Syllabus bij het college

Wiskundige Statistiek

C.A.J. Klaassen

Academisch jaar 1997–1998
tweede trimester

Korteweg-de Vries Instituut
Faculteit WINS
Plantage Muidergracht 24
1018 TV Amsterdam

Inhoud Wiskundige Statistiek

De **tentamenstof** voor dit college bestaat uit dit dictaat, in het bijzonder Deel II Statistiek. Uiteraard is beheersing van de stof (en met name de stellingen) uit het eerste deel noodzakelijk om de kanstheoretische technieken, die daar behandeld zijn, toe te kunnen passen binnen de statistiek.
De negen verdelingen genoemd op pp.13–15 worden bekend verondersteld evenals de stellingen 6.1, 6.2, 7.1 en 7.2 en hun bewijzen.

Wiskundige Statistiek

Deel I. K a n s r e k e n i n g

1. Kansruimten

Wanneer we experimenten beschouwen, kunnen we twee typen onderscheiden:
- deterministische experimenten, waarbij de uitkomst van tevoren vastligt,
- toevalsexperimenten, waarbij de uitslag van "het toeval" afhangt.

Als mathematisch model voor toevalsexperimenten gebruiken we kansruimten. Een kansruimte bestaat uit een drietal objecten:

Ω de **uitslagenruimte** of steekproefruimte.
$\mathcal{A}$ een collectie deelverzamelingen van Ω. Deze deelverzamelingen worden **eventualiteiten** of gebeurtenissen genoemd.
P een **kansmaat** op $\mathcal{A}$, die aan elke $A \in \mathcal{A}$ een **kans** $P(A)$ toevoegt.

Voorbeeld 1.1. Bij het werpen van één dobbelsteen, waarbij we geïnteresseerd zijn in het aantal ogen, kunnen we de volgende keuzes maken: $\Omega = \{1, 2, \ldots, 6\}$, $\mathcal{A}$ alle deelverzamelingen van Ω, $P(\{i\}) = \frac{1}{6}$ voor $i = 1, 2, \ldots, 6$.

Dat we de kansruimte in voorbeeld 1.1 een redelijke keuze vinden, stoelt op onze intuïtie en ervaring, die we kunnen samenvatten in de volgende **empirische wet van de grote aantallen.** We voeren een gegeven toevalsexperiment n keer uit en geven met $f_n(A)$ de relatieve frequentie aan van het optreden van de eventualiteit A. Als we dit doen voor $n = 1, 2, \ldots$, krijgen we een rij getallen $\{f_n(A)\}$, die allerlei fluctuaties zal vertonen, maar voor groter wordende n steeds dichter bij een zekere waarde $f(A)$ zal komen te liggen.

Hoewel we in werkelijkheid $f(A)$ niet kunnen bepalen, willen we toch graag in ons model $P(A)$ gelijk aan $f(A)$ kiezen. Merk op, dat de relatieve frequenties $f_n(A)$ en dus ook $f(A)$ de volgende eigenschappen hebben:

$$0 \leq f_n(A) \leq 1, \text{ voor alle } A \in \mathcal{A}, \tag{1.1}$$

$$f_n(\Omega) = 1, \tag{1.2}$$

$$f_n\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} f_n(A_i), \text{ als } A_1, A_2, \ldots \text{ disjuncte eventualiteiten zijn, d.w.z. } A_iA_j = A_i \cap A_j = \phi \text{ voor } i \neq j. \tag{1.3}$$

Dit leidt ons tot de volgende definitie.

Definitie 1.1. $(\Omega, \mathcal{A}, P)$ is een kansruimte als het volgende geldt:
(i) Ω is een niet-lege verzameling,
(ii) $\mathcal{A}$ is een collectie deelverzamelingen van Ω met

$$\Omega \in \mathcal{A}, \tag{1.4}$$

$$\text{als } A \in \mathcal{A}, \text{ dan } A^c \in \mathcal{A}, \tag{1.5}$$

$$\text{als } A_1, A_2, \ldots \in \mathcal{A}, \text{dan } \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}. \tag{1.6}$$

(iii) de kansmaat $P : \mathcal{A} \to \mathbf{R}$ is een afbeelding van $\mathcal{A}$ naar $\mathbf{R}$ met

$$0 \leq P(A) \leq 1, \text{ voor alle } A \in \mathcal{A}, \tag{1.7}$$

$$P(\Omega) = 1, \tag{1.8}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i), \text{ als } A_1, A_2, \ldots \in \mathcal{A} \text{ disjunct zijn.} \tag{1.9}$$

Uit deze definitie kunnen we de volgende eigenschappen van $\mathcal{A}$ en P afleiden:

$$\text{Als } A, B \in \mathcal{A}, \text{ dan } AB \in \mathcal{A}. \tag{1.10}$$

$$\text{Als } A_1, A_2, \ldots \in \mathcal{A}, \text{ dan} \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}. \tag{1.11}$$

$$P(\phi) = 0. \tag{1.12}$$

$$P\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} P(A_i), \text{ als } A_1, \ldots A_m \text{ disjunct zijn.} \tag{1.13}$$

$$P(A^c) = 1 - P(A), \text{ als } A \in \mathcal{A}. \tag{1.14}$$

$$\text{Als } A, B \in \mathcal{A},\ B \supset A, \text{ dan } P(B) - P(A) = P(B - A) \geq 0. \tag{1.15}$$

$$P(A \cup B) = P(A) + P(B) - P(AB), \text{ als } A, B \in \mathcal{A}. \tag{1.16}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i), \text{ als } A_1, A_2, \ldots \in \mathcal{A}. \tag{1.17}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n), \text{ als } A_1 \subset A_2 \subset \ldots. \tag{1.18}$$

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} P(A_n), \text{ als } A_1 \supset A_2 \supset \ldots. \tag{1.19}$$

Voorbeeld 1.2. Bij een toevalsexperiment met slechts twee mogelijke uitkomsten spreken we van een **alternatief** en noemen we de twee uitkomsten **succes en mislukking** of kruis en munt of 1 en 0. We kiezen $\Omega = \{0, 1\},\ \mathcal{A} = \{\phi, \{0\}, \{1\}, \Omega\}, P(\{1\}) = p$. Met de **succeskans** p ligt de kansmaat P vast.

Voorbeeld 1.3. Bij een toevalsexperiment met N mogelijke uitkomsten kiezen we $\Omega = \{1, 2, \ldots, N\}$ en $\mathcal{A}$ de collectie van alle deelverzamelingen van Ω. De kansmaat P ligt volledig vast door de $N-1$ getallen $P(\{1\}), P(\{2\}), \ldots, P(\{N-1\})$. Als alle **simpele eventualiteiten** even waarschijnlijk zijn, geldt $P(\{i\}) = \frac{1}{N},\ i = 1, \ldots, N$, en spreken we van een **aselecte trekking uit** Ω. Dan geldt voor elke $A \in \mathcal{A}$.

$$P(A) = \sum_{i \in A} P(\{i\}) = \frac{|A|}{|\Omega|}. \tag{1.20}$$

Speciale gevallen zijn $N = 2$ (voorbeeld 1.2) en $N = 6$ (voorbeeld 1.1).

Voorbeeld 1.4. Trekken we een steekproef van n elementen uit een populatie van N elementen, dan kunnen we ook denken aan het kiezen van n knikkers uit een vaas met N knikkers, die genummerd zijn van 1 tot en met N.

(i) We krijgen een **aselecte steekproef zonder teruglegging** als we blindelings n knikkers tegelijk uit de vaas nemen. Natuurlijk moet dan $n \leq N$ zijn. De mogelijke uitkomsten zijn nu alle deelverzamelingen met n elementen van $\{1, \ldots, N\}$. Er zijn $\binom{N}{n}$ van zulke verzamelingen die we uit symmetrie overwegingen alle even waarschijnlijk achten en dus alle kans $1/\binom{N}{n}$ geven.

(ii) Nemen we de knikkers één voor één uit de vaas en letten we op de volgorde, dan zijn de mogelijke uitkomsten alle geordende n-tallen $(i_1, \ldots, i_n)$ met $i_1, \ldots, i_n \in \{1, \ldots, N\}$ en alle verschillend. Elk van deze n-tallen heeft nu kans

$$\frac{1}{n!\binom{N}{n}} = \frac{(N-n)!}{N!}$$

en we spreken van een **geordende aselecte steekproef zonder teruglegging**.

(iii) Nemen we tenslotte de knikkers één voor één uit de vaas en leggen we elke knikker terug in de vaas alvorens we de volgende kiezen, dan zijn de mogelijke uitkomsten te identificeren met alle geordende n-tallen $(i_1, \ldots, i_n)$ met $i_1, \ldots, i_n \in \{1, \ldots, N\}$. Het aantal mogelijke uitkomsten bij dit experiment is dus N^n en elke uitkomst heeft dezelfde kans $1/N^n$. We spreken van een geordende **aselecte steekproef met teruglegging**.

Als, zoals in de voorbeelden tot hier toe, Ω een eindige of aftelbare verzameling is en $\mathcal{A}$ uit alle deelverzamelingen van Ω bestaat, is er sprake van **discrete** kansruimten. Als Ω niet aftelbaar is, bestaat $\mathcal{A}$ meestal **niet** uit alle deelverzamelingen van Ω.

Voorbeeld 1.5. Een wijzer kan om een verticale as draaien, waarbij de punt van de wijzer een cirkel beschrijft waarlangs een lineaire schaalverdeling is aangebracht, die van 0 tot 1 loopt. Men geeft de wijzer een flinke beginsnelheid, wacht tot hij weer stilstaat en leest dan af welk punt op de schaalverdeling wordt aangewezen. We kiezen $\Omega = [0, 1)$ en $\mathcal{A}$ de kleinste collectie deelverzamelingen van Ω, die aan (1.4) tot en met (1.6) voldoet en alle intervallen binnen Ω bevat. Op grond van de symmetrie kiezen we P zó dat $P([a, b)) = b - a$, waarbij $0 \leq a < b < 1$. In de maattheorie wordt bewezen dat P hiermee wordt vastgelegd op $\mathcal{A}$. $(\Omega, \mathcal{A}, P)$ beschrijft een **aselecte trekking uit** Ω.

Voorbeeld 1.6. Het bepalen van de levensduur van een willekeurig gekozen transistor kunnen we opvatten als een toevalsexperiment. Vaak kiest men hier $\Omega = [0, \infty)$, $\mathcal{A}$ de kleinste collectie deelverzamelingen van Ω, die aan (1.4) tot en met (1.6) voldoet en alle intervallen binnen Ω bevat, en $P([a, b)) = e^{-\lambda a} - e^{-\lambda b}$ met λ een vast positief getal en met $0 \leq a \leq b$.

Opgaven

1.1. Verifieer (1.10) tot en met (1.19).

1.2. Bewijs dat voor ieder drietal eventualiteiten A, B en C in een kansruimte

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

1.3. Hoe groot is de kans dat je, als je vijf dobbelstenen tegelijk gooit,

A. bij elke steen hetzelfde aantal ogen vindt,
B. $1, 2, 3, 4, 5$ krijgt,
C. tenminste twee keer een 1 krijgt,
D. tenminste twee keer een zelfde aantal ogen vindt?

1.4. In een kast staan N verschillende paren schoenen. Zij $n < N/2$. Als je aselect $2n$ schoenen uit deze kast pakt, hoe groot is dan de kans dat er onder deze $2n$ schoenen geen enkel paar bij elkaar passende schoenen is? Hoe groot is de kans dat je op deze manier juist één paar bij elkaar passende schoenen krijgt?

2. Onafhankelijkheid

Bij een aselect gekozen gezin met twee kinderen beschouwen we de volgende gebeurtenissen:
A. het eerste kind is een jongen,
B. het tweede kind is een meisje.
Uit symmetrie overwegingen nemen we aan dat aselect gekozen betekent dat elk van de mogelijkheden (M, M), (M, J), (J, M) en (J, J) even waarschijnlijk is en dus kans $\frac{1}{4}$ heeft. (Met (M, J) bedoelen we dat het eerste kind een meisje en het tweede een jongen is). Merk op dat nu

$$P(A) = P(B) = \tfrac{1}{2},\ P(AB) = \tfrac{1}{4}$$

en dus

$$P(AB) = P(A)P(B). \tag{2.1}$$

Definitie 2.1. Zij $(\Omega, \mathcal{A}, P)$ een kansruimte. Twee eventualiteiten A en $B \in \mathcal{A}$, waarvoor (2.1) geldt, heten **onderling onafhankelijk** (o.o.).

Zij nu in het bovenstaande voorbeeld
C. de twee kinderen zijn van hetzelfde geslacht.
Omdat $P(A) = P(B) = P(C) = \frac{1}{2}$ en $P(AB) = P(AC) = P(BC) = \frac{1}{4}$ gelden, kunnen we niet alleen concluderen dat A en B o.o. zijn maar ook dat A en C o.o. zijn en B en C. Het is echter duidelijk dat we A, B en C **niet** onderling onafhankelijk kunnen noemen. Immers, als A en B optreden dan treedt C zeker niet op. Merk op dat $P(ABC) = 0$ en $P(A)P(B)P(C) = \frac{1}{8}$.

Definitie 2.2. Zij $\mathcal{C}$ een collectie van eventualiteiten (dus $\mathcal{C} \subset \mathcal{A}$). De eventualiteiten in $\mathcal{C}$ heten **paarsgewijs onafhankelijk** als voor elk tweetal $A, B \in \mathcal{C}$ met $A \neq B$ (2.1) geldt. De eventualiteiten in $\mathcal{C}$ heten **onderling onafhankelijk** (o.o.) als voor elke k-greep $A_1, \dots, A_k \in \mathcal{C}$ van verschillende eventualiteiten

$$P(A_1 \dots A_k) = P(A_1) \dots P(A_k). \tag{2.2}$$

Laat nu $\mathcal{E}_1, \dots, \mathcal{E}_n$ experimenten zijn. Experiment $\mathcal{E}_i$ wordt beschreven door een kansruimte $(\Omega_i, \mathcal{A}_i, P_i)$ voor $i = 1, \dots, n$. Bij het **samengestelde experiment** $\mathcal{E}$ willen we nu een kansruimte $(\Omega, \mathcal{A}, P)$ kiezen. Voor Ω nemen we de verzameling rijtjes $\omega = (\omega_1, \dots, \omega_n)$ met $\omega_i \in \Omega_i$, $i = 1, \dots, n$, dat wil zeggen $\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$. Voor $\mathcal{A}$ kiezen we een collectie van deelverzamelingen van Ω, die aan (1.4) tot en met (1.6) voldoet en die de deelverzamelingen $A_1 \times A_2 \times \dots \times A_n$ bevat met $A_i \in \mathcal{A}_i$, $i = 1, \dots, n$. Voor de constructie van de kansmaat P op deze $\mathcal{A}$ merken we allereerst op, dat het al dan niet optreden van $\Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n$ met $A_i \in \mathcal{A}_i$ slechts van de uitslag van experiment $\mathcal{E}_i$ afhangt en dat het dus redelijk is te eisen

$$P(\Omega_1 \times \dots \times A_i \times \dots \times \Omega_n) = P_i(A_i). \tag{2.3}$$

Als verder de uitslag van het ene experiment dat van het andere niet beïnvloedt, willen we P zó kiezen dat de eventualiteiten $\Omega_1 \times \ldots \times A_i \times \ldots \times \Omega_n$, $i = 1, \ldots, n$, *o.o.* zijn voor iedere keuze van $A_i \in \mathcal{A}_i, i = 1, \ldots, n$. Dus eisen we

$$P\left(\bigcap_{i=1}^{n} \Omega_1 \times \ldots \times A_i \times \ldots \Omega_n\right) = \prod_{i=1}^{n} P\left(\Omega_1 \times \ldots \times A_i \times \ldots \Omega_n\right). \tag{2.4}$$

M.b.v. (2.3) kunnen we (2.4) als volgt schrijven

$$P(A_1 \times \ldots \times A_n) = P_1(A_1) \ldots P_n(A_n). \tag{2.5}$$

In de maattheorie wordt nu bewezen dat $\mathcal{A}$ geschikt gekozen kan worden en dat P door de eis (2.5) vastligt. In deze situatie zeggen we dat het samengestelde experiment $\mathcal{E}$ uit **onderling onafhankelijke deelexperimenten** $\mathcal{E}_i$ bestaat.

