# LØSNINGSFORSLAG, ROMLIG STATISTIKK

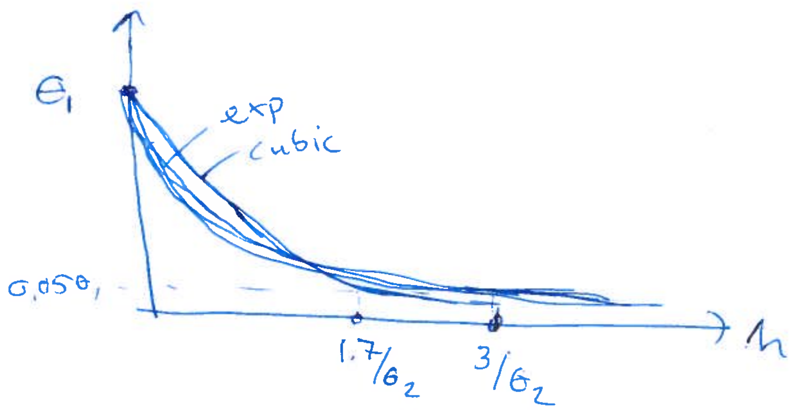
## EKSAMEN 2011, 23. MAI

1 a) Kovariansen er uavhengig av lokasjon og av retning.

$$\Sigma(s,s';\theta) = \Sigma(h;\theta), \quad h = \|s - s'\|$$


* $e^{-\theta_2 h} = 0{,}05$, $\quad h = -\frac{\ln 0{,}05}{\theta_2} \approx \underline{3/\theta_2}$

* $\frac{1}{(1+\theta_2 h)^3} = 0{,}05$ $\quad h = \frac{\left(1/\sqrt[3]{0{,}05}\right) - 1}{\theta_2} \approx \underline{1.7/\theta_2}$



1 b) $\underset{\sim}{y} = \begin{pmatrix} y(s_1) \\ \vdots \\ y(s_n) \end{pmatrix}, \quad X = \begin{pmatrix} x^t(s_1) \\ \vdots \\ x^t(s_n) \end{pmatrix}, \quad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon(s_1) \\ \vdots \\ \varepsilon(s_n) \end{pmatrix}$

$$\underset{\sim}{y} = X \cdot \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}, \quad \underset{\sim}{\varepsilon} \sim N(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} \theta_1 & \theta_1 e^{-\theta_2 h_{12}} & \cdots & \theta_1 e^{-\theta_2 h_{1n}} \\ & \ddots & & \vdots \\ e^{-\theta_2 h_{1,n}} & & \cdots & \theta_1 \end{pmatrix} = \theta_1 \cdot \begin{pmatrix} 1 & e^{-\theta_2 h_{12}} & \cdots & e^{-\theta_2 h_{1,n}} \\ & \ddots & & \\ e^{-\theta_2 h_{1n}} & & & 1 \end{pmatrix}$$

$$l(\beta, \theta_1) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(\underset{\sim}{y} - X\underset{\sim}{\beta})^t \Sigma^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta})}$$

$$\Sigma = \theta_1 \cdot R(\theta_2) = \theta_1 \cdot R$$

$$l(\beta, \theta_1) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\theta_1^{n/2} |R|^{1/2}} e^{-\frac{1}{2\theta_1}(\underset{\sim}{y} - X\underset{\sim}{\beta})^t R^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta})}$$

$$L(\beta, \theta_1) = \ln l(\beta, \theta_1) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |R| - \frac{n}{2} \ln \theta_1 - \frac{1}{2\theta_1} (\underset{\sim}{y} - X\underset{\sim}{\beta})^t R^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta})$$

