Hamiltonian Monte Carlo (Addendum)

Instructor: Prof. Jason Pacheco

Images: From M. Betancourt, “A Conceptual Introduction to HMC”
Volume vs. Dimension

Volume outside shell grows exponentially faster than volume inside.

Volume containing mode becomes negligible as dim. increases.

Number of possible directions (neighbors) is $3^d - 1$, exponential in dimension.
Intuition of Typical Set
Example: Gaussian

\[ D \text{ i.i.d Gaussian RVs } y = (y_1, \ldots, y_D)^T \in \mathbb{R}^D : \quad y \sim \mathcal{N}(0, I) \]

Squared distance is Chi-Squared RV: \[ \| y \|^2 \sim \chi^2(D) \]

[ Source: mc-stan.org ]
Weak law of large numbers. Take $x$ to be the average of $N$ independent random variables $h_1, \ldots, h_N$, having common mean $\bar{h}$ and common variance $\sigma_h^2$: $x = \frac{1}{N} \sum_{n=1}^{N} h_n$. Then

$$P((x - \bar{h})^2 \geq \alpha) \leq \sigma_h^2/\alpha N. \quad (4.32)$$

Proof: obtained by showing that $\bar{x} = \bar{h}$ and that $\sigma_{\bar{x}}^2 = \sigma_h^2/N$. \hfill \Box

- Holds for any $\alpha$ and $N$ large enough
- E.g. on average samples are similar to the mean
Applying LLN to estimate entropy we get:

\[
\frac{1}{N} \sum_{n=1}^{N} \log_2 \left( \frac{1}{p(x_n)} \right) \rightarrow H(X)
\]

Thus $X$ typically belongs to a subset of size $2^{NH(X)}$ with each element having probability $p(x)$ near $2^{-NH(X)}$.

This is the called the typical set of elements with probability:

\[
2^{-NH(X)-\epsilon} \leq p(x) \leq 2^{-NH(X)+\epsilon}
\]
Example: Random Binary String

Let $x = (x_1, \ldots, x_N)$ be $N$-length binary string

Probability of $r$ 1s and $(N-r)$ 0s:

$$P(x) = p_1^r (1 - p_1)^{N-r}$$

Number of $N$-length strings with $r$ 1s:

$$n(r) = \binom{N}{r}$$

$n(r)$ follows a Binomial distribution:

$$P(r) = \binom{N}{r} p_1^r (1 - p_1)^{N-r}$$
Example: Random Binary String

Most of the volume is concentrated away from the mean.
Ideally, MCMC dynamics should explore typical set efficiently.

In theory, HMC aligns vector field with typical set.
Conservative Dynamics Intuition

Attractive

Diffusion

Conservative
Volume Preservation

Dissipative System
(Volume Grows / Shrinks)

Conservative System
(Volume Preserved)

Conservative dynamics defined by volume preservation
Proof of Volume Preservation

Recall Hamiltonian dynamics:

\[
\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}
\]

Approximation of HMC transition for $d=1$ and time $\delta \approx 0$:

\[
T_\delta(q, p) = \begin{bmatrix} q \\ p \end{bmatrix} + \delta \begin{bmatrix} dq/dt \\ dp/dt \end{bmatrix} + \text{terms of order } \delta^2 \text{ or higher}
\]

Jacobian:

\[
B_\delta = \begin{bmatrix}
1 + \delta \frac{\partial^2 H}{\partial q \partial p} & \delta \frac{\partial^2 H}{\partial p^2} \\
-\delta \frac{\partial^2 H}{\partial q^2} & 1 - \delta \frac{\partial^2 H}{\partial p \partial q}
\end{bmatrix} + \text{terms of order } \delta^2 \text{ or higher}
\]
Proof of Volume Preservation (cont’d)

Determinant of Jacobian equals volume:

\[
\det(B_\delta) = 1 + \delta \frac{\partial^2 H}{\partial q \partial p} - \delta \frac{\partial^2 H}{\partial p \partial q} + \text{terms of order } \delta^2 \text{ or higher}
\]

\[
= 1 + \text{terms of order } \delta^2 \text{ or higher}
\]

\[
\log \det(B_\delta) \approx 0 \text{ since } \log(1 + x) \approx x \text{ for } x \text{ near zero}
\]

Consider \( \log \det(B_s) \) for \( s \) not close to zero

- Set \( \delta = s/n \) and apply \( T_\delta \) \( n \) times

\[
\log \det(B_s) = \sum_{i=1}^{n} \log \det(B_\delta) = \sum_{i=1}^{n} \left\{ \text{terms of order } 1/n^2 \text{ or smaller} \right\}
\]

\[
= \text{terms of order } 1/n \text{ or smaller} \quad \text{As } n \to \infty \text{ we have } \log \det(B_s) \to 0
\]
Proof of Volume Preservation (cont’d)

- For $d > 1$ each $dx^d$ submatrix (row $j$, col $i$) of Jacobian is:

\[
B_\delta = \begin{bmatrix}
I + \delta \left[ \frac{\partial^2 H}{\partial q_j \partial p_i} \right] & \delta \left[ \frac{\partial^2 H}{\partial p_j \partial p_i} \right] \\
-\delta \left[ \frac{\partial^2 H}{\partial q_j \partial q_i} \right] & I - \delta \left[ \frac{\partial^2 H}{\partial p_j \partial q_i} \right]
\end{bmatrix} + \text{terms of order } \delta^2 \text{ or higher}
\]

- Determinant is still $1 +$ higher order terms, remainder of argument holds