains typical curves of CKL, UBR, GACV2, and GML for $f(x) = f_5(x)$, $n = 100$, and $k = 1$ and $k = 2$. Minimum points are marked by crosses. The same plots for other simulation settings are similar. This figure shows that the UBR function is a good estimate of the CKL criterion, and that the GACV1, GACV2, and GML functions are good proxies. Minima of these functions are usually quite close.

Figures 2 and 3 show boxplots of the CKL values for $n = 200$. The CKL values of the GACV1 and the transformation methods are not shown as they are both generally larger, and much larger for the transformation method, than those obtained from other methods. The boxplots when $k = 4$ are not shown since they are similar to those when $k = 3$. The same plots for other sample sizes are similar. In general, the direct GML and UBR methods perform the best. The indirect UBR method suffers from the nonconvergence problem, and when it does converge, it performs as well as the GML and UBR methods. Although

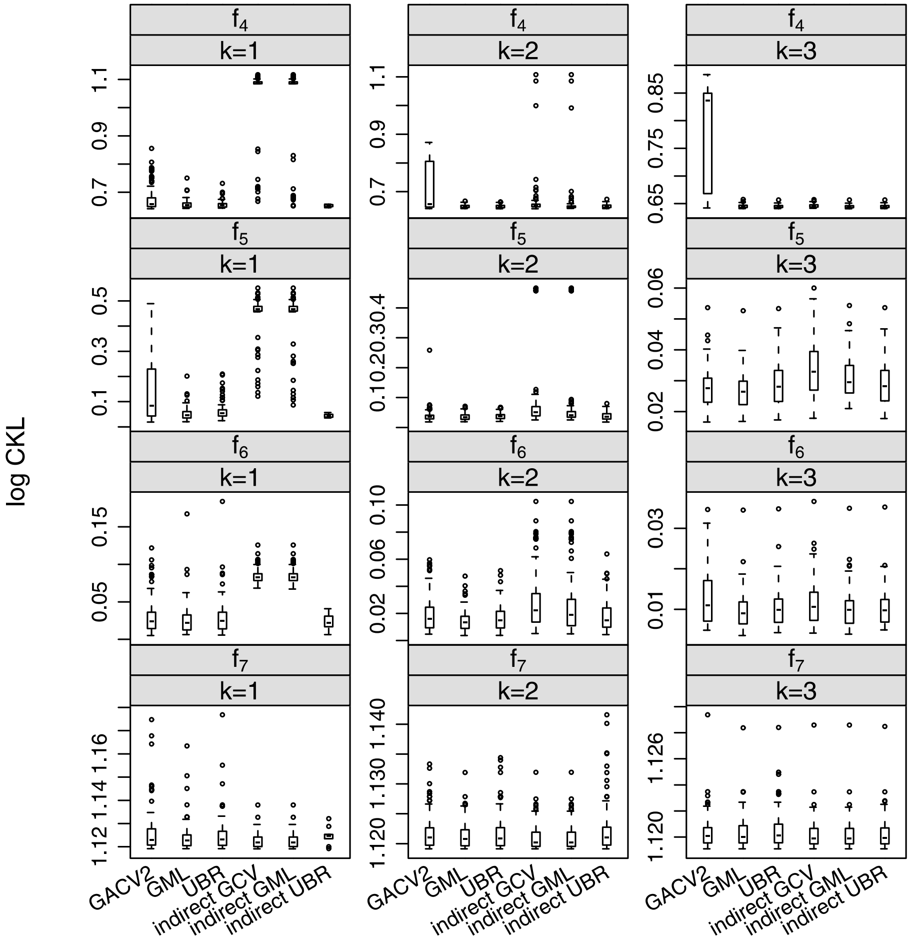

Figure 3. Boxplots of log CKL for f_4-f_7 with $n = 200$. Boxplots for indirect methods are based on converged replications.

the indirect GML and GCV converge most of the time, they tend to over-smooth which leads to larger CKL values. The GACV1 and GACV2 methods also tend to over-smooth occasionally. Even though the unbiasedness property is established for $k \geq 3$ only, we nevertheless applied the direct UBR method to the cases with $k = 1$ and $k = 2$. Performance of the direct UBR method for these cases is similar to that for larger k.