Maximum likelihood:

i) $\frac{\partial L}{\partial \beta} = -\frac{1}{\theta_1} X^t R^{-1} (\underset{\sim}{y} - X\beta) = 0 \Rightarrow \hat{\underset{\sim}{\beta}} = [X^t R^{-1} X]^{-1} X^t R^{-1} \underset{\sim}{y}$

ii) $\frac{\partial L}{\partial \theta_1} = -\frac{n}{2} \frac{1}{\theta_1} + \frac{1}{2\theta_1^2} (\underset{\sim}{y} - X\underset{\sim}{\beta})^t R^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta}) = 0$

$$\Rightarrow \hat{\theta}_1 = \frac{1}{n} (\underset{\sim}{y} - X\underset{\sim}{\beta})^t R^{-1} (\underset{\sim}{y} - X\underset{\sim}{\beta})$$

Plugger inn $\hat{\beta}$ fra i)

ii

1 c) Prediksjonsvariansen ved punkt $s_0$ er her

$$\sigma_{s_0}^2 = 1 - R_{0,1:n} R^{-1} R_{0,1:n}^t$$

A og C er nærmest datapkt.
Verdien av data betyr ingenting, kun korrelasjon regnet ut fra romlig avstander.
For A og C er $R_{0,1:n}$ 'stor' og vil minske variansen fra 1.
For C er de fleste punktene i nærheten, i samme retning av C, og nær hverandre.
Dette gjør at $R^{-1}$ blir liten, og effektiv reduksjon av varians er mindre pga datapunktene blir så korrelerte.
A har minste prediksjonsvarians. ($0{,}60^2$)
B har størst ($0.87^2$)

1 d) $\xi(s)=0,\ \xi(s')=0$ ; $h=\|s-s'\|$

$$Cov(\varepsilon(s),\varepsilon(s')|0,0) = Cov(\varepsilon_0(s),\varepsilon_0(s')) = \underline{e^{-\theta_{0,2}h}}$$

$\xi(s)=0,\ \xi(s')=1$

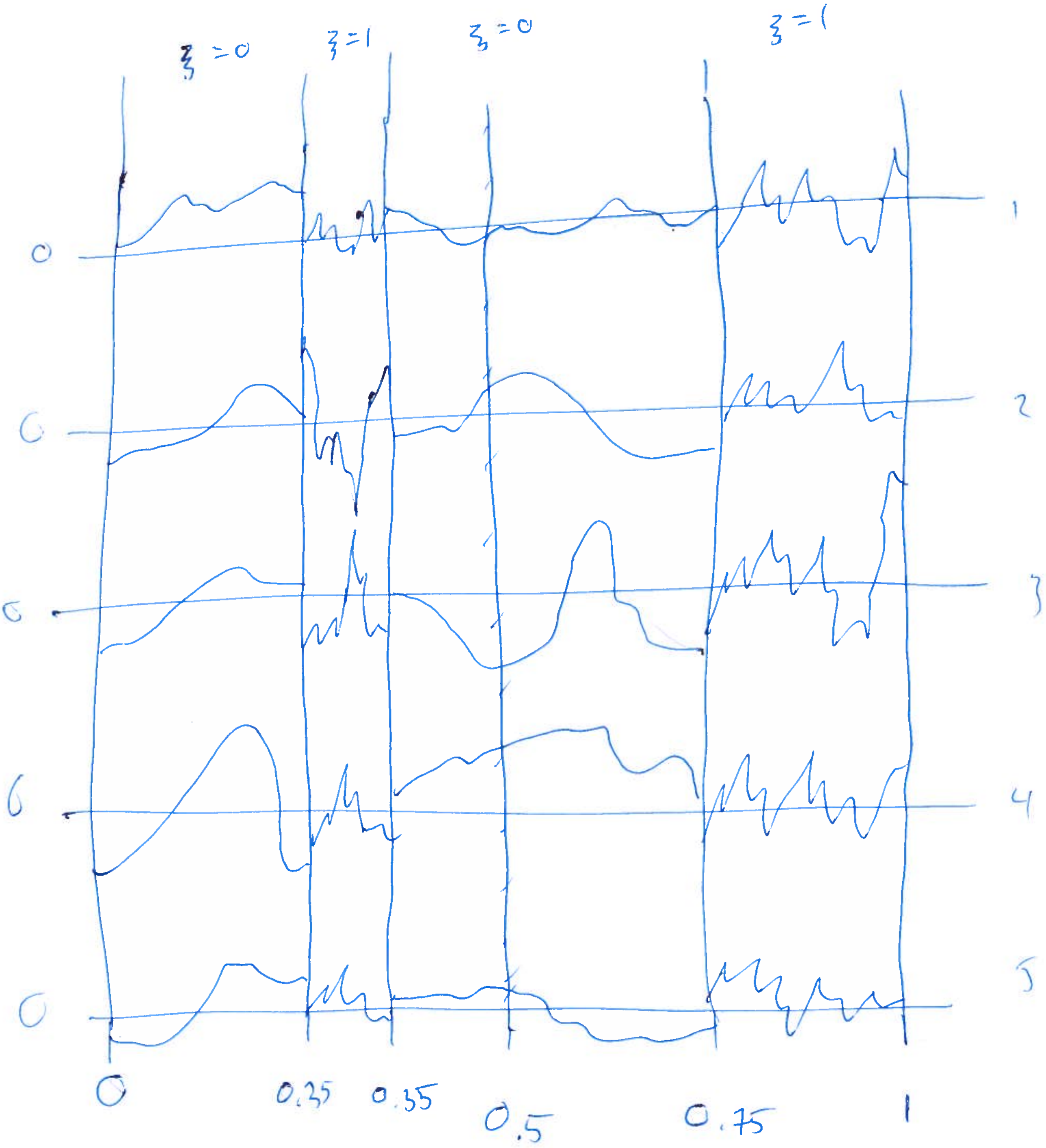
$$Cov(\varepsilon(s),\varepsilon(s')|0,1) = Cov(\varepsilon_0(s),\varepsilon_1(s')) = \underline{0}$$

