BAYESIAN MCMC ANALYSES FOR REGULATORY ASSESSMENTS OF FOOD COMPOSITION

Jay M. Harrison

Derek Culp

George G. Harrigan

BAYESIAN MCMC ANALYSES FOR REGULATORY ASSESSMENTS OF FOOD COMPOSITION

Jay M. Harrison, Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167
Derek Culp, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513
George G. Harrigan, Monsanto Company, 800 N. Lindbergh Blvd., St. Louis, MO 63167

Abstract

In order to gain regulatory approval to market a new seed product derived with biotechnology, grain and forage composition data must be collected from field trials, and summaries must be reported to various government agencies. Currently, both tests of differences in composition between a genetically modified organism (GMO) and its control and tests of equivalence of the GMO to conventional genotypes are required by regulatory agencies. Bayesian analyses offer an attractive option for regulatory assessments by expressing results that can be interpreted more easily by a wide audience and by providing more ways to examine various hypotheses of interest. In order to extend Bayesian methodology for application to different compositional analytes, and to take advantage of the information obtained in previous experiments, the use of informative prior distributions for composition studies is proposed. Methods for determining suitable informative prior distributions analytically are shown in four situations: (1) eliciting opinions from an expert, (2) finding the best fit from an overdetermined set of summary statistics from one previous study, (3) performing a meta-analysis of summary statistics from previous studies with an assumed common prior distribution, and (4) performing a different meta-analysis with the prior distribution determined by a mixture of different assumed prior distributions from previous studies. Examples from soybean composition studies are used to illustrate these techniques.

Keywords: Bayes, Markov Chain Monte Carlo, Prior Distribution, Genetically Modified Organism, Regulatory, Composition

1. Introduction

Regulatory composition trials are typically conducted by planting different genotypes of the crop plant in multiple locations using randomized complete block designs, harvesting the plants at maturity, and conducting compositional assays on the harvested grain and forage. In these trials, the newly derived genotype is compared against a control of a similar genetic background. Other genotypes, called references, can be included in the trials to represent the natural variation that occurs among commercially available or conventional genotypes. Different references may be used at different locations to accommodate for differences in growing conditions by region.

Currently, various regulatory agencies require traditional frequentist hypothesis tests from analysis of variance (ANOVA) models for evaluating variation in composition. For example, the European Food Safety Authority (EFSA) currently requires both a test for the difference in composition between a genetically modified organism (GMO) and its control and a test of

equivalence of the GMO to conventional genotypes, or references, using a different ANOVA model (EFSA, 2010). However, results of significance tests are not necessarily indicative of concerns with safety or nutritional wholesomeness, and the observed differences between the GMO and conventional comparators are small when compared to the natural variability in crop composition. A recent review of GMO corn and GMO soybean showed that over 97% of all comparisons in which a "statistically significant difference" ($p < 0.05$) between GM and conventional comparators was observed had a relative magnitude of difference less than 20% (Harrigan, et al., 2010).

Guidelines for Bayesian analyses for regulatory composition studies have been proposed (Harrison, et al., 2011; Harrigan and Harrison, 2012). Advantages of Bayesian analyses that were cited in these articles include the simplified interpretation of the results for a wide audience, the elimination of the requirement to test for differences and equivalence with different procedures, and the capability to quantify differences with meaningful summaries, such as the posterior distribution of the percentage difference in means between a GMO and its control and the probability that a GMO mean lies within the range of means of the reference genotypes. The authors cited a guidance document produced by the U. S. Food and Drug Administration (U. S. FDA, 2010) for medical devices as a precedent for Bayesian analysis in a regulatory setting. The authors also stressed the importance of model diagnostics to evaluate the goodness of fit of a Bayesian model.

Bayesian data analysis with Markov Chain Monte Carlo (MCMC) methods requires the specification of prior distributions for the parameters in the model. The prior distribution for a parameter reflects the degree of belief that the parameter will assume given values. Low-information prior distributions, such as normal distributions with very large variances, are useful when there is absolutely no prior information for a parameter or when an objective approach is desired. Harrison, et al. (2011) used the fact that amounts of fat and protein in soybean composition studies are measured as percentages to justify the use of low-information uniform prior distributions over the range (0%, 100%) for the means of the soybean genotypes and over the range (-100%, 100%) for nuisance effects in the ANOVA model, such as location effects. Improper prior distributions, such as flat prior distributions that assign a constant value to all real numbers, can be used to perform the calculations, but some authors describe problems with such distributions. For example, Christensen, et al. (2011) demonstrated that flat prior distributions assign virtually all weight to parameter values that are larger than any reasonable value, and proper posterior distributions will not always result from the use of improper prior distributions (Gelman, et al., 2004; Carlin and Louis, 2009).

