Stat
The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.100X/sta.0000

Ricean over Gaussian modeling in magnitude fMRI Analysis – Added Complexity with Negligible Practical Benefits

Daniel W. Adrian[a], Ranjan Maitra[b*], Daniel B. Rowe[c]

Received 28 September 2013; Accepted 16 October 2013

It is well-known that Gaussian modeling of functional Magnetic Resonance Imaging (fMRI) magnitude time-course data, which are truly Rice-distributed, constitutes an approximation, especially at low signal-to-noise ratios (SNRs). Based on this fact, previous work has argued that Rice-based activation tests show superior performance over their Gaussian-based counterparts at low SNRs and should be preferred in spite of the attendant additional computational and estimation burden. Here, we revisit these past studies and after identifying and removing their underlying limiting assumptions and approximations, provide a more comprehensive comparison. Our experimental evaluations using ROC curve methodology show that tests derived using Ricean modeling are substantially superior over the Gaussian-based activation tests only for SNRs below 0.6, i.e SNR values far lower than those encountered in fMRI as currently practiced.

Keywords: EM algorithm; fMRI; Likelihood Ratio Test; Maximum likelihood estimate; Newton-Raphson; Rice distribution; ROC curve; signal-to-noise ratio

1. Introduction

Over the past two decades, functional Magnetic Resonance Imaging (fMRI) has developed into a popular method for noninvasively studying the spatial characteristics and extent of human brain function. The imaging modality depends on the fact that when neurons fire in response to a stimulus or task, the blood oxygen levels in neighboring vessels change, affecting the magnetic resonance (MR) signal on the order of 2-3% (Lazar, 2008), due to the differing magnetic susceptibilities of oxygenated and deoxygenated hemoglobin. This difference causes the so-called Blood Oxygen Level Dependent (BOLD) contrast (Ogawa et al., 1990; Belliveau et al., 1991; Kwong et al., 1992), which is used as a surrogate for neural activity. Datasets collected in an fMRI study are temporal sequences of three-dimensional

[a] Research and Development Division, National Agricultural Statistics Service, Fairfax, VA, USA
[b] Department of Statistics and Statistical Laboratory, Iowa State University, Ames, Iowa, USA
[c] Department of Mathematics, Statistics and Computer Science at Marquette University, Milwaukee, Wisconsin, USA.
*Email: maitra@iastate.edu

images, in which the time course is in accordance with the presentation of a stimulus. Such images are composed of MR measurements at each voxel – or volume element – and have the same distributional and noise properties as any signal acquired using MR imaging.

A general approach for detecting regions of neural activation is to fit, at each voxel, a model — commonly a general linear model (Friston et al., 1995) — to the time course observation sequence against the expected BOLD response. This provides the setting for the application of techniques such as Statistical Parametric Mapping (SPM) (Friston et al., 1990), where the time series at each voxel is reduced to a test statistic which summarizes the association between each voxel time course and the expected BOLD response (Bandettini et al., 1993). The resulting map is then thresholded to identify voxels that are significantly activated (Worsley et al., 1996; Genovese et al., 2002; Logan & Rowe, 2004).

Most statistical analyses focus on magnitude data computed from the complex-valued measurements resulting from Fourier reconstruction (Kumar et al., 1975; Jezzard & Clare, 2001). These raw real and imaginary measurements are well-modeled as two independent normal random variables with the same variance (Wang & Lei, 1994), so the magnitude measurements follow the Rice distribution (Rice, 1944; Gudbjartsson & Patz, 1995). In recent years, there has been considerable effort in the MR community to use the Rice distribution to better understand the noise characteristics of the MR signal (Sijbers et al., 2007; Aja-Fernández et al., 2009; Maitra & Faden, 2009; Rajan et al., 2010; Maitra, 2013) and to use it to improve image restoration and reconstruction (*eg.* synthetic MRI, Maitra & Riddles, 2010). In the context of fMRI also, most standard analyses have assumed that magnitude data are Gaussian-distributed, an assumption which is only valid at high signal-to-noise ratio (SNR). This fact is increasingly important because the SNR is proportional to voxel volume (Lazar, 2008); thus an increase in the fMRI spatial resolution will correspond to a lowering of the SNR, making the Gaussian distributional approximation for the magnitude data less tenable.

Following this justification, previous work has demonstrated disadvantages of Gaussian-based modeling for simulated low-SNR, Rice-distributed time course sequences. For instance, Solo & Noh (2007) report that Gaussian-model-based maximum likelihood estimates (MLEs) of Ricean parameters are increasingly biased as the SNR decreases. Further, den Dekker & Sijbers (2005) present a Ricean-based likelihood ratio test (LRT) for activation with higher detection rate than a Gaussian-based LRT at low SNRs, and the difference in detection rates increases with decreasing SNR. Further, the paper argues that the Gaussian-based LRT "should never be used" for fMRI time series with SNRs below 10 because its false detection rate is non-constant as a function of SNR. In a similar result, Rowe (2005) derives a Ricean-approximated-based LRT statistic which has higher mean values than its Gaussian counterpart. More recently, Noh & Solo (2011) have shown that while the asymptotic power function of the Gaussian-based LRT depends on activation-to-noise ratio but not SNR, the corresponding Ricean power function appropriately depends on both.

In this paper, we argue however that the studies reported in both den Dekker & Sijbers (2005) and Rowe (2005), which provide influential evidence in favor of Ricean modeling of fMRI data, make assumptions and approximations which put their results into question. For one, den Dekker & Sijbers (2005) assumes that the noise variance is known and constant across all voxels when, typically, it is estimated separately for each voxel time series (Friston et al., 1995). Additionally, Rowe (2005) relies on a Taylor-series-based approximation of the Rice distribution, which we argue does not use the exact Rice distribution and does not yield optimal tests. We note that the assumptions of den Dekker & Sijbers (2005) or of Rowe (2005) are not needed when the Expectation Maximization (EM) algorithm (Dempster et al., 1977) is applied to the ML estimation of Ricean parameters (Solo & Noh, 2007; Zhu et al., 2009), which we make practical through the incorporation of Newton-Raphson (NR) steps into the EM calculations. However, a study comparing Ricean-based LRTs computed by this EM scheme to Gaussian-based LRTs is missing from the literature.

In this paper, we develop and report results on an updated and thorough simulation study comparing Ricean- and Gaussian-model-based LRTs for activation in low-SNR magnitude fMRI data, using testing schemes that rely on the assumptions (den Dekker & Sijbers, 2005; Rowe, 2005) discussed above as well as those that do not make these assumptions. Competing LRTs in these two sets of scenarios are described in Section 2, where we also discuss methods that can more effectively evaluate their performance. We analyze a real fMRI dataset in Section 3 to provide motivation and context behind our investigations. Section 4 presents the simulation study, and evaluates and discusses the results. We conclude in Section 5 with some concluding remarks on the implications of the findings in this paper on current fMRI practice.

