ESTIMATION OF THE BINOMIAL PARAMETER:
IN DEFENCE OF BAYES (1763)

A THESIS SUBMITTED FOR THE DEGREE OF

Doctor of Philosophy

By

Frank Tuyl

M.Ec.

School of Mathematical & Physical Sciences,

The University of Newcastle

May 2007
I hereby certify that the work embodied in this thesis is a result of original research and has not been submitted for a higher degree to any other University or Institution.

Frank Tuyl
Interval estimation of the Binomial parameter $\theta$, representing the true probability of a success, is a problem of long standing in statistical inference. The landmark work is by Bayes (1763) who applied the uniform prior to derive the Beta posterior that is the normalised Binomial likelihood function. It is not well known that Bayes favoured this ‘noninformative’ prior as a result of considering the observable random variable $x$ as opposed to the unknown parameter $\theta$, which is an important difference.

In this thesis we develop additional arguments in favour of the uniform prior for estimation of $\theta$. We start by describing the frequentist and Bayesian approaches to interval estimation. It is well known that for common continuous models, while different in interpretation, frequentist and Bayesian intervals are often identical, which is directly related to the existence of a pivotal quantity. The Binomial model, and its Poisson sister also, lack a pivotal quantity, despite having sufficient statistics. Lack of a pivotal quantity is the reason why there is no consensus on one partic-
ular estimation method, more so than its discreteness: frequentist (unconditional) coverage depends on $\theta$.

Exact methods guarantee minimum coverage to be at least equal to nominal and approximate methods aim for mean coverage to be close to nominal. We agree with what seems like the majority of frequentists, that exact methods are too conservative in practice, and show additional undesirable properties. This includes more recent ‘short’ exact intervals.

We argue that Bayesian intervals based on noninformative priors are preferable to the family of frequentist approximate intervals, some of which are wider than exact intervals for particular data values. A particular property of the interval based on the uniform prior is that its mean coverage is exactly equal to nominal. However, once committed to the Bayesian approach there is no denying that the current preferred choice, by ‘objective’ Bayesians, is the U-shaped Jeffreys prior which results from various methods aimed at finding noninformative priors. The most successful such method seems to be reference analysis which has led to sensible priors in previously unsolved problems, concerning multiparameter models that include ‘nuisance’ parameters. However, we argue that there is a class of models for which the Jeffreys/reference prior may be suboptimal and that in the case of the Binomial distribution the requirement of a uniform prior predictive distribution leads to a more reasonable ‘consensus’ prior.
Acknowledgements

I would like to thank Dr Richard Gerlach and Professor Kerrie Mengersen for their supervision of this research, and Professor John Rayner for his assistance in the final stages. What started as a distraction from financial time series, halfway through my first year, became the main focus of this thesis, with the full support of Richard and Kerrie who always shared my enthusiasm, praising any new insights I showed them. But most of all, when I kept finding more sources that I felt needed further investigation, they left me completely free, right into the final months, obviously trusting that eventually I would put it all together.

I would also like to extend my thanks to other colleagues who were happy to be a sounding board, in particular Kim Colyvas, Peter Howley and, most of all, Trevor Moffiet who is so unlucky to share an office with me.

I would like to dedicate this thesis to my wife Helen and daughters Lara, Anna and Stephanie, who had to put up with a husband/father who for several years was even more absent-minded than usual, and to my parents, who have always supported the choices I made.
Contents

Declaration ii

Abstract iii

Acknowledgements v

Chapter 1 Introduction 1

1.1 Problem ....................................................... 1

1.2 Confidence intervals vs Bayesian intervals ................. 2

1.3 Bayes (1763) .............................................. 3

1.4 Intervals for the Binomial parameter ...................... 5

1.5 Data-based inference ...................................... 8

1.6 Conditional coverage ..................................... 11

1.7 Consensus priors .......................................... 12

1.8 Summary ................................................... 14
Chapter 4  Intervals for the Binomial parameter

4.1 Introduction ...................................................... 75
4.2 Bayesian intervals ............................................. 78
4.3 Exact intervals .................................................. 81
   4.3.1 Clopper-Pearson interval ............................... 82
   4.3.2 Early short exact intervals ............................ 85
   4.3.3 The Blaker and Kabaila-Byrne intervals .............. 89
   4.3.4 Approximate is better than exact ..................... 104
4.4 Approximate intervals ........................................ 110
   4.4.1 Score-based methods .................................. 111
   4.4.2 The mid-$P$ method .................................. 112
4.5 Comparison of methods ...................................... 118
   4.5.1 Wald interval criticism ............................... 118
   4.5.2 Recommended approximate intervals .................. 122
   4.5.3 Bayes HPD interval criticism ......................... 130
4.6 Discussion ..................................................... 137

