STAT 237
Markov Chain Monte Carlo

April 1, 2022

Colin Reimer Dawson

Outline

Sampling from the Posterior

- Therefore, provided we can **sample** from our posterior distribution, $p(\theta \mid x_1, \ldots, x_N)$, we can estimate the expected value of various functions of θ
- For example, if we want $\mathbb{E}\left[p(x_{\text{new}} \mid \mu)\right]$, we could approximate it via

$$\mathbb{E}\left[p(x_{\text{new}} \mid \mu)\right] \approx \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}\left[p(x_{\text{new}} \mid \mu^{(s)})\right]$$

Approaches to sample/integrate against a given distribution

- Three approaches:
  1. Inverse CDF Method
  2. Rejection Sampling
  3. Markov Chain Monte Carlo

Non-Independent Samples

- In practice, generating independent samples is often intractable
- We can often generate **correlated** samples, however
- Idea: Use the current value to “seed” the next one

Sequential vs Independent Samples

- With **independent** sampling methods, the parameter value $\theta^{(s)}$ generated at each iteration comes **directly** from the target distribution (such as the posterior)

$$\theta^{(s)} \sim p(\theta)$$

- With **sequential** methods, we sample $\theta^{(s)}$ from a distribution that depends on $\theta^{(s-1)}$

$$\theta^{(s)} \sim p^*(\theta^{(s)} \mid \theta^{(s-1)})$$

- **Goal:** Choose p^* so that the **marginal** distribution of $\theta^{(s)}$ is the target, $p(\theta)$

"Stationary" Distributions

- **Procedure:** Sample

$$\theta^{(s)} \sim p^*(\theta^{(s)} \mid \theta^{(s-1)})$$

- **Goal:** Choose p^* so that the **marginal** distribution $p^*(\theta^{(s)})$ is the same as the target, $p(\theta)$
- For this to work, we must have

$$p(\theta^{(s)}) = \int_\Theta p^*(\theta^{(s-1)}, \theta^{(s)})\ d\theta^{(s-1)}$$

- Assuming it works at $s-1$, this becomes (product rule)

$$p(\theta^{(s)}) = \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)})\ d\theta^{(s-1)}$$

- In other words, $p^*(\theta^{(s)} \mid \theta^{(s-1)})$ **preserves** $p(\theta)$ once it "finds" it

Converging to a Stationary Distribution

- A random walk over a bounded parameter space will ultimately "converge" to a uniform distribution regardless of starting conditions
- Useful if that's our target, but what if we have a different target?
- For instance, if

$$p(\theta) = \begin{cases} 1/21 & \theta = 1 \\ 2/21 & \theta = 2 \\ 3/21 & \theta = 3 \\ 4/21 & \theta = 4 \\ 5/21 & \theta = 5 \\ 6/21 & \theta = 6 \end{cases}$$

The Metropolis Algorithm

To obtain samples from a target distribution, $p(\theta)$:

1. Pick some initial $\theta^{(0)}$ (somehow)
2. For $s = 1, \dots S$:
   (i) Generate a **proposal**, θ^* from a **symmetric random walk** distribution $q(\theta^* \mid \theta^{(s-1)})$, satisfying

   $$q(\theta' \mid \theta) = q(\theta' \mid \theta) \text{ for all pairs } \theta, \theta'$$

   (ii) Calculate the **acceptance probability**

   $$\alpha := \min\left(1, \frac{p(\theta^*)}{p(\theta)}\right)$$

   (iii) Generate $u \sim \mathrm{Unif}(0, 1)$, and set

   $$\theta^{(s)} = \begin{cases} \theta^* & u < \alpha \\ \theta^{(s-1)} & u \geq \alpha \end{cases}$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$p^*(\theta^{(s)}) = \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)})\ d\theta^{(s-1)}$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$
\begin{aligned}
p^*(\theta^{(s)}) &= \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \alpha p(\theta^{(s-1)})\ d\theta^{(s-1)}
\end{aligned}
$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$\begin{aligned}
p^*(\theta^{(s)}) &= \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \alpha p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \frac{p(\theta^{(s)})}{p(\theta^{(s-1)})} p(\theta^{(s-1)})\ d\theta^{(s-1)}
\end{aligned}$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$
\begin{aligned}
p^*(\theta^{(s)}) &= \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)}) \; d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \alpha p(\theta^{(s-1)}) \; d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \frac{p(\theta^{(s)})}{p(\theta^{(s-1)})} p(\theta^{(s-1)}) \; d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s)}) d\theta^{(s-1)}
\end{aligned}
$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$\begin{aligned}
p^*(\theta^{(s)}) &= \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)}) \; d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \alpha p(\theta^{(s-1)}) \; d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \frac{p(\theta^{(s)})}{p(\theta^{(s-1)})} p(\theta^{(s-1)}) \; d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s)}) d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s-1)} \mid \theta^{(s)}) p(\theta^{(s)}) d\theta^{(s-1)}
\end{aligned}$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$\begin{aligned}
p^*(\theta^{(s)}) &= \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \alpha p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \frac{p(\theta^{(s)})}{p(\theta^{(s-1)})} p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s)}) d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s-1)} \mid \theta^{(s)}) p(\theta^{(s)}) d\theta^{(s-1)} \\
&= p(\theta^{(s)})
\end{aligned}$$

Why does it work?

Does it **preserve** $p(\theta)$? Suppose $\theta^{(s-1)}$ has the target distribution, p. Then

$$\begin{aligned}
p^*(\theta^{(s)}) &= \int_\Theta p^*(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \alpha p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) \frac{p(\theta^{(s)})}{p(\theta^{(s-1)})} p(\theta^{(s-1)})\ d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s)} \mid \theta^{(s-1)}) p(\theta^{(s)}) d\theta^{(s-1)} \\
&= \int_\Theta q(\theta^{(s-1)} \mid \theta^{(s)}) p(\theta^{(s)}) d\theta^{(s-1)} \\
&= p(\theta^{(s)})
\end{aligned}$$

So, yes, Metropolis **preserves the target**

Metropolis-Hastings

Note that this suggests a generalization of the algorithm in case q is not symmetric: we can adjust for asymmetry through the acceptance probability

The Metropolis-Hastings Algorithm

To obtain samples from a target distribution, $p(\theta)$:

1. Pick some initial $\theta^{(0)}$ (somehow)
2. For $s = 1, \ldots S$:
   (i) Generate a **proposal**, θ^* from a distribution $q(\theta^* \mid \theta^{(s-1)})$
   (ii) Calculate the **acceptance probability**

   $$\alpha := \min\left(1, \frac{q(\theta^{(s-1)} \mid \theta^*)}{q(\theta^* \mid \theta^{(s-1)})} \frac{p(\theta^*)}{p(\theta^{(s-1)})}\right)$$

   (iii) Generate $u \sim \text{Unif}(0, 1)$, and set

   $$\theta^{(s)} = \begin{cases} \theta^* & u < \alpha \\ \theta^{(s-1)} & u \geq \alpha \end{cases}$$