Commonly Used Probability Models

Scott S. Emerson, M.D., Ph.D.

RCTdesign.org

August 13, 2012

Abstract

In this tutorial, we demonstrate the specification of the probability models most often used in clinical trial design. In each case we use a hypothetical clinical trial to illustrate the RCTdesign commands that would be used to create a `"seqDesign"` object for one and two arm trials. As the major focus of this tutorial is the specification of the probability model, we do not belabor the evaluation of the designs, though we do in a few instances illustrate some aspects of RCTdesign that are a little tangential to the probability model.

1 Introduction

RCTdesign explicitly handles six commonly used probability models. In each case, RCTdesign places greatest emphasis on the distribution free interpretation of the probability models, although each of the models was first defined in a parametric or semiparametric setting. These six models include:

- Inference about the mean (or difference in means) of a continuous random variable.
- Inference about the geometric mean (or ratio of geometric means) of a positive continuous random variable.
- Inference about the event probability (or difference in event probabilities) of a binary random variable.
- Inference about the odds of an event (or odds ratio) for a binary random variable.
- Inference about the mean rate (or ratio of rates) of a random variable counting events.
- Inference about the ratio of hazards for some event over time.

In this document we provide examples of RCTdesign clinical trial specifications under the mean, geometric mean, event probability, and hazard ratio probability models. We do not here provide the theoretical justification and details of the analytic models. (See instead the tutorials on the individual probability models or the RCTdesign Technical Overview.)

We initialize RCTdesign by typing

```
> library(RCTdesign)
```

2 Inference About Means

2.1 Hypothetical Example: Mean Change in Systolic Blood Pressure

We consider a hypothetical clinical trial of a new treatment for hypertension. The primary clinical endpoint Y_i for each subject is presumed to be the change in systolic blood pressure (SBP) measured 6 months after randomization. Important design parameters include

- *Null hypothesis*: In the absence of a treatment effect, we presume that the patients will average no change in SBP: $E[Y_i] = \mu_0 = 0$ mmHg.
- *Alternative hypothesis*: We presume that prior pilot studies have suggested that the treatment might provide a benefit that corresponds to an average decreased SBP of 10 mmHg: $E[Y_i] = \mu_1 = -10$ mmHg.
- *Variability*: We presume that prior pilot studies have suggested that the standard deviation for the change in SBP is $\sigma = 30$ mmHg within a study arm. Had we not had a direct estimate of the variability of the change in SBP, we could estimate the variability τ^2 of a single SBP within a study arm, estimate the correlation ρ between two SBP measurements taken 6 months apart, and calculated the variance σ^2 of the change in SBP as $\sigma^2 = 2\tau^2(1-\rho)$. Note that if $\tau = 27.386$ mmHg and $\rho = 0.4$, then $\sigma = 30$ mmHg.
- *Type I error*: We presume that standards of evidence demand a one-sided type I error of $\alpha = 0.025$.
- *Design power*: We presume that we desire 90% statistical power to reject the null hypothesis when the alternative hypothesis is in fact true.
- *Test statistic*: We presume that statistical inference will be based on the usual t statistic. (Such inference is distribution free in moderate to large sample sizes owing to the central limit theorem, the large sample consistency of the sample variance as an estimator of the population variance, and the fact that a t distribution with a large number of degrees of freedom is approximately standard normal.)

2.2 One Sample Inference on the Mean

Supposing that we do not want to perform any interim analyses of the data, we can design a one sample test of $H_0 : \mu \geq 0$ versus $H_1 : \mu \leq -10$ with the following RCTdesign code, which creates a `"seqDesign"` object named `dsnFixed`. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "mean",
+         arms= 1,
+         null.hypothesis= 0,
+         alt.hypothesis= -10,
+         sd= 30,
+         test.type= "less",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed

Call:
seqDesign(prob.model = "mean", arms = 1, null.hypothesis = 0,
```

```
    alt.hypothesis = -10, sd = 30, test.type = "less", size = 0.025,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N= 94.57)  -6.0464  -6.0464
```

Note that the printed description of the stopping rule identifies that

- The treatment effect θ is the mean response.
- The clinical trial design corresponds to a one-sided test of a lesser alternative.
- The null hypothesis is $\theta = 0$ is tested with a type I error of 0.025.
- The sample size corresponds to a test having 90% power to detect an alternative hypothesis of $\theta = -10$.
- The stopping boundaries are printed on the sample mean scale.
- A sample size of 94.57 would dictate that observed sample means less than -6.0464 would suggest rejection of the null hypothesis.

Because the statistical analysis used for this probability model corresponds to a parametric normal model, you could also specify `prob.model="normal"`, and that was the historic nomenclature for this probability model. However, we now greatly prefer the distribution free nomenclature.

Furthermore, several of the arguments that we supplied to `seqDesign()` were the default values, including the default value of `prob.model="mean"`. Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( arms= 1, alt.hypothesis= -10, sd= 30, power= 0.90)
> dsnFixed

Call:
seqDesign(arms = 1, alt.hypothesis = -10, sd = 30, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N= 94.57)  -6.0464  -6.0464
```

An alternative argument `variance` could have been used instead of `sd` to specify the variability of the measurements. Hence, the following code can be used to generate the exact same design:

```
> dsnFixed <- seqDesign( arms= 1, alt.hypothesis= -10, variance= 30^2, power= 0.90)
> dsnFixed

Call:
seqDesign(arms = 1, alt.hypothesis = -10, variance = 30^2, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                     Efficacy Futility
    Time 1 (N= 94.57)  -6.0464  -6.0464
```

Note that the boundaries are displayed on the scale of the estimated treatment effect by default. Had we wanted the (extremely boring) scale corresponding to the Z statistic, we could now ask for RCTdesign to print the boundary (but not the whole design) on a different scale using either of the following commands:

```
> changeSeqScale(dsnFixed,"Z")

STOPPING BOUNDARIES: Normalized Z-value scale
                          a     d
    Time 1 (N= 94.57) -1.96 -1.96

> seqBoundary(dsnFixed,"Z")

STOPPING BOUNDARIES: Normalized Z-value scale
                     Efficacy Futility
    Time 1 (N= 94.57)    -1.96    -1.96
```

Alternatively, we could have asked for the boundaries to be stored on the Z scale by creating the design using the `display.scale` argument:

```
> dsnFixed <- seqDesign( arms= 1, alt.hypothesis= -10, sd= 30, power= 0.90,
+         display.scale="Z")
> dsnFixed

Call:
seqDesign(arms = 1, alt.hypothesis = -10, sd = 30, power = 0.9,
    display.scale = "Z")

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a lesser alternative:
```

```
          Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Normalized Z-value scale
                     Efficacy Futility
    Time 1 (N= 94.57)    -1.96    -1.96
```

This is truly the exact same sequential sampling rule– it just changes the default method for display of the boundaries. (See the tutorials on the boundary scales for further information about the relative advantages and disadvantages of particular boundary scales.)

Had we wanted to create a design that would conduct up to, say, 6 equally spaced analyses of the data using O'Brien-Fleming boundaries in a one-sided symmetric design, we could execute

```
> dsn6 <- seqDesign( arms= 1, alt.hypothesis= -10, sd= 30, power= 0.90, nbr.analyses=6)
> dsn6

Call:
seqDesign(arms = 1, alt.hypothesis = -10, sd = 30, nbr.analyses = 6,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                      Efficacy Futility
    Time 1 (N=  16.74) -36.3086  24.2057
    Time 2 (N=  33.48) -18.1543   6.0514
    Time 3 (N=  50.22) -12.1029   0.0000
    Time 4 (N=  66.96)  -9.0771  -3.0257
    Time 5 (N=  83.71)  -7.2617  -4.8411
    Time 6 (N= 100.45)  -6.0514  -6.0514
```

Note that when a clinical trial design corresponds to a specific stopping rule previously described in the literature, the output will often include that reference. Though not shown in the default abbreviated output, the same is true of the boundary relationships. In the detailed printout for the design parameters, the boundary relationships will sometimes be labeled:

```
> dsn6$parameters

FULL PARAMETERIZATION:   (Emerson & Fleming (1989) symmetric test)
  Number and timing of analyses:
                          Time 1  Time 2 Time 3  Time 4  Time 5 Time 6
     Combined sample size 16.7410 33.4821  50.22 66.9642 83.7052  100.4
     Cum proportion        0.1667  0.3333   0.50  0.6667  0.8333    1.0
```

```
  Size, power, and hypothesis shift parameters:
            Lower Upper
    Alpha   0.025 0.025     (size = 0.025)
    Beta    0.975 0.975     (type I and type II errors equal)
    Epsilon 1.000 0.000     (one-sided test of lesser alternative)

  Design family based on: Sample Mean scale

  Boundary shape parameters:
      P A R G
    a 1 0 0 2.0216     (O'Brien-Fleming)
    b 1 0 0 2.0216     (O'Brien-Fleming)
    c 1 0 0 2.0216     (O'Brien-Fleming)
    d 1 0 0 2.0216     (O'Brien-Fleming)

  Constraints -
       Exact constraints:  (none)
     Minimum constraints:  (none)
     Maximum constraints:  (none)
```

(Note that the above detailed output is not for the faint of heart. See tutorials describing the full parameterization of a group sequential design.)

