Bayesian Analysis Users Guide
Release 4.00, Manual Version 1

G. Larry Bretthorst
Biomedical MR Laboratory
Washington University School Of Medicine,
Campus Box 8227
Room 2313, East Bldg.,
4525 Scott Ave.
St. Louis MO 63110
http://bayes.wustl.edu
Email: larry@bayes.wustl.edu

October 21, 2016

Contents

List of Figures

List of Tables

Chapter 19

Behrens-Fisher

The Behrens-Fisher problem is the classical medical testing problem. In this problem one has two data sets. Each data set is a repeated measurement of the same quantity, so the signal is presumed constant in both measurements. However, somewhere between measuring the first and second data set, something is changed. That change could be a change in an experimental parameter, a change in the health of the animal, or any thing else that could introduce a change into the measured data. When the package is run, the program computes the posterior probability that something has changed. The interface to the Behrens-Fisher package is shown in Fig. 19. Note that this interface does not have any package specific widgets. However, it does have two prior probabilities that must be reviewed to make sure the interface set them to reasonable values. To use this package, you must do the following:

Select the Behrens-Fisher package from the Package menu.

Load two Ascii data sets using the Files menu. When a data set is successfully loaded the data is plotted in the Ascii Data viewer.

Review the prior probabilities for the mean and standard deviation using the Prior Viewer.

Select the server that is to process the analysis.

Check the status of the selected server to determine if the server is busy, change to another server if the selected server is busy.

Run the the analysis on the selected server by activating the Run button.

Get the the results of the analysis by activating the Get Job button. If the analysis is running, this button will return the Accepted report containing the status of the current run. Otherwise, it will fetch and display the results from the current analysis.

19.1 Bayesian Calculation

In the classical medical testing problem, the question one would like to answer is, are the data sets the same or are they different? If their different, how? Did the means change? Did the standard

Figure 19.1: The Behrens-Fisher Interface

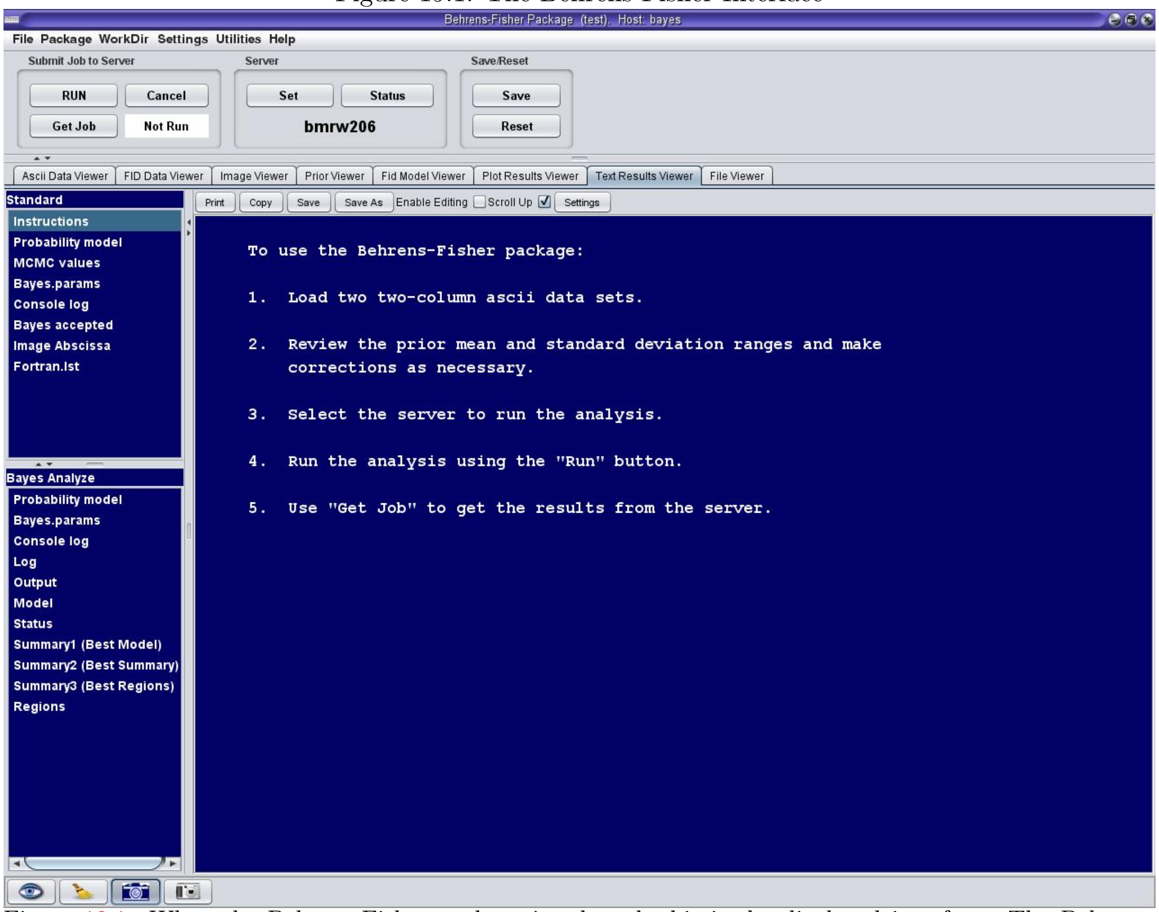

Figure 19.1: When the Behrens-Fisher package is selected, this is the displayed interface. The Behrens-Fisher package does not have any package specific widgets. However, it does have two prior probabilities that should be examined before running the analysis. These priors are for the mean and standard deviation of the data sets. The prior viewer can be used to set these prior probabilities.

Figure 19.2: Behrens-Fisher Hypotheses Tested

Hypotheses	**Abbreviation**
The means and the variances are the same	$SmSv$
The means are the same and the variances differ	$SmDv$
The means differ and the variances are the same	$DmSv$
The means and variances differ	$DmDv$
The means are the same	Sm
The means are not the same	Dm
The variances are the same	Sv
The variances are not the same	Dv
The data sets differ	$Dm + Dv$
The mean in set 1 is equal to C_1	C_1
The mean in set 2 is equal to C_2	C_2
The standard deviation in set 1 is equal to σ_1	σ_1
The standard deviation in set 2 is equal to σ_2	σ_2
The difference, $C_1 - C_2$, is equal to δ	δ
The sum, $C_1 + C_2$, is equal to γ	γ
The ratio σ_1/σ_2 is equal to w	w
The ratio σ_2/σ_1 is equal to x	x

Table 19.2: This table lists the various hypotheses that are of interest in the Behrens-Fisher problem. It has been divided into three sections. The top section is a set of model selection hypotheses and the probabilities for these hypotheses may be used to compute the probabilities in the center section. Finally, the lower section of the table is a series of parameter estimation hypotheses. Each hypotheses in this table requires a separate Bayesian calculation.

deviations change? If something changed by how much. These are only a small sample of all of the hypotheses that are of interest in the Behrens-Fisher problem, see [10] for more on this problem. The entire series of hypotheses about which we wish to make inferences are shown in Fig. 19.2. This table has been divided into three sections. The top section is a series of model selection probabilities, and they are used to determine which of four mutually exclusive and exhaustive hypotheses best describe the data. These probabilities are need to compute the probabilities in both the center and lower sections of this table. The center section of the table are hypotheses who's probabilities may be derived from the four model selection probabilities. The third section of the table is a series of parameter estimation problems. However, as we will see, even these parameter estimation problems depend on first four model selection problems. Because each section of this table produces rather different Bayesian calculations, we will address each section separately.

