MCMC Sampling with Uncertain Target Probabilities

Penalty Method and Pseudo-Likelihood Approach

Ludwig Winkler

Monte Carlo Methods in Artificial Intelligence and Machine Learning

July 25, 2017

Outline

Recap: Metropolis-Hastings

- Objective: Generate samples from the target probability distribution π given the proposal distribution Q
- Acceptance probability

$$A(x'|x) = \mathsf{min}(1, q(x'|x))$$

with

$$q(x'|x) = \frac{\pi(x')\ Q(x|x')}{\pi(x)\ Q(x'|x)}$$

Recap: Metropolis-Hastings

Metropolis-Hastings

Input: N number of samples, Q proposal distribution, π target distribution
Output: Array s of samples
Initialize s[1] $= x_0$
for $i = 2, .., N$ **do**
    $x =$ s[i-1]
    Draw $x' \sim Q(x'|x)$
    $A = \min\left(1, \frac{\pi(x')}{\pi(x)} \frac{Q(x|x')}{Q(x'|x)}\right)$
    if $u \sim \mathcal{U}(0,1) < A$ **then**
        s[i] $= x'$
    else
        s[i] $= x$
    end if
end for

Recap: Metropolis-Hastings Detailed Balance

- **Detailed balance:** Transition $(x'|x)$ is reversible
- Overall transition probability

$$\underbrace{T(x'|x)}_{\text{Transition}} = \underbrace{Q(x'|x)}_{\text{Proposal}} \; \underbrace{A(x'|x)}_{\text{Acceptance}}$$

Detailed Balance

$$\pi(x)T(x'|x) = \pi(x')T(x|x')$$

$$A(x'|x) = \frac{\pi(x')Q(x|x')}{\pi(x)Q(x'|x)}A(x|x')$$

Boltzmann Distribution as Target Distribution

- Target distribution: **Boltzmann distribution**

$$\pi(x) \propto \exp\left(-\frac{V(x)}{k_B T}\right)$$

- Symmetric proposal density $Q(x'|x) = Q(x'|x)$ yields acceptance probability

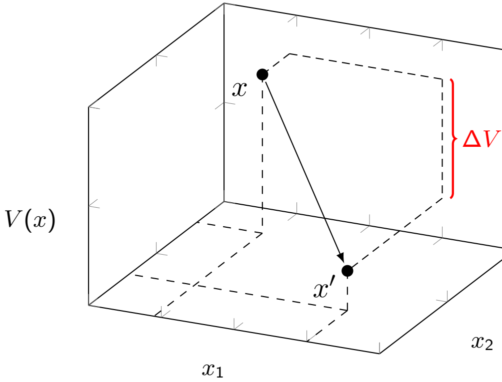
$$\begin{aligned} A(x'|x) &= \min\left(1, \frac{\pi(x')\ Q(x|x')}{\pi(x)\ Q(x'|x)}\right) \\ &= \min\left(1, \exp\left(-\frac{V(x') - V(x)}{k_B T}\right)\right) \end{aligned}$$

- Acceptance probability depends on the energy difference

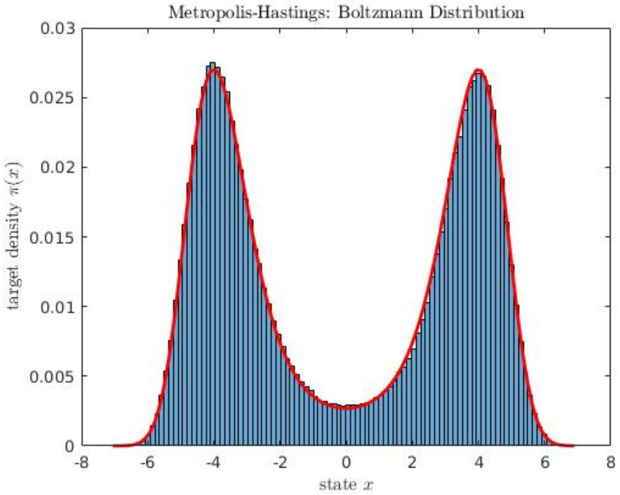
Boltzmann Distribution as Target Distribution