In the above output, the stopping boundaries at each analysis were again displayed on the sample mean scale by default. `changeSeqScale()` or `seqBoundary()` could again be used to display alternative boundary scales. (Again, see the tutorial on boundary scales for a full discussion.)

If we did not want the analyses equally spaced, we could type

```
> dsn6uneq <- seqDesign( arms= 1, alt.hypothesis= -10, sd= 30, power= 0.90,
+         nbr.analyses=6, sample.size=c(1,2,4,6,8,10))
> dsn6uneq

Call:
seqDesign(arms = 1, alt.hypothesis = -10, sd = 30, nbr.analyses = 6,
    sample.size = c(1, 2, 4, 6, 8, 10), power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                      Efficacy Futility
    Time 1 (N=  9.96) -60.5118  48.4094
    Time 2 (N= 19.93) -30.2559  18.1535
    Time 3 (N= 39.86) -15.1279   3.0256
    Time 4 (N= 59.78) -10.0853  -2.0171
    Time 5 (N= 79.71)  -7.5640  -4.5384
    Time 6 (N= 99.64)  -6.0512  -6.0512
```

Note that because we specified both an alternative hypothesis and a desired power, the `sample.size` argument is only being used to determine the relative spacing of the analyses. (See the tutorials on computing sample size, power, and alternatives for further discussion.)

2.3 Two Sample Inference on the Difference in Means

The above one sample test is not very rigorous scientifically, as it pretends that we would definitely know that in the absence of a treatment effect there would be no change in SBP. A scientifically better approach would be to conduct a randomized, double blind, placebo controlled study. In such a study, we might let μ_k be the mean change in SBP on the kth study arm, with $k = 1$ for the treatment and $k = 0$ for placebo. Our measure of treatment effect is then $\theta = \mu_1 - \mu_0$. We might presume that under the null hypothesis $\theta = \theta_0 = 0$, and under the alternative hypothesis $\theta = -10$. We continue to presume that the standard deviation of the measured change in SBP on a study arm is $\sigma = 30$ mmHg, which also corresponds to a standard deviation of a single SBP measurement of $\tau = 27.386$ mmHg and a correlation $\rho = 0.4$ between SBP measurements made six months apart.

Supposing that we desire 1:1 randomization and that we do not want to perform any interim analyses of the data, we can design a two sample test of $H_0 : \theta \geq 0$ versus $H_1 : \theta \leq -10$ with the following RCTdesign code, which creates a `"seqDesign"` object named `dsnFixed`. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "mean",
+         arms= 2,
+         ratio= c(1, 1),
+         null.hypothesis= 0,
+         alt.hypothesis= -10,
+         sd= 30,
+         test.type= "less",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed

Call:
seqDesign(prob.model = "mean", arms = 2, null.hypothesis = 0,
    alt.hypothesis = -10, sd = 30, ratio = c(1, 1), test.type = "less",
    size = 0.025, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in means (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 378.27)  -6.0464  -6.0464
```

Note that the printed description of the stopping rule now identifies that

- The treatment effect θ is the difference in mean response, where the difference is the treatment arm mean minus the control arm mean.
- The clinical trial design corresponds to a one-sided test of a lesser alternative.
- The null hypothesis is $\theta = 0$ is tested with a type I error of 0.025.
- The sample size corresponds to a test having 90% power to detect an alternative hypothesis of $\theta = -10$.
- The stopping boundaries are printed on the sample mean scale.
- A sample size of 378.28 (189.14 on each study arm) would dictate that observed difference in sample means less than -6.0464 (i.e., the decrease of SBP on the treatment arm was at least 6.0464 greater than any decrease of SBP on the control arm) would suggest rejection of the null hypothesis.

Note that the shorter printout does not include the randomization ratio. This can be obtained by obtaining the more detailed output that is the default when printing the `"seqModel"` object that is stored in every `"seqDesign"` object:

```
> dsnFixed$model

PROBABILITY MODEL:
   Two arm study of continuously distributed response variable
   Randomization scheme:  1 treatment group : 1 comparison group
   Outcome summarized by mean response
                  Treatment Comparison
             Null         0          0
      Alternative       -10          0
   Treatment effect is difference in means (Treatment - Comparison)
                       Theta
             Null          0
      Alternative       -10

   (Standardization parameters:  Mu0 0; SigmaSqr 1800; Psi 1)
```

Note that this more detailed output also displays means for each study arm. This is because in a two arm study, each of the arguments `ratio`, `null.hypothesis`, `alt.hypothesis`, `sd`, and `variance` can actually take values for each arm. In each case, the first value specified corresponds to the treatment arm, and the second value corresponds to the control arm.

- When a single value is specified for `ratio`, that value is presumed to correspond to the treatment arm, and the default value for the control arm is 1. Hence, specifying `ratio= 3` is equivalent to specifying `ratio= c(3,1)`, and specifying `ratio= 0.5` is equivalent to specifying `ratio= c(0.5,1)` (or `ratio= c(1,2)`).
- When a single value is specified for `null.hypothesis`, that value is presumed to correspond to the treatment arm under the null hypothesis, and by default it is also used for the control arm under the null hypothesis. (When no value is specified, the default value is 0.)
- When a single value is specified for `alt.hypothesis`, that value is presumed to correspond to the treatment arm under the null hypothesis, and the default value for the control arm under the alternative hypothesis is the same as is used for the control arm under the null hypothesis. (There is no default value for the alternative, and when no alternative hypothesis value is specified, RCTdesign will find the alternative that has the desired power for a specified sample size.)

- When a single value is specified for `sd` or `variance`, that value is presumed to correspond to the variability of measurements on the treatment arm (under both null and alternative hypotheses), and by default that same value is used for the control arm (under both null and alternative hypotheses).

Hence, we could have specified our design using

```
> dsnFixed <- seqDesign(
+         prob.model= "mean",
+         arms= 2,
+         ratio= c(1, 1),
+         null.hypothesis= c(0, 0),
+         alt.hypothesis= c(-10, 0),
+         sd= c(30, 30),
+         test.type= "less",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed$model

PROBABILITY MODEL:
   Two arm study of continuously distributed response variable
   Randomization scheme:  1 treatment group : 1 comparison group
   Outcome summarized by mean response
                  Treatment Comparison
            Null          0          0
       Alternative      -10          0
   Treatment effect is difference in means (Treatment - Comparison)
                       Theta
            Null           0
       Alternative      -10

   (Standardization parameters:  Mu0 0; SigmaSqr 1800; Psi 1)
```

However, there is little real reason to specify values for `null.hypothesis` and `alt.hypothesis` for each arm in the case of the mean probability model, because it is only the differences in the means that are relevant. For `sd` and `variance`, on the other hand, you may want to use both values to specify a heteroscedastic model having unequal variances. Note that the detailed printout for a `"seqModel"` object does not display the stored information about the variability of the measurements. That can be obtained by

```
> seqExtract(dsnFixed, "variance")

