Applied Likelihood methods:
With examples in R and SAS

Russell B. Millar
Department of Statistics
University of Auckland

March 4, 2009
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>viii</td>
</tr>
<tr>
<td>Part One: Preliminaries</td>
<td>i</td>
</tr>
<tr>
<td>1 Likelihood</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Motivating example</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Using SAS and R</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 Using SAS</td>
<td>9</td>
</tr>
<tr>
<td>1.3.2 Using R</td>
<td>12</td>
</tr>
<tr>
<td>1.4 Exercises</td>
<td>14</td>
</tr>
<tr>
<td>2 A taste of likelihood</td>
<td>15</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>15</td>
</tr>
<tr>
<td>2.1.1 Some necessary notation</td>
<td>16</td>
</tr>
<tr>
<td>2.1.2 MLEs of functions of the parameters</td>
<td>18</td>
</tr>
<tr>
<td>2.2 Interpretation of likelihood</td>
<td>20</td>
</tr>
<tr>
<td>2.3 IID Examples</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Exercises</td>
<td>28</td>
</tr>
<tr>
<td>Part Two: Pragmatics</td>
<td>31</td>
</tr>
<tr>
<td>3 Pragmatics</td>
<td>32</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>32</td>
</tr>
<tr>
<td>3.2 Approximate normality of MLEs</td>
<td>32</td>
</tr>
<tr>
<td>3.2.1 Estimating the approximate variance</td>
<td>33</td>
</tr>
<tr>
<td>3.2.2 Estimating variances using R and SAS</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Tests, confidence intervals and confidence regions</td>
<td>37</td>
</tr>
<tr>
<td>3.4 Wald tests, confidence intervals and regions</td>
<td>39</td>
</tr>
<tr>
<td>3.4.1 Test for a single parameter</td>
<td>39</td>
</tr>
</tbody>
</table>
3.4.2 Joint test of two or more parameters ...................................... 39
3.4.3 Using R and SAS ........................................................................ 40
3.5 Likelihood ratio tests, confidence intervals and regions .................. 42
3.5.1 Using R and SAS ........................................................................ 44
3.5.2 Using R ...................................................................................... 45
3.5.3 Using SAS .................................................................................. 46
3.6 More examples ................................................................................ 48
3.7 Exercises ....................................................................................... 51

4 What you really need to know ............................................................. 53
4.1 Introduction ................................................................................... 53
4.2 Approximate distribution of $g(\theta)$ ............................................ 54
  4.2.1 The delta method ...................................................................... 54
  4.2.2 The delta method in R and SAS ............................................. 59
4.3 Delta method examples .................................................................. 60
  4.3.1 Example 1: Variance of a product. ....................................... 60
  4.3.2 Example 2: Variance of log odds-ratio ................................. 62
  4.3.3 Example 3: Vector transformation ......................................... 64
4.4 Wald statistics - quick and dirty? ................................................. 65
  4.4.1 Wald versus likelihood ratio .................................................. 66
4.5 Profile likelihood ........................................................................... 68
4.6 Model selection ............................................................................. 71
  4.6.1 Hypothesis tests versus model selection criteria .................. 71
  4.6.2 AIC and BIC ........................................................................... 71
  4.6.3 In R and SAS ........................................................................... 71
4.7 Bootstrapping ............................................................................... 72
  4.7.1 Bootstrap simulation ............................................................... 73
  4.7.2 Bootstrap confidence intervals .............................................. 73
  4.7.3 Bootstrap estimate of variance ............................................. 75
  4.7.4 Bootstrapping the Old Faithful geyser ................................. 75
  4.7.5 How many bootstrap simulations is enough? ....................... 80
4.8 Prediction ..................................................................................... 81
  4.8.1 Prediction in Practice ............................................................. 83
    The plug-in approach ................................................................. 83
    Predictive likelihood ................................................................. 84
    Pseudo-Bayesian prediction ..................................................... 84
    Bootstrap prediction ................................................................. 85
4.9 Things that can mess you up ........................................ 86
  4.9.1 Multiple maxima of the likelihood ............................ 86
  4.9.2 Parameters on the boundary of the parameter space .... 87
  4.9.3 Non-arrival at Asymptopia .................................. 88
4.10 Tweaks ................................................................ 88
4.11 Exercises ................................................................ 88
4.12 New Chapter on regression models ............................... 90
  4.12.1 The linear-normal model ..................................... 90
  4.12.2 The maximum likelihood model ............................ 92

5 Maximizing the likelihood .............................................. 93
  5.1 Introduction .......................................................... 93
  5.2 The Newton-Raphson algorithm ................................. 94
  5.3 The EM (Expectation - Maximization) algorithm .......... 95
    5.3.1 Properties of the EM algorithm .......................... 99
    5.3.2 Inference from the EM algorithm ......................... 103
  5.4 Exercises ................................................................ 103

6 Some widely used applications of ML .............................. 106
  6.1 Box-Cox transformations .......................................... 106
    6.1.1 Case study ..................................................... 107
      Using R .............................................................. 109
      Using SAS ......................................................... 109
    6.2 Models for survival data ....................................... 110
      6.2.1 Proportional-hazards model ............................ 111
      6.2.2 Partial likelihood: Cox’s proportional-hazards model 113
    6.3 Mark-recapture models ........................................... 114
      6.3.1 Mark-recapture likelihood ............................... 115
      6.3.2 Inference ..................................................... 118
    6.4 Zero-inflated models for count data .......................... 118
    6.5 Models for time-series data ................................... 118
    6.6 Exercises .......................................................... 118