2. Methodological Development

We focus on an individual (voxel-wise) time-course sequence of magnitude measurements at a voxel, which we denote by $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)$, with n being the number of scans. (In this paper, we denote scalar quantities using regular mathematical fonts; vectors and matrices are boldfaced.) As discussed in Section 1, each measurement is computed as the magnitude $r_t = \sqrt{y_{\mathrm{Re},t}^2 + y_{\mathrm{Im},t}^2}$, $t = 1, 2, \ldots, n$, of the real and imaginary measurements $y_{\mathrm{Re},t}$ and $y_{\mathrm{Im},t}$, respectively. Upon extending findings in Wang & Lei (1994) and Sijbers (1998), it is easy to see that these complex-valued measurements are well-modeled as $y_{\mathrm{Re},t} = \boldsymbol{x}_t'\boldsymbol{\beta}\cos\theta_t + \eta_{\mathrm{Re},t}$ and $y_{\mathrm{Im},t} = \boldsymbol{x}_t'\boldsymbol{\beta}\sin\theta_t + \eta_{\mathrm{Im},t}$, where $\boldsymbol{x}_t'$ is the tth row, $t = 1, 2, \ldots, n$, of an $n \times q$ design matrix $\boldsymbol{X}$, θ_t is the phase imperfection, and $\eta_{\mathrm{Re},t}, \eta_{\mathrm{Im},t} \sim$ iid $N(0, \sigma^2)$ random variables. The Ricean probability density function (PDF) of r_t results from transforming the PDF of $(y_{\mathrm{Re},t}, y_{\mathrm{Im},t})$ to the magnitude-phase variables (r_t, ϕ_t), where $\phi_t = \arctan(y_{\mathrm{Im},t}/y_{\mathrm{Re},t})$, and "integrating out" ϕ_t, which takes the form

$$f(r_t|\boldsymbol{\beta}, \sigma^2) = \frac{r_t}{\sigma^2}\exp\left\{-\frac{r_t^2 + (\boldsymbol{x}_t'\boldsymbol{\beta})^2}{2\sigma^2}\right\}\int_{-\pi}^{\pi}\frac{1}{2\pi}\exp\left[\frac{r_t(\boldsymbol{x}_t'\boldsymbol{\beta})}{\sigma^2}\cos(\phi_t - \theta_t)\right]d\phi_t, \tag{1}$$

for $r_t \geq 0$, $\boldsymbol{x}_t'\boldsymbol{\beta} \geq 0$, and $\sigma^2 > 0$. The integral expression in (1) is equivalent to $\mathrm{I}_0(r_t\boldsymbol{x}_t'\boldsymbol{\beta}/\sigma^2)$, with $\mathrm{I}_0(\cdot)$ being the modified Bessel function of the first kind and the zeroth order(Abramowitz & Stegun, 1965). Thus, following common notation for (1), we have that $r_t \sim \mathrm{Rice}(\boldsymbol{x}_t'\boldsymbol{\beta}, \sigma)$, where the first parameter defines the deterministic signal level and the second defines the noise level; the definition of the signal-to-noise ratio (SNR) is accordingly $\boldsymbol{x}_t'\boldsymbol{\beta}/\sigma$. We note that the two parameters $\boldsymbol{x}_t'\boldsymbol{\beta}$ and σ are not the mean and the variance of the Rice distribution whose first two moments are $E(r_t; \boldsymbol{x}_t'\boldsymbol{\beta}, \sigma^2) = \sqrt{\pi\sigma^2/2}\mathrm{L}_{1/2}(-(\boldsymbol{x}_t'\boldsymbol{\beta})^2/2\sigma^2)$ and $E(r_t^2; \boldsymbol{x}_t'\boldsymbol{\beta}, \sigma^2) = (\boldsymbol{x}_t'\boldsymbol{\beta})^2 + 2\sigma^2$ (Zhu et al., 2009), where the Laguerre polynomial $\mathrm{L}_{1/2}(x) = \exp(-x/2)[(1-x)\mathrm{I}_0(-x/2) - x\mathrm{I}_1(-x/2)]$ and $\mathrm{I}_1(\cdot)$ is the modified Bessel function of the first kind and the first order (Abramowitz & Stegun, 1965).

2.1. Models for magnitude fMRI time series

In this section, we present the models and associated likelihood ratio tests (LRTs) for activation that we will compare in our investigations. Our treatment here assumes temporal independence of the magnitude time series, *e.g.* after prewhitening. To differentiate the signal and noise parameters, $\boldsymbol{\beta}$ and σ^2 respectively, and the LRT statistics Λ for the different models, we attach identifying subscripts – note, of course, that the design matrix $\boldsymbol{X}$ is the same for each model. The activation test posits $H_0 : \boldsymbol{C\beta} = 0$ (not activated) against $H_a : \boldsymbol{C\beta} \neq 0$ (activated). We illustrate the calculation of the restricted and unrestricted MLEs to correspond to the maximization of the likelihood function under the null and the alternative next: note that in all cases, the LRT statistics follow asymptotic χ_m^2 null distributions under all models, with $m = \mathrm{rank}(\boldsymbol{C})$.

2.1.1. LRTs under Gaussian Modeling We begin with the Gaussian model, widely used in fMRI (as elsewhere) due to its ease of application and the added fact that Ricean-distributed magnitudes are approximately Gaussian-distributed at high SNRs. In this setting, $\boldsymbol{r} = \boldsymbol{X\beta}_G + \boldsymbol{\epsilon}$, where the error term $\boldsymbol{\epsilon} \sim N(0, \sigma_G^2\boldsymbol{I}_n)$ with $\boldsymbol{I}_n$ denoting the identity matrix of order n. Unrestricted MLEs for the parameters $\boldsymbol{\beta}_G$ and σ_G^2 are $\hat{\boldsymbol{\beta}}_G = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{r}$ and $\hat{\sigma}_G^2 = (\boldsymbol{r} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_G)'(\boldsymbol{r} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_G)/n$, while the restricted MLEs are $\tilde{\boldsymbol{\beta}}_G = \Psi\hat{\boldsymbol{\beta}}_G$, where $\Psi = \boldsymbol{I}_q - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\left[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C}$, and $\tilde{\sigma}^2 = (\boldsymbol{r} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_G)'(\boldsymbol{r} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_G)/n$. As usual, the LRT statistic is given by $\Lambda_G = n\log(\tilde{\sigma}_G^2/\hat{\sigma}_G^2)$.

2.1.2. LRTs under the Rice model The Rice model is given by $r_t \sim$ indep Rice$(\boldsymbol{x}_t'\boldsymbol{\beta}_R, \sigma_R^2)$, $t = 1, 2, \ldots, n$, and following (1) has log-likelihood function (Rowe, 2005)

$$\log L(\boldsymbol{\beta}_R, \sigma_R^2|\boldsymbol{r}) = \sum_{t=1}^{n}\left[\log(r_t/\sigma_R^2) - \frac{r_t^2 + (\boldsymbol{x}_t'\boldsymbol{\beta}_R)^2}{2\sigma_R^2} + \log \boldsymbol{I}_0\left(\frac{r_t(\boldsymbol{x}_t'\boldsymbol{\beta}_R)}{\sigma_R^2}\right)\right]. \tag{2}$$