[1] 900 900
```

Several of the arguments that we supplied to `seqDesign()` in this instance were the default values (in particular, note that a two arm study with 1:1 randomization is the default value). Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( alt.hypothesis= -10, sd= 30, power= 0.90)
> dsnFixed
```

```
Call:
seqDesign(alt.hypothesis = -10, sd = 30, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in means (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 378.27)  -6.0464  -6.0464
```

Had we wanted to create a design that would conduct up to, say, 6 equally spaced analyses of the data using O'Brien-Fleming boundaries in a one-sided symmetric design, we could execute

```
> dsn6 <- seqDesign( alt.hypothesis= -10, sd= 30, power= 0.90, nbr.analyses=6)
> dsn6

Call:
seqDesign(alt.hypothesis = -10, sd = 30, nbr.analyses = 6, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in means (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  66.96) -36.3086  24.2057
    Time 2 (N= 133.93) -18.1543   6.0514
    Time 3 (N= 200.89) -12.1029   0.0000
    Time 4 (N= 267.86)  -9.0771  -3.0257
    Time 5 (N= 334.82)  -7.2617  -4.8411
    Time 6 (N= 401.79)  -6.0514  -6.0514
```

More detailed information about randomization ratios and sample sizes accrued to each arm at each analysis can be obtained from the detailed output for the "seqParameters" object stored in the "seqDesign" object:

```
> dsn6$parameters

FULL PARAMETERIZATION:   (Emerson & Fleming (1989) symmetric test)
  Number and timing of analyses:
                          Time 1   Time 2 Time 3   Time 4   Time 5 Time 6
     Combined sample size 66.9642 133.9284  200.9 267.8567 334.8209  401.8
(Treatment arm)           33.4821  66.9642  100.4 133.9284 167.4104  200.9
```

```
(Comparison arm)          33.4821  66.9642  100.4 133.9284 167.4104  200.9
     Cum proportion        0.1667   0.3333    0.5   0.6667   0.8333    1.0

  Size, power, and hypothesis shift parameters:
           Lower Upper
    Alpha   0.025 0.025     (size = 0.025)
    Beta    0.975 0.975     (type I and type II errors equal)
    Epsilon 1.000 0.000     (one-sided test of lesser alternative)

  Design family based on: Sample Mean scale

  Boundary shape parameters:
      P A R G
    a 1 0 0 2.0216     (O'Brien-Fleming)
    b 1 0 0 2.0216     (O'Brien-Fleming)
    c 1 0 0 2.0216     (O'Brien-Fleming)
    d 1 0 0 2.0216     (O'Brien-Fleming)

  Constraints -
       Exact constraints:  (none)
     Minimum constraints:  (none)
     Maximum constraints:  (none)
```

The above analysis of the change in SBP is not the best one to use in a RCT. Instead, more precision would be gained in an analysis of covariance (ANCOVA) that adjusts for baseline in a linear regression model. In that setting, the variability of the residual error would be $\tau^2(1-\rho^2)$, where ρ is the correlation between the baseline and final SBP measurements within a study arm and τ is the standard deviation of a single SBP measurement within each study arm. If we presume a correlation of $\rho = .4$ and a standard deviation $\tau = 27.386$ mmHg for a single SBP measurement, the following code would update the previous design `dsn6` to obtain design `dsn6ancova` by

```
> dsn6ancova <- update( dsn6, sd= sqrt(27.386^2 * (1 - 0.4^2)))
> dsn6ancova

Call:
seqDesign(alt.hypothesis = -10, sd = sqrt(27.386^2 * (1 - 0.4^2)),
    nbr.analyses = 6, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in means (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  46.87) -36.3086  24.2057
    Time 2 (N=  93.75) -18.1543   6.0514
    Time 3 (N= 140.62) -12.1029   0.0000
    Time 4 (N= 187.50)  -9.0771  -3.0257
```

```
    Time 5 (N= 234.37)  -7.2617  -4.8411
    Time 6 (N= 281.25)  -6.0514  -6.0514
```

We could also have used the `variance` argument, but then we would have had to "undefine" the previously used `sd` argument, because that takes precedence. So we could have avoided having to specify the square root by using

```
> dsn6ancova <- update( dsn6, sd= , variance= 27.386^2 * (1 - 0.4^2))
> dsn6ancova

Call:
seqDesign(alt.hypothesis = -10, variance = 27.386^2 * (1 - 0.4^2),
    nbr.analyses = 6, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in means (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                        Efficacy Futility
    Time 1 (N=  46.87) -36.3086  24.2057
    Time 2 (N=  93.75) -18.1543   6.0514
    Time 3 (N= 140.62) -12.1029   0.0000
    Time 4 (N= 187.50)  -9.0771  -3.0257
    Time 5 (N= 234.37)  -7.2617  -4.8411
    Time 6 (N= 281.25)  -6.0514  -6.0514
```

Note the smaller sample size requirements for the ANCOVA model ($N = 281.25$) compared to an analysis of the unadjusted change in SBP measurements ($N = 401.79$). In fact, given that we presumed $rho < 0.5$, the analysis of the unadjusted change in SBP measurements is larger that what would be required if we ignored the baseline and only used the final measurement:

```
> dsn6final <- update( dsn6, sd= 27.386)
> dsn6final

Call:
seqDesign(alt.hypothesis = -10, sd = 27.386, nbr.analyses = 6,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in means (Treatment - Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=   0    (size  = 0.025)
    Alternative hypothesis : Theta <= -10    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
```

```
                   Efficacy Futility
Time 1 (N=  55.80) -36.3086  24.2057
Time 2 (N= 111.61) -18.1543   6.0514
Time 3 (N= 167.41) -12.1029   0.0000
Time 4 (N= 223.21)  -9.0771  -3.0257
Time 5 (N= 279.01)  -7.2617  -4.8411
Time 6 (N= 334.82)  -6.0514  -6.0514
```

3 Inference About Geometric Means

3.1 Hypothetical Example: Change in Geometric Mean in Prostate Specific Antigen (PSA)

We consider a hypothetical early phase clinical trial of a hormonal treatment for prostate cancer. The primary analysis will use the surrogate endpoint Y_i for each subject defined to be the proportionate change (six month measurement divided by the baseline measurement) in serum prostate specific antigen (PSA) measured 3 months after randomization. The analysis will consider the geometric mean of the ratio of the 3 month measurement for each subject divided by his baseline (pretreatment) value. Important design parameters include

- *Null hypothesis*: In the absence of a treatment effect, we presume that the patients will tend to have no change in PSA as reflected in the geometric mean of the ratio of final to baseline PSA of 1: $GM[Y_i] = \mu_0 = 1.0$.
- *Alternative hypothesis*: We presume that prior pilot studies have suggested that the treatment might provide a benefit that corresponds to a tendency toward a 50% reduction in serum PSA levels over threee months in the sense that the geometric mean ratio is 0.5: $GM[Y_i] = \mu_1 = 0.5$.
- *Test statistic*: Recalling that inference about geometric means is typically effected by considering the mean of log transformed data, we presume that statistical inference will be based on the usual t statistic on log transformed person specific ratios. (Such inference is distribution free in moderate to large sample sizes owing to the central limit theorem, the large sample consistency of the sample variance as an estimator of the population variance, and the fact that a t distribution with a large number of degrees of freedom is approximately standard normal.)
- *Variability*: In order to apply the t based inference, we will need to know the variability of the log transformed ratios. We presume that prior pilot studies have suggested that the standard deviation for the log PSA ratios is $\sigma = 2.166$ within a study arm. Had we not had a direct estimate of the variability of the change in SBP, we could estimate the variability τ^2 of a single log transformed PSA within a study arm, estimate the correlation ρ between two SBP measurements taken 3 months apart, and calculated the variance σ^2 of the change in SBP as $\sigma^2 = 2\tau^2(1-\rho)$. Note that if $\tau = 1.9$ and $\rho = 0.4$, then $\sigma = 2.166$ mmHg.
- *Type I error*: We presume that standards of evidence demand a one-sided type I error of $\alpha = 0.025$.
- *Design power*: We presume that we desire 95% statistical power to reject the null hypothesis when the alternative hypothesis is in fact true.

