

Tentamen Kansrekening en Statistiek TI2216M
3 november 2014, 14.00–17.00 uur

Het tentamen bestaat uit 15 meerkeuzevragen en 3 open vragen. Een formuleblad wordt uitgedeeld.
S.v.p. na afloop weer inleveren. Normering: MC deel 2 punten per vraag. Open vragen deel: 7 + 5 + 5.
Eindcijfer = (totaal aantal punten + 3) gedeeld door 5.

1. Van de gebeurtenissen A en B is bekend dat $P(A) = 0.2$, $P(B) = 0.4$, $P(A \cup B) = 0.5$.
Bereken de voorwaardelijke kans $P(A|B)$.
a. 0.50 b. 0.05 c. 0.15 d. 0.20 e. 0.30 f. 0.35 g. 0.40 h. 0.25
2. Voor een stochast X met dichtheidsfunctie $f(x) = \frac{1}{2} \sin x$ voor $x \in [0, \pi]$ (en 0 daarbuiten), heeft de verdelingsfunctie F voor $x = \frac{1}{3}\pi$ de waarde
a. 0 b. $\frac{1}{6}\pi$ c. $\frac{1}{3}\pi$ d. $\frac{1}{4}\sqrt{3}$ e. $1/3$ f. $1/2$ g. $1/4$ h. $1/6$
3. Stel de stochast X is uniform verdeeld op het interval $[0, 4]$. Dan is $E[X\sqrt{X}]$ gelijk aan
a. $\frac{10}{3}$ b. $\frac{8}{3}$ c. $\frac{32}{15}$ d. $\frac{16}{5}$ e. $\frac{16}{3}$ f. $\frac{4\sqrt{2}}{3}$ g. $\frac{4\sqrt{2}}{5}$ h. $\frac{4\sqrt{2}}{7}$
4. X is een $Exp(\lambda)$ verdeelde stochast.
Bereken $F(x)$, de verdelingsfunctie (probability distribution function) van X^2 .
a. $1 - e^{-\lambda^2 x}$ b. $1 - e^{-2\lambda x}$ c. $1 - e^{-\lambda^2 x^2}$ d. $1 - e^{-\sqrt{\lambda x}}$
e. $1 - e^{-\lambda\sqrt{x}}$ f. $(1 - e^{-\lambda x})^2$
5. Neem $n + 2$ onafhankelijke stochasten $U_0, U_1, \dots, U_{n+1} \sim \text{Unif}(0, 1)$.
Een uitkomsten die groter is dan zijn twee burens heet een top.
Voorbeeld: de rij $(U_0, U_1, \dots, U_6) = (0.243, 0.152, 0.671, 0.367, 0.412, 0.222, 0.981)$
heeft twee toppen: 0.671 en 0.412.
De kans dat de k -de uitkomst een top is, is voor $1 \leq k \leq n$ gelijk aan
a. $\frac{1}{3}$ b. $\frac{1}{4}$ c. $\frac{1}{6}$ d. $\frac{n}{2n+4}$ e. $\frac{n}{3n+1}$ f. $\frac{1}{9}$ g. $\frac{2}{5}$ h. $\frac{n-1}{2n+2}$
6. Hoe kun je met Matlab een trekking simuleren uit een verdeling met dichtheid
$$f(x) = \begin{cases} \frac{1}{6}, & \text{als } 0 \leq x \leq 2 \\ \frac{1}{3}, & \text{als } 2 \leq x \leq 4 \end{cases}$$

a. $U = \text{rand} < 1/6;$
 $X = 2*U + \text{rand};$
b. $U = \text{rand} < 1/3;$
 $X = 2*U + \text{rand};$
c. $U = \text{rand} < 2/3;$
 $X = 2*U + \text{rand};$
d. $U = \text{rand} < 1/6;$
 $X = 2*U + 2*\text{rand};$
e. $U = \text{rand} < 1/3;$
 $X = 2*U + 2*\text{rand};$
f. $U = \text{rand} < 2/3;$
 $X = 2*U + 2*\text{rand};$
7. Stel de levensduur (in uren) van een bepaald type lamp is exponentieel verdeeld met verwachting 700. Hoe groot is de kans dat de lamp langer dan 900 uur zal branden?
a. 0.142 b. 0.276 c. 0.319 d. 0.612 e. 0.448 f. 0.500 g. 0.576 h. 0.211
8. Stel X en Y zijn onafhankelijke stochasten met allebei een $Geo(1/3)$ verdeling. Bereken $P(S = 4)$, waarbij $S = X + Y$ de som is.
(N.B. Het is niet nodig om eerste een 'algemene formule' op te stellen.)
a. 0.15 b. 0.07 c. 0.85 d. 0.43 e. 0.28 f. 0.21 g. 0.79 h. 0.34

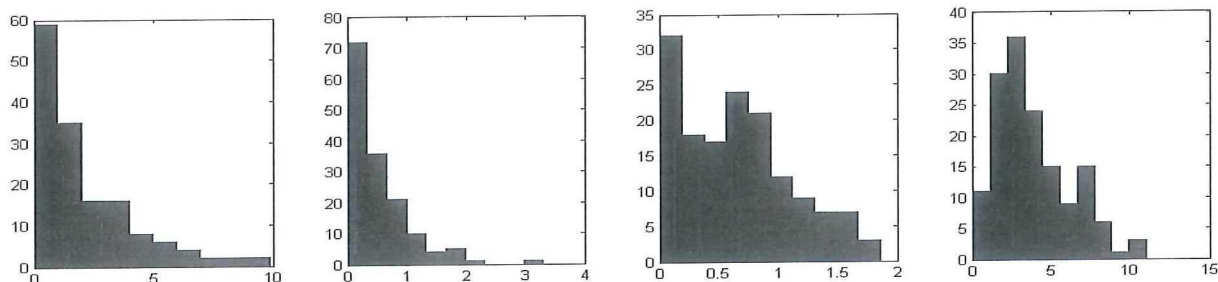
9. Stel $S = U_1 + U_2 + \dots + U_{300}$ is de som van 300 onafhankelijke stochasten U_i met uniforme verdelingen op het interval $(0,1)$. De centrale limietstelling geeft dan dat $P(150 \leq S \leq 160) \approx$
- a. 0.48 b. 0.14 c. 0.09 d. 0.25 e. 0.21 f. 0.32 g. 0.19 h. 0.06

