Logistic Regression With Brownian-Like Predictors

Martin A. LINDQUIST and Ian W. MCKEAGUE

This article introduces a new type of logistic regression model involving functional predictors of binary responses, and provides an extension of this approach to generalized linear models. The predictors are trajectories that have certain sample path properties in common with Brownian motion. Time points are treated as parameters of interest, and confidence intervals are developed under prospective and retrospective (case-control) sampling designs. In an application to functional magnetic resonance imaging data, signals from individual subjects are used to find the portion of the time course that is most predictive of the response. This allows the identification of sensitive time points specific to a brain region and associated with a certain task, which can be used to distinguish between responses. A second application concerns gene expression data in a case-control study involving breast cancer, where the aim is to identify genetic loci along a chromosome that best discriminate between cases and controls.

KEY WORDS: Brownian motion; Empirical process; Functional logistic regression; Functional magnetic resonance imaging; Gene expression; Lasso; M-estimation.

1. INTRODUCTION

This article investigates a logistic regression model involving a binary response, Y, and a predictor given by the value of the trajectory of a continuous stochastic process, $X = \{X(t), t \in [0, 1]\}$, at some unknown time point. Specifically, we consider the model

$$\text{logit}[P(Y = 1|X)] = \alpha + \beta X(\theta), \qquad (1)$$

and focus on the time point $\theta \in [0, 1]$ as the target parameter of interest. The intercept, α, and the slope, β, are scalars, and $\text{logit}(u) = \log(u/(1 - u))$. The trajectory of X is assumed to be observed over a regular grid of time points, with a sufficiently high resolution such that for statistical purposes, we can assume that it is observed continuously. We call this a *point-impact* model, because it involves only the value of X at θ, which represents a "sensitive" time point in terms of the relationship to the response. Generalized linear models (McCullagh and Nelder 1989) can be treated in a similar manner.

A motivation for using such a model arises from a functional magnetic resonance imaging (fMRI) experiment designed to explore differences between individuals based on anxiety levels (see Lindquist, Waugh, and Wager 2007; Lindquist 2008). Subjects in the experiment were classified as either resilient ($Y = 1$) or nonresilient ($Y = 0$) according to a written test. Each of the 25 subjects (13 resilient and 12 nonresilient) performed a 7-minute anxiety-provoking speech preparation task (see Figure 1), during which a series of 215 fMRI images were acquired. The design was an off-on-off design, with an anxiety-provoking period occurring between lower-anxiety resting periods. The fMRI signal, $X(t)$, from the ventromedial prefrontal cortex (a region known to be related to anxiety) is shown in Figure 2. It is of interest to furnish a time interval that most clearly distinguishes between resilient and nonresilient individuals. How can we find such a time interval? We propose the

model (1) as a natural way of approaching this problem, and in Section 2 we develop a confidence interval for the time parameter θ.

The key idea behind our approach is to exploit sample path properties of the trajectories, which from an inspection of Figure 2 appear to be locally similar to those of Brownian motion. Our results are developed for trajectories that are "Brownian-like" in the sense that $X(\theta_0 + t) - X(\theta_0)$ is a standard two-sided Brownian motion as a process in t over some neighborhood of 0, where θ_0 is the true value of θ.

Logistic regression plays an important role in case-control studies (Prentice and Pyke 1979), in which the sampling is retrospective. Our model involving Brownian-like trajectories is naturally relevant in that setting as well. A particular example arises from gene expression data, with the "time" variable corresponding to location along a chromosome. Figure 3 shows log gene expression levels from the breast tissue of 10 breast cancer patients (from a sample of 40 cases) and 6 normal subjects (controls), along a sequence of 776 loci from chromosome 1, and 518 loci from chromosome 17. The latter chromosome contains the best known breast cancer gene, the tumor suppressor BRCA1, but loci in this gene are not included; the complete data set was described by Richardson et al. (2006). Our approach can provide a framework for determining important genetic loci for discriminating between breast cancer patients and normal subjects.

A complementary approach to our proposed method is functional regression modeling, which has been extensively developed in the functional data analysis literature (see, e.g., James and Silverman 2005; Ramsay and Silverman 2006). Estimates of the regression function in such models may be difficult to interpret, however. Variable selection techniques for increasing interpretability by eliminating "unnatural wiggles" in the estimates have been recently introduced for functional linear models (James, Wang, and Zhu 2009). In contrast, our approach is based on finding interpretable *time points* that influence the response. In some applications there are scientific reasons to believe that there are only a small number of sensitive time points, which cannot be captured by the integral used in functional

Martin A. Lindquist is Associate Professor, Department of Statistics, Columbia University, New York, NY 10027 (E-mail: *martin@stat.columbia.edu*). Ian W. McKeague is Professor, Department of Biostatistics, Columbia University, New York, NY 10032 (E-mail: *im2131@columbia.edu*). Lindquist's work was supported by National Science Foundation (NSF) grant DMS-0804198 and McKeague's work was supported by NSF grant DMS-0806088 and National Cancer Institute (NCI) grant 1R01CA127532-01A1. The authors thank Bodhi Sen for discussions about the bootstrap, Tor Wager for the fMRI data, and Galina Glazko, Andrei Yakovlev, and Bill Stewart for help with the gene expression data.
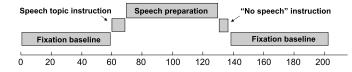
Figure 1. A schematic of the experimental task design for the fMRI study of Lindquist, Waugh, and Wager (2007). Subjects were informed that they were to be given 2 minutes to prepare a 7-minute speech, topic of which would be revealed to them during scanning. After the start of fMRI acquisition, there was 2 minutes of resting baseline. At the end of this period, subjects viewed an instruction slide for 15 seconds that described the speech topic. After 2 minutes of silent preparation, another instruction screen appeared for 15 seconds that informed subjects that they would not have to give the speech. An additional 2-minute period of resting baseline followed, completing the functional run. Images were acquired every 2 seconds throughout the course of the run.

regression. An example of such point-impact causality arises with fMRI data, in which shifts in the onset time of brain activation have been observed across different age cohorts (see D'Esposito, Deouell, and Gazzaley 2003). In such a situation, functional regression will be misleading, whereas our approach will specifically detect such shifts. Our simulation studies and real data examples given in Sections 3 and 4 confirm this. For both the fMRI and gene expression examples, our model gives results that are both sensible and interpretable in the context of application, whereas the functional estimates are difficult to interpret. There is a clear distinction between the roles of the two approaches; if the influence of the trajectories is spread over the time course, or if the aim is prediction (or classification), then functional logistic regression is suitable, but if the influence is concentrated at sensitive time points and interpretation is the overriding concern, then our approach is more suitable.

Another important area of application arises in genome-wide studies involving the expression of multiple genes, when more than one location is expected to influence the response. In such a case it is of interest to expand the point-impact model (1) to allow multiple sensitive time points, as in

$$\text{logit}[P(Y=1|X)] = \alpha + \sum_{j=1}^{p} \beta_j X(\theta_j), \qquad (2)$$

where $0 < \theta_1 < \cdots < \theta_p < 1$ and p is a (known) upper bound on the number of locations. When the β_j's correspond to values of a continuous function restricted to a fine grid, this approximates the functional logistic regression model discussed earlier. When the number of nonzero components, β_j (i.e., the number of point-impacts) is known to be small but p is large, a lasso-type penalty can be used to regularize the problem and provide a sparse collection of the θ_j. The confidence interval developed in Section 2 naturally extends to this setting, but for ease of presentation here we restrict attention to a single sensitive time point.