3.2 One Sample Inference on the Geometric Mean

Supposing that we do not want to perform any interim analyses of the data, we can design a one sample test of $H_0 : \mu \geq 1.0$ versus $H_1 : \mu \leq 0.5$ with the following RCTdesign code, which creates a `"seqDesign"` object named `dsnFixed`. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "geom.mean",
+         arms= 1,
+         null.hypothesis= 1.0,
+         alt.hypothesis= 0.5,
+         sd= 2.166,
+         test.type= "less",
+         size= 0.025,
+         power= 0.95
+         )
> dsnFixed

Call:
seqDesign(prob.model = "geom.mean", arms = 1, null.hypothesis = 1,
    alt.hypothesis = 0.5, sd = 2.166, test.type = "less", size = 0.025,
    power = 0.95)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 126.89)    0.686    0.686
```

Note that the printed description of the stopping rule identifies that

- The treatment effect θ is the geometric mean response.
- The clinical trial design corresponds to a one-sided test of a lesser alternative.
- The null hypothesis is $\theta = 1.0$ is tested with a type I error of 0.025.
- The sample size corresponds to a test having 90% power to detect an alternative hypothesis of $\theta = 0.5$.
- The stopping boundaries are printed on the scale of the estimated treatment effect (the "sample mean scale"), which in this case will correspond to the sample geometric mean.
- A sample size of 126.89 would dictate that observed sample geometric means less than 0.686 would suggest rejection of the null hypothesis.

Because the statistical analysis used for this probability model corresponds to a parametric lognormal model, you could also specify `prob.model="lognormal"`, and that was the historic nomenclature for this

probability model. In that parametric model, the geometric mean is also the median, so previous versions of S+SeqTrial referred to the treatment effect as being the median of a lognormal distribution. However, we now greatly prefer the distribution free nomenclature related to a geometric mean.

Furthermore, several of the arguments that we supplied to `seqDesign()` were the default values. Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( prob.model="geom.mean", arms= 1, alt.hypothesis= 0.5,
+         sd= 2.166, power= 0.95)
> dsnFixed

Call:
seqDesign(prob.model = "geom.mean", arms = 1, alt.hypothesis = 0.5,
    sd = 2.166, power = 0.95)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 126.89)    0.686    0.686
```

The alternative argument `variance` could have been used instead of `sd` to specify the variability of the measurements. Hence, the following code can be used to generate the exact same design (note that the first unnamed argument will be interpreted as `prob.model`):

```
> dsnFixed <- seqDesign( "geom.mean", arms= 1, alt.hypothesis= 0.5, variance= 2.166^2,
+         power= 0.95)
> dsnFixed

Call:
seqDesign(prob.model = "geom.mean", arms = 1, alt.hypothesis = 0.5,
    variance = 2.166^2, power = 0.95)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 126.89)    0.686    0.686
```

Had we wanted to create a design that would conduct up to, say, 6 equally spaced analyses of the data using Pocock boundaries (see tutorials on boundary shape functions and the unified family) in a one-sided symmetric design, we could execute

```
> dsn6 <- seqDesign( "geom.mean", arms= 1, alt.hypothesis= 0.5, sd= 2.166, power= 0.95,
+         nbr.analyses=6, P= 0.5)
> dsn6

Call:
seqDesign(prob.model = "geom.mean", arms = 1, alt.hypothesis = 0.5,
    sd = 2.166, nbr.analyses = 6, power = 0.95, P = 0.5)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean response
   One-sided hypothesis test of a lesser alternative:
          Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                      Efficacy Futility
    Time 1 (N=  31.54)   0.3943   1.1862
    Time 2 (N=  63.08)   0.5179   0.9032
    Time 3 (N=  94.62)   0.5843   0.8005
    Time 4 (N= 126.16)   0.6279   0.7449
    Time 5 (N= 157.70)   0.6596   0.7092
    Time 6 (N= 189.24)   0.6839   0.6839
```

Note that when a clinical trial design corresponds to a specific stopping rule previously described in the literature, the output will often include that reference. Though not shown in the default abbreviated output, the same is true of the boundary relationships. In the detailed printout for the design parameters, the boundary relationships will sometimes be labeled:

```
> dsn6$parameters

FULL PARAMETERIZATION:   (Emerson & Fleming (1989) symmetric test)
  Number and timing of analyses:
                           Time 1  Time 2 Time 3   Time 4   Time 5 Time 6
     Combined sample size 31.5401 63.0801  94.62 126.1603 157.7004  189.2
     Cum proportion        0.1667  0.3333   0.50   0.6667   0.8333    1.0

  Size, power, and hypothesis shift parameters:
            Lower Upper
    Alpha   0.025 0.025     (size = 0.025)
    Beta    0.975 0.975     (type I and type II errors equal)
    Epsilon 1.000 0.000     (one-sided test of lesser alternative)

  Design family based on: Sample Mean scale

  Boundary shape parameters:
      P   A R G
    a 0.5 0 0 2.4129     (Pocock)
    b 0.5 0 0 2.4129     (Pocock)
    c 0.5 0 0 2.4129     (Pocock)
    d 0.5 0 0 2.4129     (Pocock)
```

```
  Constraints -
       Exact constraints:  (none)
     Minimum constraints:  (none)
     Maximum constraints:  (none)
```

(Note that the above detailed output is not for the faint of heart. See tutorials describing the full parameterization of a group sequential design.)

In the above output, the stopping boundaries at each analysis were again displayed on the scale of the estimated treatment effect by default. `changeSeqScale()` or `seqBoundary()` could again be used to display alternative boundary scales. (See the tutorial on boundary scales for a full discussion.)

If we did not want the analyses equally spaced, we could type

```
> dsn6uneq <- seqDesign( "geom.mean", arms= 1, alt.hypothesis= 0.5, sd= 2.166,
+         power= 0.95, nbr.analyses=6, P=0.5, sample.size=c(1,2,4,6,8,10))
> dsn6uneq

Call:
seqDesign(prob.model = "geom.mean", arms = 1, alt.hypothesis = 0.5,
    sd = 2.166, nbr.analyses = 6, sample.size = c(1, 2, 4, 6,
        8, 10), power = 0.95, P = 0.5)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  19.46)   0.2994   1.5577
    Time 2 (N=  38.92)   0.4263   1.0942
    Time 3 (N=  77.84)   0.5472   0.8524
    Time 4 (N= 116.77)   0.6112   0.7631
    Time 5 (N= 155.69)   0.6529   0.7144
    Time 6 (N= 194.61)   0.6830   0.6830
```

Note that because we specified both an alternative hypothesis and a desired power, the `sample.size` argument is only being used to determine the relative spacing of the analyses. (See the tutorials on computing sample size, power, and alternatives for further discussion.)

In the above, we have chosen to parameterize our hypotheses on the natural scale of the geometric mean of the PSA measurements and the geometric means of their ratios. Some disciplines, however, are more used to reporting inference on the log transformed scale. Examples include antibody titers and gene expression, which though measured initially on a natural scale, are often analyzed according to "log change". (Other examples such as decibels, pH, and the Richter scale are measured on a logarithmic scale from the very start.) RCTdesign does afford users the option of using the log geometric mean as the fundamental summary measure of treatment effect. The following RCTdesign code will produce a clinical design in the same setting, but using the log scale for parameters and estimated treatment effects:

```
> dsn6uneqLOG <- seqDesign( "geom.mean", arms= 1, alt.hypothesis= log(0.5), sd= 2.166,
+         power= 0.95, nbr.analyses=6, P=0.5, sample.size=c(1,2,4,6,8,10),
+         log.transform=TRUE)
> dsn6uneqLOG

Call:
seqDesign(prob.model = "geom.mean", arms = 1, log.transform = TRUE,
    alt.hypothesis = log(0.5), sd = 2.166, nbr.analyses = 6,
    sample.size = c(1, 2, 4, 6, 8, 10), power = 0.95, P = 0.5)

PROBABILITY MODEL and HYPOTHESES:
   Theta is log geometric mean response
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >=  0.0000    (size  = 0.025)
    Alternative hypothesis : Theta <= -0.6931    (power = 0.950)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  19.46)  -1.2059   0.4432
    Time 2 (N=  38.92)  -0.8527   0.0900
    Time 3 (N=  77.84)  -0.6029  -0.1597
    Time 4 (N= 116.77)  -0.4923  -0.2704
    Time 5 (N= 155.69)  -0.4263  -0.3363
    Time 6 (N= 194.61)  -0.3813  -0.3813
```

