Power Analysis and Sample Size Determination

Yin Chang

Department of Mathematical Science

Montana State University

April 28, 2007

A writing project submitted in partial fulfillment
of the requirements for the degree

Master of Sciences in Statistics

APPROVAL

of a writing project submitted by Yin Chang

This writing project has been read by the writing project director and have been found to be satisfactory regarding content, English usage, format , citations, bibligraphic style, and consistency, and is ready for submission to the Statistics faculty.

5/18/2007
Date

Jim Robison-Cox
Writing Project Director

1 Introduction

Determining the appropriate sample size for an investigation, such as a clinical trial, is an essential step in the statistical design of the project. An adequate sample size helps ensure the study will yield reliable information. The motivation for writing this project is that a study with an inadequate sample size is not only futile, it also will cause unnecessary economic loss. An under-sized study can be a waste of resources for not having the capability to produce useful results, while an over-sized one uses more resources than are necessary. Furthermore, if an experiment involves human or animal subjects, sample size is a pivotal issue for ethical reasons.

In this writing project, I will first introduce a few terms that are often used in statistics: Type I error, Type II error, power. Under the frequentist approach, some examples will be given on how to calculate sample size and power for continuous response and binary response. For binary response, there are two branches, one branch in Section 4.1 is for large sample size, another branch, Section 4.2, is for small sample size, where Fisher's exact test is used. Also, two R programs for computing power and sample size are provided.

The later section offers another point of view, a Bayesian approach for sample size determination, in which I compare the difference between frequentist approach and Bayesian approach for sample size determination.

2 Basic Definitions

In a clinical trial of a new drug, the proportion of survival in the group using the ordinary drug, π_1, and the proportion of survival in the group using the new drug, π_2, are compared. Second questions may arise: Is the new drug better than what's now available to treat a specific disease? If it's not better, is it at least as good, perhaps while causing fewer side effects? A question for a statistician could be, if the proportions of survival p_1 and p_2 from two samples, differ, before we assume that a real difference exists, would it be possible that those two samples are actually from the same population and by chance produced the observed difference? Normally, we do not know the true population parameters such as the true proportions π_1 and π_2. Then, we will have to make a decision based on the observed statistical results.

2.1 Hypothesis Testing Procedures

I. Assume the null hypothesis (i.e., H_0: $\theta_1 = \theta_2$, implying no true difference, where θ is the true population parameter and it can be a true mean or a true proportion).

II. Determine whether a two-tailed test or a one-tailed test will be made. If the alternative hypothesis (H_a) specifies the direction of the difference (e.g., H_a: $\theta_1 > \theta_2$), it is a one-tailed test. If the alternative hypothesis specifies only a non-directional difference or inequality (e.g., H_a: $\theta_1 \neq \theta_2$), it is a two-tailed test.

III. Choose α, the arbitrary level of significance (usually 0.05 or 0.01).

IV. Calculate the appropriate test statistic, e.g., z or χ^2 statistic.

V. Determine the probability of the observed value or of a more extreme value if the null hypothesis is true, which is what we call the p-value.

VI. If $p-\text{value} < \alpha$, we reject the null hypothesis and accept the alternative hypothesis that a true difference exists. If $p-\text{value} > \alpha$, we fail to reject the null hypothesis.

2.2 Type I Error

A type I error occurs when one rejects the null hypothesis (H_0) when the null hypothesis is actually true. The probability of a type I error is the level of significance of the test of hypothesis, which is denoted by α. In the medical research setting, Type I error is called a false positive .

P(type I error) = significance level=α.

Back to the example mentioned above, a clinical trial of a new drug, the null hypothesis might be the new drug is no better, on average, than the current drug, that is H_0: $\pi_1 = \pi_2$. A type I error would occur if we concluded that the two drugs produced different effects, when in fact there was no difference between them.

Using significance level α is based on the assumption that the null hypothesis is true. Most commonly the value of α is taken to be $\alpha = 0.05$ or in some cases $\alpha = 0.01$ is used. It is highly unusual to have values of $\alpha > 0.05$. If $\alpha = 0.05$, then in the long run 5 in 100 times the null hypothesis will be rejected when it is true. Similarly, if $\alpha = 0.01$, only one in every 100 times a type I error would occur.

2.3 Type II Error

A type II error occurs when one fails to reject the null hypothesis (H_0) when the alternative hypothesis (H_a) is true. Again, in the medical research setting, a type II error is called false negative .The probability of a type II error, denoted by β, depends on : (1) the true difference, for example, ($\theta_1 - \theta_2$), (2) the sample size and population variance, and (3) the level chosen for α (the smaller α is , the larger β is).

P(type II error)=β.

Using the same example in section 2.2, a type II error would occur if it was concluded that the two drugs produced the same effect, $\pi_1 = \pi_2$, that is, there is no difference between the two drugs on average, when in fact they produced different ones.

Intuitively, it is clear that β should not be large, but there is no general agreement on the widespread use of a fixed β value. It is common to desire $\beta \leqslant 0.2$. A type I error is often considered to be more serious, and therefore more important to avoid, than a type II error.

2.4 Power

Power is the probability of rejecting the null hypothesis when the alternative hypothesis is true.

Power=1-P(type II error)=$1-\beta$.

Power is always computed for a particular value of the alternative value. Power is important because it indicates the chance of finding a significant difference when there really is one. A study with low power is likely to produce no significant results even when meaningful differences do indeed exist. Low power to detect important differences usually results from a situation in which the study was designed with too small a sample size. Studies with low power are a waste of resources since they do not adequately address the scientific question. Ideally we want a test to have high power, close to 1.

3 Sample size determination for continuous response

Continuous responses can be found in many clinical studies. Suppose that a drug is designed to lower cholesterol levels in the blood, the outcome, cholesterol levels (Y), in this study is continuous response, because cholesterol level is a continuous variable. In this section, we are going to talk about symmetric continuous response, more specifically, continuous response with normal distribution.

Here we specify the above example in statistical terms: the model we want to consider involves two normal populations with the following parameters:

	Sample Size	Variance	Mean
Average cholesterol level in control group ($\bar{Y}_1$)	n_1	σ_1^2	μ_1
Average cholesterol level in treatment group ($\bar{Y}_2$)	n_2	σ_2^2	μ_2

which says $\bar{Y}_1 \sim N(\mu_1, \sigma_1^2/n_1)$ and $\bar{Y}_2 \sim N(\mu_2, \sigma_2^2/n_2)$, and samples of size n_1 and n_2 are taken, respectively. Thus, $\bar{Y}_1 - \bar{Y}_2 \sim N(\mu_1 - \mu_2,\ \sigma_1^2/n_1 + \sigma_2^2/n_2)$ by independence of sampling is called the sampling distribution of the difference between means, which is the statistic of interest. The null hypothesis, H_0: $\mu_1 = \mu_2$, and the alternative hypothesis, H_a: $\mu_1 \neq \mu_2$.

3.1 To computer the power, $1-\beta$, given sample size n_1, n_2

Power is $1-\beta$, which is showed in the graph below:

To determine β , we must specify

1. σ^2, this value is ordinarily based upon previous study results ;
2. The probability, α, of a Type I error;
3. The magnitude of the difference $\delta = \mu_1 - \mu_2$ to be detected; and
4. The sample size n_1, n_2.

If $\mu_1 - \mu_2 = \delta > 0$, look at figure 3.1, the curve on the left is the curve of sample differences under the null hypothesis, H_0: $\mu_1 = \mu_2$, which is a normal curve $N(0, \sigma_1^2/n_1 + \sigma_2^2/n_2)$. The one on the right represents the distribution of differences $\mu_1 - \mu_2 = \delta$ between our samples, which is also a normal curve $N(\delta, \sigma_1^2/n_1 + \sigma_2^2/n_2)$, where we assume $\delta > 0$ for now. The area representing the value of $\alpha/2$ on the graphing is drawn in the right shaded area. The area representing β is drawn in the left stippled area.

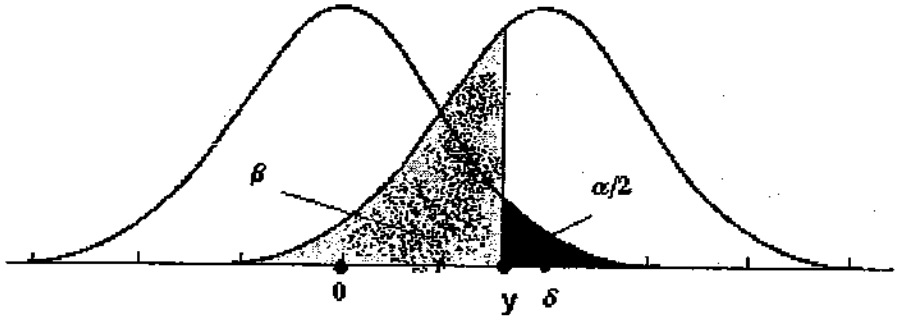

Figure 3.1

where $y = z_{1-\alpha/2} * \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$, since $\frac{(\overline{Y_1} - \overline{Y_2}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0,\ 1)$ under H_a: $\mu_1 \neq \mu_2$.

