Term problem set

Alyssa Vanderbeek (amv2187)

2 December 2019

Problem 1: (Preliminaries) For a randomized, placebo-controlled study of a statin for lowering LDL, we consider a reduction in LDL by 10mg/dL on average to be clinically meaningful, whereas patients in the placebo will experience no change on average. Assume that the variance of LDL is equal to 20mg/dL at both baseline and the 3-month follow-up in both groups.

a. What additional information or assumption(s) do you need to calculate a sample size?

We need to know the following in order to compute a sample size:

target type I error rate and power; the average LDL level in the placebo group; *the correlation between paired measurements (within-individual);* the variance of measurements at each time point (between-individual).

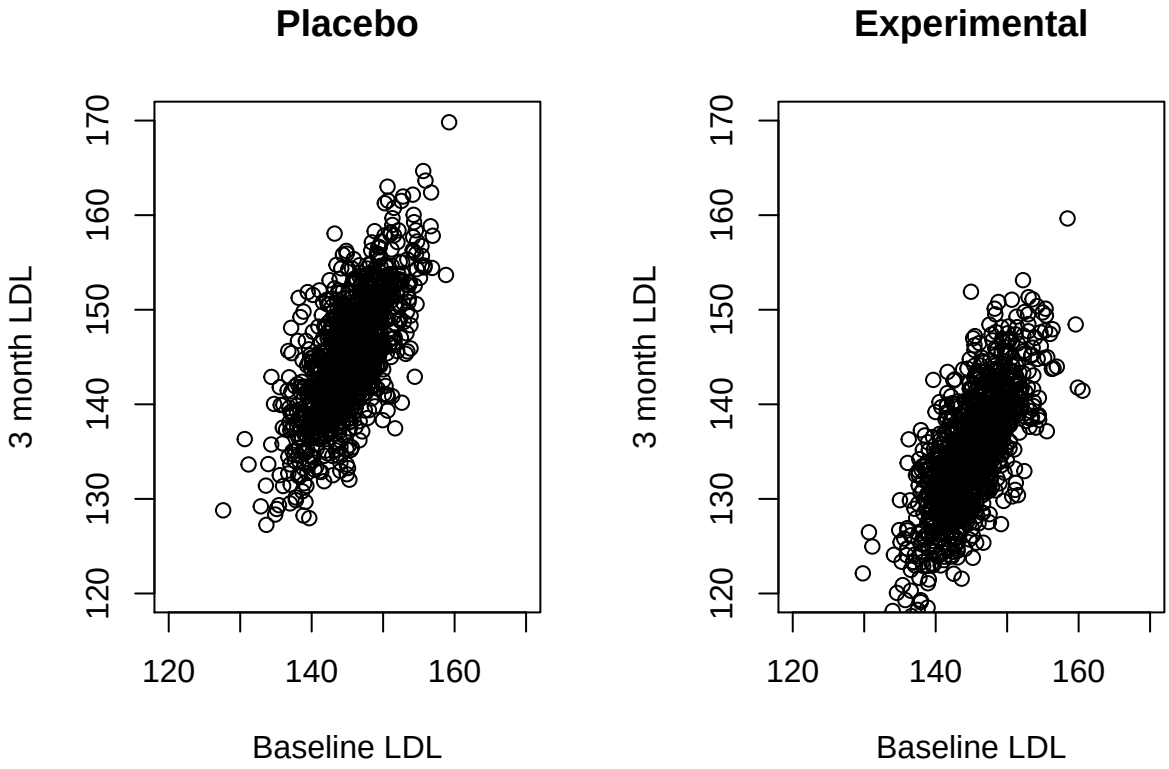
b. State the assumptions you make and calculate the sample size required for a two-sided test at 5% significance.

Let's suppose that $\alpha = 0.05$ and target power is 0.9. Consider that the population LDL is approximately normally distributed.

Since the primary outcome is the difference in differences between the two groups, we also have to make assumptions about the correlation between the measurements in each group at baseline and 3 months. I used simulation to estimate a reasonable correlation between the paired measurements in each treatment group.

```
n.sim = 1e3

d.null = 0
d.alt = 10
var = 20

ldl.null = 145
ldl.alt = ldl.null - d.alt

cor_sim = function(ldl, d, var, vals = FALSE){
  p_base = rnorm(1e3, ldl, sqrt(var))
  add = rnorm(1e3, d, sqrt(var))
  p_3 = p_base - add

  if (vals == TRUE) {
    return(list(c = cor(p_base, p_3),
                p_base = p_base,
                p_3 = p_3))
  } else {
    return(cor(p_base, p_3))
  }
}

placebo = cor_sim(ldl = ldl.null, d = d.null, var = var, vals = TRUE)
exp = cor_sim(ldl = ldl.null, d = d.alt, var = var, vals = TRUE)
```

```
par(mfrow = c(1, 2))
plot(x = placebo$p_base, y = placebo$p_3,
     xlab = "Baseline LDL", ylab = "3 month LDL", main = "Placebo",
     xlim = c(120, 170), ylim = c(120, 170))
plot(x = exp$p_base, y = exp$p_3,
     xlab = "Baseline LDL", ylab = "3 month LDL", main = "Experimental",
     xlim = c(120, 170), ylim = c(120, 170))
```




```
cor_placebo = sapply(X = 1:n.sim, function(i){
  cor_sim(ldl = ldl.null, d = d.null, var = var)
})
mean(cor_placebo)
```

```
## [1] 0.7068301
```

```
cor_exp = sapply(X = 1:n.sim, function(i){
  cor_sim(ldl = ldl.null, d = d.alt, var = var)
})
mean(cor_exp)
```

```
## [1] 0.7064271
```

Based on these simulations, it's reasonable to assume that the correlation between baseline and 3-month measurements in both groups is ~0.7. Then the distributions of the differences in each treatment group are

$$
\begin{aligned}
d &\sim N(E(\mu_{\text{base}} - \mu_{\text{3mo}}), Var(\mu_{\text{base}} - \mu_{\text{3mo}})) \\
&\sim N(E(\mu_{\text{base}}) + E(\mu_{\text{3mo}}), Var(\mu_{\text{base}}) + Var(\mu_{\text{3mo}}) - 2Cor(\mu_{\text{base}}, \mu_{\text{3mo}})sd(\mu_{\text{base}})sd(\mu_{\text{3mo}})) \\
&\sim N(0, \sigma^2 + \sigma^2 - (2)(0.7)(\sigma^2))) \\
&\sim N(0, 1.6\sigma^2)
\end{aligned}
$$