To take a closer look at the performance of various methods, we plot CKL's of the indirect UBR, indirect GML and UBR vs CKL's of the GML for $f(x) = f_6(x)$, $n = 100$, and $k = 1$ and $k = 3$ in Figure 4. The same plots for other simulation settings are similar. In general, the GML tends to have smaller CKL's than the indirect GML; the UBR has similar CKL's as the GML except for a few bad cases; and the indirect UBR has similar CKL's as the GML when it converges.

Figure 5 shows the 5th, 50th, and 95th best estimates ordered by CKL for the first four functions, with $n = 100$ and λ selected by the GML method. The same plots for other

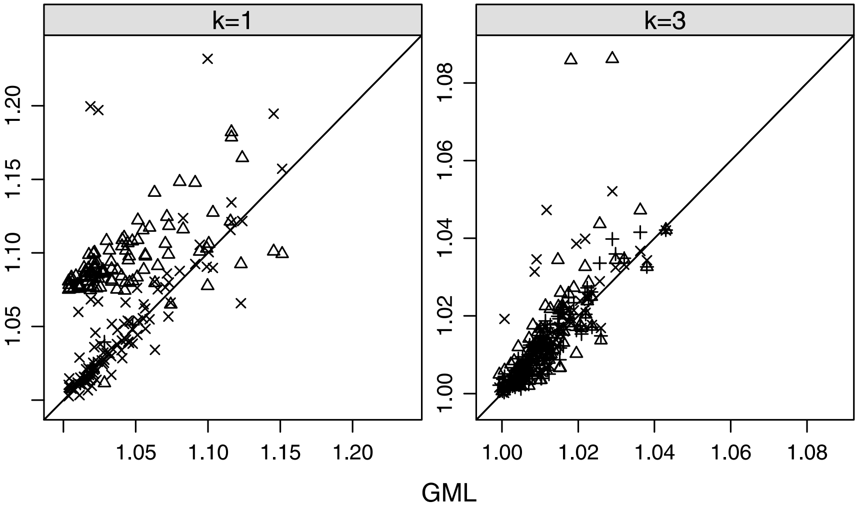

Figure 4. Plots of CKL's of the indirect UBR (+), indirect GML (△) and UBR (×) versus CKL's of GML for $f(x) = f_6(x)$, $n = 100$, and $k = 1$ and $k = 3$.

simulation settings and with the UBR method are similar. We conclude that the smoothing spline estimates of the variance functions with smoothing parameters selected by the GML and UBR methods perform well. The GML method is new for non-Gaussian data. We postulate that it can be extended to binomial and Poisson data, and can perform better than some existing methods.

5. APPLICATION

Comparative genomic hybridization (CGH) is a technique designed for detecting segmental genomic alterations. Recent advances in array-based CGH (aCGH) technology have enabled examination of chromosomal regions in unprecedented detail, revolutionizing our understanding of chromosome alterations associated with tumorigenesis and many developmental abnormalities (Albertson and Pinkel 2003; Mantripragada, Buckley, de Stahl, and Dumanski 2004). The aCGH technology uses two differentially labeled test and reference DNAs which are co-hybridized to cloned genomic fragments immobilized on glass slides. The hybridized DNAs are then detected in two different fluorochromes, and the significant deviation from unity in the ratios of the digitized intensity values is indicative of copy-number differences between the test and reference genomes (Wang and Guo 2004). Accurate identification of amplified or deleted regions requires estimates of variances (Moore, Pallavicini, Cher, and Gray 1997; Wang and Guo 2004). The number of replicated arrays is typically small, owing to the cost. Therefore, the standard sample variance estimates are unreliable and some simple improvements have been proposed (Moore et al. 1997; Tusher, Tibshirani, and Chu 2001; Huang and Pan 2002; Wang and Guo 2004). Wang and Guo (2004) considered the variance as a function of the physical locations in the genome. They

