

## 2 Markov-Zufallsfelder

### 2.1 Daten

**Fortsetzung Beispiel 8** (Ratbert)

Farbintensitäten auf regulärem Gitter.

**Fortsetzung Beispiel 7** (Disease Mapping)

- Anzahl von Erkrankungen oder Todesfällen in bestimmten Zeitperioden, geschichtet nach Geschlecht, Altersgruppen und Regionen.
- Keine individuellen Kovariablen.
- $y_{sj}$  sei die beobachtete Anzahl von Fällen in Region  $s = 1, \dots, S$  und Altersgruppe  $j = 1, \dots, J$ .
- $n_{sj}$  sei die Anzahl von Personen unter Risiko (Bevölkerungszahl in Region  $s$  und Altersgruppe  $j$ ).
- Binomialmodell

$$Y_{sj} \sim \text{Bin}(n_{sj}, \pi_{sj})$$

approximiert durch Poisson-Modell

$$Y_s \sim \text{Po}(e_s \lambda_s)$$

- $e_s$  erwartete Anzahl von Fällen in Region  $s$
- $p_j$  altersspezifische Referenzwahrscheinlichkeiten

$$e_s = \sum_j n_{sj} p_j$$

- ML-Schätzer für  $\lambda_s$  bei Unabhängigkeitsann. für  $\{Y_s, s = 1, \dots, S\}$  : Standardisierte Mortalitäts-Rate SMR

$$SMR_s = \frac{y_s}{e_s}, \quad s = 1, \dots, S$$

$$\mathbb{E}[SMR_s] = \frac{\mathbb{E}[Y_s]}{e_s} = \frac{e_s \lambda_s}{e_s} = \lambda_s$$

- Der Schätzer ist erwartungstreu, aber

$$\text{Var}(SMR_s) = \text{Var}\left(\frac{Y_s}{e_s}\right) = \frac{1}{e_s^2} \text{Var}(Y_s) = \frac{\lambda_s}{e_s}$$

$\Rightarrow$  ML-Schätzer sehr unzuverlässig bei seltenen Krankheiten ( $e_s$  klein)

**Ziel:** Stabilisierung der Schätzung durch Annahme räumlicher Korrelation und damit Berücksichtigung von Nachbarschaftsinformation.

**Beispiel 11.** (Mietspiegel)

$Y_i$  Nettomiete,  
 $x_i$  Kovariablen (Baujahr, Wohnfläche, binäre Variablen);  
 $s_i$  Subbezirksviertel der Wohnung  $i$  in München.

### 2.1.1 Allgemein

$y_i$  Wert der Zielvariablen für Individuum  $i$   
 $x_i$  Kovariablen  
 $s_i \in \{1, \dots, S\}$  räumliche Einheit in der  $i$  „lebt“  
 $Y_i | x_i, s_i \sim$  Gauß, Binomial, Poisson, ...

## 2.2 Markov-Zufallsfelder (MRF, Markov Random Fields)

Wir betrachten diskrete Zufallsfelder mit Markoveigenschaft:

**Definition 2.1.** Sei  $Y(s)$  ein Zufallsfeld mit  $s = \{1, \dots, S\}$ . Dann nennt man mit  $Y_i = Y(s_i)$  (siehe Abb. 2).

- Paarweise Markoveigenschaft:  $Y_i \perp Y_j | Y_{-ij}$  falls  $i \neq j$  und  $i$  kein Nachbar von  $j$ ,
- Lokale Markoveigenschaft  $Y_i \perp Y_{-\{i, n \in (i)\}} | Y_{n \in (i)}, n \in (i) = \{j : i \sim j\}$  Nachbarn von  $i$ ,
- Globale Markoveigenschaft:  $Y_A \perp Y_B | Y_C$  für alle disjunkten Mengen  $A, B$  und  $C$ , wobei  $C$  die (nichtleeren) Mengen  $A$  und  $B$  separiert.

Erfüllt  $Y(s)$  eine Markoveigenschaft, so nennt man  $Y(s)$  ein **Markovzufallsfeld**.

Anmerkung:

- Die Markoveigenschaften sind (im Allgemeinen) nicht äquivalent. Die globale Markoveigenschaft ist die stärkste, die paarweise die schwächste.
- Die Markoveigenschaft für zeitliche Prozesse (Gedächtnislosigkeit) ergibt sich als Spezialfall.
- Für räumliche Prozesse müssen wir erst definieren, was überhaupt ein Nachbar ist.

### 2.2.1 Nachbarschaftssysteme

$n(s)$  Menge der Nachbarn (nächste Pixel, angrenzende Regionen)

$D$  Parameterraum von  $s$ ,  $D = \{1, \dots, S\}$  oder  $D \subseteq \mathbb{Z}^d$ ,  $d = 1, 2, 3, \dots$

**Definition 2.2** ((Nachbarschaftssystem)). Eine Menge von Orten  $\partial = \{\partial(s) : s \in D\}$  ist ein Nachbarschaftssystem, wenn gilt:

$$s \notin \partial(s)$$

$$v \in \partial(s) \Leftrightarrow s \in \partial(v)$$

Alle  $v \in \partial(s)$  heißen Nachbarn von  $s$  ( $v \sim s$ ).

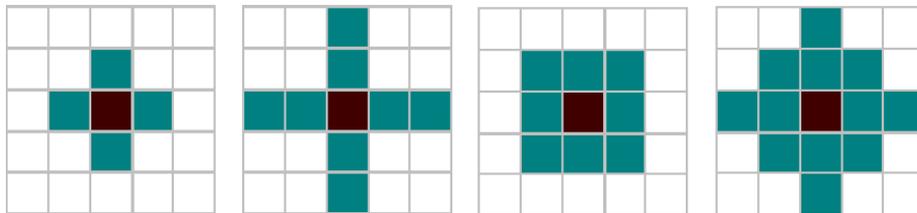


Abbildung 1: Nachbarschaftssysteme auf regulären Gittern. Von links nach rechts: Erste Nachbarn, zweite Nachbarn, zweite diagonale Nachbarn, zweite und diagonale Nachbarn.

