Metropolis-Hastings with Approximate Acceptance Ratio Calculation

Yury Svirschevsky[1*], Sergey Samsonov[1]

[1]HDI Lab, HSE

*ysvirshchevskii@hse.ru

Approximate Metropolis-Hastings

Algorithm 1 Adaptive Approximate Metropolis-Hastings Step

Input: Target density $\pi(x)$, Generative model $\mathcal{M}$

Output: Samples $Y_{1:n}$ approximating $\pi(x)$, Improved model $\mathcal{M}'$

$\hat{p}_{\mathcal{M}} \leftarrow$ Marginal likelihood estimator for $\mathcal{M}$

$X_{1:n} \leftarrow$ Draw n i.i.d. samples from $\mathcal{M}$

$Y_0 \leftarrow X_0$

for i=1 **to** n **do**

Compute acceptance probability

$$\alpha(Y_{i-1}, X_i) \leftarrow \frac{\pi(X_i)\hat{p}_{\mathcal{M}}(Y_{i-1})}{\pi(Y_{i-1})\hat{p}_{\mathcal{M}}(X_i)} \wedge 1$$

Get next sample

$$Y_i \leftarrow \begin{cases} X_i & \text{with probability } \alpha(Y_{i-1}, X_i), \\ Y_{i-1} & \text{with probability } 1 - \alpha(Y_{i-1}, X_i) \end{cases}$$

end for

$\mathcal{M}' \leftarrow$ Use $Y_{1:n}$ to train new model / fine-tune $\mathcal{M}$

- Generalization of the Metropolis-Hastings Algorithm, a popular MCMC method
- Generative model with **intractable marginal likelihood** used to model the proposal distribution
- **Estimate** of the model's marginal likelihood used in acceptance probability calculations instead of the exact value

Sample Quality is Improved

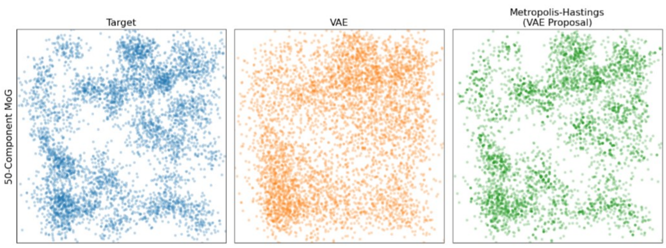

Figure 1: Demonstration on a 2D Mixture-of-Gaussians Target

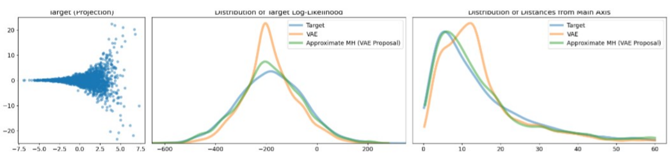

Figure 2: Approximate Metropolis-Hastings Improves Feature Distributions for a 128D Funnel

Marginal Likelihood Estimation

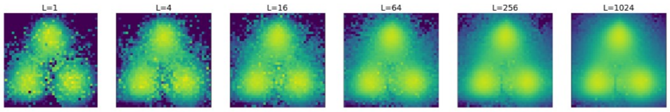

Figure 3: Importance Weighted Likelihood Estimates

The marginal likelihood of a **Variational Autoencoder** is intractable, but can be approximated. Let x and z denote the observed and latent variables respectively, $p_\theta(x, z)$ denote the joint distribution, and $q_\phi(z|x)$ denote the posterior approximation.

Importance Weighted estimator:

$$\hat{p}^{\text{IW}}(x) = \frac{1}{L}\sum_{i=1}^{L} \frac{p_\theta(x, Z^i)}{q_\phi(Z^i|x)}, \quad Z^1, \dots, Z^L \overset{\text{iid}}{\sim} q_\phi(\cdot|x)$$

Sequential Importance Sampling estimator:

$$\hat{p}^{\text{SIS}}(x) = \frac{1}{L}\sum_{i=1}^{L} \frac{p_\theta^K(x, Z^i_{0:K})}{q_\phi^K(Z^i_{0:K}|x)}, \quad Z^1, \dots, Z^L \overset{\text{iid}}{\sim} q_\phi^K(\cdot|x)$$

$$p_\theta^K(x, Z_{0:K}) = p_\theta(x, Z_k) \prod_{k=0}^{K-1} l_k(z_{k+1}, z_k)$$

$$q_\phi^K(Z^i_{0:K}|x) = q_\phi(Z_k|x) \prod_{k=1}^{K} m_k(z_{k-1}, z_k)$$

where m_k and l_k are densities of forward and reverse Markov kernels. Different choices of kernels lead to different estimators.

Comparison with Classic Metropolis-Hastings

While Normalizing Flows have tractable marginal likelihoods, which allows exact acceptance probability calculation when using them as proposals, they are less flexible than VAEs.

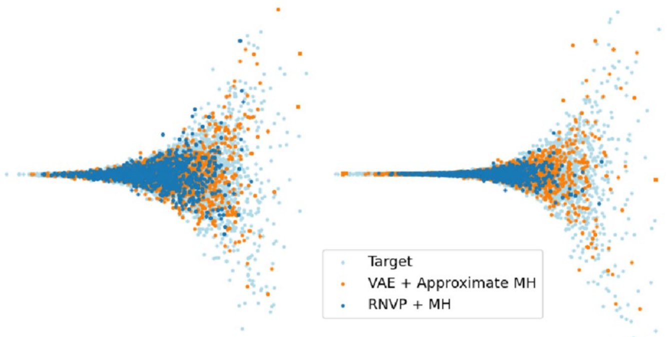

Figure 4: VAE-Proposal Approximate Metropolis-Hastings vs. Flow-Proposal Classic Metropolis-Hastings