7 Exponential Family models (including GLM’s) .................. 120
  7.1 Exponential family distributions ................................ 120
  7.2 Generalized linear models ........................................ 122
  7.3 Information calculations .......................................... 123
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4</td>
<td>Deviance and Pearson statistics</td>
<td>124</td>
</tr>
<tr>
<td>7.5</td>
<td>Binomial and Poisson models</td>
<td>125</td>
</tr>
<tr>
<td>7.6</td>
<td>Beyond binomial and Poisson models</td>
<td>126</td>
</tr>
<tr>
<td>7.6.1</td>
<td>Quasi-likelihood and quasi-AIC</td>
<td>127</td>
</tr>
<tr>
<td>7.6.2</td>
<td>Parametric alternatives</td>
<td>128</td>
</tr>
<tr>
<td>7.7</td>
<td>Case studies</td>
<td>128</td>
</tr>
<tr>
<td>7.7.1</td>
<td>Logistic regression and inverse prediction</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>Using R</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>Using SAS</td>
<td>135</td>
</tr>
<tr>
<td>7.7.2</td>
<td>Case study 2: An additive model for count data</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>Implementation differences between R and SAS</td>
<td>145</td>
</tr>
<tr>
<td>7.8</td>
<td>Exercises</td>
<td>145</td>
</tr>
<tr>
<td>8</td>
<td>Quasi-likelihood and Estimating functions</td>
<td>146</td>
</tr>
<tr>
<td>8.1</td>
<td>Quasi-likelihood</td>
<td>146</td>
</tr>
<tr>
<td>8.2</td>
<td>Case study: Quasi-likelihood in R</td>
<td>147</td>
</tr>
<tr>
<td>8.3</td>
<td>Estimating functions</td>
<td>152</td>
</tr>
<tr>
<td>9</td>
<td>Maximum likelihood estimation in the presence of nuisance parameters</td>
<td>154</td>
</tr>
<tr>
<td>9.1</td>
<td>Conditional likelihood</td>
<td>154</td>
</tr>
<tr>
<td>9.2</td>
<td>Marginal Likelihood</td>
<td>159</td>
</tr>
<tr>
<td>9.3</td>
<td>Profile likelihood</td>
<td>160</td>
</tr>
<tr>
<td>9.4</td>
<td>Penalized likelihood</td>
<td>162</td>
</tr>
<tr>
<td>9.5</td>
<td>Mixed-effects models (aka. Mixture models, Empirical Bayes models)</td>
<td>163</td>
</tr>
<tr>
<td>10</td>
<td>Models with unobserved random bits</td>
<td>169</td>
</tr>
<tr>
<td>Part Three: Theoretical foundations</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Cramér-Rao inequality and Fisher information</td>
<td>171</td>
</tr>
<tr>
<td>11.1</td>
<td>Introduction</td>
<td>171</td>
</tr>
<tr>
<td>11.2</td>
<td>The Cramér-Rao inequality for $\theta \in \mathbb{R}$</td>
<td>172</td>
</tr>
<tr>
<td>11.3</td>
<td>CR inequality for $g(\theta)$</td>
<td>174</td>
</tr>
<tr>
<td>11.4</td>
<td>Alternative formulae for $I(\theta)$</td>
<td>176</td>
</tr>
<tr>
<td>11.5</td>
<td>The iid data case</td>
<td>177</td>
</tr>
<tr>
<td>11.6</td>
<td>The multi-parameter case, $\theta \in \Theta \subset \mathbb{R}^r$</td>
<td>178</td>
</tr>
<tr>
<td>11.7</td>
<td>Exercises</td>
<td>183</td>
</tr>
</tbody>
</table>
12 Asymptotic theory

12.1 Introduction .................................................. 185

12.2 Consistency and asymptotic normality .................. 186
    12.2.1 Asymptotic normality: $\theta \in \mathbb{R}$ ....................... 190
    12.2.2 Asymptotic normality: $\theta \in \mathbb{R}^s$ ....................... 192
    12.2.3 Asymptotic normality: Non iid case ......................... 193

12.3 Approximate normality .............................. 193

12.4 Wald tests and confidence regions ................. 196
    12.4.1 Wald test statistic ........................................... 196
    12.4.2 Wald confidence intervals and regions ................. 198
    12.4.3 Wald tests and regions for $g(\theta) \in \mathbb{R}^p$ ............. 199

12.5 Likelihood ratio statistic .......................... 200
    12.5.1 Likelihood ratio test: $\theta \in \mathbb{R}$ ....................... 200
    12.5.2 Likelihood ratio test for $\theta \in \mathbb{R}^s$ and $g(\theta) \in \mathbb{R}^p$ 201

12.6 Rao-score test statistic† .......................... 201

12.7 Exercises .................................................. 203

13 Theoretical Tools ........................................ 205

13.1 Equivalence of tests and confidence intervals .......... 205

13.2 Transformation of variables ........................ 205

13.3 Relevant probability theory .......................... 206

13.4 Relevant inequalities ................................ 212
    13.4.1 Useful identities .......................................... 214

13.5 Exercises .................................................. 215

14 Fundamental paradigms and principles of inference .... 218

14.1 Introduction ................................................ 218

14.2 Sufficiency principle .................................. 219

14.3 Conditionality principle ............................... 223

14.4 The likelihood principle ................................ 225
    14.4.1 Relationship with sufficiency and conditionality .......... 226

14.5 Statistical significance versus statistical evidence † 228

14.6 Exercises .................................................. 230

15 Miscellaneous ............................................... 232

15.1 Notation ............................................... 232

15.2 Do you think like a frequentist or a Bayesian? .... 233
## 15.3 Useful distributions

15.3.1 Discrete distributions

15.3.2 Continuous distributions

## 15.4 R miscellania

15.4.1 Modification to profile likelihood function

## 15.5 SAS miscellania

15.5.1 Useful Macros

### Partial solutions to selected exercises