Statistics 641, Fall 2009
Take-home Final Exam
Solutions

1. Suppose that we have a phase II, single arm trial using a two stage design and let π be the response probability. The hypotheses of interest are H_0: $\pi \leq 0.20$ versus H_1: $\pi \geq 0.5$. Let x_1 and x_2 be the number of positive responses in stages 1 and 2, and $y_1 = x_1$, and $y_2 = x_1 + x_2$ be the total number of responses through stages 1 and 2.

   (a) We enroll 14 subjects in stage 1 and stop and accept H_0 if we observe $a_1 = 4$ or fewer responses (i.e., $y_1 \leq 4$), otherwise we continue to stage 2. Find the stopping probabilities under both H_0 and H_1.

   Let $p(x, \pi) = \binom{14}{x} \pi^x (1-\pi)^{14-x}$. The stopping probability is $\sum_{x_1=0}^{4} p(x_1, \pi)$. When $\pi = 0.2$, this sum is .8702, when $\pi = .5$, it is .0898.

   (b) At stage 2 we enroll an additional 14 subjects. Find the smallest value of a_2 so that if we reject H_0 when $y_2 > a_2$, the overall type I error probability will be at most $\alpha = .05$. For this value of a_2 what is the probability of rejecting H_0 when H_1 is true?

   If we reject at stage 2 if $y_2 > a_2$, then when H_0 is true, the probability of rejecting H_0 at the end of the trial is

$$P_R(a_2, \pi) = 1 - \sum_{x_1=0}^{4} p(x_1, \pi) - \sum_{x_1=5}^{a_2} p(x_1, \pi) \sum_{x_2=0}^{a_2 - x_1} p(x_2, \pi)$$

   By trial and error, we find that $P_R(8, 0.2) = 0.053$ and $P_R(9, 0.2) = 0.028$, so we choose $a_2 = 9$.
   When H_1 is true, we have $P_R(9, 0.5) = 0.892$

   (c) Suppose that in addition to the stopping rule from (a) we stop and reject H_0 at stage 1 if we observe $r_1 = 7$ or more responses. Find the probability of rejecting H_0 at stage 1 under H_0 and H_1.

   The probability of rejecting H_0 at stage 1 is $\sum_{x_1=7}^{14} p(x_1, \pi)$. When $\pi = 0.2$, this sum is .0116, and when $\pi = .5$, it is .6047.

   (d) If using the stopping rule from (c), we again enroll an additional 14 subjects stage 2, find the smallest value of a_2 so that if we reject H_0 when $y_2 > a_2$, the overall type I error probability will be at most $\alpha = .05$. For this value of a_2 what is the probability of rejecting H_0 when H_1 is true?

The probability of rejection is

$$P_R(a_2, \pi) = 1 - \sum_{x_1=0}^{4} p(x_1, \pi) - \sum_{x_1=5}^{6} p(x_1, \pi) \sum_{x_2=0}^{a_2 - x_1} p(x_2, \pi)$$

Similar to (b), we have $P_R(8, .2) = 0.055$ and $P_R(9, .2) = 0.033$, so we choose $a_2 = 9$. When $\pi = .5$, we have $P_R(9, .5) = 0.8940$.

(e) Compute the average sample size for the two designs above under both H_0 and H_1.

The average sample size is $14 \times \Pr\{\text{stop at stage 1}\} + 28 \times \Pr\{\text{don't stop at stage 1}\} = 28 - 14 \times \Pr\{\text{stop at stage 1}\}$.

For design (b) using (a), we have expected sample sizes of $28 - 14 \times 0.8702 = 15.8$ and $28 - 14 \times 0.0898 = 26.7$.

For design (d) using (a) and (c), we have expected sample sizes of $28 - 14 \times (0.8702 + 0.0116) = 15.7$ and $28 - 14 \times (0.0898 + 0.6047) = 18.3$.

