Vrije Universiteit Brussel

Faculteit Wetenschappen
Vakgroep Wiskunde

Statistische schatters voor data van functionele magnetische resonantie

Bachelorproef II

Susanne Blotwijk

Promotor: Dr. Lieve Lauwers

Academiejaar 2014-2015

Inhoudsopgave

1 Inleiding **2**

2 De verdeling van de modulus **3**

3 Maximum likelihood-schatters **5**

4 Bayesiaanse Schatters **15**

5 De Koay-inversie techniek **22**

6 Vergelijking van de schatters **24**

A Grafieken **25**

7 Referenties **26**

1 Inleiding

Functionele magnetische resonantie of fMRI is een MRI-techniek die wordt gebruikt om hersenactiviteit te meten. Bij fMRI worden er meerdere MRI-beelden na elkaar gemaakt met vaste tijdsintervallen om op die manier veranderingen in het brein waar te nemen. Eén van de veranderingen die fMRI kan waarnemen is die van het zuurstofpercentage in het bloed en op die manier ook de stroming van het bloed [2]. Actieve delen van het brein vereisen meer zuurstof, dus wordt er meer zuurstofrijk bloed aangevoerd en op die manier kan aan de hand van fMRI getest worden welke delen van de hersenen geactiveerd worden en in welke situaties.
Het voordeel van fMRI is dat men dit kan doen zonder dat de testpersoon een operatie, radio-actieve straling of andere potentieel schadelijke procedures moet ondergaan, terwijl men toch zeer nauwkeurig de locatie van de activiteit kan vaststellen [2]. Het nadeel van deze methode is dat men maar een beperkt aantal metingen per foto kan uitvoeren en dat de intervallen tussen de foto's noodgedwongen ook vaak relatief lang zijn, meestal tussen de 2 á 3 seconden [6].

Om de data te verkrijgen wekt de MR-scanner een sterk magnetisch veld op zodanig dat het waterstof in de hersenen magnetiseerd en zich binnen dit veld gaat richten. Vervolgens wordt er een RF-signaal doorheen gestuurd om de richting van de atomen te verstoren. Zodra dit radiosignaal wegvalt, herschikken de verstoorde atomen zich, wat zorgt voor een stroming in de ontvangende spoelen. Deze stroming is het MR-signaal. Typisch zijn er 2 spoelen onder een hoek van 90 graden om het signaal op te vangen, wat er voor zorgt dat de data die gemeten wordt een complex getal is[2]. De fase van het getal wordt over het algemeen als tijdsvariërend beschouwd, d.w.z. zonder relevante signaal informatie. Bijgevolg wordt bij het merendeel van de modellen enkel de modulus bestudeerd. Afhankelijk van de grootte van de modulus wordt vervolgens bepaald of de regio die geassocieerd met dat datapunt actief is of niet. Het verschil tussen beide situaties is over het algemeen niet zeer groot, dus is het belangrijk dat de afwijking van de schatter ook klein is.

Een andere eis voor de schatter is dat deze het ook goed doet als de signaal-tot-ruisverhouding zeer laag is, aangezien deze bij fMRI-scans van de hersenen behoorlijk laag kan gaan, typisch van 0 tot 3 [1]. Verder moet de schatter ook voor kleine steekproefgroottes betrouwbaar zijn. De duur van het MR-signaal, en daarmee ook de beschikbare tijd om metingen uit te voeren, is beperkt. Omdat men ook graag een hoge resolutie heeft voor de beelden, is

het aantal voxels, de locaties waar men de metingen neemt, ook hoog, wat het aantal metingen per voxel per instantie verder beperkt. Meestal gaat het aantal scans tot 100 per instantie. [6]
Aangezien men een 3-dimensionaal beeld van de hersenen wil, scant men de doorsnede van het hoofd op verschillende hoogtes. Het volledige aantal voxels dat gemeten moet worden per scan, kan typisch rond de 64 x 64 x 30 liggen. Als men dan 100 metingen neemt per voxel, komt dat neer op meer dan 12 miljoen datapunten per MRI-beeld. De hoeveelheid data die men moet verwerken voor de volledige fMRI-scan is dus aanzienlijk. De hoeveelheid rekenwerk die een schatter met zich meebrengt is daarom ook niet onbelangrijk.
Kort samengevat, moet de schatter dus nauwkeurig genoeg zijn om kleine verschillen op te pikken bij kleine steekproefgroottes, ook wanneer er veel storing op het signaal zit en bij voorkeur ook nog met zo min mogelijk rekenwerk. Het is dus nuttig om de kwaliteit van verschillende schatters te gaan onderzoeken om betrouwbaardere conclusies te kunnen trekken uit dezelfde data.

2 De verdeling van de modulus

Het raamwerk dat men meestal hanteert, veronderstelt dat de fouten op zowel de reële als op de imaginaire component normaal verdeeld zijn rond nul, met dezelfde variantie (σ^2) en onafhankelijk van elkaar. Als we de toevalsvariabele geassocieerd met ons model noteren als $x + iy$, krijgen we

$$x + iy \stackrel{d}{=} (A_r + \epsilon_r) + i(A_i + \epsilon_i).$$

Waarbij $A_r + iA_i$ de waarde is die men zou moeten meten en waarbij ϵ_r en ϵ_i de fouten zijn op respectievelijk het reële en het imaginaire deel. Met $\stackrel{d}{=}$ wordt bedoeld dat ze gelijk zijn in distributie. Hieruit volgt dan ook onmiddellijk dat $x \sim \mathcal{N}(A_r; \sigma^2)$ en $y \sim \mathcal{N}(A_i; \sigma^2)$. Aangezien x en y ook onafhankelijk zijn van elkaar, kunnen we meteen hun gezamenlijke dichtheidsfunctie noteren als

$$\begin{aligned} p(x, y|A_r, A_i, \sigma^2) &= p_r(x|A_r, \sigma^2)p_i(y|A_i, \sigma^2) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(A_r - x)^2}{2\sigma^2}\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(A_i - y)^2}{2\sigma^2}\right). \end{aligned}$$

Als we nu alles overzetten naar poolcordinaten via de volgende substituties,

$$\begin{aligned} A_r &= A\cos(\phi) & x &= M\cos(\psi) \\ A_i &= A\sin(\phi) & y &= M\sin(\psi) \end{aligned}$$

dan kunnen we de spoorverdeling van de modulus verkrijgen door de gezamenlijke verdeling te integreren over ϕ.

$$
\begin{aligned}
p_M(M|A,\sigma^2) &= \int_0^{2\pi} p(M,\psi|A,\phi,\sigma^2)d\psi \\
&= \frac{1}{2\pi\sigma^2}\int_0^{2\pi} \exp\left(\frac{-(A\cos(\phi)-M\cos(\psi))^2-(A\sin(\phi)-M\sin(\psi))^2}{2\sigma^2}\right) M d\psi \\
&= \frac{M}{2\pi\sigma^2}\int_0^{2\pi} \exp\left(-\frac{A^2+M^2-2MA(\cos(\phi)\cos(\psi)+\sin(\phi)\sin(\psi))}{2\sigma^2}\right) d\psi \\
&= \frac{M}{2\pi\sigma^2}\exp\left(-\frac{A^2+M^2}{2\sigma^2}\right)\int_0^{2\pi} \exp\left(\frac{MA\cos(\psi-\phi)}{\sigma^2}\right) d\psi \\
&= \frac{M}{2\pi\sigma^2}\exp\left(-\frac{A^2+M^2}{2\sigma^2}\right)\int_0^{2\pi} \exp\left(\frac{MA\cos(\psi)}{\sigma^2}\right) d\psi
\end{aligned}
$$

Deze laatste gelijkheid geldt aangezien de te integreren functie 2π-cyclisch is en het dus niet uitmaakt of men die doorloopt van $-\psi$ tot $-\psi+2\pi$ of van 0 tot 2π. Als we nu $I_0(x)$ noteren voor de nulde orde van de gemodificeerde Bessel-functie van de eerste soort in x, dan is per definitie

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp(x\cos\psi)d\psi.$$

Dit geeft ons de gelegenheid om de dichtheidsfunctie van de modulus te herschrijven als

$$p_M(M|A,\sigma^2) = \frac{M}{\sigma^2}\exp\left(-\frac{A^2+M^2}{2\sigma^2}\right) I_0\left(\frac{MA}{\sigma^2}\right).$$

Dit is de dichtheidsfunctie van de Rice-verdeling met parameters A en σ^2, dus kunnen we concluderen dat de modulus van de data (de magnitude data) steeds een Rice-verdeling volgt [1], [12]. We zullen deze dichtheidsfunctie in de rest van deze bachelorproef noteren als $f(M,A,\sigma^2)$. De uitdaging is nu om aan de hand van de data de waarde van A, σ en/of (A/σ) te schatten. Deze waarden noemt men respectievelijk de signaalsterkte (of amplitude), de standaardafwijking van de ruis en de signaal-tot-ruisverhouding (SNR, van het Engelse signal-to-noise ratio). Aangezien de Rice-verdelingen geen exponentiële familie vormen en ook geen sufficiënte statistieken hebben, is het vinden van goede schatters niet zo voor de hand liggend.

