



Norges teknisk-naturvitenskapelige universitet
 Institutt for matematiske fag

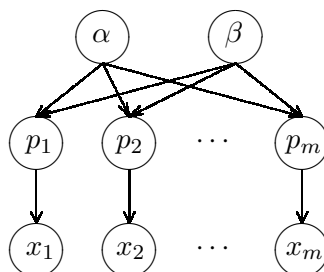
TMA4300 Mod. stat.
 metoder

Løsningsforslag - Eksamen mai 2006

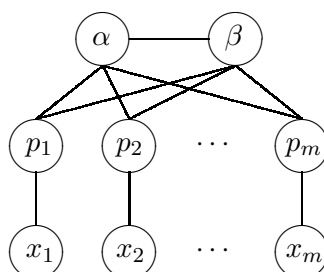
Oppgave 1

a)

DAGen blir



Fra en DAG finner man uavhengighetsgrafen ved å moralisere, dvs. ved å fjerne retningen på kantene og trekke kanter mellom/“gifte” noder med felles barn. Fra DAGen over får man dermed at uavhengighetsgrafen blir



b)

Som funksjon av p_i får man at (den første overgangen følger fra uavhengighetsgrafen)

$$f(p_i | \alpha, \beta, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_m, x_1, \dots, x_m) = f(p_i | \alpha, \beta, x_i) \\
\propto f(p_i, \alpha, \beta, x_i) = f(\alpha) f(\beta) f(p_i | \alpha, \beta) f(x_i | p_i) \propto f(p_i | \alpha, \beta) f(x_i | p_i)$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1-p_i)^{\beta-1} \binom{n}{x} p_i^{x_i} (1-p_i)^{n-x_i} \propto \underline{\underline{p_i^{\alpha+x_i-1} (1-p_i)^{\beta+(n-x_i)-1}}}.$$

Gjenkjenner dette som en betafordeling med parametre $\tilde{\alpha} = \alpha + x_i$ og $\tilde{\beta} = \beta + (n - x_i)$. Normaliseringskonstanten er dermed gitt fra formel (1) i oppgaven, dvs.

$$\frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} = \frac{\Gamma(\alpha + \beta + n)}{\Gamma(\alpha + x_i)\Gamma(\beta + n - x_i)}.$$

c)

Forslagsfordelinga er $g(p_i) = 1, p_i \in [0, 1]$. Akseptsannsynligheten blir dermed

$$a(p_i) = c \frac{f(p_i | \alpha, \beta, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_m, x_1, \dots, x_m)}{g(p_i)} = c' p_i^{\alpha+x_i-1} (1-p_i)^{\beta+n-x_i-1}$$

der c' må bestemmes slik at dette blir en sannsynlighet. Deriverer derfor uttrykket over og setter lik null

$$c' \left[(\alpha + x_i - 1) p_i^{\alpha+x_i-2} (1-p_i)^{\beta+n-x_i-1} + p_i^{\alpha+x_i-1} (\beta + n - x_i - 1) (1-p_i)^{\beta+n-x_i-2} \right] = 0$$

$$\Rightarrow \dots \Rightarrow p_i = \frac{\alpha + x_i - 1}{\alpha + \beta + n - 2}.$$

Setter så denne verdien for p_i inn i uttrykket for $a(p_i)$ og får dermed følgende krav på c'

$$c' \left(\frac{\alpha + x_i - 1}{\alpha + \beta + n - 2} \right)^{\alpha+x_i-1} \left(\frac{\beta + n - x_i - 1}{\alpha + \beta + n - 2} \right)^{\beta+n-x_i-1} \leq 1.$$

Velger c' størst mulig slik at dette kravet er oppfylt, og får (etter litt regning)

$$\underline{\underline{a(p_i) = \left(\frac{\alpha + \beta + n - 2}{\alpha + x_i - 1} p_i \right)^{\alpha+x_i-1} \left(\frac{\alpha + \beta + n - 2}{\beta + n - x_i - 1} (1-p_i) \right)^{\beta+n-x_i-1}}}.$$

Pseudokode for simuleringen er:

1. Trekk $p_i \sim \text{Unif}[0, 1]$.
2. Beregn $a(p_i)$.
3. Trekk $u \sim \text{Unif}[0, 1]$.
4. Hvis $u \leq a(p_i)$ returner p_i , hvis ikke gå til 1.

d)

Her finnes det selvfølgelig mange riktige svar. Mest naturlig er det kanskje å benytte en single-site-Metropolis-Hastings-algoritme. For eksempel spesifisert ved følgende pseudokode