Even though are assumed to know the population variance in LDL, it may be most appropriate to conduct a t-test. This may be especially appropriate for small sample sizes (n<30 in each group), where we cannot assume the observed data are normally distributed. Again, using simulations to find the sample sizes that corresponds to a type I error of at most 0.05, and power of at least 0.9:

```
alpha = 0.05
pow.target = 0.9

## Function to find sample size that meets target alpha and power
ss_calc = function(target_alpha, target_pow, null, alt, var) {
  i = 2
  t1 = 0.5
  pow = 0.5
  while (pow < target_pow || t1 > target_alpha) { # do I need to make sure that type I error is also be

    ## Under null hypothesis
    h.null = sapply(1:n.sim, function(j) {
      placebo = rnorm(i, null, sqrt(0.6*var))
      exp.null = rnorm(i, null, sqrt(0.6*var))

      return(t.test(exp.null, placebo, alternative = "greater", var.equal = TRUE)[[1]])
    })

    ## Under alternative hypothesis
    h.alt = sapply(1:n.sim, function(j) {
      placebo = rnorm(i, null, sqrt(0.6*var))
      exp.alt = rnorm(i, alt, sqrt(0.6*var))

      return(t.test(exp.alt, placebo, alternative = "greater", var.equal = TRUE)[[1]])
      #return((mean(exp.alt) - mean(placebo)) / sqrt(((2.6*var))/(2*i)))
    })

    # simulated type I error and power
    t1 = mean(h.null >= qt(1 - target_alpha, df = 2*i - 1)) # test for greater than critical value sinc
    pow = mean(h.alt >= qt(1 - target_alpha, df = 2*i - 1))

    #print(c("n" = i, "type 1" = t1, "power" = pow))
    i = i + 1
  }

  return(i - 1)
}

# simulate
set.seed(1)
t = sapply(1:100, function(i){
  ss_calc(target_alpha = alpha,
          target_pow = pow.target,
          null = d.null,
          alt = d.alt,
          var = var)
})
mean(t) # average number of subjects
```

```
## [1] 5.03
```

Rounding up, a sample size of 6 in each group is appropriate.

Problem 2: (Go/no-go) In a two-stage trial of an experimental treatment with a planned futility interim analysis, 14 patients are first enrolled and treated in the first stage. An additional of 20 patients will be enrolled and treated if there is at least 1 response in the first stage. At the end of the trial, the treatment is deemed promising ("go") when there are at least 4 responses in all enrolled patients.

a. Suppose the true response rate is 0.05. What is the expected value of the sample size?

```
p = 0.05
n1 = 14
e1 = 1
ntotal = 34
e2 = 4
n2 = ntotal - n1

# P(stopping at stage 1 | p = 0.05)
x = pbinom(e1 - 1, n1, p) # 0.48

# Expected sample size
n_expected = ceiling(n1*x + ntotal*(1 - x)) # 24.24
n_expected
```

```
## [1] 25
```

For a true response rate of 5%, the expected sample size of the trial is 25 subjects.

b. Suppose the true response rate is 0.05. Evaluate the probability of a "go" decision.

The probability of a "go" decision is given by

$$\begin{aligned} Pr(go) &= Pr(S_{14} \geq 1, S_{34} \geq 4 | p = 0.05) \\ &= ... = \Sigma_{n=1}^{14}[Pr(S_14 = n | p = 0.05) * \Sigma_{m=n}^{34}[Pr(S_{20} \geq 4 - m | p = 0.05)]]. \end{aligned}$$

```
f = c()
for (i in e1:n1) {
  f[i] = dbinom(i, n1, p) * (1 - pbinom(e2 - i - 1, n2, p))
}
pgo1 = sum(f, na.rm = T)
pgo1
```

```
## [1] 0.0803739
```

For a true response rate of 5%, the probability of a "go" decision is ~0.08.

c. Suppose the true response rate is 0.2. Evaluate the probability of a "go" decision.

```
p = 0.2
f = c()
for (i in e1:n1) {
  f[i] = dbinom(i, n1, p) * (1 - pbinom(e2 - i - 1, n2, p))
```

```
}
pgo2 = sum(f, na.rm = T)
pgo2
```

```
## [1] 0.9041092
```

Using the same formulation for calculating $Pr(go)$ as above, for a true response rate of 20%, the probability of making a "go" decision is ~0.90.

d. Using your result in (b) as type I error rate and (c) as power, evaluate the sample size required for a fixed design with null response 0.05 and alternative response 0.2.

```
p0 = 0.05
p1 = 0.2
alpha = pgo1
power = pgo2

ss = list()
for (i in 1:50) {
  n = i
  s = seq(1, i, 1)

  t1 = 1 - pbinom(s - 1, n, p0)
  pow = 1 - pbinom(s - 1, n, p1)
  index = intersect(which(t1 <= alpha), which(pow >= power))

  if (length(index) == 0) {
    ss[[i]] = NA
  } else {
    ss[[i]] = c(s[index[1]], n)
  }
}

n = ss[which(!is.na(ss))[1]][[1]][2]
s = ss[which(!is.na(ss))[1]][[1]][1]

c(n, s)
```

```
## [1] 32  4
```

Setting type I error to 0.08 and power to 0.90, we find that the sample size required for a fixed design is 32 subjects, for which we reject the null in favor of the alternative when we see a response from at least 4.

Problem 3: (Predictive distribution) Suppose X1 follows an exponential distribution with rate l.

a. Give a conjugate prior for l. Derive its posterior distribution and the corresponding prior predictive distribution of X1.

Let $\lambda \sim Gamma(\alpha, \beta)$ be the conjugate prior for λ, st. $f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$.

The prior predictive distribution of X1 is then:

$$
\begin{aligned}
f(X_1) &= \int f(X_1, \lambda) d\lambda \\
&= \int f(\lambda) f(X_1|\lambda) d\lambda \\
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \lambda e^{-\lambda x_1} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^\alpha e^{-(\beta+x_1)\lambda} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(\beta+x_1)^{\alpha+1}} \int_0^\infty \frac{(\beta+x_1)^{\alpha+1}}{\Gamma(\alpha+1)} \lambda^\alpha e^{-(\beta+x_1)\lambda} d\lambda \\
&= \frac{\beta^\alpha \Gamma(\alpha+1)}{\Gamma(\alpha)(\beta+x_1)^{\alpha+1}}.
\end{aligned}
$$

And the posterior distribution of λ is:

$$
\begin{aligned}
f(\lambda|X_1) &= \frac{f(X_1|\lambda) f(\lambda)}{f(X_1)} \\
&= \frac{\lambda e^{-\lambda x_1} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}}{\frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{(\beta+x_1)^{\alpha+1}}} \\
&= \frac{(\beta+x_1)^{\alpha+1}}{\Gamma(\alpha+1)} \lambda^\alpha e^{-(\beta+x_1)\lambda}
\end{aligned}
$$

$f(\lambda|X_1) \sim Gamma(\alpha+1, \beta+X_1)$.

b. Suppose X1 and X2 are exchangeable. Derive the posterior predictive distribution of X2 given X1.

$$
\begin{aligned}
f(X_2|X_1) &= \int f(\lambda|X_1) f(X_2|\lambda) d\lambda \\
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \lambda e^{-\lambda x_2} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \lambda^\alpha e^{-(\beta+x_1+x_2)\lambda} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{(\beta+x_1+x_2)^{\alpha+2}} \int_0^\infty \frac{(\beta+x_1+x_2)^{\alpha+2}}{\Gamma(\alpha+2)} \lambda^{\alpha+1} e^{-(\beta+x_1+x_2)\lambda} d\lambda \\
&= \frac{(\beta+x_1)^{\alpha+1} \Gamma(\alpha+2)}{\Gamma(\alpha+1)(\beta+x_1+x_2)^{\alpha+2}}.
\end{aligned}
$$

c. Derive the posterior predictive distribution of (X1 + X2)/2 given X1.