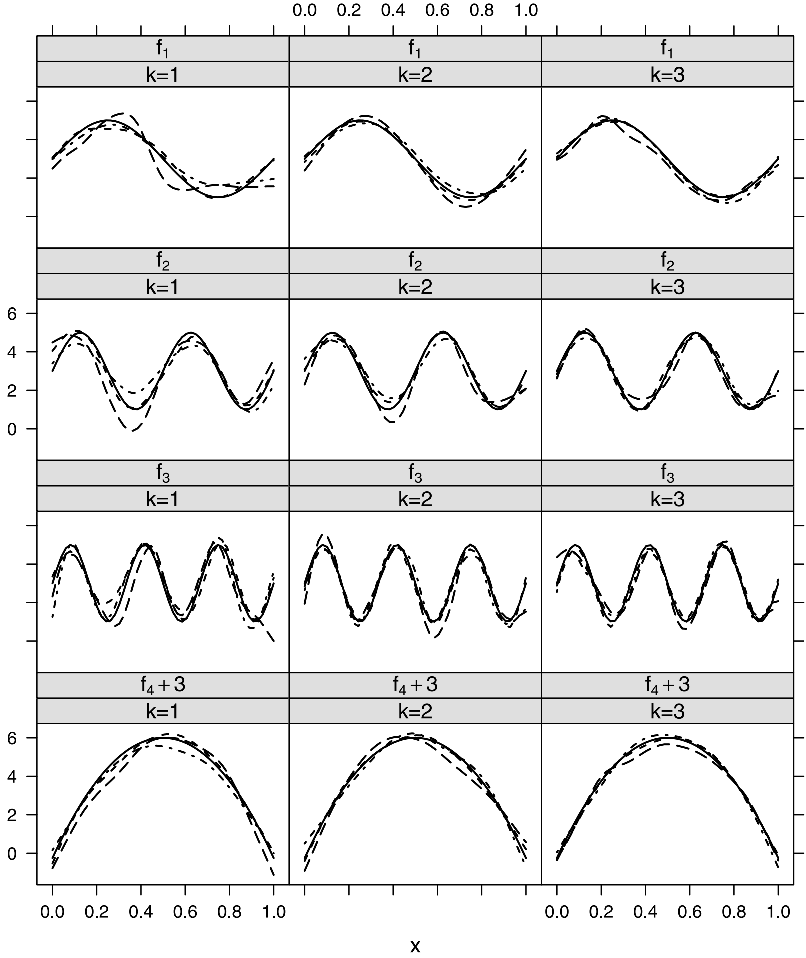

Figure 5. Plots of the true functions (solid lines), the 5th (short-dashed lines), 50th (mixed dotted-dashed lines), and 95th (long-dashed lines) best estimates ordered by CKL. $n = 100$ and λ's are selected by the GML method. To make the y-axis common for all panels, the curves in the last row are shifted up by 3.

applied the lowess smoother to the logarithm of the sample variances with a fixed smoothing parameter. They have shown that using the smoothed variances in the t test can lead to large gains in power. As we have shown in Section 4, the simple approach to smoothing the logarithm of the sample variances is less efficient when the number of replications is small. Also, the selection of the smoothing parameter in Wang and Guo (2004) is somewhat arbitrary.

To illustrate our new methods, we downloaded a well-known BAC array dataset from the Web site *http://www.nature.com/ng/journal/v29/n3/suppinfo/ng754_S1.html*. The data result from an experiment aimed at measuring copy number changes for the cell strains (test samples) against normal male reference DNAs (references), which were co-hybridized on CGH arrays containing 2,460 BAC and P1 clones in triplicate (7,380 spots) (Snijders et al. 2001).

