Cramér-Rao analysis of Inversion Recovery sequence:

Internal Report Number, April 15, 2014

Gabriel Ramos-Llordén
iMinds-Vision Lab, Department of Physics, University of Antwerp, Belgium

CONTENTS

THEORETICAL STUDY

Cramér-Rao bound Theorem

Let $\hat{\boldsymbol{\theta}}(\mathbf{x}) = \{\hat{\theta_1}, \hat{\theta_2}, ..., \hat{\theta_K}\}$ be any unbiased estimate of the unknown vector parameter $\boldsymbol{\theta} = \{\theta_1, \theta_2, ..., \theta_K\}$ and let $f(\mathbf{x};\boldsymbol{\theta})$ be the probability density function of the random vector variable $\mathbf{x} = [x_1, x_2, ...x_N]^T$, then, provided some not too strict regularity conditions (Papoulis and Pillai, 2002),

$$Cov\{\hat{\boldsymbol{\theta}}(\mathbf{x})\} \succeq \mathbf{I}^{-1}(\boldsymbol{\theta}) \qquad [1]$$

where $Cov\{\hat{\boldsymbol{\theta}}(\mathbf{x})\}$ is the covariance matrix of $\hat{\boldsymbol{\theta}}(\mathbf{x})$, $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher Information associated to $\boldsymbol{\theta}$ under the pdf $f(\mathbf{x};\boldsymbol{\theta})$ and the symbol $\succeq$ means that the matrix difference $Cov\{\hat{\boldsymbol{\theta}}(\mathbf{x})\} - \mathbf{I}^{-1}(\boldsymbol{\theta})$ is a non negative definite matrix. Using the proposition 1 in V, this straightforwardly implies that $Var\{\hat{\theta_k}(\mathbf{x})\} \geq (\mathbf{I}^{-1}(\boldsymbol{\theta}))_{kk}$, i.e, the variance of the estimators are bounded by the k^{th} entry of the diagonal of the inverse of Fisher Information.

Fisher Information to composition of models

With composition of models we refer to the case when the pdf of some random vector $\mathbf{x} \in \mathbb{R}^N$, $f(\mathbf{x};\boldsymbol{\theta})$, can be written as $f(\mathbf{x};\boldsymbol{\theta}) = p(\mathbf{x};\mathbf{S}(\boldsymbol{\theta})) = \prod_i^N p(x_i; S_i(\boldsymbol{\theta}))$ and $p(x_i;S_i)$ is known (it has been assumed that the random samples x_i are IID). This is the case when the classic parameters of a distribution, e.g, mean and standard deviation in the Gaussian distribution, envelope and σ in the Rician distribution, etc... depend themselves on some vector parameter $\boldsymbol{\theta}$. In that cases, the Fisher Information,

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E}\{\nabla_{\boldsymbol{\theta}} \log f(\mathbf{x};\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}} \log f(\mathbf{x};\boldsymbol{\theta})^T\} \quad [2]$$

where $\nabla_{\boldsymbol{\theta}} \log f(\mathbf{x};\boldsymbol{\theta}) = [\frac{\partial \log f(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_1}, \frac{\partial \log f(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_2}, ..., \frac{\partial \log f(\mathbf{x};\boldsymbol{\theta})}{\partial \theta_K}]^T$ can be decomposed as (Poot, 2010):

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^{N}\sum_{j=1}^{N} \nabla_{\boldsymbol{\theta}} S_i \nabla_{\boldsymbol{\theta}} S_j{}^T \mathbb{E}\{\frac{\partial \log p(x_i)}{\partial S_i}\frac{\partial \log p(x_j)}{\partial S_j}\} \quad [3]$$

For $i \neq j$, because x_i is independent of x_j,

$$\mathbb{E}\{\frac{\partial \log p(x_i)}{\partial S_i}\frac{\partial \log p(x_j)}{\partial S_j}\} = \mathbb{E}\{\frac{\partial \log p(x_i)}{\partial S_i}\}\mathbb{E}\{\frac{\partial \log p(x_j)}{\partial S_j}\} \quad [4]$$

The resultant expectations are the expectations of the score of S_i and S_j respectively, which under regularity conditions are always zero (Papoulis and Pillai, 2002). Finally,

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} S_i \nabla_{\boldsymbol{\theta}} S_i{}^T \mathbb{E}\{(\frac{\partial \log p(x_i)}{\partial S_i})^2\} \quad [5]$$

