

NB. Geef een duidelijke toelichting bij de antwoorden. Het gebruik van een rekenmachine is toegestaan, maar geen programmeerbare/grafische rekenmachine, mobiele telefoon of laptop. Succes!

Normering: 1a) 3, 1b) 3, 1c) 3, 2a) 3, 2b) 3, 2c) 3, 3a) 3, 3b) 3, 3c) 3.

Vraag 1 (*Markov modellen*)

De dagelijkse waterstand bij Lobith wordt met een 1^{ste} orde Markov keten gemodelleerd. Onder reguliere omstandigheden fluctueert de waterstand tussen een laag, normaal en hoog niveau. Met kans α neemt het niveau toe en met kans β af, maar het niveau kan ook gelijk blijven. Met kans γ is er een te veel aan smeltwater, welk een kritiek niveau kan veroorzaken. Het is ongewis hoe lang dit kritieke niveau aanhoudt, elke dag kan het met kans δ zakken. Merk op: de fluctuaties zijn een geleidelijk proces en sprongen van meer dan één niveau komen niet voor.

Vraag 1a)

Geef de toestandsruimte en transitie-matrix van het boven beschreven Markov proces met daarbij de parameter restrictie. Teken ook het *state diagram* met daarin bijhorende overgangskansen.

Ga bij de resterende onderdelen van deze vraag uit van de volgende transitie-matrix :

$$\mathbf{P} = \begin{pmatrix} 1 - \omega & \omega & 0 & 0 \\ \omega & 1 - 2\omega & \omega & 0 \\ 0 & \omega & 1 - 2\omega & \omega \\ 0 & 0 & \xi & 1 - \xi \end{pmatrix}.$$

waarbij rijen- en kolommen-volgorde corresponderen met oplopende waterstanden. Merk op: deze transitie-matrix kent een andere parameterizatie (wat de betekenis van de elementen niet verandert).

Vraag 1b)

Heeft dit 1^{ste} orde Markov proces een stationaire verdeling? Zo ja, geef deze.

Vraag 1c)

In januari zijn gedurende elf dagen de volgende waterstanden gemeten: **normaal, normaal, hoog, hoog, hoog, kritiek, kritiek, hoog, hoog, hoog, hoog**. Gebruik de maximum likelihood methode om de parameters ω en ξ op basis van deze observaties te schatten. Veronderstel hierbij de kans om op dag 1 met een normale waterstand te beginnen, gelijk aan 1.

Vraag 2 (*Phylogenetische bomen*)

Twee hedendaagse organismen met een binaire genetische code (bestaande uit nullen en éénen) hebben een gemeenschappelijke voorouder. Het substitutie-proces in een willekeurige locus volgt een 1^{ste} orde Markov proces. De 1-staps transitie-matrix van dit proces is:

$$\mathbf{P} = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

De stationaire verdeling van dit substitutie-proces is $(\varphi_0, \varphi_1)^\top = (\frac{4}{7}, \frac{3}{7})^\top$.

Vraag 2a)

De gemeenschappelijk voorouder is een grootouder. Wat is de kans dat de hedendaagse organismen in een locus verschillen? Voor de verdeling 0- en 1-en in gemeenschappelijke voorouder mag de stationaire verdeling worden aangenomen.

Vraag 2b)

Onder dezelfde aannames als bij vraag 2a, bereken nu de kans dat één van de hedendaagse organismen een 0 bevat gegeven dat het andere organisme in dezelfde locus een 1 heeft.

Vraag 2c)

De gemeenschappelijk voorouder leefde heel (!) lang geleden. De hedendaagse organismen verschillen in een locus. Wat is de meest aannemelijk toestand van diezelfde locus in de gemeenschappelijke voorouder?

Vraag 3 (*Netwerken*)

Beschouw een pathway van 3 genen. De expressie niveaus van genen 1 en 2, aangeduid met random variabelen Y_1 en Y_2 , volgen een bivariate verdeling:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3\frac{1}{2} & 3 \\ 3 & 4 \end{pmatrix}\right)$$

De expressie van het derde gen, Y_3 , wordt gegeven door de regressie vergelijking: $Y_3 = \beta_1 Y_1 + \beta_2 Y_2 + \varepsilon$. De random variable ε , welk de ruis representeert, volgt een standaard normale verdeling en is onafhankelijk van zowel Y_1 als Y_2 .

