Les 3 Betrouwbaarheidsintervallen

In de vorige les hebben we erna gekeken hoe we grootheden van een populatie met behulp van steekproeven kunnen schatten. We hebben daarbij gezien dat de nauwkeurigheid van een schatting met de grootte van de steekproef toeneemt, want de steekproefstandaardafwijking neemt met $\frac{1}{\sqrt{n}}$ af. We zullen in deze les bekijken hoe we uitspraken erover kunnen maken dat een schatting met een foutmarge de juiste waarde met een gegeven kans bevat. Hierbij moeten we in het bijzonder precies formuleren, wat de uitspraak dat *een waarde met een betrouwbaarheid van* 95% *in een zeker interval ligt* eigenlijk betekent.

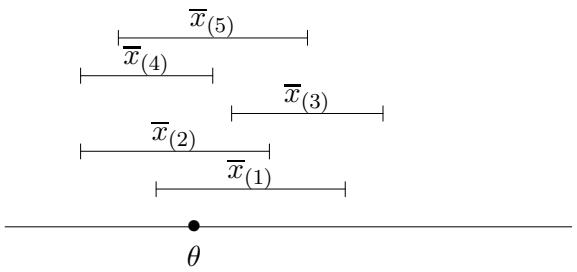
3.1 Intervalschatters

De schatters die we in de vorige les hebben bekeken, noemt men vaak *puntschatters* omdat ze voor een gegeven steekproef een precieze waarde voor een parameter opleveren. Bijvoorbeeld levert de schatter $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ voor het gemiddelde van een populatie op een gegeven steekproef $x_1, \ldots, x_n$ de schatting $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

In tegenstelling hiertoe geeft een *intervalschatter* voor een gegeven steekproef een interval aan waarin de juiste waarde θ van de parameter moet liggen. Hierbij wordt altijd een level γ van *betrouwbaarheid* geëist, waarmee het interval de juiste waarde bevat. De betrouwbaarheid γ wordt als volgt geïnterpreteerd: Voor een gegeven waarde van θ is γ de kans dat een steekproef een interval oplevert dat θ bevat. We kijken dus weer naar alle mogelijke steekproeven en analyseren de verdeling van de schattingen.

Merk op: Een betrouwbaarheid van 95% voor een interval betekent *niet* dat de juiste waarde θ met kans 95% in het interval ligt, maar dat onze methode om het interval te schatten voor 95% van de mogelijke steekproeven een interval oplevert, dat θ bevat.

Bij een betrouwbaarheid van $\gamma = 0.8$ zouden we dus bij vijf steekproeven verwachten, dat de juiste parameter vier keer in het geschatte interval ligt, bijvoorbeeld zo als in het volgende plaatje met de intervallen rond de schattingen $\overline{x}_{(i)}$ aangegeven.



In de taal van stochasten en schatters levert dit idee van betrouwbaarheid het volgende concept op. Zij X een stochast met dichtheidsfunctie $f(x) := f(x;\theta)$ en verdelingsfunctie $F(x) := F(x;\theta)$ die van een parameter θ afhangen, dan berekenen we de kansen voor X door

$$P(X \leq x) = P_\theta(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)\, dt.$$

We noemen een paar (T_1, T_2) van stochasten een *intervalschatter van betrouwbaarheid* γ *voor* θ als

$$P(T_1 \leq \theta \leq T_2) = \gamma \text{ voor elke mogelijke waarde van de parameter } \theta.$$

Een realisatie van een intervalschatter op een concrete steekproef $x_1, \ldots, x_n$ heet een *betrouwbaarheidsinterval* van betrouwbaarheid γ voor θ.

Omdat we de waarde van θ van twee zijden ingeschakeld hebben, noemen we het paar (T_1, T_2) ook een *tweezijdige intervalschatter*.