Informative prior distributions define more narrow ranges of probable values for the parameter and assign regions of higher prior density to the more probable values. Philosophically, an informative prior distribution is the mathematical formalization of the scientific principle that knowledge is accumulated through experience, and new research is always presented in the context of previous research (Kruschke, 2011). The use of informative prior distributions provides other advantages for Bayesian analysis. For example, the United States Food and Drug

Administration (U. S. FDA) (2010) cited the potential for informative prior distributions to justify the reduction of sample sizes or durations of medical device trials.

Harrison, et al. (2011) used linear models with multiple parameters to represent the genotype effects that were primarily of interest, as well as nuisance effects for locations, replicates within locations, genotype-by-location interaction, and random error. Informative prior distributions may be needed when a model has many parameters in order to obtain convergence of the posterior distributions. Gelman, et al. (2004) wrote that there are no clear general principles for defining noninformative prior distributions for models with many parameters. Carlin and Louis (2009) stated that, in models with large numbers of parameters, the information in the data may be insufficient to identify all of the parameters, so informative prior distributions are necessary for some or all of the parameters in such cases. For these reasons, informative prior distributions will be necessary in order to extend Bayesian analyses to a wider range of compositional analytes that are not bounded between 0% and 100% and to studies with large numbers of genotypes, replicates, or locations.

An informative prior distribution can be assigned in several ways. For example, if a Bayesian analysis from a previous, similar study is available, then the posterior distribution of a parameter may serve as the prior distribution of the parameter in the next study, if the exchangeability of data between the studies can be justified (U. S. FDA, 2010). Harrison, et al. (2010) started with low-information prior distributions for a study conducted in one year. Then, to model the nuisance effects in data that were collected in the following year, they used informative hierarchical prior distributions that were derived from the corresponding posterior distributions from the first year.

If Bayesian analyses from a prior study are not available, but other statistical summaries from one or more previous studies can be located, then an informative prior distribution can be formed using that information with a meta-analytic approach. Another way of constructing an informative prior distribution is with elicitation, or interviewing an expert on the subject matter about the anticipated value of the parameter and then writing equations to express those beliefs numerically. According to the U. S. FDA (2010), Bayesian methods are usually less controversial when the prior distribution is based on empirical evidence and not elicited from experts.

The statistician may also incorporate desirable mathematical features into the construction of an informative prior distribution. Physical constraints involving the parameter may be included. For example, if a parameter represents the mean concentration of a substance, then a distribution that assigns positive density only to nonnegative values should be chosen for an informative prior distribution. For many situations, unimodality is another desired characteristic of a prior distribution, since it allows a certain value to have the highest prior density. Gelman, et al. (2004) wrote that the prior distribution should include all plausible values of the parameter, but the prior distribution does not necessarily need to be concentrated around the true value of the parameter. In some situations, the statistician may choose to use a conjugate prior distribution so

that the posterior distribution will have the same parametric form as the prior distribution (Gelman, et al., 2004).

For computational simplicity, the statistician may also choose to use a prior distribution that is included in a software package for performing Bayesian analyses, although different distributions may be programmed. For example, the "zeroes trick" with a Poisson distribution and the "ones trick" with a Bernoulli distribution can be used to introduce an arbitrary prior distribution in WinBUGS software (Ntzoufras, 2009; Lunn, et al., 2000). Spiegelhalter, et al. (2007) noted problems with high autocorrelation, high Monte Carlo error, and poor convergence with these methods, so long runs are necessary to achieve suitable results with these methods in WinBUGS. With SAS PROC MCMC, the GENERAL and DGENERAL options can be used to specify a new probability distribution (SAS Institute Inc., 2009).

The impact of an informative prior distribution on the inferences from the subsequent prior distribution can be quite dramatic. A prior distribution can be too informative, in the sense that the current data have little influence on the posterior distribution. Justifications for the selection of a particular prior distribution, evaluations of the goodness of fit of the prior distribution to their contributing assumptions or previous data, and evaluations of the impact of the prior distribution on the subsequent posterior distribution are essential.

2. Informative prior distributions derived by elicitation

Elicitation involves the formation of a prior distribution of a parameter by matching the probability distribution for the parameter to the descriptions provided by interviewing a client or an expert on the subject matter. Strategies for these interviews, and software packages for deriving prior distributions based on the responses, were provided by Berger (1985), Carlin and Louis (2009), and Christensen, et al. (2011). For one example, O'Hagan, et al. (2006) recommended the elicitation of estimates of the quantiles near the center of the distribution, such as the 25^{th}, 50^{th}, and 75^{th} percentiles, in order to form a prior distribution. Carlin and Louis (2009) warn that the prior distributions obtained in this manner are not necessarily unique. For example, a standard Cauchy distribution and a normal distribution with mean 0 and variance 2.19 have the same values for these three percentiles, and the distributions seem similar, but the resulting posterior distributions can be quite different (Berger, 1985). For another example, Christensen, et al. (2011) cited an example in which two distinct beta distributions were found to share the same mode and 66^{th} percentile.

