Tentamen Kansrekening en Statistiek (191530082) voor INF
woensdag 3 juli 2013, 8.45 - 11.45 uur

> Dit tentamen bestaat uit 6 opgaven. Separaat zijn het formuleblad en de tabellen toegevoegd. Alle antwoorden dienen te worden gemotiveerd. Een programmeerbare rekenmachine is niet toegestaan, wel een gewone wetenschappelijke rekenmachine.

1. Een recent Amerikaans bevolkingsonderzoek wees uit dat 65% van alle woningen in de VS wordt bewoond door de eigenaar van die woning; de overige woningen worden gehuurd. De kansverdeling van het aantal bewoners van een willekeurige woning wordt voor beide situaties afzonderlijk weergegeven in de volgende tabel:

	1	2	3 of meer
eigendom	0.22	0.36	0.42
huur	0.36	0.26	0.38

   Begin de beantwoording van de onderstaande vragen met het definiëren van relevante gebeurtenissen en geef aan welke kansformules u gebruikt.

   a. Bepaal de kans dat een willekeurige woning een huurwoning is met precies twee bewoners.

   b. Bepaal de kans dat een willekeurige woning drie of meer bewoners heeft.

   c. Bepaal de kans dat een woning met 3 of meer bewoners wordt gehuurd.

2. Het aantal fietsendiefstallen in een zekere plaats op een willekeurige dag geven we aan met X en het aantal aangiften van fietsendiefstallen op die dag met Y. Neem aan dat X een Poissonverdeling heeft met verwachting $E(X) = 4$ en dat van een diefstal met kans $\frac{1}{2}$ aangifte wordt gedaan.

   a. Geef de voorwaardelijke verdeling van Y, gegeven $X = n$, voor $n = 0, 1, \ldots$.

   b. Bepaal de kansverdeling, de verwachtingswaarde en de variantie van Y.

   c. Bepaal de verwachtingswaarde en de variantie van $Z = X - Y$, het aantal fietsendiefstallen waarvan geen aangifte wordt gedaan.

   d. Bepaal de covariantie en de correlatiecoëfficient van X en Y.

3. Neem aan dat het resultaat van een meting van het geluidsniveau in de cockpit van een straaljager kan worden gemodelleerd als een normaal verdeelde stochastische variabele met verwachting 80 dB en standaardafwijking 5 dB. De resultaten van 16 van dergelijke metingen worden gerepresenteerd door de stochastische variabelen $X_1, X_2, \ldots, X_{16}$.

   a. Bereken $P(X_1 > 83)$, de kans dat het resultaat van de eerste meting hoger is dan 83 dB.

   b. Bereken de kans dat elk van de 16 gemeten waarden *niet* hoger is dan 83 dB.

   c. Bereken de kans dat het gemiddelde resultaat van alle 16 metingen *hoger* is dan 83 dB.

   d. Nadere bestudering van de gegevens wijst uit dat een niet-symmetrische verdeling beter past bij de waargenomen waarden dan een normale verdeling. Leg uit waarom het resultaat van de berekening bij onderdeel c toch (bij benadering) correct is.

4. De duur van een willekeurig telefoongesprek kan goed worden gemodelleerd als een exponentiële variabele met een parameter die afhankelijk is van de persoon die belt. Zij X_D de duur van een willekeurig telefoongesprek van Daan en X_S de duur van een willekeurig telefoongesprek van Sanne en neem aan dat $E(X_D) = 3$ en $E(X_S) = 4$.

   a. Bereken $P(X_D > 3 \text{ en } X_S > 3)$.

   b. Leid (m.b.v. de convolutie-integraal) de kansdichtheid van $X_D + X_S$ af.

   c. Bereken $P(X_D < X_S)$.

   d. Stel dat Daan begint te bellen op een moment dat Sanne al één minuut in gesprek is. Wat is de kans dat Daan's gesprek als eerste is afgelopen? Waarom?

5. Een geneticus is geïnteresseerd in de fractie van de bevolking met een bepaalde bloedeigenschap. Hij onderzoekt daartoe het bloed van 2000 willekeurig gekozen personen en constateert dat 600 personen de bloedeigenschap hebben.

   a. Bereken een (benaderd) 95%-betrouwbaarheidsinterval voor de fractie van de bevolking met de bloedeigenschap.

b. Hoe groot had de steekproef moeten zijn om de breedte van het betrouwbaarheidsinterval kleiner dan 0.02 te krijgen.

