Technische Universiteit Delft — Mekelweg 4
Faculteit Elektrotechniek, Wiskunde en Informatica — 2628 CD Delft

Hertentamen Kansrekening en Statistiek EE1M31
27 juli 2016, 9.00 – 12.00u

Bij dit tentamen is het gebruik van een **niet-grafische** rekenmachine toegestaan. Ook toegestaan is een formuleblad, dat wordt uitgedeeld.
Dit tentamen bestaat uit vijftien meerkeuze- en drie open vragen. De drie open vragen dient u op het tentamenvel te maken.

Cijferbepaling: Iedere meerkeuzevraag telt voor 1 punt, bij de open vragen is het aantal punten per onderdeel aangegeven. Er geldt:

$$\text{Cijfer} = \frac{MC + OV}{3} + 1,$$

waarbij MC het aantal punten voor meerkeuze-deel en OV het aantal punten voor open-vragen-deel is.
Toelichting meerkeuze-vragen: Maak op het bijgeleverde meerkeuze-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. Tenslotte, vergeet niet de versie, de vakcode, uw naam en uw studienummer in te vullen. De laatste dient u ook aan te strepen. Uw tentamen wordt anders niet gecorrigeerd.

Meerkeuze-vragen
Versie A

1. Laat A en B twee onafhankelijke gebeurtenissen (*independent events*) zijn. Er is gegeven dat $P(A) = \frac{1}{3}$ en $P(B) = \frac{1}{6}$.

   Bereken $P(A \cup B)$.

   a. $\frac{1}{3}$ **b.** $\frac{2}{3}$ **c.** $\frac{7}{9}$ **d.** $\frac{4}{9}$ **e.** $\frac{1}{2}$ **f.** $\frac{3}{8}$

2. Een chipfabrikant heeft 3 machines: A, S en M. De kans dat een machine een defecte chip produceert is respectievelijk $2\%, 3\%$ en 5%. De machines nemen respectievelijk $40\%, 35\%$ en 25% van de chipproductie voor hun rekening.

   Gegeven dat een willekeurig gekozen chip defect is, bereken in 2 decimalen nauwkeurig de voorwaardelijke kans (*conditional probability*) dat machine M deze chip produceerde.

   a. 0.25 **b.** 0.40 **c.** 0.02 **d.** 0.35 **e.** 0.33 **f.** 0.05

3. Laat $X \sim Geo(\frac{1}{4})$ en $Y \sim Geo(\frac{1}{2})$. De stochasten (*random variables*) X en Y zijn onafhankelijk.

   Bereken $P(X + Y = 3)$.

   a. $\frac{75}{512}$ **b.** $\frac{17}{64}$ **c.** $\frac{3}{64}$ **d.** $\frac{17}{128}$ **e.** $\frac{9}{64}$ **f.** $\frac{5}{32}$

4. Laat X een continue stochast zijn met kansdichtheid (*probability density*) $f_X(x) = \frac{1}{4}x^3$ als $0 \le x \le 2$ en $f_X(x) = 0$ anders. De stochast U heeft een $U(0,1)$ verdeling. Je wilt X simuleren m.b.v. U.

   Welke stochast, als functie van U, is hiervoor geschikt?

   a. $\frac{1}{4}U^3$ **b.** $\sqrt[2]{\frac{3}{4}U}$ **c.** $\frac{3}{4}U^2$ **d.** $\sqrt[2]{\frac{4}{3}U}$ **e.** $\sqrt[4]{4U}$ **f.** $2\sqrt[4]{U}$

5. X en Y zijn twee discrete stochasten met de volgende gezamenlijke kansmassafunctie (*joint probability mass function*):

$P(X=a,Y=b)$		$b=0$	$b=1$	$b=2$
a	-1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	1	0	$\frac{1}{2}$	0