2. Heart patients have a greater risk of a second heart attack (MI) immediately following the first MI, then they do later on. Suppose (simplistically) that with standard therapy, the hazard rate λ is constant .12/year during the first 6 months following MI and constant .05/year thereafter. Suppose further that a new treatment is expected to reduce these rates by 25%. We wish to perform a study of patients enrolled immediately following an MI with (uniform) enrollment and followup of either:

   (a) 1.5 year enrollment, 4 year followup

   (b) 2 year enrollment, 3.5 year followup

   Compute the required sample sizes for each of these two designs assuming equal numbers of patients in each treatment group. (*Hint: to find the probability of failure, write the cumulative hazard function as piecewise linear function, then find the survivor function.*)

First, assuming $\alpha = 0.05$ and $\beta = .1$, according to Schoenfeld's formula we have that the required number of events is

$$D = \frac{4(1.96 + 1.28)^2}{\log(.75)^2} = 508$$

For simplicity, we define the average hazard function

$$\bar{\lambda}(t) = \begin{cases} (0.12 + 0.12 \times .75) = 0.105 & \text{if } t \leq 0.5 \\ (0.05 + 0.05 \times .75) = 0.04375 & \text{if } t > 0.5 \end{cases}$$

When $t \leq 0.5$ (for t in years), the cumulative hazard function is:

$$\bar{\Lambda}(t) = \int_0^t \bar{\lambda}(t) ds = \int_0^t 0.105 ds = 0.105t.$$

For $t > .5$, the cumulative hazard function is:

$$\begin{aligned}\bar{\Lambda}(t) &= \int_0^t \bar{\lambda}(u)du = \int_0^{0.5} \bar{\lambda}(u)du + \int_{0.5}^t \bar{\lambda}(u)du \\ &= 0.105 \times 0.5 + 0.04375 \times (t - 0.5) = 0.030625 + 0.04375t\end{aligned}$$

Hence, the mean survivor function is

$$\bar{S}(t) = \begin{cases} e^{-0.105t} & \text{if } t \leq 0.5 \\ e^{-0.030625-0.04375t} & \text{if } t > 0.5 \end{cases}$$

We need to compute the probability of failure by integrating over the censoring distribution which is uniform on an interval $(F - R, F)$ where F is the maximum follow-up time, and R is the length of the enrollment period.

$$\Pr\{\text{failure}\} = \frac{1}{R}\int_{F-R}^{F} 1 - \bar{S}(u)du$$

Since for both scenarios above, $F - R > 0.5$, the integrand only needs to be evaluated for $u > 0.5$ so we don't need to break it into pieces.

$$\begin{aligned}\Pr\{\text{failure}\} &= \frac{1}{R}\int_{F-R}^{F} 1 - e^{-0.030625-0.04375u}du \\ &= 1 + \frac{e^{-0.030625}}{0.04375R} e^{-0.04375u}\Big|_{F-R}^{F} \\ &= 1 + \frac{e^{-0.030625}}{0.04375R}\left(e^{-0.04375F} - e^{-0.04375(F-R)}\right)\end{aligned}$$

Plugging in $F = 4$, $R = 1.5$, we have

$$\Pr\{\text{failure}\} = 0.15855$$

so the required number of subjects is

$$\frac{507}{0.15855} = 3198$$

Plugging in $F = 3.5$, $R = 2$, we have

$$\Pr\{\text{failure}\} = 0.13036$$

so the required number of subjects is

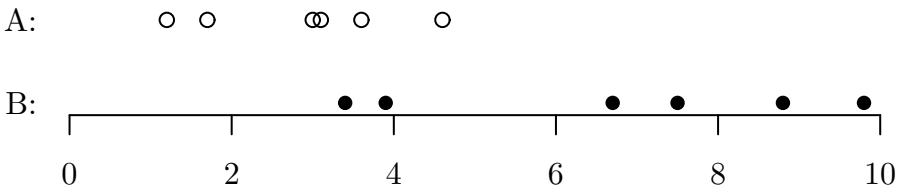
$$\frac{507}{0.13036} = 3889$$

3. Suppose that we observe the following continuous responses for groups A and B.

   A: 3.0, 1.7, 4.6, 3.6, 1.2, 3.1
   B: 3.9, 6.7, 7.5, 3.4, 8.8, 9.8

   Let $\hat{\mu}$ be then difference in means, $\bar{X}_B - \bar{X}_A$. Compute a one-sided p-value under the randomization distribution generated by the random allocation rule ($N_A = N_B$). (Note that you don't need to generate the entire randomization distribution for $\hat{\mu}$.)