- Acceptance probability depends on the energy difference

$$A(x'|x) = \min\left(1, \exp\left(-\frac{V(x') - V(x)}{k_B T}\right)\right)$$



Metropolis-Hastings with Boltzmann Distribution



$$Q(x'|x) \sim \mathcal{U}(-0.5, 0.5) \quad V(x) = -0.288x^2 + 0.009x^4 \quad k_B T = 1$$

Detailed Balance with Boltzmann Distribution

- Rewriting $q(x'|x)$

$$
\begin{aligned}
q(x'|x) &= \exp \ln \frac{\pi(x')Q(x|x')}{\pi(x)Q(x'|x)} \\
&= \exp\left(\ln \pi(x') - \ln \pi(x) - \ln Q(x'|x) + \ln Q(x|x')\right) \\
&= \exp\left(- \underbrace{\left(\frac{V(x') - V(x)}{k_B\, T} - \ln \frac{Q(x|x')}{Q(x'|x)} \right)}_{=:\Delta(x'|x)} \right)
\end{aligned}
$$

- Acceptance probability

$$A(x'|x) = \min\left(1, e^{-\Delta(x'|x)}\right)$$

- Detailed Balance

$$A(x'|x) = e^{-\Delta(x'|x)} A(x|x')$$

Noisy Energy Difference

What if $\pi(x)$ is uncertain or very noisy?

- Consider a noisy energy difference $\Delta(x'|x)$ as a random variable $\delta(x'|x)$

$$\delta(x'|x) := \Delta(x'|x) + \epsilon \qquad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

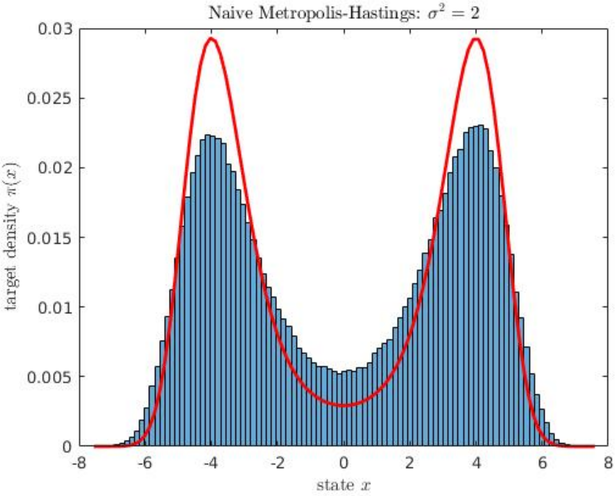
- Recall acceptance probability

$$A(x'|x) = \min\left(1, e^{-\Delta(x'|x)}\right)$$

$\Rightarrow$ Acceptance probability A is randomized
$\Rightarrow$ $A(x'|x) \rightarrow A(\delta; x'|x)$

Metropolis-Hastings with Noisy Energy Difference

Direct penalty of the energy difference



Detailed Balance with Uncertainty

- Average randomized acceptance probability

$$\mathbb{E}[A(\delta; x'|x)] = \int_{-\infty}^{\infty} f(\delta; x'|x) A(\delta; x'|x) d\delta$$

with $f(\delta; x'|x)$ the probability density function of δ

Detailed Balance with Uncertainty

$$\mathbb{E}[A(\delta; x'|x)] = e^{-\Delta(x'|x)} \mathbb{E}[A(\delta; x|x')]$$

Detailed Balance with Uncertainty

- If the density function $f(\delta)$ is symmetric, it fulfills

$$f(\delta; x|x') = f(-\delta; x'|x)$$

- Fulfilled for our assumption $f(\delta; x'|x) \sim \mathcal{N}(\Delta, \sigma^2)$

Detailed Balance with Uncertainty

$$\mathbb{E}[A(\delta; x'|x)] = e^{-\Delta(x'|x)}\mathbb{E}[A(-\delta; x'|x)]$$

Detailed Balance with Uncertainty

Detailed Balance with Uncertainty

$$\mathbb{E}[A(\delta; x'|x)] = e^{-\Delta(x'|x)}\mathbb{E}[A(-\delta; x'|x)]$$

Find $A(\delta; x'|x)$ that satisfies

$$\int_{-\infty}^{\infty} f(\delta; x'|x)\left(A(\delta; x'|x) - e^{-\Delta(x'|x)}A(-\delta; x'|x)\right) d\delta = 0$$