**Fortsetzung Beispiel 8** (Bildanalyse)

Auf regulären Gittern siehe Abb. 1.

**Fortsetzung Beispiel 7** (Irreguläre Gitter)

Auf irregulären Gittern wie Landkarten in der Regel: Nachbarn haben gemeinsame Grenze.

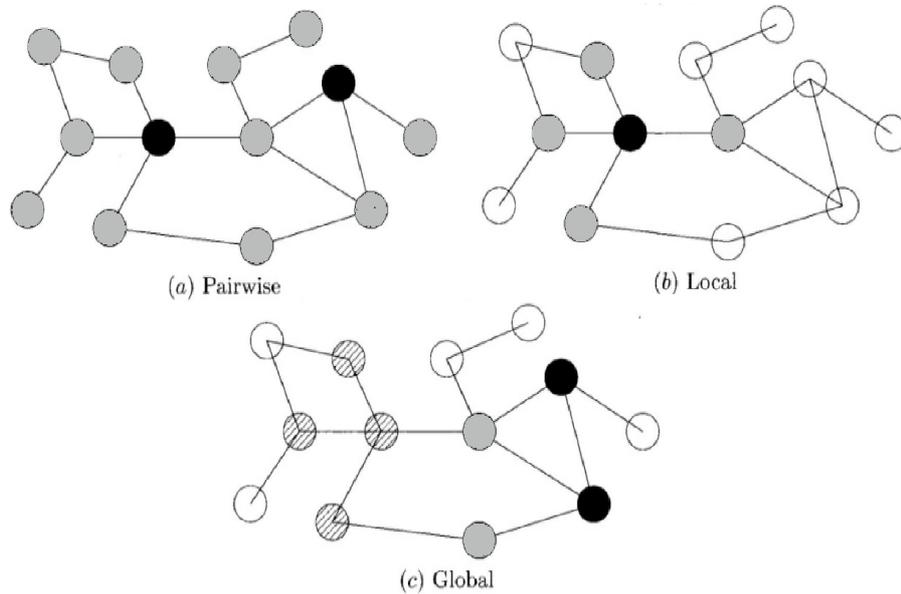
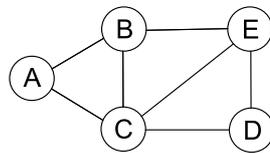


Abbildung 2: Markov-Eigenschaften: (a) Paarweise Markov-Eigenschaft, die schwarzen Knoten sind gegeben die grauen Knoten bedingt unabhängig. (b) Lokale Markov-Eigenschaft, schwarze und weiße Knoten sind bedingt unabhängig gegeben die grauen Knoten. (c) Globale Markov-Eigenschaft, schwarze und gestrichelte Knoten sind bedingt unabhängig gegeben den grauen Knoten.

**Beispiel 12.** (Ungerichtete Graphen/Markov-Netzwerke)



**Definition 2.3. (Clique)** Eine Menge  $C$  von Knoten (Indizes, Regionen, Pixeln, ...) heißt Clique:  $\Leftrightarrow$  Jedes Paar  $(i, j)$  von Knoten sind Nachbarn  $i \sim j$ .

Oft versteht man unter Clique auch die *maximale Clique*.

**Fortsetzung Beispiel 8** (Reguläre Gitter)

Erste Nachbarn

$$C_1 = \left\{ \emptyset, \square, \square-\square, \begin{array}{c} \square \\ | \\ \square \end{array} \right\}$$

Zweite Nachbarn

$$C_2 = C_1 \cup \left\{ \square-\square-\square, \begin{array}{c} \square \\ | \\ \square \\ | \\ \square \end{array} \right\}; \quad \text{nicht: } \begin{array}{c} \square-\square \\ | \\ \square \end{array}$$

**Definition 2.4** (Markov-Zufallsfeld).  $Y = \{Y_s, s \in D\}$  heißt MRF:  $\Leftrightarrow p(y_s | y_{s'}, s' \neq s) = p(y_s | y_{s'}, s' \in \partial(s))$  für alle  $s, s' \in D$

**Beispiel 13.** (Autoregressive Prozesse)

- Zeitlich:

$$Y_t = \phi Y_{t-1} + \varepsilon_t \Rightarrow Y_t | Y_{t-1}, \dots, Y_0 \sim Y_t | Y_{t-1} \sim N(\phi Y_{t-1}, \sigma^2)$$

was dem Random Walk im Fall  $\phi = 1$  entspricht.

- Räumlich:

$$Y_s = \sum_{s' \in n(s)} w_{ss'} Y_{s'} + \varepsilon_s \Rightarrow$$

$$n(s) \ni s' \sim N\left(\sum_{s' \in n(s)} w_{ss'} Y_{s'}, \sigma^2\right)$$

Wird durch Vorgabe eines räumlich autoregressiven Modells überhaupt in konsistenter Weise eine gemeinsame Verteilung definiert? Allgemeiner: Lässt sich durch Vorgabe bedingter Verteilungen in der MRF-Definition eine gemeinsame Verteilung konsistent definieren?

**Beispiel 14.** (Gegenbeispiel)

$$Y_1 | Y_2 \sim N(\alpha_0 + \alpha_1 Y_2, \sigma_1^2), \quad Y_2 | Y_1 \sim N(\beta_0 + \beta_1 Y_1^3, \sigma_2^2)$$

$$\Rightarrow \begin{aligned} \mathbb{E}[Y_1] &= \mathbb{E}[\mathbb{E}[Y_1 | Y_2]] = \mathbb{E}[\alpha_0 + \alpha_1 Y_2] = \alpha_0 + \alpha_1 \mathbb{E}[Y_2] \\ \mathbb{E}[Y_2] &= \mathbb{E}[\mathbb{E}[Y_2 | Y_1]] = \mathbb{E}[\beta_0 + \beta_1 Y_1^3] = \beta_0 + \beta_1 \mathbb{E}[Y_1^3] \end{aligned}$$

Bis auf Spezialfälle: inkompatibel! D.h.  $f(y_1|y_2)$  und  $f(y_2|y_1)$  definieren keine gemeinsame Dichte  $f(y_1, y_2)$ .

