DIPARTIMENTO DI MATEMATICA
"Francesco Brioschi"
POLITECNICO DI MILANO

Estimation approaches for the Apparent Diffusion Coefficient in Rice-distributed MR signals

Baraldo, S.; Ieva, F.; Mainardi, L.; Paganoni, A. M.

Collezione dei *Quaderni di Dipartimento*, numero **QDD 121**
Inserito negli *Archivi Digitali di Dipartimento* in data 30-3-2012

Piazza Leonardo da Vinci, 32 - 20133 Milano (Italy)

Estimation approaches for the Apparent Diffusion Coefficient in Rice-distributed MR signals

Stefano Baraldo[♯], Francesca Ieva[♯], Luca Mainardi[†] and Anna Maria Paganoni[♯]

March 30, 2012

[♯] MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
piazza Leonardo Da Vinci 32, 20133 Milano, Italy
stefano1.baraldo@mail.polimi.it
francesca.ieva@mail.polimi.it
anna.paganoni@polimi.it
[†] Dipartimento di Bioingegneria
Politecnico di Milano
piazza Leonardo Da Vinci 32, 20133 Milano, Italy
luca.mainardi@polimi.it

Keywords: diffusion tensor imaging, magnetic resonance, nonlinear regression, Rice distribution.

AMS Subject Classification: 62P10, 62F10

Abstract

The Apparent Diffusion Coefficient (ADC) is often considered in the differential diagnosis of tumors, since the analysis of a field of ADCs on a particular region of the body allows to identify regional necrosis. This quantity can be estimated from magnitude signals obtained in diffusion Magnetic Resonance (MR), but in some situations, like total body MRs, it is possible to repeat only few measurements on the same patient, thus providing a limited amount of data for the estimation of ADCs. In this work we consider a Rician distributed magnitude signal with an exponential dependence on the so-called b-value. Different pixelwise estimators for the ADC, both frequentist and Bayesian, are proposed and compared by a simulation study, focusing on issues caused by low signal-to-noise ratios and small sample sizes.

1 Introduction

Diffusion magnetic resonance (MR) is as an important tool in clinical research, as it allows to characterize some properties of biological tissues. When tumor areas are analyzed using this technique, it can be observed that the diffusion tensor, estimated from the magnetic MR magnitude signal, has reduced values in lesions with respect to surrounding physiological tissues, allowing to identify pathological areas or necrosis. When the tissue region of interest can be considered as isotropic the *Apparent Diffusion Coefficient* (ADC) is sufficient to characterize the diffusion properties of the tissue, and it is usually estimated from the exponential decay of the signal with respect to the *b-value*, the MR acquisition parameter. The assumption of isotropy is common and reasonable in various cases, like breast and prostate cancer (see for example Woodhams et al. (2005) and Sato et al. (2005)).

In many practical situations it may not be possible to collect more than few measures at different b-values, limiting the accuracy of the estimation. A reduction in the total number of measures necessary to achieve a certain accuracy is convenient in term of costs, and allows to keep the patient involved in the MR procedure for a shorter amount of time (the experience may be unpleasant, especially when total body MR must be performed). The purpose of this work is to compare different frequentist and Bayesian approaches to the estimation of the ADC, underlining their statistical properties and computational issues.

2 Rice-distributed diffusion MR signals

2.1 The Rice distribution

The random variables we deal with derive from the complex signal $w = w_r + iw_i$ measured in diffusion MR. It is usual to assume that both w_r and w_i are affected by a Gaussian noise with equal, constant variance, i.e. $w_r \sim \mathcal{N}(\nu\cos(\vartheta), \sigma^2)$ and $w_i \sim \mathcal{N}(\nu\sin(\vartheta), \sigma^2)$, with $\nu \in \mathbb{R}^+$ and $\vartheta \in [0, 2\pi)$. The quantity at hand is the modulus M of this signal, which has then a *Rice* (or *Rician*) distribution, that we will denote as $M \sim \text{Rice}(\nu, \sigma^2)$. The density of this random variable has the form

$$f_M(m|\nu, \sigma^2) = \frac{m}{\sigma^2} e^{-\frac{m^2+\nu^2}{2\sigma^2}} I_0\left(\frac{m\nu}{\sigma^2}\right) \mathbb{I}_{(0,+\infty)}(m), \tag{1}$$

where I_0 is the zeroth-order modified Bessel function of the first kind (see Abramowitz and Stegun (1964)). Using the series expression of I_0, it is possible to deduce a different, equivalent definition of a Rician random variable as $M = \sigma\sqrt{R}$, where R is a noncentral χ^2 variable that can be expressed as a mixture of $\chi^2(2P + 2)$ distributions with $P \sim Poisson(\nu^2/2\sigma^2)$. This formulation becomes particularly useful for sampling from a Rice distribution, as it allows an easy implementation of a Gibbs sampler.