$$\begin{aligned}
E[\frac{X_1+X_2}{2}|X_1] &= \frac{1}{2}E(X_1|X_1)+\frac{1}{2}E(X_2|X_1)\\
&= \frac{1}{2\lambda}+\frac{1}{2}E(E(X_2|\lambda,X_1)|X_1)\\
&= \frac{1}{2\lambda}+\frac{1}{2}E(\lambda|X_1)\\
&= \frac{1}{2\lambda}+\frac{\alpha+1}{\beta+X_1}\\
Var[\frac{X_1+X_2}{2}|X_1] &= \frac{1}{4}Var(X_1|X_1)+\frac{1}{4}Var(X_2|X_1)\\
&= \frac{1}{4\lambda^2}+\frac{1}{4}[E(var(X_2|\lambda,X_1)|X_1)+var(E(X_2|\lambda,X_1)|X_1)]\\
&= \frac{1}{4\lambda^2}+\frac{1}{4}[E(\frac{1}{\lambda^2})+Var(\lambda|X_1)]\\
&= \frac{1}{4}[\frac{2}{\lambda^2}+\frac{\alpha+1}{(\beta+X_1)^2}].
\end{aligned}$$

So the posterior predictive distribution of $\frac{X_1+X_2}{2}$ has the above mean and variance.

Problem 4: Suppose the toxicity probability of dose level 1 is 0.25. Calculate the probability that the 3+3 algorithm will declare dose level 1 safe (i.e., below the MTD).

Let S_3 be the number of DLTs observed out of 3 patients, and S_6 be the number of DLTs observed out of 6 patients. A dose is declared "safe" if either 0/3 or 1/6 DLTs are observed.

$$\begin{aligned}
Pr(safe) &= Pr(S_3=0 \cup S_6=1)\\
&= Pr(S_3=0)+Pr(S_6=1)-Pr(S_3=0 \cap S_6=1)\\
&= Pr(S_3=0)+Pr(S_6=1|S_3=1)Pr(S_3=1)+Pr(S_6=1|S_3\neq 1)Pr(S_3\neq 1)\\
&= Pr(S_3=0)+Pr(S_6=1|S_3=1)Pr(S_3=1)
\end{aligned}$$

Here, $X_i \sim Bin(1,p)$ where X designates DLT of not for patients i=1,..,n (n=3,6), and p=0.25. We can calculate this probability in R:

```
p = dbinom(0, 3, 0.25) + dbinom(1, 6,0.25)*dbinom(1, 3, 0.25)
```

Thus with a toxicity probability of 0.25, the 3+3 algorithm will declare dose safe (i.e. at or below the MTD) with probability 0.57.

Problem 5: In a phase I trial that tests 5 dose levels of a new drug, suppose the true dose toxicity curve is {p1 = 0:02; p2 = 0:04; p3 = 0:10; p4 = 0:25; p5 = 0:50} and the target rate is 0.25, that is, the MTD is dose level 4.

a. Using the uniform variates at the end of this document starting with patient 1, run the 3+3 design and estimate the MTD.

```
variate_3plus3 = function(coh.size, dlt, variates, start = NULL) {
```

```
# select variate to start with
if (is.null(start)) {
  start = sample(x = 1:200, size = 1)
} else {
  start = start
}

var = start
dose = 1
esc = 0
mtd = 0

while (esc == 0 && dose <= length(dlt)) {

  dlt.dose_3 = c()
  for (i in 1:coh.size) {
    s = i - 1
    dlt.dose_3[[i]] = qbinom(variates[[var + s]], 1, dlt[[dose]])
  }
  d3 = sum(dlt.dose_3)

  var = var + s + 1

  if (d3 == 0) {
    dose = dose + 1
  } else if (d3 > 1) {
    mtd = dose - 1
    esc = 1
  } else if (d3 == 1) {
    dlt.dose_6 = c()
    for (i in 1:coh.size) {
      s = i - 1
      dlt.dose_6[[i]] = qbinom(variates[[var + s]], 1, dlt[[dose]])
    }
    d6 = sum(dlt.dose_6)
    var = var + s + 1

    dlt_total = d3 + d6

    if (dlt_total == 1 ) {
      dose = dose + 1
    } else {
      mtd = dose - 1
      esc = 1
    }
  }

  if (dose >= length(dlt) && mtd == 0) {
    mtd = length(dlt)
  }

}
return(mtd)
```

```
}
```

```
unif_variates = c(
  0.007, 0.135, 0.772, 0.444, 0.287, 0.347, 0.777, 0.604, 0.025, 0.584,
  0.715, 0.110, 0.770, 0.405, 0.742, 0.923, 0.591, 0.567, 0.952, 0.039,
  0.342, 0.534, 0.342, 0.661, 0.829, 0.489, 0.710, 0.921, 0.055, 0.497,
  0.611, 0.118, 0.122, 0.472, 0.853, 0.931, 0.978, 0.232, 0.519, 0.333,
  0.096, 0.709, 0.985, 0.844, 0.948, 0.361, 0.061, 0.541, 0.815, 0.153,
  0.177, 0.495, 0.735, 0.872, 0.799, 0.028, 0.555, 0.763, 0.752, 0.682,
  0.228, 0.586, 0.732, 0.014, 0.753, 0.412, 0.765, 0.176, 0.919, 0.207,
  0.874, 0.178, 0.820, 0.783, 0.231, 0.541, 0.925, 0.207, 0.408, 0.808,
  0.434, 0.008, 0.382, 0.166, 0.328, 0.294, 0.635, 0.672, 0.669, 0.460,
  0.174, 0.374, 0.381, 0.600, 0.397, 0.091, 0.922, 0.872, 0.754, 0.520,
  0.977, 0.748, 0.955, 0.978, 0.531, 0.196, 0.963, 0.356, 0.061, 0.795,
  0.823, 0.731, 0.284, 0.929, 0.687, 0.858, 0.439, 0.944, 0.676, 0.189,
  0.755, 0.421, 0.357, 0.391, 0.370, 0.028, 0.866, 0.069, 0.818, 0.888,
  0.381, 0.989, 0.663, 0.491, 0.285, 0.000, 0.652, 0.341, 0.316, 0.599,
  0.977, 0.332, 0.985, 0.976, 0.695, 0.730, 0.580, 0.562, 0.674, 0.435,
  0.747, 0.521, 0.024, 0.412, 0.719, 0.819, 0.139, 0.278, 0.270, 0.877,
  0.431, 0.867, 0.723, 0.919, 0.244, 0.362, 0.442, 0.196, 0.409, 0.752,
  0.351, 0.979, 0.189, 0.523, 0.332, 0.690, 0.061, 0.552, 0.253, 0.450,
  0.403, 0.592, 0.381, 0.673, 0.182, 0.862, 0.223, 0.090, 0.729, 0.091,
  0.315, 0.763, 0.373, 0.174, 0.927, 0.264, 0.799, 0.653, 0.569, 0.109,
  0.706, 0.858, 0.357, 0.950, 0.801, 0.123, 0.019, 0.100, 0.329, 0.706,
  0.377, 0.993, 0.256, 0.964, 0.257, 0.778, 0.985, 0.849, 0.047, 0.650,
  0.932, 0.852, 0.750, 0.705, 0.965, 0.626, 0.821, 0.710, 0.988, 0.145,
  0.760, 0.296, 0.210, 0.944, 0.377, 0.880, 0.437, 0.847, 0.694, 0.069,
  0.139, 0.695, 0.444, 0.596, 0.725, 0.543, 0.580, 0.034, 0.899, 0.339
)

tox.rate = c(0.02, 0.04, 0.10, 0.25, 0.50)

variate_3plus3(coh.size = 3, dlt = tox.rate, variates = unif_variates, start = 1)
```

```
## [1] 4
```

