Kansrekening (NWI-NB004B) 31 maart 2020

Beoordelingsmodel Tentamen

In Brightspace kan je nu zien of je het tentamen wel of niet gehaald hebt (er is geen cijfer). Het doel van dit document, is om je de kans te geven om deze beoordeling te controleren.

Inzage. Als je overtuigd bent dat je de verkeerde beoordeling hebt gekregen, dan kan je dit voorleggen aan de docent. Dit doe je door je eigen werk zorgvuldig te beoordelen met dit document. Je communiceert vervolgens met de docent het aantal punten dat je denkt te halen per onderdeel, inclusief onderbouwing. De docent zal hierop reageren en uiteraard aanpassingen maken aan de beoordeling als blijkt dat er een fout is gemaakt. Deze procedure is de **inzage** voor dit tentamen, en deze duurt tot en met **8 mei 2020**.

Aanpassing beoordelingscriteria. De criteria om voor dit tentamen te slagen die vooraf waren gesteld zijn flink aangepast. Aanleiding hiervoor is de alternatieve vorm die is gebruikt in de Corona situatie. Hierdoor werd het noodzakelijk om tentamenopgaven anders vorm te geven. Een onvoorzien gevolg hiervan is dat het tentamen moeilijker is gebleken dan bedoeld. Om toch zo betrouwbaar mogelijk (binnen de middelen) vast te stellen of aan de eisen is voldaan, zijn de beoordelingscriteria aangepast. Al deze aanpassingen zijn in het voordeel van de deelnemer.

Nieuwe beoordelingscriteria.

- Bij 5 of meer van de 12 punten is de deelnemer geslaagd (voorheen moest het 6 van de 12 punten zijn).
- Er zijn wel deelscores mogelijk per vraag (voorheen zou een vraag alleen óf helemaal goed óf fout zijn).
- Een deelnemer is ook geslaagd als het gemiddelde van het huiswerkcijfer en het puntentotaal van het tentamen een 5 of hoger is (voorheen was het een gewogen gemiddelde; 25% huiswerk, 75% tentamen). Je bent dus bijvoorbeeld ook geslaagd bij 4 punten op het tentamen en een 6 voor het huiswerk.
- In andere gevallen is de deelnemer niet geslaagd.

Scoremodel opgaven. Hierna volgt per opgave een scoremodel. Je kan daarmee per vraag bepalen hoeveel punten je hebt gehaald. Let hierbij op de volgende twee zaken. (1) Deelnemers kregen de opgaven in een willekeurige volgorde. De volgorde van opgaven in dit document komt dus waarschijnlijk niet (kans $\frac{1}{6!}$ van wel) overeen met de volgorde die jij had op het tentamen. (2) De opgaven zijn willekeurig geparametriseerd. In de uitwerkingen staan daarom variabelen waar jij concrete getallen kreeg.

Geschreven uitwerking leidend. Nogmaals voor de duidelijkheid: de geschreven uitwerking is leidend. Je hebt dus wel de punten gekregen als de uitwerking goed was maar de digitale invoer in Grasple fout.

Opgave 1 (1 punt). Zij $X_1, \ldots, X_n$ een rij onafhankelijke en identiek verdeelde stochasten, waarbij X_i voor $i = 1, \ldots, n$ uniform verdeeld is over het interval $[a, b]$, waarbij a en b gegeven. Zij $\bar{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$. Gebruik de normale benadering om de kleinste gehele n te berekenen waarvoor geldt dat $\mathbb{P}[\bar{X}_n < \frac{2a+b}{3}] < p$, voor een gegeven p.

