Objective Bayes and Conditional Frequentist Inference

A thesis presented for the degree of

Doctor of Philosophy

in the

Department of Mathematics

of

Imperial College London

by

Todd Alan Kuffner

Department of Mathematics

Imperial College

180 Queen's Gate, London SW7 2BZ

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Copyright

To my parents.

Abstract

Objective Bayesian methods have garnered considerable interest and support among statisticians, particularly over the past two decades. It has often been ignored, however, that in some cases the appropriate frequentist inference to match is a conditional one. We present various methods for extending the probability matching prior (PMP) methods to conditional settings. A method based on saddlepoint approximations is found to be the most tractable and we demonstrate its use in the most common exact ancillary statistic models. As part of this analysis, we give a proof of an exactness property of a particular PMP in location-scale models. We use the proposed matching methods to investigate the relationships between conditional and unconditional PMPs. A key component of our analysis is a numerical study of the performance of probability matching priors from both a conditional and unconditional perspective in exact ancillary models. In concluding remarks we propose many routes for future research.

Acknowledgements

Firstly, I thank my supervisor, Alastair Young whose guidance, advice and patience made it possible to accomplish far more in far less time than would have otherwise been possible. I benefited tremendously from our regular discussions and feedback; I cannot imagine a better supervisor. Alastair also arranged for me to visit Tom DiCiccio at Cornell and Jens Jensen at Aarhus. Both visits contributed substantially to this thesis and resulted in ongoing collaborations and plans for future work. In particular, the collaboration with Tom DiCiccio heavily influenced the contents and direction of this thesis, and Chapter 2, Section 2 is currently in preparation for joint publication along with Alastair. Tom also acted as a second-supervisor in addition to being a co-author and I am very grateful for his hospitality and helpful discussions.

I thank the Imperial Department of Mathematics for graciously supporting my visit to Cornell University and Aarhus University for supporting my visit there. My PhD viva examiners, Professor Trevor Sweeting and Dr. Ajay Jasra contributed outstanding comments and helpful suggestions which greatly strengthened the dissertation.

My parents were unwavering in their support and patience and for that I am forever grateful. My flatmates, Anna Raute and Nathan Foley-Fisher, as well as friends Zsofia Barany and Daniel Rogger, were invaluable in times of peril and jubilant in times of triumph. Dalia Ghanem, Anna Sagan & Andrew Klise were also fundamentally important to the process; their inspiration and support are immensely appreciated. I also thank açaí and mango, those magnificent fruits which continue to sustain and inspire me.

Todd Alan Kuffner

Table of contents

List of Tables

Chapter 1

Introduction

Parametric inference often makes use of asymptotic results which are inaccurate in small samples. In such settings, statisticians have proposed many routes to higher-order accuracy. Three routes are of primary importance in the literature: frequentist asymptotic refinements, Bayesian methods and computational tools.[1] Our focus will be the study of how frequentist and Bayesian asymptotics are related to one another and, more specifically, the identification of Bayesian priors which lead to inference with desirable frequentist properties. We work in particular on settings in which the appropriate frequentist inference is a conditional one, which is something that has been largely ignored by existing literature on this topic.

The remainder of this chapter will detail the history and current state of the literature, as well as define the problem precisely. In Chapter 2, we will present four methods which may be used to identify conditional probability matching priors. Chapter 3 discusses the relationship between conditional and unconditional inference, with implications for probability matching priors. Chapter 4 is dedicated to worked examples using the matching methods and Chapter 5 presents numerical results. The final chapter offers concluding remarks and directions for future research.

[1]For an up-to-date discussion, see Young (2009).

1.1 Objective Bayes and Unconditional Inference: An Introduction

Objective Bayesian methods have a long and rich history. In fact, the argument could be made that since the advent of Bayesian statistics, objective Bayesian methods were the norm until the second half of the 20th century brought subjective Bayesian methods to the forefront of Bayesian statistics.[2] However, for the purposes of clarity and brevity, it is useful to view the identification of non-informative priors as the beginning of modern objective Bayesian methods. Jeffreys (1946) and Bernardo (1979) introduced Jeffreys' and the reference prior, respectively, which are non-informative in the sense that they add little information to the sample information [3]; i.e. the likelihood function provides nearly all of the information, as opposed to the prior providing significant information. Such priors have also been termed "objective" or "non-informative" based on the interpretation that, in contrast to subjective Bayesian methods, the subjective part of the posterior (i.e. the prior) contributes little information. The natural outcome of choosing priors which provide little information is that the posterior distribution is primarily determined by the likelihood function. Since the likelihood function is the sole determinant of likelihood-based inference, then in this sense, the less "informative" a prior is, the closer Bayesian inference should be to frequentist inference. Due to this result, the term "objective" has thus come to describe Bayesian methods which deliver results closely comparable to results derived from frequentist methods. It is well-known that Bayesian and frequentist probability statements are equivalent to order $O(n^{-1/2})$ in general for regular cases (see, for example Datta & Sweeting (2005)), where n denotes sample size. Therefore, objective Bayesian methods seek to identify priors which have the correct frequentist properties to a higher order of error, e.g. $O(n^{-1})$ or $O(n^{-3/2})$.

The study of objective Bayesian methods is important for at least four key reasons. Firstly, it is important to understand how inferences based on two different foundations of probability theory are related to one another. Secondly, as noted by Rubin (1984), con-

[2]Bernoulli and Laplace, for instance, applied the so-called "principle of insufficient reason", though they considered it too obvious to give it a name. These were some of the earliest documented uses of a uniform prior, treating all parameter values as equally likely.

[3]The term "information" used in this context is deliberately imprecise so as to include various measures of information.

sumers of statistical analysis often think in a Bayesian way, and it is important to verify that frequentist properties of such Bayesian thinking are closely related to probability computations in a repeated sampling framework. A third reason, as demonstrated below by the shrinkage argument, is that sometimes we can calculate frequentist asymptotics via a Bayesian route. Lastly, in many practical circumstances, Bayesian methods are computationally simpler and therefore, if it may be established that such methods have good frequentist properties, then the practitioner may choose which ever method is easiest and rest assured that the inferences will be the same to a higher order of error.

The field of objective Bayesian methods is too large to respectfully summarize here. Instead, we describe developments relevant to the particular type of objective Bayesian method discussed in the subsequent chapters, which is the study of Bayesian priors which deliver posterior credible sets possessing accurate frequentist coverage properties. The setting we will be concerned with is that of one-sided inference about a scalar interest parameter and the goal of the analysis will be to identify priors which deliver posterior credible sets having the correct frequentist probability interpretation up to some higher order of error as a function of the sample size.[4] We now define this more formally.

Unless otherwise noted, the setting we consider throughout is that of a sample $Y = \{Y_1, ..., Y_n\}$ from some continuous distribution indexed by a d-dimensional parameter $\theta = (\theta_1, \ldots, \theta_d) = (\psi, \lambda)$, where $\psi = \theta_1$ is a scalar interest parameter and $\lambda = (\theta_2, \ldots, \theta_d)$ is a $(d-1)$-dimensional nuisance parameter. The Bayesian posterior distribution of θ, conditional on $Y = y$, is given by

$$\pi(\theta|Y = y) \propto \pi(\theta)L(Y; \theta)$$

where $\pi(\theta)L(Y;\theta)$ is the product of the prior and likelihood function. We assume here and throughout that all prior, posterior and other density functions are with respect to Lebesgue measure.

Consider inference about the parameter θ_1 under prior $\pi(\cdot)$ and denote by $\theta_1^{1-\alpha}(\pi, Y)$

[4]As noted by Sweeting (2001), defining an objective Bayes prior on the basis of coverage probability will necessarily contravene the likelihood principle, since different sampling rules can give rise to different objective Bayesian methods. Thus from a Bayesian point of view, such objective Bayesian methods are incoherent. We will not consider alternative sampling rules here. For some discussion see Sweeting (2001)).

the $(1-\alpha)$th marginal posterior quantile for θ_1 under the prior $\pi(\cdot)$. An unconditional *probability matching prior* (PMP), $\pi(\cdot)$, is one which satisfies

$$Pr_\theta\{\theta_1 \leq \theta_1^{1-\alpha}(\pi, Y)\} = 1 - \alpha + O(n^{-m/2}) \tag{1.1}$$

$\forall \alpha \in (0,1)$ for $m = 2$ or 3 [5] which coincide respectively with second- and third-order matching, n is the sample size and Pr_θ is the frequentist probability under repeated sampling of Y. This states that the $(1-\alpha)$th quantile of the posterior density has unconditional frequentist coverage probability, $1-\alpha$, to order of error $O(n^{-m/2})$.[6]

1.1.1 Welch & Peers Matching

Most authors consider Welch & Peers (1963) to be the earliest contribution to the analysis of probability matching priors.[7] Consider the following problem. Given a sample Y from some continuous distribution with density $b(Y;\theta)$ where θ is a scalar parameter, and denoting by $\pi(\theta)$ the prior density for θ. The posterior density for θ after having observed Y is given by:

$$\pi(\theta|Y) = \frac{b(Y;\theta)\pi(\theta)}{\int b(Y;\theta)\pi(\theta)d\theta}$$

Let

$$Q(Y,\theta) = \int_{-\infty}^{\theta} \pi(t|Y)dt$$

be the cumulative distribution function corresponding to the posterior density $\pi(\theta|Y)$. We seek a prior which satisfies the property

[5] Some authors use $o(n^{-1/2})$ and $o(n^{-1})$ instead of $O(n^{-1})$ and $O(n^{-3/2})$, respectively. These are equivalent in this context, but since we are talking about *coverage*, it is more appropriate to use "big" $O(\cdot)$ notation.

[6] Yet another way of stating this, as given in Reid et al. (2003) is that a PMP is a prior for which "posterior probability statements about the parameter also have an interpretation as confidence statements in the sampling model."

[7] However, it seems that Lindley (1958) was a source and inspiration for the method.

$$Pr\{Q(Y,\theta) < \alpha|\theta\} = \alpha + O(n^{-1}),$$

where again we note that this is a frequentist probability. We are interested in a frequentist asymptotic expansion for this probability which is a function of the prior; in this way we can relate the Bayesian and frequentist asymptotics. The goal then is to find a standard normal pivot which is a function of the prior and which admits an asymptotic expansion.

We define various likelihood quantities as follows. Let $\ell(\theta) = \ell(\theta; Y)$ be the log-likelihood function for θ. We denote the joint cumulants of $\partial^k \ell(\theta)/\partial\theta^k$ by $\kappa_{rst\cdots}$ for $k = 1, 2, \ldots$ where the indices refer to the number of derivatives of the log-likelihood. Then $\kappa_{20} = var(\partial\ell(\theta)/\partial\theta)$, which is simply the expected information, and $\kappa_{30} = E\big\{\big[\partial\ell(\theta)/\partial\theta - E(\partial\ell(\theta)/\partial\theta)\big]^3\big\}$.

After some manipulation, it can be shown that we have a standard normal pivot given by

$$z(Y,\theta) = \Phi^{-1}\{Q(Y,\theta)\}.$$

An asymptotic expansion of moment generating function of this pivot is given by

$$E\{\exp(tZ)\} = \exp\left[\frac{1}{2}t^2 + \frac{t}{\sqrt{n}}\left\{\frac{1}{2}\kappa_{20}^{-3/2}\frac{\partial\kappa_{20}}{\partial\theta} - \kappa_{20}^{-1/2}\frac{\partial\log\pi(\theta)}{\partial\theta}\right\} + O(n^{-1})\right]$$

From this we have that a prior is unconditional probability matching to order $O(n^{-1})$ if and only if

$$\frac{\partial\log\pi(\theta)}{\partial\theta} = \frac{1}{2}\kappa_{20}^{-1}\frac{\partial\kappa_{20}}{\partial\theta} \tag{1.2}$$

yielding

$$\pi(\theta) \propto \kappa_{20}^{1/2}, \tag{1.3}$$

which corresponds to Jeffreys' prior.

The authors also consider matching to order $O(n^{-3/2})$ and, through some manipulations using Bartlett identities, arrive at the expansion

$$E\{\exp(tZ)\} = \exp\left[\frac{1}{2}t^2 \frac{t^2}{12n}\kappa_{20}^{-1/2}\frac{\partial}{\partial\theta}(\kappa_{30}\kappa_{20}^{-3/2}) + O(n^{-3/2})\right]$$

from which we see that Jeffreys' prior is unconditional probability matching to order $O(n^{-3/2})$ if and only if the standardized skewness of the score is independent of θ. This is an often-cited limitation of PMPs for one-parameter models. The introduction of nuisance parameters will actually enable us to achieve higher-order matching.

As noted by Johnson (1970), Welch & Peers (1963) essentially calculate an expansion for a weighted likelihood, which is of course mathematically equivalent to a posterior. Therefore, it is not surprising that the Welch & Peers method will turn out to be closely related to methods based on Johnson's expansion of the posterior.

Adapting this to the vector nuisance parameter setting, Peers (1965) considered the marginal posterior for the interest parameter θ_1. Let

$$\ell_m(\theta) = \frac{\partial \ell(\theta)}{\partial\theta_m}, \ell_{mq} = \frac{\partial^2 \ell}{\partial\theta_m\partial\theta_q}$$

so that the cumulants κ are now defined as $\kappa_{ij} = E[-\partial^2\ell(\theta)/\partial\theta_i\partial\theta_j]$ is the (i,j)th component of the expected information matrix and the indices refer to the parameters.

$$Q(Y,\psi) = \int^{\theta_1}\int\dots\int \pi(t|Y)dt_d\dots dt_2 dt_1$$

and again found a standard normal pivot

$$z(Y,\psi) = \Phi^{-1}\{Q(Y,\psi)\}$$

which is the x-axis of the standard normal distribution, and $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function.

Similarly to the scalar parameter setting, series expansion of the moment generating function for this pivot in the vector nuisance parameter setting is given by

$$E(e^{tZ}) = \exp\left[\frac{1}{2}t^2 + \frac{t}{\sqrt{n}}\left\{ -\kappa^{1j}(\kappa^{11})^{-1/2}\frac{\partial \log \pi(\theta)}{\partial \theta_j} - \frac{\partial}{\partial \theta_j}(\kappa^{1j}(\kappa^{11})^{-1/2})\right\} + O(n^{-1})\right]$$

yielding the necessary condition for a prior to be unconditional probability matching to order $O(n^{-1})$,

$$\kappa^{1j}(\kappa^{11})^{-1/2}\frac{\partial \log \pi(\theta)}{\partial \theta_j} + \frac{\partial}{\partial \theta_j}\left\{\kappa^{1j}(\kappa^{11})^{-1/2}\right\} = 0, \tag{1.4}$$

where (κ^{ij}) is the inverse of (κ_{ij}) and we are using the summation convention of summing over all repeated subscripts or superscripts.

In general there will be many solutions to these partial differential equations, though it is difficult (or impossible) to satisfy all of them at the same time to perform inference about every parameter simultaneously. Therefore, the Welch & Peers matching method is most useful when there is a particular scalar parameter of interest.

Peers (1965) also notes conditions under which the probability matching prior is the same regardless of which parameter is of particular interest. He calls this "complete equivalence for all components of θ". This requires that a solution exists to the system of partial differential equations given by

$$\kappa^{ij}(\kappa^{ii})^{-1/2}\frac{\partial \log \pi(\theta)}{\partial \theta_j} + \frac{\partial}{\partial \theta_j}\left\{\kappa^{ij}(\kappa^{ii})^{-1/2}\right\} = 0 \quad (i = 1, 2, \ldots, d) \tag{1.5}$$

Define $\psi_i = \partial \log \pi(\theta)/\partial\theta_i$. A necessary and sufficient condition for the existence of a solution to the system of equations given by (1.5) is that the equations given

$$\frac{\partial \psi_i}{\partial \theta_j} = \frac{\partial \psi_j}{\partial \theta_i} \quad (i \neq j) \tag{1.6}$$

are satisfied. As noted by Peers, this occurs in location-scale models with location parameter θ_1 and scale parameter θ_2 when the prior is chosen as $\pi \propto 1/\theta_2$. This setting will be discussed extensively in Chapter 4.

1.1.2 Parameter Orthogonality and Non-uniqueness

The literature has given some attention to a special setting of parameter orthogonality. Tibshirani (1989) extends an idea from Stein (1985) to show that when the interest parameter is orthogonal to the nuisance parameter with respect to the expected Fisher information matrix, i.e.

$$\kappa_{1,j} = E\left\{\frac{\partial \ell(\theta, Y)}{\partial \theta_1}\frac{\partial \ell(\theta, Y)}{\partial \theta_j}\right\} = 0 \qquad (j = 2, \ldots, d)$$

then any joint prior chosen as

$$\pi(\theta) \propto \kappa_{20}^{1/2} g(\theta_2, \ldots, \theta_d) \tag{1.7}$$

will be unconditional probability matching for θ_1 to order $O(n^{-1})$. A more rigorous derivation and proof of this and the results of Peers can be found in Nicolaou (1993). Berger & Bernardo (1992) show that Tibshirani's prior is the same as the reference prior (defined below) given a particular ordering of the parameters.

More recently, Datta & Mukerjee (2004) and Staicu & Reid (2008) have demonstrated that the identification of PMPs is greatly simplified under parameter orthogonality. We will return to this issue in later chapters.

1.1.3 The Shrinkage Argument

Another Edgeworth-type argument is the shrinkage argument, which was first used by Bickel & Ghosh (1990) which noted that the Bayesian posterior distribution of the likelihood ratio statistic is Bartlett-correctable and used the shrinkage argument to ascertain the validity of frequentist and Bayesian Bartlett correction factors. Ghosh & Mukerjee (1991) used the shrinkage argument to obtain explicit expressions for the frequentist Bartlett correction factors via a Bayesian route and then identified PMPs by equating the Bayesian and frequentist Bartlett correction factors. It is straightforward to extend the argument when one is interested in an arbitrary (but smooth) parametric function of θ, as was shown by

Datta & Ghosh (1995). We postpone full exposition of the shrinkage argument until Chapter 2 where we present a version modified to the conditional setting.

The shrinkage argument approach involves an Edgeworth-type expansion of the distribution of $\sqrt{n}(\hat{\theta}-\theta)$, and the partial differential equations emerging from the corresponding matching exercise have been extended to many other settings, some of which we briefly discuss below. Since its conception, it has been the most commonly applied matching method to identify PMPs. There have also been several attempts – notably Levine & Casella (2003) and Sweeting (2005) – to study the matching conditions and resulting priors numerically. We discuss these findings in comparison to our own numerical work in Chapter 5.

1.1.4 Bartlett Correction

As a first application of the shrinkage argument, Bickel & Ghosh (1990) showed that the Bayesian posterior distribution of the likelihood ratio statistic is Bartlett correctable and that frequentist Bartlett correction factors may be calculated via a Bayesian route. An obvious method, then, to identify matching priors is to set the frequentist and Bayesian Bartlett correction factors equal to some order of error. This analysis can be found in Ghosh & Mukerjee (1991). Many other authors have extended these initial results and demonstrated the usefulness of the Bartlett correction to likelihood ratio statistics in identifying matching priors, most notably DiCiccio & Stern (1993) and Sweeting (1995*b*).

1.1.5 Third-order Probability Matching Priors

The $O(n^{-1})$ term in the expansion obtained by the shrinkage argument may also be set equal to zero, which would ensure unconditional frequentist coverage of the Bayesian posterior credible set to $O(n^{-1})$. The key references include Mukerjee & Dey (1993), Datta & Ghosh (1995) and Mukerjee & Ghosh (1997). We also postpone discussion of this result to Chapter 2.

1.1.6 Other Notable PMP Approaches

We briefly turn to other approaches to the identification of PMPs. The brevity of this section should not be interpreted as a judgement that these approaches are any less valuable, but rather that they are less relevant to the content of the remaining chapters.

Non-regular Cases

Ghosal (1999) investigated PMPs in non-regular cases where the family is not smooth. The main examples are the uniform and shifted exponential. To do this, the asymptotic expansion of the posterior due to Johnson (1970) cannot be used, as it only holds for regular families. Ghosal & Samanta (1997) gives an expansion of the posterior for non-regular cases which has an exponential distribution as the leading term, as opposed to the normal distribution in Johnson's expansion. Ghosal then applies a shrinkage argument to obtain a partial differential equation which must be satisfied to ensure probability matching. It turns out that there is a unique prior in the scalar parameter setting,

$$\pi(\theta) \propto E_\theta\{\frac{\partial \log f(Y;\theta)}{\partial\theta}\},$$

which is probability matching to order $O(n^{-2})$, where Y is a set of i.i.d. observations from density $f(\cdot)$ is the density. However, Ghosal does assume that the function $f(Y;\theta)$ is jointly continuous and differentiable in θ everywhere on its support. There has also been some study of discrete cases, for instance Rousseau (2000), though asymptotic expansions for the frequentist coverage of Bayesian intervals are not possible in this setting without the introduction of some randomization.

Predictive Probability Matching

Given a set of observations $Y = \{Y_1, \ldots, Y_n\}$, we may be interested in predicting the next observation Y_{n+1}. In that case we can identify *predictive probability matching priors* by calculating the frequentist coverage of highest posterior predictive density regions. As demonstrated by Datta et al. (2000), Datta & Mukerjee (2004) and Sweeting (2008), among others, this can be done using the shrinkage argument.

Information Theoretic Methods

All objective Bayes priors are derived from the core principle that the prior should contribute little information to the likelihood. An implication of this is that Bayesian confidence sets derived from objective Bayes priors should have good frequentist properties. The nomenclature *probability matching prior* is used to distinguish priors derived through the matching of Bayesian and frequentist asymptotics for inference from other types of non-informative priors.

The most well-known of non-informative priors derived via information loss arguments are Jeffreys' prior and the reference prior first explored by Bernardo (1979). Jeffreys' prior is given by

$$\pi \propto \sqrt{\det I(\theta)}, \tag{1.8}$$

where $\det I(\theta)$ means the determinant of the Fisher information matrix, and was shown by Welch & Peers (1963) to be second-order unconditional probability matching. Berger & Bernardo (1992) give a thorough account of the reference prior approach. The reference prior attempts to maximize the Kullback-Liebler distance between the prior and the posterior. This amounts to maximizing the expected posterior information about Y given the prior $\pi(\theta)$. In this sense, the prior $\pi(\theta)$ is the *least* informative prior about Y. In the case of a scalar parameter, Jeffreys' prior and the reference prior are the same. As Jeffreys' and the reference prior are the two most common non-informative priors in the literature, we will relate them to our examples in Chapters 4 and 5.

More recently, there has been a revival of interest in using information loss criteria to identify objective Bayesian priors. For example Sweeting et al. (2006) use a posterior predictive regret criterion to minimize the predictive information contained in a prior.

1.2 Conditional Inference

All of the discussion above was in the context of unconditional frequentist inference. However, there are many common statistical settings in which the appropriate frequentist inference is a conditional one. An excellent overview of the roles of conditioning in inference can be found in Reid (1995).

1.2.1 The Motivations for Conditional Inference

Conditional inference was originally proposed by Fisher (1934) in situations where the maximum likelihood estimator is not a sufficient statistic, resulting in a loss of information in the Fisherian sense when reducing the model to the maximum likelihood estimator. This yields incorrect likelihood-based inference. In such settings, Fisher proposed that one should construct an ancillary statistic, whose distribution is free of the model parameters, such that the MLE and ancillary statistic are jointly minimal sufficient. Inference should then be done conditionally on a particular value of the ancillary statistic, in order to satisfy the principle of conditionality. Existence and uniqueness of ancillary statistics are common problems.

Well-known ancillary statistic models include the location-scale model, the exponential regression model and Fisher's gamma hyperbola. These examples are discussed in Fisher (1934), Barndorff-Nielsen & Cox (1994) and Buehler (1982) respectively. They will be of particular importance to us insofar as they are all cases in which there is an exact ancillary statistic. Non-exact ancillary statistic models and the construction of approximate ancillary statistics have been studied by Efron & Hinkley (1978), Cox (1980), Ryall (1981), McCullagh (1984), Skovgaard (1985) and Severini (1990).

Another common motivation for conditional inference is the elimination of nuisance parameters to make inferences more *relevant*. This motivation is argued in, among others, Cox & Reid (1987) and Pierce & Peters (1992), with particular reference to exponential family models.[8]

1.2.2 Conditional Probability Matching

There have been several notable contributions to conditional probability matching.

Conditional Likelihood Ratio Statistics

Ghosh & Mukerjee (1992) apply the Barlett correction arguments originated in Bickel & Ghosh (1990) to the conditional likelihood ratio statistic of Cox & Reid (1987). This relies on parametric orthogonality, which ensures that the nuisance parameters do not depend on the interest parameter[9] and therefore the nuisance parameters (or some sufficient statistic of the nuisance parameters) may serve as a conditioning variable. Ironically, this is the first serious attempt at matching Bayesian and conditional frequentist probabilities, even though the authors' motivation seems to have been to find a simpler route for calculating Bartlett corrections to the *unconditional* distribution of the likelihood ratio statistic, rather than to note that there are circumstances where the appropriate frequentist inference to match is a conditional one.

Adjusted Signed Roots of Likelihood Ratio Statistics

DiCiccio & Martin (1993) derive a Bayesian version of the adjusted signed root likelihood ratio statistic in a form that is directly comparable to Barndorff-Nielsen's R^* formula. DiCiccio & Martin show that, in general, the adjustment terms are equal to order $O(n^{-1})$. Casella et al. (1995) suggests identifying matching priors by setting the Bayesian and frequentist adjustment terms equal. This method is also discussed in Fraser & Reid (2002).

[8] A possible alternative method of eliminating nuisance parameters is through using reference priors. For example, Liseo (1993) shows how reference priors can be used to eliminate nuisance parameters and also calculates frequentist coverage of Bayesian credible sets using reference priors using Gamma and inverse Gaussian examples. He is able to obtain marginal posteriors which compare closely to conditional frequentist profile likelihoods.

[9] Usually this means the parameters are orthogonal with respect to the Fisher information.

We will discuss and extend this method in subsequent chapters.

Choosing Bayesian Interval Endpoints: Severini (1993)

Severini (1993) addresses the question of how to choose Bayesian interval endpoints such that the resulting intervals will have the correct frequentist coverage to order $O(n^{-3/2})$. This is done through comparisons of Johnson's expansion of the posterior with the Barndorff-Nielsen (1985) expansion for the conditional distribution of the maximum likelihood estimator.

Orthogonal Parameters: Nicolaou (1994)

Nicolaou (1994) considers the setting of orthogonal parameters and works with the prior identified by Stein (1985), Tibshirani (1989) and Nicolaou (1993) as being second-order unconditional probability matching, namely,

$$\pi(\theta) \propto \kappa_{20}^{1/2} g(\theta_2, \ldots, \theta_d) \tag{1.9}$$

Nicolaou shows that the frequentist coverage probabilities of Bayesian posterior credible sets derived from this prior, conditional on a second-order locally ancillary statistic, will in general differ from the nominal coverage probabilities by order $O(n^{-1})$.

Locally Ancillary Statistics: Sweeting (1995b, 1999, 2001)

Sweeting (1995a) shows how the Bayesian version of the signed root likelihood ratio statistic may be used to obtain the frequentist version using the shrinkage argument. The frequentist distribution which Sweeting considers is conditional on a locally ancillary statistic. Sweeting (1999) works on a problem closely related to Severini (1993). To construct Bayesian intervals with good frequentist coverage, Sweeting proposes to use Bayesian Bartlett correction of the signed root likelihood ratio statistic, which is a directed likelihood. Sweeting's calculations turn out to be much simpler than those required by Severini, while still yielding accurate conditional frequentist coverage to order $O(n^{-3/2})$, and throughout he conditions on a second-order locally ancillary statistic. The Bayesian Bartlett correction of the directed likelihood is, of course, another application of the shrinkage argument. Sweeting has also approached conditional probability matching from a bias perspective. Using a minimum coverage probability bias criterion, Sweeting (2001) derives a saddlepoint approximation to the signed root likelihood ratio statistic to match Bayes confidence regions with conditional frequentist coverage probabilities in settings where there is a locally ancillary statistic.

Conditional Shrinkage: Ventura et al. (2009)

The Appendix of Ventura et al. (2009) contains a partial attempt at deriving the conditional version of the shrinkage argument, though it is not very direct and only yields matching to order $O(n^{-1})$. As we will argue in Chapter 2, a higher-order conditional shrinkage argument comparable to that found in Datta & Mukerjee (2004) is straightforward to obtain, after noting the validity of the crucial step is established by Sweeting (1995*a*).

1.3 The Scope of Our Work

From this review of the literature, we observe that there are important gaps:

1. Despite the notable contributions listed above, there has been relatively little research on conditional PMPs and, in fact, it has often been ignored that in many cases, the correct frequentist inference to match is a conditional one. Moreover, even among existing studies related to conditional PMPs, there has been little attention given to exact ancillary statistic models.

2. The PMP identification/matching methods, both those already in use and the ones to be discussed in the next chapter, afford an ideal framework in which to compare conditional and unconditional inference. This can be done both directly, through asymptotic comparisons of conditional and unconditional distributional quantities,

as well as indirectly, through comparison of conditional and unconditional matching methods and the priors they identify.

3. There has been very little emphasis in the PMP literature on numerical evaluation of PMPs. It is important to understand how well these priors perform in simulations.

In addressing these gaps, we will devote our attention to both theoretical and practical aspects of matching methods. In Chapter 2 we investigate four routes to the identification of conditional PMPs: (1) the shrinkage argument approach, (2) the mean- and variance-adjusted signed root likelihood ratio statistic approach, (3) the saddlepoint approach and (4) the Edgeworth approach. We discuss the difficulties in relating these matching methods to one another and the relative merits and weaknesses of each approach. In Chapter 3, we study how these methods can be used to compare conditional and unconditional inference, and therefore how they can be used to compare conditional and unconditional PMPs. Chapter 4 contains many examples demonstrating how the saddlepoint matching method, in particular, is useful in identifying conditional PMPs in ancillary statistic models. We consider the most well-known exact ancillary models: location-scale, exponential regression and the exponential hyperbola model. We also present a direct proof of the "folk theorem" result that there is exact matching for a particular choice of prior in location-scale models.

In Chapter 5 we present numerical results comparing conditional and unconditional PMPs for these models which yields suggestive evidence about the importance of conditioning in location-scale models, as well as the performance of different PMPs in these models. We find that PMPs perform fairly well in location-scale models, both in terms of conditional and unconditional probability matching. However, in the exponential hyperbola model, we find that PMPs have better unconditional matching properties.

Chapter 6 concludes with a discussion of ongoing and future work with particular emphasis on how our results highlight existing gaps in the literature and illuminate important unanswered questions.

