Tentamen Kansrekening en Statistiek voor INF (191530082), maandag 29 juni 2015

Dit tentamen bestaat uit 7 opgaven. Het formuleblad en de tabellen zijn separaat toegevoegd.
Een gewone, niet programmeerbare rekenmachine is toegestaan (geen GR).
Studenten die de Kansrekening-toets al hebben gehaald hoeven alléén opgaven 6 en 7 te maken

1. Bij een productielijn voor laptops wordt een geautomatiseerde routine-eindtest uitgevoerd, om te controleren of de laptop voldoet aan de kwaliteitseisen. Uit de statistieken van de fabriek blijkt dat 97% van de geproduceerde laptops worden goedgekeurd. Bij een evaluatie van de eindtest is gebleken dat van de goedgekeurde laptops 99% daadwerkelijk voldoen aan de kwaliteitseisen en dat van de afgekeurde laptops 15% toch aan de kwaliteitseisen voldoen.
Bereken de volgende kansen door eerst relevante gebeurtenissen te definiëren, de gegeven en gevraagde kansen daarin uit te drukken en dan de rekenregels toe te passen:

a. de kans dat een laptop voldoet aan de eisen en
b. de kans dat een laptop die aan de eisen voldoet wordt goedgekeurd.

2. Uit een bak met 10 balletjes, genummerd 1 t/m 10, worden er lukraak en zonder terugleggen drie getrokken. Laat X het laagste en Y het hoogste nummer zijn op de drie getrokken balletjes. Bereken:

a. $P(Y = 5)$
b. $P(X = 2|Y = 5)$
c. $E(X|Y = 5)$

3. Een (pseudo) random generator geeft ons een willekeurig getal X tussen 0 en 1.

a. Laat zien dat het k-de moment $E(X^k) = \frac{1}{k+1}$ (voor $k = 1,2,3, ...$) is.
b. Gebruik a. om de variantie van X af te leiden.
c. Toon aan dat $Y = -3\ ln(X)$ exponentieel verdeeld is en bepaal $E(Y)$.

4. De bedieningstijden van 25 klanten bij een loket zijn ieder exponentieel verdeeld met een verwachting van 2 minuten. Beschouw de bedieningstijden als onderling onafhankelijke stochastische variabelen $X_1, X_2, ..., X_{25}$.

a. Bepaal de verwachting en de variantie van de totale bedieningstijd $\sum_{i=1}^{25} X_i$.
b. Bereken m.b.v. de Centrale Limiet Stelling een benadering voor de kans dat de totale bedieningstijd meer is dan een uur. Geef uw antwoord in tienden van procenten nauwkeurig.
c. Bereken de correlatiecoëfficiënt van de bedieningstijd van klant 1 (X_1) en de totale bedieningstijd $\sum_{i=1}^{25} X_i$.

5. Uit de populatie van 20-jarige mannen, van wie de gewichten $N(80, 100)$-verdeeld zijn, worden twee mannen willekeurig gekozen: hun gewichten noemen we X en Y.

a. Bereken $P(X > 90 \text{ en } Y > 90)$
b. Bereken $P(X + Y > 180)$

6. Een internet provider wil onderzoeken wat het effect is van een mogelijke wijziging in hun tariefsysteem. Daartoe wordt voor een groep van 54 personen, willekeurig gekozen uit het klantenbestand van de grootste concurrent, bepaald hoeveel van deze 54 personen bij introductie van het nieuwe tariefsysteem klant zouden worden. Zij p de fractie overstappers in de populatie en X het aantal toekomstige klanten bij het nieuwe tariefsysteem onder de groep van 54 willekeurig gekozen personen. Voor een positief besluit over invoering van het nieuwe tariefsysteem is nodig dat minstens de helft van de klanten van de concurrent willen overstappen. De vraag is of dit werkelijk het geval is, als minstens de helft van de mensen (dus 27 of meer) in de steekproef te kennen geeft te zullen overstappen. Om een antwoord op die vraag te geven gaan we er bij onderstaande vragen a. en b. vanuit dat **in werkelijkheid slechts 40%** van alle klanten van de concurrent zou overstappen.

