SOME NEW CONSIDERATIONS FOR THE STATISTICAL ANALYSIS OF AN ASSAY

A THESIS

SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY IN STATISTICS

IN THE

UNIVERSITY OF CANTERBURY

by

Alistair James O'Malley

University of Canterbury

1999

Abstract

Statistical analysis is a vital component of an assay. In an immunoassay, the diagnosis of an illness or the determination of a treatment may be at stake. In this and other assays, it is essential that the assay be analyzed with the greatest accuracy. Standard models for assays tend to have several complicating characteristics which have led to approximate rather than exact evaluation of inferences. The main focus of this thesis is the development of methods that do not compromise the accuracy of the statistical analysis of an assay.

For the most part, a Bayesian view is taken. There are many philosophical arguments in favour of the Bayesian approach. However, the involvement of the Bayesian paradigm in this thesis is through necessity rather than philosophy. The frequentist paradigm provides no apparent means of evaluating many of the calculations involved in the analysis of an assay. On the contrary, there is a formal procedure for solving all inference and decision making problems under the Bayesian paradigm.

Heterogeneity in the variance of the response is one of the complicating characteristics of assay data. Inferences for heteroscedastic regression models are strongly dependent on the fitted variance function. The estimation of the variance function for an assay is addressed in this thesis.

Fitting the assay model and drawing inferences about the parameters is only one side of the statistical analysis of an assay. The other endeavour is the assessment of the quality of the assay as a measuring device. This is needed in order to maintain the quality of the assay over time and potentially for use as a criterion for the determination of an assay's optimal analytical and statistical (i.e. experimental) designs.

The assessment of the quality of an assay has also been compromised by frequentist approximations. The development and analysis of a Bayesian model for an

assay along with the assessment of the quality of an assay join variance function estimation as the focal points of this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

An assay is a test that involves the counting of parts. In the fields of mining and prospecting, assays have been used for many years to test the purity or concentration of mineral and ore. Assays have since been used in biology[1] and more recently in clinical chemistry, pharmokinetics and toxicology (van Houweligen (1988)).

The catalyst for this thesis is the immunoassay, an assay used in clinical chemistry, conducted in the Department of Nuclear Medicine at Christchurch Public Hospital (NMCH) in Christchurch, New Zealand. Immunoassays are exquisite analytical tests used for qualitative and quantitative analysis of substances in blood fluids (Davies (1994)). The test involves the counting (assaying) of molecules, the immunoassay is thus an assay in its own right. Incidentally, the prefix "immuno" stems from the use of antibodies as reagents in these assays. (Refer to Appendix A for a detailed description of an immunoassay.)

The analysis of an assay can justly be considered a problem in statistical calibration. In an assay, samples with a known concentration (standards) and samples with an unknown concentration (unknowns) of some compound produce response counts. A regression model is fitted to these data and inferences about the concentration of the unknowns are calculated. The role of the standards is to calibrate the mean function, the relationship between the mean response and concentration. This function is often referred to as the dose-response curve of an assay. Often replicated

---

[1] Assays used for biological tests or applications are often referred to as bioassays.

response counts are taken for each sample tested.

Assays are frequently used for analyses where a high level of accuracy is demanded. For example, in immunoassay the diagnosis of an illness or a course of treatment may depend on the outcome of the test. It is therefore vital that the statistical analysis of the assay be performed with the utmost accuracy and precision.

In the literature there are two characteristics of the statistical analysis of assays that stand out:

1. Complicated models are typically used.

2. Approximations are invariably used in the calculation of inferences.

The second characteristic is cause for concern. The involvement of approximations increases the potential for inaccuracies in the statistical analysis to compromise the results from an assay.

1.2   Statistical analyses for an assay

When all assays are considered, the number of distinct statistical analyses performed is very large. It is not feasible to consider each of these within the confines of a single thesis. The analyses covered in this thesis are those performed on immunoassays conducted at NMCH. These are mainstream analyses performed in many types of assays and so there will be widespread applicability. All developments are made in a general framework with the immunoassay being used only as an example. This should allow the methods and results to be easily adapted to other assays.

There are two components to the statistical analysis of immunoassays at NMCH:

1. Estimation of the concentration in the unknown samples.

2. Assessment of the quality of the assay.

The first component is crucial to the subjects undergoing the assay. This involves fitting a model to the data from the assay. The second component is only of interest to the practitioner conducting the assay. Here the assay is thought of as a tool and its sensitivity or precision of measurement is assessed.

There is some comparison between frequentist and Bayesian methods apparent in this thesis, but the main interest is in developing new approaches for these analyses. Many of the new approaches developed use the notion of Bayesian inference.

1.2.1 Frequentist variance function estimation

The model for an assay typically involves three key components:

1. A nonlinear mean function.

2. A variance function.

3. An abundance of parameters for values of the independent variable (the concentrations of the unknowns).

In a frequentist analysis, the presence of a non-constant variance function is often the most problematic component of assay models.

Accounting for heteroscedasticity in the response is a very important and necessary task in a regression analysis of the mean response. If heteroscedasticity is neglected, standard least squares estimates of the mean function parameters can be inefficient and subsequent inferences flawed.

A variance function is a mathematical description of the variance in a response measurement. In a regression analysis of the mean response the inverses of the fitted variances can be used as weights in weighted least squares (WLS). The weights can then be re-estimated and this process repeated several times or continued until convergence is attained. This process is known as iterative reweighted least squares (IRWLS) and the result as generalized least squares (GLS) estimates. When maximum likelihood is used to estimate the mean function, the variance function is simultaneously fitted and applied to the estimation of the mean function. (See Beale and Sheiner (1988) for a review of methods for estimating a variance function.)

In an assay the responses are generally replicated. If all responses are replicated the inverses of the sample variances can be used as weights. This equates to assuming that the variance of the response does not abide to any pattern or model over the range of the independent variable. When the form of the variance function is known the inverses of the sample variances lead to hopelessly inefficient estimates

especially when the responses are only duplicated (Carroll and Cline (1988)). Pooling the information from all the samples and fitting a function that accounts for heteroscedasticity is much more parsimonious and reliable.

An alternative method of dealing with heteroscedastic responses is to transform the responses so as to induce homogeneity (Box and Cox (1964)). The downside of this approach is that the original model for the mean response often does not fit the transformed data. The model's validity is maintained if the data and model are simultaneously transformed (Carroll and Ruppert (1984)). However, the distribution of the transformed response may incur undersirable properties such as skewness. With the variance function approach no such problems are encountered because the method does not entail transforming the data or the model.

Raab (1981) proposed modified maximum likelihood (MML) for the estimation of a variance function in immunoassay. As the name suggests this is based on modifying the likelihood function. The modification to the likelihood function is just the adjustment needed for the maximum likelihood estimate of the variance of a population having a normal distribution with unknown mean to become the standard unbiased estimate. The motivation for the method is that even for simple variance functions, the method of maximum likelihood gives inconsistent estimates (Raab (1981)). In models where the number of parameters increases asymptotically, it is often the case that maximum likelihood estimates are inconsistent (Jewel and Raab (1981)). The root of the problem is the failure of the maximum likelihood method to account for the estimation of the mean responses.

Raab (1981) was perhaps the first major investigation/comparison of methods for estimating a variance function in any area of statistics. This paper was a prelude to general methodological contributions such as Davidian and Carroll (1987) and Beale and Sheiner (1988). In the immunoassay field, papers by Finney (1976), Finney and Phillips (1977) and Rodbard et al. (1976) were forerunners to Raab's paper. The maximum approximate conditional likelihood (MACL) method derived in Sadler and Smith (1986) is an offspring of Raab's MML approach[2].

MML, often in the form of MACL, has become an accepted method of estimating the variance function for immunoassays. A salient characteristic of MML is that no

---

[2]MACL was actually developed independently of MML and was used in assays at NMCH as early as 1980.

use is made of the mean function and the concentrations of the standards. It is apparent that by addressing this, an improvement can be made to MML.

Restricted maximum likelihood (REML) is an appropriate and perhaps the most strongly recommended method of variance function estimation for a regression analysis in which there are no missing values of the independent variable(s). REML was introduced in Patterson and Thompson (1971) where it was coincidentally called modified maximum likelihood! Although maximum likelihood and pseudo likelihood estimates are consistent, they still perform poorly on small amounts of data. The failure to account for the estimation of the mean responses can again be pointed to as the reason for the poor performance of these methods (Harville (1977) and Carroll and Ruppert (1988, pp. 73)). The REML method makes an allowance for the degrees of freedom lost in the estimation of the mean response by adjusting the likelihood equations. The result is that REML performs significantly better on small sets of data. Like MML, REML is based on an adjustment that leads to the usual unbiased estimate of variance when the responses have homogeneous variance.

A REML procedure for an assay that extends the classical REML procedure of Patterson and Thompson (1971) to incorporate the responses with unknown concentrations is developed in this thesis. The resulting procedure makes use of both the information in the responses for the unknowns and the information gained from the knowledge of the concentrations of the standards. This procedure is referred to as extended restricted maximum likelihood (ExREML). The ExREML procedure is also adapted so that it is in the form of MML. This procedure is called extended modified maximum likelihood (ExMML). The ExREML procedure is characterized from a Bayesian point of view and together with ExMML compared via extensive simulation to MML and REML.

1.2.2 Bayesian model for an assay

Accounting for the estimation of the mean responses is a necessary ingredient for efficient estimation of the variance function. Although empirical results indicate that adjustments used in the MML and REML procedures tend to improve the estimates of the variance function parameters (Raab (1981) and see Section 3.4), there is no formal statistical theory to validate the adjustment. Within the frequentist paradigm, however, there is no apparent means of obtaining better estimates.

From a Bayesian viewpoint there is no such problem. Indeed to evaluate an estimate or any other inference problem there is a formal procedure to follow. Lindley (1982) writes: "Once the basic step of describing uncertainty through (subjective) probability is admitted, we have a formal procedure for the solution of all inference problems." The procedure is as follows:

- Say what is known - identify the data X.

- Specify a model for the data generation - a family of distributions for X indexed by θ.

- Specify the uncertainty concerning θ.

- Consider what quantity is of interest - generally part of θ. If, for instance, $\theta = (\theta_1, \theta_2)$ we may only be interested in θ_1.

- Calculate $\pi(\theta \mid x)$, the uncertainty of the quantity of interest given the observed value x of the data. If we are interested in θ_1 integrate θ_2 out and determine inferences.

The problem of estimating a variance function therefore reduces to the calculation of the joint posterior distribution of the variance function parameters or some function thereof. The uncertainty in the mean function is automatically taken account of and a probability distribution is provided for the estimation of the variance function parameters.

In addition to the existence of a formal procedure, there are several other advantages of the Bayesian approach.

- Past experience and other prior knowledge can be incorporated.

- Inferences are conditional on the data rather than the parameters. This avoids the problem of repetition. If an experiment deviates from its plan, prespecified levels of significance do not become invalid. The data that has been observed can be conditioned on at any stage of the experiment and used to evaluate inferences.

- The statistical significance of results is easier to gauge, interpret and discuss and is not influenced by the number of inferences that may be required.

In the literature there are only a few accounts of the use of Bayesian analysis in assays. Ramsey (1972) and Antoniak (1974) used Bayesian analysis for a bioassay. However, the experiment considered by these authors was of an entirely different nature to that considered here. Racine-Poon, Weihs and Smith (1991) used a Bayesian model to analyze a radioimmunoassay where the concentrations of the standards are subject to sequential dilution errors but relied on analytical approximations.

Bayesian analysis has been used to analyze calibration problems with the models generally being more simple than the atypical assay model. See Dunsmore (1968), Hoadley (1970), Hunter and Lambroy (1981) and Brown (1982) for Bayesian approaches to the linear calibration problem and Osbourne (1991) for a review of statistical calibration. Incidentally, Osbourne (1991) lists Racine-Poon et al. (1991) as the sole example of Bayesian analysis applied to a nonlinear calibration problem.

In other fields, Bayesian analysis has been applied to models involving a nonlinear mean function, a variance function and even missing values for the independent variable. One such instance is Wakefield et al. (1994) where Bayesian analysis is used to analyze a population model.

A full Bayesian approach to the analysis of an assay is developed in this thesis. The practical side of the analysis is also addressed with computational routines and related discussion being included. The Metropolis-Hastings (M-H) algorithm (Metropolis et al. (1953) and Hastings (1970)) is successfully used to evaluate inferences for a Bayesian model of an assay. The M-H algorithm is the basis of the evaluation of all Bayesian inference problems in this thesis.

If sufficient computer power is available, Bayesian analysis is a means of determining inferences to any degree of accuracy. It will also enable practitioners to incorporate expert knowledge and their own past experiences into the analysis. These attributes will surely reduce the potential for a statistical analysis to adversely affect the results of an assay.

1.2.3 Minimal detectable concentration

The minimal detectable concentration (MDC), the smallest concentration that an assay can reliably measure, is a measure of the sensitivity of an assay. It reflects the capability of the assay to detect a small positive concentration. The MDC is commonly used to evaluate the limitations of an assay and to make comparisons

between assays. The object of measurement assumed in this thesis is a hypothetical unseen sample that is independent of the samples analyzed by the assay. The sensitivity of the assay is thus reflected in its ability to detect a positive concentration in this sample.

A plethora of statistical criteria have been used to define the term "reliable measurement" (Currie (1968), Oppenheimer et al. (1983) and Brown et al. (1996)). These fall into two categories:

- Those based on the detection of positive concentration.

- Those related to the variance or coefficient of variation of a calibrated (i.e. estimated) concentration.

The MDC usually pertains to just the first category or notion. Sometimes the MDC has encompassed both notions. The above references are instances of this. In this thesis, the MDC is associated with just the first notion of a reliable measurement. The second notion gives rise to the precision profile, another widely used diagnostic of assay performance. The MDC and the precision profile are both major focal points of this thesis. For this reason they are treated separately.

The MDC has the potential to be used as a criterion for optimizing, or at least improving, the design of an assay. The statistical design of an assay is defined by the concentrations and replications of the standards and in certain situations the replications of the unknowns. The smaller the MDC, the lower the concentration that can be reliably measured and the more powerful the assay. The MDC-optimal design for an assay is the design for which the expected value of the MDC is minimized subject to necessary constraints on the number of standards and the total number of response measurements made on the standards and if applicable, the unknowns.

In this thesis, several new measures of the MDC are developed. The following existing measures of MDC are reviewed: the critical limit (CL) and the detection limit (DL) (Currie (1968)) and the MDC as discussed in Davidian et al. (1988). This leads to the consideration of some alternative frequentist measures although the main focus is on the development of measures of MDC to accompany a Bayesian analysis of an assay.

Bayesian measures of MDC are shown to have two significant advantages over frequentist measures. Firstly, provided there is sufficient computational resource, the Bayesian measures of MDC can be precisely evaluated whereas approximations are invariably needed in order to evaluate frequentist measures. It is hypothesized that Bayesian measures of MDC more accurately depict the quality of the assay. This thesis attempts to illustrate that this hypothesis is true. The second advantage is that the MDC can be defined directly in terms of whatever decision making criteria for detection is used by the practitioner. This will lead to measures which can be easily understood by practitioners.

1.2.4 The precision profile

The precision profile is a device for displaying the quality of the assay over all concentrations, whereas the MDC considers only concentrations near zero. It was originally described in Ekins (1978) and formally defined in Ekins (1983). Imprecision has been used in place of precision by Sadler, Smith and Legge (1988), Sadler and Smith (1990a,b) and other authors. Essentially, a precision profile is a plot of the precision of the assay at a given concentration against concentration and precision is a diagnostic that quantifies the error in the calibrated concentration. The smaller the distance from the precision profile to the concentration axis the higher the quality of the assay.

To cater for five different analytical requirements of assays, Ekins (1983) defines five different precision profiles. An analytical requirement of an assay consists of the sources of error that are to be taken account of in the calculation of an inferential statement about the assay. The definitions of Ekins vary according to whether the calculation of precision involves one or multiple distinct samples of the substance being tested (intra-sample verse inter-sample), one or more assays (intra-assay verse inter-assay) and different laboratories (inter-laboratory).

In this thesis only the intra-sample, intra-assay precision profile is considered. This is the fundamental criterion of assay performance which governs assay design (Ekins (1983)). In Ekins' definition of the intra-sample, intra-assay precision profile the error in the fitted assay model is ignored. The resulting precision profile therefore reflects the error in the assay due to only the randomness of the independent response measurement. This is called "bottom line" or intrinsic assay error. Practitioners

regard Ekins' intra-sample, intra-assay precision profile as a valuable indicator of the level of intrinsic assay error.

The viewpoint taken in this thesis is that the intra-sample, intra-assay precision profile should reflect all of the error in an assay's calibration of an unknown concentration. The same stance is assumed in O'Connell, Belanger and Harland (1993). The analytical requirement of Ekin's original intra-sample, intra-assay precision profile is effectively expanded to that of the intra-sample, inter-assay precision profile when all distinguishing features of the assay are identical or, in other words, homogeneous. A precision profile along these lines appears generally desired by practitioners. The methods developed in this thesis for calculating such a precision profile can also be used to calculate the intra-sample, intra-assay precision profile of Ekins. All that needs to be done is to disregard the error in the parameter estimates of the assay parameters in the calculation of precision. In other words the parameter estimates are treated as the true values of the assay model parameters.

The general intra-sample, inter-assay precision profile only differs from the precision profile described in the preceding paragraph when the assays are heterogeneous. The calculation of the precision profile in this case and in the inter-laboratory case necessitates extensions to the assay model that are beyond the scope of this thesis. Inter-sample precision profiles are not discussed in this thesis as they appear to be neglected in the literature. Furthermore, it is envisaged that the methods developed for intra-sample precision profiles could be easily adapted to these. Unless otherwise stated, the term precision profile shall from hereon be taken to mean intra-sample, intra-assay precision profile and precision will mean intra-sample, intra-assay precision.

The calculation of precision at a specific level of concentration is fundamental to the construction of the precision profile. The precision profile is just the mechanism for graphically displaying the precision of the assay over the range of concentrations. To improve the accuracy of the precision profile the calculation of precision must be improved.

Precision has only been expressed in terms of the estimated variance, standard deviation or coefficient of variation of the calibrated concentration by Ekins and other authors (O'Connell et al. (1993), Sadler et al. (1988), Sadler and Smith (1990a)). There is no evidence in the literature of a Bayesian definition of precision.

From the frequentist viewpoint, a concern with the use of standard deviation and coefficient of variation to measure precision is that the concept of "bias" is not considered. Only the variability of the estimated concentration is considered. Hence, a constant would appear to be a perfect estimate. For variance based measures to be reliable the bias of the estimate must be negligible. If bias is not negligible the precision profile will not be an accurate reflection of the quality of the assay. This does not appear to have been noted in the literature. A corrective action is to replace variance with mean squared error in the calculation of precision. This point is made just in passing, incumbent methods are not adapted in this thesis as the effect of bias on the calculation of precision for "sensible" estimators is minimal.

To ensure that precision is a reliable indicator of an assay's quality it is apparent that precision must mirror the MDC. Precision must be based on the measurement of the concentration in a future sample. In this thesis the main development is the calculation of precision from a Bayesian viewpoint. However, currently used measures that have been reported in the literature are critiqued and suggested improvements are indicated.

A by-product from the development of Bayesian measures of precision is a class of distributions that completely describe an assay's measurement of an unknown concentration. The outstanding feature of these distributions is that they express all the information gained from the assay in terms of a probability distribution for the calibrated concentration, the entity being measured, conditional on the actual concentration. This (These) distribution(s) is (are) known as "the predictive distribution(s) of assay measurements" (denoted PDAM). At concentration x the PDAM is denoted as $PDAM(x)$. The class $\{PDAM(x), x \in [0, \infty)\}$ describes completely an assay's performance over the full range of concentrations.

The PDAM does not appear to have been developed previously[3]. These distributions can be used for more general calculations than just the calculation of the precision of an assay. The PDAM can be used to assess the performance of any measuring device. Therefore, this development is of fundamental importance.

---

[3]However, Seymour Geisser laid the foundation for the development of a similar entity in his work on Bayesian interim analysis (see Section 8.6.3).

1.2.5   Machine assays and quality diagnostics for batches of assays

In the last five years machines have been developed which automate the analysis of an assay and so reduce the level of human input[4]. A hallmark of these machines is that statistical algorithms for model fitting are embodied in them and the intermediate components of the analysis are hidden. At best only the final estimate and standard error of the quantity being measured and the associated raw response measurements are output. At worst (and most commonly) just the final estimates are divulged. Although minimizing the level of output may simplify the task of the operator, a lot of useful information about the assay is lost.

An overview of several faults with the statistical analysis performed by assays analyzed internally by a machine (machine assays) will be given. Attention is, however, mainly directed at the assessment of the quality of a machine assay and the limitations therein. It is apparent that data from a batch of two or more homogeneous assays are needed to calculate the quality of a machine assay in the manner described in Sections 1.2.3 and 1.2.4. For a reliable assessment of an assay's quality, data from a batch of many (ten or more) homogeneous assays are needed. Even then one is limited to the precision profile because the MDC cannot be calculated. A method for calculating the precision profile with data from a batch of homogeneous machine assays is outlined in this thesis.

The assessment of the overall quality of a batch of homogeneous assays in which the statistical analysis is controlled by the practitioner (manual assays) is also considered. (Note that a manual assay is known simply as an "assay" in parts of the thesis where machine assays are not discussed.) With data from a batch of homogeneous assays a more informed judgement of the quality of the assays' design can be made than from just one assay. The measures of MDC and precision discussed in the thesis for a single assay are extended to measures of the overall quality of a batch of homogeneous manual assays.

---

[4]The elimination of the human element in the analysis of experimental data is not limited to assays. In the pharmaceutical industry the number of patents being taken out on machines which automate laboratory procedures and analyses is growing (Arveson (1998)).

1.3 Overview

Before a parametric analysis of an assay can be undertaken, a statistical model for the observed data is required. A probability distribution must be specified for the observed counts given the concentration in the sample. In Chapter 2 the statistical model for the assay responses is described. Some justification of this model is given but specific details are consigned to Appendix A. The model is also expressed in a general form to enable the methods developed in this thesis to be easily generalized to other models. The notation to be used throughout the thesis is described in Section 2.2.

Frequentist methods for estimating a variance function for an assay are reviewed and improved upon in Chapter 3. In Section 3.1 the currently used MML method is reviewed. Following discussion of the short-comings of this method the ExREML procedure is developed in Section 3.2 and from this the ExMML procedure is deduced. A feature of this chapter is the characterization of the ExREML procedure in Section 3.3. The classical and extended versions of MML and REML are compared via simulation in Section 3.4. The mathematical package MATLAB is used for these simulations.

An important point to note is that while the main focus is on the estimation of the variance function, ExMML and ExREML fit every component of an assay model. These methods are therefore new techniques for the estimation of an assay model, not just the variance function.

In Chapter 4, a Bayesian model is developed for the analysis of an assay. In Section 4.1 the form and specification of the prior distribution for the assay parameters is described. This is followed in Section 4.2 with the application of the Bayesian method to the analysis of an assay. To illustrate the method the calculations are outlined for a simple model in Section 4.3. It is also shown how these calculations may be used to obtain approximate results in general situations. Exact calculation via the M-H algorithm is described in Section 4.4. The algorithm used for these calculations is described and in the associated discussion several tips for the successful implementation of the algorithm are detailed. The analysis of data from a batch of assays is considered in Section 4.5. The ANSI C programming language is used for all of the Bayesian computations reported in this thesis.

Once a procedure for fitting the assay model is in place, the assessment of the quality of an assay can be undertaken. In Chapter 5 existing measures or diagnostics for the MDC of an assay are reviewed (Section 5.2) and new measures are developed (Section 5.3). A feature of the existing measures is that they are generally able to be only approximately evaluated. Some alternative approximations to one of the existing measures are also proposed in Section 5.2. The new measures of MDC arise from theses of what a practitioner would want the MDC to represent. The thesis, definition and calculation of each new measure of MDC are contained within the section in which the measure is introduced. The new measures use the notion of Bayesian inference and may be calculated using the M-H algorithm. The necessary additions to the algorithm that fits the assay model are given. In Section 5.4 the measures are characterized and compared using simulated data. This leads to a recommendation of which measures of MDC are the best to use in practice.

Chapter 6 begins with a review of the methods currently being used to calculate the precision profile of an assay. In Section 6.2 some prospective improvements to these calculations are suggested. The main body of work is in Section 6.3 where measures of precision for a Bayesian analysis of an assay are defined. Preliminary development takes place in Section 6.3.1. In Section 6.3.3 the issue of what precision represents is addressed and one new measure of precision is developed. In Section 6.3.4 the measure developed in Section 6.3.3 is shown to give rise to the PDAM in a special case. A general PDAM based measure of precision is developed in Section 6.3.5. Further applications of the PDAM are considered in Section 6.3.5.4. Finally, in Section 6.4 the measures of precision are compared using a real data set.

The operation of a machine assay is the first point of discussion in Chapter 7. In Sections 7.2 and 7.3 the assessment of the quality of a machine assay is considered. Section 7.3 describes the involvement of batches of homogeneous assays. In Section 7.4 the measures developed in Chapters 5 and 6 are extended to diagnostics of the overall quality of batches of homogeneous manual assays.

Finally, in Chapter 8 the principle discussions and conclusions are reviewed. An outlook to future work is also given.

Chapter 2

Model and notation

The standard model for the analysis of an immunoassay is introduced and informally justified in Section 2.1. The notation used in the remainder of the thesis shall then be defined.

2.1 Statistical model for an immunoassay

In an immunoassay the response count is the number of molecules of the compound being assayed that either do or do not react with the antibody (refer to Appendix A for details). If the reaction were such that each molecule of the compound being tested had an equal chance of partaking in the reaction and acted independently of all other molecules of the compound, then the response would have a hypergeometric distribution. A variety of complicating factors in the measurement of the response count serve to make the Poisson distribution a more realistic model of the response distribution under "ideal conditions". For example, in radioimmunoassay, although the number of reacting molecules is hypergeometric the number of radioactive parts actually detected (counted or assayed) is Poisson.

In reality ideal conditions rarely exist. The molecules will be serially correlated as the binding of a molecule to the antibody tends to increase the affinity of neighbouring molecules for the antibody. Random errors from the laboratory analysis will inflate the variance of the counts. The error in the measurement of the reagents and errors incurred while separating the products of the reaction are examples of such errors. It is also unlikely that the probability of a molecule reacting will be the

same for all molecules.

Even though the count arises from a sum of events that cannot be considered to be independent, the distribution of the total number of bound (reacted) or free (unreacted) molecules will still converge to a normal distribution under fairly robust conditions. This convergence is due to a central limit theorem for correlated variables (see pages 375-379 of Billingsley (1986)). The super Poisson nature of the variances can be accommodated by the normal family of distributions by expressing the variance of the counts as a parametric function of the mean response. This function should be flexible enough to allow wide ranging behaviour in the variance of the response to be accommodated. On these grounds, the normal distribution is a reasonable model for immunoassay responses. A literature search for assay models indicated that the normal distribution is the standard distribution used for modelling immunoassay responses.

Let y_{ij} denote the jth of r_i responses for the ith of n samples analyzed by the assay and η_i the concentration of the ith sample. The standard model used for the analysis of an assay is

$$y_{ij} \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma_i^2), \ j = 1, \ldots, r_i, \ i = 1, \ldots, n,$$

where $\mu_i = m(\eta_i, \beta)$, $\sigma_i^2 = v(\mu_i, \theta)$ and $N(\mu_i, \sigma_i^2)$ denotes the normal distribution with mean μ_i and variance σ_i^2. In this model m and v are functions, and the parameters, β and θ, are p and q dimensional vectors respectively.

The sigmoid function,

$$m(x, \beta) = \beta_4 + \frac{\beta_1 - \beta_4}{1 + (x/\beta_3)^{\beta_2}}, \tag{2.1}$$

is often a suitable mean function for the analysis of an immunoassay (Healy (1972), Rodbard and Hutt (1974), Finney (1976), De Lean, Munson and Finney (1978)). The sigmoid function is a member of the rich logistic family of curves.

Common variance functions include

$$v(m, \theta) = \theta_1 m^{\theta_2} \tag{2.2}$$

and

$$v(m, \theta) = \theta_1 m + \theta_2 m^2. \tag{2.3}$$

2.1.1   General distribution of the response

The general form of the distribution of the responses assumed in this thesis is

$$y_{ij} \sim f(y_{ij} \mid \eta_i, \beta, \theta)$$

where f is any probability distribution that is completely specified by its mean and variance. In other words, f is a two parameter family of distributions. The restriction to two parameter families is solely for convenience. There is no reason why distributions with three or more parameters cannot be used. When the analysis of an assay is Bayesian, a prior distribution for the unknown parameters must also be specified. Various possibilities for the prior distribution are discussed in Section 4.1.

2.1.2   NMCH data and model

Throughout the course of this thesis, the data from a serum thyroxine radioimmuno-assay[1] conducted at NMCH on March 17, 1995 will be used for illustrations. These data shall be referred to as the NMCH data.

In the NMCH data there are 7 standards and 75 unknowns (i.e. there are 82 concentrations of which only 7 are known). The concentrations of the standards are denoted by

$$X = (0, 35, 70, 140, 210, 280, 350). \tag{2.4}$$

Each sample is duplicated except for the zero standard which is quadruplicated.

The mean function to be fitted to the NMCH data is the sigmoid function. By convention this describes the percentage bounds rather than the raw responses. The percentage bound is an affine transformation of the raw responses (see equation A.1 in the Appendix). The model for the mean response at concentration x is

$$m(x, \beta) = \frac{\text{MnTotal}}{100} \left(\beta_4 + \frac{\beta_1 - \beta_4}{1 + (x/\beta_3)^{\beta_2}} \right) + \text{NSB}, \tag{2.5}$$

where MnTotal and NSB are assay specific constants (see Appendix A) and for the NMCH data have values 35520 and 740 respectively. The variance function is the power function of the mean response given in equation 2.2. The responses will be assumed to be normally distributed. The model just described will be referred to as the NMCH model.

---

[1]This is a reagent limited assay. For details see Appendix A or Davies (1994).

The NMCH data will also serve as the basis for the computer generated data used in simulations. The concentrations of the standards are equal to the value of X in (2.4) and the mean function is the affine transformed sigmoid function given in (2.5) with $\beta = (55, 1.2, 60, 4)'$, MnTotal= 35520 and NSB= 740. The variance function will be either (2.2) or (2.3). To generate assay data β, θ, η^U and r also need to be specified. The values of these terms vary with the simulation being performed and so will be documented when and where it is appropriate.

2.2   Notation

In this thesis two sets of notation are required corresponding to the analysis of a single assay and of a batch of assays. However, there is a large overlap between the two. The notation for a batch of assays requires just a few amendments to the notation used for the analysis of a single assay.

2.2.1   A single assay

As mentioned above, f will denote the probability density function (pdf) of the observed responses. The marginal distribution function of the observed data is also indicated using the symbol f. Prior distribution functions shall be denoted using the symbol π. The output from a Bayesian analysis consists of the posterior distribution for the unknown parameters and sometimes a predictive distribution for as yet unobserved responses. Posterior distribution functions are also denoted using the symbol π while predictive distribution functions of response variables use the symbol p. All of these distributions will be described in detail in Chapter 4.

For the most part no distinction is made between vectors and scalars. A subscript will distinguish a component of a vector from the full vector. When the argument of a function is a vector as opposed to a scalar, the vector is evaluated componentwise with the dimension of the vector being maintained. For matrices or general two dimensional arrays, a single subscript refers to a row and a double subscript to individual elements.

Greek letters are used to denote parameters. A Greek letter with a hat over it is an estimate of the parameter. A vector or an array with a set as a superscript

indicates that the entity is reconstructed using just the observations in the respective set.

The expectation operator will be denoted by E and the variance operator by V. The space a definite integral is evaluated over is denoted by $\mathcal{R}(z)$ where z is the variable being integrated.

For simplicity no distinction is made between random variables and their realizations. The indices are also arranged so that the standards are first.

The notation for a single assay is as follows:

- The set of indices for the standards: S.

- The set of indices for the unknowns: U.

- The number of standards: n^s.

- The number of unknowns: n^u.

- The total number of samples: n.

- The vector containing the number of replicates for each sample: r.

- The two dimensional array containing all of the observed responses:

$$Y = \begin{pmatrix} Y^S \\ Y^U \end{pmatrix}.$$

- The vector of responses for the ith sample: Y_i. This is the ith row of Y.

- The mean response for the ith sample: $\bar{Y}_i$.

- The vector of all the concentrations:

$$\eta = \begin{pmatrix} \eta^S \\ \eta^U \end{pmatrix}.$$

- The vector of the concentrations for the standards: $X = \eta^S$.

- The assay data: (Y, X).

- An arbitrary vector of responses which, conditional on the associated concentrations, is independent of the data observed in the assay: y.

- The known or true concentration of an independent response: x.

- The unknown concentration of an independent response: ω.

- The mean function parameters: β.

- The dimension of β: p.

- The variance function parameters: θ.

- The dimension of θ: q.

- The model parameters: (β, θ).

- The unknown concentration parameters: η^U.

- The assay parameters: (β, θ, η^U).

- An arbitrary level of mean response: m.

- The vector of parameters for the mean response: $\mu = m(\eta, \beta)$.

- The vector of estimates of the mean responses: $\hat{\mu} = m(\hat{\eta}, \hat{\beta})$.

- The gradient of the mean function at concentration x: $m_\beta = m_\beta(x, \beta) = \frac{\partial}{\partial \beta} m(x, \beta)$.

- The fitted gradient of the mean function at concentration x: $\hat{m}_\beta = m_\beta(x, \hat{\beta})$.

- The Jacobian matrix of the vector mean function for the standards: $J_\beta = m_\beta(X, \beta)$. The transpose of $m_\beta(X_i, \beta)$ is the ith of n^s rows of J_β.

- The fitted Jacobian matrix of the vector mean function for the standards: $\hat{J}_\beta = m_\beta(X, \hat{\beta})$.

When required, additional notation is defined for use in specific sections of the thesis.

2.2.2 A batch of assays

When a batch of assays are analyzed, an additional subscript is required to indicate which assay the data and model belong to. For example, Y_{hij} is defined as the jth of r_{hi} replicated response measurements for the ith of n_h samples in the hth of b assays and η_{hi} is the associated concentration. The preceding rules for two dimensional arrays carryover to arrays that are two dimensions and higher. For example, Y is a three dimensional array of the responses over all assays, Y_h is the two dimensional array of all the responses in the hth assay and Y_{hi} is the vector of responses in the ith sample of the hth assay.

The extra subscript extends to other entities in an analogous fashion. For example, S_h is the set of indices for the standards in the hth assay and S contains pairs of indices indicating which samples are standards in each of the assays. If $(h, i) \in S$, or equivalently $i \in S_h$, then $\eta_{hi} = x_{hi}$, otherwise $\eta_{hi} = \eta_{hi}^U$. The vector of the concentrations of the standards in the hth assay is X_h and X is the two dimensional array of the concentrations of the standards for all of the assays in the batch.

Chapter 3

Variance function estimation

The purpose of the research reported in this chapter is to develop better frequentist methods of estimating a variance function for an assay. It is convenient that the proposed improvements involve fitting the assay model and therefore suggest new methods for estimating the mean function and calibrating the unknown concentrations. The technical developments in this chapter are made in terms of the standard immunoassay model.

3.1 Modified maximum likelihood

Modified maximum likelihood (MML) is based upon maximum likelihood estimation of the model $Y_{ij} \sim N(\mu_i, v(\mu_i, \theta))$. This is the probability density function of the responses when the mean function and the concentrations of the standards are ignored. The likelihood function for this model is given by $L = \prod_{i=1}^{n} f(Y_i \mid \mu_i, \theta)$, where

$$
\begin{aligned}
f(Y_i \mid \mu_i, \theta) &= \prod_{j=1}^{r_i} \frac{1}{\sqrt{2\pi v(\mu_i, \theta)}} \exp\left\{ -\frac{(Y_{ij} - \mu_i)^2}{2v(\mu_i, \theta)} \right\} \\
&\propto v(\mu_i, \theta)^{-r_i/2} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{r_i} \frac{(Y_{ij} - \mu_i)^2}{v(\mu_i, \theta)} \right\}.
\end{aligned}
\tag{3.1}
$$

The MML procedure for variance function estimation is to adjust the likelihood function prior to maximization. The adjustment is simply to multiply (3.1), the contribution of the ith sample to the likelihood function for all of the experimental

23

observations, by $\sqrt{v(\mu_i, \theta)}$. The resulting modified log likelihood function for the assay data is proportional to

$$\mathcal{L} = -\frac{1}{2} \sum_{i=1}^{n} \left\{ (r_i - 1) \log(v(\mu_i, \theta)) + \sum_{j=1}^{r_i} \frac{(Y_{ij} - \mu_i)^2}{v(\mu_i, \theta)} \right\}. \qquad (3.2)$$

The MML estimates of θ and μ (the nuisance parameter in this context) are obtained by maximizing $\mathcal{L}$ over the parameter space of θ and μ. The adjustment accounts for the loss of degrees in freedom resulting from the estimation of the nuisance parameter μ.

The adjustment to the likelihood function is theoretically justified on the basis that it aligns the maximum likelihood estimates with the standard parameter estimates for the model $Y_{ij} \overset{\text{ind}}{\sim} N(\mu_i, \sigma^2)$. The pseudo likelihood function[1] obtained when the maximum likelihood estimate of μ is substituted into this adjusted likelihood function is "correct" in the sense that it is "the integrated likelihood for σ^2" and is "the conditional likelihood of the observations at the observed values of the sufficient statistics $\bar{Y}_i$ for μ_i" (Raab (1981)). As the pseudo likelihood for σ^2 is equivalent to the integrated likelihood for σ^2, the MML estimate of σ^2 is also the mode of the (marginal) posterior distribution of σ^2 when the prior for (μ, σ^2) is proportional to a constant. Furthermore, MML estimates of σ^2 are unbiased whereas maximum likelihood estimates are not even consistent. If $\max\{r_i, i = 1, \ldots, n\} = k$ then maximum likelihood estimates are asymptotically negatively biased by at least σ^2/k.

The above mentioned derivations of the modified likelihood function do not generalize to the case where the variance depends on the mean (Raab (1981)). Jewell and Raab (1991) calculated a marginal likelihood for the coefficient of variation that yields consistent estimators when the variance is proportional to the square of the mean but the method also fails to generalize. However, numerical evidence suggests that MML gives consistent estimates of θ, in some cases even unbiased estimates (Raab (1981)).

Since MML determines the weights for the estimation of the mean function independently of estimates of β, only one iteration of iterative reweighted least squares (IRWLS) is required to obtain final estimates of all the parameters. This is

---

[1]A pseudo likelihood function is a likelihood function in which parameter estimates are substituted for some parameters.

at the expense of any information that would be obtained from the residuals of the fitted mean function. It seems reasonable that a more efficient estimate of θ will result if the fits from the mean function are used as estimates of μ^S in place of $\bar{Y}^S$. Only p degrees of freedom are expended as opposed to the n^s degrees of freedom needed to individually estimate each component of μ^S. With this motivation an improved method to MML is now developed.[2]

3.2 REML for an assay

In regression analyses where there are no missing values for any of the independent variables, restricted maximum likelihood (REML) is an appropriate and perhaps the most strongly recommended means of estimating a variance function. By accounting for the degrees of freedom lost in the estimation of the means, REML performs significantly better than maximum likelihood and pseudo likelihood estimates. REML has the added theoretical appeal of being equivalent to some standard non-informative Bayesian estimates. It clearly makes sense to follow the REML methodology to incorporate the mean function and use the known concentrations in the estimation of a variance function for an assay.

In order for REML to be privy to information in the replicated responses of unknown samples, η^U must be estimated in conjunction with β. This is achieved by writing the expectation of the ijth response in the extended form

$$
\begin{aligned}
E[Y_{ij} \mid \eta_i] &= \sum_{k=1}^{n} m(\eta_k, \beta) I(k = i) \\
&= m(X_i, \beta) I(i \in S) + \mu_i I(i \in U),
\end{aligned}
\tag{3.3}
$$

where I is the indicator function, $I(\text{event}) = 1$ if "event" is true and 0 otherwise, and $\mu_i = m(\eta_i, \beta)$.

By treating (3.3) as the mean function[3], generalized least squares (GLS) estimates of μ^U will be calculated along with those of β in the WLS (weighted least squares) step. The re-parameterization of η^U to μ^U greatly simplifies the exposition of the method and the computation of the estimated parameters because

---

[2]In multivariate situations where the observations are correlated, as in universal kriging, there is some disagreement as to whether or not it is better to use the fitted mean function.

[3]In classical REML the mean function would only comprise the left hand term of (3.3) causing the unknown samples to be ignored.

it uncoupled β and η^U. The estimates of the parameters are unaffected by re-parameterization since the function being maximized is unchanged.

Let $\hat{\beta}$ and $\hat{\mu}^U$ be GLS estimates of β and μ^U respectively. Since $\hat{\mu}^U$ is equal to $\bar{Y}^U$ at each iteration $\hat{\mu}_i = m(X_i, \hat{\beta}), i \in S$ and $\hat{\mu}_i = \bar{Y}_i, i \in U$.

The development in Carroll and Ruppert (1988, pp. 74) is now followed. Recall that when the argument of a function is a vector as opposed to a scalar, the vector will be evaluated componentwise with the dimension of the vector being maintained. Define $u(\mu_i, \theta) = \log(v(\mu_i, \theta))$, $u_\theta(\mu_i, \theta) = \frac{\partial}{\partial \theta} u(\mu_i, \theta)$, $e(\mu_i, \theta) = \sqrt{r_i}(\bar{Y}_i - \mu_i)/\sqrt{v(\mu_i, \theta)}$ and let $S_i^2 = \sum_{j=1}^{r_i}(Y_{ij} - \bar{Y}_i)^2$. Also let $W = W(\hat{\mu}^S, \theta) = \text{diag}\{r_i/v(\hat{\mu}_i, \theta), i \in S\}$ and $H = H(\hat{\beta}, \theta) = W^{1/2} \hat{J}_\beta (\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W^{1/2}$. The matrices W and H are commonly referred to as the weight matrix and nonlinear "hat" matrix respectively. The $n^s \times p$ first derivative matrix for the standards is $\hat{J}_\beta = m_\beta(X, \hat{\beta})$.

At $\hat{\mu}$, the pseudo likelihood in θ solves the equation

$$\sum_{i=1}^{n} \frac{S_i^2}{v(\hat{\mu}_i, \theta)} u_\theta(\hat{\mu}_i, \theta) + \sum_{i \in S} e(\hat{\mu}_i, \theta)^2 u_\theta(\hat{\mu}_i, \theta) = \sum_{i=1}^{n} r_i u_\theta(\hat{\mu}_i, \theta), \qquad (3.4)$$

since $e(\hat{\mu}_i, \theta) = 0$, $i \in U$. All the terms involving the residuals $Y_{ij} - \hat{\mu}_i$ are purposely placed on the left hand side (lhs) of (3.4). The REML procedure is to equate the lhs of (3.4) to its expectation and solve for θ. The idea is that the resulting adjustment to (3.4) accounts for the degrees of freedom lost in the estimation of μ.

The above expectations can be evaluated when $m(\eta, \beta)$ is a linear function of β and $v(m, \theta)$ is independent of m. However, if $m(\eta, \beta)$ is a nonlinear function of β or v depends on m, then approximations are required.

The procedure implied by Carroll and Ruppert (1988, pp. 74) is to approximate the expectation of the lhs of (3.4) as follows:

1. Treat $v(\hat{\mu}_i, \theta)$, $u(\hat{\mu}_i, \theta)$ and $u_\theta(\hat{\mu}_i, \theta)$ as if they did not depend on $\hat{\mu}_i$ so that $(r_i - 1)v(\hat{\mu}_i, \theta) \equiv E[S_i^2]$ and hence

$$E\left[\frac{S_i^2}{v(\hat{\mu}_i, \theta)}\right] = r_i - 1,$$

$$E\left[\frac{S_i^2}{v(\hat{\mu}_i, \theta)} u_\theta(\hat{\mu}_i, \theta)\right] = E\left[\frac{S_i^2}{v(\hat{\mu}_i, \theta)}\right] u_\theta(\hat{\mu}_i, \theta) = (r_i - 1) u_\theta(\hat{\mu}_i, \theta)$$

and

$$E[e(\hat{\mu}_i, \theta)^2 u_\theta(\hat{\mu}_i, \theta)] = E[e(\hat{\mu}_i, \theta)^2] u_\theta(\hat{\mu}_i, \theta).$$

2. Set $E[e(\hat{\mu}_i, \theta)^2] = 1 - H_{ii}, i \in S$, where $H_{ii} = H_{ii}(\hat{\beta}, \theta)$ is the ith diagonal element of H.

The details of the second approximation as presented in Carroll and Ruppert (1988) are somewhat sketchy. It is based on a first order Taylor series expansion of $\hat{\mu}^S = m(X, \hat{\beta})$ about β and the adaption of the result that $E[e(\hat{\mu}^S, \theta)e(\hat{\mu}^S, \theta)'] = I - H$ when m is linear in β and v does not depend on β. (See Carroll and Ruppert (1988, pp. 32, 33, 74) for further details.)

After incorporating the above approximations, the expectation of the lhs of (3.4) becomes

$$\sum_{i=1}^{n}(r_i - 1)u_\theta(\hat{\mu}_i, \theta) + \sum_{i \in S}(1 - H_{ii})u_\theta(\hat{\mu}_i, \theta) = \sum_{i=1}^{n}(r_i - l_i)u_\theta(\hat{\mu}_i, \theta),$$

where $l_i = l_i(\hat{\mu}_i, \theta)$ is equal to H_{ii} if $i \in S$ and 1 if $i \in U$. The ith leverage value l_i measures the amount of information in the data from the ith sample used to estimate μ (equivalently β and μ^U). The REML estimating equation for θ is thus

$$\sum_{i=1}^{n} \frac{S_i^2}{v(\hat{\mu}_i, \theta)} u_\theta(\hat{\mu}_i, \theta) + \sum_{i \in S} e(\hat{\mu}_i, \theta)^2 u_\theta(\hat{\mu}_i, \theta) = \sum_{i=1}^{n}(r_i - l(\hat{\mu}_i, \theta))u_\theta(\hat{\mu}_i, \theta). \tag{3.5}$$

The resulting estimates are referred to as extended restricted maximum likelihood (ExREML) estimates to avoid confusion with classical REML estimates. The subtraction of l_i from r_i is the only alteration to the right hand side (rhs) of (3.4). The pseudo likelihood estimating equation is in effect adjusted to account for the estimation of μ.

The difference between the rhs of (3.5) and the true expectation of the lhs of (3.5) depends on the level of variation in the data and the curvature of μ. This is discussed further in Sections 3.4 and 8.1. When the mean function is linear in β and the variance function is independent of μ the discrepancy vanishes.

The adjustment can be imposed on the pseudo likelihood function for θ at $\hat{\mu}$, a WLS estimate of μ, to give the (modified) pseudo likelihood function

$$L = \prod_{i=1}^{n} v(\hat{\mu}_i, \theta)^{-(r_i - l(\hat{\mu}_i, \theta))/2} \exp\left\{ -\frac{\sum_{j=1}^{r_i}(Y_{ij} - \hat{\mu}_i)^2}{2v(\hat{\mu}_i, \theta)} \right\}. \tag{3.6}$$

The value of θ that maximizes L is an ExREML estimate of θ. The final ExREML estimate of θ is obtained by iteratively updating the generalized least squares estimate of μ and the ExREML estimate of θ until convergence is reached. This process, iterative reweighted least squares, is formalized in the following algorithm:

1. Use the standards to obtain $\hat{\beta}$, an initial estimate of β.

2. Set $\hat{\mu}_i = m(X_i, \hat{\beta})$ for standards and $\hat{\mu}_i = \bar{Y}_i$ for unknowns.

3. Set $l_i = l(\hat{\mu}_i, \theta) = H_{ii}(\hat{\beta}, \theta)$ for standards and $l_i = 1$ for unknowns.

4. Calculate $\hat{\theta}$ by maximizing (3.6).

5. Use the standards to obtain, $\hat{\beta}$, the WLS estimate of β corresponding to $\hat{\theta}$.

6. If convergence has not yet been attained return to step 2.

3.2.1 Variation of ExREML

Instead of substituting the estimates of β and μ^U into (3.6) one could estimate β, θ and μ^U simultaneously. The procedure is to maximize the full modified likelihood function

$$L = \prod_{i=1}^{n} v(\mu_i, \theta)^{-(r_i - l(\mu_i, \theta))/2} \exp \left\{ \frac{- \sum_{j=1}^{r_i} (Y_{ij} - \mu_i)^2}{2v(\mu_i, \theta)} \right\}$$

over β, θ and μ^U, where $\mu_i = m(X_i, \beta)$ if $i \in S$. This is MML extended to take account of the mean function and the known concentrations. This method of estimation is referred to as extended modified maximum likelihood (ExMML). When $v(m, \beta)$ does not depend on m, ExMML and ExREML have the same estimating equations and so yield identical estimates. If $v(m, \beta)$ depends on m, the norm for an assay, the values of the estimates are slightly different. This is due to the use of information in the variance function about β and η in the ExMML procedure but not in the WLS step of the ExREML procedure. It is conjectured that ExMML will be slightly more efficient when the observations are normally distributed but slightly less robust to departures from normality.

The ExMML and ExREML procedures are very similar. The difference between them is analogous to the difference between MML and MACL. In MACL, the estimates of μ^U are constrained to equal the corresponding sample means, as in ExREML, as opposed to unrestricted maximization of the modified likelihood function, as in ExMML.

3.3 Bayesian characterization of ExREML

The equivalence of the ExREML estimation method to two calculations based on standard non-informative Bayesian analyses will now be illustrated. As the Bayesian paradigm is not fully empowered at this point, a full description of it is left until it is needed in Chapter 4.

As well as providing some theoretical support, an equivalent Bayesian procedure illustrates the inherent properties of an estimator. Due to the similarity of the ExMML and ExREML procedures, the Bayesian characterization can be considered to apply to both.

3.3.1 Main derivation

This derivation is presented in two segments: the integration step and the optimization step.

3.3.1.1 Integration step

Assume that the prior for the parameters β, θ and μ^U is locally uniform. (In the terminology of Chapter 4 this is an ad-hoc non-informative prior.). This assertion is approximated by $\pi(\beta, \theta, \mu^U) \propto 1$ or equivalently $\pi(\beta, \theta, \eta^U) \propto \prod_{i \in U} m_{\eta_i}(\eta_i, \beta)$. To make Bayesian inferences on θ, the posterior for θ is required (Harville (1977)). This separates the nuisance parameters from the estimation of θ. The posterior distribution for θ is proportional to

$$
\begin{aligned}
\pi(\theta \mid Y, X) &= \int_{\mathcal{R}(\eta^U)} \int_{\mathcal{R}(\beta)} f(Y \mid X, \beta, \theta, \eta) \pi(\beta, \theta, \eta^U) \, d\beta \, d\eta^U \\
&= \int_{\mathcal{R}(\beta)} f(Y^S \mid X, \beta, \theta) \left(\int_{\mathcal{R}(\eta^U)} f(Y^U \mid \eta^U, \beta, \theta) \pi(\beta, \theta, \eta^U) \, d\eta^U \right) d\beta \\
&= \int_{\mathcal{R}(\beta)} f(Y^S \mid X, \beta, \theta) \left(\int_{\mathcal{R}(\mu^U)} f(Y^U \mid \mu^U, \theta) \, d\mu^U \right) d\beta \\
&= \int_{\mathcal{R}(\beta)} f(Y^S \mid X, \beta, \theta) \, d\beta \int_{\mathcal{R}(\mu^U)} f(Y^U \mid \mu^U, \theta) \, d\mu^U.
\end{aligned}
\tag{3.7}
$$

By changing the order of integration and transforming η^U to μ^U in (3.7), β is eliminated from the portion of the likelihood function involving the responses for unknown samples thus partitioning the data. The re-parameterization has this effect because the Jacobian of the transformation is the inverse of the prior on (β, θ, η^U).