Voorbeeld 2.1. Zijn $\mathcal{E}_1, \ldots, \mathcal{E}_n$ o.o. alternatieven met succeskansen $p_1, \ldots, p_n$, dan is voor iedere $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$

$$P\left(\{\omega\}\right) = p_1^{\omega_1}(1-p_1)^{1-\omega_1} \ldots p_n^{\omega_n}(1-p_n)^{1-\omega_n}.$$

Als $p_1 = \ldots = p_n = p$, dan wordt dit

$$P\left(\{\omega\}\right) = p^{\omega_1 + \ldots + \omega_n}(1-p)^{n-(\omega_1+\ldots+\omega_n)}. \tag{2.6}$$

Als $A_k \in \mathcal{A}$ de eventualiteit is, dat precies k alternatieven een succes opleveren, dan bestaat A_k uit de $\binom{n}{k}$ punten $\omega \in \Omega$ met $\omega_1 + \ldots + \omega_n = k$. Hieruit volgt

$$P(A_k) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{2.7}$$

Voorbeeld 2.2. Het trekken van een aselecte steekproef met teruglegging van n knikkers uit een vaas met N knikkers kun je opvatten als een samengesteld experiment bestaande uit n o.o. identieke deelexperimenten, nl. het aselect trekken van 1 knikker uit een vaas met N knikkers. In de notatie van voorbeeld 1.4.(iii) geldt immers

$$P\left(\{(i_1, \ldots, i_n)\}\right) = \frac{1}{N^n} = P_1\left(\{i_1\}\right) \ldots P_n\left(\{i_n\}\right).$$

<u>Opgaven</u>

2.1. Ga na dat (2.1) impliceert dat ook A^c en B o.o. zijn en dus ook A^c en B^c en verder A en B^c.

2.2. We trekken 5 kaarten uit een spel van 52 kaarten. Welke van de volgende paren eventualiteiten zijn o.o.?

A. We trekken de schoppenaas,
We trekken precies 1 rode kaart.

B. We trekken de hartenvrouw,
We trekken meer rode dan zwarte kaarten.

C. We trekken precies 1 boer,
We trekken meer zwarte dan rode kaarten.

3. Stochastische grootheden

Stel dat we geïnteresseerd zijn in de gemiddelde leeftijd van de nederlandse vrouw. Om kosten en moeite te sparen nemen we een geordende aselecte steekproef met teruglegging van omvang n uit de populatie van omvang N van nederlandse vrouwen. Volgens voorbeeld 2.2 kunnen we dit zien als een samengesteld experiment, waarbij ieder deelexperiment bestaat uit het aselect kiezen van één vrouw uit de hele populatie van omvang N. Voor deze deelexperimenten nummeren we de populatie-elementen van 1 tot en met N. Dan is de steekproefruimte dus $\Omega = \{1, \ldots, N\}$ en $P(\{\omega\}) = \frac{1}{N}$ voor $\omega \in \Omega$. We zijn nu niet in het nummer ω geïnteresseerd, maar in de leeftijd $X(\omega)$ van het steekproefelement ω. Hiermee hebben we een functie $X : \Omega \to \mathbf{R}$ geïntroduceerd. Daarbij willen we kunnen spreken over de kans dat de aselect gekozen vrouw een leeftijd van ten hoogste x heeft. Dat betekent dat $\{\omega | X(\omega) \leq x\}$ in de collectie $\mathcal{A}$ van toegestane deelverzamelingen van Ω bevat moet zijn. Dit leidt ons tot de volgende definitie.

Definitie 3.1. Zij $(\Omega, \mathcal{A}, P)$ een kansruimte. Iedere functie $X : \Omega \to \mathbf{R}$, waarvoor $\{\omega | X(\omega) \leq x\} \in \mathcal{A}$ voor alle $x \in \mathbf{R}$, heet **stochastische grootheid**.

Zij $\mathcal{B}$ de kleinste collectie deelverzamelingen van $\mathbf{R}$ die alle intervallen van de vorm $(-\infty, a]$ met $a \in \mathbf{R}$ bevat en die voldoet aan (1.4) t/m (1.6) met $\Omega = \mathbf{R}$. $\mathcal{B}$ heet de collectie van **Borel verzamelingen** van $\mathbf{R}$. Het is niet moeilijk na te gaan (en in de maattheorie wordt dit gedaan), dat voor iedere stochastische grootheid X en elke Borel verzameling $B \in \mathcal{B}$ de verzameling $\{\omega | X(\omega) \in B\}$ tot $\mathcal{A}$ behoort. We kunnen dus spreken over de kans, dat "X in B ligt":

$$P(X \in B) = P(\{\omega | X(\omega) \in B\}). \tag{3.1}$$

Verder kunnen we nu een kansmaat P_X op $(\mathbf{R}, \mathcal{B})$ definiëren door

$$P_X(B) = P(X \in B). \tag{3.2}$$

P_X heet de **(kans)verdeling** van de stochastische grootheid X. Deze verdeling wordt vastgelegd door de **verdelingsfunctie** F_X van X, die gedefinieerd is als

$$F_X(x) = P_X((-\infty, x]) = P(X \in (-\infty, x]) = P(X \leq x). \tag{3.3}$$

Met behulp van (1.15), (1.18) en (1.19) zien we dat F_X niet-dalend en rechts-continu is en aan

$$\lim_{x \to \infty} F_X(x) = 1, \quad \lim_{x \to -\infty} F_X(x) = 0 \tag{3.4}$$

voldoet. Voor elke functie F met deze eigenschappen bestaat er een stochastische grootheid X met verdelingsfunctie $F_X = F$. Kies namelijk $\Omega = \mathbf{R}$, $\mathcal{A} = \mathcal{B}$ en P op $\mathcal{B}$ zó dat $P((-\infty, x]) = F(x)$. Dan is $(\Omega, \mathcal{A}, P)$ een kansruimte en de identiteit op Ω de gezochte stochastische grootheid! Voor de maattheoretische details zij weer verwezen naar een college maattheorie.

Men noemt X en P_X discreet als er een eindige of aftelbare deelverzameling S van $\mathbf{R}$ is zó dat $P_X(S) = P(X \in S) = 1$. Dan gelden

$$P_X(B) = \sum_{y \in B \cap S} P(X = y), \quad B \in \mathcal{B}, \tag{3.5}$$

$$F_X(x) = \sum_{y \le x, y \in S} P(X = y), \quad x \in \mathbf{R}. \tag{3.6}$$

Wij zullen X en P_X continu noemen, als er een niet-negatieve (stuksgewijs) continue functie f_X op $\mathbf{R}$ bestaat met

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy. \tag{3.7}$$

Hierbij noemen we f_X stuksgewijs continu als er een eindige (of aftelbare) vereniging van open intervallen is, waarop f_X continu is en waarbuiten f_X gelijk aan 0 is. Merk op dat

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \tag{3.8}$$

en dat continuïteit van F_X **niet** impliceert dat de corresponderende verdeling P_X continu is! Bovenstaande functie f_X heet de **dichtheid** van X.

We bespreken nu enkele veelvoorkomende discrete en continue verdelingen.

1. Binomiale verdeling, $\text{bin}(n, p), \; n \in \mathbf{N}, \; 0 \le p \le 1$.

Zij X het aantal successen in n o.o. identieke alternatieven met succeskans p. Uit voorbeeld 2.1 volgt dat

$$P_X(\{k\}) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{3.9}$$

2. Hypergeometrische verdeling, $\text{hyp}(N, R, n), \; N, R, n \in \mathbf{N}, \; N \ge R, \; N \ge n$.

Uit een vaas met N knikkers, waarvan er R rood en $N - R$ wit zijn, worden er aselect zonder teruglegging n gekozen (zie voorbeeld 1.4.(i)). Als X het aantal rode knikkers in de steekproef is, dan geldt

$$P_X(\{k\}) = P(X = k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}, \quad \max\{0, n-(N-R)\} \le k \le \min\{n, R\}. \tag{3.10}$$

3. Geometrische of Pascal verdeling, Pascal(p), $0 \leq p \leq 1$.

Beschouw o.o. herhalingen van een alternatief met succeskans p. Zij X het aantal herhalingen dat uitgevoerd moet worden tot en met het eerste succes. Dan

$$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots. \tag{3.11}$$

4. Negatief binomiale verdeling, nbin(n, p), $n \in \mathbf{N}$, $0 \leq p \leq 1$.

Zij X het aantal o.o. herhalingen van een alternatief met succeskans p, dat uitgevoerd moet worden om n successen te krijgen. Dan

$$P(X = k) = \binom{k-1}{n-1} (1-p)^{k-n} p^n, \quad k = n, n+1, \ldots. \tag{3.12}$$

Merk op dat nbin$(1, p)$ en Pascal(p) hetzelfde verdelingen zijn.

5. Poisson verdeling, Poisson(λ), $\lambda > 0$.

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \ldots. \tag{3.13}$$

Merk op dat inderdaad $\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$.

6. Homogene of uniforme verdeling, hom(a, b), $a < b$.

X heeft een homogene verdeling op het interval (a, b) als X dichtheid

$$f_X(x) = \frac{1}{b-a} 1_{(a,b)}(x), \quad x \in \mathbf{R}, \tag{3.14}$$

en dus verdelingsfunctie

$$F_X(x) = \max\left\{0,\ \min\left\{\frac{x-a}{b-a}, 1\right\}\right\}, \quad x \in \mathbf{R}, \tag{3.15}$$

heeft (vergelijk voorbeeld 1.5).

7. Exponentiële verdeling, exp(λ), $\lambda > 0$.

X heeft dichtheid

$$f_X(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x), \quad x \in \mathbf{R}, \tag{3.16}$$

en verdelingsfunctie

$$F_X(x) = \left(1 - e^{-\lambda x}\right) 1_{(0,\infty)}(x), \quad x \in \mathbf{R}, \tag{3.17}$$

(vergelijk voorbeeld 1.6).

8. Gamma verdeling, $\Gamma(n, \lambda)$, $n \in \mathbf{N}$, $\lambda > 0$.
X heeft dichtheid

$$(3.18) \qquad f_X(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} 1_{(0,\infty)}(x), \quad x \in \mathbf{R},$$

en verdelingsfunctie

$$(3.19) \qquad F_X(x) = \left(1 - \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!} e^{-\lambda x}\right) 1_{(0,\infty)}(x), \quad x \in \mathbf{R}.$$

Merk op dat $\Gamma(n) = (n-1)!$ en dat $\Gamma(1, \lambda)$ en $\exp(\lambda)$ gelijk zijn.

9. Normale verdeling, $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbf{R}$, $\sigma > 0$.
X heeft dichtheid

$$(3.20) \qquad f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbf{R}.$$

Voor de zogenaamde standaard normale verdeling $\mathcal{N}(0, 1)$ noteren we de dichtheid met

$$(3.21) \qquad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbf{R},$$

en de verdelingsfunctie met

$$(3.22) \qquad \Phi(x) = \int_{-\infty}^{x} \varphi(y)dy, \quad x \in \mathbf{R}.$$

Ga na dat als X $\mathcal{N}(\mu, \sigma^2)$ verdeeld is, dan is

$$(3.23) \qquad f_X(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbf{R},$$

en

$$(3.24) \qquad F_X(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \varphi(y)dy = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbf{R}.$$

Om na te gaan dat f_X inderdaad een dichtheid is, moeten we (3.8) verifiëren voor $f_X = \varphi$. Immers via de substitutie $y = (x - \mu)/\sigma$ zien we dat

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \varphi(y)dy.$$

Inderdaad geldt, via de overgang op poolcoördinaten,

$$\left\{\int_{-\infty}^{\infty} \varphi(x)dx\right\}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x)\varphi(y)dxdy = \int_0^{\infty}\int_0^{2\pi} \frac{1}{2\pi}e^{-\frac{1}{2}r^2}\ r\ d\varphi\ dr$$
$$= \int_0^{\infty} r\ e^{-\frac{1}{2}r^2}dr = \lim_{b\to\infty}\left[-e^{-\frac{1}{2}r^2}\right]_0^b = 1.$$

Wanneer we een aantal stochastische grootheden die gedefinieerd zijn op dezelfde kansruimte combineren tot een **vector**waardige functie, spreken we van een stochastische vector.

Definitie 3.2. Zij $(\Omega, \mathcal{A}, P)$ een kansruimte. Iedere functie $X = (X_1, \ldots, X_k) : \Omega \to \mathbf{R}^k$ heet een **stochastische vector** als elk van de X_i een stochastische grootheid is.

