Robust estimation of the noise variance from background MR data

J. Sijbers[a], A. J. den Dekker[c], D. Poot[a], R. Bos[c], M. Verhoye[a,b], N. Van Camp[b], A. Van der Linden[b],

[a]Vision Lab, [b]Bio-Imaging Lab, University of Antwerp, Belgium[a,b]
Delft Center for Systems and Control, Delft University of Technology, The Netherlands[c]

ABSTRACT

In the literature, many methods are available for estimation of the variance of the noise in magnetic resonance (MR) images. A commonly used method, based on the maximum of the background mode of the histogram, is revisited and a new, robust, and easy to use method is presented based on maximum likelihood (ML) estimation. Both methods are evaluated in terms of accuracy and precision using simulated MR data. It is shown that the newly proposed method outperforms the commonly used method in terms of mean-squared error (MSE).

Keywords: noise variance, ML estimation, histogram, Rayleigh distribution

1. INTRODUCTION

The noise variance in magnetic resonance (MR) images has always been an important parameter to account for in the processing and analysis of MRI data. Algorithms for segmentation, clustering, restoration, noise reduction etc, highly depend on the noise variance.[1–4] Also, many applications that employ statistical analysis techniques often base their conclusions on assumptions about the underlying noise characteristics. For example, the noise variance is an important parameter in functional MRI analysis, where signal variations of the order of a few percent need to be detected. Indeed, the primary obstacle in generating reliable activation maps with fMRI is the extremely low signal-to-noise ratio of the acquired data. The detection of neural activity is most commonly carried out in the form of a statistical hypothesis test, which often relies on assumptions of the underlying noise distribution. Finally, knowledge of the noise variance is useful in the quality assessment of the MR imaging system itself (e.g., to test the noise characteristics of the receiver coil, preamplifier, etc.).

In the past, many techniques have been proposed to estimate the image noise variance. These can be subdivided into two classes:

single image Commonly, the noise variance is estimated from a single magnitude image, where it is determined directly from a large, uniform signal region or from a non-signal (i.e., noise-only) region.[5–8] Thereby, the estimate is computed from the second moment of a Gauss probability density function (PDF) or from the first moment of the Rayleigh PDF, respectively. A problem with this approach is the selection of homogeneous regions that are sufficiently large to obtain a precise estimate of the noise variance. Moreover, background data points may suffer from systematic intensity variations due to streaking or ghosting artifacts. Since often many background data points are available, the noise variance can be estimated from the maximum of the image grey value histogram, which is more robust against the previous mentioned artifacts. However, as will be shown in this paper, this method yields biased results.

multiple images To cover some disadvantages of single image methods, noise variance estimation methods were developed based on two acquisitions of the same image. A standard procedure was developed by Sano in which the noise variance was estimated by subtracting two acquisitions of the same object and calculating the standard deviation of the resulting pixel values.[9, 10] The multiple acquisition methods have that advantage over the single image methods that they are relatively insensitive to structured noise such as ghosting, ringing, and DC artifacts.[11, 12] However, a strict requirement is the perfect geometrical registration of the images and temporal stationarity of the imaging proces.

Further author information: (Send correspondence to J. Sijbers)
J. Sijbers: E-mail: Jan.Sijbers@ua.ac.be, Telephone: +32 (0)3/265.34.52

In this paper, we will focus on noise variance estimation based on the background mode of the histogram of a single MR image in which the pixel grey values are represented by integers. Thereby, it is assumed that a large number of background data is available, or equivalently, that the histogram has a clear background mode, which is often a valid assumption. In Section 2.1, the paper starts by reviewing the statistics of background MR data. Next, in Section 2.2, we will describe a simple, robust, and commonly used procedure to estimated the noise variance from the histogram's background mode. In the same section, we will present a new noise variance estimation method based on maximum likelihood estimation from a partial histogram. Along with this method, a control procedure is provided that helps in determining whether or not the test will yield proper estimates of the noise variance. Subsequently, in Section 3, simulation experiments to test the performance of the proposed noise variance estimation procedure in terms of precision and accuracy are described and discussed. Finally, in Section 4, conclusions are drawn.