1. Initialiser $\alpha, \beta, p_1, \dots, p_m$ med lovlige verdier.
2. For $r = 1, \dots, N$
 - (a) For $i = 1, \dots, m$
 - i. trekk $p_i \sim f(p_i | \alpha, \beta, p_1, \dots, p_{i-1}, p_{i+1}, \dots, p_m, x_1, \dots, x_m)$ [Ved algoritma fra forrige punkt, dvs. akseptanssynligheten er lik 1 p.g.a. et Gibbs-steg.]
 - (b) Foreslå $\tilde{\alpha} \sim N(\alpha, \sigma_\alpha^2)$ [For en passende verdi av σ_α^2 .]
 - (c) Trekk $u_\alpha \sim \text{Unif}[0, 1]$.
 - (d) Hvis $u_\alpha \leq a(\tilde{\alpha} | \alpha)$ sett $\alpha := \tilde{\alpha}$, ellers behold α uforandret [Akseptanssynligheten $a(\tilde{\alpha} | \alpha)$ er spesifisert under.]
 - (e) Foreslå $\tilde{\beta} \sim N(\beta, \sigma_\beta^2)$ [For en passende verdi av σ_β^2 .]
 - (f) Trekk $u_\beta \sim \text{Unif}[0, 1]$.
 - (g) Hvis $u_\beta \leq a(\tilde{\beta} | \beta)$ sett $\beta := \tilde{\beta}$, ellers behold β uforandret [Akseptanssynligheten $a(\tilde{\beta} | \beta)$ er spesifisert under.]

Forslagene for α og β benyttet her er begge metropolisforslag slik at forslagssannsynligheten forkortes bort i uttrykket for akseptanssynligheten. Innsetting i det generelle uttrykket for akseptanssynlighet i en metropolisalgoritme gir (etter litt regning, hvor blant annet det som kan forkortes er forkortet)

$$a(\tilde{\alpha} | \alpha) = \begin{cases} \min \left\{ e^{-\lambda(\tilde{\alpha} - \alpha)} \left[\frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\alpha + \tilde{\beta})} \cdot \frac{\Gamma(\alpha)}{\Gamma(\tilde{\alpha})} \right]^m (\prod_{i=1}^m p_i)^{\tilde{\alpha} - \alpha}, 1 \right\} & \text{hvis } \tilde{\alpha} > 0, \\ 0 & \text{ellers,} \end{cases}$$

og

$$a(\tilde{\beta} | \beta) = \begin{cases} \min \left\{ e^{-\lambda(\tilde{\beta} - \beta)} \left[\frac{\Gamma(\tilde{\alpha} + \tilde{\beta})}{\Gamma(\alpha + \tilde{\beta})} \cdot \frac{\Gamma(\beta)}{\Gamma(\tilde{\beta})} \right]^m (\prod_{i=1}^m (1 - p_i))^{\tilde{\beta} - \beta}, 1 \right\} & \text{hvis } \tilde{\beta} > 0, \\ 0 & \text{ellers,} \end{cases}$$

e)

Burn-in-perioden ser ut til å være på ca. 600-700 iterasjoner. Etter dette ser den simulerte prosessen ut til å være stasjonær, før dette tidspunkt er det en klar trend for p_1 og p_3 .

Lar $\{\alpha^{(r)}, \beta^{(r)}, p_1^{(r)}, \dots, p_m^{(r)}\}_{r=1}^N$ være de simulerte verdier i iterasjon nummer r etter burn-in. Da blir

1. $\tilde{p}_5 = \frac{1}{N} \sum_{r=1}^N p_5^{(r)}$. Dette er estimat for $E(p_5|x_1, \dots, x_m)$.
2. $\hat{p} = \sum_{i=1}^m \hat{p}_i = \frac{1}{N} \sum_{r=1}^N \left[\sum_{i=1}^m p_i^{(r)} \right]$. Dette er estimat for $E(p_1 + \dots + p_m|x_1, \dots, x_m)$.
3. For hver iterasjon r må man her også simulere hvilke medlemmer som faktisk reserverer bil. Dvs. lar $s_i^{(r)}$ være indikatoren som er en dersom medlem nummer i reserverer bil (i iterasjon r), og simulerer disse (uavhengig av hverandre) ved å trekke $u_i^r \sim \text{Unif}[[0, 1]]$ og sette

$$s_i^{(r)} = \begin{cases} 1 & \text{hvis } u_i^r \leq p_i^{(r)}, \\ 0 & \text{ellers.} \end{cases}$$