For the purposes of transparency and to establish the validity and usefulness of the matching methods and resulting comparisons, we need to work with analytically feasible calculations. For this reason, we restrict our focus to ancillary statistic models for which there exist known, exact ancillary statistics. Moreover, we consider only one-sided inference about a scalar interest parameter and assume that the underlying parametric family satisfies general regularity conditions, to be discussed below.

Chapter 2

Matching Methods

We will present four methods of matching conditional frequentist and Bayesian asymptotics to identify probability matching priors. This chapter is concerned with the theoretical explanations of the matching methods. Examples and numerical evaluations are contained in later chapters.

2.1 Method 1: The Conditional Shrinkage Argument

The shrinkage argument, from its conception, was intended to be used as a method for calculating frequentist asymptotics via Bayesian asymptotics. There are two main reasons for doing this, namely that in some circumstances it may be easier to calculate the Bayesian asymptotics and that the posterior expansion used in the argument involves comparatively simpler quantities, which makes it easier to calculate than the corresponding frequentist expansion. The inspiration for this idea is the following: standard[1] first-order theory states that any prior, π in a class of priors, Π, satisfying weak regularity conditions (such as "smoothness"), will yield a posterior distribution of $X_n = I(\theta)^{1/2}\sqrt{n}(\hat{\theta} - \theta)$ which is standard normal to $O(n^{-1/2})$, where $I(\theta)$ denotes the Fisher information per observation. Now, since this distribution is independent of the data, we also have that the *marginal* distribution of X_n is standard normal to $O(n^{-1/2})$. Since this is true for every prior in the class Π, then the sampling distribution of X_n will also be standard normal to $O(n^{-1/2})$. This

[1] Bernstein von Mises Theorem

last step is called an ”unsmoothing” step, where we let the prior converge to a degenerate distribution at the true parameter value. We are left with an asymptotic expansion for the posterior which has a frequentist interpretation.

The shrinkage argument is an implication of Bickel & Ghosh (1990) which pointed out that posterior distributions admit Bartlett corrections. Most authors in the field cite Dawid (1991) as being the most eloquent presentation of the argument. Ghosh & Mukerjee (1991) further develop the unconditional frequentist version of the shrinkage argument and shows how one can identify matching priors by equating the Bayesian and frequentist Bartlett correction terms. A rigorous treatment of the final and crucial step (the ”unsmoothing step”) of the argument can be found in the appendix of Sweeting (1995*b*). In an unpublished PhD thesis at the University of Toronto, Li (1998) gives a lucid account of the steps of the unconditional shrinkage argument, and mentions that one could extend this to the conditional setting, but does not do so in any formal way. Li’s presentation is replicated by Reid & Mukerjee (2000). As argued by Sweeting (1991), the original unconditional version of the shrinkage argument is not immediately applicable to the conditional setting because we require an asymptotic analysis of the sampling distribution being used. This asymptotic analysis is provided in the Appendix of Sweeting (1995*b*,*a*), in which he showed that the ”unsmoothing” step in the conditional version of the shrinkage argument is valid whenever there exists an asymptotic expansion of the desired form. Yet what is still missing from the PMP literature is an application of Sweeting’s results to derive the conditional analogue to the Ghosh-Mukerjee-Dey-type matching conditions for the identification of conditional PMPs. Here we combine the presentation of Datta & Mukerjee (2004) with the results of Sweeting (1995*a*), to obtain a conditional version of the argument and derive matching conditions for Bayesian priors. Our exposition closely mirrors other presentations in the literature and in fact we will arrive at the same matching conditions as those for the unconditional setting, but with unconditional likelihood quantities being replaced by conditional ones.

2.1.1 The Basic Steps

We consider one-sided inference about $\sqrt{n}(\hat{\theta} - \theta)$ and the direct calculation of the frequentist coverage of posterior credible sets for the interest parameter.

Let Y be a vector-valued random variable with density $g(\cdot;\theta)$, $\theta \epsilon \Re^p$, $\hat{\theta}$ denotes the maximum likelihood estimate (MLE) for θ and A denotes an ancillary statistic. We consider the conditional density, which we express as $g(\theta; S)$, where S is the minimal sufficient statistic, so we have $g(Y;\theta|A) \equiv g(\theta; S)$. In ancillary statistic models, for example, we have $S = (\hat{\theta}, A)$. In the posterior setup, it is of no consequence whether we condition on the ancillary statistic or not, and thus there is no loss of generality by writing the full unconditional density as $g(Y;\theta) \equiv g(Y;\theta|A)g(A)$, where $g(A)$ is the marginal density of A and the definition of $g(\cdot)$ will always be clear from its arguments. Our intention is to find an expression for $E_\theta\{h(Y,\theta)|A\}$ where A is a conditioning variable (such as an ancillary statistic), h is a measurable function whose expectation exists. Taking h to be the indicator function, $E_\theta\{h(Y,\theta)|A\}$ is a conditional frequentist probability. We use an auxiliary prior in the argument and we adopt the notation $\overline{\pi}(\cdot)$ to distinguish this auxiliary prior, which we will shrink to a degenerate prior in the argument, from the prior identified by this method as a probability matching prior, $\pi(\cdot)$.

The necessary regularity conditions are from Johnson (1970), Ghosh et al. (1985) and Bickel & Ghosh (1990).[2] We assume the following regularity conditions: (1) $h(Y,\theta)$ is integrable with respect to the joint probability measure for (Y,θ) as induced by $\overline{\pi}(\cdot)$. This allows us to interchange the order of integration in the argument that follows. (2) $\overline{\pi}(\theta)$ has compact support. (3) $E_\theta\{h(Y,\theta)|A\}$ is continuous for all θ.

First, we outline the basic steps:

Step 1 Obtain the posterior density of $\theta|Y$ under prior density $\overline{\pi}(\cdot)$ for θ. Thus we obtain the expectation of $h(Y,\theta)$ with respect to the posterior density, $E_{\overline{\pi}}\{h(Y,\theta)|Y\}$.

Step 2 Next we take the expectation with respect to the conditional frequentist density of Y. Find $E_{\theta|A}E_{\overline{\pi}}\{h(Y,\theta)|Y,A\}$.

Step 3 The "unsmoothing" step is to calculate the expectation with respect to the auxiliary prior. Integrate $E_{\theta|A}E_{\overline{\pi}}\{h(Y,\theta)|Y\}$ with respect to $\overline{\pi}(\cdot)$ and allow $\overline{\pi}(\cdot)$ to converge weakly

[2] These are more or less the same as the regularity conditions in the Appendix of Nicolaou (1993)

to the degenerate prior at θ. This yields $E_\theta\{h(Y,\theta)|A\}$.

Some justification of these steps is needed. The posterior density of θ under the prior $\overline{\pi}(\cdot)$ is given by

$$\pi(\theta|Y) = \frac{g(y;\theta|A)\overline{\pi}(\theta)}{\int g(y;\theta|A)\overline{\pi}(\theta)d\theta}$$

Therefore, by Step 1, we have

$$E_{\overline{\pi}}\{h(y,\theta)|Y\} = \frac{\int h(y,\theta)g(y;\theta|A)\overline{\pi}(\theta)d\theta}{\int g(y;\theta|A)\overline{\pi}(\theta)d\theta}$$

and taking the conditional expectation in Step 2 and using $\breve{\theta}$ to clarify that this is different from θ,

$$E_{\theta|A}E_{\overline{\pi}}\{h(y,\theta)|Y\} = \int \left\{ \frac{\int h(y,\breve{\theta})g(y,\breve{\theta}|A)\overline{\pi}(\breve{\theta})d\breve{\theta}}{\int g(y,\breve{\theta}|A)\overline{\pi}(\breve{\theta})d\breve{\theta}} \right\} g(y;\theta|A)dy$$

Integrating this with respect to $\overline{\pi}(\cdot)$, as in Step 3, yields

$$E_{\overline{\pi}}E_{\theta|A}E_{\overline{\pi}}\{h(y,\theta)|Y\} = \int\int \left\{ \frac{\int h(y,\breve{\theta})g(y,\breve{\theta}|A)\overline{\pi}(\breve{\theta})d\breve{\theta}}{\int g(y,\breve{\theta}|A)\overline{\pi}(\breve{\theta})d\breve{\theta}} \right\} g(y;\theta|A)\overline{\pi}(\theta)dyd\theta$$

Since $h(Y,\theta)$ is integrable, we can change the order of integration due to the Fubini Theorem, which gives

$$E_{\overline{\pi}}E_{\theta|A}E_{\overline{\pi}}\{h(y,\theta)|Y\} = \int \left\{ \frac{\int h(y,\breve{\theta})g(y,\breve{\theta}|A)\overline{\pi}(\breve{\theta})d\breve{\theta}}{\int g(y,\breve{\theta}|A)\overline{\pi}(\breve{\theta})d\breve{\theta}} \right\} \left\{ \int g(y;\theta|A)\overline{\pi}(\theta)d\theta \right\} dy$$

and this is equal to

$$
\begin{aligned}
&\int\int h(y,\theta)g(y;\theta|A)\overline{\pi}(\theta)d\theta dy \\
&= \int \left\{ \int h(y,\theta)g(y;\theta|A)dy \right\} \overline{\pi}(\theta)d\theta \\
&= \int \left\{ E_{\theta|A}\{h(y,\theta)\} \right\} \overline{\pi}(\theta)d\theta
\end{aligned}
$$

The claim in Step 3 follows from the last line.

2.1.2 Derivation of the Matching Conditions

The shrinkage argument, as we now show, leads to a pair of conditions which characterize probability matching priors. To show this, let us formally define the problem.

We are interested in identifying a set of conditions satisfying the conditional version of equation (1.1). The parameter vector $\theta = \{(\theta_1), (\theta_2, \ldots, \theta_d)\} = (\psi, \lambda)$ where ψ is a scalar interest parameter and λ is a possibly vector-valued nuisance parameter. We use n to denote sample size and consider a sequence $\{Y_i\}_{i\epsilon\mathbb{Z}_+}$ of i.i.d. and possibly vector-valued random variables with common density $f(Y;\theta)$ with $\theta\epsilon\Re^p$. Let $\hat{\theta} = (\hat{\psi}, \hat{\lambda})^T$ denote the MLE of θ based on Y.

Additionally, we make a set of assumptions to ensure the existence of a valid Edgeworth expansion up to order $O(n^{-3/2})$ for $\sqrt{n}(\hat{\theta} - \theta)$. The assumptions are given by Johnson (1970), for the Bayesian version, and Bhattacharya & Ghosh (1978) for the frequentist version. Some further discussion can be found in Bickel & Ghosh (1990)[3].

1. θ has prior density $\pi(\cdot)$ which is positive and three times continuously differentiable

[3] A minor technical note in regards to the first assumption is that when $\pi(\cdot)$ is not a proper prior, we assume that there exists a finite N such that for $n \geq N$, the posterior $\pi(\theta|X_1, ..., X_n)$ is proper with well-defined conditional frequentist probability for all θ.

over the parameter space.

2. we consider an expansion for sample points in a set S over which the observed information matrix is positive definite and the MLE $\hat{\theta}$ is well-defined.

While $\pi(\cdot)$ is the prior of primary interest, we make use of an auxiliary prior $\overline{\pi}(\cdot)$ to carry out frequentist computations related to $\pi(\cdot)$. We make the following assumptions about $\overline{\pi}$:

1. $\overline{\pi}(\cdot)$ is a proper prior with compact rectangular support in the parameter space.

2. $\overline{\pi}(\cdot)$ vanishes on the boundary of its support.

3. the first partial derivatives of $\overline{\pi}(\cdot)$ vanish on the boundary of its support.

The Johnson (1967) asymptotic expansion for posterior distributions, while it is not a function of the cumulants of the distribution as derived by direct expansion of the moment generating function, is nonetheless an Edgeworth-type expansion in the spirit of Bhattacharya & Ghosh (1978) in that it is an expansion with terms in powers of $n^{-1/2}$. Therefore, in the literature (e.g. Bickel & Ghosh (1990)), Johnson's asymptotic expansion has been treated as the Bayesian version of an Edgeworth expansion and the conditions for its validity are treated as the conditions for the validity of an Edgeworth expansion of the posterior. We discuss this more later.

Before proceeding, we must define various likelihood quantities used in the argument. Let $\ell(\theta) = n^{-1} \sum_{i=1}^{n} \log f(Y_i; \theta)$ be the average log-likelihood, $\ell \equiv \ell(\theta)$, and let $\hat{\theta}$ denote the overall MLE for θ. Whenever we use a $\hat{\cdot}$ over a particular quantity, it means we are evaluating that quantity at $\theta = \hat{\theta}$. For the purposes of clarity and easier comparison with later results, we will proceed assuming that we are in the setting of an ancillary statistic model, for which there is a known, exact ancillary statistic, A. Then the log-likelihood function may be written as $\ell(\theta) \equiv \ell(\theta; \hat{\theta}, A)$. We then use a $\mathring{\cdot}$ over a quantity to denote that it is a *conditional* quantity, i.e. a quantity based on the *conditional* log-likelihood, e.g. $\mathring{\ell} = \ell(\theta; \hat{\theta}, A)$. The likelihood derivatives are defined as

$$\mathring{\ell}_r = \frac{\partial \mathring{\ell}(\theta)}{\partial \theta_r}, \quad \mathring{\ell}_{rs} = \frac{\partial^2 \mathring{\ell}}{\partial \theta_r \partial \theta_s},$$

$$\mathring{j}_{rs}(\theta) = -\mathring{\ell}_{rs}(\theta); \quad \mathring{j}^{rs}(\theta)$$

are the (r, s)th component of the observed information matrix $\mathring{j}(\theta)$ and its inverse $\mathring{j}^{-1}(\theta)$, respectively. We denote the expected Fisher information matrix by $\mathring{\omega}(\theta)$. Thus we have

$$\mathring{\omega}_{rs}(\theta) = -E(\mathring{\ell}_{rs}(\theta)|A)$$

denotes the (r, s)th element of the expected information matrix and $\mathring{\omega}^{rs}(\theta)$ corresponds to the (r, s)th component of the inverse of the expected information matrix. We shall then use $(\mathring{\omega}_{rs})^{-1}$ to denote $1/\mathring{\omega}_{rs}$; for notational clarity, we note this is not the same as $\mathring{\omega}^{rs}$ which is an element of a matrix inverse. The ω notation is extended as follows,

$$\mathring{\omega}_{rst}(\theta) = E(\mathring{\ell}_{rst}(\theta)), \quad \mathring{\omega}_{r,st} = E(\mathring{\ell}_r\mathring{\ell}_{st}), \quad \mathring{\omega}_{r,s,t} = E(\mathring{\ell}_r\mathring{\ell}_s\mathring{\ell}_t)$$

Further, we define

$$\mathring{m}^r = \hat{\mathring{j}}^{r1}/\hat{\mathring{j}}^{11}, \quad \mathring{k}^{rs} = \hat{\mathring{j}}^{rs} - (\hat{\mathring{j}}^{r1}\hat{\mathring{j}}^{s1}/\hat{\mathring{j}}^{11})$$

and the corresponding ”expected” quantities are

$$\mathring{\tau}^{rs} = \mathring{\omega}^{r1}\mathring{\omega}^{s1}/\mathring{\omega}^{11}, \quad \mathring{\sigma}^{rs} = \mathring{\omega}^{rs} - \mathring{\tau}^{rs}.$$

We also use the summation convention of Einstein, so that we sum over all repeated subscripts or superscripts ranging from 1 to d, for example,

$$\mathring{\ell}_{rs}\hat{\mathring{j}}^r\hat{\mathring{j}}^s = \sum_{r=1}^{d}\sum_{s=1}^{d}\mathring{\ell}_{rs}\hat{\mathring{j}}^r\hat{\mathring{j}}^s$$

We are now ready to derive the matching conditions via the shrinkage argument. Again, we follow the three step argument above.

Step 1 The first step is to get an expectation with respect to the posterior density given Y. Consider $P_{\overline{\pi}}\{\psi \leq \psi^{1-\alpha}(\pi, Y)|Y\}$. Let z be the $(1-\alpha)$th quantile of the univariate standard normal distribution and let $\mathring{j}_{11} = -\partial^2\ell(\theta)/\partial\theta_1\partial\theta_1$ be the observed information component corresponding to ψ. Correspondingly $\hat{\mathring{j}}^{11}$ is the first component of the inverse of the observed information matrix. We consider an expansion for the posterior density of $z = (n/\hat{\mathring{j}}^{11})^{1/2}(\psi - \hat{\psi})$ under the prior $\overline{\pi}$ which is given by[4]

$$\begin{aligned} Pr_{\overline{\pi}|Y}\{\psi \leq \psi^{1-\alpha}(\pi, Y)|Y\} = {} & \Phi(z) + n^{-1/2}\phi(z)\{\beta_1(\alpha, \pi, Y) - G_1(\overline{\pi}) - G_3 H_2(z)\} \\ & + n^{-1}\phi(z)\big[\beta_2(\alpha, \pi, Y) - 2z\beta_1(\alpha, \pi, Y)G_3 - \frac{1}{2}\{\beta_1(\alpha, \pi, Y)\}^2 z \\ & + \beta_1(\alpha, \pi, Y)z\{G_1(\overline{\pi}) + G_3 H_2(z)\} - G_2(\overline{\pi})H_1(z) \\ & - G_4(\overline{\pi})H_3(z) - G_6 H_5(z)\big] + O(n^{-3/2}) \end{aligned}$$

where the $H_i(\cdot)$ are the standard Hermite polynomials defined by

$$\frac{d^i}{dz^i}\phi(z) = (-1)^i H_i(z)\phi(z)$$

where $\phi(\cdot)$ is the univariate standard normal density, which yields

$$H_1(z) = z, H_2(z) = z^2 - 1, H_3(z) = z^3 - 3z, H_4(z) = z^4 - 6z^2 + 3,$$

$$H_5(z) = z^5 - 10z^3 + 15z, H_6(z) = z^6 - 15z^4 + 45z^2 - 15$$

and

$$\begin{gathered} A_1(\pi, Y) = A_{11}(\pi, Y) + A_{12}(Y), A_{11}(\pi, Y) = \hat{\pi}_j \mathring{m}^j / \hat{\pi}, \\ A_{12}(Y) = \frac{1}{2}\mathring{\ell}_{jrs}(\hat{\theta})\mathring{k}^{jr}\mathring{m}^s, A_3(Y) = \frac{1}{6}\mathring{\ell}_{jrs}(\hat{\theta})\mathring{m}^j\mathring{m}^r\mathring{m}^s, \\ A_2(\pi, Y) = A_{21}(\pi, Y) + A_{22}(Y), A_4(\pi, Y) = A_{41}(\pi, Y) + A_{42}(Y), \\ A_{21}(\pi, Y) = \frac{1}{2}(\mathring{m}^j\mathring{m}^r/\hat{\pi})(\hat{\pi}_{jr} + \mathring{\ell}_{jrs}(\hat{\theta})\mathring{k}^{su}\hat{\pi}_u) + A_{11}(\pi, Y)A_{12}(Y), \\ A_{41}(\pi, Y) = A_{11}(\pi, Y)A_3(Y) \end{gathered}$$

Note that $A_{12}(Y)$, $A_3(Y)$, $A_{22}(Y)$, $A_{42}(Y)$ and $A_6(Y)$ are of order $O(1)$ and do not depend

[4]This is the expansion from Johnson (1970).

on $\pi(\cdot)$ or its derivatives.

Also, we have

$$G_1(\pi) \equiv G_1(\pi, Y) = A_1(\pi, Y)(\hat{\mathring{j}}^{11})^{1/2} + 3A_3(Y)(\hat{\mathring{j}}^{11})^{3/2},$$

$$G_3 \equiv G_3(Y) = A_3(Y)(\hat{\mathring{j}}^{11})^{3/2},$$

$$\begin{aligned} G_2(\pi) &\equiv G_2(\pi, Y) \\ &= A_2(\pi, Y)\hat{\mathring{j}}^{11} + 6A_4(\pi, Y)(\hat{\mathring{j}}^{11})^2 + 45A_6(Y)(\hat{\mathring{j}}^{11})^3, \end{aligned}$$

$$G_4(\pi) \equiv G_4(\pi, Y) = A_4(\pi, Y)(\hat{\mathring{j}}^{11})^2 + 15A_6(Y)(\hat{\mathring{j}}^{11})^3,$$

$$G_6 \equiv G_6(Y) = A_6(Y)(\hat{\mathring{j}}^{11})^3,$$

and finally,

$$\beta_1(\alpha, \pi, Y) = G_1(\pi) + G_3H_2(z)$$

and

$$\begin{aligned} \beta_2(\alpha, \pi, Y) =& 2z\beta_1(\alpha, \pi, Y)G_3 - \frac{1}{2}\{\beta_1(\alpha, \pi, Y)\}^2 z \\ &+ G_2(\pi)H_1(z) + G_4(\pi)H_3(z) + G_6H_5(z) \end{aligned}$$

This posterior expansion is a uniformly integrable version of the expansion of Johnson (1970) and is derived by Ghosh et al. (1985).

Step 2 Now we consider the expectation with respect to the conditional frequentist distribution of Y. We first note that the expansion given above contains observed likelihood derivatives of order 2 and 3, denoted by $\hat{\mathring{j}}$ and $\mathring{\ell}_{rst}$, respectively. To calculate the conditional expectation, we will need that $E_{\theta|A}(\hat{\theta}) = \theta + o(n^{-1/2})$, and that an expansion of the relevant likelihood quantities around θ yields that $E_{\theta|A}(\hat{\mathring{j}}_{rs}) = \mathring{\omega}_{rs} + o(n^{-1/2})$ and

$E_{\theta|A}(\hat{\mathring{\ell}}_{rst}) = \mathring{\omega}_{rst} + o(n^{-1/2})$. These results are the conditional versions of their unconditional counterparts and their validity requires conditional versions of the central limit theorem and laws of large numbers to hold. In particular, once we condition on an ancillary statistic, we violate the independent distribution assumption of the usual laws of large numbers. This dependence could pose a problem. However, results such as those in Andersen (1970) regarding consistency and asymptotic normality of the conditional maximum likelihood estimator in regular cases, as well as the results of Zabell (1980) and others regarding the convergence of conditional expectations suggest that it is reasonable to apply these first-order asymptotic results in the conditional setting for regular models.

We make use of the above relationships to find that

$$\begin{aligned} G_1(\pi) - G_1(\overline{\pi}) =& \{A_{11}(\pi, Y) - A_{11}(\overline{\pi}, Y)\}(\hat{\mathring{j}}^{11})^{1/2} \\ =& \Big(\frac{\partial\pi(\hat{\theta})/\partial\theta_r}{\pi(\hat{\theta})} - \frac{\partial\overline{\pi}(\hat{\theta})/\partial\theta_r}{\overline{\pi}(\hat{\theta})}\Big)\mathring{m}^j(\hat{\mathring{j}}^{11})^{1/2} \end{aligned}$$

Also, we find that

$$G_4(\pi) - G_4(\overline{\pi}) = \{G_1(\pi) - G_1(\overline{\pi})\}G_3$$

Moreover, we find that the quantity

$$G_2(\pi) - G_2(\overline{\pi}) - \{G_1(\pi) - G_1(\overline{\pi})\}\{G_1(\pi) + 2G_3\}$$

can be expressed as

$$\begin{aligned} &\hat{\mathring{j}}^{11}[A_{21}(\pi, Y) - A_{21}(\overline{\pi}, Y)] + 6(\hat{\mathring{j}}^{11})^2 A_3(Y)[A_{11}(\pi, Y) - A_{11}(\overline{\pi}, Y)] \\ &- \hat{\mathring{j}}^{11}[A_1(\pi, Y) + 5A_3(Y)\hat{\mathring{j}}^{11}][A_{11}(\pi, Y) - A_{11}(\overline{\pi}, Y)], \end{aligned}$$

which is equal to

$$\frac{1}{2}\hat{\mathring{j}}^{11}\mathring{m}^r\mathring{m}^s\{\big(\frac{\partial^2\pi(\hat{\theta})}{\partial\theta_r\partial\theta_s}\frac{1}{\pi(\hat{\theta})}-\frac{\partial^2\overline{\pi}(\hat{\theta})}{\partial\theta_r\partial\theta_s}\frac{1}{\overline{\pi}(\hat{\theta})}\big)+\hat{\mathring{\ell}}_{rst}\mathring{k}^{tu}\big(\frac{\partial\pi(\hat{\theta})}{\partial\theta_u}\frac{1}{\pi(\hat{\theta})}-\frac{\partial\overline{\pi}(\hat{\theta})}{\partial\theta_u}\frac{1}{\overline{\pi}(\hat{\theta})}\big)\}$$
$$+\hat{\mathring{j}}^{11}\{\mathring{m}^u\big(\frac{\partial\pi(\hat{\theta})}{\partial\theta_u}\frac{1}{\pi(\hat{\theta})}-\frac{\partial\overline{\pi}(\hat{\theta})}{\partial\theta_u}\frac{1}{\overline{\pi}(\hat{\theta})}\big)\}\big(\frac{1}{6}\hat{\mathring{\ell}}_{rst}\hat{\mathring{j}}^{11}\mathring{m}^r\mathring{m}^s\mathring{m}^t-\frac{\mathring{m}^r}{\pi(\hat{\theta})}\frac{\partial\pi(\hat{\theta})}{\partial\theta_r}\big)$$

Now, for notational convenience we define the quantity,

$$J(\pi,\theta)=\frac{1}{6}\mathring{\omega}_{rst}\mathring{\tau}^{rs}\mathring{\omega}^{t1}\pi(\theta)-\mathring{\omega}^{r1}\frac{\partial\pi(\theta)}{\partial\theta_r}$$

Taking expectations and implementing our expansion, we have that, for θ on the interior of the support of $\overline{\pi}(\cdot)$, the quantity $E_\theta[Pr_{\overline{\pi}|Y}\{\psi\le\psi^{(1-\alpha)}(\pi,Y)|Y,A\}]$ is given by the expansion

$$\begin{aligned}&1-\alpha+n^{-1/2}\phi(z)\Big\{\frac{\partial\pi(\theta)}{\partial\theta_r}\frac{1}{\pi(\theta)}-\frac{\partial\overline{\pi}(\theta)}{\partial\theta_r}\frac{1}{\overline{\pi}(\theta)}\Big\}\mathring{\omega}^{r1}(\mathring{\omega}^{11})^{-1/2}\\&+n^{-1}z\phi(z)\Big\{\frac{1}{2}\mathring{\tau}^{rs}\Big[\frac{\partial^2\pi(\theta)}{\partial\theta_r\partial\theta_s}\frac{1}{\pi(\theta)}-\frac{\partial^2\overline{\pi}(\theta)}{\partial\theta_r\partial\theta_s}\frac{1}{\overline{\pi}(\theta)}+\mathring{\omega}_{rst}\mathring{\omega}^{tu}\Big(\frac{\partial\pi(\theta)}{\partial\theta_u}\frac{1}{\pi(\theta)}-\frac{\partial\overline{\pi}(\theta)}{\partial\theta_u}\frac{1}{\overline{\pi}(\theta)}\Big)\Big]\\&+\Big[\Big(\frac{\partial\pi(\theta)}{\partial\theta_u}\frac{1}{\pi(\theta)}-\frac{\partial\overline{\pi}(\theta)}{\partial\theta_u}\frac{1}{\overline{\pi}(\theta)}\Big)\frac{\mathring{\omega}^{u1}}{\mathring{\omega}^{11}}\Big]\frac{J(\pi,\theta)}{\pi(\theta)}\Big\}+O(n^{-3/2})\end{aligned}\tag{2.1}$$

Step 3 The "unsmoothing" step, which is to take the expectation with respect to the auxiliary prior, is accomplished as follows. We suppose that the true parameter value θ is an interior point in the support of the auxiliary prior $\overline{\pi}(\cdot)$. By the regularity conditions, both the density $\overline{\pi}$ and its partial derivatives vanish on the boundary of its support. Now, to obtain the desired quantity, we must integrate equation (2.1) by parts with respect to the auxiliary prior $\overline{\pi}$. This will yield the following conditional frequentist tail probability expansion,

$$Pr_{\theta|A}\{\psi \leq \psi^{(1-\alpha)}(\pi, Y)|A\} = 1 - \alpha + n^{-1/2}\frac{\phi(z)}{\pi(\theta)}\mathring{\Delta}_1(\pi, \theta) + n^{-1}z\frac{\phi(z)}{\pi(\theta)}\mathring{\Delta}_2(\pi, \theta) + O(n^{-3/2}) \tag{2.2}$$

The coefficients $\mathring{\Delta}_1$ and $\mathring{\Delta}_2$ form the matching conditions. Quite simply, when $\mathring{\Delta}_1 = 0$, we have that the conditional frequentist coverage of the posterior credible set is $1 - \alpha + O(n^{-1})$ and when both $\mathring{\Delta}_1, \mathring{\Delta}_2 = 0$, the coverage is $1 - \alpha$ to order of error $O(n^{-3/2})$ which is the definition of 3rd order probability matching.

This integration necessary to complete this step is possible due to results from Ghosh et al. (1985) and regularity conditions due to Bickel & Ghosh (1990) concerning the uniformity of the probability measure over the support of a set $\overline{\Theta}$ over which $\hat{\theta}$ is well-defined. This is the set over which posterior expansions are valid, and the assumption is made that this set has a uniform probability measure over compact subsets in the interior of the support of $\overline{\pi}(\cdot)$.[5] Integration by parts of the expansion at the end of Step 2 yields that the coefficients are equal to

$$\mathring{\Delta}_1 = \frac{\partial}{\partial\theta_r}\{\pi(\theta)\mathring{\omega}^{r1}(\mathring{\omega}^{11})^{-1/2}\} \tag{2.3}$$

and after considerable algebraic simplification through repeated use of the relationships the quantities in the expansion as detailed above, we find that

$$\mathring{\Delta}_2 = \frac{1}{3}\frac{\partial}{\partial\theta_u}\{\pi(\theta)\mathring{\tau}^{rs}\mathring{\omega}_{rst}(3\mathring{\omega}^{tu} + \mathring{\tau}^{tu})\} - \frac{\partial^2}{\partial\theta_r\partial\theta_s}\{\pi(\theta)\mathring{\tau}^{rs}\} \tag{2.4}$$

The matching conditions are then given as

$$\mathring{\Delta}_1 = 0 \tag{2.5}$$

[5] Sweeting (1991) also raises the point that an equicontinuity condition must be satisfied for the "unsmoothing" step to be valid, as we are moving to a sampling distribution from the starting point of a mixture distribution. Sweeting (1995*a*) provides more details on this. We will be working in settings for which all of these regularity conditions are satisfied.

by which we mean that any prior satisfying this condition is 2nd order conditional probability matching. The proof that such priors are 2nd order probability matching is a simple modification of the proof in the unconditional setting due to Peers (1965). The 3rd order condition, which was proved in the unconditional setting by Mukerjee & Ghosh (1997), is that in addition to the 2nd order condition, a 3rd order conditional PMP must also satisfy

$$\mathring{\Delta}_2 = 0. \tag{2.6}$$

These conditions turn out to be identical to those derived Mukerjee & Dey (1993) and Mukerjee & Ghosh (1997) for the unconditional setting, but simply replacing the unconditional quantities with the corresponding conditional ones. As we argue in Chapter 3, however, we can use the relationships between conditional and unconditional likelihood quantities to arrive at this conclusion via a much simpler route. The direct and long-winded calculation has been included for completeness of the exposition of the shrinkage argument and to confirm that all of the steps are valid in the conditional setting.

