TECHNISCHE HOGESCHOOL EINDHOVEN
Afdeling Algemene Wetenschappen
Onderafdeling der Wiskunde

KANSREKENING

en

STATISTIEK

Deel II: Statistiek

Bestemd voor WSK-IV

Voorjaarssemester 1979

Technische Hogeschool
Eindhoven

Dictaatnummer 2.242
Prijs f. 3,50

Onderafdeling der Wiskunde en Informatica

Jd8

Deel II: Statistiek

Kansrekening en Statistiek

Bestemd voor WSK-IV

TECHNISCHE HOGESCHOOL EINDHOVEN

Onderafdeling der Wiskunde

KANSREKENING EN STATISTIEK

bestemd voor WSK IV

deel 2: STATISTIEK

Voorjaarssemester 1979

Deel 2, Statistiek.

Inhoudsopgave | blz.

Inleiding

De theoretische statistiek kan beschouwd worden als een toepassing van de kansrekening. Anderzijds gaat aan de toepassing van de kansrekening meestal een toepassing van de statistiek vooraf: in de statistiek houden we ons bezig met onderzoek naar het (gedetailleerde) kanstheoretische model dat het experiment of verschijnsel dat we bestuderen beschrijft. Dit model is meestal niet in alle details bekend en er zijn waarnemingen nodig om het model volledig te maken, zonder dat in het algemeen met zekerheid het precieze model kan worden vastgesteld. Als we (bijvoorbeeld) een uitspraak willen doen over het vrouwenoverschot in 1984, zullen we (o.a.) moeten weten hoe groot de kans, zeg p, is dat een (enkelvoudige) geboorte een meisje oplevert. Deze kans p kan wel uit waarnemingen (tellingen) geschat worden, maar niet exact worden vastgesteld. Deze waarnemingen kunnen ook worden gebruikt om te concluderen dat p "wel tussen p_1 en p_2 zal liggen" of speciaal dat $p > \frac{1}{2}$ is. Methoden om te komen tot uitspraken van deze soort zijn het onderwerp van de volgende hoofdstukken.

6. Steekproef en populatie

Definitie 6.1: $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ heet een *aselecte steekproef* van de grootte n uit een *populatie* met verdelingsfunctie F, als $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onderling onafhankelijke stochastische grootheden zijn met $F_{\underline{x}_j} = F$ $(j=1,2,\ldots,n)$. Als $F = F_{\underline{x}}$, dan spreken we wel van een aselecte steekproef van $\underline{x}$ of een aselecte steekproef uit een verdeling met verdelingsfunctie F. (Een vrijwel identieke definitie geldt voor een aselecte steekproef uit een meer-dimensionale verdeling).

In deze definitie speelt het woord "populatie" geen wezenlijke rol; het wordt gebruikt om aan te geven, dat de steekproef ontstaan is door het doen van metingen of waarnemingen bij een deelverzameling van een verzameling (de populatie) waarin we geïnteresseerd zijn. Voorbeeld: we kiezen, (met teruglegging en met gelijke kansen) 100 Eindhovenaren en bepalen hun lengte; de populatie bestaat nu uit alle Eindhovenaren (of: hun lengten), de steekproef bestaat uit de 100 (te kiezen) grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{100}$, F(x) is gelijk aan de fractie van alle Eindhovenaren met een lengte niet groter dan x en $\underline{x}$ is de lengte van zo'n willekeurig te kiezen Eindhovenaar. Het is niet altijd helemaal duidelijk wat de populatie is: als we aannemen dat de kansverdeling van de lengten van bijv. Enschedeërs, toekomstige Eindhovenaren of zelfs van alle huidige en toekomstige Nederlanders ook F is, dan kunnen we de populatie dienovereenkomstig uitbreiden. Een ander voorbeeld: we gaan n keer (onafhankelijk) met een dobbelsteen gooien en daarbij telkens noteren hoeveel zessen (0 of 1) er boven komen. De steekproef bestaat dan uit de n (nog te verschijnen) aantallen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$; wat de populatie is, is in zekere mate een kwestie van smaak: alle mogelijke (gedane of nog te verrichten) worpen met die ene dobbelsteen of zelfs van alle soortgelijke dobbelstenen. We zullen ons verder niet met dit probleem bezighouden. Ook de manier waarop de steekproef tot stand komt, zullen we doorgaans in het midden laten. Problemen van deze soort komen aan de orde in de colleges Toegepaste Statistiek en Statistische Theorie van Proefopzetten.

We doen metingen alleen bij een deel van de populatie en niet bij de hele populatie omdat dat laatste meestal te kostbaar of te tijdrovend is. Sommige metingen zijn bovendien destructief: treksterktebepalingen, valproeven e.d.
Steekproeven worden gebruikt om iets over de populatie te weten te komen. In het eerste geval kunnen we bijvoorbeeld uit de steekproef een indruk krijgen van de gemiddelde lengte van de Eindhovenaren, $E\underline{x}$; in het tweede

geval kunnen we er achter komen of bij de gegeven dobbelsteen de kans op een zes (wel) 1/6 is. Het eerstgenoemde voorbeeld is een eenvoudig *schattingsprobleem*, het tweede een *toetsingsprobleem*.

In de meeste gevallen zullen we aannemen dat de populatie waar het over gaat een verdelingsfunctie $F_{\underline{x}}$ heeft van de gedaante $F_{\underline{x}}(x) = F(x;\theta)$, waarbij alleen de *waarde* van θ *onbekend* is. Hierbij zal de *parameter* θ meestal één-dimensionaal zijn. De steekproef dient dan om een indruk te krijgen van de waarde van deze parameter. Dikwijls zullen we in verband hiermee een *functie* van de steekproef beschouwen, waarvan we verwachten dat die de waarde van θ goed zal benaderen.

<u>Definitie 6.2:</u> Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$. Een *statistische grootheid* $\underline{t}$ is een stochastische grootheid (of een stochastische vector) die een *functie van de steekproef* is: $\underline{t} = t\,(\underline{x}_1,\ldots,\underline{x}_n)$. Als $F_{\underline{x}}(x) = F(x;\theta)$, dan hangt de functie t niet af van θ (de kansverdeling van $\underline{t}$ wel!).

<u>Opmerking 6.3:</u> Als de metingen of de waarnemingen gegeven zijn, dan is het resultaat een rijtje reële getallen (of vectoren): $x_1,\ldots,x_n$. We noemen dit resultaat een *steekproefrealisatie* of een *waargenomen steekproef* of ook wel "*de waarnemingen*". De statistische grootheid gaat hierbij over in een functie van deze waarnemingen: $t = t\,(x_1,\ldots,x_n)$. We noemen t dan een *realisatie* van $\underline{t}$.

We zullen in de volgende hoofdstukken drie basistechnieken uit de statistiek bespreken: (i) Het *schatten* van parameters, (ii) het *toetsen* van hypothesen en (iii) het bepalen van *betrouwbaarheidsintervallen*.
We illustreren dit aan een eenvoudig voorbeeldje.

<u>Voorbeeld 6.4:</u> We gooien 100 keer met een dobbelsteen en vinden 14 zessen. We kunnen nu (i) dit resultaat gebruiken om de kans p op een zes te schatten, bijv. $p \approx 14/100$. We kunnen ook (ii) er van uit gaan dat $p = 1/6$ zou moeten zijn en nagaan ("toetsen") of dit wel in overeenstemming is met ons resultaat van 14 zessen. Tenslotte kunnen we (iii) onderzoeken welke interval van waarden van p redelijkerwijs mogelijk is gezien dit resultaat. Het zal duidelijk zijn dat de drie problemen met elkaar verwant zijn.

<u>Opmerking 6.5</u>: Soms zullen we steekproeven $\underline{x}_1,\ldots,\underline{x}_n$ beschouwen, waarbij de $\underline{x}_j$ *niet* o.o. zijn of *niet dezelfde verdeling* hebben.

7. Schatten van parameters

7.1 Schatters en hun eigenschappen

We gaan er in dit hoofdstuk meestal van uit dat we beschikken over een aselecte steekproef $\underline{x}_1,\dots,\underline{x}_n$ van $\underline{x}$ met $F_{\underline{x}}(x) = F(x;\theta)$, waarbij $\theta \in \Omega$ met $\Omega \subset \mathbb{R}$ (of soms $\Omega \subset \mathbb{R}^k$). We noemen de verzameling Ω van parameterwaarden de *parameterruimte*. Hierbij nemen we aan dat F een bekende functie is van x en θ, waarbij θ voor de beschouwde populatie een vaste, onbekende waarde heeft; deze waarde noemt men dan wel de *ware parameterwaarde*. In een enkel geval hebben we te maken met niet gespecificeerde kansverdelingen: $F_{\underline{x}} = F$ met F onbekend, terwijl we iets willen weten over bijv. $E\underline{x}$ of var $\underline{x}$, waarvan we dan natuurlijk aannemen dat ze bestaan. In deze situatie spreekt men wel van parametervrije statistiek. Wij zullen ook in dit geval $E\underline{x}$ en var $\underline{x}$ als parameters noteren: $E\underline{x} = \mu$ en var $\underline{x} = \sigma^2$.

<u>(Meta-)Definitie 7.1.1:</u> Een *schatter* $\underline{\hat{\theta}}$ voor een parameter θ is een statistische grootheid $\hat{\theta}$ $(\underline{x}_1,\dots,\underline{x}_n)$ die bedoeld is om θ te schatten. Een realisatie $\hat{\theta} = \hat{\theta}$ $(x_1,\dots,x_n)$ van $\underline{\hat{\theta}}$ noemen we een *schatting* van θ. Soms noteren we een schatter voor μ,σ,τ, etc. als $\underline{m},\underline{s},\underline{t}$ etc.
De functie $\hat{\theta}$ hangt af van n; als n niet constant is, schrijven we $\hat{\theta}_n$ en evenzo m_n, s_n, t_n, etc.

Volgens bovenstaande definitie is het aangeven van een schatter weinig meer dan een intentieverklaring. Wat we nodig hebben zijn kwaliteitseisen en kriteria: welke eigenschappen zijn gewenst voor schatters en wanneer is de ene schatter beter dan de andere?

<u>Definitie 7.1.2:</u> Als $\underline{\hat{\theta}}_n$ een schatter is voor θ, dan heet

$$\theta - E\underline{\hat{\theta}}_n \qquad (\theta \in \Omega)$$

de *onzuiverheid* van $\underline{\hat{\theta}}_n$ (met betrekking tot θ). Als

$$E\underline{\hat{\theta}}_n = \theta \qquad (\theta \in \Omega),$$

dan heet $\underline{\hat{\theta}}_n$ een *zuivere schatter* voor θ. Als

$$\lim_{n\to\infty} E\,\underline{\hat{\theta}}_n = \theta \qquad (\theta \in \Omega),$$

dan heet (de rij schatters) $\underline{\hat{\theta}}_n$ *asymptotisch zuiver*.

Opmerking 7.1.3: $E\hat{\underline{\theta}}_n$ is een functie van θ, immers (als $F_{\underline{x}}$ absoluut continu is, en analoog in het diskrete geval)

$$E\hat{\underline{\theta}}_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \hat{\theta}_n(x_1,\ldots,x_n) f_{\underline{x}_1}(x_1;\theta)\ldots f_{\underline{x}_n}(x_n;\theta)dx_1 \ldots dx_n;$$

dit wordt soms aangegeven door de notatie $E_\theta\hat{\underline{\theta}}_n$.

Definitie 7.1.4: Als $\hat{\underline{\theta}}_n$ een schatter is voor θ, dan heet

$$\{E(\hat{\underline{\theta}}_n - \theta)^2\}^{\frac{1}{2}} \qquad (\theta \in \Omega)$$

de *onnauwkeurigheid* van $\hat{\underline{\theta}}_n$. Als $E(\hat{\underline{\theta}}_{n,o} - \theta)^2 \leq E(\hat{\underline{\theta}}_n - \theta)^2$ voor alle $\theta \in H$ en alle schatters $\hat{\underline{\theta}}_n$ in een klasse S, terwijl ook $\hat{\underline{\theta}}_{n,o} \in S$, dan heet $\hat{\underline{\theta}}_{n,o}$ een *nauwkeurigste* schatter in S (hierbij is impliciet verondersteld dat $E\hat{\underline{\theta}}_n^2 < \infty$ voor $\hat{\underline{\theta}}_n \in S$).

Definitie 7.1.5: Een rij schatters $\hat{\underline{\theta}}_n$ voor θ heet *asymptotisch nauwkeurig* als

$$\lim_{n\to\infty} E(\hat{\underline{\theta}}_n - \theta)^2 = 0 \qquad (\theta \in \Omega) \quad .$$

Definitie 7.1.6: Een rij schatters $\hat{\underline{\theta}}_n$ voor θ heet *asymptotisch raak* (of: *consistent*) als

$$\hat{\underline{\theta}}_n \xrightarrow{P} \theta \qquad (\theta \in \Omega),$$

d.w.z. als $\lim_{n\to\infty} P(|\hat{\underline{\theta}}_n - \theta| \geq \varepsilon) = 0$ voor iedere $\varepsilon > 0$.

Stelling 7.1.7: Als de rij schatters $\hat{\underline{\theta}}_n$ voor θ asymptotisch nauwkeurig is, dan is hij ook asymptotisch zuiver en consistent.

Bewijs: $E(\hat{\underline{\theta}}_n - \theta)^2 = \text{var}\,\hat{\underline{\theta}}_n + (E\,\hat{\underline{\theta}}_n - \theta)^2$ (ga na). Omdat gegeven is dat $E(\hat{\underline{\theta}}_n - \theta)^2 \to 0$ voor $n \to \infty$, geldt ook $E\hat{\underline{\theta}}_n \to \theta$ als $n \to \infty$, d.w.z. $\hat{\underline{\theta}}_n$ is asymptotisch zuiver.

Geheel analoog aan het bewijs van de ongelijkheid van Chebyshev (stelling 5.2.1) bewijst men

(7.1.1) $\quad P(|\hat{\underline{\theta}}_n - \theta| \le \varepsilon) \ge \varepsilon^{-2}E(\hat{\underline{\theta}}_n - \theta)^2.$

Volgens het gegeven gaat het rechter lid van (7.1.1) voor $n \to \infty$ naar nul.

Opmerking 7.1.8: Als $\hat{\underline{\theta}}_n$ zuiver is, dan is $E(\hat{\underline{\theta}}_n - \theta)^2 = \text{var } \hat{\underline{\theta}}_n$.

Voorbeeld 7.1.9: Zij $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een aselecte steekproef uit een Bernoulli-verdeling (binomiale verdeling met n=1) met onbekende parameter p. Als schatter voor p ligt nu voor de hand: de fractie enen in de steekproef, dus $\hat{\underline{p}}_n = \underline{k}_n/n$, waarbij $\underline{k}_n = \underline{x}_1 + \ldots + \underline{x}_n$. Nu is $\hat{\underline{p}}_n$ zuiver en asymptotisch nauwkeurig, immers (zie voorbeeld 4.1.2 b en opgave 5.2.3)

$$E\,\hat{\underline{p}}_n = \frac{1}{n} E\,\underline{k}_n = p, \quad E(\hat{\underline{p}}_n - p)^2 = \text{var } \hat{\underline{p}}_n = p(1-p)/n.$$

Definitie 7.1.10: Als $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een aselecte steekproef is van $\underline{x}$ met (onbekende) verwachting μ, dan heet

(7.1.2) $\quad \underline{m}_n := \bar{\underline{x}} = \frac{1}{n}\sum_1^n \underline{x}_j$

het *steekproefgemiddelde*;

(7.1.3) $\quad \underline{s}_n^2 := \frac{1}{n-1}\sum_1^n (\underline{x}_j - \underline{m}_n)^2$

heet de *steekproefvariantie*; hierbij nemen we aan dat $\sigma^2 := \text{var } \underline{x} < \infty$ is en mogelijk onbekend.

Stelling 7.1.11: Zij $\underline{x}_1, \ldots, \underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $E\,\underline{x} = \mu$ en $\text{var } \underline{x} = \sigma^2$, dan geldt

(7.1.4) $\quad E\,\underline{m}_n = \mu; \quad E\,\underline{s}_n^2 = \sigma^2,$

d.w.z. $\underline{\;}_n$ en $\underline{s}_n^2$ zijn zuivere schatters voor resp. μ en σ^2.

Bewijs: Dat $E\,\underline{m}_n = \mu$ zagen we al in opgave 5.2.3. Om te bewijzen dat $\underline{s}_n^2$ zuiver is beschouwen we $(n-1)E\,\underline{s}_n^2$:

$$E\sum_1^n (\underline{x}_j - \underline{m}_n)^2 = E\sum_1^n \{(\underline{x}_j - \mu) - (\underline{m}_n - \mu)\}^2 = E\sum_1^n (\underline{x}_j - \mu)^2 - n\,E(\underline{m}_n - \mu)^2 = (n-1)\sigma^2,$$

waarbij we gebruik hebben gemaakt van de identiteit

(7.1.5) $\sum_1^n (a_k - \bar{a})^2 = \sum_1^n a_k^2 - n\bar{a}^2$,

waarbij $\bar{a} = \frac{1}{n}\sum_1^n a_k$.

<u>Opgave 7.1.12:</u> Bewijs (7.1.5) door var $\underline{a}$ te beschouwen voor een stochastische grootheid $\underline{a}$ met $P(\underline{a} = a_k) = \frac{1}{n}$ $(k=1,2,\ldots,n)$.

