SUPPLEMENTARY MATERIAL

1 Proof of Theorem 3.1

Theorem 3.1. Suppose that \( \theta \) is a random variable defined on state space \( \Theta \), with probability density \( p(\theta) \). For any given \( \theta \in \Theta \), let \( y \) and \( y' \) be two random variables that are independent conditional on \( \theta \), and both follow the same distribution \( p(y|\theta) \). Now define \( z = y - y' \), and we then have,

\[
E_\theta[H(p(y|\theta))] \leq H(E_\theta[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2,
\]

where \( \text{dim}(y) \) is the dimensionality of \( y \).

Proof. From Shannon's entropy power inequality \cite{1}, we obtain,

\[
\exp(2H(p(z|\theta))/\text{dim}(y)) \\
\geq \exp(2H(p(y|\theta))/\text{dim}(y)) + \exp(2H(p(-y|\theta))/\text{dim}(y)) \\
= 2\exp(2H(p(y|\theta))/\text{dim}(y)),
\]

which implies that

\[
H(p(y|\theta)) \leq H(p(z|\theta)) - \frac{\text{dim}(y)}{2} \log 2. \tag{1}
\]

Taking expectation with respect to \( p(\theta) \) on both sides of Eq. (1) yields,

\[
E_\theta[H(p(y|\theta))] \\
\leq E_\theta[H(p(z|\theta))] - \frac{\text{dim}(y)}{2} \log 2 \tag{2}
\]

where the last inequality is due to the concavity of the entropy \cite{1}.

\[\square\]

2 Proof of Corollary 3.2

Corollary 3.2. Suppose \( p(\theta), p(y|\theta), \) and \( p(z|\theta) \) are defined as is in Theorem 3.1, and \( p(\theta) \) admits the form of,

\[
p(\theta) = \sum_{l=1}^{L} \omega_l f_l(\theta),
\]

where \( \omega_l \geq 0 \) for \( l = 1...L \), \( \sum_{l=1}^{L} \omega_l = 1 \), and \( f_l(\theta) \) are density functions. Then

\[
E_\theta[H(p(y|\theta))] \leq \sum_{l=1}^{L} \omega_l H(E_{\theta\sim f_l}[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2 \\
\leq H(E_\theta[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2.
\]

Proof. Recall that the prior takes the form of

\[
p(\theta) = \sum_{l=1}^{L} \omega_l f_l(\theta),
\]

and we have

\[
E_\theta[H(p(y|\theta))] = \int_{\Theta} p(\theta)H(p(y|\theta))d\theta \\
= \sum_{l=1}^{L} \omega_l \int_{\Theta} f_l(\theta)H(p(y|\theta))d\theta \\
\leq \sum_{l=1}^{L} \omega_l H(E_{\theta\sim f_l}[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2, \tag{3}
\]

where the inequality above is a direct consequence of Theorem 3.1. Once again, because the entropy is concave, we have

\[
\sum_{l=1}^{L} \omega_l H(E_{\theta\sim f_l}[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2 \\
\leq H(\sum_{l=1}^{L} \omega_l E_{\theta\sim f_l}[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2 \tag{4}
\]

where

\[
= H(E_\theta[p(z|\theta)]) - \frac{\text{dim}(y)}{2} \log 2.
\]

\[\square\]

3 Implementation details

This section provides the experimental setup and implementation details of the examples. Code for reproducing our experiments can be found at

[https://github.com/ziq-ao/LBKLD_estimator](https://github.com/ziq-ao/LBKLD_estimator)

The mathematical example. We estimate the expected LB-KLD utility function values with \( 3 \times 10^4 \) (i.e. \( n = 10^4 \)) model simulations. In the prior partition step, we set \( n_{\min} = 10 \) and \( L = 5 \). Averaging
was done over 100 independent runs to mitigate the random errors. Moreover we generate a larger number (10^5) of samples to estimate the KLD based expected utility function values with the nested MC method. For the D-posterior precision method, 100 samples are kept from 10^4 prior-predictive simulations to form the ABC posterior. Again, the reported results are the average over 100 runs.

**Ricker Model.** We estimate the expected LB-KLD utility with 3 \times 10^4 model simulations. In the prior partition step, we set \( n_{\text{min}} = 50 \) and \( L = 5 \). For the D-posterior precision method, 100 out of 10^4 prior-predictive samples are used to compute the posterior statistics.

**Aphid Model.** The implementation setup of the LB-KLD and the D-posterior methods is the same as that of the Ricker model. It should also be mentioned here that, for \( k = 1 \) and \( k = 2 \), the optimal solutions are obtained by exhausting all the integer grid points, while the Simultaneous Perturbation Stochastic Approximation algorithm [2] is used to optimize the expected utility functions for \( k = 3 \) and \( k = 4 \).

**References**
