Markov chain Monte Carlo (MCMC) methods

Gibbs Sampler

- The *Gibbs sampler* is a conditional sampling technique in which the acceptance-rejection step is not needed.
- The Markov transition rules of the algorithm are built upon conditional distributions derived from the target distribution.
- Suppose that the random variable can be decomposed into n components, i.e. $x = (x_1, \ldots, x_n)$. In Gibbs sampler, one randomly or systematically chooses a coordinate x_i and then substitutes it with x_i' drawn from

$$\pi(x_i \mid y, x_1, \ldots, x_{i-1}, x_{i+1}, x_{i+2}, \ldots, x_n),$$

  that is, the conditional posterior density of x_i.
- This conditional posterior usually is but does not have to be one-dimensional.

Markov chain Monte Carlo (MCMC) methods

Gibbs Sampler

Gibbs Sampler (Systematic Scan)

1. Let $k = 0$ and $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$.
2. Draw $x_i^{(k+1)}$ from the conditional density

   $$\pi(x_i \mid x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_{i+1}^{(k)}, \ldots, x_n^{(k)}),$$

   for $i = 1, \ldots, n$
3. Set $k = k + 1$, and if k is less than total number of iterations defined by the user, go back to the 2. step.

Gibbs Sampler (Random Scan)

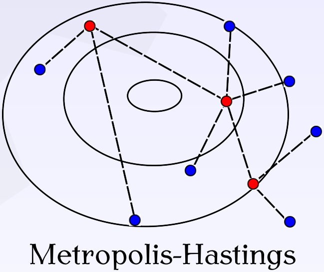
1. Let $t = 0$ and $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$.
2. Select a random uniformly distributed index i from the set $\{1, 2, \dots, n\}$.
3. Draw $x_i^{(k+1)}$ from the conditional distribution

$$\pi(x_i \mid x_1^{(k)}, \dots, x_{i-1}^{(k)}, x_{i+1}^{(k)}, \dots, x_n^{(k)}).$$

4. Set $k = k + 1$, and if k is less than total number of iterations defined by the user, go back to the 2. step.

Markov chain Monte Carlo (MCMC) methods

Gibbs Sampler



Differences between the moves of Metropolis-Hastings (M-H) and Gibbs sampler (GS). In the images, the actual samples have been visualized with red and the other proposed (M-H) and conditional sampling (GS) points with blue.

Balance Equation for Gibbs Sampler

The balance equation $\int \pi(x)A(x,y)\,dx = \int \pi(y)A(y,x)\,dx$ holds for Gibbs sampler, which can be shown as follows. Let us for simplicity assume that the systematic scan Gibbs sampler is in question and each conditional density is a one-dimensional one. The transition function is given by

$$A(x,y) = \prod_{i=1}^{n} \pi(y_i \,|\, y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n).$$

Integrating $A(y,x)$ with respect to the last coordinate x_n gives

$$\int_{\mathbb{R}} A(y,x) dx_n = \int_{\mathbb{R}} \prod_{i=1}^{n} \pi(x_i \,|\, x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n) dx_n$$

Balance Equation for Gibbs Sampler continued

$$= \prod_{i=1}^{n-1} \pi(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n) \int_{\mathbb{R}} \pi(x_n \mid x_1, \ldots, x_{n-1}) dx_n$$

$$= \prod_{i=1}^{n-1} \pi(x_i \mid x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n).$$

The last equality follows from $\int_{\mathbb{R}} \pi(x_n \mid y_1, \ldots, y_{n-1}) dx_n = 1$. Repeating this integration inductively with respect to $x_{n-1}, x_{n-2}, \ldots, x_1$, it follows that $\int_{\mathbb{R}^n} A(y, x) dx = 1$. Thus, we have

$$\int_{\mathbb{R}^n} \pi(y) A(y, x) dx = \pi(y) \int_{\mathbb{R}^n} A(y, x) dx = \pi(y).$$

Balance Equation for Gibbs Sampler continued

Considering now the right-hand side of the balance equation, and observing that $A(x,y)$ is independent of x_1, we have

$$\begin{aligned}\int_{\mathbb{R}^n} \pi(x)A(x,y)dx_1 &= A(x,y)\int_{\mathbb{R}^n} \pi(x)dx_1 \\ &= A(x,y)\pi(x_2,x_3,\ldots,x_n),\end{aligned}$$

and further $\int_{\mathbb{R}} \pi(x)A(x,y)dx_1 =$

$$\left(\prod_{i=2}^{n} \pi(y_i \,|\, y_1,\ldots,y_{i-1},x_{i+1},\ldots,x_n)\right)\pi(y_1 \,|\, x_2,\ldots,x_n)\pi(x_2,\ldots,x_n).$$

Balance Equation for Gibbs Sampler continued

$$= \left(\prod_{i=2}^{n} \pi(y_i \,|\, y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n) \right) \pi(y_1, x_2, x_3, \ldots, x_n).$$

Repeating this inductively for $x_2, x_3, \ldots, x_n$, and taking into account that

$$\int_{\mathbb{R}^i} \pi(y_1, \ldots, y_{i-1}, x_i, \ldots, x_n) \pi(y_i \mid y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n) dx_i$$
$$= \pi(y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n) \pi(y_i \mid y_1, \ldots, y_{i-1}, x_{i+1}, \ldots, x_n)$$
$$= \pi(y_1, \ldots, y_{i-1}, y_i, x_{i+1}, \ldots, x_n),$$

it follows that $\int_{\mathbb{R}^n} \pi(x) A(x, y) dx = \pi(y)$, showing that the balance equation holds.