We are going to use the rules of probability theory to compute the probability for each of the hypotheses appearing in Fig. 19.2. Unfortunately, this is a rather tedious calculation simply because the number of hypotheses of interest is large. We will simply take the entries in the table one at a time. Before we start this process we define a little notation. In the following D_1 will designate the first data set. Which data set is first and which is second is an arbitrary designation and the results will not depend on which is which; D_1 simply designate one of the data sets. In the program that implements these calculations, D_1 is literally the first data set loaded by the interface. Similarly,

D_2 will designate the other data set. We will use D to represent all of the data. The means of the first, second and combined data will be written as C_1, C_2 and C respectively. Combining or pooling the data comes about in several of the Bayesian calculations for the probability for the hypotheses shown in Fig. 19.2. For example, in the model the means and variances are the same, probability theory will lead to pooling the data into a single data set. The standard deviations of the noise prior probabilities will be designated as σ_1, σ_2 and σ for first, second and pooled data. Finally, N_1, N_2 and $N = N_1 + N_2$ will represent the total data in the first, second and pooled data.

19.1.1 The Four Model Selection Probabilities

With this notation now established we begin by factoring the probability for the hypotheses, "The means and variances are the same." This hypotheses is abbreviate as "$SmSv$" where we have used the term variance in this hypotheses. The variance would usually mean σ^2, the squared error; while it is always the standard deviation of the noise prior probability, σ, that appears in our probabilities. So the hypotheses really should have been "the means and the standard deviations of the noise are the same." This was so verbose, that we decided to use the word "variance" in these hypotheses even though it is not quite the correct expression.

We are going to attack the table from the top to the bottom. We will spend a great deal of time on the first four hypotheses as these are the model selection calculations needed to perform the calculations in the reaming sections. We will derive the probabilities in the center section, but it will turn out these are all just linear combinations of the four probabilities computed for the hypotheses in the first section of the table. Finally, to a large degree, we will only sketch how the probabilities in the bottom section of this table are computed.

The set of hypotheses: $SmSv$, $SmDv$, $DmSv$, and $DmDv$, are mutually exclusive and exhaustive given that the signals in the two data sets are constants. Let us define a model indicator ℓ. When every you see ℓ in an equation, replace it by one of the four hypotheses of interest. To perform a model selection calculation, we must compute the posterior probability for the model indication ℓ given all of the data D and whatever prior information I we might have. This posterior probability is represented symbolically by $P(\ell|DI)$. To compute it, one applies Bayes' theorem

$$P(\ell|DI) = \frac{P(\ell|I)P(D|\ell I)}{P(D|I)} \tag{19.1}$$

where $P(\ell|I)$ is the prior probability for the model, $P(D|\ell I)$ is the probability for the data given ℓ and I, and $P(D|I)$ is a normalization constant:

$$P(D|I) = \sum_{\ell} P(\ell|I)P(D|\ell I). \tag{19.2}$$

If we assign a uniform prior probability to the model indicator, then we have

$$P(\ell|DI) = \frac{P(D|\ell I)}{\sum_{\ell} P(D|\ell I)}. \tag{19.3}$$

If we compute all four probabilities represented symbolically by $P(D|\ell I)$ then we can always compute the normalization constant. So it is sufficient to compute

$$P(\ell|DI) \propto P(D|\ell I). \tag{19.4}$$

It is this discrete probability distribution that is targeted by the simulated annealing portion of the program that implements the Behrens-Fisher calculations. In the following subsections we compute these four model probabilities represented symbolically by $P(D|\ell I)$.

19.1.1.1 The Means And Variances Are The Same

Given the hypotheses, the mean and variance are the same, the model equation which relates the parameters of interest to the data is

$$d_i = C + \text{Noise of Standard Deviation } \sigma \tag{19.5}$$

where d_i is the pooled or combined data, C is the mean, and σ is the standard deviation of the noise prior probability.

The probability that is needed is the probability for the data given that the means and variances are the same, $P(D|SmSvI)$. This is a marginal probability where the mean, C, and standard deviation of the noise prior probability, σ, have been removed by marginalization:

$$P(D|SmSvI) = \int_{Low}^{High} dC \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma P(DC\sigma|SmSvI) \tag{19.6}$$

where Low and $High$ bound the mean, and similarly, σ_{Low} and σ_{High} are the bounds on the standard deviation.

The right-hand side of this equation is factored using Bayes' theorem, to obtain

$$P(D|SmSvI) \propto \int_{Low}^{High} dC \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma P(C\sigma|SmSvI)P(D|C\sigma SmSvI) \tag{19.7}$$

where $P(C\sigma|SmSvI)$ is the joint prior probability for the mean and the standard deviation given the model, and $P(D|C\sigma SmSvI)$ is the direct probability for the data given the parameters and the model.

The prior probability for the parameters, $P(C\sigma|SmSvI)$, is factored into two independent prior probabilities:

$$P(C\sigma|SmSvI) = P(C|I)P(\sigma|I) \tag{19.8}$$

where $P(C|I)$ and $P(\sigma|I)$ are the prior probabilities for the mean and the standard deviation. Note that we have assumed the prior information is independent of the model, i.e., we will assign the same prior probability for a mean or standard deviation regardless of which model we are discussing.

The prior probability for the mean, $P(C|I)$, will be assigned as a bound uniform prior probability

$$P(C|I) = \begin{cases} \dfrac{1}{R_C} & \text{If } Low \leq C \leq High \\ 0 & \text{Otherwise} \end{cases} \tag{19.9}$$

where $R_C = High - Low$.

The prior probability for the standard deviation, $P(\sigma|I)$, will be assigned as a bounded Jeffreys' prior:

$$P(\sigma|I) = \begin{cases} \dfrac{1}{\sigma R_\sigma} & \text{If } \sigma_{Low} \leq C \leq \sigma_{High} \\ 0 & \text{Otherwise} \end{cases} \tag{19.10}$$

and

$$\begin{aligned}1 &= \frac{1}{R_\sigma}\int_{\sigma_{Low}}^{\sigma_{High}} d\sigma \frac{1}{\sigma} \\ R_\sigma &= \log\left(\sigma_{High}/\sigma_{Low}\right).\end{aligned} \tag{19.11}$$

Whenever we assign a prior probability for a mean or a standard deviation in the following calculations, it will always be of the form of these two priors and these priors will use the exact same prior ranges and normalization constants.

The only term remaining in Eq. (19.7) to be assigned is the likelihood, $P(D|C\sigma SmSvI)$, and this will be assigned using a Gaussian as the noise prior probability:

$$P(D|C\sigma SmSvI) \propto \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left\{-\frac{Q_{SmSv}}{2\sigma^2}\right\} \tag{19.12}$$

where

$$Q_{SmSv} \equiv N(\overline{d^2} - 2C\overline{d} + C^2) \tag{19.13}$$

where $\overline{d^2}$ is the mean-square of the combined or pooled data and $\overline{d}$ is the average pooled data:

$$\overline{d^2} = \frac{1}{N}\sum_{i=1}^{N} d_i^2 \qquad \text{and} \qquad \overline{d} = \frac{1}{N}\sum_{i=1}^{N} d_i \tag{19.14}$$

where these sums are over all of the combined or pooled data. Together $\overline{d^2}$, $\overline{d}$ and N are sufficient statistics for computing this direct probability. The presence of these sufficient statistics is the reason that when the Behrens-Fisher program runs, its execution time is almost completely independent of the data sets being analyzed. The only component of the calculation that depends on the data is computing $\overline{d^2}$ and $\overline{d}$. Computing these averages is negligible compared to the other calculations the program must perform.

