BAYES METHODS IN GROUP SEQUENTIAL CLINICAL TRIALS

A THESIS SUBMITTED TO

THE UNIVERSITY OF KENT AT CANTERBURY

IN THE SUBJECT OF STATISTICS

FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY.

By

Wendi Qian

June 1997

Contents

Abstract

Bayesian methods for group sequential clinical trials have received increasing attention recently. They offer an approach for dealing with many difficult problems and have some practical advantages over frequentist methods. This thesis covers Bayesian methods for group sequential clinical trials comparing two treatments using both the Bayes sequential procedure and the Bayes sequential decision procedure. The main outcome measures for clinical trials are distributed as normal, binomial, and exponential and the proportional hazard model for survival time data.

Under the framework of Bayes sequential procedure for group sequential clinical trials, the student t prior distribution for the parameter of interest is proposed as a replacement for the normal prior distribution when the sample mean is very distant from the mean of the prior distribution. The framework of Bayes sequential procedure in clinical trials on normal distribution responses with variance unknown is given.

Bayes sequential decision theory is applied to group sequential clinical trials. First, Bayes sequential decision procedures with piecewise continuous loss functions are used in clinical trials on normal distribution responses. The procedures with loss functions which consider treatment efficacy and patient horizon are then given in clinical trials on binary responses. Approximation methods of Bayes sequential decision procedures are explored in clinical trials with survival time data.

Robust Bayes analysis in clinical trials is presented to address the criticism on the subjective prior distribution for parameters of interest.

Acknowledgements

I would like to thank the Institute of Mathematics and Statistics and ASRU of University of Kent for their financial help in making my PhD study possible. Thanks to Prof. John Lewis for helping me at the start of my study.

A great thank you to my supervisor, Prof. Phil Brown, for advising and guiding my study. He has provided me with many suggestions, generous assistance and continuous encouragement. I have learned a lot from him not only regarding Bayesian statistics methods but also from his way of thinking and conducting research.

Thanks Robert Lindsay for improving the English presentation of this thesis.

Many thanks to the people in the Institute of Mathematics and Statistics and ASRU, and my friends for their help and encouragement. I will have many happy memories of their friendship during this period of study.

Finally, special thanks go to my family, my parents, my sisters and sisters-in-law, my nephew and my niece, and my special friend Zhao Jinghua. Their love, care and support have given me great motivation to work diligently during my study.

Chapter 1

Introduction

1.1 Group Sequential Clinical Trial

The randomised, controlled clinical trial is the standard for evaluating new treatments and therapeutic strategies in clinical research. It consumes substantial patient, investigator and financial resources. For ethical requirements, patient resources should be deployed efficiently and necessarily. Early termination could be considered if a clinical trial shows early benefits or unexpected toxicity. To achieve this, the interim monitoring of a clinical trial has been suggested and developed by statisticians.

The mathematical theory of sequential analysis was introduced in the 1940s, motivated by industrial applications, and has continued to develop actively. Over the past 20 years, there has been extensive development in the biostatistics literature concerning the sequential monitoring of clinical trials. Classical methods of sequential analysis in clinical trials are summarised by Armitage(1975), and later by Whitehead(1982). These methods allow for continuous monitoring of paired data while they achieve the desired levels of type I and type II error rates. Though these methods are generally successful in their pursuit of a reduced sample number, they are not feasible in practice because of the difficulty of continual

monitoring, particularly in multicenter co-operative clinical trials with survival time responses.

The term "group sequential" was first introduced by Pocock(1977). The group sequential clinical trial monitors a sequence of grouped data instead of paired data one group at a time and is used to decide whether sampling should be continued or stopped based on some criteria after observing each grouped data. The interim monitoring can be at either every specified number of samples, for instance, every 20 patients, or alternatively every selected time point, for instance, every 3 months after treatment or randomisation. The group sequential clinical trial or interim monitoring of clinical trial is now widely used for ethical, scientific and economic reasons. It is generally agreed that a clinical trial could be stopped should accumulating evidence demonstrate the superiority of one of the treatments or unexpected toxicity of treatments; whilst continuing the trial would unnecessarily expose some patients in the trial to the less effective treatment and delay applying the results to patients outside the trial. With the current statistical methods, it is now recommended by FDA that planned interim analyses should be included in any clinical trial protocol.

1.2 Introduction of Frequentist Methods

A number of different frequentist statistical procedures in group sequential clinical trials have been suggested. The most popular ones are the Pocock(1977) and the O'Brien-Fleming(1979) procedures. It is widely noticed that the repeated significance tests at conventional critical values increase the overall significance level or type I error rate α. This was shown by Armitage, McPherson and Rowe(1969). Therefore their methods adjust the critical values used at interim tests of the null hypothesis by the choice of a more stringent "nominal significance level" α' such that the overall type I error rate α is controlled at some prespecified level, for

example, $\alpha = 0.05$. If the p value of test statistics Z_j, $j = 1, 2, ..., l-1$, where l is the number of groups or number of analyses, is less than the nominal significance level α' at an interim analysis, then the trial could be stopped early since the significant treatment difference under overall significance level is equal to α. Otherwise, the trial is continued to the final analysis.

Pocock and O'Brien-Fleming have given the nominal significance levels used in their procedures for various maximum number of groups and overall significance levels in their papers. Some of these nominal significance levels with overall type I error rate $\alpha = 0.05$ are shown in the following table.

Nominal Significance Level

Procedure	One interim analysis	Two interim analyses
Pocock	0.0294, 0.0294	0.0221, 0.0221, 0.0221
O'Brien-Fleming	0.0048, 0.0475	0.0005, 0.0141, 0.0451

For example, for two interim analysis clinical trials the nominal significance levels above are 0.0005, 0.0141, and 0.0451 at the first, the second and the final analysis, respectively, with the O'Brien-Fleming procedure; the levels remain at a constant value of 0.0221 with the Pocock procedure. In terms of the nominal significance level, Pocock's procedure uses a constant stopping boundary, while the boundary of O'Brien-Fleming starts from a very strict level and ends close to the overall significance level. The methods of Pocock and O'Brien-Fleming require specifying the number of groups(or number of interim analyses) in advance and monitoring a clinical trial at equal increments of information. In practice, these procedures could cause difficulties since we may change the frequency of data monitoring at some point during the course of the trial for some unforeseen reasons. Another possibility is that slower recruitment than anticipated could force extension of the trial and hence increase the number of interim analyses.

Lan and DeMets(1983) have developed a generalized group sequential procedure in clinical trials, known as the spending function (or use function) approach. It was motivated by the early termination of the Beta-Blocker Heart Attack Trial(BHAT)(1981). Their method avoids the above two restrictions, includes the approaches of Pocock and O'Brien-Fleming as special cases, and requires only the specification of a spending (or use) function $\alpha(t)$ in advance. It is briefly described as follows.

Assume completion of a trial by time T, scaled arbitrarily such that T = 1, and specify an increasing function $\alpha(t)$ such that $\alpha(0) = 0$ and $\alpha(1) = \alpha$, which is the overall significance level. This function $\alpha(t)$, which is called "spending" or "use" function, allocates the amount of type I error rate that one can "spend or use" at each interim analysis. Suppose there is a continuous stochastic process $\{W(t); 0 \leq t \leq 1\}$, for example, Brownian motion process, and a continuous boundary $b(t)$, $0 \leq t \leq 1$, with probability α of being crossed in $0 \leq t \leq 1$. More specifically,

$$\alpha(t) = P(\tau \leq t) \qquad 0 \leq t \leq 1,$$

where τ is the first exit time across the boundary $b(t)$. Assume that $W(t)$ is observed only at time points, $0 \leq t_1 < t_2 < ... < t_l \leq 1$. These are corresponding to values of test statistics Z_j, $j = 1, 2, ..., l$. Let $W_j = W(t_j)$, $j = 1, 2, ..., l$. The boundary point $b_1 = b(t_1)$ is chosen such that

$$P(|W_1| > b_1) = P(0 \leq \tau \leq t_1) = \alpha(t_1),$$

that is, to assign an accumulated boundary crossing probability $\alpha(t_1)$ to the time t_1. The $b_j = b(t_j)$, $j = 2, ..., l$, are obtained such that

$$P(|W_1| \leq b_1, ..., |W_{j-1}| \leq b_{j-1}, |W_j| > b_j) = P(t_{j-1} < \tau \leq t_j) = \alpha(t_j) - \alpha(t_{j-1}).$$

The increment $\alpha(t_j) - \alpha(t_{j-1})$ represents the additional amount of the significance level that can be spent at the time period $(t_{j-1}, t_j]$. If $b(t)$ denotes a continuously accumulated boundary with $b(t_j) = b_j$, $j = 1, 2, ..., l$, of the process $\{W(t); 0 \leq t \leq 1\}$, then the probability of being crossed in $0 \leq t \leq 1$ is α. Therefore the sum of probabilities of $\{W(t_j), j = 1, 2, ..., l\}$ exceeding $\{b_1, b_2, ..., b_l\}$ is less than or equal to α.

Implicit in this procedure is information time t_j. On the scaled $[0, 1]$ interval, t represents the fraction of patients randomised or the number of events observed. The calendar or real time can be transformed to the information time, for example, see Lan and DeMets(1989). The evaluation of b_j, $j = 1, 2, ..., l$, depends only on $\alpha(t)$ and t_1, ..., t_j, and is independent of the number of groups l. Also the group sample size, or the increment of information $t_j - t_{j-1}$, $j = 1, 2, ..., l$, in each interim analysis doesn't need to be a constant. So a clinical trial can be monitored at unequally spaced times without specifying the number of interim analyses in advance by the generalized group sequential procedure. The spending function approach needs to specify the target sample size of a clinical trial.

The fixed sample, the Pocock and the O'Brien-Fleming designs in clinical trials are special cases of the spending function approach. If the spending function is

$$\alpha(t) = \begin{cases} 0 & 0 \leq t < 1 \\ \alpha & t = 1, \end{cases}$$

then it is a fixed sample design with the significance level equal to α.

If we choose spending function α_1,

$$\alpha_1(t) = 2 - 2\Phi(z_{\frac{\alpha}{2}}/\sqrt{t}),$$

where Φ is the standard normal distribution function, then the corresponding boundary is similar to the boundary of O'Brien-Fleming procedure. However, the

spending function procedure doesn't need to specify either the number of groups l or the group sample size in advance. Note $\alpha_1(0.5) < 0.006$ for $\alpha = 0.05$. So an O'Brien-Fleming boundary is unlikely to stop very early. The $\alpha_1(t)$ may be a suitable choice when long-term treatment effect is a major concern of a clinical trial.

The spending function

$$\alpha_2(t) = \alpha \ln\{1 + (e-1)t\}$$

will give the Pocock boundary. Since $\alpha_2(0.5) = 0.62\alpha$, $\alpha_2(t)$ will generally result in earlier termination but we will suffer a reduction in power.

While not described originally as a group sequential procedure, a strategy suggested by Haybittle(1971) and later supported by Peto et al(1976) merits consideration as an *ad hoc* version of group sequential data monitoring. Most interim analyses occur periodically after the entry of an additional group of subjects or observations of an additional number of events. Haybittle proposed a very conservative critical value for all interim analyses(e.g. ± 3.0 or ± 3.5) such that type I error rate increases almost negligibly in repeated analyses. At the last scheduled analysis one could use the usual 5% critical value of ± 1.96(or ± 2.0) should the trial continue that far.

The advantages among the above stopping boundaries depend upon the needs of each clinical trial and the investigators philosophy. The Pocock boundary offers the best opportunity for early termination. However, for a trial which continues to the end with an impressive trend (e.g. the value of standardised test statistics $>$ 1.96) but does not exceed the nominal significance level, the inability to reject the null hypothesis H_0 can be awkward and difficult to explain to clinical doctors. The Haybittle-Peto boundary does not allow much opportunity for early termination but avoids the awkward situation posed above. The O'Brien-Fleming boundary offers, in some sense, a compromise. Early termination is not likely as with that

of Pocock, but becomes more possible as the clinical trial progresses. At the end of a clinical trial, the critical value of O'Brien-Fleming boundary is close to the corresponding value of the fixed sample test.

Elashoff and Reedy(1984) discuss the selection of a group sequential procedure with one interim analysis, and conclude that there is no "best" rule and they explain how the different options compare. Geller and Pocock(1987) confine their attention to a few options for clinical trials with a maximum number of interim analyses between two and five and the overall significance level $\alpha = 0.05$ for a two sided test. The clinical trials with normal distribution responses are considered in their comparison. The options are the procedures of Pocock, O'Brien-Fleming, Haybittle-Peto and the plans in Pocock(1982) which minimise the average sample number for that alternative hypothesis to be detected with given powers of 0.5, 0.75 and 0.8. Their conclusions are that the Pocock procedure has the greatest savings in an average sample number when alternative hypotheses can be detected with high power, but the O'Brien-Fleming procedure is better than the Pocock procedure for saving a maximum sample number.

Wang and Tsiatis(1987) have introduced a family of one parameter stopping boundaries, which were defined in terms of a parameter whose value affected the probability of rejection of the null hypothesis over the various analyses. Suppose a clinical trial comparing two treatments is monitored after every $2n$ observations, n for each treatment, and the maximum number of groups is l. Let Z_j, $j = 1, 2, ..., l$, be the sequence of test statistics. Assume Z_j, $j = 1, 2, ..., l$, are normally distributed with var(Z_j)=1. The group sequential test consists of rejecting null hypothesis H_0 of no treatment difference for the first j such that

$$|Z_j| \geq C(\Delta, \alpha, l) j^{\Delta},$$

where $C(\Delta, \alpha, l)$ is chosen such that the overall significance level is α, that is, under the null hypothesis H_0, the probability of failing to reject H_0, when it is

true, is

$$P(|Z_1| < C(\Delta, \alpha, l)1^{\Delta}, ..., |Z_l| < C(\Delta, \alpha, l)l^{\Delta}|H_0) = 1 - \alpha,$$

where $C(\Delta, \alpha, l)$ can be computed using the numerical recursive integration formula given by Armitage, McPherson and Rowe(1969). The discrete stopping boundary values $C(\Delta, \alpha, l)j^{\Delta}$, $j = 1, ..., l$, depend on the parameter Δ, called the shape parameter. If $\Delta = 0$, it gives the boundary of O'Brien-Fleming; and if $\Delta = 0.5$, it is the boundary of Pocock.

This family of stopping boundaries yields approximately optimal results with respect to the least number of subjects for detecting specified treatment difference at given significance level α, and power $1 - \beta$. The optimal results of Wang and Tsiatis are consistent with those of Pocock(1982) by varying nominal significance levels to minimise the average sample number(ASN) under the alternative hypothesis. So the approximately optimal boundaries within the family of stopping boundaries are approximately optimal overall. The methods of Wang-Tsiatis also need to specify a maximum number of groups in advance and analyse data at equal increments of information.

Pampallone and Tsiatis(1994) have proposed a general family of boundaries based on the boundaries of Wang-Tsiatis that allow stopping early with rejection of either the null or alternative hypothesis.

The statistical package EaSt(Early Stopping) can be used to design a group sequential clinical trial with Wang and Tsiatis' family of one parameter stopping boundaries and Pampallone and Tsiatis' general family of boundaries.

SPRT(Sequential Probability Ratio Test) designs and analyses have been summarised by Whitehead(1992). He uses a continuous boundaries approach under the assumption of a continuous sample path, which is an abstract mathematical concept, and derives distributions of test statistics and power functions of tests.

A Christmas tree adjustment is suggested at discrete interim analyses. The triangular test is the most popular one in SPRT designs. There is a statistics package PEST(Planning & Evaluation of Sequential Trials) which can be used to not only design but also analyse group sequential clinical trials of SPRT. Estimations of the treatments effects have been given as well.

Jennison and Turnbull(1984, 1989) have described the repeated confidence interval approach. A sequence of intervals that all contain the true treatment difference with a prespecified probability, 95% say, are calculated at each interim analysis. The trial will be stopped and it can be claimed that there is a significant difference between the treatments when the current repeated confidence interval excludes 0. Jennison and Turnbull formulate certain repeated confidence intervals directly. Unlike previous methods, inferences of the repeated confidence interval approach are independent of the stopping rule. Interval estimates of the treatment difference are provided at each interim analysis. They can be used in reporting interim results and serve as an adjunct to a group sequential method giving more than just the "stop/continue" information at each interim analysis. This method is especially useful in some epidemiological studies or long-term follow-up studies where the sudden ending of exposure would be impossible. Koepcke(1989) has criticised that the repeated confidence intervals are too wide compared with confidence intervals constructed at termination of a group sequential test. Pocock and Hughes(1989) have suggested that repeated confidence intervals be shrunk toward the null value of the parameter.

The stochastic curtailment approach in group sequential clinical trials was introduced by Lan, Simon and Halperin(1982). More details are given by Halperin et al (1982). The idea of stochastic curtailment is to curtail a trial as soon as an eventual conclusion of a trial is determined with high probability. At any stage of a trial, we calculate the probability of an eventual conclusion of experimental superiority, conditional on the true treatment difference of a trial and on the data

observed so far. A trial is stopped and it is concluded that the experimental treatment is superior when the probability is large. This method can be used to illustrate the effects of low accrual trials. Stochastic curtailment is a prediction method, which is a criticism on this procedure.

The frequentist methods in group sequential clinical trials have the following major difficulties.

When a clinical trial is completed, there is an impressive trend of treatment difference which is however not significant at a pre-specified significance level, and the other studies have the same result. We may wish to carry on the study. How then should we analyse the extra data after we stop a trial? This type of problem also occurs when there is a delay between the entry of patient and the assessment of response to treatment. If a trial is stopped prematurely on the basis of a stopping rule, how should the statistician deal with extra data that become available after the trial has been stopped?

Terminal inferences of the group sequential method rely on strict adherence to the specified stopping rule. The confidence intervals, and point estimates of treatments differences have been studied under some special situations only. Most frequentist methods for group sequential clinical trials do not provide any inference about the treatment difference, only about the "stop/continue" decision during the period of interim monitoring.

Sometimes there is a difficulty in explaining the result to clinical investigators. For example, take a two interim analyses clinical trial with the Pocock procedure. The nominal significance level $\alpha' = 0.021$ is used at each analysis, leading to an overall significance level $\alpha = 0.05$. Suppose a trial has evidence of a treatment difference with nominal p value equal to 0.03 at each analysis. Then according to the nominal significance level this would not be statistically significant at the 5% level, whereas an investigator with identical data carrying out a fixed size analysis would attain $p = 0.03$. It is difficult to explain this to clinical investigators, who

wonder why previous inspections of the data should affect the interpretation of the final results.

A large number of analyses give more opportunities for early stopping and will decrease the mean sample size if the treatment difference is large. On the other hand, increasing the number of analyses can actually increase the expected number of patients required for the trial under the null hypothesis, because the nominal significance level must be adjusted downward to maintain the overall type I error rate. There is a "penalty" paid for frequent interim monitoring of a clinical trial.

Reviews of frequentist methods for group sequential clinical trials can be found in DeMets(1987), Jennison and Turnbull(1990), Pocock(1992), Whitehead(1992) and Fleming and DeMets(1993).

1.3 Review of Bayesian Methods

Bayesian and the frequentist statistical approaches are based on inverse measures: one deals with probabilities of hypotheses given the data and the other involves probabilities of data sets given hypotheses. The interest of Bayesian method is on some unknown parameter δ. The notation of probability has different interpretations. The probability in Bayesian inferences is not frequentist. The $P(\delta \leq x)$ does not represent the proportion of times that δ is less than or equal to x in repeated investigations. Instead, it represents how likely the investigator thinks that δ is less than or equal to x. Berry(1987) compares Bayesian with frequentist statistical approaches based on the role of likelihood principle. The comparison is summarised here. The Bayesian approach is conditional since the posterior is a distribution of given available information. The frequentist approach is unconditional since the statistical inference is derived from a given hypothesis. The Bayesian approach is consistent with the likelihood principle since the posterior

distribution depends on the observed data only through Bayes theorem. The frequentist approach is not, because the p value or the tail probability is the probability under the null hypothesis of a result as extreme or more extreme than observed. In clinical trials, the unconditional approach disallows looking at the data if there is a possibility of stopping or otherwise modifying the study as a result, unless inferences are adjusted accordingly. The conditional approach is completely flexible in this regard.

Bayesian methods for group sequential clinical trials have received increasing attention recently, as they offer an approach for dealing with many difficult problems and have some practical advantages over frequentist methods. As we know, before designing a clinical trial to compare the experimental treatment with the standard treatment, we will acquire all possible information about the activity of both treatments. This information will give us an opinion about the treatment difference δ and can be described by a prior distribution of the treatment difference, denoted by $w(\delta)$. For example, Freedman and Spiegelhalter(1983) discuss their experience of translating doctors' opinions into subjective probability distributions. Chaloner *et. al.*(1993) describe a graphical elicitation of a prior distribution for a clinical trial. When we collect some data, we can update the opinion by Bayes theorem and get the posterior distribution of the treatment difference, denoted by $w(\delta|data)$. The Bayesian inference derives entirely from this posterior distribution of the treatment difference. Naturally the clinical trial may be stopped if either

$$\int_{-\infty}^{\delta_2} w(\delta|data)d\delta \geq 1-\varepsilon_1,$$

or

$$\int_{\delta_1}^{\infty} w(\delta|data)d\delta \geq 1-\varepsilon_2,$$

where the ε_1 and ε_2 are small positive values, the larger value of δ denotes the experimental treatment better, and the interval $(\delta_1,\ \delta_2)$ is the range of equivalence (that is, the two treatments are considered roughly equivalent.). At each interim analysis, the treatment difference can be estimated by the expectation of treatment difference δ with respect to the posterior distribution $w(\delta|data)$. Freedman and Spiegelhalter(1989, 1993) have shown that by choosing some prior distributions Bayesian boundaries can be very close to Pocock or O'Brien-Fleming boundaries. Geller and Pocock(1987) mention that the Pocock procedure has the disadvantage of undertaking the last analysis at a p value considerably smaller than 0.05 and that the O'Brien-Fleming procedure is perhaps too stringent at the first analysis, virtually assuring that the trial does not stop then. Freedman and Spiegelhalter(1989, 1993) have also shown that the Bayesian method can have a stopping rule between the Pocock and O'Brien-Fleming boundaries by the choice of some prior distributions.

The attraction of the Bayesian method lies in its simplicity of concept and the directness of its conclusions. When we collect some data at any time, we update the opinion on the treatment difference by Bayes theorem. The likelihood principle implies that interpretation of the data does not depend on the number of analyses or on the stopping rule of the trial. So no "penalty" is paid for frequent interim analyses, and extra data can be analysed after the trial has been completed. Statistical inferences on the treatment difference following the trial are derived from the posterior distribution of the treatment difference. The problems of frequentist methods described in Section 1.2 are solved by Bayesian methods. However, Bayesian methods have not been as well developed as frequentist methods, and technical difficulties arise when numerous nuisance parameters are to be considered in addition to the treatment difference δ itself. There is little corresponding software generally available which blocks the application of Bayesian methods in practice. The barrier to widespread implementation of the Bayesian method has

been its computational difficulty and the construction of prior distribution, but that in principle is no longer a problem(Whitehead 1992).

Spiegelhalter, Freedman and Parmar(1994) have reviewed and demonstrated how Bayesian methods can be applied to group sequential clinical trials. Bayes sequential methods in clinical trials have been explored by Novick and Grizzle(1965), Cornfield(1966, 1969), Berry(1985, 1989), Whitehead(1991), Freedman and Spiegelhalter(1989, 1991, 1993), Spiegelhalter and Freedman(1988), Freedman, Spiegelhalter and Parmar(1994), Parmar, Spiegelhalter and Freedman(1994), and by George *et. al.*(1994), and discussed by Jennison and Turnbull(1990).

Since Bayes theorem allows an investigator to update his subjective opinion of the treatment difference δ at any time, there is no special reason for a Bayesian to devise a stopping rule in advance. Decision theory provides the framework for combining subjective distributions with action. However, Bayes decision theory has not been widely introduced in group sequential clinical trials. Sylvester(1988) has used Bayes decision theory for a one-stage phase II clinical trial with binomial distribution response. Berry and Ho(1988) have addressed one-sided sequential stopping boundaries for clinical trials from a decision-theoretic point of view. Lewis and Berry(1994), and Lewis(1996) have studied Bayes sequential decision theory with piecewise continuous loss functions in group sequential clinical trials with binomial distribution response.

The major criticism of Bayes analyses is that it presumes an ability to completely and accurately elicit subjective information in terms of a single prior distribution. However, there has long existed (at least since Good(1959)) a robust Bayesian viewpoint which replaces the single prior distribution with a *class* of possible prior distributions. The goal of this approach is to make inferences or decisions which are robust over this class, i.e., relatively insensitive (or at least are satisfactory) to deviations as the prior distribution varies over this class. Greenhouse and Wasserman(1995) have illustrated the application of robust Bayesian

methods in clinical trials. Spiegelhalter, Freedman, and Parmar(1994) have suggested the consideration of a *community* of priors covering the perspectives of a range of individuals. This may encompass a *reference* prior intended to add as little as possible to the data and a *clinical* prior expressing reasonable opinions held by individuals or derived from overviews(meta-analyses) of similar studies. However, it is also useful to develop "off the shelf" priors corresponding to a formal expression of *sceptical* and *enthusiastic* belief. These may be thought to provide reasonable bounds to the community of priors.

1.4 Aims and Outline of the Thesis

In this thesis, Bayes methods in group sequential clinical trials comparing two treatments are studied using both Bayes sequential and Bayes sequential decision methods; and the main outcome variables for clinical trials are distributed as normal, binomial, and exponential and proportional hazard models for survival time data. The aims of the thesis are to study some unresearched problems in Bayes sequential methods, build a set of systematic Bayes sequential decision methods, and also to compare these with frequentist methods in group sequential clinical trials.

Bayes sequential methods in clinical trials are discussed in Chapter 2. Under its framework in clinical trials, in which the main outcome variables is normally distributed, the student t prior distribution is used and compared with the normal prior distribution for the treatment difference. Bayes sequential decision theory is introduced to group sequential clinical trials. The brief introduction of Bayes sequential decision theory is described in Chapter 3. In Chapter 4, Bayes group sequential decision clinical trials are set up based on normal distribution responses with piecewise continuous loss functions, and are also compared with frequentist

methods. The loss functions of considering the treatment effect and patient horizon are studied under clinical trials with binary responses, which are discussed in Chapter 5. Chapter 6 discusses the application of Bayes sequential decision theory in group sequential clinical trials with the main outcome variable being an exponential and proportional hazard model for survival time. The brief introduction of non-parametric Bayes analysis is also included in Chapter 6 in order to be applied in proportional hazard model for survival time. Robust Bayes analyses which study the uncertainty of prior information in clinical trials are described in Chapter 7. The discussion and further study on some common issues are given in Chapter 8.

Chapter 2

Bayes Sequential Methods

In this chapter, the framework of Bayes sequential methods in group sequential clinical trials is described in Section 2.1. This framework is based on clinical trials whose main outcome variable is normally distributed with known variance. The mean of the normal distribution is the treatment effect. The parameter of interest is the treatment difference which is considered to have a normal prior distribution. In Section 2.2, the prior distribution of the treatment difference which has the form of student t is discussed. This is also compared with the situation of normal prior to the treatment difference. In practice, the variance of the normal response variable is usually unknown. This issue is studied in Section 2.3.

2.1 Framework of Bayes Sequential Methods

2.1.1 The Problem

A group sequential clinical trial is designed to compare an experimental treatment with the standard treatment. The main outcome measure X for the clinical trial is normally distributed with probability density functions $N(\mu_e, \frac{\sigma^2}{2})$ and $N(\mu_s, \frac{\sigma^2}{2})$ for the experimental and the standard treatments, respectively. The value of

σ^2 is known. For the presentation of below formulas purely, the variance of main outcome measure X is assumed to have the form $\frac{\sigma^2}{2}$. The treatment is assigned by a randomised permuted block so that each consecutive group of $2n_j$, $j = 1, 2, ..., l$, patients has n_j on each treatment. The l is the maximum number of groups.

Let the parameter $\delta = \mu_e - \mu_s$ be the measure of treatment difference where large value of δ implies the superiority of the experimental treatment. The scale of treatment difference is divided into $(-\infty,\ \delta_1)$, $(\delta_1,\ \delta_2)$, and $(\delta_2,\ \infty)$. If $\delta > \delta_2$, then the experimental treatment is considered clinically superior. If $\delta < \delta_1$, then the standard treatment is superior. The interval $(\delta_1,\ \delta_2)$ is called the range of equivalence where the two treatments are considered roughly equivalent. Depending on the clinical situation, δ_1 and δ_2 will either coincide or δ_1 will be less than δ_2. The partitioning of the scale of treatment difference will be based on the relative toxicity of treatments and to a lesser extent on their cost and convenience. Assume that the treatment difference δ has the normal prior distribution, that is,

$$\delta \sim w(\delta) = N(\nu_0, \tau_0^2). \tag{2.1}$$

The variance τ_0^2 is expressed as $\tau_0^2 = \frac{\sigma^2}{n_0}$, which might suggest that there were n_0 "extra" pairs of patients in the pilot trial(Freedman and Spiegelhalter 1989). This form is useful in comparing different prior information by the change of value n_0. The n_0 is a measure of prior information on the treatment difference δ.

Let the group sample means be denoted by $\overline{X}_{js} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ijs}$ and $\overline{X}_{je} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ije}$, $j = 1, 2, ..., l$, for the standard and experimental treatments, respectively, where the group sequential sample from the standard treatment X_{ijs} is from the normal distribution $N(\mu_s, \frac{\sigma^2}{2})$ and the group sequential sample from the experimental treatment X_{ije} is from the normal distribution $N(\mu_e, \frac{\sigma^2}{2})$, $i = 1, 2, ..., n_j$. The group sample means $\overline{X}_{js}$ and $\overline{X}_{je}$ are then normally distributed with densities $N(\mu_s, \frac{\sigma^2}{2n_j})$ and $N(\mu_e, \frac{\sigma^2}{2n_j})$, respectively. The sequence of the differences between

$\overline{X}_{je}$ and $\overline{X}_{js}$, denoted by Z_j,

$$Z_j = \overline{X}_{je} - \overline{X}_{js} \sim N(\delta, \frac{\sigma^2}{n_j}), \quad j = 1, 2, ..., l, \tag{2.2}$$

are sufficient statistics of the treatment difference δ. There will be no loss of information to replace the group sequential variables $\{X_{ije},\ X_{ijs},\ i = 1, 2, ..., n_j, j = 1, 2, ..., l\}$ by this classical sequential sample $\{Z_j, j = 1, 2, ..., l\}$. The group sequential clinical trial described above becomes a sequential clinical trial whose main outcome variable Z_j, $j = 1, 2, ..., l$, is from the normal distribution $N(\delta, \frac{\sigma^2}{n_j})$, where δ is the measure of treatment difference and has the normal prior distribution $w(\delta) = N(\nu_0, \frac{\sigma^2}{n_0})$ in (2.1).

2.1.2 The framework

At each analysis j, $j = 1, 2, ..., l$, after observing the differences of group sample means $Z_1 = z_1$, $Z_2 = z_2$,..., $Z_j = z_j$, from the clinical trial, by Bayes theorem the posterior probability density function of δ is the normal distribution with mean equal to $\frac{\sum_{i=1}^{j} n_i \overline{z}_j + n_0 \nu_0}{\sum_{i=1}^{j} n_i + n_0}$ and variance equal to $\frac{\sigma^2}{\sum_{i=1}^{j} n_i + n_0}$, that is,

$$\begin{aligned} \delta \sim w(\delta | z_1, z_2, ..., z_j) &= w(\delta | \overline{z}_j) \\ &= N\left(\frac{\sum_{i=1}^{j} n_i \overline{z}_j + n_0 \nu_0}{\sum_{i=1}^{j} n_i + n_0},\ \frac{\sigma^2}{\sum_{i=1}^{j} n_i + n_0} \right), \end{aligned} \tag{2.3}$$

where $\overline{z}_j = \frac{\sum_{i=1}^{j} n_i z_i}{\sum_{i=1}^{j} n_i}$.

The sequential Bayes method may suggest termination of the clinical trial at an interim analysis j, $j = 1, 2, ..., l-1$, if either

$$P(\delta < \delta_2 | z_1, z_2, ..., z_j) = \int_{-\infty}^{\delta_2} w(\delta | z_1, z_2, ..., z_j) d\delta > 1 - \varepsilon_1, \tag{2.4}$$

resulting in a rejection of the experimental treatment, or

$$P(\delta > \delta_1 | z_1, z_2, ..., z_j) = \int_{\delta_1}^{\infty} w(\delta | z_1, z_2, ..., z_j) d\delta > 1 - \varepsilon_2, \tag{2.5}$$

resulting in a recommendation of the experimental treatment, where the posterior probability density $w(\delta | z_1, z_2, ..., z_j)$ is obtained by (2.3). Otherwise the clinical trial needs to be continued to observe the next group of patients. At the final analysis, if (2.4) or (2.5) are not satisfied, then it may be concluded that these two treatments are equivalent since $P(\delta_1 < \delta < \delta_2)$ is large. The ε_1 and ε_2 are small positive values, such as 0.05, 0.025, etc.

Conditions (2.4) and (2.5) may be written as,

$$\overline{z}_j < \frac{\sum_{i=1}^{j} n_i + n_0}{\sum_{i=1}^{j} n_i} \delta_2 - \frac{n_0}{\sum_{i=1}^{j} n_i} \nu_0 - \frac{\Phi^{-1}(1 - \varepsilon_1)}{\sum_{i=1}^{j} n_i} \sigma \sqrt{\sum_{i=1}^{j} n_i + n_0}, \tag{2.6}$$

and

$$\overline{z}_j > \frac{\sum_{i=1}^{j} n_i + n_0}{\sum_{i=1}^{j} n_i} \delta_1 - \frac{n_0}{\sum_{i=1}^{j} n_i} \nu_0 - \frac{\Phi^{-1}(\varepsilon_2)}{\sum_{i=1}^{j} n_i} \sigma \sqrt{\sum_{i=1}^{j} n_i + n_0}, \tag{2.7}$$

respectively, where $\Phi^{-1}(1 - 0.025) = 1.96$. These are the same form as boundaries of frequentist methods in terms of test statistics $\overline{Z}_j = \overline{z}_j$, $j = 1, 2, ..., l$. Freedman and Spiegelhalter(1989) have shown that by choosing some prior distributions of δ in (2.1) through some values of n_0, Bayesian boundaries of (2.6) and (2.7) can be very close to boundaries of Pocock and O'Brien-Fleming procedures. Bayesian methods provide the same desirable features as frequentist methods.

2.2 Student t Prior Distribution

2.2.1 The Framework

In Section 2.1, we assume that the treatment difference δ in the clinical trial has the normal prior distribution $\delta \sim w(\delta) = N(\nu_0, \frac{\sigma^2}{n_0})$. At each analysis j, $j = 1, 2, ..., l$, the posterior mean of the treatment difference δ from (2.3) is

$$E(\delta|z_1, z_2, ..., z_j) = \overline{z}_j - \frac{n_0}{\sum_{i=1}^{j} n_i + n_0}(\overline{z}_j - \nu_0).$$

If the sample mean $\overline{z}_j$ is very far from the mean of the prior distribution ν_0, then the posterior mean $E(\delta|z_1, z_2, ..., z_j)$ will differ considerably from $\overline{z}_j$. Dawid(1973) has shown that this undesirable behaviour would be avoided if the prior distribution of δ had the form of a student t distribution. Assume that the variance σ^2 is known. He mentions that "if $X \sim N(\theta, \sigma^2)$ given $\Theta = \theta$, while Θ has a student t prior, one obtains a limiting posterior $\Theta \sim N(x, \sigma^2)$ as $|x| \to \infty$, and $E(\Theta|x) - x \to 0$, as conjectured by Lindley.". The Bayes sequential method in the clinical trial described in Section 2.1.1 with the student t prior for the treatment difference δ is discussed below.

Without loss of generality, let the mean of the prior distribution in (2.1) be equal to 0, that is, $\nu_0 = 0$. The prior distribution for the standardised treatment difference $\delta' = \frac{\delta}{\sqrt{\frac{\sigma^2}{n_0}}}$ then becomes the standard normal distribution $N(0, 1)$. Instead, consider that δ' has a student t prior with degree of freedom v, that is,

$$\delta' \sim w(\delta'|v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}(1 + \frac{\delta'^2}{v})^{-\frac{v+1}{2}}. \tag{2.8}$$

The prior distribution of δ is then

$$\delta \sim \frac{1}{a} w(\frac{\delta}{a}|v), \tag{2.9}$$

where $a = \sqrt{\frac{\sigma^2}{n_0}}$, and $w(\cdot|v)$ is the density function of (2.8).

At each analysis j, $j = 1, 2, ..., l$, after observing $Z_1 = z_1$, $Z_2 = z_2$, ..., $Z_j = z_j$ as in Section 2.1, the posterior probability density function of δ is

$$w(\delta|z_1, z_2, ..., z_j) = w(\delta|\overline{z}_j) = \frac{e^{-\frac{1}{2}\frac{\sum_{i=1}^{j} n_i}{\sigma^2}(\delta-\overline{z}_j)^2}(1+\frac{\delta^2}{a^2 v})^{\frac{v+2}{2}}}{\int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{\sum_{i=1}^{j} n_i}{\sigma^2}(\delta-\overline{z}_j)^2}(1+\frac{\delta^2}{a^2 v})^{\frac{v+2}{2}} d\delta}, \tag{2.10}$$

where $\overline{z}_j = \frac{\sum_{i=1}^{j} n_i z_i}{\sum_{i=1}^{j} n_i}$.

The clinical trial may be terminated at analysis j, $j = 1, 2, ..., l$, if either,

$$P(\delta < \delta_2|z_1, z_2, ..., z_j) = \int_{-\infty}^{\delta_2} w(\delta|\overline{z}_j)d\delta > 1 - \varepsilon_1, \tag{2.11}$$

or,

$$P(\delta > \delta_1|z_1, z_2, ..., z_j) = \int_{\delta_1}^{\infty} w(\delta|\overline{z}_j)d\delta > 1 - \varepsilon_2, \tag{2.12}$$

where the posterior density function $w(\delta|\overline{z}_j)$ is obtained by (2.10), and ε_1 and ε_2 are small positive numbers. Otherwise the trial is continued to observe the next group of patients.

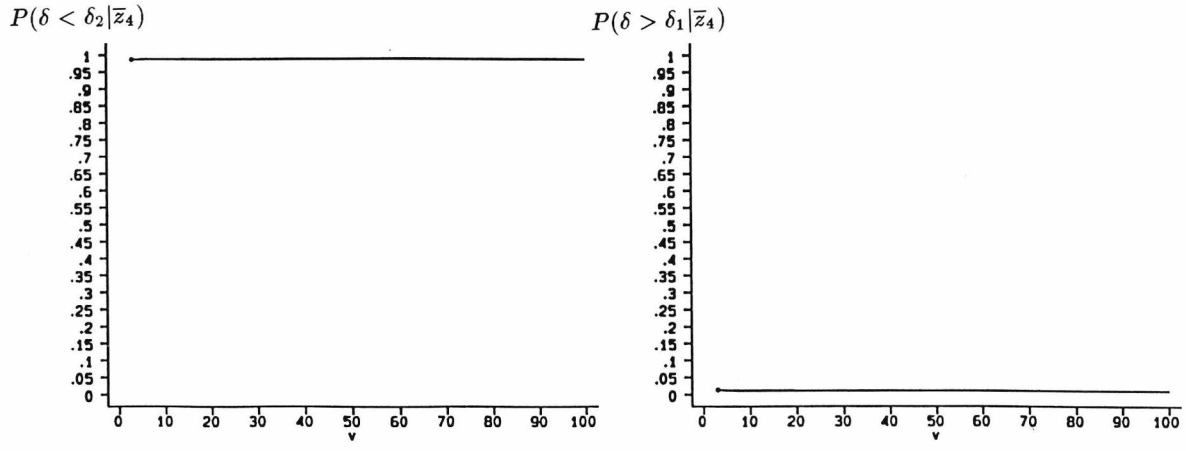
When the treatment difference δ has the form of student t prior as in (2.9), there is no closed form for the posterior probabilities $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1|z_1, z_2, ..., z_j)$ in (2.11) and (2.12), respectively. However, the numerical integration may be obtained by many mathematical and statistical packages.

2.2.2 Inferences for the Degree of Freedom v

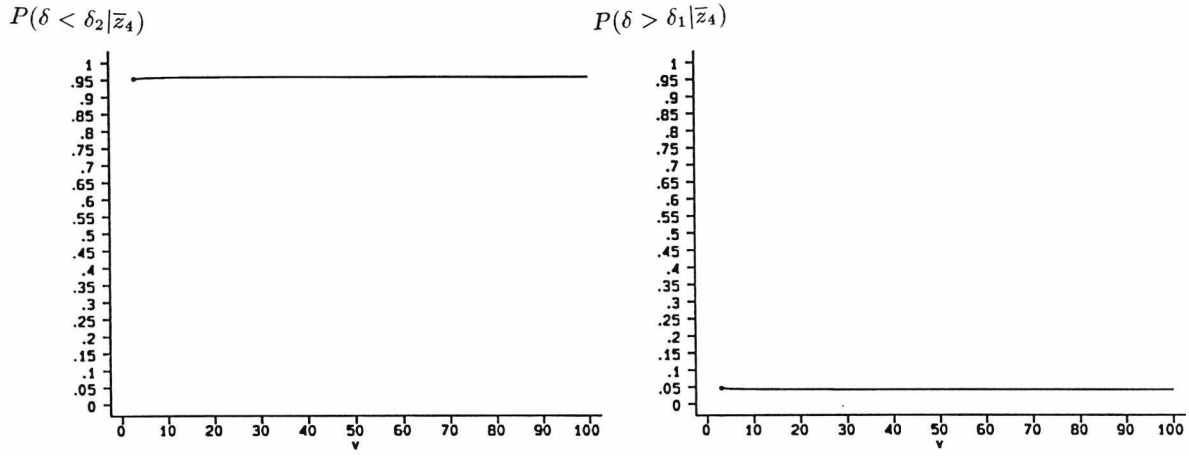
Using the example of Freedman and Spiegelhalter(1989), consider a clinical trial comparing two treatments with 200 patients, and the number of groups $l = 5$ with the equal group sample size $n_j = n = 20$, $j = 1, 2, ..., 5$. Let The test statistics Z_j, $j = 1, 2, ..., 5$, are from the normal distributions $N(\delta, \frac{\sigma^2}{n})$ given δ as in (2.2).

The $\sigma^2 = 0.5$ and $n_0 = 8, 22$, and 89 for the scanty, moderate and considerable prior information, respectively. Here suppose that the prior distribution of the treatment difference δ is the student t distribution with the form of (2.9). Let $n_0 = 22$ where prior information available is average. Assume that our interest is about departures, in either direction. Then $\delta_1 = \delta_2 = 0$. At each analysis j, $j = 1, 2, ..., 5$, if the posterior probabilities $P(\delta < \delta_2 | z_1, z_2, ..., z_j) = P(\delta < \delta_2 | \overline{z}_j)$ or $P(\delta > \delta_1 | z_1, z_2, ..., z_j) = P(\delta > \delta_1 | \overline{z}_j)$ follow the conditions (2.11) or (2.12), respectively, then stopping the clinical trial may be suggested as before.

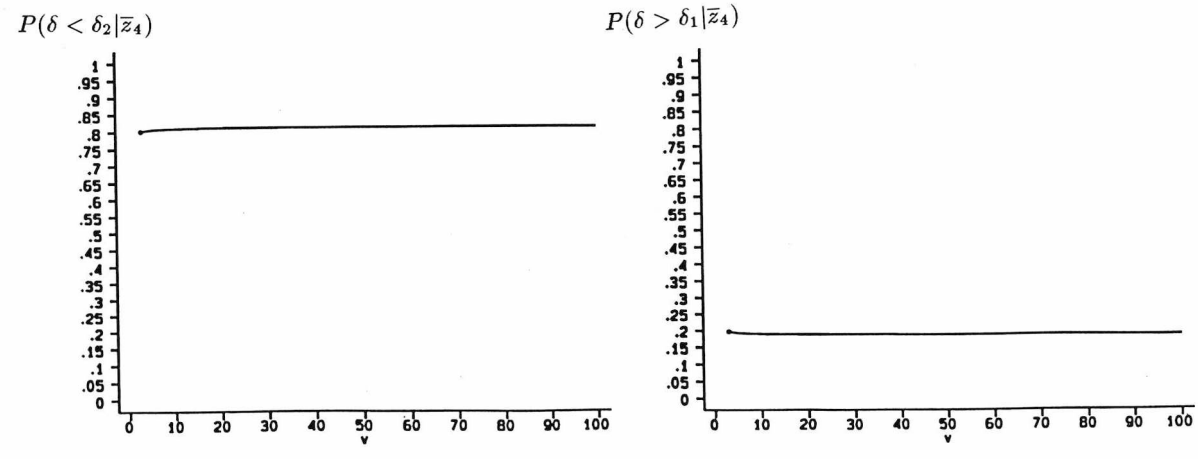
Since the $\overline{Z}_j = \frac{1}{j} \sum_{i=1}^{j} Z_i$, $j = 1, 2, ..., 5$, has the normal distribution $N(\delta, \frac{\sigma^2}{jn})$ given δ, the range of 99% possible value of $\overline{Z}_j$ is $\left(\delta + \sqrt{\frac{\sigma^2}{jn}} \times (-2.58),\ \delta + \sqrt{\frac{\sigma^2}{jn}} \times 2.58\right)$. As an example, $j = 4$ say, let $\overline{z}_4 = \sqrt{\frac{\sigma^2}{4n}} \times z_0$. By symmetry, only $\overline{z}_0 < 0$ needs to be considered. Without loss of generality, assume that z_0 = -2.58, -1.96, and -1, the corresponding posterior probabilities $P(\delta < \delta_2 | \overline{z}_4)$ of (2.11) and $P(\delta > \delta_1 | \overline{z}_4)$ of (2.12) with the degree of freedom v from 3 to 100 in the student t priors are displayed in Figure 2.1(a), Figure 2.1(b), and Figure 2.1(c), respectively. These figures show that the change of $P(\delta < \delta_2 | \overline{z}_4)$ and $P(\delta > \delta_1 | \overline{z}_4)$ with the change of the degree of freedom v of student t prior can be ignored, since these probabilities are almost constant with the degree of freedom v from 3 to 100. Similar figures can be obtained at other interim analyses. The same results are also found when the values of σ^2 and n_0 are changed. Therefore it can be concluded that the posterior inferences on the treatment difference δ is robust to the degree of freedom v of the student t prior.