Christensen, et al. (2011) provided the following example of elicitation. Suppose that a parameter θ represents the probability of success in a binomial distribution. Suppose that an expert believes that the most likely value of θ is 0.2, and the largest reasonable value that θ could assume, with 95% probability, is 0.45. The functional form of the chosen prior distribution is a beta distribution with parameters α and β. The beta distribution assigns positive probability to all values between 0 and 1, so the constraints involving the probability of success are included. The beta distribution also offers computational convenience, since it is implemented in popular Bayesian software packages.

Next, values of α and β must be chosen to reflect the opinions of the expert. These can be expressed mathematically as follows:

1. $(\alpha-1)/(\alpha+\beta-2) = 0.2$
2. $P(0 < \theta < 0.45 \mid \theta \sim \text{Beta}(\alpha,\beta)) = 0.95$
3. $\alpha>1$
4. $\beta>1$

Equation 1 is the mode of the beta distribution, and Equation 2 represents the 95th percentile. Inequalities 3 and 4 reflect the constraints that are required for the resulting distribution to be unimodal. If Inequalities 3 and 4 are both satisfied, the mode and a percentile uniquely determine a beta distribution (Christensen, et al., 2011).

PROC MODEL, which is available in the SAS/ETS package, can be used to estimate solutions of unknowns in a system of one or more nonlinear equations (SAS Institute Inc., 2009). The SAS code for the examples, including computations with PROC MODEL and graphs of results, are available by request from the third author (george.g.harrigan “at” monsanto “dot” com). All SAS procedures were performed using SAS Version 9.2 (SAS Institute Inc., 2008).

In this example, the equations that represent the mode and 95th percentile were phrased in terms of residuals with expectations of zero in PROC MODEL. For example, the equation for the mode was expressed as $(\alpha-1)/(\alpha+\beta-2) - 0.2$. Either of the functions QUANTILE or CDF in SAS could be used to specify Equation 2 in a similar way. The RESTRICT statement incorporated Inequalities 3 and 4 into the estimation routine. Starting values of $\alpha=2$ and $\beta=8$ were entered in FIT statements because the default starting values of 0.0001 do not meet the parameter restrictions, and such small values caused estimation errors in the routine. The starting values were assigned by assuming that the mode should be close to the mean of the beta distribution, which is $\alpha/(\alpha+\beta)$. PROC MODEL provided parameter estimates of $\alpha=3.3$ and $\beta=10.2$, which agreed with those derived by Christensen, et al. (2011).

Informative prior distributions should be plotted and checked for suitability. The solutions produced by PROC MODEL may not necessarily be unique, but any solution that adequately represents the prior beliefs about the parameter may be used. Figure 1 demonstrates that the Beta (3.3, 10.2) prior distribution exhibits the two properties from the elicited description. Such plots should be presented to the expert for confirmation.

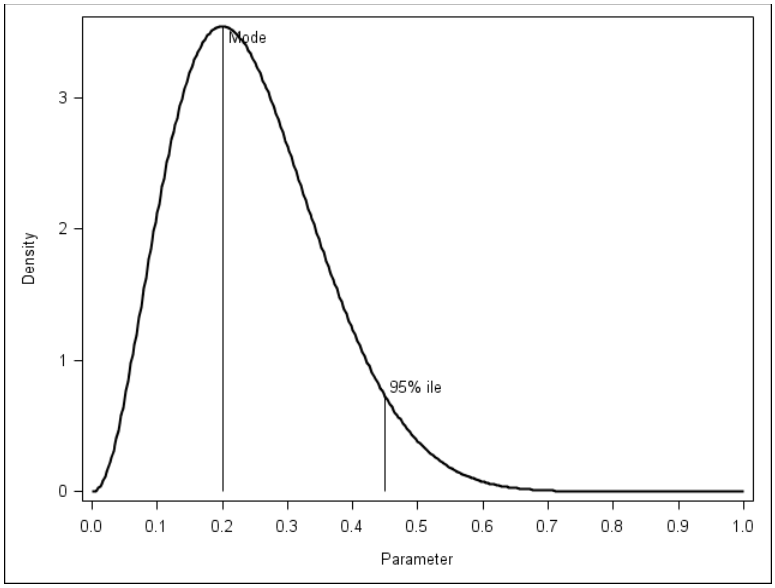

Figure 1. Beta prior distribution from elicitation

3. Informative prior distributions derived from a single study

An example from a soybean study can be used to illustrate the derivation of an informative prior distribution based on a description of compositional data that was provided in a single previous study. Hou, et al. (2009) studied the levels of various sugars in different varieties of soybeans. For sucrose, they described the distribution as normal and provided seven sets of summary statistics to describe the distribution. This information may be used to construct an informative prior distribution for a parameter representing the mean level of sucrose in a particular soybean variety. One tractable distribution will not correspond exactly to all of these details, but a reasonable solution for the overdetermined system of equations may be drawn by minimizing the errors between the properties of one distribution and the specific details in the text. PROC MODEL in SAS provides a convenient routine for this task.