In de situatie van deze definitie kunnen we dus, voor iedere $x = (x_1, \ldots, x_k) \in \mathbf{R}^k$, spreken over de kans dat $X \leq x$, dat wil zeggen $X_1 \leq x_1, \ldots, X_k \leq x_k$:

$$\begin{aligned} F_X(x) &= P_X((-\infty, x]) = P(X \leq x) \\ &= P(X_1 \leq x_1, \ldots, X_k \leq x_k) = P\left(\{\omega \mid X_1(\omega) \leq x_1, \ldots, X_k(\omega) \leq x_k\}\right). \end{aligned} \tag{3.25}$$

Hierbij is $(-\infty, x] = (-\infty, x_1] \times \ldots \times (-\infty, x_k] \subset \mathbf{R}^k$. De verdeling P_X van X heet ook wel de **simultane verdeling** van $X_1, \ldots, X_k$ en de verdelingsfunctie F_X de **simultane verdelingsfunctie** van $X_1, \ldots, X_k$. Verder worden in deze situatie P_{X_i} en F_{X_i} de **marginale verdeling** respectievelijk de **marginale verdelingsfunctie** van X_i genoemd. Uit (1.18) volgt

$$F_{X_i}(x_i) = \lim_{\substack{x_j\to\infty \\ j\neq i}} F_X(x_1, \ldots, x_k), \quad x_i \in \mathbf{R}. \tag{3.26}$$

We noemen de stochastische vector X en zijn verdeling weer **discreet** als er een **eindige** of **aftelbare** deelverzameling S van $\mathbf{R}^k$ bestaat zó dat $P_X(S) = P(X \in S) = 1$. Naast (3.6) geldt nu

$$P(X_i = x_i) = \sum_{\substack{y\in S \\ y_i = x_i}} P(X = y). \tag{3.27}$$

Voorbeeld 3.1. Bij een experiment met k mogelijke uitkomsten genummerd van 1 tot en met k is $p_i \geq 0$ de kans dat uitkomst i optreedt, $i = 1, \ldots, k$, $p_1 + \ldots + p_k = 1$. Zij bij n o.o. herhalingen van het experiment X_i het aantal malen dat uitkomst i optreedt. De stochastische vector $X = (X_1, \ldots, X_k)$ is discreet en heeft een zogenaamde **multinomiale verdeling** $(\text{mult}(n, p_1, \ldots, p_k))$ met parameters $n, p_1, \ldots, p_k$:

$$P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{x_1!, \ldots, x_k!}p_1^{x_1}\ldots p_k^{x_k}, \quad x_i \geq 0, \quad x_1 + \ldots + x_k = n. \tag{3.28}$$

Merk op dat X_i opgevat kan worden als het aantal successen in n o.o. alternatieven met succeskans p_i en dat X_i dus $\text{bin}(n, p_i)$ verdeeld is. M.a.w., de marginale verdelingen van een multinomiale verdeling zijn binomiaal.

We noemen de stochastische vector X en zijn verdeling weer **continu** als er een **dichtheid** f_X is, die niet-negatief en continu is op een open deelverzameling van $\mathbf{R}^k$ en daarbuiten 0 en die voldoet aan

$$F_X(x) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_k} f_X(y_1, \dots, y_k) dy_k \dots dy_1, \quad x \in \mathbf{R}^k. \tag{3.29}$$

Ga na dat

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_k) dy_k \dots dy_1 = 1,$$

$$F_{X_i}(x_i) = \int_{-\infty}^{x_i} \left\{ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_k) dy_k \dots dy_{i+1} dy_{i-1} \dots dy_1 \right\} dy_i$$

en dus

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_X(y_1, \dots, y_k) dy_k \dots dy_{i+1} dy_{i-1} \dots dy_1. \tag{3.30}$$

Definitie 3.2. Zij $X = (X_1, \dots, X_k)$ een stochastische vector. $X_1, \dots, X_k$ heten **onderling onafhankelijk** (o.o.) als voor alle $B_i \in \mathcal{B}$, $i = 1, \dots, k$, de eventualiteiten $\{X_i \in B_i\}$,
$i = 1, \dots, k$, o.o. zijn.

Het is eenvoudig na te gaan dat $X_1, \dots, X_k$ o.o. zijn als

$$P(X_1 \in B_1, \dots, X_k \in B_k) = P(X_1 \in B_1) \dots P(X_k \in B_k), \; B_1, \dots, B_k \in \mathcal{B}. \tag{3.31}$$

Omdat verdelingen door hun verdelingsfuncties vastgelegd worden, geldt (3.31) dan en slechts dan als

$$F_X(x) = F_{X_1}(x_1) \dots F_{X_k}(x_k), \quad x = (x_1, \dots, x_k) \in \mathbf{R}^k. \tag{3.32}$$

Voor het discrete geval geldt (3.32) dan en slechts dan als

$$P(X_1 = x_1, \dots, X_k = x_k) = P(X_1 = x_1) \dots P(X_k = x_k), \quad x_1, \dots, x_k \in \mathbf{R}, \tag{3.33}$$

en voor het continue geval geldt (3.32) dan en slechts dan als

$$f_X(x_1, \dots, x_k) = f_{X_1}(x_1) \dots f_{X_k}(x_k), \quad x_1, \dots, x_k \in \mathbf{R}. \tag{3.34}$$

Zij X een stochastische k-vector gedefinieerd op $(\Omega, \mathcal{A}, P)$ en $g : \mathbf{R}^k \to \mathbf{R}$ een functie. Als g fatsoenlijk is, dan is $Y = g \circ X$ met $Y(\omega) = g(X(\omega))$ een stochastische grootheid. Dan geldt in het discrete geval

$$P(Y = y) = \sum_{\substack{x \\ g(x)=y}} P(X = x), \quad y \in \mathbf{R}, \tag{3.35}$$

en in het continue geval

$$F_Y(y) = \int \underset{g(x)\leq y}{\dots} \int f_X(x) dx_1 \dots dx_k, \quad y \in \mathbf{R}. \tag{3.36}$$

Als $k = 2$ en $g(x_1, x_2) = x_1 + x_2$, dan is $Y = X_1 + X_2$. Als X_1 en X_2 o.o. zijn, dan geldt

$$P(Y = y) = P(X_1 + X_2 = y) = \sum_{x_1} P(X_1 = x_1) P(X_2 = y - x_1), \quad y \in \mathbf{R}, \tag{3.37}$$

in het discrete en

$$f_Y(y) = f_{X_1+X_2}(y) = \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1, \quad y \in \mathbf{R}, \tag{3.38}$$

in het continue geval.

Voorbeeld 3.2. Zijn X_1 en X_2 o.o. en zij X_i bin(n_i, p) verdeeld, $i = 1, 2$.
Dan geldt volgens (3.37)

$$\begin{aligned} &P(Y = y) \\ &\quad = \sum_{x_1} \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1 - x_1} \binom{n_2}{y - x_1} p^{y - x_1} (1-p)^{n_2 - y + x_1}, \\ &\quad y = 0, \dots, n_1 + n_2. \end{aligned} \tag{3.39}$$

Omdat X_i opgevat kan worden als het aantal successen in n_i o.o. alternatieven met succeskans p, is Y het aantal successen in $n_1 + n_2$ o.o. alternatieven met succeskans p, is dus Y bin$(n_1 + n_2, p)$ verdeeld en geldt

$$P(Y = y) = \binom{n_1 + n_2}{y} p^y (1-p)^{n_1 + n_2 - y}. \tag{3.40}$$

Wanneer we (3.39) en (3.40) vergelijken zien we dat we m.b.v. kanstheoretische argumenten hebben bewezen dat

$$\sum_{x_1} \binom{n_1}{x_1} \binom{n_2}{y - x_1} = \binom{n_1 + n_2}{y}. \tag{3.41}$$

Voorbeeld 3.3. Zij X $\mathcal{N}(0,1)$ verdeeld en $g : \mathbf{R} \to \mathbf{R}$ de affiene transformatie $g(x) = \mu + \sigma x$ met $\mu \in \mathbf{R}$, $\sigma > 0$. Nu volgt voor $Y = g(X)$ uit (3.36) en (3.24)

$$F_Y(y) = \int\limits_{\mu+\sigma x \leq y} \varphi(x)dx = \int\limits_{-\infty}^{\frac{y-\mu}{\sigma}} \varphi(x)dx = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

en dus dat Y $\mathcal{N}(\mu, \sigma^2)$ verdeeld is.

Voorbeeld 3.4. Zijn X_1 en X_2 o.o. en $\mathcal{N}(\mu, \sigma^2)$ respectievelijk $\mathcal{N}(\nu, \tau^2)$ verdeeld. Volgens het voorgaande voorbeeld bestaan er U_1 en U_2 o.o. met $X_1 = \mu + \sigma U_1$ en $X_2 = \nu + \tau U_2$. Dan geldt dat $Y = X_1 + X_2 = \mu + \nu + \sigma U_1 + \tau U_2$. We zullen nu laten zien dat $Z = \sigma U_1 + \tau U_2$ $\mathcal{N}(0, \sigma^2 + \tau^2)$ verdeeld is en dus (m.b.v. voorbeeld 3.3), dat $Y = \mu + \nu + Z$ $\mathcal{N}(\mu+\nu,\ \sigma^2+\tau^2)$ verdeeld is. Volgens (3.38) is voor zekere $c > 0$

$$\begin{aligned}
f_Z(z) &= \int\limits_{-\infty}^{\infty} \frac{1}{\sigma}\varphi\left(\frac{x_1}{\sigma}\right)\frac{1}{\tau}\varphi\left(\frac{z-x_1}{\tau}\right)dx_1 \\
&= \frac{1}{2\pi\sigma\tau}\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}\{x_1^2/\sigma^2 + (x_1^2 - 2x_1 z + z^2)/\tau^2\}}dx_1 \\
&= \frac{1}{2\pi\sigma\tau}\int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}(\sigma^{-2}+\tau^{-2})(x_1 - z\sigma^2(\sigma^2+\tau^2)^{-1})^2 - \frac{1}{2}z^2(\sigma^2+\tau^2)^{-1}}dx_1 \\
&= c\, e^{-\frac{1}{2}z^2(\sigma^2+\tau^2)^{-1}}.
\end{aligned}$$

Omdat f_Z een dichtheid moet zijn en dus aan (3.8) moet voldoen, zien we dat $c = \{2\pi(\sigma^2+\tau^2)\}^{-1/2}$ en dat Z inderdaad $\mathcal{N}(0, \sigma^2+\tau^2)$ verdeeld is.

Opgaven.

3.1. X is een stochastische grootheid met verdelingsfunctie F en a an b zijn reële getallen met $a < b$. Druk de kansen $P(X < a)$, $P(X > a)$, $P(X \geq a)$, $P(a \leq X \leq b)$, $P(a \leq X < b)$, $P(a < X < b)$ en $P(X = a)$ uit in F.

3.2. De stochastische grootheid X heeft een continue verdeling met dichtheid f. Laat zien dat $Y = a + bX$, waarin a en $b > 0$ constanten zijn, ook continu verdeeld is en bepaal de dichtheid van Y.

3.3. De continu verdeelde stochastische grootheid X heeft een dichtheid f.
  a. Bepaal de dichtheid van $Y = X^2$.
  b Bepaal de dichtheid van $Z = \sqrt{X}$ als gegeven is dat $f(x) = 0$ voor $x \leq 0$.

4. Verwachting en variantie

Zij S een eindige deelverzameling van $\mathbf{R}$ en X een stochastische grootheid met $P(X \in S) = 1$. Laat $X_1, \ldots, X_n$ o.o. en identiek verdeelde stochastische grootheden zijn alle met dezelfde verdeling als X. We bekijken voor een realisatie $(x_1, \ldots, x_n)$ van $(X_1, \ldots, X_n)$ het gemiddelde $\frac{1}{n}(x_1 + \ldots + x_n)$. Hiervoor geldt, in de notatie van hoofdstuk 1,

$$\frac{1}{n}(x_1 + \ldots + x_n) = \sum_{x \in S} \frac{x}{n} \#(x_i = x) = \sum_{x \in S} x f_n(\{x\}). \tag{4.1}$$

Volgens de empirische wet van de grote aantallen zal, voor grote waarden van n, $\frac{1}{n}(x_1 + \ldots + x_n)$ dus dichtbij $\sum_{x \in S} xP(X = x)$ liggen. Daarom definiëren we het gemiddelde of de verwachting van X als volgt.

Definitie 4.1. Als $P(X \in S) = 1$ met S eindig, dan is de **verwachting** EX van X gedefinieerd door

$$EX = \sum_{x \in S} xP(X = x). \tag{4.2}$$

Voorbeeld 4.1. Als X het aantal ogen is dat geworpen wordt met een zuivere dobbelsteen, dan is $EX = \sum_{x=1}^{6} x\frac{1}{6} = \frac{1}{6} \cdot \frac{1}{2} \cdot 6 \cdot 7 = 3\frac{1}{2}$.

Voorbeeld 4.2. Als X bin(n,p) verdeeld is, dan is

$$\begin{aligned} EX &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^{n} np \binom{n-1}{x-1} p^{x-1} (1-p)^{n-1-(x-1)} \\ &= np \sum_{y=0}^{n-1} P(Y = y) = np, \end{aligned}$$

waarbij Y bin$(n-1,p)$ verdeeld is.

Stelling 4.1. Zijn X en Y stochastische grootheden met $P(X \in S) = P(Y \in S) = 1$ en S eindig.

A. Als $P(X \geq 0) = 1$, dan $EX \geq 0$.
B. Als $P(X \geq Y) = 1$, dan $EX \geq EY$.
C. Voor alle $a \in \mathbf{R}$ geldt $E(aX) = aEX$.
D. $E(X + Y) = EX + EY$.

Bewijs.

$$\begin{aligned} E(X+Y) &= \sum_z zP(X+Y=z) = \sum_z z \sum_x P(X=x,\ Y=z-x) \\ &= \sum_x \sum_y (x+y) P(X=x,\ Y=y) \\ &= \sum_x x \sum_y P(X=x,\ Y=y) + \sum_y y \sum_x P(X=x,\ Y=y) \\ &= \sum_x xP(X=x) + \sum_y yP(Y=y) = EX + EY. \end{aligned}$$

De rest is eenvoudig. $\square$

Eigenschap C en D tonen aan dat "verwachting nemen" een lineaire operator is. Hieruit volgt dat, als $X_1, \ldots, X_n$ stochastische grootheden zijn op één kansruimte gedefinieerd met $P(X_i \in S) = 1$ voor een eindige S, dan geldt

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E X_i, \quad a_1, \ldots, a_n \in \mathbf{R}. \tag{4.3}$$

Voorbeeld 4.2 (vervolg). Als $X_1, \ldots, X_n$ o.o. alternatief verdeelde stochastische grootheden zijn met $P(X_i = 1) = p,\ P(X_i = 0) = 1 - p,\ i = 1, \ldots, n$, dan is $X = X_1 + \ldots + X_n$ bin(n, p) verdeeld. Verder is wegens (4.3)

$$EX = E(X_1 + \ldots + X_n) = EX_1 + \ldots + EX_n = nEX_1 = n(0 \cdot (1-p) + 1 \cdot p) = np.$$

Dit toont op een eenvoudige manier aan dat np de verwachting is van **iedere** stochastische grootheid met een bin(n, p) verdeling.