Note that exponentiating the boundary in `dsn6uneqLOG` yields the same boundary as in `dsn6uneq`:

```
> exp(dsn6uneqLOG$boundary)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  19.46)   0.2994   1.5577
    Time 2 (N=  38.92)   0.4263   1.0942
    Time 3 (N=  77.84)   0.5472   0.8524
    Time 4 (N= 116.77)   0.6112   0.7631
    Time 5 (N= 155.69)   0.6529   0.7144
    Time 6 (N= 194.61)   0.6830   0.6830
```

3.3 Two Sample Inference on the Ratio of Geometric Means

The above one sample test is not very rigorous scientifically, as it pretends that we would definitely know that in the absence of a treatment effect there would be no change in PSA. A scientifically better approach would be to conduct a randomized, double blind, placebo controlled study. In such a study, we might let μ_k be the geometric mean proportionate change in PSA over three months on the kth study arm, with $k = 1$ for the treatment and $k = 0$ for placebo. Our measure of treatment effect is then $\theta = \mu_1/\mu_0$. We might presume that under the null hypothesis $\theta = \theta_0 = 1.0$, and under the alternative hypothesis $\theta = 0.5$. We continue to presume that the standard deviation of the measured individual log PSA ratio on a study arm is $\sigma = 2.166$ mmHg, which also corresponds to a standard deviation of a single log transformed PSA measurement of $\tau = 1.9$ mmHg and a correlation $\rho = 0.4$ between log PSA measurements made three months apart.

Supposing that we desire 2:1 randomization (i.e., 2 subjects on the treatment arm for every 1 subject on the control arm) and that we do not want to perform any interim analyses of the data, we can design a two sample test of $H_0 : \theta \geq 1.0$ versus $H_1 : \theta \leq 0.5$ with the following RCTdesign code, which creates a "seqDesign" object named dsnFixed. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "geom.mean",
+         arms= 2,
+         ratio= c(2, 1),
+         null.hypothesis= 1.0,
+         alt.hypothesis= 0.5,
+         sd= 2.166,
+         test.type= "less",
+         size= 0.025,
+         power= 0.95
+         )
> dsnFixed

Call:
seqDesign(prob.model = "geom.mean", arms = 2, null.hypothesis = 1,
    alt.hypothesis = 0.5, sd = 2.166, ratio = c(2, 1), test.type = "less",
    size = 0.025, power = 0.95)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 571.01)    0.686    0.686
```

Note that the printed description of the stopping rule now identifies that

- The treatment effect θ is the ratio of geometric mean response, where the ratio is the treatment arm geometric mean divided by the control arm geometric mean.
- The clinical trial design corresponds to a one-sided test of a lesser alternative.
- The null hypothesis is $\theta = 1.0$ is tested with a type I error of 0.025.
- The sample size corresponds to a test having 95% power to detect an alternative hypothesis of $\theta = 0.5$.
- The stopping boundaries are printed on the scale of the estimated treatment effect.
- A sample size of 571.01 (380.74 on the treatment arm and 190.37 on the control arm) would dictate that observed ratio of geometric means less than 0.686 (i.e., the decrease of PSA on the treatment arm was at most 68.6% of any decrease of PSA on the control arm) would suggest rejection of the null hypothesis.

Note that the shorter printout does not include the randomization ratio. This can be obtained by obtaining the more detailed output that is the default when printing the "seqModel" object that is stored in every "seqDesign" object:

```
> dsnFixed$model

PROBABILITY MODEL:
   Two arm study of continuously distributed response variable
   Randomization scheme:  1 treatment group : 0.5 comparison group
   Outcome summarized by geometric mean response
                  Treatment Comparison
            Null       1.0          1
     Alternative       0.5          1
   Treatment effect is geometric mean ratio (Treatment : Comparison)
                       Theta
            Null         1.0
     Alternative         0.5

   (Standardization parameters:  Mu0 1; SigmaSqr 14.07; Psi 1)
```

Note that this more detailed output also displays geometric means for each study arm. This is because in a two arm study, each of the arguments `ratio`, `null.hypothesis`, `alt.hypothesis`, `sd`, and `variance` can actually take values for each arm. In each case, the first value specified corresponds to the treatment arm, and the second value corresponds to the control arm.

- When a single value is specified for `ratio`, that value is presumed to correspond to the treatment arm, and the default value for the control arm is 1. Hence, specifying `ratio= 3` is equivalent to specifying `ratio= c(3,1)`, and specifying `ratio= 0.5` is equivalent to specifying `ratio= c(0.5,1)` (or `ratio= c(1,2)`).
- When a single value is specified for `null.hypothesis`, that value is presumed to correspond to the treatment arm under the null hypothesis, and by default it is also used for the control arm under the null hypothesis. (When no value is specified with `prob.model="geom.mean"` and `log.transform=FALSE`, the default value is 1.0.)
- When a single value is specified for `alt.hypothesis`, that value is presumed to correspond to the treatment arm under the null hypothesis, and the default value for the control arm under the alternative hypothesis is the same as is used for the control arm under the null hypothesis. (There is no default value for the alternative, and when no alternative hypothesis value is specified, RCTdesign will find the alternative that has the desired power for a specified sample size.)
- When a single value is specified for `sd` or `variance`, that value is presumed to correspond to the variability of log measurements on the treatment arm (under both null and alternative hypotheses), and by default that same value is used for the control arm (under both null and alternative hypotheses).

Hence, we could have specified our design using

```
> dsnFixed <- seqDesign(
+         prob.model= "geom.mean",
+         arms= 2,
+         ratio= c(2, 1),
+         null.hypothesis= c(1, 1),
```

```
+         alt.hypothesis= c(0.5, 1),
+         sd= c(2.166, 2.166),
+         test.type= "less",
+         size= 0.025,
+         power= 0.95
+         )
> dsnFixed$model

PROBABILITY MODEL:
   Two arm study of continuously distributed response variable
   Randomization scheme:  1 treatment group : 0.5 comparison group
   Outcome summarized by geometric mean response
                  Treatment Comparison
            Null       1.0          1
       Alternative     0.5          1
   Treatment effect is geometric mean ratio (Treatment : Comparison)
                       Theta
            Null       1.0
       Alternative     0.5

   (Standardization parameters:  Mu0 1; SigmaSqr 14.07; Psi 1)
```

However, there is little real reason to specify values for `null.hypothesis` and `alt.hypothesis` for each arm in the case of the geometric mean probability model, because it is only the ratios in the geometric means that are relevant. For `sd` and `variance`, on the other hand, you may want to use both values to specify a heteroscedastic model having unequal variances of the log transformed data. Note that the detailed printout for a `"seqModel"` object does not display the stored information about the variability of the measurements. That can be obtained by

```
> seqExtract(dsnFixed, "variance")

[1] 4.691556 4.691556
```

Several of the arguments that we supplied to `seqDesign()` in this instance were the default values. Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( "geom.mean", ratio= 2, alt.hypothesis= 0.5, sd= 2.166,
+         power= 0.95)
> dsnFixed

Call:
seqDesign(prob.model = "geom.mean", alt.hypothesis = 0.5, sd = 2.166,
    ratio = 2, power = 0.95)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Fixed sample test)
```

```
STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 571.01)    0.686    0.686
```

Had we wanted to create a design that would conduct up to, say, 6 equally spaced analyses of the data using Pocock boundaries in a one-sided symmetric design, we could execute

```
> dsn6 <- seqDesign( "geom.mean", ratio= 2, alt.hypothesis= 0.5, sd= 2.166,
+         power= 0.95, nbr.analyses=6, P= 0.5)
> dsn6

Call:
seqDesign(prob.model = "geom.mean", alt.hypothesis = 0.5, sd = 2.166,
    ratio = 2, nbr.analyses = 6, power = 0.95, P = 0.5)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N= 141.93)   0.3943   1.1862
    Time 2 (N= 283.86)   0.5179   0.9032
    Time 3 (N= 425.79)   0.5843   0.8005
    Time 4 (N= 567.72)   0.6279   0.7449
    Time 5 (N= 709.65)   0.6596   0.7092
    Time 6 (N= 851.58)   0.6839   0.6839
```

