Inleiding

Kwalitatieve gegevens

- Ordinaal *tellen, ordenen*
- Nominaal *categorieën, niet-numerieke schaal*

Kwantitatieve gegevens

- Continue *(oneindig en niet aftelbaar)* vs discreet *(eindig en aftelbaar)*
- Intervalschaal *(zelf gekozen nulpunt)* vs ratioschaal *(natuurlijk nulpunt)*

Beschrijvende statistiek

Voorstellen gegeven

- Univariate kwalitatieve gegevens
  - Absolute en relatieve frequentie
  - Staafdiagram
  - Taartdiagram
  - Paretodiagram *ordenen volgens dalende frequentie*
- Univariate kwantitatieve gegevens
  - Stengel – en blad diagram
  - Histogram
  - Polygoon *hoogste punten van het histogram met elkaar verbinden*
- Bivariate gegevens
  - Kruistabel of meervoudig staafdiagram
  - Stapeldiagram
  - Scatterwolk (puntenwolk)

Cumulatieve frequenties

Enkel bij kwantitatieve gegevens

- Cumulatieve verdelingsfunctie
  - $F_n(x)$ = aantal waarnemingen/ n
  - = cumulatieve relatieve frequentie
  - Trapfunctie
- Kwantielfunctie
  - = inverse van de cumulatieve verdelingsfunctie
  - $Q_n(p) = \inf \{ x \in R \mid F_n(x) > p \}$ met $0 < p < 1$
  - $Q_n(p) = x_i$ met $i = \lceil np \rceil$ *(= afronden naar boven)*
  - Trapfunctie

Centrum kenmerken

- Gemiddelde
- $\bar{x} = \frac{1}{n} \sum x_i$
  - De gewogen som is dat sommige variabelen maar een aantal mogelijke waarden kunnen aannemen.
    - $\bar{x} = \frac{1}{n} \sum x_i \, m$
  - Som van de gecentreerde gegevens is steeds 0
    - $\sum(x_i - \bar{x}) = 0$
- Getrimd gemiddelde
  - Het gemiddelde van gegevens waarbij een aantal van de kleinste en grootste gegevens zijn weggelaten. Sorteren van klein naar groot!

- Mediaan
  - Med (x) = $x_{((n+1)/2)}$ als n oneven is
  - Med (x) = ($x_{(n/2)}$ + $x_{(n/2+1)}$)/2
- Modus
  - Is het element dat het meeste voorkomt, dus de hoogste frequentie heeft

Spreidingskenmerken

- Variantie en standaardafwijking
  - Var (x) = $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
  - Door het kwadraat vermijden we dat positieve en negatieve afwijkingen elkaar aflossen
  - Grote spreiding = grote variantie
  - Standaardsteekproefafwijking: s = $\sqrt{S^2}$
  - Eigenschap
    - Var (y) = a^2var(x)
  - Standaardisatie = z- score
    - $z_i = \frac{x_i - \bar{x}}{s}$
- Variatiecoëfficiënt
  - Standaardafwijking/ absolute waarde van steekproefgemiddelde
    - $c_v = \frac{s_x}{/\bar{x}/}$
  - Dimensie loze maat voor relatieve spreiding
- Spreidingsbreedte
  - Of bereik is het verschil tussen de grootste en kleinste waarneming
  - Gevoelig voor uitschieters

- Interkwartiel afstand
  - Is de lengte van het gebied rond de mediaan dat ongeveer de helft van de gegevens bevat en wordt gedefinieerd adhv kwartielen
  - Verdelen steekproef netjes in 4 delen
  - $Q_1 = Q_n(0,25)$, $Q_3 = Q_n(0,75)$ en de mediaan vormt het tweede kwantiel
  - IQR = $Q_3 - Q_1$

Boxplot

- Box is van 1ste kwartiel tem 3e kwartiel → bevat 50% van de gegevens
- Mediaan wordt weergegeven door horizontale lijn in de box
- Snorharen: de eerste reikt tot kleinste waarneming en de andere vanaf Q_3 tot de grootste waarneming. Deze kleinste en grootste waarnemers mogen geen