In general, it is impossible to evaluate the integrals in (3.7) in the form of explicit algebraic expressions. The approach taken here is to use the following approximations:

1. $v(\mu_i, \theta) \approx v(\hat{\mu}_i, \theta)$ and

2. $m(X_i, \beta) \approx m(X_i, \hat{\beta}) + m_\beta(X_i, \hat{\beta})'(\beta - \hat{\beta})$,

where $\hat{\beta}$ and $\hat{\mu}_i$ are taken to be generalized least squares estimates. These approximations mirror the approximations in Section 3.2. They allow the variance function to be treated as if it is independent of the mean and the mean function as if it is linear in β.

With the first approximation the right-hand integral in (3.7) evaluates to

$$
\begin{aligned}
\pi^U(\theta) &\propto \prod_{i \in U} \int_{\mathcal{R}(\mu_i^U)} v(\hat{\mu}_i, \theta)^{-r_i/2} \exp\left\{ \frac{-(S_i^2 + r_i(\bar{Y}_i - \mu_i)^2)}{2v(\hat{\mu}_i, \theta)} \right\} d\mu_i^U \\
&\propto \prod_{i \in U} \left(v(\hat{\mu}_i, \theta)^{-(r_i-1)/2} \exp\left\{ \frac{-S_i^2}{2v(\hat{\mu}_i, \theta)} \right\} \right),
\end{aligned}
\tag{3.8}
$$

since the normal kernels integrate to constants.

Both approximations are needed to evaluate the left-hand integral in (3.7). Let $\tilde{Y}^S = \bar{Y}^S - m(X, \hat{\beta}) + \hat{J}_\beta \hat{\beta}$ so that by the second approximation

$$
\sum_{i \in S} \frac{r_i(\bar{Y}_i - m(X_i, \beta))^2}{v(\mu_i, \theta)} \approx \sum_{i \in S} \frac{r_i(\tilde{Y}_i - m_\beta(X_i, \beta)'\beta)^2}{v(\hat{\mu}_i, \theta)}
\tag{3.9}
$$

$$
= (\tilde{Y}^S - \hat{J}_\beta \beta)'W(\tilde{Y}^S - \hat{J}_\beta \beta).
\tag{3.10}
$$

At this point, the following well known identity from weighted least squares regression is used:

$$
\begin{aligned}
(\tilde{Y}^S - \hat{J}_\beta \beta)'W(\tilde{Y}^S - \hat{J}_\beta \beta) = \\
\tilde{Y}^{S'}W^{1/2}(I - H)W^{1/2}\tilde{Y}^S + (\beta - \tilde{\beta})'(\hat{J}_\beta'W\hat{J}_\beta)(\beta - \tilde{\beta}),
\end{aligned}
\tag{3.11}
$$

where $\tilde{\beta} = (\hat{J}_\beta'W\hat{J}_\beta)^{-1}\hat{J}_\beta'W\tilde{Y}^S$. Denote the terms from left to right on the rhs of (3.11) by $q(\tilde{Y})$ and $q(\beta)$ respectively. Applying the above approximations the left-hand integral in (3.7) evaluates to

$$
\pi^S(\theta) \propto \int_{\mathcal{R}(\beta)} \left(\prod_{i \in S} v(\hat{\mu}_i, \theta)^{-r_i/2} \right)
$$

$$\cdot \exp\left\{-((\sum_{i\in S} S_i^2/v(\hat{\mu}_i,\theta)) + q(\tilde{Y}) + q(\beta))/2\right\} d\beta$$

$$= \left(\prod_{i\in S} v(\hat{\mu}_i,\theta)^{-r_i/2}\right) \frac{\exp\left\{-((\sum_{i\in S} S_i^2/v(\hat{\mu}_i,\theta)) + q(\tilde{Y}))/2\right\}}{\det(\hat{J}_\beta' W \hat{J}_\beta)^{1/2}}, \quad (3.12)$$

since $\exp\{-q(\beta)/2\}$ is a multivariate normal kernel with covariance matrix $(\hat{J}_\beta' W \hat{J}_\beta)^{-1}$.

To proceed further, the following result is needed. Although Carroll and Ruppert implicitly use this result, they do not provide a proof. Because the result is not obvious, full details of the proof are included herein.

Theorem 3.3.1 *If $\hat{\beta}$ is a GLS estimate of β at θ then*

$$\tilde{Y}^{S'} W^{1/2}(I - H) W^{1/2} \tilde{Y}^S = \sum_{i\in S} e(\hat{\mu}_i,\theta)^2.$$

Proof: Since W is diagonal

$$(I - H) W^{1/2} \tilde{Y}^S = W^{1/2}(\tilde{Y}^S - \hat{J}_\beta \tilde{\beta}).$$

(To verify this expression just substitute $\tilde{\beta}$ with $(\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W \tilde{Y}^S$.) As $I - H$ is idempotent $(I - H)^2 = I - H$ and so

$$\tilde{Y}^{S'} W^{1/2}(I - H) W^{1/2} \tilde{Y}^S = (\tilde{Y}^S - \hat{J}_\beta \tilde{\beta})' W (\tilde{Y}^S - \hat{J}_\beta \tilde{\beta}).$$

Since $\hat{\beta}$ is the GLS estimate corresponding to the weight matrix W it satisfies the following estimating equation:

$$\sum_{i\in S} W_{ii}(\bar{Y}_i - m(X_i,\hat{\beta})) m_\beta(x_i,\hat{\beta}) = \hat{J}_\beta' W (\bar{Y}^S - m(X,\hat{\beta})) = 0.$$

Hence,

$$\begin{aligned}
\tilde{\beta} &= (\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W \tilde{Y}^S \\
&= (\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W (\bar{Y}^S - m(X,\hat{\beta}) + \hat{J}_\beta \hat{\beta}) \\
&= (\hat{J}_\beta' W \hat{J}_\beta)^{-1} [\hat{J}_\beta' W (\bar{Y}^S - m(X,\hat{\beta}))] + \hat{\beta} \\
&= \hat{\beta}
\end{aligned}$$

and so

$$\begin{aligned}
\tilde{Y}^{S'} W^{1/2}(I - H) W^{1/2} \tilde{Y}^S &= (\bar{Y}^S - m(X,\hat{\beta}))' W (\bar{Y}^S - m(X,\hat{\beta})) \\
&= \sum_{i\in S} r_i(\bar{Y} - \hat{\mu}_i)^2/v(\hat{\mu}_i,\theta) \\
&= \sum_{i\in S} e(\hat{\mu}_i,\theta)^2.
\end{aligned}$$

The interpretation of the equality $\tilde{\beta} = \hat{\beta}$, is that the linearized model contains no information about β that is not modelled, or explained, by the nonlinear model. This derivation emulates an iteration of the Gauss Newton algorithm for a nonlinear model (Seber and Wild (1989), Christensen (1991)). In this case the starting value is the nonlinear least squares estimate so it is entirely reasonable that the algorithm remains stationed at this point.

By Theorem 3.3.1 the expression in (3.12) becomes

$$\left(\prod_{i \in S} v(\hat{\mu}_i, \theta)^{-r_i/2} \right) \exp \left\{ -\frac{1}{2} \sum_{i \in S} \left(\frac{S_i^2}{v(\hat{\mu}_i, \theta)} + e(\hat{\mu}_i, \theta)^2 \right) \right\} \det(\hat{J}_\beta' W \hat{J}_\beta)^{-1/2}. \quad (3.13)$$

3.3.1.2 Optimization step

The posterior distribution of θ is approximately proportional to the product, denoted $\pi(\theta)$, of the expressions in (3.8) and (3.13). It is now shown that the mode of $\pi(\theta)$ equals the ExREML estimate of θ.

The maximum value of $\log(\pi(\theta))$ satisfies $\frac{\partial}{\partial \theta} \log(\pi(\theta)) = 0$. Hence

$$- \sum_{i \in U} (r_i - 1) u_\theta(\hat{\mu}_i, \theta) + \sum_{i=1}^{N} \frac{S_i^2}{v(\hat{\mu}_i, \theta)} u_\theta(\hat{\mu}_i, \theta)$$
$$- \sum_{i \in S} r_i u_\theta(\hat{\mu}_i, \theta) + \sum_{i \in S} e(\hat{\mu}_i, \theta)^2 u_\theta(\hat{\mu}_i, \theta) - \frac{\partial}{\partial \theta} (\det(\hat{J}_\beta' W(\hat{\mu}, \theta) \hat{J}_\beta)) = 0.$$
$$(3.14)$$

From Nel (1980), $\frac{\partial}{\partial \theta_k} \det(M(\theta)) = \operatorname{tr}(M(\theta)^{-1} \frac{\partial}{\partial \theta_k} M(\theta))$ where tr is the trace operator and M is a matrix whose entries are functions of θ. Now:

$$\frac{\partial}{\partial \theta_k} (\hat{J}_\beta' W(\hat{\mu}, \theta) \hat{J}_\beta) = -\hat{J}_\beta' W(\hat{\mu}, \theta)^{1/2} U(\theta_k) W(\hat{\mu}, \theta)^{1/2} \hat{J}_\beta,$$

where $U(\theta_k) = \operatorname{diag}\{u_{\theta_k}(\hat{\mu}_i, \theta), i = 1, 2, \ldots, n^s\}$, and

$$\operatorname{tr}(-(\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W^{1/2} U(\theta_k) W^{1/2} \hat{J}_\beta)$$
$$= -\operatorname{tr}(U(\theta_k) W^{1/2} \hat{J}_\beta (\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W^{1/2})$$
$$= -\operatorname{tr}(U(\theta_k) H)$$
$$= -\sum_{i \in S} u_{\theta_k}(\hat{\mu}_i, \theta) H_{ii}.$$

Hence, $\frac{\partial}{\partial \theta} \det(\hat{J}'_\beta W(\hat{\mu}, \theta)\hat{J}_\beta) = -\sum_{i \in S} u_\theta(\hat{\mu}_i, \theta) H_{ii}(\hat{\beta}, \theta)$. Substituting this into (3.14) the ExREML procedure is obtained since

$$
\sum_{i=1}^{n} \frac{S_i^2}{v(\hat{\mu}_i, \theta)} u_\theta(i, \theta) \;+\; \sum_{i \in S} e(\hat{\mu}_i, \theta)^2 u_\theta(\hat{\mu}_i, \theta)
$$

$$
= \sum_{i \in U}(r_i - 1)u_\theta(i, \theta) + \sum_{i \in S} r_i u_\theta(\hat{\mu}_i, \theta) - \sum_{i \in S} u_\theta(\hat{\mu}_i, \theta) H_{ii}(\hat{\beta}, \theta)
$$

$$
= \sum_{i=1}^{n}(r_i - l(\hat{\mu}_i, \theta))u_\theta(\hat{\mu}_i, \theta).
$$

3.3.2 Alternative Bayesian derivation

Now suppose that the variance function has the form $v(m, \theta) = \sigma^2 g(m, \tau)$, where g is a function and $\theta = (\sigma^2, \tau)'$. The variance functions in (2.2) and (2.3) are examples of such variance functions. Let $\pi(\beta, \sigma^2, \tau, \mu^U) \propto 1/\sigma^2$ be the prior distribution. The following derivation shows that if the approximations used in Section 3.3.1 are applied to this model then the mode of the posterior distribution of τ is the ExREML estimate of τ.

For this development, g takes the place of v in the definitions of u and W. The ExREML estimates of σ^2 and τ satisfy

$$
\sigma^2 = \frac{\sum_{i=1}^{n} sst(i, \tau)}{\sum_{i=1}^{n}(r_i - 1) + n^s - p} \tag{3.15}
$$

and

$$
\frac{1}{\sigma^2} \sum_{i=1}^{n} sst(i, \tau) u_\tau(\hat{\mu}_i, \tau) = \sum_{i=1}^{n}(r_i - l(\hat{\mu}_i, \tau)) u_\tau(\hat{\mu}_i, \tau) \tag{3.16}
$$

respectively, where $sst(i, \tau) = (r_i(\bar{Y}_i - \hat{\mu}_i)^2 + S_i^2)/g(\hat{\mu}_i, \tau)$ is the weighted total sum of squares of the responses for the ith sample. Recall that $\hat{\mu}_i = \bar{Y}_i$ if $i \in U$ so the required simplification is assured for unknowns.

The estimate of τ under the alternative Bayesian formulation is constructed by replacing $v(\mu_i, \theta)$ with $\sigma^2 g(\mu_i, \tau)$ in $\pi(\theta)$, multiplying by $1/\sigma^2$ (to cater for the prior) and integrating with respect to σ^2 over the interval $(0, \infty)$. This yields

$$
\pi(\tau) \;\propto\; \prod_{i \in U} g(\hat{\mu}_i, \tau)^{-(r_i - 1)/2} \prod_{i \in S} g(\hat{\mu}_i, \tau)^{-r_i/2} \det(\hat{J}'_\beta W(\hat{\mu}, \tau)\hat{J}_\beta)^{-1/2}
$$

$$
\cdot \int_0^{\infty} \left(\frac{1}{\sigma^2}\right)^{-(\sum_{i=1}^{n}(r_i-1)+n^s-p)/2+1} \exp\left\{-\frac{1}{\sigma^2} \sum_{i=1}^{n} sst(i, \tau)\right\} d\sigma^2.
$$

After recognizing the inverse gamma kernel in σ^2 it follows immediately that

$$\pi(\tau) \quad \propto \quad \prod_{i \in U} g(\hat{\mu}_i, \tau)^{-(r_i-1)/2} \prod_{i \in S} g(\hat{\mu}_i, \tau)^{-r_i/2} \det(\hat{J}'_\beta W(\hat{\mu}, \tau) \hat{J}_\beta)^{-1/2}$$

$$\cdot \left(\sum_{i=1}^{n} sst(i, \tau) \right)^{-(\sum_{i=1}^{n}(r_i-1)+n^s-p)/2} \quad .$$

Using the results developed in Section 3.3.1, the estimating equation for the maximizing value of τ is found to be

$$\left(\sum_{i=1}^{n}(r_i - 1) + n^s - p \right) \sum_{i=1}^{n} \left(\frac{sst(i, \tau) u_\tau(\hat{\mu}_i, \tau)}{\sum_{i=1}^{n} sst(i, \tau)} \right) = \sum_{i=1}^{n}(r_i - l(\hat{\mu}_i, \tau)) u_\tau(\hat{\mu}_i, \tau). \quad (3.17)$$

Equation 3.17 is the ExREML estimating equation for τ after σ^2 is substituted with the rhs of (3.15).

This derivation extends Carroll and Ruppert's characterization of REML (Carroll and Ruppert (1988, pp. 75, 76)). Aside from the additional integration (needed to cater for σ^2) it is equivalent to the derivation in Section 3.3.1. The equivalency occurs because the estimating equation for the approximate mode of the posterior distribution of τ is the equation obtained when the REML estimate of σ^2 (as a function of $\hat{\tau}$) is substituted into the remaining estimating equation. Incidentally, there is more justification for the prior $\pi(\beta, \sigma^2, \tau, \mu^U) \propto 1/\sigma^2$ than for a flat prior when the variance function is of the form $v(m, \theta) = \sigma^2 g(m, \tau)$. In particular, when μ and τ are known this prior is the Jeffreys prior for σ^2. It makes sense to average the posterior over σ^2 because σ^2 is a scale parameter and so does not affect the weights used in WLS. In this sense, σ^2 is a nuisance parameter.

Other instances of Bayesian methods being used to suggest or to characterize estimates of variance functions include Box and Hill (1974), to estimate a homogeneity inducing transformation, and Harville (1974), to estimate variance components.

3.4   Additional topics on ExMML/REML estimation

Some further topics involving ExMML/REML estimation of a variance function are now discussed.

3.4.1 Approximate ExMML/REML estimate

It is interesting to study the extended (MML/REML) estimates when the data are homoscedastic. This is one of the rare instances where the ExMML and ExREML estimates are equal and furthermore reduce to a weighted average of an estimate based on the standards and an estimate based on the unknowns. The common ExMML and ExREML estimate of σ^2 is

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum_{i=1}^{n} S_i^2 + \sum_{i=1}^{n^s} r_i (\bar{Y}_i - \hat{\mu}_i)^2}{\sum_{i=1}^{n} (r_i - 1) + n^s - p} \\
&= \frac{(\bar{r}^u - 1) n^u \hat{\sigma}_{mm}^2 + (\bar{r}^s n^s - p) \hat{\sigma}_{re}^2}{(\bar{r}^u - 1) n^u + \bar{r}^s n^s - p},
\end{aligned}
\tag{3.18}
$$

where $\bar{r}^s = 1/n^s \sum_{i \in S} r_i$, $\bar{r}^u = 1/n^u \sum_{i \in U} r_i$, $\hat{\sigma}_{re}^2$ is the REML estimate of σ^2 and $\hat{\sigma}_{mm}^2$ is the MML estimate of σ^2 based on only the unknowns.

It seems intuitively clear that an improvement in the general case to REML and MML estimates could be made by substituting $\hat{\sigma}_{re}^2$ and $\hat{\sigma}_{mm}^2$ with $\hat{\theta}_{re}$, the REML estimate of θ, and $\hat{\theta}_{mm}$, the MML estimate of θ calculated using just the unknown samples, respectively. This may give a reasonable approximation to the extended estimate. In general, however, there is no guarantee that a weighted average is the most efficient way of combining the two estimates. The extended estimators will generally perform better than the weighted average estimator just described because, unlike the weighted average estimator, the extended estimators simultaneously use the information in the standard and unknown samples. The extended estimators benefit from a likelihood based optimal data specific weighting of the information in the standard and unknown samples while the weighted average estimator does not.

By deleting the $p\hat{\sigma}_{re}^2$ term in the numerator and the p in the denominator of (3.18) and taking $r_i = n^r > 1$ for $i = 1, \ldots, n$, it can be seen that (3.18) becomes the weighted estimate suggested by Carroll and Ruppert (1988). Since the resulting estimate takes no account of the degrees of freedom used to estimate β, the standards are weighted more heavily than in (3.18). For small n^s and $\bar{r}^s$ the standards will thus be overweighted.

3.4.2 Importance of standards and unknowns

To illustrate the relative importance of standard and unknown samples to the estimation of the variance function, the homogeneous linear assay model is considered.

Let $m(\eta_i, \beta) = z_i'\beta$ where the components of z_i are functions of η_i and $v(\mu_i, \theta) = \sigma^2$. For the purpose of simplicity, suppose that the number of replicates is the same for all samples and write $r_i = n^r$. Then the extended estimate of σ^2 is given by

$$\hat{\sigma}^2_{ex} = \frac{\sum_{i=1}^{n} S_i^2 + n^r \sum_{i \in n^s} (\bar{Y}_i^S - z_i'\hat{\beta})^2}{n(n^r - 1) + n^s - p},$$

where $n = n^s + n^u$. As S_i^2 is independent of $\bar{Y}_j$ $\forall i, j$, S_i^2 is independent of $(\bar{Y}_j^S - z_j\hat{\beta})^2$ $\forall i, j$. Therefore, $(n(n^r - 1) + n^s - p)\frac{\hat{\sigma}^2}{\sigma^2}$ has a chi-square distribution with $n(n^r - 1) + n^s - p$ degrees of freedom. Hence, the variance of $\hat{\sigma}^2_{ex}$ is

$$V(\hat{\sigma}^2_{ex}) = \left(\frac{\sigma^2}{n(n^r - 1) + n^s - p}\right)^2 \frac{n(n^r - 1) + n^s - p}{2} 2^2$$

$$= \frac{2\sigma^4}{n(n^r - 1) + n^s - p}.$$

The MML estimate of σ^2 is

$$\hat{\sigma}^2_{mml} = \frac{\sum_{i=1}^{n} S_i^2}{n(n^r - 1)}$$

and

$$V(\hat{\sigma}^2_{mml}) = \frac{2\sigma^4}{n(n^r - 1)}.$$

The REML estimate of σ^2 is

$$\hat{\sigma}^2_{reml} = \frac{\sum_{i \in n^s} S_i^2 + n^r \sum_{i \in n^s} (\bar{Y}_i^S - z_i'\hat{\beta})^2}{n^s n^r - p}$$

and

$$V(\hat{\sigma}^2_{reml}) = \frac{2\sigma^4}{n^s n^r - p}.$$

Therefore, the relative efficiency of $\hat{\sigma}^2_{mml}$ to $\hat{\sigma}^2_{ex}$ is

$$\frac{n(n^r - 1)}{n(n^r - 1) + n^s - p}$$

and the relative efficiency of $\hat{\sigma}^2_{reml}$ to $\hat{\sigma}^2_{ex}$ is

$$\frac{n^s n^r - p}{n(n^r - 1) + n^s - p} = \frac{n^s n^r - p}{n^u(n^r - 1) + n^s n^r - p}.$$

Both $\hat{\sigma}^2_{mml}$ and $\hat{\sigma}^2_{reml}$ are clearly inefficient. When n^u is large relative to n^s, the usual scenario in an assay, $\hat{\sigma}^2_{mml}$ will perform better than $\hat{\sigma}^2_{reml}$. Ignoring the unknowns will in practice lead to a very inefficient estimate.

The above model is a simplification of the standard model for an assay. In this model the values of $\eta^S = X$, η^U and other entities will have a confounding effect. However, one would still expect the results to generalize to some extent. This suggests that unknown samples significantly enhance the quality of the estimates of the parameters of a variance function. The numerical results in Section 3.5 validate this claim.

3.4.3 Some theoretical properties

To summarize the above ideas, the key theoretical properties of the extended estimators (ExMML and ExREML) for estimating a variance function are recapitulated.

- The extended estimators extract more information from the data than either MML or REML. In MML, the known concentrations of the standards are not used and in REML the unknown samples are neglected.

- Even though they use more information than MML and REML, not all of the information about θ contained in the data is used by ExMML or ExREML.

  - By holding the variance function constant while integrating over μ^U, some information about θ is discarded. From another perspective, this means that not all the uncertainty in μ has been considered. It is also the case that the likely disparity between the extended estimators and the true posterior mode will increase as the heterogeneity in the data increases.

  - Since $m(X, \beta)$ is linearized with respect to β, the curvature in the mean function is ignored. Thus a high degree of nonlinearity in this function may have a noticeable effect on the accuracy of the estimators.

- The estimate of θ corresponds to the approximate mode of the posterior distribution of θ under a flat improper prior for (β, θ, μ^U). When the variance function is of the form $v(\mu_i, \theta) = \sigma^2 g(\mu_i, \tau)$ where $\theta = (\sigma^2, \tau)'$, the ExREML estimate of τ is an approximation of the mode of the posterior distribution for τ under the prior $\pi(\beta, \sigma^2, \tau, \mu^U) \propto 1/\sigma^2$. If m is linear in β and v does not depend on m, the extended estimators coincide and are exactly equal to the posterior modes (under the above priors) of θ and τ respectively.

- The ExMML and ExREML estimates are in general not able to be represented as a weighted average of an estimate based exclusively on the standards and an estimate based exclusively on the unknowns. This is due to the simultaneous use of all the data in the calculation of the ExMML and ExREML estimates.

3.5 Simulation

Using simulated data designed to resemble immunoassay responses the behaviour of the MML, REML, ExMML and ExREML estimates were investigated. The design of the assay (X and r) is as for the NMCH data. The concentrations of the unknown samples were taken as $\eta^U = \{100, 100 + 150 \times 1/7, \ldots, 100 + 150 \times 6/7, 250\}$. This constitutes about 1/10th the unknowns that would typically be analyzed in a serum thyroxine radioimmunoassay at NMCH. This reduction was made in order to prevent the estimates being dominated by the unknowns. As differences are easier to notice, an equal number of standards and unknowns is preferable to having excessive differences in the number of standards and unknowns when comparing the estimators discussed in this chapter.

A range of values for θ were considered for the following variance functions:

$$v(m, \theta) = \theta_1 m^{\theta_2} \text{ and } v(m, \theta) = \theta_1 m + \theta_2 m^2.$$

In each case the MML, REML, ExMML and ExREML estimates were evaluated on 1000 sets of randomly generated data. The behaviour of each estimator is summarized by the mean and root mean squared error (RMSE) of the estimates over the 1000 sets of data for each parameter value and variance function. To aid in the interpretation of the results, θ is transformed so that the estimates of its components are less dependent.

To quantify the overall accuracy of the fitted variance function, the total distance between the fitted and true variance functions over the range of the assay was calculated. The measure of distance is the scaled L_2 norm defined for continuous functions h and g on the domain $[a, b]$ as

$$D_{[a,b]}(h, g) = \frac{1}{b - a} \sqrt{\int_a^b (h(u) - g(u))^2 du}. \tag{3.19}$$

This measure is referred to as the scaled L_2 distance (SL$_2$Dist). The smaller the value of SL$_2$Dist, the more accurate the fitted variance function can be said to be. In the simulation, h and g are the fitted and actual variance functions, a and b are $m(0, \beta)$ and $m(400, \beta)$ respectively and the integration is with respect to the mean response.

3.5.1 Power variance function

3.5.1.1 Reparameterization

Under the current parameterization, estimates of θ are highly correlated. This can make numerical computation of the estimates expensive. Furthermore, correlation makes individual study of the estimates difficult and perhaps futile. It is better to use a parameterization for which the parameter estimates are not as highly correlated. The parameterization of θ appropriate for the power variance function is indicated in the following theorem.

Theorem 3.5.1 *Let $\bar{\mu}$ be the geometric mean of μ. That is $\bar{\mu} = (\prod_{i=1}^{n} \mu_i^{r_i})^{1/T}$ where $T = \sum_{i=1}^{n} r_i$. Then if $Y_{ij} \sim N(\mu_i, \theta_1(\mu_i/\bar{\mu})^{\theta_2})$ and μ is known the maximum likelihood estimates of θ_1 and θ_2 are asymptotically independent.*

Proof: Let $\mathcal{L}$ denote the log-likelihood function of θ. It is easy to show that

$$-E\left[\frac{\partial^2}{\partial\theta_1\theta_2}\mathcal{L}\right] = \frac{1}{2}\sum_{i=1}^{n}\frac{r_i}{\theta_1}\log\left(\frac{\mu_i}{\bar{\mu}}\right).$$

Now

$$
\begin{aligned}
\sum_{i=1}^{n} r_i \log\left(\frac{\mu_i}{\bar{\mu}}\right) &= \sum_{i=1}^{n} r_i \left\{\log(\mu_i) - \frac{1}{T}\log\left(\prod_{i=1}^{n}\mu_i^{r_i}\right)\right\} \\
&= \log\left(\prod_{i=1}^{n}\mu_i^{r_i}\right) - \log\left(\prod_{i=1}^{n}\mu_i^{r_i}\right) \\
&= 0.
\end{aligned}
$$

Hence, the off diagonal entries of $I(\theta)$, the Fisher information matrix of θ, are equal to zero. This indicates that the asymptotic correlation of the MLEs, $\hat{\theta}_1$ and $\hat{\theta}_2$, is zero. As this is a regular model, the MLEs of θ are asymptotically normally distributed and therefore independent.

Table 3.1: Relative means and relative root mean squared errors of the estimates of the scale parameter $\tilde{\theta}_1$ in the power variance function.

Parameters		MML		REML		ExMML		ExREML	
		Relative		Relative		Relative		Relative	
θ_1	θ_2	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1000	0	0.929	0.348	0.886	0.403	0.931	0.313	0.931	0.314
5000	0	0.955	0.355	0.906	0.400	0.954	0.323	0.954	0.323
15	0.5	0.946	0.335	0.909	0.398	0.952	0.313	0.950	0.313
75	0.5	0.962	0.354	0.938	0.410	0.966	0.324	0.965	0.324
0.2	1	0.968	0.337	0.921	0.395	0.969	0.304	0.967	0.304
1	1	0.960	0.336	0.947	0.408	0.971	0.313	0.969	0.313
0.01	1.5	0.956	0.347	0.923	0.412	0.962	0.321	0.958	0.320
0.05	1.5	0.953	0.328	0.920	0.397	0.960	0.311	0.957	0.311
0.001	2	0.950	0.340	0.923	0.406	0.956	0.318	0.951	0.317
0.005	2	0.943	0.337	0.910	0.395	0.956	0.302	0.941	0.309

Since all of the estimation methods discussed in this chapter reduce to the maximum likelihood method when μ is known, Theorem 3.5.1 can be applied. To induce independence into the estimates of θ, a reasonable approach is to transform θ so that the above result applies. It is apparent from the equation

$$\theta_1 \mu_i^{\theta_2} = \theta_1 \dot{\mu}^{\theta_2} \left(\frac{\mu_i}{\dot{\mu}} \right)^{\theta_2},$$

that the required transformation is given by $\tilde{\theta}_1 = \theta_1 \dot{\mu}^{\theta_2}$. Intuitively, $\tilde{\theta}_1$ measures the magnitude of the variance while θ_2 measures the level of heteroscedasticity. The re-parameterization has uncoupled these features of the variance function.

When β and μ^U are unknown and must be estimated, numerical evidence suggests that the correlation between $\tilde{\hat{\theta}}_1$ and $\hat{\theta}_2$ is significantly less than it is between $\hat{\theta}_1$ and $\hat{\theta}_2$.

3.5.1.2 Results for $\tilde{\theta}_1$

In Table 3.1 the means and root mean squared errors of the estimates of $\tilde{\theta}_1$ are divided by the value of $\tilde{\theta}_1$. The relative values of $\tilde{\theta}_1$ are easy to display because they are always close to unit value whereas the raw values vary extensively.

The ExMML and ExREML estimators are clearly the best in terms of mean

Table 3.2: Means and root mean squared errors of the estimates of the power parameter θ_2 in the power variance function.

Parameters		MML		REML		ExMML		ExREML	
θ_1	θ_2	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
1000	0	-0.109	0.762	-0.041	0.813	-0.103	0.710	-0.106	0.729
5000	0	-0.122	0.753	-0.048	0.782	-0.111	0.699	-0.114	0.718
15	0.5	0.403	0.745	0.463	0.794	0.388	0.694	0.399	0.711
75	0.5	0.410	0.716	0.449	0.735	0.399	0.657	0.410	0.674
0.2	1	0.895	0.768	0.952	0.786	0.874	0.713	0.897	0.729
1	1	0.919	0.693	0.958	0.737	0.889	0.645	0.912	0.658
0.01	1.5	1.361	0.769	1.427	0.806	1.347	0.725	1.383	0.735
0.05	1.5	1.405	0.723	1.462	0.761	1.370	0.684	1.405	0.696
0.001	2	1.903	0.746	1.965	0.787	1.856	0.707	1.904	0.713
0.005	2	1.906	0.721	1.965	0.768	1.855	0.679	1.903	0.693

squared error and are themselves almost identical on all accounts. The estimators yield estimates of $\tilde{\theta}_1$ that are negatively biased with ExMML being the least so. It is interesting to note that the relative bias of all the estimates was least when $\theta_2 = 1$. In this case, the variance function is the variance of the Poisson distribution.

3.5.1.3 Results for θ_2

The behaviour of $\hat{\theta}_2$ varies significantly between the estimators. The extended estimators have the smallest mean squared error with ExMML performing slightly better than ExREML. It is interesting to note that the higher an estimate ranked in terms of mean squared error the lower it ranked in terms of bias. This emphasizes that bias by itself is a poor criterion for assessing the quality of an estimator.

It can also be seen that the estimates of θ_2, like $\tilde{\theta}_1$ estimates, were consistently negatively biased.[4]

3.5.1.4 Relationship to Raab's findings

At first sight these results appear to be in conflict with those of Raab (1981). Raab's results indicated that the MML estimates of θ_2 should be more or less ("to within

---

[4]The point of REML is to reduce bias relative to maximum likelihood estimates. This appears not to be the case for this problem. However, it must be remembered that the variance function is the entity for which bias should be assessed and not the individual parameters.

working precision") unbiased. However, replications of Raab's simulation have validated the results. Several variants of Raab's simulation have also been carried out. The following properties of the class of estimators of θ comprising MML, ExMML and ExREML have been observed:

- The greater the dispersion of the mean responses (the components of μ) the less biased the parameter estimates are. In Raab's paper the mean responses are uniformly distributed over a wide range of values, while for the simulation reported in this thesis the concentrations are uniformly distributed. The non-linearity of the relationship between η and μ caused the induced distribution of the mean responses to be far from uniform and not even symmetric. This accounts for the lack of bias reported in Raab (1981) and the bias observed in this simulation.

- The higher the degree of replication the lower the bias of $\hat{\theta}$. This is understandable given that the effect of estimating the mean responses dissipates as the number of replicates increases. It was, however, interesting to note that even when μ is known, the estimates of θ_1 still exhibit severe bias.

- The bias of $\hat{\theta}$ decreases as the number of unknowns increase. This is an indication that this class of estimates is consistent as $n^u \to \infty$. This result was expected given the claims about consistency in Raab (1981).

3.5.1.5   Results for SL$_2$Dist

In terms of scaled L$_2$ distance, ExMML is the best estimator of the power variance function. Both extended estimators have clearly performed better than MML and REML. It is surprising that the SL$_2$Dist scores for MML and REML are very similar given the disparity between their estimates of $\tilde{\theta}_1$ and θ_2.

3.5.2   Quadratic variance function

For the case in which the variance function is assumed to be the quadratic function of the mean given in (2.3), the Fisher information matrix for θ does not suggest a parameterization under which the MLEs are asymptotically independent. However,

Table 3.3: Mean scaled L_2 distance of the fitted power variance functions.

Parameters		MML	REML	ExMML	ExREML
θ_1	θ_2	SL$_2$Dist	SL$_2$Dist	SL$_2$Dist	SL$_2$Dist
1000	0	4.0	4.2	3.6	3.7
5000	0	20.0	21.0	18.1	18.6
15	0.5	5.6	6.0	5.3	5.4
75	0.5	28.0	29.7	26.0	26.6
0.2	1	8.4	8.9	8.2	8.4
1	1	41.3	44.0	39.1	40.4
0.01	1.5	53.5	54.8	49.7	51.7
0.05	1.5	254.7	259.1	238.1	247.7
0.001	2	678.1	673.8	619.6	647.4
0.005	2	3224.3	3256.8	3010.1	3083.6

the correlation in the estimates can be reduced by replacing μ_i^2 with

$$\tilde{\mu}_i = \mu_i^2 - \frac{\sum_{i=1}^n \mu_i^3}{\sum_{i=1}^n \mu_i^2}\mu_i.$$

This appears to be the case because the term

$$\frac{\sum_{i=1}^n \mu_i^3}{\sum_{i=1}^n \mu_i^2}\mu$$

is the projection of the vector $(\mu_1^2, \mu_2^2, \ldots, \mu_n^2)'$ onto μ, hence μ and $\tilde{\mu}$ are orthogonal.

The variance function can thus be written in the form

$$v(\mu_i, \tilde{\theta}) = \tilde{\theta}_1 \mu_i + \theta_2 \tilde{\mu}_i,$$

where

$$\tilde{\theta}_1 = \theta_1 + \frac{\sum_{i=1}^n \mu_i^3}{\sum_{i=1}^n \mu_i^2}\theta_2.$$

Unlike the power variance function the transformed parameter, $\tilde{\theta}_1$, does not seem to have an intuitive interpretation.

To ensure that the estimated variance function is non-negative over the range of concentrations that have been fitted, θ is constrained so that $\theta_1 \geq 0$ and $\theta_2 \geq -\theta_1/\max\{\bar{Y}_i\}$. This ensures that the variance function fits are non-negative over the range $[0, \max\{\bar{Y}_i\}]$.

Table 3.4: Relative means and relative root mean squared errors of the estimates of $\tilde{\theta}_1$, the parameter for the linear term in the quadratic variance function.

Parameters		MML		REML		ExMML		ExREML	
		Relative		Relative		Relative		Relative	
θ_1	$(10^2)\theta_2$	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
4	0	1.015	0.441	1.010	0.536	1.004	0.417	1.016	0.428
20	0	0.978	0.398	0.960	0.480	0.964	0.369	0.972	0.379
1	0.1	0.891	0.382	0.881	0.458	0.879	0.365	0.890	0.363
5	0.1	0.914	0.397	0.907	0.475	0.910	0.388	0.920	0.391
0.5	0.25	0.867	0.401	0.859	0.459	0.862	0.382	0.867	0.380
2.5	0.25	0.873	0.395	0.860	0.463	0.863	0.381	0.869	0.379
0.3	0.5	0.880	0.388	0.863	0.467	0.866	0.374	0.870	0.371
1.5	0.5	0.849	0.384	0.844	0.452	0.844	0.371	0.846	0.370
0.2	1	0.845	0.395	0.824	0.457	0.843	0.378	0.841	0.378
1	1	0.845	0.399	0.834	0.458	0.843	0.385	0.840	0.380

3.5.2.1   Results for $\tilde{\theta}_1$

The relative values of the estimates of $\tilde{\theta}_1$ and their root mean squared errors are reported in Table 3.4. In this case the original estimates are difficult to display because at different values of the parameters the estimates of $\tilde{\theta}_1$ have vastly different magnitudes.

The extended estimators have the smallest mean squared error. The mean squared error of the ExREML estimator appears to be smaller than the mean squared error of the ExMML estimator when $\theta_2 \neq 0$ (the only instance in Table 3.4 in which this is not the case is when $\theta_1 = 5$ and $\theta_2 = 0.001$). When $\theta_2 = 0$, the variance function is a special case of the power variance function, hence it is of no surprise that ExMML performs better in this instance. Whenever the quadratic variance function is such that $\theta_1 \neq 0$ and $\theta_2 \neq 0$, it appears that ExREML is slightly better than ExMML at estimating $\tilde{\theta}_1$.

3.5.2.2   Results for θ_2

There is a stong indication that ExREML outperforms ExMML in terms of estimating θ_2. The extended estimators outperform MML and REML but not by as much as for the power variance function (compare Table 3.2 to Table 3.5 above).

Table 3.5: Means and root mean squared errors of the estimates of θ_2 (in units of 10^{-2}), the coefficient of the quadratic term in the quadratic variance function.

Parameters		MML		REML		ExMML		ExREML	
θ_1	$(10^2)\theta_2$	Mean	RMSE	Mean	RMSE	Mean	RMSE	Mean	RMSE
4	0	0.002	0.027	0.001	0.027	0.001	0.026	0.002	0.026
20	0	-0.004	0.126	-0.008	0.132	-0.010	0.116	-0.005	0.120
1	0.1	0.072	0.067	0.073	0.070	0.070	0.065	0.073	0.063
5	0.1	0.073	0.093	0.073	0.099	0.072	0.091	0.076	0.091
0.5	0.25	0.173	0.158	0.178	0.165	0.172	0.154	0.176	0.151
2.5	0.25	0.171	0.171	0.176	0.178	0.169	0.168	0.174	0.164
0.3	0.5	0.351	0.309	0.355	0.328	0.341	0.306	0.351	0.297
1.5	0.5	0.321	0.332	0.332	0.350	0.316	0.329	0.324	0.324
0.2	1	0.638	0.650	0.646	0.672	0.634	0.634	0.642	0.626
1	1	0.653	0.640	0.678	0.650	0.646	0.631	0.656	0.619

Once again it is observed that the estimator of θ_2 which is the worst performer in the squared error sense is the least biased.

3.5.2.3 Results for SL$_2$Dist

The mean scaled L$_2$ distances cohere with the data in Tables 3.4 and 3.5. It is clear that ExREML is the best estimator at all values of the parameters except when $\theta_2 = 0$ and as expected ExMML performs slightly better. In terms of estimating θ_2, the extended estimators significantly outperform the MML and REML estimators. The extent to which MML outperforms REML is greater for this variance function than for the power variance function.

3.5.3 Summary of the findings from the simulation

The results of the simulation conclusively illustrate that the extended estimators are better than either MML or REML. However, it is not clear which extended estimator is the best. ExMML performs best when the variance function is a power of the mean response while ExREML produced the best estimates for the quadratic variance function. There is no compelling reason as to why the relative performance of ExMML and ExREML is sensitive to the form of the variance function.

Two reasons for rating ExREML ahead of ExMML are apparent. Firstly, the

Table 3.6: Mean scaled L_2 distance of the fitted quadratic variance functions.

Parameters		MML	REML	ExMML	ExREML
θ_1	$(10^2)\theta_2$	SL_2Dist	SL_2Dist	SL_2Dist	SL_2Dist
4	0	172	174	166	170
20	0	793	809	747	766
1	0.1	506	557	486	479
5	0.1	687	742	678	680
0.5	0.25	1247	1322	1186	1172
2.5	0.25	1300	1397	1258	1243
0.3	0.5	2358	2610	2284	2241
1.5	0.5	2471	2640	2394	2374
0.2	1	4900	5237	4714	4697
1	1	4844	5260	4738	4680

involvement of both β and θ in the leverage function for the standards can occasionally make full optimization of the modified likelihood function computationally unstable. Secondly, the use of weighted least squares estimates of β and η^U in the ExREML procedure makes it slightly less sensitive than ExMML to departures from normality.

When the number of standards is fixed, ExMML and ExREML clearly have the same asymptotic properties as MML. Thus, under the same conditions assumed by Raab (1981), ExMML and ExREML estimates will be consistent. Additional simulations revealed that as the number of unknowns increase, MML estimates become virtually indistinguishable from the extended estimates. The ratio of unknowns to standards is often large (ten to one or more) in practice. For such assays MML estimates will be virtually identical to the extended estimates and so continued use of MML is justified.

The disparity in performance between the MML and REML estimators illustrates the importance of the unknown samples to the estimation of the variance function and the analysis in general. With just eight unknowns the payoff from fitting the mean function and using the residuals to enhance the estimate of θ is already outweighed by the improvement obtained from utilizing the unknowns. As the degree of replication increases, MML improves relative to REML. This is due to the information contained in the replicated responses for each sample outweighing the information contained in the residuals from the fitted mean function. However,

if the number of standards increases, the leverage values decrease and so REML improves with respect to MML.

Chapter 4

Bayesian analysis of an assay

In this chapter the analysis of an assay is considered from a Bayesian viewpoint. Results are often presented in a general form with the standard immunoassay model being used as an example. The emphasis is on the evaluation of inferences about the assay parameters. The evaluation of the quality of the assay is left to Chapters 5,6 and 7.

4.1 Bayesian model for an assay

4.1.1 The likelihood function

From Section 2.1 the general form of the assay model is

$$Y_{ij} \stackrel{\text{ind}}{\sim} f(Y_{ij} \mid \eta_i, \beta, \theta), \text{ for } j = i, \ldots, r_i, \ i = 1, \ldots, n,$$

where

$$E[Y_{ij} \mid \eta_i] = \mu_i = m(\eta_i, \beta)$$

and

$$V[Y_{ij} \mid \mu_i] = \sigma_i^2 = v(\mu_i, \theta).$$

In the standard model for an assay $f(Y_{ij} \mid \eta_i, \beta, \theta)$ is a normal density.

The likelihood function for the r_i independent response measurements made on the ith sample is

$$f(Y_i \mid \eta_i, \beta, \theta) = \prod_{j=1}^{r_i} f(Y_{ij} \mid \eta_i, \beta, \theta).$$

49

The likelihood function for the complete set of assay responses is the product of the likelihood functions for the responses from each sample. This is given by

$$f(Y \mid X, \eta^U, \beta, \theta) = \prod_{i \in S} f(Y_i \mid X_i, \beta, \theta) \prod_{i \in U} f(Y_i \mid \eta_i^U, \beta, \theta). \tag{4.1}$$

4.1.2 The prior distribution

The probability distribution of the assay responses, also known as the likelihood function of the responses, has been described. To complete the specification of the Bayesian model for an assay, prior distributions for the assay parameters β, θ and η^U need to be specified.

4.1.2.1 Structure of the prior

The most general form for the prior distribution will be denoted by $\pi(\beta, \theta, \eta^U)$. However, in practice it is likely that a priori some parameters will be considered to be independent. For instance, it is unlikely that there would be a relationship between the unknown concentrations and the model parameters for a given assay. Thus we regard (β, θ) a priori independent from η^U and write

$$\pi(\beta, \theta, \eta^U) = \pi_m(\beta, \theta)\pi_u(\eta^U), \tag{4.2}$$

where π_m and π_u denote the respective prior probability densities.

A priori independence between β and θ is more difficult to justify. This is because the variance function is dependent on the regression function. In a sense, β is functioning both as a scale parameter and as a location parameter. This invalidates the usual justification that location and scale are a priori independent. The structure of the prior should reflect what is known about β and θ. Some cases for independence can certainly be imagined but because of the complexity of the assay model, there is a temptation to specify independent priors for simplicity irrespective of prior beliefs.

In practice a practitioner is typically blinded to the particulars of the subjects being tested. Without any specific information about the subjects, unknown concentrations can be regarded as independently and identically distributed, i.e. it is reasonable to expect that

$$\pi_u(\eta^U) = \prod_{i \in U} \pi_0(\eta_i), \tag{4.3}$$

where π_0 is a univariate density. The unknown concentrations can be thought of as a random sample from the distribution of the concentrations of the subjects eligible to be tested.

In many applications, the population of eligible subjects will be partitioned into "healthy" and "unhealthy" strata. The distribution of concentrations is likely to be different within each stratum. An appropriate form for the prior distribution of η_i is then

$$\pi_0(\eta_i) = p\pi_0^h(\eta_i) + (1-p)\pi_0^u(\eta_i),$$

where $\pi_0^h(\eta_i)$ and $\pi_0^u(\eta_i)$ are the distributions of the concentration in the healthy and unhealthy populations respectively and $p \in [0,1]$ is the probability that a randomly selected subject is healthy. An extreme but perhaps common situation is when $\eta = 0$ with probability 1 when a subject is healthy and $\eta > 0$ with probability 1 when a subject is unhealthy. In this case ·

$$\pi_0(\eta_i) = pI(\eta_i = 0) + (1-p)\pi_0^u(\eta_i)I(\eta_i > 0). \tag{4.4}$$

4.1.3 Informative priors

The specification of an informative prior can be a complex and time consuming process. The difficulty lies in the conversion of subjective knowledge from past experience or otherwise into probability distributions for the model parameters. The complexity of the assay model amplifies this problem. .

Methods for the elicitation of prior knowledge are difficult to discuss in any generality and the exact specification of a prior distribution based on this knowledge is highly situation oriented. Because a study of such methods is not a topic of this thesis, only some suggestions of how one might proceed are given. It should be noted that the mathematics associated with these suggestions has therefore not been totally substantiated.

4.1.3.1 Specification of $\pi_m(\beta, \theta)$

Direct elicitation of $\pi_m(\beta, \theta)$ is in theory possible. However, as β and θ have rather complicated interpretations, which may be beyond many practitioners, an easier and perhaps practical method of ascertaining the prior would be to use predictions about the responses since these are observable quantities.

For example, suppose

$$\pi_m(\beta, \theta) = N\left(\left(\begin{array}{c} \beta_0 \\ \theta_0 \end{array}\right), \Sigma_0\right).$$

If the expected response was specified at p distinct concentrations, then β_0 could be calculated as the value of β for which the mean function interpolates these responses. Similarly, by asserting the variance of the response at q distinct mean responses, θ_0 could be calculated as the value of θ for which the variance function interpolates these assertions. A convenient way of specifying Σ_0 is not so straight forward. It may be possible to obtain Σ_0 from confidence intervals for the mean and variance of the response at different concentrations. However, these predictions may themselves be too difficult to specify. The reader is referred to Bedrick, Christensen and Johnson (1996) and Bedrick, Christensen and Johnson (1997) for further ideas on inducing the prior distributions for the parameters from prior specifications of the response surface in regression problems.

In Section 4.5 it is shown how information from past assays can be used to determine a hyper-prior for a Bayesian analysis of an assay. A short cut to this method is to guess the hyper-prior. One strategy is to deliberately use a distribution that is less informative than the actual state of prior knowledge to ensure that the analysis will err on the side of conservatism (in the sense that the data will have a greater bearing on the analysis than might otherwise be the case).

4.1.3.2 Specification of $\pi_0(\eta_i)$

The task of specifying $\pi_0(\eta_i)$ is much less arduous. Firstly, the density is univariate and so dependence with other parameters does not have to be considered. Secondly, the concentration in an unknown sample is a well defined easily understood quantity. Furthermore, the prevalent concentration throughout the population of subjects eligible to be tested might be so well understood that this prior can be specified without recourse to subjective input. A convenient means of specifying $\pi_0(\eta_i)$ would be to match properties of the distribution such as the mean, variance and quantiles to an appropriate functional form.

4.1.4 Non-informative priors

Non-informative Bayesian analysis can be viewed as a method of obtaining a sensible and in a sense objective answer from statistical analyses with minimal effort. The problem of having to specify the prior is avoided through the specification of a prior that requires no prior knowledge. For these reasons and others, (see Berger (1985) and Yang and Berger (1996)) non-informative Bayesian analysis is a very powerful method in statistical analysis. The literature on non-informative priors has grown enormously over recent years (Yang and Berger (1996)). For discussion of the various approaches for developing non-informative priors, refer to the above references and also see Bernardo and Smith (1994). The non-informative prior approach is now developed for the assay model.

4.1.4.1 Jeffrey's prior

The Jeffrey's method of prior development is perhaps the most well known and widely used generic method of determining a non-informative prior. The Jeffrey's prior for the assay parameters is given by $\pi(\beta, \theta, \eta^U) = \sqrt{\det(I(\beta, \theta, \eta^U))}$, where $I(\beta, \theta, \eta^U)$ is the Fisher information matrix.

The calculation of Jeffrey's prior and variations of it are illustrated using the standard model for an assay. The log-likelihood function is

$$\mathcal{L}(\beta, \theta, \eta^U \mid Y, X) \propto -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{r_i} \left(\log(v(\mu_i, \theta)) + \frac{(Y_{ij} - \mu_i)^2}{v(\mu_i, \theta)} \right),$$

where

$$\mu_i = \begin{cases} m(X_i, \beta) & \text{if } i \in S \\ m(\eta_i, \beta) & \text{if } i \in U \end{cases}$$

It is convenient to parameterize unknowns at the response level; that is, substitute $m(\eta_i, \beta)$ for μ_i^U if $i \in U$. As Jeffrey's prior is invariant under transformations of the parameter space this substitution has no effect on the analysis. Recall that μ^U denotes the mean response corresponding to the concentrations η^U.

The first derivatives of $\mathcal{L}$ with respect to β, θ and μ^U are

$$\frac{\partial}{\partial \beta} \mathcal{L} = -\frac{1}{2} \sum_{i \in S} \sum_{j=1}^{r_i} \left\{ \left(1 - \frac{(Y_{ij} - \mu_i)^2}{v(\mu_i, \theta)} \right) v_{\mu_i} i - 2(Y_{ij} - \mu_i) \right\} \frac{m_{\beta} i}{v(\mu_i, \theta)},$$

$$\frac{\partial}{\partial \theta} \mathcal{L} = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{r_i} \left(1 - \frac{(Y_{ij} - \mu_i)^2}{v(\mu_i, \theta)} \right) \frac{v_\theta i}{v(\mu_i, \theta)}$$

and

$$\frac{\partial}{\partial \mu_i} \mathcal{L} = -\frac{1}{2} \left\{ \left(1 - \frac{(Y_{ij} - \mu_i)^2}{v(\mu_i, \theta)} \right) v_{\mu_i} i - 2(Y_{ij} - \mu_i) \right\} \frac{1}{v(\mu_i, \theta)}, \ i \in U,$$

where $m_\beta i = m_\beta(\eta_i, \beta)$, $v_{\mu_i} i = v_{\mu_i}(\mu_i, \theta)$, $v_\theta i = v_\theta(\mu_i, \theta)$ and $f_x \equiv \frac{\partial}{\partial x} f$.

