Deeltentamen II Kansrekening en Statistiek EE1M31
13 april 2015

Bij dit tentamen is het gebruik van een rekenmachine toegestaan. Ook toegestaan is een formuleblad, dat wordt uitgedeeld.
Dit deeltentamen bestaat uit zes meerkeuze- en drie open vragen.

Toelichting: Maak op het bijgeleverde antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Doorstrepen van een fout antwoord heeft geen zin: u moet het òf uitgummen, òf verwijderen met correctievloeistof òf een nieuw formulier invullen. Vergeet niet uw studienummer in te vullen èn aan te strepen.

Meerkeuze vragen

1. Stel X is een Bernoulli-verdeelde stochast met parameter $0 < p < 1$. Volgens de ongelijkheid van Chebyshev geldt dan

a. $\mathrm{P}\left(|X-p| \geq \frac{1}{2}\right) \geq \frac{1}{4}p(1-p)$
b. $\mathrm{P}\left(|X-p| \geq \frac{1}{2}\right) \leq \frac{1}{4}p(1-p)$
c. $\mathrm{P}\left(|X-p| \leq \frac{1}{2}\right) \geq \frac{1}{4}p(1-p)$
d. $\mathrm{P}\left(|X-p| \geq \frac{1}{2}\right) \leq 4p(1-p)$
e. $\mathrm{P}\left(|X-p| \geq \frac{1}{2}\right) \geq 4p(1-p)$
f. $\mathrm{P}\left(|X-p| \leq \frac{1}{2}\right) \leq 4p(1-p)$

2. Studievereniging Elektron verkoopt als service aan de studenten boeken tegen een sterk gereduceerd tarief. Aan het begin van het studiejaar moeten voldoende boeken op voorraad zijn, en aan het eind blijven er hopelijk niet te veel boeken onverkocht (want wie weet komt de uitgever weer met een nieuwe druk). Begin augustus moet inkoper Wouter beslissen hoeveel boeken Probability & Statistics hij gaat bestellen voor het eerstejaars vak kansrekening en statistiek. De ervaring leert dat van de eerstejaars gemiddeld 80 % het boek koopt. Neem aan dat elke eerstejaars student, onafhankelijk van wat de andere studenten doen, met 'kans' 0.8 het boek zal kopen (en dus met 'kans' 0.2 het boek **niet** zal kopen). Stel dat er bekend is dat de komende lichting 100 eerstejaars studenten bevat. Hoeveel boeken moet Wouter minimaal bestellen opdat de kans dat hij met een tekort te maken krijgt kleiner is dan 2 %? (Hint: gebruik de Centrale Limietstelling)

a. 50 **b.** 65 **c.** 69 **d.** 80 **e.** 89 **f.** 100

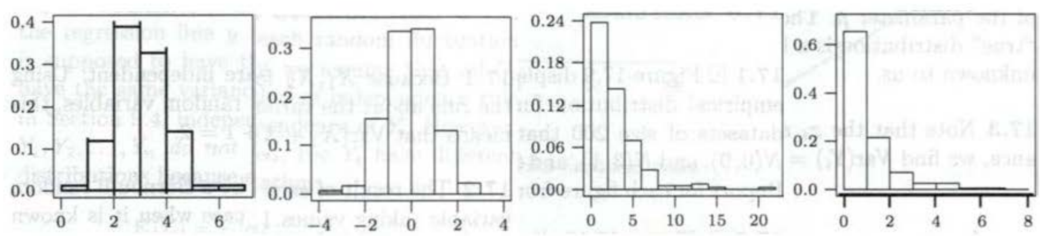
3. Een verdelingsfunctie is gegeven door

$$F(x) = \begin{cases} 0\ , & \text{als } x \leq 1 \\ \frac{1}{4}(x-1)^2, & \text{als } 1 \leq x \leq 3 \\ 1\ , & \text{als } 3 \leq x \end{cases}$$

Laat $U \sim U(0,1)$. Bepaal de stochast X, geconstrueerd met behulp van U, die verdelingsfunctie F heeft:

a. $X = \sqrt{U}$
b. $X = 1 - 2\sqrt{U}$
c. $X = 1 + 2\sqrt{U}$
d. $X = 2 + 4\sqrt{U}$
e. $X = 1 + 4\sqrt{U}$
f. $X = \frac{1}{2}\sqrt{1-U}$

4. Bekijk de onderstaande histogrammen.



Dataset I Dataset II Data set III Data set IV

Welke bewering is het meest waarschijnlijk?

