Lemma 1 Let \( p(x|\theta) = \theta_x \) be the probability mass function of the multinomial Bernoulli model with alphabet \( \mathcal{X} \). Using \( p(x|\theta) \), define a mixture model \( q(y|\theta) \) as

\[
q(y|\theta) = \sum_x \kappa(y|x)p(x|\theta),
\]

where \( \kappa(y|x) \) is a conditional probability density function of \( y \) given \( x \). Let \( \hat{\theta} \) be the maximum likelihood estimate of \( \theta \) for \( q(y|\theta) \) given \( y^n \), that is,

\[
q(y^n|\hat{\theta}) = \max_\theta q(y^n|\theta).
\]

Then, we have

\[
\forall y^n, \forall \theta, \quad \frac{q(y^n|\theta)}{q(y^n|\hat{\theta})} \geq \prod_{x \in \mathcal{X}} \left( \frac{\theta_x}{\hat{\theta}_x} \right)^{n\hat{\theta}_x}
\]

for all \( y^n \) and \( \theta \), where \( D \) denotes the Kullback-Leibler divergence.

Proof: Note that

\[
q(y^n|\theta) = \prod_{t=1}^{n} \sum_{x_t \in \mathcal{X}} \kappa(y_t|x_t)p(x_t|\theta)
\]

\[
= \sum_{x^n \in \mathcal{X}^n} \prod_{t} \kappa(y_t|x_t)p(x_t|\theta) = \sum_{x^n \in \mathcal{X}^n} \kappa(y^n|x^n)p(x^n|\theta).
\]

For a fixed \( \theta' \), define a subset \( G = G(\theta') \) of \( \mathcal{X}^n \) as

\[
G(\theta') = \{ x^n | p(x^n|\theta') \neq 0 \}.
\]
Note that $G = \mathcal{X}^n$, if $\theta'_x > 0$ for all $x \in \mathcal{X}$. We have
\[
\frac{q(y^n|\theta)}{q(y^n|\theta')} = \sum_{x^n \in \mathcal{X}^n} \frac{\kappa(y^n|x^n)p(x^n|\theta)}{\sum_{x^n \in \mathcal{X}^n} \kappa(y^n|x^n)p(x^n|\theta')} \geq \sum_{x^n \in G} \frac{p(x^n|\theta)}{p(x^n|\theta')} \sum_{z^n \in \mathcal{X}^n} \kappa(y^n|z^n)p(z^n|\theta').
\]
Here, equality holds if $G = \mathcal{X}^n$ or $\theta = \theta'$. Define $q(x^n|y^n, \theta')$ by
\[
q(x^n|y^n, \theta') = \frac{\kappa(y^n|x^n)p(x^n|\theta')}{\sum_{z^n \in \mathcal{X}^n} \kappa(y^n|z^n)p(z^n|\theta')},
\]
which is the posterior distribution of $x^n$ given $y^n$ provided $y^n$ is drawn from $q(y^n|\theta')$.
Note that the support of $q(x^n|y^n, \theta')$ is included in $G$.

Using it, we can write
\[
\frac{q(y^n|\theta)}{q(y^n|\theta')} \geq \sum_{x^n \in G} q(x^n|y^n, \theta') \frac{p(x^n|\theta)}{p(x^n|\theta')}.
\]
Then by Jensen’s inequality, we have
\[
\log \frac{q(y^n|\theta)}{q(y^n|\theta')} \geq \sum_{x^n \in G} q(x^n|y^n, \theta') \log \frac{p(x^n|\theta)}{p(x^n|\theta')} = \sum_{x^n \in \mathcal{X}^n} q(x^n|y^n, \theta') \log \frac{p(x^n|\theta)}{p(x^n|\theta')}.
\]
Letting $\theta' = \hat{\theta}$, we have
\[
0 \geq \log \frac{q(y^n|\theta)}{q(y^n|\theta')} \geq \sum_{x^n \in \mathcal{X}^n} q(x^n|y^n, \hat{\theta}) \log \frac{p(x^n|\theta)}{p(x^n|\hat{\theta})}.
\]
Both inequalities hold as equality when $\theta = \hat{\theta}$. It implies that the third side is maximized when $\theta = \hat{\theta}$. Hence the third side must be
\[
\sum_x n\hat{\theta}_x \log \frac{\theta_x}{\hat{\theta}_x}.
\]
This shows (1).

Taking expectation of both side of (3) with respect to $q(y^n|\theta')$, we have
\[
-nD(q(\cdot|\theta')|q(\cdot|\theta)) \geq \sum_{x^n \in \mathcal{X}^n} p(x^n|\theta') \log \frac{p(x^n|\theta)}{p(x^n|\theta')} = -nD(p(\cdot|\theta')|p(\cdot|\theta)).
\]
This is (2) and completes the proof of Lemma.

Let $\hat{J}(\theta)$ denote the empirical Fisher information of $q(y^n|\theta)$ given $y^n$ and $J_B(\theta)$ the Fisher information of $p(x^n|\theta)$. By (1) we have
\[
\log \frac{q(y^n|\theta)}{q(y^n|\theta')} \geq \log \prod_{x \in \mathcal{X}} \left( \frac{\theta_x}{\hat{\theta}_x} \right)^{n\theta_x}.
\]
Twice differentiating both sides, we have
\[ \hat{J}(\hat{\theta}) \leq J_B(\hat{\theta}). \]

Similarly by (2), we have
\[ J(\theta) \leq J_B(\theta), \]

where \( J(\theta) \) is the Fisher information of \( q(y|\theta) \).