How Accurately Can Parameters from Exponential Models Be Estimated? A Bayesian View

G. LARRY BRETTHORST

Department of Radiology, Washington University, One Brookings Drive, St. Louis, MO 63130

ABSTRACT: Estimating the amplitudes and decay rate constants of exponentially decaying signals is an important problem in NMR. Understanding how the uncertainty in the parameter estimates depends on the data acquisition parameters and on the "true" but unknown values of the exponential signal parameters is an important step in designing experiments and determining the amount and quality of the data that must be gathered to make good parameter estimates. In this article, Bayesian probability theory is applied to this problem. Explicit relationships between the data acquisition parameters and the "true" but unknown exponential signal parameters are derived for the cases of data containing one and two exponential signal components. Because uniform prior probabilities are purposely employed, the results are broadly applicable to experimental parameter estimation. © 2005 Wiley Periodicals, Inc. Concepts Magn Reson Part A 27A: 73–83, 2005

KEY WORDS: Bayesian probability theory; exponential parameter estimation

INTRODUCTION

Exponential signals occur in many facets of magnetic resonance and other physical and biological science arenas. Accurate estimation of the parameters characterizing exponential signal models is difficult. Because the problem is important and of broad impact, the literature is replete with reports describing various approaches to exponential parameter estimation. Though a comprehensive survey is beyond the scope of this manuscript, the interested reader is referred to a number of recent reviews and references therein (*1–4*). Missing from this literature is the formulation of (i) specific Bayesian parameter estimators for multiple parameter exponential signal models and (ii) informative concise closed-form expressions for the uncertainties in the parameter estimates; uncertainty

Received 14 June 2005; revised 18 August 2005; accepted 18 August 2005

Correspondence to: G. Larry Bretthorst; E-mail: gbretthorst@wustl.edu

Concepts in Magnetic Resonance Part A, Vol. 27A(2) 73–83 (2005)

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/cmr.a.20044

expressions that offer insight and guidance to the experimentalist. This article seeks to fill this gap through application of Bayesian probability theory. Specifically, we derive the Bayesian parameter estimators for multiple exponential signal models and concise insightful expressions for the uncertainty in the Bayesian parameter estimates. Because uniform prior probabilities are purposely employed, the results are broadly applicable to experimental parameter estimation. Traditionally, this problem is solved using the Cramer Rao lower bound. A brief discussion comparing our Bayesian approach with the Cramer Rao lower bound is provided later in this article.

Bayesian probability theory (*5–8*) is a powerful tool for parameter estimation (*9–22*) and can be used to estimate parameters in situations in which most other techniques fail. For example, the discrete Fourier transform fails when estimating frequencies in nonuniformly sampled data, yet probability theory generalizes the discrete Fourier transform and gives good parameter estimates in this situation (*20*). Probability theory can also be used for model selection (*23*, *24*), and this application is rapidly becoming one of its major uses in our laboratory. Just as important, probability theory can be used to derive the dependence of the estimated parameters on the data acquisition parameters and on the “true” but unknown signal parameters (*9*, *25*). In this article, we apply probability theory as extended logic in this latter capacity and explicitly demonstrate how the multiple parameter estimates—the estimated amplitudes and decay rate constants—depend on the “true” signal parameters, signal-to-noise ratio, data sampling interval, and total number of data values in two scenarios: data containing one and two exponential signal components.

To do these calculations, we first apply Bayesian probability theory to derive the posterior probability for the parameter to be estimated. Next, we postulate a functional form for the data. Finally, using the posterior probability and the functional form for the data, we make an estimate (mean ± standard deviation) for the parameters appearing in the exponential model. Because we are using Bayesian probability theory, there is a different, unique, calculation for each parameter in each model considered. We consider models containing one and two exponential signals, so there are a total of six different calculations, two calculations for the single exponential and four for the biexponential, that must be performed. Fortunately, in the biexponential case, the posterior probabilities are symmetric under exchange of labels on the exponential components. Because of this symmetry, we can derive the parameter estimates for the second component by a simple relabeling of the estimates from the first component. Nonetheless, there are four unique calculations to perform. For the single or monoexponential model, these calculations are straightforward and described in detail in Appendix A. The calculations using the two or biexponential model are lengthy and are outlined in Appendix B. In the following section, we summarize the results from these calculations.

SUMMARY OF RESULTS

In deriving these results, three approximations were made: high signal-to-noise ratio data (so that the projection of the model onto the noise can be ignored), uniform data sampling (so that certain sums can be approximated), and signal decays to three or more e-foldings during the data sampling interval (so that the calculations give simple, intuitive results). For data containing a single exponential, the estimate (mean ± standard deviation) of the decay rate constant, α is given by

$$(\alpha)_{est} = \hat{\alpha} \pm \frac{\sigma}{\hat{A}} \sqrt{8\hat{\alpha}^3 \Delta T} \qquad [1]$$

where $\hat{A}$ and $\hat{\alpha}$ are the “true” amplitude and decay rate constant in the data, ΔT is the sampling interval between two consecutive data points and σ is the standard deviation of the prior probability assigned to represent what is known about the noise and is assumed known and equal to the true standard deviation of the noise. Similarly, the estimated value of the amplitude, A, is given by