De meest voorkomende manier waarop men met dit probleem omgaat bij traditionele MRI, is door te negeren dat het hier om een Rice-verdeling gaat en de magnitude data te behandelen alsof ze Gaussiaans verdeeld zijn. Men schat A en σ dan aan de hand van het steekproefgemiddelde en de steekproefvariantie. Voor hoge SNR, i.e. SNR > 3, benadert de Rice-verdeling de Gauss-verdeling [12] (op figuur A van bijlage A bevinden zich grafieken van Rice-verdelingen met verschillende SNR). Zolang er dus voldoende signaal is, is dit geen al te slechte methode. Helaas moet men bij fMRI vaak wel nauwkeurige schattingen hebben bij zwakke signalen en is dit dus geen optie.

Een betere methode, maar helaas met een gelijkaardige tekortkoming, is het gebruik van de momentschatter (method of moments, MOM). De MOM-schatter heeft als voordeel dat het relatief weinig berekeningen vereist en desondanks bij hoge SNR een relatief goed resultaat geeft. Het nadeel is echter dat bij lage SNR de schatter complexe, en dus betekenisloze waarden kan opleveren. De schatter werkt ook niet zeer goed bij lage SNR tenzij er veel datapunten beschikbaar zijn [7]. Het is dus interessant om op zoek te gaan naar methoden die ook bruikbare antwoorden geven voor Rice-verdelingen met lagere SNR.

3 Maximum likelihood-schatters

De ML-schatters zijn populair aangezien ze asymptotisch voor grote steekproeven normaal verdeelde, onvertekende schatters opleveren met minimale variantie, ook voor lage waarden van de SNR [4]. Het nadeel van ML-schatters in het algemeen is echter dat men er niet noodzakelijkerwijs vanuit kan gaan dat ze bestaan en/of dat ze uniek zijn. In het geval van de familie van Rice-verdelingen is dit gelukkig wel zo.

Stelling 3.1. *Noteer $p(x|A, \sigma^2)$ voor de dichtheidsfunctie van de Rice-verdeling en laat $\boldsymbol{m}$ een n-dimensionale datavector zijn. Dan volgt er voor een vaste $\boldsymbol{m}$ dat de likelihood-functie $L_{\boldsymbol{m}}$ gegeven door*

$$L_{\boldsymbol{m}}(A, \sigma^2) = \prod_{i=1}^{n} f(m_i, A, \sigma^2), \tag{3.1}$$

een uniek maximum bereikt.

Bewijs. Als we het logaritme nemen van $L_{\mathbf{m}}$, krijgen we

$$\log\left[L_{\mathbf{m}}(A,\sigma^2)\right] = \log\left[\prod_{i=1}^{n}\frac{m_i}{\sigma^2}\exp\left(-\frac{A^2+m_i^2}{2\sigma^2}\right)I_0\left(\frac{m_iA}{\sigma^2}\right)\right]$$

$$= \sum_{i=1}^{n}\log\left(\frac{m_i}{\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(A^2+m_i^2) + \sum_{i=1}^{n}\log\left[I_0\left(\frac{m_iA}{\sigma^2}\right)\right] \tag{3.2}$$

Aangezien deze functie een samenstelling is van gladde functies (C^∞) op het domein, nl. $A > 0$ en $\sigma > 0$, dus is $\log \circ L_m$ daar ook glad. De gradiënt bestaat dus, maar voor we deze berekenen is het nuttig te vermelden dat de gemodificeerde Bessel-functie van de eerste vorm van de eerste orde (I_1) gedefinieerd is als

$$I_1(z) = \frac{1}{2\pi}\int_0^{2\pi}\exp(z\cos\psi)\cos\psi d\psi.$$

Dit stelt ons in staat om de componenten van de gradiënt te schrijven als

$$\frac{\partial}{\partial A}\log\left[L_m(A,\sigma^2)\right]$$

$$= -\frac{nA}{\sigma^2} + \sum_{i=1}^{n}\frac{1}{I_0(m_iA/\sigma^2)}\left(\int_0^{2\pi}\exp\left(\frac{m_iA}{\sigma^2}\cos\psi\right)\frac{m_i}{\sigma^2}\cos\psi d\psi\right)$$

$$= -\frac{n}{\sigma^2}\left[A - \frac{1}{n}\sum_{i=1}^{n}m_i\frac{I_1(m_iA/\sigma^2)}{I_0(m_iA/\sigma^2)}\right] \tag{3.3}$$

en

$$\frac{\partial}{\partial \sigma^2}\log\left[L_m(A,\sigma^2)\right]$$

$$= \sum_{i=1}^{n}\frac{\sigma^2}{m_i}\frac{-2m_i}{\sigma^4} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(A^2+m_i^2) - \frac{2A}{\sigma^4}\sum_{i=1}^{n}m_i\frac{I_1(m_iA/\sigma^2)}{I_0(m_iA/\sigma^2)}$$

$$= -\frac{2n}{\sigma^2}\left[1 - \frac{A^2}{2\sigma^2} - \frac{1}{N}\sum_{i=1}^{n}\frac{m_i^2}{2\sigma^2} + \frac{A}{n\sigma^2}\sum_{i=1}^{n}m_i\frac{I_1(m_iA/\sigma^2)}{I_0(m_iA/\sigma^2)}\right] \tag{3.4}$$

Indien er lokale maxima bestaan binnen het inwendige van het domein van de likelihood-functie, moet de gradiënt daar nul zijn. Door de partiële afgeleide

naar A gelijk te stellen aan nul, verkrijgen we

$$A = \frac{1}{n}\sum_{i=1}^{n} m_i \frac{I_1(m_i A/\sigma^2)}{I_0(m_i A/\sigma^2)}. \tag{3.5}$$

Als we de partiële afgeleide naar σ^2 ook gelijk stellen aan nul en (3.5) daarin invullen, krijgen we

$$\begin{aligned} 1 + \frac{A^2}{2\sigma^2} &= \frac{A^2}{2\sigma^2} + \frac{1}{n}\sum_{i=1}^{n} \frac{m_i^2}{2\sigma^2} \\ \Leftrightarrow \quad A^2 + 2\sigma^2 &= \frac{1}{n}\sum_{i=1}^{n} m_i^2 \end{aligned} \tag{3.6}$$

De likelihood-functie bereikt dus alleen een lokaal maximum als (3.5) en (3.6) voldaan zijn. De waarden van A en σ^2 die hieraan voldoen, zullen we noteren met $\overline{A} = \overline{A}(\mathbf{m})$ en $\overline{\sigma^2} = \overline{\sigma^2}(\mathbf{m})$. Voor de waarde $\sqrt{\overline{\sigma^2}}$ zullen we $\overline{\sigma}$ gebruiken. We kunnen het probleem nu opsplitsen in twee gevallen. In het eerste geval veronderstellen we dat de waarde van σ^2 gekend is en in het tweede geval doen we dat niet.

A) σ^2 is gekend

Aangezien we de likelihood-functie hier enkel als een functie van A beschouwen, moet alleen (3.3) gemaximaliseerd worden en dus is enkel (3.5) relevant. We zullen het linkerlid (LL) en het rechterlid (RL) van (3.5) beschouwen als functies van A (een visuele voorstelling hiervan kan teruggevonden worden in bijlage A). Aangezien $I_1(z)/I_0(z)$ een monotoon stijgende functie is en haar afgeleide monotoon dalend met als limiet 0 [10], volgt er meteen dat dit ook geldt voor $RL(A)$.
Als we 0 invullen in I_1 en I_0, krijgen we

$$\frac{1}{2\pi}\int_0^{2\pi} \cos\psi d\psi = 0 \quad \text{en} \quad \frac{1}{2\pi}\int_0^{2\pi} 1 d\psi = 1$$

Hieruit volgt dat $RL(0) = 0$ en als we de afgeleide van $RL(A)$ in 0 berekenen, krijgen we

$$
\begin{aligned}
RL'(0) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{m_i^2}{\sigma^2}\right)\frac{I_0(0)\int_0^{2\pi}\exp(0\cos\psi)\cos^2\psi d\psi - I_1^2(0)}{I_0^2(0)} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\frac{m_i^2}{\sigma^2}.
\end{aligned}
$$

Als we $\overline{\mathbf{m}^2}$ definiëren als $\sum_{i=1}^{n} m_i^2/n$, betekent dit dat $RL'(0) > 1$, als $\sigma^2 < \overline{\mathbf{m}^2}/2$ en $RL'(0) \leq 1$ als $\sigma^2 \geq \overline{\mathbf{m}^2}/2$. Uit het feit dat $LL(A) = A$, volgt meteen dat ook het linkerlid monotoon stijgend is en dat $LL(0) = 0$. De afgeleide van LL is constant 1, dus volgt er dat als $RL'(0) \leq 1$, de vergelijking $LL(A) = RL(A)$ een unieke oplossing heeft in het punt $A = 0$ en dat $LL(A) > RL(A)$ elders. Dit betekent dat (3.3) negatief is en dat de likelihood-functie een uniek maximum bereikt in $A = 0$ voor $A \geq 0$.
Als $RL'(0) > 1$, is $A = 0$ nog steeds een oplossing, maar omdat $RL'(A)$ monotoon daalt naar nul, moet er een tweede punt $\overline{A}$ bestaan waar $RL(A)$ en de eerste bissectrice snijden. In dit geval is $LL(A) < RL(A)$ voor $0 < A < \overline{A}$ en (3.3) dus positief, wat betekent de oplossing $A = 0$ een lokaal minimum is van de likelihood-functie. Voor $\overline{A} < A$ is $LL(A) > RL(A)$, waardoor de likelihood-functie ook hier een uniek maximum heeft, namelijk in $\overline{A}$.

