Numerical Linear Algebra and Statistical Computing:Project 1

Group 3:
Qinyi Zhou
Akash Roy
Dipnil Chakraborty
Randy Suarez Rodes

Supervisor:
Dr. Larry Ammann

UT DALLAS

Overview

- Introduction
- Methods
- Simultaneous coverage probability for Part 1
- Simultaneous coverage probability for Part 2
- Graphical Analysis
- Dispersion measures related to confidence intervals
- Some improvements
- Conclusion and References

Introduction

- During each election year we are inundated with the results of political preference polls.
- The media presents the results of these polls with little or no discussion of their accuracy.
- TV news reports typically present a set of poll percentages as if they were the actual population percentages, while newspapers may include a small-font footnote stating that the results have an error of $\pm(3-5)\%$.
- This project is intended to examine some of those issues.

Methods

A set of approximate $1-\alpha$ GB confidence intervals for $\Delta_{i,j}$, $1 \leq i < j \leq m$, is

$$\delta_{i,j} \pm \sqrt{\frac{A d_{i,j}}{N}},$$

where

$$\begin{aligned} A &= \chi^2_{M-1}(\alpha/M), \\ d_{i,j} &= \hat{p}_i + \hat{p}_j - (\delta_{i,j})^2 \end{aligned}$$

and $\chi^2_n(\alpha)$ denotes the $1-\alpha$ quantile from the chi-square distribution with n d.f. This critical value can be obtained in R by cv = qchisq(1-alpha,n)
The confidence intervals for FS method are given by

$$\delta_{i,j} \pm \frac{a}{\sqrt{N}},$$

where a is the solution to the equation:

$$1 - 2[1 - \Phi(a)] - 4[m-2][1 - \Phi(a\sqrt{2})] = 1 - \alpha,$$

Simultaneous coverage probability for Part 1

Table 1: **Simultaneous coverage probability for Part 1**

Proportions	Sample size	Method GB	Method FS
(0.5,0.3,0.2)	500	0.9860	0.9655
(0.5,0.3,0.2)	1000	0.9860	0.9690
(0.5,0.3,0.2)	1500	0.9885	0.9700
(0.5,0.48,0.02)	500	0.9865	0.9565
(0.5,0.48,0.02)	1000	0.9920	0.9620
(0.5,0.48,0.02)	1500	0.9905	0.9640

Conclusion from Part 1

- The simultaneous coverage probabilities obtained using method GB are higher than the simultaneous coverage probabilities obtained using method FS for both situations.
- For the first proportion, when the sample sizes increases the simultaneous coverage probabilities increases for both methods.
- But there are no obvious pattern for the other proportion when the sample sizes increases.

Miscoverage probability vs Sample size for Part 1.(a)

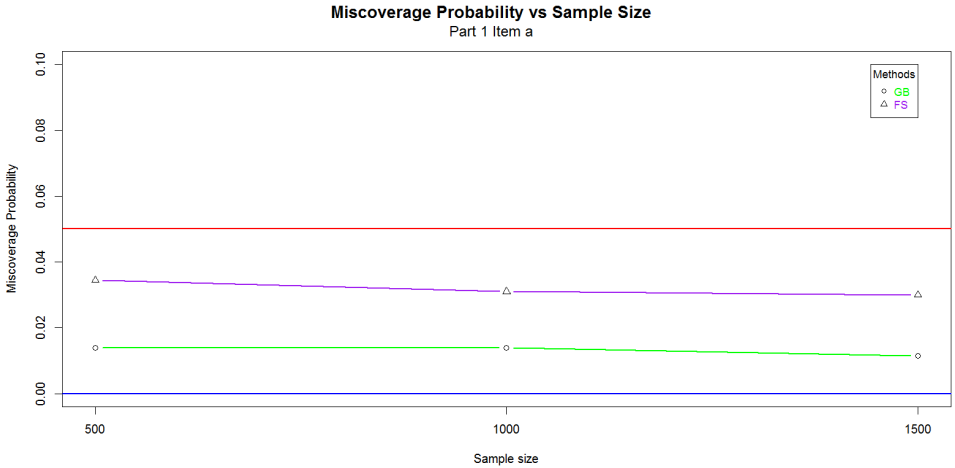


Figure 1: $(p_1, p_2, p_3) = (0.5, 0.3, 0.2)$

Miscoverage probability vs Sample size for Part 1.(b)