Empirische momenten

Zijn $\bar{x}$ en var verheven tot een bepaalde macht

- Empirische k-de orde moment
  - $m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$
- Empirische k-de orde **centrale** moment
  - $m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$
- Empirische scheefheid
  - $\hat{\gamma} = \frac{m_3'}{(m_2')^{3/2}}$
  - Hoe asymmetrisch is de steekproef
- Empirische kurtosis
  - $\hat{\kappa} = \frac{m_4'}{(m_2')^2}$
  - Hoe dik zijn de staarten van de steekproef

Verband tussen 2 variabelen

- Covariantie
  - Meet de lineaire associatie
  - $cov\,(x,y) = s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$
  - Hier kunnen de termen zowel positief als negatief zijn
  - Positieve associatie vinden we rechtsboven en linksonder
  - Negatieve associatie vinden we linksboven en rechtsonder
  - Positieve associatie lijdt tot een positieve covariantie
  - Eigenschap: $cov(a_1x + b_1, a_2x + b_2) = a_1a_2cov(x,y)$
- Correlatie
  - Helpt besluiten hoe sterk of zwak de associatie is
  - $r(x,y) = \frac{cov\,(x,y)}{\sqrt{var(x)}\sqrt{var\,(y)}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$
  - Eigenschap: $r(a_1x + b_1, a_2x + b_2) = teken(a_1a_2)r(x,y)$
- Lineaire regressie
  - Is om het verband tussen twee variabelen grafisch voor te stellen
  - Zie cursus pag 28 tem 30

Kansrekening

Kans experimenten en uitkomstenruimte

- Elke deelverzameling van de uitkomstenruimte noemen een gebeurtenis
- Soorten gebeurtenissen
  - Elementaire gebeurt. *Deelverzameling met 1 element*
  - Zekere gebeurt. *Uitkomstenruimte*
  - Onmogelijke gebeurt. *Lege verzameling*
- Begrippen
  - Complementaire gebeurtenis
    - $A^c = \Omega \setminus A$
  - Vereniging of unie
    - $A \cup B$
    - Het treedt of in A of in B
  - Exhaustief
    - Indien de unie de volledige uitkomstenruimte is
  - Doorsnede
    - $A \cap B$
    - Het element zit in A EN in B
  - Disjunct
    - Indien de doorsnede van 2 verzamelingen leeg is

Kansfunctie

<u>Axioma's van de kansrekening</u>

- $P(A) \geq 0$, voor elke gebeurtenis A
- $P(\Omega) = 1$
- Voor elke rij disjuncte gebeurtenissen: $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$

<u>De empirische definitie:</u>

Als een experimenten een groot aantal keer herhaald wordt, dan zal volgens de wet van de grote aantallen de relatieve frequentie van een gebeurtenis A zich stabiliseren rond de kans op die gebeurtenis.

<u>Kans definitie van Laplace</u>

De definitie gaat uit van een symmetrisch experiment, d.i. een experiment met een eindig aantal mogelijke uitkomsten die alle even waarschijnlijk zijn.

$$P(A) = \frac{\#A}{\#\Omega} = \frac{gunstige\ uitkomsten}{mogelijke\ uitkomsten}$$

Eigenschappen kansen

- $P(\emptyset) = 0$ dit volgt uit AX 3 en $\emptyset = \emptyset \cup \emptyset \cup \ldots$

Omschrijving	Symbolen	Bewijs/verantwoord.
Unie van een eindig aantal disjuncte gebeurtenissen	$P(A_1 \cup A_2 \cup \ldots \cup A_n)$ $= P(A_1 \cup A_2 \cup \ldots)$ $= \sum_{i=1}^{\infty} P(A_i)$ $= \sum_{i=1}^{n} P(A_i)$	
Kans van de complementaire gebeurtenis	$P(A^c) = 1 - P(A)$	$1 = P(\Omega)$ $= P(A \cup A^c)$ $= P(A) + P(A^c)$
Grenzen van de kansfunctie	$0 \leq P(A) \leq 1$	Uit ax 1
Deelverzamelingen en kansen	Als A B impliceert maw $A \subseteq B$ dan is $P(A) \leq P(B)$ en $P(B \setminus A) = P(B) - P(A)$	
Optelregel	$P(A \cup B)$ $= P(A) + P(B) - P(A \cap B)$	$P(A \cup B)$ $= P(A \cup (B \setminus (A \cap B)))$ $= P(A) + P(B \setminus (A \cap B))$ $= P(A) + P(B) - P(A \cap B)$
Ongelijkheid van Boole	$P(A \cup B) \leq P(A) + P(B)$	Omdat $P(A \cap B)$ steeds positief is

