StatsM254 Statistical Methods in Computational Biology	Lecture 11 - 05/08/2014

Lecture 11

Lecturer: Jingyi Jessica Li *Scribe: Kai Fu*

1 Markov Chain Monte Carlo (MCMC)

1. Monto Carlo Simulator
   Goal: evaluate $E[f(X)]$ for $X \sim P$ (target distribution) sample $x_1, \ldots, x_n$ as i.i.d. from P and calculate $\frac{1}{n}\sum_{i=1}^{n} f(x_i)$
   (a) vanilla MC
   (b) rejection sampling
   (c) importance sampling
2. MCMC VS MC
   Construct a i.i.d. markov chain $x_1, \ldots, x_n$. Estimate θ as $\hat{\theta} = \frac{1}{n-k}\sum_{i=k+1}^{n} f(x_i)$, the chunk that is thrown away is called the burn-in period
3. Background: First-order Markov Chain
   (a) $x_1, x_2, \ldots, x_n, x_{n+1}, \ldots$
   (b) First order
   $P(x_{n+1}|x_1, \ldots, x_n) = P(x_{n+1}|x_n)$
   (c) Invariant distribution
   Π is the probability. π is the density
   $\Pi(dy) = \int T(x, dy)\pi(x)\, dx$
   $T(x, dy)$ is called transition probability
   e.g. in the discrete case, $\Pi = \pi$, $x_i \in \{1, 2\}$,
   $\pi(x_{n+1} = 2) = \sum_{i=1}^{2} P(x_{n+1} = 2|x_n = i) \cdot \pi(x_n = i) = \sum_{i=1}^{2} T(i, 2) \cdot \pi(x_n = i)$
   (d) Transition probability
   $T(x, dy) = P(x_{n+1} \in dy|x_n = x)$
   (e) Markov chain converges to invariant distribution
   Transition probability of different orders: For starting value x, we have
   $p^{(1)}(x, A) = T(x, A)$
   $p^{(2)}(x, A) = \int p^{(1)}(x, dy)T(y, A)$
   $p^{(3)}(x, A) = \int p^{(2)}(x, dy)T(y, A)$
   $\vdots$
   $p^{(n)}(x, A) = \int p^{(n-1)}(x, dy)T(y, A) \approx \Pi(A)$
   (f) Markov Chain theory is mainly concerned about: for a given $T(x, dy)$, what is Π?
   (g) MCMC goes backwards: given a marginal distribution (target distribution) Π, can we create a Markov chain with some $T(x, dy)$ that Π is the invariant distribution?
   (h) "reversibility" criterion
   $\pi(x) \cdot t(x, y) = \pi(y) \cdot t(y, x)$, where $t(x, y) = \frac{d}{dy}T(x, dy)$
   $\Rightarrow \quad \int T(x, A)\pi(x)\, dx = \int\int_A t(x, y)dy\pi(x)dx$
   $= \int_A \int t(x, y)\pi(x)\, dx\, dy$
   $= \int_A \int t(y, x)\pi(y)\, dx\, dy$
   $= \int_A \left(\int t(y, x)\, dx\right) \pi(y)\, dy = \int_A \pi(y)dy = \pi(A)$

4. Setup of MCMC
   (a) Π is known
   (b) how to construct $T(x, dy)$?
   Suppose we take any conditional probability $q(x, y)$, e.g. $q(x,y)=f(y|x)=\phi(y-x)$ and we have
   $\pi(x) \cdot q(x, y) > \pi(y) \cdot q(y, x)$
   we "fudge" $q(x, y)$ by multiplying a "fudge" factor, $\alpha(x, y) \leq 1$ such that
   $\pi(x)q(x, y)\alpha(x, y) = \pi(y)q(y, x)\alpha(y, x)$
   (LHS) (RHS)
   Theorem $\alpha(x, y) = \min[\frac{\pi(y)\cdot q(y,x)}{\pi(x)\cdot q(x,y)}, 1]$
   Proof. When $\pi(x)q(x, y) < \pi(y)q(y, x)$
   $\Rightarrow \alpha(x, y) = 1$, $\alpha(y, 1) = \frac{\pi(x)\cdot q(x,y)}{\pi(y)\cdot q(y,x)}$
   so LHS=$\pi(x)q(x, y)$; RHS=$\pi(x)q(x, y)$
   When $\pi(x)q(x, y) > \pi(y)q(y, x)$, can prove LHS=RHS in a similar way

2 The Metropolis-Hasting algorithm (MH)

Given an (arbitrary) starting value X_1, generate X_2 as follows.

- Sample Y from the conditional density $q(x_1)$ and $U \sim Unif(0, 1)$, $Y \perp U$.
- If $U \leq \alpha(X_1, Y)$, accept the candidate Y and set $X_2 = Y$
- Else reject the candidate Y and set $X_2 = X_1$

3 The Gibbs Sampler

1. We want to samples $x = (x^{(1)}, \cdots, x^{(m)}) \sim P$, the joint distribution is complicated
2. sample each $x^{(i)}$ conditional on others, that is, in iteration $(n + 1)$,
   $x^{(1)}_{n+1} \sim P(x^{(1)}|x^{(2)}_n, x^{(3)}_n, \cdots, x^{(m)}_n)$
   $x^{(2)}_{n+1} \sim P(x^{(2)}|x^{(1)}_{n+1}, x^{(2)}_n, \cdots)$
   $\vdots$
   $x^{(m)}_{n+1} \sim P(x^{(2)}|x^{(1)}_{n+1}, \cdots, x^{(m-1)}_{n+1})$
3. Gibbs sampler is useful because conditional distributions are often much simpler
4. Relationship to Metropolis-Hasting
   Gibbs sampler is in fact an MH algorithm with the conditional distribution:
   $q((x^{(i)}_n, x^{(-i)}), (x^{(i)}_{n+1}, x^{(-i)})) = P(x^{(i)}_{n+1}|x^{(-i)})$ The "fudge" factor (acceptance probability):
   $\alpha((x^{(i)}_n, x^{(-i)}), (x^{(i)}_{n+1}, x^{(-i)}))$
   $= \frac{\pi\left(x^{(i)}_{n+1}, x^{(-i)}\right)\cdot p\left(x^{(i)}_n|x^{(-i)}\right)}{\pi\left(x^{(i)}_n, x^{(-i)}\right)\cdot p\left(x^{(i)}_{n+1}|x^{(-i)}\right)}$
   $= \frac{p\left(x^{(-i)}\right)\cdot p\left(x^{(i)}_{n+1}|x^{(-i)}\right)\cdot p\left(x^{(i)}_n|x^{(-i)}\right)}{p\left(x^{(-i)}\right)\cdot p\left(x^{(i)}_{n+1}|x^{(-i)}\right)\cdot p\left(x^{(i)}_n|x^{(-i)}\right)}$
   $= 1$

4 Critique

Draw from the points discussed in class. Write the critques in about a paragraph for each paper.

5 Possible Extensions

6 Conclusions

References

[1] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, M. and J. Crowcroft, "XORs in the air: practical wireless network coding", *IEEE/ACM Transactions on Networking,*, vol. 16, no. 3, pp. 497–510, 2008.

[2] H. Rahul, N. Kushman, D. Katabi, C. Sodini, and F. Edalat, "Learning to Share: Narrowband-Friendly Wideband Wireless Networks", *ACM SIGCOMM Computer Communication Review*, vol. 38, no. 4, pp. 147–158, 2008.