Some supplements

1 Convexity

Theorem 1.1.

$$\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b}) = \sum_{\omega\in\Omega}\sum_{\ell\in[L]} \Big\{\mathbb{1}_{\{y_\omega=\ell\}} \log\big[f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})\big]\Big\}, \text{ where } f(x) = \frac{e^x}{1+e^x}.$$

is a concave function to $(\Theta, \boldsymbol{b})$

Proof. It is enough to show that $\lambda(u,v) = \log\big[f(u) - f(v)\big]$ is a concave function to (u,v) where $u > v$. Because if $\lambda(u,v)$ is a concave function and u, v are both linear functions of $(\Theta, \boldsymbol{b})$ such that $u = \boldsymbol{a}_1^T(\Theta, \boldsymbol{b}), v = \boldsymbol{a}_2^T(\Theta, \boldsymbol{b})$, then $\lambda(u,v) = \lambda(\boldsymbol{a}_1^T(\Theta, \boldsymbol{b}), \boldsymbol{a}_2^T(\Theta, \boldsymbol{b})))$ is a concave function to $(\Theta, \boldsymbol{b})$ by the definition of the convexity. With the fact that $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ can be written as the form of summations of $\lambda(u,v)$, we can conclude that $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ is a concave function to $(\Theta, \boldsymbol{b})$ if we prove $\lambda(u,v)$ is concave. Write $\lambda(u,v) = \log\big[f(u) - f(v)\big]$ as $\log\big[\int \mathbb{1}_{(u,v)}(x) f'(x)dx\big]$. Notice that $\log \mathbb{1}_{(u,v)}(x)$ is concave to (u,v,x) and $\log f'(x)$ is concave to x because $f'(x) = \frac{e^x}{(1+e^x)^2}$. Thus, $\log \mathbb{1}_{(u,v)}(x) f'(x)$ is a concave function to (u,v,x). Therefore, the $\lambda(u,v)$ is a concave function because the following lemma says that the integral of a log concave function with respect to some of its arguments is a log concave function of its remaining argument. □

Lemma 1 (Corollary 3.5 in Brascamp & Lieb (2002))**.** *Let* $F(x,y) : \mathbb{R}^{m+n} \to \mathbb{R}$ *be an integrable function where* $x \in \mathbb{R}^m, y \in \mathbb{R}^n$. *Let*

$$G(x) = \int_{\mathbb{R}^n} F(x,y)dy.$$

If $F(x,y)$ *is a log concave function to (x,y), then* $G(x)$ *is a log concave function.*

2 Property of an estimator

Property 1. $y_\omega^{(Median)}$ *minimizes* $R(y) = \mathbb{E}_{\hat{\theta}_\omega,\hat{\boldsymbol{b}}}|y_\omega - y|$.

Proof. We can write $R(y) = \sum_{\ell\in[L]} |\ell - y|\hat{f}_\ell$ where $\hat{f}_\ell = f_\ell(\hat{\theta}_\omega, \hat{\boldsymbol{b}})$. Let us denote $\mu = y_\omega^{(\text{Median})}$. Then,

$$R(y) = \begin{cases} \big[-(\hat{f}_1 \cdots + \hat{f}_L)\big]y + \big[(L\hat{f}_L + \cdots + 1\hat{f}_1)\big], & \text{if } y \in (-\infty, 1], \\ \big[\hat{f}_1 - (\hat{f}_2 \cdots + \hat{f}_L)\big]y + \big[(L\hat{f}_L + \cdots + 2\hat{f}_2) - 1\hat{f}_1\big], & \text{if } y \in (1, 2], \\ \vdots & \vdots \\ \big[(\hat{f}_1 + \cdots + \hat{f}_{\mu-1}) - (\hat{f}_\mu \cdots + \hat{f}_L)\big]y + \big[(L\hat{f}_L + \cdots + \mu\hat{f}_\mu) - ((\mu-1)\hat{f}_{\mu-1} + \cdots + 1\hat{f}_1)\big], & \text{if } y \in (\mu-1, \mu] \\ \big[(\hat{f}_1 + \cdots + \hat{f}_\mu) - (\hat{f}_{\mu+1} \cdots + \hat{f}_L)\big]y + \big[(L\hat{f}_L + \cdots + (\mu+1)\hat{f}_{\mu+1}) - (\mu\hat{f}_\mu + \cdots + 1\hat{f}_1)\big], & \text{if } y \in (\mu, \mu+1] \\ \vdots & \vdots \\ \big[(\hat{f}_1 + \cdots + \hat{f}_{L-1}) - \hat{f}_L)\big]y + \big[(L\hat{f}_L - ((L-1)\hat{f}_{L-1} + \cdots + 1\hat{f}_1)\big], & \text{if } y \in (L-1, L] \\ \big[(\hat{f}_1 + \cdots + \hat{f}_L)\big]y + \big[(L\hat{f}_L + \cdots + 1\hat{f}_1)\big], & \text{if } y \in (L, \infty]. \end{cases}$$