### 2.2.2 Hammersley-Clifford-Theorem

Nachbarschaftsstruktur durch Graphen gegeben:  $i, j$  Nachbarn:  $\Leftrightarrow$  direkte Kante zwischen Knoten  $i \sim j$

**Definition 2.5. (Positivität)** Seien  $T_1, \dots, T_n$  die Träger von  $Y_1, \dots, Y_n$ , d.h.  $T_i = \{y : p(y_i) > 0\}$ . Falls  $T = T_1 \times \dots \times T_n$  der Träger von  $Y = (Y_1, \dots, Y_n)$ , spricht man von Positivität.

**Satz 2.1. (Hammersley-Clifford-Theorem)**  $p(y)$  besitze die Positivitätseigenschaft, d.h.  $p(y) > 0$  für beliebige  $y_1 \in T_1, \dots, y_i \in T_i, \dots, y_n \in T_n$ . Zu einer gegebenen Nachbarschaftsstruktur ist  $Y$  ein MRF  $\Leftrightarrow$  Die gemeinsame Dichte  $p(y)$  hat bis auf die Normalisierungskonstante die Form

$$p(y) \propto \exp \left( \sum_C \phi_C(y_C) \right)$$

wobei  $C$  alle einzelnen Elemente  $i \in \{1, \dots, n\}$  und alle Cliques  $C \subseteq \{1, \dots, n\}$  durchläuft,  $y_C$  der entsprechende Teilvektor von  $C$  ist und die Funktionen  $\phi_C$  invariant gegenüber Vertauschungen / Permutationen der Komponenten von  $y_C$  sind.

#### Bemerkungen:

- Damit ist zulässige Form von gemeinsamen Dichten bei MRF charakterisiert. Die explizite Berechnung der Normalisierungskonstanten ist i.A. nicht in geschlossener analytischer Form möglich.
- Bei uns bestehen Cliques meist nur aus 2 Elementen. Dann

$$p(y) \propto \exp \left( \sum_{i=1}^n \phi_i(y_i) + \sum_{i \sim j} \phi_{i,j}(y_i, y_j) \right)$$

- Dichten dieser Art werden auch Gibbs-Feld oder Gibbs-Maß genannt.  $\phi$  heißt Gibbs-Energie.

#### Beispiel 15. (Auto-logistisches Modell)

$Y_i$  binär,  $T_i = \{0, 1\}$ ,  $i = 1, \dots, n$

Nachbarschaftsstruktur so, dass nur Cliques mit zwei Elementen  $(i, j)$  auftreten.

Hammersley-Clifford-Theorem  $\Rightarrow$

$$p(y) \propto \exp \left( \sum_i \alpha_i y_i + \sum_{i \sim j} \beta_{i,j} y_i y_j \right)$$

zulässiges MRF:

$$p(y_i | y_{-i}) \propto p(y) \propto \exp \left( \alpha_i y_i + \sum_{j \in \partial(i)} \beta_{ij} y_i y_j \right)$$

$$\Leftrightarrow \log \frac{\mathbb{P}(Y_i = 1 | Y_{-i})}{\mathbb{P}(Y_i = 0 | Y_{-i})} = \alpha_i + \sum_{j \in \partial(i)} \beta_{ij} Y_j$$

- Logit Modell; mit Kovariablen:  $\alpha_i = x_i' \alpha$
- Probit-Modell nicht zulässig! ( $\Rightarrow$  latentes GMRF mit Probitlink)

**Beispiel 16.** (Ising-Modell für Schwarz-Weiß-Bilder)

4 nächste Nachbarn  $y_i \in \{0, 1\}$   
 $i = 1, \dots, n$  kodiert weiß/schwarz für Pixel  $i$

$$p(y) \propto \exp \left( \beta \sum_{i \sim j} \mathbb{1}(y_i = y_j) \right) \mathbb{1} \text{ Indikatorfunktion}$$

entspricht Bsp. 15.

**Beispiel 17.** (Potts-Modell für Farbstufen)

$y_i \in \{1, \dots, k\}$

$$p(y) \propto \exp \left( \beta \sum_{i \sim j} \mathbb{1}(y_i = y_j) \right)$$

$\Rightarrow p(y_i | y_{-i})$  multinomiales Logit-Modell zulässig. Andere multinomiale Modelle, z.B. multinomiales Probit-Modell, ordinale Modelle, unzulässig!

**Beispiel 18.** (Gauß-Markov-Zufallsfeld)

Für erste Nachbarn:

$$p(y) \propto \exp \left( -\frac{1}{2} \sum_{i \sim j} w_{ij} (y_i - y_j)^2 \right)$$

**Beispiel 19.** (Räumlich autoregressiver Prozess bei Exponential-Familie mit kanonischem Link)

$$p(y_i | y_{-i}) \propto \exp \left( \frac{\theta_i y_i - b(\theta_i)}{\phi} \right)$$

$$\theta_i = \alpha_i y_i + \sum_{j \sim i} \beta_{ij} y_i y_j$$

### 2.2.3 Besag's Lemma

$Y = (Y_1, \dots, Y_n)$   $n$ -dimensionale ZV

$p(y_1, \dots, y_n)$  gemeinsame Dichte (bzw. W.fkt.)

$p(y_i | y_{-i}) = p(y_i | y_j, j \neq i)$  vollständig bedingte Dichten

Bei gegebener gemeinsamer Dichte sind die bedingten Dichten eindeutig bestimmt. Umgekehrt:

- Bestimmen die vollständig bedingten Dichten die gemeinsame Dichte?
- Sind Restriktionen zu beachten in der Wahl eines Modells für  $p(y_i | y_{-i})$ ?
- Wie berechnet man  $p(y_1, \dots, y_n)$ ?