The calculation of $I(\beta, \theta, \mu^U)$, the information matrix of the assay parameters, is obtained as:

$$-E \left[\frac{\partial^2}{\partial \beta \beta'} \mathcal{L} \right] = \sum_{i \in S} \left\{ 1 + \frac{(v_{\mu_i} i)^2}{2v(\mu_i, \theta)} \right\} \frac{r_i}{v(\mu_i, \theta)} m_\beta i \, m_\beta i',$$

$$-E \left[\frac{\partial^2}{\partial \theta \theta'} \mathcal{L} \right] = \frac{1}{2} \sum_{i=1}^{n} \frac{r_i}{v(\mu_i, \theta)^2} v_\theta i \, v_\theta i',$$

$$-E \left[\frac{\partial^2}{\partial \mu_i^2} \mathcal{L} \right] = \left\{ 1 + \frac{(v_{\mu_i} i)^2}{2v(\mu_i, \theta)} \right\} \frac{r_i}{v(\mu_i, \theta)}, \ i \in U,$$

$$-E \left[\frac{\partial^2}{\partial \beta \theta'} \mathcal{L} \right] = \frac{1}{2} \sum_{i \in S} \frac{r_i v_\mu i}{v(\mu_i, \theta)^2} m_\beta i \, v_\theta i',$$

$$-E \left[\frac{\partial^2}{\partial \beta \mu_i} \mathcal{L} \right] = 0, \ i \in U,$$

$$-E \left[\frac{\partial^2}{\partial \theta \mu_i} \mathcal{L} \right] = \frac{1}{2} \frac{r_i v_\mu i}{v(\mu_i, \theta)^2} v_\theta i, \ i \in U,$$

$$-E \left[\frac{\partial^2}{\partial \mu_i \mu_j} \mathcal{L} \right] = 0, \ i, j \in U, \ i \neq j.$$

Jeffrey's prior is proportional to the square root of the determinant of the matrix with these components and gives a rather complicated expression especially when n^u is large.

If the parameters are a priori independent, the product of Jeffrey's priors for the independent components generally yields a prior with better properties than the Jeffrey's prior itself (Berger (1985)). Location-scale models are one instance where this is the case. If the model parameters and the unknown concentration parameters are a priori independent, then if it is used Jeffrey's method should be applied separately to (β, θ) and η^U. The determinant that then needs to be evaluated involves a matrix of dimension $(p + q)$ rather than dimension $(p + q + n^u)$. If a priori independence

between β and θ is also asserted, this amended Jeffrey's approach would stipulate that separate Jeffrey's priors should be calculated for each parameter.

Even when the parameters are not regarded independently there is some evidence in the literature to suggest that a better analysis can still be obtained by using independent priors for them. From Yang and Berger (1996), "In multi-parameter situations Jeffrey's prior frequently induces dependence between the parameters. This dependence may lead to poor performance, even inconsistency".

Jeffrey's prior for β can be expressed in the form

$$\pi(\beta) = \det(J_\beta' W(\beta) J_\beta)^{1/2},$$

where $W(\beta)$ is a diagonal matrix with

$$\left\{ 1 + \frac{v_{\mu_i}^2}{2v(\mu_i, \theta)} \right\} \frac{r_i}{v(\mu_i, \theta)}, \quad i \in S,$$

the ith diagonal element. Similarly, Jeffrey's prior for θ can be expressed in the form

$$\pi(\theta) = \det(V_\theta' W(\theta) V_\theta)^{1/2},$$

where V_θ contains $v_\theta(\mu_i, \theta)$ as its ith row and $W(\theta)$ is a diagonal matrix with

$$\frac{r_i}{v(\mu_i, \theta)}$$

being the ith diagonal element. The Jeffrey's prior for μ^U is

$$\pi(\mu^U) = \prod_{i \in U} \det \left(\left\{ 1 + \frac{v_{\mu_i}^2}{2v(\mu_i, \theta)} \right\} \frac{r_i}{v(\mu_i, \theta)} \right)^{1/2}.$$

This can also be written as a quadratic form.

Notice that these non-informative priors do not maintain a specification of prior independence. This is typical of Jeffrey's based non-informative priors.

4.1.4.2 Uniform prior

A uniform prior is any distribution, proper or improper, that is proportional to a constant. The major drawback of uniform or flat priors is that they are not invariant under transformations of the parameter space. Two natural uniform priors for η^U are $\pi_u(\eta^U) \propto 1$ and the distribution for η^U induced from $\pi_u(\mu^U) \propto 1$. These priors are clearly not equivalent.

There are preferences for both priors. An argument in favour of $\pi_u(\mu^U) \propto 1$ is that μ^U is a vector of location parameters and in the absence of other parameters the natural non-informative prior for a location parameter is the uniform prior. An argument against this prior is that it induces prior dependence between β and η and so contradicts a specification of prior independence between these parameters. A second argument in favour of $\pi_u(\mu^U) \propto 1$ is that for some simple models it gives rise to a proper posterior distribution for η^U whereas $\pi_u(\eta^U) \propto 1$ yields an improper posterior (see Hoadley (1970)). On this premise $\pi_u(\mu^U) \propto 1$ is favoured.

4.1.4.3 Reference prior

Reference priors as defined in Bernardo (1979) and Bernardo and Smith (1994) are much more difficult to obtain than Jeffrey's priors. Such reference priors have not yet been calculated for models anywhere nearly as complicated as the assay model. The task of calculating a reference prior for the assay model may in fact be impossible.

4.1.4.4 Other non-informative priors

There often exist various non-informative priors particular to m and v that have good properties but yet do not arise from a general theory. There is often no justification for the use of such priors other than evidence of good empirical results in the literature. Yang and Berger (1996) call such priors "ad-hoc" priors.

A standard non-informative prior for a model with the power variance function (2.2) is $\pi_m(\beta, \theta) \propto 1/\theta_1$. This is a natural non-informative prior for the linear regression model with variance θ_1. Since good results are obtained when m is linear in β and v is a constant there is reason to believe that this prior will give good results in more general cases. In this way $\pi_m(\beta, \theta) \propto 1/\theta_1$ is an ad-hoc prior. The prior $\pi_m(\beta, \theta) \propto 1/\tilde{\theta}_1$, where $\tilde{\theta}_1 = \theta_1 \mu^{\theta_2}$, is an alternative generalization of the prior when the variance function is a power of the mean response as opposed to a constant.

4.1.5 A generic prior

In practice it is likely that prior knowledge of η^U will outweigh prior knowledge of β and θ. The purpose of the assay is to infer the value of η^U, the entity of fundamental interest. It is desirable that all relevant information about η^U is included in the

analysis. On the other hand, β and θ are service or nuisance parameters. It is equally desirable that the priors for these parameters do not have an undue influence on the analysis. A non-informative prior for $\pi(\beta, \theta)$ and an informative prior for $\pi_0(\eta_i^U)$ would therefore be a suitable choice for a generic prior.

4.2 Bayesian Analysis

In this section the theoretical side of a Bayesian analysis of an assay is outlined.

4.2.1 Posterior distribution

The joint posterior distribution of the assay parameters is given by

$$\pi(\beta, \theta, \eta^U \mid Y, X) = \frac{f(Y \mid X, \eta^U, \beta, \theta)\pi(\beta, \theta, \eta^U)}{f(Y \mid X)},$$

where

$$f(Y \mid X) = \int_{\mathcal{R}(\eta^U)} \int_{\mathcal{R}(\beta)} \int_{\mathcal{R}(\theta)} f(Y \mid X, \eta^U, \beta, \theta)\pi(\beta, \theta, \eta^U) \, d\theta \, d\beta \, d\eta^U$$

and

$$f(Y \mid X, \eta^U, \beta, \theta) = f(Y^S \mid X, \beta, \theta)f(Y^U \mid \eta^U, \beta, \theta).$$

The above integral and those subsequently presented in this thesis may be generalized to a Riemann-Stieltjes integral to facilitate discrete or mixed distributions.

All inferences concerning the assay are based on $\pi(\beta, \theta, \eta^U \mid Y, X)$. In cases in which an inference is to be made about a subset of the parameters, it will generally be necessary to calculate the joint posterior distribution of just those parameters. In particular, this distribution is required for the calculation of the posterior mode or a credibility region for a single parameter. The posterior mean can, on the other hand, be calculated directly from $\pi(\beta, \theta, \eta^U \mid Y, X)$.

4.2.2 Inferences about the assay parameters

At the completion of the assay the joint posterior distribution of β and θ, given by

$$\pi(\beta, \theta \mid Y, X) = \int_{\mathcal{R}(\eta^U)} \pi(\beta, \theta, \eta^U \mid Y, X) \, d\eta^U,$$

contains all the information about the assay model. This distribution is referred to as the "model fit" of the assay since the distribution of parameters are of interest as opposed to a single estimate (the quantity usually referred to as the model fit or fitted model).

Inferences particular to β are based upon

$$\pi(\beta \mid Y, X) = \int_{\mathcal{R}(\theta)} \pi(\beta, \theta, \mid Y, X) \, d\theta.$$

Inferences involving only θ (for example, the estimation of the variance function) are based upon

$$\pi(\theta \mid Y, X) = \int_{\mathcal{R}(\beta)} \pi(\beta, \theta, \mid Y, X) \, d\beta.$$

Further marginalization is required for inferences involving only a component of β or θ. For example, the posterior variance or posterior standard error of a single component of θ would be calculated from the marginal posterior distribution of that component.

4.2.3   Inferences about unknown concentrations

For an inference about one unknown concentration, say η_i, the pertinent posterior distribution is

$$\pi(\eta_i \mid Y, X) = \int_{\mathcal{R}(\eta_{-i}^U)} \int_{\mathcal{R}(\beta)} \int_{\mathcal{R}(\theta)} \pi(\beta, \theta, \eta^U) \, d\theta \, d\beta \, d\eta_{-i}^U,$$

where $\eta_{-i}^U = \{\eta_j, j \neq i\}$. From the Bayesian viewpoint, this is the distribution upon which all inferences and decisions specific to η_i should be based. Given all the data that has been observed, the state of knowledge about η_i is exactly described by $\pi(\eta_i \mid Y, X)$.

In general,

$$\pi(\eta^U \mid Y, X) \neq \prod_{i \in U} \pi(\eta_i \mid Y, X)$$

because the information in each unknown sample contains information about β and θ. Even if the unknown concentration parameters are a priori independent, the common dependence of them on β and θ makes them a posteriori marginally dependent.

4.2.4 Predictive inference

At the completion of the assay, the prior and experimental information are combined in the form of the posterior distribution for the model parameters. The predictive distribution converts this information to a distribution describing the uncertainty in the value of the next response measurement at a given concentration. In Chapters 5, 6 and 7 this predictive distribution plays an important role in many calculations.

The predictive distribution of a new response measurement y at some given concentration x is

$$p(y \mid x, Y, X) = \int_{\mathcal{R}(\beta)} \int_{\mathcal{R}(\theta)} f(y \mid x, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, d\theta \, d\beta. \tag{4.5}$$

This expression follows from the theorem of total probability and the fact that y given (x, β, θ, Y, X) does not depend on (Y, X), hence $f(y \mid x, \beta, \theta, Y, X) = f(y \mid x, \beta, \theta)$. In (4.5), y can be a singleton response or a vector of replicated response measurements

A nice feature about the predictive distribution is that it translates uncertainty in the model fit to uncertainty about the quantity being observed. This provides a mechanism for measuring the error in the fitted model. The procedure is firstly to remove an observation from the calculation of the predictive distribution. Then the revised predictive distribution is used to predict the value of the deleted observation. The prediction is then compared to the actual value of the observation and the error in the prediction is calculated. These predictive distributions, known as "hold-out" predictive distributions, form the basis of Bayesian cross validation (see Schlüter (1996) or for a basic reference on predictive inference Geisser (1993)).

4.3 Analytic derivation: linear mean function with known weights case

In this section the calculations of the posterior and predictive distributions are illustrated for an assay with a linear regression function and variance known up to a constant. This is one of the few cases where closed form expressions for the distributions can be calculated.

The model to be assumed in this section is

$$f(Y_{ij} \mid \eta_i, \beta, \theta) = N(\beta_0 + \beta_1 \eta_i, w_i \theta_1),$$

where w_i is known and η_i is as indicated earlier, i.e.

$$\eta_i = \begin{cases} X_i & \text{if } i \in S \\ \eta_i^U & \text{if } i \in U. \end{cases}$$

It can be recognized that this is a special case of the linear calibration problem. In almost all practical situations this is an unrealistically simple model for an assay. Typically, the assay model is a nonlinear heterogeneous calibration model (refer to Section 8.6.1 for more details).

For purposes of this section, X is defined to be the $n^s \times 2$ matrix with ith row $X_i' = (1, X_i)$ and $W = \mathrm{diag}\{r_i/w_i, i \in S\}$ is a n^s dimensional diagonal matrix. Thus, $\mu_i = \beta_0 + \beta_1 \eta_i = X_i'\beta$ if $i \in S$ and $\mu_i = \mu_i^U$ if $i \in U$.

From (3.3) it follows that

$$E[Y_{ij} \mid \eta_i] = X_i'\beta I(i \in S) + \mu_i^U I(i \in U) \ \forall i.$$

This mean function is linear in β and μ^U. By noting that a reference prior for the linear regression model is $\pi(\beta, \theta_1) \propto 1/\theta_1$ (Yang and Berger (1996)), it is easy to show that $\pi(\beta, \theta_1, \mu^U) \propto 1/\theta_1$ is a reference prior for the model being considered here.

4.3.1   Sufficient statistics

As a prelude to deriving the posterior distribution of the parameters, the sufficient statistics of the parameters are calculated. It is well known that posterior distributions depend on experimental data only through sufficient statistics; such statistics being often easier to work with than the full data set. In particular for the given assay problem, they help to identify the posterior distribution.

Theorem 4.3.1 *A sufficient statistic for* (β, θ_1, μ^U) *is* $(\hat{\beta}, S^2, \bar{Y}^U)$, *where*

$$\hat{\beta} = (X'WX)^{-1}X'W\bar{Y}^S$$

$$S^2 = (\bar{Y} - X\hat{\beta})'W(\bar{Y} - X\hat{\beta}) + \sum_{i=1}^{n}\sum_{j=1}^{r_i} \frac{(Y_{ij} - \bar{Y}_i)^2}{w_i}$$

and

$$\bar{Y}^U = \{\bar{Y}_i\}_{i \in U}.$$

<u>Proof</u>: The sample likelihood function is proportional to

$$
\begin{aligned}
L &= \theta_1^{-\frac{1}{2}\sum_{i=1}^n r_i} \exp\left\{-\frac{1}{2\theta_1}\sum_{i \in S}\sum_{j=1}^{r_i}\frac{(Y_{ij} - X_i'\beta)^2}{w_i}\right\} \exp\left\{-\frac{1}{2\theta_1}\sum_{i \in U}\sum_{j=1}^{r_i}\frac{(Y_{ij} - \mu_i)^2}{w_i}\right\} \\
&= \exp\left\{-\frac{1}{2\theta_1}\sum_{i \in S}\frac{r_i}{w_i}(\bar{Y}_i - X_i'\beta)^2 - \frac{1}{2\theta_1}\sum_{i \in U}\frac{r_i}{w_i}(\bar{Y}_i - \mu_i^U)^2 \right. \\
&\qquad \left. -\frac{1}{2\theta_1}\sum_{i=1}^n\sum_{j=1}^{r_i}\frac{(Y_{ij} - \bar{Y}_i)^2}{w_i} - \left(\sum_{i=1}^n r_i/2\right)\log(\theta_1)\right\} \\
&= \exp\left\{-\frac{1}{2\theta_1}(\bar{Y} - X\beta)'W(\bar{Y} - X\beta) - \frac{1}{2\theta_1}\sum_{i \in U}\frac{r_i}{w_i}(\bar{Y}_i - \mu_i^U)^2 \right. \\
&\qquad \left. -\frac{1}{2\theta_1}\sum_{i=1}^n\sum_{j=1}^{r_i}\frac{(Y_{ij} - \bar{Y}_i)^2}{w_i} - \left(\sum_{i=1}^n r_i/2\right)\log(\theta_1)\right\}.
\end{aligned}
$$

Since

$$(\bar{Y} - X\beta)'W(\bar{Y} - X\beta) = (\bar{Y} - X\hat{\beta})'W(\bar{Y} - X\hat{\beta}) + (\hat{\beta} - \beta)'X'WX(\hat{\beta} - \beta)$$

it follows that

$$
\begin{aligned}
L &= \exp\left\{-\frac{1}{2\theta_1}\left((\hat{\beta} - \beta)'X'WX(\hat{\beta} - \beta) - \frac{1}{2\theta_1}\sum_{i \in S}\frac{r_i}{w_i}(\bar{Y}_i - \mu_i^U)^2\right) \right. \\
&\qquad \left. -\frac{1}{2\theta_1}S^2 - \left(\sum_{i=1}^n r_i/2\right)\log(\theta_1)\right\}.
\end{aligned}
$$

By the factorization theorem the proof is complete.

4.3.2 Joint posterior distribution

The following theorem provides a convenient expression of the joint posterior distribution of (β, θ_1, μ^U) in terms of conditional and marginal posterior distributions. This is not the only possible decomposition of the joint posterior distribution but it is the most amenable to subsequent calculations.

Theorem 4.3.2 *When $\pi(\beta, \theta_1, \mu^U) \propto 1/\theta_1$ the joint posterior distribution of the assay parameters is given by*

$$\pi(\beta, \theta_1, \mu^U \mid Y, X) = \pi(\beta \mid \theta_1, Y, X)\pi(\mu^U \mid \theta_1, Y, X)\pi(\theta_1 \mid Y, X),$$

where

$$\pi(\beta \mid \theta_1, Y, X) = N(\hat{\beta}, \theta_1(X'WX)^{-1}),$$

$$\begin{aligned}
\pi(\mu^U \mid \theta_1, Y, X) &= \prod_{i \in U} \pi(\mu_i^U \mid \theta_1, Y, X) \\
&= \prod_{i \in U} N(\bar{Y}_i, w_i\theta_1/r_i),
\end{aligned}$$

$$\pi(\theta_1 \mid Y, X) = IG(\nu/2, 2/S^2),$$

$\nu = \sum_{i=1}^n r_i - p - n^u$ and IG denotes the inverse gamma distribution.

<u>Proof</u>: It is well known that

$$\hat{\beta} \sim f(\hat{\beta} \mid \beta, \theta_0) = N(\beta, \theta_1(X'WX)^{-1})$$

and

$$\bar{Y}_i \sim f(\bar{Y}_i^U \mid \mu_i^U, \theta_0) = N(\mu_i, w_i\theta_1/r_i), \quad i \in U.$$

Furthermore, S^2/θ_1 being a sum of squares of standard normal random variables has a chi-square distribution with ν degrees of freedom. Hence,

$$S^2 \sim f(S^2 \mid \theta_0) = \theta_0^{-1}G(\nu/2, 2),$$

where G denotes the gamma distribution. Thus, by sufficiency

$$\begin{aligned}
\pi(\beta, \theta_1, \mu^U \mid Y, X) &\propto f(\hat{\beta} \mid \beta, \theta_0)f(\bar{Y}^U \mid \mu^U, \theta_0)f(S^2 \mid \theta_0)\pi(\beta, \theta_1, \mu^U) \\
&\propto \theta_1^{-p/2} \exp\left\{-\frac{1}{2\theta_1}(\beta - \hat{\beta})'X'WX(\beta - \hat{\beta})\right\} \\
&\quad \cdot \left(\prod_{i \in U} \theta_1^{-1/2} \exp\left\{-\frac{1}{2\theta_1}\frac{r_i}{w_i}(\bar{Y}_i - \mu_i^U)^2\right\}\right) \\
&\quad \cdot \theta_1^{-\nu/2} \exp\left\{-\frac{S^2}{2\theta_1}\right\} \theta_1^{-1} \\
&\propto N(\hat{\beta}, \theta_1(X'WX)^{-1}) \prod_{i \in U} N(\bar{Y}_i, w_i\theta_1/r_i)IG(\nu/2, 2/S^2).
\end{aligned}$$

$$(4.6)$$

Marginal distributions are easy to derive from (4.6) as the conditional posterior distributions for β and μ_i^U drop out when (4.6) is integrated with respect to β and μ_i^U respectively.

4.3.3 Posterior distribution of β

The posterior distribution of β is given by

$$
\begin{aligned}
\pi(\beta \mid Y, X) \;&\propto\; \int_0^\infty \theta_1^{-p/2} \exp\left\{ -\frac{1}{2\theta_1}(\beta - \hat{\beta})'X'WX(\beta - \hat{\beta}) \right\} \\
&\qquad .\,\theta_1^{-(\nu+2)/2} \exp\left\{ -\frac{S^2}{2\theta_1} \right\} d\theta_1 \\
&\propto\; \left(S^2 + (\beta - \hat{\beta})'X'WX(\beta - \hat{\beta}) \right)^{-(\nu+p)/2}.
\end{aligned}
$$

Let $s^2 = S^2/\nu$. Then

$$
\pi(\beta \mid Y, X) \propto \left(1 + \frac{(\beta - \hat{\beta})'X'WX(\beta - \hat{\beta})}{\nu s^2} \right)^{-(\nu+p)/2},
$$

which is the kernel of a p-variate t distribution. It follows that

$$
\pi(\beta \mid Y, X) = t_p(\nu, \hat{\beta}, s^2(X'WX)^{-1}),   \tag{4.7}
$$

where $t_p(\nu, \lambda, \Sigma)$ denotes a p-variate t density with ν degrees of freedom, mean λ and scale Σ.

4.3.4 Posterior distribution of μ_i^U

The posterior distribution of μ_i^U is given by

$$
\begin{aligned}
\pi(\mu_i^U \mid Y, X) \;&\propto\; \int_0^\infty \theta_1^{-1/2} \exp\left\{ -\frac{1}{2\theta_1}\frac{r_i}{w_i}(\mu_i^U - \bar{Y}_i)^2 \right\} \\
&\qquad .\,\theta_1^{-(\nu+2)/2} \exp\left\{ -\frac{S^2}{2\theta_1} \right\} d\theta_1 \\
&\propto\; \left(S^2 + \frac{r_i}{w_i}(\mu_i^U - \bar{Y}_i)^2 \right)^{-(\nu+1)/2} \\
&\propto\; \left(1 + \frac{r_i(\mu_i^U - \bar{Y}_i)^2}{\nu w_i s^2} \right)^{-(\nu+1)/2} \\
&\propto\; t(\nu, \bar{Y}_i^U, w_i s^2/r_i).   \tag{4.8}
\end{aligned}
$$

Hence, $\pi(\mu_i^U \mid Y, X) = t(\nu, \bar{Y}_i^U, w_i s^2/r_i)$.

4.3.5 Joint posterior distribution of (β, μ^U)

The joint posterior distribution of (β, μ^U) can be calculated in the following analogous fashion:

$$
\pi(\beta, \mu^U \mid Y, X) \propto
$$
$$
\int_0^\infty \theta_1^{-p/2} \exp\left\{ -\frac{1}{2\theta_1}(\beta - \hat{\beta})'X'WX(\beta - \hat{\beta}) \right\}
$$
$$
\cdot \prod_{i \in U} \theta_1^{-1/2} \exp\left\{ -\frac{1}{2\theta_1}\frac{r_i}{w_i}(\bar{Y}_i - \mu_i^U)^2 \right\} \theta_1^{-(\nu+2)/2} \exp\left\{ -\frac{S^2}{2\theta_1} \right\} d\theta_1
$$
$$
\propto \left(S^2 + (\beta - \hat{\beta})'X'WX(\beta - \hat{\beta}) + \sum_{i \in U} \frac{r_i}{w_i}(\mu_i^U - \bar{Y}_i)^2 \right)^{-(\nu+p+n^u)/2}
$$
$$
= \left(S^2 + (\beta - \hat{\beta})'X'WX(\beta - \hat{\beta}) + (\mu^U - \bar{Y}^U)'W_U(\mu^U - \bar{Y}^U) \right)^{-(\nu+p+n^u)/2},
$$

where $W_U = \text{diag}\{r_i/w_i, i \in U\}$ is a n^u dimensional diagonal matrix. Recognizing the multivariate t kernel, $\pi(\beta, \mu^U \mid Y, X)$ is readily identified as a

$$
t_{p+n^u}\left(\nu, \begin{pmatrix} \hat{\beta} \\ \bar{Y}^U \end{pmatrix}, s^2 \begin{pmatrix} (X'WX)^{-1} & 0 \\ 0 & W_U^{-1} \end{pmatrix} \right)
$$

density.

It is clear from the form of the respective densities that

$$
\pi(\beta, \mu^U \mid Y, X) \neq \pi(\beta \mid Y, X)\pi(\mu^U \mid Y, X)
$$

and

$$
\pi(\mu^U \mid Y, X) \neq \prod_{i \in U} \pi(\mu_i^U \mid Y, X).
$$

In fact all of the assay parameters are a posteriori statistically dependent.

4.3.6 Posterior distribution for η^U

To calculate the posterior distribution of η^U from the above distributions, the following steps are performed:

1. calculate the joint posterior distribution of (β, μ^U),

2. apply the parameter transformation: $(\beta, \mu^U) \to (\beta, \eta^U)$,

3. integrate out β.

The resulting expression is

$$\pi(\beta, \eta^U \mid Y, X) = \int_{\mathcal{R}(\beta)} \prod_{i \in U} \pi(\beta, m(\eta_i, \beta) \mid Y, X) |\beta_1|^{-n^u} \, d\beta,$$

since $\beta_1^{-n^u}$ is the determinant of the Jacobian matrix of the transformation from (β, μ^U) to (β, η^U). The Jacobian term makes algebraic manipulation difficult and the resulting density a non-standard distribution.

The marginal posterior distributions of the components of η^U will generally be of more interest than the posterior distribution of η^U. To calculate the posterior distribution of η_i^U, the sequence of steps given earlier are followed with the exception that $\pi(\beta, \mu_i \mid Y, X)$ is calculated in step 1. The determinant of the Jacobian matrix for the transformation of (β, μ_i^U) to (β, η_i^U) is $1/\beta_1$. An algebraic expression for the resulting probability distribution function is given in Hunter and Lambroy (1981).

4.3.7 Predictive distributions

Firstly, the predictive distribution of a singleton response measurement at concentration x is derived. Let $y \sim N(x_p'\beta, w_p\theta_1)$, where w_p is known and

$$x_p = \begin{pmatrix} 1 \\ x \end{pmatrix}.$$

Then

$$
\begin{aligned}
p(y \mid x, Y, X) &= \int_{\mathcal{R}(\beta)} \int_0^\infty f(y \mid x, \beta, \theta_1) \pi(\beta, \theta_1 \mid Y, X) \, d\theta_1 \, d\beta \\
&= \int_0^\infty \int_{\mathcal{R}(\beta)} f(y \mid x, \beta, \theta_1) \pi(\beta \mid \theta_1, Y, X) \pi(\theta_1 \mid Y, X) \, d\beta \, d\theta_1 \\
&= \int_0^\infty p(y \mid \theta_1, x, Y, X) \pi(\theta_1 \mid Y, X) \, d\theta_1
\end{aligned}
$$

As both $f(y \mid x, \beta, \theta_1)$ and $\pi(\beta \mid \theta_1, Y, X)$ are normal distributions, so too is $p(y \mid \theta_1, x, Y, X)$, the predictive distribution of y at concentration x conditional on θ_1 and (Y, X). Since $E[y \mid \theta_1, x, Y, X] = E^{\pi(\beta \mid \theta_1, Y, X)}[E[y \mid x, \beta, \theta_1]]$ and $V[y \mid \theta_1, x, Y, X] = E^{\pi(\beta \mid \theta_1, Y, X)}[V[y \mid x, \beta, \theta_1]] + V^{\pi(\beta \mid \theta_1, Y, X)}[E[y \mid x, \beta, \theta_1]]$ it follows that

$$p(y \mid \theta_1, x, Y, X) = N(x_p'\hat{\beta}, w_p\theta_1 + \theta_1 x_p'(X'WX)^{-1}x_p).$$

Therefore,

$$\pi(y \mid x, Y, X) \quad \propto \quad \int_0^\infty \theta_1^{-1/2} \exp\left\{ \frac{-(y - x_p'\hat{\beta})^2}{2\theta_1(w_p + x_p'(X'WX)x_p)} \right\}$$

$$\cdot \theta_1^{-(\nu+2)/2} \exp\left\{ -\frac{S^2}{2\theta_1} \right\} d\theta_1$$

$$\propto \quad \left(S^2 + \frac{(y - x_p'\hat{\beta})^2}{w_p + x_p'(X'WX)x_p} \right)^{-(\nu+1)/2}$$

$$\propto \quad \left(1 + \frac{(y - x_p'\hat{\beta})^2}{\nu s^2(w_p + x_p'(X'WX)x_p)} \right)^{-(\nu+1)/2} .$$

Hence,

$$p(y \mid x, Y, X) = t(\nu, x_p'\hat{\beta}, s^2(w_p + x_p'(X'WX)x_p)).$$

From the above it can be seen that if y represented the mean of r responses, the predictive distribution would be

$$p(y \mid x, Y, X) = t(\nu, x_p'\hat{\beta}, s^2(w_p/r + x_p'(X'WX)^{-1}x_p)). \tag{4.9}$$

Likewise, it is also easy to show that if y is a r dimensional replicated response measurement then

$$p(y \mid x, Y, X) = t(\nu, (x_p'\hat{\beta})1_r, s^2(w_p + x_p'(X'WX)^{-1}x_p)I_r),$$

where 1_r is a vector of r ones and I_r is the $r \times r$ identity matrix. This result can be generalized to the case in which the components of y are measured at different concentrations.

4.3.8 Approximation of the general model

Posterior and predictive distributions cannot be calculated analytically for the general assay model. As the minimal sufficient statistics are usually the order statistics, appealing to sufficiency does not lead to any simplifications. One must resort to numerical calculation or analytical approximation.

One method of obtaining analytical approximations of the relevant distributions is to use approximations to reduce the assay model to a linear regression model with known weights so that the results of Sections 4.3.1-4.3.2 can be applied. The approximations and resultant procedure used herein are outlined below:

1. Write the variance function in the form $v(m,\theta) = \sigma^2 g(m,\tau)$ where $\theta = (\sigma^2, \tau)'$. That is, assume the variance function is known up to a proportionality parameter.

2. Calculate the parameter estimates $\hat{\beta}$, $\hat{\tau}$ and $\hat{\mu}^U$ and derive $\hat{\mu}$, the vector of fits.

3. Replace the mean function, $m(\eta_i, \beta)$, with its first order Taylor series expansion about $\hat{\beta}$, i.e. assume

$$m(\eta_i, \beta) = m(\eta_i, \hat{\beta}) + m_\beta(\eta_i, \hat{\beta})'(\beta - \hat{\beta}).$$

4. Substitute $g(\mu_i, \tau)$ with $g(\hat{\mu}_i, \hat{\tau})$ and set $w_i = g(\hat{\mu}_i, \hat{\tau})$.

5. Calculate s^2 and substitute the terms found in steps 3 and 4 into (4.7) and (4.8).

6. Revert to the original parameterization by reverse transforming $\hat{\tau}$ so that $v(\hat{\mu}_i, \hat{\theta}) = s^2 g(\hat{\mu}_i, \hat{\tau})$ and $W^{-1} = \operatorname{diag}\{v(\hat{\mu}_i, \hat{\theta})/r_i, \ i = 1, \ldots, n^s\}$.

From (4.7) and (4.8), the approximate posterior distributions resulting from this procedure are

$$\pi(\beta \mid Y, X) \approx t_p(\nu, \tilde{\beta}, (\hat{J}_\beta' W \hat{J}_\beta)^{-1})$$

and

$$\pi(\mu_i \mid Y, X) \approx t(\nu, \hat{\mu}_i, v(\hat{\mu}_i, \hat{\theta})/r_i), \text{for } i \in U,$$

where from the proof of Theorem 3.3.1

$$\tilde{\beta} = (\hat{J}_\beta' W \hat{J}_\beta)^{-1} \hat{J}_\beta' W (\bar{Y} - m(X, \hat{\beta}) + \hat{J}_\beta \hat{\beta}).$$

Note that the value of s^2 used in the calculation of W satisfies

$$\nu s^2 = (\bar{Y} - m(X, \hat{\beta}) + \hat{J}_\beta(\tilde{\beta} - \hat{\beta}))' W (\bar{Y} - m(X, \hat{\beta}) + \hat{J}_\beta(\tilde{\beta} - \hat{\beta})) + \sum_{i \in U} \sum_{j=1}^{r_i} \frac{(Y_{ij} - \bar{Y}_i)^2}{g(\hat{\mu}_i, \hat{\theta})}$$

which is somewhat different from the standard estimate of σ^2. To develop an estimate of $p(y \mid x, Y, X)$, firstly use the approximation

$$f(y \mid x, \beta, \theta) \approx N(m(x, \hat{\beta}) + m_\beta(x, \hat{\beta})'(\beta - \hat{\beta}), \sigma^2 g(\hat{\mu}_i, \hat{\tau})).$$

Then, following the above procedure, the approximate predictive distribution of the mean of r responses at concentration x is

$$p(y \mid x, Y, X) \approx \tag{4.10}$$

$$t(\nu, m(x, \hat{\beta}) + m_\beta(x, \hat{\beta})'(\tilde{\beta} - \hat{\beta}), v(m(\hat{\beta}, x), \hat{\theta})/r + m_\beta(x, \hat{\beta})'(\hat{J}'_\beta W \hat{J}_\beta)^{-1} m_\beta(x, \hat{\beta})).$$

If $\hat{\beta}$ is the GLS estimate corresponding to $\hat{\theta}$ then by Theorem 3.3.1 $\tilde{\beta} = \hat{\beta}$; hence, s^2 reduces to the standard estimate of σ^2 and (4.10) simplifies.

4.4   Numerical computation for the general model

The major difficulties confronting a Bayesian analysis of an assay are the numerical evaluations required in obtaining the relevant posterior and predictive distributions. As illustrated in Section 4.2, extensive calculations in the form of integration are required. In the last 15 years there has been considerable progress in the development of techniques for approximating and evaluating integrals which arise in a Bayesian analysis. The four main strategies that have been suggested are: Laplace and related analytic techniques, adaptive quadrature based on classical numerical analysis, versions of Monte Carlo importance sampling and Markov Chain Monte Carlo (MCMC) methods. For an overview and list of references dealing with the first three methods see Smith (1991). Refer to Evans and Swartz (1995) for an overview of the last three methods. See Smith and Roberts (1993), Tierney (1994), Chib and Greenberg (1995) and Besag et al. (1995) for discussion of MCMC methods.

Of the methods that are available today, MCMC simulation appears to be the most promising general approach to Bayesian computation (Draper (1998)). In the last few years there has been a massive increase in the research being done on MCMC methods and their popularity has greatly increased. The MCMC approach is well suited to the assay problem. The abundance of unknown concentrations, which in a general context can be considered missing variables, makes the degree of parameterization very high. Other computational techniques, such as importance sampling, quadrature, and closed form approximations, tend to encounter substantial computational difficulties or become impractical when the number of parameters is large. On the other hand, MCMC methods handle missing data with virtually no difficulty (Smith and Roberts (1993)).

The numerical computations reported in the thesis use MCMC simulation, in the form of the Metropolis-Hastings (M-H) algorithm, to fit Bayesian models. The M-H algorithm was developed by Metropolis et al. (1953) and subsequently generalized by Hastings (1970). This algorithm is extremely versatile as it requires that the joint distribution of the variable being generated is known only up to proportionality. This makes it well suited to the assay model since the joint posterior distribution of the assay parameters is typically only known up to proportionality.

The other well known MCMC algorithm is the Gibbs Sampler (Geman and Geman (1984)). The Gibbs sampler is a special case of the M-H algorithm (Chib and Greenberg (1995)). (Refer to Casella and George (1992) for a simple, intuitive exposition of the Gibbs Sampler.) The Gibbs Sampler requires that independent samples are generated from the full conditional posterior distributions. For the standard assay model, this procedure is inefficient because the full conditional posterior distributions involve integrals that are themselves beyond algebraic evaluation.

Issues concerning the application of the M-H algorithm to the assay problem are discussed in the remainder of this section.

4.4.1 The M-H algorithm

A feature of the M-H algorithm is its simplicity. The general setup of the algorithm is now outlined. Suppose that $z \sim \pi(z)$ is to be sampled, where $\pi(z)$ is known as the target density. Typically, $\pi(z)$ will be a non-standard multivariate probability density and some quantity dependent on this distribution is to be evaluated.

Let $q(y \mid x)$ be a probability distribution on y conditional on the variable x, where x has the same dimension as y. This is known as the probing distribution or as the candidate generating density. It is not necessary that q depends on x.

The M-H algorithm is as follows:

1. Initialize z as $z^{(0)}$ and set $i = 0$.

2. Generate z from $q(z \mid z^{(i)})$.

3. Calculate

$$
\begin{aligned}
\alpha(z^{(i)}, z) &= \min\left[\frac{\pi(z)q(z^{(i)} \mid z)}{\pi(z^{(i)})q(z \mid z^{(i)})}, 1\right], \text{ if } \pi(z^{(i)})q(z \mid z^{(i)}) > 0 \\
&= 1, \text{ otherwise.}
\end{aligned}
$$

4. Generate u from $U(0,1)$.

5. If $u \le \alpha(z^{(i)}, z)$ set $z^{(i+1)} = z$; otherwise set $z^{(i+1)} = z^{(i)}$.

6. Return to step 2 and repeat until the quantity being evaluated has been computed with sufficient accuracy.

In the preceding algorithm, the probability $p(x,y) = \alpha(y,x)q(y \mid x)$ emulates a transition kernel in a Markov Chain. The probability of the Markov Chain moving from state x to state y is $p(x,y)$. If q is such that the Markov Chain is reversible and aperiodic, then the steady state or equilibrium distribution of the chain is π. If these conditions are met, the empirical distribution of z converges to $\pi(z)$.

The condition of reversibility requires that

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

By the choice of $\alpha(x,y)$, this condition holds if for all x and y in the domain of π, it is possible to move from x to y in a finite number of iterations with nonzero probability. Aperiodicity is usually satisfied if $q(y \mid x)$ has positive support surrounding x (Chib and Greenberg (1995)). Also see Tierney (1991) and Tierney (1994) for a complete discussion of the convergence criteria for MCMC chains.

4.4.1.1 Product of kernels principle

The following result, referred to in this thesis as the "product of kernels principle", is one of the most useful properties of MCMC algorithms.

Define a conditional transition kernel to be a transition kernel for a component of a vector z that depends on the current values or states of the other components of z. Then the product of conditional transition kernels for any partition of z has $\pi(z)$ as its invariant distribution.

As discussed in Chib and Greenberg (1995), this result implies that sub-components of z can be drawn in succession from their respective conditional transition kernels and the chain will still converge. Therefore, it is not necessary to run the conditional transition kernels to convergence for every value of the variables conditioned on. In fact only one iteration need be performed each time a conditional transitional kernel

is visited. The product of kernels principle is important because it is often far easier to find several conditional kernels that converge to their respective conditional densities than it is to find a kernel that converges to the joint distribution of z.

4.4.1.2   Calculation of inferences

The following result is the basis of MCMC inference: *If $z^1, z^2, \ldots, z^n$ is a realisation from a convergent MCMC chain then*

$$\frac{1}{n} \sum_{n=1}^{\infty} f(z^i) \xrightarrow{n \to \infty} E^{\pi}[f(z)], \quad almost\ surely,$$

where $f(z)$ is some function of z. (See for example Smith and Roberts (1993), Besag et al. (1995).)

Therefore, as more samples are generated, the empirical average (also known as the Monte Carlo average) of the function converges to the true mean of the function. Thus, to evaluate any particular inference, one merely needs to find the function(s) whose expectation(s) give the required inference. The following are examples:

- Mean: Use $f(z) = z$.

- Variance: Use $f_1(z) = z^2$ and $f_2(z) = z$ and calculate $E^{\pi}[f_1(z)]/n - E^{\pi}[f_2(z)]^2$.

- α Percentile of $\pi(z)$: Let $f(z, z_\alpha) = I(z < z_\alpha)$ and find z_α such that $E^{\pi}[f(z, z_\alpha)] = \alpha$.

4.4.2   M-H chain for an assay

For the analysis of an assay the M-H algorithm needs to be applied to $\pi(\beta, \theta, \eta^U \mid Y, X)$ or some re-parameterization thereof. All posterior and predictive inferences can be evaluated from the values of (β, θ, η^U) output from the algorithm. The ingredients of the M-H chain developed for the analysis of assay data in this thesis are now discussed.

There are three key elements of the M-H chain used to draw samples from the posterior distribution of the assay parameters. These elements are the structure of the chain, the types of the transition kernels and the choice of candidate generating densities.

4.4.2.1 Structure of chain

The product of kernels principle enables the transitional kernel to be partitioned into conditional kernels. It has been observed that this is most beneficial, in fact only beneficial, when the parameters in different partitions are close to being independent.

The assay parameters fit naturally into the following three groups: β, θ and η^U. Although there will always be a posteriori dependence between these parameters, numerical results reveal that it is not necessary to use a joint kernel. Since each component of η^U is its own entity, separate components were also used for each unknown concentration. This meant that a total of $n^u + 2$ conditional kernels were used.

4.4.2.2 Type of transition kernel

For each conditional kernel, a random walk transition kernel was used. The random walk chain is a natural way of moving about the posterior density. Unlike independence chains, chains for which $q(y \mid x)$ does not depend on x (the current state of the chain), one is not rooted to a certain location. It is often the case that x is the mean or some other measure of the location of the candidate generating density for the next step. Thus, there is no reliance on a good choice of location. However, there is no reason to believe that independence or other types of transition kernels would not be as successful. The random walk chain was used here because the implementation was quite easy. (Refer to Chib and Greenberg (1995) for descriptions on the different types of transition kernels.)

4.4.2.3 Candidate generating density

The candidate generating density or more generally distribution is the most crucial component of a M-H chain. A good candidate generating distribution is necessary for the chain to produce reliable results in realistic computing time. The main goal in choosing the candidate generating distribution is to obtain a chain that mixes well (Draper (1998)). A chain is said to mix well if all regions of the posterior density are explored efficiently. "A chain that is mixing well will move around freely, happily jumping all over the place" (Draper (1998)).

Three prominent guidelines in the literature concerning the choice of a candidate

generating density are:

1. Try to ensure that the candidate generating density is an over-dispersed version of the target density (Draper (1998)).

2. The candidate generating density should have thicker tails than the target density (Berger (1985), Section 4.9.2).

3. Choose a candidate generating density so that the current state of the chain is the expected value of the next state of the chain (Draper (1998)).

For random walk chains, the first recommendation is less important because the movement of the location of the candidate generating density naturally inflates the dispersion of the draws. See Gilks and Roberts (1996) for a more comprehensive list of guidelines for choosing a candidate generating density. In Section 4.4.2.7 the candidate generating density used for the assay parameters is described.

4.4.2.4 Parameter transformation

The manner in which a problem is parameterized often holds the key to finding good candidate generating densities. If a particular parameterization leads to a posterior distribution with nice properties such as unimodality, symmetry and parameter independence, then the search for an efficient candidate generating density is simplified. The parameterization of the standard assay model can be improved in the following ways:

1. Transforming so that the parameters are more independent.

2. Transforming so that the parameters are real valued and the model is identifiable (determined uniquely by the parameters).

These two improvements are explained and discussed below with reference to the assay model.

4.4.2.5 Independence inducing transformations

It is advantageous to induce independence both within and between partitions of parameters. In the following discussion the "conditional asymptotic covariance matrix

for a group of parameters" represents the inverse of the Fisher information matrix for that group of parameters. An independence inducing transformation is defined as one which diagonalizes the conditional asymptotic covariance matrix for a group of parameters. While such a transformation does not necessarily induce complete independence, it has proven to be a good heuristic. For the standard immunoassay model the following independence inducing transformations have been found:

$$\eta_i^U \to \mu_i^U = \beta_4 + \frac{\beta_1 - \beta_4}{1 + (\eta_i^U/\beta_3)^{\beta_2}}$$

and

$$\theta_1 \to \tilde{\theta}_1 = \theta_1 \mu^{\theta_2},$$

where μ denotes the geometric mean of μ.

The first transformation makes the conditional asymptotic covariance between β and η^U equal to zero. The proof of this result is immediate once it is noticed that the terms in the likelihood function involving both β and μ^U are separated. This transformation effectively uncouples β from η^U.

The second transformation makes the Fisher information matrix for θ diagonal (see Theorem 3.5.1). The effect of the transformation was demonstrated by noting that the convergence of the chain was several times faster than it was under the original parameterization of θ. This transformation also induced symmetry in the posterior distribution of the transformed parameter (the posterior distribution of $\tilde{\theta}_1$ is much less skewed than the posterior distribution of θ_1). When θ is known, the effect of the parameterization is even more pronounced. This is understandable as there are no confounding effects from other parameters.

In general, re-parameterizations such as that applied to the power variance function in (2.2) are not always apparent. The power variance function is in many ways a natural model for a variance function which depends on the mean response. No such parameterization has been found for the quadratic variance function in (2.3).

4.4.2.6 Constraint removing and identifiability inducing transformations

The following constraints are needed to ensure that the model is well defined:

1. $\beta_1 \geq \beta_4 > 0$

2. $\beta_3 > 0$

3. $\theta_1 > 0$

These conditions can be incorporated in the prior distribution. However, since constraints on parameters can sometimes lead to strangely shaped posteriors, from the point of view of numerical evaluation, it is beneficial transforming the parameters so that such constraints are eliminated. A model is said to be non-identifiable if the same fit can be obtained with more than one combination of parameter values. Non-identifiability is undesirable as the posterior will then be multimodal, making sampling from it more difficult. Notice that $(\beta_4, -\beta_2, \beta_3, \beta_1)$ gives the same fit as $(\beta_1, \beta_2, \beta_3, \beta_4)$ so the sigmoid function makes the assay model non-identifiable.

The following transformations remove the constraints from the parameter space and also make the MNCH model identifiable:

- $\log(\beta_1 - \beta_4) \rightarrow \beta_1$

- $\log(\beta_3) \rightarrow \beta_3$

- $\log(\beta_4) \rightarrow \beta_4$

- $\log(\theta_1) \rightarrow \theta_1$.

This step proved to be of minimal benefit when fitting the NMCH model because good candidate generating distributions were found. It may, however, be required for other assay models.

Transformations that induce linearity in β were also considered. Ratkowsky and Reedy (1986) compared such transformations for the sigmoid function. These transformations had no noticeable effect on the convergence of the chain.

4.4.2.7 Candidate generating density for assay parameters

The candidate generating function for each conditional kernel in the M-H chain used to fit the assay model is a multivariate t distribution. After the independence inducing transformations were applied the posterior distributions were approximately symmetric. It follows from the symmetry of a t distribution that the expected value of the chain at any iteration is always the current state of the chain. The multivariate t density is preferred to the multivariate normal density because it has thicker tails.

Since the candidate generating density is symmetric about the mean and because the chain is a random walk chain, the probability of moving from z^i to z^{i+1} reduces to

$$\alpha\left(z^{(i)}, z^{(i+1)}\right) = \min\left\{1, \frac{\pi\left(z^{(i+1)}\right)q\left(z^{(i)} \mid z^{(i+1)}\right)}{\pi\left(z^{(i)}\right)q\left(z^{(i+1)} \mid z^{(i)}\right)}\right\} = \min\left\{1, \frac{\pi\left(z^{(i+1)}\right)}{\pi\left(z^{(i)}\right)}\right\}.$$

To ensure that statistical dependencies between the parameters in each kernel were incorporated into the candidate generating densities, the scale parameter of the candidate generating t distribution was a Hessian based approximation of the asymptotic covariance matrix for those parameters. This is a vital step. The convergence of the chain was found to be cumbersome if posterior dependence between the parameters was ignored.

A number of values for the degrees of freedom were tried before three was selected. Following Muller (1991), a tuning constant was used to further adjust the scale of the candidate generating density.

4.4.3   Fitting the assay model

The details of the method used to fit the assay model and to calculate associated inferences are summarized in the following algorithm:

1. Transform the parameters so as to make the anticipated posterior distribution as nice as possible and partition the parameters accordingly.

2. Run the chain, sequentially drawing each group of parameters from their respective transition kernel. For each transitional kernel, use a random walk chain with candidate generating density taken as a multivariate t distribution with three degrees of freedom and with scale parameter a Hessian based estimate of the conditional covariance matrix of the parameters. Use the tuning constant to adjust the scale parameter as necessary. Define the transitional probability to be zero if any draws fall outside of the parameter space. The order in which β, θ and μ^U are generated is not important and in fact can be changed.

3. Repeat 2 until it appears that the chain is drawing from its equilibrium distribution. The usual rule of thumb is to plot the draws and their long-run average against time and continue until these plots indicate stability has been

attained. Once this point has been reached reset to zero any quantities that will be used in the calculation of inferences.

4. Continue to repeat 2 until enough draws have been made that all inferences of interest have been calculated to the required level of accuracy. This is ascertained by plotting the value of the inference over time for independent chains initialized at different starting points and using different random number streams. If the realized values of the inference on all chains are very close to one another, then this is a good indication that the required accuracy has been obtained. If this is not the case, the simulation is continued.

The methods given in the last two steps are the basis of the methods currently being used to determine when a chain reaches its steady state distribution and when inferences have been calculated accurately enough. There is a lot of ongoing research being devoted to this area. The task of determining when the effect of the initial state becomes negligible and an inference has been calculated to a specified level of accuracy are difficult problems that are magnified when the parameter has high dimensionality. Refer to Besag et al. (1995) for discussion of various diagnostic checks on the output from MCMC chains.

4.5 Analysis of the NMCH data

In this section Bayesian analysis is used to fit the NMCH model to the NMCH data. Recall that these data are real and not simulated; hence, there are no "true values" of the parameters. Prior independence is assumed between (β, θ) and η^U. The non-informative prior for (β, θ) is given by

$$\pi_m(\beta, \theta) = 1/\tilde{\theta}_1.$$

In Analysis 1, a flat improper prior is assumed for μ^U. That is

$$\pi_u(\mu^U) = 1.$$

In Analysis 2, the components of η^U are a priori independent and identically distributed with

$$\pi_0(\eta_i^U) = N(\eta_0, \tau_0^2),$$
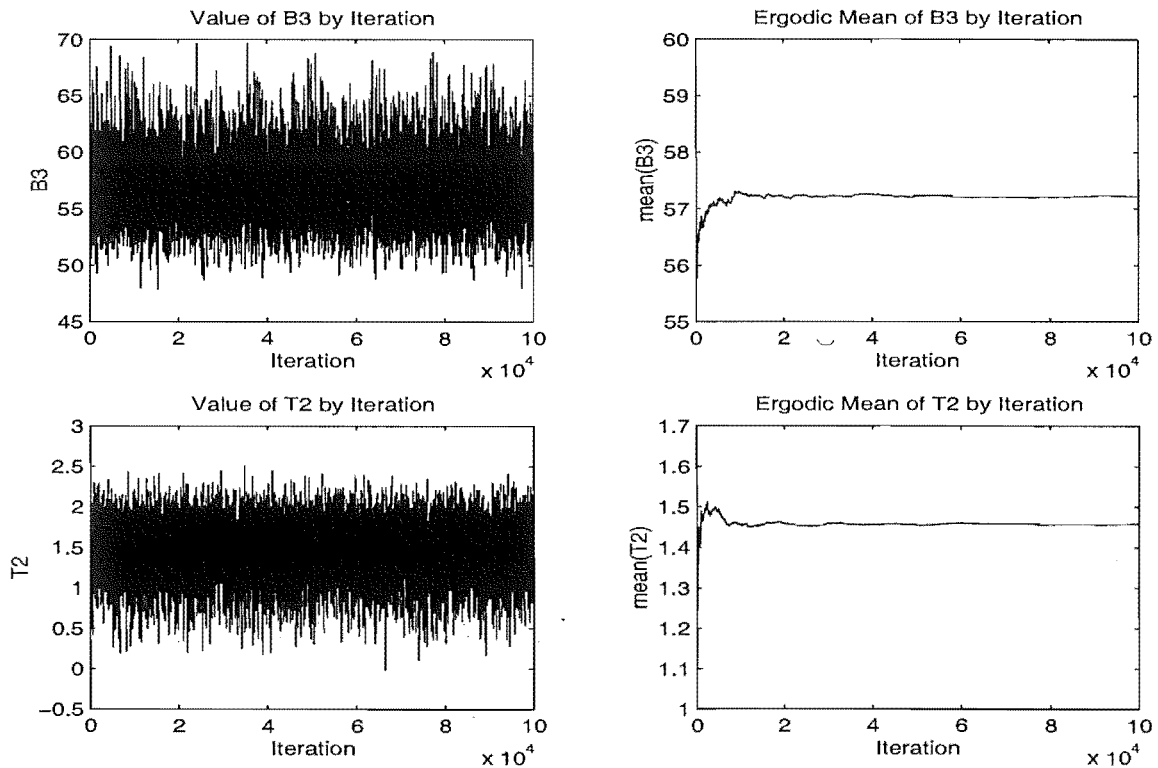
Figure 4.1: Plots of the value and ergodic average of β_3 (denoted B3) and θ_2 (denoted T2) over 100,000 iterations of the Metropolis-Hastings algorithm for analysis 1 of the NMCH data.

where $\mu_0 = 135$ and $\tau_0 = 500$. These values of μ_0 and τ_0 resemble the mean and variance of previous estimates of the concentrations in unknown samples. It should be noted that strictly speaking this is not a Bayesian approach for the selection of a prior distribution. The method has some affinity with the empirical Bayes procedure which is also not a Bayes procedure. A true Bayesian methodology is described in Section 4.6.