First, we have $\hat{\mu} = 6.683 - 2.867 = 3.817$. A plot of the data shows that most of the observations in group B are well to the right of the observations in group A.



There are only small number of treatment allocations that produce a more extreme mean difference than the one observed, specifically, we need only consider allocations such that the mean in group B is larger than the observed mean, 6.683. These are enumerated below:

Allocated to B	mean
3.9, 4.6, 6.7, 7.5, 8.8, 9.8	6.883
3.6, 4.6, 6.7, 7.5, 8.8, 9.8	6.833
3.4, 4.6, 6.7, 7.5, 8.8, 9.8	6.800
3.1, 4.6, 6.7, 7.5, 8.8, 9.8	6.750
3.0, 4.6, 6.7, 7.5, 8.8, 9.8	6.733
3.6, 3.9, 6.7, 7.5, 8.8, 9.8	6.717
3.4, 3.9, 6.7, 7.5, 8.8, 9.8	6.683

Therefore, there are 7 allocations yielding a mean difference at least as extreme as the one observed. There are a total of

$$\binom{12}{6} = 924$$

possible allocations, so the one-sided p-value is $7/924 = .0076$.

---

4. Suppose that we are using minimization as a covariate adaptive allocation scheme and we wish to balance with respect to smoking status and sex. Using the notation from class let $G_t = |x^t_{11} - x^t_{12}| + |x^t_{21} - x^t_{22}|$.

   Suppose that the next subject is a non-smoking female and we have:

	Smoker		Sex		
Group	Y	N	M	F	Total
1	15	26	19	22	41
2	16	28	21	23	44

To which treatment group should the next subject be allocated?

If the next subject is allocated to group 1, the table will be:

	Smoker		Sex		
Group	Y	N	M	F	Total
1	15	27	19	23	42
2	16	28	21	23	44

so $G_1 = |27 - 28| + |23 - 23| = 1$.

If the next subject is allocated to group 2, the table will be:

	Smoker		Sex		
Group	Y	N	M	F	Total
1	15	26	19	22	42
2	16	29	21	24	44

so $G_1 = |26 - 29| + |22 - 24| = 4$.

Therefore, the next subject should be allocated to group 1.

---

5. Suppose that a trial is designed to have 80% power at a one-sided significance level of $\alpha = 0.01$ when no interim monitoring boundaries are in use. Note that the possibility of early stopping with control of the overall type I error rate will always reduce power.

   (a) Assuming that 3 interim analysis occur at equal increments of information (25%, 50% and 75%), compute stopping boundaries and final critical values for each the following α-spending functions with one sided $\alpha = 0.01$:

      i. O'Brien-Fleming

      ii. Linear: $\alpha(t) = 0.01t$

In R:

```
> boundi <- bounds(1:4/4, iuse=1, alpha=.01)
>  summary(boundi)
Lan-DeMets bounds for a given spending function
n =  4
Overall alpha:  0.01
Type: One-Sided Bounds
alpha:  0.01
Spending function: O'Brien-Fleming
Boundaries:
   Time   Upper       Exit pr.      Diff. pr.
1  0.25  5.0201  0.00000025819  0.00000025819
```

```
2  0.50  3.4605  0.00026971696  0.00026945876
3  0.75  2.7649  0.00293646827  0.00266675132
4  1.00  2.3625  0.01000000000  0.00706353173
> boundii <- bounds(1:4/4, iuse=3, phi=1, alpha=.01)
> summary(boundii)
Lan-DeMets bounds for a given spending function
n =  4
Overall alpha:  0.01
Type: One-Sided Bounds
alpha:  0.01
Spending function: Power Family: alpha * t^phi
Boundaries:
   Time   Upper  Exit pr.  Diff. pr.
1  0.25  2.8070    0.0025     0.0025
2  0.50  2.7403    0.0050     0.0025
3  0.75  2.6724    0.0075     0.0025
4  1.00  2.6117    0.0100     0.0025
```