2. ESTIMATION OF SENSITIVE TIME POINTS

In this section we introduce estimators for sensitive time points and derive the asymptotic distribution for three separate cases. We begin with logistic regression for both prospective and retrospective sampling, and then extend the theory to generalized linear models. Finally, we develop confidence intervals.

2.1 Prospective Sampling

In this case the data consist of a random sample of n observations from the joint distribution of X and Y, and the maximum likelihood estimator of the parameters in (1) is given by

$$(\hat{\theta}_n, \hat{\alpha}_n, \hat{\beta}_n) = \arg\max_{\theta,\alpha,\beta} \mathbb{M}_n(\theta, \alpha, \beta), \qquad (3)$$

where the log-likelihood function is $\mathbb{M}_n(\theta, \alpha, \beta) = \mathbb{P}_n[m_{\theta,\alpha,\beta}]$,

$$m_{\theta,\alpha,\beta}(X, Y) = Y[\alpha + \beta X(\theta)] - \log[1 + \exp(\alpha + \beta X(\theta))], \quad (4)$$

and $\mathbb{P}_n$ is the empirical distribution of the data on (X, Y).

The large-sample distribution of $\hat{\theta}_n$ is given by the following result, in which θ_0 denotes the true value of θ.

Theorem 2.1. If $X(\theta_0 + t) - X(\theta_0)$ is a standard two-sided Brownian motion (as a process in t for $0 \le \theta_0 + t \le 1$) that is independent of $X(\theta_0)$, $0 < \theta_0 < 1$, and $\beta \neq 0$, then

$$n(\hat{\theta}_n - \theta_0) \to_d \lambda^{-1} \arg\max_{t \in \mathbb{R}}(B(t) - |t|/2),$$

where B is a standard two-sided Brownian motion and $\lambda = \beta^2 E[\text{Var}(Y|X)]$.
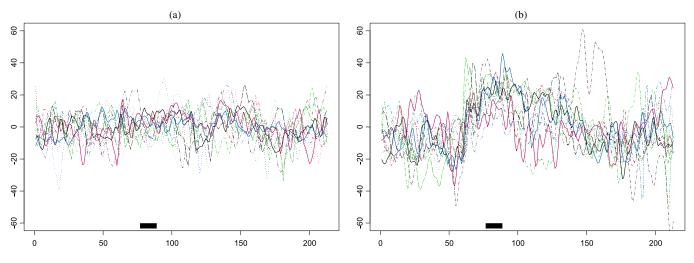

Figure 2. The fMRI signal over the ventromedial prefrontal cortex in reaction to an anxiety-provoking task for resilient (a) and nonresilient (b) subjects. The black line at the bottom of each plot indicates a 95% confidence interval for θ.
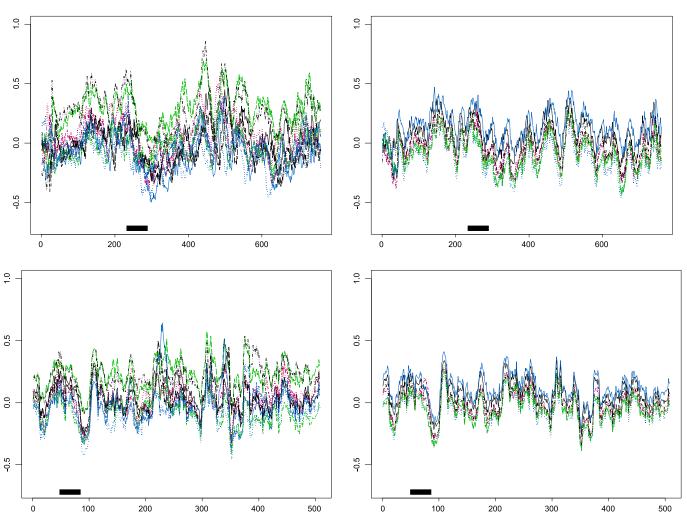
Figure 3. Log gene expression levels for 10 breast cancer cases (left column) and 6 normal controls (right column) at 776 loci along chromosome 1 (top row), and 518 loci along chromosome 17 (bottom row). The black line at the bottom of each plot indicates a 95% confidence interval for θ.

The main assumption of this theorem, that the increment of X about θ_0 is a two-sided Brownian motion, independent of $X(\theta_0)$, can be relaxed to the extent that it is needed only locally, in a neighborhood of θ_0. A standard Brownian motion X approximately satisfies this property in a *small* neighborhood of θ_0, because the behavior of the increment of X around θ_0 is affected only slightly by the constraint $X(0) = 0$ when θ_0 is sufficiently far from 0.

Another way in which the conditions of the theorem can be relaxed is that the infinitesimal variance of the two-sided Brownian motion does not need to be 1 (as with standard Brownian motion), but can take an arbitrary value $v > 0$. The estimated quadratic variation, $\hat{v}_i$, of the ith trajectory $X_i(t)$ should be used to normalize the sample paths before analysis by replacing $X_i(t)$ by $X_i(t)/\sqrt{\hat{v}_i}$. In some cases, it also may be suitable to calibrate the mean of each trajectory, as discussed in connection with the analysis of the fMRI data in Section 4.

The rate of convergence of $\hat{\theta}_n$ is controlled by the Hurst exponent of the trajectories, H, which for Brownian motion is $H = 0.5$. The Hurst exponent could be estimated (Beran 1994; Embrechts and Maejima 2002), and if it were found to deviate significantly from 0.5, either moving averages or differences could be applied before fitting the model (to bring the trajectories into accordance with the assumption involving Brownian increments). If such manipulation of the data is considered unappealing, then an alternative could be to extend our approach to the case where the increments of X are locally two-sided *fractional* Brownian motion with $0 < H \le 1$. The convergence rate would then become $n^{1/(2H)}$, and, given sufficient resolution in the data, H could be estimated locally (in the neighborhood of $\hat{\theta}_n$), leading to the construction of confidence intervals for θ_0. This extension would greatly relax the relatively restrictive assumption of Theorem 2.1, but at the cost of a more complex limiting distribution. Another alternative could be to use a model-based bootstrap as described in Section 3, which does not require the assumption of Brownian behavior and could be applied to the original trajectories without presmoothing. This approach has been developed by McKeague and Sen (2009) for point impact linear regression.

A referee raised the question of how to test the adequacy of the Brownian motion assumption. A simple procedure would be to consider increments of X over a succession of small time intervals and test whether they are uncorrelated. The multiple testing problem caused by the large number of increments can be

handled by adapting a bootstrap approach developed for high-throughput gene expression assays in which it is of interest to find sets of genes that have correlated expression profiles (see Dudoit and van der Laan 2008, p. 360).

2.2 Retrospective Sampling

In case-control studies, the predictors are sampled retrospectively for a sample of cases and a sample of controls; that is, we have a sample from the conditional distribution of X given $Y = 1$ and an independent sample from the conditional distribution of X given $Y = 0$. This gives a combined sample of size $n = n_0 + n_1$, where n_1, n_0 are the sizes of the two samples.

