Markov random fields I

Rasmus Waagepetersen

November 9, 2023

Outline:

1. specification of joint distributions
2. conditional specifications
3. conditional auto-regression
4. Brooks factorization
5. conditional independence and graphs
6. Hammersley-Clifford

Specification of joint distributions

Consider random vector $(X_1, \ldots, X_n)$.

How do we specify its joint distribution ?

1. assume $X_1, \ldots, X_n$ independent - but often not realistic
2. assume $(X_1, \ldots, X_n)$ jointly normal and specify mean vector and covariance matrix (i.e. positive definite $n \times n$ matrix)
3. use copula (e.g. transform marginal distributions of joint normal)
4. specify $f(x_1)$, $f(x_2|x_1)$, $f(x_3|x_1, x_2)$ etc.
5. specify full conditional distributions $X_i|X_{-i}$ - but what is then joint distribution - and does it exist ?
$(X_{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n))$

In this part of the course we will consider the fifth option.

Conditional auto-regressions

Suppose $X_i|X_{-i}$ is normal.

Auto-regression natural candidate for conditional distribution:

$$X_i|X_{-i} = x_{-i} \sim N(\alpha_i + \sum_{l \neq i} \gamma_{il} x_l, \kappa_i) \tag{1}$$

Equivalent (under certain conditions, see exercise) and more convenient:

$$X_i|X_{-i} = x_{-i} \sim N(\mu_i - \sum_{l \neq i} \beta_{il}(x_l - \mu_l), \kappa_i) \tag{2}$$

Is this consistent with a multivariate normal distribution $N_n(\mu, \Sigma)$ for X ?

Brook's lemma

Consider two outcomes x and y of X where X has joint density p where $p(y) > 0$.

Brooks factorization:

$$\frac{p(x)}{p(y)} = \prod_{i=1}^{n} \frac{p_i(x_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}{p_i(y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n)}$$

Note $n!$ ways to factorize !

If conditional densities consistent with joint density, we can choose fixed y and determine $p(x)$ by

$$p(x) \propto p(x)/p(y)$$

where RHS evaluated using Brook's factorization.

Application to conditional normal specification

We let $y = \mu = (\mu_1, \ldots, \mu_n)$. Then

$$\log\left(\frac{p_i(x_i|x_1,\ldots,x_{i-1},\mu_{i+1},\ldots,\mu_n)}{p_i(\mu_i|x_1,\ldots,x_{i-1},\mu_{i+1},\ldots,\mu_n)}\right)$$

$$= -\frac{1}{2\kappa_i}[(x_i - \mu_i + \sum_{l=1}^{i-1}\beta_{il}(x_l - \mu_l))^2 - (\sum_{l=1}^{i-1}\beta_{il}(x_l - \mu_l)^2]$$

$$= -\frac{1}{2\kappa_i}[(x_i - \mu_i)^2 + 2\sum_{l=1}^{i-1}\beta_{il}(x_i - \mu_i)(x_l - \mu_l)]$$

Assume now $\beta_{ij}/\kappa_i = \beta_{ji}/\kappa_j$. Then (you do the algebra)

$$\log p(x) = \log p(\mu) - \frac{1}{2}\sum_{i=1}^{n}\sum_{l=1}^{n}\frac{\beta_{il}}{\kappa_i}(x_i - \mu_i)(x_l - \mu_l) \qquad (3)$$

with $\beta_{ii} = 1$.

Easy to see that this is consistent with (2)

This is formally equivalent to a multivariate Gaussian density with mean vector μ and precision matrix $Q = \Sigma^{-1} = [q_{ij}]_{ij}$ with $q_{ij} = \beta_{ij}/\kappa_i$.

A well-defined Gaussian density provided Q is symmetric and positive definite (whereby $\Sigma = Q^{-1}$ positive definite and symmetric)

We already assumed symmetry

$$q_{ij} = q_{ji} \Leftrightarrow \beta_{ij}/\kappa_i = \beta_{ji}/\kappa_j \Leftrightarrow \beta_{ij}\kappa_j = \beta_{ji}\kappa_i$$

Positive definiteness must be checked by considering the whole of Q.

There are $n!$ factorizations. In general

$$\frac{p(x)}{p(y)} = \prod_{i=1}^{n} \frac{p_{\pi(i)}(x_{\pi(i)}|x_{\pi(1)}, \ldots, x_{\pi(i-1)}, y_{\pi(i+1)}, \ldots, y_{\pi(n)})}{p_{\pi(i)}(y_{\pi(i)}|x_{\pi(1)}, \ldots, x_{\pi(i-1)}, y_{\pi(i+1)}, \ldots, y_{\pi(n)})}$$

where $(\pi(1), \ldots, \pi(n))$ represents a permutation of $(1, 2, \ldots, n)$.