Combinatorische kansrekening

- Productregel
  - Alle mogelijke opties maal elkaar
- Variatie
  - Is een geordende keuze van k objecten uit een taal van n verschillende objecten **zonder** teruglegging van de gekozen objecten
  - Volgorde van de objecten speelt een rol
  - $V_n^k = n(n-1)(n-2)\dots(n-k+1) = \frac{n!}{(n-k)!}$ *GRT: n nPR k*
  - Bv kiezen van een raad van bestuur
- Herhalingsvariatie
  - Is een geordende keuze van k objecten uit een totaal van n verschillende objecten **met** teruglegging van de gekozen objecten
  - Volgorde van de objecten speelt een rol
  - $\overline{V_n^k} = n^k$
- Permutatie
  - Is een variatie van n uit n
  - Alle objecten worden gekozen → zien als herschikking
  - $P_n = V_n^n = n(n-1)(n-2)\dots 1 = n!$
- Herhalingspermutatie
  - Gebruiken we indien niet alle object
  - $P_n = \frac{n!}{n_1!n_2!\dots n_m!}$
  - Bv een anagram
- Combinatie
  - Is een ongeordende keuze van k objecten uit een totaal van n verschillende objecten **zonder** teruglegging van de gekozen objecten
  - Volgorde van de gekozen objecten speelt geen rol
  - $C_n^k = \frac{V_n^k}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$ *GRT: n nCr k*
  - Bv: lotto
- Herhalingscombinatie
  - Is een ongeordende keuze van k objecten uit een totaal van n verschillende objecten **met** teruglegging van de gekozen objecten
  - Volgorde van de gekozen objecten speelt geen rol
  - $C_{k+n-1}^{n-1} \binom{k+n-1}{n-1} = \frac{(k+n-1)!}{k!(n-1)!}$ *GRT: k+n-1 nCr n-1*

Voorwaardelijke kans

Voorwaardelijke kans en onafhankelijke gebeurtenissen

- Voorwaardelijke kans of a posteriori kans P(A | B) is de kans dat gebeurtenis A optreedt op voorwaarde dat gebeurtenis B optreedt.
- Onvoorwaardelijke kans is a priori kans

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad met\ P(B) > 0$$

- Vermenigvuldigingsregel
  - $P(A \cap B) = P(B)\,P(A \mid B)$
    EN $P(A \cap B) = P(A)P(B \mid A)$
- Onafhankelijke gebeurtenissen: een gebeurtenis A is onafhankelijke van gebeurtenis B indien het al dan niet optreden van B geen invloed heeft op het optreden van A
  - $P(A) = P(A \mid B)$ en $P(B) = P(B \mid A)$
  - We zeggen dat gebeurtenissen $A_1, A_2 \ldots, A_n$ onderling onafhankelijk zijn indien voor elke deelverzameling geldt dat
    $P(A_{i1} \cap A_{i2} \cap \ldots A_{ik}) = P(A_{i1})P(A_{i2})\ldots P(A_{ik})$

Wet van de totale kans en de regel van Bayes

Wet van de totale kans

Beschouwen we een gebeurtenis A met onvoorwaardelijke kans P(A). Indien de gebeurtenissen $B_1, B_2, \ldots B_n$ een partitie van Ω vormen, dan kunnen we P(A) shrijven als

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$$

Bewijs zie pagina 49

⇒ $A = A \cap \Omega = A \cap (B_1 \cup B_2 \cup .. B_n) = \ldots$
⇒ *Dus* $P(A) = P((A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n))$
⇒ *Aangezien* B_i *en* B_j *disjunct zijn voor* $i \neq j$*, zijn ook* $A \cap B_i$ *en* $A \cap B_j$ *disjunct dus*
⇒ $P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$
⇒ $\quad = P(A/B_1)P(B_1) + P(A/B_2)P(B_2) + \ldots + P(A/B_n)P(B_n)$
⇒ *Waaruit de formule onmiddellijk volgt.*
⇒ *De som van de voorwaardelijke kansen is dus niet gelijk aan P(A)*