2.2 Rice exponential regression

Diffusion MR aims at computing the diffusion tensor field on a portion of tissue, and this is achieved by analyzing the influence of water diffusion on the measured signal, under different experimental settings. In particular, the classical model for relating the magnitude signal to the acquisition parameters and the 3-dimensional diffusion tensor D is the *Stejskal-Tanner* equation

$$\nu_{\mathbf{g}} = \nu_0 \exp(-\mathbf{g}^T D \mathbf{g} b), \tag{2}$$

where $\nu_{\mathbf{g}}$ is the "real" intensity signal we want to measure, ν_0 is the signal at $b = 0$ and the vector $\mathbf{g} \in \mathbb{R}^3$ is the applied magnetic gradient. The b-value is a function of other acquisition settings, which we will omit since their description and discussion beyond the scope of this article. See for example Landini et al. (2005) for an overview on MR techniques, including diffusion MR, and a discussion of various issues and recent advances in this field.

In general, even in the ideal noiseless case, at least 6 observations are needed to determine the components of the symmetric, positive definite diffusion tensor D, by varying the direction $\mathbf{g}$ of the magnetic field gradient. However, if the tissue under study can be considered as isotropic, the diffusion tensor has the simpler form $D = \alpha I$, where α is the ADC, a scalar parameter, and I is the identity matrix. This reduces model (2) to the following

$$\nu = \nu_0 \exp(-\alpha b) \tag{3}$$

for any vector $\mathbf{g}$ (in the following, we will omit it for ease of notation).

Equation (3) describes pointwise the phenomenon on the tissue region of interest. In this study we consider the pixels of a diffusion MR sequence of images as independent, and focus on the estimation problem for a single point in space. We do not consider a spatial modeling for the ADC field: although it could be a useful way to filter noise and to capture underlying tissue structures, on the other hand for diagnostic purposes it may be preferable to submit to the physician an estimate that has not been artificially smoothed.

3 Estimation methods

In this Section we present different methods for the estimation of α, the unknown parameter of interest. We consider a sample of signal intensities on a single pixel $M_i \sim \text{Rice}(\nu_0 e^{-\alpha b_i}, \sigma^2)$, $i = 1, \dots, n$, and their respective realizations $\mathbf{m} = m_1, \dots, m_n$ at b-values $\mathbf{b} = b_1, \dots, b_n$. Since it is usual to estimate the dispersion parameter σ^2 in regions where almost pure noise gets measured, considering it as fixed in the subsequent estimates, in this work we will follow this framework, considering σ^2 as a known parameter. In Section 4 the different estimation methods for the couple (ν_0, α) presented here will be tested under different Signal-to-Noise Ratios (SNRs) ν/σ.

3.1 Nonlinear regression

A standard approach for the estimation of ν_0 and α is to solve a nonlinear least squares problem, which is equivalent to approximating $M_i = \nu_0 exp(-\alpha b_i) + \varepsilon_i$ for $i = 1, \ldots, n$, where ε_i are iid, zero mean, gaussian noise terms. The estimators $\hat{\nu}_0^{LS}$ and $\hat{\alpha}^{LS}$ are defined as

$$(\hat{\nu}_0^{LS}, \hat{\alpha}^{LS}) = \underset{(\nu_0,\alpha)}{\operatorname{argmin}} \sum_{i=1}^{n} (m_i - \nu_0 e^{-\alpha b_i})^2,$$

for $\nu_0, \alpha > 0$, which is equivalent to the solution of the following equations

$$\begin{cases} \nu_0 \sum_{i=1}^{n} e^{-2\alpha b_i} = \sum_{i=1}^{n} m_i e^{-\alpha b_i}, \\ \nu_0 \sum_{i=1}^{n} b_i e^{-2\alpha b_i} = \sum_{i=1}^{n} m_i b_i e^{-\alpha b_i}. \end{cases} \tag{4}$$

The approximation to a nonlinear regression model is inconsistent with the phenomenon under study, most evidently for the fact that in this case the noise term is symmetric and it can assume real values. This inconsistency is negligible for high SNR values, since a Rice(ν, σ^2) distribution in this case approaches a $\mathcal{N}(\nu_0, \sigma^2)$, but becomes important with "intermediate" and low SNRs. In Walker-Samuel et al. (2009), the behavior of the Rice distribution with fixing $\sigma = 1$ and varying ν is examined, observing that normality can be considered a good approximation at about $\nu/\sigma > 2.64$, but the sample variance approaches σ^2 only for SNR values greater than 5.19. Even for pixels with high SNRs at $b = 0$, for large b-values the real signal could reach the same order of magnitude of noise, depending on the unknown value of α, and this could lead to very biased estimates. However, the least squares approach is computationally simpler and quicker to carry out, since it can be seen from (4) that ν_0 can be expressed as a function of α, thus requiring just a one-dimensional optimization to compute the estimates.

