Week5 Report

Xiaohui Li, Yuxin Ma

This week we mainly focus on two parts based on the contents of the meeting last week. First, we want to find the power curve pattern when units N is increased. Next, turning from 'sparse effects' to 'dense effects', we analyze the high-dimensional Y's effect upon power.

1 Units N Analysis

We want to find that if we increase the units number N, how the power pattern will change. We use $N = 100$ to $N = 1,000,000$ increased by 10 times every time. Also, we fix K, $r_1 = \frac{Y_1}{\sigma_1}$ and $r_k = \frac{\sigma_k}{\sigma_1}$ this time where Y_1 varies from 0 to 1, and do the same test for minimum, median and mean value for the $p_{combined}$.

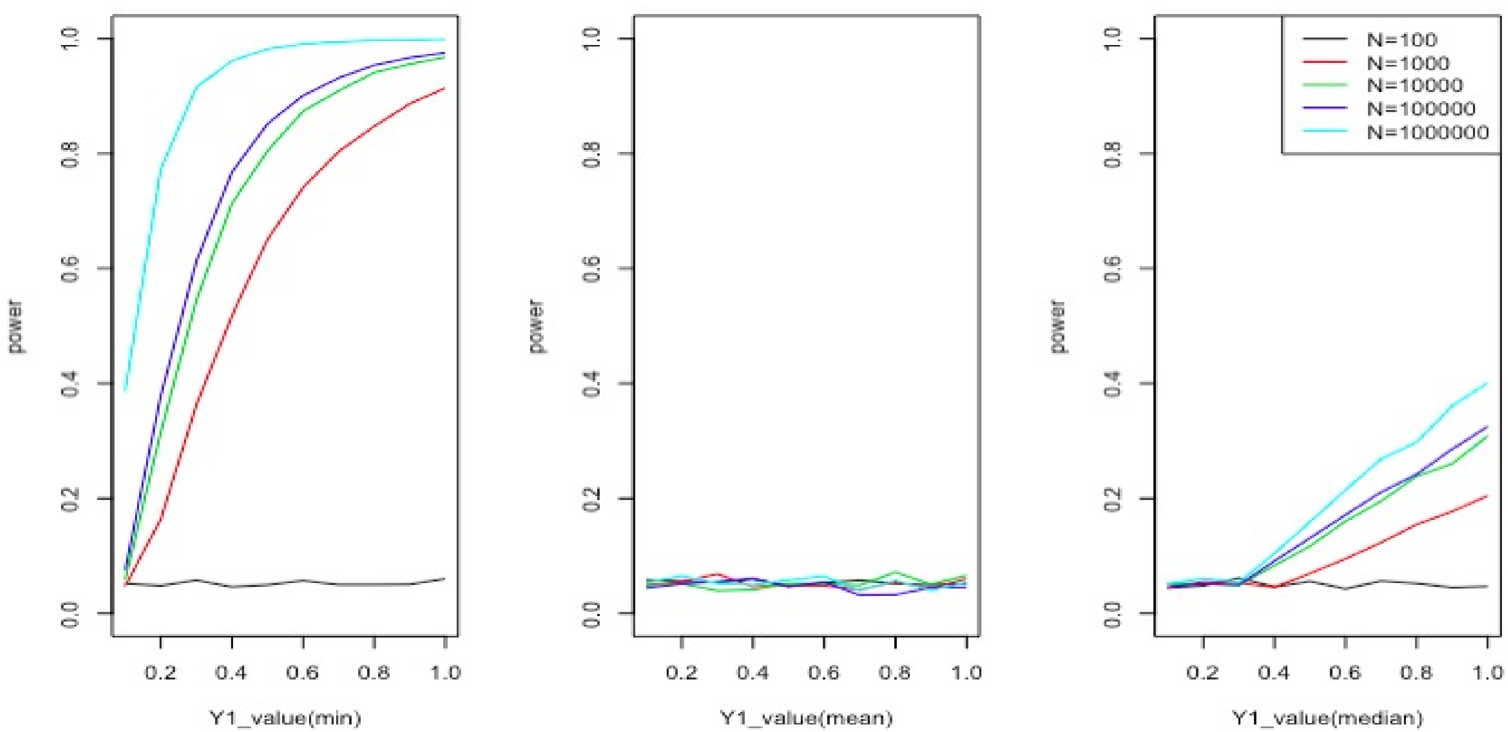

Figure 1.1: Units N analysis

From the plot we can see that, the median and mean perform not well under this condition. So for the minimum, as N increases, the power will converge to 1 faster.