In case of (2) we obtain in the same manner as before

$$\log p(x) = \log p(\mu) - \frac{1}{2} \sum_{i=1}^{n} \sum_{l=1}^{n} \frac{\beta_{\pi(i)\pi(l)}}{\kappa_{\pi(i)}} (x_{\pi(i)} - \mu_{\pi(i)})(x_{\pi(l)} - \mu_{\pi(l)})$$

which in fact coincides with (3) (just a reordering of the double sum) - hence choice of π does not matter.

Conditional distribution of X_i for $N(\mu, Q^{-1})$

$$p_i(x_i|x_{-i}) \propto \exp(-\frac{1}{2}(x_i - \mu_i)^2 Q_{ii} - \sum_{k \neq i}(x_i - \mu_i)(x_k - \mu_k) Q_{ik})$$

For a normal distribution $Y \sim N(\xi, \sigma^2)$,

$$p(y) \propto \exp(-\frac{1}{2\sigma^2} y^2 + \frac{1}{\sigma^2} y \xi)$$

Comparing the two above equations we get (again you do the algebra)

$$X_i | X_{-i} = x_{-i} \sim N(\mu_i - \frac{1}{Q_{ii}} \sum_{k \neq i} Q_{ik}(x_k - \mu_k), Q_{ii}^{-1})$$

Thus auto-regressions on slide 4 are in fact general forms of the conditional distributions for a multivariate normal distribution !

Example: Gaussian random field on 1D lattice

Consider lattice $V = \{l | l = 1, \ldots, L\}$. Define $\mu_i = 0$, $\kappa_i = \beta_{ii} = 1$ and for some $\beta \neq 0$ define

$$\beta_{ij} = \begin{cases} \beta & |i - j| \mod (L-2) = 1 \\ 0 & \text{otherwise} \end{cases}$$

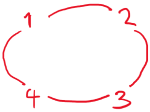
Q obviously symmetric. Q not positive definite if $\beta = -1/2$.

Q positive definite $\Leftrightarrow |\beta| < 1/2$

Note this is an example of a *circulant* precision matrix/random field: if we say i, j are neighbours if $\beta_{ij} \neq 0$ then we obtain circular graph on V.

(exercise in case $L = 4$ - consider determinant of Q)

Case $L = 4$



$\beta_{ij} \neq 0$ if i, j neighbours on circular graph.

Q in case $L=4$

$$\begin{bmatrix} 1 & \beta & 0 & \beta \\ \beta & 1 & \beta & 0 \\ 0 & \beta & 1 & \beta \\ \beta & 0 & \beta & 1 \end{bmatrix}$$

Last 3 rows obtained by shifting first row

Example: Gaussian random field on 2D lattice

Consider lattice $V = \{(l,k)|l = 1,\dots,L, k = 1,\dots,K\}$. Now indices $i,j \in V$ correspond to points (i_1,i_2) and (j_1,j_2) Define $i,j \in V$ to be neighbours $i \sim j \Leftrightarrow |i_1 - j_1| + |i_2 - j_2| = 1$ (i and j horizontal or vertical neighbours).

Tempting: define $\mu_i = 0$,

$$\beta_{ij} = \begin{cases} -1/\#N_i & i \sim j \\ 0 & \text{otherwise} \end{cases}$$

where $\#N_i$ is number of neighbours (2, 3, or 4) of i and $\kappa_i = \kappa/\#N_i > 0$ where $\kappa > 0$. Recall also $\beta_{ii} = 1$

Then

$$X_i | X_{-i} = x_{-i} \sim N\left(\frac{1}{\#N_i}\sum_{j\sim i} x_j, \kappa/\#N_i\right)$$

has conditional mean given by average of neighbours and conditional variance inversely proportional to number of neighbours.