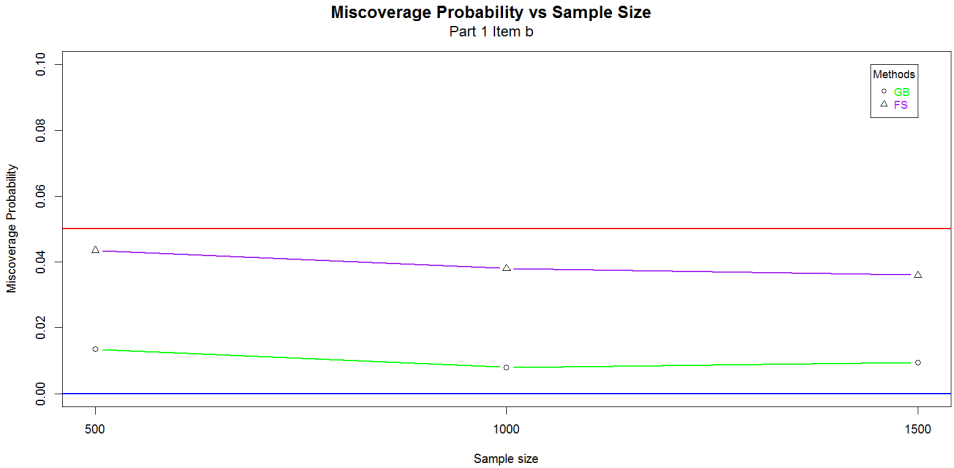
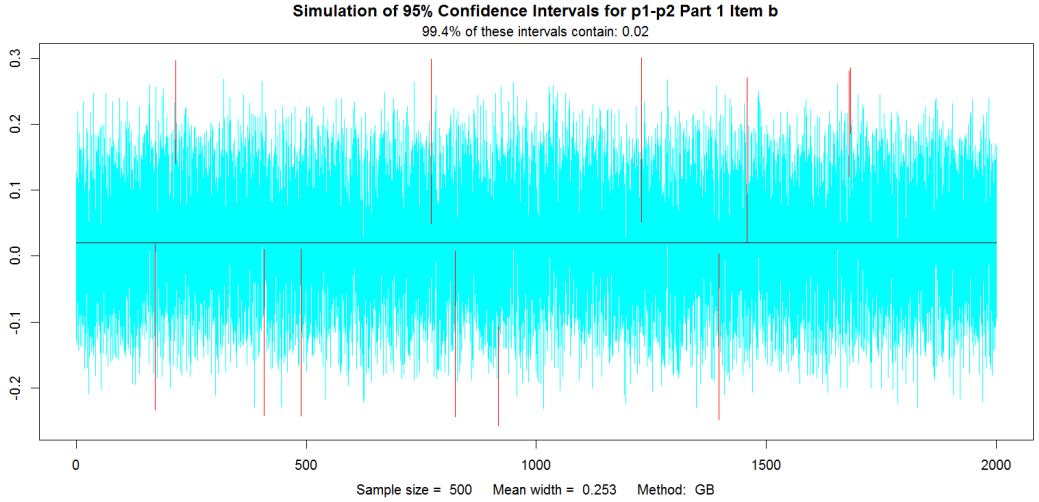
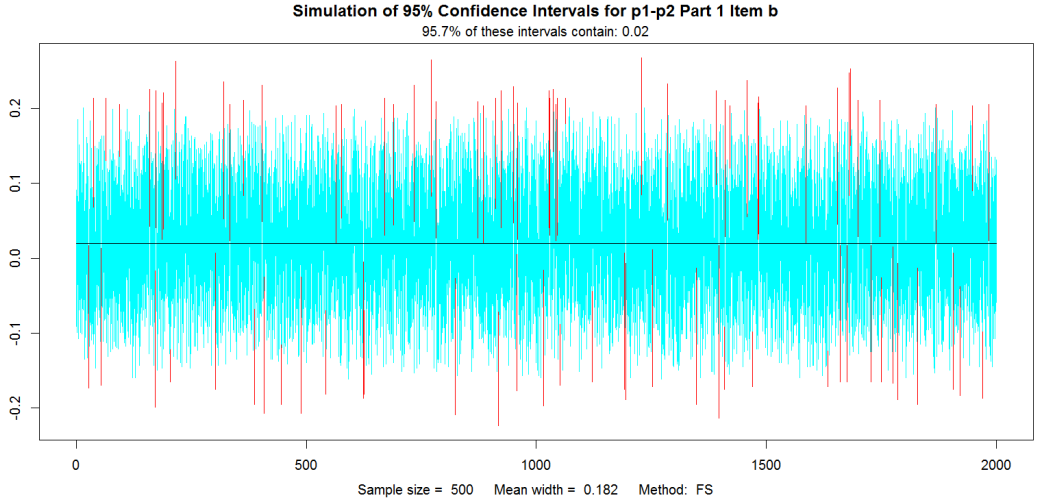