More detailed information about randomization ratios and sample sizes accrued to each arm at each analysis can be obtained from the detailed output for the `"seqParameters"` object stored in the `"seqDesign"` object:

```
> dsn6$parameters

FULL PARAMETERIZATION:   (Emerson & Fleming (1989) symmetric test)
  Number and timing of analyses:
                            Time 1   Time 2 Time 3   Time 4   Time 5 Time 6
     Combined sample size 141.9303 283.8606  425.8 567.7213 709.6516  851.6
(Treatment arm)           94.6202 189.2404  283.9 378.4808 473.1011  567.7
(Comparison arm)          47.3101  94.6202  141.9 189.2404 236.5505  283.9
     Cum proportion        0.1667   0.3333    0.5   0.6667   0.8333    1.0

  Size, power, and hypothesis shift parameters:
          Lower Upper
    Alpha   0.025 0.025     (size = 0.025)
    Beta    0.975 0.975     (type I and type II errors equal)
```

```
    Epsilon 1.000 0.000     (one-sided test of lesser alternative)

  Design family based on: Sample Mean scale

  Boundary shape parameters:
      P   A R G
    a 0.5 0 0 2.4129     (Pocock)
    b 0.5 0 0 2.4129     (Pocock)
    c 0.5 0 0 2.4129     (Pocock)
    d 0.5 0 0 2.4129     (Pocock)

  Constraints -
       Exact constraints:  (none)
     Minimum constraints:  (none)
     Maximum constraints:  (none)
```

The above analysis of the proportionate change in PSA is not the best one to use in a RCT. Instead, more precision would be gained in an analysis of covariance (ANCOVA) that adjusts for baseline log measurement in a linear regression model. In that setting, the variability of the residual error would be $\tau^2(1-\rho^2)$, where ρ is the correlation between the log baseline and log final PSA measurements within a study arm and τ is the standard deviation of a single log PSA measurement within each study arm. If we presume a correlation of $\rho = .4$ and a standard deviation $\tau = 1.9$ for a single log PSA measurement, the following code would update the previous design `dsn6` to obtain design `dsn6ancova` by

```
> dsn6ancova <- update( dsn6, sd= sqrt(1.9^2 * (1 - 0.4^2)))
> dsn6ancova

Call:
seqDesign(prob.model = "geom.mean", alt.hypothesis = 0.5, sd = sqrt(1.9^2 *
    (1 - 0.4^2)), ratio = 2, nbr.analyses = 6, power = 0.95,
    P = 0.5)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  91.74)   0.3943   1.1862
    Time 2 (N= 183.47)   0.5179   0.9032
    Time 3 (N= 275.21)   0.5843   0.8005
    Time 4 (N= 366.95)   0.6279   0.7449
    Time 5 (N= 458.69)   0.6596   0.7092
    Time 6 (N= 550.42)   0.6839   0.6839
```

We could also have used the `variance` argument, but then we would have had to "undefine" the previously used `sd` argument, because that takes precedence. So we could have avoided having to specify the square root by using

```
> dsn6ancova <- update( dsn6, sd= , variance= 1.9^2 * (1 - 0.4^2))
> dsn6ancova

Call:
seqDesign(prob.model = "geom.mean", alt.hypothesis = 0.5, variance = 1.9^2 *
    (1 - 0.4^2), ratio = 2, nbr.analyses = 6, power = 0.95, P = 0.5)

PROBABILITY MODEL and HYPOTHESES:
   Theta is geometric mean ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.0    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.5    (power = 0.950)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Efficacy Futility
    Time 1 (N=  91.74)   0.3943   1.1862
    Time 2 (N= 183.47)   0.5179   0.9032
    Time 3 (N= 275.21)   0.5843   0.8005
    Time 4 (N= 366.95)   0.6279   0.7449
    Time 5 (N= 458.69)   0.6596   0.7092
    Time 6 (N= 550.42)   0.6839   0.6839
```

Note again the smaller sample size requirements for the ANCOVA model compared to an analysis of the unadjusted ratio of PSA measurements.

Note also that if a user so desires, the argument `log.transform` can be set to `TRUE` in order to express treatment effects on the log geometric mean scale.

4 Inference About Proportions

4.1 Hypothetical Example: 28 Day Survival in Gram Negative Sepsis

We consider a clinical trial of a monoclonal antibody directed against endotoxin in the treatment of Gram negative sepsis. The primary clinical endpoint Y_i for each subject is a binary indicator of survival for 28 days after treatment. Important design parameters include

- *Null hypothesis*: In the absence of a treatment effect, we presume that 70% of patients will survive 28 days: $E[Y_i] = p_0 = 0.70$.
- *Alternative hypothesis*: We presume that prior pilot studies have suggested that the treatment might provide a benefit that corresponds to 28 day survival probability of 77%: $E[Y_i] = p_1 = 0.77$.
- *Type I error*: We presume that standards of evidence demand a one-sided type I error of $\alpha = 0.025$.
- *Design power*: We presume that we desire 90% statistical power to reject the null hypothesis when the alternative hypothesis is in fact true.
- *Test statistic*: We presume that statistical inference will be based on the asymptotic normal distribution for the sample proportion. (In two samples, such inference is equivalent to the use of Pearson's chi square statistic.)

- *Variability*: The Bernoulli distribution involves a mean-variance relationship in which a binary random variable having mean p will have variance $p(1-p)$. Because we consider alternative values for the mean, we must also consider how which hypotheses will be used to compute the variability during trial design.

4.2 One Sample Inference on a Proportion

Supposing that we do not want to perform any interim analyses of the data, we can design a one sample test of $H_0 : p \leq p_0 = 0.70$ versus $H_1 : p \geq p_1 = 0.77$ with the following RCTdesign code, which creates a `"seqDesign"` object named `dsnFixed`. Note that we have to choose whether the design will be computed under the alternative hypothesis (so the variance would be $p_1(1-p_1)$), the null hypothesis (so the variance would be $p_0(1-p_0)$, an intermediate hypothesis (so the variance would be $(p_0+p_1)(2-p_0-p_1)/4$), or some other specified value. In the one sample problem, an intermediate hypothesis is often found to provide more accurate power and sample size computations, so we initially use that. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "proportions",
+         arms= 1,
+         null.hypothesis= 0.70,
+         alt.hypothesis= 0.77,
+         variance= "intermediate",
+         test.type= "greater",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed

Call:
seqDesign(prob.model = "proportions", arms = 1, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = "intermediate", test.type = "greater",
    size = 0.025, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is event probability
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.70    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.77    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Futility Efficacy
    Time 1 (N= 417.67)   0.7423   0.7423
```

Note that the printed description of the stopping rule identifies that

- The treatment effect θ is the probability of subjects having an event.
- The clinical trial design corresponds to a one-sided test of a greater alternative.
- The null hypothesis is $\theta = 0.70$ is tested with a type I error of 0.025.

- The sample size corresponds to a test having 90% power to detect an alternative hypothesis of $\theta = 0.77$.
- The stopping boundaries are printed on the scale of the estimated treatment effect (the "sample mean scale"), which in this case is the sample proportion.
- A sample size of 417.67 would dictate that observed proportions of an event greater than 0.7423 would suggest rejection of the null hypothesis.

Two of the arguments that we supplied to `seqDesign()` were the default values. Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( prob.model= "proportions", arms= 1, null.hypothesis= 0.70,
+         alt.hypothesis= 0.77,  variance="intermediate", power= 0.90)
> dsnFixed

Call:
seqDesign(prob.model = "proportions", arms = 1, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = "intermediate", power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is event probability
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.70    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.77    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                       Futility Efficacy
    Time 1 (N= 417.67)   0.7423   0.7423
```