- **a.** Geef de kansverdeling van X (type +parameters) en bepaal $E(X)$ en $var(X)$.
- **b.** Bereken (of benader) de kans $P(X \geq 27)$, de kans dat de helft van de personen in de steekproef zou overstappen.

Ga nu weer uit van een onbekende fractie p van overstappers.

- **c.** Bepaal een 90%-betrouwbaarheidsinterval voor de fractie p van overstappers in de populatie als in de steekproef precies 27 van de 54 zouden overstappen.
- **d.** We vatten het beschreven probleem nu op als een toets op $H_0: p = 0.4$ tegen $H_1: p > 0.4$ en we nemen waar dat $X = 27$: hoe heet dan de in b. bepaalde kans $P(X \geq 27)$ in statistische termen? En zou je de nulhypothese verwerpen als $\alpha = 5\%$?
  (Bij deze laatste vraag kun je $P(X \geq 27) \approx 9\%$ nemen als je b. niet hebt opgelost.)

7. Het IQ van een persoon, gekozen uit een grote homogene groep van personen, blijkt veelal goed te modelleren met de normale verdeling. Het is dan ook niet onredelijk te veronderstellen dat dit ook geldt voor alle technische studenten in Nederland, waarbij μ = "het verwachte IQ" en σ^2 = "de variantie van de IQ's" onbekend zijn. In een aselecte steekproef van 25 technische studenten bleek het gemiddelde IQ 120 en de steekproefvariantie 121 te zijn.

- **a.** Welke kansverdeling heeft het steekproefgemiddelde?
  Geef de verwachte kwadratische fout van deze schatter van μ.
- **b.** Bepaal een 95%-betrouwbaarheidsinterval voor het gemiddelde IQ van alle technische studenten in Nederland.
- **c.** Bepaal een 95%-betrouwbaarheidsinterval voor σ.
- **d.** Uit een eerder (omvangrijk) onderzoek is gebleken dat het gemiddelde IQ in de populatie van alle Nederlandse studenten 115 is.
  Kan uit de steekproef van technische studenten met onbetrouwbaarheidsdrempel 5% worden geconcludeerd dat het gemiddelde IQ onder de technische studenten hoger is dan 115?
  Voer bij dit onderdeel een toets uit door in ieder geval te geven:
  1. de hypothesen,
  2. de kansverdeling en de waarde van de toetsingsgrootheid
  3. de overschrijdingskans of het kritieke gebied en
  4. de conclusie m.b.t. het gemiddelde IQ die daaruit getrokken kan worden.

Normering:
cijfer = $1 + 9 \times \frac{\text{punten}}{53}$

1		**2**			**3**			**4**			**5**		**6**				**7**				**Totaal**
a	b	a	b	c	a	b	c	a	b	c	a	b	a	b	c	d	a	b	c	d	
3	2	2	2	2	3	2	3	2	2	2	2	2	3	4	3	2	2	3	3	4	**53**

Uitwerkingen Kansrekening en Statistiek voor INF (191530082) d.d. 29-06-2015

Opgave 1

a. Gebeurtenissen: V treedt op als de laptop "Voldoet aan eisen" en G als de laptop "Goedgekeurd"

$P(G) = 0.97$, $P(V|G) = 0.99$ en $P\left(V\middle|\overline{G}\right) = 0.15$

$P(V) = P(VG) + P(V\,\overline{G}) = P(V|G)P(G) + P\left(V\middle|\overline{G}\right)P(\overline{G}) = 0.99 \cdot 0.97 + 0.15 \cdot 0.03 = 0.9648$

b. $P(G|V) = \frac{P(GV)}{P(V)} = \frac{0.9603}{0.9648} \approx 0.9953$ (of eerst de regel van Bayes geven)

Opgave 2

a. $P(Y=5) = \frac{\binom{4}{2}}{\binom{10}{3}} = \frac{6}{120} = 0.05$ Schematisch:

nr. 1-4	nr. 5	nr. 6-10	Totaal
4	1	5	10
↓	↓	↓	↓
2	1	0	3

(of: $P(Y=5) = \frac{1}{10} \times \frac{4}{9} \times \frac{3}{8} \times 3$)

b. $P(X=2|Y=5) = \frac{P(X=2 \text{ en } Y=5)}{P(Y=5)} = \frac{\frac{2}{120}}{\frac{6}{120}} = \frac{1}{3}$

