Supplementary tables and figures for “Asymptotic properties of Principal Component Analysis and shrinkage-bias adjustment under the Generalized Spiked Population model”

Rounak Dey[a], Seunggeun Lee[a,*]

[a]*Department of Biostatistics, University of Michigan School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109-2029, USA*

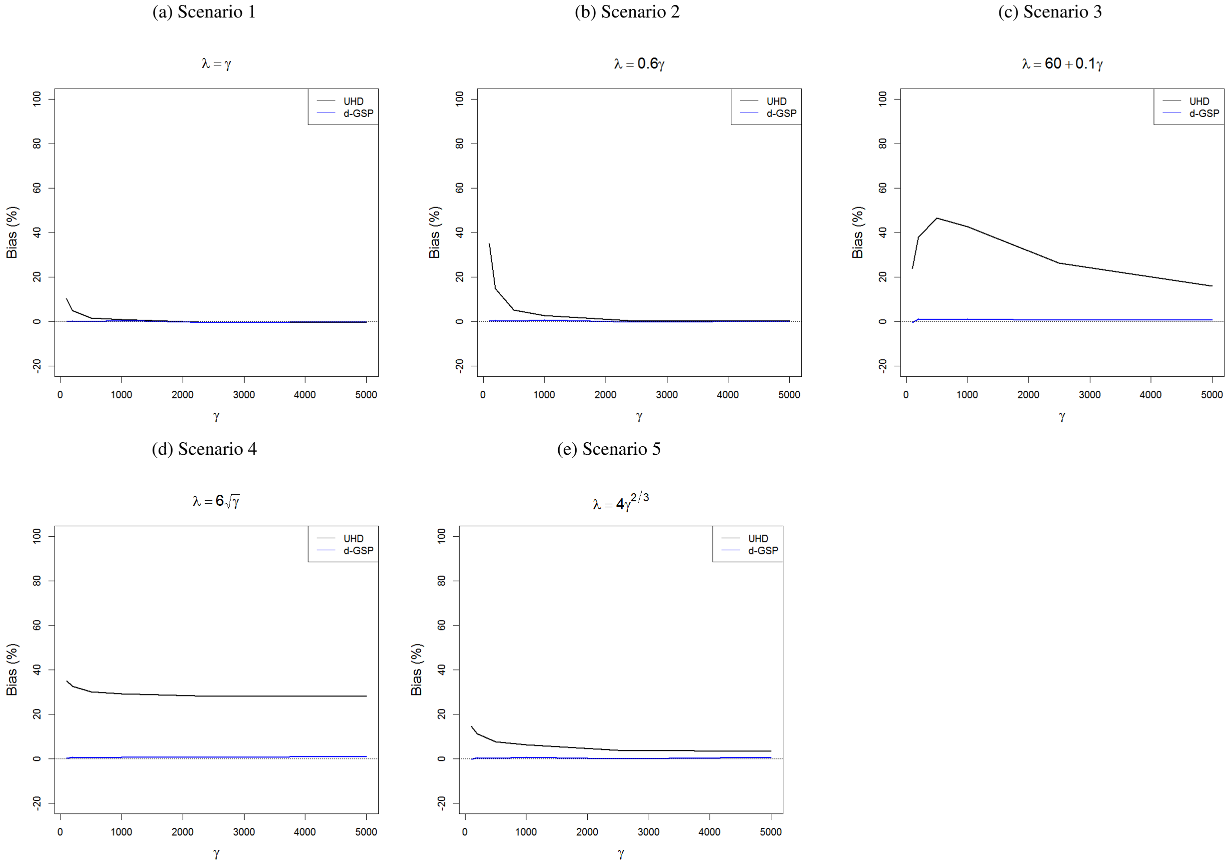
Some additional tables and figures that were referred to in “Asymptotic properties of Principal Component Analysis and shrinkage-bias adjustment under the Generalized Spiked Population model” are provided in this Online Supplement.

Table 1: Number of simulated datasets (out of 200) where the number of distant spikes were estimated to be $1, 2, 3$ or ≥ 4

Settings					**Estimated no. of distant spikes**			
No.	n	p	σ^2	ρ	1	2	3	≥ 4
1	500	5000	4	0.8	0	77	77	46
2	500	5000	1	0.7	0	139	53	8
3	500	5000	7.5	0.8	95	78	20	7
4	500	5000	4	0	0	188	12	0

[*]Corresponding author. Email address: `leeshawn@umich.edu`

Figure 1: Empirical biases (%) in estimating the shrinkage factor corresponding to the largest population eigenvalue for GSP-based and UHD-based methods. The population eigenvalues and the rate of increment of the largest population eigenvalue are assumed to be unknown.