3.2 Maximum likelihood

The maximum likelihood approach allows to take into account the asymmetry of the signal distribution, always providing admissible values of the parameters. The objective function is the log-likelihood

$$\begin{aligned} l(\nu_0, \alpha | \mathbf{m}, \mathbf{b}, \sigma^2) = \log L(\nu_0, \alpha | \mathbf{m}, \mathbf{b}, \sigma^2) &= \sum_{i=1}^{n} \log f_{M_i}(m_i | \nu_0 e^{-\alpha b_i}, \sigma^2) \\ &\propto -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \nu_0^2 e^{-2\alpha b_i} + \sum_{i=1}^{n} \log \left[I_0 \left(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2} \right) \right], \end{aligned}$$

where f_{M_i} is the Rice density (1), for $i = 1, \ldots, n$. The ML estimator is then

$$(\hat{\nu}_0^{ML}, \hat{\alpha}^{ML}) = \underset{(\nu_0,\alpha)}{\operatorname{argmax}}\, l(\nu_0, \alpha | \mathbf{m}, \mathbf{b}, \sigma^2),$$

for $\nu_0, \alpha > 0$.

Looking for stationary points of l and using the fact that $I_0'(x) = I_1(x)$, we obtain the following estimating equations

$$\begin{cases} \nu_0 \sum_{i=1}^n e^{-2\alpha b_i} = \sum_{i=1}^n \frac{I_1(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2}) m_i}{I_0(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2})} e^{-\alpha b_i}, \\ \nu_0 \sum_{i=1}^n b_i e^{-2\alpha b_i} = \sum_{i=1}^n \frac{I_1(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2}) m_i}{I_0(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2})} b_i e^{-\alpha b_i}. \end{cases}$$

Notice that these score equations differ from (4) only for the Bessel functions ratios $I_1(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2})/I_0(\frac{m_i \nu_0 e^{-\alpha b_i}}{\sigma^2})$, which multiplies the observations m_i. In particular, this factor decreases the values of observations, since $0 < I_1(x)/I_0(x) < 1$ for $x > 0$, and increases asymptotically to 1 for large SNRs, so that the score equations tend to (4).

As shown in Sijbers et al. (1998), the maximum likelihood estimator for ν obtained from an iid sample $M_1, \ldots, M_n \sim \text{Rice}(\nu, \sigma^2)$ and known σ^2 becomes exactly 0 when the moment estimator for $\mathbb{E}[M^2] = \nu^2 + 2\sigma^2$ becomes inadmissible, i.e. when $\sum_{i=1}^n M_i^2/n - 2\sigma^2 \leq 0$, even if the real value of ν is larger than 0. The case of Rice exponential regression suffers of a similar problem in a non trivial way, and would require σ^2 to be estimated with the other parameters to keep parameter values coherent with the model. Here we will not address this problem, but efforts in this direction are currently in progress.

3.3 Bayesian approaches

We consider also three different estimators based on a Bayesian posterior distribution: its mean, its median and its mode. To allow an easy implementation using BUGS code, we introduce a slightly different formulation of the model. If $M \sim \text{Rice}(\nu, \sigma^2)$, then $R = M^2/\sigma^2$ has noncentral χ^2 distribution with 2 degrees of freedom and noncentrality parameter $\lambda = \nu^2/(2\sigma^2)$. Be now $R_1, \ldots, R_n$ the random sample considered, with $R_i = M_i^2/\sigma^2$ and $M_i \sim \text{Rice}(\nu_0 e^{-\alpha b_i}, \sigma^2)$ for $i = 1, \ldots, n$, and let $\mathbf{r} = (r_1, \ldots, r_n)$ be the observations from this sample. Let $\pi(\nu_0)$ and $\pi(\alpha)$ be the prior distributions of the two unknown parameters, while the density of each R_i will be denoted as $f_{R_i}(r_i)$, with parameter $\lambda_i = \nu_0^2 e^{-2\alpha b_i}/2\sigma^2$. The joint posterior distribution of ν_0 and α is then

$$p(\nu_0, \alpha | \mathbf{r}, \mathbf{b}, \sigma^2) \propto \prod_{i=1}^n f_{R_i}(r_i | \lambda_i) \pi(\nu_0) \pi(\alpha)$$

As anticipated in Section 2.1, a noncentral χ^2 distribution of noncentrality λ can be sampled as a mixture of $\chi^2(2P + 2)$ with $P \sim \mathcal{P}(\lambda)$. This allows an easy BUGS implementation of these estimators.

