Information-Theoretic Demensionality Reduction for Poisson Models: Supplementary Material

1 Proof of Theorem 1

Proof of Theorem 1. We first establish the following Lemma.

Lemma 1. *Consider random variables $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$. Let $f^\theta_{Y|X}$ be the Radon-Nikodym derivatives of probability measure $P^\theta_{Y|X}$ with respect to arbitrary measures Q_Y provided that $P^\theta_{Y|X} \ll Q_Y$. $\theta \in \mathbb{R}$ is a parameter. f^θ_Y is the Radon-Nikodym derivatives of probability measure P^θ_Y with respect to Q_Y provided that $P^\theta_Y \ll Q_Y$. Note that in the continuous or discrete case, $f^\theta_{Y|X}$ and f^θ_Y are simply probability density or mass functions with Q_Y chosen to be the Lebesgue measure or the counting measure, respectively. We have*

$$\frac{\partial}{\partial\theta} I(X;Y) = \mathbb{E}\left[\frac{\partial \log f^\theta_{Y|X}}{\partial\theta} \log \frac{f^\theta_{Y|X}}{f^\theta_Y}\right]. \tag{1}$$

Proof of Lemma 1. Choose an arbitrary measure Q_Y such that $P^\theta_{Y|X} \ll Q_Y$ and $P^\theta_Y \ll Q_Y$.

$$\begin{aligned}
\frac{\partial}{\partial\theta} I(X;Y) &= \frac{\partial}{\partial\theta} D(P^\theta_{Y|X} \| Q_Y) - D(P^\theta_Y \| Q_Y) & (2)\\
&= \frac{\partial}{\partial\theta}\left[\mathbb{E}\left[\log \frac{P^\theta_{Y|X}}{Q_Y}\right] - \mathbb{E}\left[\log \frac{P^\theta_Y}{Q_Y}\right]\right] & (3)\\
&= \frac{\partial}{\partial\theta}\left[\int \log \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y} \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y} \mathrm{d}Q_Y \mathrm{d}P_X - \int \log \frac{\mathrm{d}P^\theta_Y}{\mathrm{d}Q_Y} \frac{\mathrm{d}P^\theta_Y}{\mathrm{d}Q_Y} \mathrm{d}Q_Y\right] & (4)\\
&= \frac{\partial}{\partial\theta}\left[\int \log \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y} \mathrm{d}P^\theta_{Y|X} \mathrm{d}P_X - \int \log \frac{\mathrm{d}P^\theta_Y}{\mathrm{d}Q_Y} \frac{\mathrm{d}P^\theta_Y}{\mathrm{d}Q_Y} \mathrm{d}Q_Y\right]. & (5)
\end{aligned}$$

We will calculate two terms in (5) separately.

$$\begin{aligned}
\frac{\partial}{\partial\theta}\left[\int \log \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y} \mathrm{d}P^\theta_{Y|X} \mathrm{d}P_X\right] &= \int \left[\frac{\partial}{\partial\theta}\left(\log \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y}\right) \mathrm{d}P^\theta_{Y|X} \mathrm{d}P_X\right] \\
&\quad + \int \left[\log \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y} \frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y}\right) \mathrm{d}Q_Y \mathrm{d}P_X\right]. & (6)
\end{aligned}$$

By Lemma 1 in [1], we have

$$\int \left[\frac{\partial}{\partial\theta}\left(\log \frac{\mathrm{d}P^\theta_{Y|X}}{\mathrm{d}Q_Y}\right) \mathrm{d}P^\theta_{Y|X} \mathrm{d}P_X\right] = 0. \tag{7}$$

Hence,

$$\frac{\partial}{\partial\theta}\left[\int \log \frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\mathrm{d}P^{\theta}_{Y|X}\mathrm{d}P_X\right] = \int \left[\log \frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}Q_Y\mathrm{d}P_X\right] \tag{8}$$

$$= \int \left[\log \frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\log\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}P^{\theta}_{Y|X}\mathrm{d}P_X\right]. \tag{9}$$

The second term in (5) can be calculated as follow.