$\xi(s)=1,\ \xi(s')=0$

$$Cov(\varepsilon(s),\varepsilon(s')|1,0) = Cov(\varepsilon_1(s),\varepsilon_0(s')) = \underline{0}$$

$\xi(s)=1,\ \xi(s')=1$

$$Cov(\varepsilon(s),\varepsilon(s')|1,1) = Cov(\varepsilon_1(s),\varepsilon_1(s')) = \underline{e^{-\theta_{1,2}h}}$$




IV

$$Z = \begin{pmatrix} Z(s) \\ Z(s') \end{pmatrix}$$

1 e)

$$p(z) = \frac{1}{2\pi} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2} z^t \Sigma^{-1} z} , \quad \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

$$\rho = e^{-\theta_{2,2} h} , \quad h = \| s - s' \|$$

$$|\Sigma| = \begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix} = 1 - \rho^2 = \alpha$$

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} , \quad \beta = \frac{\rho}{1-\rho^2}$$

$$-\frac{1}{2} z^t \begin{pmatrix} 1 & -\beta \\ -\beta & 1 \end{pmatrix} z = -\frac{1}{2} z_1^2 - \frac{1}{2} z_2^2 + \beta z_1 z_2$$
$$= -\frac{1}{2} z_2^2 - \frac{1}{2\alpha} (z_1 - \rho z_2)^2$$

$$\int_{-\infty}^{0} \int_{-\infty}^{0} \frac{1}{2\pi} \frac{1}{\sqrt{\alpha}} e^{-\frac{1}{2} z_2^2} \cdot e^{-\frac{1}{2\alpha}(z_1 - \rho z_2)^2} dz_1 \, dz_2$$

$$= P( Z(s) < 0 \cap Z(s') < 0 ) = P( \xi(s) = 0 \cap \xi(s') = 0 )$$

$$\boxed{\begin{aligned} u_1 &= \frac{z_1 - \rho z_2}{\sqrt{\alpha}} \\ du_1 &= \frac{1}{\sqrt{\alpha}} dz_1 \end{aligned}}$$