2.2 Method 2: The Mean- and Variance-Adjusted R Statistic

For the second matching method, we work with mean- and variance-adjusted versions of the signed root likelihood ratio statistic. Recall that signed-root likelihood ratio statistic is defined as

$$R(\psi) = sgn(\hat{\psi} - \psi)W(\psi)^{1/2},$$

where $W(\psi) = 2\left\{\ell(\hat{\theta}) - \ell(\psi, \hat{\lambda}_\psi)\right\}$ is the likelihood ratio statistic, $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$ denotes the global maximum likelihood estimator and $\hat{\lambda}_\psi$ is the constrained maximum likelihood estimator of λ for a fixed value of ψ. A standard normal approximation to the distribution of this test statistic has error of order $O(n^{-1/2})$. We shall require that the maximum likelihood estimator is the unique local maximum, and to that end we also note that $R(\psi)$ is a monotonic decreasing function of ψ.

We first note the following distributional results. The mean-adjusted signed root statistic, $R(\psi) - \mu(\psi)$, where $\mu(\psi)$ is the mean of $R(\psi)$, is standard normal to order $O(n^{-1})$. If we denote by σ^2 the variance of $R(\psi) - \mu(\psi)$, then $(R(\psi) - \mu(\psi))/\sigma$ is standard normal to order $O(n^{3/2})$.[6] This holds for both the frequentist and Bayesian posterior versions. Thus, denote the posterior mean of $R(\psi)$ as μ_B and the corresponding posterior variance of $R(\psi) - \mu_B$ as σ_B^2. Also the conditional frequentist mean is $\mathring{\mu}_F$. We use σ_F^2 and $\mathring{\sigma}_F^2$ to denote, respectively, the unconditional and conditional frequentist variance of $(R(\psi) - \mu_B)$, the signed root adjusted by the Bayesian mean. These Bayesian distributional results were derived in DiCiccio & Stern (1994*b*) and the corresponding frequentist distributional results can be found in DiCiccio & Stern (1994*a*).[7] In what follows, we will show that a condition ensuring 2nd order matching, after taking expectations (since the Bayesian quantity depends on observed rather than expected likelihood quantities), is given by

$$E_{Y|A}\{\mu_B\} = \mathring{\mu}_F + O(n^{-3/2})$$

and also we will find the 3rd order condition, after taking expectations,

$$E_{Y|A}\{\sigma_B^2\} = \mathring{\sigma}_F^2 + O(n^{-3/2}),$$

which turns out to be satisfied when $\mathring{\sigma}_F^2 = \sigma_F^2 + O(n^{-3/2})$.

A proof that these conditions together ensure 3rd order matching is given at the end of the section. First, these quantities need to be derived, and we now proceed to do so.

As the method we are proposing is new to both the unconditional and conditional PMP literature, then for completeness we will proceed unconditionally first and then translate our arguments to the conditional setting.

2.2.1 Unconditional Matching to 2nd Order

First we derive a condition which ensures unconditional probability matching to 2nd order.

[6]This quantity σ^2 is not the same as the σ quantity defined before or in the foregoing; it will always be clear from the context what is meant by the use of σ. For variance terms, it will include one of the subscripts B or F.

[7]Another use mean- and variance-adjusted signed roots to obtain a statistic with standard normal distribution to higher-order of error is Jensen (1986).

We will find it more convenient here to work with the full log-likelihood, so that $\ell(\theta) = \sum_{i=1}^{n} f(Y_i;\theta)$, as opposed to the definition in the shrinkage argument that this quantity was the *average* log-likelihood and we divided by n. We ask the reader to make note that this change in definition will translate into differences in the orders of magnitude of the quantities derived herein. We comment along the way on these differences, to avoid additional confusion. We do not change notation for $\ell(\theta)$ but simply redefine the corresponding quantities as being based on the full-likelihood. Aside from this key difference, we broadly follow the notation DiCiccio & Stern (1994*b*), though making every effort to be consistent with the notation of Datta & Mukerjee (2004). We again use "$^\circ$" to denote conditional quantities, e.g. $\mathring{\ell}_{\theta\theta} = \ell_{\theta\theta}(\theta;\hat{\theta},a)$. Here we work with possibly vector-valued interest and nuisance parameters. Wherever possible, we have used notation consistent with the previous section, only changing the indexes to clarify that this is a more general setting. There will be some notational redundancies, but these are included to avoid confusion.

Now consider more generally the case of an observed random vector $Y = \{Y_1,\ldots,Y_n\}$ having some continuous probability distribution that depends on unknown d- dimensional parameter $\theta = (\theta_1,\ldots,\theta_d)$. Suppose that θ is partitioned in the form $\theta = (\psi,\lambda)$ where $\psi = \theta_1$ is a scalar interest parameter and $\lambda = (\theta_2,\ldots,\theta_d)$ is a $(d-1)$-dimensional nuisance parameter. Let $\tilde{\theta}(\psi) = (\psi,\hat{\lambda}_\psi)$ be the constrained MLE of θ for a given ψ. The log profile likelihood function for ψ is $\tilde{\ell}(\psi) = \ell(\tilde{\theta}(\psi))$.

We use standard conventions for denoting arrays and summation. Namely, we use the Einstein summation convention of summing over all repeated subscripts or superscripts. The indices $i,j,k,\ldots$ range over $2,\ldots,d$, and the indices $r,s,t,\ldots$ range over $1,\ldots,d$. Differentiation of the functions $\ell(\theta)$ and $\ell(\tilde{\theta}(\psi))$ is indicated by subscripts, so $\ell_r(\theta) = \partial\ell(\theta)/\partial\theta_r$, $\ell_{rs}(\theta) = \partial^2\ell(\theta)/\partial\theta_r\theta_s$, $\tilde{\ell}_1(\psi) = \partial\tilde{\ell}_\psi/\partial\psi$, $\tilde{\ell}_{11}(\psi) = \partial^2\tilde{\ell}(\psi)/\partial\psi\partial\psi$, etc. In this notation,

$$\ell_r(\hat{\theta}) = 0 \;\; (r = 1,\ldots,d)$$

and

$$\tilde{\ell}_1(\hat{\psi}) = 0.$$

Evaluation of the derivatives of $\ell(\theta)$ at $\hat{\theta}$ and the derivatives of $\tilde{\ell}(\psi)$ at $\hat{\psi}$ is indicated by placing a "$\hat{\cdot}$" above the appropriate quantity. For example, $\hat{\ell}_r = \ell_r(\hat{\theta}) = 0$, $\hat{\ell}_{rs} = \ell_{rs}(\hat{\theta})$, $\hat{\tilde{\ell}}_1 = \tilde{\ell}_1(\hat{\theta})$, $\hat{\tilde{\ell}}_{11} = \tilde{\ell}_{11}(\hat{\theta})$, etc. The unconditional expected quantities are defined analogously to the conditional ones.[8] For instance, $\omega_r = E\{\ell_r(\theta)\} = 0$, $\omega_{rs} = E\{\ell_{rs}(\theta)\}$, $\omega_{rst} = E\{\ell_{rst}(\theta)\}$, etc. Further, we define $l_r = \ell_r(\theta) - \omega_r = \ell_r(\theta)$, $l_{rs} = \ell_{rs}(\theta) - \omega_{rs}$, $l_{rst} = \ell_{rst}(\theta) - \omega_{rst}$, etc. The constants ω_{rs}, ω_{rst}, etc. are assumed to be of order $O(n)$[9]; the variables l_r, l_{rs}, l_{rst}, etc. have expectation 0, and they are assumed to be of order $O_p(n^{1/2})$. The joint cumulants of l_r, l_{rs}, etc. are of order $O(n)$ provided that their expectations exist.

In subsequent calculations, it will be useful to extend the ω-notation: let

$$\omega_{r,s} = E\{\ell_r(\theta)\ell_s(\theta)\}, \quad \omega_{rs,t} = E\{\ell_{rs}(\theta)\ell_t(\theta)\}, \quad \omega_{r,s,t} = E\{\ell_r(\theta)\ell_s(\theta)\ell_t(\theta)\},$$

etc. Identities involving the ω's can be derived by repeated differentiation of the identity $\int \exp\{\ell(\theta)\}dy = 1$, which produces the Bartlett identities:

$$\omega_r = 0, \quad \omega_{rs} + \omega_{r,s} = 0, \quad \omega_{rst} + \omega_{rs,t}[3] + \omega_{r,s,t} = 0.$$

The bracket notation "$[k]$" is used to indicate summation over all k possible terms obtained by permutating the indices, i.e., $\omega_{rs,t}[3] = \omega_{rs,t} + \omega_{rt,s} + \omega_{st,r}$. Differentiation of the definition $\omega_{rs} = \int \ell_{rs}(\theta)\exp\{\ell(\theta)\}dy$ yields

$$\omega_{rs/t} = \omega_{rst} + \omega_{rs,t},$$

where $\omega_{rs/t} = \partial\omega_{rs}/\partial\theta_t$.

Let (ω^{rs}) and $(\hat{\ell}^{rs})$ be the $d \times d$ matrix inverses of (ω_{rs}) and $(\hat{\ell}_{rs})$, and let ν_{11} and $\hat{S}_{11}$ be the inverses of ω^{11} and $\hat{\ell}^{11}$. Define

$$\tau^{rs} = \omega^{r1}\omega^{s1}\nu_{11},$$

[8]It should not alarm the reader that here we do not take the negative expectations; we will account for this in the ensuing calculations.

[9]Contrast this to the setting of the average log-likelihood or log-likelihood for a single observation, where these quantities are $O(1)$. Similar comparisons can be made for the other quantities mentioned here.

$$\hat{T}^{rs} = \hat{\ell}^{r1}\hat{\ell}^{s1}\hat{S}_{11},$$

$$\sigma^{rs} = \omega^{rs} - \tau^{rs},$$

and

$$\hat{V}^{rs} = \hat{\ell}^{rs} - \hat{T}^{rs}.$$

Note that ω^{rs}, τ^{rs}, and σ^{rs} are all of order $O(n^{-1})$, and $\hat{\ell}^{rs}$, $\hat{T}^{rs}$, and $\hat{V}^{rs}$ are all of order $O_p(n^{-1})$. Furthermore, $\nu_{11} = \omega_{1r}\omega_{1s}\tau^{rs}$ is of order $O(n)$ and $\hat{S}_{11} = \hat{\ell}_{1r}\hat{\ell}_{1s}\hat{T}^{rs}$ is of order $O_p(n)$. It is readily seen that $\tau^{r1} = \omega^{r1}$ and $\sigma^{r1} = 0$; thus, the entries of $d \times d$ matrices (σ^{rs}) and $\hat{V}^{rs}$ are all zero except for the lower right-hand submatrices (σ^{ij}) and $(\hat{V}^{ij})$, which are the inverses of (ω_{ij}) and $(\hat{\ell}_{ij})$, respectively.

DiCiccio & Stern (1994*a*) showed that

$$\mu_F = -\frac{1}{2}\eta\omega_{rst}\omega^{r1}\omega^{st} - \frac{1}{6}\eta^3\omega_{rst}\omega^{r1}\omega^{s1}\omega^{t1} + \eta\omega_{rs/t}\omega^{r1}\omega^{st} + \frac{1}{2}\eta^3\omega_{rs/t}\omega^{r1}\omega^{s1}\omega^{t1} + O(n^{-3/2}), \tag{2.7}$$

where $\eta = (-\omega^{11})^{-1/2}$. We may derive a similar expression for μ_B in the following way.

Repeated differentiation of the definition $\tilde{\ell}(\psi) = \ell\{\tilde{\theta}(\psi)\}$ with respect to the interest parameters yields

$$\tilde{\ell}_1(\hat{\psi}) = 0,$$

$$\tilde{\ell}_{11}(\hat{\psi}) = \hat{S}_{11},$$

$$\tilde{\ell}_{111}(\hat{\psi}) = \hat{S}_{11}\hat{S}_{11}\hat{S}_{11}\hat{\ell}^{1r}\hat{\ell}^{1s}\hat{\ell}^{1t}\hat{\ell}_{rst}.$$

These derivatives require the identity $\tilde{\theta}(\hat{\psi}) = \hat{\theta}$. Also, repeated differentiation of the identity $\ell_i\{\tilde{\theta}(\psi)\} = 0$ yields $\tilde{\theta}^r_{/1}(\psi) = \ell^{r1}\{\tilde{\theta}(\psi)\}S_{11}\{\tilde{\theta}(\psi)\}$, where $\ell_{ir}\{\tilde{\theta}(\psi)\}\tilde{\theta}^r_{/1}(\psi) = 0$, and $[S_{11}\{\tilde{\theta}(\psi)\}]$ is the matrix inverse of $[\ell^{11}\{\tilde{\theta}(\psi)\}]$. In particular, $\tilde{\theta}^r_{/1}(\hat{\theta}) = \hat{\ell}^{r1}\hat{S}_{11}$ and $\hat{\ell}_{ir}\tilde{\theta}^r_{/1}(\hat{\psi}) = 0$.

Taylor expansion of the likelihood ratio statistic, $W(\psi)$, about $\hat{\psi}$ yields

$$W(\psi) = -\hat{\tilde{\ell}}_{11}(\hat{\psi} - \psi)^2 + \frac{1}{3}\hat{\tilde{\ell}}_{111}(\hat{\psi} - \psi)^3 + O_p(n^{-1}),$$

where $\hat{\hat{\ell}}_{11} = 1/\hat{\ell}^{11}$ and $\hat{\hat{\ell}}_{111} = \hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1}/(\hat{\ell}^{11})^3$. Consequently,

$$W(\psi) = \{V(\psi)\}^2 - \frac{1}{3}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1}\{V(\psi)\}^3 + O_p(n^{-1}),$$

where $V(\psi) = (-\hat{\hat{\ell}}_{11})^{1/2}(\hat{\psi} - \psi) = \hat{D}(\hat{\psi} - \psi)$ and $\hat{D} = (-\hat{\hat{\ell}}_{11})^{1/2} = (-\hat{\ell}^{11})^{-1/2}$. Thus $V(\psi)$ is the usual studentized statistic, where the standardization is by $\hat{D}$, the *observed* analogue of $\eta = (-\omega^{11})^{1/2}$. The signed root statistic $R(\psi)$ has the expansion

$$R(\psi) = V(\psi) - \frac{1}{6}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1}\{V(\psi)\}^2 + O_p(n^{-1}).$$

The Laplace approximation to the marginal posterior density function of ψ given by Tierney & Kadane (1986) can be written as

$$\pi_{\psi|Y}(\psi) = c\exp\{B(\psi) + \tilde{\ell}(\psi)\}\{1 + O_p(n^{-2})\}$$

for values of the argument ψ such that $\psi = \hat{\psi} + O_p(n^{-1})$, where c is a normalizing constant, and

$$B(\psi) = -\frac{1}{2}\log\left\{\frac{|-\ell_{ij}(\psi, \hat{\lambda}_\psi)|}{|-\ell_{ij}(\hat{\psi}, \hat{\lambda})|}\right\} + \log\left\{\frac{\pi(\psi, \hat{\lambda}_\psi)}{\pi(\hat{\psi}, \hat{\lambda})}\right\}.$$

Differentiation of $B(\psi)$ yields

$$\hat{B}_1 = \frac{1}{2}\hat{D}^2\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{rs} + \frac{1}{2}\hat{D}^4\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1} - \hat{D}^2\left.\frac{\partial \log \pi(\theta)}{\partial \theta_r}\right|_{\theta=\hat{\theta}}\hat{\ell}^{r1},$$

which is of order $O(1)$.

By Taylor expansion about $\hat{\psi}$,

$$\pi_{\psi|Y}(\psi) = (2\pi)^{-1/2}\hat{D}\exp\{-\frac{1}{2}\hat{D}^2(\hat{\psi}-\psi)^2\}\{1-\hat{B}_1(\hat{\psi}-\psi)-\frac{1}{6}\hat{\hat{\ell}}_{111}(\hat{\psi}-\psi)^3+O_p(n^{-1})\},$$

so the marginal posterior density of the studentized statistic $V(\psi)$ has the expansion

$$\pi_{V(\psi)|Y}(v) = (2\pi)^{-1/2}e^{-v^2/2}\{1 - \hat{D}^{-1}\hat{B}_1 v - \frac{1}{6}\hat{D}^{-3}\hat{\hat{\ell}}_{111}v^3 + O_p(n^{-1})\},$$

from which it follows that

$$\begin{aligned}
\mu_B &= -\hat{D}^{-1}\hat{B}_1 - \frac{1}{2}\hat{D}^{-3}\hat{\hat{\ell}}_{111} - \frac{1}{6}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1} + O(n^{-1}) \\
&= -\frac{1}{2}\hat{D}\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{rs} - \frac{1}{2}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1} + \hat{D}\frac{\partial \log \pi(\theta)}{\partial \theta_r}\Big|_{\theta=\hat{\theta}}\hat{\ell}^{r1} \\
&\quad + \frac{1}{2}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1} - \frac{1}{6}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1} + O(n^{-1}) \\
&= -\frac{1}{2}\hat{D}\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{rs} - \frac{1}{6}\hat{D}^3\hat{\ell}_{rst}\hat{\ell}^{r1}\hat{\ell}^{s1}\hat{\ell}^{t1} + \hat{D}\frac{\partial \log \pi(\theta)}{\partial \theta_r}\Big|_{\theta=\hat{\theta}}\hat{\ell}^{r1} + O_p(n^{-1}).
\end{aligned} \tag{2.8}$$

A more careful analysis, expanding the $O(n^{-1})$ term shows that the error in the preceding formula is actually $O_p(n^{-3/2})$. To be specific, a higher-order expansion of $R(\psi)$ would involve a term in $\{V(\psi)\}^3$ with coefficient that is of order $O(n^{-1})$. Moreover, a higher-order expansion of the marginal posterior density of $V(\psi)$ would involve terms in v^4 and v^6, each having coefficients of order $O(n^{-1})$. Now, when calculating the contribution to the expectation from the $\{V(\psi)\}$ term in $R(\psi)$, the v^4 and v^6 terms in the marginal posterior density of $V(\psi)$ would yield 0. When calculating the contribution to the expectation from the $\{V(\psi)\}^2$ term in $R(\psi)$, which has coefficient of order $O(n^{-1/2})$, the v^4 and v^6 terms in the marginal posterior density of $V(\psi)$, which have coefficients of order $O(n^{-1})$, would yield a term of order $O(n^{-3/2})$. When calculating the contribution to the expectation from the $\{V(\psi)\}^3$ term in $R(\psi)$, which has coefficient of order $O(n^{-1})$, the v and v^3 terms in the marginal posterior density of $V(\psi)$, which have coefficients of order $O(n^{-1/2})$, would yield terms of order $O(n^{-3/2})$, while the v^4 and v^6 terms in the marginal posterior density of $V(\psi)$ would yield 0. Similar calculations, to error of order $O(n^{-3/2})$, were given by DiCiccio & Stern (1993).

To compare μ_B and μ_F, note that

$$\mu_B = -\frac{1}{2}\eta\omega_{rst}\omega^{r1}\omega^{rs} - \frac{1}{6}\eta^3\omega_{rst}\omega^{r1}\omega^{s1}\omega^{t1} + \eta\frac{\partial \log \pi(\theta)}{\partial \theta_r}\omega^{r1} + O_p(n^{-1})$$

in the frequentist sense. Thus, the condition $\mu_B = \mu_F + O_p(n^{-1})$ is met when the prior satisfies

$$\eta \frac{\partial \log \pi(\theta)}{\partial \theta_r} \omega^{r1} = \eta \omega_{rs/t} \omega^{r1} \omega^{st} + \frac{1}{2} \eta^3 \omega_{rs/t} \omega^{r1} \omega^{s1} \omega^{t1}.$$

Since $(\omega^{uv})_{/t} = -\omega_{rs/t}\omega^{ru}\omega^{sv}$, it follows that

$$\sum_r \frac{\partial \omega^{r1}}{\partial \theta_r} = -\omega_{rs/t} \omega^{r1} \omega^{st},$$

$$\frac{\partial \eta}{\partial \theta_r} = -\frac{1}{2} \eta^3 \omega_{st/r} \omega^{s1} \omega^{t1};$$

consequently, the aforementioned condition on the prior can be written as

$$\eta \frac{\partial \log \pi(\theta)}{\partial \theta_r} \omega^{r1} = -\sum_r \frac{\partial (\eta \omega^{r1})}{\partial \theta_r},$$

which is the condition from Peers (1965) and which we will refer to as the Welch & Peers condition. Moreover, when the interest parameter is orthogonal to the nuisance paramete-terse, i.e. when $\omega^{i1} = 0$, then $\omega^{11} = 1/\omega_{11}$ and $\eta = (-\omega_{11})^{1/2}$, so the Welch & Peers condition reduces to

$$(-\omega_{11})^{-1/2} \frac{\partial \log \pi(\theta)}{\partial \psi} = -\frac{\partial (-\omega_{11})^{-1/2}}{\partial \psi},$$

which yields $\pi(\theta) \propto \sqrt{-\omega_{11}} g(\lambda)$ where $g(\lambda)$ is an arbitrary function of the nuisance parameter λ.

2.2.2 Higher-order analyses of the Bayesian mean

We now turn to a higher-order analysis of the Bayesian mean. For simplicity, we will write $\mu_B = \zeta(\hat{\theta}) + O_p(n^{-3/2})$. The new quantity, $\zeta(\theta)$ is defined by

$$\zeta(\theta) = -\frac{1}{2} D \ell_{rst} \ell^{r1} \ell^{rs} - \frac{1}{6} D^3 \ell_{rst} \ell^{r1} \ell^{s1} \ell^{t1} + D \frac{\partial \log \pi(\theta)}{\partial \theta_r} \ell^{r1},$$

where $D = (-\ell^{11})^{-1/2}$. Note that $\zeta(\theta)$ is $O_p(n^{-1/2})$. An implication of Ghosh & Mukerjee (1991), which was concerned with Bayesian and unconditional frequentist Bartlett corrections of the likelihood ratio statistic, $W(\psi)$, is that if the prior density $\pi(\theta)$ satisfies the Welch & Peers condition, then $\mu_B = \mu_F + O_p(n^{-1})$ and therefore in expectation $E(\mu_B) = \mu_F + O(n^{-1})$. This is not quite what we want, however. We are interested in a higher-order matching property. In establishing conditions under which $\pi(\theta)$ is third-order unconditional probability matching, it is necessary to show that $E(\mu_B) = \mu_F + O(n^{-3/2})$. We begin by noting that Taylor expansion of $\zeta(\hat{\theta})$ about θ yields

$$\begin{aligned}\zeta(\hat{\theta}) &= \zeta(\theta) + \zeta_{/r}(\theta)(\hat{\theta}_r - \theta_r) + O_p(n^{-3/2}) \\ &= \zeta(\theta) - \zeta_{/r}(\theta)\omega^{rs} l_s + O_p(n^{-3/2}),\end{aligned}$$

since $\hat{\theta}_r - \theta_r = \omega^{rs} l_s + O_p(n^{-1})$, where $l_s = \ell_s(\theta) - \omega_s$. Now, $\zeta_{/r}(\theta)$ is $O_p(n^{-1/2})$, and it is a function of the observed likelihood quantities denoted by ℓ's, just as $\zeta(\theta)$ is. However, in order to compare the Bayesian and frequentist means of $R(\psi)$, we will need to translate these observed quantities into expected quantities denoted by ω's.

Let $\zeta_r^{\omega}(\theta)$ be the quantity obtained when each of the ℓ's in $\zeta_r(\theta)$ is replaced by its corresponding ω, so $\zeta_r^{\omega}(\theta)$ is a nonrandom quantity depending on θ. Then $\zeta_r(\theta) = \zeta_r^{\omega}(\theta) + O_p(n^{-1})$, and it follows that

$$E\{\zeta_{/r}(\theta)\omega^{rs} l_s\} = E\{\zeta_r^{\omega}(\theta)\omega^{rs} l_s + O_p(n^{-3/2})\} = O(n^{-3/2})$$

It follows that

$$E(\mu_B) = E\{\zeta(\theta)\} + O(n^{-3/2}),$$

so it is required to show that $E\{\zeta(\theta)\} = \mu_F + O(n^{-3/2})$.

To investigate the expectation of $\zeta(\theta)$, recall that $\ell_{rs} = \omega_{rs} + l_{rs}$, where ω_{rs} is of order $O(n)$, l_{rs} is of order $O_p(n^{1/2})$ and $E(l_{rs}) = 0$. Thus,

$$\ell^{rs} = \omega^{rs} - \omega^{rt}\omega^{su} l_{tu} + O_p(n^{-2}),$$

from which it follows that

$$\ell^{r1} = \omega^{r1} - \omega^{rs}\omega^{t1} l_{st} + O_p(n^{-2}),$$

and

$$\ell^{11} = \omega^{11} - \omega^{r1}\omega^{s1} l_{rs} + O_p(n^{-2}).$$

The latter expression yields

$$D = (-\ell^{11})^{-1/2} = \eta - \frac{1}{2}\eta^3\omega^{r1}\omega^{s1} l_{rs} + O_p(n^{-1/2}),$$

and

$$D^3 = \eta^3 - \frac{3}{2}\eta^3\omega^{r1}\omega^{s1} l_{rs} + O_p(n^{-1/2})$$

Consider first the final term of $\zeta(\theta)$; the other terms can be handled similarly. It follows from the preceding equations that

$$\begin{aligned} D\frac{\partial \log \pi(\theta)}{\partial \theta_r}\ell^{r1} &= \eta\frac{\partial \log \pi(\theta)}{\partial \theta_r}\omega^{r1} - \frac{1}{2}\eta^3\frac{\partial \log \pi(\theta)}{\partial \theta_r}\omega^{r1}\omega^{s1}\omega^{t1} l_{st} \\ &\quad + \eta\frac{\partial \log \pi(\theta)}{\partial \theta_r}\omega^{rs}\omega^{t1} l_{st} + O_p(n^{-3/2}) \end{aligned}$$

Hence,

$$\begin{aligned} E\big\{D\frac{\partial \log \pi(\theta)}{\partial \theta_r}\ell^{r1}\big\} &= \eta\frac{\partial \log \pi(\theta)}{\partial \theta_r}\omega^{r1} + O(n^{-3/2}) \\ &= \eta\omega_{rs/t}\omega^{r1}\omega^{st} + \frac{1}{2}\eta^3\omega_{rs/t}\omega^{r1}\omega^{s1}\omega^{t1} + O(n^{-3/2}), \end{aligned}$$

by virtue of the condition ensuring $\mu_B = \mu_F + O_p(n^{-1})$ at the end of the previous section.[10] The other terms in $\zeta(\theta)$ have

$$E(-\frac{1}{2}D\ell_{rst}\ell^{r1}\ell^{rs}) = -\frac{1}{2}\eta\omega_{rst}\omega^{r1}\omega^{rs} + O(n^{-3/2}),$$

[10] Again, this is an implication of earlier results on Bartlett corrections of likelihood ratio statistics and the fact that the Welch & Peers prior matches the Bayesian and frequentist Bartlett correction terms to the order stated. Results such as this can be found in Ghosh & Mukerjee (1991) and others.

$$E(-\frac{1}{6}D^3\ell_{rst}\ell^{r1}\ell^{s1}\ell^{t1}) = -\frac{1}{6}\eta^3\omega_{rst}\omega^{r1}\omega^{s1}\omega^{t1} + O(n^{-3/2}),$$

and combining these expressions yields the desired result, namely,

$$E\{\zeta(\theta)\} = \mu_F + O(n^{-3/2}).$$

2.2.3 Conditional Matching to 2nd Order

When considering the conditional versions of these arguments, we make use of the fact that the conditional log-likelihood function $\mathring{\ell}(\theta)$ differs from the unconditional log-likelihood function $\ell(\theta)$ by a constant, i.e. a quantity that depends on A but not on θ. Therefore, $\mathring{\ell}_r = \ell_r$, $\mathring{\ell}_{rs} = \ell_{rs}$, etc. Thus, $\mathring{\omega}_r = E_{Y|A}(\ell_r)$, $\mathring{\omega}_{rs} = E_{Y|A}(\ell_{rs})$, etc. Note that $\mathring{\omega}_r = 0$. The quantities $\mathring{\omega}_{rs}$, $\mathring{\omega}_{rst}$, etc. are actually random variables depending on A, and they are of order $O_p(n)$ provided that their expectations exist. The variables $\mathring{l}_r = \ell_r$, $\mathring{l}_{rs} = \ell_{rs} - \mathring{\omega}_{rs}$, etc. all have conditional expectation 0, and hence they have unconditional expectation 0, and they are assumed to be of order $O_p(n^{1/2})$. The joint conditional cumulants of $\mathring{l}_r$, $\mathring{l}_{rs}$, etc. depend on A, and, provided that their expectations exist, they are of order $O_p(n)$ given A and order $O(n)$ unconditionally. The identities that hold for the ω's then immediately carry over to the $\mathring{\omega}$'s.