(a) Scenario 1

(b) Scenario 2

(c) Scenario 3

(d) Scenario 4

(e) Scenario 5

Figure 2: Linkage disequilibrium

Hapmap III CEU+TSI samples, 848 SNPs on Chr. 9, created using Haploview [1]

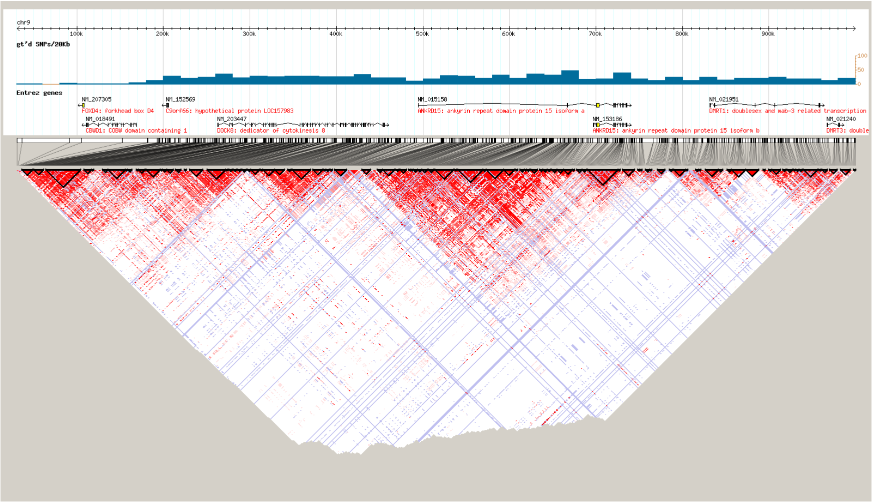


Figure 3: Sample sizes of the test samples that were included in the prediction error estimation for different values of the thresholding parameter ϵ

