Markov chain Monte Carlo

David J. H. Shih

Intended learning outcomes

- Implement statistical models in Stan

Notation

Many notations exist for specifying a statistical model.

For a discrete random variable X, the following notations are equivalent:

$$
\begin{aligned}
X &\sim g(\theta) \\
X &\sim g(\cdot \mid \theta) \\
X &\sim g(X \mid \theta) \\
X \mid \theta &\sim g(X \mid \theta) \\
P\left(X = x \mid \theta\right) &= g(x \mid \theta) \\
p\left(x \mid \theta\right) &= g(x \mid \theta) \\
p\left(x \mid \theta\right) &= g(x; \theta) \\
f_X(x) &= g(x \mid \theta)
\end{aligned}
$$

Notation

We will mostly use:

$$X \sim g(\theta)$$
$$p\left(x \mid \theta\right) = g(x \mid \theta)$$

Inverse transform sampling

We can sample $U \sim \text{Uniform}\,(0, 1)$ using a pseudorandom number generator.

We want to sample $X \sim F_X$.

Derive an invertible transform $T : [0, 1] \to \mathcal{R}$
such that $T(U) \sim F_X$.

$$
\begin{aligned}
F_X(x) &= P\,(X \leq x) && \text{(def'n of CDF)} \\
&= P\,(T(U) \leq x) && \text{(def'n of T)} \\
&= P\left(U \leq T^{-1}(x)\right) && \text{(T is invertible)} \\
&= T^{-1}(x) && (U \sim \text{Uniform})
\end{aligned}
$$

Therefore, $T(x) = F_X^{-1}(x)$.

Inverse transform sampling

Sample $X \sim F$.

procedure $\textsc{sample}(F^{-1})$
 draw $u \sim \text{Uniform}\,(0,1)$
 return $F^{-1}(u)$
end procedure

Probabilistic acceptance

Accept x with probabiliby a; otherwise, accept x'.

procedure ACCEPT(x, a, x')
  draw $u \sim$ Uniform $(0, 1)$
  if $u \leq a$ **then**
    return x
  else
    return x'
  end if
end procedure

Rejection sampling

We have proposal distribution q, and we want to sample from target distribution p.

The support of q must contain the support of p.

Choose $M > 0$ s.t. $Mq(x) \geq p(x)\ \forall x$.

Use $X \sim q$ to sample $Y \sim p$.

procedure $\textsc{Rejection}(q, p, M)$ ▷ proposal q, target p
  while $y = \emptyset$ **do**
    draw $x \sim q$
    $a \leftarrow \frac{p(x)}{Mq(x)}$
    $y \leftarrow \textsc{Accept}(x, a, \emptyset)$
  end while
  return y
end procedure

Markov chain Monte Carlo (MCMC)

MCMC is a class of algorithms for drawing samples $\boldsymbol{\theta}$ from a **target** distribution $p(\boldsymbol{\theta} \mid x)$ using an unnormalized density s.t.

$$p(\boldsymbol{\theta} \mid x) \propto \bar{p}(\theta \mid x).$$

procedure $\text{MCMC}(q, \bar{p}, k, T)$ $\quad\triangleright$ proposal q, target $\bar{p}$, cutoff k
  initialize $\boldsymbol{\theta}^{(0)}$
  for $t = 1 \ldots T$ **do**
    draw $\boldsymbol{\theta} \sim q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$
    compute acceptance probability a using $\bar{p}(\boldsymbol{\theta})$
    $\boldsymbol{\theta}^{(t+1)} \leftarrow \text{ACCEPT}(\boldsymbol{\theta}, a, \boldsymbol{\theta}^{(t)})$
  end for
  return $\boldsymbol{\theta}^{(k:T)}$
end procedure

For sufficiently large T, $\boldsymbol{\theta}^{(k:T)}$ will eventually converge to $p\left(\boldsymbol{\theta} \mid x\right)$.

Metropolis-Hastings algorithm

$$p_{\mathsf{fwd}} = \frac{\bar{p}\left(\boldsymbol{\theta} \mid x\right)}{q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right)}$$

$$p_{\mathsf{rev}} = \frac{\bar{p}\left(\boldsymbol{\theta}^{(t)} \mid x\right)}{q\left(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}\right)}$$

$$\begin{aligned} a &= \min\left(1, \frac{p_{\mathsf{fwd}}}{p_{\mathsf{rev}}}\right) \\ &= \min\left(1, \frac{\bar{p}\left(\boldsymbol{\theta} \mid x\right)\ q\left(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}\right)}{\bar{p}\left(\boldsymbol{\theta}^{(t)} \mid x\right)\ q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right)}\right) \end{aligned}$$

Gibbs sampling

Gibbs sampler always accepts ($a = 1$).

Define proposal q based on the full conditional distributions.

For parameters (θ, ϕ), we want to sample from $p(\theta, \phi \mid x)$.

procedure $\textsc{Gibbs}(p, k, T)$ $\triangleright$ full conditional p, cutoff k
  initialize $(\theta^{(0)}, \phi^{(0)})$
  for $t = 1 \ldots T$ **do**
    draw $\theta \sim p(\theta \mid x, \phi)$
    $\theta^{(t+1)} \leftarrow \theta$
    draw $\phi \sim p(\phi \mid x, \theta)$
    $\phi^{(t+1)} \leftarrow \phi$
  end for
  return $(\boldsymbol{\theta}^{(k:T)}, \boldsymbol{\phi}^{(k:T)})$
end procedure

Hamiltonian Monte Carlo

A variant of MCMC that uses Hamiltonian dynamics.

Used in the `stan` sampler.

Betancourt 2017. A conceptual introduction to Hamiltonian Monte Carlo. https://arxiv.org/abs/1701.02434

Summary

MCMC converges almost surely, as long as you are willing to wait indefinitely.

Intended learning outcomes

- Implement statistical models in Stan