Abstract
Recent advances in computational statistics have greatly simplified the implementation of Bayesian analyses in the day-to-day practice of statistics. This paper provides an overview of Bayesian methods, comparing and contrasting them with classical methods, and emphasizing the central role of the likelihood function in parametric analyses. We will outline some of the recent computational methods, particularly the Gibbs algorithm, Sampling-Importance Resampling, and Data Augmentation, and show how they can be used to generate posterior distributions for many quantities of interest. Examples will include normal theory models (including constraints) and binary regression techniques. The algorithms to be discussed make extensive use of the SAS macro language, SAS/IML™, and SAS DATA steps to augment the results of SAS PROC steps.

Introduction
My perspective is that of an applied statistician, who uses statistical methods as a tool to help scientists make decisions. In particular, I work with product developers in a competitive marketplace. Typically, an idea for product improvement or cost reduction will go through a series of refinements using various cycles of lab tests, special panels, field tests, and consumer tests before it is either dropped, put on the "back burner" or sold in the marketplace. Typically, the product developers are experienced with the product and the market. Further, since the market is quite competitive, resulting in need to get there first. (There is also considerable competition for internal resources.) In this environment I am faced with a rapidly changing series of related questions & problems. The challenge is to integrate multiple sources of uncertain information to give the best answer available, including some idea of the uncertainty of the my answer.

Classical & Bayesian Statistics
Classical frequency methods of statistics depend on two key ideas. The first of these is a frequency interpretation of probability, and the second is the notion of sampling distributions for fixed parameters. Probability is interpreted in terms of the long-run frequency of independent trials of the same event. For example, the probability of a heads in a series of coin flips is defined in terms of the limiting frequency of a heads over a long series of coin flips. This interpretation lends itself naturally to discussing the properties of statistical methods in terms of how repeated samples from a given distribution will behave. For example, a confidence interval on a parameter such as a mean is really a statement about how often a random sampled interval will cover the mean of the parent population. Similarly a "hypothesis test" and its associated "p-value" are statements about the how unusual a given value would be in repeated samples from a certain fixed population. While these methods and interpretation are all well and good in an abstract sense, one has to ask how pertinent they to question at hand.

In particular, consider the definition of probability. Using this interpretation does not allow You¹ to consider such questions as:

- "What is the probability that it will rain during tomorrow's picnic?"
- "How likely is it that we will achieve a 10% market share next quarter?"
- "What is the probability that Product A is the best of my four alternates?"
- "Are these two products at parity?"

Each of these questions, which are all quite reasonable, have no meaning if You use a classical interpretation of probability; and, using frequentist techniques, are unanswerable. The first two questions talk about a single event in the future, while the last two questions are asking about several "constants".

You can avoid these difficulties by broadening Your interpretation of probability to include subjective interpretations. Now probability becomes a statement of your beliefs about the some state of nature about which you do not have complete knowledge. In particular it can be interpreted as a statement about "fair odds" for a trivial gamble on the event. There is a long history on interpretations of probability cited in some of the references attached to this article.

¹We use the You/Your convention to refer to the entity providing or evaluating the probabilities or making decisions.
Under this interpretation, the last two questions given above now pose no particular problem. The parameters of interest can still be constant (i.e., not vary over time), but you are still uncertain about their values, and you wish to make decisions based on their uncertain values.

If we go back to the second of the two items above, the sampling distribution of a statistic, this presents another problem. While the issues addressed in the sampling distribution approach can be quite useful before you have data, when you wish to consider what might happen in some future experiment, it does not help when you wish to make inferences from observed data about an uncertain parameter. Consequently, the frequentist is in a bit of a bind when making practical inferences and must resort to talking about his/her confidence in the techniques and relying on assumptions about data that might be observed from experiments that won't be done.

How does a Bayesian approach the problem? There is a basic theorem in probability, named after Rev. T. Bayes, which demonstrates that

\[ P(B|A) = \frac{P(B)P(A|B)}{P(A)} \]

where \( P(A) = P(\text{Event } A \text{ occurs}), \)

\( P(B) = P(\text{Event } B \text{ occurs}), \)

and

\( P(B|A) = P(\text{B occurs given A}). \)

In words, it says that when you have an initial probability \( P(B) \), and you observe \( A \), you need to adjust your probability for each \( B \) by the likelihood of seeing \( A \) for that \( B \). This can be extended to continuous values as follows

\[ [\theta|x] = \frac{[\theta][x|\theta]}{[x]} , \]

where \([\theta] \) is the prior distribution of \( \theta \),

\([x|\theta] \) is the likelihood of \( x \) for fixed \( \theta \), and

\([\theta|x] \) is the pdf of \( \theta \) for observed \( x \).