Suppose that $U \sim N(\delta, \sigma_1^2/n_1 + \sigma_2^2/n_2)$. Consider a two-sided test under H_a: $\mu_1 \neq \mu_2$.

$$\beta = P\left(U \leqslant z_{1-\alpha/2} * \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\ \right) \Rightarrow$$

$$\beta = P\left(Z \leqslant \frac{z_{1-\alpha/2} * \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} - \delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right) = P\left(Z \leqslant z_{1-\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\ \right) \tag{3.1}$$

$$1 - \beta = 1 - P\left(Z \leqslant z_{1-\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\ \right) \tag{3.2}$$

where $z_{1-\alpha/2}$ is the $100^*(1-\alpha/2)$ percentile. Similarly, the $100\,(1-\beta/2)$ percentile.

If $\mu_1 - \mu_2 = \delta < 0$ as showed in figure 3.2 below, then the calculation would be slightly different.

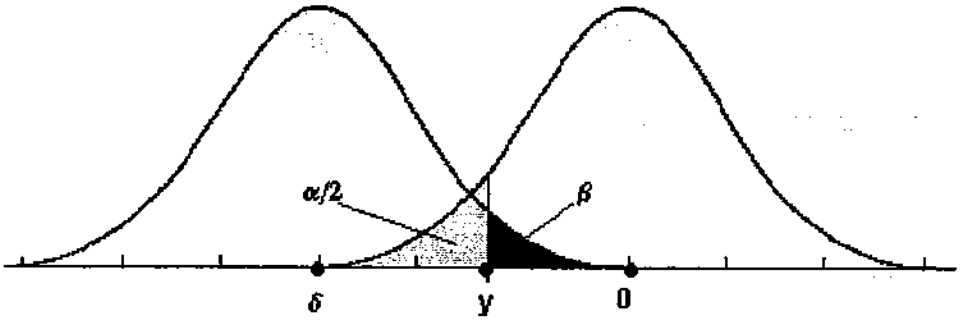

Figure 3.2

where $y = z_{\alpha/2} * \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

$$\beta = P\left(U \geqslant z_{\alpha/2} * \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right) \Rightarrow$$

$$\beta = P\left(Z \geqslant \frac{z_{\alpha/2} * \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} - \delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right) = P\left(Z \geqslant z_{\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right) \tag{3.3}$$

$$1 - \beta = 1 - P\left(Z \geqslant z_{\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right) \tag{3.4}$$

However, whether $\mu_1 - \mu_2 = \delta > 0$ or $\mu_1 - \mu_2 = \delta < 0$, as long as $|\delta|$'s are identical and everything else is kept the same for those two cases, the power would end up with the same value under the null hypothesis, H_0: $\mu_1 = \mu_2$, and the alternative hypothesis, H_a: $\mu_1 \neq \mu_2$.

One-sided test is referring to the alternative hypothesis H_a: $\mu_1 > \mu_2$ or H_a: $\mu_1 < \mu_2$, instead of H_a: $\mu_1 \neq \mu_2$. The computation for the power is very similar, the only difference is that $z_{\alpha/2}$ is replaced by z_α. The reason should be intuitive with the help of Figure 3.3, which is for the two-sided test under H_a: $\mu_1 \neq \mu_2$, and Figure 3.4, which is for the one-sided test under H_a: $\mu_1 > \mu_2$.

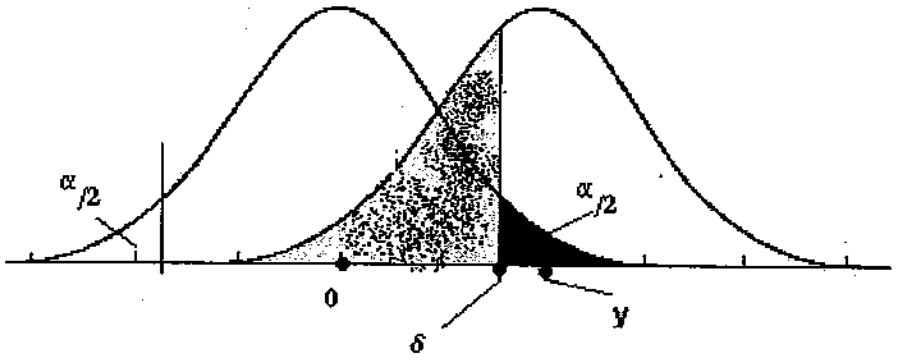

Figure 3.3

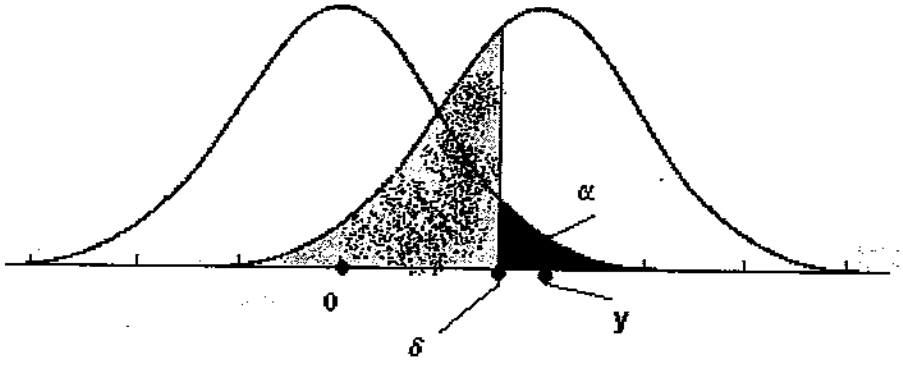

Figure 3.4

The power computed under one-sided test is:

$$1-\beta=1-P\left(Z\geqslant z_{\alpha}-\frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\right) \tag{3.5}$$

3.2 Factors effecting size of power $(1-\beta)$

A) Changes in significance level: e.g. According to formula 3.2, if α decreases, then $z_{1-\alpha/2}$ increases, β increases, power decreases. By looking at Figure 3.5, if α decreases, then β error is bigger. So, if everything else is kept constant, then the smaller α is, the smaller power is.

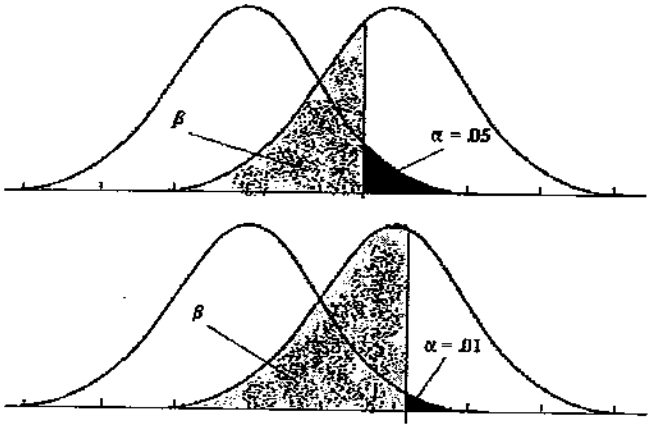

Figure 3.5

B) Changes in sample size n: e.g. When sample sizes (n_1 and n_2) increase, $\frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ increases, $P\left(Z\leqslant z_{1-\alpha/2}-\frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\right)$ decreases, by formula 3.2, power increases. In Figure 3.6, when sample sizes increases, spread decreases. Furthermore, their overlap is smaller, β is smaller. So, if everything else is kept constant, the larger sample size n is, the larger power is.

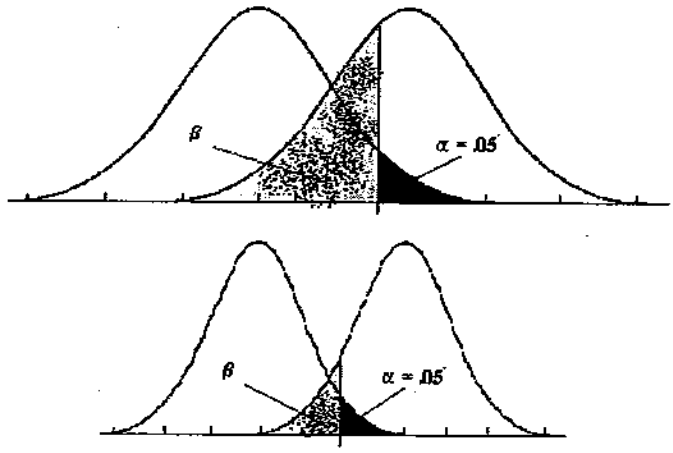

Figure 3.6

C) Changes in variance: e.g. When the variances (σ_1^2 and σ_2^2) are smaller, $\frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ increases, $P\left(Z \leqslant z_{1-\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\right)$ decreases, by formula 3.2, power increases. We still use Figure 3.6, when the variances decreases, curves have less spread. Furthermore, their overlap is smaller, β is smaller. So, if everything else is kept constant, the smaller variances are, the larger power is.

D) Changes in size of difference: e.g. If $\mu_1 - \mu_2 = \delta$ increases, $\frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ increases, $P\left(Z \leqslant z_{1-\alpha/2} - \frac{\delta}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\right)$ decreases, by formula 3.2, power increases. In Figure 3.7, curves are further apart compared to original curves and β is smaller. So, if everything else is kept constant, power $(1-\beta)$ is larger when the difference δ is larger.