B) σ^2 is onbekend

Als we A en σ overzetten in poolcordinaten t.o.v. het A-σ-vlak, d.w.z. als we de volgende substituties uitvoeren

$$
A = \rho\sin\theta, \sigma = \rho\cos\theta,
$$

dan wordt de dichtheidsfunctie

$$
\begin{aligned}
f(m_i, A, \sigma^2) &= \frac{m_i}{\rho^2\cos^2\theta}\exp\left(-\frac{m_i^2 + \rho^2\sin^2\theta}{2\rho^2\cos^2\theta}\right) I_0\left(\frac{m_i\sin\theta}{\rho\cos^2\theta}\right) \\
&=: g(m_i, A, \sigma^2).
\end{aligned}
$$

Aangezien $\sigma > 0$ op het domein, geldt $\cos\theta > 0$ en bijgevolg ook $0 \leq \theta < \pi$. Als we ρ nu naar oneindig laten gaan, dan convergeert $I_0(m_i\sin\theta/\rho\cos^2\theta)$ naar 1 en $\exp(-(m_i^2 + \rho^2\sin^2\theta)/2\rho^2\cos^2\theta)$ naar $\exp(\tan^2\theta/2) < \infty$, zodat $\lim_{\rho\to\infty} g(m_i, \rho\theta) = 0$. Hieruit kunnen we afleiden dat voor elke waarde $0 \leq t < \pi$ de functie $g(m_i, \rho, \theta)$ uniform convergeert naar 0 voor $\rho \to \infty$ op het domein $m_i \times \mathbb{R}^+ \times [0, t]$.

Om te controleren dat dit ook geldt op het volledige A-σ-vlak, moeten we het gedrag van de functie nog bestuderen in de buurt van $\theta = \pi/2$. Stel $\rho > 0$, $\theta \to \pi/2$ en benader de goniometrische functies met hun eerste orde Taylor-ontwikkeling rond $\theta = \pi/2$. Dan krijgen we

$$g(m_i, \rho, \theta) \to \frac{m_i}{\rho^2(\pi/2-\theta)^2} \exp\left(-\frac{m_i^2+\rho^2}{2\rho^2(\pi/2-\theta)^2}\right) I_0\left(\frac{m_i}{\rho(\pi/2-\theta)^2}\right) \tag{3.7}$$

We zullen nu gebruik maken van het volgende lemma, afkomstig uit [1].

Lemma 3.2. *Voor elke $\delta_0 > 0$ bestaat er een $K \in \mathbb{R}_0$ zodat voor elke $x > \delta_0$ geldt dat*

$$I_0(x) \leq K \frac{1}{\sqrt{x}} e^x \tag{3.8}$$

Als we $\delta_0 = 1$ kiezen, mogen we (3.7) herschrijven zodra $(\pi/2) - \theta \leq m_i/\rho$

$$\begin{aligned} g(m_i, \rho, \theta) &\leq \frac{m_i}{\rho^2(\pi/2-\theta)^2} \exp\left(-\frac{m_i^2+\rho^2}{2\rho^2(\pi/2-\theta)^2}\right) \frac{K \exp\left(\frac{m_i}{\rho(\pi/2-\theta)^2}\right)}{\sqrt{\frac{m_i}{\rho(\pi/2-\theta)^2}}} \\ &\leq \frac{K}{\rho(\pi/2-\theta)} \sqrt{\frac{m_i}{\rho}} \exp\left(-\frac{(m_i-\rho)^2}{2\rho^2(\pi/2-\theta)^2}\right). \end{aligned}$$

Dus geldt voor elke $0 < \rho \neq m_i$ dat $\lim_{\theta \to \pi/2} g(m_i, \rho, \theta) = 0$. Dit betekent dat $g(m_i, \rho, \theta)$ uniform convergeert naar 0 voor $\rho \to \infty$. Aangezien dit geldt voor elke waarde m_i geldt dit ook voor het product $\prod g(m_i, \rho, \theta)$ en dit is niets anders dan de likelihood-functie.

Als we het geval gaan bekijken waar $\sigma = 0$ (of equivalent $\theta = \pi/2$), dan is de kansfunctie gelijk aan de indicator-functie $I_{\{m_i\}}$. Dit is geen continue verdeling, dus kunnen we ook niet op een zinnige manier over de dichtheidsfunctie spreken. Wat we wel kunnen zeggen is dat als er indices i en j bestaan zodat $m_i \neq m_j$, dan is de kans dat $\sigma = 0$ gelijk is aan nul omdat dit zou vereisen dat we steeds dezelfde data krijgen. We mogen dus de likelihood-functie op de A-as dan ook als constant nul beschouwen. Zitten we in het triviale geval waar alle m_i's gelijk zijn, dan is $(A, \sigma) = (m_1, 0)$ duidelijk de beste schatter.

Als we enkel naar het geval kijken waarin minstens één van de m_i's verschilt van de anderen, volgt er dat de likelihood-functie haar maximum bereikt op een begrensd deel van het domein $(R^+)^2$. Voor functies die continu zijn op hun domein, kunnen lokale maxima enkel bereikt worden in punten waar de

gradiënt 0 is of in de grenspunten van het domein. De A-as hebben we hierboven al besproken, dus kijken we nu naar het geval $A = \theta = 0$ en $\sigma = \rho$. We kunnen de likelihood-functie dan schrijven als

$$\begin{aligned} L(\mathbf{m}, 0, \sigma^2) &= \prod_{i=1}^{n} g(m_i, \rho, 0) \\ &= \prod_{i=1}^{n} \frac{m_i}{\rho^2} \exp\left(-\frac{m_i^2}{2\rho^2}\right) \\ &= \left(\prod_{i=1}^{n} \frac{m_i}{\rho^2}\right) \exp\left(-\sum_{i=1}^{n} \frac{m_i^2}{2\rho^2}\right) \end{aligned}$$

Als $\rho \to 0$ en $\rho \to \infty$, dan hebben we $\prod_{i=1}^{n} g(m_i, \rho, 0) \to 0$. Als we $\sum_{i=1}^{n} m_i^2/n$ noteren als $\overline{\mathbf{m}^2}$ en g partieel af te leiden naar ρ, krijgen we

$$\frac{\partial}{\partial \rho}\left(\prod_{i=1}^{n} g(m_i, \rho, 0)\right) = \prod_{i=1}^{n} \frac{m_i}{\rho^2}\left[\frac{-2n}{\rho^{2n+1}} \exp\left(\frac{-n\overline{\mathbf{m}^2}}{2\rho^2}\right) + \frac{n\overline{\mathbf{m}^2}}{\rho^{2n+3}} \exp\left(\frac{-n\overline{\mathbf{m}^2}}{2\rho^2}\right)\right]$$

Dit is gelijk aan 0 als

$$\begin{aligned} 2n &= \frac{\overline{\mathbf{m}^2} n}{\rho^2} \\ \Leftrightarrow \rho &= \sqrt{\frac{\overline{\mathbf{m}^2}}{2}}. \end{aligned}$$

Dus wordt het maximum van de likelihood-functie op de σ-as bereikt in het punt $(0, \overline{\mathbf{m}^2}/2)$. Om te onderzoeken waar er zich lokale maxima kunnen bevinden in inwendige punten van het domein, definiëren we L_1 en L_2 als de krommen in het A-σ-vlak waarvan de punten respectievelijk voldoen aan (3.5) en (3.6).
De punten van L_1 zijn dan de maxima uit paragraaf A voor de verschillende waarden van σ. Aangezien $\lim_{z\to\infty} I_1(z)/I_0(z) = 1$ (zie [10]), is het snijpunt van L_1 met de σ-as (als $\lim_{\sigma\to 0}$ van (3.5)) gelijk aan $(0, \overline{\mathbf{m}})$, met $\overline{\mathbf{m}} = 1/n \sum_{i=1}^{n} m_i$. Als we (3.5) combineren met het feit dat $(I_1(1/x))/(I_0(1/x))$ monotoon dalend is, volgt er dat L_1 ook monotoon daalt tot $(\sqrt{\overline{\mathbf{m}^2}/2}, 0)$. Uit paragraaf A volgt namelijk dat L_1 vanaf dat punt constant 0 is.
Als we nu naar (3.6) kijken, zien we dat L_2 een ellips is met als snijpunt met de σ-as $(0, \sqrt{\overline{m^2}})$ en met de A-as $(\sqrt{\overline{\mathbf{m}^2}/2}, 0)$. Het snijpunt met $A = 0$ is dus hetzelfde als voor L_1 en dit is ook het punt waarvan we eerder al hadden

vastgesteld dat het het maximum was op de σ-as. Als het vlak geen andere lokale maxima heeft, d.w.z. als L_1 en L_2 voor de rest geen gemeenschappelijke punten hebben, dan wordt in dit punt het absolute maximum van de likelihood-functie bereikt.
We gaan nu per inductie aantonen dat

$$\left(\sum_{i=1}^{n} m_i\right)^2 \leq n \sum_{i=1}^{n} m_i^2 \tag{3.9}$$

Voor het geval $n = 1$ is dit $m_1 = m_1$ en dus triviaal. Veronderstel nu dat deze ongelijkheid geldt voor sommen met $n - 1$ waarden, dan volgt er

$$\begin{aligned}
\left(\sum_{i=1}^{n} m_i\right)^2 &= \sum_{i=1}^{n} m_i \sum_{j=1}^{n} m_j \\
&= m_n^2 + m_n \sum_{j=1}^{n-1} m_j + \sum_{i=1}^{n-1} m_i m_n + \sum_{i=1}^{n-1} m_i \sum_{j=1}^{n-1} m_j \\
&\leq m_n^2 + \sum_{i=1}^{n-1} 2 m_n m_i + (n-1) \sum_{i=1}^{n-1} m_i^2
\end{aligned}$$