Oplossing 1. We hebben

$$\begin{aligned} \mathbb{E}(\bar{X}_n) &= \mathbb{E}(X_1) = \frac{a+b}{2}, && \text{0.25 punten} \\ \sigma^2(\bar{X}_n) &= \frac{1}{n}\sigma^2(X_1) = \frac{(b-a)^2}{12n}. && \text{0.25 punten} \end{aligned}$$

$$\begin{aligned} \mathbb{P}\left[\bar{X}_n < \frac{2a+b}{3}\right] &= \mathbb{P}\left[\bar{X}_n - \mathbb{E}(\bar{X}_n) < \frac{2a+b}{3} - \frac{a+b}{2}\right] \\ &= \mathbb{P}\left[\bar{X}_n - \mathbb{E}(\bar{X}_n) < \frac{a-b}{6}\right] \\ &= \mathbb{P}\left[\frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sigma(\bar{X}_n)} < \frac{a-b}{6\sigma(\bar{X}_n)}\right] \\ &= \mathbb{P}\left[\frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\sigma(\bar{X}_n)} < -\sqrt{\frac{n}{3}}\right] \quad \text{0.25 punten} \\ &\approx \Phi\left(-\sqrt{n/3}\right). \end{aligned}$$

Dus bij benadering is

$$\begin{aligned} \mathbb{P}\left[\bar{X}_n < \frac{2a+b}{3}\right] < p \quad &\Leftrightarrow \quad -\sqrt{n/3} < \Phi^{-1}(p) \\ &\Leftrightarrow \quad n > 3\Phi^{-1}(p)^2. \qquad \text{0.25 punten} \end{aligned}$$

Er worden geen punten afgetrokken voor een kleine verschrijving waardoor je op andere getallen uitkomt, maar alleen als het duidelijk is dat het een verschrijving is in plaats van verkeerd gebruik formules.

Opgave 2 (3 punten). Zij $X_1, X_2, \ldots$ een rij onafhankelijke en identiek verdeelde stochasten, waarbij X_i voor $i = 1, 2, \ldots$ uniform verdeeld is over het interval $[0, 1]$. We definiëren de volgende twee stochasten:

- $Y = \min\{i \mid 0 \le X_i < p\}$;
- $Z = \min\{j \mid 0 \le 1 - X_j < q\}$;

voor gegeven $p > 0$ en $q > 0$ met $p + q < 1$. Bereken, bijvoorbeeld door middel van conditioneren, $\text{cov}(Y, Z)$. Rond je antwoord af op twee decimalen.

Oplossing 2. Merk op dat Y en Z geometrisch verdeeld zijn met parameters p respectievelijk q. Dus $\mathbb{E}(Y) = 1/p$ en $\mathbb{E}(Z) = 1/q$. Dit tussenresultaat is 1 punt waard.

Om $\mathbb{E}(YZ)$ te kunnen berekenen conditioneren we op X_1. We onderscheiden drie disjuncte gebeurtenissen:

$$\begin{array}{lcl} A = [X_1 < p] & \Rightarrow & Y = 1, Z > 1 \text{ en } \mathbb{E}(YZ \mid A) = \mathbb{E}(Z+1) \\ B = [1 - q < X_1] & \Rightarrow & Y > 1, Z = 1 \text{ en } \mathbb{E}(YZ \mid B) = \mathbb{E}(Y+1) \\ C = [p \le X_1 \le 1 - q] & \Rightarrow & Y > 1, Z > 1 \text{ en } \mathbb{E}(YZ \mid C) = \mathbb{E}((Y+1)(Z+1)) \end{array}$$

0.5 punten voor het benoemen van de drie gevallen en 1 punt voor de implicaties van deze gevallen bij conditioneren. Het af maken van berekening hieronder is dan nog 0.5 punten waard.

Dus
$$\begin{aligned} \mathbb{E}(YZ) &= \mathbb{P}(A)\mathbb{E}(YZ \mid A) + \mathbb{P}(B)\mathbb{E}(YZ \mid B) + \mathbb{P}(C)\mathbb{E}(YZ \mid C) \\ &= p\mathbb{E}(Z+1) + q\mathbb{E}(Y+1) + (1-p-q)\mathbb{E}((Y+1)(Z+1)) \\ &= p\mathbb{E}(Z) + p + q\mathbb{E}(Y) + q + (1-p-q)\mathbb{E}(YZ + Y + Z + 1) \\ (p+q)\mathbb{E}(YZ) &= p\mathbb{E}(Z) + p + q\mathbb{E}(Y) + q + (1-p-q)\mathbb{E}(Y + Z + 1) \\ &= p\mathbb{E}(Z) + p + -q\mathbb{E}(Z) + (1-p)\mathbb{E}(Y + Z + 1) \\ &= -p\mathbb{E}(Y) - q\mathbb{E}(Z) + \mathbb{E}(Y) + \mathbb{E}(Z) + 1 \\ &= -2 + \frac{1}{p} + \frac{1}{q} + 1 \\ \mathbb{E}(YZ) &= \frac{\frac{1}{p} + \frac{1}{q} - 1}{p+q} = \frac{p+q-pq}{pq(p+q)} \\ \text{cov}(Y,Z) &= \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z) \\ &= \frac{p+q-pq}{pq(p+q)} - \frac{1}{pq} \\ &= \frac{-pq}{pq(p+q)} \\ &= -\frac{1}{p+q} \end{aligned}$$

