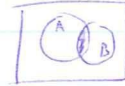


WI1295TA - STATISTIEK SAMENVATTING

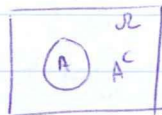
- uitkomst ruimte: Ω bv. voor dobbelsteen $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Deelruimte in de uitkomst ruimte bv. garantietijd op I-pod.

{ Door: wat hebben ze gemeen: $A \cap B$
Vereeniging: $A \cup B$

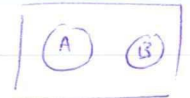


- $A \cup A^c = \Omega$

- $A \cap A^c = \emptyset$ (leeg)



$A \cap B = \emptyset$



- Disjunct: De doorsnede van A en B is leeg; A en B kunnen niet tegelijk optreden.

- Morgan's Law (p. 15)

- $(A \cup B)^c = A^c \cap B^c$

- $(A \cap B)^c = A^c \cup B^c$

- $(A^c)^c = A$

- Kans: $P(A)$ is een getal tussen 0 en 1.

↳ regels:

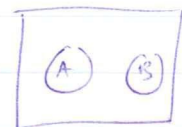
- $P(\Omega) = 1$

- $P(\emptyset) = 0$

- A en B zijn disjunct $\Rightarrow A \cap B = \emptyset$

$$P(A \cup B) = P(A) + P(B)$$

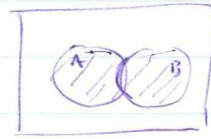
- $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$



- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

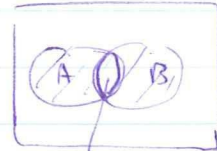
- Conditionele kans : $P(A|C) = \frac{P(A \cap C)}{P(C)}$

- $P(A) = P(A \cap B^c) + P(A \cap B)$



- $P(A \cup B) = P(A \cap B^c) + P(B)$

formule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

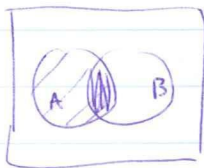


Dubbel, dus moet je nog één keer aftrekken.

- $P(A) = P(A \cap B) + P(A \cap B^c) = \frac{P(A \cap B)}{P(B)} \cdot P(B) + \frac{P(A \cap B^c)}{P(B^c)} \cdot P(B^c)$

$$= P(A|B)P(B) + P(A|B^c)P(B^c)$$

↳ wet van de totale waarschijnlijkheid.



- Regel van Bayes

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Onafhankelijkheid

A en B zijn onafhankelijk als $P(A|B) = P(A)$
(B heeft geen invloed op het resultaat van A)

verder $P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)} \left(= \frac{P(A) \cdot P(B)}{P(B)} = P(A) \right)$

(gevolg)

$\Rightarrow$ Als A en B onafhankelijk zijn dan $\frac{P(A \cap B)}{P(B)} = P(A)$

$$\Rightarrow P(A \cap B) = P(A)P(B)$$

Als de $P(B) \neq 0$ kan dit ook anders om.

A, B en C zijn onafhankelijk als:

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

$$\text{en } P(A \cap B) = P(A) \cdot P(B)$$

$$\text{en } P(A \cap C) = P(A) \cdot P(C)$$

$$\text{en } P(B \cap C) = P(B) \cdot P(C)$$

$\Rightarrow$ geldt ook voor
n onafh. stochastiken.