Under the logistic regression model, the density of $X(\theta_0)$ for cases can be expressed using Bayes formula in the form $\exp(\bar{\alpha} + \beta x)h(x)$, where $h(x)$ is the density of $X(\theta_0)$ for controls (Prentice and Pyke 1979). Here $\bar{\alpha} = \alpha + \log\{(1 - \pi)/\pi\}$, where $\pi = P(Y = 1)$ is the prevalence of cases in the population. Adapting the approach of Qin and Zhang (1997) to the present setting then leads to estimates [as in (3)] based on the following semiparametric profile log-likelihood function:

$$\mathbb{M}_n(\theta, \bar{\alpha}, \beta) = \rho \mathbb{P}_n^1[\bar{\alpha} + \beta X(\theta)]$$
$$- (\mathbb{P}_n^0 + \rho \mathbb{P}_n^1) \log[1 + \rho \exp(\bar{\alpha} + \beta X(\theta))],$$

where $\mathbb{P}_n^0$ and $\mathbb{P}_n^1$ are the empirical distributions of the control and case samples, respectively, and $\rho = n_1/n_0$ is assumed to remain fixed as $n \to \infty$. The estimates of $(\bar{\alpha}, \beta)$ for fixed θ based on this log-likelihood are identical to those of Prentice and Pyke (1979). The following result gives the large-sample behavior of $\hat{\theta}_n$.

Theorem 2.2. If the assumptions of Theorem 2.1 hold for both cases and controls, then

$$n(\hat{\theta}_n - \theta_0) \to_d \bar{\lambda}^{-1} \arg\max_{t \in \mathbb{R}}(B(t) - |t|/2),$$

where B is a standard two-sided Brownian motion and $\bar{\lambda}$ is as defined in the proof.

In contrast to the well-known result of Prentice and Pyke (1979) showing that the limit distribution of the estimator of $(\bar{\alpha}, \beta)$ is the same as if the data had been obtained via prospective sampling, the foregoing result shows that $\hat{\theta}_n$ has a different limit distribution. Although it is of the same form as in the prospective case, the nuisance parameter is different ($\bar{\lambda} \neq \lambda$). Under both prospective and retrospective sampling, $\hat{\alpha}_n$ and $\hat{\beta}_n$ converge at $\sqrt{n}$-rate, are asymptotically normal (with the same limit as though θ_0 were known), and are asymptotically independent of $\hat{\theta}_n$.

2.3 Generalized Linear Models

In this section we show how the approach of Section 2.1 can be extended to generalized linear models (McCullagh and Nelder 1989). We now model the conditional density of a scalar response Y given X by a canonical exponential family

$$p(y|X) = \exp([X(\theta)y - b(X(\theta))]/a(\phi) + r(y, \phi))$$

for some known functions $a(\cdot)$, $b(\cdot)$, and $r(\cdot, \cdot)$. Here ϕ is a dispersion parameter and $p(\cdot|X)$ is a density with respect to some given Borel measure. The cumulant function, b, is assumed to

be twice continuously differentiable, and b' is assumed to be strictly increasing. In linear regression, ϕ is the variance of the random error, whereas in logistic and Poisson regression, there is no dispersion parameter. Previously we used the more general expression $\alpha + \beta X(\theta)$ in place of $X(\theta)$, but because α, β, and ϕ can be estimated separately after estimation of θ, to keep the notation simple, here we treat θ as the only unknown parameter.

The log-likelihood, $\mathbb{M}_n(\theta) = \mathbb{P}_n[m_\theta]$, is now based on $m_\theta(X, Y) = YX(\theta) - b(X(\theta))$. As outlined in the Appendix, the limiting behavior of the corresponding maximum likelihood estimator $\hat{\theta}_n$ is the same as that in Theorem 2.1, given the same assumptions on X, except that the nuisance parameter λ is given by the ratio of the expected curvature of the cumulant function at $X(\theta_0)$ and $a(\phi)$,

$$\lambda = Eb''(X(\theta_0))/a(\phi) = E[\text{Var}(Y|X)]/a(\phi)^2.$$

2.4 Confidence Intervals

Based on the foregoing results, a Wald-type confidence interval for θ_0 having $100(1 - \gamma)\%$ nominal coverage is given by

$$\hat{\theta}_n \pm (\hat{\lambda}n)^{-1}Z_{\gamma/2}, \tag{5}$$

where Z_γ is the upper γ-quantile of $\arg\max_{t \in \mathbb{R}}(B(t) - |t|/2)$. Here $\hat{\lambda}$ is a consistent estimate of λ for prospective sampling, or of $\bar{\lambda}$ for logistic regression with retrospective sampling. Such an estimator $\hat{\lambda}$ is obtained by putting empirical distributions in place of expectations and plugging in estimates of the relevant parameters α, $\bar{\alpha}$, β, and θ into λ or $\bar{\lambda}$.

A result of Bhattacharya and Brockwell (1976) shows that the distribution function F of $\arg\max_{t \in \mathbb{R}}(B(t) - |t|/2)$ can be expressed in terms of the standard normal distribution function Φ as

$$F(x) = 1/2 + \sqrt{x}e^{-x/8}/\sqrt{2\pi} + 3e^x\Phi(-3\sqrt{x}/2)/2$$
$$- (x + 5)\Phi(-\sqrt{x}/2)/2$$

for $x \geq 0$. This distribution frequently arises in changepoint problems under "contiguous asymptotics" (Yao 1987; Stryhn 1996; Müller and Song 1997). The foregoing expression allows for the efficient computation of the upper quantiles of F and gives $Z_{0.05} = 7.687$, $Z_{0.025} = 11.033$, and $Z_{0.005} = 19.767$.

For logistic regression with prospective sampling, we can write $\lambda = \beta^2 E[A/(A + 1)^2]$, where $A = \exp[-(\alpha + \beta X(\theta_0))]$. When α and β are relatively small, λ is approximately $\beta^2/4$. Then, using the expression for the variance of F given by Stryhn (1996), the standard error of $\hat{\theta}_n$ is roughly $5/(n\beta^2)$. Figure 4 shows plots (obtained via Monte Carlo) describing the behavior of λ in the special case where $X(\theta_0) \sim N(0, \sigma^2)$, for varying values of α, β, and σ^2.

The plots indicate that the parameter β has the greatest impact on the value of λ. For fixed values of α and σ^2, λ increases with the absolute value of β. For large values of β, the increase is roughly linear. In the neighborhood of 0, λ is approximately equal to $\beta^2/4$. Thus for small values of β, the value of λ approaches 0, leading to a substantial widening of the confidence interval for θ_0. This is natural, because a value of β close to 0 implies that none of the time points has a major influence on the response; the widened confidence interval reflects this fact. Similar comments can be made for the case of retrospective sampling.
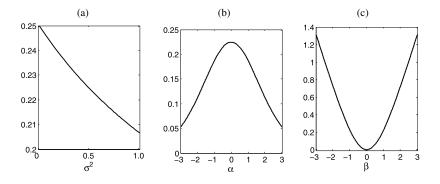
Figure 4. Plots of the value of the nuisance parameter λ as a function of the variance of $X(\theta_0)$ with $\alpha = 0$ and $\beta = 1$ (a), as a function of α with $\sigma^2 = 0.5$ and $\beta = 1$ (b), and as a function of β with $\alpha = 0$ and $\sigma^2 = 0.5$ (c).

3. SIMULATION STUDIES

In this section we report the results of five simulation studies that use standard Brownian motion to define the functional predictor. We restrict attention to prospective sampling, but the results are similar for retrospective sampling. The first simulation illustrates the behavior of the estimators of α, β, and θ_0 in repeated application of the method. The second simulation studies the coverage probabilities of the proposed confidence interval for θ_0 and compares it with model-based bootstrap confidence intervals. The third and fourth simulations are designed to explore the relationships among the point-impact (PI) model (1), the lasso, and the commonly used functional logistic regression model

$$\text{logit}[P(Y = 1 | X)] = \alpha + \int_0^1 X(t)\beta(t)\, dt, \tag{6}$$

where the regression function $\beta(t)$ is treated nonparametrically; in the sequel we refer to (6) as the *functional-impact* (FI) model. The final simulation example studies the coverage probabilities of the confidence interval for β in the PI model.