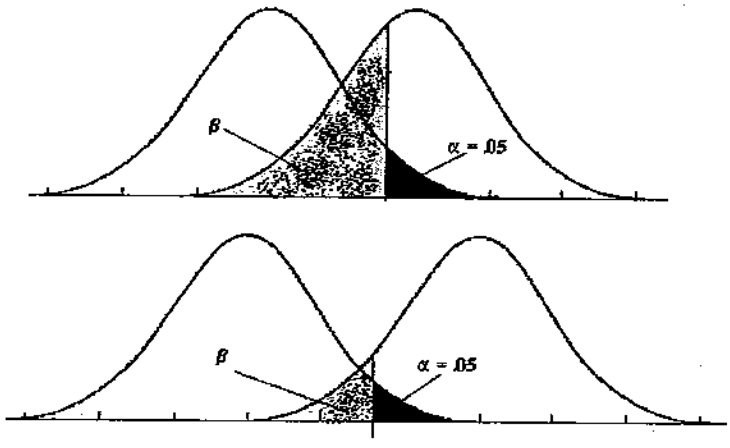

Figure 3.7

E) One-tailed versus two-tailed tests: power is greater in one-tailed tests than in comparable two-tailed test.

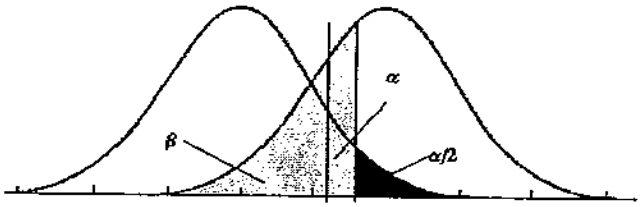

Figure 3.8

3.3 Compute sample size n given power

3.3.1 Computing sample size n for known σ^2

To determine the sample size, we must specify

1. σ^2, typically based on previous study results;
2. The probability, α, of a Type I error;
3. The magnitude of the difference $\delta = \mu_1 - \mu_2$ to be detected; and
4. The power, $1-\beta$, or equivalently the probability of a Type II error, β.

Suppose that both population in studies are sampled the same number of times, $n = n_1 = n_2$. For a two-sided test, to compute sample size $\ddot{n}$, we just need to convert Formula 3.1 in section 3.2,

$$\beta = P\left(Z \leqslant \frac{z_{1-\alpha/2} * \sqrt{\sigma_1^2/\ddot{n} + \sigma_2^2/\ddot{n}} - \delta}{\sqrt{\sigma_1^2/\ddot{n} + \sigma_2^2/\ddot{n}}}\right) \tag{3.6}$$

$$\Rightarrow z_\beta * \sqrt{\sigma_1^2/\ddot{n} + \sigma_2^2/\ddot{n}} = z_{1-\alpha/2} * \sqrt{\sigma_1^2/\ddot{n} + \sigma_2^2/\ddot{n}} - \delta$$

The required sample size *per group* is :

$$\Rightarrow \ddot{n} = \frac{(\sigma_1^2 + \sigma_2^2)(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta^2} \tag{3.7}$$

However, to get the above sample size $\ddot{n}$ result, we have to assume that σ_1^2 and σ_2^2 are known.

3.3.2 Compute sample size n, when σ^2 is unknown

If we go on to a more realistic situation in which the population standard deviation (σ_1 and σ_2) are not known, and we estimate the sample size from the sample standard deviation (s_1 and s_2).

We assume that the unknown population variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$, which makes it possible to use t-Distribution. Also, we assume $n_1 = n_2 = \ddot{n} = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta^2}$, the pooled sample variance s_p^2 is an unbiased estimator of σ^2,

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{s_1^2+s_2^2}{2} \tag{3.8}$$

The two-sample t statistic:

$$t = \frac{(\overline{Y_1}-\overline{Y_2})-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{(\overline{Y_1}-\overline{Y_2})-(\mu_1-\mu_2)}{s_p\sqrt{\frac{2}{\ddot{n}}}} \tag{3.9}$$

Referring back to Figure 3.1, the curve on the left is the curve of sample difference under the null hypothesis, $H_0 : \mu_1 - \mu_2 = 0$, which is the Student's t-Distribution curve $t_{(2\ddot{n}-2)}$. The one on the right represents the distribution under the alternative hypothesis, $H_a : \mu_1 - \mu_2 = \delta > 0$, which is the non-central Student's t-Distribution curve, $t_{(2\ddot{n}-2,\lambda)}$, where $\lambda = \frac{\delta^2}{\frac{2s_p^2}{\ddot{n}}}$.

Suppose that $T_1 \sim t_{(2\ddot{n}-2)}$, $T_2 \sim t_{(2\ddot{n}-2,\lambda)}$.

Consider a two-sided test here, in Figure 3.1, $y = t_{(1-\alpha/2,2\ddot{n}-2)} * s_p\sqrt{1/n_1+1/n_2}$.

$$\beta = P\left(T_2 \leqslant \frac{t_{(1-\alpha/2,2\ddot{n}-2)} * s_p\sqrt{1/n_1+1/n_2} - \delta}{s_p\sqrt{1/n_1+1/n_2}}\right)$$

The power is:

$$\begin{aligned} 1-\beta &= 1 - P\left(T_2 \leqslant \frac{t_{(1-\alpha/2,2\ddot{n}-2)} * s_p\sqrt{1/n_1+1/n_2} - \delta}{s_p\sqrt{1/n_1+1/n_2}}\right) \\ &= 1 - P\left(T_2 \leqslant \frac{t_{(1-\alpha/2,2\ddot{n}-2)} * s_p\sqrt{2/\ddot{n}} - \delta}{s_p\sqrt{2/\ddot{n}}}\right) \end{aligned} \tag{3.10}$$

You can find $t_{(1-\alpha/2,2\ddot{n}-2)}$ by using R code: qt(1-α/2, 2$\ddot{n}$-2). Note that $T_2 \sim t_{(2\ddot{n}-2,\lambda)}$, since $\lambda = \frac{\delta^2}{\frac{2s_p^2}{\ddot{n}}}$ is related to sample size and the desired sample is unknown yet. To solve this problem, based on Formula 3.7, we can use an approximate sample size $n_1 = n_2 = \ddot{n} = \frac{2s_p^2(z_{1-\alpha/2}+z_{1-\beta})^2}{\delta^2}$ as the starting approximate sample size and plug this approximate sample $\ddot{n}$ back into Formula 3.10 for λ, s_p and $t_{(1-\alpha/2,2\ddot{n}-2)}$. We can compute the power with the help of R code for noncentral t-Distribution, pt($\frac{t_{(1-\alpha/2,2\ddot{n}-2)} * s_p\sqrt{2/\ddot{n}} - \delta}{s_p\sqrt{2/\ddot{n}}}$, 2$\ddot{n}$-2, λ).

If the computed power is lower than the given power, we keep on increasing the approximate sample size for 1 unit each time until the computed power is higher than the given power. For the case that the computed power is higher than the given power, we keep on reducing the approximate sample size for 1 unit each time until the computed power is lower than the given power, then the desired sample size would be that stopping sample size add 1.

3.4 Factors increasing sample size

Based on Sec. 3.2, by comparison, sample size changes as a function of:

1. Power $(1-\beta)$: sample size increases as the power increases.

2. Variation in outcome (σ^2): sample size increases as variation in outcome increases.

3. Difference (effect) to be detected δ : sample size increases as this difference decreases.

4. Significance level α: sample size increases as the significance level decreases.

5. One-tailed versus two-tailed tests: sample size is smaller in one-tailed tests than in two-tailed tests.

4 Sample sizes determination for binary response

When a response takes on one and only one of two possibilities, it is called a binary response. In many clinical trials, a patient will have of two possible outcomes, such as dies/survives, the antigen presence/absence, tumor detected/non-detected. One wishes to compare proportions as estimated by samples from two populations to see if the true population parameters might be equal, where proportions could be the percentage of rats developing tumors under diets involving different doses of a food additive; or the percentage of patients experiencing pain relief by taking a drug and placebo trial.

4.1 The Normal Curve as approximation to the binomial distribution for large samples

Suppose that the first binomial variable Y_1, which represents the control group, is of size n_1 with survival rate π_1, which is estimated by the sample proportion p_1. The second binomial variable Y_2, which represents the treatment group, is of size n_2, with survival rate π_2 , which is estimated by the sample proportion p_2. It is of our interest to compare the survival rates by taking the difference $p_1 - p_2$ so that we could get some information about whether the treatment have a certain effect on the survival rate.

The mean and variance of the difference $p_1 - p_2$ are given by

$$E(p_1 - p_2) = \pi_1 - \pi_2, \tag{4.1}$$

$$\text{var}(p_1 - p_2) = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}. \tag{4.2}$$

By Central Limit Theorem, for large n_1 and n_2, p_1 and p_2 are approximately normally distributed, which implies the difference $(p_1 - p_2)$ would also be normally distributed, thus under the assumption that n_1 and n_2 are large enough, (the common rule of thumb, $n_1\pi_1$, $n_1(1-\pi_1)$, $n_2\pi_2$ and $n_2(1-\pi_2)$ all greater than 5),

$$\frac{p_1 - p_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} = z \tag{4.3}$$

is approximately $N(0,1)$.

4.1.1 Computer the power

There are four quantities needed to compute the power:

1. π_1, the proportion in the first population.
2. π_2, the proportion in the second population.
3. n_1, n_2 the number of observations to be obtained from each of the two populations.
4. The significance level α at which the statistical test will be made.