Vraag 3a)

Wat impliceert de relatie $Y_1 \perp\!\!\!\perp Y_3 | Y_2$ voor de coëfficiënten van de regressie-vergelijking $Y_3 = \beta_1 Y_1 + \beta_2 Y_2 + \varepsilon$ binnen dit 3-gen pathway?

Voor de resterende opgaven kan de relatie $Y_1 \perp\!\!\!\perp Y_3 | Y_2$ niet verondersteld worden. Wel geldt nu dat $\beta_1 = 1 = \beta_2$.

Vraag 3b)

Bepaal $\text{Cor}(Y_2, Y_3)$.

Vraag 3c)

Welke conclusies m.b.t. de conditionele onafhankelijkheden binnen het 3-gen pathway zijn gerechtvaardigd op basis van de bovenstaande informatie (vraag 3a dus niet meenemende)?

FORMULE BLAD

Bij het tentamen kunnen de volgende formules handig zijn.

De inverse van een 2×2 matrix \mathbf{A} met elementen $a_{j_1, j_2} = (\mathbf{A})_{j_1, j_2}$ is:

$$\mathbf{A}^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}^{-1} = [\det(\mathbf{A})]^{-1} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}$$

met $\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$.

De inverse van een 3×3 matrix \mathbf{A} met elementen $a_{j_1, j_2} = (\mathbf{A})_{j_1, j_2}$ is:

$$\mathbf{A}^{-1} = [\det(\mathbf{A})]^{-1} \begin{pmatrix} a_{33}a_{22} - a_{32}a_{23} & -(a_{33}a_{12} - a_{32}a_{13}) & a_{23}a_{12} - a_{22}a_{13} \\ -(a_{33}a_{21} - a_{31}a_{23}) & a_{33}a_{11} - a_{31}a_{13} & -(a_{23}a_{11} - a_{21}a_{13}) \\ a_{32}a_{21} - a_{31}a_{22} & -(a_{32}a_{11} - a_{31}a_{12}) & a_{22}a_{11} - a_{21}a_{12} \end{pmatrix}$$

met $\det(\mathbf{A}) = a_{11}(a_{33}a_{22} - a_{32}a_{23}) - a_{21}(a_{33}a_{12} - a_{32}a_{13}) + a_{31}(a_{23}a_{12} - a_{22}a_{13})$.

De dichtheidsfunctie van de multivariaat normale verdeling van random variable \mathbf{Y} is

$$f(\mathbf{Y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp[-(\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) / 2],$$

met $\boldsymbol{\mu}$ en $\boldsymbol{\Sigma}$ de verwachting en covariantie parameters, respectievelijk.

Indien een p -variante normaal verdeelde random variabele \mathbf{Z} als volgt gepartitioneerd kan worden:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{pmatrix} \right),$$

dan wordt de conditionele verdeling van $\mathbf{Y}|\mathbf{X}$ gegeven door:

$$\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_{XX}^{-1} (\mathbf{X} - \boldsymbol{\mu}_X), \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX} \boldsymbol{\Sigma}_{XX}^{-1} \boldsymbol{\Sigma}_{XY}).$$

Zij A en B twee gebeurtenissen. De regel van Bayes zegt dan: $P(A|B) = P(B|A) \cdot P(A) / P(B)$.

Zij A en B_1, \dots, B_n gebeurtenissen z.d.d. $\sum_{i=1}^n P(B_i) = 1$. De *total probability law* zegt dan:

$$P(A) = \sum_{i=1}^n P(A, B_i).$$

Zij \mathbf{W} , \mathbf{X} , \mathbf{Y} en \mathbf{Z} random vectoren, \mathbf{A} en \mathbf{B} non-random matrices van geschikte dimensies, en c een constante. Dan geldt: $\text{Cov}(c, \mathbf{Y}) = \mathbf{0}$, $\text{Cov}(\mathbf{Y}, \mathbf{Y}) = \text{Var}(\mathbf{Y})$, $\text{Cov}(\mathbf{Y}, \mathbf{Z}) = \mathbf{0}$ als \mathbf{Y} en \mathbf{Z} onafhankelijk zijn, $\text{Cov}(\mathbf{A}\mathbf{X}, \mathbf{B}\mathbf{Y}) = \mathbf{A}\text{Cov}(\mathbf{X}, \mathbf{Y})\mathbf{B}^\top$, en

