Department of Mathematics VU University Amsterdam	Herexamen: Voortgezette biostatistiek 2018, juli 11.

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, 1d), 2a), 2b) 3, 3a) 3, 3b) 3, 3c) 3.

Vraag 1 *(Markov modellen)*
Het zapp-gedrag (i.e. waar het volgende TV-programma gekeken wordt) van de TV-kijker tussen de publieke omroep en on-demand streaming diensten Netflix en HBO wordt gemodelleerd met een 1^{ste} order Markov keten. Omdat Netflix en HBO meer series aanbieden is de kans, α genoemd, dat een kijker het volgende programma daar ook kijkt vijf maal zo groot als op de publieke omroep. Het Netflix-aanbod is twee keer zo groot als dat van HBO: de kans, β genoemd, dat een kijker van HBO naar Netflix zappt is derhalve twee zo groot als de kans dat er in omgekeerde richting gezappt wordt. De kans om van de publieke omroep naar Netflix danwel HBO te zappen is even groot.

Vraag 1a)
Geef de toestandsruimte en transitie-matrix van het boven beschreven Markov proces met daarbij de parameter-restricties. Teken ook het *state diagram* met daarin bijhorende overgangskansen.

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

$$\mathbf{P} \quad = \quad \begin{pmatrix} \omega & \omega & \omega \\ \frac{1}{2} - \xi & \frac{1}{2} & \xi \\ \frac{1}{2} - 3\xi & 3\xi & \frac{1}{2} \end{pmatrix}.$$

waarbij rijen- en kolommen-volgorde corresponderen met programma-aanbieders (nu voor de eenvoud) `A`, `B`, en `C` (genoemd). Dit verandert de betekenis van de elementen van de transitie-matrix niet.

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

Vraag 1c)
Is dit 1^{ste} orde Markov proces reversibel? Neem hierbij aan dat $(\varphi_A, \varphi_B, \varphi_C)^\top = \frac{1}{6}(1, 2, 3)^\top$.

Vraag 1d)
Het zapp-gedrag van een TV-kijker is geregistreerd. De achtereenvolgende programma's werden aangeboden door aanbieder: `A`, `A`, `C`, `C`, `C`, `C`, `C`, `C`, `B`, `B`, `A`. Gebruik de maximum likelihood methode om de parameters ω en ξ op basis van deze observaties te schatten. Veronderstel hierbij de kans om het eerste programma bij aanbieder `A` te zien gelijk aan 1.

Vraag 2 *(Hidden Markov model)*
Definieer het hidden Markov model $(\boldsymbol{\pi}, \mathbf{P}, \mathbf{B})$ met toestandsruimte $\mathcal{S} = \{1, 2, 3\}$, en alfabet $\mathcal{A} =$

$\{\texttt{a}, \texttt{b}, \texttt{c}\}$ waarbij $\boldsymbol{\pi} = (1, 0, 0)^\top$,

$$\mathbf{P} \quad = \quad \begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.0 \\ 0.0 & 1.0 & 0.0 \end{pmatrix}, \text{ en } \quad \mathbf{B} \quad = \quad \begin{pmatrix} 0.5 & 0.5 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.0 & 0.5 & 0.5 \end{pmatrix}.$$

Vraag 2a)
Bereken $P((Y_1, Y_2, Y_3) = (\texttt{a}, \texttt{b}, \texttt{c}))$.

Vraag 2b)
Wat is de meest aannemelijk toestandssequentie die ten grondslag ligt aan de geobserveerde sequentie $P((Y_1, Y_2, Y_3) = (\texttt{a}, \texttt{c}, \texttt{a}))$?

Vraag 3 *(Netwerken)*
Beschouw een pathway van 3 genen. De expressie-niveaus van deze genen, aangeduid met random variabelen Y_1, Y_2 en Y_3, volgen een trivariate normale verdeling:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \quad \sim \quad \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 4 \end{pmatrix} \right).$$

Vraag 3a)
Bepaal welke partiële correlaties tussen de drie genen gelijk aan nul zijn.

Vraag 3b)
Geef de verdeling van Y_1 gegeven die van Y_2.