Voor X en Y geldt dan het volgende:

a. X en Y zijn onafhankelijk en positief gecorreleerd (*positively correlated*)
b. X en Y zijn afhankelijk en positief gecorreleerd
c. X en Y zijn onafhankelijk en negatief gecorreleerd (*negatively correlated*)
d. X en Y zijn afhankelijk en negatief gecorreleerd
e. X en Y zijn afhankelijk en ongecorreleerd (*uncorrelated*)
f. X en Y zijn onafhankelijk en ongecorreleerd

6. Laat $X \sim U(0,2)$.

Bereken $\mathrm{E}\left[e^{-X}\right]$.

a. $\frac{1}{2}(1-e^{-2})$ **b.** $(1-e^{-1})^2$ **c.** e^{-2}
d. 1 **e.** $1-e^{-2}$ **f.** e^{-1}

7. Modelleer het binnenkomen van spam-berichten in je mailbox als een Poisson proces met intensiteit 3 (aantal berichten per dag).

Gegeven dat er de afgelopen 6 uur geen spam-bericht binnenkwam, wat is de kans dat het vanaf dit moment nog tenminste 2 uur duurt voordat er een nieuw spam-bericht binnenkomt?

a. $e^{-\frac{1}{2}}$ **b.** $e^{-\frac{1}{4}}$ **c.** $1-e^{-\frac{1}{4}}$ **d.** e^{-1} **e.** $1-e^{-1}$ **f.** $1-e^{-\frac{1}{2}}$

8. Van een berucht eerstejaarsvak van de opleiding Electrical Engineering op de TU Delft kunnen we de hertentamencijfers beschouwen als een steekproef uit een (onbekende) verdeling met verwachting $\mu = 5.5$ en variantie $\sigma^2 = 4$. Aan het hertentamen doen 49 studenten mee.

Benader de kans dat het gemiddelde cijfer lager is dan een 5.

a. 0.04 **b.** 0.19 **c.** 0.55 **d.** 0.45 **e.** 0.40 **f.** Onvoldoende informatie beschikbaar

9. Je hebt 100 zuivere munten op tafel liggen en je werpt ieder van deze munten eenmaal. Het aantal keer kop noteren we met X. Vervolgens gooi je iedere munt waarvan de uitkomst kop was eenmaal opnieuw. Het aantal keer kop dat nu op tafel ligt noteren we met Y.

Wat is de verdeling van Y?

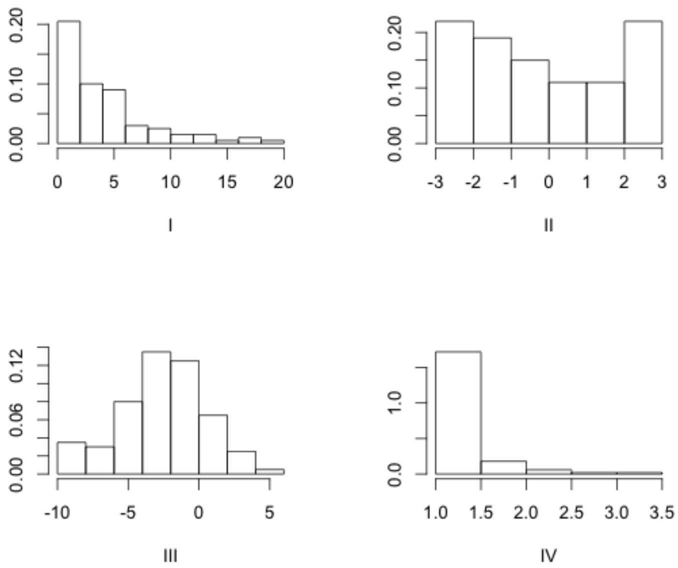
a. $Bin(50, 0.25)$ **b.** $Bin(25, 0.25)$ **c.** $Bin(25, 0.5)$
d. $Bin(100, 0.25)$ **e.** $Bin(100, 0.5)$ **f.** $Bin(50, 0.5)$

10. Laat X_1, X_2, X_3 een aselecte steekproef (*random sample*) zijn uit een verdeling met verwachting μ en variantie σ^2. Laat $T = X_1 - \frac{X_2}{2} + \frac{X_3}{2}$ een schatter (*estimator*) voor μ zijn.