Combining the prior and the likelihood, one obtains

$$P(D|SmSvI) \propto \int_{Low}^{High} dC \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma \frac{1}{\sigma R_\sigma R_C}\left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left\{-\frac{Q_{SmSv}}{2\sigma^2}\right\} \tag{19.15}$$

as the marginal direct probability for the data given the $SmSv$ hypotheses. We note in passing that the integrand of this equation is proportional to the joint posterior probability for the parameters given the $SmSv$ model, a quantity we will need in the parameter estimation calculations discussed in a later Section.

Before we used this probability in the model selection calculation, the integral over σ was evaluated. We did this for technical reasons concerning the implementation of the Markov chain. Evaluating this integral changes the functional form of the direct probability from a Gaussian to Students' t-distribution, and numerically computing the t-distribution has some advantages, because it effectively introduces a lower bound to the logarithm of the probability. Evaluating the integral over σ, one obtains

$$P(D|SmSvI) \approx \int_{Low}^{High} dC \frac{1}{2R_\sigma R_C}\Gamma\left(\frac{N}{2}\right)\left[\frac{Q_{SmSv}}{2}\right]^{-\frac{N}{2}}. \tag{19.16}$$

It the integrand of Eq. (19.16) that is used in the simulated annealing portion of the calculation to do model selection given the $SmSv$ model. In deriving this equation we made the approximation that the lower and upper bound on the standard deviation of the noise prior probability are so wide that we could effectively extend the integral from zero to infinity. It is possible to do this calculation without making this approximation, see [10]. However, experience with the full calculation shows, that under almost all reasonable circumstances the above approximation is good to many decimal places.

19.1.1.2 The Mean Are The Same And The Variances Differ

The hypotheses "the means are the same and the variances differ" implies a slightly changed model equation, one has

$$d_{1i} = C + \text{Noise of standard deviation } \sigma_1 \tag{19.17}$$

and

$$d_{2j} = C + \text{Noise of standard deviation } \sigma_2. \tag{19.18}$$

In this model we cannot use the pooled data and must explicitly take into account the difference in the two data sets.

The probability for the data given that the means are the same and the variances are different is represented symbolically by $P(D|SmDvI)$. Calculation of this probability is very similar to what was done in the previous section, and we only sketch the details here:

$$P(D|SmDvI) = \int_{Low}^{High} dC \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_1 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_2 P(DC\sigma_1\sigma_2|SmDvI). \tag{19.19}$$

Applying the product rule and Bayes' theorem to the integrand, one obtains

$$P(DC\sigma_1\sigma_2|SmDvDI) \propto P(C|I)P(\sigma_1|I)P(\sigma_2|I)P(D_1|C\sigma_1 SmDvI)P(D_2|C\sigma_2 SmDvI) \tag{19.20}$$

where $P(D_1|C\sigma_1 SmDvI)$ and $P(D_2|C\sigma_2 SmDvI)$ are the direct probability for the first and second data sets respectively. As noted earlier, the priors will be assigned using Eqs. (19.9,19.10). A Gaussian likelihood will be assigned to $P(D_1|C\sigma_1 SmDvI)$ and $P(D_2|C\sigma_2 SmDvI)$. After combining all of the terms, one obtains

$$P(D|SmDvI) \propto \int_{Low}^{High} dC \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_1 d\sigma_2 \frac{(2\pi\sigma_1^2)^{-\frac{N_1}{2}}(2\pi\sigma_2^2)^{-\frac{N_2}{2}}}{\sigma_1\sigma_2 R_\sigma^2 R_C} \exp\left\{-\frac{Q_{1SmDv}}{2\sigma_1^2} - \frac{Q_{2SmDv}}{2\sigma_2^2}\right\} \tag{19.21}$$

as the direct probability for the data given the $SmDv$ model. We note again that the integrand is proportional to the joint posterior probability for the parameters given the $SmDv$ model. This integrand will be used in the Markov chain Monte Carlo simulation to draw samples from the joint posterior probability for the parameters, and these samples used in forming the probabilities for the hypotheses given in the bottom part of Fig. 19.2. Evaluating the two integrals over σ_1 and σ_2, one obtains

$$P(DSmDvI) \approx \int_{Low}^{High} dC \frac{1}{4R_\sigma^2 R_C} \Gamma\left(\frac{N_1}{2}\right)\left[\frac{Q_{1SmDv}}{2}\right]^{-\frac{N_1}{2}} \Gamma\left(\frac{N_2}{2}\right)\left[\frac{Q_{2SmDv}}{2}\right]^{-\frac{N_2}{2}} \tag{19.22}$$

as the direct probability for the data given that the means are the same and the variances differ. We again assumed that the limits on the integral were wide compared to the location of the peak of the integrand. The two functions, Q_{1SmDv} and Q_{2SmDv} are define analogously to Q_{SmSv}:

$$Q_{1SmDv} \equiv N_1(\overline{d_1^2} - 2C\overline{d}_1 + C^2) \tag{19.23}$$

where $\overline{d}_1$ and $\overline{d_1^2}$ are the mean and mean-square of the first data set, and

$$Q_{2SmDv} \equiv N_2(\overline{d_2^2} - 2C\overline{d}_2 + C^2) \tag{19.24}$$

where $\overline{d}_2$ and $\overline{d_2^2}$ are the mean and mean-square of the second data set. It is the integrand of Eq. (19.22) that is used in the simulated annealing part of the simulations to perform model selection for the $SmDv$ model.

19.1.1.3 The Means Differ And The Variances Are The Same

The hypotheses "the means differ and the variances are the same" again implies a slight change to the model equation. The model for the first data set becomes,

$$d_{1i} = C_1 + \text{Noise of standard deviation } \sigma \tag{19.25}$$

and

$$d_{2j} = C_2 + \text{Noise of standard deviation } \sigma \tag{19.26}$$

for the second data sets.

The probability for the data given that the means differ and the variances are the same is represented symbolically by $P(D|DmSvI)$. Again this is a marginal probability and is computed by application of the sum rule:

$$P(D|DmSvI) = \int_{Low}^{High} dC_1 \int_{Low}^{High} dC_2 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma P(DC_1C_2\sigma|DmSvI). \tag{19.27}$$

The right-hand side of this equation is factored using Bayes' theorem, to obtain

$$P(DC_1C_2\sigma|DmSvI) \propto P(C_1|I)P(C_2|I)P(\sigma|I)P(D_1|C_1\sigma DmSvI)P(D_2|C_2\sigma DmSvI). \tag{19.28}$$

The priors will be assigned using Eqs. (19.9,19.10). A Gaussian likelihood will be assigned to both $P(D_1|C_1\sigma DmSvI)$ and $P(D_2|C_2\sigma DmSvI)$. Collecting these terms, one has

$$P(D|DmSvI) = \int_{Low}^{High} dC_1 \int_{Low}^{High} dC_2 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma \frac{(2\pi\sigma^2)^{-\frac{N}{2}}}{\sigma R_\sigma R_C^2} \exp\left\{-\frac{Q_{1DmSv} + Q_{2DmSv}}{2\sigma^2}\right\} \tag{19.29}$$

as the direct probability for the data given the $DmSv$ model. Note the integrand is proportional to the joint posterior probability for the parameters appearing in the $DmSv$ model. This integrand is used to generate samples from the joint posterior probability for the parameters appearing in the $DmSv$ model. These samples are then used to generate samples from posterior probabilities for parameter estimation hypotheses listed in the lower part of Fig. 19.2.