The \([\theta] \) term is called the prior and reflects your uncertainty about the parameter(s) before you obtained the data. Likewise, the \([\theta|x] \) is called the posterior and summarizes your uncertainty after you obtained the data. The final term, \([x|\theta] \), is likelihood function of the data, and reflects the agreement of the data and the parameter value(s).

If you examine the equations above, you can see how it prescribes that data should affect your uncertainty about the values that \( \theta \) can assume. The value of \([\theta] \) for any given \( \theta \) reflects the weight you would attach to \( \theta \) before observing any data. Multiplying it by \([x|\theta] \) weights it proportionally to the likelihood of the data at that point.

Alternatively, if you look at it from the perspective of the likelihood, the prior acts a smooth constraint on the values of the parameters. In classical analysis, you can use hard constraints to estimate various parameters, e.g., specifying that a parameter must be greater than some minimum, for example. This would be equivalent to a prior that is zero at or below the minimum, and some positive value above the minimum. Adopting a prior that smoothly increases above the minimum to some constant value has the effect of pulling the posterior away from that minimum.

A final way of looking at the effect of a prior is to recognize the prior as a way of specifying additional data for your analysis. In fact, this data-based approach is often used as a means of eliciting a prior. The statistician can work with scientist to think of what data would be consistent with her current knowledge and express that as data.

Where does \([\theta] \) come from? As mentioned above, it reflects what you originally know about \( \theta \). If you already know a lot, \([\theta] \) will be relatively well defined. If you know very little, it should be quite diffuse. If you wish to say you know nothing at all about it, \([\theta] \) becomes a constant, and the posterior is the likelihood. In this last situation, \([\theta] \) can be an improper prior. Typically, the prior includes data from related experiments and/or personal knowledge about the subject matter.

A common concern about this approach is that it allows two analysts with different uncertainties to reach different conclusions from the same data. This is true. However, as you are well aware, the same thing would happen using classical frequentist approaches. However, a Bayesian analysis of the data will often cause the differing opinions to converge to a common value. Box and Tiao (1973) provide an illustrative example.

A related concern is the fact that the posterior depends on some "subjective" prior, which could "unduly" influence the results. Again this is true, but a
complete analysis should now include discussion of the sensitivity of the conclusions to the particular form of the prior. In fact, Berger (1985) has argued that this form of analysis is even more objective than a frequentist result, since the analyst must specify (and justify) her assumptions as part of the analysis. In any event, even a classical analysis is "subjective" in that the analyst makes a number of assumptions about the form of the model.

Inference and Estimation

All inference and estimation is based on the posterior distribution. For example, a probability interval for some function of the parameters can be based on the interval containing the highest posterior density, similar to the construction of a confidence interval. Additionally, You can examine the entire distribution and convey considerably more information than with the simple interval.

Hypothesis testing is also based on Bayes Theorem. A further benefit of this method is that multiple hypotheses can be evaluated simultaneously, since they can be expressed as a mixture prior. For example, suppose You are comparing the means of two products and wished to know if there was a difference. You might have:

- $H_A$: Product A is better than B (prior: $\pi_A$)
- $H_B$: Product B is better than A (prior: $\pi_B$)
- $H_0$: Products A & B are at parity (prior: $\pi_0$)

(You have to define parity here.) After observing some relevant data $x$ and choosing Your likelihood functions $\lambda()$, You would have

$$P(H_A|x) = \pi_A \cdot \lambda(x|H_A),$$
$$P(H_B|x) = \pi_B \cdot \lambda(x|H_B),$$
$$P(H_0|x) = \pi_0 \cdot \lambda(x|H_0).$$

These probabilities can be combined with the relevant cost/benefit information to reach a decision.

Comparison with Frequentist Results.

How different are these results from a "Classical" analysis? This depends on the amount of prior information You add to the analysis. As we mentioned earlier, a very diffuse prior leads to the same results as a maximum likelihood analysis. In fact, numerous authors have pointed out that almost all classical models are special cases of Bayesian analyses with a flat prior. Consequently, there will be no numerical difference between a likelihood based analysis and a posterior based analysis. The difference here is only in the interpretation of the numbers.