4 Simulation study

We compared 5 estimators for α - least squares (LS), maximum likelihood (ML) and posterior mean (PMe), median (PMd) and mode (PMo) - in terms of mean and mean square error. For the two frequentist approaches, ranges for the possible parameter values have been chosen, considering $\nu_0 \in [0.1, 10]$ and $\alpha \in [0.1, 5]$, while the fixed parameter σ^2 has been taken always equal to 1. For the Bayesian point estimators we chose uninformative, uniform priors, with the same support as the ranges chosen for LS and ML. The first two estimators have been computed with R 2.12.2 (see R Development Core Team (2009)), using built-in optimization functions: `optimize` for the one-dimensional minimization required in LS and `optim`, using the L-BFGS-B method, for the likelihood maximization, with startup values $(\nu_{0\text{start}}, \alpha_{\text{start}}) = (1, 1)$. Bayesian posterior distributions have been computed using a Gibbs sampler implemented in JAGS (see Plummer (2003)). In particular, the following model code (valid for any program supporting BUGS-type language) was used:

```
model{
  for(i in 1:n){
    lambda[i]<-(nu0*nu0)*exp(-2*alpha*b[i])/(2*sigma*sigma)
    p[i]~dpois(lambda[i])
    k[i]<-2*p[i]+2
    M[i]~dchisqr(k[i])
  }
  alpha~dunif(0.1,5)
  nu0~dunif(0.1,10)
}
```

As it can be seen from the model code, uniform prior distributions have been chosen, with supports equal to the search ranges for LS and ML. 10000 Gibbs sampling iterations have been run for each different sample, with a thinning of 10, and standard diagnostics revealed a good behavior of the generated chains.

We chose b-values in a typical range for diffusion MR machine settings, i.e. from 0 to 1000s/mm^2, on equally spaced grids of $n = 5$, 10, 15, 20, 25, 30 points. Different simulations have been run with parameter values $\nu_0 = 2, 4, 8$, which represent a low, an intermediate and a high SNR, and $\alpha = 0.7, 1, 3$, typical low, intermediate and high physiological values of ADC.

It must be reported that the ML estimator, in cases of low SNR, reached the boundaries of the optimization region in various simulations. In the combination $n = 5$, $\nu_0 = 2$, $\alpha = 3$ only 45% of the simulations gave ML estimates that converged to a value inside the predefined ranges of parameters search, while in the other cases this number oscillated around 70% when $\alpha = 1$ or 100% when $\alpha = 0.7$. These degenerate results have been removed for the computation of bias and variance.

Figure 1 displays the decaying exponential curves we aim to estimate in the 9 different combinations of ν_0 and α, along with a horizontal line at level σ, to

represent the order of magnitude of noise with respect to the signal. The quality of estimates depends both on the SNR at $b = 0$ and on the ADC, as will be clear from simulations.