(a) z_0=-2.58



(b) z_0=-1.96



(c) z_0=-1

Figure 2.1 Posterior Probabilities with Different Degree of Freedom

2.2.3 Comparisons with the Normal Prior Distribution

By the framework of Bayes sequential methods in group sequential clinical trials, the stopping rules of these methods are based on the posterior probability distribution of the treatment difference δ. It is then interesting to compare the posterior probabilities $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1|z_1, z_2, ..., z_j)$, $j = 1, 2, ..., l$, for the student t prior distributions with those for normal prior distributions. The comparisons are based on the change of posterior probabilities $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1|z_1, z_2, ..., z_j)$, $j = 1, 2, ..., l$, with the change of prior information.

The variance (or n_0) of the normal prior distribution $N(\nu_0, \frac{\sigma^2}{n_0})$ is a measure of prior information on the treatment difference δ. The tail probability of the student t distribution is more sensitive than its degree of freedom. Hence, the change of prior information is considered by the change of variances and tail probabilities of normal and student t prior distributions.

The robustness to change of variances of prior distributions

Continuing the example of Section 2.2.2, assume that the treatment difference δ has the normal prior distribution $\delta \sim N(\nu_0, \frac{\sigma^2}{n_0})$ as in (2.1), where $\nu_0 = 0$. Let $n_0 = 8$, 22, and 89 for the scanty, moderate and considerable prior information, respectively. At each analysis j, $j = 1, 2, ..., 5$, after observing $Z_1 = z_1$, $Z_2 = z_2$, ..., $Z_j = z_j$, the posterior probabilities $P(\delta < \delta_2|\overline{z}_j)$ and $P(\delta > \delta_1|\overline{z}_j)$ can be calculated by (2.4) and (2.5). As an example, say $j = 4$, the posterior probabilities $P(\delta < \delta_2|\overline{z}_4)$ and $P(\delta > \delta_1|\overline{z}_4)$ with $\overline{z}_4 = \sqrt{\frac{\sigma^2}{4n}} \times z_0$, where z_0 = -2.58, -1.96, -1, and 0, are listed in Table 2.1. It shows that the change of prior information might or might not affect the decision of stopping a trial early. Assume $\varepsilon_1 = \varepsilon_2 = 0.05$. When the value of observation $\overline{z}_4 = \sqrt{\frac{\sigma^2}{4n}} \times (-2.58)$, a very strong evidence of the experimental treatment not being superior, all posterior probabilities $P(\delta < \delta_2|\overline{z}_4)$ under different prior information, $n_0 = 8$, 22, and 89, indicate that stopping the

trial would be advisable. When $z_0 = -1.96$ and with the scanty prior information $n_0 = 8$, the posterior probability $P(\delta < \delta_2|\overline{z}_4) = 0.9694 > 1 - 0.05$, which is suggesting that the trial may be stopped early, but with considerable prior information ($n_0 = 89$) of no treatment difference (since we assumed that $\delta \sim N(0, \frac{\sigma^2}{n_0})$), the posterior probability $P(\delta < \delta_2|\overline{z}_4) = 0.9115 \not> 1 - 0.05$, which is suggesting that the trial should be continued.

Table 2.1: Posterior Probabilities of Different Normal Priors (n_0)

	$P(\delta < \delta_2\|\overline{z}_4)$			$P(\delta > \delta_1\|\overline{z}_4)$		
	n_0			n_0		
z_0	8	22	89	8	22	89
-2.58	0.9933	0.9891	0.9623	0.0069	0.0112	0.0380
-1.96	0.9694	0.9589	0.9115	0.0308	0.0413	0.0888
-1	0.8300	0.8122	0.7545	0.1702	0.1880	0.2457
0	0.5001	0.5001	0.5001	0.5001	0.5001	0.5001

Alternatively, suppose that the treatment difference δ has the student t prior distribution as in (2.9). Let the variances of δ, $Var(\delta) = \frac{1}{a^2}\frac{v}{v-2}$, have the same variances as those of normal prior distributions $N(0, \frac{\sigma^2}{n_0})$ with $n_0 = 8$, 22, and 89 under the degree of freedom v from 10 to 100. The corresponding posterior probabilities $P(\delta < \delta_2|\overline{z}_4)$ and $P(\delta > \delta_1|\overline{z}_4)$ with $\overline{z}_4 = \sqrt{\frac{\sigma^2}{4n}} \times z_0$, where $z_0 = -2.58$, -1.96 and -1, are listed in Table 2.2(a), (b) and (c), respectively. It can be seen from these tables that the influence on the degree of freedom v can be ignored as in Section 2.2.2 due to the minor change of posterior probabilities with the change of the degree of freedom v. For example, at $z_0 = -1.96$, the difference of $P(\delta < \delta_2|\overline{z}_4)$ between $v = 10$ and 100 is $0.9690 - 0.9666 = 0.0024$. Similar to the results of

normal prior distributions shown in Table 2.1, Tables 2.2(a), (b), and (c) have also shown that different precision of the prior information might change the decision of stopping a trial early. For example, when $z_0 = -1.96, v = 20$, and $\varepsilon_1 = \varepsilon_2 = 0.05$, the trial is stopped and the experimental treatment is recommended if $n_0 = 8$ or 22; whereas the trial need to be continued if $n_0 = 89$.

Comparing Table 2.1 with Tables 2.2(a), (b), and (c), it can be seen that the change of posterior probabilities $P(\delta < \delta_2|\overline{z}_4)$ and $P(\delta > \delta_1|\overline{z}_4)$ is slightly bigger to the student t priors than those to normal priors with the same change of the precision of prior information. For example, at $z_0 = -2.58$, for the normal prior, the change of the posterior probability $P(\delta < \delta_2|\overline{z}_4)$ from the scanty prior information ($n_0 = 8$) to considerable prior information ($n_0 = 89$) is,

$$\begin{aligned} P(\delta < \delta_2|\overline{z}_4, n_0 = 8) - P(\delta < \delta_2|\overline{z}_4, n_0 = 89) &= 0.9933 - 0.9623 \\ &= 0.0310; \end{aligned}$$

and for the student t prior, the change is,

$$\begin{aligned} &P(\delta < \delta_2|\overline{z}_4, n_0 = 8) - P(\delta < \delta_2|\overline{z}_4, n_0 = 89) \\ &= \begin{cases} 0.9921 - 0.9533 = 0.0389 & v = 10 \\ 0.9930 - 0.9613 = 0.0317 & v = 100. \end{cases} \end{aligned}$$

The same results above are found at other interim analyses.

Table 2.2(a): Posterior Probabilities of Different Student t Priors (v)

		$P(\delta < \delta_2 \mid \overline{z}_4)$			$P(\delta > \delta_1 \mid \overline{z}_4)$		
		n_0			n_0		
z_0	v	8	22	89	8	22	89
-2.58	10	0.9921	0.9863	0.9533	0.0079	0.0137	0.0467
	20	0.9926	0.9877	0.9581	0.0074	0.0123	0.0419
	30	0.9928	0.9881	0.9595	0.0072	0.0119	0.0405
	40	0.9929	0.9883	0.9602	0.0071	0.0117	0.0398
	50	0.9929	0.9884	0.9606	0.0071	0.0116	0.0394
	60	0.9929	0.9885	0.9608	0.0071	0.0115	0.0392
	70	0.9929	0.9885	0.9610	0.0071	0.0115	0.0390
	80	0.9930	0.9886	0.9611	0.0070	0.0114	0.0389
	90	0.9930	0.9886	0.9613	0.0070	0.0114	0.0388
	100	0.9930	0.9886	0.9613	0.0070	0.0114	0.0387

Table 2.2(b): Posterior Probabilities of Different Student t Priors (v)

		$P(\delta < \delta_2 \mid \overline{z}_4)$			$P(\delta > \delta_1 \mid \overline{z}_4)$		
		n_0			n_0		
z_0	v	8	22	89	8	22	89
-1.96	10	0.9666	0.9526	0.8966	0.0334	0.0475	0.1034
	20	0.9680	0.9559	0.9045	0.0320	0.0441	0.0955
	30	0.9684	0.9569	0.9068	0.0316	0.0431	0.0932
	40	0.9686	0.9574	0.9080	0.0314	0.0426	0.0920
	50	0.9687	0.9576	0.9087	0.0313	0.0424	0.0913
	60	0.9688	0.9578	0.9091	0.0312	0.0422	0.0909
	70	0.9689	0.9579	0.9094	0.0311	0.0421	0.0906
	80	0.9689	0.9580	0.9097	0.0311	0.0420	0.0904
	90	0.9689	0.9581	0.9098	0.0311	0.0419	0.0902
	100	0.9690	0.9582	0.9100	0.0310	0.0418	0.0900

Table 2.2(c): Posterior Probabilities of Different Student t Priors (v)

		$P(\delta < \delta_2 \mid \overline{z}_4)$			$P(\delta > \delta_1 \mid \overline{z}_4)$		
		n_0			n_0		
z_0	v	8	22	89	8	22	89
-1	10	0.8249	0.8022	0.7383	0.1751	0.1978	0.2617
	20	0.8276	0.8075	0.7467	0.1724	0.1925	0.2533
	30	0.8284	0.8091	0.7493	0.1716	0.1909	0.2507
	40	0.8288	0.8099	0.7506	0.1712	0.1902	0.2494
	50	0.8290	0.8103	0.7513	0.1710	0.1897	0.2487
	60	0.8291	0.8106	0.7518	0.1709	0.1894	0.2482
	70	0.8292	0.8108	0.7522	0.1708	0.1892	0.2478
	80	0.8293	0.8110	0.7524	0.1707	0.1890	0.2476
	90	0.8294	0.8111	0.7526	0.1706	0.1889	0.2474
	100	0.8294	0.8112	0.7528	0.1706	0.1888	0.2472

The robustness to change of tail probabilities of prior distributions

We continue using the above example to discuss the inferences on the change of tail probabilities of normal priors and student t priors. Assume that the tail probabilities of normal priors and student t priors, denoted by P_{tail}, are equal to 0.01, 0.02, 0.03, 0.04, and 0.05. The posterior probabilities $P(\delta < \delta_2 \mid \overline{z}_4)$ and $P(\delta > \delta_1 \mid \overline{z}_4)$ for normal priors and student t priors are listed in Table 2.3 and Tables 2.4(a), (b), and (c). The $\overline{z}_4 = \sqrt{\frac{\sigma^2}{4n}} \times z_0$, where z_0 = -2.58, -1.96, and -1.

Comparing Table 2.3 with Tables 2.4(a), (b) and (c), conclusions are consistent with those of the same change of variances of normal priors and student t priors,

that is, the changes of posterior probabilities $P(\delta < \delta_2|\overline{z}_4)$ and $P(\delta > \delta_1|\overline{z}_4)$ are less to normal prior distributions than those to student t prior distributions. For example, at $z_0 = -2.58$, the change of the posterior probability $P(\delta < \delta_2|\overline{z}_4)$ from $P_{tail} = 0.01$ to $P_{tail} = 0.05$, for the normal prior is,

$$\begin{aligned} P(\delta < \delta_2|\overline{z}_4, P_{tail} = 0.01) - P(\delta < \delta_2|\overline{z}_4, P_{tail} = 0.05) &= 0.9874 - 0.9771 \\ &= 0.0103; \end{aligned}$$

for the student t prior is,

$$\begin{aligned} &P(\delta < \delta_2|\overline{z}_4, P_{tail} = 0.01) - P(\delta < \delta_2|\overline{z}_4, P_{tail} = 0.05) \\ &= \begin{cases} 0.8083 - 0.7227 = 0.0856 & v = 10 \\ 0.8214 - 0.7481 = 0.0733 & v = 100. \end{cases} \end{aligned}$$

2.2.4 Summary

The student t prior distribution (2.9) for the treatment difference δ is discussed. It can be concluded that the posterior inferences on the treatment difference δ is robust to the degree of freedom v of the student t prior. By comparing the student t prior with the normal prior, it can be obtained that posterior probabilities $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1|z_1, z_2, ..., z_j)$, $j = 1, 2, ..., l$, are more robust to the normal prior than those to the student t prior with the same change of variances or tail probabilities of the prior distributions. Hence, it is reasonable to assume that the treatment difference δ has the normal prior distribution if the sample mean is not far from the mean of this prior distribution.

Table 2.3: Posterior Probabilities of Different(tail) Normal Priors

z_0	P_{tail}	$P(\delta < \delta_2 \mid \overline{z}_4)$	$P(\delta > \delta_1 \mid \overline{z}_4)$
-2.58	0.01	0.9771	0.0232
	0.02	0.9818	0.0184
	0.03	0.9844	0.0158
	0.04	0.9861	0.0141
	0.05	0.9874	0.0129
-1.96	0.01	0.9352	0.0650
	0.02	0.9439	0.0564
	0.03	0.9489	0.0513
	0.04	0.9524	0.0478
	0.05	0.9551	0.0451
-1	0.01	0.7802	0.2200
	0.02	0.7910	0.2093
	0.03	0.7977	0.2026
	0.04	0.8026	0.1976
	0.05	0.8065	0.1937

Table 2.4(a): Posterior Probabilities of Different(tail) Student t Priors(v)

z_0	P_{tail}	v	$P(\delta < \delta_2 \mid \overline{z}_4)$	$P(\delta > \delta_1 \mid \overline{z}_4)$
-2.58	0.01	10	0.7227	0.2773
		50	0.7454	0.2546
		100	0.7481	0.2519
	0.02	10	0.7536	0.2464
		50	0.7718	0.2282
		100	0.7740	0.2260
	0.03	10	0.7754	0.2246
		50	0.7909	0.2091
		100	0.7927	0.2073
	0.04	10	0.7931	0.2069
		50	0.8065	0.1935
		100	0.8081	0.1919
	0.05	10	0.8083	0.1917
		50	0.8200	0.1800
		100	0.8214	0.1786

Table 2.4(b): Posterior Probabilities of Different(tail) Student t Priors(v)

z_0	P_{tail}	v	$P(\delta < \delta_2 \mid \overline{z}_4)$	$P(\delta > \delta_1 \mid \overline{z}_4)$
-1.96	0.01	10	0.6725	0.3275
		50	0.6918	0.3082
		100	0.6941	0.3059
	0.02	10	0.6978	0.3022
		50	0.7141	0.2859
		100	0.7160	0.2840
	0.03	10	0.7161	0.2839
		50	0.7305	0.2695
		100	0.7322	0.2678
	0.04	10	0.7312	0.2688
		50	0.7442	0.2558
		100	0.7457	0.2543
	0.05	10	0.7445	0.2555
		50	0.7563	0.2437
		100	0.7577	0.2423

Table 2.4(c): Posterior Probabilities of Different(tail) Student t Priors(v)

z_0	P_{tail}	v	$P(\delta < \delta_2 \mid \overline{z}_4)$	$P(\delta > \delta_1 \mid \overline{z}_4)$
-1	0.01	10	0.5898	0.4102
		50	0.6008	0.3992
		100	0.6021	0.3979
	0.02	10	0.6037	0.3963
		50	0.6134	0.3866
		100	0.6146	0.3854
	0.03	10	0.6140	0.3860
		50	0.6229	0.3771
		100	0.6240	0.3760
	0.04	10	0.6227	0.3774
		50	0.6310	0.3690
		100	0.6320	0.3680
	0.05	10	0.6304	0.3696
		50	0.6383	0.3617
		100	0.6392	0.3608

2.3 Clinical Trials on Normally Distributed Response with Unknown Variance

The frameworks of Bayes sequential method in group sequential clinical trials above assume that the main outcome variable for a clinical trial is from the normal distribution $N(\mu, \frac{\sigma^2}{2})$ with known variance σ^2. The μ is a measure of treatment effect. However, it is often difficult to know the exact value of σ^2 in practice. This issue is considered as follows.

2.3.1 The Framework

The prior distribution

Consider the clinical trial as described in Section 2.1. The main outcome measure X for the clinical trial is normally distributed with probability density functions $N(\mu_e, \frac{\sigma^2}{2})$ and $N(\mu_s, \frac{\sigma^2}{2})$ for the experimental and the standard treatments, respectively, where the variance σ^2 is unknown. The parameter of interest is the treatment difference $\delta = \mu_e - \mu_s$. For the convenience of notation, assume that the clinical trial is monitored at every $2n$ patients with n for each treatment. At each analysis j, $j = 1, 2, ..., l$,

$$Z_j = \overline{X}_{je} - \overline{X}_{js} \sim N(\delta, \frac{\sigma^2}{n}), \tag{2.13}$$

is a sufficient statistic of the treatment difference δ given σ^2, where $\overline{X}_{je} \sim N(\mu_e, \frac{\sigma^2}{2n})$ and $\overline{X}_{js} \sim N(\mu_s, \frac{\sigma^2}{2n})$ are group sample means for the experimental and standard treatments, respectively.

Let $r = \frac{1}{\sigma^2}$ and let R be the corresponding random variable of r. Suppose the prior distribution of δ given $R = r$ is the normal distribution,

$$\delta | R = r \sim w(\delta | r) = N(\nu_0, \frac{1}{n_0 r}), \tag{2.14}$$

and the marginal prior distribution of R is the gamma distribution,

$$R \sim w(r) = \Gamma(\alpha, \beta), \tag{2.15}$$

where parameters $\alpha > 0$ and $\beta > 0$.

The Posterior Distribution

It is known that for the normal distribution $N(\delta, \frac{\sigma^2}{n})$(or $N(\delta, \frac{1}{nr})$) likelihood, the normal-gamma prior is conjugate prior density for parameters $(\delta, \frac{1}{\sigma^2})$(or (δ, r)). At each analysis j, $j = 1, 2, ..., l$, assume values $Z_1 = z_1$, $Z_2 = z_2$, ..., $Z_j = z_j$ have been observed and $f(z_1, z_2, ..., z_j|\delta, r)$ is the probability density function of $(z_1, z_2, ..., z_j)$ given (δ, r), that is,

$$f(z_1, z_2, ..., z_j|\delta, r) = \left(\sqrt{\frac{nr}{2\pi}}\right)^j e^{-\frac{nr}{2}\sum_{i=1}^{j}(z_i-\delta)^2} \tag{2.16}$$

Let $\overline{z}_j = \frac{1}{j}\sum_{i=1}^{j} z_i$. The posterior probability distribution of δ given $R = r$ by Bayes theorem is

$$\begin{aligned} w(\delta|z_1, z_2, ..., z_j, r) &\propto f(z_1, z_2, ..., z_j|\delta, r)\ w(\delta|r) \\ &\propto e^{-\frac{jnr(\overline{z}_j-\delta)^2}{2}} e^{-\frac{n_0 r(\delta-\nu_0)^2}{2}} \\ &\propto e^{-\frac{(jn+n_0)r}{2}(\delta-\frac{jn\overline{z}_j+n_0\nu_0}{jn+n_0})^2}. \end{aligned}$$

It then follows

$$\delta \sim w(\delta|z_1, z_2, ..., z_j, r) = N\left(\frac{jn\overline{z}_j + n_0\nu_0}{jn + n_0},\ \frac{1}{(jn + n_0)r}\right). \tag{2.17}$$

The distribution of R given $Z_1 = z_1$, $Z_2 = z_2$, ..., $Z_j = z_j$ is

$$w(r|z_1, z_2, ..., z_j) \propto \int f(z_1, z_2, ..., z_j|\delta, r)\ w(\delta|r)\ w(r)\ d\delta$$

$$
\begin{aligned}
&\propto w(r)\int(\sqrt{r})^j e^{-\frac{nr}{2}\sum_{i=1}^{j}(z_i-\delta)^2}\sqrt{r}e^{-\frac{n_0 r}{2}(\delta-\nu_0)^2}d\delta \\
&= w(r)\, r^{\frac{j}{2}} e^{-\frac{nr}{2}\sum_{i=1}^{j} z_i^2-\frac{n_0 r}{2}\nu_0^2}\int\sqrt{r}e^{-\frac{r}{2}[(nj+n_0)\delta^2-2(nj\overline{z}_j+n_0\nu_0)\delta]}d\delta \\
&\propto w(r)\, r^{\frac{j}{2}} e^{-\frac{nr}{2}\sum_{i=1}^{j} z_i^2-\frac{n_0 r}{2}\nu_0^2} e^{\frac{r}{2}\frac{(nj\overline{z}_j+n_0\nu_0)^2}{nj+n_0}} \\
&= w(r)\, r^{\frac{j}{2}} e^{-\frac{nr}{2}\sum_{i=1}^{j}(z_i-\overline{z}_j)^2-\frac{n_0 njr}{2(nj+n_0)}(\overline{z}_j-\nu_0)^2} \\
&\propto r^{\alpha+\frac{j}{2}-1} e^{-\left(\beta+\frac{n}{2}\sum_{i=1}^{j}(z_i-\overline{z}_j)^2+\frac{1}{2}\frac{jnn_0}{jn+n_0}(\overline{z}_j-\nu_0)^2\right)r}.
\end{aligned}
$$

Hence, $w(r|z_1, z_2, ..., z_j)$ is the gamma distribution with shape parameter equal to $\alpha + \frac{j}{2}$, and scale parameter equal to $\beta + \frac{n}{2}\sum_{i=1}^{j}(z_i - \overline{z}_j)^2 + \frac{1}{2}\frac{jnn_0}{jn+n_0}(\overline{z}_j - \nu_0)^2$. Let

$$
\begin{aligned}
a &= \frac{jn\overline{z}_j + n_0\nu_0}{jn + n_0}, \\
\beta_1 &= \beta + \frac{n}{2}\sum_{i=1}^{j}(z_i - \overline{z}_j)^2 + \frac{n_0 nj}{2(nj + n_0)}(\overline{z}_j - \nu_0)^2. \qquad (2.18)
\end{aligned}
$$

Then,

$$
R \sim w(r|z_1, z_2, ..., z_j) = \Gamma(r|\alpha + \frac{j}{2},\ \beta_1). \qquad (2.19)
$$

The posterior probability density function of (δ, r) by (2.17) and (2.19) is

$$
\begin{aligned}
w(\delta, r|z_1, z_2, ..., z_j) &= w(\delta|z_1, z_2, ..., z_j, r)\ w(r|z_1, z_2, ..., z_j) \\
&= N(\delta\ |a,\ \frac{1}{(jn + n_0)r})\ \Gamma(r\ |\alpha + \frac{j}{2},\ \beta_1) \\
&= \frac{\sqrt{(jn + n_0)r}}{\sqrt{2\pi}} e^{-\frac{(jn+n_0)r}{2}(\delta-a)^2}\frac{\beta_1^{\alpha+\frac{j}{2}}}{\Gamma(\alpha + \frac{j}{2})} r^{\alpha+\frac{j}{2}-1} e^{-\beta_1 r} \\
&= \frac{\sqrt{jn + n_0}}{\sqrt{2\pi}}\frac{\beta_1^{\alpha+\frac{j}{2}}}{\Gamma(\alpha + \frac{j}{2})}\ r^{\alpha+\frac{j}{2}+\frac{1}{2}-1} e^{-(\frac{(jn+n_0)(\delta-a)^2}{2}+\beta_1)r}. \qquad (2.20)
\end{aligned}
$$

The posterior probability density function of δ is then,

$$\begin{aligned} w(\delta|z_1, z_2, ..., z_j) &= \int_0^\infty w(\delta, r|z_1, z_2, ..., z_j)\, dr \\ &= \frac{\sqrt{jn+n_0}}{\sqrt{2\pi}} \frac{\beta_1^{\alpha+\frac{j}{2}}}{\Gamma(\alpha+\frac{j}{2})} \frac{\Gamma(\alpha+\frac{j}{2}+\frac{1}{2})}{(\frac{(jn+n_0)(\delta-a)^2}{2}+\beta_1)^{\alpha+\frac{j}{2}+\frac{1}{2}}} \\ &= \frac{\left(\frac{(jn+n_0)(2\alpha+j)}{2\beta_1}\right)^{\frac{1}{2}} \Gamma(\frac{(2\alpha+j)+1}{2})}{((2\alpha+j)\pi)^{\frac{1}{2}}\Gamma(\frac{2\alpha+j}{2})} \left[1+\frac{\frac{(jn+n_0)(2\alpha+j)}{2\beta_1}}{2\alpha+j}(\delta-a)^2\right]^{-\frac{(2\alpha+j)+1}{2}} \\ &= t(\delta \mid 2\alpha+j,\ a,\ \frac{(jn+n_0)(2\alpha+j)}{2\beta_1}), \end{aligned} \tag{2.21}$$

which is a t distribution with degree of freedom $2\alpha+j$, location parameter a, and precision $\frac{(jn+n_0)(2\alpha+j)}{2\beta_1}$. The values of a and β_1 are obtained by (2.18).

The stopping rule

At each analysis j, $j = 1, 2, ..., l$, the clinical trial is suggested being stopped if either

$$P(\delta < \delta_2|z_1, z_2, ..., z_j) = \int_{-\infty}^{\delta_2} w(\delta|z_1, z_2, ..., z_j)\, d\delta > 1-\varepsilon_1, \tag{2.22}$$

or

$$P(\delta > \delta_1|z_1, z_2, ..., z_j) = \int_{\delta_1}^{\infty} w(\delta|z_1, z_2, ..., z_j)\, d\delta > 1-\varepsilon_2, \tag{2.23}$$

where $w(\delta|z_1, z_2, ..., z_j)$ is obtained by (2.21) and ε_1 and ε_2 are small positive numbers; Otherwise the trial needs to be continued.

2.3.2 The Prior Information

Influences of the prior information of (δ, σ^2) on the posterior distribution of δ is studied by continuing the example of Section 2.2.2. Let the prior distribution of δ given σ^2 be the normal distribution $N(0, \frac{\sigma^2}{n_0})$ as in (2.14), where $n_0 = 8$, 22, and 89; and let the marginal prior distribution of $R = \frac{1}{\sigma^2}$ be the gamma distribution $\Gamma(\alpha, \frac{1}{2})$ as in (2.15), where $\alpha = 0.2$, 0.5, 0.8, 1, 1.5 and 2. If $\alpha = \frac{1}{2}$ or an even natural number, then the gamma distribution $\Gamma(\alpha, \frac{1}{2})$ is a $\chi^2(2\alpha)$ distribution. The Monte Carlo simulations are used in the study.

The inference of the parameter α

Let the sequential sample Z_j, $j = 1, 2, ..., 5$, of the example be from the normal distribution $N(-0.01, \frac{0.5}{n})$. The value $\delta = -0.01$ is the mean of 1000 δ which are derived from the normal prior distribution $N(0, \frac{\sigma^2}{n_0})$ with $\sigma^2 = 0.5$. The $\sigma^2 = 0.5$ is the average of 1000 σ^2 where $\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$. The average posterior probabilities of δ as in (2.22) and (2.23) from 1000 simulations are denoted by Mean $P(\delta < \delta_2 | j = 4)$ and Mean $P(\delta > \delta_1 | j = 4)$, respectively, at the analysis $j = 4$ in Table 2.5. The corresponding standard errors are given in the brackets. For each $n_0 = 8$, 22, and 89, the differences among the posterior probabilities within two consecutive values of α in Table 2.5 are around 0.002. These differences might be negligible. The differences in the average posterior probabilities $P(\delta < \delta_2 | z_1, z_2, z_3, z_4)$ with $\alpha =$ 0.2 and 2 are equal to 0.5589 - 0.5497 = 0.0092 when $n_0 = 8$ for the scanty prior information of δ; equal to 0.5566 - 0.5470 = 0.0096 when $n_0 = 22$ for the moderate prior information of δ; and equal to 0.5483 - 0.5383 = 0.01 when $n_0 = 89$ for the considerable prior information of δ. These differences are close to each other. The same results are found when the values of δ and σ^2 are changed. The example shows that the posterior inferences on δ is reasonable robust to the parameter α.

Table 2.5: Posterior Probabilities of δ under Different n_0 and α

n_0	α	Mean $P(\delta < \delta_2 \mid j = 4)$ (se)	Mean $P(\delta > \delta_1 \mid j = 4)$ (se)
8	0.2	0.5497 (0.0080)	0.4503 (0.0080)
	0.5	0.5518 (0.0084)	0.4482 (0.0084)
	0.8	0.5536 (0.0087)	0.4464 (0.0087)
	1	0.5547 (0.0089)	0.4453 (0.0089)
	1.5	0.5570 (0.0094)	0.4430 (0.0094)
	2	0.5589 (0.0097)	0.4411 (0.0097)
22	0.2	0.5470 (0.0076)	0.4530 (0.0076)
	0.5	0.5492 (0.0080)	0.4508 (0.0080)
	0.8	0.5511 (0.0083)	0.4489 (0.0083)
	1	0.5522 (0.0085)	0.4478 (0.0085)
	1.5	0.5546 (0.0089)	0.4454 (0.0089)
	2	0.5566 (0.0093)	0.4434 (0.0093)
89	0.2	0.5383 (0.0061)	0.4617 (0.0061)
	0.5	0.5405 (0.0065)	0.4595 (0.0065)
	0.8	0.5424 (0.0068)	0.4576 (0.0068)
	1	0.5435 (0.0070)	0.4565 (0.0070)
	1.5	0.5461 (0.0075)	0.4539 (0.0075)
	2	0.5483 (0.0078)	0.4517 (0.0078)

The comparison with the situation of σ^2 known

Assume the sequential sample Z_j, $j = 1, 2, ..., 5$, in the example is from the normal distribution $N(-0.01, \frac{0.5}{n})$. Table 2.6 lists average posterior probabilities $P(\delta < \delta_2|z_1, z_2, z_3, z_4)$ and $P(\delta > \delta_1|z_1, z_2, z_3, z_4)$ with the variance of the normal response known $\sigma^2 = 0.5$, denoted by $P(\cdot|\overline{z}_4)$, and the variance unknown $\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$, denoted by $P(\cdot|j = 4)$ from the Monte Carlo simulations. We choose $\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$ since $E(\frac{1}{\sigma^2}) = 2$. The values in brackets are corresponding standard errors. It can been seen from Table 2.6 that the posterior probabilities $P(\delta < \delta_2|z_1, z_2, z_3, z_4)$ and $P(\delta > \delta_1|z_1, z_2, z_3, z_4)$ with $\sigma^2 = 0.5$ are similar to those with $\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$. This example shows that it is reasonable to assume $\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$.

Table 2.6: Posterior Probabilities of δ with σ^2 Known and Unknown

	$\sigma^2 = 0.5$		$\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$	
	Mean	Mean	Mean	Mean
n_0	$P(\delta < \delta_2\|\overline{z}_4)$	$P(\delta > \delta_1\|\overline{z}_4)$	$P(\delta < \delta_2\|j = 4)$	$P(\delta > \delta_1\|j = 4)$
8	0.5294 (0.0090)	0.4706 (0.0090)	0.5294(0.0092)	0.4706 (0.0092)
22	0.5284 (0.0086)	0.4716 (0.0086)	0.5281(0.0087)	0.4719 (0.0087)
89	0.5244 (0.0073)	0.4756 (0.0073)	0.5233(0.0072)	0.4767 (0.0072)

The inference of the change of σ^2

Assume that the sequential sample Z_{1j} and Z_{2j} are from the normal distribution $N(-0.01, \frac{\sigma_1^2}{n})$ and $N(-0.01, \frac{\sigma_2^2}{n})$, respectively. Let $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 1$. The corresponding posterior probabilities are denoted by $P_1(\cdot|\cdot)$ and $P_2(\cdot|\cdot)$ respectively in Table 2.7. Table 2.7 only lists the posterior probabilities $P_k(\delta < \delta_2|z_1, z_2, z_3, z_4, \sigma_k^2)$ since $P_k(\delta > \delta_1|z_1, z_2, z_3, z_4, \sigma_k^2) = 1 - P_k(\delta < \delta_2|z_1, z_2, z_3, z_4, \sigma_k^2)$, $k = 1, 2$, when $\delta_2 = \delta_1 = 0$.

Table 2.7 shows that the change of the posterior probabilities $P(\delta < \delta_2|z_1, z_2, z_3, z_4)$ with the change of σ^2 is bigger at σ^2 known than those at $\sigma^2 \sim \Gamma(\alpha, \frac{1}{2})$.

Table 2.7: Posterior Probabilities of δ with Different σ^2

	$\sigma^2 = 0.5$ known		$\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$	
	Mean	Mean	Mean	Mean
n_0	$P_1(\delta < \delta_2\|\overline{z}_4)$	$P_2(\delta < \delta_2\|\overline{z}_4)$	$P_1(\delta < \delta_2\|j = 4)$	$P_2(\delta < \delta_2\|j = 4)$
8	0.5294 (0.0090)	0.5194 (0.0090)	0.5294(0.0092)	0.5204 (0.0099)
22	0.5284 (0.0086)	0.5187 (0.0086)	0.5281(0.0087)	0.5196 (0.0094)
89	0.5244 (0.0073)	0.5161 (0.0073)	0.5233(0.0072)	0.5165 (0.0079)

The summary

Under the framework described in Section 2.3.1, we have that the posterior probabilities $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1|z_1, z_2, ..., z_j)$ are reasonably robust to the parameter α; when we choose the prior distribution $\frac{1}{\sigma^2} \sim \Gamma(\alpha, \frac{1}{2})$, where $E(\frac{1}{\sigma^2}) = 2\alpha$, the posterior probabilities $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1|z_1, z_2, ..., z_j)$ are similar to those of σ^2 being known and equal to $\frac{1}{2\alpha}$; if we change σ^2 from 0.5 to 1, the corresponding change of posterior probability $P(\delta < \delta_2|z_1, z_2, ..., z_j)$ is smaller in assuming $\frac{1}{\sigma^2} \sim \Gamma(1, \frac{1}{2})$ than that in assuming $\sigma^2 = 0.5$. Therefore the normal-gamma prior described in Section 2.3.1 is recommended when we do not have enough prior information on the variance σ^2.

2.4 Discussion

The framework of Bayes sequential methods in group sequential clinical trials described in Section 2.1 is based on the work of Freedman and Spiegelhalter(1989). It can be generalized to clinical trials with binomially distributed responses and

survival time data, although these clinical trials can be approximated by clinical trials with normal distribution responses.

In the framework, at each analysis j, $j = 1, 2, ..., l$, after observing $Z_1 = z_1$, $Z_2 = z_2$, ..., $Z_j = z_j$, we look at the posterior probabilities $P(\delta < \delta_2 | z_1, z_2, ..., z_j)$ as in (2.4), and $P(\delta > \delta_1 | z_1, z_2, ..., z_j)$ as in (2.5). If the posterior probability $P(\delta < \delta_2 | z_1, z_2, ..., z_j)$ (or $P(\delta > \delta_1 | z_1, z_2, ..., z_j)$) is greater than some specified value, then the clinical trial may be stopped. The calculation of corresponding posterior probabilities could be found in Chapter 5 for clinical trials with binary response and in Chapter 6 for clinical trials with survival time data.

The framework of Bayes sequential methods in groups sequential clinical trials which we have discussed are based on posterior probabilites $P(\delta < \delta_2 | z_1, z_2, ..., z_j)$ and $P(\delta > \delta_1 | z_1, z_2, ..., z_j)$. Other criteria have also been suggested, for example, at each analysis j, $j = 1, 2, ..., l$, we can also look at the posterior expectation of the treatment difference, denoted by $E(\delta | z_1, z_2, ..., z_j)$. The clinical trial may be suggested to be stopped if the expectation $E(\delta | z_1, z_2, ..., z_j)$ is greater than some specified value. More generally, let $g(\delta)$ be some quantity of interest. At each analysis j, $j = 1, 2, ..., l$, we may look at the posterior expectation of $g(\delta)$, that is, $E(g(\delta) | z_1, z_2, ..., z_j)$, and decide whether to stop the trial based on the value of $E(g(\delta) | z_1, z_2, ..., z_j)$. If $g(\delta)$ is the indicator function for the interval $(-\infty, \delta_1)$ (or (δ_2, ∞)), then $E(g(\delta) | z_1, z_2, ..., z_j) = P(\delta < \delta_2 | z_1, z_2, ..., z_j)$ (or $= P(\delta > \delta_1 | z_1, z_2, ..., z_j)$).

Chapter 3

Bayes Sequential Decision Theory

3.1 Introduction

Decision theory provides the framework for combining subjective distributions with actions. The method of a sequential decision procedure is to look at a sequence of observations one at a time and to decide after each observation whether to stop sampling and make a decision immediately or to continue sampling and make a decision sometime later.

Bayes sequential decision theory used in group sequential clinical trials is briefly described in this chapter. Details can be found in Berger(1985), Degroot(1970), and Ferguson(1967), etc.

3.1.1 Basic Elements of A Sequential Decision Procedure

The basic elements of a sequential decision procedure considered in the study are

1) a parameter δ whose values are in the parameter space Δ and its prior distribution $w(\delta)$ which is from the space of prior distributions Δ^*;

Consider a clinical trial comparing two treatments described in Section 2.1. The parameter of interest is the treatment difference $\delta = \mu_e - \mu_s$. Its space Δ is equal to the real line $\mathcal{R}$. The prior distribution of the parameter of interest δ is

assumed to be the normal distribution

$$\delta \sim w(\delta) = N(\nu_0, \tau_0^2).$$

The space of prior distributions is the normal distribution family, $\Delta^* = \{N(\delta, \tau^2) : \delta \in \mathcal{R}, \tau^2 > 0\}$.

2) a decision d which is from decision space D;

When the above clinical trial is terminated, a decision d will be chosen from the decision space $D = \{$experimental treatment, standard treatment$\}$.

3) a sequential random sample X_1, X_2,...; assume that the conditional g.p.d.f of each X_m, $m = 1, 2...$, is $f(\cdot|\delta)$ for every $\delta \in \Delta$;

In the clinical trial described in Section 2.1, we have the sequential sample Z_j, $j = 1, 2, ..., l$, which are from the normal distribution $N(\delta, \frac{\sigma^2}{n_j})$ given δ with the variance $\frac{\sigma^2}{n_j}$ known.

4) a loss function $L(\delta, d)$, a real value function defined on $\Delta \times D$, which represents the loss when δ is true and decision d is chosen;

5) the cost functions are denoted by $\{c_m(\delta, x_1, x_2, ..., x_m), m = 1, 2, ...\}$; the value of $c_m(\delta, x_1, x_2, ..., x_m)$ represents the cost of taking observations $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$ and stopping sampling when δ is the true value of the parameter.

3.1.2 Loss Function and Cost Function

Bayes sequential decision theory has not been widely used in clinical trials because of the computational complexity of Bayes inferences and the difficulty of specifying loss and cost functions which can describe or measure the cost of decisions and the cost of carrying out a clinical trial. Lewis and Berry(1994), Lewis(1996) have applied Bayes sequential decision theory with piecewise continuous loss functions in group sequential clinical trials of binomial response variables. R.J.Sylvester(1988)

used Bayes decision theory in a one stage Phase II clinical trial with binary response outcome and a two-point prior distribution for a new drug response rate. He has suggested a loss function which involves the patient horizon and the amount of the difference between the new drug response rate and the standard rate.

Generally, the aims of the ideas considered in loss functions and cost functions are to maximise the expected experimental treatment benefit over a patient horizon and the loss in efficacy will be taken to be proportional to the magnitude of the advantage of the treatment difference(Anscombe 1963, Berry et al 1992, Whitehead 1992). This area has been little studied.

The simplest form of the loss function is piecewise continuous. This will be used in group sequential clinical trials comparing two treatments with normal distribution response variables in Chapter 4. Although we may not rely on such a simple loss function to make decisions in real clinical trials, this is a start to introduce Bayes sequential decision theory into group sequential clinical trials.

Following this, the more complicated loss and cost functions will be discussed. For example, suppose a group sequential clinical trial is designed to compare an experimental treatment with the standard treatment. The parameter of interest is the treatment difference δ. Assume that the experimental treatment is to be regarded as better than the standard treatment if the treatment difference $\delta \geq \delta_0 > 0$ and that the experimental treatment is not to be recommended otherwise. The δ_0 is the break-even value of the treatment difference δ. Let z_j, $j = 1, 2, ..., l$, be the observation values of the group sequential sample Z_j, $j = 1, 2, ..., l$, respectively, which are used to test the treatment difference δ. The l is the number of groups or analyses. Let d_e and d_s be decisions of choosing the experimental and the standard treatments respectively, after the clinical trial is

terminated. The loss functions can be defined as,

$$L(\delta, d_s) = \begin{cases} 0 & \delta \leq \delta_0 \\ K(\delta - \delta_0)t & \delta > \delta_0, \end{cases} \tag{3.1}$$

$$L(\delta, d_e) = \begin{cases} -K(\delta - \delta_0)t & \delta < \delta_0 \\ -K(\delta - \delta_0)t & \delta \geq \delta_0, \end{cases} \tag{3.2}$$

where K denotes the difference in cost of further treatment between a patient who takes the experimental treatment and a patient who takes the standard treatment (assume $K > 0$); t expresses the patient horizon, ie. the average number of patients who are treated with the experimental treatment after the trial before a second experimental treatment, which is as least as good, is found; n_j is the group sample size in each treatment at analysis j, $j = 1, 2, ..., l$.

The loss functions $L(\delta, d_s)$ in (3.1) and $L(\delta, d_e)$ in (3.2) show that if the treatment difference $\delta \leq \delta_0$, that is, the experimental treatment is not better than the standard treatment, then there is no loss in choosing the standard treatment, but there is a cost $-K(\delta - \delta_0)t > 0$ in choosing the experimental treatment; if the treatment difference $\delta \geq \delta_0$, that is, the experimental treatment is better than the standard treatment, then there is a gain(negative cost) $-K(\delta - \delta_0)t < 0$ in choosing the experimental treatment, but there is a cost $K(\delta - \delta_0)t > 0$ in choosing the standard treatment. The cost or gain of making a decision is proportional to the patient horizon t and the treatment efficacy $\delta - \delta_0$.

Chapter 5 will apply Bayes sequential decision theory with the form of loss functions as in (3.1) to group sequential clinical trials with binomial distribution response variables.

3.1.3 Bayes Risk, Bayes Decision and the Expected Risk

Consider a sequential decision procedure with the basic elements specified in Section 3.1.1. Suppose the sequential samples $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$, $m = 1, 2, ...$, have been observed. Let $w^m = w(\delta|x_1, x_2..., x_m)$ be the posterior distribution of δ after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$, which is,

$$\delta|x_1, x_2, ..., x_m \sim w(\delta|x_1, x_2, ..., x_m) = \frac{f(x_1, x_2, ..., x_m|\delta)\ w(\delta)}{\int_\Delta f(x_1, x_2, ..., x_m|\delta)\ w(\delta)\ d\delta}. \quad (3.3)$$

The *Bayes risk* of stopping sampling, denoted by $r_0(w^m, m)$, is defined as the greatest lower bound of expected losses, or risks, with respect to the posterior distribution $w(\delta|x_1, x_2, ..., x_m)$ among decisions $d \in D$, that is,

$$r_0(w^m, m) = \inf_{d \in D} E_{w(\delta|x_1, x_2, ..., x_m)}(L(\delta, d) + c_m(\delta, x_1, x_2, ..., x_m)). \quad (3.4)$$

After sampling is terminated, a decision $d \in D$ is called a *Bayes decision* if its risk $E_{w(\delta|x_1, x_2, ..., x_m)}(L(\delta, d) + c_m(\delta, x_1, x_2, ..., x_m))$ is equal to the Bayes risk $r_0(w^m, m)$ in (3.4). In clinical trials of comparing different treatments, the decision space D is a set of finite treatments, that is, a set of finite elements, so we can always get a Bayes decision in clinical trials.

