

A General Framework

A randomized clinical trial asks questions about the effect of an intervention on an outcome defined by a continuous, dichotomous, or time-to-failure variable. While the test statistics associated with these outcomes may appear quite disparate, they share a common thread—all behave like standardized sums of independent random variables. In fact, they all have the same asymptotic joint distribution over time, provided that we define the time parameter appropriately. Understanding the distribution of the test statistic over time is essential because typically we monitor data several times throughout the course of a trial, with an eye toward stopping if data show convincing evidence of benefit or harm. In clinical trials, the term “monitoring” often refers to a procedure for visiting clinical sites and checking that the investigators are carrying out the protocol faithfully and recording the data accurately. In statistics, and in this book, “monitoring” refers to the statistical process of assessing the strength of emerging data for making inferences or for estimating the treatment effect.

This chapter distinguishes between hypothesis testing (Section 2.1) and parameter estimation (Section 2.2). We begin with simple settings in which the test statistic and treatment effect estimator are a sum and mean, respectively, of independent and identically distributed (i.i.d.) random variables. We show that in less simple settings, the test statistic and treatment effect estimator behave *as if* they were a sum and mean, respectively, of i.i.d. random variables. This leads naturally to the concept of a sum process (S-process) behaving like a sum and an estimation process (E-process) behaving like a sample mean. Following the approach of Lan and Zucker (1993) [LZ93] and Lan and Wittes (1988) [LW88], we show the connection between S-processes, E-processes, and Brownian motion. We use Brownian motion to approximate the joint distribution of repeatedly computed test statistics over time for many different trial settings, including comparisons of means, proportions, and survival times, with or without adjustment for covariates. Because of our extensive use of Brownian motion, we were tempted to subtitle this chapter “Brown v. the Board of Data Monitoring.”

This chapter, which presents the general framework for the rest of the book, is necessarily long. The reader may prefer to read the first three sections containing the essential ideas applied to tests of means, proportions, and survival, and then proceed to the next chapter showing how to apply Brownian motion to compute conditional power. The reader may then return to this chapter to see how to use the same ideas in more complicated settings such as maximum likelihood or minimum variance estimation, or even mixed models. While digesting the next sections, the reader should keep in mind the essential idea throughout this chapter—test statistics and estimators behave like sums and sample means, respectively, of i.i.d. random variables.

Lest the reader get the wrong impression that Brownian motion, like gravity, always works, we close the chapter with an example in which Brownian motion fails to provide a good approximation to the joint distribution of a test statistic over time.

2.1 Hypothesis Testing: The Null Distribution of Test Statistics Over Time

This section focuses on the null distribution of test statistics over time, while the next section deals with the distribution under an alternative hypothesis. We begin with paired data assuming the paired differences are independent and identically distributed normals with known variance. Because this ideal setting rarely holds in clinical trials, we then back away from these assumptions, one by one, to see which are really necessary.

2.1.1 Continuous Outcomes

Imagine a trial with a continuous outcome, and suppose first that the data are paired. For example, the data might come from a crossover trial studying the effects of two diets on blood pressure, or from a trial comparing two different treatments applied directly to the eyes, one to the left eye and the other to the right. Let X_i and Y_i be the control and treatment observations, respectively, for patient i and let $D_i = Y_i - X_i$. Assume that the D_i are normally distributed with mean δ and known variance σ^2 . We wish to test whether $\delta = 0$.

At the end of the trial the z -score is

$$Z_N = v_N^{-1/2} \sum_{i=1}^N D_i, \quad (2.1)$$

where $S_N = \sum_{i=1}^N D_i$ and $v_N = \text{var}(S_N) = N \text{var}(D_1)$. Treatment is declared beneficial if $Z_N > z_{\alpha/2}$, where z_a , for $0 < a < 1$, denotes the $100(1 - a)$ th percentile of a standard normal distribution.

Now imagine an interim analysis after n of the planned N observations in each arm have been evaluated. Note that

$$\begin{aligned} Z_N &= \{S_n + S_N - S_n\}/\sqrt{v_N} \\ &= S_n/\sqrt{v_N} + (S_N - S_n)/\sqrt{v_N} \end{aligned} \quad (2.2)$$

is the sum of two independent components. We call the first term of (2.2) the *B-value* because of its connection to Brownian motion established later in this chapter. We term the ratio

$$t = v_n/v_N = \text{var}(S_n)/\text{var}(S_N) \quad (2.3)$$

the *trial fraction* because it measures how far through the trial we are. In this simple case, t simplifies to n/N , the fraction of participants evaluated thus far; $t = 0$ and $t = 1$ correspond to the beginning and end of the trial, respectively.

Denote the interim z-score $S_n/v_n^{1/2}$ at trial fraction t by $Z(t)$. Define the B-value $B(t)$ at trial fraction t by

$$B(t) = \frac{S_n}{\sqrt{v_N}} \quad (2.4)$$

$$= \sqrt{t}Z(t). \quad (2.5)$$

We could monitor using either the z-score or the B-value; in this book we use both. We use z-scores for setting boundaries (i.e., calculations assuming the null hypothesis is true), whereas for deciding whether observed results follow the expected trend (i.e., calculations assuming the alternative hypothesis is true), we find it advantageous to think in terms of B-values.

At the end of the trial, $B(1) = Z(1) = S_N/v_N^{1/2}$, so (2.2) becomes

$$B(1) = B(t) + \{B(1) - B(t)\}. \quad (2.6)$$

The decomposition (2.2) leading to (2.6) clearly implies that $B(t)$ and $B(1) - B(t)$ are independent (note, however, that the forthcoming derivation of the covariance structure of $B(t)$ is valid even when $B(t)$ and $B(1) - B(t)$ are uncorrelated, but not independent). At trial fraction t , $B(t)$ reflects the past while $B(1) - B(t)$ lies in the future.

More generally, let $t_0 = 0, t_1 = n_1/N, \dots, t_k = n_k/N$ and let $B(t_0) = 0, B(t_1) = S_{n_1}/v_N^{1/2}, \dots, B(t_k) = S_{n_k}/v_N^{1/2}$ be interim B-values at trial fractions $t_0 = 0, t_1, \dots, t_k$. Then the successive increments $B(t_1) - B(t_0) = S_{n_1}/v_N^{1/2}, B(t_2) - B(t_1) = (S_{n_2} - S_{n_1})/v_N^{1/2}, \dots, B(t_k) - B(t_{k-1}) = (S_{n_k} - S_{n_{k-1}})/v_N^{1/2}$ are independent because they involve nonoverlapping sums. Further, (2.5) implies that

$$\text{var}\{B(t)\} = t \text{var}\{Z(t)\} = t.$$

For $t_i \leq t_j$,

$$\begin{aligned}
\text{cov}\{B(t_i), B(t_j)\} &= \text{cov}\{S_{n_i}/v_N^{1/2}, S_{n_j}/v_N^{1/2}\} \\
&= v_N^{-1} \text{cov}\{S_{n_i}, S_{n_i} + S_{n_j} - S_{n_i}\} \\
&= v_N^{-1} \{\text{cov}(S_{n_i}, S_{n_i}) + \text{cov}(S_{n_i}, S_{n_j} - S_{n_i})\} \\
&= v_N^{-1} \{\text{var}(S_{n_i}) + 0\} = v_{n_i}/v_N = t_i.
\end{aligned} \tag{2.7}$$

Thus, the distribution of $B(t)$ has the following structure:

- B1: $B(t_1), B(t_2), \dots, B(t_k)$ have a multivariate normal distribution.
- B2: $E\{B(t)\} = 0$.
- B3: $\text{cov}\{B(t_i), B(t_j)\} = t_i$ for $t_i \leq t_j$.

Properties B1-B3 and relationship (2.5) confer the following properties to z-scores:

- Z1: $Z(t_1), Z(t_2), \dots, Z(t_k)$ have a multivariate normal distribution.
- Z2: $E\{Z(t)\} = 0$.
- Z3: $\text{cov}\{Z(t_i), Z(t_j)\} = (t_i/t_j)^{1/2}$ for $t_i \leq t_j$.

We have been somewhat loose in that we have defined $B(t)$ only at trial fraction values $t = 0, 1/N, \dots, N/N = 1$. That the set of points at which we defined the B-value depends on N suggests that we really should use the notation $B_N(t)$. The natural way to extend the definition of $B_N(t)$ to the entire unit interval is by linear interpolation: if $t = \lambda(i/N) + (1 - \lambda)\{(i + 1)/N\}$, we define $B_N(t)$ to be $\lambda B_N(i/N) + (1 - \lambda)B_N\{(i + 1)/N\}$. This makes $B_N(t)$ continuous on $t \in (0, 1)$ but nondifferentiable at the “sharp” points $t = 0, 1/N, \dots, N/N = 1$. As $N \rightarrow \infty$, the set of t at which $B_N(t)$ is nondifferentiable becomes more and more dense. In the limit, we get *standard Brownian motion*, a random, continuous, but nondifferentiable, function $B(t)$ satisfying B1-B3 (Figure 2.1).

The approach we take throughout the book is first to transform a probability involving z-scores $Z_N(t)$ to one involving B-values $B_N(t) = t^{1/2}Z_N(t)$, and then to approximate that probability by one involving the limiting Brownian motion process, $B(t) = \lim_{N \rightarrow \infty} B_N(t)$. A major advantage to this approach is that properties and formulas involving Brownian motion are well known, having been studied extensively by mathematicians and physicists. The following example demonstrates in detail the process of using Brownian motion to approximate probabilities of interest. In the future, we jump right to $B(t)$, eliminating the intermediate step of arguing that probabilities involving $B_N(t)$ can be approximated by those of $B(t)$.

Example 2.1. Consider a trial comparing two different treatments for the eye. Each volunteer receives treatment 1 in one randomly selected eye and treatment 2 in the other. The outcome for each volunteer is the difference between the results from the eye treated with treatment 1 and the eye treated with treatment 2. Suppose we take an interim analysis after 50 of the 100 planned patients are evaluated, and the paired t-statistic is 1.44. The sample size is sufficiently large to regard the t-statistic as a z-score.

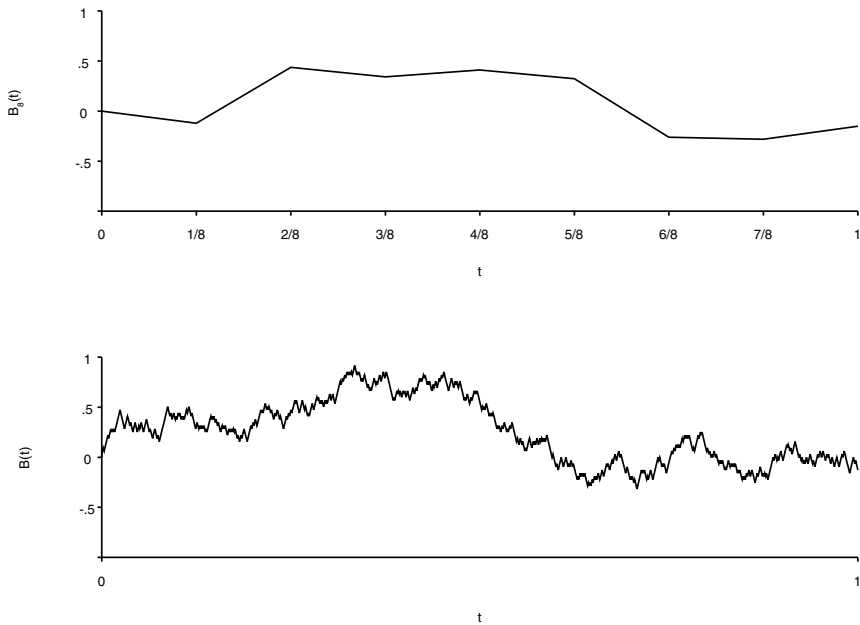


Fig. 2.1. Top panel: The B-value $B_N(t)$ for a trial with $N = 8$ pairs; $B_8(t)$ is defined by linear interpolation for t other than $i/8$, $i = 0, \dots, 8$. The resulting random function is continuous everywhere but not differentiable at the “sharp” points $t = i/8$, $i = 0, \dots, 8$. Bottom panel: As the sample size N increases, the set of points at which $B_N(t)$ is not differentiable becomes denser. The limiting case of $B_N(t)$ as $N \rightarrow \infty$ is Brownian motion, a random function continuous everywhere but differentiable nowhere, satisfying B1-B3. This nondifferentiability reflects the zigzagging Brown noted when he looked at pollen through his microscope (see the end of Chapter 1).

The trial fraction is $t = 50/100 = 0.50$, so $Z(0.50) = 1.44$. The B-value is $B(0.50) = (0.50)^{1/2}(1.44) = 1.018$. We can approximate the joint distribution of the interim and final B-values, $B_{100}(0.50)$ and $B_{100}(1)$, by those of $B(0.50)$ and $B(1)$, where $B(t)$ is Brownian motion. For example, we could compute boundaries a_1 and a_2 such that $\Pr(B(0.50) \geq a_1) = 0.01$ and $\Pr(B(0.50) \geq a_1 \cup B(1) \geq a_2) = 0.05$ (equivalently, z-score boundaries c_1 and c_2 such that $\Pr(Z(0.50) \geq c_1) = 0.01$ and $\Pr(Z(0.50) \geq c_1 \cup Z(1) \geq c_2) = 0.05$). We can also use Brownian motion to compute more complicated probabilities such as the effect on type 1 error rate of monitoring continuously from now to the end of the trial without adjusting for multiple looks (i.e., using criti-

cal value 1.96). The actual type 1 error rate, $\Pr(Z_{100}(i/N) \geq 1.96 \text{ for some } i = 50, 51, \dots, 100)$, can be approximated by $\Pr(B(t)/t^{1/2} \geq 1.96 \text{ for some } 1/2 \leq t \leq 1)$.

