Kansrekening (NWI-NB004B) 24 maart 2021

Nakijkmodel bij het tentamen

Gebruik van een (grafische) rekenmachine is wel toegestaan. Gebruik van het dictaat, een lesboek of aantekeningen is niet toegestaan.

Dit tentamen duurt twee uur.

Het tentamen bestaat uit 4 opgaven. In totaal zijn er 29 punten te verdienen.

Het cijfer dat je krijgt is 1 + 9 × aantal gehaalde punten / 29. Hierbij wordt de Python-inleveropdrachten-bonus opgeteld. Deze som wordt afgerond op halve punten. Uitzondering zijn cijfers tussen de 5 en de 6: deze cijfers worden afgerond op hele punten. Je slaagt voor dit vak als je een 6 of hoger haalt.

Opgave 1 (4 + 1 + 5 punten). Ik ga langs 10 verkeerslichten die onafhankelijk van elkaar opereren. Bij elk van deze verkeerslichten is de kans dat het licht op groen, oranje of rood staat respectievelijk 50%, 20%, of 30% (deze drie gebeurtenissen sluiten elkaar natuurlijk uit). Laat G het aantal van de 10 verkeerslichten dat op groen staat. Net zo, laat O het aantal van de 10 verkeerslichten dat op oranje staat.

(a) Bereken $\mathbb{P}[G = 5, O \leq 1]$.

(b) Beredeneer hoe G en O gecorreleerd zijn (positief-, negatief-, of ongecorreleerd) met daarbij een intuïtieve verklaring zonder berekeningen.

(c) Bereken de correlatiecoëfficient $\rho_{G,O}$.

Oplossing 1.

(a) Merk op dat $G \sim \text{Bin}(10, 0.5)$ en $O \sim \text{Bin}(10, 0.2)$ $\boxed{1}$ (dit punt is ook te verdienen door een zinnige "binomiale-pmf-kans" uit te rekenen), maar natuurlijk niet onafhankelijk van elkaar. Het is daarom handig om deze kans via conditioneren te berekenen:

$$\begin{aligned}\mathbb{P}[G = 5, O \leq 1] &= \mathbb{P}[O \leq 1 \mid G = 5]\mathbb{P}[G = 5] && \boxed{1}\\ &= (\mathbb{P}[O = 0 \mid G = 5] + \mathbb{P}[O = 1 \mid G = 5])\binom{10}{5}(0.5)^{10} && \boxed{1}\end{aligned}$$

Merk op dat $(O \mid G = 5)$ ook weer binomiaal verdeeld is, maar nu met nog maar $n = 5$ overgebleven verkeerslichten, en met succeskans $p = \frac{2}{5}$ (want de overgebleven verkeerslichten staan niet op groen), dus $(O \mid G = 5) \sim \text{Bin}(5, 0.4)$ $\boxed{1}$. We krijgen dus

$$\begin{aligned}\mathbb{P}[G = 5, O \leq 1] &= \left(\binom{5}{0}(0.6)^5 + \binom{5}{1}(0.4)(0.6)^4\right)\binom{10}{5}(0.5)^{10}\\ &\approx 0.082924\end{aligned}$$

Alternatief.

$$\begin{aligned}\mathbb{P}[G = 5, O \leq 1] &= \mathbb{P}[G = 5, O = 0, R = 5] + \mathbb{P}[G = 5, O = 1, R = 4] && \boxed{1}\\ &= \underbrace{\binom{10}{5}(0.5)^5(0.2)^0(0.3)^5}_{\boxed{1}} + \underbrace{\binom{10}{5}\binom{5}{1}(0.5)^5(0.2)^1(0.3)^4}_{\boxed{2}}\end{aligned}$$

Je gebruikt hier een generalisatie van de binomiale formule. De machten verklaren de kans dat de stoplichten achtereenvolgens op een gegeven aantal kleuren staan. De binomiaalcoëfficiënten tellen alle rangschikkingen. Er zijn $\binom{10}{5}$ manieren om er 5 op groen te zetten. Als er van de niet groene dan k op oranje staan, dan kan dat vervolgens (onafhankelijk) nog op $\binom{5}{k}$ manieren. We hebben dit hierboven gebruikt voor $k = 0$ en $k = 1$ (maximaal twee punten als je de $\binom{5}{k}$ vergeet).