To fit the FI model, we use the S-PLUS 7.0 function fGLM in the functionalData library, with a B-spline basis of order 4 (piecewise cubic). The uniform grid of observation times provides the knots, and the roughness penalty for $\beta(t)$ is taken as the L_2-norm of its second derivative, with the smoothing parameter selected by leave-one-out cross-validation; no smoothing is used in the initial step of representing the trajectories

in terms of the B-spline basis. For the lasso, we use the fast and efficient coordinate descent algorithm implemented in the R package glmnet (Friedman, Hastie, and Tibshirani 2008) to calculate the lasso path diagram, in which the estimates of β_j in (2) are plotted against the magnitude of the constraint on their ℓ_1-norm.

Simulation I. The data are generated from the PI model with $\alpha = 0$, $\beta = 3$, $\theta_0 = 0.5$, and $n = 40$. We restrict θ to a uniform grid of 101 points in the interval $[0, 1]$, and the Brownian predictors were generated over this grid using the R function fbmSim from the fSeries package. The deviance $(-2 \log\text{-likelihood})$ is calculated along the grid, with α and β successively replaced by their estimates corresponding to each value of θ; the grid point minimizing the deviance is then taken as the estimate $\hat{\theta}_n$. The results, displayed in Figure 5, highlight the faster rate of convergence for $\hat{\theta}_n$ compared with $\hat{\alpha}_n$ and $\hat{\beta}_n$.

Simulation II. We next repeated Simulation I using a variety of choices for α, β, and n, keeping the value of θ_0 fixed at 0.5. For each combination, we calculated $100(1 - \gamma)\%$ confidence intervals according to (5) and determined the coverage probability based on 10,000 replications; the results are given in Table 1. At small sample sizes (e.g., $n = 40$), the coverage probabilities are somewhat less than their nominal values. Accuracy naturally improves with larger sample sizes and as β increases. Table 2 gives corresponding results for the model-based bootstrap in which the fitted PI model is used to create bootstrap samples of the response; the coverage probabilities now fall on the conservative side, but have a similar pattern of accuracy.
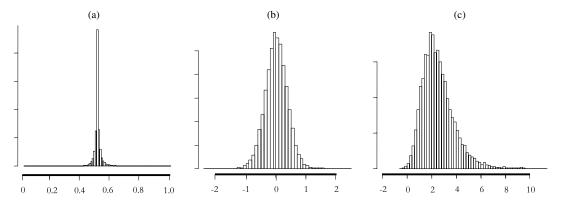

Figure 5. Simulation I: histograms of estimates of $\hat{\theta}_n$ (a), $\hat{\alpha}_n$ (b), and $\hat{\beta}_n$ (c) from 10,000 replications.

Table 1. Simulation II: Coverage probabilities of the proposed confidence intervals for θ_0 having nominal coverage of 0.90, 0.95, and 0.99

		$\gamma = 0.90$		$\gamma = 0.95$		$\gamma = 0.99$	
n	α	$\beta = 3$	$\beta = 6$	$\beta = 3$	$\beta = 6$	$\beta = 3$	$\beta = 6$
40	0	0.868	0.847	0.920	0.911	0.980	0.969
	1	0.858	0.850	0.912	0.914	0.980	0.968
80	0	0.879	0.902	0.928	0.935	0.979	0.982
	1	0.884	0.897	0.928	0.954	0.980	0.984

Table 2. Simulation II: Coverage probabilities of (percentile) bootstrap confidence intervals for θ_0 having nominal coverage of 0.90, 0.95, and 0.99, based on 1000 replications and 1000 bootstrap samples

		$\gamma = 0.90$		$\gamma = 0.95$		$\gamma = 0.99$	
n	α	$\beta = 3$	$\beta = 6$	$\beta = 3$	$\beta = 6$	$\beta = 3$	$\beta = 6$
40	0	0.977	0.923	0.988	0.968	1.000	0.999
	1	0.971	0.938	0.994	0.974	0.997	0.993
80	0	0.949	0.942	0.982	0.971	0.998	0.993
	1	0.956	0.912	0.984	0.952	1.000	0.996

Simulation III. A single sample was generated from the PI model in Simulation I with $n = 40$. Figure 6 shows the results of fitting the PI and FI models along with the lasso path diagram. The 95% confidence interval for θ_0 is 0.5 ± 0.071, which is very accurate, as expected. The deviance plotted at each possible value of θ on the grid of time points has a remarkably sharp global minimum at θ_0. The estimate of $\beta(t)$ achieves its maximum at θ_0, but gives the misleading impression that the effect of the predictor is spread out over much of the time course, rather than being concentrated at θ_0. This is perhaps not surprising, because cross-validation is a prediction error rate criterion, and so the smoothing causes the estimate to use as much of the information along the time course as possible. The lasso performs well, immediately picking out θ_0, as indicated by the arrow in the path diagram [Figure 6(c)].

Simulation IV. Consider the FI model for the spike-shaped regression functions displayed in the first column of Figure 7. In each case the estimate $\hat{\theta}_n$ ($n = 40$) coincides with one of the initial selections of the lasso, and both are either identical or close to the point $t = 0.5$ at which $\beta(t)$ achieves its maximum. The 95% confidence intervals based on $\hat{\theta}_n$ are 0.5 ± 0.034 for the narrower spikes and on 0.54 ± 0.042 for the wider spikes. The estimates based on the FI model, even though it is correctly specified, wrongly suggest that the influence of the predictor is significant over the whole time course. The estimates of $\beta(t)$ have maxima located close to $t = 0.5$, the location of the spikes, but have no other features in common with $\beta(t)$.

Attempts at using a higher-order derivative penalty for estimating $\beta(t)$ produced similar results to those shown in Figure 7. Features of $\beta(t)$ conceivably might be captured more accurately

using wavelet bases and thresholding, but we have restricted attention to the most commonly used approach to functional regression.

Simulation V. Data were generated in the same way as in Simulation II, except with $X(t) = B(t + \theta_0) - B(\theta_0)$, where B is two-sided Brownian motion; $\theta_0 = 0.5$ and $X(\theta_0) \sim N(0, 0.5)$. Confidence intervals for β are based on the $\sqrt{n}$-rate asymptotic normality of $\hat{\beta}_n$. The results, reported in Table 3, show that these confidence intervals have accurate coverage except when $\beta = 0$, in which case there is severe undercoverage. The case $\beta = 0$, while important for testing whether there is *any* effect of X, is outside the scope of our results, however. An implicit simultaneous inference problem caused by minimizing the deviance over θ appears when $\beta = 0$, but not otherwise. The reason why simultaneous inference is not an issue when $\beta \neq 0$ is that $\hat{\theta}_n$ converges at a much faster rate than $\hat{\beta}_n$, and thus the inference for β is concentrated in a very small neighborhood of θ_0, and hypothesis testing is not needed over the whole range of θ.