> Als we in de praktijk een betrouwbaarheidsinterval voor de verwachtingswaarde $\mu := E[X]$ schatten, zal het interval bijna altijd symmetrisch rond het steekproefgemiddelde $\overline{x}$ liggen. Dit is geen noodzakelijke voorwaarde maar wel heel gebruikelijk. Er laat zich aantonen dat voor een normaal verdeelde stochast X het symmetrische interval rond $\overline{x}$ de kleinste lengte van alle intervallen met betrouwbaarheid γ heeft.

Soms is het interessant om alleen maar een boven- of een benedengrens voor een parameter te schatten. Dit levert *éénzijdige intervalschatters*. We noemen een stochast T_1 een *rechtséénzijdige intervalschatter* van betrouwbaarheid γ als

$$P(T_1 \leq \theta) = \gamma \text{ voor elke mogelijke waarde van de parameter } \theta$$

en we noemen een stochast T_2 een *linkséénzijdige intervalschatter* van betrouwbaarheid γ als

$$P(\theta \leq T_2) = \gamma \text{ voor elke mogelijke waarde van de parameter } \theta.$$

De reden waarom de stochast T_1 met $P(T_1 \leq \theta) = \gamma$ *rechts*éénzijdig heet, hangt met de éénzijdige toetsen samen die we in de volgende les gaan behandelen.

3.2 Betrouwbaarheidsintervallen bij gegeven variantie

Als belangrijk voorbeeld zullen we naar een intervalschatter kijken die voor een normaal verdeelde stochast X met bekende variantie σ^2 een betrouwbaarheidsinterval voor de verwachtingswaarde μ van X geeft.

Hetzelfde principe werkt bij benadering voor de verwachtingswaarde van niet normaal verdeelde stochasten, in het bijzonder voor de verwachte kans op succes bij een binomiale verdeling.

> De centrale limietstelling zegt dat de som van onafhankelijke stochasten goed benaderd wordt door een normale verdeling. Hieruit volgt dat de vorm van de onderzochte stochast X geen grote rol speelt als de steekproefgrootte n niet te klein is. Maar er zijn wel andere problemen, waardoor de verdeling van schattingen van de normale verdeling afwijkt. Deze hebben vooral met de veronderstelling te maken dat we een *aselecte* steekproef hebben genomen. Dit is in de praktijk vaak

lastig, omdat mensen bijvoorbeeld een enquête weigeren, maar dit niet representatief over de populatie gebeurt. Ook is het vaak niet realistisch, dat de verschillende steekproefelementen onafhankelijk van elkaar genomen worden. Het is de kunst van de instituten voor opinieonderzoek deze factoren zo ver mogelijk te onderdrukken of de resultaten navenant te corrigeren.

Stel we hebben een normaal verdeelde stochast $X \in \mathcal{N}(\mu, \sigma^2)$ dan weten we dat $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ een zuivere schatter voor μ is. Omdat X normaal verdeeld is, geldt dit ook voor $\overline{X}$ (de som van onafhankelijke normaal verdeelde stochasten is weer normaal verdeeld) en we weten dat $Var(\overline{X}) = \frac{\sigma^2}{n}$. Hieruit volgt dat de stochast

$$Z := \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

standaard-normaal verdeeld is.

Als X een niet-normaal verdeelde stochast met verwachtingswaarde μ en variantie σ^2 is, geldt voor $\overline{X}$ nog steeds dat $E[\overline{X}] = \mu$ en $Var(\overline{X}) = \frac{\sigma^2}{n}$, maar $\overline{X}$ is niet meer normaal verdeeld. Uit de Centrale limietstelling volgt echter dat voor een niet te kleine n de verdeling van $\overline{X}$ sterk op een normale verdeling lijkt en hierdoor goed benaderd kan worden.