Based on the above formula, we can find out that $R(y)$ is minimized when $y = \mu$ because $(\hat{f}_1 + \cdots + \hat{f}_{\mu-1}) - (\hat{f}_\mu \cdots + \hat{f}_L) < 0$ and $(\hat{f}_1 + \cdots + \hat{f}_\mu) - (\hat{f}_{\mu+1} \cdots + \hat{f}_L) > 0$ by the definition of μ. □

3 Theorems in the case when the cut points is unknown

When the cut points $\boldsymbol{b}$ is unknown, we estimate $(\hat{\Theta}, \hat{\boldsymbol{b}})$ by alternately fixing each factor and proceed until each factor does not change from each update. We have the following relationship between $\hat{\Theta}$ and $\hat{\boldsymbol{b}}$

$$\hat{\Theta} = \arg\max_{\Theta \in \mathcal{P}} \mathcal{L}_{\mathcal{Y},\Omega}(\Theta|\hat{\boldsymbol{b}}) \quad \text{and} \quad \hat{\boldsymbol{b}} = \arg\max_{\boldsymbol{b} \in \mathcal{B}} \mathcal{L}_{\mathcal{Y},\Omega}(\boldsymbol{b}|\hat{\Theta}).$$

We assess the estimation accuracy using the mean squared error (MSE):

$$\text{MSE}(\hat{\Theta}, \hat{\boldsymbol{b}}) = \frac{1}{\prod_k d_k} \|(\hat{\Theta}, \hat{\boldsymbol{b}}) - (\Theta^{\text{true}}, \boldsymbol{b}^{\text{true}})\|_F.$$

We show that $\frac{1}{\prod_k d_k}\|(\hat{\Theta}, \hat{\boldsymbol{b}}) - (\hat{\Theta}, \boldsymbol{b}^{\text{true}})\|_F$ is bounded by $o(\frac{1}{\prod_k d_k})$ with high probability so that

$$\frac{1}{\prod_k d_k}\|(\hat{\Theta}, \hat{\boldsymbol{b}}) - (\Theta^{\text{true}}, \boldsymbol{b}^{\text{true}})\|_F \leq \frac{1}{\prod_k d_k}\|\hat{\Theta} - \Theta^{\text{true}}\|_F + o(\frac{1}{\prod_k d_k}).$$

Therefore, we can utilize Theorem 4.1 to establish the upper bound for the proposed $(\hat{\Theta}, \hat{\boldsymbol{b}})$. We define a few key quantities that will be used in the proof.

$$\begin{aligned}
U_\alpha &= \max_{\ell \in [L-1], |\theta| \leq \alpha} \max\left(\frac{\dot{f}(\theta + b_\ell)}{f(\theta + b_\ell) - f(\theta + b_{\ell-1})}, \frac{\dot{f}(\theta + b_\ell)}{f(\theta + b_{\ell+1}) - f(\theta + b_\ell)}\right) \\
L_\alpha &= \min_{\ell \in [L-1], |\theta| \leq \alpha} \min\Bigg(-\frac{\ddot{f}(\theta + b_\ell)[f(\theta + b_\ell) - f(\theta + b_{\ell-1})] - \dot{f}(\theta + b_\ell)[\dot{f}(\theta + b_\ell) - \dot{f}(\theta + b_{\ell-1})]}{[f(\theta + b_\ell) - f(\theta + b_{\ell-1})]^2}, \\
&\qquad \frac{\ddot{f}(\theta + b_\ell)[f(\theta + b_{\ell+1}) - f(\theta + b_\ell)] - \dot{f}(\theta + b_\ell)[\dot{f}(\theta + b_{\ell+1}) - \dot{f}(\theta + b_\ell)]}{[f(\theta + b_{\ell+1}) - f(\theta + b_\ell)]^2}\Bigg)
\end{aligned}$$