On the other hand, if sampling needs to be continued after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$, the *expected risk* from continuing sampling to observe the next observation X and to choose a decision $d \in D$ later, expressed by $E^* r_0(w^m(X), m + 1)$, is the expectation of the Bayes risk $r_0(w^m(X), m + 1)$ with respect to the predictive density of x, $f(x|x_1, x_2, ..., x_m)$, that is,

$$E^* r_0(w^m(X), m + 1) = \int_{\mathcal{X}} r_0(w^m(X = x), m + 1)\ f(x|x_1, x_2, ..., x_m)\ dx, \quad (3.5)$$

where $w^m(X = x) = w(\delta|x_1, x_2, ..., x_m, x)$ is the posterior distribution of δ after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$ and $X = x$; $r_0(w^m(X = x), m + 1)$ is

the Bayes risk after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$ and $X = x$, and $m+1$ means that the risk is from these $m+1$ observations; $f(x|x_1, x_2, ..., x_m)$ is the *predictive density function* of x after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$, that is,

$$f(x|x_1, x_2, ..., x_m) = \int_\Delta f(x|\delta)\, w(\delta|x_1, x_2, ..., x_m)\, d\delta. \tag{3.6}$$

3.2 Bayes Sequential Decision Theory

3.2.1 Bayes Sequential Decision Procedure

A sequential decision procedure involves looking at a sequence of observations one at a time and deciding after each observation whether to stop sampling and make a decision immediately or to continue sampling and make a decision sometime later. It has two components. One component is called a *stopping rule*, or a sampling plan, which specifies whether sampling should be stopped and a decision $d \in D$ should be chosen without further observations or whether another sample X should be observed after observing values $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$, m=1, 2, ... The second component of a sequential decision procedure may be called a *decision rule*. It specifies the decision $d(x_1, x_2, ..., x_m) \in D$ to be chosen for each possible set of observed values $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$ after which sampling might be terminated.

A Bayes sequential decision procedure, or an optimal sequential decision procedure, is a procedure for which the total risk(at least one observation is to be taken in clinical trials) is minimised. For a bounded sequential procedure, in which there is a fix number of observations N_0 that can be taken, at each analysis, after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$, $m = 1, 2, ..., N_0 - 1$, the stopping rule

of a Bayes sequential decision procedure is applied by comparing the Bayes risk from stopping sampling $r_0(w^m, m)$, obtained by (3.4), with the expected risk from the optimal continuation of sampling and then choosing a decision $d \in D$ later, which is denoted by $r_{N_0-m}(w^m, m)$ and discussed in Section 3.2.2. If the Bayes risk from stopping sampling is less than the risk from the optimal continuing sampling, that is,

$$r_0(w^m, m) \leq r_{N_0-m}(w^m, m), \tag{3.7}$$

then sampling may be stopped and a decision $d \in D$ would need to be chosen. Otherwise the sampling is continued.

The decision rule of a Bayes sequential decision procedure requires that decision functions $d(x_1, x_2, ..., x_m)$, $m = 1, 2, ...$, are always specified by Bayes decisions in D. That is, if sampling is to be terminated after values $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$ have been observed and a decision $d = d(x_1, x_2, ..., x_m) \in D$ is chosen, then the risk of this decision d is equal to the Bayes risk, that is,

$$E_{w(\delta|x_1,x_2,...,x_m)}(L(\delta, d) + c_m(\delta, x_1, x_2, ..., x_m)) = r_0(w^m, m). \tag{3.8}$$

3.2.2 Bounded Bayes Sequential Decision Procedure

The bounded sequential decision procedure involves stopping sampling after, at most, N_0 samples. Corresponding to group sequential clinical trials comparing two treatments, let the maximum number of groups be l and the group sample size be n for each treatment, we have $N_0 = 2nl$. If the value of N_0 is large enough, then the bounded sequential procedure is in fact also the unbounded procedure. For example, if we can reach the point that as the trial progresses the cost of enrolling any additional groups of patients is greater than the cost savings achieved by any possible decrease in making Bayes decision, then we should terminate the trial

at this stage, j, say, and make a decision. Let $N_0 \geq 2nj$, the optimum bounded sequential decision procedure will be the optimum unbounded sequential decision procedure.

Backward induction(Berger 1985, Degroot 1970) is used to construct bounded Bayes sequential decision procedures. At each observation $X_m, m = 1, 2, ..., N_0-1$, $X_1 = x_1$, $X_2 = x_2$, ..., $X_m = x_m$ have been observed, there are not more than N_0-m observations which can be taken; we need to compare the risk from stopping the trial, denoted by $r_0(w^m, m)$, with the risk from the optimal continuation of the trial, denoted by $r_{N_0-m}(w^m, m)$, in which not more than $N_0 - m$ observations can be taken. If

$$r_0(w^m, m) \leq r_{N_0-m}(w^m, m), \tag{3.9}$$

then we make a decision without further observations; otherwise we continue sampling to observe the next sample X. The calculations of $r_{N_0-m}(w^m, m)$ can be obtained by the following recursive relationships,

$$\begin{aligned} r_0(w^m, m) &= \inf_{d \in D} E_{w(\delta|x_1,x_2,...,x_m)}(L(\delta, d) + c_m(\delta, x_1, x_2, ..., x_m)), \\ &\qquad m = 1, 2, ..., N_0, \end{aligned} \tag{3.10}$$

$$\begin{aligned} r_k(w^m, m) &= \min\{r_0(w^m, m),\ E^* r_{k-1}(w^m(X), m+1)\}, \\ &\qquad m, k = 1, ..., N_0 - 1,\ m + k \leq N_0. \end{aligned} \tag{3.11}$$

After stopping sampling, the decision $d \in D$ with the Bayes risk as in (3.8) is chosen.

As an example, a Bayes sequential decision procedure with two interim analyses, that is, $N_0 = 3$, is described as follows.

the first interim analysis

After observing $X_1 = x_1$, we need to compare the Bayes risk from stopping sampling, $r_0(w^1, 1)$ with the expected risk from continuing sampling with not more than two observations $r_2(w^1, 1)$. If $r_0(w^1, 1) \leq r_2(w^1, 1)$, then we stop sampling and make a Bayes decision. Otherwise we continue sampling and observe the next observation X. The Bayes risk from stopping sampling $r_0(w^1, 1)$ is obtained by (3.4) with $m = 1$. The expected risk $r_2(w^1, 1)$ is calculated by using recursive relationships of (3.10) and (3.11), that is,

$$\begin{aligned} r_1(w^1, 1) &= \min\{r_0(w^1, 1), \ E^* r_0(w^1(X), 2)\}, \\ r_2(w^1, 1) &= \min\{r_0(w^1, 1), \ E^* r_1(w^1(X), 2)\}. \end{aligned}$$

the second interim analysis

After observing $X_1 = x_2$ and $X_2 = x_2$, the Bayes risk from stopping sampling is $r_0(w^2, 2)$, obtained by (3.4) and the risk from continuing sampling with not more than one observation is $r_1(w^2, 2)$. If $r_0(w^2, 2) \leq r_1(w^2, 2)$, then we stop sampling. Otherwise we continue sampling to observe the next observation X. The risk $r_1(w^2, 2)$ is obtained by (3.11) with $m = 2, k = 1$.

the final analysis

After observing $X_1 = x_2$, $X_2 = x_2$ and $X_3 = x_3$, we need to choose a decision with the Bayes risk $r_0(w^3, 3)$ that is calculated by (3.4) with m=3.

3.2.3 Group Sequential Decision Procedure

To use a group sequential decision procedure one looks at observations one group at a time instead of one observation at a time as in a classical sequential decision procedure. It is more practical to carry out a group sequential procedure than a classical sequential decision procedure in clinical trials because of the difficulty in continual monitoring, particularly in multicenter co-operative clinical trials with

survival time response.

It is known that there is no information loss of statistical inference on the parameters of interest if sufficient statistics(if they exist) of the parameters of interest are used. Therefore group sequential samples may be replaced by a sequence of sufficient statistics of the parameters of interest. Bayes sequential decision procedure can be applied to the sequence of sufficient statistics. Chapters 4, 5, and 6 will apply Bayes sequential decision theory into group sequential clinical trials comparing two treatments with normal, binomial and survival time response variables, respectively.

Chapter 4

Bayes Group Sequential Decision Clinical Trials on Normal Response

4.1 The Problem

Consider a clinical trial comparing an experimental treatment with the standard treatment. Assume that the main outcome variable of the clinical trial X is normally distributed with known variance σ^2 and unknown means μ_e and μ_s for the experimental and standard treatments, respectively. Let $\delta = \mu_e - \mu_s$ denote the treatment difference. The conventional hypotheses are

$H_0 : \delta \leq 0$(the experimental treatment is not better) vs $H_1 : \delta \geq \delta_0 > 0$

in which δ_0 is a break-even value of δ.

Suppose that the treatment is assigned by a randomised permuted block so that each consecutive group of $2n$ patients has n patients on each treatment, and the maximum number of groups is l. The group sample size n might be different in each group. This is discussed in Section 4.4. Let the group sequential sample be denoted by X_{ije}, X_{ijs}, $i = 1, 2, ..., n$, $j = 1, 2, ..., l$, for the experimental and

standard treatments respectively. Random variables X_{ije}, X_{ijs}, $i = 1, 2, ..., n$, $j = 1, 2, ..., l$, are independent and identically distributed as the normal distribution $N(\mu_e, \sigma^2)$ for the experimental treatment and the normal distribution $N(\mu_s, \sigma^2)$ for the standard treatment. The group sequential samples can be expressed as

$$\begin{array}{ll} X_{11e}, X_{21e}, ..., X_{n1e}; & X_{11s}, X_{21s}, ..., X_{n1s}; \\ X_{12e}, X_{22e}, ..., X_{n2e}; & X_{12s}, X_{22s}, ..., X_{n2s}; \\ ..... & ..... \\ X_{1le}, X_{2le}, ..., X_{nle}. & X_{1ls}, X_{2ls}, ..., X_{nls}. \end{array} \tag{4.1}$$

At each analysis j, $j = 1, 2, ..., l$, the Z_j is defined to be the difference of group sample means of the experimental and the standard treatments, which is

$$Z_j = \overline{X}_{je} - \overline{X}_{js}, \tag{4.2}$$

where,

$$\overline{X}_{je} = \frac{1}{n} \sum_{i=1}^{n} X_{ije},$$
$$\overline{X}_{js} = \frac{1}{n} \sum_{i=1}^{n} X_{ijs}.$$

The sequence of new random variables Z_1, Z_2,..., Z_l constitutes a classical sequential random sample, and they are independent and identically distributed as the normal distribution with mean equal to δ and variance equal to $\frac{2\sigma^2}{n}$.

Since Z_1, Z_2,..., Z_l are sufficient statistics of the treatment difference δ, they can be used instead of the group sequential sample of (4.1) in statistical inferences on the treatment difference δ without losing information. Then the Bayes sequential decision theory can be applied to the classical sequential random sample Z_j, $j = 1, 2, ..., l$.

4.2 Bayes Group Sequential Decision Procedure

4.2.1 Basic Elements of Bayes Sequential Decision Procedure of the Study

Corresponding to Section 3.1.1, the basic elements of the Bayes sequential decision procedure applied to the problem described in Section 4.1 are,

1) The parameter of interest is the treatment difference $\delta = \mu_e - \mu_s$. Assume that the prior distribution of δ is the normal distribution with mean equal to ν_0 and variance equal to τ_0^2, that is,

$$\delta \sim w(\delta) = N(\nu_0, \tau_0^2). \tag{4.3}$$

2) The decision d is chosen from the standard treatment and the experimental treatment, that is,

$$d \in \{standard\ treatment\ d_s,\ \ experimental\ treatment\ d_e\}. \tag{4.4}$$

3) the sequential sample

It has been shown in Section 4.1 that the classical sequential sample Z_1, Z_2,..., Z_l of (4.2) can be used instead of the original group sequential sample X_{ije}, X_{ijs}, $i = 1, 2, ..., n$, $j = 1, 2, ..., l$, of (4.1) without loss of information concerning statistical inferences on the treatment difference δ. The probability density function of Z_j, $j = 1, 2, ..., l$, is

$$Z_j \ \sim \ f(z|\delta) = N(\delta, \frac{2\sigma^2}{n}). \tag{4.5}$$

3) the loss and cost functions

To facilitate comparison with frequentist methods in group sequential clinical trials, and also for the simplicity of computation, the piecewise continuous

loss function is used. The loss functions $L(\delta, d_e)$ and $L(\delta, d_s)$ for choosing the experimental and standard treatments, respectively, are defined as,

$$\begin{aligned} L(\delta, d_s) &= \begin{cases} 0 & \delta \leq \delta_0 \\ K & \delta > \delta_0, \end{cases} \\ L(\delta, d_e) &= \begin{cases} K & \delta < 0 \\ 0 & \delta \geq 0, \end{cases} \end{aligned} \tag{4.6}$$

that is, if the treatment difference $\delta < 0$ – the experimental treatment is not better than the standard treatment, then there is no loss in choosing the standard treatment, but there is a cost K in choosing the experimental treatment; if the treatment difference $\delta > \delta_0$ – the experimental treatment is better than the standard treatment, then there is no loss in choosing the experimental treatment but there is a cost K in choosing the standard treatment. The loss function implies a "zone of indifference" or "range of equivalence" that extends from 0 to δ_0, in which benefis from the experimental treatment are balanced by increased toxicity, inconvenience or cost. If the true difference δ lies in this region, then there is no loss associated with accepting or rejecting the null hypothesis. The loss functions of (4.6) are displayed in Figure 4.1.

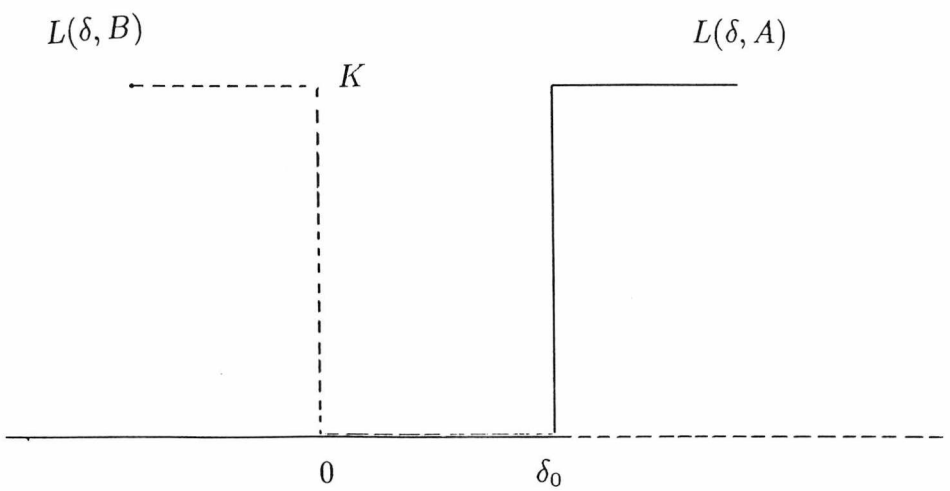


Figure 4.1 Loss Function $L(\delta, A)(—)$ and Loss Function $L(\delta, B)(--)$

Suppose the cost of enrolling a patient into the trial is 1 unit. This cost is constant through the trial. The cost function is then,

$$C_j(\delta, z_1, z_2, ..., z_j) = 2nj, \quad j = 1, 2, ..., l. \tag{4.7}$$

The total cost, for example, will be $L(\delta, d)+2n$ if the trial stops at the first interim analysis.

4.2.2 Posterior Distribution of δ

At each analysis j, $j = 1, ..., l$, after observing $Z_1 = z_1$, $Z_2 = z_2$,..., $Z_j = z_j$, the joint probability density function of Z_1, Z_2,..., and Z_j, from (4.5) is

$$\begin{aligned} f(z_1, z_2, ..., z_j|\delta) &= \prod_{i=1}^{j} f(z_i|\delta) \\ &= \left(\sqrt{\frac{n}{4\pi\sigma^2}}\right)^j e^{-\frac{n}{4\sigma^2}\sum_{i=1}^{j}(z_i-\delta)^2}. \end{aligned} \tag{4.8}$$

The posterior distribution of δ, $w^j = w(\delta|z_1, z_2, ..., z_j)$, is obtained by the Bayes theorem, that is,

$$\begin{aligned} w(\delta|z_1, z_2, ..., z_j) &= \frac{f(z_1, z_2, ..., z_j|\delta)\ w(\delta)}{\int_\Delta f(z_1, z_2, ..., z_j|\delta)\ w(\delta)\ d\delta} \\ &= \frac{\left(\sqrt{\frac{n}{4\pi\sigma^2}}\right)^j e^{-\frac{n}{4\sigma^2}\sum_{i=1}^{j}(z_i-\delta)^2}\ \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{(\delta-\nu_0)^2}{2\tau_0^2}}}{\int_\Delta \left(\sqrt{\frac{n}{4\pi\sigma^2}}\right)^j e^{-\frac{n}{4\sigma^2}\sum_{i=1}^{j}(z_i-\delta)^2}\ \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{(\delta-\nu_0)^2}{2\tau_0^2}} d\delta} \\ &= \frac{1}{\sqrt{2\pi}\tau_j} e^{-\frac{(\delta-\nu_j)^2}{2\tau_j^2}} \\ &= N(\nu_j, \tau_j^2), \end{aligned} \tag{4.9}$$

where,

$$\begin{aligned} \nu_j &= \frac{jn\tau_0^2\overline{z}_j + 2\nu_0\sigma^2}{jn\tau_0^2 + 2\sigma^2}, \\ \tau_j^2 &= \frac{2\tau_0^2\sigma^2}{jn\tau_0^2 + 2\sigma^2}, \\ \overline{z}_j &= \frac{1}{j}\sum_{i=1}^{j} z_i. \end{aligned} \tag{4.10}$$

4.2.3 Predictive Density Function

At each analysis j, $j = 1, 2, ..., l$, after observing $Z_1 = z_1$, $Z_2 = z_2$,..., $Z_j = z_j$, the predictive density function of Z given $Z_1 = z_1$, $Z_2 = z_2$,..., $Z_j = z_j$, $f(z|z_1, z_2, ..., z_j)$, defined by (3.6), is,

$$\begin{aligned} f(z|z_1, z_2..., z_j) &= \int_{-\infty}^{\infty} f(z|\delta)\, w(\delta|z_1, z_2, ..., z_j)\, d\delta \\ &= \int_{-\infty}^{\infty} \sqrt{\frac{n}{4\pi\sigma^2}} e^{-\frac{n}{4\sigma^2}(z-\delta)^2} \frac{1}{\sqrt{2\pi}\tau_j} e^{-\frac{(\delta-\nu_j)^2}{2\tau_j^2}}\, d\delta \\ &= \frac{1}{\sqrt{2\pi(\tau_j^2 + \frac{2\sigma^2}{n})}} e^{-\frac{1}{2(\tau_j^2+\frac{2\sigma^2}{n})}(\delta-\nu_j)^2} \\ &= N(\nu_j, \tau_j^2 + \frac{2\sigma^2}{n}), \end{aligned} \tag{4.11}$$

where ν_j, τ_j^2, and $\overline{z}_j$ are obtained by (4.10).

4.2.4 Sequential Decision Procedure

Suppose the study defined in Section 4.1 is designed to have two interim analyses. The Bayes sequential decision procedure of the study may have the first interim analysis, the second interim analysis and the final analysis. The Bayes group sequential decision procedure of the clinical trial is described as follows by backward induction(see Section 3.2.2).

The first interim analysis

At the first interim analysis, after $Z_1 = z_1$ has been observed, we need to get $r_0(w^1, 1)$ which is the Bayes risk from stopping the trial, and $r_2(w^1, 1)$ which is the risk from optimally continuing the trial with not more than two observations. Since the cost of enrolling patients, which is $C_1(\delta, z_1) = 2n$ at the first interim analysis, is independent of δ, and the risk $r_0(w^1, 1)$ is used to compare with the risk $r_2(w^1, 1)$ only, the cost $C_1(\delta, z_1) = 2n$ may be ignored in calculations of $r_0(w^1, 1)$ and $r_2(w^1, 1)$.

Using the formulae (3.4) with $m = 1$, we have,

$$\begin{aligned} r_0(w^1, 1) &= \min_{d \in \{d_s d_e\}} E_{\delta|z_1} L(\delta, d) \\ &= \min\{E_{\delta|z_1} L(\delta, d_s), \quad E_{\delta|z_1} L(\delta, d_e)\} \\ &= \min\{K(1 - \Phi(\frac{\delta_0 - \nu_1}{\tau_1})), \quad K\Phi(\frac{-\nu_1}{\tau_1})\}, \end{aligned} \tag{4.12}$$

where $\Phi(\cdot)$ is the standard normal cumulative probability distribution function and ν_1 and τ_1 are obtained by (4.10).

From the formula (3.11) with $m = 1$ and $k = 2$, we get,

$$r_2(w^1, 1) = \min\{r_0(w^1, 1), \quad E^* r_1(w^1(Z), 2) + 2n\}, \tag{4.13}$$

where $r_0(w^1, 1)$ is obtained by (4.12) and

$$E^* r_1(w^1(Z), 2) = \int_{-\infty}^{\infty} r_1(w^1(z), 2) * f(z|z_1) dz. \tag{4.14}$$

In (4.14), the predictive density $f(z|z_1)$ is the normal distribution $N(\nu_1, \tau_1^2 + \frac{2\sigma^2}{n})$ by (4.11), and the calculation of $r_1(w^1(z), 2)$ is obtained as follows.

By (3.11) and (4.12), first we have,

$$\begin{aligned} r_1(w^1,1) &= \min\{r_0(w^1,1),\ \ E^* r_0(w^1(Z'),2)+2n\} \\ &= \begin{cases} \min\{K(1-\Phi(\frac{\delta_0-\nu_1}{\tau_1})),\ \ E^* r_0(w^1(Z'),2)+2n\} & z_1 \le M_1 \\ \min\{K\Phi(\frac{-\nu_1}{\tau_1}),\ \ E^* r_0(w^1(Z'),2)+2n\} & z_1 > M_1, \end{cases} \end{aligned}$$

where

$$M_1 = \frac{\delta_0(n\tau_0^2+2\sigma^2)}{2n\tau_0^2} - \frac{2\nu_0\sigma^2}{n\tau_0^2}.$$

If $w^1 = w(\delta|z_1)$ is replaced by $w^1(z) = w(\delta|z_1,z)$, then we can get,

$r_1(w^1(z),2) =$

$$\begin{cases} \min\{K(1-\Phi(\frac{\delta_0-\nu_2(z_1,z)}{\tau_2})),\quad K\int_{-\infty}^{M'}(1-\Phi(\frac{\delta_0-\nu_3(z_1,z,z')}{\tau_3}))\frac{1}{\sqrt{2\pi(\tau_2^2+\frac{2\sigma^2}{n})}}e^{-\frac{(z'-\nu_2(z_1,z))^2}{2(\tau_2^2+\frac{2\sigma^2}{n})}}dz' \\ \qquad +K\int_{M'}^{\infty}\Phi(\frac{-\nu_3(z_1,z,z')}{\tau_3})\frac{1}{\sqrt{2\pi(\tau_2^2+\frac{2\sigma^2}{n})}}e^{-\frac{(z'-\nu_2(z_1,z))^2}{2(\tau_2^2+\frac{2\sigma^2}{n})}}dz'+2n\} & y \le M, \\ \min\{K\Phi(\frac{-\nu_2(z_1,z)}{\tau_2}),\quad K\int_{-\infty}^{M'}(1-\Phi(\frac{\delta_0-\nu_3(z_1,z,z')}{\tau_3}))\frac{1}{\sqrt{2\pi(\tau_2^2+\frac{2\sigma^2}{n})}}e^{-\frac{(z'-\nu_2(z_1,z))^2}{2(\tau_2^2+\frac{2\sigma^2}{n})}}dz' \\ \qquad +K\int_{M'}^{\infty}\Phi(\frac{-\nu_3(z_1,z,z')}{\tau_3})\frac{1}{\sqrt{2\pi(\tau_2^2+\frac{2\sigma^2}{n})}}e^{-\frac{(z'-\nu_2(z_1,z))^2}{2(\tau_2^2+\frac{2\sigma^2}{n})}}dz'+2n\} & y > M, \end{cases}$$

where,

$$\begin{aligned} \nu_1 &= \frac{n\tau_0^2 z_1 + 2\nu_0\sigma^2}{n\tau_0^2+2\sigma^2}, \\ \tau_1^2 &= \frac{2\tau_0^2\sigma^2}{n\tau_0^2+2\sigma^2}, \\ \nu_2(z_1,z) &= \frac{n\tau_0^2(z_1+z)+2\nu_0\sigma^2}{2n\tau_0^2+2\sigma^2}, \\ \tau_2^2 &= \frac{2\tau_0^2\sigma^2}{2n\tau_0^2+2\sigma^2}, \end{aligned}$$

$$\begin{aligned}
\nu_3(z_1, z, z') &= \frac{n\tau_0^2(z_1 + z + z') + 2\nu_0\sigma^2}{3n\tau_0^2 + 2\sigma^2}, \\
\tau_3^2 &= \frac{2\tau_0^2\sigma^2}{3n\tau_0^2 + 2\sigma^2}, \\
M' &= \frac{1}{n\tau_0^2}(-2\nu_0\sigma^2 - n\tau_0^2(z_1 + z) + \frac{1}{2}\delta_0(3n\tau_0^2 + 2\sigma^2)), \\
M &= \frac{1}{n\tau_0^2}(-2\nu_0\sigma^2 - n\tau_0^2 z_1 + \delta_0(n\tau_0^2 + \sigma^2)).
\end{aligned}$$

After $r_0(w^1, 1)$ and $r_2(w^1, 1)$ have been calculated, we compare these two risks. If $r_0(w^1, 1) \leq r_2(w^1, 1)$, then we stop the trial and choose a treatment which has the Bayes risk $r_0(w^1, 1)$. The statistical inferences on the treatment difference δ can be obtained based on the posterior distribution $\delta|z_1 \sim w^1 = w(\delta|z_1) = N(\nu_1, \tau_1^2)$. Otherwise we continue the trial to observe the next observation Z and need to go to the second interim analysis.

The second interim analysis

At the second interim analysis, after $Z_1 = z_1$ and $Z_2 = z_2$ have been observed, we need to calculate the Bayes risk from stopping the trial $r_0(w^2, 2)$, and $r_1(w^2, 2)$ which is the expected risk from observing not more than one observation. Similar to the first interim analysis, the risks $r_0(w^2, 2)$ and $r_1(w^2, 2)$ are calculated by (3.4) and the recursive relationships of (3.10) and (3.11), respectively. The Bayes risk from stopping the trial is,

$$\begin{aligned}
r_0(w^2, 2) &= \inf_{d \in \{d_s, d_e\}} E_{\delta|z_1,z_2} L(\delta, d) \\
&= \min\{E_{\delta|z_1,z_2} L(\delta, d_s), \quad E_{\delta|z_1,z_2} L(\delta, d_e)\}, \\
&= \min\{K(1 - \Phi(\frac{\delta_0 - \nu_2}{\tau_2})), \quad K\Phi(\frac{-\nu_2}{\tau_2})\}.
\end{aligned}$$

The expected risk is,

$$r_1(w^2, 2) = \min\{r_0(w^2, 2), \quad E^* r_0(w^2(Z), 3) + 2n\},$$

where ν_2 and τ_2 are obtained by (4.10), and

$$\begin{aligned} E^* r_0(w^2(Z),3) &= \int_{-\infty}^{\infty} r_0(w^2(Z=z),3) f(z|z_1,z_2)dz \\ &= K\int_{-\infty}^{M_2}(1-\Phi(\frac{\delta_0-\nu_3(z)}{\tau_3}))\frac{1}{\sqrt{2\pi(\tau_2^2+\frac{2\sigma^2}{n})}}e^{-\frac{(z-\nu_2)^2}{2(\tau_2^2+\frac{2\sigma^2}{n})}}dz \\ &\quad +K\int_{M_2}^{\infty}\Phi(\frac{-\nu_3(z)}{\tau_3})\frac{1}{\sqrt{2\pi(\tau_2^2+\frac{2\sigma^2}{n})}}e^{-\frac{(z-\nu_2)^2}{2(\tau_2^2+\frac{2\sigma^2}{n})}}dz, \end{aligned}$$

where

$$\begin{aligned} \nu_3(z) &= \frac{n\tau_0^2(z_1+z_2+z)+2\nu_0\sigma^2}{3n\tau_0^2+2\sigma^2}, \\ \tau_3^2 &= \frac{2\tau_0^2\sigma^2}{3n\tau_0^2+2\sigma^2}, \\ M_2 &= \frac{1}{n\tau_0^2}(-2\nu_0\sigma^2-n\tau_0^2(z_1+z_2)+\frac{1}{2}\delta_0(3n\tau_0^2+2\sigma^2)). \end{aligned}$$

If $r_0(w^2,2) \leq E^* r_0(w^2(Z),3)+2n$, then we should stop the trial to choose a decision with the Bayes risk $r_0(w^2,2)$. The statistical inferences on the treatment difference δ may be based on the posterior distribution $\delta \sim w^2 = w(\delta|z_1,z_2) = N(\nu_2,\tau_2^2)$. Otherwise we need to continue the trial to observe the next observation Z.

The Final Analysis

At the final analysis, after observing $Z_1 = z_1$, $Z_2 = z_2$ and $Z_3 = z_3$, we should stop sampling and choose the treatment with the Bayes risk $r_0(w^3,3)$, which is,

$$r_0(w^3,3) = \min\{E_{\delta|z_1,z_2,z_3}L(\delta,d_s),\ \ E_{\delta|z_1,z_2,z_3}L(\delta,d_e)\}.$$

The decision d may be chosen by,

$$d = \begin{cases} d_s & E_{\delta|z_1,z_2,z_3} L(\delta, d_s) \leq E_{\delta|z_1,z_2,z_3} L(\delta, d_e) \\ d_e & E_{\delta|z_1,z_2,z_3} L(\delta, d_s) \geq E_{\delta|z_1,z_2,z_3} L(\delta, d_e). \end{cases}$$

The statistical inferences on the treatment difference δ could be obtained from the posterior distribution $w(\delta|z_1, z_2, z_2) = N(\nu_3, \tau_3^2)$ as in (4.9). For example, the posterior expectation of δ, $E(\delta|z_1, z_2, z_3) = \frac{n\tau_0^2(z_1+z_2+z_3)+2\nu_0\sigma^2}{3n\tau_0^2+2\sigma^2}$, is a point estimator of the treatment difference δ.

4.2.5 Mean Sample Size

Suppose random variable J is the number of analyses. For two interim analyses, values of J may be 1, 2, or 3, and the expected number of analyses $E(J)$ is

$$\begin{aligned} E(J) &= 1 \times P(r_0(w^1, 1) \leq r_2(w^1, 1)) \\ &\quad +2 \times P(r_0(w^1, 1) > r_2(w^1, 1),\ r_0(w^2, 2) \leq E^* r_0(w^2(Z), 3) + 2n) \\ &\quad +3 \times P(r_0(w^1, 1) > r_2(w^1, 1),\ r_0(w^2, 2) > E^* r_0(w^2(Z), 3) + 2n) \end{aligned}$$

If the group of sample size of each treatment is n, then the mean sample size of the clinical trial will be $2n * E(J)$.

Although the group sequential clinical trial with three analyses was used in the study to describe the Bayes group sequential decision procedure, this procedure is similar when applied to a group sequential clinical trial with a greater number of analyses. Pocock(1982) has suggested that the number of analyses in group sequential clinical trials should not be more than five; otherwise there is little advantage in carrying out a group sequential clinical trial to reduce the sample size in frequentist methods. The group sequential clinical trial with three analyses will be used to compare Bayes sequential decision procedure with procedures of Pocock and O'Brien-Fleming in Section 4.3 by Monte Carlo simulations.

4.3 Monte Carlo Simulations

4.3.1 Introduction

Two types of Monte Carlo simulations were used to compare Bayes procedures with frequentist procedures on group sequential clinical trials. The first type of simulation compared type I and type II error rates and mean sample sizes of Bayes group sequential decision procedures with one-tailed frequentist group sequential procedures of the Pocock and the O'Brien-Fleming. The second type of simulation was the comparison of Bayesian characteristics – the mean cost C – which is equal to the mean sample size plus the loss K times error rate. Two interim analyses were used in simulations.

4.3.2 Simulations on Type I, Type II error rates and Mean Sample Size

We start by considering the first type of simulation. Type I error rates α were determined with the treatment difference δ equal to 0, that is, the sequential sample of (4.5) are from the normal distribution $N(0, \frac{2\sigma^2}{n})$ with mean equal to zero; The type I error rate α is the ratio of the number of decisions that there is a treatment difference to the total times of simulations. The mean sample size under these conditions is denoted by *Mean* N_α. Type II error rates β were determined with the treatment difference δ equal to 0.25σ, that is, the sequential sample of (4.5) are from the normal distribution $N(0.25\sigma, \frac{2\sigma^2}{n})$; The type II error rate β is the ratio of the number of decisions that there is no treatment difference to the total times of simulations. The corresponding mean sample size is *Mean* N_β.

We assumed that the prior distribution of the treatment difference δ was the normal distribution $N(0, \frac{\sigma^2}{n_0})$. The mean of the prior distribution is equal to 0 which means that there was no treatment difference to our prior knowledge. The

variance of the prior distribution is equal to $\frac{\sigma^2}{n_0}$ which might suggest that there were $2n_0$ "extra" pairs of patients in the pilot trial(Freedman and Spiegelhalter 1989). The number n_0 can be regarded as a measurement of the prior information and was changed (from 1% of maximum sample size to 50% of maximum sample size) in our simulations to study inferences from different prior information.

Group sample sizes n obtained by the Pocock's design and the O'Brien-Fleming's design with type I error rate $\alpha = 0.05$, powers $1 - \beta = 80\%$, and 90% were used in the simulations. The loss functions of (4.6), K values of 5,000 and 10,000, were chosen to yield type I error rates close to 0.05, type II error rates close to 0.20 or 0.10.

We ran 1000 simulations on Bayes procedures with various prior information. Results of simulations, which are the type I error rate α, the type II error rate β, the mean sample size under no treatment difference and the mean sample size under the treatment difference $= 0.25\sigma$, denoted by Mean N_α and Mean N_β respectively, are listed in Table 4.1, Table 4.2, Table 4.3, Table 4.4 and Table 4.5. Table 4.1 and Table 4.2 are the comparison of the Pocock procedures(powers 1-β = 80%, 90%) with Bayes procedures (loss functions $K = 5,000$, $K = 10,000$). Table 4.3 and Table 4.4 are the comparison of the O'Brien-Fleming procedures(powers 1-β = 80%, 90%) with Bayes procedures (loss functions $K = 5,000$, $K = 10,000$). They show that with some prior distribution of the treatment difference δ, there is a Bayes procedure(bold print in tables) with type I error rates α and type II error rates β similar to those of the Pocock and the O'Brien-Fleming procedures, but with smaller mean sample sizes (Mean N_α and Mean N_β) than those of the Pocock and the O'Brien-Fleming procedures. Bayes procedures stop the trial earlier under null hypothesis than the Pocock and the O'Brien-Fleming procedures since Bayes procedures have a much smaller mean sample size(Mean N_α). The more prior information we have concerning treatment difference (when n_0 is large), the earlier the trial stops under the null hypothesis(for Mean N_α is small), but the

lower power is(1-β is small). This is because the prior distribution of treatment difference δ has the mean equal to 0. Since we did 1,000 simulations in Bayes procedures, the type I error rate α and type II error rate β were random variables with binomial distributions. If we compare type I error rates α of Bayes procedures with the $\alpha = 0.05$ of frequentist procedures, the 95% confidence interval of α is (0.036, 0.064). The 95% confidence intervals for powers 1-β=80%, 90% are (0.775, 0.825), (0.881, 0.919), respectively. This is the reason that we are able to say that the bold print of type I error rate α and type II error rate β in Table 4.1, Table 4.2, Table 4.3 and Table 4.4 are similar to 0.05, 0.80(or 0.90), respectively.

Table 4.5 is the comparison between Bayes procedures with different loss functions(loss functions K=5,000, and 10,000) and the Pocock procedures with the type I error rate $\alpha = 0.05$ and the power 1-$\beta = 90\%$. Results on the comparison of Bayes procedures with the Pocock procedure are the same as those of Tables 4.1 and 4.2. By looking at different loss functions in Bayes procedures, Table 4.5 also shows that the bigger the cost of making a wrong decision(when $K = 10,000$), the larger the mean sample sizes needed in trials in order to make a decision with greater accuracy, that is, with smaller type I error rate and bigger power.

4.3.3 Simulations on Costs

The second type of simulation considered the comparison of Bayes characteristics – the mean cost C – which is equal to the mean sample size plus the loss K times error rate. Instead of giving values of treatment difference δ, the value of treatment difference δ used for each simulation was from the prior distribution $N(0, \frac{\sigma^2}{n_0})$ and the sequential samples $Z's$ were from the normal distribution $N(\delta, \frac{2\sigma^2}{n})$ with the mean equal to δ obtained from the prior distribution $N(0, \frac{\sigma^2}{n_0})$. The Table 4.6 shows the comparison of Bayes procedures of the loss K=5,000 with the Pocock procedure of type I error rate $\alpha = 0.05$ and type II error rate $\beta = 0.20$.

In Table 4.6, *Bayes procedure I* is the result of 10 simulations on treatment

differences δ and 500 simulations on sequential samples $Z's$ for each treatment difference δ, and *Bayes procedure II* is the result of 100 simulations on treatment differences δ and 20 times on sequential samples for each δ. It is very interesting to note that the mean cost(Mean C) in Bayes procedures is lower than that in the Pocock procedures. The same results are found in the comparison of Bayes procedures with the O'Brien-Fleming procedures.

4.3.4 Discussion

In the simulations, we used the break-even value of the treatment difference δ equal to 0.25σ, that is, $\delta_0 = 0.25\sigma$, and got a large sample size saving in Bayes procedures. If $\delta_0 = 0.50\sigma$, then the sample size saving in Bayes procedures would not be as big as in $\delta_0 = 0.25\sigma$. However, the reason for carrying out sequential clinical trials is to be able to detect small treatment differences as early as possible. Snapinn(1992) has also shown that monitoring clinical trials with a conditional probability stopping rule can achieve a large reduction in expected sample size without greatly affecting either the significance level or power of the trial. Another reason for the large sample size saving in Bayes procedures is that the Pocock and O'Brien-Fleming procedures stop a clinical trial when there is a treatment difference. But Bayes procedures stop a clinical trial either there is a treatment difference or there is no treatment difference when it is demonstrated by enough accumulating data.

Table 4.1 Frequentist Characteristics Comparison of Bayes($K = 5,000$) with the Pocock Procedures

Procedure	n	Max$N$	Prior n_0	α	Mean N_α	$1-\beta$	Mean N_β
Pocock	78	468		0.050	456	0.800	320
Bayes			2(1%)	0.105	239	0.874	247
			12(5%)	0.078	223	0.821	252
			23(10%)	**0.055**	**206**	**0.763**	**257**
			47(20%)	0.023	176	0.564	240
			117(50%)	0.001	157	0.060	160
Pocock	106	636		0.050	626	0.900	388
Bayes			3(1%)	0.094	279	0.883	291
			16(5%)	**0.067**	**258**	**0.822**	**293**
			31(10%)	0.044	241	0.742	285
			63(20%)	0.011	218	0.482	256
			157(50%)	0.001	212	0.030	212

Table 4.2 Frequentist Characteristics Comparison of Bayes($K = 10,000$) with the Pocock Procedures

Procedure	n	Max$N$	Prior n_0	α	Mean N_α	$1-\beta$	Mean N_β
Pocock	78	468		0.050	456	0.800	320
Bayes			2(1%)	0.080	273	0.898	286
			12(5%)	0.069	254	0.860	295
			23(10%)	**0.047**	**234**	**0.796**	**303**
			47(20%)	0.025	194	0.641	292
			117(50%)	0.000	157	0.114	181
Pocock	106	636		0.050	626	0.900	388
Bayes			3(1%)	0.064	323	0.918	338
			16(5%)	**0.049**	**296**	**0.875**	**346**
			31(10%)	0.031	269	0.801	353
			63(20%)	0.010	229	0.584	331
			157(50%)	0.000	213	0.037	216

Table 4.3 Frequentist Characteristics Comparison of Bayes($K = 5,000$) with the O'Brien-Fleming($O - F$) Procedures

Procedure	n	Max$N$	Prior n_0	α	Mean N_α	$1-\beta$	Mean N_β
$O - F$	68	408		0.050	404	0.800	332
Bayes			2(1%)	0.107	221	0.864	230
			10(5%)	0.083	209	0.818	235
			20(10%)	**0.061**	**192**	**0.758**	**239**
			41(20%)	0.028	163	0.588	228
			102(50%)	0.001	137	0.099	150
$O - F$	93	558		0.050	556	0.900	425
Bayes			3(1%)	0.096	260	0.881	273
			14(5%)	**0.072**	**245**	**0.830**	**274**
			28(10%)	0.046	224	0.743	273
			56(20%)	0.014	197	0.511	249
			140(50%)	0.001	186	0.038	186

Table 4.4 Frequentist Characteristics Comparison of Bayes($K = 10,000$) with the O'Brien-Fleming($O - F$) Procedures

Procedure	n	Max$N$	Prior n_0	α	Mean N_α	$1-\beta$	Mean N_β
$O-F$	68	408		0.050	404	0.800	332
Bayes			2(1%)	0.089	247	0.886	264
			10(5%)	0.072	236	0.850	272
			20(10%)	**0.059**	**219**	**0.793**	**277**
			41(20%)	0.029	182	0.653	274
			102(50%)	0.000	139	0.159	172
$O-F$	93	558		0.050	556	0.900	425
Bayes			3(1%)	0.071	300	0.913	316
			14(5%)	**0.056**	**277**	**0.870**	**325**
			28(10%)	0.034	253	0.794	330
			56(20%)	0.014	212	0.614	319
			140(50%)	0.000	187	0.055	195

Table 4.5 Comparison of Bayes Procedures with Different Loss Functions

Procedure	n	Max$N$	Prior n_0	α	Mean N_α	$1-\beta$	Mean N_β
Pocock	106	636		0.050	626	0.900	388
				$K = 5,000$			
Bayes			3(1%)	0.094	279	0.883	291
			16(5%)	0.067	258	0.822	293
			31(10%)	0.044	241	0.742	285
			63(20%)	0.011	218	0.482	256
			157(50%)	0.001	212	0.030	212
				$K = 10,000$			
Bayes			3(1%)	0.064	323	0.918	338
			16(5%)	0.049	296	0.875	346
			31(10%)	0.031	269	0.801	353
			63(20%)	0.010	229	0.584	331
			157(50%)	0.000	213	0.037	216

Table 4.6 Costs Comparison of Bayes($K = 5,000,\ n = 78$) with Pocock($n = 78,\ \alpha = 0.05,\ \beta = 0.2$) Procedures

Procedure	Prior n_0	Mean N	Error Rate	Mean C
Pocock		304	0.0196	402
Bayes procedure I	2(1%)	185	0.0126	248
Pocock		356	0.0060	386
Bayes procedure II	2(1%)	183	0.0070	218
Pocock		410	0.0086	453
Bayes procedure I	12(5%)	187	0.0176	274
Pocock		398	0.0180	480
Bayes procedure II	12(5%)	203	0.0215	310
Pocock		428	0.0228	542
Bayes procedure I	23(10%)	207	0.0304	359
Pocock		426	0.0050	452
Bayes procedure II	23(10%)	208	0.0090	253
Pocock		447	0.0048	471
Bayes procedure I	47(20%)	177	0.0022	187
Pocock		438	0.0105	490
Bayes procedure II	47(20%)	185	0.0105	237
Pocock		461	0.0178	550
Bayes procedure I	117(50%)	156	0.0004	158
Pocock		451	0.0105	504
Bayes procedure II	117(50%)	157	0.0100	207

4.4 Conclusion and Discussion

In the group sequential clinical trial comparing an experimental treatment with the standard treatment, where the main outcome measure X for the clinical trial is normally distributed and the mean of its normal distribution is the measure of treatment effect, there is a sequence of sufficient statistics of the treatment difference δ, Z_1, Z_2,...,Z_l, which constitutes a classical sequential random sample. The Z_j, $j = 1, 2, ..., l$, is normally distributed. The original group sequential random sample can be replaced by this sequence of efficient statistics without loss of information of statistical inferences on the treatment difference δ. The Bayes sequential decision theory is then applied to the classical sequential random sample Z_1, Z_2,...,Z_l. Monte Carlo simulations have shown that by choosing proper prior distributions and loss functions, there are Bayes group sequential decision procedures with type I error rate and type II error rate similar to those of the Pocock and the O'Brien-Fleming procedures, but with smaller expected sample sizes and costs than those of the Pocock and the O'Brien-Fleming procedures.

In the above study, we assumed that the clinical trial was monitored at every equal group sample size, which was every $2n$ patients. This is not the requirement of the Bayes sequential decision procedure in clinical trials. A clinical trial can be monitored at unequal group sample size. It is explained as follows.