**Satz 2.2. (Besag's Lemma/Brook's Lemma)**

$$\frac{p(y)}{p(\tilde{y})} = \prod_{i=1}^n \frac{p(y_i | y_1, \dots, y_{i-1}, \tilde{y}_{i+1}, \dots, \tilde{y}_n)}{p(\tilde{y}_i | y_1, \dots, y_{i-1}, \tilde{y}_{i+1}, \dots, \tilde{y}_n)}$$

**Bemerkungen:**

- Setze  $\tilde{y}$  fest, dann kann  $p(y)$  aus bedingten Verteilungen (bis auf die Normalisierungskonstante) berechnet werden.
- Satz ist unabhängig von Anordnung ( $n!$  Möglichkeiten)  
 $\Rightarrow$  Kompatibilitätsbedingung.

**Beweis:**

$$\frac{p(y_n | y_1, \dots, y_{n-1}) p(y_1, \dots, y_{n-1})}{p(\tilde{y}_n | y_1, \dots, y_n) p(y_1, \dots, y_{n-1})} = \frac{p(y_1, \dots, y_{n-1}, y_n)}{p(y_1, \dots, y_{n-1}, \tilde{y}_n)}$$

Daraus folgt

$$p(y_1, \dots, y_n) = \frac{p(y_n | y_1, \dots, y_{n-1})}{p(\tilde{y}_n | y_1, \dots, y_{n-1})} p(y_1, \dots, y_{n-1}, \tilde{y}_n)$$

Entwickle analog

$$p(y_1, \dots, y_{n-1}, \tilde{y}_n) = \frac{p(y_{n-1} | y_1, \dots, y_{n-2}, y_n)}{p(\tilde{y}_{n-1} | y_1, \dots, y_{n-2}, \tilde{y}_n)} p(y_1, \dots, y_{n-2}, \tilde{y}_{n-1}, \tilde{y}_n)$$

□

## 2.3 Gauß-Markov Zufallsfelder (GMRF)

### 2.3.1 Definition

**Definition 2.6.** Sei  $Y \sim N(\mu, \Sigma)$ . Dann heißt  $P = \Sigma^{-1}$  Präzisionsmatrix. Dabei darf  $\Sigma$  auch singular sein, also  $\text{Rang}(\Sigma) < n$ . Dann ist  $\Sigma^{-1}$  als verallgemeinerte Inverse aufzufassen.

**Satz 2.3.**  $Y$  sei normalverteilt mit  $\mathbb{E}[Y] = \mu$  und  $P = \Sigma^{-1} > 0$

$$\Rightarrow Y_i \perp Y_j | Y_{-ij} \Leftrightarrow p_{ij} = 0 \quad \text{für alle } i \neq j$$

Die Elemente  $p_{ij} \neq 0$  in  $P$  bestimmen also den Nachbarschaftsgraphen.

**Definition 2.7** (Gauß-Markov-Zufallsfeld).  $Y \sim N(\mu, P^{-1})$ , heißt GMRF bezüglich der Nachbarschaftsstruktur eines Graphen:  $\Leftrightarrow$

$$p(y) = (2\pi)^{-n/2} |P|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(y - \mu)' P (y - \mu)\right)$$

und  $p_{ij} \neq 0 \Leftrightarrow i \sim j$  ( $i$  und  $j$  sind Nachbarn).  $P > 0$  heißt Präzisionsmatrix (Inverse der Kovarianzmatrix).

**Beispiel 20.** (GMRF auf deutschen Ländern)

**Definition 2.8** (Kanonische Schreibweise der Normalverteilung).

$$\begin{aligned} Y \sim N^k(m, P) &= N(P^{-1}m, P^{-1}) \\ \Leftrightarrow p(y) &\propto |P|^{\frac{1}{2}} \exp\left(-\frac{1}{2}y' P y + m'y\right) \end{aligned}$$

### 2.3.2 Eigenschaften

**Satz 2.4** (Bedingte Verteilungen eines GMRF).  $Y \sim N(\mu, P^{-1})$  sei GMRF bezüglich eines Nachbarschaftsgraphen. Dann gilt

$$\begin{aligned} \mathbb{E}[Y_i | Y_{-i}] &= \mu_i - \frac{1}{p_{ii}} \sum_{j:j \sim i} p_{ij} (Y_j - \mu_j) \\ \text{Var}(Y_i | Y_{-i}) &= \frac{1}{p_{ii}} \\ \text{Cov}(Y_i, Y_j | Y_{-ij}) &= -\frac{p_{ij}}{\sqrt{p_{ii} p_{jj}}} \end{aligned}$$

**Satz 2.5** (Markov-Eigenschaften von GMRF). Sei  $Y(s)$  ein GMRF. Dann besitzt  $Y$  die in Def. 2.1 genannten Markov-Eigenschaften. Siehe Abb. 2)

**Satz 2.6** (Spezifikation durch vollständig bedingte Dichten).

$$\begin{aligned}\mathbb{E}[Y_i | Y_{-i}] &= \mu_i + \sum_{j:i \sim j} c_{ij}(Y_j - \mu_j) \\ \text{Var}(Y_i | Y_{-i}) &= \sigma_i^2 > 0 \\ \Rightarrow Y &\sim N(\mu, P^{-1}), \text{ also GMRF mit} \\ p_{ij} &= \begin{cases} -c_{ij}/(\sigma_i \sigma_j), & i \neq j \\ 1/\sigma_i^2, & i = j \end{cases}\end{aligned}$$

wobei  $c_{ij}$  den Gewichtungsfaktor bezeichnet.

