

A characteristic function approach to the biased sampling model, with application to robust logistic regression

Howard D. Bondell

Department of Statistics, North Carolina State University

Raleigh, NC 27695-8203, U.S.A.

(email: bondell@stat.ncsu.edu)

Abstract

Many sampling problems from multiple populations can be considered under the semiparametric framework of the biased, or weighted, sampling model. Included under this framework is logistic regression under case-control sampling. For any model, atypical observations can greatly influence the maximum likelihood estimate of the parameters. Several robust alternatives have been proposed for the special case of logistic regression. However, some current techniques can exhibit poor behavior in many common situations. In this paper a new family of procedures are constructed to estimate the parameters in the semiparametric biased sampling model. The procedures incorporate a minimum distance approach, but are instead based on characteristic functions. The estimators can also be represented as the minimizers of quadratic forms in simple residuals, thus yielding straightforward computation. For the case of logistic regression, the resulting estimators are shown to be competitive with the existing robust approaches in terms of both robustness and efficiency, while maintaining affine equivariance. The approach is developed under the case-control sampling scheme, yet is shown to be applicable under prospective sampling logistic regression as well.

Keywords: Biased sampling problem; Case-control data; Logistic regression; Minimum distance; Robust regression.

1 Introduction

Many sampling problems with independent samples from multiple populations can be considered under the framework of the biased, or weighted, sampling model. In this semiparametric model, the density of each sample is assumed to be a weighted version of some baseline density. This paper focuses on the case of a parametric weight function with known form, but unknown parameters, and completely unspecified baseline density. The most common example of this model is logistic regression under case-control sampling.

Vardi (1985) discusses the nonparametric maximum likelihood estimator of the baseline cumulative distribution function for a fully specified weight function. Gilbert et al. (1999) use maximum likelihood to jointly estimate this distribution function along with unknown parameters in the weight function. Qin and Zhang (1997) use the biased sampling view of logistic regression to construct a Kolmogorov-Smirnoff statistic as a goodness-of-fit test, while Gilbert (2004) explores the use of other measures of fit based on distribution functions.

In any model, atypical observations can greatly influence the maximum likelihood estimate of the parameters. For logistic regression, several robust alternatives have been proposed in the literature, particularly in terms of systematically downweighting observations. Examples of these procedures include Pregibon (1982), Copas (1988), Künsch et al. (1989), Carroll and Pederson (1993), and Bianco and Yohai (1996).

Bondell (2005) views logistic regression via the case-control viewpoint as a special case of the biased sampling model and demonstrates some potential pitfalls of some standard downweighting approaches. This is due to the covariate distribution being a mixture, while the weighted approaches tend to rely on a homogeneous distribution, as they downweight via elliptical contours around some notion of overall center. It is shown that in the extreme case of one group much smaller than the other, with

sufficiently separated centers, the entire smaller group can be systematically considered as outliers.

The goodness-of-fit tests of Qin and Zhang (1997) and Gilbert (2004) are derived via a measure of discrepancy between two competing estimates of the underlying baseline distribution. In addition to using a discrepancy measure to test for fit, minimum distance estimation approaches have been shown to be highly efficient and robust in numerous situations, particularly in the parametric setting. There are numerous discrepancy measures between distributions that may be used as the basis for the minimum distance approach, see for example Donoho and Liu (1988). Among these distances are the class of weighted Cramér-von Mises distance measures (Parr and Schucany, 1980; Boos, 1981; Ozturk and Hettmansperger, 1997). This class of distance measures are used by Bondell (2005) to construct highly robust and efficient estimators for logistic regression. However, an additional complication appears in this setting, in that any distance based on the cumulative distribution function yields a procedure that, while can be robust, lacks affine equivariance. Heathcote (1977), Feuerverger and McDunnough (1981), and Besbeas and Morgan (2001) discuss the merits of using a distance measure based on the characteristic function in a minimum distance approach for a fully parametric model, including the robustness of the resulting procedure.

In this paper a new family of procedures is constructed to estimate the parameters in the general semiparametric biased sampling model. The procedures incorporate the minimum distance approach using a measure of discrepancy based on the characteristic functions. These estimators are shown to be computed as the minimizer of quadratic forms in appropriate residuals. This representation allows for the straightforward computation of the estimators, along with the derivation of the asymptotic theory.

Applications of the estimators are given for the special case of logistic regression, where they are shown to be competitive with existing procedures, while retaining the

property of affine equivariance.

The remainder of the paper is organized as follows. In section 2, the semiparametric biased sampling model is reviewed, and likelihood estimation of the baseline distribution is discussed. The proposed minimum distance estimation procedure is then introduced in section 3. The special case of logistic regression is investigated in detail in section 4, where asymptotics, computation, and robustness issues are discussed. The new procedure is then compared to existing robust logistic regression procedures via a simulation study and real data example.

2 Model and likelihood

In the semiparametric biased sampling model, a sample of size n_k is taken from population k independently for each $k = 0, \dots, s - 1$, with $n = \sum_{k=0}^{s-1} n_k$. Let $(\mathbf{x}_{k1}, \dots, \mathbf{x}_{kn_k})$ denote the sample from population k , with each $\mathbf{x}_{kj} = (x_{kj1}, \dots, x_{kjp})'$, a $p \times 1$ vector of measurements. Under this model it is assumed that the probability density function corresponding to sampling from population k is a weighted version of a baseline and is given by:

$$f_k(\mathbf{x}) = C_k w_k(\boldsymbol{\theta}, \mathbf{x}) f_0(\mathbf{x}), \text{ for each } k = 0, \dots, s - 1, \quad (2.1)$$

with f_0 some unknown baseline density function. Each $w_k(\cdot, \cdot)$ is a weight function of known form, while $\boldsymbol{\theta}$ is an unknown parameter vector. Furthermore, $w_k(\boldsymbol{\theta}, \mathbf{x}) > 0$ for all $\boldsymbol{\theta}$ and \mathbf{x} , with $C_k = \{\int w_k(\boldsymbol{\theta}, \mathbf{x}) f_0(\mathbf{x}) d\mathbf{x}\}^{-1}$. For simplicity and identifiability it is assumed that $w_0(\boldsymbol{\theta}, \mathbf{x}) \equiv 1$, and hence $C_0 \equiv 1$, for all $\boldsymbol{\theta}$ and \mathbf{x} .

The typical approach to estimation in this model is maximum likelihood. To simplify the notation, assume that there are no ties in the data. The likelihood can be

written as:

$$\mathcal{L}(f_0, \boldsymbol{\theta}) = \prod_{k=0}^{s-1} C_k^{n_k} \prod_{j=1}^{n_k} w_k(\boldsymbol{\theta}, \mathbf{x}_{kj}) f_0(\mathbf{x}_{kj}). \quad (2.2)$$

As in Vardi (1985) and Gilbert et al. (1999), $\mathcal{L}(f_0, \boldsymbol{\theta}) = 0$ if any \mathbf{x}_{kj} is a continuity point of the distribution F_0 . Additionally, for any $\boldsymbol{\theta}$, if F_0 assigns mass to any set outside the discrete set of data values, then F_0 can be replaced with a distribution G_0 with that mass shifted to the data points, with $\mathcal{L}(f_0, \boldsymbol{\theta}) \leq \mathcal{L}(g_0, \boldsymbol{\theta})$. Hence it suffices to consider only discrete distributions which have positive jumps at the data points.