Neem nu twee waarden m_k en m_l met $0 \leq k, l \leq n$, dan mogen we veronderstellen dat één van de twee waarden groter is of gelijk aan de andere, bijvoorbeeld $m_k \geq m_l$. Dan kunnen we daaruit concluderen dat

$$\begin{aligned}
m_k(m_k - m_l) &\geq m_l(m_k - m_l) \\
\Leftrightarrow m_k(m_k - m_l) + m_l(m_l - m_k) &\geq 0 \\
\Leftrightarrow m_k^2 + m_l^2 &\geq 2 m_l m_k.
\end{aligned}$$

en dus

$$\begin{aligned}
\left(\sum_{i=1}^{n} m_i\right)^2 &\leq m_n^2 + (n-1) m_n^2 + \sum_{i=1}^{n-1} m_i^2 + (n-1) \sum_{i=1}^{n-1} m_i^2 \\
&\leq n \sum_{i=1}^{n} m_i^2.
\end{aligned}$$

Waarbij de gelijkheid enkel geldt als alle m_i gelijk zijn, d.w.z. $\sigma = 0$. De bovenstaande vergelijking is equivalent met

$$\begin{aligned}
\frac{1}{n^2} \left(\sum_{i=1}^{n} m_i\right)^2 &\leq \frac{1}{n} \sum_{i=1}^{n} m_i^2 \\
\Leftrightarrow \overline{\mathbf{m}} &\leq \sqrt{\overline{\mathbf{m}^2}}.
\end{aligned}$$

Hieruit volgt dat L_2 de A-as later snijdt dan L_1, behalve in het triviale geval waar alle m_i gelijk zijn. Op de A-as bevinden zich dus geen snijpunten van L_1 en L_2.
Veronderstel dat er een tweede snijpunt $(\overline{A}, \overline{\sigma})$ bestaat, dan kunnen we daarvoor de Hessiaan van $\log \circ L$ berekenen aan de hand van de volgende recurrentieformule [9],

$$I'_\nu(z) = I_{\nu-1} - \frac{\nu}{x} I_\nu.$$

Als we dan nu de partiële afgeleides berekenen van (3.3) en (3.4) in $(\overline{A}, \overline{\sigma})$, krijgen we

$$\begin{aligned}
H_{11} &= \frac{\partial^2(\log \circ L)}{\partial A^2}(\overline{A}, \overline{\sigma}) \\
&= \frac{-n}{\overline{\sigma^2}} \left[1 - \frac{1}{n} \sum_{i=1}^{n} \frac{m_i^2}{\overline{\sigma^2}} \left(1 - \frac{\overline{\sigma^2}}{\overline{A} m_i} \frac{I_1(m_i \overline{A}/\overline{\sigma^2})}{I_0(m_i \overline{A}/\overline{\sigma^2})} - \frac{I_1^2(m_i \overline{A}/\overline{\sigma^2})}{I_0^2(m_i \overline{A}/\overline{\sigma^2})} \right) \right] \\
&= \frac{-n}{\overline{\sigma^2}} \left[2 - \frac{1}{n} \sum_{i=1}^{n} \frac{m_i^2}{\overline{\sigma^2}} \left(1 - \frac{I_1^2(m_i \overline{A}/\overline{\sigma^2})}{I_0^2(m_i \overline{A}/\overline{\sigma^2})} \right) \right]
\end{aligned}$$

$$\begin{aligned}
H_{12} &= \frac{\partial^2(\log \circ L)}{\partial A \partial \sigma^2}(\overline{A}, \overline{\sigma}) \\
&= \frac{n}{(\overline{\sigma^2})^2} \left[\overline{A} - \overline{A} - \frac{1}{n} \sum_{i=1}^{n} \frac{-m_i^2 \overline{A}}{\overline{\sigma^2}} \left(1 - \frac{\overline{\sigma^2}}{\overline{A} m_i} \frac{I_1(m_i \overline{A}/\overline{\sigma^2})}{I_0(m_i \overline{A}/\overline{\sigma^2})} - \frac{I_1^2(m_i \overline{A}/\overline{\sigma^2})}{I_0^2(m_i \overline{A}/\overline{\sigma^2})} \right) \right] \\
&= \frac{n\overline{A}}{(\overline{\sigma^2})^2} \left[1 - \frac{1}{n} \sum_{i=1}^{n} \frac{m_i}{\overline{\sigma^2}} \left[1 - \frac{I_1^2(m_i \overline{A}/\overline{\sigma^2})}{I_0^2(m_i \overline{A}/\overline{\sigma^2})} \right) \right] \\
&= H_{21}
\end{aligned}$$

$$\begin{aligned}
H_{22} &= \frac{\partial^2(\log \circ L)}{\partial (\sigma^2)^2}(\overline{A}, \overline{\sigma^2}) \\
&= \frac{-n}{(\overline{\sigma^2})^2} \left[1 + \frac{\overline{A}^2}{\overline{\sigma^2}} \left(1 - \frac{1}{n} \sum_{i=1}^{n} \frac{m_i}{\overline{\sigma^2}} \left[1 - \frac{I_1^2(m_i \overline{A}/\overline{\sigma^2})}{I_0^2(m_i \overline{A}/\overline{\sigma^2})} \right] \right) \right]
\end{aligned}$$

We hebben hierbij gebruikt dat voor $(\overline{A}, \overline{\sigma})$ vergelijkingen (3.5) en (3.6) gelden. Als we dat gebruiken om de determinant van de Hessiaan uit te rekenen,

verkrijgen we uiteindelijk

$$\begin{aligned}\det[\mathbf{H}] &= H_{11}H_{22} - H_{12}^2 \\ &= \frac{-n}{(\overline{\sigma^2})^3}\left[2 - \frac{1}{n}\sum_{i=1}^{n}\frac{m_i^2}{\overline{\sigma^2}}\left(1 - \frac{I_1^2(m_i\overline{A}/\overline{\sigma^2})}{I_0^2(m_i\overline{A}/\overline{\sigma^2})} + \frac{\overline{A}^2}{\overline{\sigma^2}}\right)\right]\end{aligned}$$

Als we α definiëren als

$$\alpha = 2 - \frac{1}{n}\sum_{i=1}^{n}\frac{m_i^2}{\overline{\sigma}^2}\left(1 - \frac{I_1^2(m_i\overline{A}/\overline{\sigma})}{I_0^2(m_i\overline{A}/\overline{\sigma})}\right), \tag{3.10}$$

dan worden H_{11} en de determinant

$$H_{11} = -\frac{n\alpha}{(\overline{\sigma^2})^2}$$

$$\det[\mathbf{H}] = \frac{-n}{(\overline{\sigma^2})^3}\left[\alpha + \frac{\overline{A}^2}{\overline{\sigma^2}}\right]$$

Als we nu (3.6) toepassen in (3.10), krijgen we

$$\alpha = \frac{1}{\overline{\sigma^2}}\left[(\overline{A})^2 - \frac{1}{n}\sum_{i=1}^{n}\left(m_i\frac{I_1(m_i\overline{A}/\overline{\sigma^2})}{I_0(m_i\overline{A}/\overline{\sigma^2})}\right)^2\right]. \tag{3.11}$$

Wegens (3.9), volgt er ook dat

$$\frac{1}{n}\sum_{i=1}^{n}\left(m_i\frac{I_1(m_i\overline{A}/\overline{\sigma^2})}{I_0(m_i\overline{A}/\overline{\sigma^2})}\right)^2 > \left(\frac{1}{n}\sum_{i=1}^{n}m_i\frac{I_1(m_i\overline{A}/\overline{\sigma^2})}{I_0(m_i\overline{A}/\overline{\sigma^2})}\right)^2 = (\overline{A})^2$$

Uit deze ongelijkheid kunnen we dus concluderen dat $\alpha > 0$ en dus $H_{11} < 0$ en $\det[\mathbf{H}] > 0$. De Hessiaan is dus negatief definiet en elk snijpunt van L_1 en L_2 met $A > 0$, is een lokaal maximum.