We select dose 4 as the MTD.

b. Repeat (a) several times (say 10) and record the distribution of the MTD estimates.

```
set.seed(2)
var_sims = sapply(1:10, function(i){
  variate_3plus3(coh.size = 3, dlt = tox.rate, variates = unif_variates)
})
table(var_sims)
```

```
## var_sims
## 1 2 3 4 5
## 1 1 4 3 1
```

The distribution of MTD selection is variable depending on which uniform variate the simulated trial starts at. In this run, dose 3 is selected as the MTD most often, closely followed by dose 4.

Problem 6: Download R from 'cran.r-project.org' and install the package 'dfcrm'. The following R code will simulate a CRM trial under the dose-toxicity curve (1). Try the code and record the output of the trial.

```
library(dfcrm)
target = 0.1
prior = c(0.05, 0.12, 0.25, 0.40, 0.55)
trueP = c(0.02, 0.04, 0.10, 0.25, 0.50)
N = 20
crmoutput = crmsim(trueP, prior, target, N, 3, model = "logistic")
crmoutput$MTD
```

```
## [1] 2
```

In this simulation, dose 2 is selected as the MTD.

Problem 7: Use 'crmsim' to generate 10 simulated trials, by specifying the argument 'nsim'. Record the distribution of the MTD estimated by the CRM.

```
crm.sim = crmsim(PI = trueP, prior = prior, target = target, n = N, x0 = 3, nsim = 10)
```

```
## simulation number: 1
## simulation number: 2
## simulation number: 3
## simulation number: 4
## simulation number: 5
## simulation number: 6
## simulation number: 7
## simulation number: 8
## simulation number: 9
## simulation number: 10
```

```
crm.sim$MTD
```

```
## [1] 0.0 0.1 0.7 0.1 0.1
```

Dose 3 is selected as MTD 7/10 times.

Problem 8: Let X1,..,Xn be iid survival times with an exponential distribution with rate l1, and another independent sample Y1,..,Ym be iid from exponential with rate l2. Use the prior you choose in a previous question. Evaluate the posterior probability Pr(l1 > l2 | x1,..,xn; y1,..,ym).

$$\begin{aligned}
Pr(\lambda_2 > \lambda_1) &= Pr(\lambda_2 - \lambda_1 > 0) \\
E(\lambda_2 - \lambda_1) &= E(\lambda_2) - E(\lambda_1) \\
&= \frac{1}{\lambda_2} - \frac{1}{\lambda_1} \\
&= \frac{\lambda_1 - \lambda_2}{\lambda_2 \lambda_1} \\
Var(\lambda_2 - \lambda_1) &= Var(\lambda_2) + Var(\lambda_1) \\
&= \frac{1}{\lambda_2^2} - \frac{1}{\lambda_1^2} \\
&= \frac{(\lambda_1 + \lambda_2)(\lambda_1 - \lambda_2)}{\lambda_2^2 \lambda_1^2} \\
Pr(\lambda_2 - \lambda_1 > 0) &= Pr(\frac{\lambda_2 - \lambda_1 - E(\lambda_2 - \lambda_1)}{\sqrt{Var(\lambda_2 - \lambda_1)}} > \frac{-E(\lambda_2 - \lambda_1)}{\sqrt{Var(\lambda_2 - \lambda_1)}}) \\
&= Pr(Z > \frac{\lambda_2 - \lambda_1}{\sqrt{\lambda_1^2 \lambda_2^2}}).
\end{aligned}$$

Problem 9: Let X1,..,Xn be iid standard normal variables with mean mu and variance 1, and assume a standard normal prior on mu.

a. Evaluate the posterior distribution of mu given x1,..,xn.

$$\begin{aligned}
f(\mu|x_1,..,x_n) &\propto f(\mu) f(x_1,..,x_n|\mu) \\
&\propto \frac{1}{sqrt2\pi} e^{-\frac{\mu^2}{2}} [\Pi_{i=1}^n \frac{1}{sqrt2\pi} e^{-\frac{(X_i-\mu)^2}{2}}] \text{since X1,..,Xn are iid} \\
&\propto \frac{1}{sqrt2\pi} e^{-\frac{\mu^2}{2}} [\frac{1}{(sqrt2\pi)^n} e^{-\frac{\Sigma(X_i-\mu)^2}{2}}] \\
&\propto (2\pi)^{-\frac{n_1}{2}} e^{-\frac{\Sigma(X_i-\mu)^2+\mu^2}{2}}
\end{aligned}$$

where the exponent becomes

$$\begin{aligned}
\frac{-1}{2}[\Sigma(X_i^2 - 2\mu X_i + \mu^2) + \mu^2] &= \frac{-1}{2}[\Sigma X_i^2 - 2\mu n\bar{X} + n\mu^2 + \mu^2] \\
&= \frac{-1}{2}[\Sigma X_i^2 - 2\mu n\bar{X} + (n+1)\mu^2] \\
&= \frac{-1}{2}(n+1)[\frac{\Sigma X_i^2}{n+1} - \frac{2\mu n\bar{X}}{n+1} + \mu^2] \\
&= \frac{-1}{2}(n+1)[\mu^2 - \frac{2\mu n\bar{X}}{n+1} + (\frac{n\bar{X}}{n+1})^2 + \frac{\Sigma X_i^2}{n+1} - (\frac{n\bar{X}}{n+1})^2] \\
&= \frac{-1}{2(\frac{1}{n+1})}[(\mu - \frac{n\bar{X}}{n+1})^2]
\end{aligned}$$

Therefore, $f(\mu|x_1,..,x_n) \propto N(\frac{n\bar{X}}{n+1}, \frac{1}{n+1})$.

b. Suppose another independent sample Y1,..,Ym arises from normal with mean v and variance 1. Assuming a standard normal prior on v, evaluate the posterior probability Pr(mu > v | x1,..,xn; y1,..,ym).

