Some Inequality for Models with Hidden Variables

Jun'ichi Takeuchi and Andrew R. Barron

January 19th, 2014

The following is some extension of the inequalities given in [1].

Let $p(y|\theta) = \exp(\theta^t y - \psi(\theta))$ be a probability density function of an exponential family, where y is a random variable over $\mathcal{Y} \subseteq \Re^k$ and θ is the natural parameter. Let Θ denote the range of θ.

Using $p(y|\theta)$, define a model with hidden variable $q(x|\theta)$ as

$$q(x|\theta) = \int \kappa(x|y) p(y|\theta) dy$$

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

$$q(x^n|\hat{\theta}) = \max_{\theta} q(x^n|\theta).$$

Let $\hat{J}(\theta, x^n)$ denote the empirical Fisher information of θ for $q(x^n|\theta)$:

$$\hat{J}_{ij}(\theta, x^n) = \frac{-1}{n} \frac{\partial^2 \log q(x^n|\theta)}{\partial\theta_i \partial\theta_j}.$$

Let $I(\theta)$ denote the Fisher information of θ for $p(x|\theta)$:

$$I_{ij}(\theta) = -E_{\theta} \frac{\partial^2 \log p(x|\theta)}{\partial\theta_i \partial\theta_j} = \frac{\partial^2 \psi(\theta)}{\partial\theta_i \partial\theta_j}.$$

Lemma 1 The following holds.

$$\forall x^n,\ \forall \theta,\ \ \frac{1}{n} \log \frac{q(x^n|\hat{\theta})}{q(x^n|\theta)} \le D(p(\cdot|\hat{\theta})|p(\cdot|\theta)) \tag{1}$$

and

$$\forall \theta \in \Theta,\ \ \hat{J}(\theta, x^n) \le I(\theta) \tag{2}$$

where $D(p(\cdot|\hat{\theta})|p(\cdot|\theta))$ denotes the Kullback-Leibler divergence from $p(\cdot|\hat{\theta})$ to $p(\cdot|\theta)$.

In particular, when $p(y|\theta)$ is the Bernoulli model, the following holds

$$\frac{1}{n}\log\frac{q(x^n|\hat{\theta})}{q(x^n|\theta)} \leq D(p(\cdot|\hat{\theta})|p(\cdot|\theta)) = \log\prod_{y\in\mathcal{Y}}\frac{\hat{\eta}_y^{n\hat{\eta}_y}}{\eta_y^{n\hat{\eta}_y}},$$

where $\eta_y = p(y|\theta)$ and $\hat{\eta}_y = p(y|\hat{\theta})$.

Proof: Note that

$$\begin{aligned} q(x^n|\theta) &= \prod_{t=1}^n \int \kappa(x_t|y_t)p(y_t|\theta)dy_t \\ &= \int\prod_t \kappa(x_t|y_t)p(y_t|\theta)dy^n = \int \kappa(x^n|y^n)p(y^n|\theta)dy^n. \end{aligned}$$

We have

$$\frac{q(x^n|\theta)}{q(x^n|\theta')} = \frac{\int\kappa(x^n|y^n)p(y^n|\theta)dy^n}{\int\kappa(x^n|y^n)p(y^n|\theta')dy^n} = \int\frac{p(y^n|\theta)}{p(y^n|\theta')}\frac{\kappa(x^n|y^n)p(y^n|\theta')}{\int\kappa(x^n|z^n)p(z^n|\theta')dz^n}dy^n.$$

Define $q(y^n|x^n,\theta')$ by

$$q(y^n|x^n,\theta') = \frac{\kappa(x^n|y^n)p(y^n|\theta')}{\int\kappa(x^n|z^n)p(z^n|\theta')dz^n},$$

which is the posterior distribution of y^n given x^n provided x^n is drawn from $q(x^n|\theta')$.