2. METHODS

2.1. Noise properties of MR data

In MRI, the acquired complex data is known to be polluted by white noise which is characterized by a Gaussian PDF. After inverse Fourier transformation, the real and imaginary data is still corrupted with Gaussian noise because of the orthogonality of the Fourier transform. However, it is common practice to transform the complex valued images into magnitude and phase images. Since computation of a magnitude (or phase) image is a non-linear operation, the PDF of the data under concern changes. It is well known that the data in a magnitude image is no longer Gaussian but Rician distributed[5, 13]:

$$p\left(m|A,\sigma\right) = \frac{m}{\sigma^2} e^{-\frac{m^2+A^2}{2\sigma^2}} I_0\left(\frac{Am}{\sigma^2}\right) \epsilon(m) \quad , \tag{1}$$

with I_0 denoting the 0th order modified Bessel function of the first kind, A the noiseless signal level, σ^2 the noise variance, and m the MR magnitude variable. The unit step Heaviside function $\epsilon(\cdot)$ is used to indicate that the expression for the PDF of m is valid for non-negative values of m only.

For low signal-to-noise ratio (SNR), where SNR is defined as A/σ, the modified Bessel function is given by:

$$I_\nu(z) \sim \left(\frac{z}{2}\right)^\nu \Gamma(\nu+1) \quad \text{for} \quad z \to 0 \quad . \tag{2}$$

Hence, it is easy to show that, when the SNR is zero, the Rician PDF leads to a Rayleigh PDF:

$$p\left(m|\sigma\right) = \frac{m}{\sigma^2} e^{-\frac{m^2}{2\sigma^2}} \epsilon(m) \quad \text{for} \quad \text{SNR} \to 0 \quad . \tag{3}$$

The Rayleigh PDF characterizes the random intensity distribution of non-signal background areas.

2.2. Noise estimation

In many MR applications, a large number of background data (where the true MR signal is assumed to be zero) is available to estimate the noise variance. Commonly, a region of interest (ROI) is selected from the background data from which σ^2 is estimated using (3).[5] However, such a procedure is not recommended because the ROI has to be manually selected and is highly sensitive to errors such as ghosting artifacts. Moreover, since only a ROI is exploited for estimation, the precision of the estimated parameter is relatively low. Instead, the image noise variance can be more robustly estimated from the image histogram. Indeed, since generally a large number of background data points are available, a histogram often shows a mode that is clearly distinguishable (see for example Fig. 1).

2.2.1. Maximum of the background mode of the histogram

From the Rayleigh PDF, given in (3), the noise variance can be estimated by searching for the value of m for which the Rayleigh PDF attains a maximum:

$$\frac{\partial p}{\partial m} = 0 \qquad \Leftrightarrow \qquad 1 - \frac{m^2}{\sigma^2} = 0 \tag{4}$$

From this, it is clear that an estimate of the noise standard deviation is simply given by:

$$\widehat{\sigma} = m_{\max} \tag{5}$$

In practice, $m_{\max}$ can be found by searching for the magnitude value at which the background mode in the histogram attains a maximum. Since the background mode is always located at the left of the histogram, finding this maximum is trivial.

2.2.2. ML estimation from the background mode of the histogram

Alternatively, σ can be estimated by ML estimation using a partial histogram (i.e., the left side of the histogram). If $\{l_i\}$ with $i = 0, ..., K$ denotes the set of boundaries of histogram bins, and n_i represents the number of observations (counts) within the bin $[l_{i-1}, l_i]$, then the joint PDF of the corresponding data is given by[14]:

$$p(\{n_i\}|\sigma, \{l_i\}) = \frac{N_K!}{\prod_{i=1}^{K} n_i!} \prod_{i=1}^{K} p_i^{n_i}(\sigma) \tag{6}$$

with $N_K = \sum_{i=1}^{K} n_i$ the total number of observations within the partial histogram and

$$p_i(\sigma) = \frac{\int_{l_{i-1}}^{l_i} \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2\sigma^2}\right) dm}{\sum_{i=1}^{K} \int_{l_{i-1}}^{l_i} \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2\sigma^2}\right) dm} \tag{7}$$

Since

$$\int_{a}^{b} \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2\sigma^2}\right) dm = e^{-\frac{a^2}{2\sigma^2}} - e^{-\frac{b^2}{2\sigma^2}} \tag{8}$$

it is easy to show that

$$p_i(\sigma) = \left(e^{-\frac{l_{i-1}^2}{2\sigma^2}} - e^{-\frac{l_i^2}{2\sigma^2}}\right) \left(e^{-\frac{l_0^2}{2\sigma^2}} - e^{-\frac{l_K^2}{2\sigma^2}}\right)^{-1} \tag{9}$$

If the set of observations $\{n_i\}$ is fixed and σ is regarded as a variable, the joint PDF given in (6) is called a likelihood function. The ML estimate is then found by maximizing this likelihood function L with respect to σ:

$$\widehat{\sigma}_{ML,K} = \arg\max_{\sigma} L(\sigma|\{n_i\}, \{l_i\}) \tag{10}$$