$$\frac{\partial}{\partial\theta}\left[\int \log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\mathrm{d}Q_Y\right] = \int\left[\frac{\partial}{\partial\theta}\left(\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\right)\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\mathrm{d}Q_Y\right] + \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\right)\mathrm{d}Q_Y\right] \tag{10}$$

$$= \int\left[\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\right)\mathrm{d}Q_Y\right] + \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\right)\mathrm{d}Q_Y\right] \tag{11}$$

$$= \frac{\partial}{\partial\theta}\int \mathrm{d}P^{\theta}_{Y} + \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\right)\mathrm{d}Q_Y\right] \tag{12}$$

$$= 0 + \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\int\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\mathrm{d}P_X\right)\mathrm{d}Q_Y\right] \tag{13}$$

$$= \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\int\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}P_X\mathrm{d}Q_Y\right] \tag{14}$$

$$= \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}P_X\mathrm{d}Q_Y\right] \tag{15}$$

$$= \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\log\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}P^{\theta}_{Y|X}\mathrm{d}P_X\right]. \tag{16}$$

Plugging (9) and (16) back to (5), we have

$$\frac{\partial}{\partial\theta}I(X;Y) = \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\log\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}P^{\theta}_{Y|X}\mathrm{d}P_X\right] - \int\left[\log\frac{\mathrm{d}P^{\theta}_{Y}}{\mathrm{d}Q_Y}\frac{\partial}{\partial\theta}\left(\log\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\mathrm{d}P^{\theta}_{Y|X}\mathrm{d}P_X\right] \tag{17}$$

$$= \int\left[\frac{\partial}{\partial\theta}\left(\log\frac{\mathrm{d}P^{\theta}_{Y|X}}{\mathrm{d}Q_Y}\right)\log\frac{\mathrm{d}P^{\theta}_{Y|X}/\mathrm{d}Q_Y}{\mathrm{d}P^{\theta}_{Y}/\mathrm{d}Q_Y}\mathrm{d}P^{\theta}_{Y|X}\mathrm{d}P_X\right] \tag{18}$$

$$= \mathbb{E}\left[\frac{\partial\log f^{\theta}_{Y|X}}{\partial\theta}\log\frac{f^{\theta}_{Y|X}}{f^{\theta}_{Y}}\right]. \tag{19}$$

□

By Lemma 1, we have

$$\frac{\partial I(\boldsymbol{x};\boldsymbol{y})}{\partial\mathbf{M}_{ij}} = \mathbb{E}\left(\frac{\partial}{\partial\mathbf{M}_{ij}}\log p^{\Phi_{ij}}_{\boldsymbol{y}|\boldsymbol{x}}(y|x)\times\log\frac{p^{\Phi_{ij}}_{\boldsymbol{y}|\boldsymbol{x}}}{p^{\mathbf{M}_{ij}}_{\boldsymbol{y}}}\right) \tag{20}$$

$$= \mathbb{E}\left(\frac{\frac{\partial}{\partial\mathbf{M}_{ij}}p^{\mathbf{M}_{ij}}_{\boldsymbol{y}|\boldsymbol{x}}(y|x)}{p^{\mathbf{M}_{ij}}_{\boldsymbol{y}|\boldsymbol{x}}(y|x)}\times\log\frac{p^{\mathbf{M}_{ij}}_{\boldsymbol{y}|\boldsymbol{x}}}{p^{\mathbf{M}_{ij}}_{\boldsymbol{y}}}\right). \tag{21}$$

Given the Poisson model, we can get that

$$\frac{\partial}{\partial \mathbf{M}_{ij}} p_{\boldsymbol{y}|\boldsymbol{x}}^{\mathbf{M}_{ij}} = \frac{\partial}{\partial \mathbf{M}_{ij}} \mathrm{Pois}(\mathbf{M}\boldsymbol{x}) \tag{22}$$