In the calculations that follow, it is necessary to take into account the differences between the ω's and the $\mathring{\omega}$'s. To describe the difference between ω_{rs} and $\mathring{\omega}_{rs}$, first note that

$$E_A(\mathring{\omega}_{rs}) = E_A\{E_{Y|A}(\ell_{rs})\} = E_Y(\ell_{rs}) = \omega_{rs}$$

Moreover,

$$\begin{aligned}
V_A(\mathring{\omega}_{rs}) &= V_A\{E_{Y|A}(\ell_{rs})\} \\
&= V_Y(\ell_{rs}) - E_A\{V_{Y|A}(\ell_{rs})\} \\
&= C_Y(l_{rs}, l_{rs}) - E_A\{C_{Y|A}(\mathring{l}_{rs}, \mathring{l}_{rs})\} \\
&= O(n) - E_A\{O_p(n)\} \\
&= O(n)
\end{aligned}$$

where E, V and C denote expectation, variance and cumulant, respectively. (Recall that the first and second cumulants of a random variable X are its mean and variance, that is, $E(X) = C(X)$ and $V(X) = C(X, X)$.) Since, with respect to the distribution of A, $\mathring{\omega}_{rs}$ has mean ω_{rs} and variance of order $O(n)$, it follows that $\mathring{\omega}_{rs} = \omega_{rs} + O_p(n^{1/2})$.[11] Identical arguments apply for ℓ_{rst}, etc., so that $\mathring{\omega}_{rst} = \omega_{rst} + O_p(n^{1/2})$, etc.

Differentiation of the identity $\mathring{\omega}_{rs} = \omega_{rs} + O_p(n^{1/2})$ yields $\mathring{\omega}_{rs/t} = \omega_{rs/t} + O_p(n^{1/2})$. Hence,

$$\mathring{\omega}_{rs,t} = (\mathring{\omega}_{rs})_{/t} - \mathring{\omega}_{rst} = (\omega_{rs})_{/t} - \omega_{rst} + O_p(n^{1/2}) = \omega_{rs,t} + O_p(n^{1/2}).$$

By working with the conditional density of Y given A in place of the marginal density of Y, it follows that the conditional distribution of the signed root $R(\psi)$ is standard normal to error of order $O(n^{-1/2})$ and that the error in the standard normal approximation to the conditional distribution of $R(\psi)$ can be reduced to order $O(n^{-1})$ by adjusting for the conditional mean of $R(\psi)$. Denote the conditional mean by $\mathring{\mu}_F$; then $R(\psi) - \mathring{\mu}_F$ is standard normal to error of order $O(n^{-1})$.

The calculations of DiCiccio & Stern (1994*b*) can be applied to the conditional distribution of $R(\psi)$ to show that

[11] As we note later in the thesis, had we standardized the log-likelihood by n, this difference is of order $O_p(n^{-1/2})$. We again note this to avoid confusion about why this differs from other results in the thesis.

$$\mathring{\mu}_F = -\frac{1}{2}\mathring{\eta}\mathring{\omega}_{rst}\mathring{\omega}^{r1}\mathring{\omega}^{st} - \frac{1}{6}\mathring{\eta}^3\mathring{\omega}_{rst}\mathring{\omega}^{r1}\mathring{\omega}^{s1}\mathring{\omega}^{t1} + \mathring{\eta}\mathring{\omega}_{rs/t}\mathring{\omega}^{r1}\mathring{\omega}^{st} + \frac{1}{2}\mathring{\eta}^3\mathring{\omega}_{rs/t}\mathring{\omega}^{r1}\mathring{\omega}^{s1}\mathring{\omega}^{t1} + O_p(n^{-3/2}), \tag{2.9}$$

where $\mathring{\eta} = (-\mathring{\omega}^{11})^{-1/2}$. Furthermore, the preceding comparisons of the ω's and the corresponding $\mathring{\omega}$'s shows that $\mathring{\mu}_F = \mu_F + O_p(n^{-1})$, and hence, $\mu_B = \mathring{\mu}_F + O_p(n^{-1})$. It follows that the Welch & Peers priors produce approximate confidence limits having conditional coverage error of order $O(n^{-1})$, which was shown by DiCiccio & Martin (1993).

Since $\mu_B = \mathring{\mu}_F + O_p(n^{-1})$, it follows that $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-1})$. In resolving conditional properties of second-order probability matching priors, the crucial question is: under what circumstances might $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$ hold? Since $\mathring{\omega}_r = 0$, an argument analogous to one given previously shows that $E_{Y|A}(\mu_B) = E_{Y|A}\{\zeta(\theta)\} + O(n^{-3/2})$, so that the crucial criterion reduces to $E_{Y|A}\{\zeta(\theta)\} = \mathring{\mu}_F + O(n^{-3/2})$. In the same way as was argued previously,

$$E_{Y|A}\Big\{D\frac{\partial \log \pi(\theta)}{\partial \theta^r}\ell^{r1}\Big\} = \mathring{\eta}\frac{\partial \log \pi(\theta)}{\partial \theta_r}\mathring{\omega}^{r1} + O(n^{-3/2})$$
$$E_{Y|A}(-\frac{1}{2}D\ell_{rst}\ell^{r1}\ell^{rs}) = -\frac{1}{2}\mathring{\eta}\mathring{\omega}_{rst}\mathring{\omega}^{r1}\mathring{\omega}^{rs} + O(n^{-3/2})$$
$$E_{Y|A}(-\frac{1}{6}D^3\ell_{rst}\ell^{r1}\ell^{s1}\ell^{t1}) = -\frac{1}{6}\mathring{\eta}^3\mathring{\omega}_{rst}\mathring{\omega}^{r1}\mathring{\omega}^{s1}\mathring{\omega}^{t1} + O(n^{-3/2})$$

Consequently, the crucial criterion holds provided that

$$\mathring{\eta}\frac{\partial \log \pi(\theta)}{\partial \theta_r}\mathring{\omega}^{r1} = \mathring{\eta}\mathring{\omega}_{rs/t}\mathring{\omega}^{r1}\mathring{\omega}^{st} + \frac{1}{2}\mathring{\eta}^3\mathring{\omega}_{rs/t}\mathring{\omega}^{r1}\mathring{\omega}^{s1}\mathring{\omega}^{t1} + O(n^{-3/2}). \tag{2.10}$$

2.2.4 Unconditional Matching to 3rd Order

We now consider 3rd order unconditional probability matching, which can be achieved by further adjusting the signed root statistic by its variance.

DiCiccio & Stern (1993) showed that the posterior expectation of $\{R(\psi)\}^2$ is $1 + a_B + O(n^{-3/2})$, where

$$
\begin{aligned}
a_B &= \frac{1}{4}(\hat{\ell}^{rs}\hat{\ell}^{tu} - \hat{V}^{rs}\hat{V}^{tu})\hat{\ell}_{rstu} - \frac{1}{4}(\hat{\ell}^{ru}\hat{\ell}^{st}\hat{\ell}^{vw} - \hat{V}^{ru}\hat{V}^{st}\hat{V}^{vw})\hat{\ell}_{rst}\hat{\ell}_{uvw} \\
&- \frac{1}{6}(\hat{\ell}^{ru}\hat{\ell}^{sw}\hat{\ell}^{tv} - \hat{V}^{ru}\hat{V}^{sw}\hat{V}^{tv})\hat{\ell}_{rst}\hat{\ell}_{uvw} + (\hat{\ell}^{rs}\hat{\ell}^{tu} - \hat{V}^{rs}\hat{V}^{tu})\hat{\ell}_{rst}\hat{\Pi}_u \\
&- (\hat{\ell}^{rs} - \hat{V}^{rs})\hat{\Pi}_{rs},
\end{aligned}
$$

and[12] $\hat{\Pi}_r = \Pi_r(\hat{\theta})$, $\hat{\Pi}_{rs} = \Pi_{rs}(\hat{\theta})$, with $\Pi_r = \pi_r(\theta)/\pi(\theta)$, $\Pi_{rs}(\theta) = \pi_{rs}(\theta)/\pi(\theta)$, $\pi_r(\theta) = \partial\pi(\theta)/\partial\theta_r$, $\pi_{rs}(\theta) = \partial^2\pi(\theta)/\partial\theta_r\partial\theta_s$. It follows then that the posterior variance of $R(\psi) - \mu_B$ is

$$\sigma_B^2 = 1 + a_B - \mu_B^2 + O(n^{-3/2}).$$

From a frequentist perspective, since μ_B is of order $O_p(n^{-1/2})$ and $\mu_B = \mu_F + O_p(n^{-1})$, where μ_F is of order $O(n^{-1/2})$, it follows that $\mu_B^2 = \mu_F^2 + O_p(n^{-3/2})$, where μ_F^2 is of order $O(n^{-1})$. Hence, the posterior variance satisfies

$$\sigma_B^2 = 1 + a_B - \mu_F^2 + O_p(n^{-3/2}).$$

DiCiccio & Stern (1994*b*) also showed that the unconditional frequentist expectation of $\{R(\psi)\}^2$ is $1 + a_F + O(n^{-3/2})$, where

$$
\begin{aligned}
a_F &= (\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\{\frac{1}{4}\omega_{rstu} - \omega_{rst/u} + \omega_{rt/su}\} \\
&- (\omega^{ru}\omega^{st}\omega^{vw} - \sigma^{ru}\sigma^{st}\sigma^{vw})\{\frac{1}{4}\omega_{rst}\omega_{uvw} - \omega_{rst}\omega_{uv/w} + \omega_{rs/t}\omega_{uv/w}\} \\
&- (\omega^{ru}\omega^{sw}\omega^{tv} - \sigma^{ru}\sigma^{sw}\sigma^{tv})\{\frac{1}{6}\omega_{rst}\omega_{uvw} - \omega_{rst}\omega_{uv/w} + \omega_{rs/t}\omega_{uv/w}\} + O(n^{-3/2});
\end{aligned}
$$

and since $E(\mu_B) = \mu_F + O(n^{-3/2})$, it follows from their results that the frequentist variance of $R(\psi) - \mu_B$ is

[12]This $\hat{\Pi}_{rs}$ type of notation is equivalent to notation used in the shrinkage argument, but we adopt this notation to better handle larger dimensions.

$$\sigma_F^2 = 1 + a_F + 2\eta\mu_{F/r}\omega^{r1} - \mu_F^2 + O(n^{-3/2}),$$

where $\mu_{F/r} = \partial\mu_F/\partial\theta_r$.

From both Bayesian and frequentist perspectives, the third- and higher-order cumulants of $R(\psi) - \mu_B$ are $O(n^{-3/2})$ or smaller. Thus, the marginal distribution of $\{R(\psi) - \mu_B\}/\sigma_F$ and the posterior distribution of $\{R(\psi) - \mu_B\}/\sigma_B$ are both standard normal to error of order $O(n^{-3/2})$; moreover, if $\pi(\theta)$ is a prior density such that $\sigma_B^2 = \sigma_F^2 + O_p(n^{-3/2})$, then $\pi(\theta)$ is a third-order unconditional probability matching prior. The sufficient condition for $\pi(\theta)$ to be a third-order PMP is that

$$a_B = a_F + 2\eta\mu_{F/r}\omega^{r1} + O_p(n^{-3/2}).$$

Since

$$\begin{aligned} a_B = &\frac{1}{4}(\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\omega_{rstu} - \frac{1}{4}(\omega^{ru}\omega^{st}\omega^{vw} - \sigma^{ru}\sigma^{st}\sigma^{vw})\omega_{rst}\omega_{uvw} \\ &- \frac{1}{6}(\omega^{ru}\omega^{sw}\omega^{tv} - \sigma^{ru}\sigma^{sw}\sigma^{tv})\omega_{rst}\omega_{uvw} + (\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\omega_{rst}\Pi_u \\ &- (\omega^{rs} - \sigma^{rs})\Pi_{rs} + O_p(n^{-3/2}), \end{aligned}$$

the third-order condition can be expressed as

$$\begin{aligned} a_F + 2\eta\mu_{F/r}\omega^{r1} = &\frac{1}{4}(\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\omega_{rstu} - \frac{1}{4}(\omega^{ru}\omega^{st}\omega^{vw} - \sigma^{ru}\sigma^{st}\sigma^{vw})\omega_{rst}\omega_{uvw} \\ &- \frac{1}{6}(\omega^{ru}\omega^{sw}\omega^{tv} - \sigma^{ru}\sigma^{sw}\sigma^{tv})\omega_{rst}\omega_{uvw} + (\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\omega_{rst}\Pi_u \\ &- (\omega^{rs} - \sigma^{rs})\Pi_{rs}, \end{aligned}$$

that is

$$
\begin{aligned}
-(\omega^{rs} - \sigma^{rs})\Pi_{rs} + (\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\omega_{rst}\Pi_u = \\
&- (\omega^{rs}\omega^{tu} - \sigma^{rs}\sigma^{tu})\{\omega_{rst/u} - \omega_{rt/su}\} \\
&+ (\omega^{ru}\omega^{st}\omega^{vw} - \sigma^{ru}\sigma^{st}\sigma^{vw})\{\omega_{rst}\omega_{uv/w} - \omega_{rs/t}\omega_{uv/w}\} \\
&+ (\omega^{ru}\omega^{sw}\omega^{tv} - \sigma^{ru}\sigma^{sw}\sigma^{tv})\{\omega_{rst}\omega_{uv/w} - \omega_{rs/t}\omega_{uv/w}\} + 2\eta\mu_{F/r}\omega^{r1}.
\end{aligned}
$$

By assuming that the prior density $\pi(\theta)$ satisfies the second-order probability matching condition, the third-order condition reduces to

$$
\begin{aligned}
\tau^{rs}\Pi_{rs} - \tau^{rs}\omega^{tu}\omega_{rst}\Pi_u &= (\tau^{rs}\omega^{tu} - \frac{2}{3}\tau^{rs}\tau^{tu})\omega_{rst/u} - \omega^{ru}\tau^{st}\omega^{vw}\omega_{rst}\omega_{uv/w} \\
&- (2\omega^{ru}\omega^{sw}\tau^{tv} - 2\omega^{ru}\tau^{sw}\tau^{tv} - \tau^{ru}\omega^{sw}\tau^{tv} + \tau^{ru}\tau^{sw}\tau^{tv})\omega_{rst}\omega_{uv/w} \\
&+ (\tau^{ru}\omega^{st}\omega^{vw} - \tau^{ru}\omega^{st}\tau^{vw} + \frac{1}{2}\tau^{ru}\tau^{st}\tau^{vw})\omega_{rs/t}\omega_{uv/w} \\
&+ (\tau^{ru}\omega^{sw}\omega^{tv} - \tau^{ru}\omega^{sw}\tau^{tv})\omega_{rs/t}\omega_{uv/w},
\end{aligned}
$$

which can be written somewhat more succinctly as

$$
\begin{aligned}
\tau^{rs}\Pi_{rs} - \tau^{rs}\omega^{tu}\omega_{rst}\Pi_u = (\tau^{rs}\sigma^{tu} + \frac{1}{3}\tau^{rs}\tau^{tu})\omega_{rst/u} - \omega^{ru}\tau^{st}\omega^{vw}\omega_{rst}\omega_{uv/w} \\
- (\omega^{ru}\sigma^{sw}\tau^{tv} + \sigma^{ru}\sigma^{sw}\tau^{tv})\omega_{rst}\omega_{uv/w} \\
+ (\tau^{ru}\omega^{st}\sigma^{vw} + \frac{1}{2}\tau^{ru}\tau^{st}\tau^{vw})\omega_{rs/t}\omega_{uv/w} + \tau^{ru}\omega^{sw}\sigma^{tv}\omega_{rs/t}\omega_{uv/w}.
\end{aligned}
\tag{2.11}
$$

Welch & Peers (1963) showed that in a scalar parameter model, Jeffreys' prior is the unique second-order matching prior, and this prior is only third-order probability matching if the standardized skewness of the score function, $\omega_{1,1,1}(-\omega_{11})^{-3/2}$ is a constant, i.e. it does not depend on the model parameter.[13] For a one-parameter location model, Jeffreys'

[13]This is trivially satisfied for a standard normal location model, but there is a notable absence in the

prior is a constant which does not depend on the model parameter, and this is a case where we have 3rd order matching, according to the above formula. In fact, the matching is exact, as was shown by Welch & Peers (1963).

2.2.5 Conditional Matching to 3rd Order

Suppose that π is a second-order matching prior satisfying the condition $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$. Let $\mathring{\sigma}_F^2$ denote the conditional frequentist variance of $R(\psi) - \mu_B$. From arguments similar to the ones given previously that showed $\mathring{\mu}_F = \mu_F + O_p(n^{-1})$, it follows that $\mathring{\sigma}_F^2 = \sigma_F^2 + O(n^{-3/2})$. To be specific, recall that

$$\sigma_F^2 = 1 + a_F + 2\eta\mu_{F/r}\omega^{r1} - \mu_F^2 + O(n^{-3/2})$$

where $a_F + 2\eta\mu_{F/r}\omega^{r1} - \mu_F^2$ is of order $O(n^{-1})$ and can be expressed, to error of order $O(n^{-3/2})$, as a function of the ω's. By applying the identical calculations to the conditional distribution, it follows that

$$\mathring{\sigma}_F^2 = 1 + \mathring{a}_F + 2\mathring{\eta}\mathring{\mu}_{F/r}\mathring{\omega}^{r1} - \mathring{\mu}_F^2 + O(n^{-3/2}),$$

where $\mathring{a}_F + 2\mathring{\eta}\mathring{\mu}_{F/r}\mathring{\omega}^{r1} - \mathring{\mu}_F^2$ is of order $O_p(n^{-1})$ and can be expressed, to error of order $O_p(n^{-3/2})$, as the identical function as can its unconditional version, with each ω being replaced by its corresponding $\mathring{\omega}$. Since by assumption each $\mathring{\omega}$ differs by its corresponding ω by $O_p(n^{-1/2})$, it follows that $\mathring{\sigma}_F^2 = \sigma_F^2 + O_p(n^{-3/2})$. Hence, the condition that ensures $\pi(\theta)$ is 3rd order matching in the marginal frequentist sense also ensures that it is 3rd matching in the conditional frequentist sense.

Then the conditions for third-order conditional matching are the condition for *unconditional* 3rd order matching combined with the condition for 2nd order conditional matching given by equation (2.10) or, simplifying slightly,

$$\frac{\partial \log \pi(\theta)}{\partial \theta_r}\mathring{\omega}^{r1} = \mathring{\omega}_{rs/t}\mathring{\omega}^{r1}\mathring{\omega}^{st} + \frac{1}{2}\mathring{\eta}^2\mathring{\omega}_{rs/t}\mathring{\omega}^{r1}\mathring{\omega}^{s1}\mathring{\omega}^{t1} + O_p(n^{-3/2}), \tag{2.12}$$

literature of models in which this condition is satisfied.

What this means is that, since $\mathring{\mu}_F = \mu_F + O_p(n^{-1})$, it follows that $\mathring{\sigma}_F^2 = \sigma_F^2 + O_p(n^{-3/2})$. Therefore, a prior satisfying the 3rd order unconditional matching condition will also satisfy the 3rd order conditional matching condition provided that $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$.

2.2.6 Proof of the Probability Matching Property

We now formally state the theorem concerning the probability matching property of priors identified via this route.

Theorem 2.2.1 *Suppose that the prior π is such that $E_{Y|A}(\mu_B) = \mathring{\mu}_F + O(n^{-3/2})$. Then if the prior also satisfies $\sigma_B^2 = \mathring{\sigma}_F^2 + O(n^{-3/2})$, the Bayesian quantile is 3rd order conditional probability matching.*

Proof: Given Y, let $\psi_l \equiv \psi^{1-\alpha}(\pi, Y)$ be the posterior $1-\alpha$ quantile for the interest parameter, i.e. $Pr(\psi \leq \psi^{1-\alpha}(\pi, Y)|Y) = 1-\alpha$. Also, given Y, $R(\psi)$ is a monotonic decreasing function of ψ. Therefore,

$$Pr(R_Y(\psi) \geq R_Y(\psi_l)|Y) = 1-\alpha,$$

where $R_Y(\psi)$ is the signed root statistic constructed from Y. That is,

$$Pr\Big(\frac{R_Y(\psi)-\mu_B}{\sigma_B} \geq \frac{R_Y(\psi_l)-\mu_B}{\sigma_B}|Y\Big) = 1-\alpha$$

Since $\frac{R_Y(\psi)-\mu_B}{\sigma_B} \sim N(0,1) + O_p(n^{-3/2})$, where $N(0,1)$ denotes the standard normal distribution, then, by the delta method from Section 2.7 of Hall (1992),

$$Pr\Big(N(0,1) \geq \frac{R_Y(\psi_l)-\mu_B}{\sigma_B}\Big) = 1-\alpha + O(n^{-3/2}),$$

so that

$$\frac{R_Y(\psi_l)-\mu_B}{\sigma_B} = z_\alpha + O(n^{-3/2}),$$

in terms of the $N(0,1)$ quantile z_α defined by $\Phi(z_\alpha) = \alpha$.

From a conditional frequentist perspective, given an ancillary statistic $A = a$ and noting that given Y, the event $\psi \leq \psi_l$ is equivalent to the event $R_Y(\psi) \geq R_Y(\psi_l)$. Thus,

$$
\begin{aligned}
Pr(\psi \leq \psi_l | a) &= Pr(R_Y(\psi) - \mu_B \geq R_Y(\psi_l) - \mu_B | a) \\
&= Pr\Big(\frac{R_Y(\psi) - \mu_B}{\sigma_B} \geq \frac{R_Y(\psi_l) - \mu_B}{\sigma_B} | a\Big) \\
&= Pr\Big(\frac{R_Y(\psi) - \mu_B}{\mathring{\sigma}_F} + O_p(n^{-3/2}) \geq \frac{R_Y(\psi_l) - \mu_B}{\sigma_B} | a\Big) \\
&= Pr\Big(\frac{R_Y(\psi) - \mu_B}{\mathring{\sigma}_F} \geq z_\alpha + O_p(n^{-3/2}) | a\Big) + O(n^{-3/2}) \\
&= Pr\Big(N(0,1) \geq z_\alpha\Big) + O(n^{-3/2}) \\
&= 1 - \alpha + O(n^{-3/2}).
\end{aligned}
$$

The first equality follows since $R(\psi)$ is monotonically decreasing in ψ, the second equality simply adjusts by the variance, the third equality follows from the assumption that $\sigma_B = \mathring{\sigma}_F + O(n^{-3/2})$, the fourth equality follows by the delta method, the fifth equality follows by the distributional result for the Bayesian mean- and variance-adjusted signed root statistic combined with the delta method and the last equality is just a re-expression of the preceding line.

■

2.3 Method 3: The Saddlepoint Approach

The saddlepoint approach to identifying conditional PMPs was suggested by Casella et al. (1995) and further developed by DiCiccio & Young (2010). This approach relies on Barndorff-Nielsen (1986) for the conditional frequentist saddlepoint approximation and DiCiccio & Martin (1993) for the Bayesian counterpart. Since this method already exists in the literature, we do not present the general derivation for parameters of arbitrary length, but instead present only the results we will use, i.e. the setting of a scalar interest and scalar nuisance parameter.

In this section, we can simplify notation to emphasize the interest-nuisance parameter structure of the parameter vector, $\theta = (\theta_1, \theta_2) = (\psi, \lambda)$. We will do so by replacing the numbered indices in the likelihood quantities with the corresponding scalar parameter, for example

$$\ell_1 = \ell_\psi, \quad \ell_2 = \ell_\lambda, \quad \ell_{21} = \ell_{\lambda\psi},$$

etc.

Barndorff-Nielsen (1986) showed that a modified signed square root likelihood ratio statistic, called R^*, has a distribution which may be approximated by a standard normal distribution with error of order $O(n^{-3/2})$, conditional on an ancillary statistic A. This statistic is defined as

$$R^* = R(\psi) + R^{-1}(\psi)\log\big(U_F(\psi)/R(\psi)\big) \tag{2.13}$$

where the adjustment term $U_F(\psi)$ is such that $U_F(\psi) = R(\psi) + O_p(n^{-1/2})$. The adjustment term is also parameterization invariant and does not depend on the nuisance parameter λ. The error in approximating R^* by the standard normal is a relative error and holds in small deviation regions.[14] A tail-area approximation is given by,

$$Pr(R^* \le r; \theta|A) = \Phi(r) + \phi(r)[U_F^{-1}(\psi) - r^{-1} + O_p(n^{-3/2})]. \tag{2.14}$$

The conditional frequentist adustment term is given by

$$U_F(\psi) = \frac{\begin{vmatrix} \ell_{;\hat\theta}(\hat\psi,\hat\lambda) - \ell_{;\hat\theta}(\psi,\hat\lambda_\psi) & \ell_{\lambda;\hat\theta}(\psi,\hat\lambda_\psi)\end{vmatrix}}{\left\{\left|\ell_{\lambda\lambda}(\psi,\hat\lambda_\psi)\right|\left|\ell_{\theta\theta}(\hat\psi,\hat\lambda)\right|\right\}^{1/2}}, \tag{2.15}$$

where all terms are simply log-likelihood derivatives. More specifically,

$$\ell_{;\hat\theta}(\hat\psi,\hat\lambda) = \begin{pmatrix} \frac{\partial \ell(\theta)}{\partial\hat\psi} \\ \frac{\partial \ell(\theta)}{\partial\hat\lambda} \end{pmatrix}_{\theta=\left(\hat\psi,\hat\lambda\right)},$$

[14]The absolute error would likely be $O(n^{-1})$, and would hold in large-deviation regions. Since we are working with the maximum likelihood estimator, which is $\sqrt{n}$-consistent, we are in a small-deviation region. Jensen (1995) gives an alternative approximation for large-deviation regions, which turns out to be exactly equal to R^* only when there is an exact ancillary statistic.

$$\ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}_{\psi}) = \begin{pmatrix} \frac{\partial \ell(\theta)}{\partial \hat{\psi}} \\ \frac{\partial \ell(\theta)}{\partial \hat{\lambda}} \end{pmatrix}_{\theta=\left(\psi,\hat{\lambda}_{\psi}\right)},$$

$$\ell_{\lambda;\hat{\theta}}(\hat{\psi}, \hat{\lambda}_{\psi}) = \begin{pmatrix} \ell_{\lambda;\hat{\psi}}(\psi, \hat{\lambda}_{\psi}) \\ \ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_{\psi}) \end{pmatrix}_{\theta=\left(\psi,\hat{\lambda}_{\psi}\right)},$$

$$\ell_{\theta\theta}(\hat{\psi}, \hat{\lambda}) = \begin{pmatrix} \ell_{\psi\psi} & \ell_{\psi\lambda} \\ \ell_{\lambda\psi} & \ell_{\lambda\lambda} \end{pmatrix}_{\theta=\left(\hat{\psi},\hat{\lambda}\right)}.$$

The terms $\ell_{\psi\psi}$,$\ell_{\lambda\lambda}$, $\ell_{\lambda;\hat{\lambda}}$ and $\ell_{\lambda,\hat{\psi}}$ are second-order partial derivatives of the full log-likelihood.[15] The terms $\ell_{\lambda;\hat{\lambda}}$ and $\ell_{\lambda;\hat{\psi}}$ involve partial derivatives of the average log-likelihood with respect to the maximum likelihood estimator holding the ancillary statistic fixed. Such partial derivatives are called sample space derivatives. The disadvantage of this approach is that in order to calculate the sample space derivatives, we must explicitly specify an ancillary statistic. As argued by Cox & Hinkley (1974), existence and uniqueness of ancillary statistics can be problematic and hence construction of ancillary statistics can prove quite difficult. However, as Severini (2000) points out, in cases where an exact ancillary statistic, A exists, such that the maximum likelihood estimator and ancillary statistic are jointly minimal sufficient, then the log-likelihood function may be written as

$$\ell(\psi, \lambda) = \ell(\psi, \lambda; \hat{\psi}, \hat{\lambda}, A).$$

Whenever it is possible to express the likelihood in this form, the relevant likelihood quantities can be easily derived.

Using asymptotic results of DiCiccio & Martin (1991), DiCiccio & Martin (1993) extends the Barndorff-Nielsen adjusted signed root likelihood ratio statistic to the Bayesian setting. The key advantage of this approach is that it is not necessary to explicitly specify an ancillary statistic. The authors give the Bayesian form of the adjustment term as

[15] Again we ask the reader to note that we are using the full log-likelihood as in the previous section.

$$U_B(\psi) = \ell_\psi\big(\psi, \hat{\lambda}_\psi\big) \frac{\Big| - \ell_{\lambda\lambda}\big(\psi, \hat{\lambda}_\psi\big)\Big|^{1/2} \pi(\hat{\theta})}{\Big| - \ell_{\theta\theta}\big(\hat{\theta}\big)\Big|^{1/2} \pi(\psi, \hat{\lambda}_\psi)}. \quad (2.16)$$

For a Welch & Peers prior, it can be shown that

$$U_B = U_F + O_p(n^{-1}).$$

and this is true to order $O_p(n^{-1/2})$ in general (for a general prior). The statistic $R + R^{-1}\log(U_B R^{-1})$ has a posterior distribution which is standard normal to $O(n^{-3/2})$ in general, and a frequentist distribution which is standard normal to order $O(n^{-1})$, and also to $O(n^{-1})$, U_B is parameterization invariant. Casella et al. (1995) argue that matching priors to order $O(n^{-3/2})$ may be found by choosing the prior to satisfy

$$U_F = U_B + O(n^{-3/2}). \quad (2.17)$$

Unlike the shrinkage argument, which calculates the frequentist coverage of the Bayesian quantiles directly and therefore does not require further proof of the probability matching achieved, the saddlepoint approach compares separate asymptotic expansions and thus it must be established that $U_F = U_B + O(n^{-3/2})$ ensures probability matching to 3rd order. A formal and rigorous proof of this result can be constructed as follows.

Theorem 2.3.1 *If a prior is such that $U_F = U_B + O_p(n^{-3/2})$, then the Bayesian posterior credible sets match the conditional frequentist coverage probabilities to order $O(n^{-3/2})$.*

Proof: Suppose the prior is such that $U_F = U_B + O_p(n^{-3/2})$. Fix ψ and then following Casella et al. (1995), for values ψ such that $\hat{\psi} - \psi$ is of order $O(n^{-1/2})$, we have that

$$Pr\big(\psi \geq \psi_0 | Y\big) = \Phi(R) + \phi(R)\big\{R^{-1} - U_B^{-1}\big\} + O(n^{-3/2})$$

where Φ and ϕ are the standard normal distribution and density functions respectively. This Bayesian asymptotic expansion was first derived by DiCiccio & Martin (1991) and DiCiccio & Martin (1993).

From a frequentist perspective, Barndorff-Nielsen (1986) showed that

$$\Phi(R^*) = \Phi(R) + \phi(R)\big\{R^{-1} - U_F^{-1}\big\} + O(n^{-3/2}).$$

Combining these conditions, we see that, for given Y , the event $\psi \leq \psi^{(1-\alpha)}(\pi, Y)$ is the same as the condition that $\Phi(R^*) + O(n^{-3/2}) \geq \alpha$. Therefore, by the delta method of Hall (1992), we have from a repeated sampling perspective that

$$\begin{aligned} Pr_\theta\big\{\psi \leq \psi^{(1-\alpha)}(\pi, Y)\,|A = a\big\} &= Pr_\theta\big[\Phi(R^*) + O_p(n^{-3/2}) \geq \alpha|A = a\big] \\ &= Pr_\theta\big[R^* + O_p(n^{-3/2}) \geq z_\alpha|A = a\big] \\ &= 1 - \alpha + O_p(n^{-3/2}), \end{aligned}$$

where $\Phi(z_\alpha) = \alpha$. Thus the Bayesian confidence limits have conditional frequentist coverage error of order $O(n^{-3/2})$, which is the definition of third-order conditional probability matching.