Als er dus zo een $(\overline{A}, \overline{\sigma})$ bestaat, dan zijn zowel dit punt als $(0, \sqrt{\overline{\mathbf{m}^2}/2})$ mogelijke kandidaten voor het maximum. Veronderstel dat $(\overline{A}, \overline{\sigma})$ van alle gemeenschappelijke punten van L_1 en L_2 degene is met de hoogste waarde van A en dat $\overline{A} \neq 0$. Als we nu over de ellips L_2 bewegen van $(0, \sqrt{\overline{\mathbf{m}^2}/2})$ naar $(\overline{A}, \overline{\sigma})$, waarbij we σ^2 substitueren door $(\overline{\mathbf{m}^2} - A^2)/2$, dan wordt (3.2)

$$\begin{aligned}\log\left[L_{\mathbf{m}}(A, \sigma^2(A))\right] = \sum_{i=1}^{n}\log\left(\frac{2m_i}{\overline{\mathbf{m}^2} - A^2}\right) - \sum_{i=1}^{n}\frac{A^2 + m_i^2}{\overline{\mathbf{m}^2} - A^2} \\ + \sum_{i=1}^{n}\log\left[I_0\left(\frac{2m_iA}{\overline{\mathbf{m}^2} - A^2}\right)\right]\end{aligned}$$

De afgeleide naar A van deze functie wordt na enkele berekeningen

$$\frac{d}{dA}\left(\log\left[L_{\mathbf{m}}(A,\sigma^2(A))\right]\right)$$
$$= \frac{nA^2}{\overline{m^2}-A^2}\left[(\overline{m^2+A^2})\left(\frac{1}{n}\sum_{i=1}^{n} m_i \frac{I_1((2Am_i)/(\overline{m^2-A^2})}{I_0((2Am_i)/(\overline{m^2-A^2})} - A\right)\right]. \tag{3.12}$$

Als we nu bewegen van het punt $(0,\sqrt{\overline{\mathbf{m}^2}/2})$ naar $(\overline{A},\overline{\sigma})$, dan is $0 \leq A \leq \overline{A}$ en $A^2 < \overline{m^2}$ (omdat we hadden aangetoond dat het snijpunt van de ellips met de A-as $(\overline{m^2},0)$ is). Dit betekent dat het teken van de afgeleide van (3.12) gelijk is aan

$$\frac{1}{n}\sum_{i=1}^{n} m_n \frac{I_1\left(\frac{2Am_i}{\overline{m^2}-A^2}\right)}{I_0\left(\frac{2Am_i}{\overline{m^2}-A^2}\right)} - A \tag{3.13}$$

Aangezien we aangetoond hebben dat $(\overline{A},\overline{\sigma})$ een lokaal maximum is, moet de afgeleide positief zijn op een omgeving $]\overline{A}-\epsilon_1,\overline{A}[$, waarbij $0<\epsilon_1$. We kunnen nu met elke waarde van A twee waarden van σ associëren, namelijk die op L_2 en die op L_1. Van de eerste weten we dat deze waarde gelijk is aan $\sqrt{(\overline{m^2}-A^2)/2}$, de tweede zullen we noteren met $\tilde{\sigma}$. We weten dat L_2 boven L_1 ligt voor $A>\overline{A}$, stel nu dat L_1 en L_2 elkaar niet snijden, maar slechts raken in $(\overline{A},\overline{\sigma})$. Dan betekent dit dat voor $\overline{A}-\epsilon_2 < A < \overline{A}$

$$(\tilde{\sigma})^2 \leq \frac{\overline{m^2}-A^2}{2},$$

voor een $\epsilon_2>0$. Wegens het monotoon zijn van $I_1(z)/I_0(z)$, krijgen we

$$\frac{1}{n}\sum_{i=1}^{n} m_n \frac{I_1\left(\frac{2Am_i}{\overline{m^2}-A^2}\right)}{I_0\left(\frac{2Am_i}{\overline{m^2}-A^2}\right)} - A \leq \frac{1}{n}\sum_{i=1}^{n} m_n \frac{I_1\left(\frac{Am_i}{(\tilde{\sigma})^2}\right)}{I_0\left(\frac{Am_i}{(\tilde{\sigma})^2}\right)} - A = A$$

Deze laatste gelijkheid geldt omdat het punt $(A,\tilde{\sigma})$ tot L_2 behoort. Deze ongelijkheid betekent echter dat voor $\overline{A}-\min\{\epsilon_1,\epsilon_2\} < A < \overline{A}$ vergelijking (3.13) negatief wordt. Dit is echter in contradictie met het feit dat (3.12) hier positief moet zijn, dus snijden L_1 en L_2 elkaar wel degelijk en

$$(\tilde{\sigma})^2 > \frac{\overline{m^2}-A^2}{2},$$

tussen $\overline{A}$ en het volgende gemeenschappelijke punt van L_1 en L_2. Dit betekent dus ook dat de afgeleide (3.12) op dit volledige stuk positief is. Dit volgende gemeenschappelijke punt kan dus onmogelijk een (lokaal) maximum zijn. Behalve $(0, \sqrt{\overline{\mathbf{m}^2}/2})$ en $(\overline{A}, \overline{\sigma})$ hebben L_1 en L_2 geen mogelijke gemeenschappelijke snijpunten. Als er een $(\overline{A}, \overline{\sigma})$ bestaat is dit het unieke maximum, zoniet is $(0, \sqrt{\overline{\mathbf{m}^2}/2})$ het maximum. □

Een nadeel bij het berekenen van ML-schatters is dat men hierbij vaak impliciete, niet-lineaire vergelijkingen moet oplossen, wat behoorlijk wat rekenwerk met zich mee kan brengen. Verder hebben we bij fMRI vaak niet de luxe om veel metingen uit te voeren per voxel per instantie. Dit is ten eerste vanwege de beperkte tijd waarin de metingen voor één foto gemaakt kunnen worden en ten tweede vanwege de grote hoeveelheid voxels waarvan de data geregistreerd moet worden binnen deze tijd.
Verder geven de algoritmes waarmee we het maximum zoeken niet noodzakelijkerwijs de juiste oplossing, in de zin dat ze afhankelijk van de startwaarde soms ook naar de andere oplossing van vergelijkingen (3.3) en (3.4) convergeren, namelijk $(0, \sqrt{\overline{\mathbf{m}^2}/2})$. Dit betekent dat het nodig kan zijn om het algoritme verschillende keren te herstarten met andere beginwaarden om alsnog het correcte maximum te verkrijgen [3], [1].
Ondanks de mooie asymptotische eigenschappen van ML-schatters is het dus toch interessant om te kijken naar technieken die het beter doen voor minder grote steekproeven en met algoritmes die gegarandeerd naar de juiste waarde convergeren.

4 Bayesiaanse Schatters

Bij Bayesiaanse statistiek is het centrale idee dat men vantevoren gekende informatie ook verwerkt in het wiskundige raamwerk. Er wordt hier dus uitgegaan van een zogenaamde a-priori kansverdeling ($\Pi(\theta)$) van de onbekende parameters θ met een a-priori dichtheidsfunctie $\pi(\theta)$, waarbij in ons geval $\theta = (A, \sigma)$. Deze dichtheid kunnen we dan benutten om een a-postiori verdeling $f(\theta|\mathbf{x})$ te construeren aan de hand van de stelling van Bayes voor dichtheden, waarbij ook de metingen $\mathbf{x}$ verwerkt worden:

$$f(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{f(\mathbf{x})}. \tag{4.1}$$

Hierbij is $f(x|\theta)$ de kans van de metingen $\mathbf{x}$ voor de verdeling met parameter θ, dus niets anders dan de likelihood functie (3.1). De functie $f(\mathbf{x})$

is hier de marginale dichtheid van $\mathbf{x}$, aangezien we er van uitgaan dat de metingen gegeven zijn en kan dit dus vervangen worden door een normaliserende constante $1/c$ die er voor zorgt dat de integraal van de a-postiori dichtheidsfunctie gelijk wordt aan 1. Deze kunnen we dus uitrekenen zodra $\pi(\theta)$ bekend is.
Voor het bepalen van welke kennis er in de a-priori verdeling verwerkt wordt, bestaan er twee stromingen binnen de Bayesiaanse statistiek, die van de objectivisten en die van de subjectivisten. De objectivisten eisen dat deze kennis theoretisch gefundeerd moet zijn of formeel afgeleid van algemeen geaccepteerde eigenschappen van het statistische probleem [11]. Subjectivisten zijn hier wat soepeler in en accepteren ook persoonlijke overtuigingen als mogelijke bronnen van informatie. Wij zullen hier enkel informatie benutten conform het standpunt van de objectivisten.
De eigenschappen van het probleem waaruit we deze informatie zullen afleiden zijn:

- $A > 0$ en $\sigma > 0$,
- A en σ zijn onafhankelijk.

Voor de a-priori dichtheidsfunctie betekent dit

$$\pi(A, \sigma) = 0 \quad \text{voor } A \leq 0 \text{ of } \sigma \leq 0 \tag{4.2}$$

$$\pi(A, \sigma) = g(A)h(\sigma). \tag{4.3}$$

De functies g en h zijn hierbij de marginale dichtheidsfuncties van A en σ. Veronderstel nu dat X $Rice(A, \sigma)$-verdeeld is, dan kunnen we voor kX afleiden dat

$$\begin{aligned}
\mathbb{P}(kX \geq x) &= \mathbb{P}(X \geq \frac{x}{k}) \\
&= \int_0^{\frac{x}{k}} \frac{t}{\sigma^2} \exp\left(-\frac{A^2 + t^2}{2\sigma^2}\right) I_0\left(\frac{At}{\sigma^2}\right) dt
\end{aligned}$$

Substitueer $kt = u$

$$= \int_0^x \frac{u}{k^2\sigma^2} \exp\left(-\frac{k^2A^2 + u^2}{2k^2\sigma^2}\right) I_0\left(\frac{kAu}{k^2\sigma^2}\right) du$$

$\Rightarrow kX \sim Rice(kA, k\sigma)$.