Equivalently, the optimal value of σ is found by finding the minimum of $-\ln L$, i.e.:

$$\widehat{\sigma}_{ML,K} = \arg\min_{\sigma} \left[N_K \ln\left(e^{-\frac{l_0^2}{2\sigma^2}} - e^{-\frac{l_K^2}{2\sigma^2}}\right) - \sum_{i=1}^{K} n_i \ln\left(e^{-\frac{l_{i-1}^2}{2\sigma^2}} - e^{-\frac{l_i^2}{2\sigma^2}}\right) \right] \quad . \tag{11}$$

Eq. (11) is the ML estimate of the noise standard deviation from K bins.

2.2.3. Bin selection procedure

ML estimation of σ by fitting a (discrete) Rayleigh PDF to the partial histogram implies that we have to select the number of bins we will take into account during estimation. This number, denoted as K, is found by testing the null hypothesis that the bins less than l_K are Rayleigh distributed using a chi-square test. Since the background mode is located on the left side of the histogram, only the first K bins are used to estimated σ. In order to determine how many bins should be taken into account (i.e., how large K is), for each value of K a χ^2 value λ_K is computed. Thereby, λ_K is computed from the expected number of bin counts f_i (obtained from the theoretical Rayleigh distribution and the estimated noise variance) and the actually observed number of counts n_i for each bin:

$$\lambda_K = \sum_{i=1}^{K} \frac{(f_i - n_i)^2}{f_i^2} \tag{12}$$

where

$$f_i = p_i(\widehat{\sigma}) N_K \tag{13}$$

denotes the expected number of counts ($\widehat{\sigma} \equiv \widehat{\sigma}_{ML,K}$).

The statistical distribution of the random variable λ_K can be analyzed as follows. In theory, the number of counts n_i are multinomially distributed, and the marginal distribution of the number of counts in each bin is a binomial distribution with parameters N_K and p_i. This means that the expected value of n_i is $p_i(\widehat{\sigma}) N_K$ and its variance is $p_i(\widehat{\sigma})(1 - p_i(\widehat{\sigma})) N_K$. However, since in general, N_K is large (and p_i is small), the binomial distribution may be approximated by a normal distribution, with expectation value $p_i(\widehat{\sigma}) N_K$ and variance is $p_i(\widehat{\sigma}) N_K$. Therefore, λ_K is approximately χ^2_{K-2} distributed (i.e., is chi-squared distributed with $K - 2$ degrees of freedom). For each λ_K, a probability value $p_K = p(\chi^2_{K-2} > \lambda_K)$ is computed. The best fit is expected to have the highest probability.

At first sight, it seems sensible to select that value of K that gives the highest probability value (p-value) p_K. This approach has been tested by means of simulations. Simulation results revealed that the ML noise variance estimation could be made more robust by refining the bin selection procedure by three (heuristic) rules:

1. During simulation experiments, it has been observed that the bin selection procedure often favored a number of bins that was very small. Indeed, for such a small number of bins, the fit of the Rayleigh PDF to those bins may result in a high p-value. However, in that case, the estimated value for σ was often clearly too large (judging from the counts in bins that were not used for the estimation and the expected number of counts in those bins).

   Therefore, a second χ^2 value is computed, but now on m bins that were *not* used for the estimation of σ and of which the predicted counts were larger than the observed counts. The reasoning behind this is that the Rayleigh PDF composed from $\widehat{\sigma}_{ML,K}$ should never exceed the actual number of counts in the histogram bins (apart from statistical fluctuations) since histogram contributions from the underlying, noiseless signal can only *increase* the number of counts in the bins.

   Under the hypothesis that the data are perfectly Rayleigh distributed and the normal approximation for the number of counts within each bin holds, this second $\lambda_{2,K}$ value is χ^2 distributed with m degrees of freedom:

$$\lambda_{2,K} = \sum_{i>K} \frac{[\max(0, p_i(\widehat{\sigma}) N_K - n_i)]^2}{p_i(\widehat{\sigma}) N_K} \tag{14}$$

   From $\lambda_{2,K}$, a second probability $p_{2,K}(\chi^2_m > \lambda_{2,K})$ is computed. The resulting probabilities are multiplied and the number of bins with the maximum probability is selected as providing the best estimate of σ.

