ACT Research Report Series 2002-2

Bayesian Statistical Inference for Coefficient Alpha

Jun Corser Li

David J. Woodruff

ACT

January 2002

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243-0168

Bayesian Statistical Inference for Coefficient Alpha

Jun Corser Li
David J. Woodruff

Abstract

Coefficient alpha is a simple and very useful index of test reliability that is widely used in educational and psychological measurement. Classical statistical inference for coefficient alpha is well developed. This paper presents two methods for Bayesian statistical inference for a single sample alpha coefficient. An approximate analytic method based on conjugate distributions is derived. This method is easy to compute. A second method uses MCMC methodology as implemented by the computer program WinBUGS. WinBUGS may be downloaded for free from the Internet, and this paper includes WinBUGS code for making Bayesian inferences about coefficient alpha. Psuedo-randomly generated data are use to compare the two Bayesian methods to each other and both of those methods to the classical method. The results indicate that the two Bayesian methods work well so long as the number of items and examinees are not too small.

Acknowledgements

The authors thank the following persons for their comments, suggestions, and support: Dr. Hiroshi Watanabe, Dr. Shin-ichi Mayekawa, and Dr. Jim Sconing.

Bayesian Statistical Inference for Coefficient Alpha

Frequentist inferential procedures for coefficient alpha are well developed. The paper by Feldt, Woodruff, and Salih (1987) presents a summary of the different methods, gives complete references to the area, and also discusses the robustness of the procedures to violations of their assumptions. A more recent paper by Hakstian and Barchard (2000) also evaluates the robustness of the procedures and references other recent studies on the robustness of the procedures. The methods depend on normal distribution theory for random and mixed effects ANOVA. Even though the methods are often employed using dichotomous item response data, the procedures generally perform well when there are reasonable numbers of items and examiness. Two papers by Feldt and Ankenmann (1998, 1999) contain classical power curves and tables for testing the difference between two independent sample alpha coefficients or two dependent sample alpha coefficients. A more recent paper on the sampling distribution of coefficient alpha is VanZyl, Neudecker, and Nel (2000). Their paper takes a multivariate approach instead of the ANOVA approach used in the earlier papers.

The present paper derives two Bayesian procedures for making inference about a single alpha coefficient. Our procedures also depend on normal distribution theory for random and mixed effects ANOVA. Our first procedure is an approximate analytic method based on conjugate distributions. This method is relatively simple and easy to compute. Our second method uses MCMC methodology as implemented by the WinBUGS computer program. Example WinBUGS code is included in Appendix B. We compare the two Bayesian procedures to each other and both to a frequentist method.

One advantage gained from using Bayesian techniques is the ability to incorporate disparate but relevant information into the analysis by way of the prior distribution. Another advantage is the ability to combine data from different analyses by using the posterior distribution obtained from a previous analysis as the prior distribution for the next analysis. These properties can be useful to test developers creating new tests. If the new tests are related to previous tests, then information about the reliability of the earlier tests can be used to develop an initial prior distribution for the new tests. In addition, test developers initially may have to use shorter versions of the tests and administer these versions at different times and to different small groups of students. In these situations inference can be updated by using the posterior distribution obtained from an earlier analysis as the prior distribution for the next analysis. Test users may want to calculate the reliability of a test in a specific group of examinees of special interest to them, and the test may be administered to only a few examinees at any one time, but at regular intervals over time. The Bayesian process of updating the inference from prior to posterior to new prior could prove convenient in such situations.

Feldt's inferential procedures for alpha is based on the demonstration by Hoyt (1941) that the sample alpha coefficient can be computed from the observed student mean square and the observed interaction mean square of a two-way students by items ANOVA with one observation per cell. Hoyt's result is an algebraic identity. It is true for any two-way table of numbers with one observation per cell.

Using Hoyt's 1941 result, Feldt (1965) developed a frequentist procedure for a single alpha coefficient based on ANOVA normal distribution theory. He considered the two-way examinees by items random effects ANOVA model with one observation per cell, but his results also can be valid under a mixed model. In practice examinees often are randomly sampled from a large population of examinees and the same is sometimes true for items, though the sampling of items is not always strictly random. In these situations the random effects ANOVA model is most appropriate. In other situations only the items actually administered are of interest so items are treated as a fixed effect and the mixed ANOVA model should be used.

VanZyl, Neudecker, and Nel (2000) derive the same result as Feldt, but their derivation is based on the multivariate form of the mixed ANOVA model that assumes a compound-symmetric covariance matrix. The derivation of our analytic Bayesian method is similar to the derivation used by VanZyl et. al. (2000), but we use the two-way random effects ANOVA model with WinBUGS.