$$(A)_{est} = \hat{A} \pm \sigma \sqrt{2\hat{\alpha}\Delta T}. \qquad [2]$$

In both cases, as the noise goes to zero, the parameter estimates go smoothly to the “true” values of the parameters. Each parameter estimate depends on the sampling interval. The sampling interval is the total data acquisition time divided by the number of data values, N, so the precision of each estimate improves by the square root of N criteria. In addition, the uncertainty of each estimate depends on the “true” decay rate constant of the signal. Consequently, the more rapidly the signal decays, the poorer the estimates become for the amplitude and decay rate constant. However, the dependence on the “true” decay rate constant is much stronger for the uncertainty in the estimated decay rate constant than for the estimated amplitude. For the estimated decay rate constant, one can increase the precision of the estimate in two ways: increase the signal-to-noise ratio or acquire more data during the time period in which the signal is present (i.e., make ΔT smaller). For the estimated amplitude, a similar result holds for acquiring more data but not for increasing the signal-to-noise ratio. To reduce the uncertainty in the estimated amplitudes, the noise level must be reduced; the relative or fractional uncertainty depends on the signal-to-noise ratio but the uncertainty in the amplitude estimate does not depend on the "true" amplitude of the signal.

For data containing two exponential signal components that both decay by at least three or more e-foldings during the data acquisition period, the estimates (mean ± standard deviation) are given by

$$(\alpha)_{est} = \hat{\alpha} \pm \frac{\sigma}{\hat{A}}\left[\frac{\hat{\alpha}+\hat{\beta}}{\hat{\alpha}-\hat{\beta}}\right]^2 \sqrt{8\hat{\alpha}^3\Delta T} \qquad [3]$$

for the estimated decay rate constant, and

$$(A)_{est} = \hat{A} \pm \sigma\left|\frac{\hat{\alpha}+\hat{\beta}}{\hat{\alpha}-\hat{\beta}}\right| \sqrt{2\hat{\alpha}\Delta T} \qquad [4]$$

for the estimated amplitude, where $\hat{A}$ and $\hat{\alpha}$ are the "true" amplitude and decay rate constant of one of the exponential signal components, and $\hat{B}$ and $\hat{\beta}$ are the "true" amplitude and decay rate constant of the other signal component. The parameter estimates for $\hat{B}$ and $\hat{\beta}$ are obtained by exchanging the role of the two exponentials. Except for the bracketed ratio, which we call the interaction ratio, these estimates are identical to the single exponential case. Consequently, all of the previous comments on how to improve the parameter estimates apply to the biexponential case. The interaction ratio tells us how the presence of the additional exponential component interferes with the estimation process. If the "true" decay rate constants are as close as a factor of 2, the uncertainty in the estimated decay rate constants are a factor of 9 larger than the single exponential case; whereas the uncertainty in the amplitude estimates are a factor of 3 larger.

DISCUSSION

We illustrate the results of this analysis using numerical examples that replicate exactly the conditions under which these formulas were derived. In these examples, we generate several data sets, apply these formulas, and then compare the results with the appropriate posterior probabilities. These formulas and the posterior probabilities assume that the standard deviation of the noise is known. Consequently, we assume σ is known and arbitrarily set σ = 0.1 in these examples. The amplitude of the exponential signal components will be 100 throughout most of the examples, so the signal-to-noise ratio of each component is 1000:1.

The exact results, the posterior probabilities, are obtained by computing the marginal posterior probability for each of the parameters of interest. In these examples, we apply the formulas for the estimated decay rate constants. We do not apply the formulas for the estimated amplitudes. However, if the formulas for the decay rate constants behave as predicted, the formulas for the amplitudes will also behave as predicted. In data containing a single exponential, Eq. [19] in the appendix was used to compute the posterior probability for the decay rate constant. For biexponential data, no closed-form solution exists for the marginal posterior probability for one of the decay rate constants. To compute this marginal posterior probability, we first computed the joint posterior probability for the decay rate constants, Eq. [37] in the appendix, on a rectangular grid. Next, we normalized this joint probability and then projected the two-dimensional grid onto each axis, i.e., we numerically integrated the joint posterior probability to obtain the needed marginal probability. The joint distribution was evaluated on a fine grid to make the error in the integrals about one part in 1,000.

First, we illustrate that the formulas correctly characterize the estimation process. Suppose we have a single exponential data set given by

$$d(t_i) = 100\exp\{-4t_i\} \qquad [5]$$

with sampling interval 0.005 seconds and 200 data values. Under these conditions, Eq. [1] predicts that the decay rate constant will be estimated to be

$$(\alpha_1)_{est} = 4 \pm \frac{0.1}{100}\sqrt{8\times 64\times 0.005} = 4 \pm 0.0016. \qquad [6]$$

Next we compare this to the posterior probability for the decay rate constant, shown as the sharp peak in Fig. 1(A). When we computed this posterior probability, we also computed its mean and standard deviation, obtaining

$$(\alpha_1)_{est} = 4.000 \pm 0.0016; \qquad [7]$$

a result identical to the predicted mean and standard deviation.