Voor een stochast $Z \in \mathcal{N}(0, 1)$ met standaard-normale verdeling definiëren we nu de z-waarde z_α van level $\alpha := 1 - \gamma$ door

$$P(Z > z_\alpha) = \alpha.$$

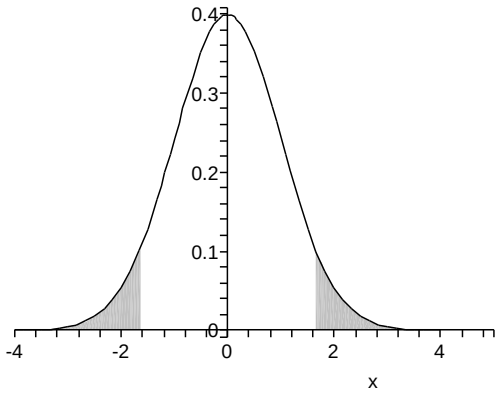
Voor een betrouwbaarheid van 95% is dus $\alpha = 0.05 = 1 - 0.95$ en geeft z_α de waarde aan, waarvoor slechts 5% van de waarden van Z boven z_α liggen en de waarden van Z dus met betrouwbaarheid 95% hoogstens z_α zijn. De level $\alpha = 1 - \gamma$ wordt ook wel de *onbetrouwbaarheid* genoemd.

Omdat de normale verdeling symmetrisch rond 0 is, geldt $P(Z < -z_\alpha) = \alpha$ en dus $P(|Z| > z_\alpha) = 2\alpha$. Hieruit volgt in het bijzonder:

$$P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha = \gamma.$$

De waarden van de standaard-normale verdeling liggen dus met kans $\gamma = 1 - \alpha$ tussen $-z_{\frac{\alpha}{2}}$ en $z_{\frac{\alpha}{2}}$. In Figuur 15 is dit voor $\gamma = 0.9$ aangeduid. Het witte stuk onder de grafiek bevat 90% van de totale oppervlakte onder de grafiek, de resterende 10% liggen in de grijze staarten, dus telkens 5% in de linker- en rechterstaart. De z-waarde $z_{0.05}$ is dus juist het punt waar de rechterstaart begint.

Als we de relatie $P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = \gamma$ nu op de standaard-normaal verdeelde stochast $Z = \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ toepassen, krijgen we voor de betrouwbaarheid



Figuur 15: Standaard-normale verdeling met betrouwbaarheidsinterval voor $\gamma = 0.9$.

γ en onbetrouwbaarheid $\alpha := 1 - \gamma$:

$$\begin{aligned}
&P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = \gamma \Leftrightarrow P(-z_{\frac{\alpha}{2}} \leq \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \leq z_{\frac{\alpha}{2}}) = \gamma \\
\Leftrightarrow\ & P(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = \gamma \\
\Leftrightarrow\ & P(\mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = \gamma \\
\Leftrightarrow\ & P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = \gamma.
\end{aligned}$$

We weten dus dat het steekproefgemiddelde met kans γ niet meer dan $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ van de juiste waarde μ afwijkt. Als intervalschatter voor het gemiddelde nemen we dus (T_1, T_2) met

$$T_1 := \overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{en} \quad T_2 := \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

en een betrouwbaarheidsinterval is een realisatie van de intervalschatter voor een concrete steekproef, dus het interval

$$\left[\overline{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

Omdat $P(\mu - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}) = P(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$ is dit precies het interval van de waarden van μ waarvoor $\overline{x}$ binnen het symmetrische interval rond μ met kansmassa γ valt. Merk op dat de lengte van het betrouwbaarheidsinterval alleen maar van de gekozen betrouwbaarheid γ, de grootte n van de steekproef en de variantie σ^2 van de stochast X afhangt.

Voor éénzijdige betrouwbaarheidsintervallen kunnen we op dezelfde manier als bij de tweezijdige intervallen argumenteren. Voor een rechtséénzijdig interval

met betrouwbaarheid γ en $\alpha := 1 - \gamma$ krijgen we:

$$P(Z \leq z_\alpha) = \gamma \Leftrightarrow P(\frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \leq z_\alpha) = \gamma \Leftrightarrow P(\overline{X} - \mu \leq z_\alpha \frac{\sigma}{\sqrt{n}}) = \gamma$$
$$\Leftrightarrow P(\overline{X} \leq \mu + z_\alpha \frac{\sigma}{\sqrt{n}}) = \gamma \Leftrightarrow P(\overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}} \leq \mu) = \gamma$$

dus is

$$T_1 := \overline{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$$

een rechtséénzijdige intervalschatter en een concrete steekproef geeft het rechtsééénzijdige betrouwbaarheidsinterval

$$\left[\overline{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, \infty\right].$$

Dit is precies het interval van de waarden van μ waarvoor $\overline{x}$ binnen het naar *rechts* begrensde en naar links open interval rond μ met kansmassa γ valt.