Consider the clinical trial as in Section 4.1, here suppose that the group sample size is n_j, $j = 1, 2, ..., l$. Corresponding to (4.2), the classical sequential sample $Z_j = \overline{X}_{je} - \overline{X}_{js}$, where $\overline{X}_{je} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ije}$, $\overline{X}_{js} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ijs}$, $j = 1, 2, .., l$. The $\{Z_1, Z_2, ..., Z_j\}$ are the sequence of differences of group sample means and from the normal distributions $f(z|\delta) = N(\delta, \frac{2\sigma^2}{n_j})$, $j = 1, 2, ..., l$, respectively. The variances of this sequence of random variables $\{Z_j, j = 1, 2, ..., l\}$ are different, which is the only difference from the above clinical trial with equal group sample size. However, the variance σ^2 is known, at each analysis j, $j = 1, 2, ..., l$, the sequential sample Z_j may be standardised to have its variance equal to 1. The Bayes sequential

decision procedure of this clinical trial with unequal group sample size is then the same as the above procedure with equal group sample size. The Bayes sequential decision procedure can also be applied to the sequential sample $\{Z_j, j = 1, 2, ..., l\}$ itself like the above clinical trial, which is monitored at every $2n$ patients, because the corresponding posterior distribution $w(\delta|z_1, z_2, ..., z_j)$, $j = 1, 2, ..., l$, shown in (4.15) and predictive distribution $f(z|z_1, z_2, ..., z_j)$, $j = 1, 2, ..., l$, shown in (4.17) are still normal distributions.

At each analysis j, $j = 1, 2, ..., l$, after observing $Z_1 = z_1$, $Z_2 = z_2$,..., $Z_j = z_j$, the posterior distribution of the treatment difference δ is

$$\begin{aligned} w(\delta|z_1, z_2, ..., z_j) &= \frac{\prod_{i=1}^{j} f(z_i|\delta)\, w(\delta)}{\int_{-\infty}^{\infty} \prod_{i=1}^{j} f(z_i|\delta)\, w(\delta)\, d\delta} \\ &= \frac{\prod_{i=1}^{j} \sqrt{\frac{n_i}{4\pi\sigma^2}} e^{-\sum_{i=1}^{j} \frac{n_i}{4\sigma^2}(z_i-\delta)^2} \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{(\delta-\nu_0)^2}{2\tau_0^2}}}{\int_{-\infty}^{\infty} \prod_{i=1}^{j} \sqrt{\frac{n_i}{4\pi\sigma^2}} e^{-\sum_{i=1}^{j} \frac{n_i}{4\sigma^2}(z_i-\delta)^2} \frac{1}{\sqrt{2\pi}\tau_0} e^{-\frac{(\delta-\nu_0)^2}{2\tau_0^2}} d\delta} \\ &= \frac{1}{\sqrt{2\pi}\tau_j} e^{-\frac{(\delta-\nu_j)^2}{2\tau_j^2}} \\ &= N(\nu_j, \tau_j^2), \end{aligned} \tag{4.15}$$

where,

$$\begin{aligned} \nu_j &= \frac{\sum_{i=1}^{j} n_j \tau_0^2 \overline{z}_j + 2\nu_0\sigma^2}{\sum_{i=1}^{j} n_i\tau_0^2 + 2\sigma^2}, \\ \tau_j^2 &= \frac{2\tau_0^2\sigma^2}{\sum_{i=1}^{j} n_i\tau_0^2 + 2\sigma^2}, \\ \overline{z}_j &= \frac{\sum_{i=1}^{j} n_i z_i}{\sum_{i=1}^{j} n_i}. \end{aligned} \tag{4.16}$$

The predictive density function of z with the group sample size equal to n given

$Z_1 = z_1$, $Z_2 = z_2$,..., $Z_j = z_j$, $f(z|z_1, z_2, ..., z_j)$, is,

$$\begin{aligned} f(z|z_1, z_2..., z_j) &= \int_{-\infty}^{\infty} f(z|\delta)\ w(\delta|z_1, z_2, ..., z_j)\ d\delta \\ &= \int_{-\infty}^{\infty} \sqrt{\frac{n}{4\pi\sigma^2}} e^{-\frac{n}{4\sigma^2}(z-\delta)^2}\ \frac{1}{\sqrt{2\pi}\tau_j} e^{-\frac{(\delta-\nu_j)^2}{2\tau_j^2}}\ d\delta \\ &= \frac{1}{\sqrt{2\pi(\tau_j^2 + \frac{2\sigma^2}{n})}} e^{-\frac{1}{2(\tau_j^2+\frac{2\sigma^2}{n})}(\delta-\nu_j)^2} \\ &= N(\nu_j, \tau_j^2 + \frac{2\sigma^2}{n}), \end{aligned} \tag{4.17}$$

where ν_j, τ_j^2, and $\overline{z}_j$ are obtained by (4.16).

This detailed procedure of monitoring a clinical trial with unequal group sample size is discussed in Chapter 6, where the clinical trial with survival time response may be approximated by the clinical trial with normal distribution response with unequal variance in each analysis.

In the previous study, we also assumed that the main outcome variable in the clinical trial is from the normal distribution with variance σ^2 known. When the variance σ^2 is unknown, the computation required by the Bayes sequential decision procedure is complicated. This needs to be further studied.

Chapter 5

Bayes Group Sequential Decision Clinical Trials on Binary Response

This chapter will look at the Bayes sequential decision procedure of clinical trials with main outcome variables being binary response. If a random variable X is from the binomial distribution $B(n, p)$, where p is the parameter of interest, then the set of possible values of this random variable X are $\{0, 1, 2, ..., n\}$, which is a finite set. Consequently the computational difficulties of Bayesian inferences on the parameter p may be partly overcome. Therefore, instead of using the simple piecewise continuous loss function as in Chapter 4, the loss function which considers the treatment effect and patient horizon will be used in the Bayes sequential decision procedure of clinical trials with binary response.

5.1 The Problem

A clinical trial is designed to test a new drug response rate p. The conventional hypotheses are

$H_0 : p \leq p_0$(reject the new drug) vs $H_1 : p > p_0$ (accept the new drug),
in which p_0 is the break-even value of the response rate.

Suppose the clinical trial is monitored at every n observations and the total number of analyses is l. Let X_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ..., l$, be the group sequential random variables, where X_{ij} are independent identically distributed Bernoulli random variables with unknown parameter p. At each analysis j, $j = 1, 2, ..., l$, the group sequential sample, $X_{1j} = x_{1j}$, $X_{2j} = x_{2j}$, ..., $X_{nj} = x_{nj}$, are observed. Let Y_j be defined to be the sum of the group sample $X_{ij}, i = 1, 2, ..., n$. The sequence of random variables $Y_j, j = 1, 2, ..., l$, constitutes a classical sequential sample, and the distribution of Y_j is the binomial distribution $B(n, p)$, that is,

$$Y_j = \sum_{i=1}^{n} X_{ij} \sim B(n, p). \tag{5.1}$$

Since Y_j, $j = 1, 2, ..., l$, are sufficient statistics of the parameter p, the group sequential sample X_{ij}, $i = 1, 2, ..., n$, $j = 1, 2, ..., l$, can be replaced by the classical sequential sample Y_j, $j = 1, 2, ..., l$, without losing information on statistical inferences of the parameter p. The group sequential procedure is then replaced by the classical sequential procedure. Bayes sequential decision theory described in Chapter 3 can therefore be applied to the study. It is discussed as follows.

The basic elements of Bayes sequential decision procedure of the study are,

1) The parameter of interest is the new drug response rate p. A two-point prior distribution and beta prior distribution for p are studied in Section 5.2 and Section 5.3, respectively.

2) The decision space of the study is $D = \{d_0$(reject the new drug), d_1(accept the new drug)$\}$.

3) the loss and cost functions

At each analysis $j, j = 1, 2, ..., l$, the loss function is defined as,

$$\begin{aligned} L(p, d_0) &= \begin{cases} 0 & p \leq p_0 \\ K(p - p_0)t & p > p_0, \end{cases} \\ L(p, d_1) &= \begin{cases} K(p_0 - p)t & p \leq p_0 \\ K(p_0 - p)t & p > p_0, \end{cases} \end{aligned} \tag{5.2}$$

where p = the true underlying response rate of the new drug;

p_0 = the break-even response rate;

t = the patient horizon;

K = the difference in cost(monetary or ethical) of further treatment between a patient who does not respond to the new drug and a patient who does respond($K > 0$).

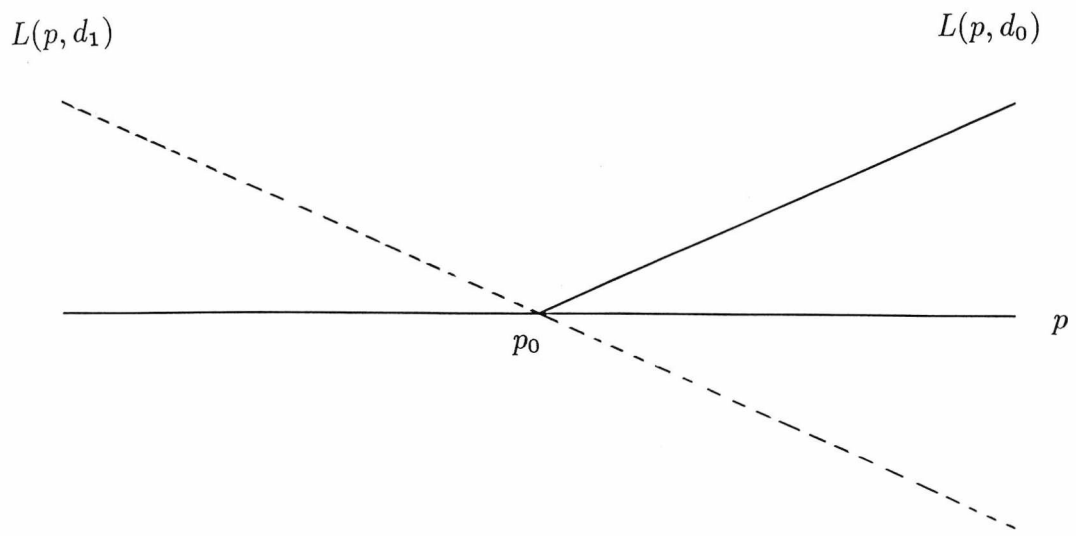


Figure 5.1 Loss Function $L(p, d_0)$(——) and Loss Function $L(p, d_1)$(−−)

The loss function is shown in figure 5.1. The patient horizon t and the treatment effect $p - p_0$ are considered in the loss function. If the response rate of new drug p is less than p_0, where the new drug should not be recommended, then there is no loss($L(p, d_0) = 0$) in rejecting the new drug, and there is a loss $L(p, d_1) = K(p_0 - p)t$ in accepting the new drug. If the response rate of new drug p is greater than p_0, where the new drug could be recommended, then there is a loss $L(p, d_0) = K(p - p_0)t$ in rejecting the new drug, and there is a gain $L(p, d_1) = K(p_0 - p)t$ (< 0) in accepting the new drug.

Suppose that the unit of $K(> 0)$ is the cost of enrolling a patient into the trial. The cost function at analysis j, $j = 1, 2, ..., l$, is then

$$C_j = jn. \tag{5.3}$$

4) The sequential sample Y_j, $j = 1, 2, ..., l$, are obtained by (5.1) in which $Y_j \sim f(y|p) = B(n, p)$.

5.2 Bayes Sequential Decision Procedure with Two-Point Prior Distribution

Consider the Bayes sequential decision procedure of the clinical trial described in Section 5.1. Suppose that the response rate of new drug p has the two-point prior distribution,

$$w(p) = \begin{cases} w_1 & p = p_1 \\ w_2 & p = p_2, \end{cases} \tag{5.4}$$

where $w_1 + w_2 = 1$, $0 < p_1 < p_0 < p_2 < 1$. In practical, this prior distribution would be suggested only when we have very strong prior information to show this

form of prior.

At each analysis j, $j = 1, 2, ..., l$, the posterior distribution of p after observing $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$, denoted by $w^j = w(p|y_1, y_2, ..., y_j)$, is

$$
\begin{aligned}
& w(p|y_1, y_2, ..., y_j) \\
&= \frac{f(y_1, y_2, ..., y_j|p)\ w(p)}{f(y_1, y_2, ..., y_j|p_1)\ w(p_1) + f(y_1, y_2, ..., y_j|p_2)\ w(p_2)} \\
&= \frac{\prod_{i=1}^{j} f(y_i|p)\ w(p)}{\prod_{i=1}^{j} f(y_i|p_1)\ w(p_1) + \prod_{i=1}^{j} f(y_i|p_2)\ w(p_2)} \\
&= \frac{p^{\sum_{i=1}^{j} y_i}(1-p)^{jn-\sum_{i=1}^{j} y_i}\ w(p)}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1 + p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2} \\
&= \begin{cases} \dfrac{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1 + p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2} & p = p_1 \\ \dfrac{p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1 + p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2} & p = p_2. \end{cases}
\end{aligned} \tag{5.5}
$$

At each analysis j, $j = 1, 2, ..., l$, the predictive density function of Y after observing $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$, expressed by $f(y|y_1, y_2, ..., y_j)$, is

$$
\begin{aligned}
& f(y|y_1, y_2, ..., y_j) \\
&= E_{p|y_1, y_2, ..., y_j} f(y|p) \\
&= f(y|p_1) w(p_1|y_1, y_2, ..., y_j) + f(y|p_2) w(p_2|y_1, y_2, ..., y_j) \\
&= \frac{\binom{n}{y} p_1^{\sum_{i=1}^{j} y_i + y}(1-p_1)^{(j+1)n-(\sum_{i=1}^{j} y_i + y)} w_1 + \binom{n}{y} p_2^{\sum_{i=1}^{j} y_i + y}(1-p_2)^{(j+1)n-(\sum_{i=1}^{j} y_i + y)} w_2}{p_1^{\sum_{i=1}^{j} y_i}(1-p_1)^{jn-\sum_{i=1}^{j} y_i} w_1 + p_2^{\sum_{i=1}^{j} y_i}(1-p_2)^{jn-\sum_{i=1}^{j} y_i} w_2}.
\end{aligned} \tag{5.6}
$$

5.2.1 Bayes Sequential Decision Procedure

The Bayes sequential decision procedure of a clinical trial compares the Bayes risk from stopping the clinical trial with the risk from the optimal continuation

of the clinical trial after each observation. During any analysis, if the Bayes risk from stopping the trial is less than the risk from the optimally continuing the trial, then the clinical trial is stopped; otherwise the trial is continued. After the clinical trial is terminated, a decision with the Bayes risk is chosen. These are described in Chapter 3. However, the stopping and decision rules of the Bayes sequential decision procedure of the clinical trial defined in Section 5.1 with the two-point prior distribution as in (5.4) can be described in the form of test statistics as can those of frequentist methods. This is explained as follows by these Bayes sequential decision procedures with one interim analysis and two interim analyses.

The Stopping Rule

One Interim Analysis

In the interim analysis, after the value $Y_1 = y_1$ has been observed, ignoring the constant cost function of (5.3), the Bayes risk from stopping the trial $r_0(w^1, 1)$ is,

$$
\begin{aligned}
& r_0(w^1, 1) \\
& = \min_{d\in\{d_0,d_1\}} E_{p|y_1} L(p, d) \\
& = \min\{E_{p|y_1} L(p, d_0),\ \ E_{p|y_1} L(p, d_1)\} \\
& = \begin{cases} K(p_2 - p_0)t\, w(p_2|y_1) & y_1 \le \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^n]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} \\ K(p_0 - p_1)t\, w(p_1|y_1) + K(p_0 - p_2)t\, w(p_2|y_1) & y_1 > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^n]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]}. \end{cases}
\end{aligned}
\tag{5.7}
$$

On the other hand, if the next sample Y needs to be observed, the expected risk from observing Y, denoted by $E^* r_0(w^1(Y), 2)$, is

$$
E^* r_0(w^1(Y), 2) \ = \ \sum_{y=0}^{n} r_0(w^1(Y = y), 2)\, f(y|Y_1 = y_1),
$$

where

$$
\begin{aligned}
& r_0(w^1(Y=y),2) \\
&= \min_{d\in\{d_0,d_1\}} E_{p|y_1,y} L(p,d) \\
&= \min\{E_{p|y_1,y} L(p,d_0),\ \ E_{p|y_1,y} L(p,d_1)\} \\
&= \begin{cases} K(p_2-p_0)t\, w(p_2|y_1,y) & y \le \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} - y_1 \\ K(p_0-p_1)t\, w(p_1|y_1,y) + K(p_0-p_2)t\, w(p_2|y_1,y) & y > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} - y_1, \end{cases}
\end{aligned}
$$

and the predictive density function $f(y|y_1)$ is obtained by (5.6). Let M be the integer portion of value $\max\{\frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} - y_1,\ 0\}$. It is obtained that

$$
\begin{aligned}
E^* r_0(w^1(Y),2) &= \textstyle\sum_{y=0}^{M} K(p_2-p_0)t\, w(p_2|y_1,y)\, f(y|y_1) \\
&\quad + \sum_{y=M+1}^{n} [K(p_0-p_1)t\ w(p_1|y_1,y) + K(p_0-p_2)t\, w(p_2|y_1,y)]\, f(y|y_1).
\end{aligned} \tag{5.8}
$$

Let

$$
D(y_1|n,K,t,w) \ = \ r_0(w^1,1) - (E^* r_0(w^1(Y),2) + n). \tag{5.9}
$$

If $D(y_1|n,K,t,w) \le 0$, then the clinical trial might be stopped; otherwise the clinical trial needs to be continued. Assume that the clinical trial is terminated at the interim analysis. The value of y_1 should be small if the drug is not effective; and the value of y_1 would not be close to 0 if the response rate of the drug is high. The structure of stopping region is then,

$$
\begin{aligned}
&\{y_1 : D(y_1|n,K,t,w) \le 0\} = \\
&\qquad \{y_1 : y_1 \le c_1(n,K,t,w)\} \cup \{y_1 : y_1 \ge c_2(n,K,t,w)\},
\end{aligned} \tag{5.10}
$$

where c_1 and $c_2 (0 \le c_1 \le c_2 \le n)$ might be the roots of the equation

$$
D(y_1|n,K,t,w) = 0. \tag{5.11}
$$

Although the form of (5.10) is obtained under the assumption of the clinical trial being terminated at the interim analysis, it can be also used in the situation when the clinical trial needs to be continued by some special values of c_1 and c_2. The stopping rule at the interim analysis is then described as below.

At the interim analysis, if there are no roots in $[0, n]$ of the equation (5.11), then the clinical trial is either continued when $D(y_1|n, K, t, w) > 0$ or stopped when $D(y_1|n, K, t, w) < 0$, no matter what value y_1 has been observed. If there are roots of the equation (5.11), then

1) when $c_1 = c_2 = n$ or $c_1 = c_2 = 0$, that is, $\{y_1 : D(y_1|n, K, t, w) \leq 0\} = \{0, 1, 2, ..., n\}$, the clinical trial is stopped at the interim analysis no matter what value of the first observation Y_1 is observed;

2) when $c_1 < 0$ and $c_2 > n$, while $\{y_1 : D(y_1|n, K, t, w) \leq 0\}$ is an empty set, the clinical trial needs to be continued;

3) when $c_1 < 0$ and $c_2 \leq n$, the trial could be stopped and the new drug accepted if $y_1 \in \{y_1 : y_1 \geq c_2(n, K, t, w)\}$, otherwise the trial needs to be continued;

4) when $c_2 > n$ and $c_1 \geq 0$, the trial could be stopped and the new drug rejected if $y_1 \in \{y_1 : y_1 \leq c_1(n, K, t, w)\}$, otherwise the trial needs to be continued;

5) when $0 \leq c_1 \leq c_2 \leq n$, the trial could be stopped with the decision of rejecting the new drug if $y_1 \in \{y_1 : y_1 \leq c_1(n, K, t, w)\}$ or accepting the new drug if $y_1 \in \{y_1 : y_1 > c_2(n, K, t, w)\}$, otherwise the trial is continued to the final analysis.

Two Interim Analyses Procedure

Say at the first interim analysis, $Y_1 = y_1$ has been observed. The next observation should be taken if, and only if, the risk from stopping the trial, denoted by $r_0(w^1, 1)$, is greater than the risk of continuing the trial with not more than two observations, denoted by $r_2(w^1, 1)$. The risk $r_0(w^1, 1)$ is obtained by (5.7). Using

the method of backward induction described in Chapter 3, the risk $r_2(w^1, 1)$ is,

$$r_2(w^1, 1) = \min\{r_0(w^1, 1),\ E^* r_1(w^1(Y), 2) + n\}, \tag{5.12}$$

where,

$$E^* r_1(w^1(Y), 2) = \sum_{y=0}^{n} r_1(w^1(Y = y), 2)\, f(y|y_1). \tag{5.13}$$

In (5.13) the predictive density function $f(y|y_1)$ is obtained by (5.6), and the calculation of $r_1(w^1(Y = y), 2)$ is described as below.

From (3.11) it is obtained,

$$\begin{aligned} r_1(w^1, 1) &= \min\{r_0(w^1, 1),\ E^* r_0(w^1(Y'), 2) + n\} \\ &= \begin{cases} \min\{E_{p|y_1} L(p, d_0),\ E^* r_0(w^1(Y'), 2) + n\} & y_1 \le \frac{\log[\frac{p_0 - p_1}{2(p_2 - p_0)} \frac{w_1}{w_2} (\frac{1-p_1}{1-p_2})^n]}{\log[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}]} \\ \min\{E_{p|y_1} L(p, d_1),\ E^* r_0(w^1(Y'), 2) + n\} & y_1 > \frac{\log[\frac{p_0 - p_1}{2(p_2 - p_0)} \frac{w_1}{w_2} (\frac{1-p_1}{1-p_2})^n]}{\log[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}]}, \end{cases} \end{aligned}$$

where,

$$\begin{aligned} E^* r_0(w^1(Y'), 2) &= \sum_{y'=0}^{n} r_0(w^1(Y' = y'), 2)\, f(Y' = y'|y_1). \\ &= \sum_{y'=0}^{M} K(p_2 - p_0) t\, w(p_2|y_1, y')\, f(y'|y_1) \\ &+ \sum_{y'=M+1}^{n} [K(p_0 - p_1) t\, w(p_1|y_1, y') + K(p_0 - p_2) t\, w(p_2|y_1, y')]\, f(y'|y_1), \end{aligned}$$

in which the M is the integer portion of value $\max\{\frac{\log[\frac{p_0 - p_1}{2(p_2 - p_0)} \frac{w_1}{w_2} (\frac{1-p_1}{1-p_2})^{2n}]}{\log[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}]} - y_1,\ 0\}$.
If $w^1 = w(p|y_1)$ is replaced by $w^1(y) = w(p|y_1, y)$, then $r_1(w^1(Y = y), 2)$ in (5.13)

is obtained, that is,

$$
\begin{aligned}
& r_1(w^1(Y=y),2) \\
& = \min\{r_0(w^1(Y=y)),\ E^* r_0(w^1(Y')|Y=y)) + n\} \\
& = \begin{cases} \min\{E_{p|y_1,y} L(p,d_0),\ E^* r_0(w^1(Y')|Y=y)) + n\} & y \le \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} - y_1 \\ \min\{E_{p|y_1,y} L(p,d_1),\ E^* r_0(w^1(Y')|Y=y)) + n\} & y > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{2n}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} - y_1, \end{cases}
\end{aligned}
$$

where,

$$
\begin{aligned}
E^* r_0(w^1(Y')|Y=y)) = & \sum_{y'=0}^{M_1} K(p_2-p_0)t\, w(p_2|y_1,y,y')\, f(y'|y_1,y) \\
& + \sum_{y'=M_1+1}^{n} \{K(p_0-p_1)t\, w(p_1|y_1,y,y') + Kt(p_0-p_2)\, w(p_2|y_1,y,y')\} f(y'|y_1,y),
\end{aligned}
$$

in which the M_1 is the integer portion of $\max\{\frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{3n}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2},]} - (y_1+y),\ 0\}$.

Let

$$
D_1(y_1|n,K,t,w) \;=\; r_0(w^1,1) - r_2(w^1,1). \tag{5.14}
$$

If $D_1(y_1|n,K,t,w) \le 0$, then the clinical trial could be stopped; otherwise the trial needs to be continued. As described in the procedure of one interim analysis, the stopping region with the form of y_1 has the structure,

$$
\begin{aligned}
& \{y_1 : D_1(y_1|n,K,t,w) \le 0\} = \\
& \qquad \{y_1 : y_1 \le c_1^1(n,K,t,w)\} \cup \{y_1 : y_1 \ge c_2^1(n,K,t,w)\},
\end{aligned} \tag{5.15}
$$

where $c_1^1(n,K,t,w)$ and $c_2^1(n,K,t,w)$ $(c_1^1 \le c_2^1)$ might be the roots of equation $D_1(y_1|n,K,t,w) = 0$.

If the equation $D_1(y_1|n,K,t,w) = 0$ has no roots, the clinical trial could be either stopped when $D_1(y_1|n,K,t,w) < 0$, or continued when $D_1(y_1|n,K,t,w) > 0$ no matter what observed value $Y_1 = y_1$ is. If there are roots of the equation

$D_1(y_1|n, K, t, w) = 0$, then clinical trial could be stopped when $0 \leq y_1 \leq c_1^1 \leq n$ or $n \geq y_1 \geq c_2^1 \geq 0$; otherwise the trial needs to be continued.

At the second interim analysis, after $Y_1 = y_1$ and $Y_2 = y_2$ have been observed, there is only one more sample which can be observed. The calculation of risks at the second interim analysis is therefore same as those at the interim analysis of one interim analysis procedure except that the posterior distribution of p, $w(p|y_1)$, is replaced by $w(p|y_1, y_2)$. The Bayes risk from stopping the trial is,

$$\begin{aligned} r_0(w^2, 2) &= \min_{\{d_0, d_1\}} E_{p|y_1,y_2} L(p, d) \\ &= \min\{E_{p|y_1,y_2} L(p, d_0), \quad E_{p|y_1,y_2} L(p, d_1)\}, \end{aligned} \tag{5.16}$$

where,

$$\begin{aligned} E_{p|y_1,y_2} L(p, d_0) &= K(p_2 - p_0)t\, w(p_2|y_1, y_2), \\ E_{p|y_1,y_2} L(p, d_1) &= K(p_0 - p_1)t\, w(p_1|y_1, y_2) + K(p_0 - p_2)t\, w(p_2|y_1, y_2). \end{aligned}$$

The expected risk from observing the last observation is,

$$E^* r_0(w^2(Y), 3) = \sum_{y=0}^{n} r_0(w^2(Y = y), 3)\, f(y|y_1, y_2), \tag{5.17}$$

where,

$$\begin{aligned} r_0(w^2(Y = y), 3) &= \min_{d \in \{d_0, d_1\}} E_{p|y_1,y_2,y} L(p, d) \\ &= \min\{E_{p|y_1,y_2,y} L(p, d_0), \quad E_{p|y_1,y_2,y} L(p, d_1)\}. \end{aligned}$$

It is easy to show that the $r_0(w^2, 2)$ and $E^* r_0(w^2(Y), 3) + n$ are functions of $y_1 + y_2$ given the group sample size n, parameters of the loss function K, t and the prior

distribution $w(p)$. Let

$$D_2(y_1 + y_2|n, K, t, w) = r_0(w^2, 2) - (E^* r_0(w^2(Y), 3) + n). \tag{5.18}$$

If $D_2(y_1 + y_2|n, K, t, w) \leq 0$, then the trial should be stopped; otherwise the trial needs to be continued. The stopping region with the form of $y_1 + y_2$ is

$$\begin{aligned} &\{y_1 + y_2 : D_2(y_1 + y_2|n, K, t, w) \leq 0\} = \\ &\quad \{y_1 + y_2 : y_1 + y_2 \leq c_1^2(n, K, t, w)\} \cup \{y_1 + y_2 : y_1 + y_2 > c_2^2(n, K, t, w)\}. \end{aligned} \tag{5.19}$$

where $c_1^2(n, K, t, w)$, $c_2^2(n, K, t, w)$ can be the roots of the equation $D_2(y_1+y_2|n, K, t, w) = 0$.

Hence the stopping rule of the clinical trial with two interim analyses in the form of test statistics is that at each interim analysis $j = 1$ (or $j = 2$), after $Y_1 = y_1$ (or $Y_1 = y_1$, $Y_2 = y_2$) being observed, if the equation $D_1(y_1|n, K, t, w) = 0$ (or $D_2(y_1 + y_2|n, K, t, w)$) have no roots, then the clinical trial should be stopped when $D_1(y_1|n, K, t, w) < 0$ (or $D_2(y_1 + y_2|n, K, t, w) < 0$), and the clinical trial is continued when $D_1(y_1|n, K, t, w) > 0$ (or $D_2(y_1 + y_2|n, K, t, w) > 0$); if the equation $D_1(y_1|n, K, t, w) = 0$ (or $D_2(y_1 + y_2|n, K, t, w) = 0$) have the roots, then the decision of stopping the trial early is based on the values of the roots listed in the following table.

The Stopping Rule of Clinical Trials Two Interim Analyses

first interim analysis	$D_1(y_1\|n,K,t,w)=0$	
	no roots in $[0,n]$	roots c_1^1, c_2^1
stop the trial	$D_1(y_1\|n,K,t,w)<0$	$0\le y_1\le c_1^1\le n$
		$0\le c_2^1\le y_1\le n$
continue the trial	$D_1(y_1\|n,K,t,w)>0$	$c_1^1<0$ and $c_2^1>n$
second interim analysis	$D_2(y_1+y_2\|n,K,t,w)=0$	
	no roots in $[0,2n]$	roots c_2^1, c_2^2
stop the trial	$D_2(y_1+y_2\|n,K,t,w)<0$	$0\le y_1+y_2\le c_1^2\le n$
		$0\le c_2^2\le y_1+y_2\le n$
continue the trial	$D_2(y_1+y_1\|n,K,t,w)>0$	$c_1^2<0$ and $c_2^2>n$

The above results can be easily generalized to the Bayes sequential decision procedure of the clinical trial with more than two interim analyses. Consider the Bayes sequential decision procedure of the clinical trial described in Section 5.1. At each analysis j, $j=1,2,...,l$, after $Y_1=y_1$, $Y_2=y_2$, ..., $Y_j=y_j$ have been observed, the Bayes risk from stopping the trial is denoted by $r_0(w^j,j)$ and the expected risk of continuing the trial with not more than $l-j$ observations is denoted by

$r_{l-j}(w^j, j)$. Let

$$D_j(\sum_{i=1}^{j} y_i|n, K, t, w) = r_0(w^j, j) - r_{l-j}(w^j, j), \tag{5.20}$$

If $D_j(\sum_{i=1}^{j} y_i|n, K, t, w) \leq 0$, then the trial could be stopped; otherwise the trial needs to be continued. The stopping region with the form of statistics $\sum_{i=1}^{j} Y_i$ is

$$\begin{aligned} &\{\sum_{i=1}^{j} y_i : D_j(\sum_{i=1}^{j} y_j|n, K, t, w) \leq 0\} = \\ &\quad \{\sum_{i=1}^{j} y_i : \sum_{i=1}^{j} y_i \leq c_1^j(n, K, t, w)\} \cup \{\sum_{i=1}^{j} y_i : \sum_{i=1}^{j} y_i \geq c_2^j(n, K, t, w\} \end{aligned} \tag{5.21}$$

where $c_1^j(n, K, t, w)$ and $c_2^j(n, K, t, w)$ $(c_1^j \leq c_2^j)$ may be the roots of the equation $D_j(\sum_{i=1}^{j} y_i|n, K, t, w) = 0$. The stopping rule at analysis j can be summarised as the following table.

Stopping rule at analysis j

jth interim analysis	$D_j(\sum_{i=1}^{j} y_i\|n, K, t, w) = 0$	
	no roots in $[0, jn]$	roots c_1^j, c_2^j
stop the trial	$D_j(\sum_{i=1}^{j} y_i\|n, K, t, w) < 0$	$0 \leq \sum_{i=1}^{j} y_i \leq c_1^j \leq nj$
		$0 \leq c_2^j \leq \sum_{i=1}^{j} y_i \leq nj$
continue the trial	$D_j(\sum_{i=1}^{j} y_i\|n, K, t, w) > 0$	$c_1^j < 0$ and $c_2^j > nj$

The Decision Rule

At each analysis j, $j = 1, 2, ..., l$, after values $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$ have been observed, if the clinical trial is terminated, the Bayes risk from stopping the trial, $r_0(w^j, j)$, is,

$$\begin{aligned} r_0(w^j, j) &= \min_{d\in\{d_0,d_1\}} E_{p|y_1,y_2,...,y_j} L(p,d) \\ &= \min\{E_{p|y_1,y_2,...,y_j} L(p,d_0),\ \ E_{p|y_1,y_2,...,y_j} L(p,d_1)\} \\ &= \begin{cases} E_{p|y_1,y_2,...,y_j} L(p,d_0) & \sum_{i=1}^{j} y_i \le \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} \\ E_{p|y_1,y_2,...,y_j} L(p,d_1) & \sum_{i=1}^{j} y_i > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]}, \end{cases} \end{aligned} \tag{5.22}$$

where

$$\begin{aligned} E_{p|y_1,y_2,...,y_j} L(p,d_0) &= Kt(p_2-p_0)\, w(p_2|y_1,y_2,...,y_j), \\ E_{p|y_1,y_2,...,y_j} L(p,d_1) &= Kt(p_0-p_1)\, w(p_1|y_1,y_2,...,y_j) + Kt(p_0-p_2)\, w(p_2|y_1,y_2,...,y_j). \end{aligned}$$

The Bayes risk $r_0(w^j, j)$ is a function of $\sum_{i=1}^{j} y_i$. The decision rule after the trial be terminated is then

$$d = \begin{cases} d_0(reject\ the\ new\ drug) & \sum_{i=1}^{j} y_i \le \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]} \\ d_1(accept\ the\ new\ drug) & \sum_{i=1}^{j} y_i > \frac{\log[\frac{p_0-p_1}{2(p_2-p_0)}\frac{w_1}{w_2}(\frac{1-p_1}{1-p_2})^{jn}]}{\log[\frac{p_2}{p_1}\frac{1-p_1}{1-p_2}]}. \end{cases} \tag{5.23}$$

5.2.2 Group Sample Size

Consider the group sequential clinical trial with the number of analyses equal to l. At the final analysis, let $f(y_1, y_2, ..., y_l)$ be the joint probability density function

of Y_1, Y_2,..., Y_l, that is,

$$\begin{aligned} f(y_1, y_2, ..., y_l) &= f(y_1, y_2, ..., y_l|p_1)\, w(p_1) + f(y_1, y_2, ..., y_l|p_2)\, w(p_2) \\ &= \prod_{i=1}^{l} \binom{n}{y_i} p_1^{\sum_{i=1}^{l} y_i} (1-p_1)^{nl-\sum_{i=1}^{l} y_i}\, w_1 \\ &+ \prod_{i=1}^{l} \binom{n}{y_i} p_2^{\sum_{i=1}^{l} y_i} (1-p_2)^{nl-\sum_{i=1}^{l} y_i}\, w_2. \end{aligned} \tag{5.24}$$

Since the Bayes risk $r_0(w^l, l)$ as in (5.22) with $j = l$ is a function of $\sum_{i=1}^{l} y_i$ where the constant cost function C_l as in (5.3) was ignored, the average Bayes risk, $E_{(Y_1,Y_2,...,Y_l)} r_0(w^l, l) + nl$, is,

$$\begin{aligned} E_{(Y_1,Y_2,...,Y_l)} r_0(w^l, l) + nl &= \sum_{y_1,y_2,...,y_l} r_0(w^l, l)\, f(y_1, y_2, ..., y_l) + nl \\ &= \sum_{\sum y_i} r_0(w^l, l) f(\sum y_I) + nl \\ &= \sum_{\sum y_i=0}^{nl} r_0(w^l, l) \left[\binom{nl}{\sum y_i} p_1^{\sum y_i} (1-p_1)^{nl-\sum y_i}\, w_1 \right. \\ &\quad \left. + \binom{nl}{\sum y_i} p_2^{\sum y_i} (1-p_2)^{nl-\sum y_i}\, w_2 \right] + nl \\ &= \sum_{\sum y_i=0}^{M_0} E_{p|y_1,y_2,...,y_l} L(p, d_0) \left[\binom{nl}{\sum y_i} p_1^{\sum y_i} (1-p_1)^{nl-\sum y_i}\, w_1 \right. \\ &\quad \left. + \binom{nl}{\sum y_i} p_2^{\sum y_i} (1-p_2)^{nl-\sum y_i}\, w_2 \right] + \\ &\quad \sum_{\sum y_i=M_0+1}^{nl} E_{p|y_1,y_2,...,y_l} L(p, d_1) \left[\binom{nl}{\sum y_i} p_1^{\sum y_i} (1-p_1)^{nl-\sum y_i}\, w_1 \right. \\ &\quad \left. + \binom{nl}{\sum y_i} p_2^{\sum y_i} (1-p_2)^{nl-\sum y_i}\, w_2 \right] + nl \end{aligned} \tag{5.25}$$

where the M_0 is the integer portion of $\frac{\log[\frac{p_0-p_1}{2(p_2-p_0)} \frac{w_1}{w_2} (\frac{1-p_1}{1-p_2})^{nl}]}{\log[\frac{p_2}{p_1} \frac{1-p_1}{1-p_2}]}$.

It can been seen from above that given the prior distribution $w(p)$, the loss function $L(p, d)$, and the cost function $C_j, j = 1, 2, ..., l$, the average Bayes risk

$E_{(Y_1,Y_2,...,Y_l)}r_0(w^l, l)$ is a function of nl, which is the maximum sample size of the group sequential clinical trial. Bayes inferences are independent of the number of analyses after the trial is terminated. Therefore in the view of Bayes sequential decision theory the maximum sample size $N = nl$ can be designed to reach the prespecified average Bayes risk level, say R, and

$$N = \min\{N : E_{(Y_1,Y_2,...,Y_l)}r_0(w^l, l) \leq R\}. \tag{5.26}$$

If the clinical trial is designed to have equal group size, then the group sample size is $n = \frac{N}{l}$.

5.2.3 Comparison with Frequentist Methods

In this section, the procedures of Pocock and O'Brien-Fleming are used for comparisons with the Bayes sequential decision procedure in the group sequential clinical trial described in Section 5.2.1. The comparisons are based on type I error rates α, type II error rates β, and their corresponding expected sample sizes, *Mean* N_α, *Mean* N_β, respectively,

Since the response rate of new drug p is assumed to have the two-point prior distribution as in (5.4), the corresponding conventional hypotheses test is H_0 : $p = p_1 < p_0$(reject the new drug), vs H_1 : $p = p_2 > p_0$(accept the new drug). Hence, the type I error rate α in the study is the probability of accepting the drug if the drug is not effective, that is, the probability of choosing decision d_1 if the sequential samples Y_j, $j = 1, 2, ..., l$, are from the binomial distribution $B(n, p_1)$. The type II error rate β in the study is the probability of rejecting the drug while the drug is effective, that is, the probability of choosing decision d_0 if the sequential samples Y_j, $j = 1, 2, ..., l$, are from the binomial distribution $B(n, p_2)$. Since the number of all possible values of Y_j, $j = 1, 2, ..., l$, is $n + 1$, the type I error rate α and type II error rate β can be calculated.

Assume the clinical trial described in Section 5.1 is studied under the Bayes sequential decision theory. Suppose the break-even value of response rate of the new drug $p_0 = 0.20$, the new drug is rejected if the response rate of the drug $p = p_1 = 0.01$, 0.05, and 0.10, and the new drug is accepted if the response rate of the drug $p = p_2 = 0.25$. Let the two-point prior distribution of p as in (5.4) have $w_1 = 0.3$, 0.5, 0.7, and 0.9. The group sample sizes n are designed to be the same as those used in the procedures of Pocock and O'Brien-Fleming with type I error rate $\alpha = 0.05$ and type II error rate $\beta = 0.20$ and 0.10. Assume that the patient horizon $t = 500$, 1000, 5000, 10000. The two interim analyses procedure is used as an example. The Table 5.1 and Table 5.2 are the results of the type I error rate α and its expected sample size *Mean* N_α, and the type II error rate β and its expected sample size *Mean* N_β, respectively, with the group sample size designed by the procedure of Pocock with type I error rate $\alpha = 0.05$ and type II error rate $\beta = 0.2$.

It has been shown that there are Bayes sequential decision procedures which have the type I error rates α, type II error rates β close to the procedures of Pocock, but their corresponding expected sample sizes are smaller than those of Pocock's procedures, which are in bold print in the tables. The more prior belief in $p = p_1$ of the new drug being ineffective, that is, the larger value of w_1, then the lower type I error rate α and the smaller expected sample size Mean N_α but the higher type II error rate β and the bigger expected sample size Mean N_β. The less prior belief in $p = p_1$ or the more prior belief in $p = p_2$, that is, the larger the value of $w_2 = 1 - w_1$, then the smaller type II error rate β is. The larger the patient horizon t is, the larger the expected sample size is needed to be able to make a decision with greater accuracy(that is, both type I and type II error rates are small). The bigger the difference of response rate(that is, the value of p_1 is smaller because the value of p_2 is fixed to be 0.25.) is, the earlier the trial is stopped since the expected sample size is smaller. The same results are found

when the type II error rate $\beta = 0.1$.

The Table 5.1' and Table 5.2' are the results of the type I error rate α and its expected sample size *Mean* N_α, and the type II error rate β and its expected sample size *Mean* N_β, respectively, with the group sample size designed by the procedure of O'Brien-Fleming with type I error rate $\alpha = 0.05$ and type II error rate $\beta = 0.2$. The results of Bayesian decision procedures comparing with the procedure of O'Brien-Fleming are same as those of Bayesian decision procedures comparing with the procedure of Pocock.

Hence, it is obtained that Bayesian sequential decision procedures in clinical trials could be based on the statistics $\sum_{i=1}^{j} Y_i$, $j = 1, 2, ..., l$, as those of frequentist methods. There are Bayesian sequential decision procedures with type I and type II error rates similar to those of the Pocock and the O'Brien-Fleming procedures, but with smaller mean sample size than those of the Pocock and the O'Brien-Fleming procedures.