Two sets of statistical properties are incorporated into the solution. To use the information about the two smallest and two largest observations, properties of order statistics can be employed. Let $\{X_1, \ldots, X_n\}$ represent a random sample of size n from a population with a continuous cumulative distribution function F(x), and let $\{U_1, \ldots, U_n\}$ represent a random sample of size n from the standard uniform distribution. Then, $F(X_{i:n})$ is equal in distribution to $U_{i:n}$, where i:n represents the i^{th}-smallest order statistic. Since equality in distribution implies equality of moment-generating functions, the expected value of $F(X_{i:n})$ is equal to the expected value of $U_{i:n}$, which follows a beta distribution with parameters $\alpha=i$ and $\beta=n-i+1$. Thus, the expected value of $F(X_{i:n})$ is $i/(n+1)$ (Arnold, et al., 1992). For example, the minimum observed value for sucrose out of 241 soybean varieties was 1.6 mg g^{-1}, so one contribution to the derivation of the informative prior distribution F can be obtained by approximating a solution for $F(1.6) = 1/(241+1)$.

Another statistical property to be incorporated involves the prior distribution. Hou, et al. (2009) claimed that the sucrose levels followed a normal distribution. This would assign slight probability to values below zero, but, physically, the level of a component must be at least zero. In this situation, a truncated normal distribution that assigns zero probability to negative values would maintain the general appearance of normality while preventing parameter values below zero. To facilitate the use of this distribution with PROC MODEL, PROC FCMP in SAS can be used to write a new function that provides the cumulative density function of a truncated normal distribution (SAS Institute Inc., 2009). For values of X truncated to be above zero, the cumulative distribution function for the truncated normal distribution can be expressed as $F(x) = \{\Phi[(x-\mu)/\sigma] - \Phi(-\mu/\sigma)\} / [1 - \Phi(-\mu/\sigma)]$, where Φ represents the cumulative distribution function for the standard normal distribution (Johnson, et al., 1994). In order to use this new function with PROC MODEL, the FUNCDIFFERENCING option must be used to allow numerical differentiation with multiple applications of the new function.

The resulting parameter estimates were μ=43.7 and σ=16.2. The ESTIMATE statement was used to provide an estimate of the mean (43.9 mg g^{-1}) and standard deviation (16.0 mg g^{-1}) of the truncated normal distribution as functions of the parameters μ and σ (Johnson, et al., 1994). PROC MODEL also produced residuals with respect to each equation. These residuals showed that the fitted model agreed fairly closely with the cited summary statistics. The plot of the derived truncated normal distribution in Figure 2 includes illustrations of the summary statistics from the article. The plot shows general agreement between the statistics from the previous study and the prior distribution that was derived. PROC MODEL provides options to accommodate other model features, such as correlation among the residuals.

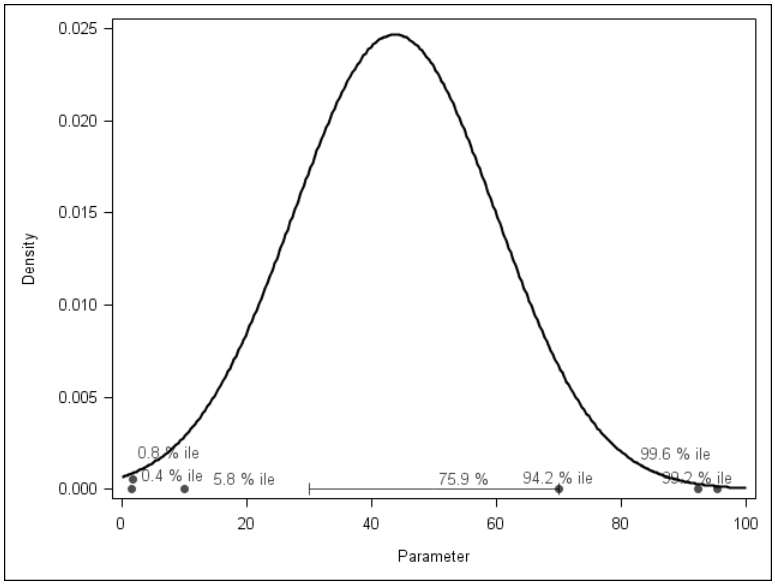

Figure 2. Truncated normal prior distribution from summary statistics

4. Informative prior distributions from a mixture of distributions over multiple studies

Often, the parameters for an informative prior distribution will not be as thoroughly described as in the sucrose example, and the prior distribution must be derived with minimal information from multiple studies. The following method estimates a prior distribution as an approximation of the distribution of the mixture of results from previous studies.

A search was conducted for literature that reported isoflavone levels in soybean seed. The search was conducted objectively using the following terms: soybean, isoflavones, composition, and nutrients. As isoflavone levels in soybeans have been the subject of many published papers in the last 40 years, the literature obtained was reviewed to ensure that consistent methodology was used to determine total isoflavone levels (aglycones). Briefly, citations that utilized acid hydrolysis procedures to reduce all isoflavones down to their aglycone structures (daidzein, glycitein, and genistein) conducted during the last ten years were selected for inclusion (Hutabarat, et al., 2001; Seguin, et al., 2004; Primomo, et al., 2005; Lundry, et al., 2008; Morrison, et al., 2008; Berman, et al., 2009; Devi, et al., 2009; Zhou, et al., 2011). This is the most routinely used methodology in regulatory studies. In this example, the ranges and sample sizes of values that were reported within each of the cited studies were used to form an informative prior distribution for the parameters representing genotype means of daidzein in an upcoming study. The units of measurement were $\mu g\ g^{-1}$ of dry weight.

