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

Deeltentamen II Kansrekening en Statistiek EE1M31
13 april 2016, 9.00 – 11.00u

Bij dit tentamen is het gebruik van een **niet-grafische** rekenmachine toegestaan. Ook toegestaan is een formuleblad, dat wordt uitgedeeld.
Dit tentamen bestaat uit negen 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:

$$\text{Cijfer} = \frac{MC + OV}{2} + 1,$$

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. $X_1, X_2, \ldots, X_{100}$ zijn onafhankelijke en gelijk verdeelde stochasten (*independent, identically distributed random variables*) met een $U(0,6)$-verdeling.
   De som $\sum_{i=1}^{100} X_i$ heeft dan bij benadering de volgende verdeling:
   a. $N(300,300)$ **b.** $N(200,600)$ **c.** $U(300,600)$
   d. $N(300,200)$ **e.** $N(200,300)$ **f.** $U(0,600)$

2. Laat X een *Exp*(2) verdeling hebben. Volgens de ongelijkheid van Chebyshev (*Chebyshev's inequality*) geldt dan:
   a. $\mathrm{P}\big(|X-\frac{1}{2}|<2\big) \le \frac{1}{4}$ **b.** $\mathrm{P}(|X-2|<2) \ge \frac{15}{16}$
   c. $\mathrm{P}\big(|X-\frac{1}{2}|<2\big) \le \frac{1}{16}$ **d.** $\mathrm{P}(|X-2|<2) \ge \frac{3}{4}$
   e. $\mathrm{P}\big(|X-\frac{1}{2}|<2\big) \ge \frac{15}{16}$ **f.** $\mathrm{P}(|X-2|<2) < \frac{1}{16}$

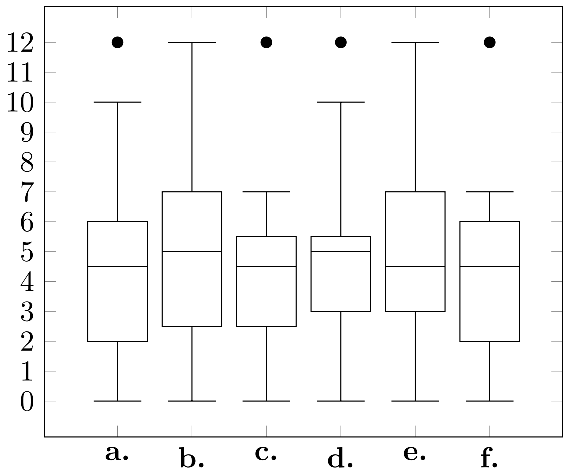
3. Beschouw betrouwbaarheidsintervallen (*confidence intervals*) voor het gemiddelde (*mean*) van normale data met bekende variantie σ^2. Welk van de beweringen is/zijn waar:
   het betrouwbaarheidsinterval wordt kleiner als
   (A) de steekproefomvang (*sample size*) groter wordt
   (bij gelijkblijvende betrouwbaarheid en variantie);
   (B) de betrouwbaarheid (*confidence*) groter wordt
   (bij gelijkblijvende steekproefomvang en variantie);
   (C) de variantie (*variance*) groter wordt
   (bij gelijkblijvende steekproefomvang en betrouwbaarheid).

   a. alle drie onwaar **b.** alleen (A) waar **c.** alleen (B) waar **d.** alleen (C) waar
   e. (A) en (B) waar **f.** (A) en (C) waar **g.** (B) en (C) waar **h.** alle drie waar

4. Bekijk de volgende geordende dataset:

`0 1 3 3 4 5 5 5 7 12`

Welke van onderstaande boxplots is de boxplot behorende bij deze dataset?



5. Van een dataset zijn enkele waarden van de empirische verdelingsfunctie (*empirical distribution function*) gegeven:

t	0	1	2	3	4	5	6
$F_n(t)$	0	0.20	0.30	0.55	0.70	0.90	1.0

Als we een histogram van deze dataset maken, wat is dan de hoogte op de cel (*bin*) $(2, 4]$?

a. 0.20 **b.** 0.30 **c.** 0.50 **d.** 0.40 **e.** 0.35 **f.** 0.25

6. Voor een bivariate dataset $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, bestuderen we het lineaire regressiemodel

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

waarbij de U_i onafhankelijk en $N(0, \sigma^2)$ verdeeld zijn. Beschouw de volgende twee beweringen:

(A) In de kleinste-kwadratenmethode (*least squares method*) wordt de som van de residuen (*residuals*) geminimaliseerd.

(B) De toetsingsgrootheid (*test statistic*) $T_b = \frac{\hat{\beta}-\beta_0}{S_b}$ heeft onder $H_0 : \beta = \beta_0$ een t-verdeling met $n-2$ vrijheidsgraden (*degrees of freedom*)

Welke van deze beweringen is/zijn waar?

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

7. Laat T_1 en T_2 twee zuivere schatters (*unbiased estimators*) zijn voor een onbekende parameter θ. Er geldt dat $\mathrm{Var}(T_1) = 1$ en $\mathrm{Var}(T_2) = 3$. Bovendien zijn T_1 en T_2 onafhankelijk. Bekijk nu de schatter $T_3 = \alpha T_1 + (1-\alpha)T_2$, waarbij $\alpha \in [0, 1]$.

Voor welke waarde van α heeft T_3 de kleinste gemiddelde kwadratische fout (*mean squared error*)?