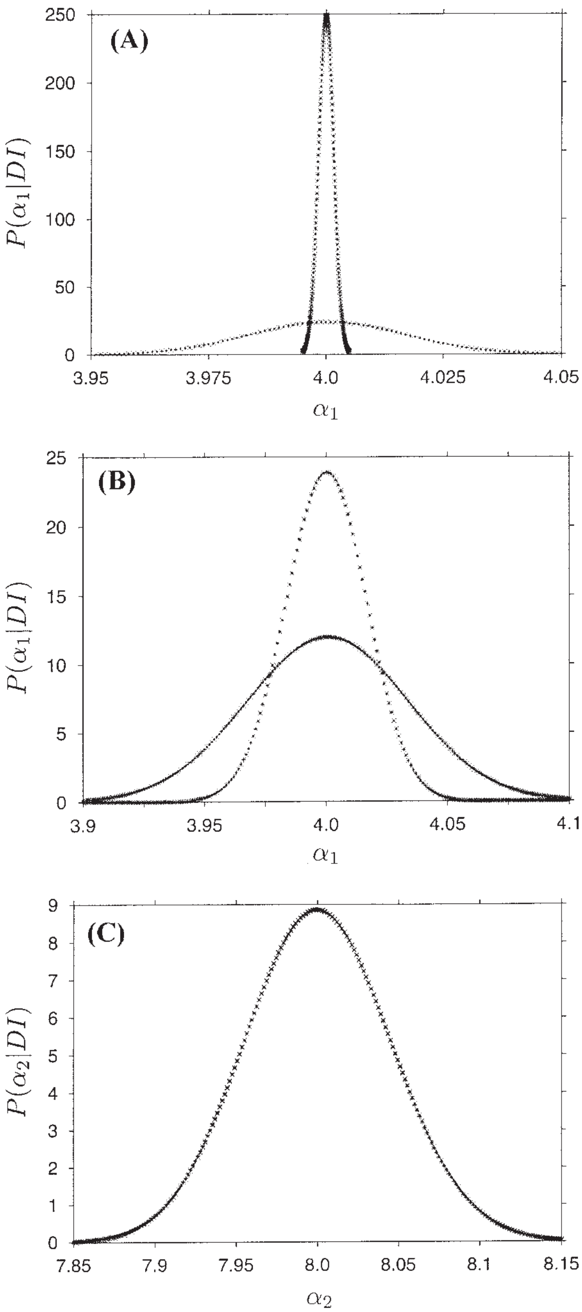

Figure 1 Estimating the decay rate constant. Panel (A) is $P(\alpha_1|D\ I)$ when the rapidly decaying second component is not present, narrow feature, and when it is present, broad feature. The change in width is roughly a factor of ten; the predicted factor is 9. Panel (B) is $P(\alpha_1|D\ I)$ before, narrow feature, and after, broad feature, halving the signal-to-noise ratio of the first component. The change in width is the predicted factor of two. Panel (C) is $P(\alpha_2|D\ I)$ computed before and after halving the first component's signal-to-noise ratio. Note the probability density functions are, as predicted, identical.

Next we illustrate what happens to the estimated decay rate constant when a second exponential signal component is present. In this case, we generated data using

$$d(t_i) = 100 \exp\{-4t_i\} + 100 \exp\{-8t_i\}. \quad [8]$$

The same sampling time and number of data values were generated. Note that the first component is identical to the component used in the previous example. Using Eq. [3], we should be able to estimate the first decay rate constant to

$$(\alpha_1)_{est} = 4 \pm \frac{0.1}{100} 3^2 \sqrt{8 \times 64 \times 0.005} = 4 \pm 0.0144. \quad [9]$$

Using these simulated data, Eq. [8], we computed the marginal posterior probability for the first decay rate constant, shown as the broad curve in Fig. 1(A). An expansion of this posterior probability is shown as the tall curve in panel (B). When we computed this marginal posterior probability density, we also computed its mean and standard deviation,

$$(\alpha_1)_{est} = 4.000 \pm 0.0166, \quad [10]$$

in good agreement with the predicted value. The ratio of standard deviations, $0.0166/0.0016 \approx 10$ is in good agreement with the predicted factor of 9. Note that Fig. 1(A) clearly demonstrates that the mere presence of the second exponential is enough to appreciably widen the posterior probability for the decay rate constant, and thus make the parameter estimates much worse than when only a single exponential is present.

One of the predictions derived from Eq. [3] is that the parameter estimates depend only on the signal-to-noise ratio of the component being estimated and not on the signal-to-noise ratio of the other component. To illustrate this effect, a third data set was analyzed using

$$d(t_i) = 50 \exp\{-4t_i\} + 100 \exp\{-8t_i\} \quad [11]$$

to generate the data.

Note that nothing has changed in these simulated data except that the amplitude of the first exponential has been halved. Uniformly sampled data were generated using the same sampling time and interval. According to Eq. [3], halving this amplitude should result in doubling the width of the posterior probability for the first decay rate constant. We have plotted this posterior probability in Fig. 1(B), the broader of the two curves. Again, when we computed this marginal posterior probability we also computed the mean and standard deviation for the decay rate constant:

$$(\alpha_1)_{est} = 4.000 \pm 0.0329. \qquad [12]$$

Taking the ratio of the standard deviations computed from the posterior probability for the decay rate constants given the data generated from Eq. [11] and Eq. [12], we find 0.0329/0.0166 $\approx$ 1.98, in good agreement with the predicted value of 2.

Finally, note that Eq. [3] predicts that the parameter estimates for the exponential component that we did not change will not change. We can verify this fact if we plot the posterior probability for α_2 using the data generated from Eq. [11] and Eq. [12] respectively, Fig. 1(C). Note that these distributions are not just indistinguishable from one another; they are identical to one another. Thus, the marginal posterior probability for the decay rate constant whose amplitude did not change does not depend on the signal-to-noise ratio of the signal component that did change, exactly as predicted.