Table 5.1 Type I error rate α and expected sample size Mean N_α

	t=500		t=1000		t=5000		t=10000	
w_1	α	Mean N_α	α	Mean N_α	α	Mean N_α	α	Mean N_α
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 17.5								
Bayes sequential decision procedure:								
0.3	**0.0082**	**14.7**	**0.0082**	**14.7**	0.0085	28.8	0.0024	29.0
0.5	**0.0082**	**14.7**	**0.0082**	**14.7**	**0.0007**	**14.9**	0.0008	26.4
0.7	**0.0072**	**13**	**0.0082**	**14.7**	**0.0007**	**14.9**	**0.0007**	**14.9**
0.9	**0.0004**	**13.1**	**0.0004**	**13.1**	**0.0007**	**14.9**	**0.0007**	**14.9**
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 28								
Bayes sequential decision procedure:								
0.3	0.0849	21	0.0898	25.4	0.0277	37.2	0.0149	37.8
0.5	0.0849	21	**0.0205**	**22.5**	**0.0080**	**27.8**	0.0066	37.9
0.7	**0.0189**	**21**	**0.0205**	**22.5**	**0.0061**	**27.8**	0.0038	27.9
0.9	0.0032	21	0.0036	21.4	0.0011	23.0	0.0012	23.4
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 55.5								
Bayes sequential decision procedure:								
0.3	0.1332	43	0.1333	43	0.0728	60.6	0.0373	65.0
0.5	0.0607	43	0.0607	43	**0.0318**	**55.0**	0.0162	64.2
0.7	0.0607	43	0.0607	43	**0.0124**	**49.4**	0.0067	56.7
0.9	0.0087	43	0.0087	43	**0.0044**	**45.9**	0.0024	49.4

Table 5.2 Type II error rate β and expected sample size $MeanN_\beta$

	t=500		t=1000		t=5000		t=10000	
w_1	β	Mean N_β	β	Mean N_β	β	Mean N_β	β	Mean N_β
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 20.5								
Bayes sequential decision procedure:								
0.3	**0.0275**	**14.5**	**0.0275**	**14.5**	0.0042	14.9	0.0048	17.7
0.5	**0.0275**	**14.5**	**0.0275**	**14.5**	**0.0292**	**17.8**	0.0092	18.2
0.7	**0.1267**	**13**	**0.0275**	**14.5**	**0.0292**	**17.8**	**0.0292**	**17.8**
0.9	**0.1321**	**16.0**	**0.1321**	**16.0**	**0.0292**	**17.8**	**0.0292**	**17.8**
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 33.5								
Bayes sequential decision procedure:								
0.3	0.0745	21	0.0236	22.3	0.0080	26.0	0.0085	30.0
0.5	0.0745	21	**0.0843**	**23.7**	**0.0313**	**29.3**	0.0151	31.1
0.7	**0.1917**	**21**	**0.0843**	**23.7**	**0.0323**	**30.4**	0.0329	34.9
0.9	0.3674	21	0.2063	25.1	0.0976	34.6	0.0913	34.8
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 66								
Bayes sequential decision procedure:								
0.3	0.0612	43	0.0612	43	0.0161	48.6	0.0147	53.8
0.5	0.1237	43	0.1237	43	**0.0339**	**52.6**	0.0217	60.5
0.7	0.1237	43	0.1237	43	**0.0740**	**57.1**	0.0420	67.6
0.9	0.3390	43	0.3390	43	**0.1411**	**61.0**	0.0916	73.4

Table 5.1' Type I error rate α and expected sample size N_α

	t=500		t=1000		t=5000		t=10000	
w_1	α	Mean N_α	α	Mean N_α	α	Mean N_α	α	Mean N_α
O'Brien-Fleming procedure: $p_1 = 0.01$, Max N = 30, Mean N = 20								
Bayes sequential decision procedure:								
0.3	0.0956	10	**0.0134**	**11.7**	0.0066	21.73	0.0066	21.7
0.5	**0.0055**	**11.0**	**0.0055**	**11.0**	0.0066	21.73	0.0032	21.8
0.7	**0.0055**	**11.0**	**0.0055**	**11.0**	**0.0024**	**11.90**	0.0032	21.8
0.9	0.0043	10	**0.0010**	**11.1**	**0.0003**	**11.08**	**0.0003**	**11.1**
O'Brien-Fleming procedure: $p_1 = 0.05$, Max N = 48, Mean N = 32								
Bayes sequential decision procedure:								
0.3	0.0582	19.7	0.0594	20.1	**0.0200**	**30.26**	0.0205	37.7
0.5	0.0503	17	0.0542	20.2	**0.0188**	**27.96**	**0.0128**	**30.6**
0.7	0.0503	17	**0.0112**	**17.8**	**0.0041**	**21.59**	**0.0045**	**28.3**
0.9	0.0088	17	0.0025	17.9	**0.0038**	**20.85**	**0.0030**	**21.6**
O'Brien-Fleming procedure: $p_1 = 0.10$, Max N = 94, Mean N = 62								
Bayes sequential decision procedure:								
0.3	0.2115	32	0.2115	32	0.0593	55.38	0.0605	58.7
0.5	0.0944	32	0.0944	32	0.0503	45.62	**0.0352**	**57.0**
0.7	0.0359	32	0.0358	32	**0.0222**	**40.24**	**0.0148**	**47.4**
0.9	0.0117	32	0.0117	32	**0.0062**	**35.34**	**0.0057**	**40.0**

Table 5.2' Type II error rate β and expected sample size N_β

	t=500		t=1000		t=5000		t=10000	
w_1	β	Mean N_β	β	Mean N_β	β	Mean N_β	β	Mean N_β
O'Brien-Fleming procedure: $p_1 = 0.01$, Max N = 30, Mean N = 24.5								
Bayes sequential decision procedure:								
0.3	0.0563	10	**0.0589**	**12.0**	0.0112	13.2	0.0112	13.2
0.5	**0.0689**	**12.2**	**0.0689**	**12.2**	0.0112	13.2	0.0121	16.1
0.7	**0.0689**	**12.2**	**0.0689**	**12.2**	**0.0618**	**15.3**	0.0121	16.1
0.9	0.2440	10	**0.0794**	**15.2**	**0.0853**	**16.3**	**0.0853**	**16.3**
O'Brien-Fleming procedure: $p_1 = 0.05$, Max N = 48, Mean N = 39								
Bayes sequential decision procedure:								
0.3	0.0687	18.9	0.0579	19.2	**0.0218**	**24.4**	0.0164	24.7
0.5	0.1637	17	0.0590	19.5	**0.0290**	**24.3**	**0.0233**	**30.0**
0.7	0.1637	17	**0.1767**	**20.6**	**0.0781**	**28.6**	**0.0473**	**29.8**
0.9	0.3530	17	0.2088	25.0	**0.0927**	**28.3**	**0.0797**	**34.2**
O'Brien-Fleming procedure: $p_1 = 0.10$, Max N = 94, Mean N = 76								
Bayes sequential decision procedure:								
0.3	0.0698	32	0.0698	32	0.0243	42.7	0.0210	42.9
0.5	0.1530	32	0.1530	32	0.0465	42.7	**0.0269**	**49.6**
0.7	0.2779	32	0.2779	32	**0.0911**	**46.3**	**0.0597**	**56.2**
0.9	0.4325	32	0.4325	32	**0.1976**	**51.0**	**0.1201**	**61.1**

5.3 Bayes Sequential Decision Procedure with Beta Prior Distribution

Consider the study described as in Section 5.1. Assume that the prior distribution of the response rate p in the study has the beta distribution with parameters $u(u > 0)$ and $v(v > 0)$. The probability density function of p is,

$$\begin{aligned} p \sim w(p) &= f(p|u,v) \\ &= \frac{1}{\beta(u,v)} p^{u-1}(1-p)^{v-1}. \end{aligned} \tag{5.27}$$

At each analysis j, $j = 1, 2, ..., l$, the posterior distribution of p after observing $Y_1 = y_1$, $Y_2 = y_2$, ..., $Y_j = y_j$, denoted by $w^j = w(p|y_1, y_2, ..., y_j)$, is still a beta distribution with parameters $\sum_{i=1}^{j} y_i + u$ and $jn - \sum_{i=1}^{j} y_i + v$. It is shown as follows by Bayes theorem.

$$\begin{aligned} w(p|y_1, y_2, ..., y_j) &= \frac{f(y_1, y_2, ..., y_j|p)\, w(p)}{\int_0^1 f(y_1, y_2, ..., y_j|p)\, w(p)\, dp} \\ &= \frac{\prod_{i=1}^{j} f(y_i|p)\, w(p)}{\int_0^1 \prod_{i=1}^{j} f(y_i|p)\, w(p)\, dp} \\ &= \frac{\prod_{i=1}^{j} \binom{n}{y_i} p^{\sum_{i=1}^{j} y_i} (1-p)^{jn - \sum_{i=1}^{j} y_i} \frac{1}{\beta(u,v)} p^{u-1}(1-p)^{v-1}}{\int_0^1 \prod_{i=1}^{j} \binom{n}{y_i} p^{\sum_{i=1}^{j} y_i} (1-p)^{jn - \sum_{i=1}^{j} y_i} \frac{1}{\beta(u,v)} p^{u-1}(1-p)^{v-1} dp} \\ &= \frac{p^{\sum_{i=1}^{j} y_i + u - 1} (1-p)^{jn - \sum_{i=1}^{j} y_i + v - 1}}{\beta(\sum_{i=1}^{j} y_i + u,\ jn - \sum_{i=1}^{j} y_i + v)}. \end{aligned} \tag{5.28}$$

At each analysis j, $j = 1, 2, ..., l$, the predictive density function of Y after observing $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$, expressed by $f(y|y_1, y_2, ..., y_j)$, is

$$\begin{aligned}
& f(y|y_1, y_2, ..., y_j) \\
&= E_{p|y_1,y_2,...,y_j} f(y|p) \\
&= \int_0^1 f(y|p)\, w(p|y_1, y_2, ..., y_j)\, dp \\
&= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} \frac{p^{\sum_{i=1}^{j} y_i + u - 1} (1-p)^{jn - \sum_{i=1}^{j} y_i + v - 1}}{\beta(\sum_{i=1}^{j} y_i + u,\ jn - \sum_{i=1}^{j} y_i + v)}\, dp \\
&= \binom{n}{y} \frac{\beta(\sum_{i=1}^{j} y_i + y + u,\ (j+1)n - \sum_{i=1}^{j} y_i - y + v)}{\beta(\sum_{i=1}^{j} y_i + u,\ jn - \sum_{i=1}^{j} y_i + v)}. \qquad (5.29)
\end{aligned}$$

5.3.1 Bayes Sequential Decision Procedure

In this section, Bayes sequential decision procedures of the study are described by the one interim analysis and the two interim analyses.

The Stopping Rule

One Interim Analysis

At the interim analysis, suppose that the value $Y_1 = y_1$ has been observed, the Bayes risk from stopping the trial, $r_0(w^1, 1)$, is,

$$\begin{aligned}
r_0(w^1, 1) &= \min_{d \in \{d_0, d_1\}} E_{p|y_1} L(p, d) \\
&= \min\{E_{p|y_1} L(p, d_0),\ E_{p|y_1} L(p, d_1)\}, \qquad (5.30)
\end{aligned}$$

where

$$\begin{aligned}
E_{p|y_1} L(p, d_0) &= \int_0^1 L(p, d_0)\, w(p|y_1)\, dp \\
&= \int_{p_0}^1 K(p - p_0) t \frac{p^{y_1+u-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u,\ n-y_1+v)}\, dp \\
&= Kt \left(\frac{y_1+u}{n+u+v} \int_{p_0}^1 \frac{p^{y_1+u+1-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u+1,\ n-y_1+v)}\, dp \right. \\
&\qquad \left. -p_0 \int_{p_0}^1 \frac{p^{y_1+u-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u,\ n-y_1+v)}\, dp \right), \\
E_{p|y_1} L(p, d_1) &= \int_0^1 L(p, d_1)\, w(p|y_1)\, dp \\
&= \int_0^1 K(p_0 - p) t \frac{p^{y_1+u-1}(1-p)^{n-y_1+v-1}}{\beta(y_1+u,\ n-y_1+v)}\, dp \\
&= Ktp_0 - Kt \frac{y_1+u}{n+u+v}.
\end{aligned}$$

There is no closed form for the Bayes risk $r_0(w^1, 1)$ as in (5.7) for the study with the two-point prior distribution.

The expected risk from observing the next observation Y, denoted by $E^* r_0(w^1(Y), 2)$, is,

$$\begin{aligned}
E^* r_0(w^1(Y), 2) &= \sum_{y=0}^{n} r_0(w^1(Y = y), 2)\, f(y|y_1) \\
&= \sum_{y=0}^{n} \left(\min_{d \in \{d_0, d_1\}} E_{p|y_1, y} L(p, d) \right) f(y|y_1) \\
&= \sum_{y=0}^{n} \left(\min_{d \in \{d_0, d_1\}} \int_0^1 L(p, d)\, w(p|y_1, y)\, dp \right) f(y|y_1), \quad (5.31)
\end{aligned}$$

where the posterior density function $w(p|y_1, y)$ and the predictive density function $f(y_1|y)$ are obtained by (5.28) and (5.29), respectively.

If the Bayes risk from stopping the trial $r_0(w^1, 1)$ is less than $E^* r_0(w^1(Y), 2) + n$, the clinical trial could be stopped; otherwise the trial needs to be continued. Although the stopping rule of the Bayes sequential decision procedure in the study can be described as a form of statistics Y_1 as in Section 5.2, it is not easy to get

roots from the following equation (5.32) and it is unnecessary for this discussion.

$$D(y_1|n,K,t,w) \;=\; r_0(w^1,1) - (E^*r_0(w^1(Y),2)+n) = 0. \qquad (5.32)$$

Two Interim Analyses

At the first interim analysis, say $Y_1 = y_1$ has been observed. The Bayes risk from stopping the trial $r_0(w^1,1)$ is obtained by (5.30). Using the backward induction, the risk from continuing the trial with not more than two observations $r_2(w^1,1)$ is,

$$r_2(w^1,1) \;=\; \min\{r_0(w^1,1),\;\; E^*r_1(w^1(Y),2)+n\}, \qquad (5.33)$$

where,

$$E^*r_1(w^1(Y),2) \;=\; \sum_{y=0}^{n} r_1(w^1(Y=y),2)f(y|y_1). \qquad (5.34)$$

In (5.34) the predictive density function $f(y|y_1)$ is obtained by (5.29), and the risk $r_1(w^1(Y=y),2)$ may be calculated as the risk $r_1(w^1,1)$ by replacing the posterior distribution $w^1 = w(p|y_1)$ with the posterior distribution $w^1(Y) = w(p|y_1,y)$. It is described as follows.

From (3.11) it is obtained,

$$r_1(w^1,1) \;=\; \min\{r_0(w^1,1),\;\; E^*r_0(w^1(Y'),2)+n\},$$

where,

$$E^*r_0(w^1(Y'),2) \;=\; \sum_{y=0}^{n} r_0(w^1(Y'=y'),2)f(Y'=y'|y_1).$$

Hence,

$$r_1(w^1(Y=y),2) \;=\; \min\{r_0(w^1(Y=y)),\; E^* r_0(w^1(Y')|Y=y)) + n\},$$

where,

$$r_0(w^1(Y=y)) = \min\{E_{p|y_1,y}L(p,d_0),\; E_{p|y_1,y}L(p,d_1)\},$$
$$E^* r_0(w^1(Y')|Y=y)) = \sum_{y'=0}^{n} \min\{E_{p|y_1,y,y'}L(p,d_0),\; E_{p|y_1,y,y'}L(p,d_1)\}f(y'|y_1,y)$$

If the Bayes risk from stopping the trial $r_0(w^1,1)$ is less than the expected risk of continuing the trial with not more than two observations $r_2(w^1,1)$, then the clinical trial could be stopped; Otherwise it needs to be continued to the second interim analysis.

At the second interim analysis, after $Y_1 = y_1$ and $Y_2 = y_2$ have been observed, there is only one more sample which can be observed. The risk from stopping the trial is,

$$\begin{aligned} r_0(w^2,2) &= \min_{\{d_0,d_1\}} E_{p|y_1,y_2}L(p,d) \\ &= \min\{E_{p|y_1,y_2}L(p,d_0),\;\; E_{p|y_1,y_2}L(p,d_1)\}. \end{aligned} \tag{5.35}$$

The expected risk from observing the last observation is,

$$E^* r_0(w^2(Y),3) \;=\; \sum_{y=0}^{n} r_0(w^2(Y=y),3)f(y|y_1,y_2), \tag{5.36}$$

where,

$$\begin{aligned} r_0(w^2(Y=y),3) &= \min_{d\in\{d_0,d_1\}} E_{p|y_1,y_2,y}L(p,d) \\ &= \min\{E_{p|y_1,y_2,y}L(p,d_0),\;\; E_{p|y_1,y_2,y}L(p,d_1)\}. \end{aligned}$$

If the Bayes risk from stopping the trial $r_0(w^2,2)$ is less than the expected risk of

continuing the trial $E^* r_0(w^2(Y), 3) + n$, then the clinical trial could be stopped; Otherwise the trial needs to be continued.

The Decision Rule

At each analysis j, $j = 1, 2, ..., l$, in the study, after values $Y_1 = y_1$, $Y_2 = y_2$,..., $Y_j = y_j$ have been observed, if the trial is terminated, then the Bayes risk from stopping the trial $r_0(w^j, j)$ is,

$$\begin{aligned} r_0(w^j, j) &= \min_{d \in \{d_0, d_1\}} E_{p|y_1, y_2, ..., y_j} L(p, d) \\ &= \min\{E_{p|y_1, y_2, ..., y_j} L(p, d_0), \ E_{p|y_1, y_2, ..., y_j} L(p, d_1)\}. \end{aligned} \tag{5.37}$$

Since the $E_{p|y_1, y_2, ..., y_j} L(p, d_0)$ and the $E_{p|y_1, y_2, ..., y_j} L(p, d_1)$ are functions of $\sum_{i=1}^{j} y_i$, and the drug is not accepted if the response rate is not high enough, that is, the $\sum_{i=1}^{j} y_i$ is small. The decision rule after the trial being terminated has the form,

$$d = \begin{cases} d_0(reject\ the\ new\ drug) & \sum_{i=1}^{j} y_i \leq k_1 \\ d_1(accept\ the\ new\ drug) & \sum_{i=1}^{j} y_i \geq k_2 \end{cases} \tag{5.38}$$

where $0 \leq k_1 \leq k_2 \leq jn$.

5.3.2 Prior Information

In the study the prior distribution of the response rate p of the new drug is assumed to be a beta distribution as in (5.27), that is,

$$p \sim w(p) = \frac{1}{\beta(u, v)} p^{u-1} (1-p)^{v-1}.$$

The expectation and variance of the prior distribution of p are,

$$\begin{aligned} E(p) &= \frac{u}{u+v}. \\ Var(p) &= \frac{uv}{(u+v)^2(u+v+1)} \\ &= \left(\frac{u}{u+v}\right)\left(\frac{v}{u+v}\right)\left(\frac{1}{u+v+1}\right) \\ &= \left(\frac{u}{u+v}\right)\left(1-\frac{u}{u+v}\right)\left(\frac{1}{u+v+1}\right) \\ &= E(p)(1-E(p))\left(\frac{1}{u+v+1}\right). \end{aligned}$$

These expressions show that the variance $Var(p)$ is a function of $u+v$ if the expectation $E(p)$ is fixed. Therefore the value of $u+v$ may be regarded as a measure of prior information.

If $E(p) < p_0$, then there is prior belief that the new drug is not effective. The larger value of $u+v$ (or the more prior information), the more likely that the new drug would be rejected, and hence the smaller type I error rate α would be. When $E(p) > p_0$, it is assumed that the new drug is effective by the prior information. The larger value of $u+v$, that is, the more prior information, the more chance that the new drug be accepted, and hence the smaller type II error rate β would be. These results can be shown by the following example.

Consider a one interim analysis with group sample size $n = 20$ and the break-even value of response rate $p_0 = 0.20$. The type I error rate α and its corresponding mean sample size $Mean\ N_\alpha$ are obtained under the assumption that the sequential sample Y is from the binomial distribution $B(n, p_1)$, where $p_1 = 0.05$. The type II error rate β and its corresponding mean sample size $Mean\ N_\beta$ are obtained under the assumption that the sequential sample Y is from the binomial distribution $B(n, p_2)$, where $p_2 = 0.25$. The parameters u and v of the beta prior distribution of (5.27) are selected to have the $E(p) = 0.01$ and 0.5. As an example, assume that the value of u =0.2, 0.25, 0.33, 0.5, 1, 2, 3, and 4. The corresponding values

of v are obtained by $E(p) = 0.01$ and 0.5. The type I error rate α and its expected sample size *Mean* N_α, and the type II error rate β and its expected sample size *Mean* N_β are listed in Table 5.3.

The Table 5.3 shows that when $E(p) < p_0$, then the more prior information, such as the value of $u + v$ from 2 to 40, the smaller the type I error rate α, which is 0.0159 to < 0.0001, but the bigger type II error rate β, which is from 0.2252 to 0.8982. The type I error rate α and type II error rate β are quite stable when the value of $u + v$ is from 2 to 2.5, 3.3 to 10, and 20 to 40. When $E(p) > p_0$, then the more prior information, while the value of $u + v$ is from 2 to 40, the smaller the type II error rate β is, but the type I error rate α is increased. The type I error rate α and type II error rate β are relatively stable when the value of $u + v$ is from 0.4 to 2.

Table 5.3 The Beta Prior Information

u	v	$E(p)$	$Var(p)$	α	Mean N_α	β	Mean N_β
			$E(p) < p_0$				
0.2	1.8	0.01	0.03	0.0159	20	0.2252	20
0.25	2.25	0.01	0.0257	0.0159	20	0.2252	20
0.33	3	0.01	0.0208	0.0026	20.3	0.3422	23.8
0.5	4.5	0.01	0.015	0.0026	20.3	0.3422	23.8
1	9	0.01	0.0082	0.0026	20	0.4148	20
2	18	0.01	0.0043	0.0003	20	0.6172	20
3	27	0.01	0.0029	< 0.0001	20	0.7858	20
4	36	0.01	0.0022	< 0.0001	20	0.8982	20
			$E(p) > p_0$				
0.2	0.2	0.5	0.1786	0.0159	20	0.2252	20
0.25	0.25	0.5	0.1667	0.0159	20	0.2252	20
0.33	0.33	0.5	0.15	0.0159	20	0.2252	20
0.5	0.5	0.5	0.125	0.0159	21.2	0.1739	22.7
1	1	0.5	0.0833	0.0161	21.2	0.1468	22.7
2	2	0.5	0.05	0.0755	20	0.0913	20
3	3	0.5	0.0357	0.2642	20	0.0243	20
4	4	0.5	0.0278	0.2926	27.5	0.0051	20.4

5.3.3 Comparison with the Two-point Prior Distribution

The comparisons are based on type I, type II error rates, their corresponding expected sample sizes, and average Bayes risks. The values of parameters u, v of the beta prior distribution are selected to have the same expectation and variance as those of the two-point prior distributions in Table 5.1 and Table 5.2. The loss and cost functions are also same as those of the two-point prior distributions in Table 5.1 and Table 5.2. The results of the type I error rate α and its corresponding expected sample size *Mean* N_α, and the type II error rate β and its corresponding expected sample size *Mean* N_β on the beta prior distribution are listed in Table 5.4 and Table 5.5. The expected Bayes risk *Mean Risk* of two-point prior distribution and beta prior distribution are listed in Table 5.6 and Table 5.7, respectively.

Comparing Table 5.1 with Table 5.4, and Table 5.2 with Table 5.5, it can been seen that the type I error rates α of Bayes sequential decision procedures with the beta prior distribution are smaller than those of Bayes sequential decision procedures with the two-point prior distribution; the type II error rates β of Bayes sequential decision procedures with the beta prior distribution are larger than those of Bayes sequential decision procedures with two-points prior distribution, but can still be smaller than those of procedures of Pocock in some situations. Comparing Table 5.6 with Table 5.7, it can been seen that the average Bayes risks of Bayes sequential decision procedures with beta prior distributions are bigger than those with two-point prior procedures. This is because the two-point prior distribution could be assumed when we have very strong prior information.

Table 5.4 Type I error rate α and expected sample size Mean N_α

		t=500		t=1000		t=5000		t=10000	
u	v	α	MeanN_α	α	MeanN_α	α	MeanN_α	α	MeanN_α
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 17.5									
Bayes sequential decision procedure:									
1.98	9.12	.0003	13	.0003	13.1	$<$ .0001	14.6	$<$ .0001	14.6
0.89	5.96	.0003	13	$<$ .0001	13.0	$<$ .0001	13.1	$<$ .0001	14.6
0.43	4.79	.0003	13	$<$ .0001	13.0	$<$ .0001	13.1	$<$ .0001	14.6
0.18	5.15	$<$ .0001	13	$<$ .0001	13.0	$<$ .0001	13.1	$<$ .0001	13.1
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 28									
Bayes sequential decision procedure:									
1.98	9.12	.0189	21	.0189	21	.0006	22.8	.0005	27.0
0.89	5.96	.0032	21	.0032	21	.0005	22.8	.0001	22.9
0.43	4.79	.0032	21	.0032	21	.0001	21.4	.0001	22.9
0.18	5.15	.0004	21	.0004	21	$<$ .0001	21.1	$<$ .0001	21.4
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 55.5									
Bayes sequential decision procedure:									
1.98	9.12	.0607	43	.0607	43	.0283	47.7	.0117	54.3
0.89	5.96	.0244	43	.0244	43	.0105	45.2	.0041	48.9
0.43	4.79	.0087	43	.0087	43	.0091	43.7	.0032	45.6
0.18	5.15	.0008	43	.0008	43	.0002	43.1	$<$ .0001	43.2

Table 5.5 Type II error rate β and expected sample size Mean N_β

		t=500		t=1000		t=5000		t=10000	
u	v	β	MeanN_β	β	MeanN_β	β	MeanN_β	β	MeanN_β
Pocock's procedure: $p_1 = 0.01$, Max N = 39, Mean N = 20.5									
Bayes sequential decision procedure:									
1.98	9.12	0.3326	13	0.1952	15.7	0.2175	28.3	0.2175	28.3
0.89	5.96	0.3326	13	0.3817	16.9	0.2622	26.2	0.2175	28.3
0.43	4.79	0.3326	13	0.3817	16.9	0.2622	26.2	0.2175	28.3
0.18	5.15	0.5843	13	0.4163	16.3	0.2927	25.2	0.2686	30.1
Pocock's procedure: $p_1 = 0.05$, Max N = 63, Mean N = 33.5									
Bayes sequential decision procedure:									
1.98	9.12	0.1917	21	0.1917	21	0.1479	34.1	0.1261	37.7
0.89	5.96	0.3674	21	0.3674	21	0.2120	34.9	0.1990	40.4
0.43	4.79	0.3674	21	0.3674	21	0.2845	35.8	0.2077	42.8
0.18	5.15	0.5666	21	0.5666	21	0.4607	34.7	0.3769	42.2
Pocock's procedure: $p_1 = 0.10$, Max N = 129, Mean N = 66									
Bayes sequential decision procedure:									
1.98	9.12	0.1237	43	0.1237	43	0.0864	49.7	0.0632	58.6
0.89	5.96	0.2175	43	0.2175	43	0.1591	52.3	0.1158	63.8
0.43	4.79	0.3390	43	0.3390	43	0.2439	48.2	0.1794	61.3
0.18	5.15	0.6145	43	0.6145	43	0.6273	53.5	0.5177	73.7

Table 5.6 Expected Bayes Risk for the Two-point Prior

	t=500	t=1000	t=5000	t=10000
w_1	Mean Risk	Mean Risk	Mean Risk	Mean Risk
$p_1 = 0.01$,	Max N = 39,	Mean $N_\alpha = 20.5$,		Mean N_β=23.0
0.3	-13.04	-26.08	-136.90	-276.22
0.5	-9.14	-18.28	-93.91	-195.75
0.7	-4.09	-10.47	-56.14	-112.27
$p_1 = 0.05$,	Max N = 63,	Mean $N_\alpha = 33.5$,		Mean N_β=26.5
0.3	-10.38	-23.44	-132.77	-269.89
0.5	-5.96	-15.40	-91.34	-189.99
0.7	-2.91	-8.26	-53.57	-108.93
$p_1 = 0.10$,	Max N = 129,	Mean $N_\alpha = 66$,		Mean N_β=72
0.3	-10.69	-21.38	-126.77	-262.83
0.5	-6.31	-12.63	-86.85	-184.83
0.7	-2.82	-5.63	-47.66	-106.17

Table 5.7 Expected Bayes Risk for the Beta Prior

	t=500	t=1000	t=5000	t=10000
w_1	Mean Risk	Mean Risk	Mean Risk	Mean Risk
$p_1 = 0.01$,	Max N = 39,	Mean $N_\alpha = 20.5$,		Mean N_β=23.0
0.3	-7.61	-18.02	-110.53	-221.05
0.5	-5.73	-13.30	-79.83	-165.41
0.7	-3.41	-8.19	-48.54	-100.15
$p_1 = 0.05$,	Max N = 63,	Mean $N_\alpha = 33.5$,		Mean N_β=26.5
0.3	-7.66	-15.32	-99.79	-206.49
0.5	-4.82	-9.64	-67.85	-140.27
0.7	-2.12	-4.24	-32.61	-71.77
$p_1 = 0.10$,	Max N = 129,	Mean $N_\alpha = 66$,		Mean N_β=72
0.3	-8.26	-16.52	-91.41	-195.74
0.5	-4.68	-9.35	-55.01	-121.97
0.7	-1.50	-3.00	-18.97	-48.94

Chapter 6

Bayes Group Sequential Decision Clinical Trials on Survival Time Data

In this chapter, Bayes sequential decision theory is introduced into clinical trials with survival time data. The exponential distributions and proportional hazard models for survival time are studied. The one treatment clinical trial with an exponential distribution response is discussed first to give a picture of using Bayes sequential decision theory. Following this, clinical trials comparing two treatments with exponential distribution responses are studied. After a brief introduction of the non-parametric Bayes analysis, several approaches are discussed for group sequential clinical trials with proportional hazard models for survival time data.

6.1 One Treatment Clinical Trials with Exponential Distribution Responses

6.1.1 The Problem

A clinical trial is designed to test the effect of a new treatment. The major outcome of the clinical trial is an exponentially distributed random variable. Its probability density function is,

$$f(t) = \lambda e^{-\lambda t}, \tag{6.1}$$

and the survival function is,

$$S(t) = P(T \geq t) = e^{-\lambda t},$$

where the hazard rate $\lambda > 0$.

Assume that λ_0 is a break-even value of λ. The new treatment is not considered effective if $\lambda \geq \lambda_0$. The conventional hypotheses will be, H_0: $\lambda \geq \lambda_0$(no treatment effect), vs H_1: $\lambda < \lambda_0$(treatment effect).

Suppose a size n random sample t_1, t_2, ..., t_m, c_1, c_2, ..., c_{n-m}, are from the exponential distribution (6.1), where t_1, t_2,..., t_m are the failure times and c_1, c_2,..., c_{n-m} are the censored times. Let S be the random variable of total survival time and s be its observed value, that is,

$$s = \sum_{j=1}^{m} t_j + \sum_{k=1}^{n-m} c_k. \tag{6.2}$$

The likelihood function of λ is then

$$L(\lambda; t_1, ..., t_m, c_1, ..., c_{n-m}) = \prod_{j=1}^{m} f(t_j) \prod_{k=1}^{n-m} S(c_k)$$

$$
\begin{aligned}
&= \prod_{j=1}^{m}(\lambda e^{-\lambda t_j}) \prod_{k=1}^{n-m} e^{-\lambda c_k} \\
&= \lambda^m e^{-\lambda(\sum_{j=1}^{m} t_j + \sum_{k=1}^{n-m} c_k)} \\
&\propto \frac{\lambda^m}{\Gamma(m)}(\sum_{j=1}^{m} t_j + \sum_{k=1}^{n-m} c_k)^{m-1} e^{-\lambda(\sum_{j=1}^{m} t_j + \sum_{k=1}^{n-m} c_k)} \\
&= \frac{\lambda^m}{\Gamma(m)} s^{m-1} e^{-\lambda s}, \qquad (6.3)
\end{aligned}
$$

which is proportional to the gamma probability density function with shape parameter m and scale parameter λ. The likelihood function shows that given the number of failures m, the total survival time S is a sufficient statistic for λ, and the statistic S follows the gamma distribution with shape parameter equal to the number of failures m and scale parameter equal to the hazard rata λ, denoted by $\Gamma_{S|m}(s; m, \lambda)$. The statistical inferences on λ based on the data $t_1, ..., t_m, c_1, ..., c_{n-m}$ can therefore be replaced by the datum s defined as in (6.2) without loss of information.

6.1.2 Design of the Clinical Trial

Assume patient accrual is uniform in period $(0, s_a)$ with a constant rate R. The maximum number of patients in the clinical trial is then Rs_a. The clinical trial is monitored at either **I)** selected times after treatment or randomisation, u_1, u_2, ..., u_l, every 6 months say, or **II)** specified number of new failures m_1, m_2,...,m_l, say, every 10 failures. The l is the maximum number of analyses. Suppose patients are observed until failed or censored by the termination of the clinical trial.

I) The clinical trial is monitored at times u_1, u_2, ..., u_l.

Let the total number of failures until times u_1, u_2, ..., u_l be denoted by $d(u_1)$, $d(u_2)$,..., $d(u_l)$, respectively. At each analysis j, $j = 1, 2, ..., l$, the total observed

survival time s_j is,

$$s_j = \sum_{i=1}^{d(u_j)} t_i + \sum_{k=1}^{R\min\{u_j, s_a\} - d(u_j)} c_{kj}, \tag{6.4}$$

where $t_i, i = 1, 2, ..., d(u_j)$, are the failure times at the analysis j, and c_{kj}, $k = 1, 2, ..., R\min\{u_j, s_a\} - d(u_j)$, are censored times until u_j. Let S_j, $j = 1, 2, ..., l$, be the corresponding random variables of s_j, $j = 1, 2, ..., l$. Given the number of failures $d(u_j)$, the S_j, $j = 1, 2, ..., l$, are from the gamma distributions with shape parameter equal to $d(u_j)$ and scale parameter equal to λ by Section 6.1.1. Let

$$u_0 = 0, \quad d(u_0) = 0,$$
$$m_i = d(u_i) - d(u_{i-1}).$$

The m_i, $i = 1, 2, ..., j$, $j = 1, 2, ..., l$, are the number of failures at the period (u_{i-1}, u_i). The total number of failures at analysis j, denoted by $d(u_j)$, is equal to $\sum_{i=1}^{j} m_i$. The gamma random variable S_j can be then decomposed as a sum of the gamma random variables X_i, that is,

$$S_j = \sum_{i=1}^{j} X_i, \tag{6.5}$$

where random variables X_i, $i = 1, 2, ..., j$, are from the gamma distributions with shape parameters equal to m_i, $j = 1, 2, ..., l$, respectively, and the constant scale parameter equal to λ.

II) The clinical trial is monitored at number of new failures m_1, m_2, ..., m_l.

Let the corresponding monitoring times at the cumulated number of failures m_1, $m_1 + m_2$, ..., $\sum_{i=1}^{l} m_i$ be denoted by u_1, u_2,...,u_l. At each analysis j, $j = 1, 2, ..., l$, let S_j be the total survival time random variable and s_j be its observed value,

that is,

$$s_j = \sum_{i=1}^{m_1+...+m_j} t_i + \sum_{k=1}^{R\min\{u_j,s_a\}-(m_1+...+m_j)} c_{kj}, \tag{6.6}$$

where $t_i, i = 1, 2, ..., \sum_{i=1}^{j} m_i$, are the failure times until analysis j, and $c_{kj}, k = 1, 2, ..., R\min\{u_j, s_a\} - (m_1 + ... + m_j)$ are censored time until analysis j. The $S_j, j = 1, 2, ..., l$, are from the gamma distributions with probability density functions $\Gamma(s; \sum_{i=1}^{j} m_i, \lambda)$ by Section 6.1.1. As in the case **I)**, the S_j can be decomposed as,

$$S_j = \sum_{i=1}^{j} X_i, \tag{6.7}$$

where $X_i, i = 1, 2, ...j$ are from the gamma distributions $\Gamma(x; m_i, \lambda)$, respectively.

Hence, the group sequential samples in both cases **I)** and **II)** can be replaced by the classical sequential sample X_j, $j = 1, 2, ..., l$, where X_j, $j = 1, 2, ..., l$, are from the gamma distributions with probability density function $\Gamma(x; m_j, \lambda)$. The m_j is the number of new failures at analysis j and $\sum_{i=1}^{j} m_i$ is the total number of failures at analysis j.

6.1.3 Basic Elements of Bayes Sequential Decision Theory

Bayes sequential decision theory is applied in the clinical trial described above in Section 6.1.1 and Section 6.1.2. The basic elements of Bayes sequential decision theory in this study are,

1) the parameter of interest λ is assumed to have the gamma prior distribution with parameters α and β, that is,

$$\lambda \sim w(\lambda) = \Gamma(\lambda; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} & \lambda > 0 \\ 0 & otherwise. \end{cases} \tag{6.8}$$

2) the decision space is

$D = \{d_0$ (no treatment effect), d_1(treatment effect)$\}$.

3) the loss and the cost functions

For computational simplicity, piecewise continuous loss functions are used here. They are defined as,

$$\begin{aligned} L(\lambda, d_0) &= \begin{cases} K & \lambda < \lambda_0 \\ 0 & \lambda \geq \lambda_0, \end{cases} \\ L(\lambda, d_1) &= \begin{cases} 0 & \lambda < \lambda_0 \\ K & \lambda \geq \lambda_0. \end{cases} \end{aligned} \tag{6.9}$$

Suppose the unit of $K(> 0)$ is the cost of enrolling a patient into the trial. This cost is constant through the trial.

4) the sequential sample X_j, $j = 1, 2, ..., l$, obtained by Section 6.1.2, are from the gamma distributions, which are,

$$X_j \sim f(x|\lambda, m_j) = \Gamma(x; m_j, \lambda), \tag{6.10}$$

where the m_j is the number of failures between the $(j-1)$th analysis and jth analysis, and the $\sum_{i=1}^{j} m_i$ are the total number of failures at the jth analysis.

From above elements, the posterior distribution of λ and predictive distribution of x are obtained as follows. At each analysis j, $j = 1, 2, ..., l$, after observing $X_1 = x_1$, $X_2 = x_2$, ..., $X_j = x_j$ with the number of new failures m_1, m_2, ..., m_j, respectively, the posterior distribution of λ, denoted by $w^j = w(\lambda|x_1, x_2, ..., x_j)$, is,

$$\begin{aligned} w(\lambda|x_1, x_2, ..., x_j) &= \frac{f(x_1, x_2, ..., x_j|\lambda)w(\lambda)}{\int_0^\infty f(x_1, x_2, ..., x_j|\lambda)w(\lambda)\, d\lambda} \\ &= \frac{\prod_{i=1}^{j} \frac{\lambda^{m_i}}{\Gamma(m_i)} x_i^{m_i-1} e^{-\lambda x_i} (\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda})}{\int_0^\infty \prod_{i=1}^{j} \frac{\lambda^{m_i}}{\Gamma(m_i)} x_i^{m_i-1} e^{-\lambda x_i} (\frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda})\, d\lambda} \end{aligned}$$

$$= \frac{\lambda^{\alpha+\sum_{i=1}^{j} m_i-1} e^{-(\beta+\sum_{i=1}^{j} x_i)\lambda}}{\int_0^\infty \lambda^{\alpha+\sum_{i=1}^{j} m_i-1} e^{-(\beta+\sum_{i=1}^{j} x_i)\lambda}\, d\lambda}$$

$$= \frac{(\beta+\sum_{i=1}^{j} x_i)^{\alpha+\sum_{i=1}^{j} m_i}}{\Gamma(\alpha+\sum_{i=1}^{j} m_i)} \lambda^{\alpha+\sum_{i=1}^{j} m_i-1} e^{-(\beta+\sum_{i=1}^{j} x_i)\lambda}, \quad (6.11)$$

which is still the gamma probability density function with parameters $\alpha+\sum_{i=1}^{j} m_i$ and $\beta+\sum_{i=1}^{j} x_i$.

At each interim analysis j, $j = 1, 2, ..., l-1$, after observing $X_1 = x_1$, $X_2 = x_2$,..., $X_j = x_j$, the predictive density function of x with number of new failures equal to m, denoted by $f(x|x_1, x_2, ..., x_j)$, is,

$$\begin{aligned}
&f(x|x_1, x_2, ..., x_j) \\
&= E_{\lambda|x_1,x_2,...,x_j} f(x|\lambda) \\
&= \int_0^\infty f(x|\lambda) w(\lambda|x_1, x_2, ..., x_j)\, d\lambda \\
&= \int_0^\infty \left(\frac{\lambda^m}{\Gamma(m)} x^{m-1} e^{-\lambda x}\right)\left(\frac{(\beta+\sum_{i=1}^{j} x_i)^{(\alpha+\sum_{i=1}^{j} m_i)}}{\Gamma(\alpha+\sum_{i=1}^{j} m_i)} \lambda^{\alpha+\sum_{i=1}^{j} m_i-1} e^{-(\beta+\sum_{i=1}^{j} x_i)\lambda}\right) d\lambda \\
&= \frac{x^{m-1}(\beta+\sum_{i=1}^{j} x_i)^{\alpha+\sum_{i=1}^{j} m_i}}{\Gamma(m)\Gamma(\alpha+\sum_{i=1}^{j} m_i)} \int_0^\infty \lambda^{\alpha+m+\sum_{i=1}^{j} m_i-1} e^{-(\beta+x+\sum_{i=1}^{j} x_i)\lambda}\, d\lambda \\
&= \frac{x^{m-1}(\beta+\sum_{i=1}^{j} x_i)^{\alpha+\sum_{i=1}^{j} m_i}}{\Gamma(m)\Gamma(\alpha+\sum_{i=1}^{j} m_i)} \frac{\Gamma(\alpha+m+\sum_{i=1}^{j} m_i)}{(\beta+x+\sum_{i=1}^{j} x_i)^{\alpha+m+\sum_{i=1}^{j} m_i}} \\
&= \frac{\Gamma(\alpha+m+\sum_{i=1}^{j} m_i)}{\Gamma(\alpha+\sum_{i=1}^{j} m_i)\Gamma(m)} \frac{x^{m-1}(\beta+\sum_{i=1}^{j} x_i)^{\alpha+\sum_{i=1}^{j} m_i}}{(\beta+x+\sum_{i=1}^{j} x_i)^{\alpha+m+\sum_{i=1}^{j} m_i}}. \qquad (6.12)
\end{aligned}$$

6.1.4 Bayes Sequential Decision Procedure

The one interim analysis is first considered to describe Bayes sequential decision procedure of the study described in Sections 6.1.1, 6.1.2, and 6.1.3. The two types of monitoring are studied separately.

I) Monitoring the clinical trial at the selected times u_1 and u_2

In the interim analysis, at the time u_1, the number of failures equal to m_1 and the total survival time X_1 equal to x_1 are observed. The X_1 is from the gamma distribution with parameters m_1 and λ, denoted by $\Gamma(x;\ m_1, \lambda)$. The risk from stopping the trial, denoted by $r_0(w^1, 1)$, is

$$\begin{aligned} r_0(w^1, 1) &= \min_{\{d_0, d_1\}} E_{\lambda|x_1} L(\lambda, d) \\ &= \min\{E_{\lambda|x_1} L(\lambda, d_0),\ E_{\lambda|x_1} L(\lambda, d_1)\}, \end{aligned}$$

where, by the loss function (6.9) and posterior distribution function (6.11),

$$\begin{aligned} E_{\lambda|x_1} L(\lambda, d_0) &= \int_0^{\infty} L(\lambda, d_0)\ w(\lambda|x_1)\ d\lambda \\ &= \int_0^{\lambda_0} K w(\lambda|x_1)\ d\lambda \\ &= K \int_0^{\lambda_0} \Gamma(\lambda;\ \alpha + m_1, \beta + x_1) d\ \lambda, \\ E_{\lambda|x_1} L(\lambda, d_1) &= \int_0^{\infty} L(\lambda, d_1)\ w(\lambda|x_1)\ d\lambda \\ &= \int_{\lambda_0}^{\infty} K w(\lambda|x_1)\ d\lambda \\ &= K \int_{\lambda_0}^{\infty} \Gamma(\lambda;\ \alpha + m_1, \beta + x_1)\ d\lambda. \end{aligned}$$

Let M_1 be the median of the gamma distribution $\Gamma(\lambda; \alpha + m_1, \beta + x_1)$, then

$$r_0(w^1, 1) = \begin{cases} KP(\lambda \le \lambda_0 | \alpha + m_1, \beta + x_1) & \lambda_0 \le M_1 \\ K[1 - P(\lambda \le \lambda_0 | \alpha + m_1, \beta + x_1)] & \lambda_0 > M_1, \end{cases} \tag{6.13}$$

where $P(\lambda \le \lambda_0 | a, b)$ is the probability of event $\{\lambda \le \lambda_0\}$ of the gamma distribution with the shape parameter equal to a and the scale parameter equal to b.

On the other hand, if the next observation $X = x$ with the number of failures m is observed, then the Bayes risk from stopping the trial after observing $X_1 = x_1$

and $X = x$ is,

$$
\begin{aligned}
& r_0(w^1(\lambda|X=x),2) \\
&= r_0(w^2,2) \\
&= \min_{\{d_0,d_1\}} E_{\lambda|x_1,x} L(\lambda,d) \\
&= K \min\{P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x),\ 1-P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x)\}.
\end{aligned}
$$

Since

$$
\begin{aligned}
& P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x) \\
&\qquad = \textstyle\int_0^{\lambda_0} \frac{(\beta+x_1+x)^{(\alpha+m_1+m)}}{\Gamma(\alpha+m_1+m)} \lambda^{\alpha+m_1+m-1} e^{-(\beta+x_1+x)\lambda} d\lambda \\
&\qquad = \textstyle\int_0^{(\beta+x_1+x)\lambda_0} \frac{1}{\Gamma(\alpha+m_1+m)} \tilde{\lambda}^{\alpha+m_1+m-1} e^{-\tilde{\lambda}} d\tilde{\lambda},
\end{aligned}
$$

which is a monotonic increasing function of x given λ_0, α, β, m and x_1. So there is an unique value $M = \frac{M_2}{\lambda_0} - \beta - x_1$, where M_2 is the median of the gamma distribution $\Gamma(\tilde{\lambda}; \alpha+m_1+m, 1)$, such that

$$
r_0(w^1(\lambda|X=x),2) = \begin{cases} K\ P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x) & x \le M \\ K[1-P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x)] & x > M. \end{cases}
$$

Hence, the expected risk from observing the next sample X given the number of failures m, denoted by $E^*_{|m} r_0(w^1(\lambda|X),2)$, is

$$
\begin{aligned}
& E^*_{|m} r_0(w^1(\lambda|X),2) \\
&= \int_0^\infty r_0(w^1(\lambda|X=x),2) f(x|X_1=x_1)dx \\
&= K\{\int_0^M P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x) \frac{\Gamma(\alpha+m_1+m)}{\Gamma(\alpha+m_1)\Gamma(m)} \frac{(\beta+x_1)^{\alpha+m_1} x^{m-1}}{(\beta+x_1+x)^{\alpha+m_1+m}} dx \\
&+ \int_M^\infty [1-P(\lambda \le \lambda_0|\alpha+m_1+m,\beta+x_1+x)] \frac{\Gamma(\alpha+m_1+m)}{\Gamma(\alpha+m_1)\Gamma(m)} \frac{(\beta+x_1)^{\alpha+m_1} x^{m-1}}{(\beta+x_1+x)^{\alpha+m_1+m}} dx\} \\
&= K\{\int_{\frac{M}{\beta+x_1+M}}^1 \frac{1}{\beta(\alpha+m_1,\ m)} \left(\frac{x}{\beta+x_1+x}\right)^{\alpha+m_1-1} \left(\frac{\beta+x_1}{\beta+x_1+x}\right)^{m-1} d\left(\frac{x}{\beta+x_1+x}\right) \\
&- \frac{(\beta+x_1)^{\alpha+m_1}}{\Gamma(\alpha+m_1)} \int_0^{\lambda_0} \lambda^{\alpha+m_1-1} e^{-(\beta+x_1)\lambda} d\lambda
\end{aligned}
$$

$$+\frac{2(\beta+x_1)^{\alpha+m_1}}{\Gamma(\alpha+m_1)\Gamma(m)}\int_0^M\int_0^{\lambda_0}\lambda^{\alpha+m_1+m-1}e^{-(\beta+x_1+x)\lambda}x^{m-1}d\lambda dx\}. \tag{6.14}$$

Let $g(m)$ be the probability distribution function of m. The expected risk from observing the next sample X is,

$$E^*r_0(w^1(\lambda|X),2)=\sum_m E^*_{|m}r_0(w^1(\lambda|X),2)\ g(m). \tag{6.15}$$

If the Bayes risk from stopping the trial is less than the expected risk from observing the next sample X, that is, $r_0(w^1,1)\le E^*r_0(w^1(\lambda|X),2)+R\ \min\{u_2-u_1,s_a-u_1,0\}$, then the clinical trial is terminated. Otherwise the clinical trial is continued to the final analysis.