Evaluating the integral over the standard deviation of the noise prior probability, one obtains

$$P(D|DmSvI) \approx \int_{Low}^{High} dC_1 \int_{Low}^{High} dC_2 \frac{1}{2R_\sigma R_C^2} \Gamma\left(\frac{N}{2}\right) \left[\frac{Q_{1DmSv} + Q_{2DmSv}}{2}\right]^{-\frac{N}{2}} \tag{19.30}$$

where

$$Q_{1DmSv} \equiv N_1(\overline{d_1^2} - 2C_1\overline{d}_1 + C_1^2) \tag{19.31}$$

and

$$Q_{2DmSv} \equiv N_2(\overline{d_2^2} - 2C_2\overline{d}_2 + C_2^2). \tag{19.32}$$

In evaluating the integral over σ we again assumed that the peak in the integrand was such that we make only a small error in extending the limits of the integral from zero to infinity. It is this probability that is used in the simulated annealing portion of the calculations to do model selection for the $DmSv$ model.

19.1.1.4 The Means And Variances Differ

The hypotheses "the means and variances differ" again implies a new model equation,

$$d_{1i} = C_1 + \text{Noise of standard deviation } \sigma_1 \tag{19.33}$$

for the first data set, and

$$d_{2j} = C_2 + \text{Noise of standard deviation } \sigma_2 \tag{19.34}$$

for the second data set.

The probability for the data given that the means and the variances differ is represented symbolically by $P(D|DmDvI)$. Again this is a marginal probability and is computed by application of the sum rule:

$$P(D|DmDvI) = \int_{Low}^{High} dC_1 \int_{Low}^{High} dC_2 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_1 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_2 P(D_1D_2C_1C_2\sigma_1\sigma_2|DmDvI). \tag{19.35}$$

This calculation is just the one done for the same means and variance, but now computed for each data set separately. We do not give the details of the calculation:

$$\begin{aligned} P(D|DmDvI) \quad &\propto \quad \int_{Low}^{High} dC_1 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_1 \frac{(2\pi\sigma_1^2)^{-\frac{N_1}{2}}}{\sigma_1 R_\sigma R_C} \exp\left\{-\frac{Q_{1DmDv}}{2\sigma_1^2}\right\} \\ &\times \quad \int_{Low}^{High} dC_2 \int_{\sigma_{Low}}^{\sigma_{High}} d\sigma_2 \frac{(2\pi\sigma_2^2)^{-\frac{N_2}{2}}}{\sigma_2 R_\sigma R_C} \exp\left\{-\frac{Q_{2DmDv}}{2\sigma_2^2}\right\}. \end{aligned} \tag{19.36}$$

Note that the integrand of this equation is proportional to the joint posterior probability for the parameters appearing in the $DmDv$. This integrand is used in the program to generate samples for this probability. These samples are in turn used to generate samples from the probabilities for the hypotheses described in the bottom section of Fig. 19.2.

Evaluating the integrals over the standard deviations, one obtains

$$P(D|DmDvI) \approx \int_{Low}^{High} dC_1 \int_{Low}^{High} \frac{dC_2}{4R_\sigma^2 R_C^2} \Gamma\left(\frac{N_1}{2}\right) \left[\frac{Q_{1DmDv}}{2}\right]^{-\frac{N_1}{2}} \Gamma\left(\frac{N_2}{2}\right) \left[\frac{Q_{2DmDv}}{2}\right]^{-\frac{N_2}{2}} \tag{19.37}$$

where

$$Q_{1DmDv} \equiv N_1(\overline{d_1^2} - 2C_1\overline{d}_1 + C_1^2) \tag{19.38}$$

and

$$Q_{2DmDv} \equiv N_2(\overline{d_2^2} - 2C_2\overline{d}_2 + C_2^2) \tag{19.39}$$

as the direct probability for data given the $DmDv$ model. It is this direct probability that is used in the simulated annealing portion of the calculation that performs model selection.

In the Markov chain Monte Carlo simulation that implements the Bayesian calculation it is the probability for the model, Eq. (19.4), that is targeted by the Markov chain. To target this distribution, there are four probabilities that must be computed, Eqs (19.16,19.22,19.30,19.37). The program that performs the Monte Carlo simulation runs multiple independent simulations. The various simulations start out uniformly distributed over the four different models. One of the routines in the program, varies the model indicator. It does this randomly, for example it might arbitrarily try to switch the indicator from the same means and variances model to the different means and variance model. When it does this, it generates a new simulation having new parameter values unrelated to those in the original simulation. It then varies the parameters in this model using a sub-Markov chain to decorrelate this simulation for the others. Finally, the routine compares the probability for original model indication to the probability for this new model indicator. The new model indicator and parameters are accepted or rejected using the acceptance criteria for a Metropolis-Hastings Markov chain. When the program completes the annealing phase, the distribution of simulations is an approximation to the posterior probability for the model indicator.

19.1.2 The Derived Probabilities

The center section of the Fig. 19.2 is in some ways simpler than the first section. By simpler we mean that these probabilities may be computed from the probabilities derived in the previous section. So while we have to apply the rules of probability theory to determine how to compute these, after we do that we should find combinations of things we have already computed.

The probability that the means are the same, $P(Sm|DI)$, is a marginal probability where we marginalize over all of the different ways that the means could be the same. There are only two ways the means could be the same. So the probability the means are the same is a sum of two terms:

$$P(Sm|DI) = P(SmSv|DI) + P(SmDv|DI) \tag{19.40}$$

where $P(SmSv|DI)$ is the probability the means and variances are the same, and $P(SmDv|DI)$ is the probability the means are the same and the variances differ. But these are just two of the probabilities for the model indicator, $P(\ell|DI)$, so after normalizing the probability for the model indicator Eq. (19.40) can be trivially computed for it.

Next the probability that the means are not the same is trivially computed from $P(Sm|DI)$ because the means are the same or differ are mutually exclusive and exhaustive, so

$$P(Dm|DI) = 1 - P(Sm|DI). \tag{19.41}$$

Like the issue of the means being the same, there are only two ways that the variances could be the same. So this is a marginal probability where we marginalize over the different ways the variances could be the same:

$$P(Sv|DI) = P(SmSv|DI) + P(DmSv|DI). \tag{19.42}$$

These again are computed from the probability for the model indicator $P(\ell|DI)$. The probability that the variances differ or are the same from a mutually exclusive and exhaustive set. So the probability that the variances differ, $P(Dv|DI)$, is given by

$$P(Dv|DI) = 1 - P(Sv|DI). \tag{19.43}$$

The final probability in the center section is the probability that the data sets differ. There are only two ways the data sets could differ, there could be different means or different variances. The probability for different means or different variances is represented symbolically as $P(Dm+Dv|DI)$. To compute this probability we apply the sum rule:

$$P(Dm + Dv|DI) = P(Dm|DI) + P(Dv|DI) - P(DmDv|DI) \tag{19.44}$$

the last term in the sum rule, the joint probability for the two hypotheses doesn't show up in most problems because the hypotheses are mutually exclusive, so the term is frequently zero. However, here the probability for different means and variances is not zero and so this term must be subtracted form the sum of the probability for different means and the probability for different variances.

19.1.3 Parameter Estimation

The probabilities in the lower section of Fig. 19.2 are all parameter estimation problems. The program that implements these calculations computes these probabilities by a somewhat roundabout way. It does four parameter estimation calculations, one for each of the four basic models. In these parameter estimation calculations a Markov chain Monte Carlo simulation is run using the integrand of Eqs. (19.12,19.21,19.29,19.36) as the target distributions. From these samples and the probabilities computed in the previous sections, it is possible for the program to form samples for each probability associated with the hypotheses listed in Fig. 19.2.