> We zien hier dus de reden waarom de stochast T_1 met $P(T_1 \leq \mu) = \gamma$ een *rechts*ééénzijdig betrouwbaarheidsinterval geeft. De waarden van μ die in dit éénzijdige betrouwbaarheidsinterval liggen, zijn namelijk juist de waarden waarvoor $\overline{x}$ een plausibele schatting aangeeft, als we met plausibel bedoelen, dat de schatting $\overline{x}$ niet de ver *rechts* van de ware waarde ligt.

Analoog krijgen we voor het linksééénzijdige betrouwbaarheidsinterval met betrouwbaarheid γ de schatter

$$T_2 := \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}} \quad \text{met} \quad P(\mu \leq T_2) = P(\mu \leq \overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}) = \gamma$$

en het linksééénzijdige betrouwbaarheidsinterval

$$\left[-\infty, \overline{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right].$$

Typische waarden voor de betrouwbaarheid γ zijn 90%, 95% en 99%, Tabel 1 geeft de z_α- en $z_{\frac{\alpha}{2}}$-waarden voor een paar gebruikelijke betrouwbaarheden:

γ	α	z_α	$\frac{\alpha}{2}$	$z_{\frac{\alpha}{2}}$
0.80	0.20	0.8416	0.10	1.2816
0.90	0.10	1.2816	0.05	1.6449
0.95	0.05	1.6449	0.025	1.9600
0.98	0.02	2.0537	0.01	2.3263
0.99	0.01	2.3263	0.005	2.5758
0.999	0.001	3.0902	0.0005	3.2905

Tabel 1: Kritieke waarden voor de standaard-normale verdeling.

We hebben gezien dat betrouwbaarheidsintervallen door drie parameters beschreven worden:

(i) De grote n van de steekproef.

(ii) De gewenste betrouwbaarheid γ.

(iii) De lengte van het betrouwbaarheidsinterval.

Als we de betrouwbaarheid willen verhogen, moeten we of de steekproef vergroten of een groter interval accepteren. Omgekeerd kunnen we het betrouwbaarheidsinterval alleen maar kleiner maken door of de steekproef te vergroten of een lagere level van betrouwbaarheid te kiezen. Bij een gegeven grootte van de steekproef zijn dus de lengte van het betrouwbaarheidsinterval en de betrouwbaarheid parameters, die elkaar tegenstrijdig beïnvloeden.

Bij het opzetten van een experiment (bijvoorbeeld een enquête) heeft men vaak andere voorwaarden: Voor een gegeven level γ van betrouwbaarheid is er een maximale lengte $2l$ van het betrouwbaarheidsinterval dat als acceptabel beschouwd wordt. Hierdoor wordt de noodzakelijke grootte van de steekproef bepaald, namelijk door:

$$z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq l \Rightarrow n \geq \left(z_{\frac{\alpha}{2}} \frac{\sigma}{l} \right)^2 = z_{\frac{\alpha}{2}}^2 \frac{\sigma^2}{l^2}.$$

Betrouwbaarheidsinterval voor relatieve frequenties

Als we de kans p schatten waarmee een Bernoulli-experiment een succes oplevert, tellen we het aantal k van successen bij n pogingen een nemen $\overline{p} := \frac{k}{n}$ als schatting voor p. De stochast X die de verdeling van de aantallen van successen beschrijft, is binomiaal verdeeld met parameter p en er geldt $E[X] = np$ en $Var(X) = np(1-p)$. Voor de stochast $\overline{P} := \frac{X}{n}$ die de verdeling van de relatieve aantallen beschrijft, geldt dus $E[\overline{P}] = p$ en $Var(\overline{P}) = \frac{p(1-p)}{n}$. Als n niet te klein en p niet te dicht bij 0 of 1 is, kunnen we met de normale benadering van de binomiale verdeling werken, d.w.z. we kunnen aannemen dat de stochast

$$Z := \frac{\overline{P} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{(\overline{P} - p)\sqrt{n}}{\sqrt{p(1-p)}}$$

goed door de standaard-normale verdeling benaderd wordt. In dit geval kunnen we de redenering van de normale verdeling weer toepassen en we krijgen

$$P\left(\overline{P} - z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \overline{P} + z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right) = \gamma.$$