■

In the PMP literature, this condition has also been called ”strong matching”.[16] This method of identifying priors has several advantages which will become apparent in subsequent calculations, namely that it does not involve partial differential equations or expected likelihood quantities but rather simpler observed likelihood quantities.

As argued by Efron & Hinkley (1978), when using normal approximations to the distributions of maximum likelihood estimators–and by implication other likelihood quantities–it is better to use observed information rather than expected Fisher information. The expected information is an average measure of information, whereas the observed information is for a particular data set and therefore more ”relevant” to the data. Particularly in small samples, the expected information may greatly differ from the observed information. In light of these type of arguments, we think this is another advantage of using the matching conditions derived via this saddlepoint approach.

[16]See, for instance, Fraser & Reid (2002).

The saddlepoint approach highlights another advantage of Bayesian inference, which is that the Bayesian adjustment term does not contain sample space derivatives and therefore does not require the specification of an exact ancillary.[17]

2.4 Method 4: The Edgeworth Approach

A complete picture of objective Bayesian methods for conditional inference would include all of the most common asymptotic expansions. Thus far, we have considered posterior expansions of the Laplace type and also more general saddlepoint approximations. We now consider what is possible using Edgeworth expansions.

In the shrinkage argument, we did not actually derive both a frequentist and Bayesian expansion of a test statistic. Rather, we derived a posterior expansion for $\sqrt{n}(\hat{\theta} - \theta)$, a quantity which is standard normally distributed to first-order. We then considered the frequentist properties of this expansion, thus obtaining a frequentist expansion via a Bayesian route. Here, we take a more direct approach. Firstly, there are statistics which are more natural for parametric inference, such as the signed root likelihood ratio statistic. Second, instead of calculating the frequentist properties of a Bayesian expansion, we will investigate whether it is helpful or illuminating to derive the frequentist and Bayesian expansions separately and then compare them. The idea is that, if we use the same expansion for both the frequentist and posterior densities of the same test statistic, then we obtain two expansions which are standard normal to the same order of error and have coefficients on the expansion terms which can be directly compared. The validity of such comparisons may be established by this simple theorem.

Theorem 2.4.1 *Suppose we have valid asymptotic expansions for both the posterior and conditional frequentist distribution of the signed root statistic, which are both asymptotically standard normal to order $O(n^{-3/2})$. Then if the expansions are equal to order $O(n^{-3/2})$, we will have that the tail probabilities for that test statistic match to order $O(n^{-3/2})$.*

[17] It may occur to the reader that it might be possible to compare the mean and variance-adjusted Bayesian signed root statistic with the modified Bayesian signed root statistic from this section. The problem is that in general they are standard normal to different orders of error. We discuss this further in Chapter 6.

Proof: Let $G(Y)$ be a function approximating the posterior distribution function of the signed root statistic to order $O(n^{-3/2})$. For given Y, the event $\psi \leq \psi_l$ is equivalent to the event $G(Y) + O_p(n^{-3/2}) \leq 1 - \alpha$, where ψ_l is the Bayesian $1 - \alpha$ quantile.

Let $F(\cdot)$ be a frequentist expansion for $H(\cdot)$, the true conditional distribution function of the signed root statistic R, such that $F(R) = H(R) + O_p(n^{-3/2})$. From a conditional frequentist perspective, the conditional probability

$$\begin{aligned} Pr_{Y|a}(\psi \leq \psi_l) &= Pr_{Y|a}(G(Y) + O_p(n^{-3/2}) \leq 1 - \alpha) \\ &= Pr_{Y|a}(G(Y) \leq 1 - \alpha) + O(n^{-3/2}) \\ &= Pr_{Y|a}(F(R) \leq 1 - \alpha) + O(n^{-3/2}) \\ &= Pr_{Y|a}(H(R) + O_p(n^{-3/2}) \leq 1 - \alpha) + O(n^{-3/2}) \\ &= Pr_{Y|a}(H(R) \leq 1 - \alpha) + O(n^{-3/2}) \\ &= 1 - \alpha + O(n^{-3/2}), \end{aligned}$$

where the first equality is a definition, the second equality follows by the delta method. The third equality, which relates the Bayesian and frequentist approximations, follows by the mild assumption that we can choose the prior appropriately such that we can recognize $G(Y)$ as the leading terms in a frequentist expansion of $H(R)$. This is valid whenever both expansions are in powers of $n^{-1/2}$ and accurate to the same order of error. The fourth equality is a definition, the fifth equality follows by the delta method. The final equality follows by the probability integral transform, since $H(R)$ is uniformly distributed on $[0, 1]$.

■

We will obtain an Edgeworth expansion of the conditional distribution of a general likelihood-based test statistic, T, following Severini (1990). The signed root statistic will be a particular case. We will do this for T rather than only for the signed root statistic because we will use the more general result later for comparing conditional and unconditional frequentist inference. When we consider the Bayesian asymptotics, however, we will consider only the special case of the signed root statistic. The expansion of the posterior distribution of the signed root statistic requires further comment. There currently does not exist a general direct Edgeworth expansion for posterior densities. The earliest posterior expansions due to Lindley (1961), Johnson (1967) and Ghosh et al. (1985) are essentially Laplace expansions, which take the posterior as a ratio and expand the numerator and denominator separately. The resulting expansions have the Edgeworth property of being functions of powers of $n^{-1/2}$ and the advantage noted in the shrinkage argument literature that the coefficients in the expansion depend only on low-order derivatives of the log-likelihood and prior density. However, as noted by Weng (2010), one major shortcoming of this simplistic nature is that, by not depending on orthogonal Hermite polynomials and moments or cumulants, these expansions do not efficiently store information about the distribution and are not directly comparable to the corresponding Edgeworth expansions in frequentist settings. Thus, while the Johnson-type expansions are the closest relative to Edgeworth expansions which currently exist for posterior densities, they are not suitable for comparison to Edgeworth expansions of frequentist densities as needed in the present context.

It is beyond our scope[18] to prove the existence and validity of Edgeworth expansions for posterior densities. Rather, we will assume that such expansions are valid only for the posterior density of the signed root likelihood ratio statistic. Finally we will compare the terms of the Edgeworth expansions for the conditional and posterior densities of similar order of error to obtain conditions which a prior must satisfy to be probability matching to that respective order of error.

Severini (1990) compares Edgeworth expansions to conditional and unconditional densities of test statistics to order $O(n^{-1})$. We wish to extend this work to the Bayesian setting for the purposes of identifying 2nd order probability matching priors. Extensions to the higher-order setting are discussed at the end of the section.

In this section, we make a further simplification of notation because we are only considering scalar parameter models. Thus all of the indices in the likelihood quantities run

[18] Personal communication with J.K. Ghosh, one of the leading experts on posterior expansions, has confirmed that such expansions have not yet been proven to be valid and that such a proof would be quite difficult. We are currently studying the existence of posterior Edgeworth expansions, though so far without success. On the other hand, Weng (2010) has derived posterior expansions via Stein's identity which have all of the properties of an Edgeworth expansion, though they are derived directly by expansion of the moment generating function, and therefore we would not call them true Edgeworth expansions.

over a single parameter. In this setting, it is more convenient to change notation slightly. We have

$$\ell_1 = \frac{\partial \ell}{\partial \theta}, \quad \ell_2 = \frac{\partial^2 \ell}{\partial \theta \partial \theta},$$

$$i_{10} = E(\ell_1), \quad i_{01} = E(\ell_2),$$

$$i_{20} = E(\ell_1^2), \quad i_{11} = E(\ell_1 \ell_2),$$

etc. Let $Y = Y_1, \ldots, Y_n$ be identically distributed random variables. The Y_j have common density $f(y; \theta)$.

2.4.1 General Derivation for Likelihood-Based Statistic, T

In what follows, we rely heavily on Cox (1980) and Ryall (1981) and we refer the reader to Severini (1990) for a proof of the validity of this approach. The general likelihood-based statistic, T, which we consider here, is a function of likelihood quantities such as the maximum likelihood estimator and score function. Furthermore, we suppose that the statistic T is such that its null distribution admits an Edgeworth expansion of the form

$$Pr(T \leq t; \theta_0) = \Phi(t) - \frac{1}{\sqrt{n}} \phi(t) \frac{\rho_3}{6} H_2(t) + O(n^{-1}),$$

where θ_0 is some specified value, $\rho_3/\sqrt{n}$ is the third standardized cumulant of T and $H_j(\cdot)$ is the jth Hermite polynomial. We also suppose that the conditional null distribution of T given $A = a$ admits an Edgeworth expansion of the form

$$\begin{aligned} Pr(T \leq t | A = a; \theta_0) = &\Phi(t) - \frac{1}{\sqrt{n}} \Big\{ \frac{\rho_{21}}{2} H_1(a) H_1(t) + \frac{\rho_{12}}{2} H_2(a) H_0(t) + \frac{\rho_{30}}{6} H_2(t) \Big\} \\ &+ O(n^{-1}) \end{aligned}$$

where $\rho_{ij}/\sqrt{n}$ represents the standardized (i, j)th cumulant of (T, A) following Barndorff-Nielsen & Cox (1979).

Now, let

$$L_T(t) = Pr(T \geq t; \theta_0),$$

which ensures that for all $0 \leq \alpha \leq 1$, $Pr\{L_T(T) \leq \alpha; \theta_0\} \leq \alpha$, and which holds with equality when T has a continuous distribution. We also let

$$\alpha(a) = Pr\{L_T(T) \leq \alpha | A = a; \theta_0\}$$

be the corresponding conditional probability relationship.

We proceed in the following way: let

$$U_j = \partial \log f(Y_j; \theta)/\partial\theta,$$

$$V_j = \partial^2 \log f(Y_j; \theta)/\partial\theta\partial\theta,$$

and

$$\bar{U} = n^{-1} \sum U_j,$$
$$\bar{V} = n^{-1} \sum V_j.$$

Write $i_{lm} = E(U_j^l V_j^m; \theta_0)$ Also, let

$$\bar{S}_1 = \frac{\bar{U}}{\sqrt{i_{20}}}, \quad \bar{S}_2 = \frac{1}{\sigma}\{\bar{V} + i_{20} - \bar{U}(\frac{i_{11}}{i_{20}})\},$$

where

$$\sigma^2 = i_{02} - i_{20}^2 - \frac{i_{11}^2}{i_{20}}$$

and we define

$$S_1 = \sqrt{n}\bar{S}_1, \quad S_2 = \sqrt{n}\bar{S}_2.$$

Now, suppose we have an expansion to order $O(n^{-1})$ for a likelihood-based statistic, T, in terms of S_1 and S_2 with expansion denoted by S^*, such that

$$S^* = S_1 + \frac{1}{\sqrt{n}}(\alpha_1 S_1 S_2 + \alpha_2 S_1^2)$$

where α_1 and α_2 are arbitrary constants which correspond to particular choices of T. The necessary regularity conditions and proof of the validity of this asymptotic expansion, for a group of likelihood-based test statistics of which the signed root is a particular case, is given in the Appendix of Severini (1990). The regularity conditions necessary for these expansions to be valid are:

1. The true value θ_0 is an interior point of the parameter space

2. The function $\log f(y;\theta)$ is four times differentiable in a neighbourhood of θ_0 for almost all y and

$$E[\{\partial^j \log f(y;\theta_0)/\partial\theta_0^j\}^2;\theta_0] < \infty \quad (j = 1, 2, 3, 4).$$

   Furthermore, there exists a square-integrable function H such that for all sufficiently small $|\delta|$,

$$|[\partial^4 \log f(y;\theta)/\partial\theta^4]_{\theta=\theta_0+\delta}| \leq H(y).$$

3. The order of integration and differentiation can be interchanged in computing

$$\partial^j E\{\log f(y;\theta);\theta\}/\partial^j \quad (j = 1, 2, 3, 4).$$

4. The random variables (U_j, V_j) satisfy general regularity conditions given by Cox (1980).

5. The maximum likelihood estimator $\hat{\theta}$ is consistent.

Now, using an Edgeworth expansion for the distribution of (S^*, A), we have an expansion of the conditional (on a particular value of A) distribution of $S^* - \alpha_2/\sqrt{n}$, with ancillary statistic $A = S_2 + O_p(n^{-1/2})$, given by

$$f(s|a) = \phi(s)\Big\{1 + \frac{1}{6}\{3\kappa^*_{21}H_2(s)H_1(a) + \kappa^*_{30}H_3(s) + 3\kappa^*_{12}H_1(s)H_2(a)\Big\} + O(n^{-1}). \tag{2.18}$$

Letting κ^*_{lm} denote the joint (l,m)th cumulant of (S^*, A),

$$\kappa^*_{10} = \frac{\alpha_2}{\sqrt{n}}, \quad \kappa^*_{20} = \kappa^*_{02} = 1, \quad \kappa^*_{11} = \kappa^*_{12} = 0,$$
$$\kappa^*_{21} = \frac{1}{\sqrt{n}}(2\alpha_1 - \frac{\sigma}{i_{20}}), \quad \kappa^*_{30} = \frac{1}{\sqrt{n}}(6\alpha_2 + \frac{i_{30}}{i_{20}^{3/2}}),$$

to order $O(n^{-1})$, where we have used the fact that to order $O(n^{-1/2})$, S_1 and S_2 are bivariate normal.

Letting $T = S^* - \alpha_2/\sqrt{n}$, we obtain

$$\alpha(a) = \alpha - \phi(t)\Big\{\frac{1}{6\sqrt{n}}\{3\kappa^*_{21}H_2(t)H_1(a) + 3\kappa^*_{12}H_1(t)H_2(a)\Big\} \tag{2.19}$$

2.4.2 The Signed Root Likelihood Ratio Statistic

For the signed root statistic, it is convenient to re-express in an equivalent form depending on the score function, i.e.

$$R = sgn(\bar{U})[2\{\ell(\hat{\theta}) - \ell(\theta_0)\}]^{1/2},$$

and the expansion, to $O(n^{-1})$ is given by[19]

$$S^* = S_1 + \frac{1}{\sqrt{n}}\Big(\frac{\sigma}{2i_{20}}S_1S_2 - \frac{i_{30}}{6i_{20}^{3/2}}S_1^2\Big) + O_p(n^{-1}).$$

[19] At the end of this section, we discuss the difficulties in extending this to $O(n^{-3/2})$.

2.4.3 The Bayesian Asymptotics

We now look at the Bayesian asymptotics. Let $\pi(r|Y)$ denote the posterior density of $R(\theta)$, the usual signed root likelihood statistic. As shown by Woodroofe (1992) and Sweeting (1995*b*)[20], the posterior density is given by

$$\pi(r|Y) \propto \phi(r)\frac{\pi(\theta)}{\pi(\hat{\theta})}\frac{\sqrt{2[\ell(\hat{\theta})-\ell(\theta)]}}{|-\ell_\theta(\theta)|}. \tag{2.20}$$

Then, assuming the regularity conditions necessary for the validity of an Edgeworth expansion, as well as the Bickel-Ghosh regularity conditions for the prior, we assume the validity of an Edgeworth expansion of the posterior distribution given by

$$Pr_{\pi(r|Y)}\{R \le r\} = \Phi(r) - \phi(r)\frac{b_3}{6\sqrt{n}}H_2(r) + O(n^{-1}), \tag{2.21}$$

where b_j denotes the ith standardized cumulant of the posterior density, i.e. the cumulants, κ_j^B are the coefficients of the log of the moment generating function for a random variable with density function given by the posterior density, and the standardized cumulants are defined in the following way: b_1 is the mean, b_2 is the variance and $b_j = \kappa_j^B/b_2^{1/2},\ \ j \ge 3$.

We could then identify probability matching priors by setting the coefficients of the same order of n from the different Edgeworth expansions equal to one another. For example, if the prior is chosen such that

$$3\kappa_{21}^* H_2(r)H_1(a) + \kappa_{30}^* H_3(r) + 3\kappa_{12}^* H_1(r)H_2(a) = b_3 H_2(r), \tag{2.22}$$

then we will have 2nd order matching of posterior and conditional frequentist tail area probabilities. This corresponds to matching the coefficients from equations (2.18) and (2.21).

Calculation of the standardized cumulants of the posterior density is analytically challenging and we have not yet found a good example. A subject of future work is the numerical solution of these conditions. While the comparison of conditional frequentist and Bayesian Edgeworth expansions is intractable, it might occur to the reader that one can

---

[20] Woodroofe does not include the factor $1/\hat{\theta}$, and his derivation is less general than Sweeting's

compare the Edgeworth expansion of the conditional density to the Laplace expansion of the posterior density. As noted by Johnson (1970), Ghosh et al. (1985) and Ghosh & Mukerjee (1991) among others, the Laplace expansion of the posterior has a simpler form than Edgeworth expansions in that it only depends on derivatives of the likelihood and prior. A Laplace expansion of the posterior density of the signed root statistic can be found in Bickel & Ghosh (1990) or, more accessibly, in Li (1998). However, the Laplace expansion is in terms of observed rather than expected likelihood quantities. Therefore, due to the approximation error in translating these quantities, we cannot compare the expansions directly to the order of error required in general.

2.4.4 Extension To Higher-Order Matching

There are several significant obstacles to extending this analysis to $O(n^{-3/2})$ to identify 3rd order PMPs. The first and most obvious is that the posterior Edgeworth expansion needs to be established as valid. The second obstacle is that the expansion of the signed root statistic in terms of S_1 and S_2 would not be possible to order $O(n^{-3/2})$ without introducing additional variables S_3 and S_4 corresponding to the third and fourth likelihood derivatives, in the spirit of Ryall's (1981) third-order ancillary. The introduction of these new variables poses a further complication in that the dimension of the sufficient statistic has increased, and so to perform inference on a scalar parameter, we could need a higher-dimensional ancillary statistic. The validity of the Edgeworth expansion, conditional on these new variables, would need to be established. Thus, there are many new theoretical results which are missing and which would be required to adopt the Edgeworth approach to 3rd order objective Bayesian analysis.

2.5 Comparison of Matching Conditions

Any attempt to study the relationship between these matching conditions is mired by significant challenges. The most obvious, though perhaps not the most daunting, is that three of the methods use expected likelihood quantities while the fourth (the saddlepoint approach) uses observed information quantities. A more general study based on the relationship between expected and observed likelihood geometries using likelihood yokes is ongoing, but is not included here. The second major problem in comparing these matching conditions is that when there are many parameters, it is more difficult to find algebraic simplifications and less realistic to justify the necessary assumptions which make it practical to compare the different matching methods. For simplicity, I will focus on the case where both the interest parameter is scalar. It is non-trivial to extend the results for vector interest parameters, both theoretically and in practice. This is the subject of ongoing research.

2.5.1 Obstacles to Comparison

There are two key difficulties here. The first is that the saddlepoint approach uses observed likelihood quantities while the other methods use expected likelihood quantities. A full investigation of how these matching conditions are related would require repeated application of the results of Barndorff-Nielsen & Blaesild (1993), who use likelihood yokes to derive the relationships between expected and observed likelihood derivatives. The problem, however, is that, in general, the standardized observed and expected likelihood quantities differ by order $O_p(n^{-1/2})$, both conditionally and unconditionally, which means, for instance, that

$$\hat{\mathring{\ell}}^{11} = \mathring{\omega}^{11} + O_p(n^{-1/2}), \quad \hat{\mathring{\ell}}^{12}(a) = \mathring{\omega}^{12} + O_p(n^{-1/2}),$$

and so forth. This will not allow us to achieve accurate results to order $O(n^{-3/2})$ unless there exist expansions of these quantities to higher order of accuracy. We were not able to obtain more accurate results in any of the exact ancillary examples considered.[21]

A second difficulty in comparing the saddlepoint approach to the other matching conditions is that the frequentist adjustment term U_F is a function of derivatives of the sample log-likelihood with respect to the MLE, holding the ancillary statistic fixed. These sample-space derivatives do not appear in the expected likelihood quantities. To relate the sample-

[21] Of course in exponential families with the canonical parameterization, the observed and expected information are equal, since the relevant log-likelihood derivatives are non-random. However, this is not generally true in curved exponential or transformation models. For example, Hinkley (1978) studies the variance of the MLE in location-scale models using both expected and observed information matrices and confirms numerically that they are different.

space derivatives to the expected likelihood quantities, we must rely on approximations to the sample space derivatives. Skovgaard (1996) and Severini (1998, 1999) proposed approximations based on covariances.[22] In principle, it is possible to derive approximations to sample-space derivatives which are accurate to the order $O(n^{-3/2})$.[23] However, the other approximations used–translation between observed and expected likelihood quantities–are only accurate to order $O_p(n^{-1/2})$ and thus there is no gain in overall accuracy in general unless we can improve the error in approximating expected likelihood quantities by their observed counterparts and vice versa. [24]

We now turn to the only setting where we have found it possible to compare the matching conditions.

2.5.2 The Shrinkage Argument vs. the Mean & Variance Adjusted Signed Root

First, we note that (2.5) and (2.10) are identical, which can be seen by re-expressing the r.h.s. of (2.10) using the conditional expression corresponding to Peers (1965) condition. This was presented at the end of Section 2.2.1.

We can also show that the 3rd order condition derived via the adjusted signed root is equivalent to the 3rd order condition obtained via the shrinkage arguement. To obtain the 3rd order condition derived from the shrinkage argument, note that if the prior density satisfies the second-order probability matching condition, then

[22] Fraser & Reid (1995) and Fraser et al. (1999) also studied expansions based on approximate ancillary statistics which are rather intractable in most settings, and less relevant for us as we consider exact ancillary models.

[23] This also seems to be missing from the literature, but is possible by extending the results of Skovgaard and Severini. It does not actually help us here, however, which is why it is omitted.

[24] Another good treatment of the covariance approach is given in Severini (2000) and Pierce & Bellio (2006).

$$\sum_r \frac{\partial\{\pi(\mathring{\sigma}^{rs}+\frac{1}{3}\mathring{\tau}^{rs})\mathring{\tau}^{tu}\mathring{\omega}_{stu}\}}{\partial\theta_r} = \mathring{\tau}^{rs}\mathring{\omega}^{tu}\mathring{\omega}_{rst}\pi_u + \pi(\mathring{\tau}^{rs}\mathring{\sigma}^{tu}+\frac{1}{3}\mathring{\tau}^{rs}\mathring{\tau}^{tu})\mathring{\omega}_{rst/u}$$
$$- \pi\mathring{\omega}^{ru}\mathring{\tau}^{st}\mathring{\omega}^{vw}\mathring{\omega}_{rst}\mathring{\omega}_{uv/w}$$
$$- \pi(\mathring{\omega}^{ru}\mathring{\sigma}^{sw}\mathring{\tau}^{tv}+\mathring{\sigma}^{ru}\mathring{\sigma}^{sw}\mathring{\tau}^{tv})\mathring{\omega}_{rst}\mathring{\omega}_{uv/w}$$

and

$$\sum_r\sum_s \frac{\partial^2(\pi\mathring{\tau}^{rs})}{\partial\theta_r\partial\theta_s} = -\mathring{\tau}^{rs}\pi_{rs} + \pi(\mathring{\tau}^{ru}\mathring{\omega}^{st}\mathring{\sigma}^{vw}+\frac{1}{2}\mathring{\tau}^{ru}\mathring{\tau}^{st}\mathring{\tau}^{vw})\mathring{\omega}_{rs/t}\mathring{\omega}_{uv/w}$$
$$+ \pi\mathring{\tau}^{ru}\mathring{\omega}^{sw}\mathring{\sigma}^{tv}\mathring{\omega}_{rs/t}\mathring{\omega}_{uv/w}.$$

Consequently, the 3rd order condition derived via the adjusted signed root can also be expressed as

$$\sum_r \frac{\partial\{\pi(\mathring{\sigma}^{rs}+\frac{1}{3}\mathring{\tau}^{rs})\mathring{\tau}^{tu}\mathring{\omega}_{stu}}{\partial\theta_r} + \sum_r\sum_s \frac{\partial^2(\pi\mathring{\tau}^{rs})}{\partial\theta_r\partial\theta_s} = 0,$$

which is the 3rd order condition derived via the shrinkage argument.

Chapter 3

Routes for Comparison of Conditional and Unconditional PMPs

We now discuss several routes for comparing conditional and unconditional PMPs. We have found the Edgeworth expansion to be useful for comparing conditional and unconditional asymptotics in simple one-parameter models, for which we give an example. Due to well-documented concerns about Edgeworth expansions for tail probabilities, we would prefer to study conditional and unconditional inference in the context of the saddlepoint approximation to the marginal and conditional density of the signed root statistic. This is currently not possible in general, though we briefly discuss how this could be done and illustrate the mechanics with an example.

3.1 Conditional and Unconditional Probability Matching PMPs

For the sake of completeness, we first argue that a prior which is conditional probability matching to some order will also be unconditional probability matching to the same order.

Theorem 3.1.1 *A conditional matching prior of order m is also an unconditional matching prior of order m.*

This may be shown by simple application of the law of iterated expectations. Consider

$$
\begin{aligned}
Pr_\theta\big[\psi \le \psi^{(1-\alpha)}(\pi, Y)\big] &= E_A Pr\big[\psi \le \psi^{(1-\alpha)}(\pi, Y)|A = a\big] \\
&= E_A\big[1 - \alpha + O(n^{-m/2})\big] \\
&= 1 - \alpha + O(n^{-m/2})
\end{aligned}
$$

.

To determine whether an unconditional matching prior delivers conditional matching to a given order of error, we need a condition relating the conditional and unconditional asymptotics. The analysis in the previous chapter for the adjusted signed root statistic led to matching conditions which, at least in one- and two-parameter models, are essentially conditions on the prior and the information of the model. We now seek to relate conditional and unconditional matching priors by deriving a relationship between the conditional and unconditional information.

Consider the expansion for the conditional density from the Edgeworth argument, to order $O(n^{-1})$. Recalling the notation from Chapter 2, we have that $S = S^* - \alpha_2/\sqrt{n}$ and $S = S_1 + (\alpha_1 S_1 S_2 + \alpha_2 S_1^2)/\sqrt{n}$. Let U be the score function from a single observation so that $\bar{U} = n^{-1}\ell_\theta$ is the sample average of the score function. First, we note that i_{20} is the information component at the true value for a single observation, but if we condition on an ancillary statistic, then we want the information from the entire likelihood using all observations. Let $-\omega_2 = E(\ell_\theta^2) = n i_{20} = nE(U^2)$, where the last two equalities follow because of the $i.i.d.$ assumption. This quantity is of order $O(n)$. Since α_1 and α_2 are arbitrary constants, we set them equal to zero. Then $S = S^* = S_1$. Using the definitions for these quantities, namely that $S_1 = \sqrt{n}\bar{S}_1$, $\bar{S}_1 = \bar{U}/\sqrt{i_{20}}$ and $\bar{U} = \ell_\theta/n$, we have that

$$
S_1 = \frac{\ell_\theta}{\sqrt{n}}\frac{1}{\sqrt{i_{20}}} = \frac{\ell_\theta}{\sqrt{-\omega_2}}.
$$

We are interested in the conditional information $\mathring{\omega}_2$, where

$$
\mathring{\omega}_2 = -E(\ell_\theta^2|a) = \omega_2 E(S_1^2|a).
$$

We have, plugging in for the cumulants and Hermite polynomials,

$$f(s|a) = \phi(s)\Big\{1 - \frac{\sigma a(s^2-1)}{2\sqrt{n}i_{20}} + \frac{i_{30}(s^3-3s)}{6\sqrt{n}i_{20}^{3/2}}\Big\} + O(n^{-1})$$

and integrating over s to get the conditional expectation of s^2 yields

$$E(S^2|a) = 1 - \frac{a\sigma}{\sqrt{n}i_{20}} + O(n^{-1}).$$

Plugging this in to our definition above, we have

$$\mathring{\omega}_2 = \omega_2\{1 - \frac{a\sigma}{\sqrt{n}i_{20}} + O(n^{-1})\},$$

where we again note that ω_2 is of order $O(n)$.

We recall that Efron's curvature is defined as

$$\frac{\sigma^2}{i_{20}^2} = \frac{(i_{02} - i_{20}^2 - \frac{i_{11}^2}{i_{20}})}{i_{20}^2} = \gamma^2,$$

and thus

$$\mathring{\omega}_2 = \omega_2\{1 - \frac{a\gamma}{\sqrt{n}} + O(n^{-1})\},$$

from which we can see that a prior which is *unconditional* probability matching to order $O(n^{-1})$ will also be *conditional* probability matching to order $O(n^{-1})$ if the Efron curvature is constant. This is because the term in the middle of the brackets will not depend on the parameter and thus can be taken outside the brackets. Therefore, we have that

$$\mathring{\omega}_2 \propto \omega_2$$

to $O(n^{-1})$ when the Efron curvature is constant. We can also see this by directly substituting the above relationship into the unconditional matching conditions derived either via the shrinkage argument or adjusted signed root in the unconditional setting for a scalar parameter model.

Taking this a step further, we can relate this more directly to probability matching priors. Using the shrinkage argument for the scalar parameter setting, we arrive at the condition for 2nd order unconditional matching

$$\begin{aligned}\frac{d\log\pi(\theta)}{d\theta} &= -\frac{1}{\omega_2^{-1/2}}\frac{d(\omega_2)^{-1/2}}{d\theta} \\ &= -\frac{d}{d\theta}\log(\omega_2)^{-1/2}.\end{aligned}$$

Integrating both sides, we have

$$\log\pi(\theta) = \log(\omega_2)^{1/2} + c,$$

where c is a constant. Therefore, the condition for $2nd$ order unconditional matching is

$$\pi(\theta) \propto (\omega_2)^{1/2}.$$

Repeating the exercise above for the conditional setting,

$$\frac{d\log\pi(\theta)}{d\theta} = -\frac{1}{(\mathring{\omega}_2)^{-1/2}}\frac{d(\mathring{\omega}_2)^{-1/2}}{d\theta} + O(n^{-1}),$$

and integrating both sides yields

$$\log\pi(\theta) = \log(\mathring{\omega}_2)^{1/2} + c + \frac{d}{n},$$

where c and d are constants. Therefore,

$$\pi(\theta) \propto (\mathring{\omega}_2)^{1/2}\exp(d/n) = (\mathring{\omega}_2)^{1/2}(1 + O(n^{-1})).$$

Thus we need

$$\mathring{\omega}_2 = \omega_2(1 + O(n^{-1}))$$

and

$$\omega_2 = \mathring{\omega}_2(1 + O(n^{-1})),$$

which are both satisfied when Efron's curvature is constant. Therefore, to establish that Jeffreys' prior is a 2nd order conditional PMP, we need only show that Efron's curvature is constant.