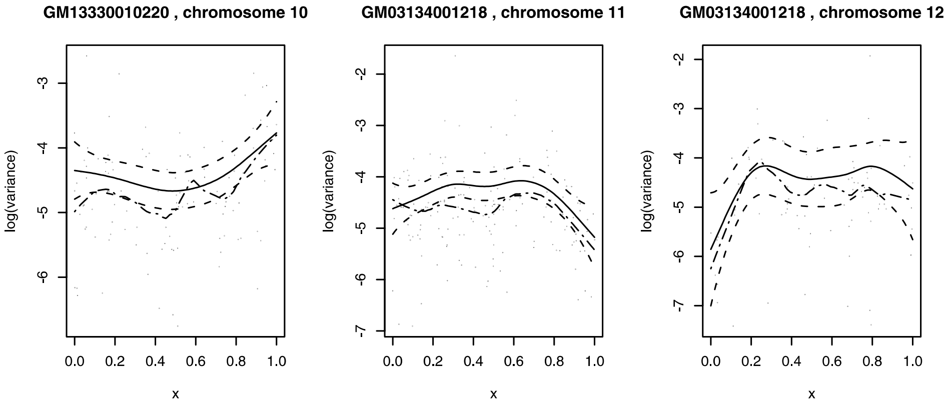

Figure 6. Plots of observations and fits for chromosome 10 of the cell strain GM13330010220 (left), and chromosomes 11 (middle) and 12 (right) of the cell strain GM03134001218. Points are sample variances on the logarithm scale. Solid and mixed dotted-dashed lines are smoothing spline fits and lowess fits, respectively. Two dashed lines in each plot are 95% Bayesian confidence intervals of the smoothing spline fit.

For each chromosome of a cell strain, we assume model (1.2) where z_{ij} is the jth replicate of $\log_2$ fluorescent ratio of clone i, μ_i is the mean fluorescent ratio, x_i is the physical position along the chromosome, and $k = 2$. A nonzero μ_i corresponds to an alteration. Wang and Guo (2004) introduced a smoothed t statistic with variances smoothed along the genome. They have shown that the smoothed t-statistic always improves the performance over the standard t-statistic. We focus on estimation of the variance function. For simplicity we transform the genome position x_i into the interval [0, 1].

We fitted variance functions to all chromosomes of all cell strains. Figure 6 shows observations, fits, and confidence intervals for chromosome 10 of the cell strain GM13330010220, and chromosomes 11 and 12 of the cell strain GM03134001218. For the lowess fits, as in Wang and Guo (2004), we used 30% of the data for smoothing at each position. For the smoothing spline fit, we used the direct GML method to select the smoothing parameter. The approximate Bayesian confidence intervals were computed using the methods in Wahba et al. (1995). Note that the lowess fits are based on the logarithm of the sample variances. Therefore, visually, they are in the center of the data in Figure 6 since variances are plotted on the logarithm scale. However, it is known that naive estimates based on the logarithm of the sample variances are biased downward (Cui et al. 2005; Tong and Wang 2007). A bias correction factor which is greater than one should have been applied before fitting. The lowess estimates are more rough since the smoothing span has been set at 30%, which is the somewhat arbitrary choice used by Wang and Guo (2004).

Some variances are relatively large and some variances are very small (Figure 6). Replacing variances by their smoothed estimates also reduces the effect of outliers and prevents clones with very small variances from dominating the results. Other approaches have been proposed to deal with very small variances (Tusher, Tibshirani, and Chu 2001).

6. CONCLUSIONS

Nonparametric estimation of variance functions has been well-studied in the literature, while the related problem of selection of the smoothing parameters has received less attention. Within the smoothing spline framework, the existing indirect UBR, GCV, and GML methods for selecting the smoothing parameter do not guarantee convergence. In this article we have proposed the direct UBR, GACV, and GML methods and found that the direct UBR and GML methods have the best performance in general.

APPENDIX A

We show that when $k \geq 3$, (a) $v(t)\ \exp(-kt/2\exp(f_i))|_0^\infty = 0$, and (b) $\mathrm{E}(\exp(-f_{\lambda i})) < \infty$.

For (a), we show that $\lim_{t\to 0} v(t) = 0$ and $\lim_{t\to\infty} (v(t)\exp(-kt/2\exp(f_i))) = 0$. We use the following fact that for any $a > 0$ and x,

$$a\exp(-x) + x \geq 1 + \ln a. \tag{A.1}$$

The above inequality can be verified by finding the minimum of the left-hand side as a function of x.