Based on Formula 3.6, for one-sided hypothesis test, where H_o: $\pi_1 = \pi_2$ and H_a: $\pi_1 > \pi_2$, the power $(1-\beta)$ is :

$$1-\beta = 1 - P(Z \leqslant \frac{z_{1-\alpha}\sqrt{\pi_1(1-\pi_1)/n_1 + \pi_2(1-\pi_2)/n_2} - \delta}{\sqrt{\pi_1(1-\pi_1)/n_1 + \pi_2(1-\pi_2)/n_2}}) \tag{4.4}$$

where $\delta = \pi_1 - \pi_2 > 0$.

4.1.2 Finding sample sizes needed for testing the difference between proportions

It is common to suppose that both population in studies are sampled the same number of times, $n = n_1 = n_2$. To determine the sample size n, we often specify π_1, π_2, α and β.

Recall in section 3.3, Formula 3.6 : $n = \frac{2\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta^2}$.

Using the same approach, for one-sided hypothesis test, an approximation for n here is :

$$\ddot{n} = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\pi_1 - \pi_2}\right)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2)) \tag{4.5}$$

For two-sided test, we just need to change $z_{1-\alpha}$ in Formula 4.2 into $z_{1-\alpha/2}$; the rest stays the same.

$$\ddot{n} = \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{\pi_1 - \pi_2}\right)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2)) \tag{4.6}$$

Example 4.1 : Suppose the investigator wishes to do a study on the drugs for hypertension. The design considered is that hypertensive patients are assigned to either treatment and placebo randomly. We compare the proportions (π_1 and π_2) of patients who are still hypertensive in each group by taking the difference of the two proportions $(\pi_1 - \pi_2)$. Suppose that we use a two-tailed test with a 0.05 significance level, where H_o: $\pi_1 = \pi_2$ and H_a: $\pi_1 > \pi_2$. How many subjects would be required to have 90% power of detecting a difference in the proportions $(\pi_1 - \pi_2)$ of 0.04?

Suppose that from other studies we know that an estimate of the proportion of patients in the placebo group still hypertensive (π_1) is 0.9. Thus, we have that $\pi_2 = 0.86$ because $\pi_1 - \pi_2 = 0.04$.

The required sample size in each of the two groups is:

$$\ddot{n} = \frac{(1.96+1.282)^2(0.9\times 0.1+0.86\times 0.15)}{(0.04)^2} = 1438.6 \approx 1439$$

Thus, more than 2878 hypertensive patients would need to be enrolled in this parallel groups study.

If the power was reduced to 80%, given everything else fixed, then

$$\ddot{n} = \frac{(1.96+0.841)^2(0.9\times 0.1+0.86\times 0.15)}{(0.04)^2} = 1066.5 \approx 1067$$

It is easy to notice that when the power decreases, the required sample size decreases.

Now, we keep the power to be 80% still, but we are detecting a difference (δ) of 0.25, where $\pi_1 = 0.9$ and $\pi_2 = 0.65$, then the corresponding sample size is:

$$\ddot{n} = \frac{(1.96+0.841)^2(0.9\times 0.1+0.65\times 0.35)}{(0.25)^2} = 39.9 \approx 40$$

We need a sample size of approximately 80 subjects in total to detect a very large effect (a difference of 0.25) with 80% power, which is very much smaller than the sample size (2134) to detect a small effect (a difference of 0.04).

4.2 Fisher's exact test for small samples

In section 4.1, we use central limit theorem to approximate the distribution of $p_1 - p_2$ as normal distribution, given the sample size is large enough. While in situations where a large sample approximation is inappropriate, we may prefer to use Fishers Exact test to return exact one-tailed and two-tailed p-values for a given frequency table. Then based on the exact p-value, we can determine whether the proportions of those falling into each category differ by group.

Example 4.2 : In clinical trials, survival/absence of a certain disease could be considered as a success, while death/presence of a certain disease could be considered as a failure. Assume that the control group has n_{11} successes in $n_{11} + n_{12}$ trials; treatment group has n_{21} successes in $n_{21} + n_{22}$ trials. Let $n_{1.} = n_{11} + n_{12}$. Similarly define $n_{2.}$, $n_{.1}$, and $n_{.2}$. Let $n_{..} = n_{11} + n_{12} + n_{21} + n_{22}$. Writing the table with row and column totals gives:

	Success	Failure	
Control group	n_{11}	n_{12}	$n_{1.}$
Treatment group	n_{21}	n_{22}	$n_{2.}$
	$n_{.1}$	$n_{.2}$	$n_{..}$

Let X represent the total number of success, X_1 represent the number of successes in control group, and X_2 represent the number of successes in treatment group, where $X \sim \text{Bin}(n_{..}, \pi)$, $X_1 \sim \text{Bin}(n_{1.}, \pi_1)$ and $X_2 \sim \text{Bin}(n_{2.}, \pi_2)$. Further, Fisher's exact test assumes that both row and column totals are fixed, which means $n_{1.}, n_{2.}, n_{.1}$, and $n_{.2}$ are fixed, but *not* n_{11}, n_{12}, n_{21}, and n_{22}. Let the null hypothesis H_o: $\pi_1 = \pi_2$ and the alternative hypothesis H_o: $\pi_1 > \pi_2$.

The probability distribution of n_{11}:

$$P[n_{11} = k | \pi = \pi_1 = \pi_2] = \frac{\binom{n_{1.}}{k}\pi_1^k(1-\pi_1)^{n_{1.}-k}\binom{n_{2.}}{n_{.1}-k}\pi_2^{n_{.1}-k}(1-\pi_2)^{n_{2.}-n_{.1}+k}}{\binom{n_{..}}{n_{.1}}\pi^{n_{.1}}(1-\pi)^{n_{..}-n_{.1}}}$$

$$\Rightarrow P[n_{11} = k | \pi_1 = \pi_2] = \binom{n_{1.}}{k}\binom{n_{2.}}{n_{.1}-k}/\binom{n_{..}}{n_{.1}} \tag{4.7}$$

for $k \leqslant n_{1.}$, $k \leqslant n_{.1}$ and $n_{.1} - k \leqslant n_{2.}$.

The probability distribution of n_{11} is called the *hypergeometric distribution.*

A *p-value* represents the probability of obtaining values of the test statistic that are equal or greater than the observed test statistic. The exact *p-value* for this test is:

$$p - \text{value} = \sum_{i=k}^{n_{.1}} \frac{\binom{n_{1.}}{i}\binom{n_{2.}}{n_{.1}-i}}{\binom{n_{..}}{n_{.1}}} \tag{4.8}$$

where k is the observed value. If $n_{1.} < n_{.1}$, then $\sum_{i=k}^{n_{.1}}$ should be changed into $\sum_{i=k}^{n_{1.}}$.

A *p-value* close to zero suggests that we reject the null hypothesis H_o: $\pi_1 = \pi_2$, and there is enough evidence that the difference $(\pi_1 = \pi_2)$ exists. While a *p-value* close to 1 would suggest that we fail to reject the null hypothesis H_o: $\pi_1 = \pi_2$.

Most statistical software programs will compute Fishers Exact test. There are also several web pages that will compute this test. Here's an example: http://www.psych.ku.edu/preacher/fisher/fisher.html

4.2.1 Computing power by simulation

When sample size is small, one way to compute the power is to find out the critical value c under given α level and the hypergeometric distribution, then also, based on the distribution under H_a: $\pi_1 \neq \pi_2$, we can could compute the exact power with the critical value c. For more information, see: Suissa and Shuster (1985).

Another way to compute the power: we can use simulation to compute the power according to different values of π_1 (the probability of success in control group), π_2 (the probability of success in treatment group), n_1 (the fixed total number of control group), n_2 (the fixed total number of treatment group), α (the significance level) and N (the total number of simulations creating tables for Fisher's test).

The brief explanation for Program 1 (fisher.power) is the following:

First, we randomly generate the number of successes a/b in control group/treatment group based on that $a \sim \text{binomial}(n_1, \pi_1)$ and $b \sim \text{binomial}(n_2, \pi_2)$. Then we create a 2×2 table of those numbers based on fixed margins and use Fisher's test to obtain p-value. The above simulation is repeated up to N times, where you could define how large N is when you apply Program 1. Provided that $\pi_1 \neq \pi_2$, we count how many times H_o: $\pi_1 = \pi_2$ is rejected when H_o: $\pi_1 = \pi_2$ is false, which is denoted by k in the R code. Thus, power can be calculated by $\frac{k}{N}$.

In Figure 4.1, we suppose that both populations are sampled the same number of times, $n = n_1 = n_2$, and $N = 500$.