<u>Opgave 7.1.13:</u> Als $E\underline{x} = \mu$ *bekend* is, dan is

(7.1.6) $\underline{s}_0^2 := n^{-1} \sum_1^n (\underline{x}_k - \mu)^2$

een zuivere schatter voor $\sigma^2 = \text{var } \underline{x}$.

<u>Stelling 7.1.14:</u> *$\underline{m}_n$ is de nauwkeurigste schatter voor μ onder alle zuivere lineaire schatters voor μ, d.w.z. onder alle zuivere schatters $\hat{\underline{\mu}}_n$ van de vorm*

(7.1.7) $\hat{\underline{\mu}}_n = \sum_1^n \alpha_k \underline{x}_k$.

Bewijs: Omdat $E\,\hat{\underline{\mu}}_n = \mu$, volgt uit (7.1.7) dat $\sum_1^n \alpha_k = 1$. De onnauwkeurigheid van $\hat{\underline{\mu}}_n$ wordt gegeven door (zie opmerking 7.1.8) $\sigma(\hat{\underline{\mu}}_n)$. Deze onnauwkeurigheid is minimaal, als $\sum_1^n \alpha_k^2$ minimaal is onder de nevenvoorwaarde dat $\sum_1^n \alpha_k = 1$. Zij nu $\delta_k = \alpha_k - 1/n$,dan is $\sum_1^n \delta_k = 0$ en (ga na)

$$\sum_1^n \alpha_k^2 = \sum_1^n \delta_k^2 + \frac{1}{n};$$

dit is minimaal als $\delta_k = 0$, d.w.z. $\alpha_k = \frac{1}{n}$ voor $k=1,2,\ldots,n$.

<u>Opmerking 7.1.15:</u> Het is soms mogelijk om de nauwkeurigheid te vergroten ten koste van de zuiverheid. Als we μ schatten met

$$\hat{\underline{\mu}}_n = a\,\underline{m}_n = a/n \sum_1^n \underline{x}_k,$$

dan geldt

$$(7.1.8) \qquad E\,\underline{\hat{\mu}}_n = a\mu \;,\; E(\underline{\hat{\mu}}_n-\mu)^2 = a^2\,\frac{\sigma^2}{n} + (1-a)^2\,\mu^2 .$$

De onnauwkeurigheid is minimaal en *kleiner* dan σ^2/n voor $a = \mu^2/(\mu^2+\sigma^2/n)$.

<u>Opmerking 7.1.16:</u> Het kan voorkomen dat er voor een parameter geen zuivere schatter bestaat. Voorbeeld: zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $P(\underline{x}=1)=1/\theta$ en $P(\underline{x}=0)=1-1/\theta$ met $\theta \in (1,\infty)$.
Stel nu dat $\underline{\hat{\theta}} = \hat{\theta}\,(\underline{x}_1,\ldots,\underline{x}_n)$ een zuivere schatter is voor θ, dan geldt

$$(7.1.9) \qquad E\underline{\hat{\theta}} = \sum \ldots \sum \hat{\theta}(x_1,\ldots,x_n)\,\left(\frac{1}{\theta}\right)^{\sum_1^n x_k}\left(1-\frac{1}{\theta}\right)^{n-\sum_1^n x_k} = \theta ,$$

waarbij gesommeerd wordt over alle (2^n) mogelijkheden voor $(x_1,\ldots,x_n)$.
Nu staat in het linker lid van (7.1.9) een polynoom in $1/\theta$; dit kan onmogelijk voor alle $\theta \in (1,\infty)$ gelijk zijn aan θ.

<u>Opgave 7.1.17:</u> Ga na dat voor n o.o. grootheden $\underline{x}_1,\ldots,\underline{x}_n$ met $P(\underline{x}_k=0)=1-p$ en $P(\underline{x}_k=1)=p$ $(k=1,2,\ldots,n)$ geldt

$$(7.1.10) \qquad P((\underline{x}_1,\ldots,\underline{x}_n) = (x_1,\ldots,x_n)) = p^{x_1+\ldots+x_n}(1-p)^{n-(x_1+\ldots+x_n)} .$$

Soms is het mogelijk om van een schatter voor θ te laten zien dat hij de nauwkeurigste is onder alle zuivere schatters. Dit kan op grond van de volgende stelling.

<u>Stelling 7.1.18</u> (ongelijkheid van Cramér -Rao): Zij $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $F_{\underline{x}}(x) = F(x;\theta)$ en zij $\underline{t} = t\,(\underline{x}_1,\ldots,\underline{x}_n)$ een schatter voor θ met $E\,\underline{t} = \alpha(\theta)$, dan geldt (onder weinig beperkende regulariteitsvoorwaarden)

$$(7.1.11) \qquad \operatorname{var} \underline{t} \geq \frac{\{\alpha'(\theta)\}^2}{n\,E\left\{\frac{\partial \log p(\underline{x};\theta)}{\partial\theta}\right\}^2} = \frac{-\{\alpha'(\theta)\}^2}{nE\,\frac{\partial^2 \log p(\underline{x};\theta)}{\partial\theta^2}} ,$$

waarin

$$(7.1.12)\quad p(x;\theta) = \begin{cases} f_{\underline{x}}(x) & \text{als } F_{\underline{x}} \text{ absoluut continu is} \\ P(\underline{x}=x) & \text{als } F_{\underline{x}} \text{ diskreet is.} \end{cases}$$

Gelijkheid geldt in (7.1.11) dan en slechts dan als er een constante $c_n(\theta)$ bestaat zó dat met kans 1

$$(7.1.13)\quad \sum_{k=1}^{n} \frac{\partial \log p(\underline{x}_k;\theta)}{\partial\theta} = c_n(\theta)\ (\underline{t} - \alpha(\theta)).$$

Bewijs: voor het bewijs van een wat algemenere en preciezer geformuleerde stelling verwijzen we naar de Appendix (stelling A 3.1).

Relatie (7.1.13) kunnen we gebruiken om een nauwkeurigste *zuivere* schatter te vinden (vergelijk definitie 7.1.4). We komen hier in § 7.2 op terug.

Voorbeeld 7.1.19: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $P(\underline{x} = x) = e^{-\lambda}\lambda^x/x!$ $(x=0,1,2,\ldots)$, d.w.z. in de notatie van stelling 7.1.18 dat

$$\log p(x;\lambda) = -\lambda + x \log \lambda - \log(x!)$$

en dat

$$(7.1.14)\quad \frac{\partial \log p(x;\lambda)}{\partial\lambda} = -1 + \frac{x}{\lambda}\ ;\ \frac{\partial^2 \log p(x;\lambda)}{\partial\lambda^2} = -\frac{x}{\lambda^2}\,,$$

en dus

$$-\, n\, E\, \frac{\partial^2 \log p(x;\lambda)}{\partial\lambda^2} = n/\lambda.$$

Omdat $\alpha(\lambda) = \lambda$ en dus $\alpha'(\lambda) = 1$, is de kleinst mogelijke variantie gelijk aan λ/n. Het is nu duidelijk dat $\underline{t} := \underline{m}_n = n^{-1}\sum_1^n \underline{x}_k$ een zuivere schatter is met minimale variantie. Uit (7.1.13) volgt ook met (7.1.14) dat

$$\underline{t} = \lambda + c_n^{-1}(\lambda)\{\sum_1^n \underline{x}_k/\lambda - n\};$$

omdat $\underline{t}$ een statistische grootheid moet zijn en dus niet van λ mag afhangen is de keuze $c_n(\lambda) = n/\lambda$ de enig mogelijke.

Dit levert ook $\underline{t}_n = n^{-1} \sum_1^n \underline{x}_k$.

Opgave 7.1.20: Zij $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een aselecte steekproef van $\underline{x}$, waarbij $\underline{x}$ normaal verdeeld is met E $\underline{x} = \mu$(onbekend) en var $\underline{x} = 1$. Laat zien dat $\underline{m}_n$ (zie (7.1.2)) een zuivere schatter is voor μ met minimale variantie.

Opgave 7.1.21: Als $\underline{t}$ aan (7.1.13) voldoet, dan geldt $c_n^2(\theta) = (\alpha'(\theta))^2/(\text{var } \underline{t})^2$.

7.2 Methoden voor het vinden van schatters

In eenvoudige gevallen is het niet moeilijk om een schatter te bedenken voor een parameter. Zo ligt $\underline{m}_n$ voor de hand als schatter voor E $\underline{x}$ en $\underline{s}_0^2$ als schatter voor σ^2 als E $\underline{x} = \mu$ bekend is (zie(7.1.6)). In ingewikkelder situaties is het van belang een formele methode te hebben om schatters voor een parameter te vinden. Eén zo'n methode berust op het zgn. principe van de *grootste aannemelijkheid* (*maximum likelihood principle*).

Definitie 7.2.1: Als $(\underline{x}_1, \ldots, \underline{x}_n)$ een stochastische vector is met verdelingsfunctie

$$F_{\underline{x}_1, \ldots, \underline{x}_n}(x_1, \ldots, x_n) = F(x_1, \ldots, x_n; \theta),$$

waarin θ een (mogelijk meer-dimensionale) parameter is, dan heet de functie L gedefinieerd door

$$\text{(7.2.1)} \qquad L(x_1, \ldots, x_n; \theta) = \begin{cases} f_{\underline{x}_1, \ldots, \underline{x}_n}(x_1, \ldots, x_n) & \text{als } F_{\underline{x}_1, \ldots, \underline{x}_n} \text{ abs.kont. is} \\ P(\underline{x}_1 = x_1, \ldots, \underline{x}_n = x_n) & \text{als } F_{\underline{x}_1, \ldots, \underline{x}_n} \text{ diskreet is,} \end{cases}$$

de *aannemelijkheidsfunctie* van $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$. Als $\underline{x}_1, \ldots, \underline{x}_n$ een aselecte steekproef is van $\underline{x}$ met $F_{\underline{x}}(x) = F(x;\theta)$, dan gaat (7.2.1) over in (zie (7.1.12))

$$\text{(7.2.2)} \qquad L(x_1, \ldots, x_n; \theta) = \prod_1^n p(x_j; \theta).$$

Definitie 7.2.2: Zij $(x_1,\dots,x_n)$ een realisatie van een stochastische vector met verdelingsfunctie $F(x_1,\dots,x_n;\theta)$. Een *meest aannemelijke schatting* van θ, gebaseerd op de waarneming $(x_1,\dots,x_n)$, is een waarde $\hat\theta = \hat\theta(x_1,\dots,x_n)$ met de eigenschap dat

$$\max_{\theta\in\Omega} L(x_1,\dots,x_n;\theta) = L(x_1,\dots,x_n;\hat\theta). \tag{7.2.3}$$

Als $\hat\theta$ voor iedere $(x_1,\dots,x_n)$ eenduidig bepaald is, en als $\hat\theta(\underline{x}_1,\dots,\underline{x}_n)$ een stochastische grootheid (stochastische vector) is, dan heet

$$\underline{\hat\theta} := \hat\theta\ (\underline{x}_1,\dots,\underline{x}_n)$$

de *meest aannemelijke schatter* (*maximum likelihood estimator*) voor θ. We zullen $\underline{\hat\theta}$ bij wijze van afkorting wel de ML-schatter noemen.

Bovenstaande definitie is gebaseerd op de overweging dat een meest aannemelijke schatting voor θ die waarde van θ is, die aan de gevonden waarnemingen $x_1,\dots,x_n$ in het diskrete geval de grootst mogelijke kans geeft en in het absoluut kontinue geval de grootst mogelijke kansdichtheid.

Voorbeeld 7.2.3: Zij $\underline{x}_1, \underline{x}_2,\dots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sqrt{2\pi}}\, e^{-(x-\mu)^2/2} \qquad (\mu \in \mathbb{R}).$$

Wat is de ML-schatter $\underline{\hat\mu}$ voor μ?

Antwoord: $L(x_1,\dots,x_n;\mu)$ is maximaal als $\log L(x_1,\dots,x_n;\mu)$ maximaal is, en

$$\log L(x_1,\dots,x_n;\mu) = -\frac{n}{2}\log(2\pi) - \tfrac{1}{2}\sum_1^n (x_k-\mu)^2.$$

Men gaat eenvoudig na dat L maximaal is voor $\mu = n^{-1}\sum_1^n x_k$, d.w.z. dat de meest aannemelijke schatter voor μ gegeven wordt door $\underline{\hat\mu} = \underline{m}_n$ (zie definitie 7.1.10).

Voorbeeld 7.2.4: Zij $\underline{x}_1,\dots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met

$$P(\underline{x} = 0) = 1 - p \;;\quad P(\underline{x} = 1) = p \qquad\qquad (0 \le p \le 1).$$

Bepaal de ML-schatter voor p.

Antwoord: $L(x_1,\dots,x_n;\, p) = p^{\sum_1^n x_k}(1-p)^{n-\sum_1^n x_k}$ (zie opgave 7.1.17).

Als $0 < \sum_1^n x_k < n$, dan is $L = 0$ voor $p = 0$ en $p = 1$. Voor $0 < p < 1$ geldt dat L maximaal is als log L maximaal is, terwijl

$$\log L(x_1,\dots,x_n) = \sum_1^n x_k \log p + (n - \sum_1^n x_k) \log (1 - p).$$

Dit is maximaal voor $p = n^{-1} \sum_1^n x_k$. Als $\sum_1^n x_k = 0$, dan is L maximaal (gelijk aan 1) voor p=0; als $\sum_1^n x_k = n$, dan is L maximaal voor p=1.

Conclusie: de ML-schatter voor p is

$$\underline{\hat{p}} = \underline{k}/n,$$

met $\underline{k} = \underline{x}_1 + \dots + \underline{x}_n$.

Opgave 7.2.5: Zij $\underline{x}_1, \underline{x}_2,\dots, \underline{x}_n$ een aselecte steekproef uit een homogene verdeling op $[0,\theta]$, $\theta > 0$. Bepaal de ML-schatter voor θ.

Voorbeeld 7.2.6: Zij $\underline{x}_1, \underline{x}_2,\dots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad\qquad (\mu \in \mathbb{R}\,;\ \sigma^2 > 0).$$

$L(x_1,\dots,x_n;\, \mu,\sigma^2)$ is maximaal, als log L maximaal is. We schrijven $\lambda = 1/(2\sigma^2)$ en $\ell = \log L$. We krijgen dan

$$\ell(x_1,\dots,x_n;\, \mu\,,\lambda) = \text{const.} + \frac{n}{2} \log \lambda - \lambda \sum_1^n (x_j-\mu)^2$$

met $\lambda > 0$ en $\mu \in \mathbb{R}$. We vinden de stationaire punten uit

$$\frac{\partial\ell}{\partial\lambda} = \frac{n}{2\lambda} - \sum_1^n (x_j - \mu)^2 = 0$$

$$\frac{\partial \ell}{\partial \mu} \equiv 2\lambda \sum_{1}^{n} (x_j - \mu) = 0.$$

De enige oplossing hiervan is $\mu = \hat{\mu} = \frac{1}{n}\sum_1^n x_j$ en $\lambda = \hat{\lambda} = (2\sum_1^n (x_j-\hat{\mu})^2/n)^{-1}$, d.w.z. $\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_j - \hat{\mu})^2$. Verder vinden we voor $\lambda = \hat{\lambda}$ en $\mu = \hat{\mu}$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -n/(2\lambda^2); \ \frac{\partial^2 \ell}{\partial\lambda\partial\mu} = \frac{\partial^2 \ell}{\partial\mu\partial\lambda} = 0 \ ; \ \frac{\partial^2 \ell}{\partial \mu^2} = -2n\lambda.$$

Dit betekent dat L een uniek maximum heeft in $(\hat{\lambda},\hat{\mu})$, zodat de ML-schatter voor (μ,σ^2) gegeven wordt door $(\underline{\hat{\mu}}, \underline{\hat{\sigma}}^2)$ met

$$\underline{\hat{\mu}} = \frac{1}{n}\sum_1^n \underline{x}_j = \underline{m}_n$$

$$\underline{\hat{\sigma}}^2 = \frac{1}{n}\sum_1^n (\underline{x}_j - \underline{m}_n)^2,$$

waarbij dus $\underline{\hat{\sigma}}^2$ blijkens stelling 7.1.11 niet zuiver is (wel asymptotisch zuiver; $\underline{\hat{\mu}}$ en $\underline{\hat{\sigma}}^2$ zijn natuurlijk functies van n).

Opgave 7.2.7: Laat het volgende zien. Als $\underline{\hat{\theta}}$ de meest aannemelijke schatter is voor θ en als u: $\Omega \to \Omega'$ een een-eenduidige funktie is, zō dat $u(\underline{\hat{\theta}})$ een stochastische grootheid is, dan is $u(\underline{\hat{\theta}})$de ML-schatter voor $u(\theta)$ (deze eigenschap wordt wel *invariantie* genoemd).

Opmerking 7.2.8: Onder vrij algemene voorwaarden geldt voor ML-schatters

a. $E(\underline{\hat{\theta}}_n-\theta)^2 \to 0$ (en zelfs $\underline{\hat{\theta}}_n \overset{a.s.}{\to} \theta$) $\qquad (n \to \infty;\ \theta \in \Omega)$

b. $\underline{\hat{\theta}}_n$ is asymptotisch normaal (vergelijk definitie 5.4.4).

Opgave 7.2.9: De schatter $\underline{s}_n^2$ (zie 7.1.3) is zuiver voor σ^2. Laat zien dat $\underline{s}_n$ niet zuiver is voor σ.