Using the Gaussian-model estimates as starting values, we propose calculating MLEs with a hybrid scheme that utilizes both EM and Newton-Raphson (NR) iterations (McLachlan & Krishnan, 2008), thus capitalizing on the stability of the former algorithm and the superior speed of convergence of the latter. Under unrestricted maximization, EM iterates update the kth step estimates $\hat{\boldsymbol{\beta}}_R^{(k)}$ and $\hat{\sigma}_R^{2(k)}$ to $\hat{\boldsymbol{\beta}}_R^{(k+1)} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{u}}^{(k)}$ and $\hat{\sigma}_R^{2(k+1)} = [\boldsymbol{r}'\boldsymbol{r} - (\boldsymbol{X}'\hat{\boldsymbol{u}}^{(k)})'(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\hat{\boldsymbol{u}}^{(k)})]/(2n)$ respectively, where $\hat{\boldsymbol{u}}^{(k)}$ is a vector of length n with tth entry $\hat{u}_t^{(k)} = r_t\mathrm{A}(\boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_R^{(k)}r_t/\hat{\sigma}_R^{2(k)})$, $t = 1, 2, \ldots, n$ and $\mathrm{A}(\cdot) = \mathrm{I}_1(\cdot)/\mathrm{I}_0(\cdot)$ (Solo & Noh, 2007). Under restricted maximization, EM updates are provided by $\tilde{\boldsymbol{\beta}}_R^{(k+1)} = \Psi(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{u}}^{(k)}$ and $\tilde{\sigma}_R^{2(k+1)} = [\boldsymbol{r}'\boldsymbol{r} - (\boldsymbol{X}'\tilde{\boldsymbol{u}}^{(k)})'\Psi(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\tilde{\boldsymbol{u}}^{(k)})]/(2n)$, where Ψ is as defined before in Section 2.1.1 and $\tilde{\boldsymbol{u}}^{(k)}$ has tth entry $\tilde{u}_t^{(k)} = r_t\mathrm{A}(\boldsymbol{x}_t'\tilde{\boldsymbol{\beta}}_R^{(k)}r_t/\tilde{\sigma}_R^{2(k)})$, $t = 1, 2, \ldots, n$. The NR iterations are derived from (2) using the derivative forms $\mathrm{I}_0'(\cdot) = \mathrm{I}_1(\cdot)$ and $A'(x) = 1 - \mathrm{A}(x)/x - A^2(x)$, for $x \neq 0$, $A'(0) = 0.5$ (Schou, 1978). In our implementation, we used a hybrid scheme with up to 1000 EM iterations, which brought about convergence – as measured by the change in (2) – in most cases. In case our algorithm had not converged by then, as was the case (only) for very low-SNR data (*i.e.* data with SNR < 1.5), we followed these EM iterations with a combination of NR and EM iterations to speed up convergence. An additional difficulty in the low-SNR case is that the constraints $\boldsymbol{x}_t'\boldsymbol{\beta}_R \geq 0$, $t = 1, 2, \ldots, n$ are harder to enforce and require quadratic programming methods. In all cases, the LRT statistic is given by $\Lambda_R = 2[\ell_R(\hat{\boldsymbol{\beta}}_R, \hat{\sigma}_R^2) - \ell_R(\tilde{\boldsymbol{\beta}}_R, \tilde{\sigma}_R^2)]$, where $\ell_R(\cdot,\cdot)$ is shorthand for (2). We conclude discussion in this section by noting, as in Solo & Noh (2007), that the Gaussian and Ricean estimates for $\boldsymbol{\beta}$ differ only by the "weight" function $\mathrm{A}(\cdot)$. Also, since $\mathrm{A}(z) \uparrow 1$ as $z \uparrow \infty$ and the argument increases with SNR, Solo & Noh (2007) recommend using $\mathrm{A}(\hat{\mu}_t r_t/\hat{\sigma}^2)$ as an indicator of whether measurements represent low or high SNR and whether the normal approximation is appropriate.

2.1.3. Alternate Approximate LRT derivations As mentioned in Section 1, den Dekker & Sijbers (2005) derives Gaussian- and Ricean-model-based LRT statistics under the assumption of known noise parameters. Notationally, we add asterisks to the parameters and LRT statistics under this assumption to distinguish them from their counterparts under estimated noise. For the Gaussian model, $\hat{\boldsymbol{\beta}}_G^* = \hat{\boldsymbol{\beta}}_G$ and $\tilde{\boldsymbol{\beta}}_G^* = \tilde{\boldsymbol{\beta}}_G$, and the LRT statistic is given by

$$\Lambda_G^* = [(\boldsymbol{r} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_G^*)'(\boldsymbol{r} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_G^*) - (\boldsymbol{r} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_G^*)'(\boldsymbol{r} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_G^*)]/\sigma_G^{2*}, \tag{3}$$

where σ_G^{2*} is the assumed variance. For the Ricean model, we calculate MLEs via an EM-NR hybrid scheme similar to the estimated variance case, except that σ_R^{2*}, the assumed (known) value of the Ricean noise parameter, is substituted for all iterates $\hat{\sigma}_R^{2(k)}$ and $\tilde{\sigma}_R^{2(k)}$. The LRT statistic is given by $\Lambda_R^* = 2[\ell_R(\hat{\boldsymbol{\beta}}_R^*, \sigma_R^{2*}) - \ell_R(\tilde{\boldsymbol{\beta}}_R^*, \sigma_R^{2*})]$.

The alternative "Taylor model" approach of Rowe (2005) approximates the Rice distribution by replacing the cosine term in (1) by the first two terms of its Taylor series expansion. The paper illustrates an iterative approach for maximizing the resulting log-likelihood, but in our investigations, we find that it does not produce exact MLEs. So, we utilize NR iterations instead. In addition, we find that the Taylor-model "PDF" does not integrate to one for low-SNR parameter values

Prepared using staauth.cls

as shown in Figure 1. Though this is cause for concern, for comparability with other published studies in the literature, we do not correct for this shortcoming in calculating the LRT statistic Λ_T. Further, because the Gaussian distribution also does not integrate to one over positive support, with the discrepancy especially acute at low SNRs, we also consider a Gaussian model truncated at zero and normalized to integrate to one, with PDF $f(r_t; \boldsymbol{\beta}_{TG}, \sigma^2_{TG}) = (2\pi)^{-1/2}\sigma^{-1}_{TG}\exp[-(r_t - \boldsymbol{x}_t'\boldsymbol{\beta}_{TG})^2/(2\sigma^2_{TG})][1 - \Phi(-\boldsymbol{x}_t'\boldsymbol{\beta}_{TG}/\sigma_{TG})]^{-1}$, for $r_t \geq 0$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF). The LRT statistic under this model, Λ_{TG}, can be computed using NR iterations. Table 1 provides a ready summary and reference of the different models and LRT statistics presented in this paper. We now discuss methods of evaluating these statistics.

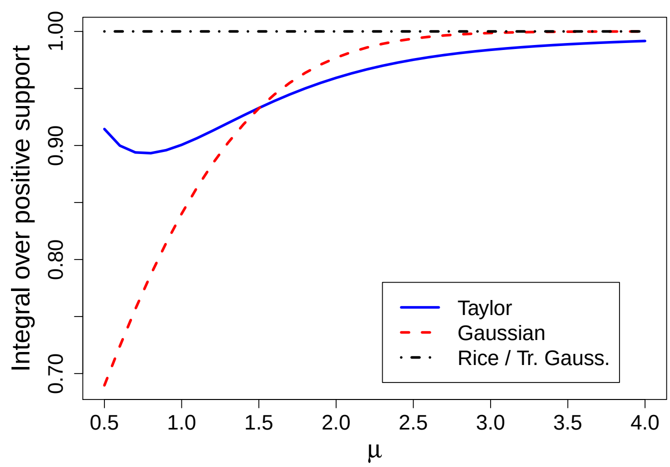


Figure 1. Integrals of Taylor, Gaussian, Ricean, and truncated Gaussian PDFs over positive support for different signal parameters μ and noise parameter $\sigma^2 = 1.0$.

Table 1. Summary of the models and LRT statistics presented in Section 2.1.

LRT Statistic	Model Description
Λ_G	Gaussian model with estimated variance
Λ_R	Ricean model with estimated noise parameter
Λ^*_G	Gaussian model with assumed variance
Λ^*_R	Ricean model with assumed noise parameter
Λ_T	Taylor model
Λ_{TG}	Truncated Gaussian model