Methodology

In what follows, the true and sample values of coefficient alpha can take values between $-\infty$ and one. Negative values for coefficient alpha can occur when the inter-item correlations are negative. See VanZyl, Neudecker, and Nel (2000) for additional discussion of this issue. We denote coefficient alpha by ρ and we take as prior distribution for ρ

$$p(\rho) = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')}(1-\rho)^{(\alpha'-1)}e^{-\beta'(1-\rho)} \quad \alpha' > 0,\ \beta' > 0,\ -\infty < \rho < 1. \tag{1}$$

This is a gamma type distribution and we denote it $G_\rho(\alpha', \beta')$. It is convenient to indicate our prior knowledge about ρ by specifying a likely value for ρ. We then can indicate our confidence in this value by specifying a hypothetical prior sample size upon which the value is based. We denote the prior mean of ρ as r', and we assign it our prior estimate of the value of ρ. We next define $n'+1$ as the size of a hypothetical prior sample that indicates the strength of our prior belief. Finally we take $\alpha' = n'/2$ and $\beta' = n'/[2(1-r')]$. Using E, M, and V to denote mean, mode, and variance we find that

$$E(\rho) = r', \tag{2}$$

$$M(\rho) = \left(\frac{n'-2}{n'}\right)r' + \frac{2}{n'}, \text{ and} \tag{3}$$

$$V(\rho) = \frac{2(1-r')^2}{n'}. \tag{4}$$

Having specified values for n' and r' we then can compute from equations (3) and (4) the prior mode and prior variance for ρ. If n' is less than three the prior density will be J shaped and without a well-defined mode. Taking $n'=0$ yields an improper non-informative prior density for ρ. Note that coefficient alpha cannot by computed from a sample of size one.

We recommend specifying the prior mean of ρ and not the prior mode of ρ. When n' is small the prior mean and prior mode can have quite different values and then better results are obtained by specifying a value for the prior mean rather than the prior mode.

Before combining our prior with the likelihood, it is convenient to transform ρ to τ where $\tau = 1/(1-\rho)$. This yields as prior distribution for τ

$$p(\tau) = \frac{(\beta')^{\alpha'}}{\Gamma(\alpha')}(\tau)^{-(\alpha'+1)}e^{-\beta'(\tau^{-1})}, \quad \tau > 0. \tag{5}$$

This is an inverse gamma distribution and we denote it $\mathrm{IG}_\tau(\alpha', \beta')$.

As previously mentioned Feldt (1965) and VanZyl, Neudecker, and Nel (2000) derived a frequentist distribution theory for coefficient alpha. Let $t = 1/(1-r)$ where r denotes the sample alpha coefficient. For a sample that has $n+1$ examinees and $m+1$ items they showed that

$$\frac{t}{\tau} = \frac{(1-\rho)}{(1-r)} \sim \mathrm{F}_{n,nm}. \tag{6}$$

From a Bayesian perspective this F distribution is a marginal or integrated likelihood (Bernardo & Smith, 1994), and we derive it under a Bayesian model in Appendix A.

For nm large the $\mathrm{F}_{n,nm}$ distribution is well approximated by a χ^2_n/n distribution and our Bayesian method of inference is based on this approximation. Considered as a function of τ we write the χ^2_n based likelihood as

$$l(\tau \mid t) = k(n,t)\tau^{-n/2}e^{-nt/(2\tau)} \tag{7}$$

where $k(n,t)$ is a function of n and t only and does not depend on τ. Inspection of equation (7) shows that it is the kernel of an inverse gamma distribution for τ. The inverse gamma distribution is closed under multiplication; consequently, because our prior distribution for τ is also inverse gamma, it follows that application of Bayes theorem yields a posterior inverse gamma distribution for τ. We denote the posterior $\mathrm{IG}_\tau(\alpha'', \beta'')$. Multiplying (5) and (7) together as dictated by Bayes theorem gives the following values for α'' and β'':

$$\alpha'' = \frac{(n' + n)}{2} \text{ and} \tag{8}$$

$$\beta'' = \frac{1}{2}\left[\frac{n'}{(1-r')} + \frac{n}{(1-r)}\right]. \tag{9}$$

We calculate the posterior distribution of ρ by making the transformation $\rho = 1/(1-\tau)$. This yields posterior distribution $G_\rho(\alpha'', \beta'')$ for ρ. Posterior summary indices for ρ can be calculated from equations (8) and (9). They are:

$$\mathrm{E}(\rho \mid r) = \frac{n'r'(1-r) + nr(1-r')}{n'(1-r) + n(1-r')}, \tag{10}$$

$$\mathrm{M}(\rho \mid r) = \frac{n'r'(1-r) + nr(1-r') + 2(1-r')(1-r)}{n'(1-r) + n(1-r')}, \text{ and} \tag{11}$$

$$\mathrm{V}(\rho \mid r) = \frac{2(n'+n)(1-r')^2(1-r)^2}{\left[n'(1-r) + n(1-r')\right]^2}. \tag{12}$$

If an improper non-informative prior is selected by taking $n' = 0$ then equations (10), (11), and (12) will reduce to equations (2), (3), and (4) but with n' and r' replaced by n and r.