(b) Repeat (a) if there are 6 planned interim analyses at equally spaced intervals (increments of 1/7 information)

```
> boundi6 <- bounds(1:7/7, iuse=1, alpha=.01)
> summary(boundi6)
Lan-DeMets bounds for a given spending function
n =  7
Overall alpha:  0.01
Type: One-Sided Bounds
alpha:  0.01
Spending function: O'Brien-Fleming
Boundaries:
      Time   Upper    Exit pr.   Diff. pr.
1  0.14286  6.7146  9.4262e-12  9.4262e-12
2  0.28571  4.6918  1.4433e-06  1.4433e-06
3  0.42857  3.7667  8.3320e-05  8.1877e-05
4  0.57143  3.2258  6.5560e-04  5.7228e-04
5  0.71429  2.8635  2.3055e-03  1.6499e-03
6  0.85714  2.6004  5.3989e-03  3.0934e-03
7  1.00000  2.3985  1.0000e-02  4.6011e-03
> boundii6 <- bounds(1:7/7, iuse=3, phi=1, alpha=.01)
> summary(boundii6)
Lan-DeMets bounds for a given spending function
n =  7
Overall alpha:  0.01
```

```
Type: One-Sided Bounds
alpha:  0.01
Spending function: Power Family: alpha * t^phi
Boundaries:
      Time   Upper   Exit pr.  Diff. pr.
1  0.14286  2.9827  0.0014286  0.0014286
2  0.28571  2.9267  0.0028571  0.0014286
3  0.42857  2.8672  0.0042857  0.0014286
4  0.57143  2.8134  0.0057143  0.0014286
5  0.71429  2.7654  0.0071429  0.0014286
6  0.85714  2.7221  0.0085714  0.0014286
7  1.00000  2.6828  0.0100000  0.0014286
```

(c) For each of the boundaries in (a) and (b) compute power under the assumed alternative hypothesis.

Under the assumed alternative hypothesis, the drift parameter is $Z_{1-\alpha}+Z_{1-\beta} = Z_{1-.01}+Z_{1-.2} = 3.168$.

```
> d <- qnorm(.99) + qnorm(.8)
> summary(drift(zb=boundi$upper, t=boundi$time, drft=d))
[snip]
Drift parameters:  3.167969
   Time  Lower probs  Upper probs    Exit pr.  Cum exit pr.
1  0.25   1.9996e-11   0.00029504  0.00029504    0.00029504
2  0.50   5.9667e-09   0.11086267  0.11086268    0.11115771
3  0.75   1.7427e-08   0.38212772  0.38212774    0.49328545
4  1.00   1.3752e-08   0.30119314  0.30119315    0.79447861
> summary(drift(zb=boundii$upper, t=boundii$time, drft=d))
[snip]
Drift parameters:  3.167969
   Time  Lower probs  Upper probs  Exit pr.  Cum exit pr.
1  0.25   5.6411e-06      0.11066   0.11066       0.11066
2  0.50   2.6891e-07      0.21749   0.21749       0.32815
3  0.75   2.2218e-08      0.23059   0.23059       0.55873
4  1.00   2.4032e-09      0.18154   0.18154       0.74027
> summary(drift(zb=boundi6$upper, t=boundi6$time, drft=d))
[snip]
Drift parameters:  3.167969
      Time  Lower probs  Upper probs    Exit pr.  Cum exit pr.
1  0.14286   1.2662e-15   1.7216e-08  1.7216e-08    1.7216e-08
2  0.28571   8.5588e-11   1.3566e-03  1.3566e-03    1.3566e-03
3  0.42857   2.5832e-09   4.3961e-02  4.3961e-02    4.5317e-02
4  0.57143   8.9744e-09   1.5948e-01  1.5948e-01    2.0479e-01
```

```
5  0.71429   1.2912e-08   2.2772e-01  2.2772e-01    4.3252e-01
6  0.85714   1.2066e-08   2.0784e-01  2.0784e-01    6.4036e-01
7  1.00000   8.9180e-09   1.4974e-01  1.4974e-01    7.9009e-01
> summary(drift(zb=boundii6$upper, t=boundii6$time, drft=d))
[snip]
Drift parameters:  3.167969
      Time  Lower probs  Upper probs  Exit pr.  Cum exit pr.
1  0.14286   1.4570e-05     0.037104  0.037119      0.037119
2  0.28571   1.6462e-06     0.084283  0.084285      0.121403
3  0.42857   2.8733e-07     0.119798  0.119799      0.241202
4  0.57143   6.2134e-08     0.135430  0.135430      0.376632
5  0.71429   1.5331e-08     0.133177  0.133177      0.509809
6  0.85714   4.1438e-09     0.119040  0.119040      0.628849
7  1.00000   1.1978e-09     0.099172  0.099172      0.728021
```