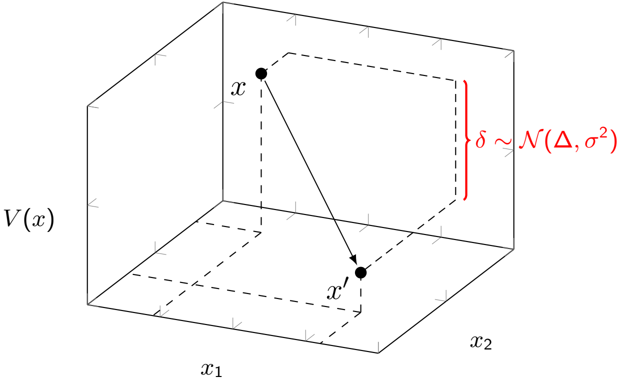
for all $f(\delta; x'|x)$ and $\Delta(x'|x)$ subject to $0 \leq A(\delta; x'|x) \leq 1$

1. Find $A(\delta; x'|x)$ with unknown Δ and known σ^2
2. Find $A(\delta; x'|x)$ with unknown Δ and unknown σ^2

Case 1: Acceptance Probability with Known Variance

- Detailed Balance is fulfilled for $f(\delta; x'|x)$ and known σ^2 with

$$A(\delta; x'|x) := \min\left[1, \exp\left(-\delta - \frac{\sigma^2}{2}\right)\right]$$



Case 1: Acceptance Probability with Known Variance

- Known variance term $\frac{\sigma^2}{2}$ causes a reduction in the acceptance

$$A(\delta; x'|x) := \min\left[1, \exp\left(-\delta \underbrace{-\frac{\sigma^2}{2}}_{\text{noise penalty}}\right)\right]$$

- Acceptance probability in Metropolis-Hastings is unpenalized

$$A_{MH}(\delta; x'|x) := \min\left[1, \exp\left(-\delta\right)\right]$$

Case 1: Acceptance Probability with Known Variance

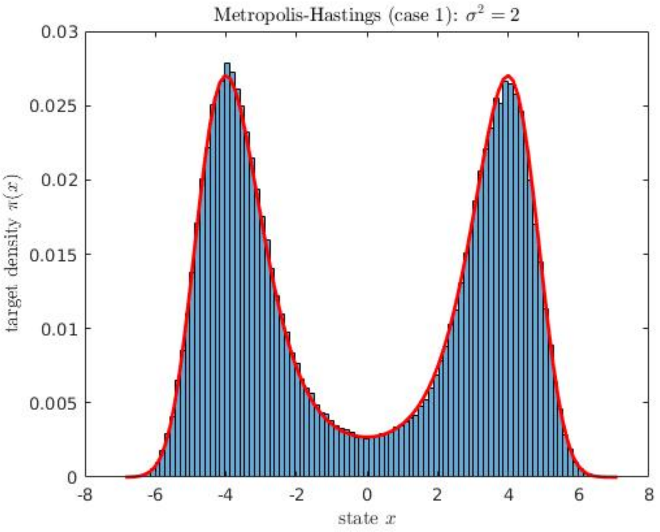
- Known variance term $\frac{\sigma^2}{2}$ causes a reduction in the acceptance

$$A(\delta; x'|x) := \min\left[1, \exp\left(-\delta \underbrace{-\frac{\sigma^2}{2}}_{\text{noise penalty}}\right)\right]$$