Denote the combined sample by $(\mathbf{z}_1, \dots, \mathbf{z}_n)$, and for each $i = 1, \dots, n$ let $y_i \in \{0, \dots, s-1\}$ be the group indicator and let $p_i = f_0(\mathbf{z}_i)$ denote the jump at \mathbf{z}_i . The likelihood can then be written as:

$$\mathcal{L}(\mathbf{p}, \boldsymbol{\theta}) = \prod_{k=1}^{s-1} c_k^{n_k} \prod_{i=1}^n w_{y_i}(\boldsymbol{\theta}, \mathbf{z}_i) p_i, \quad (2.3)$$

with $c_k = \{\sum_{i=1}^n w_k(\boldsymbol{\theta}, \mathbf{z}_i) p_i\}^{-1}$ and \mathbf{p} denotes the $n \times 1$ vector of probabilities.

For a fixed $\boldsymbol{\theta}$, the nonparametric maximum likelihood estimator of F_0 can be computed via a Lagrange multiplier technique as follows. The approach used here is most similar to those of Vardi (1985) and Qin and Lawless (1994). It is desired to maximize the likelihood (2.3) subject to $p_i > 0$ for all $i = 1, \dots, n$ and $\sum_{i=1}^n p_i = 1$. From (2.3), it suffices to maximize:

$$H(\mathbf{p}) = \sum_{k=1}^{s-1} n_k \log c_k + \sum_{i=1}^n \log p_i + \lambda \left(1 - \sum_{j=1}^n p_j\right), \quad (2.4)$$

where λ is a Lagrange multiplier. Using (2.4) and the definition of c_k , it must be that for each i , the maximum likelihood vector of probabilities, $\tilde{\mathbf{p}}$, satisfies

$$\frac{\partial H}{\partial p_i} = p_i^{-1} - \lambda - \sum_{k=1}^{s-1} n_k c_k w_k(\boldsymbol{\theta}, \mathbf{z}_i) = 0. \quad (2.5)$$

It follows that

$$\sum_{i=1}^n p_i \frac{\partial H}{\partial p_i} = n - \lambda - \sum_{k=1}^{s-1} n_k = 0. \quad (2.6)$$

From (2.6) one gets $\lambda = n_0$ and using (2.5) yields

$$\tilde{p}_i(\boldsymbol{\theta}, \mathbf{c}) = n_0^{-1} \left\{ 1 + \sum_{k=1}^{s-1} \rho_k c_k w_k(\boldsymbol{\theta}, \mathbf{z}_i) \right\}^{-1}, \quad (2.7)$$

where $\rho_k \equiv n_k/n_0$.

Maximum likelihood estimation of $\boldsymbol{\theta}$ would then proceed by substituting the vector $\tilde{\mathbf{p}}$ into (2.3) and maximizing with respect to $\boldsymbol{\theta}$. As is the case with most likelihood based techniques, the resulting parameter estimates are greatly affected by atypical observations. A more robust procedure is proposed in the next section.

3 The procedure

Based on the data, the baseline distribution F_0 can be estimated in two ways. One is to use the empirical distribution of the first (baseline) sample, and the other is to use all samples by exploiting the model (2.1) and use the nonparametric maximum likelihood estimator. If the model were true, and $\boldsymbol{\theta}$ were known, the latter should be used. However, if the model were not true, the empirical distribution of the first sample is the only information regarding F_0 contained in the combined sample. Intuitively, if the model were at least approximately true, the two estimates of F_0 should be close. This idea coupled with the use of measures of discrepancy based on the distribution function is used for goodness-of-fit and estimation in the logistic regression model by Qin and Zhang (1997) and Bondell (2005), respectively, and for goodness-of-fit of the more general biased sampling model by Gilbert (2004). As noted previously, minimum distance approaches to estimation can be highly efficient and robust. Standard minimum distance estimation is based on a measure of discrepancy between the empirical distribution and a parametric family. In this semiparametric setting, both distributions are actually data-based.

The minimum distance approach in this model proceeds as follows. Explicitly, for a fixed $\boldsymbol{\theta}$, F_0 can be estimated in the following two ways:

$$\begin{aligned}\tilde{F}_0(\mathbf{t}; \boldsymbol{\theta}, \mathbf{c}) &= \sum_{i=1}^n \tilde{p}_i(\boldsymbol{\theta}, \mathbf{c}) I\{\mathbf{z}_i \leq \mathbf{t}\}, \\ \hat{F}_0(\mathbf{t}) &= \sum_{i=1}^n \hat{p}_i I\{\mathbf{z}_i \leq \mathbf{t}\},\end{aligned}\tag{3.1}$$

where $\hat{p}_i = n_0^{-1} I\{y_i = 0\}$, $\mathbf{c} = (c_1, \dots, c_{s-1})$, and $\tilde{p}_i(\boldsymbol{\theta}, \mathbf{c})$ is as in (2.7). By the usual definition of the cumulative distribution function, for two p -dimensional vectors, $(\mathbf{x} \leq \mathbf{y}) \equiv (x_1 \leq y_1, \dots, x_p \leq y_p)$.

In this setting, choosing the discrepancy based on the cumulative distribution function leads to a procedure that is not affine equivariant. An alternative choice that has been used in other settings is a discrepancy based on the characteristic function (Heathcote, 1977; Feuerverger and McDunnough, 1981; Besbeas and Morgan, 2001). The use of an empirical characteristic function-based procedure is now proposed for robust estimation in this semiparametric model where two alternative empirical characteristic functions can be computed.

Under the model assumptions in (2.1), the empirical characteristic function can be computed by using either of the two distributions in (3.1), \tilde{F}_0 or \hat{F}_0 . The characteristic functions of these two empirical distributions are given by

$$\hat{f}_{d\tilde{F}_0}(\mathbf{t}) = \sum_{i=1}^n \exp(it'\mathbf{z}_i) \tilde{p}_i(\boldsymbol{\theta}, \mathbf{c}) = n_0^{-1} \sum_{i=1}^n \exp(it'\mathbf{z}_i) \left\{1 + \sum_{k=1}^{s-1} \nu_k w_k(\boldsymbol{\theta}, \mathbf{z}_i)\right\}^{-1}\tag{3.2}$$

and

$$\hat{f}_{d\hat{F}_0}(\mathbf{t}) = \sum_{i=1}^n \exp(it'\mathbf{z}_i) \hat{p}_i = n_0^{-1} \sum_{i=1}^n \exp(it'\mathbf{z}_i) I\{y_i = 0\},\tag{3.3}$$

with $\nu_k \equiv \rho_k c_k$.