Figure 2: $(p_1, p_2, p_3) = (0.5, 0.48, 0.02)$

GB Confidence interval for p_1-p_2 for N=500(PART 1(b))



FS Confidence interval for p_1-p_2 for N=500(PART 1(b))



Simultaneous coverage probability for Part 2

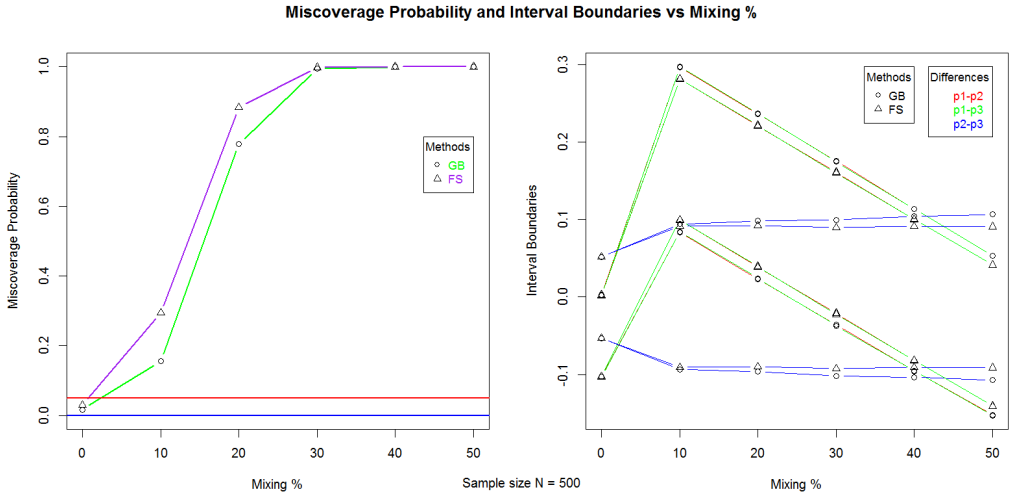
Table 2: **Simultaneous Coverage probability for Part 2**

Sample Size	Method	q=0.1	q=0.2	q=0.3	q=0.4	q=0.5
500	GB	0.8445	0.2215	0.0055	0.0000	0.0000
	FS	0.7060	0.1165	0.0020	0.0000	0.0000
1000	GB	0.6075	0.0160	0.0000	0.0000	0.0000
	FS	0.4215	0.0040	0.0000	0.0000	0.0000
1500	GB	0.3950	0.0010	0.0000	0.0000	0.0000
	FS	0.2260	0.0000	0.0000	0.0000	0.0000

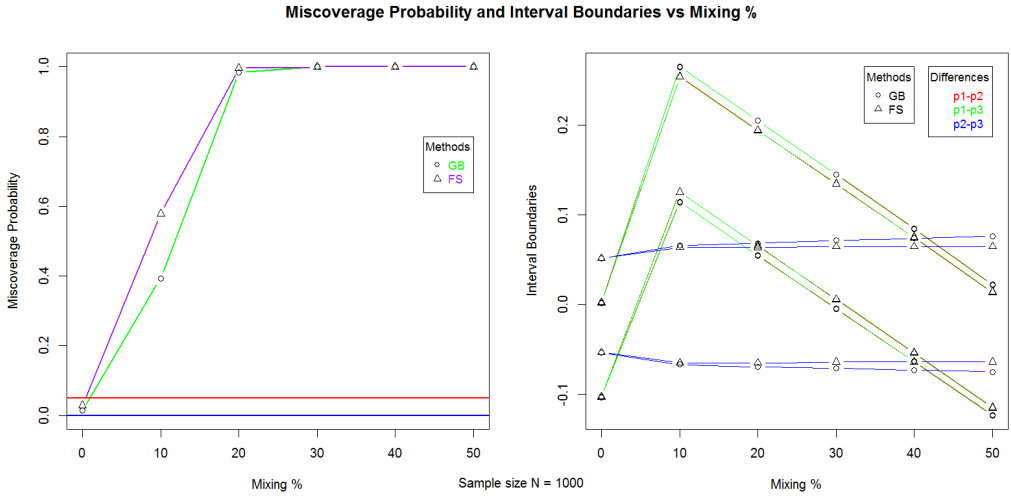
Conclusion from Part 2

- For each q and N, method GB gives higher simultaneous coverage probabilities than method FS.
- When q increases, simultaneous coverage probabilities decrease for both the methods.
- When N increases, simultaneous coverage probabilities decrease for both the methods.