CONCLUSIONS

These equations show how the uncertainty in the estimated amplitudes and decay rate constants depends on the experimental parameters. The equations also demonstrate how the addition of a second exponential markedly increases the difficulty of estimating the parameters. These estimates are valid for high signal-to-noise ratio data containing single and biexponentially decaying signal components that decay away during the data acquisition period. They are not valid for truncated signals or for data that contain more than two exponential signal components, nor are they valid for data containing a constant offset.

The problem of determining the uncertainty in the parameters for a single exponential plus a constant would, at first glance, seem to be a trivial extension of the results obtained in this article. Unfortunately, the approximations used in these calculations are not adequate for the single exponential plus a constant problem, and when the problem is solved using better approximations, the resulting expressions are too cumbersome to be useful, i.e., they do not give any physical insight.

In deriving these formulas we used the Laplace approximation, as described in the appendices. The resulting formulas are valid, provided this approximation is valid. There are two regimes in which this approximation breaks down. First, in the limit that one exponential decay rate constant approaches the other; one has a single exponential, not two. Consequently, only the sum of the decay rates constants can be estimated. The joint posterior probability has a peak when the decay rate constants are well separated. However, as they approach each other, this peak becomes a ridge line corresponding to the sum of the decay rates equal to a constant. As a consequence, the Laplace approximation is not valid and the given formulas do not apply. The second circumstance under which the Laplace approximation breaks down is when the amplitude of one of the exponentials is much smaller than the other. In this case, for all practical purposes, the data are single exponential, and the posterior probability will not depend strongly on the decay rate constant of the small amplitude exponential. But if the posterior probability does not depend strongly on this decay rate constant, then again we have a broad peak that is developing into a ridge line, and the Laplace approximation cannot be applied.

Because we ignored the projection of the model onto the noise in the calculations, these parameter estimates should be considered as a lower bound on the estimated uncertainties. The actual parameter estimates obtained for any given data set will essentially never be better than these estimates, and will almost certainly be worse.

As noted in the introduction, the traditional way to obtain lower bounds on parameter estimates is using the Cramer Rao lower bound. The Cramer Rao lower bound is a theoretical result that specifies the minimum variance for a parameter estimate, given an unbiased, single parameter estimator (*26*, *27*). This can be extremely useful. However, the Cramer Rao lower bound does not provide the estimator. An estimator must be guessed and then tested to see if it achieves the Cramer Rao lower bound. Further, the Cramer Rao lower bound was derived for single parameter estimators. It was not derived for multiple parameter estimators, which are required for exponential signal models, nor does it apply when biased estimators are used. Often, the multiple parameter estimator problem is resolved by constraining the other parameters to their "best" estimate and then evaluating the Cramer Rao lower bound for the parameter of interest, as if those other parameters were not present. Bayesian probability theory uses biased estimators. The bias is introduced by the prior probabilities and by marginalization. The Cox theorem (*8* [chaps. 1–3], *28*) guarantees that the Bayesian estimate is the best estimate one can make. Any other technique will either do worse, or reproduce the Bayesian results, but it will not outperform the Bayesian calculation. To determine how a Bayesian posterior probability will perform, procedures other than calculating the Cramer Rao lower bound are required (*29*). The calculations presented in this article are a specific example. In these calculations, our intention was to derive closed-form expressions for how the actual Bayesian posterior probabilities perform. The numerical example given in the discussion illustrates that the derived expressions do reflect how the actual Bayesian posterior probabilities perform.

In this article, uniform prior probabilities were purposely employed. Thus, the results presented will approximate those expected from an optimal unbiased multiple parameter estimator and the insights offered by Eqs. [1–4] are broadly applicable to exponential parameter estimation.

ACKNOWLEDGMENTS

I thank Joseph J. H. Ackerman, Jeffrey J. Neil, Joel R. Garbow, Dmitriy Yablonskiy, and Alex Sukstanskyy for encouragement, support, and helpful comments. This work was supported by a contract with Varian NMR Systems, Palo Alto, CA; by the Small Animal Imaging Resources Program (SAIRP) of the National Cancer Institute, grant R24 CA83060; and by the National Institute of Neurological Disorders and Stroke grants NS35912 and NS41519.

APPENDIX

A. Data Containing a Single Exponential Signal

As explained in the text, we make three simplifying assumptions in the process of performing these calculations. First, we assume that the exponential signals decay by three or more e-foldings. Second, we assume that the projection of the noise onto the model is small compared with the projection of the signal onto the model, i.e., high signal-to-noise ratio data. Third, we repeatedly project one exponential onto another. These projections result in sums of the form $\sum_{i=1}^{N} \exp\{-\delta t_i\}$, where δ is typically a simple function of the decay rate constants (e.g., the sum of two decay rate constants). The third simplifying assumption is related to this sum. If we assume uniform sampling of N data points with sampling interval ΔT between two consecutive data points, this sum may be approximated as

$$\sum_{j=0}^{N-1} \exp\{-\delta\Delta Tj\} = 1 + X + \cdots + X^{N-1} = \frac{1 - X^N}{1 - X} \approx \frac{1}{\delta\Delta T}, \quad [13]$$

where $X \equiv \exp\{-\delta\Delta T\}$. In deriving this approximation, we assume that $\exp\{-\delta\Delta T\ N\}$ is small compared with one, and that $\delta\Delta T \ll 1$, so that X may be approximated by a two-term Taylor series.