2.2. *Methods for evaluating activation statistics*

We utilize three criteria in evaluating the LRTs. The first two are the rates of true and false (activation) detection – the rates of rejecting the null H_0 when it is in fact false and true, respectively. We compute the true and false detection rates from time series simulated under H_a and H_0 respectively; in both cases, for a significance level α, the detection rate is the proportion of LRT statistics greater than the $(1-\alpha)$th χ^2_m quantile. The third criterion, the area under the receiver operating characteristic (ROC) curve, or AUC, considers both null and alternative statistics at all significance levels. Denoting the kth-model test statistics, $k = 1, 2, \ldots, m$, computed under H_0 and H_a as $\{T^{(k)}_{0i}\}^{n_0}_{i=1}$ and $\{T^{(k)}_{aj}\}^{n_a}_{j=1}$ respectively, Bamber (1975) computes the AUC as $\hat{\tau}^{(k)} = \frac{1}{n_0 n_a}\sum^{n_0}_{i=1}\sum^{n_a}_{j=1} I(T^{(k)}_{0i} < T^{(k)}_{aj})$, where the indicator function $I(B)$ is 1 if B is true and 0 otherwise. A test with higher AUC has greater ability to discriminate statistics computed under H_0 and H_a, as the AUC above can be thought of as the proportion of null-alternative statistic pairs in which the rule $I(T^{(k)}_{0i} < T^{(k)}_{aj})$ discriminates the null and alternative statistics correctly. DeLong et al. (1988) develops significance tests for comparing AUCs based on the fact that the sample-based AUCs $\hat{\boldsymbol{\tau}} = (\hat{\tau}^{(1)}, \hat{\tau}^{(2)}, \ldots, \hat{\tau}^{(m)})$ are asymptotically normal, unbiased for the population AUCs $\boldsymbol{\tau}$, and have covariance matrix $\boldsymbol{S}$. As a result, the test $H_0: \tau^{(k)} = \tau^{(l)}$ vs. $H_a: \tau^{(k)} \neq \tau^{(l)}$ has the common z-score test statistic $z^{(kl)} = \hat{\tau}^{(k)} - \hat{\tau}^{(l)}/\sqrt{\boldsymbol{e}'_{kl}\boldsymbol{S}\boldsymbol{e}_{kl}}$ which asymptotically, under the null, has a standard normal distribution, with $\boldsymbol{e}_{kl}$ as a vector of length m with zeroes at all the coordinates but for the kth and lth positions which are 1 and -1, respectively.

To evaluate the six LRTs in our simulation study, we first disqualify any with false detection rates that deviate significantly from the nominal significance level. Then, for each two-way comparison of the remaining tests, we compute n_b replicates of the z-statistic ($z^{(kl)}$) based on n_b batches of $n_0 + n_a$ simulated time series. The proportion of significant z-statistics $\{z_b^{(kl)}\}_{b=1}^{n_b}$ at the α_1 level is $\hat{p}^{(kl)} = (1/n_b)\sum_{b=1}^{n_b} I(|z_b^{(kl)}| > z_{1-\alpha_1/2})$, where z_γ is the γth quantile of the standard normal distribution. Under $H_0 : \tau^{(k)} = \tau^{(l)}$, $n_b\hat{p}^{(kl)}$ follows a Binomial(n_b, α_1) distribution. Thus, we conclude that tests k and l are significantly different at the α_2 level if $\hat{p}^{(kl)} > U_{1-\alpha_2}$, the α_2th upper quantile of the Binomial(n_b, α_1) distribution divided by n_b.

3. A Motivating Example: Detecting Activation in a Finger-Tapping Experiment

We motivate our simulation study by analyzing a commonly-performed bilateral sequential finger-tapping experiment. The data are from Rowe & Logan (2004) and have been pre-processed, as detailed in that paper. In this case, the MR

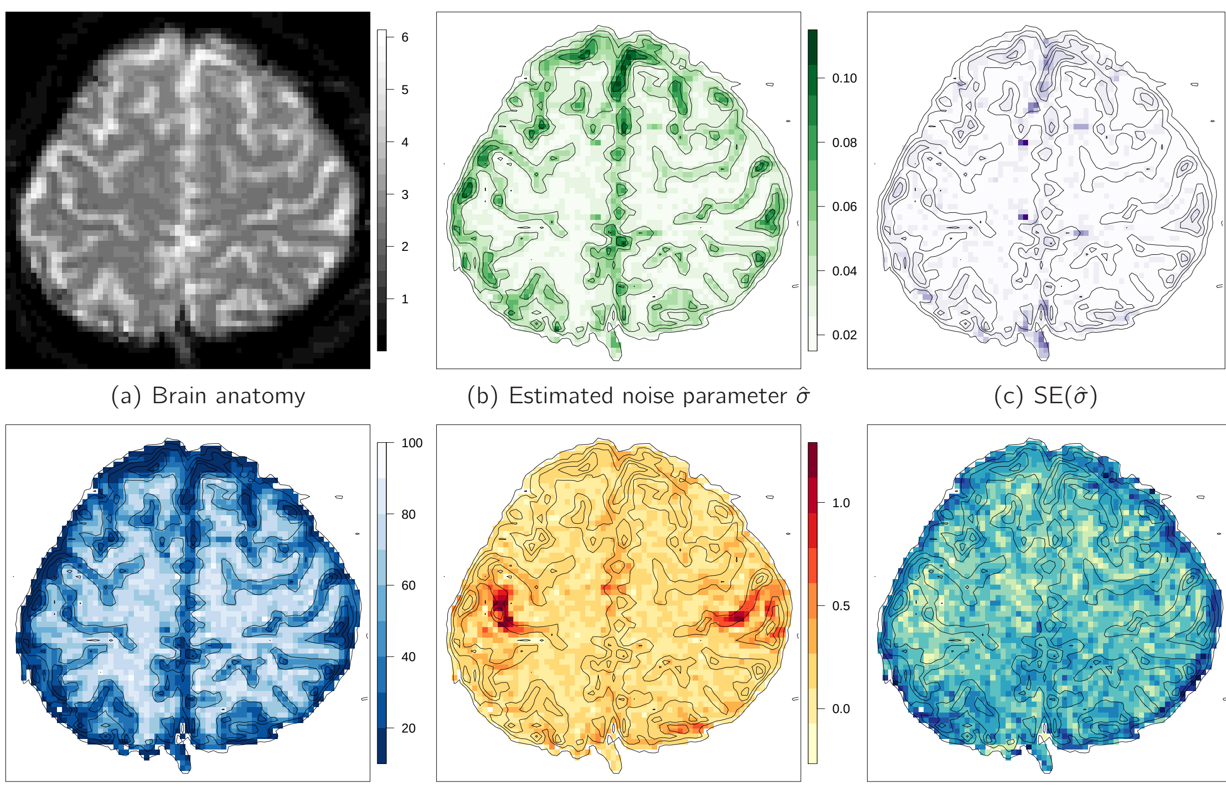
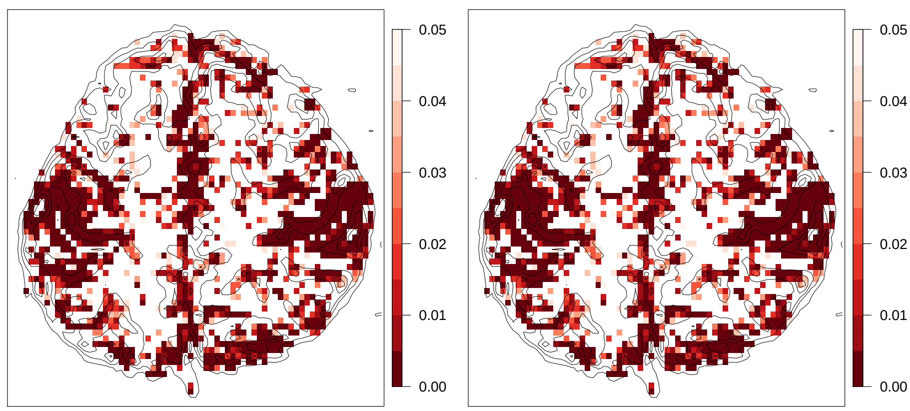


Figure 2. Images concerning the finger-tapping experiment presented in Section 3. (a) Anatomical image, which is shown as a contour plot in (b)-(f). (b),(c) Images of the estimated noise parameter $\hat{\sigma}$ and its standard error. (d)-(f) Images of the signal-, contrast-, and drift-to-noise ratios, respectively.

images were acquired while the (normal healthy male) volunteer subject was instructed to either lie at rest or to rapidly tap fingers of both hands at the same time. The fingers were tapped sequentially in the order of index, middle, ring, and little fingers. The experiment consisted of a block design with 16 s of rest followed by eight "epochs" of 16 s tapping alternating with 16 s of rest. MR scans were acquired once every second, resulting in 272 images. For simplicity, we restrict attention to a single axial slice through the motor cortex consisting of 128×128 voxels. In our study, steady state magnetization was not achieved until at least the fourth time-point: to guard against lingering effects, we delete the first 16 images and analyze the dataset based on a time-course sequence of the remaining 256 images. Magnitude time-course sequences at each voxel were fit using the Gaussian and Ricean models with estimated noise parameters presented in Section 2.1. The design matrix $\boldsymbol{X}$ contained three columns: an intercept representing the baseline MR signal level, a ± 1 square wave (lagged five points from the stimulus time course) representing the BOLD contrast, and an arithmetic sequence from -1 to 1 representing linear drift in the MR signal. Correspondingly, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ represents the size of the baseline, activation, and drift effects, respectively. Since only β_1 is activation-related, the activation test is $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$, and the LRT statistics have χ^2_1 null distributions.