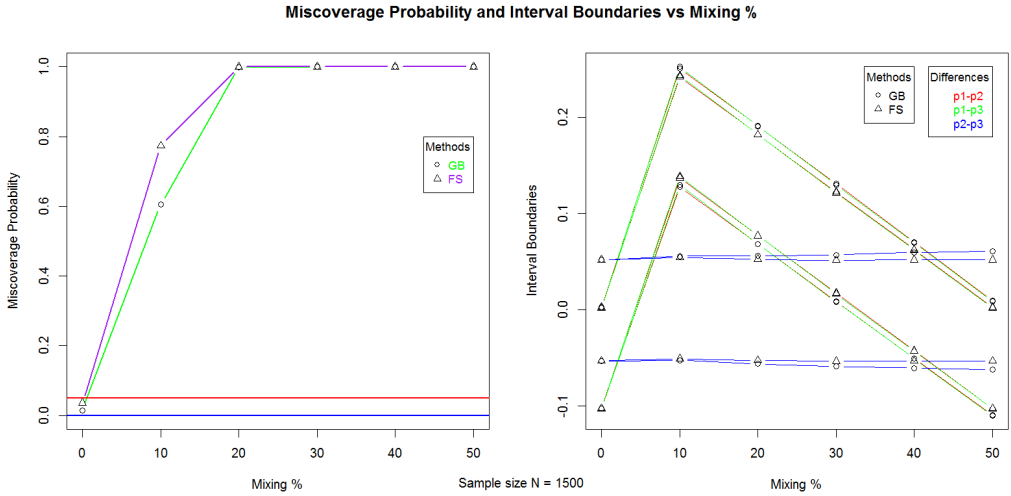
Miscoverage Probability vs Mixing rate for N=500



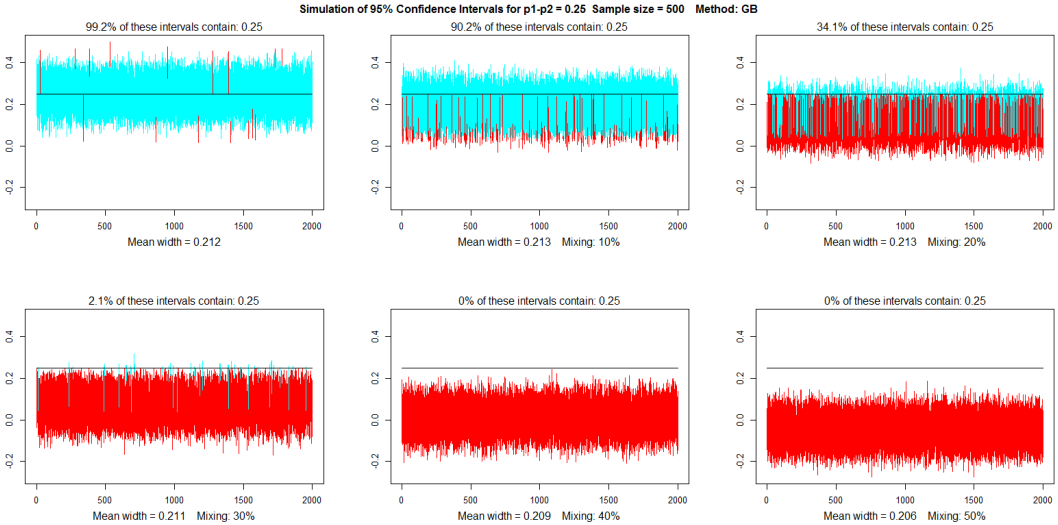
Miscoverage Probability vs Mixing rate for N=1000



