Parameter estimation in the magnitude-only and complex-valued fMRI data models

Daniel B. Rowe*

Department of Biophysics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, USA

In functional magnetic resonance imaging, voxel time courses are complex-valued data but are traditionally converted to real magnitude-only data ones. At a large signal-to-noise ratio (SNR), the magnitude-only data Ricean distribution is approximated by a normal distribution that has been suggested as reasonable in magnitude-only data magnetic resonance images for an SNR of 5 and potentially as low as 3. A complex activation model has been recently introduced by Rowe and Logan [Rowe, D.B., and Logan, B.R. (2004). A complex way to compute fMRI activation. Neuro-Image, 23 (3):1078–1092] that is valid for all SNRs. The properties of the parameter estimators and activation statistic for these two models and a more accurate Ricean approximation based on a Taylor series expansion are characterized in terms of bias, variance, and Cramer–Rao lower bound. It was found that the unbiased estimators in the complex model continued to be unbiased for lower SNRs while those of the normal magnitude-only data model became biased as the SNR decreased and at differing levels for the regression coefficients. The unbiased parameter estimators from the approximate magnitude-only Ricean Taylor model were unbiased to lower SNRs than the magnitude-only normal ones. Further, the variances of the parameter estimators achieved their minimum value in the complex data model regardless of SNR while the magnitude-only data normal model and Ricean approximation using a Taylor series did not as the SNR decreased. Finally, the mean activation statistic for the complex data model was higher and not SNR dependent while it decreased with SNR in the magnitude-only data models but less so for the approximate Ricean model. These results indicate that using the complex data model and not approximations to the true magnitude-only Ricean data model is more appropriate at lower SNRs. Therefore, since the computational cost is relatively low for the complex data model and since the SNR is not inherently known a priori for all voxels, the complex data model is recommended at all SNRs.
© 2004 Elsevier Inc. All rights reserved.

Keywords: fMRI; Signal-to-noise ratio; Magnitude-only data model

Introduction

Recently, a complex-valued time course model was introduced by Rowe and Logan (2004) to determine functional brain activation (Rowe and Logan, 2004). This model builds upon previous work in which pre magnitude complex-valued voxel time courses were used to determine brain activation (Lai and Glover, 1997; Nan and Nowak, 1999). This model showed improved power of detection at low signal-to-noise ratios (SNRs) and low contrast-to-noise ratios (CNRs) for three distinct thresholding methods described in Logan and Rowe (2004).

The improvement in power of the complex data model over the magnitude-only data model may be a combination of having more quantities to estimate the error variance thus increasing the significance of an activation statistic and not making a distributional approximation. If the improvement were solely due to not making a distributional approximation, then it would be expected that a magnitude-only data model that does not approximate the Ricean distribution with the normal distribution would perform nearly identical to the complex data model. If the improvement in power of the complex data model over the magnitude-only data model were solely due to a reduction in the error variance, then a model that does not approximate the Ricean distribution with the normal distribution would perform nearly identical to the magnitude-only data model. In this paper, a good approximation to the Ricean distribution is introduced using a truncated Taylor series expansion. It will be seen that the performance of the Taylor approximation model is between that of the magnitude-only normal and complex data model.

Subsequently, Rowe and Logan (2005) generalized the complex model to have no restrictions on the phase time courses. They showed that this unrestricted phase model was mathematically equivalent to the usual magnitude-only data model including regression coefficients and voxel activation statistic but philosophically different due to its derivation from complex data (Rowe and Logan, 2005).

Models

As previously noted, fMRI voxel time courses are complex-valued. Recently, Rowe and Logan (2004) introduced a complex-

* Fax: +1 414 456 6512.
  E-mail address: dbrowe@mcw.edu.
  Available online on ScienceDirect (www.sciencedirect.com).

valued model to characterize task related magnitude changes. This model will be briefly summarized and the magnitude-only data Ricean model derived from it with a Taylor series approximation and the usual normal approximation.

Complex model

Voxelwise complex-valued fMRI data can be characterized by a nonlinear multiple regression model with $n \times (q + 1)$ dimensional design matrix $X = (x_1, \ldots, x_n)'$ as described in Rowe and Logan (2004) where n is the number of time points and q is the number of non-baseline regressors. For example, a simple model could have a baseline, a linear trend, and a reference function. In this simple model, X would have three columns, the first being a column of ones, the second a column of counting numbers, and the third a column containing the reference function. In this model which includes a phase imperfection θ, the complex-valued voxel observations at time t are given by

$$y_t = [(x_t'\beta)\cos\theta + \eta_{Rt}] + i[(x_t'\beta)\sin\theta + \eta_{It}] \tag{2.1}$$

where $x_t'\beta = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_q x_{qt}$, $(\eta_{Rt}, \eta_{It})' \sim \mathcal{N}(0, \Sigma)$ and $\Sigma = \sigma^2 I_2$.

The distributional specification is on the real and imaginary parts of the observed signal and not on the magnitude. The phase imperfection in Eq. (2.1) is assumed to be fixed and unknown, but may be estimated voxel-by-voxel as in Rowe and Logan (2004).

Alternatively, the observed data can be represented at time point t as a 2×1 vector instead of as a complex number

$$\begin{pmatrix} y_{Rt} \\ y_{It} \end{pmatrix} = \begin{pmatrix} x_t'\beta\cos\theta \\ x_t'\beta\sin\theta \end{pmatrix} + \begin{pmatrix} \eta_{Rt} \\ \eta_{It} \end{pmatrix}, \qquad t = 1, \ldots, n.$$

This model can also be written to describe the observations at all time points simultaneously as

$$\begin{array}{cccc} y & = & \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix} & \begin{pmatrix} \beta\cos\theta \\ \beta\sin\theta \end{pmatrix} + & \eta \\ {\scriptstyle 2n \times 1} & & {\scriptstyle 2n \times 2(q+1)} \quad {\scriptstyle 2(q+1) \times 1} & {\scriptstyle 2n \times 1} \end{array} \tag{2.2}$$

where it is specified that the observed vector of data $y = (y_R', y_I')'$ is the vector of observed real values stacked on the vector of observed imaginary values and the vector of errors $\eta = (\eta_R', \eta_I')' \sim \mathcal{N}(0, \Sigma \otimes \Phi)$ is similarly defined. Here, it is assumed that $\Sigma = \sigma^2 I_2$ and $\Phi = I_n$.