To measure the discrepancy between the two distributions, an integrated weighted squared difference between the two characteristic functions is chosen. For notational convenience, set

$$Q_w(\mathbf{z}; \boldsymbol{\theta}, \boldsymbol{\nu}) \equiv \frac{\sum_{k=1}^{s-1} \nu_k w_k(\boldsymbol{\theta}, \mathbf{z})}{1 + \sum_{k=1}^{s-1} \nu_k w_k(\boldsymbol{\theta}, \mathbf{z})}.\tag{3.4}$$

Treating the vector $\boldsymbol{\nu}$ as a nuisance parameter and using the two characteristic functions in (3.2) and (3.3), the estimator $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\nu}})$ is then defined as the solution to the following minimization problem.

$$\begin{aligned} (\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\nu}}) &= \arg \min_{(\boldsymbol{\theta}, \boldsymbol{\nu})} D(\boldsymbol{\theta}, \boldsymbol{\nu}), \\ \text{with } D(\boldsymbol{\theta}, \boldsymbol{\nu}) &\equiv \int |\hat{f}_{d\hat{F}_0}(\mathbf{t}) - \hat{f}_{d\hat{F}_0}(\mathbf{t})|^2 dW_{F_Z}(\mathbf{t}) \\ &= \int \left| \sum_{i=1}^n \exp(it'\mathbf{z}_i) [I\{y_i \neq 0\} - Q_w(\mathbf{z}_i; \boldsymbol{\theta}, \boldsymbol{\nu})] \right|^2 dW_{F_Z}(\mathbf{t}), \end{aligned} \quad (3.5)$$

for some choice of weight function on the frequencies, $dW_{F_Z}(\mathbf{t}) \geq 0$ for all \mathbf{t} , with $\int dW_{F_Z}(\mathbf{t}) < \infty$, which, as written, can depend on the mixture distribution of \mathbf{Z} . Without loss of generality, it is assumed that W_{F_Z} is a distribution, i.e. $\int dW_{F_Z}(\mathbf{t}) = 1$.

PROPOSITION 1 *The minimization problem using the characteristic function can be expressed as the following nonlinear least squares problem.*

$$(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\nu}}) = \arg \min_{(\boldsymbol{\theta}, \boldsymbol{\nu})} \mathbf{r}' S \mathbf{r},$$

with the i^{th} component of the ‘residual’ vector, \mathbf{r} , given by $r_i = I\{y_i \neq 0\} - Q_w(\mathbf{z}_i; \boldsymbol{\theta}, \boldsymbol{\nu})$, and the matrix

$$S = (S)_{ij} = \text{Re}\{\hat{f}_{dW_{F_Z}}(\mathbf{z}_i - \mathbf{z}_j)\} = \int \cos\{\mathbf{t}'(\mathbf{z}_i - \mathbf{z}_j)\} dW_{F_Z}(\mathbf{t}),$$

i.e. the real part of the fourier transform of the weight function evaluated at the pairwise differences.

Proof. By definition,

$$\begin{aligned} D(\boldsymbol{\theta}, \boldsymbol{\nu}) &\equiv \int [\{\sum_{i=1}^n r_i \cos(\mathbf{t}'\mathbf{z}_i)\}^2 + \{\sum_{i=1}^n r_i \sin(\mathbf{t}'\mathbf{z}_i)\}^2] dW_{F_Z}(\mathbf{t}) \\ &= \sum_{i=1}^n \sum_{j=1}^n r_i r_j \int \{\cos(\mathbf{t}'\mathbf{z}_i) \cos(\mathbf{t}'\mathbf{z}_j) + \sin(\mathbf{t}'\mathbf{z}_i) \sin(\mathbf{t}'\mathbf{z}_j)\} dW_{F_Z}(\mathbf{t}) \\ &= \sum_{i=1}^n \sum_{j=1}^n r_i r_j \int \cos\{\mathbf{t}'(\mathbf{z}_i - \mathbf{z}_j)\} dW_{F_Z}(\mathbf{t}), \end{aligned}$$

where the identity $\cos x \cos y + \sin x \sin y = \cos(x - y)$ is used to obtain the last equality.

REMARK: The value of the quadratic form at its minimum can be used as an alternative to the goodness-of-fit test statistic of Gilbert (2004) to test the model assumptions. The null and alternative distributions of this statistic are beyond the scope of this paper.

4 Application to logistic regression

4.1 Estimation and asymptotic distribution

Consider a binary variable Y and a $p \times 1$ vector \mathbf{Z} of covariates. Then the prospective logistic regression model states that

$$\text{pr}(Y = 1 | \mathbf{Z} = \mathbf{z}) = \frac{\exp(\alpha^* + \mathbf{z}'\boldsymbol{\beta})}{1 + \exp(\alpha^* + \mathbf{z}'\boldsymbol{\beta})} \equiv Q(\alpha^* + \mathbf{z}'\boldsymbol{\beta}), \quad (4.1)$$

say, where α^* is a scalar parameter and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters. Under this model formulation, no specification regarding the marginal distribution of \mathbf{Z} is made.

Robust procedures for logistic regression are typically constructed via this prospective formulation given by (4.1). However, the work of Carroll et al. (1995) has shown that for most of these procedures, point estimators and the corresponding asymptotic standard errors for the slope parameters, $\boldsymbol{\beta}$, derived from this sampling scheme remain valid if the sampling is actually of the case-control form. This result matches that of Prentice and Pyke (1979) for the maximum likelihood estimator.

Case-control sampling, also known as retrospective sampling, refers to independently sampling from F_0 , the conditional distribution of $\mathbf{Z} | (Y = 0)$, and F_1 , defined similarly. As in Qin and Zhang (1997) and Bondell (2005), letting f_0 and f_1 denote the corresponding conditional densities and using (4.1), an application of Bayes' rule yields the equivalent semiparametric biased sampling model (2.1) with $s = 2$ groups, $w_1(\boldsymbol{\beta}, \mathbf{z}) = \exp(\mathbf{z}'\boldsymbol{\beta})$, $C_1 = \exp[\alpha^* + \log\{(1 - \pi)/\pi\}]$, and $\pi = \text{pr}(Y = 1)$.

Hence for the case of logistic regression, letting

$$\alpha \equiv \alpha^* + \log[\{(1 - \pi)/\pi\} \rho_1],$$

the ‘residual’ in Proposition 1 is exactly the residual under the prospective formulation with $\alpha^* = \alpha$, i.e. $r_i = y_i - Q(\alpha + \mathbf{z}'_i \boldsymbol{\beta})$.