Als we nu Jaynes' invariantieprincipe gebruiken, is het mogelijk om aan de hand van al deze gegevens $\pi(A, \sigma)$ op te schrijven. Jaynes' invariantieprincipe stelt dat als een transformatie het onderliggende probleem niet verandert (zoals bijvoorbeeld bij een herparametrisatie), dan verandert onze voorafgaande kennis van het probleem ook niet [3]. Als onze a-priori kennis niet verandert, dan blijft ook de a-priori verdeling hetzelfde, d.w.z. $\Pi(A, \sigma) = \Pi(kA, k\sigma)$ met $k > 0$. Aangezien de dichtheidsfunctie $\pi(A, \sigma) = \frac{\partial^2 \Pi(A,\sigma)}{\partial A \partial \sigma}$, is het gemakkelijk in te zien dat $\pi(A, \sigma) = k^2 \pi(kA, k\sigma)$ met $k > 0$. Als we nu (4.3) gebruiken, volgt er

$$\frac{1}{k^2} g(A) h(\sigma) = g(kA) h(k\sigma).$$

Of equivalent

$$\frac{1}{k^{1-\epsilon}} g(A) \frac{1}{k^{1+\epsilon}} h(\sigma) = g(kA) h(k\sigma) \text{ voor elke } \epsilon \in \mathbb{R}.$$

Bijgevolg zijn alle mogelijke oplossingen van deze vergelijking van de vorm

$$\begin{cases} \frac{1}{k^{1-\epsilon}} g(A) = g(kA) \\ \frac{1}{k^{1+\epsilon}} h(\sigma) = h(k\sigma) \end{cases}$$

Dit betekent natuurlijk dat de enige mogelijke keuze voor g en h gelijk is aan

$$\begin{cases} g(A) = \frac{1}{A^{1-\epsilon}}) \\ h(\sigma) = \frac{1}{\sigma^{1+\epsilon}} \end{cases}$$

Als we hierbij ook (4.2) nog benutten, krijgen we de volgende familie van a-priori verdelingen

$$\pi_\epsilon(A, \sigma) = \begin{cases} \frac{1}{A\sigma} \left(\frac{A}{\sigma}\right)^\epsilon & \text{als } A, \sigma > 0 \\ 0 & \text{anders} \end{cases}$$

Merk op dat deze dichtheidsfunctie geen dichtheidsfunctie is van een klassieke kansverdeling aangezien ze niet integreerbaar is. Dit is geen probleem zolang de a-postiori verdeling uit (4.1) dat wel is. Deze wordt dus

$$\begin{aligned} f(\theta|x) &= c L_x(\theta) \pi_\epsilon(\theta) \\ &= c \frac{\prod_{i=1}^n x_i}{A^{1-\epsilon} \sigma^{2n+1+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + A^2}{2\sigma^2}\right) \prod_{i=1}^n I_0\left(\frac{x_i A}{\sigma^2}\right) \end{aligned} \tag{4.4}$$

Tot nu toe mocht ϵ een willekeurig getal zijn, maar nu moet deze gekozen worden zodat er een constante c bestaat waarvoor

$$\iint_0^\infty f(\theta|x)dAd\sigma = 1.$$

Laten we $f_0(\theta|x)$ definiëren als

$$f_0(\theta|x) := \frac{\prod_{i=1}^n x_i}{A^{1-\epsilon}\sigma^{2n+1+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + A^2}{2\sigma^2}\right) \prod_{i=1}^n I_0\left(\frac{x_i A}{\sigma^2}\right).$$

Dan is het duidelijk dat het bestaan van een c samenvalt met het integreerbaar zijn van f_0.

Stelling 4.1. *De functie $f_0(\theta|x)$ is integreerbaar als $\epsilon \geq 1$.*

Bewijs. Om dit te bewijzen, zullen we gebruik maken van lemma 3.2. Kies daarvoor een $\delta_0 > 0$ willekeurig maar vast en beschouw de volgende partitie met $\delta = \delta_0 / \min_{1\leq i\leq n}\{x_i\}$:

$$V_1 = \left\{(A,\sigma)|\frac{A}{\sigma^2} \geq \delta\right\}$$
$$V_2 = \left\{(A,\sigma)|\frac{A}{\sigma^2} < \delta\right\}$$

Als we dan nu lemma 3.2 toepassen, verkrijgen we

$$\begin{aligned} f_0(\theta|x) &\leq \frac{\prod_{i=1}^n x_i}{A^{1-\epsilon}\sigma^{2n+1+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n x_i^2 + A^2}{2\sigma^2}\right) \prod_{i=1}^n K\sqrt{\frac{\sigma^2}{x_i A}} \exp\left(\frac{x_i A}{\sigma^2}\right) \\ &= K^n \frac{\sqrt{\prod_{i=1}^n x_i}}{A^{(n/2)+1-\epsilon}\sigma^{n+1+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2}\right). \end{aligned}$$

Als we deze functie bekijken voor de eerste verzameling V_1, dan is $A/\sigma^2 \geq \delta$, wat hetzelfde is als $A \geq \delta\sigma^2$. Door $\bar{K} := K^n\sqrt{\prod_{i=1}^n x_i}$ te noteren, wordt dit

$$\begin{aligned} f_0(\theta|x) &\leq \bar{K}\frac{1}{\delta^{(n/2)+1-\epsilon}\sigma^{2n+3+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2}\right) \\ &=: \frac{\bar{K}}{\delta^{(n/2)+1-\epsilon}} g_1(\theta|x). \end{aligned}$$

We kunnen nu de integreerbaarheid van $g_1(\theta|x)$ aantonen door de stelling van Fubini toe te passen voor niet-negatieve, meetbare functies [13].

$$\iint_{V_1} g_1(\theta|x)dAd\sigma = \int_0^\infty \left(\int_0^{\sqrt{\frac{A}{\delta}}} g_1(\theta|x)d\sigma \right) dA$$

Om deze integraal te berekenen, maken we gebruik van de onvolledige gamma functie, die gedefinieerd is als [9]

$$\begin{aligned}\Gamma(a,b) &= \Gamma(a) - \int_0^b t^{a-1}e^{-t}dt \\ &= \int_b^\infty t^{a-1}e^{-t}dt\end{aligned}$$

Nu kan worden aangetoond dat

$$\int_0^{\sqrt{\frac{A}{\delta}}} g_1(\theta|x)d\sigma = \int_0^{\sqrt{\frac{A}{\delta}}} \frac{1}{\sigma^{2n+3-\epsilon}} \exp\left(-\frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2} \right) d\sigma$$

$$\text{Substitueer } t = \frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2} \text{ en } dt = -\frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2}$$

$$\begin{aligned}\int_0^{\sqrt{\frac{A}{\delta}}} g_1(\theta|x)d\sigma &= \int_{t_0}^{t_1} \frac{2^{n-\frac{\epsilon}{2}}}{\left(\sum_{i=1}^n (x_i - A)^2\right)^{\tilde{n}}} t^{\tilde{n}-1} \exp(-t)dt \\ &= -\frac{2^{n-\frac{\epsilon}{2}}}{\left(\sum_{i=1}^n (x_i - A)^2\right)^{\tilde{n}}} \left[\Gamma\left(\tilde{n}, t\right)\right]_{t_1}^{t_0} \\ &= \frac{2^{n-\frac{\epsilon}{2}}}{\left(\sum_{i=1}^n (x_i - A)^2\right)^{\tilde{n}}} \left[\Gamma\left(\tilde{n}, \frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2}\right)\right]_0^{\sqrt{\frac{A}{\delta}}}\end{aligned}$$

met $\tilde{n} = n + 1 - \frac{\epsilon}{2}$. Door de grenzen in te vullen, verkrijgen we

$$\int_0^{\sqrt{\frac{A}{\delta}}} g_1(\theta|x)d\sigma = 2^{n-\frac{\epsilon}{2}} \left(\sum_{i=1}^n (x_i - A)^2 \right)^{-\tilde{n}} \Gamma\left(\tilde{n}, \frac{\delta \sum_{i=1}^n (x_i - A)^2}{2A}\right)$$

aangezien $\Gamma(\tilde{n}, \infty) = 0$. Uit de definitie van $\Gamma(.,.)$ volgt meteen dat $\Gamma(a,b) \leq \Gamma(a,0) = \Gamma(a)$. Dit geeft ons meteen een bovengrens, namelijk

$$\int_0^{\sqrt{\frac{A}{\delta}}} g_1(\theta|x)d\sigma \leq 2^{n-\frac{\epsilon}{2}} \left(\sum_{i=1}^n (x_i - A)^2 \right)^{-\tilde{n}} \Gamma(\tilde{n})$$

Als we de situatie uitsluiten waar alle x_i hetzelfde zijn, is $\sum_{i=1}^n (x_i - A)^2$ een tweedegraadsvergelijking zonder nulpunten. We weten ook dat $\tilde{n} > 1$, dus de volgende integraal over A bestaat

$$\int_0^\infty \left(\sum_{i=1}^n (x_i - A)^2 \right)^{-\tilde{n}} dA.$$

Bijgevolg zijn ook $g_1(\theta|x)$ en $f_0(\theta|x)$ integreerbaar over V_1.
Voor de tweede verzameling V_2 geldt $\frac{A}{\sigma^2} < \delta$, dus kunnen we niet langer gebruik maken van lemma 3.2, maar wel van het stijgend zijn van I_0. We krijgen dus

$$f_0(\theta|x) \leq \frac{\prod_{i=1}^n x_i}{A^{1-\epsilon}\sigma^{2n+1+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n (x_i^2 + A^2)}{2\sigma^2} \right) \prod_{i=1}^n I_0(x, \delta).$$