With the aforementioned distributional specifications, the likelihood of the complex-valued data model is

$$p(y|X, \beta, \theta, \sigma^2) = (2\pi\sigma^2)^{-n}$$
$$\times \exp\left\{ -\frac{1}{2\sigma^2}\left[y - \begin{pmatrix} X\beta\cos\theta \\ X\beta\sin\theta \end{pmatrix}\right]'\right.$$
$$\left.\times \left[y - \begin{pmatrix} X\beta\cos\theta \\ X\beta\sin\theta \end{pmatrix}\right]\right\} \tag{2.3}$$

from which maximum likelihood estimators (MLEs) and Cramer–Rao lower bounds (CRLBs) can be computed.

Ricean model

In fMRI, repeated measurements are taken over time while a subject is performing a task. In each voxel, we usually compute a measure of association between the observed time course and a preassigned reference function that characterizes the experimental paradigm. The typical method to compute activations (Bandettini et al., 1993; Cox et al., 1995) is to use only the magnitude at time t denoted by r_t and written as

$$r_t = \left[(x_t'\beta\cos\theta + \eta_{Rt})^2 + (x_t'\beta\sin\theta + \eta_{It})^2\right]^{\frac{1}{2}}. \tag{2.4}$$

As previously outlined (Rowe and Logan, 2004), the magnitude of a complex valued observation at time t is Ricean distributed (Gudbjartsson and Patz, 1995; Rice, 1944) and given by

$$p(r_t|x_t, \beta, \sigma^2) = \frac{r_t}{\sigma^2}\exp\left\{ -\frac{1}{2\sigma^2}\left[r_t^2 + (x_t'\beta)^2\right]\right\}$$
$$\times \int_{\phi_t = -\pi}^{\pi} \frac{1}{2\pi}\exp\left\{\frac{x_t'\beta r_t}{\sigma^2}\cos(\phi_t - \theta)\right\}d\phi_t, \tag{2.5}$$

where a general linear model is assumed and the integral factor often denoted I_o $(x_t'\beta r_t/\sigma^2)$ is the zeroth order modified Bessel function of the first kind. It is well known that for "large" SNRs, the Ricean distribution of the magnitude r_t in Eq. (2.5) is approximately normal with mean $x_t'\beta$ and variance σ^2. When the SNR is zero, the Ricean distribution is a Rayleigh distribution. It is intermediate SNR values that are of interest along with guidelines for what is considered a "large" SNR.

Maximum likelihood estimates of the parameters (β, σ^2) can be determined for both restricted null and unrestricted alternative hypotheses. For example with unrestricted null and alternative hypotheses $H_0: C\beta = \gamma$ versus $H_1: C\beta \neq \gamma$ where C is a full row rank matrix where the hypotheses to be tested are in the form of linear constraints in the rows and γ is a column vector of dimension equal to the full row rank of C. The likelihood

$$p(r|X, \beta, \sigma^2) = (\sigma^2)^{-n}\left(\prod_{t=1}^{n} r_t\right)\exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{t=1}^{n} r_t^2 + \sum_{t=1}^{n}(x_t'\beta)^2\right)\right\}$$
$$\times \prod_{t=1}^{n} I_o(x_t'\beta r_t/\sigma^2),$$

or the logarithm of the likelihood

$$LL = -n\log(\sigma^2) + \sum_{t=1}^{n}\log r_t$$
$$-\frac{1}{2\sigma^2}\left(\sum_{t=1}^{n} r_t^2 + \sum_{t=1}^{n}(x_t'\beta)^2\right) + \sum_{t=1}^{n}\log I_o(x_t'\beta r_t/\sigma^2) \tag{2.6}$$

is maximized where $r = (r_1, \ldots, r_n)'$. When maximizing the log likelihood under the restricted null hypothesis, the Lagrange multiplier constraint $\psi'(C\beta - \gamma)$ is added. If the parameter estimates under the restricted null hypothesis are denoted $(\tilde{\beta}, \tilde{\sigma}^2)$ and those under the unrestricted alternative hypothesis $(\hat{\beta}, \hat{\sigma}^2)$, then substituting back into the likelihood and the ratio λ of restricted null over unrestricted alternative leads to the large sample χ^2 distributed statistic

$$-2\log\lambda = 2n\log(\tilde{\sigma}^2/\hat{\sigma}^2) + \frac{1}{\tilde{\sigma}^2}\left(\sum_{t=1}^{n} r_t^2 + \sum_{t=1}^{n}(r_t\tilde{\beta})^2\right)$$
$$-\frac{1}{\hat{\sigma}^2}\left(\sum_{t=1}^{n} r_t^2 + \sum_{t=1}^{n}(x_t'\hat{\beta})^2\right)$$
$$-2\sum_{t=1}^{n}\log\left[I_o(x_t'\tilde{\beta}r_t/\tilde{\sigma}^2)/I_o(x_t'\hat{\beta}r_t/\hat{\sigma}^2)\right]$$

with degrees of freedom equal to the full row rank of C.