The representation as a nonlinear least squares problem in Proposition 1 allows some insight into the robustness of the resulting procedure. If one sets the matrix $S = I$, the resulting procedure is to perform least squares estimation for logistic regression, i.e., minimize

$$\sum_{i=1}^n \{y_i - Q(\alpha + \mathbf{z}'_i \boldsymbol{\beta})\}^2.$$

The score function for this procedure is given by

$$\sum_{i=1}^n Q(\alpha + \mathbf{z}'_i \boldsymbol{\beta}) \{1 - Q(\alpha + \mathbf{z}'_i \boldsymbol{\beta})\} \{y_i - Q(\alpha + \mathbf{z}'_i \boldsymbol{\beta})\} \mathbf{z}_i,$$

and is exactly the form of the prediction downweighting estimator discussed by Carroll and Pederson (1993) which weights according to the variance of Y given \mathbf{z} and handles extreme predicted probabilities. This is also the estimator derived by Beran (1982) as the asymptotically minimax robust estimator for contamination neighborhoods based on Hellinger distance. In the next section, it is shown how the matrix S can be made arbitrarily close to the identity by choosing the distribution $W_{F_{\mathbf{Z}}}$ appropriately. In this sense, the objective function for the proposed estimator can be viewed as smoothed version of the least squares objective function.

Taking the prospective viewpoint, the representation of Proposition 1 further allows for the straightforward computation and asymptotic distribution theory using standard techniques for nonlinear least squares. As a consequence of Theorem 1, which follows from Carroll et al. (1995), inference for the slope parameters remains valid under case-control sampling.

Specifically, assuming the true $q = p + 1$ dimensional parameter $\boldsymbol{\xi}_0 \equiv (\alpha_0, \boldsymbol{\beta}_0)$, for large samples, the solution to the optimization problem, $\tilde{\boldsymbol{\xi}} \equiv (\tilde{\alpha}, \tilde{\boldsymbol{\beta}})$, has the following

asymptotic representation (see, for example, Gallant, 1987).

$$\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \doteq (M' SM)^{-1} M' S \mathbf{r}, \quad (4.2)$$

where $M = M(\mathbf{z}, \boldsymbol{\xi}_0)$ is the $n \times q$ matrix whose i^{th} row is given by

$$\mathbf{m}_i = \frac{\partial}{\partial \boldsymbol{\xi}'} Q(\mathbf{v}'_i \boldsymbol{\xi}) = Q(\mathbf{v}'_i \boldsymbol{\xi}) \{1 - Q(\mathbf{v}'_i \boldsymbol{\xi})\} \mathbf{v}'_i \quad (4.3)$$

with $\mathbf{v} \equiv (1, \mathbf{z}')'$, and S and \mathbf{r} are as in Proposition 1.

Hence, an asymptotic approximation to the variance is directly calculated as:

$$Var(\tilde{\boldsymbol{\xi}}) = (M' SM)^{-1} M' S V S M (M' SM)^{-1}, \quad (4.4)$$

where the diagonal matrix, V has i^{th} diagonal element given by the variance of the i^{th} residual,

$$(V)_{ii} = Q(\mathbf{v}'_i \boldsymbol{\xi}) \{1 - Q(\mathbf{v}'_i \boldsymbol{\xi})\}. \quad (4.5)$$

THEOREM 1. *Under the logistic regression model, $\tilde{\boldsymbol{\beta}}$ consistently estimates $\boldsymbol{\beta}$, and the portion of $Var(\tilde{\boldsymbol{\xi}})$ corresponding to $\tilde{\boldsymbol{\beta}}$ consistently estimates the variance matrix of $\tilde{\boldsymbol{\beta}}$, regardless of whether the sampling was done prospectively or retrospectively.*

Proof. The resulting estimating equation from the least squares formulation given by Proposition 1 is prospectively conditionally unbiased, i.e. its expectation conditional on \mathbf{Z} is identically zero, as the matrix S does not depend on the response, Y . Hence the theorem follows directly from the lemma of Carroll et al. (1995, sec. 5).

4.2 Choosing the weight function

It is desirable that the resulting estimator be both equivariant under changing of the class labels, and under affine transformations of the covariate vector. Hence, $\tilde{\boldsymbol{\xi}}$ should be equivariant in that it satisfies the following two properties (A1) and (A2).

A1. Under the transformation given by $T : y \rightarrow (1 - y)$, $\tilde{\boldsymbol{\xi}}$ should transform to $-\tilde{\boldsymbol{\xi}}$.

A2. Let A be a $p \times p$ nonsingular matrix and \mathbf{b} be a $p \times 1$ vector. Under the transformation given by $T : \mathbf{z} \rightarrow (A'\mathbf{z} + \mathbf{b})$, $\tilde{\boldsymbol{\xi}}$ should transform to $(\tilde{\boldsymbol{\alpha}} - \mathbf{b}'A^{-1}\tilde{\boldsymbol{\beta}}, A^{-1}\tilde{\boldsymbol{\beta}})$.

The estimator of Bondell (2005) lacks the affine equivariance property (A2), while the goodness-of-fit tests of Qin and Zhang (1997) and Gilbert (2004) lack the corresponding property of affine invariance. An appropriate family of weight functions for the proposed estimator of this paper ensures equivariance.

Let $\Sigma = \text{Cov}(\mathbf{Z})$. Choose a constant, σ^2 , and set

$$dW_{F_{\mathbf{Z}}}(\mathbf{t}) \equiv g(\sigma^2 \mathbf{t}'\Sigma\mathbf{t}) d\mathbf{t}$$

for some function g such that $\int dW_{F_{\mathbf{Z}}}(\mathbf{t}) = 1$. Note that Σ is used, not Σ^{-1} in the standardization for \mathbf{t} . The value of σ^2 plays the role of a tuning constant and yields a tradeoff between efficiency and robustness.

PROPOSITION 2 *If the weight function is chosen as above, then both equivariance properties (A1) and (A2) are satisfied.*

Proof. Clearly (A1) holds since the objective function depends on y only through the residual. For (A2), first note that the vector of residuals, \mathbf{r} , remains unchanged under the transformation and corresponding transformation of the parameter vector, so that from Proposition 1 it suffices to show that $\int \cos(\mathbf{z}'A\mathbf{t}) dW_{F_{A'\mathbf{z}+\mathbf{b}}}(\mathbf{t}) \propto \int \cos(\mathbf{z}'\mathbf{t}) dW_{F_{\mathbf{Z}}}(\mathbf{t})$, where the proportionality constant is the same for all \mathbf{z} . Clearly, since under the transformation on the covariates, Σ is transformed to $A'\Sigma A$, it follows that $dW_{F_{A'\mathbf{z}+\mathbf{b}}}(\mathbf{t}) \propto dW_{F_{\mathbf{Z}}}(A\mathbf{t})$. After a change of variable, the desired result is obtained. Hence under the transformation, the optimization problem remains unchanged, up to proportionality.

In practice, Σ must be estimated from the data. Any choice of covariance estimator is sufficient for the equivariance in Proposition 2 to hold, provided that the covariance estimator is affine equivariant.