Chapter 5  Data-based inference

5.1 Introduction ................................................... 147
5.2 The case of a single success or failure .................... 148
5.3 The case of no success or failure .......................... 155
5.4 The Rule of Three ............................................ 159
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.1</td>
<td>Credible probability of the confidence interval</td>
<td>223</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Coverage probability of the credible interval</td>
<td>226</td>
</tr>
<tr>
<td>6.5</td>
<td>Conditional coverage based on noninformative priors</td>
<td>229</td>
</tr>
<tr>
<td>6.6</td>
<td>Exact conditional coverage</td>
<td>234</td>
</tr>
<tr>
<td>6.7</td>
<td>Conditional confidence</td>
<td>237</td>
</tr>
<tr>
<td>6.7.1</td>
<td>Confidence and probability again</td>
<td>239</td>
</tr>
<tr>
<td>6.7.2</td>
<td>Exact intervals</td>
<td>242</td>
</tr>
<tr>
<td>6.7.3</td>
<td>Approximate intervals</td>
<td>244</td>
</tr>
<tr>
<td>6.7.4</td>
<td>Comparison</td>
<td>246</td>
</tr>
<tr>
<td>6.8</td>
<td>Discussion</td>
<td>250</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>255</td>
</tr>
<tr>
<td>7.2</td>
<td>Negative binomial sampling</td>
<td>258</td>
</tr>
<tr>
<td>7.3</td>
<td>Example continued: Hardy-Weinberg equilibrium</td>
<td>262</td>
</tr>
<tr>
<td>7.4</td>
<td>Ratio of Multinomial parameters</td>
<td>265</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Reference analysis</td>
<td>265</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Standard analysis</td>
<td>267</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Straightforward alternative derivations</td>
<td>270</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Comparison</td>
<td>272</td>
</tr>
<tr>
<td>7.5</td>
<td>Ratio of independent Binomial parameters</td>
<td>278</td>
</tr>
<tr>
<td>7.5.1</td>
<td>Marginal prior of $\phi$</td>
<td>279</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>The Blaker method applied to the Poisson parameter; threshold values for equivalence of the Blaker and one-sided C-P lower limits: maximum values of $\alpha$ for given $x$.</td>
<td>93</td>
</tr>
<tr>
<td>4.2</td>
<td>The Blaker method applied to the Binomial parameter; threshold values for equivalence of the Blaker and one-sided C-P lower limits: minimum values of $n$ for given $x$ and $\alpha$.</td>
<td>94</td>
</tr>
<tr>
<td>4.3</td>
<td>Four intervals based on $n = 10$ and $\alpha = 0.05$.</td>
<td>96</td>
</tr>
<tr>
<td>4.4</td>
<td>Poisson example: four intervals when $x = 10$.</td>
<td>102</td>
</tr>
<tr>
<td>4.5</td>
<td>Coverage probabilities and interval widths when $1 - \alpha = 0.95$: Clopper-Pearson, Blaker, Agresti-Coull and Score methods. (Source: Blaker (2000))</td>
<td>127</td>
</tr>
<tr>
<td>4.6</td>
<td>Coverage probabilities and interval widths when $1 - \alpha = 0.95$: Mid-$P$, Jeffreys (HPD and ‘equal-tailed’) and Bayes HPD methods.</td>
<td>127</td>
</tr>
<tr>
<td>4.7</td>
<td>Four methods based on $n = 5$ and $\alpha = 0.05$.</td>
<td>129</td>
</tr>
<tr>
<td>4.8</td>
<td>Four methods based on $n = 8$ and $\alpha = 0.01$.</td>
<td>130</td>
</tr>
</tbody>
</table>
4.9 Various coverage probabilities of four methods based on $n = 8$ and
\[ \alpha = 0.01. \]

