

TECHNISCHE UNIVERSITEIT DELFT

TECHNISCHE WISKUNDE
FACULTEIT ELEKTROTECHNIEK, WISKUNDE EN
INFORMATICA

Frequentistische en Bayesiaanse statistiek in de forensische statistiek

Jade Swanenburg (4565614)
Begeleider: Jakob Söhl

21 januari 2020

Samenvatting

In de statistiek zijn er verschillende methodes voor het uitvoeren van model selectie. Het verschil in deze methodes komt voort uit het verschil in stromingen. Voor niet-geneste model selectie zijn de meest gangbare stromingen de Bayes Factor en de likelihood ratio. D. M. Ommen en C. P. Saunders presenteerden theoretische resultaten voor de relatie tussen de Bayes Factor en de likelihood ratio, waardoor de resultaten van beide paradigma's met elkaar kunnen worden vergeleken [5]. De Bayes Factor wordt uitgedrukt in de verwachting van likelihoodratio functie met betrekking tot de posterior verdeling van de parameters. In de bewijzen van deze theoriën ontbraken een aantal belangrijke aspecten. Om deze resultaten volledig te kunnen bewijzen, wordt in dit verslag aangetoond dat eerdere theoretische resultaten moeten worden uitgebreid met extra aannames of volledig moeten worden aangepast. Aan de hand van deze aannames en aanpassingen worden de theoretische resultaten bewezen. De theoretische resultaten zijn belangrijk in de toepassing van niet-geneste model selectie, omdat er nu gecommuniceerd kan worden tussen experts vanuit een ander paradigma.

Inhoudsopgave

1	Inleiding Forensische Statistiek	1
1.1	Model selectie in de forensische statistiek	1
1.2	Model selectie kaders	1
1.2.1	Gemeenschappelijke-bron probleem	2
1.2.2	Specifieke-bron probleem	2
1.3	Concurrerende modellen	3
2	Wiskundige interpretatie van de model selectie kaders met bijbehorende likelihood functies	5
2.1	Algemene begrippen, verdelingen en likelihood functies voor de wiskundige interpretatie	5
2.2	Gemeenschappelijke-bron probleem	6
2.2.1	Probleemschets	6
2.2.2	Likelihood functie gemeenschappelijke-bron probleem .	10
2.3	Specifieke-bron probleem	11
2.3.1	Probleemschets	12
2.3.2	Likelihoodfunctie specifieke-bron probleem	14
2.4	Toepassingen buiten de forensische statistiek	15
3	Model selectie methodes	17
3.1	Model selectie	17
3.2	Frequentisme	17
3.2.1	Likelihood ratio functie	17
3.2.2	Ware likelihood ratio	18
3.3	Bayesiaanse modelselectie	18
4	Asymptotische resultaten van de relatie tussen de Bayes Fac- tor en de likelihood ratio functie	21
4.1	Van de likelihood ratio functie naar de Bayes Factor	21
4.2	Asymptotische statistiek	22
4.3	De Bernstein-von Mises Theorie	23
4.4	Convergentie van de Bayes Factor naar de ware likelihood ratio voor het gemeenschappelijke-bron probleem	24
4.5	Toegevoegde aannames voor Theorie 1	25
4.5.1	Verschil in de aannames	25
4.5.2	Theoriën voor het bewijs van Theorie 1	26
4.6	Convergentie van de Bayes Factor naar de ware likelihood ratio voor het specifieke-bron probleem	27

5	Asymptotische resultaten voor de parameter intervallen	28
5.1	Soorten parameterintervallen	28
5.1.1	Betrouwbaarheids interval	28
5.1.2	Geloofwaardigheids interval	29
5.2	Geloofwaardigheids interval voor de likelihood ratio	29
5.3	Verandering in aannames	31
5.3.1	Aanpassingen Lemma 3.2	31
5.3.2	Aanpassing Theorie 3	32
5.4	Gevolgen van Theorie 3	33
5.4.1	De vorm van het geloofwaardigheids interval	33
5.4.2	Nadelen van het geloofwaardigheids interval voor de likelihood ratio	33
5.4.3	Toepassing Theorie 3 in de praktijk	34
6	Conclusie	35
A	Bayes Factor afleidingen	36
A.1	Vergelijking (11)	36
A.2	Vergelijking (12)	37
A.3	Vergelijking (13)	38
A.4	Vergelijking (14)	39
B	Bewijs van Theorie 1	41
B.1	Convergentie van de Bayes Factor naar de likelihoodratio . . .	41
B.2	Toepassing van de Bernstein–von Mises Theorie	42
B.3	Toepassing van het Lemma van Portmanteau	45
B.3.1	Het Lemma van Portmanteau	45
B.3.2	De eerste voorwaarde van het Lemma van Portmanteau aantonen	45
B.3.3	Een conclusie uit de tweede voorwaarde van het Lemma van Portmanteau	51
B.4	Conclusie Theorie 1	52
C	Bewijs van Lemma 3.1	53
C.1	Convergentie van de restterm van de Taylorreeks	54
C.2	Bernstein–von Mises Theorie	55
C.3	Conclusie Lemma 3.1	56
D	Bewijs van Lemma 3.2	57
D.1	Convergentie van de Bayes Factor en de likelihood ratio functie	57
D.2	Convergentie van $\gamma_{n M_2}^2$ en σ_n^2	57

D.3	Conclusie Lemma 3.2	58
E	Bewijs van Theorie 3	60
E.1	Convergentie van σ_n	60
E.2	Conclusie Theorie 3	61
F	De aannames van de Theoriën en Lemma's	62
F.1	Aannames Theorie 1	62
F.2	Aannames Theorie 2	62
F.3	Aannames Theorie 3	62
F.4	Aannames Lemma 3.1	63
F.5	Aannames Lemma 3.2	63

Lijst van figuren

1	<i>Schematische weergave van subject x_{a_1} met bijbehorende samples in het gemeenschappelijke-bron model [5].</i>	8
2	<i>Schematische weergave van een gemeenschappelijke-bron databank [5].</i>	8
3	<i>Schematische weergave van model 1 in het gemeenschappelijke-bron probleem. [5]</i>	9
4	<i>Schematische weergave van model 2 in het gemeenschappelijke-bron probleem. [5]</i>	9
5	<i>Schematische weergave van het volledige gemeenschappelijke-bron probleem [5].</i>	10
6	<i>Schematische weergave van een specifieke-bron databank [5].</i>	12
7	<i>Schematische weergave van model 1 van het specifieke-bron probleem [5].</i>	13
8	<i>Schematische weergave van model 2 van het specifieke-bron probleem [5].</i>	14
9	<i>Schematische weergave van model 2 van het specifieke-bron probleem [5].</i>	15

1 Inleiding Forensische Statistiek

In de forensische statistiek worden statistische modellen gebruikt in strafrechtelijke zaken. Zo wordt er door middel van modelvergelijking gekeken naar een oplossing voor een forensisch probleem. In deze paragraaf wordt uitgelegd wat model selectie is en hoe dit wordt toegepast in de forensische statistiek.

1.1 Model selectie in de forensische statistiek

De statistiek die wordt gebruikt binnen strafrechtelijke zaken kan worden verdeeld in het oplossen van twee type problemen:

- Het gemeenschappelijke-bron probleem en
- Het specifieke-bron probleem

Wat het verschil tussen deze twee problemen precies is, staat beschreven in Paragraaf 1.2 verder in dit hoofdstuk. De *problemen* worden ook wel *kaders* genoemd. In dit verslag wordt met ‘probleem’ en ‘model selectie kader’ hetzelfde bedoeld.

Beide problemen kennen elk twee mogelijke uitkomsten: schuldig of onschuldig. Deze uitkomsten worden binnen de statistiek aangeduid met de term *modellen*. Om een uitspraak te doen over eventuele schuld of onschuld, wordt gekeken welk model het best wordt ondersteund door het bewijsmateriaal. Om deze vraag te beantwoorden, wordt er zogeheten *modelselectie* toegepast. In Paragraaf 1.3 verder in dit hoofdstuk wordt dieper ingegaan op de modellen en de modelselectie.

1.2 Model selectie kaders

In de forensische statistiek wordt onderscheid gemaakt tussen twee soorten problemen: gemeenschappelijke- en specifieke-bron problemen. Stel er wordt op de plaats-delict bewijsmateriaal gevonden. Dan hebben beide problemen als doel de bron van het bewijsmateriaal te identificeren. Het verschil zit in de identificatie. Bij het onderzoek met het gemeenschappelijke-bron probleem worden twee verschillende bewijsmaterialen met beide een onbekende bron met elkaar vergeleken. Bij het onderzoek met het specifieke-bron probleem wordt bewijsmateriaal van een onbekende bron vergeleken met bewijsmateriaal van een bekende bron.

In deze paragraaf zullen beide problemen worden uitgelegd en toegelicht aan de hand van een voorbeeld. De model selectie problemen uit dit verslag zijn

afkomstig uit het artikel ‘Reconciling the Bayes Factor and Likelihood Ratio for Two Non-Nested Model Selection Problems’ van Danica M. Ommen en Christopher P. Saunders [5].

Om de type problemen goed uit te kunnen leggen, worden hier eerst wat begrippen gedefinieerd. Wanneer er op de plaats-delict bewijsmateriaal wordt gevonden, heet dit bewijsmateriaal een *sample*. Meerdere samples die bij elkaar horen heten een *observatie*. De samples uit een observatie behoren tot dezelfde bron. Deze bron wordt het *subject* genoemd.

Om een observatie van het bewijsmateriaal te kunnen identificeren, wordt er een *databank* opgesteld. Een databank is een verzameling van verschillende observaties die bestaan uit meerdere samples. Zoals eerder gedefinieerd, hoort elke observatie bij een subject (de bron van de observatie). Alle subjects samen wordt een *populatie* genoemd.

1.2.1 Gemeenschappelijke-bron probleem

Bij het gemeenschappelijke-bron probleem wordt er onderzocht of twee observaties met een onbekende bron afstammen van hetzelfde onbekende subject. Dit onbekende subject maakt deel uit van een grote populatie. Bij dit probleem wordt er niet gekeken naar welk subject dit is, maar puur of beide observaties van ditzelfde subject afstammen en dus of beide observaties een gemeenschappelijke bron hebben.

Voorbeeld

In een stapel handgeschreven brieven wordt er gekeken of twee brieven door dezelfde persoon zijn geschreven. Er staat een databank met heel veel handgeschreven brieven ter beschikking. De subjects zijn dan alle brieven uit de populatie met als bron de persoon die ze heeft geschreven. Uit dit probleem komt dus alleen een antwoord op de vraag of ze door dezelfde persoon zijn geschreven. Over welke persoon dit is, wordt geen uitspraak gedaan.

1.2.2 Specifieke-bron probleem

In tegenstelling tot het gemeenschappelijke-bron probleem, wordt er hier gekeken of een observatie afstamt van een bekend (specifiek) subject. Er is hier net zoals bij het gemeenschappelijke-bron probleem sprake van maar één observatie. Bij het specifieke-bron probleem wordt er onderscheid gemaakt tussen twee populaties: de grote populatie net zoals in het gemeenschappelijke-bron probleem en een populatie die alleen bestaat uit het specifieke subject. Er wordt dus eigenlijk gekeken of de observatie afstamt van

Voorbeeld

Als er DNA materiaal wordt gevonden op de plaats-delict, wordt er gekeken of dit materiaal overeenkomt met het DNA van een verdachte. Als dit niet het geval is, zal het materiaal afstammen van een ander onbekend persoon uit een grote populatie. Er staat dus enerzijds een DNA-profiel van de verdachte, het specifieke subject, ter beschikking. Anderzijds is er een databank met DNA-profielen van andere personen. Uit dit probleem komt dus een antwoord op de vraag of het DNA afkomstig is van de verdachte die bekend is bij de politie.

1.3 Concurrerende modellen

Om een uitspraak te kunnen doen over de problemen, worden er twee concurrerende modellen opgesteld. De uitspraak wordt gebaseerd op de zogeheten model selectie. De model selectie wordt nader toegelicht in Hoofdstuk 3. De modellen zijn voor beide problemen verschillend. Deze modellen zijn voor beide problemen niet-genest. Dat betekent dat slechts één van de twee overeen kan komen met de werkelijkheid. Bij het toepassen van de forensische statistiek worden deze twee modellen tegen elkaar getest. Dat betekent dat er een conclusie wordt getrokken over welk model het best ondersteund wordt door het gevonden bewijs.

Gemeenschappelijke-bron probleem

Zoals hiervoor beschreven, wordt er gekeken of twee observaties van hetzelfde subject afkomstig zijn. Het eerste model, ook wel M_1 , beschrijft de situatie dat beide observaties inderdaad van hetzelfde onbekende subject afstammen. Het tweede model, M_2 , beschrijft de situatie dat beide observaties afkomstig zijn van twee verschillende subjects. Het is duidelijk dat als het ene model overeenkomt met de werkelijkheid, het andere model dit niet kan doen.

Specifieke-bron probleem

Er wordt hier gekeken of een observatie afkomstig is van een specifiek subject. Het eerste model, M_1 , beschrijft de situatie dat de observatie inderdaad afkomstig is van het specifieke subject. Het tweede model M_2 beschrijft de situatie dat de observatie afkomstig is van een ander subject uit een populatie. Deze populatie mag het specifieke subject niet bevatten. Ook deze twee modellen kunnen niet tegelijkertijd overeenkomen met de werkelijkheid.

Om op de problemen model selectie toe te passen, wordt eerst een wiskundige interpretatie van de modellen opgesteld. Deze wiskundige interpretatie wordt beschreven in het volgende hoofdstuk.

2 Wiskundige interpretatie van de model selectie kaders met bijbehorende likelihood functies

Om de model selectie kaders voor zowel het gemeenschappelijke-bron- als het specifieke-bron probleem wiskundig te beschrijven, wordt eerst een aantal begrippen gedefinieerd. Vanuit deze begrippen worden bepaalde parameter verdelingen beschreven. De wiskundige interpretatie van de modellen kan worden beschreven aan de hand van deze begrippen en verdelingen. De wiskundige interpretatie wordt in dit hoofdstuk voor het gemeenschappelijke-bron probleem uitgelegd in Paragraaf 2.2 en voor het specifieke-bron probleem in Paragraaf 2.3. De model selectie kaders en de wiskundige interpretatie in dit hoofdstuk zijn afkomstig uit het artikel van Ommen en Saunders [5].

2.1 Algemene begrippen, verdelingen en likelihood functies voor de wiskundige interpretatie

Een probleem begint bij een *populatie* die bestaat uit *subjects*. Een subject bestaat uit *samples* die allemaal afkomstig zijn van dezelfde bron. Een subject kan dus ook worden gezien als een bron van de samples in het subject. Een voorbeeld is de populatie van DNA-materiaal van heel veel verschillende personen. De subjects van deze populatie zijn alle verschillende personen. De samples van een subject zijn alle verzamelingen van DNA-materiaal die afkomstig zijn van dat specifieke subject.

De *databank* of *dataverzameling* is de verzameling van alle samples van alle subjects samen. Dit zijn dus alle stukjes DNA-materiaal samen.

Verdelingen

Een sample kan wiskundig worden weergegeven met een of meerdere stochastische variabelen. Wanneer er meerdere samples zijn, kan er worden gesproken van een verdeling op de samples. Deze verdeling bevat verdelingsparameters: de verdeling van de samples van één subject.

Een voorbeeld van een sample als een stochastische variabele is de lengte van een haar van een persoon. Elke haar van een persoon in een sample met als sample waarde de lengte van die haar. Wanneer de lengte van de haren van een persoon bekend zijn, kan er worden gesproken van een verdeling van de lengtes van de haren. Het zou kunnen zijn dat de lengte van de haren van een persoon normaal verdeeld is. De verdelingsparameters van de haren

(samples) van deze persoon (subject) zijn dan bijvoorbeeld het gemiddelde en de standaardafwijking. Deze parameters zijn specifiek voor een subject, elk subject heeft dus zijn eigen verdelingsparameters.

De verdelingsparameters van een subject kunnen ook gezien worden als stochastische variabelen. In een populatie zitten meerdere subjects. De parameters van de subjects geven dus een nieuwe verdeling: de verdeling van de subject parameters. Deze verdeling heeft opnieuw eigen verdelingsparameters. Deze parameters worden genoteerd als de vector θ .

Likelihood functies

Om de modellen zoals beschreven in Paragraaf 1.3 te kunnen vergelijken, moet de parameter θ bekend zijn. Echter, de verdelingen op zowel de samples als de subjectparameters zijn onbekend. Hierdoor is ook de waarde voor θ onbekend. Wel kan er een likelihood functie voor θ worden opgesteld.

Laat X een stochastische vector zijn met kansdichtheidfunctie $f(X|\theta)$, afhankelijk van de parameter $\theta \in \Theta$. Voor een vaste X geldt dat de functie

$$\mathcal{L}(\theta) = \mathcal{L}(\theta|X) = f(X|\theta)$$

de likelihoodfunctie is. De likelihood functie wordt gezien als een functie van $\theta \in \Theta$ waar Θ de parameterruimte is [2]. In het geval van een discrete kansverdeling, wordt f vervangen door p ; alle kansen met elkaar vermenigvuldigd. De likelihood functies worden gebruikt om uiteindelijk de modellen met elkaar te vergelijken.

2.2 Gemeenschappelijke-bron probleem

De opbouw van dit probleem bestaat uit één populatie die zowel een databank als beide modellen bevat. Zoals eerder beschreven, wordt er bij dit probleem onderzocht of twee observaties van een onbekende bron afstammen van hetzelfde subject. Met andere woorden, of twee verzamelingen van samples afkomstig zijn van hetzelfde subject. Dit subject is afkomstig uit de databank.

2.2.1 Probleemschets

De begrippen zijn hiervoor al uitgelegd, maar hier worden er nog notaties aan toegevoegd, zodat het overzichtelijk in een schets kan worden weergegeven.

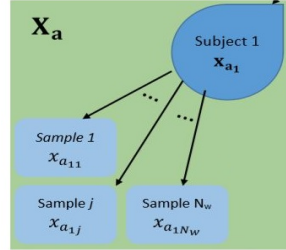
De begrippen sample, subject, databank, populatie en parameter worden als volgt genoteerd:

- \mathcal{X}_a , de populatie van subjects;
- a , de index van de subjects, $a \in A$;
- X_a , de databank, de verzameling van alle samples;
- x_{a_i} , subject i uit de populatie met $i = 1, \dots, N_a$;
- $x_{a_{ij}}$, sample j van subject i uit de databank met $j = 1, \dots, N_w$;
- x_b, x_c , de twee verzamelingen van samples met onbekende bron. Dit zijn de samples die worden gebruikt als bewijsmateriaal en deze zitten dus niet in de databank;
- x_{b_i} en x_{c_i} , de samples i uit de verzamelingen van samples met onbekende bron met $i = 1, \dots, N_b$ en respectievelijk $i = 1, \dots, N_c$;
- x_{a_u} , het subject waar onder M_1 de twee verzamelingen van samples met onbekende bron eventueel gezamenlijk vanaf stammen;
- x_{a_b} , het subject waar onder M_2 een van de twee verzamelingen van samples met onbekende bron eventueel vanaf stamt;
- x_{a_c} , het subject waar onder M_2 de ander van de twee verzamelingen van samples met onbekende bron eventueel vanaf stamt;
- $X_N = \{X_a, x_b, x_c\}$, de gehele data verzameling, met N de index voor de verschillende sample groottes.