REMARK 1: It can be shown that for logistic regression, matching the first moments of the two distributions in (3.1) yields the likelihood score equations.

REMARK 2: Two characteristic functions agreeing in a neighborhood of zero, forces the moments of the two distributions to agree. As $\sigma^2 \rightarrow \infty$, dW moves towards point mass at zero, hence the two characteristic functions are forced to agree more closely on a small neighborhood of zero, thus forcing the first moments to become closer. Hence, intuitively, as $\sigma^2 \rightarrow \infty$, the estimator should tend towards maximum likelihood.

Further insight into how the tuning constant affects the robustness properties of the estimator is gained by inspection of the estimating equation that results from the quadratic form given by Proposition 1. One can see that the estimator $\tilde{\boldsymbol{\xi}}$ is the solution to the estimating equation

$$\sum_{i=1}^n \mathbf{a}_i r_i = \mathbf{0},$$

where the $q \times 1$ vector \mathbf{a}_i is given by

$$\mathbf{a}_i \equiv \sum_{j=1}^n \hat{f}_{dW_{F_{\mathbf{Z}}}}(\mathbf{z}_j - \mathbf{z}_i) Q(\mathbf{v}'_j \boldsymbol{\xi}) \{1 - Q(\mathbf{v}'_j \boldsymbol{\xi})\} \mathbf{v}_j. \quad (4.6)$$

This shows that the weight given to each residual can be viewed as a smoothed version of the products of its neighboring covariate values and variances. Since the original function $dW_{F_{\mathbf{Z}}}$ was elliptical, its fourier transform, $\hat{f}_{dW_{F_{\mathbf{Z}}}}$ acts as an elliptical distribution centered at each observation to determine the weighting for that observation. This centering at each observation is in sharp contrast to the typical approach of centering an elliptical distribution at either zero (for the score function, as in Künsch et al., 1989), or the overall mean of the covariates (Carroll and Pederson, 1993). Now, smaller

values of the tuning constant, σ^2 , yield more spread in the original function, hence less spread in the fourier transform. This peakedness in $\hat{f}_{dW_{F_Z}}$ would then tend to reduce the inclusion of more extreme values in the weights for the bulk of the data, and hence smaller values of σ^2 would lead to more robustness, with larger values leading towards maximum likelihood.

Now, a convenient choice of the function g in the weight function is to let $g \propto \phi$, where ϕ represents the density of a standard normal. Since the characteristic function of a normal distribution is of the same form, this choice of g yields a simple closed form for the characteristic function and, hence, the matrix S . This is the choice of weight function that is used in the examples that follow.

5 Examples

5.1 Some existing robust estimators

The minimum characteristic function (MCF) proposal of this paper is compared to the maximum likelihood estimator and three currently implemented robust estimators for logistic regression in a simulation study and real data example. The robust estimators considered are: the conditionally unbiased bounded influence (CUBIF) estimator of Künsch et al. (1989), the Mallows-type estimator with weights depending on a robust Mahalanobis distance (Carroll and Pederson, 1993), both computed as implemented in the Robust package of S-Plus, and the estimator of Bianco and Yohai (1996) with choice of objective function and implementation as in Croux and Haesbroeck (2003).

The first two downweight via elliptical contours, whereas the third downweights in terms of extreme predicted probability. As heuristically discussed earlier, the new proposal can be viewed as weighting by a smoothed version of extreme predicted probability using elliptical contours at each data point, thus it may exhibit some aspects

of both of the previous types of procedures. All of these estimators, except for the maximum likelihood estimator, are indexed by a tuning constant, variation of which yields a trade-off between robustness and efficiency. As previously mentioned, the weight function for the new characteristic function-based procedure is chosen as the normal density. The covariance matrix Σ is estimated via the Minimum Covariance Determinant (MCD) estimator (Rousseeuw and van Driessen, 1999). Other choices of robust covariance estimator were also considered and the simulation results were similar. These alternative choices are also examined in the real data example.

5.2 Simulation study: normal covariate

To begin the simulation study, 5000 samples of 50 controls and 50 cases are generated from individual bivariate normal distributions with $\Sigma = I$ and the controls having mean, $\boldsymbol{\mu}_0 = (0, 0)$, while the cases, $\boldsymbol{\mu}_1 = (2, 2)$. This yields a true slope parameter vector, $\boldsymbol{\beta} = (2, 2)$.

The tuning constants for the estimators were set to yield 92% efficiency as compared to maximum likelihood with no contamination. This efficiency was computed empirically from the 5000 samples. This resulted in the tuning constants $\sigma^2 = 2.5$ for the minimum characteristic function proposal, $c = 3$ for the Bianco and Yohai estimator, and $c = 1$ and $c = 3.5$ for the CUBIF and Mallows estimators, respectively. To examine robustness properties, a single case, $y = 1$, is added at a given value of the two-dimensional covariate \mathbf{z} to each of the 5000 samples and the bias vector for each of the estimators is empirically estimated. A typical configuration of the simulation study is depicted in Figure 1.

*** (FIGURE 1 GOES AROUND HERE) ***

Figure 2 shows a plot of the Euclidean norm of this bias vector for each of the

estimators as a function of contamination location along the direction of the parameter vector. Note that moving along this line yields the quickest change in predicted probability. Under this contamination the proposed procedure performs similarly to the CUBIF and Mallows estimators which are based on elliptical contours, while the Bianco and Yohai gives better performance in terms of bias.

*** (FIGURE 2 GOES AROUND HERE) ***

A second simulation using the normal assumptions was set up to examine the unbalanced situation. Again 5000 samples of size 100 were generated from the normal distributions as in the previous setup. However, here the proportions were changed, 80 controls and 20 cases were used to demonstrate a highly unbalanced situation.

The tuning constants for the estimators were kept the same, and now both the efficiency at the uncontaminated model as well as the bias under point mass contamination is examined. The efficiencies relative to maximum likelihood for the two estimators based on elliptical contours are much lower in this setup due to the fact that any measure of center and covariance is heavily influenced by the larger group, hence the smaller group receives little weight. Recall that the tuning parameters in each of the estimators were set to yield 92% efficiency in the previous setting. The efficiencies for the Künsch et al. and Mallows estimators using the same tuning constants are now 83.6% and 82.8%, respectively. The Bianco and Yohai estimator and the newly proposed characteristic function-based estimator maintain their efficiencies of 93.6% and 93.0%, respectively, in this setting.

*** (FIGURE 3 GOES AROUND HERE) ***

Figure 3 shows the plot of the Euclidean norm of the bias vector for a single point of contamination along the direction of the parameter vector, as before. This plot

clearly shows that, not only does the new proposal maintain high efficiency, it, along with the Bianco and Yohai estimator, is also more robust than those based on elliptical contours, as the effect of contamination is greatly reduced. The major issue here is that the elliptical contours in the unbalanced situation do not detect the outlying point from the smaller group when it is mixed in with the bulk of the larger group as it is now close to the overall ‘center’. Although its performance is clearly better, the new estimator does retain some of the characteristics of those based on elliptical contours as its bias curve peaks and redescends much more quickly than that of the Bianco and Yohai estimator. In this sense, the Bianco and Yohai estimator is less affected by mid-range outliers than the new estimator, but more affected by extreme outliers.

