Appendix to “Sampling for Conditional Inference on Contingency Tables” Published in the Journal of Computational and Graphical Statistics

Robert D. Eisinger and Yuguo Chen

Appendix A: Proof of Proposal 1

The approximation of Good (1976) implies the number of tables after sampling the first column is approximately

$$|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}| \approx \Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^G = \frac{\prod_{i=1}^m \binom{n+r_i-\alpha_{i1}-2}{r_i-\alpha_{i1}} \prod_{j=2}^n \binom{m+c_j-1}{c_j}}{\binom{M-c_1+m(n-1)-1}{M-c_1}},$$

and the approximation to the total number of tables $|\Sigma_{\mathbf{rc}}| \approx \Delta_{\mathbf{rc}}^G$ is given in (6). Consequently, the proposal SIS-G is

$$q(t_1 = (\alpha_{11}, \dots, \alpha_{m1})) \propto \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^G}{\Delta_{\mathbf{rc}}^G} \propto \prod_{i=1}^m \binom{n+r_i-\alpha_{i1}-2}{r_i-\alpha_{i1}}.$$

Appendix B: Proof of Proposal 2

Using the approximation $\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}$ from Greenhill and McKay (2008), we have

$$|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}| \sim \Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}} = \frac{\prod_{i=1}^m \binom{n+r_i-\alpha_{i1}-2}{r_i-\alpha_{i1}} \prod_{j=2}^n \binom{m+c_j-1}{c_j}}{\binom{M-c_1+m(n-1)-1}{M-c_1}} \exp\{\boldsymbol{\alpha}(\mathbf{r}^{(2)}, \mathbf{c}^{(2)})\},$$

and

$$|\Sigma_{\mathbf{rc}}| \sim \Delta_{\mathbf{rc}}^{\text{GM}} = \frac{\prod_{i=1}^m \binom{n+r_i-1}{r_i} \prod_{j=1}^n \binom{m+c_j-1}{c_j}}{\binom{M+mn-1}{M}} \exp\{\boldsymbol{\alpha}(\mathbf{r}, \mathbf{c})\}.$$

So the proposal SIS-GM is

$$q(t_1 = (\alpha_{11}, \dots, \alpha_{m1})) \propto \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{\Delta_{\mathbf{rc}}^{\text{GM}}} \propto \prod_{i=1}^m \binom{n+r_i-\alpha_{i1}-2}{r_i-\alpha_{i1}} \exp\{\boldsymbol{\alpha}(\mathbf{r}^{(2)}, \mathbf{c}^{(2)})\}.$$

Appendix C: Proof of Theorem 2

Since the conditions of Theorem 1 hold, we have

$$\left| \frac{|\Sigma_{\mathbf{rc}}|}{\Delta_{\mathbf{rc}}^{\text{GM}}} - 1 \right| = \left| \exp \left\{ O \left(\frac{r^3 c^3}{M^2} \right) \right\} - 1 \right| = O \left(\frac{r^3 c^3}{M^2} \right).$$

Following Zhang and Chen (2013), the approximation error can be written as

$$\begin{aligned} \left| \frac{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|}{|\Sigma_{\mathbf{rc}}|} - \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{\Delta_{\mathbf{rc}}^{\text{GM}}} \right| &= \left| \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{|\Sigma_{\mathbf{rc}}|} \left(\frac{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|}{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}} - 1 \right) - \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{|\Sigma_{\mathbf{rc}}|} \left(\frac{|\Sigma_{\mathbf{rc}}|}{\Delta_{\mathbf{rc}}^{\text{GM}}} - 1 \right) \right| \\ &\leq \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{|\Sigma_{\mathbf{rc}}|} \left(\left| \frac{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|}{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}} - 1 \right| + \left| \frac{|\Sigma_{\mathbf{rc}}|}{\Delta_{\mathbf{rc}}^{\text{GM}}} - 1 \right| \right). \end{aligned}$$

Combine the above results with the fact that

$$\frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{|\Sigma_{\mathbf{rc}}|} = \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|} \frac{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|}{|\Sigma_{\mathbf{rc}}|} \leq \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|} \leq 2$$

for large M , then we have

$$\left| \frac{|\Sigma_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}|}{|\Sigma_{\mathbf{rc}}|} - \frac{\Delta_{\mathbf{r}^{(2)}\mathbf{c}^{(2)}}^{\text{GM}}}{\Delta_{\mathbf{rc}}^{\text{GM}}} \right| = O \left(\frac{r^3 c^3}{M^2} \right).$$

References

- Good, I. J. (1976), “On the application of symmetric Dirichlet distributions and their mixtures to contingency tables,” *The Annals of Statistics*, 4, 1159–1189.
- Greenhill, C., and McKay, B. D. (2008), “Asymptotic enumeration of sparse nonnegative integer matrices with specified row and column sums,” *Advances in Applied Mathematics*, 41, 459–481.
- Zhang, J., and Chen, Y. (2013), “Sampling for conditional inference on network data,” *Journal of the American Statistical Association*, 108, 1295–1307.