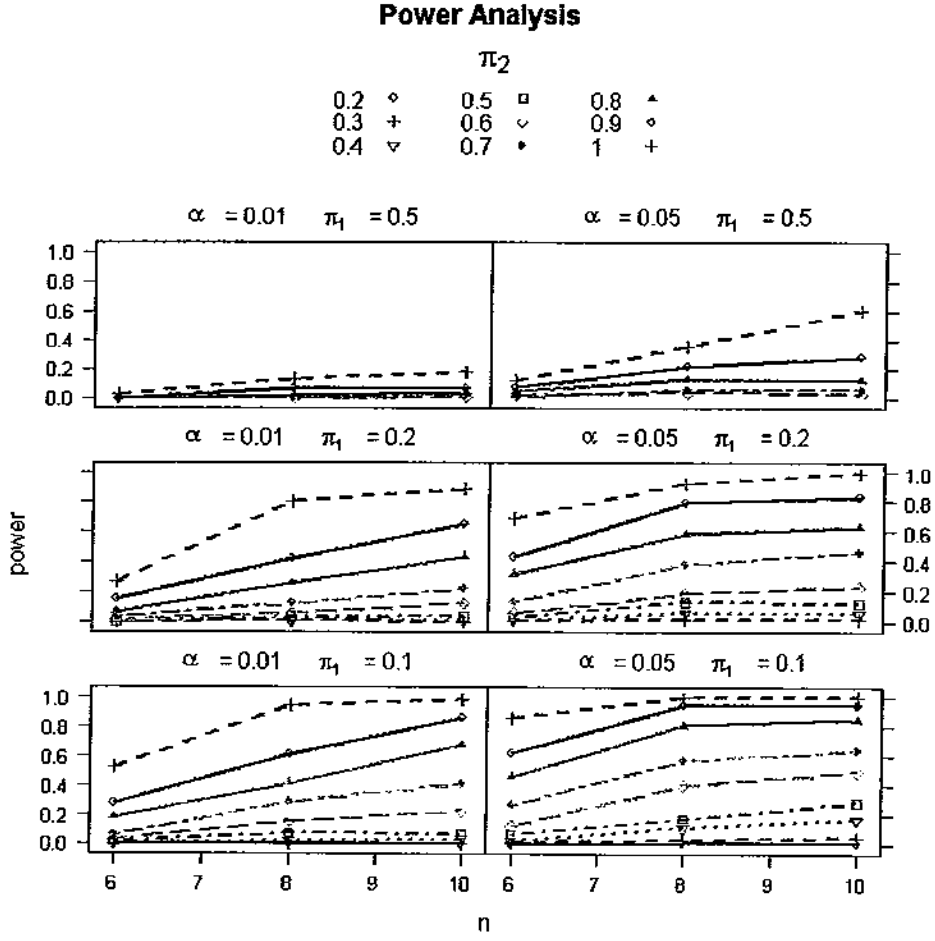

Figure 4.1

Based on Figure 4.1, we could conclude that:

1. Given everything else fixed, when α increases, power increases.

2. Given everything else fixed, when the difference $\delta = \pi_1 - \pi_2$ increases, power increases.

3. Given everything else fixed, when the sample size n increases, power increases.

4.2.2 Finding sample sizes for given power

Just like computing power by simulation, we could also compute sample size by simulation according to different values of power, π_1 (the probability of success in control group), π_2 (the probability of success in treatment group), α (the significance level) and N (the total number of simulations creating tables for Fisher's test). Here, we need to assume the unknown sample size to be small and for both populations the sample size is the same.

The brief explanation for Program 2 (fisher.ssd) is the following:

Based on the given power, α, π_1, π_2, although the sample size is small that it is not appropriate to use the Central Limit Theorem to compute the sample size directly, we could still use normal distribution to get an approximate sample size n by using the formula 4.2:

$$\tilde{n} = \left(\frac{z_{1-\alpha} + z_{1-\beta}}{\pi_1 - \pi_2}\right)^2 (\pi_1(1-\pi_1) + \pi_2(1-\pi_2)),$$

After obtaining the approximate $\tilde{n}$, we use Program 1 (fisher.power) to keep on computing power until we achieve the power that is equal to or just above the given power. If you reduce the sample size for 1 unit, then the power would be smaller than the given power. In Figure 4.2, we assume that power is 0.9, $\alpha = 0.05$, π_1 and π_2 vary.

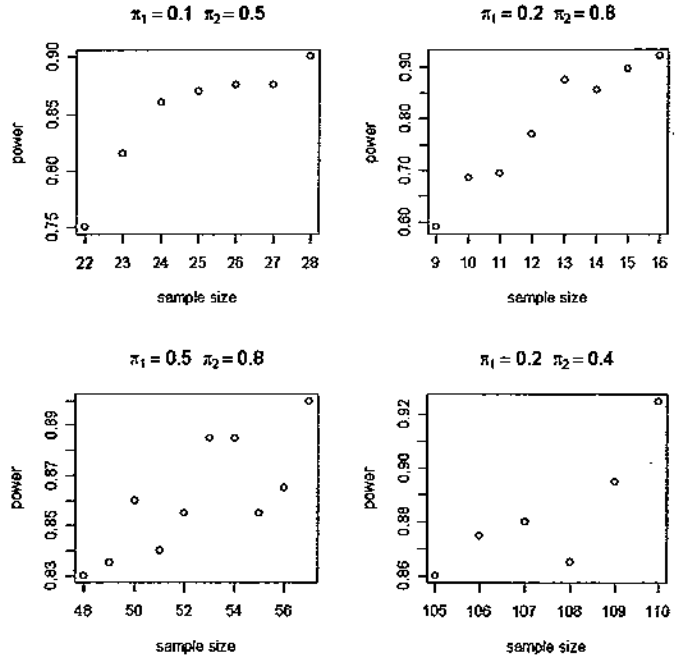

Figure 4.2

In Figure 4.3, we assume that power is 0.6, $\alpha = 0.05$, π_1 and π_2 vary.

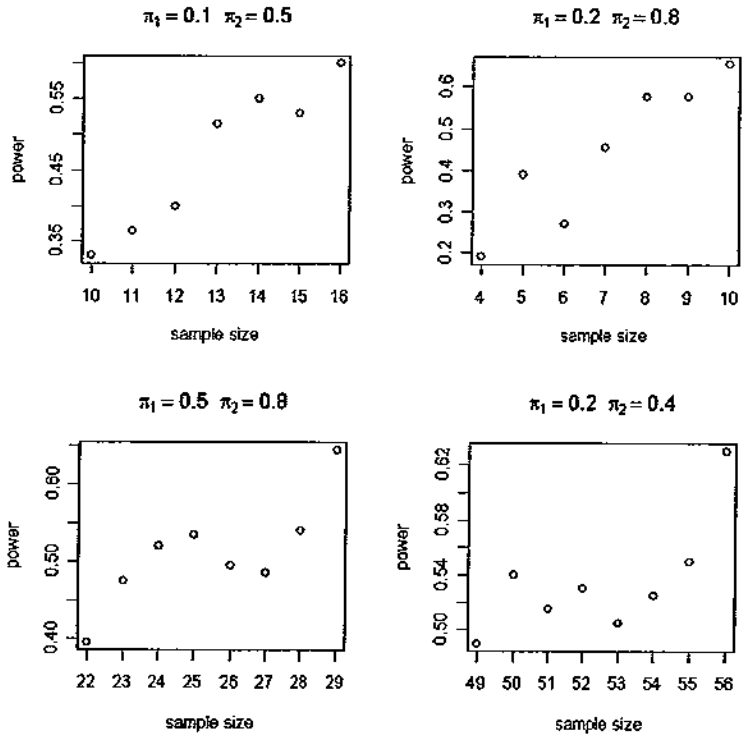

Figure 4.3

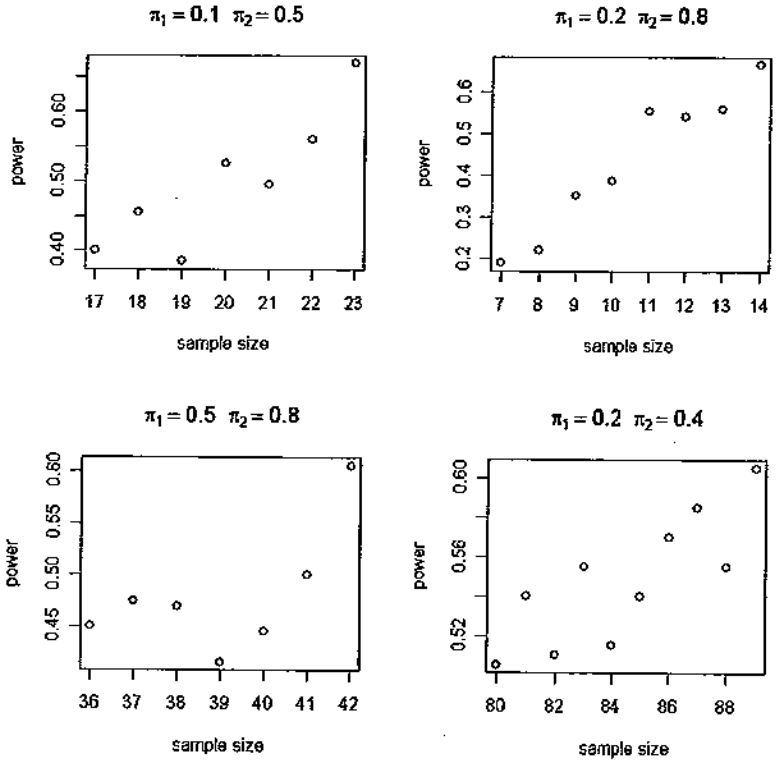

Figure 4.4

In Figure 4.4, we assume that power is 0.6, $\alpha = 0.01$, π_1 and π_2 vary.
We could conclude that:

1. Given everything else fixed, when power increases, sample size increases.
2. Given everything else fixed, when the difference $\delta = \pi_1 - \pi_2$ increases, sample size decreases.
3. Given everything else fixed, when α increases, sample size decreases.

5 Bayesian sample size determination

In the previous sections, for a frequentist, it is the goal to seek the smallest sample size that is sufficient to achieve a desired power at a specified significance level. However, a full Bayesian approach to sample size determination (SSD) has a different goal———to minimize the expected posterior loss. For hypothesis testing, the loss function is 0-1 loss function, in which the loss is 0 only if the hypotheses are both correctly classified, i.e., we fail to reject the null hypothesis when the null hypothesis is true, or we reject the null hypothesis when the alternative hypothesis is true, otherwise, we set the loss to be 1. Since 0-1 loss function assigns the equal loss on both types of error, Bayesian SSD is aiming for minimizing the sum of Type I error and Type II error.

