Applied algebraic geometry: algebraic statistics

Jan Draisma, Universität Bern

Metropolis-Hastings

Setup

- finite, connected graph $G = (V, E)$ without loops or mult edges
- a target probability distribution $\pi : V \to \mathbb{R}_{>0}$ on V.

Goal

Sample vertices according to π.

Algorithm

- for each $x \in V$ choose a probability distribution q_x on $\{y \in V \mid y \sim x\}$ such that $q_x(y) > 0$ for all adjacent vertices x, y.
- choose any $x_0 \in V$ and set $t := 0$
- repeat indefinitely: - set $x := x_t$, choose $y \sim x$ according to q_x

- set $\alpha_x(y) := \min\{\pi(y)q_y(x)/\pi(x)q_x(y), 1\}$, the *acceptance ratio*
- set $x_{t+1} := y$ with probability α and $x_{t+1} := x$ otherwise
- increase t by 1.

Metropolis-Hastings

Theorem
Under mild conditions, $\lim_{t\to\infty} \text{Prob}(x_t = x) = \pi(x)$.

Proof is not so difficult, but needs a bit of Perron-Frobenius theory. This yields that there is a unique limiting distribution π' on V, independent of the starting point x_0. Let's argue why it equals π:

$\text{Prob}(x_{t+1} = y) = \sum_{x\sim y} \text{Prob}(x_t = x) q_x(y)\alpha_x(y)$
$+\text{Prob}(x_t = y)(1 - \sum_{x\sim y} q_y(x)a_y(x))$.
So π' satisfies
$\pi'(y) = \sum_{x\sim y} \pi'(x)q_x(y)\alpha_x(y) + \pi'(y)(1 - \sum_{x\sim y} q_y(x)a_y(x))$.
We claim that π is the (unique) probabilitity distribution satisfying this equation:

$\sum_{x\sim y} \pi(x)q_x(y)\alpha_x(y) = \sum_{x\sim y} \min\{\pi(x)q_x(y), \pi(y)q_y(x)\} =$
$\sum_{x\sim y} \pi(y)q_y(x)\alpha_y(x) = \pi(y) \sum_{x\sim y} q_y(x)\alpha_y(x)$ □

Metropolis-Hastings

Remarks

- Very important: the acceptance ratio $\alpha_x(y)$ only depends on q_x, q_y and the *ratio* $P(y)/P(x)$. In applications, P is often hard to evaluate but the ratio isn't.

- For $t \gg 0$, x_t is a good sample from π. How good, is not so easy to analyse—much work being done.

- We will discuss an application of MH due to Diaconis-Sturmfels, where the *graph* is constructed using commutative algebra.

The model of independence

Setting
Y, Z random variables taking values in $[r], [s]$, respectively.

Joint distribution given by the $m \times n$-matrix $P = (p_{ij})_{ij}$ defined by $p_{ij} = \mathrm{Prob}(Y = i, Z = j)$. This is a matrix $P \in \mathbb{R}_{\geq 0}^{r \times s}$ with $\sum_{i,j} p_{ij} = 1$.

Observation
X, Y are *independent*, i.e., $p_{ij} = \mathrm{Prob}(Y = i) \cdot \mathrm{Prob}(Z = j)$ if and only if $\mathrm{rk} P = 1$.

Definition
The *model of independence* is the semi-algebraic set $X := \{P \in \mathbb{R}_{\geq 0}^{r \times s} \mid \mathrm{rk}(P) \leq 1, \sum_{i,j} p_{ij} = 1\}$. Similarly for d random variables (get certain rank-one d-tensors).

Fisher's test

Example from wikipedia

	Men	Women	Row total
Studying	1	9	10
Not-studying	11	3	14
Column total	12	12	24

Think of these numbers as independent draws from a fixed, but unknown joint distribution P. **Question**: should we believe that the two variables Y, Z are independent, in view of so few men among the studiers?

Fischer's approach: assuming that they are, this gives an (unknown) probability distribution on the set of 2×2 tables with values in $\mathbb{Z}_{\geq 0}$ and total entry sum 24.

