Markov Chain Monte Carlo, contingency tables and Gröbner bases

Gunnar Englund

Diaconis, P., Sturmfels, B. (1998).

Algebraic algorithms for sampling from

conditional distributions.

Annals of Statistics Vol. 26.

MCMC: A method to simulate from complicated distributions.

The distributions need not be normed.

Great applications in Bayesian statistics

$$P(\Theta = \theta | X = x) = \frac{P(X = x|\Theta = \theta)P(\Theta = \theta)}{P(X = x)} \propto$$

$$\propto P(X = x|\Theta = \theta)P(\Theta = \theta) = P(X = x; \Theta = \theta)$$

i.e.

a-posteriori-distribution $\propto$

likelihoodfunction $\times$ apriori-distribution

"Hot" area

<u>About MCMC:</u>

We want to simulate distribution

$\propto \pi(x),\ x \in E$

For simplicity:

E finite (but gigantic)

$$P(X = x) = \frac{\pi(x)}{\sum_{z \in E} \pi(z)}$$

Norming difficult but unneccessary

Metropolis-Hastings algorithm:

Gives Markov-chain with $p_X(x),\ x \in E$ (or equivalently $\pi(x)$, $x \in E$) as stationary distribution.

1) If $X_n = x$: Proposal y for X_{n+1} according to proposal distribution $q(x,y)$.

Great freedom in choice of $q(x,y)$.

2) Accept proposal y given from x with probability

$$\alpha(x,y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right)$$

else $X_{n+1} = x$

If $x \neq y$ the tranisition probability

$$P(x,y) = q(x,y)\alpha(x,y)$$

$$P(x,x) = 1 - \sum_{y \in E;\ y \neq x} P(x,y)$$

The chain is reversible i.e.

$$\pi(x)P(x,y) = \pi(y)P(y,x), \quad \text{for all } x,y$$

Trivial for $x = y$. For $x \neq y$

$$\text{LHS} = \pi(x)q(x,y)\alpha(x,y) =$$

$$= \min(\pi(x)q(x,y), \pi(y)q(y,x)) = \pi(y)P(y,x)$$

Reversability yields (summation over x)

$$\sum_{x \in E} \pi(x)P(x,y) = \sum_{x \in E} \pi(y)P(y,x) =$$

$$= \pi(y) \sum_{x \in E} P(y,x) = \pi(y)$$

Hence π is stationary distribution.

(Compare $\boldsymbol{p}^{(1)} = \boldsymbol{p}^{(0)}\boldsymbol{P}$)

NB! π appears both in denominator and numerator in $\alpha(x,y)$! Need not be normed!!

If chain is ergodic:

$$\frac{1}{B}\sum_{i=1}^{B} g(X_i) \to E_\pi(g(X)) =$$

$$= \frac{\sum_{x \in E} g(x)\pi(x)}{\sum_{x \in E} \pi(x)} \text{ when } B \to \infty$$

for arbitrary g. Everything can be calculated.

Ergodic if irreducable, aperiodic (not needed if you do not take subsequences!) Easy to verify in specific situations.

Puts restrictions on proposal distribution $q(x,y)$. Gröbner-bases can be used to verify irreducability.

But: $q(x,y)$ should be chosen cleverly so that B need not be too large.

Example: Let $\chi = I \times J$-lattice of zeroes and ones.

$E = \{x \in \chi$ two neighbours not both 1$\}$.

π distribution on χ but 0 outside E, e.g. $\pi(x)$=constant for $x \in E$ and 0 elsewhere (uniform distribution on E)

Proposal distribution (i.e. $q(x,y)$):

Choose node (i,j) at random with probability $1/IJ$. Choose value 0 or 1 in the node with probability 1/2 each.

Possibly $y \notin E$, but then $\pi(y) = 0$ and therefore $\alpha(x,y) = 0$.

If $y \in E$ proposal is accepted with probability

$$\alpha(x,y) = \min\left(1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right) = 1$$

since $q(x,y) = q(y,x)$ and $\pi(x) = \pi(y)$.

Chain is obviously irreducable and aperiodic and hence ergodic.

E.g. with

$g(x)$ =number of 1:s in x we get average number of 1:s!

With $I = J = 10$ we get ($B = 10000$)

E(number of 1:s) ≈ 23.40,

E(max number of 1:s in a row) ≈ 3.88.

Their correlation ≈ 0.4240.