5.1 Fitting and sampling priors

In the frequentist approach to SSD problems, it would be of interest to investigate the power of the SSD procedure when the true parameter assumes some particular values. But, for a Bayesian, it would not be considered to be satisfactory.

All Bayesian model fitting exercises need a prior distribution for the unknown parameters in the model. This is the prior distribution which would have been used for model fitting if the sample data were available, which is called the fitting prior. The fitting prior is to be used to obtain the posterior distribution for making inference. The fitting prior is often assumed to be non-informative, thus it does not influence the sample size much.

Although from a Bayesian perspective the unknown parameter is assumed to be random, to perform power analysis in a Bayesian framework, it is natural to assume that the parameter follows an informative prior distribution concentrated around some specific values of the parameter which are of particular interest to the practitioner. This is called the sampling prior, used after full consideration of all the available prior information. The sampling prior addresses the sensitivity scenarios, such as what if a small change in a parameter results in relatively large changes in the outcomes. And if that is the case, then the parameter has to be determined very accurately or that the alternative has to be redesigned for low sensitivity. Unlike common fitting priors, the sampling prior has a large influence on the optimal sample size, similar to

5.2 The sum of Type I error and Type II error

Let $X^{(n)} = (X_1, ..., X_n)$ denote a random sample size of n from a population with density $f(x|\theta)$ and let $\pi(\theta)$ denote the prior distribution for the unknown parameter θ. Let $\pi(\theta|x^{(n)})$ denote the posterior distribution of θ given the observed sample $x^{(n)}$.

We also set up the hypotheses: $H_o: \theta \in \Theta_o$ versus $H_a: \theta \in \Theta_a$, here we shall take $\Theta_o = \{\theta: \theta \leqslant \theta_o\}$ and $\Theta_a = \{\theta: \theta > \theta_o\}$.

Let A_i denote the action of accepting H_i for $i = o, a$ and $L(\theta, A_i)$ denote the loss for taking decision A_i when the θ is the true value. The Bayes decision rule, denote by δ_n^{π}, is to select A_o if the average posterior loss under A_o is less than that under A_a, i.e. if

$$\int_{\Theta_a} L(\theta, A_o)\pi(\theta|x^{(n)})d\theta < \int_{\Theta_o} L(\theta, A_a)\pi(\theta|x^{(n)})d\theta \qquad (5.1).$$

Under some parametric assumptions, it is often possible to find a suitable function $g(x^{(n)})$ such than inequality (5.1) holds if and only if $g(x^{(n)}) < k^{\pi}(n)$, where $k^{\pi}(n)$ is the value of $g(x^{(n)})$ for which equality holds in expression 5.1.

Due the distinction between the sampling prior ($\pi^{(s)}$) and fitting prior ($\pi^{(f)}$), the Bayes decision risk will have the form

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}}) = \int_{\Theta_a} L(\theta, A_o)P\{g(X^{(n)}) < k^{\pi^{(f)}}(n)|\theta\}\pi^{(s)}(\theta)d\theta + \int_{\Theta_o} L(\theta, A_a)P\{g(X^{(n)}) \geqslant k^{\pi^{(f)}}(n)|\theta\}\pi^{(s)}(\theta)d\theta \qquad (5.2).$$

Assume the constant loss function $L(\theta, a_o) = L_o$ for $\theta > \theta_o$ and $L(\theta, A_a) = L_a$ for $\theta \leqslant \theta_o$, the ratio of losses, L_o/L_a, or, $\eta = \frac{L_o}{L_o + L_a}$.

We can simplify formula 5.2 into the following form:

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}}) = L_o[\int_{\Theta_a} P\{g(X^{(n)}) < k^{\pi^{(f)}}(n)|\theta\}\pi^{(s)}(\theta)d\theta + \frac{1-\eta}{\eta}\int_{\Theta_o} P\{g(X^{(n)}) \geqslant k^{\pi^{(f)}}(n)|\theta\}\pi^{(s)}(\theta)d\theta] \qquad (5.3).$$

Also, note that $\int_{\Theta_a} P\{g(X^{(n)}) < k^{\pi^{(f)}}(n)|\theta\}\pi^{(s)}(\theta)d\theta$ is the probability of Type I error, α, and $\int_{\Theta_o} P\{g(X^{(n)}) \geqslant k^{\pi^{(f)}}(n)|\theta\}\pi^{(s)}(\theta)d\theta$ is the probability of Type II error, β. Based on Formula 5.3, we will get:

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}})/L_o = \alpha + \frac{1-\eta}{\eta}\beta \qquad (5.4).$$

The sample size determination (SSD) problem would be considered as one of finding the minimum n such that

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}})/L_o \leqslant M(\eta) \qquad (5.5).$$

for a given values of η and $M(\eta)$.

If the losses are equal for the two possible wrong decisions, then $L_o = L_a$ so that $\eta = 1$. The equality 5.4 would be $\alpha + \beta \leqslant M(\eta)$. Thus the quantity to be bounded for the SSD is the sum of two error probabilities. $M(\eta)$ is sometimes called the total error rate, and Sahu (2006) list a table including $M(\eta)$.

Sahu, S.K. (2006) said that in some experiment with 0.25, 0.15 and 0.1 as values of $M(\eta)$, which can be interpreted as follows: the test of H_o is carried out at the 5% significance level and it is required to have 80%, 90% and 95% power respectively.

5.3 A clinical trial example

Fayers (2000) discussed the SSD problem for a trial for surgery for gastric cancer where a radical surgery (new treatment) is compared with conventional surgery (standard treatment). The log-hazard-ratio of death (X_1, ..., X_n) is the outcome of the trial and it follows an approximate normal distribution with mean θ and standard deviation $\sigma = 2$; see Spiegelhalter (2004), page 198, for justification of this assumption. The values of $\theta > 0$ favor the new treatment.

5.3.1 Frequentist Approach

For frequentists, the SSD problem is to determine n such that the test of simple null versus simple alternative, i.e.:

$H_o: \theta = 0$ versus $H_a: \theta = \theta_a$

where θ_a is a fixed value specified as the alternative, at 5% significance level achieves 90% power. The new surgery will be selected if the mean log-hazard-ratio is positive.

Let us choose $\theta_a = 0.39$, Under $H_o: \theta = 0$, we have $\bar{X} \sim N(0, 2/\sqrt{n})$. Based on Formula 3.7, we have

$$\beta = P\left(Z \leqslant \frac{z_{1-\alpha} * \sigma/\sqrt{n} - \delta}{\sigma/\sqrt{n}}\right) = P\left(Z \leqslant z_{1-\alpha} - \frac{\delta}{\sigma/\sqrt{n}}\right) \Rightarrow$$

$$1 - 90\% = P\left(Z \leqslant z_{1-5\%} - \frac{0.39}{2/\sqrt{n}}\right) \Rightarrow$$

$$-1.28 = 1.65 - \frac{0.39}{2/\sqrt{n}}$$

$$n \approx 226$$

5.3.2 Bayesian approach

As the form of the hypotheses mentioned in Section 5.2, for this example, our hypotheses are: $H_o: \theta \leqslant 0$ versus $H_a: \theta > 0$. We also assume that $L_o = L_a = 1$, that a wrong decision in either direction will occur the same amount of loss.

Now we still want to consider that the same example mentioned in this section. Fayers *et al.* (2000) reported prior opinions of 26 surgeons who were experienced in gastric surgery . By fitting a normal distribution on an appropriate transformed scale Spiegelhalter *et al.* (2004) concluded that the surgeons' opinion can be summarized by the $N(0.12, 0.19^2)$ prior distribution for θ, which is an enthusiastic prior the new treatment (radical surgery). This corresponds to $n_s = 111$ approximately since $\tau_s^2 = \sigma^2/n_s$, where $\sigma = 2$. Sahu, S.K. (2006) says that when $\theta_a = 0.39$, the sample size that we obtain by using the Bayesian method proposed is 287, given that $M(\eta) = 0.15$.