- **a.** Dataset I komt uit een N(0,1)-verdeling en dataset III uit een Exp(1/3)-verdeling
- **b.** Dataset II komt uit een N(0,1)-verdeling en dataset III uit een Exp(1)-verdeling
- **c.** Dataset I komt uit een N(0,1)-verdeling en dataset IV uit een Exp(1/3)-verdeling
- **d.** Dataset I komt uit een N(3,1)-verdeling en dataset II uit een Norm(2,4)-verdeling
- **e.** Dataset I komt uit een N(3,1)-verdeling en dataset III uit een Exp(1/3)-verdeling
- **f.** Dataset I komt uit een N(3,1)-verdeling en dataset IV uit een Exp(1/3)-verdeling

5. Zij $\bar{X}_n$ and $\bar{Y}_m$ de steekproefgemiddelden van twee onafhankelijke steekproeven van omvang n respectievelijk m uit dezelfde kansverdeling met verwachting μ en variantie σ^2. Zoals bekend uit Hoofdstuk 17 van het boek is het steekproefgemiddelde een schatter voor μ. We combineren de twee schatters tot een nieuwe schatter

$$T = r\bar{X}_n + (1-r)\bar{Y}_m, \qquad \text{waarbij } r \text{ een getal is tussen 0 en 1.}$$

Een schatter T heet *zuiver* voor de parameter μ van een verdeling als geldt

$$\mathrm{E}\left[T\right] = \mu.$$

Voor welke r geldt dat T een zuivere schatter is voor μ?

- **a.** voor alle $0 < r < 1$
- **b.** voor geen enkele $0 < r < 1$
- **c.** alleen voor $r = \frac{1}{4}$
- **d.** alleen voor $r = \frac{3}{4}$
- **e.** alleen voor $r = \frac{1}{m} + \frac{1}{n}$
- **f.** alleen voor $r = |\frac{1}{n} - \frac{1}{m}|$

6. Een Casino in een beroemde badplaats staat terecht in de volgende zaak: Een geliefd spel in het Casino is dat een klant de uitkomst van een muntworp voorspelt (K of M). Dit spel is legaal zolang er met een zuivere munt wordt geworpen. Een controleur van de Gokautoriteit Nederland vermoedt echter dat het casino een onzuivere munt gebruikt die de uitkomst K vaker laat voorkomen en heeft het casino aangeklaagd. Tijdens de daarop volgende rechtszaak gooit de rechter 10 keer met de gewraakte munt, met als uitkomst: KKKKKKKMKM.
Stel $p = P(K)$. Welke toetsopzet geeft de beschreven situatie correct weer?

- **a.** $H_0 : p = \frac{1}{2}$, $H_1 : p < \frac{1}{2}$ en $p-$ waarde $= \mathrm{P}\left(X \geq 8 | X \sim Bin(10, \frac{1}{2})\right)$
- **b.** $H_0 : p = \frac{1}{2}$, $H_1 : p \neq \frac{1}{2}$ en $p-$ waarde $= \mathrm{P}\left(X \geq 8 | X \sim Bin(10, \frac{1}{2})\right)$
- **c.** $H_0 : p = \frac{1}{2}$, $H_1 : p > \frac{1}{2}$ en $p-$ waarde $= \mathrm{P}\left(X \geq 8 | X \sim Bin(10, \frac{1}{2})\right)$
- **d.** $H_0 : p = \frac{1}{2}$, $H_1 : p < \frac{1}{2}$ en $p-$ waarde $= \mathrm{P}\left(X \geq 8 | X \sim Bin(8, \frac{1}{2})\right)$
- **e.** $H_0 : p = \frac{1}{2}$, $H_1 : p \neq \frac{1}{2}$ en $p-$ waarde $= \mathrm{P}\left(X \geq 8 | X \sim Bin(8, \frac{1}{2})\right)$
- **f.** $H_0 : p = \frac{1}{2}$, $H_1 : p > \frac{1}{2}$ en $p-$ waarde $= \mathrm{P}\left(X \geq 8 | X \sim Bin(8, \frac{1}{2})\right)$

Open vragen

Toelichting: Een antwoord alleen is *niet* voldoende: er dient een berekening, toelichting en/of motivatie aanwezig te zijn. Dit alles goed leesbaar en in goed Nederlands.