5.1 Several intervals when $\alpha = 0.05$ and $x = 1$: Binomial ($n = 10$) and
\text{Poisson.}

5.2 Congdon example ($x = 0$ and $n = 8$): 95% credible intervals based
\text{on five ‘noninformative’ priors.}

5.3 Variations on the Rule of Three.

5.4 Cumulative predictive distribution $p(y|x = 0)$ for the Binomial model
\text{when } n \to \infty.

5.5 Beta-binomial scenarios: $p(y|m, x, n)$.

5.6 Multinomial general laws based on remaining non-zero classes.

5.7 Glass breakage example: posterior results based on Berry and mix-
ture priors ($x = 3, n = 3$).

5.8 Glass breakage example: posterior results based on Berry and mix-
ture priors ($x = 0, n = 3$).

5.9 Contrast agent example ($x = 0$ and $n = 167$): aspects of six priors
\text{and corresponding posteriors (between brackets).}

6.1 Price (1763) example based on $x = n = 10$ and the interval $\left[ \frac{2}{3}, \frac{16}{17} \right]$;
\text{the scenario } x = n = 9 \text{ is added to the right.}

6.2 Price (1763) example based on the interval $\left[ \frac{2}{3}, \frac{16}{17} \right]$:
\text{adapted for } x = 9 \text{ and } n = 10.
7.1 Hardy-Weinberg example: prior predictive probabilities for three ‘noninformative’ priors. ........................................... 263

7.2 Comparison of conditional reference, Jeffreys and Bayes priors \( p(\lambda|\phi) \) for \( \phi \to 0 \) and \( \phi = 1 \). ........................................... 273

7.3 Comparison of integrated likelihoods for \( n = 4 \) and \( \phi \to 0 \), based on reference, Jeffreys and Bayes priors. ......................... 274

7.4 Comparison of integrated likelihoods for \( n = 4 \) and \( \phi = 1 \), based on reference, Jeffreys and Bayes priors. ......................... 274

7.5 Comparison of integrated likelihoods for \( n = 10 \) and \( \phi = 1 \), based on reference, Jeffreys and Bayes priors. ......................... 275

7.6 Comparison of prior predictive distributions for \( n = 4 \), based on reference, Jeffreys and Bayes priors. ......................... 276

7.7 Comparison of various 95% intervals for the odds ratio; upper limits are \( \infty \). ........................................... 308

7.8 Comparison of various 95% intervals for the odds ratio. ............ 311

7.9 Conviction of like-sex twins of criminals. ................................. 314

7.10 Comparison of various intervals for an example by Fisher (1935)... 315

xv
List of Figures

2.1 Hardy-Weinberg example: Beta posterior based on the uniform prior, i.e. the normalised likelihood function; 95% credible interval shown. 29

2.2 Muon decay example: unconstrained posterior (solid curve) and constrained posterior (dotted curve). 33

2.3 Scale uniform parameter example: posterior distribution given $y_n = 5$ when $n = 20$; 95% credible upper limit shown. 35

2.4 Cauchy example: posterior distribution given $x_1 = 3$ and $x_2 = 5$; 90% credible interval shown. 41

2.5 Fisher’s Nile problem example: posterior distribution given $X = 1$ and $Y = 6$; two intervals are shown: HPD (solid) and Fisher’s (dotted). 46

4.1 Coverage of the Binomial parameter for $n = 20$ and $\alpha = 0.05$: Wald (solid and thick), Clopper-Pearson (solid) and Bayes HPD (dotted) methods. 76
4.2 Confidence curves for the Binomial parameter when $x = 1$ and $n = 10$: Clopper-Pearson (dotted), and Wilson/Sterne (solid, left) and Blaker (solid, right) intervals. ........................................... 87

4.3 Coverage of the Binomial parameter for $n = 10$ and $\alpha = 0.05$: Clopper-Pearson (solid) and Blaker (dotted) methods. ............... 95

4.4 Coverage of the Binomial parameter vs nominal for $\theta = 0.1$ and $n = 10$: Clopper-Pearson (solid) and Blaker (dotted) methods. . . . 98

4.5 Confidence curves for the Poisson parameter when $x = 6$: Clopper-Pearson (dashed), Kabaila-Byrne (solid) and Blaker (dotted) intervals.100