Almost as easy with general π but then acceptance probability

$$\alpha(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$$

Example: Contingency table

$I \times J$-table of counts with

rowsums $\boldsymbol{r} = (r_1, r_2, \cdots, r_I)$

and columnsums $\boldsymbol{c} = (c_1, c_2, \cdots, c_J)$.

$$A(\boldsymbol{r}, \boldsymbol{c}) = \{I \times J\text{-tables with } \boldsymbol{r} \text{ and } \boldsymbol{c}\}$$

We want to study $\boldsymbol{X} = (x_{ij}, (i, j) \in I \times J)$ with given distribution on $A(\boldsymbol{r}, \boldsymbol{c})$.

Especially interesting is the "hypergeometric" distribution on $A(\boldsymbol{r}, \boldsymbol{c})$ i.e. with $N = \sum_{i,j} x_{ij}$

$$P(\boldsymbol{X} = \boldsymbol{x}) = \frac{\prod_{j=1}^{J} \frac{c_j!}{x_{1j}! x_{2j}! \cdots x_{Ij}!}}{\frac{N!}{r_1! r_2! \cdots r_I!}}$$

Compare 2×2-table with cell counts

$x_{11}, x_{12}, x_{21}, x_{22}$ where

$$P(X_{11} = x_{11}) = \frac{\binom{c_1}{x_{11}}\binom{c_2}{x_{12}}}{\binom{N}{r_1}}$$

If H_0 : "rows and columns are independent" is true $\boldsymbol{X}$ has a "hypergeometric" distribution.

Test variable

$$Q = \sum_{i,j} \frac{(x_{ij} - Np^*_{ij})^2}{Np^*_{ij}}$$

is approximately $\chi^2((I-1)(J-1))$-distributed where $p^*_{ij} = r_i/N \cdot c_j/N$.

Good approximation?

With MCMC: If $X_n = x \in A(\boldsymbol{r}, \boldsymbol{c})$ we let proposal distribution be

$q(x, y)$: Choose two rows and two columns at random

Choose patterns

$$\begin{pmatrix} + & - \\ - & + \end{pmatrix} \text{ och } \begin{pmatrix} - & + \\ + & - \end{pmatrix}$$

with probability 1/2 each.

If $y \notin A(\boldsymbol{r}, \boldsymbol{c})$ (negative entries) we let $X_{n+1} = x$.

We want to generate "Hypergeometric" distribution, i.e.

$$\pi(x) \propto \frac{1}{\prod_{i,j}(x_{ij}!)}.$$

Acceptance probability is (since $q(x,y) = q(y,x)$)

$$\alpha(x,y) = \min\left(1, \frac{\prod_{i,j}(x_{ij}!)}{\prod_{i,j}(y_{ij}!)}\right)$$

and $\alpha(x,y) = 0$ if $y \notin A(\boldsymbol{r}, \boldsymbol{c})$.

Irreducable? Can be shown with Gröbner-bases!

Aperiodic? Yes, since if any entry=0 we stay there with positive probability.

Therefore ergodic and we can simulate distribution of the test variable Q arbitrarily well.

Works also for other distributions than "hyper-geometric".

<u>More general tables</u> (with $I \times J$ as illustration)

$\mathcal{H}$=finite set

For $I \times J$ we have

$\mathcal{H} = \{(i,j) : 1 \le i \le I, 1 \le j \le J\}$ i.e. $\mathcal{H} = I \times J$. An $x \in \mathcal{H}$ corresponds to a cell.

Generate observation $X_1 \in \mathcal{H}$ by the exponential family distribution

$$P(X_1 = x) = Z(\theta)e^{\boldsymbol{\theta}\cdot T(x)}, \;\; x \in \mathcal{H}$$

where $\boldsymbol{\theta} \in R^d$ is d-dimensional parameter and

$T : \mathcal{H} \to N^d - 0^d$, (often $T : \mathcal{H} \to \{0,1\}^d - 0^d$)

T is a d-dimensional vector of suffcient statistics.

For independent $(X_1, X_2, \cdots, X_N) = \boldsymbol{X}$ with that distribution

$$P(\boldsymbol{X} = \boldsymbol{x}) = (Z(\boldsymbol{\theta}))^N \exp(\boldsymbol{\theta} \sum_{i=1}^{N} T(x_i))$$

depending on data only through value of $\sum_1^N T(x_i)$ (i.e. d sufficient statistics).