Alternatief. Je dit ook oplossen door andersom te conditioneren:

$$\mathbb{P}[G = 5, O \leq 1] = \mathbb{P}[G = 5 \mid O = 0]\mathbb{P}[O = 0] + \mathbb{P}[G = 5 \mid O = 1]\mathbb{P}[O = 1].$$

(b) Meer verkeerslichten op groen is natuurlijk minder verkeerslichten op oranje (gemiddeld genomen), want elk stoplicht kan maar één kleur aannemen. We verwachten dus een negatieve correlatie. $\boxed{1}$

Een antwoord zonder uitleg is $\boxed{1}$ punt waard. Het goede antwoord met een uitleg die niet expliciet over de stoplichten gaat (denk aan: "als G groot, dan O klein, dus negatief") is $\boxed{0.5}$ punten waard.

(c) We moeten berekenen

$$\rho_{G,O} = \frac{\mathrm{cov}(G,O)}{\sqrt{\mathrm{var}(G)\,\mathrm{var}(O)}}, \qquad \boxed{0.5}$$

Omdat $G \sim \mathrm{Bin}(10, 0.5)$ en $O \sim \mathrm{Bin}(10, 0.2)$ verkrijgen we direct

$$\mathrm{var}(G) = 10 \cdot 0.5 \cdot 0.5 = 2.5, \qquad \mathrm{var}(O) = 10 \cdot 0.2 \cdot 0.8 = 1.6. \qquad \boxed{0.5}$$

De covariantie is lastiger omdat G en O natuurlijk afhankelijk zijn. Er zijn grofweg drie manieren om de covariantie te berekenen. Daarna is het een kwestie van getallen invullen om ρ te berekenen. Dat invullen levert geen verdere punten op, dus voor de berekening van cov zijn nog 4 punten te verdienen. Je kan maar punten uit één van de manieren krijgen.

Manier 1. (de slimme versie)

Er geldt dat

$$\mathrm{cov}(G,O) = \frac{1}{2}(\mathrm{var}(G+O) - \mathrm{var}(G) - \mathrm{var}(O)). \qquad \boxed{1}$$

Merk op dat $(G+O) \sim \mathrm{Bin}(10, 0.7)$ $\boxed{2}$. We verkrijgen dus dat

$$\begin{aligned} \mathrm{var}(G+O) &= 10 \cdot 0.7 \cdot (1 - 0.7) = 2.1 \\ \mathrm{cov}(G,O) &= \frac{1}{2}(\mathrm{var}(G+O) - \mathrm{var}(G) - \mathrm{var}(O)) \\ &= \frac{1}{2}(2.1 - 2.5 - 1.6) = -1 \qquad \boxed{1} \end{aligned}$$

Manier 2. (de brute kracht versie)

Er geldt dat

$$\mathrm{cov}(G,O) = \mathbb{E}(GO) - \mathbb{E}(G)\mathbb{E}(O) \qquad \boxed{0.5}$$

waarbij

$$\mathbb{E}(G) = 10 \cdot 0.5 = 5, \qquad \mathbb{E}(O) = 10 \cdot 0.2 = 2. \qquad \boxed{0.5}$$

De berekening van $\mathbb{E}(GO)$ kan door middel van opsplitsen naar de individuele lichten: $G = \sum_{i=1}^{10} G_i$ en $O = \sum_{j=1}^{10} O_j$, waarbij $G_i \sim \mathrm{Bernoulli}(0.5)$ en

$O_j \sim$ Bernoulli(0.2) (indicatie per verkeerslicht of die op een betreffende kleur staat).

$$\begin{aligned}\mathbb{E}(GO) &= \mathbb{E}((\sum G_i)(\sum O_j)) \\ &= \sum_i \mathbb{E}(G_iO_i) + \sum_{i\neq j} \mathbb{E}(G_iO_j) \qquad \boxed{1}\end{aligned}$$

gebruik dat verschillende verkeerslichten onafhankelijk opereren

$$\mathbb{E}(GO) = \sum_i \mathbb{E}(G_iO_i) + \sum_{i\neq j} \mathbb{E}(G_i)\mathbb{E}(O_j) \qquad \boxed{0.5}$$

gebruik dat de 10 verkeerslichten identiek opereren

$$\mathbb{E}(GO) = 10\mathbb{E}(G_1O_1) + 90\mathbb{E}(G_1)\mathbb{E}(O_1) \qquad \boxed{0.5}$$

een verkeerslicht kan maar op één kleur tegelijk staan, dus $\mathbb{E}(G_1O_1) = \mathbb{P}[G_1 = 1, O_1 = 1] = 0$, dus