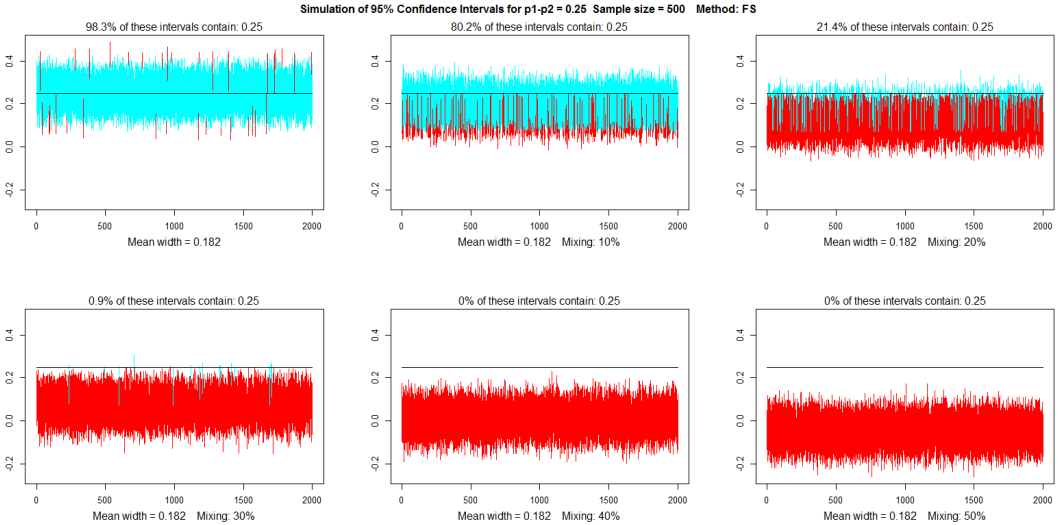
Miscoverage Probability vs Mixing rate for N=1500



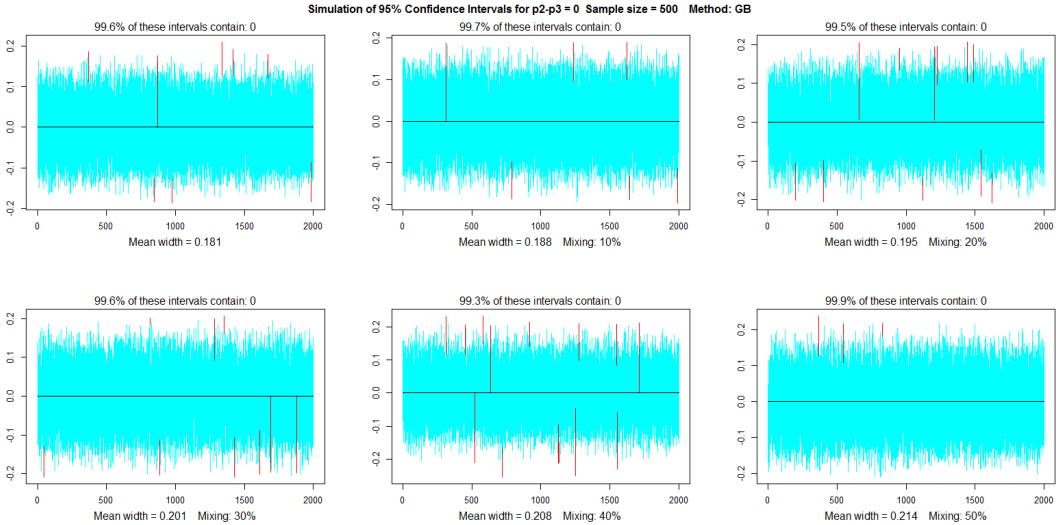
GB Confidence interval for p_1-p_2 for N=500(PART 2)