Regel van Bayes

Voor een partitie $B_1, B_2, \ldots B_n$ van Ω en een willekeurige gebeurtenis A geldt

$$P(B_i \mid A) = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}$$

Dit volgt onmiddellijk uit de definitie van de voorwaardelijke kans want

$$P(B_i \mid A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}$$

enkel bij ***voorwaardelijke kans***

Univariate kansvariabelen

Kansvariabelen en verdelingsfunctie

Een functie X die een uitkomstenruimte afbeeldt op R noemen we een kansvariabele of stochastische variabele X.
X = x interpreteren we als *"de functie X heeft als waarde x"*

De kans dat een kansvariabele een bepaalde x aanneemt, kunnen we afleiden uit de kans van de corresponderende gebeurtenis

⇒ Kansfunctie: P(X = x) = P(A) met A = $\{\omega \in \Omega \mid X(\omega) = x\}$

Cumulatieve verdelingsfunctie:

⇒ $F_X(X) = P(X \leq x)$, $x \in R$

Omschrijving	Symbolen	Bewijs/verantwoord.
Kans op een uitkomst tussen 2 waarden	$P(x_1 < X \leq x_2)$ $= F_X(x_2) - F_X(x_1)$	$F_X(x_2) = P(X \leq x_2)$ $= P(X \leq x_1) + P(x_1 < X \leq x_2)$ $= F_X(x_1) + P(x_1 < X \leq x_2)$
Voorwaarden cumulatieve verdelingsfunctie	$x_1 < x_2 \Rightarrow F_X(x_2) \leq F_X(x_1)$ $\lim_{n \to -\infty} F_X(x) = 0$ $\lim_{n \to +\infty} F_X(x) = 1$	

Discrete kansvariabele

Een discrete kansvariabele is een kansvariabele die slecht een **eindig** of aftelbaar oneindig aantal waarden kan aannemen.

$p_X(m_i) = P(X = m_i)$ met i = 1, 2, …, k

Omschrijving	Symbolen	Bewijs/verantwoord.
Kansverdeling	$p_X(m_i) \geq 0$ met i = 1, …, k en $\sum_{i=1}^{k} p_x m_i = 1$	⇒ Kans functie steeds + ⇒ uitkomst vh experiment steeds op 1 van de waarden wordt afgebeeld en $P(\Omega) = 1$

Cumulatieve verdelingsfunctie van een discrete kansvariabele:

⇒ $F_X(X) = P(X \leq x) = \sum_{m_i \leq x} p_x m_i$, $x \in R$

Continue kansvariabele

Een continue kansvariabele is een kansvariabele die een **oneindig** aantal waarden kan aannemen.

We kunnen GEEN kansverdeling meer opstellen omdat de variabele oneindig is. De kans dat een specifieke waarde zal voorkomen is dan bijgevolg ook 0.
→ Kansdichtheid $f_X(x)$ is de afgeleide van de verdelingsfunctie $F_X(x)$.

$$f_X(x) = \frac{d}{dx}F_X(x)$$

Merk op dat de kansdichtheid hier > 1 kan zijn.

Omschrijving	Symbolen	Bewijs/verantwoord.
Verdelingsfunctie bepalen uit kansdichtheid	$F_X(x') = P(X \leq x')$ $= \int_{-\infty}^{x'} f_X(x)dx$	De verdeling stemt dus overeen met opp onder kansdichtheid links van het punt x'
Voorwaarden kansdichtheid	* $f_X(x) \geq 0, \forall x \in R$ * $\int_{-\infty}^{\infty} f_X(x)dx = \lim_{x \to +\infty} F_X(x) = 1$	*afgeleide is een stijgende functie *opp tussen kansdichtheid en horizontale as steeds gelijk aan 1
Kans op X = x	$P(X = x) = \int_x^x f_X(y)dy = 0$	$P(X \leq x) = P(X < x)$ En dus $P((x_1 < X \leq x_2)$ $= \int_{x_1}^{x_2} f_X(x)dx$

Kenmerken van populatie verdelingen

Verwachtingswaarde van een kansvariabele

Discrete variabele	Continue variabele
$\mu_X = E(X) = \sum_{i=1}^{k} m_i p_X(m_i)$	$\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$

Omschrijving	Symbolen	Bewijs/verantwoord.
Lineaire transformatie	Y = aX + b → E(Y) = aE(X) + b	$E(Y) = \int_{-\infty}^{\infty}(ax+b)f_X(x)dx$ $= a\int_{-\infty}^{\infty} x f_X(x)dx + b\int_{-\infty}^{\infty} x f_X(x)dx$ =aE(X) + b
Verwachtingswaarde van een constante	E(b) = b	De VW ve constante is de constant zelf
Variabele centreren	Y = X – E(X) = 0	E(Y) = E(X – E(X)) $= E(X - \mu_X)$