Een iets primitievere methode om schatters te krijgen is de *momentenmethode*. We geven eerst een definitie.

Definitie 7.2.10: Als $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef is van $\underline{x}$, dan heet

(7.2.4) $\underline{m}^{(k)} = \underline{m}_n^{(k)} := \frac{1}{n}\sum_1^n \underline{x}_j^k$

het k-de *steekproefmoment*. Een realisatie van $\underline{m}^{(k)}$ geven we aan met $m^{(k)}$.

Definitie 7.2.11: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $\mu_k = E\,\underline{x}^k$ $(k=1,2,\ldots,r)$ en $F_{\underline{x}}(x) = F(x;\,\theta_1,\ldots,\,\theta_r)$. Als de vergelijkingen

(7.2.5) $\mu_k = \mu_k(\theta_1,\ldots,\theta_r) = m^{(k)}$ $(k=1,\ldots,r)$

één oplossing $(\hat\theta_1,\ldots,\hat\theta_r)$ hebben zó dat $\underline{\hat\theta}_k := \hat\theta_k(\underline{m}^{(1)},\ldots,\underline{m}^{(r)})$ een stochastische grootheid is $(k=1,\ldots,r)$, dan heet $(\underline{\hat\theta}_1,\ldots,\underline{\hat\theta}_r)$ de *momentenschatter* voor $(\theta_1,\ldots,\theta_r)$.

Voorbeeld 7.2.12: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $E\,\underline{x} = \mu$ en var $\underline{x} = \sigma^2$. Voor r=2 vinden we uit (7.2.5)

$$m^{(1)} = \mu_1 \text{ en } m^{(2)} = \mu_2 = \sigma^2 + \mu_1^2,$$

omdat $\sigma^2 = \mu_2 - \mu_1^2$. De momenten schatters voor μ_1 en σ^2 zijn dus

$$\underline{\hat\mu}_1 = \underline{m}^{(1)} \quad ; \quad \underline{\hat\sigma}^2 = \underline{m}^{(2)} - (\underline{m}^{(1)})^2.$$

Opgave 7.2.13: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef uit een homogene verdeling op (θ_1,θ_2). Bereken de momentenschatter en de ML-schatter voor (θ_1,θ_2).

Opmerking 7.2.14: Ook voor momentenschatters geldt onder vrij algemene voorwaarden dat zij asymptotisch nauwkeurig zijn en asymptotisch normaal.

Opmerking 7.2.15: Beide hierboven behandelde schattingsmethoden kunnen ook gebruikt worden als de steekproef niet bestaat uit onafhankelijke grootheden (vergelijk definities 7.2.1 en 7.2.2). Afhankelijkheid van de waarnemingen heeft natuurlijk wel invloed op de eigenschappen van de schatters. Bovendien moet (in ieder geval in het geval van de ML-schatters) de simultane kansverdeling bekend zijn.

7.3 Voldoende schatters

Als $\underline{x}_1$, $\underline{x}_2,\ldots,\underline{x}_n$ diskrete stochastische grootheden zijn, die een kansverdeling hebben die *niet* van θ afhangt, d.w.z. als

$$P(\underline{x}_1 = x_1,\ldots,\underline{x}_n = x_n)$$

niet van θ afhangt, dan kan men uit de steekproef $\underline{x}_1,\ldots,\underline{x}_n$ geen informatie krijgen over deze parameter. Immers eigenschappen als $\underline{t} = t(\underline{x}_1,\ldots,\underline{x}_n) \xrightarrow{P} \theta$ voor $\theta \in \Omega$ kunnen alleen gelden als de verdeling van $\underline{t}$ afhangt van θ. Op een soortgelijke overweging berust de volgende definitie.

Definitie 7.3.1: Zij $(\underline{x}_1,\ldots,\underline{x}_n)$ een diskrete stochastische vector, waarvan de kansverdeling afhangt van een parameter θ ($\theta \in \Omega$). Een statistische grootheid $\underline{t} = t(\underline{x}_1,\ldots,\underline{x}_n)$ heet *voldoende* voor θ, als

$$(7.3.1) \qquad P(\underline{x}_1 = x_1,\ldots,\underline{x}_n = x_n \mid \underline{t} = t) \quad \text{niet afhangt van } \theta,$$

d.w.z. als de voorwaardelijke kansverdeling van $\underline{x}_1,\ldots,\underline{x}_n$ gegeven $\underline{t} = t$ niet van θ afhangt.

In analogie met de in de eerste alinea beschreven situatie zegt men nu dat $\underline{t}$ alle in de steekproef $\underline{x}_1,\ldots,\underline{x}_n$ aanwezige informatie over θ bevat: als $t(\underline{x}_1,\ldots,\underline{x}_n)$ gegeven is, dan geeft verdere precisering van $(\underline{x}_1,\ldots,\underline{x}_n)$ geen nadere informatie over θ: $\underline{t}$ is voldoende voor θ.

Voorbeeld 7.3.2: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $P(\underline{x}=1) = \theta$ en $P(\underline{x}=0) = 1-\theta$, waarbij θ onbekend is. Nu is $\underline{k}_n = \underline{x}_1 + \ldots + \underline{x}_n$ voldoende voor θ, immers (ga na)

$$(7.3.2) \qquad P(\underline{x}_1=x_1,\ldots,\underline{x}_n = x_n \mid \underline{k}_n=k) = \begin{cases} \binom{n}{k}^{-1} & \text{als } x_1 + \ldots + x_n = k \\ 0 & \text{anders} \end{cases},$$

d.w.z. de voorwaardelijke kansverdeling is een diskrete homogene verdeling over de $\binom{n}{k}$ mogelijke punten. We kunnen nu het idee dat $\underline{k}_n$ alle informatie over θ bevat ook als volgt duidelijk maken: als we de kansverdeling van $\underline{k}_n$ kennen, dan kunnen we de kansverdeling van de hele steekproef $\underline{x}_1,\ldots,\underline{x}_n$ terugvinden zonder dat we θ kennen, immers

$$P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n) = \sum_{k=0}^{n} P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n \mid \underline{k}_n=k)P(\underline{k}_n=k),$$

waarbij blijkens het voorgaande alleen $P(\underline{k}_n=k)$ van θ afhangt.

Stelling 7.3.3 (*Factorisatiestelling*): Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef uit een diskrete verdeling met verdelingsfunctie $F(x;\theta)$. Een statistische grootheid $\underline{t}$ is dan en slechts dan voldoende voor θ als

$$P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n) = h(t(x_1,\ldots,x_n);\theta)k(x_1,\ldots,x_n), \tag{7.3.3}$$

waarbij k niet afhangt van θ.

Bewijs: Als $\underline{t}$ voldoende is voor θ, dan geldt

$$P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n) = P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n \mid \underline{t}=t(x_1,\ldots,x_n))\, P(\underline{t}=t(x_1,\ldots,x_n))$$

waarbij o.g.v. definitie 7.3.1 de eerste factor niet van θ afhangt.
Als (7.3.3) geldt, dan is

$$P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n \mid \underline{t}=t)=\begin{cases} 0 & \text{als } t \neq t(x_1,\ldots,x_n) \\ \dfrac{P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n)}{P(\underline{t}=t)} & \text{als } t = t(x_1,\ldots,x_n) \end{cases}$$

onafhankelijk van θ, omdat $P(\underline{t}=t) = \sum\limits_{t(x_1,\ldots,x_n)=t} P(\underline{x}_1=x_1,\ldots,\underline{x}_n=x_n)$

alleen via de factor $h(t;\theta)$ van θ afhangt.

Definitie 7.3.4: Zij $(\underline{x}_1,\ldots,\underline{x}_n)$ een stochastische vector met kansdichtheid $f(x_1,\ldots,x_n;\theta)$. Een statistische grootheid $\underline{t}=t(\underline{x}_1,\ldots,\underline{x}_n)$ heet *voldoende* voor θ, als f in de volgende vorm gebracht kan worden:

$$f(x_1,\ldots,x_n;\theta) = h(t(x_1,\ldots,x_n);\theta)k(x_1,\ldots,x_n), \tag{7.3.4}$$

waarbij k niet van θ afhangt.

Opmerking 7.3.5: Men definieert ook wel: $\underline{t}$ is voldoende voor θ als $F_{\underline{x}_1,\ldots,\underline{x}_n}(x_1,\ldots,x_n \mid \underline{t}=t)$ niet van θ afhangt, analoog aan definitie 7.3.1.

Een eigenschap als (7.3.4) kan dan uit deze definitie worden afgeleid.
Dit leidt tot voor ons te grote wiskundige moeilijkheden.

Voorbeeld 7.3.6: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met $f_{\underline{x}}(x) = \theta e^{-\theta x}$ $(x > 0)$. Nu is $\underline{x}_1+\ldots+\underline{x}_n$ voldoende voor θ, immers

$$f_{\underline{x}_1,\ldots,\underline{x}_n}(x_1,\ldots,x_n) = \theta^n e^{-\theta(x_1+\ldots+x_n)} \quad (x_1 > 0,\ldots,x_n > 0)$$

voldoet aan (7.3.4) met $k(x_1,\ldots,x_n) = 1$.

Stelling 7.3.7: Als $\underline{t} = t(\underline{x}_1,\ldots,\underline{x}_n)$ voldoende is voor θ en als $\underline{\hat{\theta}}$ de ML-schatter is voor θ, dan is $\underline{\hat{\theta}}$ een functie van $\underline{t}$.

Bewijs: We hebben (zie definitie 7.2.1) omdat $\underline{t}$ voldoende is

$$(7.3.5) \qquad L(x_1,\ldots,x_n;\theta) = h(t(x_1,\ldots,x_n);\theta)k(x_1,\ldots,x_n),$$

waarbij k niet van θ afhangt. Uit (7.3.5) volgt dat L maximaal is voor die waarde van θ, waarvoor $h(t;\theta)$ maximaal is. Dit betekent dat $\hat{\theta}$ een functie is van t, d.w.z. $\hat{\theta}(x_1,\ldots,x_n)$ hangt alleen via $t=t(x_1,\ldots,x_n)$ van $x_1,\ldots,x_n$ af.

Stelling 7.3.8: Zij $\underline{t}$ voldoende voor θ.

a. Als u: $\mathbb{R} \to \mathbb{R}$ een omkeerbare funktie is zo dat $\underline{u} = u(\underline{t})$ een stochastische grootheid is, dan is ook $\underline{u}$ voldoende voor θ.

b. Als v: $\mathbb{R} \to \mathbb{R}$ een omkeerbare funktie is, dan is $\underline{t}$ voldoende voor $v(\theta)$.

Bewijs: a. zij τ de inverse funktie van u, dan geldt

$$L(x_1,\ldots,x_n;\theta) = h(t;\theta)k(x_1,\ldots,x_n) = h(\tau(u);\theta)k(x_1,\ldots,x_n).$$

Het bewijs van b is analoog.

Opgave 7.3.9: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2}.$$

Laat zien dat de ML-schatter voor μ voldoende is voor μ.

Opgave 7.3.10: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}.$$

Laat zien dat de ML-schatter voor σ voldoende is voor σ.

Voorbeeld 7.3.11: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}\ .$$

We zagen in voorbeeld 7.2.6 dat de ML-schatter voor (μ,σ^2) gelijk is aan $(\hat{\underline{\mu}},\ \hat{\underline{\sigma}}^2)$ met

$$\hat{\underline{\mu}} = \frac{1}{n}\sum_1^n \underline{x}_j \quad ; \ \hat{\underline{\sigma}}^2 = \frac{1}{n}\sum_1^n (\underline{x}_j - \hat{\underline{\mu}})^2.$$

We laten zien dat het paar $(\hat{\underline{\mu}},\hat{\underline{\sigma}}^2)$ voldoende is voor (μ,σ^2). We hebben

$$\begin{aligned} L(x_1,\ldots,x_n;\mu,\sigma^2) &= (\frac{1}{2\pi})^{n/2}\frac{1}{\sigma^n}\exp[-\frac{1}{2\sigma^2}\sum_1^n (x_j-\mu)^2] = \\ &= (\frac{1}{2\pi})^{n/2}\frac{1}{\sigma^n}\exp[-\frac{n}{2\sigma^2}\{\hat{\sigma}^2 + (\hat{\mu}-\mu)^2\}] = \\ &= (\frac{1}{2\pi})^{n/2}\, h(\hat{\mu},\hat{\sigma}^2;\mu,\sigma^2). \end{aligned}$$

Opmerking 7.3.12: Het begrip "voldoende" komt duidelijk tot uitdrukking in het volgende voorbeeld. Laat $\underline{k}_1,\ \underline{k}_2,\ldots,\underline{k}_n$ waarnemingen zijn van aantallen gebeurtenissen die optreden bij een Poisson-proces (zie § A1) gedurende disjuncte tijdsintervallen van (bekende) lengte $t_1,t_2,\ldots,t_n$. We willen de onbekende parameter λ van het Poisson-proces schatten. We beschouwen

$$\begin{aligned} P(\underline{k}_1=k_1,\ldots,\underline{k}_n=k_n) &= e^{-\lambda(t_1+\ldots+t_n)}\prod_1^n\{(\lambda t_j)^{k_j}/k_j!\} = \\ &= e^{-\lambda(t_1+\ldots+t_n)}\lambda^{k_1+\ldots+k_n}\prod_1^n\frac{t_j^{k_j}}{k_j!}\ , \end{aligned}$$

waaruit blijkt dat $\underline{k}_1+\ldots+\underline{k}_n$ voldoende is voor λ. Om λ te schatten is het dus "voldoende" om alleen dit totaal aantal gebeurtenissen dat in de tijd $t_1+\ldots+t_n$ optreedt te tellen. De ML-schatter voor λ is (ga na)

$\underline{\hat{\lambda}} = \sum_1^n \underline{k}_j / \sum_1^n t_j.$

Stelling 7.3.13: Als $\underline{t}$ een schatter is voor θ met minimale variantie, d.w.z. als in (7.1.11) het gelijkheidsteken geldt, dan is $\underline{t}$ voldoende voor θ.

Bewijs: Blijkens (7.1.13) is met kans 1 (zie (7.2.1) en (7.2.2))

$$\frac{\partial \log L\ (\underline{x}_1,\dots,\underline{x}_n;\theta)}{\partial\theta} = c_n(\theta)\{\ t\ (\underline{x}_1,\dots,\underline{x}_n) - \alpha(\theta)\},$$

d.w.z. (bijna) overal waar $L(x_1,\dots,x_n;\theta)$ positief is geldt

$$\frac{\partial \log L(x_1,\dots,x_n;\theta)}{\partial\theta} = c_n(\theta)\ \{t(x_1,\dots,x_n) - \alpha(\theta)\}.$$

Dit betekent dat $\log L(x_1,\dots,x_n;\theta)$ de volgende vorm heeft:

$$\log L(x_1,\dots,x_n;\theta) = q(t(x_1,\dots,x_n);\theta) + c(x_1,\dots,x_n)$$

en dus L de vorm

$$L(x_1,\dots,x_n;\theta) = h(t(x_1,\dots,x_n);\theta)k(x_1,\dots,x_n),$$

d.w.z. $\underline{t}$ is voldoende voor θ (zie (7.3.3) en (7.3.4)).

Stelling 7.3.14 (Rao-Blackwell): Zij $\underline{t}$ voldoende voor θ en zij $\underline{u}$ een zuivere schatter voor $h(\theta)$. Zij verder $\underline{v}$ gedefinieerd door (zie p.57)

$$\underline{v} := E(\underline{u}|\underline{t}).$$

Dan geldt

(i) $\underline{v}$ is een statistische grootheid; $\underline{v} = v(\underline{t})$,

(ii) $E\underline{v} = h(\theta)$,

(iii) $\text{var}\ \underline{v} \le \text{var}\ \underline{u}$ met $\text{var}\ \underline{v} = \text{var}\ \underline{u}$ d.e.s.d. als $P(\underline{u}=\underline{v})=1$.

Bewijs: We beperken het bewijs tot het geval dat $(\underline{x}_1,\dots,\underline{x}_n)$ en dus ook $\underline{u}$ en $\underline{t}$ een diskrete verdeling hebben.

(i) $v(t) := E(\underline{u}|\underline{t}=t) = \sum\dots\sum u(x_1,\dots,x_n)\ P(\underline{x}_1=x_1,\dots,\underline{x}_n=x_n|\underline{t}=t).$

Omdat $\underline{t}$ voldoende is voor θ hangt deze uitdrukking niet van θ af, maar alleen van t. Uit de definities van $\underline{v}$ en van $E(\underline{u}\ |\ \underline{t})$ (zie blz. 57) volgt nu dat $\underline{v} = v(\underline{t})$.