4. APPLICATIONS

In this section we illustrate our approach by applying it to two real data sets. The first data set comes from the fMRI study described in Section 1. Because only the relative change in signal is important, the individual mean over the first resting period was removed from the entire time course for calibration purposes. In addition, each trajectory was normalized by the square root of its estimated quadratic variation. Finally, each trajectory was smoothed with a moving average window
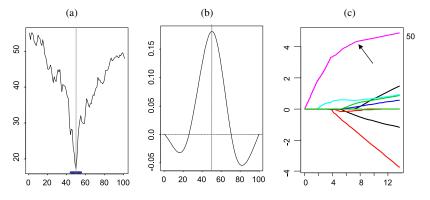

Figure 6. Simulation III: Deviance calculated as a function of θ. (a) The 95% confidence interval for θ_0 is depicted by the solid line at the bottom of the plot. (b) The estimated $\beta(t)$ using the FI model with cross-validated roughness penalty. In each panel the vertical line indicates the location of θ_0. (c) In the lasso path diagram, the arrow indicates the path corresponding to the grid point indicated to its right.
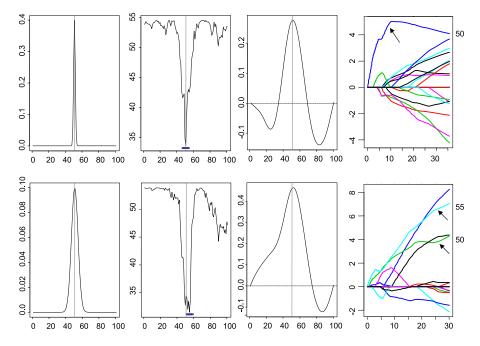
Figure 7. Simulation IV: The regression function $\beta(t)$ is taken as two separate Gaussian probability distribution functions centered at $t = 0.5$ in the first column. The deviance as a function of θ and the 95% confidence interval based on $\hat{\theta}_n$ is shown in the second column. The estimated $\beta(t)$ with cross-validated roughness penalty is in the third column; the vertical line indicates the time point at which $\beta(t)$ achieves its maximum. The lasso path diagram (fourth column) is labeled as before.

width of 3 time units. This width was chosen to given an estimated Hurst exponent of approximately 0.5 (corresponding to Brownian motion). The resulting trajectories are displayed in Figure 2. The trajectories of the 13 resilient subjects remain stable over the whole time course, but the nonresilient trajectories show a clear increase around the time of the anxiety-provoking task.

Figure 8 shows the results. The sensitive time point obtained using the proposed model corresponds to the 84th time point, which is 28 seconds into the anxiety-provoking period of the task. Inspecting the trajectories for subjects in the nonresilient group shown in Figure 2, it appears that this time point coincides with the peak activity in the ventromedial prefrontal cortex. The 95% confidence interval for θ_0 is 84 ± 5.4, as superimposed onto the bottom portion of Figure 8(a). The 95% confidence interval for the regression parameter β is -14.9 ± 13.5.

The FI model-based estimate of $\beta(t)$ has a local extremum just after the start of the anxiety-provoking period, but the influence of the predictor appears to be spread out over most of the time course, even though the anxiety-provoking period does not start immediately. The lasso first selects 87, then quickly

adds 84 (the PI selection), but is slow to add any further points (and those that are added are widely dispersed over the time course), suggesting that the PI model provides an adequate fit to the data.

We next consider the case-control study involving breast cancer patients, as described in the Introduction. For chromosome 1, before analysis, we took the natural logarithm of the gene expression level and smoothed each of the resulting trajectories with a moving average window of width 17. The top row of Figure 3 shows the trajectories of a subsample of the transformed data with breast cancer patients and normal subjects separated. Results of the analysis are shown in the top row of Figure 9. The 95% confidence interval for θ_0 is 260 ± 27.8, and that for β is 9.0 ± 8.0. The largest peak in the estimate of $\beta(t)$ again closely matches the estimate of θ_0. The lasso path diagram confirms the PI selection of 260 but suggests that several more loci may be involved as well.

For chromosome 17, the data were handled similarly, but with a window of width 11 used in the smoothing step. The results are shown in the bottom row of Figure 9. The 95% confidence interval for θ_0 is 76 ± 16.7, and that for β is 10.9 ± 9.6. The lasso path diagram, along with the presence of multiple

Table 3. Simulation V: Coverage probabilities of confidence intervals for β having nominal coverage of 0.90, 0.95, and 0.99

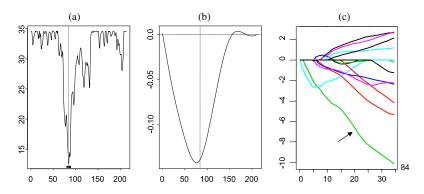|		$\gamma = 0.90$			$\gamma = 0.95$			$\gamma = 0.99$		
n	α	$\beta = 0$	$\beta = 3$	$\beta = 6$	$\beta = 0$	$\beta = 3$	$\beta = 6$	$\beta = 0$	$\beta = 3$	$\beta = 6$
40	0	0.468	0.860	0.931	0.701	0.920	0.954	0.954	0.958	0.982
	1	0.487	0.865	0.943	0.720	0.928	0.967	0.964	0.960	0.980
80	0	0.477	0.879	0.898	0.695	0.935	0.953	0.935	0.974	0.982
	1	0.474	0.868	0.900	0.684	0.925	0.951	0.928	0.961	0.987

Figure 8. Results for the fMRI data. (a) The deviance as a function of θ; the 95% confidence interval for θ_0 is depicted by a solid line along the bottom of the plot. (b) The estimate of the regression function $\beta(t)$ in the functional logistic regression model with cross-validated roughness penalty; the vertical line indicates the location of $\hat{\theta}_n$. (c) The lasso path diagram is labeled as before.

peaks in the estimate of $\beta(t)$, now suggest that numerous loci (beyond the PI selection) are involved.

5. DISCUSSION

In this article we have developed a point-impact (PI) logistic regression model for use with "Brownian-like" predictors. We expect this model to be useful when there are one or more sensitive time points at which the trajectory has a strong effect on the response. We have derived the rate of convergence, as well as the explicit limiting distribution of estimators of such time parameters in prospective and retrospective (case-control) settings, and used these results to construct Wald-type confidence intervals.

Our approach is complementary to standard functional logistic regression, which, although well adapted to classification

(prediction) problems, tends to oversmooth the estimate of the regression function when there are localized effects, due to the roughness penalty and the cross-validated choice of smoothing parameter. In contrast, our approach allows the estimation of point impact effects that would not be seen otherwise. It also enhances interpretation of the lasso path diagram by providing confidence intervals around sensitive time points selected by the lasso. In contrast to the lasso, however, our approach is not designed to search for a sparse collection of sensitive time points, because it only applies when X is known to have *some* effect on the response, that is, $\beta \neq 0$. In our approach, the implicit multiple testing problem concerning β is avoided, because of the fast rate of convergence of $\hat{\theta}_n$.