(als $Y=5$ zijn er slechts $\binom{4}{2} = 6$ mogelijkheden: (1,2,5), (1,3,5), (1,4,5), (2,3,5), (2,4,5) en (3,4,5).
Twee van die 6 mogelijkheden hebben 2 als laagste nummer: dan is $X=2$ en $Y=5$)

c. Analoog b vinden we: $P(X=1|Y=5) = \frac{3}{6}$ en $P(X=3|Y=5) = \frac{1}{6}$

Dus $E(X|Y=5) = \sum_x xP(X=x|Y=5) = 1 \times \frac{3}{6} + 2 \times \frac{2}{6} + 3 \times \frac{1}{6} = \frac{5}{3}$

Opgave 3

a. $E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx = \int_0^1 x^k \cdot 1dx = \left[\frac{1}{k+1}x^{k+1}\right]_{x=0}^{x=1} = \frac{1}{k+1}$, voor $k = 1,2,\ldots$,

b. Volgens a is $E(X) = \frac{1}{2}$ en $E(X^2) = \frac{1}{3}$. Dus $var(X) = E(X^2) - (EX)^2 = \frac{1}{2+1} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$

c. $F_Y(y) = P(Y \le y) = P(-3\ln(X) \le y) = P\left(X \ge e^{-\frac{1}{3}y}\right) = 1 - P\left(X \le e^{-\frac{1}{3}y}\right) = 1 - F_X\left(e^{-\frac{1}{3}y}\right)$

$f_Y(y) = \frac{d}{dy}F_Y(y) = +\frac{1}{3}e^{-\frac{1}{3}y} f_X\left(e^{-\frac{1}{3}y}\right)$

Omdat $f_X(x) = 1$ als $0 \le x \le 1$, is $f_X\left(e^{-\frac{1}{3}y}\right) = 1$ als $y \ge 0$, dus:

$f_Y(y) = \frac{1}{3}e^{-\frac{1}{3}y}$, voor $y \ge 0$, dus Y is exponentieel verdeeld met $\lambda = \frac{1}{3}$

Dus $E(Y) = \frac{1}{\lambda} = 3$

Opgave 4

a. Omdat $E(X_i) = 2 = \frac{1}{\lambda}$ en $var(X_i) = \frac{1}{\lambda^2} = 4$, geldt: $E\left(\sum_{i=1}^{25} X_i\right) = \sum_{i=1}^{25} E(X_i) = 25 \times 2 = 50$

en $var\left(\sum_{i=1}^{25} X_i\right) = \sum_{i=1}^{25} var(X_i) = 25 \times 4 = 100$.

b. Volgens de Centrale Limiet Stelling geldt dat $\sum_{i=1}^{25} X_i$ bij benadering N(25×2, 25×4)-verdeeld is:

$P\left(\sum_{i=1}^{25} X_i > 60\right) = P\left(\frac{\sum_{i=1}^{25} X_i - 50}{\sqrt{100}} > \frac{60-50}{\sqrt{100}}\right) \approx 1 - \Phi(1) = 1 - 0.8413 = 15.87\%$

c. $\varrho\left(X_1, \sum_{i=1}^{25} X_i\right) = \frac{cov\left(X_1,\ \sum_{i=1}^{25} X_i\right)}{\sqrt{varX_1}\sqrt{\sum_{i=1}^{25} X_i}}$

Er geldt: $cov\left(X_1, \sum_{i=1}^{25} X_i\right) = cov(X_1, X_1) + \cdots + cov(X_1, X_{25}) = var(X_1) + 0 + \cdots + 0 = 4$

en (zie a.) $var(X_1) = 4$ resp. $var\left(\sum_{i=1}^{25} X_i\right) = 100$.

Dus $\varrho\left(X_1, \sum_{i=1}^{25} X_i\right) = \frac{4}{\sqrt{4}\sqrt{100}} = \frac{1}{5}$

Opgave 5

a. $P(X > 90 \text{ en } Y > 90) \overset{\text{o.o.}}{=} P(X > 90) \cdot P(Y > 90) = \left(1 - \Phi\left(\frac{90-80}{10}\right)\right)^2 \approx 2.52\%$

b. $X + Y$ is N(80+80, 100+100)-verdeeld, dus

$P(X + Y > 180) = P\left(Z > \frac{180-160}{\sqrt{200}}\right) \approx 1 - \Phi(1.41)$ = 7.93%

Opgave 6

a. Zij X het aantal toekomstige klanten bij het nieuwe tariefsysteem onder de groep van n = 54 willekeurig gekozen klanten van de concurrent: $X \sim B(54, 0.40)$.