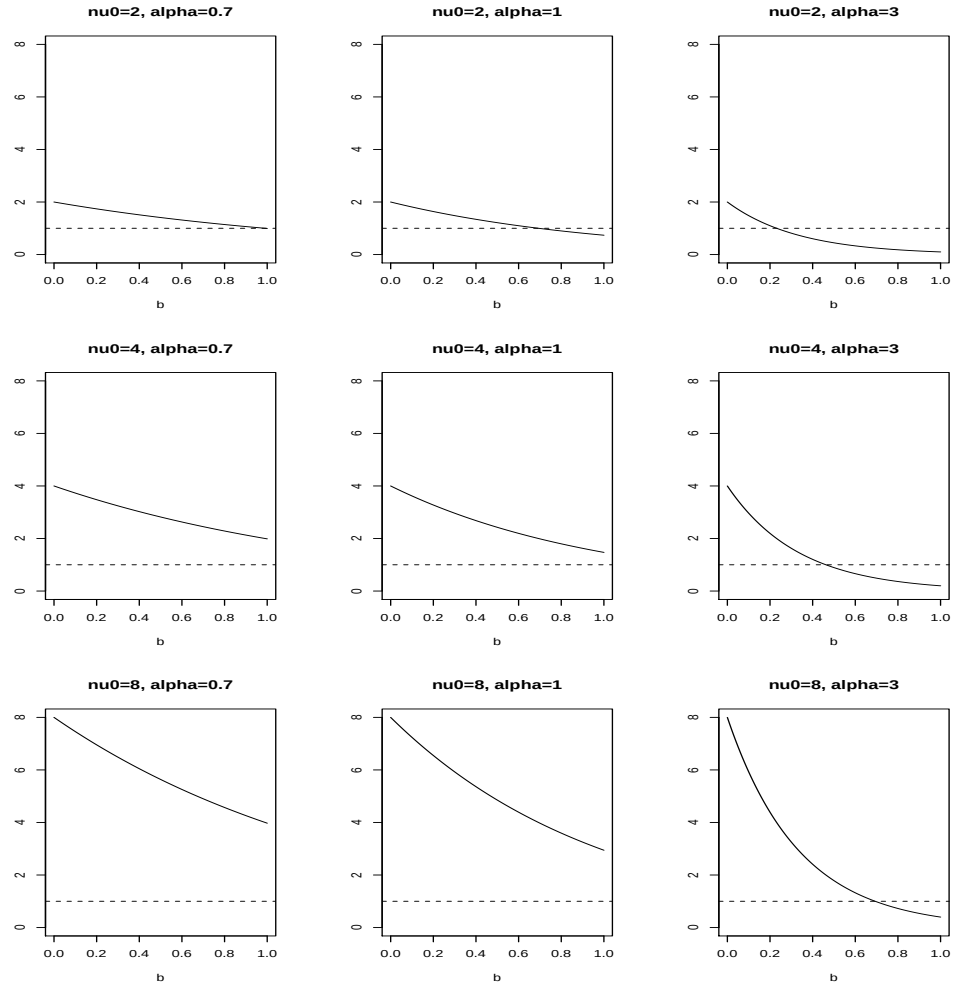

Figure 1: Stejskal-Tanner model in simulation parameter combinations. b-values are expressed in $1000 s/mm^2$.

Figure 2 shows the behavior of bias for the estimators of α with different sample sizes n. For what concerns the frequentist estimators (LS and ML), there is no uniform ordering through the considered values of n when the signal decays slowly ($\alpha = 0.7$), but in the other cases, when noise is stronger along the curve, the maximum likelihood estimate is always less biased than the least squares one; notice also that the least squares estimates do not seem to have a decreasing bias when n increases among the considered values. Concerning the 3 Bayesian estimators, no striking differences arise among them, while with

respect to the frequentist estimators in many cases they have comparable or higher bias, with the exception of the "worst case" $\nu_0 = 2$, $\alpha = 3$, where they are uniformly more accurate.

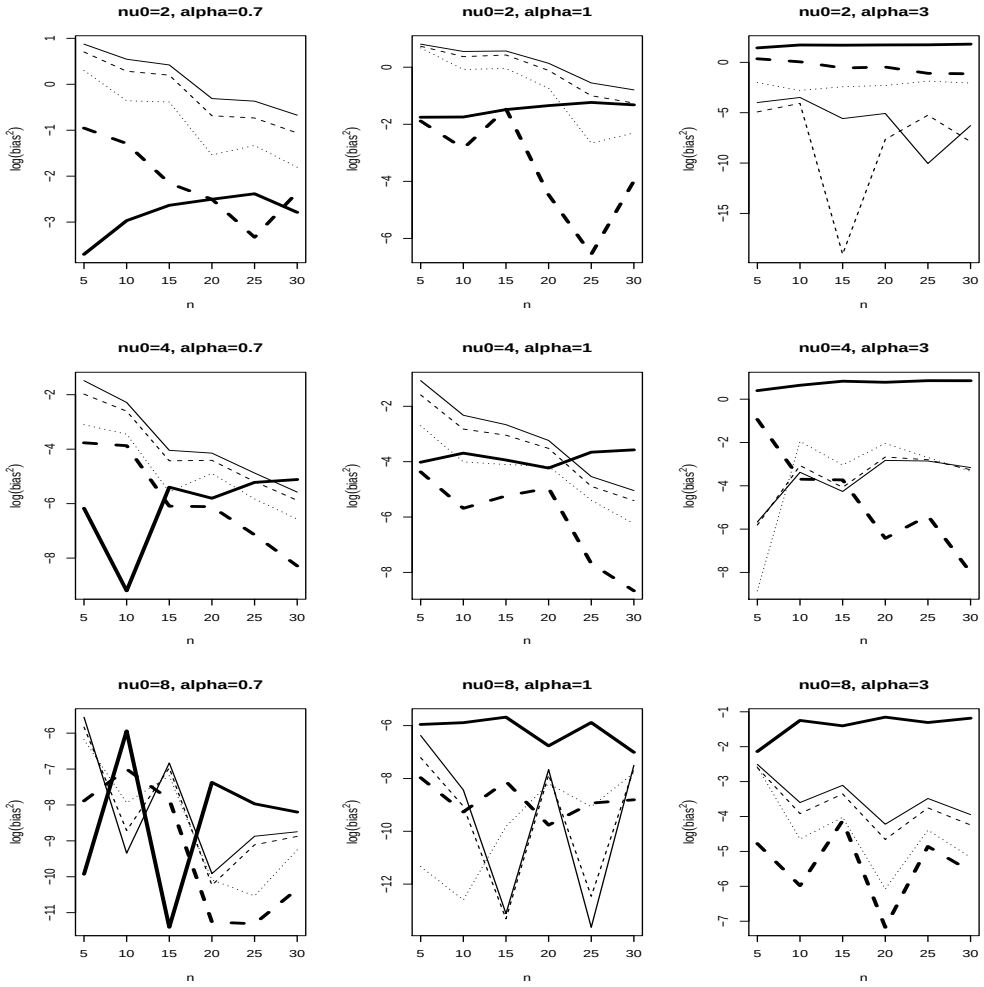

Figure 2: Bias of estimators for α. Bold lines: solid=LS, dashed=ML; slim lines: solid=PMe, dashed=PMd, dotted=PMo.