Our next step is to show that Brownian motion approximates the null distribution over t for many other testing scenarios. We reexamine the assumptions in Section 2.1.1 to see which ones we can relax.

First, the differences need not be normally distributed. Even if D is not normally distributed, the increments are independent and, by the central limit theorem (CLT), each increment is approximately normally distributed. Consequently, the joint distribution of partial sums is approximately multivariate normal even if the individual observations are not normally distributed.

Second, the sample variance need not be known. As we argued in the example above, Brownian motion holds approximately even if v_n is a consistent estimate of $\text{var}(S_n)$ (that is, $\text{var}(S_n)/v_n$ tends to 1 in probability—see Section 2.9.1 for a formal proof).

Third, we do not need paired observations, as we illustrate in the next section.

2.1.2 Dichotomous Outcomes

Consider a parallel arm trial with a dichotomous outcome such as 28-day mortality. Denote by $I(A)$ the indicator function taking the value 1 if the event A occurs and 0 otherwise. Although the data are not paired differences, we can view the difference in proportions after n patients per arm as S_n/n , where S_n is the sum of n paired differences (we get the same difference in proportions irrespective of how we pair treatment and control observations). The observations $D_i = I(\text{patient } i \text{ of treatment arm has an event}) - I(\text{patient } i \text{ of control arm has an event})$, $i = 1, \dots, N$ are i.i.d. with null mean 0 and variance $2p(1-p)$, where p is the null probability that a randomly selected patient has an event. The z-statistic at the end of the trial is given by (2.1), where $v_N = \text{var}(S_N) = 2Np(1-p)$ is the null variance of S_N . As the true p is unknown, to compute the z-score one replaces p by the sample proportion of all patients with events. The result is the usual (unpaired) z-statistic for a test of proportions. Decomposition (2.2) still holds. Define t by (2.3), which again simplifies to n/N . Brownian motion is again a good approximation for $B(t)$ defined by (2.4). Also, the joint distribution of z-scores is asymptotically the same for a dichotomous outcome trial as for a continuous outcome trial. We can use the same boundaries to monitor either type of trial.

Of course, we do not actually pair the data from a parallel arm trial. In fact, it is unusual for the control and treatment sample sizes to be exactly the same even at the end of a trial, let alone at all interim analyses. Later we will see how to use Brownian motion even in the unequal sample size setting.

Example 2.2. Suppose we design a trial of 200 breast cancer patients randomly assigned in a 1:1 ratio to the standard treatment plus a new treatment or to

the standard treatment plus placebo. We want to compare the proportion of patients whose tumor regresses by 3 months after randomization. Interim analyses occur after 50, 75, and 100 patients per arm have been evaluated. The corresponding trial fractions are $t_1 = 50/100 = 0.50$, $t_2 = 75/100 = 0.75$, and $t_3 = 100/100 = 1$. If the z-scores for the usual test of proportions are $Z(0.50) = 0.55$, $Z(0.75) = -0.20$, and $Z(1) = 0.23$, the B-values are $B(0.50) = (0.50)^{1/2}(0.55) = 0.389$, $B(0.75) = (0.75)^{1/2}(-0.20) = -0.173$, and $B(1) = (1)^{1/2}(0.23) = 0.230$. The joint distribution of $B(0.50)$, $B(0.75)$, and $B(1)$, and therefore the joint distribution of $Z(0.50)$, $Z(0.75)$, and $Z(1)$, is the same as for a trial with a continuous outcome monitored at those trial fractions. Any boundary developed for continuous outcome trials would be valid for this dichotomous outcome trial as well. For any z-score boundary c_1 , c_2 , and c_3 we could compute the probability of crossing at various times. For example, suppose the upper boundary is $c_1 = 2.963$, $c_2 = 2.359$, and $c_3 = 2.014$. The probability of crossing the boundary at $t = 0.50$ is $\Pr(Z(0.50) \geq 2.963) = 1 - \Phi(2.963) = 0.0015$. The cumulative probability of crossing by the second look depends on the joint distribution of $Z(0.50)$ and $Z(0.75)$, which by properties Z1-Z3 is bivariate normal with zero means, unit variances, and covariance $(0.50/0.75)^{1/2} = 0.816$. We can use numerical integration (described in Section 4.7) to show that the cumulative crossing probability by $t = 0.75$ is $\Pr[\{Z(0.50) \geq 2.963\} \cup \{Z(0.75) \geq 2.359\}] = 0.0097$. Similarly, for the cumulative crossing probability by $t = 1$, we use the fact that

$$\begin{aligned}\text{cov}\{Z(0.50), Z(0.75)\} &= 0.816 \\ \text{cov}\{Z(0.50), Z(1)\} &= (0.50/1)^{1/2} = 0.707 \\ \text{cov}\{Z(0.75), Z(1)\} &= (0.75/1)^{1/2} = 0.866.\end{aligned}$$

The cumulative crossing probability by $t = 1$ is $\Pr[\{Z(0.50) \geq 2.963\} \cup \{Z(0.75) \geq 2.359\} \cup \{Z(1) \geq 2.014\}] = 0.025$.

We next relax the assumption of independent observations. Notice that the steps leading to (2.7) remain valid even if the D_i s are merely uncorrelated. Thus, even when the observations are uncorrelated but not independent, the B-values have the same correlation structure as Brownian motion. If we are willing to accept that the joint distribution of the B-values is asymptotically multivariate normal, then it must be that of Brownian motion. In the next section, we apply this idea to comparison of survival curves using the logrank statistic.

2.1.3 Survival Outcomes

In many clinical trials, the outcome is the time to some event. For simplicity, assume the event is death so that each person can only have one event; the same ideas apply for events that can recur, but in those cases we restrict attention to the first event for each patient. We use the logrank statistic to

compare the treatment and control arms. Assume for now that all patients are randomized simultaneously. We show that the logrank statistic is also of the form (2.1) for uncorrelated, mean 0 random variables D_i . Brownian motion can approximate its null joint distribution at different analysis times. See Chapter 13 for further discussion of the logrank and related tests.

Let N be the total number of deaths at the end of the trial instead of the per-arm sample size. The numerator of the logrank statistic at the end of the trial is $\sum_{i=1}^N D_i$, where $D_i = O_i - E_i$, O_i is the indicator that the i th death occurred in a treatment patient, and $E_i = m_{1i}/(m_{0i} + m_{1i})$ is the null expectation of O_i given the respective numbers, m_{0i} and m_{1i} , of control and treatment patients at risk just prior to the i th death. Conditioned on m_{0i} and m_{1i} , O_i has a Bernoulli distribution with parameter E_i . The null conditional mean and variance of D_i are 0 and $V_i = E_i(1 - E_i)$, respectively.

We show in Section 2.9.3 that, unconditionally, the D_i are uncorrelated, mean 0 random variables with variance $E(V_i)$ under the null hypothesis. Thus, conditioned on N , $v_N = \text{var}(S_N) = \sum_{i=1}^N \text{var}(D_i) = \sum_{i=1}^N E(V_i) = E(\sum_{i=1}^N V_i)$. The logrank statistic is given by (2.1), where v_N is replaced by its estimate $\sum_{i=1}^N V_i$.

In the setting of survival, we should define the trial fraction in terms of patients with events rather than patients evaluated. Suppose we examine the data after n deaths. If we condition on N and n and define the trial fraction by (2.3), the covariance structure of Brownian motion holds. For now, assume that the joint distribution of $B(t_1), \dots, B(t_k)$ is approximately multivariate normal. Then Brownian motion is again a good approximation to the process $B(t)$. A practical problem is that at the interim analysis, we would not know v_N even if we knew with certainty the number, N , of patients with an event by the end of the trial. We can, however, approximate v_N as follows. Under the null hypothesis, $E(V_i) = E\{E_i(1 - E_i)\} \approx (1/2)(1 - 1/2) = 1/4$. We find this result quite remarkable—without making any assumption about the form of the survival curve, this simple argument shows that the variance of D_i is approximately $1/4$. It follows that $v_N \approx N/4$. This calculation leads to the familiar estimate $t = n/N$. In other words, for the logrank test, the trial fraction is the ratio of the number of patients with an event thus far to the number expected by trial's end.

Example 2.3. Consider a trial comparing mortality of lung cancer patients on a new treatment plus the standard treatment compared to placebo plus the standard treatment. Assume 200 deaths expected over the 2-year trial, and monitoring every 6 months. The total numbers of deaths at the first three looks were 20, 50, and 122, so the estimated trial fractions were $t_1 = 20/200 = 0.10$, $t_2 = 50/200 = 0.25$, and $t_3 = 122/200 = 0.61$. The values of the logrank statistic at these looks were $Z(0.10) = -0.162$, $Z(0.25) = 0.258$, and $Z(0.61) = 1.384$, so the B-values were $(0.10)^{1/2}(-0.162) = -0.051$, $B(0.25) = (0.25)^{1/2}(0.258) = 0.129$, and $B(0.61) = (0.61)^{1/2}(1.384) = 1.081$. Under the null hypothesis, these B-values behave like Brownian motion. Sup-

pose we constructed boundaries c_1 , c_2 , and c_3 such that

$$\Pr(Z(0.10) \geq c_1 \cup Z(0.25) \geq c_2 \cup Z(0.61) \geq c_3) = 0.01.$$

But imagine that when we reached the end of the trial, we had 190 instead of the expected 200 deaths. Thus, the “right” trial fractions at earlier looks should have been $t_1 = 20/190 = 0.105$, $t_2 = 50/190 = 0.263$, and $t_3 = 122/190 = 0.642$. The actual probability of crossing at least one earlier boundary should have been

$$\Pr(Z(0.105) \geq c_1 \cup Z(0.263) \geq c_2 \cup Z(0.642) \geq c_3). \quad (2.8)$$

Fortunately, this discrepancy does not present a problem because the null joint distribution of $Z(t_1), Z(t_2), Z(t_3)$ is multivariate normal with marginal mean 0 and variance 1, and $\text{cov}\{Z(t_i)/Z(t_j)\} = (t_i/t_j)^{1/2}$. This distribution depends on the trial fractions only through their ratios. The ratio of trial fractions is invariant to how many events we thought there would be at the end; e.g., $(20/200)/(50/200) = (20/190)/(50/190) = 20/50$. Thus, the correct probability of crossing an earlier boundary, (2.8), is also 0.01. We will see this invariance property many more times.

We used some sleight of hand in concluding that $(B(t_1), \dots, B(t_k))$ is approximately multivariate normal in the survival setting. Because $\sum_{i=1}^N D_i$ is a sum of uncorrelated but not independent observations, we can no longer rely on the central limit theorem to conclude that the asymptotic marginal distribution of $\sum_{i=1}^N D_i$ is normal. Furthermore, asymptotic marginal normality of $\sum_{i=1}^N D_i$ does not necessarily imply asymptotic multivariate normality of $(\sum_{i=1}^{n_1} D_i, \dots, \sum_{i=1}^{n_k} D_i)$, as it did in the clinical trial scenarios in which the D_i s were independent. Things get even more complicated if we account for the fact that in most trials participants are recruited over time (staggered entry) instead of all at once. A more rigorous treatment accounting for these factors requires a stochastic process formulation. Using such a formulation, one can show that the simple result obtained above holds under staggered entry as well. That is, $B(t) = t^{1/2}Z(t)$ behaves asymptotically like Brownian motion, where the trial fraction t is the ratio of the number of patients with an event thus far to the number expected by trial’s end, and $Z(t)$ is the logrank statistic at trial fraction t .

2.1.4 Summary of Sums

In the clinical trial scenarios considered thus far, the test statistic was a sum of either independent or uncorrelated observations. In either case, we adopted the following approach to convert the statistic to a B-value:

Approach 1. *We transform a sum of independent or uncorrelated random variables to a B-value $B(t)$ having the same correlation structure as Brownian*

motion by dividing the current sum S_n by the standard deviation of the sum S_N at the end of the trial. The time parameter t of $B(t)$ is the trial fraction $t = \text{var}(S_n)/\text{var}(S_N)$.

If the random variables are i.i.d., the same force that causes the z-statistic to be asymptotically standard normal—namely the central limit theorem—also causes the asymptotic joint distribution of B-values to be that of Brownian motion. In fact, the result holds even if the random variables are independent but not identically distributed (proof in Section 2.9.2).