2. During specific simulation experiments it has been observed that, occasionally, an estimate of σ was selected that was clearly too small. This occurred when K was very small. This class of errors is removed by comparing the estimated σ with the maximum of the histogram. As will be shown in the results section, the maximum can be found in a robust way, but it has a high variance. When $\widehat{\sigma}_{ML,K}$ is less than 70%

of the position of the maximum of the histogram, a penalty factor is introduced, so the elements with an occasionally good fit on the first bins are filtered out:

$$p_{4,K} = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(10\frac{\widehat{\sigma}_{ML,K}}{\widehat{\sigma}_{\max}} - 5\right) \tag{15}$$

3. Finally, When σ estimates are large compared to the last bin used, the estimate is unreliable, even when the estimated value matches the distribution accurately. To make sure that a sufficient number of bins is selected for a reliable estimate, a penalty is introduced when the estimated σ is large compared to the last bin:

$$p_{3,K} = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(10\frac{\widehat{\sigma}}{l_K} - 9\right) \tag{16}$$

The resulting probabilities are multiplied and the number of bins with the maximum probability is selected as providing the best estimate of σ.

3. EXPERIMENTS AND DISCUSSION

Simulation experiments were set up to test the performance of both methods discussed in section 2.2 in terms of precision (i.e., variance), accuracy (i.e., bias), and root mean-squared-error (RMSE). The first estimator is the one that takes the maximum of the histogram as estimate for σ, as explained in subsection 2.2.1. The second estimator is the ML estimator as outlined in subsection 2.2.2.

3.1. Simulated noise-only images

First, the performance of the estimators was compared using simulated Rayleigh data. Thereby, the parameter σ was fixed to 30. In all simulations, the histogram was composed from the integer values of the simulated real-valued data where the width of each bin was fixed to 1.

In Fig. 2, the performance of the σ estimators as a function of K is shown in terms of precision, accuracy, and RMSE. Each point of the curves was generated by averaging 1500 estimations.

precision Fig. 2(a) shows the standard deviation of both estimators as a function of the number of bins used in this process (K). We recall that the true value for σ was 30 and that the bin width equals 1. Hence, from the figure it is clear that the standard deviation of the maximum estimator increases until the number of bins is 30 after which it remains constant with increasing number of bins used during estimation. On the other hand, the standard deviation of the ML estimator decreases fast with the number of bins. However, one can also note that if the number of bins is small (generally smaller than the true sigma), the standard deviation of the ML estimator increases drastically. Although it depends slightly on the number of voxels of which the histogram is composed, the minimum number of bins necessary for a valid estimation (in terms of precision) is about the true σ.

accuracy Fig. 2(a) shows the expectation values of both estimators as a function of K. As can be observed from the figure, both estimators perform more or less the same in terms of accuracy is K is sufficiently large. If K is less than the true σ, both estimators show a significant bias (though for the ML estimator the bias is significant in a smaller interval than the maximum estimator).

RMSE Finally, the RMSE is shown in Fig. 3(c). From the figure, it is clear that in terms of the RMSE, the ML estimator performs generally better than the maximum estimator. However, also in this case, it is clear that the number of bins taken into account for both estimators should not be significantly smaller than the true σ.

3.2. MR image with noise

Next, a simulated noise free MRI image was obtained from[15] with the following simulation parameters: normal brain, T1, 3mm, 0% noise, 20% intensity non-uniformity, TR = 18 ms, TE = 10 ms using a FLASH imaging sequence (flip angle = 30°). This simulated noise-free magnitude MR image was corrupted with Rician distributed noise with $\sigma = 30$. The resulting real-valued image was rounded to integer values and the noise variance was estimated with the estimators described above. The results are presented in Fig. 3.

precision By visual inspection of Fig. 3(a) it is clear that the standard deviation of both estimators as a function of the number of bins is similar to the noise-only images. Hence, also in this case, the ML estimator seems to outperform the maximum estimator with respect to precision, at least if the number of bins is sufficiently large.

accuracy Fig. 2(a) shows the expectation values of both estimators as a function of K. In contrast to the noise-only case, the bias of the ML estimator now increases with increasing K. This is because when K increases, the chance that the histogram bins also contains significant contributions from the noiseless underlying signal becomes more and more likely. Obviously, as can be seen from the figure, the maximum estimator is not sensitive to this phenomenon (at least as long as the background histogram mode is much more pronounced than the signal mode). Whether or not significant signal contributions are present in the background mode can be tested as was outlined in subsection 2.2.3.

RMSE The RMSE as a function of K is shown in Fig. 3(c).