Dit geeft het betrouwbaarheidsinterval

$$\left[\overline{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}, \overline{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right]$$

voor de schatting van de parameter p.

Het probleem bij de binomiale verdeling is, dat de variantie $\frac{p(1-p)}{n}$ en dus ook de lengte van het betrouwbaarheidsinterval van de gezochte parameter p

afhangt. In de praktijk wordt dit meestal opgelost door p gewoon door $\overline{p}$ te vervangen, men gebruikt hiervoor de *standaard fout* (standard error)

$$SE(\overline{p}) := \sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

van $\overline{p}$. De standaard fout is dus een schatting voor de standaardafwijking $\sqrt{Var(\overline{P})}$ van de schatter $\overline{P}$. Met behulp van de standaard fout krijgt men het betrouwbaarheidsinterval

$$\left[\overline{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}, \overline{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right] = \left[\overline{p} - z_{\frac{\alpha}{2}}SE(\overline{p}), \overline{p} + z_{\frac{\alpha}{2}}SE(\overline{p})\right].$$

Bij een precieze analyse komt men erachter dat de zuivere grenzen voor het betrouwbaarheidsinterval

$$\frac{\overline{p} + \frac{z_{\frac{\alpha}{2}}^2}{2n} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n} + \frac{z_{\frac{\alpha}{2}}^2}{4n^2}}}{1 + \frac{z_{\frac{\alpha}{2}}^2}{n}}$$

zijn, maar voor $n\overline{p} \geq 50$ en $n(1-\overline{p}) \geq 50$ kunnen de correctie termen veilig verwaarloosd worden.

Ook in het geval van de relatieve frequenties kan men de benodigde grootte van de steekproef afschatten om een betrouwbaarheid γ en een maximale lengte van $2l$ voor het betrouwbaarheidsinterval te bereiken. Er geldt dezelfde relatie als bij de normale verdeling, met σ^2 vervangen door $p(1-p)$, dus

$$n \geq z_{\frac{\alpha}{2}}^2 \frac{p(1-p)}{l^2}.$$

Merk op dat we hierbij ook weer de gezochte relatieve frequentie p nodig hebben. Omdat we juist willen bepalen, hoe groot we de steekproef moeten kiezen, kunnen we hier niet eens de schatting $\overline{p}$ voor p invullen, maar we kunnen natuurlijk wel een gok doen wat voor een waarde van p we verwachten.

Voorbeeld: Bij een enquête onder 1000 mensen hebben 52% aangegeven *voor* de Europese grondwet te stemmen. Een betrouwbaarheidsinterval op de level 99% geeft een nauwkeurigheid van $z_{\frac{\alpha}{2}}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} = 2.5758 \cdot \sqrt{\frac{0.2496}{1000}} \approx 0.041$ voor de schatting $\overline{p} = 0.52$ van de echte proportie van toestemming. Het betrouwbaarheidsinterval is dus $[47.9\%, 56.1\%]$.

Natuurlijk is de interessante vraag, of de toestemming boven de 50% ligt. Om hierover een uitspraak met betrouwbaarheid 99% te kunnen doen, moet de lengte van het betrouwbaarheidsinterval tot 4% worden beperkt. De benodigde grootte van de steekproef hiervoor is $n \geq z_{\frac{\alpha}{2}}^2 \frac{p(1-p)}{l^2} = 2.5758^2 \cdot \frac{0.25}{0.02^2} \approx 4147$. Hierbij hebben we voor p de schatting $p = 0.5$ ingevuld, voor $p = 0.52$ zouden we $n \geq 4140$ krijgen, dus bijna hetzelfde.