The N matrices $\nabla_{\boldsymbol{\theta}} S_i \nabla_{\boldsymbol{\theta}} S_i{}^T \quad i = 1, ..., N$ are called structure tensor and represent the variations of the 2D-functions S_i, $i = 1, ...N$. An inspection of eq.([7]) reveals that the Fisher information depends on the deterministic model, (structure tensor) and the probabilistic one : $\mathbb{E}\{(\frac{\partial \log p(x_i)}{\partial S_i})^2\}$. In order to produce higher values of the Fisher information and then, decrease the Cramér Rao bound, the deterministic model has to be the more sensitive to the parameters to be estimated as possible (Kay, 1993) (e.g, high derivatives). However, the structure tensors are weighted by the expectation which depends on the type of probabilistic distribution. A rigorous analysis may imply to make an spectral analysis in terms of eigenvalues and eigenvectors to the structure tensors, because it will reveal the directions of maximum variation. Moreover, the term *higher values of Fisher information* is not rigorous because we are not dealing with scalar functions. Nevertheless, this is out of scope of this work. Therefore, we will retain the intuitively and reasonably idea that the highest the derivatives of the model are with respect to the parameters, the *higher* the Fisher information matrix is .

APPLICATION TO THE TWO-PARAMETER INVERSION RECOVERY (IR) SEQUENCE

This section is focused on the application of previous results to the two parameter Inversion Recovery (IR) sequence,

$$S_i(\boldsymbol{\theta}) = S_i(\rho, T_1) = \rho|1 - 2e^{-\frac{TI_n}{T_1}}|, \quad i = 1, ..., N \quad [6]$$

The N samples are obtained when sampling the $S(\rho, T_1, t)$ at the vector inversion time $\mathbf{TI} = [TI_1, TI_2, ..., TI_N]^T$. Considering that the random samples x_i, $i = 1, ..., N$ follow a Rician model with magnitude $S_i(\rho, T_1)$ and noise σ, according to (Poot, 2010), the Fisher Information is:

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} S_i \nabla_{\boldsymbol{\theta}} S_i{}^T I_{Rice}(S_i, \sigma) \quad [7]$$

where

$$I_{Rice}(S_i, \sigma) = \int_0^\infty \frac{x_i}{\sigma^2} e^{-\frac{x_i^2+S_i^2}{2\sigma^2}} I_0(\frac{x_i S_i}{\sigma^2})(\frac{x_i I_1(\frac{x_i S_i}{\sigma^2})}{\sigma^2 I_0(\frac{x_i S_i}{\sigma^2})} - \frac{S_i}{\sigma^2})^2 \, dx_i \quad [8]$$

This integral does not possess an analytical closed solution so we have to resort to numerical implementations. Introducing the substitutions $s = \frac{S_i}{\sigma}$ and $\bar{x}_i = \frac{x_i}{\sigma}$, $I_{Rice}(S_i, \sigma) = I_{Rice_{Norm}}(s, \sigma) = \frac{1}{\sigma^2} I_{Rice_{Norm}}(s, 1)$ (Poot, 2010). This last expression, $I_{Rice_{Norm}}(s, 1)$, has been tabulated by Dirk Poot using polynomial interpolation with an accuracy of 2^{-52} approximately.

Influence of Inversion Times

The $I_{Rice}(S_i, \sigma)$ term in the Fisher matrix depends on the level of noise: σ and the values of magnitude S_i. More interesting is the effect of the inversion time vector $\mathbf{TI}$ which determines the behavior of $\nabla_{\boldsymbol{\theta}} S_i \nabla_{\boldsymbol{\theta}} S_i{}^T, i = 1, ..., N$ and therefore performs a crucial role in the goal of maximizing the Fisher information. For that point the analysis of the variability of $S_i(\rho, T_1)$ deserves a deep study. In this work, we will focus on the derivatives of the first order,

$$\nabla_{\boldsymbol{\theta}} S_i{}^T = [\frac{\partial S_i(\rho, T_1)}{\partial \rho}, \frac{\partial S_i(\rho, T_1)}{\partial T_1}] \quad [9]$$

with

$$\frac{\partial S_i(\rho, T_1)}{\partial \rho} = |1 - 2e^{-\frac{TI_i}{T_1}}| \quad [10]$$

and

$$\frac{\partial S_i(\rho, T_1)}{\partial T_1} = \begin{cases} -\frac{2TI_i\rho}{T_1^2} e^{-\frac{TI_i}{T_1}} & \text{if } (1 - 2e^{-\frac{TI_i}{T_1}}) > 0 \\ \frac{2TI_i\rho}{T_1^2} e^{-\frac{TI_i}{T_1}} & \text{if } (1 - 2e^{-\frac{TI_i}{T_1}}) < 0 \end{cases} \quad [11]$$