The frequentist sampling distribution for the sample alpha coefficient can be derived from (6). It is

$$p(r \mid \rho) = \frac{\Gamma\left(\frac{n}{2}(m+1)\right) m^{(-n/2)}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{nm}{2}\right)(1-\rho)} \left[\frac{1-\rho}{1-r}\right]^{\left(\frac{n}{2}+1\right)} \left[1 + \frac{(1-\rho)}{m(1-r)}\right]^{\left(-\frac{n}{2}(m+1)\right)} \quad -\infty < r < 1. \tag{13}$$

and sampling distribution summary indices for r are:

$$\mathrm{E}(r \mid \rho) = \left(\frac{n}{n-2}\right)\rho - \left(\frac{2}{n-2}\right) \approx \rho \text{ for large } n, \tag{14}$$

$$\begin{aligned} \mathrm{M}(r \mid \rho) &= \left[\left(\frac{n}{n+2}\right)\left(1 - \frac{2}{nm}\right)\right]\rho + \left[1 - \left(\frac{n}{n+2}\right)\left(1 - \frac{2}{nm}\right)\right] \\ \mathrm{M}(r \mid \rho) &\approx \left(\frac{n}{n+2}\right)\rho + \left(\frac{2}{n+2}\right) \text{ for large } nm, \text{ and} \end{aligned} \tag{15}$$

$$\begin{aligned} \mathrm{V}(r \mid \rho) &= \left(\frac{2n(nm+n-2)}{m(n-4)(n-2)^2}\right)(1-\rho)^2 \\ \mathrm{V}(r \mid \rho) &\approx \frac{2(m+1)(1-\rho)^2}{nm} \text{ for large } n. \end{aligned} \tag{16}$$

Our Bayesian method is based on an approximation to the likelihood given in (6), but the sampling distribution of alpha given in (13) is derived exactly from the likelihood given in (6). Some idea of the accuracy of our approximation can be obtained from a comparison between the posterior distribution of alpha based on a non-informative prior and the exact sampling distribution of alpha. Both densities depend only on the data and on the data only through the sufficient statistic r. Comparison of results between the two procedures will be presented in the Results section.

Our second Bayesian procedure uses the MCMC method as implemented by the computer program WinBUGS. We use the normal theory two-way examinees by items ANOVA model with one observation per cell—the same model originally used by Feldt (1965). We treat both examinees and items as random effects. The model is

$$y_{ij} = \mu + a_i + b_j + e_{ij} \tag{17}$$

where a_i denotes the effect of the i-th examinee, b_j denotes the effect of the j-th item, and e_{ij} denotes the the ij-th error component. Following Hoyt (1941) we note that

$$r = \frac{MS_a - MS_e}{MS_a} \tag{18}$$

and from Feldt(1965) we have

$$\rho = \frac{(m+1)\sigma_a^2}{(m+1)\sigma_a^2 + \sigma_e^2}\,. \tag{19}$$

The symbols just introduced have their usual meanings under the two-way random effects ANOVA model (See Box and Tiao (1973) or Sahai and Ageel (2000)) and they are defined in appendix A of this paper.

To use WinBUGS we first needed to generate pseudo-random samples of observations based on the model in equation (17). We took $\mu = 0$ and generated values of a from a Normal(0, σ_a^2) distribution, values of b from a Normal(0, σ_b^2) distribution, and values of e from a Normal(0, σ_e^2) distribution. The values of the two variance components, σ_a^2 and σ_e^2, were chosen so as to give specific values for ρ. Nine such samples were generated, and their characteristics are presented in Table 1. The computer program *JMP* from the SAS Institute was used to generate the samples.

TABLE 1

Characteristics of the Nine Pseudo-Randomly Generated Samples

$p=m+1$	$N=n+1$	ρ	r	m_y	s_y
10	10		0.48	0.057	0.46
10	20	0.5	0.58	0.056	0.44
10	40		0.53	0.082	0.44
20	10		0.65	-0.059	0.46
20	20	0.7	0.73	-0.047	0.46
20	40		0.71	-0.015	0.45
35	10		0.92	-0.081	0.52
35	20	0.9	0.92	-0.070	0.51
35	40		0.91	-0.018	0.52

We input into WinBUGS the model given in equation (17) along with the accompanying normal distribution assumptions for the model effects. The basic model parameters are μ and the three variance components (though WinBUGS uses precisions which are the inverses of the variances). The parameter of interest is ρ and as can be seen in equation (19) it is a function of the two variance components σ_a^2 and σ_e^2. WinBUGS uses the distributions of the basic model parameters and the data to generate the marginal posterior distribution of the functional parameter ρ.

We wanted to compare the WinBUGS method to the frequentist method so we gave the basic model parameters very diffuse, nearly non-informative prior distributions. We gave μ a Normal(0.0, 10,000) prior distribution and the inverses of the three variance components Gamma(0.0001, 0.0001) distributions which have means of unity and standard deviations of 100. An annotated example of our WinBUGS code is presented in Appendix B.

Results

We compare the classical sampling distribution of r, as given in equation (13), with the posterior distribution of ρ obtained from our approximate Bayesian method using an improper non-informative prior. In Figures 1 through 9 in

Appendix C both densities are plotted for the values of alpha, m, and n given in Table 1. Inspection of Figures 1 through 9 show that when n is small both tails of the sampling distribution of alpha are shifted left in comparison to the tails of the Bayesian posterior distribution. The left tail of the sampling distribution can be especially elongated when n is small. This causes the Bayesian posterior variance to be smaller than the sampling distribution variance even though the shapes of the two distributions are closely matched around their nearly equal modes.

Inferential statistics for ρ from the classical sampling distribution method (F), our approximate Bayesian Gamma method (G), and the WinBUGS program (W) are given in Table 2. These statistics are for the nine pseudo-randomly generated samples. The WinBUGS results in Table 2 appear to match the classical method results more closely than the results of our approximate Bayesian method (with the exception of the standard deviation in the 10 item, 10 examinee sample). But, as just noted, Figures 1 through 9 show that our approximate Bayesian density and the classical sampling density have nearly identical modes, and all three methods quickly converge as m and n increase.