Discrete stochast

- Elke waarde kan uitgeschreven worden

$$F_x(a) = P(X \leq a) \quad a \in \mathbb{R}$$

bv:

a	2	3	...	12
$P_x(X=a)$	$1/36$	$2/36$	...	$1/36$

↑ kansmassafunctie

Som: 1

↓ Discrete verdelingen
zie formuleblad (p. 429)

- Bernoulli verdeling

$$X \sim \text{Ber}(p)$$

kan alleen de waarden 0 en 1 aannemen
(bv. goed of fout)

- Binomiale verdeling

$$X \sim \text{Bin}(n, p)$$

$$\text{NB. } \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Geometrische verdeling

$$X \sim \text{Geo}(p)$$

(bv.) trekken tot goed

- Poisson verdeling

$$X \sim \text{Pois}(\mu) \quad \mu > 0$$

Continue stochast

- kunnen alle waarden tussen 0 en 1 aannemen.

$$F_x = P(a \leq X \leq b) = \int_a^b f(x) dx$$

$$\text{en } f(x) = \frac{d}{dx} F_x(x)$$

met:

$f(x)$: kansmassafunctie

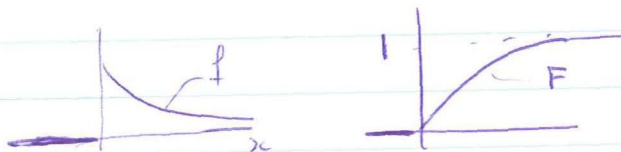
$F(x)$: verdelingsfunctie

↓ Continue verdelingen
zie formuleblad (p. 429)

- Uniforme verdeling $X \sim [a, b]$

- Exponentiële verdeling

$$X \sim \exp(\lambda)$$



- Normaalverdeling

$$\left. \begin{array}{l} X \sim N(\mu, \sigma^2) \\ Y \sim N(\mu, \sigma^2) \end{array} \right\} \Rightarrow X+Y \sim N(\mu+\mu, \sigma^2+\sigma^2)$$

gok normaal verdeeld.

- Standaard normaalverdeling

$$X \sim N(0, 1)$$

$$\text{standaardiseren: } X = \frac{X - \mu}{\sigma}$$

voor: $f(x)$, $F(x)$, $E[X]$, $\text{var}(X)$

$\Rightarrow$ zie formuleblad
of p. 429

Discrete stochast

- verwachtingswaarde

$$E[X] = \sum_i a_i \cdot P(X=a_i) \\ = \sum_i a_i \cdot P(a_i)$$

^{v.b.} a	-1	0	1
P(X=a)	1/5	2/5	2/5

$$E[X] = (-1) \cdot \frac{1}{5} + 0 \cdot \frac{2}{5} + 1 \cdot \frac{2}{5} = \frac{1}{5}$$

- The change of variable formula

$$E[Y] = \sum_x g(x) \cdot P(X=x)$$

- Wet van de Bemiddelde stochastische (WBS):

$$E[aX+b] = aE[X] + b \longleftrightarrow \text{idem}$$

- Variantie

$$\text{Var}(X) = E[X^2] - (E[X])^2 \longleftrightarrow \text{idem}$$

$$\text{NB. } E[X^2] > (E[X])^2$$

- $\text{Var}(bX \pm a) = b^2 \text{Var}(X)$
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
 $\hookrightarrow X \text{ en } Y \text{ onafh.}$

Continue stochast

- verwachtingswaarde:

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

en ook $\Rightarrow$

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f(x) dx$$

=

- The change of variable formula

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx$$

- Als X een continue stochast met kansdichtheid f_x en a, b en b getallen, dan is, als

$$Y = aX + b$$

de kansdichtheid van Y gelijk aan:

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{1}{a}(y-b)\right)$$

- Ongelijkheid van Jensen (§8.3) → formule blad.

Laat X een stochast zijn, en $g: \mathbb{R} \rightarrow \mathbb{R}$ een ~~convex~~ functie zijn dan geldt:
 L_1 (kennelje)

$$g(E[X]) \leq E[g(X)]$$

- Extremen
 - maximum: $F_X(a) = (F(a))^n$
 - minimum: $F_Y(a) = 1 - (1 - F(a))^n$

