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

Vraag 1 *(Markov modellen)*
Dagelijks wordt de fijnstof-concentratie in Amsterdam gemeten. Deze concentratie, geclassificeerd als laag, gemiddeld of hoog, kan m.b.v. een 1^{ste} orde Markov keten gemodelleerd worden. Op reguliere dagen staat er een zachte danwel stevige bries, maar af en toe is het windstil danwel stormachtig. Bij een zachte bries daalt de concentratie niet en stijgt (met kans α) één niveau. Staat er daarentegen een stevige bries, dan stijgt de concentratie niet maar daalt (met kans β) één niveau. Eens in de honderd dagen trekt een storm over de stad en veegt de lucht volledig schoon van fijnstof. Evenzo is het eens in de honderd dagen windstil en stijgt de fijnstof-concentratie naar het hoogste niveau. Natuurlijk kan (bijv. bij aanhoudendend stormachtig weer) de fijnstof-concentratie gelijk blijven.

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 :

$$\mathbf{P} = \begin{pmatrix} 1-\gamma-\frac{1}{5} & \frac{1}{5} & \gamma \\ \gamma & 1-\frac{1}{5}-\delta & \frac{1}{5}+\delta-\gamma \\ \frac{1}{5} & \delta & 1-\frac{1}{5}-\delta \end{pmatrix},$$

waarbij rijen- en kolommen-volgorde correspondeert met fijnstof-concentraties `laag`, `gemiddeld` en `hoog`. 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 (veronderstel hierbij enkel voor deze deelvraag dat $\gamma = \frac{1}{5}$ and $\delta = \frac{2}{5}$).

Vraag 1c)
Neem aan dat het antwoord op 1b een uniforme verdeling is. Is het 1^{ste} orde Markov proces dan reversibel? Motiveer je antwoord. Belicht tevens of je reversibiliteit noodzakelijk vindt voor een zinvolle beschrijving van het voornoemde fijnstof-verloop.

Vraag 1d)
Over een periode van tien dagen is de volgende fijnstof-concentratie gemeten: `laag`, `laag`, `gemiddeld`, `laag`, `gemiddeld`, `laag`, `hoog`, `gemiddeld`, `gemiddeld`, `laag`. Gebruik de maximum likelihood methode om de parameters α en β op basis van dit verloop te schatten. Veronderstel hierbij de kans om op dag één een lage concentratie aan te treffen gelijk aan 1.

Vraag 2 *(Hidden Markov model)*
Het menselijk DNA van een baarmoederhalskankercel bevat viraal DNA maar kent ook veel mutaties. Het DNA van zo'n cel is opgedeeld in opeenvolgende stukken en van elk stuk is de overeenkomst met het humane referentie-genoom bepaald. De mate van overeenkomst duidt op de DNA origine. De geobserveerde overeenkomst-sequentie $\{Y_t\}_{t=1}^T$ kan worden verklaard vanuit de origine van het corresponderende stuk DNA m.b.v. een hidden Markov model (HMM), waarbij de verschillende origines de toestanden van de onderliggende 1^{ste} order Markov keten representeren. De parameters $(\boldsymbol{\pi}, \mathbf{P}, \mathbf{B})$ (resp. de startverdeling, de transitie- en emissie-matrix) van het HMM worden gegeven door $\boldsymbol{\pi} = (1, 0)^\top$ (de kans op `humaan` danwel `viraal`, resp.),

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}, \qquad \text{en} \qquad \mathbf{B} = \begin{pmatrix} 0.0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0.0 \end{pmatrix}$$

De rijen van de matrices $\mathbf{P}$ en $\mathbf{B}$ representeren de origines (`humaan`/`viraal`, resp.) en de kolommen van $\mathbf{B}$ corresponderen met een `lage`/`gemiddelde`/`hoge` (resp.) overeenkomst.

Vraag 2a
Bereken $P((Y_1, Y_2, Y_3) = (\texttt{hoog}, \texttt{middel}, \texttt{laag}))$.

Vraag 2b
Wat is de meest aannemelijk toestandssequentie die ten grondslag ligt aan de geobserveerde overeenkomst-sequentie $(Y_1, Y_2, Y_3) = (\texttt{hoog}, \texttt{laag}, \texttt{middel})$?