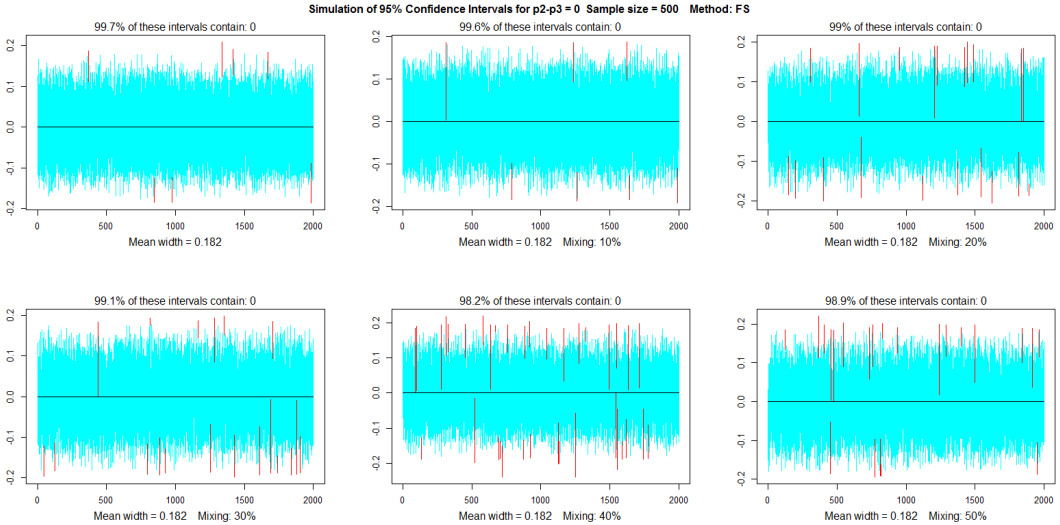
FS Confidence interval for p_1-p_2 for N=500(PART 2)



GB Confidence interval for p_2-p_3 for N=500(PART 2)



FS Confidence interval for p_2-p_3 for N=500(PART 2)



Dispersion measures related to Part 1.(a) from Method GB

Difference	Sample size	mean_CI	sd_CI	IQR_CI	MAD_CI
p1-p2_part-a	500	0.2227	0.0033	0.0049	0.0036
p1-p3_part-a	500	0.1996	0.0042	0.0061	0.0045
p2-p3_part-a	500	0.1786	0.0044	0.0058	0.0043
p1-p2_part-a	1000	0.1577	0.0016	0.0022	0.0016
p1-p3_part-a	1000	0.1412	0.002	0.0027	0.002
p2-p3_part-a	1000	0.1266	0.0021	0.0028	0.0021
p1-p2_part-a	1500	0.1288	0.0011	0.0015	0.0011
p1-p3_part-a	1500	0.1154	0.0014	0.0019	0.0014
p2-p3_part-a	1500	0.1033	0.0013	0.0018	0.0013

Final Conclusion

- Polls need to be careful about how they include or exclude people based on their response to the question “are you voting or not?”.
- They should ask more questions so as to have a better prediction. According to the results it is clear that contamination affects the results of the polls.
- If we compute individual coverage probabilities then only it is possible to explain whether the poll is favourable to any candidate; otherwise not.

Simultaneous coverage probability for Part 1 with Improvement

Table 3: **Coverage probabilities Part 1**

		N=500	N=1000	N=1500
Item a)	GB	0.9860	0.9860	0.9885
	FS	0.9655	0.9690	0.9700
	GB.logratio	0.9890	0.9880	0.9895

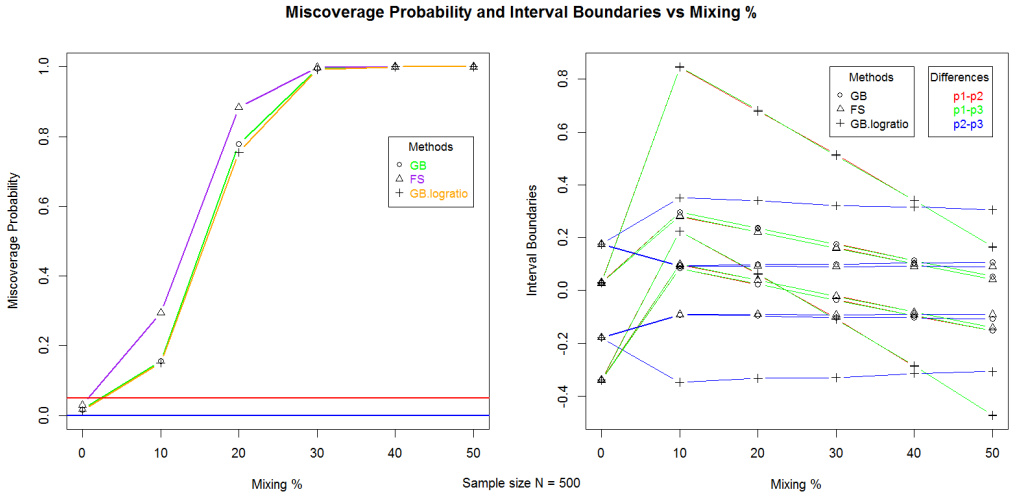
		N=500	N=1000	N=1500
Item b)	GB	0.9865	0.9920	0.9905
	FS	0.9565	0.9620	0.9640
	GB.logratio	0.9870	0.9910	0.9880