Het geven van een formule voor covariantie levert geen punten op. Het geven van een formule met voorwaardelijke verwachting of iets met conditioneren levert geen punten op als het verder niet gebruikt wordt of tot niets leidt.

Opgave 3 (1 punt). We verdelen een goed geschud pak van 52 kaarten over 4 spelers. Van de 52 kaarten zijn er 13 van het soort schoppen. Bereken de kans dat speler 1 precies n schoppen (met $2 \leq n \leq 4$) en speler 2 precies m schoppen (met $3 \leq m \leq 6$) heeft gekregen. Rond je antwoord af op 3 decimalen.

Oplossing 3. Dit is simpelweg twee hypergeometrische experimenten na elkaar. De kans is

$$\frac{\binom{13}{n}\binom{39}{13-n}}{\binom{52}{13}} \cdot \frac{\binom{13-n}{m}\binom{26+n}{13-m}}{\binom{39}{13}}.$$

0.5 punten voor de "linker breuk" en 0.5 punten als de rest ook helemaal goed is. Dus bijvoorbeeld 0.5 punten als je de twee breuken worden opgeteld i.p.v. vermenigvuldigd. Ook bij een kleine fout in de tweede breuk, zoals vergeten de n op te tellen bij 26 in de term $\binom{26+n}{13-m}$, maximaal 0.5 punten voor de vraag.

Je krijgt wel 1 punt als de formule goed is maar het uitgerekende getal fout.

Je krijgt ook 1 punt als je per ongeluk een typo maakt en daardoor op de verkeerde formule uitkomt. Het moet dan wel duidelijk zijn dat het om een verschrijving gaat en dat het niet mogelijk een verkeerd gebruik is van de formule.

Opgave 4 (2 punten)**.** Zij X een stochast die uniform is verdeeld over het interval $[0,1]$. Zij $Y = -\log(X^n)$, voor een gegeven $n > 0$. Dan is Y exponentieel verdeeld, zeg met parameter a. Bereken a. Rond je antwoord af op 2 decimalen.

Oplossing 4. De kansverdelingsfunctie van X is $F_X(t) = t$ voor $0 \leq t \leq 1$. De kansverdelingsfunctie van Y is voor $t > 0$

$$\begin{aligned} F_Y(t) &= \mathbb{P}[Y \leq t] \\ &= \mathbb{P}[-\log(X^n) \leq t] \\ &= \mathbb{P}[n \log(X) \geq -t] && \text{0.5 punten voor teken omdraaien} \\ &= \mathbb{P}[X \geq e^{-t/n}] && \text{0.5 punten} \\ &= 1 - \mathbb{P}[X < e^{-t/n}] && \text{0.5 punten} \\ &= 1 - F_X(e^{-t/n}) \\ &= 1 - e^{-t/n}. && \text{0.5 punten} \end{aligned}$$

Hieruit volgt dat de parameter van Y gelijk is aan $a = 1/n$.

Opmerking. Sommige deelnemers hebben de log opgevat als de logaritme met basis 10. In dat geval krijg je (zelfde idee als hierboven)

$$\begin{aligned} F_Y(t) &= 1 - 10^{-t/n} \\ &= 1 - e^{-t\ln(10)/n}. \end{aligned}$$

En dan is dus $a = \ln(10)/n$. Omdat het onduidelijk was dat de natuurlijke logaritme bedoeld was, levert een goede uitwerking, maar dan met basis 10, ook punten op analoog als boven.