5.3 Simulation study: gamma covariate

Two normal distributions yield the logistic regression model with symmetric conditional distributions. However, if each distribution were skewed, the logistic regression model can still be appropriate. For example, if the conditional distributions of the covariates were each to follow a bivariate gamma distribution, in the sense that each coordinate had an independent gamma distribution, the logistic regression model holds as well. This results in a skewed distribution for each group.

In this simulation setup, as before, 5000 samples of 50 controls and 50 cases were generated. The conditional distributions were generated according to the gamma distributional assumptions with the true parameter again $\beta = (2, 2)$. Specifically for the controls, each component is given by a Gamma distribution with rate = 3 and for the cases rate = 1. The shape parameters for both groups is set to 3 so that there is a fair amount of skewness to the data. These choices give a mean of (1, 1) for the controls and (3, 3) for the cases, with a skewness of $2\sqrt{3}/3$ for each component. Figure 4 shows a typical data configuration for this setting.

*** (FIGURE 4 GOES AROUND HERE) ***

Under the uncontaminated model, the Mallows estimator loses a great deal of efficiency relative to maximum likelihood as it is now only 79.4% efficient. This is due to the fact that the estimator measures outliers directly in terms of elliptical contours for the covariate, while the gamma distribution is highly skewed. The efficiencies of the estimators of Künsch et al. and Bianco and Yohai, along with the characteristic function-based estimator remain stable at 92.5%, 93.1% and 92.1%, respectively.

The robustness properties of the estimators for contaminating points along the direction of the parameter vector are similar to that of the balanced case for the normal distributions and are not shown.

Figure 5 shows the norm of the bias vector for contaminating points along a line orthogonal to the direction of the parameter vector, the constant probability contour. As expected, due to the fact that the Bianco and Yohai estimator determines influence in terms of predicted probability, its behavior parallels maximum likelihood for this contamination setup whereas the new proposal of this paper in addition to the other two robust estimators correctly downweight this outlying observation. This type of contamination results in similar performances in all of the simulation setups and is therefore omitted elsewhere.

*** (FIGURE 5 GOES AROUND HERE) ***

As previously discussed, the unbalanced situation is where the existing procedures based on elliptical contours tend to behave extremely poorly. To examine this behavior, again 5000 samples of size 100 are generated according to the gamma assumptions using 80 controls and 20 cases.

In this highly skewed unbalanced situation, the Bianco and Yohai estimator and the newly proposed procedure greatly outperform the other procedures. The efficiencies for the characteristic function-based proposal and the Bianco and Yohai estimator remain high at 90.8% and 92.8%, respectively. The estimator of Künsch et al. loses a great deal of efficiency coming in at 82.8%, but the Mallows estimator is completely unstable and its efficiency dropped to an extremely poor 24.6%.

Again the norm of the bias due to contamination is shown in Figure 6. Clearly, the Bianco and Yohai and the characteristic function-based approaches behave much better than the other methods under contamination. Noting that the scale on the vertical axis is shrunk by more than 50% in order to accommodate the plots, it is clear that the existing proposals, particularly the Mallows, are not downweighting the observations that should be downweighted, but are instead downweighting the observations that actually fit the model, this is due to the use of elliptical contours in the downweighting scheme.

*** (FIGURE 6 GOES AROUND HERE) ***

Other configurations were investigated, such as changing the dimension of the covariate, varying the covariance structure, changing the parameters, and contamination of more than a single point. The results were similar to those shown, and thus omitted.

5.4 Leukemia data

The leukemia data set from Cook and Weisberg (1982) was analyzed by Carroll and Pederson (1993), among others. There are 33 observations with the response, Y , indicating survival for 52 weeks, and the two predictors, Z_1 and Z_2 , are white blood cell count (WBC) and indicator of acute granuloma (AG). The median value for

WBC in the original data set is 10500, while the median absolute deviation (rescaled to be consistent for the standard deviation at the normal distribution) is 12602.1. Before performing the analysis, this covariate was standardized to have median zero and median absolute deviation of one, as suggested by the associate editor.

It is well documented that a possible response outlier with $Y = 1$ lies in a cluster of five points with identical white blood cell counts, three with acute granuloma and two without. The five estimators discussed in the simulation studies using the identical tuning constants as calibrated previously are compared on this data set. Table 1 gives the estimated coefficients and estimated asymptotic standard errors for the various procedures keeping the tuning constants the same as in the simulations. Shown for comparison is the maximum likelihood estimator obtained after removing the suspected outlier and thus using only 32 observations, denoted by MLE_{32} .

*** (TABLE 1 GOES AROUND HERE) ***

The Künsch et al. and Mallows estimators, along with the characteristic function-based estimator of this paper, appear to downweight the suspected outlier in that the estimated coefficient for white blood cell count all move towards that of MLE_{32} . However, the estimator of Künsch et al. does not downweight as severely as the other two in that its estimated coefficient remains closer to the full data maximum likelihood estimator than to MLE_{32} . This seems to imply that, in this setting, the effect of a single outlier is greater on this estimator than on the other two robust estimators. In contrast, the Bianco and Yohai estimator does not downweight the outlier at all and is almost identical to maximum likelihood for this data. The tuning constant in this estimator was then varied from the one used in the simulations in order to make the estimator more robust, until the resulting estimate was somewhat in line with the other robust estimates. This change in tuning constant from $c = 3$ to $c = 0.5$ reduces the

efficiency at the uncontaminated model so that the resulting estimator is now a mere 61.2% efficient under the uncontaminated equal proportion normal setup, whereas all of the others are set to 92%.

Although the Bianco and Yohai estimator performed well in terms of robustness in the simulations, particularly with the normal distributions, its nonrobustness in the real data example for the same tuning constant appears to indicate a strong dependence on the covariate distribution even when the dimension is unchanged.

As previously mentioned, the minimum characteristic function procedure uses a covariance matrix estimate so it is reasonable to consider the effect of choosing alternative robust covariance matrices in addition to the MCD estimator. Table 2 gives the estimated coefficients and estimated asymptotic standard errors for the MCF procedure using various choices of covariance estimators keeping the same tuning constant. The estimators are the MCD, an S-estimator (Rocke, 1996), the Donoho-Stahel projection-based estimator (Maronna and Yohai, 1995), and a robust estimator based on pairwise correlations (Maronna and Zamar, 2002), all implemented in the Robust package of S-Plus. From the table, one can see that regardless of the choice of covariance estimator, the newly proposed procedure tends to behave similarly to the MLE with the outlier removed. For this data, the use of the S-estimator agrees more closely with this MLE, so it appears to give the most satisfactory solution, but this will not be the case in every situation.