(a) Gaussian model (b) Ricean model

Figure 3. Activation maps of q-values under (a) Gaussian and (b) Ricean LRTs.

Figure 2 displays images — aligned with anatomical contour plots — of the Ricean model estimates of the noise parameter σ, its standard error, and the ratios $(\beta_0, \beta_1, \beta_2)/\sigma$, the signal-, contrast-, and drift-to-noise ratios, respectively, or the SNR, CNR, and DNR. (We consider such ratios instead $\boldsymbol{\beta}$ itself because fMRI data is unitless.) First, we note that the varying estimates of σ shown in Figure 2(b), whose variation cannot alone be explained through the standard errors in Figure 2(c), are at odds with the assumption of a known (and thus constant) noise parameter as in den Dekker & Sijbers (2005). We use the estimated SNRs, CNRs, and DNRs in developing representative fMRI simulations in Section 4. We note that the SNRs for the finger-tapping dataset are above 10, a region for which den Dekker & Sijbers (2005) claims that Gaussian and Ricean activation tests should not have significantly different results. Our results support this claim. In fact, the largest absolute difference between the voxelwise Gaussian- and Ricean-model-based LRT statistics is less than 0.002. As a result, the Gaussian- and Ricean-based activation maps shown in Figure 3 — which consist of q-values, the analog of p-values in false discovery rate thresholding (Benjamini & Hochberg, 1995; Storey, 2002) — are essentially identical. The largest absolute difference between the voxelwise Gaussian- and Ricean-based q-values is 1.1×10^{-5}.

4. Experimental Evaluations

We assume that the simulated fMRI magnitude time series are generated from a block-design experiment such as that analyzed in Section 3. The time series follow $r_t \sim$ indep Rice$(\boldsymbol{x}_t'\boldsymbol{\beta}, \sigma^2)$, $t = 1, \ldots, 256$, where the design matrix $\boldsymbol{X}$ has the same columns as before. We fixed the noise parameter $\sigma^2 = 1.0$ across all simulations for easy interpretation of the SNR, CNR, and DNR. As in den Dekker & Sijbers (2005), we assume that $\sigma_R^{2*} = \sigma_G^{2*} = \sigma^2 = 1.0$. After applying each of the six models discussed in Section 2, we examine the parameter estimates in Section 4.1 and evaluate the activation statistics in Section 4.2.

4.1. Properties of parameter estimates

Figure 4 shows plots of bias, standard error, and root mean squared error (RMSE) against SNR for the MLEs of

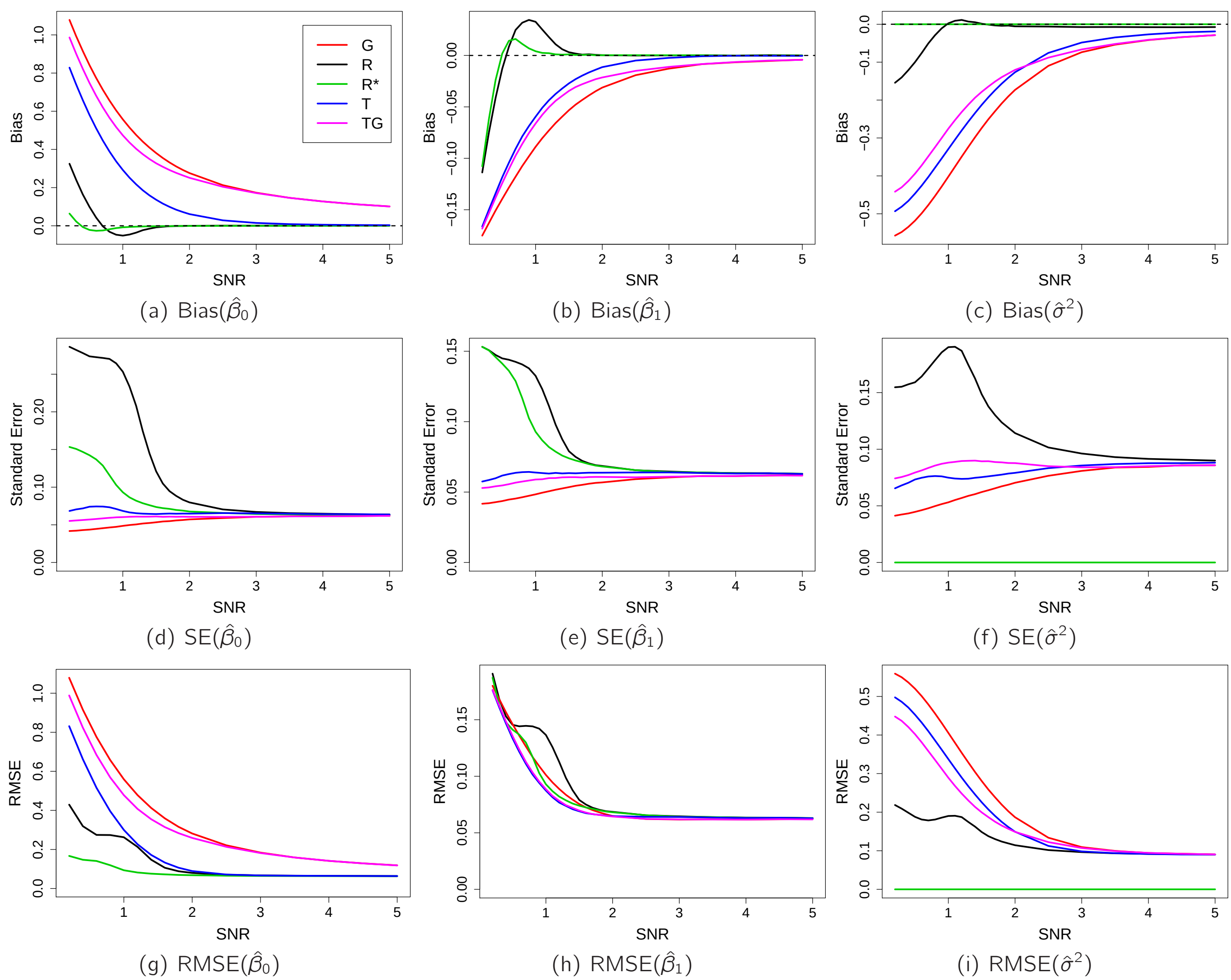
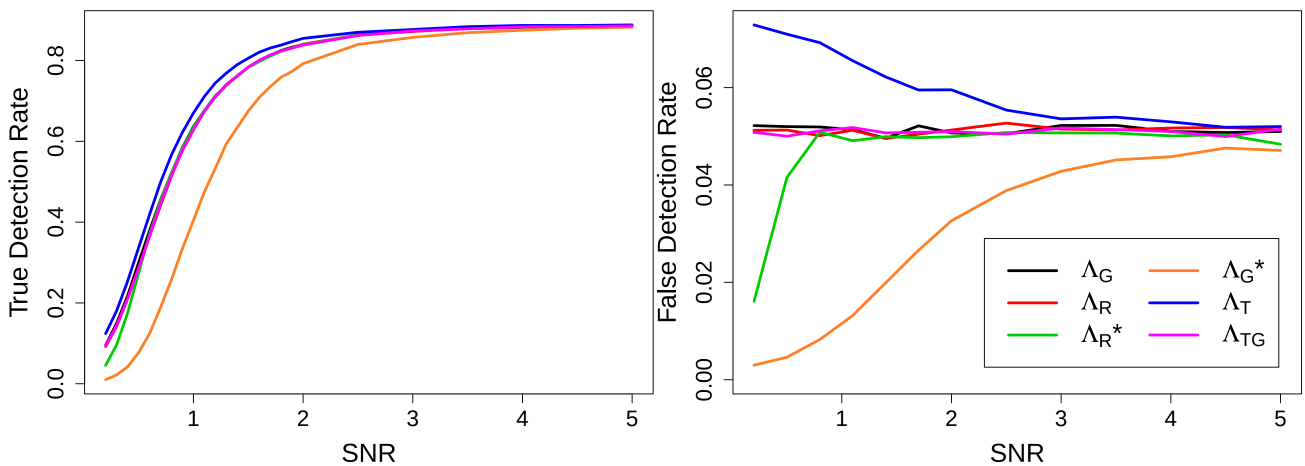