Fixed T_1 and ρ, both derivatives are functions of the $TI_i, i = 1, ..., N$. It is mandatory to remark that $\frac{\partial S_i(\rho, T_1)}{\partial T_1}$ does not exist for $TI_i = T_1 \log 2$.

Application to the two-parameter Squared Inversion Recovery (IR) sequence

One approach to obtain a continuous derivable probabilistic model consist of avoiding the absolute value of $S_i(\boldsymbol{\theta})$, for example, making the parameters of the model dependent on $S_i^2(\boldsymbol{\theta})$. This is possible by taking the square of the Rician samples $\mathbf{x}$, i.e $\mathbf{r} = \mathbf{x}^2$. According to (Moser, 2007), r_i, for $i = 1, ..., N$, follows a Non Central Chi squared distribution up to a constant, i.e $r_i = \sigma^2 y_i$ where y_i distributed as a Non Central Chi Squared Random Variable with two degrees of freedom and centrality parameter, $\lambda = \frac{S_i^2(\boldsymbol{\theta})}{\sigma^2}$, i.e

$$f(y_i; \lambda) = \frac{1}{2} e^{-\frac{\lambda + y_i}{2}} I_0(\sqrt{\lambda y}) = \frac{1}{2} e^{-\frac{\frac{S_i^2(\boldsymbol{\theta})}{\sigma^2} + y_i}{2}} I_0(\sqrt{\frac{S_i^2(\boldsymbol{\theta})}{\sigma^2} y_i}) \quad y_i \geq 0 \quad [12]$$

Therefore, considering the parameters of the distribution to be $S_i^2(\boldsymbol{\theta})$ and σ, the probabilistic model is derivable everywhere, because $\nabla_{\boldsymbol{\theta}} S_i^2$ exists. The Fisher information turns out to be,

$$\mathbf{I}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \nabla_{\boldsymbol{\theta}} S_i^2 \nabla_{\boldsymbol{\theta}} S_i^{2\,T} I_{NonCentral\chi^2}(S_i^2, \sigma) \quad [13]$$

where $I_{NonCentral\chi^2}(S_i^2, \sigma)$ is the equivalent expression to eq.([8]) but applied to Non Central Chi Square Data (derived in the appendix) and

$$\nabla_{\boldsymbol{\theta}} S_i^{2\,T} = [\frac{\partial S_i^2(\rho, T_1)}{\partial \rho}, \frac{\partial S_i^2(\rho, T_1)}{\partial T_1}] \quad [14]$$