Aangezien $\frac{A}{\sigma^2} < \delta$ equivalent is met $A < \sigma^2\delta$, verkrijgen we als $\epsilon \geq 1$

$$\begin{aligned} f_0(\theta|x) &< \frac{\delta^{\epsilon-1}\sigma^{2\epsilon-2}}{\sigma^{2n+1+\epsilon}} \exp\left(-\frac{\sum_{i=1}^n (x_i^2 + A^2)}{2\sigma^2} \right) \prod_{i=1}^n x_i I_0(x, \delta) \\ &= \tilde{K} \frac{1}{\sigma^{2n+3-\epsilon}} \exp\left(-\frac{\sum_{i=1}^n (x_i^2 + A^2)}{2\sigma^2} \right) \qquad =: \tilde{K} g_2(\theta|x). \end{aligned}$$

Hierbij is $\tilde{K} = \delta^{\epsilon-1} \prod_{i=1}^n (x_i I_0(x, \delta))$. Door opnieuw gebruik te maken van de stelling van Fubini, krijgen we

$$\iint_{V_2} g_2(\theta|x) dA d\sigma = \int_0^\infty \left(\int_{\sqrt{\frac{A}{\delta}}}^\infty g_2(\theta|x) d\sigma \right) dA$$

Volledig analoog als voor g_1, kunnen we aantonen dat

$$\begin{aligned} \int_{\sqrt{\frac{A}{\delta}}}^\infty g_2(\theta|x) d\sigma &= \frac{2^{n-\frac{\epsilon}{2}}}{\left(\sum_{i=1}^n (x_i - A)^2\right)^{\tilde{n}}} \left[\Gamma\left(\tilde{n}, \frac{\sum_{i=1}^n (x_i - A)^2}{2\sigma^2} \right) \right]_{\sqrt{\frac{A}{\delta}}}^\infty \\ &= \frac{2^{n-\frac{\epsilon}{2}}}{\left(\sum_{i=1}^n (x_i - A)^2\right)^{\tilde{n}}} \left(\Gamma(\tilde{n}) - \Gamma\left(\tilde{n}, \frac{\delta \sum_{i=1}^n (x_i - A)^2}{2A} \right) \right) \\ &\leq \frac{2^{n-\frac{\epsilon}{2}}}{\left(\sum_{i=1}^n (x_i - A)^2\right)^{\tilde{n}}} \Gamma(\tilde{n}) \end{aligned}$$

Aangezien ook $\sum_{i=1}^{n}(x_i - A)^2$ een tweedegraadsvergelijking zonder nulpunten is, volgt volgens dezelfde redenering als hiervoor dat g_2 en f_0 integreerbaar zijn over V_2 als $\epsilon \geq 1$. $\square$

Deze stelling geeft ons de waarden van ϵ waar we uit mogen kiezen, maar zegt ons niet welke schatters deze waarden opleveren of welke van deze schatters optimaal is. De a-postiori dichtheidsfunctie uit vergelijking (4.4) kan men interpreteren als een soort likelihood-functie en de schatter die we hieruit afleiden is dan ook analoog

$$\hat{\theta}_{MAP} = \arg\max_{\theta} f(\theta|x). \tag{4.5}$$

We noemen deze schatter de maximum a-postiori schatter (MAP-schatter). Om deze te vinden kijken we waar de gradiënt gelijk is aan 0 voor de functie

$$\log f(\theta|x) = \log(c) - \log(A^{1-\epsilon}\sigma^{2n+1+\epsilon}) - \sum_{i=1}^{n} \frac{x_i^2 + A^2}{2\sigma^2} + \sum_{i=1}^{n} \log\left(x_i I_0\left(\frac{x_i A}{\sigma^2}\right)\right)$$

Als we $m_i = x_i A/\sigma^2$ noteren, geeft dit ons de volgende partiële afgeleiden

$$\begin{aligned}
\frac{\partial \log f(\theta|x)}{\partial A} &= (\epsilon - 1)\frac{1}{A} - n\frac{A}{\sigma^2} + \sum_{i=1}^{n} \frac{x_i}{\sigma^2}\frac{I_1(m_i)}{I_0(m_i)} \\
\frac{\partial \log f(\theta|x)}{\partial \sigma} &= -\frac{2n+1+\epsilon}{\sigma} + \sum_{i=1}^{n} \frac{x_i^2 + A^2}{\sigma^3} - 2\sum_{i=1}^{n} \frac{x_i A}{\sigma^3}\frac{I_1(m_i)}{I_0(m_i)}
\end{aligned}$$

Als we deze twee dan gelijk stellen aan 0, is de MAP-schatter de oplossing van

$$\begin{cases} (\epsilon - 1)\frac{\sigma^2}{n} - A^2 + \frac{A}{n}\sum_{i=1}^{n} x_i \frac{I_1(m_i)}{I_0(m_i)} & = 0 \\ (-2n - 3 + \epsilon)\frac{\sigma^2}{n} - A^2 + \frac{1}{n}\sum_{i=1}^{n} x_i^2 & = 0 \end{cases} \tag{4.6}$$

Waarbij we de tweede vergelijking verkrijgen door de eerste hier in te substitueren. Als we even terugkijken naar de vergelijkingen (3.5) en (3.6) voor de ML-schatter, dan kun je zien dat het verschil tussen de vergelijkingen gelijk is aan $(\epsilon - 1)\frac{A\sigma^2}{n}$ en $(3 - \epsilon)\frac{\sigma^2}{n}$. Hieruit volgt dat de MAP-schatter b.o. convergeert naar de ML-schatter voor $n \to \infty$, wat betekent dat ze dezelfde mooie asymptotische eigenschappen hebben.
Voor de vergelijkingen (4.6) bestaat er een hypothese dat deze een unieke oplossing hebben als $\epsilon > 1$, [3], [1]. Met deze beperking op ϵ kunnen we dus

volledig het risico uitsluiten dat we zouden convergeren naar de (verkeerde) nul-oplossing van A. Anderzijds zullen deze vergelijkingen dan ook nooit $A = 0$ als antwoord geven wanneer dit daadwerkelijk het juiste antwoord zou zijn voor de ML-schatter. Of men dan kiest voor $\epsilon = 1$ of $\epsilon > 1$, ligt eraan of men het geval $A = 0$ al dan niet wenst uit te sluiten.
Een universeel optimale waarde vinden voor ϵ is nog een open probleem [1], dus wordt er meestal met een heuristiek gewerkt. Deze wordt het beste gekozen tussen 1.01 en 2 [1].

5 De Koay-inversie techniek

De laatste schatter die we zullen bespreken, is degene die verkrijgen wordt via de Koay-inversietechniek. We gaan ervan uit dat het gemiddelde en de variantie gekend zijn, aangezien we deze kunnen vervangen door het steekproefgemiddelde en de steekproefvariantie. De parameters die we willen schatten zijn nog steeds A en σ. Hierbij zullen we gebruik maken van het eerste en het tweede moment van $X \sim Rice(A, \sigma)$ (genoteerd als $\mathbb{E}[X]$ en $\mathbb{E}[X^2]$). Volgens [5] zijn dit

$$\mathbb{E}[X] = \sigma \left(\exp\left(-\frac{A^2}{4\sigma^2}\right) \sqrt{\frac{\pi}{2}} \left[\left(\frac{A^2}{2\sigma^2} + 1 \right) I_0 \left(\frac{A^2}{4\sigma^2} \right) + \frac{A^2}{2\sigma^2} I_1 \left(\frac{A^2}{4\sigma^2} \right) \right] \right)$$

$$\mathbb{E}[X^2] = 2\sigma^2 + A^2$$

Per definitie is de variantie $Var[X]$ gelijk aan $\mathbb{E}[X^2] - \mathbb{E}[X]^2$. Dus als we $\theta = A/\sigma$ noteren voor de SNR, dan kunnen we de variantie uitdrukken als

$$Var[X] = \sigma^2 \xi(\theta) \tag{5.1}$$

met

$$\xi(\theta) = 2 + \theta^2 - \frac{\pi}{8} \exp\left(-\frac{\theta^2}{2}\right) \left((2 + \theta^2) I_0 \left(\frac{\theta^2}{4} \right) + \theta^2 I_1 \left(\frac{\theta^2}{4} \right) \right)^2$$

Hierbij noemen we $\xi(\theta)$ de correctiefactor. Deze functie geeft de relatie weer tussen de variantie van de Rice-verdeling en de variantie van de onderliggende normale verdeling. Aangezien $Var[X]$ ook kunnen schrijven als $2\sigma^2 + A^2 - \mathbb{E}[X]^2$, krijgen we

$$2\sigma^2 + A^2 - \mathbb{E}[X]^2 = \xi(\theta)\sigma^2 \tag{5.2}$$

$$\Leftrightarrow 2 + \theta^2 - \frac{\mathbb{E}[X]^2\xi(\theta)}{Var[X]} = \xi(\theta)$$

$$\Leftrightarrow \theta = g(\theta) := \sqrt{\xi(\theta)\left(1 + \frac{\mathbb{E}[X]^2}{Var[X]}\right) - 2}. \tag{5.3}$$

Er kan bewezen worden dat de functie g uit deze laatste vergelijking een uniek fixpunt heeft zolang $(\mathbb{E}[X]^2/Var[X] \geq \pi/(4-\pi)$. Deze ondergrens is de waarde die men bekomt als men θ gelijk stelt aan zijn ondergrens, namelijk 0.