**Beweis:** Besag's Lemma. Sei  $Y_i | Y_{-i}$  normalverteilt mit

$$\mathbb{E}[Y_i | Y_{-i}] = \mu_i + \sum_{j=1}^n c_{ij}(Y_j - \mu_j) \quad \text{und} \quad \text{Var}(Y_i | Y_{-i}) = \sigma_i^2$$

Wir zeigen mit Besag's Lemma, dass  $Y = (Y_1, \dots, Y_n)'$  multivariat normalverteilt ist. Sei  $C = (c_{ij})$ ,  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ ,  $Y_0 = \mu = (\mu_1, \dots, \mu_n)'$ . Dann gilt für die (logarithmierte) Besag's Zerlegung

$$\begin{aligned}
\log\left(\frac{p(y)}{p(y_0)}\right) &= \log\left(\prod_{i=1}^n \frac{p(y_i | \mu_1, \dots, \mu_{i-1}, y_{i+1}, \dots, y_n)}{p(\mu_i | \mu_1, \dots, \mu_{i-1}, y_{i+1}, \dots, y_n)}\right) = \\
&= \log\left(\prod_{i=1}^n \frac{p(y_i | y_1, \dots, y_{i-1}, \mu_{i+1}, \dots, \mu_n)}{p(\mu_i | y_1, \dots, y_{i-1}, \mu_{i+1}, \dots, \mu_n)}\right) = \\
&= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left( y_i - \mu_i - \sum_{j=1}^{i-1} c_{ij}(y_j - \mu_j) \right)^2 + \\
&\quad + \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left( \sum_{j=1}^{i-1} c_{ij}(y_j - \mu_j) \right)^2 = \\
&= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left[ (y_i - \mu_i)^2 - 2(y_i - \mu_i) \sum_{j=1}^{i-1} c_{ij}(y_j - \mu_j) + \right. \\
&\quad \left. + \left( \sum_{j=1}^{i-1} c_{ij}(y_j - \mu_j) \right)^2 \right] + \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \left( \sum_{j=1}^{i-1} c_{ij}(y_j - \mu_j) \right)^2 = \\
&= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} (y_i - \mu_i)^2 + \sum_{i=1}^n \sum_{j=1}^{i-1} \frac{1}{\sigma_i^2} c_{ij}(y_i - \mu_i)(y_j - \mu_j) = \\
&= -\frac{1}{2} (y - \mu)' \Sigma^{-1} (y - \mu) + (y - \mu)' \Sigma^{-1} C (y - \mu) = \\
&= -\frac{1}{2} (y - \mu)' \Sigma^{-1} (\mathbb{I} - C) (y - \mu)
\end{aligned}$$

$\Rightarrow Y$  multivariat normalverteilt mit  $\mathbb{E}[Y] = \mu$  und Präzisionsmatrix

$$P = \Sigma^{-1}(\mathbb{I} - C), \quad p_{ii} = \frac{1}{\sigma_i^2}, \quad p_{ij} = -\frac{c_{ij}}{\sigma_i^2}, \quad i \neq j$$

Vertauscht man die Reihenfolge der Komponenten  $y_i$  und  $y_j$ , muss in der Zerlegung  $i$  und  $j$  vertauscht werden  $\Rightarrow c_{ij} = c_{ji}$ . Man kann zeigen, dass keine weiteren Restriktionen auftreten.  $\square$

### 2.3.3 Arten des GMRF

#### Conditional autoregressive model (CAR)

$$\mu = 0 \quad \Rightarrow \quad Y_i | Y_{-i} \sim N\left(\sum_{j \sim i} \frac{\lambda w_{ij} Y_j}{w_{i\bullet}}, \frac{\sigma^2}{w_{i\bullet}}\right) \quad \text{mit} \quad w_{i\bullet} = \sum_{j \sim i} w_{ij}$$

$$Y \sim N(0, \sigma^2 P^{-1}) \quad \text{mit} \quad p_{ij} = \begin{cases} -\lambda w_{ij} & i \sim j \\ w_{i\bullet} & i = j \end{cases}$$

Üblicherweise  $w_{ij} \equiv 1 \Rightarrow \text{CAR}$

**Intrinsic autoregressive model (IAR/IGMRF)** CAR mit

$$\lambda = 1 \quad \Rightarrow \quad p_{ij} = \begin{cases} -1 & i \sim j \\ |\partial_i| & i = j \end{cases}$$

$\text{Rg}(P) = n - 1!$  Es liegt also eine partiel impropere Verteilung vor.

$$p(y) \propto \sigma^{-\text{Rg}(P)} \exp\left(-\frac{1}{2\sigma^2} y' P y\right)$$

Vgl. Bsp.20.

**Simultaneous autoregressive model (SAR)** Sei  $\nu \sim N(0, \Sigma)$  mit  $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$

$$Y(s_i) = \mu(s_i) + \sum_{j=1}^n b_{ij}(Y(s_j) - \mu(s_j)) + \nu(s_i)$$

mit  $b_{ii} = 0$ ;  $\mu$  ist der Erwartungswertvektor  $B$  muss nicht symmetrisch sein, aber  $(\mathbb{I} - B)$  muss invertierbar sein.

$$\Rightarrow Y \sim N(\mu, (\mathbb{I} - B)^{-1} \Sigma (\mathbb{I} - B)^{-1}),$$

also:  $P = (\mathbb{I} - B) \Sigma^{-1} (\mathbb{I} - B)$

### 2.3.4 Ziehen aus GMRF

Sei  $\theta \sim N^K(b, Q)$

**Rue-Block-Algorithmus**

1. Berechne die Cholesky-Zerlegung  $Q = LL'$
2. Löse  $Lw = b$
3. Löse  $L'u = w$
4. Ziehe einen Zufallsvektor  $z \sim N(0, \mathbb{I}_n)$
5. Löse  $L'v = z$
6. Berechne  $\theta = u + v$

Effiziente Algorithmen zur Cholesky-Zerlegung existieren insbesondere für Bandmatrizen. Für räumliche Nachbarschaftsmatrizen läßt sich durch Umordnen der Regionen eine Reduzierung der Bandweite erreichen.

**Restringierte Ziehung** Eine restringierte Ziehung unter  $A\theta = e$  erhält man durch

$$\theta^* = \theta - Q^{-1}A'(AQ^{-1}A')^{-1}(A\theta - e),$$

wobei  $\theta$  die unrestringierte Zufallsvariable ist.

**2.3.5 GMRF auf regulären Gittern**

Auf regulären Gittern lassen sich GMRF als Interaktionen von Random Walks darstellen (im Folgenden o.B.d.A. intrinsische GMRF).

