# Tentamen toegepaste statistiek voor TA (WI1275TA)

4 juli 2012, 14:00-17:00

*Afdeling Toegepaste Wiskunde, Faculteit Elektrotechniek, Wiskunde en Informatica*

---

**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.
Rekenmachine toegestaan. Formuleblad/boek/aantekeningen niet toegestaan.

---

1. De gebeurtenissen $A$ en $B$ zijn onafhankelijk en bovendien is bekend dat $\mathrm{P}(A \mid B) = 1/2$ en $\mathrm{P}(B \mid A) = 1/3$. Bereken $\mathrm{P}(A \cup B)$.

2. Gegeven is de stochastische variable $X$ met verdelingsfunctie

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}x & 0 \leq x < 1 \\ \frac{1}{2} + \frac{1}{4}x & 1 \leq x \leq 2 \\ 1 & x > 2 \end{cases}.$$

   (a) Welke waarden kan $X$ aannemen?

   (b) Bereken $\mathrm{P}\big(\frac{1}{2} \leq X \leq 1\frac{1}{2}\big)$.

   (c) Geef aan hoe je de stochastische variable $X$ kan simuleren met behulp van een Uniform$(0,1)$-verdeelde stochastische variable.
   *Hint: maak eerst een schets van de grafiek van $F$.*

3. Stel $X$ en $Y$ zijn onafhankelijke stochasten met kansverdelingen gegeven door

$$\mathrm{P}(X=0) = \mathrm{P}(X=1) = \frac{1}{2} \qquad \mathrm{P}(Y=0) = \frac{1}{3}, \quad \mathrm{P}(Y=1) = \frac{2}{3}.$$

   (a) Bereken de kansmassafunctie van $Z = X + Y$.

   (b) Toon aan dat $\mathrm{Cov}(2X, Z) = 2\mathrm{Var}(X)$.

4. Stel dat bits worden verzonden over een communicatiekanaal. Met kans 0.7 wordt een 1 verzonden (en dus met kans 0.3 een 0). Door storingen wordt het signaal niet altijd correct ontvangen. In 90% van de gevallen gaat het echter goed. Dat wil zeggen, als een 1 verzonden wordt, dan wordt ook een 1 ontvangen, en als een 0 verzonden wordt, dan wordt ook een 0 ontvangen.

   (a) Eén bit wordt verzonden. Stel dat we een 1 ontvangen, wat is dan de kans dat er een 0 verzonden is?
   *Hint: Definieer de gebeurtenissen*

$$A = \{\textit{een 1 wordt ontvangen}\}, \qquad B = \{\textit{een 1 wordt verzonden}\}.$$

   *en gebruik de regel van Bayes'.*

   (b) Stel nu dat we 1024 bits verzenden. Definieer

$$Y_i = \begin{cases} 1 & \text{als de } i\text{-de bit incorrect wordt ontvangen} \\ 0 & \text{als de } i\text{-de correct wordt ontvangen} \end{cases},$$

zodat

$$T = \sum_{i=1}^{1024} Y_i$$

het totaal aantal incorrect ontvangen bits geeft. Benader de kans dat minder dan 100 van de bits incorrect ontvangen worden met behulp van de centrale limietstelling.

5. We construeren een histogram met cellen [0,1], (1,3], (3,5], (5,8], (8,11], (11,14], en (14,18]. Gegeven zijn de waarden van de empirische verdelingsfunctie in de randen van de cellen:

| $t$ | 0 | 1 | 3 | 5 | 8 | 11 | 14 | 18 |
|---|---|---|---|---|---|---|---|---|
| $F_n(t)$ | 0 | 0.225 | 0.445 | 0.615 | 0.735 | 0.805 | 0.910 | 1.000 |

   Bereken de hoogte van het histogram op de cel $(8, 11]$.

