Biometrics 000, 000–000 DOI: 000
000 0000

Supporting Information for "Scalable Estimation and Regularization for the Logistic Normal Multinomial Model" by Jingru Zhang and Wei Lin

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A. Proof of Proposition 1

By direct calculation, the Hessian of $\log g(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$ is $\nabla^2 \log g(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) = \text{diag}(\boldsymbol{H}_1,\ldots,\boldsymbol{H}_n)$ with $\boldsymbol{H}_i = -m_i\boldsymbol{A}_i - \boldsymbol{\Sigma}^{-1}$, where $\boldsymbol{A}_i = \text{diag}(\widetilde{\boldsymbol{\pi}}_i) - \widetilde{\boldsymbol{\pi}}_i\widetilde{\boldsymbol{\pi}}_i^T$ and $\widetilde{\boldsymbol{\pi}}_i = (\pi_{i1},\ldots,\pi_{i,p-1})^T$. For any $\boldsymbol{v} = (v_1,\ldots,v_{p-1})^T \neq \boldsymbol{0}$, noting that $\pi_{ij} > 0$, $j = 1,\ldots,p-1$, and $\sum_{j=1}^{p-1}\pi_{ij} < 1$, we have

$$\begin{aligned}
\boldsymbol{v}^T\boldsymbol{A}_i\boldsymbol{v} &= \sum_{j=1}^{p-1}\pi_{ij}v_j^2 - \sum_{j=1}^{p-1}\sum_{k=1}^{p-1}\pi_{ij}\pi_{ik}v_jv_k \\
&> \left(\sum_{j=1}^{p-1}\pi_{ij}v_j^2\right)\left(\sum_{k=1}^{p-1}\pi_{ik}\right) - \sum_{j=1}^{p-1}\sum_{k=1}^{p-1}\pi_{ij}\pi_{ik}v_jv_k \\
&= \frac{1}{2}\sum_{j=1}^{p-1}\sum_{k=1}^{p-1}\pi_{ij}\pi_{ik}(v_j^2 + v_k^2 - 2v_jv_k) \\
&= \frac{1}{2}\sum_{j=1}^{p-1}\sum_{k=1}^{p-1}\pi_{ij}\pi_{ik}(v_j - v_k)^2 \geqslant 0.
\end{aligned}$$

Thus, $\boldsymbol{A}_i$ are positive definite and $\nabla^2 \log g(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta})$ is negative definite, verifying the log-concavity.

Web Appendix B. Split HMC for the LNM Model

Following Shahbaba et al. (2014), we may speed up the HMC algorithm by splitting the Hamiltonian and exploiting partial analytic solutions. To this end, we write $U(\boldsymbol{y}) = U_0(\boldsymbol{y}) + U_1(\boldsymbol{y})$, where

$$\begin{aligned}
U_0(\boldsymbol{y}) &= \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}), \\
U_1(\boldsymbol{y}) &= -\sum_{j=1}^{p-1}x_jy_j + m\log\left(1 + \sum_{k=1}^{p-1}e^{y_k}\right),
\end{aligned}$$

and decompose the Hamiltonian into three parts,

$$H(\boldsymbol{y},\boldsymbol{p}) = \frac{1}{2}U_1(\boldsymbol{y}) + \{U_0(\boldsymbol{y}) + K(\boldsymbol{p})\} + \frac{1}{2}U_1(\boldsymbol{y}),$$

where $K(\boldsymbol{p}) = \boldsymbol{p}^T\boldsymbol{\Sigma}\boldsymbol{p}/2$. The first and third parts yield updates similar to those in (8) and (10), while the second part gives the Hamiltonian equations

$$\frac{d\boldsymbol{y}}{dt} = \boldsymbol{\Sigma}\boldsymbol{p},$$
$$\frac{d\boldsymbol{p}}{dt} = -\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}),$$

which admit an analytic solution

$$\boldsymbol{y}(t) = \boldsymbol{\Sigma}\boldsymbol{p}(0)\sin t + \{\boldsymbol{y}(0) - \boldsymbol{\mu}\}\cos t + \boldsymbol{\mu},$$
$$\boldsymbol{p}(t) = \boldsymbol{p}(0)\cos t - \boldsymbol{\Sigma}^{-1}\{\boldsymbol{y}(0) - \boldsymbol{\mu}\}\sin t.$$