To increase the flexibility of our approach, it would be of interest to preclude the need to presmooth the trajectories by
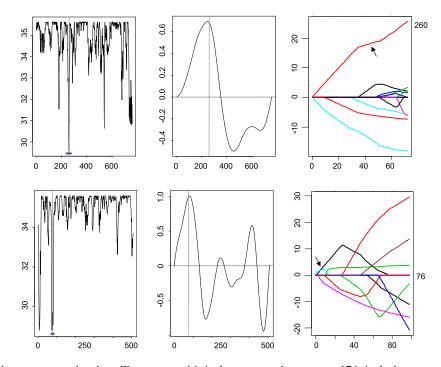

Figure 9. Results for the gene expression data. Chromosome 1 is in the top row; chromosome 17 is in the bottom row. The deviance calculated as a function of θ (first column); 95% confidence intervals for θ_0 are depicted by solid horizontal lines along the bottom of each plot. The estimate of $\beta(t)$ in functional logistic regression with cross-validated roughness penalty (second column); the vertical lines indicate the location of $\hat{\theta}_n$. The lasso path diagrams (third column) are labeled as before.

extending our results to fractional Brownian motion locally in the neighborhood of θ_0, as discussed in Section 2.1. Instead of Wald-type confidence intervals, however, in this case it would be preferable to pursue a model-based bootstrap approach, because the rate of convergence of $\hat{\theta}_n$ depends on the Hurst exponent, which is unlikely to be known in practice; see McKeague and Sen (2009).

Going beyond the PI model, it also would be interesting to allow for the estimation of sensitive *domains*, rather than sensitive time points, in the time course. In this situation, we would use

$$\text{logit}[P(Y = 1|X)] = \alpha + \sum_{j=1}^{p} \int_{l_j}^{r_j} \beta_j(t) \, dX(t),$$

where $\beta_j(t)$ is a nonparametric regression function with support on $[l_j, r_j]$, where $l_j < r_j$ are parameters. If the time intervals $[l_j, r_j]$ are small or the $\beta_j(t)$ are relatively constant, then this model essentially reduces to (2). Otherwise, this model is not covered by our approach and would require a separate development.

APPENDIX: PROOFS

The proofs presented herein are based on the theory of M-estimation (see van der Vaart and Wellner 1996, chap. 3.2) and involve (a) establishing the rate of convergence, (b) establishing the weak convergence of a suitably localized version of the empirical criterion function $\mathbb{M}_n$, and (c) applying the arg max continuous mapping theorem.

It can be shown that $\hat{\theta}_n$ is asymptotically independent of $\hat{\alpha}_n$ and $\hat{\beta}_n$, which converge at $\sqrt{n}$-rate, and that its limiting distribution is the same as if α and β were known. The proof of this involves mixed rates asymptotics (cf. Radchenko 2008). Similar results arise in changepoint problems (see, e.g., Koul, Qian, and Surgailis 2003). From now on, we fix α and β, and treat the log-likelihood function as a function solely of θ. We start with the proof of Theorem 2.1, and then explain what modifications are needed in the other two settings.

Rate of Convergence

The first step is to identify a nonnegative function, $d(\cdot, \theta_0)$, on the parameter space so that the criterion function $\mathbb{M}(\theta) = E[m_\theta]$ satisfies

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0) \tag{A.1}$$

for all $\theta \in [0, 1]$, where $\lesssim$ means "is bounded above up to a universal constant."

Recall that in maximum likelihood estimation, the expected log-likelihood, $\mathbb{M}$, is usually twice-differentiable, $\mathbb{M}'(\theta_0) = 0$, and the Fisher information $\mathbb{M}''(\theta_0) > 0$. Thus a Taylor expansion shows that $\mathbb{M}$ is approximately parabolic in the neighborhood of θ_0, and that the best choice for d is the usual Euclidean distance. In the present setting, however, the Brownian-like trajectories X are not smooth enough to ensure that $\mathbb{M}$ is differentiable.

Using the model (1) to find the expectation of the first term in m_θ, we have

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = E\left(\frac{\beta[X(\theta) - X(\theta_0)]e^{\alpha + \beta X(\theta_0)}}{1 + e^{\alpha + \beta X(\theta_0)}} \right)$$
$$- E \log\left(\frac{1 + e^{\alpha + \beta X(\theta)}}{1 + e^{\alpha + \beta X(\theta_0)}} \right).$$

The first term above vanishes by the assumption that the increments of X about θ_0 are independent of $X(\theta_0)$, leading to

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = -E \log\left(\frac{A + e^{\beta \sigma Z}}{A + 1} \right) \equiv -g(\sigma)$$

for $\sigma = \sqrt{|\theta - \theta_0|} \geq 0$, where $Z \sim N(0, 1)$ and $A = \exp[-(\alpha + \beta X(\theta_0))]$ are independent. Note that g is twice continuously differentiable with $g(0) = 0$,

$$g'(\sigma) = E\left(\frac{\beta Z e^{\beta \sigma Z}}{A + e^{\beta \sigma Z}} \right) \geq 0,$$
$$g''(\sigma) = E\left(\frac{A\beta^2 Z^2 e^{\beta \sigma Z}}{(A + e^{\beta \sigma Z})^2} \right) > 0 \tag{A.2}$$

for $\sigma \geq 0$. It follows that $g(\sigma) \gtrsim \sigma^2$ for $\sigma \in [0, 1]$, and (A.1) holds with the Hölder metric

$$d(\theta, \theta_0) = \sqrt{|\theta - \theta_0|}. \tag{A.3}$$

We next apply the following special case of a result of van der Vaart and Wellner (1996, thm. 3.2.5), giving a lower bound on the rate of convergence of the M-estimator $\hat{\theta}_n$ in terms of the continuity modulus $w_n(\delta) = \sup_{d(\theta, \theta_0) < \delta} |\mathbb{G}_n(m_\theta - m_{\theta_0})|$, where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ is the empirical process. In this result, outer expectation E^* and outer probability P^* are used to avoid measurability problems.

Proposition A.1. Suppose that (A.1) holds, and that $E^*[w_n(\delta)] \lesssim \delta^\alpha$ for every $\delta > 0$, where $0 < \alpha < 2$. Then $n^{1/(4-2\alpha)} d(\hat{\theta}_n, \theta_0) = O_p^*(1)$.

Note that $\alpha = 1$ gives the usual $n^{1/2}$-rate with respect to the metric d. The foregoing moment condition can be checked using an inequality from empirical process theory,

$$E^*[w_n(\delta)] \lesssim J_{[]}(1, \mathcal{M}_\delta, L^2(P))\{EM_\delta^2\}^{1/2}, \tag{A.4}$$

where $J_{[]}(1, \mathcal{M}_\delta, L^2(P))$ is the bracketing entropy integral of the class of functions $\mathcal{M}_\delta = \{m_\theta - m_{\theta_0} : d(\theta, \theta_0) < \delta\}$ and M_δ is an envelope function for $\mathcal{M}_\delta$ (cf. van der Vaart and Wellner 1996, p. 291).

The following lemma shows that m_θ is "Lipschitz in parameter" and consequently that $J_{[]}(1, \mathcal{M}_\delta, L^2(P)) < \infty$ for all $\delta > 0$ (see van der Vaart and Wellner 1996, p. 294).

Lemma A.1. Under the conditions of Theorem 2.1, if $0 < \alpha < 1/2$, then there is a random variable L with finite second moment such that

$$|m_{\theta_1} - m_{\theta_2}| \leq L|\theta_1 - \theta_2|^\alpha \tag{A.5}$$

for all $\theta_1, \theta_2 \in [0, 1]$ a.s.