**(a) Random Walk erster Ordnung (1D-GMRF)**

$$Y_i|Y_{i-1} \sim N(Y_{i-1}, \tau^{-1}) \quad \forall i = 2, \dots, n$$

$$p(y_1) \propto \text{const} \Leftrightarrow p(y) = \left(\frac{1}{\sqrt{2\pi}}\right)^{n-1} \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2} \sum_{i=2}^n (y_i - y_{i-1})^2\right)$$

$$\Leftrightarrow Y \sim N\left(0, (\tau R^{(1)})^{-1}\right)$$

$$\Leftrightarrow p(y) \propto \tau^{(n-1)/2} \exp\left(-\frac{\tau}{2} y' R^{(1)} y\right)$$

wobei

$$R^{(1)} = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & & \\ & & & \ddots & \\ & & & & \ddots \end{pmatrix}$$

Eigenwerte von  $R^{(1)}$ :

$$\lambda_i = 2 \left[ 1 - \cos \left( \frac{\pi(i-1)}{n} \right) \right]$$

vgl. [Künsch \[1994\]](#).  $\text{Rg}(R^{(1)}) = n - 1$ .

(b) **Random Walk zweiter Ordnung**

$$Y_i | Y_{i-1}, Y_{i-2} \sim N(2Y_{i-1} - Y_{i-2}, \tau^{-1}) \quad \forall i \geq 3$$

$$\begin{aligned} p(y_1) \propto p(y_2) \propto \text{const} &\Leftrightarrow Y \sim N(0, (\tau R^{(2)})^{-1}) \\ &\Leftrightarrow p(y) \propto \tau^{(n-2)/2} \exp\left(-\frac{\tau}{2} y' R^{(2)} y\right) \end{aligned}$$

$$R^{(2)} = \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \ddots & \\ & & & & & & & & \ddots \end{pmatrix}$$

wobei  $\text{Rg}(R^{(2)}) = n - 2$ .

(c) **2D-reguläres Gitter, erste Nachbarn**

Nach Hammersley-Clifford:

$$\begin{aligned} p(y) &= \left[ \left( \frac{1}{\sqrt{2\pi}} \right) \tau \right]^{(IJ-1)/2} \cdot \\ &\exp \left( -\frac{\tau}{2} \left( \sum_{i=1}^I \sum_{j=2}^J (y_{ij} - y_{ij-1})^2 + \sum_{i=2}^I \sum_{j=1}^J (y_{ij} - y_{i-1j})^2 \right) \right) \\ &\propto \tau^{(IJ-1)/2} \exp \left( -\frac{\tau}{2} \left( \sum_i y'_{i\bullet} R_J^{(1)} y_{i\bullet} + \sum_j y'_{\bullet j} R_I^{(1)} y_{\bullet j} \right) \right) \\ &\propto \tau^{(IJ-1)/2} \exp \left( -\frac{\tau}{2} \left( y' \left( R_J^{(1)} \oplus R_I^{(1)} \right) y \right) \right) \end{aligned}$$

Mit  $Y_{\bullet j}, Y_{i\bullet}$  und  $Y$  passend definiert. Es gilt:

$$\begin{aligned} \text{Rg}(R_J^{(1)} \oplus R_I^{(1)}) &= IJ - 1 \\ R_J^{(1)} \oplus R_I^{(1)} &:= \mathbb{I}_I \otimes R_J^{(1)} + R_I^{(1)} \otimes \mathbb{I}_J \end{aligned}$$

Besag and Higdon [1999] schlagen Richtungswichtung vor

$$Q = \gamma R_J^{(1)} \oplus (2 - \gamma) R_I^{(1)}$$

Es gilt: Erwartungswerte von  $Q$  sind Kroneckersummen der Erwartungswerte von  $\gamma R_J^{(1)}$  und  $(2 - \gamma) R_I^{(1)}$  als Vektoren.

Vorsicht: Modelle mit unterschiedlichen Gewichten zwischen allen Paaren lassen sich so nicht darstellen.

(d) **2D-reguläres Gitter mit zweiten Nachbarn**

$$\text{Präzisionsmatrix: } P^{(2)} = R_J^{(2)} \oplus R_I^{(2)}, \quad \text{Rg}(P^{(2)}) = IJ - 4$$

(e) **2D-reguläres Gitter mit diagonalen Nachbarn**

$$P^{(D)} = R_J^{(1)} \otimes R_I^{(1)}, \quad \text{Rg}(P^{(D)}) = (I - 1)(J - 1) \quad (\text{starker Rangabfall})$$

(f) **Kombination von (d) und (e)**

$$P^{(12)} = R_J^{(2)} \otimes R_I^{(2)} - R_J^{(1)} \otimes R_I^{(1)}, \quad \text{Rg}(P^{(12)}) = IJ - 1$$

(g) **3D-Gitter mit ersten Nachbarn**

$$P^{(3D)} = R_K^{(1)} \otimes \mathbb{I}_J \otimes \mathbb{I}_I + \mathbb{I}_K \otimes R_J^{(1)} \otimes \mathbb{I}_I + \mathbb{I}_K \otimes \mathbb{I}_J \otimes R_I^{(1)}$$

Analog für andere Nachbarschaften oder höhere Dimensionen.

(h) **Räumlich-zeitliche Felder**

Nach Clayton [1996] lassen sich Interaktion von (diskretem) Raum und (diskreter) Zeit als Kroneckerprodukt schreiben. Knorr-Held [2000] unterscheidet vier Typen:

I	unabh. Interaktion	$\text{Rg}(I_n \otimes I_T) = n \cdot T$
II	unabh. Random Walks pro Region	$\text{Rg}(I_n \otimes R) = n \cdot \text{Rg}(R)$
III	unabh. GMRF pro Zeitpunkt	$\text{Rg}(Q \otimes I_T) = \text{Rg}(Q) \cdot T$
IV	abhängige Interaktion	$\text{Rg}(Q \otimes R) = \text{Rg}(Q) \cdot \text{Rg}(R)$

Dies ist für beliebige Präzisionsmatrizen  $Q$  und  $R$  anwendbar.