Had we wanted to create a design that would conduct up to, say, 6 equally spaced analyses of the data using O'Brien-Fleming boundaries in a one-sided symmetric design, we could execute

```
> dsn6 <- seqDesign( prob.model= "proportions", arms= 1, null.hypothesis= 0.70,
+         alt.hypothesis= 0.77, variance="intermediate", power= 0.90, nbr.analyses=6)
> dsn6

Call:
seqDesign(prob.model = "proportions", arms = 1, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = "intermediate", nbr.analyses = 6,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is event probability
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.70    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.77    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Futility Efficacy
```

```
    Time 1 (N=  73.94)   0.5306   0.9542
    Time 2 (N= 147.88)   0.6576   0.8271
    Time 3 (N= 221.82)   0.7000   0.7847
    Time 4 (N= 295.76)   0.7212   0.7635
    Time 5 (N= 369.70)   0.7339   0.7508
    Time 6 (N= 443.64)   0.7424   0.7424
```

We could also have considered a design derived using the variance under the alternative hypothesis

```
> dsn6alt <- seqDesign( prob.model= "proportions", arms= 1, null.hypothesis= 0.70,
+         alt.hypothesis= 0.77, variance="alternative", power= 0.90, nbr.analyses=6)
> dsn6alt

Call:
seqDesign(prob.model = "proportions", arms = 1, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = "alternative", nbr.analyses = 6,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is event probability
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.70    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.77    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                       Futility Efficacy
    Time 1 (N=  67.23)   0.5306   0.9542
    Time 2 (N= 134.46)   0.6576   0.8271
    Time 3 (N= 201.69)   0.7000   0.7847
    Time 4 (N= 268.92)   0.7212   0.7635
    Time 5 (N= 336.15)   0.7339   0.7508
    Time 6 (N= 403.38)   0.7424   0.7424
```

Under the null hypothesis

```
> dsn6null <- seqDesign( prob.model= "proportions", arms= 1, null.hypothesis= 0.70,
+         alt.hypothesis= 0.77, variance="null", power= 0.90, nbr.analyses=6)
> dsn6null

Call:
seqDesign(prob.model = "proportions", arms = 1, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = "null", nbr.analyses = 6,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is event probability
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.70    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.77    (power = 0.900)
```

```
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                      Futility Efficacy
    Time 1 (N=  79.72)   0.5306   0.9542
    Time 2 (N= 159.44)   0.6576   0.8271
    Time 3 (N= 239.16)   0.7000   0.7847
    Time 4 (N= 318.88)   0.7212   0.7635
    Time 5 (N= 398.60)   0.7339   0.7508
    Time 6 (N= 478.32)   0.7424   0.7424
```

Or under a worst case variance of 0.25

```
> dsn6worst <- seqDesign( prob.model= "proportions", arms= 1, null.hypothesis= 0.70,
+         alt.hypothesis= 0.77, variance=0.25, power= 0.90, nbr.analyses=6)
> dsn6worst

Call:
seqDesign(prob.model = "proportions", arms = 1, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = 0.25, nbr.analyses = 6,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is event probability
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.70    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.77    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                      Futility Efficacy
    Time 1 (N=  94.90)   0.5306   0.9542
    Time 2 (N= 189.81)   0.6576   0.8271
    Time 3 (N= 284.71)   0.7000   0.7847
    Time 4 (N= 379.62)   0.7212   0.7635
    Time 5 (N= 474.52)   0.7339   0.7508
    Time 6 (N= 569.42)   0.7424   0.7424
```

In the one sample binomial setting, it is generally very important to evaluate the true operating characteristics of the design. RCTdesign provides functions that can be used to compute exact operating characteristics in the one sample binomial setting, as well as to simulate RCT. (See the tutorials on computing and simulating operating characteristics.)

4.3 Two Sample Inference on the Difference in Proportions

The above one sample test is not very rigorous scientifically, as it pretends that we would definitely know that in the absence of a treatment effect there would 70% 28 day survival. A scientifically better approach would be to conduct a randomized, double blind, placebo controlled study. In such a study, we might let p_k be the probability of 28 day survival on the kth study arm, with $k = 1$ for the treatment and $k = 0$ for placebo. Our measure of treatment effect is then $\theta = p_1 - p_0$. We might presume that under the null

hypothesis $\theta = \theta_0 = 0$, and under the alternative hypothesis $\theta = 0.07$. We again have to consider whether the design should be created assuming the null, alternative, intermediate, or some other hypothesis. In the two sample case, it is probably more common to assume the alternative hypothesis during study design, though the impact of that decision should be evaluated.

Supposing that we desire 1:1 randomization and that we do not want to perform any interim analyses of the data, we can design a two sample test of $H_0 : \theta \leq 0$ versus $H_1 : \theta \geq 0.07$ with the following RCTdesign code, which creates a `"seqDesign"` object named `dsnFixed`. In the proportions probability model, we must specify the actual hypothesized proportions to allow the computation of the variance. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "proportions",
+         arms= 2,
+         ratio= c(1, 1),
+         null.hypothesis= 0.70,
+         alt.hypothesis= 0.77,
+         variance= "alternative",
+         test.type= "greater",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed

Call:
seqDesign(prob.model = "proportions", arms = 2, null.hypothesis = 0.7,
    alt.hypothesis = 0.77, variance = "alternative", ratio = c(1,
        1), test.type = "greater", size = 0.025, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in probabilities (Treatment - Comparison)
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.00    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.07    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                        Futility Efficacy
    Time 1 (N= 1660.17)   0.0423   0.0423
```

Note that the printed description of the stopping rule now identifies that

- The treatment effect θ is the difference in event probabilities, where the difference is the treatment arm probability minus the control arm probability.
- The clinical trial design corresponds to a one-sided test of a greater alternative.
- The null hypothesis is $\theta = 0$ is tested with a type I error of 0.025.
- The sample size corresponds to a test having 90% power to detect an alternative hypothesis of $\theta = 0.07$.
- The stopping boundaries are printed on the scale of the estimated treatment effect.

- A sample size of 1660.17 (189.14 on each study arm) would dictate that observed difference in event probabilities greater than 0.0423 (i.e., the observed proportion surviving 28 days on the treatment arm was at least 0.0423 greater than the 28 day survival probability on the control arm) would suggest rejection of the null hypothesis. (It should be noted that for these numbers to be accurate, the true event probabilities on each of the study arms would need to be close to the hypothesized values. Otherwise, the precision of the study might be markedly different.

Note again that the shorter printout does not include the randomization ratio, though this can be obtained by obtaining the more detailed output that is the default when printing the `"seqModel"` object that is stored in every `"seqDesign"` object as demonstrated with the mean and geometric mean probability models.

In a two arm study, each of the arguments `ratio`, `null.hypothesis`, and `alt.hypothesis` can actually take values for each arm. In each case, the first value specified corresponds to the treatment arm, and the second value corresponds to the control arm.

- When a single value is specified for `ratio`, that value is presumed to correspond to the treatment arm, and the default value for the control arm is 1. Hence, specifying `ratio= 3` is equivalent to specifying `ratio= c(3,1)`, and specifying `ratio= 0.5` is equivalent to specifying `ratio= c(0.5,1)` (or `ratio= c(1,2)`).
- When a single value is specified for `null.hypothesis`, that value is presumed to correspond to the treatment arm under the null hypothesis, and by default it is also used for the control arm under the null hypothesis. (When no value is specified, the default value is 0.5.)
- When a single value is specified for `alt.hypothesis`, that value is presumed to correspond to the treatment arm under the null hypothesis, and the default value for the control arm under the alternative hypothesis is the same as is used for the control arm under the null hypothesis. (There is no default value for the alternative, and when no alternative hypothesis value is specified, RCTdesign will find the alternative that has the desired power for a specified sample size.)

Hence, we could have specified our design using

```
> dsnFixed <- seqDesign(
+         prob.model= "proportions",
+         arms= 2,
+         ratio= c(1, 1),
+         null.hypothesis= c(0.7, 0.7),
+         alt.hypothesis= c(0.77, 0.7),
+         variance= "alternative",
+         test.type= "greater",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed$model

PROBABILITY MODEL:
   Two arm study of binary response variable
   Randomization scheme:  1 treatment group : 1 comparison group
   Outcome summarized by event probability
                   Treatment Comparison
              Null      0.70        0.7
       Alternative      0.77        0.7
```