$$= \left(\frac{1}{\boldsymbol{y}_i!} \boldsymbol{y}_i \boldsymbol{x}_j \times (\boldsymbol{m}_i \boldsymbol{x})^{\boldsymbol{y}_i - 1} \times e^{-(\boldsymbol{m}_i \boldsymbol{x})} + \frac{1}{\boldsymbol{y}_i!} (\boldsymbol{m}_i \boldsymbol{x})^{\boldsymbol{y}_i} (-\boldsymbol{x}_j) e^{-(\boldsymbol{m}_i \boldsymbol{x})} \right) \times \prod_{k \neq i} \frac{1}{\boldsymbol{y}_k!} (\boldsymbol{m}_i \boldsymbol{x})^{\boldsymbol{y}_k} e^{-(\boldsymbol{m}_i \boldsymbol{x})} \tag{23}$$

$$= \frac{1}{\boldsymbol{y}_i!} \boldsymbol{x}_j \times (\boldsymbol{m}_i \boldsymbol{x})^{\boldsymbol{y}_i} e^{-(\boldsymbol{m}_i \boldsymbol{x})} \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \prod_{k \neq i} \frac{1}{\boldsymbol{y}_k!} (\boldsymbol{m}_i \boldsymbol{x})^{\boldsymbol{y}_k} e^{-(\boldsymbol{m}_i \boldsymbol{x})} \tag{24}$$

$$= \boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times p_{\boldsymbol{y}|\boldsymbol{x}}^{\boldsymbol{m}_{ij}}, \tag{25}$$

where $\boldsymbol{m}_i$ is the i-th row of $\mathbf{M}$.

Therefore, we have

$$\frac{\partial I(\boldsymbol{x}; \boldsymbol{y})}{\partial \mathbf{M}_{ij}} = \mathbb{E} \left(\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \log \frac{p_{\boldsymbol{y}|\boldsymbol{x}}^{\mathbf{M}_{ij}}}{p_{\boldsymbol{y}}^{\mathbf{M}_{ij}}} \right) \tag{26}$$

$$= \mathbb{E} \left(\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \log p_{\boldsymbol{y}|\boldsymbol{x}}^{\mathbf{M}_{ij}} \right) \tag{27}$$

$$- \mathbb{E} \left(\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \log p_{\boldsymbol{y}}^{\mathbf{M}_{ij}} \right). \tag{28}$$

We will calculate (27) and (28) separately.

$$(27) = \mathbb{E} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \sum_k \log \left(\frac{1}{\boldsymbol{y}_k!} (\boldsymbol{m}_i \boldsymbol{x})^{\boldsymbol{y}_k} e^{-(\boldsymbol{m}_i \boldsymbol{x})} \right) \right] \tag{29}$$

$$= \sum_k \mathbb{E} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \log \frac{1}{\boldsymbol{y}_k!} \right] \tag{30}$$

$$+ \sum_k \mathbb{E} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \boldsymbol{y}_k \times \log(\boldsymbol{m}_i \boldsymbol{x}) \right] \tag{31}$$

$$- \sum_k \mathbb{E} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times (\boldsymbol{m}_i \boldsymbol{x}) \right]. \tag{32}$$

Claim (32)=0. Since

$$(32) = \sum_k \mathbb{E}_{\boldsymbol{x}} \left[\mathbb{E}_{\boldsymbol{y}|\boldsymbol{x}} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times (\boldsymbol{m}_i \boldsymbol{x}) \right] \right] \tag{33}$$

$$= \sum_k \mathbb{E}_{\boldsymbol{x}} \left[\boldsymbol{x}_j \left(\frac{\mathbb{E}_{\boldsymbol{y}|\boldsymbol{x}}[\boldsymbol{y}_i]}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times (\boldsymbol{m}_i \boldsymbol{x}) \right] \tag{34}$$

$$= \sum_k \mathbb{E}_{\boldsymbol{x}} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{m}_i \boldsymbol{x}}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times (\boldsymbol{m}_i \boldsymbol{x}) \right] \tag{35}$$

$$= 0, \tag{36}$$

we have

$$(27) = \sum_k \mathbb{E} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \log \frac{1}{\boldsymbol{y}_k!} \right] + \sum_k \mathbb{E} \left[\boldsymbol{x}_j \left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i \boldsymbol{x}} - 1 \right) \times \boldsymbol{y}_k \times \log(\boldsymbol{m}_i \boldsymbol{x}) \right]. \tag{37}$$