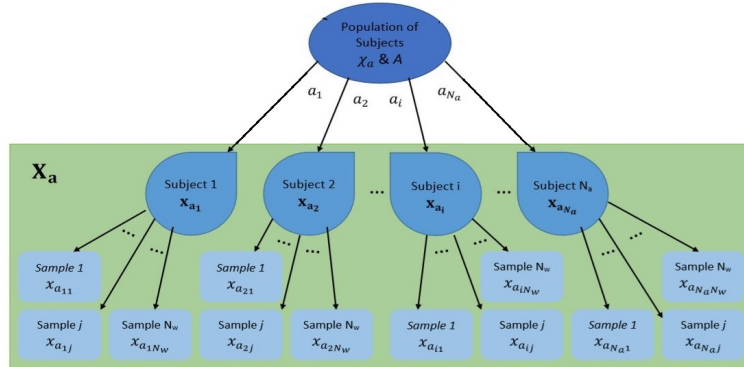
Dataverzameling

Voor het overzicht wordt het probleem schematisch weergegeven. Het probleem begint bij een verzameling samples die hoort bij een subject. Figuur 1 is een schematische weergave van subject 1 (genoteerd als x_{a_1}) met bijbehorende samples. Er zijn in totaal N_w samples per subject.

Zoals eerder beschreven, bestaat een populatie uit meerdere subjects met hun samples. De schematische weergave van de populatie en de databank staat in Figuur 2. Hier is duidelijk te zien dat de samples zijn geordend per subject.



Figuur 1: Schematische weergave van subject x_{a1} met bijbehorende samples in het gemeenschappelijke-bron model [5].

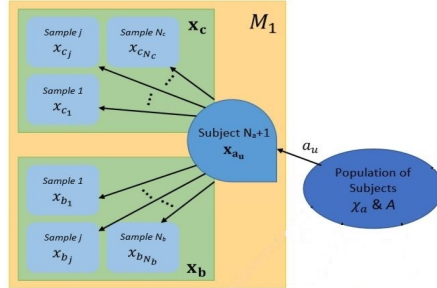


Figuur 2: Schematische weergave van een gemeenschappelijke-bron databank [5].

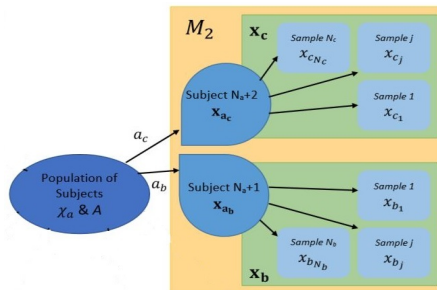
Concurrerende modellen

In Paragraaf 1.3 staan de twee concurrerende modellen beschreven. Het eerste model gaat ervanuit dat beide sample verzamelingen die dienen als bewijsmateriaal, afkomstig zijn van hetzelfde subject. Dit subject wordt genoteerd als x_{a_u} . De u staat voor ‘unknown’, omdat het niet bekend is om welk subject het gaat. Het onbekende subject x_{a_u} wordt ‘genummerd’ met $N_a + 1$, omdat het buiten de databank valt. In Figuur 3 staat model 1 schematisch weergegeven.

Het tweede model zoals in Paragraaf 1.3 gaat ervanuit dat beide sample verzamelingen afkomstig zijn van een ander subject. Deze subjects worden genoteerd met x_{a_b} en x_{a_c} . Omdat ze beide buiten de databank vallen en omdat model 1 en model 2 niet tegelijk ‘waar’ kunnen zijn, worden de subjects genummerd als $N_a + 1$ en $N_a + 2$. Figuur 4 is een schematische weergave van model 2 van het gemeenschappelijke-bron probleem.



Figuur 3: Schematische weergave van model 1 in het gemeenschappelijke-bron probleem. [5]



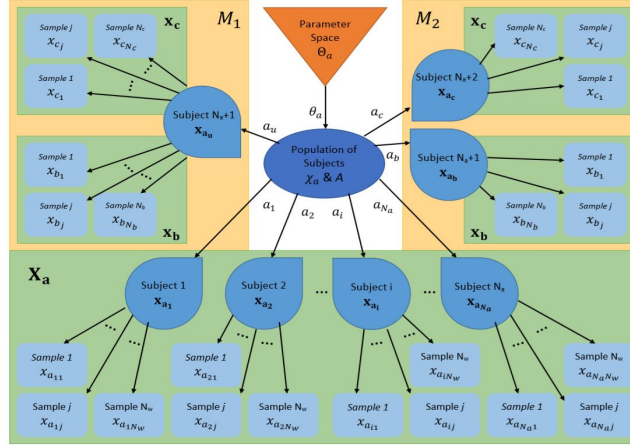
Figuur 4: Schematische weergave van model 2 in het gemeenschappelijke-bron probleem. [5]

Verdelingsparameters

Zoals beschreven aan het begin van dit hoofdstuk, spelen verdelingen een grote rol in het testen van de modellen tegenover elkaar. In het gemeenschappelijke-bron model zijn er meerdere verdelingen. De ‘eerste’ verdeling is de verdeling van de subjects in een populatie. Elk subject kan worden weergegeven met een random variabele. Dit kan ook een vector van meerdere random variabelen zijn. Alle random variabelen van de subjects in een populatie volgen een bepaalde verdeling. Deze verdeling wordt genoteerd met G_{θ_a} . Hier is θ_a de parameter(vector). Deze parameter is dus de grondlegging van de populatie. Θ_a is de parameterruimte van θ_a .

De random variabelen in de subjects zijn ook parameters van een kansverdelingsfunctie, namelijk de functie voor de verdeling van de samples in een subject. Elk sample kan worden weergegeven als een random variabele, of een vector hiervan.

Het hele gemeenschappelijke-bron probleem kan worden weergegeven in één schets, te zien in Figuur 5.



Figuur 5: Schematische weergave van het volledige gemeenschappelijke-bron probleem [5].

2.2.2 Likelihood functie gemeenschappelijke-bron probleem

Om de modelselectie toe te kunnen passen, kan voor beide modellen kan de likelihoodfunctie worden gegeven. Deze likelihood functie is voor beide modellen anders.

Model 1

De likelihood functie van model 1 is een functie van θ_a , de parameter van de verdeling op de populatie, dus de verdeling van (de parameters van) de subjects. Gegeven zijn beide verzamelingen van samples van het bewijsmateriaal, x_b en x_c , en het model dat wordt gebruikt, in dit geval M_1 .

Wanneer zowel het model als het bewijsmateriaal als gegeven worden beschouwd, dan kan de likelihood functie van model 1 gegeven worden door

$$\mathcal{L}(\theta_a | x_b, x_c, M_1) = f(x_b, x_c | \theta_a, M_1).$$

Het model test of beide sampleverzamelingen afkomstig zijn van een onbekend subject. Dit kan elk willekeurig subject zijn en dit moet dus worden getest voor elk subject. Daarom kan deze kansverdelingsfunctie ook worden geschreven als som van alle kansverdelingsfuncties per subject a_u waarop wordt getest. Omdat dit voor zowel continue als discrete kansverdelingen geldt, wordt hier de algemene vorm gebruikt. De kansverdelingsfunctie kan worden geschreven als integraal van alle kansverdelingsfuncties gegeven a_u geïntegreerd over de parameters van a_u : $G_{\theta_a}(a_u)$. Er wordt dus voor elk subject uit de databank gekeken of dit model ‘het best passend’ is. Voor

model 1 wordt de likelihood functie voor de onbekende bron observaties dus gegeven door

$$\mathcal{L}(\theta_a | \mathbf{x}_b, \mathbf{x}_c, M_1) = f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1) = \int f(\mathbf{x}_b, \mathbf{x}_c | a_u, \theta_a) dG_{\theta_a}(a_u). \quad (1)$$

De likelihood functie zal in een later hoofdstuk worden gebruikt voor de model selectie.

Model 2

De likelihood functie van model 2 is ook een functie van de parameter θ_a . In dit model wordt aangenomen dat beide sampleverzamelingen afkomstig zijn van een ander subject en dus een andere bron hebben. Hieruit volgt dat de sample verdelingen en dus ook de parameters van de subjects onafhankelijk zijn van elkaar. De likelihoodfuncties kunnen dus met elkaar worden vermenigvuldigd

$$\mathcal{L}(\theta_a | \mathbf{x}_b, \mathbf{x}_c, M_2) = \mathcal{L}(\theta_a | \mathbf{x}_b, M_2) \mathcal{L}(\theta_a | \mathbf{x}_c, M_2).$$

Ook hier zijn de likelihood functies voor een bepaalde waarde van θ_a te schrijven als de kansverdelingsfuncties van \mathbf{x}_b en \mathbf{x}_c . Net zoals bij het vorige model, wordt er gekeken per subject. Het verschil is dat hier niet één subject a_u wordt gebruikt, maar twee verschillende subjects a_b en a_c . Er wordt dan ook over de parameters van deze subjects geïntegreerd. De likelihood functie van het tweede model wordt gegeven door

$$\begin{aligned} \mathcal{L}(\theta_a | \mathbf{x}_b, \mathbf{x}_c, M_2) &= \mathcal{L}(\theta_a | \mathbf{x}_b, M_2) \mathcal{L}(\theta_a | \mathbf{x}_c, M_2) = f(\mathbf{x}_b | \theta_a) f(\mathbf{x}_c | \theta_a) \\ &= \int f(\mathbf{x}_b | a_b, \theta_a) dG_{\theta_a}(a_b) \int f(\mathbf{x}_c | a_c, \theta_a) dG_{\theta_a}(a_c). \end{aligned} \quad (2)$$

De likelihood functies zoals in Vergelijking (1) en (2) worden in een volgend hoofdstuk gebruikt om de model selectie uit te voeren.

2.3 Specifieke-bron probleem

De opbouw van dit probleem bestaat uit een populatie met een databank én een specifiek subject uit een andere populatie. Zoals eerder beschreven wordt er bij dit probleem onderzocht of de verzameling samples uit het bewijsmateriaal afstamt van het specifieke subject of van een willekeurig subject uit de populatie waaruit de databank is opgebouwd.

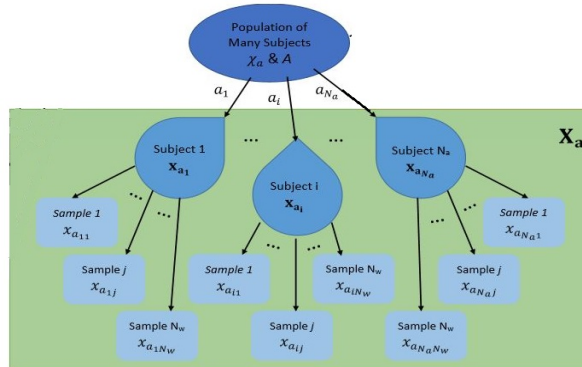
2.3.1 Probleemschets

De notaties voor de populatie, de index, de databank, de subjects en de samples zijn hetzelfde. De andere begrippen worden in het specifieke probleem genoteerd als:

- \mathcal{X}_b , de populatie die bestaat uit alleen het specifieke subject;
- x_b , de verzameling van samples van het specifieke subject;
- x_{b_i} , sample i van het specifieke subject, met $i = 1, \dots, N_b$;
- x_c , de verzameling van samples uit het bewijsmateriaal;
- x_{c_i} , sample i van het bewijsmateriaal, met $i = 1, \dots, N_c$;
- x_{a_c} , het subject uit de databank waar het bewijsmateriaal mogelijk vanaf stamt;
- $X_N = \{X_a, x_b, x_c\}$, de gehele data verzameling, met N de index voor de verschillende sample groottes.

Dataverzameling

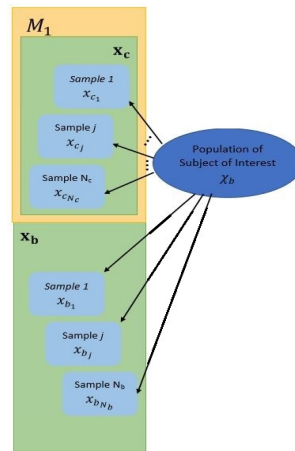
De opbouw van de populatie waaruit de databank is opgebouwd, is hetzelfde als de opbouw van het gemeenschappelijke-bron probleem. De schematische weergave van de populatie en de databank staat in Figuur 6.



Figuur 6: Schematische weergave van een specifieke-bron databank [5].

Concurrerende modellen

Het eerste model gaat er vanuit dat de samples van het bewijsmateriaal afkomstig zijn van een specifiek subject dat niet voorkomt in de populatie zoals hiervoor beschreven. Hiervoor is een tweede populatie \mathcal{X}_b met daarin slechts één subject: x_b . In Figuur 7 is te zien hoe deze populatie schematisch wordt weergegeven. Hierin is ook model 1 geschematiseerd. De sampleverzameling x_c is hier dus afkomstig van hetzelfde subject en dus van dezelfde populatie \mathcal{X}_b .



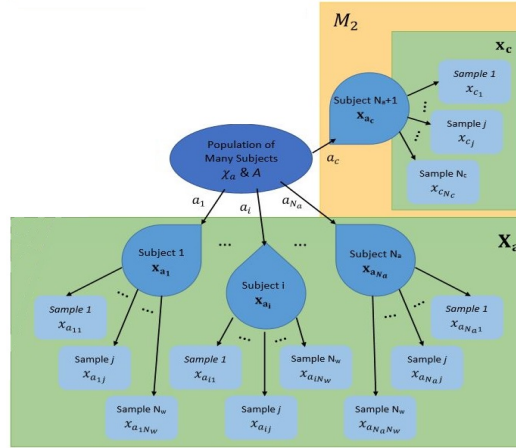
Figuur 7: Schematische weergave van model 1 van het specifieke-bron probleem [5].

Het tweede model gaat er vanuit dat de samples afkomstig zijn van een subject uit de populatie waaruit ook de databank is opgebouwd. Dit subject wordt genummerd als $N_a + 1$, omdat het niet in de databank zelf zit. Het subject wordt genoteerd als x_{ac} . Figuur 8 is een schematische weergave van model 2 van het specifieke-bron probleem.

Verdelingsparameters

In dit probleem zijn er twee populaties. Beide populaties hebben hun eigen verdeling van de subject parameters en dus ook hun eigen parameter(vector). Voor populatie \mathcal{X}_a en \mathcal{X}_b zijn de parameters respectievelijk θ_a en θ_b uit parameter ruimte Θ_a en Θ_b . De verdelingen van de subject parameters uit de populaties worden gegeven door G_{θ_a} en G_{θ_b} . Figuur 9 is een schematische weergave van het volledige specifieke-bron probleem.

De populatie van model 1 bestaat slechts uit één subject. Dit betekent dat de parameter θ_b direct de parameter is voor de verdeling van de samples van het



Figuur 8: Schematische weergave van model 2 van het specifieke-bron probleem [5].

specifieke subject. Er is hier dus ‘een verdeling minder’ dan bij het tweede model.

2.3.2 Likelihoodfunctie specifieke-bron probleem

Net zoals bij het gemeenschappelijke-bron probleem, worden model 1 en model 2 tegen elkaar getest om model selectie toe te passen. Hiervoor zijn beide likelihood functies nodig.

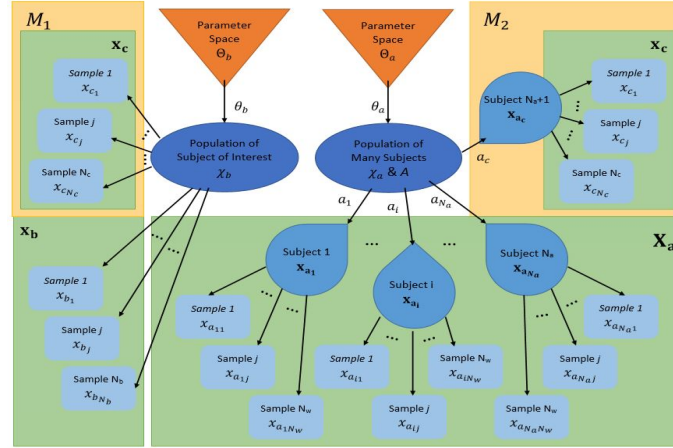
Model 1

De likelihood functie van model 1 is een functie van θ_b , de parameter op de populatie van het specifieke subject. Deze likelihood functie kan per definitie worden geschreven als kansverdelingsfunctie,

$$\mathcal{L}(\theta_b | x_c, M_1) = f(x_c | \theta_b, M_1).$$

De populatie van model 1 bestaat uit een enkel subject. Zoals beschreven aan het einde van Paragraaf 2.3.1, is θ_b de parameter die direct op de verdeling van de samples van toepassing is. De kansverdelingsfunctie van x_c is dan ook het product van de kansverdelingsfuncties van de samples van x_c , mits deze onafhankelijk zijn van elkaar.

$$\mathcal{L}(\theta_b | x_c, M_1) = f(x_c | \theta_b, M_1) = \prod_{i=1}^{N_c} f(x_{c_i} | \theta_b) \quad (3)$$



Figuur 9: Schematische weergave van model 2 van het specifieke-bron probleem [5].

Model 2

De likelihood functie van model 2 is een functie van θ_a , de parameter van de populatie waaruit de databank is opgebouwd. Net zoals in model 1 van het gemeenschappelijke-bron probleem, wordt hier per subject uit de databank gekeken of het waarschijnlijk is dat de samples afkomstig zijn van dat subject. De waarschijnlijkheid dat de samples afkomstig zijn van één van de subjects uit de databank, is dan de integraal van de kansverdelingsfuncties per subject. Er wordt geïntegreerd over de parameters van het subject waar naar wordt gekeken (a_c), gegeven θ_a . De likelihood functie van model 2 van het specifieke-bron probleem wordt gegeven door

$$\mathcal{L}(\theta_a|x_c, M_2) = f(x_c|\theta_a, M_2) = \int f(x_c|a_c, \theta_a)dG_{\theta_a}(a_c). \quad (4)$$

2.4 Toepassingen buiten de forensische statistiek

In dit verslag wordt de niet-geneste model selectie gespecificeerd op de forensische statistiek. Echter, er zijn buiten de forensische statistiek nog talloze vakgebieden waar de methode, beschreven in de voorgaande paragrafen, toegepast kunnen worden. Wanneer er sprake is van het herkennen van statistische patronen, kunnen model selectie kaders worden toegepast. Toch zit er wel een belangrijk verschil in de toepassing van model selectie in de forensische statistiek. Dit verschil komt voort uit de volgorde van het verkrijgen van de data.

In forensische toepassingen is de onbekende bron data vaak afkomstig van de plaats delict. Pas daarna zal er data worden verzameld van eventuele andere bronnen. Onderzoekers zullen geen samples verzamelen van verdachten voordat er een misdaad is gepleegd. Dit betekent dat in de forensische toepassingen, eerst de onbekende-bron data wordt verzameld en pas daarna de bekende-bron data. In traditionele toepassingen is dit in het algemeen andersom. De bekende-bron data zal als eerst worden verzameld. Pas daarna wordt de onbekende-bron data geanalyseerd [5]. In de resultaten van dit verslag is er vanuit gegaan de de onbekende-bron observaties als eerst worden geobserveerd.

3 Model selectie methodes

3.1 Model selectie

Model selectie is gericht op het vinden van het best-passende model vanuit een gegeven verzameling van modellen [2]. In het kader van de modellen die zijn geschetst in dit verslag, gaat het om het vergelijken van twee concurrerende modellen, M_1 en M_2 . Het vergelijken van deze modellen gebeurt aan de hand van de basisvraag:

Welk model wordt beter ondersteund door de data?

Er zijn verschillende methodes om de twee modellen van een probleem met elkaar te vergelijken. De methodes laten zich opsplitsen in de Bayesiaanse methode en de frequentistische methode. Waar de Bayesiaanse methode gebaseerd is op een stochastische parametervector, gaat de frequentistische methode uit van een vaste ware parameter. De toepassing van de frequentistische modelselectie is uitgewerkt in Paragraaf 3.2 en de toepassing van de Bayesiaanse model selectie is uitgewerkt in Paragraaf 3.3.

3.2 Frequentisme

In de frequentistische statistiek wordt aangenomen dat er ware sampleverdelingen bestaan die worden geïmpliceerd door de modellen. Voor elke ware sampleverdeling behorend bij een model bestaat er een ‘ware parameter’ θ . Deze ware parameter heeft een vaste waarde. Deze waarde is onbekend. De ware parameter behorend bij een verdeling onder een model wordt genoteerd met θ_{a_0} of θ_{b_0} .