2 High-dimensional Analysis

2.1 Power Analysis

For the further discuss, we do not only vary Y_1 for the treatment. Except fixing $Y_0 = 0$, we vary all the rest Y by multivariable normal distribution this time. We use the 'mvtnorm' package in **R** to generate Y, and since we control

$||Y|| = \epsilon$, we further generate $Y_{norm} = \frac{Y}{\sqrt{\sum Y^2}} \cdot \epsilon$.

For the test, we control all $\sigma = 1$ and vary ϵ from 0 to 1 to compute the power. To get more information, we also do the same test for minimum, median and mean value for the p_combined. Meanwhile, we try different units number N as well.

2.1.1 Test of power variation

As a pretest we test for the variation of power given fixed epsilon norm but random direction level $\vec{Y}$, we assigned the ϵ with value 0.1.

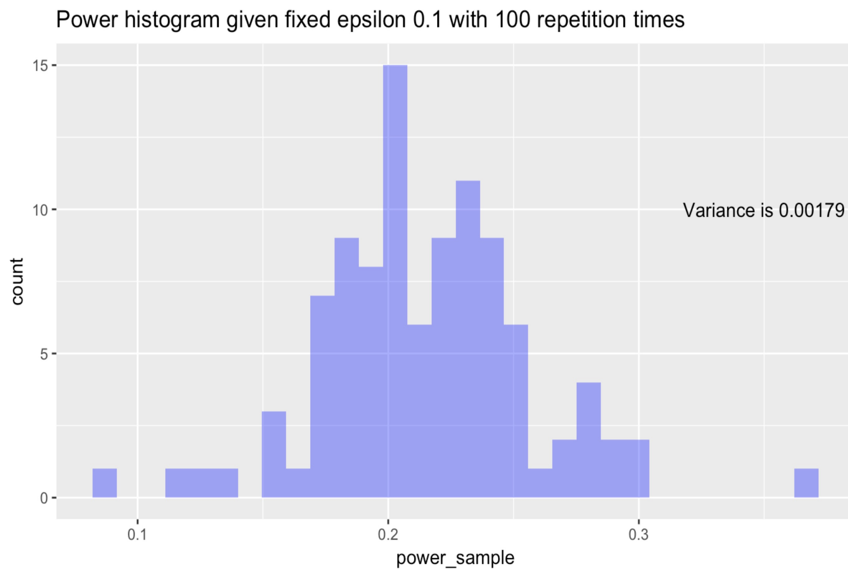

Figure 2.1: Histogram of Power given fixed epsilon

We find that as repetition times increase the sample variance gets less and the mean is more , from pre test, 100 reptition times is accurate enough.

2.1.2 Power curve

The power curves respectively for **minimum**, **median** and **maximum** are shown below.