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

8. Laat $X_1, \ldots, X_n$ een aselecte steekproef (*random sample*) zijn uit een $Pois(\mu)$-verdeling. Bekijk de volgende twee schatters (*estimators*) voor $e^{-\mu} = P(X_i = 0)$:

$$S = \frac{\#\{X_i = 0\}}{n} \quad \text{en} \quad T = e^{-\bar{X}_n}.$$

Hierbij gebruiken we de notatie $\#A$ voor het aantal elementen in de verzameling A.

Definieer de bias van een schatter T voor een parameter θ als $\text{Bias}[T] = \text{E}[T] - \theta$.

Dan geldt:

a. S en T zijn beide zuiver voor p
b. S is zuiver en T heeft een negatieve bias
c. T is zuiver en S heeft een negatieve bias
d. S is zuiver en T heeft een positieve bias
e. T is zuiver en S heeft een positieve bias
f. S en T zijn beide onzuiver

9. Stel een meting X van de snelheid van een passerende auto op de snelweg wordt gemodelleerd als een normaal verdeelde stochast met parameters $\mu = v$ (in km/u) en $\sigma^2 = 25$, waarbij v de exacte snelheid is van de auto. Om te toetsen of een automobilist harder rijdt dan de toegestane 130 km/u wordt getoetst H_0: $v = 130$ tegen H_1: $v > 130$. Als toetsingsgrootheid wordt de meting X gebruikt.
Stel verder dat de nulhypothese (*null hypothesis*) wordt verworpen (en dus: de automobilist krijgt een boete) als de gemeten snelheid 136 km/u of hoger is.

Wat geldt dan voor P_{II}, de kans op een fout van de tweede soort (*error of type II*), voor een automobilist die 132 km/u rijdt?

a. $P_{II} = 0.8888$	**b.** $P_{II} = 0.7881$	**c.** $P_{II} = 0.4364$
d. $P_{II} = 0.1112$	**e.** $P_{II} = 0.5636$	**f.** $P_{II} = 0.2119$

Deeltentamen II Kansrekening en Statistiek EE1M31 13 april 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. Om de slaagkans van een experiment te onderzoeken wordt door honderd proefpersonen een serie experimenten uitgevoerd tot ze voor het eerst succes hebben. Neem aan dat het aantal pogingen N dat iemand nodig heeft een $Geo(p)$-verdeling heeft. We willen p schatten met de maximum likelihood methode. De gegevens:

aantal pogingen	1	2	3	≥ 4
frequentie	40	30	10	20

(a) (1.5 punten) Toon aan dat de likelihoodfunctie voor p gegeven wordt door $L(p) = C \cdot p^{80}(1-p)^{110}$, waarbij C een constante is die het aantal volgordes aanggeeft.

(b) (1.5 punten) Bepaal de meest aannemelijke schatting (*maximum likelihood estimate*) voor p.

2. In een fabriek worden zakken gevuld met (aardappel)chips. Het is gegeven dat de hoeveelheid chips die vulmachine Apollo in een zak stopt een normale verdeling heeft met onbekende verwachting μ_A (gram) en bekende variantie $\sigma_A^2 = 16$. Voor de interne kwaliteitsbewaking neemt men wekelijks een aselecte steekproef met omvang 25 van zakken die door machine Apollo gevuld zijn.

   Het steekproefgemiddelde (*sample mean*) van de dataset van vorige week is $\bar{x}_{25} = 148.22$ gram en de steekproefstandaarddeviatie (*sample standard deviation*) is $s_{25} = 3.62$ gram.

   (a) (1.5 punten) Bepaal een tweezijdig 90%-betrouwbaarheidsinterval voor μ_A.

De fabriek heeft sinds kort ook een andere vulmachine, deze heet Bello. Voor deze machine geldt dat de hoeveelheid chips die in een zak gaat een normale verdeling hebben met verwachting μ_B gram, maar in dit geval is de variantie σ_B^2 nog onbekend. De fabrikant claimt 150 gram chips in de zakken te stoppen.

Om machine Bello te testen werd vorige week een steekproef met omvang 25 genomen. Het steekproefgemiddelde van deze dataset is $\bar{x}_{25} = 148.47$ gram en de steekproefstandaarddeviatie is $s_{25} = 4.25$ gram.

   (b) (2 punten) Formuleer de relevante hypothesen en toetsingsgrootheid (met verdeling!) en onderzoek vervolgens m.b.v. een geschikte statistische toets of machine Bello te weinig chips in een zak stopt bij een significantieniveau van $\alpha = 0.05$.

3. Laat de stochast X de wachttijd zijn tot de eerstvolgende student naar het toilet moet. We nemen aan dat X een $Exp(\lambda)$-verdeling heeft met onbekende parameter λ. We willen $H_0 : \lambda = 0.2$ toetsen tegen $H_1 : \lambda < 0.2$, waarbij we X als toetsingsgrootheid gebruiken.

   (a) (1.5 punten) Bepaal het kritieke gebied voor deze toets behorende bij significantieniveau $\alpha = 0.05$.

   (b) (1 punt) Stel dat de waargenomen waarde $x = 10$ is. Bereken de p-waarde voor deze realisatie.

Extra ruimte

Antwoorden:

1 **a.**

2 **e.**

3 **b.**

4 **c.**

5 **a.**

6 **c.**

7 f.

8 d.

9 b.

b $\hat{p} = 80/190$

a $(146.90, 149.54)$

b $H_0 : \mu_B = 150, H_1 : \mu_B < 150$, toetsingsgrootheid is $T = \frac{\bar{X}-\mu_B}{S_n/\sqrt{n}} \sim t(24)$ onder H_0. $t = -1.80$, kritieke waarde is $t_{24,0.95} = -1.711$ dus we verwerpen H_0.

a $K = [14.98, \infty)$.

b p-waarde=0.135