Bij de frequentistische modelselectie wordt er gebruik gemaakt van de likelihood ratio functie. Dit is de ratio tussen de twee likelihood functies van de twee concurrerende modellen. Omdat er wordt aangenomen dat er een ware verdeling bestaat onder elk model, geldt dat er een ‘ware likelihood ratio’ bestaat [5].

3.2.1 Likelihood ratio functie

De algemene vergelijking voor de likelihood ratio functie is

$$\lambda(\theta|X) = \frac{\mathcal{L}(\theta|X, M_1)}{\mathcal{L}(\theta|X, M_2)}.$$

De likelihood ratio functie voor het gemeenschappelijke-bron model is een functie van de multivariate parameter vector $\theta_a \in \Theta_a$. De likelihood ratio

functie van het gemeenschappelijke-bron model wordt gegeven door

$$\lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) = \frac{\mathcal{L}(\theta_a | \mathbf{x}_b, \mathbf{x}_c, M_1)}{\mathcal{L}(\theta_a | \mathbf{x}_b, M_2) \mathcal{L}(\theta_a | \mathbf{x}_c, M_2)}. \quad (5)$$

De likelihood ratio functie van het specifieke-bron model is een multivariate functie van de gemeenschappelijke parametervector van θ_a en θ_b , met de onbekende-bron observatie gegeven. De likelihood ratio functie van het specifieke-bron probleem wordt gegeven door

$$\lambda_{ss}(\theta_a, \theta_b | \mathbf{x}_c) = \frac{\mathcal{L}(\theta_b | \mathbf{x}_c, M_1)}{\mathcal{L}(\theta_a | \mathbf{x}_c, M_2)}. \quad (6)$$

3.2.2 Ware likelihood ratio

Zoals beschreven in de inleiding van dit hoofdstuk, wordt er bij het frequentisme uitgegaan van een ware parameter behorend bij een ware sampleverdeling gegenereerd onder een van de modellen. De likelihood ratio functie die hoort bij de waarden van de parameters die worden geassocieerd met de ware sample verdeling geïmpliceerd door model 1 en model 2 kan dus ook worden gezien als de ‘ware likelihood ratio’. Omdat de ware parameter wel is vastgesteld, maar onbekend is, geldt dat de ware likelihood ratio ook wel is vastgesteld, maar niet bekend is.

De ware likelihood voor het gemeenschappelijke-bron model wordt gegeven door

$$\lambda_{cs}(\theta_{a_0} | \mathbf{x}_b, \mathbf{x}_c) = \frac{\mathcal{L}(\theta_{a_0} | \mathbf{x}_b, \mathbf{x}_c, M_1)}{\mathcal{L}(\theta_{a_0} | \mathbf{x}_b, M_2) \mathcal{L}(\theta_{a_0} | \mathbf{x}_c, M_2)}. \quad (7)$$

En de ware likelihood ratio voor het specifieke-bron probleem wordt gegeven door

$$\lambda_{ss}(\theta_{a_0}, \theta_{b_0} | \mathbf{x}_c) = \frac{\mathcal{L}(\theta_{b_0} | \mathbf{x}_c, M_1)}{\mathcal{L}(\theta_{a_0} | \mathbf{x}_c, M_2)}. \quad (8)$$

Bij het bepalen van de ware likelihood wordt het aantal onbekende bron observaties vastgesteld. De ware likelihood ratio neemt waarden aan in het interval $(0, \infty)$.

3.3 Bayesiaanse modelselectie

De Bayesiaanse modelselectie is gebaseerd op de Bayes Factor uit de Theorie van Bayes. De Bayes Factor is een relatieve maat voor hoezeer de data een

van de twee modellen de voorkeur over het andere model geeft. Volgens de Theorie van Bayes verhouden de modellen zich volgens [5]:

$$\frac{\mathbb{P}(M_1|X)}{\mathbb{P}(M_2|X)} = \frac{\mathbb{P}(X|M_1)}{\mathbb{P}(X|M_2)} \times \frac{\mathbb{P}(M_1)}{\mathbb{P}(M_2)},$$

Posterior Kansen = Bayes Factor \times Prior Kansen.

Hierin geeft posterior kans de verhouding van de waarschijnlijkheid van de modellen, gegeven het bewijsmateriaal en de prior kansen. De prior kansen geeft de verhouding van de modellen ongeacht het bewijsmateriaal. De posterior kansen zullen worden gebruikt om een van de modellen te kiezen. Een posterior kans van groter dan 1 wijst erop dat model 1 verkozen wordt boven model 2. Een posterior kans kleiner dan 1 wijst erop dat model 2 verkozen wordt. Een posterior kans dicht bij 1 wijst erop dat er geen onderscheid kan worden gemaakt tussen de twee concurrerende modellen.

De prior kansen in de Bayesiaanse modelselectie worden bepaald door de rechtbank. In de forensische statistiek wordt aan de hand van wiskundige theoriën de Bayes Factor bepaald. In dit verslag gaat het om modellen die parameter verdelingen bevatten. In de volgende paragraaf wordt uitgelegd hoe de Bayes Factor van een model met een parametrische verdeling kan worden bepaald.

Bayes Factor in parametrische verdelingen

In de Bayesiaanse statistiek is er geen sprake van een ware parameter zoals in de frequentistische statistiek. De parameterwaarde heeft een bepaalde onzekerheid. Deze onzekerheid kan worden uitgedrukt in een verdeling [2]. In deze verdeling wordt de parameterwaarde gezien als een stochastische vector $\theta \in \Theta$. In de Bayesiaanse methode wordt er onderscheid gemaakt tussen twee verdelingen van θ ,

- de prior verdeling: $\Pi(\theta)$, de verdeling van de parametervector zonder dat de observaties worden gebruikt;
- de posterior verdeling: $\Pi(\theta|X, M)$, de verdeling van de parametervector waarin de observaties onder een bepaald model zijn meegenomen.

In de situatie dat de modellen parameter verdelingen bevatten, wordt de Bayes Factor gegeven door

$$\beta(X) = \frac{\int f(X|\theta, M_1)d\Pi(\theta|M_1)}{\int f(X|\theta, M_2)d\Pi(\theta|M_2)}.$$

Voor beide concurrerende modellen kan de Bayes Factor worden afgeleid.

De Bayes Factor voor het gemeenschappelijke-bron probleem

Voor het gemeenschappelijke-bron probleem wordt aangenomen dat de prior verdeling $\Pi(\theta_a)$ gelijk is onder beide modellen. In het vervolg wordt het ‘ \equiv ’-teken gebruikt om een definitie aan te geven. De Bayes Factor van het gemeenschappelijke-bron probleem wordt gegeven door

$$\begin{aligned}\beta_{cs}(X_N) &= \frac{\int f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1) d\Pi(\theta_a | X_a)}{\int f(\mathbf{x}_b | \theta_a, M_2) f(\mathbf{x}_c | \theta_a, M_2) d\Pi(\theta_a | X_a)} \\ &\equiv \frac{m(\mathbf{x}_b, \mathbf{x}_c | X_a, M_1)}{m(\mathbf{x}_b, \mathbf{x}_c | X_a, M_2)}.\end{aligned}\quad (9)$$

De functie $m(\mathbf{x}_b, \mathbf{x}_c | X_a, M_2)$ is dus gedefinieerd zoals gegeven in Vergelijking (9).

De Bayes Factor voor het specifieke-bron probleem

Voor het specifieke-bron probleem wordt aangenomen dat de prior verdeling $\Pi(\theta_a)$ statistisch onafhankelijk is van de prior verdeling van $\Pi(\theta_b)$. De Bayes Factor van het specifieke-bron probleem wordt gegeven door

$$\begin{aligned}\beta_{ss}(X_N) &= \frac{\int f(\mathbf{x}_c | \theta_b, M_1) d\Pi(\theta_b | \mathbf{x}_b)}{\int f(\mathbf{x}_c | \theta_a, M_2) d\Pi(\theta_a | X_a)} \\ &\equiv \frac{m(\mathbf{x}_c | \mathbf{x}_b, M_1)}{m(\mathbf{x}_c | X_a, M_2)}.\end{aligned}\quad (10)$$

In deze paragraaf is beschreven hoe er tussen de twee modellen kan worden gekozen aan de hand van de basisvraag:

Welk model wordt beter ondersteund door de data?

Het beantwoorden van deze vraag kan op twee verschillende manieren: via de frequentistische methode aan de hand van de likelihood ratio met een ‘ware parameter’ óf via de Bayesiaanse methode aan de hand van de verdeling op de parameter. In het volgende hoofdstuk zal worden beschreven wat de relatie tussen deze twee methodes is.

4 Asymptotische resultaten van de relatie tussen de Bayes Factor en de likelihood ratio functie

Wanneer twee experts vanuit twee verschillende paradigma's, frequentistisch en Bayesiaans, met elkaar willen communiceren over de uitkomst van hun model selectie methode, is het nodig om de resultaten van de ene methode uit te drukken in een resultaat in de vorm van de andere methode. Het blijkt dat wanneer het aantal observaties blijft groeien, de resultaten aan elkaar gerelateerd kunnen worden. In dit hoofdstuk wordt beschreven hoe de Bayes Factor zich gedraagt wanneer het aantal observaties tegen oneindig gaat. Het resultaat van dit gedrag is beschreven aan de hand van de likelihood ratio functie in Paragraaf 4.4. Het bewijs van deze theorie is gebaseerd op de asymptotische statistiek. De belangrijkste theorie die wordt gebruikt in dit bewijs, is de Bernstein–von Mises Theorie. Deze theorie en haar toepassing in het bewijs worden uitgelegd in Paragraaf 4.3. De resultaten uit dit hoofdstuk zijn grotendeels gebaseerd op het artikel van Ommen en Saunders [5]. De aanpassingen ten opzichte van de resultaten uit het artikel van Ommen en Saunders staan beschreven in de laatste paragraaf van dit hoofdstuk.

4.1 Van de likelihood ratio functie naar de Bayes Factor

Om de relatie tussen de Bayes Factor en de likelihood ratio functie te kunnen beschrijven, is er een uitdrukking nodig waar beide aspecten in voorkomen. Het blijkt dat de Bayes Factor voor het model selectie probleem kan worden uitgedrukt als de verwachting van de likelihood ratio functie met betrekking tot de Bayesiaanse posterior verdeling van de parameters van de sample verdelingen. Hier wordt de gehele verzameling van data gegenereerd onder het tweede model. Voor zowel het gemeenschappelijke- als het specifieke-bron probleem is deze uitdrukking geformaliseerd.

In dit verslag wordt de maximum likelihood schatter, in het vervolg MLE, gebruikt als consistente schatter voor de parameter θ . De MLE wordt genoteerd als $\hat{\theta}$. De MLE voor $\theta_{a|M_2}$ wordt dus genoteerd als $\hat{\theta}_{a|M_2}$.

De Bayes Factor onder het tweede model

De Bayes Factor van het gemeenschappelijke-bron probleem onder het tweede model wordt gegeven door

$$\beta_{cs}(X_N) = \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d\Pi(\theta_a | X_N, M_2). \quad (11)$$

De Bayes Factor van het specifieke-bron probleem onder het tweede model wordt gegeven door

$$\beta_{ss}(X_N) = \int \lambda_{ss}(\theta_a, \theta_b | \mathbf{x}_c) d\Pi(\theta_a, \theta_b | X_N, M_2). \quad (12)$$

De afleidingen van Vergelijking (11) en Vergelijking (12) staan in Appendix A.1 respectievelijk A.2.

De Bayes Factor onder het eerste model

Een soortgelijke uitdrukking kan worden afgeleid wanneer posterior verdeling gebruikt maakt van de data onder het eerste model. De Bayes Factor van het gemeenschappelijke-bron probleem onder het eerste model wordt gegeven door

$$\beta_{cs}(X_N) = \left[\int \frac{1}{\lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c)} d\Pi(\theta_a | X_N, M_1) \right]^{-1}. \quad (13)$$

De Bayes Factor van het specifieke-bron probleem onder het eerste model wordt gegeven door

$$\beta_{ss}(X_N) = \left[\frac{1}{\lambda_{ss}(\theta_a, \theta_b | \mathbf{x}_c)} d\Pi(\theta_a, \theta_b | X_N, M_1) \right]^{-1}. \quad (14)$$

De afleidingen van Vergelijking (13) en Vergelijking (14) staan in Appendix A.3 respectievelijk Appendix A.4.

4.2 Asymptotische statistiek

Asymptotische statistiek is gebaseerd op het principe dat het aantal van observaties steeds maar groter wordt. Er wordt aangenomen dat er oneindig lang data kan worden verzameld. Het wordt gebruikt om eigenschappen van schatters en statistieke testen te bepalen, die niet kunnen worden bepaald met dataverzamelingen van eindige grootte. Het kan ook zijn dat bepaalde verdelingen van statistieken analytisch niet handelbaar zijn of dat statistieke problemen geen optimale oplossing hebben [8].

Laat n_a de index zijn dat het stijgend aantal subjects in dataverzameling X_{a,n_a} aangeeft. Neem ook aan dat elk subject in de dataverzameling een vast aantal elementen heeft. Met asymptotisch wordt bedoeld dat n_a naar oneindig gaat, dus $n_a \rightarrow \infty$.

Convergentie

Binnen de asymptotische statistiek wordt er onderscheid gemaakt tussen verschillende soorten convergentie. Een van de vormen van convergentie is zwakke convergentie [8]. Een rij van stochastische vectoren X_n convergeert zwak naar een stochastische vector X wanneer

$$\mathbb{P}(X_n \leq x) \rightarrow \mathbb{P}(X \leq x),$$

voor elke x waarin de limiet van de verdelingsfunctie $x \mapsto P(X \leq x)$ continu is. Deze zwakke convergentie wordt genoteerd als $X_n \rightsquigarrow X$.

Een andere vorm van convergentie is convergentie in kans. Een rij van stochastische vectoren X_n convergeert in kans naar X wanneer $\forall \epsilon > 0$ geldt

$$\mathbb{P}(d(X_n, X) > \epsilon) \rightarrow 0.$$

De convergentie in kans wordt genoteerd als $X_n \xrightarrow{P} X$.

4.3 De Bernstein–von Mises Theorie

Zoals beschreven in Paragraaf 3.3, is er bij de Bayesiaanse methode geen sprake van een vaste ware parameter, maar is de parameter θ een stochastische variabele die verdeeld is volgens een prior of posterior verdeling Π . De posterior verdeling ($\Pi(\theta|X, M)$) van θ is afhankelijk van de gedane observaties en het model waaronder de dataverzameling is gegenereerd. Wanneer het aantal van observaties tegen oneindig gaat, resulteert dit in een interessant resultaat voor de posterior verdeling. Het blijkt dat een Bayesiaanse schatter asymptotisch normaal verdeeld is. Dat betekent dat asymptotisch gezien, de posterior verdeling convergeert naar een normaal verdeling. Verder blijkt ook dat de afstand tussen een Bayesiaanse schatter en een maximum likelihood schatter verdwijnt wanneer het aantal observaties tegen oneindig gaat.

Een belangrijke theorie die dit uitlegt is de Bernstein–von Mises Theorie [8]. Deze theorie zegt dat de posterior verdeling van θ asymptotisch normaal verdeeld is, gecentreerd in de maximum likelihood schatter. Formeel gezegd convergeert de totale variantie afstand tussen de posterior verdeling van θ en

een zekere normaal verdeling naar nul in kans wanneer het aantal observaties tegen oneindig gaat. De normaal verdeling heeft als parameters

- $\hat{\theta}_n$, de maximum likelihood schatter voor θ , als gemiddelde;
- $\frac{1}{n}I_\theta^{-1}$, waarin I_θ^{-1} de inverse van de Fisher informatie matrix is, als variantie.

De aannames van de Bernstein–von Mises Theorie gaan buiten de stof in dit verslag en worden gegeven in het boek *Asymptotic Statistics* van A.W. Van der Vaart [9]. Het resultaat van de Bernstein–von Mises Theorie wordt gebruikt in de bewijzen van Theorie 1 en Theorie 2 in dit verslag.

4.4 Convergentie van de Bayes Factor naar de ware likelihood ratio voor het gemeenschappelijke-bron probleem

Wanneer het mogelijk is om het aantal observaties tegen oneindig te laten gaan, kan er worden aangetoond dat onder bepaalde omstandigheden de Bayes Factor convergeert naar de ware likelihood ratio. Deze convergentie is geformaliseerd in Theorie 1 verder in dit verslag.

Theorie 1 (Gemeenschappelijke-bron Bayes Factor consistentie)

Gegeven een vaste observatie van x_b en x_c , stel dat de likelihood ratio functie $\lambda_{cs}(\theta_a|x_b, x_c)$ begrensd is in een omgeving van θ_{a_0} en dat $\hat{\theta}_{a|M_2}$ een consistente schatter is voor θ_{a_0} . Stel verder dat de aannames van de Bernstein–von Mises Theorie voldoen. Dan, als $n_a \rightarrow \infty$, convergeert de Bayes Factor in $P_{\theta_a}^{n_a}$ -waarschijnlijkheid naar de ware likelihood ratio,

$$\beta_{cs}(X_{a,n_a}, x_b, x_c) \xrightarrow{P_{\theta_a}^{n_a}} \lambda_{cs}(\theta_{a_0}|x_b, x_c).$$

Ter verduidelijking staat in Appendix F.1 een lijst met alle aannames die gelden voor Theorie 1.

Het bewijs van Theorie 1 begint bij het stellen van een bovengrens voor de afstand tussen de Bayes Factor en de likelihood. In deze bovengrens wordt de normaal verdeling verwerkt zodat later de Bernstein–von Mises Theorie kan worden toegepast. Een deel van de bovengrens convergeert naar nul in kans wegens de Bernstein–von Mises Theorie. Het andere deel van de bovengrens convergeert naar nul wegens de zwakke convergentie van de normaal verdeling naar de discrete Delta-Dirac verdeling. Het precieze bewijs van Theorie 1 staat in de appendix in Appendix B. In dit bewijs wordt de Bernstein–von

Mises Theorie toegepast. Deze theorie is uitgelegd in Paragraaf 4.3.

Theorie 1 is gebaseerd op het artikel van Ommen en Saunders [5]. In het artikel van Ommen en Saunders staat een soortgelijke theorie. Het bewijs van Theorie 1 is aangepast ten opzichte van het bewijs uit het artikel, omdat er een aantal aannames moesten worden toegevoegd om de theorie te kunnen bewijzen. Dit wordt uitgelegd in de volgende paragraaf.

4.5 Toegevoegde aannames voor Theorie 1

Het bewijs van Theorie 1 in Appendix B verschilt van het bewijs uit de appendix van het artikel van Ommen en Saunders [5]. In deze paragraaf wordt uitgelegd waarom het bewijs in dit verslag is aangepast.

4.5.1 Verschil in de aannames

Het verschil tussen de bewijzen zit in de convergentie van

$$\left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d [\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right|$$

naar nul. In het artikel wordt de convergentie aangetoond aan de hand van de totale variatie norm,

$$\|\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)\|_{TV} \xrightarrow{P_{\theta_a}^{n_a}} 0.$$

Echter, zoals in het bewijs is gedefinieerd, is $\Phi(\theta | X, M)$ een absoluut continue verdeling van $\theta_a \in \Theta_a$ en is $\delta_{\theta_{a_0}}(\theta)$ een discrete verdeling op het punt θ_{a_0} . Hierdoor kan er geen conclusie worden getrokken over de convergentie in totale variatie norm. Dit wordt aangetoond aan de hand van de definitie van de totale variatie norm in het volgende tegenvoorbeeld.

Wanneer de totale variatie norm in bovenstaande convergentie met de definitie wordt omschreven, geldt het volgende.

$$\begin{aligned} \int_{\Theta_a} |d(\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a))| &= \int_{\Theta_a \setminus \{\theta_{a_0}\}} |d(\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a))| \\ &\quad + \int_{\theta_{a_0}} |d(\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a))| \\ &= |1| + |1| \\ &\neq 0. \end{aligned}$$

Het bewijs dat de convergentie toch plaatsvindt, is aangepast. In Appendix B.3 wordt de convergentie bewezen aan de hand van het Lemma van Portmanteau.