At the final analysis, after observing $X_1=x_1$ and $X_2=x_2$, the decision with the Bayes stopping risk is chosen.

The simulation on the expected risk $E^*(w^1(\lambda|X),2)$

The distribution of the number of failures $g(m)$ is very complicated. Alternatively the following simulation method may be used to get the expected risk $E^*r_0(w^1(\lambda|X),2)$ in (6.15).

At the interim analysis, assume that the clinical trial is continued and that the next sample $X=x$ with number of new failures m is observed. The expected risk given m, $E^*_{|m}(w^1(\lambda|X),2)$, can be calculated by using (6.14). The expected risk $E^*(w^1(\lambda|X),2)$ is then

$$E^*(w^1(\lambda|X),2)=\frac{\sum_N E^*_{|m}(w^1(\lambda|X),2)}{N},$$

where N is the number of simulations.

II) Monitoring the clinical trial at number of failures m_1 **and** m_1+m_2

At the interim analysis, after observing the total survival time $X_1=x_1$ with

number of failures equal to m_1, the Bayes risk from stopping the trial $r_0(w^1, 1)$ is obtained by (6.13). The expected risk from continuing the trial $E^*(w^1(\lambda|X), 2)$ can be obtained directly from (6.14) with

$$E^*(w^1(\lambda|X), 2) = E^*_{|m=m_2}(w^1(\lambda|X), 2),$$

as the clinical trial is expected to be monitored at the next number of failures equal to m_2.

Let u be the corresponding monitoring time with number of failures equal to $m_1 + m_2$. The computation of the expected monitoring time $E(u)$ is complicated. The simulation method is used to get the $E(u)$.

If the Bayes risk from stopping the trial is less than the expected risk from optimally continuing the trial, that is, $r_0(w^1, 1) \leq E^* r_0(w^1(\lambda|X), 2) + R \min\{E(u) - u_1, s_a - u_1, 0\}$, then the clinical trial is terminated. Otherwise the clinical trial is continued to the final analysis.

At the final analysis, after observing $X_1 = x_1$ and $X_2 = x_2$, the decision with the Bayes stopping risk is chosen.

If the Bayes sequential decision procedure with more than one interim analysis is designed in the clinical trial, then at each interim analysis j, the computation of the Bayes risk from stopping the trial is similar to that of one interim analysis, where the posterior density $w^1 = w(\lambda|x_1)$ is replaced by the posterior density $w^j = w(\lambda|x_1, x_2, ..., x_j)$; and the expected risk from optimally continuing the trial would need be obtained by simulations as that of one interim analysis. The clinical trial could be stopped at the interim analysis j when the Bayes risk from stopping the trial is less than the expected risk from continuing the trial.

6.2 Two Treatments Clinical Trials on Exponential Distribution Responses

6.2.1 The Problem

Consider a clinical trial comparing an experimenatal treatment with the standard treatment. The main outcome of treatments is exponentially distributed with hazard rates λ_e and λ_s for the experimental and standard treatments respectively. Let the treatment difference in efficacy is $\gamma = \frac{\lambda_e}{\lambda_s}$. The conventional hypotheses are

$$H_0 : \gamma \geq 1 \quad vs \quad H_1 : \gamma < \gamma_0, \tag{6.16}$$

where $\gamma_0 (< 1)$ is a break-even value of γ. The H_0 corresponds to the experimental treatment not better and H_1 to the experimental treatment better.

Patients are uniformly enrolled into the trial in period $(0, s_a)$ at a constant rate R and are allocated randomly and equally to each treatment. The clinical trial is monitored at either **I)** selected times u_1, u_2, ..., u_l or **II)** selected number of new failures $m_{1e} + m_{1s}$, $m_{2e} + m_{2s}$, ...,$m_{le} + m_{ls}$. The l is the maximum number of analyses. Sections 6.1.1 and 6.1.2 have shown that given the number of failures, the group sequential sample from an exponential distribution can be regarded as a classical sequential sample from gamma distributions. Therefore, the group sequential sample from these exponential distributions with hazard rates λ_e for the experimental treatment and λ_s for the standard treatment can be replaced by classical sequential samples of gamma distributions without loss of information on inference of λ_e and λ_s. Let the X_{jk}, $k = e, s$, $j = 1, 2, ..., l$, be the sequences of classical samples, then

$$X_{jk} \sim f(x|m_{jk}, \lambda_k) = \Gamma(x_k; m_{jk}, \lambda_k), \quad k = e, s, \ j = 1, 2, ..., l, \tag{6.17}$$

where $\sum_{i=1}^{j} X_{ik}$ and $\sum_{i=1}^{j} m_{ik}$ are the total survival time and number of failures, respectively, for treatment k, $k = e, s$, at analysis j, $j = 1, 2, ..., l$.

Suppose hazard rates λ_e and λ_s are independent and the prior distributions of λ_k, $k = e, s$, are the gamma distributions which are,

$$\lambda_k \sim w(\lambda_k) = \Gamma(\lambda_k; \alpha_k, \beta_k). \tag{6.18}$$

At each analysis j, $j = 1, 2, ..., l$, after observing $X_{1k} = x_{1k}, X_{2k} = x_{2k}, ..., X_{jk} = x_{jk}$, with number of failures m_{1k}, m_{2k},..., m_{jk}, respectively, $k = e, s$, by (6.10) the posterior distribution of λ_k, denoted by $w(\lambda_k | x_{1k}, x_{2k}, ..., x_{jk})$, are the gamma distributions with parameters $\alpha_k + \sum_{i=1}^{j} m_{ik}$ and $\beta_k + \sum_{i=1}^{j} x_{ik}$, that is

$$w(\lambda_k | x_{1k}, x_{2k}, ..., x_{jk}) = \Gamma(\lambda_k; \alpha_k + \sum_{i=1}^{j} m_{ik}, \beta_k + \sum_{i=1}^{j} x_{ik}).$$

The posterior distribution of γ, denoted by $w(\gamma | (x_{1e}, x_{1s}), (x_{2e}, x_{2s}), ..., (x_{je}, x_{js}))$, is then,

$$\begin{aligned}
&w(\gamma | (x_{1e}, x_{1s}), (x_{2e}, x_{2s}), ..., (x_{je}, x_{js})) \\
&= \int_0^\infty \lambda_s w(\gamma\lambda_s | x_{1e}, x_{2e}, ..., x_{je})\ w(\lambda_s | x_{1s}, x_{2s}, ..., x_{js})\ d\lambda_s \\
&= \int_0^\infty \lambda_s \Gamma(\gamma\lambda_s; \alpha_e + \sum_{i=1}^{j} m_{ie}, \beta_e + \sum_{i=1}^{j} x_{ie}) \Gamma(\lambda_s; \alpha_s + \sum_{i=1}^{j} m_{is}, \beta_s + \sum_{i=1}^{j} x_{is})\ d\lambda_s \\
&= \int_0^\infty \{ \frac{(\beta_e + \sum_{i=1}^{j} x_{ie})^{\alpha_e + \sum_{i=1}^{j} m_{ie}}}{\Gamma(\alpha_e + \sum_{i=1}^{j} m_{ie})} (\gamma\lambda_s)^{\alpha_e + \sum_{i=1}^{j} m_{ie} - 1} e^{-(\beta_e + \sum_{i=1}^{j} x_{ie})\gamma\lambda_s} \\
&\quad \times \frac{(\beta_s + \sum_{i=1}^{j} x_{is})^{\alpha_s + \sum_{i=1}^{j} m_{is}}}{\Gamma(\alpha_s + \sum_{i=1}^{j} m_{is})} \lambda_s^{\alpha_s + \sum_{i=1}^{j} m_{is} + 1} e^{-(\beta_s + \sum_{i=1}^{j} x_{is})\lambda_s} \} d\lambda_s \\
&= \frac{(\beta_e + \sum_{i=1}^{j} x_{ie})^{\alpha_e + \sum_{i=1}^{j} m_{ie}} (\beta_s + \sum_{i=1}^{j} x_{is})^{\alpha_s + \sum_{i=1}^{j} m_{is}}}{\Gamma(\alpha_e + \sum_{i=1}^{j} m_{ie}) \Gamma(\alpha_s + \sum_{i=1}^{j} m_{is})}\ \gamma^{\alpha_e + \sum_{i=1}^{j} m_{ie} - 1} \\
&\quad \times \int_0^\infty \lambda_s^{\alpha_e + \sum_{i=1}^{j} m_{ie} + \alpha_s + \sum_{i=1}^{j} m_{is}}\ e^{-[(\beta_e + \sum_{i=1}^{j} x_{ie})\gamma + (\beta_s + \sum_{i=1}^{j} x_{is})]\lambda_s} d\lambda_s \\
&= \frac{\Gamma(\alpha_e + \sum_{i=1}^{j} m_{ie} + \alpha_s + \sum_{i=1}^{j} m_{is} + 1)}{\Gamma(\alpha_e + \sum_{i=1}^{j} m_{ie}) \Gamma(\alpha_s + \sum_{i=1}^{j} m_{is})}
\end{aligned}$$

$$\times \frac{(\beta_e + \sum_{i=1}^{j} x_{ie})^{\alpha_e + \sum_{i=1}^{j} m_{ie}} \ (\beta_s + \sum_{i=1}^{j} x_{is})^{\alpha_s + \sum_{i=1}^{j} m_{is}} \ \gamma^{\alpha_e + \sum_{i=1}^{j} m_{ie} - 1}}{[(\beta_e + \sum_{i=1}^{j} x_{ie})\gamma + (\beta_s + \sum_{i=1}^{j} x_{is})]^{\alpha_e + \sum_{i=1}^{j} m_{ie} + \alpha_s + \sum_{i=1}^{j} m_{is} + 1}} \quad (6.19)$$

At each interim analysis $j, j = 1, 2, ..., l-1$, the predictive density function of (X_e, X_s) with number of new failures (m_e, m_s) after observing $X_{1k} = x_{1k}, X_{2k} = x_{2k}, ..., X_{jk} = x_{jk}$, $k = e, s$, is,

$$\begin{aligned}
&f(x_e, x_s|(x_{1e}, x_{1s}), (x_{2e}, x_{2s}), ..., (x_{je}, x_{js})) \\
&= f(x_e|x_{1e}, x_{2e}, ..., x_{je}) \ f(x_s|x_{1s}, x_{2s}, ..., x_{js}) \\
&= E_{\lambda_e|x_{1e}, x_{2e}, ..., x_{je}} f(x_e|\lambda_e) \ E_{\lambda_s|x_{1s}, x_{2s}, ..., x_{js}} f(x_s|\lambda_s) \\
&= \prod_{k=e,s} \int_0^\infty f(x_k|\lambda_k) w(\lambda_k|x_{1k}, x_{2k}, ..., x_{jk}) \ d\lambda_k \\
&= \prod_{k=e,s} \frac{x_k^{m_k - 1}}{\Gamma(m_k)} \frac{(\beta_k + \sum_{i=1}^{j} x_{ik})^{\alpha_k + \sum_{i=1}^{j} m_{ik}}}{\Gamma(\alpha_k + \sum_{i=1}^{j} m_{ik})} \\
&\quad \times \prod_{k=e,s} \int_0^\infty \lambda_k^{m_k + \alpha_k + \sum_{i=1}^{j} m_{ik} - 1} e^{-(\beta_k + \sum_{i=1}^{j} x_{ik} + x_k)\lambda_k} \ d\lambda_k \\
&= \prod_{k=e,s} \frac{\Gamma(\alpha_k + m_k + \sum_{i=1}^{j} m_{ik})}{\Gamma(\alpha_k + \sum_{i=1}^{j} m_{ik})\Gamma(m_k)} \frac{{x_k}^{m_k - 1}(\beta_k + \sum_{i=1}^{j} x_{ik})^{\alpha_k + \sum_{i=1}^{j} m_{ik}}}{(\beta_k + x_k + \sum_{i=1}^{j} x_{ik})^{\alpha_k + m_k + \sum_{i=1}^{j} m_{ik}}}.
\end{aligned} \quad (6.20)$$

6.2.2 Bayes Sequential Decision Procedure

Corresponding to the conventional hypotheses (6.16), piecewise continuous loss functions for choosing the experimental treatment and the standard treatment are defined as,

$$L(\gamma, d_e) = \begin{cases} 0 & \gamma \leq 1 \\ K & \gamma > 1, \end{cases}$$

$$L(\gamma, d_s) = \begin{cases} K & \gamma < \gamma_0 \\ 0 & \gamma \geq \gamma_0, \end{cases} \tag{6.21}$$

in which the unit of $K(>0)$ is the cost of enrolling a patient into the trial. This cost is constant through the trial. There is no loss in the range of equivalence $(\gamma_0,\ 1)$.

One interim analysis is considered first. At the interim analysis, let m_{1k}, $k = e, s$, be the number of failures and u_1 be the monitoring time. The Bayes risk from stopping the clinical trial after observing $X_{1k} = x_{1k}$, $k = e, s$, is

$$r_0(w^1, 1) = \min\{E_{\gamma|x_{1e},x_{1s}} L(\gamma, d_e),\ E_{\gamma|x_{1e},x_{1s}} L(\gamma, d_s)\}, \tag{6.22}$$

where,

$$\begin{aligned}
&E_{\gamma|x_{1e},x_{1s}} L(\gamma, d_e) \\
&\quad = \int_1^\infty K w(\gamma|x_{1e}, x_{1s}) d\gamma \\
&\quad = K \int_1^\infty \frac{\Gamma(\alpha_e+m_{1e}+\alpha_s+m_{1s}-1)}{\Gamma(\alpha_e+m_{1e})\Gamma(\alpha_s+m_{1s})} \frac{(\beta_e+x_{1e})^{\alpha_e+m_{1e}}\ (\beta_s+x_{1s})^{\alpha_s+m_{1s}}\ \gamma^{\alpha_e+m_{1e}}}{[(\beta_e+x_{1e})\gamma+(\beta_s+x_{1s})]^{\alpha_e+m_{1e}+\alpha_s+m_{1s}-1}} d\gamma, \\
&E_{\gamma|x_{1e},x_{1s}} L(\gamma, d_s) \\
&\quad = \int_0^{\gamma_0} K w(\gamma|x_{1e}, x_{1s}) d\gamma \\
&\quad = K \int_0^{\gamma_0} \frac{\Gamma(\alpha_e+m_{1e}+\alpha_s+m_{1s}-1)}{\Gamma(\alpha_e+m_{1e})\Gamma(\alpha_s+m_{1s})} \frac{(\beta_e+x_{1e})^{\alpha_e+m_{1e}}\ (\beta_s+x_{1s})^{\alpha_s+m_{1s}}\ \gamma^{\alpha_e+m_{1e}}}{[(\beta_e+x_{1e})\gamma+(\beta_s+x_{1s})]^{\alpha_e+m_{1e}+\alpha_s+m_{1s}-1}} d\gamma.
\end{aligned}$$

On the other hand, the expected risk from observing the next samples $X_k, k = e, s$, with the number of new failures m_k, $k = e, s$, at the monitoring time u, denoted by $E^* r_0(w^1(\gamma|X_e, X_s), 2)$, is

$$\begin{aligned}
&E^* r_0(w^1(\gamma|X_e, X_s), 2) \\
&= \int_0^\infty \int_0^\infty r_0(w^1(\gamma|X_e = x_e, s = x_s), 2)\ f(x_e, x_s|x_{1e}, x_{1s})\ dx_e dx_s,
\end{aligned} \tag{6.23}$$

where the predictive density $f(x_e, x_s|x_{1e}, x_{1s})$ is obtained by (6.20), and the risk $r_0(w^1(\gamma|X_e = x_e, X_s = x_s), 2)$ is the stopping risk after observing $X_{1k} = x_{1k}$ and

$X_k = x_k$ with the number of failures m_{1k}, m_k, respectively, $k = e, s$, that is,

$$\begin{aligned} & r_0(w^1(\gamma|X_e = x_e, X_s = x_s), 2) \\ & \quad = \min\{E_{\gamma|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_e),\ E_{\gamma|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_s)\}, \end{aligned} \tag{6.24}$$

where

$$\begin{aligned} & E_{\gamma\,|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_e) \\ & = \int_1^\infty K w(\gamma|(x_{1e}, x_{1s}), (x_e, x_s)) d\gamma \\ & = K \frac{\Gamma(\alpha_e+m_{1e}+m_e+\alpha_s+m_{1s}+m_s+1)}{\Gamma(\alpha_e+m_{1e}+m_e)\Gamma(\alpha_s+m_{1s}+m_s)} (\beta_e + x_{1e} + x_e)^{\alpha_e+m_{1e}+m_e}\ (\beta_s + x_{1s} + x_s)^{\alpha_s+m_{1s}+m_s} \\ & \quad \times \int_1^\infty \frac{\gamma^{\alpha_e+m_{1e}+m_e-1}}{[(\beta_e+x_{1e}+x_e)\gamma+(\beta_s+x_{1s}+x_s)]^{\alpha_e+m_{1e}+m_e+\alpha_s+m_{1s}+m_s+1}}\, d\gamma, \\ & E_{\gamma\,|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_s) \\ & = \int_0^{\gamma_0} K w(\gamma|(x_{1e}, x_{1s}), (x_e, x_s)) d\gamma \\ & = K \frac{\Gamma(\alpha_e+m_{1e}+m_e+\alpha_s+m_{1s}+m_s+1)}{\Gamma(\alpha_e+m_{1e}+m_e)\Gamma(\alpha_s+m_{1s}+m_s)} (\beta_e + x_{1e} + x_e)^{\alpha_e+m_{1e}+m_e}\ (\beta_s + x_{1s} + x_s)^{\alpha_s+m_{1s}+m_s} \\ & \quad \times \int_0^{\gamma_0} \frac{\gamma^{\alpha_e+m_{1e}+m_e-1}}{[(\beta_e+x_{1e}+x_e)\gamma+(\beta_s+x_{1s}+x_s)]^{\alpha_e+m_{1e}+m_e+\alpha_s+m_{1s}+m_s+1}}\, d\gamma. \end{aligned}$$

If $r_0(w^1, 1) \le E^* r_0(w^1(\gamma|X_e, X_s), 2) + R\min\{u - u_1, s_a - u_1, 0\}$, then the clinical trial is terminated. Otherwise the clinical trial is continued to the final analysis.

At the final analysis, after observing $X_1 = x_{1k}$ and $X_2 = x_{2k}$, the decision with the Bayes risk is chosen.

The computation of the expected risk $E^* r_0(w^1(\gamma|X_e, X_s), 2)$ in (6.23) is difficult even given the number of failures m_k, $k = e, s$. It is very complicated to divide the two dimensional space $\{(x_e, x_s); x_e \ge 0, x_s \ge 0\}$ in (6.24) into two parts, say x_e^2 and x_s^2 ($\{(x_e, x_s); x_e \ge 0, x_s \ge 0\} = x_e^2 + x_s^2$), as follows,

$$\begin{aligned} & r_0(w^1(\gamma|X_e = x_e, X_s = x_s), 2) \\ & \quad = \min\{E_{\gamma|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_e),\ E_{\gamma|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_s)\} \\ & \quad = \begin{cases} E_{\gamma|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_e) & (x_e, x_s) \in x_e^2 \\ E_{\gamma|(x_{1e},x_{1s}),(x_e,x_s)} L(\gamma, d_s) & (x_e, x_s) \in x_s^2, \end{cases} \end{aligned}$$

in order to get $E^* r_0(w^1(\gamma|x_e, x_s), 2)$ by (6.23). Therefore, it is necessary to find some approximate approaches to be able to use Bayes sequential decision theory in the study.

6.2.3 Log Gamma Approximation

The approximation method

If the random variable X is from the gamma distribution, which is $X \sim \Gamma(\alpha, \beta)$, then

$$\begin{aligned} 2\beta X &\sim \chi^2(2\alpha), \\ \ln(2\beta X) &\dot{\sim} N(\ln(2\alpha) - \frac{1}{2\alpha} - \frac{1}{12\alpha^2}, \ \frac{2}{2\alpha - 1}), \\ \ln(2X) - \ln(2\alpha) + \frac{1}{2\alpha} + \frac{1}{12\alpha^2} &\dot{\sim} N(-\ln(\beta), \ \frac{2}{2\alpha - 1}), \end{aligned}$$

that is, the distribution of log gamma random variable can be approximated by the normal distribution. In terms of the sequential samples X_{jk}, $k = e, s$, $j = 1, 2..., l$, in (6.17) from the study described in Sections 6.2.1 and 6.2.2, it is obtained that

$$\ln(2X_{jk}) - \ln(2m_{jk}) + \frac{1}{2m_{jk}} + \frac{1}{12m_{jk}^2} \ \dot{\sim} \ N(-\ln(\lambda_k), \ \frac{2}{2m_{jk} - 1}).$$

Let

$$\begin{aligned} Y_{jk} &= \ln(2X_{jk}) - \ln(2m_{jk}) + \frac{1}{2m_{jk}} + \frac{1}{12m_{jk}^2}, \\ Z_j &= Y_{js} - Y_{je}, \end{aligned} \tag{6.25}$$

then

$$Z_j \dot{\sim} N(\ln(\frac{\lambda_e}{\lambda_s}), \ \frac{2}{2m_{je} - 1} + \frac{2}{2m_{js} - 1}). \tag{6.26}$$

The mean of the normal distribution in (6.26) is equal to $\ln(\frac{\lambda_e}{\lambda_s})$, which is a measure of the difference between the experimental treatment and the standard treatment, the sequence of random variables Z_j, $j = 1, 2, ..., l$, may be used to test the treatment difference instead of the sequential sample X_{jk}, $j = 1, 2, ..., l$, $k = e, s$, in (6.17). Again using the log transformation in the hazard rates λ_k, $k = e, s$ in (6.18), let $\delta = \ln\gamma = \ln(\frac{\lambda_e}{\lambda_s})$, then $\delta = \ln(\lambda_e) - \ln(\lambda_s)$ has an asymptotic normal distribution.

Hence, the problem in Section 6.2.1 becomes the study that a classical sequential clinical trial with the normal distribution responses Z_j, $j = 1, 2, ..., l$, is designed to test the hypotheses $H_0 : \delta = \ln(\frac{\lambda_e}{\lambda_s}) \geq 0$, $H_1 : \delta < \delta_0 = \ln\gamma_0$. The Bayes sequential decision procedure on normal responses in Chapter 4 can be used.

The variances of the sequential samples Z_j, $j = 1, 2, ..., l$, are equal to $\frac{2}{2m_{je}-1} + \frac{2}{2m_{js}-1})$, $j = 1, 2, ..., l$, respectively, which are not constant. The procedure discussed in Section 4.4 should be used. This is described as follow.

The Bayes sequential decision procedure

the basic elements

The Bayes sequential decision procedure is based on the following basic elements. 1) the parameter of interest is $\delta = \ln(\frac{\lambda_e}{\lambda_s})$. Assume that δ has the normal prior distribution, that is,

$$\delta \sim w(\delta) = N(\nu_0, \tau_0^2), \tag{6.27}$$

where,

$$\begin{aligned} \nu_0 &= \ln(\frac{\alpha_e\beta_s}{\alpha_s\beta_e}) - \frac{1}{2\alpha_e} - \frac{1}{12\alpha_e^2} + \frac{1}{2\alpha_s} + \frac{1}{12\alpha_s^2}, \\ \tau_0^2 &= \frac{2}{2\alpha_e - 1} + \frac{2}{2\alpha_s - 1}. \end{aligned}$$

The above α_e, β_e, α_s, β_s are obtained by (6.18). The posterior distribution of δ given $Z_1 = z_1$,...,$Z_j = z_j$, $j = 1, ..., l$, is the normal distribution,

$$w(\delta|z_1, ..., z_j) = N(\nu_j, \tau_j^2), \tag{6.28}$$

where,

$$\begin{aligned}
\nu_j &= \frac{\sum_{i=1}^{j} \frac{z_i}{\sigma_i^2} + \frac{\nu_0}{\tau_0^2}}{\sum_{i=1}^{j} \frac{1}{\sigma_i^2} + \frac{1}{\tau_0^2}}, \\
\tau_j^2 &= \frac{1}{\sum_{i=1}^{j} \frac{1}{\sigma_i^2} + \frac{1}{\tau_0^2}}, \\
\sigma_i^2 &= \frac{2}{2m_{ie} - 1} + \frac{2}{2m_{is} - 1}.
\end{aligned}$$

2) the decision space.

After the trial is terminated, the decision $d \in D$ with the Bayes risk is chosen, where the decision space D is defined as

$$D = \{ \text{ experimental treatment } d_e, \text{ standard treatment } d_s \}.$$

3) the loss and cost functions

Corresponding to the loss functions of (6.21), the loss functions here are

$$\begin{aligned}
L(\delta, d_e) &= \begin{cases} 0 & \delta \leq 0 \\ K & \delta > 0, \end{cases} \\
L(\delta, d_s) &= \begin{cases} K & \delta < \delta_0 \\ 0 & \delta \geq \delta_0, \end{cases}
\end{aligned}$$

in which the unit of $K(> 0)$ is the cost of enrolling a patient into the trial. This cost is constant through the trial.

4) the sequential sample

The sequential sample $Z_j, j = 1, 2, ..., l$, are from normal distributions $N(\delta, \frac{2}{2m_{je}-1}+$

$\frac{2}{2m_{js}-1}$), respectively.

At each interim analysis j, $j = 1, 2, ..., l-1$, given $Z_1 = z_1$, $Z_2 = z_2$,...,$Z_j = z_j$, the predictive density function of Z with number of failures m_k, $k = e, s$, is still the normal distributions,

$$\begin{aligned} f(z|z_1, z_2, ..., z_j) &= E_{\delta|z_1,z_2,...,z_j} f(z|\delta) \\ &= N(\nu_j, \tau_j^2 + \frac{2}{2m_e - 1} + \frac{2}{2m_s - 1}) \end{aligned} \tag{6.29}$$

where ν_j and τ_j^2 are the same as those in (6.28).

the stopping risk and the expected risk

At each interim analysis j, $j = 1, 2, ..., l-1$, after observing $Z_1 = z_1$, $Z_2 = z_2$,...,$Z_j = z_j$, the risk from stopping the trial is,

$$\begin{aligned} r_0(w^j, j) &= \min\{E_{\delta|z_1,z_2,...,z_j} L(\delta, d_e),\ E_{\delta|z_1,z_2,...,z_j} L(\delta, d_s)\} \\ &= \min\{K(1 - \Phi(\frac{-\nu_j}{\tau_j})),\ K\Phi(\frac{\delta_0 - \nu_j}{\tau_j})\}. \end{aligned}$$

If the clinical trial is monitored at **I)** selected times u_1, u_2, ..., u_l, then the expected risk from observing the next sample Z at time $u = u_{j+1}$, is

$$E^*(r_0(w^j(Z), j+1) = \sum_{m_e, m_s} E^*_{|m_e,m_s}(r_0(w^j(Z), j+1)\ g(m_e, m_s).$$

where $E^*_{|m_e,m_s}(r_0(w^j(Z), j+1)$ is the expected risk given m_e and m_s, and $g(m_e, m_s)$ is the distribution of (m_e, m_s). The distribution $g(m_e, m_s)$ is very complicated. The $E^*(r_0(w^j(Z), j+1)$ may be obtained by the simulation method used in Section 6.1.4.

If the clinical trial is monitored at **II)** selected number of failures $m_{1e} + m_{1s}$, $m_{2e} + m_{2s}$, ..., $m_{le} + m_{ls}$, then the expected risk from observing the next sample

with number of new failures $m_e + m_s = m$ is

$$E^*(r_0(w^j(Z), j+1) = \sum_{m_e+m_s=m} E^*_{|m_e,m_s}(r_0(w^j(Z), j+1)\ g(m_e, m_s).$$

Let u be the corresponding monitoring time with number of new failures equal to $m_e + m_s = m$. The $E(u_j)$ may be obtained by simulation as in Section 6.1.4.

By backward induction described in Section 3.2.2, the risk from optimally continuing the trial with not more than $l - j$ groups of observations, denoted by $r_{l-j}(w^j, j)$. is obtained based on the above stopping risk and expected risk.

the procedure

At each interim analysis j, $j = 1, 2, ..., l-1$, if the Bayes risk from stopping the trial is less than the risk of continuing the trial, that is,

$$r_0(w^j, j) \leq r_{l-j}(w^j, j),$$

then we stop the trial. Otherwise we continue the trial to observe the next group of samples.

6.2.4 Log-rank Statistics Approximation

Consider the study described in Section 6.2.1. Let $\theta = \ln(\frac{\lambda_e}{\lambda_s})$ and $\lambda_s = \lambda_0$. The θ is a log hazard ratio and $\lambda_e = \lambda_0 exp(\theta)$. A proportional hazard model is defined with the hazard rate,

$$\lambda = \lambda_0 exp(\theta z), \tag{6.30}$$

where z is an indicator variable of treatments, that is, when $z = 1$, then the patient is from the experimental treatment with the hazard rate $\lambda = \lambda_0 exp(\theta) = \lambda_e$; whereas when $z = 0$, the patient is from the standard treatment with the hazard

rate $\lambda = \lambda_0 = \lambda_s$. Under this model, $\theta = 0$ means that there is no treatment difference between the experimental and standard treatments.

Tsiatis(1981,1982) has shown that the log rank test of a proportional hazard model computed over time indeed behaves like a partial sum of independent normal variables, with the variance proportional to the number of failures observed.

Suppose the clinical trial is monitored at arbitrarily selected calendar times $u_j, j = 1, 2, ..., l$, where the total number of failures at these time points are denoted by $d(u_1), d(u_2), ..., d(u_l)$. Let $V(u_j)$, $j = 1, 2, ..., l$, denote the value of the log rank test computed at calendar time u_j. Tsiatis(1981,1982) has derived that,

$$V(u_j) \approx X_1 + X_2 + ... + X_j, \tag{6.31}$$

where $X_1, X_2, ..., X_j$ are independent normal random variables with mean $E(X_i)$ and variance $Var(X_i)$, $i = 1, 2..., j$, respectively, as follows.

$$\begin{aligned} E(X_i) &= \theta[d(u_i) - d(u_{i-1})]p(1-p), \\ Var(X_i) &= [d(u_i) - d(u_{i-1})]p(1-p), \end{aligned} \tag{6.32}$$

in which p denotes the proportion of failures in one of the treatments.

At each analysis j, $j = 1, 2, ..., l$, let

$$\begin{aligned} Y_j &= \frac{X_j}{[d(u_j) - d(u_{j-1})]p(1-p)}, \\ Y_j &\dot{\sim} N(\theta, \frac{1}{[d(u_j) - d(u_{j-1})]p(1-p)}). \end{aligned} \tag{6.33}$$

When the number of failures is large enough, the sequence of random variables Y_j, $j = 1, 2, ..., l$, may be used to test the hypotheses $H_0 : \theta = 0$. The sequential sample in (6.17) can be then replaced by the sequential sample Y_j, $j = 1, 2, ..., l$, in (6.33), where Y_j's are from the normal distributions. The Bayes sequential decision procedure in clinical trials with normal distribution responses in Chapter 4 can be

used to approximate the Bayes sequential decision clinical trials on exponential distributions for survival time data.

6.2.5 Comparison Between Log gamma Approximation and Log-rank Statistics Approximation

The asymptotic normal distributions

The log transformation of the gamma random variable in Section 6.3.2 and the log-rank statistics in Section 6.3.4 are asymptotically normal distributed with the probability density functions (6.26) and (6.33), respectively. The mean of the normal distribution (6.26) equal to $\delta = \ln(\frac{\lambda_e}{\lambda_s})$ is the same as the mean of the normal distribution (6.33) which is $\theta = \ln(\frac{\lambda_e}{\lambda_s})$. If the number of failures m_{jk}, $k = e, s$, $j = 1, 2, ..., l$, is large enough, then the variance of the normal distribution (6.26) is

$$\begin{aligned} \frac{2}{2m_{je}-1} + \frac{2}{2m_{js}-1} &\approx \frac{1}{m_{je}} + \frac{1}{m_{js}} \\ &= \frac{1}{(m_{je}+m_{js})\frac{m_{je}}{m_{je}+m_{js}}\frac{m_{js}}{m_{je}+m_{js}}} \\ &= \frac{1}{[d(u_j)-d(u_{j-1})]p(1-p)}, \end{aligned}$$

which is the variance of the normal distribution (6.33). Therefore these two asymptotic normal distributions are almost the same.

Monte Carlo simulations

Monte Carlo simulations are used to compare these two approximation methods based on the following example.

Assume 300 patients enter a clinical trial for comparing an experimental treatment with the standard treatment. Patients are uniformly enrolled into the clinical trial and are allocated randomly and equally for each treatment during a year. The

major outcome of the clinical trial is the survival time from treatment. Suppose the survival time from the standard treatment has the exponential distribution with hazard rate λ_s equal to 0.8. The experimental treatment is considered effective if the hazard ratio is less than 0.5. The hazard rate λ_e is then equal to $0.5 \times 0.8 = 0.4$. Suppose this is the one interim analysis and the clinical trial is monitored at the middle of the study year. Let λ_e and λ_s have the same prior gamma distribution $\Gamma(\lambda; 1, b)$. The $b > 0$ was changed in simulations to study the inferences of the prior information.

Simulation 1. The log gamma approximation. After observing the group sequential samples from the exponential distributions with hazard rates λ_e and λ_s, the corresponding classical sequential samples, X_{jk}, $j = 1, 2$, which are from the gamma distributions $\Gamma(m_{jk}, \lambda_k)$, $j = 1, 2$, are calculated by (6.2) for each treatment k, $k = e, s$. Following (6.25) and (6.26), the sequence of asymptotic normal random variables Z_j, $j = 1, 2$, are obtained. The Bayes sequential decision procedure described in Section 6.2.3 is used into the simulation.

Simulation 2. The log-rank statistics approximation. The log-rank statistics are calculated after observing the group sequential samples from exponential distributions with hazard rate λ_e and λ_s. Using the result of (6.33), the log-rank statistics are treated as from the normal distributions. The Bayes sequential decision procedure in clinical trials with normal distribution responses may be used in the simulation.

The comparisons are based on the type I error rate α, its expected sample size Mean N_α, type II error rate β, its expected sample size Mean N_β and the average Bayes risk. The results are listed in Table 6.1 with the loss function $K = 2000$ and parameter of prior distribution $b = 1.5$, where the 95% confidence intervals are listed in the brackets. It shows that these two approximation methods are reasonably close. The same conclusion is found when the values of K and b are changed.

The prior gamma distribution was changed in simulations. The result of simulations on log gamma transformation is listed in Table 6.2. The type I error rate α, its mean sample size Mean N_α, type II error rate β, its mean sample size Mean N_β, and average Bayes risk are quite stable with the change of b from 1 to 2.5. This is the same for the log-rank statistics approximation, which is shown in Table 6.3.

Table 6.1 The comparison of two approximation methods

	log gamma transformation	log-rank statistics
type I error α	0.1058 (0.0953, 0.1162)	0.0820 (0.0650, 0.0990)
Mean N_α	190.1 (188.3, 191.9)	185.7 (181.7, 189.7)
Average Bayes risk	40.14 (38.88, 41.40)	37.93 (35.06, 40.80)
type II error β	0.1168 (0.1079, 0.1257)	0.1740 (0.1505, 0.1975)
Mean N_β	202.6 (200.6, 204.6)	203.7 (199.2, 208.2)
Average Bayes risk	50.65 (49.30, 52.00)	54.57 (51.58, 57.56)

Table 6.2 The log gamma approximation with different prior distributions

b	α	Mean N_α	Mean Risk	β	Mean N_β	Mean Risk
1	0.1040	190.5	39.78	0.1170	202.9	50.26
1.25	0.1054	190.2	40.01	0.1168	202.7	50.44
1.5	0.1058	190.1	40.14	0.1168	202.6	50.65
2	0.1060	190.0	40.32	0.1172	202.2	51.04
2.25	0.1066	189.8	40.52	0.1176	202.0	51.23
2.5	0.1064	189.7	40.57	0.1176	201.8	51.50

Table 6.3 Log-rank statistics approximation with different prior distributions

b	α	Mean N_α	Mean Risk	β	Mean N_β	Mean Risk
1	0.0850	186.0	36.48	0.1870	202.2	54.68
1.25	0.0930	190.2	36.24	0.1680	199.7	56.13
1.5	0.0820	185.7	37.93	0.1740	203.7	54.57
2	0.0810	186.9	37.09	0.1790	206.6	53.05
2.25	0.0990	184.5	39.70	0.1850	200.1	52.77
2.5	0.0950	187.5	38.57	0.1790	201.2	52.00

6.2.6 Conclusion

Bayes sequential decision theory was applied to the clinical trials comparing two treatments with survival time data. When the survival time is from an exponential distribution with hazard rate equal to λ and the prior distribution of the λ is a gamma distribution, then the Bayes sequential decision procedure in clinical trials with normal distribution responses can be used as an approximation. If the gamma prior distribution has the form $\Gamma(1, b)$, then the inferences of the average sample size and the average Bayes risk from changing the value of b are small.

6.3 Non-Parametric Bayes Analysis

6.3.1 Prior Distributions on Spaces of Probability Measures

Ferguson(1973, 1974) has said that the Bayes approach in treating non-parametric problems has not been very successful, and this is due primarily to the difficulty in finding workable prior distributions on the parameter space. There are two desirable properties of a prior distribution for nonparametric problems.

(1) The support of the prior distribution should be large – with respect to some suitable topology on the space of probability distributions on a given sample space.

(2) Posterior distributions given a sample of observations from the true probability distribution should be manageable analytically.

Ferguson(1973) first introduced the Dirichlet process and used it as prior for an unknown cumulative distribution function. Later Doksum(1974) and Ferguson(1974) have addressed the tail-free processes and processes neutral to the right as prior probability distributions on spaces of probability measures or distribution functions. These three processes have often been used in the non-parametric Bayes analysis.

Here the prior distributions on the space of all probability measures are restricted on $(R, \mathcal{B})$ where R is the real line and $\mathcal{B}$ is the σ-algebra of Borel subsets of R. Let

$$\mathcal{F} = \{P : P \ \textit{is a probability measure on} \ (R, \mathcal{B})\}.$$

The definitions of the Dirichlet process, the process neutral to the right and the tail-free process

The Dirichlet process

Let $\alpha(\cdot)$ be a finite non-null measure on $(R, \mathcal{B})$, and let $P(\cdot)$ be a stochastic process indexed by elements of $\mathcal{B}$. The P is called a Dirichlet process with parameter α and written $P \in \mathcal{D}(\alpha)$, if for every finite measurable partition $\{B_1, ..., B_m\}$ of R, the random vector $(P(B_1), ..., P(B_m))$ has a Dirichlet distribution with parameter $(\alpha(B_1), ..., \alpha(B_m))$. In particular, for every $B \in \mathcal{B}$, $P(B) \in \mathcal{B}e(\alpha(B), \alpha(R) - \alpha(B))$, therefore $E[P(B)] = \frac{\alpha(B)}{\alpha(R)}$.

Equivalently, the Dirichlet process can be defined as follows. Let $\alpha(t) = \alpha((-\infty, t])$ and $F(t) = P((-\infty, t])$. The P is called a Dirichlet process with parameter α and written $P \in \mathcal{D}(\alpha)$ (or $F \in \mathcal{D}(\alpha)$), if the process $F(t)$ may be written as $\frac{Z_t}{Z_\infty}$, where Z_t is a process with independent increments, $Z_t \in \Gamma(\alpha(t), 1)$, and $Z_\infty = \lim_{t\to\infty} Z_t \in \Gamma(\alpha(R), 1)$, $F(t) \in \mathcal{B}e(\alpha(t), \alpha(R) - \alpha(t))$.

The process neutral to the right

A random distribution function $F(t)$ on the real line is said to be neutral to the right if for every m and $t_1 < t_2 < \cdots < t_m$, there exist independent random variables $V_1, V_2, ..., V_m$, such that $(1 - F(t_1), 1 - F(t_2), ..., 1 - F(t_m))$ has the same distribution as $(V_1, V_1V_2, ..., \Pi_1^m V_i)$.

Essentially, F is said to be neutral to the right, if $(1-F(t_1), \frac{1-F(t_2)}{1-F(t_1)}, ..., \frac{1-F(t_m)}{1-F(t_{m-1})})$ are independent when the denominators are non zero.

The tail-free process

Let $\{\pi_m;\ m = 1, 2, ..\}$ be a tree of measurable partitions of $(R, \mathcal{B})$. The distribution of a random probability P on $(R, \mathcal{B})$ is said to be tail-free with respect to $\{\pi_m\}$ if there exist a family of non-negative independent random variables $\{V_{m,B};\ m = 1, 2, ...,\ B \in \pi_m\}$ such that for every $m = 1, 2, ...,$ if $B_j \in \pi_j, j = 1, ..., m$ and $B_m \subset ... \subset B_1$, then $P(B_m) = \Pi_{j=1}^m V_{j,B_j}$.

A random distribution function F is tail-free with respect to the tail (s, ∞) if for all $s = t_0 < ... < t_k$ there exist non-negative independent random variables

V_1,...,V_k independent of $\{F(t) : t \leq s\}$ such that

$$(F(t_1), ..., F(t_k)) = (F(s) + [1 - F(s)][1 - \Pi_{j \leq i}(1 - V_j)],\ i = 1, ..., k)$$

If F is neutral to the right then $Y_t = -log(1 - F(t))$ has independent increments. Let Y_t be a process with independent increments, non-decreasing a.s., right continuous a.s., $\lim_{t \to -\infty} Y_t = 0$ a.s. and $\lim_{t \to \infty} Y_t = \infty$ a.s. Then $F(t) = 1 - e^{-Y_t}$ is a random distribution function neutral to the right.

Characterisations of three processes

Let three trivial types of processes be

T_1. P non-random ($F \equiv F_0$);

T_2. P degenerate at a random point ($F = I_{[X,\infty)}$) where X has distribution F_0);

T_3. P concentrated on two non-random points ($F = UI_{[a,\infty)} + (1 - U)I_{[b,\infty)}$ where U has an arbitrary distribution on [0,1], and $a < b$).

1. If F is a Dirichlet process ($F \in \mathcal{D}(\alpha)$), then with probability one F is discrete.

The limitations of the Dirichlet process stem mainly from the fact that it chooses discrete distributions with probability one, so that it is expected to have some observations repeated exactly. To avoid these limitations, we should try to find some workable priors that choose continuous distribution with probability one. There are some among the tailfree processes.

2. The support of $\mathcal{D}(\alpha)$ with respect to the topology of weak convergence is the set of all distribution whose support is contained in the support of α. Therefore, if the support of α is R, then the support of $\mathcal{D}(\alpha)$ with respect to convergence in law is $\mathcal{F}$.

3. If P is neutral with respect to every finite measurable partition, then P is

either a Dirichlet process or of types T_1, T_2 or T_3.

4. If P is tail-free with respect to every tree of partitions, then P is either a Dirichlet process or of types T_1, T_2 or T_3.

The Dirichlet process is essentially the only random probability that is independent of the defined partitions in the sense of having the desired independence properties for all sequences of partitions.

5. If for every measurable set B, the posterior distribution of $P(B)$ given a sample $X_1, ..., X_n$ from P, depends on $X_1, ..., X_n$ only through the number of observations that fall in B (and not on where they fall within or outside of B), then P is either a Dirichlet process or of types T_1, T_2 or T_3.

