Sample Size Estimation while Controlling False Discovery Rate for Microarray Experiments Using the ssize.fdr Package

by Megan Orr and Peng Liu

Abstract

Sample size estimation is important in microarray or proteomic experiments because biologists can typically afford only a few repetitions. Classical procedures to calculate sample size are based on controlling type I error, e.g., family-wise error rate (FWER, the probability of concluding at least one false positive in multiple testing). In the context of microarray and other large-scale genomic data, tests that control false discovery rate (FDR) seem more reasonable and more powerful (Storey and Tibshirani, 2003). Liu and Hwang (2007) proposed a procedure to calculate sample size while controlling FDR. In this paper, we implement this procedure with the **ssize.fdr** R package, which is straightforward to apply and requires minimal computation. This R package is now available from CRAN. Applications include sample size calculations for one-sample, two-sample, and multi-sample experimental designs. Effect sizes and variances can be assumed to be identical or vary among genes.

Note

For more in-depth details and evaluation of this sample size calculation method, please refer to the article: P. Liu and J. T. G. Hwang. Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics*, 23 (6):739–746, 2007.

Introduction

Microarray experiments are becoming more and more popular and critical in many biological disciplines. As in any statistical experiment, appropriate experimental design is essential for reliable statistical inference, and sample size has a crucial role in experimental design. Because microarray experiments are rather costly, it is important to have an adequate sample size that will achieve a desired power without wasting resources.

For a given microarray data set, thousands of hypotheses, one for each gene, are simultaneously tested. Storey and Tibshirani (2003) argued that controlling false discovery rate (FDR) is more powerful than controlling family-wise error rate (FWER) in genomic data. However, the most common procedure used to calculate sample size involves controlling FWER, not FDR.

Liu and Hwang (2007) describe a method for a quick sample size calculation for microarray experiments while controlling FDR. In this paper, we introduce the R package **ssize.fdr** which implements the method proposed by Liu and Hwang (2007). This package can be applied for designing one-sample, two-sample, or multi-sample experiments. The practitioner defines the desired power, the level of FDR to control, the proportion of non-differentially expressed genes, as well as effect size and variance. More specifically, the effect size and variance can be set as fixed or random quantities coming from appropriate distributions. Based on user-defined parameters, this package creates plots of power vs. sample size. These plots allow for visualization of trade-offs between power and sample size, in addition to the changes between power and sample size when the user-defined quantities are modified.

Method

For a given microarray experiment, for each gene, let $H = 0$ if the null hypothesis is true and $H = 1$ if the alternative hypothesis is true. In a microarray experiment, $H = 1$ represents differential expression for a gene, whereas $H = 0$ represents no differential expression. As in Storey and Tibshirani (2003), we assume each test is Bernoulli distributed with the probability $\Pr(H = 0) = \pi_0$, where π_0 is the proportion of non-differentially expressed genes. Liu and Hwang (2007) derived that the following must hold to control FDR at the level of α:

$$\frac{\alpha}{1-\alpha}\frac{1-\pi_0}{\pi_0} \geq \frac{\Pr(T \in \Gamma \mid H = 0)}{\Pr(T \in \Gamma \mid H = 1)} \qquad (1)$$

where T is the test statistic and Γ is the rejection region of the test. For each proposed hypothesis (test) with given user-defined quantities, the sample size is calculated using the following steps:

1. Solve for Γ using (1) for each sample size.
2. Calculate the power corresponding to Γ with appropriate formula for $\Pr(T \in \Gamma \mid H = 1)$.
3. Determine sample size based on desired power.

For specific experimental designs, the numerator and denominator of the right hand side of (1) is replaced by corresponding functions that calculate the type I error and power of the test, respectively.