We give only a sketch of how this is done. For example the hypotheses, the difference in means is δ. Designating the difference in means by, $\delta \equiv C_1 - C_2$, then the probability for δ is given by

$$P(\delta|DI) = P(\delta Sv|DI) + P(\delta Dv|DI) \tag{19.45}$$

where we have restricted oneself to the model subspace that permits the means to be different, i.e., the different means and same or different variance. Factoring the right-hand side of this equation we obtain

$$P(\delta|DI) = P(Sv|DI)P(\delta|DmSvDI) + P(Dv|DI)P(\delta|DmDvDI). \tag{19.46}$$

Note that $P(Sv|DI)$ and $P(Dv|DI)$ are things that have already been calculated in the previous sections. But the other two probabilities, $P(\delta|DmSvDI)$ and $P(\delta|DmDvDI)$, have not been calculated. We can compute samples from from these probabilities if we have samples from Eqs. (19.29,19.36). Designating $(C_1 - C_2|DmSv)$ as the difference in means computed from the parameter estimation samples using the $DmSv$ model, and $(C_1 - C_2|DmDv)$ as the difference in means computed from the parameter estimation samples drawn from the $DmDv$ model, then a sample for the difference in means, independent of the model are given by

$$\text{Sample from:} P(\delta|DI) = P(Sv|DI)(C_1 - C_2|DmSv) + P(Dv|DI)(C_1 - C_2|DmDv). \tag{19.47}$$

So computationally the program takes the samples generated from the $DmSv$ model, computes the difference in means for each sample, multiples this by $P(Sv|DI)$ and adds this to the difference in mean computed from the samples generated from the $DmDv$ model multiplied by the probability for different variances, $P(Dv|DI)$. These samples are computed in the output portion of the program. In a similar way the probability for the sum of the means is computed by simply computing the sum of the means rather than the difference.

The samples from the probability for the ratio of the standard deviations are computed in an analogous way

$$\text{Sample from:} P(\sigma_1/\sigma_2|DI) = P(Sm|DI)(\sigma_1/\sigma_2|SmDv) + P(Dm|DI)(\sigma_1/\sigma_2|DmDv). \quad (19.48)$$

Again these samples are used to construct mean and standard deviation estimates for the ratio of standard deviations and to construct an histogram that approximates $P(\sigma_1/\sigma_2|DI)$. Samples from the probabilities for the remaining hypotheses given in the bottom section of Fig. 19.2 are computed in analogous fashions.

19.2 Outputs From Behrens-Fisher Package

The Behrens-Fisher package is sufficiently different from the other package in the Bayesian Analysis software, that we are going to outline how the program works and describe the outputs of the program in a little more detail than is typical of other packages. First, however the package does have the standard reports. The Text outputs files from the Behrens-Fisher Packages consist of: "Bayes.prob.model," "BayesBF.mcmc.values," "Bayes.params," "Console.log," "Bayes.accepted" and a "Bayes.Condensed.File." These output files can be viewed using the Text Viewer or they can be viewed using File Viewer by navigating to the current working directory and then selecting the files. The format of the mcmc.values report is discussed in Appendix D and the other reports are discussed in Chapter 3. Additionally, the "Plot Results Viewer" can be used to view the output probability density functions. In addition to the standard data, model and residual plots there are four model independent posterior probabilities, two mean values and two standard deviation. These plots are named C1, C2, Sigma1, and Sigma2 respectively in the Plot Results Viewer. They are model independent in that they are sum over all of the parameter estimates weighted by the posterior probability for the model. Additionally, there are output plots for the sums, differences and ratios of the mean values. Finally, there are output plots for the parameters given each of the four models.

The Behrens-Fisher package starts by running a Markov chain Monte Carlo simulation with simulated annealing to compute the posterior probability for the model. These four probabilities are given by the integrand of Eqs. (19.16,19.22,19.30,19.37) respectively. The simulation varies the model much like any other parameter in a Markov chain Monte Carlo simulation. To do this, the Markov chain Monte Carlo simulation postulates a change in the model by sampling a uniform prior probability for the model. Say, for example, the package postulates a change from the same mean and variance model to a same mean and different variance model. The Markov chain Monte Carlo simulation will then simulate the parameters in this new model until it reaches equilibrium at the current value of the annealing parameter. It will then either accept or reject this change of model using a Metropolis-Hastings acceptance criteria. Multiple simulations are run parallel at a fixed annealing parameter. Between annealing steps the package outputs the number of simulations in each of the four models. An example of this output is shown in Figure 19.3. This output serves as a visual picture of the posterior probability for the model indicator as it is evolving under simulated annealing. When the package first starts this distribution is initialized using a uniform prior probability for the model. As the annealing parameter is increased the distribution of simulations goes into the posterior probability for the model indicator. The posterior probabilities for the four models, shown in Fig. 19.4, are then output, These four probabilities can then be used to compute the probabilities shown in the center section of Fig. 19.4 and these probabilities are also output.

Figure 19.3: Behrens-Fisher Console Log

```
Phase         % <Prior> <Likelihood>  SmSv  SmDv  DmSv  DmDv
Prob Model  80  -5.39 -6.183379E+01   18    12     0     0
            82  -5.66 -6.198974E+01   18    11     0     1
            84  -5.79 -6.160440E+01   15    15     0     0
            86  -5.89 -6.203856E+01   16    14     0     0
            88  -6.11 -6.172997E+01   14    16     0     0
            90  -6.16 -6.177755E+01   16    14     0     0
            92  -6.11 -6.204847E+01   20    10     0     0
            94  -6.29 -6.175797E+01   19    11     0     0
            96  -6.42 -6.205258E+01   19    11     0     0
            98  -6.56 -6.194344E+01   21     8     1     0
```

Figure 19.3: Between annealing steps, the Behrens-Fisher package counts the number of simulations in each of the four models and outputs this information to the console log. When the annealing parameter is zero these counts should be roughly equal for the four models. As the annealing parameter increases, the distribution of simulations goes into the posterior probability for the model. Here the program is nearing the completion of the annealing phase, and you can see that the same mean, same variance model is preferred in this data.

After outputting the probabilities from the center section of the table, the program proceeds to sample the joint posterior probability for the parameters for each of the four models. Four separate Markov chain Monte Carlo simulations are run without simulated annealing. The target distributions for these simulations are the integrands of Eqs. (19.12,19.21,19.29,19.36). These four integrands are proportional to joint posterior probability for the parameters given the model indicators. These four sets of samples are then used, along with the model probabilities, to generate samples from the probabilities for the hypotheses given in the bottom section of Fig. 19.2. As each of these phases occur, their passage is noted on in the text window, Fig 19.4 shows an example of this output.

Figure 19.5 is an example of the first section of the BayesBF.mcmc.values report. This first section is pretty standard for McMC values reports, it contains the configuration parameters and the prior probabilities used in the calculations, the names of the input files, the prior probabilities used for the mean and standard deviations. Finally, the last three lines are the names of the original input files. Note that to get this figure to paginate correctly we truncated the priors to three significant digits, the original contained 5.