(ii) $E\,\underline{v} = E\,E(\underline{u}\,|\underline{t}) = E\,\underline{u} = h(\theta)$ (zie(4.3.4))

(iii) $\mathrm{var}\,\underline{u} = E(\underline{u} - h(\theta))^2 = E(\underline{u}-\underline{v}+(\underline{v}-h(\theta)))^2 = E(\underline{u}-\underline{v})^2 + \mathrm{var}\,\underline{v}$,
omdat $E\,(\underline{u}-\underline{v})(\underline{v}-h(\theta)) = E(\underline{u}-\underline{v})\underline{v} = E\;\;E((\underline{u}-\underline{v})\underline{v}|\underline{t})$, met

(7.3.6) $E((\underline{u}-\underline{v})\underline{v}|\underline{t}=t) = v(t)\;E(\underline{u}-\underline{v}\;|\underline{t}=t) = v(t)\;\{E(\underline{u}|\underline{t}=t) - v(t)\} = 0.$

Nu is dus $\mathrm{var}\,\underline{u} \geq \mathrm{var}\,\underline{v}$. Er geldt zelfs (zie stelling 4.2.6) dat $\mathrm{var}\,\underline{u} > \mathrm{var}\,\underline{v}$, tenzij $P(\underline{u}=\underline{v}) = 1$. Dit laatste betekent dat $\underline{u}$ zelf al een funktie van $\underline{t}$ is. Dan geldt natuurlijk (zie p. 57a of stelling 4.3.3)

$$\underline{v} = E(\underline{u}|\underline{t}) = E(g(\underline{t})|\underline{t}) = g(\underline{t}) = \underline{u}.$$

<u>Voorbeeld 7.3.15:</u> Laat $\underline{x}_1$ en $\underline{x}_2$ onafhankelijk zijn en exponentieel verdeeld met $E\,\underline{x} = \theta$. Nu is $\underline{t} = \underline{x}_1+\underline{x}_2$ voldoende voor θ en $\underline{u}=\underline{x}_1$ is een zuivere schatter voor θ, zodat $\underline{v}=E(\underline{u}|\underline{t}) = \underline{t}/2$. We hebben dus $\mathrm{var}\,\underline{v} = \theta^2/2$, terwijl $\mathrm{var}\,\underline{u} = \theta^2$.

<u>Opgave 7.3.16:</u> Bewijs dat $E(\underline{x}_1\;|\;\underline{x}_1+\underline{x}_2 = t) = t/2$, als $\underline{x}_1$ en $\underline{x}_2$ o.o. zijn en exponentieel verdeeld met dezelfde verwachting.

<u>Opmerking 7.3.17:</u> Als $\underline{u} = u(\underline{x}_1,\ldots,\underline{x}_n)$ een funktie is van alle grootheden $\underline{x}_1,\ldots,\underline{x}_n$, dan is de voorwaardelijke kansverdeling van $\underline{u}$ gegeven $\underline{t}=t$ geconcentreerd op het oppervlak $t(x_1,\ldots,x_n) = t$. Dit soort kansverdelingen geeft aanleiding tot wiskundige moeilijkheden (zie opmerking 3.8.7). Voorwaardelijke verwachtingen van de vorm $E(\underline{x}_1^2+\underline{x}_2^2\;|\underline{x}_1+\underline{x}_2 = t)$ kunnen we met de door ons ingevoerde begrippen niet direkt berekenen (de uitkomst is $2t^2/3$ - ga na).

<u>Definitie 7.3.18:</u> Een verzameling verdelingsfuncties $\{F_{\underline{t}}(t;\theta);\theta \in \Omega\}$ heet *volledig*, als voor iedere (continue) functie v de volgende implicatie geldt

$$\{\forall_{\theta\in\Omega}\;[E_\theta v(\underline{t})=0]\} \Rightarrow \{\forall_{\theta\in\Omega}[P_\theta\,(v(\underline{t})=0)=1]\}.$$

Hierbij geeft E_θ en P_θ aan dat de verwachting resp. de kans wordt uitgerekend met behulp van $F(t;\theta)$. Verder wordt aangenomen dat $E_\theta\;v(\underline{t})$ inderdaad bestaat voor $\theta \in \Omega$.

Voorbeeld 7.3.19: De verzameling kansdichtheden $f(t;\theta) = \theta e^{-\theta t}(t > 0)$ met $\theta \in (0,\infty)$ is volledig: als $v:(0,\infty) \to \mathbb{R}$ continu is en als

$$E\, v(\underline{t}) = \theta \int_0^\infty v(t)\, e^{-\theta t} dt = 0$$

voor alle $\theta \in (0,\infty)$, dan is $v(t) = 0$ voor alle $t > 0$ o.g.v. de eenduidigheid van de Laplace-transformatie. Dus is $P(v(\underline{t})=0)=1$.

Voorbeeld 7.3.20: De verzameling verdelingen met

$$P(\underline{k}=k;\mu) = e^{-\mu} \frac{\mu^k}{k!} \quad (k=0,1,\ldots)$$

voor $\mu \in (0,\infty)$ is volledig, immers als

$$E\, v(\underline{k}) = e^{-\mu} \sum_{k=0}^{\infty} v(k) \frac{\mu^k}{k!} = 0$$

voor alle μ in een rechter-omgeving van $\mu = 0$ dan is $v(k) = 0$ voor $k \in \{0,1,\ldots\}$ o.g.v. de eenduidigheidsstelling voor machtreeksen.

Stelling 7.3.21: Als $\underline{t}$ voldoende is voor θ $(\theta \in \Omega)$ en als de verzameling verdelingsfuncties $F_{\underline{t}}(t;\theta)$ volledig is en als bovendien $v(\underline{t})$ een zuivere schatter is voor $h(\theta)$, dan is $v(\underline{t})$ de (unieke) nauwkeurigste zuivere schatter voor $h(\theta)$.

Bewijs: Omdat $F_{\underline{t}}(t;\theta)$ $(\theta \in \Omega)$ volledig is geldt

$$E_\theta\{v_1(\underline{t}) - v_2(\underline{t})\} \equiv 0 \Rightarrow P_\theta(v_1(\underline{t})=v_2(\underline{t})) \equiv 1\ .$$

Er bestaat dus hoogstens één zuivere schatter voor $h(\theta)$, die een functie is van $\underline{t}$. Stel nu dat $\underline{u}$ een zuivere schatter is voor $h(\theta)$, die geen functie is van $\underline{t}$ (en dus ongelijk aan $v(\underline{t})$), dan geldt wegens stelling 7.3.14 dat $\operatorname{var} \underline{u} > \operatorname{var} v(\underline{t})$, omdat nu $E(\underline{u}|\underline{t}) = v(\underline{t})$.

Voorbeeld 7.3.22: Zij $\underline{k}_1,\underline{k}_2,\ldots,\underline{k}_n$ een aselecte steekproef van $\underline{k}$ met $P(\underline{k}=k) = e^{-\mu} \frac{\mu^k}{k!}(k=0,1,\ldots;\mu > 0)$. De grootheid $\underline{t} = \sum_1^n \underline{k}_j$ is voldoende voor μ (vergelijk opmerking 7.3.12). Verder is $\underline{t}/n$ een zuivere schatter voor μ. Tenslotte is de verzameling verdelingsfunkties $F_{\underline{t}}(t;\mu)$ volledig, immers als

$$E\, v(\underline{t}) = e^{-n\mu} \sum_{k=0}^{\infty} v(k) \frac{(n\mu)^k}{k!} = 0$$

voor alle $\mu > 0$, dan is $v(k) = 0$ voor alle $k \in \{0,1,\ldots\}$. Op grond van stelling 7.3.21 is $\underline{t}/n$ de nauwkeurigste zuivere schatter voor μ. Dit is in overeenstemming met voorbeeld 7.1.19 (zie ook voorbeeld 7.3.20).

Opmerking 7.3.23: Het begrip voldoende is altijd gerelateerd aan het beschikbare waarnemingsmateriaal. Zo is in het bovenstaande voorbeeld $\underline{k}_1+\ldots+\underline{k}_n$ voldoende t.a.v. $(\underline{k}_1,\ldots,\underline{k}_n)$, maar niet t.a.v. $(\underline{k}_1,\ldots,\underline{k}_{n+1})$.

8. Toetingstheorie

8.1 Algemene principes

We beginnen met een eenvoudig voorbeeld.

Voorbeeld 8.1.1: We hebben een geldstuk dat bij iedere worp K ("kruis") of M ("munt") oplevert. We vermoeden dat $p := P(M) = \frac{1}{2}$ is en we willen dit vermoeden *toetsen aan de hand van de uitkomsten van 100 worpen.* Het ligt (ook gezien de resultaten bij de schattingstheorie; zie voorbeeld 7.3.2) voor de hand om het aantal keren munt als kriterium te gebruiken. We noemen dit aantal $\underline{k}$. Stel nu dat we als realisatie vinden $k = 40$. We kunnen dan als volgt redeneren. Onder de hypothese $p = \frac{1}{2}$ lijkt de uitkomst $\underline{k} = 40$ tamelijk klein. We berekenen nu de kans op een zo kleine waarde of een nog kleinere: $P(\underline{k} \leq 40;\ p = \frac{1}{2}) \approx \Phi(-2) = 0{,}0228$. We kunnen nu deze kans (te) klein vinden en de hypothese $p = \frac{1}{2}$ verwerpen Hierbij is het principe dat we niet geloven in het optreden van gebeurtenissen met heel kleine kansen. Wat in dit verband "klein" genoemd wordt is een kwestie van smaak - en van traditie. Het is natuurlijk altijd mogelijk dat een dergelijke gebeurtenis toch optreedt (opgetreden is); dit is een onvermijdbaar risico.

We geven nu een algemenere beschrijving van het toetsen van een hypothese.

Definitie 8.1.2: Zij $F_n(x_1,\dots,x_n;\theta)$ de verdelingsfunktie van een stochastische vector $\underline{x}_1,\dots,\underline{x}_n$, waarbij $\theta \in \Omega \subset \mathbb{R}^k$ een onbekende parameter is. Onder een *hypothese* H betreffende θ verstaan we een uitspraak van de vorm $\theta \in \Omega_0$ met $\Omega_0 \subset \Omega$. Als Ω_0 uit één punt bestaat, dan heet H *enkelvoudig*, anders *meervoudig*.

Definitie 8.1.3: Zij $\underline{x}_1,\dots,\underline{x}_n$ een steekproef met verdelingsfunktie $F(x_1,\dots,x_n;\theta)$ $(\theta \in \Omega)$. Een *toets* voor de hypothese $H_0 : \theta \in \Omega_0$, is een voorschrift dat voor iedere realisatie $x_1,\dots,x_n$ aangeeft of H_0 al of niet verworpen wordt. Met andere woorden: een toets is een splitsing van $\mathbb{R}^n$ in twee delen: $\mathbb{R}^n = R \cup A$, met $R \cap A = \emptyset$, waarbij H_0 verworpen wordt als $(x_1,\dots,x_n) \in R$ en H_0 niet verworpen wordt als $(x_1,\dots,x_n) \in A = R^*$. De hypothese H_0 heet ook wel *nulhypothese* en de hypothese H_1, gedefinieerd door $H_1 :\ \theta \in \Omega_1 := \Omega^*_0$ (t.o.v. Ω) heet de *alternatieve hypothese*. Het gebied R heet het *kritieke gebied*, A heet wel het aanvaardingsgebied (R van "reject", A van "accept").

Opmerking 8.1.4: Bij een enkelvoudige hypothese van de vorm $\theta = \theta_0$ worden kansen dikwijls als volgt genoteerd: $P((\underline{x}_1,\ldots,\underline{x}_n) \in R; \theta_0)$, om aan te geven dat de betreffende kans wordt uitgerekend onder de aanname dat $\theta = \theta_0$ is, dus met de verdelingsfunktie $F(x_1,\ldots,x_n;\theta_0)$.

Bij het toetsen van een hypothese kunnen twee fouten optreden: we kunnen H_0 verwerpen terwijl deze hypothese juist is en we kunnen H_0 niet verwerpen, terwijl de hypothese onjuist is. Deze fouten worden resp. *fout van de eerste soort* en *fout van de tweede soort* genoemd. Toetsen worden i.h.a. zó geconstrueerd, dat de kans op een fout van de eerste soort niet boven een gegeven grens α, de *onbetrouwbaarheidsdrempel*, komt. Veel gebruikte waarden van α zijn 0,05 en 0,01. Men probeert dan onder deze restrictie de kans op een fout van de tweede soort zo klein mogelijk te maken, d.w.z. de kans op verwerpen van H_0 voor waarden van θ in Ω_1 zo groot mogelijk te maken.

Definitie 8.1.5: Zij R het kritieke gebied voor een toets van de nulhypothese H_0: $\theta \in \Omega_0$ tegen het alternatief H_1: $\theta \in \Omega_1$. Dan heet

$$\alpha_0 := \sup_{\theta\in\Omega_0} P((\underline{x}_1,\ldots,\underline{x}_n) \in R;\theta) \qquad \text{de } \textit{onbetrouwbaarheid}$$

van de toets. Voor $\theta \in \Omega_1$ heet de funktie β, gedefinieerd door

$$\beta(\theta) := P((\underline{x}_1,\ldots,\underline{x}_n) \in R;\theta), \quad \text{het } \textit{onderscheidingsvermogen}$$

van de toets. (We eisen dat $\alpha_0 \le \alpha$, de onbetrouwbaarheidsdrempel, terwijl we zo mogelijk $\alpha_0 = \alpha$ kiezen; dit kan i.h.a. als $(\underline{x}_1,\ldots,\underline{x}_n)$ een absoluut continue verdeling heeft; bij diskrete verdelingen kiezen we α_0 zo groot mogelijk, maar $\le \alpha$).

Opmerking 8.1.6: In de meeste gevallen wordt het kritieke gebied R bepaald door de waarden van een statistische grootheid $t = t(x_1,\ldots,x_n)$, waarbij $t(\underline{x}_1,\ldots,\underline{x}_n)$ bijvoorbeeld een schatter is voor θ (vergelijk voorbeeld 8.1.1). De functie t heet in dit verband een *toetsingsgrootheid*. Het kritieke gebied R heeft dan de gedaante

$$(8.1.1) \qquad R = R(\alpha,n) = \{ (x_1,\ldots,x_n) \mid t(x_1,\ldots,x_n) \in K\}$$

met $K \subset \mathbb{R}$. De verzameling K (dikwijls het complement van een interval) wordt nu ook wel kritiek gebied (voor t) genoemd.

We illustreren nu de boven ingevoerde begrippen aan de hand van een voorbeeld.

Voorbeeld 8.1.7: Zij $x_1, x_2, \ldots, x_n$ de realisatie van een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\tfrac{1}{2}(x-\theta)^2\} \qquad (\theta \in \mathbb{R}).$$

We toetsen de enkelvoudige hypothese H_0: $\theta = \theta_0$ tegen het alternatief H_1: $\theta \neq \theta_0$ (d.w.z. $\Omega_0 = \{\theta_0\}$ en $\Omega_1 = \mathbb{R}\setminus\{\theta_0\}$). Het ligt voor de hand om $\frac{1}{n}(\underline{x}_1+\ldots+\underline{x}_n)$ (of een functie daarvan) als toetsingsgrootheid te gebruiken, d.w.z. om een kritiek gebied te gebruiken van de vorm

$$R = \{(x_1,\ldots,x_n) \mid (x_1+\ldots+x_n)/n \in K\}$$

met $P((\underline{x}_1,\ldots,\underline{x}_n) \in R;\, \theta_0) = \alpha = 0{,}05$. Door deze eisen ligt de toets nog allerminst vast. We schrijven $t = t(x_1,\ldots,x_n) = (x_1+\ldots+x_n)/n$ en beschouwen drie mogelijkheden voor K (en dus voor R):

(i) $K_1 = \{t \mid |t-\theta_0| \geq t_1\}$ met t_1 zó dat

$$P((\underline{x}_1,\ldots,\underline{x}_n) \in R_1;\theta_0) = P(|\underline{t}-\theta_0| \geq t_1;\theta_0) = \alpha = 0{,}05.$$

Hieruit volgt (ga na), dat $t_1 = 1{,}96/\sqrt{n}$, zodat

$$R_1 = \{(x_1,\ldots,x_n) \mid \left|\frac{x_1+\ldots+x_n}{n} - \theta_0\right| \geq 1{,}96/\sqrt{n}\}.$$

(ii) $K_2 = \{t \mid t \geq \theta_0 + t_2\}$ met t_2 zó dat

$$P((\underline{x}_1,\ldots,\underline{x}_n) \in R_2;\theta_0) = P(\underline{t} \geq \theta_0 + t_2;\theta_0) = 0{,}05.$$

We vinden dan $t_2 = 1{,}645/\sqrt{n}$, d.w.z.

$$R_2 = \{(x_1,\ldots,x_n) \mid \frac{x_1+\ldots+x_n}{n} \geq \theta_0 + 1{,}645/\sqrt{n}\}.$$

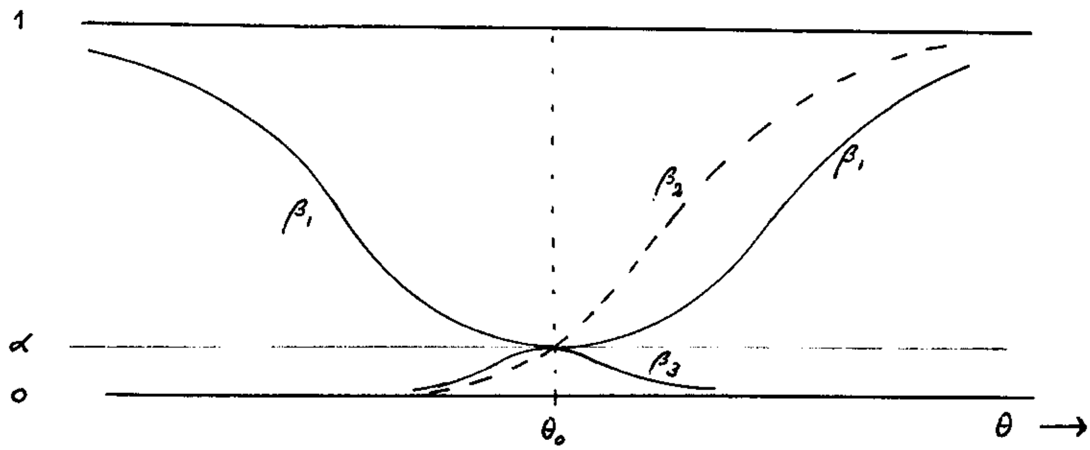
(iii) $K_2 = \{t \mid |t-\theta_0| \leq t_3\}$ met t_3 zó dat

$$P((\underline{x}_1,\ldots,\underline{x}_n) \in R_3;\theta_0) = P(|\underline{t}-\theta_0| \leq t_3;\theta_0) = 0{,}05.$$

We vinden $t_3 = 0{,}063/\sqrt{n}$, d.w.z.

$$R_3 = \{(x_1,\ldots,x_n) \mid |\frac{x_1+\ldots+x_n}{n} - \theta_0| \leq 0{,}063/\sqrt{n}\}.$$

De kans op een fout van de eerste soort is voor elk van de drie toetsen 0,05. Het verschil zit in het onderscheidingsvermogen (zie definitie 8.1.5). In het onderstaande plaatje is het onderscheidingsvermogen voor elk van de toetsen geschetst (en aangegeven met β_1, β_2 en β_3) als funktie van θ.



