Markov random fields

The Markov property

Discrete time:

\((X_k | X_{k-1}, X_{k-2}, \ldots) = (X_k | X_{k-1})\)

A time symmetric version:

\((X_k | X^{-k}) = (X_k | X_{k-1}, X_{k+1})\)

A more general version:

Let \(A\) be a set of indices \(>k\), \(B\) a set of indices \(<k\). Then

\(X_A \perp X_B | X_k\)

These are all equivalent.
On a spatial grid

Let $\delta_i$ be the neighbors of the location $i$. The Markov assumption is

$$P(Z_i = z_i \mid Z^{-i} = z^{-i}) = P(Z_i = z_i \mid Z_{\delta_i} = z_{\delta_i})$$

$$= p_i(z_i \mid z_{\delta_i})$$

Equivalently for $i \notin \delta_i$ $Z_i \perp Z_{\delta_i} \mid Z^{-i,j}$

The $p_i$ are called *local characteristics*. They are stationary if $p_i = p$.

A potential assigns a number $V_A(z)$ to every subconfiguration $z_A$ of a configuration $z$. (There are lots of them!)

Graphical models

Neighbors are nodes connected with edges.

Given 2, 1 and 4 are independent.
**Gibbs measure**

The *energy* $U$ corresponding to a potential $V$ is $U(z) = \sum_A V_A(z)$.

The corresponding *Gibbs measure* is

$$P(z) = \frac{\exp(-U(z))}{C}$$

where $C = \sum_z \exp(-U(z))$

is called the *partition function*.

---

**Nearest neighbor potentials**

A set of points is a *clique* if all its members are neighbours.

A potential is a *nearest neighbor potential* if $V_A(z) = 0$ whenever $A$ is not a clique.
**Markov random field**

Any nearest neighbour potential induces a Markov random field:

\[
p_i(z_i | z_{-i}) = \frac{P(z)}{\sum_{z'} P(z')} = \frac{\exp(- \sum_{C \text{ clique}} V_C(z))}{\sum_{z'} \exp(- \sum_{C \text{ clique}} V_C(z'))}
\]

where \( z' \) agrees with \( z \) except possibly at \( i \), so \( V_C(z) = V_C(z') \) for any \( C \) not including \( i \).

---

**The Hammersley-Clifford theorem**

Assume \( P(z) > 0 \) for all \( z \). Then \( P \) is a MRF on a (finite) graph with respect to a neighbourhood system \( \Delta \) iff \( P \) is a Gibbs measure corresponding to a nearest neighbour potential.

Does a given nn potential correspond to a unique \( P \)?
The Ising model

Model for ferromagnetic spin (values +1 or -1). Stationary nn pair potential
\[ V(i,j)=V(j,i); \quad V(i,i)=V(0,0)=v_0; \]
\[ V(0,e_N)=V(0,e_E)=v_1. \]

\[ \log P(Z_i=1|Z^{-1}) = -(v_0 + v_1(z_{i+e_n} + z_{i-e_n} + z_{i+e_E} + z_{i-e_E})) \]

so \[ L(v) = \frac{\exp(t_0 v_0 + t_1 v_1)}{C(v)} \]
where
\[ t_0 = \sum_i z_i; \quad t_1 = \sum_i \sum_{j \neq i} z_i z_j \]

Interpretation

\( v_0 \) is related to the external magnetic field (if it is strong the field will tend to have the same sign as the external field)

\( v_1 \) corresponds to inverse temperature (in Kelvins), so will be large for low temperatures.
Phase transition

At very low temperature there is a tendency for spontaneous magnetization. For the Ising model, the boundary conditions can affect the distribution of $x_0$. In fact, there is a critical temperature (or value of $v_1$) such that for temperatures below this the boundary conditions are felt. Thus there can be different probabilities at the origin depending on the values on an arbitrary distant boundary!

Simulated Ising fields
Tomato disease

Data on spotted wilt from the Waite Institute 1929. 16 plots in 4x4 Latin square, each 6 rows with 15 plants each. Occurrence of the viral disease 23 days after planting.

Exponential tilting

$$\log\left( \frac{L(v)}{L(u)} \right) = t_0(v_0 - u_0) + t_1(v_1 - u_1)$$
$$-\log\left( \frac{C(v)}{C(u)} \right)$$

$$\frac{C(u)}{C(v)} = E_v(\exp(T_0(v_0 - u_0) + T_1(v_1 - u_1)))$$

Simulate from the model u and estimate the expectation by an average.
Fitting the tomato data

t_0 = -834 \ t_1 = 2266
Condition on boundary and simulate 100,000 draws from u = (0, 0.5).
Mle \ \hat{\nu} = (0, 0.5)
The simulated values of t_0 are half positive and half negative (phase transition).

The auto-models

Let Q(x) = log(P(x)/P(0)). Besag’s auto-models are defined by

\[ Q(z) = \sum_{i=1}^{n} z_i G_i(z_i) + \sum_{i=1}^{n} \sum_{j \neq i} \beta_{ij} z_i z_j \]

When \ z_i \in \{0, 1\} \ and \ G_i(z_i) = \alpha_i \ we \ get \ the \ autologistic \ model

When \ G_i(z_i) = \alpha_i z_i - \log(z_i!) \ and \ \beta_{ij} \leq 0 \ we \ get \ the \ auto-Poisson \ model
Coding schemes

In order to estimate parameters, it can be easier to not use all the data. Besag suggested a coding scheme in which one only uses data at points which are conditionally independent (given all the other data):

Pseudolikelihood

Another approximate approach is to write down a function of the data which is the product of the $p_i(x_i)$, i.e., acting as if the neighborhoods of each point were independent.