6. Gegeven een dataset $x_1, \dots, x_{32}$ die we opvatten als realisatie van onafhankelijke stochastische variabelen $X_1, \dots, X_{32}$. Veronderstel dat alle $X_i$ dezelfde $N(\mu, 2)$-verdeling hebben. We willen de nulhypothese $H_0 : \mu = 1$ toetsen tegen de alternatieve hypothese $H_1 : \mu \neq 1$. Beschouw de toetsingsgrootheid $T = |\bar{X}_{32} - 1|$. Bereken het kritieke gebied voor deze toets bij significantieniveau $\alpha = 0.01$.

7. Stel een meting $X$ van de snelheid van een passerende auto op de snelweg wordt gemodelleerd als een normale variable met parameters $\mu = v$ (in km/u) en $\sigma = 5$, waarbij $v$ de exacte snelheid is van de auto. Om te toetsen of een automobilist harder rijdt dan de toegestane 120 km/u wordt getoetst $H_0$: $v = 120$ tegen $H_1$: $v > 120$.
   Stel verder dat de nulhypothese wordt verworpen (en dus: de automobilist krijgt een boete) als de gemeten snelheid 128 km/u of hoger is. Bereken de kans op een fout van de tweede soort voor een automobilist die 133 km/u rijdt.

8. Stel dat we binaire data $x_1, \dots, x_n$ hebben. Dus iedere $x_i$ is ofwel gelijk aan 1, ofwel gelijk aan 0. We vatten $x_1, \dots, x_n$ op als realisaties van onafhankelijke stochasten $X_1, \dots, X_n$ uit een Bernoulli verdeling met parameter $p$. Bereken de Mean Squared Error (MSE) van de schatter $S = \bar{X}_n$.

9. We gooien een onzuivere munt achtereenvolgens totdat we voor het eerst kop gooien. We herhalen dit experiment 3 keer (met dezelfde munt). De aldus verkregen data zijn:

   - eerste experiment: eerste keer kop bij de 3de worp,
   - tweede experiment: eerste keer kop bij de 3de worp,
   - derde experiment: eerste keer kop bij de 4de worp.

   Noem $p$ de kans op kop bij één keer gooien met de munt. Bereken de meest aannemelijke (maximum likelihood) schatting voor $p$ op grond van de gegeven data.

**Puntenverdeling bij open vragen:**

| Opgave: | 1 | 2a | 2b | 2c | 3a | 3b | 4a | 4b | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Punten: | 3 | 1 | 1 | 2 | 2 | 2 | 3 | 2 | 2 | 4 | 3 | 2 | 3 |

## Beknopte uitwerkingen

1. Door onafhankelijkheid geldt $\mathrm{P}(A \mid B) = \mathrm{P}(A) = \frac{1}{2}$ en $\mathrm{P}(B \mid A) = \mathrm{P}(B) = \frac{1}{3}$. Merk op dat $\mathrm{P}(A \cap B) = \mathrm{P}(A \mid B)\,\mathrm{P}(B) = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$. Dus $\mathrm{P}(A \cup B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}$.

2. (a) $X$ neemt waarden tussen 0 en 2 aan.
   (b) $\mathrm{P}\left(\frac{1}{2} \leq X \leq 1\frac{1}{2}\right) = F(1\frac{1}{2}) - F(\frac{1}{2}) = \frac{7}{8} - \frac{1}{4} = \frac{5}{8}$.
   (c) Als $U \in (0, 1/2)$, dan $X = 2U$. Als $U \in [1/2, 3/4)$, dan $X = 1$. Als $U \in (3/4, 1)$, dan $X = 4U - 2$.