Proof. Two-sided Brownian motion B has trajectories that are Lipschitz of any order $\alpha < 1/2$, in the sense that

$$|B(t) - B(s)| \leq K|t - s|^\alpha \qquad \forall t, s \in [-1, 1] \tag{A.6}$$

a.s., where K has moments of all orders. This is a consequence of the proof of Kolmogorov's continuity theorem, see theorem 2.2 of Revuz and Yor (2006). With m_θ given by (4), and writing $B(t) = X(\theta_0 + t) - X(\theta_0)$, which is a two-sided Brownian motion, by hypothesis,

$$|m_{\theta_1} - m_{\theta_2}| \leq 2|\beta||X(\theta_1) - X(\theta_2)|$$
$$= 2|\beta||B(\theta_1 - \theta_0) - B(\theta_2 - \theta_0)|$$
$$\leq 2K|\beta||\theta_1 - \theta_2|^\alpha, \tag{A.7}$$

where the first inequality uses the fact that the derivative of $x \mapsto \log(1 + e^x)$ is bounded between 0 and 1. We can then take $L = 2K|\beta|$.

From the foregoing lemma, Proposition A.1, and (A.4) we can see that the rate of convergence is controlled solely by the L^2-norm of the envelope function M_δ, which we now evaluate. We first bound the second moment of the continuity modulus $F_\delta = \sup_{|\theta - \theta_0| < \delta} |m_\theta - m_{\theta_0}|$. Using the first inequality in (A.7), we have

$$EF_\delta^2 \leq 4|\beta|E \sup_{|\theta - \theta_0| < \delta} |X(\theta) - X(\theta_0)|^2$$
$$= 4|\beta|E \sup_{|t| < \delta} |B(t)|^2 \lesssim \delta, \tag{A.8}$$

where the last step uses Doob's inequality. In view of (A.3), the envelope function is $M_\delta = F_{\delta^2}$, and we find $\{EM_\delta^2\}^{1/2} \lesssim \delta^2$, which translates to rate $r_n = n$ with respect to the usual Euclidean distance. Once we determine the rates of convergence, the next step is to identify the limit distribution in each case by localizing the criterion function.

Localizing the Criterion Function

Given the rate of convergence r_n, write $r_n(\hat{\theta}_n - \theta_0) = \hat{h}_n = \arg\max_{h \in \mathbb{R}} \widetilde{\mathbb{M}}_n(h)$, where

$$\widetilde{\mathbb{M}}_n(h) = s_n[\mathbb{M}_n(\theta_0 + h/r_n) - \mathbb{M}_n(\theta_0)], \qquad h \in \mathbb{R}. \quad (A.9)$$

We need to show that there exists an appropriate scaling s_n such that $\widetilde{\mathbb{M}}_n$ converges weakly to a nondegenerate limit process $\widetilde{\mathbb{M}}$ in the space $B_{\text{loc}}(\mathbb{R})$ of locally bounded functions on $\mathbb{R}$ equipped with the topology of uniform convergence on compacta. Then the argmax continuous mapping theorem, which is applicable because $\hat{h}_n = O_p^*(1)$, implies that $\hat{h}_n$ converges in distribution to the (unique) maximizer of $\widetilde{\mathbb{M}}$.

Setting $s_n = r_n = n$ and centering $\mathbb{M}_n$ by its mean gives

$$\widetilde{\mathbb{M}}_n(h) = n(\mathbb{P}_n - P)(m_{\theta_0 + h/n} - m_{\theta_0}) + nP(m_{\theta_0 + h/n} - m_{\theta_0})$$

$$= \beta\mathbb{G}_n[YZ_n(h)] - \sqrt{n}\mathbb{G}_n \log\left[\frac{A + e^{\beta Z_n(h)/\sqrt{n}}}{A + 1}\right]$$

$$- ng(\sqrt{|h|/n}), \quad (A.10)$$

where $Z_n(h) \equiv \sqrt{n}[X(\theta_0 + h/n) - X(\theta_0)]$. Using the hypothesis of the theorem, $Z_n(h) =_d \sqrt{n}B(h/n) =_d B(h)$, as processes, where the last step follows from the self-similarity property of two-sided Brownian motion, it follows that the second term in (A.10) (without the minus sign) can be written as

$$\sqrt{n}\mathbb{G}_n \log\left[1 + \frac{e^{\beta Z_n(h)/\sqrt{n}} - 1}{A + 1}\right] = \beta\mathbb{G}_n[Z_n(h)/(A+1)] + o_p(1),$$

where we have used $\log(1 + x) = x + O(x^2)$ and $e^x = 1 + x + O(x^2)$ as $x \to 0$. The difference between the first term in (A.10) and first term in the foregoing display is

$$\beta\mathbb{G}_n Z_n(h)[Y - 1/(A+1)] =_d \beta B(h)\left(\frac{1}{n}\sum_{i=1}^n [Y_i - 1/(A_i + 1)]^2\right)^{1/2}$$

$$\to_d \beta c B(h),$$

where $c^2 = E[\text{Var}(Y|X)]$ and we have used the fact that (A, Y) is independent of Z_n. The third term in (A.10) (without the minus sign) tends to $g''(0)|h|/2$. Noting that

$$E[\text{Var}(Y|X)] = E\left[\frac{1}{(A+1)}\left(1 - \frac{1}{(A+1)}\right)\right]$$

$$= E\left[\frac{A}{(A+1)^2}\right] = g''(0)/\beta^2,$$

with the last step following from (A.2), we conclude that $\widetilde{\mathbb{M}}_n$ converges weakly to $\widehat{\mathbb{M}}$ in the space $B_{\text{loc}}(\mathbb{R})$, where $\widehat{\mathbb{M}}(h) = \beta c B(h) - \beta^2 c^2|h|/2$. This completes the proof of Theorem 2.1.

Proof of Theorem 2.2

The rate of convergence is again n, which can be seen using essentially the same argument as before. Putting $s_n = r_n = n_1$ in the case-control version of the localized criterion function (A.9) gives, along

the lines of (A.10),

$$\widetilde{\mathbb{M}}_n(h) = \beta\rho\mathbb{G}_n^1[Z_{n_1}(h)] - \rho\sqrt{n_1}\mathbb{G}_n^1 \log\left[\frac{A + \rho e^{\beta Z_{n_1}(h)/\sqrt{n_1}}}{A + \rho}\right]$$

$$- \frac{n_1}{\sqrt{n_0}}\mathbb{G}_n^0 \log\left[\frac{A + \rho e^{\beta Z_{n_1}(h)/\sqrt{n_1}}}{A + \rho}\right]$$

$$- n_1\rho g_1(\sqrt{|h|/n_1}) - n_1 g_0(\sqrt{|h|/n_1}), \quad (A.11)$$

where $\mathbb{G}_n^j = \sqrt{n_j}(\mathbb{P}_n^j - P_j)$, $j = 0, 1$, are the empirical processes for the two samples (n_0 controls, n_1 cases), and

$$g_j(\sigma) = P_j \log\left(\frac{A + \rho e^{\beta\sigma Z}}{A + \rho}\right).$$