1. Bekijk de data punten

$$(0,0),\ (1,0),\ (2,1),\ (2,2).$$

We gaan op zoek naar de rechte lijn die het beste past bij de data in de zin van de kleinste kwadratenoplossing. Het model is

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

De matrixformulering voor dit regressiemodel is

$$\mathbf{y} = X\beta + \epsilon,$$

waarbij $\epsilon_i \sim N(0, \sigma^2)$.

a. Geef de designmatrix X en de observatievector $\mathbf{y}$.

b. Geef de regressielijn door deze datapunten, door kleinste kwadratenschattingen voor het startgetal en de richtingscoëfficiënt te berekenen met behulp van de matrixformulering.

2. Bij een groep van 25 proefpersonen wordt de reactietijd in seconden op een bepaalde taak vastgesteld, zowel onmiddelijk vóór als een half uur na de consumptie van één liter bier.
Per persoon wordt het verschil D in reactietijd voor en na het drinken bepaald. Het vermoeden is dat de alcoholconsumptie de reactietijd doet toenemen, met andere woorden, men vermoedt dat D positief is.
De uitkomsten van de 25 proefpersonen geven een gemiddeld verschil $\bar{D}_{25}$ van 1.32 sec met steekproefstandaarddeviatie $s_{25} = 1.02$.
Toets het vermoeden op significantieniveau $\alpha = 0.1$:

a. Stel de relevante hypotheses op.

b. Bereken de waarde van de toetsingsgrootheid.

c. Geef de conclusie van de toets.

3. Een rij onafhankelijke continue stochasten $X_1, X_2, \cdots, X_{225}$ hebben allemaal dezelfde verdelingsfunctie

$$F_X(x) = \begin{cases} 0\ , & \text{als } x \le 0 \\ \sqrt{x}, & \text{als } 0 \le x \le 1 \\ 1\ , & \text{als } 1 \le x \end{cases}$$

a. Bereken $\mathrm{E}\left[X_1\right]$

b. Bereken $\mathrm{Var}\left(X_1\right)$

c. Gebruik de Centrale Limietstelling om $\mathrm{P}\left(X_1 + X_2 + \cdots X_{225} \le 75\right)$ te benaderen.

Antwoorden multiple choice:

1 **d.**

2 **e.**

3 **c.**

4 **e.**

5 **a.**

6 **c.**

Antwoorden open vragen:

1a De vector $\mathbf{y} = [0, 0, 1, 2]^T$ en de matrix X is:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}.$$

1b $y = -\frac{3}{11} + \frac{9}{11}x.$

2a $H_0 : D = 0,\ H_1 : D > 0.$

2b $t = 9.354$

2c Er is wel effect op de reactietijd

3a $\mathrm{E}\left[X_1\right] = \frac{1}{3}$

3b $\mathrm{Var}\left(X_1\right) = \frac{4}{45}$

3c 0.5

Uitwerkingen multiple choice:

1 De ongelijkheid van Chebyshev zegt dat voor een stochast X en $a > 0$

$$\mathrm{P}\left(|X - \mathrm{E}[X]| \geq a\right) \leq \frac{1}{a^2}\mathrm{Var}(X).$$

Neem $a = \frac{1}{2}$, $\mathrm{E}[X] = p$ en $\mathrm{Var}(X) = p(1-p)$.

2 We moeten de minimale gehele c berekenen zodat $\mathrm{P}(X \geq c) \leq 0.02$. $X \sim Ber(0.8)$. Volgens de Centrale Limietstelling kunnen we deze kans benaderen als volgt:

$$\begin{aligned}\mathrm{P}(X_1 + X_2 + \cdots + X_{100} \geq c) &= \mathrm{P}\left(\bar{X}_{100} \geq \frac{c}{100}\right)\\ &= \mathrm{P}\left(\frac{\bar{X}_{100} - \mathrm{E}[X]}{\sigma/\sqrt{n}} \geq \frac{\frac{c}{100} - 0.8}{\sqrt{0.16/10}}\right) \approx \mathrm{P}\left(Z \geq \frac{\frac{c}{100} - 0.8}{0.04}\right)\end{aligned}$$

Dus er moet gelden dat

$$\frac{\frac{c}{100} - 0.8}{0.04} \geq 2.05 \text{ en dus } c \geq 88.2.$$

Aangezien c een geheel aantal boeken is: $c = 89$.