*** (TABLE 2 GOES AROUND HERE) ***

6 Discussion

This paper has proposed a new procedure for robust estimation under the semi-parametric biased sampling model. As a special case, the logistic regression model was investigated in detail. Overall, the new characteristic function-based approach is shown to be competitive with existing robust estimators for the logistic regression model under the various simulation scenarios and the real data example. As shown in the simulations and example, there is no one best robust estimator for logistic regression, as they each have their strengths and weaknesses. Some existing estimators are shown to perform extremely poorly in certain scenarios, so it appears that the newly proposed method of this paper for robust estimation in logistic regression can be a worthwhile addition to a careful data analysis.

As mentioned earlier, the covariance matrix used to ensure affine equivariance must be estimated from the data. One can choose any root- n consistent robust estimator for Σ , and the asymptotic distribution remains unchanged regardless of choice. The robustness of the resulting estimator does not appear to be heavily affected by the choice of robust covariance matrix estimator, as its main purpose is to determine the shape of the contours that are placed around each data point. In the simulation examples, the Minimum Covariance Determinant estimator (Rousseeuw and van Driessen, 1999) was used for Σ . Other choices were examined and the results were similar. For the leukemia data, small differences in the performance when using alternative estimators were noted, as is typically the case when comparing procedures based on plugging-in robust covariance estimates. The direct comparison of the covariance estimators themselves is outside the scope of this paper.

The proposed procedure also yields a natural goodness-of-fit test statistic to test the validity of the model. The properties of this test statistic under both prospective and retrospective sampling deserve further investigation.

Acknowledgement

The author would like to thank David E. Tyler and Butch Tsiatis for their helpful comments. The author is also grateful to an associate editor and two referees for their feedback which helped to improve this manuscript. This work was partially funded by a grant from the National Science Foundation.

References

- Beran, R. (1982) Robust estimation in models for independent non-identically distributed data. *Ann. Statist.*, **10**, 415-428.
- Besbeas, P. and Morgan, B. J. T. (2001) Integrated squared error estimation of Cauchy parameters. *Statist. Probab. Lett.*, **55**, 397-401.
- Bianco, A. M. and Yohai, V. J. (1996) Robust estimation in the logistic regression model, in *Robust Statistics, Data Analysis, and Computer Intensive Methods, Lecture Notes in Statistics*, **109**, Ed. H. Rieder, pp. 17-34. New York: Springer-Verlag.
- Bondell, H. D. (2005) Minimum distance estimation for the logistic regression model. *Biometrika*, **92**, 724-731.
- Boos, D. D. (1981) Minimum distance estimators for location and goodness of fit. *J. Am. Statist. Assoc.*, **76**, 663-670.
- Carroll, R. J. and Pederson, S. (1993) On robust estimation in the logistic regression model. *J. R. Statist. Soc. B*, **55**, 693-706.
- Carroll, R. J., Wang, S. and Wang, C. Y. (1995) Prospective analysis of logistic case-control studies. *J. Am. Statist. Assoc.*, **90**, 157-169.
- Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*. London: Chapman and Hall.
- Copas, J. B. (1988) Binary regression models for contaminated data (with discussion). *J. R. Statist. Soc. B*, **50**, 225-265.

- Croux, C. and Haesbroeck, G. (2003) Implementing the Bianco and Yohai estimator for logistic regression. *Comp. Statist. Data Anal.*, **44**, 273-295.
- Donoho, D. L. and Liu, R. C. (1988) The automatic robustness of minimum distance functionals. *Ann. Statist.*, **16**, 552-586.
- Feuerverger, A. and McDunnough, P. (1981) On some Fourier methods for inference. *J. Am. Statist. Assoc.*, **76**, 379-387.
- Gallant, A. R. (1987) *Nonlinear Statistical Models*. New York: Wiley.
- Gilbert, P. B. (2004) Goodness-of-fit tests for semiparametric biased sampling models. *J. Statist. Plan. Infer.*, **118**, 51-81.
- Gilbert, P. B., Lele, S. R. and Vardi, Y. (1999) Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*, **86**, 27-43.
- Heathcote, C. R. (1977) The integrated squared error estimation of parameters. *Biometrika*, **64**, 255-264.
- Künsch, H. R., Stefanski, L. A. and Carroll, R. J. (1989) Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models. *J. Am. Statist. Assoc.*, **84**, 460-466.
- Maronna, R. A. and Yohai, V. J. (1995) The behavior of the Stahel-Donoho robust multivariate estimator. *J. Am. Statist. Assoc.*, **90**, 330-341.
- Maronna R. and Zamar R. (2002) Robust estimation of location and dispersion for high-dimensional datasets. *Technometrics*, **44**, 307317.
- Ozturk, O. and Hettmansperger, T. P. (1997) Generalised weighted Cramér-von Mises distance estimators. *Biometrika*, **84**, 283-294.
- Parr, W. C. and Schucany, W. R. (1980) Minimum distance and robust estimation. *J. Am. Statist. Assoc.*, **75**, 616-624.
- Pregibon, D. (1982) Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, **38**, 485-498.

- Prentice, R. L. and Pyke, R. (1979) Logistic disease incidence models and case-control studies. *Biometrika*, **66**, 403-411.
- Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating equations. *Ann. Statist.*, **22**, 300-325.
- Qin, J. and Zhang, B. (1997) A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*, **84**, 609-618.
- Rocke, D. M. (1996) Robustness properties of S-estimators of multivariate location and shape in high dimension. *Ann. Statist.*, **24**, 1327-1345.
- Rousseeuw, P. J. and van Driessen, K. (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41**, 212-223.
- Vardi, Y. (1985) Empirical distributions in selection bias models. *Ann. Statist.*, **13**, 178-203.

Table 1: *Estimates (standard errors) for the Leukemia data.*

	Intercept	WBC	AG
MLE	-1.641 (0.772)	-0.400 (0.232)	2.261 (0.950)
CUBIF $c=1$	-1.731 (0.853)	-1.567 (0.838)	2.325 (1.078)
Mallows $c=3.5$	-1.945 (1.150)	-2.349 (1.774)	2.396 (1.199)
BY $c=3$	-1.634 (0.754)	-0.397 (0.335)	2.250 (0.878)
BY $c=0.5$	-1.573 (0.862)	-1.435 (2.194)	2.158 (1.021)
MCF $\sigma^2=2.5$	-1.957 (1.082)	-2.413 (1.584)	2.387 (1.192)
MLE_{32}	-2.260 (1.153)	-2.967 (1.705)	2.558 (1.234)

Table 2: *Comparison of the proposed characteristic function-based procedure for different covariance estimators for the Leukemia data.*

	Intercept	WBC	AG
MCD	-1.957 (1.082)	-2.413 (1.584)	2.387 (1.192)
S	-2.183 (1.119)	-2.810 (1.636)	2.512 (1.216)
Donoho-Stahel	-2.140 (1.111)	-2.732 (1.624)	2.490 (1.211)
Pairwise	-2.159 (1.115)	-2.769 (1.630)	2.501 (1.213)
MLE_{32}	-2.260 (1.153)	-2.967 (1.705)	2.558 (1.234)

Figure 1: Typical setup for the simulation study with normal covariates. Each covariate is bivariate normal, with 50 cases and 50 controls. The solid line represents the classifying boundary. The dashed line represents the direction of the parameter vector. A single observation with $y = 1$ is placed at the point $z_1 = z_2 = -0.5$, denoted by a + in the figure.