Result 2.1 *Let S_N be a sum of independent (not necessarily identically distributed) random variables with mean 0, and let $n_i \rightarrow \infty$ and $N \rightarrow \infty$ such that $v_{n_i}/v_N \rightarrow t_i$, $i = 1, \dots, k$. Then the joint distribution of the B-values from Approach 1 is asymptotically that of Brownian motion if and only if the marginal distribution of the z-statistic is asymptotically standard normal.*

2.2 An Estimation Perspective

2.2.1 Information

In each scenario above, we were able to write the test statistic in terms of a sum $S_n = \sum_{i=1}^n D_i$, but testing whether the treatment effect is 0 is only one facet of inference; we are also interested in estimating the size of the treatment effect. Thus, we must determine the joint distribution of the treatment effect estimator $\hat{\delta}$ across different interim analyses. In the simplest setting, which involves paired data D_1, \dots, D_n , the treatment effect estimator $\hat{\delta}$ is a sample mean \bar{D} . The joint distribution of $\hat{\delta}_1, \dots, \hat{\delta}_k$ with n_1, \dots, n_k pairs is multivariate normal with marginal mean δ and covariance

$$\begin{aligned}
 \text{cov}(\hat{\delta}_i, \hat{\delta}_j) &= (n_i n_j)^{-1} \text{cov} \left(\sum_{r=1}^{n_i} D_r, \sum_{r=1}^{n_j} D_r \right) \\
 &= (n_i n_j)^{-1} \text{cov} \left(\sum_{r=1}^{n_i} D_r, \sum_{r=1}^{n_i} D_r + \sum_{r=n_i+1}^{n_j} D_r \right) \\
 &= (n_i n_j)^{-1} \left\{ \text{cov} \left(\sum_{r=1}^{n_i} D_r, \sum_{r=1}^{n_i} D_r \right) + \text{cov} \left(\sum_{r=1}^{n_i} D_r, \sum_{r=n_i+1}^{n_j} D_r \right) \right\} \\
 &= (n_i n_j)^{-1} \left\{ \text{var} \left(\sum_{r=1}^{n_i} D_r \right) + 0 \right\} \\
 &= (n_i n_j)^{-1} n_i \sigma^2 = \sigma^2 / n_j \\
 &= \text{var}(\hat{\delta}_j).
 \end{aligned} \tag{2.9}$$

Equation (2.9) shows the covariance of $\hat{\delta}$ over time when $\hat{\delta}$ is a sample mean; however, when the treatment and control sample sizes differ, the treatment

effect estimator $\hat{\delta} = \bar{Y}_T - \bar{X}_T$ is not a sample mean. Can we nonetheless view $\hat{\delta}$ as being *like* a sample mean even when the numbers n_T and n_C of treatment and control observations differ? If so, then a mean of how many observations? Let us assume that $\hat{\delta}$ behaves like a sample mean of, say, I i.i.d. observations with mean δ and variance 1. Then $E(\hat{\delta}) = \delta$ and $\text{var}(\hat{\delta}) = 1/I$. Solving for I yields

$$I = 1/\text{var}(\hat{\delta}). \quad (2.10)$$

Think of $\hat{\delta}$ as a sample mean and I as its sample size, even though I need not be an integer. Note that $\hat{\delta}$ has the same expectation and variance as a sample mean of I i.i.d. observations with mean δ and variance 1. We will show later that $\hat{\delta}$ computed at different interim analyses also has the same covariance as a sample mean computed at those analysis times. I defined by (2.10) is called the *information* contained in $\hat{\delta}$, which can be interpreted as the number of independent observations with expectation δ and variance 1 whose sample mean has the same precision as $\hat{\delta}$.

In the continuous outcome scenario with treatment and control sample sizes n_T and n_C , the information contained in $\hat{\delta} = \bar{Y} - \bar{X}$ is

$$I = \{\sigma^2(1/n_T + 1/n_C)\}^{-1} = n_T n_C / \{(n_T + n_C)\sigma^2\}.$$

I decreases as σ^2 increases, and for a fixed total sample size $n_T + n_C$, I increases as the disparity between n_T and n_C decreases.

Although information is interesting in its own right, we return to our goal of showing that $\hat{\delta}$ behaves like a sample mean of I i.i.d. random variables with mean δ and variance 1. We showed that this holds marginally, but we now show that the covariance over time of $\hat{\delta}$ is also that of a sample mean. The covariance over time for a sample mean was given by (2.9), which in view of (2.10) may be rewritten as

$$\text{cov}(\hat{\delta}_i, \hat{\delta}_j) = 1/I_j. \quad (2.11)$$

That is, the covariance between sample means at two different times is the inverse of the information at the later time.

Returning to the estimator $\hat{\delta} = \bar{Y} - \bar{X}$, let (n_{Ti}, n_{Ci}) and I_i be the (Treatment, Control) sample sizes and information, respectively, at the i th interim analysis. Then for $i \leq j$,

$$\begin{aligned} \text{cov}(\hat{\delta}_i, \hat{\delta}_j) &= \text{cov} \left\{ \frac{1}{n_{Ti}} \sum_{r=1}^{n_{Ti}} Y_r - \frac{1}{n_{Ci}} \sum_{r=1}^{n_{Ci}} X_r, \frac{1}{n_{Tj}} \sum_{r=1}^{n_{Tj}} Y_r - \frac{1}{n_{Cj}} \sum_{r=1}^{n_{Cj}} X_r \right\} \\ &= \frac{1}{n_{Ti} n_{Tj}} \text{cov} \left\{ \sum_{r=1}^{n_{Ti}} Y_r, \sum_{r=1}^{n_{Tj}} Y_r \right\} - \frac{1}{n_{Ti} n_{Cj}} \text{cov} \left\{ \sum_{r=1}^{n_{Ti}} Y_r, \sum_{r=1}^{n_{Cj}} X_r \right\} \\ &\quad - \frac{1}{n_{Ci} n_{Tj}} \text{cov} \left\{ \sum_{r=1}^{n_{Ci}} X_r, \sum_{r=1}^{n_{Tj}} Y_r \right\} + \frac{1}{n_{Ci} n_{Cj}} \text{cov} \left\{ \sum_{r=1}^{n_{Ci}} X_r, \sum_{r=1}^{n_{Cj}} X_r \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{n_{Ti}\sigma^2}{n_{Ti}n_{Tj}} - 0 - 0 + \frac{n_{Ci}\sigma^2}{n_{Ci}n_{Cj}} \\
&= \sigma^2 \left(\frac{1}{n_{Tj}} + \frac{1}{n_{Cj}} \right) \\
&= \text{var}(\hat{\delta}_j) \\
&= 1/I_j.
\end{aligned} \tag{2.12}$$

Equation (2.12) shows that, just as with a sample mean, the covariance of $\hat{\delta}$ computed at different times is the inverse of the information at the later time.

The same thing happens with binary data (Section 2.1.2), where the information in $\hat{\delta} = \hat{p}_T - \hat{p}_C$ is $\{p_T(1 - p_T)/n_T + p_C(1 - p_C)/n_C\}^{-1} = n_T n_C / \{n_C p_T(1 - p_T) + n_T p_C(1 - p_C)\}$. Again, (2.11) holds.

No estimator was immediately apparent for survival data (Section 2.1.3), but one was actually lurking in the background. For each i , $(O_i - E_i)/V_i$ is an estimate of the log hazard ratio (see the Statistical Appendix of Yusuf et al., 1985 [YPL85] for a heuristic justification of a closely related odds ratio estimate) with estimated variance $1/V_i$. We combine these uncorrelated estimates by weighting inversely proportionally to their variance:

$$\hat{\delta} = \frac{\sum_{r=1}^n V_r \{(O_r - E_r)/V_r\}}{\sum_{r=1}^n V_r} = S_n / \hat{v}_n,$$

where $S_n = \sum_r (O_r - E_r)$ and $\hat{v}_n = \sum_{r=1}^n V_r$ is an estimate of $v_n = \sum_{r=1}^n E(V_r)$. It can be shown that \hat{v}_n/n converges to a constant just as in Sections 2.1.1 and 2.1.2. Thus, we can treat \hat{v}_n as if it were v_n ;

$$\text{var}(\hat{\delta}) \approx v_n^{-2} \text{var}(S_n) = v_n^{-2} v_n = 1/v_n,$$

and information is approximately $I = v_n$, estimated by \hat{v}_n . Again $\hat{\delta}$ behaves like a mean of I i.i.d. observations with expectation δ and variance 1; $\hat{\delta}$ has mean δ and variance $1/I$. Furthermore, for $I_i = v_{n_i} \leq I_j = v_{n_j}$,

$$\begin{aligned}
\text{cov}(\hat{\delta}_i, \hat{\delta}_j) &= \text{cov} \left((1/I_i) \sum_{r=1}^{n_i} D_r, (1/I_j) \sum_{r=1}^{n_j} D_r \right) \\
&= (I_i I_j)^{-1} \text{cov} \left(\sum_{r=1}^{n_i} D_r, \sum_{r=1}^{n_i} D_r + \sum_{r=n_i+1}^{n_j} D_r \right) \\
&= (I_i I_j)^{-1} \left\{ \text{var} \left(\sum_{r=1}^{n_i} D_r \right) + \text{cov} \left(\sum_{r=1}^{n_i} D_r, \sum_{r=n_i+1}^{n_j} D_r \right) \right\} \\
&\approx (I_i I_j)^{-1} \{v_{n_i} + 0\} \\
&= (I_i I_j)^{-1} I_i \\
&= 1/I_j.
\end{aligned} \tag{2.13}$$

Equation (2.13) shows that the covariance of log hazard ratio estimators computed at two different times is the same as for a sample mean, namely the inverse of the information at the later time.

The reason for the \approx in the fourth line of the derivation of (2.13) is that we are no longer assuming the null hypothesis, and the D_r are not uncorrelated under the alternative hypothesis. Still, under a local alternative (loosely speaking, an alternative “near” the null hypothesis—see Section 2.9.4), the D_r are approximately uncorrelated.

2.2.2 Summary of Treatment Effect Estimators

With the t-test, the test of proportions, or the logrank test, the treatment effect estimator computed at k different interim analyses behaves just like cumulative sample means. It is cumbersome and vague to repeat each time we discuss estimation that the treatment effect estimator “behaves like” a sample mean of i.i.d. observations with expectation δ and variance 1. Instead, we follow the approach of Lan and Zucker (1993) [LZ93], spelling out precisely what we mean by “behaves like” a sample mean, and attaching a name to processes with these properties. Let τ be any measure of how far through the trial we are, scaled such that $\tau = 0$ and $\tau = 1$ at the beginning and end of the trial, respectively. For example, τ may be the calendar fraction (e.g., the 6-month point of a 5-year trial corresponds to $\tau = 1/10$). Let the increasing function $I(\tau)$ denote the information at time τ . What we mean when we say that $\hat{\delta}(\tau)$ “behaves like” a sample mean of $I(\tau)$ random variables with expectation δ and variance 1 is that $\hat{\delta}(\tau)$ satisfies—at least asymptotically—the following properties:

- E1: $\hat{\delta}(\tau_1), \dots, \hat{\delta}(\tau_k)$ have a multivariate normal distribution,
- E2: $E\{\hat{\delta}(\tau)\} = \delta$, and
- E3: $\text{cov}\{\hat{\delta}(\tau_i), \hat{\delta}(\tau_j)\} = \text{var}\{\hat{\delta}(\tau_j)\} = 1/I(\tau_j)$ for $i \leq j$.

Lan and Zucker called an estimator satisfying E1-E3 an *E-process* (E standing for estimator or estimation) with parameter δ and information function $I(\tau)$. An arguably better term might be *sample mean process* because properties E1-E3 are those of cumulative sample means of $I(\tau_1), \dots, I(\tau_k)$ observations. We will soon see that many other estimators are also E-processes.

2.3 Connection Between Estimators, Sums, Z-Scores, and Brownian Motion

Because the treatment effect estimator for the comparison of means, proportions, or log hazard ratios behaves like a sample mean of I i.i.d. random variables with expectation δ and variance 1, it stands to reason that $I\hat{\delta}$ should behave like a sum of I i.i.d. observations with expectation δ and variance 1. That is, if $\hat{\delta}(\tau)$ is an E-process, then $S(\tau) = I(\tau)\hat{\delta}(\tau)$ “behaves like” a sum of $I(\tau)$ i.i.d. random variables with expectation δ and variance 1. By “behaves like” a sum of $I(\tau)$ i.i.d. random variables with expectation δ and variance 1, we mean that $S(\tau)$ satisfies—at least asymptotically—

- S1: $S(\tau_1), \dots, S(\tau_k)$ have a multivariate normal distribution.
- S2: $E\{S(\tau)\} = I(\tau)\delta$.
- S3: For $\tau_i \leq \tau_j$, $\text{cov}\{S(\tau_i), S(\tau_j)\} = \text{var}\{S(\tau_i)\} = I(\tau_i)$.

Lan and Zucker (1993) [LZ93] termed $S(\tau)$ an *S-Process* because it behaves like a sum. The following result formalizes the notion that the estimator $\hat{\delta}(\tau)$ behaves like a sample mean if and only if $I(\tau)\hat{\delta}(\tau)$ behaves like a sum. We omit the straightforward proof.

Result 2.2 *If $\hat{\delta}$ is an unbiased estimator with information $0 < I(\tau) < \infty$ for $\tau > 0$, then $\hat{\delta}$ is an E-process iff $I(\tau)\hat{\delta}$ is an S-process.*

To emphasize that $I\hat{\delta}(\tau)$ behaves like a sum of $I(\tau)$ random variables, we use the more suggestive notation $S_{I(\tau)}$ for $I(\tau)\hat{\delta}(\tau)$. Because $S_{I(\tau)}$ behaves like a sum, we try to use Approach 1 to convert to Brownian motion, where $I(\tau)$ plays the role of the sample size. In Approach 1 we divide the current “sum” $S_{I(\tau)} = I(\tau)\hat{\delta}(\tau)$ by the standard deviation of the “sum” $S_{I(1)}$ at the end of the trial: $\{\text{var}(S_{I(1)})\}^{1/2} = \{I(1)\}^{1/2}$. The trial fraction and B-value are

$$\begin{aligned} t &= \text{var}\{S_{I(\tau)}\} / \text{var}\{S_{I(1)}\} \\ &= I(\tau) / I(1) \end{aligned} \quad (2.14)$$

and

$$B(t) = I(\tau)\hat{\delta}(\tau) / \{I(1)\}^{1/2}. \quad (2.15)$$

We call expression (2.14) the *information fraction* or *information time*. It is a generalization of the trial fraction, which was defined only for actual sums, not S-processes. Henceforth, we dispense with the notion of trial fraction in favor of the more general information fraction.