4.5.2 Theoriën voor het bewijs van Theorie 1

Om het de theoriën toe te lichten, worden in deze paragraaf een aantal definities uitgewerkt. Deze definities en theoriën zijn afkomstig uit het boek *Asymptotic Statistics* van Van der Vaart [10]. De belangrijkste definities zijn

- een *stochastische vector* in \mathbb{R}^k is een vector $X = (X_1, \dots, X_k)$ met reële stochastische variabelen;
- een *verdelingsfunctie* van X is de afbeelding $x \mapsto P(X \leq x)$.

Een rij van stochastische vectoren X_n *convergeert zwak* naar een stochastische vector X wanneer $P(X_n \leq x) \rightarrow P(X \leq x)$ voor elke x waarin de limiet verdelingsfunctie $x \mapsto P(X \leq x)$ continu is. Dit wordt genoteerd als $X_n \rightsquigarrow X$.

Een rij van stochastische vectoren X_n *convergeert in kans* naar een stochastische vector X wanneer voor alle $\epsilon > 0$, $P(d(X_n, X) > \epsilon) \rightarrow 0$. Dit wordt genoteerd als $X_n \xrightarrow{P} X$.

Een verzameling van stochastische vectoren is *uniform begrensd* wanneer voor elke $\epsilon > 0$ er een constante M bestaat zodat $\sup_{\alpha} P(\|X_{\alpha}\| > M) < \epsilon$.

Het Lemma van Portmanteau geeft een aantal equivalenten met betrekking tot zwakke convergentie. In dit verslag worden maar twee van deze equivalenten genoemd, omdat de rest niet wordt toegepast. Prohorov's Theorie zegt iets over uniforme begrensdheid en zwakke convergentie.

Het Lemma van Portmanteau

Voor willekeurige stochastische vectoren X_n en X zijn de volgende uitdrukkingen equivalent.

1. $P(X_n \leq x) \rightarrow P(X \leq x)$ voor alle continue punten van $x \mapsto P(X \leq x)$;
2. $Ef(X_n) \rightarrow Ef(X)$ voor alle begrensde, continue functies f .

Prohorov's Theorie

Laat X_n stochastische vectoren zijn in \mathbb{R}^k . Als $X_n \rightsquigarrow X$ voor een X , dan $\{X_n : n \in \mathbb{N}\}$ is uniform begrensd.

Toepassing

Het bewijs van de convergentie in Theorie 3 staat in Appendix B. Aan de hand van zwakke convergentie wordt de eerste voorwaarde van het Lemma van Portmanteau aangetoond. Uit de tweede voorwaarde kan vervolgens de convergentie worden afgeleid.

4.6 Convergentie van de Bayes Factor naar de ware likelihood ratio voor het specifieke-bron probleem

Niet alleen voor het gemeenschappelijke-bron probleem, maar ook voor het specifieke-bron probleem kan worden bewezen dat de Bayes Factor in kans convergeert naar de ware likelihood ratio. Deze convergentie is geformaliseerd in Theorie 2 als volgt. Hierin geldt weer dat $\hat{\theta}_{n|M_2}$ de MLE is voor de parameter $\theta_{n|M_2}$.

Theorie 2 (Specifieke-bron Bayes Factor consistentie)

Gegeven een vaste observatie van x_c , stel dat de likelihood ratio functie $\lambda_{ss}(\theta|x_c)$ begrensd is in een omgeving van θ_0 en dat $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 . Stel verder dat de aannames van de Bernstein-von Mises Theorie voldoen. Dan, als $n \rightarrow \infty$, convergeert de Bayes Factor in P_θ^n -waarschijnlijkheid naar de ware likelihood ratio,

$$\beta_{cs}(X_{a,n_a}, x_b, x_c) \xrightarrow{P_{\theta_a}^{n_a}} \lambda_{cs}(\theta_{a_0}|x_b, x_c).$$

Ter verduidelijking staat in Appendix F.2 een lijst met alle aannames die gelden voor Theorie 2.

Het bewijs van Theorie 2 is hetzelfde als het bewijs van Theorie 1 en kan dus worden afgeleid uit Appendix B.

Het resultaat van Theorie 1 en Theorie 2 zorgt voor een vergroting van de effectiviteit van de communicatie tussen leden van een van beide paradigma's, zowel Bayesiaans als frequentistisch. Zo kan een resultaat bestaand uit de Bayes Factor worden gepresenteert aan een frequentist in de vorm van de limiet van deze Bayes Factor, wanneer meer en meer informatie wordt gegenereerd.

In de realiteit is het vaak onhaalbaar om meer informatie te verzamelen. De sample groottes zijn zelden groot genoeg om de waarde van de Bayes Factor te gebruiken als een groot-sample vervanging voor de waarde van de ware likelihood ratio [5].

Het volgende hoofdstuk beschrijft een theorie die wel toe te passen is in de realiteit.

5 Asymptotische resultaten voor de parameter intervallen

Zowel in de Bayesiaanse statistiek als in de frequentistische statistiek is er sprake van parameter intervallen. Omdat de basis van beide paradigma's verschillend is, zullen ook de parameter intervallen voor beide paradigma's van elkaar verschillen. In dit hoofdstuk wordt beschreven wat de samenhang is tussen een Bayesiaans geloofwaardigheids interval en de likelihood ratio. In tegenstelling tot Theorie 1 en Theorie 2 in het voorgaande hoofdstuk, zal de theorie uit dit hoofdstuk meer praktische toepassingen hebben.

Net als in het voorgaande hoofdstuk wordt voor de theorie, lemma's en toepassingen in dit hoofdstuk de MLE $\hat{\theta}$ gebruikt als consistente schatter voor θ .

5.1 Soorten parameterintervallen

In de Bayesiaanse statistiek is er sprake van een geloofwaardigheids interval. In de frequentistische statistiek is er sprake van een betrouwbaarheids interval. Deze intervallen verschillen van elkaar. Beide intervallen en de verschillen staan uitgelegd in onderstaande paragrafen.

5.1.1 Betrouwbaarheids interval

Een betrouwbaarheids interval is gebaseerd op de frequentistische statistiek. Zoals eerder vermeld, gaat de frequentistische statistiek er vanuit dat er een ware parameter(vector) bestaat. Dit is een vaste waarde en deze waarde is meestal onbekend. Er kan bij een betrouwbaarheids interval dan ook niet gesproken worden van een kans dat de parameter binnen het interval valt, omdat dit of wel, of niet zo is.

De betrouwbaarheid van het interval geeft aan in hoeveel gevallen de ware parameter binnen het betrouwbaarheids interval ligt, wanneer deze intervallen telkens op dezelfde manier worden bepaald. Een 95% betrouwbaarheids interval voor bijvoorbeeld het gemiddelde van de normaalverdeling, betekent dat asymptotisch voor 95% van de berekende intervallen, het ware gemiddelde binnen het interval ligt.

Er wordt ook wel gezegd dat de grenzen van een betrouwbaarheids interval gezien worden als stochastische variabelen [4].

5.1.2 Geloofwaardigheids interval

Een geloofwaardigheids interval is gebaseerd op de Bayesiaanse statistiek. Zoals eerder vermeld, is er bij de Bayesiaanse statistiek sprake van een verdeling op de parameter in plaats van een vaste, ware parameter. Er kan bij een geloofwaardigheids interval dus niet worden bepaald of de parameter wel of niet binnen een interval ligt. Wel kan er worden gesproken van de kans dat de parameter binnen het interval valt [2].

Het geloofwaardigheids interval is een interval in het domein van de posterior verdeling op de parameter. Zoals beschreven in Paragraaf 3.3, is er een prior verdeling nodig om deze posterior verdeling op te stellen. Om een geloofwaardigheids interval te bepalen, is deze prior verdeling dus ook nodig.

Bij een geloofwaardigheids interval wordt er gebruik gemaakt van vaste grenzen. De parameter wordt, net zoals in de gehele Bayesiaanse statistiek, beschouwd als stochastische variabele. Een 95% geloofwaardigheids interval is dan ook een interval dat 95% van de oppervlakte onder de posterior verdeling bevat. Dus er is asymptotisch 95% kans dat de parameter daadwerkelijk in dit interval valt, gegeven het bewijs.

5.2 Geloofwaardigheids interval voor de likelihood ratio

Met de resultaten uit Theorie 1 en Theorie 2 in het voorgaande hoofdstuk, kan er een geloofwaardigheids interval worden opgesteld voor de waarde van de likelihood ratio. Dit interval is afgeleid van de posterior verdeling van de parameters. Het precieze resultaat staat hieronder gegeven in Theorie 3. Een resultaat zoals in Theorie 3 is ook beschreven in het artikel van Ommen en Saunders [5]. Om de theorie te kunnen bewijzen, waren er een aantal extra aannames nodig. Dit staat beschreven in de volgende paragraaf.

Theorie 3 (Benadering $1 - \alpha$ geloofwaardigheids interval voor de likelihood ratio)

Laat de aannames van Lemma 3.1 en Lemma 3.2 gelden en laat

$$\mathcal{I}_n = \beta(X_n) \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sigma_n,$$

waar $\beta(X_n)$ de rij van de Bayes Factoren van of de gemeenschappelijke-bron of de specifieke-bron representeert, $0 < \alpha < 1$ het gewenste significantie niveau is, $\sqrt{n}\sigma_n$ de rij van kwantielen $q_{\Phi(1),n}$ voor de posterior verdeling van $\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2}))$ is en Φ^{-1} de standaard normale kwantiel functie is. Als $n \rightarrow \infty$, dan

$$\Pi(\lambda(\theta) \in \mathcal{I}_n | X_n, M_2) \rightarrow 1 - \alpha.$$

Ter verduidelijking staat in Appendix F.3 een lijst met alle aannames die gelden voor Theorie 3 inclusief de aannames van de Lemma's 3.1 en 3.2. Het bewijs van Theorie 3 wordt gegeven in Appendix E. In het bewijs van Theorie 3 worden twee lemma's toegepast. Lemma 3.1 stelt dat de totale variatie norm tussen de posterior verdeling van $\lambda(\theta)$ en een normaalverdeling convergeert naar nul in P_θ^n -kans. Lemma 3.2 stelt dat de totale variatie norm tussen twee specifieke normaalverdelingen ook convergeert naar nul in P_θ^n -kans. Verder wordt voor beide lemma's $\gamma_{n|M_2}^2$ gedefinieerd als

$$\gamma_{n|M_2}^2 = \lambda'(\hat{\theta}_{n|M_2})^T I_{\hat{\theta}_{n|M_2}}^{-1} \lambda'(\hat{\theta}_{n|M_2}).$$

Lemma 3.1 *Stel voor een gegeven observatie van de onbekende-bron verzameling(en) van observaties dat de likelihood ratio functie, $\lambda(\theta)$, is twee keer continu differentieerbaar, en stel dat $\hat{\theta}_{n|M_2}$ een consistente schatter voor θ_0 is onder P_θ^n -kans. Als de aannames van de Bernstein-von Mises Theorie gelden en $n \rightarrow \infty$, dan*

$$\left\| \Pi(\lambda(\theta)|X_n, M_2) - \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \frac{1}{n}\gamma_{n|M_2}^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Lemma 3.2 *Stel voor een gegeven observatie van de onbekende-bron verzameling(en) van observaties dat $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 onder P_θ^n -kans en stel dat $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$. Stel dat de Fisher Informatie Matrix en continue afbeelding en positief definit is. Stel verder dat $\lambda'(\theta_0)$ ongelijk is aan de nulvector. Als de aannames van de Bernstein-von Mises Theorie gelden en $n \rightarrow \infty$, dan*

$$\left\| \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2) - \mathcal{N}(\beta(X_n), n\sigma_n^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Ter verduidelijking staan de aannames die gelden voor Lemma 3.1 en Lemma 3.2 in Appendix F.4 en Appendix F.5 respectievelijk. De bewijzen van Lemma 3.1 en Lemma 3.2 worden respectievelijk gegeven in Appendix C en Appendix D. De resultaten van de lemma's hebben op zichzelf geen praktische toepassing, maar dienen ter ondersteuning van het bewijs van Theorie 3.

Net zoals voor de theoriën, geldt ook voor Lemma 3.1 en Lemma 3.2 dat er soortgelijke theoretische resultaten zijn gepubliceerd in het artikel van Ommen en Saunders [5]. Om de lemma's te bewijzen, zijn er een aantal aannames veranderd en toegevoegd.

5.3 Verandering in aannames

Lemma 3.2 en Theorie 3 zijn aangepast ten opzichte van het artikel van Ommen en Saunders [5]. Voor Lemma 3.2 was dit nodig om het lemma te kunnen bewijzen.

In Lemma 3.2 zijn de volgende aannames toegevoegd of aangepast,

- $\sigma_n^2 \rightarrow 0$ wordt $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$;
- beide voorwaarden voor de Fisher Informatie Matrix zijn toegevoegd;
- de voorwaarde $\lambda'(\theta_0) \neq 0$ is toegevoegd.

In Theorie 3 is een aanname zodanig veranderd dat de theorie ook anders is. In Theorie 3 is de definitie van σ_n aangepast.

5.3.1 Aanpassingen Lemma 3.2

Om de convergentie van $\gamma_{n|M_2}^2$ aan te tonen, is de Continue Afbeeldingstelling gebruikt. Hiervoor is de voorwaarde dat de Fisher Informatie Matrix een continue afbeelding is, toegevoegd. In het artikel tonen ze de convergentie aan aan de hand van de voorwaarden van de Bernstein–von Mises Theorie. De aannames van de Bernstein–von Mises Theorie gaan buiten de stof van dit verslag om en om die reden is deze extra aanname toegevoegd.

De voorwaarde dat $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$ in Lemma 3.2 werd door het artikel gegeven door $\sigma_n^2 \rightarrow 0$ als $n \rightarrow \infty$. Deze voorwaarde wordt in het artikel gebruikt om een conclusie te trekken over de convergentie van

$$\left\| \mathcal{N}\left(\lambda(\hat{\theta}_{n|M_2}), \frac{1}{n}\gamma_{n|M_2}^2\right) - \mathcal{N}\left(\beta(X_n), \sigma_n^2\right) \right\|_{TV} \xrightarrow{P_{\theta}^n} 0.$$

Deze conclusie is gebaseerd op de convergentie van $\lambda(\hat{\theta}_{n|M_2})$ en $\beta(X_n)$ naar dezelfde waarde en de convergentie van $\frac{1}{n}\gamma_{n|M_2}^2$ en σ_n^2 naar nul. Het bewijs is in dit verslag aangepast, omdat, wanneer de varianties convergeren naar nul, er niks gezegd kan worden over de convergentie in totale variatie van de normaalverdelingen naar nul. Dit blijkt uit het volgende tegenvoorbeeld:

$$\left\| \mathcal{N}\left(0, \frac{1}{n}\right) - \mathcal{N}\left(0, \frac{2}{n}\right) \right\|_{TV} = \left\| \mathcal{N}(0, 1) - \mathcal{N}(0, 2) \right\|_{TV} \neq 0.$$

Wanneer de varianties naar een vaste waarde, ongelijk aan nul, convergeren, kan er wel een conclusie getrokken worden met betrekking tot de convergentie van de totale variatie norm.

Om die reden is de voorwaarde aangepast naar $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als

$n \rightarrow \infty$. Om vervolgens te zorgen dat $\lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0) > 0$, zijn de voorwaarde over de positief definitie I_{θ_0} en de voorwaarde $\lambda'(\theta_0) \neq 0$ toegevoegd. Vervolgens wordt er aan de hand van deze voorwaarde aangetoond dat

$$\left\| \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2) - \mathcal{N}(\beta(X_n), n\sigma_n^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Zie Appendix D voor het bewijs van Lemma 3.2.

5.3.2 Aanpassing Theorie 3

In het artikel van Ommen en Saunders [5] wordt σ_n gedefinieerd als de rij van posterior standaard fouten voor de likelihood ratio. Vervolgens wordt aangenomen dat de convergentie van de posterior verdeling van $\lambda(\theta)$ naar de normaal verdeling $\mathcal{N}(\beta(X_n), \sigma_n^2)$ in totale variatie norm impliceert dat de standaard afwijkingen van beide verdelingen convergeren.

Voor het bepalen van de standaard afwijking σ^2 in een normaal verdeling geldt dat de tweede momentschatter wordt gebruikt [3]. Echter, er kan niks geconcludeerd worden over de convergentie van tweede momentschatters vanuit de convergentie van de verdelingen in totale variatie norm.

De definitie van σ_n is in dit artikel zodanig aangepast dat er vanuit de convergentie in totale variatie norm wél een conclusie kan worden getrokken met betrekking tot de convergentie van σ_n . Dit is gedaan door $\sqrt{n}\sigma_n$ te definiëren als de rij van kwantielen $q_{\Phi(1),n}$ voor de posterior verdeling van $\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2}))$. Wanneer twee verdelingen convergeren in totale variatie norm, convergeren de kwantielen gedefinieerd op deze verdelingen ook [7]. Deze achtergrond informatie wordt gebruikt voor het bewijs van Theorie 3 in Appendix E.

Theorie over kwantielen

Om de aanpassing in Theorie 3 te bewijzen, volgt hier de algemene achtergrond informatie over kwantielen die in het bewijs wordt gebruikt.

Laat voor een punt $A \in (0, 1)$ gelden

$$q_A = \inf\{x : f(x) \geq A\}.$$

Dan is q_A het A -kwantiel gedefinieerd op een verdeling met kansdichtheidsfunctie $f(x)$ [2]. Oftewel, voor een getal $A \in (0, 1)$ gedefinieerd op een verdeling $\mathcal{N}(\mu, \sigma^2)$ geeft de waarde a zodat voor een stochastische variabele

$X \sim \mathcal{N}(\mu, \sigma^2)$ geldt $P(X \leq a) = A$.

Voor de standaard normale kwantiel functie Φ geldt dat $\Phi(a)$ gelijk is aan de kans dat voor een stochastische variabele $X \sim \mathcal{N}(0, 1)$ geldt $P(X \leq a)$.

Het kwantiel $q_{\Phi(a)}$ gedefinieerd op $\mathcal{N}(0, 1)$ geeft dan $q_{\Phi(a)} = a$.

Verder geldt voor elke verdeling $\mathcal{N}(\mu, \sigma^2)$ dat

$$q_{\Phi(1)} = \mu + \sigma.$$

5.4 Gevolgen van Theorie 3

5.4.1 De vorm van het geloofwaardigheids interval

Het geloofwaardigheids interval $\mathcal{I}_n = \beta(X_n) \pm \Phi^{-1}(1 - \frac{\alpha}{2})\sigma_n$ heeft een niet-standaard vorm. Deze vorm is gekozen, zodat het interval gegarandeerd het gemiddelde van de posterior verdeling bevat. Dit gemiddelde is in dit specifieke geval de Bayes Factor voor het niet-geneste model selectie probleem.

De reden dat dit interval zo is gekozen, is omdat de posterior verdeling een scheve verdeling is. Hierdoor valt de Bayes Factor niet in het lichaam van deze verdeling. Wanneer er een standaard vorm wordt gekozen voor het geloofwaardigheids interval, kan het zo zijn dat de Bayes Factor buiten het interval valt. Het interval voor de likelihood ratio komt dan niet overeen met de Bayes Factor. Dit kan serieuze problemen opleveren in de toepassing van Theorie 3 op het gebied van forensische statistiek [5].

5.4.2 Nadelen van het geloofwaardigheids interval voor de likelihood ratio

In Theorie 3 is een methode beschreven om een geloofwaardigheids interval op te stellen voor de likelihood ratio. Dit interval is sterk afhankelijk van zowel de Bayes Factor als de prior verdeling van de parameters, dus Bayesiaanse kwantiteiten. Een interval voor een klassieke statistiek wordt dus sterk beïnvloed door Bayesiaanse kwantiteiten.