Functions

The package **ssize.fdr** has six functions which includes three new functions that have recently been developed as well as three functions translated from the previous available **Matlab** codes. The six functions and their descriptions are listed below.

```
ssize.oneSamp(delta, sigma, fdr = 0.05,
   power = 0.8, pi0 = 0.95, maxN = 35,
   side = "two-sided")

ssize.oneSampVary(deltaMean, deltaSE, a, b,
   fdr = 0.05, power = 0.8, pi0 = 0.95,
   maxN = 35, side = "two-sided")

ssize.twoSamp(delta, sigma, fdr = 0.05,
   power = 0.8, pi0 = 0.95, maxN = 35,
   side = "two-sided")

ssize.twoSampVary(deltaMean, deltaSE, a, b,
   fdr = 0.05, power = 0.8, pi0 = 0.95,
   maxN = 35, side = "two-sided")

ssize.F(X, beta, L = NULL, dn, sigma,
   fdr = 0.05, power = 0.8, pi0 = 0.95,
   maxN = 35)

ssize.Fvary(X, beta, L = NULL, dn, a, b,
   fdr = 0.05, power = 0.8, pi0 = 0.95,
   maxN = 35)
```

ssize.oneSamp and **ssize.twoSamp** compute appropriate sample sizes for one- and two-sample microarray experiments, respectively, for fixed effect size (`delta`) and standard deviation (`sigma`). For one-sample designs, the effect size is defined as the difference in the true mean expression level and its proposed value under the null hypothesis for each gene. For two-sample designs, the effect size is defined as the difference in mean expression levels between treatment groups for each gene. In the two-sample case, `sigma` is the pooled standard deviation of treatment expression levels.

ssize.oneSampVary and **ssize.twoSampVary** compute appropriate sample sizes for one- and two-sample microarray experiments, respectively, in which effect sizes and standard deviations vary among genes. Effect sizes among genes are assumed to follow a normal distribution with mean `deltaMean` and standard deviation `deltaSE`, while variances (the square of the standard deviations) among genes are assumed to follow an Inverse Gamma distribution with shape parameter `a` and scale parameter `b`.

ssize.F computes appropriate sample sizes for multi-sample microarray experiments. Additional inputs include the design matrix (`X`), the parameter vector (`beta`), the coefficient matrix for linear contrasts of interest (`L`), a function for the degrees of freedom of the experimental design (`dn`), and the pooled standard deviation of treatment expression levels (`sigma`), assumed to be identical for all genes.

ssize.Fvary computes appropriate sample sizes for multi-sample microarray experiments in which the parameter vector is fixed for all genes, but the variances are assumed to vary among genes and follow an Inverse Gamma distribution with shape parameter `a` and scale parameter `b`.

All functions contain a default value for π_0 (`pi0`), among others quantities. The value of π_0 can be obtained from a pilot study. If a pilot study is not available, a guess based on a biological system under study could be used. In this case, we recommend using a conservative guess (bigger values for π_0) so that the desired power will still be achieved. π_0 can take a vector input, in which case separate calculations are performed for each element of the vector. This allows one to assess the changes of power due to the changes in the π_0 estimate(s).

Each function outputs the following results: a plot of power vs. sample size, the smallest sample size that achieves the desired power, a table of calculated powers for each sample size, and a table of calculated critical values for each sample size.

Examples

In the following examples, we will use Δ_g and σ_g to denote the effect size and the standard deviation of expression levels for gene g, respectively.

ssize.twoSamp

Suppose that we are interested in doing an Affymetrix microarray experiment to compare gene expression levels in samples from a treatment group to those from a control group, and we are only interested in up-regulated genes in the treatment group. In this case, we will perform a one-sided hypothesis test with samples from two groups. Suppose a 2-fold change is desired ($\Delta_g = log_2(2) = 1$) and the standard deviation is know to be 0.5 ($\sigma_g = 0.5$) from previous experiments with the same setup.

The following code will calculate appropriate samples sizes for this example. Please note that we perform calculations for three values of π_0 because we are unsure of an appropriate value for the proportion of non-differentially expressed genes.

```
>p0<-c(0.5,0.9,0.95)
>ts<-ssize.twoSamp(delta=1,sigma=0.5,
   fdr=0.05,power=0.8,pi0=p0,maxN=20,
```

```
side="upper")
```