Taylor model

However, an exact solution that maximizes the logarithm of the likelihood given in Eq. (2.6) under both hypotheses is difficult in practice and requires nonstandard numerical maximization. An approximation to the Ricean distribution will be explored that is valid at lower SNRs than the usual normal approximation. A more accurate approximation to the Ricean distribution is

$$p(r_t|x_t, \beta, \sigma^2) = \sqrt{\frac{r_t}{x_t'\beta}} \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\times \exp\left\{ -\frac{1}{2\sigma^2}\left[r_t - (x_t'\beta)\right]^2 \right\}, \tag{2.7}$$

which is found by approximating the cosine term in Eq. (2.5) by the first two terms of its Taylor series expansion $\cos(\phi_t - \theta) = 1 - (\phi_t - \theta)^2/2$ and integrating.

Normal model

As described in Rowe and Logan (2004), the Ricean distribution in Eq. (2.5) can be approximated by the normal distribution

$$p(r_t|x_t, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}\left[r_t - (x_t'\beta)\right]^2 \right\}, \tag{2.8}$$

at high SNR. This can also be seen from Eq. (2.7) by approximating the leading square root factor by unity. Then the likelihood function

$$p(r|X, \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2}(r - X\beta)'(r - X\beta) \right\},$$

is formed, from which MLEs and CRLBs can be computed.

Statistics

For each of the three models, MLE statistics can be found under both the restricted null and unrestricted alternative hypotheses. Unbiased versions of the MLEs and activation statistics can be found from a generalized likelihood ratio test. Subscripts will be used to associate statistical quantities with the three models introduced in the previous section: N and T for the Normal and Taylor series approximations to the Ricean distribution, respectively, and C for the complex data bivariate normal distribution. The MLEs for the three models under the unrestricted alternative and restricted null hypotheses denoted with hats and tildes respectively are

Complex model

$$\hat{\theta}_C = \frac{1}{2}\tan^{-1}\left[\frac{\hat{\beta}_R'(X'X)\hat{\beta}_I}{\left(\hat{\beta}_R'(X'X)\hat{\beta}_R - \hat{\beta}_I'(X'X)\hat{\beta}_I\right)/2}\right] \qquad \tilde{\theta}_C = \frac{1}{2}\tan^{-1}\left[\frac{\hat{\beta}_R'\Psi(X'X)\hat{\beta}_I}{\left(\hat{\beta}_R'\Psi(X'X)\hat{\beta}_R - \hat{\beta}_I'\Psi(X'X)\hat{\beta}_I\right)/2}\right]$$

$$\hat{\beta}_C = \hat{\beta}_R\cos\hat{\theta}_C + \hat{\beta}_I\sin\hat{\theta}_C \qquad\qquad \tilde{\beta}_C = \Psi\left(\hat{\beta}_R\cos\tilde{\theta}_C + \hat{\beta}_I\sin\tilde{\theta}_C\right)$$

$$\hat{\sigma}_C^2 = \frac{1}{2n}\begin{pmatrix} y_R - X\hat{\beta}\cos\hat{\theta}_C \\ y_I - X\hat{\beta}\sin\hat{\theta}_C \end{pmatrix}'\begin{pmatrix} y_R - X\hat{\beta}\cos\hat{\theta}_C \\ y_I - X\hat{\beta}\sin\hat{\theta}_C \end{pmatrix} \qquad \tilde{\sigma}_C^2 = \frac{1}{2n}\begin{pmatrix} y_R - X\tilde{\beta}\cos\tilde{\theta}_C \\ y_I - X\tilde{\beta}\sin\tilde{\theta}_C \end{pmatrix}'\begin{pmatrix} y_R - X\tilde{\beta}\cos\tilde{\theta}_C \\ y_I - X\tilde{\beta}\sin\tilde{\theta}_C \end{pmatrix} \tag{3.1}$$

Taylor model

$$\hat{\beta}_T = \hat{\beta}_N - \frac{\tilde{\sigma}_T^2}{2}(X'X)^{-1}\sum_{t=1}^{n} x_t/\left(x_t'\hat{\beta}_T\right) \qquad\qquad \tilde{\beta}_T = \tilde{\beta}_N - \Psi\frac{\tilde{\sigma}_T^2}{2}(X'X)^{-1}\sum_{t=1}^{n} x_t/\left(x_t'\tilde{\beta}_T\right)$$

$$\hat{\sigma}_T^2 = \frac{1}{n}\left(r - X\hat{\beta}_T\right)'\left(r - X\hat{\beta}_T\right) \qquad\qquad \tilde{\sigma}_T^2 = \frac{1}{n}(r - X\tilde{\beta}_T)'(r - X\tilde{\beta}_T) \tag{3.2}$$

Normal model

$$\hat{\beta}_N = (X'X)^{-1}X'r \qquad\qquad\qquad \tilde{\beta}_N = \Psi(X'X)^{-1}X'r$$

$$\hat{\sigma}_N^2 = \frac{1}{n}\left(r - X\hat{\beta}_N\right)'\left(r - X\hat{\beta}_N\right) \qquad \tilde{\sigma}_N^2 = \frac{1}{n}(r - X\tilde{\beta}_N)'(r - X\tilde{\beta}_N) \tag{3.3}$$

where $\Psi = I_{q+1} - (X'X)^{-1} C'[C(X'X)^{-1}C']^{-1}C$, $\hat{\beta}_R = (X'X)^{-1} X'y_R$, and $\hat{\beta}_I = (X'X)^{-1} X'y_I$, while y_R and y_I are the $n \times 1$ vectors of real and imaginary observations. The parameters for the approximate Taylor series model are iteratively estimated (Rowe, 2001, 2003).

Unrestricted alternative hypothesis estimators will be evaluated in terms of the deviation of their mean and variance from the true value and their CRLB. An estimator τ of θ is said to be unbiased if $E(\tau) = \theta$ for all θ (Hogg and Craig, 1978). The deviation of the mean of an estimator from the true value is called the bias, $b(\tau) = E(\tau) - \theta$. However, since the CRLB is for unbiased parameter estimators, the unrestricted alternative hypothesis estimators for the variance are multiplied by $n/(n - q - 1)$ for the magnitude-only models and $2n/(2n - q - 2)$ for the complex model.

