HGEN 48600/STAT 35450: Fundamentals of Computational Biology

Lecture 7 — January 26

Lecturer: Prof. Stephens *Scribe: Rahul S. and Daniel Z.*

Note: These lecture notes are still rough, and have only have been mildly proofread.

7.1 Markov Chain Monte Carlo

Given $X, Y \in 0, 1$ such that $P(X = x, Y = y) = :$

X / Y	0	1
0	0.6	0.1
1	0.15	0.15

An alternative way to simulate from this joint distribution using the conditional probabilities would be:

$$Pr(X = 0|Y = 0) = \frac{0.6}{0.75} = 0.8 \tag{7.1}$$

$$Pr(X = 1|Y = 0) = \frac{0.6}{0.75} = 0.2 \tag{7.2}$$

$$Pr(X = 0|Y = 1) = \frac{0.1}{.25} = 0.4 \tag{7.3}$$

$$Pr(X = 1|Y = 1) = \frac{.3}{.5} = .6 \tag{7.4}$$

Method:
Set $x_0 = 0$, $y_0 = 0$,
Iterate:
at step $i = 1, 2, 3..$

- simulate x_i from the conditional distribution $Pr(x|y = y_{i-1})$
- simulate x_i from the conditional distribution $Pr(y|x = x_{i-1})$

The theory: for large enough N: $Pr(X_n = x_n, Y_n = y_n) = Pr(X = x, Y = y)$ based only on the conditional probabilities.
In this example we simulated a Markov Chain. The stationary distribution of this Markov Chain converges to $Pr(X = x, Y = y)$.
Chain $c_1, c_2, c_3,$. where $c_1 = (x_1, y_1), c_2 = (x_2, y_2), c_3 = (x_3, y_3)$. $c_i$only depends on c_{i-1} and is therefore a Markov Chain.

7.2 Proof

Proof of convergence to the stationary distribution $Pr(X = x, Y = y)$ (intuitive explanation):
Once we reach a stationary distribution π we remain there: $\pi * P = P$. Suppose we reached the stationary distribution s.t. $Pr(X = x, Y = y)$: At the first step of the iteration:

$$Pr(X_i = x, Y_i = y) = Pr(X_{i=x}|Y_{i=y})Pr(Y_{i=y}) \tag{7.5}$$

$$= Pr_{x|y}(X_i = x|Y_i = y)Pr_y(Y = y) \tag{7.6}$$

= (Conditional distribution from the table)*(marignal distribution from table)

$$Pr(X_i = x, Y_i = y) = Pr(X_i = x|Y_{i=y})Pr(Y_{i=y}) \tag{7.7}$$

$$= Pr_{x,y}(X = x, Y = y) \tag{7.8}$$

As long as x and y come from the right marginal after step 1 and step 2 x_{i+1}, y_{i+1} are from the correct target distribution. As such, if we start the algorithm from the stationary distribution we stay at the joint distribution.
note: this Markov Chain is irreducible, recurrent, finite and aperiodic therefor it will converge to its stationary distribution. There are different definitions for the convergence of distributions such as almost surely.
This proof can be extended to N variables e.g. x,y,z: a) simulate x conditional on y and z b) simulate y conditional on x and z c) simulate z conditional on x and y
alternatively, we can use the algorithms for subsets of variables: a)simulate x given y and z b)simulate y and z given x
our proof still works when we add conditionals which is useful for Bayesian methods.
Example: Genetic data on elephant tusks: Some elephants are from a savanna population some from a forest population, no prior knowledge on which is which. Cluster data based on a mixture model using Haploid Elephants markers:
Data: $x = x_1, .x_n$ n tusks π_1, π_2= proportion from each of the two groups F_1, F_2 = marker allele frequency in the two groups Latent variables: $z = z_1, .z_n$ where z_i is the group origin of tusk i

"Complete" Data Likliehood: $Pr(x, z|F, \pi) =$

$$= \prod_{i=1}^{n} P(X_i|Z_i, F, \pi) * P(Z_i|f, \pi) \tag{7.9}$$

$$Likelihood = P(x|f, \pi) = \pi_i * \sum_k \pi_k P(X_i|Z_i, F, \pi) \tag{7.10}$$

7.3

7.3.1 Likelihood calculation

Complete data likelihood.

$$\prod_{i=1}^{n} P(z_i|F, \pi) \cdot P(x_i|z_i) = \prod_{i=1}^{n} \pi_{z_i} \cdot \prod_j F_{k1j}^{k_j} \cdot (1 - F)^{1-k_j} \tag{7.11}$$

product across individual i marker j where F_{k1j} is the frequency of an allele of a of a marker in a given population group k Likelihood.

$$Likelihood = P(x|F, \pi) = \prod_{i=1} (\sum_k \pi_k \cdot P(x_i|z_i = k, F, \pi)) \tag{7.12}$$

to simulate data decide on group 1 or 2, and simulate alleles.

7.3.2 Gibbs sampling

Gibbs sampling for $P(z, pi, Fx)$:
1) sample from $z|\pi, F, (x)$ (use current value of π and F to generate z group origins)
2) sample from $\pi, F|z, (x)$

where $\pi, F|z, (x) \propto P(\pi, F|Z, x) \propto P(\pi)P(F)P(z, x|\pi, F)$
if $\pi \sim Be(\alpha_\pi, \beta_\pi)$ is the prior then:
$\pi|z, x \sim Be(\alpha_{pi} + \#k_i = 1, \beta_\pi = \sum z_i = z$

$\pi|z, x \sim Be(\alpha_\# Fj + \#, \beta_{k,j} + \#0_s...$ (allele number j from pop. k

$$p(\pi, f|z, x) \alpha p(\pi) * p(f) * p(z, x|pi, f) \tag{7.13}$$

this algorithm converges to a distribution rather than a point estimate which allows for calculating confidence bounds.

Gibbs sampling is one way of generating a markov chain when we can directly sample from the conditional probabilities but not from the joint distribution

7.4 Metropolis Hastings Algorithm

How do we generate a sample from a Markov chain with stationary distribution $\pi(x)$

1) π is reffered to as the target distribution and has to be defined up to a constant of proprtion α
2) Makrov transition matrix Q (also known as the "kernel") from which it is possible to simulate from: s.t. for any given current state of x you can simulate the next state of the Markov Chain with transition matrix Q. We have to know how to compute Q or what is the probability of transitioning.

7.4.1 algorithm

Providing Q, π
i) initialize $x \in X$
ii) at step i=1,2 ....

- let x be the current value of $x = x_{i-1}$
- generate a proposed value of x, x' by simulating one step of Q
- with probablity A set $x = x'$ otherwise set $x_i = x(x_i = x_{i-1})$ where

$$A = \frac{\pi(x')Q(x' -> x)}{pi(x)Q(x -> x')} \bigwedge 1 \tag{7.14}$$

the $\frac{Q()}{Q()}$ part is called the hastings part
this is a bit reminiscent of the detalied balance equation

7.4.2 example

Example of Q: add a small random deviate to x if $\pi \sim exp(1)$ and Q adds random normal the probability of moving back and forth $+ - 1$ is the same since Q is normal which is symetric