**Tentamen Statistische methoden**
**4052STAMEY**
**19 april 2013, 9:00–12:00**

**Studienummers:** Vult u alstublieft op het MC formulier uw *Delftse* studienummer in; en op het open vragen formulier graag beide, naar volgend voorbeeld: 1234567(D), 7654321(L).

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft. Normering: De meerkeuzevragen tellen voor één derde en de open vragen voor twee derde van het cijfer. Bij de open vragen telt elk (vraag)onderdeel even zwaar.

## Meerkeuzevragen

**Toelichting:** In het algemeen zijn niet altijd vijf van de zes alternatieven 100% fout, het juiste antwoord is het meest volledige antwoord. Maak op het bijgeleverde antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Doorstrepen van een fout antwoord heeft geen zin: u moet het òf uitgummen, òf verwijderen met correctievloeistof òf een nieuw formulier invullen. Vergeet niet uw studienummer in te vullen èn aan te strepen.

**1.** Drie vrienden gaan een avondje naar het casino. Ze spelen daar een spel met kans 1/4 om te winnen. Nadat ze alle drie dit spel drie keer hebben gespeeld, wat is de kans dat precies twee van de drie vrienden geen enkele keer het spel hebben gewonnen?

**a.** 0.052 **b.** 0.103 **c.** 0.206 **d.** 0.309 **e.** 0.412 **f.** 0.515

**2.** Een ijscoman staat met zijn kar langs de kant van de weg. De momenten waarop iemand stopt om een ijsje te kopen (eentje tegelijk) vormen een Poisson proces met intensiteit 7 per uur, waaruit volgt dat het totaal aantal klanten op een 8-urige dag een $Pois(56)$ verdeling heeft. Een ijsje kost 2 Euro. Noem de (bruto) inkomsten van een 8-urige werkdag $Y$. Dan is de variantie van $Y$ (ongeveer) gelijk aan:

**a.** 7.5 **b.** 15 **c.** 56 **d.** 112 **e.** 224 **f.** 6272

**3.** Dertien studenten stappen in een lift, die maximaal 1200 kg kan vervoeren. Stel dat we de dertien gewichten modelleren als onafhankelijke trekkingen uit een $N(90, 40)$ verdeling. Wat is de kans dat de lift overbelast is?

**a.** 0.09 **b.** 0.14 **c.** 0.24 **d.** 0.29 **e.** 0.36 **f.** 0.47

**4.** Zelfde contekst als het vorige probleem, maar nu hanteren we als verdeling de $U(75, 105)$. Wat is de kans dat de lichtste van de 13 minder weegt dan 80 kilo?

**a.** 0.09 **b.** 0.14 **c.** 0.31 **d.** 0.47 **e.** 0.79 **f.** 0.91

**5.** Een tentamen bestaat uit 30 driekeuzevragen. We nemen aan dat een student met kans $p$ het antwoord op een vraag weet (en dan dus goed antwoordt), en indien hij het antwoord niet weet gokt (met kans 1/3 op het goede antwoord). Een docent wil $p$ schatten op grond van $X$, het aantal correct beantwoorde vragen. Welke van de volgende schatters is een zuivere schatter voor $p$?

**a.** $X/30 - 1/3$ **b.** $X/45$ **c.** $X/30$
**d.** $(X + 10)/30$ **e.** $(X - 10)/20$ **f.** $(10 + \frac{2}{3}X)/30$

**6.** Een fabrikant van verf wenst de gemiddelde droogtijd van een nieuwe type muurverf te bepalen. Voor 12 testmuren van dezelfde oppervlakte vond hij een gemiddelde droogtijd van 66.3 minuten en een steekproefvariantie van 8.4. De 12 gemeten droogtijden vat men op als een realisatie van een steekproef uit een $N(\mu, \sigma^2)$ verdeling. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd $\mu$ wordt gegeven door

**a.** (61.944 , 70.655) **b.** (61.978 , 70.621) **c.** (64.730 , 67.869)
**d.** (64.742 , 67.857) **e.** (64.797 , 67.803) **f.** (64.809 , 67.790)

**7.** Stel dat $X_1, X_2, \ldots, X_n$ een steekproef is uit een normale verdeling met verwachting $\mu$ en variantie 1. We toetsen de nulhypothese $\mu = 0$ tegen de alternatieve hypothese $\mu > 0$. We nemen als toetsingsgrootheid $T = \bar{X}_n$ en als kritiek gebied $[1, \infty)$. De kans op een type 2 fout als $\mu = 1$, is gelijk aan:

**a.** 0.05 **b.** 0.14 **c.** 0.31 **d.** 0.36 **e.** 0.41 **f.** 0.50

**8.** Vervolg op voorgaande opgave. Stel dat voor een zekere dataset we de waarde $t = 1$ krijgen (dus $t$ is een realisatie van $T$). Als $n = 3$, dan wordt de $p$-waarde voor deze toets gegeven door

**a.** 0.001 **b.** 0.042 **c.** 0.081 **d.** 0.103 **e.** 0.159 **f.** 0.411

**9.** Stochasten $X$ en $Y$ zijn onafhankelijk, $X$ heeft een $N(0,1)$ verdeling en $Y$ een $N(1,3)$. Dan is $\mathrm{P}(-1 \leq X+Y \leq 2)$ ongeveer

**a.** 0.467 **b.** 0.533 **c.** 0.631 **d.** 0.718 **e.** 0.841 **f.** 0.933

**10.** Op een pot Albert Heijn pindakaas staat dat er 350 gram in zit. We vragen ons af of AH zó goed op de kleintjes let dat ze er wat minder instoppen. We nemen de proef op de som: we kopen 120 potten pindakaas, die we elk wegen, vervolgens leegeten, goed afwassen, en dan nogmaals wegen. Zo verkrijgen we—naast een lichte misselijkheid—een dataset bestaande uit 120 gewichtsbepalingen. We vinden als gemiddelde 348.87 en als standaardafwijking 5.10. De $p$-waarde van de passende toets is ongeveer:

**a.** 0.0008 **b.** 0.0078 **c.** 0.0397 **d.** 0.1038 **e.** 0.3446 **f.** 0.4400

## Open vragen

**Toelichting:** 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.

**1.** Een bron zendt onafhankelijk de symbolen 0 of 1 uit, met gelijke kans. Een ontvanger leest wat de bron uitzendt en maakt hierbij fouten: de kans dat de ontvanger een 1 leest als de bron een 0 heeft uitgezonden is 1/10; de kans dat de ontvanger een 0 leest als de bron een 1 heeft uitgezonden is 1/20.

**a.** Maak een tabel van de gezamenlijke kansverdeling van het uitgezonden symbool $X$ en het ontvangen symbool $Y$; voeg ook de marginale kansverdelingen toe aan de tabel.

**b.** Gegeven dat de ontvanger een 0 leest, wat is de kans dat een 1 werd uitgezonden?

**c.** Indien 10 symbolen worden uitgezonden wat is dan de kans dat precies 2 symbolen fout gelezen worden?

**2.** Voor de kansdichtheid van de stochastische variabele $Z$ geldt $f(z) = 3z^2/19$ voor $2 \leq z \leq 3$ en $f(z) = 0$ elders.

**a.** Bepaal de verdelingsfunctie van $Z$.

**b.** Laat zien hoe $Z$ gesimuleerd kan worden met behulp van een $U(0,1)$ stochast $U$; anders gezegd: bepaal de functie $g$ zodat $g(U)$ dezelfde verdeling heeft als $Z$.

**3.** Voor de gezamenlijke kansdichtheid van de stochastische variabelen $X$ en $Y$ geldt $f(x,y) = 2x^2/19 + y^2/19$ voor $2 \leq x \leq 3$ en $2 \leq y \leq 3$ en $f(x,y) = 0$ elders.

**a.** Bepaal $\mathrm{E}[X]$.

**b.** Zijn $X$ en $Y$ onafhankelijk? Licht uw antwoord toe.

**4.** We beschikken over twee onafhankelijke schatters, $S$ en $T$, die beide zuiver zijn voor een parameter $\theta$, met respectievelijke (bekende) varianties $\sigma_S^2$ en $\sigma_T^2$. Beschouw de schatter $V = aS + (1-a)T$, waarbij $a$ een willekeurige constante is.

**a.** Laat zien dat $V$ een zuivere schatter is voor $\theta$.

**b.** Is voor $a = 0.5$ de schatter $V$ altijd de beste van het drietal $S$, $T$, $V$? Licht uw antwoord duidelijk toe.

**5.** Door allerlei invloeden lukt een bepaalde conversieprocedure niet altijd; onder herhaalbare experimentele condities is de kans van slagen $p$. In drie laboratoria wordt de procedure net zolang herhaald totdat hij slaagt. Dit zijn de data:

- Lab 1: eerste succes bij de 3de poging;
- Lab 2: eerste succes bij de 3de poging;
- Lab 3: eerste succes bij de 4de poging.