As outlined in Appendix B, the CRLB for the complex data activation model of Rowe and Logan (2004) can be found to be

$$CRLB_C = \begin{array}{c} \beta \\ \sigma^2 \\ \theta \end{array} \begin{array}{ccc} \beta & \sigma^2 & \theta \\ \left[\begin{array}{ccc} \sigma^2(X'X)^{-1} & 0 & 0 \\ 0 & \sigma^4/n & 0 \\ 0 & 0 & \sigma^2/\beta'\,(X'X)\beta \end{array} \right], \end{array} \tag{3.4}$$

for the magnitude-only data Taylor series approximation to the Ricean distribution model is found to be

$$CRLB_T = \begin{array}{c} \beta \\ \sigma^2 \end{array} \left[\begin{array}{cc} \sigma^2(X'X)^{-1}\left[I - \frac{\hat{\sigma}^2}{2}(X'X)^{-1}\sum_{t=1}^{n} x_t/(x'_t\beta)\right]^{-1} & 0 \\ 0 & 2\sigma^4/n \end{array} \right], \tag{3.5}$$

and for the magnitude-only data normal activation model is found to be

$$CRLB_N = \begin{array}{c} \beta \\ \sigma^2 \end{array} \begin{array}{cc} \beta & \sigma^2 \\ \left[\begin{array}{cc} \sigma^2(X'X)^{-1} & 0 \\ 0 & 2\sigma^4/n \end{array} \right]. \end{array} \tag{3.6}$$

Note that the CRLB for the variance of an unbiased estimate of the observation variance is two times larger in the magnitude-only data models than in the complex data model. On average, the mean of the variance of an estimator may not achieve the CRLB for finite sample sizes. If the variance of an unbiased parameter estimator achieves the CRLB for any sample size, then the estimator is said to be efficient, while if it achieves the CRLB asymptotically as the sample size increases, then it is said to be asymptotically efficient. For comparability between the three models, activation statistics are taken to be $-2\log\lambda$ with λ being the likelihood ratio statistic described in Appendix A.

Simulation

Data is generated to simulate voxel activation from a bilateral finger tapping fMRI block design experiment as in Rowe and Logan (2004). The simulation consisted of $n = 256$ points where the true activation structure is known so that the three models can be evaluated.

Simulated fMRI data is constructed according to a general non-linear regression model which consists of an intercept β_0, a coefficient β_1 for a time trend t for all voxels and a coefficient β_2 for a reference function x_{2t} related to a block experimental design where $x_t = (1, x_{1t}, x_{2t})'$. This model dictates that at time t,

$$y_t = [(\beta_0 + \beta_1 t + \beta_2 x_{2t})\cos\theta + \eta_{Rt}]$$
$$+ i[(\beta_0 + \beta_1 t + \beta_2 x_{2t})\sin\theta + \eta_{It}],$$

where η_{Rt} and η_{It} are independent identically distributed normal random variates with mean zero and variance σ^2 simply denoted i.i.d. N(0, σ^2).

For all voxels in this simulation study, the phase was generically selected to be $\theta = \pi/6$, while $\beta_1 = 0.00001$, and $\sigma = 0.04909$ which are values taken from a "highly active" voxel (Rowe and Logan, 2004). The coefficient for the reference function β_2 has a value determined by a contrast-to-noise ratio (CNR = β_2/σ) of 1/2. Therefore, since the variance is held fixed, the SNR is parameterized by varying β_0 so that the ratio SNR = β_0/σ takes on values 1, 2.5, 5, 7.5, 10, 12.5, and 15. For each SNR

combination, 10^6 simulated voxel time courses were generated and the parameters estimated by all three models.

The means (Fig. 1) and variances (Fig. 2) of the unbiased parameter estimates and activation statistics from the simulations were computed for the various SNRs. Note that the mean and variance of the parameter estimates in the magnitude-only data normal approximate model (in red) and Taylor series approximate model (in cyan) deviate from the true value (in magenta) as the SNR decreases while the parameter estimates in the complex data model (in blue) achieve their correct value (in magenta) and remain fairly constant. Also included in Fig. 1 are the true parameter values and in Fig. 2 the CRLBs for the magnitude-only data approximate normal model, the magnitude-only data Taylor series approximate model, and complex data model, in green, yellow, and magenta, respectively. Since the true mean values for the parameters are the same for the three models considered, only the true value curve for the complex data model in magenta is shown in Fig. 1. In panel (e) of Figs. 1 and 2, the approximate normal and Taylor magnitude-only data models do not include a phase parameter and thus curves are not included. In panel (f) of Figs. 1 and 2, the parameter and activation statistic means or variances are not analytically known for the three models and thus curves are not included.