With $\mathcal{H} = I \times J$ and $T((i,j)) \in N^{I+J}$ where

$T((i,j)) = (0, \cdots, 0, 1, 0, \cdots, 0 | 0 \cdots, 0, 1, 0, \cdots, 0)$

with 1:s in position i and $I + j$.

$\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_I, \theta'_1, \theta'_2, \cdots, \theta'_J)$ and

$$P(X = (i,j)) \propto e^{\theta_i} e^{\theta'_j}$$

i.e. independence between row and column. The $d = I + J$ sufficient statistics

$$\sum_1^N T(X_i) = \boldsymbol{t}$$

for $\mathcal{H} = I \times J$ corresponds to $\boldsymbol{t} = (\boldsymbol{r}, \boldsymbol{c})$, i.e. the marginals.

With

$$\mathcal{Y}_{\boldsymbol{t}} = \{(x_1, x_2, \cdots, x_N) \in \mathcal{H}^N : \sum_1^N T(x_i) = \boldsymbol{t}\}$$

we have uniform distribution on $\mathcal{Y}_{\boldsymbol{t}}$ with the exponential model.

Conveniently rewritten as

$$\boldsymbol{t} = \sum_{i=1}^{N} T(x_i) = \sum_{x \in \mathcal{H}} g(x)T(x)$$

where $g(x) =$ (number of $x_i = x$)

We let

$$\mathcal{F}_{\boldsymbol{t}} = \{g : \mathcal{H} \to N \text{ where } \sum_{x \in \mathcal{H}} g(x)T(x) = \boldsymbol{t}\}$$

and $g(x) =$number in cell x.

If $g \in \mathcal{F}_{\boldsymbol{t}}$ then $g(x)$, $x \in \mathcal{H}$ is a possible table with "correct" sufficient statistics $\boldsymbol{t}$.

We are interested in "hypergeometric" distribution on $\mathcal{F}_{\boldsymbol{t}}$

$$\pi(g) = H_{\boldsymbol{t}}(g) = \frac{\frac{N!}{\prod_{x \in \mathcal{H}}(g(x)!)}}{|\mathcal{Y}_{\boldsymbol{t}}|}.$$

which is the distribution on $\mathcal{F}_{\boldsymbol{t}}$ if we have uniform distribution on $\mathcal{Y}_{\boldsymbol{t}}$. True for model with exponential family.

Note that

$$\pi(g) \propto \frac{1}{\prod_{x \in \mathcal{H}}(g(x)!)}.$$

We want to simulate $\pi(g)$, $g \in \mathcal{F}_{\boldsymbol{t}}$ with Metropolis-Hastings algorithm. Convenient to choose proposal distribution $q(g, g')$ so that $q(g, g') = q(g', g)$ and with such simple moves that $\pi(g')/\pi(g)$ is simple to calculate so acceptance probability

$$\alpha(g, g') = \min\left(1, \frac{\pi(g')q(g', g)}{\pi(g)q(g, g')}\right) = \min\left(1, \frac{\pi(g')}{\pi(g)}\right)$$

is simple to calculate.

We want simple candidates for moves f where we from $g \in \mathcal{H}$ go to $g \pm f$ which gives an irreducable, aperiodic Markov chain.

Note: These (say) L simple moves $f_1, f_2, \cdots, f_L$ must not change the d sufficient statistics $\boldsymbol{t}$. Therefore

$$\sum_{x \in \mathcal{H}} f_i(x)T(x) = \mathbf{0}, \; i = 1, 2, , \cdots, L$$

since then

$$\sum_{x \in \mathcal{H}} (g(x) + f_i(x))T(x) = \boldsymbol{t}, \; i = 1, 2, , \cdots, L$$

and therefore $g + f_i \in \mathcal{F}_{\boldsymbol{t}}$ (if $g(x) + f_i(x) \geq 0$ for all $x \in \mathcal{H}$).

Move from g to $g - f_i$ also keeps value of $\boldsymbol{t}$.

We choose f_i, $i = 1, 2, \cdots, L$ with probability $1/L$ each and then (independently) ε as $+1$ or -1 with probability 1/2 each and then go from g to $g' = g + \varepsilon f_i$ with probability $\alpha(g, g')$.

We have $q(g, g') = q(g', g)$!