Note that $f(\mu|x_1,..,x_n) \propto N(\frac{n\bar{X}}{n+1}, \frac{1}{n+1})$ and $f(\mu|y_1,..,y_m) \propto N(\frac{m\bar{Y}}{m+1}, \frac{1}{m+1})$.

$$\begin{aligned}
Pr(\mu > v) &= Pr(\mu - v > 0) \\
E(\mu - v) &= E(\mu) - E(v) \\
&= \frac{n\bar{X}}{n+1} - \frac{m\bar{Y}}{m+1} \\
Var(\mu - v) &= Var(\mu) + Var(v) \\
&= \frac{1}{n+1} + \frac{1}{m+1} \\
Pr(\mu - v > 0) &= Pr(\frac{\mu - v - E(\mu - v)}{\sqrt{Var(\mu - v)}} > \frac{-E(\mu 0 v)}{\sqrt{Var(\mu - v)}}) \\
&= Pr(Z > \frac{\frac{m\bar{Y}}{m+1} - \frac{n\bar{X}}{n+1}}{\sqrt{\frac{1}{n+1} + \frac{1}{m+1}}}), \text{ where } Z \sim N(0,1)
\end{aligned}$$

Problem 10: Consider simple linear regression model Yi = a + b*xi + ei, where ei's are iid normal noise with variance o2, for i = 1,..,n.

a. Derive the maximum likelihood estimate of a, b, and o2.

Because of the assumptions needed for simple linear regression, we know $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$. Then

$$\begin{aligned}
L(Y) &= \Pi_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{[\frac{-(y_i - (\alpha+\beta x_i))^2}{2\sigma^2}]} \\
&= (2\pi\sigma^2)^{-n/2} e^{[\frac{-\Sigma(y_i - (\alpha+\beta x_i))^2}{2\sigma^2}]} \\
l(Y) &= log(L(Y)) = -nlog(\sigma) + \frac{n}{2}log(2\pi) - \frac{1}{2\sigma^2}\Sigma(y_i - (\alpha + \beta x_i))^2
\end{aligned}$$

MLE of α:

$$\begin{aligned}
\frac{\partial}{\partial\alpha} l(Y) &= 0 \\
...\frac{\partial}{\partial\alpha}\Sigma(y_i - (\alpha + \beta x_i))^2 &= 0 \\
...\text{expand}... - 2\bar{Y} + 2\alpha + 2\beta\bar{X} &= 0 \\
\Rightarrow \hat{\alpha} &= \bar{Y} - \beta\bar{X}
\end{aligned}$$

MLE of β:

$$\frac{\partial}{\partial\beta}l(Y) = 0$$
$$...\frac{\partial}{\partial\beta}\Sigma(y_i - (\alpha + \beta x_i))^2 = 0$$
$$...\frac{\partial}{\partial\beta}\Sigma y_i^2 - 2\alpha\Sigma y_i - 2\beta\Sigma x_i y_i + n\alpha^2 + 2\alpha\beta\Sigma x_i + \beta^2\Sigma x_i^2 = 0$$
$$\Rightarrow \hat{\beta} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$$

MLE of σ^2:

$$\frac{\partial}{\partial\sigma^2}l(Y) = 0$$
$$...\frac{\partial}{\partial\sigma^2} - nlog(\sqrt{\sigma^2}) + \frac{n}{2}log(2\pi) - \frac{1}{2\sigma^2}\Sigma(y_i - (\alpha + \beta x_i))^2 = 0$$
$$...\frac{\partial}{\partial u} - nlog(u) + \frac{n}{2}log(2\pi) - \frac{1}{2u}\Sigma(y_i - (\alpha + \beta x_i))^2 = 0$$
$$...\frac{-n}{u} + \frac{1}{u^2}\Sigma(y_i - (\alpha + \beta x_i))^2 = 0$$
$$\Rightarrow \hat{\sigma^2} = \frac{1}{n}\Sigma(y_i - (\alpha + \beta x_i))^2$$

b. Suppose o2 is known. Assume an improper non-informative prior on (a; b), i.e., having f(a; b) / 1. Derive the posterior distribution of (a; b) given the data.

$$\begin{aligned} f(\alpha, \beta|Y) &\propto f(\alpha, \beta)f(Y|\alpha, \beta) \\ &\propto (1)f(Y|\alpha, \beta) \\ &\propto (2\pi\sigma^2)^{-n/2}e^{[\frac{-\Sigma(y_i-(\alpha+\beta x_i))^2}{2\sigma^2}]} \end{aligned}$$

c. Suppose (a; b) are known. What is a conjugate prior for o2? Derive its posterior distribution. [Hint: Consider o2^(-2)]

Let $f(\sigma^2) \propto 1/\sigma^2$ be the conjugate prior for σ^2. Then the posterior given the data is

$$\begin{aligned} f(\alpha, \beta|Y) &\propto f(\alpha, \beta)f(Y|\alpha, \beta) \\ &\propto \frac{1}{\sigma^2}(2\pi\sigma^2)^{-n/2}e^{[\frac{-\Sigma(y_i-(\alpha+\beta x_i))^2}{2\sigma^2}]} \\ &\propto \frac{1}{(2\pi)^{n/2}(\sigma^2)^{\frac{n+2}{2}}}e^{[\frac{-\Sigma(y_i-(\alpha+\beta x_i))^2}{2\sigma^2}]} \end{aligned}$$

d. Derive the marginal posterior distribution of o2 when (a; b) are unknown and have a non-informative prior.

To find the marginal posterior distribution of σ^2 when (α, β) are unknown, we need to integrate the joint posterior density over all possible values of (α, β). Using the results in (b) and (c):

$$
\begin{aligned}
f(\sigma^2|Y) &= \int f(\alpha, \beta, \sigma^2|Y) d(\alpha, \beta) \\
&= \int f(\sigma^2|\alpha, \beta, Y) f(\alpha, \beta|\sigma^2, Y) d(\alpha, \beta) \\
&= \int_0^\infty \int_0^\infty \frac{1}{(2\pi)^n (\sigma^2)^{n+2}} e^{[\frac{-\Sigma(y_i - (\alpha + \beta x_i))^2}{\sigma^2}]} d\alpha d\beta
\end{aligned}
$$

Problem 11: Consider a study comparing treatments A and B. Suppose that in truth the response rate of treatment A is 0.2, and treatment B 0.8.

a. Suppose we use Zelen's play-the-winner (PTW) rule with 1:1 randomization ratio for the first patient. Evaluate the distribution of the number of subjects allocated to treatment B among the first four patients. How does the distribution compare to to a balanced design?

```
ptw = function(p_a, p_b, r_a, r_b, n) {

  # 1:1 randomization of first patient
  alloc = rbinom(1, 1, p_a)
  pts = c()
  outcome = c()

  for (i in 1:n) {

    # If allocated to treatment A
    if (alloc == 1) {
      o = rbinom(1, 1, r_a)
      pts[[i]] = 0

      # If success, stay on trt A, otherwise switch to trt B
      if (o == 1) {
        alloc = alloc
      } else {
        alloc = 0
      }

    # If allocated to trt B
    } else {
      o = rbinom(1, 1, r_b)
      pts[[i]] = 1

      # If success, stay on trt B, otherwise switch to trt A
      if (o == 1) {
        alloc = alloc
      } else {
        alloc = 1
      }

    }
    outcome[[i]] = o
  }
```

```
  return(pts)
}

response.a = 0.2
response.b = 0.8
n.sim = 1e4

set.seed(1)
ptw_sims = lapply(1:n.sim, function(i){
  ptw(0.5, 0.5, response.a, response.b, 4)
})

prop.B.ptw = mean(sapply(ptw_sims, mean))
prop.B.ptw
```