Het is duidelijk dat de derde toets slecht is: de kans op verwerpen van H_0 is hoogstens α (voor $\theta=\theta_0$) en neemt bovendien af naar mate θ verder van θ_0 verwijderd is. De eerste toets lijkt het meest redelijk: de kans op verwerpen van H_0 neemt (snel) toe als θ verder van θ_0 ligt, zowel links als rechts. Bij de tweede toets is de kans op verwerpen van H_0 klein bij waarden van θ kleiner dan θ_0, maar $\beta_2(\theta) > \beta_1(\theta)$ als $\theta > \theta_0$. In sommige gevallen verdient de toets met R_2 daarom de voorkeur boven die met R_1. We komen hier later op terug, als we zgn. eenzijdige toetsen bespreken.

Uit dit voorbeeld blijkt dat we pas over de "beste" toets kunnen praten, in de zin van de toets met het grootste onderscheidingsvermogen, als we van te voren hebben afgesproken, voor welke θ het onderscheidingsvermogen groot moet zijn. Gezien definitie 8.1.5 komt dit neer op het kiezen van de nulhypothese en de alternatieve hypothese. Ook hierdoor ligt een toets echter niet vast.

Opgave: Geef een formule voor het onderscheidingsvermogen van de toets in voorbeeld 8.1.7 (i).

Definitie 8.1.8: Een toets T^* voor de hypothese H_0: $\theta \in \Omega_0$ tegen het alternatief H_1: $\theta \in \Omega_1$ met onbetrouwbaarheid α_0 heet *uniform meest onderscheidend*, als

$$\beta^*(\theta) \geq \beta(\theta) \qquad (\theta \in \Omega_1)$$

voor iedere toets T met onbetrouwbaarheid hoogstens α_0 (hierbij hebben we het onderscheidingsvermogen van T^* en T aangegeven met resp. β^* en β).

Definitie 8.1.9: Een toets voor de hypothese H_0: $\theta \in \Omega_0$ tegen H_1: $\theta \in \Omega_1$ heet *zuiver* (met onbetrouwbaarheid α_0) als

$$\alpha_0 = \sup_{\theta\in\Omega_0} P((\underline{x}_1,\ldots,\underline{x}_n) \in R;\theta) \leq \inf_{\theta\in\Omega_1} P((\underline{x}_1,\ldots,\underline{x}_n) \in R;\theta).$$

Definitie 8.1.10: Een toets heet(uniform)meest onderscheidend zuiver, als hij onder de zuivere toetsen (uniform) meest onderscheidend is, d.w.z. (uniform) het grootste onderscheidingsvermogen heeft.

Opmerking 8.1.11: De toets onder (i) in voorbeeld 8.1.7 is uniform meest onderscheidend zuiver voor de nulhypothese $\theta = \theta_0$ tegen het alternatief $\theta \neq \theta_0$ (geen bewijs).

Voor enkelvoudige hypothesen hebben we

Stelling 8.1.12 (Lemma van Neyman en Pearson): Zij $(x_1,\ldots,x_n)$ de realisatie van een steekproef $(\underline{x}_1,\ldots,\underline{x}_n)$ met aannemelijkheidsfunctie $L(x_1,\ldots,x_n;\theta)$ (zie definitie 7.2.1) met $\theta \in \{\theta_0,\theta_1\}$. De meest onderscheidende toets voor de hypothese H_0: $\theta = \theta_0$ tegen het alternatief H_1: $\theta = \theta_1$ wordt verkregen door een kritiek gebied R te nemen van de vorm

$$(8.1.2) \qquad R = \{(x_1,\ldots,x_n) \mid \frac{L(x_1,\ldots,x_n;\theta_0)}{L(x_1,\ldots,x_n;\theta_1)} \leq k\}.$$

(Voor een toets met onbetrouwbaarheidsdrempel α wordt k zó gekozen dat $\alpha_0 \leq \alpha$ en zo groot mogelijk)

Bewijs: We beschouwen het geval, waar $(\underline{x}_1,\ldots,\underline{x}_n)$ een absoluut continue verdeling heeft met dichtheid (in vectornotatie) $L(x;\theta)$. Zij R' het kritieke gebied van een willekeurige andere toets (met dezelfde onbetrouwbaarheid). We schrijven

$$R = U + D;\ R' = V + D$$

met $D = RR'$, $U = R(R')^*$ en $V = R^*R'$. Dan geldt $\alpha = \int_R L(x;\theta_0)dx$ en $\alpha = \int_{R'} L(x;\theta_0)dx$, dus

$$(8.1.3) \qquad \int_U L(x;\theta_0)dx = \int_V L(x;\theta_0)dx.$$

Omdat op R en dus op U geldt dat $L(x;\theta_0) \le kL(x;\theta_1)$ en op V dat $L(x;\theta_0) > kL(x;\theta_1)$, hebben we

$$(8.1.4) \qquad \int_U L(x;\theta_0)dx \le k \int_U L(x;\theta_1)dx,$$

$$(8.1.5) \qquad \int_V L(x;\theta_0)dx \ge k \int_V L(x;\theta_1)dx.$$

Uit (8.1.3) t.m. (8.1.5) volgt dat $\int_U L(x;\theta_1)dx \ge \int_V L(x;\theta_1)dx$ en dus dat

$$\int_R L(x;\theta_1)dx \ge \int_{R'} L(x;\theta_1)dx,$$

d.w.z. $\beta(\theta_1) \ge \beta'(\theta_1)$, dus de toets met R is minstens even onderscheidend als die met R'.

<u>Opgave 8.1.13:</u> Bewijs stelling 8.1.12 voor diskrete $(\underline{x}_1,\dots,\underline{x}_n)$.

<u>Opmerking 8.1.14</u>: Het quotient $L(x_1,\dots,x_n;\theta_0)/L(x_1,\dots,x_n;\theta_1)$ wordt *aannemelijkheidsquotient* genoemd (engels: likelihood ratio); de bijbehorende toets en toetsen van een verwant type heten *likelihood-ratio-toetsen*.

<u>Opmerking 8.1.15:</u> Punten met $L(x_1,\dots,x_n;\theta_0) = 0$ komen in R terecht; punten met $L(x_1,\dots,x_n;\theta_1) = 0$ (en $L(x_1,\dots,x_n;\theta_0) \ne 0$) in R^*.

Opgave 8.1.16: De toets onder (ii) in voorbeeld 8.1.7 is uniform meest onderscheidend (en zuiver) voor H_0: $\theta = \theta_0$ tegen H_1: $\theta > \theta_0$. Bewijs dit door toepassing van stelling 8.1.12 met (telkens) één vaste $\theta_1 > \theta_0$.

8.2 Constructie van toetsen

In veel gevallen worden toetsen "bedacht" door uit te gaan van een redelijke toetsingsgrootheid, zoals in de voorbeelden 8.1.1 en 8.1.7.

We geven nu een aantal voorbeelden van toetsen die veel worden gebruikt en die op zo'n "redelijke" toetsingsgrootheid berusten.

Voorbeeld 8.2.1: Zij $x_1,\ldots,x_n$ een realisatie van een aselecte steekproef uit een normale verdeling met verwachting μ en variantie σ^2. We willen toetsen H_0: $\mu = \mu_0$ tegen H_1: $\mu \neq \mu_0$. We onderscheiden twee gevallen:

A: σ^2 bekend. Dan zijn we terug bij voorbeeld 8.1.7: we gebruiken als toetsingsgrootheid $\underline{x}_1+\ldots+\underline{x}_n$ of liever de grootheid

$$(8.2.1) \qquad \underline{u} = \frac{\underline{x}_1+\ldots+\underline{x}_n - n\mu_0}{\sigma\sqrt{n}} = \frac{\bar{\underline{x}} - \mu_0}{\sigma}\sqrt{n} ,$$

die onder de hypothese H_0 normaal verdeeld is met $E\underline{u} = 0$ en var $\underline{u} = 1$. Bij voorgeschreven α verwerpen we nu H_0, als $u = \frac{\bar{x}-\mu_0}{\sigma}\sqrt{n}$ zó is dat

$$(8.2.2) \qquad P(\underline{u} \geq u;\ \mu_0) \leq \alpha/2$$

of zó dat

$$(8.2.3) \qquad P(\underline{u} \leq u;\ \mu_0) \leq \alpha/2,$$

d.w.z. als $u \geq u_{\alpha/2}$ of $u \leq -u_{\alpha/2}$, waar u_γ gedefinieerd is door

$$(8.2.4) \qquad 1 - \Phi(u_\gamma) = \gamma.$$

B: σ^2 onbekend. Dan vervangen we σ^2 in (8.2.1) door een goede schatter: $\underline{s}^2 = (n-1)^{-1}\sum_1^n(\underline{x}_j-\bar{\underline{x}})^2$. De toetsingsgrootheid wordt dan

$$(8.2.5) \qquad \underline{t} = \frac{\bar{\underline{x}} - \mu_0}{\underline{s}}\sqrt{n}.$$

Nu zijn (zie Appendix 4) $\underline{\bar{x}}$ en $\underline{s}^2$ (en dus ook $\underline{\bar{x}}$ en $\underline{s}$) onafhankelijk, zodat $\underline{t}$ in (8.2.5) onder H_0 een student-verdeling heeft met n-1 vrijheidsgraden (vergelijk Appendix 2, blz. 80 en opgave 8.2.3). We verwerpen nu H_0 als t zō is dat

$$P(\underline{t} \geq t;\ \mu_0) \leq \alpha/2$$

of zō dat

$$P(\underline{t} \leq t;\ \mu_0) \leq \alpha/2,$$

d.w.z. als $t \geq t_{\alpha/2}$ of $t \leq -t_{\alpha/2}$ met

$$(8.2.6) \qquad \int_{t_\gamma}^{\infty} f_{\underline{t}_{n-1}}(x)\,dx = \gamma.$$

Opmerking 8.2.2: De bovengenoemde toetsen zijn *tweezijdig*: we verwerpen H_0 zowel bij grote waarden van u en t als bij kleine. We kunnen ook *eenzijdig* toetsen. We toetsen dan H_0: $\mu = \mu_0$ tegen H_1: $\mu > \mu_0$ of ook wel H_0: $\mu \leq \mu_0$ tegen $\mu > \mu_0$. Bij deze *rechts-eenzijdige* toets verwerpen we H_0 alleen bij grote waarden van u of t. In plaats van beide *overschrijdingskansen* (8.2.2) en (8.2.3) beschouwen we alleen de *rechter overschrijdingskans* $P(\underline{u} \geq u;\mu_0)$ en we verwerpen als $u \geq u_\alpha$, d.w.z. als

$$P(\underline{u} \geq u;\mu_0) \leq \alpha.$$

Analoog kunnen we *links-eenzijdig* toetsen. Eenzijdig toetsen doen we in die gevallen waar we alleen willen verwerpen (en dus ingrijpen of protesteren) bij een steekproefresultaat dat naar één bepaalde kant t.o.v. de nulhypothese afwijkt, bijvoorbeeld als we een partij goederen niet willen kopen als een steekproef te veel (vergeleken met een opgegeven percentage) defecten bevat.

Opgave 8.2.3: Ga na dat $\sum_1^n (\underline{x}_j - \underline{\bar{x}})^2/\sigma^2$ een χ^2_{n-1}-verdeling heeft, als $\underline{x}_1,\ldots,\underline{x}_n$ o.o. zijn en $N(\mu,\sigma^2)$-verdeeld.

Voorbeeld 8.2.4: Zij $x_1,\dots,x_n$ een realisatie van een aselecte steekproef van $\underline{x}$ met een $N(\mu,\sigma^2)$-verdeling. We willen toetsen H_0: $\sigma^2 = \sigma_0^2$ tegen H_1: $\sigma^2 \neq \sigma_0^2$.

A. Als μ bekend is, kunnen we als toetingsgrootheid nemen

$$(8.2.7) \qquad \underline{\chi}_n^2 = \sum_1^n (\underline{x}_j - \mu)^2/\sigma_0^2.$$

Het is duidelijk dat $\underline{\chi}_n^2$ onder H_0 een χ_n^2-verdeling heeft. We verwerpen H_0 dan als χ^2 zo is dat

$$P(\underline{\chi}_n^2 \geq \chi^2;\sigma_0^2) \leq \alpha/2$$

of zó dat

$$P(\underline{\chi}_n^2 \leq \chi^2;\ \sigma_0^2) \leq \alpha/2,$$

d.w.z. als $\chi^2 \geq \chi^2_{\alpha/2}$ of $\chi^2 \leq \chi^2_{1-\alpha/2}$, waarbij χ^2_γ gedefinieerd is in analogie met (8.2.4) en (8.2.6).

B. Als μ onbekend is, kunnen we dezelfde toets gebruiken, nu met

$$(8.2.7') \qquad \underline{\chi}_{n-1}^2 = \sum_1^n (\underline{x}_j - \underline{\bar{x}})^2/\sigma_0^2.$$

Deze grootheid is onder H_0 χ^2_{n-1}-verdeeld (zie opgave 8.2.3).

De volgende voorbeelden hebben betrekking op twee steekproeven.

Voorbeeld 8.2.5: $x_1,\dots,x_m$ en $y_1,\dots,y_n$ zijn realisaties van twee onafhankelijke aselecte steekproeven van $\underline{x}$ en $\underline{y}$ die resp. $N(\mu,\sigma_1^2)$- en $N(\lambda,\sigma_2^2)$-verdeeld zijn. We toetsen H_0: $\lambda = \mu$ tegen H_1 : $\lambda \neq \mu$.

A: σ_1^2 en σ_2^2 bekend. Nu is $(\underline{\bar{x}}-\underline{\bar{y}})\left(\frac{mn}{n\sigma_1^2+m\sigma_2^2}\right)^{\frac{1}{2}}$ $N(0,1)$-verdeeld (onder H_0).

Als $\sigma_1 = \sigma_2 = \sigma$ wordt dit

$$\frac{\underline{\bar{x}}-\underline{\bar{y}}}{\sigma}\left(\frac{mn}{m+n}\right)^{\frac{1}{2}}.$$

B: $\sigma_1^2 = \sigma_2^2 = \sigma^2$ onbekend. We kunnen nu de, (onder H_0) t_{m+n-2}-verdeelde, toetsingsgrootheid

(8.2.8) $\quad \underline{t} = \dfrac{\bar{\underline{x}}-\bar{\underline{y}}}{\underline{s}} \left(\dfrac{mn}{m+n}\right)^{\frac{1}{2}}$

gebruiken, waarbij $\underline{s}^2 = \frac{1}{m+n-2}\{\sum_1^m(\underline{x}_j-\bar{\underline{x}})^2 + \sum_1^n(\underline{y}_k-\bar{\underline{y}})^2\}$.

Opmerking: Als $m=n$ en als de paren $(\underline{x}_j,\underline{y}_j)$ o.o. zijn en simultaan normaal verdeeld, dan passen we de toets uit voorbeeld 8.2.1 toe op de *verschillen* $\underline{x}_1-\underline{y}_1,\ldots,\underline{x}_n-\underline{y}_n$.

Voorbeeld 8.2.6: $x_1,\ldots,x_m$ en $y_1,\ldots,y_n$ zijn aselecte steekproeven van $\underline{x}$ en $\underline{y}$, waarbij $\underline{x}$ en $\underline{y}$ o.o. zijn en resp. $N(\mu_1,\sigma_1^2)$-en $N(\mu_2,\sigma_2^2)$-verdeeld We toetsen H_0: $\sigma_1^2 = \sigma_2^2$ tegen H_1: $\sigma_1^2 \neq \sigma_2^2$.

A. μ_1 en μ_2 bekend. Als toetsingsgrootheid ligt voor de hand

(8.2.9) $\quad \underline{F} = \dfrac{\underline{s}^2_{0,1}}{\underline{s}^2_{0,2}}$,

met $\underline{s}^2_{0,1} = \frac{1}{m}\sum_1^m(\underline{x}_j-\mu_1)^2$ en $\underline{s}^2_{0,2} = \frac{1}{n}\sum_1^n(\underline{y}_k-\mu_2)^2$. Onder H_0 heeft F een $F_{m,n}$-verdeling (zie appendix 2, blz. 80).