The above property makes the posterior distribution of the Dirichlet process easy to handle. However, this is not necessarily a desirable property since the posterior distribution is rather insensitive to the values of the sample. From this point of view, the tail-free process prior that chooses absolutely continuous distributions with probability one would seem to be more appropriate. But the tail-free with continuous singular or absolutely continuous with probability one has new drawbacks. A minor one is that the expectation of $F(t)$ is now more difficult to compute. The main drawback is that the dyadic points of subdivision play a strong role in the posterior distribution.

6. F is neutral to the right if and only if F is tail-free with respect to (s, ∞) for all s in R.

Posterior distributions of three processes

If F is a Dirichlet process $F \in \mathcal{D}(\alpha)$ and if $X_1, ..., X_n$ is a sample from F, then the posterior distribution of F given $X_1, ..., X_n$ is $\mathcal{D}(\alpha + \sum_1^n \delta_{X_i})$, where δ_x is the measure giving mass one to x.

If F is neutral to the right, and if $X_1, ..., X_n$ is a sample from F, then the posterior distribution of F given $X_1, ..., X_n$ is neutral to the right.

If the distribution of P is tail-free with respect to $\{\pi_m\}$ and if $X_1, ..., X_n$ is a sample from P, then the posterior distribution of P given $X_1, ..., X_n$ is tail-free with respect to $\{\pi_m\}$.

6.3.2 Non-parametric Bayes Analysis of the Proportional Hazard Model

The non-parametric Bayes estimation of a survival or reliability function has been considered by several authors (for example, see Susarla and Van Ryzin, 1976, Ferguson and Phadia, 1979, Dykstra and Laud, 1981, Padgett and Wei,1981, and Berliner and Hill, 1988). Kalbfleisch(1978) and Hjort(1990) have discussed the non-parametric Bayes analysis of the proportional hazard model.

Let $T > 0$ represent the failure time of an individual. The covariate variable z is equal to 0 or 1 to indicate two treatments. The distribution function $F(t)$ and the survival function $S(t)$ are

$$1 - F(t|z) = S(t|z) = exp\{-\Lambda(t)exp(z\beta)\} \tag{6.34}$$

where $exp\{-\Lambda(t)\}$ is a base line survival function and is left unspecified. In the continuous case, the cumulative hazard function $\Lambda(t) = \int_0^t \lambda(u)du$.

Kalbfleisch(1978) has treated the $\Lambda(t)$ as a nuisance parameter with a gamma process, denoted by $\Lambda(t) \sim \mathbf{\Gamma}(c\Lambda^*(t), c)$, where the c is a positive real number and $exp\{-\Lambda^*(t)\}$ is a completely specified survival function. The gamma process that he constructed is a non-decreasing process with independent increments, which is similar to the Dirichlet process. Estimation of β is carried out by determining the marginal probability distribution of data as a function of β, after $\Lambda(t)$ having been eliminated.

Suppose (t_1, z_1),...,(t_n, z_n) are observed from the proportional hazard model (6.34), where t_1, t_2, ..., t_n are observed failure times with $t_1 < t_2 < ... < t_n$, and

z_1, z_2, ..., z_n are values of covariate variable z, then,

$$\begin{aligned} P(T_1 \geq t_1, ..., T_n \geq t_n | \beta, z, \Lambda) &= exp\{-\sum \Lambda(t_i) exp(z_i\beta)\} \\ &= exp\{-\sum_1^n r_i A_i\}, \end{aligned} \tag{6.35}$$

where $r_i = \Lambda(t_i) - \Lambda(t_{i-1})$, and $A_i = \sum_{l \in R(t_i)} exp(z_l\beta)$, $i = 1, ..., n$. Let $t_0 = 0$ and $t_{n+1} = \infty$.

If $q_i = P(T \in [t_{i-1}, t_i) | T \geq t_{i-1}, \Lambda)$, then $\Lambda(t_i) = \sum_{u=1}^{i} -\log(1 - q_u)$. It has been shown by Doksum(1974) that a probability distribution can be specified on the space $\{\Lambda(t)\}$ by specifying the finite dimensional distributions of q_1, q_2,...,q_{n+1} for each partition $[t_{i-1}, t_i)$ (i=1,2,...,n+1). Accordingly, independent prior probability densities can be specified for q_1, q_2,...,q_{n+1} subject to some consistency conditions and the resulting processes are called tailfree or nuetral to the right by Doksum. The $\Lambda(t)$ is by this construction a non-decreasing process with independent increments. The problem then reduces to the specification of a non-decreasing independent increments process for $\Lambda(t)$. To do this, one need only specify independent prios for the r_i's or (q_i's) subject to the condition that the distribution of $r_i + r_{i+1}$ must be the same as would be obtained by direct application of the rules to the combined interval. If the $q_i's$ have independent beta prior distributions, then the resulting process $\Lambda(t)$ is Dirichlet process. The gamma process specifies that $r_i = -\log(1 - q_i)$ have the independent gamma distributions

$$\begin{aligned} r_i &= \Lambda(t_i) - \Lambda(t_{i-1}) \\ &\sim \Gamma\{c\Lambda^*(t_i) - c\Lambda^*(t_{i-1}), c\} \quad (i = 1, ..., n+1), \end{aligned} \tag{6.36}$$

Integrating (6.35) with respect to the distribution of r_i in (6.36) gives,

$$P(T_1 \geq t_1, T_2 \geq t_2, ..., T_n \geq t_n | \beta, z) = exp\{-\sum cB_i\Lambda^*(t_i)\}$$

then the likelihood function of β is

$$L(\beta) = c^n exp\{-\sum cB_i\Lambda^*(t_i)\}\prod\{\lambda^*(t_i)B_i\}, \tag{6.37}$$

where

$$\begin{aligned}\lambda^*(t) &= \frac{d}{dt}\Lambda^*(t),\\ B_i &= -\log\{1 - exp(z_i\beta)/(c + A_i)\},\\ A_i &= \sum_{u\in R(t_i)} exp(z_u\beta), \quad i = 1, ..., n,\end{aligned}$$

and $R(t_i)$ is the set of individuals at risk at time $t_i - 0$.

Although the likelihood function $L(\beta)$ in (6.37) is considered under failure times, right censoring is easily accommodated which is

$$L(\beta) = c^n exp\{-\sum cB_i\Lambda^*(t_i)\}\prod_1^n\{\lambda^*(t_i)B_i\}^{d_i}. \tag{6.38}$$

where $d_i = 0$ or 1 for censored or failure times t_i respectively.

Let $\pi(\beta)$ be a prior distribution of parameter β. The posterior distribution of β is,

$$w(\beta|T, z) \propto \pi(\beta)exp\{-\sum cB_i\Lambda^*(t_i)\}\prod\{\lambda^*(t_i)B_i\}^{d_i}. \tag{6.39}$$

Statistical inferences for β can be derived based on this posterior distribution.

Hjort(1990) has considered a beta process on the cumulative hazard function $\Lambda(t)$. Using the product integral, the distribution function $F(t|z)$ obtained from $\Lambda(t)$ is

$$F(t|z) = 1 - \prod_{[0,t]}\{1 - d\Lambda(s)\}^{exp(z\beta)}, \qquad t \geq 0 \tag{6.40}$$

which is equivalent to the distribution function of (6.34) if and only if $\Lambda(t)$ is continuous.

Hjort(1990) said that the cumulative hazard function $\Lambda(t)$ in (6.40) was more easily interpreted and generalized than that in (6.34), and also is the desire to parallel the construction and results of non-parametric time discreta survival analysis. Loosely speaking, a beta process on a cumulative hazard function is a process which produces cumulative hazard rates whose increments are independent and approximately beta distributed. A particular transformation of a given Dirichlet process produces a special case of the beta process, but the beta process forms a much larger and more flexible class.

Let $\Lambda_0(t)$ be a cumulative hazard function with a finite number of jumps at t_1, t_2,...,t_n and let $c(t)$ be a piecewise continuous, non-negative function on $[0, \infty)$. The Levy process $\Lambda(t)$, ie., one having independent non-negative increments, is called a beta process with parameters $c(t)$, $\Lambda_0(t)$, and denoted by

$$\Lambda(t) \sim beta\{c(t), \Lambda_0(t)\}, \tag{6.41}$$

if the following equation holds,

$$E\left(exp\{-\theta\Lambda(t)\}\right) = \left[\prod_{j:t_j \le t} E(exp(-\theta S_j))\right] exp\{-\int_0^\infty (1 - e^{\theta s}) dL_t(s)\},$$

where

$$\begin{aligned} S_j &= \Lambda(t_j) \sim beta\{c(t_j), \Lambda_0(t_j)\}, \\ dL_t(s) &= \int_0^t c(z) s^{-1} (1-s)^{c(z)-1} d\Lambda_{0,c}(z) ds, \end{aligned}$$

in which $\Lambda_{0,c}(t) = \Lambda_0(t) - \sum_{t_j \le t} \Lambda_0(t_j)$ is $\Lambda_0(t)$ with its jumps removed, then the $\Lambda(t)$ has Levy representation.

If $\Lambda(t)$ is a beta process $beta\{c(t), \Lambda_0(t)\}$ defined by (6.41), then the independent increments have

$$d\Lambda(s) \sim beta\{c(s)d\Lambda_0(s), c(s)(1 - d\Lambda_0(s))\}$$

Suppose the data (t_1, z_1),...,(t_n, z_n) have been observed from (6.40) with $t_1 < t_2 < ... < t_n$. Let $w(\beta)$ be a prior distribution of β. The posterior distribution of β by Hjort(1990) is,

$$\begin{aligned} w(\beta|data) = const.exp\{-\textstyle\int_0^\infty [\psi(c(s) + R(s,\beta)) - \psi(c(s))]c(s)dA_0(s)\} \\ \times \textstyle\prod_{i:\delta_i=1} [\psi(c(t_i) + R(t_i,\beta)) - \psi(c(t_i) + R(t_i,\beta) - \Delta(t_i,\beta))]\pi(\beta), \end{aligned} \tag{6.42}$$

where

$$\begin{aligned} \psi(z) &= \frac{\Gamma'(z)}{\Gamma(z)} = \log z - \frac{1}{2}/z - \frac{1}{12}/z^2 + ..., \\ R(s,\beta) &= \sum_{j=1}^{n} exp(\beta z_j) I_{\{t_j \geq s\}}, \\ \delta_i &= 1 \ failure; \quad or \quad 0 \ censored, \\ \Delta(s,\beta) &= \sum_{j=1}^{n} exp(\beta z_j) I_{\{s=t_j, \delta_j=1\}}. \end{aligned}$$

Both posterior distributions of β in (6.39) and (6.42) are related to data of all individuals. It is, however, very difficult to get the expected risk from observing the next group of samples in using Bayes sequential decision theory.

6.4 Bayes Group Sequential Decision Clinical Trials on Proportional Hazard Model for Survival Time Data

6.4.1 The Problem

Consider a clinical trial comparing an experimental treatment with the standard treatment. The main outcome of the clinical trial is a survival time random variable with the following proportional hazard model. Let T be the survival time random variable. The hazard function $h(t|z)$ and the survival function $S(t|z)$ of T are

$$\begin{aligned} h(t|z) &= \lambda(t)exp(\beta z), \\ S(t|z) &= exp\{-\Lambda(t)exp(\beta z)\}, \end{aligned} \tag{6.43}$$

where $\lambda(t)$ is a baseline hazard function which is left unspecified and $\Lambda(t) = \int_0^t \lambda(s)ds$ is the cumulative hazard function of $\lambda(t)$. The covariate variable z is a 0-1 variable. The $z = 0$ for the outcome from the standard treatment; and $z = 1$ for the outcome from the experimental treatment. The β is the parameter of interest. Assume that the experimental treatment is not considered better if $\beta \geq \beta_2$ and that the experimental treatment is considered better if $\beta \leq \beta_1$. The interval $(\beta_1,\ \beta_2)$ is the range of equivalence. Suppose β has the normal prior distribution with mean ν_0 and variance τ_0^2, that is, $\beta \sim w(\beta) = N(\nu_0, \sigma_0^2)$.

Assume patients accrual is uniform in period $(0,\ s_a)$ with a constant rate R and the allocation of patients is random and equal for each treatment. The clinical trial is monitored at either **I)** selected times $u_1,\ u_2,...,\ u_l$ or **II)** total number of new failures $m_1 = m_{1e} + m_{1s},\ m_2 = m_{2e} + m_{2s},\ ...,\ m_l = m_{le} + m_{ls}$. The l is the maximum number of analyses. For simplity of computation, the loss functions are

defined as

$$
\begin{aligned}
L(\beta, d_s) &= \begin{cases} K & \beta < \beta_1 \\ 0 & \beta \geq \beta_1, \end{cases} \\
L(\beta, d_e) &= \begin{cases} 0 & \beta \leq \beta_2 \\ K & \beta > \beta_2, \end{cases}
\end{aligned} \tag{6.44}
$$

in which the unit of K is the cost of enrolling a patient into the trial. This cost is constant through the trial.

6.4.2 Method 1. Beta Process on the Cumulative Hazard Function $\Lambda(t)$

Let the nuisance parameter $\Lambda(t)$ in the proportional hazard model (6.43) have a prior beta process $beta\{c, \Lambda_0(t)\}$ defined in (6.41), where the c is a constant and $\Lambda_0(t) = \lambda_0 t$.

The posterior distribution of the parameter β

At each analysis j, $j = 1, 2, ..., l$, suppose the values $(\mathbf{t},\ \mathbf{z}) = \{(t_1, z_1), (t_2, z_2), ..., (t_{m_1+m_2+...+m_j},\ z_{m_1+m_2+...+m_j})\}$ have been observed from the proportional hazard model (6.43). The $\sum_{i=1}^{j} m_i$ is the total number of failures at analysis j. The posterior distribution of β, $w(\beta|\mathbf{t}, \mathbf{z})$, is given by (6.42).

The predictive density function

At each interim analysis j, $j = 1, 2, ..., l-1$, let $\mathbf{s} = \{s_1^e, s_2^e, ..., s_{m_e}^e; s_1^s, s_2^s, ..., s_{m_s}^s\}$ be the observed survival time from the next group of observations, where $s_1^k < s_2^k < ... < s_{m_k}^k$ are the survival time from treatment k, $k = e, s$, and $m = m_e + m_s$ is the total number of new failures.

Let $r_i^k = \Lambda(s_i^k) - \Lambda(s_{i-1}^k)$, which is from $beta\{c(\Lambda_0(s_i^k) - \Lambda_0(s_{i-1}^k)), c(1-(\Lambda_0(s_i^k) - \Lambda_0(s_{i-1}^k)))\}$; $R_i^k = (m_k - (i-1))exp(\beta z^k)$, where $z^s = 0$ and $z^e = 1$; and let $G_{r_i^k}(.)$ be the moment generating function of the beta distribution r_i^k, $k = e, s$, $i = 1, ..., m_k$.

Since

$$\begin{aligned}
P(S_1^k \geq s_1^k, ..., S_{m_k}^k \geq s_{m_k}^k|\beta) &= E_\Lambda[P(S_1^k \geq s_1^k, ..., S_{m_k}^k \geq s_{m_k}^k|\beta, \Lambda)] \\
&= E_\Lambda[\prod_{i=1}^{m_k} e^{-\Lambda(s_i^k)exp(\beta z^k)}] \\
&= E_\Lambda[e^{-\sum_{i=1}^{m_k} \Lambda(s_i^k)exp(\beta z^k)}] \\
&= E_\Lambda[exp\{-\sum_{i=1}^{m_k} r_i^k R_i^k\}] \\
&= \prod_{i=1}^{m_k} E_{r_i^k}[exp(-r_i^k R_i^k)] \\
&= \prod_{i=1}^{m_k} G_{r_i^k}(-R_i^k),
\end{aligned}$$

and,

$$\begin{aligned}
&P(S_1^s \geq s_1^s, ..., S_{m_s}^s \geq s_{m_s}^s,\ S_1^e \geq s_1^e, ..., S_{m_e}^e \geq s_{m_e}^e|\mathbf{t}, \mathbf{z}) \\
&= E_{\beta|\mathbf{t},\mathbf{z}}[P(S_1^s \geq s_1^s, ..., S_{m_s}^s \geq s_{m_s}^s,\ S_1^e \geq s_1^e, ..., S_{m_e}^e \geq s_{m_e}^e|\beta)] \\
&= E_{\beta|\mathbf{t},\mathbf{z}}[P(S_1^s \geq s_1^s, ..., S_{m_s}^s \geq s_{m_s}^s|\beta)]\ E_{\beta|\mathbf{t},\mathbf{z}}[P(S_1^e \geq s_1^e, ..., S_{m_e}^e \geq s_{m_e}^e|\beta)] \\
&= E_{\beta|\mathbf{t},\mathbf{z}}[\textstyle\prod_{i=1}^{m_s} G_{r_i^s}(-m_s + i - 1)]\ E_{\beta|\mathbf{t},\mathbf{z}}[\prod_{i=1}^{m_e} G_{r_i^e}((-m_e + i - 1)exp(\beta))] \\
&= \textstyle\prod_{i=1}^{m_s} G_{r_i^s}(-m_s + i - 1)\ E_{\beta|\mathbf{t},\mathbf{z}}[\prod_{i=1}^{m_e} G_{r_i^e}((-m_e + i - 1)exp(\beta))] \\
&= \textstyle\prod_{i=1}^{m_s} G_{r_i^s}(-m_s + i - 1)\ \int_{-\infty}^{\infty} w(\beta|\mathbf{t}, \mathbf{z})[\prod_{i=1}^{m_e} G_{r_i^e}((-m_e + i - 1)exp(\beta))]d\beta \\
&= \textstyle\prod_{i=1}^{m_s} G_{r_i^s}(-m_s + i - 1) \\
&\times \textstyle\int_{-\infty}^{\infty}\{const.exp\{-\frac{(\beta-\nu_0)^2}{2\sigma_0^2} - \sum_{i=0}^{m} \lambda c[\psi(c + \sum_{u=i}^{m} exp(\beta z_u)) - \psi(c)](t_i - t_{i-1})\} \\
&\times \textstyle\prod_{i=1}^{m} [\psi(c + \sum_{u=i}^{m} exp(\beta z_u)) - \psi(c + \sum_{u=i+1}^{m} exp(\beta z_u))] \\
&\times [\textstyle\prod_{i=1}^{m_e} G_{r_i^e}((-m_e + i - 1)exp(\beta))]\}d\beta,
\end{aligned}$$

then, the predictive density function is,

$$f(\mathbf{s}|\mathbf{t},\mathbf{z}) \;=\; -\frac{\mathbf{d}}{\mathbf{ds}}P(S_1^s \geq s_1^s,...,S_{m_s}^s \geq s_{m_s}^s, S_1^e \geq s_1^e,...,S_{m_e}^e \geq s_{m_e}^e|\mathbf{t},\mathbf{z}).$$

Bayes sequential decision procedure: one interim analysis

At the interim analysis, the values $(\mathbf{t},\ \mathbf{z}) = ((t_1,\ z_1),\ (t_2,\ z_2),...,(t_{m_1},\ z_{m_1}))$ have been observed. The risk from stopping the trial is,

$$r_0(w^1,1) \;=\; \min\{E_{\beta|\mathbf{t}}L(\beta,d_s),\quad E_{\beta|\mathbf{t}}L(\beta,d_e)\},$$

where,

$$\begin{aligned}
&E_{\beta|\mathbf{t}}L(\beta,d_s)\\
&= K\int_{-\infty}^{\beta_1} w(\beta|\mathbf{t})d\beta\\
&= K\int_{-\infty}^{\beta_1} const.exp\{-\tfrac{(\beta-\nu_0)^2}{2\tau_0^2} - \textstyle\sum_{i=0}^{m_1}\lambda c[\psi(c+\sum_{u=i}^{m_1}exp(\beta z_u)) - \psi(c)](t_i - t_{i-1})\}\\
&\quad\times\textstyle\prod_{i=1}^{m_1}[\psi(c+\sum_{u=i}^{m_1}exp(\beta z_u)) - \psi(c+\sum_{u=i+1}^{m_1}exp(\beta z_u))]d\beta,\\
&E_{\beta|\mathbf{t}}L(\beta,d_e)\\
&= K\int_{\beta_2}^{\infty} w(\beta|\mathbf{t})d\beta\\
&= K\int_{\beta_2}^{\infty} const.exp\{-\tfrac{(\beta-\nu_0)^2}{2\tau_0^2} - \textstyle\sum_{i=0}^{m_1}\lambda c[\psi(c+\sum_{u=i}^{m_1}exp(\beta z_u)) - \psi(c)](t_i - t_{i-1})\}\\
&\quad\times\textstyle\prod_{i=1}^{m_1}[\psi(c+\sum_{u=i}^{m_1}exp(\beta z_u)) - \psi(c+\sum_{u=i+1}^{m_1}exp(\beta z_u))]d\beta.
\end{aligned}$$

The expected risk from continuing the trial, denoted by $E^*r_0(w^1(\mathbf{s}),2)$, is

$$\begin{aligned}
&E^*_{\mathbf{s}|\mathbf{t},\mathbf{z}}\min\{E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_s),\ E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_e)\}\\
&= \textstyle\int_{\mathbf{s}}\min\{E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_s),\ E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_e)\}\ f(\mathbf{s}|\mathbf{t},\mathbf{z})d\mathbf{s}\\
&= \int\min\{E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_s), E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_e)\}f(\mathbf{s}|\mathbf{t},\mathbf{z})ds_1^s...ds_{m_s}^s ds_1^e...ds_{m_e}^e, \quad (6.45)\\
&\quad s_1^s < ... < s_{m_s}^s\\
&\quad s_1^e < ... < s_{m_e}^e
\end{aligned}$$

where

$$
\begin{aligned}
E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_s) &= const.\int_{-\infty}^{\beta_1} exp\{-\frac{(\beta-\nu_0)^2}{2\sigma_0^2}-\int_0^{\infty}[\psi(c+R(t,\beta))-\psi(c)]c\lambda dt\} \\
&\quad\times\prod_{i=1}^{m_1}[\psi(c)+R(t_i,\beta))-\psi(c+R(t_i,\beta)-\Delta(t_i,\beta))] \\
&\quad\times\prod_{u=1}^{m_s}[\psi(c)+R(s_u^s,\beta))-\psi(c+R(s_u^s,\beta)-\Delta(s_u^s,\beta))] \\
&\quad\times\prod_{v=1}^{m_e}[\psi(c)+R(s_v^e,\beta))-\psi(c+R(s_v^e,\beta)-\Delta(s_v^e,\beta))]d\beta, \\
E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}}L(\beta,d_e) &= const.\int_{\beta_2}^{\infty} exp\{-\frac{(\beta-\nu_0)^2}{2\sigma_0^2}-\int_0^{\infty}[\psi(c+R(t,\beta))-\psi(c)]c\lambda dt\} \\
&\quad\times\prod_{i=1}^{m_1}[\psi(c)+R(t_i,\beta))-\psi(c+R(t_i,\beta)-\Delta(t_i,\beta))] \\
&\quad\times\prod_{u=1}^{m_s}[\psi(c)+R(s_u^s,\beta))-\psi(c+R(s_u^s,\beta)-\Delta(s_u^s,\beta))] \\
&\quad\times\prod_{v=1}^{m_e}[\psi(c)+R(s_v^e,\beta))-\psi(c+R(s_v^e,\beta)-\Delta(s_v^e,\beta))]d\beta,
\end{aligned}
$$

in which $\psi(.)$, $\Delta(.,\beta)$ are the same as those in (6.42), and

$$
R(t,\beta) = \sum_{i=1}^{m} exp(\beta z_i)I_{\{t_i\geq t\}}+\sum_{u=1}^{m_s} I_{\{s_u^s\geq t\}}+\sum_{v=1}^{m_e} exp(\beta)I_{\{s_v^e\geq t\}}.
$$

The computation of the expected risk (6.45) is extremely difficult even for a one interim analysis. It is the same situation when a gamma process (Kalbfleisch 1978) is considered on the cumulative hazard function $\Lambda(t)$. It is necessary to develop some approximation methods.

6.4.3 Method 2. Cox partial likelihood method

The Cox partial likelihood function of the parameter β ignores the nuisance parameter $\Lambda(t)$ in (6.43). It is worthwhile trying to apply Bayes sequential decision theory based on the Cox partial likelihood function.

Suppose the values $(\mathbf{t},\ \mathbf{z}) = (t_1, z_1), (t_2, z_2), ..., (t_m, z_m)\}$ have been observed from the proportional hazard model (6.43). The Cox partial likelihood function is

$$L(\beta, \mathbf{t}, \mathbf{z}) = \prod_{i=1}^{m} \frac{exp(\beta z_{(i)})}{\sum_{u \in R(t_{(i)})} exp(\beta z_u)}, \tag{6.46}$$

where $R(t_{(i)})$ is the set of individuals at risk of failing just prior to $t_{(i)}$.

The posterior distribution of β is

$$\begin{aligned} w(\beta|\mathbf{t}, \mathbf{z}) &= C_1 w(\beta) L(\beta, \mathbf{t}, \mathbf{z}) \\ &= C_1 exp\{-\frac{(\beta - \nu_0)^2}{2\tau_0^2}\} \prod_{i=1}^{m} \frac{exp(\beta z_{(i)})}{\sum_{u \in R(t_{(i)})} exp(\beta z_u)}, \end{aligned} \tag{6.47}$$

where C_1 is a constant such that $\int w(\beta|\mathbf{t}, \mathbf{z}) d\beta = 1$.

Consider a one interim analysis clinical trial. Suppose the trial is monitored at total number of failures m_1 and $m_1 + m_2$. At the interim analysis, after observing $(\mathbf{t},\ \mathbf{z}) = ((t_1,\ z_1),\ (t_2,\ z_2), ..., (t_{m_1},\ z_{m_1}))$, the risk from stopping the trial is,

$$r_0(w^1, 1) = \min\{E_{\beta|\mathbf{t}} L(\beta, d_s),\ \ E_{\beta|\mathbf{t}} L(\beta, d_e)\},$$

where,

$$\begin{aligned} E_{\beta|\mathbf{t}} L(\beta, d_s) &= K \int_{-\infty}^{\beta_1} C_1 exp\{-\frac{(\beta - \nu_0)^2}{2\tau_0^2}\} \prod_{i=1}^{m_1} \frac{exp(\beta z_{(i)})}{\sum_{u \in R(t_{(i)})} exp(\beta z_u)} d\beta, \\ E_{\beta|\mathbf{t}} L(\beta, d_e) &= K \int_{\beta_2}^{\infty} C_1 exp\{-\frac{(\beta - \nu_0)^2}{2\tau_0^2}\} \prod_{i=1}^{m_1} \frac{exp(\beta z_{(i)})}{\sum_{u \in R(t_{(i)})} exp(\beta z_u)} d\beta. \end{aligned}$$

Let $s_1^s < ... < s_{m_{2s}}^s$ be the next group of observed survival time from the standard treatment, and $s_1^e < ... < s_{m_{2e}}^e$ be the next group of observed survival time from the experimental treatment with $m_2 = m_{2s} + m_{2e}$. The predictive density function

is,

$$
\begin{aligned}
& f(\mathbf{s}|\mathbf{t},\mathbf{z}) \\
& = E_{\beta|\mathbf{t},\mathbf{z}} f(\mathbf{s}|\beta) \\
& = C_2 \int_{-\infty}^{\infty} exp\{-\tfrac{(\beta-\nu_0)^2}{2\tau_0^2}\} \prod_{i=1}^{m_1} \tfrac{exp(\beta z_{(i)})}{\sum_{h\in R(t_{(i)})} exp(\beta z_h)} \prod_{u=1}^{m_{2s}} \tfrac{1}{\sum_{h\in R(s_u^s)} 1} \prod_{v=1}^{m_{2e}} \tfrac{exp\beta}{\sum_{h\in R(s_v^e)} exp\beta} d\beta.
\end{aligned}
$$

The expected risk from continuing the trial is

$$
\begin{aligned}
& E^*_{\mathbf{s}|\mathbf{t},\mathbf{z}} \min\{E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_s),\ E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_e)\} \\
& = \int_{\mathbf{s}} \min\{E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_s),\ E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_e)\} f(\mathbf{s}|\mathbf{t},\mathbf{z}) d\mathbf{s} \\
& = \int \min\{E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_s), E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_e)\} f(\mathbf{s}|\mathbf{t},\mathbf{z}) ds_1^s ... ds_{m_{2e}}^s ds_1^e ... ds_{m_{2s}}^e, \quad (6.48) \\
& s_1^s < ... < s_{m_{2e}}^s \\
& s_1^e < ... < s_{m_{2s}}^e
\end{aligned}
$$

where

$$
\begin{aligned}
& E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_s) \\
& = C_3 \int_{-\infty}^{\beta_1} exp\{-\tfrac{(\beta-\nu_0)^2}{2\tau_0^2}\} \prod_{i=1}^{m_1} \tfrac{exp(\beta z_{(i)})}{\sum_{h\in R(t_{(i)})} exp(\beta z_h)} \prod_{u=1}^{m_{2s}} \tfrac{1}{\sum_{h\in R(s_u^s)} 1} \prod_{v=1}^{m_{2e}} \tfrac{exp\beta}{\sum_{h\in R(s_v^e)} exp\beta}, \\
& E_{\beta|\mathbf{t},\mathbf{z};\mathbf{s}} L(\beta, d_e) \\
& = C_3 \int_{\beta_2}^{\infty} exp\{-\tfrac{(\beta-\nu_0)^2}{2\tau_0^2}\} \prod_{i=1}^{m_1} \tfrac{exp(\beta z_{(i)})}{\sum_{h\in R(t_{(i)})} exp(\beta z_h)} \prod_{u=1}^{m_{2s}} \tfrac{1}{\sum_{h\in R(s_u^s)} 1} \prod_{v=1}^{m_{2e}} \tfrac{exp\beta}{\sum_{h\in R(s_v^e)} exp\beta}.
\end{aligned}
$$

Since the risk sets $R(s_u^s)$ and $R(s_v^e)$ involve the data of all failures, then the computation of the expected risk in (6.48) is multiple integration. It is impractical to compute it and also difficult to get it by simulations.

6.4.4 Method 3. Log-rank statistics

Suppose the clinical trial is monitored at times u_1, u_2, ..., u_l. Let $V(u_j)$, $j = 1, 2, ..., l$, denote the value of log rank statistics computed at time u_j from the proportional hazard model (6.43).

Tsiatis(1981, 1982) has derived that the asymptotic joint distribution of log rank statistics at different time points u_j, $j = 1, 2, ..., l$, is the same as the distribution of a sequence of partial sums of independent normal random variables. That is,

$$V(u_j) \approx X_1 + X_2 + ... + X_j,$$

where X_1, X_2,...,X_j are independent normal random variables with

$$\begin{aligned} E(X_j) &= \beta(d(u_j) - d(u_{j-1}))p(1-p), \\ Var(X_j) &= (d(u_j) - d(u_{j-1}))p(1-p). \end{aligned}$$

The p is the proportion of failures on one of the treatments. At each analysis j, $j = 1, 2, ..., l$, let

$$\begin{aligned} Y_j &= \frac{X_j}{[d(u_j) - d(u_{j-1})]p(1-p)}, \\ Y_j &\dot{\sim} N\left(\beta, \frac{1}{[d(u_j) - d(u_{j-1})]p(1-p)}\right). \end{aligned}$$

The original group sequential data may be replaced by the sequence of random variables $Y_j, j = 1, 2, ..., l$ in Bayes sequential decision procedure.

The Bayes sequential decision procedure: one interim analysis

Suppose the clinical trial is monitored at **I)** u_1 and u_2. At time u_1, number of failures $d(u_1) = d^s(u_1) + d^e(u_1)$ have been observed, the log rank statistics is $V(u_1)$. The value of Y_1 is

$$Y_1 = y_1 = \frac{V(u_1)}{d(u_1)p(1-p)} \dot{\sim} N(\beta, \frac{1}{d(u_1)p(1-p)}),$$

where $p = \frac{d^s(u_1)}{d(u_1)}$.

Using Bayes theorem, the posterior distribution of β given $Y_1 = y_1$ is a normal distribution, denoted by $N(\nu_1, \tau_1^2)$. The risk from stopping the trial is

$$r_0(w^1, 1) = \min\{E_{\beta|y_1} L(\beta, d_s),\ E_{\beta|y_1} L(\beta, d_e)\},$$

where

$$\begin{aligned} E_{\beta|U(c_1)} L(\beta, d_s) &= K \int_{-\infty}^{\beta_1} \frac{1}{\sqrt{2\pi}\tau_1} e^{-\frac{(\beta-\nu_1)^2}{2\tau_1^2}} d\beta, \\ E_{\beta|U(c_1)} L(\beta, d_e) &= K \int_{\beta_2}^{\infty} \frac{1}{\sqrt{2\pi}\tau_1} e^{-\frac{(\beta-\nu_1)^2}{2\tau_1^2}} d\beta. \end{aligned}$$

On the other hand, the predictive distribution of Y with the number of new failures $m = d(u_2) - d(u_1)$ given $Y_1 = y_1$ is,

$$\begin{aligned} f(y|y_1) &= E_{\beta|y_1} f(y|\beta) \\ &= \int_{-\infty}^{\infty} \sqrt{\frac{(d(u_2) - d(u_1))p(1-p)}{2\pi}} e^{-\frac{(d(u_2)-d(u_1))p(1-p)(y-\beta)^2}{2}} \frac{1}{\sqrt{2\pi}\tau_1} e^{-\frac{(\beta-\nu_1)^2}{2\tau_1^2}} d\beta. \\ &= N(\nu_1,\ \tau_1^2 + \frac{1}{(d(u_2) - d(u_1))p(1-p)}) \end{aligned}$$

The expected risk from continuing the trial given the number of new failures $m = d(u_2) - d(u_1)$ is

$$E^*_{|m} r_0(w^1(Y), 2) = \int_{-\infty}^{\infty} r_0(w^1(Y = y), 2) f(y|y_1) dy,$$

where,

$$r_0(w^1(Y = y), 2) = \min\{E_{\beta|y_1,y} L(\beta, d_s),\ E_{\beta|y_1,y} L(\beta, d_e)\}$$

Let $g(m)$ be the distribution of m. The expected risk from continuing the trial is,

$$E^* r_0(w^1(Y), 2) \;=\; \sum_m E^*_{|m} r_0(w^1(Y), 2)\; g(m).$$

If the risk from stopping the trial, $r_0(w^1, 1)$, is less than the risk from continuing the trial, $E^* r_0(w^1(X), 2) + R \min\{u_2 - u_1, s_a - u_1, 0\}$, then the trial is terminated. Otherwise the trial is continued to the final analysis.

Suppose the clinical trial is monitored **II)** at total number of new failures m_1, m_2. In the Bayes sequential decision procedure, the risk from stopping the trial is obtained in the same way as in case I); the expected risk from continuing the trial is $E^* r_0(w^1(X), 2) + R \min\{E(u) - u_1, s_a - u_1, 0\}$, where $E(u)$ is the expected monitoring time which is corresponding to having total number of deaths equal to $m_1 + m_2$.

Hence, if the number of failures are large enough, the group sequential sample from the proportional hazard model (6.43) may be replaced by the log-rank statistics calculated at each analysis. The Bayes sequential decision procedure in clinical trials with normal distribution responses in Chapter 4 may be used.

6.4.5 Conclusion

Bayes sequential decision theory was applied to the clinical trials with survival time data, where the survival time is the proportional hazard model. As a nuisance parameter, the cumulative hazard function of the proportional hazard model was assumed to have a prior beta process from the idea of Hjort(1990) and a gamma process from Kalbfleisch(1978). Since we are interested in the sequential clinical trials, there were computational problems of obtaining the Bayes stopping risk and the expected risk from continuing the trial, which was discussed in Section 6.4.2. The Cox partial likelihood doesn't involve the cumulative hazard function, but there was still difficulty to get the expected risk from continuing the trial

which was described in Section 6.4.3. However, the asymptotic joint distribution of log-rank statistics at different time points is the same as the distribution of a sequence of partial sums of independent normal random variable. Therefore, the Bayes sequential decision procedure in clinical trials with normal distribution responses can be used as an approximation. This was discussed in Section 6.4.4. The discussion on the non-parametric Bayes analysis of proportional hazard model is used in the estimation of treatment difference after the trial is terminated. This is also used in Chapter 2 for clinical trials with proportional hazard model for survival time.

Chapter 7

Robust Bayes Analysis in Clinical Trials

Bayesian methods in clinical trials have received increasing attention recently as they offer an approach for dealing with many difficult problems which arise in practice. This was described in Chapter 1. A major criticism of Bayes analysis is that it presumes an ability to completely and accurately elicit subjective information in terms of a single prior distribution for parameters of interest. Investigators are concerned that inferences based on a posterior distribution will be sensitive to the specification of the prior distribution. However, there has long existed a robust Bayesian viewpoint to address this criticism. Good(1959, 1961, 1962) has started the robust Bayesian approach to inference. The robust Bayesian approach replaces the single prior distribution with a *class* of possible prior distributions. The goal of this approach is to make inferences or decisions which are robust within this class, that is, the inferences or decisions are relatively insensitive (or at least are satisfactory) to deviations as the prior distribution varies over this class. There are reviews by Berger(1984, 1985, 1990, and 1994) and Wasserman(1992). Greenhouse and Wasserman(1995) have illustrated the application of robust Bayes methods to the analysis of clinical trials.

Although a complete robust Bayes analysis would consider the sensitivity of posterior inferences not only to the specification of the prior distribution but also to the specification of the likelihood function, in this chapter only the sensitivity of posterior inferences to the prior distribution is discussed.

The framework of robust Bayes analysis is described in Section 7.1. Following this, the applications of robust Bayes analysis to clinical trials are discussed. The robust Bayes analysis of clinical trials with main outcome variable from normal distribution is considered in Section 7.2. Section 7.3 discusses the robust analysis of clinical trials with main outcome variable from binomial distribution. Chapter 6 has shown that analysis of clinical trials with survival time data could be approximated by analysis of clinical trials with normal distribution response. Therefore the robust Bayes analysis of clinical trials with survival time data is not discussed here.

7.1 The Framework of Robust Bayes Analysis

7.1.1 The $\varepsilon-$contamination class

To formulate a class of possible prior distributions for a parameter of interest has not been resolved completely in the Bayesian robust approach. The $\varepsilon-$contamination class is used as the class of possible prior distributions in the robust Bayes analysis here. This class has been considered by many people, including Huber(1973), Berger(1984), Berger and Berliner(1986), Lavine et al(1991) and Greenhouse and Wasserman(1995).

Let X denote the observable random variable or random vector $X = (X_1, X_2, ..., X_n)$ which is assumed to have a distribution function $f(x|\theta)$, where θ is an unknown parameter lying in a parameter space Θ. Assume that θ has the prior probability density function $w_0 = w_0(\theta)$ on Θ. The posterior probability density function of θ given $X = x$ (assuming it exists) is denoted by $w_0(\theta|x)$. To be more precise,

we should work with probability distributions not densities because the classes of priors we deal with do not necessarily have densities, but for simplicity, we will write expressions in terms of density functions. Let $\mathcal{P}$ denote the space of all probability distributions on Θ. The $\varepsilon-$contamination class is defined as

$$\Gamma_\varepsilon = \{w : w = (1-\varepsilon)w_0 + \varepsilon q,\ q \in \mathcal{Q}\}, \tag{7.1}$$

where $0 \leq \varepsilon \leq 1$ represents the amount of uncertainty relating the accuracy of the specified prior distribution w_0, the value $(1-\varepsilon)$ reflects degree of confidence in the accuracy of the prior distribution w_0, and $\mathcal{Q}$ is a class of reasonable alternative priors which is some subset of $\mathcal{P}$.

Berger and Berliner(1986) have given several reasons for consideration of this class Γ_ε, and suggested that the class of contaminations $\mathcal{Q}$ may be

I) $\mathcal{Q} = \mathcal{P}$(all probability distributions on Θ);

II) $\mathcal{Q}$={all symmetric and unimodal distributions on Θ };

III) the class of contaminations such that the resulting w is unimodal(assuming that w_0 is unimodal);

IV) $\mathcal{Q}$= { mixtures of various classes }.

The $\varepsilon-$contamination class Γ_ε in (7.1) with $\mathcal{Q} = \mathcal{P}$ is computationally easy to work with. Although this class is large and contains many more distributions than we would consider reasonable in practice, if the posterior distribution is not sensitive to this class, then we should have considerable confidence in the posterior robustness to the prior w_0.

7.1.2 The framework

Let $g(\theta)$ be some quantity of interest. The upper and lower bounds on the posterior expectation of $g(\theta)$ over the class Γ_ε defined by (7.1) are

$$\sup_{w\in\Gamma_\varepsilon} E_w(g(\theta)|x) = \sup_{w\in\Gamma_\varepsilon} \int_\Theta g(\theta)w(\theta|x)d\theta, \tag{7.2}$$

$$\inf_{w\in\Gamma_\varepsilon} E_w(g(\theta)|x) = \inf_{w\in\Gamma_\varepsilon} \int_\Theta g(\theta)w(\theta|x)d\theta. \tag{7.3}$$

If the differences $\sup_{w\in\Gamma_\varepsilon} E_w(g(\theta)|x) - \inf_{w\in\Gamma_\varepsilon} E_w(g(\theta)|x)$ are small for some values of $\varepsilon(0 \leq \varepsilon \leq 1)$, that is, the inferences or decisions based on the posterior expectation $E_w(g(\theta)|x)$ do not change as the prior varies over this class Γ_ε, it may be concluded that the inferences for $g(\theta)$ are not sensitive to the specified prior distribution w_0.

When $g(\theta)$ is the indicator function for a measurable set $A \subseteq \Theta$, then the posterior expectation $E_w(g(\theta)|x) = P_w(\theta \in A|x)$, which is the posterior probability of A after observing $X = x$. Corresponding to Chapter 2, let the $\theta \in \Theta = \mathcal{R}^1$(real line) be the measure of treatment difference in a clinical trial comparing two treatments and let $A = \{\theta,\ \theta \leq \theta_1\}$ (or $\{\theta,\ \theta \geq \theta_2\}$), where the interval $(\theta_1,\ \theta_2)(\theta_1 \leq \theta_2)$ is the range of equivalence of the clinical trial. When the posterior probability $P_{w_0}(\theta \in A|x) = P_{w_0}(\theta \leq \theta_1|x)$ (or $P_{w_0}(\theta \geq \theta_2|x)$) is big, the clinical trial is suggested being stopped early as discussed in Chapter 2. If the difference $\sup_{w\in\Gamma_\varepsilon} P_w(\theta \in A|x) - \inf_{w\in\Gamma_\varepsilon} P_w(\theta \in A|x)$ is small, in other words, the decisions based on posterior probabilities $P_w(\theta \in A|x)$ over this class Γ_ε are same, then we do not need to worry about the choice of the single "correct" prior distribution w_0 and can be concluded that the decision of stopping the clinical trial early is robust to the prior w_0.

From a mathematical perspective, the calculations $\sup_{w\in\Gamma_\varepsilon} E_w(g(\theta)|x)$ in (7.2) and $\inf_{w\in\Gamma_\varepsilon} E_w(g(\theta)|x)$ in (7.3) are very complicated. Huber(1973) has derived the calculation of upper and lower bounds of $P_w(\theta \in A|x)$ over the class Γ_ε of

(7.1) with $\mathcal{Q} = \mathcal{P}$.

Huber Theorem Assume that the random variable X has the density function $f(x|\theta)$, and the parameter of interest $\theta \in \Theta$ has the prior distribution w_0. Let A be a measurable subset of Θ, and β_0 be the posterior probability of A under w_0, that is, $\beta_0 = P_{w_0}(\theta \in A|X = x)$. It is obtained that

$$\inf_{w \in \Gamma_\varepsilon} P(\theta \in A|X = x) = \beta_0 \left\{ 1 + \frac{\varepsilon \sup_{\theta \notin A} f(x|\theta)}{(1-\varepsilon)m(x|w_0)} \right\}^{-1}, \tag{7.4}$$

$$\sup_{w \in \Gamma_\varepsilon} P(\theta \in A|X = x) = \frac{(1-\varepsilon)m(x|w_0)\beta_0 + \varepsilon \sup_{\theta \in A} f(x|\theta)}{(1-\varepsilon)m(x|w_0) + \varepsilon \sup_{\theta \in A} f(x|\theta)}, \tag{7.5}$$

where $m(x|w_0) = \int_\Theta f(x|\theta)w_0(\theta)d\theta$ is the marginal density of x under the prior w_0, and Γ_ε is the $\varepsilon-$contamination class defined as in (7.1) with $\mathcal{Q} = \mathcal{P}$.

7.2 Clinical Trials with Normal Distribution Response

Consider a clinical trial comparing two treatments. The main outcome variable is normally distributed with variance known and mean μ_i, $i = 1, 2$, for each treatment respectively. Let $\theta = \mu_1 - \mu_2$. The parameter θ is a measure of difference of these two treatments. The comparison of these two treatments could be based on statistical inferences of the parameter θ. Let random variable X be the test statistic of the difference θ. The X is from the normal distribution with mean equal to θ and variance known, denoted by σ^2, that is, $X \sim N(\theta, \sigma^2)$.