TABLE 2

Means and SD's for the Three Different Methods: F, G, and W

$p=m+1$	$N=n+1$	r	F mean	G mean	W mean	F sd	G sd	W sd
10	10	0.48	0.33	0.48	0.31	0.44	0.25	0.25
10	20	0.58	0.53	0.58	0.50	0.18	0.14	0.19
10	40	0.53	0.51	0.53	0.50	0.12	0.11	0.13
20	10	0.65	0.55	0.65	0.56	0.29	0.17	0.22
20	20	0.73	0.70	0.73	0.71	0.11	0.09	0.10
20	40	0.71	0.69	0.71	0.70	0.08	0.07	0.07
35	10	0.92	0.89	0.92	0.91	0.07	0.04	0.04
35	20	0.92	0.91	0.92	0.92	0.03	0.03	0.03
35	40	0.91	0.91	0.91	0.91	0.02	0.02	0.02

These results indicate that our approximate Bayesian method performed reasonably well when there were at least ten items and ten examinees. Its performance was excellent when there were at least twenty items and twenty examinees. Whenever a Bayesian approach is preferred we recommend the use of our approximate Bayesian method so long as m and n are not too small. Our approximate method is very easy to compute, but the WinBUGS computer program, which can be down loaded for free from the Internet, can also be used to obtain a marginal posterior distribution for alpha. Our Bayesian inferential procedures are based on the same models used for frequentist inference for alpha. Past studies previously mentioned have shown the frequentist methods to be robust to violations of their assumptions. It therefore is reasonable to expect our Bayesian methods to be similarly robust to violations of their assumptions.

References

Bernardo, J. M., & Smith, A., F., M. (1994). *Bayesian theory.* New York: Wiley.

Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Massachusetts: Addison-Wesley.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika, 30*, 357-370.

Feldt, L. S., Woodruff, D J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93-103.

Feldt, L. S. & Ankenmann, R. D. (1998). Appropriate sample size for comparing alpha reliabilities. *Applied Psychological Measurement, 22*, 170-178.

Feldt, L. S. & Ankenmann, R. D. (1999). Determining sample size for a test of the equality of alpha coefficients when the number of part-tests are small. *Psychological Methods, 4*, 366-377.

Hakstian, A. R. & Barchard, K. A. (2000). Toward more robust inferential procedures for coefficient alpha under sampling of both subjects and conditions. *Multivariate Behavioral Research, 35*, 427-456.

Hoyt, C. (1941). Test reliability estimated by analysis of variance. *Psychometrika, 6*, 153-160.

Press, S. J. (1982). *Applied multivariate analysis* (2nd Ed.). Malabar, Florida: Krieger.

Sahai, H. & Ageel, M. I. (2000). *The analysis of variance: fixed, random, and mixed models.* Boston: Birkhauser.

Van Zyl, J. M., Nuedecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika, 65*, 271-280.

Appendix A

Derivation of the Marginal Likelihood

When nuisance parameters are present in a Bayesian analysis they can be carried through to the posterior distribution and then integrated out. Bernardo and Smith (p. 245, 1994) present an alternative approach. They suggest integrating the nuisance parameters out of the likelihood and then using the resulting marginal or integrated likelihood for a Bayesian analysis. We now follow their method to derive the likelihood given in equation (6).

For notational convenience let $p = m + 1$ and $N = n + 1$. Let the column vector, $\mathbf{x}_i$, contain examinee i's responses to the p items. We assume that the $\mathbf{x}_i$, i=1, ..., N, constitute a random sample from a $\text{MVN}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ distribution where $\boldsymbol{\Sigma}$ has the compound symmetric form

$$\boldsymbol{\Sigma} = \phi^2(1-\rho_o)\mathbf{I} + \phi^2\rho_o\mathbf{jj}' = \sigma_e^2\mathbf{I} + \sigma_a^2\mathbf{jj}'. \tag{20}$$

In equation (20) $\boldsymbol{\Sigma}$ is the inter-item covariance matrix, $\mathbf{I}$ is a p by p identity matrix, and $\mathbf{jj}'$ is a p by p matrix of all ones. All the items have variance ϕ^2 and ρ_o is the common inter-item correlation. The two variance components, σ_e^2 and σ_a^2, relate to the ANOVA model of Feldt (1965). We include both parameterizations so the reader can relate the final results to the likelihood given in (6) and the model given in equations (17), (18), and (19). More specifically, $\sigma_e^2 = \phi^2(1-\rho_o)$ is the error variance component, and $\sigma_a^2 = \phi^2\rho_o$ is the examinee variance component and also the common inter-item covariance. Under this model ρ_o is the correlation between any two items and also the reliability of a single item.