6. Op 10 verschillende landbouwuniversiteiten in de VS zijn experimenten gedaan met twee nieuwe tarwevariëteiten. Beide variëteiten werden op elk van de 10 universiteiten ingezaaid op even grote percelen. De opbrengsten, in kilo's per perceel, zijn als volgt:

	universiteit									
variëteit	1	2	3	4	5	6	7	8	9	10
1	38	23	35	41	44	29	37	31	38	35
2	45	25	31	38	50	33	36	40	43	37

7 2 -4 -3 6 4 -1 9 5 2

Het verschil in opbrengst tussen de beide variëteiten op een willekeurige universiteit modelleren we als een normaal verdeelde stochastische variabele met onbekend gemiddelde μ en onbekende variantie σ^2.

a. Geef de gebruikelijke schattingen van μ en σ^2.

b. Bepaal een numeriek 95%-betrouwbaarheidsinterval voor μ.

c. Bepaal een 95%-voorspellingsinterval voor een meting van het verschil in opbrengst op een nieuwe universiteit.

d. Voordat de tests werden uitgevoerd bestond het vermoeden dat variëteit 2 gemiddeld meer zou opbrengen dan variëteit 1. Toets of dit vermoeden door de resultaten van de experimenten wordt bevestigd. Neem $\alpha = 0.05$.

normering:

1			2				3				4				5		6			
a	b	c	a	b	c	d	a	b	c	d	a	b	c	d	a	b	a	b	c	d
2	2	2	1	3	2	3	2	2	2	1	2	3	3	2	2	2	2	2	2	3

tentamencijfer: $\frac{\text{totaal}}{45} \times 9 + 1$ (afgerond)

Formuleblad Tentamen

Enkele formules Kansrekening en Statistiek (191530082)

Hypergeometrische verdeling: $\text{var}(X) = n\frac{R}{N}\left(1 - \frac{R}{N}\right)\frac{N-n}{N-1}$

Geometrische verdeling: $\text{var}(X) = \frac{1-p}{p^2}$

Uniforme verdeling: $\text{var}(X) = \frac{(b-a)^2}{12}$

Erlang verdeling: $f_X(x) = \frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!} \quad (x \geq 0)$

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\text{var}(X_i) + \sum\sum_{i \neq j}\text{cov}(X_i, X_j)$$

Betrouwbaarheidsintervallen:

$$\left(\bar{X} - c\frac{S}{\sqrt{n}}, \bar{X} + c\frac{S}{\sqrt{n}}\right) \quad \text{met} \quad P(T_{n-1} \leq c) = 1 - \frac{1}{2}\alpha$$

$$\left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right) \quad \text{met} \quad P(\chi^2_{n-1} \leq c_1) = \frac{1}{2}\alpha \quad \text{en} \quad P(\chi^2_{n-1} \geq c_2) = \frac{1}{2}\alpha$$

$$\left(\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \quad \text{met} \quad \Phi(c) = 1 - \frac{1}{2}\alpha$$

Voorspellingsinterval:

$$\left(\bar{X} - cS\sqrt{1+\frac{1}{n}}, \bar{X} + cS\sqrt{1+\frac{1}{n}}\right) \quad \text{met} \quad P(T_{n-1} \leq c) = 1 - \frac{1}{2}\alpha$$

Schatters:

verdeling	m.a. schatters	momentenschatter
Discreet		
Poissonverdeling met par. μ	$\hat{\mu} = \bar{X}$	$\bar{X}$
Geometrische verd. met par. p	$\hat{p} = 1/\bar{X}$	$1/\bar{X}$
$B(1,p)$-verdeling	$\hat{p} = \bar{X}$	$\bar{X}$
X is $B(n,p)$-verdeeld	$\hat{p} = \frac{X}{n}$	$\frac{X}{n}$
Continu		
Exponentiële verd. met par. λ	$\hat{\lambda} = 1/\bar{X}$	$1/\bar{X}$
Uniforme verdeling op $[0,\theta]$	$\hat{\theta} = \max\{X_1, X_2, \ldots, X_n\}$	$2\bar{X}$
Normale verdeling μ	$\hat{\mu} = \bar{X}$	$\bar{X}$
Normale verdeling σ^2	$\hat{\sigma^2} = \frac{1}{n}\sum(X_i - \bar{X})^2$	$\frac{1}{n}\sum X_i^2 - \bar{X}^2$