Tentamen Kanstat
WI1275TA
28 maart 2008, 14 − 17 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Een formuleblad wordt bijgeleverd.
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. Een bak bevat twee rode en drie witte ballen. Er wordt een bal getrokken en teruggelegd met een extra bal van dezelfde kleur.

a. Bereken de kans dat de tweede bal rood is, gegeven dat de eerste bal rood is.

b. Bereken de kans dat er twee keer een bal met dezelfde kleur wordt getrokken

c. Als je alleen weet dat de tweede bal wit is, wat is dan de kans dat de eerste bal ook wit was?

2. Zes mensen staan op de begane grond voor de lift in een warenhuis met vijf verdiepingen. De verdiepingen waar zij heen willen zijn onafhankelijk van elkaar, en we nemen aan dat elk van de verdiepingen dezelfde 'kans' heeft om door een willekeurige bezoeker gekozen worden. We voeren drie stochasten in:

X: het aantal verschillende etages waar de zes personen heen willen.
Y: het aantal personen dat naar de tweede etage wil.
Z: de hoogste etage die wordt gekozen.

a. Welk van de drie heeft een binomiale verdeling, en wat zijn de parameters daarvan?

b. Bereken voor elk van de drie stochasten de kans op uitkomst 2.

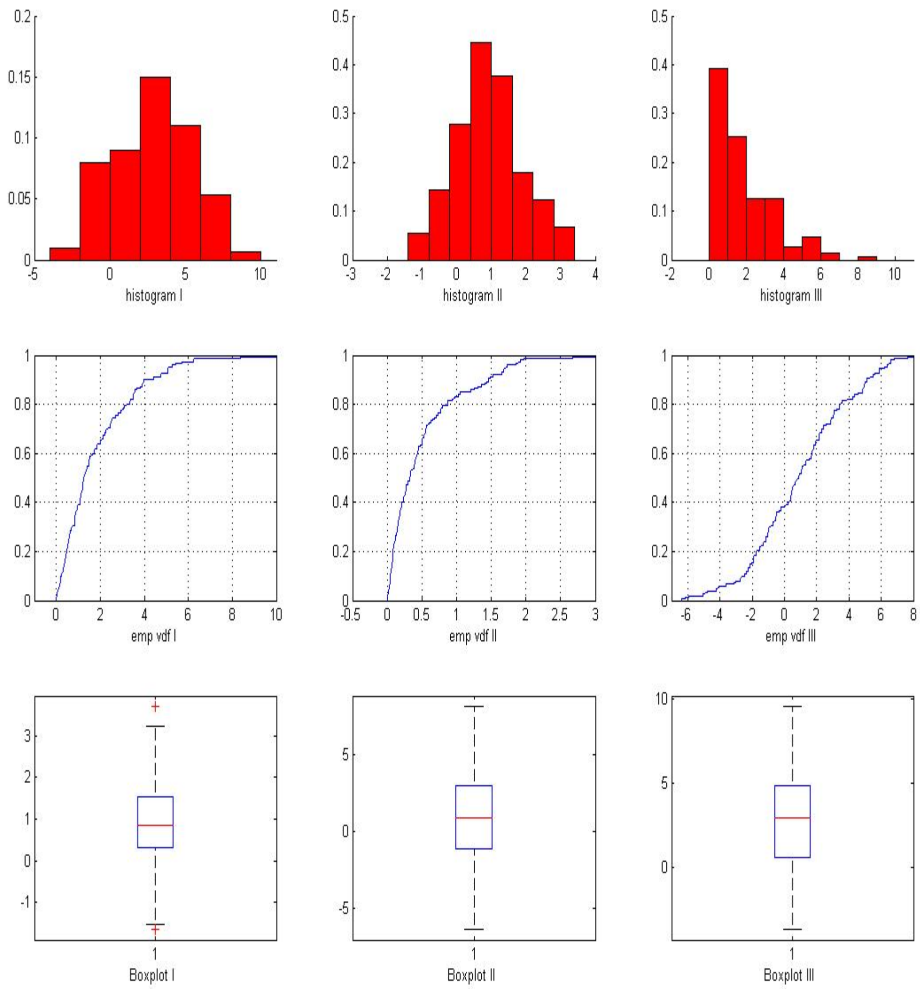
3. Een stochast X heeft verdelingsfunctie $F(x) = \begin{cases} 0\ , & \text{als } x \leq 0 \\ 1\ , & \text{als } x \geq 4 \\ x^2/16, & \text{als } 0 \leq x \leq 4 \end{cases}$
Bereken achtereenvolgens

a. $\mathrm{P}(1 \leq X \leq 3)$ en $\mathrm{P}(3.5 \leq X \leq 4.5)$.

b. Het eerste kwartiel $q_{0.25}$ van X.

c. De variantie van X.