Possibly $g' \notin \mathcal{F}_{\boldsymbol{t}}$ (negative entry) but then we stay in g.

Markov-base: The moves $f_1, f_2, \cdots, f_L$ are a Markov-base if they fulfill

1) $\sum_{x \in \mathcal{H}} f_i(x) T(x) = \mathbf{0}$

2) If g and $g' \in \mathcal{F}_{\boldsymbol{t}}$ we can go from g to g' by a sequence of simple moves $\varepsilon_i f_{i_j}$ where we are in $\mathcal{F}_{\boldsymbol{t}}$ after each step (irreducable chain)

Therefore: There exists an A such that with correct choices of $i_1, i_2, \cdots, i_A$ and $\varepsilon_j,\ j = 1, 2, \cdots, A$ (which are $+1$ or -1) we have

$$g' = g + \sum_{j=1}^{A} f_{i_j} \varepsilon_j$$

and

$$g + \sum_{j=1}^{a} f_{i_j} \varepsilon_j \geq 0,\ 1 \leq a \leq A$$

The proposal distribution $q(g, g')$ means: Choose an f from $f_1, f_2, \cdots, f_L$ with equal probabilities and then independently ε as $+1$ or -1 with probability 1/2 each and let $g' = g + \varepsilon f$

A proposal $g + \varepsilon f$ is accepted with probability

$$\alpha(g, g + \varepsilon f_i) = \min(1, \frac{H_{\boldsymbol{t}}(g + \varepsilon f)}{H_{\boldsymbol{t}}(g)})$$

where $H_{\boldsymbol{t}}$ is the target distribution on $\mathcal{F}_{\boldsymbol{t}}$.

If $g + \varepsilon f \notin \mathcal{F}_{\boldsymbol{t}}$ we remain in g.

Can be modified to make larger jumps.

How to find these simple moves? Given by a Gröbner base!

Example: $3 \times 3 \times 3$-table i.e.

$\mathcal{H} = \{(i, j, k) : 1 \leq i \leq 3, 1 \leq j \leq 3, 1 \leq k \leq 3\}$

We let

$p_{ijk} = P(\text{cell } (i, j, k))$.

As a model we have "no three-way-interaction"

$\ln p_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$

where we have no $(\alpha\beta\gamma)_{ijk}$-terms.

Is a model of the type "exponential family"!

With N_{ijk} =count in cell (i, j, k) we have under the model that $N_{ij\cdot}$, $N_{i\cdot k}$ and $N_{\cdot jk}$ are sufficient statistics where

$$N_{ij\cdot} = \sum_{k=1}^{3} N_{ijk},$$

$$N_{i\cdot k} = \sum_{j=1}^{3} N_{ijk},$$

$$N_{\cdot jk} = \sum_{i=1}^{3} N_{ijk}.$$

What do simple moves look like?

27 of the type $2 \times 2 \times 2$

$$
\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}
\qquad
\begin{array}{ccc} 0 & 0 & 0 \\ 0 & + & - \\ 0 & - & + \end{array}
\qquad
\begin{array}{ccc} 0 & 0 & 0 \\ 0 & - & + \\ 0 & + & - \end{array}
$$

and 54 of the type

$$
\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}
\qquad
\begin{array}{ccc} 0 & - & + \\ + & 0 & - \\ - & + & 0 \end{array}
\qquad
\begin{array}{ccc} 0 & + & - \\ - & 0 & + \\ + & - & 0 \end{array}
$$

and 28 of the type

$$
\begin{array}{ccc} 0 & 0 & 0 \\ 0 & - & + \\ 0 & + & - \end{array}
\qquad
\begin{array}{ccc} + & 0 & - \\ - & + & 0 \\ 0 & - & + \end{array}
\qquad
\begin{array}{ccc} - & 0 & + \\ + & 0 & - \\ 0 & 0 & 0 \end{array}
$$

and 1 of the type

$$
\begin{array}{ccc} -2 & + & + \\ + & 0 & - \\ + & - & 0 \end{array}
\qquad
\begin{array}{ccc} + & 0 & - \\ 0 & 0 & 0 \\ - & 0 & + \end{array}
\qquad
\begin{array}{ccc} + & - & 0 \\ - & 0 & + \\ 0 & + & - \end{array}
$$

110 simple moves in all!

Example: Logistic regression.

Answer "yes" (1) or "no" (0) to question.