This as an estimating equation, but not an optimal one. In fact, in cases of high dependence it tends to be biased.
Recall the Gaussian formula

If
\[
\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)
\]
then
\[
(Y \mid X) \sim N(\mu_Y + \Sigma_{YX} \Sigma_{XX}^{-1} (X - \mu_X), \\
\Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY})
\]

Let \( Q = \Sigma^{-1} \) be the **precision matrix**.
Then the conditional precision matrix is
\[
\left( \Sigma_{YY} - \Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \right)^{-1} = Q_{YY}
\]

Gaussian MRFs

We want a setup in which \( Z_i \perp Z_j \mid \mathcal{Z}_{-ij} \) whenever \( i \) and \( j \) are not neighbors.
Using the Gaussian formula we see that the condition is met iff \( Q_{ij} = 0 \).
Typically the precision matrix of a GMRF is sparse where the covariance is not. This allows fast computation of likelihoods, simulation etc.
An AR(1) process

Let \(X_t \mid X_{t-1} = \phi X_{t-1} + \varepsilon_1\). The lag \(k\) autocorrelation is \(\phi^k\). The precision matrix has \(Q_{ij} = \phi\) if \(|i-j|=1\), \(Q_{ii} = Q_{nn} = 1\), and \(Q_{ii} = 1 + \phi^2\) elsewhere. Thus \(\Sigma\) has \(n^2\) non-zero elements, while \(Q\) has \(n+2(n-1) = 3n-2\) non-zero elements.

Using the Gaussian formula we see that

\[ \mu_{i-l} = \frac{\phi}{1+\phi^2} (x_{i-l-1} + x_{i+l+1}) \quad Q_{i-l} = 1 + \phi^2 \]

Conditional autoregression

Suppose that

\[ Z_i \mid Z^{-i} \sim N(\mu_i + \sum_{j \neq i} \beta_{ij} (x_j - \mu_j), \kappa_i^{-1}) \]

This is called a Gaussian conditional autoregressive model. WLOG \(\mu_i = 0\). If also \(\kappa_i \beta_{ij} = \kappa_j \beta_{ji}\) these conditional distributions correspond to a multivariate joint Gaussian distribution, mean 0 and precision \(Q\) with \(Q_{ii} = \kappa_i\) and \(Q_{ij} = -\kappa_i \beta_{ij}\), provided \(Q\) is positive definite. If the \(\beta_{ij}\) are nonzero only when \(i \sim j\) we have a GMRF.
Likelihood calculation

The Cholesky decomposition of a pd square matrix $A$ is a lower triangular matrix $L$ such that $A=LL^T$.
To solve $Ay = b$ first solve $Lv = b$ (forward substitution), then $L^Ty = v$ (backward substitution).
If a precision matrix $Q = LL^T$, 
$\log \det(Q) = 2 \sum \log(L_{ii})$. The quadratic form in the likelihood is $w^Tu$ where 
$u = Qw$ and $w = (z - \mu)$. Note that 
$$u_i = Q_{ii}w_i + \sum_{j\neq i} Q_{ij}w_j$$

Simulation

Let $x \sim N(0, I)$, solve $L^Tv = x$ and set $z = \mu + v$.
Then $E(z) = \mu$ and $\text{Var}(z) = (L^T)^{-1}IL^{-1} = (LL^T)^{-1} = Q^{-1}$.
Spatial covariance

Whittle (1963) noted that the solution to the stochastic differential equation
\[ \left( \Delta - \frac{1}{\phi} \right)^{(k+1)/2} Z(s) = \varepsilon(s) \]
has covariance function
\[ \text{Cov}(Z(s), Z(s + h)) \propto \left( \frac{\|h\|}{\phi} \right)^{k} \mathcal{K} \left( \frac{\|h\|}{\phi} \right) \]
Rue and Tjelmeland (2003) show that one can approximate a Gaussian random field on a grid by a GMRF.

Unequal spacing

Lindgren and Rue show how one can use finite element methods to approximate the solution to the sde (even on a manifold like a sphere) on a triangulization on a set of possibly unequally spaced points.
The African Sahel region (south of Sahara) suffered severe drought in the 70s through 90s. There is recent evidence of recovery. Data from GHCN at 550 stations 1982-1996 (monthly, aggregated to annual).
Model

Precipitation is determined by a latent (hidden) process, modelled as a Gaussian process on the sphere. This is approximated by a GMRF $Z(s, t)$ on a discrete number of points:

$$Y(s, t) = P(s, t)^{1/3}(1-0.13P(s, t)^{1/3}) = Z(s, t) + \epsilon$$

Mean is a linear combination of basis functions (B-splines).
Temporal structure is AR(1).
Fitting by MCMC.
Results

$E(x | Y)$ for 1982

Lower 95% confidence interval for 1982

Upper 95% confidence interval for 1982

Reserved data

Model check by reserving 10 stations.