For roughly 100 data values and a signal decaying to three e-foldings, this approximation introduces an error of about 5%, more than sufficient for our purposes.

We first consider data containing a single exponential signal. In this calculation, the data value that was sampled at time t_i is designated as d_i, and the data are related to the parameters to be estimated by

$$d_i = A \exp\{-\alpha t_i\} + n_i \quad (1 \le i \le N), \quad [14]$$

where n_i represents the noise at time t_i. The two parameters to be estimated are A and α. All of the information in the data relevant to these two parameters is summarized in the joint posterior probability for the amplitude and decay rate constant. Symbolically, this posterior probability is written as $P(A\alpha|D\sigma I)$. This expression should be read as the joint posterior probability for the amplitude and decay rate constant given all of the data, D, the standard deviation, σ, and the prior information, I. The standard deviation is really the standard deviation of a Gaussian prior probability that is assigned to represent what is known about the noise. In the following calculations, we assume this standard deviation is known and equal to the "true" standard deviation of the noise. We do this so that we can explicitly show how the parameter estimates depend on the noise level. Finally, the prior information represents everything known about the parameters, the model, and the noise.

To compute the joint posterior probability, $P(A\alpha|D\sigma I)$, we apply Bayes' theorem [5] and the product rule to obtain

$$P(A\alpha|D\sigma I) \propto P(\alpha|I)P(A|I)P(D|\alpha A\sigma I) \quad [15]$$

where $P(\alpha|I)$ and $P(A|I)$ are the prior probabilities for the parameters and $P(D|\alpha A\sigma I)$ is the direct probability for the data and is proportional to a likelihood function. In the high signal-to-noise ratio approximation, any uninformative prior probability assigned, regardless of its functional form, will essentially be a constant over the high-likelihood region and so cancel when the distributions are normalized. Consequently, we simply assign uniform prior probabilities to these parameters. Finally, the direct probability, $P(D|\alpha A\sigma I)$, is assigned using a Gaussian noise prior probability having zero mean, and standard deviation, σ. Thus, we obtain

$$P(A\alpha|D\sigma I) \propto \exp\left\{-\frac{Q}{2\sigma^2}\right\} \quad [16]$$

as the joint posterior probability for the parameters (note that a normalization constant has been dropped). The quantity Q, defined by

$$Q \equiv \sum_{i=1}^{N} (d_i - A\exp\{-\alpha t_i\})^2, \quad [17]$$

is essentially chi-squared.

A.1. Estimating the Decay Rate Constant

To estimate the decay rate constant, we apply the sum rule of probability theory to the joint posterior probability for the parameters, Eq. [16], to obtain

$$P(\alpha|D\sigma I) \propto \int dA\, \exp\left\{-\frac{Q}{2\sigma^2}\right\}. \quad [18]$$

The integral over A is a Gaussian integral. It is evaluated by completing the square in Q, followed by a change of variables to transform the integral into a Gaussian. One then obtains

$$P(\alpha|D\sigma I) \propto \exp\left\{\frac{\overline{h^2}}{2\sigma^2}\right\}, \quad [19]$$

from which we have again dropped some constants. The quantity $\overline{h_2}$ is called a sufficient statistic, and it summarizes all of the information in the data relevant to estimating the decay rate constant. This sufficient statistic is given by

$$\overline{h^2} = \frac{(d \cdot G)^2}{G \cdot G} \quad [20]$$

where " · " means sum over discrete times; and we are using G as a place holder for the exponential in the model. Consequently,

$$d \cdot G \equiv \sum_{i=1}^{N} d_i G(t_i) = \sum_{i=1}^{N} d_i \exp\{-\alpha t_i\}, \quad [21]$$

and

$$G \cdot G \equiv \sum_{i=1}^{N} G(t_i)^2 = \sum_{i=1}^{N} \exp\{-2\alpha t_i\} \approx \frac{1}{2\alpha\Delta T}. \quad [22]$$

As already noted, the approximation in this last equation assumes that the exponential decays into the noise and that $\exp\{-2\alpha\Delta T\} \cong 1 - 2\alpha\Delta T$.

We were able to simplify the $G \cdot G$ term because all of the quantities in this sum were known, but the term $d \cdot G$ could not be simplified because we do not know the numerical values of the data. However, suppose the data are given by

$$d_i \equiv \hat{A}\exp\{-\hat{\alpha} t_i\}, \quad [23]$$

where $\hat{A}$ and $\hat{\alpha}$ are the "true" amplitude and decay rate constant of the uniformly sampled signal, then

$$d \cdot G = \hat{A} \sum_{i=1}^{N} \exp\{-(\alpha + \hat{\alpha})t_i\} \approx \frac{\hat{A}}{(\alpha + \hat{\alpha})\Delta T}. \quad [24]$$

In making this approximation, the projection of the model onto the noise has been ignored.

For the postulated data, the marginal posterior probability density for the decay rate constant becomes

$$P(\alpha|D\sigma I) \propto \exp\left\{\frac{\alpha \hat{A}^2}{\sigma^2(\alpha + \hat{\alpha})^2 \Delta T}\right\}. \quad [25]$$

The maximum of this posterior probability occurs at $\alpha = \hat{\alpha}$. Taylor expanding the exponent about this maximum to second order gives

$$P(\alpha|D\sigma I) \approx \exp\left\{-\frac{(\alpha - \hat{\alpha})^2 \hat{A}^2}{16\sigma^2 \hat{\alpha}^3 \Delta T}\right\}, \quad [26]$$

from which we obtain

$$(\alpha)_{est} = \hat{\alpha} \pm \frac{\sigma}{\hat{A}} \sqrt{8\hat{\alpha}^3 \Delta T} \quad [27]$$

as the estimate (mean ± standard deviation) for the decay rate constant.