We have covariates (prognostic factors) $\boldsymbol{z}_j \in Z^d$ for individual j

$\boldsymbol{z}_j = (z_{1j}, z_{2j}, \cdots, z_{dj})$.

Logistic regression :

$$P(Y_j = 1 | \boldsymbol{z}_j) = \frac{e^{\boldsymbol{\beta} \cdot \boldsymbol{z}_j}}{1 + e^{\boldsymbol{\beta} \cdot \boldsymbol{z}_j}}$$

and

$$P(Y_j = 0 | \boldsymbol{z}_j) = \frac{1}{1 + e^{\boldsymbol{\beta} \cdot \boldsymbol{z}_j}}$$

$\boldsymbol{\beta} \in R^d$ is a d-dimensional parameter.

E.g. "simple linear regression" with

$\boldsymbol{\beta} = (\alpha, \beta)$ as parameter and covariates $\boldsymbol{z}_j = (1, z_j)$ where e.g. z_j =number of years of school for person j, $j = 1, 2, \cdots, N$.

We have

$$P(Y_j = 1 | z_j) = \frac{e^{\alpha + \beta z_j}}{1 + e^{\alpha + \beta z_j}}$$

and

$$P(Y_j = 0 | z_j) = \frac{1}{1 + e^{\alpha + \beta z_j}}.$$

Likelihood for $Y_1, Y_2, \cdots, Y_N$ is

$$\prod_{j=1}^{N} \frac{e^{Y_j \boldsymbol{\beta} \cdot \boldsymbol{z_j}}}{1 + e^{\boldsymbol{\beta} \cdot \boldsymbol{z_j}}}$$

Let $n(\boldsymbol{z})$ =(number with $\boldsymbol{z}_j = \boldsymbol{z}$)

$n_1(\boldsymbol{z})$=(number with $\boldsymbol{z_j} = \boldsymbol{z}$ and $Y_j = 1$).

If the possible $\boldsymbol{z}$-values are $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_M$ we have the likelihood

$$\frac{e^{\boldsymbol{\beta} \sum_{i=1}^{M} n_1(\boldsymbol{a}_i) \boldsymbol{a}_i}}{\prod_{i=1}^{M} (1 + e^{\boldsymbol{a}_i \boldsymbol{\beta}})^{n(\boldsymbol{a}_i)}}$$

i.e.

$n(\boldsymbol{a}_1), n(\boldsymbol{a}_2), \cdots, n(\boldsymbol{a}_M)$ och $\sum_{i=1}^{M} n_1(\boldsymbol{a}_i) \boldsymbol{a}_i$

are sufficient statistics which we want to condition on and keep fixed for all moves.

Corresponds to

$T(0, \boldsymbol{a}_i) = (0; 0, 0, \cdots, 1, 0, \cdots, 0)$

and

$T(1, \boldsymbol{a}_i) = (\boldsymbol{a}_i; 0, 0, \cdots, 1, 0, \cdots, 0)$

where the 1 is in the i:th position.

We have

$$\mathcal{H} = \{(0, \boldsymbol{a}_i), (1, \boldsymbol{a}_i),\ i = 1, 2, \cdots, M\}$$

and with $g(x)$ =(number of observations $= x$)

$$\boldsymbol{t} = \sum_{i=1}^{N} T(x_i) = \sum_{x \in \mathcal{H}} g(x) T(x) =$$

$$= (\sum_{i=1}^{M} n_1(\boldsymbol{a}_i)\boldsymbol{a}_i, n(\boldsymbol{a}_1), n(\boldsymbol{a}_2), \cdots, n(\boldsymbol{a}_M))$$

Simple moves?

Number with covariate$= \boldsymbol{a}_i$ unchanged and in addition $\sum_{i=1}^{M} n_1(\boldsymbol{a}_i)\boldsymbol{a}_i$ unchanged.

For "simple linear regression" with

$\boldsymbol{a_i} = (1, i),\ i = 1, 2, \cdots 12$ we have

$\mathcal{H} = \{(0, 1, i), (1, 1, i),\ i = 1, 2, \cdots 12\}$

and therefore

$n(i),\ i = 1, 2, \cdots, 12$

$\sum_{i=1}^{12} n_1(i)$

$\sum_{i=1}^{12} n_1(i) i$

are fixed.

With $M = 12$ there are 16968 simple moves for the specific example.