MARKOV

Voor een willekeurige positieve kans variabele X op $[0, \infty]$ geldt dat

$$P(X \geq a) \leq \frac{E(X)}{a}, \quad \forall a > 0$$

Bewijs zie cursus pag 61

Variantie en standaardafwijking

⇒ VARIANTIE

$$\sigma_X^2 = Var(X) = E[(X - E(X))^2]$$

Discrete variabele	Continue variabele
$Var(X) = \sum_{i=1}^{k} (m_i - E(X))^2 p_X(m_i)$	$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx$

⇒ STANDAARDAFWIJKING

$$\sigma_X = \sqrt{Var(X)}$$

Omschrijving	Symbolen	Bewijs/verantwoord.
Rekenformule	$Var(X) = E(X^2) - (E(X))^2$	Zie pag 62
Lineaire transformatie	$Y = aX + b$ $Var(Y) = a^2 Var(X)$ En $\sigma_y = \lvert a \rvert \, \sigma_X$	Zie pag 63
Variantie van een constante	$Var(b) = 0$	Variantie van een constant is 0
Gestandaardiseerde kansvariabele	$Z = \frac{X - \mu_X}{\sigma_X}$ met $E(Z) = 0$ & $Var(Z) = 1$	Zie pag 63

CHEBYSHEV

Voor een willekeurige kans variabele X geldt dat

$$P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}, \quad \forall t > 0$$

Bewijs zie cursus pag 64

Kwantielen, mediaan en kwartielen

⇒ KWANTIELFUNCTIE

- Is de inverse van de verdelingsfunctie
- De waarde van deze functie in een punt p is het p-kwantiel
- $Q_X(p) = \inf \{x \in R \mid F_X(x) \geq p\}$ met $0 < p < 1$
- $\int_{-\infty}^{Q_x(p)} f_X(x)dx = p, \quad 0 < p < 1$

⇒ MEDIAAN

- Med (X)= Q_X (0,5)

Andere kenmerken

⇒ MODUS

- Waarde waarvoor de kansverdeling of kansdichtheid haar (lokaal) maximum bereikt

⇒ SCHEEFHEID

- $\gamma_X = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$
- Waarde < 0 dan linksscheef (links niets, rechts 'bult'
- Waarde > 0 dan rechtsscheef

⇒ KURTOSIS

- $\kappa_X = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$
- Geeft aan hoe dik de staarten zijn

Multivariate kans variabelen

Indien meerdere reële getallen geassocieerd worden met elke mogelijke uitkomst

Discrete multivariate kans variabelen

- Oneindig, aftelbaar aantal waarden kan aannemen
- $p(x_i,y_j) = P(X = x_i, Y = y_l) = P((X = x_i) \cap P(Y = y_j)$
- bv gooien met 2 dobbelstenen

Omschrijving	Symbolen	Bewijs/verantwoord.
Gezamenlijke kansverdeling	$p_X(x_i,y_i) \geq 0$ en $\sum_i \sum_j p(x_i y_i) = 1$	
Marginale kansverdelingen	$p_x(x_i) = \sum_i p\,(x_i, y_i)$ en $p_y(y_i) = \sum_i p\,(x_i, y_i)$	Voor x sommeren over alle mogelijke waarden van y en omgekeerd

Continue multivariate kans variabalen

- Oneindig, onaftelbaar veel waarden kan aan nemen
- $p(x_i,y_j) = P(X = x_i, Y = y_l) = P((X = x_i) \cap P(Y = y_j)$

Omschrijving	Symbolen	Bewijs/verantwoord.
Gezamenlijke kansverdeling	$f(x, y) \geq 0$ en $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$	
Marginale kansverdelingen	$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$ $f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$	Om kansdichtheid $f_x(x)$ te bepalen integreren we over gezamenlijke kansdichtheid f(x,y) met alle mogelijk waarden voor Y