By applying the fact that the expected value of $F(X_{i:n})$ is $i/(n+1)$ for a given cumulative distribution function F (Arnold, et al., 1992), a set of estimates for the n individual values within each study can be generated. For this example, a lognormal distribution was assumed for each study. The lognormal distribution assigns a positive prior density to the set of nonnegative numbers and is skewed to the right. In SAS, the lognormal distribution has two parameters, μ and σ, that are the mean and standard deviation, respectively, of the natural logarithms of the values. SAS was used to estimate the values of μ and σ separately for each study from the minimum and maximum values with quantiles $1/(n+1)$ and $n/(n+1)$, respectively. Next, the values of a new variable were generated to represent the quantiles $2/(n+1)$, …, $(n-1)/(n+1)$ from the corresponding lognormal distribution. Finally, all of the original extremes and the new observations were pooled, and summary statistics and graphics were used to determine a suitable functional form for a prior distribution from this mixture of imputed observations from previous studies. PROC UNIVARIATE was used to provide these summaries and to evaluate the goodness of fit of an assumed lognormal distribution. PROC UNIVARIATE returned parameter estimates of $\mu=6.37$ and $\sigma=0.53$, corresponding to a mean of 669 $\mu g\ g^{-1}$ and a standard deviation of 378 $\mu g\ g^{-1}$. Figure 3 shows the ranges from previous studies, the imputed values based on percentiles, and the derived prior distribution.

The mixture of distributions method can be extended in several ways. For example, if a uniform distribution is assumed for each previous study, the imputed values may be produced with simple linear interpolation between the extremes. Instead of using the expected values of the order statistics, a set of n imputed observations may be randomly generated from the assumed distribution for each study. The previous studies do not necessarily need to share the same

distributional form. Finally, other distributions, such as a gamma distribution, may be fitted to the imputed data and compared to each other for goodness of fit.

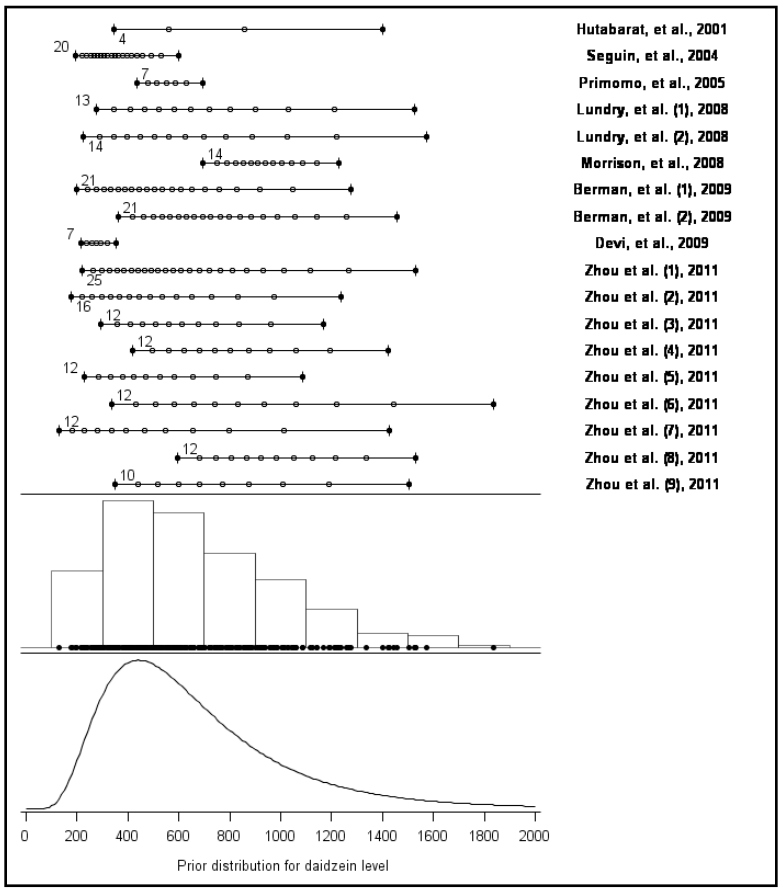

Figure 3. Lognormal prior distribution from mixture of distributions

5. Informative prior distributions from a common distribution over multiple studies

PROC MODEL in SAS can also be used to perform another type of meta-analysis to generate an informative prior distribution. Under the assumption that the same prior distribution applies to these previous studies and the next study, and that all of these studies are exchangeable, then the following approach may be suitable. Unlike the mixture approach that was described previously, this approach assigns equal weight to each study, regardless of the number of genotypes that were observed in each study. This technique is more computationally intensive than the mixture approach; however, it also provides some useful diagnostic analyses and extensions.