Gröbner-basers:

For $x \in \mathcal{H}$ we define a formal variable u_x and let $g : \mathcal{H} \to N$ be represented by the monomial

$$\prod_{x \in \mathcal{H}} u_x^{g(x)}$$

denoted $\mathcal{H}^g$. We substitute u_x for x

$$\mathcal{H}^g = \prod_{x \in \mathcal{H}} x^{g(x)}$$

For 2×2-table

3	2
7	4

with $\mathcal{H} = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ we get the monomial

$$u_{(1,1)}^3 u_{(1,2)}^2 u_{(2,1)}^7 u_{(2,2)}^4$$

We study (finite) polynomials in these monomials with coefficients from a field k and denote these as $k[\mathcal{H}]$.

A $T : \mathcal{H} \to N^d$ where

$T(x) = (T(x)_1, T(x)_2, \cdots, T(x)_d)$ is represented by

$$\phi_T : k[\mathcal{H}] \to k[t_1, t_2, \cdots, t_d]$$

where

$$x \mapsto t_1^{T(x)_1} t_2^{T(x)_2} \cdots t_d^{T(x)_d}$$

ϕ_T is extended to $k[\mathcal{H}]$ by $\phi_T(xy) = \phi_T(x)\phi_T(y)$ and $\phi_T(x + y) = \phi_T(x) + \phi_T(y)$

Hence, for $g : \mathcal{H} \to N$

$$\phi_T(\mathcal{H}^g) = \phi_T(\prod_{x\in\mathcal{H}} x^{g(x)}) = \prod_{x\in\mathcal{H}} (\phi_T(x))^{g(x)} =$$

$$= \prod_{x\in\mathcal{H}} (t_1^{T(x)_1} t_2^{T(x)_2} \cdots t_d^{T(x)_d})^{g(x)} =$$

$$= \prod_{x\in\mathcal{H}} t_1^{T(x)_1 g(x)} t_2^{T(x)_2 g(x)} \cdots t_d^{T(x)_d g(x)} =$$

$$= t_1^{\sum_x T(x)_1 g(x)} t_2^{\sum_x T(x)_2 g(x)} \cdots t_d^{\sum_x T(x)_d g(x)}.$$

If the variable list is $Y = (t_1, t_2, \cdots, t_d)$ we can write

$$\phi_T(\mathcal{H}^g) = Y^{\sum_{x\in\mathcal{H}} g(x)T(x)}$$

Of interest is the ideal

$\mathcal{J}_T = \{p \in k[\mathcal{H}] : \phi_T(p) = 0\} = \ker \phi_T$ which is an ideal generated by binomials (in fact by monomial-differences).

A move $f_i : \mathcal{H} \to Z$ is split up in

"positive" and "negative" changes

$$f_i = f_i^+ - f_i^-$$

where $f_i^\pm : \mathcal{H} \to N$.

Corresponds to monomial difference

$\mathcal{H}^{f_i^+} - \mathcal{H}^{f_i^-}$.

We have $\sum_{x\in\mathcal{H}} f(x)T(x) = 0$ iff

$\mathcal{H}^{f_i^+} - \mathcal{H}^{f_i^-}$ is in $\mathcal{J}_T$, i.e.

$$\phi_T(\mathcal{H}^{f_i^+} - \mathcal{H}^{f_i^-}) = 0.$$

Theorem: $f_1, f_2, \cdots, f_L$ is a Markov-base iff $\mathcal{H}^{f_i^+} - \mathcal{H}^{f_i^-}$ for $i = 1, 2, \cdots, L$ generates the ideal $\mathcal{J}_T = \{p \in k[\mathcal{H}] : \phi_T(p) = 0\}$.

Hilbert's basis-theorem: Every ideal in a polynomial ring is generated by a finite number of polynomials.

Gröbner-bases are a special choice of bases which generate the ideal. Depends on the "ordering" of monomials. If the ideal is generated by monomial differences the Gröbner basis also consists of monomial differences.

Gives an algorithm (in Maple!) giving Gröbner-bases corresponding to the Markov-base, i.e. the simple moves making the chain ergodic.

But: The Markov-base can be too large.

For $I \times J$ this proves that the simple moves of type

$$\begin{pmatrix} + & - \\ - & + \end{pmatrix} \text{ och } \begin{pmatrix} - & + \\ + & - \end{pmatrix}$$

makes the chain irreducable on $A(\boldsymbol{r}, \boldsymbol{c})$