Combining the fact that $\mathbb{E}[\boldsymbol{y}_i|\boldsymbol{x}] = \boldsymbol{m}_i\boldsymbol{x}$ and $\mathrm{var}[\boldsymbol{y}_i|\boldsymbol{x}] = \boldsymbol{m}_i\boldsymbol{x}$, the latter term can be calculated as follow.

$$
\begin{aligned}
&\sum_k \mathbb{E}\left[\boldsymbol{x}_j\left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i\boldsymbol{x}} - 1\right) \times \boldsymbol{y}_k \times \log(\boldsymbol{m}_i\boldsymbol{x})\right] \\
&= \sum_k \int p_{\boldsymbol{x}}(x) \times \prod_l p_{\boldsymbol{y}_l|\boldsymbol{x}}(y_l|x) \times x_j \left(\frac{y_i}{\phi_i x} - 1\right) y_k \log(\boldsymbol{m}_i x) && (38) \\
&= \int p_{\boldsymbol{x}}(x) \times p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x) \times x_j \left(\frac{y_i^2}{\boldsymbol{m}_i x} - y_i\right) \log(\boldsymbol{m}_i x) \\
&+ \sum_{k \neq i} \int p_{\boldsymbol{x}}(x) \times p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x) p_{\boldsymbol{y}_k|\boldsymbol{x}}(y_k|x) \times x_j \left(\frac{y_i}{\boldsymbol{m}_i x} - 1\right) y_k \log(\boldsymbol{m}_i x) && (39) \\
&= \int p_{\boldsymbol{x}}(x) \times p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x) \times x_j \left(\frac{(\boldsymbol{m}_i x)^2 + (\boldsymbol{m}_i x)}{\boldsymbol{m}_i x} - \boldsymbol{m}_i x\right) \log(\boldsymbol{m}_i x) \\
&+ \sum_{k \neq i} \int p_{\boldsymbol{x}}(x) \times p_{\boldsymbol{y}_k|\boldsymbol{x}}(y_k|x) \times x_j \left(\frac{\boldsymbol{m}_i x}{\boldsymbol{m}_i x} - 1\right) y_k \log(\boldsymbol{m}_i x) && (40) \\
&= \mathbb{E}[\boldsymbol{x}_j \log(\boldsymbol{m}_i \boldsymbol{x})] + 0. && (41)
\end{aligned}
$$

The following two Lemmas that will be useful can be established.

Lemma 2.

$$
\mathbb{E}\left[\frac{\boldsymbol{x}_j}{\boldsymbol{m}_i\boldsymbol{x}} \middle| \boldsymbol{y} = y\right] = \frac{1}{y_i}\frac{p_{\boldsymbol{y}}(y_i - 1, y_i^c)}{p_{\boldsymbol{y}}(y)}. \tag{42}
$$

Proof of Lemma 2.

$$
\begin{aligned}
\mathbb{E}\left[\frac{\boldsymbol{x}_j}{\boldsymbol{m}_i\boldsymbol{x}} \middle| \boldsymbol{y} = y\right] &= \mathbb{E}\left[\frac{1}{y_i} \times \frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i - 1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)} \middle| \boldsymbol{y} = y\right] && (43) \\
&= \frac{1}{y_i}\int_x \frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i - 1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)} p_{\boldsymbol{x}|\boldsymbol{y}}(x|y) && (44) \\
&= \frac{1}{y_i}\int_x \frac{p_{\boldsymbol{y}_i-1|\boldsymbol{x}}(y_i - 1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)} p_{\boldsymbol{x}|\boldsymbol{y}}(x|y) && (45) \\
&= \frac{1}{y_i}\int_x \frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i - 1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)} \frac{p_{\boldsymbol{y}|\boldsymbol{x}}(y|x)}{P_{\boldsymbol{y}}(y)} p_{\boldsymbol{x}}(x) && (46) \\
&= \frac{1}{y_i}\frac{1}{P_{\boldsymbol{y}}(y)}\int_x \frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i - 1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)} \prod_k p_{\boldsymbol{y}_k|\boldsymbol{x}}(y_k|x) \times p_{\boldsymbol{x}}(x) && (47) \\
&= \frac{1}{y_i}\frac{1}{P_{\boldsymbol{y}}(y)}\int_x p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i - 1|x) \prod_{k \neq i} p_{\boldsymbol{y}_k|\boldsymbol{x}}(y_k|x) \times p_{\boldsymbol{x}}(x) && (48) \\
&= \frac{1}{y_i}\frac{P_{\boldsymbol{y}}(y_i - 1, y_i^c)}{P_{\boldsymbol{y}}(y)}. && (49)
\end{aligned}
$$