- For $V(x) \geq V(x')$ sampled energy difference $\delta \leq 0$
- Moves with $-\delta - \frac{\sigma^2}{2} \geq 0$ will always be accepted

$$A(\delta; x'|x)\Big|_{\delta \leq -\frac{\sigma^2}{2}} = 1$$

Case 1: Boltzmann Distribution



Case 2: Acceptance Probability with Unknown Variance

- Estimate $\Delta = \mathbb{E}[\delta]$ and $\sigma^2 = \mathbb{V}[\delta]$ from i.i.d. samples $\{\delta_1, ..., \delta_N\}$

$$\delta_i \sim \mathcal{N}(\Delta, N\sigma^2) \quad \forall i = 1, ..., N$$

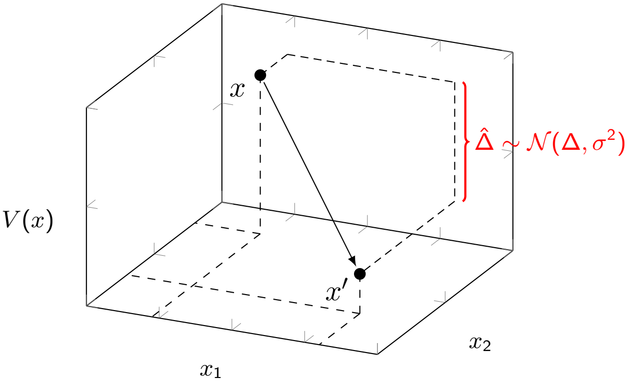
- Estimators $\hat{\Delta}$ and $\hat{\sigma}^2$ are **random variables**

$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N} \delta_i \qquad \hat{\sigma}^2 = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\delta_i - \hat{\Delta})^2$$

Case 2: Acceptance Probability with Unknown Variance

- Energy difference distribution Δ estimated from state transitions

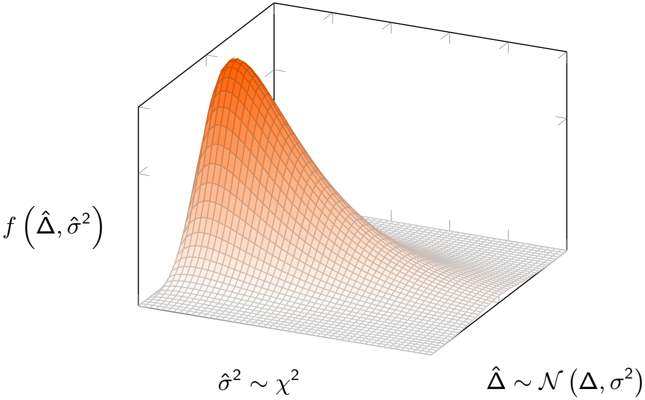
$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N}\delta_i \qquad \hat{\sigma}^2 = \frac{1}{N(N-1)}\sum_{i=1}^{N}(\delta_i - \hat{\Delta})^2$$



Case 2: Acceptance Probability with Unknown Variance

- PDF over energy difference Δ and variance σ^2

$$f(\hat{\Delta}, \hat{\sigma}^2) = \mathcal{N}(\hat{\Delta}|\Delta, \sigma^2) \cdot \chi^2$$



Case 2: Acceptance Probability with Unknown Variance

- Average Acceptance probability for the transition $(x'|x)$ is then

$$\mathbb{E}\left[A(\hat{\Delta},\hat{\sigma}^2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\hat{\Delta},\hat{\sigma}^2)\ A(\hat{\Delta},\hat{\sigma}^2) d\hat{\Delta} d\hat{\sigma}^2$$

Detailed balance

$$\mathbb{E}\left[A(\hat{\Delta},\hat{\sigma}^2)\right] = \exp(-\Delta)\mathbb{E}\left[A(-\hat{\Delta},\hat{\sigma}^2)\right]$$

Case 2: Acceptance Probability with Unknown Variance

- Detailed Balance is fulfilled $\forall\ \hat{\Delta}$ and $\forall\ \hat{\sigma}^2 \geq 0$ if

$$A(\hat{\Delta}, \hat{\sigma}^2) = \min\left[1, \exp(-\hat{\Delta} - u_B)\right]$$