The sample mean vector and covariance matrix are:

$$\bar{\mathbf{x}} = N^{-1}\sum_i^N \mathbf{x}_i \text{ and} \tag{21}$$

$$\mathbf{S} = N^{-1}\sum_i^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \tag{22}$$

The sum of squares and cross-products matrix is $\mathbf{V} = N\mathbf{S}$. The statistics $\bar{\mathbf{x}}$ and $\mathbf{V}$ are independent and constitute sufficient statistics for $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. The distribution of $\bar{\mathbf{x}}$, denoted $f(\bar{\mathbf{x}} \mid \boldsymbol{\theta}, \boldsymbol{\Sigma})$, is $\text{MVN}(\boldsymbol{\theta}, N^{-1}\boldsymbol{\Sigma})$ and $\mathbf{V}$ has a $\text{Wishart}(\boldsymbol{\Sigma}, n)$ distribution that we write as $f(\mathbf{V} \mid \boldsymbol{\Sigma}, n)$. The posterior distribution of a parameter depends on the data only through its sufficient statistic. So following

Press (1982, pp. 186-187) we take as the likelihood in this situation the joint distribution of the sufficient statistics $\bar{\mathbf{x}}$ and $\mathbf{V}$. Because they are statistical independent their joint distribution is just the product of their marginal distributions. Hence,

$$f(\bar{\mathbf{x}},\mathbf{V}\mid\boldsymbol{\theta},\boldsymbol{\Sigma}) = f(\mathbf{V}\mid\boldsymbol{\Sigma},n)f(\bar{\mathbf{x}}\mid\boldsymbol{\theta},\boldsymbol{\Sigma})$$

$$= \left[\frac{k(p,n)\mathbf{V}^{(n-p-1)/2}\exp(\frac{-\mathrm{tr}\boldsymbol{\Sigma}^{-1}\mathbf{V}}{2})}{|\boldsymbol{\Sigma}|^{n/2}}\right] * \left[\frac{\exp\left(-{}^{N}\!/_{2}(\boldsymbol{\theta}-\bar{\mathbf{x}})'\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\bar{\mathbf{x}})\right)}{(2\pi)^{p/2}\left|N^{-1}\boldsymbol{\Sigma}\right|^{1/2}}\right] \quad (23)$$

where $k(p,n)$ is the constant term for the Wishart distribution.

There is symmetry between $\bar{\mathbf{x}}$ and $\boldsymbol{\theta}$ in the above normal distribution. If we take the improper non-informative prior distribution $f(\boldsymbol{\theta}\mid\boldsymbol{\Sigma})=1$ for the nuisance parameter $\boldsymbol{\theta}$ and then integrate equation (23) with respect to it, all that will remain is the Wishart marginal likelihood for $\mathbf{V}$ and that is free of $\boldsymbol{\theta}$. The proper conjugate prior for $\boldsymbol{\theta}$ in the above likelihood is $\mathrm{MVN}(\boldsymbol{\mu},\nu^{-1}\boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is our best apriori estimate of $\boldsymbol{\theta}$ and ν measures the strength of our belief in that apriori estimate. Integrating equation (23) with respect to such a MVN prior for $\boldsymbol{\theta}$ will not give the desired result. The improper non-informative prior must be used. Our method forgoes specifying prior information for $\boldsymbol{\theta}$ so as to make simple the specifying of prior information for alpha.

We next follow VanZyl, Neudecker, and Nel (2000) and consider two scalar transformations of $\mathbf{V}$ along with their sampling distributions:

$$MS_a = \frac{\mathbf{j}'\mathbf{V}\mathbf{j}}{np} \sim \frac{\chi^2_n}{n}[1+m\rho_o]\phi^2 \text{ and} \quad (24)$$

$$MS_e = \frac{\mathrm{tr}(\mathbf{V})-nMS_a}{nm} \sim \frac{\chi^2_{nm}}{nm}[1-\rho_o]\phi^2. \quad (25)$$

These two random variables are statistically independent and constitute sufficient statistics for ϕ^2 and ρ_o. Again, because the posterior distribution of a parameter depends on the data only through its sufficient statistic, our marginal likelihood for $\mathbf{V}$ reduces to the product of the two independently distributed χ^2 variables MS_a and MS_e.

Under the covariance structure given in equation (20) coefficient alpha equals reliability rather than just being a lower bound for reliability, and ρ can be expressed as a function of ρ_o using the Spearman-Brown formula:

$$\rho = \frac{p\rho_o}{1+(p-1)\rho_o}. \tag{26}$$

Using the inverse of equation (26) one finds that the ratio of the expected values of MS_a and MS_e is

$$\frac{\mathrm{E}(MS_a)}{\mathrm{E}(MS_e)} = \frac{1+m\rho_o}{1-\rho_o} = \frac{1}{1-\rho} = \tau. \tag{27}$$

Using equation (18) shows that

$$\frac{MS_a}{MS_e} = \frac{1}{1-r} = t. \tag{28}$$

Dividing equation (28) by equation (27) gives the final result

$$\frac{MS_a / MS_e}{\mathrm{E}(MS_a)/\mathrm{E}(MS_e)} = \frac{t}{\tau} \sim \mathrm{F}_{n,nm}. \tag{29}$$

Hence, t is a sufficient statistic for τ (and r sufficient for ρ), and when interest focuses solely on ρ this $\mathrm{F}_{n,nm}$ distribution can be used as the likelihood in a Bayesian analysis of alpha.