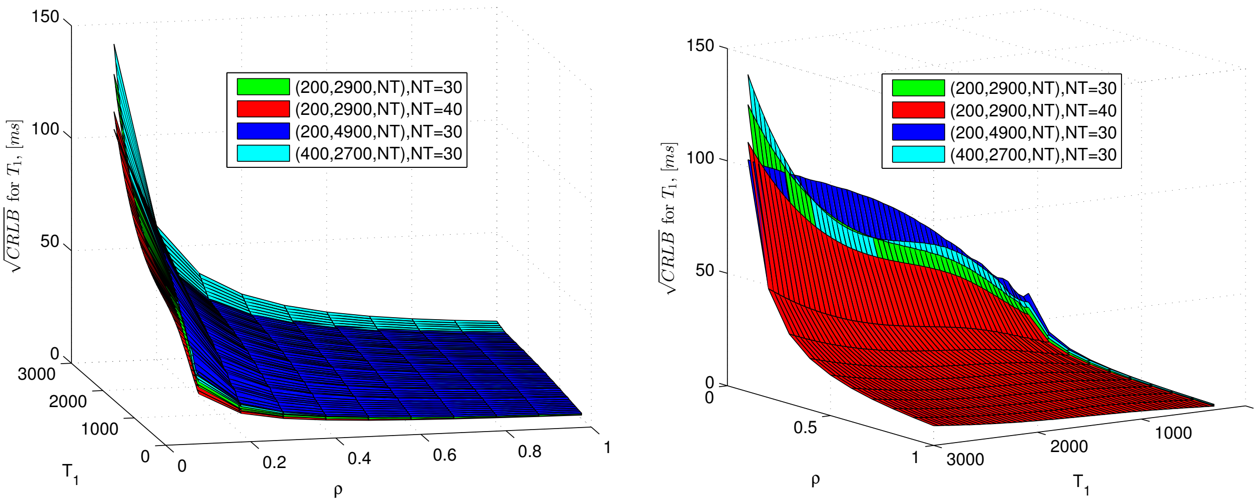
with

$$\frac{\partial S_i^2(\rho, T_1)}{\partial \rho} = 2\rho(1 - 2e^{-\frac{TI}{T_1}})^2 \quad [15]$$

$$\frac{\partial S_i^2(\rho, T_1)}{\partial T_1} = -\frac{4\rho^2 TI}{T_1^2} e^{-\frac{TI}{T_1}} (1 - 2e^{-\frac{TI}{T_1}}) \quad [16]$$

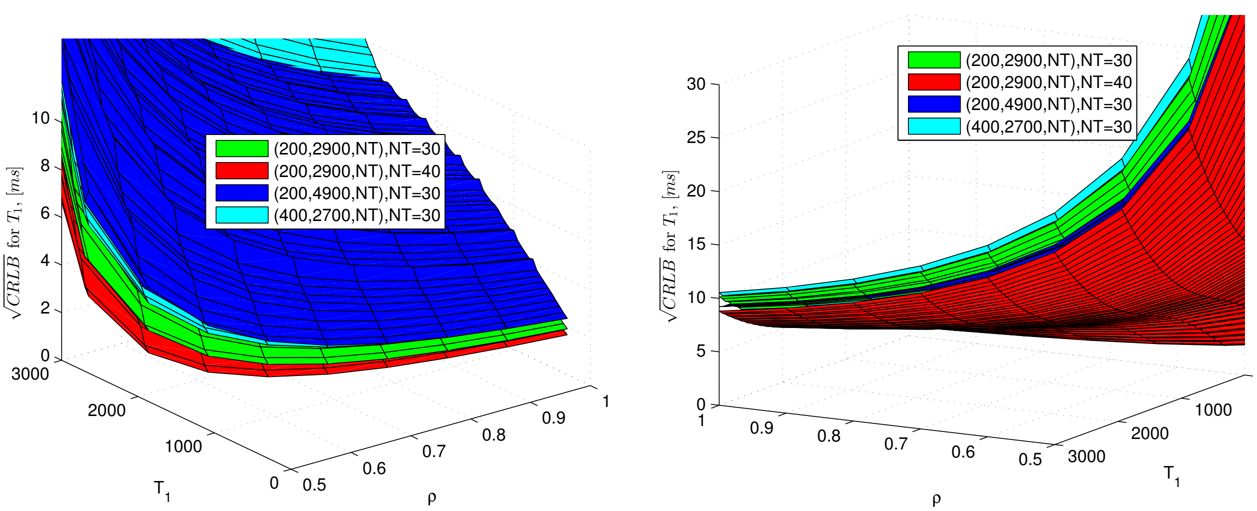
Visualizing the CRLB

The first experiment shows the CRLB as a function of T_1 and ρ for $\sigma = 0.01$. Four different acquisition configurations are performed,i.e, different Inversion Times. The purpose of this experiment is to illustrate the effect of the inversion time vector $\mathbf{TI}$ produces in the final results, both taking into account its length and the sampling points. The higher is the number of samplings, the lower will be the CRLB. This is not surprising. What is not clear is the optimal configuration when the number of Inversion Times are fixed. In Fig.1, the square root of CRLB (the units are in ms) is shown for the four configurations. The values of the square root of CRLB are also shown in Fig.2 for the region $\rho \in [0.5, 1]$.



(a) Global view (b) Bottom view

Figure 1. CRLB visualization for four different configurations of inversion times. $\sigma = 0.01$.



(a) Global view (b) Bottom view

Figure 2. Previous graphics restricted to $\rho \in [0.5, 1]$.