Afhankelijkheid en voorwaardelijke verdeling

- Afhankelijk vs onafhankelijk
  - $p(x_i,y_i) = p_X(x_i)p_Y(y_i)$
  - $f(x,y) = f(x)f(y)$
    - Als het voldaan is **on**afhankelijk
    - Als het niet voldaan voor minstens 1 koppel is afhankelijk
- Voorwaardelijke kansverdeling
  - $p_{X|Y}(x_i|y_i) = P(X = x_i|Y = y_i) = \frac{P(X=x_i, Y=y_i)}{P(Y=y_i)} = \frac{p(x_i,y_i)}{p_Y(y_i)}$
- Voorwaardelijke kansdichtheden
  - $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ en $f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$

Kenmerken

Verwachtingswaarde

Discrete variabele	Continue variabele
$E[g(X,Y)] = \sum_i \sum_j g(x_i, y_i) p(x_i, y_i)$	$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dxdy$

Omschrijving	Symbolen	Bewijs/verantwoord.
VW van lineaire combinatie	E(aX + bY + C) = aE(X) + bE(Y) + c	Zie pag 74
Factorisatie	Zijn X en Y onafhankelijk dan E(XY) = E(X)E(Y)	Zie pag 75

(Co)variantie en correlatie

⇒ COVARIANTIE

- $cov(X,Y) = E[(X - E(X))(Y - E(Y))]$

Discrete variabele	Continue variabele
$Cov(X,Y) = \sum_i \sum_j (x_i - E(X))(y_i - E(Y))\, p(x_i, y_i)$	$Cov(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X)).(y - E(Y))\, f(x,y) dxdy$

⇒ CORRELATIECOEFFIECIENT

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

Omschrijving	Symbolen	Bewijs/verantwoord.
VW van lineaire combinatie	E(aX + bY + C) = aE(X) + bE(Y) + c	Zie pag 74
Factorisatie	Zijn X en Y onafhankelijk dan E(XY) = E(X)E(Y)	Zie pag 75
Lineaire transformatie	Cov (a_1X + b_1, a_2Y + b_2) = a_1a_2Cov(X, Y)	Zie pag 77
Cov tussen onafhankelijke variabelen	Indien X en Y onafhankelijk zijn is Cov(X, Y) = 0	Volgt uit rekenformule ! omgekeerde niet waar!
Variantie van een lineaire combinatie	Var(aX + bY + c) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)	Zie pag 78
Voor onafhankelijke variabelen gelt	Var(X + Y) = Var(X) + Var(Y)	
Covariantie en lineaire combinaties	Cov(X + Y, Z) = Cov (X,Z) + Cov(Y,Z)	Zie pag 78

Discrete verdelingen

Uniforme verdeling

⇒ Als elke mogelijke waarde dezelfde kans heeft op te treden

Bernoulli verdeling

⇒ Hebben we als we een kansverdeling hebben die maar 2 waarden kan aannemen = binaire variabele
⇒ X ~ B (1, p)

Binomiale verdeling

⇒ Is een opeenvolging van identieke en onafhankelijke Bernoulli experimenten
⇒ X ~ B (n, p)
⇒ Kansverdeling P (X = k) = Binompdf (n, p, k)
⇒ Cumulatieve verdelingsfunctie P (X ≤ k)= Binomcdf(n, p, k)

Omschrijving	Symbolen	Bewijs/verantwoord.
Som van binomiale verdeling	$Z = X_1 + X_2 \sim B(n_1 + n_2, p)$	2 onafhankelijke binominaal verdeelde kans variabelen met zelfde parameter, daarvan is de som ook binomiaal verdeeld.
Aantal mislukkingen	$Y = n - X \sim B(n, 1 - p)$	
Wet van de grote aantallen	$\lim_{n\to\infty} P\left(\left\lvert\frac{Y_N}{n} - p\right\rvert \geq \epsilon\right) = 0$	Gebaseerd op Chebyshev. Zie pag 88

Poisson verdeling

⇒ Kan beschouwd worden als de limiet van de binomiale verdeling met λ = np
⇒ X ~ Poisson (λ)
⇒ Kansverdeling P (X = k) = Poissonpdf (λ, k)
⇒ Cumulatieve verdelingsfunctie P (X ≤ k) = Poissoncdf (λ, k)
⇒ Voorwaarden
- λ is de intensiteit = gemiddeld aantal gebeurtenissen in gekozen eenheid
- gebeurtenissen komen niet in groep voor m.a.w. de kans op 2 of meer gebeurtenissen binnen eenzelfde interval is verwaarloosbaar t.o.v. de kans op 1 of geen gebeurtenis
- de kans dat zich voordoet in een bepaald interval is dezelfde voor alle intervallen
- of een gebeurtenis zich voordoet in een bepaald interval is onafhankelijk van wat in de andere intervallen gebeurt