However, the addition of a prior allows much richer models to be treated and estimation of effects that could not normally be estimated. For example, in a logistic regression with class effects, some parameters might be impossible to estimate because very few events took place. The addition of a prior can effectively add data to the experiment and allow You to assess the associated parameters. Additionally You can now consider complex hierachical models that allow very rich applications. Lange et al. (1992) give an example in AIDS modeling where they estimated about the same number of parameters as data points. They were able to do this through the use of prior data from related studies and by building a hierachical structure into the prior, so that parameters on one level constrain parameters on other levels. Others have been quite successful in modeling complex systems using this hierachical approach.

One area where the two approaches can give different results in hypothesis testing. Non-informative one sided hypothesis tests give similar results to a one-sided "p-value". However, two sided approaches can yield quite different results, with the frequentist tests producing wildly over-optimistic results. In this case, the results depend critically on the specification of the distribution under the alternate hypothesis and the prior probability of the null hypothesis. See Berger (1985) or Lee (1989) for details.

Computation Methods

All of these benefits aren't free. First, You have to spend more time thinking about the physical problem that You are analyzing so as to construct a reasonable prior/model for the data. A side benefit we've observed is that we sometimes discover that we already have the data we need and can avoid the experiment entirely. A second cost is the computational and programming time required to reach and evaluate a solution. There are very few packages of routines, such as those in SAS, which do it all for You. Currently, You need to spend some time carefully setting up Your analysis. We will consider a number of algorithms to implement these analyses.
The available methods can be grouped into several categories:

- Non-informative analyses of classical models
- Conjugate methods & Approximations
- Numerical Quadrature
- Sampling-Importance Resampling (SIR)
- Data Augmentation
- Monte Carlo Markov Chain

As we mentioned above, the simplest case of a Bayesian Analysis is with a non-informative prior. For most normal theory models, these are the same as their classical counterparts. The interpretation is quite different.

**Conjugate methods & Approximations.**

Certain priors are simple to analyze for many standard models. A prior distribution is conjugate if the posterior has the same (simple) form as the prior. For example, if you assume the data is from a normal distribution with a known variance and choose a normal prior, the posterior is also normal. If you have binomial data and use a beta form for the prior, the posterior is also beta. These methods can actually be extended quite far, since a mixture of conjugate priors will give a mixture of conjugate posteriors. In the SAS System, these methods can be implemented by capturing the estimates in an OUTEST type data set and manipulating them in a SAS DATA or SAS/IML step. Alternatively, you can develop approximations to the posterior by using methods such as Laplace approximations (Tanner, 1991).

**Numerical Quadrature.**

An alternate method, suitable for models involving a few parameters, is to numerically evaluate the likelihood. The solution is quite specific to the problem and can involve considerable programming. Dellaportas & Wright (1991) is a starting point for further reading.

The next three methods involve Monte Carlo techniques, which use various algorithms to produce samples from the posterior distribution. An disadvantage of these techniques is that the can consume a fair amount of computer time. An advantage is that the resulting data can be easily processed to yield marginal and conditional distributions for further analysis and summarization.

**Sampling-Importance Resampling (SIR).**

The Sampling-Importance Resampling (SIR) algorithm (Rubin, 1987) is a two pass Monte Carlo algorithm for generating random samples from the posterior. This method starts with an initial distribution that is easy to sample and is close to the posterior \( \theta \). Sample \( \theta_i (i = 1, 2, \ldots, n) \) from this starting distribution. For each \( \theta_i \), evaluate the density \( f(\theta_i) \), the posterior \( \pi(\theta_i) \), and compute and save the ratios \( r_i = \pi(\theta_i) f(\theta_i) / f(\theta_i) \). During a second pass, take the desired number of random samples from this list, sampling proportionately to the (normalized) weights. Alternatively, treat the data set as if it were a weighted random sample from the posterior. Gelfand and Smith (1992) and Tanner (1991) provide further details. Example 1 shows an implementation of this method for logistic regression.

**Data Augmentation**

Tanner & Wong (1987) suggested a method that is analogous to the EM algorithm. This method depends on the posterior having a certain form such that the problem would be easy to solve if certain data were observed. They noted that this could be solved iteratively by first sampling that produce a complete data set and then solving the complete problem. This process is then repeated multiple times and the resulting samples combined to form the posterior distribution. Strictly speaking, this is a form of Markov Chain Monte Carlo.