This technique was applied with PROC MODEL using the same data that were used in the previous example. A lognormal distribution was chosen for the informative prior distribution. Two equations were defined to represent the minimum and maximum from each study, expressed as functions of the parameters μ and σ with the CDF function. PROC MODEL

provided best-fit estimates of μ=6.4 and σ=0.6, corresponding to a mean of 738 $\mu g\ g^{-1}$ and a standard deviation of 493 $\mu g\ g^{-1}$. For comparison, the mixture of distributions method in the previous example returned a mean of 669 $\mu g\ g^{-1}$ and a standard deviation of 378 $\mu g\ g^{-1}$ for a lognormal prior distribution, illustrating the difference between weighting previous observations equally and weighting previous studies equally.

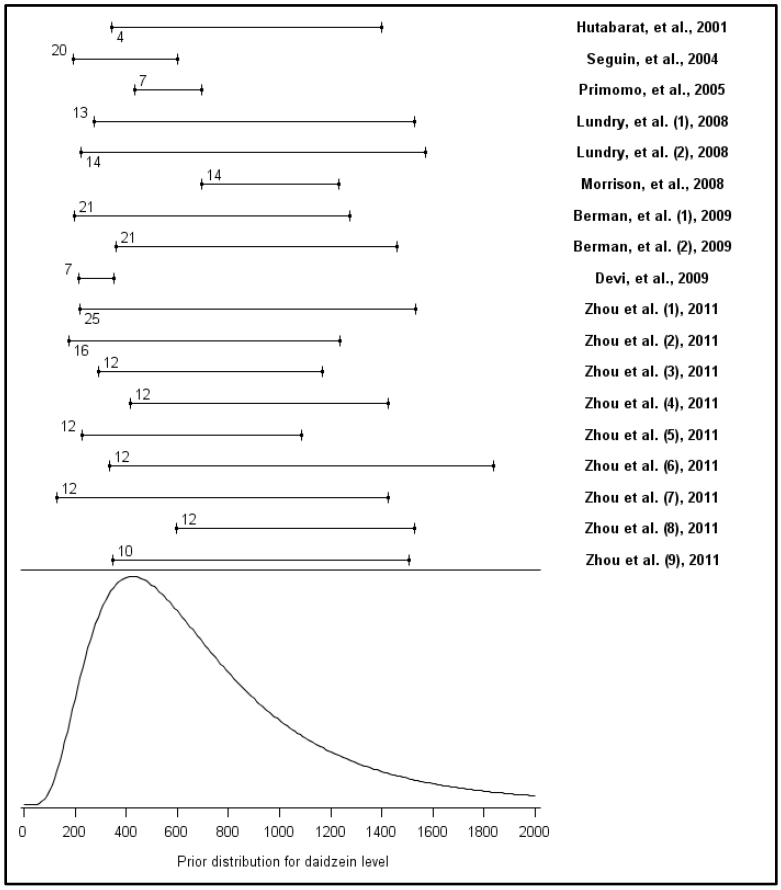

Figure 4. Lognormal prior distribution from assumed common distribution

Along with the parameter estimates of μ and σ, PROC MODEL provided their associated standard errors. If desired, these standard errors may be used in a Bayesian analysis to indicate the uncertainty of the parameters of the prior distribution. For example, instead of assuming μ=6.4 for the prior distribution, the parameter μ could be assigned a normal hyperprior distribution with mean 6.4 and standard deviation 0.1. Likewise, an informative prior distribution that allows correlations among the parameters in the model may be used. The COVB option in PROC MODEL provides a listing of the covariance matrix for the estimates, which could be used to specify a bivariate prior distribution for μ and σ.

Various diagnostic results involving the residuals were provided by PROC MODEL, such as the mean square error of the residuals for the equations for the minima and maxima. The PLOTS(UNPACKPANEL) option produced a series of plots for assessing the goodness of fit of

the estimates, including plots of the studentized residuals for both equations. Such diagnostics could be used to identify unusual studies from the list of previous studies that were used to form the prior distribution.

6. Facilitation of Bayesian sensitivity analyses

The derivation of an informative prior distribution in an objective, analytical manner, as shown in the previous examples, should help to alleviate any concerns that an informative prior distribution was chosen specifically to support a certain predetermined conclusion with a Bayesian analysis. The sensitivity of the analyses to the choice of the prior distribution may be assessed by using two or more prior distributions and comparing the results from the resulting analyses (Gelman, et al., 2004). Using the residual analyses that are provided by PROC MODEL, alternative prior distributions may also be compared before fitting a Bayesian model to new data.

To illustrate the approach with two prior distributions, a truncated normal distribution that is common to all studies may be used instead of a lognormal distribution. The new function for a truncated normal cumulative density function was defined using PROC FCMP and applied to the daidzein data, and the results were compared to the results from the previous example. The results are illustrated in Figure 5.