$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z_2^2} \cancel{\frac{1}{\sqrt{\alpha}}} \int_{-\infty}^{-\rho/\sqrt{\alpha} z_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} u_1^2} \cdot \cancel{\sqrt{\alpha}} \, du_1 \, dz_2$$

$$= \int_{-\infty}^{0} \varphi(z_2) \cdot \Phi\left(-\frac{\rho}{\sqrt{\alpha}} z_2\right) dz_2$$

$$= \frac{\arctan\left(-\frac{1}{\rho/\sqrt{\alpha}}\right)}{2\pi} = \underline{\underline{\frac{\arctan\left(-\frac{\sqrt{\alpha}}{\rho}\right)}{2\pi}}}$$

v

1 f)

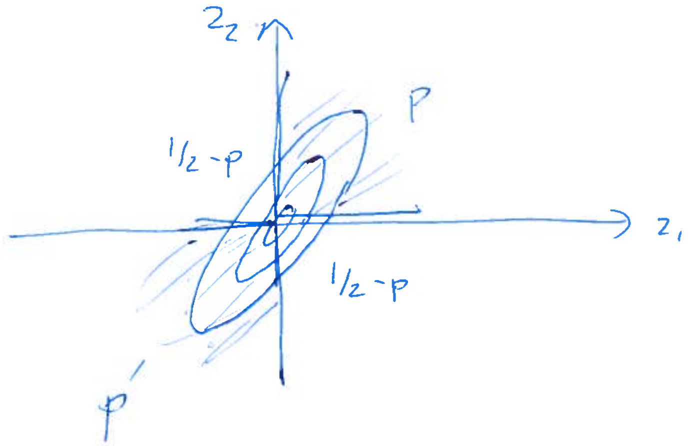
Dobbel kovarians:

$$\mathrm{Cov}(\varepsilon(s),\varepsilon(s')) = E\left(\mathrm{Cov}(\varepsilon(s),\varepsilon(s') \mid \xi(s),\xi(s'))\right) + \mathrm{Cov}\big(\underbrace{E(\varepsilon(s) \mid \xi(s),\xi(s'))}_{0},\ \underbrace{E(\varepsilon(s') \mid \xi(s),\xi(s'))}_{0}\big)$$

$$= \sum_{k=0}^{1} \sum_{l=0}^{1} \mathrm{Cov}\left(\varepsilon(s),\varepsilon(s') \mid \xi(s)=k,\ \xi(s')=l\right) \cdot P(\xi(s)=k \cap \xi(s')=l)$$

$$= e^{-\theta_{0,2}\cdot h}\cdot p + e^{-\theta_{1,2}\cdot h}\cdot p + (0+0)\left(\tfrac{1}{2}-p\right)$$

der $p = \dfrac{\arctan\left(-\frac{\sqrt{x}}{g}\right)}{2\pi}$




$$= p\left(e^{-\theta_{0,1}h} + e^{-\theta_{1,2}h}\right)$$

$$= \frac{1}{2\pi}\arctan\left(-\frac{\sqrt{1-e^{-2\theta_{2,2}h}}}{e^{-\theta_{2,2}\cdot h}}\right)\cdot\left(e^{-\theta_{0,1}h} + e^{-\theta_{1,2}h}\right)$$

2 a) $N_1$ = antall pkt i $(0,0.1)\times(0,0.1)$

$$P(0 \text{ pkt i } (0,0.1)\times(0,0.1)) = P(N_1=0) = e^{-0.1^2\lambda} = \underline{\underline{e^{-\frac{\lambda}{100}}}}$$

P(minst ett av rutene har 0 punkt)

$= 1 - P(\text{alle rutene har mer enn 0 punkt})$

$$= 1 - \left[P(N_1>0)\cdot P(N_2>0) \cdots P(N_{100}>0)\right]$$

$$= 1 - \left[1 - P(N_1=0)\right]^{100}$$

$$= \underline{1 - \left[1 - e^{-\frac{\lambda}{100}}\right]^{100}}$$

$$1 - \left(1 - e^{-\frac{\lambda}{100}}\right)^{100} = 0.05$$

$$\ln 0.95 = 100 \cdot \ln\left(1 - e^{-\frac{\lambda}{100}}\right)$$

$$\frac{\ln 0.95}{100} = -0,0005 = \ln\left(1 - e^{-\frac{\lambda}{100}}\right)$$