A.2. Estimating the Amplitude

As noted previously, all of the information in the data relevant to estimating the amplitude is summarized in a posterior probability. This posterior probability is represented symbolically by P(A|DσI), which should be read as the posterior probability for the amplitude given all of the data, the standard deviation, and the prior information. Just as the posterior probability for the decay rate constant was computed from the joint posterior probability for the parameters by integrating over the amplitude, the posterior probability for the amplitude is computed from the joint posterior probability by integrating over the decay rate constant:

$$P(A|D\sigma I) \propto \int d\alpha \, \exp\left\{-\frac{Q}{2\sigma^2}\right\}, \qquad [28]$$

where Q was defined in Eq. [17]. The integral in this equation has no closed-form solution. However, the function Q is a simple function of the amplitude and decay rate constant. One can locate the maximum of Q as a function of α, Taylor expand about this maximum, and evaluate the integral using the Laplace approximation. This Laplace approximation is the real version of the method of stationary phase in complex analysis. Locating this maximum and evaluating this integral, one obtains

$$P(A|D\sigma I) \propto \exp\left\{-\frac{(A-\hat{A})^2}{4\sigma^2\hat{\alpha}\Delta T}\right\} \qquad [29]$$

as the marginal posterior probability density for the amplitude. In deriving this expression, several constants were dropped. Examining Eq. [29], the estimated amplitude (mean ± standard deviation) is given by

$$(A)_{est} = \hat{A} \pm \sigma\sqrt{2\hat{\alpha}\Delta T}. \qquad [30]$$

B. Data Containing a Biexponential Signal

The results derived in the previous appendix show explicitly how the uncertainty in the parameter estimates depends on the acquisition parameters for data containing a single exponential signal component. In this Appendix, we treat the case of data containing a biexponential signal. Because the mathematics is complicated, we simply sketch how the calculations are performed. The actual calculations were performed using a symbolic mathematics package.

The model considered in this section is the two component or biexponential model:

$$d_i = G_i + n_i, \qquad [31]$$

where G_i is a biexponential:

$$G_i = A \exp\{-\alpha t_i\} + B \exp\{-\beta t_i\}, \qquad [32]$$

with $1 \leq i \leq N$, and B and β are the amplitude and decay rate constant of the second exponential model component. As in the single exponential case, the posterior probability for each of the parameters is computed from the joint posterior probability for all of the parameters. The joint posterior probability, represented symbolically by $P(A\alpha B\beta|D\sigma I)$, is computed by application of Bayes' theorem and the product rule. As in the single exponential case, all of the prior probabilities are assigned using uniform prior probabilities, so the joint posterior probability is proportional to the direct probability:

$$P(A\alpha B\beta|D\sigma I) \propto P(D|A\alpha B\beta\sigma I). \qquad [33]$$

If we assign the direct probability using a Gaussian prior probability for the noise, we obtain

$$P(A\alpha B\beta|D\sigma I) \propto \exp\left\{-\frac{Q}{2\sigma^2}\right\}, \qquad [34]$$

where Q, for this problem, is given by

$$Q \equiv \sum_{i=1}^{N} (d_i - G_i)^2, \qquad [35]$$

and is essentially chi-squared.

B.1. Estimating the Decay Rate Constants

All of the information in the data relevant to estimating the decay rate constants is contained in the joint posterior probability for these rate constants. This joint posterior probability is represented symbolically by $P(\alpha\beta|D\sigma I)$, which should be read as the joint posterior probability for the decay rate constants given the data, the standard deviation, and the prior information. This probability is computed from Eq. [34] by application of the sum rule:

$$P(\alpha\beta|D\sigma I) \propto \int dAdB \, \exp\left\{-\frac{Q}{2\sigma^2}\right\}. \qquad [36]$$

The integral over A and B is Gaussian. Again, evaluating such integrals is straightforward and we omit the details here, to obtain

$$P(\alpha|D\sigma I) \propto \exp\left\{\frac{\overline{h^2}}{2\sigma^2}\right\} \quad [37]$$

as the joint posterior probability for the decay rate constants. The sufficient statistic, $\overline{h^2}$, is given by

$$\overline{h^2} = \frac{2\alpha\beta\Delta T(\alpha + \beta)^2}{(\alpha - \beta)^2}\left\{\frac{T_1^2}{2\beta} - \frac{2T_1T_2}{\alpha + \beta} + \frac{T_2^2}{2\alpha}\right\} \quad [38]$$

with

$$T_1 = \sum_{i=1}^{N} d_i \exp\{-\alpha t_i\} \quad [39]$$

and

$$T_2 = \sum_{i=1}^{N} d_i \exp\{-\beta t_i\}. \quad [40]$$

As we did in the single exponential case, we now postulate a data set containing two exponential signal components given by

$$d_i \equiv \hat{A}\exp\{-\hat{\alpha}t_i\} + \hat{B}\exp\{-\hat{\beta}t_i\}, \quad [41]$$

where $\hat{B}$ and $\hat{\beta}$ are the "true" amplitude and decay rate constant of the second exponential signal component. Substituting the data, Eq. [41], into Eqs. [39, 40], T_1 and T_2 are approximated by

$$T_1 = \frac{\hat{A}}{(\alpha + \hat{\alpha})\Delta T} + \frac{\hat{B}}{(\alpha + \hat{\beta})\Delta T} \quad [42]$$

and

$$T_2 = \frac{\hat{A}}{(\beta + \hat{\alpha})\Delta T} + \frac{\hat{B}}{(\beta + \hat{\beta})\Delta T}. \quad [43]$$