We next show that $B(t)$ defined by (2.15) has the properties of Brownian motion, except that its mean is not 0 under the alternative hypothesis. To see that $B(t)$ has the covariance structure of Brownian motion, note that for $t_i = I(\tau_i)/I(1) \leq t_j = I(\tau_j)/I(1)$,

$$\begin{aligned} \text{cov}\{B(t_i), B(t_j)\} &= \text{cov}[S_{I(\tau_i)} / \{I(1)\}^{1/2}, S_{I(\tau_j)} / \{I(1)\}^{1/2}] \\ &= \{I(1)\}^{-1} \text{cov}(S_{I(\tau_i)}, S_{I(\tau_j)}) \\ &= \{I(1)\}^{-1} I(\tau_i) \\ &= t_i. \end{aligned}$$

The mean of $B(t)$ is different from the mean under the null hypothesis. Under the alternative hypothesis,

$$\begin{aligned} E\{B(t)\} &= E[I(\tau)\hat{\delta}(\tau) / \{I(1)\}^{1/2}] \\ &= I(\tau)\delta / \{I(1)\}^{1/2} \\ &= [\{I(1)\}^{1/2}\delta] \{I(\tau)/I(1)\} \\ &= \theta t, \end{aligned}$$

where $\theta = \{I(1)\}^{1/2}\delta$ is the expected value of the z-score $\hat{\delta}(1)/[\text{var}\{\hat{\delta}(1)\}]^{1/2} = \{I(1)\}^{1/2}\hat{\delta}(1)$ at the end of the trial. $B(t)$ is said to be a Brownian motion with *drift* θ . The standard Brownian motion has drift 0.

Instead of beginning with the estimator $\hat{\delta}(\tau)$, transforming to a sum, then transforming to Brownian motion, we could have begun with the z-score $Z(t) = \hat{\delta}(\tau)/[\text{var}\{\hat{\delta}(\tau)\}]^{1/2} = \{I(\tau)\}^{1/2}\hat{\delta}(\tau)$ and multiplied by $t^{1/2} = \{I(\tau)/I(1)\}^{1/2}$ to obtain (2.15). We have essentially proven the following result.

Result 2.3 (*Summary*) *Let $I(\tau)/I(1)$ be the information fraction. We can convert an E-process, S-process, or Z-process to Brownian motion with drift θ , the expected value of the z-score at the end of the trial, as follows:*

$$\text{E to B : } B(t) = I(\tau)\hat{\delta}(\tau)/\{I(1)\}^{1/2}$$

$$\text{S to B : } B(t) = S(\tau)/\{I(1)\}^{1/2}$$

$$\text{Z to B : } B(t) = t^{1/2}Z(t).$$

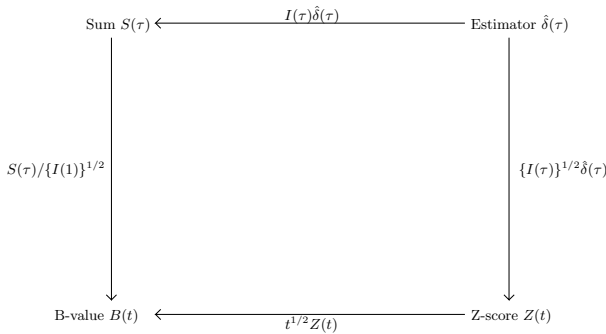


Fig. 2.2. Relationship between S-processes, E-processes, z-scores, and Brownian motion with drift θ , where θ is the expected value of the z-score at the end of the trial, $I(\tau)$ is the information function, and t is the information fraction $I(\tau)/I(1)$.

Figure 2.2 summarizes the relationships between S-processes, E-processes, z-scores, and Brownian motion.

Now that we are not restricting ourselves to the null hypothesis, we see the advantage of using the B-value instead of the z-score to monitor data. Because $E\{B(t)\} = \theta t$, it follows that $B(t)/t$ estimates the drift parameter, a simple transformation of the treatment effect estimate. Geometrically, $B(t)/t$ is the slope of the line joining the origin to $(t, B(t))$ (Figure 2.3). We can easily see whether the treatment effect estimate increases from one interim look to the next by seeing whether the slope of the line increases. Chapter 3 on conditional power uses the B-value approach extensively.

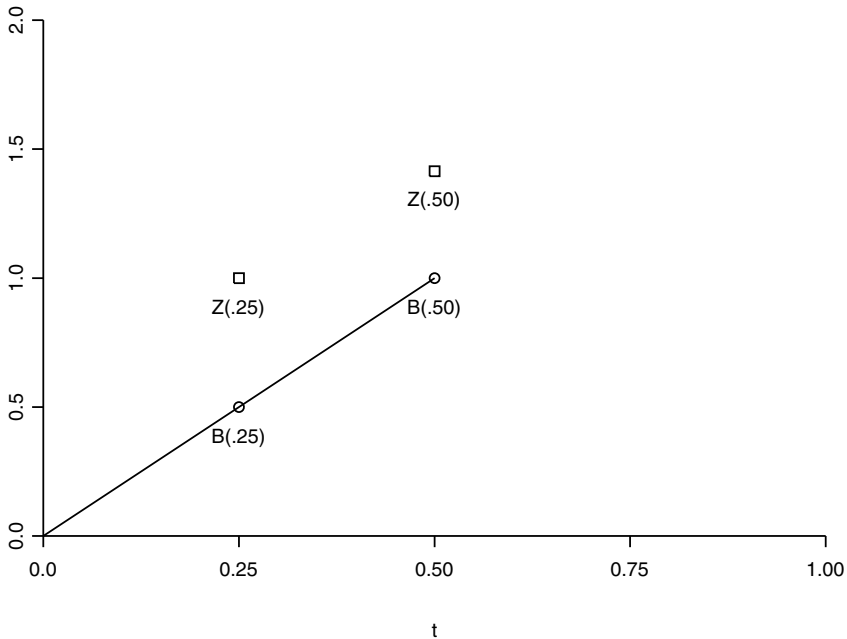


Fig. 2.3. Summarizing data with $B(t)$ instead of $Z(t)$ makes it easy to see whether results are improving over time. The slope of the line segment connecting the origin to $(t, B(t))$ is the drift parameter estimate, which is a simple transformation of the treatment effect estimate; the treatment effect estimate is larger at t_{i+1} than at t_i if and only if the slope of the line connecting the origin to $(t_{i+1}, B(t_{i+1}))$ is larger than the slope of the line connecting the origin to $(t_i, B(t_i))$. For the data shown in this graph, the line segments joining the origin to the circle at $(0.50, B(0.50))$ and the origin to the circle at $(0.25, B(0.25))$ have the same slope, so the treatment effect estimate at $t = 0.50$ is the same as at $t = 0.25$. Deducing this information from the z-scores (squares) is more difficult.

2.4 Maximum Likelihood Estimation

As discussed previously, many clinical trials use a difference in means or proportions to compare treatments; in other trials, the treatment effect is estimated by maximum likelihood in a model that adjusts for covariates. Analysis of covariance and logistic regression are the covariate-adjusted analogs of differences in means or proportions. To deal with these situations, assume that we have independent observations X_1, \dots, X_n from a distribution with density $f(x, \delta)$. We will show that the maximum likelihood estimator (MLE)

$\hat{\delta}$ of the treatment effect is asymptotically an E-process, and therefore can be converted to Brownian motion. This allows us to apply the results of Sections 2.1 through 2.3. In fact, as we shall demonstrate below, for the same set of information times, the monitoring boundaries for a trial that uses an MLE as the outcome are the same as the boundaries of the t-test, a test of proportions, or the logrank test.

First we review the arguments leading to asymptotic normality of an MLE at a single time point. Let $L(\delta)$ be the log likelihood function:

$$L(\delta) = \sum_{i=1}^n (\partial/\partial\delta) \{\ln f(X_i, \delta)\}.$$

Using a familiar technique, we expand the log likelihood in a Taylor series:

$$\begin{aligned} 0 = L(\hat{\delta}) &\approx L(\delta) + L'(\delta)(\hat{\delta} - \delta) \\ &= L(\delta) + \sum_{i=1}^n (\partial^2/\partial\delta^2) \{\ln f(X_i, \delta)\}(\hat{\delta} - \delta), \end{aligned}$$

and hence

$$\begin{aligned} \hat{\delta} - \delta &\approx \frac{-L(\delta)}{\sum_{i=1}^n (\partial^2/\partial\delta^2) \{\ln f(X_i, \delta)\}} \\ &= \frac{\sum_{i=1}^n (\partial/\partial\delta) \{\ln f(X_i, \delta)\}}{\sum_{i=1}^n -(\partial^2/\partial\delta^2) \{\ln f(X_i, \delta)\}} \\ &\approx \frac{\sum_{i=1}^n (\partial/\partial\delta) \{\ln f(X_i, \delta)\}}{\mathcal{I}_n}. \end{aligned} \tag{2.16}$$

In the last step, we replaced the denominator by its expectation, $\mathcal{I}_n = -nE[(\partial^2/\partial\delta^2) \{\ln f(X, \delta)\}]$, the Fisher information contained in X_1, \dots, X_n . Multiplying both sides of (2.16) by \mathcal{I}_n results in

$$\mathcal{I}_n(\hat{\delta} - \delta) = S_n + R_n, \tag{2.17}$$

where $S_n = L(\delta) = \sum_{i=1}^n (\partial/\partial\delta) \{\ln f(X_i, \delta)\}$ is a sum of i.i.d. mean 0 random variables and R_n is a remainder term. It is not difficult to show that, under mild conditions, $\text{var}(S_n) = \mathcal{I}_n$. Thus, from (2.17),

$$\begin{aligned} \frac{\mathcal{I}_n(\hat{\delta} - \delta)}{\mathcal{I}_n^{1/2}} &= \frac{S_n}{\mathcal{I}_n^{1/2}} + \frac{R_n}{\mathcal{I}_n^{1/2}} \\ \mathcal{I}_n^{1/2}(\hat{\delta} - \delta) &= \frac{S_n}{\sqrt{\text{var}(S_n)}} + \mathcal{I}_n^{-1/2} R_n. \end{aligned} \tag{2.18}$$

The first term on the right side of (2.18) is asymptotically standard normal by the central limit theorem, while the second term tends to 0 in probability

under regularity conditions, so $\mathcal{I}_n^{1/2}(\hat{\delta} - \delta)$ is asymptotically standard normal. In other words, $\hat{\delta}$ is asymptotically normal with mean δ and variance $1/\mathcal{I}_n$. Marginally at least, $\hat{\delta}$ and $\mathcal{I}_n\hat{\delta}$ behave like an E-process and S-process, respectively, with mean δ and information $I_n = \mathcal{I}_n$.

Now consider the MLE monitored over time. Equation (2.17) shows that $\mathcal{I}_n(\hat{\delta} - \delta)$ is essentially a sum, and Approach 1 suggests we can convert it to Brownian motion by dividing by the standard deviation of the sum at the end of the trial, $I_N^{1/2} = \{\text{var}(S_N)\}^{1/2}$. Let $\hat{\delta}_i$ denote the MLE at look i , $i = 1, \dots, k$. By (2.17),

$$\frac{\mathcal{I}_{n_i}(\hat{\delta}_i - \delta)}{I_N^{1/2}} = \frac{S_{n_i}}{\sqrt{\text{var}(S_N)}} + I_N^{-1/2}R_{n_i}. \quad (2.19)$$

Now let $n_i \rightarrow \infty$ and $N \rightarrow \infty$ such that $n_i/N \rightarrow t_i$, $i = 1, \dots, k$. Each remainder term $I_N^{-1/2}R_{n_i}$ of (2.19) converges to 0 in probability because

$$\begin{aligned} I_N^{-1/2}R_{n_i} &= (I_{n_i}/I_N)^{1/2}I_{n_i}^{-1/2}R_{n_i} \\ &= (n_i/N)^{1/2}I_{n_i}^{-1/2}R_{n_i} \\ &\rightarrow (t_i^{1/2})(0) = 0 \end{aligned}$$

in probability. Thus, $\mathcal{I}_{n_1}(\hat{\delta}_1 - \delta)/I_N^{1/2}, \dots, \mathcal{I}_{n_k}(\hat{\delta}_k - \delta)/I_N^{1/2}$ behaves asymptotically like $S_{n_1}/\{\text{var}(S_N)\}^{1/2}, \dots, S_{n_k}/\{\text{var}(S_N)\}^{1/2}$, which, in turn, behaves asymptotically like standard Brownian motion by Result 2.1 and the central limit theorem. Note that we can rewrite $\mathcal{I}_{n_i}(\hat{\delta}_i - \delta)/\mathcal{I}_N^{1/2}$ as $t_i^{1/2}(\hat{\delta}_i - \delta)/\hat{\sigma}_{\hat{\delta}_i}$.