3. (a) Merk op dat $Z$ waarden $0, 1, 2$ aanneemt.

$$\mathrm{P}(Z = 0) = \mathrm{P}(X = 0, Y = 0) = \tfrac{1}{6}.$$

   Net zo

$$\mathrm{P}(Z = 1) = \mathrm{P}(X = 0, Y = 1) + \mathrm{P}(X = 1, Y = 0) = \tfrac{1}{3} + \tfrac{1}{6} = \tfrac{1}{2}$$

   en

$$\mathrm{P}(Z = 2) = \mathrm{P}(X = 1, Y = 1) = \tfrac{1}{3}.$$

   (b)

$$\mathrm{Cov}(2X, Z) = \mathrm{Cov}(2X, X + Y) = 2\mathrm{Var}(X) + 2\mathrm{Cov}(X, Y) = 2\mathrm{Var}(X)\,,$$

   want $X$ en $Y$ zijn onafhankelijk.

4. (a) Gevraagd is $P(B^c|A)$. Gegeven zijn

$$\begin{aligned} &P(B) = 0.7 \\ &P(A|B) = 0.9 \qquad \text{(een 1 wordt correct ontvangen)} \\ &P(A^c|B^c) = 0.9 \qquad \text{(een 0 wordt correct ontvangen)} \end{aligned}$$

   Bayes' regel zegt dat

$$P(B^c|A) = \frac{P(A|B^c)P(B^c)}{P(A|B^c)P(B^c) + P(A|B)P(B)} = \frac{0.1 \cdot 0.3}{0.1 \cdot 0.3 + 0.9 \cdot 0.7} \approx 0.045.$$

   (b) De gevraagde kans is $\mathrm{P}(T < 100)$. De normale benadering voor de binomiale verdeling geeft vervolgens

$$\begin{aligned} \mathrm{P}(T < 100) &= \mathrm{P}\left(\frac{T - 1024 \cdot 0.1}{\sqrt{1024 \cdot 0.1(1 - 0.1)}} < \frac{100 - 1024 \cdot 0.1}{\sqrt{1024 \cdot 0.1(1 - 0.1)}}\right) \\ &\approx \mathrm{P}(Z < -0.25) \approx 0.4013, \end{aligned}$$

   waarbij Z een standaard normaal verdeelde random variabele is.

5. De fractie waarnemingen in cel $(8, 11]$ is gelijk aan $0.805 - 0.735 = 0.07$. De oppervlakte onder het histogram in deze cel is hieraan gelijk. De hoogte van cel moet dus zijn $0.07/\text{breedte cel} = 0.07/3 = 0.0233$.

6. De vorm van het kritieke gebied is $[c, \infty)$. We moeten $c$ nu bepalen. Dit gaat als volgt:

   - Kies $c$ zdd de kans op een type I fout $\leq \alpha$ is, maw

$$P_{\mu=1}(T \in K) \leq \alpha \quad \text{oftwel} \quad P_{\mu=1}(|\bar{X}_{32} - 1| > c) \leq 0.01.$$

- Kies vervolgens $K$ zo groot mogelijk, dwz $c$ zo klein mogelijk.

Nu geldt
$$P_{\mu=1}(|\bar{X}_{32} - 1| > c) = P_{\mu=1}(4|\bar{X}_{32} - 1| > 4c) = P(|Z| > 4c).$$
Hierbij is $Z \sim N(0,1)$. Blijkbaar moet dus gelden
$$4c \geq z_{0.005} \approx 2.57.$$
We concluderen dat $c = 2.57/4 \approx 0.64$.

7. Het kritieke gebied is gelijk aan $K = [128, \infty)$. De gevraagde kans is gelijk aan
$$\begin{aligned} \mathrm{P}(X \notin K \mid \mu = 133) &= \mathrm{P}(X < 128 \mid \mu = 133) \\ &= \mathrm{P}\left(\frac{X - 133}{5} \leq \frac{128 - 133}{5}\right) \\ &= \mathrm{P}(Z \leq -1) \approx 0.16. \end{aligned}$$

8. Merk eerst op dat de schatter zuiver is voor $p$. Dwz $\mathrm{E}\left[\bar{X}_n\right] = \mathrm{E}[X_1] = p$.. De MSE is dus gelijk aan $\mathrm{Var}\left(\bar{X}_n\right) = p(1-p)/n$.

9. 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 $\hat{p} = 0.3$.