Appendix B

WinBUGS Code For Inference about Coefficient Alpha Based on a Two-Way Random Effects ANOVA Model

In the WinBUGS code below we use the following conventions. WinBUGS code words are in all capital letters. The names of user supplied constants or initial values are in all lower case italicized letters. **xx** is a place-holder for where the values of the constants and initial values need to be specified. The names of variables and parameters are in regular all lower case letters. Note that for simplicity we have given a value of zero to the grand mean, μ, instead of treating it as a parameter and giving it a prior distribution.

```
MODEL
{
   FOR(i IN 1:nostuds)
     { a[i]~DNORM(0, tau.a)   }
   FOR(j IN 1:noitems)
     { b[j]~DNORM(0, tau.b)   }
   FOR(i IN 1:nostuds)
     { FOR(j IN 1:noitems)
        { m[i,j] ← 0+a[i]+b[j]
          y[i,j]~DNORM(m[i,j],tau.c) } }
   sigma2.a ← 1/tau.a
   sigma2.b ← 1/tau.b
   sigma2.c ← 1/tau.c
   alpha ← ( sigma2.a )/( sigma2.a + (sigm2.c/noitems) )
   tau.a~DGAMMA(xx,xx)
   tau.b~DGAMMA(xx,xx)
   tau.c~DGAMMA(xx,xx)
}

LIST(nostuds=xx, noitems=xx,
y=STRUCTURE(.DATA=
C(… data values separated by commas …),
.DIM=c(xx,xx)
)
)
LIST(tau.a=xx, tau.b=xx, tau.c=xx)
```

The last `LIST` statement contains initial values for the three variance components. DNORM denotes a normal distribution and DGAMMA denotes a gamma distribution.

Appendix C

Figures

Figure 1. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=9 and n=9.

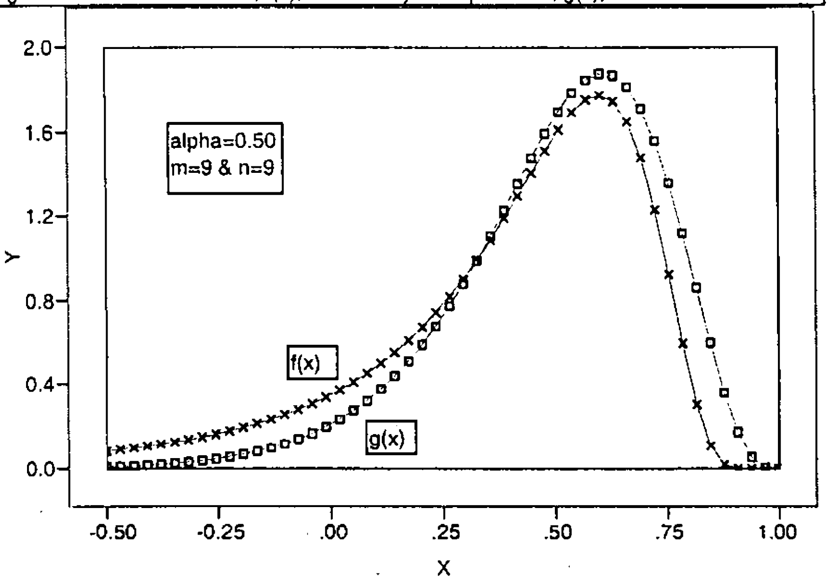

Figure 2. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=9 and n=19.

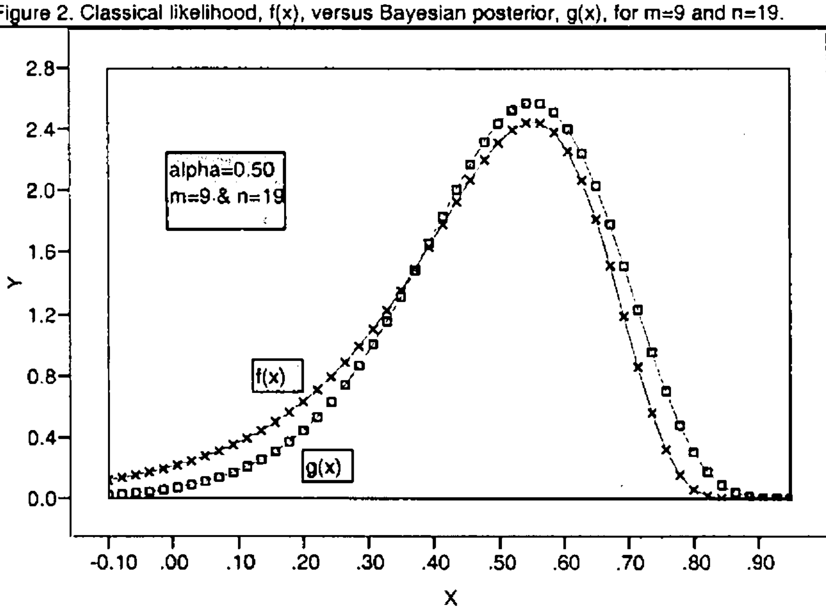

Figure 3. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=9 and n=39.