B. μ_1 en μ_2 onbekend. We vervangen nu μ_1 en μ_2 door hun schatters en vinden

(8.2.10) $\quad \underline{F} = \dfrac{\underline{s}_1^2}{\underline{s}_2^2}$,

met $\underline{s}_1^2 = \frac{1}{m-1}\sum_1^m(\underline{x}_j-\bar{\underline{x}})^2$ en $\underline{s}_2^2 = \frac{1}{n-1}\sum_1^n(\underline{y}_j-\bar{\underline{y}})^2$. De grootheid $\underline{F}$ in (8.2.10) heeft een $F_{m-1,n-1}$-verdeling (onder H_0).

Opmerking 8.2.7: De toetsingsgrootheden in de bovenstaande voorbeelden kunnen natuurlijk zowel voor éénzijdige als voor tweezijdige toetsen gebruikt worden.

In de volgende definitie wordt een procedure gegeven om toetsen te construeren. Alle hierboven behandelde toetsingsgrootheden kunnen met deze procedure gevonden worden.

<u>Definitie 8.2.8:</u> Zij $x_1, x_2, \ldots, x_n$ een steekproef uit een verdeling met likelihood functie $L(x_1, \ldots, x_n; \theta)$. De (gegeneraliseerde) *likelihood-ratio-toets* met onbetrouwbaarheidsdrempel α voor de hypothese H_0: $\theta \in \Omega_0$ tegen H_1: $\theta \in \Omega_1 = \Omega_0^*$ is de toets gebaseerd op het volgende kritieke gebied R. We definiëren eerst het *aannemelijkheidsquotient* (*likelihoodratio*) λ door

$$(8.2.11) \quad \lambda(x_1, \ldots, x_n) = \frac{\sup\limits_{\theta \in \Omega_0} L(x_1, \ldots, x_n; \theta)}{\sup\limits_{\theta \in \Omega} L(x_1, \ldots, x_n; \theta)}.$$

Vervolgens definiëren we het kritieke gebied R door

$$(8.2.12) \quad R = \{(x_1, \ldots, x_n) \mid \lambda(x_1, \ldots, x_n) \le k(\alpha)\},$$

waarbij $k(\alpha)$ vast ligt door de eis dat $\sup\limits_{\theta \in \Omega_0} P((\underline{x}_1, \ldots, \underline{x}_n) \in R; \theta)$ zo groot mogelijk is, maar $\le \alpha$.

Het is duidelijk dat $0 \le \lambda \le 1$. Het idee achter deze procedure, die verwant is aan het maximum likelihood principe, is dat λ i.h.a. groot zal zijn (dicht bij 1 of gelijk aan 1) als de werkelijke waarde van θ in Ω_0 ligt, d.w.z. als H_0 waar is.

<u>Opmerking 8.2.9:</u> Een iets algemenere toets ontstaat als we λ vervangen door λ_1 gedefineerd door

$$\lambda_1(x_1, \ldots, x_n) = \frac{\sup\limits_{\theta \in \Omega_0} L(x_1, \ldots, x_n; \theta)}{\sup\limits_{\theta \in \Omega_1} L(x_1, \ldots, x_n; \theta)}.$$

Deze toets is een echte generalisatie van de toets in stelling 8.1.12. Hij gaat hier in over als H_0 en H_1 enkelvoudig zijn, d.w.z. als $\Omega_0 = \{\theta_0\}$ en $\Omega_1 = \{\theta_1\}$.

<u>Voorbeeld 8.2.10:</u> Zij $x_1, \ldots, x_n$ de realisatie van een aselecte steekproef van $\underline{x}$ met

$$f_{\underline{x}}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}.$$

We bepalen de likelihood-ratio-toets voor de hypothese H_0: $\mu = \mu_0$ tegen het alternatief H_1: $\mu \neq \mu_0$ met onbetrouwbaarheid α. We bepalen eerst $\lambda(x_1,\ldots,x_n)$; we hebben $\Omega = \{(\mu,\sigma^2)\,|\,\mu\in\mathbb{R},\sigma^2>0\}$, $\Omega_0=\{(\mu,\sigma^2)\,|\,\mu=\mu_0,\sigma^2>0\}$ en dus (zie opgave 7.3.10 en voorbeeld 7.3.11), met $\theta = (\mu,\sigma^2)$,

$$\sup_{\theta\in\Omega_0} L(x_1,\ldots,x_n;\theta) = \max_{\sigma^2} \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma^2}\sum_1^n (x_j-\mu_0)^2\} =$$

$$= \left(\frac{1}{2\pi\hat{\sigma}_0^2}\right)^{n/2} \exp\{-\frac{1}{\hat{\sigma}_0^2}\sum_1^n (x_j-\mu_0)^2\},$$

met $\hat{\sigma}_0^2 = \frac{1}{n}\sum_1^n (x_j-\mu_0)^2$, terwijl

$$\sup_{\theta\in\Omega} L(x_1,\ldots,x_n;\theta) = \max_{\mu,\sigma^2} \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma^2}\sum_1^n (x_j-\mu)^2\} =$$

$$= \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{n/2} \exp\{-\frac{1}{2\hat{\sigma}^2}\sum_1^n (x_j-\bar{x})^2\}$$

met $\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_j-\bar{x})^2$. Voor λ vinden we dus

$$\lambda(x_1,\ldots,x_n) = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2} = \left\{\frac{1}{1+(\hat{\sigma}_0^2-\hat{\sigma}^2)/\hat{\sigma}^2}\right\}^{n/2} = \frac{1}{\left\{1+\frac{(\bar{x}-\mu_0)^2}{\hat{\sigma}^2}\right\}^{n/2}}.$$

Nu is λ dus gelijk aan (zie (8.2.5))

$$\lambda(x_1,\ldots,x_n) = \left\{\frac{1}{1+t^2/(n-1)}\right\}^{n/2},$$

zodat (ga na) het kritieke gebied (8.2.11) de volgende vorm krijgt:

$$R = \{(x_1,\ldots,x_n)\ |\ t^2 \geq c(\alpha)\},$$

met $t = (\bar{x}-\mu_0)\frac{\sqrt{n}}{s}$, zodat $\underline{t} = (\bar{\underline{x}}-\mu_0)\frac{\sqrt{n}}{\underline{s}}$, een t_{n-1}-verdeling heeft (zie App. 4). Het kritieke gebied K voor t wordt dus

$$K = \{t|\ \ |t| \geq t_{\alpha/2}\},$$

d.w.z. we vinden de toets van voorbeeld 8.2.1 B terug.

Voorbeeld 8.2.11: In de situatie van voorbeeld 8.2.10 toetsen we nu de nulhypothese H_0: $\sigma^2 = \sigma_0^2$ tegen H_1: $\sigma^2 \neq \sigma_0^2$. We vinden nu met $\Omega_0 = \{(\mu,\sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\}$

$$\sup_{\theta\in\Omega_0} L(x_1,\ldots,x_n;\theta) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma_0^2}\sum_1^n (x_j-\bar{x})^2\}$$

$$\sup_{\theta\in\Omega} L(x_1,\ldots,x_n;\theta) = \left(\frac{1}{2\pi\hat{\sigma}^2}\right)^{n/2} \exp\{-\frac{1}{2\hat{\sigma}^2}\sum_1^n (x_j-\bar{x})^2\},$$

met weer $\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (x_j-\bar{x})^2$. We vinden dus voor λ de uitdrukking (ga na)

$$\lambda(x_1,\ldots,x_n) = e^{n/2}\left\{\frac{\hat{\sigma}^2}{\sigma_0^2}\, e^{-\hat{\sigma}^2/\sigma_0^2}\right\}^{n/2}.$$

Het kritieke gebied wordt nu van de vorm

$$(8.2.13)\quad R = \{(x_1,\ldots,x_n) \mid \frac{\hat{\sigma}^2}{\sigma_0^2}\, e^{-\hat{\sigma}^2/\sigma_0^2} \le c(\alpha)\},$$

met $c(\alpha)$ zó dat $P((\underline{x}_1,\ldots,\underline{x}_n) \in R;\ \sigma_0^2) = \alpha$. Als toetsingsgrootheid kunnen we dus gebruiken $\hat{\underline{\sigma}}^2/\sigma_0^2$ of

$$\underline{\chi}^2 = \frac{n\hat{\underline{\sigma}}^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_1^n (\underline{x}_j-\bar{\underline{x}})^2.$$

Het kritieke gebied K voor χ^2 behorende bij deze likelihood-ratio-toets wordt nu gegeven door

$$(8.2.14)\quad K = \{\chi^2 \mid \chi^2 \le \chi_1^2 \text{ of } \chi^2 \ge \chi_2^2\},$$

waarbij χ_1^2 en χ_2^2 voldoen aan $\chi^2 e^{-\chi^2/n} = nc(\alpha)$, terwijl $c(\alpha)$ zó is dat

$$P(\chi_1^2 < \underline{\chi}^2 < \chi_2^2) = 1-\alpha.$$

De grootheid $\underline{\chi}^2$ is dezelfde als die in voorbeeld 8.2.4 B en heeft dus een χ_{n-1}^2-verdeling. Omdat de grenzen χ_1^2 en χ_2^2 zoals boven gedefinieerd niet zo makkelijk te berekenen zijn, gebruikt men in de practijk toch vaak de toets uit voorbeeld 8.2.4 B, d.w.z. men gebruikt een kritiek gebied van de vorm (8.2.14), maar met χ_1^2 en χ_2^2 zō dat

$$P(\underline{\chi}^2 \le \chi_1^2) = P(\underline{\chi}^2 \ge \chi_2^2) = \alpha/2.$$

Dit is dan geen echte likelihood-ratio-toets.

Opmerking 8.2.12: In voorbeeld 8.2.10 hangt de kansverdeling van $\underline{t}$ niet van σ^2 af; in voorbeeld 8.2.11 hangt de kansverdeling van $\underline{\chi}^2$ niet van μ af (vergelijk definitie 8.2.8).

Opgave 8.2.13: $x_1,\ldots,x_{100}$ is de realisatie van een aselecte steekproef uit een normale verdeling. Men vindt $\bar{x} = 4{,}27$ en $s^2 = 0{,}975$.

a. Toets H_0: $\mu = 4{,}00$ tegen H_1: $\mu \ne 4{,}00$ ($\alpha = 0{,}01$)

b. Toets H_0: $\sigma^2 = 0{,}810$ tegen H_1: $\sigma^2 \ne 0{,}810$ ($\alpha = 0{,}01$) en bereken het onderscheidingsvermogen van deze toets voor $\sigma = 1{,}00$.

Opgave 8.2.14: Laat zien dat de toetsen in voorbeeld 8.2.1 A en in voorbeeld 8.2.5 A en B likelihood-ratio-toetsen zijn.

Voorbeeld 8.2.15: Zij $x_1,\ldots,x_n$ de realisatie van een aselecte steekproef van $\underline{x}$ met

$$P(\underline{x}=0) = 1-p, \quad P(\underline{x}=1) = p.$$

We construeren de likelihood-ratio-toets voor de nulhypothese H_0: $p=p_0$ tegen H_1: $p\ne p_0$. We vinden gemakkelijk (ga na) dat

$$\lambda(x_1,\ldots,x_n) = \frac{p_0^k(1-p_0)^{n-k}}{(\frac{k}{n})^k(1-\frac{k}{n})^{n-k}}, \tag{8.2.15}$$

met $k = x_1+\ldots+x_n$. Het kritieke gebied is dus

$$R = \{(x_1,\ldots,x_n) \mid (\frac{np_0}{k})^k \{\frac{n(1-p_0)}{n-k}\}^{n-k} \le c(\alpha)\}.$$

Nu is het rechterlid van (8.2.15) stijgend voor $k < np_0$ en dalend voor $k > np_0$, zodat het kritieke gebied voor k de volgende vorm heeft:

$$K = \{k \mid k \le k_1 \text{ of } k \ge k_2\}. \tag{8.2.16}$$

Ook hier is het weer lastig om k_1 en k_2 te vinden. In de practijk kiest men k_1 en k_2 zo groot, resp. zo klein, mogelijk terwijl

(8.2.17) $P(\underline{k} \le k_1;p_0) \le \alpha/2$, $P(\underline{k} \ge k_2;p_0) \le \alpha/2$.

Opgave 8.2.16: Men gooit 200 keer met een dobbelsteen en vindt 40 "zessen". Toets de hypothese P("zes") = 1/6 tegen het alternatief P("zes") $\neq$ 1/6 (zie opmerking 5.4.7). Kies α = 0,05.

Opgave 8.2.17: Laat zien dat de eenzijdige toets met kritiek gebied van de vorm

$$K = \{k \mid k \ge k_1\}$$

uniform meest onderscheidend is voor de hypothese H_0: $p=p_0$ tegen H_1: $p > p_0$ (vergelijk opgave 8.1.16).

Voorbeeld 8.2.18: Twee analisten bepalen telkens beiden de concentratie van stof Z in n monsters. Het resultaat bestaat uit een realisatie $(x_1,y_1),\ldots,(x_n,y_n)$ van een aselecte steekproef van $(\underline{x},\underline{y})$, waarbij $\underline{x}$ en $\underline{y}$ niet noodzakelijk onafhankelijk zijn. We willen toetsen of beide analisten even grote concentraties vinden. We toetsen de hypothese H_0: $P(\underline{x} > \underline{y}) = P(\underline{x} < \underline{y}) = \frac{1}{2}$ tegen het alternatief H_1: $P(\underline{x}>\underline{y}) \neq P(\underline{x}<\underline{y})$ (we nemen gemakshalve aan dat $(\underline{x},\underline{y})$ een absoluut continue verdeling heeft, zodat $P(\underline{x}=\underline{y})=0$).
Als toetsingsgrootheid gebruiken we $\underline{k}$, het aantal positieve verschillen onder $\underline{x}_1-\underline{y}_1$, $\underline{x}_2-\underline{y}_2,\ldots,\underline{x}_n-\underline{y}_n$. Onder H_0 heeft $\underline{k}$ dus een binomiale verdeling met parameters n en $\frac{1}{2}$. Men noemt deze toets, waarbij alleen op het teken van $\underline{x}-\underline{y}$ wordt gelet, *tekentoets*.
Men gebruikt hierbij voor de tweezijdige toets een kritiek gebied bepaald door (8.2.16) en (8.2.17), weer met de conditie dat k_1 zo groot mogelijk is en k_2 zo klein mogelijk; we kunnen natuurlijk ook eenzijdig toetsen.

Opgave 8.2.19: Hoe vaak moeten we met een munt gooien om met een betrouwbaarheid van 0,01 te kunnen beslissen dat hij onzuiver is (d.w.z. niet met kansen $\frac{1}{2}$ kruis en munt geeft)?

8.3. Betrouwbaarheidsintervallen

Betrouwbaarheidsintervallen en ,iets algemener, betrouwbaarheidsgebieden kunnen worden opgevat als een generalisatie van schatters, in dit verband ook wel puntschatters genoemd. Een schatter (of een schatting) geeft één enkele aannemelijke waarde voor een onbekende parameter, een betrouwbaarheidsgebied is een *verzameling* van parameterwaarden, die met een zekere aannemelijkheid, betrouwbaarheid de werkelijke (onbekende) waarde bevat. Het begrip aannemelijkheid kwamen we tegen bij de schattingstheorie,het begrip onbetrouwbaarheid bij de toetsingstheorie (zie definities 7.2.1 en 3.2.5). We geven eerst een formele definitie.

Definitie 8.3.1: Zij $\underline{x}_1,\ldots,\underline{x}_n$ een steekproef met aannemelijkheidsfunctie $L(x_1,\ldots,x_n;\theta)$ $(\theta\in\Omega)$. Een *betrouwbaarheidsgebied* voor de (mogelijk meerdimensionale) parameter θ met betrouwbaarheid $1-\alpha$ is een gebied $C = C(x_1,\ldots,x_n)$ met de eigenschap dat

$$(8.3.1) \qquad P_\theta(C(\underline{x}_1,\ldots,\underline{x}_n) \ni \theta) = 1-\alpha \qquad (\theta\in\Omega).$$

Speciaal, als θ een-dimensionaal is, dan is een betrouwbaarheidsinterval voor θ een interval (θ_1,θ_2) met $\theta_1 = \theta_1(x_1,\ldots,x_n)$ en $\theta_2 = \theta_2(x_1,\ldots,x_n)$ zó, dat

$$P_\theta(\underline{\theta}_1 < \theta < \underline{\theta}_2) = 1-\alpha.$$

Opmerking 8.3.2: Bij gegeven waarnemingen, d.w.z. bij een realisatie van de steekproef is een betrouwbaarheidsgebied en vaste verzameling, waarover geen kansuitspraken kunnen worden gedaan. Op grond van de wet van de grote aantallen, zal wel bij dikwijls herhalen van het waarnemen in een fractie $1-\alpha$ van de gevallen dit gebied de gezochte parameterwaarde bevatten.