It can be seen in Fig. 1a that the estimated intercept coefficient or baseline β_0 appears to be unbiased for the magnitude-only data normal model (in red) to about an SNR of 7.5 and the magnitude-only data Taylor series approximate model (in cyan) down to an SNR of about 2.5 but is unbiased for the complex data model (in blue) for all SNRs considered.
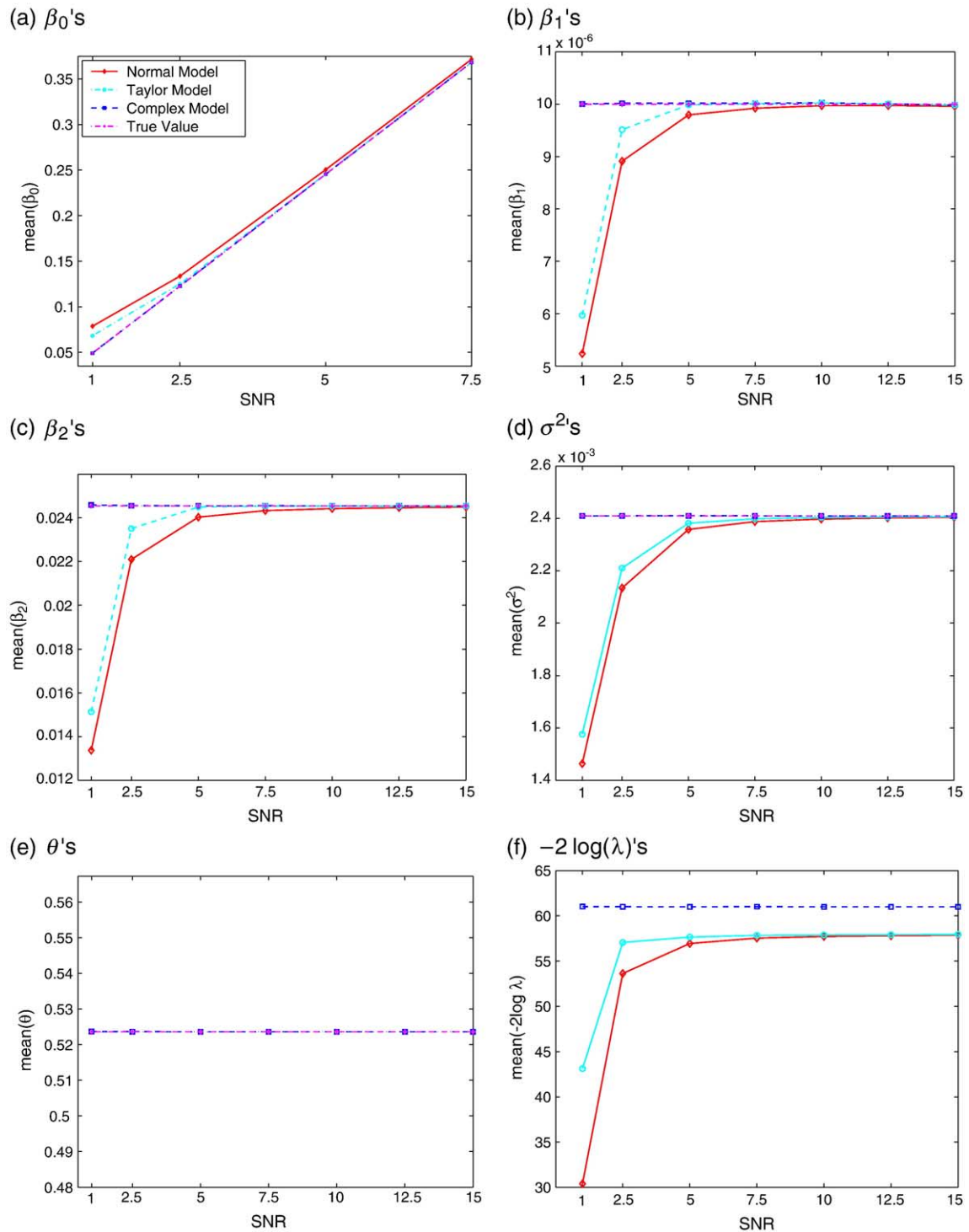
Fig. 1. Plot of estimated parameter means with varying SNR.

In Fig. 1b, the coefficient β_1 for the linear trend appears to be unbiased for the magnitude-only data normal model to about an SNR of 10 and the magnitude-only data Taylor model down to an SNR of about 5 but unbiased for the complex data model for all SNRs considered. Additionally, it can be seen in Fig. 1c that the reference function coefficient β_2 appears to be unbiased for the magnitude-only data normal model to about an SNR of 10 and the magnitude-only data Taylor model down to an SNR