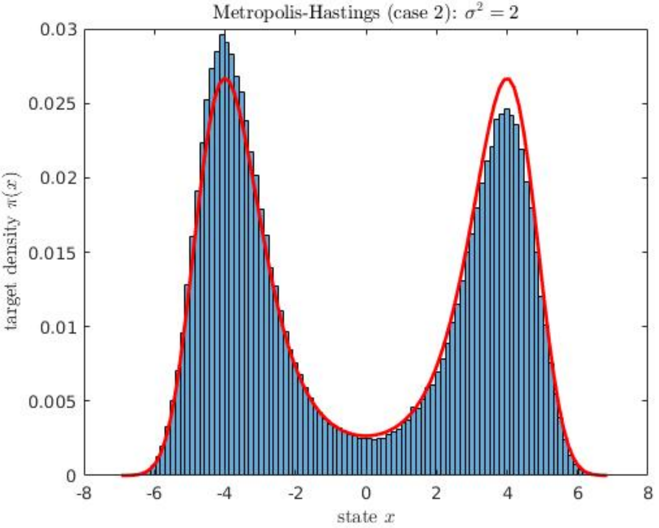
with

$$u_B = \frac{\hat{\sigma}^2}{2} + \frac{(\hat{\sigma}^2)^2}{4(N+1)} + \frac{(\hat{\sigma}^2)^3}{3(N+1)(N+3)} + \ldots$$

- Recall acceptance ratio with known variance

$$A(\delta; x'|x) := \min\left[1, \exp\left(-\delta - \frac{\sigma^2}{2}\right)\right]$$

Case 2: Boltzmann Distribution



Motivation

- Direct sampling from $\pi(x)$ can be difficult
- Approximation of $\pi(x)$ with estimate $r(x)$
- Naive MH with $r(x)$ used as target, often inaccurate

- Auxiliary variable W for estimator randomness
- Ratio between estimate and target is a random variable

$$W = \frac{r(x)}{\pi(x)} \quad \text{with} \quad \mathbb{E}[W|x] = c$$

Pseudo-Marginal MCMC

- Sampling from the joint target distribution of x and auxiliary w

$$f(x, w) = f(w|x)r(x)$$

- Conditional dependency allows for factorization

$$f(x, w) = f(w|x)r(x) = f(w|x)\ w\ \pi(x)$$

- Factorization of proposal density over joint target distribution

$$Q(x', w'|x, w) = Q(x'|w, x)\ Q(w'|x', x, w) = Q(x'|x)\ f(w'|x')$$

Pseudo-Marginal MCMC - Acceptance Ratio

Acceptance Ratio with Estimated Joint Target Distribution

$$\begin{aligned} A(x', w'; x, w) &= \min\left[1, \frac{f(x', w')Q(x, w|x', w')}{f(x, w)\ Q(x', w'|x, w)}\right] \\ &= \min\left[1, \frac{\pi(x')w' f(w'|x')\ Q(x|x')f(w|x)}{\pi(x)w f(w|x)\ \ Q(x'|x)f(w'|x')}\right] \\ &= \min\left[1, \frac{\pi(x')Q(x|x')w'}{\pi(x)Q(x'|x)w}\right] \end{aligned}$$

- Adjustment of acceptance ratio with $\frac{w'}{w}$
- Asymptotic distribution is $w f(w|x)\pi(x)$

Pseudo-Marginal MCMC - Marginalization

- Marginalize over w to obtain $\pi(x)$

$$\int_{-\infty}^{\infty} f(x,w)dw = \pi(x)\int_{-\infty}^{\infty} \underbrace{w\ \ f(w|x)}_{\mathbb{E}[W|x]=c} dw = c\pi(x)$$

- For $N \to \infty$ marginalization will converge to $\pi(x)$

Simulation

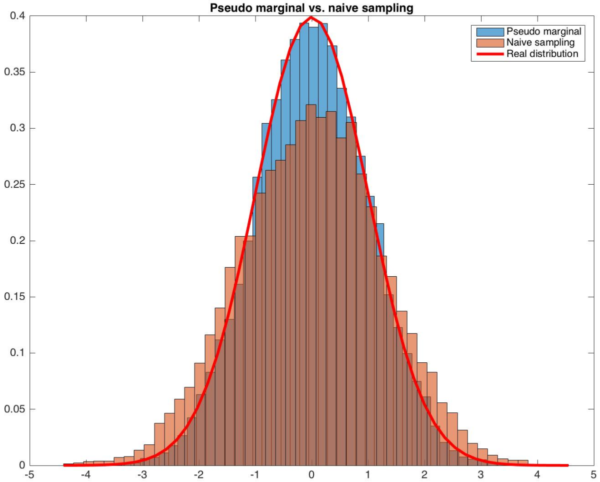
- Sampling from normal distribution
- Access only to estimator

$$r(x) = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}}_{\pi(x)} \cdot \underbrace{(1 + 0.9(E - 1))}_{\text{estimator noise } w} \quad \text{with} \quad E \sim \text{Exp}(1)$$