4.6 Coverage of the Poisson parameter vs nominal for $\lambda = 5$: Kabaila-Byrne (solid) and Blaker (dotted) methods. ....................... 101

4.7 Confidence curves for the Poisson parameter when $x = 10$ for $0.05 \leq \alpha \leq 0.15$: Clopper-Pearson (solid & thick), Kabaila-Byrne (solid), Blaker (dotted) and Bayes HPD (dashed) intervals. ................. 102

4.8 Coverage of the Binomial parameter based on $\alpha = 0.05$ and averaging $n = 10$ and $n = 20$: Clopper-Pearson (solid) and Blaker (dotted) methods. .............................................................. 106

4.9 Fiducial density (solid), Bayes posterior (dotted) and Jeffreys posterior (dashed) when $n = 10$, for $x = 0, 1, 3, 5$. ......................... 116

4.10 Coverage of the Wald interval as a function of $n$, with $\theta = 0.005$ and $\alpha = 0.05$.............................................................. 119
4.11 Unconditional coverage of the Wald interval (lower) and the Wald
interval when modified for $x \leq 5$ (upper) for $\theta = 0.005$ and $\alpha = 0.05$. 122

4.12 Coverage probabilities for six methods, with $\theta = 0.005$ and $\alpha = 0.05$. 123

4.13 Coverage probabilities for six methods, with $n = 50$ and $\alpha = 0.05$. 125

4.14 Coverage probabilities for four methods, with $n = 4$ (left) and $n = 5$
(right), and $\alpha = 0.05$. 128

4.15 Coverage probabilities for four methods, with $n = 8$ and $\alpha = 0.01$. 131

4.16 Coverage averaged between $n = 8$ and $n = 16$, $\alpha = 0.01$: Clopper-
Pearson (solid) and Bayes HPD (dotted) methods. 135

4.17 Coverage vs nominal level for $n = 8$ and $\theta = 0.55$: Clopper-Pearson
(solid) and Bayes HPD (dotted) methods. 137

4.18 Confidence curves for the Binomial parameter when $x = 4$ and $n = 20$
(top) and for $x = 16$ and $n = 80$ (bottom): Wald/Score (left/right,
dotted) and Bayes HPD (solid) intervals. 142

5.1 Binomial likelihood function (normalised) when $x = 1$ and $n = 10$. 153

5.2 Congdon example ($x = 0$ and $n = 8$): the normalised likelihood
function, and upper limits of various 95% intervals. 158

5.3 Glass breakage example: Berry prior (dotted curve) and posterior
(solid curve) based on $x = 0, n = 3$. 194
5.4 Glass breakage example. Left two graphs: left and right side of Berry prior (solid curve) and mixture prior (dotted curve); Right graph: Berry posterior (solid curve) and mixture posterior (dotted curve) based on $x = n = 3$.

5.5 Contrast agent example ($x = 0$ and $n = 167$); Bottom left corner: Beta(0.042, 27.96) prior (dotted curve) and Beta(0.042, 194.96) posterior (solid curve). ‘Top’ right: Beta(1, 398) prior (dotted) and Beta(1, 565) posterior (solid).

5.6 Contrast agent example: posterior probability of an improvement given the number of patients with no reaction. Solid curve: based on Beta($\frac{1}{2}$, $\frac{1}{2}$) prior; dotted curves, bottom to top: based on Beta(1,1), Beta(1,398) and Beta(1,666) priors.

6.1 ‘Absolute sample average’ example for $n = 4$ and $\alpha = 0.1$: intervals based on limits $y \pm z_{\alpha/2}/\sqrt{n}$ (dashed); horizontal reference lines (dotted) at $\pm z_{\alpha/2}/\sqrt{n} = \pm 0.82$.