Markov chain Monte Carlo (MCMC) methods

Gibbs Sampler

- In general, one step of Gibbs sampler (GS) requires more work than that of the Metropolis-Hastings (M-H) algorithm, since the former is likely to require more point evaluations of the posterior density.
- However, subsequent points produced by GS are usually less mutually correlated than those produced by M-H, i.e. the sample ensemble of a given size is typically better distributed according to the posterior in the case of GS than that of M-H.
- Sampling from a conditional density in Gibbs Sampler typically requires findind the essential part of the density due to which implementation can be difficult.

Sampling from one-dimensional density

One can draw x distributed according to the probability distribution $\Phi : \mathbb{R} \to [0,1]$, $\Phi(t) = \int_{-\infty}^{t} \pi(\tau)\, d\tau$, through the following steps:

1. Draw u from Uniform($[0,1]$).
2. Set $x = \Phi^{-1}(u)$.

- If the integral $\Phi(t) = \int_{-\infty}^{t} \pi(\tau)\, d\tau$ is computed for a set of points $t_1, t_2, \ldots, t_m$ covering the essential part of the support of π, the above algorithm can be used to numerical sampling.

Numerical sampling from one-dimensional density

One can draw x distributed according to the probability distribution $\Phi : \mathbb{R} \to [0, 1]$, $\Phi(t) = \int_{-\infty}^{t} \pi(\tau)\, d\tau$, through the following steps

1. Find a set of points $t_1, t_2, \ldots, t_m$ covering the essential part of the support of π.
2. Compute the integral $\Phi(t) = \int_{-\infty}^{t} \pi(\tau)\, d\tau$ for each point $t_1, t_2, \ldots, t_m$.
3. Draw u from Uniform([0, 1]).
4. Find the smallest index i for which $\Phi(t_i) > u$ and set $x = t_i$.

Example 9 (Matlab)

Draw and visualize a sample ensemble consisting of 100000 points from a probability density

(a) $\pi(\tau) \propto \cos(\pi\tau/2)$

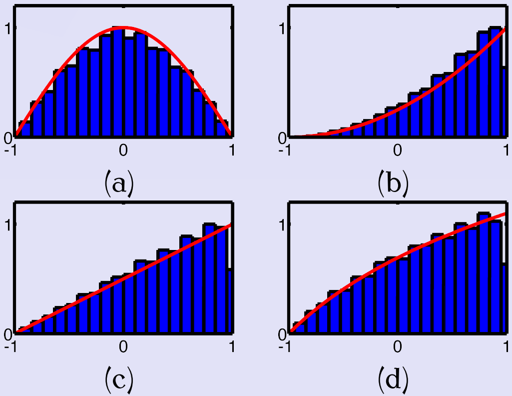
(b) $\pi(\tau) \propto (\tau + 1)^2/4$

(c) $\pi(\tau) \propto (\tau + 1)/2$

(d) $\pi(\tau) \propto \log(\tau + 2)$

with $\tau \in [-1, 1]$. Use the algorithm of the previous page. Visualize the results using histograms.

Example 9 (Matlab) continued



Blue = Histogram, Red = Exact.

Example 10 (Matlab)

Repeat the sampling procedures of Example 9 using Gibbs Sampler. Compare the results to the ones obtained with the random walk Metropolis with Gaussian proposals. Observe that the Gibbs sampler produces faster moving Markov chain than Metropolis-Hastings.

Solution

Samples from each one-dimensional conditional density were produced numerically by dividing intersection of the unit disk and the line corresponding to conditional density into a 200 equally spaced points, and then the previously given one-dimensional sampling algorithm was used.

Markov chain Monte Carlo (MCMC) methods

Gibbs Sampler

Example 10 (Matlab) continued

Case I: Three sensors, $\sigma = 0.1$

Posterior density Sample ensemble

horizontal component vertical component

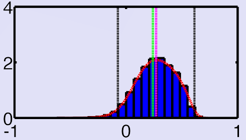
purple = exact, green = conditional mean, black = 95 % credibility

Markov chain Monte Carlo (MCMC) methods

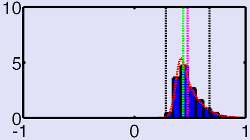
Gibbs Sampler

Example 10 (Matlab) continued

Case I: Three sensors, $\sigma = 0.1$

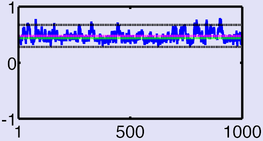


Marginal density (horiz.)