**Markov Chain Monte Carlo (MCMC)**

This is an area where there is much active research. Several other papers in this session also address this topic. The SIR algorithm doesn't directly involve recent samples from the posterior in the computation and sampling of the next sample. The MCMC methods do. Methods like these were initially proposed by Hastings (1970), who adapted the Metropolis method to statistical estimation problems. In general, these methods involve iteratively sampling from the posterior distribution where the sampled distribution depends on previously sampled values. A special case, known as Gibbs Sampling, provides a good starting point.

Gibbs Sampling can be used when the posterior can be broken down into a number of simple components, each of which is simple to sample from when the values of the other parameters are known. Denote the parameter vector by \( \theta \), and its components by \( \theta_i \), and their conditional posteriors by \( \pi(\theta_i | \theta_{-i}) \). The components may also be vector valued. Then the algorithm proceeds as follows:

1. **Initialization:** Set \( \theta \) to an initial value.
2. **Iteration:** For each component \( i = 1, 2, \ldots, p \), update \( \theta_i \) from the current value of \( \theta_{-i} \) using its conditional posterior distribution.
3. **Sampling:** Repeat the iteration until convergence is achieved.

This process allows for efficient sampling from complex distributions, often leading to more accurate estimates compared to alternative methods.
1. Set \( \theta \) to an initial value \( \theta^{(0)} \), and set \( j=1 \).

2. Sample

\[
\begin{align*}
\theta_1^{(j)} & \text{ from } [\theta_1|x, \theta_2^{(j-1)}, \theta_3^{(j-1)}, \theta_4^{(j-1)}], \\
\theta_2^{(j)} & \text{ from } [\theta_2|x, \theta_1^{(j)}, \theta_3^{(j-1)}, \theta_4^{(j-1)}], \\
\vdots
\end{align*}
\]

to yield \( \theta^{(j)} \).

3. Repeat Step 2 for \( j=2, \ldots, k \).

4. Keep every \( r \)th sample for the final estimate of \( [\theta|x] \).

The posterior can then be summarized using functions of the form \( \Sigma g(\theta|x) \), where \( g(.) \) denotes the particular statistic of interest and \( j \) indexes the final sample. There are a number of considerations having to do with how big \( k \) and \( r \) should be, whether to restart the algorithm occasionally, how long an initial burn-in period should be, and so on. These are discussed in detail in the references given below. We also note in passing that this method can be used for likelihood based analyses.

**Examples**

In the following sections we give examples showing how these methods may be implemented in SAS, using SAS/IML and the SAS data steps.

**Example 1. Logistic Regression using SIR.**

This example uses the vaso data set from Example 3 of Chapter 37 (The LOGISTIC Procedure) of the SAS/STAT manual. That example uses this data to show the how to examine influential observations. We use it to show the SIR algorithm and to compare the maximum likelihood estimates with the posterior distributions.

```sas
/* Use MLE for starting values */
PROC LOGISTIC data=vaso outest=outest covout;
model y=logvol lograte; run;

PROC IML; /* generates samples from posterior */
use outest var{intercep logvol lograte};
need = 20000; inflate=3 ;
eps = 0.000001; eps2 = 1-eps ;
read all into X; /* import LOGISTIC values */
mu = X[1,] ; /* separate estimates and */
vc = X[(2 3 4)].; /* variance covariance matrix */
/* generate estimates from asymptotic N(\mu,\Sigma) */
/* keeping MLE */
sd= half(inflate*vc) ;
z = [0 0 0] / normal((need,3,0));
est = (mu@((need+1,1,1) + z*sd)) ;
dens = -z[##] ; /* density of sampled values */
llike= J((need+1),1,0) ; /* storage for log-likelihood*/
/* evaluate log-likelihood for sampled parameters*/
use vaso var{y logvol lograte} ;
do data ;
read next into tmp ;
resp = tmp[1,1] ;
tmp[1,1]=1; /*intercept*/
xbeta=est@tmp ;
pred = exp(xbeta)/(1+exp(xbeta)) ;
pred = ((pred>eps)>eps2) ;
llike=llike+(1-resp)#log(pred)+resp#log(1-pred) ;
end ;
/* ratio is the weight for the posterior */
ratio = llike/dens ;
prob = ratio/sum(prob) ; /* normalize for sampling*/
/* create data set for later processing */
out = est || prob ;
names = {"Int" "LogVol" "LogRate" "Prob"};
create postr from out [colname=names];
append from out ;
close postr ;
quit;
```
Figure 1, attached, compares the marginal posterior calculated using this method (solid line) with that calculated using the asymptotic distribution (dotted line). The maximums agree for the three parameters, but the posteriors are noticeably skewed compared to the usual MLE estimates. This may be due to the outliers discussed in the SAS/STAT manual, but this clearly indicates the difference in the two approaches.