From what concerns variance, analyzed in Figure 3, the LS estimator shows almost always the best performance, excepted for low sample sizes when $\alpha = 3$. The other estimators have similar performances and behaviors at different sample sizes n, with ML and PMe having strikingly higher variance in some noisy cases. As expected, variance notably decreases for all estimators at increasing n in most combinations of parameters, but with very low SNR ($\nu_0 = 2$) the only one showing empirical convergence of variance to 0 is LS.

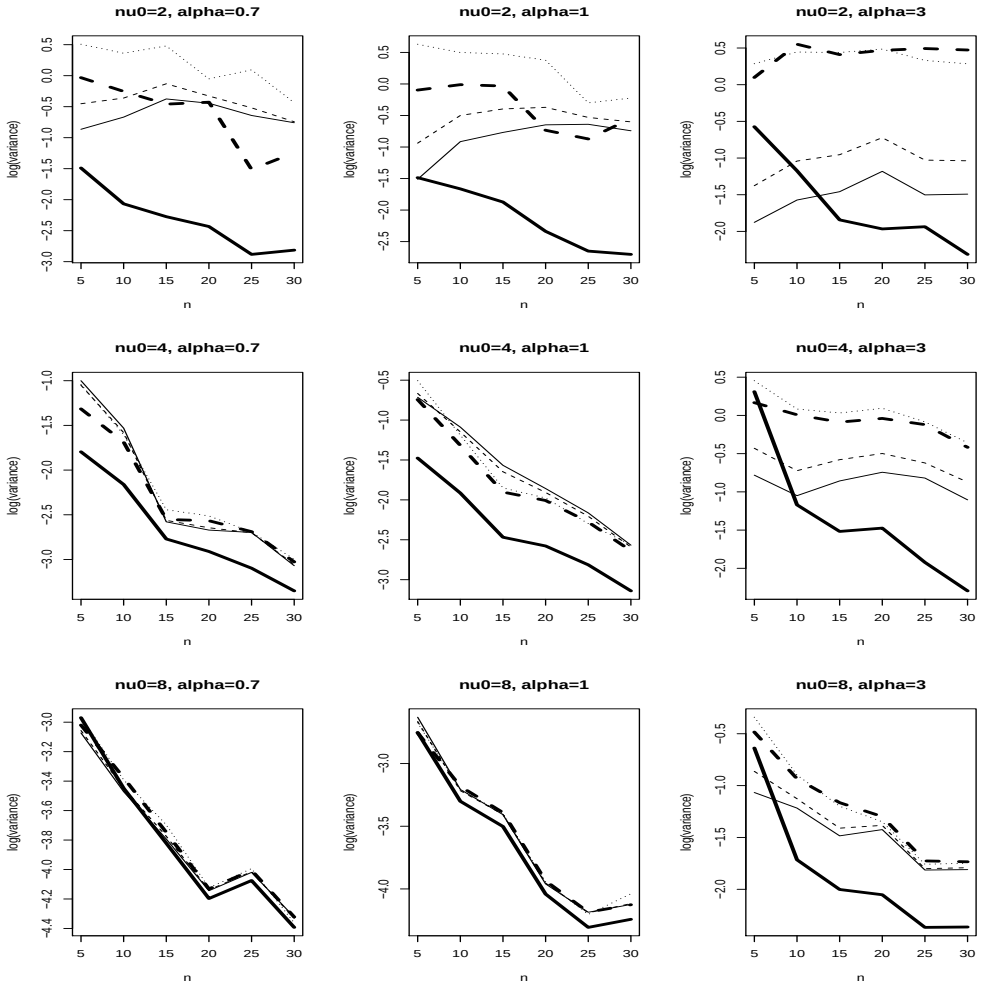

Figure 3: Variance of estimators for α. Bold lines: solid=LS, dashed=ML; slim lines: solid=PMe, dashed=PMd, dotted=PMo.

An overall index of estimator performance can be evaluated by the mean square error (MSE). Since the MSE is the sum of square bias and variance, the orders of magnitude of these two characteristics assume an important role. As it can be seen from Figure 4, the LS estimator has the lowest MSE when $\alpha = 0.7, 1$, but exhibits the worst performances in the critical cases of high ADC, where Bayesian estimators seem to work better.