Finally, the RMSE as a function of σ was computed for both estimators. The results are shown in Fig. 4. Each point was generated by averaging 250 estimations. From the figure, it is clear that in terms of the RMSE the proposed ML estimator has an overall better performance than the maximum estimator.

4. CONCLUSIONS

In this paper, a comparison of two noise variance estimation methods that employ the image histogram was given. Simulation experiments showed that the Maximum Likelihood (ML) estimator outperformes the estimator that uses the maximum of the histogram. However, selecting how many bins are used for the ML estimator is essential when low intensity signal contributions, for example caused by ghosting, are present in the MRI image. A selection procedure has been devised that automatically selects the number of bins to be used for estimation.

ACKNOWLEDGMENTS

REFERENCES

1. G. Z. Yang, P. Burger, D. N. Firmin, and S. R. Underwood, "Structure adaptive anisotropic filtering for magnetic resonance image enhancement," in *Proceedings of CAIP: Computer Analysis of Images and Patterns*, pp. 384–391, 1995.
2. G. Gerig, O. Kubler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Transactions on Medical Imaging* **11**(2), pp. 221–232, 1992.
3. S. J. Garnier and G. L. Bilbro, "Magnetic resonance image restoration," *Journal of Mathematical Imaging and Vision* **5**, pp. 7–19, 1995.
4. M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE signal processing magazine* **3**, pp. 24–41, 1997.
5. R. M. Henkelman, "Measurement of signal intensities in the presence of noise in MR images," *Medical Physics* **12**(2), pp. 232–233, 1985.
6. L. Kaufman, D. M. Kramer, L. E. Crooks, and D. A. Ortendahl, "Measuring signal-to-noise ratios in MR imaging," *Radiology* **173**, pp. 265–267, 1989.

7. J. P. De Wilde, J. A. Hunt, and K. Straughan, "Information in magnetic resonance images: evaluation of signal, noise and contrast," *Medical and Biological Engineering and Computing* **35**, pp. 259–265, 1997.

8. A. J. den Dekker and J. Sijbers, *Advanced Image Processing in Magnetic Resonance Imaging*, vol. 27 of *Signal Processing and Communications*, ch. 4: Estimation of signal and noise from MR data, pp. 85–143. CRC press, October 2005. ISBN: 0824725425.

9. R. M. Sano, *MRI: Acceptance Testing and Quality Control - The Role of the Clinical Medical Physicist*, Medical Physics Publishing Corporation, Madison, Wisconsin, April 1988.

10. B. W. Murphy, P. L. Carson, J. H. Ellis, Y. T. Zhang, R. J. Hyde, and T. L. Chenevert, "Signal-to-noise measures for magnetic resonance imagers," *Magnetic Resonance Imaging* **11**, pp. 425–428, 1993.

11. J. Sijbers, P. Scheunders, N. Bonnet, D. Van Dyck, and E. Raman, "Quantification and improvement of the signal-to-noise ratio in a magnetic resonance image acquisition procedure," *Magnetic Resonance Imaging* **14**(10), pp. 1157–1163, 1996.

12. J. Sijbers, A. J. den Dekker, M. Verhoye, J. Van Audekerke, and D. Van Dyck, "Estimation of noise from magnitude MR images," *Magnetic Resonance Imaging* **16**(1), pp. 87–90, 1998.

