Supplementary Materials to *Factorized Asymptotic Bayesian Inference for Factorial Hidden Markov Models*

Lemma 1. *Suppose $\{z_{n,1}, \cdots, z_{n,T_n}\}_{n=1}^N$ are N sequences of Bernoulli random variables, whose means are $\{p_{n,1}, \cdots, p_{n,T_n}\}_{n=1}^N$. For the n-th sequence $\{z_{n,t}\}$, $z_{n,1}, \cdots, z_{n,T_n}$ are independent with each other. Let $y_n = \sum_i z_{n,i}$, $\bar{y}_n = E[y_n] = \sum_i p_{n,i}$. Suppose further that $\bar{y}_n \to \infty$ as $T_n \to \infty$. Besides, there are N numbers $\{\hat{y}_n\}$, $\forall n, \hat{y}_n \approx \bar{y}_n$. When all T_n are large enough, the following bounds hold:*

1)

$$\mathbf{E}[\log(y_n + 1)] = \log(\hat{y}_n + 1) + \frac{\bar{y}_n - \hat{y}_n}{\hat{y}_n + 1} + \epsilon_1;$$

2)

$$\mathbf{E}[y_n \log y_n] = \bar{y}_n \log \hat{y}_n + (\bar{y}_n - \hat{y}_n) + \epsilon_2.$$

Here ϵ_1, ϵ_2 are small bounded errors.

Proof.

1) Using the convexity of $-\log(y_n + 1)$, we easily obtain

$$\mathbf{E}[\log(y_n + 1)] \leq \log(\mathbf{E}[y_n] + 1) = \log(\bar{y}_n + 1). \tag{1}$$

We proceed to derive a lower bound of $\mathbf{E}[\log(y_n + 1)]$.

The following lower bound of the logarithm function is well know:

$$\log x > 1 - \frac{1}{x} \quad \text{for all } x > 0.$$

Substituting x with $\dfrac{y_n + 1}{\bar{y}_n + 1}$, we have

$$\log(y_n + 1) - \log(\bar{y}_n + 1) > 1 - \frac{\bar{y}_n + 1}{y_n + 1}.$$

After taking the expectation of both sides, it becomes

$$\mathbf{E}[\log(y_n + 1)] - \log(\bar{y}_n + 1) > 1 - (\bar{y}_n + 1)\mathbf{E}\left[\frac{1}{y_n + 1}\right]. \tag{2}$$

$\mathbf{E}\left[\frac{1}{y_n+1}\right]$ is commonly referred to as the *negative moment* of y_n [1]. We apply the corollary in [1] that

$$\mathbf{E}\left[\frac{1}{y_n+1}\right] = \int_0^1 G_n(t)dt, \tag{3}$$

where $G_n(t)$ is the probability generating function of y_n. It is known that $G_n(t) = \prod_{i=1}^{T_n}(q_{n,i} + p_{n,i} \cdot t)$, in which $q_{n,i} = 1 - p_{n,i}$. Thus (3) becomes

$$\mathbf{E}\left[\frac{1}{y_n+1}\right] = \int_0^1 \prod_{i=1}^{T_n}(q_{n,i} + p_{n,i} \cdot t)dt. \tag{4}$$

We apply the *Inequality of arithmetic and geometric means* to $\prod_{i=1}^{T_n}(q_{n,i} + p_{n,i} \cdot t)$:

$$\begin{aligned}\text{For all } t \geq 0: \quad \prod_{i=1}^{T_n}(q_{n,i} + p_{n,i} \cdot t) &\leq \left(\frac{\sum_i q_{n,i}}{T_n} + \frac{\sum_i p_{n,i}}{T_n}t\right)^{T_n} \\ &= (1 - \frac{\bar{y}_n}{T_n} + \frac{\bar{y}_n}{T_n}t)^{T_n}.\end{aligned}$$