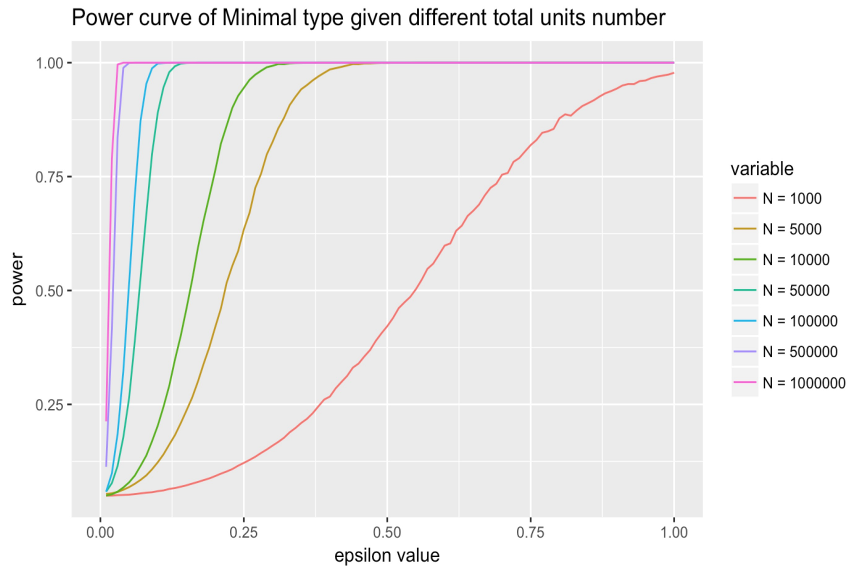

Figure 2.2: High-dimensional analysis for minimum

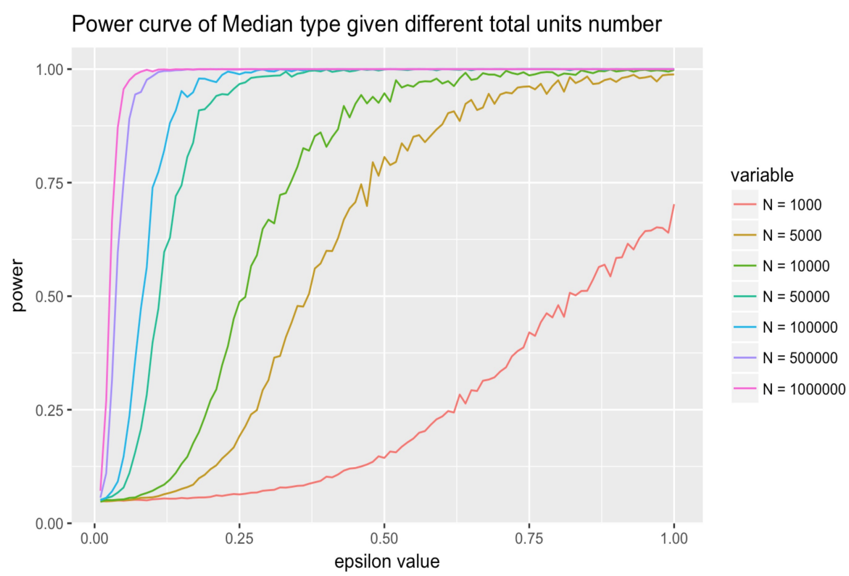

Figure 2.3: High-dimensional analysis for median

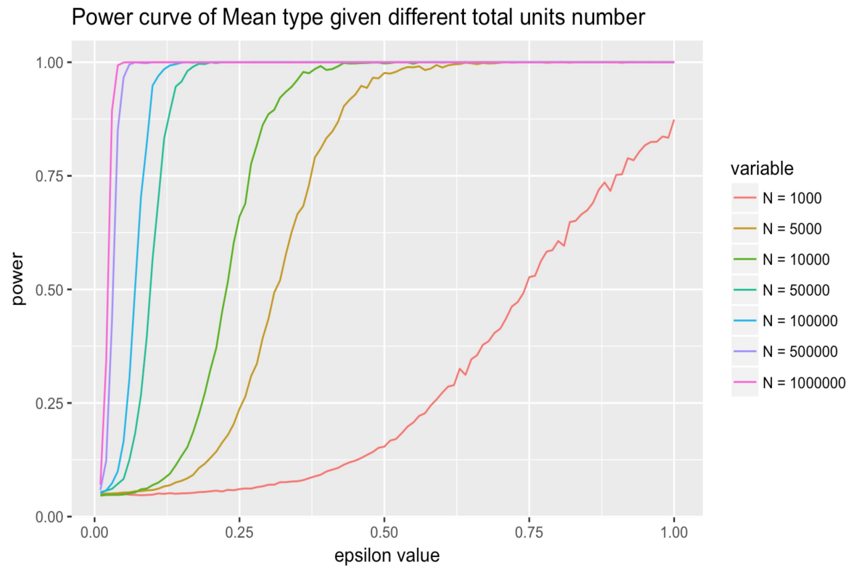

Figure 2.4: High-dimensional analysis for mean

We can see that the units number N has the same trend as before: the larger N, the faster the speed at which the power converges to 1. But unlike in the single normal distribution, 1-dimensional analysis before, all three plots perform pretty well under the high-dimensional analysis. But the minimum seems to be smoother and more stable than the other two. And mean is better than median.