3.3 Betrouwbaarheidsintervallen bij onbekende variantie

We zijn er tot nu toe van uitgegaan dat we het met een normaal verdeelde stochast X met *bekende* variantie te maken hebben. Omdat dit in de praktijk niet realistisch is, kijken we nu naar het geval van een stochast met onbekende variantie. In dit geval hebben we helaas niets meer aan de stochast $Z := \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$, omdat we de variantie σ^2 gewoon niet kennen. Maar we weten wel, dat $S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ een zuivere schatter voor σ^2 is, dus kunnen we proberen de onbekende variantie σ^2 door de schatter S^2 te vervangen. Dit geeft de stochast

$$T := \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{(\overline{X} - \mu)\sqrt{n}}{S}$$

die we al in de laatste les zijn tegengekomen: Voor een normaal verdeelde stochast X heeft T de *Student-t verdeling met $n-1$ vrijheidsgraden.* We weten dat deze verdeling voor kleine n meer uitgespreid is dan de standaard-normale verdeling en voor grote n steeds meer op de standaard-normale verdeling lijkt.

Met dezelfde argumenten als in het geval van bekende variantie komen we nu weer naar betrouwbaarheidsintervallen, als we de standaard-normale verdeling altijd door de Student-t verdeling met $n-1$ vrijheidsgraden vervangen.

Analoog met de standaard-normale verdeling definiëren we de t-waarde $t_\alpha := t_{n-1,\alpha}$ van level $\alpha = 1 - \gamma$ door

$$P(T > t_\alpha) = \alpha$$

waarbij het aantal $n-1$ van vrijheidsgraden meestal niet aangeven wordt, omdat het uit de samenhang duidelijk is.

Een soortgelijke berekening als boven geeft:

$$\begin{aligned}
&P(-t_{\frac{\alpha}{2}} \leq T \leq t_{\frac{\alpha}{2}}) = \gamma \Leftrightarrow P(-t_{\frac{\alpha}{2}} \leq \frac{(\overline{X} - \mu)\sqrt{n}}{S} \leq t_{\frac{\alpha}{2}}) = \gamma \\
\Leftrightarrow\ &P(\mu - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \overline{X} \leq \mu + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}) = \gamma \\
\Leftrightarrow\ &P(\overline{X} - t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}) = \gamma.
\end{aligned}$$

Voor een steekproef $x_1, \ldots, x_n$ met steekproefgemiddelde $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ en steekproefstandaardafwijking $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ noemen we (net als bij de binomiale verdeling) de schatting $\frac{s}{\sqrt{n}}$ voor de standaardafwijking $\sqrt{Var(\overline{X})}$ van de schatter $\overline{X}$ de *standaard fout* van $\overline{x}$ en noteren dit met $SE(\overline{x})$. Hiermee krijgen we het betrouwbaarheidsinterval

$$\left[\overline{x} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \overline{x} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right] = \left[\overline{x} - t_{\frac{\alpha}{2}}SE(\overline{x}), \overline{x} + t_{\frac{\alpha}{2}}SE(\overline{x})\right]$$

van betrouwbaarheid γ voor μ.

Net zo als bij de standaard-normale verdeling worden de t-waarden voor de meest gebruikelijke levels van betrouwbaarheid en voor de verschillende vrijheidsgraden in tabellen opgeslagen. Inmiddels worden in plaats van tabellen meestal software pakketten gebruikt, die de t-waarden voor een gewenste betrouwbaarheid γ en een gegeven aantal van vrijheidsgraden uitrekenen. Typische waarden van $t_{n,\alpha}$ zijn in Tabel 2 te zien (waarbij we met $n = \infty$ de waarden voor de standaard-normale verdeling aangeven):

$n \backslash \alpha$	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
5	1.476	2.015	2.571	3.365	4.032
10	1.372	1.812	2.228	2.764	3.169
30	1.310	1.697	2.042	2.457	2.750
∞	1.282	1.645	1.960	2.326	2.576

Tabel 2: Kritieke waarden $t_{n,\alpha}$ voor de Student-t verdelingen met n vrijheidsgraden.