Het geloofwaardigheids interval wordt vaak geïnterpreteerd als een bereik voor waarschijnlijke waarden voor de ware likelihood ratio [5]. Door de sterke invloed van de keuze van de prior verdeling, kan het zijn dat, wanneer de prior verdeling verkeerd wordt gekozen, het interval zoals beschreven in Theorie 3 de ware likelihood ratio niet bevat. Om deze reden moet het geloofwaardigheids interval zoals beschreven in Theorie 3 worde geïnterpreteerd als een bereik voor waarschijnlijke waarden voor de geschatte likelihood ratio, gegeven de gekozen prior verdeling en de eindige data.

Het geloofwaardigheids interval in Theorie 3 is afhankelijk van de posterior verdeling van $\sqrt{n} \left(\lambda\theta - \lambda(\hat{\theta}_{n|M_2}) \right)$. In dit verslag is aangenomen dat $\hat{\theta}_{n|M_2}$ de maximum likelihood schatter is voor θ_0 . Om de posterior verdeling op te stellen wordt zowel de posterior verdeling van $\lambda(\theta)$ (Bayesiaans) als de MLE $\hat{\theta}_{n|M_2}$ (frequentistisch) gebruikt. Het interval in Theorie 3 bevat dus zowel frequentistische als Bayesiaanse elementen.

De frequentistische MLE kan worden vervangen door elke efficiënte schatter. De theoriën zoals gegeven in dit verslag hoeven niet te worden aangepast, wanneer $\hat{\theta}_{n|M_2}$ een andere efficiënte schatter is, omdat de voorwaarde dat $\hat{\theta}_{n|M_2}$ niet nodig is om de theoriën te bewijzen. Door een Bayesiaanse efficiënte schatter te gebruiken, wordt het interval uit Theorie 3 opgesteld puur aan de hand van de Bayesiaanse statistiek.

5.4.3 Toepassing Theorie 3 in de praktijk

In dit verslag is Theorie 3 bewezen aan de hand van wiskundige definities, lemma's en theoriën. De toepassing van Theorie 3 blijkt in de praktijk een stuk lastiger. Dit komt doordat het bepalen van de posterior verdelingen en de Bayes Factor aan de hand van Vergelijkingen (11), (12), (13) en (14) in de praktijk complex blijkt te zijn.

De aanpassing ten opzichte van de theorie uit het artikel van Ommen en Saunders [5] heeft geen invloed op deze berekening, omdat beide theoriën het berekenen van de posterior verdelingen en de Bayes Factor vereisen.

De posterior verdelingen en de Bayes Factor kunnen op een andere manier worden bepaald. Het artikel van Ommen et al. [6] gebruikt bijvoorbeeld de Monte Carlo methode voor het bepalen van hoe de posterior verdelingen en de Bayes Factor. Door een andere methode te gebruiken om de posterior verdelingen en de Bayes Factor te bepalen, wordt de toepassing van Theorie 3 in de praktijk eenvoudiger. Wel geldt dan, zoals ook beschreven in het artikel van Ommen et al., dat het niet duidelijk is hoe en of de berekende fout moet worden meegenomen in het interval zoals beschreven in Theorie 3.

6 Conclusie

In dit verslag is de relatie tussen de Bayes Factor en de likelihood ratio onderzocht aan de hand van model selectie toegepast op de forensische statistiek.

In dit verslag zijn de theoretische resultaten uit het artikel ‘Reconciling the Bayes Factor and Likelihood Ratio for Two Non-Nested Model Selection Problems’ van Ommen en Saunders [5] geanalyseerd. Aan de hand van deze analyse wordt in dit verslag aangetoond dat de Bayes Factor kan worden uitgedrukt als de verwachte waarde van de bijbehorende likelihood ratio en dat er geloofwaardigheids intervallen kunnen worden opgesteld voor de likelihood ratio.

De theorie waarin deze geloofwaardigheids intervallen worden geformaliseerd, verschilt van de resultaten uit het artikel van Ommen en Saunders. In dit verslag wordt uitgelegd welke aanpassingen ten opzichte van de resultaten van Ommen en Saunders nodig zijn om de theorie te kunnen bewijzen.

Door de theoretische resultaten uit dit verslag, kan er in de forensische statistiek voortaan door aanhangers van zowel de Bayesiaanse als de frequentistische statistiek worden gecommuniceerd over een conclusie met betrekking tot de likelihood ratio.

A Bayes Factor afleidingen

A.1 Vergelijking (11)

$$\beta_{cs}(X_N) = \frac{\int f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1) d\Pi(\theta_a | X_a)}{\int f(\mathbf{x}_b | \theta_a, M_2) f(\mathbf{x}_c | \theta_a, M_2) d\Pi(\theta_a | X_a)} \quad (15a)$$

$$= \frac{1}{m(\mathbf{x}_b, \mathbf{x}_c | X_a, M_2)} \times \int f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1) d\Pi(\theta_a | X_a) \quad (15b)$$

$$= \int \frac{f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1)}{m(\mathbf{x}_b, \mathbf{x}_c | X_a, M_2)} d\Pi(\theta_a | X_a) \quad (15c)$$

$$= \int \frac{f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1)}{f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_2)} \times \frac{f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_2)}{m(\mathbf{x}_b, \mathbf{x}_c | X_a, M_2)} d\Pi(\theta_a | X_a) \quad (15d)$$

$$= \int \frac{f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_1)}{f(\mathbf{x}_b, \mathbf{x}_c | \theta_a, M_2)} d\Pi(\theta_a | X_a, \mathbf{x}_b, \mathbf{x}_c, M_2) \quad (15e)$$

$$= \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d\Pi(\theta_a | X_N, M_2). \quad (15f)$$

A.2 Vergelijking (12)

$$\beta_{ss}(X_N) \tag{16a}$$

$$= \frac{\int f(x_c|\theta_b, M_1)d\Pi(\theta_b|x_b)}{\int f(x_c|\theta_a, M_2)d\Pi(\theta_a|X_a)} \tag{16b}$$

$$= \int \frac{f(x_c|\theta_b, M_1)}{\int f(x_c|\theta_a, M_2)d\Pi(\theta_a|X_a)} d\Pi(\theta_b|x_b) \tag{16c}$$

$$= \int \frac{f(x_b|\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(x_c|\theta_b, M_1)}{\int f(x_c|\theta_a, M_2)d\Pi(\theta_a|X_a)} d\Pi(\theta_b) \tag{16d}$$

$$= \int \frac{\int f(X_a|\theta_a)d\Pi(\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_b|\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(x_c|\theta_b, M_1)}{\int f(x_c|\theta_a, M_2)d\Pi(\theta_a|X_a)} d\Pi(\theta_b) \tag{16e}$$

$$= \int \frac{\int f(X_a|\theta_a)d\Pi(\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_b|\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(x_c|\theta_b, M_1)d\Pi(\theta_b)}{\int \frac{f(X_a|\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} f(x_c|\theta_a, M_2)d\Pi(\theta_a)} \tag{16f}$$

$$= \int \frac{\int f(X_a|\theta_a)d\Pi(\theta_a)}{1} \frac{f(x_b|\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(x_c|\theta_b, M_1)d\Pi(\theta_b)}{\int f(X_a|\theta_a)f(x_c|\theta_a, M_2)d\Pi(\theta_a)} \tag{16g}$$

$$= \int \frac{f(X_a|\theta_a)}{1} \frac{f(x_b|\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(x_c|\theta_b, M_1)}{\int f(X_a|\theta_a)f(x_c|\theta_a, M_2)d\Pi(\theta_a)} d\Pi(\theta_a, \theta_b) \tag{16h}$$

$$= \int \frac{f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_b, M_1)}{\int f(x_b|\theta_b)d\Pi(\theta_b) \int f(X_a|\theta_a)f(x_c|\theta_a, M_2)d\Pi(\theta_a)} d\Pi(\theta_a, \theta_b) \tag{16i}$$

$$= \int \frac{f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_b, M_1)}{\int f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_a, M_2)d\Pi(\theta_a, \theta_b)} d\Pi(\theta_a, \theta_b) \tag{16j}$$

$$\equiv \int \frac{f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_b, M_1)}{m(X_a, x_b, x_c|M_2)} d\Pi(\theta_a, \theta_b). \tag{16k}$$

Het ‘ \equiv ’- teken wordt gebruikt om een definitie aan te geven. In bovenstaande vergelijking wordt $m(X_a, x_b, x_c|M_2)$ gedefinieerd in de laatste regel.

Hieruit kan vervolgens de likelihood worden afgeleid:

$$\beta_{ss}(X_N) = \int \frac{f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_b, M_1)}{m(X_a, x_b, x_c|M_2)} d\Pi(\theta_a, \theta_b) \quad (17a)$$

$$= \int \frac{f(x_c|\theta_a, M_2)}{f(x_c|\theta_a, M_2)} \times \frac{f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_b, M_1)}{m(X_a, x_b, x_c|M_2)} d\Pi(\theta_a, \theta_b) \quad (17b)$$

$$= \int \frac{f(x_c|\theta_b, M_1)}{f(x_c|\theta_a, M_2)} \times \frac{f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_a, M_2)}{m(X_a, x_b, x_c|M_2)} d\Pi(\theta_a, \theta_b) \quad (17c)$$

$$= \int \frac{f(x_c|\theta_b, M_1)}{f(x_c|\theta_a, M_2)} d\Pi(\theta_a, \theta_b|X_a, x_b, x_c, M_2) \quad (17d)$$

$$= \int \lambda_{ss}(\theta_a, \theta_b|x_c) d\Pi(\theta_a, \theta_b|X_N, M_2) \quad (17e)$$

A.3 Vergelijking (13)

$$\frac{1}{\beta_{cs}(X_N)} = \frac{\int f(x_b|\theta_a, M_2)f(x_c|\theta_a, M_2) d\Pi(\theta_a|X_a)}{\int f(x_b, x_c|\theta_a, M_1) d\Pi(\theta_a|X_a)} \quad (18a)$$

$$= \int \frac{f(x_b|\theta_a, M_2)f(x_c|\theta_a, M_2)}{\int f(x_b, x_c|\theta_a, M_1) d\Pi(\theta_a|X_a)} d\Pi(\theta_a|X_a) \quad (18b)$$

$$= \int \frac{f(x_b, x_c|\theta_a, M_2)}{\int f(x_b, x_c|\theta_a, M_1) d\Pi(\theta_a|X_a)} d\Pi(\theta_a|X_a) \quad (18c)$$

$$= \int \frac{f(x_b, x_c|\theta_a, M_1)}{f(x_b, x_c|\theta_a, M_1)} \frac{f(x_b, x_c|\theta_a, M_2)}{\int f(x_b, x_c|\theta_a, M_1) d\Pi(\theta_a|X_a)} d\Pi(\theta_a|X_a) \quad (18d)$$

$$= \int \frac{f(x_b, x_c|\theta_a, M_2)}{f(x_b, x_c|\theta_a, M_1)} \frac{f(x_b, x_c|\theta_a, M_1)}{\int f(x_b, x_c|\theta_a, M_1) d\Pi(\theta_a|X_a)} d\Pi(\theta_a|X_a) \quad (18e)$$

$$= \int \frac{f(x_b, x_c|\theta_a, M_2)}{f(x_b, x_c|\theta_a, M_1)} d\Pi(\theta_a|X_a, x_b, x_c, M_1) \quad (18f)$$

$$= \int \frac{1}{\lambda_{cs}(\theta_a|x_b, x_c)} d\Pi(\theta_a|X_N, M_1). \quad (18g)$$

Hieruit volgt dat de Bayes Factor als inverse van de verwachting kan worden geschreven,

$$\beta_{cs}(X_N) = \left[\int \frac{1}{\lambda_{cs}(\theta_a|x_b, x_c)} d\Pi(\theta_a|X_N, M_1) \right]^{-1}.$$

A.4 Vergelijking (14)

$$\frac{1}{\beta_{ss}(X_N)} \quad (19a)$$

$$= \frac{\int f(x_c|\theta_a, M_2)d\Pi(\theta_a|X_a)}{\int f(x_c|\theta_b, M_1)d\Pi(\theta_b|x_b)} \quad (19b)$$

$$= \int \frac{f(x_c|\theta_a, M_2)}{\int f(x_c|\theta_b, M_1)d\Pi(\theta_b|x_b)} d\Pi(\theta_a|X_a) \quad (19c)$$

$$= \int \frac{f(X_a|\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_c|\theta_a, M_2)}{\int f(x_c|\theta_b, M_1)d\Pi(\theta_b|x_b)} d\Pi(\theta_a) \quad (19d)$$

$$= \int \frac{\int f(x_b|\theta_b)d\Pi(\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(X_a|\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_c|\theta_a, M_2)}{\int f(x_c|\theta_b, M_1)d\Pi(\theta_b|x_b)} d\Pi(\theta_a) \quad (19e)$$

$$= \int \frac{\int f(x_b|\theta_b)d\Pi(\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} \frac{f(X_a|\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_c|\theta_a, M_2)d\Pi(\theta_a)}{\int \frac{f(x_b|\theta_b)}{\int f(x_b|\theta_b)d\Pi(\theta_b)} f(x_c|\theta_b, M_1)d\Pi(\theta_b)} \quad (19f)$$

$$= \int \frac{\int f(x_b|\theta_b)d\Pi(\theta_b)}{1} \frac{f(X_a|\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_c|\theta_a, M_2)}{\int f(x_b|\theta_b)f(x_c|\theta_b, M_1)d\Pi(\theta_b)} d\Pi(\theta_a) \quad (19g)$$

$$= \int \frac{f(x_b|\theta_b)}{1} \frac{f(X_a|\theta_a)}{\int f(X_a|\theta_a)d\Pi(\theta_a)} \frac{f(x_c|\theta_a, M_2)}{\int f(x_b|\theta_b)f(x_c|\theta_b, M_1)d\Pi(\theta_b)} d\Pi(\theta_a, \theta_b) \quad (19h)$$

$$= \int \frac{f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_a, M_2)}{\int f(X_a|\theta_a)d\Pi(\theta_a) \int f(x_b|\theta_b)f(x_c|\theta_b, M_1)d\Pi(\theta_b)} d\Pi(\theta_a, \theta_b) \quad (19i)$$

$$= \int \frac{f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_a, M_2)}{\int f(X_a|\theta_a)f(x_b|\theta_b)f(x_c|\theta_b, M_1)d\Pi(\theta_a, \theta_b)} d\Pi(\theta_a, \theta_b) \quad (19j)$$

$$\equiv \int \frac{f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_a, M_2)}{m(X_a, x_b, x_c|M_1)} d\Pi(\theta_a, \theta_b). \quad (19k)$$

Deze uitdrukking kan verder worden herleid tot de inverse van de likelihood ratio verwachting:

$$\frac{1}{\beta_{ss}(X_N)} = \int \frac{f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_a, M_2)}{m(X_a, x_b, x_c|M_1)} d\Pi(\theta_a, \theta_b) \quad (20a)$$

$$= \int \frac{f(x_c|\theta_b, M_1)}{f(x_c|\theta_b, M_1)} \frac{f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_a, M_2)}{m(X_a, x_b, x_c|M_1)} d\Pi(\theta_a, \theta_b) \quad (20b)$$

$$= \int \frac{f(x_c|\theta_a, M_2)}{f(x_c|\theta_b, M_1)} \frac{f(x_b|\theta_b)f(X_a|\theta_a)f(x_c|\theta_b, M_1)}{m(X_a, x_b, x_c|M_1)} d\Pi(\theta_a, \theta_b) \quad (20c)$$

$$= \int \frac{f(x_c|\theta_a, M_2)}{f(x_c|\theta_b, M_1)} \frac{f(X_a, x_b, x_c|\theta_a, \theta_b, M_1)}{m(X_a, x_b, x_c|M_1)} d\Pi(\theta_a, \theta_b) \quad (20d)$$

$$= \int \frac{f(x_c|\theta_a, M_2)}{f(x_c|\theta_b, M_1)} d\Pi(\theta_a, \theta_b|X_a, x_b, x_c, M_1) \quad (20e)$$

$$= \int \frac{1}{\lambda_{ss}(\theta_a, \theta_b|x_c)} d\Pi(\theta_a, \theta_b|X_N, M_1). \quad (20f)$$

Hieruit volgt dat de Bayes Factor van het specifieke-bron probleem gelijk is aan de inverse van de verwachting. De Bayes Factor wordt dan gegeven door

$$\beta_{ss}(X_N) = \left[\frac{1}{\lambda_{ss}(\theta_a, \theta_b|x_c)} d\Pi(\theta_a, \theta_b|X_N, M_1) \right]^{-1}.$$

B Bewijs van Theorie 1

Ter ondersteuning van het bewijs van Theorie 1 worden hier de notaties gegeven. Laat X_{a,n_a} een rij van stochastische variabelen zijn die hoort bij de generatie van de databanken X_{a,n_a} . Hierin is n_a het indexnummer dat de groeiende hoeveelheid subjects in de databank aantoond. Let wel op dat elk subject een vast aantal elementen bezit. Verder is $P_{\theta_a}^{n_a}$ de gezamenlijke kansmaat op X_{a,n_a} voor alle $\theta_a \in \Theta_a$ inclusief θ_{a_0} . De twee verzamelingen van onbekende bron worden genoteerd als x_b en x_c .

Laat $\hat{\theta}_{a|M_2}$ de maximum likelihood schatter zijn voor θ_{a_0} , berekend met de gehele databank X_{a,n_a} wanneer x_b en x_c worden gegenereerd onder M_2 . Laat tot slot $\beta_{cs}(X_{a,n_a}, x_b, x_c)$ de stochastische functie zijn die hoort bij de Bayes Factor gegeven in Vergelijking (11) voordat de waarde van X_{a,n_a} wordt geobserveerd [5].

Theorie 1 (Gemeenschappelijke-bron Bayes Factor consistentie)

Gegeven een vaste observatie van x_b en x_c , stel dat de likelihood ratio functie $\lambda_{cs}(\theta_a|x_b, x_c)$ begrensd is in een omgeving van θ_{a_0} en dat $\hat{\theta}_{a|M_2}$ een consistente schatter is voor θ_{a_0} . Stel verder dat de aannames van de Bernstein-von Mises Theorie voldoen. Dan, als $n_a \rightarrow \infty$, convergeert de Bayes Factor in $P_{\theta_a}^{n_a}$ -waarschijnlijkheid naar de ware likelihood ratio,

$$\beta_{cs}(X_{a,n_a}, x_b, x_c) \xrightarrow{P_{\theta_a}^{n_a}} \lambda_{cs}(\theta_{a_0}|x_b, x_c).$$

B.1 Convergentie van de Bayes Factor naar de likelihoodratio

Om te laten zien dat de Bayes factor convergeert naar de likelihood als $n_a \rightarrow \infty$, wordt er met behulp van een bovengrens afgeleid dat

$$|\beta_{cs}(X_{a,n_a}, x_b, x_c) - \lambda_{cs}(\theta_{a_0}|x_b, x_c)|$$

naar nul convergeert wanneer $n_a \rightarrow \infty$.

Laat X_{n_a} de random variabele zijn die hoort bij de generatie van de gehele databank. Hierin genereert X_{a,n_a} de observaties van de populatie van bronnen en hierin zijn de observaties x_b en x_c vastgesteld. Laat $\delta_{\theta_{a_0}}$ de waarschijnlijkheids maat zijn, gegeven door

$$\int_{\Theta_a} d\delta_{\theta_{a_0}}(\theta_a) = \begin{cases} 1 & \text{als } \theta_{a_0} \in \Theta_a \\ 0 & \text{als } \theta_{a_0} \notin \Theta_a \end{cases}.$$

Hieruit volgt $\int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d\delta_{\theta_{a_0}}(\theta_a) = \lambda_{cs}(\theta_{a_0} | \mathbf{x}_b, \mathbf{x}_c)$.