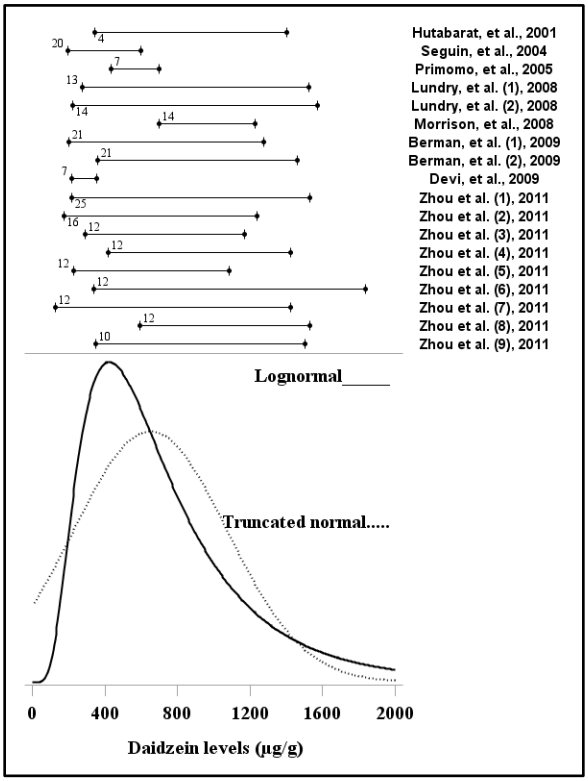

Figure 5. Comparison of lognormal and truncated normal distributions

The mean square errors of the residuals were compared, with lower values indicating a closer fit. The derived truncated normal distribution matched the minima from previous studies more closely than the lognormal distribution, but it did not match the previous maxima as closely. The same observations having standardized residuals of large magnitude with the lognormal distribution were also identified as unusual observations for the truncated normal distribution. In this example, the truncated normal prior distribution did not offer a clear advantage over the lognormal prior distribution.

Comparisons of the posterior distributions resulting from multiple models could be conducted to examine the sensitivity of the analyses to the choice of the prior distribution. Such comparisons could include: (1) overlaid density plots of the competing prior distributions for evaluation of features such as skewness and kurtosis, (2) density plots of the prior distributions overlaid with the corresponding sample statistics from the new study for evaluation of features such as outliers, (3) density plots of the prior distributions overlaying a histogram of the samples from the corresponding posterior distributions after fitting the Bayesian model, (4) plots of the posterior means and credible intervals of the same parameter from the competing distributions, and (5) plots of the posterior predictive distributions of sample statistics, such as the sample mean and variance (Gelman, et al., 2004). Chen (2010) provided SAS code for plotting posterior predictive distributions and Bayesian posterior p-values from PROC MCMC in SAS.

7. Summary

Bayesian methods provide the advantages of expanded and simplified interpretation in the analyses of crop composition data for regulatory reviews. In order to extend Bayesian analyses to a wide variety of compositional analytes, informative prior distributions will be necessary. Analytical methods are available for deriving informative prior distributions for subsequent Bayesian analysis from information that is elicited from an expert or by using previous results of one or more studies. These analyses allow evaluations of goodness of fit, as well as construction of sophisticated models with hyperparameters, nonstandard prior distributions, and multivariate prior distributions. Sensitivity analyses may be conducted by using the same routines to generate different prior distributions, then comparing the results from the corresponding Bayesian models. Graphical analyses are encouraged to assess the appropriateness of the chosen prior distribution. By following objective, transparent construction of prior distributions and providing thorough evaluations of goodness of fit, Bayesian approaches have the potential to be incorporated into regulatory reviews of crop composition data.

8. Acknowledgements

The authors wish to thank the following contributors. Kristina Berman, Denise Lundry, and Matt Breeze of Monsanto Company provided literature reviews. Shi Zhao and Rubin Wei provided assistance with the computations during their internships with Monsanto Company. Sanjay Matange of SAS Institute Inc. and Susan Riordan of Monsanto Company provided samples of SAS Graph Template Language code for use in constructing graphs for meta-analysis results. Fang Chen of SAS Institute Inc. provided helpful comments and suggestions, including the idea

for the mixture of distributions method for constructing a prior distribution. Rob Agnelli of SAS Institute Inc. provided assistance with the FCMP procedure. Two anonymous reviewers provided beneficial comments and suggestions.

9. References

Arnold, B. C., Balakrishnan, N., Nagaraja, H. N., 1992. A First Course in Order Statistics. John Wiley & Sons, New York.

Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis (second ed.), Springer-Verlag, New York (1985).

Berman, K.H., Harrigan, G. G., Riordan, S. G., Nemeth, M. A., Hanson, C., Smith, M., Sorbet, R., Zhu, E., Ridley, W. P., 2009. Compositions of seed, forage, and processed fractions from insect-protected soybean MON 87701 are equivalent to those of conventional soybean. Journal of Agricultural and Food Chemistry 57, 11360-11369.