In veel gevallen vindt men betrouwbaarheidsintervallen op de volgende manier: $x_1,\ldots,x_n$ is de realisatie van $\underline{x}_1,\ldots,\underline{x}_n$ met aannemelijkheidsfunctie $L(x_1,\ldots,x_n;\theta)$ (θ een-dimensionaal). De grootheid $\underline{T}$ is een schatter voor θ met de eigenschap dat $P(\underline{T} \le T;\theta)$ monotoon,bijvoorbeeld dalend, is in θ (en stijgend in T) terwijl dan $P(\underline{T} \ge T;\theta)$ stijgend is in θ (en dalend in T). Men bepaalt θ_1 en θ_2 uit definitie 8.3.1 door van de overschrijdingskansen voor de waargenomen waarde T te eisen dat

$$(8.3.2)\qquad P(\underline{T} \le T;\theta_2) = P(\underline{T} \ge T;\theta_1) = \alpha/2,$$

d.w.z. die waarden van θ, waarvoor de kans op overschrijding van de gevonden waarde T gelijk is aan $\alpha/2$, vormen de grenzen van het betrouwbaarheidsinterval. We hebben dan $\theta_2=\theta_2(T)$ en $\theta_1=\theta_1(T)$, waarbij θ_1 en θ_2 stijgend zijn in T (ga na), zodat de volgende identiteit in T geldt:

$$P(\underline{T}\le T;\theta_2)=P(\underline{T}\le T;\theta_2(T))=P(\theta_2(\underline{T})\le\theta_2(T);\theta_2(T))=\alpha/2,$$

en dus voor willekeurige $\theta \in \Omega$ met $\theta = \theta_2(T)$ geldt $P_\theta(\underline{\theta}_2\le\theta):=P_\theta(\theta_2(\underline{T})\le\theta)=\alpha/2$ en evenzo $P_\theta(\underline{\theta}_1\ge\theta)=\alpha/2$, d.w.z.

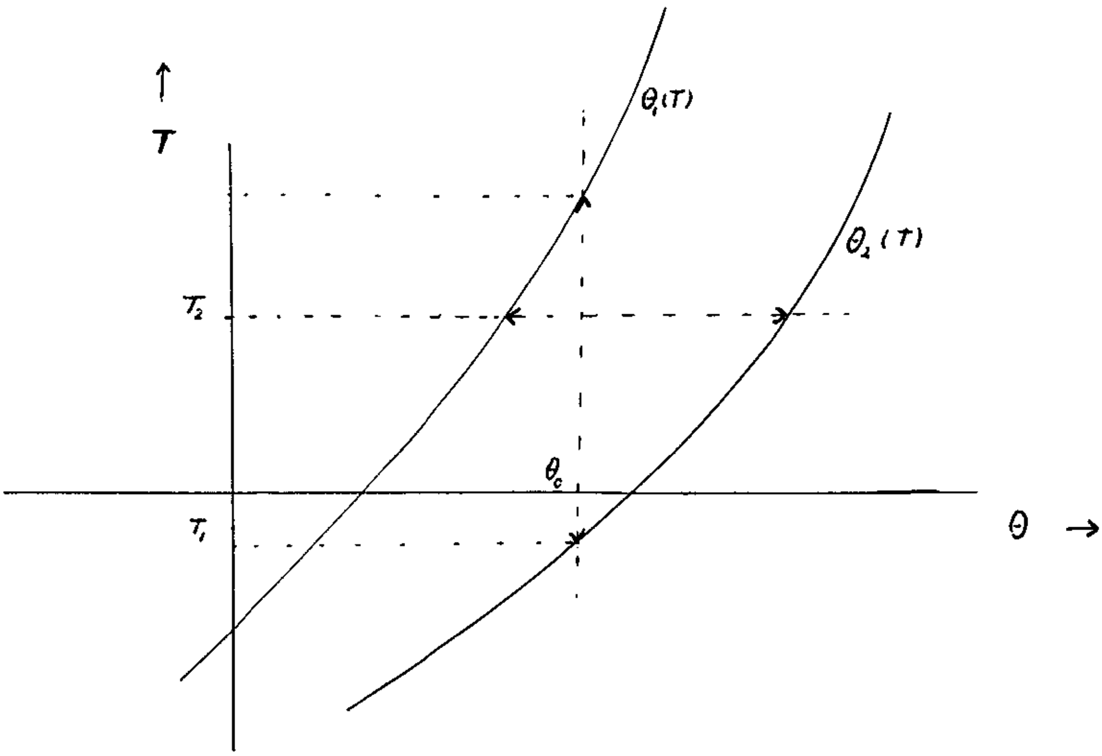
$$(8.3.3)\qquad P_\theta(\underline{\theta}_1 < \theta < \underline{\theta}_2) = 1-\alpha.$$

Dus $(\underline{\theta}_1,\underline{\theta}_2)$ is een betrouwbaarheidsinterval voor θ met betrouwbaarheid $1-\alpha$. De samenhang met een toets voor θ is als volgt: voor vaste $\theta = \theta_0$ hebben we $\theta_0 = \theta_1(T_2) = \theta_2(T_1)$ met $T_1 < T_2$. Voor $\theta = \theta_0$ gaat (8.3.3) dan over in

$$1-P_{\theta_0}(\underline{\theta}_1<\theta_0<\underline{\theta}_2)=P_{\theta_0}(\theta_1(\underline{T})\ge\theta_1(T_2) \text{ of } \theta_2(\underline{T})\le\theta_2(T_1)) =$$

$$=P_{\theta_0}(\underline{T}\le T_1 \text{ of } \underline{T}\ge T_2) = \alpha.$$

Het betrouwbaarheidsinterval met betrouwbaarheid $1-\alpha$ correspondeert dus met een toets voor θ met kritiek gebied $(T_1,T_2)^*$, waarbij T_1 en T_2 functies zijn van θ bepaald door $\theta = \theta_1(T_2) = \theta_2(T_1)$; bij vaste T bestaat het betrouwbaarheidsinterval uit al die waarden van θ die bij toetsing niet worden verworpen: $\theta \in (\theta_1(T), \theta_2(T)) \Leftrightarrow T \in (T_1(\theta), T_2(\theta))$. De correspondentie tussen betrouwbaarheidsinterval en aanvaardingsinterval is geschetst in de onderstaande figuur.



Voorbeeld 8.3.3: Zij $x_1,\ldots,x_n$ de realisatie van een aselecte steekproef van $\underline{x}$ met $\underline{x}$ $N(\mu,\sigma^2)$-verdeeld. We zoeken een betrouwbaarheidsinterval met betrouwbaarheid $1-\alpha$ voor μ.

A. σ^2 bekend. Als toetsingsgrootheid gebruiken we $\bar{\underline{x}} = (\underline{x}_1+\ldots+\underline{x}_n)/n$. We weten dat *bij gegeven* μ de grootheid $\underline{u}=(\bar{\underline{x}}-\mu)\sqrt{n}/\sigma$ een $N(0,1)$-verdeling heeft. We vinden dus de grenzen van het betrouwbaarheidsinterval uit de voorwaarden

$$P(\bar{\underline{x}} \leq \bar{x};\ \mu_2) = \alpha/2 \quad ; \quad P(\bar{\underline{x}} \geq \bar{x};\ \mu_1) = \alpha/2$$

en dus (vergelijk voorbeeld 8.2.1 A)

$$\frac{\bar{x}-\mu_1}{\sigma}\sqrt{n} = u_{\alpha/2} \quad ; \quad \frac{\bar{x}-\mu_2}{\sigma}\sqrt{n} = -u_{\alpha/2},$$

zodat

$$(\mu_1,\mu_2) = (\bar{x} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2},\ \bar{x} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}).$$

B. σ^2 onbekend. Bij gegeven μ gebruiken we de toetsingsgrootheid (zie voorbeeld 8.2.1 B) $\underline{t} = (\bar{\underline{x}}-\mu)\sqrt{n}/\underline{s}$, die (bij gegeven μ) een t_{n-1}-verdeling heeft. Geheel analoog vinden we nu het interval

$$(\mu_1,\mu_2) = (\bar{x} - \frac{s}{\sqrt{n}}\, t_{\alpha/2},\ \bar{x} + \frac{s}{\sqrt{n}}\, t_{\alpha/2}).$$

Opgave 8.3.4: Zij $x_1,\ldots,x_{100}$ de realisatie van een aselecte steekproef van $\underline{x}$ met $\underline{x}$ $N(\mu,\sigma^2)$ verdeeld. Zij $\bar{x} = 1{,}24$ en $s^2 = 0{,}0081$.
Bereken een betrouwbaarheidsinterval met betrouwbaarheid 0,96 voor σ^2,

A. als μ bekend is: $\mu = 1{,}20$,

B. als μ onbekend is (vergelijk voorbeeld 8.2.4).

Zoals al door het bovenstaande wordt gesuggereerd, geldt de volgende stelling.

Stelling 8.3.5: Zij $x_1,\ldots,x_n$ de realisatie van een steekproef $\underline{x}_1,\ldots,\underline{x}_n$ met aannemelijkheidsfunctie $L(x_1,\ldots,x_n;\theta)$ $(\theta\in\Omega)$. Zij $A = A(\theta) \subset \mathbb{R}^n$ het aanvaardingsgebied van een toets voor de hypothese θ met onbetrouwbaarheidsdrempel α, d.w.z.

$$P_\theta((\underline{x}_1,\ldots,\underline{x}_n) \in A(\theta)) \geq 1-\alpha,$$

dan is de verzameling parameterwaarden $C = C(x_1,\ldots,x_n) \subset \Omega$ gedefinieerd door

$$C(x_1,\ldots,x_n) = \{\theta\in\Omega \mid (x_1,\ldots,x_n)\in A(\theta)\}$$

een betrouwbaarheidsgebied voor θ met betrouwbaarheid $\geq 1-\alpha$, d.w.z. de verzameling van alle parameterwaarden die bij toetsing niet verworpen worden vormen een betrouwbaarheidsgebied voor θ (vergelijk definitie van A en R, definitie 8.1.3).

Bewijs: $P_\theta(C(\underline{x}_1,\ldots,\underline{x}_n) \ni \theta) = P_\theta((\underline{x}_1,\ldots,\underline{x}_n) \in A(\theta)) \geq 1-\alpha.$

Opmerking 8.3.6: Stelling 8.3.5 geeft ook bij een eenzijdige toets een (eenzijdig) betrouwbaarheidsgebied. Ook analoog aan (8.3.2) vinden we een eenzijdig betrouwbaarheidsinterval $(-\infty,\theta_2')$ of (θ_1',∞) door te eisen

$$P(\underline{T} \leq T;\theta_2') = \alpha$$

of

$$P(\underline{T} \geq T;\theta_1') = \alpha.$$

In voorbeeld 8.3.3 A krijgen we dan

$$\text{linkseenzijdig: } \mu_2' = \bar{x} + \frac{\sigma}{\sqrt{n}} u_\alpha$$

$$\text{rechtseenzijdig: } \mu_1' = \bar{x} - \frac{\sigma}{\sqrt{n}} u_\alpha .$$

Opmerking 8.3.7: Stelling 8.3.5 houdt in, dat een betrouwbaarheidsinterval ook een toets levert: alle $\theta \notin C(x_1,\ldots,x_n)$ worden verworpen.

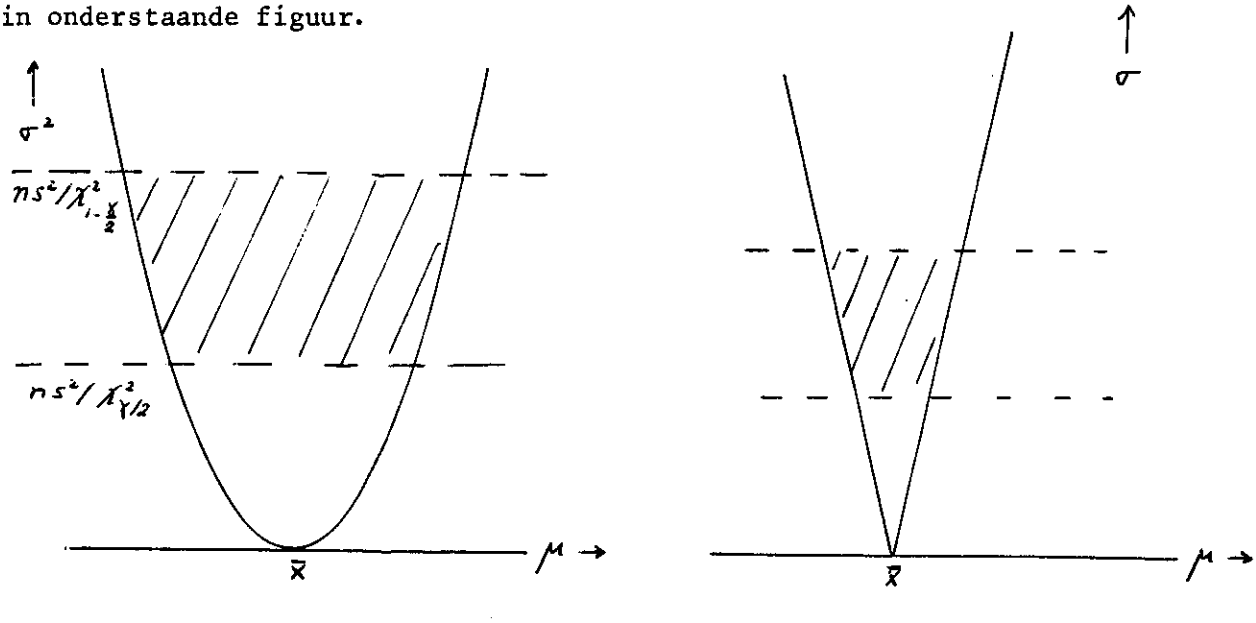
Voorbeeld 8.3.8: Zij $x_1,\ldots,x_n$ de realisatie van een aselecte steekproef van $\underline{x}$ met $\underline{x}$ $N(\mu,\sigma^2)$-verdeeld. We construeren een betrouwbaarheidsgebied voor de parameter (μ,σ^2). We gebruiken de toetsingsgrootheden $\bar{\underline{x}}$ en $\underline{s}^2$, waarvan we weten dat ze onafhankelijk zijn, terwijl $\bar{\underline{x}}$ een $N(\mu,\sigma^2/n)$-verdeling heeft en $(n-1)\underline{s}^2/\sigma^2 = \sigma^{-2}\sum_1^n(\underline{x}_j-\bar{\underline{x}})^2$ een χ^2_{n-1}-verdeling. Omdat $\bar{\underline{x}}$ en $\underline{s}^2$ onafhankelijk zijn, geldt (zie voorbeeld 8.2.4)

$$P(|\frac{\bar{\underline{x}}-\mu}{\sigma}\sqrt{n}| < u_{\beta/2},\ \chi^2_{1-\gamma/2} < n\underline{s}^2/\sigma^2 < \chi^2_{\gamma/2}\ ;\ \mu,\sigma^2) = (1-\beta)(1-\gamma).$$

Voor $(1-\beta)(1-\gamma) = 1-\alpha$ vinden we o.g.v. stelling 8.3.5 als betrouwbaarheidsgebied met betrouwbaarheid $1-\alpha$ de volgende verzameling van (μ,σ^2)-waarden:

$$\{(\mu,\sigma^2)\ |\ |\frac{\bar{x}-\mu}{\sigma}\sqrt{n}| < u_{\beta/2}\ ,\ \chi^2_{1-\gamma/2} < ns^2/\sigma^2 < \chi^2_{\gamma/2}\}.$$

Dit gebied en het corresponderende gebied in het (μ,σ)-vlak is geschetst in onderstaande figuur.

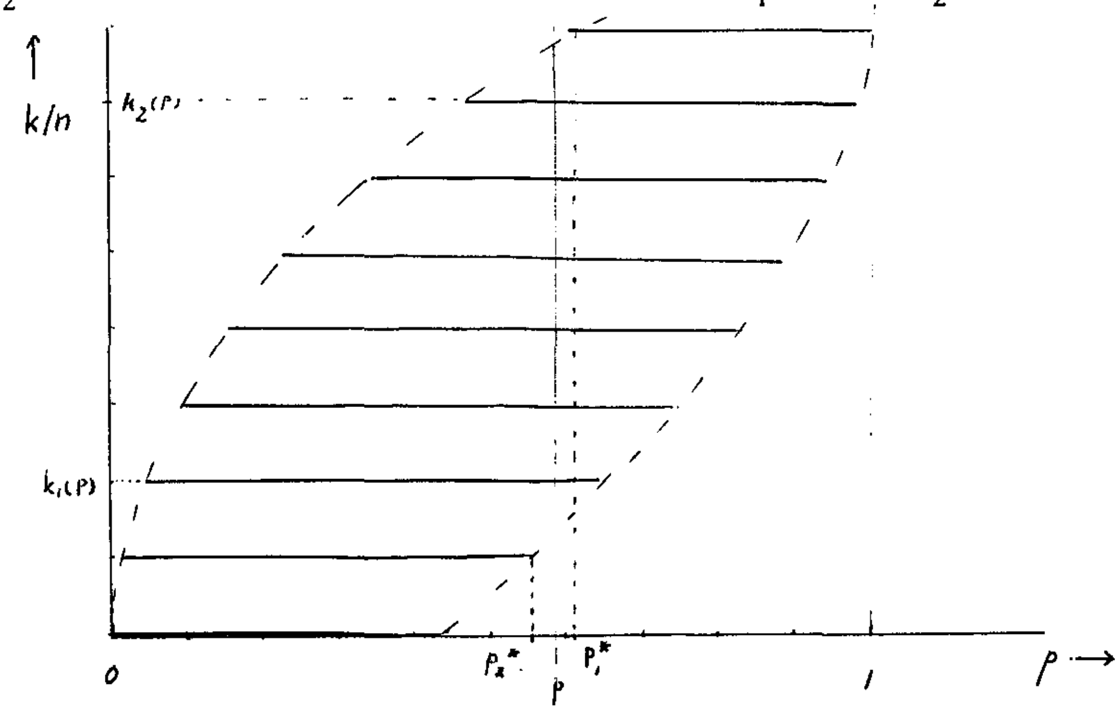


Als we te maken hebben met diskrete verdelingen, dan hebben toetsen met onbetrouwbaarheidsdrempel α dikwijls een onbetrouwbaarheid die kleiner is dan α; als gevolg hiervan hebben de corresponderende betrouwbaarheidsgebieden dan een betrouwbaarheid die groter is dan $1-\alpha$. Als voorbeeld beschouwen we betrouwbaarheidsintervallen voor de kans p op succes.