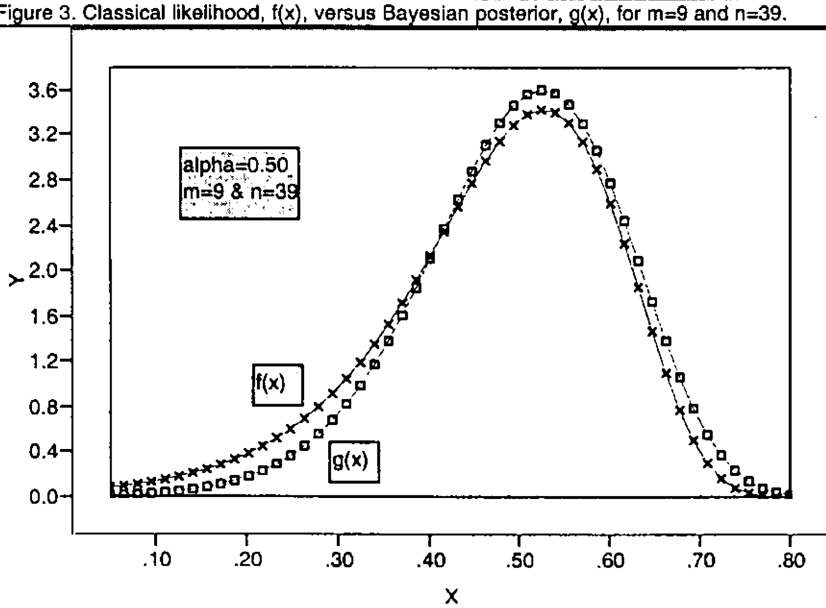

Figure 4. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=19 and n=9.

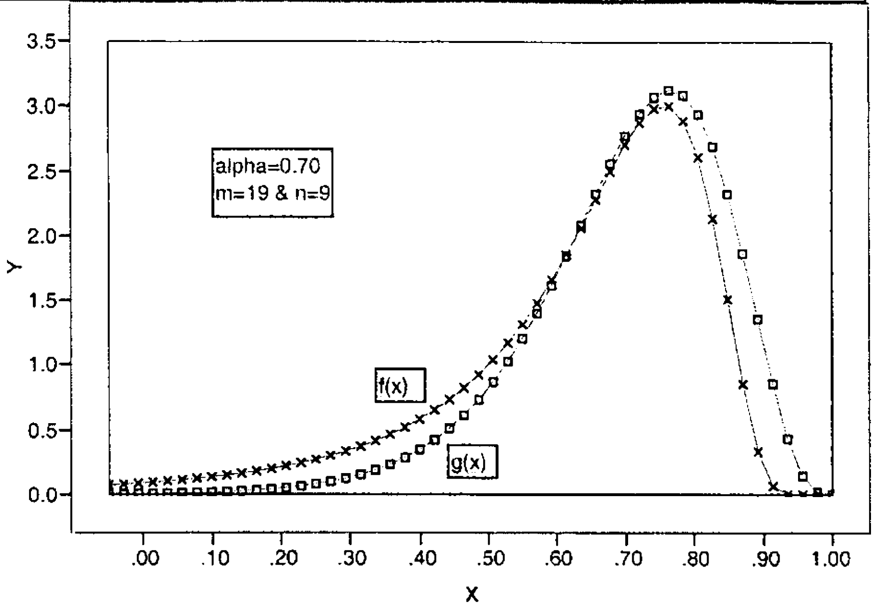

Figure 5. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=19 and n=19.

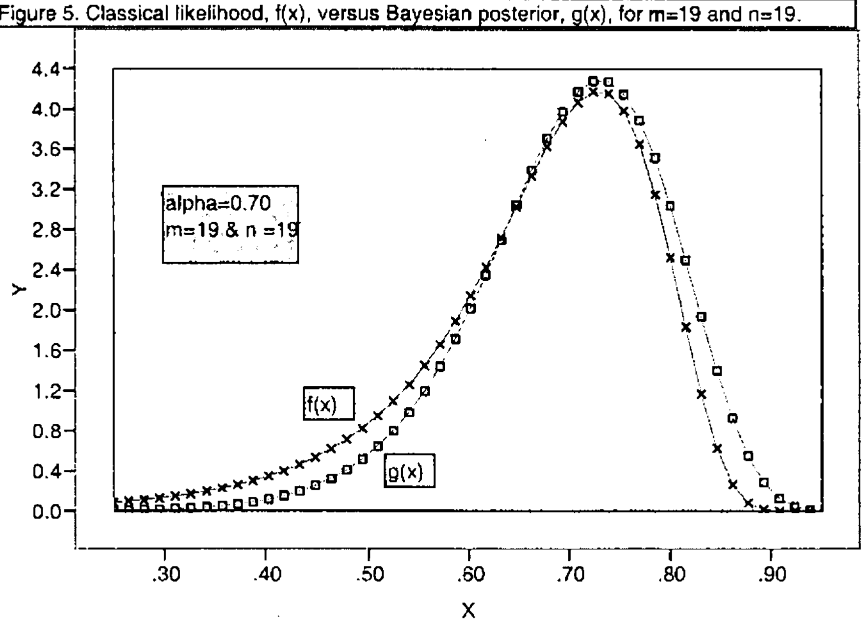

Figure 6. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=19 and n=39.