We now consider a simple example.

Example 3.1.1 *Exponential Hyperbola*
Consider the exponential hyperbola model where X_i and Y_i are distributed independently with exponential distributions. Further suppose that X_i has mean $1/\theta$ and Y_i has mean θ. The exact ancillary statistic for this model is given by $A = \sqrt{\bar{X}\bar{Y}}$ and the MLE is given by $\hat{\theta} = \sqrt{\bar{Y}/\bar{X}}$. We first consider the unconditional likelihood

$$\ell(\theta) = -n(\theta\bar{X} + \frac{\bar{Y}}{\theta})$$

and we calculate

$$i_{20}(\theta) = -i_{01}(\theta) = -E(\ell_{\theta\theta}(\theta)) = -E(-\frac{2n\bar{Y}}{\theta^3}) = \frac{2n}{\theta^2}.$$

It will be helpful to re-express Efron's curvature in the following way. For $j, k = 1, 2$, we have the moments

$$v_{jk} = E\Big\{\big[\ell_\theta(\theta)\big]^j \Big[(\ell_{\theta\theta}(\theta)) + E\big[\ell_\theta(\theta)\big]^2\Big]^k\Big\},$$

and then Efron's curvature $\gamma(\theta)$ is defined by

$$\gamma(\theta)^2 = \frac{v_{02}v_{20} - v_{11}^2}{v_{20}^2}. \tag{3.1}$$

In this example, we have $\gamma(\theta)^2 = 1/2$, which is obviously constant and therefore we have 2nd order unconditional matching priors will also be 2nd order conditional matching priors.

Example 3.1.2 *Cauchy Location Model*

As another example, consider the Cauchy location family. Let Y be random samples from the Cauchy location-model with location parameter θ and density

$$\frac{1}{\pi\left(1+\left[Y_i-\theta\right]^2\right)}$$

The exact ancillary statistic in this model is $A = (Y_1 - \hat{\theta}, \ldots, Y_n - \hat{\theta})$, so that the conditional log-likelihood may be expressed as:

$$\ell(\theta; \hat{\theta}, A) = -\sum \log\left\{\pi\left[1 + (A_i + \hat{\theta} - \theta)^2\right]\right\}$$

From this we can derive Efron's curvature to be $\gamma(\theta)^2 = 2.5$ which agrees with Efron & Hinkley (1978). Efron (1975) showed that, in fact, the Efron curvature is constant for any location model. Thus again, in any location model, a 2nd order unconditional PMP is also a 2nd order conditional PMP.

3.2 More On Conditional vs. Unconditional Inference

The previous section and its reliance on Efron's curvature and Edgeworth expansions is possible to extend to multi-dimensional parameter models, though as evidenced from the discussion, it is difficult even in one-parameter models. The saddlepoint machinery from the previous chapter is better-suited to handling multiple parameters, at least in the conditional setting for which there are results. The extension of the saddlepoint approximation to the marginal density of the signed root is the next step. The saddlepoint approach cannot currently be used in general models to compare conditional and unconditional matching priors. The reason is that there currently does not exist a marginal or unconditional version of the adjusted signed root likelihood ratio statistic of the R^* type, or otherwise. This is the subject of ongoing joint work with Jens Jensen at Aarhus University. Jensen (1995) gives a marginal version of a saddlepoint approximation to the signed root likelihood ratio statistic which is valid for some simple, low-dimensional parameter models such as a curved $(2, 1)$ exponential family. While it is a conventional saddlepoint approximation, it is not exactly the same approach as the R^* approximation of Barndorff-Nielsen (1986). Jensen is concerned with large-deviation properties to accommodate parameter values on the boundary of the support. This caveat must be noted before proceeding. We present an example to demonstrate that this route is analytically tractable and promising for future research, even if not particularly enlightening in the example considered.

We wish to investigate the relationship between the conditional and marginal adjustment terms of the saddlepoint approximation to the signed root statistic, with the disclaimer that, while we are comparing two saddlepoint approximations to the distribution of the same statistic and they are both accurate to at least $O(n^{-3/2})$, the marginal approximation is derived for large-deviation regions, whereas the conditional approximation is for small-deviation regions, such as one containing the MLE. Therefore, any comparison of these approximations and their resulting adjustment terms should not be interpreted as a straight comparison of conditional and marginal asymptotics. We again present the comparison for the exponential hyperbola model because it is the simplest curved $(2, 1)$ exponential family.[1]

Example 3.2.1 *Exponential Hyperbola (continued)*
Using the saddlepoint approximation to the marginal density of the signed root statistic as derived by Jensen (1995), we calculate that

$$Pr(R \leq r; \theta) = \Phi(r) + \phi(r)\Big[\frac{\theta^2 + \hat{\theta}^2}{\sqrt{2n}(\hat{\theta}^2 - \theta^2)} - \frac{1}{r} + O(n^{-3/2})\Big],$$

where we have the adjustment term given as the reciprocal of the first term in the brackets and to calculate conditional adjustment term, we use the likelihood quantities from the exponential hyperbola model,

$$\ell(\theta; \hat{\theta}, a) = -na(\frac{\theta}{\hat{\theta}} + \frac{\hat{\theta}}{\theta}),$$

and

$$\ell_{;\hat{\theta}}(\hat{\theta}) - \ell_{;\hat{\theta}}(\theta) = na\big(\frac{1}{\theta} - \frac{\theta}{\hat{\theta}^2}\big).$$

We calculate the conditional frequentist adjustment term,

[1] Kumon (2009) gives a nice explanation of how to construct, and ascertain the existence of, ancillary statistics in exponential transformation models, such as the gamma hyperbola in Example 3. These results are of particular interest as we study other curved exponential models in future work.

$$u_F = \frac{\sqrt{na}}{\sqrt{2}}\left[\frac{\hat{\theta}^2 - \theta^2}{\hat{\theta}\theta}\right]$$

and rearrange the unconditional frequentist adjustment term,

$$u_{UF} = \sqrt{2n}\left[\frac{\hat{\theta}^2 - \theta^2}{\hat{\theta}^2 + \theta^2}\right].$$

Since many of these orders of magnitude are raised to powers, we make use of the binomial series result that, for a variable x,

$$(1+x)^{-1/2} = 1 - \frac{1}{2}x + \frac{3}{8}x^2 + \dots$$

so, taking $x = O(n^{-1/2})$ yields

$$[1 + O(n^{-1/2})]^{-1/2} = 1 - \frac{1}{2}O(n^{-1/2}) + \dots$$

Now, using first-order theory for the MLE, we have that $\hat{\theta} = \theta + cn^{-1/2}$, where $c = O_p(1)$, so that if we take the ratio U_F/U_{UF}, we have

$$\frac{\frac{\sqrt{na}}{\sqrt{2}}\left[\frac{\hat{\theta}^2-\theta^2}{\hat{\theta}\theta}\right]}{\sqrt{2n}\left[\frac{\hat{\theta}^2-\theta^2}{\hat{\theta}^2+\theta^2}\right]}$$

and plugging in the first-order asymptotic result for the MLE, we have that

$$\begin{aligned}\frac{U_F}{U_{UF}} &= \frac{\sqrt{A}\left[2(\theta^2 + \theta cn^{-1/2}) + c^2n^{-1}\right]}{2(\theta^2 + \theta cn^{-1/2})} \\ &= \sqrt{A}[1 + O_p(n^{-1})]\end{aligned}$$

which can be seen from repeated application of the binomial series expansion. Thus we have that the ratio is $1 + O_p(n^{-1/2})$, since A converges in probability to a constant, or

$1 + O_p(n^{-1})$ *if* A *is fixed such that* $A = 1 + O(n^{-1})$.

Moreover, we note that

$$\frac{1}{U_{UF}} - \frac{1}{r} = O(n^{-1/2}),$$

and

$$\frac{1}{U_F} - \frac{1}{r} = O(n^{-1/2}).$$

Therefore, if $U_F = U_{UF} + O_p(n^{-1})$*, and* A *is fixed such that* $A = 1 + O(n^{-1})$*, then the marginal and conditional tail probabilities based on the adjusted signed root statistic will agree to order* $O(n^{-1})$.

We explore this example further in Chapter 4 via comparison to the Bayesian adjustment term.

Severini (1990) found that for the usual signed root statistic, conditional and unconditional inference agree to order $O(n^{-1})$. The result in the above example is for the adjusted signed root statistic, though we again note that caution should be exercised in the interpretation of this result. The conditional frequentist tail area expansion is for R^*, while the marginal tail area expansion is for some R^*-type statistic with good large-deviation properties but which is not the same as R^*. Nonetheless, as is apparent from the above example, this route for comparison of conditional and marginal frequentist asymptotics is promising and research on this topic is underway. In the next chapter, we demonstrate the desirable properties of the saddlepoint approach in particular ancillary statistic models, again confirming that of the four matching methods discussed in Chapter 2, the saddlepoint approach is the most practical to implement analytically in these settings.

3.3 Conditional and Unconditional Shrinkage Arguments

Lastly, we give a result relating conditional and unconditional PMPs via the shrinkage argument matching conditions.

This sort of argument makes use of the shortcut made possible by the results of Barndorff-Nielsen & Blaesild (1993) relating (for a single observation) conditional and unconditional expected (or observed) likelihood quantities to order $O_p(n^{-1/2})$ in general. Using the unconditional version of the shrinkage matching conditions, for each quantity ω_{rs}, we simply plug in the relationship that, conditional on A, $\omega_{rs} = \mathring{\omega}_{rs} + O(n^{-1/2})$, and similar relationships hold for the other quantities involved. For example, using the binomial series expansion given above, we have that when $\omega_{rs} = \mathring{\omega}_{rs} + O(n^{-1/2})$, then

$$\omega_{rs}^{-1/2} = \mathring{\omega}_{rs}[1 + O(n^{-1/2})].$$

Repeated application of this tool leads to a relationship between $\mathring{\Delta}_i$ and Δ_i for $i = 1, 2$. In particular, we have that, conditional on $A = a$, $\mathring{\Delta}_2 = \Delta_2 + O(n^{-1/2})$, which means that if we substitute the unconditional Δ_2 for the conditional $\mathring{\Delta}_2$, we have

$$n^{-1}\mathring{\Delta}_2 = n^{-1}[\Delta_2 + O(n^{-1/2})] = n^{-1}\Delta_2 + O(n^{-3/2})$$

Therefore to get 3rd order matching, what we really need is the *conditional* version of the 2nd order condition and either the *conditional* or *unconditional* version of the 3rd order condition.

Then the conditions required for 3rd order conditional matching are

$$\mathring{\Delta}_1 = 0$$

and

$$\Delta_2 = 0.$$

Chapter 4

Theoretical Examples

As evidenced above, the shrinkage, adjusted signed root and Edgeworth matching methods are very difficult to use analytically. The shrinkage argument has been studied in terms of numerical solutions of the partial differential equations by Levine & Casella (2003) and Sweeting (2005) and the analytical solution of these partial differential equations has been thoroughly summarized in Datta & Mukerjee (2004). The overwhelming evidence is that the partial differential equations are difficult to work with in practice. One would encounter the same problem working with the adjusted signed root matching conditions, which are also partial differential equations. Recently, Zhang (2008) studied numerical algorithms for the solution of these partial differential equations and then investigated the performance of the DiCiccio & Martin (1993) saddlepoint approximation to the posterior density of the signed root statistic. The results were very good, but there has still not been an investigation of priors identified using the saddlepoint matching conditions, with the notable exception of DiCiccio & Young (2010) for exponential families. In this chapter, we demonstrate that the saddlepoint approach is a comparatively simpler method which works well in all examples considered for the identification of conditional PMPs.

4.1 Location-Scale and Other Examples

The location-scale model specifies an i.i.d. sample, $Y = \{Y_1, \ldots, Y_n\}$, from the location-scale family

$$\sigma^{-1} f\left(\frac{y-\mu}{\sigma}\right),$$

where $f(\cdot)$ is a known density. In the location-scale model, the ancillary statistic is the configuration statistic, $A = \{A_1, \ldots, A_n\}$, with

$$A_i = \frac{Y_i - \hat{\mu}}{\hat{\sigma}}.$$

As shown in Barndorff-Nielsen & Cox (1994), when the density $f(\cdot)$ is symmetric, then σ and μ are expected orthogonal. In such settings, taking σ as the interest parameter and μ as a nuisance parameter, we have that when $\sigma - \hat{\sigma} = O_p(n^{-1/2})$, as is the case taking $\hat{\sigma}$ as the global maximum likelihood estimator, then

$$\hat{\mu}_\sigma = \hat{\mu} + O_p(n^{-1}),$$

where $\hat{\mu}_\sigma$ is the maximum likelihood estimator for μ given a fixed value of σ and $\hat{\mu}$ is the global MLE for μ. Without parameter orthogonality, maintaining that $\sigma - \hat{\sigma} = O_p(n^{-1/2})$, we have only that

$$\hat{\mu}_\sigma = \hat{\mu} + O_p(n^{-1/2}).$$

Therefore, symmetry in the location-scale model, which ensures parameter orthogonality with respect to the Fisher information, is particularly useful in achieving higher-order matching. In some special cases, such as when $f(\cdot)$ is a normal distribution, due to Basu's Theorem, we have that $\hat{\mu}_\sigma = \hat{\mu}$ exactly. This affords considerable simplification in the algebra, making it easier to identify matching priors.[1]

We now apply the saddlepoint matching method to the general location-scale model. To keep notation consistent throughout, we let $\psi = \sigma$ and $\lambda = \mu$. The log-likelihood function

[1] Also noted by Staicu & Reid (2008).

has the form

$$\ell(\theta) = -n\log\psi + \sum \log f\left(\frac{y_i - \lambda}{\psi}\right).$$

Define $g(\cdot) = -\log f(\cdot)$ and rewriting in terms of the ancillary statistic, we have

$$\ell(\theta;\hat{\theta},a) = -n\log\psi - \sum g\left\{\frac{\hat{\psi}}{\psi}\left(A_i + \frac{\hat{\lambda}-\lambda}{\hat{\psi}}\right)\right\}.$$

Let $\delta_i = \frac{\hat{\psi}}{\psi}\left(A_i + \frac{\hat{\lambda}-\lambda}{\hat{\psi}}\right)$ so that $A_i + \frac{\hat{\lambda}-\lambda}{\hat{\psi}} = \frac{\psi}{\hat{\psi}}\delta_i$, and also let $\tilde{\delta}_i = \frac{\hat{\psi}}{\psi}\left(A_i + \frac{\hat{\lambda}-\hat{\lambda}_\psi}{\hat{\psi}}\right)$. To calculate the adjustment terms in the saddlepoint approach, we will use the following quantities

$$\ell_\psi = -\frac{n}{\psi} + \sum \frac{1}{\psi}\delta_i g'(\delta_i), \quad \ell_\lambda = \sum \frac{1}{\psi}g'(\delta_i),$$

where we note that the maximum likelihood conditions and their implications are:

$$\ell_\psi(\hat{\theta}) = -\frac{n}{\hat{\psi}} + \sum \frac{A_i}{\hat{\psi}}g'(A_i) = 0,$$

which implies $\sum A_i g'(A_i) = n$;

$$\ell_\lambda(\hat{\theta}) = \sum \frac{1}{\hat{\psi}}g'(A_i) = 0,$$

which implies $\sum g'(A_i) = 0$; and

$$\ell_\lambda(\psi,\hat{\lambda}_\psi) = \sum \frac{1}{\psi}g'(\tilde{\delta}_i) = 0,$$

which implies $\sum g'(\tilde{\delta}_i) = 0$. These conditions will be used repeatedly in the calculations that follow to obtain simplifications. Now,

$$\ell_{\psi\psi}(\hat{\theta}) = \sum -\frac{A_i^2}{\hat{\psi}^2}g''(A_i),$$

$$\ell_{\lambda\lambda}(\hat{\theta}) = \sum -\frac{1}{\hat{\psi}^2} g''(A_i),$$

$$\ell_{\psi\lambda}(\hat{\theta}) = \ell_{\lambda\psi}(\hat{\theta}) = \sum -\frac{A_i}{\hat{\psi}^2} g''(A_i),$$

$$\ell_{\psi}(\psi, \hat{\lambda}_\psi) = -\frac{n}{\psi} + \sum \frac{\tilde{\delta}_i}{\psi} g'(\tilde{\delta}_i) = -\frac{n}{\psi} + \sum \frac{\hat{\psi}}{\psi^2} A_i g'(\tilde{\delta}_i),$$

$$\ell_{;\hat{\theta}}\big(\hat{\psi}, \hat{\lambda}\big) = \begin{pmatrix} -\frac{n}{\hat{\psi}} \\ 0 \end{pmatrix},$$

$$\ell_{;\hat{\theta}}\big(\psi, \hat{\lambda}_\psi\big) = \begin{pmatrix} \sum -\frac{A_i}{\psi} g'(\tilde{\delta}_i) \\ 0 \end{pmatrix},$$

$$\ell_{\lambda;\hat{\theta}}\big(\psi, \hat{\lambda}_\psi\big) = \begin{pmatrix} \ell_{\lambda;\hat{\psi}}(\psi, \hat{\lambda}_\psi) \\ \ell_{\lambda;\hat{\lambda}}(\psi, \hat{\lambda}_\psi) \end{pmatrix} = \begin{pmatrix} \sum \frac{A_i}{\psi^2} g''(\tilde{\delta}_i) \\ \sum \frac{1}{\psi^2} g''(\tilde{\delta}_i) \end{pmatrix}.$$

Now,

$$\ell_{\theta\theta}\big(\hat{\psi}, \hat{\lambda}\big) = \begin{pmatrix} \sum -\frac{A_i^2}{\hat{\psi}^2} g''(A_i) & -\sum \frac{A_i}{\hat{\psi}^2} g''(A_i) \\ -\sum \frac{A_i}{\hat{\psi}^2} g''(A_i) & -\sum \frac{1}{\hat{\psi}^2} g''(A_i) \end{pmatrix},$$

and we note in particular that $\ell_{\lambda\lambda}\big(\psi, \hat{\lambda}_\psi\big) = -\sum \frac{1}{\psi^2} g''(\tilde{\delta}_i)$ is a submatrix of $\ell_{\theta\theta}\big(\psi, \hat{\lambda}_\psi\big)$. Now, the determinant of the partitioned matrix is

$$\big|\ell_{;\hat{\theta}}\big(\hat{\psi}, \hat{\lambda}\big) - \ell_{;\hat{\theta}}\big(\psi, \hat{\lambda}_\psi\big) \quad \ell_{\lambda;\hat{\theta}}\big(\psi, \hat{\lambda}_\psi\big)\big| = \big[-\frac{n}{\hat{\psi}} - \sum \frac{A_i}{\psi} g'(\tilde{\delta}_i)\big]\big[\sum \frac{1}{\psi^2} g''(\tilde{\delta}_i)\big].$$

Therefore, we have the adjustment terms, ignoring the $|\ell_{\theta\theta}(\hat{\theta})|^{1/2}$ term which is common to both U_F and U_B,

$$U_F = \frac{\big[-\frac{n}{\hat{\psi}} + \sum \frac{A_i}{\psi} g'(\tilde{\delta}_i)\big]\big[\sum \frac{1}{\psi^2} g''(\tilde{\delta}_i)\big]}{\Big(\sum \frac{1}{\psi^2} g''(\tilde{\delta}_i)\Big)^{1/2}}$$

and

$$U_B = \Big[-\frac{n}{\psi} + \sum \frac{\hat{\psi}}{\psi^2} A_i g'(\tilde{\delta}_i) \Big] \Big(\sum \frac{1}{\psi^2} g''(\tilde{\delta}_i) \Big)^{1/2} \frac{\pi(\hat{\theta})}{\pi(\psi, \hat{\lambda}_\psi)}.$$

Now, setting the adjustment terms equal, we have that when

$$\frac{\pi(\hat{\theta})}{\pi(\psi, \hat{\lambda}_\psi)} \propto \frac{\psi}{\hat{\psi}}$$

then the Bayesian credible set has conditional frequentist coverage probability $1 - \alpha$ to error of order $O(n^{-3/2})$. In the original parameterization of the model, this implies that a third-order probability matching prior is of the form

$$\pi(\theta) \propto \frac{1}{\sigma}.$$

Fernandez & Steel (1999) showed that this prior is Bernardo's reference prior as well as independence Jeffreys prior, though not the usual Jeffreys prior.[2] This prior is the right-invariant Haar prior, whereas as Jeffreys prior is the left-invariant Haar prior, $\pi(\theta) \propto \frac{1}{\sigma^2}$.[3]

We now show how this matching method works in a few specific examples

Example 4.1.1 *Normal Location-Scale*

Consider a location-scale model with normal density and let $\psi = \sigma^2$ *and* $\lambda = \mu$. *Then the log-likelihood function has the form*

$$\ell(\theta; \hat{\theta}, a) = -\frac{1}{2} n \Big\{ \log \psi + \frac{\hat{\psi}}{\psi} + \frac{(\hat{\lambda} - \lambda)^2}{\psi} \Big\}.$$

To calculate the adjustment terms, we will use the following quantities

$$\ell_\psi = -\frac{n}{2} \Big\{ \frac{1}{\psi} - \frac{\hat{\psi}}{\psi^2} - \frac{(\hat{\lambda} - \lambda)^2}{\psi^2} \Big\}, \quad \ell_\lambda = \frac{n(\hat{\lambda} - \lambda)}{\psi},$$

$$\ell_{\psi\psi} = -\frac{n}{2} \Big\{ -\frac{1}{\psi^2} + \frac{2\hat{\psi}}{\psi^3} + \frac{2(\hat{\lambda} - \lambda)^2}{\psi^3} \Big\}, \quad \ell_{\lambda\lambda} = -\frac{n}{\psi},$$

---

[2]The independence Jeffreys prior assumes that the parameters are *a priori* independent and as such the independence Jeffreys prior is the product of the independent Jeffreys rule priors for each parameter.

[3]Peers (1965) claims that in a location-scale model, the choice of $\pi \propto 1/\sigma$ yields exact matching, a result which we prove later in this chapter.

$$\ell_{\psi\lambda} = \ell_{\lambda\psi} = -\frac{n(\hat{\lambda}-\lambda)}{\psi^2}, \quad \ell_{\psi}(\psi, \hat{\lambda}_{\psi}) = -\frac{n}{2}\left\{\frac{1}{\psi} - \frac{\hat{\psi}}{\psi^2} - \frac{(\hat{\lambda}-\hat{\lambda}_{\psi})^2}{\psi^2}\right\},$$

$$\ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}) = \begin{pmatrix} -\frac{n}{2\hat{\psi}} \\ 0 \end{pmatrix}, \quad \ell_{;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) = \begin{pmatrix} -\frac{n}{2\psi} \\ \frac{n(\hat{\lambda}-\hat{\lambda}_{\psi})}{\psi} \end{pmatrix},$$

$$\ell_{\lambda;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) = \begin{pmatrix} 0 \\ \frac{n}{\psi} \end{pmatrix}, \quad \ell_{\theta\theta}(\hat{\psi}, \hat{\lambda}) = \begin{pmatrix} -\frac{n}{2\hat{\psi}^2} & 0 \\ 0 & -\frac{n}{\hat{\psi}} \end{pmatrix},$$

where we note that $\ell_{\lambda\lambda}(\psi, \hat{\lambda}_{\psi})$ is a submatrix of $\ell_{\theta\theta}(\psi, \hat{\lambda}_{\psi})$. Using these quantities, we have that

$$\left| -\ell_{\theta\theta}(\hat{\psi}, \hat{\lambda})\right| = \frac{n^2}{2\hat{\psi}^3}$$

and

$$\left|\ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}) - \ell_{;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) \quad \ell_{\lambda;\hat{\theta}}(\psi, \hat{\lambda}_{\psi})\right| = \frac{n^2}{2\psi^2} - \frac{n^2}{2\psi\hat{\psi}}.$$

This yields the two adjustment terms

$$U_F = \frac{\frac{n^2(\hat{\psi}-\psi)}{2\psi^2\hat{\psi}}}{\left\{\frac{n^3}{2\psi\hat{\psi}^3}\right\}^{1/2}}$$

and

$$U_B = \left[\frac{n(\hat{\psi}-\psi)}{2\psi^2}\right] \frac{(\frac{n}{\psi})^{1/2}}{(\frac{n^2}{2\hat{\psi}^3})^{1/2}} \frac{\pi(\hat{\theta})}{\pi(\psi, \hat{\lambda}_{\psi})},$$

where we have used the fact that in the normal case, the global and constrained MLE are the same, thus $\hat{\lambda} - \hat{\lambda}_{\psi}$=0. Setting the adjustment terms equal, we have that when

$$\frac{\pi(\hat{\theta})}{\pi(\psi, \hat{\lambda}_{\psi})} \propto \frac{\psi}{\hat{\psi}},$$

then such a prior is third-order conditional probability matching. This implies that the prior is of the form

$$\pi(\theta) \propto \frac{1}{\sigma^2}.$$

Recall that here, to simplify the calculations, we have taken the interest parameter to be σ^2. Therefore if we instead take σ as the interest parameter, then we have that the prior

$$\pi(\theta) \propto \frac{1}{\sigma}$$

is identified as being third-order probability matching.

While the location-scale model with normal distribution is an ancillary statistic model, it is a special case. Under the normal distribution, the MLE is a sufficient statistic and thus there is no need to condition on an ancillary statistic. The results has been included for illustrative purposes.

Example 4.1.2 *Gumbel Location-Scale*

Again consider a location-scale model but with Gumbel, or Extreme Value Type I, distribution, having the density function

$$\sigma^{-1} \exp\left\{\frac{\mu - y}{\sigma} - \exp\left(\frac{\mu - y}{\sigma}\right)\right\}.$$

Let $\psi = \sigma$ and $\lambda = \mu$. The log-likelihood function is

$$\ell(\psi, \lambda; \hat{\psi}, \hat{\lambda}, a) = -n \log \psi + \sum_{i=1}^{n} \frac{\lambda - \hat{\lambda} - \hat{\psi} a_i}{\psi} - \exp\left(\frac{\lambda - \hat{\lambda} - \hat{\psi} a_i}{\psi}\right).$$

We again suppress the limits of the sums. To calculate the adjustment terms, we will use the following quantities,

$$\ell_\psi = \sum -\frac{n}{\psi} - \frac{1}{\psi^2}(\lambda - \hat{\lambda} - \hat{\psi} a_i) + \frac{1}{\psi^2}(\lambda - \hat{\lambda} - \hat{\psi} a_i) \exp\left(\frac{\lambda - \hat{\lambda} - \hat{\psi} a_i}{\psi}\right),$$

$$\ell_\lambda = \sum \frac{1}{\psi}\left[1 - \exp\left(\frac{\lambda - \hat{\lambda} - \hat{\psi} a_i}{\psi}\right)\right],$$

where we note that the maximum likelihood conditions $\ell_\psi(\hat{\theta}) = 0$, $\ell_\lambda(\hat{\theta}) = 0$ and $\ell_\lambda(\psi, \hat{\lambda}_\psi) = 0$ imply the following three results, respectively,

$$\sum a_i - a_i \exp(-a_i) = n,$$

$$\sum \exp(-a_i) = 1,$$

and

$$\sum \exp\big(\frac{\hat{\lambda}_\psi - \hat{\lambda} - \hat{\psi} a_i}{\psi}\big) = 1.$$

These results are used repeatedly to obtain the quantities which follow. For convenience, we will let $B_i = \frac{\hat{\lambda}_\psi - \hat{\lambda} - \hat{\psi} a_i}{\psi}$. Now,

$$\ell_{\psi\psi}(\hat{\theta}) = \frac{3n-1}{\hat{\psi}^2},$$

$$\ell_{\lambda\lambda}(\hat{\theta}) = -\frac{1}{\hat{\psi}^2},$$

$$\ell_{\psi\lambda}(\hat{\theta}) = \ell_{\lambda\psi}(\hat{\theta}) = \sum -\frac{a_i}{\hat{\psi}^2} \exp(-a_i),$$

$$\ell_\psi(\psi, \hat{\lambda}_\psi) = -\frac{n}{\psi} + \sum \frac{\hat{\psi}}{\psi^2} a_i[1 - \exp(B_i)],$$

$$\ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}) = \begin{pmatrix} \frac{-n}{\hat{\psi}} \\ 0 \end{pmatrix},$$

$$\ell_{;\hat{\theta}}(\psi, \hat{\lambda}_\psi) = \begin{pmatrix} \sum -\frac{a_i}{\psi}[1 - \exp(B_i)] \\ 0 \end{pmatrix},$$

$$\ell_{\lambda;\hat{\theta}}(\psi, \hat{\lambda}_\psi) = \begin{pmatrix} \sum \frac{a_i}{\psi^2} \exp(B_i) \\ \frac{1}{\psi^2} \end{pmatrix},$$

and where we note that

$$\ell_{\lambda\lambda}(\psi, \hat{\lambda}_\psi) = -\frac{1}{\psi^2}$$

is a submatrix of $\ell_{\theta\theta}(\psi, \hat{\lambda}_\psi)$*. We can now calculate that*

$$\left| \ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}) - \ell_{;\hat{\theta}}(\psi, \hat{\lambda}_\psi) \quad \ell_{\lambda;\hat{\theta}}(\psi, \hat{\lambda}_\psi) \right| = \frac{1}{\psi^2}\big[-\frac{n}{\hat{\psi}} + \sum \frac{a_i}{\psi}[1 - \exp(B_i)]\big]$$

Thus we have the adjustment terms, ignoring the term which is common to both U_F *and* U_B*,*

$$U_F = \frac{\frac{1}{\psi^2}\big[-\frac{n}{\hat{\psi}} + \sum \frac{a_i}{\psi}[1 - \exp(B_i)]\big]}{\left(\frac{1}{\psi^2}\right)^{1/2}}$$

and

$$U_B = \Big[-\frac{n}{\psi^2} + \sum \frac{\hat{\psi}}{\psi^3} a_i [1 - \exp(B_i)]\Big] \frac{\pi(\hat{\theta})}{\pi(\psi, \hat{\lambda}_\psi)}$$

Setting these adjustment terms equal, we identify the prior

$$\pi(\theta) \propto \frac{1}{\sigma}$$

as being third-order conditional probability matching.