Vraag 3c)
De expressie-niveau's van een vierde gen, genoteerd met Y_4, hangen af van de anderen genen middels de relatie: $Y_4 = 1 + 3Y_1 - 2Y_3 + \varepsilon_4$ met ε_4 standaard normaal verdeeld en onafhankelijk van Y_1, Y_2 en Y_3. Bepaal de variantie van Y_4.

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 dichtsheidsfunctie 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-variate 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 $B_1, \ldots, B_n$ 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: $\mathrm{Cov}(c, \mathbf{Y}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{Y}, \mathbf{Y}) = \mathrm{Var}(\mathbf{Y})$, $\mathrm{Cov}(\mathbf{Y}, \mathbf{Z}) = \mathbf{0}$ als $\mathbf{Y}$ en $\mathbf{Z}$ onafhankelijk zijn, $\mathrm{Cov}(\mathbf{AX}, \mathbf{BY}) = \mathbf{A}\mathrm{Cov}(\mathbf{X}, \mathbf{Y})\mathbf{B}^\top$, en

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

Antwoorden
Antwoord op vraag 1
Vraag 1a
De toestandsruimte is $\mathcal{S} = \{\texttt{Publieke omroep}, \texttt{HBO}, \texttt{Netflix}\}$. Of, afgekort $\mathcal{S} = \{\texttt{P}, \texttt{H}, \texttt{N}\}$. Met betrekking tot de overgangskansen leren we uit het gegeven het volgende. Zij $P(X_{t+1} = \texttt{P} \,|\, X_t = \texttt{P}) = \alpha$, dan $P(X_{t+1} = \texttt{H} \,|\, X_t = \texttt{H}) = 5\alpha = P(X_{t+1} = \texttt{N} \,|\, X_t = \texttt{N})$. Verder, $P(X_{t+1} = \texttt{H} \,|\, X_t = \texttt{P}) = P(X_{t+1} = \texttt{N} \,|\, X_t = \texttt{P})$. En ook $P(X_{t+1} = \texttt{N} \,|\, X_t = \texttt{H}) = 2P(X_{t+1} = \texttt{H} \,|\, X_t = \texttt{N})$. Gebruiken we tot slot dat de rijen van een transitie-matrix tot een sommeren, dan levert dat:

$$\mathbf{P} \;=\; \begin{pmatrix} \alpha & \frac{1}{2}(1-\alpha) & \frac{1}{2}(1-\alpha) \\ 1-5\alpha-2\beta & 5\alpha & 2\beta \\ 1-5\alpha-2\beta & \beta & 5\alpha \end{pmatrix}.$$

Alles kansen dienen in het interval $[0, 1]$ te liggen. De vrije parameter α en β dienen derhalve te voldoen aan: $0 \leq \alpha, \beta \leq 1$ zodanig dat ook $5\alpha + 2\beta \leq 1$. Includeer ook het *state diagram*.

Vraag 1b
Ja, heeft stationaire verdeling: irreducibel (elke toestand kan in eindige tijd bereikt worden vanuit elke andere) en aperiodiek (overgangskansen zijn niet gelijk aan 1, dus toestanden worden niet met een vast ritme aangedaan). Gebruik dan: $\boldsymbol{\varphi}^\top \mathbf{P} = \boldsymbol{\varphi}^\top$ en $\varphi_A + \varphi_B + \varphi_C = 1$. Met weglating van de eerste vergelijking geeft het volgende stelsel van vergelijkingen:

$$\begin{aligned} \omega\varphi_A + \tfrac{1}{2}\varphi_B + 3\xi\varphi_C &= \varphi_B, \\ \omega\varphi_A + \xi\varphi_B + \tfrac{1}{2}\varphi_C &= \varphi_C, \\ \varphi_A + \varphi_B + \varphi_C &= 1. \end{aligned}$$

Herschrijf de vergelijkingen:

$$\begin{aligned} \omega\varphi_A + 3\xi\varphi_C &= \tfrac{1}{2}\varphi_B, \\ \omega\varphi_A + \xi\varphi_B &= \tfrac{1}{2}\varphi_C, \\ \varphi_A + \varphi_B + \varphi_C &= 1. \end{aligned}$$