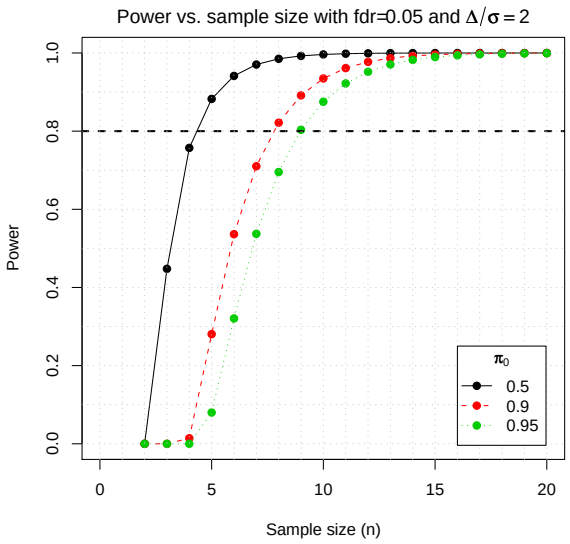

Figure 1: Sample size vs. power for two-sample one-sided upper t-test with user-defined quantities.

Figure 1 shows that the smallest sample sizes to reach the desired power of 0.8 for π_0 values of 0.5, 0.9, and 0.95 are 5, 8, and 9, respectively. This information, along with the calculated powers and critical values can be obtained using the following code:

```
>ts$ssize
>ts$power
>ts$crit.vals
```

ssize.twoSampVary

Suppose that from the pilot study, we know that effect size and variances vary among genes. In addition, a Normal distribution with a mean of 2 and a standard deviation of 1 is appropriate for the effect sizes (Δ_g) and the Inverse Gamma distribution with shape parameter 3 and scale parameter 1 fits well for the variances (σ_g^2). We then can apply the following code to calculate sample size or assess power:

```
>p0<-c(0.5,0.9,0.95)
>N<-20
>tsv<-ssize.twoSampVary(deltaMean=2,
   deltaSE=1,a=3,b=1,fdr=0.05,
   power=0.8,pi0=p0,maxN=20,
   side="two-sided")
```

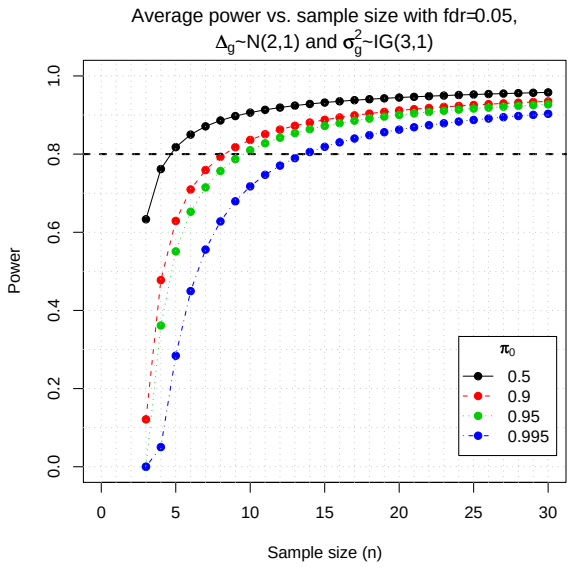

Figure 2: Sample size vs. power for two-sample two-sided t-test with varying Δ/σ. quantities.

Please note that we are doing a two-sided test in this case, which corresponds to interest in both up-regulated and down-regulated genes. From Figure 2, it is clear that the sample sizes of 5, 9, 10, and 14 are the smallest samples sizes to obtain desired power for the π_0 values of 0.5, 0.9, 0.95, and 0.995, respectively. Similar to all the other functions, we can obtain the calculated sample sizes, powers, and critical values using the following code:

```
>tsv$ssize
>tsv$power
>tsv$crit.vals
```

ssize.Fvary

In this example, we want to compare gene expressions among three independent treatment groups with two-color microarray. If we apply a loop design as shown in Figure 3, one "set" requires three slides. The question is how many sets of slides are adequate to obtain a sufficient power at a controlled FDR.