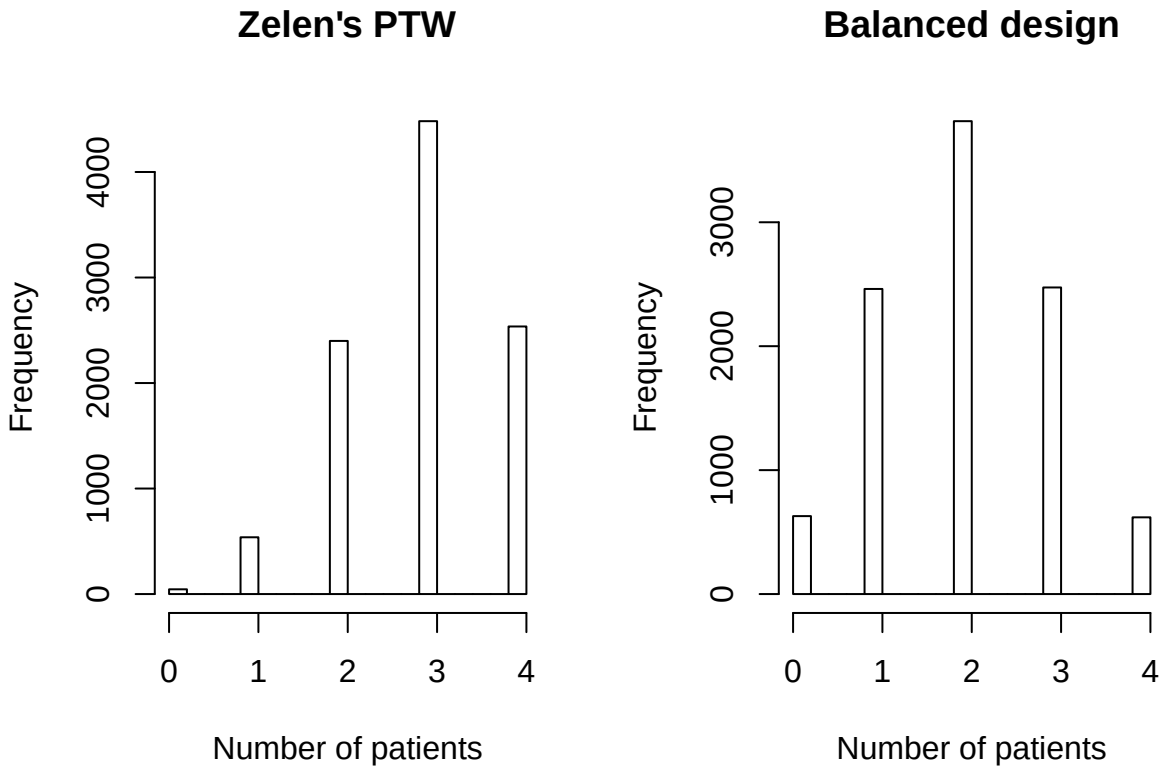
```
## [1] 0.72315
```

```
####

balanced_design = function(r_a, r_b, n){

  p_a = 0.5
  pts = c()
  outcome = c()

  for (i in 1:n) {

    # 1:1 randomization of first patient
    alloc = rbinom(1, 1, p_a)

    # If allocated to treatment A
    if (alloc == 1) {
      o = rbinom(1, 1, r_a)
      pts[[i]] = 0

    # If allocated to trt B
    } else {
      o = rbinom(1, 1, r_b)
      pts[[i]] = 1

    }
    outcome[[i]] = o
  }

  return(pts)

}

set.seed(1)
balanced_sims = lapply(1:n.sim, function(i){
  balanced_design(response.a, response.b, 4)
})

prop.B.balanced = mean(sapply(balanced_sims, mean))
```

```
prop.B.balanced
```

```
## [1] 0.4998
```

```
par(mfrow = c(1, 2))
hist(sapply(ptw_sims, sum), main = "Zelen's PTW", xlab = "Number of patients")
hist(sapply(balanced_sims, sum), main = "Balanced design", xlab = "Number of patients")
```



Using Zelen's play the winner rule, the first 4 patients are allocated to treatment B about at a rate of 0.72, on average. The distribution of patients to treatment B is shown in the histogram above, where we give 3 out of 4 patients trt B. Meanwhile, the balanced design (1:1 randomization for all patients) allocates 2 patients to trt B on average (rate of 0.5).

b. Repeat (a) under the scenario where both treatments have the same response rate at 0.3.

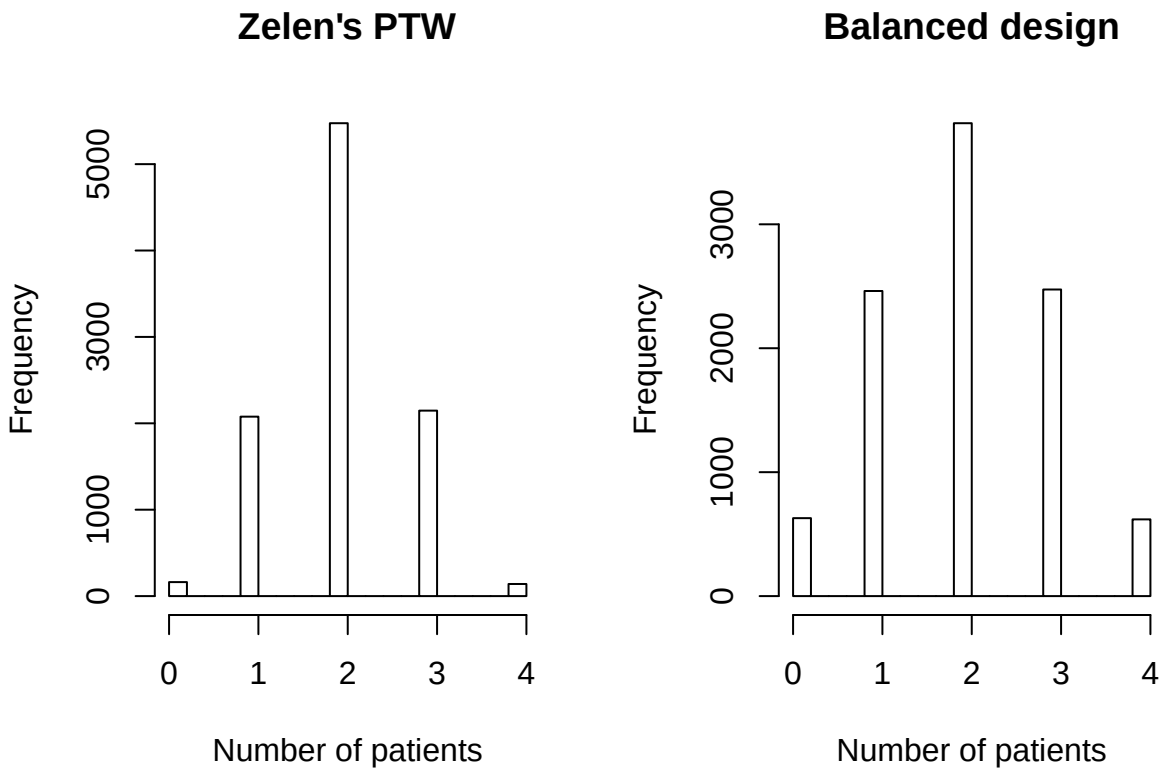
```
response.a = 0.3
response.b = 0.3
n.sim = 1e4

set.seed(1)
ptw_sims = lapply(1:n.sim, function(i){
  ptw(0.5, 0.5, response.a, response.b, 4)
})

prop.B.ptw = mean(sapply(ptw_sims, mean))
prop.B.ptw
```

```
## [1] 0.50065
```

```
####

set.seed(1)
balanced_sims = lapply(1:n.sim, function(i){
  balanced_design(response.a, response.b, 4)
})

prop.B.balanced = mean(sapply(balanced_sims, mean))
prop.B.balanced
```

```
## [1] 0.4998
```

```
par(mfrow = c(1, 2))
hist(sapply(ptw_sims, sum), main = "Zelen's PTW", xlab = "Number of patients")
hist(sapply(balanced_sims, sum), main = "Balanced design", xlab = "Number of patients")
```