Here $Z \sim N(0, 1)$ and $A = \exp[-(\bar{\alpha} + \beta X(\theta_0))]$ are independent under P_j by the hypothesis of the theorem. Note the slightly different definition of A in the case-control setting. Using similar steps as in the previous proof, we see that the combined first three terms in (A.11) are asymptotically equivalent to

$$\beta\rho\mathbb{G}_n^1 Z_{n_1}(h)[1 - \rho/(A + \rho)] - \beta\sqrt{\rho}\mathbb{G}_n^0 Z_{n_1}(h)[\rho/(A + \rho)]$$

$$=_d \beta\rho B_1(h)\{\mathbb{P}_n^1[1 - \rho/(A + \rho)]^2\}^{1/2}$$

$$- \beta\sqrt{\rho}B_0(h)\{\mathbb{P}_n^0[\rho/(A + \rho)]^2\}^{1/2}$$

$$\to_d \beta\sqrt{\rho}c_1 B(h),$$

where B_0 and B_1 are independent two-sided Brownian motions and

$$c_1^2 = \rho P_1[1 - \rho/(A + \rho)]^2 + P_0[\rho/(A + \rho)]^2.$$

Note that

$$n_1 g_j(\sqrt{|h|/n_1}) \to g_j''(0)|h|/2 = \beta^2\rho P_j[A/(A + \rho)^2]|h|/2,$$

giving the limits of the last two terms in (A.11), so

$$\widetilde{\mathbb{M}}_n(h) \to_d \beta\sqrt{\rho}c_1 B(h) - \beta^2\rho c_2|h|/2,$$

where

$$c_2 = (P_0 + \rho P_1)[A/(A + \rho)^2].$$

We conclude that

$$n(\hat{\theta}_n - \theta_0) = (1 + 1/\rho)n_1(\hat{\theta}_n - \theta_0)$$

$$\to_d \bar{\lambda}^{-1} \arg\max_{t \in \mathbb{R}}(B(t) - |t|/2),$$

where $\bar{\lambda} = \beta^2\rho^2 c_2^2/[(1 + \rho)c_1^2]$. This completes the proof of Theorem 2.2.

Generalized Linear Models

To extend Theorem 2.1 to the GLM setting, we make use of two well-known formulas from the theory of canonical exponential families: $E(Y|X) = b'(X(\theta))$ and $\text{var}(Y|X) = a(\phi)b''(X(\theta))$. From the first of these formulas, the criterion function $\mathbb{M}(\theta) = E[m_\theta]$ satisfies

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = E\big[(X(\theta) - X(\theta_0))b'(X(\theta_0))\big]$$

$$- E\big[b(X(\theta)) - b(X(\theta_0))\big].$$

The first expectation above vanishes using the hypothesis about X in the statement of the theorem. The second expectation requires an extra argument beyond that needed for the proof of Theorem 2.1. From Itô's formula,

$$b(X(\theta)) - b(X(\theta_0)) = \int_{\theta_0}^\theta b'(X(u))\, dX(u) + \frac{1}{2}\int_{\theta_0}^\theta b''(X(u))\, du,$$

and because the Itô integral above has zero expectation (under mild conditions to ensure that it exists), we obtain

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = -\frac{1}{2}\int_0^{\sigma^2} Eb''(\sqrt{u}Z + X(\theta_0))\, du \equiv -g(\sigma),$$

where $\sigma = \sqrt{|\theta - \theta_0|} \geq 0$ and $Z \sim N(0, 1)$. Note that $g(0) = g'(0) = 0$ and $g''(0) = Eb''(X(\theta_0)) = c^2/a(\phi)$, where $c^2 = E[\text{Var}(Y|X)]$. The remaining steps to obtain the rate of convergence are similar to those in the logistic regression case, except that Lemma A.1 and (A.8) must be extended. This can be done under mild conditions using Itô's formula, applying theorem 2.1 of Revuz and Yor (2006), and bounding the higher-order moments of the Itô integral using the Burkholder–Davis–Gundy inequality.

For the last part of the proof, the localized criterion function (A.10) now decomposes as

$$
\begin{aligned}
\widetilde{\mathbb{M}}_n(h) &= \mathbb{G}_n[YZ_n(h)] - \sqrt{n}\,\mathbb{G}_n\big[b(X(\theta_0) + Z_n(h)/\sqrt{n}) - b(X(\theta_0))\big] \\
&\quad - ng(\sqrt{|h|/n}) \\
&= \mathbb{G}_n\big[Z_n(h)(Y - b'(X(\theta_0)))\big] - g''(0)|h|/2 + o_p(1) \\
&\to_d cB(h) - c^2|h|/(2a(\phi)),
\end{aligned}
$$

where the second line is based on a first-order Taylor expansion of b around $X(\theta_0)$ and a second-order expansion of g around 0, and the last line uses the independence of Z_n and $(b'(X(\theta_0)), Y)$.

[Received September 2008. Revised April 2009.]

REFERENCES

Beran, J. (1994), *Statistics for Long-Memory Processes*, New York: Chapman & Hall/CRC.

Bhattacharya, P. K., and Brockwell, P. J. (1976), "The Minimum of an Additive Process With Applications to Signal Estimation and Storage Theory," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 37, 51–75.

D'Esposito, M., Deouell, L. Y., and Gazzaley, A. (2003), "Alterations in the BOLD fMRI Signal With Ageing and Disease: A Challenge for Neuroimaging," *Nature Reviews Neuroscience*, 4, 863–872.

Dudoit, S., and van der Laan, M. J. (2008), *Multiple Testing Procedures With Applications to Genomics*, New York: Springer.

Embrechts, P., and Maejima, M. (2002), *Selfsimilar Processes*, Princeton, NJ: Princeton University Press.

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Regularization Paths for Generalized Linear Models via Coordinate Descent," available at *http://www-stat.stanford.edu/~hastie/Papers/glmnet.pdf*.

James, G. M., and Silverman, B. W. (2005), "Functional Adaptive Model Estimation," *Journal of the American Statistical Association*, 100, 565–576.

James, G. M., Wang, J., and Zhu, J. (2009), "Functional Linear Regression That's Interpretable," *The Annals of Statistics*, 37, 2083–2108.

Koul, H. L., Qian, L., and Surgailis, D. (2003), "Asymptotics of M-Estimators in Two Phase Linear Regression Models," *Journal of Stochastic Processes and Applications*, 103, 123–154.

Lindquist, M. A. (2008), "The Statistical Analysis of fMRI Data," *Statistical Science*, 23 (4), 439–464.

Lindquist, M. A., Waugh, C., and Wager, T. D. (2007), "Modeling State-Related fMRI Activity Using Change-Point Theory," *NeuroImage*, 35, 1125–1141.

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

McKeague, I. W., and Sen, B. (2009), "Fractals With Point Impact in Functional Linear Regression," *The Annals of Statistics*, to appear.

Müller, H.-G., and Song, K.-S. (1997), "Two-Stage Change-Point Estimators in Smooth Regression Models," *Statistics & Probability Letters*, 34, 323–335.

Prentice, R. L., and Pyke, R. (1979), "Logistic Disease Incidence Models and Case-Control Studies," *Biometrika*, 66, 403–411.

Qin, J., and Zhang, B. (1997), "A Goodness of Fit Test for Logistic Regression Models Based on Case-Control Data," *Biometrika*, 84, 609–618.

Radchenko, P. (2008), "Mixed-Rates Asymptotics," *The Annals of Statistics*, 36, 287–309.

Ramsay, J. O., and Silverman, B. W. (2006), *Functional Data Analysis* (2nd ed.), New York: Springer.

Revuz, D., and Yor, M. (2006), *Continuous Martingales and Brownian Motion* (3rd ed.), New York: Springer.

Richardson, A. L., Wang, Z. C., De Nicolo, A., Lu, X. et al. (2006), "X Chromosomal Abnormalities in Basal-Like Human Breast Cancer," *Cancer Cell*, 9, 121–132.

Stryhn, H. (1996), "The Location of the Maximum of Asymmetric Two-Sided Brownian Motion With Triangular Drift," *Statistics & Probability Letters*, 29, 279–284.

van der Vaart, A., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Yao, Y.-C. (1987), "Approximating the Distribution of the Maximum Likelihood Estimate of the Change-Point in a Sequence of Independent Random Variables," *The Annals of Statistics*, 15, 1321–1328.