Note that $\mathbf{y}_{-i}$ is fixed and the ith observation is replaced by z. The penalized likelihood (2.3) is equivalent to

$$I(f) = z\exp(-f_i) + f_i + \sum_{j\neq i}^{n} \big(y_j\exp(-f_j) + f_j\big) + \frac{n\lambda}{k}\int_0^1 \Big(f^{(m)}\Big)^2\,dx. \tag{A.2}$$

Note that $h_\lambda(i, z, \cdot)$ is the minimizer of (A.2) and $g_{\lambda i}(z) = h_\lambda(i, z, x_i)$.

We assume that the constant function, $f_0(x) = 0$, belongs to the model space. This is true when the model space is $W_m([0, 1])$. $I(f_0) = z + \sum_{j\neq i} y_j$. Since $h_\lambda(i, t, \cdot)$ is the minimizer of $I(f)$, we have

$$\begin{aligned} z\exp(-g_{\lambda i}(z)) + g_{\lambda i}(z) \quad &\leq \quad I(h_\lambda(i, z, \cdot)) - \sum_{j\neq i} \big(y_j\exp(-h_\lambda(i, z, x_j)) + h_\lambda(i, z, x_j)\big) \\ &\leq \quad I(f_0) - \sum_{j\neq i} \big(y_j\exp(-h_\lambda(i, z, x_j)) + h_\lambda(i, z, x_j)\big) \\ &\leq \quad z + \sum_{j\neq i} y_j - \sum_{j\neq i}(1 + \ln y_j), \end{aligned} \tag{A.3}$$

where the last inequality is based on (A.1). Therefore, as $z \to 0$, $z\exp(-g_{\lambda i}(z)) + g_{\lambda i}(z)$ is bounded above.

We claim that for any power $p > 1$,

$$\lim_{z\to 0} z^p \exp(-g_{\lambda i}(z)) = 0. \tag{A.4}$$

Otherwise, there exists some $p_1 > 1$ such that $\limsup_{z\to 0} z^{p_1}\exp(-g_{\lambda i}(z)) > 0$. Let $p_2 = (p_1 + 1)/2$. Then $\limsup_{z\to 0} z^{p_2}\exp(-g_{\lambda i}(z)) = \infty$. There exists a sequence $z_m \to 0$ such that $-g_{\lambda i}(z_m) + p_2 \ln z_m \to \infty$. Also, $-g_{\lambda i}(z_m) + \ln z_m \geq -g_{\lambda i}(z_m) + p_2 \ln z_m \to \infty$. Therefore, for large enough m, we have $(-g_{\lambda i}(z_m) + \ln z_m)/2 \geq 1/(p_2 - 1) > 0$. Thus, for large enough m,

$$\begin{aligned} z_m \exp(-g_{\lambda i}(z_m)) + g_{\lambda i}(z_m) &= \exp\left(-g_{\lambda i}(z_m) + \ln z_m\right) + g_{\lambda i}(z_m) \\ &\geq 1 - g_{\lambda i}(z_m) + \ln z_m + (-g_{\lambda i}(z_m) \\ &\quad + \ln z_m)^2/2 + g_{\lambda i}(z_m) \\ &\geq \ln z_m + \left(-g_{\lambda i}(z_m) + \ln z_m\right)/(p_2 - 1) \\ &= \left(-g_{\lambda i}(z_m) + p_2 \ln z_m\right)/(p_2 - 1). \end{aligned}$$

The left-hand side is bounded above, while the right-hand side approaches ∞, a contradiction.

Taking $p = 1.1$ in (A.4), when $k \geq 3$, we have

$$\begin{aligned} \limsup_{t\to 0} v(t) &= \limsup_{t\to 0} \int_0^t \left(z^{1.1} \exp(-g_{\lambda i}(z))\right) z^{\frac{k}{2}-2.1} dz \\ &\leq \limsup_{t\to 0} \int_0^t z^{\frac{k}{2}-2.1} dz = 0. \end{aligned} \tag{A.5}$$

Therefore, $\lim_{t\to 0} v(t) = 0$ since $v(t) \geq 0$.