⇒ bv aantal oproepen in een callcenter, aantal tikfouten in een krant, aantal voortuigen dat voorbij rijdt in een bepaalde tijd

Omschrijving	Symbolen	Bewijs/verantwoord.
Som van poisson verdelingen	$Z = X_1 + X_2 \sim$ Poisson $(\lambda_1 + \lambda_2)$	2 onafhankelijke poisson verdeelde kans variabelen, daarvan is de som ook poisson verdeeld.

Geometrische verdeling

⇒ Een opeenvolging van identieke of onafhankelijk Bernoulli experimenten met kans p of succes. We zijn nu echter niet geïnteresseerd op aantal successen, maar **in het aantal experimenten dat nodig is om een eerste succes te observeren**

⇒ Als het eerste succes optreedt in experiment k, moeten we bij de k-1 experimenten immers een mislukking hebben geobserveerd.

⇒ $F_x(k) = 1 - (1-p)^k$

⇒ Bv sollicitatie, lotto

⇒ Heeft geen geheugen, de kansen blijven gelijk.

Omschrijving	Symbolen	Bewijs/verantwoord.
Geheugenloosheid	Indien X geometrisch verdeeld is, dan geldt voor $k_1>0$ en $k_2>0$ dat $P(X > k_1 + k_2 \mid X > k_1)$ $= P(X > k_2)$	Zie pag 95

Hypergeometrische verdeling

⇒ Eindige populatie van grootte N bestaande uit r elementen met uitkomst 1 (succes) en N - r elementen met uitkomst 0 (mislukking). Uit die populatie trekken we lukraak en zonder teruglegging een steekproef van $n \leq$ N-elementen. Het aantal successen X in deze steekproef heeft dan een hypergeometrische verdeling

⇒ Bv capture – recapture methode

- Wordt gebruikt om de grootte van een populatie in te schatten o.b.v. 2 of meer onvolledige lijsten
  - Eerste groep vangen en kenmerken dan weer vrijlaten
  - Tweede groep vangen en kijken hoeveel er gekenmerkt zijn
  - Schatten aantal dieren, drugsgebruikers, HIV-patiënten, aantal personen die de belastingen ontduiken

Continue verdelingen

Uniforme verdeling

⇒ De kansvariabele X is uniform verdeeld over het interval (a, b) als de kansdichtheid constant is over (a, b)

⇒ $f_x(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & elders \end{cases}$

Exponentiële verdeling

⇒ $f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & elders \end{cases}$

⇒ bv wachtrij in een postkantoor

Omschrijving	Symbolen	Bewijs/verantwoord.
Verband met Poisson verdeling		Zie pag 99-110 EXAMEN!
Geheugenloosheid	Indien X exp verdeeld is, $P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2)$	Zie pag 100 Belangrijk!

Normale verdeling

Definitie en kenmerken

⇒ $f_x(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

⇒ Standaardnotatie: $X \sim N(\mu, \sigma^2)$

⇒ De verwachtingswaarde E(X)

- Geeft de kansdichtheid en de symmetrie rond μ weer
- Is locatieparameter
- Geeft locatie van de klokcurve weer

⇒ Variantie geeft de schaalcurve weer

- Als sigma groot is, is er een grote spreiding
- Platte of hoge klokcurve

Standaard normale verdeling

⇒ Is de normale verdeling met $\mu = 0$ en $\sigma = 1$

⇒ $\mu < 0$ naar links en $\mu > 0$ naar rechts verschuiven

⇒ $\sigma < 1$ uitrekken naar boven, wordt smaller en $\sigma > 1$ inkrimpen naar beneden, wordt breder

⇒ Normalpdf en normalcdf

Omschrijving	Symbolen	Bewijs/verantwoord.
Negatieve waarden	$F(-z) = 1 - F(z)$	zie pag 102
Kwantielfunctie p < 0,5	$F^{-1}(1-p) = -F^{-1}(p)$	Zie pag 103 & invNorm

Willekeurige normale verdeling

Omschrijving	Symbolen	Bewijs/verantwoord.
Lineaire transformatie	$Y \sim N(a\mu + b, a^2\sigma^2$	zie pag 104
Omzetten standaardnormale verdeling	$Z = \dfrac{X - \mu}{\sigma}$	

Centrale limietstelling

- ⇒ Is een van de basisresultaten uit de statistiek
- ⇒ Ze stelt dat de som van n onafhankelijke kans variabelen X_i $voor\ n \to\ \infty$ normaal verdeeld is, ongeacht de verdelingen van de kans variabelen.