$$\text{Cov}(\mathbf{W} + \mathbf{X}, \mathbf{Y} + \mathbf{Z}) = \text{Cov}(\mathbf{W}, \mathbf{Y}) + \text{Cov}(\mathbf{W}, \mathbf{Z}) + \text{Cov}(\mathbf{X}, \mathbf{Y}) + \text{Cov}(\mathbf{X}, \mathbf{Z}).$$

Antwoorden

Antwoord op vraag 1

Vraag 1a

De toestandsruimte is $\mathcal{S} = \{\text{laag, normaal, hoog, kritiek}\}$. De transitie-matrix wordt:

$$\mathbf{P} = \begin{pmatrix} 1 - \alpha & \alpha & 0 & 0 \\ \beta & 1 - \alpha - \beta & \alpha & 0 \\ 0 & \beta & 1 - \beta - \gamma & \gamma \\ 0 & 0 & \delta & 1 - \delta \end{pmatrix}.$$

De rijen sommeren tot 1, zoals vereist. Verder, uit het feit dat elk element van \mathbf{P} in het interval $[0, 1]$ moet liggen, volgt de parameter-restrictie: $\alpha, \beta, \gamma, \delta \in [0, 1]$, alsook $\alpha + \beta \leq 1$ en $\beta + \gamma \leq 1$. Includeer ook het *state diagram*.

Vraag 1b

Ja, heeft stationaire verdeling: irreducibel (na verloop van tijd kan elke toestand bereikt worden) en aperiodiek (geen toestand wordt met vaste regelmaat aangedaan). Gebruik dan: $\varphi^\top \mathbf{P} = \varphi^\top$ en $\varphi_\ell + \varphi_n + \varphi_h + \varphi_k = 1$. Dit geeft het volgende stelsel van vergelijkingen:

$$\begin{aligned} (1 - \omega)\varphi_\ell + \omega\varphi_n &= \varphi_\ell, \\ \omega\varphi_\ell + (1 - 2\omega)\varphi_n + \omega\varphi_h &= \varphi_n, \\ \omega\varphi_n + (1 - 2\omega)\varphi_h + \xi\varphi_k &= \varphi_h, \\ \omega\varphi_h + (1 - \xi)\varphi_k &= \varphi_k, \\ \varphi_\ell + \varphi_n + \varphi_h + \varphi_k &= 1. \end{aligned}$$

Laat de middelste vergelijking vallen en herschrijf de eerste drie overgeblevene:

$$\begin{aligned} \omega\varphi_n &= \omega\varphi_\ell, \\ \omega\varphi_h &= -\omega\varphi_\ell + 2\omega\varphi_n, \\ \omega\varphi_h &= \xi\varphi_k, \\ \varphi_\ell + \varphi_n + \varphi_h + \varphi_k &= 1. \end{aligned}$$

Uit de eerste vergelijking volgt $\varphi_\ell = \varphi_n$. Substitueer dit in de tweede vergelijking en concludeer: $\varphi_\ell = \varphi_n = \varphi_h$. Gebruik dit vervolgens in de vierde vergelijking. We hebben nu het stelsel:

$$\begin{aligned} \omega\varphi_h &= \xi\varphi_k, \\ \varphi_k &= 1 - 3\varphi_\ell. \end{aligned}$$

Oplossen levert: $\varphi_\ell = \varphi_n = \varphi_h = \omega/(\xi + 3\omega)$ en $\varphi_k = \xi/(\xi + 3\omega)$. Deze kansen liggen in het interval $[0, 1]$ en sommeren tot een.