Voorbeeld 8.3.9: Zij $x_1,\ldots,x_n$ een realisatie van een aselecte steekproef van $\underline{x}$ met $P(\underline{x}=1)=1-P(\underline{x}=0)=p$. We bepalen een tweezijdig betrouwbaarheidsinterval voor p uitgaande van (8.3.2): we bepalen p_1 en p_2 zó dat bij de realisatie k van de toetsingsgrootheid $\underline{k} = \underline{x}_1+\ldots+\underline{x}_n$ geldt

$$(8.3.4) \qquad P(\underline{k} \le k; p_2) = P(\underline{k} \ge k;\, p_1) = \alpha/2 \quad (1 \le k \le n-1);$$

dit is mogelijk, omdat beide kansen *continu* afhangen van p. Omdat (zie lemma 8.3.10) $P(\underline{k} \ge k;p)$ stijgend is in p voor alle k met $1 \le k \le n-1$, en dus $P(\underline{k} \le k;p)$ dalend, zijn p_1 en p_2 in (8.3.4) eenduidig bepaald: $p_1=p_1(k)$ en $p_2=p_2(k)$ (zie onderstaande figuur). We definiëren $p_1(0)=0$ en $p_2(n) = 1$.



We berekenen nu de betrouwbaarheid van het interval (p_1,p_2) met $p_1=p_1(k)$ en $p_2 = p_2(k)$. We hebben

$$P_p(\underline{p}_1 < p < \underline{p}_2) = 1-P_p(\underline{p}_1 \ge p) - P_p(\underline{p}_2 \le p).$$

Bij een gegeven waarde van p is er een kleinste waarde van k, $k_1(p)$, waarbij p niet wordt verworpen, d.w.z. waarvoor p in het bijbehorende betrouwbaarheidsinterval ligt; evenzo is er een grootste waarde van k, $k_2(p)$, waarvoor p in het daarbij behorende betrouwbaarheidsinterval ligt. Op grond van de monotonie geldt dan dat voor $k_1(p) \le k \le k_2(p)$ de waarde p in het betrouwbaarheidsinterval van k ligt (en niet in het betrouwbaarheidsinterval behorende bij andere waarden van k), d.w.z.

$$P_p(\underline{p}_1 < p < \underline{p}_2) = P_p(k_1(p) \le \underline{k} \le k_2\ (p)).$$

Zij nu $p_1^* = p_1(k_2(p)+1)$ en $p_2^* = p_2(k_1(p)-1)$, dan is i.h.a. (zie figuur) $p_1^* > p$ en $p_2^* < p$, zodat o.g.v. de monotonie (zie lemma 8.3.10)

$$P_p(\underline{k} \le k_1(p)-1) < P_{p^*}(\underline{k} \le k_1(p)-1) = \alpha/2$$

$$P_p(\underline{k} \ge k_2(p)+1) < P_{p^*}(\underline{k} \ge k_2(p)+1) = \alpha/2.$$

Dan is ook $P_p(\underline{p}_1 < p < \underline{p}_2) > 1-\alpha$.

<u>Lemma 8.3.10:</u>

$$P(\underline{k} \ge k;p) := \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

is stijgend in p voor k=1,2,... .

Bewijs: door differentiatie verifieert men gemakkelijk dat

$$\sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \frac{1}{B(k,n-k+1)} \int_0^p t^{k-1}(1-t)^{n-k}dt$$

(vergelijk voorbeeld 3.5.13).

<u>Opgave 8.3.11:</u> $x_1+\ldots+x_{10}$ is een realisatie van een steekproef van $\underline{x}$ met $P(\underline{x}=1)=1-P(\underline{x}=0)=p$. Bereken de betrouwbaarheidsintervallen voor p als $x_1+\ldots+x_n=k$ voor $k=0,1,\ldots,10$.
Schets het verloop van de werkelijke betrouwbaarheid als functie van p door $P_p(\underline{k}\le k_1(p)-1)$ en $P_p(\underline{k}\ge k_2(p)+1)$ uit te rekenen.

Voor grote waarden van n benaderen we de (binomiale) verdeling van $\underline{k}$ weer door een normale verdeling. We bepalen de benaderende grenzen van het betrouwbaarheidsinterval dan als volgt (zie opmerking 5.4.7):

$$P(\underline{k} \geq k; p_1) \approx 1 - \Phi\left(\frac{k - np_1 - \frac{1}{2}}{\sqrt{np_1(1-p_1)}}\right) = \alpha/2$$

$$P(\underline{k} \leq k; p_2) \approx \Phi\left(\frac{k - np_2 + \frac{1}{2}}{\sqrt{np_2(1-p_2)}}\right) = \alpha/2.$$

We vinden nu p_1 als de kleinste oplossing van

$$(k-np-\tfrac{1}{2})^2 = (u_{\alpha/2})^2\, np(1-p)$$

en p_2 als de grootste oplossing van

$$(k-np+\tfrac{1}{2})^2 = (u_{\alpha/2})^2\, np(1-p).$$

Voor grotere n (bijv. n > 100) kunnen we de continuïteitscorrectie weglaten; we vinden dan p_1 en p_2 als de oplossingen van

$$(k-np)^2 = (u_{\alpha/2})^2\, np(1-p).$$

Een benaderende oplossing hiervan is (ga na)

$$p_{1,2} \approx \tilde{p} \mp u\sqrt{\tilde{p}(1-\tilde{p})/n},$$

met $u = u_{\alpha/2}$ en $\tilde{p} = (k+\frac{u^2}{2})/(n+u^2)$. Voor zeer grote n gebruikt men de (verdere) benadering

$$\tilde{p}_{1,2} \approx \frac{k}{n} \mp u\sqrt{\frac{k}{n}(1-\frac{k}{n})\frac{1}{n}}.$$

Opgave 8.3.12: Bij 10000 enkelvoudige geboorten zijn 5150 van de babies jongens. Geef een betrouwbaarheidsinterval met een betrouwbaarheid van 0,99 voor p:= P(jongen).

Opgave 8.3.13: Hoe bepaalt men een betrouwbaarheidsinterval voor de parameter μ van een Poisson-verdeling; hoe ziet de benadering voor grote steekproeven er uit?

Opmerking 8.3.14: Eenzijdige betrouwbaarheidsintervallen voor p vinden we door oplossing van

$$P(\underline{k} \geq k;\ p_1') = \alpha \ ;$$

dit levert de linkergrens voor een rechts-eenzijdig betrouwbaarheidsinterval $(p_1', 1]$. De normale benadering is analoog aan die in het tweezijdige geval.

APPENDIX

A3 Ongelijkheid van Cramér-Rao

We geven hier een iets algemenere versie van stelling 7.1.18.

Stelling A3.1 (Ongelijkheid van Cramér-Rao): Zij $\underline{x}_1,\ldots,\underline{x}_n$ een steekproef met likelihood-functie $L(x_1,\ldots,x_n;\theta)$ met $(x_1,\ldots,x_n) \in X \subset \mathbb{R}^n$ en $\theta \in \Omega$ (zie definitie 7.2.1) en zij $\underline{t} = t(\underline{x}_1,\ldots,\underline{x}_n)$ een schatter voor θ met $E\,\underline{t} = \alpha(\theta)$ en $E\,\underline{t}^2 < \infty$ voor alle $\theta \in \Omega$. Als voldaan is aan de volgende voorwaarden (hierbij korten we $L(x_1,\ldots,x_n;\theta)$ af tot $L(x;\theta)$):

(i) $\frac{\partial}{\partial\theta} \log L(x;\theta)$ bestaat voor alle $x \in X$ en $\theta \in \Omega$,

(ii) $\frac{\partial}{\partial\theta} \int_X L(x;\theta)dx = \int_X \frac{\partial}{\partial\theta} L(x;\theta)dx \quad (\theta \in \Omega)$,

(iii) $\frac{\partial}{\partial\theta} \int_X t(x)L(x;\theta)dx = \int_X t(x) \frac{\partial}{\partial\theta} L(x;\theta)dx \quad (\theta \in \Omega)$,

(iv) $0 < E_\theta \{\frac{\partial}{\partial\theta} \log L(\underline{x};\theta)\}^2 < \infty \quad (\theta \in \Omega)$,

dan geldt

$$\text{(A3.1)} \qquad \operatorname{var} \underline{t} \ge \frac{\{\alpha'(\theta)\}^2}{E\{\frac{\partial}{\partial\theta} \log L(\underline{x};\theta)\}^2} \qquad (\theta \in \Omega).$$

Bewijs: de voorwaarden betekenen dat onder het integraalteken gedifferentieerd mag worden. We hebben (o.g.v. (i) is $L(x;\theta) > 0$ voor $x \in X$ en $\theta \in \Omega$)

$$\text{(A 3.2)} \qquad 1 \equiv \int_X L(x;\theta)dx \Rightarrow 0 \equiv \int_X \frac{\partial}{\partial\theta} L(x;\theta)dx = E \frac{\partial}{\partial\theta} \log L(\underline{x};\theta),$$

omdat $\int \frac{\partial}{\partial\theta} L(x;\theta)dx = \int \frac{\partial}{\partial\theta}\{\log L(x;\theta)\}L(x;\theta)dx = E \frac{\partial}{\partial\theta} \log L(\underline{x};\theta)$. Verder is op dezelfde manier

$$\text{(A 3.3)} \qquad \alpha'(\theta) = \int_X t(x) \frac{\partial}{\partial\theta} L(x;\theta)dx = E\, \underline{t} \frac{\partial}{\partial\theta} \log L(\underline{x};\theta).$$

Als we afkorten $\underline{u}:= \frac{\partial}{\partial\theta} \log L(\underline{x};\theta)$ (dit is o.g.v. (i) en (ii) een stochastische grootheid) dan geldt

$$\alpha'(\theta) = E\,\underline{t}\,\underline{u} = E\,(\underline{t} - E\underline{t})\underline{u},$$

omdat $E\,\underline{u} = 0$ (zie (A.3.2)) . De ongelijkheid van Schwarz (stelling 4.2.8) levert nu

$$\{\alpha'(\theta)\}^2 = \{E(\underline{t} - E\underline{t})\,\underline{u}\}^2 \leq \text{var}\,\underline{t}.\ E\,\underline{u}^2,$$

equivalent met (A 3.1).

<u>Opmerking A 3.2:</u> als nogmaals onder het integraalteken gedifferentieerd mag worden, dan blijkt (ga na) dat

$$- E\,\frac{\partial^2 \log L(\underline{x};\theta)}{\partial\theta^2} = E\,\{\frac{\partial \log L(\underline{x};\theta)}{\partial\theta}\}^2.$$

Ongelijkheid (A 3.1) gaat dan over in

$$\text{(A.3.4)} \qquad \text{var}\,\underline{t} \geq \frac{-\{\alpha'(\theta)\}^2}{E\,\frac{\partial^2}{\partial\theta^2} \log L(\underline{x};\theta)}\,.$$

<u>Opmerking A 3.3:</u> Gelijkheid geldt in (A 3.1) d.e.s.d. als met kans 1

$$\text{(A 3.5)} \qquad \frac{\partial}{\partial\theta} \log L(\underline{x}_1,\ldots,\underline{x}_n;\theta) = c_n(\theta)(\underline{t}-E\underline{t}) \equiv c_n(\theta)(\underline{t}-\alpha(\theta));$$

dit volgt ook uit stelling 4.2.8. Door de uiterste leden van (A 3.5) te kwadrateren en vervolgens de verwachting ($E=E_\theta$) te nemen blijkt dat

$$E\,\{\frac{\partial}{\partial\theta} \log L(\underline{x}_1,\ldots,\underline{x}_n;\theta)\}^2 = (c_n(\theta))^2\ \text{var}\,\underline{t}\ ,$$

zodat (zie (A 3.1) en vergelijk opgave 7.1.21)

$$|c_n(\theta)| = \frac{|\alpha'(\theta)|}{\text{var}\,\underline{t}}\ .$$

A 4 Onafhankelijkheid van $\underline{s}^2$ en $\bar{\underline{x}}$

Stelling A 4.1: Als $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onderling onafhankelijk zijn en N(0,1)-verdeeld, dan zijn de grootheden $\bar{\underline{x}} = (\underline{x}_1 + \ldots + \underline{x}_n)/n$ en $\underline{s}^2 = (n-1)^{-1} \sum_1^n (\underline{x}_j - \bar{\underline{x}})^2$ onderling onafhankelijk. Bovendien heeft de grootheid $(n-1)\underline{s}^2$ een χ^2_{n-1}-verdeling.

Bewijs: we maken gebruik van het feit dat er een orthonormale matrix A bestaat, met elementen a_{ij} ($i=1,2,\ldots,n$; $j=1,2,\ldots n$), die behalve de eigenschappen $\sum_j a_{ij}^2 = 1$ en $\sum_j a_{ij}a_{kj} = 0$ $(i \neq k)$ nog de eigenschap heeft dat

$$\text{(A 4.1)} \qquad a_{nj} = \frac{1}{\sqrt{n}} \quad (j=1,2,\ldots,n).$$

Dat zo'n A bestaat is een gevolg van het feit dat we n-1 orthonormale vectoren kunnen vinden in de ruimte $\{x \in \mathbb{R}^n \mid x_1 + \ldots + x_n = 0\}$, die isomorf is met $\mathbb{R}^{n-1}$.

We definiëren nu $\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_n$ door (in (kolom)vectornotatie)

$$\text{(A 4.2)} \qquad \underline{u} = A\,\underline{x}.$$

Dan is $\sum_1^n \underline{u}_j^2 = \underline{u}^T\underline{u} = \underline{x}^T A^T A \underline{x} = \underline{x}^T\underline{x} = \sum_1^n \underline{x}_j^2$, omdat wegens de orthonormaliteit van A geldt dat $A^TA = I$. Voor de kansdichtheid van $\underline{u}_1, \ldots, \underline{u}_n$ hebben we dan (vergelijk opgave 3.7.4)

$$f_{\underline{u}_1,\ldots,\underline{u}_n}(u_1,\ldots,u_n) = f_{\underline{x}_1,\ldots,\underline{x}_n}(x_1(u_1,\ldots,u_n),\ldots,x_n(u_1,\ldots,u_n)) \left| \frac{\partial(x_1,\ldots,x_n)}{\partial(u_1,\ldots,u_n)} \right|$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\tfrac{1}{2} \sum_1^n x_j^2(u_1,\ldots,u_n)\}.\det A$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\{-\tfrac{1}{2} \sum_1^n u_j^2\},$$

d.w.z. $\underline{u}_1, \ldots, \underline{u}_n$ zijn o.o. en N(0,1) verdeeld. Nu is wegens (A 4.1)

$$\text{(A 4.3)} \qquad \underline{u}_n = \sum_1^n \underline{x}_j/\sqrt{n} = \sqrt{n}\,\bar{\underline{x}},$$

zodat

$$\text{(A 4.4)} \qquad (n-1)\underline{s}^2 = \sum_1^n (\underline{x}_j - \bar{\underline{x}})^2 = \sum_1^n \underline{x}_j^2 - n\bar{\underline{x}}^2 = \sum_1^{n-1} \underline{u}_j^2 .$$

Uit (A 4.3) en (A 4.4) volgt nu direct dat $\bar{\underline{x}}$ en $\underline{s}^2$ onafhankelijk zijn en dat $(n-1)\underline{s}^2 = \sum_1^{n-1} \underline{u}_j^2$ een χ^2_{n-1} verdeling heeft (vergelijk definitie A 2.1.1).

Gevolg A 4.2: Als $\underline{x}_1,\ldots,\underline{x}_n$ o.o. zijn en $N(\mu,\sigma^2)$-verdeeld, dan zijn $\bar{\underline{x}}$ en $\underline{s}^2$ o.o., terwijl

$$\frac{1}{\sigma^2} \sum_1^n (\underline{x}_j - \bar{\underline{x}})^2$$

een χ^2_{n-1} verdeling heeft.

Opmerking A 4.3: Voor de matrix A in het bewijs van stelling A 4.1 kunnen we nemen

$$A = \begin{pmatrix}
\frac{1}{\sqrt{1.2}} & -\frac{1}{\sqrt{1.2}} & 0 & \cdots & \cdots & 0 \\
\frac{1}{\sqrt{2.3}} & \frac{1}{\sqrt{2.3}} & -\frac{2}{\sqrt{2.3}} & 0 & \cdots & 0 \\
\vdots & & & & & \vdots \\
 & & & & & 0 \\
\frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \cdots & \frac{1}{\sqrt{(n-1)n}} & \frac{1}{\sqrt{(n-1)n}} & \frac{-(n-1)}{\sqrt{(n-1)n}} \\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}}
\end{pmatrix} .$$

Opgave A 4.4: Bewijs dat

$$\frac{(\bar{\underline{x}} - \mu)\sqrt{n(n-1)}}{\sqrt{\sum_1^n (\underline{x}_j - \bar{\underline{x}})^2}}$$

een t_{n-1}-verdeling heeft.