Example 2 - Probit Regression Using Data Augmentation and Gibbs Sampling

This example shows how two related approaches can be combined to yield usable estimates. In this case, we note that if we knew the latent continuous variable $Z$ that gave rise to the observed binary variable $Y$, the problem would reduce to a simple regression analysis, the solution to which is well known for non-informative and conjugate priors (Zellner, 1971). This can be extended to include censored regression, ordinal responses, other link functions, and other error distributions, such as a $t$ distribution with unknown degrees of freedom (Albert and Chibbs, 1993).

Note that it is not necessary to assume that $Z$ exists, as the same results can be achieved by assuming that the $Z$'s are transforms of a parameter vector $\pi$, the values of which are related by a linear model, i.e. we are fitting a hierarchical model with $\pi$ as the first level and assuming that $\text{probit}(\pi) \sim N(X^*\beta, \Sigma)$, with a non-informative prior on $\beta$.

We use the same data as in Example 1. Only the SAS/IML code is shown below, as the results are substantially identical to those presented in Example 1 (except for a change of sign.)

Zegler and Karim (1991) provide details on the implementation of the Gibbs algorithm in generalized linear models with random effects and apply it to two logistic regression examples.

%MACRO BRProbit(Data=_LAST_,
    Resp=Y, Predict=X, outer=50, inner=20);
...
This macro samples the posterior distribution of the coefficients for a probit binary regression using the Albert-Chib algorithm. (Data Augmentation and Gibbs sampling from the Augmented posterior.)

Output Data Set

The function creates a data set called OUTBETA containing the $\text{OUTER}$ samples from the posterior

```
%MACRO BRProbit(Data=_LAST_,
    Resp=Y, Predict=X, outer=50, inner=20);

PROC IML ; read all var{&Predict} into X[ColName=XcName];
read all var{&Resp } into y[ColName=YcName];
Start Estim{X,Y,MLE) ;
    Maxiter = 50 ; Converge = 10e-6 ; iter = 1 ;
    BLS = Inv(X"Xj*X'*Y ;
    Do While (Iter < Maxiter) ;
        XB= X*BLS ;
        EXB = Exp(XB)/(1+Exp{XB);
        DLMLE = X"*(Y-EXB) ;
        Exb2 = EXB/(1+Exp(XB)) ;
        DL2MLE = - X"*(Exb2 # X) ;
        MLE = BLS - Inv(DL2MLE)*DLMLE ;
        if Max(Abs(MLE-BLS)) < Converge then do;
            enditer = iter ; iter = Maxiter ;
        end;
        else do; BLS = MLE ; iter=iter+1 ; end ;
    end;
Finish ;
```
Example 3 • Constraints On Normal Theory Models

Our final example will demonstrate the use of the Gibbs algorithm to sample values the posterior distribution for a heteroscedastic ANOVA model with order constraints on the means. This example was given in the article by Gelfand et al. (1990). This article provides further discussion on the Gibbs Sampler and additional examples.

In this case, the Gibbs algorithm allows you to easily handle the constraints that would be difficult using standard methods: Each parameter is estimated conditional on the others, so that the problem reduces to sampling from a truncated Normal distribution. The model is:

\[ Y_t | \theta_1, \sigma_1^2, \ldots, \mu_j, \tau_j^2, \sigma_j^2 \sim N(\mu_j, \tau_j^2) \]

where \( \theta_1, \sigma_1^2, \mu_j, \tau_j^2, \sigma_j^2 \) are assumed known. The sufficient statistics are used in this model. In the example from the paper \( k=5, n_i = 2^{i+4} \) and the observations were drawn from a \( \mathcal{N}(\mu, \tau^2) \) distribution. The data are presented in Table 1:

<table>
<thead>
<tr>
<th>Sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_i )</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>mean, ( \bar{y} )</td>
<td>0.3191</td>
<td>2.0343</td>
<td>3.539</td>
<td>6.398</td>
<td>4.811</td>
</tr>
<tr>
<td>( s_i^2 )</td>
<td>0.2356</td>
<td>2.471</td>
<td>5.761</td>
<td>8.758</td>
<td>19.670</td>
</tr>
</tbody>
</table>

Figure 2 shows the results of this method. The scatterplot matrix clearly shows the effect of the constraints on the joint distribution, as does the marginal plot. This plot also shows the non-normal characteristics of the resulting posteriors.