Vraag 3
Beschouw een pathway van 3 genen. De expressie-niveau's van de drie genen worden gerepresenteerd door de random variabelen Y_1, Y_2 en Y_3.

Vraag 3a)
Zij $Y_2 = Y_1 + \varepsilon_2$, $Y_3 = Y_1 - Y_2 + \varepsilon_3$ met Y_1, ε_2 en ε_3 onderling onafhankelijk en elk standaard normaal verdeeld. Bereken de correlatie tussen Y_1 en Y_3.

Vraag 3b)
Neem nu aan dat $(Y_1, Y_2, Y_3)^\top \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ met $\boldsymbol{\mu} = (0, 0, 0)^\top$ en:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 4 & 1 & 3 \\ 1 & 3 & 2 \\ 3 & 2 & 3 \end{pmatrix},$$

Reken de partiële correlatie tussen Y_1 en Y_2 gegeven Y_3 uit.

Vraag 3c)
Neem aan dat de bij 3b gevraagde partiele correlatie gelijk aan nul is. Data van de expressie-niveau's zijn beschikbaar. Een regressie-model, dat de expressie-niveaus's van Y_2 in termen van Y_3 verklaart, wordt m.b.v. deze data gefit, nl.: $Y_2 = \beta_2 Y_3 + \varepsilon$. De `R`-output voor deze regressie fit is:

```
     Estimate  Std.  Error  t value  Pr(>|t|)
Y3   0.96649       0.08687    11.13    <2e-16
```

Kan uit bovenstaande `R`-output geconcludeerd worden dat Y_2 conditioneel afhankelijk is van Y_3?

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:

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

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:

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

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:

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

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 $\mathcal{S}$ bestaat uit toestanden $\{\texttt{laag}, \texttt{middel}, \texttt{hoog}\}$. De kans op storm danwel windstil is $\frac{1}{100}$, dus de kans op een reguliere dag is $\frac{98}{100}$. Oftewel: $P(\texttt{stevige bries}) + P(\texttt{zachte bries}) = \frac{98}{100}$. Als $P(\texttt{stevige bries}) = p_{sb}$ dan $P(\texttt{zachte bries}) = \frac{98}{100} - p_{bz}$. Merk op: de interpretatie $P(\texttt{stevige bries}) = \frac{49}{100} = P(\texttt{zachte bries})$ wordt goed gerekend. Dan, mbv de 'total probability law':

$$\begin{aligned} P(X_{t+1} = \texttt{laag} \,|\, X_t = \texttt{laag}) &= P(X_{t+1} = \texttt{laag}, \texttt{windstil} \,|\, X_t = \texttt{laag}) \\ &\quad + P(X_{t+1} = \texttt{laag}, \texttt{zachte bries} \,|\, X_t = \texttt{laag}) \\ &\quad + P(X_{t+1} = \texttt{laag}, \texttt{stevige bries} \,|\, X_t = \texttt{laag}) \\ &\quad + P(X_{t+1} = \texttt{laag}, \texttt{storm} \,|\, X_t = \texttt{laag}) \\ &= 0 \times \tfrac{1}{100} + (1-\alpha) \times (\tfrac{98}{100} - p_{sb}) + 1 \times p_{sb} + 1 \times \tfrac{1}{100} \\ &= \tfrac{98}{100}(1-\alpha) + \alpha p_{sb} + \tfrac{1}{100}. \end{aligned}$$

Evenzo, of gelijke wijze:

$$\begin{aligned} P(X_{t+1} = \texttt{middel} \,|\, X_t = \texttt{laag}) &= 0 \times \tfrac{1}{100} + \alpha \times (\tfrac{98}{100} - p_{sb}) + 0 \times p_{sb} + 0 \times \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{hoog} \,|\, X_t = \texttt{laag}) &= 1 \times \tfrac{1}{100} + 0 \times (\tfrac{98}{100} - p_{sb}) + 0 \times p_{sb} + 0 \times \tfrac{1}{100} = \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{laag} \,|\, X_t = \texttt{middel}) &= 0 \times \tfrac{1}{100} + 0 \times (\tfrac{98}{100} - p_{sb}) + \beta \times p_{sb} + 1 \times \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{middel} \,|\, X_t = \texttt{middel}) &= 0 \times \tfrac{1}{100} + (1-\alpha) \times (\tfrac{98}{100} - p_{sb}) + (1-\beta) \times p_{sb} + 0 \times \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{hoog} \,|\, X_t = \texttt{middel}) &= 1 \times \tfrac{1}{100} + \alpha \times (\tfrac{98}{100} - p_{sb}) + 0 \times p_{sb} + 0 \times \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{laag} \,|\, X_t = \texttt{hoog}) &= 0 \times \tfrac{1}{100} + 0 \times (\tfrac{98}{100} - p_{sb}) + 0 \times p_{sb} + 1 \times \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{middel} \,|\, X_t = \texttt{hoog}) &= 0 \times \tfrac{1}{100} + 0 \times (\tfrac{98}{100} - p_{sb}) + \beta \times p_{sb} + 0 \times \tfrac{1}{100}, \\ P(X_{t+1} = \texttt{hoog} \,|\, X_t = \texttt{hoog}) &= 1 \times \tfrac{1}{100} + 1 \times (\tfrac{98}{100} - p_{sb}) + (1-\beta) \times p_{sb} + 0 \times \tfrac{1}{100}, \end{aligned}$$

Samengevoegd:

$$\mathbf{P} = \begin{pmatrix} \frac{98}{100}(1-\alpha) + \alpha p_{sb} + \frac{1}{100} & \alpha(\frac{98}{100} - p_{sb}) & \frac{1}{100} \\ \beta p_{sb} + \frac{1}{100} & (1-\alpha)(\frac{98}{100} - p_{sb}) + (1-\beta)p_{sb} & \frac{1}{100} + \alpha(\frac{98}{100} - p_{sb}) \\ \frac{1}{100} & \beta p_{sb} & \frac{1}{100} + (\frac{98}{100} - p_{sb}) + (1-\beta)p_{sb} \end{pmatrix}.$$

De rijen sommeren inderdaad tot een. Verder, uit het feit dat elk element van $\mathbf{P}$ in het interval $[0,1]$ moet liggen, volgen de parameter-restricties. Direct volgt $0 \leq \alpha, \beta$. Specifieker, middels expliciet uitrekenen geeft: $\alpha \leq \min\{1, \frac{99}{100}(\frac{98}{100} - p_{sb})^{-1}\}$ en $\beta \leq \min\{1, \frac{99}{100} p_{sb}^{-1}\}$. Includeer ook het *state diagram.*

Vraag 1b
De transitie matrix wordt dus:

$$\mathbf{P} = \begin{pmatrix} \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \end{pmatrix},$$

Ja, heeft stationaire verdeling: irreducibel en aperiodiek. Gebruik dan: $\boldsymbol{\varphi}^\top \mathbf{P} = \boldsymbol{\varphi}^\top$ en $\varphi_{\texttt{L}} + \varphi_{\texttt{M}} + \varphi_{\texttt{H}} = 1$. Dit geeft het volgende stelsel van vergelijkingen:

$$\begin{aligned} 3\varphi_{\texttt{L}} + \varphi_{\texttt{M}} + \varphi_{\texttt{H}} &= 5\varphi_{\texttt{L}}, \\ \varphi_{\texttt{L}} + 2\varphi_{\texttt{M}} + 2\varphi_{\texttt{H}} &= 5\varphi_{\texttt{M}}, \\ \varphi_{\texttt{L}} + 2\varphi_{\texttt{M}} + 2\varphi_{\texttt{H}} &= 5\varphi_{\texttt{H}}. \end{aligned}$$

Laat de laatste formule vervallen en herschrijf de eerste twee:

$$\begin{aligned}(\varphi_{\mathtt{M}} - \varphi_{\mathtt{L}}) + (\varphi_{\mathtt{H}} - \varphi_{\mathtt{L}}) &= 0 \\ (\varphi_{\mathtt{L}} - \varphi_{\mathtt{M}}) + 2(\varphi_{\mathtt{H}} - \varphi_{\mathtt{M}}) &= 0, \\ \varphi_{\mathtt{L}} + \varphi_{\mathtt{M}} + \varphi_{\mathtt{H}} &= 1.\end{aligned}$$

