A Systematic Relationship Between Minimum Bias and Generalized Linear Models

Stephen Mildenhall

CAS Spring Meeting
May 1999

Use Statistics!

- But, what is relationship between statistics and minimum bias methods (MBM)?
- Brown's *Proceedings* paper showed MBMs are maximum likelihood solutions to a statistical model for some examples
  - Statistical models specify loss distributions and modeled values from them, and an objective function to solve for parameters

Overview of Talk

- Background
- Three problematic points
- Main results of paper
- So what?

Background

- Work grew out of class plan study for CNA Personal Lines
- Struggled with “problematic points”
- Wanted to use statistics
- Stumbled across result that all “linear” minimum bias methods correspond naturally to a statistical Generalized Linear Model (MGM = GLM)

Background

MBMs and GLMs have symbiotic relationship

MBM users:

- Faster computation method using SAS etc.
- Underlying assumptions expressed in explicit statistical model
- Model fit and other statistics to interpret results
- Simple intro to GLMs

GLM users:

- Intuitive framework for GLMs
- Mathematically and notational more elementary description of GLMs

Problematic Point #1

- Cents or Percents?

Cents!

Problematic Point #2

- Uniqueness of parameters in a class plan
  - Class plans make hard-to-see choices for base classes to ensure parameters are unique
    - Married, 30-49, pleasure use has a factor of 1.00
    - Plan has deviations for business use, single, different age groups
    - **No** factors for pleasure use, married, age 30-49 etc.

Problematic Point #2

- Uniqueness of parameters in a class plan
  - Naïve statistical models have too many parameters so parameter values are not unique
    - **Predicted** values are unique though **parameters** are not:

      $$x_{\text{age } i} + y_{\text{vehicle use } j} = (x_{\text{age } i} + a) + (y_{\text{vehicle use } j} - a)$$

    - SAS GLM message
    - Choice of base classes corresponds to deleting columns from design matrix
    - It is possible to get unique parameter estimates from statistical models

Problematic Point #3

- Statistical notation and minimum bias notation
  - ANOVAs and MBMs tend to use notation

$$\text{rate}_{\text{age i, vehicle use j}} = x_{\text{age i}} + y_{\text{vehicle use j}}$$

  - Regression and linear models tend to use different notation

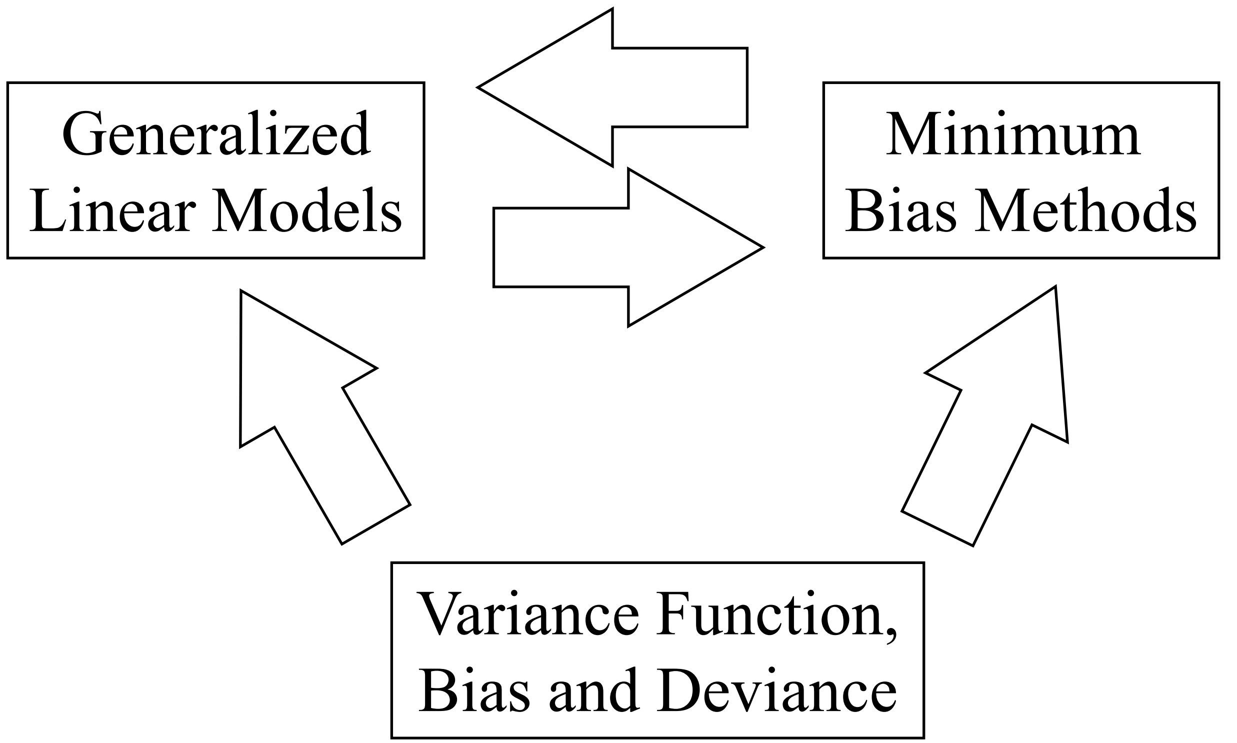
$$\text{rate}_k = a + b\,'\,x_k \quad (k=1,\ldots,n)$$