Substituting these values for T_1 and T_2 into the joint posterior probability for the decay rate constants, we discover that the joint distribution is completely symmetric with respect to the two exponentials. Consequently, we do not need the posterior probability for α and β, we need this distribution for only one of the two decay rate constants. By an appropriate change of labels, we can then obtain the results for the other decay rate constant. Below, we compute the posterior probability for α the decay rate constant of the first exponential.

The posterior probability for this decay rate constant is computed from the joint posterior probability for the two decay rates, $P(\alpha\beta|D\sigma I)$, using the sum rule of probability theory. One then obtains

$$P(\alpha|D\sigma I) = \int d\beta \exp\left\{\frac{\overline{h^2}}{\sigma^2}\right\}. \quad [44]$$

The maximum of the integrand occurs at $\alpha = \hat{\alpha}$ and $\beta = \hat{\beta}$. If we Taylor expand about this maximum to second order, the integral can be evaluated in the Laplace approximation:

$$P(\alpha|D\sigma I) \propto \exp\left\{-\frac{(\hat{\alpha} - \alpha)^2\hat{A}^2(\hat{\alpha} - \hat{\beta})^4}{16\sigma^2\hat{\alpha}^3\Delta T(\hat{\alpha} + \hat{\beta})^4}\right\}, \quad [45]$$

from which several constants have been dropped. From Eq. [45] one obtains

$$(\alpha)_{est} = \hat{\alpha} \pm \frac{\sigma}{\hat{A}}\left[\frac{\hat{\alpha} + \hat{\beta}}{\hat{\alpha} - \hat{\beta}}\right]^2 \sqrt{8\hat{\alpha}^3\Delta T} \quad [46]$$

as the estimate (mean $\pm$ standard deviation) of the decay rate constant. To obtain the estimate for the second decay rate constant, one simply exchanges the role of the two exponentials:

$$(\beta)_{est} = \hat{\beta} \pm \frac{\sigma}{\hat{\beta}}\left[\frac{\hat{\alpha} + \hat{\beta}}{\hat{\alpha} - \hat{\beta}}\right]^2 \sqrt{8\hat{\beta}^3\Delta T}. \quad [47]$$

B.2. Estimating the Amplitudes

As with the other examples, all of the information in the data relevant to estimating the amplitude is contained in the posterior probability for the amplitude. Because we have already described three of these calculations, we omit most of the details and go directly to the results. The posterior probability for the amplitude of the first component is represented symbolically by $P(A|D\sigma I)$, which is a marginal posterior probability. It is computed from the joint posterior probability for all of the parameters, Eq. [34], by application of the sum rule

$$P(A|D\sigma I) = \int dBd\alpha d\beta \exp\left\{\frac{Q}{2\sigma^2}\right\}, \quad [48]$$

where Q is given by Eq. [35]. If we assume data of the form of Eq. [41], the posterior probability for the amplitude may be written as

$$P(A|D\sigma I) \propto \int d\alpha d\beta \, \exp\left\{-\frac{(\mathscr{A} - A)^2}{2\sigma^2}\right\}, \quad [49]$$

where the integral over B was evaluated analytically and $\mathscr{A}$ is given by

$$\mathscr{A} \equiv 2\alpha\Delta T\left[\frac{\alpha + \beta}{\alpha^2 + \beta^2}\right]\left[T_1 + \left(\frac{2\beta}{\alpha + \beta}\right)T_2\right], \quad [50]$$

where T_1 and T_2 were defined previously, Eqs. [39, 40]. If we now Taylor expand the integrand about its maximum to second order and evaluate the two integrals using the Laplace approximation, the posterior probability for A is given by

$$P(A|D\sigma I) \propto \exp\left\{-\frac{(\hat{A} - A)^2}{4\sigma^2\hat{\alpha}\Delta T}\left[\frac{\hat{\alpha} - \hat{\beta}}{\hat{\alpha} + \hat{\beta}}\right]^2\right\}. \quad [51]$$

Examining Eq. [51] the estimated (mean ± standard deviation) amplitude estimates are given by

$$(A)_{est} = \hat{A} \pm \sigma\left|\frac{\hat{\alpha} + \hat{\beta}}{\hat{\alpha} - \hat{\beta}}\right|\sqrt{2\hat{\alpha}\Delta T} \quad [52]$$

and

$$(B)_{est} = \hat{B} \pm \sigma\left|\frac{\hat{\alpha} + \hat{\beta}}{\hat{\alpha} - \hat{\beta}}\right|\sqrt{2\hat{\beta}\Delta T}. \quad [53]$$