In all cases, the CRLB tends to increase when $\rho \to 0$ because the SNR becomes small. The worst configuration is that whose inversion times are between 200 ms and 4900 ms (blue surface). One can ask himself that for high inversion times, the SNR is high enough to make the CRLB smaller. However and because the number of inversion times are fixed, the sampling scheme is not dense enough to sample near to the cross by zero point, points that provide high information. As it was expected, the best configuration is that with the higher number of inversion times (red surface). The green and light blue surfaces represent two similar configuration (same number of inversion times) but starting with a different initial value. It cannot be assured that one configuration is better than the other, because depending on the region $T_1 - \rho$, light surface crosses the green one and vice versa. This can be more clearly appreciated in the Fig.3.

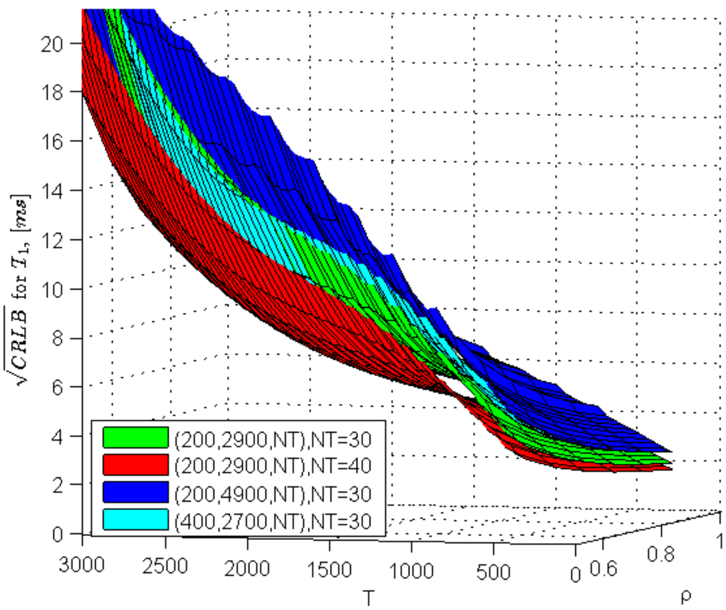
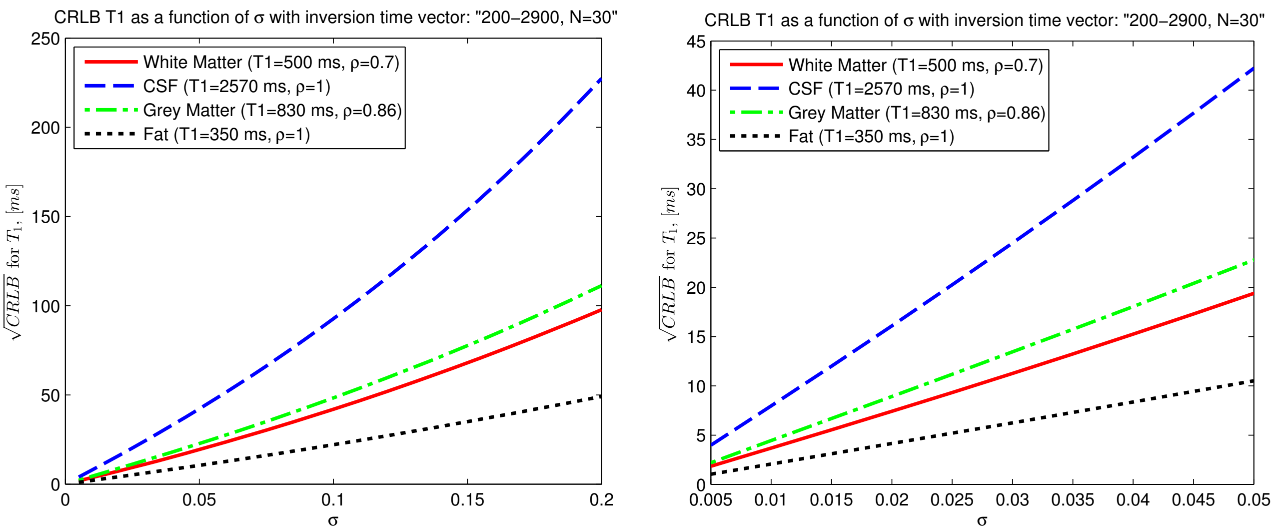


Figure 3. Illustrative view where the crosses between CRLB surfaces can be observed.