$$0 = \sqrt{\xi(0)\left(1 + \frac{\mathbb{E}[X]^2}{Var[X]}\right) - 2}.$$

$$\Leftrightarrow 0 = \left(2 - \frac{\pi}{2}\right)\left(1 + \frac{\mathbb{E}[X]^2}{Var[X]}\right) - 2$$

$$\Leftrightarrow \frac{\mathbb{E}[X]^2}{Var[X]} = \frac{\pi}{4-\pi}$$

Het fixpunt kan men onder andere vinden door g herhaaldelijk toe te passen op een initiële waarde θ_0 [5]. Zodra we een schatter hebben voor θ, kunnen we hieruit schatters voor A en σ halen omdat uit (5.1) en (5.2) volgt dat

$$\sigma^2 = \frac{Var[X]}{\xi(\theta)}$$

$$A^2 = \mathbb{E}[X]^2 + (\xi(\theta) - 2)\sigma^2$$

$$\text{of } A^2 = \mathbb{E}[X]^2 + (1 - \frac{2}{\xi(\theta)})Var[X].$$

Helaas convergeert deze fixpuntmethode vrij langzaam [5], dus is er een sneller algoritme ontwikkeld op basis van vergelijking (5.3). In plaats van het fixpunt van $g(\theta)$ te zoeken, zoekt men het nulpunt van $f(\theta) = g(\theta) - \theta$ aan de hand van de methode van Newton. Bij deze methode wordt voor een gegeven gok θ_i een volgende, betere gok θ_{i+1} geconstrueerd door

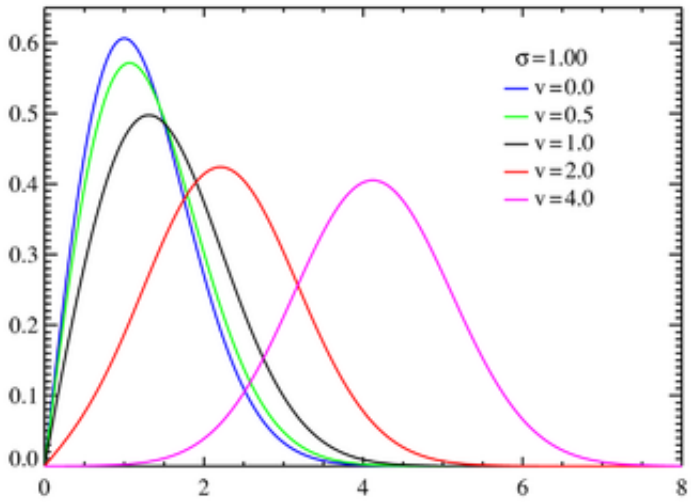
$$\theta_{i+1} = \theta_i - \frac{f(\theta_i)}{f'(\theta_i)}$$

Een voorwaarde om deze methode te mogen gebruiken is dat $f(\theta)$ twee keer continu afleidbaar moet zijn. Als samenstelling van gladde functies is $f(\theta)$ zelf ook glad, dus is deze voorwaarde meteen voldaan.
Men kan zich natuurlijk afvragen wat de kwaliteit van deze schatter is en welke eigenschappen deze heeft. Via simulaties waarbij men de gemiddelde kwadratische fout voor verschillende schatters bestudeerde, is gebleken dat deze schatter goed presteert en zelfs van vergelijkbare kwaliteit als de MAP-schatter [1]. Het voordeel van deze schatter is dat er hier slechts het nulpunt wordt gezocht in één vergelijking met één parameter in tegenstelling tot zowel de ML- als de MAP-schatter, waar we de oplossing zoeken van twee, meestal impliciete vergelijkingen met twee verschillende parameters. Deze oplossing is uniek en niet betekenisloos, d.w.z. negatief of complex. Het nadeel van deze schatter is dat deze geen bewezen statistische eigenschappen heeft [1].

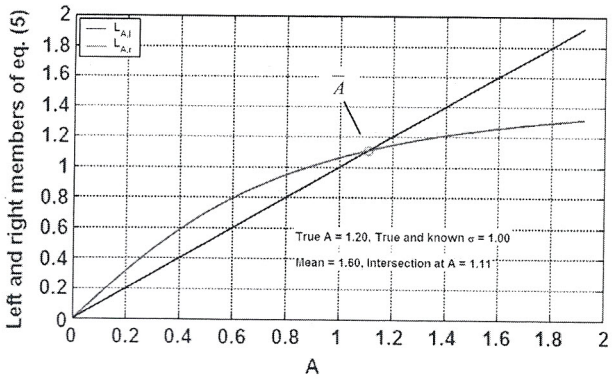
6 Vergelijking van de schatters

We hebben in deze bachelorproef vijf verschillende schatters besproken. Het steekproefgemiddelde en de MOM-schatter hebben we verworpen op basis van het feit dat ze slecht presteren voor lagere SNR. Via simulaties is aangetoond in [1] dat voor zeer lage SNR (0,5 voor steekproefgrootte n=100; 0,75 voor n=20) de ML-schatter beter presteert dan zowel de MAP-schatter als de methode van Koay ten opzichte van de root mean square error (RMSE). Een gedeeltelijke verklaring hiervoor is dat voor lage SNR de echte amplitude niet veraf ligt van $A = 0$ zodat als de ML-schatter convergeert naar de foute oplossing met $A = 0$ de afwijking niet zeer groot is. Desondanks presteert de ML-schatter voor lage SNR nog steeds beter dan de MAP-schatter, ook als men deze enkel vergelijkt met de correcte ML-oplossing.
Voor lage tot gematigde SNR presteren de MAP- en de Koay-schatter beter dan de ML-schatter. Voor σ presteert de Koay-inversie techniek duidelijk slechter dan de twee andere, maar aangezien men hoofdzakelijk geïnteresseerd is in de amplitude is dat geen argument om deze methode niet te gebruiken. Zowel de Bayesiaanse methode als de methode van Koay garanderen ook een oplossing onafhankelijk van de kwaliteit van de startwaarde [5], [1]. Het nadeel van Koay ten opzichte van de MAP-schatter is dat de RMSE voor deze schatter niet monotoon daalt, waardoor men niet kan garanderen dat de kwaliteit van de schatter voor hogere SNR niet afneemt.

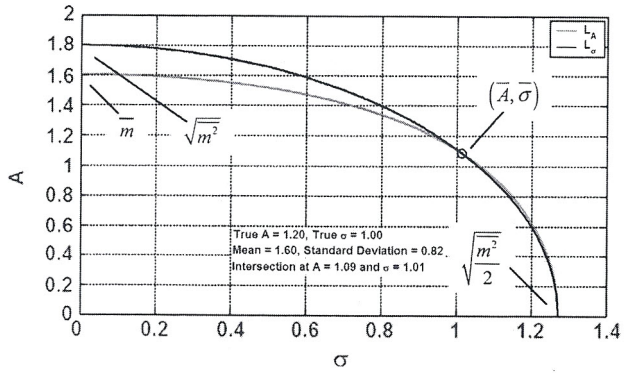
Bijlage A Grafieken



Figuur 1: De Rice-verdeling voor verschillende SNR, bron: Wikipedia



Figuur 2: Plot $LL(A)$ en $RL(A)$, hier respectievelijk genoteerd als $L_{A,l}$ en $L_{A,r}$, bron: [4]



Figuur 3: Plot van L_1 en L_2, hier respectievelijk genoteerd als L_A en L_σ, bron: [4]

7 Referenties

[1] L. Lauwers, *"Some practical applications of the best linear approximation in nonlinear block-oriented modelling"*, University Press, Zelzate, Hoodstuk 6: Parameter Estimation of the Rice Distribution, 2011.

[2] M.A. Lindquist, "The Statistical Analysis of fMRI Data", *Statistical Science*, Vol. 23, No.4, pp. 439-464, 2008

[3] L. Lauwers, K. Barbé, W. Van Moer, R. Pintelon, "Analyzing Rice distributed functional Magnetic Resonance Imaging (fMRI) data: a Bayesian approach", *Measurement Science and Technology*, Vol. 21, No. 11, 2010.

[4] C.F.M. Carobbi, M. Cati, "The Absolute Maximum of the Likelihood Function of the Rice Distribution: Existence and Uniqueness", *IEEE Transactions on Instrumentation and Measurement*, Vol. 57, pp. 682-689, 2008.

[5] C.G. Koay, P.J. Basser, "Analytically Exact Correction Scheme for Signal Extraction from Noisy Magnitude MR Signals", *Journal of Magnetic Resonance*, Volume 179, No. 2, p. 317322, 2006.

[6] K. Barbé, W. Van Moer, L. Lauwers, "Functional Magnetic Resonance Imaging: An Improved Short Record Signal Model", *IEEE Transactions on Instrumentation and Measurement*, Vol. 60, No. 5, pp. 1724-1731, 2011.

[7] J. Sijbers, A. J. den Dekker, P. Scheunders, D. Van Dyck, "Maximum Likelihood Estimation of Rician Distribution Parameters", *IEEE Transactions on Medical Imaging*, Vol. 17, No. 3, pp.357-361, 1998

[8] G.B. Arfken, H.J. Weber, *"Essential Mathematical Methods for Physicist"*, Academic Press, London, 2003.

[9] I.A. Stegun, M. Abramowitz, *"Handbook of Mathematical Functions"*, 10de ed., Dover Publications, New York, 1965.

[10] D.E. Amos, "Computation of Modified Bessel Functions and Their Ratios", *Mathematics of Computation*, Vol. 28, No. 125, pp.239-251, 1974.

[11] E.T. Jaynes, *"Probability Theory: the Logic of Science"*, Cambridge University Press, Cambridge, 2003.

[12] J. Sijbers, *"Signal and Noise Estimation from Magnetic Resonance Images"*, UA, Antwerpen, 1998.

[13] U. Einmahl, *"Maattheorie"*, VUB, Brussel, 2013.