For (a), power is 79.45% and 74.03% for boundaries (i) and (ii) respectively. For (b), the corresponding power is 79.01% and 72.80%. For the more conservative OBF spending function (i), there is very little power loss due to interim monitoring, even for 6 interim analyses. For the linear spending function (ii), power loss is greater, although the additional interim analyses yield little additional power loss. (For spending function (ii) power loss is roughly equivalent to a reduction in sample size of 12% and 14%: $1-(Z_{1-\alpha}+Z_{1-\beta^*})^2/(Z_{1-\alpha}+Z_{1-\beta})^2$ where $\beta=.2$ is the planned type II error rate and β^* is the actual type II error rate after accounting for interim analyses.)

6. Suppose you are designing a trial whose primary outcome is all-cause (ACM) mortality and that the number of deaths required to achieve the desired power is 750. Construct interim monitoring boundaries using α-spending functions satisfying the following conditions:

   (I) Since the new drug is not likely to be approved unless the evidence for benefit is sufficiently compelling, the trial should not be stopped for benefit unless the p-value for ACM is quite small—in particular, the test-statistic should be required to exceed a boundary defined by a spending function $\alpha_U(t)$, where $\alpha_U(t) < 0.005$ for $t < 1$. (Note that this guarantees that the stage-wise adjusted p-value is less than 0.005 if the trial is stopped early.)

   (II) The lower boundary should be considered a safety boundary, so that compelling evidence of harm is not required—the goal is to stop if there is evidence of harm that is unlikely to occur by chance alone. The overall probability of stopping for harm should be no greater than 0.01 under the null hypothesis ($\alpha_L(t) \leq 0.01$ for $t < 1$).

   (a) Define upper and lower spending functions and compute the critical values if analysis will occur at information fractions $t = 0.25, 0.5, 0.75, 1.0$.

   There are many acceptable answers, so I illustrate one.

   For the upper boundary, the requirement that $\alpha_U(t) < 0.005$ for $t < 1$ forces the boundary to be fairly conservative regardless of the spending function that we might use, so a

more aggressive shape (e.g., linear) might be appropriate.

$$\alpha_U(t) = \begin{cases} 0.005t & \text{for } t < 1 \\ 0.025 & \text{for } t = 1 \end{cases}$$

For the lower boundary because we are more concerned about potential harm, we want a less conservative boundary (easier to stop). A reasonable choice is

$$\alpha_L(t) = 0.01t$$

```
## Define upper alpha-spending function
> upper <- function(t) ifelse(t < 1, .005*t/.025, 1)
> bound6U <- bounds(1:4/4, iuse=5, alpha=.025, asf=upper)
> summary(bound6U)
Lan-DeMets bounds for a given spending function
n =  4
Overall alpha:  0.025
Type: One-Sided Bounds
alpha:  0.025
Spending function: User-specified spending function
Boundaries:
   Time   Upper  Exit pr.  Diff. pr.
1  0.25  3.0233   0.00125    0.00125
2  0.50  2.9696   0.00250    0.00125
3  0.75  2.9118   0.00375    0.00125
4  1.00  1.9833   0.02500    0.02125
> bound6L <- bounds(1:4/4, iuse=3, phi=1, alpha=.01)
> summary(bound6L)
Lan-DeMets bounds for a given spending function
n =  4
Overall alpha:  0.01
Type: One-Sided Bounds
alpha:  0.01
Spending function: Power Family: alpha * t^phi

Boundaries:
   Time   Upper  Exit pr.  Diff. pr.
1  0.25  2.8070    0.0025     0.0025
2  0.50  2.7403    0.0050     0.0025
3  0.75  2.6724    0.0075     0.0025
4  1.00  2.6117    0.0100     0.0025
```

(Note that because I've computed the upper and lower boundaries separately, the critical values for the lower boundary are given as positive numbers. The signs for these should be negative.)