□

Lemma 3.

$$
\mathbb{E}(\boldsymbol{m}_i\boldsymbol{x}|\boldsymbol{y} = y) = (y_i + 1)\frac{P_{\boldsymbol{y}}(y_i + 1, y_i^c)}{P_{\boldsymbol{y}}(y)}. \tag{50}
$$

Proof of Lemma 3.

$$\mathbb{E}(\boldsymbol{m}_i\boldsymbol{x}|\boldsymbol{y}=y) = (y_i+1)\mathbb{E}\left[\frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i+1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)}\middle|\,\boldsymbol{y}=y\right] \tag{51}$$

$$= (y_i+1)\int_x \frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i+1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|\boldsymbol{x})} \times p_{\boldsymbol{x}|\boldsymbol{y}}(x|y) \tag{52}$$

$$= \frac{y_i+1}{P_{\boldsymbol{y}}(y)}\int_x \frac{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i+1|x)}{p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i|x)} \times p_{\boldsymbol{y}|\boldsymbol{x}}(y|x) \times p_{\boldsymbol{x}}(x) \tag{53}$$

$$= \frac{y_i+1}{P_{\boldsymbol{y}}(y)}\int_x p_{\boldsymbol{y}_i|\boldsymbol{x}}(y_i+1|x)\prod_{k\neq i} p_{\boldsymbol{y}_k|\boldsymbol{x}}(y_k|x) \times p_{\boldsymbol{x}}(x) \tag{54}$$

$$= (y_i+1)\frac{P_{\boldsymbol{y}}(y_i+1, y_i^c)}{P_{\boldsymbol{y}}(y)} \tag{55}$$

□

Combing previous derivations, we get

$$\frac{\partial I(\boldsymbol{x};\boldsymbol{y})}{\partial \mathbf{M}_{ij}} = \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) - \mathbb{E}\left\{\boldsymbol{x}_j\left(\frac{\boldsymbol{y}_i}{\boldsymbol{m}_i\boldsymbol{x}} - 1\right)\log\left((\prod_k \boldsymbol{y}_k!)p_{\boldsymbol{y}}\right)\right\}$$

$$= \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) - \mathbb{E}\left\{\left(\mathbb{E}\left(\frac{\boldsymbol{x}_j}{\boldsymbol{m}_i\boldsymbol{x}}|\boldsymbol{y}\right)\boldsymbol{y}_i - \mathbb{E}(\boldsymbol{x}_j|\boldsymbol{y})\right)\log\left((\prod_k \boldsymbol{y}_k!)p_{\boldsymbol{y}}(y)\right)\right\} \tag{56}$$

$$\begin{aligned} &= \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) \\ &- \mathbb{E}\left\{\frac{p_{\boldsymbol{y}}(y_i-1, y_i^c)}{p_{\boldsymbol{y}}(y)} \times \mathbb{E}\left[\boldsymbol{x}_j|\boldsymbol{y}=(y_i-1,y_i^c)\right] \times \log\left(((\boldsymbol{y}_i+1)!\prod_{k\neq i}\boldsymbol{y}_k!)p_{\boldsymbol{y}}(y_i+1,y_i^c)\right)\right\} \\ &+ \mathbb{E}\left\{(\mathbb{E}(\boldsymbol{x}_j|\boldsymbol{y}))\log\left((\prod_k \boldsymbol{y}_k!)p_{\boldsymbol{y}}(y)\right)\right\} \end{aligned} \tag{57}$$

