Lecture 3

04.11.2019.

Recap: A, h, A is a $k \times r$ integer matrix with $\mathbb{1}$ in its rowspan $h \in \mathbb{R}^r_{>0}$.

Want to test $H_0: p \in \mathcal{M}_{A,h}$ vs. $H_1: p \notin \mathcal{M}_{A,h}$.

To do: Compute p-value. asymptotic vs. exact approach.

Prop 1: If $p \in \mathcal{M}_A$, and $p(x) = \theta_1^{a_{1x}} \cdots \theta_k^{a_{kx}}$, $x \in [r]$, then $P(U = u) = \frac{n!}{\prod_{i \in [r]} u_i!} \theta^{Au}$ and the probability $P(U = u \mid AU = Au)$ does not depend on p.

- Based on Prop 1, we generalize Fisher's exact test by computing the p-value

$$P(X^2(U) \geq X^2(u) \mid AU = Au)$$

Here $$X^2(U) = \sum_{i \in} \frac{(U_i - \hat{u}_i)^2}{\hat{u}_i},$$ $\hat{u}_i = n\hat{p}_i$ where $\hat{p}_i$ is the MLE. → Lecture 4.

$\hat{p}_i$ is the same for all tables in the fiber.

Hence the p-value is:

$$\frac{\sum_{v \in \mathcal{F}(u)} 1_{X^2(v) \geq X^2(u)} \Big/ \left(\prod_{i \in r} v_i!\right)}{\sum_{v \in \mathcal{F}(u)} 1 \Big/ \left(\prod_{i \in r} v_i!\right)}$$

→ Exact computation of this quantity is prohibitive. Thus we sample from elements in the fiber.

Def: Let $\mathcal{M}_A$ be the log-linear model associated with a matrix A, $\mathbb{1} \in \text{rowspan}(A)$. A finite subset $\mathcal{B} \subset \text{Ker}_{\mathbb{Z}} A$ is a Markov basis for $\mathcal{M}_A$ if for all $u \in \mathcal{T}(n)$ and all pairs $v, v' \in \mathcal{F}(u)$ there exists a sequence $u_1, \ldots, u_L \in \mathcal{B}$ such that

$$v' = v + \sum_{k=1}^{L} u_k \quad \text{and} \quad \underbrace{v + \sum_{k=1}^{l} u_k \geq 0}_{} \quad \text{for all } l = 1, \ldots, L$$

This is saying entries in the transition tables always need to be positive

Algorithm: (Metropolis-Hastings)

INPUT: A table $u \in \mathcal{T}(n)$ and a Markov basis for $\mathcal{M}_A$

OUTPUT: A sequence $(X^2(v_1), X^2(v_2), \ldots \to \infty)$ for tables v_t in the fiber $\mathcal{F}(u)$.

STEP 1: $v_1 = u$.

STEP 2: For $t = 1, 2, \ldots$

(i) Select uniformly at random a move $u_t \in \mathcal{B}$

(ii) If any coordinate of $v_t + u_t$ is negative, set $v_{t+1} = v_t$, else set

$$v_{t+1} = \begin{cases} v_t + u_t \\ v_t \end{cases} \quad \text{with probability} \quad \begin{cases} q \\ 1 - q \end{cases}$$

where $q = \min \left\{ 1, \underbrace{\frac{P(U = v_t + u_t \mid AU = Au)}{P(U = v_t \mid AU = Au)}}_{} \right\}$

(iii) Compute $X^2(v_t)$

↳ The fact that this is a ratio means we don't have to comput $\sum_{v \in \mathcal{F}(u)} \frac{1}{\prod v_i!}$

Thm: The output $(X^2(v_t))_{t=1}^{\infty}$ of ALG. is an aperiodic, reversible and irreducible Markov chain that has stationary distribution equal to the conditional distribution of $X^2(U)$ given $AU = Au$.

Coroll1: With probability one, the output sequence $(X^2(v_t))_{t=1}^{\infty}$ of ALG satisfies

$$\lim_{M \to \infty} \frac{1}{M} \sum_{t=1}^{M} 1_{\{X^2(v_t) \geq X^2(u)\}} = P(X^2(U) \geq X^2(u) \mid AU = Au)$$

Thm: (Fundamental Thm of Markov bases).
A finite subset $\mathcal{B} \subseteq \mathrm{Ker}_{\mathbb{Z}} A$ is a Markov basis for A
$\Leftrightarrow$ the corresponding set of binomials

$\{ p^{v^+} - p^{v^-} : v \in \mathcal{B} \}$ is a generating set for the toric ideal I_A.

Conditional independence models — Ch. 4. Sullivant

Let $X = (X_1, \ldots, X_m)$ be a random vector with state space $\mathcal{X} = \prod_{i=1}^{m} \mathcal{X}_i$.

$f(x) = f(x_1, \ldots, x_m)$ is the joint pdf of X.