- Simultane kansverdeling

Discreet

onafh. $P(X=x, Y=y) = P(X=x)P(Y=y)$

net als $P(A \cap B) = P(X=x) \cdot P(Y=y)$

- continu

Simultane vdf in punt (x,y) = meng vdf naar X meng. vdf. Y

$$f(x,y) = f_X(x)f_Y(y)$$

- X en Y onafh. $E[XY] = E[X] \cdot E[Y]$

$E[X], E[Y]$ uitrekenen zo als v.b. vorige p.

$$E[XY] = \sum_x \sum_y xy (P(X=x, Y=y))$$

- Covariantie : hoe X en Y samenhangen

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \\ \text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$$

- correlatie
 - positief gecorreleerd: $\text{Cov}(X, Y) > 0$
meting X en die is boven zijn verwachte waarde dan ligt ook Y boven zijn verwachte waarde
 - negatief gecorreleerd: $\text{Cov}(X, Y) < 0$
 X boven de verwachte waarde dan ligt Y waarschijnlijk onder zijn verwachte waarde
 - ongecorreleerd: $\text{Cov}(X, Y) = 0$

- Correlatie coëfficiënt

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

$$\text{en } \rho(aX + b, cY + d) = \frac{ac \text{Cov}(X, Y)}{\sqrt{a^2 \text{Var}(X) c^2 \text{Var}(Y)}} = \underbrace{\frac{ac}{|a||c|}}_{\pm 1} \underbrace{\frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}}_{\rho(X, Y)}$$

$$\text{Er geldt } -1 \leq \rho(X, Y) \leq 1$$

- Chebyshev's ongelijkheid :

Laat Y een stochast zijn met $\mu = E[Y] \in \mathbb{R}$ en $\sigma^2 = \text{Var}(Y) \in \mathbb{R}$. Laat $a \geq 0$ een willekeurig getal zijn, dan geldt :

$$P(|Y - E[Y]| \geq a) \leq \frac{\text{Var}(Y)}{a^2}$$

- Zwakke wet van de grote aantallen:

Neem $Y = \bar{X}_n$, dan geldt er voor elke vaste $a \geq 0$

$$P(|\bar{X}_n - \mu| \geq a) \leq \frac{\frac{\sigma^2}{n}}{a^2} = \frac{\sigma^2}{n \cdot a^2} \xrightarrow{n \rightarrow \infty} 0$$

- Centrale limitstelling

Laat $X_1, X_2, \dots$ onafh. gelijk verdeelde stochasten zijn met $\mu = E[X_i] \in \mathbb{R}$ en $\sigma^2 = \text{Var}(X) \in \mathbb{R}$.

Dan geldt: $F_{Z_n}(z) = P(Z_n \leq z) \xrightarrow{n \rightarrow \infty} \int_{-\infty}^z \frac{1}{\sigma} e^{-\frac{1}{2}x^2} dx$

$$Z_n = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

↳ tabel: $P(Z \geq \dots)$

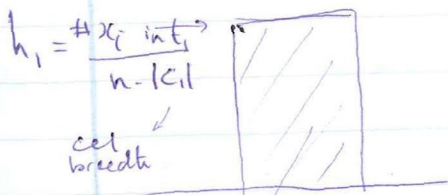
standaardnormaal
verdeelde stochast

- order of statistics: op volgorde van grootte zetten

$$X_{(1)} = 0 \leq X_{(2)} = 1$$

volgorde van grootte

- Histogram



De kans dat je hier terecht komt is de opp: $\text{opp} = h \times b = \frac{\#x_i - m_{C_i}}{n}$

- Empirische verdelingsfunctie: $F_n(x) = \frac{\# x_i \leq x}{n}$
 Logeft een idee van de echte F

- gemiddelde: $\bar{x}_n = \frac{x_1 + x_2 + \dots + x_n}{n}$

- mediaan: (midden van de dataset)

$$= \begin{cases} \frac{x_{(k)} + x_{(k+1)}}{2}, & \text{als } n=2k \quad (k \geq 1) \\ x_{(k+1)}, & n=2k+1 \quad (k \geq 0) \end{cases}$$