Using it, we can write

$$\frac{q(x^n|\theta)}{q(x^n|\theta')} = \int q(y^n|x^n,\theta')\frac{p(y^n|\theta)}{p(y^n|\theta')}dy^n.$$

Then by Jensen's inequality, we have

$$\frac{1}{n}\log\frac{q(x^n|\theta)}{q(x^n|\theta')} \geq \frac{1}{n}\int q(y^n|x^n,\theta')\log\frac{p(y^n|\theta)}{p(y^n|\theta')}dy^n. \tag{3}$$

Let $f(\theta,\theta')$ denote the left side, and $g(\theta,\theta')$ the right side. Then, we have

$$\forall\theta,\theta'\in\Theta,\ \ f(\theta,\theta')-g(\theta,\theta')\geq 0, \tag{4}$$

where equality holds when $\theta=\theta'$. Hence, Hessian of the left side is semi positive-definite. That is, the matrix whose ij entry is

$$\frac{\partial^2\log f(\theta,\theta')}{\partial\theta_i\partial\theta_j} - \frac{\partial^2\log g(\theta,\theta')}{\partial\theta_i\partial\theta_j} \tag{5}$$

is semi positive definite. Note that

$$g(\theta, \theta') = \theta\bar{\eta}^t - \psi(\theta) - (\theta'\bar{\eta}^t - \psi(\theta')), \tag{6}$$

where

$$\bar{\eta} = \frac{1}{n}\int q(y^n|x^n, \theta')\sum_{t=1}^{n} y_t dy^n.$$

From (6), we have

$$-\frac{\partial^2 \log g(\theta, \theta')}{\partial\theta_i\partial\theta_j} = \frac{\partial^2\psi(\theta)}{\partial\theta_i\partial\theta_j} = I_{ij}(\theta).$$

Hence, semi positive-definiteness of (5) implies

$$\forall\theta \in \Theta,\ \ \hat{J}(\theta, x^n) \le I(\theta),$$

Plugging in $\hat{\theta}$ to θ' in (4) and noting $f(\theta, \hat{\theta}) \le 0$, we have

$$\forall\theta \in \Theta,\ \ 0 \ge f(\theta, \hat{\theta}) \ge g(\theta, \hat{\theta}), \tag{7}$$

where both inequality hold as equality, when $\theta = \hat{\theta}$. That is,

$$g(\theta, \hat{\theta}) \le g(\hat{\theta}, \hat{\theta}) = 0.$$

Together with (6) the following holds

$$g(\theta, \hat{\theta}) = \theta\bar{\eta}^t - \psi(\theta) - (\hat{\theta}\bar{\eta}^t - \psi(\hat{\theta})) \le \hat{\theta}\bar{\eta}^t - \psi(\hat{\theta}) - (\hat{\theta}\bar{\eta}^t - \psi(\hat{\theta})) = 0, \tag{8}$$

which implies $\bar{\eta} = \hat{\eta}$. Here $\hat{\eta}$ denotes the coo responding value of expectation parameter η to $\hat{\theta}$.

Note that

$$g(\theta, \hat{\theta}) = -D(p(\cdot|\hat{\theta})|p(\cdot|\theta)),$$

where $D(p(\cdot|\hat{\theta})|p(\cdot|\theta))$ is the Kullback-Leibler divergence from $p(y|\hat{\theta})$ to $p(y|\theta)$ defined as

$$D(p(\cdot|\hat{\theta})|p(\cdot|\theta)) = \int p(y|\hat{\theta})\log\frac{p(y|\hat{\theta})}{p(y|\theta)}dy.$$

Hence from (3), we have

$$\frac{1}{n}\log\frac{q(x^n|\hat{\theta})}{q(x^n|\theta)} \le D(\hat{\theta}|\theta). \tag{9}$$

Acknowledgment: The authors give sincere gratitude to Hiroshi Nagaoka, who pointed out that the inequality in [1] holds for the model with hidden variables according to an exponential family. He gave an information geometrical proof for the inequality for it.

References

[1] J. Takeuchi & A. R. Barron, "Some Inequality for Mixture Families," http://www-kairo.csce.kyushu-u.ac.jp/~tak/papers/memo_r2.pdf, October 2013.