We can see $U_\alpha > 0$ and $L_\alpha > 0$ from the assumption. Especially when $f(x) = \frac{1}{1+e^{-x}}$, we have

$$L_\alpha = \min_{\ell \in [L-1], |\Theta| \leq \alpha} \min\left(\frac{e^{\theta + b_\ell}(e^{\theta + b_\ell} - e^{\theta + b_{\ell-1}})^2}{(e^{\theta + b_\ell} + 1)^4 (e^{\theta + b_{\ell-1}} + 1)^2}, \frac{e^{\theta + b_\ell}(e^{\theta + b_{\ell+1}} - e^{\theta + b_\ell})^2}{(e^{\theta + b_{\ell+1}} + 1)^2 (e^{\theta + b_\ell} + 1)^4}\right).$$

Theorem 3.1. *From the same setting in Theorem 4.1 except the knowledge of $\boldsymbol{b}$, with very high probability, the estimator satisfies*

$$\frac{1}{\prod_k d_k}\|(\hat{\Theta}, \hat{\boldsymbol{b}}) - (\hat{\Theta}, \boldsymbol{b}^{true})\|_F = o(\frac{1}{\prod_k d_k}).$$

Proof. It follows from the expression of $\mathcal{L}_{\mathcal{Y},\Omega}(\Theta, \boldsymbol{b})$ that

$$\begin{aligned}
\frac{\partial \mathcal{L}_\mathcal{Y}}{\partial \beta_\ell} &= \sum_{\omega \in \Omega}\left[\mathbb{1}_{\{y_\omega = \ell\}} \frac{\dot{f}(\theta_\omega + b_\ell)}{f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})} - \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\dot{f}(\theta_\omega + b_\ell)}{f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)}\right], \\
\frac{\partial^2 \mathcal{L}_\mathcal{Y}}{\partial b_\ell^2} &= \sum_{\omega \in \Omega}\Bigg[\mathbb{1}_{\{y_\omega = \ell\}} \frac{\ddot{f}(\theta_\omega + b_\ell)[f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})] - \dot{f}(\theta_\omega + b_\ell)^2}{[f(\theta_\omega + b_\ell) - f(\theta_\omega + b_{\ell-1})]^2} \\
&\qquad - \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\ddot{f}(\theta_\omega + b_\ell)[f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)] + \dot{f}(\theta_\omega + b_\ell)^2}{[f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)]^2}\Bigg] \text{ for } \ell \in [L-1],
\end{aligned}$$

$$\frac{\partial^2 \mathcal{L}_\mathcal{Y}}{\partial b_\ell \partial b_{\ell+1}} = \sum_{\omega \in \Omega} \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\dot{f}(\theta_\omega + b_\ell)\dot{f}(\theta_\omega + b_{\ell+1})}{[f(\theta_\omega + b_{\ell+1}) - f(\theta_\omega + b_\ell)]^2} \text{ for } \ell \in [L-2] \text{ and } \frac{\partial^2 \mathcal{L}_\mathcal{Y}}{\partial b_i \partial b_j} = 0 \text{ if } |i-j| > 1.$$

Therefore, all entries in $\frac{1}{\prod_k d_k}\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}$ are upper bounded $U > 0$, and $\frac{1}{\prod_k d_k}\nabla^2_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}$ is a tridiagonal matrix.

By the second-order Taylor's expansion of $\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}|\hat{\Theta})$ around $\boldsymbol{b}^{\text{true}}$, we obtain

$$\mathcal{L}_{\mathcal{Y}}(\hat{\boldsymbol{b}}|\hat{\Theta}) = \mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}|\hat{\Theta}) + (\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}})^T\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}) + (\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}})^T\nabla^2_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\check{\boldsymbol{b}})(\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}}), \tag{1}$$

$\check{\boldsymbol{b}} = \gamma\boldsymbol{b}^{\text{true}} + (1-\gamma)\hat{\boldsymbol{b}}$ for some $\gamma \in [0,1]$, and $\nabla^2_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\check{\boldsymbol{b}})$ denotes the $(L-1)$-by-$(L-1)$ Hession matrix evaluated at $\check{\boldsymbol{b}}$.