Example 4.1.3 *Exponential Regression*

Now consider the exponential regression model. Suppose we have a sample of independently and exponentially distributed variables with mean $\lambda e^{-\psi z_i}$ *and covariates* z_i *with* $\sum_{i=1}^n z_i = 0$*. As noted by Cox & Reid (1989), the parameters in this model are orthogonal. The log-likelihood is of the form*

$$\ell(\psi, \lambda; \hat{\psi}, \hat{\lambda}, a) = -n \log \lambda - n\hat{\lambda}_\psi \lambda^{-1},$$

where $a = (a_1, \ldots, a_n)$ *is exactly ancillary with* $a_i = \log y_i - \log \hat{\lambda} + \psi z_i$, $\hat{\psi}$ *satisfies the equation* $\sum_{i=1}^{n} z_i y_i e^{\hat{\psi} z_i} = 0$ *and*

$$\hat{\lambda}_\psi = \frac{1}{n} \sum_{i=1}^{n} y_i e^{\psi z_i} = \frac{1}{n} \sum_{i=1}^{n} \hat{\lambda} \exp\left\{a_i + (\psi - \hat{\psi}) z_i\right\}.$$

Suppressing limits of sums and plugging in the ancillary statistic, we have

$$\ell(\theta; \hat{\theta}, a) = -n \log \lambda - \frac{1}{\lambda} \sum \hat{\lambda} \exp\left\{a_i + (\psi - \hat{\psi}) z_i\right\},$$

we will define $c_i = a_i + (\psi - \hat{\psi}) z_i$ *and use the following quantities to calculate the adjustment terms*

$$\ell_\psi = \sum -\frac{\hat{\lambda}}{\lambda} z_i \exp(c_i), \quad \ell_\lambda = -\frac{n}{\lambda} + \sum \frac{\hat{\lambda}}{\lambda^2} \exp(c_i),$$

where we note the maximum likelihood conditions and their implications are

$$\ell_\psi(\hat{\theta}) = \sum -z_i \exp(a_i) = 0,$$

which implies $\sum z_i \exp(a_i) = 0$*;*

$$\ell_\lambda(\hat{\theta}) = -\frac{n}{\hat{\lambda}} + \sum \frac{1}{\hat{\lambda}} \exp(a_i) = 0,$$

which implies $\sum \exp(a_i) = n$*; and*

$$\ell_\lambda(\psi, \hat{\lambda}_\psi) = -\frac{n}{\hat{\lambda}_\psi} + \sum \frac{\hat{\lambda}}{\hat{\lambda}_\psi^2} \exp(c_i) = 0,$$

which implies $\sum \exp(c_i) = n\hat{\lambda}_\psi / \hat{\lambda}$*. These conditions will be used repeatedly in the following calculations to obtain simplifications. We further note that the term* $|\ell_{\theta\theta}(\hat{\theta})|^{1/2}$ *is common to both* U_B *and* U_F *and cancels out when setting the adjustment terms equal. Thus we can ignore this term in our calculations. The remaining quantities necessary for identifying the* 3*rd order conditional PMP are*

$$\ell_{\psi}(\psi, \hat{\lambda}_{\psi}) = \sum -\frac{\hat{\lambda}}{\hat{\lambda}_{\psi}} z_i \exp(c_i),$$

$$\ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}) = \begin{pmatrix} 0 \\ -\frac{n}{\hat{\lambda}} \end{pmatrix},$$

$$\ell_{;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) = \begin{pmatrix} \sum \frac{\hat{\lambda}}{\hat{\lambda}_{\psi}} z_i \exp(c_i) \\ -\frac{n}{\hat{\lambda}} \end{pmatrix},$$

$$\ell_{\lambda;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) = \begin{pmatrix} \sum -\frac{\hat{\lambda}}{\hat{\lambda}_{\psi}^2} z_i \exp(c_i) \\ \frac{n}{\hat{\lambda}\hat{\lambda}_{\psi}} \end{pmatrix},$$

and we note that

$$\ell_{\lambda\lambda}(\psi, \hat{\lambda}_{\psi}) = -\frac{n}{\hat{\lambda}_{\psi}^2}$$

is a submatrix of $\ell_{\theta\theta}(\psi, \hat{\lambda}_{\psi})$. Thus the determinant of the partitioned matrix is equal to

$$\left| \ell_{;\hat{\theta}}(\hat{\psi}, \hat{\lambda}) - \ell_{;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) \quad \ell_{\lambda;\hat{\theta}}(\psi, \hat{\lambda}_{\psi}) \right| = \left[\frac{n}{\hat{\lambda}\hat{\lambda}_{\psi}} \right] \left[\sum -\frac{\hat{\lambda}}{\hat{\lambda}_{\psi}} z_i \exp(c_i) \right]$$

and, ignoring the term common to both U_F and U_B, the adjustment terms are given by

$$U_F = \frac{\left[\frac{n}{\hat{\lambda}\hat{\lambda}_{\psi}} \right] \left[\sum -\frac{\hat{\lambda}}{\hat{\lambda}_{\psi}} z_i \exp(c_i) \right]}{\left(\frac{n}{\hat{\lambda}_{\psi}^2} \right)^{1/2}},$$

and

$$U_B = \left[\sum -\frac{\hat{\lambda}}{\hat{\lambda}_{\psi}} z_i \exp(c_i) \right] \left(\frac{n}{\hat{\lambda}_{\psi}^2} \right)^{1/2} \frac{\pi(\hat{\theta})}{\pi(\psi, \hat{\lambda}_{\psi})}.$$

Now, by setting the adjustment terms equal, we identify the prior

$$\pi(\theta) \propto \frac{1}{\lambda}$$

as being third-order probability matching. This is also a reference prior.[4]

[4] Ye & Berger (1991) studied reference priors for exponential regression models and found that there are

As noted by Lawless (1976), Fraser & Reid (1995) and Severini (2000), the exponential regression model is a composite transformation model which may be rewritten as a location model on the log scale, with scale parameter equal to 1*. To do so, let* $\lambda = \exp\{\gamma + \beta z_i\}$*, so that* $\log \lambda = \gamma + \beta z_i$*, where* $\gamma, \beta \epsilon \Re$ *are unknown scalar parameters. This is a composite transformation model with group paramater* γ *and index* β*. Plugging this into our model yields the result.*

Example 4.1.4 *Exponential Hyperbola*

The exponential hyperbola model was introduced in the previous chapter. Here we give more details of the calculations and identify the conditional PMP. Let

$$\ell(\theta; \hat{\theta}, a) = -na(\frac{\theta}{\hat{\theta}} + \frac{\hat{\theta}}{\theta})$$

and therefore we have,

$$\ell_{;\hat{\theta}}(\hat{\theta}) - \ell_{;\hat{\theta}}(\theta) = na\big(\frac{1}{\theta} - \frac{\theta}{\hat{\theta}^2}\big)$$

Then we can calculate the conditional frequentist adjustment term

$$u_F = \frac{\sqrt{na}}{\sqrt{2}} \left[\frac{\hat{\theta}}{\theta} - \frac{\theta}{\hat{\theta}} \right]$$

and the Bayesian adjustment term

$$u_B = \frac{\sqrt{na}}{\sqrt{2}} \big[\big(\frac{\hat{\theta}}{\theta}\big)^2 - 1 \big] \frac{\pi(\hat{\theta})}{\pi(\theta)}.$$

By setting these terms equal, we identify the prior $\pi(\theta) \propto 1/\theta$ *as being third-order probability matching.*

4.2 Exact Matching in Location-Scale Models

The result that there is exact matching for a scalar parameter of interest in the location-scale model for the prior $\pi(\theta) \propto 1/\sigma$ has been mentioned in several previous works, such

---

many.

as Peers (1965), Lawless (1982) and DiCiccio & Martin (1993). However, these authors did not formally prove this result. A more general result about exact matching for predictive regions using right-invariant Haar priors in group transformation models may be found in Severini et al. (2002). This result is due to an invariance property of the highest predictive density region, and is essentially an extension of invariance results derived in Hora & Buehler (1966) and Hora & Buehler (1967).

Here we present a proof which has the advantages of being more transparent, more intuitive and more directly suited to the problem of probability matching for inference.

Theorem 4.2.1 *The prior* $\pi(\theta) \propto 1/\sigma$ *yields exact conditional probability matching in location-scale models.*

Proof: Given a sample $\{Y_1, \ldots, Y_n\}$ from the family $\sigma^{-n} f\{(y-\mu)/\sigma\}$, we define

$$p(y; \mu, \sigma) = \sigma^{-n} \prod_{i=1}^{n} f\{(y-\mu)/\sigma\},$$

where $p(\cdot)$ is a probability density function and $f(\cdot)$ is a known probability density function defined on $\Re$. It is assumed that the maximum likelihood estimators for (μ, σ) are $(\hat{\mu}, \hat{\sigma})$ which are unique and exist with probability one. As noted before, the ancillary statistic for this model is the configuration statistic, defined as

$$A = (A_1, \ldots, A_n) = \left(\frac{Y_1 - \hat{\mu}}{\hat{\sigma}}, \ldots, \frac{Y_n - \hat{\mu}}{\hat{\sigma}} \right).$$

This statistic is distribution constant, and a transformation of the result shown by Fisher (1934), letting $(Q_1, Q_2) = \left(\frac{\hat{\mu}-\mu}{\hat{\sigma}}, \frac{\hat{\sigma}}{\sigma} \right)$, yields that the exact conditional joint distribution of (Q_1, Q_2) is

$$p(y; q_1, q_2 | a) = c(a) q_2^{n-2} p(y; q_1, q_2),$$

where $c(a)$ is a normalizing constant. Rearranging, we have

$$p(y; q_1, q_2 | a) = c(a) q_2^{n-1} \prod_{i=1}^{n} f(q_1 q_2 + q_2 a_i).$$

Since $p(\cdot)$ is a proper density function, the normalizing constant $c(a)$ is defined by

$$c(a)\int_0^\infty\int_{-\infty}^\infty q_2^{n-1}\prod_{i=1}^n f(q_1q_2+q_2a_i)dq_1dq_2=1$$

or

$$c(a)^{-1}=\int_0^\infty\int_{-\infty}^\infty q_2^{n-1}\prod_{i=1}^n f(q_1q_2+q_2a_i)dq_1dq_2$$

From a Bayesian perspective, the joint posterior is

$$\pi(\mu,\sigma|Y=y)\propto\pi(\mu,\sigma)\exp\left\{\log\sigma^{-n}\prod_{i=1}^n f\{\frac{\hat\sigma}{\sigma}(a_i+\frac{\hat\mu-\mu}{\hat\sigma})\}\right\}.$$

Using the prior identified by our matching method, $\pi\propto 1/\sigma$, we have

$$\begin{aligned}\pi(\mu,\sigma|Y=y)&\propto\frac{1}{\sigma}\exp\left\{\log\sigma^{-n}\prod_{i=1}^n f\{\frac{\hat\sigma}{\sigma}(a_i+\frac{\hat\mu-\mu}{\hat\sigma})\}\right\}\\&\propto s\sigma^{-n-1}\prod_{i=1}^n f\{\frac{\hat\sigma}{\sigma}(a_i+\frac{\hat\mu-\mu}{\hat\sigma})\}.\end{aligned}$$

The normalizing constant s is determined by

$$s\int_0^\infty\int_{-\infty}^\infty\sigma^{-n-1}\prod_{i=1}^n f\{\frac{\hat\sigma}{\sigma}(a_i+\frac{\hat\mu-\mu}{\hat\sigma})\}d\mu d\sigma=1.$$

Now, note that the marginal posterior for μ is $\pi(\mu|Y=y)=\int_0^\infty\pi(\mu,\sigma|Y=y)d\sigma$ which, upon making the substitution $q_2=\hat\sigma/\sigma$ equals

$$\frac{s}{\hat\sigma^n}\int_0^\infty q_2^{n-1}\prod_{i=1}^n f(\frac{\hat\mu-\mu}{\hat\sigma}q_2+q_2a_i)dq_2.$$

This marginal posterior for μ integrates to one, so we have that

$$\frac{s}{\hat\sigma^n}\int_{-\infty}^\infty\int_0^\infty q_2^{n-1}\prod_{i=1}^n f(\frac{\hat\mu-\mu}{\hat\sigma}q_2+q_2a_i)dq_2d\mu=1.$$

Upon making the substitution $q_1=(\hat\mu-\mu)/\hat\sigma$ we find the relationship between the normalizing constants:

$$s \equiv c(a)\hat{\sigma}^{n-1}.$$

The Bayesian $1-\alpha$ confidence limit, $\beta \equiv \beta(y)$, for σ is such that

$$\int_0^\beta \pi(\sigma|Y=y)d\sigma = 1-\alpha$$

in terms of the marginal posterior $\pi(\sigma|Y=y)$ for σ. Using the form of the joint posterior and making the substitution $(\mu,\sigma) \to (q_1,q_2)$ we have that

$$\int_0^\beta \int_{-\infty}^\infty \pi(\mu,\sigma|Y=y)d\mu d\sigma = 1-\alpha$$

which gives

$$c(a)\int_{\frac{\hat{\sigma}}{\beta}}^\infty \int_{-\infty}^\infty q_2^{n-1}\prod_{i=1}^n f(q_1q_2+q_2a_i)dq_1dq_2 = 1-\alpha \tag{4.1}$$

from which we can see that as $\hat{\sigma}$ increases, so does β and vice versa.

Fix σ. The event $\{\sigma \leq \beta\} \equiv \{\hat{\sigma} \geq \hat{\sigma}_0\}$ where $\hat{\sigma}_0$ is the value of $\hat{\sigma}$ for which $\beta \equiv \sigma$. Then $Pr\{\sigma \leq \beta|a\} \equiv Pr\{\hat{\sigma} \geq \hat{\sigma}_0;\sigma|a\}$ and this conditional frequentist coverage is

$$Pr_{\mu,\sigma}\{\sigma \leq \beta|a\} = Pr_{\mu,\sigma}\{Q_2 \geq \hat{\sigma}_0;\sigma/\hat{\sigma}|a\}.$$

Using the result above, this is equal to

$$\int_{\frac{\hat{\sigma}_0}{\sigma}}^\infty \int_{-\infty}^\infty p(y;q_1,q_2|a)dq_1dq_2,$$

which gives

$$c(a)\int_{\frac{\hat{\sigma}_0}{\sigma}}^\infty \int_{-\infty}^\infty q_2^{n-1}\prod_{i=1}^n f(q_1q_2+q_2a_i)dq_1dq_2$$

but the Bayes limit $\beta \equiv \sigma$ when $\hat{\sigma} = \hat{\sigma}_0$. Therefore, by equation (4.1) , $Pr\{\sigma \leq \beta|a\} \equiv 1-\alpha$, i.e. the Bayes limit is exact conditional frequentist probability matching.

■

Chapter 5

Numerical Examples

For the first 40 years after Welch & Peers (1963) proposed conditions to identify PMPs, there was no serious attempt to investigate the numerical solution of these conditions nor was there any numerical study of the performance of PMPs identified by matching methods. Levine & Casella (2003) and Sweeting (2005) were the first to study numerical solution of the partial differential equations arising from the Welch & Peers (1963) and shrinkage matching methods.

We will not be concerned here with the numerical solution of the matching conditions, but rather the performance of priors identified by the matching conditions. Thus we take as given the the first and more difficult step and simply ask the question, ”How do matching priors perform?”

There have thus far been three somewhat related assessments of the performance of matching priors. Levine & Casella (2003) presented an algorithm for the implementation of PMPs by solving the partial differential equations derived via the shrinkage approach for the case of a scalar interest parameter in the presence of a scalar nuisance parameter. Sweeting (2005) improved upon this result, making it considerably simpler by implementing a local solution. Both papers also applied the PMPs obtained numerically to real data with fairly good results. Neither paper considered conditional inference.

As a third related study, in her recent PhD thesis at Rutgers University, Zhang (2008) argued strongly in favour of using the DiCiccio-Martin approximation. Zhang was concerned with the numerical solution of the matching condition, however, and did not consider general numerical validation of the accuracy of matching priors.

Therefore, what we now present marks the first simulation study of the conditional frequentist properties of Bayesian quantiles obtained from PMPs.

We also compare our conditional coverage results to the unconditional ones. We believe this is the first time in the literature that conditional and unconditional inference comparisons have been made via a probability matching prior route.

5.1 Our Algorithm

We investigate the conditional coverage properties of Bayesian quantiles via the following algorithm.

1. Given a particular model (and density, where required), specify the ”true” parameter values for model. For the location-scale model, we set these as $\mu = 0$, $\sigma = 1$.

   Generate a sample from the model of size n, e.g. $n = 10$.

2. Calculate the MLE for the model parameters, $\hat{\theta}$. This, together with the original sample will define the ancillary statistic. For the location-scale model, we have $A_i = (Y_i - \hat{\mu})/\hat{\sigma}$, $i = 1, \ldots, n$.

   We repeat steps 1 and 2 many times, comparing the norms of the ancillary statistics to find a ”good” value for the ancillary statistic.[1]

3. We sample from Barndorff-Nielsen's p^* formula for the conditional density of the MLE, which is exact for transformation models and accurate to order $O(n^{-3/2})$ in general.

   We use a random walk Metropolis-Hastings to sample from this p^* density, and this will give us a sequence of $\hat{\theta}^*$, which in the location-scale model is a vector of length two. Following a burn-in period, we sample only every 20th point of the Markov chain. We do this until we have 2000 samples of $\hat{\theta}^*$.

[1] That is, we try to avoid ”extreme” conditioning by not choosing an extreme value of the ancillary statistic. ”Extreme” conditioning is something we will consider in future work.

4. Now, for each of these samples $\hat{\theta}^*$, we can construct a sample Y. From the location-scale model, for example, this would be $Y = A\hat{\sigma}^* + \hat{\mu}^*$. This gives us 2000 data sets Y^* of size n.

5. From each of these samples Y^*, we construct the posterior density after designating a particular prior π. Now we have 2000 posterior densities for our model parameters, each conditional on observing n data points.

6. We now use a random walk Metropolis-Hastings to sample from each posterior. After a burn-in period, we sample every 20th point of the Markov chain and stop after 5,000 samples have been taken.

7. Using 5000 samples from each of the 2000 posteriors, we calculate the $1-\alpha$ quantile and ask the question: *Is this* $1-\alpha$ *posterior quantile for the parameter of interest bigger than the true parameter value?* More concretely, let $\pi(\theta|Y)$ be the posterior density of θ and $F(\theta|Y)$ denote the posterior cumulative distribution function. We calculate $F^{-1}(1-\alpha)$, i.e. the $1-\alpha$ posterior quantile for each posterior and simply compare this to the true parameter value, with output "yes" if $F^{-1}(1-\alpha) \geq \sigma_0$ and "no" otherwise. This gives us a sequence of "yes" and "no" for each of the 2000 posteriors.

8. The conditional coverage probability is simply the frequency that the Bayesian quantile covers the true parameter value, i.e. the number of "yes" outcomes divided by 2000.

All Metropolis-Hastings sampling schemes were tuned to give an acceptance rate near the optimal 23% as suggested by Roberts et al. (1997) and we have taken only every 20th point in the Markov chain to avoid dependence in the simulated sample.

5.2 Examples

First we examine location-scale models. While Fisher (1934) and Cox (1958) argued strongly in favour of conditional inference for location-scale models from a sufficiency standpoint, Fraser & McDunnough (1980) have found that other theoretical motivations, such as mean length arguments pertaining to confidence intervals or power in inference, are in general not suitable criteria for determining whether conditional or unconditional inference should be performed. Lawless (1972, 1982) made an admirable effort to assess the properties of conditional and unconditional inference in location-scale models, but was hindered by the computing technology of his era.[2] It is then worthwhile to revisit the comparison of conditional and unconditional inference, though we will do so in the context of coverage of posterior quantiles resulting from probability matching priors. To the best of our knowledge, this is the first appearance of such comparisons via this route in the literature.[3]

We do not specifically consider the exponential regression model. There does exist a formula for the exact conditional density of the MLE in the exponential regression model as derived by Hillier & O'Brien (1999) and Hillier & Armstrong (1999). Once the dimension of the covariates increases beyond one or two, it becomes analytically intractable. It is a location-scale model on log-scale with scale parameter equal to one, however, the presence of covariates means that the location model results cannot be strictly interpreted as carrying over to the exponential regression model.

We now present numerical results for 3 specific location-scale models: the normal, Cauchy and Gumbel. The formula for the exact conditional density of the MLE can be found in the proof at the end of Chapter 4. One final caveat before proceeding is to note that the priors $\pi \propto \mu/\sigma$ and $\pi \propto \mu/\sigma^2$ are not invariant under linear transformation and thus the numerical results obtained here are not "universal".

Example 5.2.1 *Normal Location-Scale*

As noted in Chapter 4, the normal location-scale model is a case where the MLE is a sufficient statistic and thus there is no need to condition on the ancillary. In theory,

[2]Simple examples of conditional inference in location-scale models can also be found in Table 4 of Fraser (1976) and Example 4.3 of Morgenthaler & Nicolaou (1997).

[3]Sweeting (1984) actually foreshadowed the PMP literature in noting that approximations to posterior distributions in linear location-scale regression models, using improper Bayesian priors, have a conditional frequency interpretation. He also provided some numerical results for the Weibull and t distributions, though his focus was on approximation of the posterior distribution, rather than calculating the frequentist properties of posterior quantiles.

conditional and unconditional inference should be exactly equal in the normal location-scale model.

Table 5.1: *Conditional Coverage, Normal Location-Scale*

	$\pi = 1/\sigma$		$\pi = \mu/\sigma$		$\pi = 1/\sigma^2$		$\pi = \mu/\sigma^2$	
n	5%	95%	5%	95%	5%	95%	5%	95%
5	5.8	94.6	19.6	98.1	2.5	86.6	11.2	94.9
10	5.5	94.9	11.1	97.5	3.4	91.3	7.2	94.8
15	5.3	94.6	10.1	96.3	3.7	91.7	7.3	95.0

Interest Parameter: σ; n=5: a=(-0.072035, 0.380166, -1.89159, 0.69578, 1.259878) ; n=10: a=(-1.528164, -0.553271, 1.348682, -1.14644, -0.091283, 0.15096, 0.432003, 0.849907, -1.02141, -1.497798); n=15: a=(-1.22088, -2.335605, 0.2490885, 0.5621225, 0.8719466, 0.11653339, 0.2091285, -1.015707, 1.285798, -0.493713, -0.9813167, 0.1033543, 1.514438, 0.6655885, 0.4693618)

Table 5.2: *Unconditional Coverage, Normal Location-Scale*

	$\pi = 1/\sigma$		$\pi = \mu/\sigma$		$\pi = 1/\sigma^2$		$\pi = \mu/\sigma^2$	
n	5%	95%	5%	95%	5%	95%	5%	95%
5	4.4	95.2	2.7	96.6	6.9	92.5	5.5	96.9
10	4.3	95.4	2.7	97.0	5.2	92.7	6.0	97.2
15	4.4	94.7	2.8	96.5	6.4	91.4	6.0	97.2

The results for the normal case confirm that our algorithm is working. Keep in mind that our simulation sizes were relatively small (2000 simulations). Increasing the simulation size did show a slight improvement, but at a large cost in terms of time required.

One interesting thing to note here is that the theory predicts that

$$\pi(\theta) \propto \frac{g(\mu)}{\sigma}$$

is 3rd order probability matching (see for instance, Datta & Mukerjee (2004)), where $g(\cdot)$ is an arbitrary function of the nuisance parameter. We consider the simplest such arbitrary function, $g(\mu) = \mu$ and investigate the frequentist coverage of posterior limits based on such joint priors. These priors do not, in fact, perform as well as priors which do not depend on the nuisance parameter.

Example 5.2.2 *Cauchy Location-Scale*

The Cauchy location-scale model is a unique example. As noted by Datta & Mukerjee (2004), the Cauchy model enjoys strong parameter orthogonality and thus the 3rd order matching prior may actually depend on an arbitrary function of the nuisance parameter, as in the normal case above, without affecting the matching result. We can investigate this numerically. The other quirky aspects of the Cauchy location-scale model are that there is no analytical solution for the MLE and the exact ancillary is actually not unique, as noted by McCullagh (1992). This point is reiterated in Kass & Wasserman (1996) and Sundberg (2003) where they note that we can get different right-invariant Haar priors depending on how we label the sample space.

Howlader & Weiss (1988) compared Bayesian estimates with MLEs for Cauchy location-scale, both with and without assuming orthogonality. They use the right-invariant prior (our usual prior, which McCullagh calls the Pitman prior) instead of Jeffreys prior but find that Bayesian procedures often lead to extreme values for the scale parameter, whereas the usual MLE route underestimates the scale parameter.

Table 5.3: *Conditional Coverage, Cauchy Location-Scale*

	$\pi = 1/\sigma$		$\pi = \mu/\sigma$		$\pi = 1/\sigma^2$		$\pi = \mu/\sigma^2$	
n	5%	95%	5%	95%	5%	95%	5%	95%
5	7.0	95.9	10.7	97.8	2.2	87.5	2.4	92.4
10	4.6	95.0	7.6	97.7	0.5	85.6	2.2	90.1
15	5.7	95.2	5.7	96.6	2.0	90.8	2.7	92.9

Interest Parameter: σ; n=5: a=(0.3446336, 1.948575, 1.144016, -1.144508, -0.6239359); n=10: a=(-1.424294, 0.5148966, 16.20368, 0.2690624, -1.868927, 0.1940811, -17.95261, 1.439158, 0.0623819, -1.146699); n=15: a=(-3.673128, 1.532387, 0.983479, -2.457987, 0.98798, 2.5723487, -4.3523987, 1.435232, -1.2342078, -1.6784234, 5.2350786, 3.14097, -0.177453)

These results are based on 10,000 simulations.

Table 5.4: *Unconditional Coverage, Cauchy Location-Scale*

	$\pi = 1/\sigma$		$\pi = \mu/\sigma$		$\pi = 1/\sigma^2$		$\pi = \mu/\sigma^2$	
n	5%	95%	5%	95%	5%	95%	5%	95%
5	4.8	95.1	9.8	96.3	0.7	81.6	1.5	85.4
10	5.2	94.5	7.8	96.3	1.5	87.4	2.7	89.9
15	5.1	95.0	7.4	96.1	2.2	89.0	2.9	91.4

These results are broadly consistent with what the theory predicts. Both the conditional and unconditional coverage probabilities are nearly exact when using the exact matching prior, and other priors (such as Jeffreys) perform slightly worse.

Example 5.2.3 *Extreme Value (Gumbel) Location-Scale*

Table 5.5: *Conditional Coverage, Gumbel Location-Scale*

	$\pi = 1/\sigma$		$\pi = \mu/\sigma$		$\pi = 1/\sigma^2$		$\pi = \mu/\sigma^2$	
n	5%	95%	5%	95%	5%	95%	5%	95%
5	8.9	94.8	30.6	98.1	6.4	88.5	22.2	92.2
10	6.3	95.1	23.4	98.3	3.3	91.0	16.7	96.8
15	4.8	94.5	15.0	97.5	1.8	90.8	11.6	96.3

Interest Parameter: σ; n=5: a=(-2.756895, -6.946514, 1.698959, 0.6130725, 1.305158) ; n=10: a=(1.528164, -0.5532712, 1.348682, -1.14644, -0.09128283, 0.150959, 0.4320034, 0.8499072, -1.02141, -1.497798) ; n=15: a=(1.280454, -2.36309, -1.609262, -2.082839, -3.814035, -2.646271, 3.076648, -2.359961, -1.802601, -2.894048, 4.485054, -0.5286729, -1.801257, 2.199035, -4.663153)

Table 5.6: *Unconditional Coverage, Gumbel Location-Scale*

	$\pi = 1/\sigma$		$\pi = \mu/\sigma$		$\pi = 1/\sigma^2$		$\pi = \mu/\sigma^2$	
n	5%	95%	5%	95%	5%	95%	5%	95%
5	1.6	93.2	3.0	95.7	0.8	81.6	0.6	87.0
10	3.8	91.4	3.3	92.3	2.8	86.8	2.3	86.7
15	5.1	88.9	4.2	89.9	3.2	86.7	2.9	85.3

These results are somewhat puzzling. The conditional coverage seems to become more concentrated as the sample size increases. We have tried many permutations of the algorithms and ancillary statistics of various norms (some mild, some extreme) and the results are fairly stable. However, the unconditional coverage is slightly better. The lower tail coverage behaves as it should. In both the conditional and unconditional sense, the exact matching prior yields the best results.

Example 5.2.4 *Exponential Hyperbola*

An approximation to the conditional density of the MLE is given by

$$p(\hat{\theta}|a;\theta) = \frac{1}{2\hat{K}_0(2e^{a/2})}\frac{1}{\hat{\theta}}\exp\Big\{-na\big[\frac{\hat{\theta}^2+\theta^2}{\hat{\theta}\theta}\big]\Big\} \tag{5.1}$$

where $\hat{\theta} > 0$, and $\hat{K}_0$ is a Bessel K function defined by

$$\hat{K}_0(z) = e^{-z}(\pi/2z)^{1/2}$$

As noted by Butler (2007), this approximation is exact. We had considerable difficulty with this example. The ancillary statistic is scalar and the results seem to be quite sensitive to the value of the ancillary statistic. The results we present are representative; better results can be achieved by severe manipulation of the algorithms, but the validity of such results is doubtful. As we can see, the small sample performance is not very good, and the choice of prior is not terribly important. We consider the reference prior (our 3rd order conditional PMP from Example 4.1.4) as well as several variations.[4]

Table 5.7: *Conditional Coverage, Exponential Hyperbola*

	$\pi = 1/\theta$		$\pi = 1/\theta^2$		$\pi = (\theta)^{-1/2}$	
n	5%	95%	5%	95%	5%	95%
5	5.1	95.1	2.7	91.3	-	-
10	5.5	95.3	3.4	92.5	-	-
15	5.2	95.5	3.8	93.2	6.1	96.4

$n = 5$: *a=1.13734; n=10: a=1.09326;* $n = 15$: *a=1.06872*

For low values of n, the third prior considered seems to yield an improper posterior, which is why the results for $n = 5, 10$ are missing. These figures are based on simulations of size 5000.

Table 5.8: *Unconditional Coverage, Exponential Hyperbola*

	$\pi = 1/\theta$		$\pi = 1/\theta^2$		$\pi = (\theta)^{-1/2}$	
n	5%	95%	5%	95%	5%	95%
5	5.3	93.8	2.8	89.3	6.9	95.6
10	5.1	93.7	3.7	90.0	5.8	94.9
15	4.9	95.4	4.5	90.2	5.6	94.5

The conditional coverages are moving in the right direction, but clearly all of the priors considered yield better unconditional than conditional matching for the sample sizes considered. Also, in the unconditional setting, it is more readily apparent that the 3rd order PMP identified via the saddlepoint route does perform slightly better than the other priors.