As $z \to \infty$, from (A.3), we have $z\exp(-g_{\lambda i}(z)) + g_{\lambda i}(z) \leq 2z$. On the other hand, from (A.1) with $a = z/2$, $-g_{\lambda i}(z) \leq z\exp(-g_{\lambda i}(z))/2$ when $z \geq 2$. Therefore, $z\exp(-g_{\lambda i}(z)) \leq 2z - g_{\lambda i}(z) \leq 2z + z\exp(-g_{\lambda i}(z))/2$. There exists some large T such that

$$\exp(-g_{\lambda i}(z)) \leq 4, \quad z \geq T. \tag{A.6}$$

Now,

$$v(t) = \int_0^T \exp(-g_{\lambda i}(z)) z^{\frac{k}{2}-1} dz + \int_T^t \exp(-g_{\lambda i}(z)) z^{\frac{k}{2}-1} dz \leq v(T) + \frac{8}{k} t^{\frac{k}{2}}.$$

Then

$$\limsup_{t\to\infty} \left(v(t)\exp\left(\frac{-kt}{2\exp(f_i)}\right)\right) \leq \limsup_{t\to\infty} \left(v(T) + \frac{8}{k}t^{\frac{k}{2}}\right) \exp\left(\frac{-kt}{2\exp(f_i)}\right) = 0.$$

Therefore, $\lim_{t\to\infty} (v(t)\exp(-kt/2\exp(f_i))) = 0$.

Now we prove (b) $\mathrm{E}(\exp(-f_{\lambda i})) < \infty$. From (A.4), when $k \geq 3$, there exists a small enough ϵ_1 such that $\mathrm{E}(\exp(-f_{\lambda i}) I_{\{y_i < \epsilon_1\}}) \leq \mathrm{E}(y_i^{-1.1} I_{\{y_i < \epsilon_1\}}) < \infty$. Similarly, from (A.6), there exists a large enough ϵ_2 such that $\mathrm{E}(\exp(-f_{\lambda i}) I_{\{y_i > \epsilon_2\}}) < \infty$. Then $\mathrm{E}(\exp(-f_{\lambda i})) = \mathrm{E}(\exp(-f_{\lambda i}) I_{\{y_i < \epsilon_1\}}) + \mathrm{E}(\exp(-f_{\lambda i}) I_{\{y_i > \epsilon_2\}}) + \mathrm{E}(\exp(-f_{\lambda i}) I_{\{\epsilon_1 \leq y_i \leq \epsilon_2\}}) < \infty$.

APPENDIX B

Lemma 1. *(Leave-one-out lemma.)* $h_\lambda(i, \exp(f_{\lambda i}^{(-i)}), x) = f_\lambda^{(-i)}(x)$.

The proof is omitted since it is similar to that in Xiang and Wahba (1996). Lemma 1 states that the estimated function based on $\mathbf{y}^{-i} = (y_1, \ldots, y_{i-1}, \exp(f_{\lambda i}^{(-i)}), y_{i+1}, \ldots, y_n)^T$ is the same as $f_\lambda^{(-i)}$. It is easy to check that $\partial^2 J/\partial \mathbf{f} \partial \mathbf{f}^T = \mathbf{W}_\lambda + n\lambda \boldsymbol{\Omega}$ and $\partial^2 J/\partial \mathbf{y} \partial \mathbf{f}^T = -\mathbf{V}$. Using Taylor expansions as in Xiang and Wahba (1996), we have $(f_{\lambda i} - f_{\lambda i}^{(-i)})/(y_i - \exp(f_{\lambda i}^{(-i)})) \approx d_{ii}$. Thus by (3.7),