We first bound the linear term in (1). Note that, by Cauchy-Schwartz inequality,

$$(\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}})^T\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}) \leq \|\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}}\|\|\nabla_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}})\| \leq \|\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}}\|U_\alpha\sqrt{L}\prod_k d_k.$$

The last inequality is followed by

$$\frac{\partial\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell}\Big|_{\boldsymbol{b}=\boldsymbol{b}^{\text{true}}} \leq U_\alpha \quad \text{for all} \;\; l \in [L-1].$$

We next bound the quadratic term in (1). Note that

$$\begin{aligned}
&(\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}}))^T\nabla^2_{\boldsymbol{b}}\mathcal{L}_{\mathcal{Y}}(\check{\boldsymbol{b}})(\boldsymbol{b}^{\text{true}} - \hat{\boldsymbol{b}})\\
&= \sum_{\ell\in[L-1]}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat{b}_\ell - b_{\text{true},\ell})^2 + 2\sum_{\ell\in[L-1]-\{1\}}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell\partial b_{\ell-1}}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat{b}_\ell - b_{\text{true},\ell})(\hat{b}_{\ell-1} - b_{\text{true},\ell-1})\\
&\leq \sum_{\ell\in[L-1]}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat{b}_\ell - b_{\text{true},\ell})^2 + \sum_{\ell\in[L-1]-\{1\}}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell\partial b_{\ell-1}}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)\left[(\hat{b}_\ell - b_{\text{true},\ell})^2 + (\hat{b}_{\ell-1} - b_{\text{true},\ell-1})^2\right]\\
&= \left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_1^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_1\partial b_2}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat{b}_1 - b_{\text{true},1})^2 + \left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{L-1}^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{L-2}\partial b_{L-1}}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat{b}_{L-1} - b_{\text{true},L-1})^2\\
&+ \sum_{\ell\in[L-2]-\{1\}}\left(\frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell^2} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_\ell\partial b_{\ell-1}} + \frac{\partial^2\mathcal{L}_{\mathcal{Y}}}{\partial b_{\ell+1}\partial b_\ell}\Big|_{\boldsymbol{b}=\check{\boldsymbol{b}}}\right)(\hat{b}_\ell - b_{\text{true},\ell})^2\\
&\leq -L_\alpha\prod_k d_k\sum_{\ell\in[L-1]}(\hat{b}_\ell - b_{\text{true},\ell})^2\\
&= -L_\alpha\prod_k d_k\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F^2,
\end{aligned}$$

Therefore, combining (1) and the above linear term and quadratic term results, we have that,

$$\mathcal{L}_{\mathcal{Y}}(\hat{\boldsymbol{b}}|\hat{\Theta}) \leq \mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}|\hat{\Theta}) + U_\alpha\sqrt{L}\prod_k d_k\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F - \frac{L_\alpha}{2}\prod_k d_k\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F^2.$$

Since $\hat{\boldsymbol{b}} = \arg\max_{\boldsymbol{b}\in\mathcal{B}}\mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}|\hat{\Theta})$, $\mathcal{L}_{\mathcal{Y}}(\hat{\boldsymbol{b}}|\hat{\Theta})) - \mathcal{L}_{\mathcal{Y}}(\boldsymbol{b}^{\text{true}}|\hat{\Theta}) \geq 0$, which gives

$$U_\alpha\sqrt{L}\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F - \frac{L_\alpha}{2}\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F^2 \geq 0.$$

Henceforth,

$$\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F \leq \frac{2U_\alpha\sqrt{L}}{L_\alpha}.$$

Finally, this complete the proof.

$$\frac{1}{\prod_k d_k}\|(\hat{\Theta}, \hat{\boldsymbol{b}}) - (\hat{\Theta}, \boldsymbol{b}^{\text{true}})\|_F = \frac{1}{\prod_k d_k}\|\hat{\boldsymbol{b}} - \boldsymbol{b}^{\text{true}}\|_F = o(\frac{1}{\prod_k d_k}).$$

.

□

References

Brascamp, H. J. and Lieb, E. H. On extensions of the brunn-minkowski and prékopa-leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. In *Inequalities*, pp. 441–464. Springer, 2002.