```
   Treatment effect is difference in probabilities (Treatment - Comparison)
                       Theta
              Null      0.00
       Alternative      0.07

   (Standardization parameters:  Mu0 0; SigmaSqr 0.3871; Psi 1)
```

Several of the arguments that we supplied to `seqDesign()` in this instance were the default values (in particular, note that a two arm study with 1:1 randomization is the default value). Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( "proportions", null.hypothesis= 0.7, alt.hypothesis= 0.77,
+         power= 0.90)
> dsnFixed

Call:
seqDesign(prob.model = "proportions", null.hypothesis = 0.7,
    alt.hypothesis = 0.77, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in probabilities (Treatment - Comparison)
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.00    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.07    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                        Futility Efficacy
    Time 1 (N= 1660.17)   0.0423   0.0423
```

Had we wanted to create a design that would conduct up to, say, 4 equally spaced analyses of the data using O'Brien-Fleming boundaries in a one-sided symmetric design, we could execute

```
> dsn4 <- seqDesign( "proportions", null.hypothesis= 0.7, alt.hypothesis= 0.77,
+         power= 0.90, nbr.analyses= 4)
> dsn4

Call:
seqDesign(prob.model = "proportions", null.hypothesis = 0.7,
    alt.hypothesis = 0.77, nbr.analyses = 4, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is difference in probabilities (Treatment - Comparison)
   One-sided hypothesis test of a greater alternative:
           Null hypothesis : Theta <= 0.00    (size  = 0.025)
    Alternative hypothesis : Theta >= 0.07    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                        Futility Efficacy
```

```
Time 1 (N=  432.95)  -0.0847   0.1694
Time 2 (N=  865.90)   0.0000   0.0847
Time 3 (N= 1298.85)   0.0282   0.0565
Time 4 (N= 1731.80)   0.0424   0.0424
```

5 Inference About Hazard Ratios

5.1 Hypothetical Example: Time to Death in Non-Small Cell Lung Cancer

We consider a clinical trial of a new chemotherapeutic agent in the treatment of second-line non-small cell lung cancer (NSCLC). The primary clinical endpoint T_i for each subject is the time of death, though observations of some subjects may be censored in that the subject is still alive at the time of data anaysis. The planned analysis is based on the hazard ratio as might be tested with the logrank test or in a proportional hazards regression model. Important design parameters include

- *Null hypothesis*: In the absence of a treatment effect, we presume that the (average) ratio of hazard functions will be 1.0.
- *Alternative hypothesis*: We presume that prior pilot studies have suggested that the treatment might provide a 25% decrease in the risk of death at each instant of death. Hence, under the alternative hypothesis we anticipate a hazard ratio of 0.75.
- *Type I error*: We presume that standards of evidence demand a one-sided type I error of $\alpha = 0.025$.
- *Design power*: We presume that we desire 90% statistical power to reject the null hypothesis when the alternative hypothesis is in fact true.
- *Test statistic*: We presume that statistical inference will be based on the asymptotic normal distribution for the estimated hazard ratio and/or the logrank statistic.
- *Variability*: In the proportional hazards model, statistical information is proportional to the number of events, and the constant of proportionality is estimated using only the randomization ratio. (See the tutorial on time to event analytic models.)

5.2 Two Sample Inference on the Hazard Ratio

We consider a randomized, double blind, placebo controlled study in which the treatment effect θ is summarized by the hazard ratio. We might presume that under the null hypothesis $\theta = \theta_0 = 1.0$, and under the alternative hypothesis $\theta = 0.75$.

Supposing that we desire 1:1 randomization and that we do not want to perform any interim analyses of the data, we can design a two sample test of $H_0 : \theta \geq 1.0$ versus $H_1 : \theta \leq 0.75$ with the following RCTdesign code, which creates a `"seqDesign"` object named `dsnFixed`. We obtain a printout by merely typing the object name.

```
> dsnFixed <- seqDesign(
+         prob.model= "hazard",
+         arms= 2,
+         ratio= c(1, 1),
+         null.hypothesis= 1.0,
+         alt.hypothesis= 0.75,
```

```
+         test.type= "less",
+         size= 0.025,
+         power= 0.90
+         )
> dsnFixed

Call:
seqDesign(prob.model = "hazard", arms = 2, null.hypothesis = 1,
    alt.hypothesis = 0.75, ratio = c(1, 1), test.type = "less",
    size = 0.025, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is hazard ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.00    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.75    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                        Efficacy Futility
    Time 1 (NEv= 507.84)   0.8403   0.8403
```

Note that the printed description of the stopping rule now identifies that

- The treatment effect θ is the hazard ratio, where the ratio is the treatment arm hazard (instantaneous rate of failure) divided by the control arm hazard.
- The clinical trial design corresponds to a one-sided test of a lesser alternative.
- The null hypothesis is $\theta = 1.0$ is tested with a type I error of 0.025.
- The sample size corresponds to a test having 90% power to detect an alternative hypothesis of $\theta = 0.75$.
- The stopping boundaries are printed on the scale of the estimated treatment effect, which is the hazard ratio estimate as might be estimated in a proportional hazards regression.
- After observing 507.84 events (on both arms combined), we would be able to reject the null hypothesis if the observed hazard ratio estimate were less than 0.8403.

Note again that the shorter printout does not include the randomization ratio, though this can be obtained by obtaining the more detailed output that is the default when printing the `"seqModel"` object that is stored in every `"seqDesign"` object as demonstrated with the mean and geometric mean probability models.

Unlike in other probability models, the arguments `null.hypothesis` and `alt.hypothesis` can only take a scalar value: the hazard ratio is the comparison across study arms.

Several of the arguments that we supplied to `seqDesign()` in this instance were the default values (in particular, note that a two arm study with 1:1 randomization is the default value). Hence, we could have obtained the exact same design by specifying

```
> dsnFixed <- seqDesign( "hazard", alt.hypothesis= 0.75, power= 0.90)
> dsnFixed
```

```
Call:
seqDesign(prob.model = "hazard", alt.hypothesis = 0.75, power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is hazard ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.00    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.75    (power = 0.900)
   (Fixed sample test)

STOPPING BOUNDARIES: Sample Mean scale
                         Efficacy Futility
    Time 1 (NEv= 507.84)   0.8403   0.8403
```

Had we wanted to create a design that would conduct up to, say, 4 equally spaced analyses of the data using O'Brien-Fleming boundaries in a one-sided symmetric design, we could execute

```
> dsn4 <- seqDesign( "hazard", alt.hypothesis= 0.75, power= 0.90,
+         nbr.analyses= 4)
> dsn4

Call:
seqDesign(prob.model = "hazard", alt.hypothesis = 0.75, nbr.analyses = 4,
    power = 0.9)

PROBABILITY MODEL and HYPOTHESES:
   Theta is hazard ratio (Treatment : Comparison)
   One-sided hypothesis test of a lesser alternative:
           Null hypothesis : Theta >= 1.00    (size  = 0.025)
    Alternative hypothesis : Theta <= 0.75    (power = 0.900)
   (Emerson & Fleming (1989) symmetric test)

STOPPING BOUNDARIES: Sample Mean scale
                         Efficacy Futility
    Time 1 (NEv= 132.44)   0.4984   1.4164
    Time 2 (NEv= 264.88)   0.7060   1.0000
    Time 3 (NEv= 397.32)   0.7929   0.8904
    Time 4 (NEv= 529.76)   0.8402   0.8402
```

As with the geometric mean probability model, a user may specify `log.transform=TRUE` in order to use the log hazard ratio as the primary summary measure for specifying hypotheses and critical values.

It should be noted that when using the above code, `seqDesign()` returned the number of observed events needed to provide the desired precision. In practice, however, we will also want to determine the number of subjects that should be accrued. The RCTdesign functions `seqPHSubjects`, `seqAccrual()`, and `seqPHNSubjects` can be used to explore different accrual patterns that would yield the desired number of events in a specified calendar time. While arguments to `seqAccrual()` and `seqPHNSubjects` can be supplied to `seqDesign()` in order to make the trial design a one-step process, we do not here present the details regarding those arguments. Instead, see the tutorials on specification of accrual in time to event probability models.