Voorbeeld 4.3. Zij X_i het aantal rode knikkers bij de i-de trekking tijdens het trekken van een aselecte steekproef van omvang n zonder teruglegging uit een vaas met R rode en $N-R$ witte knikkers. Dan is X_i alternatief verdeeld met parameter $\frac{R}{N}$ en $X = X_1 + \ldots + X_n$ heeft een hyp(N, R, n) verdeling. Dus

$$EX = EX_1 + \ldots + EX_n = nEX_1 = \frac{nR}{N}.$$

We breiden nu definitie 4.1 als volgt uit.

Definitie 4.2. Als X een discreet verdeelde stochastische grootheid is met

$$\sum_x |x| P(X = x) < \infty, \tag{4.4}$$

dan wordt de verwachting EX van X gedefinieerd door

$$EX = \sum_x x P(X = x). \tag{4.5}$$

Als X een continu verdeelde stochastische grootheid is met

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty, \tag{4.6}$$

dan wordt de verwachting EX van X gedefinieerd door

$$EX = \int_{-\infty}^{\infty} x f_X(x) dx. \tag{4.7}$$

Zonder bewijs vermelden we dat ook voor bovenstaande gevallen stelling 4.1 en formule (4.3) gelden.

Voorbeeld 4.4. Als X Poisson(λ) verdeeld is, dan geldt

$$EX = \sum_{k=0}^{\infty} k\ e^{-\lambda}\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Voorbeeld 4.5. Als X Pascal(p) verdeeld is, dan geldt

$$\begin{aligned}
EX &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p = -p\lim_{K\to\infty}\sum_{k=1}^{K}\frac{d}{dp}(1-p)^k = -p\lim_{K\to\infty}\frac{d}{dp}\sum_{k=1}^{K}(1-p)^k p \\
&= -p\lim_{K\to\infty}\frac{d}{dp}\frac{(1-p)-(1-p)}{p}^{K+1} \\
&= -p\lim_{K\to\infty}\frac{-1+(K+1)(1-p)^K}{p} - \frac{(1-p)-(1-p)^{K+1}}{p^2} \\
&= -p\left[\frac{-1}{p} - \frac{1-p}{p^2}\right] = \frac{1}{p}.
\end{aligned}$$

Voorbeeld 4.6. Als X hom(a,b) verdeeld is, dan geldt

$$EX = \int_{-\infty}^{\infty}\frac{x}{b-a}1_{(a,b)}(x)dx = \frac{1}{b-a}\int_a^b x\ dx = \frac{1}{b-a}\frac{1}{2}(b^2-a^2) = \frac{1}{2}(a+b).$$

Voorbeeld 4.7. Als X $\Gamma(n.\lambda)$ verdeeld is, dan geldt

$$\begin{aligned}
EX &= \int_{-\infty}^{\infty} x\frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}1_{(0,\infty)}(x)dx \\
&= \frac{1}{(n-1)!}\int_0^{\infty} y^n e^{-y}\frac{1}{\lambda}dy = \frac{n!}{\lambda(n-1)!} = \frac{n}{\lambda}.
\end{aligned}$$

Voorbeeld 4.8. Als U $\mathcal{N}(0,1)$ verdeeld is, dan geldt

$$EU = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u\ e^{-\frac{1}{2}u^2}du = 0,$$

omdat $\int_{-\infty}^{\infty}|u|e^{-\frac{1}{2}u^2}du < \infty$ is en $u\ e^{-\frac{1}{2}u^2}$ een oneven functie is. Nu volgt, dat $X = \mu + \sigma U$ $\mathcal{N}(\mu,\sigma^2)$ verdeeld is en verwachting

$$EX = \mu + \sigma EU = \mu$$

heeft.

Als X een stochastische k-vector is en $Y = g(X)$ een stochastische grootheid, is het voor de berekening van EY volgens definitie 4.2 niet nodig eerst de verdeling van Y te bepalen wegens de volgende stelling.

Stelling 4.2. Als in de hierboven beschreven situatie X discreet verdeeld is, dan geldt

$$EY = \sum_x g(x) P(X = x), \tag{4.8.}$$

mits deze som absoluut convergent is. Is X continu verdeeld, dan is

$$EY = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x) f_X(x) dx_1 \dots dx_k, \tag{4.9.}$$

mits deze integraal absoluut convergent is.

Bewijs voor het eindige discrete geval.

$$\begin{aligned} EY &= \sum_y y P(g(X) = y) = \sum_y y \Big\{ \sum_{\substack{x \\ g(x)=y}} P(X = x) \Big\} \\ &= \sum_y \sum_{\substack{x \\ g(x)=y}} g(x) P(X = x) = \sum_x g(x) P(X = x). \end{aligned}$$

▯

Voorbeeld 4.9. Zij X $\mathcal{N}(0,1)$ verdeeld, dus $k = 1$, en $Y = X^2$. Dan

$$\begin{aligned} EY = EX^2 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{1}{2}x^2} dx = \frac{-x}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Big]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx \\ &= \int_{-\infty}^{\infty} \varphi(x) dx = 1. \end{aligned}$$

Definitie 4.3. Als X een stochastische grootheid is waarvoor EX^2 welgedefinieerd is, dan heet

$$\operatorname{var} X = \sigma^2(X) = E(X - EX)^2$$

de **variantie** van X en $\sigma(X) = \sqrt{\operatorname{var} X}$ de **standaardafwijking** van X.

Merk op dat $2|x| \leq x^2 + 1$ en dus $E|X| \leq \frac{1}{2}EX^2 + \frac{1}{2}$. Dit betekent dat eindigheid van EX^2 impliceert dat ook EX welgedefinieerd is.

Stelling 4.3. Als $EX^2 < \infty, \quad a, b \in \mathbf{R}$, dan

$$\sigma^2(X) = EX^2 - (EX)^2, \tag{4.10}$$

$$\sigma^2(aX + b) = a^2\sigma^2(X). \tag{4.11}$$

Bewijs.

$$\begin{aligned} \sigma^2(X) &= E(X - EX)^2 = E\{X^2 - 2(EX)X + (EX)^2\} \\ &= EX^2 - 2(EX)EX + (EX)^2 = EX^2 - (EX)^2. \end{aligned}$$

$$\begin{aligned} \sigma^2(aX + b) &= E(aX + b - E(aX + b))^2 \\ &= E(a(X - EX))^2 = a^2E(X - EX)^2 = a^2\sigma^2(X). \end{aligned}$$

∎

Voorbeeld 4.10. Als X Poisson(λ) verdeeld is, dan is

$$EX(X-1) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = \sum_{x=2}^{\infty} e^{-\lambda}\frac{\lambda^{x-2}}{(x-2)!}\lambda^2 = \lambda^2.$$

Dus $\sigma^2(X) = EX(X-1) + EX - (EX)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Voorbeeld 4.11. Als X $\mathcal{N}(0,1)$ verdeeld is, dan geldt volgens voorbeeld 4.9 dat $\sigma^2(X) = EX^2 - (EX)^2 = 1$. Verder is $Y = \mu + \sigma X$ $\mathcal{N}(\mu, \sigma^2)$ verdeeld met $\sigma^2(Y) = \sigma^2(\mu + \sigma X) = \sigma^2\sigma^2(X) = \sigma^2$.

Definitie 4.4. Als X en Y stochastische grootheden zijn, die op één kansruimte gedefinieerd zijn en waarvoor $E|X|$, $E|Y|$ en $E|X - EX||Y - EY|$ eindig zijn, dan heet

$$\mathrm{cov}(X, Y) = E(X - EX)(Y - EY)$$

de **covariantie** van X en Y.

Stelling 4.4. Als X en Y stochastische grootheden zijn met $EX^2 < \infty$ en $EY^2 < \infty$, dan is $\mathrm{cov}(X, Y)$ gedefinieerd en

$$\mathrm{cov}(X, Y) = EXY - EXEY. \tag{4.12}$$

Zijn bovendien X en Y o.o. dan geldt

$$\mathrm{cov}(X, Y) = 0. \tag{4.13}$$

Het omgekeerde geldt niet.

Bewijs. Omdat $2|(x-EX)(y-EY)| \leq (x-EX)^2+(y-EY)^2$, is $\operatorname{cov}(X,Y)$ welgedefinieerd en geldt

$$\begin{aligned} E(X-EX)(Y-EY) &= E(XY - XEY - (EX)Y + EXEY) \\ &= EXY - EXEY. \end{aligned}$$

Als X en Y een (eindige) discrete verdeling hebben en o.o. zijn, dan geldt

$$\begin{aligned} EXY &= \sum_x \sum_y xy\ P(X=x,\ Y=y) = \sum_x \sum_y xy\ P(X=x)P(Y=y) \\ &= \left(\sum_x x\ P(X=x)\right)\left(\sum_y y\ P(Y=y)\right) = EXEY. \end{aligned}$$

Voor het algemene discrete geval en het continue geval bewijzen we (4.13) niet. ∎

Stelling 4.5. Als $X_1, \ldots, X_n$ stochastische grootheden zijn, die op één kansruimte gedefinieerd zijn, dan is

$$(4.14)\qquad \begin{aligned} &\operatorname{var}\left(\sum_{i=1}^n \alpha_i X_i\right) \\ &= \sum_{i=1}^n \alpha_i^2 \operatorname{var} X_i + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n \alpha_i\alpha_j \operatorname{cov}(X_i, X_j) \end{aligned}$$

voor alle $\alpha_1, \ldots, \alpha_n \in \mathbf{R}$. Als $X_1, \ldots, X_n$ o.o. zijn, geldt dus

$$(4.15)\qquad \operatorname{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{var} X_i.$$

Bewijs. Uitschrijven. ∎

Voorbeeld 4.12. Als $X_1, \ldots, X_n$ o.o. alternatief verdeeld zijn met succeskans p, dan is $X = \sum_{i=1}^n X_i$ $\operatorname{bin}(n,p)$ verdeeld en geldt volgens (4.15)

$$\operatorname{var} X = \sum_{i=1}^n \operatorname{var} X_i = n \operatorname{var} X_1 = n(EX_1^2 - (EX_1)^2) = n(EX_1 - (EX_1)^2) = np(1-p).$$

Stelling 4.6. (ongelijkheid van Cauchy-Schwarz). Als X en Y stochastische grootheden zijn met $EX^2 < \infty$ en $EY^2 < \infty$, dan geldt

$$(4.16)\qquad |\operatorname{cov}(X,Y)| \leq \sigma(X)\sigma(Y).$$

Bewijs. Voor elke $\alpha \in \mathbf{R}$ geldt volgens (4.14)

$$0 \leq \operatorname{var}(X + \alpha Y) = \operatorname{var} X + \alpha^2 \operatorname{var} Y + 2\alpha \operatorname{cov}(X, Y).$$

Deze kwadratische functie in α kan alleen niet negatief zijn als de discriminant niet positief is, d.w.z. als

$$4 \operatorname{cov}^2(X, Y) - 4 \operatorname{var} X \operatorname{var} Y \leq 0.$$

Maar dit is equivalent met (4.16). ▯

Opgaven

4.1. Bereken EX als X negatief binomiaal verdeeld is met parameters n en p.

4.2. Stel X heeft een Laplace of dubbel exponentiële verdeling met dichtheid

$$f_X(x) = \frac{1}{2\sigma} e^{-\left|\frac{x-\mu}{\sigma}\right|}, \quad \mu \in \mathbf{R}, \ \sigma > 0.$$

Bepaal EX.

4.3. Beschouw voor $\epsilon > 0$ de functies $\epsilon \, 1_{[\epsilon,\infty)}(|\cdot|)$ en $|\cdot|$.
Bewijs de ongelijkheid van Markov

$$P(|X| \geq \epsilon) \leq \frac{E|X|}{\epsilon}.$$

Literatuur

Dit eerste deel van het dictaat is sterk geïnspireerd door de eerste vier hoofdstukken van

Fabius, J. (1984), Inleiding kansrekening, Syllabus RUL.

Engelstalige en uitgebreidere besprekingen van de behandelde onderwerpen kunnen gevonden worden in

Chung, K.L. (1974), Elementary probability theory with stochastic processes, Springer, Berlin.

Mood, A.M., F.A. Graybill and **D.C. Boes** (1974), Introduction to the theory of statistics, McGraw-Hill, New York.

Deel II. S t a t i s t i e k

5. De kansrekening als basis voor de statistiek

In de kansrekening worden kansruimten en daarop gedefinieerde stochastische grootheden met een **bekende** kansverdeling bestudeerd. Wanneer deze kansverdeling **onbekend** is en we een aantal realisaties van de stochastische grootheden ter beschikking hebben, kan de statistiek ons helpen uitspraken te doen over deze kansverdeling. Er zijn twee belangrijke typen van uitspraken, die in de **schattingstheorie** respectievelijk **toetsingstheorie** thuis horen en die we als volgt kunnen illustreren:

- "de kansverdeling is deze",
- "de kansverdeling is **niet** deze".

Het zal duidelijk zijn, dat de kansrekening het fundament is, waarop de statistiek rust.

6. Schattingstheorie

Voorbeeld 6.1. Beschouw een grote partij transistoren van een zeker type.

De levensduur X van een aselect gekozen transistor uit deze partij is stochastisch en wordt verondersteld $\exp(\lambda)$ verdeeld te zijn. Dan is $EX = \frac{1}{\lambda}$, $P(X > 1) = e^{-\lambda}$ etc. In feite hebben we te maken met een grote collectie kansverdelingen als de parameter $\lambda > 0$ **onbekend** is. Wanneer we iets willen zeggen over λ moeten we ons baseren op **waarnemingen**. Deze waarnemingen moeten dan **realisaties** zijn van $X_1, \ldots, X_n$. Hierbij zijn $X_1, \ldots, X_n$ o.o. identiek verdeelde stochastische grootheden alle $\exp(\lambda)$ verdeeld. $X_1, \ldots, X_n$ zijn de stochastische levensduren van n aselect gekozen transistoren uit de partij. We willen nu λ of $\frac{1}{\lambda}$, $e^{-\lambda}$ etc. **schatten** op basis van de realisaties $x_1, \ldots, x_n$ van $X_1, \ldots, X_n$.

Voorbeeld 6.2. Verondersteld wordt dat de logaritme van de prijs van een bepaald aandeel op een zekere dag $\mathcal{N}(\mu, \sigma^2)$ verdeeld is. Hierbij is σ^2 bekend en μ onbekend. We willen μ schatten. Daarvoor kunnen we ons baseren op realisaties $x_1, \ldots, x_n$ van $X_1, \ldots, X_n$, de logaritmen van de prijzen van het aandeel op n zó ver van elkaar af liggende dagen, dat we redelijkerwijs kunnen veronderstellen, dat $X_1, \ldots, X_n$ o.o. zijn.

Dit zijn voorbeelden van de volgende algemene situatie.