Laat $\Phi(\theta_a | X_{n_a}, M_2)$ een kansmaat zijn die correspondeert met de normaalverdeling op θ_a met als gemiddelde $\hat{\theta}_{a|M_2}$ en variantie $\frac{1}{n_a} I_{\hat{\theta}_{a|M_2}}^{-1}$. Uit $\Phi(\theta_a | X_{n_a}, M_2)$, Vergelijking (11) en de driehoeksongelijkheid, volgt

$$\begin{aligned}
& |\beta_{cs}(X_{n_a}) - \lambda_{cs}(\theta_{a_0} | \mathbf{x}_b, \mathbf{x}_c)| \\
&= \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d\Pi(\theta_a | X_{n_a}, M_2) - \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d\delta_{\theta_{a_0}}(\theta_a) \right| \\
&= \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Pi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right| \\
&= \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2) \right. \\
&\quad \left. + \Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right| \\
&\leq \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)] \right| \\
&\quad + \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right|.
\end{aligned}$$

Om aan te tonen dat de absolute afstand tussen de Bayes Factor en de ware likelihood ratio naar nul convergeert, wordt aangetoond dat de bovengrens naar nul convergeert. De bovengrens voor $|\beta_{cs}(X_{n_a}) - \lambda_{cs}(\theta_{a_0} | \mathbf{x}_b, \mathbf{x}_c)|$ bestaat, zoals hierboven te zien is, uit twee delen:

1. $\left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)] \right|$;
2. $\left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right|$.

Voor eerste deel wordt door middel van de Bernstein–von Mises Theorie, zie Paragraaf 4.3, aangetoond dat als $n_a \rightarrow \infty$, de norm naar nul convergeert. Dit deel van het bewijs is beschreven in Appendix B.2. Voor het tweede deel wordt er gebruik gemaakt van zwakke convergentie en het Lemma van Portmanteau. Het tweede deel van het bewijs staat in Appendix B.3.

B.2 Toepassing van de Bernstein–von Mises Theorie

Om de Bernstein–von Mises Theorie toe te passen op

$$\left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d[\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)] \right|,$$

moet er een afleiding worden gemaakt naar de totale variatie norm. Eerst wordt de driehoeksongelijkheid toegepast,

$$\begin{aligned} & \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)] \right| \\ & \leq \int |\lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c)| |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]| \\ & = \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]|. \end{aligned}$$

De laatste stap uit deze ongelijkheid volgt uit $\lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) > 0$. Dit geldt, omdat de likelihoodfuncties altijd positief zijn en de likelihood ratio functie een verhouding tussen likelihoodfuncties geeft.

In de aannames is gesteld dat $\lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c)$ begrensd is. Dat betekent dat er een reëel getal $C > 0$ bestaat zodat $\lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) \leq C$. Hieruit volgt

$$\begin{aligned} & \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]| \\ & \leq \int C |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]| \\ & = C \int |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]|. \end{aligned}$$

Dit is een asymptotische verdeling. Omdat $n_a \rightarrow \infty$, geldt dat $\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)$ een rij van gesigeneerde maten is. Definieer de totale variatie norm voor gesigeneerde maten door $\|\rho\|_{TV} = \int |d\rho|$. Voor de rij van gesigeneerde maten geldt dan

$$\int |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]| = \|\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)\|_{TV}$$

en dus ook

$$\begin{aligned} & C \int |d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)]| \\ & = C \|\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)\|_{TV}. \end{aligned}$$

Deze afleiding geeft de ongelijkheid

$$\begin{aligned} & \left| \int \lambda_{cs}(\theta_a | \mathbf{x}_b, \mathbf{x}_c) d [\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)] \right| \\ & \leq C \|\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)\|_{TV}, \end{aligned} \tag{21}$$

voor een reëel getal $C > 0$. De bovengrens is nu uitgedrukt in de totale variatie norm. Hierin is $\Pi(\theta_a)$ de prior verdeling op θ_a . De posterior verdeling wordt gegeven door $\Pi(\theta_a|X_{n_a}, M_2)$. Verder geldt dat $\Phi(\theta_a|X_{n_a}, M_2)$ de kansmaat is die correspondeert met de normaalverdeling met gemiddelde $\hat{\theta}_{a|M_2}$ en variantie $\frac{1}{n_a} I_{\hat{\theta}_{a|M_2}}^{-1}$. Het rechterlid uit de ongelijkheid in Vergelijking (21) kan worden herschreven als

$$C \left\| \Pi(\theta_a|X_{n_a}, M_2) - \mathcal{N}(\hat{\theta}_{a|M_2}, \frac{1}{n_a} I_{\hat{\theta}_{a|M_2}}^{-1}) \right\|_{TV}.$$

Zoals beschreven in Paragraaf 4.3, geldt wegens de Bernstein–von Mises Theorie dat wanneer $n \rightarrow \infty$, de posterior verdeling van θ convergeert naar de normaal verdeling met als gemiddelde $\hat{\theta}_n$ en als variantie $\frac{1}{n} I_{\hat{\theta}_n}^{-1}$. Echter, in bovenstaande vergelijking wordt er iets gezegd over de totale variatie norm tussen een normaal verdeling en de posterior verdeling van θ_a . Dus moet er een afleiding worden gemaakt van de Bernstein–von Mises Theorie om hier iets over te kunnen zeggen.

Voor het experiment $P_{\theta_a} : \theta_a \in \Theta_a$ wordt de prior verdeling van θ_a gegeven door $\Pi(\theta_a)$. De posterior verdeling van θ_a wordt gegeven door $\Pi(\theta_a|X_{n_a}, M_2)$. Verder geldt dat $\hat{\theta}_{a|M_2}$ voor $n_a \in \mathbb{N}$ een rij van maximum likelihood schatters voor θ_a is. Er is aangenomen dat $\hat{\theta}_{a|M_2}$ een consistente schatter is voor θ_{a_0} . Ook is er aangenomen dat de Fisher Informatie Matrix een continue functie van θ is. Wegens de Continue Afbeelding Stelling geldt dan

$$I_{\hat{\theta}_{a|M_2}}^{-1} \xrightarrow{P_{\theta_a}^{n_a}} I_{\theta_{a_0}}^{-1}.$$

Wegens de aanname dat alle aannames van de Bernstein–von Mises Theorie gelden, kan de Bernstein–von Mises Theorie op de posterior verdeling van θ_a worden toegepast. Er geldt dat de posterior verdeling van θ_a gegeven door $\Pi(\theta_a|X_{n_a}, M_2)$, wanneer $n_a \rightarrow \infty$, in totale variatie norm convergeert naar een normaalverdeling met gemiddelde de maximum likelihood schatter voor θ_a ; $\hat{\theta}_{a|M_2}$ en variantie $\frac{1}{n_a} I_{\theta_{a_0}}^{-1}$. Omdat $I_{\hat{\theta}_{a|M_2}}^{-1} \xrightarrow{P_{\theta_a}^{n_a}} I_{\theta_{a_0}}^{-1}$ en $n_a \rightarrow \infty$, mag hetzelfde worden gezegd over een normaalverdeling met variantie $\frac{1}{n_a} I_{\hat{\theta}_{a|M_2}}^{-1}$. Wegens de Bernstein–von Mises Theorie geldt dan

$$\left\| \Pi(\theta_a|X_{n_a}, M_2) - \mathcal{N}(\hat{\theta}_{a|M_2}, \frac{1}{n_a} I_{\hat{\theta}_{a|M_2}}^{-1}) \right\|_{TV} \xrightarrow{P_{\theta_a}^{n_a}} 0,$$

als $n_a \rightarrow \infty$. Uit de ongelijkheid in Vergelijking (21) blijkt vervolgens dat

als $n_a \rightarrow \infty$,

$$\left| \int \lambda_{cs}(\theta_a | x_b, x_c) d[\Pi(\theta_a | X_{n_a}, M_2) - \Phi(\theta_a | X_{n_a}, M_2)] \right| \xrightarrow{P_{\theta_a}^{n_a}} 0. \quad (22)$$

Dus er is aangetoond dat het eerste deel van de bovengrens, zoals gegeven aan het begin van dit bewijs, convergeert naar nul wanneer $n_a \rightarrow \infty$.

B.3 Toepassing van het Lemma van Portmanteau

Aan de hand van zwakke convergentie en het Lemma van Portmanteau kan worden aangetoond dat

$$\left| \int \lambda_{cs}(\theta_a | x_b, x_c) d[\Phi(\theta_a | X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right| \quad (23)$$

convergeert naar nul in $P_{\theta_a}^{n_a}$ -kans wanneer $n_a \rightarrow \infty$.

B.3.1 Het Lemma van Portmanteau

Het Lemma van Portmanteau is beschreven in Paragraaf 4.5.2. Om het lemma toe te passen op Vergelijking (23), geldt voor de stochastische vectoren $\hat{\theta}_{a|M_2}$ en θ_{a_0} dat de volgende uitdrukkingen equivalent zijn:

1. $P(\theta_a \leq \theta) \rightarrow P(\theta_{a_0} \leq \theta)$ voor alle continue punten van $\theta \mapsto P(\theta_{a_0} \leq \theta)$;
2. $E\lambda(\theta_a) \rightarrow E\lambda(\theta_{a_0})$ voor alle begrensde, continue functies λ .

Door het aantonen van de eerste voorwaarde, kan uit de tweede voorwaarde worden afgeleid dat de uitdrukking in Vergelijking (23) naar nul convergeert wanneer $n_a \rightarrow \infty$.

B.3.2 De eerste voorwaarde van het Lemma van Portmanteau aantonen

In deze paragraaf wordt aangetoond dat voor de stochastische vectoren $\hat{\theta}_{a|M_2}$ en θ_{a_0} voor alle continue punten van $\theta \mapsto P(\theta_{a_0} \leq \theta)$ er geldt dat

$$P(\theta_a \leq \theta) \rightarrow P(\theta_{a_0} \leq \theta).$$

$P(\theta_{a_0} \leq \theta)$ wordt gegeven door

$$P(\theta_{a_0} \leq \theta) = \begin{cases} 1 & \text{als voor elke component geldt } \theta \geq \theta_{a_0} \\ 0 & \text{als minstens één component voldoet aan } \theta < \theta_{a_0}. \end{cases} \quad (24)$$

Volgens het Lemma van Portmanteau hoeft de eerste voorwaarde zoals gegeven in het eerste deel van Paragraaf B.3 slechts te gelden voor alle continue punten van $\theta \mapsto P(\theta_{a_0} \leq \theta)$. De continue punten van $P(\theta_{a_0} \leq \theta)$ zijn alle punten waarvoor in elke component geldt dat $\theta \neq \theta_{a_0}$. Vanuit de definitie van $P(\theta_{a_0} \leq \theta)$ geldt dit wanneer $P(\theta_{a_0} \leq \theta) = 1$ of $P(\theta_{a_0} \leq \theta) = 0$ én in beide gevallen elke component voldoet aan $\theta \neq \theta_{a_0}$.

Het geval dat $P(\theta_{a_0} \leq \theta) = 1$ én voor elk component geldt $\theta \neq \theta_{a_0}$

Neem aan dat $P(\theta_{a_0} \leq \theta) = 1$. We willen aantonen dat wanneer $n_a \rightarrow \infty$, de kans dat $P(\theta_a \leq \theta)$ convergeert naar 1. Dit bewijs bestaat uit vier stappen. In de eerste stap wordt er aangetoond dat voor alle $\delta > 0$ geldt

$$P(\hat{\theta}_{a|M_2} \leq \theta) \geq 1 - \delta,$$

wanneer n_a groot genoeg.

Er is aangenomen dat $P(\theta_{a_0} \leq \theta) = 1$ en elke component voldoet aan $\theta \neq \theta_{a_0}$. Vanuit de definitie in Vergelijking (24) blijkt dat hieruit volgt dat voor elke component geldt $\theta_{a_0} < \theta$. Dus bestaat er een $\epsilon > 0$ zodat $\theta_{a_0} + 3\epsilon < \theta$ in elk component.

Er is aangenomen dat $\hat{\theta}_{a|M_2}$ een consistente schatter is voor θ_{a_0} . Hieruit volgt dat voor elke $\epsilon > 0$ geldt $P(d(\hat{\theta}_{a|M_2}, \theta_{a_0}) > \epsilon) \rightarrow 0$ wanneer n_a groot genoeg. In andere woorden, als n_a groot genoeg, dan geldt $\forall \epsilon > 0, \forall \delta > 0$:

$$\begin{aligned} P(d(\hat{\theta}_{a|M_2}, \theta_{a_0}) > \epsilon) &\leq \delta \\ &\Leftrightarrow \\ P(d(\hat{\theta}_{a|M_2}, \theta_{a_0}) \leq \epsilon) &\geq 1 - \delta. \end{aligned}$$

Kies een $\epsilon > 0$ zodat $\theta_{a_0} + 3\epsilon < \theta$. Per component geldt dan, wanneer n_a groot genoeg,

$$\begin{aligned}
P(|\hat{\theta}_{a|M_2} - \theta_{a_0}| \leq \epsilon) &\geq 1 - \delta \\
&\implies \\
P(\hat{\theta}_{a|M_2} - \theta_{a_0} \leq \epsilon) &\geq 1 - \delta \\
&\iff \\
P(\hat{\theta}_{a|M_2} - \theta_{a_0} + 2\epsilon \leq 3\epsilon) &\geq 1 - \delta \\
&\iff \\
P(\hat{\theta}_{a|M_2} + 2\epsilon \leq \theta_{a_0} + 3\epsilon) &\geq 1 - \delta \\
&\implies \\
P(\hat{\theta}_{a|M_2} + 2\epsilon \leq \theta) &\geq 1 - \delta.
\end{aligned}$$

Dus geldt er voor alle $\epsilon > 0$ en voor alle $\delta > 0$ dat er een $n_a > 0$ bestaat die groot genoeg is, zodat $P(\hat{\theta}_{a|M_2} + 2\epsilon \leq \theta) \geq 1 - \delta$.

In de tweede stap wordt er gekeken naar verdelingen van algemene stochastische vectoren. Laat X een stochastische vector zijn die verdeeld is volgens $X \sim \mathcal{N}(0, I_{\theta_{a_0}}^{-1})$, met $I_{\theta_{a_0}}^{-1}$ de inverse van de Fisher Informatie Matrix van θ_{a_0} . Dan is er $\forall \delta > 0$ een $M > 0$, zodat

1. $P(\|X\| > M) \leq \frac{\delta}{2}$;
2. $P(\|X\| \leq M) \geq 1 - \frac{\delta}{2}$.

Kies voor $\delta > 0$ een $M > 0$ zodat bovenstaande ongelijkheid voldoet. Laat Y een stochastische vector zijn die verdeeld is volgens $Y \sim \mathcal{N}(0, I_{\hat{\theta}_{a|M_2}}^{-1})$. Er is aangenomen dat de Fisher Informatie Matrix een continue afbeelding is op heel Θ . Omdat $\hat{\theta}_{a|M_2} \xrightarrow{P_{\theta}^n} \theta_{a_0}$, geldt wegens de Continue Afbeeldingstelling dat $I_{\hat{\theta}_{a|M_2}}^{-1} \xrightarrow{P_{\theta}^n} I_{\theta_{a_0}}^{-1}$. Hieruit volgt voor de vastgestelde $M > 0$, dat als n_a groot genoeg is, dat:

1. $P(\|Y\| > 2M) \leq \delta$;
2. $P(\|Y\| \leq 2M) \geq 1 - \delta$.

In de derde stap wordt aangetoond dat de kans dat θ_a en $\hat{\theta}_{a|M_2}$ meer dan ϵ uit elkaar liggen, naar nul convergeert. De Bernstein–von Mises Theorie geeft

dat θ_a asymptotisch normaal verdeeld is volgens $\theta_a \sim \mathcal{N}(\hat{\theta}_{a|M_2}, \frac{1}{n_a} I_{\hat{\theta}_{a|M_2}}^{-1})$. Na het centreren in $\hat{\theta}_{a|M_2}$ en schalen met $\sqrt{n_a}$ geldt

$$\sqrt{n_a}(\theta_a - \hat{\theta}_{a|M_2}) \sim \mathcal{N}(0, I_{\hat{\theta}_{a|M_2}}^{-1}).$$

Uit stap 2 volgt dat er voor alle $\delta > 0$ een $M > 0$ bestaat zodat

$$P(\sqrt{n_a} \left\| \theta_a - \hat{\theta}_{a|M_2} \right\| > 2M) \leq \delta.$$

En dus

$$P\left(\left\| \theta_a - \hat{\theta}_{a|M_2} \right\| > \frac{2M}{\sqrt{n_a}}\right) \leq \delta.$$

Kies nu $n_a > 0$ zo groot dat $\frac{2M}{\sqrt{n_a}} < \epsilon$:

$$P\left(\left\| \theta_a - \hat{\theta}_{a|M_2} \right\| > \epsilon\right) \leq \delta. \quad (25)$$

Hieruit volgt dat de kans convergeert naar nul.

In de laatste stap wordt aangetoond dat $P(\theta_a \leq \theta)$ naar 1 convergeert, ofwel $P(\theta_a \leq \theta) \geq 1 - \delta$ voor alle $\delta > 0$, als n_a groot genoeg is. Laat $\delta > 0$. Wegens de driehoeksongelijkheid geldt

$$\begin{aligned} P(\theta_a \leq \theta) &= P(\theta_a - \hat{\theta}_{a|M_2} + \hat{\theta}_{a|M_2} \leq \theta) \\ &\geq P(\theta_a - \hat{\theta}_{a|M_2} + \hat{\theta}_{a|M_2} \leq \theta, \left\| \theta_a - \hat{\theta}_{a|M_2} \right\| > \epsilon) \\ &\quad + P(\theta_a - \hat{\theta}_{a|M_2} + \hat{\theta}_{a|M_2} \leq \theta, \left\| \theta_a - \hat{\theta}_{a|M_2} \right\| \leq \epsilon) \\ &\geq P(\theta_a - \hat{\theta}_{a|M_2} + \hat{\theta}_{a|M_2} \leq \theta, \left\| \theta_a - \hat{\theta}_{a|M_2} \right\| \leq \epsilon). \end{aligned}$$

De laatste ongelijkheid volgt uit dat een kans altijd een waarde tussen 0 en 1 aanneemt.

Als $\left\| \theta_a - \hat{\theta}_{a|M_2} \right\| \leq \epsilon$, geldt $\theta_a - \hat{\theta}_{a|M_2} \leq \epsilon$ voor elke component. Hieruit volgt

$$\theta_a - \hat{\theta}_{a|M_2} + \hat{\theta}_{a|M_2} \leq \epsilon + \hat{\theta}_{a|M_2}$$

en dus geldt er

$$\begin{aligned} P(\theta_a \leq \theta) &\geq P\left(\theta_a - \hat{\theta}_{a|M_2} + \hat{\theta}_{a|M_2} \leq \theta, \left\| \theta_a - \hat{\theta}_{a|M_2} \right\| \leq \epsilon\right) \\ &\geq P\left(\epsilon + \hat{\theta}_{a|M_2} \leq \theta, \left\| \theta_a - \hat{\theta}_{a|M_2} \right\| \leq \epsilon\right). \end{aligned}$$

In Vergelijking (25) is aangetoond dat $P\left(\left\|\theta_a - \hat{\theta}_{a|M_2}\right\| > \epsilon\right)$ kleiner of gelijk is aan δ . Dus geldt er

$$\begin{aligned} P\left(\epsilon + \hat{\theta}_{a|M_2} \leq \theta\right) &= P\left(\epsilon + \hat{\theta}_{a|M_2} \leq \theta, \left\|\theta_a - \hat{\theta}_{a|M_2}\right\| \leq \epsilon\right) \\ &\quad + P\left(\epsilon + \hat{\theta}_{a|M_2} \leq \theta, \left\|\theta_a - \hat{\theta}_{a|M_2}\right\| > \epsilon\right) \\ &\leq P\left(\epsilon + \hat{\theta}_{a|M_2} \leq \theta, \left\|\theta_a - \hat{\theta}_{a|M_2}\right\| \leq \epsilon\right) + \delta. \end{aligned}$$

Hieruit volgt door $\delta > 0$ naar de andere kant te halen,

$$P(\epsilon + \hat{\theta}_{a|M_2} \leq \theta, \left\|\theta_a - \hat{\theta}_{a|M_2}\right\| \leq \epsilon) \geq P(\epsilon + \hat{\theta}_{a|M_2} \leq \theta) - \delta.$$

Dus geldt er

$$\begin{aligned} P(\theta_a \leq \theta) &\geq P\left(\epsilon + \hat{\theta}_{a|M_2} \leq \theta, \left\|\theta_a - \hat{\theta}_{a|M_2}\right\| \leq \epsilon\right) \\ &\geq P(\epsilon + \hat{\theta}_{a|M_2} \leq \theta) - \delta. \end{aligned} \tag{26}$$

In de eerste stap is aangetoond dat voor alle $\delta > 0$ er een $n_a > 0$ bestaat die groot genoeg is, zodat $P(\hat{\theta}_{a|M_2} \leq \theta) \geq 1 - \delta$, ofwel

$$P(\epsilon + \hat{\theta}_{a|M_2} < \theta) \geq 1 - \delta.$$

Hieruit volgt dat voor n_a groot genoeg, er geldt

$$\left[P(\epsilon + \hat{\theta}_{a|M_2} \leq \theta) - \delta\right] \geq 1 - 2\delta.$$

Nu volgt uit Vergelijking (26) dat $P(\theta_a \leq \theta) \geq 1 - 2\delta$. Omdat $\delta > 0$ willekeurig gekozen was, geldt dat $P(\theta_a \leq \theta) \rightarrow 1$, bij aanname van $P(\theta_{a_0} \leq \theta) = 1$ en elke component voldoet aan $\theta \neq \theta_{a_0}$.