- For $A \subseteq [m]$, $X_A = (X_a)_{a \in A}$, $\mathcal{X}_A = \prod_{a \in A} \mathcal{X}_a$.

Def: • Let $A \subseteq [m]$. The marginal density $f_A(x_A)$ of X_A is obtained by integrating out $x_{[m]\setminus A}$:

$$f_A(x_A) := \int_{\mathcal{X}_{[m]\setminus A}} f(x_A, x_{[m]\setminus A}) \, d\nu_{[m]\setminus A}(x_{[m]\setminus A}), \quad x_A \in \mathcal{X}_A$$

- Let $A, B \subseteq [m]$ be disjoint and $x_B \in \mathcal{X}_B$. The conditional density of X_A given $X_B = x_B$ is

$$f_{A|B}(x_A | x_B) := \begin{cases} \dfrac{f_{A \cup B}(x_A, x_B)}{f_B(x_B)}, & \text{if } f_B(x_B) > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Let $A, B, C \subseteq [m]$ be pairwise disjoint. The random vector X_A is conditionally independent of X_B given X_C if and only if

$$f_{A \cup B | C}(x_A, x_B | x_C) = f_{A|C}(x_A | x_C) \cdot f_{B|C}(x_B | x_C)$$

for all x_A, x_B, x_C. Notation: $X_A \perp\!\!\!\perp X_B \mid X_C$

$$f_{A|B \cup C}(x_A | x_B, x_C) = \frac{f_{A \cup B | C}(x_A, x_B | x_C)}{f_{B|C}(x_B | x_C)} = f_{A|C}(x_A | x_C)$$

"Given X_C, knowing X_B does not give any information about X_A"

Suppose $X=(X_1,\dots,X_m)$ is discrete.
X_j has outcome space $[r_j]$. $\mathcal{X}=\prod_{j\in[m]}[r_j]$

Proposition: If X is a discrete random vector, then the condition. independence statement $X_A \perp\!\!\!\perp X_B \mid X_C$ holds if and only if

$$p_{i_A,i_B,i_C,+}\cdot p_{j_A,j_B,i_C,+} - p_{i_A,j_B,i_C,+}\cdot p_{j_A,i_B,i_C,+} = 0$$

for all $i_A, j_A \in \mathcal{X}_A$, $i_B, j_B \in \mathcal{X}_B$, $i_C \in \mathcal{X}_C$.

NOTE: This is a binomial if $A\cup B\cup C=[m]$.

These can be realized as minors of matrices.

$p_{i_A,i_B,i_C,+} = P(X_A=i_A, X_B=i_B, X_C=i_C)$

$p_x = P(X=x) \quad x\in\mathcal{X} \quad x=(i_1,i_2,\dots,i_m) \quad i_k\in[r_k]$

$$p_{i_A,i_B,i_C,+} = \sum p_{i_A,i_B,i_C,j}$$

over all subindices j in $\prod_{i\in[m]\setminus(A\cup B\cup C)} X_i$

pf. $P(X_A=i_A, X_B=i_B, X_C=i_C)\cdot P(X_A=j_A, X_B=j_B, X_C=i_C)$

$= P(X_A=i_A \mid X_B=i_B, X_C=i_C)\, P(X_B=i_B, X_C=i_C)$

swap swap

$\cdot P(X_A=j_A \mid X_B=j_B, X_C=i_C)\, P(X_B=j_B, X_C=i_C)$ ∎

Def: The conditional independence ideal $I_{A\perp\!\!\!\perp B\mid C} \subseteq \mathbb{C}[p]$ is generated by all the quadratic polynomials in the Prop. The discrete conditional independence model, $r=|\mathcal{X}|$

$\mathcal{M}_{A\perp\!\!\!\perp B\mid C} = V_\Delta(I_{A\perp\!\!\!\perp B\mid C}) \subseteq \Delta_{r-1}$ is the set of all probability distributions in Δ_{r-1} satisfying the polyn. in Prop.

If $\mathcal{C} = \{A_1 \perp\!\!\!\perp B_1 \mid C_1,\ A_2 \perp\!\!\!\perp B_2 \mid C_2, \dots\}$ is a set of conditional independence statements, we construct the conditional indep. ideal

$$I_{\mathcal{C}} = \sum_{A \perp\!\!\!\perp B \mid C \in \mathcal{C}} I_{A \perp\!\!\!\perp B \mid C}.$$

which is the sum of ideals generated by quadrics
The model $\mathcal{M}_{\mathcal{C}} := V_{\Delta}(I_{\mathcal{C}}) \subseteq \Delta_r$ consists of all probability distributions satisfying the constraints in $\mathcal{C}$.

<u>Example</u>: Binary contraction axiom.

$$\mathcal{C} = \{1 \perp\!\!\!\perp 2 \mid 3,\ 2 \perp\!\!\!\perp 3\} \Rightarrow \{2 \perp\!\!\!\perp \{1,3\}\}$$

$$I_{\mathcal{C}} = \langle p_{111}p_{221} - p_{121}p_{211}$$