  - Expressing the change of notation using matrices shows that an additive MBM is a General Linear Model (Section 4)

Main Points of Paper

- ANOVA, regression and General Linear Models
- General Linear Models
- Generalized Linear Models (GLMs)
- Minimum Bias Framework
- Bias, Variance Functions and Deviance
- Exponential Family Distributions
- Main Results

Main Points of Paper



ANOVA, Linear Regression, and General Linear Models

- ANOVA and linear regression are both special cases of General Linear Models
- Response = linear combination of continuous variables + linear combination of discrete variables + Normal Error term
- $\text{Weight}_i = a + b \times \text{age}_i + c_{\text{sex}(i)} + \text{error}_i$
- Discrete variables are **effects**
  - "sex effect", "controlling for other effects"
- Values of effects are called **levels**

General Linear Models

- $\text{Response}_i \sim \text{Normal}(\text{Mean}_i, \text{Variance})$
  - Mean = linear combination of continuous and discrete variables
  - Variance = constant (up to weights)
- Objective: maximum likelihood
- Generalized Linear Models allow three important extensions

Generalized Linear Models

- $Response_i \sim ExpDist(Mean_i, Variance_i)$
  - Mean = **function of** linear combination of variables
  - Variance = **function of** fitted mean
  - ExpDist = Member of **exponential family** of distributions
    - Family includes normal, gamma, inverse Gaussian, Poisson, and binomial distributions
  - Objective: still maximum likelihood

Minimum Bias Framework

- Minimize weighted average bias over all other classes, for each class in turn
- Iterate until results converge
- Minimum bias is equivalent to zero bias when bias can be positive or negative
- MBM becomes “balanced by class”
  - “And in the aggregate” comes for free

Bias, Variance Functions and Deviance

- Minimum bias is an objective
- How should bias be measured?
- How should individual biases be added to get total bias needed in objective?
- Bias generally proportional to predicted value minus observed value
  - Can be positive or negative

Bias, Variance Functions and Deviance

- Variance function, V, defines a bias:
  Bias = (Predicted-Observed)/V(Predicted)
  - E.g. $V(x)=1$, or $V(x)=x^2$
- Variance functions allow less weight to be given observations considered to have high variance
- Allows biases to be added in reasonable manner