$X_1, \ldots, X_n$ zijn onderling onafhankelijke en identiek verdeelde stochastische grootheden met kansverdeling P_θ. Hierbij is θ onbekend. Bekend is wel dat $\theta \in \Theta$, waarbij $\Theta \subset \mathbf{R}$ een open interval is. Zij $\psi : \Theta \to \mathbf{R}$. We willen $\psi(\theta)$ schatten. Gegeven zijn de waarnemingen $x_1, \ldots, x_n$, die realisaties zijn van $X_1, \ldots, X_n$. Als $t : \mathbf{R}^n \to \mathbf{R}$, dan heet

$t(x_1, \ldots, x_n)$ een **schatting** (estimate) van $\psi(\theta)$ en

$T = t(X_1, \ldots, X_n)$ de bijbehorende **schatter** (estimator) van $\psi(\theta)$.

Merk op dat T een stochastische grootheid is waarvan de verdeling rond $\psi(\theta)$ van belang is. Deze verdeling is afhankelijk van θ, $\psi(\cdot)$ en $t(\cdot)$, de functie die de schattingsmethode beschrijft.

Voorbeeld 6.1 (vervolg). $\overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ is een schatting van $\frac{1}{\lambda}$. Over de vraag, hoever bij zekere realisaties $x_1, \dots x_n$ de schatting $\overline{x}_n$ van $\frac{1}{\lambda}$ af ligt, valt **niets** te zeggen, omdat λ onbekend is. Van de schatter $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ valt bijvoorbeeld op te merken dat $E_\lambda \overline{X}_n = \frac{1}{\lambda}$ (met E_λ geven we aan dat de verwachting wordt genomen met P_λ als de kansverdeling van $X_1, \dots, X_n$).

Voorbeeld 6.2 (vervolg). Ook hier lijkt $\overline{x}_n$ een redelijke schatting. Weer valt over $\overline{x}_n - \mu$ niets te zeggen. Wel is eenvoudig na te gaan dat $\overline{X}_n - \mu$ onder P_μ $\mathcal{N}(0, \sigma^2/n)$ verdeeld is.

Als $\theta_0 \in \Theta$ een "vaste" parameterwaarde is en $t : \mathbf{R}^n \to \mathbf{R}$ wordt gedefinieerd door $t(x_1, \dots, x_n) = \psi(\theta_0)$, dan schat $T = t(X_1, \dots, X_n)$ de waarde van $\psi(\theta)$ uitstekend als $\theta = \theta_0$. Zodra echter $\theta \neq \theta_0$ of preciezer $\psi(\theta) \neq \psi(\theta_0)$, dan schat T de waarde van $\psi(\theta)$ ronduit slecht. We zullen daarom van een schatter eisen dat deze "gemiddeld genomen" de juiste waarde schat.

Definitie 6.1. De schatter T van $\psi(\theta)$ heet **zuiver** (engels: unbiased; vlaams: onvertekend) als

$$E_\theta T = \psi(\theta) \text{ voor alle } \theta \in \Theta. \tag{6.1}$$

Voorbeeld 6.2 (vervolg). $\overline{X}_n$ is een zuivere schatter van μ omdat $E_\mu \overline{X}_n = \mu$. Als $\alpha_1, \dots, \alpha_n$ reële getallen zijn met $\alpha_1 + \ldots + \alpha_n = 1$, dan is ook

$$T = \alpha_1 X_1 + \ldots + \alpha_n X_n \tag{6.2}$$

zuiver. Immers,

$$\begin{aligned} E_\mu T &= \alpha_1 E_\mu X_1 + \ldots + \alpha_n E_\mu X_\mu \\ &= \alpha_1 \mu + \ldots + \alpha_n \mu = (\alpha_1 + \ldots + \alpha_n)\mu = \mu. \end{aligned}$$

Aan dit voorbeeld zien we dat er verscheidene zuivere schatters kunnen zijn. Om hiertussen onze schatter te kiezen, hanteren we het kriterium van het gemiddelde kwadratische verschil tussen de schatter en de te schatten waarde oftewel de verwachte kwadratische fout:

$$E_\theta \left(T - \psi(\theta)\right)^2. \tag{6.3}$$

We zoeken dan onder de zuivere schatters T die schatter waarvoor (6.3) minimaal is voor alle $\theta \in \Theta$. Merk op, dat voor zuivere schatters (6.3) gelijk is aan $\mathrm{var}_\theta T$. We zoeken dus zuivere schatters met minimale variantie.

Voorbeeld 6.2 (vervolg). Het is niet zo moeilijk onder de zuivere **lineaire** (zie (6.2)) schatters van μ de schatter te vinden met minimale variantie. Zij T als in (6.2) dan is de variantie

$$\begin{aligned} \mathrm{var}_\mu T = \mathrm{var}_\mu \sum_{i=1}^n \alpha_i X_i &= \sum_{i=1}^n \mathrm{var}_\mu \alpha_i X_i \\ &= \sum_{i=1}^n \alpha_i^2 \mathrm{var}_\mu X_i = \sigma^2 \sum_{i=1}^n \alpha_i^2. \end{aligned} \tag{6.4}$$

Neem even aan dat alle α_i verschillend zijn. Zij nu A een discrete stochastische grootheid met $P(A = \alpha_i) = 1/n$ voor $i = 1, \ldots, n$. Dan is $EA = \frac{1}{n}\sum_{i=1}^n \alpha_i$, $EA^2 = \frac{1}{n}\sum_{i=1}^n \alpha_i^2$ en $\operatorname{var} A = EA^2 - (EA)^2 \geq 0$. Dus

$$\frac{1}{n}\sum_{i=1}^n \alpha_i^2 = EA^2 \geq (EA)^2 = \left(\frac{1}{n}\sum_{i=1}^n \alpha_i\right)^2 \quad \text{en} \quad \sum_{i=1}^n \alpha_i^2 \geq \frac{1}{n}\left(\sum_{i=1}^n \alpha_i\right)^2 = \frac{1}{n}.$$

Het is eenvoudig in te zien dat dit idee uitgebreid kan worden tot het geval waarin er gelijken onder de α_i voorkomen. Hieruit en uit (6.4) volgt dat

$$E_\mu(T - \mu)^2 \geq \frac{\sigma^2}{n} = E_\mu(\overline{X}_n - \mu)^2.$$

Dus $\overline{X}_n$ heeft de kleinste verwachte kwadratische fout onder alle zuivere lineaire schatters van μ.

Definitie 6.2. Als T een zuivere schatter is van $\psi(\theta)$ zó dat voor elke zuivere schatter $\tilde{T}$ van $\psi(\theta)$

$$E_\theta(T - \psi(\theta))^2 \leq E_\theta(\tilde{T} - \psi(\theta))^2, \quad \theta \in \Theta, \tag{6.5}$$

dan heet T **uniform minimum variantie zuiver** (uniformly minimum variance unbiased; UMVU).

Een uniform minimum variantie zuivere schatter hoeft niet te bestaan, maar als zij bestaat, is zij “in essentie” uniek zoals blijkt uit de volgende stelling.

Stelling 6.1. Als T en $\tilde{T}$ beide uniform minimum variantie zuiver zijn, dan geldt

$$P_\theta(T = \tilde{T}) = 1 \text{ voor alle } \theta \in \Theta. \tag{6.6}$$

Bewijs. Zij $\sigma_\theta^2 = \operatorname{var}_\theta T = \operatorname{var}_\theta \tilde{T}$, $\theta \in \Theta$. Voor elke $\alpha \in \mathbf{R}$ is ook $\alpha T + (1-\alpha)\tilde{T}$ een zuivere schatter van $\psi(\theta)$ en dus

$$\sigma_\theta^2 \leq \operatorname{var}_\theta\left(\alpha T + (1-\alpha)\tilde{T}\right) = \{\alpha^2 + (1-\alpha)^2\}\sigma_\theta^2 + 2\alpha(1-\alpha)\operatorname{cov}_\theta(T, \tilde{T}).$$

Hieruit volgt $2\alpha(1-\alpha)\sigma_\theta^2 \leq 2\alpha(1-\alpha)\operatorname{cov}_\theta(T, \tilde{T})$. Omdat $\alpha(1-\alpha)$ zowel positief als negatief kan zijn, impliceert dit $\operatorname{cov}_\theta(T, \tilde{T}) = \sigma_\theta^2$ en dus

$$E_\theta(T - \tilde{T})^2 = \operatorname{var}_\theta(T - \tilde{T}) = \sigma_\theta^2 + \sigma_\theta^2 - 2\operatorname{cov}_\theta(T, \tilde{T}) = 0.$$

Hieruit volgt (6.6). ▯

Intuïtief is duidelijk dat in het algemeen de varianties $E_\theta(T - \psi(\theta))^2$, $\theta \in \Theta$, van een zuivere schatter T van $\psi(\theta)$ niet willekeurig klein kunnen zijn. Immers de $X_1, \ldots, X_n$

bevatten slechts “eindig” veel informatie over θ respectievelijk $\psi(\theta)$. Dit intuïtieve inzicht is formeel gemaakt in de volgende stelling, waarvoor we enige notatie invoeren.

$$l_\theta(x) = \begin{cases} \log P_\theta(X_1 = x) & \text{als } X_1 \text{ discreet verdeeld is} \\ \log f_\theta & \text{als } X_1 \text{ continu verdeeld is met dichtheid } f_\theta \end{cases} \tag{6.7}$$

$$\dot{l}_\theta(x) = \frac{\partial}{\partial\theta} l_\theta(x) \tag{6.8}$$

heet de score functie, als zij bestaat, en

Definitie 6.3.

$$I(\theta) = E_\theta[\dot{l}_\theta(X_1)]^2 \tag{6.9}$$

heet de Fisher informatie over θ bevat in één waarneming.

Stelling 6.2. Cramér-Rao ongelijkheid (1946)
$X_1, \ldots, X_n$ zijn onderling onafhankelijke en identiek verdeelde stochastische grootheden met kansverdeling P_θ. θ is een onbekend element van de parameterverzameling Θ, een open interval. Zij T een zuivere schatter van $\psi(\theta)$, waarbij ψ differentieerbaar is naar θ. Als $\dot{l}_\theta$ bestaat, $I(\theta)$ en $E_\theta T^2$ eindig zijn,

$$\psi'(\theta) = \frac{d}{d\theta}\psi(\theta) = E_\theta T \sum_{i=1}^{n} \dot{l}_\theta(X_i), \tag{6.10}$$

$$0 = E_\theta\, \dot{l}_\theta(X_1), \tag{6.11}$$

dan

$$E_\theta(T - \psi(\theta))^2 = \mathrm{var}_\theta T \geq \frac{(\psi'(\theta))^2}{nI(\theta)}. \tag{6.12}$$

Bewijs. Zij $S_n = \sum_{i=1}^n \dot{l}_\theta(X_i)$. Uit (6.11) volgt $E_\theta S_n = 0$ en $\mathrm{var}_\theta S_n = nI(\theta)$. Hiervoor is de eindigheid van $I(\theta)$ nodig. Samen met de eindigheid van $E_\theta T^2$ geeft dit volgens Stelling 4.4

$$\mathrm{cov}_\theta(T, S_n) = E_\theta\, T\, S_n.$$

Hieruit volgt met (6.10) en Stelling 4.6

$$(\psi'(\theta))^2 = (E_\theta\, T\, S_n)^2 = \mathrm{cov}_\theta^2(T, S_n) \leq \mathrm{var}_\theta\, T \mathrm{var}_\theta\, S_n = nI(\theta)\mathrm{var}_\theta\, T,$$

waarmee de stelling bewezen is. ∎

Merk op dat de Cramér-Rao ongelijkheid niet voor elke zuivere schatter T geldt. Met name (6.10) moet steeds geverifieerd worden. Als de X_i (eindige) discrete verdelingen hebben, is dit niet lastig:

$$\begin{aligned}
\psi'(\theta) &= \frac{d}{d\theta}E_\theta T = \frac{d}{d\theta}\sum_{x_1,\ldots,x_n} t(x_1,\ldots,x_n)\prod_{i=1}^n P_\theta(X_i = x_i) \\
&= \sum_{x_1,\ldots,x_n} t(x_1,\ldots,x_n)\frac{d}{d\theta}e^{\sum_{i=1}^n l_\theta(x_i)} \\
&= \sum_{x_1,\ldots,x_n} t(x_1,\ldots,x_n)\sum_{i=1}^n \dot{l}_\theta(x_i)e^{\sum_{i=1}^n l_\theta(x_i)} \\
&= \sum_{x_1,\ldots,x_n} t(x_1,\ldots,x_n)\sum_{i=1}^n \dot{l}_\theta(x_i)\prod_{i=1}^n P_\theta(X_i = x_i) \\
&= E_\theta T\sum_{i=1}^n \dot{l}_\theta(X_i).
\end{aligned}$$

Voorbeeld 6.3. Stel $X_1,\ldots,X_n$ zijn o.o. en alternatief verdeeld met succeskans p. Dan $\theta = p$, $l_p(x) = \log(p^x(1-p)^{1-x}) = x\log p + (1-x)\log(1-p)$ en $\dot{l}_p(x) = x/p - (1-x)/(1-p) = (x-p)/(p(1-p))$. Dus

$$E_p\dot{l}_p(X_1) = 0 \;\text{ en }\; I(p) = E_p\left(\frac{X_1 - p}{p(1-p)}\right)^2 = \frac{\mathrm{var}_p X_1}{p^2(1-p)^2} = \frac{1}{p(1-p)}.$$

Als E_pT^2 eindig is, zijn alle voorwaarden van Stelling 6.2 vervuld en dus geldt voor elke zuivere schatter T van p met $E_pT^2 < \infty$, dat

$$\mathrm{var}_p T \geq \frac{p(1-p)}{n}. \tag{6.13}$$

Omdat $\overline{X}_n$ een zuivere schatter van p is met variantie $p(1-p)/n$, volgt uit (6.13) dat $\overline{X}_n$ de uniform minimum variantie zuivere schatter van p is.

Voorbeeld 6.2 (vervolg). Laat $X_1,\ldots,X_n$ o.o. en $\mathcal{N}(\mu,\sigma^2)$ verdeeld zijn met $\theta = \mu$ onbekend en $\sigma^2 > 0$ bekend. Nu is $l_\mu(x) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}(x-\mu)^2/\sigma^2$ en $\dot{l}_\mu(x) = (x-\mu)/\sigma^2$. Dus $I(\mu) = E_\mu\left(\frac{X-\mu}{\sigma^2}\right)^2 = \frac{\mathrm{var}_\mu X}{\sigma^4} = \frac{1}{\sigma^2} < \infty$.