In the second experiment, the CRLB for four different brain tissues: white matter, grey matter, CSF and Fat, are shown as a function of σ. This allows to understand the effect of noise in the precision of the estimates. This is illustrated in Fig.4 where the same curves are shown in different graphics varying the axis. The purpose is to make a zoom in the region of reasonable noise level, $\sigma < 0.1$.



(a) Region $\sigma \in [0.005, 0.2]$ (b) Region $\sigma \in [0.005, 0.05]$

Figure 4. CRLB for white matter, grey matter, CSF and fat depending of σ.

The higher the value of T_1 is, the higher is the CRLB. This do not probably represent a serious problem, because coarser errors can be allowed for high T_1 than for small T_1. The higher σ is, the higher CRLB turns out to be.

APPLICATION TO THE THREE-PARAMETER SQUARED INVERSION RECOVERY (IR) SEQUENCE

APPENDIX

Diagonal entries of a semidefinite positive matrix are always non negative

Let $\mathbf{A}$ a non negative definite matrix, i.e $\mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0$ for any $\mathbf{v}$. Then if $\mathbf{v} = [1, \mathbf{0}]^T$, $\mathbf{A}\mathbf{v} = [a_{11}, a_{21}, ..., a_{N1}]^T$ and hence $\mathbf{v}^T\mathbf{A}\mathbf{v} = a_{11}$. But $\mathbf{v}^T\mathbf{A}\mathbf{v} \geq 0$, so $a_{11} \geq 0$. In a similar way, let $\mathbf{v}_k$ the k^{th} canonical vector, i.e, it contains zeros except in the k^{th} position, where it is 1. Therefore, $\mathbf{A}\mathbf{v}_k = [a_{1k}, a_{2k}, ..., a_{Nk}]^T$, hence $\mathbf{v}_k^T\mathbf{A}\mathbf{v}_k = a_{kk}$, which again with the non negative definite property, it leads to $a_{kk} \geq 0$. Therefore, the diagonal entries of a non negative definite matrix are always non negative. Applying this result to $\mathbf{A} = Cov\{\hat{\boldsymbol{\theta}}(\mathbf{x})\} - \mathbf{I}^{-1}(\boldsymbol{\theta})$ the bound of the variance is proven.

Fisher Information for Non Central Chi Squared

The expression $I_{NonCentral\chi^2}(S_i^2, \sigma))$ takes the form of

$$I_{NonCentral\chi^2}(S_i^2, \sigma) = \int_0^\infty \left(\frac{\partial \log f(y_i;\lambda)}{\partial S_i^2}\right)^2 f(y_i;\lambda)dy_i = \frac{1}{\sigma^4}\int_0^\infty \left(\frac{\partial \log f(y_i;\lambda)}{\partial \lambda}\right)^2 f(y_i;\lambda)dy_i = \quad [17]$$

$$\frac{1}{\sigma^4}\int_0^\infty \left(\frac{I_1(\sqrt{\lambda y_i})}{I_0(\sqrt{\lambda y_i})}\frac{y_i}{2\sqrt{\lambda y_i}} - \frac{1}{2}\right)^2 \frac{1}{2}e^{-\frac{\lambda+y_i}{2}} I_0(\sqrt{\lambda y_i})dy_i = \frac{1}{\sigma^4}\mathcal{I}(\lambda)$$

The integral $\mathcal{I}(\lambda)$ is approximated with the trapezoidal rule in an interval $[0, y_i^*]$ where the truncation error (defined as the integral in $[y_i^*, \infty))$ can be bounded by Marcum Q function and made it negligible. In practice, $y_i^* = 50\lambda$ has been used.

REFERENCES

(Kay, 1993) Kay, S. M. (1993). Fundamentals of statistical signal processing, volume i: Estimation theory (v. 1).

(Moser, 2007) Moser, S. M. (2007). Some expectations of a non-central chi-square distribution with an even number of degrees of freedom. In *TENCON 2007-2007 IEEE Region 10 Conference*, pages 1–4. IEEE.

(Papoulis and Pillai, 2002) Papoulis, A. and Pillai, S. U. (2002). Probability, random variables and stochastic processes with errata sheet. *New York, NY, McGraw-Hill Education*.

(Poot, 2010) Poot, D. H. (2010). *Advances in the reconstruction and statistical processing of Magnetic Resonance Images*. PhD thesis, University of Antwerp.