Case $K = L = 3$

$L = K = 3$

$$q_{ij} = \frac{\beta_{ij}}{k_i} = \frac{\frac{-1}{\#N_i}}{\frac{k}{\#N_i}} = \frac{-1}{k}$$

$$q_{ii} = \frac{1}{k_i} = \frac{1}{k} \#N_i$$

Case $K = L = 3$ continued

$V = \{1, \ldots, 9\}$

$$\begin{tikzcd}
1 \arrow[r, dash] \arrow[d, dash] & 2 \arrow[r, dash] \arrow[d, dash] & 3 \arrow[d, dash] \\
4 \arrow[r, dash] \arrow[d, dash] & 5 \arrow[r, dash] \arrow[d, dash] & 6 \arrow[d, dash] \\
7 \arrow[r, dash] & 8 \arrow[r, dash] & 9
\end{tikzcd}$$

$$kQ = \begin{array}{c|ccccccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
1 & 2 & -1 & & -1 & & & & & \\
2 & -1 & 3 & -1 & & -1 & & & & \\
3 & & -1 & 2 & & & -1 & & & \\
4 & -1 & & & 3 & -1 & & -1 & & \\
5 & & -1 & & -1 & 4 & -1 & & -1 & \\
6 & & & -1 & & -1 & 3 & & & -1 \\
7 & & & & -1 & & & 2 & -1 & \\
8 & & & & & -1 & & -1 & 3 & -1 \\
9 & & & & & & -1 & & -1 & 2
\end{array}$$

Problem: resulting Q is positive semi definite:
$x^{\mathsf{T}} Q x = 0 \Leftrightarrow x = a 1_n$ for some $a \in \mathbb{R}$.

Why: ith row of $\kappa^{-1} Q$ has ith entry $\# N_i$ and N_i entries equal to -1. Rest zero. Consider e.g. specific case $L = K = 3$ to see what happens.

We can modify by $Q := Q + \frac{\tau}{\kappa} I$ where $\tau > 0$.
Then modified Q is positive definite and we obtain modified conditional distributions

$$X_i | X_{-i} = x_{-i} \sim N(\mu_i + \frac{1}{\# N_i + \tau} \sum_{k \neq i} (x_k - \mu_k), \frac{\kappa}{\# N_i + \tau})$$

which are consistent with joint multivariate distribution.

For the record: we can make sense of multivariate distributions with positive-semidefinite Σ as distributions on lower dimensional subspaces. Then Q is generalized inverse of Σ.

Conditional vs. marginal modeling

Usually we model multivariate normal in terms of Σ - i.e. covariance structure well understood.

Modeling in terms of conditional distributions (or equivalently Q) appealing but downside is that we do not necessarily know structure of Σ - although it can be computed numerically by inverting Q.

You are invited to invert Q from previous example $L = K = 3$ with $\tau > 0$.

Markov random fields

Let V denote a finite set of vertices and E a set of edges where an element e in E is of the form $\{i, j\}$ for $i \neq j \in V$. (i.e. an edge is a unordered pair of vertices). $G = (V, E)$ is a graph.

$i, j \in V$ are neighbours, $i \sim j$, if $\{i, j\} \in E$.

A random vector $X = (X_i)_{i \in V}$ is a Markov random field with respect to G if

$$p_i(x_i | x_{-i}) = p_i(x_i | x_{N_i})$$

where N_i is the set of neighbours of i and for $x = (x_l)_{l \in V}$ and $A \subseteq V$, $x_A = (x_i)_{i \in A}$.

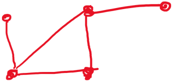
In other words, X_i and X_j are conditionally independent given $X_{-\{i,j\}}$ if i and j are not neighbours.

Markov random field is another word for graphical model.

Graphical model

Model where conditional dependence structure specified by graph.

Graphical models



Nodes V represent random variables.

For $i,j \in V$: i,j not neighbours

$(i \not\sim j,\ \{i,j\} \notin G)$

$\Rightarrow X_i \perp\!\!\!\perp X_j \mid X_{-\{i,j\}}$

Hammersley-Clifford theorem

Consider a positive density $p(\cdot) > 0$ for $X = (X_i)_{i \in V}$ and a graph $G = (V, E)$. Then the following statements are equivalent:

1. X is a MRF wrt G.
2. $$p(x) = \prod_{C \subseteq V} \phi_C(x_C)$$

   for interaction functions ϕ_C where $\phi_C = 1$ unless C is a clique wrt. G. We can further introduce the constraint $\phi_C(x_C) = 1$ if $x_l = y_l$ for $l \in C$ and some fixed y. Then the interaction functions are uniquely determined.

   Notation: for ease of notation we often write i for $\{i\}$ and (x_A, y_B) will denote a vector with entries x_i for $i \in A$ and y_j for $j \in B$, $A \cap B = \emptyset$ (this is a convenient but not rigorous notation)

   Clique: $C \subseteq V$ is a clique if $i \sim j$ for all $i \neq j \in C$ (in particular, all singletons $C = \{i\}$ are cliques)

Proof: 2. $\Rightarrow$ 1.

$$p_i(x_i|x_{-i}) \propto \prod_{C \subseteq V : C \cap i \neq \emptyset} \phi_C(x_C)$$

RHS depends only on $x_j \in N_i$: if $l \in C$ is not a neighbour of i then C can not be a clique. Then $\phi_C(x_C) = 1$ so it does not depend on x_l.