$$\mathbb{E}(GO) = 90\mathbb{E}(G_1)\mathbb{E}(O_i) = 90 \cdot 0.5 \cdot 0.2 = 9 \qquad \boxed{1}$$

$$\text{cov}(G, O) = \mathbb{E}(GO) - \mathbb{E}(G)\mathbb{E}(O) = 9 - 5 \cdot 2 = -1$$

$$\rho_{G,O} = \frac{\text{cov}(G, O)}{\sqrt{\text{var}(G)\,\text{var}(O)}} = \frac{-1}{\sqrt{2.5 \cdot 1.6}} = -\frac{1}{2}$$

Manier 3. (rekenmachine versie) Een aantal studenten heeft gebruik gemaakt van een sommatie functie van een grafische rekenmachine (GR). Deze manier van oplossen was eigenlijk niet de bedoeling. Ik heb vooraf echter niet aangegeven welke functionaliteit van een GR wel of niet toegestaan is, en daarom moet ik het toch goed keuren. Dit gaat ook via de formule

$$\begin{aligned}\text{cov}(G, O) &= \mathbb{E}(GO) - \mathbb{E}(G)\mathbb{E}(O) && \boxed{0.5} \\ \mathbb{E}(GO) &= \frac{1}{2}(\mathbb{E}((G + O)^2) - \mathbb{E}(G^2) - \mathbb{E}(O^2)) && \boxed{1} \\ \mathbb{E}(O) &= \sum_{k=0}^{10} k\binom{10}{k}0.2^k 0.8^{10-k} = \text{GR} = 2 && \boxed{0.5} \\ \mathbb{E}(O^2) &= \sum_{k=0}^{10} k^2\binom{10}{k}0.2^k 0.8^{10-k} = \text{GR} = 5.6 && \boxed{0.5}\end{aligned}$$

Net zo: $\mathbb{E}(G) = 5$, en $\mathbb{E}(G^2) = 27.5$

$$\begin{aligned}\mathbb{E}((G + O)^2) &= \sum_{k=0}^{10} k^2\mathbb{P}[G + O = k, R = 10 - k] && \boxed{1} \\ &= \sum_{k=0}^{10} k^2\binom{10}{k}0.7^k 0.3^{10-k} = \text{GR} = 51.1 && \boxed{0.5}\end{aligned}$$

De rest is invullen. Een ander alternatief met de GR is via de formule van manier 1.

Opgave 2 (6 punten). We beschouwen onafhankelijke continue stochasten $X_1, X_2, \ldots$. Neem voor $j = 1, 2, \ldots$ aan dat $X_j \geq 1$ en dat voor $t \geq 1$ geldt dat $\mathbb{P}[X_j \geq t] = t^{-3}$. Bereken de met behulp van de normale benadering $\mathbb{P}[X_1 + X_2 + \ldots + X_{100} \geq 160]$.

Oplossing 2. Laat Y de normale benadering van $S = X_1 + X_2 + \ldots + X_{100}$. Gevraagd is dus

$$\begin{aligned}
\mathbb{P}[Y \geq 160] &\approx \texttt{normcdf(160, inf, } \mathbb{E}(S)\texttt{, } \sqrt{\mathrm{var}(S)}\texttt{)} && \boxed{1} \\
\mathbb{E}(S) &= 100\mathbb{E}(X_1) && \boxed{0.5} \\
\mathrm{var}(S) &= 100\,\mathrm{var}(X_1) && \boxed{0.5}
\end{aligned}$$

Voordat we de $\mathbb{E}$ en σ van een X_j kunnen berekenen, zullen we eerst de cdf en de pdf van een X_j moeten bepalen. Als $x \geq 1$, dan is

$$\begin{aligned}
F(x) &= \mathbb{P}[X_j \leq x] = 1 - x^{-3} && \boxed{1} \\
f(x) &= F'(x) = 3x^{-4} && \boxed{1}
\end{aligned}$$

Hiermee berekenen de momenten van X_1 (onderstaande punten alleen als de juiste pdf wordt gebruikt, dus deze punten kan je alleen halen als de twee punten direct hierboven gelukt zijn).