$$
\begin{aligned}
\mathrm{CV}(\lambda) &= L(\lambda) + \sum_{i=1}^n y_i \left(\exp(-f_{\lambda i}^{(-i)}) - \exp(-f_{\lambda i})\right) \\
&\approx L(\lambda) + \sum_{i=1}^n y_i \exp(-f_{\lambda i}) \left(f_{\lambda i} - f_{\lambda i}^{(-i)}\right) \\
&= L(\lambda) + \sum_{i=1}^n y_i \exp(-f_{\lambda i}) \frac{f_{\lambda i} - f_{\lambda i}^{(-i)}}{y_i - \exp(f_{\lambda i}^{(-i)})} \frac{y_i - \exp(f_{\lambda i})}{1 - \frac{\exp(f_{\lambda i}) - \exp(f_{\lambda i}^{(-i)})}{y_i - \exp(f_{\lambda i}^{(-i)})}} \\
&\approx L(\lambda) + \sum_{i=1}^n y_i \exp(-f_{\lambda i}) \frac{f_{\lambda i} - f_{\lambda i}^{(-i)}}{y_i - \exp(f_{\lambda i}^{(-i)})} \frac{y_i - \exp(f_{\lambda i})}{1 - \exp(f_{\lambda i}) \frac{f_{\lambda i} - f_{\lambda i}^{(-i)}}{y_i - \exp(f_{\lambda i}^{(-i)})}} \\
&= L(\lambda) + \sum_{i=1}^n \frac{y_i \exp(-f_{\lambda i}) (y_i - \exp(f_{\lambda i}))}{\frac{y_i - \exp(f_{\lambda i}^{(-i)})}{f_{\lambda i} - f_{\lambda i}^{(-i)}} - \exp(f_{\lambda i})} \\
&\approx L(\lambda) + \sum_{i=1}^n \frac{d_{ii} y_i \exp(-f_{\lambda i})(y_i - \exp(f_{\lambda i}))}{1 - d_{ii} \exp(f_{\lambda i})}.
\end{aligned}
$$

ACKNOWLEDGMENTS

This research was supported by NIH Grant R01 GM58533. The authors thank the editor, the associate editor and two referees for their constructive comments and suggestions that have led to a substantial improvement in the article.

[Received November 2005. Revised September 2006.]

REFERENCES

Albertson, D. G., and Pinkel, D. (2003), "Genomic Microarray in Human Genetic Disease and Cancer," *Human Molecular Genetics*, 12, R145–R152.

Andersen, T. G., and Lund, J. (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–377.

Berger, J. O. (1980), "Improving on Inadmissible Estimators in Continuous Exponential Families with Applications to Simultaneous Estimation of Gamma Scale Parameters," *The Annals of Statistics*, 8, 545–571.

Box, G. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations," *Technometrics*, 30, 1–17.

Carroll, R. J. (1982), "Adapting for Heteroscedasticity in Linear Models," *The Annals of Statistics*, 10, 1224–1233.

——— (1987), "The Effect of Variance Function Estimation on Prediction Intervals," in *Proceedings of the 4th Purdue Symposium on Statistical Decision Theory and Related Topics*, eds. J. O. Berger and S. S. Gupta, Vol II.

Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*, New York: Chapman & Hall.

Cui, X., Hwang, J. T. G., Qiu, J., Blades, N. J., and Churchill, G. A. (2005), "Improved Statistical Tests for Differential Gene Expression by Shrinking Variance Components Estimates," *Biostatistics*, 6, 59–75.

Dai, M., and Guo, W. (2005), "Heteroscedastic Smoothing Spline Models," unpublished manuscript.

Fan, J., and Yao, Q. (1998), "Efficient Estimation of Conditional Variance Functions in Stochastic Regression," *Biometrika*, 85, 645–660.

Gallant, A. R., and Tauchen, G. (1997), "Estimation of Continuous Time Models for Stock Returns and Interest Rates," *Macroeconomic Dynamics*, 1, 135–168.

Gu, C. (1992), "Cross-Validating Non Gaussian Data," *Journal of Computational and Graphical Statistics*, 2, 169–179.

——— (2002), *Smoothing Spline ANOVA Models*, New York: Springer-Verlag.

Gu, C., and Xiang, D. (2001), "Cross-Validating Non-Gaussian Data: Generalized Approximate Cross-Validation Revisited," *Journal of Computational and Graphical Statistics*, 10, 581–591.

Hall, P., and Carroll, R. J. (1989), "Variance Function Estimation in Regression: The Effect of Estimating the Mean," *Journal of the Royal Statistical Society*, Series B, 51, 3–14.