Vraag 1c

De likelihood van deze sequentie wordt gegeven door: $P(X_1 = \mathbf{n}) \prod_{t=2}^{10} P(X_t | X_{t-1})$. Oftewel, gebruikmakende van de gegeven transitiematrix \mathbf{P} : $1 \times (1 - 2\omega) \times \omega \times (1 - 2\omega) \times (1 - 2\omega) \times \omega \times (1 - \xi) \times \xi \times (1 - 2\omega) \times (1 - 2\omega) \times (1 - 2\omega) = (1 - 2\omega)^6 \omega^2 (1 - \xi) \xi$. Neem de logaritme, deze is proportioneel aan: $6 \log(1 - 2\omega) + 2 \log(\omega) + \log(1 - \xi) + \log(\xi)$. Stel eerste orde afgeleiden naar ω en ξ gelijk aan nul: $\frac{-12}{1-2\omega} + \frac{2}{\omega} = 0$ en $\frac{-1}{1-\xi} + \frac{1}{\xi} = 0$. Oplossen levert: $\hat{\omega} = 1/8$ en $\hat{\xi} = 1/2$.

Antwoord op vraag 2

Vraag 2a

De grootouder vormt $t = 0$, de hedendaagse soorten leven dus op $t = 2$. We kijken naar 2-staps overgangskansen. Die worden gegeven door:

$$\mathbf{P}^2 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}.$$

Gebruikmakende van de totale kans wet en de definitie van de conditionele kans:

$$\begin{aligned} P(X_{t=2}^{(1)} \neq X_{t=2}^{(2)}) &= P(X_{t=2}^{(1)} = 0, X_{t=2}^{(2)} = 1) + P(X_{t=2}^{(1)} = 1, X_{t=2}^{(2)} = 0) \\ &= \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 0, X_{t=2}^{(2)} = 1, X_{t=0}^{(ca)} = x) \\ &\quad + \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 1, X_{t=2}^{(2)} = 0, X_{t=0}^{(ca)} = x) \\ &= \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 0, X_{t=2}^{(2)} = 1 | X_{t=0}^{(ca)} = x) P(X_{t=0}^{(ca)} = x) \\ &\quad + \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 1, X_{t=2}^{(2)} = 0 | X_{t=0}^{(ca)} = x) P(X_{t=0}^{(ca)} = x) \\ &= \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 0 | X_{t=0}^{(ca)} = x) P(X_{t=2}^{(2)} = 1 | X_{t=0}^{(ca)} = x) P(X_{t=0}^{(ca)} = x) \\ &\quad + \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 1 | X_{t=0}^{(ca)} = x) P(X_{t=2}^{(2)} = 0 | X_{t=0}^{(ca)} = x) P(X_{t=0}^{(ca)} = x) \\ &= 2 \sum_{x \in \{0,1\}} P(X_{t=2}^{(1)} = 0 | X_{t=0}^{(ca)} = x) P(X_{t=2}^{(2)} = 1 | X_{t=0}^{(ca)} = x) P(X_{t=0}^{(ca)} = x) \\ &= 2 \times \left[\frac{4}{7} \times 0.61 \times 0.39 + \frac{3}{7} \times 0.51 \times 0.48 \right] = 0.4858286. \end{aligned}$$

Vraag 2b Impliciet gebruikmakende van een onderdeel uit onder 2a, tesamen met de definitie van conditionele kans en de totale kans wet:

$$\begin{aligned} P(X_{t=2}^{(1)} = 0 | X_{t=2}^{(2)} = 1) &= P(X_{t=2}^{(1)} = 0, X_{t=2}^{(2)} = 1) / P(X_{t=2}^{(2)} = 1) \\ &= \frac{1}{2} 0.4858286 / \left[\sum_{x \in \{0,1\}} P(X_{t=2}^{(2)} = 1 | X_{t=0}^{(ca)} = x) P(X_{t=0}^{(ca)} = x) \right] \\ &= \frac{1}{2} \times 0.4858286 / \left[\frac{4}{7} \times 0.39 + \frac{3}{7} \times 0.48 \right] = 0.5668. \end{aligned}$$

Vraag 2c

Gevraagd:

$$\arg \max_{x \in \{0,1\}} P(X_{t=0}^{(ca)} = x | X_t^{(1)} = 0, X_t^{(2)} = 1)$$