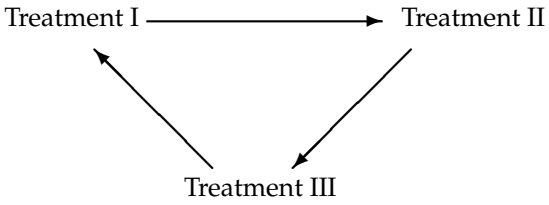

Figure 3: A design example of a three-sample microarray experiment. Each arrow represents one two-color array with the green-labeled sample at the tail and the red-labeled sample at the arrow head.

Because the response is the log-ratio of expression levels of the red-labeled sample to the green-labeled

sample for each slide, the corresponding design matrix for this experiment is (Liu and Hwang, 2007)

$$X = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

If we have n sets, the degrees of freedom are then $(3n - 2)$ because we have $3n$ slides which correspond to $3n$ responses of log-ratios (one for each slide) and 2 parameters to estimate. Suppose the true parameter vector is β=(1,-0.5). In other words, the difference of log-transformed expression levels between the first and second treatments is assumed to be 1 (2-fold change) and the difference between the second and third treatments is assumed to be -0.5 (0.7-fold change) for all genes. Suppose standard deviations among genes differ and σ_g^2 is assumed to follow an Inverse Gamma distribution with shape parameter 3 and scale parameter 1. Running the following code obtains the desired results, including the plot in Figure 4.

```
>des<-matrix(c(1,-1,0,0,1,-1),ncol=2,
>   byrow=FALSE)
>b<-c(1,-0.5)
>df<-function(n){3*n-2}
>p0<-c(0.5,0.9,0.95,0.995)
ftv<-ssize.Fvary(X=des,beta=b,dn=df,
   a=3,b=1,fdr=0.05,power=0.8,
   pi0=p0,maxN=20)
```

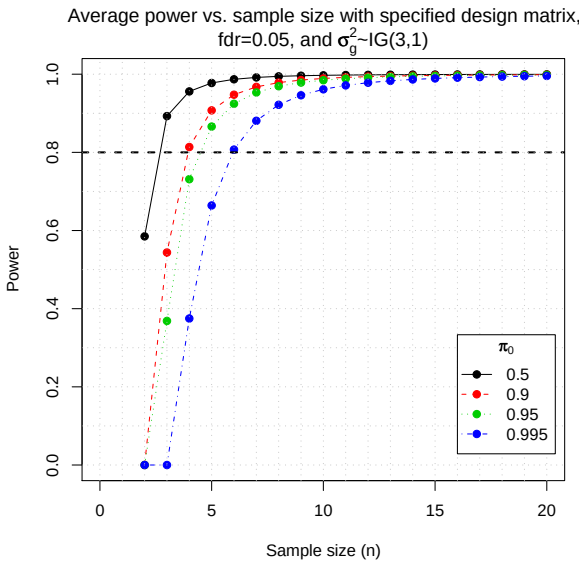

Figure 4: Sample size vs. power for three-sample loop design F-test with varying σ among genes.

ssize.F

This last example looks at a 2x2 factorial design which is commonly encountered in scientific research. For example, suppose we have two groups of plants with different genotypes, wild type and mutant, that are each exposed to two different conditions, nonstress and stress (Nettleton, 2006). This results in four treatments: wild type with nonstress, wild type stress, mutant nonstress, and mutant stress. The means (of normalized log expressions) for each gene may be defined with the following table where α represents the difference in gene expression levels between the stress condition and nonstress condition for the wild type, for example. Note that the parameter θ represents the interaction between genotype and condition (stress or nonstress).

Treatment	Mean
wild type, nonstress	μ_g
wild type, stress	$\mu_g + \alpha$
mutant, nonstress	$\mu_g + \gamma$
mutant, stress	$\mu_g + \alpha + \gamma + \theta$

For this example, we could have one "set" of 2-color microarray slides as illustrated in Figure 5.

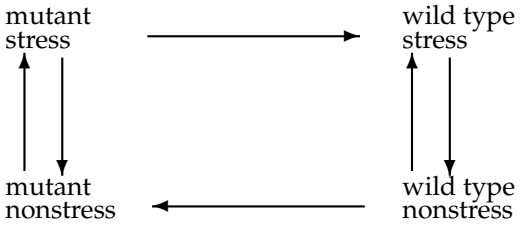

Figure 5: An example of a microarray experiment with a 2x2 factorial design. Each arrow represents one two-color array with the green-labeled sample at the tail and the red-labeled sample at the arrow head.