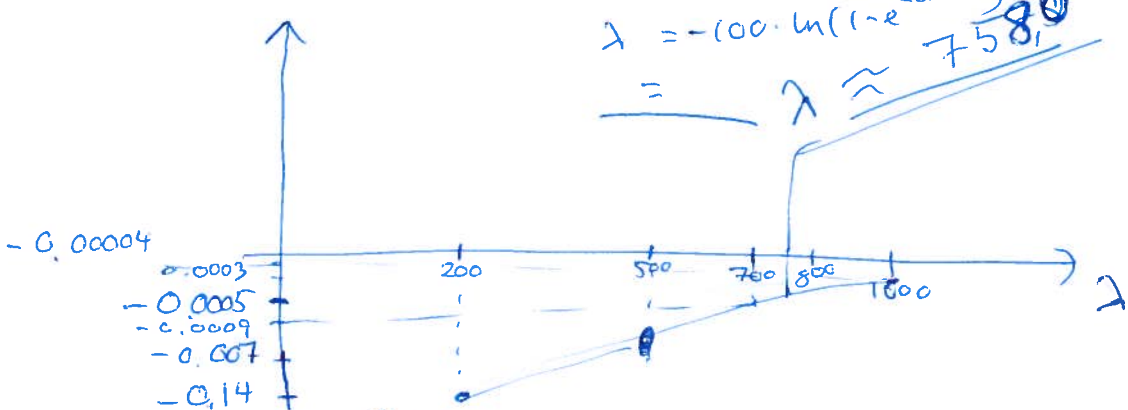
$$e^{-0,005} = 1 - e^{-\lambda/100}$$

$$\frac{\lambda}{100} = \ln(1 - e^{-0,0005})$$

$$\lambda = -100 \cdot \ln(1 - e^{-0.0005})$$

$$= \lambda \approx 758,0$$

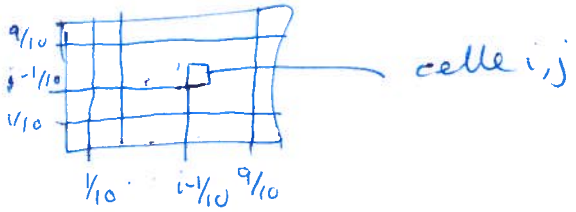
$$\ln\left(1 - e^{-\frac{\lambda}{100}}\right)$$




$\lambda \approx$

vii

2b)




$$\int_a^b \int_c^d \lambda_0 xy \, dx\, dy = \frac{\lambda_0}{4}(b^2-a^2)(d^2-c^2)$$

$$\boxed{\begin{matrix} a = i-1, b = i \\ c = j-1, d = j \end{matrix}}$$

$$= \frac{\lambda_0}{4}\left(\frac{i^2}{100} - \frac{(i-1)^2}{100}\right)\left(\frac{j^2}{100} - \frac{(j-1)^2}{100}\right)$$

$$= \frac{\lambda_0}{4\,100^2}(2i-1)(2j-1)$$

$N_{ij}$ = antall pkt i celle i,j

$$P(N_{ij}=0) = \exp\left(-\frac{\lambda_0}{4 \cdot 100^2}(2i-1)(2j-1)\right)$$

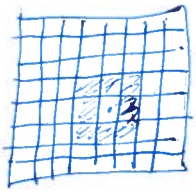
Bruk tynning av en homogen Poisson-prosess for å trekke en ikke-homogen prosess.

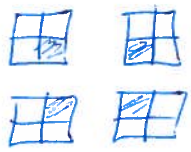
Homogen prosess med $\lambda = \lambda_0$ fungerer.
Dette gir altfor mange pkt i celle 1,1, og de tynnes med sannsynlighet $x_i y_i$.