In summary:

Result 2.4 (*Brownian motion for MLEs with i.i.d. data*) Let X_j be i.i.d. with density $f(x_j; \delta)$, and let $\hat{\delta}_i$ and $\hat{\sigma}_{\hat{\delta}_i}$ denote the MLE and its estimated standard error, respectively, after n_i patients are evaluated, $i = 1, \dots, k$. Suppose that $n_i \rightarrow \infty$ and $N \rightarrow \infty$ such that $n_i/N \rightarrow t_i$. Under the same regularity conditions that imply marginal asymptotic normality of the MLE, $t_1^{1/2}(\hat{\delta}_1 - \delta)/\hat{\sigma}_{\hat{\delta}_1}, \dots, t_k^{1/2}(\hat{\delta}_k - \delta)/\hat{\sigma}_{\hat{\delta}_k}$ have the asymptotic distribution of standard Brownian motion at t_1, \dots, t_k . Equivalently, the B-values $B(t_i) = t_i^{1/2}\hat{\delta}_i/\hat{\sigma}_{\hat{\delta}_i}$ behave approximately like Brownian motion with drift θ , where $\theta = I_N^{1/2}\delta$ is the expected z-score at the end of the trial.

Essentially the same arguments leading to Result 2.4 can be used even if the underlying observations X_i are independent but not identically distributed because Result 2.1 does not require identical distributions. A result analogous to Result 2.4 holds when the parameter is a vector (Jennison and Turnbull, 1997 [JT97] or Jennison and Turnbull, 2000 [JT00]).

Example 2.4. Consider a trial in which the outcome was the presence of at least one episode of cardiac ischemia on a Holter monitor—a device recording

the electrical activity of the heart over a 24-hour period—12 weeks following randomization. Patients were also monitored with the Holter at baseline, and investigators wanted to use logistic regression to adjust the 12-week results for differences in the baseline number of ischemic episodes. The model is

$$\ln\{p/(1-p)\} = \alpha + \beta u + \delta x,$$

where p is the probability of having ischemia at 12 weeks, u is the baseline number of episodes, and x is the treatment indicator. We parameterize such that positive z-scores indicate that the treatment is beneficial, so we take $x = 1$ to mean the control condition. We are interested in testing whether $\delta = 0$ (no treatment effect). After 200 of the planned 600 patients are evaluated, the estimated information fraction is $t = 200/600 = 1/3$. For simplicity, rather than using two different time scales τ and t for calendar fraction and information fraction, we use only information fraction. Thus, we denote the current treatment effect estimator and its estimated standard error by $\hat{\delta}(1/3)$ and $\hat{\sigma}_{\hat{\delta}(1/3)}$. Suppose that $\hat{\delta}(1/3) = 0.180$ and $\hat{\sigma}_{\hat{\delta}(1/3)} = 0.153$. The z-score and B-value are $Z(1/3) = 0.180/0.153 = 1.176$ and $B(1/3) = (1/3)^{1/2}(1.176) = 0.679$. Because $Z(1/3)$ has a standard normal distribution under the null hypothesis, we can easily determine a critical value c_1 such that $P_0\{|Z(1/3)| \geq c_1\} = 0.01$, where P_0 denotes a probability computed under the null hypothesis. We find that $c_1 = 2.576$. Suppose that at the end of the trial, the estimated slope and standard error are $\hat{\delta}(1) = 0.120$ and $\hat{\sigma}_{\hat{\delta}(1)} = 0.095$. The approximate joint distribution of the interim and final B-values under true log odds ratio δ is that of $B(1/3)$ and $B(1)$, where $B(t)$ is Brownian motion with drift $\theta = E\{Z(1)\} = \delta/\sigma_{\hat{\delta}(1)}$. We estimate θ by $\delta/0.095$, where δ is the true log odds ratio.

Having reached the end of the trial, we can obtain a more precise estimate of the information fraction at the first look: $t_1 = \{\text{var}(\hat{\delta}(\frac{1}{3}))\}^{-1}/\{\text{var}(\hat{\delta}(1))\}^{-1} = (0.153)^{-2}/(0.095)^{-2} = 0.386$ rather than $1/3$. Thus, the approximate joint distribution of the interim and final B-values is that of $B(0.386)$ and $B(1)$, where $B(t)$ is Brownian motion with drift θ . As we have seen before, this correcting of information fractions does not cause a problem for previous boundaries because the z-score at previous analyses has the same null distribution whether or not we correct the information times. Thus, the correct null probability of crossing the boundary at the first look, $P_0\{|Z(0.386)| \geq 2.576\} = 0.01$, is the same as $P_0\{|Z(1/3)| \geq 2.576\}$. The advantage of using the slightly more accurate estimate $t_1 = 0.386$ lies in computation of the boundary at the next look at the end of the trial. We determine c_2 such that $P_0\{(|Z(0.386)| \geq 2.576) \cup (|Z(1)| \geq c_2)\} = 0.05$. Numerical integration can be used to obtain $c_2 = 2.014$.

Importantly, the boundaries $c_1 = 2.576$ and $c_2 = 2.014$ for the z-scores associated with the MLE are the same as for a t-test, test of proportions, or logrank test at information fractions $t_1 = 0.386$ and $t_2 = 1$.

2.5 Other Settings Leading to E-Processes and Brownian Motion

We have seen that many estimators frequently used in clinical trials are E-processes when monitored over time. Other broad classes of estimators monitored over time are also E-processes, and can therefore be transformed to Brownian motion using Result 2.3. Sometimes it is possible to argue directly that $\hat{\delta}(\tau)$ satisfies E3, as we now show.

2.5.1 Minimum Variance Unbiased Estimators

Consider a minimum variance unbiased estimator $\hat{\delta}$ in a nonmonitoring setting (i.e., $\hat{\delta}$ has the smallest variance among all unbiased estimators of δ). Let $\hat{\delta}(\tau)$ denote the corresponding minimum variance unbiased estimator monitored over time τ , $0 \leq \tau \leq 1$. Jennison and Turnbull (1997) [JT97] gave a simple argument by contradiction that $\hat{\delta}$ must satisfy E3. Note first that condition E3 can be written in the equivalent way

$$\begin{aligned} 0 &= 1/I(\tau_j) - \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_i)\} \\ &= \text{var}\{\hat{\delta}(\tau_j)\} - \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_i)\} \\ &= \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j)\} - \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_i)\} \\ &= \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\}. \end{aligned} \tag{2.20}$$

Thus, E3 is equivalent to

$$\text{E3}' : \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} = 0.$$

Suppose E3' did not hold for a minimum variance unbiased estimator $\hat{\delta}$ monitored over time. For example, suppose that $\text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} > 0$. Jennison and Turnbull argued that for small $\epsilon > 0$, the estimator $\tilde{\delta}_\epsilon(\tau_j) = \hat{\delta}(\tau_j) - \epsilon\{\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\}$ has smaller variance than $\hat{\delta}(\tau_j)$. To see this, note that $\text{var}\{\tilde{\delta}_\epsilon(\tau_j)\} = \text{var}\{\hat{\delta}(\tau_j)\} + \epsilon^2 \text{var}\{\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} - 2\epsilon \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\}$, so

$$\begin{aligned} &\lim_{\epsilon \rightarrow 0} [\text{var}\{\tilde{\delta}_\epsilon(\tau_j)\} - \text{var}\{\hat{\delta}(\tau_j)\}] / \epsilon \\ &= \lim_{\epsilon \rightarrow 0} \epsilon \text{var}\{\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} - 2 \lim_{\epsilon \rightarrow 0} \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} \\ &= 0 - 2 \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} \\ &= -2 \text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} \\ &< 0. \end{aligned}$$

But this implies that $\text{var}\{\tilde{\delta}_\epsilon(\tau_j)\} < \text{var}\{\hat{\delta}(\tau_j)\}$ for sufficiently small ϵ , which contradicts the fact that $\hat{\delta}(\tau_j)$ is a minimum variance unbiased estimator. Similarly, if we had begun with the assumption that $\text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} < 0$, we could show that the estimator $\hat{\delta}(\tau_j) + \epsilon\{\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\}$ has smaller variance

than the minimum variance unbiased estimator $\hat{\delta}(\tau_j)$ for sufficiently small ϵ . This would again be a contradiction. In other words, we can find a contradiction whenever $\text{cov}\{\hat{\delta}(\tau_j), \hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)\} \neq 0$. Thus, property E3', and therefore E3, must hold.

Implicit in the above argument is the assumption that $\text{cov}\{\hat{\delta}(\tau_i), \hat{\delta}(\tau_j)\}$ does not depend on the parameter δ . If it did, then $\tilde{\delta}_\epsilon(\tau_j)$ would depend on δ . The fact that it has smaller variance than $\hat{\delta}(\tau_j)$ would not cause a contradiction because $\tilde{\delta}_\epsilon(\tau_j)$ would not be a bona-fide estimator. The arguments above prove the following result.

Result 2.5 *Let $\hat{\delta}$ be a minimum variance unbiased estimator of δ in a non-monitoring setting, and let $\hat{\delta}(\tau)$ denote $\hat{\delta}$ monitored at time τ , $0 \leq \tau \leq 1$. If $\text{cov}\{\hat{\delta}(\tau_i), \hat{\delta}(\tau_j)\}$ does not depend on δ for any $\tau_i < \tau_j$, then $\hat{\delta}(\tau)$ satisfies E3.*

While Result 2.5 does not establish condition E1 (multivariate normality) for a minimum variance unbiased estimator $\hat{\delta}(\tau)$ over time, it does show that $\hat{\delta}(\tau)$ must have the same mean and covariance structure of an E-process. Thus, if we can establish through other arguments that $\hat{\delta}(\tau)$ has an approximate multivariate normal distribution, we can convert to Brownian motion as we did for other estimators.

2.5.2 Complete Sufficient Statistics

This subsection concerns complete sufficient statistics, so we briefly review the concepts of sufficiency and completeness. If a vector (X_1, \dots, X_n) of observations has distribution function $F(x_1, \dots, x_n; \delta)$ depending on a parameter δ , a statistic $S(X_1, \dots, X_n)$ (which could be a vector) is called *sufficient* if the conditional distribution of the data X_1, \dots, X_n given $S = s$ does not depend on δ . We could generate data X_1, \dots, X_n from $F(x_1, \dots, x_n; \delta)$ by first generating a value of S from its distribution—which depends on δ —and then generating (X_1, \dots, X_n) from its conditional distribution given $S = s$. The latter generation is a random draw of n numbers from a distribution that has nothing to do with δ . In that sense, once we condition on the value of the sufficient statistic S , no further information about δ can be gleaned from the data.

A statistic S is called *complete* if the condition $E\{f(S)\} = 0$ for all δ implies that $f(S) = 0$ with probability 1 for all δ . Completeness is typically used to show that there is at most one unbiased function of S , for if both $g_1(S)$ and $g_2(S)$ were unbiased for δ , then $E\{g_2(S) - g_1(S)\} = 0$, which would mean that $g_2(S) - g_1(S) = 0$; i.e., $g_2(S) = g_1(S)$ with probability 1 for all δ .

We now consider condition E3 of an E-process and relate it to a complete sufficient statistic. By (2.20), we can consider the equivalent condition E3'. Note that condition E3' would be satisfied if $\hat{\delta}(\tau_j)$ and $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$ were independent. Moreover, under E1, E3' is equivalent to $\hat{\delta}(\tau_j)$ being independent of $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$. Independence of $\hat{\delta}(\tau_j)$ and $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$ is in some sense

natural. Think of the comparison of means: $\hat{\delta}(\tau_j)$ is complete and sufficient for δ , whereas $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$ is *ancillary*, meaning that its distribution does not depend on δ . In a sense, $\hat{\delta}_j$ and $\hat{\delta}_j - \hat{\delta}_i$ contain all of the information and none of the information, respectively, about δ . Not surprisingly, $\hat{\delta}(\tau_j)$ is independent of $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$. In fact, this is a special case of a beautiful theorem due to Basu (1955) [B55]. Basu's theorem states that if $\hat{\delta}$ is sufficient and complete and A is ancillary, then $\hat{\delta}$ and A are independent (see Section 2.9.5 for proof). Thus, condition E3 will hold for any complete sufficient statistic such that $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$ is ancillary.

Result 2.6 *Let $\hat{\delta}$ be a complete sufficient statistic in a nonmonitoring setting, and let $\hat{\delta}(\tau)$ denote $\hat{\delta}$ monitored over time, $0 \leq \tau \leq 1$. If $\hat{\delta}(\tau_j) - \hat{\delta}(\tau_i)$ is ancillary for every $\tau_i \leq \tau_j$, then*

1. E3 holds.
2. E1 holds iff $\hat{\delta}$ is marginally normal.
3. E1 holds asymptotically iff $\hat{\delta}$ is asymptotically marginally normal.

2.6 The Normal Linear and Mixed Models

2.6.1 The Linear Model

Some clinical trials analyze results using a normal linear model. For example, in the nonmonitoring setting, the analysis of covariance model that adjusts the end of study blood pressure Y for baseline blood pressure x may be written as

$$\underline{Y} = \begin{pmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_n \\ 1 & 1 & x_{n+1} \\ \vdots & \vdots & \vdots \\ 1 & 1 & x_{2n} \end{pmatrix} \begin{pmatrix} \alpha_C \\ \delta \\ \lambda \end{pmatrix} + \underline{\epsilon},$$

where α_C is the intercept in the control arm, $\delta = \alpha_T - \alpha_C$ is the difference between treatment and control intercepts (i.e., δ is the treatment effect), and λ is the slope—assumed the same in the treatment and control arms—of the relationship between baseline and end of study blood pressure. (Y_1, \dots, Y_n and Y_{n+1}, \dots, Y_{2n} are end-of-study blood pressures for control and treatment patients, respectively.) More generally, the normal linear model may be written as $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$, where X is a design matrix of dimension $n \times p$, $\underline{\beta}$ is a p -dimensional parameter vector, and $\underline{\epsilon}$ is an n -dimensional vector of i.i.d. $N(0, \sigma^2)$ errors.