- kwantiel: $q_n(p) = x_{(k)} + \alpha(x_{(k+1)} - x_{(k)})$

$$k = \lfloor p(n+1) \rfloor$$

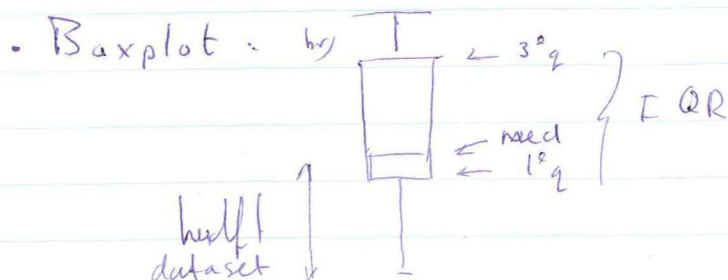
noemt getallen
 achter de komma α

alles achter de komma weg

bv. $\lfloor 1.7 \rfloor = 1$
 $\lfloor 2 \rfloor = 2$
 $\lfloor -\frac{1}{2} \rfloor = -1$

$$\alpha = p(n+1) - k$$

$$\begin{cases} p = \frac{1}{4} \rightarrow \text{eerste kwantiel} \\ p = \frac{3}{4} \rightarrow \text{derde kwantiel} \end{cases}$$



Data set: $x_1, x_2, \dots, x_n$

$\left. \begin{array}{c} \{ \\ \text{realisatie} \\ \} \end{array} \right\}$

Stochastic: $X_1, X_2, \dots, X_n \leftarrow$ Random sample

• Schatting $\swarrow$ $\nwarrow$ T (Belangrijk veel v.b maken)

• Zuivere schatter: de verwachtingswaarde is precies de
(EN: unbiased) gezochte waarde

• Onzuivere schatter: $\begin{cases} E[X] > \text{dan gezocht} \\ E[X] < \text{dan gezocht.} \end{cases}$
(EN: biased)

$\hookrightarrow$ de afwijking.

voor onzuivere schatter:

$$\text{Var}(T) \stackrel{\text{def}}{=} E[(T - E(T))^2] = E[T^2] - (E[T])^2$$

$$\text{mean squared error: } \text{MSE}(T) = E[(T - \theta)^2] = \text{Var}(T) + (E[T] - \theta)^2$$

als T een zuivere schatter is voor $\theta \Rightarrow \text{MSE}(T) = \text{Var}(T)$

• Schatting: Een waarde berekend met je schatter.

• kleinste kwadraten methode:

Zoek regressielijn door wolk gegevens (dataset):

$$Y_i = a + bX_i + U_i$$

a en b bepalen
(schatting)
↓
getal

$$\begin{cases} a = \bar{y}_n - b\bar{x}_n \\ b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \end{cases}$$

(schatte)
↓
methode om
getallen te vinden

$$\begin{cases} \hat{\alpha} = \bar{y}_n - \hat{\beta} \bar{x}_n \\ \hat{\beta} = \frac{\sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} \end{cases}$$

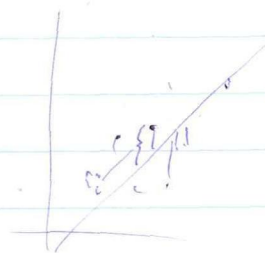
$$\begin{cases} E(\hat{\alpha}) = a \\ E(\hat{\beta}) = b \end{cases} \Rightarrow \text{juiste schattingen}$$

controle of gevonden regressielijn correct is.

• $(\bar{x}_n, \bar{y}_n)$ ligt op $y = \hat{a} + \hat{b}x$

$$\Rightarrow \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i) = 0$$

||
 $\hat{e}_i$: residu



• Betrouwbaarheidsinterval (p. 343)

def. Zeg we hebben een dataset $x_1, x_2, \dots, x_n$ die we opvatten als realisaties van onafh. en gelijk verdeelde stochasten $X_1, X_2, X_3, \dots, X_n$. Laat θ de parameter zijn waar we in geïnteresseerd zijn.