Carlin, B. P., Louis, T. A., 2009. Bayesian Methods for Data Analysis, third ed. CRC Press, Boca Raton, FL.

Chen, F., 2011. Practical Bayesian computation. 2011 Joint Statistical Meetings, Miami Beach, FL.

Christensen, R., Johnson, W., Branscum, A., Hanson, T. E., 2011. Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians. CRC Press, Boca Raton, FL.

Devi, M.K.A., Gondi, M., Sakthivelu, G., Giridhar, P., Rajasekaran, T., Ravishankar, G. A., 2009. Functional attributes of soybean seeds and products, with reference to isoflavone content and antioxidant activity. Food Chemistry 114, 771-776.

European Food Safety Authority (EFSA) Panel on Genetically Modified Organisms (GMO), 2010. Scientific opinion on statistical considerations for the safety evaluation of GMOs. EFSA Journal 8(1): 1250.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, second ed. Chapman & Hall/CRC, Boca Raton, FL.

Harrigan, G.G., Lundry, D., Drury, S., Berman, K., Riordan, S. G., Nemeth, M. A., Ridley, W. P., Glenn, K. C., 2010. Natural variation in crop composition and the impact of transgenesis. Nature Biotech 28, 402–404.

Harrigan, G. G., Harrison, J. M., 2012. Assessing compositional variability through graphical analysis and Bayesian statistical approaches: Case studies on transgenic crops. Biotechnology and Genetic Engineering Reviews 28, 15-32.

Harrison, J. M., Breeze, M. L., and Harrigan, G. G., 2011. Introduction to Bayesian statistical approaches to compositional analyses of transgenic crops 1. Model validation and setting the stage. Regulatory Toxicology and Pharmacology 60, 381-388.

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., Shi, A., 2009. Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Science 49, 903-912.

Hutabarat, L.S., Greenfield, H., Mulholland, M., 2001. Isoflavones and coumestrol in soybeans and soybean products from Australia and Indonesia. Journal of Food Composition and Analysis 14, 43-58.

Kruschke, J. K., 2011. Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press, Burlington, MA.

Johnson, N. L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, Volume 1, second ed. John Wiley & Sons, New York.

Lundry, D. R., Ridley, W. P., Meyer, J. J., Riordan, S. G., Nemeth, M. A., Trujillo, W. A., Breeze, M. L., Sorbet, R., 2008. Composition of grain, forage, and processed fractions from second-generation glyphosate-tolerant soybean, MON 89788, is equivalent to that of conventional soybean (Glycine max L.). Journal of Agricultural and Food Chemistry 56, 4611-4622.

Lunn, D. J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS -- A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing 10, 325-337.

Morrison, M. J., Cober, E. R., Saleem, M. F., McLaughlin, N. B., Fregeau-Reid, J., Ma, B. L., Yan, W., Woodrow, L., 2008. Changes in isoflavone concentration with 58 years of genetic improvement of short-season soybean cultivars in Canada. Crop Science 48, 2201-2208.

Ntzoufras, I., 2009. Bayesian Modeling Using WinBUGS. John Wiley & Sons, Hoboken, NJ.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., Rakow, T., 2006. Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons, Chichester, UK.

Primomo, V.S., Poysa, V., Ablett, G. R., Jackson, C.-J., Rajcan, I., 2005. Agronomic performance of recombinate inbred line populations segregating for isoflavone content in soybean seeds. Crop Science 45, 2203-2211.

SAS Institute Inc., 2008. SAS Software Release 9.2 (TS2M3). SAS Institute Inc., Cary, NC.

SAS Institute Inc., 2009. The FCMP Procedure. In SAS/STAT 9.2 Base SAS 9.2 Procedures Guide. SAS Institute Inc., Cary, NC.

SAS Institute Inc., 2009. The MODEL Procedure. In SAS/ETS 9.22 User's Guide. SAS Institute Inc., Cary, NC.

SAS Institute Inc., 2009. The MCMC Procedure. In SAS/STAT 9.2 User's Guide. SAS Institute Inc., Cary, NC.

Seguin, P., Zheng, W., Smith, D. L., Deng, W., 2004. Isoflavone content of soybean cultivars grown in eastern Canada. Journal of the Science of Food and Agriculture 84, 1327-1332.

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2007. WinBUGS User Manual, Version 1.4.3. http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf

U.S. Food and Drug Administration (Center for Devices and Radiological Health), 2010. Guidance for the use of Bayesian statistics in medical device clinical trials. U.S. Food and Drug Administration, Rockville, MD. http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf

Zhou, J., Harrigan, G. G., Berman, K. H., Webb, E. G., Klusmeyer, T. H., Nemeth, M. A., 2011. Stability in the composition equivalence of grain from insect-protected corn and seed from glyphosate-tolerant soybean over multiple seasons, locations and breeding germplasms. Journal of Agricultural and Food Chemistry 59, 8822-8828.