**a.** Bepaal de meest aannemelijke (dwz, maximum likelihood) schatting voor $p$ op grond van de gegeven data.

**b.** Stel dat in Lab 3 ook de vierde poging mislukte en men het zo zat was dat men ermee ophield en naar de kroeg ging. Is toepassing van de maximum likelihood methode met behulp van de resulterende data dan nog mogelijk? Leg uit en als het antwoord ja is: geef de likelihoodfunctie voor dit geval.

## Antwoorden multiple choice:

**1 d.** De kans dat vriend nummer 1 geen een keer wint, is gelijk aan $p = (1 - \frac{1}{4})^3 = 0.422$. De kans dat precies twee van de drie vrienden geen een keer wint, is dus de binomiale kans

$$\mathrm{P}(X = 2) = \binom{3}{2} p^2 (1-p)^1 = 0.309,$$

waarbij $X$ een $Bin(3, p)$ verdeling heeft.

**2 e.** Als $N$ het aantal ijsklanten in 8 uur is, dan heeft $N$ een Poisson verdeling met verwachting $8 \cdot 7 = 56$. Verder geldt $Y = 2N$, dus $\mathrm{Var}(Y) = 4\mathrm{Var}(N) = 4 \cdot 56 = 224$.

**3 a.** $T = S_1 + \cdots + S_{13}$ heeft een normale verdeling, met parameters $\mathrm{E}[T] = 13 \cdot 90 = 1170$ en $\mathrm{Var}(T) = 13 \cdot 40 = 520$. Dus $\mathrm{P}(T > 1200) = \mathrm{P}\big(Z > (1200 - 1170)/\sqrt{520}\big) = \mathrm{P}(Z > 1.32) = 0.0934$.

**4 f.** Dat is 1 min de kans dat ze allemaal zwaarder zijn dan 80 kilo. Voor elk van de studenten is die kans $25/30 = 5/6$, dus $\mathrm{P}(\text{allemaal zwaarder dan 80 kilo}) = (5/6)^{13} \approx 0.0935$. De gevraagde kans is dus ongeveer 0.9065.

**5 e.** Er geldt $X \sim Bin(30, \theta)$ met $\theta = p + \frac{1}{3}(1-p) = (1+2p)/3$. Nu geldt dat $X/30$ zuiver is voor $\theta$. Dus $(X/10 - 1)/2 = (X - 10)/20$ is zuiver voor $p$.

**6 e.** Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd $\mu$ wordt gegeven door $\Big(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\Big)$. Hierbij is $t_{n-1,\alpha/2} = t_{11,0.05} = 1.796$ en $s_n = \sqrt{8.4} = 2.898$, zodat het 90% betrouwbaarheidsinterval wordt gegeven door $\Big(66.3 - 1.796\frac{2.898}{\sqrt{12}}, 66.3 + 1.796\frac{2.898}{\sqrt{12}}\Big) = (64.797, 67.803)$.

**7 f.** $P_{\mu=1}(T \notin K) = P_{\mu=1}(\bar{X}_n < 1) = 0.5$

**8 b.** $P_{\mu=0}(T > t) = P_{\mu=0}(\bar{X}_n > 1) = P_{\mu=0}(\sqrt{n}\bar{X}_n > \sqrt{n}) = P(Z > \sqrt{n}) \approx 0.0418$.

**9 b.** $S = X + Y$ heeft een $N(1, 4)$ verdeling, dus $Z = \frac{S-1}{2}$ is $N(0, 1)$, en we vinden

$$\mathrm{P}(0 \le S \le 3) = \mathrm{P}\left(-\frac{1}{2} \le \frac{S-1}{2} \le \frac{3-1}{2}\right) = \mathrm{P}\left(-\frac{1}{2} \le Z \le 1\right)$$

en

$$\mathrm{P}\left(-\frac{1}{2} \le Z \le 1\right) = \mathrm{P}(Z \le 1) - \mathrm{P}\left(Z \le -\frac{1}{2}\right) = 1 - 0.1587 - 0.3085 = 0.533$$

**10 b.** We doen de $t$-toets voor $H_0 : \mu = 350$ tegen $H_1 : \mu < 350$, waarbij $\mu$ de verwachte hoeveelheid in een pot is. We vinden: $t = (348.87 - 350)/(5.1/\sqrt{120}) = -2.42$. Omdat 120 groot is, kunnen we de $p$-waarde uit de tabel van de standaardnormale verdeling halen: $p = \mathrm{P}(Z < -2.42) = 0.0078$.