Simultaneous coverage probability for Part 2 with Improvement

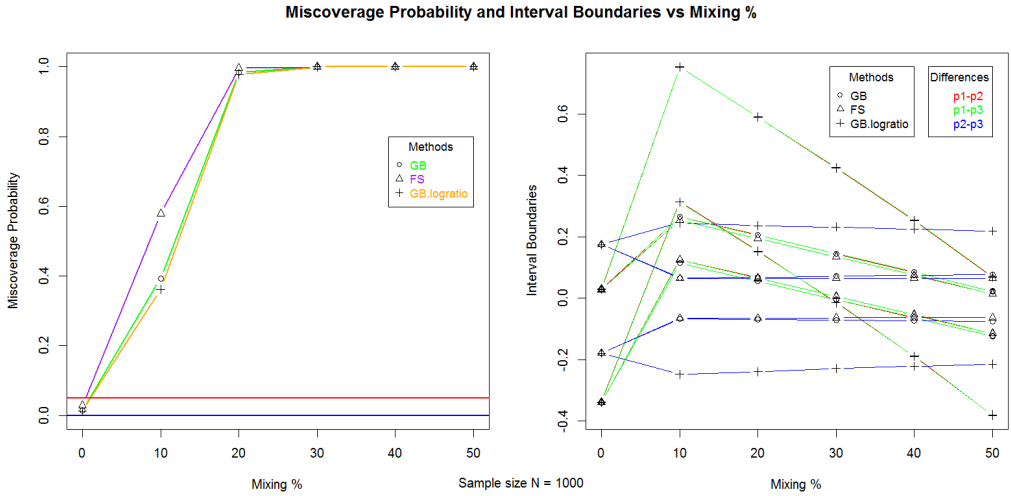
Table 4: **Coverage probabilities Part 2**

Sample Size	Method	q=0.1	q=0.2	q=0.3	q=0.4	q=0.5
500	GB	0.8445	0.2215	0.0055	0.0000	0.0000
	FS	0.7060	0.1165	0.0020	0.0000	0.0000
	GB.logratio	0.8495	0.2455	0.0080	0.0000	0.0000
1000	GB	0.6075	0.0160	0.0000	0.0000	0.0000
	FS	0.4215	0.0040	0.0000	0.0000	0.0000
	GB.logratio	0.6380	0.0220	0.0000	0.0000	0.0000
1500	GB	0.3950	0.0010	0.0000	0.0000	0.0000
	FS	0.2260	0.0000	0.0000	0.0000	0.0000
	GB.logratio	0.4285	0.0015	0.0000	0.0000	0.0000

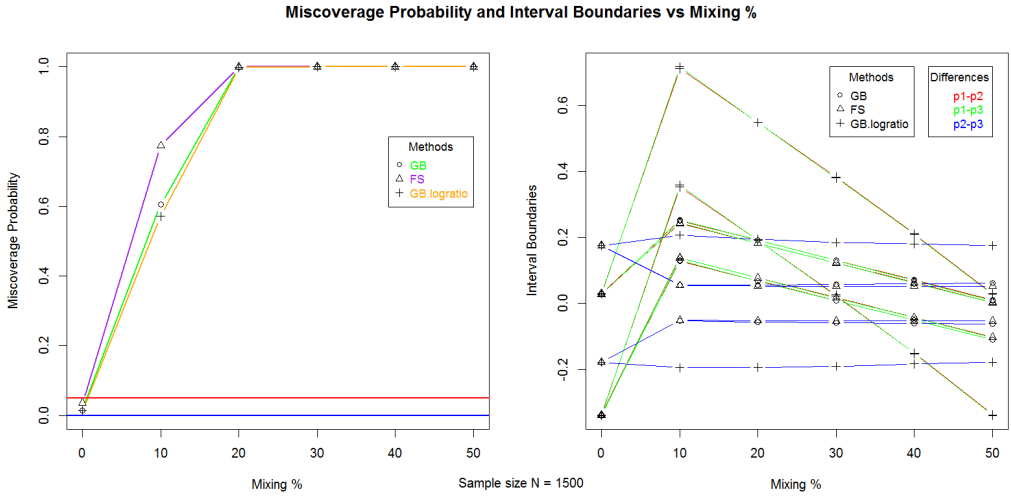
Miscoverage Probability vs Mixing rate with improvement for N=500