4.5.1 Convergence diagnostics

Analysis of the draws indicated that the Markov Chain appeared to have passed through its transient phase by the 1000th iteration. However, the first 10,000 variates drawn were excluded from the calculation of inferences to be sure that there were no lingering transient effects.
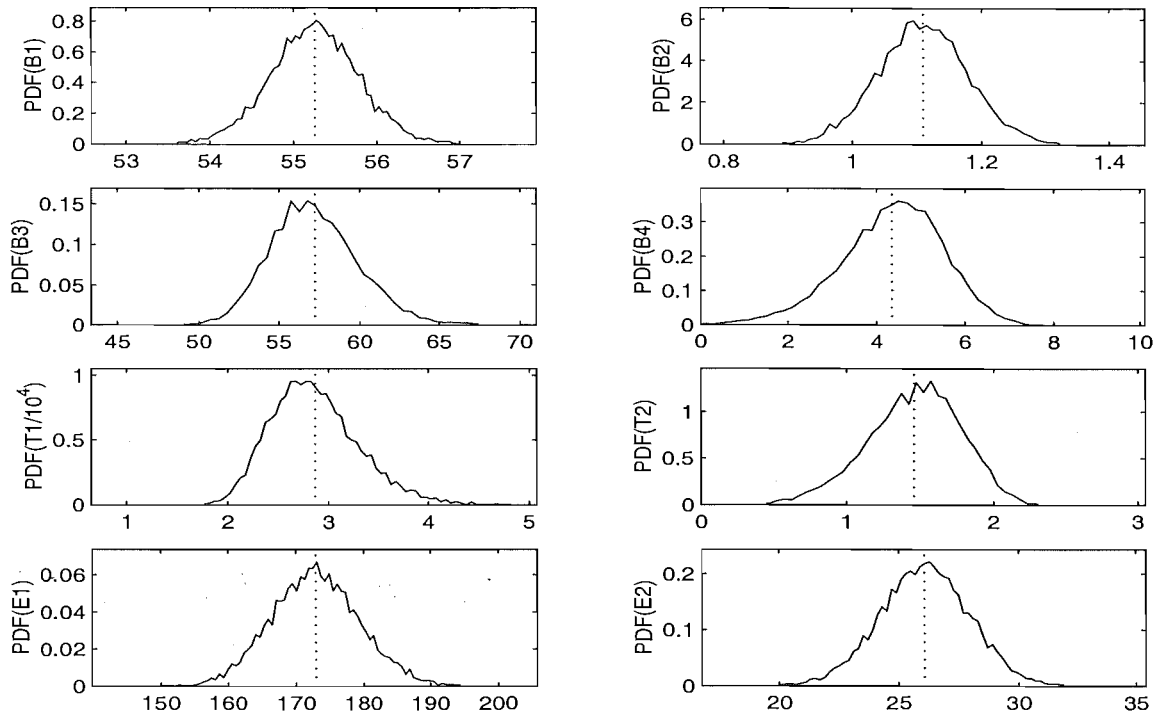
Figure 4.2: Posterior distributions of the model parameters and two unknown concentrations for the NMCH model with a fully non-informative prior on the NMCH data (the x-axis is the parameter value, the y-axis is the value of the density).

The point at which one can be assured that inferences have attained a specified level of accuracy is more difficult to determine. After 100,000 iterations the ergodic means of the parameters appeared to be sufficiently stable for reliable evaluation of inferences (see the plots on the rhs of Figure 4.1).

4.5.2 Fitted models

The posterior distributions of β (B in the plots) and θ (T in the plots) and for two unknown concentrations (E in the plots) are shown in Figures 4.2 and 4.3.
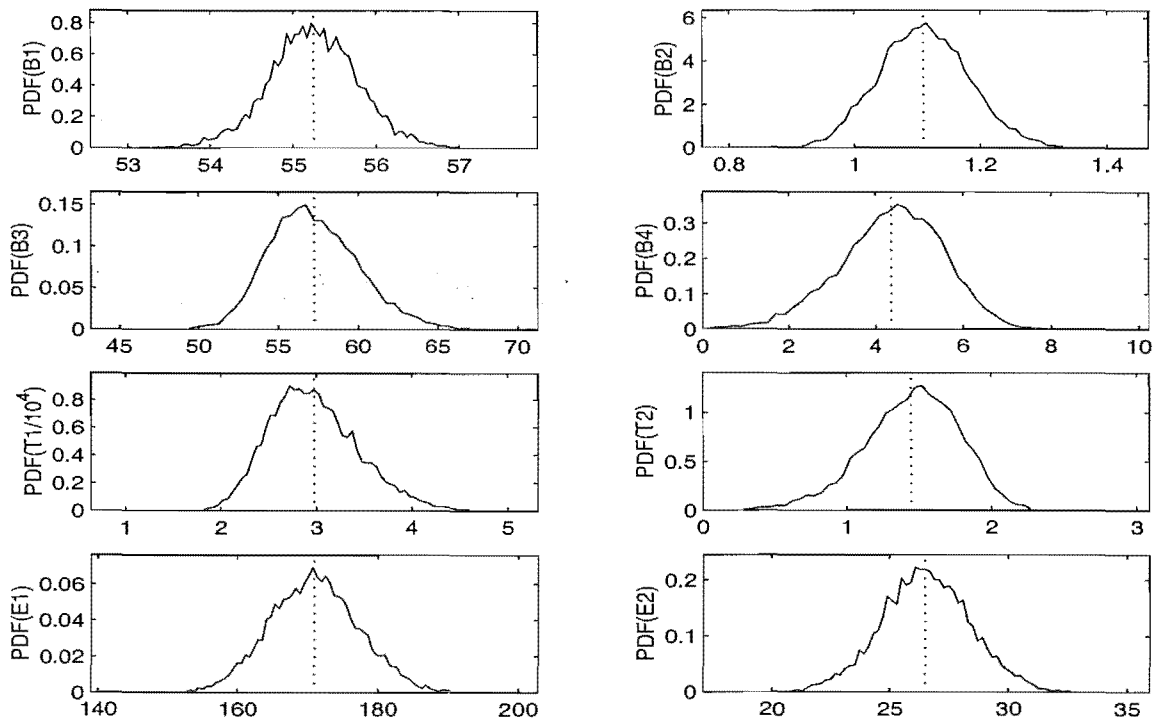
Figure 4.3: Posterior distributions of the model parameters and two unknown concentrations for the NMCH model with a non-informative prior for β and θ and an informative prior for η on the NMCH data (the x-axis is the parameter value, the y-axis is the value of the density).

Table 4.1: Posterior means, posterior standard deviations and extended REML estimates of the model parameters and two unknown concentrations for the NMCH model on the NMCH data.

Parameter	Non-inf. Bayes		Inf. Bayes		ExREML
	Mean	Std. Dev.	Mean	Std. Dev.	Estimate
β_1	55.25	0.53	55.25	0.54	55.26
β_2	1.11	0.07	1.11	0.07	1.11
β_3	57.21	2.76	57.24	2.81	56.66
β_4	4.35	1.15	4.35	1.18	4.54
θ_1	28660	4417	29764	4690	27130
θ_2	1.4578	0.3188	1.4494	0.3276	1.52
η_1	172.91	6.5785	170.90	6.3889	172.62
η_2	26.0761	1.8709	26.5109	1.8873	26.13

4.5.2.1 Discussion

The distributions displayed in Figures 4.2 and 4.3 have some interesting features. The posterior distribution for each parameter is reasonably symmetric with virtually no mass close to boundary points. This is the reason why transition kernels based on the t-distribution work so well for this model.

To the precision of graphical inspection, the (marginal) posterior distributions for β_1 and β_2 appear symmetric. The posterior distribution of β_3 is clearly right skewed while $\pi(\beta_4 \mid Y, X)$ is left skewed. It is also the case that β_3 and β_4 are a posteriori negatively correlated.

The posterior correlation between $\tilde{\theta}_1$ and θ_2 is negative and the skewness of their posterior distributions is in opposite directions. The posterior distribution of $\tilde{\theta}_1$, the scale parameter, is skewed to the right whereas θ_2, the shape parameter, is skewed to the left.

There are no noticeable differences in the shape of the posterior distributions between the two analyses. For a more precise inspection, the posterior means and standard deviations of the model parameters refer to Table 4.1. For the purpose of comparison the extended REML estimates are also shown in Table 4.1.

The posterior means from Analysis 1 in particular, are very similar to the extended REML estimates. From the perspective of point estimation, this validates the use of extended REML as an approximation to the posterior mean, at least for immunoassays data such as this. Incidentally, exact calculation of the estimation

errors of the extended REML estimates is intrinsically compromised since the true values of the assay parameters are not known. The posterior standard deviation can, however, be calculated along with the posterior mean and reported as a measure of the uncertainty in the value of the parameter.

It is evident from Figures 4.2 and 4.3 and Table 4.1 that the informative prior on η^U has little effect on the posterior distributions of β and θ. This is understandable given that this prior contains no information about β and θ. The informative prior has the effect of pulling the posterior means of η_1 and η_2 a little towards the prior mean. This effect, known as shrinkage, is a trait that frequently occurs when an informative prior is used in Bayesian analysis.

4.6   Grand immunoassay model

In the final section of this chapter, a hierarchical Bayesian model is developed for the analysis of data from a batch of assays of the same type. Assays of the same type are assays homogeneous in all departments except the statistical design; i.e. the values of the concentrations of both standards and unknowns may be different. The substance being analyzed, the procedure and the reagents used are the same in all assays. In the final part of this section a procedure is developed for passing information from completed assays on to future assays.

4.6.1   Model for a batch of assays

The same data generating process is experienced by assays that are deemed to be of the same type. This allows information to be shared between the assays.

Within a single assay the unknown concentrations can be viewed as an independent sample of the concentrations in the population of eligible subjects. In a similar vein, the model parameters across a batch of assays can be thought of as independent realizations from a population of parameter values for assays of that type. It makes sense that the parameter values of the model are subject to variation as uncontrollable factors or hidden error processes will surely act between assays. The behaviour of the assay parameters is described by their marginal distribution. This is commonly known as the prior distribution.

Let Y_{hij} denote the jth replicate, for the ith sample in the hth of k assays. Let η_{hi} be the associated concentration and (β_h, θ_h) be the model parameters for assay h. The model for the data from a batch of assays is then

$$Y_{hij} \sim f(Y_{hij} \mid, \eta_{hi}, \beta_h, \theta_h),$$

$$(\beta_h, \theta_h) \sim \pi_m(\beta_h, \theta_h \mid \beta_\pi, \theta_\pi)$$

and

$$\eta_{hi} \sim \pi_u(\eta_{hi} \mid \eta_\pi),$$

for $j = 1, \ldots, r_{hi}$, $i = 1, \ldots, n_i$ and $h = 1, \ldots, k$.

The likelihood function involves just the observed data and the assay parameters. This is the first stage of the model. The prior distributions are the second stage of the model. The prior distributions are dependent on a further set of parameters, η_π, β_π and θ_π. These parameters are known as prior parameters. These do not necessarily have the same dimensions as η_{hi}, β_h and θ_h respectively. The joint prior distribution of $(\beta_h, \theta_h, \eta_h)$ will be denoted by $\pi_1(\beta_h, \theta_h, \eta_h \mid \beta_\pi, \theta_\pi, \eta_\pi)$ in this section.

To perform the analysis, the prior parameters must be specified. One method of doing this is direct elicitation. Then the prior distribution is completely specified. Thus no information is able to be shared between the assays unless the same samples are analyzed on multiple assays. In other words, the assays may as well be analyzed independently.

A hierarchical Bayesian model is a natural way of describing the data generating process when a batch of assays is analyzed. Instead of assigning specific values to the prior parameters, distributions that describe the uncertainty in the value of the prior parameters are specified. Such distributions are called stage II priors or hyperpriors. In a hierarchical Bayesian model, the stage I prior is often thought of as the structural component of the prior and the hyperprior as the subjective component of the prior (Berger (1985)).

Since the model parameters and the unknown concentration parameters are regarded independently, it makes sense that the associated prior parameters are also regarded independently. The hyper-priors for the assay model are denoted as $\pi_2(\beta_\pi, \theta_\pi)$ and $\pi_2(\eta_\pi)$. These densities may depend on known hyper-prior parameters.

4.6.2 Analysis of a batch of assays

The hierarchical Bayesian model for a batch of assays resembles a union of the models from many assays. Instead of one assay there are a collection of assays and therefore many more unknown parameters.[1]

Let $P = \{\beta_h, \theta_h, \eta_{hi}, h = 1, \ldots, k, i \in U\}$, $Q = (\beta_\pi, \theta_\pi, \eta_\pi)$ and $D = \{D_h, h = 1, \ldots, k\}$ where $D_h = (Y_h, X_h)$ denotes the data for assay h. Define Z_{-i} to mean $\{Z_j, j \neq i\}$, $Z_{<i}$ to mean $\{Z_j, j < i\}$ and $Z_{\leq i}$ to mean $\{Z_j, j \leq i\}$ for an arbitrary entity Z.

Since D given P and Q does not depend on Q and hence $f(D \mid P, Q) = f(D \mid P)$ the posterior distribution of the model parameters is

$$\pi_1(P \mid D) = \frac{1}{f(D)} \int_{\mathcal{R}(Q)} f(D \mid P) \pi_1(P \mid Q) \pi_2(Q) dQ,$$

where

$$f(D) = \int_{\mathcal{R}(P)} \int_{\mathcal{R}(Q)} f(D \mid P) \pi_1(P \mid Q) \pi_2(Q) dQ dP.$$

To obtain the posterior distribution for any parameter of interest, the remaining components of P are simply integrated out. For inferences concerning the hth assay the quantity of interest is

$$\pi_1(P_h \mid D) = \int_{\mathcal{R}(P_{-h})} \pi_1(P \mid D) dP_{-h}.$$

Further integration yields the marginal distribution of specific components of P_h. Notice that the posterior distribution for the hth assay depends on D, the data from the whole batch of assays. If only the data from assay h were used, then inferences would be based on $\pi_1(P_h \mid D_h)$ alone and there would be no pooling or sharing of information between the assays.

In theory this is how the data from a batch of assays could be simultaneously analyzed and inferences for any particular assay evaluated. The computation will, however, be difficult because it is huge, especially for large k.

4.6.3 Practical application

In practice it is unlikely that data from a full batch of assays will be available for the analysis of each individual assay. Because results are usually required urgently,

---

[1]This is in fact the set up for an empirical Bayes problem.

assays tend to be analyzed as the data are received. The analysis cannot be delayed until all the assays in the batch have been completed, since assays are usually carried out one at a time. Information from the analysis of assays preceding the current assay can, however, be incorporated in the analysis of the current assay. The method is illustrated in the following theorem.

Theorem 4.6.1 *Suppose that h out of the k assays have been completed. The posterior distribution for the hth assay can be written in the form*

$$\pi_1(P_h \mid D_{\leq h}) = \frac{1}{p(D_h \mid D_{<h})} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q)\pi_2(Q \mid D_{<h})dQ,$$

where $\pi_2(Q \mid D_{<h})$ is the posterior distribution of Q given $D_{<h}$ and

$$p(D_h \mid D_{<h}) = \int_{\mathcal{R}(P_h)} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q)\pi_2(Q \mid D_{<h})dQdP_h$$

is the predictive distribution of D_h given $D_{<h}$.

Proof: From the theorem of total probability and Bayes' theorem

$$
\begin{aligned}
\pi_1(P_h \mid D_{\leq h}) &= \int_{\mathcal{R}(P_{<h})} \pi_1(P_{\leq h} \mid D_{\leq h})dP_{<h} \\
&= \frac{1}{f(D_{\leq h})} \int_{\mathcal{R}(P_{<h})} \int_{\mathcal{R}(Q)} f(D \mid P_{\leq h})\pi_1(P_{\leq h} \mid Q)\pi_2(Q)dQdP_{<h} \\
&= \frac{1}{f(D_{\leq h})} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q) \int_{\mathcal{R}(P_{<h})} f(D_{<h} \mid P_{<h})\pi_1(P_{<h} \mid Q) \\
&\qquad\qquad .\pi_2(Q)dP_{<h}dQ \\
&= \frac{f(D_{<h})}{f(D_{\leq h})} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q) \\
&\qquad\qquad .\int_{\mathcal{R}(P_{<h})} \frac{f(D_{<h} \mid P_{<h})\pi_1(P_{<h} \mid Q)\pi_2(Q)}{f(D_{<h})}dP_{<h}dQ,
\end{aligned}
$$

where

$$f(D_{<h}) = \int_{\mathcal{R}(P_{<h})} \int_{\mathcal{R}(Q)} f(D_{<h} \mid P_{<h})\pi_1(P_{<h} \mid Q)\pi_2(Q)dQdP_{<h}.$$

Hence,

$$
\begin{aligned}
\int_{\mathcal{R}(P_{<h})} \frac{f(D_{<h} \mid P_{<h})\pi_1(P_{<h} \mid Q)\pi_2(Q)}{f(D_{<h})}dP_{<h} &= \int_{\mathcal{R}(P_{<h})} \pi_1(P_{<h}, Q \mid D_{<h})dP_{<h} \\
&= \pi_2(Q \mid D_{<h}).
\end{aligned}
$$

Since $f(D_{\leq h}) = p(D_h \mid D_{<h})f(D_{<h})$ it follows that

$$\pi_1(P_h \mid D_{\leq h}) = \frac{1}{p(D_h \mid D_{<h})} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q)\pi_2(Q \mid D_{<h})dQ. \quad (4.11)$$

The fact that

$$p(D_h \mid D_{<h}) = \int_{\mathcal{R}(P_h)} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q)\pi_2(Q \mid D_{<h})dQdP_h$$

is all that is needed to complete the proof. This result follows easily by conditioning on P_h and Q and applying the theorem of total probability.

Theorem 4.6.1 is an important result. It says that all experimental information from past assays of relevance to subsequent assays in the batch is contained in the posterior distribution of the hyper-parameters. After each assay is analyzed $\pi_2(Q \mid D_{\leq h})$ merely needs to be stored and then used as the hyper-prior in the analysis of the next assay. Information can be passed from one assay to the next (see Section 4.6.4) without affecting the complexity of any of the individual analyses.

Note that after integrating Q out of the numerator of (4.11), the corresponding result at the level of the stage I prior is obtained

$$\pi_1(P_h \mid D_{\leq h}) = \frac{1}{f(D \mid D_{<h})} f(D_h \mid P_h)\pi_1(P_h \mid D_{<h}).$$

It might be thought that $\pi_1(P_h \mid D_{\leq h})$ can be used to pass information onto the $(h+1)$th assay. However, the stage I prior is the wrong medium for passing information between assays. This is because P_h is specific to the hth assay; hence, $\pi_1(P_h \mid D_{\leq h})$ has no bearing on the $(h+1)$th assay. On the other hand the analysis of every assay depends on the state of knowledge about Q.

4.6.4 Sequential analysis of a batch of assays

Firstly, the posterior distribution of the prior parameters given all the data, $\pi_2(Q \mid D_{\leq h})$, can be computed using $\pi_2(Q \mid D_{<h})$, the posterior of the prior parameters given all the data up to the data from the last assay, as:

$$
\begin{aligned}
\pi_2(Q \mid D_{\leq h}) &= \frac{f(D_{\leq h} \mid Q)\pi_2(Q)}{f(D_{\leq h})} \\
&= \frac{f(D_h \mid Q)f(D_{<h} \mid Q)\pi_2(Q)}{p(D_h \mid D_{<h})f(D_{<h})}
\end{aligned}
\quad (4.12)
$$

$$= \frac{f(D_h \mid Q)\pi_2(Q \mid D_{<h})}{p(D_h \mid D_{<h})}$$

$$= \left(\int_{\mathcal{R}(P_h)} f(D_h \mid P_h)\pi_1(P_h \mid Q)dP_h \right) \frac{\pi_2(Q \mid D_{<h})}{p(D_h \mid D_{<h})}$$

Equation (4.12) follows from the fact that if given P, the D_h are independent and given Q, the P_h are indentically independent then given Q, the D_h are independent.

The procedure for analyzing data from a batch of assays using the hierarchical Bayesian model described in this section is therefore:

1. Specify an initial prior for Q, subjectively or otherwise. Set $D_0 = \{\}$ and h=1;

2. Analyze the hth assay using $\pi_2(Q \mid D_{<h})$ as the hyper-prior.

   - Calculate

   $$\pi_1(P_h \mid D_{\leq h}) = \frac{1}{p(D_h \mid D_{<h})} \int_{\mathcal{R}(Q)} f(D_h \mid P_h)\pi_1(P_h \mid Q)\pi_2(Q \mid D_{<h})dQ,$$

   the posterior distribution of P_h

   - Use $\pi_1(P_h \mid D_{\leq h})$ to evaluate the inferences of interest.

3. Update the posterior distribution of Q, the hyper-prior parameters, according to

   $$\pi_2(Q \mid D_{\leq h}) = \left(\int_{\mathcal{R}(P_h)} f(D_h \mid P_h)\pi_1(P_h \mid Q)dP_h \right) \frac{\pi_2(Q \mid D_{<h})}{p(D_h \mid D_{<h})}.$$

4. Return to step 2 and continue until all the assays in the batch have been analyzed.

4.6.5 Discussion

Bayes' theorem has provided a means of updating knowledge about the data generating process of an assay. The analysis of a sequence of assays of the same type is an illustration of the function of Bayes' theorem as a learning tool.

After the specification of the structural and subjective components of the prior for the analysis of the first assay, no further subjective input is required. Therefore, this approach is an automatic method of conducting an informative Bayesian analysis for the second assay onwards in a batch of assays. As more assays are

analyzed the hyper-prior will become more developed and most likely more informative; hence, it is expected that the precision of the inferences would improve as more assays are analyzed.

The hyper-prior must be specified for the first assay in the batch. An informative hyper-prior can be elicited as if a single assay were being analyzed. Alternatively, a non-informative hyper-prior may be used. The use of the continuously updated posterior will continue for as long as the data generating process remains the same. As the number of assays increases the influence of the initial prior will decrease and eventually become negligible. If the data generating process changes, the hyper-prior should of course be re-specified and the analysis of a new batch of assays initiated.

Chapter 5

Minimum detectable concentration

5.1 Definition and preliminary remarks

The minimum detectable concentration (MDC) of an assay is the smallest concentration (in an unseen sample) the assay can reliably measure. The MDC is a very general quantity in that the term "reliable measurement" has no universally accepted definition. The key ingredient and distinguishing feature of a measure of the MDC is the definition of a reliable measurement, several of which will be presented in this chapter. Reliable detection is a synonym that will at times be used for reliable measurement.

The MDC is usually calculated using all of the observations from the assay. However, there are specific applications where it would be appropriate to only use the standards. One such example is when the MDC is used to determine the best concentrations at which the standards should be set. To calculate the MDC on a subset of the data, the necessary observations need only be excluded.

5.2 Existing measures of MDC

The objective of this section is to set the basis and motivation for the remainder of the chapter by reviewing the most widely used measures of MDC and identifying areas of potential improvement. This is by no means a comprehensive literature review. Testimony to this is the fact that one existing measure is not mentioned until the end of the chapter (Section 5.3.4). Notwithstanding, it is fair to say that

the most respected and widely used measures are covered.

In this chapter the notation $\hat{\beta}$ and $\hat{\theta}$ will be used to indicate estimators arising from generalized least squares, maximum likelihood or some variant such as extended REML and extended MML. For the measures of MDC discussed in this section, $\bar{y}$ represents the mean of r hypothetical response measurements made on an independent sample having concentration x.

5.2.1 Critical limit (CL)

The CL was introduced in Currie (1968). It was adapted to a regression set-up by Rodbard (1978). The following definition is based on Rodbard (1978).

When m is a decreasing function of concentration, the CL is defined as the concentration that interpolates (i.e. is backfitted from) the left hand end point of a one sided $(1 - \alpha)100\%$ prediction interval for $\bar{y}$ when $x = 0$.

The lack of a pivotal quantity or other formal procedure renders calculation of this prediction interval and subsequently of the MDC impossible. Approximations must be used. The usual procedure (Rodbard (1978) and Davidian et al. (1988)) is to suppose that

$$\frac{\bar{y} - m(x, \hat{\beta})}{\sqrt{v(m(x, \hat{\beta}), \hat{\theta})/r + \widehat{\text{var}}(m(x, \hat{\beta}))}} \tag{5.1}$$

has a Students t distribution with degrees of freedom based on the amount of information used to estimate θ. This approximation stems from the underlying normality of the observations and the fact that $\bar{y}$ and $m(x, \hat{\beta})$ (the fitted mean response at concentration x) are independent. When θ is estimated using all of the observations, the degrees of freedom are given by $\nu = \sum_{i=1}^{n}(r_i - 1) + n^s - p$. Instead, if only the residuals from the regression of the mean responses on the mean function were used to estimate θ, as in Rodbard (1978) and Davidian et al. (1988), then $\nu = n^s - p$.

The variance of $m(x, \hat{\beta})$ is given by $\widehat{\text{var}}(m(x, \hat{\beta}))$. This is usually approximated by:

$$\widehat{\text{var}}(m(x, \hat{\beta})) \approx m_\beta(x, \hat{\beta})' \text{var}(\hat{\beta}) m_\beta(x, \hat{\beta}), \tag{5.2}$$

where

$$\text{var}(\hat{\beta}) \approx \frac{1}{n - p} \left\{ \sum_{i \in S} \frac{r_i m_\beta(X_i, \hat{\beta}) m_\beta(X_i, \hat{\beta})'}{v(m(X_i, \hat{\beta}), \hat{\theta})} \right\}^{-1}. \tag{5.3}$$

Recall from Section 2.2 that m_β is the gradient of m with respect to β. The rhs of (5.2) is the variance of the first order Taylor series expansion of the non-linear function $m(x, \beta)$. The rhs of (5.3) is the asymptotic variance of $\hat{\beta}$ when the dependence of v on β is ignored and $\hat{\beta}$ is substituted for β. If the dependence of v on β is not ignored, the analogous information matrix based approximation for $\text{var}(\hat{\beta})$ is

$$\text{var}(\hat{\beta}) \simeq \frac{1}{\nu} \left\{ \sum_{i \in S} \left(1 + \frac{v_m(m(X_i, \hat{\beta}), \hat{\theta})^2}{2v(m(X_i, \hat{\beta}), \hat{\theta})} \right) \frac{r_i m_\beta(X_i, \hat{\beta}) m_\beta(X_i, \hat{\beta})'}{v(m(X_i, \hat{\beta}), \hat{\theta})} \right\}^{-1}, \qquad (5.4)$$

where v_m denotes the derivative of v with respect to m. When v is $O(m^{1/2})$ or greater (5.4) should be used instead of (5.3). In an immunoassay this is usually not the case (see Sadler and Smith (1986)) and (5.3) is used.

The measure of MDC that results when the above approximations are used, denoted by x^{cl}, is obtained by solving the equation

$$m(x^{cl}, \hat{\beta}) = m(0, \hat{\beta}) - t_{(\alpha, \nu)} \sqrt{v(m(0, \hat{\beta}), \hat{\theta})/r + \widehat{\text{var}}(m(0, \hat{\beta}))}, \qquad (5.5)$$

where $t_{(\alpha, \nu)}$ is the $1 - \alpha$ percentile of the Students t distribution with ν degrees of freedom. If m were an increasing function of concentration, then x^{cl} would backfit the upper end point of a one-sided prediction interval for $\bar{y}$ and the terms on the rhs of (5.5) would be added together.

Irrespective of whether m is increasing or decreasing, the monotone nature of the mean function ensures that when out of range responses are backfitted to whichever of 0 or ∞ is appropriate, $[0, x^{cl}]$ is the equivalent prediction interval for the backfitted concentration, $\hat{\omega} = m^{-1}(\bar{y}, \hat{\beta})$, when $x = 0$.

The CL is the concentration above which detection is said to occur. In this scenario a reliable measurement is deemed to be a response measurement associated with a backfitted concentration that exceeds x^{cl}. When m is a decreasing function of concentration, this will be the case if and only if the response is below the lower limit point of the prediction interval for $\bar{y}$ when $x = 0$. If the measurement is reliable then concentration is said to be detected. This criterion for detection approximates the rejection region of the α level test of the hypotheses $H_0 : \omega = 0$ verse $H_1 : \omega > 0$ that would be carried out if $\bar{y}$ were actually observed.

In Rodbard's terms, x^{cl} is the concentration "which would have an expected response statistically significantly different from the fitted response for zero dose

(concentration)". Statistical significance is specified by a one sided level α Students t-test for the difference of means. The hypotheses are $H_0 : E[\bar{y} \mid x^{cl}] = m(0, \beta)$ and $H_1 : E[\bar{y} \mid x^{cl}] < m(0, \beta)$.

If the t distribution with ν degrees of freedom were the exact distribution of (5.1), then the type I error probability for detection under both of the preceding criteria would be

$$
\begin{aligned}
pr\left(\bar{y} \;\leq\; m(0, \hat{\beta}) - t_{(\alpha,\nu)}\sqrt{v(m(0, \hat{\beta}), \hat{\theta})/r + \widehat{\mathrm{var}}(m(0, \hat{\beta}))} \mid x = 0, X\right) \\
= \; \mathrm{pr}(m^{-1}(\bar{y}, \hat{\beta}) \geq x^{cl} \mid x = 0, X) \\
= \; \alpha.
\end{aligned}
$$

The ease of calculation makes the CL a widely used measure of MDC (Brown et al. 1996). The major shortcoming of the CL is that the term "reliable measurement" is defined solely in terms of the uncertainty at zero concentration. No account is taken of the likelihood of detecting a positive concentration. In fact, it would be expected that about 50% of samples with concentrations equal to the CL would be detected.

5.2.2 Detection limit (DL)

This measure considers the uncertainty at positive concentrations by comparing responses at positive concentrations to the CL. The DL, like the CL, originates from Currie (1968). It is defined as the concentration for which the upper (lower) limit of a $1 - \gamma$ one-sided prediction interval for $\bar{y}$ equals $m(x^{cl}, \hat{\beta})$, in the case that m is a decreasing (increasing) function of concentration.

The value of the DL is denoted by x^{dl}, where x^{dl} solves

$$
m(x^{dl}, \hat{\beta}) = m(x^{cl}, \hat{\beta}) \overset{(+)}{-} t_{(\gamma,\nu)}\sqrt{v(m(x^{dl}, \hat{\beta}), \hat{\theta})/r + \widehat{\mathrm{var}}(m(x^{dl}, \hat{\beta}))}, \tag{5.6}
$$

when m is a decreasing (increasing) function of concentration.

Even when (5.1) holds, the DL is often incorrectly said to be the concentration at which the type II error probability for the hypothesis test related to the CL is equal to γ. The type II error probability of this test at x^{dl} is

$$
\mathrm{pr}\left(\frac{\bar{y} - m(0, \hat{\beta})}{\sqrt{v(m(0, \hat{\beta}), \hat{\theta})/r + \widehat{\mathrm{var}}(m(0, \hat{\beta}))}} > -t_{(\alpha,\nu)} \mid x = x^{dl}, X\right).
$$

This cannot be evaluated without assuming specific values for β and θ. Even then, the exact distribution of the quotient is complicated. Among well known families of distributions, a non-central t distribution may be the closest approximation.

In the following it is shown that $\mathrm{pr}(\bar{y} > \mu^c \mid x = x^{dl})$ is in general not equal to γ in the case that (5.1) is true[1]. Let $\mu^c = m(0,\hat{\beta}) - t_{(\alpha,\nu)}\sqrt{v(m(0,\hat{\beta}),\hat{\theta})/r + \widehat{\mathrm{var}}(m(0,\hat{\beta}))}$. Then

$$\mathrm{pr}(\bar{y} > \mu^c \mid x = x^{dl}) = \mathrm{pr}\left(\frac{\bar{y} - m(x^{dl},\hat{\beta})}{\sqrt{v(m(x^{dl},\hat{\beta}),\hat{\theta})/r + \widehat{\mathrm{var}}(m(x^{dl},\hat{\beta}))}} > \right.$$
$$\left. \frac{\mu^c - m(x^{dl},\hat{\beta})}{\sqrt{v(m(x^{dl},\hat{\beta}),\hat{\theta})/r + \widehat{\mathrm{var}}(m(x^{dl},\hat{\beta}))}} \mid x = x^{dl}, X\right).$$

If (5.1) holds, it is clear that the probability is equal to γ if and only if

$$t_{\gamma,\nu} = \frac{\mu^c - m(x^{dl},\hat{\beta})}{\sqrt{v(m(x^{dl},\hat{\beta}),\hat{\theta})/r + \widehat{\mathrm{var}}(m(x^{dl},\hat{\beta}))}}.$$

Since the right hand term is a function of random quantities when m is nonlinear in β or v is a function of m this is certainly not true in general.

Hence, x^{dl} is merely a concentration at which the type II error of the related hypothesis test is approximately γ. There is no guarantee that the approximation will be particularly good.

Historically the DL has been viewed as the concentration at which detection can be "expected". The built in allowance for type I error makes it accountable for both error types (although the exact levels of the type I and II error probabilities are unknown) and uncertainty at both zero and positive concentrations. Currie (1968) states that "the DL is an a priori measure of an assay's reliability. It reflects an assay's ability to detect a concentration when a sensible allowance has been made to ensure reliable non-detection." It is the limit of reliable or expected detection. On the other hand, the role of the CL is to classify a sample as detected or non-detected after a response measurement has been observed. It is in this sense that it is referred to as an a posteriori measure.

Within the frequentist paradigm the DL is one of the most reasonable measures of the MDC. However, although the DL accounts for uncertainty in the response

---

[1]When (5.1) is not true the type I error probability does not in general equal α; hence, there is no point in even trying to prove that the type II error probability is equal to γ.

at both zero and positive concentrations, it does not do so simultaneously (the uncertainty at zero concentration is firstly considered in the type I error probability calculation, then the uncertainty at positive concentrations is considered in the type II error probability calculation). Furthermore, the rather complicated nature of its definition makes it seem improbable that a practitioner would instinctively associated the DL with the MDC of an assay. In Sections 5.3.2 and 5.3.3 measures of MDC will be developed that simultaneously account for the uncertainty in the measurement of both zero and nonzero concentrations and which have definitions that are easy to understand.

5.2.3   Estimated limit (EL)

For the case in which m is a decreasing function of concentration, Davidian et al. (1988) defines the MDC as the smallest value of x for which

$$\mathrm{pr}(\bar{y} < m(0, \beta) \mid x) \geq 1 - \alpha. \tag{5.7}$$

This measure has a realistic and understandable foundation. Consider firstly the case in which β and θ are known, i.e. everything is known about the assay. The maximum likelihood estimate of the unknown concentration is zero if the observed mean response exceeds $m(0, \beta)$. This is because such a response does not backfit to a positive concentration. A concentration can thus be said to be "reliably able to be measured" if the likelihood of backfitting to a positive concentration is sufficiently high.

Since (5.7) can only be evaluated when β and θ are known (i.e. when everything is known about the assay's properties) this measure is referred to as the MDC in the presence of perfect information (x^{pi}). If detection is the event that the mean response backfits to a positive concentration then x^{pi} is the concentration at which detection occurs with probability $1 - \alpha$.

In a realistic situation, β and θ are unknown and thus x^{pi} must be generalized. The uncertainty in the assay model must be accounted for if the measure of the MDC is to reflect the quality of the assay. Davidian et al. (1988) suggested the smallest value of x for which

$$\mathrm{pr}(\bar{y} < m(0, \hat{\beta}) \mid x, X) \geq 1 - \alpha. \tag{5.8}$$

This measure of the MDC is denoted x^{el}. It is interesting to note that Davidian et al. (1988) implies that x^{el} is an estimator of x^{pi}. It will be shown in Section 5.4.2 that x^{el} is a terrible estimator of x^{pi}.

In general x^{el} cannot be evaluated exactly and must be approximated. The procedure used by Davidian et al. (1988) is to solve the following equation for x^{el}:

$$m(x^{el}, \hat{\beta}) = m(0, \hat{\beta}) - t_{(\alpha,\nu)} \sqrt{v(m(x^{el}, \hat{\beta}), \hat{\theta})/r + \widehat{var}(m(0, \hat{\beta}))}.$$

To obtain the above equation it is necessary to apply (5.1) and to act as if $m(x, \hat{\beta}) = m(0, \hat{\beta})$. Only after this equation has been found is this "temporary" restriction removed.

The number of approximations used in evaluating x^{el} suggests that improvements are possible. The "temporary" restriction is particularly alarming since, in the context of the measure of MDC being developed, it does not make any sense to assume that $m(x, \hat{\beta}) = m(0, \hat{\beta})$. Two alternative frequentist measures are developed in the remainder of this section, followed by a Bayesian measure in Section 5.3.1. The proximity of all of these measures to x^{pi} is investigated in Section 5.4.2.

5.2.3.1 Direct substitution approach

For comparison with other methods a measure of the MDC based on a sensible estimate of x^{pi} is considered. This is not a serious measure of the MDC but rather is provided because it is a sensible estimator of x^{pi} and so is useful for comparative purposes.

An obvious estimator of x^{pi} is found by substituting β and θ with $\hat{\beta}$ and $\hat{\theta}$ in the definition of x^{pi}. The resulting concentration, denoted x^{es}, is given by

$$m(x^{es}, \hat{\beta}) = m(0, \hat{\beta}) - z_\alpha \sqrt{v(m(x^{es}, \hat{\beta}), \hat{\theta})/r},$$

where z_α is the $1 - \alpha$ percentile of the standard normal distribution.

5.2.3.2 Prediction interval approach

In this approach, the measure of MDC is based on an approximate $1 - \alpha$ prediction interval for $\bar{y} - m(0, \beta)$. The procedure is to approximate (5.8) without pretending that $m(x, \hat{\beta}) = m(0, \hat{\beta})$ at any point of the calculation.

The value of x is sought for which the interval $(-\infty, 0)$ would be a 95% prediction interval for $\bar{y} - m(0, \beta)$ when $E[m(x, \hat{\beta})] = m(x, \beta)$. Firstly, note that

$$\bar{y} - m(0, \beta) \sim N(m(x, \beta) - m(0, \beta), v(m(x, \beta), \theta)/r).$$

Then, as $\bar{y}$ is independent of $\hat{\beta}$, it follows that

$$\bar{y} - m(0, \beta) - (m(x, \hat{\beta}) - m(0, \hat{\beta})) \sim N(0, v(m(x, \beta), \theta)/r + \text{var}(m(x, \hat{\beta}) - m(0, \hat{\beta}))).$$

However, the variance of the above distribution is dependent on the unknown parameter θ so inference is based on the approximate pivotal quantity

$$\frac{\bar{y} - m(0, \beta) - (m(x, \hat{\beta}) - m(0, \hat{\beta}))}{\sqrt{v(m(x, \hat{\beta}), \hat{\theta})/r + \widehat{\text{var}}(m(x, \hat{\beta}) - m(0, \hat{\beta}))}}.$$

This quantity, like the quantity in (5.1), is distributed approximately according to a Students t distribution with ν degrees of freedom. It thus follows that

$$
\begin{aligned}
\text{pr}\,\{\bar{y} - m(0, \beta) \quad &< \quad m(x, \hat{\beta}) - m(0, \hat{\beta}) + \\
&\quad t_{(\alpha, \nu)}\sqrt{v(m(x, \hat{\beta}), \hat{\theta})/r + \widehat{\text{var}}(m(x, \hat{\beta}) - m(0, \hat{\beta}))}\,\} \\
&\simeq \quad 1 - \alpha.
\end{aligned}
$$

Since the upper end point of the required prediction interval for $\bar{y} - f(0, \beta)$ is 0, it is therefore reasonable to define the MDC by solving

$$m(x, \hat{\beta}) - m(0, \hat{\beta}) + t_{(\alpha, \nu)}\sqrt{v(m(x, \hat{\beta}), \hat{\theta})/r + \widehat{\text{var}}(m(x, \hat{\beta}) - m(0, \hat{\beta}))} = 0.$$

for x. The solution is denoted x^{ei}. Thus, x^{ei} solves

$$m(x^{ei}, \hat{\beta}) = m(0, \hat{\beta}) - t_{(\alpha, \nu)}\sqrt{v(m(x^{ei}, \hat{\beta}), \hat{\theta})/r + \widehat{\text{var}}(m(x^{ei}, \hat{\beta}) - m(0, \hat{\beta}))}.$$

In the above expression the term $\widehat{\text{var}}(m(x^{ei}, \hat{\beta}) - m(0, \hat{\beta}))$ can be approximated by

$$(m_\beta(x^{ei}, \hat{\beta}) - m_\beta(0, \hat{\beta}))'\text{var}(\hat{\beta})(m_\beta(x^{ei}, \hat{\beta}) - m_\beta(0, \hat{\beta}))$$

and $\text{var}(\hat{\beta})$ may be evaluated using (5.3), (5.4) or some other approximation.

The prediction interval version of the EL is the concentration x^{ei} for which zero is the upper end point of a one sided $1 - \alpha$ prediction interval for the difference between an independent mean response based on r observations at x^{ei} and the mean response

at zero concentration. Under repeated sampling $1-\alpha$ of such intervals would contain the true difference between $\bar{y}$ and $m(0, \beta)$. In this way, x^{ei} may be interpreted as the concentration for which the uncertainty associated with an independent response measurement is significantly segregated from the uncertainty in the mean response at zero concentration.

Since $\widehat{\text{var}}(m(x^{ei}, \hat{\beta}) - m(0, \hat{\beta}))$ and $\widehat{\text{var}}(m(0, \hat{\beta}))$ may be quite different in value, x^{el} and x^{ei} may also be very different. It is unclear which of x^{el} and x^{ei} is superior. Simulations have indicated that x^{el} appears better for some values of the model parameters while for others x^{ei} seems to do a better job (see Tables 5.1 - 5.6).

Like x^{el}, x^{ei} is an extension of x^{pi} to the case in which β and θ are unknown. It can be seen that the solution for x^{ei} given above reduces to x^{pi} when the assay model is known. This follows by substituting $\hat{\beta}$ and $\hat{\theta}$ with β and θ respectively and setting $\widehat{\text{var}}(m(x, \hat{\beta}) - m(0, \hat{\beta})) = 0$.

The EL is based on an event which would be evaluated if the assay parameters were known. In this regard, the EL relates to what one really wants to know. This philosophy is reflected in the measures of MDC developed in the remainder of this chapter.

5.3 Bayesian measures of MDC

In this section measures of MDC are developed that use the notion of Bayesian inference. In the Bayesian paradigm, fitting a model consists of obtaining the posterior distribution of the unknown parameters. The posterior distribution quantifies the uncertainty in the underlying model at the completion of the assay. Inferences are derived from these distributions and so are conditional on the observed data. This differs from the frequentist approach where inferences are based on the sampling distributions of point estimates of the model parameters. These inferences are made in reference to all the data that could possibly have been observed in the experiment as opposed to just the data that were observed.

Recall that the superscripts S and U indicate that the entity is calculated using just the standards or the unknowns respectively. The development in this section is almost exclusively presented using the general assay model. The likelihood function

of the observed data is given by

$$f(Y \mid X, \eta^U, \beta, \theta) = \prod_{i \in S} f(Y_i \mid X_i, \beta, \theta) \prod_{i \in U} f(Y_i \mid \eta_i, \beta, \theta)$$

and the prior distribution of the assay parameters by

$$\pi(\beta, \theta, \eta^U) = \pi_m(\beta, \theta) \prod_{i \in U} \pi_0(\eta_i).$$

Throughout this section it is assumed that $\pi(\beta, \theta \mid Y, X)$ is determined at least to the extent that representative samples (a sample that reflects the population) are able to be drawn. The computational feasibility of any quantity involving $\pi(\beta, \theta \mid Y, X)$ is subject to efficient generation of a representative sample from $\pi(\beta, \theta \mid Y, X)$.

After the observed data has been analyzed, all of the information known about the assay and the underlying data generating process is contained in the following three distributions:

- $\pi(\beta, \theta \mid Y, X)$, the joint posterior distribution of β and θ. This contains all of the experimental information about the assay model; it may be appropriately termed the model fit.

- The prior (or marginal) distribution of the level of concentration that prevails throughout the population of unknowns. Recall that ω denotes a general value of concentration and $\pi_0(\omega)$ is the prior distribution for ω.

- $f(y \mid \omega, \beta, \theta)$, the distribution function of y, a vector of independent (i.e. hypothetical) responses at concentration ω.

Any definition of MDC from a Bayesian framework must be constructed entirely from these three distributions. Furthermore, if the MDC is to be used to assess the "quality" of the assay just performed information from measurements made on the unseen sample cannot be used to update the posterior distribution of β and θ. Several Bayesian definitions of MDC are now developed.

5.3.1   Response level MDC (RL MDC)

In the first instance the logical Bayesian generalization of Davidian's MDC is considered. This is to compute $\text{pr}(\bar{y} < m(0, \beta) \mid x)$ as a Bayesian predictive probability.

The resulting measure of the MDC is denoted by x^{rl}. Formally, x^{rl} is defined as the smallest value of x for which

$$\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X) \geq 1 - \alpha, \tag{5.9}$$

where $\bar{y}$ denotes the mean of y, a vector of r hypothetical responses. Thus x^{rl} is the smallest concentration for which the predictive probability of the event $(\bar{y} < m(0, \beta))$ is at least $1 - \alpha$. This probability is of course conditioned on the data (Y, X) of the current assay.

The lhs of (5.9) is now expanded to reveal the procedure for evaluating x^{rl}. By definition of the Bayesian predictive distribution it follows that

$$
\begin{aligned}
\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X) &= \int_{\mathcal{R}(\beta)} \int_0^{m(0,\beta)} \pi(\bar{y}, \beta \mid x, Y, X)\, d\bar{y}\, d\beta \\
&= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \int_0^{m(0,\beta)} \pi(\bar{y}, \beta, \theta \mid x, Y, X)\, d\bar{y}\, d\beta\, d\theta \\
&= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \int_0^{m(0,\beta)} f(\bar{y} \mid x, Y, X, \beta, \theta) \pi(\beta, \theta \mid x, Y, X) \\
&\qquad\qquad .d\bar{y}\, d\beta\, d\theta.
\end{aligned}
$$

Since $\bar{y}$ given (x, β, θ) is independent of Y and X and (β, θ) given (Y, X) does not depend on the concentration x it follows that

$$f(\bar{y} \mid x, Y, X, \beta, \theta) = f(\bar{y} \mid x, \beta, \theta)$$

and

$$\pi(\beta, \theta \mid x, Y, X) = \pi(\beta, \theta \mid Y, X).$$

Hence,

$$
\begin{aligned}
\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X) &= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \int_0^{m(0,\beta)} f(\bar{y} \mid x, \beta, \theta)\, d\bar{y}\, \pi(\beta, \theta \mid Y, X)\, d\beta\, d\theta \\
&= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \text{pr}(\bar{y} < m(0, \beta) \mid x, \beta, \theta) \pi(\beta, \theta \mid Y, X)\, d\beta\, d\theta
\end{aligned}
$$

$$\tag{5.10}$$

The above derivation indicates that the predictive probability is an average with respect to the uncertainty (or error) in the values of β and θ. The rhs of (5.10) is a weighted sum of the conditional probabilities, $\text{pr}(\bar{y} < m(0, \beta) \mid x, \beta, \theta)$, for each

value of β and θ. It can be seen from this expression that all the uncertainty in the parameters has been simultaneously considered.

It is easy to see that $x^{rl} = x^{pi}$ if β and θ were known. This follows by observing that given β and θ, the data (Y, X) has no influence; hence, $\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X) = \text{pr}(\bar{y} < m(0, \beta) \mid x)$.

If normality can be assumed, i.e. $\bar{y} \sim N(m(x, \beta), v(m(x, \beta), \theta)/r)$, $\text{pr}(\bar{y} < m(0, \beta) \mid x, \beta, \theta)$ is equal to the cumulative distribution function (cdf) of the standard normal distribution evaluated at

$$\frac{m(0, \beta) - m(x, \beta)}{\sqrt{v(m(x, \beta), \theta)/r}}.$$

Hence, in this case

$$\text{pr}(\bar{y} < m(0, \beta) \mid Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \Phi\left(\frac{m(0, \beta) - m(x, \beta)}{\sqrt{v(m(x, \beta), \theta)/r}}\right) \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta.$$

An interesting feature of this measure of MDC is that the desired predictive probability involves a model parameter. Bayesian inferences usually pertain separately to functions of parameters or functions of future responses. This makes the RL MDC somewhat unique. See Section 8.6.2 for more comments.

5.3.1.1   Numerical evaluation

In general, the calculation of x^{rl} relies on the calculation of $\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X)$ as given in equation (5.10) above. The strategy is to generate $(y, \beta) \sim \pi(y, \beta \mid x, Y, X)$ and then to calculate $\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X)$ as a Monte-Carlo average of $z = I(\bar{y} < m(0, \beta))$. By the strong law of large numbers this empirical average converges to its expectation which is precisely $\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X)$ (see Smith and Roberts (1993), Besag et al. (1995) for the general result).

The crucial step is the generation of (y, β) from $\pi(y, \beta \mid x, Y, X)$. To accomplish this firstly write

$$\pi(y, \beta \mid x, Y, X) = \int_{\mathcal{R}(\theta)} f(y \mid x, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, d\theta$$

and note that $y \sim f(y \mid x, \beta, \theta)$ can be easily generated. Thus it remains to generate a representative sample of (β, θ) from $\pi(\beta, \theta \mid Y, X)$. As discussed in Section 4.4,

the Metropolis-Hastings algorithm is a suitable method for generating β and θ from their joint posterior distribution.

The algorithm given below embodies the calculation of $\text{pr}(\bar{y} < m(0, \beta) \mid x, Y, X)$ into a M-H scheme for generating $(\beta, \theta) \sim \pi(\beta, \theta \mid Y, X)$. This is not the only method which could be used but is one that is simple to implement and efficient to run. The algorithm can be easily adjusted to be compatible with other methods of generating posterior variates of β and θ.

1. Run the M-H algorithm on β and θ until the burn-in period is over.

2. Generate $(\beta^{(i)}, \theta^{(i)}) \sim \pi(\beta, \theta \mid Y, X)$ using the M-H algorithm.

3. Calculate $m(0, \beta^{(i)})$.

4. Generate an r dimensional response vector $y^{(ij)} \sim f(y \mid x, \beta^{(i)}, \theta^{(i)})$. Let $\bar{y}^{(ij)}$ be the sample mean of $y^{(ij)}$.