## Antwoorden open vragen:

**1a** Neem $X$ is uitzenden 0 of 1 en $Y$ ontvangen van 0 or 1. De tabel die de gezamenlijke kansen geeft is:

| | $X = a$ | | |
|---|---|---|---|
| Y=b | 0 | 1 | $P(Y = b)$ |
| 0 | 9/20 | 1/40 | 19/40 |
| 1 | 1/20 | 19/40 | 21/40 |
| $P(X = a)$ | 1/2 | 1/2 | 1 |

Men leest direct af dat $P(Y = 1) = 21/40 = 0.525$. Gezamenlijk: in body van de tabel; de marginale kansverdeling voor het verzenden is op onderste rij en de kansverdeling voor het ontvangen is op de rechter zijde.

**1b** Zie tabel:

$$P(X = 1|Y = 0) = \frac{1/40}{19/40} = \frac{1}{19}.$$

**1c** De kans op een fout is $1/20 + 1/40 = 3/40 = 0.075$. Dan heeft het aantal fouten $X$ in 10 signalen een $Bin(10, 0.075)$ verdeling. $P(X = 2) = \binom{10}{2} \times 0.075^2 \times 0.925^8 = 0.1357$.

**2a** $F(z) = (z^3 - 8)/19$ voor $2 \leq z \leq 3$; $F(z) = 0$ voor $z < 2$ en $F(z) = 1$ voor $z > 3$.

**2b** Zie §6.2. Los op $F(z) = u$ voor $0 \leq u \leq 1$. We vinden $z = \sqrt[3]{19u + 8}$, dus $\sqrt[3]{19U + 8}$ heeft dezelfde verdeling als $Z$.

**3a**

$$\mathrm{E}[X] = \int_2^3 \int_2^3 x f(x, y) dx dy = 65/38 + 5/6.$$

Alternatief is om eerst $f_X(x) = \int_2^3 f(x, y) dy = 2/19x^2 + 1/3$ en vervolgens

$$\mathrm{E}[X] = \int_2^3 x f_X(x) dx = 65/38 + 5/6.$$

**3b** Bij het eenvoudigste antwoord stel je dat bij onafhankelijkheid $f(x, y) = f_X(x) f_Y(y)$ (zie §9.4). De gegeven gezamenlijke dichtheid is nooit te schrijven als het product van een functie van alleen $x$ met een functie van alleen $y$.
Alternatieven: leidt af $f_X$ en $f_Y$ en laat dan het verschil zien; leidt af $F(x, y)$, $F_X(x)$ en $F_Y(y)$; of laat zien[1] dat $\mathrm{Cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y] \neq 0$.

**4a** Dan moeten we laten zien dat de verwachtingswaarde gelijk is aan $\theta$:

$$\mathrm{E}[V] = a\mathrm{E}[S] + (1 - a)\mathrm{E}[T] = a\theta + (1 - a)\theta = \theta.$$

Hier maken we expliciet gebruik van het feit dat de twee schatters zuiver zijn.

**4b** Laat $\sigma_S^2$ de kleinste variantie zijn en $\sigma_T^2 = k\sigma_S^2$ met $k \geq 1$. De variantie van $V$ is:

$$\sigma_V^2 = a^2\sigma_S^2 + (1 - a)^2\sigma_T^2 = (a^2 + k(1 - a)^2)\sigma_S^2,$$

waarbij gebruik is gemaakt van de onafhankelijkheid van de twee schatters. $\sigma_V^2$ is de kleinste variantie wanneer

$$a^2 + k(1 - a)^2 < 1 \Rightarrow k < \frac{1 + a}{1 - a}.$$

Dit betekent dat voor $a = 0.5$ de schatter $V$ een betere schatter is wanneer $k < 3$ en anders is het de schatter $S$ of $T$ die de kleinste variantie heeft.

**5a** Zie MIPS opgave 21.2 De likelihood is

$$L(p) = (1 - p)^2 p (1 - p)^2 p (1 - p)^3 p = (1 - p)^7 p^3.$$

Neem de logaritme

$$\ell(p) = 7 \log(1 - p) + 3 \log p$$

en vervolgens de afgeleide naar $p$. Gelijkstellen aan nul en verifiëren van het tekenoverzicht geeft $\hat{p} = 0.3$.

**5b** De kans op de data kan gewoon nog bepaald worden, rekening houdend met $\mathrm{P}(4 \text{ mislukkingen}) = (1 - p)^4$:

$$L(p) = (1 - p)^2 p (1 - p)^2 p (1 - p)^4 = (1 - p)^8 p^2.$$

[1] Wanneer de bewijsvoering werd begonnen met 'als $\mathrm{Cov}(X, Y) = 0$ dan zijn $X$ en $Y$' onafhankelijk, is dit als fout aangemerkt.