$$\begin{aligned}
\mathbb{E}(X_1) &= \int_1^\infty x \cdot 3x^{-4}\,\mathrm{d}x = \left[-\frac{3}{2}x^{-2}\right]_1^\infty = \frac{3}{2} && \boxed{1} \\
\mathbb{E}(X_1^2) &= \int_1^\infty x^2 \cdot 3x^{-4}\,\mathrm{d}x = \left[-3x^{-1}\right]_1^\infty = 3 \\
\mathrm{var}(X_1) &= \mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2 = \frac{3}{4} && \boxed{1}
\end{aligned}$$

Nu nog alles invullen

$$\begin{aligned}
\mathbb{E}(S) &= 100\mathbb{E}(X_1) = 150 \\
\mathrm{var}(S) &= 100\,\mathrm{var}(X_1) = 75
\end{aligned}$$

Nu kunnen we het antwoord berekenen

$$\mathbb{P}[S \geq 160] \approx \texttt{normcdf(160, inf, 150, sqrt(75))} \approx 0.124$$

Alternatief voor de laatste stap voor als je geen rekenmachine hebt. $Z = (S - 150)/\sqrt{75}$ is bij benadering standaardnormaal verdeeld.

$$\begin{aligned}
\mathbb{P}[S \geq 160] &\approx \mathbb{P}[Z \geq (160 - 150)/\sqrt{75}] \\
&= 1 - \Phi\left(\tfrac{10}{\sqrt{75}}\right) = 1 - \Phi\left(\tfrac{2}{3}\sqrt{3}\right)
\end{aligned}$$

Als je de continuïteitscorrectie toepast $\boxed{-1}$.

Opgave 3 (5 punten). Gooi een eerste keer met een dobbelsteen. Gooi vervolgens herhaald met de dobbelsteen totdat je een aantal gooit dat kleiner of gelijk is aan de eerste worp. Bereken de verwachtingswaarde van het totaal aantal worpen dat je doet (inclusief de eerste worp).

Oplossing 3. Zij X de uitkomst van de eerste worp. Zij Y het aantal worpen dat vervolgens nodig is om $\leq X$ te gooien $\boxed{1}$. Gevraagd is om $\mathbb{E}(1+Y)$ te berekenen. Merk op dat Y geometrisch is verdeeld $\boxed{1}$, met een parameter $p = \frac{X}{6}$ die afhangt van X (je krijgt de eerste 2 punten ook als je de stochasten niet netjes definieert maar de berekeningen wel goed doet). Er geldt dus dat $\mathbb{E}(Y) = \frac{1}{p} = \frac{6}{X}$ $\boxed{1}$. Omdat de parameter p afhangt van X kunnen we de gevraagde verwachtingswaarde bepalen door te conditioneren naar X.

$$\begin{aligned}
\mathbb{E}(1+Y) &= 1 + \sum_{k=1}^{6} \mathbb{E}(Y \mid X = k)\mathbb{P}[X = k] && \boxed{2} \\
&= 1 + \sum_{k=1}^{6} \frac{6}{k} \cdot \frac{1}{6} \\
&= 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} \\
&= \frac{69}{20} = 3.45
\end{aligned}$$

Als je $\mathbb{E}(Y) = 2.45$ als antwoord geeft dan krijg je ook alle punten. Het is dus niet erg als je allereerste worp vergeet mee te rekenen.

Alternatief (voor als je de verwachtingswaarde van de geometrische verdeling niet kent). Het eerste punt is, net als boven, voor de definitie van Y.

$$\begin{aligned}
\mathbb{E}(1+Y) &= 1 + \frac{1}{6} \sum_{k=1}^{6} \mathbb{E}(Y \mid X = k) && \boxed{2} \\
&= 1 + \frac{1}{6} \sum_{k=1}^{6} \sum_{j \geq 0} \mathbb{P}[Y > j \mid X = k] && \boxed{1}
\end{aligned}$$

$\mathbb{P}[Y > j \mid X = k] = \mathbb{P}[j \text{ dobbelstenen} > k] = (\frac{6-k}{6})^j$

$$\begin{aligned}
&= 1 + \frac{1}{6} \sum_{k=1}^{6} \sum_{j \geq 0} \left(\frac{6-k}{6}\right)^j \\
&= 1 + \frac{1}{6} \sum_{k=1}^{6} \frac{1}{1 - (\frac{6-k}{6})} \\
&= 1 + \frac{1}{6} \sum_{k=1}^{6} \frac{6}{k} && \boxed{1}
\end{aligned}$$

Opgave 4 (4 + 4 punten). Stel de stochasten X_1 en X_2 zijn onafhankelijk en allebei exponentieel verdeeld met parameter a voor een zekere $a > 0$. Dus $f_X(t) = ae^{-at}$ is een dichtheidsfunctie voor $t \geq 0$ voor zowel X_1 als X_2. Zij $Y = \max(X_1, X_2)$, en zij $Z = \min(X_1, X_2)$.