De laatste vergelijking levert $\varphi_{\mathtt{H}} = 1 - \varphi_{\mathtt{L}} - \varphi_{\mathtt{M}}$. Substitueer dit in de eerste twee vergelijkingen:

$$\begin{aligned}(\varphi_{\mathtt{M}} - \varphi_{\mathtt{L}}) + (1 - 2\varphi_{\mathtt{L}} - \varphi_{\mathtt{M}}) &= 1 - 3\varphi_{\mathtt{L}} &= 0 \\ (\varphi_{\mathtt{L}} - \varphi_{\mathtt{M}}) + 2(1 - \varphi_{\mathtt{L}} - 2\varphi_{\mathtt{M}}) &= 2 - \varphi_{\mathtt{L}} - 5\varphi_{\mathtt{M}} &= 0.\end{aligned}$$

Dus: $(\varphi_{\mathtt{L}}, \varphi_{\mathtt{M}}, \varphi_{\mathtt{H}}) = (1, 1, 1)/3$. Deze kansen liggen in het interval $(0, 1)$ en sommeren tot een.

Merk op: dit had ook direct uit de symmetrie van $\mathbf{P}$ geconcludeerd kunnen worden: een irreducibele, aperiodieke Markov keten met een symmetrische transitie matrix heeft een uniforme stationaire verdeling.

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

$$\varphi_{\mathtt{L}}(\mathbf{P})_{\mathtt{L},\mathtt{M}} \;=\; \tfrac{1}{3}\tfrac{1}{5} \;\neq\; \tfrac{1}{3}\gamma \;=\; \varphi_s(\mathbf{P})_{\mathtt{M},\mathtt{L}}.$$

Het model is derhalve niet reversibel voor alle keuze vd parameters. Dit is ook nietnoodzakelijk daar de actuele staat van het weer en de fijnstof-concentratie direct geobserveerd kunnen worden.

Vraag 1d
De likelihood van deze sequentie wordt gegeven door: $P(X_1 = \mathtt{1}) \prod_{t=2}^{10} P(X_t \,|\, X_{t-1})$. Ofwel: $1 \times (1 - \gamma - \frac{1}{5}) \times \frac{1}{5} \times \gamma \times \frac{1}{5} \times \gamma \times \gamma \times \delta \times (1 - \frac{1}{5} - \delta) \times \gamma = \frac{1}{5}^2 \delta\gamma^4(1 - \frac{1}{5} - \gamma)(1 - \frac{1}{5} - \delta)$. Neem de logaritme, deze is proportioneel aan: $\log(\delta) + 4\log(\gamma) + \log(1 - \frac{1}{5} - \gamma) + \log(1 - \frac{1}{5} - \delta)$. Stel eerste orde afgeleiden (naar γ en δ) gelijk aan nul: $1/\delta - 1/(1 - \frac{1}{5} - \delta) = 0$ en $4/\gamma - 1/(1 - \frac{1}{5} - \gamma) = 0$. Oplossen levert: $\hat{\delta} = 2/5$ en $\hat{\gamma} = 16/25$.

Antwoord op vraag 2
Vraag 2a
Slechts twee onderliggende toestandssequenties kunnen de geobserveerde data verklaren: $\{X_1 = \mathtt{Humaan}, X_2 = \mathtt{Humaan}, X_3 = \mathtt{Viraal}\}$ en $\{X_1 = \mathtt{Humaan}, X_2 = \mathtt{Viraal}, X_3 = \mathtt{Viraal}\}$. Om de gevraagde kans uit te rekenen gebruik de total probability law:

$$\begin{aligned}
&P(Y_1 = \mathtt{H}, Y_2 = \mathtt{M}, Y_3 = \mathtt{L}) \\
&= P(Y_1 = \mathtt{H}, Y_2 = \mathtt{M}, Y_3 = \mathtt{L} \,|\, X_1 = \mathtt{Humaan}, X_2 = \mathtt{Humaan}, X_3 = \mathtt{Viraal}) \\
&\quad \times P(X_1 = \mathtt{Humaan}, X_2 = \mathtt{Humaan}, X_3 = \mathtt{Viraal}) \\
&\quad + P(Y_1 = \mathtt{H}, Y_2 = \mathtt{M}, Y_3 = \mathtt{L} \,|\, X_1 = \mathtt{Humaan}, X_2 = \mathtt{Viraal}, X_3 = \mathtt{Viraal}) \\
&\quad\; P(X_1 = \mathtt{Humaan}, X_2 = \mathtt{Viraal}, X_3 = \mathtt{Viraal}) \\
&= P(Y_1 = \mathtt{H} \,|\, X_1 = \mathtt{Humaan}) P(Y_2 = \mathtt{M} \,|\, X_2 = \mathtt{Humaan}) P(Y_3 = \mathtt{L} \,|\, X_3 = \mathtt{Viraal}) \\
&\quad \times P(X_1 = \mathtt{Humaan}) P(X_2 = \mathtt{Humaan} \,|\, X_1 = \mathtt{Humaan}) P(X_3 = \mathtt{Viraal} \,|\, X_2 = \mathtt{Humaan}) \\
&\quad + P(Y_1 = \mathtt{H} \,|\, X_1 = \mathtt{Humaan}) P(Y_2 = \mathtt{M} \,|\, X_2 = \mathtt{Viraal}) P(Y_3 = \mathtt{L} \,|\, X_3 = \mathtt{Viraal}) \\
&\quad \times P(X_1 = \mathtt{Humaan}) P(X_2 = \mathtt{Viraal} \,|\, X_1 = \mathtt{Humaan}) P(X_3 = \mathtt{Viraal} \,|\, X_2 = \mathtt{Viraal}) \\
&= (0.9 \times 0.1 \times 0.6) \times (1.0 \times 0.9 \times 0.1) + (0.9 \times 0.4 \times 0.6) \times (1.0 \times 0.1 \times 0.5) \\
&= 0.01566.
\end{aligned}$$

In het bovenstaande is de Markov eigenschap van de onderliggende keten gebruikt alsook de conditionele onafhankelijkheid van de observaties gegeven de onderliggende toestanden.

Vraag 2b
Gebruik de regel van Bayes danwel definitie van conditionele kans om het gevraagde te herschrijven:

$$\begin{aligned}
&\arg\max_{X_1,X_2,X_3} P(X_1, X_2, X_3 \,|\, Y_1 = \mathtt{H}, Y_2 = \mathtt{L}, Y_3 = \mathtt{M}) \\
&= \arg\max_{X_1,X_2,X_3} \frac{P(X_1, X_2, X_3, Y_1 = \mathtt{H}, Y_2 = \mathtt{L}, Y_3 = \mathtt{M})}{P(Y_1 = \mathtt{H}, Y_2 = \mathtt{L}, Y_3 = \mathtt{M})} \\
&= \arg\max_{X_1,X_2,X_3} P(X_1, X_2, X_3, Y_1 = \mathtt{H}, Y_2 = \mathtt{L}, Y_3 = \mathtt{M}) \\
&= \arg\max_{X_1,X_2,X_3} P(Y_1 = \mathtt{H}, Y_2 = \mathtt{L}, Y_3 = \mathtt{M} \,|\, X_1, X_2, X_3) P(X_1, X_2, X_3) \\
&= \arg\max_{X_1,X_2,X_3} P(Y_1 = \mathtt{H} \,|\, X_1) P(Y_2 = \mathtt{L} \,|\, X_2) P(Y_3 = \mathtt{M} \,|\, X_3) P(X_1) P(X_2 \,|\, X_1) P(X_3 \,|\, X_2).
\end{aligned}$$

Merk nu op dat wederom slechts twee toestandssequenties de observaties kunnen genereren: $\{X_1 = \mathtt{Humaan}, X_2 = \mathtt{Viraal}, X_3 = \mathtt{Humaan}\}$ en $\{X_1 = \mathtt{Humaan}, X_2 = \mathtt{Viraal}, X_3 = \mathtt{Viraal}\}$. Evalueer voor beide toestandssequenties de bovenstaande kans. De eerst levert: $(0.9 \times 0.6 \times 0.1) \times (1 \times 0.1 \times 0.5)$, terwijl de tweede $(0.9 \times 0.6 \times 0.4) \times (1 \times 0.1 \times 0.5)$ geeft. De tweede sequentie is dus meer aannemelijk.