Opgave 5 (3 punten). Zij Z een stochast die standaardnormaal verdeeld is. Zij $X = e^{sZ+m}$, en zij $f(x)$ een dichtheidsfunctie voor X die differentieerbaar is voor $x > 0$. Bepaal de modus van X: dat is de $x > 0$ die $f(x)$ maximaliseert. Rond je antwoord af op 2 decimalen.

Oplossing 5. We bepalen eerst $f(x)$ via de verdelingsfunctie $F(x)$ van X.

$$\begin{aligned} F(x) &= \mathbb{P}[X \leq x] \\ &= \mathbb{P}[e^{sZ+m} \leq x] \\ &= \mathbb{P}[sZ + m \leq \log(x)] \\ &= \mathbb{P}\left[Z \leq \frac{\log(x) - m}{s}\right] \\ &= \Phi\left(\frac{\log(x) - m}{s}\right) \end{aligned}$$

Dit tussenresultaat is 1 punt waard.

$$\begin{aligned} f(x) &= F'(x) \\ &= \frac{1}{sx}\varphi\left(\frac{\log(x) - m}{s}\right) \end{aligned}$$

Dit tussenresultaat is 1 punt waard.

$$f'(x) = \frac{-1}{sx^2}\varphi\left(\frac{\log(x) - m}{s}\right) + \frac{1}{(sx)^2}\varphi'\left(\frac{\log(x) - m}{s}\right)$$

omdat $\varphi(t) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}t^2)$ is $\varphi'(t) = -t\varphi(t)$

$$= \frac{-1}{sx^2}\varphi\left(\frac{\log(x) - m}{s}\right) - \frac{1}{(sx)^2}\frac{\log(x) - m}{s}\varphi\left(\frac{\log(x) - m}{s}\right)$$

Dit tussenresultaat is 0.5 punten waard. nu zoeken naar het maximum van f

$$\begin{aligned} 0 &= f'(x) \\ 0 &= \frac{-1}{s} - \frac{1}{(s)^2}\frac{\log(x) - m}{s} \\ \log(x) - m &= -s^2 \\ x &= e^{m-s^2} \quad \text{0.5 punten} \end{aligned}$$

De gedachtegang dat $Z = \frac{\log(X)-m}{s}$ en dus $f(x) = \phi\left(\frac{\log(x)-m}{s}\right)$ klopt totaal niet (kijk maar boven: je mist de kettingregel). Bij dit soort fouten 0 punten voor de opgave.

Opgave 6 (2 punten). Zij X een stochast die uniform verdeeld is over het interval $[0,1]$. We definiëren de volgende twee stochasten met de indicator functie:

- $Y = 1_{[X<p]}$;
- $Z = 1_{[1-X<q]}$;

voor gegeven $p > 0$ en $q > 0$ met $p + q < 1$. Bereken de correlatie-coëfficiënt van Y en Z. Rond je antwoord af op 2 decimalen.

Oplossing 6. Merk op dat $YZ = 0$ met kans 1, want $X < p$ sluit uit dat $1 - X < q$. Dus $\mathbb{E}(YZ) = 0$. Dit tussenresultaat is 1 punt waard.

Y en Z zijn Bernoulli verdeeld met parameter p respectievelijk q. Hiervan weten we dat

$$\begin{aligned} \mathbb{E}(Y) &= p, & \sigma^2(Y) &= p(1-p), \\ \mathbb{E}(Z) &= q, & \sigma^2(Z) &= q(1-q). \end{aligned}$$

Elk van bovenstaande 4 vergelijkingen levert 0.25 punten op.

Merk op deze tussenresultaten al optellen tot de 2 punten die de opgave waard is. Invullen in de formules hieronder is nu een trivialiteit die dus geen extra punten oplevert (maar ook niet kost als het mis gaat).

De covariantie is dan

$$\mathrm{cov}(YZ) = \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z) = 0 - pq = -pq.$$

De correlatie is

$$\begin{aligned} \rho(Y,Z) &= \frac{\mathrm{cov}(Y,Z)}{\sigma(Y)\sigma(Z)} \\ &= \frac{-pq}{\sqrt{p(1-p)q(1-q)}} \\ &= -\sqrt{\frac{pq}{(1-p)(1-q)}}. \end{aligned}$$