Het geval dat $P(\theta_{a_0} \leq \theta) = 0$ én voor elk component geldt $\theta \neq \theta_{a_0}$

Neem aan dat $P(\theta_{a_0} \leq \theta) = 0$. We willen aantonen dat wanneer $n_a \rightarrow \infty$, de kans dat $P(\theta_a \leq \theta)$ convergeert naar 0. Dit bewijs bestaat uit vier stappen. De eerste stap toont aan dat als $P(\theta_{a_0} \leq \theta) = 0$, er voor alle $\delta > 0$ geldt

$$P(\hat{\theta}_{a|M_2} - 2\epsilon \leq \theta) \leq \delta$$

wanneer n groot genoeg.

Er is aangenomen dat $P(\theta_{a_0} \leq \theta) = 0$. Er is dus minimaal één component waarvoor geldt $\theta_{a_0} > \theta$. Stel dat voor component i geldt $\theta_{a_0} > \theta$. Voor

dit component is er een $\epsilon > 0$, zodat $\theta_{a_0} > \theta + 3\epsilon$. Wegens de consistentie van $\hat{\theta}_{a|M_2}$ voor θ_{a_0} geldt voor component i dat $\forall \epsilon > 0, \forall \delta > 0$, wanneer n_a groot genoeg:

$$\begin{aligned}
 P(|\theta_{a_0} - \hat{\theta}_{a|M_2}| \leq \epsilon) &\geq 1 - \delta \\
 &\implies \\
 P(\theta_{a_0} - \hat{\theta}_{a|M_2} \leq \epsilon) &\geq 1 - \delta \\
 &\implies \\
 P(\theta + 3\epsilon - \hat{\theta}_{a|M_2} \leq \epsilon) &\geq 1 - \delta \\
 &\iff \\
 P(\theta + 2\epsilon \leq \hat{\theta}_{a|M_2}) &\geq 1 - \delta \\
 &\iff \\
 P(\hat{\theta}_{a|M_2} < \theta + 2\epsilon) &\leq \delta.
 \end{aligned}$$

Dus geldt er voorvoor component i dat voor alle $\epsilon > 0$, voor alle $\delta > 0$ en voor alle $n_a > 0$ groot genoeg geldt dat $P(\hat{\theta}_{a|M_2} \leq \theta + 2\epsilon) \leq \delta$. Omdat i een willekeurig component is waarin $\theta_{a_0} > \theta$, geldt voor elk component waarin $\theta_{a_0} > \theta$ dat voor alle $\epsilon > 0$, voor alle $\delta > 0$ en voor alle $n_a > 0$ groot genoeg geldt

$$P(\hat{\theta}_{a|M_2} \geq \theta + 2\epsilon) \geq 1 - \delta. \quad (27)$$

Er is aan het begin van deze paragraaf bewezen dat, wanneer er geldt $P(\theta_{a_0} \leq \theta) = 1$ en elke component voldoet aan $\theta \neq \theta_{a_0}$, er voor alle $\epsilon > 0$, voor alle $\delta > 0$ en alle n_a groot genoeg geldt

$$P\left(\left|\theta_a - \hat{\theta}_{a|M_2}\right| > \epsilon\right) \leq \delta. \quad (28)$$

In het bewijs dat Vergelijking (28) geldt, wordt de aanname $P(\theta_{a_0} \leq \theta) = 1$ en elke component voldoet aan $\theta \neq \theta_{a_0}$ niet gebruikt. Dit betekent dat in het algemeen geldt dat Vergelijking (28) voldoet, dus ook nu er is aangenomen dat $P(\theta_{a_0} \leq \theta) = 0$ en elke component voldoet aan $\theta \neq \theta_{a_0}$.

Met de resultaten uit Vergelijking (27) en Vergelijking (28) kan in de vierde stap worden bewezen dat $P(\theta_a \leq \theta)$ naar 0 convergeert, ofwel $P(\theta_a \leq \theta) \leq \delta$, voor alle $\delta > 0$ wanneer n_a groot genoeg.

Laat voor component i gelden $(\theta_{a_0})_i > (\theta)_i$. Uit Vergelijking (27) en Vergelijking (28) volgt dan voor dit i -de component dat er een n_a bestaat die

groot genoeg is, zodat

$$\begin{aligned}
P\left(\left(\hat{\theta}_{a|M_2}\right)_i \geq \theta_i + 2\epsilon \text{ en } \left\|\theta_a - \hat{\theta}_{a|M_2}\right\| \leq \epsilon\right) &\geq 1 - 2\delta \\
&\implies \\
P\left(\left(\hat{\theta}_{a|M_2}\right)_i \geq \theta_i + 2\epsilon \text{ en } \left|(\theta_a - \hat{\theta}_{a|M_2})_i\right| \leq \epsilon\right) &\geq 1 - 2\delta \\
&\implies \\
P\left(\left(\hat{\theta}_{a|M_2}\right)_i \geq \theta_i + 2\epsilon \text{ en } (\hat{\theta}_{a|M_2} - \theta_a)_i \leq \epsilon\right) &\geq 1 - 2\delta \\
&\implies \\
P\left((\theta_a)_i + \epsilon \geq \left(\hat{\theta}_{a|M_2}\right)_i \geq \theta_i + 2\epsilon\right) &\geq 1 - 2\delta \\
&\implies \\
P\left((\theta_a)_i \geq \theta_i + \epsilon\right) &\geq 1 - 2\delta,
\end{aligned}$$

voor alle $\delta > 0$ en alle $\epsilon > 0$. Met andere woorden, voor alle componenten i waarvoor $\theta_{a_0} > \theta$, geldt voor alle $\delta > 0$ en voor alle n_a groot genoeg dat

$$P(\theta_a < \theta + \epsilon) \leq \delta. \quad (29)$$

Hieruit volgt dat de kans $P(\theta_a \leq \theta)$ convergeert naar nul in dit i -de component.

Wanneer voor minstens één component van een stochastische vector geldt dat $P(\theta_a \leq \theta)$ convergeert naar nul, dan geldt voor de hele vector ook

$$P(\theta_a \leq \theta).$$

Met het samenvoegen van de twee bovenstaande gevallen is de eerste voorwaarde van het Lemma van Portmanteau aangetoond. Er geldt voor alle continue punten

$$P(\theta_a \leq \theta) \rightarrow P(\theta_{a_0} \leq \theta).$$

B.3.3 Een conclusie uit de tweede voorwaarde van het Lemma van Portmanteau

In Paragraaf B.3.2 is de eerste voorwaarde van het Lemma van Portmanteau aangetoond voor de stochastische vectoren θ_a en θ_{a_0} . Het lemma geeft dan dat wegens de zwakke convergentie ook de tweede voorwaarde geldt. Er geldt voor alle begrensde, continue functies λ dat $E\lambda(\theta_a) \rightarrow E\lambda(\theta_{a_0})$.

$\lambda_{cs}(\theta_a|X_{n_a}, M_2)$ is continu op heel Θ_a , omdat beide likelihood ratio functies continu zijn op heel Θ_a . $\lambda_{cs}(\theta_a|X_{n_a}, M_2)$ is begrensd wegens de aanname. Dus de functie $\lambda_{cs}(\theta_a|X_{n_a}, M_2)$ voldoet aan begrensdheid en continuïteit en dus geldt volgens de tweede voorwaarde van het Lemma van Portmanteau,

$$E\lambda_{cs}(\theta_a|X_{n_a}, M_2) \rightarrow E\lambda_{cs}(\theta_{a_0}|X_{n_a}, M_2), \quad (30)$$

wanneer $n_a \rightarrow \infty$. Verder geldt dat $E\lambda(\theta_a)$ is gegeven door

$$E\lambda(\theta_a) = \int \lambda_{cs}(\theta_a|X_{n_a}, M_2) d\Phi(\theta_a|X_{n_a}, M_2).$$

$E\lambda(\theta_{a_0})$ is gegeven door

$$E\lambda(\theta_{a_0}) = \int \lambda_{cs}(\theta_a|X_{n_a}, M_2) d\delta_{\theta_{a_0}}(\theta_a) = \lambda_{cs}(\theta_{a_0}).$$

De convergentie uit Vergelijking (30) kan ook worden geschreven als

$$\int \lambda_{cs}(\theta_a|x_b, x_c) d\Phi(\theta_a|X_{n_a}, M_2) \rightarrow \int \lambda_{cs}(\theta_a|x_b, x_c) d\delta_{\theta_{a_0}}(\theta_a).$$

Hieruit volgt dat wanneer $n_a \rightarrow \infty$,

$$\left| \int \lambda_{cs}(\theta_a|x_b, x_c) d[\Phi(\theta_a|X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right| \rightarrow 0. \quad (31)$$

B.4 Conclusie Theorie 1

In het begin van het bewijs voor Theorie 1 is gesteld dat

$$|\beta_{cs}(X_{n_a}) - \lambda_{cs}(\theta_{a_0}|x_b, x_c)|$$

van boven begrensd is:

$$\begin{aligned} & |\beta_{cs}(X_{n_a}) - \lambda_{cs}(\theta_{a_0}|x_b, x_c)| \leq \\ & \left| \int \lambda_{cs}(\theta_a|x_b, x_c) d[\Pi(\theta_a|X_{n_a}, M_2) - \Phi(\theta_a|X_{n_a}, M_2)] \right| \\ & + \left| \int \lambda_{cs}(\theta_a|x_b, x_c) d[\Phi(\theta_a|X_{n_a}, M_2) - \delta_{\theta_{a_0}}(\theta_a)] \right|. \end{aligned}$$

In Paragraaf B.2 is met behulp van de Bernstein–von Mises aangetoond dat het eerste deel van deze bovengrens convergeert naar nul. Het resultaat staat in Vergelijking (22). In Paragraaf B.3 is met behulp van het Lemma van Portmanteau aangetoond dat het tweede deel van deze bovengrens convergeert naar nul. Dit resultaat staat in Vergelijking (31). Hieruit volgt dat de absolute afstand tussen $\beta_{cs}(X_{n_a})$ en $\lambda_{cs}(\theta_{a_0}|x_b, x_c)$ ook naar nul convergeert. Hiermee is Theorie 1 bewezen. \square

C Bewijs van Lemma 3.1

Ter ondersteuning van het bewijs van Lemma 3.1, Lemma 3.2 en Theorie 3 worden hier de notaties gegeven. Laat $P_{\theta_a}^{n_a}$ de gezamenlijke kansmaat zijn op X_{a,n_a} voor alle $\theta_a \in \Theta_a$, inclusief θ_{a_0} . Voor het specifieke-bron probleem, laat $P_{\theta_b}^{n_b}$ de gezamenlijke kansmaat zijn op X_{b,n_b} voor alle $\theta_b \in \Theta_b$, inclusief θ_{b_0} .

Laat vervolgens X_n de stochastische variabele noteren die hoort bij het genereren van de gehele databank onder een van beide model selectie problemen. Voor het gemeenschappelijke-bron probleem geldt dat X_{a,n_a} de observaties van de populatie van bronnen genereert en dat de observaties x_b en x_c vast zijn. Voor het specifieke-bron probleem geldt dat X_{a,n_a} de observaties van de grote populatie van bronnen genereert en dat X_{b,n_b} de observaties van de populatie van het specifieke subject genereert. Ook geldt dat de observatie x_c vast is.

Laat P_θ^n de gezamenlijke kansmaat op X_n zijn voor alle $\theta \in \Theta$, inclusief θ_0 waar Θ de gezamenlijke parameter ruimte is. Definieer tot slot

$$\gamma_{n|M_2}^2 = \lambda'(\hat{\theta}_{n|M_2})^T I_{\hat{\theta}_{n|M_2}}^{-1} \lambda'(\hat{\theta}_{n|M_2})$$

Lemma 3.1, Lemma 3.2 en Theorie 3 zijn voor het gemeenschappelijke-bron probleem en het specifieke-bron probleem aan elkaar gelijk. De lemma's en de theorie zijn dan ook gedefinieerd met algemene notaties. Hierin geldt dat $\lambda(\theta)$ de likelihood ratio is voor zowel het gemeenschappelijke-bron probleem als het specifieke-bron probleem. Verder geldt dat $\hat{\theta}_{n|M_2}$ de MLE is, berekend aan de hand van de gehele dataverzameling X_n onder het tweede model. Tot slot is $\beta(X_n)$ de Bayes Factor van zowel het specifieke- als het gemeenschappelijke-bron probleem.

Lemma 3.1 *Stel voor een gegeven observatie van de onbekende-bron verzameling(en) van observaties dat de likelihood ratio functie, $\lambda(\theta)$, is twee keer continu differentieerbaar, en stel dat $\hat{\theta}_{n|M_2}$ een consistente schatter voor θ_0 is onder P_θ^n -kans. Als de aannames van de Bernstein-von Mises Theorie gelden en $n \rightarrow \infty$, dan*

$$\left\| \Pi(\lambda(\theta)|X_n, M_2) - \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \frac{1}{n} \gamma_{n|M_2}^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Het bewijs van Lemma 3.1 bestaat uit de convergentie van de restterm van de Taylorreeks van $\lambda(\theta)$ en de convergentie van de totale variatie norm van twee verdelingen met behulp van de Bernstein-von Mises Theorie. Deze bewijzen volgen in Paragraaf C.1 en respectievelijk Paragraaf C.2.

C.1 Convergentie van de restterm van de Taylorreeks

Het bewijs van Lemma 3.1 is gebaseerd op de Taylorreeks expansie van $\lambda(\theta)$ in de maximum likelihood schatter $\hat{\theta}_{n|M_2}$. Deze Taylorreeks wordt gegeven door

$$\lambda(\theta) = \lambda(\hat{\theta}_{n|M_2}) + (\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}) + \frac{1}{2}(\theta - \hat{\theta}_{n|M_2})^T \lambda''(\tilde{\theta}_{n|M_2})(\theta - \hat{\theta}_{n|M_2}).$$

Hierin is $\tilde{\theta}_{n|M_2}$ een waarde op de lijn tussen θ en $\hat{\theta}_{n|M_2}$, λ' de vector van de eerste partiële afgeleiden van λ en λ'' de matrix met tweede partiële afgeleiden van λ . De Taylorreeks kan ook worden geschreven als

$$\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2}) = (\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}) + \frac{1}{2}(\theta - \hat{\theta}_{n|M_2})^T \lambda''(\tilde{\theta}_{n|M_2})(\theta - \hat{\theta}_{n|M_2}).$$

Beide kanten vermenigvuldigen met \sqrt{n} geeft

$$\begin{aligned} \sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2})) \\ = \sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}) + \frac{\sqrt{n}}{2}(\theta - \hat{\theta}_{n|M_2})^T \lambda''(\tilde{\theta}_{n|M_2})(\theta - \hat{\theta}_{n|M_2}). \end{aligned}$$

De foutterm van deze expansie wordt gegeven door

$$\frac{1}{2} \left[\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) \right]^T \left[\frac{1}{\sqrt{n}} \lambda''(\tilde{\theta}_{n|M_2}) \right] \left[\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) \right]. \quad (32)$$

Met een resultaat van de Bernstein-von Mises Theorie, de aannames, de Continue Afbeelding Stelling en de Sandwich Theorie kan worden aangetoond dat de foutterm van de Taylorreeks convergeert naar nul in P_θ^n -kans wanneer $n \rightarrow \infty$.

Wegens de Bernstein-von Mises Theorie geldt als $n \rightarrow \infty$,

$$\left\| \Pi \left(\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) | X_n, M_2 \right) - N \left(0, I_{\theta_0}^{-1} \right) \right\|_{TV} \xrightarrow{P} 0.$$

Wegens Prohorov's Theorie, zie Paragraaf 4.5.2, geldt er nu dat $\{\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) : n \in \mathbb{N}\}$ begrensd is in P_θ^n -kans.

De aanname dat $\hat{\theta}_{n|M_2}$ een consistente schatter is, geeft $\hat{\theta}_{n|M_2} \xrightarrow{P_\theta^n} \theta_0$ als $n \rightarrow \infty$. Er geldt ook $\theta \xrightarrow{P_\theta^n} \theta_0$ als $n \rightarrow \infty$.

Omdat $\tilde{\theta}_{n|M_2}$ op de lijn tussen θ en $\hat{\theta}_{n|M_2}$ ligt, geldt wegens de Insluitstelling dat

$$\tilde{\theta}_{n|M_2} \xrightarrow{P_\theta^n} \theta_0$$

als $n \rightarrow \infty$. Omdat $\lambda(\theta)$ continu is voor elk punt, geldt wegens de Continue Afbeelding Stelling dat als $n \rightarrow \infty$,

$$\lambda(\tilde{\theta}_{n|M_2}) \xrightarrow{P_\theta^n} \lambda(\theta_0).$$

Er is aangenomen dat $\lambda(\theta)$ twee keer continu differentieerbaar is. Uit de Continue Afbeeldingstelling geldt dan

$$\lambda''(\tilde{\theta}_{n|M_2}) \xrightarrow{P_\theta^n} \lambda''(\theta_0),$$

als $n \rightarrow \infty$. Omdat θ_0 een onbekende, maar vaste waarde is, geldt dat $\lambda''(\theta_0)$ ook een vaste waarde is. Dus geldt er

$$\frac{1}{\sqrt{n}} \lambda''(\tilde{\theta}_{n|M_2}) \xrightarrow{P_\theta^n} 0$$

als $n \rightarrow \infty$.