Antwoord op vraag 3
Vraag 3a
De definitie van correlatie geeft:

$$\rho(Y_1, Y_3) \quad = \quad \frac{\mathrm{Cov}(Y_1, Y_3)}{\sqrt{\mathrm{Var}(Y_1)}\sqrt{\mathrm{Var}(Y_3)}}.$$

Reken de individuele termen uit:

$$\begin{aligned}
\mathrm{Var}(Y_1) &= 1 \\
\mathrm{Var}(Y_3) &= \mathrm{Cov}(Y_3, Y_3) \quad = \quad \mathrm{Cov}(Y_1 - Y_2 + \varepsilon_3, Y_1 - Y_2 + \varepsilon_3) \\
&= \mathrm{Cov}(Y_1 - Y_1 - \varepsilon_2 + \varepsilon_3, Y_1 - Y_1 - \varepsilon_2 + \varepsilon_3) \quad = \quad \mathrm{Cov}(-\varepsilon_2 + \varepsilon_3, -\varepsilon_2 + \varepsilon_3) \\
&= \mathrm{Cov}(Y_1 - Y_1 - \varepsilon_2 + \varepsilon_3, Y_1 - Y_1 - \varepsilon_2 + \varepsilon_3) \quad = \quad \mathrm{Cov}(-\varepsilon_2 + \varepsilon_3, -\varepsilon_2 + \varepsilon_3) \\
&= \mathrm{Cov}(-\varepsilon_2, -\varepsilon_2) + \mathrm{Cov}(-\varepsilon_2, \varepsilon_3) + \mathrm{Cov}(\varepsilon_3, -\varepsilon_2) + \mathrm{Cov}(\varepsilon_3, \varepsilon_3) \\
&= \mathrm{Var}(\varepsilon_2) + \mathrm{Var}(\varepsilon_3) \quad = \quad 2 \\
\mathrm{Cov}(Y_1, Y_3) &= \mathrm{Cov}(Y_1, Y_1 - Y_2 + \varepsilon_3) \\
&= \mathrm{Cov}(Y_1 + \varepsilon_3, Y_1 - Y_1 - \varepsilon_2 + \varepsilon_3) \quad = \quad \mathrm{Cov}(Y_1 + \varepsilon_3, -\varepsilon_2 + \varepsilon_3) \quad = \quad 0.
\end{aligned}$$

De gevraagde correlatie is dus 0.

Vraag 3b
Voor de partiële correlatie matrix, neem eerst de inverse:

$$\mathbf{\Sigma}^{-1} \quad = \quad \frac{1}{2} \begin{pmatrix} 5 & 3 & -7 \\ 3 & 3 & -5 \\ -7 & -5 & 11 \end{pmatrix}^{-1}.$$

Standardizeer deze matrix tot:

$$\begin{pmatrix} 1 & 3/\sqrt{15} & -7/\sqrt{55} \\ 3/\sqrt{15} & 1 & -5/\sqrt{33} \\ -7/\sqrt{55} & -5/\sqrt{33} & 1 \end{pmatrix}.$$

Rest nog de 'off-diagonal' elementen met -1 te vermenigvuldigen en de gevraagde partiële correlatie af te lezen: $\rho(Y_1, Y_2 \mid Y_3) = -3/\sqrt{15}$.

Vraag 3c

Gegeven: $\rho(Y_1, Y_2 \mid Y_3) = 0$. Daar partiele correlaties 1-1 gerelateerd zijn aan regressie-coefficienten weten we dat (bijv.) $\beta_2 = 0$ in $Y_1 = \beta_2 Y_2 + \beta_3 Y_3 + \varepsilon$, i.e. de variable Y_2 voegt niets toe aan het verklaren van de variatie in Y_1. Kortom, om nu vast te stellen dat Y_1 en Y_3 gegeven Y_2 onhankelijk zijn volstaat om vast te stellen dat $\beta_3 \neq$ in $Y_1 = \beta_3 Y_3 + \varepsilon$. Dit blijkt uit de regressie output.