4. Hieronder tref je negen plaatjes aan van trekkingen van 150 elementen uit de volgende vijf verdelingen:
$\mathcal{N}(1, 3^2)$, $\mathcal{N}(3, 3^2)$, $\mathcal{N}(1, 1)$, $\text{Exp}(2)$ en $\text{Exp}(1/2)$.
Geef voor elk van de negen aan welke verdeling het meest waarschijnlijk is **en waarom**.



5. Stel $X_1, X_2, \ldots, X_{12}$ zijn zuivere schatters voor een parameter c, met variantie 3, en $Y_1, Y_2, \ldots, Y_{15}$ zijn eveneens zuivere schatters voor c, maar met $\text{Var}(Y_i) = 6$.

a. Toon aan dat $T = \frac{2}{3}\bar{X}_{12} + \frac{1}{3}\bar{Y}_{15}$ een zuivere schatter is voor c.

b. Bereken de variantie van T.

6. 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 $\sigma = 0.025$ 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 0.02 gram?

c. Na een nieuwe ijking willen we een toets uitvoeren om te zien of deze goed heeft plaatsgevonden.
We nemen als nulhypothese $\mu = 10$ en als alternatieve hypothese $H_1 : \mu \neq 10$, en als significantieniveau $\alpha = 2\,\%$.
Er worden opnieuw tien (onafhankelijke) metingen gedaan. Deze hebben gemiddelde 9.986 gram.
Onderzoek of de nulhypothese wel of niet verworpen wordt.

7. Een geoloog wil onderzoeken of in een bepaald gebied de sporen van een zeker materiaal zich op 'willekeurige' locaties bevinden. Daartoe heeft hij het (rechthoekige) gebied in 100 gelijke deelgebieden verdeeld. De tabel bevat de aantallen sporen per deelgebied

# sporen	0	1	2	3	4	5	6	7
# gebieden	34	26	17	10	5	4	3	1

Dus: er zijn 34 deelgebieden zonder sporen, 26 deelgebieden met 1 spoor, enz.
Laat X de stochast zijn die beschrijft: het aantal sporen in een deelgebied.

a. Wat is het gemiddelde aantal sporen per deelgebied?

b. Leg uit waarom onder de aanname van 'willekeurige locaties' de stochast X (bij benadering) een Poissonverdeling zal hebben. Geef ook een *beargumenteerde* schatting van de paramater μ op grond van de data.

c. Geef een tabel waarin je de empirische kansen vergelijkt met de theoretische kansverdeling van X, met de in het vorige onderdeel bepaalde schatting van μ. (Als je geen schatting gevonden hebt, reken dan met de waarde $\mu = 1.25$.)

d. Geef de likelihoodfunctie en de loglikelihoodfunctie van μ op grond van de data.
Vereenvoudig de laatste zover mogelijk en bereken de maximum-likelihood schatting van μ.

Antwoorden open vragen

1a Voer in: R_i: i-de bal is rood, W_i: i-de bal is wit.
Als de eerste bal rood is, dan bevat de zak bij de tweede trekking drie rode en drie witte ballen. De kans dat er dan een rode bal getrokken worden, dat is $\mathrm{P}(R_2 \mid R_1)$, is dan 3/6.

1b Gevraagd: $\mathrm{P}((R_1 \cap R_2) \cup (W_1 \cap W_2))$.
De gevraagde kans is gelijk aan $\mathrm{P}(R_1 \cap R_2) + \mathrm{P}(W_1 \cap W_2)$.
$\mathrm{P}(R_1 \cap R_2) = \mathrm{P}(R_1)\,\mathrm{P}(R_2 \mid R_1) = \dfrac{2}{5} \cdot \dfrac{3}{6} = \dfrac{1}{5}$, en
$\mathrm{P}(W_1 \cap W_2) = \mathrm{P}(W_1)\,\mathrm{P}(W_2 \mid W_1) = \dfrac{3}{5} \cdot \dfrac{4}{6} = \dfrac{2}{5}$.
Het antwoord wordt dan $\dfrac{1}{5} + \dfrac{2}{5} = \dfrac{3}{5}$.

1c $\mathrm{P}(W_1 \mid W_2) = \dfrac{\mathrm{P}(W_1 \cap W_2)}{\mathrm{P}(W_2)} = \dfrac{2/5}{3/5} = \dfrac{2}{3}$,
want $\mathrm{P}(W_2) = \mathrm{P}(W_1)\,\mathrm{P}(W_2 \mid W_1) + \mathrm{P}(R_1)\,\mathrm{P}(W_2 \mid R_1) = \dfrac{2}{5} + \dfrac{2}{5} \cdot \dfrac{3}{6} = \dfrac{3}{5}$.

