Tentamen Kanstat
WI1275TA
9 april 2010, 09 − 12 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Een formuleblad wordt bijgeleverd.
Een antwoord moet voorzien zijn van een berekening, toelichting en/of motivatie. Dit alles goed leesbaar en in goed Nederlands (of Engels).

---

1. Met een ingewikkelde test kan worden nagegaan of een patiënt een bepaald virus onder de leden heeft. In negentig procent van de gevallen zal de test dit aangeven, terwijl bij de test bij een op de twintig patiënten die het virus niet dragen ten onrechte een positieve uitslag geeft. Verder is gegeven dat twee procent van de bevolking drager is van het virus.

a. Voer in de gebeurtenissen V: een persoon draagt het virus, en T: een persoon krijgt een positieve test. Vertaal de (*drie*) in de inleiding genoemde getallen naar (voorwaardelijke) kansen betreffende T en V.

b. Bereken de kans dat een persoon die positief test werkelijk het virus draagt.

Sommige patiënten vragen een 'second opinion' aan: een tweede onafhankelijke test (dat wil zeggen: met dezelfde kansen op fouten in beide richtingen).

c. Bereken de kans dat een persoon die de ziekte niet heeft twee keer positief test.

d. Bereken de kans dat een persoon die twee keer positief test de ziekte niet heeft.

2. Stel, de tijdstippen waarop grote overstromingen plaatsvinden in een gebied kunnen gemodelleerd worden als een Poissonproces met intensiteit 1 per 15 jaar.

a. Wat is de kans dat er in een periode van 20 jaar twee of meer overstromingen plaatsvinden?

b. Wat is de kans dat er in twee achtereenvolgende perioden van tien jaar telkens één of meer overstromingen plaatsvinden?

c. Wat is de (voorwaardelijke) kans dat als er in een periode van twintig jaar twee overstromingen plaatsvinden deze allebei in de laatste tien jaar plaatsvinden?

3. Gegeven is de volgende dataset van 23 getallen :

66	70	69	68	67	72	73	70	57	63	70	78
67	53	67	75	70	81	76	79	75	76	58	

a. Als een histogram wordt gemaakt met cellen $(a, b]$ van lengte 5, beginnend vanaf 50, wat is dan de hoogte van het histogram op de eerste twee cellen?

b. Bereken de waarde van de empirische verdelingsfunctie in het punt 76.

c. Bereken de MAD van deze dataset.

4. Bladeren komen voor in vier verschillende typen: stijfselgroen, suikerwit, stijfselwit en suikergroen, welke wij afkorten met StG, SuW, StW respectievelijk SuG. Deze komen voor met de volgende kansen:

StG	SuW	StW	SuG
$\frac{1}{4}(\theta+2)$	$\frac{1}{4}\theta$	$\frac{1}{4}(1-\theta)$	$\frac{1}{4}(1-\theta)$

Stel een bioloog heeft n bladeren verzameld (en dit beschouwd kan worden als een random steekproef).

a. Laat N_1 het aantal bladeren in zo'n steekproef zijn van type StG. Welke verdeling heeft N_1?

b. Laat nu $T = 1 - \dfrac{4}{n}N_3$, met N_3 het aantal bladeren van type StW. Toon aan dat T een zuivere schatter is voor θ en bereken $\mathrm{Var}(T)$.

c. Geef een schatter S, gebaseerd op N_1 die zuiver is voor θ.

Stel nu dat een steekproef van $n = 420$ de volgende aantallen geeft:

type	aantal
StG	200
SuW	20
StW	110
SuG	90

c. Geef de likelihood $L(\theta)$ die hoort bij deze data.

5. Iemand doet metingen om het smeltpunt m van een nieuw materiaal te bepalen. Stel, de metingen kunnen gezien worden als onafhankelijke stochasten X_i met verwachting c en standaardafwijking $\sigma = 0.5\,(^oK)$.

a. Benader met de centrale limietstelling de kans dat het gemiddelde van 25 metingen een waarde oplevert die minder dan $0.1\,^oK$ afwijkt van het 'echte' smeltpunt.

b. Hoeveel metingen moeten er gedaan worden opdat de kans dat het gemiddelde minder dan $0.1\,^oK$ afwijkt van het 'echte' smeltpunt minstens $90\,\%$ is?

6. Stel U en V zijn onafhankelijke stochasten, U is uniform verdeeld op (0,4) en V is uniform verdeeld op (0,2). (Je kunt ook zeggen: (U,V) is uniform verdeeld op de rechthoek [0,4] $\times$ [0,2].)

a. Bereken $\mathrm{P}(U \geq 2V)$ en $\mathrm{P}(U + V \leq 2)$.

b. Bereken $\mathrm{E}[U + V]$ en $\mathrm{E}[UV]$

Laat Z het minimum zijn van U en V.

c. Welke waarden a kan Z aannemen?
Bereken voor elk van deze waarden $\mathrm{P}(Z \geq a)$.

d. Bereken $\mathrm{E}[Z]$.

7. Docenten van het Allwise college willen onderzoeken of leerlingen beter woordjes leren per computer dan met pen en papier. De 120 leerlingen van 2 vwo worden aselect in twee gelijke groepen gesplitst en krijgen een week de tijd om een lijst van vijfhonderd Duitse woordjes uit hun hoofd te leren. Tijdens een schriftelijke overhoring moeten ze twintig woordjes vertalen. De leerlingen die met de computer geleerd hebben behalen een gemiddelde van 14.35 met een standaardafwijking van 1.4; de andere groep komt op een gemiddelde van 13.77 met een standaardafwijking van 1.6.

We gaan nu de nulhypothese H_0: "er is geen verschil" toetsen tegen H_1: "er is wel verschil".
Als toetsingsgrootheid nemen we het verschil V van de gemiddelde scores.
De gerealiseerde waarde hiervan is dus $v = 14.35 - 13.87 = 0.48$.

a. Beredeneer dat V bij benadering een normale verdeling zal hebben.

Ga in het vervolg uit van een normale verdeling voor V.

b. Wat zegt de nulhypothese over de parameters μ_V en σ_V van V?

c. Geef op grond van de data een schatting voor σ_V.

d. Bereken met behulp van de schatting uit het vorige onderdeel de p-waarde bij de gevonden data. (Als je het vorige onderdeel niet hebt beantwoord: neem als schatting 0.18.)

e. Wordt de nulhypothese bij significantieniveau $5\,\%$ verworpen?