As discussed in section 5.2, for hypothesis testing, the loss function is 0-1 loss function. SSD problem would be considered as one of finding the minimum n such that

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}})/L_o \leqslant M(\eta) \qquad (5.5).$$

For the particular example mentioned in this section, I simply wish to verify the result. Using $r(\pi^{(s)}, \delta_n^{\pi^{(f)}})$ equation in Appendix A of Sahu, S.K. (2006), also, assume that the fitting prior and the sampling prior are identical,

$$k^{\pi^{(f)}}(n) = \frac{\sigma^2}{n}\left(\frac{\theta_0 - q\lambda_f}{\lambda_f^2} - \frac{\mu_f}{\tau_f^2}\right) = \frac{2^2}{n}\left(\frac{0 - 0\lambda_f}{\lambda_f^2} - \frac{0.12}{0.19^2}\right) = -\frac{0.48}{0.0361n}$$

$$r(\pi^{(s)}, \delta_n^{\pi^{(f)}})/L_o$$

$$= \int_{\frac{(\theta_0-\mu_s)}{\tau_s}}^{\infty} \Phi\left\{\frac{k^{\pi^{(f)}}(n) - \mu_s - \tau_s u}{\sigma/\sqrt{n}}\right\}\phi(u)\mathrm{du} + \int_{-\infty}^{\frac{(\theta_0-\mu_s)}{\tau_s}} \left[1 - \Phi\left\{\frac{k^{\pi^{(f)}}(n) - \mu_s - \tau_s u}{\sigma/\sqrt{n}}\right\}\right]\phi(u)\mathrm{du}$$

$$= \int_{-0.631579}^{\infty} \Phi\left\{\frac{-\frac{0.48}{0.0361n} - 0.12 - 0.19u}{2/\sqrt{n}}\right\}\phi(u)\mathrm{du} +$$

$$\int_{-\infty}^{-0.631579} \left[1 - \Phi\left\{\frac{-\frac{0.48}{0.0361n} - 0.12 - 0.19u}{2/\sqrt{n}}\right\}\right]\phi(u)\mathrm{du}$$

The following are some results we found by using R , when $n = 287$, $r(\pi^{(s)}, \delta_n^{\pi^{(f)}})/L_o =$ 0.0908953+0.05961983=0.1505151 $\approx M(\eta) = 0.15$, while when $n = 290$, $r(\pi^{(s)}, \delta_n^{\pi^{(f)}})/L_o =$ 0.09042767=0.1498407 $\leqslant M(\eta) = 0.15$, which are consistent with the results mentioned in Sahu, S.K. (2006).

6 Discussion

In this article, we have discussed simple methods of estimating the SSD for two types of variables under Frequentist approach and Bayesian approach. Theoretically, it is obvious that the larger samples are better. But practically, it is also true that resources and finances have a major influence on the final sample size chosen, for example, health insurance and managed care providers often do not cover the patient care costs associated with a clinical trial. If participants have to pay for a certain treatment, tests, and other charges, then they might stop taking part in this clinical trial due to cost issues, which will directly influence the final sample size chosen. On the other hand, if a specified organization will sponsor a clinical trial, then there should be more people volunteering for this clinical trial.

To avoid unnecessary loss due to over-sized or under-sized sample size , it is very important to choose the power, the significant level or the total sum of Type I Error and Type II Error.

Typically, calculated sample size should be inflated by 10-20% because possible dropouts will happen in these groups during the process of study. Thus, although the design of an experiment is simplified for the purposes of estimating sample size, it should be noted that using a more sophisticated design and statistical analysis usually provides more power to detect a difference.

Bibliography

1. Lee, S. J. and Zelen, M. (2000) Clinical Trials and Sample Size Considerations: Another Perspective. *Statistical Science*, **Vol. 15, No.2,** 95-110

2. Bahn, Anita K. (1972) Basic medical statistics. *Grune & Stratton*

3. Sahu, S.K. and Smith, T.M.F. (2006) A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society, Series A: Statistics in Society,* **169**,235-253

4. Inoue, L.Y.T., Berry, D.A. and Parmigiani, G. (2005) Relationship between Bayesian and frequentist sample size determination.

5. Suissa, S. and Shuster, J. (1985) Exact unconditional sample sizes for the 2 x 2 binomial trial. *Journal of the Royal Statistical Society, Series A*, Vol **148**, 317-327.

APPENDIX : R CODE

```
########### Program 1 ###########
########## fisher:power ##########
fisher:power <- function(p1; p2; n1; n2; alpha; N)
#p1 is the probability of success in group1
#p2 is the probability of success in group2
#p1 must not be equal to p2; which will guarantee Ho is false
#n1 is the fixed total number of group1
#n2 is the fixed total number of group2
#alpha is the significance level
#N is the number of simulations creating tables for fisher:test
{
k = 0
if (p1 == p2) stop("p1 cannot equal p2")
  for (i in 1: N)
  {
   a = rbinom(1; n1; p1) #a is the number of successes in group1
   b = rbinom(1; n2; p2) #b is the number of successes in group2
   table = matrix(c(a; b; n1 - a; n2 - b); nr = 2; dimnames = list(Group = c
     ("Group1"; "Group2"); Results = c("Success"; "Failure")))
   #table for fisher:test
   p = fisher:test(table)$p:value
    if (p < alpha)
     k = k + 1
  }
power = k/N
#k is the number of times to reject Ho when Ho is false
#k/N is the probability of rejecting Ho when Ho is false; which is power
power
}

#######creating original data set########
alpha = :01
 p1 = 0.1
 for (n1 in c(6; 8; 10)){
  for (delta in 1: 9/10) {
  p2 = p1 + delta
   cat(c(alpha; p1; p2; n1; fisher:power(p1; p2; n1; n1; alpha; 500)); "\n")
 }
 }


alpha = :01
 p1 = 0.2
 for (n1 in c(6; 8; 10)){
  for (delta in 1: 8/10) {
  p2 = p1 + delta
```

```
  cat(c(alpha; p1; p2; n1; fisher:power(p1; p2; n1; n1; alpha; 500)); "\n")
 }
}

alpha = :01
 p1 = 0.5
for (n1 in c(6; 8; 10)){
 for (delta in 1: 5/10) {
  p2 = p1 + delta
  cat(c(alpha; p1; p2; n1; fisher:power(p1; p2; n1; n1; alpha; 500)); "\n")
 }
}


alpha = :05
 p1 = 0.1
for (n1 in c(6; 8; 10)){
 for (delta in 1: 9/10) {
  p2 = p1 + delta
  cat(c(alpha; p1; p2; n1; fisher:power(p1; p2; n1; n1; alpha; 500)); "\n")
 }
}

alpha = 0.05
p1 = 0.2
for (n1 in c(6; 8; 10)){
 for (delta in 1: 8/10)
  {p2 = p1 + delta
  cat(c(alpha; p1; p2; n1; fisher:power(p1; p2; n1; n1; alpha; 500)); "\n")
  }
}


alpha = 0.05
p1 = 0.5
for (n1 in c(6; 8; 10)){
 for (delta in 1: 5/10)
  {p2 = p1 + delta
  cat(c(alpha; p1; p2; n1; fisher:power(p1; p2; n1; n1; alpha; 500)); "\n")
  }
}

## Power Analysis Graph Program#######

output1 = read:table(file:choose(); head = T)
##file name is called "original data"
require(lattice)
## delete the the top stick on the graph
```

APPENDIX

```
tmp <- trellis.par.get("axis.components");
tmp$top$tck <- 0
trellis.par.set("axis.components", tmp)
## make graph backgroud white
trellis.par.set(col.whitebg())
myStrip <- function(which.panel, ...){
    ltext(.25, .5, expression(alpha))
    ltext(.4, .5, paste(" = ", c(.01, .05)[which.panel[1]]))
    ltext(.6, .4, expression(pi[1]))
    ltext(.75, .5, paste(" = ", c(.1, .2, .5)[which.panel[2]]))
}
```

```
xyplot(power ~ n | factor(alpha) factor(prob1), group = prob2,
type = "o", auto.key = list(title = expression(pi[2]), columns = 3),
data = output1, lwd = 2, strip = myStrip, main = "Power Analysis", scale = list(
x = list(alternating = c(1, 1))))
```

```
########### Program 2 ##########
########### fisher.ssd ##########
fisher.ssd <- function(power, p1, p2, alpha, N){
```

```
#power is 1 - type II error
#sd is the standard deviation for the normal approximation
#p1 is the probability of success in group1
#p2 is the probability of success in group2
#p1 must not be equal to p2, which will guarantee Ho is false
#alpha is the significance level
#N is the number of simulations creating tables for fisher.test
#Use normal approximation to estimate the sample size n = n1 = n2

delta = p1 - p2
#delta is difference between the p1 and p2
Var = p1 (1 - p1) + p2 (1 - p2)
n = round(Var (qnorm(1 - alpha/2) + qnorm(power))^2/(delta^2))
#print(n)
power0 = fisher.power(p1, p2, n, n, alpha, N)
ssdTable = c(n, power0)
#print(power0)
if (power0 <power){
while (power0 <power){
n = n + 1
power0 = fisher.power(p1, p2, n, n, alpha, N)
```

```
  ssdTable = rbind( ssdTable; c(n ; power0))
 #print(power0)
 }
 return(list(n = n ; ssdtable = ssdTable))
}
else{
 while (power0 > power)
 { n = n - 1
  power0 = fisher:power(p1; p2; n ; n ; alpha; N)
  ssdTable = rbind( ssdTable; c(n ; power0))
 #print(power0)
 }
 list(n = n + 1; ssdtable = ssdTable)
}
}
```

```
## Plotting the sample size based on given power#######
ssd0.9:
par(mfrow = c(2; 2))
plot(fisher:ssd(0.9; 0.1; 0.5; 0.05; 200)$ssdtable;
main = expression(pi[1] = = 0.1; pi[2] = = 0.5);
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.9; 0.2; 0.8; 0.05; 200)$ssdtable; m
main = expression(pi[1] = = 0.2; pi[2] = = 0.5);
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.9; 0.5; 0.8; 0.05; 200)$ssdtable;
ma
main = "p1 = 0.5 vs p2 = 0.8";
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.9; 0.2; 0.4; 0.05; 200)$ssdtable;

main = "p1 = 0.2 vs p2 = 0.4";
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")


ssd0.6:
par(mfrow = c(2; 2))
plot(fisher:ssd(0.6; 0.1; 0.5; 0.05; 200)$ssdtable;
main = "p1 = 0.1 vs p2 = 0.5";
xlab = "sample size"; ylab = "power";
```

```
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.6; 0.2; 0.8; 0.05; 200)$ssdtable;
main = "p1 = 0.2 vs p2 = 0.8";
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.6; 0.5; 0.8; 0.05; 200)$ssdtable;


main = "p1 = 0.5 vs p2 = 0.8";
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.6; 0.2; 0.4; 0.05; 200)$ssdtable;
main = "p1 = 0.2 vs p2 = 0.4";
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

ssd0.01:
par(mfrow = c(2; 2))
plot(fisher:ssd(0.6; 0.1; 0.5; 0.01; 200)$ssdtable;
main = expression(paste(pi[1] = = 0.1; " "; pi[2] = = 0.5));
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.6; 0.2; 0.8; 0.01; 200)$ssdtable;
main = expression(paste(pi[1] = = 0.2; " "; pi[2] = = 0.8));
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.6; 0.5; 0.8; 0.01; 200)$ssdtable;
main = expression(paste(pi[1] = = 0.5; " "; pi[2] = = 0.8));
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")

plot(fisher:ssd(0.6; 0.2; 0.4; 0.01; 200)$ssdtable;
main = expression(paste(pi[1] = = 0.2; " "; pi[2] = = 0.4));
xlab = "sample size"; ylab = "power";
col:axis = "sky blue"; col:lab = "thistle")


##R code for Bayesian SSD

n = 287
k = (4/n) ( - 12/19^2)
integrand1 <- function(u){pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n))) dnorm(u)}
integrate(integrand1; lower = - 0.631579; upper = Inf)
integrand2 <- function(u){(1 - pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n)))) dnorm(u)}
```