13. H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magnetic Resonance in Medicine* **34**, pp. 910–914, 1995.

14. A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, Tokyo, 3rd ed., 1974.

15. C. A. Cocosco, V. Kollokian, R.-S. Kwan, and A. C. Evans, "Brainweb: Online interface to a 3D MRI simulated brain database; http://www.bic.mni.mcgill.ca/brainweb/," *NeuroImage* **5**(4), p. S425, 1997.
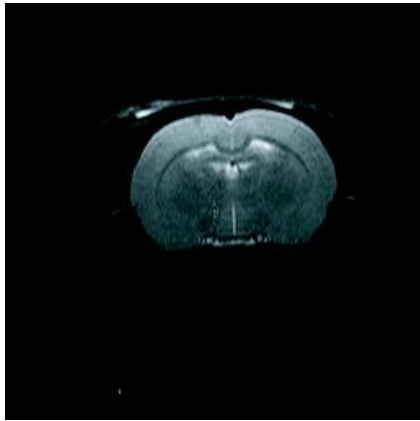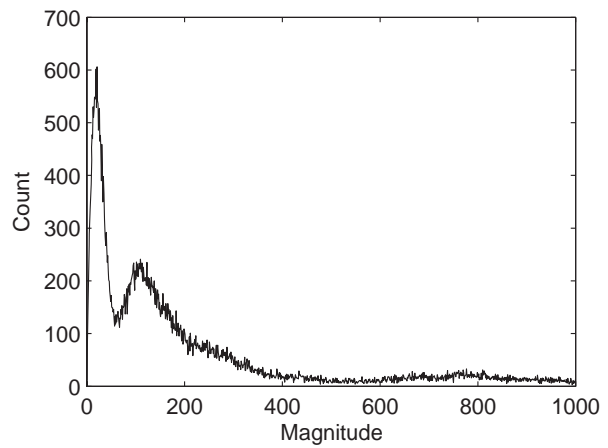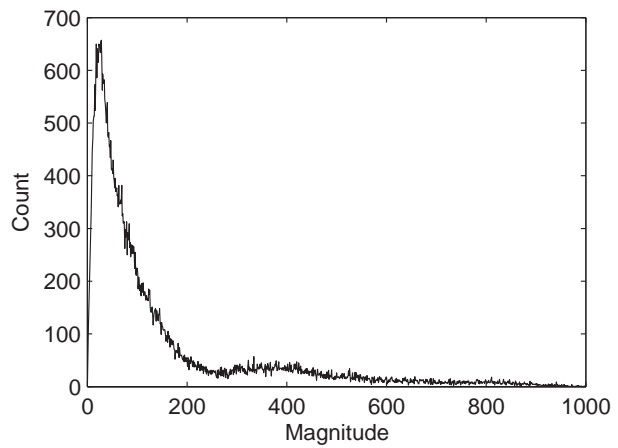
(a) T$_2$ weighted
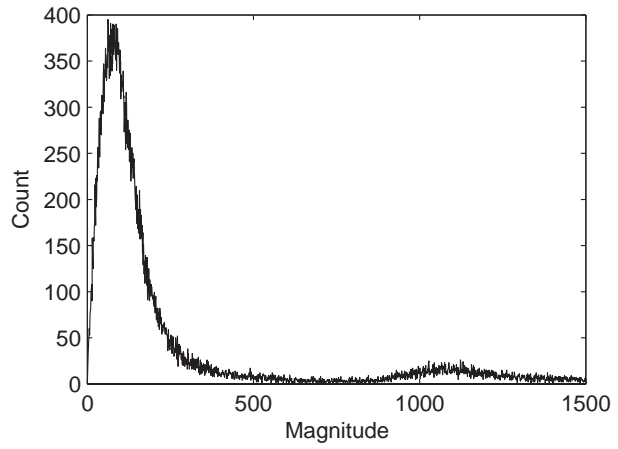
(b) Histogram
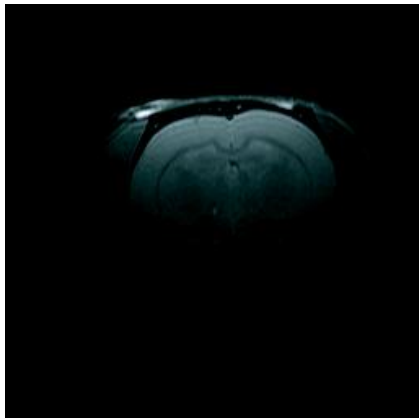
(c) Proton density