of about 5 but unbiased for the complex data model for all SNRs considered. In Fig. 1d, the variance σ^2 appears to be unbiased for the magnitude-only data normal model to about an SNR of 10 and the magnitude-only data Taylor model down to an SNR of about 7.5 but unbiased for the complex data model for all SNRs considered. In Fig. 1e, the phase imperfection θ only present in the complex data model appears to be unbiased for all SNRs considered. In Fig. 1f, the activation statistic
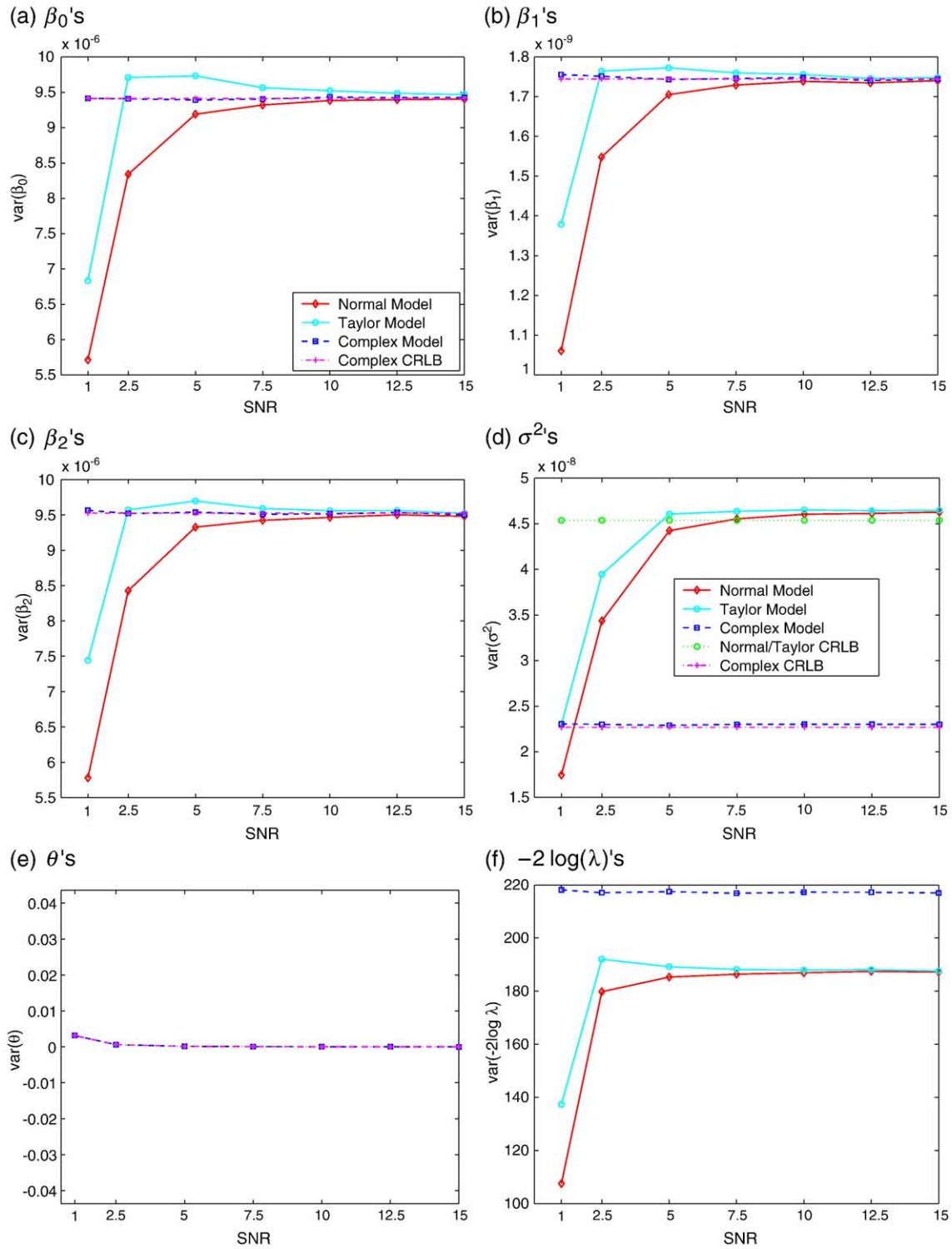
Fig. 2. Plot of estimated parameter variances with varying SNR.

$-2\log(\lambda)$ is uniformly lower for the two magnitude-only data models than the complex data model however it appears to be unbiased for the magnitude-only data normal model to about an SNR of 10 and the magnitude-only data Taylor model down to an SNR of about 7.5 but it is uniformly higher and appears to be unbiased for the complex data. Reiterating the results in this figure, the complex data model produces higher activation

statistics or a better ability to detect activation than the magnitude-only data models that does not decrease with SNR. In general, the parameter estimates for the complex data model appear to have less bias than the magnitude-only data models for all SNRs considered.

It can be seen in Fig. 2a that the estimated intercept coefficient or baseline β_0 appears to be efficient for the

magnitude-only data normal model (in red) to about an SNR of 10, is above the CRLB for the magnitude-only data Taylor model (in cyan) down to an SNR of about 2.5, then decreases rapidly when the SNR is decreased to 1. The unbiased value for the complex data model (in blue) achieves the CRLB for all SNRs considered. In Fig. 2b, the variance of the coefficient β_1 for the linear trend appears to be efficient for the magnitude-only data normal (in red) and Taylor (in cyan) models to about an SNR of 10. The unbiased value for the complex data model (in blue) achieves the CRLB for all SNRs considered. Additionally, it can be seen in Fig. 2c that the variance of the coefficient β_2 for the linear trend appears to be efficient for the magnitude-only data normal (in red) and Taylor (in cyan) models to about an SNR of 10. The unbiased value for the complex data model (in blue) achieves the CRLB for all SNRs considered.

In Fig. 2d, the variance of the variance σ^2 does not appear to be efficient or achieve the CRLB for the three models considered, the magnitude-only data normal model (green), the magnitude-only data Taylor model (omitted because it is not pictorially different from the normal model CRLB), or the complex data model (in magenta).

However, there is a value that each model achieves for high SNR which may tend to the CRLB or be asymptotically efficient at a larger sample size. The unbiased estimator for σ^2 appears to achieve its asymptotic CRLB value for the magnitude-only data normal model (in red) at the largest SNR value considered of 15 then decrease with SNR while the magnitude-only data Taylor model (in cyan) achieves its asymptotic CRLB value down to an SNR of about 7.5 then decrease with SNR. The unbiased estimator for σ^2 achieves its asymptotic CRLB value for the complex data model (in blue) for all SNRs considered. Also note that the error variance estimates are approximately twice as large for the magnitude-only data models than for the complex data model, even for the largest value of SNR. This factor of two disparity is as stated by the CRLBs. In Fig. 2e, the variance of the unbiased estimator for the phase imperfection θ only present in the complex data model (in blue) appears to be efficient for all SNRs considered and achieve its CRLB (in magenta) for all SNRs considered. In Fig. 2f, the variance of the activation statistic $-2\log(\lambda)$ is smaller for the two magnitude-only data models as compared to the complex data model but this is about a lower mean activation statistic. However, the variance of the activation statistic appears to be constant for the magnitude-only data normal model (in red) to about an SNR of 7.5 and the magnitude-only data Taylor model (in cyan) down to an SNR of about 5. The variance of the activation statistic is slightly higher but is constant for the complex data model for all SNRs considered. In general, the parameter estimates for the complex data model appear to be more efficient than the magnitude-only data models for all SNRs considered.

Conclusions

A recently proposed complex data fMRI model by Rowe and Logan (2004) as an alternative to the typical magnitude-only data normal data model was outlined along with a model that uses a Taylor series approximation in the Ricean

distribution. Maximum likelihood parameter estimates and Cramer–Rao lower bounds for the parameters of each model were also described. The CRLB for the variance of the observation variance was half as large in the complex data model as it is in the two magnitude-only data models.