Figure 4. (a)-(c) Biases, (d)-(f) standard errors (SE), and (g)-(i) root mean squared errors (RMSE) of the unrestricted MLEs under each model plotted against SNR. The models are labeled in (a) as in Table 1. Note that estimates for the Gaussian model with assumed variance (G*) are not shown because they coincide with other models: that is, $\hat{\boldsymbol{\beta}}_G^* = \hat{\boldsymbol{\beta}}_G$ and $\sigma_G^* = \sigma_R^*$.

β_0, β_1, and σ^2 (β_2 is generally not of interest, and consequently not estimated, in typical fMRI experiments) under each model, which are based on 100,000 simulated time series at each of $\beta_0 = 0.2, 0.4, \ldots, 5.0$, with $\beta_1 = 0.2$ and

$\beta_2 = 0.0$. Overall, we note that the MLEs under each model differ most at low SNRs; however, as the SNR increases, their properties become more similar. Denoting the parameter vector by $\boldsymbol{\theta}$, we note that the Ricean-model MLEs $\hat{\boldsymbol{\theta}}_R$ and $\hat{\boldsymbol{\theta}}^*_R$ show the least amount of bias, the Gaussian-model MLEs $\hat{\boldsymbol{\theta}}_G$, $\hat{\boldsymbol{\theta}}^*_G$, and $\hat{\boldsymbol{\theta}}_{TG}$ show the most, and the biases of Taylor-model MLEs are in between. This result should not be surprising because the Ricean model parameters correspond exactly to those of the generated data while those in the Taylor and Gaussian models correspond only approximately. However, there seems to be a trade-off between the bias and variance of the estimates, as the Ricean-model MLEs (which are numerically calculated) show larger standard errors than the Gaussian-model MLEs (which are analytically obtained in closed form). The results for the RMSE, which encompasses both bias and variance, are mixed: for instance, the MLEs $\hat{\beta}_{0R}$ and $\hat{\sigma}^2_R$ have the lowest RMSEs of all models, but $\hat{\beta}_{1R}$ has the highest RMSE.

4.2. Evaluation of activation tests



(a) True detection rate (b) False detection rate

Figure 5. (a) True and (b) false detection rates of the different LRT statistics, according to an $\alpha = 0.05$ significance level, plotted against SNR. The legend in (b) follows Table 1. In (a), the lines for Λ_G and Λ_R are not visible because they coincide with the line for Λ_{TG}.

Figure 5 shows plots against the SNR of the true and false detection rates of the LRT statistics for each model for a significance level of $\alpha = 0.05$, which are based on 100,000 simulated time series at each of $\beta_0 = 0.2, 0.4, \ldots, 5.0$, with $\beta_1 = 0.2$ and 0.0 (for true and false detection, respectively) and $\beta_2 = 0.0$. As seen in den Dekker & Sijbers (2005), the true detection rates of Λ^*_R are greater than Λ^*_G with a difference that increases with decreasing SNR; also, as noted in the paper, the false detection rates of Λ^*_G fail to adhere to the significance level and are not constant with SNR. However, results differ for their counterparts with estimated variance parameters: the true detection rates of Λ_R and Λ_G are more comparable, and the false detection rate of Λ_G is closer than Λ^*_G to $\alpha = 0.05$. We attribute the above differences to the assumption $\sigma^{2*}_G = \sigma^2_R$. When the Gaussian model is applied to the simulated Rice-distributed data, σ^2_G represents the variance of the Rice-distributed data, which, as discussed in Section 2, differs from the Ricean parameter σ^2_R. To illustrate, we plot the variance of the Rice$(\mu, 1)$ distribution and the middle 95% of the estimates $\hat{\sigma}^2_G$ for simulated Rice$(\mu, 1)$ data for different μ in Figure 6. At low SNRs, the estimates $\hat{\sigma}^2_G$ are smaller than the assumed value $\sigma^{2*}_G = \sigma^2_R$. Because σ^{2*}_G is over-specified at low SNRs, by the form of (3), Λ^*_G takes lower values than Λ_G, which results in the former's lower true and false detection rates. Further, as suggested by Rowe

(2005), the true detection rates of Λ_T are greater than Λ_G. However, this may be explained by the former's higher false detection rate, perhaps due to the improper Taylor model PDF, which prevent Λ_T from being a usable test. The true and false detection rates of Λ_G and Λ_{TG} are similar at low SNRs so that it appears that the impropriety of the Gaussian model PDF may not be an issue then. We see no similar problems with the Gaussian model PDF at low SNR which also has a higher false detection rate than Λ_G, perhaps because the Taylor model PDF does not integrate to one. As a result, Λ_T, like Λ_G^*, is not a usable test. Because the false detection rates of Λ_T and Λ_G^* fail to adhere to significance level, we remove these tests from further comparisons.