## 2.4 Latente Markov-Zufallsfelder

Wie in 2.2 gesehen, sind Nicht-Gauß-MRF oft schwierig zu konstruieren. Statt dessen werden oft latente (nicht direkt beobachtbare) GMRFs benutzt. **Fortsetzung Beispiel 7** (Disease Mapping)

$y_i$  beobachtete Fälle in Region  $i$  (Zählvariable)

$e_i$  erwartete Fälle in Region  $i$  (gegeben)

offensichtlich sinnvoll:  $Y_i \sim \text{Po}(\lambda_i e_i)$   $\lambda_i$  Rate mit räumlicher Abhängigkeit  
 $\Rightarrow$  transformiere  $\lambda_i \rightarrow \log(\lambda_i) = \eta_i$  und benutze GMRF für (latentes)  $\eta_i$   
(Linkfunktion)

Als Hierarchisches Modell:

$$\begin{aligned} Y_i | \lambda_i &\stackrel{\text{i.n.d.}}{\sim} \text{Po}(\lambda_i e_i) \quad \forall i = 1, \dots, n \\ \log(\lambda_i) &\sim \text{GMRF}(\tau) \quad \forall i = 1, \dots, n \quad \text{bzw.} \quad \log(\lambda) \sim N_n(0, \tau^{-1} Q^{-1}) \end{aligned}$$

Alternativ als GAM darstellbar, dazu mehr in Kapitel 8.

### 2.4.1 Besag-York-Mollie Modell

Nach [Besag et al. \[1991\]](#) als BYM-Modell benannt:

$$\begin{aligned} Y_i | \lambda_i &\sim \text{Po}(\lambda_i e_i) \\ \lambda_i &= \exp(\gamma_i) = \exp(u_i + v_i) \\ u_i &\sim N\left(\frac{1}{|\partial_i|} \sum_{j \sim i} u_j, \frac{1}{\tau |\partial_i|}\right) \quad \text{Strukturierter Effekt} \\ v_i &\stackrel{\text{i.i.d.}}{\sim} N(0, \nu^{-1}) \quad \text{Unstrukturierter Effekt} \\ \tau &\sim \text{Ga}(a, b) \\ \nu &\sim \text{Ga}(c, d) \end{aligned}$$

Im Folgenden  $\theta = (u, v, \tau, \nu)$ . Die Posteriori ergibt sich zu

$$\begin{aligned} p(\theta | y) &\propto \prod_{i=1}^n (e_i \lambda_i)^{y_i} \exp(-e_i \lambda_i) \\ &\quad \cdot \tau^{\text{Rg}(Q)/2} \exp\left(-\frac{\tau}{2} u' Q u\right) \\ &\quad \cdot \nu^{n/2} \exp\left(-\frac{\nu}{2} v' v\right) \\ &\quad \cdot \tau^{a-1} \exp(-b\tau) \\ &\quad \cdot \nu^{c-1} \exp(-d\nu) \end{aligned}$$

#### (a) Inferenz über MCMC

$$\tau | \cdot \sim \text{Ga}, \quad \nu | \cdot \sim \text{Ga} \quad \rightarrow \quad \text{Gibbs-Schritte}$$

$$\begin{aligned} p(u | \cdot) &\propto \prod_{i=1}^n \exp\left(y_i u_i - e_i \exp(u_i + v_i) - \frac{\tau}{2} u' Q u\right) \\ p(u_i | \cdot) &\propto \exp\left(\underbrace{y_i u_i - \frac{\tau}{2} \sum_{j \sim i} (u_i - u_j)^2}_{\text{passt zu Gaußverteilung}} - e_i \exp(u_i + v_i)\right) \end{aligned}$$

Metropolis-Hastings-Schritt: verwende als Vorschlagsdichte eine Normalverteilung, die „nahe“ an Full conditional liegt, z.B. über Taylor-Entwicklung (analog für andere Verteilungen möglich):

$$f(u_i) := e_i \exp(u_i + v_i) \quad \text{für ein } i$$

$$\tilde{f}(u_i) = f(u_0) + (u_i - u_0)f'(u_0) + \frac{1}{2}(u_i - u_0)^2 f''(u_0)$$

Alternativ: IWLS-Proposal.

Damit Akzeptanzwahrscheinlichkeiten nahe 1. Analog für  $v$ . Aus  $q(u_i|\cdot)$  lässt sich  $q(u|\cdot)$  konstruieren mit identischer Struktur der Präzisionsmatrix:

$$q(u|u_0) \propto \exp\left(-\frac{1}{2}u'(\tau Q + H(u_0))u + u'(y - f(u))\right)$$

(b) **Reparametrisierungstrick**

$\tilde{\theta} = (u, \gamma, \tau, \nu)$ . Full conditionals:

$$\begin{array}{llll} u|\gamma, \tau & \propto & N_n(\cdot, \cdot) & \rightarrow \text{Gibbs-Schritt} \\ \gamma_i|u_i, \nu_i & \stackrel{\text{i.i.d.}}{\sim} & (\cdot, \cdot) & \rightarrow \text{Metropolis-Hastings-Schritt} \\ \tau|u & \propto & \text{Ga}(\cdot, \cdot) & \\ \nu|\gamma & \propto & \text{Ga}(\cdot, \cdot) & \end{array}$$

$$\begin{aligned} p(u|\gamma, \tau) & \propto \exp\left(-\frac{\tau}{2}u'Qu - \frac{\nu}{2}(\gamma - u)'I(\gamma - u)\right) \\ & \propto \exp\left(-\frac{1}{2}u'(\tau Q + \nu I) + \nu\gamma'u\right) \end{aligned}$$

$$u|\cdot \sim N^K(\nu\gamma, \underbrace{\tau Q + \nu I}_{\text{voller Rang für } \nu > 0}) \quad \text{Rue-Block-Update (vgl. 2.3.4)}$$

**Bemerkungen**

- Ziehen aus  $p(u_i|\cdot) \propto p(u_i|\partial u_i, \gamma_i, \tau)$  (single-site). Erlaubt parallelisieren, z.B. Schachbrettmuster bei regulärem Gitter.
- Gemeinsamer Update von Parametern verbessert Mixing, siehe **Knorr-Held and Rue [2002]**: z.B. ziehe  $\tilde{\tau}|u$  und  $\tilde{u}|\tilde{\tau}, \gamma$  und akzeptiere/verwerfe gemeinsam.
- $u$  und  $v$  sind nicht identifizierbar! Nur  $\gamma = u + v$  kann interpretiert werden. I.d.R. Priorerwartungswert für  $\nu$  hoch, damit für  $v$  niedrig.