(d) Histogram

(e) T$_1$ weighted

(f) Histogram

Figure 1. Histogram of a 2D MR image of the mouse brain: (a-b) T$_2$-, (c-d) proton density, (e-f) T$_1$-weighted image with corresponding histogram.
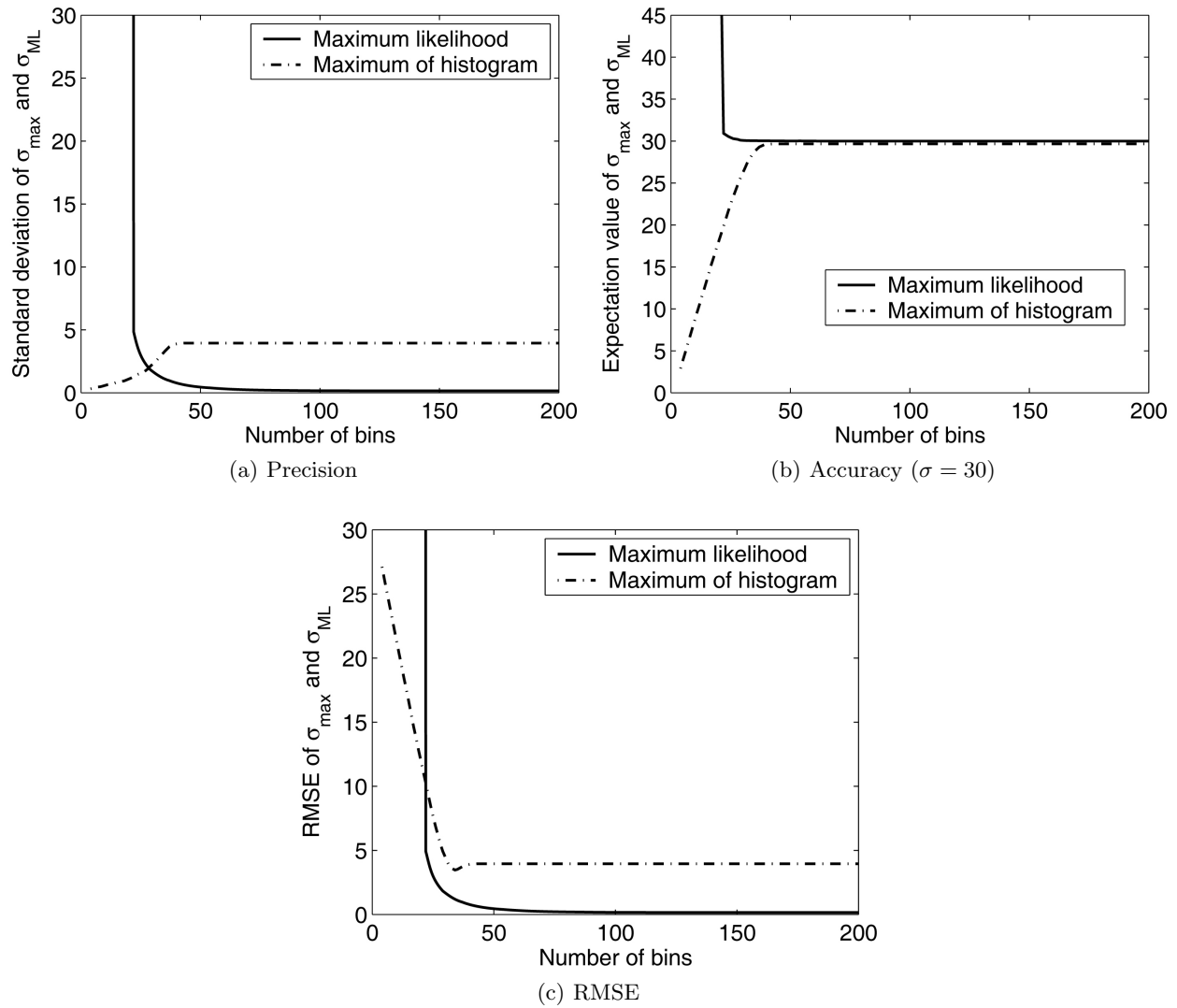
(a) Precision

(b) Accuracy ($\sigma = 30$)