Let $\bar{p}_n$ denote $\frac{\bar{y}_n}{T_n}$, and $\bar{q}_n$ denote $1 - \frac{\bar{y}_n}{T_n}$. Then (4) becomes

$$\mathbf{E}\left[\frac{1}{y_n+1}\right] \leq \int_0^1 (\bar{q}_n + \bar{p}_n t)^{T_n} dt = \frac{1 - \bar{q}_n^{T_n+1}}{(T_n+1)\bar{p}_n} < \frac{1}{\bar{y}_n}. \tag{5}$$

Plugging (5) into (2), we have

$$\mathbf{E}[\log(y_n+1)] - \log(\bar{y}_n+1) > 1 - \frac{\bar{y}_n+1}{\bar{y}_n} = -\frac{1}{\bar{y}_n}.$$

Thus $-\frac{1}{\bar{y}_n}$ is the lower bound of the approximation error, which tends to 0 as $T_n \to \infty$. Hence

$$\exists \epsilon_{1a} > 0, s.t.\ \forall T_n, \mathbf{E}[\log(y_n+1)] > \log(\bar{y}_n+1) - \epsilon_{1a}. \tag{6}$$

Combining (1) and (6), we have

$$\left|\mathbf{E}[\log(y_n+1)] - \log(\bar{y}_n+1)\right| < \epsilon_{1a}. \tag{7}$$

On the other hand, as $\hat{y}_n \approx \bar{y}_n$, the first order approximation of $\log(\bar{y}_n+1)$ about $\hat{y}_n$ is good:

$$\log(\bar{y}_n+1) = \log(\hat{y}_n+1) + \frac{\bar{y}_n - \hat{y}_n}{\bar{y}_n+1} + \epsilon_{1b}, \tag{8}$$

where ϵ_{1b} is an error of order $o(\bar{y}_n^{-1})$ that tends to 0 when T_n increases.

Combining (7) and (8), we have

$$\left|\mathbf{E}[\log(y_n+1)] - \log(\hat{y}_n+1) - \frac{\bar{y}_n - \hat{y}_n}{\bar{y}_n+1}\right| < |\epsilon_{1a}| + |\epsilon_{1b}|.$$

In other words, $\log(\hat{y}_n + 1) + \frac{\bar{y}_n - \hat{y}_n}{\bar{y}_n+1}$ approximates $\mathbf{E}[\log(y_n + 1)]$ with a small bounded error. In addition, this error tends to 0 quickly as the sequence length T_n increases.

2) It is easy to verify $y_n \log y_n$ is convex, and thus

$$\mathbf{E}[y_n \log y_n] \geq \bar{y}_n \log \bar{y}_n. \tag{9}$$

We proceed to derive an upper bound of $\mathbf{E}[y_n \log y_n]$.

Applying the inequality $\log x \leq x - 1$ by substituting x with $\frac{y_n}{\bar{y}_n}$, we have

$$y_n \log \frac{y_n}{\bar{y}_n} \leq y_n(\frac{y_n}{\bar{y}_n} - 1) = \frac{y_n^2}{\bar{y}_n} - y_n. \tag{10}$$

Taking the expectation of both sides of (10), we have

$$\mathbf{E}[y_n \log y_n] - \bar{y}_n \log \bar{y}_n \leq \frac{\mathbf{E}[y_n^2]}{\bar{y}_n} - \bar{y}_n = \frac{\mathbf{Var}(y_n)}{\bar{y}_n} = \frac{\sum_i p_{n,i}^2}{\sum_i p_{n,i}} \leq 1. \tag{11}$$

Combining (9) and (11), we have

$$\left|\mathbf{E}[y_n \log y_n] - \bar{y}_n \log \bar{y}_n\right| \leq 1. \tag{12}$$

Furthermore, when $\bar{y}_n \approx \hat{y}_n$, the first order approximation of $\log \bar{y}_n$ about $\hat{y}_n$ is good, i.e.:

$$\begin{aligned}\bar{y}_n(\log \bar{y}_n - \log \hat{y}_n) &= \bar{y}_n(\bar{y}_n - \hat{y}_n)/\hat{y}_n + \epsilon_{2a} \\ &= (\bar{y}_n - \hat{y}_n) + \frac{(\bar{y}_n - \hat{y}_n)^2}{\hat{y}_n} + \epsilon_{2a}.\end{aligned} \tag{13}$$

where ϵ_{2a} is a small bounded error.