[4]We also tried priors such as $\pi \propto \theta$, $\pi \propto \sqrt{\theta}$ but encountered numerical problems.

Chapter 6

Ongoing and Future Work

We conclude by discussing various topics under research at present and in future.

6.1 General Asymptotic Results

Posterior Edgeworth Expansions

The first result needed in the literature, as discussed in Chapter 2, is the establishment of the validity of Edgeworth expansions of posterior densities. Moreover, it must be shown how Edgeworth expansion works in the posterior setup, since the posterior is usually expressed as a ratio and existing posterior expansions analyze the asymptotics of the numerator and denominator separately and then simply take the ratio. It is not clear if this is the way to proceed for Edgeworth expansions or if there is a more direct route.

A Marginal Version of R^*

The second general asymptotic result which is missing from the literature is a marginal frequentist version of R^*. Jensen (1995) gives the only attempt at this problem and derives a saddlepoint approximation to R which has good large-deviation properties and is valid for a curved $(2, 1)$ exponential family. It is very difficult to extend this result to models with higher-dimensional parameters. The author is currently working with Jens Jensen on this problem.

Higher Order Analysis of Expected and Observed Likelihood Quantities

Barndorff-Nielsen & Blaesild (1993) derived relationships between expected and observed likelihood geometries based on likelihood yokes. In general these are only accurate to $O_p(n^{-1/2})$. However, it would be helpful to find examples where equivalence can be established to a higher order of accuracy. This is necessary for the comparison of the saddlepoint matching conditions with the other matching methods. It is also likely related to the "sufficiency" motivation for conditional inference; for example, in exponential families with the canonical parameterization, expected and observed information are the same and conditioning in such settings is done not because the MLE is not a sufficient statistic, but rather to make inference more relevant to the parameter of interest. Higher-order comparison of expected and observed likelihood quantities is planned for future research.

Higher Order Approximations to Sample Space Derivatives

We noted that one of the major obstacles to comparing the matching methods is that the saddlepoint approach requires the calculation of sample-space derivatives, which would require explicit specification of an ancillary statistic. Alternatively, we could use approximations to these quantities. Skovgaard (1996) proposed an approximation to the sample-space derivatives which is accurate to order $O(n^{-1})$ and proportional to the statistical curvature of the model. We note that we are considering situations where $\theta = \theta_0 + \delta/\sqrt{n}$, i.e. small deviation regions which include the usual maximum likelihood estimates. Severini (1999) approaches the problem in a similar way to Skovgaard, and extends the analysis to incorporate empirical covariances.

Let the relationship between the constrained and global MLE be such that $\hat{\lambda}_\psi = \hat{\lambda} + O(n^{-1/2})$, which we will write more conveniently as $\hat{\theta}_\psi = \hat{\theta} + O(n^{-1/2})$. We also use the fact that $\hat{\jmath} = -I(\hat{\theta}) + O_p(n^{-1/2})$. The basic idea behind these approximations is that the sample space derivatives may be approximated by covariances of the log-likelihood and score functions.

The Skovgaard-Severini approximations to the relevant sample space derivatives for the computation of the conditional frequentist adjustment term of the R^* are given by:

$$\Sigma(\psi) = \ell_{\theta;\hat{\theta}}(\psi, \hat{\lambda}_\psi) = \frac{\partial^2 \ell(\psi, \hat{\lambda}_\psi)}{\partial\theta\partial\hat{\theta}} = cov\Big(\ell_\theta(\hat{\theta}), \ell_\theta(\psi, \hat{\lambda}_\psi)\Big)\hat{I}^{-1}\hat{\jmath},$$

$$\Gamma(\psi) = \ell_{\hat{\theta}}(\psi, \hat{\lambda}_\psi) - \ell_{\hat{\theta}}(\hat{\theta}) = cov\Big(\ell_\theta(\hat{\theta}), [\ell(\psi, \hat{\lambda}_\psi) - \ell(\hat{\theta})]\Big)\hat{I}^{-1}\hat{\jmath},$$

where $\hat{I}$ and $\hat{\jmath}$ are the expected and observed information evaluated at $\hat{\theta}$.

These are accurate to order $O(n^{-1})$, though asymptotic expansions (for instance Taylor expansion to a higher order) of the likelihood quantities is possible. For the time being, this is useful only for general interest since we are hindered by the approximations in the previous point (observed vs. expected quantities).

6.2 Objective Bayes and Conditional Inference

Necessary and Sufficient Conditions?

As it currently stands, we have necessary and sufficient conditions for 3rd order matching via the shrinkage argument and, since it provides identical conditions, we also have necessary and sufficient conditions for 3rd order matching via mean- and variance-adjusted signed roots. The matching conditions via the saddlepoint approach are thus far only demonstrated to be sufficient to ensure 3rd order matching. A proof that these matching conditions are also necessary has thus far eluded us, but it is currently being studied.

More General Model Settings

We are already working on extending the simple framework to cover models with vector nuisance parameters as well as models for which there is only an approximate ancillary statistic. Much of this has been considered in the unconditional PMP literature already and an excellent review of these findings is in Datta & Sweeting (2005).

Relationships Between Conditional and Unconditional PMPs

The analysis of the relationship between conditional and unconditional PMPs is far from complete. Thus far we have established some conditions in Chapter 3 under which unconditional PMPs are also conditional PMPs to the same order, yet these results are fairly specific. One of the goals of future research is to establish necessary and sufficient conditions for which unconditional PMPs are also conditional PMPs, with the hope that these turn out to be conditions on the model being considered.

The Importance of Test Statistic

The suggestive results of Chapter 3 that, for a particular choice of test statistic, conditional and unconditional inference in the exponential hyperbola model and Cauchy location model agree to order $O(n^{-1})$, highlight a more important point about inference in general and probability matching in particular: the choice of test statistic does seem to matter. In addition to pursuing a better understanding of conditional and unconditional inference and addressing the fundamental question, ”does conditioning make any difference?”, it is also important from a practical standpoint. Since few applied statisticians are using conditional inference in situations where theory tells us to do so, then we must be realistic. If we cannot convince practitioners to do inference conditionally, then at least we could provide a guide for choosing test statistics for parametric inference which are more robust to a failure to do conditional inference.

6.3 Numerical Work

Cases To Be Studied

While we have examined the most common exact ancillary statistic models, there are many more to be considered, as evidenced by the list in Buehler (1982).

Numerical Solution of Edgeworth Approach Matching Conditions

As noted in Chapter 2, the Edgeworth approach is analytically extremely challenging to implement due to the necessity of calculating standardized cumulants of the posterior density. In the future we will endeavour to solve these matching conditions numerically to identify conditional PMPs numerically. Given the well-documented problems with Edgeworth expansions for tail probabilities, we are not hopeful that this route will compare favourably against the saddlepoint approach, but nonetheless it is under investigation in the interest of completeness.

The Importance of Test Statistic

Further to the point above about the theoretical importance of the choice of test statistic, we must also investigate this numerically. In particular, since we are basing so much of the theory on the signed-root likelihood ratio statistic, we must establish that this is a ”good” statistic to use in practice. We would also like to determine which test statistics are such that some measure of the ”distance” between the marginal and conditional distributions of the test statistic is maximised. Such test statistics would be illuminating for studying the differences between conditional and unconditional PMPs in particular, and inference in general.

Sample R Code

Here is a sample R code used to produce Table 5.3, Conditional Coverage in the Cauchy Location-Scale Model.

```
"Cauchy Conditional Location-Scale

require(MASS)

Defining the functions

#this function computes conditional density of the MLE given
#an ancillary to use it in the Metropolis-Hastings algorithm

conditionalDensMLE <- function(eLocation, eScale,
ancillary, tLocation, tScale){
if (eScale <= 0){
target <- 0
} else {
n <- length(ancillary) # size of the ancillary vector
temp <- prod(dcauchy(eLocation+eScale*ancillary,
location=tLocation, scale=tScale, log = FALSE))
target <- temp*eScale^(n-2)
}
return(target)
}

# this function computes the posterior density
# to use it in the Metropolis-Hastings algorithm
```

```
posteriorDensMLE <- function(xx, yy, ancillary, rmu, rsig){
if (yy <= 0){
target <- 0
} else {
n <- length(ancillary) # size of the ancillary vector
temp <- prod(dcauchy(rmu+rsig*ancillary))
target <- temp*1/yy # the prior part
}
return(target)
}

sampleSize <- 10 # sample size

### Part 1 ###  Get initial MLE estimates
# generate random samples from the Cauchy distr.
tLocation <- 0 # "true" location parameter
tScale <- 1 # "true" scale parameter
y <- rcauchy(sampleSize, location=tLocation, scale=tScale)
# maximum-likelihood fit of Cauchy distr.
fit <- fitdistr(y, dcauchy, list(location = tLocation,
scale = tScale))
eLocation <- fit\$estimate[1] # estimated location parameter
eScale <- fit\$estimate[2] # estimated scale parameter
# compute ancillary statistic (vector of size 'sampleSize')
ancillary <- (y-eLocation)/eScale

cat("ancillary values", ancillary, "\n")


Part 2 Sampling from p*
### output is rmu and rsig, vectors of size nSamples


# nSamples <- 2000 # number of samples


ntest <- 1 # number of tests
```

```
nsim <- 2000 # number of simulations
# set parameters of distributions

locationSigma <- 1
scaleSigma <- 1
alim <- 0
blim <- 1 # upper limit of the uniform distr.

rmu <- rsig <- 0
for (i in 1:ntest){
nacc <- 0
curr <- c(0.08, 1.30)
icount <- 0
jcount <- 0

# now do the simulations
y <- rcauchy(sampleSize, tLocation, tScale)

for (kkk in 1:nsim){
# generate random sample from the Normal distr.
val1 <- rnorm(1, curr[1], locationSigma)
val2 <- rnorm(1, curr[2], scaleSigma)
one <- conditionalDensMLE(val1, val2, ancillary,
tLocation, tScale)
two <- conditionalDensMLE(curr[1], curr[2], ancillary,
tLocation, tScale)
ratio <- one/two
factor <- min(1, ratio)
# generate random sample from the Uniform distr.
unif <- runif(1, alim, blim)
if (unif <= factor){
nacc <- nacc+1
curr[1] <- val1
curr[2] <- val2
}
if (kkk\%\%50 == 0){
#  cat("carr[1] and carr[2]", curr[1], curr[2], "\n")
```

```
}
rmu[kkk] <- curr[1]
rsig[kkk] <- curr[2]
}
acc <- nacc/nsim
cat("sigma & acc", locationSigma, scaleSigma, acc, "\n")
}

# define vector (1:9 is a sequence 1,2,...,9)
rlevel <- c(0.01, 0.025, 0.05, 0.1*1:9, 0.95, 0.975, 0.99)

##### Part 3 Sampling from the posterior

# replicate zero 15 times
kcount <- rep.int(0,15)
ntest2 <- 2000  # number of tests
cat("simulation size", ntest, "\n")
nsim2 <- 2000 # number of simulations
# set parameter values of distributions
locationSigma <- 2
scaleSigma <- 2
tLocation <- 0
tScale <- 1
alim <- 0
blim <- 1


for (i in 1:ntest2){
curr <- c(0.01, 0.26)
icount <- 0
jcount <- 0

# generate n random samples from the Cauchy distr.
y <- rcauchy(sampleSize, tLocation, tScale)

for (kkk in 1:nsim2){
# generate random sample from the Normal distr.
```

```
val1 <- rnorm(1, curr[1], locationSigma)
val2 <- rnorm(1, curr[2], scaleSigma)
one <- posteriorDensMLE(val1, val2, ancillary, rmu[i], rsig[i])
two <- posteriorDensMLE(curr[1], curr[2], ancillary, rmu[i], rsig[i])
ratio <- one/two
if (two == 0) ratio <- 1
factor <- min(1, ratio)
# generate random sample from the Uniform distr.
unif <- runif(1, alim, blim)
if (unif <= factor){
nacc <- nacc+1
curr[1] <- val1
curr[2] <- val2
}
if (kkk\%\%20 == 0){
jcount <- jcount+1
if (curr[2] < tScale) icount <- icount+1
}
}
tail <- icount/jcount
# cat("i:", i, ", tail:", tail, "\n")
for (jj in 1:15){
if(tail < rlevel[jj]) kcount[jj] <- kcount[jj]+1
}
}
for (jj in 1:15){
cat("rlevel & kcount/ntest:", rlevel[jj], kcount[jj]/ntest2, "\n")
}
acc <- nacc/(ntest2*nsim2)
cat("locationSigma, scaleSigma", locationSigma, scaleSigma, "\n")
cat("acceptance rate", acc, "\n")"
```

References

Andersen, E. B. (1970), ‘Asymptotic Properties of Conditional Maximum-Likelihood Estimators.’, *Journal of the Royal Statistical Society, Series B* **32**(2), 283–301.

Barndorff-Nielsen, O. (1985), ‘Confidence limits from $c|\jmath|^{1/2}\bar{L}$ in the single-parameter case.’, *Scandinavian Journal of Statistics* **12**, 83–87.

Barndorff-Nielsen, O. (1986), ‘Inference on full or partial parameters based on the standardized signed log likelihood ratio.’, *Biometrika* **73**(2), 307–322.

Barndorff-Nielsen, O. & Blaesild, P. (1993), ‘Likelihood Yokes and Stable Combinations of Cumulants.’, *Scandinavian Journal of Statistics* **20**, 83–87.

Barndorff-Nielsen, O. & Cox, D. R. (1979), ‘Edgeworth and saddlepoint approximations with statistical applications (with discussion).’, *Journal of the Royal Statistical Society, Series B* **41**, 279–312.

Barndorff-Nielsen, O. & Cox, D. R. (1994), *Inference and Asymptotics.*, Vol. 52, Chapman & Hall, Monographs on Statistics and Applied Probability.

Berger, J. & Bernardo, J. M. (1992), On the development of reference priors (with discussion)., *in* ‘Bayesian Statistics 4: eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith’, Oxford University Press, pp. 35–60.

Bernardo, J. M. (1979), ‘Reference posterior distributions for Bayesian inference (with discussion).’, *Journal of the Royal Statistical Society, Series B* **41**, 113–147.

Bhattacharya, R. N. & Ghosh, J. K. (1978), ‘On the Validity of the Formal Edgeworth Expansion.’, *Annals of Statistics* **6**(2), 434–451.

Bickel, P. J. & Ghosh, J. K. (1990), 'A decomposition for the likelihood ratio statistic and the Bartlett correction - a Bayesian argument.', *Annals of Statistics* **18**, 1070–1090.

Buehler, R. J. (1982), 'Some Ancillary Statistics and Their Properties.', *Journal of the American Statistical Association* **77**(379), 581–589.

Butler, R. W. (2007), *Saddlepoint Approximations With Applications.*, Cambridge University Press.

Casella, G., DiCiccio, T. J. & Wells, M. T. (1995), '[Inference based on estimation functions in the presence of nuisance parameters]: Comment: Alternative aspects of conditional inference.', *Statistical Science* **10**, 179–185.

Cox, D. R. (1958), 'Some problems connected with statistical inference.', *Annals of Mathematical Statistics* **29**, 357–372.

Cox, D. R. (1980), 'Local ancillarity.', *Biometrika* **67**, 279–286.

Cox, D. R. & Hinkley, D. V. (1974), *Theoretical Statistics.*, Chapman & Hall.

Cox, D. R. & Reid, N. (1987), 'Parameter orthogonality and approximate conditional inference (with discussion).', *Journal of the Royal Statistical Society, Series B* **53**, 79–109.

Cox, D. R. & Reid, N. (1989), 'On the stability of maximum-likelihood estimators of orthogonal parameters.', *The Canadian Journal of Statistics* **17**(2), 229–233.

Datta, G. S. & Ghosh, J. K. (1995), 'On priors providing frequentist validity for Bayesian inference.', *Biometrika* **82**, 37–45.

Datta, G. S. & Mukerjee, R. (2004), *Probability Matching Priors: Higher Order Asymptotics*, Springer Lecture Notes in Statistics, New York.

Datta, G. S., Mukerjee, R., Ghosh, M. & Sweeting, T. J. (2000), 'Bayesian prediction with approximate frequentist validity.', *Annals of Statistics* **28**(5), 1414–1426.

Datta, G. S. & Sweeting, T. J. (2005), 'Probability matching priors.', *Handbook of Statistics, 25: Bayesian Thinking: Modeling and Computation* pp. 91–114.

Dawid, A. P. (1991), 'Fisherian inference in likelihood and prequential frames of reference (with discussion).', *Journal of the Royal Statistical Society, Series B* **53**, 79–109.

DiCiccio, T. J. & Martin, M. A. (1991), 'Approximations of marginal tail probabilities for a class of smooth functions with applications to Bayesian and conditional inference.', *Biometrika* **78**, 891–902.

DiCiccio, T. J. & Martin, M. A. (1993), 'Simple modifications for signed roots of likelihood ratio statistics.', *Journal of the Royal Statistical Society, Series B* **55**, 305–316.

DiCiccio, T. J. & Stern, S. E. (1993), 'On Bartlett adjustments for approximate Bayesian inference.', *Biometrika* **80**, 731–740.

DiCiccio, T. J. & Stern, S. E. (1994*a*), 'Constructing Approximately Standard Normal Pivots from Signed Roots of Adjusted Likelihood Ratio Statistics.', *Scandinavian Journal of Statistics* **21**(4), 447–460.

DiCiccio, T. J. & Stern, S. E. (1994*b*), 'Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihoods.', *Journal of the Royal Statistical Society, Series B* **56**, 397–408.

DiCiccio, T. J. & Young, G. A. (2010), 'Objective Bayes and conditional inference in exponential families.', *Biometrika* **97**(2), 497–504.

Efron, B. (1975), 'Defining the curvature of a statistical problem (with applications to second order sufficiency)(with discussion)', *Annals of Statistics* **3**, 1189–1242.

Efron, B. & Hinkley, D. V. (1978), 'Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information (with discussion)', *Biometrika* **65**, 457–487.

Fernandez, C. & Steel, M. (1999), 'Reference priors for the general location-scale model.', *Statistics & Probability Letters* **43**(4), 377–384.

Fisher, R. A. (1934), 'Two new properties of mathematical likelihood.', *Proceedings of the Royal Society of London, Series A* **134**, 285–307.

Fraser, D. A. S. (1976), ‘Necessary Analysis and Adaptive Inference.’, *Journal of the American Statistical Association* **71**(353), 99–113.

Fraser, D. A. S. & McDunnough, P. (1980), ‘Some remarks on conditional and unconditional inference for location-scale models.’, *Statistiche Hefte* **21**, 224–231.

Fraser, D. A. S. & Reid, N. (1995), ‘Ancillaries and third-order significance.’, *Utilitas Math.* **7**, 33–53.

Fraser, D. A. S. & Reid, N. (2002), ‘Strong matching of frequentist and Bayesian parametric inference.’, *Journal of Statistical Planning and Inference* **103**, 263–285.

Fraser, D. A. S., Reid, N. & Wu, J. (1999), ‘A simple general formula for tail probabilities for frequentist and Bayesian inference.’, *Biometrika* **86**(2), 249–264.

Ghosal, S. (1999), ‘Probability matching priors for non-regular cases.’, *Biometrika* **86**, 956–964.

Ghosal, S. & Samanta, T. (1997), ‘Asymptotic Expansions of Posterior Distributions in Nonregular Cases.’, *Annals of the Institute of Statistical Mathematics* **49**(1), 181–197.

Ghosh, J. K., B.Sinha & Joshi, S. (1985), ‘Expansions for posterior distributions and integrated Bayes risk.’, *Proceedings of the Third Purdue Symposium Statist. Dec. Th. Rel. Topics* **1**, 403–456.

Ghosh, J. K. & Mukerjee, R. (1991), ‘Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case.’, *Journal of Multivariate Analysis* **38**, 385–393.

Ghosh, J. K. & Mukerjee, R. (1992), ‘Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio tests.’, *Journal of the Royal Statistical Society, Series B* **54**, 867–875.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion.*, Springer Series in Statistics.

Hillier, G. H. & Armstrong, M. (1999), ‘The Density of the Maximum Likelihood Estimator.’, *Econometrica* **67**(6), 1459–1470.

Hillier, G. H. & O'Brien, R. J. (1999), 'Exact Properties of the Maximum Likelihood Estimator in Exponential Regression Models: A Differential Geometric Approach.', *Applications of Differential Geometry to Econometrics, ed. by M. Salmon and P. Marriot* pp. 85–118.

Hinkley, D. V. (1978), 'Likelihood inference about location and scale parameters.', *Biometrika* **65**, 253–261.

Hora, R. B. & Buehler, R. J. (1966), 'Fiducial theory and invariant estimation.', *Annals of Mathematical Statistics* **37**, 643–656.

Hora, R. B. & Buehler, R. J. (1967), 'Fiducial theory and invariant prediction.', *Annals of Mathematical Statistics* **38**, 795–801.

Howlader, H. A. & Weiss, G. (1988), 'On Bayesian Estimation of the Cauchy Parameters.', *Sankhya* **50**(3), 350–361.

Jeffreys, H. (1946), 'An Invariant Form for the Prior Probability in Estimation Problems.', *Proceedings of the Royal Society of London, Series A* **186**(1007), 453–461.

Jensen, J. L. (1986), 'Similar tests and the standardized log likelihood ratio statistic.', *Biometrika* **73**, 567–572.

Jensen, J. L. (1995), *Saddlepoint Approximations.*, Oxford University Press.

Johnson, R. A. (1967), 'An asymptotic expansion for posterior distributions.', *Annals of Mathematical Statistics* **38**, 1899–1907.

Johnson, R. A. (1970), 'Asymptotic expansions associated with posterior distributions.', *Annals of Mathematical Statistics* **41**, 851–864.

Kass, R. E. & Wasserman, L. (1996), 'The selection of prior distributions by formal rules.', *Journal of the American Statistical Association* **91**, 1343–1370.

Kumon, M. (2009), 'On the conditions for the existence of ancillary statistics in a curved exponential family.', *Statistical Methodology* **6**, 320–335.

Lawless, J. (1976), 'Confidence Interval Estimation in the Inverse Power Law Model.', *Journal of the Royal Statistical Society Series C (Applied Statistics)* **25**(2), 128–138.

Lawless, J. F. (1972), 'Conditional Confidence Interval Procedures for the Location and Scale Parameters of the Cauchy and Logistic Distributions.', *Biometrika* **59**(2), 377–386.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data.*, John Wiley & Sons, New York.

Levine, R. A. & Casella, G. (2003), 'Implementing matching priors for frequentist inference.', *Biometrika* **90**, 127–137.

Li, J. (1998), 'Some Aspects of Bayesian and Frequentist Asymptotics.', *PhD Thesis, University of Toronto* pp. 1–129.

Lindley, D. V. (1958), 'Fiducial distributions and Bayes' theorem.', *Journal of the Royal Statistical Society, Series B* **20**, 102–107.

Lindley, D. V. (1961), 'The Use of Prior Probability Distributions in Statistical Inference and Decisions.', *Proceedings of the Fourth Berkeley Symposium* **1**, 453–468.

Liseo, B. (1993), 'Elimination of nuisance parameters with reference priors.', *Biometrika* **80**, 295–304.

McCullagh, P. (1984), 'Local sufficiency.', *Biometrika* **71**, 233–244.

McCullagh, P. (1992), 'Conditional inference and Cauchy models.', *Biometrika* **79**(2), 247–259.

Morgenthaler, S. & Nicolaou, A. (1997), 'Composite transformation models: A fiducial perspective.', *Journal of Statistical Planning and Inference* **57**, 49–57.

Mukerjee, R. & Dey, D. K. (1993), 'Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher-order asymptotics.', *Biometrika* **80**, 499–505.

Mukerjee, R. & Ghosh, M. (1997), 'Second-order probability matching priors.', *Biometrika* **84**, 970–975.

Nicolaou, A. (1993), 'Bayesian intervals with good frequency behaviour in the presence of nuisance parameters.', *Journal of the Royal Statistical Society, Series B* **55**, 377–390.

Nicolaou, A. (1994), 'Conditional Properties of Bayesian Interval Estimates.', *Annals of the Institute of Statistical Mathematics* **16**(1), 21–28.

Peers, H. W. (1965), 'On confidence sets and Bayesian probability points in the case of several parameters.', *Journal of the Royal Statistical Society, Series B* **27**, 9–16.

Pierce, D. A. & Bellio, R. (2006), 'Effects of the reference set on frequentist inferences.', *Biometrika* **93**(2), 425–438.

Pierce, D. A. & Peters, D. (1992), 'Practical use of higher-order asymptotics for multiparameter exponential families (with discussion).', *Journal of the Royal Statistical Society, Series B* **54**, 701–738.

Reid, N. (1995), 'The Roles of Conditioning in Inference.', *Statistical Science* **10**(2), 138–199.

Reid, N. & Mukerjee, R. (2000), 'On the Bayesian Approach for Frequentist Computations.', *Brazilian Journal of Probability and Statistics* **14**, 159–166.

Roberts, G. O., Gelman, A. & Gilks, W. R. (1997), 'Weak convergence and optimal scaling of random walk Metropolis algorithms.', *Annals of Applied Probability* **7**(1), 110–120.

Rousseau, J. (2000), 'Coverage Properties of One-Sided Intervals in the Discrete Case and Application to Matching Priors.', *Annals of the Institute of Statistical Mathematics* **52**(1), 28–42.

Rubin, D. B. (1984), 'Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician.', *Annals of Statistics* **12**(4), 1151–1172.

Ryall, T. A. (1981), 'Extensions of the concept of local ancillarity.', *Biometrika* **68**(3), 677–683.

Severini, T. A. (1990), 'Conditional properties of likelihood-based significance tests.', *Biometrika* **77**(2), 343–352.

Severini, T. A. (1993), ‘Bayesian Interval Estimates which are also Confidence Intervals.’, *Journal of the Royal Statistical Society, Series B* **55**(2), 533–540.

Severini, T. A. (1998), ‘An approximation to the modified profile likelihood function.’, *Biometrika* **85**, 403–411.

Severini, T. A. (1999), ‘An empirical adjustment to the likelihood ratio statistic.’, *Biometrika* **86**, 235–247.

Severini, T. A. (2000), *Likelihood Methods in Statistics*, Vol. 22, Oxford Statistical Science Series.

Severini, T. A., Mukerjee, R. & Ghosh, M. (2002), ‘On an exact probability matching property of right-invariant priors.’, *Biometrika* **89**, 952–957.

Skovgaard, I. M. (1985), ‘A second-order investigation of asymptotic ancillarity.’, *Annals of Statistics* **13**, 534–551.

Skovgaard, I. M. (1996), ‘An explicit large-deviation approximation to one-parameter tests.’, *Bernoulli* **2**, 145–165.

Staicu, A. & Reid, N. (2008), ‘On Probability Matching Priors.’, *Canadian Journal of Statistics* **36**, 613–622.

Stein, C. M. (1985), On the coverage probability of confidence sets based on a prior distribution., *in* ‘Sequential Methods in Statistics’, Vol. 16, PWN-Polish Scientific Publishers, Warsaw, pp. 485–514.

Sundberg, R. (2003), ‘Conditional statistical inference and quantification of relevance.’, *Journal of the Royal Statistical Society, Series B* **65**, 299–315.

Sweeting, T. J. (1984), ‘Approximate Inference in Location-Scale Regression Models.’, *Journal of the American Statistical Association* **79**(388), 847–852.

Sweeting, T. J. (1991), ‘Discussion of A.P. Dawid: Fisherian inference in likelihood and prequential frames of reference.’, *Journal of the Royal Statistical Society, Series B* **53**, 106–107.

Sweeting, T. J. (1995*a*), 'A Bayesian approach to approximate conditional inference.', *Biometrika* **82**, 25–36.

Sweeting, T. J. (1995*b*), 'A framework for Bayesian and likelihood approximations in statistics.', *Biometrika* **82**, 1–23.

Sweeting, T. J. (1999), 'On the construction of Bayes-confidence regions.', *Journal of the Royal Statistical Society, Series B* **61**, 849–861.

Sweeting, T. J. (2001), 'Coverage probability bias, objective Bayes and the likelihood principle.', *Biometrika* **88**, 657–675.

Sweeting, T. J. (2005), 'On the implementation of local probability matching priors for interest parameters.', *Biometrika* **92**, 47–58.

Sweeting, T. J. (2008), On predictive probability matching priors., *in* B.Clarke & S. Ghosal, eds, 'IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh.', Vol. 3, Institute of Mathematical Statistics, pp. 46–59.

Sweeting, T. J., Datta, G. S. & Ghosh, M. (2006), 'Nonsubjective priors via predictive relative entropy loss.', *Annals of Statistics* **34**, 441–468.

Tibshirani, T. A. (1989), 'Noninformative priors for one parameter of many.', *Biometrika* **76**, 604–608.

Tierney, L. & Kadane, J. B. (1986), 'Accurate Approximations for Posterior Moments and Marginal Densities.', *Journal of the American Statistical Association* **81**, 82–86.

Ventura, L., Cabras, S. & Racugno, W. (2009), 'Prior Distributions from Pseudo-Likelihoods in the Presence of Nuisance Parameters.', *Journal of the American Statistical Association* **104**(486), 668–774.

Welch, B. L. & Peers, H. W. (1963), 'On formulae for confidence points based on integrals of weighted likelihoods.', *Journal of the Royal Statistical Society, Series B* **35**, 318–329.

Weng, R. C. (2010), 'A Bayesian Edgeworth expansion by Stein's Identity.', *Bayesian Analysis* **5**(4), 619–641.

Woodroofe, M. (1992), 'Integrable Expansions for Posterior Distributions for One-Parameter Exponential Families.', *Statistica Sinica* **2**, 91–111.

Ye, K. & Berger, J. O. (1991), 'Noninformative Priors for Inferences in Exponential Regression Models.', *Biometrika* **78**(3), 645–656.

Young, G. A. (2009), 'Routes to higher-order accuracy in parametric inference.', *Australian & New Zealand Journal of Statistics* **51**, 115–126.

Zabell, S. L. (1980), 'Rates of convergence for conditional expectations.', *Annals of Probability* **8**(5), 928–941.

Zhang, J. (2008), 'Higher Order Conditional Inference Using Parallels with Approximate Bayesian Techniques.', *PhD Thesis, Rutgers University* pp. 1–74.