(c) RMSE

Figure 2. Performance comparison of the Maximum and the ML estimator of σ in terms of precision (a), accuracy (b), and RMSE (c) simulated noise only data.
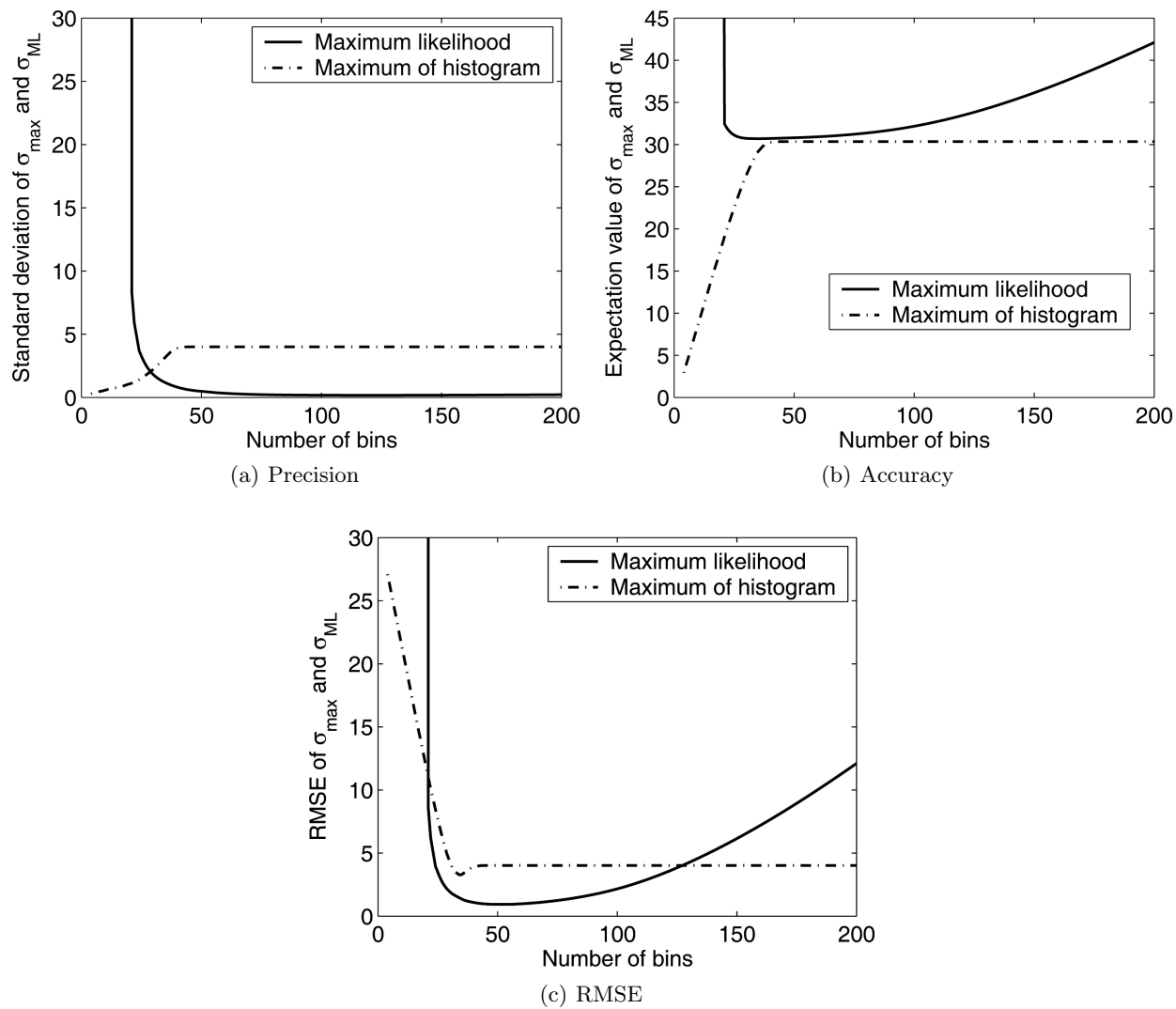
(a) Precision

(b) Accuracy

(c) RMSE

Figure 3. Performance comparison of the Maximum and the ML estimator of σ in terms of precision (a), accuracy (b), and RMSE (c) for simulated MR data.
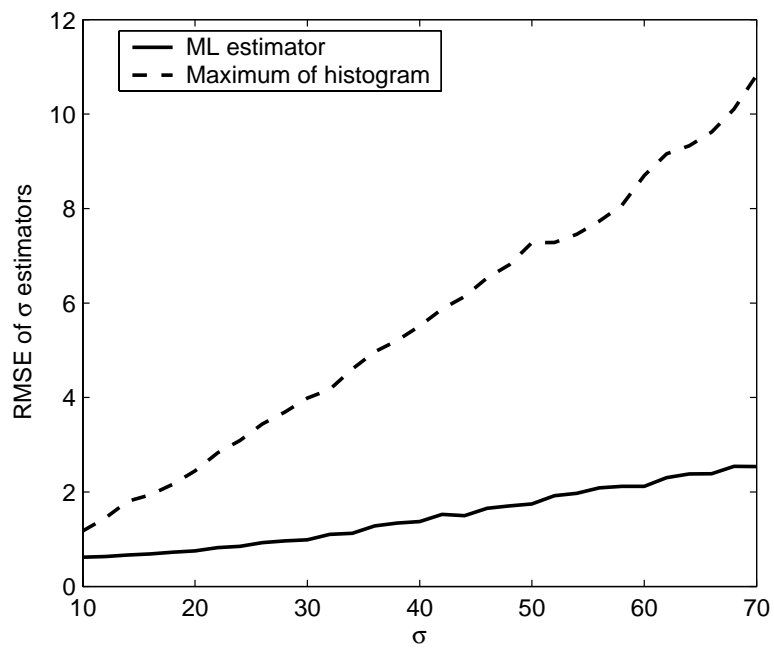
Figure 4. Performance comparison of the Maximum and the ML estimator of σ in terms of the RMSE as a function of σ for simulated MR data. For each point, 250 simulations were used.