Fra B Monte Carlo realisasjoner kan vi se andelen av realisasjoner som har 0 punkter i minst en av cellene. Andelen er et estimat på sannsynligheten.

$$\hat{p} = \frac{\#\text{ realisasjoner med 0 i minst en celle}}{B}$$

3 a)

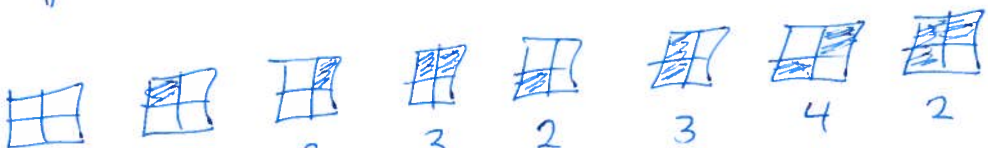




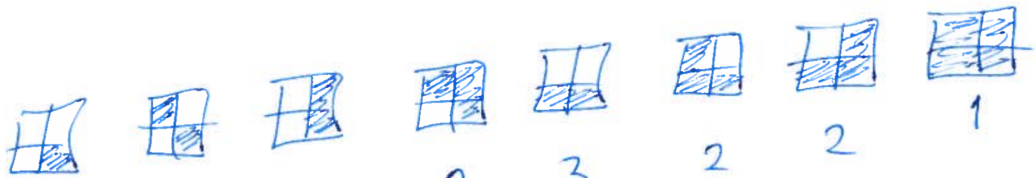
cliquer i 2 ordens naboskap.

cliquer er definert ved at celler inngår i hverandres naboskap.

Clique konfigurasjoner



1 2 2 3 2 3 4 2



2 4 3 2 3 2 2 1

| | | |
|---|---|---|
| Ensfarget | = 1 | { 2 stk } |
| Hjørne | = 2 | { 8 stk } |
| Rad/kolonne | = 3 | { 4 stk } |
| Diagonal | = 4 | { 2 stk } |

3 b)

$$PL(\varphi) = \prod_{i=1}^{nm} P(x_i \mid x_j\, ; j \in N_i)$$

der $N_i$ er naboskapet til celle $i$

$$P(x_i \mid x_j\, ; j \in N_i) \propto \exp\Big( \sum_{k \in C} \varphi(x_k) \Big)$$

$$\propto \exp\Big( \sum_{\substack{k \in C \\ i \in k}} \varphi(x_k) \Big)$$

$$P(x_i \mid x_j\, ; j \in N_i) = \frac{e^{\sum_{k \in C, i \in k} \varphi(x_k \mid x_i)}}{e^{\sum_{k \in C, i \in k} \varphi(x_k \mid x_i)} + e^{\sum_{k \in C, i \in k} \varphi(x_k \mid x_i^c)}}$$

$k \in C$, $i \in k$ betyr alle cliquer der celle $i$ inngår

 = 4 stk cliquer

$x_i^c$ er komplementet av dataverdi $x_i$.

Dersom alle naboer er 0, og $x_i = 0$ blir

$$P(x_i \mid x_j\, ; j \in N_i) = \frac{e^{4\varphi_1}}{e^{4\varphi_1} + e^{4\varphi_2}}$$

pga $\varphi_1$ er ensfarget potensial, $\varphi_2$ er hjørne potensial.

Dette produktet gjentar seg $n_{1,1}$ = antall naboskap med bare 0-ere.

vi får tilsvarende med bare 1-ere.
Argumentet videreføres til hjørner, kolonner, men totalt er $PL(\varphi)$ et vanskelig uttrykk av

$$\frac{e^{\sum_k \varphi(x_k \mid x_i)}}{e^{\sum_k \varphi(x_k \mid x_i)} + e^{\sum_k \varphi(x_k \mid x_i^c)}}$$ som maksimeres <u>numerisk</u>.

Typisk Newton's metode, eller annen optimering.

X