Tentamen toegepaste statistiek voor TA (WI1275TA)

1 juli 2013, 14:00-17:00

Afdeling Toegepaste Wiskunde, Faculteit Elektrotechniek, Wiskunde en Informatica

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. Rekenmachine toegestaan. Formuleblad/boek/aantekeningen niet toegestaan.

1. Stel dat een stochastische variabele X kansdichtheid f heeft, waarbij

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1/2 \\ 1/4 & 2 < x \leq 4 \\ 0 & \text{in alle andere gevallen} \end{cases} .$$

   (a) Bereken de verdelingsfunctie $F(x)$ van X voor alle $x \in \mathbb{R}$.

   (b) Bereken de verwachting van X.

2. Stel $X_1, X_2, \ldots, X_{100}$ zijn onafhankelijke Poisson-verdeelde stochastische variabelen met verwachting en variantie elk gelijk aan 2. De centrale limietstelling geeft

$$\mathrm{P}(X_1 + X_2 + \ldots + X_{100} \leq 210) \approx \mathrm{P}(Z \leq a),$$

   waarbij Z een standaard normaal verdeelde stochastische variabele is.

   Leid af wat de waarde van a moet zijn.

3. Een systeem bestaat uit onderdelen A en B. De kans dat onderdeel A kapot gaat is 0.1. De kans dat onderdeel B kapot gaat is 0.2. Als B kapot gaat, dan is de kans dat A kapot gaat 0.4.

   Wat is de kans dat precies één van de 2 onderdelen kapot gaat?

   Hint: Definieer de gebeurtenissen

$$\mathcal{A} = \{\text{onderdeel A gaat kapot}\}$$
$$\mathcal{B} = \{\text{onderdeel B gaat kapot}\}$$

   en druk de gevraagde kans uit met behulp van doorsnijdingen, verenigingen, complement, etc. van deze gebeurtenissen.

4. Stel dat X en Y stochastische variabelen zijn.

   (a) Druk $\mathrm{Cov}(X, X+Y)$ uit in $\mathrm{Var}(X)$ en $\mathrm{Cov}(X,Y)$.

   (b) Stel dat $Y = \alpha X$, met $\alpha \in \mathbb{R}$. Voor welke waarden van α zijn X en $X+Y$ negatief gecorreleerd? Je kunt bij het beantwoorden van dit onderdeel het resultaat van onderdeel (a) gebruiken, maar dit hoeft niet.

5. De stochasten X en Y hebben simultane kansdichtheid

$$f(x, y) = e^{-x-y} \quad \text{voor} \quad x, y \geq 0.$$

Beargumenteer of X en Y afhankelijk, dan wel onafhankelijk zijn.

6. In een laboratorium worden op de ochtend van dag 1 van een experiment 100 ratten ingespoten met een virus. Aan het begin van dag 2, 3 en 4 wordt gekeken hoeveel ratten er de vorige dag overleden zijn. Dit levert de volgende gegevens:

dag 1	dag 2	dag 3	overlevend na drie dagen
31	16	18	35

We werken met het volgende model: elke rat heeft kans p om op een dag te sterven, onafhankelijk van de andere ratten en onafhankelijk van hoeveel dagen hij/zij al overleefd heeft. We willen de kans p schatten op grond van de gegeven data.

Beargumenteer dat de likelihood $L(p)$ gegeven wordt door

$$L(p) = p^{65}(1-p)^{157}.$$

7. Een dataset wordt gemodelleerd als een realisatie van een steekproef $X_1, X_2, \ldots, X_n$ uit een $Exp(\lambda)$ verdeling, waarbij $\lambda > 0$ onbekend is. Zij μ de verwachting van de modelverdeling (dus $\mu = 1/\lambda$), en M_n het minimum van $X_1, X_2, \ldots, X_n$. Het blijkt (dit mag u aannemen) dat in dit geval M_n een $Exp(\lambda n)$ verdeling heeft. We beschouwen de schatter

$$T = c_n M_n,$$

waarbij c_n een constante is die van n afhangt.

Bereken voor welke waarden van c_n de schatter T zuiver is voor μ.

8. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, beschouwen we een lineair regressie model zonder intercept:

$$Y_i = \beta x_i + U_i \qquad \text{voor } i = 1, 2, \ldots, n,$$

waarbij $U_1, U_2, \ldots, U_n$ onafhankelijk zijn met verwachting 0 en variantie σ^2.

Bereken de kleinste-kwadraten-schatting voor β.

9. Gegeven is een getal t dat de realisatie is van een stochast T. Verder is er gegeven dat T een $N(\mu, 1)$-verdeling heeft. Om de nulhypothese $H_0 : \mu = 0$ te testen tegen de alternatieve hypothese $H_1 : \mu \neq 0$ gebruiken we T als toetsingsgrootheid. Er is besloten (dus dat mag u ook als gegeven beschouwen) om H_0 te verwerpen ten gunste van H_1 als $|t| \geq \frac{3}{2}$.

Bepaal met bovenstaande gegevens de kans op een fout van de eerste soort. Rond af op tenminste 2 decimalen.

Puntenverdeling bij open vragen:

Opgave:	1a	1b	2	3	4a	4b	5	6	7	8	9
Punten:	2	1	2	3	1	1	1	2	1	2	2