Voorbeeld: Men neemt aan dat het aantal lijnen die in een grote telefooncentrale tijdens het spitsuur in gebruik zijn normaal verdeeld is. Uit een steekproef over 11 dagen blijkt een steekproefgemiddelde van $\overline{x} = 120$ voor het aantal lijnen, met een steekproefstandaardafwijking van $s = 10$. Als we een betrouwbaarheidsinterval op level 99% voor het gemiddelde aantal μ van lijnen in gebruik willen bepalen, hebben we de t-waarde $t_{10,0.005}$ nodig, want $n = 11$ en $\alpha = 0.01$. In de tabel vinden we $t_{10,0.005} = 3.169$, dus is de afwijking $t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 3.169 \cdot \frac{10}{\sqrt{11}} \approx 9.6$ en we krijgen het betrouwbaarheidsinterval $[110.4, 129.6]$ voor μ.

3.4 Betrouwbaarheidsintervallen voor de variantie

We hebben in de vorige les aangegeven dat voor een standaard-normaal verdeelde stochast X de stochast

$$Y := \frac{n-1}{\sigma^2} S^2 = \sum_{i=1}^{n} (\frac{X_i - \overline{X}}{\sigma})^2$$

een χ^2-verdeling met $n-1$ vrijheidsgraden heeft. Deze stochast Y is nu geschikt om een betrouwbaarheidsinterval voor de variante aan te geven.

Analoog met de z-waarde voor de standaard-normale verdeling en de t-waarde voor de Student-t verdeling definiëren we de χ^2-waarde $\chi^2_\alpha := \chi^2_{n-1,\alpha}$ door

$$P(Y > \chi^2_\alpha) = \alpha$$

waarbij de index voor het aantal vrijheidsgraden weer weggelaten is.

Omdat de χ^2-verdeling niet symmetrisch is, kunnen we niet meer zo makkelijk uit χ^2_α een waarde χ^2_β afleiden zo dat $P(Y < \chi^2_\beta) = P(Y > \chi^2_\alpha) = \alpha$ is.

Maar uit $P(Y > \chi^2_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$ volgt dat tussen $\chi^2_{1-\frac{\alpha}{2}}$ en $\chi^2_{\frac{\alpha}{2}}$ de kansmassa $(1 - \frac{\alpha}{2}) - \frac{\alpha}{2} = 1 - \alpha = \gamma$ ligt.

> Bij symmetrische verdelingen zo als de normale verdeling laat zich aantonen dat de symmetrische betrouwbaarheidsintervallen de intervallen van minimale lengte voor een gegeven betrouwbaarheid zijn. De χ^2-verdeling is niet symmetrisch, en men kan voor het interval rond Y dat de kansmassa γ bevat ook een willekeurig interval van de vorm $[\chi^2_{\gamma+c}, \chi^2_c]$ kiezen. Zo'n interval heeft inderdaad niet voor $c = \frac{\alpha}{2}$ de minimale lengte, maar de waarde c waarvoor de lengte minimaal is ligt in de praktijk meestal zo dicht bij $\frac{\alpha}{2}$ dat men dit verwaarloost.

Met een analoge redenering als eerder krijgen we voor de stochast Y:

$$P(\chi^2_{1-\frac{\alpha}{2}} \leq Y \leq \chi^2_{\frac{\alpha}{2}}) = 1 - \alpha = \gamma \Leftrightarrow P(\chi^2_{1-\frac{\alpha}{2}} \leq \frac{n-1}{\sigma^2} S^2 \leq \chi^2_{\frac{\alpha}{2}}) = \gamma$$
$$\Leftrightarrow P(\chi^2_{1-\frac{\alpha}{2}} \frac{\sigma^2}{n-1} \leq S^2 \leq \mu + \chi^2_{\frac{\alpha}{2}} \frac{\sigma^2}{n-1}) = \gamma$$
$$\Leftrightarrow P(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}) = \gamma.$$