Fisher's test

Explicitly: $\text{Prob}(\begin{bmatrix} a & b \\ c & d \end{bmatrix}) = t_1^{a+b} t_2^{c+d} u_1^{a+c} u_2^{b+d} \underbrace{\binom{a+b+c+d}{a; b; c; d}}_{(*)}$

where $t_i = \text{Prob}(Y = i)$ and $u_j = \text{Prob}(Z = j)$

Now *condition* on the row sums and column sums. This gives the probability of observing the table *conditional* on seeing $a + c$ men, $b + d$ women, $a + b$ studiers and $c + d$ non-studiers. To do so, you divide the expression above by the sum over those expressions with fixed row and column sums. All that matters for us later, is that the result is proportional to the multinomial coefficient above (and independent of t_i, u_j).

Fisher proposes: *compute the probability of at most* 1 *male studier when drawing from this conditional distribution*. If this is small, reject the null hypothesis of independence.

Enter Diaconis-Sturmfels

In the example, this can be done explicitly, since a is either 0 or 1 and this determines the entire table. The probability computed by Fisher equals roughly 0.001379728, so one probably rejects.

But what if the table is much larger, and the set of tables in Fisher's approach is too large to enumerate completely?

Proposal by Diaconis-Sturmfels: run the Metropolis-Hastings algorithm on the set of such tables, with probability proportional to (*) (remember that only *ratios* of probabilities are needed).

Then the proportion of the time that x_t has fewer than a studying men is a good estimate for the probability in Fisher's test.

They did this for a table of birth months and death months for 82 descendants of Queen Victoria. Here $r = s = 12$.

A Markov basis

So from the graph $G = (V, E)$ we know V: the $r \times s$-tables with fixed row and column sums. But what are the *edges*?

Simple observation
Adding/subtracting to a table
a table of the form:
does not change the row/column sums.

$$\begin{bmatrix} +1 & -1 \\ -1 & +1 \end{bmatrix}$$

Naturally, may only do this when it does not create negative entries. Take these moves as edges of the graph.

Proposition
This graph is connected. In other words, any two tables $M, M' \in \mathbb{Z}_{\geq 0}^{r \times s}$ with the same row and column sums are connected by a path inside $\mathbb{Z}_{\geq 0}^{r \times s}$ of such moves.

A Markov basis

Proof

Encode M as a sequence:

(i_1, j_1)
(i_2, j_2)
$\vdots$

where $i_1 \leq i_2 \leq \ldots$ and where $m_{ij} = \#\{l : (i_l, j_l) = (i, j)\}$

Suppose $j_l > j_{l+1}$. Then if $i_l = i_{l+1}$ swap the l-th and $(l + 1)$st entry; the sequence represents the same M. If, on the other hand, $i_l < i_{l+1}$, subtract 1 from positions $(i_l, j_l), (i_{l+1}, j_{l+1})$ in M and add 1 to positions (i_l, j_{l+1}) and (i_{l+1}, j_l). In the sequence, this corresponds to swapping j_l, j_{l+1}, and in the matrix it is an allowed move.

Eventually, the j's are also ordered weakly. So we have connected M to the unique M'' with same row and column sums and increasing sequences i_l and j_l. $\square$

The Diaconis-Sturmfels theorem

In more complicated models, do such connected graphs exist?

Definition
Let $A \in \mathbb{Z}^{m\times n}_{\geq 0}$. A *Markov basis* for A is a subset S of $\ker A : \mathbb{Z}^n \to \mathbb{Z}^m$ with the property that if $u, v \in \mathbb{Z}^n_{\geq 0}$ satisfy $Au = Av$, then there exists a sequence $u_0 = u, u_1, \ldots, u_k = v$ in $\mathbb{Z}^n_{\geq 0}$ such that $u_i - u_{i+1} \in \pm S$ for all i.

In our 2 × 2-table example, $m = n = 4$ and A looks like this:

	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$
row 1	1	1	0	0
row 2	0	0	1	1
column 1	1	0	1	0
column 2	0	1	0	1