Uit de laatste vergelijking volgt: $\varphi_A = 1 - \varphi_B - \varphi_C$. Trek nu de tweede vergelijking van de eerste af. Dit resulteert in $\varphi_B = (1+6\xi)(1+2\xi)^{-1}\varphi_C$. Merk nu op dat $\omega = \frac{1}{3}$ (omdat de rijen van $\mathbf{P}$ tot een sommeren). Substitueer dit in de eerste vergelijking:

$$\tfrac{1}{3} + (3\xi - \tfrac{1}{3})\varphi_C \;=\; (\tfrac{1}{2} + \tfrac{1}{3})(1+6\xi)(1+2\xi)^{-1}\varphi_C.$$

Ofwel:

$$\begin{aligned} 2 &= (1+2\xi)^{-1}[5(1+6\xi) - (1+2\xi)(18\xi - 2)]\varphi_C \\ &= (1+2\xi)^{-1}(7 + 16\xi - 36\xi^2)\varphi_C. \end{aligned}$$

Dus:

$$\begin{aligned} \varphi_C &= 2(1+2\xi)(7+16\xi-36\xi^2)^{-1}, \\ \varphi_B &= 2(1+6\xi)(7+16\xi-36\xi^2)^{-1}, \\ \varphi_A &= 1 - 2(2+8\xi)(7+16\xi-36\xi^2)^{-1}. \end{aligned}$$

Evident: $\varphi_A + \varphi_B + \varphi_C = 1$. Rest te verifieren dat de waarden allemaal tussen nul en een liggen. Gebruik hiervoor dat $\xi < \frac{1}{6}$ (volgt uit de parameterrestricties die garanderen dat $\mathbf{P}$ goed gedefinieerd is. Dan: $0 < 6+16\xi < 7+16\xi-36\xi^2$, waaruit het nu duidelijk is dat $\varphi_A, \varphi_B, \varphi_C \in [0,1]$

Vraag 1c
Reversibiliteit toetst men met behulp van de detailed balance equations:

$$\varphi_A(\mathbf{P})_{A,B} \quad = \quad \tfrac{1}{6}\omega \quad \neq \quad \tfrac{2}{6}(\tfrac{1}{2}-\xi) \quad = \quad \varphi_B(\mathbf{P})_{B,A}.$$

Het model is derhalve niet reversibel.

Vraag 1d
De likelihood van deze sequentie wordt gegeven door: $P(X_1 = \mathtt{A}) \prod_{t=2}^{11} P(X_t \,|\, X_{t-1})$. Ofwel: $1 \times \omega \times \omega \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times 3\xi \times \frac{1}{2} \times (\frac{1}{2} - \xi)$. Neem de logaritme (dit verandert de lokatie van het maximum niet), deze is proportioneel aan: $2\log(\omega) + \log(\frac{1}{2} - \xi) + \log(2\xi)$. Zoals eerder opgemerkt $\omega = \frac{1}{3}$ (zie onderdeel b). Rest ξ te schatten. Stel de afgeleide naar ξ gelijk aan nul: $\frac{-1}{1/2-\xi} + \frac{1}{\xi} = 0$. Oplossen levert: $\hat{\xi} = 1/4$.

Antwoord op vraag 2
Vraag 2a
Vanuit de gegeven parameters van het hidden Markov model valt af te leiden dat er is slechts ééen onderliggende sequentie die deze obervatie mogelijk maakt, en dat is $(X_1, X_2, X_3) = (\mathtt{1}, \mathtt{3}, \mathtt{2})$. Gebruikt nu de totale kans wet:

$$\begin{aligned}
P((Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{b}, \mathtt{c})) \quad &= \quad \sum_{x_1, x_2, x_3 \in \mathcal{S}} P((Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{b}, \mathtt{c}) \,|\, (X_1, X_2, X_3) = (x_1, x_2, x_3)) \\
&\qquad \times P((X_1, X_2, X_3) = (x_1, x_2, x_3)) \\
&= \quad P((Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{b}, \mathtt{c}) \,|\, (X_1, X_2, X_3) = (\mathtt{1}, \mathtt{3}, \mathtt{2})) \\
&\qquad \times P((X_1, X_2, X_3) = (\mathtt{1}, \mathtt{3}, \mathtt{2})) \\
&= \quad P(Y_1 = \mathtt{a} \,|\, X_1 = \mathtt{1}) P(Y_2 = \mathtt{b} \,|\, X_2 = \mathtt{3}) P(Y_3 = \mathtt{c} \,|\, X_3 = \mathtt{2}) \\
&\quad \times P(X_3 = \mathtt{2} \,|\, X_2 = \mathtt{3}) P(X_2 = \mathtt{3} \,|\, X_1 = \mathtt{1}) P(X_1 = \mathtt{1}) \\
&= \quad \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} \times 1 \times \tfrac{1}{2} \times 1 \quad = \quad \tfrac{1}{16}.
\end{aligned}$$

Vraag 2b
Concludeer op basis van de gegeven parameters dat slechts drie onderliggende sequenties deze obervatie kunnen genereren: $(X_1, X_2, X_3) = (\mathtt{1}, \mathtt{2}, \mathtt{1})$, $(X_1, X_2, X_3) = (\mathtt{1}, \mathtt{2}, \mathtt{2})$ en $(X_1, X_2, X_3) = (\mathtt{1}, \mathtt{3}, \mathtt{2})$. Welk van deze onderliggende sequenties maximalizeert de kans $P((X_1, X_2, X_3) \,|\, (Y_1, Y_2, Y_3))$. Bereken hiertoe voor elke van deze sequenties:

$$\begin{aligned}
&P((X_1, X_2, X_3) = (x_1, x_2, x_3) \,|\, (Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a})) \\
&\quad = \quad P((Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a}) \,|\, (X_1, X_2, X_3) = (x_1, x_2, x_3)) \\
&\qquad \times P((X_1, X_2, X_3) = (x_1, x_2, x_3)) / P((Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a}).
\end{aligned}$$

De laatste term, $P((Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a})$, kan hierbij genegeerd worden, daar ie in elk van deze kansen voorkomt. Als bij onder 2a:

$$\begin{aligned}
P((X_1, X_2, X_3) = (\mathtt{1}, \mathtt{2}, \mathtt{1}) \,|\, (Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a})) \quad &\propto \quad \tfrac{1}{32}, \\
P((X_1, X_2, X_3) = (\mathtt{1}, \mathtt{2}, \mathtt{2}) \,|\, (Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a})) \quad &\propto \quad \tfrac{1}{32}, \\
P((X_1, X_2, X_3) = (\mathtt{1}, \mathtt{3}, \mathtt{2}) \,|\, (Y_1, Y_2, Y_3) = (\mathtt{a}, \mathtt{c}, \mathtt{a})) \quad &\propto \quad \tfrac{1}{16}.
\end{aligned}$$

Dus, $(X_1, X_2, X_3) = (\mathtt{1}, \mathtt{3}, \mathtt{2})$ is de meest aannemelijke onderliggende sequentie.

Vraag 2c
Net als bij vraag 1b, bepaal de stationaire verdeling, nu van het onderliggende proces. Kies de vergelijking uit $\boldsymbol{\varphi}^\top \mathbf{P} = \boldsymbol{\varphi}$ slim, dan volgt direct dat: $\frac{1}{2}\varphi_2 = \varphi_1$ en $\frac{1}{2}\varphi_1 = \varphi_3$. Tesamen met $\varphi_1 + \varphi_2 + \varphi_3 = 1$ levert dat: $(\varphi_1, \varphi_2, \varphi_3) = \frac{1}{7}(2, 4, 1)$. In de limiet, e.g.:

$$\begin{aligned} P(Y_t = \mathtt{a}) &= P(Y_t = \mathtt{a} \,|\, X_t = \mathtt{1}) P(X_t = \mathtt{1}) + P(Y_t = \mathtt{a} \,|\, X_t = \mathtt{2}) P(X_t = \mathtt{2}) \\ &\quad + P(Y_t = \mathtt{a} \,|\, X_t = \mathtt{3}) P(X_t = \mathtt{3}). \end{aligned}$$

Dus: $P(Y_t = \mathtt{a}) = \frac{2}{7} \cdot \frac{1}{2} + \frac{4}{7} \cdot \frac{1}{2}$, $P(Y_t = \mathtt{b}) = \frac{2}{7} \cdot \frac{1}{2} + \frac{1}{7} \cdot \frac{1}{2}$ en $P(Y_t = \mathtt{c}) = \frac{4}{7} \cdot \frac{1}{2} + \frac{1}{7} \cdot \frac{1}{2}$. De meest voorkomende observatie in een oneindig lange sequentie afkomstig model is derhalve de $\mathtt{a}$.