Now consider monitoring. At the first interim analysis with n_1 observations per arm, the dimension of \underline{Y} and the number of rows of the design matrix is

$2n_1$. At future interim analyses, \underline{Y} will be appended by additional observations and the design matrix will be appended by additional rows; each new patient contributes a new Y and a new row to the design matrix.

We now argue that the treatment effect estimator is an E-process. To see this, assume for the moment that σ^2 is known. The least squares estimators at different interim analyses are linear combinations of the Y s, and therefore have a multivariate normal distribution. Furthermore, in a nonmonitoring setting, the least squares estimator $\underline{\hat{\beta}}$ is complete and sufficient (Arnold, 1981 [A81] contains a similar result when σ^2 is unknown). Moreover, if $\underline{\hat{\beta}}(\tau_i)$ and $\underline{\hat{\beta}}(\tau_j)$ denote the least squares estimators at interim analyses at times τ_i and τ_j , then $\underline{\hat{\beta}}(\tau_j) - \underline{\hat{\beta}}(\tau_i)$ is ancillary because it has a multivariate normal distribution with zero mean vector (because both estimators are unbiased) and covariance matrix not depending on $\underline{\beta}$. It follows from Basu's theorem that $\underline{\hat{\beta}}(\tau_j)$ and $\underline{\hat{\beta}}(\tau_j) - \underline{\hat{\beta}}(\tau_i)$ are independent.

Now consider the case when σ^2 is unknown. The least squares estimator $\underline{\hat{\beta}}$ is exactly the same as in the case of known σ^2 . It follows that $\underline{\hat{\beta}}(\tau_j)$ and $\underline{\hat{\beta}}(\tau_j) - \underline{\hat{\beta}}(\tau_i)$ are independent in the case of unknown σ^2 as well. In summary:

Result 2.7 *In the normal linear model, $\underline{\hat{\beta}}(\tau_1), \dots, \underline{\hat{\beta}}(\tau_k)$ are multivariate normal and $\underline{\hat{\beta}}(\tau_j)$ is independent of $\underline{\hat{\beta}}(\tau_j) - \underline{\hat{\beta}}(\tau_i)$, $i = 2, \dots, k$. Consequently, the treatment effect estimator, its associated z-score, and its associated B-value behave like E-, Z-, and B-processes, respectively.*

A consequence of Result 2.7 is that we may use the same boundaries for the z-scores (treatment effect estimators divided by their standard errors) from a linear model that we used for the t-test.

2.6.2 The Mixed Model

Thus far we have dealt with either independent or uncorrelated observations Y , but sometimes data from clinical trials are correlated. Common examples are trials with continuous, longitudinal data reflecting each patient's progression of disease over time. For example, the model for an observation Y_{ij} at time x_j for patient i might be

$$Y_{ij} = \alpha_C + \beta_C x_j + (\gamma + \delta x_j)u_i + a_i + b_i x_j + \epsilon_{ij}, \quad (2.21)$$

where α_C and β_C are the mean intercept and slope in the control arm, $u_i = 0, 1$ is the treatment indicator, $\gamma = \alpha_T - \alpha_C$ and $\delta = \beta_T - \beta_C$ are differences between treatment and control mean intercepts and slopes, respectively, and a_i and b_i are random, patient-specific intercepts and slopes. The patient-specific intercepts reflect the fact that patients have different baseline values, whereas the patient-specific slopes measure the patients' improvement or deterioration over time. The quantity $\gamma = \alpha_T - \alpha_C$ reflects the between-arm difference in

baseline values. The parameter on which we gauge the success of the treatment is the between-arm difference in slopes, $\delta = \beta_T - \beta_C$.

More generally, for an arbitrary mixed model, the observation vector \underline{Y} is normally distributed with mean vector $X\underline{\beta}$, where $\underline{\beta} = (\beta_1, \dots, \beta_p)^T$ is the vector of fixed effects and X its design matrix. The design matrix is similar to that of the linear model of the preceding subsection except that each patient contributes multiple rows. Each additional time point for a patient contributes a new row to the design matrix.

Now consider monitoring. Both the number of patients and the number of time points per patient may differ from one interim analysis to the next. The effect of incorporating data between successive analyses is to append observations to the \underline{Y} vector and rows to the design matrix. Observations from one patient to the next are independent, but observations on the same patient over time are correlated. Nonetheless, we shall see that the Brownian motion paradigm still holds if the covariance matrix of \underline{Y} is known.

In the known Σ case, we can transform to the model of the preceding subsection:

$$\begin{aligned}\Sigma^{-1/2}\underline{Y} &= \Sigma^{-1/2}\underline{\beta} + \Sigma^{-1/2}\underline{\epsilon} \\ \underline{Y}' &= X'\underline{\beta} + \underline{\epsilon}',\end{aligned}$$

where $\underline{\epsilon}' = \Sigma^{-1/2}\underline{\epsilon}$ has covariance matrix $\Sigma^{-1/2}\Sigma\Sigma^{-1/2} = I$ (Arnold, 1981 [A81]). As noted earlier, the least squares estimator in this transformed model is complete and sufficient. The arguments of the preceding subsection imply that $\hat{\underline{\beta}}(\tau_j)$ and $\hat{\underline{\beta}}(\tau_j) - \hat{\underline{\beta}}(\tau_i)$ are independent.

Result 2.8 *Result 2.7 holds for the mixed model if the covariance between every pair of Y observations is known.*

A similar result holds in the unknown covariance case provided that the number of distinct covariances to be estimated is small compared to the number of participants.

Result 2.8 means that the null joint distribution of z-statistics (treatment effect estimates divided by their estimated standard errors) at different information fractions $t_i = [\text{var}\{\hat{\delta}(\tau_i)\}]^{-1}/[\text{var}\{\hat{\delta}(1)\}]^{-1}$ in a trial analyzed with a mixed model is the same as for a simple t-test. Therefore, any z-score boundaries developed for continuous outcome trials can be applied to trials employing a mixed model. We have not yet addressed how to estimate $\text{var}\{\hat{\delta}(1)\}$, but as we saw for other clinical trial scenarios, accurate estimation of $\text{var}\{\hat{\delta}(1)\}$ is not important for calculating probabilities under the null hypothesis. Accurate estimation of $\text{var}\{\hat{\delta}(1)\}$ does become important for probability calculations assuming the alternative hypothesis.

Because tests of treatment effects from mixed models are more complicated than t-tests and tests of proportions, we give a more detailed explanation of probability calculations assuming the alternative hypothesis is true. To use Brownian motion we must know the drift parameter $\theta = E\{Z(1)\}$, which

means we must have a representation for the z-statistic at the end of the trial, $Z(1)$. Suppose participant i has observations at M_i time points x_{i1}, \dots, x_{iM_i} by the time the trial ends, and let $\bar{x}_i(1) = (1/M_i) \sum_{j=1}^{M_i} x_{ij}$. Assuming the M_i are similar across participants (which they typically are in trials that use longitudinal models), the z-statistic using the mixed model is approximately the same as the z-statistic for a test of means applied to participants' end of study least squares slope estimates $b_i(1)$,

$$\hat{b}_i(1) = SSXY_i(1)/SSX_i(1),$$

where $SSXY_i(1) = \sum_{j=1}^{M_i} (x_{ij} - \bar{x}_i(1))(Y_{ij} - \bar{Y}_i(1))$ and $SSX_i(1) = \sum (x_{ij} - \bar{x}_i(1))^2$. The expected z-score at the end of the trial is roughly $E\{Z(1)\} = \delta/[2\text{var}\{\hat{b}_i(1)\}/N]^{1/2}$, where δ is the difference between treatment and control population mean slopes. We can determine the variance of $\hat{b}_i(1)$ by first conditioning on the patient's true intercept and slope, a_i and b_i , and then using the formula $\text{var}(U | \underline{V}) = E\{\text{var}(U | \underline{V})\} + \text{var}\{E(U | \underline{V})\}$, valid for any random variable U (with finite variance) and random vector \underline{V} . The conditional variance of $\hat{b}_i(1)$ given the patient-specific intercept and slope is $\sigma_e^2/SSX_i(1)$. The unconditional variance of $\hat{b}_i(1)$ is $E\{\text{var}(\hat{b}_i(1) | a_i, b_i)\} + \text{var}\{E(\hat{b}_i(1) | a_i, b_i)\} = \sigma_e^2/SSX_i + \sigma_b^2$, where σ_e^2 is the within-patient residual variability about his/her regression line and σ_b^2 is the variability of the patient-specific true slopes b_i . We can estimate σ_e^2 and σ_b^2 from the data at an interim analysis. For example, suppose at an interim analysis at information fraction t (we will show how to estimate the information fraction shortly) there are n patients, patient i having measurements at times x_{i1}, \dots, x_{im} , and let $\bar{x}_i(t) = (1/m_i) \sum_{j=1}^{m_i} x_{ij}$. Then $\text{var}\{\hat{b}_i(t)\} = \sigma_e^2/SSX_i(t) + \sigma_b^2$. Averaging over the number of patients gives us an estimate of $\text{var}\{\hat{b}(t)\}$ for a randomly selected patient:

$$\text{var}\{\hat{b}(t)\} = \sigma_e^2(1/n) \sum_{i=1}^n 1/SSX_i(t) + \sigma_b^2. \quad (2.22)$$

We can estimate σ_e^2 as follows. For patient i , we perform least squares regression and compute the residual sum of squares $RSS_i(t) = \sum_{j=1}^{m_i} \{Y_{ij} - (\hat{a}_i + \hat{b}_i x_{ij})\}^2$. We estimate σ_e^2 by pooling over patients:

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^n RSS_i(t)}{\sum_{i=1}^n (m_i - 2)}. \quad (2.23)$$

We can substitute this $\hat{\sigma}_e^2$ into the right side of (2.22) and the sample variance of the \hat{b}_i pooled across arms into the left side. We estimate σ_b^2 by subtraction: pooled $\text{var}(\hat{b}_i) - \hat{\sigma}_e^2(1/n) \sum_{i=1}^n 1/SSX_i$. Because this estimate can be negative, we take the maximum of this estimate and 0:

$$\hat{\sigma}_b^2 = \max\left(0, \text{pooled var}(\hat{b}_i) - \hat{\sigma}_e^2(1/n) \sum_{i=1}^n 1/SSX_i(t)\right), \quad (2.24)$$

where $\hat{\sigma}_e^2$ is given by (2.23). We use $\hat{\sigma}_e^2$ and $\hat{\sigma}_b^2$ to estimate $\text{var}\{\hat{b}(1)\}$. Once we have $\text{var}\{\hat{b}(1)\}$, we estimate $\text{var}\{\hat{\delta}(1)\}$ and $I(1)$ by

$$\hat{\text{var}}\{\hat{\delta}(1)\} = 2\text{var}\{\hat{b}(1)\}/N, \quad I(1) = 1/\hat{\text{var}}\{\hat{\delta}(1)\}$$

The current information is much easier. It is simply the inverse of the variance of the treatment effect estimator at the interim analysis, which we can compute from the standard output of a mixed model program. Specific details of these calculations are given in the following example:

Example 2.5. Consider a trial randomizing overweight patients to an advice-only control arm versus a treatment arm with advice plus an exercise program. Each patient is followed for 7 weeks. Weights are recorded at baseline and weekly thereafter (eight weights total). Data are analyzed according to the mixed model (2.21), where Y_{ij} is the weight of participant i at week x_{ij} (week 0 denotes the baseline period). The planned sample size is 80 patients, 40 in each treatment group.

The interim analysis is to include data from the first 4 weeks of follow-up for the first 20 patients randomized. Table 2.1 shows the data for this cohort. The weights \underline{y}_i for participant i are regressed on the participant's times \underline{x}_i , and a least squares line is fit. The table shows, for each participant's data \underline{x}_i , \underline{y}_i , the slope estimate \hat{b}_i , the residual sum of squares RSS_i , $1/SSX_i = [\sum_j \{x_{ij} - (1/m_i) \sum_{r=1}^{m_i} x_{ir}\}^2]^{-1}$, and the degrees of freedom $m_i - 2$. We estimate σ_e^2 by $\sum_{i=1}^{20} RSS_i / \sum_{i=1}^{20} (m_i - 2)$. From Table 2.1, $\sum_{i=1}^{20} RSS_i = 45.6373$ and $\sum_{i=1}^{20} (m_i - 2) = 42$, so

$$\hat{\sigma}_e^2 = 45.6373/42 = 1.0866.$$

The sample variances of the slopes in column 5 for control and treatment patients are 1.1628 and 1.4957, for a pooled variance of $\{9(1.1628) + 9(1.4957)\}/18 = 1.3293$. From (2.24), we estimate σ_b^2 by

$$\hat{\sigma}_b^2 = 1.3293 - 1.0866(1/20)(2.2666) = 1.2062.$$

At the end of the trial, participants will have data for a maximum of 8 weeks, though some data may be missing. At the interim analysis, everyone had a baseline value, but 18 of the 80 possible follow-up weights for the 20 participants were missing (22.5 percent). If we assume the same percentage missing for the seven follow-up weights by the end of the trial as for the follow-up weights thus far, participants will have an average of $0.225(7) = 1.575$ missing observations among the 7 follow-up weeks. Thus, the average participant will have one baseline measurement and $7 - 1.575 = 5.425$ follow-up measurements, for a total of 6.425 measurements. The variance of x values (using M instead of $M - 1$ in the denominator) for a participant with no missing data will be $(1/8) \sum_{j=0}^7 \{j - (0 + 1 + \dots + 7)/8\}^2 = 5.25$. We expect

Table 2.1. Interim data from a trial using a mixed model. Twenty patients have been randomized, and up to five measurements (the baseline and first four follow-up measurements) are available. For participant i , vector \underline{x}_i is the number of weeks since randomization and \underline{y}_i contains the weights at weeks \underline{x}_i . Ordinary least squares regression is used for each participant's data, and the intercept \hat{a}_i and slope \hat{b}_i are computed. Shown are the slope estimate \hat{b}_i , $RSS_i = \sum_j (y_{ij} - \hat{a}_i - \hat{b}_i x_{ij})^2$, and $SSX_i = \sum_j (x_{ij} - (1/m_i) \sum_{j=1}^{m_i} x_i)^2$, where m_i is the number of observations per participant. Also shown are the degrees of freedom for each participant, namely $df_i = m_i - 2$.