When both treatments have the same response rate, the first 4 patients are allocated to treatment B at a rate of about 0.5, on average. This allocation rate is about identical to the balanced design, however the actual distribution of number of patients varies between designs (see histograms).

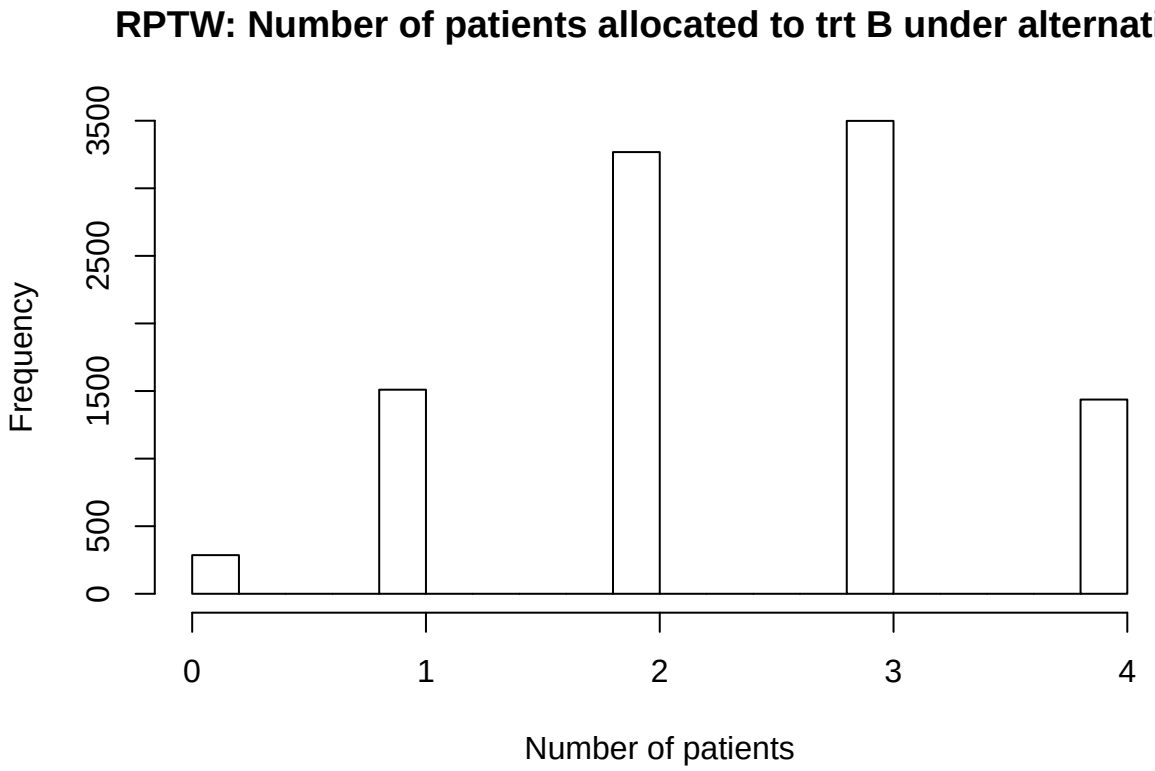
Problem 12: Repeat the previous question with the randomized PTW with an initial urn of one 'A' and one 'B' ball. Recall that the RPTW adds a ball of the same type following a success, and a ball of the opposite type following a failure.

```
rptw = function(r_a, r_b, n) {

  na = nb = 1
  pts = c()
```

```
  outcome = c()

  for (i in 1:n) {

    p_a = na/(na + nb)
    p_b = nb/(na + nb)

    # Randomize pt
    alloc = rbinom(1, 1, p_a)

    # If allocated to treatment A
    if (alloc == 1) {
      o = rbinom(1, 1, r_a)
      pts[[i]] = 0

      # If success, stay on trt A, otherwise switch to trt B
      if (o == 1) {
        na = na + 1
      } else {
        nb = nb + 1
      }

    # If allocated to trt B
    } else {
      o = rbinom(1, 1, r_b)
      pts[[i]] = 1

      # If success, stay on trt B, otherwise switch to trt A
      if (o == 1) {
        nb = nb + 1
      } else {
        na = na + 1
      }

    }
    outcome[[i]] = o
  }

  return(pts)
}

response.a = 0.2
response.b = 0.8
n.sim = 1e4

set.seed(1)
rptw_sims = lapply(1:n.sim, function(i){
  rptw(response.a, response.b, 4)
})

prop.B = mean(sapply(rptw_sims, mean))
prop.B
```

```
## [1] 0.607275
```