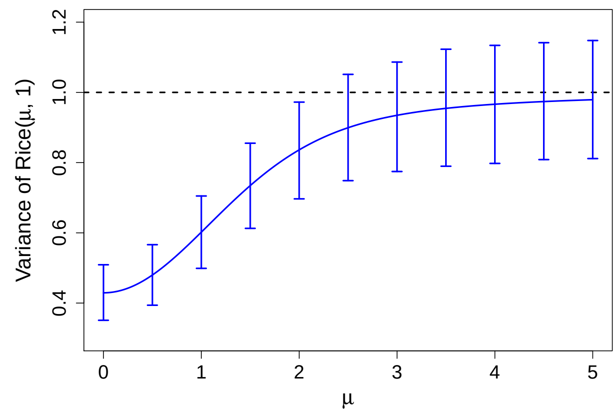
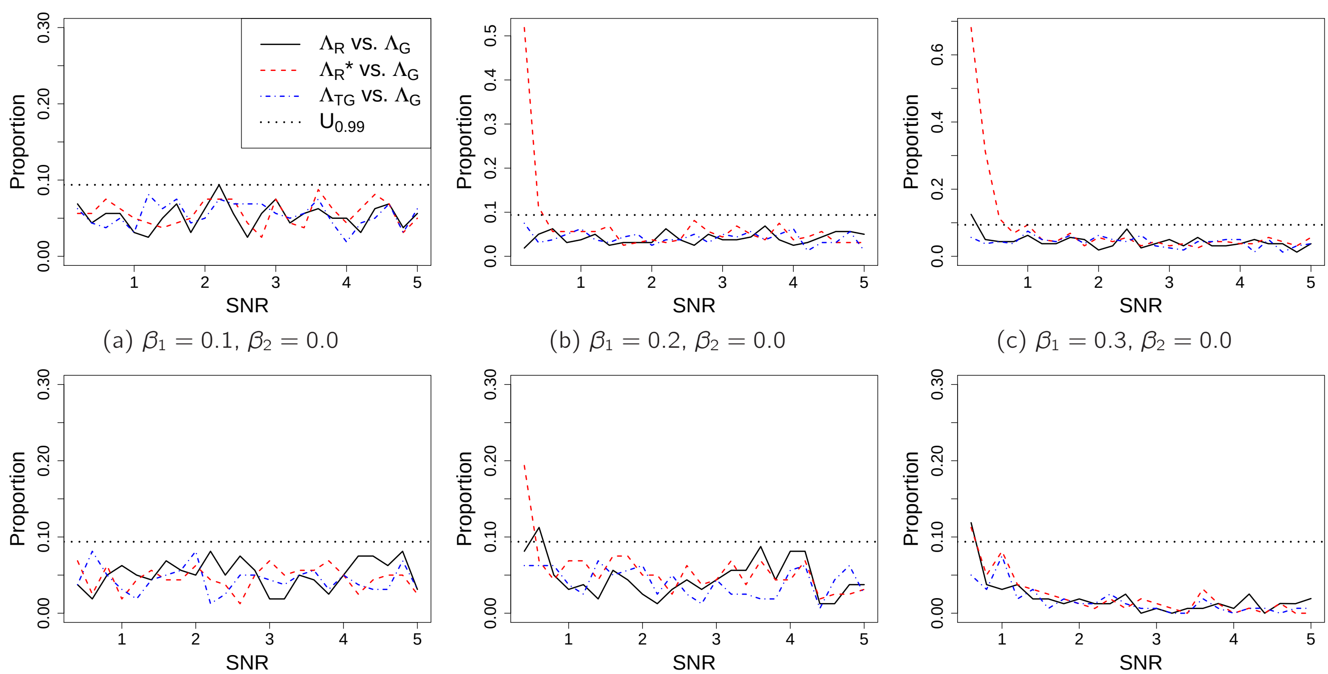


Figure 6. The variance of the Rice$(\mu, 1.0)$ distribution plotted against μ (or alternatively, SNR), with estimates of the middle 95% of the distributions of $\hat{\sigma}_G^2$ (obtained from simulation) at $\mu = 0.0, 0.5, \ldots, 5.0$. A horizontal line at $\sigma_G^{2*} = \sigma_R^2 = 1.0$ is given for comparison.

We evaluate the remaining LRTs using the AUC-based analysis described in Section 2.2. Because the Gaussian model is most commonly used in practice, we use it as a baseline, computing $z^{(k,\Lambda_G)}$ for $k = \Lambda_R, \Lambda_R^*, \Lambda_{TG}$. We compute $n_b = 160$ batches of z-statistics, each based on $n_0 = n_a = 1000$ null and alternative LRT statistics, at SNR levels β_0 from 0.2 to 5.0, activation levels $\beta_1 = 0.1$, 0.2, and 0.3, and drift levels $\beta_2 = 0.0$ and 0.2. Figure 7 plots $\hat{p}^{(k,\Lambda_G)}$, $k = \Lambda_R, \Lambda_R^*, \Lambda_{TG}$, for $\alpha_1 = 0.05$ against SNR for the various activation and drift levels and displays $U_{0.99}$



(a) $\beta_1 = 0.1$, $\beta_2 = 0.0$ (b) $\beta_1 = 0.2$, $\beta_2 = 0.0$ (c) $\beta_1 = 0.3$, $\beta_2 = 0.0$

(d) $\beta_1 = 0.1$, $\beta_2 = 0.2$ (e) $\beta_1 = 0.2$, $\beta_2 = 0.2$ (f) $\beta_1 = 0.3$, $\beta_2 = 0.2$

Figure 7. Plots of $\hat{p}^{(k,\Lambda_G)}$, $k = \Lambda_R, \Lambda_R^*, \Lambda_{TG}$, for an $\alpha_1 = 0.05$ significance level, against SNR for the various activation (β_1) and drift (β_2) levels; for comparison, we display $U_{0.99}$, the upper 99% quantile of $\hat{p}^{(k,\Lambda_G)}$ under AUC equality.

for comparison. At all activation/drift levels and SNR ≤ 0.6, $\hat{p}^{(k,\Lambda_G)} \leq U_{0.99}$, indicating that the AUCs of the Rician- and the truncated-Gaussian-model-based LRTs are not significantly different from the Gaussian LRT.

5. Conclusion

In this paper, we have studied the effects of Gaussian and Ricean modeling of low-SNR fMRI magnitude time series. Noting that previous studies showing improved performance of Ricean-based activation tests were based on assumptions and approximations, our simulation study included both these previous tests and tests which we developed further and removed the assumptions. It became apparent that some of the previous comparisons of Ricean- and Gaussian-based tests were flawed. Specifically, we argue that the Gaussian-based test in den Dekker & Sijbers (2005) is based on an incorrect assumption and that the Ricean-approximated test in Rowe (2005) is not usable because its false detection rate is incompatible with its desired significance level. After addressing these issues, we found that the performances of Ricean- and Gaussian-model activation tests, as measured by the AUC, are significantly different at SNRs much lower than earlier results indicated (SNR ≤ 0.6 versus 10.0), perhaps too low a range for Ricean-based activation tests to be practically beneficial. Therefore, based on the Gaussian model's simple implementation and low computational expense, we recommend it over the Ricean model at all SNR for activation tests based on fMRI magnitude time series.

A few comments are in order. We note that our simulation experiments have used pre-whitened time series and then proceeded with the testing under assumptions of independence. This is not just a matter of simplicity, but because parameter estimation of the time series under the Ricean model remains intractable. It would be of interest to see if suitable estimates of Ricean time series can be developed. However, there is some reason to doubt that our recommendation will be overturned, given our findings on how much lower SNR's have to be than seen in fMRI as currently practiced, for Ricean-based tests to have a clear preference over the Gaussian-based ones. A second, but important, issue involves the (sometimes ad hoc) pre-processing that is often done in real-world fMRI experiments (such as in Section 3) to account for distortions owing to bias fields, imaging modality used, scanner drift, subject motion, physiological factors and so on (Buxton, 2002; Lazar, 2008). There are thus several steps, such as slice timing correction, image registration, etc. that are performed on the acquired (raw) Rice-distributed magnitude data. While these preprocessing steps are difficult to capture in a simulation setting, we note that many of the common corrections (*e.g., registration*) are essentially linear so that the resulting data are really linear combinations of Rice-distributed data. However, given that our idealized simulation scenario does not recommend Ricean- over Gaussian-modeling, it is unlikely that our findings will be overturned in a situation where the (pre-processed) data are (mostly linear) transformations of the raw acquired magnitude measurements. (This is because, as commonly known, linear transformations respect the Gaussian distribution: for other transformations, this relationship is asymptotic – upon appealing to the Delta method.) Finally, we note that our tests have been framed in the context of fMRI as currently practiced. We have not discussed the recommendations of Nan & Nowak (1999) or Rowe & Logan (2004) who have argued for fMRI analysis using both the magnitude and phase information in the original acquired data. It would be interesting to include an analysis using these models. Thus, we see that while we have a clear recommendation in favor of the Gaussian model for fMRI as currently practiced, a few issues meriting further attention remain.

Acknowledgement

The National Science Foundation (NSF) partially supported the research of the first and the second authors under its Grant No. DMS-0502347 and its CAREER Grant No. DMS-0437555, respectively. The research of the second author was also supported in part by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the

National Institutes of Health (NIH) under its Award No. R21EB016212. The content of this paper however is solely the responsibility of the authors and does not represent the official views of either the NSF or the NIH.