Results for ν_0 are not detailed here, but it is worth mentioning that, since it is necessary to estimate the two parameters jointly, the precisions and accuracies of their estimators are mutually influenced. Anyway, estimators for ν_0 show a more classical behavior: the LS estimator is in all cases less accurate but more precise (high bias and low variance), and the consistency of all estimators is

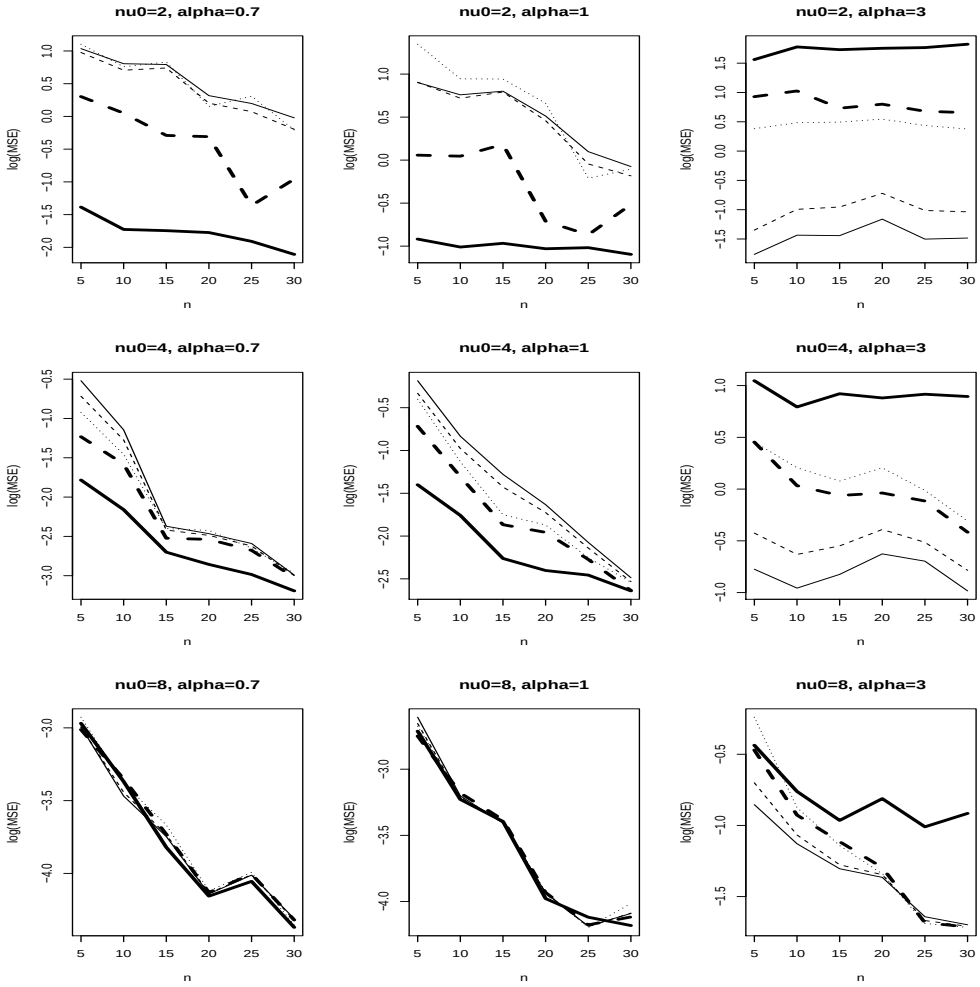

Figure 4: MSE of estimators for α. Bold lines: solid=LS, dashed=ML; slim lines: solid=PMe, dashed=PMd, dotted=PMo.

evident when increasing n. The summary plots for the MSE of the estimators for ν_0 can be seen in Figure 5.

5 Conclusions

In this work, we proposed different methods for estimating pixelwise the ADC from diffusion MR signals, following the Rice noise model and the Stejskal-Tanner equation for magnitude decay. The presented estimators exhibit different features that should be taken into account when approaching real data. The least squares approach is the fastest and has low variance, but becomes less accurate when the conditional signal distribution at different b-values is more

