Tentamen Kansrekening en Statistiek (191530082) voor INF
donderdag 7 april 2011, 13.45 – 16.45 uur

Dit tentamen bestaat uit 6 opgaven. Separaat zijn het formuleblad en tabellen toegevoegd. Vermeld ook uw studentnummer en studierichting op uw werk. Een GR (programmeerbare rekenmachine) is **niet** toegestaan, wel een gewone wetenschappelijke rekenmachine.

1. Bij de dorpsloterij heeft Gait 5 van de 100 verkochte loten bemachtigd. Er zijn 10 prijzen door de plaatselijke middenstand beschikbaar gesteld. We noemen X het aantal prijzen dat Gait bij de verloting in de wacht sleept.
Bereken de kans dat Gait minstens 1 prijs wint als:
a. Op elk lot maximaal één prijs kan vallen
b. Op elk lot meer dan één prijs kan vallen (bij elke trekking doen alle 100 loten mee).
(je mag bij deze opgave volstaan met een rekenkundige uitdrukking als antwoord)

2. Gegeven is de volgende simultane kansverdeling , d.w.z. de kansen $P(X = x$ en $Y = y)$ van twee stochastische variabelen X en Y:

$x \backslash y$	0	1	2
0	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$
1	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$
2	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$

a. Bepaal de kansverdeling van X en bepaal $E(X)$ en var(X).
b. Bereken de correlatiecoëfficiënt van X en Y.
c. Zijn X en Y onderling onafhankelijk? Motiveer je antwoord.
d. Bepaal de voorwaardelijke verdeling van Y gegeven $X = 1$ en bereken $E(Y \mid X = 1)$.

3. Een (pseudo) random generator geeft ons een willekeurig getal X tussen 0 en 1
a. Laat zien dat $\frac{1}{k+1}$ het k-de moment $E(X^k)$ van X is ($k = 1, 2, 3, \ldots$).
b. Leid (met behulp van onderdeel a.) de variantie van X af.
c. Toon aan dat $Y = -2\ ln(X)$ exponentieel verdeeld is en bereken $E(Y)$

Zij $X_1, X_2, \ldots, X_{300}$ is een reeks van 300 o.o. random getallen tussen 0 en 1.
d. Bereken de kans dat de som van deze getallen groter is dan 158.

4. Bij waterpompen in ontwikkelingslanden is de cilinder het kritieke onderdeel: als de pomp kapot gaat moet (bijna) altijd de cilinder vervangen worden. De ervaring leert dat een cilinder **gemiddeld 5 jaar** meegaat, maar de variantie van de levensduur is groot. Een ontwikkelingsorganisatie plaatst daarom pompen met een reservecilinder.
Noem X de levensduur van de eerste cilinder die gebruikt wordt en Y van de tweede.
a. Bereken $P(X + Y > 12)$, de kans dat de pomp na 12 jaar nog in gebruik is, als de levensduur van een pomp normaal verdeeld is met $\sigma = 2.5$ jaar
b. Bereken $P(X + Y > 12)$, als de levensduur van een pomp exponentieel verdeeld is.

5. De UT wil marktonderzoek doen naar de belangstelling onder middelbare scholieren met een technisch profiel voor haar nieuwe *Science bachelor*. Met deze nieuwe opleiding hoopt men meer studenten te trekken dan met de opleidingen die binnen de nieuwe opleiding als major zouden worden ondergebracht. Uit vroeger onderzoek is bekend dat de huidige opleidingen tezamen kunnen rekenen op serieuze belangstelling onder 25% van de middelbare scholieren met het juiste profiel in Noord-Oost Nederland.

Na een informatiecampagne wordt een onderzoek uitgezet om de serieuze belangstelling onder de betreffende leerlingen te peilen. X is het aantal serieus belangstellenden in de steekproef van n leerlingen en p de fractie (percentage) serieus belangstellenden onder alle betreffende leerlingen (in NO-Nederland).

a. Welke verdeling heeft X ?
b. Geef de gebruikelijke schatter van p en laat zien dat deze zuiver is.
c. Druk de verwachte kwadratische fout van de schatter uit in n en p, en leg aan de hand daarvan uit dat een grotere steekproef een betere schatting oplevert.
d. Hoe groot moet je n kiezen zodat het 95%-betrouwbaarheidsinterval van p maximaal 0.10 (10%) breed is. Ga voor de berekening van n uit van een waarde van p (in de buurt) van 0.25.
e. Stel dat in een steekproef van $n = 192$ leerlingen er 60 serieuze belangstelling hadden. Toon aan dat de kans dat je in een steekproef van deze omvang 60 of meer serieus belangstellenden aantreft (afgerond) gelijk is aan 3%. Ga uit van de veronderstelling dat $p = 0.25$ en pas bij berekening van de kans continuïteitscorrectie toe.
f. Hoe heet de in e. berekende kans in statistische termen, als we met de steekproef de toets op H_0: $p \leq 0.25$ tegen H_1: $p > 0.25$ uitvoeren?

6. De automatiseringsafdeling van Stork Ketels heeft na klachten van gebruikers over traagheid van het computernetwerk besloten de responsietijden bij een bepaald type Cadcam-opdrachten te meten op willekeurige momenten (gedurende normale werktijden). Zo tracht men de wachttijden te kwantificeren. Bij een aselecte steekproef van 16 responsietijden werd een gemiddelde responsietijd van 15.10 seconden berekend en de steekproefstandaardafwijking bedroeg 5.06 sec.

a. Bepaal op grond daarvan een numeriek 90%-betrouwbaarheidsinterval voor de verwachte responsietijd (voor dergelijke opdrachten op willekeurige momenten).
b. Het management interpreteert de uitkomst van het onderzoek als volgt: zo'n 90% van de responsietijden liggen tussen de in onderdeel a bepaalde grenzen. Is dit een correcte interpretatie? Leg kort uit waarom (niet)?
c. Voorafgaand aan het onderzoek is gesteld dat een gemiddelde responsietijd van 13 seconden (of minder) acceptabel is. Is met deze steekproef aangetoond dat de verwachte responsietijd hoger is dan 13 seconden.
   Voer een toets uit met $\alpha = 0.05$ en geef daarbij
   1. de modelveronderstellingen,
   2. de hypothesen,
   3. de toetsingsgrootheid en zijn verdeling,
   4. De waargenomen waarde
   5. het kritieke gebied (of overschrijdingskans)
   6. de statistische en
   7. de verwoorde conclusie.

Normering:

1		2				3				4		5						6			**Totaal**
a	b	a	b	c	d	a	b	c	d	a	b	a	b	c	d	e	f	a	b	c	
2	2	2	2	2	2	2	2	3	3	3	3	2	2	2	3	3	1	3	1	5	**50**

Tentamencijfer = 1 + 9 × (aantal punten)/50
Eindcijfer = het gemiddelde van het tentamencijfer en de opdrachtcijfers, die hoger zijn dan het tentamencijfer. Elk opdrachtcijfer dat hoger is dan het tentamencijfer telt voor 5% mee.