$$\begin{aligned} &= \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) \\ &- \int_{\boldsymbol{y}}\left\{p_{\boldsymbol{y}}(y_i-1, y_i^c) \times \mathbb{E}\left[\boldsymbol{x}_j|\boldsymbol{y}=(y_i-1,y_i^c)\right] \times \log\left(((\boldsymbol{y}_i+1)!\prod_{k\neq i}\boldsymbol{y}_k!)p_{\boldsymbol{y}}(y_i+1,y_i^c)\right)\right\} \\ &+ \mathbb{E}\left\{(\mathbb{E}(\boldsymbol{x}_j|\boldsymbol{y}))\log\left((\prod_k \boldsymbol{y}_k!)p_{\boldsymbol{y}}(y)\right)\right\} \end{aligned} \tag{58}$$

$$\begin{aligned} &= \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) \\ &- \mathbb{E}\left\{\mathbb{E}\left[\boldsymbol{x}_j|\boldsymbol{y}\right] \times \log\left(((\boldsymbol{y}_i+1)!\prod_{k\neq i}\boldsymbol{y}_k!)p_{\boldsymbol{y}}(y_i+1,y_i^c)\right)\right\} \\ &+ \mathbb{E}\left\{(\mathbb{E}(\boldsymbol{x}_j|\boldsymbol{y}))\log\left((\prod_k \boldsymbol{y}_k!)p_{\boldsymbol{y}}(y)\right)\right\} \end{aligned} \tag{59}$$

$$= \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) - \mathbb{E}\left\{\mathbb{E}(\boldsymbol{x}_j|\boldsymbol{y})\log(y_i+1)\frac{p_{\boldsymbol{y}}(y_i+1,y_i^c)}{p_{\boldsymbol{y}}(y)}\right\} \tag{60}$$

$$= \mathbb{E}\left(\boldsymbol{x}_j \log \boldsymbol{m}_i\boldsymbol{x}\right) - \mathbb{E}[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y}]\log(\mathbb{E}[\boldsymbol{m}_i\boldsymbol{x}|\boldsymbol{y}])], \tag{61}$$

where (57) follows from Lemma (2) and (3). (61) follows from Lemma (3).

Hence, we have

$$\nabla_{\mathbf{M}} I(\boldsymbol{x};\boldsymbol{y})_{ij} = \mathbb{E}\left[\boldsymbol{x}_j \log((\mathbf{M}\boldsymbol{x})_i)\right], \\ - \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y}] \log \mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y}]\right]. \quad (62)$$

□

2 Proof of Theorem 2

Proof. First we notice that

$$I(c;\boldsymbol{y}) = H(\boldsymbol{y}) - H(\boldsymbol{y}|c) \quad (63)$$
$$= H(\boldsymbol{y}) - H(\boldsymbol{y}|\boldsymbol{x}) + H(\boldsymbol{y}|\boldsymbol{x},c) - H(\boldsymbol{y}|c) \quad (64)$$
$$= I(\boldsymbol{x};\boldsymbol{y}) - I(\boldsymbol{x};\boldsymbol{y}|c). \quad (65)$$

Using the fact that $c \to \boldsymbol{x} \to \boldsymbol{y}$ forms a Markov chain and $p_{\boldsymbol{y}|\boldsymbol{x},c} = p_{\boldsymbol{y}|\boldsymbol{x}}$, followed by the similar steps in the proof of Theorem 1, we have

$$[\nabla_{\mathbf{M}} I(\boldsymbol{x};\boldsymbol{y}|c)_{ij}] = \left[\mathbb{E}\left[\boldsymbol{x}_j \log((\mathbf{M}\boldsymbol{x})_i)\right] - \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y}] \log \mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y},c]\right]\right].$$

Hence,

$$[\nabla_{\mathbf{M}} I(\boldsymbol{x};c)_{ij}] = -\mathbb{E}\left[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y}] \log \mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y}]\right] + \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y}] \log \mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y},c]\right] \quad (66)$$
$$= \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y},c] \log \frac{\mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y},c]}{\mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y}]}\right] \quad (67)$$