(c) **Inferenz über penalisierte Likelihood**GLM-Ansatz:  $\log(\mathbb{E}[Y_i]) = u_i + v_i$ Penalisierung auf Differenzen von  $u_i$ :  $\text{Pen}(u) = \lambda_u \sum_{i \sim j} (u_i - u_j)^2$ Penalisierung auf Differenzen von  $v_i$ :  $\text{Pen}(v) = \lambda_v \sum_{i=1}^n v_i^2$  (Ridge!) $\Rightarrow$  Pen. log-Likelihood entspricht marginaler log-Posteriori  $\log(p(u, v|y))$ .Schätze  $\lambda_u$  und  $\lambda_v$  über Kreuzvalidierung.(d) **Inferenz über empirischen Bayes/REML**Schätze  $\nu$  und  $\tau$  aus den Daten, d.h. aus marginaler Posteriori, dann ML für  $u$  und  $v$  „einfach“.(e) **Inferenz über INLA**Idee: – Approximiere marginale Posteriori für  $\nu$  und  $\tau$  numerisch.– Bedingte Posteriori  $u, v | \nu, \tau, Y$  „fast“ normal.

Details zu INLA später.

**2.4.2 Hilfsvariablenansatz**Formulierung eines Binomialmodells auf Basis der Individuen (nach [Holmes and Held \[2006\]](#)):

$$\begin{aligned}
 Y_j &\sim \text{Bernoulli}(\exp(\eta_j)) \\
 \eta_j &= u_{i(j)} + v_{i(j)} \\
 u &\sim \text{N}(0, (\tau M)^{-1}) \\
 v_i &\sim \text{N}(0, \delta) \quad \text{für alle } i = 1, \dots, n
 \end{aligned}$$

mit  $Y_j \in \{0, 1\}$  für  $j = 1, \dots, n$  binäre Responsevariable des Individuums  $j$ ,  $i(j)$  Index der Region des Individuums  $j$ .

Das Modell läßt sich wie folgt mit Hilfsvariablen formulieren:

$$\begin{aligned}
 Y_j &= \begin{cases} 1 & \text{für } z_j > 0 \\ 0 & \text{sonst} \end{cases} \\
 z_j &= u_{i(j)} + v_{i(j)} + \varepsilon_j \\
 \varepsilon_j &\sim \text{N}(0, \lambda_j) \\
 \lambda_j &= (2\psi_j)^2 \\
 \psi_j &\sim \text{Kolmogorov-Smirnov} \\
 u &\sim \text{N}(0, (\tau M)^{-1}) \\
 v_i &\sim \text{N}(0, \sigma_v^2) \quad \text{für alle } i = 1, \dots, n
 \end{aligned}$$

mit  $\lambda_j = 1$  für  $j = 1, \dots, n$  erhält man ein Probitmodell. Durch geeignete Wahl der Verteilung von  $\lambda$  sind auch andere Links möglich. Wichtig: Die marginale Likelihood  $L(u, v|y)$  ist in beiden Modellformulierungen BYM und Holmes/Held-Modell gleich.

Full conditionals:

$$\begin{aligned}
u|v, z, \lambda, Y &\sim N(V^{-1}b, V^{-1}) \\
b &= (b_1, \dots, b_I) \\
b_i &= \sum_{j \in R_i} \frac{z_j - v_i}{\lambda_j} \\
V &= \tau M + W \\
W &= \text{diag} \left( \sum_{j \in R_1} \lambda_j^{-1}, \dots, \sum_{j \in R_I} \lambda_j^{-1} \right) \\
v|u, z, \lambda, Y &\sim N(\tilde{V}^{-1}\tilde{b}, \tilde{V}^{-1}) \\
\tilde{b} &= (\tilde{b}_1, \dots, \tilde{b}_I) \\
\tilde{b}_i &= \sum_{j \in R_i} \frac{z_j - u_i}{\lambda_j} \\
\tilde{V} &= \delta I + W \\
\tau &\sim \text{Ga} \left( a + \frac{n-1}{2}, b + \frac{1}{2} u' M u \right) \\
\delta &\sim \text{Ga} \left( a + \frac{n}{2}, b + \frac{1}{2} \sum_{i=1}^n v_i^2 \right)
\end{aligned}$$

Im Unterschied zum BYM-Modell kann  $v_i$  hier in jeder Region  $i \in \{1, \dots, I\}$  geschätzt werden. Liegen in einer Region keine Daten vor, so ist die Posteriori von  $v_i$  gleich der Priori.

### 2.4.3 Latente Ising/Potts-Felder

**Beispiel 21.** (Segmentierung)

Auf regulärem (Pixel-)Gitter:  $Y_i = \delta_i \alpha_i + (1 - \delta_i) \beta_i; \quad \delta_i \in \{0, 1\}$

$$\begin{aligned}
\alpha_i &\sim N(\mu_\alpha, \sigma_\alpha^2) \\
\beta_i &\sim N(\mu_\beta, \sigma_\beta^2) \\
p(\delta) &\propto \exp \left( \beta_1 \sum_{i \sim j} \delta_i \delta_j + \beta_2 \sum_{i \sim j} (1 - \delta_i)(1 - \delta_j) \right)
\end{aligned}$$

Inferenz über:

- MCMC
- Simulated Annealing ( $\beta_{\bullet} \rightarrow \infty$ )
- MAP (Maximum a Posteriori)
- EM-Algorithmus (Gemischte Verteilung)

Vgl. Vorlesung BioImaging.