3 We gebruiken de methode van de inverse verdelingsfunctie: $U = \frac{1}{4}(X-1)^2 \iff X = 1 + 2\sqrt{U}$.

4 Dit is een deel van opgave 17.1

5 Allereerst geldt $\mathrm{E}\left[\bar{X}_n\right] = \mu$ en $\mathrm{E}\left[\bar{Y}_m\right] = \mu$. Vanwege de lineariteit van verwachting geldt dan

$$\mathrm{E}[T] = \mathrm{E}\left[r\bar{X}_n + (1-r)\bar{Y}_m\right] = r\mathrm{E}\left[\bar{X}_n\right] + (1-r)\mathrm{E}\left[\bar{Y}_m\right] = r\mu + (1-r)\mu = \mu.$$

Dus T is zuiver voor μ, onafhankelijk van r.

6 We toetsen of de munt zuiver is ($p = \frac{1}{2}$) tegen de hypothese dat hij K bevoordeelt($p > \frac{1}{2}$).
In de beschikbare data heeft de rechter 8 keer K gegooid. Hoe waarschijnlijk is dat als de munt zuiver zou zijn? Dit is de p-waarde die je kunt berekenen onde de nulhypothese, dus $\mathrm{P}\left(X \geq 8 | X \sim Bin(10, \frac{1}{2})\right)$.

Uitwerkingen open vragen:

1a De vector $\mathbf{y} = [0, 0, 1, 2]^T$ en de matrix X is:

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix}.$$

1b Er geldt

$$\mathbf{X^T X} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix}$$

$$\mathbf{X^T y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

De normaalvergelijkingen zijn

$$\begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

De oplossing van de normaal vergelijkingen is

$$\begin{aligned} \mathbf{b} = (\mathbf{X^TX})^{-1}\mathbf{X^Ty} &= \begin{pmatrix} 4 & 5 \\ 5 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ 6 \end{pmatrix} \\ &= \frac{1}{11} \begin{pmatrix} 9 & -5 \\ -5 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix} \\ &= \frac{1}{11} \begin{pmatrix} -3 \\ 9 \end{pmatrix} = \begin{pmatrix} -3/11 \\ 9/11 \end{pmatrix} \end{aligned}$$

Dus de lijn is

$$y = -\frac{3}{11} + \frac{9}{11}x.$$

2a $H_0 : D = 0$, $H_1 : D > 0$.

2b

$$T = \frac{\bar{D} - \mu_0}{s_n/\sqrt{n}} = \frac{1.32 - 0}{1.02/\sqrt{25}} = 6.471$$

2c De kritieke waarde waarmee we de uitkomst van (b) moeten vergelijken is $t_{n-1,\alpha} = t_{24,0.1} = 1.318$. De gevonden waarde van T ligt dus in het kritieke gebied, dus we verwerpen H_0 en concluderen dat er inderdaad een effect is van het drinken van alcohol op de reactietijd.

3a Om de verwachting te berekenen hebben we de dichtheid f_X nodig, en we weten $f_X(x) = F'_X(x) = \frac{1}{2\sqrt{x}}$.
Dus

$$\mathrm{E}\left[X\right] = \int_0^1 x\frac{1}{2\sqrt{x}}\,dx = \frac{1}{2}\left[\frac{2}{3}x^{3/2}\right] = \frac{1}{3}$$

3b

$$\mathrm{E}\left[X^2\right] = \int_0^1 x^2\frac{1}{2\sqrt{x}}\,dx = \frac{1}{2}\left[\frac{2}{5}x^{5/2}\right] = \frac{1}{5}$$

Dus $\mathrm{Var}\left(X\right) = \frac{1}{5} - \frac{1}{9} = \frac{4}{45}$.

3c De gevraagde kans is

$$\begin{aligned} \mathrm{P}\left(X_1 + X_2 + \cdots + X_{225} \leq 75\right) &= \mathrm{P}\left(\bar{X}_{225} \leq \frac{75}{225}\right) \\ &\approx \mathrm{P}\quad Z \leq \frac{\frac{75}{225} - \frac{1}{3}}{\sqrt{4/45}/15}\Bigg) \\ &= \mathrm{P}\left(Z \leq 0\right) \\ &= 0.5 \end{aligned}$$