Combining these pieces leads to the modified leapfrog updates

$$\boldsymbol{p} \leftarrow \boldsymbol{p} - \frac{\varepsilon}{2}\nabla U_1(\boldsymbol{y}),$$
$$\boldsymbol{y} \leftarrow \boldsymbol{\Sigma}\boldsymbol{p}\sin\varepsilon + (\boldsymbol{y} - \boldsymbol{\mu})\cos\varepsilon + \boldsymbol{\mu},$$
$$\boldsymbol{p} \leftarrow \boldsymbol{p}\cos\varepsilon - \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\sin\varepsilon,$$
$$\boldsymbol{p} \leftarrow \boldsymbol{p} - \frac{\varepsilon}{2}\nabla U_1(\boldsymbol{y}).$$

It is important to emphasize that our decomposition of the Hamiltonian comes directly from the specific form of the likelihood, rather than from a normal approximation around the posterior mode as suggested by Shahbaba et al. (2014); thus, computation of the posterior mode is not required.

Web Appendix C. Proof of Proposition 2

By Property 5.3 of Aitchison (2003), the generalized eigenproblems $(\boldsymbol{\Sigma}_\rho - \lambda\boldsymbol{H})\boldsymbol{v}_\rho = \boldsymbol{0}$ and $(\boldsymbol{\Sigma} - \lambda\boldsymbol{H})\boldsymbol{v} = \boldsymbol{0}$ have identical sets of eigenvalues with corresponding eigenvectors related by $\boldsymbol{v}_\rho = \boldsymbol{Q}_\rho\boldsymbol{v}$ for some suitably defined matrix $\boldsymbol{Q}_\rho$. Since $\boldsymbol{H} = \boldsymbol{K}^2$, the ordinary eigenproblems $(\boldsymbol{K}^{-1}\boldsymbol{\Sigma}_\rho\boldsymbol{K}^{-1} - \lambda\boldsymbol{I}_{p-1})\widetilde{\boldsymbol{v}}_\rho = \boldsymbol{0}$ and $(\boldsymbol{K}^{-1}\boldsymbol{\Sigma}\boldsymbol{K}^{-1} - \lambda\boldsymbol{I}_{p-1})\widetilde{\boldsymbol{v}} = \boldsymbol{0}$ have the same set of eigenvalues with corresponding eigenvectors $\widetilde{\boldsymbol{v}}_\rho = \boldsymbol{K}\boldsymbol{v}_\rho$ and $\widetilde{\boldsymbol{v}} = \boldsymbol{K}\boldsymbol{v}$. The equality of condition numbers then follows.

Web Appendix D. Algorithm for Finding τ^*

Following the idea of Won et al. (2013), define

$$\tau_{ab} = \frac{\sum_{j=1}^{b} \lambda_j/\kappa + \sum_{j=a}^{p-1} \lambda_j}{b + p - a}$$

and consider the rectangles $R_{ab} = \{(\tau, u) : \lambda_a < \tau \leqslant \lambda_{a-1}, \lambda_{b+1} \leqslant u < \lambda_b\}$, $a = 2, \ldots, p-1$, $b = 1, \ldots, p-2$, in the (τ, u)-plane. Algorithm 1 keeps track of the rectangles R_{ab} that intersect the line $u = \kappa\tau$, and check if the point $(\tau_{ab}, \kappa\tau_{ab})$ lies in the current rectangle. If so, τ_{ab} is the desired τ^*; otherwise, by determining the position of the top-right vertex of the current rectangle relative to the line $u = \kappa\tau$, the algorithm decides which of the two next rectangles to follow.

Algorithm 1 Algorithm for finding τ^*

$a \leftarrow p-1$; find b such that $\lambda_{b+1} \leqslant \kappa\lambda_{p-1} < \lambda_b$.
while $a \geqslant 2$ and $b \geqslant 1$ **do**
  if $\lambda_a < \tau_{ab} \leqslant \lambda_{a-1}$ and $\lambda_{b+1} \leqslant \kappa\tau_{ab} < \lambda_b$ **then**
    return τ_{ab}
  else if $\lambda_b > \kappa\lambda_{a-1}$ **then**
    $a \leftarrow a - 1$
  else
    $b \leftarrow b - 1$
  end if
end while