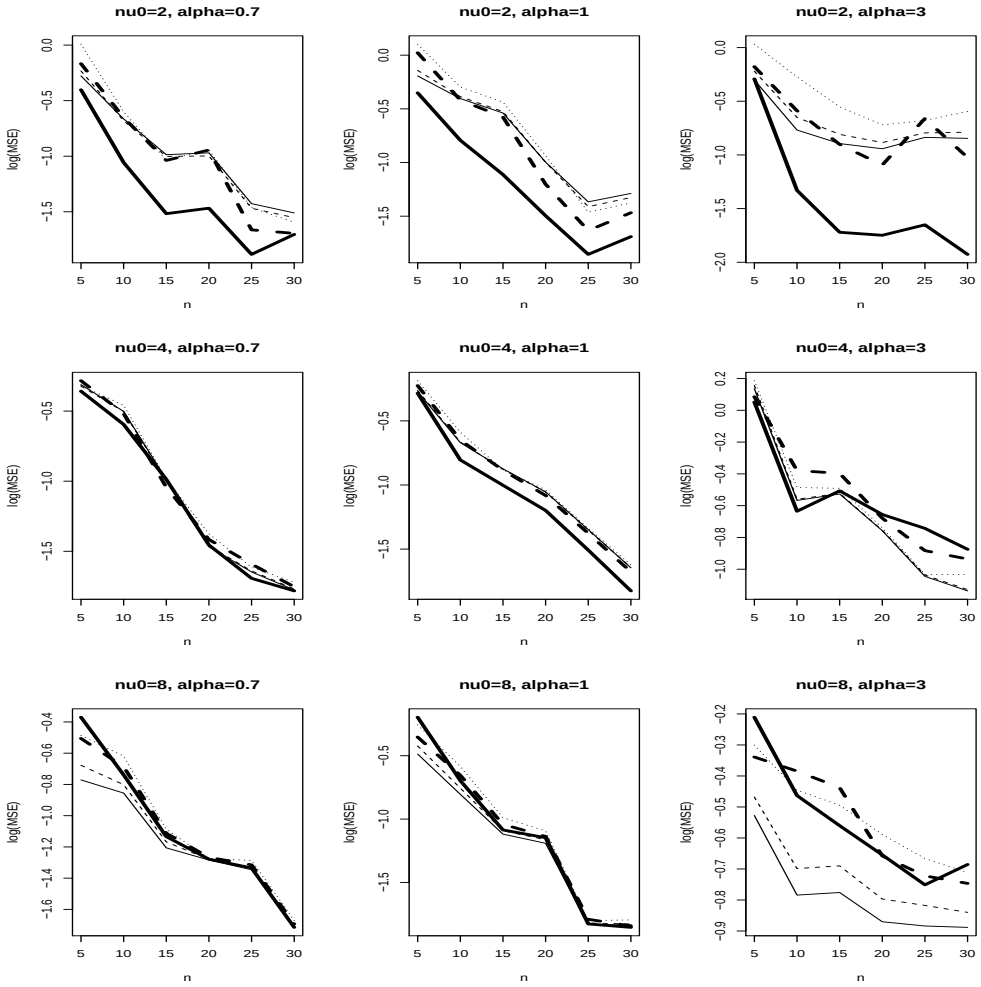

Figure 5: MSE of estimators for ν_0. Bold lines: solid=LS, dashed=ML; slim lines: solid=PMe, dashed=PMd, dotted=PMo.

distant from normality. The maximum likelihood estimator is slightly slower, requiring a nonlinear maximization on 2 variables, and has the lowest bias in many cases, but, as pointed out before, it may diverge with samples from noisy signals. Bayesian estimators are the most expensive in terms of computational costs, and may require further tuning for improving their performances; they are the best in terms of mean square error at the high ADC here tested, and offer the advantage of providing the whole posterior distribution for inferential purposes, while inferential tools regarding LS and ML should rely, at present time, on normal approximations, which may not be reliable with low sample sizes and SNRs. Future studies concentrate on the inferential aspect, while extending in efficient ways these estimation methods to full MR images.

References

Abramowitz, M. and I. A. Stegun (Eds.) (1964). *Handbook of Mathematical Functions*. Dover Publications.

Landini, L., V. Positano, and M. F. Santarelli (Eds.) (2005). *Advanced Image Processing in Magnetic Resonance Imaging*. CRC Press.

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: `http://cran.r-project.org/`.

Sato, C., S. Naganawa, T. Nakamura, H. Kumada, S. Miura, O. Takizawa, and T. Ishigaki (2005). Differentiation of noncancerous tissue and cancer lesions by apparent diffusion coefficient values in transition and peripheral zones of the prostate. *Journal of Magnetic Resonance Imaging 21*(3), 258–62.

Sijbers, J., A. J. den Dekker, P. Scheunders, and D. Van Dyck (1998). Maximum Likelihood estimation of Rician distribution parameters. *IEEE Transactions on Medical Imaging 17*, 357–361.

Walker-Samuel, S., M. Orton, L. D. McPhail, and S. P. Robinson (2009). Robust estimation of the apparent diffusion coefficient (ADC) in heterogeneous solid tumors. *Magnetic Resonance in Medicine 62*(2), 420–429.

Woodhams, R., K. Matsunaga, K. Iwabuchi, S. Kan, H. Hata, M. Kuranami, M. Watanabe, and K. Hayakawa (2005). Diffusion-Weighted Imaging of Malignant Breast Tumors: The Usefulness of Apparent Diffusion Coefficient (ADC) Value and ADC Map for the Detection of Malignant Breast Tumors and Evaluation of Cancer Extension. *Journal of Computer Assisted Tomography 29*(5), 644–649.