Supplementary Web Materials for "A New Monte Carlo Method for Estimating Marginal Likelihoods"

Yu-Bo Wang *, Ming-Hui Chen †, Lynn Kuo ‡, and Paul O. Lewis §

*Department of Statistics, University of Connecticut, Storrs, CT 06269, USA yu-bo.wang@uconn.edu

†Department of Statistics, University of Connecticut, Storrs, CT 06269, USA ming-hui.chen@uconn.edu

‡Department of Statistics, University of Connecticut, Storrs, CT 06269, USA lynn.kuo@uconn.edu

§Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA paul.lewis@uconn.edu

1 Proof of Theorem 3.1.

Under certain ergodic conditions and Assumptions 1-2, we have

$$
\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\frac{w_k}{q(\boldsymbol{\theta}_t)}1\{\boldsymbol{\theta}_t \in A_k\}\\
=&\sum_{k=1}^{K} w_k \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{q(\boldsymbol{\theta}_t)}1\{\boldsymbol{\theta}_t \in A_k\}\\
=&\sum_{k=1}^{K} w_k \int_{\{\boldsymbol{\theta}\in A_k\}}\frac{1}{q(\boldsymbol{\theta})}\frac{q(\boldsymbol{\theta})}{c}d\boldsymbol{\theta}\\
=&\frac{1}{c}\sum_{k=1}^{K} w_k V(A_k),
\end{aligned}
$$

which implies that $\hat{d} \xrightarrow{a.s.} d = 1/c$. Let $q_{k,min} = \min_{\{\boldsymbol{\theta}_t \in A_k\}} q(\boldsymbol{\theta}_t)$. Under Assumption 2, we have $q_{k,min} > 0$. Write $g(\boldsymbol{\theta}_t) = \sum_{k=1}^{K} w_k/q(\boldsymbol{\theta}_t)1\{\boldsymbol{\theta}_t \in A_k\}$. Under Assumptions 1 and 2, we have

$$
\begin{aligned}
\mathrm{E}[g(\boldsymbol{\theta}_t)]^2 =& \sum_{k=1}^{K}\mathrm{E}\Big(\Big[\frac{w_k}{q(\boldsymbol{\theta}_t)}\Big]1\{\boldsymbol{\theta}_t \in A_k\}\Big)^2\\
\le& \sum_{k=1}^{K}\frac{w_k^2}{q_{k,min}}\mathrm{E}\Big(\Big[\frac{1}{q(\boldsymbol{\theta}_t)}\Big]1\{\boldsymbol{\theta}_t \in A_k\}\Big)\\
\le& \sum_{k=1}^{K}\frac{w_k^2 V(A_k)}{q_{k,min}c} < \infty,
\end{aligned} \tag{1}
$$

which implies that $\mathrm{Var}[g(\boldsymbol{\theta}_i)] < \infty$. Using Cauchy–Schwarz Inequality, we obtain

$$
\begin{aligned}
\mathrm{Var}\Big[\frac{1}{T}\sum_{t=1}^{T} g(\boldsymbol{\theta}_t)\Big] =& \frac{1}{T^2}\mathrm{Var}\Big[\sum_{t=1}^{T} g(\boldsymbol{\theta}_t)\Big]\\
=& \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathrm{Var}[g(\boldsymbol{\theta}_t)] + 2\sum_{t'<t''}\sum \mathrm{Cov}[g(\boldsymbol{\theta}_{t'}), g(\boldsymbol{\theta}_{t''})]\Big\}\\
\le& \frac{1}{T^2}\Big\{\sum_{t=1}^{T}\mathrm{Var}[g(\boldsymbol{\theta}_t)] + 2\sum_{t'<t''}\sum \sqrt{\mathrm{Var}[g(\boldsymbol{\theta}_{t'})]\mathrm{Var}[g(\boldsymbol{\theta}_{t''})]}\Big\}.
\end{aligned} \tag{2}
$$