Merk op:

$$\mathbf{P}^t = \begin{pmatrix} \varphi_0 & \varphi_1 \\ \varphi_0 & \varphi_1 \end{pmatrix}.$$

voor grote t . Dus:

$$\begin{aligned}
 P(X_{t=0}^{(ca)} = x | X_t^{(1)} = 0, X_t^{(2)} = 1) \\
 &= P(X_t^{(1)} = 0, X_t^{(2)} = 1 | X_{t=0}^{(ca)} = x) \frac{P(X_{t=0}^{(ca)} = x)}{P(X_t^{(1)} = 0, X_t^{(2)} = 1)} \\
 &= P(X_t^{(1)} = 0 | X_{t=0}^{(ca)} = x) P(X_t^{(2)} = 1 | X_{t=0}^{(ca)} = x) \frac{P(X_{t=0}^{(ca)} = x)}{P(X_t^{(1)} = 0, X_t^{(2)} = 1)} \\
 &= \frac{\varphi_0 \varphi_1 \varphi_x}{P(X_t^{(1)} = 0, X_t^{(2)} = 1)}.
 \end{aligned}$$

Dit bereikt een maximum als $x = 0$ daar $\varphi_0 > \varphi_1$.

Antwoord op vraag 3

Vraag 3a

Dit conditioneel onafhankelijkheidsstatement impliceert dat de corresponderende partiële correlatie nul is. Daar de partiële correlaties en de regressie-coëfficiënten een één-op-één relatie hebben (i.h.b. als de een gelijk aan nul is, is de ander dat ook), is de corresponderende regressie-coëfficiënt (hier β_1) ook gelijk aan nul. Merk op: over de ander β_2 weten we niets.

Vraag 3b

De gevraagde covariantie:

$$\begin{aligned}
 \text{Cov}(Y_2, Y_3) &= \text{Cov}(Y_2, Y_1 + Y_2 + \varepsilon_1) \\
 &= \text{Cov}(Y_2, Y_1) + \text{Cov}(Y_2, Y_2) + \text{Cov}(Y_2, \varepsilon_1) \\
 &= \text{Cov}(Y_2, Y_1) + \text{Var}(Y_2, Y_2) \\
 &= 3 + 4 = 7.
 \end{aligned}$$

en de variantie:

$$\begin{aligned}
 \text{Var}(Y_3) &= \text{Cov}(Y_1 + Y_2 + \varepsilon_1, Y_1 + Y_2 + \varepsilon_1) \\
 &= \text{Var}(Y_1, Y_1) + \text{Var}(Y_2, Y_2) + \text{Var}(\varepsilon_1, \varepsilon_1) + 2\text{Cov}(Y_1, Y_2) + 2\text{Cov}(Y_1, \varepsilon_1) + 2\text{Cov}(Y_2, \varepsilon_1) \\
 &= 3\frac{1}{2} + 4 + 1 + 2 \times 3 = 14\frac{1}{2}.
 \end{aligned}$$

Bovenstaande afleidingen gebruiken de onafhankelijkheid tussen ε_1 en Y_1 als wel Y_2 .

De correlatie wordt nu gegeven door:

$$\text{Cor}(Y_2, Y_3) = \frac{\text{Cov}(Y_2, Y_3)}{\sqrt{\text{Var}(Y_2)}\sqrt{\text{Var}(Y_3)}} = \frac{7}{\sqrt{4}\sqrt{14\frac{1}{2}}} \approx 0.919145.$$

Vraag 3c

Uit de opgave en bovenstaande opgaven volgt:

$$\Sigma = \begin{pmatrix} 3\frac{1}{2} & 3 & 6\frac{1}{2} \\ 3 & 4 & 7 \\ 6\frac{1}{2} & 7 & 14\frac{1}{2} \end{pmatrix}.$$

Eenvoudig rekenwerk (welk beperkt kan worden tot de boven-diagonaal) levert:

$$\Sigma^{-1} = \begin{pmatrix} 1.8 & -0.4 & -1.0 \\ -0.4 & 1.1 & -1.0 \\ -1.0 & -1.0 & 1.0 \end{pmatrix}$$

Daar de inverse covariantie matrix 1-op-1 relateert aan de partiele correlaties corresponderen de conditionele onafhankelijkheden met nullen in Σ^{-1} . Dus, $Y_1 \perp\!\!\!\perp Y_3 \mid Y_2$, $Y_1 \perp\!\!\!\perp Y_3 \mid Y_2$, en $Y_2 \perp\!\!\!\perp Y_3 \mid Y_1$. Ofwel, er zijn geen conditionele onafhankelijkheden.