□

3 Proof of Theorem 3

Proof. We first show the Poisson case. Notice that $DF = (\log(\mathbf{M}\boldsymbol{x}))^T \otimes I$ under the standard basis. Hence,

$$\mathbb{E}\left[D_F(\boldsymbol{x}, \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])\right] = \mathbb{E}\left[\boldsymbol{x}(\log \mathbf{M}\boldsymbol{x})^T - G(\boldsymbol{x}) - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T + G(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])\right] \\ - \mathbb{E}\left[\boldsymbol{x}(\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T\right]. \quad (68)$$

$G(\boldsymbol{x})$ is a linear in $\boldsymbol{x}$, thus

$$\mathbb{E}[G(\boldsymbol{x})] = \mathbb{E}[G(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])]. \quad (69)$$

Therefore,

$$\mathbb{E}\left[D_F(\boldsymbol{x}, \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])\right] = \mathbb{E}\left[\boldsymbol{x}(\log \mathbf{M}\boldsymbol{x})^T - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T\right] \\ - \mathbb{E}\left[\boldsymbol{x}(\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T\right] \quad (70)$$
$$= \mathbb{E}\left[\boldsymbol{x}(\log \mathbf{M}\boldsymbol{x})^T - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T\right] \\ - \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T + \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T\right] \quad (71)$$
$$= \mathbb{E}\left[\boldsymbol{x}(\log \mathbf{M}\boldsymbol{x})^T - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\log \mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T\right] \quad (72)$$
$$= \nabla_{\mathbf{M}} I(\boldsymbol{x};\boldsymbol{y}) \quad (73)$$

For the Gaussian case, notice that $DF = 2I \otimes \boldsymbol{x}$. Hence,

$$\mathbb{E}\left[D_F(\boldsymbol{x}, \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])\right] = \mathbb{E}\left[\boldsymbol{\Sigma}\mathbf{M}\boldsymbol{x}\boldsymbol{x}^T - \boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]^T - 2\boldsymbol{\Sigma}\mathbf{M}\boldsymbol{x}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]^T + 2\boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]^T\right] \quad (74)$$
$$= \boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])(\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T] \quad (75)$$
$$= \nabla_{\mathbf{M}} I(\boldsymbol{x};\boldsymbol{y}). \quad (76)$$

The last equality follows from the gradient result for Gaussian classification problem in [2]. □

4 Proof of Theorem 4

Proof. We first show the Poisson case. Notice that

$$DF = (I \otimes \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])\nabla(\log(\mathbf{M}\boldsymbol{x}))^T - (I \otimes \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])\nabla(\log(\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]))^T \tag{77}$$

under the standard basis. Therefore, $DF(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]) = \mathbf{0}$. We have

$$\begin{aligned}\mathbb{E}\left[D_F(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}], \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])\right] = \mathbb{E}&\left[\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c](\log(\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]))^T - L(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])\right] \\ &- \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c](\log(\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]))^T - L(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])\right],\end{aligned} \tag{78}$$

where $L(\boldsymbol{x}) := (I \otimes \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])\nabla(\log(\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]))^T\boldsymbol{x}$. $L(\boldsymbol{x})$ is a linear in $\boldsymbol{x}$, thus

$$\mathbb{E}[L(\boldsymbol{x})] = \mathbb{E}[L(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])]. \tag{79}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[D_F(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c], \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])\right] &= \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c](\log(\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]))^T - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c](\log(\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]))^T\right] &(80)\\ &= \mathbb{E}\left[\mathbb{E}[\boldsymbol{x}_j|\boldsymbol{y}, c]\log\frac{\mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y}, c]}{\mathbb{E}[(\mathbf{M}\boldsymbol{x})_i|\boldsymbol{y}]}\right] &(81)\\ &= \nabla_{\mathbf{M}} I(c; \boldsymbol{y}). &(82)\end{aligned}$$