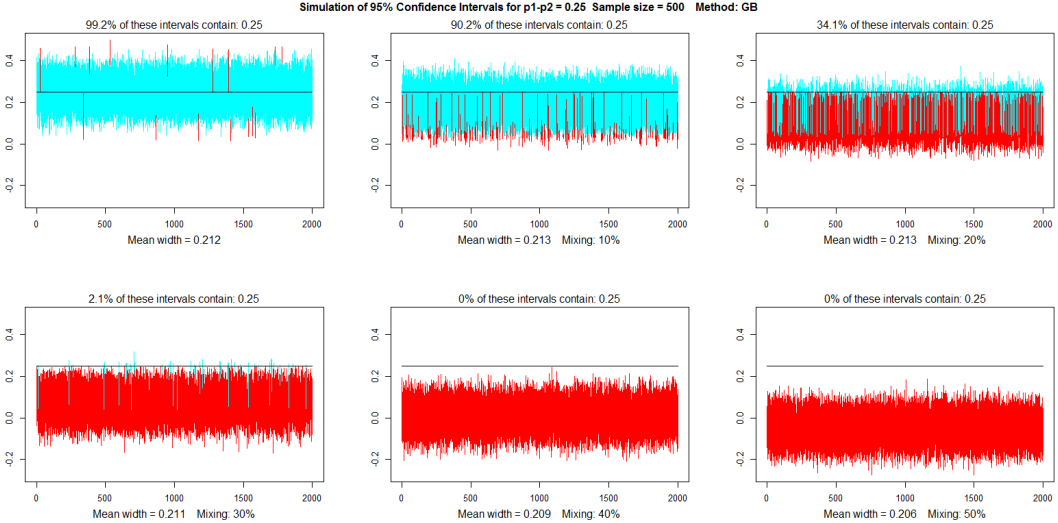
Miscoverage Probability vs Mixing rate with improvement for N=1000



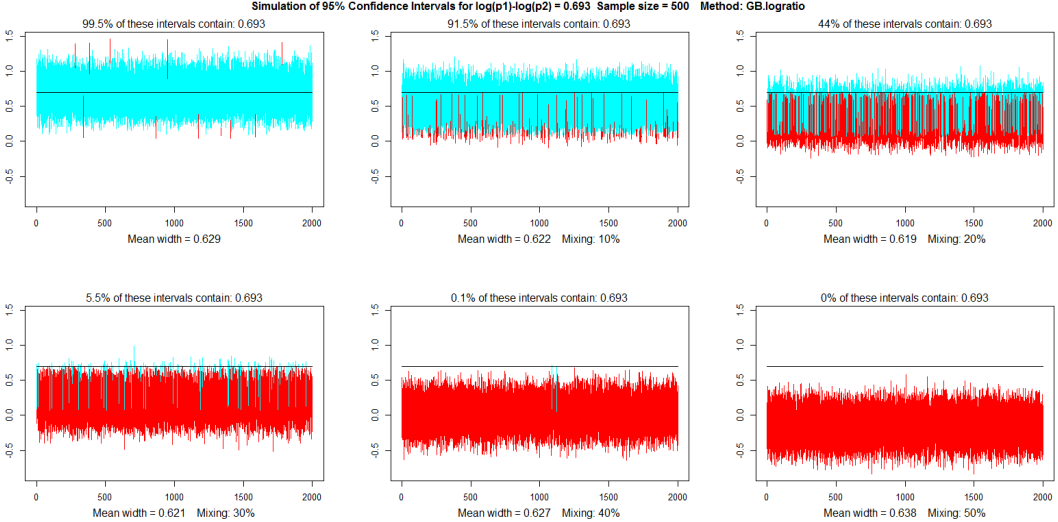
Miscoverage Probability vs Mixing rate with improvement for N=1500



GB CIs for p_1-p_2 for N=500(PART 2)



GB Logratio Confidence interval for p_1-p_2 for N=500(PART 2)



References

1 Piegorsch, W.W. and Richwine, K.A. (2001) Large-Sample Pairwise Comparisons Among Multinomial Proportions with an Application to Analysis of Mutant Species. J. Agricultural, Biological, and Environmental Statistics 6, 3, pp 305-325.

2 Class Notes.

Thank You!