(b) Suppose that the true hazard ratio is 0.78 (22% reduction in hazard) in favor of the new treatment. Find power for this alternative hypothesis given the stopping boundaries defined in the previous part.

To compute power we need the drift parameter which we can find by inverting Schoenfeld's formula:

$$\text{drift } = Z_{1-\alpha/2} + Z_{1-\beta} = \frac{\sqrt{750}}{2}|\log(0.78)| = 3.402$$

```
> summary(drift(zb=bound6U$upper, t=bound6U$time, drft=d))
[snip]
Drift parameters:  3.402197
   Time  Lower probs  Upper probs  Exit pr.  Cum exit pr.
1  0.25   1.1538e-06     0.093044  0.093045      0.093045
2  0.50   3.3165e-08     0.210005  0.210005      0.303050
3  0.75   1.7690e-09     0.237660  0.237660      0.540710
4  1.00   3.4347e-08     0.382186  0.382186      0.922895
```

Hence, power is 92.3%.

(c) Suppose that interim analyses have occurred when 105, 230 and 465 deaths have been observed. At the third analysis, we observe a Z-score for ACM of $Z = 3.04$ (in favor of the new treatment). Should the trial stop based on a finding of benefit?

```
> summary(bounds(c(105, 230, 465)/750, iuse=5, alpha=.025, asf=upper))
Lan-DeMets bounds for a given spending function
n =  3
Overall alpha:  0.025
Type: One-Sided Bounds
alpha:  0.025
Spending function: User-specified spending function
Boundaries:
      Time   Upper   Exit pr.   Diff. pr.
1  0.14000  3.1947  0.0007000  0.00070000
2  0.30667  3.1079  0.0015333  0.00083333
3  0.62000  2.9092  0.0031000  0.00156667
```

The upper critical value at the third analysis using this spending function is 2.909. Hence the boundary has been crossed that the study can be stopped for benefit. (If a more conservative boundary had been used, the boundary may not have been crossed, so the answer here depends on your choice of spending function in (a).)

(d) Suppose instead that the Z-score for ACM is .41 after 520 deaths have been observed. What is the conditional power for showing benefit for ACM if the alternative hypothesis is that there is a 22% reduction in hazard?

To use the formula given in class, we need to convert the Z statistic to the B statistic $B = Z\sqrt{t} = 0.41\sqrt{520/750} = 0.341$.

Conditional power will also depend on the expected final critical value, which depends on the spending function and the interim analysis times, however, because we have used a conservative boundary, this dependence will be minimal.

E.g., if we assume the analysis times in (c) plus the one at 520 events (69.3% information) and no additional analyses, we would have:

```
> summary(bounds(c(105, 230, 465, 520,750)/750, iuse=5, alpha=.025, asf=upper))
[snip]
      Time   Upper   Exit pr.   Diff. pr.
...
5  1.00000  1.9841  0.0250000  0.02153333
```

for a final critical value of 1.984 (if we add an intermediate analysis at, say 600, events, the critical value goes up to 1.9844, so the dependence on the analysis times is quite small).

Conditional power is therefore

$$\begin{aligned} 1 - \Phi\left(\frac{b(1) - B(t_0) - (1-t_0)\theta}{\sqrt{1-t_0}}\right) &= 1 - \Phi\left(\frac{1.984 - 0.341 - (1-.693)3.402}{\sqrt{1-.693}}\right) \\ &= 14.0\% \end{aligned}$$

It is relatively unlikely that we will be able to reject H_0 at the conclusion of the trial.