2.2 Critical Epsilon Analysis

We generate a new 80% power critical epsilon function with binary algorithm. And we fix the N to be 10000 and vary the K from 2 to 100 to see the relationship between critical epsilon and K.

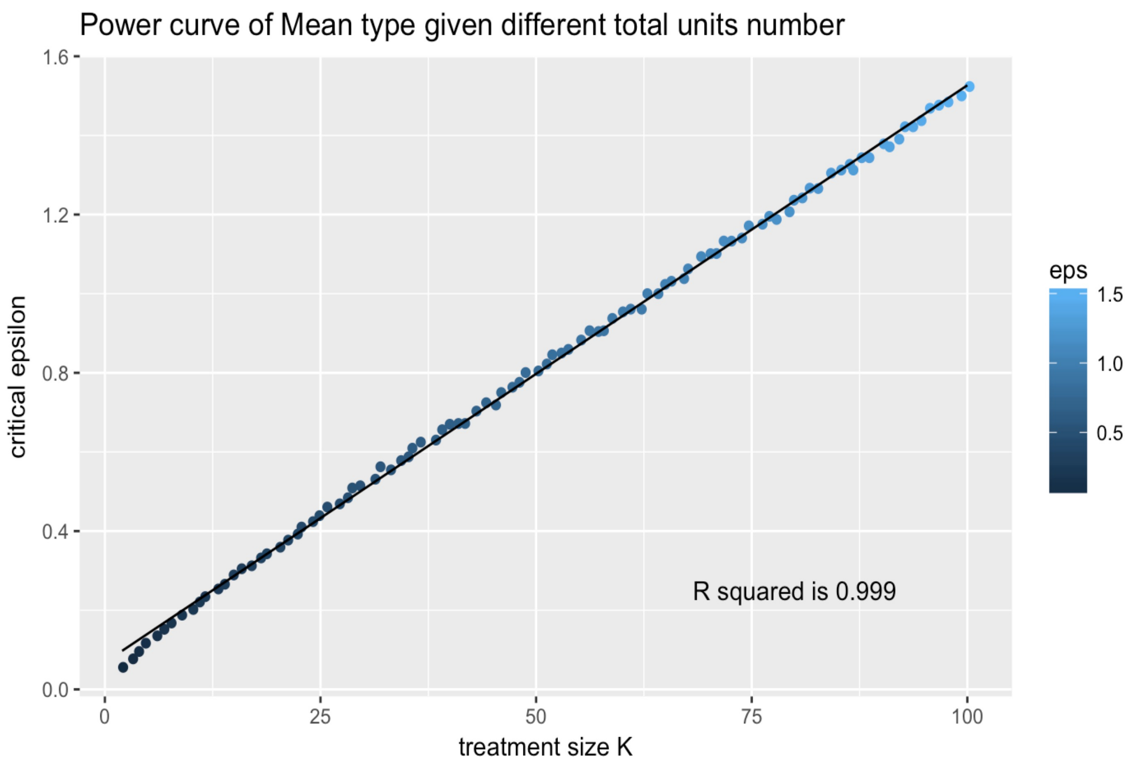

Figure 2.5: Critical epsilon analysis for treatment size K

```
Call:
lm(formula = level_eps ~ k)

Residuals:
      Min        1Q    Median        3Q       Max
-0.041726 -0.007848  0.002534  0.009056  0.027575

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.822e-02  2.738e-03   24.91   <2e-16 ***
k           1.458e-02  4.684e-05  311.37   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01332 on 97 degrees of freedom
Multiple R-squared:  0.999,     Adjusted R-squared:  0.999
F-statistic: 9.695e+04 on 1 and 97 DF,  p-value: < 2.2e-16
```

Figure 2.6: Linear regression summary of Epsilon against treatment K

We find that the critical epsilon is linearly increasing along with K and the R squared is enormously large as 0.999, suggesting theoretical linear model. $\epsilon_{\{P(Y_\epsilon)=0.8)\}} = 1.48K$