Antall medlemmer som reserverer bil i iterasjon r blir dermed

$$\sum_{i=1}^m s_i^{(r)}$$

og et estimat for forventet antall medlemmer som reserverer bil men ikke får bil til disposisjon blir dermed

$$\frac{1}{N} \sum_{r=1}^N \max \left\{ \sum_{i=1}^m s_i^{(r)} - k, 0 \right\}.$$

Oppgave 2

a)

$$S = \sum_{i=1}^{478} [0 \cdot I(N_i = 0) + 300 \cdot I(N_i = 1) + 400 \cdot I(N_i = 2)],$$

hvor $I(A) = 1$ hvis A er sann og $I(A) = 0$ ellers. Har videre at

$$\begin{aligned} E(S) &= \sum_{i=1}^{478} \{0 \cdot E[I(N_i = 0)] + 300 \cdot E[I(N_i = 1)] + 400 \cdot E[I(N_i = 2)]\} \\ &= 478[0 \cdot P(N = 0) + 300 \cdot P(N = 1) + 400 \cdot P(N = 2)]. \end{aligned}$$

Estimerer sannsynlighetsfordelingen for N ved

$$\hat{P}(N = 0) = \frac{3}{10}, \quad \hat{P}(N = 1) = \frac{4}{10} \quad \text{og} \quad \hat{P}(N = 2) = \frac{4}{10}.$$

Estimat for forventet verdi av S blir dermed

$$\hat{E}(S) = 478[0 \cdot \hat{P}(N=0) + 300 \cdot \hat{P}(N=1) + 400 \cdot \hat{P}(N=2)] = \dots = 114720.$$

b)

Pseudokode for å estimere $\text{Var}[\hat{E}(S)]$ ved hjelp av ikke-parametrisk bootstrapping blir

1. For $b = 1, \dots, B$

(a) Trekk $N_1^b, \dots, N_{10}^b$ tilfeldig (med tilbakelegging) fra 0, 1, 1, 2, 2, 0, 1, 1, 0, 2.

(b) Beregn

$$\hat{P}^b(N=0) = \frac{1}{10} \sum_{i=1}^{10} I(N_i^b=0) = \frac{1}{10} M_0^b,$$

$$\hat{P}^b(N=1) = \frac{1}{10} \sum_{i=1}^{10} I(N_i^b=1) = \frac{1}{10} M_1^b,$$

$$\hat{P}^b(N=2) = \frac{1}{10} \sum_{i=1}^{10} I(N_i^b=2) = \frac{1}{10} M_2^b$$

og

$$\hat{E}^b(S) = 478[0 \cdot \hat{P}^b(N=0) + 300 \cdot \hat{P}^b(N=1) + 400 \cdot \hat{P}^b(N=2)]$$

2. Beregn estimert varians ved

$$\widehat{\text{Var}}[\hat{E}(S)] = \frac{1}{B} \sum_{b=1}^B \left[\hat{E}^b(S) - \frac{1}{B} \sum_{b=1}^B \hat{E}^b(S) \right]^2.$$

c)

Benytter at

$$(M_0^b, M_1^b, M_2^b) \sim \text{Multinomisk}(10, 0.3, 0.4, 0.4)$$

og

$$\hat{E}_B(S) = 478[0 \cdot \frac{M_0}{10} + 300 \cdot \frac{M_1}{10} + 400 \cdot \frac{M_2}{10}],$$

slik at

$$\begin{aligned} \text{Var}[\hat{E}_B(S)] &= \left(\frac{478}{10} \right)^2 [0^2 \cdot \text{Var}(M_0) + 300^2 \cdot \text{Var}(M_1) + 400^2 \cdot \text{Var}(M_2) \\ &\quad + 2 \cdot 0 \cdot 300 \cdot \text{Cov}(M_0, M_1) + 2 \cdot 0 \cdot 400 \cdot \text{Cov}(M_0, M_2) + 2 \cdot 300 \cdot 400 \cdot \text{Cov}(M_1, M_2)] \\ &= \left(\frac{478}{10} \right)^2 [300^2 \cdot 10 \cdot 0.4 \cdot 0.6 + 400^2 \cdot 10 \cdot 0.3 \cdot 0.7 + 2 \cdot 300 \cdot 400 \cdot (-10 \cdot 0.3 \cdot 0.4)] = \underline{\underline{603197760}} = \underline{\underline{24560.08^2}}. \end{aligned}$$