The design matrix that corresponds to Figure 5 is:

$$X = \begin{bmatrix} 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & -1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

with parameter vector $\beta = (\delta, \alpha, \gamma, \theta)$. The parameter δ represents the difference in dye effects (red and green), and the other parameters are as defined in the above table. Because there are six slides for one set as shown in Figure 5, there will be $6n$ responses of log-ratios for n sets. With four parameters to estimate, there are $6n - 4$ degrees of freedom. The design of Figure 5 could certainly be modified to contain more or less slides, or with different pairings of the samples. In those cases, the degrees of freedom should be adjusted accordingly.

Let the true parameter vector be $\beta = (0.2, 1, 1.5, -0.5)$. To find the number of sets of slides that assures 80% power while controlling FDR at 0.05, we perform the following in **R**:

```
des<-matrix(c(1,1,1,1,1,1,0,-1,0,1,-1,1,
   -1,0,1,0,0,0,-1,0,0,1,-1,0),
```

```
   nrow=6,byrow=FALSE)
B<-c(0.2,1,1.5,-0.5)
df<-function{n}{6*n-4}
sig<-1
p0<-c(0.50,0.90,0.95,0.99)
ft<-ssize.F(X=des, beta=B,dn=df,sigma=sig,
   pi0=p0)
```

Please note that, for simplicity, we assume the same standard deviation (`sigma=1`) for all genes in this calculation. The function `ssize.Fvary` should be the one to use if we assume σ_g^2 follows an Inverse Gamma distribution. Running this code produces Figure 6.

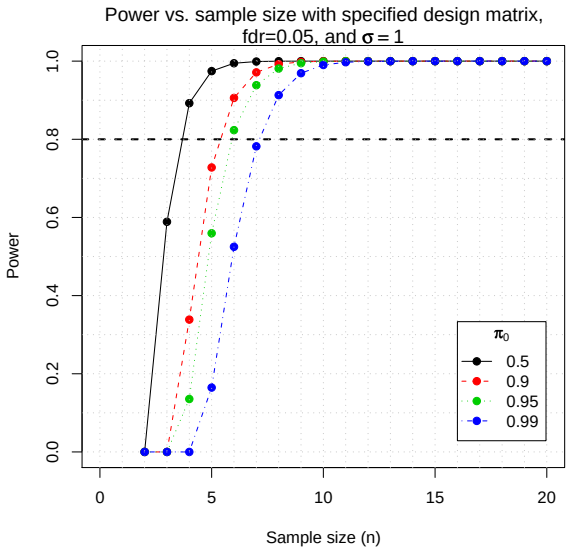

Figure 6: Sample size vs. power for 2x2 factorial design F-test.

We can clearly tell from Figure 6, that the number of sets of slides needed to ensure 80% power are 3, 5, 5, and 6 for π_0 values of 0.5, 0.90, 0.95, and 0.99, respectively. To obtain this information from **R**, along with specific power values and critical values, use the commands `ft$ssize`, `ft$power` and `ft$crit.vals`.

Modifications

Functions **ssize.oneSampVary**, **ssize.twoSampVary**, and **ssize.Fvary** calculate sample sizes using the assumption that at least one parameter is random. Currently, these assumptions are that effect sizes follow Normal distributions and variances follow Inverse Gamma distributions. If users desire to assume that parameters follow other distributions, the code can be modified accordingly.

Contributions

Contributions and discussion are welcome.

Acknowledgement

This material is based upon work partially supported by the National Science Foundation under Grant Number 0714978.

Bibliography

P. Liu and J. T. G. Hwang. Quick calculation for sample size while controlling false discovery rate with application to microarray analysis. *Bioinformatics*, 23(6):739–746, 2007.

D. Nettleton. A discussion of statistical methods for design and analysis of microarray experiments for plant scientists. *The Plant Cell*, 18:2112–2121, 2006.

J. D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA*, 100(16):9440–9445, 2003.