- Unbiased and positive estimator

Pseudo-Marginal MCMC vs. Naive MH

- Normal distribution with exponential noise



Simulation

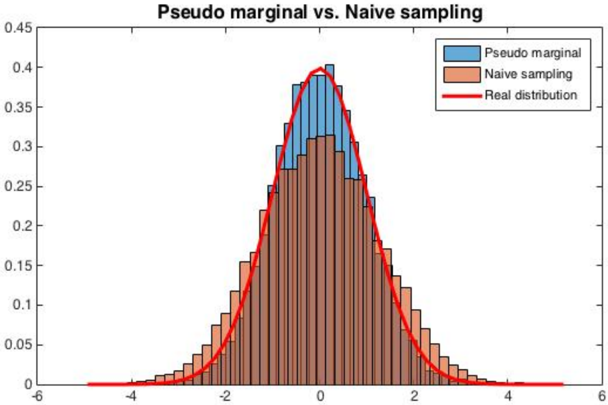
- Sampling from normal distribution
- Access only to estimator

$$r(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot L \quad \text{with} \quad L \sim \text{Laplace}(0,1)$$

- $\mu_L = 0$ and estimator can be negative

Pseudo-Marginal MCMC vs. Naive MH

- Normal distribution with Laplace noise



Simulation

- Sampling from Boltzmann distribution described earlier
- Target distribution given by

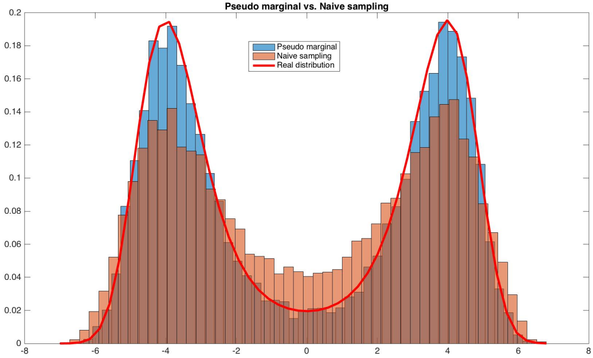
$$\pi(x) \propto \exp\left(-V(x)\right)$$

- Access only to estimator

$$\begin{aligned} r(x) &\propto e^{-V(x)+\eta} && \text{with} \quad \eta \sim \mathcal{N}(0,2) \\ &= e^{-V(x)} \cdot Z && \text{with} \quad Z \sim \text{Lognormal}(0,2) \end{aligned}$$

Pseudo-Marginal MCMC vs. Naive MH

- Boltzmann distribution with Log-Normal noise



Take Home Message

Random Walks for Uncertain Target Distribution

- Gaussian noise in energy difference yields randomized acceptance probability
- Detailed balance holds for the randomized acceptance probability **on average**
- Estimate the parameters of the energy difference density (normal distributed)

Pseudo-Marginal for Uncertain Target Distributions

- Auxiliary variable for the randomness of the estimator
- Joint PDF of estimator noise and target
- Marginalize estimator to obtain target distribution

References

- Ceperley, D. M. and M. Dewing (1999). "The penalty method for random walks with uncertain energies". In: *Journal of Chemical Physics* 110.20, pp. 9812–9820.
- Nicholls, G. K., C. Fox, and A. Muir Watt (2012). "Coupled MCMC with a randomized acceptance probability". In: *ArXiv e-prints*.
- Wilkinson, D. (2010). *The pseudo-marginal approach to exact approximate MCMC algorithms*. URL: https://darrenjw.wordpress.com/2010/09/20/the-pseudo-marginal-approach-to-exact-approximate-mcmc-algorithms/ (visited on 07/17/2017).