(a) Bepaal een dichtheidsfunctie voor Y.

(b) Bewijs of weerleg: Y is onafhankelijk van Z.

Oplossing 4.

(a) De verdeling van de X_k is $F_X(t) = 1 - e^{-at}$ $\boxed{1}$ (als je dit niet uit het hoofd weet kan je het ook via een integraal berekenen). Voor de dichtheid f_Y van $Y = \max(X_1, X_2)$ bepalen we eerst de verdeling F_Y. Voor $t \geq 0$ hebben we

$$\begin{aligned} F_Y(t) &= \mathbb{P}[Y \leq t] \\ &= \mathbb{P}[\max(X_1, X_2) \leq t] && \boxed{0.5} \\ &= \mathbb{P}[X_1 \leq t \text{ en } X_2 \leq t] && \boxed{1} \end{aligned}$$

gebruik dat $X_1 \perp\!\!\!\perp X_2$

$$\begin{aligned} &= \mathbb{P}[X_1 \leq t]\mathbb{P}[X_2 \leq t] && \boxed{0.5} \\ &= F_X(t)^2 \end{aligned}$$

$$\begin{aligned} f_Y(t) &= F_Y(t)' \\ &= (F_X(t)^2)' \\ &= 2f_X(t)F_X(t) \\ &= 2ae^{-at}(1 - e^{-at}) && \boxed{1} \end{aligned}$$

(b) Het "goede" antwoord inclusief een intuïtieve verklaring leveren geen punten op. Met intuïtieve verklaring bedoelen we iets als: "het minimum geeft informatie over het maximum dus ze zijn afhankelijk". Je kan de intuïtieve idee trouwens wel gebruiken in een bewijs als je het concreter maak, zie het alternatief onderaan.

We laten zien dat ze afhankelijk zijn door s en t te geven zodat

$$\mathbb{P}[Y \leq s \text{ en } Z \leq t] \neq \mathbb{P}[Y \leq s]\mathbb{P}[Z \leq t]. \qquad \boxed{1}$$

Voor een tegenvoorbeeld volstaat het om dit voor specifieke s en t aan te tonen. Wij doen dat hier voor $s = t = 37$.

Merk op dat $Z \leq Y$ $\boxed{1}$. Dus $[Y \leq 37] \subset [Z \leq 37]$. Hieruit volgt dat

$$\begin{aligned} \mathbb{P}[Y \leq 37 \text{ en } Z \leq 37] &= \mathbb{P}[Y \leq 37] && \boxed{1} \\ &\neq \mathbb{P}[Y \leq 37]\mathbb{P}[Z \leq 37] && \boxed{1} \end{aligned}$$

Dus $Y \not\perp\!\!\!\perp Z$.

In het bovenstaande bluf je natuurlijk enigszins. De volgende argumenten zijn niet noodzakelijk voor punten.

$$\mathbb{P}[Y \leq 37] = F_X(37)^2 = (1 - e^{-37a})^2 > 0,$$

en

$$\begin{aligned}
\mathbb{P}[Z \leq 37] &= 1 - \mathbb{P}[Z > 37] \\
&= 1 - \mathbb{P}[X_1 > 37 \text{ en } X_2 > 37] \\
&= 1 - \mathbb{P}[X_1 > 37]\mathbb{P}[X_2 > 37] \\
&= 1 - (1 - F_X(37))^2 \\
&= 1 - e^{-2 \cdot 37a} \\
&< 1
\end{aligned}$$

Alternatief. Er geldt dat $Z \leq Y$ $\boxed{1}$. Dus bepaalde waarden van Z zullen bepaalde waarden van Y uitsluiten $\boxed{1}$. We maken dit concreet. Als bijvoorbeeld $Z > t$ voor een $t > 0$, dan is het uitgesloten dat $Y \leq t$ $\boxed{1}$, terwijl zonder informatie over Z het niet is uitgesloten dat $Y \leq t$ $\boxed{1}$.

Ditzelfde argument in formules:

$$\mathbb{P}[Y \leq t \mid Z > t] = 0 \neq \mathbb{P}[Y \leq t].$$