10. Stel $X_1, \dots, X_n$ is een steekproef uit dezelfde verdeling van een stochast X . Definieer

$$Y = \frac{\text{aantal } X_i \text{ dat in } (-2,2) \text{ ligt}}{4n}.$$

De wet van de grote aantallen zegt dan dat Y voor grote n met grote kans ongeveer gelijk is aan

- a. $E[X]$ b. $\frac{E[X]}{4}$ c. $4P(-2 < X < 2)$ d. $4E[X]$
e. $P(-2 < X < 2)$ f. $\frac{P(-2 < X < 2)}{4}$
11. Welke van de onderstaande vier histogrammen (nummer ze a. t/m d.) past het best bij een steekproef (van grootte 150) uit een $\text{Exp}(2)$ verdeling?



12. Bereken de maximum-likelihoodschatting van de parameter α op grond van een dataset $x_1, x_2, \dots, x_n$ die geacht wordt uit een $\text{Pareto}(\alpha)$ -verdeling afkomstig te zijn. Voor uw gemak: $f(x) = \alpha x^{-\alpha+1}$, voor $x \geq 1$.

- a. $\sqrt[n]{x_1 + \dots + x_n}$ b. $\sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$ c. $\sqrt[n]{\ln x_1 + \dots + \ln x_n}$
d. $\frac{n}{x_1 + \dots + x_n}$ e. $\frac{n}{x_1 \cdot x_2 \cdot \dots \cdot x_n}$ f. $\frac{n}{\ln x_1 + \dots + \ln x_n}$

13. Een histogram van een dataset kan worden opgevat als een grafische representatie van de:

- a. verwachting b. variantie c. standaardafwijking
d. percentielen e. dichtheidsfunctie f. verdelingsfunctie

14. De *sample mean* $\bar{x}_n$ en de *sample median* Med_n , die in het Nederlands respectievelijk het gemiddelde en de mediaan heten, zijn allebei representatief voor het centrum van de dataset. Wat is het voordeel van Med_n boven $\bar{x}_n$?

- a. zuiverheid b. lage MSE c. ongevoelig voor outliers
d. maximum likelihood e. hoge variantie f. eenvoudig berekenbaar

15. Herinner je de tekst over het aantal Duitse tanks uit het boek. Bekend zijn de zeven serienummers

2121, 72613, 128172, 287851, 42210, 228095, 18219

die kunnen worden opgevat als zeven getallen die random zijn getrokken uit de verzameling van de getallen 1 tot en met N . Daarbij is N het (onbekende) aantal Duitse tanks. De hypothese $H_0: N = 500000$ wordt getoetst tegen de alternatieve hypothese $H_1: N < 500000$ via de toetsingsgrootheid $T = \max\{X_1, \dots, X_7\}$. Het significantieniveau is 0.05. De p -waarde van de toets wordt bij deze zeven serienummers gegeven door

- a. 0.23 b. 0.021 c. 0.171 d. 0.011 e. 0.046 f. 0.05 g. 0.14 h. 0.0014

Open vragen. 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 (of Engels).

1. Stel X_1, X_2, X_3, X_4, X_5 zijn onafhankelijke stochasten met $P(X_i = 1) = P(X_i = 0) = 1/2$.
Definieer

$$U = X_1 + X_2 + X_3, \quad V = X_2 + X_3 + X_4, \quad W = X_3 + X_4 + X_5$$

- a. Bereken $P(W = 0 | U = 2)$.
 - b. Bereken $E[U + V + W]$.
 - c. Bereken $\text{Cov}(U, V)$.
2. Om de nauwkeurigheid van een weeginstrument te onderzoeken wordt een serie metingen van een standaardgewicht van 10 gram uitgevoerd. De aannamen zijn dat de metingen normaal verdeeld zijn met (onbekende) verwachting μ , en standaardafwijking $\sigma = 0.0025$ gram.
- a. Stel het ijkgewicht wordt tien keer gewogen en het gemiddelde van de tien metingen is 10.016 gram. Geef een 98% betrouwbaarheidsinterval voor μ .
 - b. Hoeveel metingen moeten er (minimaal) gedaan worden om een 98% betrouwbaarheidsinterval te krijgen dat korter is dan 0.002 gram?
3. Het aantal patiënten dat dagelijks wordt opgenomen op de Intensive Care van het Reinier de Graafziekenhuis wordt gemodelleerd via een Poisson verdeling met parameter μ . Gemiddeld worden er 25 personen per dag opgenomen. Op vrijdag 13 mei 2011 gaat per ongeluk het luchtalarm af in Delft. Die dag worden 36 personen opgenomen op de Intensive Care. Is deze toename van 10 personen te wijten aan het luchtalarm?
- a. Formuleer de nulhypothese en de alternatieve hypothese.
 - b. Bereken de kans op een fout van de eerste soort (type I error) in de hypothesetoets voor onderdeel A. Gebruik dat de Poisson verdeling kan worden benaderd door de normale verdeling met verwachting μ en variantie μ .