Bereken de gemiddelde kwadratische fout (*mean squared error*) van T.

a. $\frac{3\sigma}{2}$ **b.** $\frac{3\sigma^2}{2}$ **c.** $2\sigma^2$ **d.** σ^2 **e.** σ **f.** 2σ

11. Van vier datasets ter grootte 100 zijn hieronder de histogrammen weergegeven. De datasets zijn steekproeven uit 4 verschillende verdelingen: $U(-3,3), N(-3,-3), Exp(\frac{1}{4})$ en $Par(4)$.



Welke dataset komt uit de $U(-3,3)$-verdeling en welke komt uit de $Exp(\frac{1}{4})$-verdeling?

a. $U(-3,3)$: II; $Exp(\frac{1}{4})$: I

b. $U(-3,3)$: III; $Exp(\frac{1}{4})$: I

c. $U(-3,3)$: I; $Exp(\frac{1}{4})$: IV

d. $U(-3,3)$: IV; $Exp(\frac{1}{4})$: III

e. $U(-3,3)$: II; $Exp(\frac{1}{4})$: IV

f. $U(-3,3)$: III; $Exp(\frac{1}{4})$: II

12. Beschouw de volgende twee beweringen:

   (A) Als $(100, 200)$ een 95% betrouwbaarheidsinterval (*confidence interval*) is voor θ, dan geldt $P(\theta \in (100, 200)) = 0.95$.

   (B) Het significantieniveau (*significance level*) is de kans dat de nulhypothese juist is.

   Welke van deze beweringen is/zijn waar?

   a. allebei onwaar **b.** alleen (A) waar **c.** alleen (B) waar **d.** allebei waar

13. Laat $x_1, \ldots, x_n$ een dataset zijn die we beschouwen als een realisatie van een aselecte steekproef $X_1, \ldots, X_n$ uit een verdeling met kansdichtheid $f_\theta(x)$ gegeven door

$$f_\theta(x) = \begin{cases} 2\theta x e^{-\theta x^2} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

   Bereken de meest aannemelijke schatting (*maximum likelihood estimate*) van θ.

   a. $\frac{n}{\sum_{i=1}^n x_i}$ **b.** $\frac{\sum_{i=1}^n x_i}{n}$ **c.** $\frac{n\sum_{i=1}^n x_i^2}{\prod_{i=1}^n x_i}$ **d.** $\frac{\sum_{i=1}^n x_i^2}{n}$ **e.** $\frac{n}{\sum_{i=1}^n x_i^2}$ **f.** $\frac{\sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2}$

14. Bekijk de volgende geordende dataset

$$12 \quad 14 \quad 14 \quad 20 \quad 21 \quad 22 \quad 23 \quad 23 \quad 27 \quad 28.$$

Bepaal de waarde van de empirische verdelingsfunctie $F_{10}(x)$ in $x = 23$.

a. 2/5 **b.** 1/5 **c.** 4/5 **d.** 7/10 **e.** 3/5 **f.** 3/10

15. Laat x een realisatie zijn van een $Pois(\lambda)$-verdeling. We toetsen $H_0 : \lambda = 1$ tegen $H_1 : \lambda > 1$ en we verwerpen H_0 ten gunste van H_1 als $x > 2$.

Bereken de kans op een Type II-fout als de ware waarde van λ gelijk aan 2 is.

a. $\frac{3}{2}e^{-1}$ **b.** $5e^{-2}$ **c.** $1 - 5e^{-2}$ **d.** $1 - \frac{5}{2}e^{-1}$ **e.** $1 - 3e^{-2}$ **f.** $\frac{5}{2}e^{-1}$

Hertentamen Kansrekening en Statistiek EE1M31 27 juli 2016

Naam: Groep: *Cijfer:*

Studentnummer:

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. Laat X en Y twee continue stochasten zijn. De gezamenlijke dichtheid van X en Y wordt gegeven door $f(x, y) = Kxy$, als $0 \le x \le y \le 1$ en $f(x, y) = 0$ anders.