5. Record $z^{(ij)} = I(\bar{y}^{(ij)} \geq m(0, \beta^{(i)}))$.

6. Repeat steps 4-5 K times.

7. Let $p^{(i)} = \frac{1}{K} \sum_{j=1}^{K} z^{(ij)}$.

8. Repeat steps 2-7 until, in addition to the satisfaction of all convergence criteria of the M-H chain, the overall mean of the $p^{(i)}$ has converged.

In the case of normally distributed responses, steps 4 and 5 of the algorithm may be circumvented with direct evaluation of the cumulative probability using a function that closely approximates Φ. Adapting the approximation of the error function in Cody (1969) results in an excellent approximation of Φ. Other good approximating functions are given on pages 932-3 of Abramowitz and Stegun (1970).

The RL MDC is the value of x for which the proportion in step 5 converges to $1 - \alpha$. This is found by simultaneously carrying out steps 2 to 8 for a range of concentrations judiciously chosen so as to include x^{rl}. The extreme values of x should give rise to proportions on either side of $1 - \alpha$. If this is not the case, the range of x needs to be shifted. When the proportions encompass $1 - \alpha$, an appropriate curve can be fitted to the proportions and x^{rl} estimated via interpolation at $1 - \alpha$. Alternatively, the algorithm can be re-run with the concentrations more condensed thus providing more concise information from which to interpolate x^{rl}.

5.3.1.2   Optimal value of K

The algorithm calculates binary variables $z^{(ij)} = I(\bar{y}^{(ij)} < m(0, \beta^{(i)}))$ corresponding to the jth of K values of $\bar{y}$ generated using the ith generated values of β and θ. Note that because $\bar{y}^{(ij)}$ is different for each $z^{(ij)}$ these binary variables are not Bernoulli random variables even if they were independent. Stratum are defined by the binary variables arising from the same value of β and θ. By the strong law of large numbers for any fixed value of K (strata size) the marginal mean of these indicator variables converges to the probability being sought. It makes sense to choose K so that convergence occurs in the smallest possible computer (cpu) time. This is equivalent to determining K such that the variance of the marginal mean (proportion) is minimized subject to a finite amount of cpu time. There are two aspects to consider here. The statistical aspect is that the information contained in two binary variables within the same stratum is less than that contained in two binary variables from different strata. However, the between strata generation (of β and θ) is more computer intensive than the within strata generation (of $\bar{y}$).

Consider the following simplified representation. Let $z^{(ij)}$ be such that $E[z^{(ij)} \mid \mu_i] = \mu_i$ and $\text{var}[z^{(ij)} \mid \mu_i] = \sigma^2$ where $E[\mu_i] = \mu$ and $\text{var}[\mu_i] = \tau^2$. In the interest of simplicity the fact is ignored that because $z^{(ij)}$ is binary, σ^2 and τ^2 are likely to be functions of μ_i and μ respectively. Suppose further that both the generation of $z^{(ij)}$ given μ_i and the generation of μ_i are independent processes. The strata means $p^{(i)} = \sum_{j=1}^{K} z^{(ij)}/K$ are then independent random variables with variance given by

$$
\begin{aligned}
V[p^{(i)}] &= E[V[p^{(i)} \mid \mu_i]] + V[E[p^{(i)} \mid \mu_i]] \\
&= E[\sigma^2/K] + V[\mu_i] \\
&= \sigma^2/K + \tau^2.
\end{aligned}
$$

The variance of the overall (marginal) proportion $\bar{p} = \sum_{i=1}^{n} p^{(i)}/n$ is therefore given by

$$
V(\bar{p}) = \frac{\sigma^2/K + \tau^2}{n}.
$$

Let t_I and t_J represent the respective average cpu times required to generate each μ_i and $z^{(ij)} \mid \mu_i$. The cpu time taken to generate n clusters with K replicates is thus $n(t_I + Kt_J)$. The problem is to find the values of n and K which minimize $V(\bar{p})$

subject to the constraint $n(t_I + Kt_J) \leq T$. A simple calculation yields the solution

$$n = \frac{T}{t_I + Kt_J} \text{ and } K = \sqrt{\frac{\sigma^2 t_I}{\tau^2 t_J}}.$$

Of course K and n must be rounded to integer values in the set of natural numbers.

Clearly, it is only worthwhile to sample multiple $z^{(ij)}$ at each value of μ_i if the variance-cpu time ratio for $z^{(ij)}$ exceeds the variance-cpu time ratio for the generation of the μ_i.

Even though this is a simplified representation, some insight is gained into the optimal value for K in step 6 of the algorithm for calculating x^{rl}. It can be construed that the optimal value of K in the calculation of x^{rl} will increase if

1. the within strata variation caused by the variability of $\bar{y}^{(ij)}$ increases relative to the between stratum variation caused by the variability of $(\beta^{(i)}, \theta^{(i)})$;

2. the ratio of the CPU time taken to generate $(\beta^{(i)}, \theta^{(i)})$ increases relative to the CPU time required to generate $\bar{y}^{(ij)}$.

Since the generation time for $(\beta^{(i)}, \theta^{(i)})$ is much greater than for $y^{(ij)}$, the optimal value of K is likely to be quite large. In other words it is beneficial to estimate the conditional probability at each value of $(\beta^{(i)}, \theta^{(i)})$ with a high degree of accuracy. Informal experiments have indicated that $K \in [10, 100]$ works quite well. The effect of moderate to high serial correlation in the values of $\beta^{(i)}$ and $\theta^{(i)})$ remains an unresolved issue.

5.3.2 Probability MDC (Pr MDC)

The event of interest in the preceding section, $(\bar{y} - m(0, \beta) < 0)$, was motivated by a frequentist perspective rather than from a Bayesian perspective. In this section a measure of MDC based on how a Bayesian practitioner could be expected to act is developed.

Suppose for a moment that an unknown sample, not already analyzed by the assay, existed. Let y denote the observed response for this sample and ω the (unknown) concentration. Under a Bayesian framework, inferences about ω would be based upon $\pi(\omega \mid y, Y, X)$, the posterior distribution of the unknown concentration given y and (Y, X) and an initial prior for the assay parameters. As alluded to in

Section 4.1.2.1, in the assays where the MDC will be a concern it only makes sense to start with a prior that has some positive probability for zero concentrations.

If the mean response exceeded $m(0, \beta)$, the estimate of concentration would not necessarily be zero. Instead, it would be determined from $\text{pr}(\omega = 0 \mid y, Y, X)$, the posterior probability that the concentration is zero, and the posterior distribution over concentrations close to zero. A measure of the MDC based on $\text{pr}(\omega > 0 \mid y, Y, X) = 1 - \text{pr}(\omega = 0 \mid y, Y, X)$ is now developed.

5.3.2.1 Definition of detection

As $\text{pr}(\omega > 0 \mid y, Y, X)$ is the quantity of interest it seems reasonable that the concentration associated with the response y would be declared as detected (greater than zero) if

$$\text{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha$$

for some small probability α. That is, detection is defined to be the event $(\text{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha)$.

5.3.2.2 Definition of reliable measurement

Note that the event "detection" depends on the response y. But for a given concentration, the response is a random variable and hence "detection" may or may not occur depending on the value of y. For a given concentration x, such dependence can be removed by using the predictive distribution of y given (x, Y, X) to determine the probability of detection.[2] If this predictive probability were sufficiently large, the concentration x would be said to be reliably able to be measured or detected. Accordingly the concentration x is defined as reliably able to be measured if the probability of detection is greater than or equal to $1 - \gamma$, where $1 - \gamma$ is a "sufficiently high" probability level.

A convenient way of expressing the mathematical definition of "reliable detection" is the following. Let $\mathcal{Y}_C$ be the set of responses for which concentration would be detected given the data (Y, X). That is

$$\mathcal{Y}_C = \{y : \text{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha\}.$$

---

[2] If such a calculation were being performed in a frequentist framework the probability would be evaluated by averaging with respect to $f(y, Y \mid x, X)$.

A concentration x is then deemed to be reliably measured if the predictive probability of the set $\mathcal{Y}_C$ given (x, Y, X), $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$, is at least $1 - \gamma$.

The corresponding measure of MDC is x^{pr}, the smallest concentration for which the predictive probability of observing a response in the set $\mathcal{Y}_C$ thus detecting positive concentration, is at least $1 - \gamma$. This is the smallest value of x for which

$$\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X) \geq 1 - \gamma.$$

The quantity $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$ may be referred to as a predictive posterior probability since the quantity being predicted is based on a posterior probability.

5.3.2.3 Fine points

In the above definition the predictive distribution rids the MDC of any dependence on data, namely y, that is not a part of the assay data. This ensures that the MDC can be calculated in practice. However, there is one subtlety that must first be addressed. If y were actually observed, then $\mathrm{pr}(\omega > 0 \mid y, Y, X)$ would be calculated as

$$\begin{aligned}
\mathrm{pr}(\omega > 0 \mid y, Y, X) &= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > 0 \mid y, Y, X, \beta, \theta) \pi(\beta, \theta \mid y, Y, X) \, d\beta \, d\theta \\
&= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > 0 \mid y, \beta, \theta) \pi(\beta, \theta \mid y, Y, X) \, d\beta \, d\theta. \quad (5.11)
\end{aligned}$$

Since y is an additional response measurement, standard Bayesian analysis would automatically incorporate it into the posterior distribution of (β, θ). However, y has not been observed so it can be assigned any arbitrary value. The term $\mathrm{pr}(\omega > 0 \mid y, \beta, \theta)$ is effectively computing the probability that the unknown concentration is larger than zero as a function of y. In this sense y does not represent "data" and hence should not be used to compute the posterior distribution of (β, θ); that is $\pi(\beta, \theta \mid Y, X)$ should replace $\pi(\beta, \theta \mid y, Y, X)$ in (5.11); hence, in this section it is defined that

$$\mathrm{pr}(\omega > 0 \mid y, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > 0 \mid y, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta.$$

There is further justification for the amendment to (5.11) described above. In the defining equation of x^{pr} the predictive distribution of y is determined from $\pi(\beta, \theta \mid Y, X)$, the current state of knowledge about the assay. The same level of

knowledge should also be assumed in the calculation of $\text{pr}(\omega > 0 \mid y, Y, X)$; that is, the value of y should not be used to update the posterior distribution of (β, θ) or in other words, the model fit should be independent of y.

If $\pi(\beta, \theta \mid y, Y, X)$ were used, then a slightly optimistic picture of the assay would be portrayed as being a product of the model fit y reinforces it. The future quality of the assay would be predicted as opposed to the current quality of the assay being measured. The prediction of the state of the assay in the future is an extension of the ideas of Geisser (1992) and Geisser (1993) on Bayesian interim analysis. See Section 8.6.3 for more comments.

5.3.2.4 Formal definition of probability MDC

With the remarks in Section 5.3.2.3 in mind the probability MDC is now formally defined. The probability MDC, x^{pr}, is the smallest value of x for which

$$\text{pr}(\mathcal{Y}_C \mid x, Y, X) \geq 1 - \gamma$$

where

$$\text{pr}(\mathcal{Y}_C \mid x, Y, X) = \int_{\mathcal{R}(y)} I(y \in \mathcal{Y}_C) p(y \mid x, Y, X) \, dy \tag{5.12}$$

$$= \int_{\mathcal{R}(y)} I(\text{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha) p(y \mid x, Y, X) \, dy,$$

$$\text{pr}(\omega > 0 \mid y, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \text{pr}(\omega > 0 \mid y, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta, \tag{5.13}$$

and

$$p(y \mid x, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} f(y \mid x, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta.$$

5.3.2.5 MDC when β, θ are known

If β and θ are known, the data from the assay are not needed. The Pr MDC is the smallest concentration for which

$$\int_{\mathcal{R}(y)} I(\text{pr}(\omega > 0 \mid y, \beta, \theta) \geq 1 - \alpha) f(y \mid x, \beta, \theta) \, dy \geq 1 - \gamma.$$

Recall from Section 5.3.1 that the predictive probability involved in the RL MDC could be expressed as an average of itself conditional on (β, θ) and the model fit.

This is not the case for x^{pr} because (5.12) does not simplify to

$$\int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \int_{\mathcal{R}(y)} I(\text{pr}(\omega > 0 \mid y, \beta, \theta) \geq 1 - \alpha) f(y \mid x, \beta, \theta) \pi(\beta, \theta \mid Y, X)\, dy\, d\beta\, d\theta.$$
(5.14)

Furthermore, it seems unlikely that (5.12) would reduce to (5.14) in any circumstance. To illustrate the point let $h(x) = I(x \geq 1 - \alpha)$. Then

$$\begin{aligned}
I(\text{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha) &= h(\text{pr}(\omega > 0 \mid y, Y, X)) \\
&= h(E^{\pi(\beta, \theta \mid Y, X)}[\text{pr}(\omega > 0 \mid y, \beta, \theta)]).
\end{aligned}$$

A sufficient condition for equality between (5.12) and (5.14) is

$$h(E^{\pi(\beta, \theta \mid Y, X)}[\text{pr}(\omega > 0 \mid y, \beta, \theta)]) = E^{\pi(\beta, \theta \mid Y, X)}[h(\text{pr}(\omega > 0 \mid y, \beta, \theta))].$$

However, for $0 < \alpha < 1$ the lhs equals 0 or 1 with probability 1 but the rhs will almost certainly lie between 0 and 1. Hence, this sufficient condition is not attained. However, a necessary condition still remains to be found.

From the computational point of view it is disappointing that (5.12) and (5.14) are not equivalent. It is easier to evaluate (5.14) than (5.12) because the order of integration in (5.14) is opposite to the order in which β, θ and y are naturally generated. Refer to Section 5.3.2.8 for further discussion.

The Pr MDC has some similarity to the DL discussed in Section 5.2.2. A decision making interpretation of this measure is, "x^{pr} is the minimum concentration for which the probability of rejecting $H_0 : \omega = 0$ in favour of $H_1 : \omega > 0$ using the decision rule: reject H_0 if $(\text{pr}(H_1 \mid y, Y, X) \geq 1 - \alpha)$, is predicted with probability at least $1 - \gamma$". The Bayesian analogues to the type I and II error probabilities are α and γ respectively. A natural choice for α is 0.5 since under zero-one loss this is the Bayes action.

The calculation of x^{pr} emulates the inference about the unknown concentration which would be performed if y were actually observed. This makes it a suitable diagnostic for measuring the performance of the assay. Furthermore, the chance of it being understood and interpreted correctly by practitioners is likely to be higher than that for the preceding definitions.

The conversion of uncertainty in the response to uncertainty in the concentration through the use of Bayes theorem is an integral component of the calculation of x^{pr}.

Such a conversion is not seen in any of the preceding definitions. In a frequentist context such a conversion would be difficult because the sampling distribution of an estimator of concentration does not lend itself to explicit calculation.

5.3.2.6   Prior distributions for ω

To implement the Pr MDC, a prior distribution with non-zero probability at $\omega = 0$ must to be specified for the concentration associated with the independent response y. Over positive concentrations it is reasonable that the prior be continuous. Both requirements can be achieved using the following split prior distribution introduced in Section 4.1.2.1, i.e.

$$\pi_0(\omega) = p_0 I(\omega = 0) + (1 - p_0)\pi_0^u(\omega)I(\omega > 0), \qquad (5.15)$$

where $p_0 \in [0, 1]$. As discussed in Section 4.1.2.1, in many assays (especially those in which the occurrence of zero concentrations are routine) such a distribution is more realistic than a purely continuous distribution. Purely continuous prior distributions do not admit positive probability at a point, in particular $\omega = 0$. The main reason they are used is only that the model is easier to fit. For the split prior given in (5.15)

$$\mathrm{pr}(\omega > 0 \mid y, \beta, \theta) = \frac{(1 - p_0) \int_0^\infty f(y \mid \omega, \beta, \theta)\pi_0^u(\omega)\, d\omega}{p_0 f(y \mid 0, \beta, \theta) + (1 - p_0) \int_0^\infty f(y \mid \omega, \beta, \theta)\pi_0^u(\omega)d\omega}. \qquad (5.16)$$

It is easy to see that calculation of (5.16) and hence (5.12) is routine, provided of course that p_0 and $\pi_0^u(\omega)$ have been specified.

5.3.2.7   Non-informative prior

It could be desired to evaluate the MDC without the addition of any further subjective input. However, it is not immediately clear how a priori ignorance of ω is represented in a split prior distribution. It seems reasonable that $\pi_0^u(\omega)$ be uniform so as to reflect impartiality over positive concentrations. It is tempting to assert that $p_0 = 0.5$ and $\pi_0^u(\omega)$ is uniform over some finite range, $(0, M)$ say. However, this imbalances the prior at zero concentration; the result being that as M increases posterior inferences about ω become increasingly tighter about zero concentration, a phenomena known as Lindley's paradox. A more reasonable representation is to

set $\pi_0^u(\omega) = U(0, M)$ where $U(a, b)$ is a uniform distribution over the interval (a, b) and $p_0 = 1/(M + 1)$. Then as $M \to \infty$

$$\text{pr}(\omega > 0 \mid y, \beta, \theta) \to \frac{\int_0^\infty f(y \mid \omega, \beta, \theta)\, d\omega}{f(y \mid 0, \beta, \theta) + \int_0^\infty f(y \mid \omega, \beta, \theta)\, d\omega}. \tag{5.17}$$

In (5.17) the posterior probability is calculated when the prior is a proper distribution (i.e. M is finite) before the limit as $M \to \infty$ is taken. Note that, in (5.15) $p_0 = 1/(M + 1) \to 0$ as $M \to \infty$. Hence, if the order of the calculation of the posterior probability and the evaluation of the limit in the derivation of (5.17) were reversed, the essential concept of a split prior would thus be nullified.

The rhs of (5.17) has the intuitive appeal of being the ratio of the likelihood of y when $\omega > 0$ to the total likelihood of y (the likelihood when $\omega = 0$ plus the likelihood when $\omega > 0$).

5.3.2.8 Numerical calculation

Recall that in the context of the Pr MDC a concentration x is "reliably" able to be measured if $\text{pr}(\mathcal{Y}_C \mid x, Y, X) \geq 1 - \gamma$, where $\mathcal{Y}_C = \{y : \text{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha\}$ and where $\text{pr}(\omega > 0 \mid y, Y, X)$ is calculated according to the amendment described in Section 5.3.2.3; that is,

$$\text{pr}(\omega > 0 \mid y, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \text{pr}(\omega > 0 \mid y, \beta, \theta) \pi(\beta, \theta \mid Y, X)\, d\beta\, d\theta,$$

as in (5.13).

To evaluate x^{pr} a method for the evaluation of the predictive probability $\text{pr}(\mathcal{Y}_C \mid x, Y, X)$ is required. It is seen from the preceding paragraph that this requires that:

1. One can generate $y \sim p(y \mid x, Y, X)$.

2. The probability $\text{pr}(\omega > 0 \mid y, \beta, \theta)$ can be evaluated.

The generation of $y \sim p(y \mid x, Y, X)$ is described in the calculation of x^{rl} (see Section 5.3.1). It is to be noted that the term, $\text{pr}(\omega > 0 \mid y, \beta, \theta)$, depends on just a one dimensional integral in ω and so can be easily evaluated using some numerical means of integration (the trapezium rule is used for calculations reported in this thesis).

The probability $\mathrm{pr}(\omega > 0 \mid y, Y, X)$ can be evaluated as the empirical average of $\mathrm{pr}(\omega > 0 \mid y, \beta^{(j)}, \theta^{(j)})$ where $\beta^{(j)}$ and $\theta^{(j)}$ are drawn from $\pi(\beta, \theta \mid Y, X)$ using the M-H algorithm. Thus for each y the value of the indicator function $I(\mathrm{pr}(\omega > 0 \mid y, Y, X) \geq 1 - \alpha)$ is able to be evaluated. By generating many values of y from $p(y \mid x, Y, X)$ and taking the average of the values of the indicator function, the desired quantity, $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$, is obtained.

It is important to note that in the evaluation of $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$ two independent sequences of values of $(\beta, \theta) \sim \pi(\beta, \theta \mid Y, X)$ are required. This is because for evaluation $\mathrm{pr}(\omega > 0 \mid y, Y, X)$ and $p(y \mid x, Y, X)$ are decomposed into expressions involving $\pi(\beta, \theta \mid Y, X)$. The first set of values are used to generate the predictive responses $y \sim p(y \mid x, Y, X)$. The second set of values are used to evaluate $\mathrm{pr}(\omega > 0 \mid y, Y, X)$ for each such value of y. If the same set of values were used for both calculations the full uncertainty in ω would not be captured. This is apparant since the value of y being conditioned on in the evaluation of $\mathrm{pr}(\omega > 0 \mid y, \beta_1, \theta_1)$ would be generated from the distribution $f(y \mid x, \beta_1, \theta_1)$ as opposed to $p(y \mid x, Y, X)$.

The strategy for calculating $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$ is thus:

1. Generate $(\beta^1, \theta^1) \sim \pi(\beta, \theta \mid Y, X)$.

2. Generate $y \sim f(y \mid x, \beta^1, \theta^1)$.

3. Generate $(\beta^2, \theta^2) \sim \pi(\beta, \theta \mid Y, X)$ where (β^1, θ^1) is independent of (β^2, θ^2).

4. Calculate $\mathrm{pr}(\omega > 0 \mid y, \beta^2, \theta^2)$.

5. Repeat steps 1-4 until the empirical mean of $z = I(\mathrm{pr}(\omega > 0 \mid y, \beta^2, \theta^2) \geq \alpha)$ has converged to its limiting value.

Since z is a binary variable with marginal mean $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$ these steps need only be iterated as many times as are needed for the empirical mean of z to estimate $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$ to the required accuracy.

As the M-H algorithm does not provide an independent sequence of values of $(\beta, \theta) \sim \pi(\beta, \theta \mid Y, X)$, independent M-H chains are needed to generate (β^1, θ^1) and (β^2, θ^2) in the above. An outline of the algorithm for calculating $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$ follows:

1. Run two independent M-H simulations on $\pi(\beta, \theta \mid Y, X)$ until each chain reaches a steady state (i.e. passes through the transient phase).

2. Generate $y \sim p(y \mid x, Y, X)$.

   (a) Generate $(\beta^{(i)}, \theta^{(i)}) \sim \pi(\beta, \theta \mid Y, X)$ as described for the RL MDC.

   (b) Generate an r dimensional response vector $y^{(i)} \sim f(y \mid x, \beta^{(i)}, \theta^{(i)})$.

3. Generate z for which $E[z] = \mathrm{pr}(\mathcal{Y}_C \mid x, Y, X)$.

   (a) Generate $(\beta^{(j)}, \theta^{(j)}) \sim \pi(\beta, \theta \mid Y, X)$.

   (b) Use the Trapezium method (the approach taken in this thesis) or some other numerical method to evaluate the one dimensional integral

   $$L_A = \int_0^\infty f(y^{(i)} \mid x, \beta^{(j)}, \theta^{(j)}) \pi_0^u(\omega) \, d\omega.$$

   Then calculate the probability

   $$p^{(ij)} = \mathrm{pr}(\omega > 0 \mid y^{(i)}, \beta^{(j)}, \theta^{(j)}) = \frac{(1 - p_0) L_A}{p_0 L_0 + (1 - p_0) L_A},$$

   where $L_0 = f(y^{(i)} \mid 0, \beta^{(j)}, \theta^{(j)})$.

   (c) Record $z^{(ij)} = I(p^{(ij)} \geq \alpha)$.

4. Repeat step 3 K times.

5. Repeat steps 2-4 until all convergence diagnostics of the MCMC chain are satisfied and the overall mean of the $z^{(ij)}$ has converged.

To determine x^{pr} this algorithm can be used within a scheme like that described for x^{rl}. The probabilities for a range of values of x are evaluated simultaneously. A smooth curve is fitted to the empirical probabilities and x^{pr} is found by projecting the point where $\mathrm{pr}(\mathcal{Y}_C \mid x, Y, X) = 1 - \gamma$ onto the x axis.

Setting K at a value other than 1 may improve the running time of the algorithm. Step 3 makes the calculation involving each value of y more time consuming than for x^{rl}. For this reason it is probably better to use a smaller value of K than the value that is optimal for the calculation of x^{rl}.

5.3.3   Discriminant MDC (Ds MDC)

In this section another definition of a reliable measurement of concentration is introduced. This is done in terms of the likelihood of correctly ordering two samples by their levels of concentration when one sample has zero concentration (unbeknown to the practitioner). In a Bayesian framework the decision on which sample has the greater concentration would be based on the joint posterior distribution for the concentration in each sample given the observed responses. A valid decision criterion is to select the sample for which the posterior probability that the level of concentration it contains is greater than that of the other sample exceeds 0.5.

5.3.3.1   Preliminary definitions

The probability MDC can be adapted to yield a measure of the MDC based on the scenario detailed above. This measure is known as the discriminant MDC and will be denoted by x^{ds}.

Let y and y_0 represent response measurements from two samples with unknown concentration. Detection of the concentration in one sample as distinct from the concentration in the other sample is thus said to occur if $\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) \geq 1 - \alpha$ with the posterior distribution $\pi(\omega, \omega_0 \mid y, y_0, Y, X)$ being used to make the computation. Detection thus depends on the values y and y_0; the set of responses for which detection occurs is defined to be

$$\mathcal{Y}_D = \{(y, y_0) : \mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) \geq 1 - \alpha\}.$$

The predictive probability of $\mathcal{Y}_D$ can be computed at any values of concentration of the hypothetical responses, y and y_0, using the same method as was used for $\mathcal{Y}_C$. For the discriminant MDC the concentration generating y_0 is, of course, set at 0.

A concentration x is said to be reliably measurable if the predictive probability of $\mathcal{Y}_D$ is at least $1 - \gamma$ assuming that y is a response at concentration x and y_0 is a response at zero concentration. In this case the MDC, x^{ds}, is defined to be the smallest of such x values; i.e. the smallest value of x for which:

$$\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X) \geq 1 - \gamma. \tag{5.18}$$

Before proceeding the sequence of steps is reviewed. The set $\mathcal{Y}_D$ has been defined in the space of vector pairs of the form (y, y_0). It is simply a set with a certain

property, namely that the probability $\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X)$ is at least $1 - \alpha$. The predictive probability of this set is then computed at a pair of concentrations x and 0. If this probability is at least $1 - \gamma$, it is said that x has been reliably measured. The smallest of these reliably measurable concentrations is defined to be x^{ds}.

For computational purposes it is helpful to expand the lhs of (5.18) as follows:

$$\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$$
$$= \int_{\mathcal{R}(y_0)} \int_{\mathcal{R}(y)} I((y, y_0) \in \mathcal{Y}_D) p(y, y_0 \mid x, 0, Y, X) \, dy \, dy_0$$
$$= \int_{\mathcal{R}(y_0)} \int_{\mathcal{R}(y)} I(\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) \geq 1 - \alpha) p(y, y_0 \mid x, 0, Y, X) \, dy \, dy_0.$$

The term $\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X)$ requires some attention. In the first instance this can be written

$$\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) \tag{5.19}$$
$$= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X, \beta, \theta) \pi(\beta, \theta \mid y, y_0, Y, X) d\beta d\theta$$
$$= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > \omega_0 \mid y, y_0, \beta, \theta) \pi(\beta, \theta \mid y, y_0, Y, X) d\beta d\theta.$$

$$\tag{5.20}$$

The term $\mathrm{pr}(\omega > \omega_0 \mid y, y_0, \beta, \theta)$ is simply computing the probability that one unknown concentration is larger than another as a function of unknown responses, y and y_0, to which values could be assigned. As noted in Section 5.3.2.3 such responses are in this sense not data and should not be used to compute the term $\pi(\beta, \theta \mid y, y_0, Y, X)$. Thus $\pi(\beta, \theta \mid Y, X)$ is used instead of $\pi(\beta, \theta \mid y, y_0, Y, X)$ in (5.20); hence, in this section it is defined that

$$\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) = \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > \omega_0 \mid y, y_0, \beta, \theta) \pi(\beta, \theta \mid Y, X) d\beta d\theta.$$

5.3.3.2 Formal definition of discriminatory MDC

The discriminatory MDC, x^{pr}, is thus formally defined as the smallest value of x for which

$$\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X) \geq 1 - \gamma,$$

where

$$\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X) = \tag{5.21}$$

$$\int_{\mathcal{R}(y_0)} \int_{\mathcal{R}(y)} I(\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) \geq 1 - \alpha) p(y, y_0 \mid x, 0, Y, X) \, dy \, dy_0,$$

$$\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) = \int_{\mathcal{R}(\beta)} \mathrm{pr}(\omega > \omega_0 \mid y, y_0, \beta, \theta) \pi(\beta, \theta \mid Y, X) d\beta d\theta \quad (5.22)$$

and

$$p(y, y_0 \mid x, 0, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} f(y \mid x, \beta, \theta) f(y_0 \mid 0, \beta, \theta) \pi(\beta, \theta \mid Y, X) d\beta d\theta.$$

The Ds MDC is interpreted as the smallest concentration at which the predictive probability of rejecting the hypothesis $H_0 : \omega \leq \omega_0$ in favour of $H_1 : \omega > \omega_0$ using the decision rule: reject the null if $\mathrm{pr}(H_1) \geq 1 - \alpha$, is at least $1 - \gamma$. The probability $1 - \gamma$ is interpretated as the minimum acceptable probability of correctly ordering a sample with zero concentration and a sample with positive concentration.

If the measurement error at zero concentration is neglected; that is, if ω_0 is replaced with the fixed point 0, x^{ds} reduces to x^{pr}. Due to the extra component of variation it is clear that $x^{ds} \geq x^{pr}$.

The Ds MDC is an extended version of the Pr MDC. It is constructed from the same basis and with the same formulation. There are numerous variants of the probability and discriminant MDCs that could also be developed. The Bayesian paradigm provides the freedom to develop measures of MDC that are directly relevant to a particular situation and are also able to be realized (can be evaluated without the need for mathematical approximations). The frequentist paradigm is not nearly as amenable.

A prior distribution specifying non-zero probability at zero concentration is not required for x^{ds} to make sense. However, if samples with zero concentration can occur this possibility should be reflected in the prior. If (4.4) is used as the prior distribution then

$$\mathrm{pr}(\omega > \omega_0 \mid y, y_0, Y, X) =$$
$$\mathrm{pr}(\omega > 0, \omega_0 = 0 \mid y, y_0, Y, X) + \mathrm{pr}(\omega > 0, \omega_0 > 0 \mid y, y_0, Y, X)$$
$$\cdot \int_0^\infty \int_{\omega_0}^\infty \pi(\omega, \omega_0 \mid \omega > 0, \omega_0 > 0, y, y_0 Y, X) \, d\omega_0 \, d\omega.$$

The integral in the above expression is the probability that would be obtained if the prior distribution were purely continuous. This fact is made use of in the calculation of x^{ds}. It is important to note that when this prior is used it matters whether $\omega > \omega_0$ or $\omega \geq \omega_0$ in the definition of x^{ds}. If the latter were the case, the first term in the summation becomes $\mathrm{pr}(\omega_0 = 0 \mid y_0, Y, X)$.

5.3.3.3 Numerical calculation

To evaluate x^{ds} a calculation much like that for x^{pr} is required. A sample of $z = I(\text{pr}(\omega > \omega_0 \mid y, y_0, Y, X) > 1 - \alpha)$ are generated. Since $E[z] = \text{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$, $\text{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$ can be computed as the long run empirical average of M-H draws of z. The final step is to search over a range of values of x until the MDC is found.

It can be seen from (5.21) that a sequence $(y^{(i)}, y_0^{(i)})$, $i = 1, 2, \ldots$ must be generated from $p(y, y_0 \mid x, 0, Y, X)$. This is the first step in the computation. As

$$p(y, y_0 \mid x, 0, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} f(y, y_0 \mid x, 0, Y, X, \beta, \theta) \pi(\beta, \theta \mid x, 0, Y, X) d\beta d\theta$$

$$= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} f(y \mid x, \beta, \theta) f(y_0 \mid 0, \beta, \theta) \pi(\beta, \theta \mid Y, X) d\beta d\theta,$$

a sequence of $(y^{(i)}, y_0^{(i)})$ can be obtained by repeating the following steps:

1. Generate $(\beta^{(i)}, \theta^{(i)}) \sim \pi(\beta, \theta \mid Y, X)$.

2. Generate $y^{(i)} \sim f(y \mid x, \beta^{(i)}, \theta^{(i)})$ and $y_0^{(i)} \sim f(y_0 \mid 0, \beta^{(i)}, \theta^{(i)})$.

It must then be determined which of these $(y^{(i)}, y_0^{(i)})$'s belong to the given set $\mathcal{Y}_D$. This requires that the probability $\text{pr}(\omega > \omega_0 \mid y, y_0, Y, X)$ is evaluated. If for the particular pair $(y^{(i)}, y_0^{(i)})$, the value of this probability is at least $1 - \alpha$ it is recorded that the pair belongs to $\mathcal{Y}_D$. Note that this step "acts as though" this particular pair was generated at "unknown" concentrations ω and ω_0 respectively, which of course they were not.

Thus the second computational step is to compute $\text{pr}(\omega > \omega_0 \mid y^{(i)}, y_0^{(i)}, Y, X)$. This can be done by generating a sequence of $(\omega^{(ij)}, \omega_0^{(ij)}) \sim \pi(\omega, \omega_0 \mid y^{(i)}, y_0^{(i)}, Y, X)$ $j = 1, 2, \ldots$ for each pair $(y^{(i)}, y_0^{(i)})$ and computing the proportion of times that the event $(\omega^{(ij)} > \omega_0^{(ij)})$ is true. Alternatively, the probability can be evaluated directly using a two dimensional version of the Trapezium method or some other method from numerical analysis. The former approach was used in this thesis.

To generate $(\omega^{(ij)}, \omega_0^{(ij)}) \sim \pi(\omega, \omega_0 \mid y^{(i)}, y_0^{(i)}, Y, X)$ the fact is used that

$$\pi(\omega, \omega_0 \mid y, y_0, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \pi(\omega, \omega_0 \mid y, y_0, \beta, \theta) \pi(\beta, \theta \mid Y, X) d\beta d\theta$$

$$= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \pi(\omega \mid y, \beta, \theta) \pi(\omega_0 \mid y_0, \beta, \theta) \pi(\beta, \theta \mid Y, X) d\beta d\theta,$$

the latter step following from the fact that given (β, θ), ω and ω_0 are independent. Note that in line with the definition of $\pi(\omega, \omega_0 \mid y, y_0, Y, X)$ being used in this section, the posterior distribution for (β, θ) does not depend on (y, y_0).

It is clear from the above decomposition of $\pi(\omega, \omega_0 \mid y, y_0, Y, X)$ that a sequence of $(\omega^{(ij)}, \omega_0^{(ij)}) \sim \pi(\omega, \omega_0 \mid y^{(i)}, y_0^{(i)}, Y, X)$ can be generated by repeated application of the following steps:

1. Generate $(\beta^{(j)}, \theta^{(j)}) \sim \pi(\beta, \theta \mid Y, X)$.

2. Generate $\omega^{(ij)} \sim \pi(\omega \mid y^{(i)}, \beta^{(j)}, \theta^{(j)})$ and $\omega_0^{(ij)} \sim \pi(\omega \mid y_0^{(i)}, \beta^{(j)}, \theta^{(j)})$.

It is crucial that the sequence $(\beta^{(j)}, \theta^{(j)})$ is independent of the sequence $(\beta^{(i)}, \theta^{(i)})$ used for generating $(y^{(i)}, y_0^{(i)}) \sim p(y, y_0 \mid x, 0, Y, X)$. As noted in Section 5.3.2.8 failure to observe this rule would lead to the full uncertainty in (ω, ω_0) not being accounted for.

As $\pi(\omega \mid y, \beta, \theta)$ and $\pi(\omega_0 \mid y_0, \beta, \theta)$ are only known up to proportionality, numerical methods must be relied upon to generate samples from them. Accept-reject importance sampling and the M-H algorithm are two options with importance sampling probably the preferred choice as the densities are univariate.

To recapitulate, the strategy for the calculation of $\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$ is:

1. Generate $(\beta^1, \theta^1) \sim \pi(\beta, \theta \mid Y, X)$.

2. Generate $y \sim f(y \mid x, \beta^1, \theta^1)$ and $y_0 \sim f(y_0 \mid 0, \beta^1, \theta^1)$.

3. Generate $(\beta^2, \theta^2) \sim \pi(\beta, \theta \mid Y, X)$ where (β^1, θ^1) is independent of (β^2, θ^2).

4. Generate $\omega \sim \pi(\omega \mid y, \beta^2, \theta^2)$ and $\omega_0 \sim \pi(\omega \mid y_0, \beta^2, \theta^2)$.

5. Calculate $\mathrm{pr}(\omega > \omega_0 \mid y, y_0, \beta^2, \theta^2)$.

6. Calculate $z = I(\mathrm{pr}(\omega > \omega_0 \mid y, y_0, \beta^2, \theta^2) > 1 - \alpha)$ and repeat steps 1-5 until the empirical mean of the values of z has converged to its limiting value, $\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$.

A skeleton algorithm for the calculation of $\mathrm{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$ is given below:

1. Run two independent M-H simulations on $\pi(\beta, \theta \mid Y, X)$ until each chain reaches a steady state (i.e. the transient phase is over).

2. Generate $(y, y_0) \sim p(y, y_0 \mid x, 0, Y, X)$.

   (a) Generate $(\beta^{(i)}, \theta^{(i)}) \sim \pi(\beta, \theta \mid Y, X)$ using the M-H algorithm.

   (b) Generate r dimensional response vectors $y^{(i)} \sim f(y \mid x, \beta^{(i)}, \theta^{(i)})$ and $y_0^{(i)} \sim f(y \mid 0, \beta^{(i)}, \theta^{(i)})$.

3. Generate z for which $E[z] = \text{pr}(\mathcal{Y}_D \mid x, 0, Y, X)$.

   (a) Set $k = 1$.

   (b) Generate $(\beta^{(j)}, \theta^{(j)}) \sim \pi(\beta, \theta \mid Y, X)$.

   (c) Use the M-H algorithm (this was used in this thesis), importance sampling or some other method to generate $\omega^{(ijk)} \sim \pi(\omega \mid y^{(i)}, \beta^{(j)}, \theta^{(j)})$ and $\omega_0^{(ijk)} \sim \pi(\omega \mid y_0^{(i)}, \beta^{(j)}, \theta^{(j)})$.

   (d) Record $t^{(k)} = I(\omega^{(ijk)} > \omega_0^{(ijk)})$

   (e) Repeat steps 3b and 3c until the mean of the sequence $\{t^{(1)}, t^{(2)}, \ldots\}$ has converged to its limiting value, denoted $p^{(ij)}$.

   (f) Record $z^{(ij)} = I(p^{(ij)} \geq 1 - \alpha)$.

4. Repeat steps 2-4 until all convergence diagnostics of the MCMC chain are satisfied and the grand mean of the $z^{(ij)}$ has converged to its limiting value.

When $p_0 > 0$ in $\pi_0(\omega)$, it might be more efficient to proceed as if, for the most part, $p_0 = 0$. To cater for the fact that $p_0 = 0$, all that needs to be done is to update $p^{(ij)}$ to

$$\text{pr}(\omega > 0 \mid y^{(i)}, \beta^{(j)}, \theta^{(j)})[\text{pr}(\omega_0 = 0 \mid y_0^{(i)}, \beta^{(j)}, \theta^{(j)}) + p^{(ij)}.\text{pr}(\omega_0 > 0 \mid y_0^{(i)}, \beta^{(j)}, \theta^{(j)})]$$

in between steps 3e and 3f. The probabilities on the rhs of the above equation can be calculated at each iteration by the method described for the calculation of x^{pr}.

The generation of posterior samples of ω and ω_0 in step 3b of the algorithm makes the CPU time needed to calculate x^{ds} greater than for x^{rl} or x^{pr}. An optimal value for K has not been determined.

5.3.4   Browns unified measure of MDC

The final measure of MDC to be studied is Brown's unified measure (Brown et al. (1996)). This has purposely not been introduced until this point because it warrants comparison with x^{ds}.

Brown's unified MDC is denoted by x^{un}. It is calculated as follows:

1. Using the data from the assay and a reliable means of estimation (frequentist or Bayesian) obtain point estimates, $\hat{\beta}$ and $\hat{\theta}$, of β and θ.

2. Find all r dimensional response vectors y for which

$$\int_0^\infty \int_{\omega_{(0)}}^\infty \pi(\omega \mid y, \hat{\beta}, \hat{\theta})\pi(\omega_0 \mid Y_0, \hat{\beta}, \hat{\theta})\, d\omega\, d\omega_0 = 1 - \alpha, \qquad (5.23)$$

   where Y_0 is the vector of responses associated with the zero standard and $\pi(\omega_0 \mid y, Y_0, \hat{\beta}, \hat{\theta})$ is the posterior distribution of ω_0 for the response Y_0, conditional on $\hat{\beta}$ and $\hat{\theta}$ being the values of the model parameters. $\pi(\omega \mid y, \hat{\beta}, \hat{\theta})$ is similarly interpreted in terms of the response y.

3. For each response vector y satisfying (5.23) calculate the median of $\pi(\omega \mid y, \hat{\beta}, \hat{\theta})$. The maximum medium is x^{un}.

In step 2 of the calculation Brown recommends using a prior distribution for ω that is uniform over the range of 0 to 1.5 times the maximum concentration of the standards.

If the response measurements are singletons there will be only one solution to (5.23). When the responses are replicated and the sufficient statistic for these observations has dimension greater than 1, the possibility of more than one solution arises. The normal distribution is one such case.

The major differences between x^{un} and x^{ds} are:

1. The uncertainty in β and θ is not accounted for in x^{un} whereas it is accounted for in x^{ds}.

2. The response at zero concentration in x^{ds} is an independent measurement. Although Y_0 is not independent of the assay data it is used as the response at zero concentration in x^{un}.

3. The posterior distributions of the unknown concentrations are not averaged with respect to the distributions of the responses they are conditioned on when calculating x^{un}. All hypothetical response measurements are averaged over in the calculation of x^{ds}.

These are all deficiences with x^{un}. The fact that x^{un} treats the estimates of β and θ as the true parameter values means that it is insensitive to the design of the assay. No account is taken of the uncertainty in β and θ and so x^{un} will be an overly optimistic measure, particularly when the number of standards is small.

The response Y_0 is used twice; in the first instance to estimate β and θ and secondly as data that is conditioned on to develop the posterior of ω_0. This makes the procedure ad-hoc and x^{un} devoid of any statistical properties.

By not averaging over the distribution of y, the estimating equation is left as a function of hypothetical data. A secondary criterion, in this case the posterior median, must then be used to convert these responses into concentrations. The fact that many, perhaps infinite, values of y may satisfy (5.23) makes x^{un} more ad-hoc and unreliable. The rule of thumb that x^{un} be the maximum possible posterior median is likely to associate the MDC with a hypothetical response that is unusual and extreme.

In summary x^{un}, is difficult to interpret and appears to take little account of the quality of the assay. The name "unified definition" assigned to this measure in Brown et al. (1996) is misleading. Other measures of MDC are only unified in the sense that they are approximated by features such as percentiles of $\pi(\omega \mid y, \hat{\beta}, \hat{\theta})$ (see Brown et al. (1996)). The measures of MDC developed in this chapter are not encompassed in this way by x^{un}.

5.4 Some numerical properties of the measures of MDC

The measures of MDC that have been discussed are now compared using simulated data based on the NMCH data. The behaviour of the measures of MDC when the quality of the assay is changed is of particular interest. To study this, the sampling distribution of the measures of MDC are calculated using simulated data based on

the NMCH data and model.

The non-informative prior $\pi(\beta, \theta, \mu^U) \propto 1/\tilde{\theta}_1$, where $\tilde{\theta}_1 = \theta_1 \dot{\mu}^{\theta_2}$ and $\dot{\mu}$ is the geometric mean of μ, is used for Bayesian model fitting and (4.4), the non-informative split prior, is used in the calculation of x^{pr} and x^{ds}.

5.4.1   Comparative performance

Since the MDC varies according to which definition of a reliable measurement is used, it is not possible to say which measure of MDC is better in an absolute sense. Unlike standard estimation problems, a "true value" to which each measure can be compared does not exist. In essence, different measures of MDC relate to different quantities. The difference in the nature of these quantities needs to be taken account of before the measures of MDC can be compared.

Imagine a situation in which two assays have been performed and one wishes to determine which is the better assay. The MDC is promoted as a diagnostic of the quality of an assay and the assay with the smaller value of the MDC is deemed to be the better assay.

If the two assays were repeated many times, sampling distributions of measures of MDC could be obtained. It is intuitively clear that a large segregation of these sampling distributions implies a large difference in the quality of the assays. In this thesis the "relative quality (RQ)" of two assays for a particular measure of MDC is defined as the probability that a randomly drawn value from the sampling distribution of the MDC for one assay exceeds a randomly drawn value from the sampling distribution of the MDC for the other assay. A probability close to 0.5 indicates that there is little difference between the quality of the assays. If the probability is close to 1 or 0 then one assay is much better than the other.

The RQ of two assays as defined above is a common language effect size (CLES) statistic in the sense of McGraw and Wong (1992); a CLES statistic is in general a probability that a value from one population is less than (or greater than depending on the definition) a value drawn from another population. A value drawn from the sampling distribution of a measure of MDC for a particular assay is referred to as a score for that assay.

When the ordering of the quality of two or more assays is known, the RQ of any two assays can be used to compare different measures of MDC and to identify

the best one. In general, the best measure of MDC is the one which gives the most accurate assessment of the comparative quality for two assays. In the following discussion it will be illustrated that the RQ is not necessarily a true indicator of the comparative quality of two assays. In situations where the RQ is a true measure of the comparative quality of two assays, the best measure of MDC has the highest RQ value and so is easy to identify. Such situations are now described.

The quality of an assay is influenced by the values of β, θ, X, η^U and r. The RQ for two assays is, however, a well defined measure of the performance of the MDC only in the situations in which the disparity between the assays is in the assays' data generating processes; i.e. the value of β and θ. This follows from the fact that the other entities are either known at the time of analysis or information is known about them (for instance, n^u is known in the case of η^U) and this knowledge is such that it can be used to "rig" the performance of the MDC in terms of the RQ criterion. For example, consider two assays that are identical except that in one, assay responses are measured in duplicate and in the other responses are measured in quadruplicate. Then the measure of MDC defined to equal the inverse of the total number of observations will identify the better assay with probability 1; hence RQ= 1. Realistically though, this ridiculous measure of the MDC is of no practical value.

The measures of MDC considered in this chapter are in general much more practical than the one discussed above. However, approximations are used in the evaluation of the frequentist measures which may result in these measures being over-sensitive to changes in the values of X, n^u and r; hence, a distorted picture of the comparative quality of the assays may be portrayed. In the case in which assays differ in the value of X, n^u or r, identifying the best measure of MDC becomes a matter of common sense.

When the RQ is a well defined measure of the performance of the MDC, one only needs to know which assay is better in order to be able to be able to identify the best measure of MDC. There are certain factors which when altered the quality of the modified assay can be assessed as better or worse. For example:

1. Increasing either component of θ in the power variance function increases the variance of the responses and so the quality of the assay is reduced.

2. Increasing either the number of standards, unknowns or the degree of replication improves the quality of the assay.

The consequences of changes in β, X and η^U are not in general clear so such changes are not considered herein. Recall from the above that only in first instance is the RQ is a a well defined measure of the performance of the MDC.

In Section 5.4.1.2 the measures of MDC are evaluated in the controlled environment in which only θ varies. The behaviour of the measures of MDC when n^s, n^u and r change is then investigated.

5.4.1.1   Calculation

Suppose that two assays, A and B, have been defined by specifying the values of X, r and (β, θ, η^U) for each of them. The RQ for assays A and B is then calculated as follows:

1. For each assay generate simulated sets of responses and then the corresponding MDC scores. Denote the ith scores for the assays A and B by s_i^A and s_i^B respectively.

2. Calculate

$$\hat{p} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} I(s_i^A < s_j^B).$$

This is the proportion of times a score for assay A is less than a score for assay B.

Alternatively, the calculation in step 2 can be performed by summing the ranks of the scores as in the calculation of the Mann-Whitney U statistic (Randles and Wolfe (1979)).

Unless n is extremely large, the estimation error of $\hat{p}$ will be non-trivial and should be reported. The variance of $\hat{p}$, which can be derived from the variance of the Mann-Whitney U statistic (Randles and Wolfe (1979)), is a function of p. But in this case the sample sizes are equal and hence the variance is maximized at $p = 0.5$. Then

$$\text{var}(\hat{p} \mid p = 0.5) = \frac{2n + 1}{12n^2}.$$

Since many independent sets of simulated data are generated, the Central Limit Theorem implies that the distribution of $\hat{p}$ will be very closely approximated by a

normal distribution with mean p and variance less than var$(\hat{p} \mid p = 0.5)$. Thus, an upper bound on a $(1 - \alpha)100\%$ confidence margin (CM), or half-width of a $(1 - \alpha)100\%$ confidence interval, for p is almost certain to be

$$\frac{z_{\alpha/2}}{2n} \sqrt{(2n + 1)/3}.$$

5.4.1.2 Change in response variability

The effect of changes in the variability of the assay responses on the measures of MDC will be investigated for the power variance function having the following values of θ:

V1: $\theta = (0.5, 1.1)$.

V2: $\theta = (1, 1)$.

V3: $\theta = (1, 1.1)$.

V4: $\theta = (1, 1.2)$.

V5: $\theta = (2, 1.1)$.

Models V1 to V5 are otherwise identical, namely 75 unknowns are set at concentrations evenly spaced from 50 to 300, β is fixed at $(55, 1.2, 60, 4)'$ and the remaining components are as for the NMCH model.

The partial orders of quality for these assays are V1 > V3 > V5 and V2 > V3 > V4, where VA > VB indicates that assay A is better than assay B. The event (RQ> 0.5) can thus be predicted for the following pairs of assays:

For change in θ_1: V1 v. V3, V3 v. V5, V1 v. V5.

For change in θ_2: V2 v. V3, V3 v. V4, V2 v. V4.

For change in θ_1 and θ_2: V1 v. V4, V2 v. V5.

The RQ for each of the above comparisons are shown in Tables 5.1 to 5.3. The sample size and 95% confidence margins are also given. Note that RQ(A,B) is taken to mean the "estimated RQ of assays A and B"; i.e. the estimate of pr$(S^A > S^A)$ in which S^A denotes a randomly generated score (value of MDC) for assay A. Furthermore, in each instance A will denote the better assay. Thus, the value of RQ(A,B) is expected to exceed 0.5.

Some examples of the empirical distributions of the simulated values of MDC are shown in Figure 5.1. To make this figure more presentable the data has been fitted to a gamma distribution. The gamma distribution was used because this family

Table 5.1: Estimated values of the RQ for measures of MDC when θ_1 changes in the NMCH model.

MDC Procedure	n	95% CM	RQ(V1,V3)	RQ(V3,V5)	RQ(V1,V5)
Critical limit (x^{cl})	6403	0.01	0.7983	0.7899	0.9441
Detection limit (x^{dl})	6403	0.01	0.8280	0.8338	0.9694
Davidian's EL (x^{el})	6403	0.01	0.7992	0.7907	0.9447
Substitution EL (x^{es})	6403	0.01	0.7930	0.7815	0.9380
Prediction interval EL (x^{ei})	6403	0.01	0.8164	0.8211	0.9613
Response level MDC (x^{rl})	1028	0.025	0.7718	0.7785	0.9365
Probability MDC (x^{pr})	1028	0.025	0.8117	0.8098	0.9630
Discriminant MDC (x^{ds})	257	0.05	0.8248	0.8177	0.9690

Table 5.2: Estimated values of the RQ for measures of MDC when θ_2 changes in the NMCH model.