Als $\psi(\mu) = \mu$, dan is $\psi'(\mu) = 1$ en kan voor elke zuivere schatter T van $\psi(\mu)$ met $E_\mu T^2 < \infty$

bewezen worden – hoewel dat niet zo eenvoudig is – dat

$$
\begin{aligned}
1 = \frac{\partial}{\partial\mu} E_\mu T &= \frac{\partial}{\partial\mu} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} t(x_1, \dots, x_n)(2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} dx_1 \dots dx_n \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} t(x_1, \dots, x_n) \frac{\partial}{\partial\mu} \left\{ (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} \right\} dx_1 \dots dx_n \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} t(x_1, \dots, x_n) \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} dx_1 \dots dx_n \\
&= E_\mu T \sum_{i=1}^n \dot{l}_\mu(X_i).
\end{aligned}
$$

Ook geldt $E_\mu \dot{l}_\mu(X_1) = E_\mu \frac{X_1-\mu}{\sigma^2} = 0$. Dus voor elke zuivere schatter T van μ met $E_\mu T^2 < \infty$ geldt

$$var_\mu T \geq \frac{\sigma^2}{n}$$

en deze ondergrens wordt bereikt voor $\overline{X}_n$. Dus het steekproefgemiddelde $\overline{X}_n$ is de uniform minimum variantie zuivere schatter van μ.

Opmerking 6.1. De restrictie tot schatters met $E_\theta T^2 < \infty$ is een natuurlijke, omdat voor schatters met $E_\theta T^2 = \infty$ ook $E_\theta(T - \psi(\theta))^2 = \infty$ geldt.

De bekendste en meest gebruikte methode om goede schatters te vinden is die der **meest aannemelijke schatters**. Stel X is bin(n, p) verdeeld met onbekende succeskans $p \in [0, 1]$. Dan is

$$P_p(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ voor } x = 0, 1, \dots, n.$$

Als nu x een gegeven waarneming , d.w.z. realisatie van X, is, beschouwen we de zogenaamde **aannemelijkheidsfunctie**. Dit is $P_p(X = x)$ maar nu gezien als functie in p op $[0, 1]$. De waarde van p waar deze aannemelijkheidsfunctie maximaal is, is de waarde van p waarbij de kans dat X de gegeven realisatie x geeft zo groot mogelijk is en deze waarde komt ons als de meest aannemelijke voor bij de gegeven waarneming x. Daarom noemen we deze waarde van p de **meest aannemelijke schatting** van p. Omdat in dit geval $P_p(X = x)$ een “gladde” functie van p is, kan de waarde van p waar deze maximaal is, gevonden worden onder de stationaire punten (of de randpunten 0 en 1). Dus we lossen op $\frac{\partial}{\partial p} P_p(X = x) = 0$ ofwel

$$\binom{n}{x} p^{x-1} (1-p)^{n-x-1} \{x(1-p) - (n-x)p\} = 0$$

en we vinden $p = x/n$ (en eventueel $p = 0$ en $p = 1$). Enige studie leert, dat het maximum bereikt wordt in x/n. Omdat dit geldt voor alle mogelijke realisaties x, is hier $\hat{p} = \frac{X}{n}$ de **meest aannemelijke schatter** van p. Uit Voorbeeld 6.3 kunnen we concluderen dat dit een goede schatter is.

We geven nu een algemene definitie voor het geval van n waarnemingen. Als $X_1, \ldots, X_n$ o.o. zijn en identiek verdeeld en wel discreet met verdeling P_θ, $\theta \in \Theta$, en als $x_1, \ldots, x_n$ waarnemingen zijn, d.w.z. realisaties van $X_1, \ldots, X_n$, dan wordt een waarde van θ waarvoor $P_\theta(X_1 = x_1, \ldots, X_n = x_n)$ maximaal is een meest aannemelijke schatting van θ genoemd. Een bij deze methode horende schatter $\hat{\theta}$ heet een meest aannemelijke schatter.

Rekenkundig is het vaak eenvoudiger $\log P_\theta(X_1 = x_1, \ldots, X_n = x_n)$ te maximaliseren dan $P_\theta(X_1 = x_1, \ldots, X_n = x_n)$ zelf. Vaak zullen we dan ook een meest aannemelijke schatting bepalen door

$$(6.14) \qquad \sum_{i=1}^{n} l_\theta(x_i) \text{ te maximaliseren (zie (6.7)).}$$

Als l_θ differentieerbaar is naar θ doen we dat door eerst

$$(6.15) \qquad \sum_{i=1}^{n} \dot{l}_\theta(x_i) = 0 \text{ op te lossen (zie (6.8)).}$$

Dit kunnen we ook doen als $X_1, \ldots, X_n$ continu verdeeld zijn.

Definitie 6.4. Als $X_1, \ldots, X_n$ o.o. en identiek verdeeld zijn en wel continu of discreet met kansverdeling P_θ, $\theta \in \Theta$, dan is een meest aannemelijke schatter $\hat{\theta}$ van θ een (stochastische) waarde van θ waarvoor $\sum_{i=1}^{n} l_\theta(X_i)$ maximaal is.

Voorbeeld 6.4. Stel $X_1, \ldots, X_n$ zijn o.o. $\mathcal{N}(\mu, \sigma^2)$ verdeeld met σ^2 bekend en μ onbekend. In Voorbeeld 6.2 hebben we gezien dat $\dot{l}_\mu(x) = (x - \mu)/\sigma^2$. Om de meest aannemelijke schatter $\hat{\mu}$ te bepalen moeten we dus $\sum_{i=1}^{n}(X_i - \mu)/\sigma^2 = 0$ oplossen. Maar de enige oplossing hiervoor is $\overline{X}_n$. We zien dat $\hat{\mu} = \overline{X}_n$ en we weten al dat dit de minimum variantie zuivere schatter van μ is.

Voorbeeld 6.5. Stel X_1 is Laplace verdeeld met dichtheid $\frac{1}{2}e^{-|x-\theta|}$ op $(-\infty, \infty)$. Om de meest aannemelijke schatter te bepalen moeten we $-\log 2 - |X_1 - \theta|$ maximaliseren . Dus $\hat{\theta} = X_1$. Voor n waarnememingen zie Opgave 6.4.

Voorbeeld 6.6. Stel $X_1, \ldots, X_n$ zijn o.o. en hom$(0, \theta)$ verdeeld met dichtheid $\frac{1}{\theta} 1_{[0,\theta]}(\cdot)$. Voor de meest aannemelijke schatting bij waarnemingen $x_1, \ldots, x_n$ moeten we

$$\sum_{i=1}^{n} l_\theta(x_i) = -n \log \theta + \sum_{i=1}^{n} \log 1_{[0,\theta]}(x_i)$$

maximaliseren. Deze uitdrukking is alleen dan groter dan $-\infty$ als $\theta \geq x_i$ is voor $i = 1, \ldots, n$. De waarde van de uitdrukking is dan $-n \log \theta$ en deze is maximaal als θ

minimaal is. Conclusie: de meest aannemelijke schatting van θ is $\max\{x_1, \ldots, x_n\}$ en $\hat{\theta} = \max\{X_1, \ldots, X_n\}$.

Opgaven

6.1. $X_1, \ldots, X_n$ zijn o.o. alternatief verdeeld met succeskans p. Hier is $n \geq 2$ en $X = \sum_{i=1}^n X_i$. Bewijs dat $X(X-1)/\{n(n-1)\}$ een zuivere schatter van p^2 is en $X(n-X)/\{n(n-1)\}$ een zuivere schatter van $p(1-p)$.

6.2. $X_1, \ldots, X_n$ zijn o.o. en homogeen verdeeld op $[0, \theta]$ met $\theta > 0$ en θ onbekend. Laat zien dat $T_1 = \frac{2}{n}(X_1 + \ldots + X_n)$ en $T_2 = \frac{n+1}{n}\max\{X_1, \ldots, X_n\}$ beide zuivere schatters voor θ zijn. Bereken $E_\theta(T_1 - \theta)^2$ en $E_\theta(T_2 - \theta)^2$. Aan welke schatter geeft U de voorkeur?

6.3. Laat $X_1, \ldots, X_n$ o.o. $\mathcal{N}(0, \sigma^2)$ verdeeld zijn met $\sigma^2 > 0$ onbekend. Bepaal de meest aannemelijke schatter van σ^2.

6.4. Laat $X_1, \ldots, X_n$ o.o. en identiek verdeeld zijn en wel Laplace met dichtheid $\frac{1}{2}e^{-|x-\theta|}$ op $\mathbf{R}$. Bepaal **een** meest aannemelijke schatter van θ.

7. Toetsingstheorie

Stel $X_1, \ldots, X_n$ zijn o.o. en identiek verdeeld en wel $\mathcal{N}(\mu, \sigma^2)$ met μ onbekend en σ^2 bekend. Dit is een (gebrekkig)stochastisch model voor de logaritme van de prijs van een zeker aandeel op 13 juni van n voorgaande jaren. Op 13 juni van dit jaar willen we beslissen over de aankoop van zulke aandelen met het doel deze op 13 juni volgend jaar weer te verkopen. De prijs op 13 juni van dit jaar is e^{μ_0}.

Alleen als de (onbekende) μ groter is dan μ_0 willen we kopen, omdat dan de verkoopprijs volgend jaar naar verwachting groter is dan de koopprijs nu. De realisaties $x_1, \ldots, x_n$ van $X_1, \ldots, X_n$ bevatten informatie over μ en deze waarnemingen kunnen we gebruiken om onze aankoopbeslissing op te baseren. Absolute zekerheid over μ is niet te krijgen. We zullen dus steeds de kans lopen dat we beslissen tot aankoop, d.w.z. uit de gegevens concluderen dat $\mu > \mu_0$, terwijl, in feite, $\mu \leq \mu_0$ is. Deze fout kan ernstige financiële consequenties hebben. De andere fout, nl. concluderen dat $\mu \leq \mu_0$ en dus niet kopen, terwijl, in feite, wèl $\mu > \mu_o$ is, kan ons slechts een mogelijke winst kosten. Daarom willen we de kans, dat we de eerstgenoemde fout begaan begrenzen door α_0, zeg, en tegelijkertijd de kans op de tweede fout minimaliseren.

In het vorige hoofdstuk hebben we gezien dat $\overline{X}_n$ de optimale schatter voor μ is. Daarom besluiten we tot de volgende procedure. We kiezen een getal c en concluderen tot $\mu > \mu_0$, d.w.z. tot aankoop, als de realisatie $\overline{x}_n$ van $\overline{X}_n$ groter dan of gelijk aan c is. We moeten nu c zó kiezen dat

$$(7.1) \qquad P_\mu(\overline{X}_n \geq c) \leq \alpha_0 \text{ voor alle } \mu \leq \mu_0$$

(de kans op de eerste fout wordt begrensd door α_0) en

$$(7.2) \qquad P_\mu(\overline{X}_n < c) \text{ minimaal is voor alle } \mu > \mu_0$$

(de kans op de tweede fout willen we minimaliseren). Uiteraard is (7.2) equivalent met

$$(7.3) \qquad P_\mu(\overline{X}_n \geq c) \text{ maximaal voor alle } \mu > \mu_0.$$

Nu is, omdat $\overline{X}_n$ onder P_μ $\mathcal{N}(\mu, \sigma^2/n)$ verdeeld is,

$$(7.4) \qquad P_\mu(\overline{X}_n \geq c) = P_\mu\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \geq \frac{c - \mu}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma}\right)$$

met Φ de standaard normale verdelingsfunctie. De eis (7.1) wordt dus

$$1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma}\right) \leq \alpha_0 \text{ voor alle } \mu \leq \mu_0$$

ofwel

$$c \geq \mu + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha_0) \text{ voor alle } \mu \leq \mu_0$$

ofwel

$$c \geq \mu_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha_0). \tag{7.5}$$

Verder is, in het licht van (7.4), de eis (7.3) equivalent met

$$1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma}\right) \text{ maximaal voor alle } \mu > \mu_0$$

ofwel

$$\Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma}\right) \text{ minimaal voor alle } \mu > \mu_0$$

ofwel

$$c - \mu \text{ minimaal voor alle } \mu > \mu_0$$

ofwel

$$c \text{ minimaal.} \tag{7.6}$$

De eisen (7.5) en (7.6) leiden ons tot de keuze

$$c = \mu_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha_0). \tag{7.7}$$

We beslissen dus tot aankoop over te gaan, als

$$\overline{x}_n \geq \mu_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1-\alpha_0). \tag{7.8}$$

Deze aanpak is in veel statistische situaties toepasbaar en is in de dertiger jaren ontwikkeld door voornamelijk J. Neyman en E. Pearson. Om deze aanpak/methode/filosofie nauwkeurig te beschrijven voeren we enkele begrippen in voor de algemene situatie van $X_1, \ldots, X_n$ o.o. en identiek verdeeld met kansmaat P_θ, $\theta \in \Theta$. Zij $\Theta_0 \subset \Theta$ en $\Theta_1 = \Theta \setminus \Theta_0$. We willen de **nulhypothese** $H_0 : \theta \in \Theta_0$ **toetsen** tegen de **alternatieve hypothese** $H_1 : \theta \in \Theta_1$, d.w.z. we willen tot één van de volgende conclusies komen:

- H_0 wordt verworpen,
- H_0 wordt **niet** verworpen.

Hierbij kunnen we twee soorten fouten maken:

- H_0 wordt verworpen, terwijl $\theta \in \Theta_0$ en H_0 dus juist is; dit is een **fout van de eerste soort** (en de ernstigste fout die we kunnen maken),
- H_0 wordt niet verworpen, terwijl $\theta \notin \Theta_0$ en H_0 dus niet juist is; dit heet een **fout van de tweede soort**.

We kiezen een zogenaamde **onbetrouwbaarheidsdrempel** α_0 en eisen van onze **toets** dat de kans op een fout van de eerste soort ten hoogste α_0 is. Verder willen we dat de kans op een fout van de tweede soort zo klein mogelijk is.

Een toets kunnen we beschrijven door middel van een **kritiek gebied** $K \subset \mathbf{R}^n$. Als $x = (x_1, \ldots, x_n)$ de (vector van) realisatie(s) van $X = (X_1, \ldots, X_n)$ is, dan is het toetsingsvoorschrift:

– verwerp H_0 als $x \in K$,
– verwerp H_0 niet als $x \notin K$.

Als generalisatie van (7.1)–(7.3) hebben we dan: we willen K zó kiezen, dat

$$P_\theta(X \in K) \le \alpha_0 \text{ voor alle } \theta \in \Theta_0 \tag{7.9}$$

en

$$P_\theta(X \notin K) \text{ minimaal is voor } \theta \notin \Theta_0 \tag{7.10}$$

ofwel

$$P_\theta(X \in K) \text{ maximaal is voor } \theta \notin \Theta_0. \tag{7.11}$$

Verder noemen we de functie $\theta \mapsto P_\theta(X \in K)$ het **onderscheidend vermogen** van de toets met kritiek gebied K en

$$\sup_{\theta \in \Theta_0} P_\theta(X \in K) = \alpha$$

de **onbetrouwbaarheid** van deze toets. Een toets met $\alpha \le \alpha_0$ heet **meest onderscheidend** in θ_1 (met $\theta_1 \in \Theta_1$), als er geen enkele toets met onbetrouwbaarheid $\le \alpha_0$ is met een hoger onderscheidend vermogen in θ_1. Een toets met $\alpha \le \alpha_0$ heet **uniform meest onderscheidend** als deze meest onderscheidend is in elk punt θ_1 van Θ_1.