   (a) (1.5 punten) Laat zien dat $K = 8$.

   (b) (1.5 punten) Bereken $\mathrm{E}\left[\frac{1}{Y}\right]$.

   (c) (1.5 punten) Bereken de marginale dichtheden (*marginal densities*) van X en Y en bepaal hiermee of X en Y onafhankelijk zijn.

2. Voor een bivariate dataset $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, bestuderen we eerst het lineaire regressiemodel (*linear regression model*)

$$Y_i = \alpha + \beta x_i + U_i,$$

waarbij de U_i onafhankelijk en $N(0, \sigma^2)$-verdeeld zijn. Het is bekend (zie het formuleblad) dat de kleinste kwadratenschatter (*least squares estimator*) voor α gelijk is aan $\hat{\alpha} = \bar{Y}_n - \hat{\beta}\bar{x}_n$, waarbij $\hat{\beta}$ de kleinste kwadratenschatter voor β is.

(a) (1 punt) Toon aan dat $\hat{\alpha}$ een zuivere (*unbiased*) schatter is voor α. U mag hierbij aannemen dat $\hat{\beta}$ een zuivere schatter is voor β.

Bekijk nu het volgende lineaire regressiemodel zonder intercept:

$$Y_i = \beta x_i + U_i,$$

waarbij de U_i onderling onafhankelijk en $N(0, \sigma^2)$-verdeeld zijn.

(b) (1.5 punten) Bepaal de kleinste kwadratenschatter $\hat{\beta}$ voor β in dit model zonder intercept.

3. (a) (1 punt) Laat de stochast X een $Bin(n, p)$-verdeling hebben. Toon aan dat de stochast $Z = \dfrac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}$ bij benadering een standaard normale verdeling heeft.

Vervolg vraag 3!

De bekende opiniepeiler Maurice de Kadt heeft een peiling gehouden over de vraag of Nederland in de Europese Unie dient te blijven. Uit een aselecte steekproef onder 1024 burgers blijkt dat 250 van hen voor het vertrek van Nederland ('Nexit') uit de Europese Unie is. Laat p de onbekende fractie van burgers voor een Nexit onder de gehele stemgerechtigde bevolking zijn.

(b) (2 punten) Bepaal m.b.v. onderdeel (a) het tweezijdig 95% -betrouwbaarheidsinterval voor p.

Uit dezelfde peiling bleek dat 800 van de 1024 ondervraagden zou gaan stemmen bij een referendum over een vertrek van Nederland uit de Europese Unie. Laat q het werkelijke (nog onbekende) opkomstpercentage zijn bij een referendum over een Nexit.

(c) (2 punten) Onderzoek of het opkomstpercentage q hoger dan 75% is bij een significantieniveau van 1%: formuleer de relevante hypothesen, toetsingsgrootheid met verdeling (gebruik wederom onderdeel (a)!) en gemotiveerde conclusie.

Extra ruimte

Antwoorden:

1 d.

2 b.

3 f.

4 f.

5 **e.**

6 a.

7 **b.**

8 **a.**

9 d.

10 b.

11 a.

12 **a.**

13 **e.**

14 c.

15 b.

b $\frac{4}{3}$

c $f_X(x) = 4(x - x^3), f_Y(y) = 4y^3$, dus afhankelijk

b $\hat{\beta} = \frac{\sum x_i Y}{\sum x_i^2}$

b $(0.22, .27)$

c $H_0 : q = 0.75$ tegen $H_1 : q > 0.75$. $z = \frac{800-768}{\sqrt{(32)}} \approx 2.31$, dus p-waarde is ongeveer 0.0104 en we verwerpen H_0 niet.