The next section of the McMC values report, Fig. 19.6, contains a plot of the probability for the four models. It also contains the bounds on both the amplitudes and standard deviations. It contains the computed mean and standard deviations for the first, second and combined data sets. Finally, it contains the posterior probability for each of the four models: $P(SmSv|DI)$, $P(SmDv|DI)$, $P(DmSv|DI)$ and $P(DmDv|DI)$ respectively. These are followed by three sets of three lines. The first set of three lines are the probability the means are the same, $P(Sm|DI)$; the probability the means differ, $P(Dm|DI)$; and the odd ratio either $P(Sm|DI)/P(Dm|DI)$ or $P(Dm|DI)/P(Sm|DI)$ depending on which is greater. The second set of three lines is essentially the same as the first but for the standard deviations rather than the means. Finally, the third set of three lines is the same set of calculations but for the data sets themselves. The reaming lines in Fig. 19.6 are the model

Figure 19.4: Behrens-Fisher Status Listing

```
Sampling  P(C V|SmSv,D1,D2,I)          Same mean,      Same variance
Sampling  P(C V1 V2|SmDv,D1,D2,I)      Same mean,      Different variance
Sampling  P(C1 C2 V|DmSv,D1,D2,I)      Different mean, Same variance
Sampling  P(C1 C2 V1 V2|DmDv,D1,D2,I) Different mean, Different variance
Computing P(C1|D1,D2,I)                The mean in the first set
Computing P(C2|D1,D2,I)                The mean in the second set
Computing P(Sigma 1|D1,D2,I)           The Sd in the first set
Computing P(Sigma 2|D1,D2,I)           The Sd in the second set
Computing P(C1-C2|D1,D2,I)             The Difference in means
Computing P(C1+C2|D1,D2,I)             The Sum of means
Computing P(S1/S2|D1,D2,I)             The Ratio Of Standard Deviations
Computing P(S2/S1|D1,D2,I)
```

Figure 19.4: As each of the probability density functions is sampled the program outputs an indication of what it is doing on console log. Here is a sample of this output. Note in these outputs "V" means Variance. So in the first line the program is trying to tell one that it is sampling the mean and variance of the same mean and same variance model. In the first four lines the program is sampling the parameters for the four basic model. Finally the remaining lines indicate that the program is computing the samples for the probability for C_1, C_2 etc.

independent parameter estimates expressed as $\langle x \rangle \pm \sigma_x$, where x is any one of the parameters. The third column, are the parameters which maximized the joint posterior probability for the parameters.

The remaining sections of this report are the parameter estimates given a particular model. There are four models, so there are four sections. Each section contains the estimated parameters given one of the four models. These estimates will be different for each of the four models, and they may be very different depending on the data. An example of the output for the same mean and different variances is shown in Fig. 19.7. The first four lines of this section summarizes the state of the probability density functions at the end of the Markov chain Monte Carlo simulations. The parameter estimates are the mean, standard deviation, and maximum posterior probability estimates of the parameters. In this model there is a mean and two standard deviation, so there are three parameter estimates.

In addition to outputting this information to the BayesBF.mcmc.values file, histograms computed from the Markov chain Monte Carlo simulations are also output. These histograms may be accessed using the Plot Results Viewer. The plotted output consists of the probability for the model, the model independent parameter estimates, the parameter estimates given each model, and a plot of the logarithm of the posterior probabilities for each simulations for each of the four models as a function of the repeat number. This last set of plots, are meant as aids in determining if the Markov chain Monte Carlo simulations have converged. When the simulations are mixing correctly, these plots should show no upward trend, and the trajectories of the individual simulations should overlap or mix together.

Figure 19.5: Behrens-Fisher McMC Values File, The Preamble

```
Parameter File Listing for the Behrens-Fisher package

! BayesBF Package
! Created 13-Feb-2012 14:17:28  by larry
!
                    Output Dir = BayesOtherAnalysis
            Number Of Abscissa = 1
             Number Of Columns = 1
                Number Of Sets = 2
                     File Name = BayesOtherAnalysis/001.dat
                     File Name = BayesOtherAnalysis/002.dat
              McMC Simulations = 48
                  McMC Repeats = 21
       Minimum Annealing Steps = 21
                Histogram Type = Binned
            Outlier Detection  = Disabled
            Total Mcmc Samples = 1008
                    Kill Count = 4
              Number Of Priors = 2

Param Name    Low      Mean     High    Std Dev    Norm    Prior    Ordered    Param Type
Mean        4.69E+01 5.00E+01 5.31E+01 6.25E-01 -3.22E+00 Gaussian NotOrdered NonLinear
StdDev      8.33E-01 1.04E+00 1.25E+00 4.16E-02 -3.22E+00 Gaussian NotOrdered NonLinear

            Package Parameters = 2
              Input Data Set 1 = BF.smsv.01.dat
              Input Data Set 2 = BF.smsv.02.dat
```

Figure 19.5: The preamble of the BayesBF.mcmc.values file is pretty standard for most packages, it contains the configuration parameters used during the run and it contains the prior probabilities used.

Figure 19.6: Behrens-Fisher McMC Values File, The Middle

```
                    Probability For The Model
          Same Mean   Same Mean   Diff Mean   Diff Mean
          Same Var    Diff Var    Same Var    Diff Var
     Prob.____v___________v___________v___________v_______.Prob
     1.0|                                                 |1.0
     0.9|                                                 |0.9
     0.8|     #                                           |0.8
     0.7|     #                                           |0.7
     0.6|     #                                           |0.6
     0.5|     #                                           |0.5
     0.4|     #                                           |0.4
     0.3|     #                                           |0.3
     0.2|     #                                           |0.2
     0.1|     #           #                               |0.1
     Model.___^___________^___________^___________^_______.Model
          Same Mean   Same Mean   Diff Mean   Diff Mean
          Same Var    Diff Var    Same Var    Diff Var


 The First Input File Is:  BF.smsv.01.dat
 The Second Input File Is: BF.smsv.02.dat

 The Amplitude Lower Bound: 46.92
 The Amplitude Upper Bound: 53.18

 The Standard Deviation Lower Bound: 0.8339
 The Standard Deviation Upper Bound: 1.251

    No.  Standard Deviation    Average      Data Set
    50        1.1505           49.951       F.smsv.01.dat
    50       0.93520           50.146       F.smsv.02.dat
   100        1.0477           50.049       Combined

 ---------------Model----------------    Probability
 Same Mean,      Same Standard Dev        0.8422619
 Different Mean, Same Standard Dev        0.0059524
 Same Mean,      Different Standard Dev   0.1507937
 Different Mean, Different Standard Dev   0.0009921

 The probability the means are the same is:  0.9931
 The probability the means are different is: 0.6944E-02
 The odds ratio is 143. to 1 in favor of the same means

 The probability the standard deviations are the same is:  0.8482
 The probability the standard deviations are different is: 0.1518
 The odds ratio is 5.59 to 1 in favor of the same standard deviations

 The probability the data sets are the same is:  0.8423
 The probability the data sets are different is: 0.1577
 The odds ratio is 5.34 to 1 in favor of the data sets being the same
```

Figure 19.6: The middle section of this report contains a plot of the normalized posterior pro abilities for the four models, It contains the means and standard deviations of the first, second and combined data sets. It contains the posterior probability for the four models. It contains the posterior probability the means are the same. It contains the posterior probability the standard deviations are the same. Finally, it contains the posterior probability the data sets are the same.

Figure 19.7: Behrens-Fisher McMC Values File, The End

```
Model Independent Estimates
Parameter Description              Mean Value     Std. Dev.    Peak Value
C1                                 5.00501E+01    9.28245E-02   5.00431E+01
C2                                 5.00514E+01    9.27814E-02   5.00445E+01
Sigma 1                            1.06436E+00    6.00040E-02   1.04995E+00
Sigma 2                            1.04186E+00    6.02953E-02   1.02866E+00
Diff In Means                     -1.36972E-03    1.25270E-03  -1.36794E-03
Sum Of Means                       1.00102E+02    1.85602E-01   1.00088E+02
Ratio Sig1/Sig2                    1.02446E+00    1.90638E-02   1.02229E+00
Ratio Sig2/Sig1                    9.80543E-01    1.47886E-02   9.70275E-01

Model: SmSv                                          Avg.          Sd.
The Average Log Posterior Probability Was:        -153.1989      0.92696
             The Average Log Prior Params:          -6.1835      0.06631
               The Average Log Likelihood:        -147.0154      0.91732
                        Total Simulations:             1008
                   Probability For the Model:        0.8423

Parameter Description              Mean Value     Std. Dev.    Peak Value
Mean Given Model SmSv              5.00482E+01    1.08521E-01   5.00468E+01
Sd Given Model SmSv                1.05480E+00    7.01560E-02   1.04063E+00
```

Figure 19.7: The final sections of the McMC values file contains a set of model independent parameter estimates. Each estimate consists of the parameter being estimated, the mean value, standard deviation and peak value of the posterior probability. Finally, after these model independent parameter estimates, there are four additional sets of parameter estimates, one estimate for each of the four models. Show here are the estimates from the "SmSv" model: same mean, same variance. These estimates are mean and standard deviation estimates.