De keuze (7.7), (7.8) is in deze terminologie

$$K = \left\{ x = (x_1, \ldots, x_n) \mid \frac{1}{n} \sum_{i=1}^n x_i \ge \mu_0 + \frac{\sigma}{\sqrt{n}} \Phi^{-1}(1 - \alpha_0) \right\}$$

en hier is $\alpha = \alpha_0$. We zullen later zien dat deze toets uniform meest onderscheidend is.

Het basisresultaat in de toetsingstheorie is het fundamentele lemma van Neyman en Pearson. Dit geeft de oplossing voor het eenvoudigste toetsingsprobleem, nl. het toetsen van een enkelvoudige nulhypothese tegen een enkelvoudige alternatieve hypothese: $\Theta_0 = \{\theta_0\},\ \Theta_1 = \{\theta_1\}$.
Stel nu dat $X = (X_1, \ldots, X_n)$ discreet verdeeld is met $P_\theta(X = x) = p_\theta(x),\ x \in \mathbf{R}^n$.

Stelling 7.1. Fundamentele lemma van Neyman en Pearson (1933).
Als er in bovenstaande situatie en constante $c \geq 0$ bestaat zó dat voor

$$K = \Big\{x \in \mathbf{R}^n \mid p_{\theta_1}(x) - cp_{\theta_0}(x) \geq 0\Big\},$$

$$P_{\theta_0}(X \in K) = \alpha_0$$

geldt, dan is de toets met kritiek gebied K meest onderscheidend voor het toetsen van $H_0 : \theta = \theta_0$ tegen het alternatief $H_1 : \theta = \theta_1$ bij onbetrouwbaarheidsdrempel α_0.

Bewijs. Beschouw een toets met kritiek gebied $\tilde{K}$ en onbetrouwbaarheid $\leq \alpha_0$. Met $1_K(x)$ noteren we de indicatorfunctie van K,

$$1_K(x) = \begin{cases} 1 & \text{als x} \in \text{K.} \\ 0 & \text{als x} \notin \text{K.} \end{cases}$$

Nu geldt

$$(a) \qquad \sum_x 1_{\tilde{K}}(x) p_{\theta_0}(x) = P_{\theta_0}(X \in \tilde{K}) \leq \alpha_0 = P_{\theta_0}(X \in K) = \sum_x 1_K(x) p_{\theta_0}(x)$$

en dus

$$\sum_x \Big\{1_K(x) - 1_{\tilde{K}}(x)\Big\} p_{\theta_0}(x) \geq 0$$

en, omdat $c \geq 0$ is,

$$(b) \qquad \sum_x \Big\{1_K(x) - 1_{\tilde{K}}(x)\Big\}\Big\{-cp_{\theta_0}(x)\Big\} \leq 0.$$

Hieruit volgt

$$(c) \qquad \begin{aligned} P_{\theta_1}(X \in K) - P_{\theta_1}(X \in \tilde{K}) &= \sum_x 1_K(x) p_{\theta_1}(x) - \sum_x 1_{\tilde{K}}(x) p_{\theta_1}(x) \\ &= \sum_x \{1_K(x) - 1_{\tilde{K}}(x)\}\, p_{\theta_1}(x) \\ &\geq \sum_x \{1_K(x) - 1_{\tilde{K}}(x)\}\, \{p_{\theta_1}(x) - cp_{\theta_0}(x)\}\,. \end{aligned}$$

Uit onderstaande tabel

	$1_K(x)$	$1_K(x) - 1_{\tilde{K}}(x)$	$p_{\theta_1}(x) - cp_{\theta_0}(x)$	$\{1_K(x) - 1_{\tilde{K}}(x)\}\{p_{\theta_1}(x) - cp_{\theta_0}(x)\}$
$x \in K$	1	≥ 0	≥ 0	≥ 0
$x \notin K$	0	≤ 0	< 0	≥ 0

volgt

$$(d) \qquad \sum_x \{1_K(x) - 1_{\tilde{K}}(x)\}\{p_{\theta_1}(x) - cp_{\theta_0}(x)\} \geq 0.$$

Combinatie van (c) en (d) levert

$$(e) \qquad P_{\theta_1}(X \in K) - P_{\theta_1}(X \in \tilde{K}) \geq 0$$

en dus dat de toets met kritiek gebied K meer onderscheidend is dan die met kritiek gebied $\tilde{K}$. Omdat $\tilde{K}$ willekeurig is, mits de resulterende onbetrouwbaarheid niet groter is dan α_0, bewijst dit dat de toets met kritiek gebied K meest onderscheidend is. ▯

Als X continu verdeeld is met f_θ als de simultane dichtheid van $X_1, \ldots, X_n$, geldt in essentie dezelfde stelling.

Stelling 7.2. Fundamentele lemma van Neyman en Pearson; continue geval
Als er een constante $c \geq 0$ bestaat zó dat voor

$$K = \{x \in \mathbf{R}^n \mid f_{\theta_1}(x) - cf_{\theta_0}(x) \geq 0\},$$

$$P_{\theta_0}(X \in K) = \alpha_0$$

geldt, dan is de toets met kritiek gebied K meest onderscheidend voor het toetsen van $H_0 : \theta = \theta_0$ tegen het alternatief $H_1 : \theta = \theta_1$ bij onbetrouwbaarheidsdrempel α_0.

Bewijs. Beschouw een willekeurige toets met kritiek gebied $\tilde{K}$ en onbetrouwbaarheid $\leq \alpha_0$. Nu is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1_{\tilde{K}}(x) f_{\theta_0}(x) dx_1 \ldots dx_n \leq \alpha_0 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1_K(x) f_{\theta_0}(x) dx_1 \ldots dx_n$$

en dus

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \{1_K(x) - 1_{\tilde{K}}(x)\}\{-cf_{\theta_0}(x)\}\, dx_1 \ldots dx_n \leq 0.$$

Hieruit volgt met behulp van een tabel als in het vorige bewijs, dat

$$P_{\theta_1}(X \in K) - P_{\theta_1}(X \in \tilde{K}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \{1_K(x) - 1_{\tilde{K}}(x)\} f_{\theta_1}(x) dx_1 \dots dx_n$$

$$\geq \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \{1_K(x) - 1_{\tilde{K}}(x)\} \{f_{\theta_1}(x) - c f_{\theta_0}(x)\} dx_1 \dots dx_n \geq 0$$

en dus

$$P_{\theta_1}(X \in K) \geq P_{\theta_1}(X \in \tilde{K}).$$

$\square$

Voorbeeld 7.1. Stel $X_1, \dots, X_n$ zijn o.o. $\mathcal{N}(\mu, \sigma^2)$ verdeelde stochastische grootheden met σ^2 bekend en μ onbekend. Zij $\mu_0 \in \mathbf{R}$ vast. We willen de nulhypothese $H_0 : \mu \leq \mu_0$ toetsen tegen de alternatieve hypothese $H_1 : \mu > \mu_0$ bij onbetrouwbaarheidsdrempel α_0. Eerst beschouwen we een eenvoudig deelprobleem, nl. het toetsen van $\tilde{H}_0 : \mu = \mu_0$ versus $\tilde{H}_1 : \mu = \mu_1$ met $\mu_1 > \mu_0$ vast gekozen. Volgens het lemma van Neyman en Pearson moeten we de simultane dichtheden f_{μ_1} en f_{μ_0} met elkaar vergelijken en $\tilde{H}_0$ verwerpen voor grote waarden van $f_{\mu_1}(x)/f_{\mu_0}(x)$. Nu is

$$\frac{f_{\mu_1}(x)}{f_{\mu_0}(x)} = \frac{(2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_1)^2}}{(2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_0)^2}}$$

groot, d.e.s.d.a.

$$e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left\{(x_i-\mu_0)^2-(x_i-\mu_1)^2\right\}}$$

groot is d.e.s.d.a.

$$\sum_{i=1}^{n}\left\{(x_i - \mu_0)^2 - (x_i - \mu_1)^2\right\} = 2(\mu_1 - \mu_0)\sum_{i=1}^{n} x_i + n(\mu_0^2 - \mu_1^2)$$

groot is d.e.s.d.a.

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}_n$$

groot is. Voor het optimaal toetsen van $\tilde{H}_0$ tegen $\tilde{H}_1$ moeten we dus als kritiek gebied K kiezen met $K = \{x \in \mathbf{R}^n \mid \overline{x}_n \geq c\}$ en hierbij moeten we c zó kiezen dat $P_{\mu_0}(X \in K) = \alpha_0$. Dit probleem hebben we al opgelost in (7.4)–(7.8) en dus is

(7.12) $$K = \left\{x \in \mathbf{R}^n \mid \overline{x}_n \geq \mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1 - \alpha_0)\right\}.$$

Merk op dat K en dus de toets niet van de keuze μ_1 afhangt. Dit betekent dat één en dezelfde toets voor het toetsen van $\tilde{H}_0$ meest onderscheidend is in elk punt μ_1 met $\mu_1 > \mu_0$. Dus de toets met kritiek gebied K is uniform meest onderscheidend voor het toetsen van $\tilde{H}_0 : \mu = \mu_0$ tegen $H_1 : \mu > \mu_0$ bij onbetrouwbaarheidsdrempel α_0. Tenslotte merken we op dat voor alle $\mu \leq \mu_0$

$$
\begin{aligned}
P_\mu(X \in K) &= 1 - \Phi\left(\frac{\sqrt{n}}{\sigma}\left(\mu_0 + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha_0) - \mu\right)\right) \\
&= 1 - \Phi\left(\Phi^{-1}(1-\alpha_0) + \frac{\sqrt{n}}{\sigma}(\mu_0 - \mu)\right) \\
&\leq 1 - \Phi\left(\Phi^{-1}(1-\alpha_0)\right) = 1 - \alpha_0
\end{aligned}
\tag{7.13}
$$

en dat dus de toets met kritiek gebied K van (7.12) onbetrouwheid α_0 heeft voor het toetsen van $H_0 : \mu \leq \mu_0$ tegen $H_1 : \mu > \mu_0$. Maar dan is deze toets dus ook een toegestane toets voor dit toetsingsprobleem.
Conclusie: de toets met kritiek gebied K uit (7.12) is uniform meest onderscheidend voor het toetsen van $H_0 : \mu \leq \mu_0$ tegen $H_1 : \mu > \mu_0$ bij onbetrouwbaarheidsdrempel α_0.

Voorbeeld 7.2. $X_1, \ldots, X_n$ zijn o.o. alternatief verdeeld met succeskans $p \in (0,1)$.

Kies $p_0 \in (0,1)$ vast en toets $H_0 : p \leq p_0$ tegen $H_1 : p > p_0$. We bestuderen weer eerst een eenvoudiger toetsingsprobleem met $\tilde{H}_0 : p = p_0$ tegen $\tilde{H}_1 : p = p_1$, waarbij $p_1 > p_0$ vast is gekozen. Volgens stelling 7.1 moeten we $\tilde{H}_0$ verwerpen voor grote waarden van

$$
\begin{aligned}
\frac{P_{p_1}(X_1 = x_1, \ldots, X_n = x_n)}{P_{p_0}(X_1 = x_1, \ldots, X_n = x_n)} &= \left(\frac{p_1}{p_0}\right)^{x_1+\ldots+x_n} \left(\frac{1-p_1}{1-p_0}\right)^{(1-x_1)+\ldots+(1-x_n)} \\
&= \left(\frac{1-p_1}{1-p_0}\right)^n \left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)^{\sum_{i=1}^n x_i}.
\end{aligned}
$$

Omdat $p_1/p_0 > 1$ en $(1-p_0)/(1-p_1) > 1$, betekent dit dat we moeten verwerpen voor grote waarden van $\sum_{i=1}^n x_i$. Kies nu $k \in \mathbf{N}$ en beschouw de toets die verwerpt bij k of meer successen, d.w.z. als $\sum_{i=1}^n x_i \geq k$. Deze toets heeft onbetrouwbaarheid

$$
\alpha = P_{p_0}\left(\sum_{i=1}^n X_i \geq k\right) = \sum_{x=k}^n \binom{n}{x} p_0^x (1-p_0)^{n-x}.
$$

Als we de onbetrouwbaarheidsdrempel α_0 gelijk kiezen aan deze α, is volgens stelling 7.1 de toets met kritiek gebied

$$
K = \left\{x \in \mathbf{R}^n \mid \sum_{i=1}^n x_i \geq k\right\}
$$

meest onderscheidend bij het toetsen van $\tilde{H}_0$ tegen $\tilde{H}_1$. Omdat K en dus de toets niet van p_1 afhangen, is de toets uniform meest onderscheidend voor het toetsen van $\tilde{H}_0 : p = p_0$ tegen $H_1 : p > p_0$. Verder is

$$\begin{aligned} \frac{d}{dp} P_p(X \in K) &= \frac{d}{dp} P_p\left(\sum_{i=1}^n X_i \geq k\right) = \frac{d}{dp} \sum_{x=k}^n \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=k}^n \left\{\frac{x}{p} - \frac{n-x}{1-p}\right\} \binom{n}{x} p^x (1-p)^{n-x} \\ &= E_p\left\{\frac{X}{p} - \frac{n-X}{1-p}\right\} 1_{[k,\infty)}(X). \end{aligned} \tag{7.14}$$

Nu is eenvoudig te zien dat

$$\left\{\frac{x}{p} - \frac{n-x}{1-p}\right\} 1_{[k,\infty)}(x) \geq \begin{cases} 0 & \text{als } \frac{k}{p} - \frac{n-k}{1-p} \geq 0. \\ \frac{x}{p} - \frac{n-x}{1-p} & \text{als } \frac{k}{p} - \frac{n-k}{1-p} < 0. \end{cases} \tag{7.15}$$

Omdat

$$E_p\left\{\frac{X}{p} - \frac{n-X}{1-p}\right\} = 0,$$

volgt uit (7.14) en (7.15) dat $\frac{d}{dp} P_p(X \in K) \geq 0$ en dus dat

$$\sup_{p \leq p_0} P_p(X \in K) = P_{p_0}(X \in K) = \alpha = \alpha_0.$$

Net als in Voorbeeld 7.1 kunnen we concluderen dat de toets met kritiek gebied K uniform meest onderscheidend is voor het toetsen van $H_0 : p \leq p_0$ tegen $H_1 : p > p_0$ bij onbetrouwheidsdrempel α_0.