1. $\Rightarrow$ 2.

We choose an arbitrary reference outcome y for X. We then define $\phi_\emptyset = p(y)$ and, recursively,

$$\phi_C(x_C) = \begin{cases} 1 & C \text{ not a clique or } x_l = y_l \text{ for some } l \in C \\ \frac{p(x_C, y_{-C})}{\prod_{B \subset C} \phi_B(x_B)} & \text{otherwise} \end{cases}$$

Let $x = (x_A, y_{-A})$ where $x_l \neq y_l$ for all $l \in A$. We show 2. by induction in the cardinality $|A|$ of A. If $|A| = 0$ then $x = y$ and $p(y) = \phi_\emptyset$ so 2. holds. Assume now that 2. holds for $|A| = k - 1$ where $k \leq |V|$ and consider A with $|A| = k$.

Assume A is a clique. Then by construction,

$$p(x_A, y_{-A}) = \phi_A(x_A) \prod_{B \subset A} \phi_B(x_B)$$

and we are done since for $C \subseteq V$ which is not a subset of A we have $\phi_C((x_A, y_{-A})_C) = 1$ by construction

NB: don't need induction hypothesis in this case.

Assume A is not a clique, i.e. there exist $l, j \in A$ so that $l \not\sim j$. Then

$$\begin{aligned}
p(x_A, y_{-A}) &= \frac{p_l(x_l | x_{A\setminus l}, y_{-A})}{p_l(y_l | x_{A\setminus l}, y_{-A})} p(x_{A\setminus l}, y_{-A}, y_l) \\
&= \frac{p_l(x_l | x_{A\setminus\{l,j\}}, y_j, y_{-A})}{p(y_l | x_{A\setminus\{l,j\}}, y_j, y_{-A})} p(x_{A\setminus l}, y_{-A}, y_l) \\
&= \frac{p_l(x_l, x_{A\setminus\{l,j\}}, y_j, y_{-A})}{p_l(y_l, x_{A\setminus\{l,j\}}, y_j, y_{-A})} p(x_{A\setminus l}, y_{-A}, y_l) \\
&= \frac{\prod_{C\subseteq A\setminus j} \phi_C(x_C)}{\prod_{C\subseteq A\setminus\{l,j\}} \phi_C(x_C)} \prod_{C\subseteq A\setminus l} \phi_C(x_C) \\
&= \prod_{C\subseteq A} \phi_C(x_C)
\end{aligned}$$

where second "$=$" by 1. and fourth "$=$" by induction. Thus 2. also holds in this case.

NB: At the expense of further tecnicalities HC-theorem can be generalized to the case of a not strictly positive $p(\cdot)$.

Exercises

1. Show that two parametrizations (1) and (2) are equivalent under the condition that the matrix A with $A_{ii} = 1$ and $A_{ij} = -\gamma_{ij}$ is invertible. In other words, show that there is an invertible mapping between the parameter vectors $(\alpha_1, \dots, \alpha_n, \gamma_{12}, \dots, \gamma_{(n-1)n})$ and $(\mu_1, \dots, \mu_n, \beta_{12}, \dots, \beta_{(n-1)n})$.

   Hint: equate the conditional means for all $i = 1, \dots, n$.
2. Verify Brook's Lemma.
3. Perform derivations left to the reader at slides 6 and 9
4. Show that a precision matrix, if it exists, is positive definite.
5. Check in case $L = 4$ for circulant Gaussian that Q is positive definite if and only if $|\beta| < 1/2$ (one criterion for this is that all leading principal submatrices have positive determinants)
6. Compute numerically the inverse of Q for circulant Gaussian ($L = 4$, $\beta = -0.3, 0.3$) and inverse of $Q + \tau I$ for various $\tau = 0.01, 0.1, 1$ in case of 2D Gaussian with $L = K = 3$ (slides 10 and 12). Also consider the correlation matrix.

Conditional independence

Suppose X, Y, Z are random variables (or vectors). Then we define X and Y to be conditionally independent given Z if

$$p(x, y|z) = p(x|z)p(y|z)$$

The following statements are equivalent:

1. $p(x, y|z) = p(x|z)p(y|z)$
2. $p(x, y, z) = f(x, z)g(y, z)$ for some functions f and g
3. $p(x|y, z) = p(x|z)$
4. $p(y|x, z) = p(y|z)$

($p(\cdot)$ generic notation for (possibly conditional) probability densities)