Thus, $\mathrm{Var}(\hat{d}) < \infty$ directly follows from (2). □

2 Proof of Theorem 3.2.

First,

$$
\begin{aligned}
\mathrm{Var}\left(\hat{d}_t\right) =& \mathrm{Var}\Big[\sum_{k=1}^{K} \frac{w_k^*}{q(\boldsymbol{\theta}_t)} 1\{\boldsymbol{\theta}_t \in A_k\}\Big] \\
=& \mathrm{E}\Big[\sum_{k=1}^{K} \frac{w_k^*}{q(\boldsymbol{\theta})} 1\{\boldsymbol{\theta} \in A_k\}\Big]^2 - \Big(\mathrm{E}\Big[\sum_{k=1}^{K} \frac{w_k^*}{q(\boldsymbol{\theta})} 1\{\boldsymbol{\theta} \in A_k\}\Big]\Big)^2 \\
=& \mathrm{E}\Big[\sum_{k=1}^{K} \frac{w_k^{*2}}{q^2(\boldsymbol{\theta})} 1\{\boldsymbol{\theta} \in A_k\}\Big] - \Big(\sum_{k=1}^{K} w_k^* \mathrm{E}\Big[\frac{1}{q(\boldsymbol{\theta})} 1\{\boldsymbol{\theta} \in A_k\}\Big]\Big)^2 \\
=& \sum_{k=1}^{K} w_k^{*2} \mathrm{E}\Big[\frac{1}{q^2(\boldsymbol{\theta})} 1\{\boldsymbol{\theta} \in A_k\}\Big] - \Big[\sum_{k=1}^{K} \frac{w_k^* V(A_k)}{c}\Big]^2 \\
=& \sum_{k=1}^{K} w_k^{*2} \alpha_k - \frac{1}{c^2}.
\end{aligned}
$$

Secondly, with the constraint $\sum_{k=1}^{K} w_k^* V(A_k) = 1$, the optimal weights directly follow from the Lagrange multiplier method,

$$
\begin{aligned}
&\frac{\partial}{\partial w_k^*}\Big[\Big(\sum_{k=1}^{K} w_k^{*2}\alpha_k - \frac{1}{c^2}\Big) - \lambda\Big(\sum_{k=1}^{K} w_k^* V(A_k) - 1\Big)\Big] = 0 \\
\Rightarrow & 2w_k^*\alpha_k - \lambda V(A_k) = 0 \\
\Rightarrow & w_k^* = \frac{\lambda V(A_k)}{2\alpha_k}, \quad \text{for } k = 1, ..., K.
\end{aligned}
$$

Replacing w_k^* by $\lambda V(A_k)/(2\alpha_k)$ in the constraint, we can obtain $\lambda = 1/\Big[\sum_{k=1}^{K} V^2(A_k)/(2\alpha_k)\Big]$ and $w_{k,opt}^* = V(A_k)/\Big[2\alpha_k \sum_{k=1}^{K} V^2(A_k)/(2\alpha_k)\Big] = V(A_k)/\Big[\alpha_k \sum_{k=1}^{K} V^2(A_k)/\alpha_k\Big]$, for $k = 1, ..., K$. So, $w_{k,opt}$ is proportional to $V(A_k)/\alpha_k$ for $k = 1, ..., K$. Under this setting, the variance can be simplified to

$$
\mathrm{Var}\left(\hat{d}_t\right) = \frac{1}{\sum_{k=1}^{K} V^2(A_k)/\alpha_k} - \frac{1}{c^2}. \tag{3}
$$

3 Proof of Theorem 4.1.

Under certain ergodic conditions and Assumption 3, the consistency property can be proven similarly as that in the general PWK estimator in (10). Specifically, we have

$$
\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{k=1}^{K^*} \frac{w_k(\boldsymbol{\theta}_t)}{q(\boldsymbol{\theta}_t)} 1\{\boldsymbol{\theta}_t \in A_k\} \\
=& \sum_{k=1}^{K^*} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{w_k(\boldsymbol{\theta}_t)}{q(\boldsymbol{\theta}_t)} 1\{\boldsymbol{\theta}_t \in A_k\} \\
=& \sum_{k=1}^{K^*} \int_{\{\boldsymbol{\theta}\in A_k\}} \frac{w_k(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \frac{q(\boldsymbol{\theta})}{c} d\boldsymbol{\theta} \\
=& \frac{1}{c} \sum_{k=1}^{K^*} \int_{\{\boldsymbol{\theta}\in A_k\}} w_k(\boldsymbol{\theta}) d\boldsymbol{\theta},
\end{aligned}
$$

which implies that $\hat{d}^* \xrightarrow{a.s.} d$.