In hoofdstukken 6 en 7 zijn op een elementaire manier de basisideeën van de hoofdstellingen van de schattings– en toetsingstheorie afgeleid. Uiteraard kunnen deze ideeën verder uitgediept en uitgebreid worden en zijn er binnen de statistiek veel toepassingen van te vinden. Overigens worden er in de mathematische statistiek verscheidene uiteenlopende wiskundige theorieën toegepast en blijft het niet bij de Cramér-Rao ongelijkheid als toepassing van de ongelijkheid van Cauchy-Schwarz en het Neyman-Pearson lemma als eenvoudig maximalisatie probleem.

Opgaven

7.1. $X_1, \ldots, X_n$ zijn o.o. $\mathcal{N}(\mu, \sigma^2)$ verdeeld. Toets op basis van $X_1, \ldots, X_n$ bij onbetrouwheidsdrempel α_0 de nulhypothese $H_0 : \mu \geq \mu_0$ tegen het alternatief $H_1 : \mu < \mu_0$.

7.2. Laat $X_1, \ldots, X_n$ o.o. $\text{bin}(k, p)$ verdeeld zijn. Beschouw het toetsen van de nulhypothese $H_0 : p \geq p_0$ tegen het alternatief $H_1 : p < p_0$. Voor welke onbetrouwbaarheidsdrempels kunt U de uniform meest onderscheidende toets aangeven?

7.3. Stel $X_1, \dots, X_n$ zijn o.o. hom$(0, \theta)$ verdeeld, $H_0 : \theta \leq \theta_0$ en $H_1 : \theta > \theta_0$. Vind de uniform meest onderscheidende toets voor het toetsen van H_0 versus H_1 bij onbetrouwbaarheidsdrempel 0.

7.4. Zijn $X_1, \dots, X_n$ o.o. en Poisson(λ) verdeeld. Zijn $H_0 : \lambda \leq \lambda_0$ en $H_1 : \lambda > \lambda_0$ de nul– en alternatieve hypothese. Op welke toetsingsgrootheid zijn de uniform meest onderscheidende toetsen gebaseerd?

Literatuur

Een gedegen uitwerking van de geboden stof met vele opgaven vindt U in

Mood, A.M., F.A. Graybill and **D.C. Boes** (1974). Introduction to the Theory of Statistics, McGraw-Hill, Singapore.

Wiskundig interessanter is

Bickel, P.J. and **K.A. Doksum** (1977). Mathematical Statistics, Basic ideas and Selected Topics, Holden-Day, San Francisco.

Proeftentamen Statistiek D1

1. Formuleer en bewijs de Cramér-Rao ongelijkheid.
2. Bereken de verwachting en variantie van een stochastische grootheid met een hypergeometrische verdeling.
3. Laat $X_1, \dots, X_n$ o.o. en identiek verdeeld zijn. Bepaal de meest aannemelijke schatter van λ als de X_i
   a. Poisson(λ)
   b. $\Gamma(k, \lambda)$
   verdeeld zijn.
4. Stel $X_1, \dots, X_n$ zijn o.o. hom$(\theta - \frac{1}{2}, \theta + \frac{1}{2})$ verdeeld. Zij $H_0 : \theta \leq 0$ en $H_1 : \theta > 0$. Vind de uniform meest onderscheidende toets voor het toetsen van H_0 versus H_1 bij onbetrouwbaarheidsdrempel 0.

8. Betrouwbaarheidsintervallen

In Voorbeeld 6.2 van Hoofdstuk 6 hebben we gezien dat het gemiddelde $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ de uniform minimum variantie zuivere schatter is van de parameter μ van de normale verdeling $\mathcal{N}(\mu,\sigma^2)$, als $X_1,\ldots,X_n$ o.o. en identiek verdeeld zijn. Dit betekent dat de realisatie $\bar{x}_n$ van de schatter $\bar{X}_n$ in de buurt van de ware parameterwaarde μ zal liggen. Hoe groot $|\bar{x}_n-\mu|$ precies is, weten we niet. In ieder geval is de kans nul dat $\bar{X}_n$ gelijk wordt aan μ als μ de ware waarde is; in formule,

$$P_\mu\big(\bar{X}_n=\mu\big)=P_\mu\bigg(\sqrt{n}\Big(\frac{\bar{X}_n-\mu}{\sigma}\Big)=0\bigg)=\Phi(0)-\Phi(0-)=0. \tag{8.1}$$

De sterke bewering $\mu=\bar{X}_n$ kunnen we vervangen door de zwakkere bewering $|\bar{X}_n-\mu|\le c$. Analoog aan (8.1) geldt voor deze bewering,

$$\begin{aligned}P_\mu\big(|\bar{X}_n-\mu|\le c\big)&=P_\mu\bigg(\sqrt{n}\Big|\frac{\bar{X}_n-\mu}{\sigma}\Big|\le\frac{c}{\sigma}\sqrt{n}\bigg)=\Phi\Big(\frac{c}{\sigma}\sqrt{n}\Big)-\Phi\Big(-\frac{c}{\sigma}\sqrt{n}\Big)\\&=2\Phi\Big(\frac{c}{\sigma}\sqrt{n}\Big)-1,\ \text{voor alle } \mu\in\mathbf{R}.\end{aligned} \tag{8.2}$$

Voor de keuze $c=\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{1}{2}\alpha)$ levert dit op

$$P_\mu\bigg(\Big|\bar{X}_n-\mu\Big|\le\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\tfrac{1}{2}\alpha)\bigg)=1-\alpha,\quad \mu\in\mathbf{R}. \tag{8.3}$$

We zeggen dat de bewering $|\bar{X}_n-\mu|\le\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{1}{2}\alpha)$ betrouwbaarheid $1-\alpha$ of onbetrouwbaarheid α heeft. Duidelijk zal zijn dat de bewering $|\bar{X}_n-\mu|\le\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{1}{2}\alpha)$ zwakker wordt naarmate de betrouwbaarheid $1-\alpha$ groter wordt. Er moet dus een evenwicht gevonden worden tussen de kracht en de betrouwbaarheid van de bewering. Hoe sterker, hoe onbetrouwbaarder. Merk in dit verband op dat (8.1) het speciale geval van (8.3) is met $\alpha=1$. De bewering $|\bar{X}_n-\mu|\le\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\frac{1}{2}\alpha)$ kan ook als volgt geformuleerd worden:

$$\mu\in\Big[\bar{X}_n-\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\tfrac{1}{2}\alpha),\ \bar{X}_n+\frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\tfrac{1}{2}\alpha)\Big]=I_{n,\alpha}(X), \tag{8.4}$$

met $X=(X_1,\ldots,X_n)^T$. Het interval $I_{n,\alpha}(X)$ noemen we een betrouwbaarheidsinterval met onbetrouwbaarheid α. Hoe groter α, hoe korter het stochastische interval $I_{n,\alpha}(X)$ is en hoe sterker de bewering over μ.

Hoewel de theorie over betrouwbaarheidsintervallen qua onderwerp sterk gelieerd is aan de schattingstheorie is het toch de toetsingstheorie die ons een algemene methode geeft om betrouwbaarheidsintervallen te construeren. Alvorens een algemene stelling te formuleren bestuderen we eerst het toetsingsvoorbeeld 7.1 voor normaal verdeelde stochastische grootheden wat nauwkeuriger.

Voorbeeld 8.1. Stel $X_1,\dots,X_n$ zijn o.o. $\mathcal{N}(\mu,\sigma^2)$ verdeelde stochastische grootheden met σ^2 bekend en μ onbekend. Zij $\mu_0\in\mathbf{R}$ vast. In Voorbeeld 7.1 hebben we gezien dat de toets die verwerpt als $\bar X_n-\mu_0\ge\frac{\sigma}{\sqrt n}\Phi^{-1}(1-\alpha_0)$ uniform meest onderscheidend is voor het toetsen van de nulhypothese $H_0:\mu\le\mu_0$ tegen $H_1:\mu>\mu_0$ bij onbetrouwbaarheidsdrempel α_0. Analoog hebben we in Opgave 7.1 gezien dat de toets die verwerpt als $\bar X_n-\mu_0\le-\frac{\sigma}{\sqrt n}\Phi^{-1}(1-\alpha_0)$ uniform meest onderscheidend is voor het toetsen van de nulhypothese $H_0:\mu\ge\mu_0$ tegen $H_1:\mu<\mu_0$ bij onbetrouwbaarheidsdrempel α_0. Het ligt nu voor de hand om de nulhypothese $H_0:\mu=\mu_0$ te verwerpen ten gunste van de alternatieve hypothese $H_1:\mu\ne\mu_0$ als $|\bar X_n-\mu_0|\ge c$. Met de berekening (8.2) zien we dat $P_{\mu_0}(|\bar X_n-\mu_0|\ge c)=2(1-\Phi(\frac{c}{\sigma}\sqrt n))=\alpha_0$ als we $c=\frac{\sigma}{\sqrt n}\Phi^{-1}(1-\frac12\alpha_0)$ kiezen. We zien dat de toets met kritiek gebied

$$K_{n,\alpha}(\mu_0)=\big\{x\in\mathbf{R}^n\mid|\bar x_n-\mu_0|>\frac{\sigma}{\sqrt n}\Phi^{-1}(1-\tfrac12\alpha)\big\}\tag{8.5}$$

onbetrouwbaarheid α heeft voor het toetsen van $H_0:\mu=\mu_0$ tegen $H_1:\mu\ne\mu_0$. Merk op, dat we $>$ hebben geschreven in (8.5) i.p.v. $\ge$. ▯

Vergelijking van (8.4) en (8.5) toont aan dat μ_0 in het betrouwbaarheidsinterval $I_{n,\alpha}(X)$ zit als het bij toetsing van $H_0:\mu=\mu_0$ tegen $H_1:\mu\ne\mu_0$ **niet** wordt verworpen. We hebben dus

$$\mu_0\in I_{n,\alpha}(X)\iff X\notin K_{n,\alpha}(\mu_0).\tag{8.6}$$

We kunnen dit als volgt formuleren: "*Het betrouwbaarheidsinterval bestaat uit de bij toetsing* **niet** *verworpen waarden.*"

Dit levert een algemene methode om betrouwbaarheidsintervallen te construeren, zoals we nu zullen nagaan. Daarvoor kiezen we net als in de voorgaande hoofdstukken weer de algemene situatie van $X_1,\dots,X_n$ o.o. en identiek verdeeld met kansmaat P_θ, $\theta\in\Theta$. Noteer $X=(X_1,\dots,X_n)^T$. Beschouw de nulhypothese $H_0:\theta=\theta_0$ met alternatieve hypothese $H_1:\theta\ne\theta_0$. Stel dat we een toets hebben met kritiek gebied $K_{n,\alpha}(\theta_0)$ en onbetrouwbaarheid α. Dan geldt dus

$$P_{\theta_0}(X\in K_{n,\alpha}(\theta_0))=\alpha,\quad\text{ofwel}\quad P_{\theta_0}(X\notin K_{n,\alpha}(\theta_0))=1-\alpha.\tag{8.7}$$

Als we voor *elke* $\theta_0\in\Theta$ zo een kritiek gebied hebben, dan kunnen we definiëren

$$I_{n,\alpha}(x)=\{\theta\in\Theta\mid x\notin K_{n,\alpha}(\theta)\}.\tag{8.8}$$

Dan geldt dus

$$\theta\in I_{n,\alpha}(x)\iff x\notin K_{n,\alpha}(\theta),\tag{8.9}$$

en daarom met (8.7)

$$P_{\theta_0}(\theta_0\in I_{n,\alpha}(X))=P_{\theta_0}(X\notin K_{n,\alpha}(\theta_0))=1-\alpha,\ \theta_0\in\Theta,$$

of zonder index 0

$$P_\theta(\theta \in I_{n,\alpha}(X)) = 1 - \alpha, \;\; \theta \in \theta. \tag{8.10}$$

We vatten dit alles als volgt samen.

Stelling 8.1 *Als er in bovenstaande situatie voor elke* $\theta_0 \in \Theta$ *een toets bestaat met kritiek gebied* $K_{n,\alpha}(\theta_0)$ *bij onbetrouwbaarheid* α *voor het toetsen van de nulhypothese* $H_0 : \theta = \theta_0$ *tegen* $H_1 : \theta \neq \theta_0$, *dan is* $I_{n,\alpha}(X)$ *met* $I_{n,\alpha}(x)$ *gedefiniëerd in (8.8) een betrouwbaarheidsgebied voor* θ *met betrouwbaarheid* $1 - \alpha$, *dat wil zeggen dat voor alle* $\theta \in \Theta$ *(8.10) geldt.*

Bewijs. Het eenvoudige bewijs is hierboven in (8.7) tot en met (8.10) gegeven. ▯

Merk op, dat $I_{n,\alpha}(x)$ niet noodzakelijkerwijze een interval is en dat daarom de naam betrouwbaarheidsgebied preferabel is.

Opgaven

8.1. $X_1, \ldots, X_n$ zijn o.o. $\mathcal{N}(\mu, \sigma^2)$ verdeeld. Laat zien dat

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha), \infty\right) \text{ en } \left(-\infty, \bar{X}_n + \frac{\sigma}{\sqrt{n}}\Phi^{-1}(1-\alpha)\right]$$

(oneigenlijke) betrouwbaarheidsintervallen voor μ zijn met betrouwbaarheid $1 - \alpha$.

8.2. $X_1, \ldots, X_n$ zijn o.o. alternatief (Bernoulli) verdeeld met succeskans $p \in (0, 1)$. Voorbeeld 7.2 heeft laten zien dat $K_{n,\alpha}(p_0) = \{x \in \mathbf{R}^n \mid \sum_{i=1}^n x_i \geq k(p_0, \alpha)\}$ een toets met onbetrouwbaarheid $\leq \alpha$ oplevert voor $\tilde{H}_0 = H_0 : p = p_0$. Hierbij wordt $k(p_0, \alpha)$ zo klein mogelijk gekozen. Laat zien dat Stelling 8.1 een betrouwbaarheids*interval* voor p oplevert.

Repetitie-opgaven

Voer voor elk van de verdelingen uit Hoofdstuk 3 de volgende opgaven uit.

R.1.a. Bepaal meest aannemelijke schatters voor de parameters.

b. Ga na of met behulp van de Cramér-Rao ongelijkheid de optimaliteit van deze schatters bewezen kan worden.

R.2. Kunt u met behulp van de fundamentele Neyman-Pearson lemmata uit Hoofdstuk 7 optimale toetsen construeren voor het toetsen van

a. $H_0 : \theta = \theta_0$ tegen $H_1 : \theta = \theta_1$,
b. $H_0 : \theta \leq \theta_0$ tegen $H_1 : \theta > \theta_0$,
c. $H_0 : \theta \geq \theta_0$ tegen $H_1 : \theta < \theta_0$,
d. $H_0 : \theta = \theta_0$ tegen $H_1 : \theta \neq \theta_0$?

R.3. Tot welke betrouwbaarheidsgebieden leiden de resultaten van R.2?