MDC Procedure	n	95% CM	RQ(V2,V3)	RQ(V3,V4)	RQ(V2,V4)
Critical limit (x^{cl})	6403	0.01	0.8845	0.8730	0.9873
Detection limit (x^{dl})	6403	0.01	0.9092	0.9140	0.9957
Davidian's EL (x^{el})	6403	0.01	0.8852	0.8734	0.9875
Substitution EL (x^{es})	6403	0.01	0.8793	0.8629	0.9844
Prediction interval EL (x^{ei})	6403	0.01	0.8985	0.9023	0.9931
Response level MDC (x^{rl})	1028	0.025	0.8689	0.8642	0.9898
Probability MDC (x^{pr})	1028	0.025	0.8974	0.8865	0.9916
Discriminant MDC (x^{ds})	257	0.05	0.8967	0.9038	0.9954

Table 5.3: Estimated values of the RQ for measures of MDC when both θ_1 and θ_2 change in the NMCH model.

MDC Procedure	n	95% CM	RQ(V1,V4)	RQ(V2,V5)
Critical limit (x^{cl})	6403	0.01	0.9720	0.9726
Detection limit (x^{dl})	6403	0.01	0.9880	0.9875
Davidian's EL (x^{el})	6403	0.01	0.9723	0.9729
Substitution EL (x^{es})	6403	0.01	0.9672	0.9683
Prediction interval EL (x^{ei})	6403	0.01	0.9833	0.9824
Response level MDC (x^{rl})	1028	0.025	0.9690	0.9736
Probability MDC (x^{pr})	1028	0.025	0.9810	0.9838
Discriminant MDC (x^{ds})	257	0.05	0.9893	0.9848
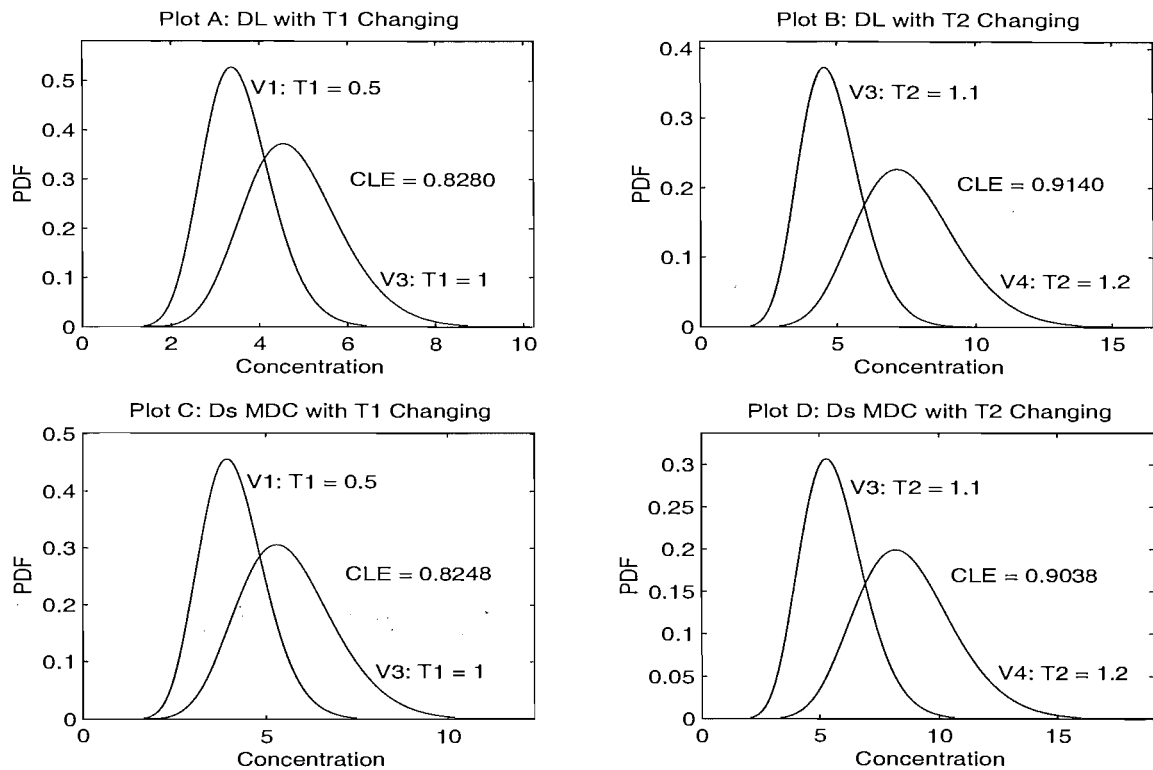
Figure 5.1: Some examples of the sampling distributions of measures of MDC when θ (denoted by T) changes in the NMCH model.

of distributions seemed to adequately describe the important features of the MDC sampling distributions.

5.4.1.3 Comments: changes in θ

There is very little difference in the RQ values between the measures of MDC that have been considered. One reason why the Bayesian measures did not perform better is that they could not be calculated as accurately as desired because it was not possible to run the simulation for the required length of time. Thus, the additional sensitivity gained from exact calculation may have been lost or at least compromised.

Despite the above comments, the Ds MDC still performed close to the level of the DL, which from inspecting Tables 5.1 to 5.3 can be seen to be the best performing measure of MDC. It can also be observed from inspecting these tables that the

prediction interval EL and the Pr MDC were the next highest scoring measures of MDC. A feature exclusive to x^{ds} and the DL is that they are based on measurements at both zero and positive concentrations. This suggests that it is advantageous to use measures of MDC that are based on the discrimination of samples at zero concentration from samples at positive concentrations when comparing assays for which the variability of the responses is different.

The values of RQ are only slightly lower for x^{es} than for other measures. In this case, incorporating the error due to estimation appears to have only a small effect on the comparative quality of assays having different values of θ.

5.4.1.4   Changes in the number of observations

The behaviour of the measures of MDC when the number of observations changes between assays is now investigated. To standardize the data generating process, the model parameters, β and θ, are fixed at $(55, 1.2, 60, 4)'$ and $(1, 1.1)'$ respectively. In cases where there are more than seven standards, the seven values of concentrations in the NMCH model are kept and the remaining values are evenly spaced from 50 to 250. The values of concentration of the unknowns are again evenly spaced from 50 to 300. The degree of replication is as indicated except for the zero standard for which there are always four replicates. The seven models to be used in this simulation are given below:

D1: $n^s = 7$, $n^u = 0$, $r = 2$.
D2: $n^s = 7$, $n^u = 75$, $r = 1$.
D3: $n^s = 7$, $n^u = 75$, $r = 2$.
D4: $n^s = 14$, $n^u = 75$, $r = 2$.
D5: $n^s = 28$, $n^u = 75$, $r = 2$.
D6: $n^s = 7$, $n^u = 75$, $r = 4$.
D7: $n^s = 7$, $n^u = 150$, $r = 2$.

These assays have the following partial orders of quality: D7 > D3 > D1, D5 > D4 > D3 and D6 > D3 > D2. The event (RQ> 0.5) can thus be predicted for the following pairs of assays:

For change in n^u: D1 v. D3, D3 v. D7, D1 v. D7.
For change in n^s: D3 v. D4, D4 v. D5, D3 v. D5.
For change in r: D2 v. D3, D3 v. D6, D2 v. D6.

Table 5.4: Estimated values of the RQ for measures of MDC when n^u changes in the NMCH model.

MDC Procedure	n	95% CM	RQ(D3,D1)	RQ(D7,D3)	RQ(D7,D1)
Critical limit (x^{cl})	6403	0.01	0.5200	0.4863	0.5114
Detection limit (x^{dl})	6403	0.01	0.5245	0.4867	0.5167
Davidian's EL (x^{el})	6403	0.01	0.5198	0.4862	0.5113
Substitution EL (x^{es})	6403	0.01	0.4643	0.4810	0.4484
Prediction interval EL (x^{ei})	6403	0.01	0.5236	0.4865	0.5154
Response level MDC (x^{rl})	1028	0.025	0.5831	0.5003	0.5891
Probability MDC (x^{pr})	1028	0.025	0.5988	0.5211	0.6187
Discriminant MDC (x^{ds})	257	0.05	0.6650	0.5808	0.7316

Table 5.5: Estimated values of the RQ for measures of MDC for the NMCH data when n^s changes in the NMCH model.

MDC Procedure	n	95% CM	RQ(D4,D3)	RQ(D5,D4)	RQ(D5,D3)
Critical limit (x^{cl})	6403	0.01	0.6077	0.5406	0.6489
Detection limit (x^{dl})	6403	0.01	0.6467	0.5497	0.6958
Davidian's EL (x^{el})	6403	0.01	0.6095	0.5413	0.6513
Substitution EL (x^{es})	6403	0.01	0.4839	0.4983	0.4814
Prediction interval EL (x^{ei})	6403	0.01	0.5412	0.5149	0.5570
Response level MDC (x^{rl})	1028	0.025	0.5267	0.4544	0.4817
Probability MDC (x^{pr})	1028	0.025	0.5291	0.5058	0.5365
Discriminant MDC (x^{ds})	257	0.05	0.5541	0.5388	0.5867

For change in n^s and n^u: D1 v. D4, D1 v. D5.

For change in n^s and r: D2 v. D3, D2 v. D4.

For change in n^u and r: D1 v. D6.

The estimated RQ for each pair of assays when just one of n^u, n^s and r change are given in Tables 5.4 - 5.6 along with 95% confidence margins. In RQ(A,B), A is once again the better assay; hence, the value of RQ(A,B) is expected to exceed 0.5. The sampling distributions of the MDC in some of these instances are also displayed. The results for comparisons between assays where two of n^u, n^s and r change are not reported.

Figure 5.2: Some examples of the sampling distributions of measures of MDC when n^u changes in the NMCH model.

Table 5.6: Estimated values of the RQ for measures of MDC when r changes in the NMCH model.

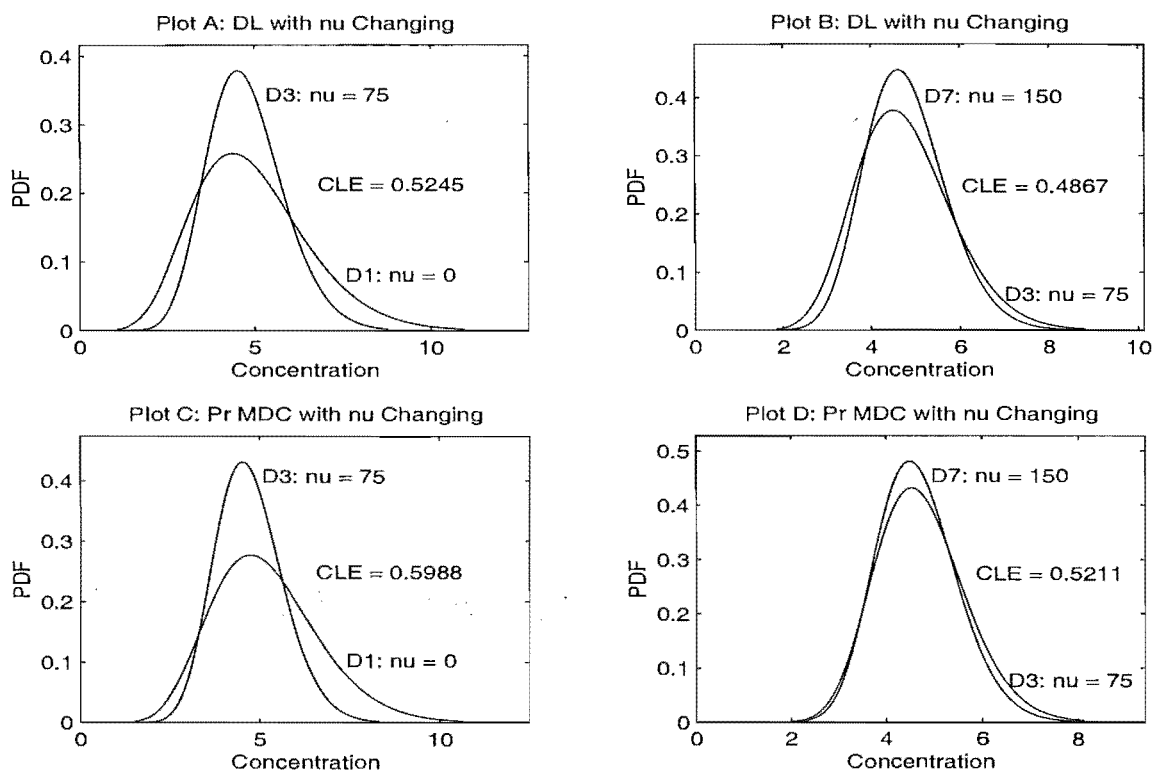| MDC Procedure	n	95% CM	RQ(D3,D2)	RQ(D6,D3)	RQ(D6,D2)
Critical limit (x^{cl})	6403	0.01	0.6759	0.6279	0.7516
Detection limit (x^{dl})	6403	0.01	0.7320	0.6963	0.8233
Davidian's EL (x^{el})	6403	0.01	0.6746	0.6264	0.7501
Substitution EL (x^{es})	6403	0.01	0.6720	0.8077	0.8368
Prediction interval EL (x^{ei})	6403	0.01	0.7860	0.8341	0.9111
Response level MDC (x^{rl})	1028	0.025	0.7882	0.8308	0.9211
Probability MDC (x^{pr})	1028	0.025	0.8165	0.8091	0.9381
Discriminant MDC (x^{ds})	257	0.05	0.7926	0.8561	0.9374
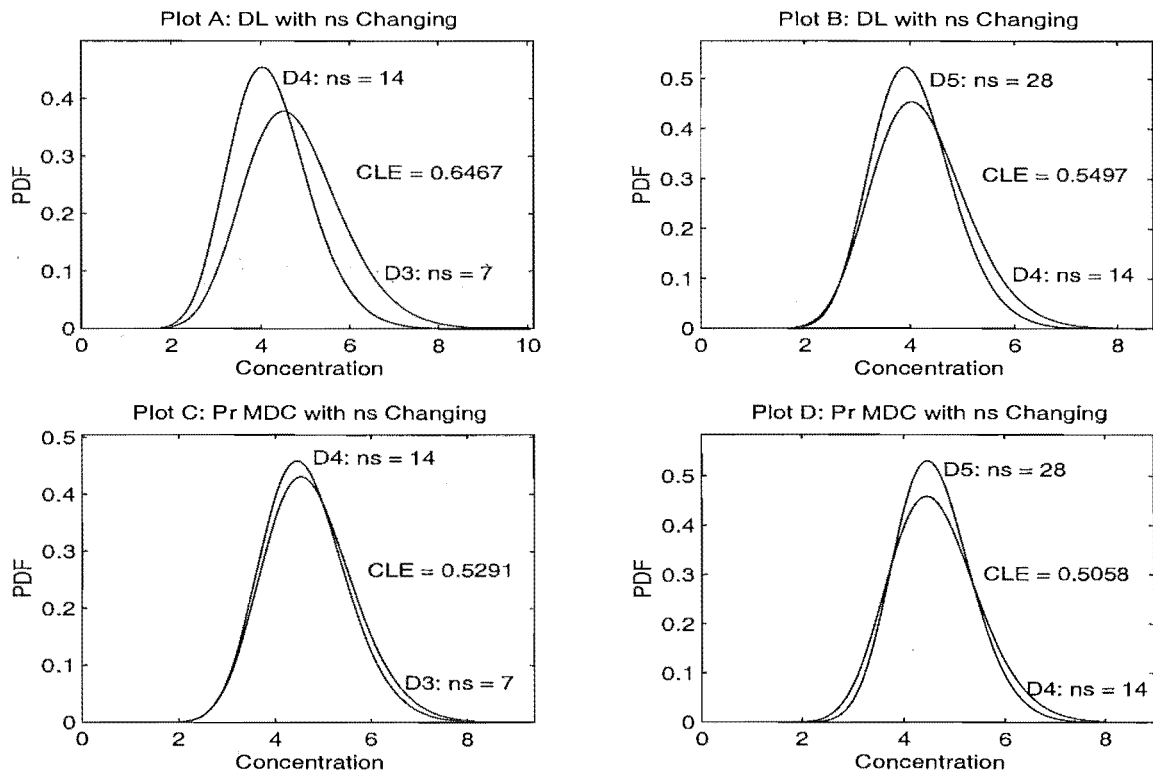
Figure 5.3: Some examples of the sampling distributions of measures of MDC when n^s changes in the NMCH model.
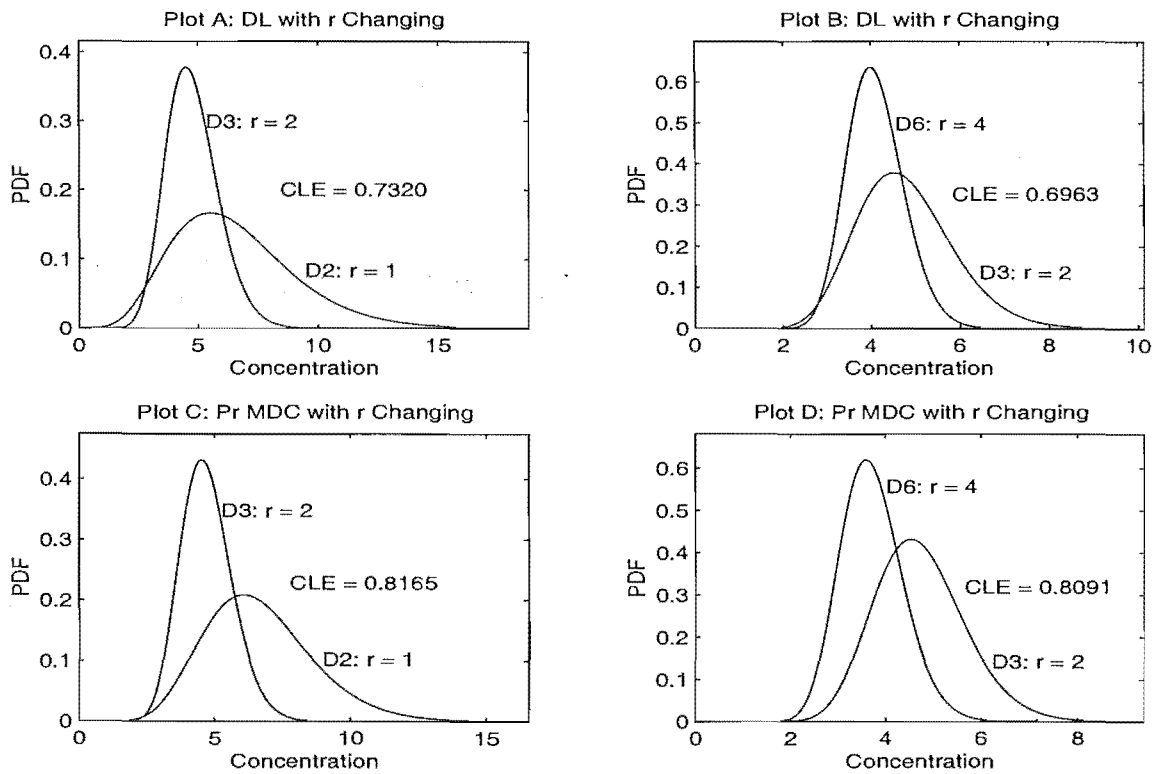
Figure 5.4: Some Examples of the Sampling Distributions of Measures of MDC when r changes in the NMCH model.

5.4.1.5 Comments: changes in n^u, n^s and r

It is apparent from Tables 5.4 to 5.6 that there is a greater difference in the RQ values for different measures of MDC when n^u, n^s and r change than when θ changes. Although the confidence intervals of the RQ values for different measures of the MDC frequently overlap, the observed differences are often consistent within and between tables. For example, in each of the Tables 5.4, 5.5 and 5.6 the Ds MDC is always greater than the RL MDC. It is therefore would be fair to say that there is substantial evidence that the RQ value is on the whole higher for the Ds MDC than for the RL MDC.

The frequentist/Bayesian foundation is the main contributor to differences in the RQ values between the measures of MDC. By inspecting Tables 5.4 to 5.6 it is obvious that the variation between the frequentist and Bayesian measures of MDC far exceeds the variation within either group. Since the Bayesian measures involve no analytical approximations, it is reasonable to assume that these react more sensibly to changes in the assay than do the frequentist measures. The following discussion is based on this supposition; i.e. that the RQ is being reliably assessed by Bayesian measures of MDC and not so by frequentist measures of MDC.

The nature of the disparity between the frequentist and Bayesian measures indicates that the approximations underlying the frequentist measures have the following effects on the assessment of assay quality:

- An increase in the number of standards is over-weighted. This is seen from the fact that in Table 5.5 the RQ values for frequentist measures is generally somewhat larger than the RQ values for the Bayesian measures. Note that this is not the case for x^{es} and x^{ei} but these are special cases and will be discussed later.

- An increase in the number of unknowns is under-weighted. This conclusion is immediately reached upon observing Table 5.4; in every instance the smallest RQ value for the Bayesian measures exceeds the largest RQ value of the frequentist measures.

- An increase in the number of replicates is under-weighted. From observing Table 5.6 one can see that RQ values for the Bayesian measures generally

exceed those of the frequentist measures. It should be noted that x^{ei} is again the exception to the rule.

The three effects listed above can be attributed to the analytical approximations used in evaluating the frequentist measures of MDC. Firstly note that these approximations cause information in the unknown samples and from the dependence of the variance function on the mean function to be discarded. Then note that the sample variances contain information about both the variance function and mean function of a model. Therefore, an increase in the number of unknowns or the number of replicates will improve the quality of the fitted variance and mean functions. Likewise, an increase in the number of standards will be tempered by the fact that a lot of information about the mean function is already provided by replicated unknowns.

Figures 5.2 - 5.4 show that as the quality of the assay is reduced, the sampling distributions lose their symmetry and become right skewed. When there is a decrease in n^u or r, the sampling distributions for the detection limit stretch outwards at both the right and left hand tails. Therefore, even though the quality of the assay has decreased in this situation, values of the DL (and other frequentist measures) which indicate that the assay is extremely good, become more likely.

The behaviour of x^{ei} is quite unique. From Tables 5.4 to 5.6 it can be observed that x^{ei} is rather insensitive to the number of unknowns but behaves like a Bayesian measure when the number of standards or replicates are changed. There is no apparent explanation for this. When n^s is changed, the substitution EL, x^{es}, also behaves differently than the other frequentist measures of MDC (see Table 5.5). However, this is explained by the fact that the error in the fitted assay model is not accounted for in x^{es}.

While the frequentist measures of MDC have been somewhat validated in situations in which only the variability of the response is changed, the same cannot be said when n^u, n^s or r change. By the arguments presented in this section, the Bayesian measures of MDC are more reliable criteria to use when comparing the quality of assays that differ on account of n^u, n^s or r.

Among Bayesian measures, it appears that x^{ds} is the best discriminator between assays with a different number of unknowns or standards and that x^{pr} is superior to x^{rl}. The scores for x^{ds} are highest in each of these cases (see Table 5.4 and Table 5.5). In the case of models D3 and D6, the RQ value for x^{ds} is the highest of

all the measures of MDC (see Table 5.6). This suggests that the Ds MDC may also be the most sensitive to changes in the number of replicates. However, for models D2 and D3 the score for x^{pr} is slightly higher implying that there is less certainty in the superiority of Ds MDC in terms of the degree of replication.

The scores for Bayesian measures are generally larger when the number of replicates changes than they are when the number of unknowns changes. This suggests that the benefit gained from increasing the number of replicates is greater than the benefit gained from analyzing more unknowns.

5.4.1.6 Best generic measure of MDC

In the light of these results and the results from Section 5.4.1.2, it can be said that x^{ds} is the best generic measure of the MDC and x^{pr} is the second best. These measures performed well in all of the situations that have been considered. It is also expected that these will be easier to interpret than existing measures since they are based on probabilities that are easy to interpret as opposed to an obscure prediction interval.

5.4.2 Comparison of measures related to Davidian's MDC

The relationships between the measures of MDC associated with Davidian's MDC (x^{pi}, the MDC in the presence of perfect information) are now explored further. Although there is some comment on the closeness of these measures to x^{pi}, the primary intention is to reinforce the points made in Section 5.4.1.

The measures of MDC to be compared are:

1. Davidians EL: x^{el} (denoted EL in Figure 5.5).

2. Substitution EL: x^{es} (denoted ES in Figure 5.5).

3. Prediction interval EL: x^{ei} (denoted EI in Figure 5.5).

4. Response level MDC: x^{rl} (denoted RL MDC in Figure 5.5).

For each model, the sampling distributions of these measures of MDC and x^{pi} (in the form of a dotted vertical line), are shown.
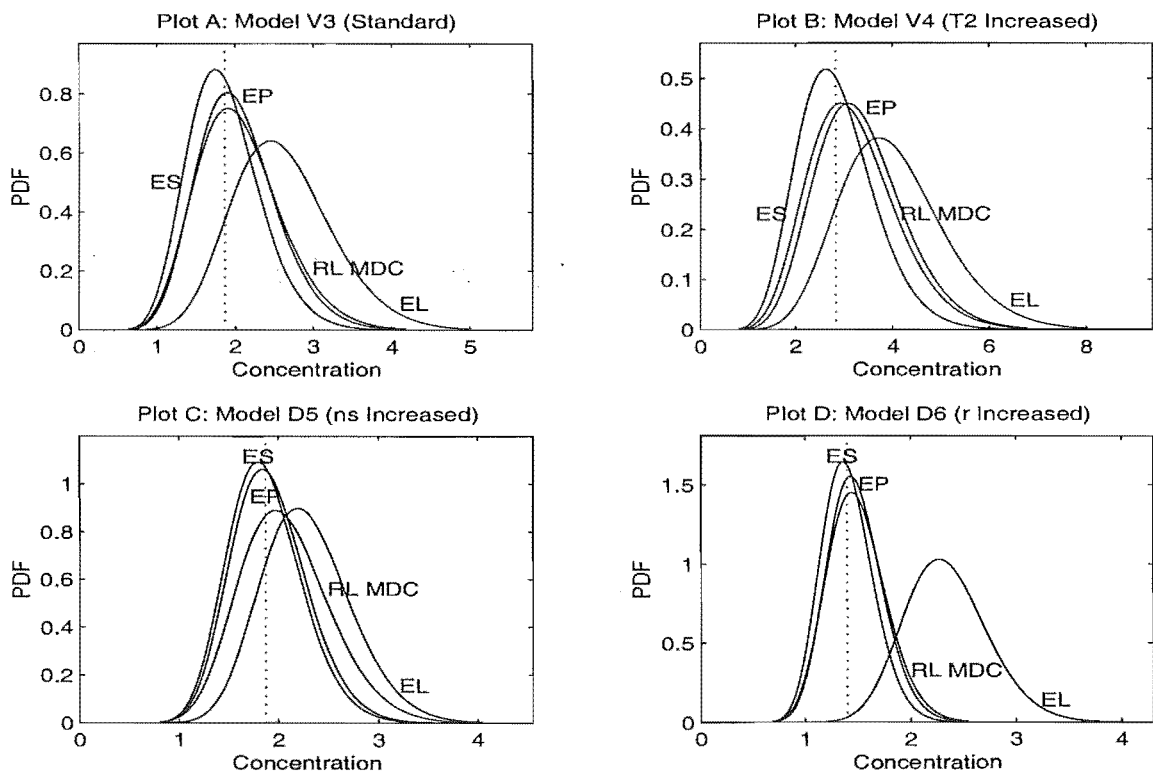
Figure 5.5: Sampling distributions of the measures of MDC related to Davidian's MDC.

5.4.2.1 Comments on Figure 5.5

The behaviour of x^{ei} and x^{rl} is very similar except when $n^s = 28$ (plot C in Figure 5.5). In this case, the sampling distribution of x^{ei} shifts to the left and behaves like x^{pi} while the sampling distribution of x^{rl} remains to the right of x^{ei}. This is an illustration of the fact that the frequentist measures of MDC are overly sensitive to changes in n^s. Very similar characteristics are observed for x^{el}.

The sampling distribution for Davidians EL, x^{el}, is in every case located well to the right of x^{pi} and of the sampling distributions of the other measures. Clearly this is a very poor estimator of x^{pi}. The substitution EL, x^{es}, appears to be a good estimator of x^{pi} but as previously noted this is not a good measure of an assay's quality and so should not be used as a measure of MDC.

It is also apparent from Figure 5.5 that the sampling distributions of x^{ei} and x^{rl} are very similar in comparison to the other distributions. This indicates that x^{ei} emulates x^{rl} to some degree whereas the other frequentist measures are much less responsive to a change in the degree of replication.

Chapter 6

The precision profile

6.1 Definition and preliminary remarks

A precision profile (PP) is the relationship between the concentration of a substance and its measured precision. The expression "precision of an assay" is taken to mean a quantitative measure of the error in the calibrated concentration of an independent sample. The formal definition of precision is the key ingredient of a precision profile. The role of the independent sample in the definition of precision emulates its role in the definition of MDC; it makes precision, like MDC, a measure of the quality of an assay's "next" measurement.

In Section 6.2 the measures of precision that have been reported in the literature are reviewed. In Section 6.3 measures of precision based on the notion of Bayesian inference are developed. In Section 6.3.6 there is some discussion of the various measures of precision and in Section 6.4 the results of some numerical computations are presented.

6.2 Current method of calculating precision

Historically, the error in the calibrated concentration has been expressed either in terms of its standard deviation or the coefficient of variation. Let $\hat{\omega} = \hat{\omega}(y, Y, X)$ be some estimator of an unknown concentration. The precision at concentration x is thus based on the variance of $\hat{\omega}$, where y is a response measurement from an independent sample with assumed concentration x. In addition, to calculate the

coefficient of variation, the mean of $\hat{\omega}$ at concentration x is also required. The sampling distribution of $\hat{\omega}$ is constructed by averaging over all possible realizations of Y and y.

Exact calculation of the mean and variance of $\hat{\omega}$ requires that β and θ are known which, of course, is not the case in practice. The mean and variance must therefore be approximated. The original derivation of Ekins (1983) is the most widely used method but relative to the definition of precision assumed in this thesis (see Section 1.2.4) it involves particularly harsh approximations. These will be exhibited shortly.

The precision profile can be constructed by either an analytical method, i.e. determining the functional relationship between precision and the hypothetical concentration x, or by calculating precision directly from a data set of fitted concentrations generated from the fitted assay model. In the latter case, it is necessary that the concentration values are spaced finely enough to ensure that no important features of the precision profile are missed.

6.2.1 Analytical method

The procedures of Ekins (1983) and O'Connell et al. (1993) are special cases of the analytical method. It is called the analytical method because the precision profile is calculated by analytically approximating the functional relationship between precision and concentration. This requires that $\hat{\omega}$ must be expressed as a closed form function of y and $(\hat{\beta}, \hat{\theta})$, the estimate of (β, θ). For such estimators, it is assumed that $\hat{\omega} = \hat{\omega}(y, \hat{\beta}, \hat{\theta})$ is expressible in a closed form. The analytical method involves using a truncated Taylor series expansion (TSE) of $\hat{\omega}(y, \hat{\beta}, \hat{\theta})$ about the mean values of y, $\hat{\beta}$ and $\hat{\theta}$, to derive approximate closed form expressions for the mean and variance of $\hat{\omega}$ at an arbitrary concentration x.

The inverse estimator defined as $\hat{\omega} = I(y, \hat{\beta}) = m^{-1}(y, \hat{\beta})$ for y in the range of $m(x, \hat{\beta})$, otherwise 0 or ∞ as appropriate, is the estimator to which the analytical method has been exclusively applied to in the literature. When m is the sigmoid

function (equation 2.1) and $\hat{\beta}_2 > 0$ this is given by

$$\hat{\omega}(y, \hat{\beta}) = \begin{cases} \hat{\beta}_3 \left(\frac{\hat{\beta}_1 - y}{y - \hat{\beta}_4}\right)^{1/\hat{\beta}_2} & \text{if } y \in (\hat{\beta}_4, \hat{\beta}_1] \\ 0 \text{ if } y > \hat{\beta}_1 \\ \infty \text{ if } y \le \hat{\beta}_4 \end{cases} \tag{6.1}$$

The inverse estimator has the intuitive appeal of being the concentration which backfits or interpolates the fitted mean function at the response y. Of course y must be a singleton or a one dimensional summary statistic of a replicated response. This will be assumed whenever the inverse estimator is being discussed.

The procedure for deriving a precision profile using the analytical method is as follows:

1. Derive the Taylor series expansion for the function $\hat{\omega}(y, \hat{\beta}, \hat{\theta})$ about the point $(m(x, \beta), E[\hat{\beta}], E[\hat{\theta}])$.

2. Truncate the expansion after a finite number of terms and denote this by TSE.

3. Compute expressions for the expectation and variance of TSE.

4. Replace β and θ by their respective estimates $\hat{\beta}$ and $\hat{\theta}$ to obtain approximate expressions for the expectation and variance of the TSE.

5. Compute the approximated precision at concentration x.

Additional approximations are typically needed to yield analytical expressions in step 3 of the algorithm.

This procedure is now illustrated by reviewing the calculations of Ekins (1983) and O'Connell et al. (1993). New approaches are then developed.

6.2.1.1 First order methods

To reduce the number of terms in the expansion suppose that $\hat{\omega}$ is a function of y and $\hat{\beta}$ as for the inverse estimator. That is, $\hat{\omega} = \hat{\omega}(y, \hat{\beta})$. Let $\hat{\omega}_y$ denote the first derivative of $\hat{\omega}$ with respect to y and $\hat{\omega}_\beta$ the gradient of $\hat{\omega}$ with respect to β. The method used by O'Connell et al. (1993) is based on the following TSE of $\hat{\omega}(y, \hat{\beta})$:

$$\hat{\omega}(\mu_x, E[\hat{\beta}]) + \hat{\omega}_y(\mu_x, E[\hat{\beta}])(y - \mu_x) + \hat{\omega}_\beta(\mu_x, E[\hat{\beta}])'(\hat{\beta} - E[\hat{\beta}]), \tag{6.2}$$

where $\mu_x = m(x, \beta)$. Note that $\hat{\omega}_y(\mu_x, E[\hat{\beta}]) = 1/m_x(\hat{\omega}(\mu_x), E[\hat{\beta}])$. Since (6.2) is linear in y and $\hat{\beta}$ and y and $\hat{\beta}$ are independent (this follows from the independence of y and Y), the mean and variance of (6.2) are easy to calculate. In the evaluation of the expectations, y is conditioned on concentration x. The mean of (6.2) is

$$E[\hat{\omega} \mid x, X] \simeq \hat{\omega}(\mu_x, E[\hat{\beta}]) + \hat{\omega}_\beta(\mu_x, \beta)'\text{bias}(\hat{\beta})$$

When $\hat{\beta}$ is an unbiased estimate of β

$$E[\hat{\omega} \mid x, X] \simeq \hat{\omega}(\mu_x, \beta) = x \qquad (6.3)$$

and the complication of having to estimate the bias of $\hat{\beta}$ is removed. Throughout this section it is assumed that $E[\hat{\beta}] = \beta$ or at least that the bias is negligible - an assumption made by authors in the past.

The variance operator is similarly applied to (6.2). By the independence of y and $\hat{\beta}$, $\text{cov}(y, \hat{\beta}_i) = 0 \forall i$ and so the variance of (6.2) reduces to

$$\begin{aligned}
V[\hat{\omega} \mid x, X] &\simeq \hat{\omega}_y(\mu_x, \beta)^2 v(\mu_x, \theta) + \hat{\omega}_\beta(\mu_x, \beta)'\Sigma\hat{\omega}_\beta(\mu_x, \beta) \\
&\simeq \hat{\omega}_y(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta}) + \hat{\omega}_\beta(\hat{\mu}_x, \hat{\beta})'\hat{\Sigma}\hat{\omega}_\beta(\hat{\mu}_x, \hat{\beta}),
\end{aligned} \qquad (6.4)$$

where $\Sigma = \text{cov}(\hat{\beta})$ and the terms $\hat{\mu}_x$ and $\hat{\Sigma}$ are obtained by respectively evaluating μ_x and Σ using $(\hat{\beta}, \hat{\theta})$ for (β, θ). Recall from Section 5.2.1 that the usual estimate of $\hat{\Sigma}$ is

$$\frac{1}{n-p}\left\{\sum_{i \in S}\frac{r_i m_\beta(X_i, \hat{\beta})m_\beta(X_i, \hat{\beta})'}{v(m(X_i, \hat{\beta}), \hat{\theta})}\right\}^{-1},$$

where m_β is the gradient of m with respect to β and $p = \dim(\beta)$.

The estimated precision at concentration x is given by $S[\hat{\omega} \mid x, X] = \sqrt{V[\hat{\omega} \mid x, X]}$, when precision is measured in terms of standard deviation and $S[\hat{\omega} \mid x, X]/x$, when precision is measured in terms of the coefficient of variation.

The calculations used in Ekins method are the same as in (6.3) and (6.4) but with $\hat{\beta}$ being treated as a constant function of the data; i.e. the randomness of $\hat{\beta}$ is ignored. The resulting expressions for the mean and variance of $\hat{\omega}$ at concentration x are

$$E[\hat{\omega} \mid x, X] \simeq x$$

and

$$V[\hat{\omega} \mid x, X] \simeq \hat{\omega}_y(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta}). \qquad (6.5)$$

These are the formulae initially used in Ekins (1983) to calculate the precision profile. Since $\hat{\beta}$ and $\hat{\theta}$ are treated as fixed quantities, the precision profile is completely insensitive to the quality of the fitted assay model. As mentioned in Section 1.2.4 this precision profile is a measure of intrinsic assay error (the component of error due to the independent response).

6.2.1.2 Higher order methods

A concern with the above calculations is that the approximate mean and variance of the first order TSE may be poor approximations of the actual mean and variance of $\hat{\omega}$. Together with other approximations used to estimate $E[\hat{\omega} \mid x, X]$ and in particular, $V[\hat{\omega} \mid x, X]$ this may cause the measure of precision to be misleading.

A potential means of improving the accuracy of these calculations is to base the estimates of the mean and variance on a higher order TSE of $\hat{\omega}(y, \hat{\beta})$. The second order Taylor series expansion of $\hat{\omega}(y, \hat{\beta})$ about $\hat{\mu}_x$ and β is (6.2) augmented with

$$\frac{1}{2}\hat{\omega}_{yy}(\mu_x, \beta)(y - \mu_x)^2 + \hat{\omega}_{y\beta}(\mu_x, \beta)'(\hat{\beta} - \beta)(y - \mu_x) + \frac{1}{2}(\hat{\beta} - \beta)'\omega_{\beta\beta}(\mu_x, \beta)(\hat{\beta} - \beta),$$

where $\hat{\omega}_{yy}$, $\hat{\omega}_{y\beta}$ and $\hat{\omega}_{\beta\beta}$ denote the second order derivatives of $\hat{\omega}$ with respect to y and β. The analogous estimate of $E[\hat{\omega} \mid x, X]$, obtained by substituting β with $\hat{\beta}$ in the derivative terms, is

$$E[\hat{\omega} \mid x, X] \simeq \hat{\omega}(\hat{\mu}_x, \hat{\beta}) + \frac{1}{2}\hat{\omega}_{yy}(\hat{\mu}_x, \hat{\beta})v(\hat{\mu}_x, \hat{\theta}) + \frac{1}{2}\mathrm{tr}(\hat{\omega}_{\beta\beta}(\hat{\mu}_x, \hat{\beta})\hat{\Sigma}),$$

where $\mathrm{tr}(A)$ denotes the trace of the matrix A.

The general expression for the variance of the second order TSE involves third and fourth order moments of the distributions of y and $\hat{\beta}$. Since only the first two moments of $y \sim f(y \mid x, \beta, \theta)$ are specified by the mean and variance functions, the distribution of the responses must be used to determine higher order moments. Procedures based on expansions that are of second order or higher are therefore parametric or in other words, distribution dependent.

Since the sampling distribution of $\hat{\beta}$ will in general be unknown, it is difficult to calculate the higher order moments. An intuitively simple solution is to use a well known distribution as an approximation. A reasonable choice, especially when the responses are normally distributed, is to assume that $\hat{\beta}$ is normally distributed. This

has some justification since maximum likelihood type estimators (M-estimators) of β are asymptotically normal.

Using the moment generating function, it is easy to show that if $z \sim N_p(0, \Sigma)$ where $p \geq 4$ then

$$E[z_i z_j] = \Sigma_{ij},$$

$$E[z_i z_j z_k] = 0, \forall i, j, k$$

and

$$E[z_i z_j z_k z_l] = \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}.$$

Thus, $\text{cov}(z_i z_j, z_k z_l) = \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}$ and $\text{var}(z_i^2) = 2\text{var}(z_i)^2$. If it is assumed that $E[y - \hat{\mu}_x \mid x, X] = 0$, then together with the earlier supposition that $E[\hat{\beta} - \beta \mid X] = 0$ the above results may be used to obtain the approximation

$$
\begin{aligned}
V[\hat{\omega} \mid x, X] \simeq\ & \hat{\omega}_y(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta}) + \hat{\omega}_\beta(\hat{\mu}_x, \hat{\beta})'\hat{\Sigma}\hat{\omega}_\beta(\hat{\mu}_x, \hat{\beta}) \\
& + \frac{1}{2}\hat{\omega}_{yy}(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta})^2 + v(\hat{\mu}_x, \hat{\beta})\hat{\omega}_{y\beta}(\hat{\mu}_x, \hat{\beta})'\hat{\Sigma}\hat{\omega}_{y\beta}(\hat{\mu}_x, \hat{\beta}) \\
& + \frac{1}{4}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\hat{\omega}_{\beta_i\beta_j}(\hat{\mu}_x, \hat{\beta})\hat{\omega}_{\beta_k\beta_l}(\hat{\mu}_x, \hat{\beta})(\hat{\Sigma}_{ik}\hat{\Sigma}_{jl} + \hat{\Sigma}_{il}\hat{\Sigma}_{jk}).
\end{aligned}
$$

The second order expressions for the expectation and variance can be used in place of their first order equivalents to evaluate the precision at concentration x.

It is clear that as the number of terms in the TSE expansion of $\hat{\omega}$ increases the number of terms in the expressions for the mean and variance will escalate. The number of terms can be reduced if second order and higher derivatives of $\hat{\omega}$ with respect to β are negligible thus enabling terms involving these can be eliminated. For example, if terms involving second order derivatives in β are excluded from the second order expansions, the expressions for the mean and variance of $\hat{\omega}(y, Y, X)$ reduce to

$$E[\hat{\omega} \mid \omega] \simeq \hat{\omega}(\hat{\mu}_x, \hat{\beta}) + \frac{1}{2}\hat{\omega}_{yy}(\hat{\mu}_x, \hat{\beta})v(\hat{\mu}_x, \hat{\theta})$$

and

$$
\begin{aligned}
V[\hat{\omega} \mid x, X] \simeq\ & \hat{\omega}_y(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta}) + \hat{\omega}_\beta(\hat{\mu}_x, \hat{\beta})'\hat{\Sigma}\hat{\omega}_\beta(\hat{\mu}_x, \hat{\beta}) \\
& + \frac{1}{2}\hat{\omega}_{yy}(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta})^2 + v(\hat{\mu}_x, \hat{\beta})\hat{\omega}_{y\beta}(\hat{\mu}_x, \hat{\beta})'\hat{\Sigma}\hat{\omega}_{y\beta}(\hat{\mu}_x, \hat{\beta})
\end{aligned}
$$

respectively. It is envisaged that there would be little, if any, benefit to be gained from using a TSE with derivatives of order greater than two. This comment stems

from a concern that the higher the order of the expansion, the greater the potential magnitude of the error from assuming that $\hat{\beta}$ is normally distributed. A decrease in the robustness of the procedure will offset the increased accuracy of the TSE as the number of terms in the expansion increases.

It should be noted that O'Malley (1996) suggested an improvement to Ekins' calculation of precision. This is to use a second order approximation but with $\hat{\beta}$ treated as a constant function of the data. If the variance and covariance terms involving $\hat{\beta}$ are set equal to zero in the general second order approximations of $E[\hat{\omega} \mid x, X]$ and $V[\hat{\omega} \mid x, X]$, one obtains

$$E[\hat{\omega} \mid x, X] \simeq \hat{\omega}(\hat{\mu}_x, \hat{\beta}) + \frac{1}{2}\hat{\omega}_{yy}(\hat{\mu}_x, \hat{\beta})v(\hat{\mu}_x, \hat{\theta})$$

and

$$V[\hat{\omega} \mid x, X] \simeq \hat{\omega}_y(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta}) + \frac{1}{2}\hat{\omega}_{yy}(\hat{\mu}_x, \hat{\beta})^2 v(\hat{\mu}_x, \hat{\theta})^2,$$

as in O'Malley (1996). Some improvement over Ekins' method was observed (see O'Malley (1996)).

6.2.2 Empirical method

The second approach that has been used to calculate the precision profile is now reviewed. Suitable aliases for this method are "memoryless method" or "direct method". This is because the analysis is developed solely on the basis of the fitted concentrations which in a way ignores or forgets the assay model for the responses.

The general procedures for calculating $E[\hat{\omega} \mid x, X]$ and $V[\hat{\omega} \mid x, X]$ using the empirical method are as follows:

1. Obtain $\hat{\beta}$ and $\hat{\theta}$ by fitting the assay model.

2. For every Y_{ij}, $i = 1, \ldots, n$ and $j = 1, \ldots, r_i$ evaluate the estimate of concentration that would result if Y_{ij} were actually observed as an independent singleton response. Denote the ijth fitted concentration by $\hat{\eta}_{ij}$. The only requirement is that $\hat{\eta}_{ij}$ is defined for every Y_{ij}.

3. Using a statistical model express $\hat{\eta}_{ij}$ as a stochastic function of η_i, the actual concentration.

4. Obtain an expression for $E(\hat{\omega} \mid x, X)$ and $V(\hat{\omega} \mid x, X)$ by fitting the model in step 3.

The above methodology is a generalization of the approach of Sadler et al. (1988). Also refer to Sadler and Smith (1990a,b) for details of this method. The model used in step 3 of the algorithm by these authors is

$$\hat{\eta}_{ij} \sim N(\eta_i, (\alpha_1 + \alpha_2\eta_i)^{\alpha_3}),$$

where α is a vector of unknown parameters. This model asserts that $\hat{\eta}_{ij}$ is an unbiased estimate of η_i.

In Sadler et al. (1988) an analytical argument based on (6.5) is used to justify the use of the variance function

$$V(\hat{\omega} \mid x, X) \doteq (\alpha_1 + \alpha_2 x)^{\alpha_3}.$$

when $m(x, \beta)$ is the sigmoid function. This variance function accommodates wide ranging behaviour in $\hat{\omega}$ and so may be suitable in many situations.

To estimate the vector α, Sadler and Smith (1986) use MACL. When there are relatively few standards, as is typically the case in immunoassays, this method of estimation is endorsed. In general, though, a method that distinguishes standards from unknowns would be preferred.

Unfortunately, with data from just a single assay, the empirical method can only account for the intrinsic error of an assay. Since Y is fixed in the calculation of $\{\hat{\eta}_{ij}, i = 1, \ldots, n, j = 1, \ldots, r_i\}$, this approach does not consider the error in the fitted assay model when estimating the variance of the fitted concentrations. Hence, only the variation due to the randomness of y is reflected in these data. The same concentrations must be estimated on two or more homogeneous assays to account for the error of estimation in $\hat{\beta}$ and $\hat{\theta}$. Refer to Chapter 7 for further comments.

The empirical method has wider applicability than the analytical method. It can be used in modern immunoassay laboratories in which analyses are performed by a machine that outputs only the estimates of concentration for each sample. The fitted assay model and response counts are often not revealed. It is clear that the analytical method cannot be used in this scenario. Again refer to Chapter 7 for further comments.

6.2.3 Exact calculation of precision

Two contrasting approaches are currently being used to calculate the precision profile of an assay. Both the analytical and empirical methods are characterized by their use of approximations which must by definition compromise the sensitivity of the precision profile.

Using simulation, precision may be calculated exactly at specific values of β, θ and η^U. The procedure for the calculation of precision at concentration x and the parameter values $(\beta_1, \theta_1, \eta_1^U)$ is as follows:

1. For given x and X, generate $y \sim f(y \mid x, \beta_1, \theta_1)$ and $Y \sim f(Y \mid \eta_1, \beta_1, \theta_1)$, where $\eta_1 = (X, \eta_1^U)$.

2. Calculate $\hat{\beta}$ and $\hat{\theta}$ as though (β, θ, η^U) were unknown by fitting the assay model to the data (Y, X).

3. Calculate $\hat{\omega} = \hat{\omega}(y, \hat{\beta}, \hat{\theta})$ and update the mean and variance of the values of $\hat{\omega}$ that have so far been generated.

4. Repeat steps 1 to 3 until the mean and variance in step 3 has converged.

5. Calculate the precision as the standard deviation or coefficient of variation and relate to the known value x.

A feature of this calculation is that the observed data are never utilized. Since the purpose of the analysis is to assess the quality of the measurements that have been made, this seems somewhat unreasonable.

A further problem with the calculation is that in practice β, θ and η^U are unknown. Some arbitrary choice for the value of these parameters thus needs to be made or, alternatively, some criterion such as the maximum or average precision with respect to (β, θ, η^U) could in theory be used. But since (β, θ, η^U) is a $p + q + n^u$ vector, these latter calculations appear to be unfeasible. It should be noted that averaging over the (β, θ, η^U) space requires that some distribution be asserted for (β, θ, η^U); hence, such a procedure is somewhat related to the calculation of precision under a Bayesian framework.

6.3 Bayesian approach to calculating precision

In this section, measures of precision are developed that arise naturally with a Bayesian analysis of an assay. In this development, the procedure for estimating an unknown concentration under a Bayesian model will be important and hence is now briefly reviewed.

6.3.1 Bayesian estimation of an unknown concentration

In a decision theoretic context the Bayesian procedure for estimating ω is based on the minimization of a loss function. Let $l(a,\omega)$ denote the loss of deciding that a is the concentration when the true concentration is ω. Examples of common loss functions are: squared error loss, $l(a,\omega) = (a - \omega)^2$; absolute error loss: $l(a,\omega) = |a - \omega|$; and weighted squared error loss: $l(a,\omega) = (a-\omega)^2/\omega$. The Bayes procedure for the estimation of the concentration ω associated with the response y is as follows:

1. Calculate $\pi(\omega \mid y, Y, X)$, the posterior distribution of ω given y and (Y, X).

2. Estimate ω as the value of a which minimizes

$$E^{\pi}[l(a,\omega) \mid y, Y, X] = \int_0^{\infty} l(a,\omega)\pi(\omega \mid y, Y, X)\, d\omega,$$

the expected value of $l(a,\omega)$ with respect to $\pi(\omega \mid y, Y, X)$.

The first step can be thought of as representing all pre-experimental and experimental information about ω in the form of a probability distribution, $\pi(\omega \mid y, Y, X)$, which describes the uncertainty about the value of ω.

In the second step, $\pi(\omega \mid y, Y, X)$ is converted to a one number summary known as an estimate. This is denoted by $\hat{\omega} = \hat{\omega}(y, Y, X)$ and is called the Bayes rule associated with the loss function $l(a,\omega)$.

The error of estimation of $\hat{\omega}$ is the posterior expected loss. This is given by

$$EL(\hat{\omega} \mid y, Y, X) = E^{\pi}[l(\hat{\omega},\omega) \mid y, Y, X] = \int_0^{\infty} l(\hat{\omega},\omega)\pi(\omega \mid y, Y, X)\, d\omega. \qquad (6.6)$$

This is reported in conjunction with the estimate and interpreted as a measure of the reliability or precision of the estimate just as the standard error is often reported in the frequentist paradigm.

In general however, it is not necessary to estimate ω by way of a loss function as in the general theoretic framework. In fact, $\hat{\omega}$ can be any function of (y, Y, X). When no loss function is specified, the estimation error of $\hat{\omega}$ is typically reported as the posterior mean squared error of $\hat{\omega}$, given by

$$E^\pi[(\hat{\omega} - \omega)^2 \mid y, Y, X] = \int_0^\infty (\hat{\omega} - \omega)^2 \pi(\omega \mid y, Y, X)\, d\omega.$$

This is defined as the posterior variance of $\hat{\omega}$ (see Berger (1985)) and is denoted here as $V^\pi[\hat{\omega} \mid y, Y, X]$. The mean squared error of $\hat{\omega}$ is minimized when $\hat{\omega} = E^\pi[\omega \mid y, Y, X]$, the posterior mean. In this case the expected loss is $V^\pi[\omega \mid y, Y, X]$, the posterior variance of ω. The posterior mean is a frequently used point estimate in Bayesian analysis.