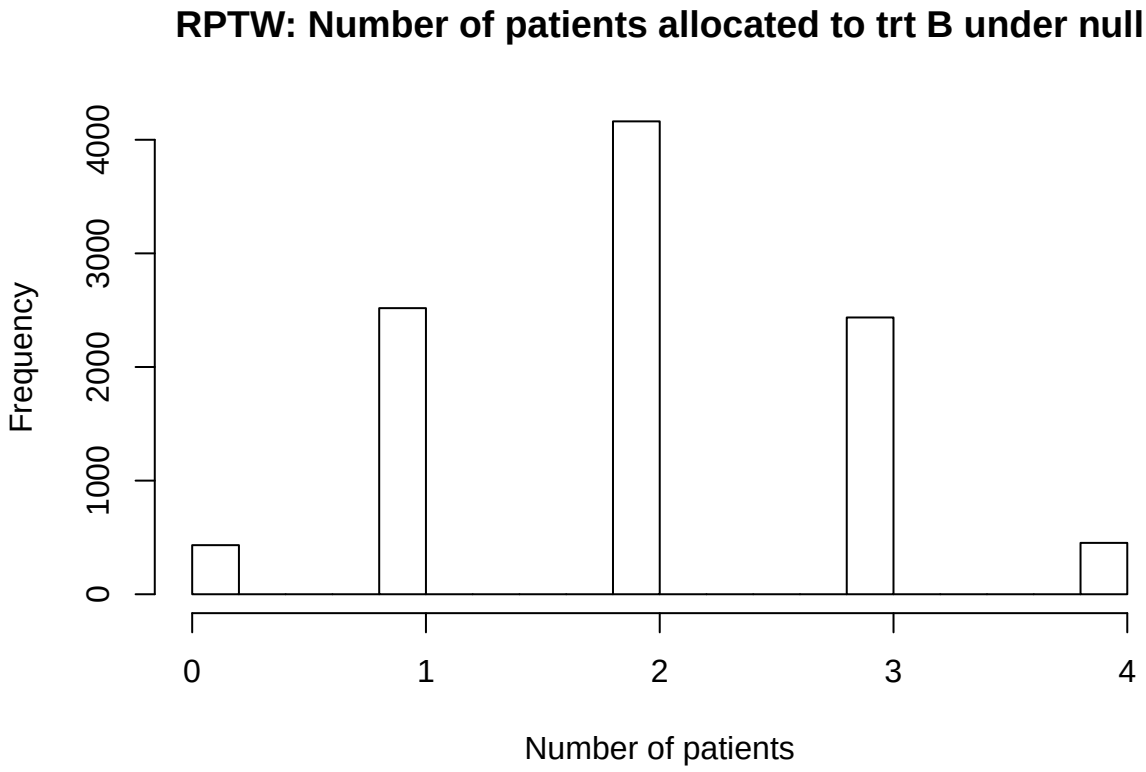
```
hist(sapply(rptw_sims, sum), main = "RPTW: Number of patients allocated to trt B under alternative", xla
```




```
response.a = 0.3
response.b = 0.3
n.sim = 1e4

set.seed(1)
rptw_sims = lapply(1:n.sim, function(i){
  rptw(response.a, response.b, 4)
})

prop.B = mean(sapply(rptw_sims, mean))
prop.B
```

```
## [1] 0.49895
```

```
hist(sapply(rptw_sims, sum), main = "RPTW: Number of patients allocated to trt B under null", xlab = "Nu
```



Problem 13: Let X1, X2 be independent normal with respective means mu.t1 and mu.t2 and variance 1, and define Z = w1X1 + w2X2 where w1^2 + w2^2 = 1 and the weights are chosen a priori. Let f{abs(Z) >= z0.025} be the rejection region of a test against the null hypothesis mu = 0.

a. What is the type I error rate of the test?

Under the null,

$$\begin{aligned} E(Z) &= E(w_1X_1 + w_2X_2) = w_1E(X_1) + w_2E(X_2) = 0 \\ Var(Z) &= Var(w_1X_1 + w_2X_2) \\ &= w_1^2Var(X_1) + w_2^2Var(X_2) \qquad\qquad Z \sim N(0,1) \\ &= w_1^2 + w_2^2 \\ &= 1 \end{aligned}$$

Therefore, $Z \sim N(0,1) \Rightarrow Pr[|Z| \geq z_{0.025}] = 0.05$.

b. What is the power of the test?

Under the alternative,

$$
\begin{aligned}
E(Z) &= E(w_1X_1 + w_2X_2) \\
&= w_1E(X_1) + w_2E(X_2) \\
&= \mu(w_1t_1 + w_2t_2) \\
Var(Z) &= Var(w_1X_1 + w_2X_2) \qquad Z \sim N(\mu(w_1t_1 + w_2t_2), 1). \\
&= w_1^2Var(X_1) + w_2^2Var(X_2) \\
&= w_1^2 + w_2^2 \\
&= 1
\end{aligned}
$$

$$
\begin{aligned}
Power &= Pr[|Z| \geq z_{0.025}] \\
&= 2Pr(|Z| \geq z_{0.025}), \text{ since the normal dist is symmetric} \\
&= 2Pr(Z - \mu(w_1t_1 + w_2t_2) \geq z_{0.025} - \mu(w_1t_1 + w_2t_2)) \\
&= 2Pr(U \geq z_{0.025} - \mu(w_1t_1 + w_2t_2)), \text{ where } U \sim N(0,1) \\
&= 2\Phi(z_{0.025} - \mu(w_1t_1 + w_2t_2)).
\end{aligned}
$$

c. Determine the optimal weights w1 and w2 that maximize the power.

$$
\begin{aligned}
Power &= 2\Phi(z_{0.025} - \mu(w_1t_1 + w_2t_2)) \\
&= 2\Phi(z_{0.025} - \mu(w_1t_1 + (1 - w_1^2)t_2)) \\
\frac{\partial}{\partial w_1}power &= 2(-\mu t_1 + \frac{w_1}{\sqrt{1 - w_1^2}}\mu t_2)\phi(z_0.025 - c) \\
...\frac{t_1}{t_2} &= \frac{w_1}{\sqrt{1 - w_1^2}} \\
... \Rightarrow w_1 &= \frac{t_1/t_2}{\sqrt{1 + (\frac{t_1}{t_2})^2}} \\
w_1^2 + w_2^2 = 1 \Rightarrow w_2 &= \sqrt{1 - \frac{(t_1/t_2)^2}{1 + (t_1/t_2)^2}}
\end{aligned}
$$

So we can see that the selection of optimal weights that maximize the power is dependent on the ratio of t_1 to t_2, or the ratio of the means of X_1, X_2.

d. Prove the inequality under the null.

We can show this by looking at each side of the inequality, starting with the left side.

$$
\begin{aligned}
Pr(X_1 \geq z_{0.005} \cup Z \geq z_{0.02}) &\leq 0.025 \\
Pr(X_1 \geq z_{0.005}) + Pr(\geq z_{0.02}) - Pr(X_1 \geq z_{0.005} \cap Z \geq z_{0.02}) &\leq 0.025 \\
0.005 + 0.02 - Pr(X_1 \geq z_{0.005}, Z \geq z_{0.02}) &\leq 0.025 \\
Pr(X_1 \geq z_{0.005}, Z \geq z_{0.02}) &\geq 0.
\end{aligned}
$$

And now the right side.

$$
\begin{aligned}
Pr(X_1 \geq z_{0.005} \cup Z \geq z_{0.025}) &\geq 0.025 \\
Pr(X_1 \geq z_{0.005}) + Pr(\geq z_{0.025}) - Pr(X_1 \geq z_{0.005} \cap Z \geq z_{0.025}) &\geq 0.025 \\
0.005 + 0.025 - Pr(X_1 \geq z_{0.005}, Z \geq z_{0.02}) &\geq 0.025 \\
Pr(X_1 \geq z_{0.005}, Z \geq z_{0.02}) &\leq 0.05.
\end{aligned}
$$

QED

.