For the Gaussian case, notice that $DF = 2I \otimes \boldsymbol{x}$ under the standard basis. Hence,

$$\begin{aligned}\mathbb{E}\left[D_F(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}], \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])\right] &= \mathbb{E}\left[\boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}](\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}])^T - \boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]^T\right] \\ &\quad - \mathbb{E}\left[2\boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}]^T - 2\boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c]^T\right] &(83)\\ &= \boldsymbol{\Sigma}\mathbf{M}\mathbb{E}[(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}] - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])(\mathbb{E}[\boldsymbol{x}|\boldsymbol{y}] - \mathbb{E}[\boldsymbol{x}|\boldsymbol{y}, c])^T] &(84)\\ &= \nabla_{\mathbf{M}} I(\boldsymbol{x}; \boldsymbol{y}). &(85)\end{aligned}$$

The last equality follows from the result in [3]. □

5 Variational Bayesian Updates for Topic Models

Given the model, $\boldsymbol{y}_d \sim \text{Pois}(\mathbf{F}\boldsymbol{x}_d)$ with each column of $\mathbf{F}$, $\boldsymbol{f}_k$, drawn from a $\text{Dir}(\eta, ..., \eta)$ and each entry $x_{dk} \sim \text{Gamma}(\alpha_0, \beta_0)$. We let $\boldsymbol{y}_d \in \mathbb{Z}^n$, $\boldsymbol{x}_d \in \mathbb{R}_+^K$ and $\mathbf{F} \in \mathbb{R}_+^{n\times K}$. We use Variational Bayesian updates to estimate the posterior distribution q:

$$\begin{aligned}q(\boldsymbol{\phi}_{dj}) &\sim \text{Multi}(y_{dj;\boldsymbol{\pi}_{dj}}) &(86)\\ \pi_{djk} &\propto \exp\left(\psi(\gamma_{dk}) - \log(\beta_0 + 1) + \psi(\zeta_{kj}) - \psi(\sum_{i=1}^n \zeta_{ki})\right) &(87)\\ q(x_{dk}) &\sim \text{Gamma}(\gamma_{dk}, \beta_0 + 1) &(88)\\ \gamma_{dk} &= \alpha_0 + \sum_{j=1}^n y_{dj}\pi_{djk} &(89)\\ q(\boldsymbol{f}_k) &\sim \text{Dir}(\boldsymbol{\zeta}_k) &(90)\\ \zeta_{kj} &= \eta + \sum_d y_{dj}\pi_{djk} &(91)\end{aligned}$$

where ψ represents the digamma function.

When we consider the compressive measurements $\boldsymbol{y}_d|\mathbf{F} \sim \text{Pois}(\mathbf{M}\mathbf{F}\boldsymbol{x}_d)$ where each $x_{dk} \sim \text{Gamma}(\alpha_0, \beta_0)$.

In this case, we let $\mathbf{\Phi} = \mathbf{MF}$ and $\mathbf{\Phi} \in \mathbb{R}_+^{m\times K}$, and we have $\boldsymbol{y}_d \in \mathbb{Z}^m$ and $\boldsymbol{x}_d \in \mathbb{K}$. We use Variational Bayeisans to estimate the posterior distribution q:

$$\begin{aligned}
q(\boldsymbol{\phi}_{dj}) &\sim \text{Multi}(y_{dj;\boldsymbol{\pi}_{dj}}) && (92)\\
\pi_{djk} &\propto \Phi_{jk} \exp\left(\psi(\gamma_{dk}) - \log(\beta'_{dk})\right)) && (93)\\
q(x_{dk}) &\sim \text{Gamma}(\gamma_{dk}, \beta'_{dk}) && (94)\\
\gamma_{dk} &= \alpha_0 + \sum_{j=1}^{n} y_{dj}\pi_{djk} && (95)\\
\beta'_{dk} &= \beta_0 + \sum_{j=1}^{m} \Phi_{kj} && (96)
\end{aligned}$$

References

[1] D.P. Palomar and S. Verdú. Representation of mutual information via input estimates. *IEEE Transactions on Information Theory*, 53(2):453–470, Feb. 2007.

[2] D.P. Palomar and S. Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Transactions on Information Theory,*, 52(1):141–154, Jan. 2006.

[3] M. Chen, W. Carson, M. Rodrigues, R. Calderbank, and L. Carin. Communications inspired linear discriminant analysis. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2012.