Bibliography

[1] Rev. Thomas Bayes (1763), "An Essay Toward Solving a Problem in the Doctrine of Chances," *Philos. Trans. R. Soc. London*, **53,** pp. 370-418; reprinted in *Biometrika*, **45**, pp. 293-315 (1958), and *Facsimiles of Two Papers by Bayes,* with commentary by W. Edwards Deming, New York, Hafner, 1963.

[2] G. Larry Bretthorst (1988), "Bayesian Spectrum Analysis and Parameter Estimation," in *Lecture Notes in Statistics,* **48**, J. Berger, S. Fienberg, J. Gani, K. Krickenberg, and B. Singer (eds), Springer-Verlag, New York, New York.

[3] G. Larry Bretthorst (1990), "An Introduction to Parameter Estimation Using Bayesian Probability Theory," in *Maximum Entropy and Bayesian Methods,* Dartmouth College 1989, P. Fougère *ed.,* pp. 53-79, Kluwer Academic Publishers, Dordrecht the Netherlands.

[4] G. Larry Bretthorst (1990), "Bayesian Analysis I. Parameter Estimation Using Quadrature NMR Models" *J. Magn. Reson.*, **88**, pp. 533-551.

[5] G. Larry Bretthorst (1990), "Bayesian Analysis II. Signal Detection And Model Selection" *J. Magn. Reson.*, **88**, pp. 552-570.

[6] G. Larry Bretthorst (1990), "Bayesian Analysis III. Examples Relevant to NMR" *J. Magn. Reson.*, **88**, pp. 571-595.

[7] G. Larry Bretthorst (1991), "Bayesian Analysis. IV. Noise and Computing Time Considerations," *J. Magn. Reson.*, **93**, pp. 369-394.

[8] G. Larry Bretthorst (1992), "Bayesian Analysis. V. Amplitude Estimation for Multiple Well-Separated Sinusoids," *J. Magn. Reson.*, **98**, pp. 501-523.

[9] G. Larry Bretthorst (1992), "Estimating The Ratio Of Two Amplitudes In Nuclear Magnetic Resonance Data," in *Maximum Entropy and Bayesian Methods,* C. R. Smith *et al.* (eds.), pp. 67-77, Kluwer Academic Publishers, the Netherlands.

[10] G. Larry Bretthorst (1993), "On The Difference In Means," in Physics & Probability Essays in honor of Edwin T. Jaynes, W. T. Grandy and P. W. Milonni (eds.), pp. 177-194, Cambridge University Press, England.

[11] G. Larry Bretthorst (1996), "An Introduction To Model Selection Using Bayesian Probability Theory," in *Maximum Entropy and Bayesian Methods*, G. R. Heidbreder, ed., pp. 1-42, Kluwer Academic Publishers, Printed in the Netherlands.

[12] G. Larry Bretthorst (1999), "The Near-Irrelevance of Sampling Frequency Distributions," in *Maximum Entropy and Bayesian Methods,* W. von der Linden *et al.* (eds.), pp. 21-46, Kluwer Academic Publishers, the Netherlands.

[13] G. Larry Bretthorst (2001), "Nonuniform Sampling: Bandwidth and Aliasing," in *Maximum Entropy and Bayesian Methods in Science and Engineering,* Joshua Rychert, Gary Erickson and C. Ray Smith *eds.*, pp. 1-28, American Institute of Physics, USA.

[14] G. Larry Bretthorst, Christopher D. Kroenke, and Jeffrey J. Neil (2004), "Characterizing Water Diffusion In Fixed Baboon Brain," in Bayesian Inference And Maximum Entropy Methods In Science And Engineering, Rainer Fischer, Roland Preuss and Udo von Toussaint *eds.*, AIP conference Proceedings, **735**, pp. 3-15.

[15] G. Larry Bretthorst, William C. Hutton, Joel R. Garbow, and Joseph J.H. Ackerman (2005), "Exponential parameter estimation (in NMR) using Bayesian probability theory," *Concepts in Magnetic Resonance*, 27A, Issue 2, pp. 55-63.

[16] G. Larry Bretthorst, William C. Hutton, Joel R. Garbow, and Joseph J. H. Ackerman (2005), "Exponential model selection (in NMR) using Bayesian probability theory," *Concepts in Magnetic Resonance*, 27A, Issue 2, pp. 64-72.

[17] G. Larry Bretthorst, William C. Hutton, Joel R. Garbow, and Joseph J.H. Ackerman (2005), "How accurately can parameters from exponential models be estimated? A Bayesian view," *Concepts in Magnetic Resonance*, 27A, Issue 2, pp. 73-83.

[18] G. Larry Bretthorst, W. C. Hutton, J. R. Garbow, and Joseph J. H. Ackerman (2008), "High Dynamic Range MRS Time-Domain Signal Analysis," *Magn. Reson. in Med.*, **62**, pp. 1026-1035.

[19] V. Chandramouli, K. Ekberg, W. C. Schumann, S. C. Kalhan, J. Wahren, and B. R. Landau (1997), "Quantifying gluconeogenesis during fasting," *American Journal of Physiology*, **273**, pp. H1209-H1215.

[20] R. T. Cox (1961), "The Algebra of Probable Inference," Johns Hopkins Univ. Press, Baltimore.

[21] André d'Avignon, G. Larry Bretthorst, Marlyn Emerson Holtzer, and Alfred Holtzer (1998), "Site-Specific Thermodynamics and Kinetics of a Coiled-Coil Transition by Spin Inversion Transfer NMR," *Biophysical Journal*, **74**, pp. 3190-3197.

[22] André d'Avignon, G. Larry Bretthorst, Marlyn Emerson Holtzer, and Alfred Holtzer (1999), "Thermodynamics and Kinetics of a Folded-Folded Transition at Valine-9 of a GCN4-Like Leucine Zipper," *Biophysical Journal*, **76**, pp. 2752-2759.

[23] David Freedman, and Persi Diaconis (1981), "On the histogram as a density estimator: L_2 theory," *Zeitschrift f¨r Wahrscheinlichkeitstheorie und verwandte Gebiete*, **57**, 4, pp. 453-476.

[24] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (1996), "Markov Chain Monte Carlo in Practice," Chapman & Hall, London.

[25] Paul M. Goggans, and Ying Chi (2004), "Using Thermodynamic Integration to Calculate the Posterior Probability in Bayesian Model Selection Problems," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 23rd International Workshop*, **707**, pp. 59-66.

[26] Marlyn Emerson Holtzer, G. Larry Bretthorst, D. André d'Avignon, Ruth Hogue Angelette, Lisa Mints, and Alfred Holtzer (2001), "Temperature Dependence of the Folding and Unfolding Kinetics of the GCN4 Leucine Lipper via 13C alpha-NMR," *Biophysical Journal*, **80**, pp. 939-951.

[27] E. T. Jaynes (1968), "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, pp. 227-241; reprinted in [30].