Voor een concrete steekproef $x_1, \ldots, x_n$ met steekproefvariantie s^2 krijgen we hieruit als betrouwbaarheidsinterval van betrouwbaarheid γ voor σ^2 het interval

$$\left[\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}\right].$$

We kunnen ook een betrouwbaarheidsinterval voor de standaardafwijking σ aangeven, want worteltrekken geeft

$$P\left(\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2}}}}\, S \leq \sigma \leq \sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}}}}\, S\right) = P\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}\right) = \gamma$$

en hieruit krijgen we het betrouwbaarheidsinterval

$$\left[\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2}}}}\, s, \sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2}}}}\, s\right]$$

van betrouwbaarheid γ voor de standaardafwijking σ.

Belangrijke begrippen in deze les

- betrouwbaarheid
- tweezijdige / éénzijdige intervalschatter
- betrouwbaarheidsintervallen

- z-waarde, t-waarde, χ^2-waarde
- standaard fout

Opgaven

14. Zij X een uniform verdeelde stochast op het interval $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$ en zij $x_1, \ldots, x_n$ een steekproef voor deze stochast. Laat zien dat $[\min(x_1, \ldots, x_n), \max(x_1, \ldots, x_n)]$ een betrouwbaarheidsinterval voor θ is (dus de realisatie van een intervalschatter) en bepaal de level γ van betrouwbaarheid van dit interval.

15. Bij het bedrijf *Bonanza Banana* heeft een steekproef van 225 aanvragen een gemiddelde verwerkingstijd van $\overline{x} = 7$ `jerks` opgeleverd. Uit langdurige ervaring is bekend dat de standaardafwijking voor de verwerkingstijd $\sigma = 3$ `jerks` bedraagt.

    (i) Bepaal een betrouwbaarheidsinterval voor de level 95% voor de gemiddelde verwerkingstijd.

    (ii) Hoe groot moet de steekproef minstens zijn om op level 95% een betrouwbaarheidsinterval van lengte hoogstens 0.5 `jerks` te hebben?

16. In een aselecte steekproef van 100 studenten geven 18 studenten aan dat ze bekend met de binomiale verdeling zijn.

    (i) Bepaal betrouwbaarheidsintervallen op de levels 90%, 95% en 99% voor het relatieve aantal $\overline{p}$ van studenten die de binomiale verdeling kennen.

    (ii) Hoe groot moet voor ieder van de drie levels uit (i) de steekproef zijn om de lengte van het betrouwbaarheidsinterval op hoogstens 0.05 te beperken?

17. Gegeven is een aselecte steekproef (12.05, 12.71, 12.25, 12.40, 12.15, 12.94, 12.00, 12.40, 12.49, 12.33, 12.37) van 11 waarnemingen van een normaal verdeelde stochast met onbekende verwachtingswaarde μ en (bekende) standaardafwijking $\sigma = 0.3$.

    (i) Bereken een betrouwbaarheidsinterval op level 95% voor μ.

    (ii) Bereken een linkséénzijdig betrouwbaarheidsinterval op level 90% voor μ.

    (iii) Vergelijk het betrouwbaarheidsinterval uit (i) met het betrouwbaarheidsinterval op level 95% bij onbekende standaardafwijking σ.

18. Een onderzoek naar het atoomgewicht van thallium leverde de volgende waarden op: 203.628, 203.636, 203.639, 203.644, 203.650, 203.666.

    (i) Bereken een betrouwbaarheidsinterval van level 95% voor het atoomgewicht.

    (ii) Hoeveel waarnemingen moeten er extra worden gedaan om op level 95% het atoomgewicht met een nauwkeurigheid van 0.002 te kunnen bepalen?

19. Iemand werpt 600 keer met een dobbelsteen en vindt 70 keer een `6`. Geef een betrouwbaarheidsinterval op level 95% voor de kans op een `6` bij deze dobbelsteen. Doe hetzelfde voor de levels 99% en 99.9%. Lijkt je dit een eerlijke dobbelsteen?