References

Abramowitz, M & Stegun, I (1965), *Handbook of Mathematical Functions*, Dover Publications.

Aja-Fernández, S, Tristán-Vega, A & Alberola-Lòpez, C (2009), 'Noise estimation in single- and multiple-coil Magnetic Resonance data based on statistical models,' *Magnetic Resonance Imaging*, **27**(10), pp. 1397–1409.

Bamber, D (1975), 'The area above the ordinal dominance graph and the area below the receiver operating characteristic graph,' *Journal of Mathematical Psychology*, **12**, pp. 387–415.

Bandettini, PA, Jesmanowicz, A, Wong, EC & Hyde, JS (1993), 'Processing strategies for time-course data sets in functional MRI of the human brain,' *Magnetic Resonance in Medicine*, **30**, pp. 161–173.

Belliveau, JW, Kennedy, DN, McKinstry, RC, Buchbinder, BR, Weisskoff, RM, Cohen, MS, Vevea, JM, Brady, TJ & Rosen, BR (1991), 'Functional mapping of the human visual cortex by Magnetic Resonance imaging,' *Science*, **254**, pp. 716–719.

Benjamini, Y & Hochberg, Y (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing,' *Journal of the Royal Statistical Society. Series B (Methodology)*, **57**(1), pp. 289–300.

Buxton, RB (2002), *Introduction to functional Magnetic Resonance Imaging: Principles and techniques*, Cambridge University Press.

DeLong, ER, DeLong, DM & Clarke-Pearson, DL (1988), 'Comparing the areas under two or more correlated receiver operating characterist curves: a nonparameteric approach,' *Biometrics*, **44**, pp. 837–845.

Dempster, AP, Laird, NM & Rubin, D (1977), 'Maximum likelihood from incomplete data via the EM algorithm,' *Journal of Royal Statistical Society Series B*, **23**, pp. 1–38.

den Dekker, AJ & Sijbers, J (2005), 'Implications of the Rician distribution for fMRI generalized likelihood ratio tests,' *Magnetic Resonance Imaging*, **23**, pp. 953–959.

Friston, KJ, Frith, CD, Liddle, PF, Dolan, RJ, Lammertsma, AA & Frackowiak, RSJ (1990), 'The relationship between global and local changes in PET scans,' *Journal of Cerebral Blood Flow and Metabolism*, **10**, pp. 458–466.

Friston, KJ, Holmes, AP, Worsley, KJ, Poline, JB, Frith, CD & Frackowiak, RSJ (1995), 'Statistical parametric maps in functional imaging: A general linear approach,' *Human Brain Mapping*, **2**, pp. 189–210.

Genovese, CR, Lazar, NA & Nichols, TE (2002), 'Thresholding of statistical maps in functional neuroimaging using the false discovery rate,' *NeuroImage*, **15**, pp. 870–878.

Gudbjartsson, H & Patz, S (1995), 'The Rician distribution of noisy data,' *Magnetic Resonance in Medicine*, **34**(6), pp. 910–914.

Jezzard, P & Clare, S (2001), 'Principles of Nuclear Magnetic Resonance and MRI,' in Jezzard, P, Matthews, PM & Smith, SM (eds.), *Functional MRI: An Introduction to Methods*, Oxford University Press, chap. 3, pp. 67–92.

Kumar, A, Welti, D & Ernst, RR (1975), 'NMR Fourier zeugmatography,' *Journal of Magnetic Resonance*, **18**, pp. 69–83.

Kwong, KK, Belliveau, JW, Chesler, DA, Goldberg, IE, Weisskoff, RM, Poncelet, BP, Kennedy, DN, Hoppel, BE, Cohen, MS, Turner, R, Cheng, HM, Brady, TJ & Rosen, BR (1992), 'Dynamic Magnetic Resonance imaging of human brain activity during primary sensory stimulation,' *Proceedings of the National Academy of Sciences of the United States of America*, **89**, pp. 5675–5679.

Lazar, NA (2008), *The Statistical Analysis of Functional MRI Data*, Springer.

Logan, BR & Rowe, DB (2004), 'An evaluation of thresholding techniques in fMRI analysis,' *NeuroImage*, **22**, pp. 95–108.

Maitra, R (2013), 'On the Expectation-Maximization algorithm for Rice-Rayleigh mixtures with application to noise parameter estimation in magnitude MR datasets,' *Sankhyä*, p. to appear, doi:10.1007/s13571-012-0055-y.

Maitra, R & Faden, D (2009), 'Noise estimation in magnitude MR datasets,' *IEEE Transactions on Medical Imaging*, **28**(10), pp. 1615–1622.

Maitra, R & Riddles, JJ (2010), 'Synthetic Magnetic Resonance Imaging revisited,' *IEEE Transactions on Medical Imaging*, **29**(3), pp. 895 –902, doi:10.1109/TMI.2009.2039487.

McLachlan, GJ & Krishnan, T (2008), *The EM Algorithm and Extensions*, Wiley.

Nan, FY & Nowak, RD (1999), 'Generalized likelihood ratio detection for fmri using complex data,' *IEEE Transactions on Medical Imaging*, **18**, pp. 320–329.

Noh, J & Solo, V (2011), 'Rician distributed fMRI: Asymptotic power analysis and Cramér-Rao lower bounds,' *IEEE Transactions on Signal Processing*, **59**, pp. 1322–1328.

Ogawa, S, Lee, TM, Nayak, AS & Glynn, P (1990), 'Oxygenation-sensitive contrast in Magnetic Resonance image of rodent brain at high magnetic fields,' *Magnetic Resonance in Medicine*, **14**, pp. 68–78.

Rajan, J, Poot, D, Juntu, J & Sijbers, J (2010), 'Noise measurement from magnitude MRI using local estimates of variance and skewness,' *Physics in Medicine and Biology*, **55**, pp. N441–449.

Rice, SO (1944), 'Mathematical analysis of random noise,' *Bell Systems Technical Journal*, **23**, p. 282.

Rowe, DB (2005), 'Parameter estimation in the magnitude-only and complex-valued fMRI data models,' *NeuroImage*, **25**, pp. 1124–1132.

Rowe, DB & Logan, BR (2004), 'A complex way to compute fMRI activation,' *NeuroImage*, **23**, pp. 1078–1092.

Schou, G (1978), 'Estimation of the concentration parameter in von Mises-Fisher distributions,' *Biometrika*, **65**, pp. 369–377.

Sijbers, J (1998), *Signal and Noise Estimation from Magnetic Resonance Images*, Ph.D. thesis, University of Antwerp.

Sijbers, J, Poot, D, den Dekker, AJ & Pintjens, W (2007), 'Automatic estimation of the noise variance from the histogram of a Magnetic Resonance image,' *Physics in Medicine and Biology*, **52**, pp. 1335–1348.

Solo, V & Noh, J (2007), 'An EM algorithm for Rician fMRI activation detection,' in *ISBI*, pp. 464–467.

Storey, JD (2002), 'A direct approach to false discovery rates,' *Journal of the Royal Statistical Society B*, **64**, pp. 479–498.

Wang, T & Lei, T (1994), 'Statistical analysis of MR imaging and its applications in image modeling,' *Proceedings of the IEEE International Conference on Image Processing and Neural Networks*, **1**, pp. 866–870.

Worsley, KJ, Marrett, S, Neelin, P, Vandal, AC, Friston, KJ & Evans, AC (1996), 'A unified statistical approach for determining significant voxels in images of cerebral activation,' *Human Brain Mapping*, **4**, pp. 58–73.

Zhu, H, Li, Y, Ibrahim, JG, Shi, X, An, H, Chen, Y, Gao, W, Lin, W, Rowe, DB & Peterson, BS (2009), 'Regression models for identifying noise sources in Magnetic Resonance images,' *Journal of the American Statistical Association*, **104**, pp. 623–637.