6.2 ‘Absolute sample average’ example for $n = 4$ and $\alpha = 0.1$: unconditional coverage $C(\theta)$ (solid curve) and conditional coverage $C(y)$ (dotted curve).
6.3 Truncated Exponential example with $n = 3$ and $\alpha = 0.1$: conditional coverage, as a function of the sample mean minus the sample minimum, of the ‘bad’ confidence interval (solid curve) and of the ‘good’ confidence, or credible, interval (dotted line). .......................... 217

6.4 Cauchy example with $n = 2$ and $\alpha = 0.1$: conditional coverage of the fixed width interval (solid curve) and of the credible interval (dotted line). Also shown: the fraction of samples with half-range greater than $|y|$ (dashed curve). .................................................. 219

6.5 Casella & Berger Poisson example with $\alpha = 0.1$: conditional coverage of Clopper-Pearson (+), Bayes HPD (o) and Gamma(1,1) based (*) intervals when Nature’s prior is Gamma(1,1) (top) and close to Uniform (bottom). .......................... 225

6.6 Casella & Berger Poisson example with $\alpha = 0.1$. Top: unconditional coverage of the Clopper-Pearson interval (solid and thick) and the credible interval based on the Gamma(1,1) prior (solid); bottom: unconditional coverage of the Bayes HPD interval. .......................... 228

6.7 Conditional coverage of the Poisson parameter: 95% HPD interval estimation based on Haldane, Jeffreys and Bayes priors (horizontally) vs parameter ‘generation’ by $p(\lambda) \propto \lambda^{-1}, \lambda^{-\frac{1}{2}}, 1$ (vertically). .......................... 230
6.8 Conditional coverage of the Binomial parameter \((n = 20)\): 95\% HPD interval estimation based on Haldane, Jeffreys and Bayes priors (horizontally) vs parameter ‘generation’ by \(p(\theta) \propto \theta^{-1}(1-\theta)^{-1}, \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}, 1\) (vertically). ........................................... 232

6.9 Conditional coverage of the Binomial parameter when \(n = 20\) and \(\alpha = 0.05\), based on a uniform Nature’s prior. Top: Clopper-Pearson (+), Blaker (*) and mid-\(P\) (o); bottom: Agresti-Coull (+), Score (*) and Wald (o). ........................................... 233

6.10 ‘Exact’ conditional coverage of the Binomial parameter when \(n = 20\) and \(\alpha = 0.05\). Top: Clopper-Pearson (+), Blaker (*) and mid-\(P\) (o); middle: Agresti-Coull (+), Score (*) and Wald (o); bottom: Haldane HPD (+), Jeffreys HPD (*) and Bayes HPD (o). ........................................... 235

6.11 Confidence curves for the Binomial parameter when \(n = 20\) and \(x = 1\) (top left), \(x = 5\) (top right), \(x = 9\) (bottom left), \(x = 10\) (bottom right): Wardell (solid), Clopper-Pearson (dotted) and Bayes HPD (dashed) intervals. ........................................... 238

6.12 Confidence curves for the Binomial parameter when \(x = 10\) and \(n = 10\) (left), and \(x = 9\) and \(n = 10\) (right): Blaker (solid), Clopper-Pearson (dotted) and Bayes HPD (dashed) intervals. ................. 249
7.1  Marginal prior of the ratio of two Multinomial parameters: reference/Jeffreys prior (solid curve) and Bayes prior based on a uniform prior for the original Multinomial parameters (dotted curve). . . . . . . . 269

7.2  Marginal prior of the relative risk parameter: reference/Jeffreys prior (solid curve) and Bayes prior (dotted) based on uniform priors for the individual Binomial parameters. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 280

7.3  Infinite discrete case: pdf for $\theta = 1$ (solid curve) and $\theta = 10$ (dotted). 284

7.4  Infinite discrete case: likelihood for $z = (0.2, 0.4)$ (solid curve), $z = (0.2, 0.6)$ (dotted) & $z = (0.2, 0.8)$ (broken). . . . . . . . . . . . . . . . . . . . . . . . . . . 285

7.5  Infinite discrete case: likelihood (solid curve) and reference posterior (dotted) for a sample of size 1000 from pdf with $\theta = 100$. . . . . . . . . . . . . 288

7.6  Infinite discrete case: ‘normalised’ likelihood (solid curve) and reference posterior (dotted) for a sample of size 250,000 from pdf with $\theta = 10$. . . . . . . 290

7.7  Finite discrete case ($N = 200$): uniform reference posterior (o) and ‘continuous’ reference posterior (+) for a sample of size 1000 from pdf with $\theta = 100$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 294

7.8  Finite discrete case ($N = 2000$): uniform reference posterior (o) and ‘continuous’ reference posterior (+) for a sample of size 250,000 from pdf with $\theta = 10$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 295