As $T_n \to \infty$, both $\bar{y}_n$ and $\hat{y}_n$ tend to ∞. Therefore $\frac{(\bar{y}_n - \hat{y}_n)^2}{\hat{y}_n}$ is a small bounded error ϵ_{2b}. Then (13) becomes

$$\bar{y}_n \log \bar{y}_n - \bar{y}_n \log \hat{y}_n = (\bar{y}_n - \hat{y}_n) + \epsilon_{2a} + \epsilon_{2b}. \tag{14}$$

Combining (12) and (14), we have

$$\mathbf{E}[y_n \log y_n] = \bar{y}_n \log \hat{y}_n + (\bar{y}_n - \hat{y}_n) + \epsilon_2,$$

where $|\epsilon_2| \leq |\epsilon_{2a}| + |\epsilon_{2b}| + 1$ is also a small bounded error.

Corollary 2. *$\{y_n\}$, $\{\bar{y}_n\}$ and $\{\hat{y}_n\}$ are defined as the same as in* Lemma 1*. Then the following bounds hold:*

1)

$$\mathbf{E}[\log \Gamma(y_n)] = \bar{y}_n(\log \hat{y}_n - \frac{1}{2\hat{y}_n}) - (\hat{y}_n + \frac{1}{2}\log \hat{y}_n) + \frac{1}{2}(\log 2\pi + 1) + \epsilon_3;$$

2)

$$\begin{aligned}
&\mathbf{E}\left[\sum_n \log\Gamma(y_n) - \log\Gamma(\sum_n y_n)\right] \\
=&\sum_n \bar{y}_n \left[\log(\frac{\hat{y}_n}{\sum_m \hat{y}_m}) + \frac{1}{2\sum_m \hat{y}_m} - \frac{1}{2\hat{y}_n}\right] + \frac{1}{2}\log\big(\sum_n \hat{y}_n\big) - \frac{1}{2}\sum_n \log \hat{y}_n \\
&+ \frac{1}{2}(N-1)(\log 2\pi + 1) + \epsilon_4.
\end{aligned}$$

Here ϵ_3, ϵ_4 are small bounded errors.

Proof.

1) The Stirling's approximation for the log-Gamma function is:

$$\log \Gamma(y_n) = (y_n - \frac{1}{2})\log y_n - y_n + \frac{1}{2}\log 2\pi + \epsilon_{3a}, \tag{15}$$

where ϵ_{3a} is an error term of $O(y^{-1})$, which goes to zero quickly as T_n increases. Therefore $\mathbf{E}[\epsilon_{3a}]$, denoted as $\bar{\epsilon}_{3a}$, is also a small bounded number.

Taking the expectation of both sides of (15), and applying Lemma 1, we obtain

$$\mathbf{E}[\log \Gamma(y_n)] = \bar{y}_n(\log \hat{y}_n - \frac{1}{2\hat{y}_n}) - (\hat{y}_n + \frac{1}{2}\log \hat{y}_n) + \frac{1}{2}(\log 2\pi + 1) + \epsilon_3. \tag{16}$$

2) Obtained by repeatedly applying 1), and combining terms involving $\bar{y}_n$.

References

[1] M. T. Chao and W. E. Strawderman, "Negative moments of positive random variables," *Journal of the American Statistical Association*, vol. 67, no. 338, pp. 429–431, 1972. [Online]. Available: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1972.10482404