Nu is er afgeleid dat de $\left[\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) \right]$ termen begrensd zijn in P_θ^n -kans en dat de $\left[\frac{1}{\sqrt{n}} \lambda''(\tilde{\theta}_{n|M_2}) \right]$ term in P_θ^n kans convergeert naar nul wanneer $n \rightarrow \infty$. In het algemeen geldt dat wanneer rij termen in kans convergeert naar nul en een andere rij termen in kans begrensd is, dat het product van deze rijen ook in kans convergeert naar nul [10]. De foutterm van de Taylorreeks gegeven in Vergelijking (32) is een product van twee in P_θ^n -kans begrensde termen en een term die in P_θ^n -kans convergeert naar nul als $n \rightarrow \infty$. Hieruit volgt dat de gehele foutterm gegeven in Vergelijking (32) in P_θ^n -kans naar nul convergeert wanneer $n \rightarrow \infty$. Dus als $n \rightarrow \infty$,

$$\left(\frac{1}{2} \left[\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) \right]^T \left[\frac{1}{\sqrt{n}} \lambda''(\tilde{\theta}_{n|M_2}) \right] \left[\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) \right] \right) \xrightarrow{P_\theta^n} 0.$$

C.2 Bernstein–von Mises Theorie

Er is aangenomen dat de aannames van de Bernstein–von Mises Theorie gelden. Hieruit volgt dat θ asymptotisch normaal verdeeld is;

$$\left\| \Pi(\theta|X_n, M_2) - \mathcal{N}\left(\hat{\theta}_{n|M_2}, \frac{1}{n} I_{\hat{\theta}_{n|M_2}}^{-1}\right) \right\|_{TV} \xrightarrow{P_\theta^n} 0, \text{ als } n \rightarrow \infty.$$

Opnieuw centreren en schalen geeft

$$\left\| \Pi\left(\sqrt{n}(\theta - \hat{\theta}_{n|M_2})|X_n, M_2\right) - \mathcal{N}\left(0, I_{\hat{\theta}_{n|M_2}}^{-1}\right) \right\|_{TV} \xrightarrow{P_\theta^n} 0, \text{ als } n \rightarrow \infty.$$

In het algemeen geldt voor een stochastische vector $X \sim \mathcal{N}(\mu, \Sigma)$ en een stochastische vector $Y = BX$, met B een constante $M \times N$ matrix en X een $N \times 1$ vector, dat $Y \sim \mathcal{N}(B\mu, B\Sigma B^T)$.

Voor de vector $\lambda'(\hat{\theta}_{n|M_2})^T$ geldt

$$\lambda'(\hat{\theta}_{n|M_2})^T \sqrt{n}(\theta - \hat{\theta}_{n|M_2}) = \sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}).$$

Uit deze transformatie en de verdeling van $\sqrt{n}(\theta - \hat{\theta}_{n|M_2})$ volgt als $n \rightarrow \infty$,

$$\left\| \Pi \left(\sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}) | X_n, M_2 \right) - \mathcal{N} \left(0, \lambda'(\hat{\theta}_{n|M_2})^T I_{\hat{\theta}_{n|M_2}}^{-1} \lambda'(\hat{\theta}_{n|M_2}) \right) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Uit de definitie van $\gamma_{n|M_2}^2$ volgt dat dit te schrijven is als

$$\left\| \Pi \left(\sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}) | X_n, M_2 \right) - \mathcal{N} \left(0, \gamma_{n|M_2}^2 \right) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Hierin is $\Pi \left(\sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2}) | X_n, M_2 \right)$ de posterior verdeling van $\sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2})$.

C.3 Conclusie Lemma 3.1

In Paragraaf C.1 is bewezen dat de restterm van de Taylorreeks van $\lambda(\theta)$ in $\hat{\theta}_{n|M_2}$ convergeert naar nul in P_θ^n -kans. In Paragraaf C.2 is bewezen dat $\sqrt{n}(\theta - \hat{\theta}_{n|M_2})^T \lambda'(\hat{\theta}_{n|M_2})$ asymptotisch normaal verdeeld is met gemiddelde 0 en variantie $\gamma_{n|M_2}^2$. Uit Slutsky's Lemma volgt dan

$$\left\| \Pi \left(\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2})) | X_n, M_2 \right) - \mathcal{N}(0, \gamma_{n|M_2}^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0$$

Schalen met \sqrt{n} en centreren in $\lambda(\theta)$ geeft

$$\left\| \Pi(\lambda(\theta) | X_n, M_2) - \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \frac{1}{n} \gamma_{n|M_2}^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Hiermee is Lemma 3.1 bewezen. \square

D Bewijs van Lemma 3.2

Voor het bewijs van Lemma 3.2 wordt de notatie gehanteerd die staat beschreven aan het begin van Appendix C. Ook wordt dezelfde definitie van $\gamma_{n|M_2}^2$ gebruikt.

Lemma 3.2 *Stel voor een gegeven observatie van de onbekende-bron verzameling(en) van observaties dat $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 onder P_θ^n -kans en stel dat $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$. Stel dat de Fisher Informatie Matrix en continue afbeelding en positief definit is. Stel verder dat $\lambda'(\theta_0)$ ongelijk is aan de nulvector. Als de aannames van de Bernstein-von Mises Theorie gelden en $n \rightarrow \infty$, dan*

$$\left\| \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2) - \mathcal{N}(\beta(X_n), n\sigma_n^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

D.1 Convergentie van de Bayes Factor en de likelihood ratio functie

Uit Theorie 1 en Theorie 2 blijkt

$$\beta(X_n) \xrightarrow{P_\theta^n} \lambda(\theta_0),$$

als $n \rightarrow \infty$, voor zowel het gemeenschappelijke-bron- als het specifieke-bron probleem. Uit de consistentie van $\hat{\theta}_{n|M_2}$ volgt $\hat{\theta}_{n|M_2} \xrightarrow{P_\theta^n} \theta_0$. Wegens de Continue Afbeeldingstelling geldt voor de continue functie $\lambda(\theta)$ dat wanneer $n \rightarrow \infty$,

$$\lambda(\hat{\theta}_{n|M_2}) \xrightarrow{P_\theta^n} \lambda(\theta_0).$$

Hieruit volgt als $n \rightarrow \infty$,

$$\left| \beta(X_n) - \lambda(\hat{\theta}_{n|M_2}) \right| \xrightarrow{P_\theta^n} 0. \quad (33)$$

D.2 Convergentie van $\gamma_{n|M_2}^2$ en σ_n^2

Wegens de continuïteit van de Fisher Informatie Matrix uit de aanname en de consistentie van $\hat{\theta}_{n|M_2}$ voor θ_0 , blijkt uit de Continue Afbeeldingstelling dat

$$I_{\hat{\theta}_{n|M_2}}^{-1} \xrightarrow{P_\theta^n} I_{\theta_0}^{-1}, \text{ als } n \rightarrow \infty.$$

Uit de continuïteit van $\lambda'(\theta)$ en de consistentie van $\hat{\theta}_{n|M_2}$ volgt

$$\lambda'(\hat{\theta}_{a|M_2}) \xrightarrow{P_\theta^n} \lambda'(\theta_0), \text{ als } n \rightarrow \infty.$$

Uit de convergentie van de Fisher Informatie Matrix $I_{\hat{\theta}_{n|M_2}}^{-1}$ en de convergentie van $\lambda'(\hat{\theta}_{a|M_2})$ blijkt

$$\lambda'(\hat{\theta}_{a|M_2})^T I_{\hat{\theta}_{a|M_2}}^{-1} \lambda'(\hat{\theta}_{a|M_2}) \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0).$$

Omdat θ_0 een vaste, maar onbekende waarde is, geldt ook dat $\lambda(\theta_0)$, $\lambda'(\theta_0)$ en $\lambda''(\theta_0)$ vaste, maar onbekende waarden zijn en dat $I_{\theta_0}^{-1}$ uit vaste, onbekende waarden bestaat. Uit de voorwaarde dat I_{θ_0} positief definitief is volgt dat $I_{\theta_0}^{-1}$ positief definitief is. Omdat $\lambda'(\theta_0) \neq 0$ en omdat $I_{\theta_0}^{-1}$ positief definitief is, volgt dat $\lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ een constante, onbekende waarde $c > 0$ aanneemt. Wegens de definitie van $\gamma_{n|M_2}^2$ geldt voor deze onbekende, constante $c > 0$,

$$\gamma_{n|M_2}^2 \rightarrow c.$$

Er is aangenomen dat $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$. Dus geldt voor dezelfde constante $c > 0$ dat

$$n\sigma_n^2 \rightarrow c.$$

Hieruit volgt als $n \rightarrow \infty$,

$$|\gamma_{n|M_2}^2 - n\sigma_n^2| \xrightarrow{P_\theta^n} 0. \quad (34)$$

D.3 Conclusie Lemma 3.2

Uit Vergelijking (33) en Vergelijking (34) zal worden afgeleid dat wanneer $n \rightarrow \infty$,

$$\left\| \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2) - \mathcal{N}(\beta(X_n), n\sigma_n^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Uit de definitie van de totale variatie norm geldt dat de totale variatie norm uit bovenstaande convergentie gelijk is aan

$$\int \left| d\left(\Phi(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2) - \Phi(\beta(X_n), n\sigma_n^2) \right) \right|.$$

Hierin is $\Phi(\mu, \sigma^2)$ de kansdichtheidfunctie van de normaalverdeling $\mathcal{N}(\mu, \sigma^2)$.

Uit Vergelijking (33) en Vergelijking (34) blijkt dat $\Phi\left(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2\right) - \Phi(\beta(X_n), n\sigma_n^2)$ puntsgewijs convergeert naar nul. Door middel van de Dominante Convergence Theorie kan de convergentie in totale variatie norm worden aangetoond. Voor een $\epsilon > 0$ willekeurig, geldt

$$\left| \Phi\left(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2\right) - \Phi(\beta(X_n), n\sigma_n^2) \right| \leq \max_{[\mu-\epsilon, \mu+\epsilon], [\sigma-\epsilon, \sigma+\epsilon]} 2\Phi(\mu, \sigma^2),$$

Voor een algemene kansdichtheidfunctie $\Phi(\mu, \sigma^2)$ geldt dat deze exponentieel convergeert naar nul. Er volgt dan automatisch dat het maximum van deze $\Phi(\mu, \sigma^2)$ ook exponentieel naar nul convergeert en dus integreerbaar is. Dus bestaat er een dominante, integreerbare, positieve functie voor de afstand tussen de twee kansverdelingsfuncties. Uit de Dominante Convergence Theorie volgt dan,

$$\int \left| d\left(\Phi\left(\lambda(\hat{\theta}_{n|M_2}), \gamma_{n|M_2}^2\right) - \Phi(\beta(X_n), n\sigma_n^2)\right) \right| \rightarrow 0.$$

Hiermee is Lemma 3.2 bewezen. \square

E Bewijs van Theorie 3

Voor het bewijs van Theorie 3 wordt de notatie gehanteerd die staat beschreven aan het begin van Appendix C. Ook wordt dezelfde definitie van $\gamma_{n|M_2}^2$ gebruikt.

Theorie 3 (Benadering $1 - \alpha$ geloofwaardigheids interval voor de likelihood ratio)

Laat de aannames van Lemma 3.1 en Lemma 3.2 gelden en laat

$$\mathcal{I}_n = \beta(X_n) \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sigma_n,$$

waar $\beta(X_n)$ de rij van de Bayes Factoren van of de gemeenschappelijke-bron of de specifieke-bron representeert, $0 < \alpha < 1$ het gewenste significantie niveau is, $\sqrt{n}\sigma_n$ de rij van kwantielen $q_{\Phi(1),n}$ voor de posterior verdeling van $\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2}))$ is en Φ^{-1} de standaard normale kwantiel functie is. Als $n \rightarrow \infty$, dan

$$\Pi(\lambda(\theta) \in \mathcal{I}_n | X_n, M_2) \rightarrow 1 - \alpha.$$

In Theorie 3 is $\sqrt{n}\sigma_n$ gedefinieerd als rij van kwantielen $q_{\Phi(1),n}$ voor de posterior verdeling van de geschaalde likelihood ratio. Om Lemma 3.2 toe te passen met $\sqrt{n}\sigma_n$ gedefinieerd als rij van kwantielen, moet er worden aangetoond dat $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$. Dit wordt bewezen in Paragraaf E.1.

E.1 Convergentie van σ_n

Uit de Bernstein–von Mises Theorie blijkt dat $\sqrt{n}(\theta - \hat{\theta}_{n|M_2})$ asymptotisch is verdeeld volgens

$$\left\| \Pi \left(\sqrt{n}(\theta - \hat{\theta}_{n|M_2}) | X_n, M_2 \right) - \mathcal{N} \left(0, I_{\theta}^{-1} \right) \right\|_{TV} \xrightarrow{P_{\theta}^n} 0.$$

Uit de consistentie van $\hat{\theta}_{n|M_2}$ voor θ_0 en uit Appendix C.3 blijkt

$$\left\| \Pi \left(\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta})) | X_n, M_2 \right) - \mathcal{N} \left(0, \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0) \right) \right\|_{TV} \xrightarrow{P_{\theta}^n} 0. \quad (35)$$

Uit de achtergrond informatie in Paragraaf 5.3.2 blijkt dat hieruit ook geldt dat de kwantielen van de verdelingen uit bovenstaande convergentie convergeren. Dus er geldt dat $q_{\Phi(1),n}$ gedefinieerd op $\Pi \left(\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta})) | X_n, M_2 \right)$

convergeert naar $q_{\Phi(1)}$ gedefinieerd op $\mathcal{N}(0, \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0))$ wanneer $n \rightarrow \infty$. Uit de achtergrond informatie blijkt ook voor $q_{\Phi(1)}$ dat

$$q_{\Phi(1)} = 0 + \sqrt{\lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)}.$$

Uit de convergentie volgt dan

$$q_{\Phi(1),n} \rightarrow \sqrt{\lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)}, \text{ als } n \rightarrow \infty.$$

In Theorie 3 is $\sqrt{n}\sigma_n$ gedefinieerd als kwantiel $q_{\Phi(1),n}$ van de posterior verdeling Π zoals gegeven in Vergelijking (35). Dus voor $n\sigma_n^2$ geldt

$$n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0), \text{ als } n \rightarrow \infty.$$

E.2 Conclusie Theorie 3

Uit Lemma 3.1 weet je wanneer $n \rightarrow \infty$,

$$\left\| \Pi(\lambda(\theta)|X_n, M_2) - \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \frac{1}{n} \gamma_{n|M_2}^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Uit Lemma 3.2 weet je wanneer $n \rightarrow \infty$,

$$\left\| \mathcal{N}(\lambda(\hat{\theta}_{n|M_2}), \frac{1}{n} \gamma_{n|M_2}^2) - \mathcal{N}(\beta(X_n), \sigma_n^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0.$$

Hieruit volgt wanneer $n \rightarrow \infty$,

$$\left\| \Pi(\lambda(\theta)|X_n, M_2) - \mathcal{N}(\beta(X_n), \sigma_n^2) \right\|_{TV} \xrightarrow{P_\theta^n} 0. \quad (36)$$

Hierin is $\mathcal{N}(\beta(X_n), \sigma_n^2)$ de rij van kansmaten die hoort bij de normaalverdeling met $\beta(X_n)$ als gemiddelde en σ_n^2 als covariantie. Dus

$$\mathcal{I}_n = \beta(X_n) \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) \sigma_n$$

is een rij van bijbehorende geloofwaardigheids intervallen met $100(1 - \alpha)\%$ geloofwaardigheid. Uit Vergelijking (36) volgt dan dat

$$\Pi(\lambda(\theta) \in \mathcal{I}_n | X_n, M_2) \rightarrow 1 - \alpha,$$

wanneer $n \rightarrow \infty$. Hiermee is Theorie 3 bewezen. \square

F De aannames van de Theoriën en Lemma's

Voor het overzicht is in deze appendix een lijst van de aannames van de theoriën en lemma's opgenomen.

F.1 Aannames Theorie 1

De aannames van Theorie 1 zijn dat voor vaste observaties x_b en x_c , geldt dat

1. de likelihood ratio functie $\lambda_{cs}(\theta_a|x_b, x_c)$ begrensd is in een omgeving van θ_0 ;
2. de MLE $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 ;
3. de aannames van de Bernstein–von Mises Theorie voldoen.

F.2 Aannames Theorie 2

De aannames van Theorie 2 zijn dat voor vaste observatie x_c , geldt dat

1. de likelihood ratio functie $\lambda_{ss}(\theta|x_c)$ begrensd is in een omgeving van θ_{a_0} ;
2. de MLE $\hat{\theta}_{a_0}$ een consistente schatter is voor θ_{a_0} ;
3. de aannames van de Bernstein–von Mises Theorie voldoen.

F.3 Aannames Theorie 3

De aannames van Theorie 3 worden gegeven inclusief de aannames van Lemma 3.1 en Lemma 3.2. Voor een gegeven observatie van onbekende bron verzameling(en) van observaties, geldt dat

1. de likelihood ratio functie, $\lambda(\theta)$, twee keer continu differentieerbaar is;
2. de MLE $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 onder P_θ^n -kans;
3. $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$;
4. de Fisher Informatie Matrix een continue afbeelding is;
5. de Fisher Informatie Matrix positief definit is;
6. $\lambda'(\theta_0) \neq 0$;

7. de aannames van de Bernstein–von Mises Theorie voldoen;

Voor $\mathcal{I}_n = \beta(X_n) \pm \Phi^{-1}(1 - \frac{\alpha}{2})\sigma_n$ geldt verder dat

1. $\beta(X_n)$ de rij van Bayes Factoren van een van de bronnen;
2. $0 < \alpha < 1$ het gewenste significantie niveau;
3. $\sqrt{n}\sigma_n$ de rij van kwantielen $q_{\Phi(1),n}$ voor de posterior verdeling van $\sqrt{n}(\lambda(\theta) - \lambda(\hat{\theta}_{n|M_2}))$;
4. Φ^{-1} de standaard normaal kwantiel functie is.

F.4 Aannames Lemma 3.1

De aannames van Lemma 3.1 zijn dat voor een gegeven observatie van de onbekende bron verzameling(en) van observaties, geldt dat

1. de likelihood ratio functie, $\lambda(\theta)$, twee keer continu differentieerbaar is;
2. de MLE $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 onder P_{θ}^n -kans;
3. de aannames van de Bernstein–von Mises Theorie voldoen.

F.5 Aannames Lemma 3.2

De aannames van Lemma 3.2 zijn dat voor een gegeven observatie van de onbekende bron verzameling(en) van observaties, geldt dat

1. de MLE $\hat{\theta}_{n|M_2}$ een consistente schatter is voor θ_0 onder P_{θ}^n -kans;
2. $n\sigma_n^2 \rightarrow \lambda'(\theta_0)^T I_{\theta_0}^{-1} \lambda'(\theta_0)$ als $n \rightarrow \infty$;
3. de Fisher Informatie Matrix een continue afbeelding is;
4. de Fisher Informatie Matrix positief definit is;
5. $\lambda'(\theta_0) \neq 0$;
6. de aannames van de Bernstein–von Mises Theorie voldoen.

Referenties

- [1] C. G. G. Aitken and D. Lucy. Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 53(1):109–122, Januari 2004.
- [2] F. Bijma, M. Jonker, and A. van der Vaart. *An Introduction to Mathematical Statistics*. Amsterdam University Press B.V., Amsterdam, 2017.
- [3] F. Bijma, M. Jonker, and A. van der Vaart. *An Introduction to Mathematical Statistics*, chapter 3.4 Method of Moments Estimators, pages 72–73. Amsterdam University Press B.V., Amsterdam, first edition, 2017.
- [4] G. Grimmett and D. Welsh. *Probability: An Introduction*. Oxford University Press, New York, 2014.
- [5] D. M. Ommen and C. P. Saunders. Reconciling the Bayes Factor and Likelihood Ratio for Two Non-Nested Model Selection Problems. Januari 2019. <https://arxiv.org/abs/1901.09798>.
- [6] D. M. Ommen, C. P. Saunders, and C. Neumann. The characterization of Monte Carlo errors for the quantification of the value of forensic evidence. *Journal of Statistical Computation and Simulation*, 87(8):1608–1643, Januari 2017.
- [7] E. Parzen. Quantile Functions, Convergence in Quantile, and Extreme Value Distribution Theory. November 1980. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a093000.pdf>.
- [8] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge series. Cambridge University Press, first edition, 1998.
- [9] A. W. van der Vaart. *Asymptotic Statistics*, chapter 10. Bayes Procedures, pages 138–152. Cambridge University Press, Cambridge, first edition, 1998.
- [10] A. W. van der Vaart. *Asymptotic Statistics*, chapter 2. Stochastic Convergence, pages 5–24. Cambridge University Press, Cambridge, first edition, 1998.