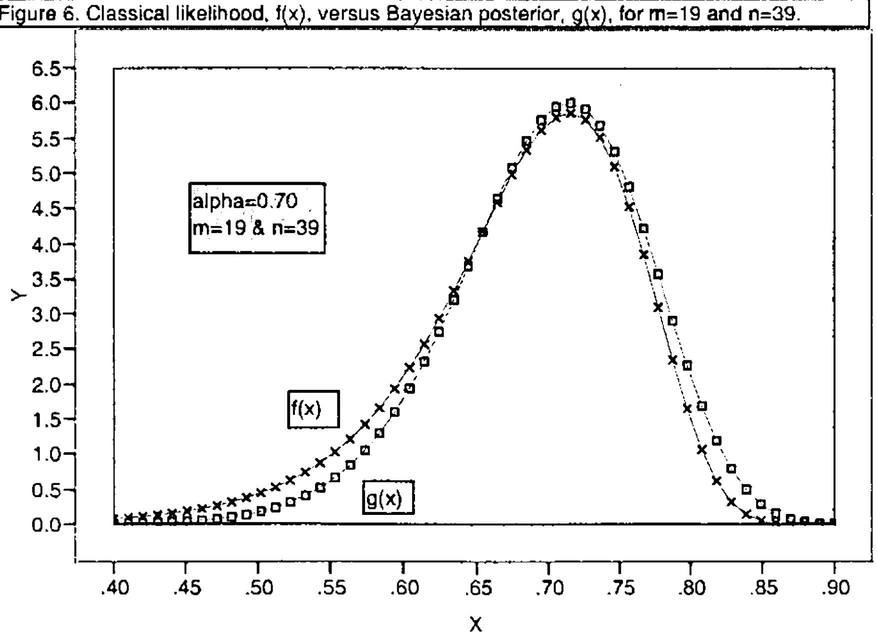

Figure 7. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=34 and n=9.

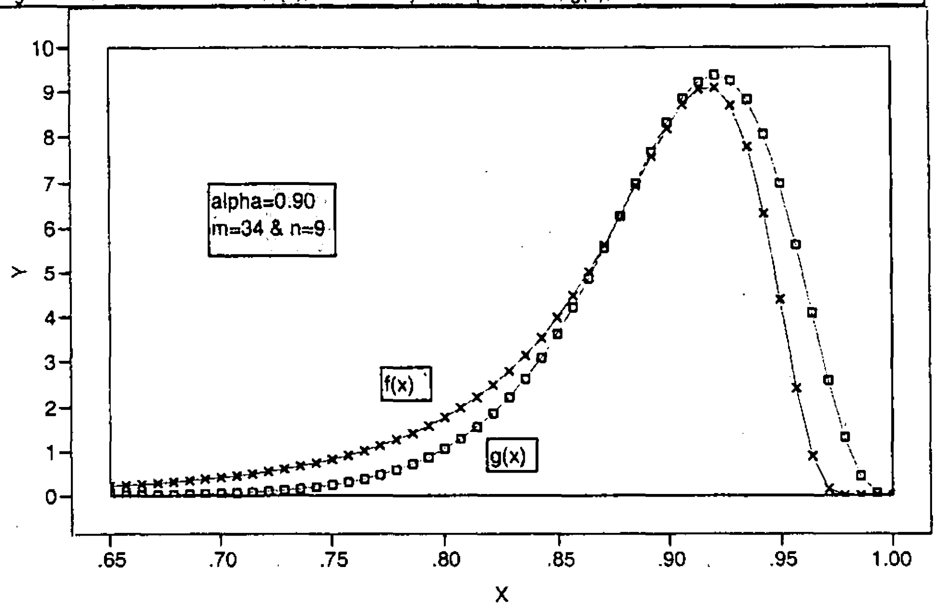

Figure 8. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=34 and n=19.

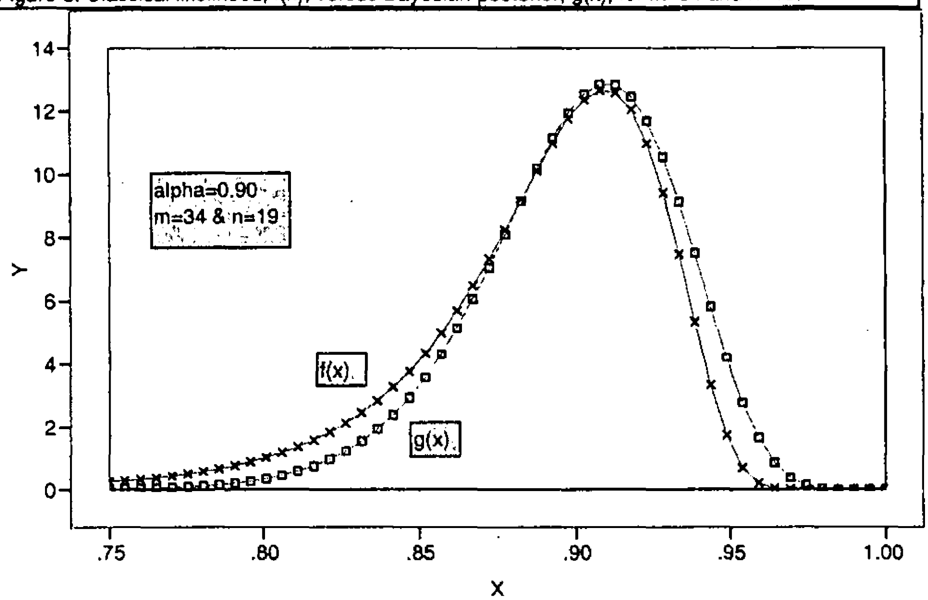

Figure 9. Classical likelihood, f(x), versus Bayesian posterior, g(x), for m=34 and n=39.