2a X en Z kunnen niet de waarde nul aannemen, dus hebben zeker geen binomiale verdeling.
Y heeft dat wel: het 'experiment' op een knop in de lift drukken, met als 'succes' drukken op de knop van de 2de etage (met kans 1 op 5), wordt zes keer herhaald. Het aantal 'successen', dus Y, heeft dan een Bin(6, 0.2) verdeling.

2b $\mathrm{P}(X = 2) = \mathrm{P}(X \leq 2) - \mathrm{P}(X = 1) = \mathrm{P}(X = 2) - (1/5)^5$. Resteert $\mathrm{P}(X \leq 2)$.
Het aantal manieren om twee etages te kiezen: $\binom{5}{2} = 10$.
De kans dat alle zes personen naar twee aangewezen etages gaan: $(2/5)^6$.
De kans dat de zes personen niet meer dan twee etages kiezen: $\binom{5}{2} \cdot (2/5)^6$.
Het antwoord: $\binom{5}{2} \cdot (2/5)^6 - (1/5)^5 = 127/5^5 = 127/3125 \approx 0.0406$.
$\mathrm{P}(Y = 2) = \binom{6}{2}(1/5)^2(4/5)^4 \approx 0.2458$.
Dan nog $\mathrm{P}(Z = 2) = \mathrm{P}(Z \leq 2) - \mathrm{P}(Z \leq 1) = (2/5)^6 - (1/5)^6 \approx 0.0040$.

3a $\mathrm{P}(1 \leq X \leq 3) = F(3) - F(1) = \frac{1}{16}(9 - 1) = 0.5$, en analoog
$\mathrm{P}(3.5 \leq X \leq 4.5) = F(4.5) - F(3.5) = 1 - 3.5^2/16 = 0.234$.

3b Gevraagd wordt de m waarvoor $\mathrm{P}(X \leq m) = \frac{1}{4}$.
Op te lossen: $F(m) = m^2/16 = \frac{1}{4}$. Duidelijk: $m = 2$.

3c De dichtheid van X: $f(x) = F'(x) = \frac{1}{8}x$, voor $0 \leq x \leq 4$.
$\mathrm{E}[X] = \int_0^4 x \cdot \frac{1}{8}x\,dx = \int_0^4 \frac{1}{8}x^2\,dx = ... = \frac{8}{3} = 2.666....$
Analoog $\mathrm{E}[X^2] = \int_0^4 \frac{1}{8}x^3\,dx = 8$.
Dan volgt: $\mathrm{Var}(X) = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = 8 - (8/3)^2 = 8/9 = 0.888....$

4 Het onderscheid normaal $\leftrightarrow$ exponentieel gaat via symmetrie: histogram III en empirische vdfs I en II horen bij exponentiële verdeling. (De boxplots passen trouwens ook niet vanwege de aanwezigheid van negatieve datapunten.)
Verder geldt voor een Exp(2) variabele X dat $\mathrm{P}(X \geq 4) = e^{-2 \cdot 4} \approx 0.003$, wat (verreweg) het best past bij emp vdf II, terwijl de andere twee beter passen bij een Exp(1/2) verdeling: als $X \sim \mathrm{Exp}(1/2)$, dan $\mathrm{P}(X \geq 4) \approx 0.14$.
De normale datasets zijn te onderscheiden via mediaan en spreiding/kwartielen. histogram I, boxplot III: $\mathcal{N}(3, 3^2)$, histogram II, boxplot I: $\mathcal{N}(1, 1)$, en de overgeblevenen: emp.vdf. III en boxplot II: $\mathcal{N}(1, 3^2)$.
Kort samengevat:

	I	II	II
histogram	$\mathcal{N}(3, 3^2)$	$\mathcal{N}(1, 1)$	Exp(1/2)
emp. vdf	Exp(1/2)	Exp(2)	$\mathcal{N}(1, 3^2)$
boxplot	$\mathcal{N}(1, 1)$	$\mathcal{N}(1, 3^2)$	$\mathcal{N}(3, 3^2)$

5a

$$\mathrm{E}[T] = \mathrm{E}\left[\frac{2}{3}\bar{X}_{12} + \frac{1}{3}\bar{Y}_{15}\right] = \frac{2}{3}\mathrm{E}\left[\bar{X}_{12}\right] + \frac{1}{3}\mathrm{E}\left[\bar{Y}_{15}\right] = \frac{2}{3}c + \frac{1}{3}c = c,$$

dus T is een zuivere schatter voor c.