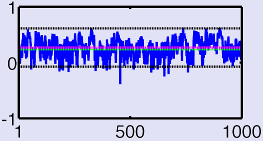


Marginal density (vertical)

blue = sample based, red = exact



Burn in sequence (horiz.)

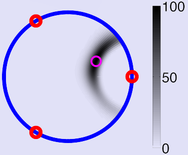


Burn in sequence (vertical)

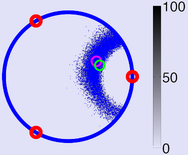
purple = exact, green = conditional mean, black = 95 % credibility

Example 10 (Matlab) continued

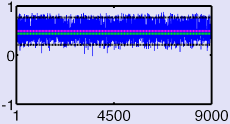
Case II: Three sensors, $\sigma = 0.2$



Posterior density

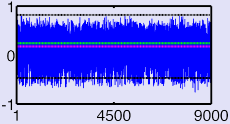


Sample ensemble



STD = 0.13864

horizontal component



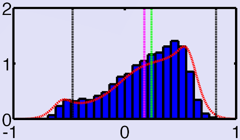
STD = 0.32186

vertical component

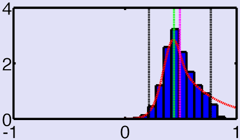
purple = exact, green = conditional mean, black = 95 % credibility

Example 10 (Matlab) continued

Case II: Three sensors, $\sigma = 0.2$

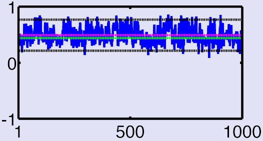


Marginal density (horiz.)

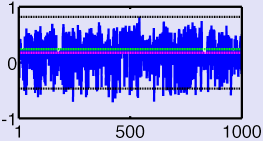


Marginal density (vertical)

blue = sample based, red = exact



Burn in sequence (horiz.)



Burn in sequence (vertical)

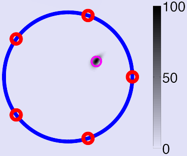
purple = exact, green = conditional mean, black = 95 % credibility

Markov chain Monte Carlo (MCMC) methods

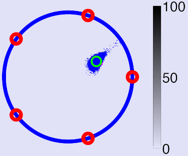
Gibbs Sampler

Example 10 (Matlab) continued

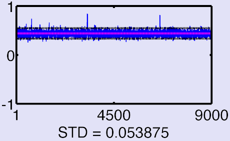
Case III: Five sensors, $\sigma = 0.1$



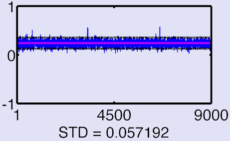
Posterior density



Sample ensemble



horizontal component



vertical component

purple = exact, green = conditional mean, black = 95 % credibility

Markov chain Monte Carlo (MCMC) methods

Example 10 (Matlab) continued

Case III: Five sensors, $\sigma = 0.1$

Marginal density (horiz.) Marginal density (vertical)

blue = sample based, red = exact

Burn in sequence (horiz.) Burn in sequence (vertical)

purple = exact, green = conditional mean, black = 95 % credibility

Markov chain Monte Carlo (MCMC) methods

Example 10 (Matlab) continued

Case IV: Five sensors, $\sigma = 0.2$

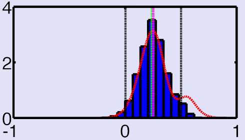
Posterior density   Sample ensemble

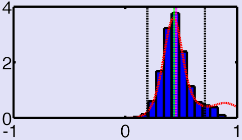
horizontal component   vertical component

purple = exact, green = conditional mean, black = 95 % credibility

Example 10 (Matlab) continued

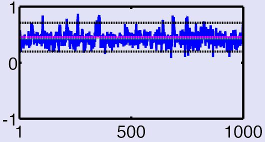
Case IV: Five sensors, $\sigma = 0.2$

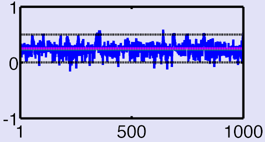




Marginal density (horiz.) Marginal density (vertical)

blue = sample based, red = exact





Burn in sequence (horiz.) Burn in sequence (vertical)

purple = exact, green = conditional mean, black = 95 % credibility

Markov chain Monte Carlo (MCMC) methods

Gibbs Sampler

Example 10 (Matlab) continued

- Based on the burn in sequence, it is clear that the Gibbs sampler produces a faster moving Markov chain than the Metropolis-Hastings, i.e. the sample points are less mutually correlated in the case of Gibbs sampler.
- A rule of thumb is that the more the sampling history for a single coordinate looks like a "fuzzy worm" the more independent are the sampling points and the better is the Markov chain in general with regard to convergence of the estimates.