Centrale limietstelling

Beschouw n onafhankelijke kansvariabelen X_1, X_2, …, X_n met willekeurige kansverdeling of dichtheid en met E(X) = μ_i en Var(X_i) = σ_i^2. Onder bepaalde technische voorwaarden geldt voor $Y_n = X_1 + X_2 + \ldots + X_n$ dat

$$\lim_{n \to \infty} P\left(\frac{Y_n - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}\right)$$

Voor voldoende grote waarvan van *n* is Y_n bij benadering normaal verdeeld. Als de kans variabelen alle dezelfde verdeling hebben met E(X_i) = μ en Var(X_i) = σ^2, dan

$$\lim_{n \to \infty} P\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}}\right)$$

Enkele vuistregels als de kans variabelen dezelfde verdeling hebben

- ⇒ De kansdichtheid wijkt niet te sterk van de normale als n ≥ 5
- ⇒ De kansverdeling of –dichtheid vertoont geen al te grote pieken: n ≤ 12
- ⇒ Voor andere continue kans variabelen die in de praktijk voorkomen: n ≥ 30

Omschrijving	Symbolen	Bewijs/verantwoord.
Normale verdeling van de binomiale verdeling	$B(n, p) \approx N(np, np(1-p)$	De binomiale verdeling voor voldoende grote n kan normaal verdeeld benaderd worden

- ⇒ Continuïteitscorrectie
  - Als we een binomverdeling hebben waarbij np > 5 en n(1-p) > 5 dan kunnen we de verdeling normaal benaderen met normale verdeling
  - Continuïteitscorrectie is dan +1/2 of -1/2 of beide

χ^2-verdeling

- ⇒ Chi-kwadraat verdeling
- ⇒ Belangrijk in verklarende statistiek
- ⇒ n = aantal vrijheidsgraden

Omschrijving	Symbolen	Bewijs/verantwoord.
Karakterisatie	$Z_1^2 + Z_2^2 + \cdots + Z_n^2 \sim \chi_n^2$	Een KV die chi-kwadraatische verdeling heeft kunnen we schrijven als een som van n onafhankelijke kwadraten

- ⇒ χ^2pdf, χ^2cdf en INVCHI2

t- verdeling

- ⇒ Belangrijk in verklarende statistiek
- ⇒ n = aantal vrijheidsgraden
- ⇒ Als n oneindig groot wordt, vinden we de standaard normale dichtheid terug
- ⇒ De t-verdeling is dus een symmetrische verdeling met zwaardere staarten dan de normale verdeling

Omschrijving	Symbolen	Bewijs/verantwoord.
Karakterisatie	$\frac{X}{\sqrt{Y/n}} \sim t_n$ met $X \sim N(0,1) en\ Y \sim \chi_n^2$ onafhankelijk	KV is t-verdeeld asa de V kan geschreven worden als quotiënt van (zie formule)

- ⇒ De variantie zal steeds groter zijn dan bij de standaard normale verdeling
- ⇒ Tcdf, tpdf en invT

F-verdeling

- ⇒ Heeft twee parameters waarvan de volgorde BELANGRIJK is!
- ⇒ $X \sim F_{n,\,m}$
- ⇒ Wordt gebruikt voor de vergelijking van 2 groepen

Omschrijving	Symbolen	Bewijs/verantwoord.
Karakterisatie	$\frac{X/n}{Y\ /m} \sim F_{n,m}$	En KV is F-verdeeld asa de variabele geschreven kan worden als quotiënt van twee onafhankelijke chikwadraat verdeelde KV

- ⇒ Fcdf, Fpdf en InvF

Bivariate normale verdeling

- ⇒ Multivariate uitbreiding van de normale verdeling
- ⇒ Kansvariabelen X en Y allebei normaal verdeeld
- ⇒ $Cov(X, Y) = \rho\sigma_X\sigma_Y$ met ρ de correlatie
- ⇒ Zie cursus pag 111 ev