REFERENCES

1. Istratov AA, Vyvenko OF. 1999. Exponential analysis in physical phenomena. Rev Sci Instrum 70:1233–1257.
2. Holmstrom K, Petersson J. 2002. A review of the parameter estimation problem of fitting positive exponential sums to empirical data. Appl Math Comput 126:31–36.
3. Bromage GE. 1983. A quantification of the hazards of fitting sums of exponentials to noisy data. Comput Phys Commun 30:229–233.
4. Shrager RI, Hendler RW. 1998. Some pitfalls in curve-fitting and how to avoid them: a case in point. J Biochem Biophys Methods 36:157–173.
5. Bayes Rev. T. 1763. An essay toward solving a problem in the doctrine of chances. Philos Trans R Soc London 53:370–418; reprinted in 1958. Biometrika 45:293–315; and 1963. Facsimiles of two papers by Bayes with commentary by Deming WE. New York: Hafner.
6. Laplace PS. 1951 [1814]. A philosophical essay on probabilities. [Unabridged and unaltered reprint of the Truscott and Emory translation.] New York: Dover Publications.
7. Jefferys H. 1983. Theory of probability. Oxford: Clarendon Press.
8. Jaynes ET. 2003. Probability theory: the logic of science. In: Bretthorst GL, editor. Cambridge: Cambridge University Press.
9. Bretthorst GL. 1988. Bayesian spectrum analysis and parameter estimation. In: Berger J, et al. Lecture notes in statistics, vol. 48. New York: Springer-Verlag.
10. Bretthorst GL. 1989. An introduction to parameter estimation using Bayesian probability theory. In: Fougère PF, editor. Maximum entropy and Bayesian methods. Dordrecht, The Netherlands: Kluwer Academic Publishers p 53–70.
11. Bretthorst GL. 1990. Bayesian analysis. I. Parameter estimation using quadrature NMR models. J Magn Reson 88:533–551.
12. Bretthorst GL. 1990. Bayesian analysis. II. Signal detection and model selection. J Magn Reson 88:552–570.
13. Bretthorst GL. 1990. Bayesian analysis. III. Applications to NMR signal detection, model selection, and parameter estimation. J Magn Reson 88:571–595.
14. Bretthorst GL. 1991. Bayesian analysis. IV. Noise and computing time considerations. J Magn Reson 93:369–394.
15. Bretthorst GL. 1992. Bayesian analysis. V. Amplitude estimation for multiple well-separated sinusoids. J Magn Reson 98:501–523.
16. Bretthorst GL. 1992. Estimating the ratio of two amplitudes in nuclear magnetic resonance data. In: Erickson GL, Neudorfer PO, editors. Maximum entropy and Bayesian methods. Dordrecht, The Netherlands: Kluwer Academic Publishers p 66–77.
17. d'Avignon DA, Bretthorst GL, Holtzer ME, Holtzer 1998. A. Site-specific thermodynamics and kinetics of a coiled-coil transition by spin inversion transfer NMR. Biophys J 74:3190–3197.
18. d'Avignon DA, Bretthorst GL, Holtzer ME, Holtzer A. 1999. Thermodynamics and kinetics of a folded-folded transition at valine-9 of a GCN4-like leucine zipper. Biophys J 76:2752–2759.
19. Holtzer ME, Bretthorst GL, d'Avignon DA, Angeletti RH, Mints L, Holtzer A. 2001. Temperature dependence of the folding and unfolding kinetics of the GCN4 leucine zipper via ^{13}C(alpha)-NMR. Biophys J 80:939–951.
20. Bretthorst GL. 2001. Nonuniform sampling: bandwidth and aliasing. In: Rychert J, Erickson G, Smith CR, editors. Maximum entropy and Bayesian methods in science and engineering. College Park, MD: American Institute of Physics. p 1–28.
21. Bretthorst GL. 2001. Generalizing the Lomb-Scargle periodogram. In: Mohammad-Djafari A, editor. Maximum entropy and Bayesian methods in science and engineering. College Park, MD: American Institute of Physics p 241–245.

22. Bretthorst GL. 2001. The Lomb-Scargle periodogram—the nonsinusoidal case. In: Mohammad-Djafari A, editor. Maximum entropy and Bayesian methods in science and engineering. College Park, MD: American Institute of Physics p 246–251.
23. Bretthorst GL. 1996. An introduction to model selection using probability theory as logic. In: Heidbreder GR, editor. Maximum entropy and Bayesian methods. Dordrecht, The Netherlands: Kluwer Academic Publishers p 1–42.
24. Bretthorst GL. 1993. On the difference in means. In: Physics and probability essays in honor of Edwin T. Jaynes. Grandy WT, Milonni PW, editors. Cambridge: Cambridge University Press p 177–194.
25. Ramaswami A, Bretthorst GL. Probability 1996. Theory and multiexponential signals, how accurately can they be determined. In: Heidbreder GR, editor. Maximum entropy and Bayesian methods. Dordrecht, The Netherlands: Kluwer Academic Publishers p 265–273.
26. Cramer H. 1946. Mathematical methods of statistics. Princeton, NJ: Princeton University Press.
27. Rao C. R. 1945. Information and the accuracy attainable in the estimation of statistical parameters. Bull Calcutta Math Soc 37:81–91.
28. Cox RT. 2002. The algebra of probable inference. Baltimore: Johns Hopkins University Press p 1–29.
29. Nicholson D, Vecchio A. 1998. A Bayesian bounds on parameter estimation accuracy for compact coalescing binary gravitational wave signals. Phys Rev D 57: 4588–4599.

BIOGRAPHY

G. Larry Bretthorst is a senior scientist at Washington University. He has written extensively on parameter estimation and model selection using Bayesian probability theory, including a book titled *Bayesian Spectrum Analysis and Parameter Estimation* and more recently as editor of *Probability Theory: The Logic of Science*.