APPENDIX

```
integrate(integrand2; lower = - Inf; upper = - 0.631579)
```

```
n = 286
k = (4/n) ( - 12/19^2)
integrand1 <- function(u) {pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n))) dnorm(u)}
integrate(integrand1; lower = - 0.631579; upper = Inf)
integrand2 <- function(u) {(1 - pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n)))) dnorm(u)}
integrate(integrand2; lower = - Inf; upper = - 0.631579)

n = 250
k = (4/n) ( - 12/19^2)
integrand1 <- function(u) {pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n))) dnorm(u)}
integrate(integrand1; lower = - 0.631579; upper = Inf)
integrand2 <- function(u) {(1 - pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n)))) dnorm(u)}
integrate(integrand2; lower = - Inf; upper = - 0.631579)

n = 290
k = (4/n) ( - 12/19^2)
integrand1 <- function(u) {pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n))) dnorm(u)}
integrate(integrand1; lower = - 0.631579; upper = Inf)
integrand2 <- function(u) {(1 - pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n)))) dnorm(u)}
integrate(integrand2; lower = - Inf; upper = - 0.631579)

n = 289
k = (4/n) ( - 12/19^2)
integrand1 <- function(u) {pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n))) dnorm(u)}
integrate(integrand1; lower = - 0.631579; upper = Inf)
integrand2 <- function(u) {(1 - pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n)))) dnorm(u)}
integrate(integrand2; lower = - Inf; upper = - 0.631579)

n = 288
k = (4/n) ( - 12/19^2)
integrand1 <- function(u) {pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n))) dnorm(u)}
integrate(integrand1; lower = - 0.631579; upper = Inf)
integrand2 <- function(u) {(1 - pnorm((k - 0.12 - 0.19 u)/(2/sqrt(n)))) dnorm(u)}
integrate(integrand2; lower = - Inf; upper = - 0.631579)
```

List of tables

Using Fisher's exact test

α	π_1	π_2	n	power
0.01	0.1	0.2	6	0
0.01	0.1	0.3	6	0
0.01	0.1	0.4	6	0.004
0.01	0.1	0.5	6	0.014
0.01	0.1	0.6	6	0.032
0.01	0.1	0.7	6	0.058
0.01	0.1	0.8	6	0.18
0.01	0.1	0.9	6	0.278
0.01	0.1	1	6	0.528
0.01	0.1	0.2	8	0.002
0.01	0.1	0.3	8	0.002
0.01	0.1	0.4	8	0.018
0.01	0.1	0.5	8	0.072
0.01	0.1	0.6	8	0.148
0.01	0.1	0.7	8	0.282
0.01	0.1	0.8	8	0.412
0.01	0.1	0.9	8	0.614
0.01	0.1	1	8	0.942
0.01	0.1	0.2	10	0
0.01	0.1	0.3	10	0.004
0.01	0.1	0.4	10	0.032
0.01	0.1	0.5	10	0.068
0.01	0.1	0.6	10	0.226
0.01	0.1	0.7	10	0.408
0.01	0.1	0.8	10	0.674
0.01	0.1	0.9	10	0.856
0.01	0.1	1	10	0.982

α	π_1	π_2	n	power
0.01	0.2	0.3	6	0
0.01	0.2	0.4	6	0.002
0.01	0.2	0.5	6	0.002
0.01	0.2	0.6	6	0.012
0.01	0.2	0.7	6	0.028
0.01	0.2	0.8	6	0.066
0.01	0.2	0.9	6	0.15
0.01	0.2	1	6	0.266
0.01	0.2	0.3	8	0.006
0.01	0.2	0.4	8	0.006
0.01	0.2	0.5	8	0.026
0.01	0.2	0.6	8	0.06
0.01	0.2	0.7	8	0.12
0.01	0.2	0.8	8	0.258
0.01	0.2	0.9	8	0.42
0.01	0.2	1	8	0.802
0.01	0.2	0.3	10	0.002
0.01	0.2	0.4	10	0.01
0.01	0.2	0.5	10	0.036
0.01	0.2	0.6	10	0.12
0.01	0.2	0.7	10	0.22
0.01	0.2	0.8	10	0.432
0.01	0.2	0.9	10	0.652
0.01	0.2	1	10	0.884

α	π_1	π_2	n	power
0.01	0.5	0.6	6	0.002
0.01	0.5	0.7	6	0.002
0.01	0.5	0.8	6	0.002
0.01	0.5	0.9	6	0.004
0.01	0.5	1	6	0.024
0.01	0.5	0.6	8	0.002
0.01	0.5	0.7	8	0.004
0.01	0.5	0.8	8	0.02
0.01	0.5	0.9	8	0.064
0.01	0.5	1	8	0.136
0.01	0.5	0.6	10	0.006
0.01	0.5	0.7	10	0.022
0.01	0.5	0.8	10	0.036
0.01	0.5	0.9	10	0.062
0.01	0.5	1	10	0.18

α	π_1	π_2	n	power
0.05	0.1	0.2	6	0
0.05	0.1	0.3	6	0.008
0.05	0.1	0.4	6	0.02
0.05	0.1	0.5	6	0.064
0.05	0.1	0.6	6	0.122
0.05	0.1	0.7	6	0.264
0.05	0.1	0.8	6	0.462
0.05	0.1	0.9	6	0.624
0.05	0.1	1	6	0.854
0.05	0.1	0.2	8	0.008
0.05	0.1	0.3	8	0.026
0.05	0.1	0.4	8	0.11
0.05	0.1	0.5	8	0.176
0.05	0.1	0.6	8	0.404
0.05	0.1	0.7	8	0.572
0.05	0.1	0.8	8	0.806
0.05	0.1	0.9	8	0.946
0.05	0.1	1	8	0.996
0.05	0.1	0.2	10	0.006
0.05	0.1	0.3	10	0.05
0.05	0.1	0.4	10	0.174
0.05	0.1	0.5	10	0.284
0.05	0.1	0.6	10	0.494
0.05	0.1	0.7	10	0.642
0.05	0.1	0.8	10	0.844
0.05	0.1	0.9	10	0.942
0.05	0.1	1	10	1

α	π_1	π_2	n	power
0.05	0.2	0.3	6	0.008
0.05	0.2	0.4	6	0.008
0.05	0.2	0.5	6	0.032
0.05	0.2	0.6	6	0.056
0.05	0.2	0.7	6	0.13
0.05	0.2	0.8	6	0.318
0.05	0.2	0.9	6	0.432
0.05	0.2	1	6	0.686
0.05	0.2	0.3	8	0.016
0.05	0.2	0.4	8	0.058
0.05	0.2	0.5	8	0.134
0.05	0.2	0.6	8	0.192
0.05	0.2	0.7	8	0.38
0.05	0.2	0.8	8	0.582
0.05	0.2	0.9	8	0.788
0.05	0.2	1	8	0.924
0.05	0.2	0.3	10	0.02
0.05	0.2	0.4	10	0.054
0.05	0.2	0.5	10	0.122
0.05	0.2	0.6	10	0.226
0.05	0.2	0.7	10	0.454
0.05	0.2	0.8	10	0.628
0.05	0.2	0.9	10	0.826
0.05	0.2	1	10	0.994

α	π_1	π_2	n	power
0.05	0.5	0.6	6	0.014
0.05	0.5	0.7	6	0.008
0.05	0.5	0.8	6	0.044
0.05	0.5	0.9	6	0.078
0.05	0.5	1	6	0.12
0.05	0.5	0.6	8	0.034
0.05	0.5	0.7	8	0.054
0.05	0.5	0.8	8	0.134
0.05	0.5	0.9	8	0.214
0.05	0.5	1	8	0.36
0.05	0.5	0.6	10	0.032
0.05	0.5	0.7	10	0.058
0.05	0.5	0.8	10	0.122
0.05	0.5	0.9	10	0.286
0.05	0.5	1	10	0.612