Huang, X., and Pan, W. (2002), "Comparing Three Methods for Variance Estimation with Duplicated High Density Oligonucleotide Arrays," *Functional & Integrative Genomics*, 2, 126–133.

Liu, A., Meiring, W., and Wang, Y. (2005), "Testing Generalized Linear Models Using Smoothing Spline Methods," *Statistica Sinica*, 15, 235–256.

Mantripragada, K. K., Buckley, P. G., de Stahl, T. D., and Dumanski, J. P. (2004), "Genomic Microarrays in the Spotlight," *Trends in Genetics*, 20, 87–94.

Moore, D. H., Pallavicini, M., Cher, M. L., and Gray, J. W. (1997), "A t-Statistic for Objective Interpretation of Comparative Genomic Hybridization (CGH) Profiles," *Cytometry*, 28, 183–190.

Müller, H. G., and Stadtmüller, U. (1987), "Estimation of Heteroscedasticity in Regression Analysis," *The Annals of Statistics*, 15, 610–635.

Raab, G. M. (1981), "Estimation of a Variance Function, With Application to Immunoassay," *Applied Statistics*, 30, 32–40.

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997), "Local Polynomial Variance-Function Estimation," *Technometrics*, 39, 262–273.

Silverman, B. W. (1985), "Some Aspects of the Spline Smoothing Approach to Nonparametric Regression Curve Fitting" (with discussion), *Journal of the Royal Statistical Society*, Series B, 47, 1–21.

Snijders, A. M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A. K., Huey, B., Kimura, K., Law, S., Myambo, K., Palmer, J., Ylstra, B., Yue, J. P., Gray, J. W., Jain, A. N., Pinkel, D., and Albertson, D. G. (2001), "Assembly of Microarrays for Genome-Wide Measurement of DNA Copy Number," *Nature Genetics*, 29, 263–264.

Tong, T., and Wang, Y. (2007), "Optimal Shrinkage Estimation of Variances with Applications to Microarray Data Analysis," *Journal of the American Statistical Association*, 102, 113–122.

Tusher, V., Tibshirani, R., and Chu, G. (2001), "Significance Analysis of Microarrays Applied to the Ionizing Radiation Response," in *Proceedings of the National Academy of Sciences U.S.A.*, 98, 5116–5121.

Wahba, G., Wang, Y., Gu, C., Klein, R., and Klein, B. (1995), "Smoothing Spline ANOVA for Exponential Families, With Application to the Wisconsin Epidemiological Study of Diabetic Retinopathy," *The Annals of Statistics*, 23, 1865–1895.

Wang, Y., and Guo, S. W. (2004), "Statistical Methods for Detecting Genomic Alterations Through Array-Based Comparative Genomic Hybridization (CGH)," *Frontiers in Bioscience*, 9, 540–549.

Wang, Y., Wahba, G., Chappell, R., and Gu, C. (1995), "Simulation Studies of Smoothing Parameter Estimates and Bayesian Confidence Intervals in Bernoulli SS ANOVA Models," *Communications in Statistics: Simulation and Computation*, 24, 1037–1059.

Watters, R. L., J., Carroll, R. J., and Spiegelman, C. H. (1987), "Error Modeling and Confidence Interval Estimation for Inductively Coupled Plasma Calibration Curves," *Analytical Chemistry*, 59, 1639–1643.

Wong, W. (2006), "Estimation of the Loss of an Estimate," *Frontiers in Statistics*, eds. J. Fan and H. L. Koul, World Scientific.

Xiang, D., and Wahba, G. (1996), "A Generalized Approximate Cross-Validation for Smoothing Splines with Non-Gaussian Data," *Statistica Sinica*, 6, 675–692.

Yao, Q., and Tong, H. (1994), "Quantifying the Influence of Initial Values on Nonlinear Prediction," *Journal of the Royal Statistical Society*, Series B, 56, 701–725.

Yuan, M. (2005), "Automatic Smoothing for Poisson Regression," *Communications in Statistics: Theory and Methods*, 34, 603–617.

Yuan, M., and Wahba, G. (2004), "Doubly Penalized Likelihood Estimator in Heteroscedastic Regression," *Statistics and Probability Letters*, 69, 11–20.