Web Appendix E. Data-Driven Choice of κ

The tuning parameter κ can be chosen adaptively by K-fold cross-validation. The approach divides the data $\boldsymbol{x}$ into K groups of approximately equal sample size, and treats in turn $K-1$ groups as the training set and the remaining group as the test set. Let $\widehat{\boldsymbol{\eta}}^{(-k)}(\kappa)$ denote

the estimate obtained with tuning parameter κ while leaving out the kth group, and $\widetilde{\ell}^{(k)}(\boldsymbol{\eta})$ the log-likelihood evaluated on the kth group. The optimal κ is then chosen as the smallest value that maximizes the average predictive log-likelihood

$$\mathrm{CV}(\kappa) = \frac{1}{K}\sum_{k=1}^{K} \widetilde{\ell}^{(k)}(\widehat{\boldsymbol{\eta}}^{(-k)}(\kappa)).$$

A Monte Carlo approximation is used to evaluate the predictive log-likelihood.

Web Appendix F. Simulations with Varying Zero Proportions

We further investigate the performance of different methods with varying zero proportions, in particular, when the count data are dense or extremely sparse. For the dense settings, we sampled $\boldsymbol{\xi}$ uniformly from the intervals $[0, 4.5]$ and $[0, 6]$, resulting in proportions of zeros 10.5% and 20.1%, respectively. To generate extremely sparse data, we sampled the first 95 components of $\boldsymbol{\xi}$ from $[0, 4]$ and the rest from $[10, 12]$, resulting in a zero proportion of 80.9%. The simulation results are summarized in Web Table 1. We see that the LNM and LNM+ methods still consistently outperform the other methods. The performance gaps tend to shrink as the data become denser, and tend to grow as the data become sparser. We also observe that the performance of the LN2 method does not improve substantially in the dense settings, since it adds 1 to all counts even when only a small proportion of them are zeros.

[Table 1 about here.]

Web Appendix G. Comparison of Microbiome Compositions

We applied the two-sample test of Cao, Lin, and Li (2018) to the compositions obtained by using different estimation methods. Since the test requires strictly positive proportions, the multinomial method is not applicable. The p-values for the DM, LN1, LN2, LNM, and LNM+ methods are 0.0009, 0.0010, 0.0012, 0.0004, and 0.0001, respectively, with LNM+ giving the strongest evidence for composition changes in the gut microbiome. To assess

the stability of the tests, we maintained a proportion varying from 0.1 to 0.9 of subjects within each group, and randomly assigned the remaining subjects to either group in order to retain the sample sizes. We repeated the resampling procedure 100 times and for each method recorded the empirical power as the proportion of rejections at the 0.05 level. We also obtained the empirical size as the proportion of rejections under completely random assignment. The results are presented in Web Figure 1. We see that using the LNM and LNM+ methods for estimation can substantially boost the statistical powers of the tests, while the sizes of the tests are still controlled at the nominal level.

[Figure 1 about here.]

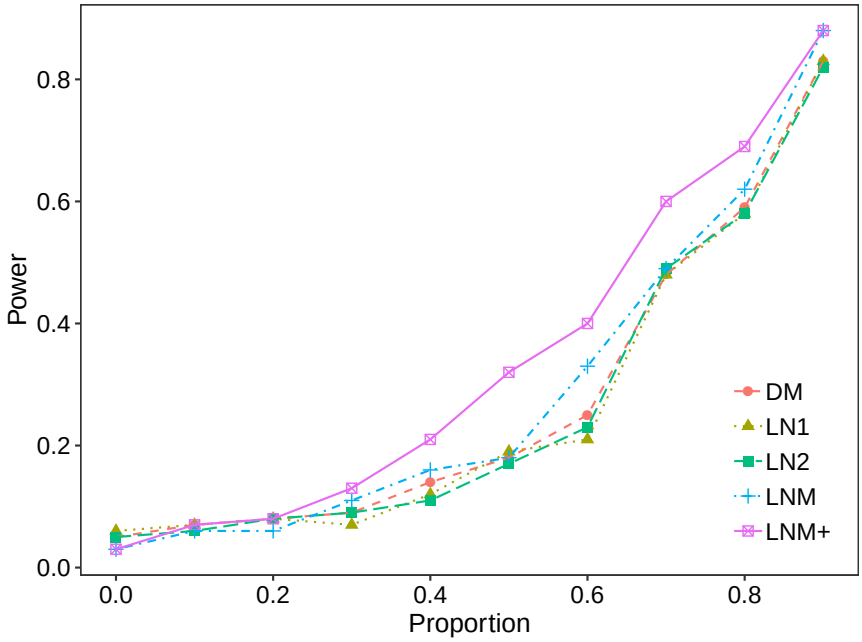
References

Aitchison, J. (2003). *The Statistical Analysis of Compositional Data*. Blackburn Press, Caldwell, NJ.