Laat γ een getal tussen 0 en 1 $\in, n$, en veronderstel dat er stochasten

$$L_n = g(X_1, X_2, \dots, X_n)$$

$$R_n = h(X_1, X_2, \dots, X_n)$$

bestaan z.d.d.

$$P(L_n < \theta < R_n) = \gamma$$

dan is (l, r) een $\gamma 100\%$ betrouwbaarheids interval

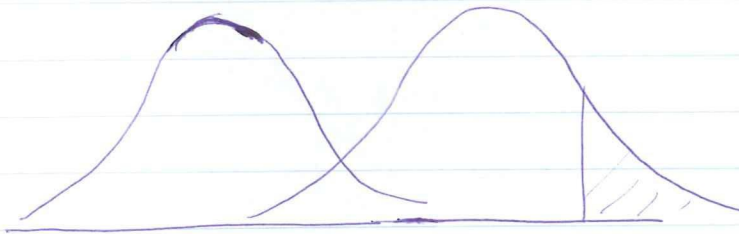
$\uparrow$ realisatie $\nwarrow$ realisatie
 L_n R_n

Toetsingstheorieën

schema (p. 378)

		True state of nature	
		H_0 is true	H_1 is true
our decision	H_0 verwerpen	Type I error	o.k.
	H_0 niet verwerpen	o.k.	Type II error

$$\begin{array}{l} \text{skel} \\ \text{op:} \end{array} \left\{ \begin{array}{l} H_0: \mu = 147.2 \\ H_1: \mu \left\{ \begin{array}{l} \geq \\ \leq \end{array} \right\} \text{op} \end{array} \right.$$



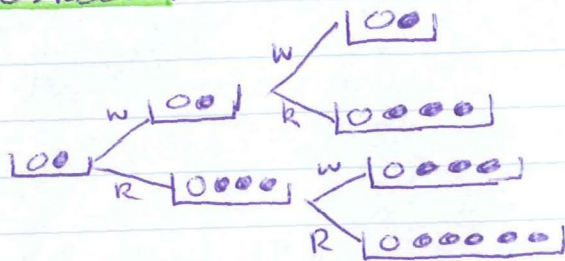
- Urausmodel (in vorm van vb; som 3.5 p.37)

Uraus: één rode en één witte bal

Trek willekeurig één bal; als deze wit is terug leggen.

Als hij rood is, terug leggen samen met 2 extra rode ballen.

Kansboom:



a) VRAAG: Wat is de kans dat je allebei de keren rood trekt?

• of: $P(RR) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$

• of: $P(R_1 R_2) = \frac{P(R_2 | R_1) \cdot P(R_1)}{P(R)} = P(R_2 | R_1) P(R_1) = \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{8}$

b) VRAAG: Wat is de kans dat de tweede bal rood is?

• of: $P(2^{de} \text{ rode bal}) = P(WR) + P(RR) = \frac{1}{2} \cdot \frac{1}{2} + \frac{3}{8} = \frac{5}{8}$

• of: $P(R_2) = P(R_2 | R_1) + P(R_2 | R_1^c)$
 $= \frac{3}{8} + \frac{P(R_2 | R_1^c) \cdot P(R^c)}{P(R^c)}$
 $= \frac{3}{8} + P(R_2 | R^c) \cdot P(R^c)$
 $= \frac{3}{8} + \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{8}$

c) VRAAG: Gegeven dat de tweede bal rood was; wat is de kans op de eerste bal is rood?

• $P(R_1 | R_2) = \frac{P(R_1 R_2)}{P(R_2)} = \frac{3/8}{5/8} = \frac{3}{8} \cdot \frac{8}{5} = \frac{3}{5}$