Patient	Arm	\underline{x}_i	\underline{y}_i	\hat{b}_i	RSS_i	$1/SSX_i$	$m_i - 2$
1	C	(0, 2, 3, 4)	(248, 250, 251, 251)	0.8000	0.4000	0.1143	2
2	C	(0, 1, 2, 3, 4)	(216, 214, 215, 214, 213)	-0.6000	1.6000	0.1000	3
3	C	(0, 1, 2, 3, 4)	(217, 218, 220, 217, 216)	-0.3000	8.3000	0.1000	3
4	C	(0, 1, 4)	(195, 195, 191)	-1.0769	0.6154	0.1154	1
5	C	(0, 2, 3, 4)	(197, 200, 200, 199)	0.5714	3.1429	0.1143	2
6	C	(0, 1, 2, 3, 4)	(251, 250, 252, 253, 254)	0.9000	1.9000	0.1000	3
7	C	(0, 1, 2, 3, 4)	(187, 187, 187, 188, 186)	-0.1000	1.9000	0.1000	3
8	C	(0, 1, 2, 4)	(208, 208, 207, 206)	-0.5429	0.1714	0.1143	2
9	C	(0, 1, 3)	(231, 234, 239)	2.6429	0.0714	0.2143	1
10	C	(0, 1, 3, 4)	(188, 190, 191, 192)	0.9000	0.6500	0.1000	2
11	T	(0, 1, 3, 4)	(231, 228, 224, 222)	-2.2000	0.3500	0.1000	2
12	T	(0, 1, 2, 4)	(200, 200, 202, 203)	0.8286	0.7429	0.1143	2
13	T	(0, 1, 2, 3, 4)	(271, 269, 267, 262, 261)	-2.7000	3.1000	0.1000	3
14	T	(0, 1, 2, 3, 4)	(226, 227, 225, 222, 226)	-0.5000	12.3000	0.1000	3
15	T	(0, 1, 2, 3, 4)	(182, 178, 176, 175, 170)	-2.7000	3.9000	0.1000	3
16	T	(0, 2, 4)	(212, 213, 213)	0.2500	0.1667	0.1250	1
17	T	(0, 1, 2, 4)	(208, 203, 201, 198)	-2.3429	4.9714	0.1143	2
18	T	(0, 3, 4)	(178, 174, 173)	-1.2692	0.0385	0.1154	1
19	T	(0, 2, 4)	(257, 255, 252)	-1.2500	0.1667	0.1250	1
20	T	(0, 1, 3, 4)	(203, 200, 198, 196)	-1.6000	1.1500	0.1000	2
					45.6373	2.2666	42

the variance for a patient with missing data to be similar. Thus, a typical participant's SSX at the end of the trial will be

$$SSX(1) = 6.425(5.25) = 33.7313.$$

Thus, we estimate the variance of $\hat{b}(1)$ for a typical participant to be

$$\text{var}\{\hat{b}(1)\} = \hat{\sigma}_e^2/33.7313 + \hat{\sigma}_b^2 = 1.0866/33.7313 + 1.2062 = 1.2384.$$

We estimate the variance of the treatment effect estimate at the end of the trial with 40 participants per arm by

$$\text{var}\{\hat{\delta}(1)\} = 2(1.2384)/40 = 0.0619.$$

From this we calculate the information at the end of the trial to be

$$I(1) = 1/0.0619 = 16.1551.$$

The fitted linear model using SAS's Proc Mixed is $y = 214.0587 + 2.5513u + (0.3008 - 1.6544u)x$, where u is 1 for a treatment patient and 0 for a control patient. Thus, the control slope minus the treatment slope is estimated to be $\hat{\delta} = 1.6544$, with an estimated standard error of 0.5078. We estimate the current information and information fraction to be

$$I(t) = 1/(0.5078)^2 = 3.8781, \quad t = 3.8781/16.1551 = 0.24.$$

The current z-score and B-value are $Z(0.24) = 1.6544/0.5078 = 3.258$ and $B(0.24) = (0.24)^{1/2}(3.258) = 1.596$.

Calculations like these are useful for computing the conditional probability that the final z-score will be at least 1.96. This probability, called *conditional power*, is very useful for deciding whether there is any hope of seeing a significant treatment benefit by trial's end (see Chapter 3).

This example was instructive because although the interim analysis occurred with data from only one quarter of the patients, each with only between three and five of the eight observations expected by trial's end, the estimated information fraction was 0.24. In other words, the information fraction was almost the same as the fraction of participants evaluated, even though participants had data for only about half the total number of weeks. This occurred because the variance of $\hat{\delta}(t)$ depends on, in addition to the sample size, 1) the number of observations per participant (reflected through SSX_i) and 2) the random effects variance of the true slopes of different participants. If the random effects variance is large enough, it will dominate, and we will not appreciably decrease the variance of $\hat{\delta}(t)$ regardless of the number of weeks of data. That is what occurred in this example. If the random effects variance had been very small, then the number of weeks of data would have contributed mightily to the amount of information.

2.7 When Is Brownian Motion Not Appropriate?

Sometimes observations in clinical trials are not i.i.d. For example, early in a trial clinicians or laboratories may not completely understand the protocol. Early patients may differ from later patients because once certain patient sources (e.g., a catheterization laboratory) are exhausted, other sources for patients must be used. These changes could make the drift nonlinear in t . Nonlinear drift also occurs with survival analysis when the proportional hazards model does not hold. Nonetheless, these things have little to no effect on the null distribution of the test statistic over time. Clinical trialists are most concerned about threats to type 1 error rate, so they do not worry much about the effect of drift.

In almost all realistic settings, we must estimate standard errors from the data. When the sample size is large, we can treat estimated standard errors as though they were constants (see Section 2.9.1). We cannot do this with a small sample size even in a nonmonitoring setting. For example, we know that the t-distribution differs substantially from the standard normal distribution if the number of degrees of freedom is small. Generally speaking, we need large sample sizes to use Brownian motion, although Chapter 8 shows that applying boundaries to p -values instead of z -scores works well unless the sample size is extremely small.

It is not immediately clear what a large sample size means in a complicated mixed model. Consider Example 2.5. Does the number of patients or the number of observations per patient need to be large? Suppose we had only two observations per patient, one at baseline and one at the end. Then each patient's slope would essentially reduce to a change in score from baseline to end of study. With enough patients, the Brownian motion paradigm would still apply. On the other hand, suppose the trial included only two patients per arm, each with a huge number of observations. We would be very confident about slopes of the four individuals in the study, but not at all about the mean slopes in the entire populations. Because we aim to make inferences about all patients in the population, we need a large number of patients, not a large number of observations per patient.

Another way to determine what must be large in Example 2.5 is to examine the expression for the variance of the treatment effect estimate. The variance contains parameters such as σ_e^2 and σ_b^2 that must be estimated from the data. The weakest link is the random effect variance σ_b^2 . Consider a best case scenario with an infinite number of observations per participant, so we could estimate each participant's slope perfectly. In that case the best estimate of σ_b^2 would be the sample variance of those n patient-specific slopes. If n were small, that sample variance would be a very poor estimate of σ_b^2 , and so the Brownian motion approximation would also be poor.

Example 2.6. The Rapid Early Action for Coronary Treatment (REACT) [LRO00] was a trial that randomized communities instead of individual patients. The intervention consisted of a media campaign intended to reduce the delay time between the onset of symptoms of a heart attack and the patient's arrival at the hospital emergency room. Control communities received no intervention. The data within each community consisted of delay times as a function of calendar time, and the slope of the relationship between calendar time and the logarithm of delay time summarized the trend in a given community.

This example is similar to Example 2.5 in certain respects. Both involved multiple correlated observations on the same randomized unit. The difference is that the number of randomized units is necessarily small in a community randomized trial. The primary analysis in REACT was a paired t-test with

only 9 degrees of freedom. Brownian motion provides a very poor approximation to the joint distribution of this paired t-statistic over time. Indeed, the Brownian motion approximation would treat the B-value $B(1)$ at the end of the trial as a standard normal deviate instead of a t-deviate with only 9 degrees of freedom.

2.8 Summary

This chapter showed that commonly used test statistics of the form $\hat{\delta}/\{\text{var}(\hat{\delta})\}^{\frac{1}{2}}$ behave like standardized sums of independent random variables with mean δ and variance 1. In these settings we measure the proportion of the trial completed in terms of information rather than chronological time. Information, the inverse of the variance of the treatment effect estimator $\hat{\delta}$, can be interpreted as the number of i.i.d. observations with expectation δ and variance 1 whose average has the same variance as $\hat{\delta}$. The information fraction t , the ratio of the current information to that at the end of the trial, is used to define the B-value $B(t) = t^{1/2}Z(t)$. The B-value is used to monitor the trial. Tables 2.2 and 2.3 summarize the B-value approach to monitoring.

Table 2.2. Brownian motion framework for four testing scenarios. For survival, n and N are the numbers of patients with an event at calendar fraction τ and the end of the trial ($\tau = 1$), respectively. For the other three scenarios, they are the numbers of patients evaluated at those times. The expressions given for information and information fraction assume equal per-arm sample sizes for means, proportions, and survival.

	Means	Proportions	Survival	MLE
Parameter δ	$\mu_T - \mu_C$	$p_T - p_C$	$\ln(\lambda_T/\lambda_C)$	arbitrary
Estimator $\hat{\delta}(\tau)$	$\bar{Y}_T - \bar{Y}_C$	$\hat{p}_T - \hat{p}_C$	$\frac{\sum_{i=1}^n (O_i - E_i)}{\sum_{i=1}^n V_i}$	MLE
$I(\tau) = [\text{var}\{\hat{\delta}(\tau)\}]^{-1}$	$\frac{n}{2\sigma^2}$	$\frac{n}{2p(1-p)}$	$\sum_{i=1}^n V_i$	Fisher info.
Info. fraction t	n/N	n/N	$\approx n/N$	$\approx n/N$
$Z(t)$	$\{I(\tau)\}^{1/2}\hat{\delta}(\tau)$	$\{I(\tau)\}^{1/2}\hat{\delta}(\tau)$	$\{I(\tau)\}^{1/2}\hat{\delta}(\tau)$	$\{I(\tau)\}^{1/2}\hat{\delta}(\tau)$
Drift $\theta = E\{Z(1)\}$	$\{I(1)\}^{1/2}\delta$	$\{I(1)\}^{1/2}\delta$	$\{I(1)\}^{1/2}\delta$	$\{I(1)\}^{1/2}\delta$

The advantage of monitoring the trial using the B-value instead of the more commonly used z-score is that its mean is a linear function of t . In fact, $E\{B(t)\} = \theta t$, where $\theta = E\{Z(1)\}$ is the expected z-score at the end of the

Table 2.3. Distribution and relationship between $B(t)$ and $Z(t)$.

B-value		Relationship between $B(t)$ and $Z(t)$	Z-score	
$E\{B(t)\}$	$\text{cov}\{B(s), B(t)\}$ $s \leq t$		$E\{Z(t)\}$	$\text{cov}\{Z(s), Z(t)\}$ $s \leq t$
θt	s	$B(t) = t^{1/2} Z(t)$	$\theta t^{1/2}$	$(s/t)^{1/2}$

trial. Plotting the B-value against θt makes it very easy to see whether, and to what degree, the current trend in the data is better or worse than expected.

2.9 Appendix

2.9.1 Asymptotic Validity of Using Estimated Standard Errors

In Section 2.1.1, the variance of $\hat{\delta}$ depended on σ^2 , which we treated as known. In practice we estimate σ^2 by the sample variance s^2 . We know that in a nonmonitoring setting, we can substitute s^2 for σ^2 and treat it as fixed if the sample size is large because $\hat{\delta}/(2s^2/N)^{1/2} = Z_N + R_N$, where $Z_N = \hat{\delta}/(2\sigma^2/N)^{1/2}$ converges in distribution to a standard normal deviate Z and $R_N = \{\hat{\delta}/(2\sigma^2/N)^{1/2}\}(\sigma/s - 1)$ converges in probability to 0. Similarly, in a nonmonitoring situation we treat the standard error of the MLE as if it were a fixed constant instead of being estimated from the data because $\hat{\delta}/\hat{\sigma}_{\hat{\delta}} = Z_N + R_N$, where $Z_N = \hat{\delta}/\sigma_{\hat{\delta}}$ converges in distribution to a standard normal deviate and $R_N = (\hat{\delta}/\sigma_{\hat{\delta}})(\sigma_{\hat{\delta}}/\hat{\sigma}_{\hat{\delta}} - 1)$ converges in probability to 0. Both these cases relied on Slutsky's theorem (Cramér, 1946 [C46]), which says that if Z_N converges in distribution to Z and R_N converges in probability to 0, then $Z_N + R_N$ converges in distribution to Z .