This technique can be extended, for example, to deal with multiple rankings of the data by assigning a prior that gives a some probability to each ranking. The analysis could either estimate the posterior values of these rankings, or treat the prior weights as fixed sampling probabilities.
%let ninner = 50;
%let nsamples = 1000;
%let ncells = 5;
data postr;
  set gelfand end=eof;
  array y[&ncells] _temporary_;
  array v[&ncells] _temporary_;
  array n[&ncells] _temporary_;
y mean = avg;
  v mean = var;
n mean = SampSize;
  if eof;
  nC = &ncells;
*[- constants to prevent under/overflows ];
eps = 0.00000001; epsrange = 1-2*eps;
*[-Prior values for hyperparameters ];
mustart = 0;
  sigma0sq = 10000;
a1 = 0.5; b1 = 1; a2 = 0.5; b2 = 1;
taustart = b2/(a2);
*[- Starting Values for the cell values ];
array theta[&ncells];
do i = 1 to &ncells; theta[i] = y[i];
  array sigysq[&ncells];
do i = 1 to &ncells; sigysq[i] = v[i];
  do sample=1 to &nsamples ;
  *[- Gibbs outer loop ];
  *[- initialize hyperpars for Gibbs ];
  mu = mustart;
  tausq = taustart;
  do inner = 1 to &ninner ;
  *[- Gibbs inner loop ];
  do i = 1 to &ncells ;
    *[- cell means ];
    *[- first, the unconstrained prior ];
    c = n[i]*y[i]*tausq + mu*sigysq[i];
    c = c/(sigysq[i]+n[i]*tausq);
    d=sqrt(sigysq[i]*tausq/(n[i]*tausq+sigysq[i]));
    if i = 1 then do ;
      *[- enforce the constraints ];
      lower = 0;
      upper = probnorm((theta[i+1]-c)/d);
      end;
    else if i eq &ncells then do;
      lower = probnorm((theta[i-1]-c)/d);
      upper = 1;
end;
else do;
  upper = probnorm((theta[i+1]-c)/d);
  lower = probnorm((theta[i-1]-c)/d);
end;
U = lower + Ranuni(0)*(upper-lower);
U = eps + U*epsrange;
theta[i] = c+d*probit(U)*[constrained sample];
  *[- cell variances ];
  a = a1+ 0.5*n[i];
  b = b1+ ((n[i]-1)*v[i]+n[i]*(y[i]-theta[i])**2)/2;
  *[- sigysq ~ IG(a,b. );
  *[- sample as b/z for [z]=G(a) ];
  sigysq[i] = b/(rangam(0,a));
end;
  *[- generate the next mu ];
  c = mean(of theta1-theta&nCells);
  mu = c + sqrt(tausq/&ncells)*ranor(0);
  *[- generate the next tau**2 ];
  dsq = var(of theta1 - theta&nCells);
  a = a2+&ncells/2;
  b = b2+ ((nC-1)*dsq+nC*(c-mu)**2)/2;
  tausq = b/rangam(0,a);
  end;
  *[- end inner loop ];
output; *[- save estimate ];
  keep theta1-theta&nCells mu tausq
  sigysq1-sigysq&nCells;
end;
run;

Conclusions

Bayesian methods allow You to adopt a coherent approach to data analysis, that permits You to incorporate all of the relevant information into Your analyses and decisions. A Bayesian approach to experimentation and decision making forces You to think about what You know and the relevancy of Your data to the problem at hand. An added benefit of this method is that it allows You to map initial beliefs from other decision makers through the same data to evaluate their post-data agreement and/or disagreement.

The SAS System provides a wealth of tools, particularly SAS/IML and the SAS DATA step, to implement these techniques.
Acknowledgments

I have benefitted from discussions with my colleagues R. Gopalan and P. Ramgopal. The initial version of the macro presented in Example 2 was written by R. Gopalan.

SAS, SAS/IML, and SAS/GRAPH are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration.

References


Scatterplot Matrix Of Posterior Values for Example 3 - Order Constraints

Marginal Posteriors for Example 3 - Order Constraints