Throughout this chapter the estimation error of $\hat{\omega}$ is written $EL(\hat{\omega} \mid y, Y, X)$. This shall be taken to mean $V^\pi(\hat{\omega} \mid y, Y, X)$ when $\hat{\omega}$ is not determined from the minimization of a loss function.

The estimate of ω conditional on β and θ, written $\hat{\omega}(y, \beta, \theta)$, is the value of a that minimizes

$$\int_0^\infty l(a, \omega) \pi(\omega \mid y, \beta, \theta)\, d\omega.$$

Note that in general $\hat{\omega}(y, Y, X)$ is not necessarily equal to

$$\int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \hat{\omega}(y, \beta, \theta) \pi(\beta, \theta \mid y, Y, X)\, d\beta\, d\theta.$$

One important exception is the case in which $\hat{\omega}$ is the posterior mean of ω. From well known results

$$
\begin{aligned}
\hat{\omega}(y, Y, X) &= E^\pi[\omega \mid y, Y, X] \\
&= E^{\pi(\beta, \theta \mid y, Y, X)}[E^\pi[\omega \mid y, \beta, \theta]] \\
&= E^{\pi(\beta, \theta \mid y, Y, X)}[\hat{\omega}(y, \beta, \theta)].
\end{aligned}
$$

The estimate in this case can be calculated by averaging the conditional estimate over the posterior distribution of (β, θ) (see Section 6.3.3.4). However, the same is not true of the estimation error since

$$
\begin{aligned}
EL(\hat{\omega} \mid y, Y, X) &= V^\pi[\omega \mid y, Y, X] \\
&= E^{\pi(\beta, \theta \mid y, Y, X)}[V^\pi[\omega \mid y, \beta, \theta]] + V^{\pi(\beta, \theta \mid y, Y, X)}[E^\pi[\omega \mid y, \beta, \theta]] \\
&= E^{\pi(\beta, \theta \mid y, Y, X)}[EL(\hat{\omega} \mid y, \beta, \theta)] + V^{\pi(\beta, \theta \mid y, Y, X)}[\hat{\omega}(y, \beta, \theta)] \\
&> E^{\pi(\beta, \theta \mid y, Y, X)}[EL(\hat{\omega} \mid y, \beta, \theta)].
\end{aligned}
$$

6.3.2 General form of a Bayesian measure of precision

As with Bayesian measures of MDC, Bayesian measures of precision are based on emulating what would be done if y were actually observed. As discussed in the previous section, an estimate ω is obtained and the estimation error is reported as a measure of accuracy or precision. To calculate the precision profile of the assay, the procedure parallels that derived for the MDC.

The same conditions that apply to the calculation of $\text{pr}(\omega > 0 \mid y, Y, X)$, for the probability MDC (x^{pr}), and $\pi(\omega \mid y, Y, X)$, for the discriminant MDC (x^{ds}), apply here. Since y is a predicted response based on $\pi(\beta, \theta \mid Y, X)$, it cannot be allowed to influence $\pi(\beta, \theta \mid Y, X)$ at any stage of the calculation of precision (see Section 5.3.2.3). Hence, $\pi(\omega \mid y, Y, X)$ is defined as

$$\pi(\omega \mid y, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \pi(\omega \mid y, \beta, \theta)\pi(\beta, \theta \mid Y, X)\, d\beta\, d\theta. \tag{6.7}$$

6.3.3 Estimator specific precision

By the expression "estimator specific precision" a two component measure of precision is implied. This measure reflects both the quality of the assay and the quality of the estimator. Estimator specific (ES) precision encapsulates a predictive loss component and a predictive error component. These components are firstly described before giving the formal definition of ES precision.

6.3.3.1 Predictive loss

The predictive loss at concentration x will be taken to mean the predictive estimation error of $\hat{\omega}$ at concentration x. This is the average of $EL(\hat{\omega} \mid y, Y, X)$ with respect to $p(y \mid x, Y, X)$. Let $PL(\hat{\omega} \mid x, Y, X)$ denote the predictive estimation error at concentration x. Then

$$PL(\hat{\omega} \mid x, Y, X) = \int_{\mathcal{R}(y)} EL(\hat{\omega} \mid y, Y, X)p(y \mid x, Y, X)dy, \tag{6.8}$$

where from (6.6) and (6.7)

$$EL(\hat{\omega} \mid y, Y, X) = \int_0^\infty \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} l(\hat{\omega}, \omega)\pi(\omega \mid y, \beta, \theta)\pi(\beta, \theta \mid Y, X)\, d\beta\, d\theta\, d\omega,$$

and

$$p(y \mid x, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} f(y \mid x, \beta, \theta)\pi(\beta, \theta \mid Y, X)\, d\beta\, d\theta.$$

A key feature of this measure is that the uncertainty about ω is expressed as a one number summary, $EL(\hat{\omega} \mid y, Y, X)$, before the hypothetical response is averaged over in (6.8). This one number is just the predicted posterior variance of ω at concentration x when $l(\hat{\omega}, \omega)$ is taken to be mean squared error.

6.3.3.2 Predictive error

This second component of ES precision makes use of the fact that x is known. The distance between $\hat{\omega}$ and x is an assessment of the quality of $\hat{\omega}$. However, this information is not used in the calculation of the predictive loss. Let $d(\hat{\omega}, x)$ be some metric of the distance of $\hat{\omega}$ from x; two common metrics being the squared distance, $d(\hat{\omega}, x) = (\hat{\omega} - x)^2$, and the absolute distance, $d(\hat{\omega}, x) = |\hat{\omega} - x|$.

The quantity $d(\hat{\omega}, x)$ is a measure of the error in $\hat{\omega}(y, Y, X)$. It can be predicted in the same way as the predicted loss. Let $PE(\hat{\omega} \mid x, Y, X)$ denote the predictive error of $\hat{\omega} = \hat{\omega}(y, Y, X)$ at concentration x; that is

$$PE(\hat{\omega} \mid x, Y, X) = \int_{\mathcal{R}(y)} d(\hat{\omega}(y, Y, X), x) p(y \mid x, Y, X) dy.$$

There is no restriction on the function d. However, it seems reasonable that d should have the same form as l in a decision theoretic setting and that $d(\hat{\omega}, x) = (\hat{\omega} - x)^2$ otherwise.

Like the predictive loss, the predictive error can also be expressed as a one number summary, namely $d(\hat{\omega}(y, Y, X), x)$, before averaging over the hypothetical response.

The predictive error appears at first sight to be exactly what is required of a measure of precision. It describes how close the estimated concentration is to the actual concentration. However, this is not the full picture. The predictive error does not account for the uncertainty in the value of concentration when y is actually observed and x is unknown. In other words, the uncertainty about the value of ω, the concentration that would be measured in practice, is not accounted for. The uncertainty about ω is in fact what the predictive loss quantifies. By combining predictive loss and predictive error into a single measure, all sources of error in $\hat{\omega}(y, Y, X)$ are taken into account.

6.3.3.3 Definition of ES precision

Define the quantity $ES(\hat{\omega} \mid x, Y, X)$ to be an ES measure of precision at concentration x if

$$ES(\hat{\omega} \mid x, Y, X) = h(PL(\hat{\omega} \mid x, Y, X), PE(\hat{\omega} \mid x, Y, X))$$

for some function h.

This is the most general form for this definition. In practice, it is difficult to think of a situation when it would not be appropriate for h to simply be a function of the arithmetic sum of the predictive loss and the predictive error, i.e.

$$ES(\hat{\omega} \mid x, Y, X) = h(PL(\hat{\omega} \mid x, Y, X) + PE(\hat{\omega} \mid x, Y, X)). \tag{6.9}$$

6.3.3.4 Numerical evaluation of ES precision

For simplicity, it is supposed that estimation is by squared error loss, so that $\hat{\omega} = E^{\pi}[\omega \mid y, Y, X]$ and $EL[\hat{\omega} \mid y, Y, X] = V^{\pi}[\omega \mid y, Y, X]$. Also let $d(\hat{\omega}, x) = (\hat{\omega} - x)^2$. In this case the algorithm for the calculation of x^{ds} may be easily adapted to the predictive loss (PL) and the predictive error (PE). Samples of $\omega \sim \pi(\omega \mid y, Y, X)$ are drawn in exactly the same manner as $(\omega, \omega_0) \sim \pi(\omega, \omega_0 \mid y, y_0, Y, X)$ variates were generated to calculate x^{ds}. The only difference to the structure of the algorithm is that, instead of evaluating a probability in the inner loop, statistics that enable the calculation of $PL(\hat{\omega} \mid x, Y, X)$ and $PE(\hat{\omega} \mid x, Y, X)$ are enumerated.

The following pseudo code forms the backbone of a suitable algorithm for the calculation of the PL and the PE.

1. Run two independent M-H simulations on $\pi(\beta, \theta \mid Y, X)$ until each chain has surpassed its transient phase.

2. Generate $y \sim p(y \mid x, Y, X)$.

   (a) Generate $(\beta^{(i)}, \theta^{(i)}) \sim \pi(\beta, \theta \mid Y, X)$ using the M-H algorithm.

   (b) Generate an r dimensional response vector $y^{(i)} \sim f(y \mid x, \beta^{(i)}, \theta^{(i)})$.

3. Generate $\omega \sim \pi(\omega \mid y, Y, X)$.

   (a) Set $j = 1$.

(b) Generate $(\beta^{(j)}, \theta^{(j)}) \sim \pi(\beta, \theta \mid Y, X)$ using an independent M-H simulation to that used in step 2.

(c) Use MCMC simulation or importance sampling to generate $\omega^{(j)} \sim p(\omega \mid y^{(i)}, \beta^{(j)}, \theta^{(j)})$. If the M-H algorithm is used, a warm-up run will be necessary.

4. Continue step 3 until both the long run averages of $\omega^{(j)}$ and $(\omega^{(j)})^2$ have converged.

5. Let $\bar{\omega}$ and $PL^{(i)}$ be the long-run averages of $\omega^{(j)}$ and $(\omega^{(j)} - \bar{\omega})^2$ respectively and $PE^{(i)} = (\bar{\omega} - x)^2$.

6. Repeat steps 2-5 until the long run averages of the $PL^{(i)}$ variates and the $PE^{(i)}$ values in step 5 converge and all other convergence criteria of the two M-H chains are satisfied. The converged averages of these sequences will be $PL(\hat{\omega} \mid x, Y, X)$ and $PL(\hat{\omega} \mid x, Y, X)$ respectively.

To develop the precision profile, this algorithm must be applied to values of x that span the range of concentrations of interest. The resulting values of ES precision are then plotted against x.

When criteria other than minimum expected squared error loss are used to estimate ω, steps 4 and 5 of the algorithm need to be adjusted. The calculation of the posterior mean needs to be substituted with the calculations needed to obtain the estimate of concentration corresponding to the estimation criterion being used. If the estimator cannot be expressed in closed algebraic form, numerical minimization of the loss function may have to be embedded in the algorithm.

6.3.3.5 Alternative methods of calculation

The independence required of $y^{(i)}$ and $(\beta^{(j)}, \theta^{(j)})$ makes this algorithm, like those for x^{pr} and x^{ds}, rather complicated. The computation would be more efficient if y and ω variates could be generated from the same value of β and θ.

As for the calculation of x^{pr} and x^{ds} this is not possible, at least for loss and distance functions that are strictly convex. This is because neither the predictive loss or the predictive error are able to be expressed in the form

$$E^{\pi(\beta,\theta|Y,X)}[E^{f(y|x,\beta,\theta)}[P(y,\beta,\theta)]],$$

where $P(y, \beta, \theta)$ is some function of y, β and θ. If this were the case, numerical evaluation would be easy and efficient to evaluate because the order of integration is in the reverse order to the order in which (β, θ), y and ω are naturally generated. Here the natural order of generation is (β, θ) followed by y and then ω. Hence, each variate only needs to be generated once within each iteration of a Monte Carlo simulation.

Partial representation in the above form is possible for some loss functions. Again, consider the case where l and d represent squared error loss and distance respectively. Then

$$
\begin{aligned}
EL(\hat{\omega} \mid y, Y, X) &= V^{\pi}[\omega \mid y, Y, X] \\
&= E^{\pi(\beta, \theta \mid Y, X)}[V^{\pi}[\omega \mid y, \beta, \theta]] + V^{\pi(\beta, \theta \mid Y, X)}[E^{\pi}[\omega \mid y, \beta, \theta]]
\end{aligned}
$$

The second term on the rhs is not in the required form since it depends on $(E^{\pi}[\omega \mid y, Y, X])^2$. Therefore, the inner expectation/integration with respect to β and θ will not cancel when the expectation of $EL(\hat{\omega} \mid y, Y, X)$ is taken with respect to $p(y \mid x, Y, X)$. However, the sought after cancellation does take place for the first term since

$$
\begin{aligned}
E^{p(y \mid x, Y, X)}[E^{\pi(\beta, \theta \mid Y, X)}[V^{\pi}[\omega \mid y, \beta, \theta]]] &= E^{\pi(\beta, \theta \mid Y, X)}[E^{f(y \mid x, \beta, \theta)}[V^{\pi}[\omega \mid y, \beta, \theta]]] \\
&= E^{\pi(\beta, \theta \mid Y, X)}[PL(\hat{\omega} \mid x, \beta, \theta)]. \quad (6.10)
\end{aligned}
$$

The efficiency of the algorithm may be able to be improved by embedding a more efficient calculation of $E^{\pi(\beta, \theta \mid Y, X)}[V^{\pi}[\omega \mid y, \beta, \theta]]$ in the full algorithm or by calculating this component of ES precision separately.

6.3.3.6 Alternative (incorrect) expression

An alternative expression for the precision of an estimator is obtained by averaging $E(\hat{\omega} \mid x, \beta, \theta)$, the predicted loss conditional on β and θ, over $\pi(\beta, \theta \mid Y, X)$. The resulting expression

$$
\int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \left[\int_{\mathcal{R}(y)} l(\hat{\omega}(y, \beta, \theta), x) f(y \mid x, \beta, \theta) \, dy \right] \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta, \quad (6.11)
$$

is just the general form of (6.10). Hence, although the above expression and $EL[\hat{\omega} \mid x, Y, X]$ are functions of the same variables (x, Y and X), this expression does not

account for all of the sources of uncertainty in the measurement process. It will therefore yield an overly optimistic view of the assay. Only in the extreme case where β and θ are known does this equal ES precision.

6.3.4 The predictive distribution of assay measurements (PDAM)

In the calculation of the predictive loss and the predictive error it was noted that the uncertainty in ω is summarized before averaging over the hypothetical response. This is a consequence of quantifying the uncertainty about ω in terms of the error in $\hat{\omega}(y, Y, X)$. However, the general definition of precision does not specify that precision must be based upon the properties of an estimator of the unknown concentration. The calculation of an estimator of an unknown concentration is a feature that is only required when the analysis is frequentist. When the analysis of the assay is Bayesian it is not necessary to be so restricted, in fact precision can be based upon any measure of the quality of an assay.

6.3.4.1 Derivation

If the involvement of an estimator of ω is taken out of the calculation of precision then just two terms are left, namely the posterior distribution of ω, $\pi(\omega \mid y, Y, X)$ and the predictive distribution of y, $p(y \mid x, Y, X)$, in which x is arbitrary. The uncertainty in ω is completely described by $\pi(\omega \mid y, Y, X)$. Since y is a hypothetical response, the dependence of $\pi(\omega \mid y, Y, X)$ on y needs to be removed. The procedure is to average $\pi(\omega \mid y, Y, X)$ with respect to the distribution of y given x and (Y, X); that is, with respect to $p(y \mid x, Y, X)$, thus obtaining

$$\pi(\omega \mid x, Y, X) = \int_{\mathcal{R}(y)} \pi(\omega \mid y, Y, X) p(y \mid x, Y, X) \, dy.$$

Of course, as y is hypothetical $\pi(\omega \mid y, Y, X)$ is calculated as

$$\pi(\omega \mid y, Y, X) = \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \pi(\omega \mid y, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta,$$

to avoid the undesirable phenomena of having the response y, predicted from the fitted model, reinforcing the model fit.

The resulting distribution, $\pi(\omega \mid x, Y, X)$, is the predictive posterior distribution of the unknown concentration at concentration x. It is called the predicted distribution of assay measurements at concentration x (PDAM(x)). The abbreviation, PDAM, is used for both singular and plural versions of "predicted distribution(s) of assay measurements". As the full distribution of the uncertainty in the measurement of ω is predicted, the PDAM contain all of the information about an assay's measurement of ω at any value of x.

6.3.4.2 Expansion

It is instructive to express PDAM(x) as the following mixtures of its component distributions:

$$
\pi(\omega \mid x, Y, X)
$$
$$
= \int_{\mathcal{R}(y)} \pi(\omega \mid y, Y, X) p(y \mid x, Y, X) \, dy \tag{6.12}
$$
$$
= \int_{\mathcal{R}(y)} \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \pi(\omega \mid y, \beta, \theta) \pi(\beta, \theta \mid Y, X) p(y \mid x, Y, X) \, d\beta \, d\theta \, dy \tag{6.13}
$$
$$
= \int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \int_{\mathcal{R}(y)} \pi(\omega \mid y, Y, X) f(y \mid x, \beta, \theta) \pi(\beta, \theta \mid Y, X) \, dy \, d\beta \, d\theta. \tag{6.14}
$$

Note that, (6.14) is obtained directly from (6.12) not (6.13). The two mixtures in (6.13) and (6.14) cannot be combined because $\pi(\omega \mid y, Y, X)$ and $p(y \mid x, Y, X)$ must be conditioned on independent values of (β, θ); i.e. they cannot be conditioned on the same value of (β, θ). It is a consequence of this fact that PDAM(x) cannot be expressed as a posterior expectation of $\pi(\omega \mid x, \beta, \theta)$, a quantity that would be known as PDAM(x) given (β, θ).

It is important not to fall into the trap of conditioning $\pi(\omega \mid x, Y, X)$ on (β, θ) before conditioning on y. This would lead to the expression

$$
\int_{\mathcal{R}(\theta)} \int_{\mathcal{R}(\beta)} \left[\int_{\mathcal{R}(y)} \pi(\omega \mid y, \beta, \theta) f(y \mid x, \beta, \theta) \, dy \right] \pi(\beta, \theta \mid Y, X) \, d\beta \, d\theta, \tag{6.15}
$$

in which it can be observed that $\pi(\omega \mid y, Y, X)$ and $p(y \mid x, Y, X)$ are conditioned on the same value of (β, θ). This expression is an invalid expansion for the PDAM because $\pi(\omega \mid y, \beta, \theta)$ is, in effect, averaged with respect to $f(y \mid x, \beta, \theta)$ as opposed to $p(y \mid x, Y, X)$. Incidently, the error of using (6.15) to calculating PDAM(x) is analogous to error of using (6.11) to calculate ES precision.

6.3.5 PDAM precision

The PDAM provide precisely what is needed to form a precision profile, namely a probability distribution on ω (the quantity being measured) that is conditioned on x (the true value of this quantity) and the information in the assay, (Y, X).

6.3.5.1 Definition of PDAM precision

A PDAM measure of precision at concentration x is defined to be any one-dimension measure of the spread of $PDAM(x)$. Examples of such measures are variance, range, squared error and relative squared error.

PDAM precision is a one number summary of the uncertainty that surrounds the measurement of ω. The property of the PDAM which best reflects the quality of an assay's measurements at a given concentration will be situation and application dependent. In the majority of situations, it is envisaged that a measure based on mean squared error (MSE) will be appropriate. The calculation in the case of relative root mean squared error (RRMSE) is

$$RRMSE^{\pi}[\omega \mid x, Y, X] = \frac{\sqrt{MSE^{\pi}(\omega \mid x, Y, X)}}{x},$$

where

$$\begin{aligned}
MSE^{\pi}[\omega \mid x, Y, X] &= \int_{\mathcal{R}(\omega)} (\omega - x)^2 \pi(\omega \mid x, Y, X) d\omega \\
&= \int_{\mathcal{R}(\omega)} \int_{\mathcal{R}(y)} (\omega - x)^2 \pi(\omega \mid y, Y, X) p(y \mid x, Y, X) \, dy \, d\omega \\
&= \int_{\mathcal{R}(y)} \int_{\mathcal{R}(\omega)} (\omega - x)^2 \pi(\omega \mid y, Y, X) \, d\omega \, p(y \mid x, Y, X) \, dy \\
&= \int_{\mathcal{R}(y)} MSE^{\pi}(\omega \mid y, Y, X) p(y \mid x, Y, X) \, dy. \quad (6.16)
\end{aligned}$$

Equation 6.16 indicates that when it is measured in terms of RRMSE, PDAM precision is based on the predicted value of $MSE^{\pi}(\omega \mid y, Y, X)$, a quantity that would be reported if the response y was actually observed. A similar expansion to (6.16) holds for all measures of precision based on moments of $PDAM(x)$ about x.

The advantage of mean squared error based measures over variance based measures is that both location and spread are considered. Thus, there is no need to constrain bias when measuring the spread of the measurements.

6.3.5.2   Relationship between ES and PDAM precision

The ES and PDAM measures of precision are different quantities. However, in a very important case they coincide.

Theorem 6.3.1 *Suppose that estimator specific precision is*

$$ES(\hat{\omega} \mid x, Y, X) = PL(\hat{\omega} \mid x, Y, X) + PE(\hat{\omega} \mid x, Y, X),$$

that $\hat{\omega}$ is the Bayes rule under squared error loss and that $d(\hat{\omega}, x) = (\hat{\omega} - x)^2$ in the calculation of $PE(\hat{\omega} \mid x, Y, X)$. Then, if the PDAM precision at concentration x is the mean squared error of $PDAM(x)$, PDAM precision equals ES precision.

Proof: As the posterior mean is the Bayes rule under squared error loss

$$\hat{\omega} = \hat{\omega}(y, Y, X) = E^{\pi}[\omega \mid y, Y, X]$$

and

$$
\begin{aligned}
EL(\hat{\omega} \mid y, Y, X) &= V^{\pi}[\omega \mid y, Y, X] \\
&= E^{\pi}[(\omega - E^{\pi}[\omega \mid y, Y, X])^2].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
PL(\hat{\omega} \mid x, Y, X) &= \int_{\mathcal{R}(y)} V^{\pi}[\omega \mid y, Y, X] p(y \mid x, Y, X)\, dy \\
&= E^{p(y \mid x, Y, X)}[V^{\pi}[\omega \mid y, Y, X]]
\end{aligned}
$$

and

$$
\begin{aligned}
PE(\hat{\omega} \mid x, Y, X) &= \int_{\mathcal{R}(y)} (E^{\pi}[\omega \mid y, Y, X] - x)^2 p(y \mid x, Y, X)\, dy \\
&= E^{p(y \mid x, Y, X)}[(E^{\pi}[\omega \mid y, Y, X] - x)^2]
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
ES(\hat{\omega} \mid x, Y, X) &= E^{p(y \mid x, Y, X)}[V^{\pi}[\omega \mid y, Y, X]] + E^{p(y \mid x, Y, X)}[(E^{\pi}[\omega \mid y, Y, X] - x)^2] \\
&= E^{p(y \mid x, Y, X)}\{E^{\pi}[(\omega - E^{\pi}[\omega \mid y, Y, X])^2] + (E^{\pi}[\omega \mid y, Y, X] - x)^2\} \\
&= E^{p(y \mid x, Y, X)}\{E^{\pi}[(\omega - x)^2 \mid y, Y, X]\} \\
&= \int_{\mathcal{R}(y)} \int_0^{\infty} (\omega - x)^2 \pi(\omega \mid y, Y, X) p(y \mid x, Y, X)\, d\omega\, dy \\
&= \int_0^{\infty} (\omega - x)^2 \pi(\omega \mid x, Y, X)\, d\omega \\
&= MSE^{\pi}[\omega \mid x, Y, X].
\end{aligned}
$$

Under the mean squared error criterion, $\text{MSE}^\pi[\omega \mid x, Y, X]$ is the PDAM precision at concentration x and so the proof is complete.

This result to some extent "characterizes" PDAM precision. It is a re-assuring result that the two Bayesian measures of precision are equivalent under the widely used squared error/squared distance metrics. It is clear that the two measures will also coincide under these metrics if $ES(\hat{\omega} \mid x, Y, X) = h(PL(\hat{\omega} \mid x, Y, X) + PE(\hat{\omega} \mid x, Y, X))$ and $\text{PDAM}(x) = h(\text{MSE}^\pi[\omega \mid x, Y, X])$ for any function h.

6.3.5.3 Numerical evaluation of PDAM precision

To calculate any measure of PDAM precision, one just needs to be able to sample from $\pi(\omega \mid x, Y, X)$. Recall that $\text{PDAM}(x)$ is defined as:

$$\text{PDAM}(x) = \int_{\mathcal{R}(y)} \pi(\omega \mid y, Y, X) p(y \mid x, Y, X)\, dy.$$

This indicates that to evaluate $\text{PDAM}(x)$ samples from the following distributions must be able to be generated:

1. $y \sim p(y \mid x, Y, X)$.

2. $\omega \sim \pi(\omega \mid y, Y, X)$.

Methods of generating variates from each of these distributions have been discussed earlier. The order in which samples are drawn from these distributions is identical to that needed to calculate ES precision. In fact, the only change required to adapt the algorithm used to calculate ES precision to the calculation of PDAM precision is the nature of the summary statistics being enumerated and the locations where enumeration occurs.

The following algorithm outlines a method for the calculation of the PDAM precision at concentration x when it is measured in terms of the RRMSE of $\text{PDAM}(x)$.

1. Run two independent M-H simulations on $\pi(\beta, \theta \mid Y, X)$ until each chain has surpassed its transient phase and set SSQ$= 0$

2. Generate $y \sim p(y \mid x, Y, X)$.

   (a) Generate $(\beta^{(i)}, \theta^{(i)}) \sim \pi(\beta, \theta \mid Y, X)$ using the M-H algorithm.

   (b) Generate an r dimensional response vector $y^{(i)} \sim f(y \mid x, \beta^{(i)}, \theta^{(i)})$.

3. Generate $\omega \sim \pi(\omega \mid y, Y, X)$.

    (a) Set $j = 1$.

    (b) Generate $(\beta^{(j)}, \theta^{(j)}) \sim \pi(\beta, \theta \mid Y, X)$ using an independent M-H simulation to that used in step 2.

    (c) Use importance sampling or MCMC simulation to generate $\omega^{(j)} \sim p(\omega \mid y^{(i)}, \beta^{(j)}, \theta^{(j)})$ for $j = 1, \ldots, K1$. If the M-H algorithm is used, a warm-up run will be necessary.

    (d) Calculate SSQ $=$ SSQ $+ \sum_{j=1}^{K1} (\omega^{(j)} - x)^2$.

4. Calculate $\text{SSQ}^{(i)} = SSQ/K1$.

5. Repeat steps 2-4 until all convergence diagnostics of the MCMC chain are satisfied and it is evident that the mean of the $\text{SSQ}^{(i)}$ values in step 4 has converged.

6. Set $\text{RRMSE}(\omega \mid x, Y, X) = \sqrt{(\text{SSQ}/K2)}/x$, where $K2$ is the total number of times the algorithm passed through step 5.

To develop the precision profile the algorithm needs to be applied to values of x that span the range of concentrations of interest. The resulting values of precision can then be plotted against x to form the precision profile.

The nature of the calculation in steps 3d, 4 and 6 will have to be altered if other measures of PDAM precision are used. For example, to calculate the variance of PDAM(x), running totals of $\omega^{(j)}$ and $(\omega^{(j)})^2$ need to be maintained at step 3d and the averages at steps 4 and 6 adjusted accordingly. It should be noted that it is possible to adapt the algorithm so that it will calculate any property of PDAM(x).

6.3.5.4   Further applications of the PDAM

The PDAM completely describe the quality of an assay's measurement of the concentration in an unseen sample. This completeness makes the PDAM candidates for the calculation of other aspects of an assay's quality. For example, probability and discriminant type measures of MDC can be defined in terms of the PDAM.

The predictive posterior probability at concentration x that ω, the calibrated concentration, is in some set can be obtained from PDAM(x). A concentration

might then be said to be "reliably able to be measured" if the predicted posterior probability of this set is sufficiently high (see Chapter 5 for various other definitions of the term "reliable measurement").

This leads to the definition of the MDC as the smallest value of x for which

$$\text{pr}(\omega > 0 \mid x, Y, X) \geq 1 - \alpha,$$

where $\omega \sim \text{PDAM}(x)$. Unlike the probability MDC of Chapter 5, the act of deciding if $\omega > 0$ for all possible values of y is removed. This makes this measure of MDC a function of the PDAM. Of course for this measure to make sense it is necessary that a priori $\text{pr}(\omega = 0) > 0$, as for the probability MDC.

The PDAM can also be used to develop a measure of the MDC based on discrimination of a sample with positive concentration from one with zero concentration. The resulting MDC is the smallest value of x for which

$$\text{pr}(\omega > \omega_0 \mid x, 0, Y, X) \geq 1 - \alpha,$$

where $(\omega, \omega_0) \sim \pi(\omega, \omega_0 \mid x, 0, Y, X)$ - a bivariate PDAM.

Finally, the PDAM can be used for model selection. When $\pi(\omega \mid y, Y, X)$ was used for the derivation of measures of MDC or precision it was noted that y must not be allowed to influence $\pi(\beta, \theta \mid Y, X)$. In other words y is withdrawn or "held out" from $\pi(\beta, \theta \mid Y, X)$. Hence, there is some similarity between $\pi(\omega \mid y, Y, X)$ and the "hold-out" distributions used in Bayesian cross-validation. The major distinction is that a PDAM predicts the independent variable whereas hold-out distributions predict the response.

In any case, the PDAM could be used for model selection whenever there is a basis for judging a model in terms of its prediction of the value of an independent variable. Such applications will be those where the primary intention of the analysis is the estimation of values of the independent variable from observed responses, such as in assays.

6.3.6 Discussion

6.3.6.1 Bayesian verse frequentist precision profiles

In the frequentist framework, precision is defined in terms of a property of the sampling distribution of some estimator of ω denoted by $\hat{\omega}(y, Y, X)$. The calculation

of precision therefore involves averaging with respect to $f(y, Y \mid x, X)$; i.e. averaging over both past and future observations. In the Bayesian framework precision is defined using the notion of predictive inference. Instead of averaging over all values of Y the averaging is with respect to $p(y \mid x, Y, X)$ which clearly conditions on the present observations.

The role of $p(y \mid x, Y, X)$ in a Bayesian framework parallels the role of the sampling distribution of $(y, \hat{\beta}, \hat{\theta})$ in a frequentist framework; i.e. it serves the same purpose. Both the uncertainty in β and θ and the randomness of y are taken account of in $p(y \mid x, Y, X)$.

Bayesian analysis allows more freedom in the construction of the precision profile. An equivalent to the PDAM and hence PDAM precision clearly does not exist under the frequentist paradigm since uncertainty about ω cannot be expressed in the form of a probability distribution about ω. Probability distributions can be defined for estimates of a parameter but not for the parameter itself. In the frequentist paradigm uncertainty about ω can only be represented in terms of a sampling distribution for $\hat{\omega}$, as if the present assay could be repeated over and over again under the exact same conditions. Thus, when the analysis is frequentist, an estimator of the concentration ω must be relied upon and the calculation of precision will be specific to this estimator.

6.3.6.2 ES verse PDAM based precision profiles

In the Bayesian paradigm one is able to choose between ES and PDAM precision. These measures are now compared. The former, ES precision, takes account of both the quality of the assay experiment and the method of estimating the unknown concentrations. If the estimator of an unknown concentration has undesirable properties, ES precision will be adversely affected. It should be used if the quality of an assay's estimation of an unknown concentration is of fundamental concern (for example, when deciding which of two estimators of η^U is best).

The fundamental feature of PDAM precision is that there is no reliance on an estimator of the unknown concentration. PDAM precision therefore reflects the quality of the assay in its purest form as opposed to being tailored to the specific demands of an estimator. PDAM precision is a more general measure of quality than ES precision. If the quality of the assay alone is of interest, then PDAM precision

is the better measure. This is the case whenever aspects of an assay's design (for example, the number, replication and concentration of the standards) are being studied.

6.3.6.3 Use of ES and PDAM precision with Bayesian measures of MDC

ES and PDAM precision profiles may be used in conjunction with the probability and discriminant measures of MDC to present a summary of the quality of the assay. Practitioners often use the precision profile to determine the range of concentrations measured or calibrated within a certain level of precision. This is called the working range. The working range together with x^{pr} and x^{ds} would constitute a three number summary of the limit of reliable detection of an assay. This is in the spirit of the three number summary originally proposed by Currie (1968).

6.4 Numerical illustration

In this section the precision profiles are calculated using the NMCH data and model. For a range of values of x PDAM(x) is plotted. The non-informative prior $\pi(\beta, \theta, \mu^U) \propto 1/\tilde{\theta}_0$ is used for Bayesian model fitting. These plots provide insight into the relationship between the various measures of precision.

The performance of the measures of precision could have been assessed in the same manner as the measures of MDC were in Section 5.4.1. Given a rule for, on the basis of their precision profiles, determining the better of two assays the methodology can be directly applied. The smallest area under the precision profile on an interval of interest, the lowest value of precision at a concentration of particular importance and the largest width of the working range (the range of concentrations that can be measured within some specified level of precision) are all potential rules for selecting the better of two assays based from their precision profiles. Due to the similarity in the construction of the MDC and the precision at a given concentration it is expected that the results from such a simulation would be similar in spirit to the results for the frequentist and Bayesian measures of MDC in Section 5.4.1.
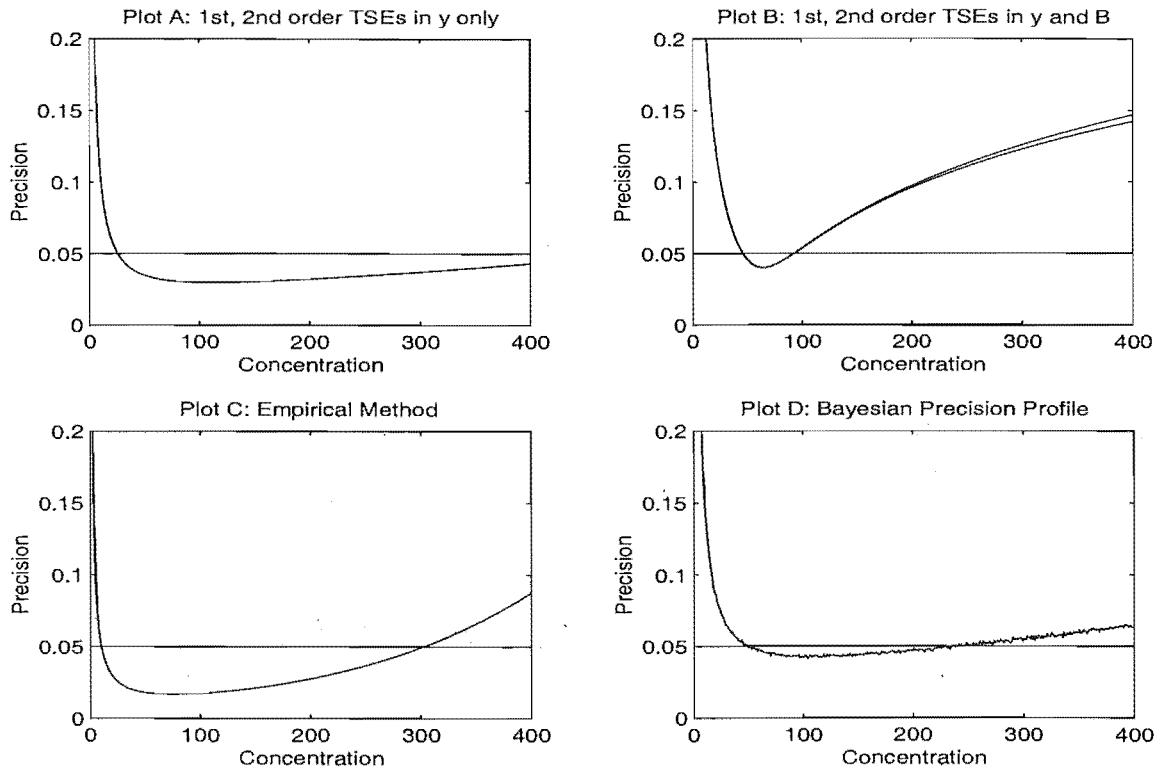
Figure 6.1: The precision profiles for the NMCH data.

6.4.1 Precision profiles for the NMCH data

The precision profiles for the NMCH data and model are shown in Figure 6.1. A Bayesian precision profile is developed for the RRMSE measure of PDAM precision. Note that, this is equivalent to the measure of ES precision given by:

$$ES(\hat{\omega} \mid x, Y, X) = \frac{\sqrt{PL(\hat{\omega} \mid x, Y, X) + PE(\hat{\omega} \mid x, Y, X)}}{x},$$

where $PL(\hat{\omega} \mid x, Y, X)$ and $PE(\hat{\omega} \mid x, Y, X)$ are calculated using the squared error loss and squared distance criteria respectively. Frequentist measures of precision are based on the coefficient of variation and Bayesian measures are based on relative root mean squared error. The inverse estimator is used to calculate the frequentist measures of precision.

6.4.1.1 Comments

From Figure 6.1 it is clear that the precision profiles fall into the following four categories:

1. Precision profiles developed using the analytical method applied to only y. Both the first and second order versions of these precision profiles are shown in plot A.

2. Precision Profiles developed using the analytical method applied to y and $\hat{\beta}$. The first and second order versions of these precision profiles are shown in plot B.

3. The precision profile developed using the empirical method. This precision profile is shown in plot C.

4. The Bayesian precision profile. The ES/PDAM precision profile is shown in plot D.

As the precision profiles in plot A cannot be distinguished and those in plot B can only just be distinguished, it is strikingly apparent that the addition of the second order terms in the analytical approach has little effect on the calculation of precision. For the remainder of this discussion the first and second order analytical methods will be discussed as one.

The difference between the precision profiles in plots A and B reveals that the incorporation of the error in the model fit has a significant effect on the precision profiles derived using the analytical approach. Accounting for only the error in y results in a precision profile that is a lot more optimistic than when the error in the model fit is also accounted for. This is reflected in the profound difference in the working ranges at level 0.05, the range of concentrations for which precision is less than 0.05, between plots A and B.

The analytical methods involving expansions in y and β yield precision profiles that are generally substantially more conservative than the Bayesian precision profile (to observe this compare plots B and D). There is only a small interval around the minimum point of the precision profile in which the analytical method's precision

profiles are less conservative than the Bayesian precision profiles[1]. The working range at level 0.05 is indeed significantly greater for the Bayesian precision profile. The discrepancy is likely to have been caused by the fact that in the frequentist calculation of the error in the model fit, the information in the unknowns and the dependence of v on m are not adequately incorporated in the analysis. Thus, an overly pessimistic view of the assay will be given if the analytical method is applied to both y and β. Of course, it is assumed here that the Bayesian precision profile is the standard against which other precision profiles should be judged, but this seems reasonable since it appeared as though this was the case for the MDC (see the results and discussion in Section 5.4.1).

The empirical method yields a precision profile (see plot C) that exhibits variable behaviour. Over low and mid-range concentrations it gives the most favourable impression of the assay. However, compared to the Bayesian precision profile (see plot D), it is significantly less favourable at high concentrations. The consequence of this is that the precision profile produced by the empirical method and the Bayesian precision profile give rise to similar working ranges at level 0.05. It is clear that this is, however, a fortuitous result because for the most part the empirical method's precison profile is quite different from the Bayesian precision.

An interesting feature of the precision profiles is that they are decreasing functions over some concentrations and increasing functions over others. Even when the measure of precision is based on variance or squared error as opposed to a relative measure, it is still often the case that the precision profile decreases before it increases. This happens because the decrease in the variability of y temporarily offsets the increasing effect of other sources of variability in the calibrated value of ω.

6.4.2 The PDAM for the NMCH data

The PDAM(x) for $x = 0, 2, 5, 10, 100$ and 300 are shown in Figure 6.2.

---

[1]It can be construed from plots B and D that this interval is approximately from 50 to 80 on the horizontal axis, i.e. for $x \in [50, 80]$.
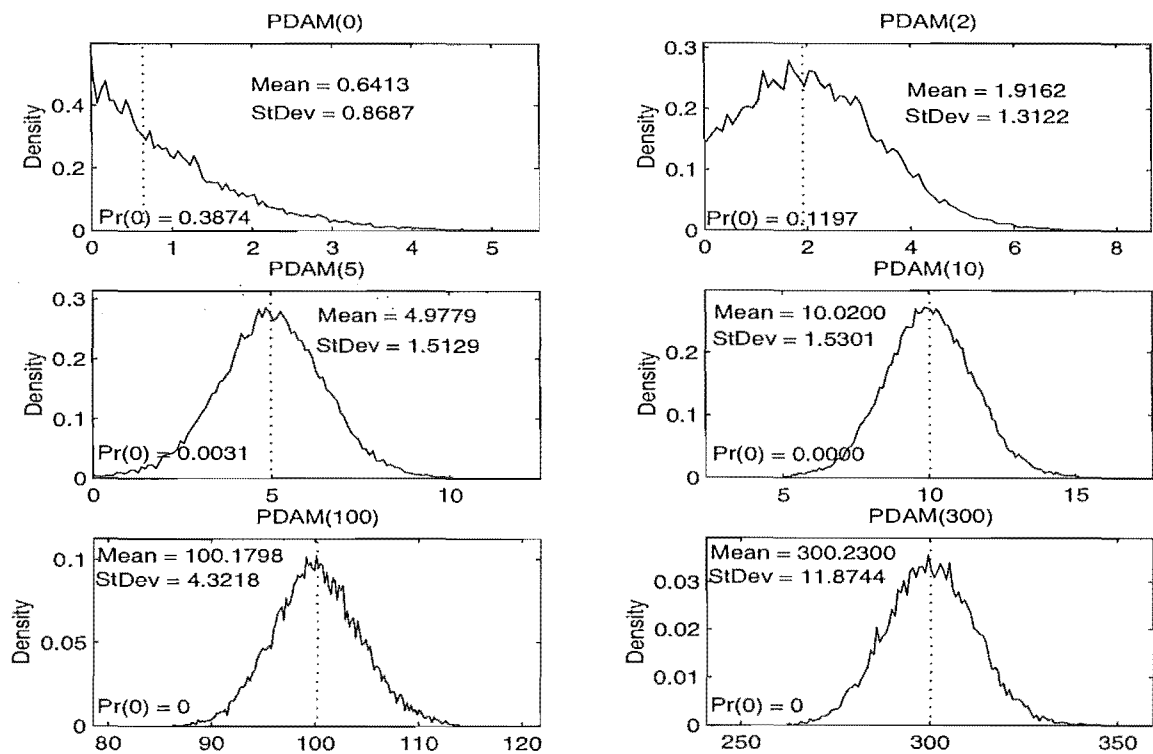
Figure 6.2: The predictive distribution of assay measurements at the following concentrations: 0, 2, 5, 10, 100 and 300 for the NMCH data (the horizontal axis is the concentration, the mean of PDAM(x) is the dotted vertical line).

6.4.2.1   Comments

As the concentration x increases the spread of the PDAM increases. At values of x close to zero, PDAM(x) is right skewed. At large values of x the skewness dissipates, the probability of zero concentration vanishes and the distribution takes on a very symmetric appearance. As would be expected, PDAM(x) is centered at a value very close to x when x is large.

Chapter 7

Machine assays and quality diagnostics for batches of assays

In this chapter the situation in which batches of assays are available for analysis is the main focus of discussion. In the first part of the chapter, Sections 7.1 to 7.3, the statistical analysis of a machine assay, an assay performed by a machine that has a statistical algorithm embodied in it, is discussed. The main point of emphasis arising from this discussion is that data from other homogeneous assays are needed in order to incorporate the error in the model fit in an assessment of the assay's quality. In Section 7.4, measures of MDC and precision are derived for a batch of homogeneous manual assays using the techniques discussed in earlier chapters. Note that, the term "manual assay" is used in this chapter to distinguish these assays from machine assays and that earlier use of the word assay in this thesis has meant manual assay.

7.1 Machine assays

The fundamental characteristic of a machine assay is its black box behaviour. Standard and unknown samples are supplied to the machine and estimates of concentration are returned. The intermediate steps of the analysis are not included in the output. In many instances the raw response measurements are not even provided. Machine assays thus give the appearance that they directly measure the concentration.

When using a machine assay, a practitioner merely indicates which samples are standards and which are unknowns. Presently, since there is no facility for obtaining a common estimate for groups of replicated samples, in a machine assay the fitted concentrations for every individual sample are returned. The response measurements and the fitted assay model are not revealed.

The following consequences are noteworthy:

1. The practitioner must combine the estimate for each replicate into a single estimate. The resulting estimate is unlikely to be as efficient as a single estimate based on all the data since the information in each of the replicates is not pooled.

2. Since the assay model is not disclosed, there is insufficient information to calculate reliable measures of assay performance such as the MDC and the precision profile.

Another concern with machine assays is that a limitation on the models that can be fitted to the assay will necessitate the use of a model that may be inappropriate. With the intermediate analysis not being provided, detection of an inappropriate model will also be difficult.

It is clear that machine assays as they currently exist, are unable to provide an analysis of an assay of the quality or completeness that can be achieved manually. The output from a machine assay should be amended so that it emulates the output from a manual assay. The raw data and full details of the model fit should at least be given. A further improvement would require machine assays to calculate inferences of choice, measures of MDC and precision profiles upon request from the practitioner.

7.2   Assessing the quality of a single machine assay - intrinsic assay error

Despite the above remarks, there are instances in which the difference in the estimates of the unknown concentrations between a manual and a machine assay are minimal. However, there is a large difference in the ability to ascertain the quality

of the assay. Fundamentally, machine assays do not yield the information needed to make an accurate assessment of an assay's quality.

It is clear from the onset that there is no practical way of estimating the MDC for a machine assay. This is because the output from a machine assay does not allow inferences about the response, y, or the concentration, ω, of an unseen sample to be calculated. However, as described in Section 6.2.2 one measure of quality that can be calculated for a machine assay is its intrinsic error (see Section 1.2.4 for definition of intrinsic assay error). The method applied to a machine assay is briefly reviewed.

Sadler et al. (1988) proposed a method for the calculation of intrinsic assay error that can be applied to machine assays. The method, known as the empirical method, is the only method presently available which can calculate intrinsic assay error for a machine assay. The approach is to fit a parametric model that relates the estimated concentrations to the actual concentration. A generalized version of the model they proposed is

$$\hat{\eta}_{ij} \sim N(\eta_i, g(\eta_i, \alpha)),$$

where $\hat{\eta}_{ij}$ is the fitted concentration for the jth replicate of the ith solution and $g(\eta_i, \alpha)$ is the variance function with parameter α. The intrinsic assay error at concentration η_i is then calculated as either the standard deviation or coefficient of variation at η_i.

This model asserts that estimates of concentration are unbiased. When this assumption holds, the variance function equals the squared error of an estimate of concentration. For the reasons discussed in Section 6.4.1.1, $g(\eta_i, \alpha)$ is not necessarily a monotone function of η_i.

It is not necessary to be restricted to the above model. In fact any model can be used to describe the relationship between $\hat{\eta}$ and η. If a prior distribution is specified for α, then the model can be fitted using Bayesian methods.

7.3 The precision profile for batches of homogeneous machine assays

In Section 6.2.2 it was pointed out that given the data from a single assay, the empirical method can only incorporate the intrinsic assay error in the calculation of

precision. This is because all of the estimates of concentration are derived from the same fitted assay model; hence, the error in the fitted assay model is ignored or in other words, only error due to the randomness of the response is modelled. In this section it is seen that with data from a batch of homogeneous machine assays the error in the fitted assay model may be incorporated in the calculation of precision.

Recall that a batch of homogeneous assays is a group of assays with identical experimental designs that are designed to analyze the same substance using the same reagents and procedures. If data from a batch of homogeneous assays were available, then the error in the fitted assay model may be considered by calculating a precision profile for the batch of assays.

The estimated concentrations provide the basis from which the precision profile for a machine assay is calculated. Therefore, the only way to reliably calculate a precision profile for a machine assay is to make the error in the fitted assay model an inherent feature of the estimated concentrations. This requires that some samples are analyzed by more than one assay in the batch.

Such data are typically provided by the standard and QC (quality control) samples. The same batches are used on each assay in succession until the QC specimen indicate that the assay is out of control. Then fresh standards and QC specimen are made up and a new batch of assays is begun.

7.3.1   Model and method

If singleton measurements are made within each assay the empirical method can be applied to these data as if a single assay were being analyzed. When response measurements are replicated within an assay the estimates of concentration are not marginally independent (i.e. over the batch of assays). The estimates of concentration made on the same assay are positively correlated. Any valid calculation of precision must consider this fact.

For the purpose of illustration, suppose that the fitted concentrations are normally distributed. The approach can be easily generalized to a general distribution. Let $g(\eta_{hi}, \alpha)$, where α is an unknown parameter, denote the variance of the calibrated concentration for the ith sample in the hth assay. Define $\rho(\eta_{hi}, \tau)$, where τ is an unknown parameter, to be the correlation between estimates of η_{hi}.

If there are two replicates for the ith sample in the hth assay, the contribution

to the sample likelihood function is

$$f(\hat{\eta}_{hi} \mid \eta_{hi}, \alpha, \tau) = N\left(\begin{pmatrix} \eta_{hi} \\ \eta_{hi} \end{pmatrix}, g(\eta_{hi}, \alpha) \begin{pmatrix} 1 & \rho(\eta_{hi}, \tau) \\ \rho(\eta_{hi}, \tau) & 1 \end{pmatrix}\right), \qquad (7.1)$$

where

$$(\eta_{hi}, \alpha, \tau) \in \{(\eta_{hi}, \alpha, \tau) : \eta_{hi} \geq 0,\ g(\eta_{hi}, \alpha) \geq 0,\ 0 \leq \rho(\eta_{hi}, \tau) \leq 1\}.$$

When the same sample is analyzed by more than one assay, some components of the unknown concentration parameters are obviously redundant. Note also that, the known values of the concentration of any standards can be substituted into the sample likelihood function, thus eliminating the need to estimate these concentrations.

If the final estimate of concentration is given by the mean $\hat{\bar{\eta}}_{hi} = \sum_{j=1}^{r_{hi}} \hat{\eta}_{hij}/r_{hi}$, the estimated total or marginal variance of $\hat{\bar{\eta}}_{hi}$ is

$$\text{var}(\hat{\bar{\eta}}_{hi}) = g(\eta_{hi}, \hat{\alpha})(1 + \rho(\eta_{hi}, \hat{\tau}))/r_{hi},$$

where $\hat{\alpha}$ and $\hat{\tau}$ are estimates of α and τ respectively and r_{hi} is the number of replicates for the ith sample in the hth assay. For a given concentration, this expression estimates the total variation in the fitted concentrations and so can be used to form the precision profile, i.e. the procedure is just to plot $g(x, \hat{\alpha})(1 + \rho(x, \hat{\tau}))/r$, or some function thereof, against x for some degree of replication r. It is clear that $\text{var}(\hat{\bar{\eta}}_{hi})$ will be under-estimated if no account is taken of the positive correlation between the estimates of η_{hi}.

7.3.2 Simplified calculation

An alternative method of calculating precision is to use the mean estimates as the responses and proceed as if there were no replication within each assay (assuming that the number of replicates is the same for all samples). The fitted concentrations are independent conditional on η and so the model

$$\hat{\eta}_{hi} \sim N(\eta_i, g(\eta_i, \alpha)) \qquad (7.2)$$

can be used.