$E(X) = np = 54 \cdot 0.4 = 21.6$ en $var(X) = np(1-p) = 54 \cdot 0.4 \cdot 0.6 = 12.96$

b. Omdat $n > 25$, en ook $np = 21.6 > 5$ en $n(1-p) = 32.4 > 5$, kunnen we X normaal benaderen met $\mu = 21.6$ en $\sigma = \sqrt{var(X)} = \sqrt{12.96} = 3.6$

$P(X \geq 27) \overset{\text{cont. corr.}}{=} P(X \geq 26.5) \approx P\left(Z \geq \frac{26.5-21.6}{3.6}\right) \approx 1 - \Phi(1.36) = 1 - 0.9131 = 8.69\%$

c. 90%-BI(p) = $\left(\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = (0.5 - 0.112,\ 0.5 + 0.112) = (0.388, 0.612)$

$c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645\sqrt{\frac{0.5^2}{54}} \approx 0.112$, want $n = 54$, $\hat{p} = \frac{27}{54}$ en $\Phi(c) = 1 - \frac{1}{2}\alpha = 0.95$, zodat $c = 1.645$.

d. $P(X \geq 27 | p = 0.4) \approx 8.69\%$ is de (rechter) overschrijdingskans bij de waargenomen $X = 27$.

de overschrijdingskans ≈ 9% > α = 5%, dus H_0 verwerpen:

"Met onbetrouwbaarheidsdrempel 5% is de fractie overstappers niet aantoonbaar groter dan 40%."

Opgave 7

a. $\overline{X}$ is $N\left(\mu, \frac{\sigma^2}{n}\right)$-verdeeld (met $n = 25$)

Omdat $\overline{X}$ een zuivere schatter is, is de verwachte kwadratische fout: $E(\overline{X} - \mu)^2 = var(\overline{X}) = \frac{\sigma^2}{25}$

b. (Modelveronderstellingen onafhankelijkheid en normaliteit zijn gegeven)

95%-BI(μ)= = $(\overline{x} - c\frac{s}{\sqrt{n}},\ \overline{x} + c\frac{s}{\sqrt{n}}) \approx (115.5, 124.5)$

waarin n = 25, $\overline{x} = 120$, s^2 = 121 en c uit de t_{25-1}–tabel, zodat $P(T_{24} < c)$= 0.975 , dus c = 2.06

c. 95%-BI(σ)= $\left(\sqrt{\frac{(n-1)s^2}{c_2}}, \sqrt{\frac{(n-1)s^2}{c_1}}\right) \approx (8.6, 15.3)$. Hierin zijn:

n = 25, s^2 = 121 en c_1= 12.4 en c_2= 39.4 zodat $P(\chi^2_{24} \leq c_1) = 0.025$ en $P(\chi^2_{24} \leq c_2) = 0.975$.

d. 1. Toets $H_0 : \mu = 115$ tegen $H_1 : \mu > 115$ (met α = 0.05).

2. Toetsingsgrootheid $T = \frac{\overline{X} - 115}{S/\sqrt{25}}$ is onder H_0 t_{24}-verdeeld.

Waargenomen waarde: $t = \frac{120-115}{11/\sqrt{25}} \approx 2.27$

3. Kritieke Gebied: **verwerp $\boldsymbol{H_0}$, als $\boldsymbol{T \geq c}$**, met c = 1.71zodat $P(T \geq c | H_0) = 0.05$.

4. t = 2.27 ligt in het KG dus H_0 verwerpen: met onbetrouwbaarheid van 5% is statistisch aangetoond dat het gemiddeld IQ van technische studenten hoger is dan 115.

Alternatief: beslissing met de (rechter)overschrijdingskans $P(T_{24} \geq 2.27)$ < 2.5% < 5% = α, dus H_0 verwerpen. (zie de t_{24}-tabel: $P(T_{24} \geq 2.06)$ = 2.5%)