1. A group of researchers is running an experiment over the course of 30 months, with a single observation collected at the end of each month. Let $X_1, ..., X_{30}$ denote the observations for each month. From prior studies, the researchers know that

$$X_i \sim f_X(x),$$

but the mean μ_X is unknown, and they wish to conduct the following test

$$H_0 : \mu_X = 0$$
$$H_a : \mu_X > 0.$$

At month k, they have accumulated data $X_1, ..., X_k$ and they have the t-statistic

$$T_k = \frac{\bar{X} - 0}{S_k/\sqrt{n}}.$$

The initial plan was to test the hypotheses after all data was collected (at the end of month 30), at level $\alpha = 0.05$. However, conducting the experiment is expensive, so the researchers want to "peek" at the data at the end of month 20 to see if they can stop it early. That is, the researchers propose to check whether t_{20} provides statistically discernible support for the alternative. If it does, they will stop the experiment early and report support for the researcher's alternative hypothesis. If it does not, they will continue to month 30 and test whether t_{30} provides statistically discernible support for the alternative.

(a) What values of t_{20} provide statistically discernible support for the alternative hypothesis?
Because we are doing a one-sided t-test with a significance level of 0.05, we need to compare T_{20} to the critical value with $n - 1 = 20 - 1 = 19$ degrees of freedom, denoted as $t_{0.95,19}$. This is the value in which 95% of the t-distribution will lie to the left of it, and 5% will lie to the right of it.

```r
(crit.value.20 = qt(0.95, df=19))

## [1] 1.729133
```

So we got that $t_{0.95,19} = 1.729$, and because the alternative hypothesis is that the mean is greater than 0, we want the T_{20} values that are greater than 1.729. So there is statistically discernible support for the alternative hypothesis if $T_{20} > 1.729$.

(b) What values of t_{30} provide statistically discernible support for the alternative hypothesis?
Here we are going to do the same as before, but now with $n-1 = 30-1 = 29$ degrees of freedom, denoted as $t_{0.95,29}$. This again is the value in which 95% of the t-distribution will lie to the left of it, and 5% will lie to the right of it.

```r
(crit.value.30 = qt(0.95, df=29))

## [1] 1.699127
```

So we got that $t_{0.95,29} = 1.699$, and because the alternative hypothesis is that the mean is greater than 0, we want the T_{30} values that are greater than 1.699. So there is statistically discernible support for the alternative hypothesis if $T_{30} > 1.699$.

(c) Suppose $f_X(x)$ is a Laplace distribution with $a = 0$ and $b = 4.0$. Conduct a simulation study to assess the Type I error rate of this approach.
Note: You can use the `rlaplace()` function from the `VGAM` package for `R` (Yee, 2010).
To simulate the study, I am doing 10000 simulations, where each simulation creates 30 observations of the Laplace(0, 4) distribution. First, I calculate the test statistic for the first 20 observations, and reject if it meets the criteria from part a. If not, then I calculate the test statistic for all 30 observations, and reject if it meets the criteria from part b.

```r
library(VGAM)
set.seed(7272)
simulations = 10000 #number of simulations
n = 30 #total sample size
n.peek = 20 #sample size for peeking
rejections = 0 #counter for number of rejections (rejecting the null)
for(i in 1:simulations){
  x = rlaplace(n, location = 0, scale = 4) #30 observations of Laplace(0, 4) distribution
  x.peek = x[1:20] #first 20 observations
  t.peek = mean(x.peek)/(sd(x.peek)/sqrt(n.peek)) #t statistic after peeking
  if(t.peek > crit.value.20){
    rejections = rejections + 1 #reject if T20 is greater than the critical value with df=19
  }
  else {
    t.final = mean(x)/(sd(x)/sqrt(n)) #t statistic for all 30 observations
    if(t.final > crit.value.30){
      rejections = rejections + 1 #reject if T30 is greater than the critical value with df=29
    }
  }
}
#every rejection is a type 1 error because we are assuming the null is true
(type1.error.rate = rejections / simulations)

## [1] 0.0763
```

The Type I error rate is represented by the number of rejections divided by the number of simulations. The reason that each rejection is an occurrence of a Type I error is because we are under the assumption that the null hypothesis is true ($\mu_X = 0$). So based on the seed that I set, I got a Type I error rate of 0.0763, which is greater than our significance level of 0.05. This means that this study increases the chance of getting a Type I error. This is due to the fact that we are checking for rejection twice. If we only checked after 30 observations, then the Type I error rate should be around 0.05, but the probability of rejection goes up when you are testing for it twice per simulation.

(d) **Optional Challenge:** Can you find a value of $\alpha < 0.05$ that yields a Type I error rate of 0.05? Yes, if you change α to 0.033, which changes the critical values to $t_{0.967,19} = 1.951$, and $t_{0.967,29} = 1.911$. Because both critical values have increased, it will be less likely for the test statistic to be greater than the critical value, which then decreases the Type I error rate. This significance level of 0.03 yields a Type I error rate of exactly 0.05 given the seed that I set (7272).

2. Perform a simulation study to assess the robustness of the T test. Specifically, generate samples of size $n = 15$ from the Beta(10,2), Beta(2,10), and Beta(10,10) distributions and conduct the following hypothesis tests against the actual mean for each case (e.g., $\frac{10}{10+2}$, $\frac{2}{10+2}$, and $\frac{10}{10+10}$).

   (a) What proportion of the time do we make an error of Type I for a left-tailed test?

   (b) What proportion of the time do we make an error of Type I for a right-tailed test?

   (c) What proportion of the time do we make an error of Type I for a two-tailed test?

   (d) How does skewness of the underlying population distribution effect Type I error across the test types?

References

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.