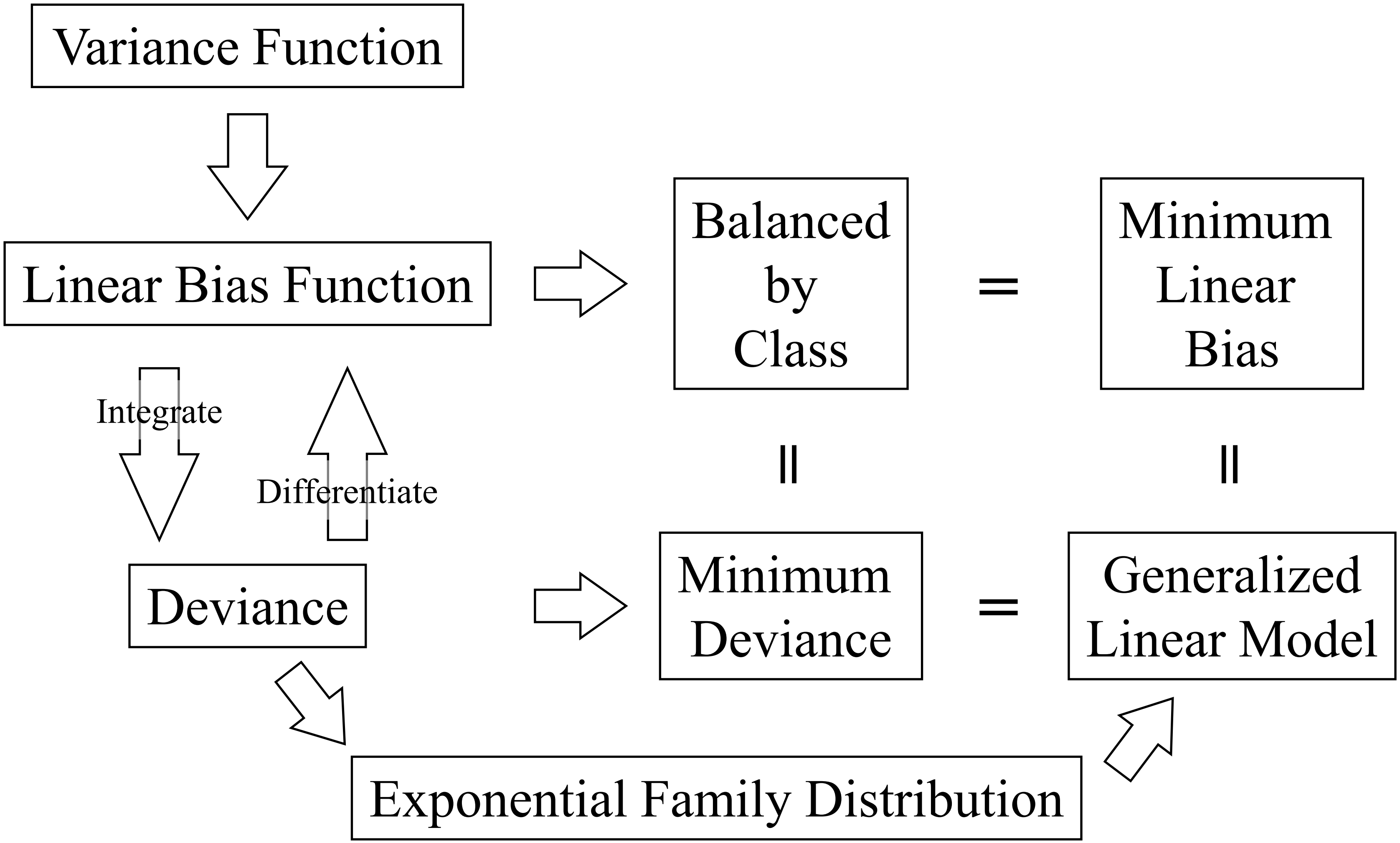
Bias, Variance Functions and Deviance

- Deviance is a measure of overall model goodness of fit
  - Like distance, greater than or equal to zero
  - E.g. Sum of squared differences
- Can define a deviance from a bias function, providing key link between the two concepts

Exponential Family Distributions

- Normal distribution in linear models
  - Form of density implies maximum likelihood = minimum squared differences
- Exponential family generalizes normal distribution with similar defining property
  - Distributions correspond to deviance functions
  - Maximum likelihood = minimum deviance

Main Results



Main Results

- Can write down relationship between MBMs and GLMs based on variance function
  - V(x) Distribution

V(x)	Distribution
$V(x)=1$	Normal
$V(x)=x$	Poisson
$V(x)=x^2$	Gamma
$V(x)=x^3$	Inverse Gaussian
$V(x)=x(1-x)$	Binomial

Main Results

- Minimum bias parameters equal maximum likelihood parameters to the corresponding generalized linear model
- MBM parameters can be obtained using iterative method, or using other methods related to GLM

So What?

- Algorithms available to solve GLMs are much quicker than iterating the MBM
  - Moreover, GLMs are programmed into SAS and other statistical packages
  - GLM solution always applies, even when iterative paradigm not available
- User of GLM has a statistical model which can be tested for reasonableness for given application

So What?

- User of GLM has statistical output from model to assess
  - Model fit and comparison of different models
  - Significance of individual effects and selection of variables in class plan
  - Significance of different levels of an effect
    - Should males be rated higher/lower than females?

So What?

- GLMs offer greatly increased flexibility over general linear models
  - Choice of error distribution suitable for insurance applications (positively skewed)
  - Independent choice of link function to make effects additive
    - In General Linear Models use of log transformation forces lognormal errors

So What?

Stop using MBMs and start using GLMs!