5b

$$\mathrm{Var}(T) = \mathrm{Var}\big(\tfrac{2}{3}\bar{X}_{12}\big) + \mathrm{Var}\big(\tfrac{1}{3}\bar{Y}_{15}\big) = (\tfrac{2}{3})^2\mathrm{Var}\big(\bar{X}_{12}\big) + (\tfrac{1}{3})^2\mathrm{Var}\big(\bar{Y}_{15}\big) = \frac{4}{9}\frac{3}{12} + \frac{1}{9}\frac{6}{15} = \frac{7}{45}.$$

Bij het eerste en het derde =-teken is gebruikt dat de stochasten onafhankelijk zijn.

6a Het gemiddelde van 10 metingen heeft een $\mathcal{N}(\mu, \sigma^2/10)$ verdeling. Een 98 % betrouwbaarheidsinterval heeft de grenzen

$$\bar{x}_{10} \pm z_{0.01} \cdot \frac{0.025}{\sqrt{10}}, \text{ waarbij } z_{0.01} = 2.326 \text{ (afgerond: 2.33)}$$

dus dat wordt $[9.998, 10.034]$.

6b Het interval wordt bij n metingen kleiner dan 0.02 gram als

$$2z_{0.01}\frac{0.025}{\sqrt{n}} \leq 0.02$$

dus als

$$\sqrt{n} \geq \frac{2z_{0.01}0.025}{0.02} \approx 5.82$$

Oftewel, als $n \geq 5.82^2 = 33.8...$, dus, aangezien n geheel moet zijn: $n \geq 34$.

6c De gevonden waarde 9.986 ligt in de linkerstaart van de verdeling (van $\bar{X}_{10}$). We moeten kijken of de P-waarde, dus $\mathrm{P}\big(\bar{X}_{10} \leq 9.986\big)$ groter dan of kleiner dan 0.01 ($= \frac{1}{2}\alpha$) is. Onder de nulhypothese geldt:

$$\mathrm{P}\big(\bar{X}_{10} \leq 9.986\big) = \mathrm{P}\left(\frac{\bar{X}_{10} - 10}{0.025/\sqrt{10}} \leq \frac{9.986 - 10}{0.025/\sqrt{10}}\right) = \mathrm{P}(Z \leq -1.77) \approx 0.038 > 0.01$$

dus de nulhypothese wordt niet verworpen.

Alternatief: het 98 % betrouwbaarheidsinterval bij 9.986 wordt

$$[\bar{x}_{10} \pm z_{0.01} \cdot \frac{0.025}{\sqrt{10}}, \bar{x}_{10} \pm z_{0.01} \cdot \frac{0.025}{\sqrt{10}}] = [9.986 - 0.0184, 9.986 + 0.0184] = [9.9676, 10.0044]$$

en dat bevat de waarde 10, dus H_0 wordt niet verworpen.

7a $(34 \cdot 0 + 26 \cdot 1 + 17 \cdot 2 + \ldots + 1 \cdot 7)/100 = 1.55.$

7b Elk spoor heeft kans 1/100 om in een gegeven gebied terecht te komen. Bij een groot aantal, zeg n, sporen (hier 155) zal het aantal sporen in een gegeven deelgebied een binomiale verdeling hebben met parameters n en 0.01. Bij benadering is dat een Poissonvariable met parameter $\mu = 0.01n$. Hier $\mu = 1.55$.
Andere redenering: als je de deelgebieden op een strook naast elkaar legt dan is er onder aanname van onafhankelijkheid en 'constante intensiteit' sprake van een Poissonproces. Als de lengte-eenheid langs de strook gelijk is aan de lengte van een deelgebiedje dan heeft X een Poisson(μ) verdeling, met μ het verwachte aantal sporen in een gebiedje. Een schatting voor μ is het gemiddelde over de honderd gebiedjes, en dat is net berekend: 1.55.

7c

k	Pois(1.55)	rel. freq.
0	0.2122	0.34
1	0.3290	0.26
2	0.2550	0.17
3	0.1317	0.10
4	0.0510	0.05
5	0.0158	0.04
6	0.0041	0.03
7	0.0009	0.01

7d

$$L(\mu) = C \cdot (e^{-\mu})^{34} \left(\mu e^{-\mu}\right)^{26} \left(\frac{\mu^2}{2!} e^{-\mu}\right)^{17} \cdots \left(\frac{\mu^7}{7!} e^{-\mu}\right)^{1} = C' \mu^{155} e^{-100\mu}$$

$$\ell(\mu) = \ln\left(C' \mu^{155} e^{-100\mu}\right) = K + 155 \ln \mu - 100\mu$$

$$\ell'(\mu) = \frac{155}{\mu} - 100 = 0 \Leftrightarrow \mu = \hat{\mu} = \frac{155}{100} = 1.55.$$