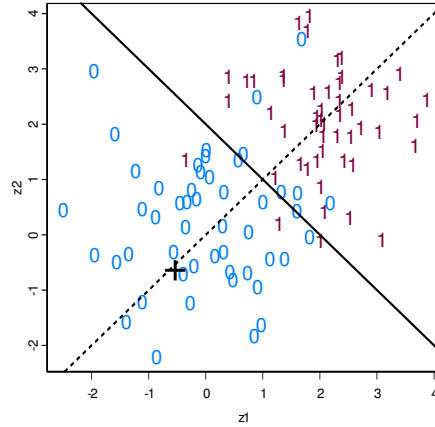


Figure 2: Plots of the norm of the bias incurred from a single contaminating point with $y = 1$ at various locations along the line $z_1 = z_2$ with 50 controls and 50 cases under bivariate normal distributions. The horizontal axis represents the value of z_1 .

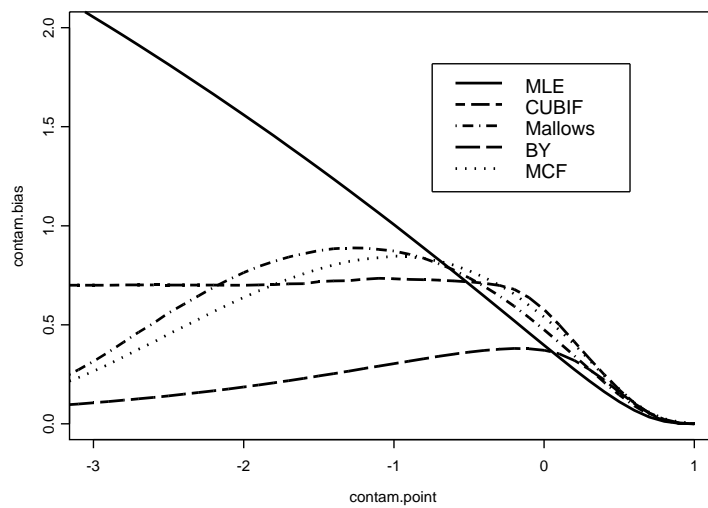


Figure 3: Plots of the norm of the bias incurred from a single contaminating point with $y = 1$ at various locations along the line $z_1 = z_2$ with 80 controls and 20 cases under bivariate normal distributions. The horizontal axis represents the value of z_1 .

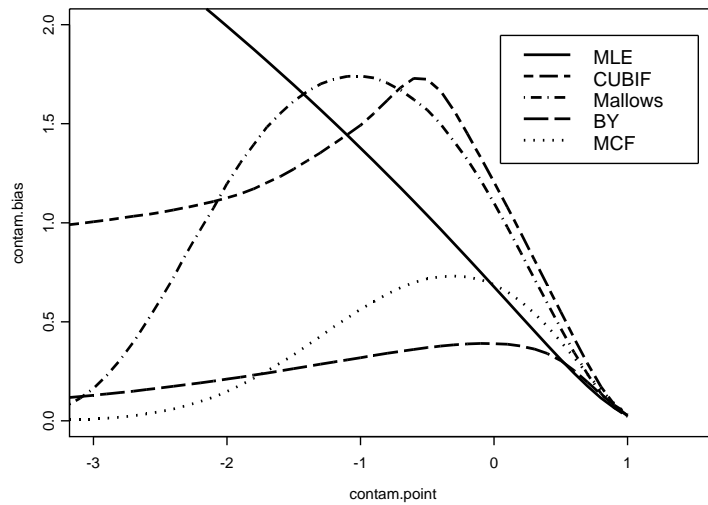


Figure 4: Typical setup for the simulation study with gamma covariates. Each covariate is bivariate gamma, with 50 cases and 50 controls. The solid line represents the classifying boundary. The dashed line represents the direction of the parameter vector. A single observation with $y = 1$ is placed at the point $z_1 = z_2 = 0.25$, denoted by a + in the figure.

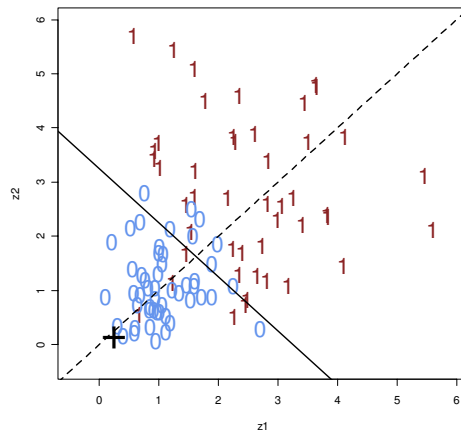


Figure 5: *Plots of the norm of the bias incurred from a single contaminating point with $y = 1$ at various locations along the line $z_1 = -z_2$ with 50 controls and 50 cases under bivariate gamma distributions. The horizontal axis represents the value of z_1 .*

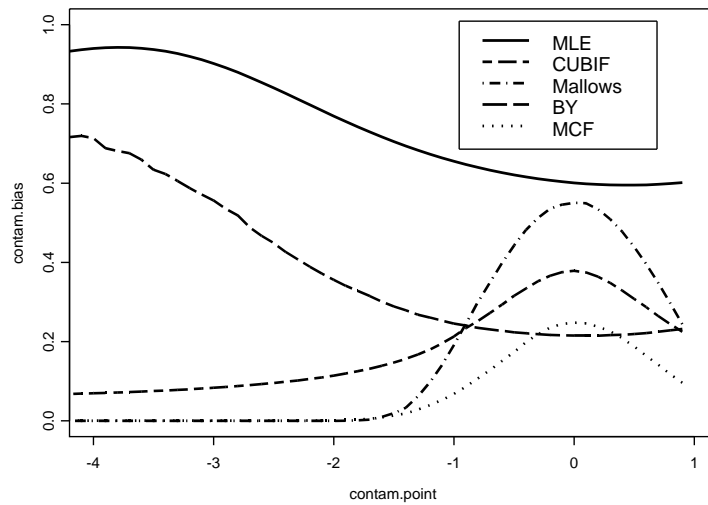


Figure 6: Plots of the norm of the bias incurred from a single contaminating point with $y = 1$ at various locations along the line $z_1 = z_2$ with 80 controls and 20 cases under bivariate gamma distributions. The horizontal axis represents the value of z_1 .