The disadvantage of this approach is that only standard and QC samples can be used. This is because taking the mean leaves only one degree of freedom with each unknown sample and this is used to estimate the unknown concentration. Another casualty is the information contained in the replicated responses, from the same assay, of the standard samples and the QC specimen. Since the standard and QC samples typically make up only a small proportion of the samples, a very large pool of information will not be used in the analysis; hence, the procedure will not be fully efficient.

7.3.3 Discussion

In machine assays, QC and unknown samples are often not replicated within an assay. However, the responses for the standards are usually measured at least in duplicate. If the estimated concentrations are normally distributed, the model in (7.1) could be used for the standards and the model in (7.2) for the QC samples. In practice, the model in (7.2) tends to be used for both the standards and QC samples despite this not being a fully efficient procedure.

In some practical situations, the standards are not used in the calculation of the precision profile. This is due to a concern that standards have different properties than unknowns, in particular variability. This concern arises from the fact that the standards are made up under controlled laboratory conditions rather than being actual real-life samples. This issue is not considered relevant to this thesis. If standards and unknowns did have different properties, then this should be reflected in the assay model, not just in the calculation of a precision profile.

7.4 Quality assessment for batches of manual assays

When data from a batch of machine assays are used to form a precision profile, the result is an average or batch-wide measure of the quality of the assays. When the assays are manual assays batch-wide measures of both MDC and precision can be developed.

In this section measures of MDC and precision are developed for a batch of

homogeneous manual assays. These are based on the definitions of precision and the MDC used in Chapters 5 and 6 respectively. The purpose here is to review the calculations that can be made given the output from a batch of homogeneous manual assays.

For each assay in a batch of assays, measures of MDC and precision may have been calculated. A quick and easy method of combining the measures for each assay is to average them. However, this is ad-hoc because there is no assurance that such averages will have desirable properties. A more reliable approach is now described.

7.4.1   Annexing a batch of assays

A batch of assays is converted to a single assay through the random assignment of the unseen sample (i.e. the sample the independent response y is generated from) to one of the assays for measurement. In the calculation of measures of MDC and precision, there are now two unknowns, namely the assay the hypothetical response y is measured on and the unknown concentration.

Since a batch of assays is involved, the notation of Section 2.2.2 is used. Recall that, $f(y \mid x, \beta_h, \theta_h)$ is the likelihood function for an independent response from a sample with concentration x in the hth of k assays. The data from the hth assay are denoted by (Y_h, X_h), $\hat{\beta}_h$ and $\hat{\theta}_h$ denote the estimates of β_h and θ_h and $(Y, X) = \{(Y_h, X_h), h = 1, \ldots, k\}$.

Let $(\beta, \theta) = \{(\beta_h, \theta_h), h = 1, 2, \ldots, k\}$ and $(\hat{\beta}, \hat{\theta}) = \{(\hat{\beta}_h, \hat{\theta}_h), h = 1, 2, \ldots, k\}$. Let p_h be the probability that the unseen sample is analyzed by the hth assay. Then the probability distribution for y at concentration x is

$$f(y \mid x, \beta, \theta) = \sum_{h=1}^{b} p_h f(y \mid x, \beta_h, \theta_h). \tag{7.3}$$

Typically, p_h equals the proportion of samples out of the total number of samples that are analyzed by the hth assay. If each assay analyses the same number of samples then $p_h = 1/k$.

7.4.2 Definition of batch-wide MDC and precision

The batch-wide equivalent of a measure of MDC or precision is obtained by evaluating the measure of precision using (7.3) for the distribution function of the independent response measurement and the model fits of the individual assays for the overall model fit.

A feature of this definition is that each assay in the batch is fitted using just the data observed for that assay. The model fit for each assay is therefore independent of the model fits for the other assays in the batch. This coincides with the way in which assays are fitted in practice. An alternative approach would be to use the grand immunoassay model of Section 4.6 to determine the overall model fit. Under this model, data from the current assay is supplemented with information from the preceding assays in the batch. If the model assumptions are valid, resultant inferences will be more efficient. However, as the fitted assay models would no longer be independent, the calculation of batch-wide measures of MDC and precision would be more arduous.

7.4.3 Measures of batch-wide MDC

7.4.3.1 Frequentist measures

Under a frequentist framework, the MDC for the hth assay can be expressed as the value of x for which

$$m(x, \hat{\beta}_h) = c(\alpha, z, Y_h, X_h), \tag{7.4}$$

where $c(\alpha, z, Y_h, X_h)$ is an approximate $1 - \alpha$ level critical value of a prediction interval of an independent response y at concentration $z \in \{0, x\}$. For the CL $z = 0$ and

$$c(\alpha, 0, Y_h, X_h) = m(0, \hat{\beta}_h) - t_{(\alpha, \nu)} \sqrt{v(m(0, \hat{\beta}_h), \hat{\theta}_h)/r + \widehat{\mathrm{var}}(m(0, \hat{\beta}_h))}$$

and for the DL

$$c(\gamma, z, Y_h, X_h) = c(\alpha, 0, Y_h, X_h) - t_{(\gamma, \nu)} \sqrt{v(m(z, \hat{\beta}_h), \hat{\theta}_h)/r + \widehat{\mathrm{var}}(m(z, \hat{\beta}_h))}.$$

If the MDC for the hth assay is based on backfitting $c(\alpha, z, Y_h, X_h)$ to the concentration axis, then the batch-wide MDC should be based on backfitting $c(\alpha, z, Y, X)$,

the α level critical value for the annexed batch of assays, to the concentration axis. Recall that (Y, X) is taken to mean $\{(Y_1, X_1), \ldots, (Y_k, X_k)\}$ in dicussions of batches of assays.

Under a frequentist framework, calculation of $c(\alpha, z, Y, X)$ is impossible without knowledge of β and θ. Since y has a mixture distribution, possibly having more than one mode, accurate approximations of prediction intervals and error probabilities will even be difficult to find.

The following measure of frequentist batch-wide MDC may be used instead of the above. Define as the MDC the value of x for which

$$\sum_{h=1}^{k} p_h \alpha_h = \alpha,$$

where α_h solves

$$m(x, \hat{\beta}_h) = c(\alpha_h, z, Y_h, X_h).$$

Under this definition, the MDC is interpreted as the value of x at which $1 - \alpha$ is the average probability level of the prediction intervals defined by (7.4) for the assays in the batch.

The above system of equations consists of $k + 1$ equations and the $(k + 1)$ unknowns, (x and $\{\alpha_1, \ldots, \alpha_k\}$). Hence, a unique value for the MDC can be obtained. Allowing the probability levels of (7.4) to vary over the individual assays, allows the constraint $m(x, \hat{\beta}_h) = c(\alpha_h, z, Y_h, X_h)$ $\forall h$ to be satisfied. Note that, if this constraint were not enforced, the measure of MDC being considered here would not have a unique value.

For the CL, the approximate batch-wide MDC is the value of x for which

$$\sum_{h=1}^{k} p_h \alpha_h = \alpha,$$

where α_h solves

$$m(x, \hat{\beta}_h) = m(0, \hat{\beta}_h) - t_{(\alpha_h, \nu)} \sqrt{v(m(0, \hat{\beta}_h), \hat{\theta}_h)/r + \widehat{\operatorname{var}}(m(0, \hat{\beta}_h))}.$$

7.4.3.2 Bayesian measures

In a Bayesian framework batch-wide measures of the MDC can be computed directly from the definition. For example, the batch-wide response level MDC is the smallest

concentration x for which

$$
\begin{aligned}
1 - \alpha &= \mathrm{pr}(y < m(0, \beta) \mid x, Y, X) \\
&= \sum_{h=1}^{k} p_h \mathrm{pr}(y < m(0, \beta_h) \mid x, Y_h, X_h).
\end{aligned}
$$

Recall that $\mathrm{pr}(y < m(0, \beta_h) \mid x, Y_h, X_h)$ is the probability that determines the response level MDC on the hth assay.

For the batch-wide probability MDC let $\mathcal{Y}_C$ denote the set of values of y for which

$$
\begin{aligned}
\pi(\eta > 0 \mid y, Y, X) &= \sum_{h=1}^{k} p_h \mathrm{pr}(\eta > 0 \mid y, Y_h, X_h) \\
&\geq 1 - \alpha.
\end{aligned}
$$

Then x^{pr} is the smallest value of x for which

$$
1 - \gamma = \mathrm{pr}(y \in \mathcal{Y}_C \mid x, Y, X).
$$

The batch-wide discriminant MDC is defined similarly.

Unlike the frequentist framework, it is generally the case that the Bayesian definitions can be evaluated. The computational requirements for Bayesian batch-wide measures of MDC are not much greater than for the evaluation of the MDC for a single assay.

7.4.4 Measures of batch-wide precision

7.4.4.1 Frequentist measures

Frequentist measures of precision are based on the standard error of $\hat{\omega}(y, Y, X)$, the estimated concentration for the response y. Let $E[\hat{\omega} \mid x, X_h]$ and $\mathrm{var}[\hat{\omega} \mid x, X_h]$ denote the respective expectation and variance (standard error) of $\hat{\omega}$ conditional on y being measured by the hth assay. Then for a batch of assays

$$
\begin{aligned}
\mathrm{var}[\hat{\omega} \mid x, X] &= E[\mathrm{var}[\hat{\omega} \mid x, X_h]] + \mathrm{var}[E[\hat{\omega} \mid x, X_h]]] \\
&= \sum_{h=1}^{k} p_h \left[\mathrm{var}(\hat{\omega} \mid x, X_h) + (E[\hat{\omega} \mid x, X_h] - \bar{E}[\hat{\omega} \mid x, X_h])^2 \right],
\end{aligned}
$$

where $\bar{E}[\hat{\omega} \mid x, X_h] = \frac{1}{k} \sum_{h=1}^{k} E[\hat{\omega} \mid x, X_h]$. Estimates of $E[\hat{\omega} \mid x, X_h]$ and $\mathrm{var}[\hat{\omega} \mid x, X_h]$ can be obtained using the analytical method on each assay or otherwise.

These are substituted into the above expression to yield an approximate expression for $\text{var}[\hat{\omega} \mid x, X]$. The approximate precision at concentration x for the batch of assays is then evaluated. If the estimates of $\hat{\omega}$ are unbiased, or are assumed as such, then the variance of $\hat{\omega}$ reduces to the mean of the variances for the individual assays.

7.4.4.2 Bayesian measures

PDAM precision is used to illustrate the calculation in the Bayesian sense. In this case, the precision at concentration x for a batch of assays is based on the PDAM at concentration x when (7.3) is the distribution function of y. PDAM(x) is determined from the predictive distribution of the response y at concentration x, which is given by

$$
\begin{aligned}
p(y \mid x, Y, X) &= \sum_{h=1}^{k} p_h \int_{\mathcal{R}(\theta_h)} \int_{\mathcal{R}(\beta_h)} f(y \mid x, \beta_h, \theta_h) \pi(\beta_h, \theta_h \mid Y_h, X_h) \, d\beta_h \, d\theta_h \\
&= \sum_{h=1}^{k} p_h p(y \mid x, Y_h, X_h)
\end{aligned}
$$

It follows that PDAM(x) is a weighted average of the PDAM at concentration x for each individual assay. If $\pi(\omega \mid x, Y_h, X_h)$ denotes the PDAM for the hth assay then

$$
\text{PDAM}(x) = \pi(\omega \mid x, Y, X) = \sum_{h=1}^{k} p_h \pi(\omega \mid x, Y_h, X_h).
$$

In fact, whenever precision is based on a moment of $\pi(\omega \mid x, Y, X)$ about x, batch-wide precision is the weighted average precision of the individual assays.

7.4.4.3 An important point

A feature of this Bayesian measure of batch-wide precision is that it is the average precision of the assays. On the other hand, the value of a frequentist measure of batch wide precision is greater than the (weighted) average precision of the assays. The second term in the frequentist expression for batch-wide precision accounts for the variation between assays. In a frequentist measure of precision for a single assay, this term is not present; hence, the magnitude of the average precision is smaller. This is a further illustration of the fact that variance based frequentist measures of precision do not account fully for the error in the estimated concentration. A better criterion is mean squared error.

Chapter 8

Summary

The main observations and findings of the thesis are summarized in this chapter. Some general comments concerning the application of the methods that have been developed to other statistical analyses and some thoughts on future research are included.

8.1   Frequentist variance function estimation

8.1.1   ExMML and ExREML

In this thesis two new procedures (ExMML and ExREML) for the estimation of a variance function in an assay arose by observing that Raab's MML (modified maximum likelihood) procedure and the REML (restricted maximum likelihood) procedure could be improved. The ExMML and ExREML procedures incorporate the best aspects of the MML and REML procedures; i.e. by pooling information from both standard and unknown samples as in MML, and by using the fitted regression function to estimate the mean responses for the standards as in REML.

The methods were compared using simulated data for two variance functions and a range of parameters values. The mean squared errors of the estimated parameters were smaller for the extended estimators (ExMML and ExREML) than for either MML or REML. The average scaled L_2 distance between the actual variance function and the fitted variance function was also smaller for the extended estimators. These numerical results reinforce the notion that the extended estimators perform better

179

because they extract more information from the data.

For general use, the extended estimators are superior to both MML and REML. It should be noted that in practice it is usually the case that there are many more standards than unknowns. When the response measurements are replicated the extended estimators and the MML estimate will be virtually indistinguishable. This validates the use of MML in this circumstance. However, if the responses for the unknowns are only measured in singleton then REML will perform better than MML and the extended estimators reduce to the REML estimate.

Since the extended estimators have simultaneous access to all of the information in the data, they can also be expected to perform better (if only slightly) than any weighted average of the MML estimate and the REML estimate (see Section 3.4.1).

8.1.2   Practical application

Among the class of likelihood based estimators of the variance function parameters, the ExMML and ExREML procedures may not be able to be surpassed. Furthermore, they should appeal to both frequentist and Bayesian statisticians.

The extended estimators should appeal to frequentist statisticians because of their properties concerning consistency. Standard frequentist procedures for estimating a variance function in a regression of the mean response do not take account of the estimation of the mean responses. These procedures must be adjusted if they are to produce estimates of the variance function parameters that will be consistent in even simple models (Raab (1981)). This is true in all such problems where the dimension of the parameter space for the mean responses is proportional to the number of observations. The extended procedures should be used instead of standard frequentist estimates if consistency is desired.

On the Bayesian side, the ExREML procedure is equivalent to an analytical approximation of a non-informative Bayesian analysis. This provides a platform upon which results can be interpreted.

Although minor improvements to the quality of variance function estimates may not lead to much better point estimates of the mean function parameters, such improvements can significantly affect interval estimates and other inferences that depend on the error of estimation. In Davidian (1989) it was shown that the accuracy and reliability of frequentist measures of the MDC are very dependent on the error

in the fitted variance function. If the variance function is itself of interest, then it should be estimated as accurately as possible. In such circumstances the use of either the ExMML or ExREML estimator is clearly warranted.

The ExMML and ExREML procedures can be applied to a broader class of variance functions from those typically used in immunoassay. Since the form of the adjusted likelihood remains the same, these procedures are in fact applicable to any analysis where a variance function is estimated and a proportion of the observations have missing values for some independent variables. This includes analyses where the variance function depends directly on explanatory variables as well as through the mean function.

8.1.3 Further work

The development of exact procedures (i.e. those void of approximations) with desirable properties such as consistency, for the estimation of a variance function is a largely untouched area. Within the frequentist framework, it is unclear how to proceed. Substituting the full likelihood function with the conditional likelihood function for the variance function parameters given the estimates of the mean responses is a natural way of accounting for the estimation of the mean responses. However, since the conditional likelihood function can only be calculated explicitly in a few simple cases, the method cannot be applied generally.

These difficulties are overcome by the Bayesian paradigm. The procedure is to apply a full Bayesian analysis to the data and then to use numerical simulation or otherwise to evaluate the estimate to the required level of accuracy. Although this approach is more time consuming than solving a few estimating equations, the procedure is always clear and well defined. The availability of the necessary computational resource is the only concern, but this is generally an issue in high dimensional problems only.

A preliminary study of the frequentist properties of Bayesian estimates of the variance function indicated that in the case of the power variance function and the non-informative prior $\pi(\beta, \theta, \mu^U) \propto 1/\tilde{\theta}_1$, the posterior mean of θ_2 appeared to perform better than the ExREML estimator. In one instance a reduction of 9% in the squared error of $\hat{\theta}_2$ was noticed. These results, however, are still very much at the preliminary stage. Conclusive results for this and other variance functions have

yet to be determined.

8.2   Bayesian analysis of an assay

A Bayesian analysis of an assay is a challenging endeavour which offers many rewards. The most outstanding reward is that it provides a formal procedure for the solution of all inference problems. Unlike frequentist methods, any inference problem can be reduced to one of computation. With the recent evolution in Bayesian computation, the numerical evaluation of problems once considered intractable can now be solved in reasonable periods of time.

A second advantage is that a Bayesian analysis of an assay provides a formula for exact computation of the model fit and related inferences. In Chapter 4 the formulae are given for various posterior and predictive distributions pertinent to the analysis of an assay. From these distributions any inference of interest can be evaluated. As alluded to in Section 8.1.3, the Bayesian approach is perhaps the only means of calculating exact inferences about a variance function.

A final advantage is that Bayesian analysis enables prior knowledge of unknown parameters to be incorporated into the model. Practitioners tend to carry out many assays per week and so are likely to have expert knowledge of entities such as the distribution of the concentrations of the substance being tested throughout the population of eligible subjects and the behaviour of the mean and variance functions. This knowledge has the potential to improve the analysis.

A generic model for a Bayesian analysis of an assay has been described and procedures for the calculation of inferences of interest have been suggested.

8.2.1   Numerical implementation

Numerical evaluation of a Bayesian analysis of an assay has proven to be surprisingly feasible. Several techniques have been proposed for making the M-H (Metropolis-Hastings) algorithm an efficient means of fitting the model in an assay. These include the use of structural re-parameterizations, based on the Fisher information matrix or otherwise, to lessen the a posteriori dependence between the parameters and using separate kernels for groups of parameters that are independent or nearly so.

Transformations based on the Fisher information matrix have the potential to uncouple complex interactions between parameters in addition to statistical dependence making them very powerful tools. The use of different kernels for distinct groups of parameters increased the probability of a change in the state of the chain in each step causing the chain to explore the posterior more rapidly.

One of the main observations noted from the trials and tribulations with the implementation of the M-H algorithm was the harmful effect of dependence between the parameters. If statistical dependence between parameters was not reflected in the candidate generating density, then the algorithm tended to perform very badly as the chain took a long time to move about the posterior density. This necessitated the use of multivariate probing distributions for all groups of parameters between which there was significant statistical dependence. The inverse hessian based approximation of the posterior covariance matrix for a group of parameters proved to be an adequate approximation of their posterior covariance.

Another factor that contributed to the success of the M-H algorithm on real and simulated NMCH data was that after transformation, the posterior distribution of the parameters was reasonably symmetric and had virtually no mass close to boundaries of the parameter space. This meant that the multivariate t distribution was a good approximation of the posterior distribution being sampled from.

8.2.2 Further Work

There is much work to be done on the use of Bayesian methods in the analysis of an assay. For example, suitable non-informative and informative priors need to be developed for the variety of situations encountered in practice.

Some comparisons between existing methods and the Bayesian methods developed in this thesis for analyzing an assay were made herein. As mentioned earlier a comprehensive investigation of Bayesian estimates of the variance function is warranted.

8.3   Minimum detectable concentration

Although it has elusive definitions depending on the perspective, the MDC (minimum detectable concentration) is an important diagnostic of an assay's performance.

Many measures of the MDC of an assay have been presented in this thesis. Anomalies with the interpretation of some incumbent measures have been resolved and several new measures have been developed.

8.3.1   Bayesian measures of MDC

The notion of Bayesian inference facilitates development of measures of MDC that emulate decision making criteria used in practice. This is possible because the value of quantities that would be calculated in practice can be predicted using Bayesian analysis. The procedure is just to express such quantities as a function of an independent response y and then at the last step average these quantities with respect to the predictive distribution of y. Only under the Bayesian paradigm is a calculation of this form possible.

There is also a strong connection to quantities that would be evaluated if the assay model were known. Any quantity that would be evaluated if the assay model were known has a counterpart that can be evaluated when the model is unknown. The procedure is to average the quantity with respect to the joint posterior distribution of all the unknown parameters in the evaluation of the MDC.

Under the frequentist paradigm, the lack of a formal procedure for the solution of all problems of inference inhibits the calculation of the MDC under either of the above principles. The problem is two-fold. Firstly, a procedure with the required probability level must be determined. Secondly, the procedure must not involve approximations if it is to accurately reflect the quality of the assay. The impossibility of finding a pivotal quantity based on a minimal sufficient statistic for the model parameters renders both impossible. Frequentist measures of MDC must be estimated by approximate confidence procedures. The direct correspondence to and meaning of the measure of MDC is jeopardized by the use of these approximations. The MDC of an assay is an example of a difficult problem for which the Bayesian approach is a very good computational tool.

The Bayesian paradigm also enables uncertainty in the response to be transferred to uncertainty in the concentration. This enables measures of the MDC to be expressed in terms of the distribution of an unknown concentration. It is conjectured that it is in these terms that practitioners naturally think of the MDC.

8.3.2 Numerical results

When the value of the variance function parameter is the only difference between two assays, the performance of a measure of MDC is evaluated by the probability that the MDC will identify the better assay. The best performing measures of MDC were those for which detection is based on discrimination from a sample with zero concentration, namely the detection limit and the discriminant MDC.

The numerical simulations in Section 5.4.1.4 indicated that the RQ (relative quality) values for the detection limit and other frequentist measures of MDC were misleading when the amount of data changed in some way. When the number of unknowns or the number of replicates changed, the frequentist measures of MDC were on the whole less sensitive than the Bayesian measures. However, when the number of standards changed, they were more sensitive than Bayesian measures. Since the Bayesian measures are based on exact calculation, it is reasonable to assume that these behave the way a measure of the MDC should. It is therefore concluded that in these situations the frequentist measures of MDC provide incorrect assessments of the relative quality of rival assays.

The behaviour of the Bayesian measures of MDC indicated that the number of unknown samples and the number of replicates have a significant influence on the value of the MDC. A substantial increase in the number of unknowns or an increase in the degree of replication leads to a more precise fit of the variance function. This in turn reduces the MDC. Due to the dependence of the variance function on the mean function, the sample variances contain information about the mean function in addition to the variance function. While adding to the disparity in the quality between assays that differ in the number of unknowns or the degree of replication, this lessens the influence of the standards on the quality of an assay.

8.3.3 Generalization

Although the MDC uses zero or null concentration as a benchmark, this can be generalized to any concentration. The resultant diagnostic would reflect the ability of an assay to detect concentrations significantly greater than the benchmark concentration. Further generalization would lead to quantities that assess an assay's ability to detect concentrations significantly less than or different from a benchmark.

(See Sadler, Murray and Turner (1992) for discussion of the relationship of this type of measure and the precision profile.) In this way the measures of MDC developed in this thesis can be applied to any assay, not just those where the objective is the detection of a substance.

8.4   Precision profiles

The precision profile is another tool for measuring and displaying the quality of an assay. It may be reported in conjunction with measures of MDC to give an overview of an assay's quality.

The calculation of precision in the frequentist sense, like MDC, encounters difficulties because the average over the sample space must be computed. To evaluate frequentist measures of precision, approximations are needed. Alternative approximations, which may offer better accuracy in some situations, have been proposed, but these procedures are still inexact.

Two general measures of precision that use the notion of Bayesian inference were developed in Chapter 5. These were called ES (estimator specific) and PDAM (predictive distribution(s) of assay measurements) precision. In Section 6.3.5.2 it was shown that the mean squared error of the PDAM is equivalent to a natural measure of ES precision. This is a key result in the sense that it characterizes the two forms of Bayesian precision.

The advantages these have over frequentist measures parallel the advantages that Bayesian measures of MDC have over their frequentist counterparts. Exploratory investigations not reported in the thesis have revealed that frequentist precision profiles suffer the same deficiences as the frequentist measures of MDC. They are overly sensitive to changes in the number of standards but do not reveal the full importance of unknowns or the degree of replication.

The fact that much of the information in the data is not used in the calculation of frequentist precision profiles was apparent when they were plotted along side the Bayesian precision profile for the NMCH data (Figure 6.1). This was seen from the fact that the Bayesian precision profiles offered a more optimistic view of the assay than the comparable frequentist precision profiles.

The PDAM (predictive distribution(s) of assay measurements) is a significant

development in itself. These distributions give a complete representation of the accuracy and precision of an assay's measurement of an unknown concentration. The use of the PDAM is clearly not confined to the development of precision profiles. They can be used as the basis for any quality definition involving the measurement of an unknown concentration. A PDAM based measure of MDC is one application.

8.5 Machine assays and quality diagnostics for batches of assays

The embodiment of statistical algorithms in a machine that analyses an assay has the potential to significantly simplify the analysis of the assay. Unfortunately existing machines do not analyze the data as efficiently as they could, nor do they provide the level of output needed to account for the error in the fitted assay model. This makes it impossible to reliably assess the quality of the assay. It would be desirable for the interface of a machine assay to be amended so that replications of the same sample can be grouped for more efficient estimation and full details of the analysis can be obtained upon request.

The situation improves from the point of view of the assessment of an assay's quality if a batch of homogeneous assays have been analyzed. The estimates of concentration for the same sample on different assays contain the error due to fitting the assay model. A new method of calculating the precision profile from these data has been described. The precision profile that results is known as a batch-wide precision profile since it is based on the data from the whole batch of assays.

Measures of the batch wide MDC and batch wide precision profiles for frequentist and Bayesian analyses of manual assays have been described. The generalization of the measures for a single assay to a batch of assays is more easily accomplished when the analysis of the assays is Bayesian than when it is frequentist. Unlike frequentist measures, mathematical approximations are not needed to develop the Bayesian measures. Accordingly, it is envisioned that the Bayesian measures will provide a more accurate assessment of the overall quality of a batch of assays.

When the assays are homogeneous in the sense that each assay can be considered a re-sampled version of any other assay, batch-wide measures of MDC and precision reflect the average performance of the assays in the batch. These measures

are important in their own right because, compared to measures developed on a single assay, they provide superior assessments of the quality of the analytical and statistical (or in other words, experimental) designs used in the assays.

8.6   General extensions

The techniques developed in this thesis have application beyond the analysis of an assay.

8.6.1   The calibration problem

The extension of methods developed in this thesis to the calibration problem is immediate. Furthermore, if the concentrations of the standards are themselves measured with error, the general calibration problem becomes embodied in the model for an assay.

Let the observed or measured value (X_i) of the concentration of the ith standard vary about the true value (η_i) according to the probability distribution function

$$X_i \sim f(X_i \mid \eta_i, \alpha), i \in S,$$

where α is an unknown vector of parameters. It is likely that errors in the measurement of the standard concentrations would be independent of the response counts. The likelihood function then becomes

$$f(Y \mid \eta, \beta, \theta) f(X \mid \eta^S, \alpha).$$

It would be expected that X will be tightly distributed about η^S, otherwise there would be little point in having standards. The model used for the estimation of relative potency in a radioimmunoassay with sequential dilution errors by Racine-Poon et al. (1991) is of this form. In this case the observed concentrations are perturbations of the actual concentrations of the standards due to errors incurred when performing the dilutions.

A further extension is the case where the response and concentration measurements were correlated. In this situation, the likelihood function for the ith standards would be a multivariate distribution function of (Y_i, X_i).

In all of these instances a prior distribution for any additional unknown parameters only needs to be specified and all of the methods developed in this thesis can be applied.

8.6.2 Ranking and selection

The type of inference at the basis of the response level MDC is not typically seen in Bayesian analysis. The population of subjects with zero concentration and the population of subjects with some positive concentration are being compared in terms of the probability that the predicted response for a subject with positive concentration exceeds the mean response for the subjects with zero concentration.

A general setting for this type of problem is that of determining which of k lots is best. Let D denote the observed data, θ denote the vector of means for the lots and z a vector of future measurements made on the lots. The selection strategy is often based on the posterior probabilities

$$\mathrm{pr}(\theta_i > \theta_j \mid D) \text{ for } i \neq j$$

or the predictive probabilities

$$\mathrm{pr}(z_i > z_j \mid D) \text{ for } i \neq j.$$

If the comparison made in the response level MDC is incorporated in this problem the predictive probabilities

$$\mathrm{pr}(z_i > \theta_j \mid D) \text{ for } i \neq j \tag{8.1}$$

form the basis of the selection strategy.

It is difficult to conceive of a scenario where one would be interested in using using (8.1) for ranking or selection. However, the idea behind the response level MDC may be able to be applied in a more general way to the field of ranking and selection.

8.6.3 Bayesian interim analysis

The calculation of x^{pr} and x^{ds} resembles the calculations of Seymour Geisser in the field of Bayesian interim analysis (see Geisser (1992) and Geisser (1993)). The connection is now explored.

Suppose that the interim point of an assay occurs after the data (Y, X) has been observed and there is just one more unknown sample to analyze. Let y denote the response for this sample and ω the unknown concentration. At the end of the assay, inference about ω will be based upon $\pi(\omega \mid y, Y, X)$. Suppose that one is interested in whether or not there will be sufficient evidence at the completion of the assay to conclude that $\omega > 0$ when in fact the actual concentration is x. Let $1 - \alpha$ be the critical value that $\mathrm{pr}(\omega > 0 \mid y, Y, X)$ must exceed in order to reach the conclusion that $\omega > 0$. Then the predictive probability of the event $I(y) = (\mathrm{pr}(\omega > 0 \mid y, Y, X) > 1 - \alpha)$ for $y \sim f(y \mid x, \beta, \theta)$ is of interest. This is

$$E[I(y) \mid x, Y, X] = \int_{\mathcal{R}(y)} I(y) p(y \mid x, Y, X) dy,$$

where $p(y \mid x, Y, X)$ is the predictive distribution of y at concentration x given (Y, X). This is the interim analysis for the detection of positive concentration for a sample having concentration x. In the above, Geisser's procedure has been extended to a regression framework by conditioning the future response on the independent variable x. Geisser's procedure is also extended in the sense that there are many nuisance parameters and the parameter of interest ω is aligned only with a future observation and not a parameter such as β or θ that is associated with all of the observations.

The only difference between the above calculation and that of x^{pr} is in the calculation of $\mathrm{pr}(\omega > 0 \mid y, Y, X)$. In the interim analysis setting

$$\mathrm{pr}(\omega > 0 \mid y, Y, X) = \int_{\mathcal{R}(\beta)} \int_{\mathcal{R}(\theta)} \mathrm{pr}(\omega > 0 \mid y, \beta, \theta) \pi(\beta, \theta \mid y, Y, X) d\beta d\theta,$$

where $\pi(\beta, \theta \mid y, Y, X) = \int_0^\infty \pi(\beta, \theta, \omega \mid y, Y, X) \, d\omega$, while for the calculation of x^{pr}, $\pi(\beta, \theta \mid y, Y, X)$ is replaced with $\pi(\beta, \theta \mid Y, X)$. As alluded to in Section 5.3.2.3 this is necessary to ensure that x^{pr} reflects the current quality of the assay, not the predicted quality of the assay one observation into the future.

In an analogous way x^{ds} and ES precision can also be formulated as a Bayesian interim analysis problem. A further extension is to characterize $\mathrm{PDAM}(x)$ and hence PDAM precision using Bayesian interim analysis.

Note that, the derivations in Chapters 5 and 6 had been developed by the author before he became aware of Geisser's work.

8.7 Final remarks

The results of assays are used to make important decisions in many fields. Hence, it is desirable that the analysis be as accurate and precise as possible. To accomplish the correspondent statistical analysis many interesting and challenging problems are suggested.

In this thesis it has been shown that procedures based on the notion of Bayesian inference allow precise evaluation of the statistical analyses of an assay. It has also been shown that frequentist procedures do not take account of all of the information in the data; hence, these have the potential to be inaccurate. It is perhaps unfortunate that these inadequate frequentist procedures are the methods currently being used to analyze assays. When important decisions are made from the results of an assay, even small improvements to the quality of information being provided are invaluable. Thus, Bayesian methods should be seriously considered by practitioners.

It is the case that evaluation of Bayesian inferences requires intensive CPU simulations. In the past this was a seemingly insurmountable obstacle in the way of Bayesian analysis. With recent advances in Bayesian computation and the ever increasing power of modern computers, computational issues are no longer a valid reason for overlooking Bayesian methods. However, it still remains to be seen if practitioners will implement the Bayesian methods developed in this thesis.

Acknowledgements

The journey this thesis has taken me on has certainly been an interesting one. As such there are many people that I wish to thank.

My most sincere appreciation is extended to my two supervisors, Prof. John Deely and Dr. Murray Smith, for their input and guidance. They have certainly helped to make the undertaking of this thesis an interesting and rewarding period of my life.

I would also like to thank Mr. Bill Sadler of the Department of Nuclear Medicine at Christchurch Public Hospital for providing me with real immunoassay data and also for motivating some of the research. Special thanks goes to Dr. Frank Lad, Dr. Peter Renaud and Dr. Allan McInnes for looking after my affairs at the University of Canterbury while I was studying at Purdue University. I also thank my student colleagues Robert Ware, University of Canterbury, and Lin Chen, Purdue University, for their assistance. In addition, I extend my gratitude to the computer technicians in the department of Mathematics and Statistics at the University of Canterbury, in particular Mr. John Spain, and in the department of Statistics at Purdue University, in particular Mr. Doug Crabill, for their advice and dealing with my many "disasters" and enquiries. In today's world one would certainly be lost without people who know about computers! Finally, I wish to thank my parents, Peter and Joan O'Malley, my brother and sister in law, Justin and Heather O'Malley, other family members and my friends for their support and encouragement. Special thanks is extended to Heather for helping with the proof-reading!

References

Abramowitz, M. and Stegun, I. A. (1970). *Handbook of Mathematical Functions.* Government Printing Office: United States Department of Commerce, Washington D. C.

Amemiya, T. (1973). Regression Analysis When the Variance of the Dependent Variable is Proportional to the Square of its Expectation. *Journal of the American Statistical Association,* 68, 928-34.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Annals of Statistics,* 2, 1152-74.

Arveson, J (1998). How to Get a Patent with Regression Analysis. *Seminar in Biostatistics/Statistical Genetics, March 31, 1998.* Department of Statistics: Purdue University.

Beal, S. L. and Sheiner, L. B. (1988). Heteroscedastic nonlinear regression. *Technometrics,* 30, 327-38.

Bedrick, E. J., Christensen, R. R. and Johnson, W. O. (1996). A New Perspective on Priors for Generalized Linear Models. *Journal of the American Statistical Association,* 91, 1450-60.

Bedrick, E. J. and Christensen, R. R. and Johnson, W. O. (1997). Bayesian Binomial Regression: Predicting Survival at a Trauma Center. *American Statistician,* 51, 211-8.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd edn.). New York: Springer Verlag.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inferences (with discussion). *Journal of the Royal Statistical Society* (Series B), 41, 113-47.

Bernardo, J. M. and Smith , A. F. M. (1994). *Bayesian Theory*. New York: Wiley.

Besag, J., Green, P., Higdon, D. and Mengersen, K. (1995). Bayesian Computation and Stochastic Systems. *Statistical Science*, 10, 3-66.

Billingsley, P. (1986). *Probability and Measure* (2nd edn.). New York: Wiley.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society* (Series B), 26, 211-46.

Box, G. E. P. and Hill, W. J. (1974). Correcting Inhomogeneity of Variance With Power Transformation Weighting. *Technometrics*, 16, 385-9.

Brown, E. N., McDermott, T. J., Bloch, K. J. and McCollom, A. D. (1996). Defining the smallest analyte concentration an immunoassay can measure. *Clinical Chemistry*, 42, 893-903.

Brown, P. J. (1982). Multivariate Calibration (with discussion). *Journal of the Royal Statistical Society* (Series B), 44, 287-321.

Carroll, R. J. and Cline, D. B. H. (1988). An asymptotic theory for weighted least squares with weights estimated by replication. *Biometrika*, 75, 35-43.

Carroll, R. J. and Ruppert, D (1984). Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association*, 79, 321-8.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. New York: Chapman and Hall.

Casella, G. and George, E. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46, 167-74.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49, 327-35.

Christensen, R. R. (1991). *Linear Models for Multivariate, Time Series and Spatial Data*, New York: Springer-Verlag. pp. 217-20.

Cody, W. J. (1969). Rational Chebyshev approximations for the error function. *Mathematical Computing*, 631-8.

Currie, L. A. (1968). Limits for Qualitative Detection and Quantitative Determination. *Analytical Chemistry*, 40, 586-93.

Davidian, M. (1989). Optimism in Estimation of Minimum Detectable Concentration in Assays. *Unpublished Technical Report.* Department of Statistics: North Carolina State University.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association,* 82, 1079-91.

Davidian, M., Carroll, R. J. and Smith, W. (1988). Variance functions and the minimum detectable concentration in assays. *Biometrika* 75, 549-56.

Davies, C. J. (1994). Principles. *The Immunoassay Handbook* (Wild, D. G., ed.), New York: Stockton Press, Chapter 1, 3-47.

De Lean, A., Munson, P. J. and Rodbard, D. (1978). Simultaneous analysis of families of sigmoidal curves: Application to bioassay, radioligand assay, and physiological dose-response curves. *American Journal of Physiology,* 235, E97-E102.

Draper, D. (1998). *Bayesian Hierarchical Modeling* (Draft II). Department of Mathematical Sciences: University of Bath, UK.

Dunsmore, I. R. (1968). A Bayesian approach to Calibration. *Journal of the Royal Statistical Society* (Series B), 30, 396-405.

Ekins, R. P. (1978). Basic Concepts in Quality Control. *International Symposium on Radioimmunoassay and Related Procedures in Medicine 1977,* Vienna: International Atomic Energy Agency, 2, 6-20.

Ekins, R. P. (1983). The Precision Profile: Its use in assay design, assessment and quality control. *Immunoassays for Clinical Chemistry* (2nd edn. Hunter, W. M. and Corrie, J. E. T., eds.), Edinburgh: Churchill Livingston, 76-105.

Evans, M and Swartz, T. (1995). Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems. *Statistical Science,* 10, 254-72.

Finney, D. J. (1976). Radioligand assay. *Biometrics,* 32, 721-740.

Finney, D. J. and Phillips, P. (1977). The Form and Estimation of a Variance Function with Particular Reference to Radioimmunoassay. *Applied Statistics,* 26, 312-20.

Geisser, S. (1992). On the curtailment of sampling. *The Canadian Journal of Statistics*, 20, 297-309.

Geisser, S. (1993). Bayesian interim analysis of censored exponential observations. *Statistics and Probability Letters*, 18, 163-8.

Geisser, S. (1993). *Predictive Inference. An Introduction.* London: Chapman and Hall.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-41.

Gilks, W. R. and Roberts, G. O. (1996). Strategies for improving MCMC. *Markov Chain Monte Carlo in Practice* (Gilks, W. R., Richardson, S. and Spiegelhalter, D. J., eds.), London: Chapman and Hall, 89-114.

Harville, D. S. (1974). Bayesian inference for error components using only error contrasts. *Biometrika*, 61, 383-5.

Harville, D. S. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-40.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97-109.

Healy, M. J. R. (1972). Statistical analysis of radioimmunoassay data. *Biochemical Journal*, 130, 207-10.

Hoadley, B. (1970). A Bayesian look at Inverse Linear Regression. *Journal of the American Statistical Association*, 65, 356-69.

Hunter W. G. and Lambroy, W. F. (1981). A Bayesian Analysis of the Linear Calibration Problem (with discussion). *Technometrics*, 23, 323-50.

Jewell, N. P. and Raab, G. M. (1981). Difficulties in obtaining consistent estimators of variance parameters. *Biometrika*, 68, 221-6.

Lindley, D. V. (1982). Bayesian Inference. *International Encyclopedia of Statistics* (Kotz, S. and Johnson, N. L. eds.), 1, 197-224.

McGraw, K. O. and Wong, S. P. (1992). A Common Language Effect Size Statistic.

Psychological Bulletin, 361-5.

Mengersen, K. C. and Tweedie, R. L. (1996). Rates of Convergence of the Hastings and Metropolis algorithms. *The Annals of Statistics*, 24, 101-21.

Metropolis, N, Rosenbluth, A. W., Rosenbluth, M, N, Teller, A. H. and Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087-92.

Muller, P. (1991). A Generic Approach To Posterior Integration And Gibbs Sampling. *Technical Report*, 91-09. Department of Statistics: Purdue University.

Nel, D. G. (1980). On Matrix Differentation in Statistics. *South African Statistical Journal*, 14, 137-93.

O'Connell, M. A., Belanger, B. A. and Harland, P. D. (1993). Calibration and assay development using the four-parameter logistic model. *Chemometrics and Intelligent Laboratory Systems*, 20, 97-114.

O'Malley, A. J. (1996). Radioimmunoassay Analysis: An Overview. *Proceedings of the 1996 NZSA Conference: Research in the Learning of Statistics* (Neyland, J. and Clark, M., eds.), 62-6.

Oppenheimer, L., Capizzi, T. P., Weppelman, R. M. and Mehta, H. (1983). Determining the Lowest Limit of Reliable Assay Measurement. *Analytical Chemistry*, 55, 638-43.

Osbourne, C. (1991). Statistical Calibration: A Review. *International Statistical Review*, 59, 309-36.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545-54.

Raab, G. M. (1981). Estimation of a variance function, with application to Immunoassay. *Applied Statistics*, 30, 32-40.

Racine-Poon, A., Weihs, C., and Smith, A. F. M. (1991). Estimation of Relative Potency with Sequential Dilution Errors in Radioimmunoassay. *Biometrics*, 47, 1235-46.

Ramsey, F. L. (1972). A Bayesian Approach to Bioassay. *Biometrics*, 28, 841-58.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the Theory of Non-Parametric Statistics*. New York: Wiley. pp 90-4.

Ratkowsky, D. A. and Reedy, T. J. (1986). Choosing Near-Linear Parameters in the Four Parameter Logistic Model for Radioligand and Related Assays. *Biometrics*, 42, 575-82.

Robinson-Cox, J. F. (1995). Multiple estimation of concentrations in immunoassay using logistic models. *Journal of Immunological Methods*, 186, 79-88.

Rodbard, D. (1978). Statistical Estimation of the Minimal Detectable Concentration ("Sensitivity") for Radioligand Assays. *Analytical Biochemistry*, 90, 1-12.

Rodbard, D. and Hutt D. M. (1974). Statistical analysis of radioimmunoassays and immunoradiometric (labeled antibody) assays: a generalized weighted, iterative, least squares method for logistic curve fitting. *Radioimmunoassay and Related Procedures in Medicine*, Vienna: International Atomic Energy Agency, 1, 165-92.

Rodbard, D., Lenox, R. H., Wray, H. L. and Ramseth, D. (1976). Statistical characterization of the random errors in the radioimmunoassay dose-response variable. *Clinical Chemsitry*, 22, 350-8.

Sadler, W. A., Murray, L. M. and Turner, J. G. (1992). Minimum Distinguishable Difference in Concentration: A Clinical Oriented Translation of Assay Precision Summaries. *Clinical Chemistry*. 38, 1773-8.

Sadler, W. A. and Smith, M. H. (1986). A reliable method of estimating the variance function in immunoassay. *Computational Statistics and Data Analysis*, 3, 227-239.

Sadler, W. A. and Smith, M. H. (1990a). Use and Abuse of Imprecision Profiles: Some Pitfalls Illustrated by Computing and Plotting Confidence Intervals. *Clinical Chemistry*, 36, 1346-50.

Sadler, W. A. and Smith, M. H. (1990b). A computer program for estimating Imprecision characteristics of Immunoassays. *Computer and Biomedical Research*. 23, 105-14.

Sadler, W. A., Smith, M. H. and Legge, H. M. (1988). A method for direct estimation of imprecision profiles, with reference to immunoassay data. *Clinical Chemistry*. 34, 1058-61.

Schlüter, P. J. (1996). *Identification of Hazardous Motor Vehicle Accident Sites: Some Bayesian Considerations.* Ph.D. Thesis: University of Canterbury, Christchurch, New Zealand.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression.* New York: Wiley, pp 25-6, 619-23.

Smith, A. F. M. (1991). Bayesian computational methods. *Philosophical Transactions of the Royal Society of London* (Series A), 337, 369-86.

Smith, A. F. M. and Roberts, G. O. (1993). Bayesian Computation via Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society* (Series B), 55, 3-23.

Tierney, L. (1991). Exploring posterior distributions using Markov Chains. *Computer Science and Statistics: 23rd Symposium on the Interface* (Keramidas, E. M., ed.), 563-70.

Tierney, L. (1994). Markov Chains For Exploring Posterior Distributions. *The Annals of Statistics*, 22, 1701-28.

van Houweligen, J.C. (1988). Use and Abuse of Variance Models in Regression. *Biometrics*, 44, 1075-81.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian Analysis of Linear and Non-linear Population Models by using the Gibbs Sampler. *Applied Statistics*, 43, 201-21.

Yang, R. and Berger, J. O. (1996). *A Catalog of Non-informative Priors* (Preliminary Draft). Department of Statistics: Purdue University.

Chapter A

The immunoassay experiment

The content of the first two sections of this appendix is largely based on Davies (1994).

Immunoassays are exquisite analytical tests that utilize antibodies. They are used for qualitative and quantitative analysis of substances in blood fluids. The experiment is based on the assaying (counting) of molecules.

The fundamental property of an immunoassay is an unparalleled specificity of the antibody for the substance to which they bind and the strength of the binding once formed. The unparalleled specificity of the antibody is necessary in order for minute concentrations of analyte to be assayed in the presense of many closely related substances such as in blood samples. The strength of the binding enables accurate and precise quantification of concentrations.

A.1 Components of an immunoassay

The three major components of an immunoassay are:

- the antigen

- the antibody

- the labeling agent

In an immunoassay experiment the substance being analyzed is the antigen. An antigen is a molecule that can bind to an antibody. The antibody used in an

immunoassay needs to have a single binding site that recognizes and binds to the antigen.

The labeling agent is either a labeled version of the antigen or the antibody. The number of bound or free (unbound) molecules of the labeling agent are counted at the completion of the reaction. These counts are referred to as the activity bound and the activity free respectively.

A.2   Immunoassay mechanics

The reaction between the antigen and antibody is an equilibrium reaction. The bonds that form are non-covalent and there is a balance between the molecules in the reaction complex (i.e. bound) and those which are not. As the concentration of the reagents alter so do the rates of the forward and backward reactions and hence the equilibrium constant. If the concentration of one reagent increases then the rate of the reaction in the reverse direction increases and the equilibrium constant adjusts accordingly.

There are two major types of immunoassays. The important features of each are reviewed.

A.2.1   Competitive assays

In a competitive assay the labeling agent is a labeled version of the antigen. For example, in a radioimmunoassay a radioactive isotope is covalently attached to the antigen. The concentrations of the antibody and the labeled antigen are small relative to the expected concentration of the antigen. Consequently, this type of assay is sometimes referred to as a reagent limited assay. Despite an excess of antibody binding sites, the equilibrium nature of the reaction ensures that there will always be some molecules of antigen that are bound and some that are free.

Assuming that all molecules of the antigen, labeled and unlabeled, have an equal chance of binding, the relative proportions of molecules bound and free will be the same for the labeled and unlabeled versions of the antigen. As the concentration of the unlabeled antigen increases the equilibrium constant moves in favour of the reverse reaction. So, although the total number of bound antigen will increase the overall proportion of bound antigen decreases. Therefore, the expected number of

molecules of bound antigen that are labeled decreases as the concentration of unlabeled antigen increases. Since the rate of decrease is also a decreasing function of the concentration of unlabeled antigen, the dominating shape of the relationship between the activity bound and the concentration of unlabeled antigen is a hyperbola. The activity bound tends towards an asymptote on a small positive count at very large concentrations of antigen.

A.2.2 Immunometric assays

Immunometric assays use an antibody as the labeling agent. The most common design of this assay is the two-site immunometric assay, also called a sandwich assay.

Sandwich assays are appropriate when the antigen has two well separated binding sites. Two antibodies are used as reagents. One antibody captures the antigen, the other detects or quantifies it. Only the detecting antibody need be labeled. Excessive doses of both antibodies are added to the antigen. A sandwich assay is therefore a reagent excess assay. The experimental procedure is as follows:

1. The antigen is incubated with the capture antibody, which is usually attached to some solid phase (for example, a magnetic particle). Bonds form between the antibody and one of the available binding sites on the antigen.

2. The solid is washed to remove unreacted components.

3. The solid is incubated with the (labeled) detection antibody. Bonds form between the detection antibody and the solid complex from the first reaction.

4. The solid is washed to remove unreacted components.

5. The molecules of antigen bound to the detection antibody in the solid phase are counted.

Often steps 1 and 3 are performed simultaneously and the solid is only washed at the completion of the reaction.

In this assay, the activity bound increases as the concentration of the antigen increases. However, as the concentration of the antigen continues to increase, the increments become smaller because the equilibrium constant moves more in favour

of the reverse reaction. The relationship between the activity bound and the concentration of antigen is again hyperbolic. Note that, if the two reactions between the antibody and the antigen did not settle at equilibrium but rather went to completion, the relationship between the activity bound and the concentration of antigen would be a straight line.

A.2.3   Some additional terminology

The following are standard terms used in immunoassay:

- **Non-specific binding (NSB).** The activity bound for the assay when the concentration of antigen is zero. This estimates the residual free antigen in the activity bound after the free antigen has been separated. This needs to be subtracted from the raw counts.

- **Mean total activity (MnTotal).** The sum of the activity bound and the activity free in the presense of only labeled reagent (antigen or antibody depending on assay type). This measures the total activity of the assay.

- **Percentage bound.** The standardized form of the raw counts given by

$$\%\text{Bound} = \frac{\text{Activity bound} - \text{NSB}}{\text{MnTotal}} \times 100. \tag{A.1}$$

  In radioimmunoassays it is standard practice for the raw counts to be transformed into their percentage bounds prior to model fitting.

A.3   Assay design

Thus far, attention has been directed at the qualitative and quantitative aspects of the assay reaction. We now look at the process of using the output from this reaction to estimate the concentration of the substance being tested.

In an immunoassay it is usual for a batch of samples to be calibrated in the one experiment. These samples are referred to as unknowns. The procedure is as follows:

1. Obtain the raw count for each unknown.

2. Estimate the mean function and make point estimates and other inferences about the unknown concentrations.

It is important to estimate the mean function as accurately as possible. The closer the fitted mean function is to the actual relationship between the mean counts and the underlying concentration of the unknowns, the better the estimates of the unknown concentrations will be. The counts for the unknowns contain almost no information about the mean function. This is because there is no way of calibrating the measurements. To bolster the pool of information about the mean function, the unknowns are usually supplemented with samples containing a known concentration of the substance being analyzed. These samples are called standards. To maximize the information about the mean function contained in the standards their concentrations should be strategically spread over the full range of concentrations that can realistically occur.

Another feature of the design of an assay is the degree of replication for each sample. Standards are usually measured in duplicate. At the most critical or influential concentrations (in particular zero concentration) the degree of replication may be greater than two. In modern laboratories it is common for just one measurement of the response of an unknown sample to be taken. However, replicated response measurements have many benefits. In addition to obtaining a better estimate of the mean response, replicated responses yield more information about the variance function and thus allow it to be fitted with greater accuracy. When measured in singleton the counts for the unknowns contain virtually no information about the variance function.

A third type of sample that is often used in an immunoassay is the quality control (QC) specimen. These are unknowns which are analyzed in more than one assay. The purpose of QC specimen is to provide a checking mechanism for possible deterioration in the assay materials. Of primary concern is the degradation of the concentration of the standards over time. If the estimated concentrations of the QC samples vary too much between assays, it is likely that either the standards or the QC specimen have become corrupted. To guard against the first possibility, the current assay is repeated with a fresh batch of standards and QC specimen. The QC samples also allow a model to be fitted to the systematic movement or drift in the mean function from assay to assay. A potential means of improving the estimation

within each assay is to pool the information in separate assays. If there is a model accounting for the drift between assays, the same mean function can be assumed for each assay in the batch; hence, information from the assays can be pooled.