Simulations were performed for several SNRs and the parameters for each model estimated along with an activation statistic. The mean and variance of the estimated parameter values and activation statistics were computed and compared with the true values and CRLBs where applicable.

It was found that the complex model performed extremely well at estimating the true parameter values and achieving its CRLBs even for very low SNR. The magnitude-only data models did not perform as well as the complex model. Additionally, even for very large SNRs, the variance of the error variance was twice as large for the magnitude-only models as for the complex model. These results indicate that using the complex data model instead of approximations to the Ricean distribution of the magnitude-only data are more useful at low SNR and yield higher activation statistics regardless of SNR. Since the SNR in a voxel is not known a priori and the models are equivalent at high SNR, the complex data model is recommended over the magnitude-only data models.

Appendix A. MLEs for Taylor model

The logarithm of the likelihood for the Taylor series approximation of the Ricean distribution of magnitude-only data observations is

$$
LL = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\sum_{t=1}^{n}\log r_t - \frac{n}{2}\log(\sigma^2)
$$
$$
- \frac{1}{2}\sum_{t=1}^{n}\log(x_t'\beta) - \frac{1}{2\sigma^2}(r - X\beta)'(r - X\beta).
$$

Unrestricted alternative hypothesis MLEs

Unrestricted alternative hypothesis MLEs will be found by maximizing the unrestricted logarithm of the likelihood given in Eq. (A.1) with respect to the parameters and yields

$$
\frac{\partial LL}{\partial \beta} = -\frac{1}{2\sigma^2}\left[2(X'X)\beta - 2X'r\right] - \frac{1}{2}\sum_{t=1}^{n}\frac{1}{x_t'\beta}x_t
$$

$$
\frac{\partial LL}{\partial \sigma^2} - \frac{2n}{2}\frac{1}{\sigma^2} + \frac{h}{2}\frac{1}{(\sigma^2)^2}
$$

where $h = (r - X\beta)'(r - X\beta)$. By setting these derivatives equal to zero and solving, we get the alternative hypothesis MLEs under the unrestricted model given in Eq. (3.2).

Restricted null hypothesis MLEs

Restricted null hypothesis MLEs will be found by maximizing the logarithm of the likelihood in Eq. (A.1) with respect to the

parameters with the Lagrange multiplier term $\psi'(C\beta - \gamma)$ added for the alternative hypothesis restriction and yields

$$\frac{\partial LL}{\partial \beta} = -\frac{1}{2\sigma^2}\left[2(X'X)\beta - 2X'r\right] - \frac{1}{2}\sum_{t=1}^{n}\frac{1}{x_t'\beta}x_t - C'\psi$$

$$\frac{\partial LL}{\partial \psi} = (C\beta - \gamma)$$

$$\frac{\partial LL}{\partial \sigma^2} - \frac{2n}{2}\frac{1}{\sigma^2} + \frac{h}{2}\frac{1}{(\sigma^2)^2}.$$

By setting these derivatives equal to zero and solving, we get the null hypothesis MLEs under the restricted model given in Eq. (3.2).

Note that the maximum likelihood variance estimators are $\hat{\sigma}^2 = \hat{h}/(2n)$ and $\tilde{\sigma}^2 = \tilde{h}/(2n)$. Then the generalized likelihood ratio statistic is

$$\lambda = \frac{p(r|\tilde{\beta}, \tilde{\sigma}^2, \tilde{\theta}, X)}{p(r|\hat{\beta}, \hat{\sigma}^2, \hat{\theta}, X)}$$

$$= \frac{(\tilde{\sigma}^2)^{-2n/2}e^{-2\tilde{h}n/(2\tilde{h})}}{(\hat{\sigma}^2)^{-2n/2}e^{-2\hat{h}n/(2\hat{h})}}$$

$$= (\hat{h}/\tilde{h})^n. \tag{A.2}$$

This is the ratio of residual sums of squares for the alternative and null hypotheses raised to the nth power. It should be noted that the likelihood ratio statistic is bounded, $0 \le \lambda \le 1$ because the residual sum of squares will always be smaller for the unrestricted alternative hypothesis. For example, say that we have a regression model with two regressors besides the mean, so that $\beta = (\beta_0, \beta_1, \beta_2)'$, and would like to test to see if $\beta_2 = 0$. The null hypothesis can be described with $C = (0,0,1)$ and $\gamma = 0$. Under the null hypothesis, we essentially estimate the model parameters with β_2 fixed to be zero and under the alternative hypothesis estimate the parameters with β_2 unrestricted or free to take on any value. If the unrestricted estimate of β_2 were zero, then the experimental results tend to confirm the null hypothesis and the likelihood ratio statistic is unity. If the unrestricted estimate of β_2 deviated considerably from zero, then the experimental results tend to confirm the alternative hypothesis and the likelihood ratio statistic is "close" to zero. So if λ were used as a test criterion, then a natural procedure is to reject the null hypothesis if λ were less than some critical value λ_0 that is statistically determined. Often some algebra can be performed and the likelihood ratio statistic reduced to a familiar t of F statistic. Occasionally, a useful simplification cannot be found; however, it has been noted that under very general conditions, $-2\log\lambda$ is approximately χ^2 distributed in large samples (Wald, 1943).

Appendix B. Cramer–Rao lower bounds

The CRLB for the variance of an unbiased estimate of a model parameter requires the second derivatives of the log likelihoods LL with respect to the model parameters.