With monitoring, we know that $(\hat{\delta}_1/\sigma_{\hat{\delta}_1}, \dots, \hat{\delta}_k/\sigma_{\hat{\delta}_k})$ converges in distribution, and we want to show that $(\hat{\delta}_1/\hat{\sigma}_{\hat{\delta}_1}, \dots, \hat{\delta}_k/\hat{\sigma}_{\hat{\delta}_k})$ converges in distribution to the same thing. We need the following generalization of Slutsky's theorem.

Result 2.9 *Suppose that $\underline{X}_n = (X_{n1}, \dots, X_{np})$ converges in distribution to $\underline{X} = (X_1, \dots, X_p)$.*

1. *If $\underline{Y}_n = (Y_{n1}, \dots, Y_{np})$ converges to $\underline{0}$ in probability, then $\underline{X}_n + \underline{Y}_n$ converges in distribution to \underline{X} .*
2. *If A_n is an $m \times p$ dimensional matrix of random variables, each converging in probability to the corresponding element of the constant matrix A , then $A_n \underline{X}_n$ converges in distribution to $A \underline{X}$.*

Proof of 1: By the Cramer-Wold device (see, for example, page 18 of Serfling, 1980 [S80]), it suffices to prove that $\underline{a} \cdot (\underline{X}_n + \underline{Y}_n)$ converges in distribution to $\underline{a} \cdot \underline{X}$ for every p -dimensional nonrandom vector \underline{a} . But $\underline{a} \cdot \underline{X}_n$ converges in

distribution to $\underline{a} \cdot \underline{X}$, and $\underline{a} \cdot \underline{Y}_n$ converges to 0 in probability. By Slutsky's theorem for one-dimensional random variables, $\underline{a} \cdot (\underline{X}_n + \underline{Y}_n)$ converges in distribution to $\underline{a} \cdot \underline{X}$, completing the proof of 1. ||

Proof of 2: $A_n \underline{X}_n = A \underline{X}_n + (A_n - A) \underline{X}_n$. It is clear that $A \underline{X}_n$ converges in distribution to $A \underline{X}$ because $f(\underline{x}) = A \underline{x}$ is a continuous function of \underline{x} . Furthermore, because each element of the matrix $A_n - A$ converges in probability to 0 and \underline{X}_n converges in distribution, $(A_n - A) \underline{X}_n$ converges in probability to the m -dimensional zero vector. It follows from part 1 that $A_n \underline{X}_n$ converges in distribution to $A \underline{X}$. ||

Result 2.9 shows that when the sample sizes are large, we can treat the estimated standard errors of $\hat{\delta}_1, \dots, \hat{\delta}_k$ as if they were exact because

$$\left(\frac{\hat{\delta}_1}{\hat{\sigma}_{\hat{\delta}_1}}, \dots, \frac{\hat{\delta}_k}{\hat{\sigma}_{\hat{\delta}_k}} \right) = \left(\frac{\hat{\delta}_1}{\sigma_{\hat{\delta}_1}}, \dots, \frac{\hat{\delta}_k}{\sigma_{\hat{\delta}_k}} \right) \left(\frac{\sigma_{\hat{\delta}_1}}{\hat{\sigma}_{\hat{\delta}_1}} - 1, \dots, \frac{\sigma_{\hat{\delta}_k}}{\hat{\sigma}_{\hat{\delta}_k}} - 1 \right) + \left(\frac{\hat{\delta}_1}{\sigma_{\hat{\delta}_1}}, \dots, \frac{\hat{\delta}_k}{\sigma_{\hat{\delta}_k}} \right) \quad (2.25)$$

and each $\sigma_{\hat{\delta}_i} / \hat{\sigma}_{\hat{\delta}_i} - 1$ converges to 0 in probability.

2.9.2 Proof of Result 2.1

One direction is obvious, so we prove that if $S_N / v_N^{1/2}$ is asymptotically standard normal, then the asymptotic distribution of $(S_{n_1} / v_N^{1/2}, \dots, S_{n_k} / v_N^{1/2})$ is that of $B(t_1), \dots, B(t_k)$.

We first prove that the asymptotic distribution of $(S_M - S_m) / (v_M - v_m)^{1/2}$ is standard normal for $m < M$, $m \rightarrow \infty$, $M \rightarrow \infty$ such that $v_m / v_M \rightarrow t$. Write

$$(S_M / v_M^{1/2}) \left(\frac{v_M}{v_M - v_m} \right)^{1/2} = \frac{S_M - S_m}{(v_M - v_m)^{1/2}} + (S_m / v_m^{1/2}) \left(\frac{v_m}{v_M - v_m} \right)^{1/2}$$

$$W_{m,M} = U_{m,M} + V_{m,M},$$

where $U_{m,M}$ and $V_{m,M}$ are independent, $W_{m,M}$ converges in distribution to $N(0, (1-t)^{-1})$ and $V_{m,M}$ converges in distribution to $N(0, t/(1-t))$. Because $U_{m,M}$ is independent of $V_{m,M}$,

$$E(e^{isW_{m,M}}) = E(e^{isU_{m,M}})E(e^{isV_{m,M}}). \quad (2.26)$$

The left side of (2.26) converges to $\exp[-s^2 / \{2(1-t)\}]$, while $E(e^{isV_{m,M}})$ converges to $\exp[-s^2 t / \{2(1-t)\}]$. It follows that $E(e^{isU_{m,M}})$ converges to $\exp[(-s^2/2)\{1/(1-t) - t/(1-t)\}] = \exp(-s^2/2)$, the characteristic function of a standard normal deviate. Hence, $(S_M - S_m) / (v_M - v_m)^{1/2}$ is asymptotically standard normal as $m \rightarrow \infty$, $M \rightarrow \infty$, $v_m / v_M \rightarrow t$.

Let $n = 10^{n_1} + \dots + 10^{n_k}$, so that each (n_1, \dots, n_k) corresponds to a unique integer n . Let $\underline{Z}_n^T = (S_{n_1} / v_{n_1}, (S_{n_2} - S_{n_1}) / (v_{n_2} - v_{n_1})^{1/2}, \dots, (S_{n_k} - S_{n_{k-1}}) / (v_{n_k} - v_{n_{k-1}})^{1/2})$. The Z_{ni} are independent, and we have shown that

each converges in distribution to a standard normal, so the asymptotic distribution of \underline{Z}_n is that of i.i.d. standard normals $\underline{Z} = (Z_1, \dots, Z_k)^T$. Moreover, $(S_{n_1}/v_N^{1/2}, \dots, S_{n_k}/v_N^{1/2})^T = A_n \underline{Z}_n$, where the (i, j) th element of the $k \times k$ matrix A_n is $\{(v_{n_j} - v_{n_{j-1}})/v_N\}^{1/2}$ if $j \leq i$ and 0 if $j > i$, where $v_{n_0} = 0$. The (i, j) th element of A_n converges to $A_{ij} = (t_j - t_{j-1})^{1/2}$ for $j \leq i$ and 0 for $j > i$, where $t_0 = 0$, so by Result 2.9, $(S_{n_1}/v_N^{1/2}, \dots, S_{n_k}/v_N^{1/2})^T$ converges in distribution to $A\underline{Z}$. The joint distribution of $A\underline{Z}$ is multivariate normal with zero means and covariance matrix AA^T . Direct calculation shows that (i, j) th component of AA^T is t_j for $j \leq i$ and t_i for $j > i$. Thus, $(S_{n_1}/v_N^{1/2}, \dots, S_{n_k}/v_N^{1/2})$ converges in distribution to $(B(t_1), \dots, B(t_k))$. ||

2.9.3 Proof that for the Logrank Test, $D_i = O_i - E_i$ Are Uncorrelated Under H_0

To show that the D_i are uncorrelated, mean 0 random variables, we use the identity $\text{var}(Y) = E\{\text{var}(Y | \underline{X})\} + \text{var}\{E(Y | \underline{X})\}$ for a random variable Y with finite variance and a random vector \underline{X} . The unconditional mean and variance of D_i are $E(D_i) = E\{E(D_i | m_{0i}, m_{1i})\} = E(0) = 0$ and $\text{var}(D_i) = E\{\text{var}(D_i | m_{0i}, m_{1i})\} + \text{var}\{E(D_i | m_{0i}, m_{1i})\} = E(V_i) + \text{var}(0) = E(V_i)$. The D_i are uncorrelated because $\text{cov}(D_i, D_j) = E(D_i D_j) = E\{E(D_i D_j | D_i, m_{0j}, m_{1j})\} = E\{D_i E(D_j | D_i, m_{0j}, m_{1j})\}$.

Now consider $E\{D_i E(D_j | D_i, m_{0j}, m_{1j})\}$. Just prior to the j th death, D_i is relevant only in that it provides information about the numbers m_{0j} and m_{1j} of patients at risk at that time. Therefore, once we condition on m_{0j} and m_{1j} , the additional variable D_i becomes irrelevant so $E(D_j | D_i, m_{0j}, m_{1j}) = E(D_j | m_{0j}, m_{1j}) = 0$.

2.9.4 A Rigorous Justification of Brownian Motion with Drift: Local Alternatives

Up to now we have not been completely rigorous in our use of Brownian motion with drift. Consider the t-test for a continuous outcome trial. Ordinarily, we think of the treatment effect δ as a fixed constant (e.g., a 3 mm Hg blood pressure difference between the treatment and control arms). But then the expected final z-score,

$$\theta = \frac{\delta}{\sqrt{2\sigma^2/N}},$$

would tend to ∞ as $N \rightarrow \infty$, reflecting the obvious fact that power tends to 1 as the sample size tends to ∞ . To avoid having the power tend to 1, we must consider local alternatives (i.e., treatment effects δ_N that approach 0 as $N \rightarrow \infty$). The situation is analogous to the Poisson approximation to the binomial (n, p) distribution; for fixed p , the number of successes tends to ∞ as $n \rightarrow \infty$, but if $p = p_n$ tends to 0 such that $np_n \rightarrow \lambda$, the number of successes has an approximate Poisson distribution with mean λ .

Returning to the t-test, consider the location shift setting in which the $2N$ observations at the end of the trial in the control and treatment arms are i.i.d. $F(x)$ and i.i.d. $F(x - \delta_N)$, respectively, for some distribution function $F(x)$ and location parameter δ_N . We can imagine generating such data by generating $2N$ i.i.d. observations Y_1, \dots, Y_{2N} from F and adding δ_N to the first N . The interim z-statistic after n observations/arm is

$$\begin{aligned} Z_n &= \frac{\sum_{i=1}^n (Y_i + \delta_N) - \sum_{i=1}^n Y_{N+i}}{\sqrt{2n\sigma^2}} \\ &= \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n Y_{N+i}}{\sqrt{2n\sigma^2}} + \sqrt{\frac{n}{2\sigma^2}} \delta_N. \end{aligned}$$

Converting to B-values gives

$$B_n = \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n Y_{N+i}}{\sqrt{2N\sigma^2}} + \sqrt{\frac{N}{2\sigma^2}} \delta_N (n/N) \quad (2.27)$$

The first term of (2.27) is the B-value under the null hypothesis, whose joint distribution over information time is asymptotically standard Brownian motion. Let $\theta_N = \{N/(2\sigma^2)\}^{1/2} \delta_N$ and suppose that as $N \rightarrow \infty$, $n/N \rightarrow t$ and $\delta_N \rightarrow 0$ such that $\theta_N \rightarrow \theta$ for some constant θ . The rightmost term of the right side of (2.27) converges in probability to θt , so the multivariate version of Slutsky's theorem implies that the joint distribution of B_{n_1}, \dots, B_{n_k} is that of a Brownian motion with drift θ .

A similar technique can be used with dichotomous outcome trials. A rigorous justification of local alternatives in survival analysis is beyond the scope of this book. An excellent reference for the required martingale approach is Helland (1982) [H82].

2.9.5 Basu's Theorem

Result 2.10 *Basu (1955) [B55]. If $\hat{\delta} = (\hat{\delta}_1, \dots, \hat{\delta}_p)$ is a complete sufficient statistic for $(\delta_1, \dots, \delta_p)$ and $\underline{A} = (A_1, \dots, A_m)$ is ancillary, then $\hat{\delta}$ and \underline{A} are independent.*

Proof: Let $f(\underline{A})$ be any function with finite expectation, and let $\psi(\hat{\delta}) = E\{f(\underline{A}) \mid \hat{\delta}\}$. Then $E\{\psi(\hat{\delta})\} = E\{f(\underline{A})\}$, so $E[\psi(\hat{\delta}) - E\{f(\underline{A})\}] = 0$. Because \underline{A} is ancillary, $E\{f(\underline{A})\}$ does not depend on δ , so $\psi(\hat{\delta}) - E\{f(\underline{A})\}$ is a statistic and a function of $\hat{\delta}$. Completeness of $\hat{\delta}$ implies that $\psi(\hat{\delta}) = E\{f(\underline{A})\}$. Thus, $E\{f(\underline{A}) \mid \hat{\delta}\} = E\{f(\underline{A})\}$ for any function f with finite expectation. Taking $f(\underline{A}) = I(A_1 \leq a_1, \dots, A_m \leq a_m)$ shows that \underline{A} and $\hat{\delta}$ are independent. ||

Statistical Monitoring of Clinical Trials
A Unified Approach

Proschan, M.A.; Lan, K.K.G.; Wittes, J.T.

2006, XIV, 268 p. 32 illus.,

ISBN: 978-0-387-44970-8