1. The Problem
Take some distribution \( p(z, x) \) with \( x \) fixed.

Observation: If \( \mathbb{E} R = p(x) \) then \( \mathbb{E} \log R \leq \log p(x) \).

Example: Take \( R = \frac{q(z)}{p(z)} \) for \( z \sim q \) Gaussian, optimize \( q \):
\[
\log R = -0.257 \quad \text{(p(z), naive)}
\]

Decomposition: \( K L(q(z)||p(z|x)) = \log p(x) - \mathbb{E} \log R \).

- Likelihood bound: √
- Posterior approximation: √

Recent work: Better Monte Carlo estimators \( R \).

Second Example: Antithetic Sampling: Let \( T(z) \) “flip” \( z \) around mean of \( q \).

\[
R = \frac{q(z)}{p(z|x)}, \text{antithetic}
\]

- Likelihood bound: √
- Posterior approximation: x

This paper: Is some other distribution close to \( p \)?

2. Main Framework

Contribution of this paper: Given estimator with \( \mathbb{E} R = p(x) \), we show how to construct \( q(z) \) such that
\[
K L(Q(z)||p(z|x)) \leq \log p(x) - \mathbb{E} \log R.
\]

Intuition:
- An unbiased estimator \( \mathbb{E}_w R(w) = p(x) \) is not enough!
- We suggest: Need a coupling: \( \mathbb{E}_w R(w) | z(w) = p(z, x) \).
- We define augmented distributions in state space \( (z, w) \).
- Tightening \( \log p(x) - \mathbb{E} \log R \) is equivalent to VI on the augmented distributions.

2.1 Divide
1. Random variable \( R \) with underlying sample space \( \omega \)
\[
R(\omega), \omega \sim Q(\omega).
\]
2. Assume unbiased
\[
\mathbb{E}_\omega [R(\omega) | z(\omega)] = p(z, x).
\]
3. Define “extended target”
\[
P^{BIC}(z, x) = Q(\omega) R(\omega)
\]
4. Then
\[
K L(Q(\omega)||P^{BIC}(z, x)) = \log p(x) - \mathbb{E} \log R.
\]

2.2 Couple
1. Assume coupling
\[
\mathbb{E}_\omega [R(\omega) | z(\omega)] = p(z, x).
\]
2. Define
\[
P^{\text{BIC}}(z, w) = Q(\omega) R(\omega)
\]
3. Then
\[
K L(Q(\omega)||P^{\text{BIC}}(z, w)) = \log p(x) - \mathbb{E} \log R.
\]

Fine... But where is \( z \)?

2.3 Summary
- Tightening a bound \( \log p(x) - \mathbb{E} \log R \) is equivalent to VI in an augmented state space \( (\omega, z) \).
- To sample from \( Q(z) \) draw \( \omega \) then \( z \sim Q(z|\omega) \).
- We gives couplings for:
  - Antithetic sampling
  - Stratified sampling
  - Quasi Monte Carlo
  - Latin hypercube sampling
  - Arbitrary recursive combinations of above

3. Transforming Couplings

<table>
<thead>
<tr>
<th>Description</th>
<th>( R(\cdot) )</th>
<th>( a(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>IID Mean</td>
<td>( \omega \sim q_0 \ldots q_M ) i.i.d.</td>
<td>( 1 )</td>
</tr>
<tr>
<td>Stratified Sampling</td>
<td>( \Omega = \Omega_1 \ldots \Omega_M ) partition ( Q ); ( \omega \sim \ldots \sim Q_M, \omega \sim Q_0 ) restricted to ( \Omega_0 ), ( \mathbb{E}<em>\omega R(\omega) = \mathbb{E}</em>\omega R_0(\omega))</td>
<td>( \sum_{m=0}^M \sum_{\omega \in \Omega_m} R_0(\omega) \sum_{\omega \in \Omega_m} R(\omega) )</td>
</tr>
<tr>
<td>Antithetic Sampling</td>
<td>( \omega \sim q_0 ) for all ( m ); ( T_M(\omega) ) i.i.d.</td>
<td>( \sum_{m=0}^M \sum_{\omega \in \Omega_m} R(\omega) Q(T_M(\omega)) )</td>
</tr>
<tr>
<td>Randomized Quasi Monte Carlo</td>
<td>( \omega \sim \text{Unif}(0,1)^M, \omega_1, \ldots, \omega_M ) fixed. ( T_M(\omega) = F^{-1}(\omega + \omega (\text{mod } 1)) )</td>
<td>( \sum_{m=0}^M \sum_{\omega \in \Omega_m} R(\omega) Q(T_M(\omega)) )</td>
</tr>
<tr>
<td>Latin Hypercube Sampling</td>
<td>( \omega_1, \ldots, \omega_M ) jointly sampled from Latin hypercube ( [0,1]^M )</td>
<td>( \sum_{m=0}^M \sum_{\omega \in \Omega_m} R(\omega) Q(T_M(\omega)) )</td>
</tr>
</tbody>
</table>

4. Example Implementation
Generate “batches” of samples from \([0,1] \) hypercubes and then transform.

5. Example Results

Better likelihood bounds ⇒ better posteriors √

---


Justin Domke and Daniel Sheldon, University of Massachusetts, Amherst