Antwoord op vraag 3
Vraag 3a
Een partiële correlatie is een geschaald element van de inverse covariantie matrix. Die inverse is:

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 0.4 & -0.6 & -0.4 \\ -0.6 & 1.9 & 1.1 \\ -0.4 & 1.1 & 0.9 \end{pmatrix}$$

De schaling beinvloedt het feit dat een element nul is niet. Uit de inverse, die dus 1-op-1 relateert aan de partiële correlaties, kunnen we dus concluderen dat geen partiële correlatie gelijk aan nul is. (Ofwel, er zijn geen conditionele onafhankelijkheden.)

Vraag 3b
Gebruik de formule van het formuleblad:

$$Y_1 \,|Y_2 \sim \mathcal{N}[\mu_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(Y_2 - \mu_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}].$$

Invullen geeft: $Y_1 \,|Y_2 \sim \mathcal{N}(\frac{1}{2}Y_2, 4\frac{1}{2})$.

Vraag 3c
Gebruik de rekenregels voor co- en variantie:

$$\begin{aligned} \mathrm{Var}(Y_4) &= \mathrm{Cov}(Y_4, Y_4) \\ &= \mathrm{Cov}(1 + 3Y_1 - 2Y_3 + \varepsilon_4, 1 + 3Y_1 - 2Y_3 + \varepsilon_4) \\ &= \mathrm{Cov}(1, 1) + \mathrm{Cov}(3Y_1, 3Y_1) + \mathrm{Cov}(-2Y_3, -2Y_3) + \mathrm{Cov}(\varepsilon_4, \varepsilon_4) + 2\mathrm{Cov}(1, 3Y_1) \\ &\quad + 2\mathrm{Cov}(1, -2Y_3) + 2\mathrm{Cov}(1, \varepsilon_4) + 2\mathrm{Cov}(3Y_1, -2Y_3) + 2\mathrm{Cov}(3Y_1, \varepsilon_4) + 2\mathrm{Cov}(-2Y_3, \varepsilon_4) \\ &= 9\mathrm{Var}(Y_1) + 4\mathrm{Var}(Y_3, Y_3) + \mathrm{Var}(\varepsilon_4) \\ &\quad - 12\mathrm{Cov}(Y_1, Y_3) + 6\mathrm{Cov}(Y_1, \varepsilon_4) - 4\mathrm{Cov}(Y_3, \varepsilon_4) \\ &= 9\mathrm{Var}(Y_1) + 4\mathrm{Var}(Y_3, Y_3) + \mathrm{Var}(\varepsilon_4) - 12\mathrm{Cov}(Y_1, Y_3) \\ &= 9 \cdot 5 + 4 \cdot 4 + 1 - 12 \cdot 1 \;=\; 50. \end{aligned}$$

Vraag 3d
Gebruik wederom de formule van het formuleblad. Deze geeft $\mathbb{E}(Y_3 \,|\, Y_1, Y_2) = \mu_3 + \boldsymbol{\Sigma}_{3,12}\boldsymbol{\Sigma}_{12,12}^{-1}(\mathbf{Y}_{12} - \boldsymbol{\mu}_{12})$. Invullen geeft: $\mathbb{E}(Y_3 \,|\, Y_1, Y_2) = \frac{1}{9}Y_1 + \frac{4}{9}Y_2$. Dus elke expressie niveau combinatie voor de eerste twee genen in de verzameling $\{(Y_1, Y_2) \in \mathbb{R}^2 \text{ z.d.d. } \frac{1}{9}Y_1 + \frac{4}{9}Y_2 = 1\}$ volstaat.