Complex model

The second derivatives of the log likelihood for the complex data model are

$$H_C(1,1) = \frac{\partial^2 LL}{\partial \beta \partial \beta'} = -\frac{1}{2\sigma^2}\left[2(X'X)\right]$$

$$H_C(2,2) = \frac{\partial^2 LL}{\partial \theta^2} = -\frac{1}{2\sigma^2}\left[-2\beta'(-X'y_R\cos\theta - X'y_I\sin\theta)\right]$$

$$H_C(3,3) = \frac{\partial^2 LL}{\partial(\sigma^2)^2} = -\frac{2n}{2}(-1)(\sigma^2)^{-2} + \frac{h}{2}(-2)(\sigma^2)^{-3}$$

$$H_C(1,2) = \frac{\partial^2 LL}{\partial \beta \partial \theta} = -\frac{1}{\sigma^2}\left[X'y_R\sin\theta - X'y_I\cos\theta\right]$$

$$H_C(2,1) = \frac{\partial^2 LL}{\partial \theta \partial \beta'} = H_C(1,2)'$$

$$H_C(1,3) = \frac{\partial^2 LL}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4}\left[X'y_R\cos\theta - X'y_I\sin\theta - (X'X)\beta\right]$$

$$H_C(3,1) = \frac{\partial^2 LL}{\partial \sigma^2 \partial \beta'} = H_C(1,3)'$$

$$H_C(2,3) = \frac{\partial^2 LL}{\partial \theta \partial \sigma^2} = -\frac{1}{\sigma^4}\beta'X'[y_R\sin\theta - y_I\cos\theta]$$

$$H_C(3,2) = \frac{\partial^2 LL}{\partial \sigma^2 \partial \theta} = H_C(2,3).$$

Taylor model

The second derivatives of the log likelihood for the Taylor approximate model are

$$H_T(1,1) = \frac{\partial^2 LL}{\partial \beta \partial \beta'} = -\frac{1}{2\sigma^2}\left[2(X'X)\right]$$
$$- \frac{1}{2}\sum_{t=1}^{n}x_t(-1)(x_t'\beta)^{-2}x_t'$$

$$H_T(2,2) = \frac{\partial^2 LL}{\partial(\sigma^2)^2} = -\frac{n}{2}(-1)(\sigma^2)^{-2} - 2\frac{h}{2}(\sigma^2)^{-3}$$

$$H_T(1,2) = \frac{\partial^2 LL}{\partial \beta \partial \sigma^2} = \left[X'r - (X'X)\beta\right](-1)(\sigma^2)^{-2}$$

$$H_T(2,1) = \frac{\partial^2 LL}{\partial \sigma^2 \partial \beta'} = H_T(1,2)'.$$

Normal model

The second derivatives of the log likelihood for the normal approximate model are

$$H_N(1,1) = \frac{\partial^2 LL}{\partial \beta \partial \beta'} = -\frac{1}{2\sigma^2}\left[2(X'X)\right]$$

$$H_N(2,2) = \frac{\partial^2 LL}{\partial(\sigma^2)^2} = -\frac{n}{2}(-1)(\sigma^2)^{-2} - 2\frac{h}{2}(\sigma^2)^{-3}$$

$$H_N(1,2) = \frac{\partial^2 LL}{\partial \beta \partial \sigma^2} = \left[X'r - (X'X)\beta\right](-1)(\sigma^2)^{-2}$$

$$H_N(2,1) = \frac{\partial^2 LL}{\partial \sigma^2 \partial \beta'} = H_N(1,2)'.$$

The matrix of CRLBs is the inverse of the Fisher information matrix. By taking expectations of the block elements of the Hessian matrices, the CRLBs in Eqs. (3.4) and (3.5) are found. The symmetric Hessian matrices generically denoted H are formed from the second derivatives which are seen to be negative definite and therefore the estimated values from the first derivatives are maxima and not minima. The Fisher information matrices generically denoted I are $-E(H|\beta,\sigma^2)$ for the normal and Taylor models, that is, the expectation of the Hessian matrices with respect to r given β and σ^2. The Fisher information matrix I_C is $-E(H_C|\beta,\theta,\sigma^2)$, that is, the expectation of the Hessian matrix H_C with respect to y_R and y_I given β, θ, and σ^2.

References

Bandettini, P., Jesmanowicz, A., Wong, E., Hyde, J.S., 1993. Processing strategies for time-course data sets in functional MRI of the human brain. Magn. Reson. Med. 30 (2), 161–173.

Cox, R.W., Jesmanowicz, A., Hyde, J.S., 1995. Real-time functional magnetic resonance imaging. Magn. Reson. Med. 33 (2), 230–236.

Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy data. Magn. Reson. Med. 34 (6), 910–914.

Hogg, R.V., Craig, A.T., 1978. Introduction to Mathematical Statistics, fourth ed. Macmillan, New York, NY, USA.

Lai, S., Glover, G.H., 1997. Detection of BOLD fMRI signals using complex data. Proc. ISMRM, p. 1671.

Logan, B.R., Rowe, D.B., 2004. An evaluation of thresholding techniques in fMRI analysis. NeuroImage 22 (1), 95–108.

Nan, F.Y., Nowak, R.D., 1999. Generalized likelihood ratio detection for fMRI using complex data. IEEE Trans. Med. Imaging 18 (4), 320–329.

Rice, S.O., 1944. Mathematical analysis of random noise. Bell Syst. Tech. 23, 282 (Reprinted by N. Wax, Selected papers on Noise and Stochastic Process, Dover Publication, 1954. QA273W3).

Rowe, D.B., 2001. Bayesian source separation for reference function determination in fMRI. Magn. Reson. Med. 45 (5), 374–378.

Rowe, D.B., 2003. Multivariate Bayesian Statistics. CRC Press, Boca Raton, FL, USA.

Rowe, D.B., Logan, B.R., 2004. A complex way to compute fMRI activation. NeuroImage 23 (3), 1078–1092.

Rowe, D.B., Logan, B.R., 2005. Complex fMRI analysis with unrestricted phase is equivalent to a magnitude-only model. NeuroImage 24 (2), 603–606.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observation is large. Trans. Am. Math. Soc. 54, 426–483.