[28] E. T. Jaynes (1978), "Where Do We Stand On Maximum Entropy?" in *The Maximum Entropy Formalism*, R. D. Levine and M. Tribus *Eds.*, pp. 15-118, Cambridge: MIT Press, Reprinted in [30].

[29] E. T. Jaynes (1980), "Marginalization and Prior Probabilities," in *Bayesian Analysis in Econometrics and Statistics*, A. Zellner *ed.,* North-Holland Publishing Company, Amsterdam; reprinted in [30].

[30] E. T. Jaynes (1983), "Papers on Probability, Statistics and Statistical Physics," a reprint collection, D. Reidel, Dordrecht the Netherlands; second edition Kluwer Academic Publishers, Dordrecht the Netherlands, 1989.

[31] E. T. Jaynes (1957), "How Does the Brain do Plausible Reasoning?" unpublished Stanford University Microwave Laboratory Report No. 421; reprinted in *Maximum-Entropy and Bayesian Methods in Science and Engineering* **1,** pp. 1-24, G. J. Erickson and C. R. Smith *Eds.,* 1988.

[32] E. T. Jaynes (2003), "Probability Theory—The Logic of Science," edited by G. Larry Bretthorst, Cambridge University Press, Cambridge UK.

[33] Sir Harold Jeffreys (1939), "Theory of Probability," Oxford Univ. Press, London; Later editions, 1948, 1961.

[34] John G. Jones, Michael A. Solomon, Suzanne M. Cole, A. Dean Sherry, and Craig R. Malloy (2001) "An integrated ^{2}H and ^{13}C NMR study of gluconeogenesis and TCA cycle flux in humans," *American Journal of Physiology, Endocrinology, and Metabolism,* **281,** pp. H848-H856.

[35] John Kotyk, N. G. Hoffman, W. C. Hutton, G. Larry Bretthorst, and J. J. H. Ackerman (1992), "Comparison of Fourier and Bayesian Analysis of NMR Signals. I. Well-Separated Resonances (The Single-Frequency Case)," *J. Magn. Reson.*, **98**, pp. 483–500.

[36] Pierre Simon Laplace (1814), "A Philosophical Essay on Probabilities," John Wiley & Sons, London, Chapman & Hall, Limited 1902. Translated from the 6th edition by F. W. Truscott and F. L. Emory.

[37] N. Lartillot, and H. Philippe (2006), "Computing Bayes Factors Using Thermodynamic Integration," *Systematic Biology*, **55** (2), pp. 195-207.

[38] D. Le Bihan, and E. Breton (1985), "Imagerie de diffusion in-vivo par rsonance," Comptes rendus de l'Acadmie des Sciences (Paris), **301** (15), pp. 1109-1112.

[39] N. R. Lomb (1976), "Least-Squares Frequency Analysis of Unevenly Spaced Data," *Astrophysical and Space Science*, **39**, pp. 447-462.

[40] T. J. Loredo (1990), "From Laplace To SN 1987A: Bayesian Inference In Astrophysics," in *Maximum Entropy and Bayesian Methods,* P. F. Fougere (ed), Kluwer Academic Publishers, Dordrecht, The Netherlands.

[41] Craig R. Malloy, A. Dean Sherry, and Mark Jeffrey (1988), "Evaluation of Carbon Flux and Substrate Selection through Alternate Pathways Involving the Citric Acid Cycle of the Heart by 13C NMR Spectroscopy," *Journal of Biological Chemistry*, **263** (15), pp. 6964-6971.

[42] Craig R. Malloy, Dean Sherry, and Mark Jeffrey (1990), "Analysis of tricarboxylic acid cycle of the heart using ^{13}C isotope isomers," *American Journal of Physiology*, **259**, pp. H987-H995.

[43] Lawrence R. Mead and Nikos Papanicolaou, "Maximum entropy in the problem of moments," *J. Math. Phys.* **25**, 2404–2417 (1984).

[44] K. Merboldt, Wolfgang Hanicke, and Jens Frahm (1969), "Self-diffusion NMR imaging using stimulated echoes," Journal of Magnetic Resonance, **64** (3), pp. 479-486.

[45] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller (1953), "Equation of State Calculations by Fast Computing Machines," Journal of Chemical Physics. The previous link is to the Americain Institute of Physics and if you do not have access to Science Sitations you many not be able to retrieve this paper.

[46] Radford M. Neal (1993), "Probabilistic Inference Using Markov Chain Monte Carlo Methods," technical report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.

[47] Jeffrey J. Neil, and G. Larry Bretthorst (1993), "On the Use of Bayesian Probability Theory for Analysis of Exponential Decay Data: An Example Taken from Intravoxel Incoherent Motion Experiments," *Magn. Reson. in Med.*, **29**, pp. 642–647.

[48] H. Nyquist (1924), "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, **3**, pp. 324-346.

[49] H. Nyquist (1928), "Certain Topics in Telegraph Transmission Theory," *Transactions AIEE*, **3**, pp. 617-644.

[50] William H. Press, Saul A. Teukolsky, William T. Vetterling and Brian P. Flannery (1992), "Numerical Recipes The Art of Scientific Computing Second Edition," Cambridge University Press, Cambridge UK.

[51] Emanuel Parzen (1962), "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics* **33**, 1065–1076

[52] Karl Pearson (1895), "Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material," *Phil. Trans. R. Soc. A* **186**, 343–326.

[53] Murray Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function," *Annals of Mathematical Statistics* **27**, 832–837 (1956).

[54] Jeffery D. Scargle (1981), "Studies in Astronomical Time Series Analysis I. Random Process In The Time Domain," *Astrophysical Journal Supplement Series*, **45**, pp. 1-71.

[55] Jeffery D. Scargle (1982), "Studies in Astronomical Time Series Analysis II. Statistical Aspects of Spectral Analysis of Unevenly Sampled Data," *Astrophysical Journal*, **263**, pp. 835-853.

[56] Jeffery D. Scargle (1989), "Studies in Astronomical Time Series Analysis. III. Fourier Transforms, Autocorrelation Functions, and Cross-correlation Functions of Unevenly Spaced Data," *Astrophysical Journal*, **343**, pp. 874-887.

[57] Arthur Schuster (1905), "The Periodogram and its Optical Analogy," *Proceedings of the Royal Society of London*, **77**, p. 136-140.

[58] Claude E. Shannon (1948), "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, **27**, pp. 379-423.

[59] John E. Shore, and Rodney W. Johnson (1981), "Properties of cross-entropy minimization," *IEEE Trans. on Information Theory*, **IT-27,** No. 4, pp. 472-482.

[60] John E. Shore and Rodney W. Johnson (1980), "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. on Information Theory*, **IT-26** (1), pp. 26-37.

[61] Devinderjit Sivia, and John Skilling (2006), "Data Analysis: A Bayesian Tutorial," Oxford University Press, USA.

[62] Edward O. Stejskal and Tanner, J. E. (1965), "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient." Journal of Chemical Physics, **42** (1), pp. 288-292.

[63] D. G. Taylor and Bushell, M. C. (1985), "The spatial mapping of translational diffusion coefficients by the NMR imaging technique," Physics in Medicine and Biology, **30** (4), pp. 345-349.

[64] Myron Tribus (1969), "Rational Descriptions, Decisions and Designs," Pergamon Press, Oxford.

[65] P. M. Woodward (1953), "Probability and Information Theory, with Applications to Radar," McGraw-Hill, N. Y. Second edition (1987); R. E. Krieger Pub. Co., Malabar, Florida.

[66] Arnold Zellner (1971), "An Introduction to Bayesian Inference in Econometrics," John Wiley and Sons, New York.