Assume that the prior distribution of θ, $w_0(\theta)$, is the normal distribution with mean μ and variance τ^2, that is, $\theta \sim w_0(\theta) = N(\mu, \tau^2)$. The posterior distribution of θ given $X = x$ is,

$$w_0(\theta|x) = N(\delta(x), V^2),$$

where $\delta(x) = x - \frac{\sigma^2}{\sigma^2+\tau^2}(x-\mu)$, $V^2 = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}$.

The usual $100(1-\alpha)\%$ Bayes credible region for θ, denoted by A, is

$$A = \{\theta : \delta(x) - K < \theta < \delta(x) + K\}, \tag{7.6}$$

where $K = z_{\alpha/2}V$, and $z_{\alpha/2}$ is the $100(1-\alpha/2)$ upper percentile of the standard normal distribution. The robustness of posterior probability $P(\theta \in A|x)$ to the specified prior distribution $w_0(\theta) = N(\mu, \tau^2)$ is discussed below.

For the simplicity of computation, only the class of contaminations $\mathcal{Q} = \mathcal{P}$(all distributions on Θ) and $\mathcal{Q} = \{$all symmetric and unimodal distributions on $\Theta\}$ in (7.1) are considered in Section 7.2.1 and Section 7.2.2, respectively.

7.2.1 Arbitrary contaminations $\mathcal{Q} = \mathcal{P}$

Corresponding to the definition of $\varepsilon-$contamination class described in Section 7.1.1, the ε-contamination class of prior distributions for the difference of these two treatments θ in the clinical trial is,

$$\Gamma_\varepsilon = \{w : w = (1-\varepsilon)N(\mu, \tau^2) + \varepsilon q,\ q \in \mathcal{P}\}. \tag{7.7}$$

The lower and upper bounds of $P_w(\theta \in A|X = x)$ over this class Γ_ε may be calculated by the Huber theorem.

Consider the example in Berger and Berliner(1986). Suppose $\sigma^2 = 1$, $\tau^2 = 2$, $\mu = 0$, and $\varepsilon = 0.2$. If $x = 0.5$ is observed, then 95% Bayes credible interval for θ is $A = (-1.27, 1.93)$, and

$$\begin{aligned} \inf_{w\in\Gamma_{0.2}} P_w(-1.27 < \theta < 1.93|X = 0.5) &= 0.817, \\ \sup_{w\in\Gamma_{0.2}} P_w(-1.27 < \theta < 1.93|X = 0.5) &= 0.966. \end{aligned}$$

The above results show that the difference between the lower and upper bounds

of the posterior probability $P_w(-1.27 < \theta < 1.93|X = 0.5)$ is small. The posterior probability of this standard credible set is reasonably robust to the prior distribution of θ, which is $w_0(\theta) = N(0, 2)$.

If $x = 4$ is observed, then the usual 95% credible set is $(1.07, 4.27)$, and it is obtained that

$$\begin{aligned} \inf_{w \in \Gamma_1} P_w(1.07 < \theta < 4.27|X = 4) &= 0.1355, \\ \sup_{w \in \Gamma_1} P_w(1.07 < \theta < 4.27|X = 4) &= 0.99. \end{aligned}$$

Since the posterior probability can get as low as 0.1355, and as high as 0.99 for $x = 4$, the procedure is not robust to the prior distribution $N(0, 2)$ with respect to the class Γ_ε defined by (7.7).

The example has shown that the robustness with respect to Γ_ε depends significantly on the x observed. A lack of robustness may be due to the fact that the $\varepsilon-$contamination class Γ_ε in (7.7) with arbitrary contaminations $\mathcal{Q} = \mathcal{P}$ is "too large".

7.2.2 Symmetric unimodal contaminations

Let $\Theta \subseteq \mathcal{R}^1$(real line). The class of symmetric unimodal contaminations can be expressed by $\mathcal{Q} = \{$densities of the form $q(|\theta - \theta_0|)$, q nonincreasing$\}$, where $\theta = \theta_0$ is the symmetric axis of density q. This class may be approximated by $\mathcal{Q}' = \{$Uniform $(\theta_0 - a, \theta_0 + a)$ densities, $a \geq 0\}$ as a considerable simplicity to work with (Berger and Berliner 1986). The ε-contamination class of prior distributions of θ in the study is then

$$\Gamma_\varepsilon = \{w : w = (1 - \varepsilon)N(\mu, \tau^2) + \varepsilon U(\mu - a, \mu + a), a \geq 0\}. \tag{7.8}$$

The expression of a prior distribution from this class Γ_ε in (7.8) can be derived as follows.

Suppose θ has the prior distribution from this class Γ_ε defined by (7.8), then $\theta = (1-\varepsilon)U_1 + \varepsilon U_2$, where U_1 and U_2 are independent with $U_1 \sim N(\mu, \tau^2)$ and $U_2 \sim U(\theta_0 - a, \theta_0 + a)$. The probability distribution function of θ is,

$$\begin{aligned} P(\theta < b) &= P((1-\varepsilon)U_1 + \varepsilon U_2 < b) \\ &= \iint_{\substack{u_2 \in (\theta_0 - a, \theta_0 + a) \\ (1-\varepsilon)u_1 + \varepsilon u_2 < b}} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{(u_1-\mu)^2}{2\tau^2}} \frac{1}{2a}\, du_1 du_2 \\ &= \frac{1}{2a} \int_{\theta_0 - a}^{\theta_0 + a} \int_{-\infty}^{\frac{b - \varepsilon u_2}{1-\varepsilon}} \frac{1}{\sqrt{2\pi}\tau} e^{-\frac{(u_1-\mu)^2}{2\tau^2}}\, du_1 du_2. \end{aligned}$$

The prior density function of θ is then,

$$w(\theta) = \frac{1}{2a\tau(1-\varepsilon)} \int_{\theta_0 - a}^{\theta_0 + a} \phi\left(\frac{\frac{\theta - \varepsilon u_2}{1-\varepsilon} - \mu}{\tau}\right) du_2. \tag{7.9}$$

After observing $X = x$, the posterior probability density function of θ is,

$$\begin{aligned} w(\theta|x) &= \frac{f(x|\theta)w(\theta)}{\int_{-\infty}^{\infty} f(x|\theta)w(\theta)d\theta} \\ &= \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \frac{1}{2a\tau(1-\varepsilon)} \int_{\theta_0 - a}^{\theta_0 + a} \phi\left(\frac{\frac{\theta - \varepsilon u_2}{1-\varepsilon} - \mu}{\tau}\right) du_2}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \frac{1}{2a\tau(1-\varepsilon)} \int_{\theta_0 - a}^{\theta_0 + a} \phi\left(\frac{\frac{\theta - \varepsilon u_2}{1-\varepsilon} - \mu}{\tau}\right) du_2 d\theta} \\ &= \frac{e^{-\frac{(x-\theta)^2}{2\sigma^2}} \int_{\theta_0 - a}^{\theta_0 + a} e^{-\frac{1}{2}\left(\frac{\frac{\theta - \varepsilon u_2}{1-\varepsilon} - \mu}{\tau}\right)^2} du_2}{\int_{-\infty}^{\infty} e^{-\frac{(x-\theta)^2}{2\sigma^2}} \int_{\theta_0 - a}^{\theta_0 + a} e^{-\frac{1}{2}\left(\frac{\frac{\theta - \varepsilon u_2}{1-\varepsilon} - \mu}{\tau}\right)^2} du_2 d\theta}. \end{aligned} \tag{7.10}$$

The lower and upper bounds of $P_w(\theta \in A|X = x)$ over the class Γ_ε in (7.8) are

$$\inf_{w \in \Gamma_\varepsilon} P_w(\theta \in A|X = x) = \inf_{a \geq 0} \int_A w(\theta|x)d\theta, \tag{7.11}$$

$$\sup_{w \in \Gamma_\varepsilon} P_w(\theta \in A|X = x) = \sup_{a \geq 0} \int_A w(\theta|x)d\theta, \tag{7.12}$$

where the posterior probability density function $w(\theta|x)$ is obtained by (7.10).

If the difference $\sup_{w\in\Gamma_\varepsilon} P_w(\theta \in A|X = x) - \inf_{w\in\Gamma_\varepsilon} P_w(\theta \in A|X = x)$ are small for some values of $\varepsilon(0 \leq \varepsilon \leq 1)$, then it may be concluded that the inference on the posterior probability $P(\theta \in A|x)$ is robust to the prior with respect to the $\varepsilon-$contamination class Γ_ε in (7.8).

Continuing the example in Section 7.2.1, it was shown that the posterior probability of 95% credible set at $x = 4$ is not robust to the prior with respect to the $\varepsilon-$contamination class Γ_ε defined by (7.7) with $\varepsilon = 0.2$. Consider the $\varepsilon-$contamination class defined as in (7.8) with $\mathcal{Q}' = \{$ Uniform(0 - a, 0 + a) densities, $a \leq 5\}$. At $x = 4$,

$$\begin{aligned} \inf_{w\in\Gamma_{0.2}} P_w(1.07 < \theta < 4.27|X = 4) &= 0.139, \\ \sup_{w\in\Gamma_{0.2}} P_w(1.07 < \theta < 4.27|X = 4) &= 0.218. \end{aligned}$$

The difference of these upper and lower bounds is small over this class Γ_ε, which shows that the posterior probability $P_{w_0}(1.07 < \theta < 4.27|X = 4)$ is much less sensitive over the class Γ_ε defined by (7.8) than over the class Γ_ε defined by (7.7). The example shows that it is worthwhile to check the robustness with respect to some reasonable subset of $\mathcal{P}$ when a procedure is not robust with respect to the $\varepsilon-$contamination class with the arbitrary contamination class $\mathcal{P}$.

7.3 Clinical Trials with Binomial Distribution Response

7.3.1 One Sample

Consider a clinical trial with main outcome variable X from the binomial distribution $B(n,p)$. The rate p is the parameter of interest. Greenhouse and Wasserman(1995) have discussed the robustness of $P(p > p_0|x)$ to the prior with respect to the following ε-contamination class Γ_ε defined by (7.13), in which p_0 is the break-even value of p.

$$\Gamma_\varepsilon = \{w : w = (1-\varepsilon)Beta(\nu_1,\nu_2) + \varepsilon q, q \in \mathcal{P}\}, \tag{7.13}$$

where they have assumed that the p has the beta prior distribution $w_0(p) = Beta(\nu_1,\nu_2)$.

However, sometimes it may be suggested that the $\theta = \log\frac{p}{1-p}$ has the normal prior distribution $w_0(\theta) = N(\mu,\tau^2)$. The robustness of $P_w(\theta \leq \theta_0|X = x)$(or $P_w(\theta \geq \theta_0|X = x)$ to this prior is discussed below, where $\theta_0 = \log\frac{p_0}{1-p_0}$ is the break-even value of θ.

Assume that the ε-contamination class with arbitrary contamination class is used and that it is defined as Γ_ε in (7.14).

$$\Gamma_\varepsilon = \{w : w = (1-\varepsilon)N(\mu,\tau^2) + \varepsilon q, q \in \mathcal{P}\}. \tag{7.14}$$

The $\inf_{w\in\Gamma_\varepsilon} P_w(\theta \leq \theta_0|X = x)$ and $\sup_{w\in\Gamma_\varepsilon} P_w(\theta \leq \theta_0|X = x)$ are calculated by the Huber theorem. Corresponding to the Huber theorem, the density function $f(x|\theta)$, $\sup_{\theta>\theta_0} f(x|\theta)$, $\sup_{\theta\leq\theta_0} f(x|\theta)$ and the marginal density function $m(x|w_0)$

under the prior distribution $w_0(\theta) = N(\mu, \tau^2)$ are,

$$
\begin{aligned}
f(x|\theta) &= \binom{n}{x}\left(\frac{e^\theta}{1+e^\theta}\right)^x\left(1-\frac{e^\theta}{1+e^\theta}\right)^{n-x}, \\
\sup_{\theta>\theta_0} f(x|\theta) &= \begin{cases} \binom{n}{x} p_0^x(1-p_0)^{n-x} & x \le np_0 \\ \binom{n}{x}(\frac{x}{n})^x(1-\frac{x}{n})^{n-x} & x > np_0, \end{cases} \\
\sup_{\theta\le\theta_0} f(x|\theta) &= \begin{cases} \binom{n}{x}(\frac{x}{n})^x(1-\frac{x}{n})^{n-x} & x \le np_0 \\ \binom{n}{x} p_0^x(1-p_0)^{n-x} & x > np_0, \end{cases} \\
m(x|w_0) &= \int_{-\infty}^{\infty}\binom{n}{x}\left(\frac{e^\theta}{1+e^\theta}\right)^x\left(1-\frac{e^\theta}{1+e^\theta}\right)^{n-x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}d\theta.
\end{aligned}
$$

If $\sup_{w\in\Gamma_\varepsilon} P_w(\theta \le \theta_0|X=x) - \inf_{w\in\Gamma_\varepsilon} P_w(\theta \le \theta_0|X=x)$ are small for some values of $\varepsilon(0 \le \varepsilon \le 1)$, then the posterior probability $P(\theta \le \theta_0|X=x)$ is robust to the prior w_0.

Consider the example given by Greenhouse and Wasserman(1995). A clinical trial is designed to test the effect of a new treatment. If the failure rate of the new treatment p is greater than $p_0 = 0.2$, then the new treatment is not accepted. They assumed that p had the beta prior distribution $w_0(p) = Beta(1.56, 8.44)$ with mean equal to 0.16 and variance equal to 0.012. When three of the first four patients in the trial failed, it was obtained that the lower and upper bounds of the posterior probability $P(p > 0.2|x)$ over the class Γ_ε in (7.13) at $\varepsilon = 0.2$ were

$$
\begin{aligned}
&\inf_{w\in\Gamma_{0.2}} P_w(p > 0.2|x) = 0.67, \\
&\sup_{w\in\Gamma_{0.2}} P_w(p > 0.2|x) = 0.98;
\end{aligned}
$$

and at $\varepsilon = 0.5$ were

$$
\begin{aligned}
&\inf_{w\in\Gamma_{0.5}} P_w(p > 0.2|x) = 0.41, \\
&\sup_{w\in\Gamma_{0.5}} P_w(p > 0.2|x) = 0.99.
\end{aligned}
$$

For $\varepsilon = 0.2$, where there was 80% belief in the specified prior distribution $w_0(p) = Beta(1.56, 8.44)$, the posterior probability that the failure rate was greater than 20% was fairly large, which was between 0.67 and 0.98. For $\varepsilon = 0.5$, where the belief in the prior $w_0(p) = Beta(1.56, 8.44)$ and the class of all other priors q was equally split, this posterior probability was still large, which was from 0.41 to 0.99. They also considered the expected failure rate $E_w(p|x)$. For $\varepsilon = 0.5$, the bounds for $E_w(p|x)$ with respect to the $\varepsilon-$contamination class Γ_ε defined by (7.13) were 0.30 and 0.80 respectively. Based on these analyses, they stated that they had considerable confidence in supporting a recommendation to stop the trial and not to accept the new treatment.

Alternatively, suppose that the $\theta = \log \frac{p}{1-p}$ has the normal prior distribution $w_0(\theta) = N(\mu, \tau^2)$ and that the $\varepsilon-$contamination class Γ_ε defined by (7.14) is used. The mean and variance of the normal prior distribution are equal to $\log \frac{0.16}{1-0.16}$ and $(\frac{1}{0.16} + \frac{1}{1-0.16})^2 \times 0.012$ respectively, which are approximatedly equal to the mean and variance of the beta prior distribution $Beta(1.56, 8.44)$, respectively, by the Taylor expansions. The corresponding lower and upper bounds of $P_w(p > 0.2|x) = P_w(\theta > \log \frac{0.2}{1-0.2}|x)$ over the class of Γ_ε in (7.14) at $\varepsilon = 0.2$ are,

$$\inf_{w \in \Gamma_{0.2}} P_w(\theta > \log \frac{0.2}{1-0.2}|x) = 0.755,$$
$$\sup_{w \in \Gamma_{0.2}} P_w(\theta > \log \frac{0.2}{1-0.2}|x) = 0.967;$$

and at $\varepsilon = 0.5$ are,

$$\inf_{w \in \Gamma_{0.5}} P_w(\theta > \log \frac{0.2}{1-0.2}|x) = 0.532,$$
$$\sup_{w \in \Gamma_{0.5}} P_w(\theta > \log \frac{0.2}{1-0.2}|x) = 0.990.$$

The differences of upper and lower bounds of the posterior probability at $\varepsilon = 0.2$ and 0.5 are 0.212 and 0.458, respectively, over the class Γ_ε in (7.14), while

differences are 0.31 and 0.58 at $\varepsilon = 0.2$ and 0.5, respectively, over the class Γ_ε in (7.13). The example shows that posterior inferences on $\theta = \log \frac{p}{1-p}$ to the normal prior sometimes are more robust than those of p to the beta prior for binomial response variables.

7.3.2 Two Samples

Consider a clinical trial comparing two treatments A and B with n patients in each treatment. The main outcome variable of the clinical trial is from the binomial distribution $X_A \sim B(n, p_A)$ and $X_B \sim B(n, p_B)$ for treatments A and B, respectively. Assume that the p_A and p_B are independent and have beta prior distributions $p_A \sim Beta(\nu_{1A}, \nu_{2A})$ and $p_B \sim Beta(\nu_{1B}, \nu_{2B})$. Let $\delta = \frac{p_A}{p_B} \in \Theta = (0, \infty)$ be the measure of the treatment difference. Assume that p_A and p_B are successful rates of treatments A and B, respectively, and that δ_1 and δ_2 are break-even values of δ and the interval $(\delta_1,\ \delta_2)(\delta_1 \leq \delta_2)$ is the range of equivalence. If $\delta \leq \delta_1$, then treatment B is better than treatment A; if $\delta \geq \delta_2$, then treatment A is better than treatment B. The prior distribution of δ is,

$$\begin{aligned} w_0(\delta) &= \int_0^1 p_B w_0(\delta p_B) w_0(p_B) dp_B \\ &= \int_0^1 \frac{p_B}{\beta(\nu_{1A}, \nu_{2A})} (\delta p_B)^{\nu_{1A}-1} (1-\delta p_B)^{\nu_{2A}-1} \\ &\quad \times \frac{1}{\beta(\nu_{1B}, \nu_{2B})} p_B^{\nu_{1B}-1} (1-p_B)^{\nu_{2B}-1} dp_B. \end{aligned} \tag{7.15}$$

After observing $X_A = x_A$ and $X_B = x_B$, before making inferences or decisions, it is important to know the robustness of $P(\delta \leq \delta_1 | X_A = x_A, X_B = x_B)$(or $P(\delta \geq \delta_2 | X_A = x_A, X_B = x_B)$) to the prior $w_0(\delta)$.

Since $f(x_A|p_A) f(x_B|p_B) = f(x_A, x_B|p_A, p_B) = f(x_A, x_B|\delta p_B, p_B)$, then the probability density function of (x_A, x_B) is,

$$f(x_A, x_B|\delta) = \int_0^1 f(x_A, x_B|\delta p_B, p_B) w_0(p_B) dp_B$$

$$= \int_0^1 \binom{n}{x_A} (\delta p_B)^{x_A}(1-\delta p_B)^{n-x_A} \binom{n}{x_B} p_B^{x_B}(1-p_B)^{n-x_B} \times \frac{1}{\beta(\nu_{1B},\nu_{2B})} p_B^{\nu_{1B}-1}(1-p_B)^{\nu_{2B}-1} dp_B. \quad (7.16)$$

Consider the below ε-contamination class with $\mathcal{Q} = \mathcal{P}$, defined by (7.17),

$$\Gamma_\varepsilon = \{w : w = (1-\varepsilon)w_0 + \varepsilon q,\ q \in \mathcal{P}\}, \quad (7.17)$$

where the prior distribution w_0 is defined by (7.15). The lower and upper bounds of $P_w(\delta \leq \delta_1 | X_A = x_A, X_B = x_B)$(or $P_w(\delta \geq \delta_2 | X_A = x_A, X_B = x_B)$) over this class Γ_ε can be calculated by the Huber theorem from the prior distribution $w_0(\delta)$ in (7.15), the density function $f(x_A, x_B|\delta)$ in (7.16) and the $\varepsilon-$contamination class Γ_ε in (7.17). Let $C = \{\delta,\ \delta \leq \delta_1\}$ (or $\{\delta,\ \delta \geq \delta_2\}$). The robustness is based on the difference

$$\sup_{w\in\Gamma_\varepsilon} P_w(\delta \in C | X_A = x_A, X_B = x_B) - \inf_{w\in\Gamma_\varepsilon} P_w(\delta \in C | X_A = x_A, X_B = x_B).$$

If the procedure above is not robust, it is worthwhile to study that the $\theta_A = \log \frac{p_A}{1-p_A}$ and $\theta_B = \log \frac{p_B}{1-p_B}$ have the normal prior distributions as the case of one sample clinical trial in Section 7.3.1. The treatment difference may be measured by $\eta = \frac{\theta_A}{\theta_B}$ since $\log \frac{p}{1-p}$ is a monotonic increasing function of p. The discussion on this robustness is then similar to the study of p_A and p_B with the beta prior distributions, which is to find the prior distribution $w_0(\eta)$ and density function $f(x_A, x_B|\eta)$ first and then calculate the lower and upper bounds over the corresponding class of possible prior distributions for η.

7.4 Discussion

The robust Bayes analysis methods applied to clinical trials were illustrated. For the simplicity of computation, only the types I)(all probability distributions for the

parameters of interest) and II) (all symmetric and unimodal distributions for the parameters of interest) of the class of contaminations Q for the $\varepsilon-$contamination class defined as in (7.1) were considered.

The idea of carrying out the robust Bayes analysis in clinical trials is straightforward, that is, to analyse the sensitivity of posterior inferences to the specified prior distribution for the parameters of interest. However, to implement the robust Bayes analysis, we need to use numerical integration to get $E(g(\theta)|x)$ and $m(x|w_0)$, and sometimes may need a numerical optimization routine to obtain the superma and infima. Although the Markov Chain Monte Carlo method is currently very popular to obtain integration, from a mathematical perspective, this method does not necessarily overcome the complications of computations involving the robust analysis.

For the analysis of sensitivity of posterior inferences to the prior in clinical trials, Spiegelhalter, Freedman, and Parmar(1994) have suggested considering a *community* of priors covering the perspectives of a range of individuals in clinical trials. This may encompass a "reference" prior intended to add as little as possible to the data and a "clinical" prior expressing reasonable opinions held by individuals or derived from overviews(meta-analyses) of similar studies. It is also useful to develop "off the shelf" priors corresponding to a formal expression of "sceptical" and "enthusiastic" belief- these may be thought to provide reasonable bounds to the community of priors. However, the robust Bayes methods do not mean to imply that the single prior Bayes approach is necessarily bad. It usually works very well.

The framework of robust Bayes analysis may be used in the sequential sampling, which is corresponding the sequential clinical trials. Assume that $\{X_i, i = 1, 2, ...\}$ is a sequential sample from the density function $f(x|\theta)$ and that the parameter of interest θ has the prior distribution w_0. Let Γ be the class of possible prior distributions of θ. A sequential robust Bayes analysis looks at the sequence

of differences $\sup_{w\in\Gamma} E_w(g(\theta)|x_i) - \inf_{w\in\Gamma} E_w(g(\theta)|x_i)$ one at a time after observing $X_i = x_i$, $i = 1, 2, ...$ If the sampling is terminated at analysis l and the sequence of the differences

$$\{\sup_{w\in\Gamma} E_w(g(\theta)|x_i) - \inf_{w\in\Gamma} E_w(g(\theta)|x_i),\ i = 1, 2, ..., l\}$$

are small, then it may be concluded that the sequential inferences for $g(\theta)$ are not sensitive to the prior. The robust Bayes analysis in sequential clinical trials would give us considerable confidence in suggesting that whether the trial should be stopped early or not. Greenhouse and Wasserman(1995) give a sequential robust Bayes analysis of an efficacy trial. It was the Harvard ECMO clinical trial, in which the treatment ECMO was compared with the treatment CMT. The probabilities of success on ECMO and CMT were denoted by p_E and p_C, respectively. They plotted bounds of $P(p_E > p_C|x_1, ..., x_i)$ and $E(p_E - p_C|x_1, ..., x_i)$ over the $\varepsilon-$contamination class after each patient $x_i, i = 1, 2, ..., 19$, for some values of ε(=0.1, 0.2, 0.3, 0.4). The plots showed that ECMO appeared consistently superior to CMT starting with the tenth patient for $\varepsilon = 0.1$ which was a small degree of uncertainty in the specified prior; however the evidence did not favor ECMO with sufficient strength to stop the trial early for slightly large values of ε.

The robust analysis has not been studied in Bayes sequential decision methods in clinical trials because of computational difficulty. It needs to be developed in the future.

Chapter 8

Discussion and Further Study

The corresponding discussions are given in chapters. The following are several common issues.

8.1 Prior Information

In the previous study, we have assumed that the prior distribution for parameters of interest could be obtained. Freedman and Spiegelhalter(1983) and Kadane(1986) discuss their experience of translating doctors opinions into subjective probability distributions. Chaloner *et. al.*(1993) describe a graphical elicitation of a prior distribution for a clinical trial. In clinical trials comparing two treatments, the normal likelihood with its mean being the parameter of interest covers many situations. The mean is often assumed to have a normal prior distribution. This is not only mathematically convenient, but also reasonably realistic. Spiegelhalter, Freedman and Parmar(1994) have listed sources of evidence for clinical priors, which are,

1) Evidence from other randomised trials.

For example, in the design of the $\beta-$Blocker heart attack trial(BHAT 1982), it was assumed that the drug propranalol would reduce mortality by 28% which

was based on the results of five European trials of different $\beta-$Blocker drugs. In their paper, they suggest that results from previous randomised trials should generally form the basis for a prior distribution but should not specify the distribution completely and that random effects models in meta-analysis might be an appropriate tool, in which case the prior distribution would correspond to the predictive distribution for the effect in a new trial(for example, see Carlin, 1992).

2) Evidence from non-randomised studies

It is often the situation that results of relevant randomised trial are not available, but non-randomised studies may have been conducted. For instance, Byar *et al.*(1976) have given some examples.

3) Subjective clinical opinion

This is emphasized in their paper even when evidence in the form of randomised studies is available. They have suggested that one approach to eliciting opinion is to conduct individual interviews with clinicians who would participate in the trial(for examples, Spiegelhalter, Freedman and Parmar, 1994, Chaloner *et al.,* 1993, MRC Urological Working Party 1985, Spiegelhalter and Freedman 1988, and Freedman and Spiegelhalter 1983, etc.). Genest and Zidek(1986) have proposed many methods to combine these individual distributions to arrive at a prior distribution for the group.

The subject of prior distributions is extensively studied by theoretical statistician(for example, see Bernardo and Smith, 1994). However, the issue of how we can collect proper prior information and to formulate a prior distribution for parameters of interest in clinical trials needs to be further studied in practice.

8.2 Decision Theory

Decision theory provides the framework for combining subjective distributions with action. The barriers to the use of Bayesian decision theory in clinical trials

include the choice and formulation of loss and cost functions, and the lack of statistical packages. There is a discussion on whether decision theory should be used in sequential clinical trials. Spiegelhalter, Freedman and Parmar(1994) have said that "when the decision is whether or not to discontinue the trial coupled with whether or not to recommend one treatment in preference to the other, the consequences of any particular course of action are so uncertain that they make the meaningful specification of utilities rather speculative.". However, in contrast, Berry(1994) has said that speculation and assessing uncertainty are the stuff of the Bayesian approach and that deciding whether to stop a trial requires considering why we are running it in the first place, and this means assessing utilities. He has also pointed out that the value of ε in Bayes sequential methods in clinical trials as described in Chapter 2 should be obtained by decision theory. Berry, Wolff, and Sack(1992, 1994) have given an example in which the various uncertainties are explicitly considered.

Lindley(1994) has said that "it must be recognized that clinical trials are not there for inference but to reach a decision, and the omission of their *raison d'être* is serious. In the long term, expected utility is realistic and , indeed, necessary.".

We are interested in applying decision theory in clinical trials as in Chapter 4, Chapter 5, and Chapter 6. There are problems in the formulation of proper loss and cost functions for clinical trials and in the computational difficulties.

Whitehead(1992) has said that, "There is a more fundamental concern about their use: is a Phase III clinical trial really a decision procedure? Who is making the decision? In reality, a clinical trial provides data for a whole series of decision makers. The investigators must decide whether to recommend an experimental drug, regulatory bodies must decide whether to licence it, and individual clinicians must decide whether to use it. In addition, future investigators will use the trial results to help in the design of their own studies. Each decision maker will have different prior opinions, different assessments of loss and different supplementary

sets of data available." There is a need to develop decision theory for multiple decision makers.

Thall and Simon(1994), and Thall, Simon, and Estey(1995) use frequentist criteria to avoid the specification of costs and a loss function intentionally although a loss function certainly is defined implicitly in their formulation. It is interesting to combine Bayesian methods with frequentist methods.

8.3 Computation

Bayesian approaches have not been widely used in clinical trials as frequentist methods due to the computational difficulties and the lack of statistical packages. In Chapter 6, we developed some approximate methods to be able to use the decision theory in clinical trials with survival time data. Recently Markov Chain Monte Carlo has become popular in using Bayesian approaches. Tanner and Wong(1987), and Tierney(1991) have used the way of Markov Chain Monte Carlo to find posterior distributions. Clydo, Muller and Parmigian(1996) explore expected utility surface by Markov Chain Monte Carlo. The computational methods concerning Bayesian inference need to be further explored. The corresponding statistical packages need to be developed.

Bibliography

Anscombe, F. (1963). Sequential medical trials. *J. Amer. Statist. Assoc. 58*, 365–387.

Armitage, P. (1975). *Sequential Medical Trials.* Oxford: Blackwell.

Armitage, P., C. McPherson, and B. Rowe (1969). Repeated significance tests on accumulating data. *J. Roy. Statist. Soc. A132*, 235–244.

Berger, J. (1984). The robust Bayesian viewpoint. In J. Kadane (Ed.), *Robustness of Bayesian Analysis*, pp. 63–144. Amsterdam: North-Holland.

Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Berlin: Springer.

Berger, J. (1990). Robust Bayesian analysis: Sensitivity to the prior. *J. Statist. Planning and Inference 25*, 303–328.

Berger, J. (1994). An overview of robust Bayesian analysis. *Test 3*, 5–124.

Berger, J. and L. Berliner (1986). Robust Bayes and empirical Bayes analysis with $\varepsilon-$contaminated priors. *Ann. Statist. 14*, 461–486.

Berliner, L. and B. Hill (1988). Bayesian nonparametric survival analysis. *J. Amer. Statist. Assoc. 83*, 772–784.

Bernardo, J. and A. Smith (1994). *Bayesian Theory.* Chichester: Wiley.

Berry, D. (1985). Interim analysis in clinical trials: Classical vs Bayesian approaches. *Statist. Med. 4*, 521–526.

Berry, D. (1987). Interim analysis in clinical trials: The role of the likelihood principle. *Amer. Statist. 41*, 117–122.

Berry, D. (1989). Monitoring accumulating accumulating data in a clinical trial. *Biometrics 45*, 1197–1211.

Berry, D. (1994). Comments on Bayesian approaches to randomized trials. *J. Roy. Statist. Soc. A132*, 399.

Berry, D. and C.-H. Ho (1988). One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics 44*, 219–227.

Berry, D., M. Wolff, and D. Sack (1992). Public health decision making: a sequential vaccine trial. In J. Bernado, J. Berger, A. David, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 483–502. Oxford: University Press.

Berry, D., M. Wolff, and D. Sack (1994). Decision making during a phase III randomized controlled trial. *Controlled Clin. Trials 15*, 360–378.

β-Blocker Heart Attack Trial Research Group (1981a). β-Blocker heart attack trial: Design features. *Controlled Clin. Trials 2*, 275–285.

β-Blocker Heart Attack Trial Research Group (1982b). A randomized trial of popranolol in patients with acute myocardial infarction. *J. Amer. Med. Assoc. 247*, 1707–1714.

Byar, D., R. Simon, W. Friedewald, J. Schlesselman, D. DeMets, J. Ellenberg, M. Gail, and J. Ware (1976). Randomized clinical trials: perspective on some recent ideas. *New Engl. J. Med. 295*, 74–80.

Carlin, B., K. Chaloner, T. Church, T. Louis, and J. Matts (1993). Bayesian approaches for monitoring clinical trials with an application to toxiplasmic encephalitis prophylaxis. *Statistician 42*, 355–367.

Carlin, J. (1992). Meta-analysis for 2×2 tables: a Bayesian approach. *Statist. Med. 11*, 141–158.

Chaloner, K., T. Church, T. Louis, and J. Matts (1993). Graphical elicitation of a prior distribution for a clinical trial. *Statistician 42*, 341–353.

Chernoff, H. and A. Petkau (1985). Sequential medical trials with ethical cost. In L. M. L. Cam and R. A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jersy Neyman and Jack Kiefer, Volume II.*

Clyde, M., P. Muller, and G. Parmigian (1996). Exploring expected utility surfaces by markov chains. Technical report, Institute of Statistics and Decision Sciences, Duke University.

Cornfield, J. (1966a). A Bayesian test of some classical hypotheses - with applications to sequential clinical trials. *J. Amer. Statist. Assoc. 61*, 577–594.

Cornfield, J. (1966b). Sequential trials, sequential analysis and the likelihood principle. *Amer. Statist. 20*, 18–23.

Cornfield, J. (1969). The Bayesian outlook and its application. *Biometrics 24*, 617–657.

Cornfield, J. (1976). Recent methodological contributions to clinical trials. *Amer. J. Epidemiol. 104*, 408–421.

Dawid, A. (1973). Posterior expectations for large observations. *Biometrika 60*, 664–667.

DeGroot, M. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

DeMets, D. (1987). Practical aspects in data monitoring: A brief review. *Statist. Med. 6*, 753–760.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distribution. *Ann. Probability 2*, 183–201.

Donner, A. (1977). The use of auxiliary information in the design of a clinical trial. *Biometrics 33*, 305–314.

Dykstra, R. and P. Laud (1981). A Bayesian non-parametric approach to reliability. *Ann. Statist. 9*, 356–367.

Elashoff, J. and T. Reedy (1984). Two-stage clinical trial stopping rules. *Biometrics 40*, 791–796.

Ferguon, T. (1967). *Mathematical Statistics: A Decision Theoretic Approach.* New York: Academic Press.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist. 1*, 209–230.

Ferguson, T. (1974). Prior distributions on spaces of probability measures. *Ann. Statist. 2*, 615–629.

Ferguson, T. and E. Phadia (1979). Bayesian nonparametric estimation based on censored data. *Ann. Statist. 7*, 163–186.

Fleming, T. and D. DeMets (1993). Monitoring of clinical trials: Issues and recommendations. *Controlled Clin. Trials 14*, 183–197.

Freedman, L., D. Lowe, and P. Macaskill (1984). Stopping rules for clinical trials. *Biometrics 40*, 575–586.

Freedman, L. and D. Spiegelhalter (1983). The assessment of subjective opinion and its use in relation to stopping rules for clinical trials. *The Statistician 32*, 153–160.

Freedman, L. and D. Spiegelhalter (1989). Comparison of Bayesian with group sequential methods for monitoring clinical trials. *Controlled Clin. Trials 10*, 357–367.

Freedman, L. and D. Spiegelhalter (1991). Response to letter to the editor. *Controlled Clin. Trials 12*, 346–350.

Freedman, L. and D. Spiegelhalter (1993). Application of Bayesian statistics to decision making during a clinical trial. In *Yearbook of Medical Informatics*,

pp. 353–365.

Freedman, L., D. Spiegelhalter, and M. Parmar (1994). The what, why and how of Bayesian clinical trials monitoring. *Statist. Med. 13*, 1317–1383.

Geller, N. and S. Pocock (1987). Interim analyses in randomized clinical trials: Ramifications and guidelines for practitioners. *Biometrics 43*, 213–223.

Genest, C. and J. Zidek (1986). Combining probability distributions: a critique and annotated bibliography. *Statist. Sci. 1*, 114–148.

George, S., C. Li, and D. Berry (1994). Stopping a clinical trial early: frequentist and Bayesian approaches applied to a calgb trials in non-small-cell lung cancer. *Statist. Med. 13*, 1313–1327.

Good, I. (1959). Could a machine make probability judgements ? *Computers and Automation 8*, 14–16,24–26.

Good, I. (1961). Discussion of C.A.B. Smith "Consistency in statistical inference and decision ". *J. Roy. Statist. Soc. 23*, 28–29.

Good, I. (1962). Subjective probability as the measure of a non-measurable set. In *Logic, Methodology and the Philosphy of Science*, pp. 319–329. Standford: University Press.

Greenhouse, J. and L. Wasserman (1995). Robust Bayesian methods for monitoring clinical trials. *Statist. Med. 14*, 1379–1391.

Halperin, M., K. Lan, J. Ware, N. Johnson, and D. DeMets (1982). An aid to data monitoring in long-term clinical trials. *Controlled Clin. Trials 3*, 311–323.

Haybittle, J. (1971). Repeated assessment of results in clinical trials of cancer treatment. *Br. J. Radiol. 44*, 793–797.

Hjort, N. (1990). Nonparametric Bayes estimators based on beta processes in model for life history data. *Ann. Statist. 18*, 1259–1294.

Huber, P. (1973). The use of choquet capacities in statistics. *Bulletin of the International Statistics Institute 45*, 181–191.

Jennison, C. and B. Turnbull (1984). Repeated confidence intervals for group sequential trials. *Controlled Clin. Trials 5*, 33–45.

Jennison, C. and B. Turnbull (1989). Interim analysis: the repeated confidence interval approach. *J. Roy. Statist. Soc. B51*, 305–361.

Jennison, C. and B. Turnbull (1990). Statistical approaches to interim monitoring of medical trials: a review and commentary. *Statistical Science 5*, 299–317.

Kadane (1986). Progress toward a more ethical method for clinical trials. *J. Med. Philos. 11*, 385–404.

Kalbfleisch, J. (1978). Non-parametric Bayesian analysis of survival time data. *J. R. Statist. Soc. B40*, 214–221.

Karlin, S. and H. Taylor (1975). *A First Course in Stochastic Processes*. New York: Academic Press.

Kass, R. and J. Greenhouse (1989). Comments on "investigating therapies of potentially great benefit: Emco" (by J.H. Ware). *Statist. Sci. 4*, 310–317.

Kim, K. and D. DeMets (1987). Design and analysis of group sequential tests based on the type I error spending rate function. *Biometrika 74*, 149–154.

Kim, K. and A. Tsiatis (1990). Study duration for clinical trials with survival responses and early stopping rule. *Biometrics 46*, 81–92.

Koepcke, W. (1989). Analyses of group sequential clinical trials. *Controlled Clin. Trials 10*, 222S–230S.

Lan, K. and D. DeMets (1983). Discrete sequential boundaries for clinical trials. *Biometrika 70*, 659–663.

Lan, K. and D. DeMets (1989). Group sequential procedures: calendar versus in formation time. *Statistics in Medicine 8*, 1191–1198.

Lan, K., R. Simon, and M. Halperin (1982). Stochastically curtailed tests in long-term clinical trials. *Commun. Statist. C1*, 207–219.

Lavine, M., L. Wasserman, and R. Wolpert (1991). Bayesian inference with specified prior marginals. *J. Amer. Statist. Assoc. 86*, 964–971.

Lewis, R. (1996). Bayesian hypothesis testing: Interim analysis of a clinical trial evaluating phenytoin for the prophylaxis of early post-traumatic seizures in children. In D. Berry and D. Stangl (Eds.), *Bayesian Biostatistics*, pp. 279–296. New York: Marrel Dekker.

Lewis, R. and D. Berry (1994). Group sequential clinical trials - a classical evaluation of Bayesian decision theoretic designs. *J. Amer. Statist. Assoc. 89*, 1528–1534.

Lindley, D. (1994). Comments on Bayesian approaches to randomized trials. *J. Roy. Statist. Soc. A132*, 393.

MRC Urological Working Party (1985). Intravesical thiotepa for superficial bladder tumors: an MRC randomized study. *Br. J. Urol. 57*, 680–689.

Novick, M. and J. E. Grizzle (1965). A Bayesian approach to the analysis of data from clinical trials. *J. Amer. Statist. Assoc. 60*, 81–96.

O'Brien, P. and T. Fleming (1979). A multiple testing procedure for clinical trials. *Biometrics 35*, 549–556.

Padgett, W. and L. Wei (1981). A Bayesian nonparametric estimator of survival probability assuming increasing failure rate. *Commun. Statist.- Theor. Meth. A10*, 49–63.

Papallona, S. and A. Tsiatis (1994). Group sequential designs for one-sided and two sided hypothesis testing with provision for early stopping in favor of the

null hypothesis. *J. of Statist. Planning and Inference 42*, 19–35.

Parmar, M., D. Spiegelhalter, and L. Freedman (1994). The chart trials: Bayesian design and monitoring in practice. *Statist. Med. 13*, 1297–1312.

Peto, R., M. Pike, P. Armitage, N. Breslow, D. Cox, S. Howard, N. Mantel, K. McPherson, J. Peto, and P. Smith (1976). Design and analysis of randomised clinical trials requiring prolonged observation of each patient. I. introduction and design. *Br. J. Cancer 34*, 586–612.

Peto, R., M. Pike, P. Armitage, N. Breslow, D. Cox, S. Howard, N. Mantel, K. McPherson, J. Peto, and P. Smith (1977). Design and analysis of randomised clinical trials requiring prolonged observation of each patient. II. analysis and examples. *Br. J. Cancer 35*, 1–39.

Pocock, S. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika 64*, 191–199.

Pocock, S. (1982). Interim analyses for randomised clinical trials: The group sequential approach. *Biometrics 38*, 153–162.

Pocock, S. (1983). *Clinical Trials: A Practical Approach.* New York: Wiley.

Pocock, S. (1992). When to stop a clinical trial. *Br. Med. J. 305*, 235–240.

Pocock, S. and M. Hugher (1989). Practical problems in interim analyses, with particular regard to estimation. *Controlled Clin. Trials 10*, 209S–221S.

Raiffa, H. and R. Schlaifer (1961). *Applied Statistical Decision Theory.* Boston: Harvard University Press.

Rosner, G. and D. Berry (1995). A Bayesian group sequential design for a multiple arm randomized clinical trials. *Statist. Med. 14*, 381–394.

Schaid, D., S. Wieand, and T. Therneau (1990). Optimal two-stage screening designs for survival comparisons. *Biometrika 77*, 507–513.

Snapinn, S. (1992). Monitoring clinical trials with a conditional probability stopping rule. *Statist. Med. 11*, 659–672.

Spiegelhalter, D. (1986). Probabilistic prediction in patient management and clinical trials. *Statist. Med. 5*, 421–433.

Spiegelhalter, D. and L. Freedman (1988). Bayesian approaches to clinical trials. In J. Bernardo, M. DeGroot, D. Lindley, and A. Smith (Eds.), *Bayesian Statistics 3*, pp. 453–477. Oxford: University Press.

Spiegelhalter, D., L. Freedman, and M. Parmar (1994). Bayesian approaches to randomised trials. *J. R. Statist. Soc. A157*, 357–461.

Susarla, V. and J. Van-Ryzin (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations. *J. Amer. Statist. Assoc. 71*, 897–902.

Sylvester, R. (1988). A Bayesian approach to the design of phase II clinical trials. *Biometrics 44*, 823–836.

Sylvester, R. (1990). Response to reader reaction. *Biometrics 46*, 537–538.

Thall, P. and R. Simon (1994). Practical Bayesian guidelines for phaseIIB clinical trials. *Biometrics 50*, 337–349.

Thall, P., R. Simon, and E. H. Estey (1995). Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes. *Statist. Med. 14*, 357–379.

Tsiatis, A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika 68*, 311–315.

Tsiatis, A. (1987). Repeated significance testing for a general class of statistics used in censored survival analysis. *J. Amer. Statist. Assoc. 77*, 855–861.

Wang, S. and A. Tsiatis (1987). Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics 43*, 193–199.

Wasserman, L. (1992). Recent methodological advances in robust Bayesian inference. In J. Bernado, J. Berger, A. David, and A. Smith (Eds.), *Bayesian Statistics 4*, pp. 483–502. Oxford: University Press.

Whitehead, J. (1982). *The Design and Analysis of Sequential Clinical Trials.* Chichester: Ellis Horwood.

Whitehead, J. (1991). Sequential methods in clinical trials. In B. Ghosh and P. Sen (Eds.), *Handbook of Sequential Analysis*, pp. 593–622. New York: Dekker.

Whitehead, J. (1992). *The Design and Analysis of Sequential Clinical Trials.* Chichester: Ellis Horwood.

Whitehead, J. and I. Stratton (1983). Group sequential clinical trials with triangular continuation regions. *Biometrics 39*, 227–236.

Wieand, H. and S. Cha (1992). Description of the statistical aspects of a study for advanced colorectal cancer patients. *Statist. Med. 11*, 5–12.