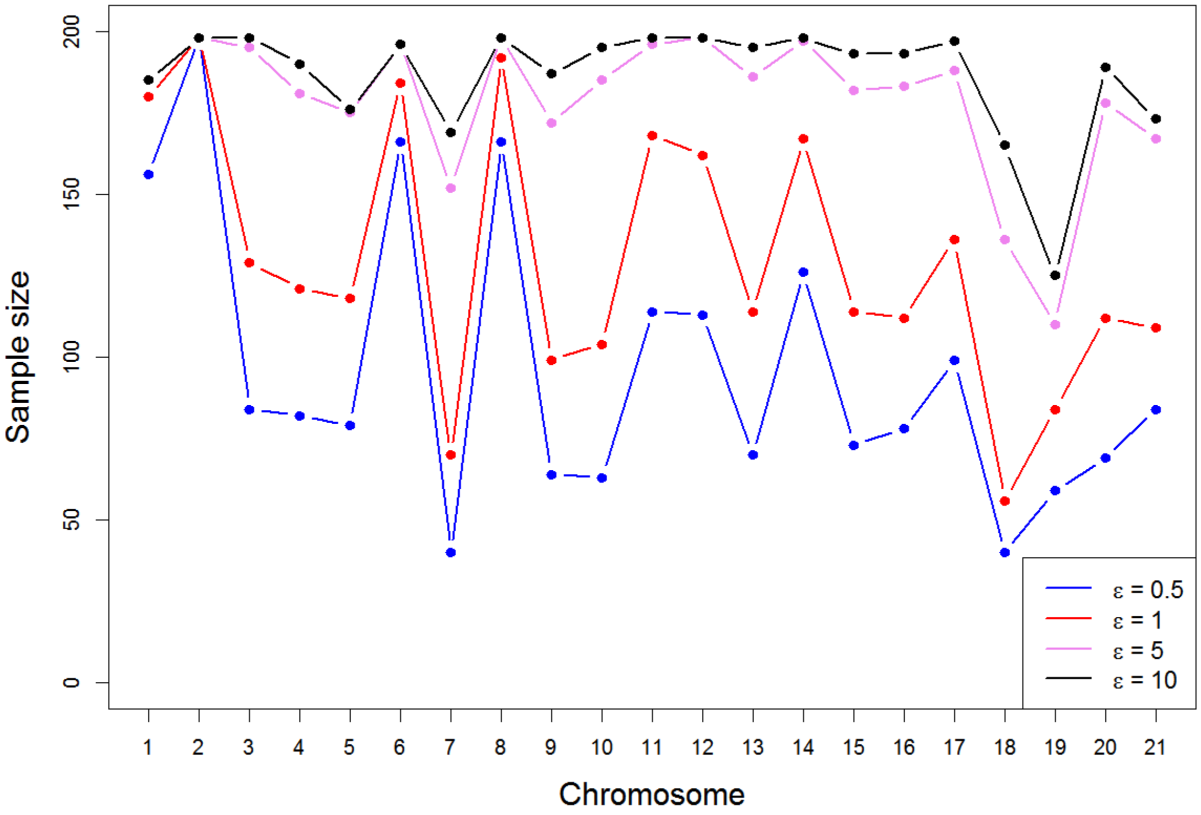


Figure 4: Distribution of the number of markers across different chromosomes

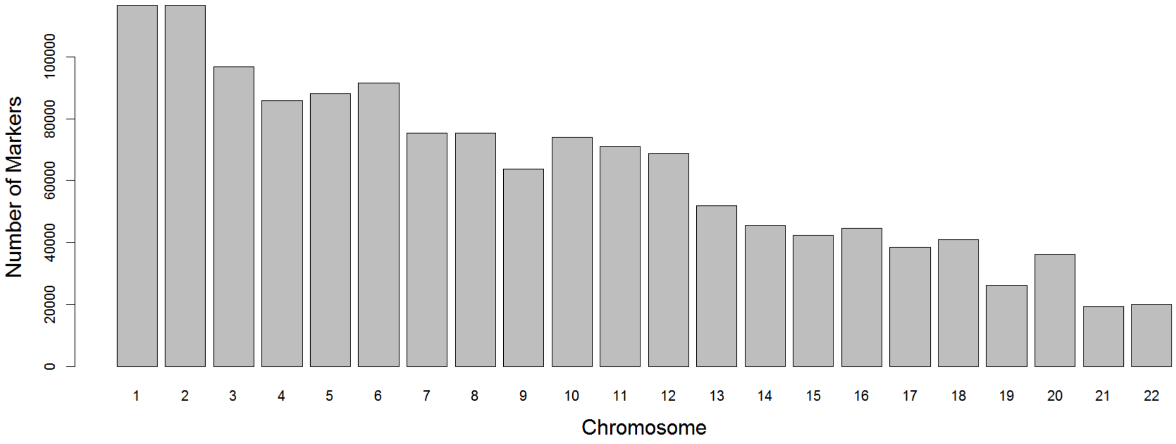
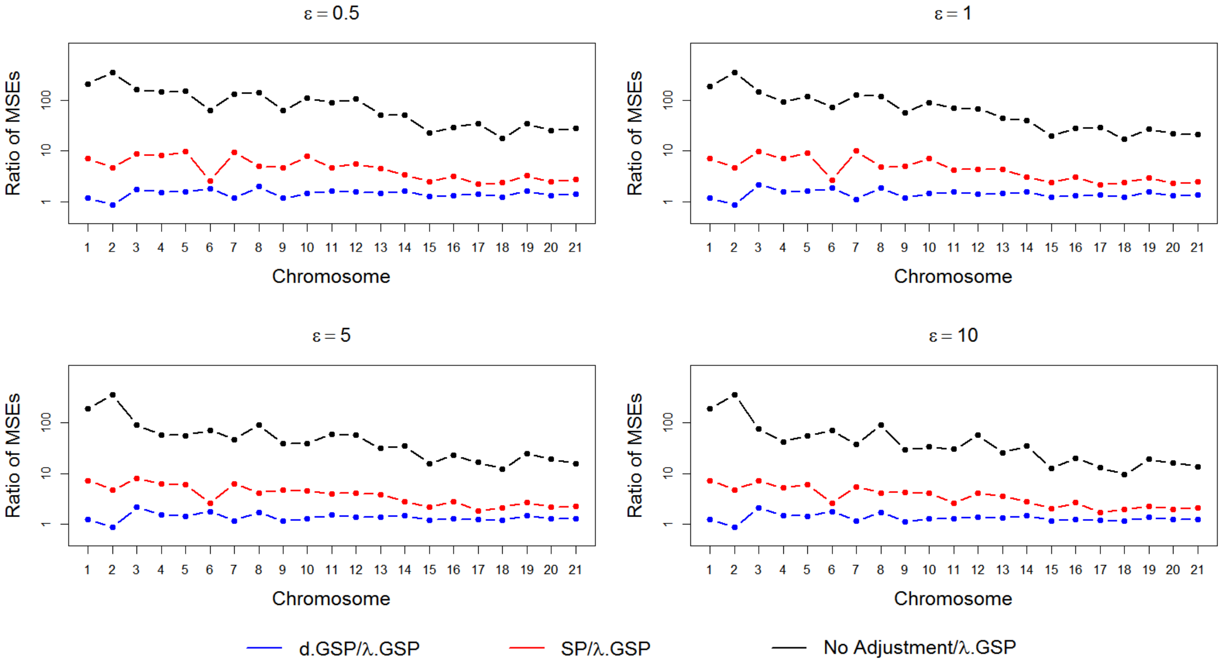


Figure 5: Comparison of the mean squared errors (MSE) of the unadjusted and adjusted PC scores based on the d-GSP and SP methods with the adjusted PC scores based on the λ-GSP method. The ratios of the MSEs are presented for chromosome 1-21 using different values of the thresholding parameter ϵ. The y-axis is presented in a logarithmic scale.



References

[1] J. C. Barrett, B. Fry, J. Maller, M. J. Daly, Haploview: analysis and visualization of ld and haplotype maps, Bioinformatics 21 (2005) 263–265.