Cao, Y., Lin, W., and Li, H. (2018). Two-sample tests of high-dimensional means for compositional data. *Biometrika* **105**, 115–132.

Shahbaba, B., Lan, S., Johnson, W. O., and Neal, R. M. (2014). Split Hamiltonian Monte Carlo. *Statistics and Computing* **24**, 339–349.

Won, J.-H., Lim, J., Kim, S.-J., and Rajaratnam, B. (2013). Condition-number-regularized covariance estimation. *Journal of the Royal Statistical Society, Series B* **75**, 427–450.



Web Figure 1. Empirical power curves of the two-sample tests after estimation using different methods for the gut microbiome data.

Web Table 1

Means and standard errors (in parentheses) of the relative errors of parameter estimates for various methods with $n = p = 100$ and varying zero proportions based on 100 simulations. All values have been multiplied by 100.

Zero prop.	Method	$\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_1/\|\boldsymbol{\mu}\|_1$	$\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2/\|\boldsymbol{\mu}\|_2$	$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_2/\|\boldsymbol{\Sigma}\|_2$	$\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_F/\|\boldsymbol{\Sigma}\|_F$	$\|\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_1/\|\boldsymbol{\pi}\|_1$	$\|\widehat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_2/\|\boldsymbol{\pi}\|_2$
10.5%	Mult	—	—	—	—	11.99 (0.02)	8.72 (0.04)
	DM	—	—	—	—	11.88 (0.02)	8.81 (0.04)
	LN1	9.2 (0.2)	9.4 (0.2)	26.4 (1.4)	29.6 (1.2)	11.81 (0.02)	8.68 (0.04)
	LN2	19.2 (0.5)	19.6 (0.4)	27.0 (1.5)	29.3 (1.3)	12.27 (0.02)	9.07 (0.04)
	LNM	8.3 (0.2)	9.2 (0.2)	30.2 (1.6)	34.7 (1.4)	11.91 (0.03)	8.74 (0.04)
	LNM+	7.0 (0.1)	7.2 (0.1)	20.3 (0.7)	22.1 (0.6)	11.55 (0.03)	8.60 (0.04)
20.1%	Mult	—	—	—	—	10.99 (0.03)	7.70 (0.04)
	DM	—	—	—	—	10.89 (0.03)	7.84 (0.04)
	LN1	15.2 (0.6)	16.6 (0.5)	31.3 (1.9)	33.9 (1.8)	10.80 (0.03)	7.67 (0.04)
	LN2	26.0 (0.7)	27.7 (0.6)	32.4 (2.4)	34.7 (2.3)	11.68 (0.02)	8.23 (0.04)
	LNM	10.3 (0.3)	11.4 (0.3)	32.2 (1.9)	38.6 (1.6)	10.81 (0.03)	7.70 (0.04)
	LNM+	7.1 (0.3)	7.2 (0.2)	20.3 (0.6)	22.4 (0.5)	10.49 (0.03)	7.60 (0.04)
80.9%	Mult	—	—	—	—	3.68 (0.02)	2.89 (0.02)
	DM	—	—	—	—	3.64 (0.02)	2.92 (0.02)
	LN1	26.0 (0.2)	28.4 (0.2)	30.3 (1.4)	32.6 (1.2)	4.94 (0.01)	3.29 (0.02)
	LN2	33.8 (0.2)	35.6 (0.2)	30.9 (1.4)	33.2 (1.2)	8.10 (0.01)	4.75 (0.01)
	LNM	5.6 (0.1)	6.7 (0.1)	31.1 (0.6)	48.3 (0.4)	3.40 (0.02)	2.88 (0.02)
	LNM+	4.2 (0.1)	5.5 (0.1)	26.5 (0.7)	28.5 (0.6)	3.32 (0.02)	2.88 (0.02)