DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 10 | Issue 1 Article 25

5-1-2011

Estimating Internal Consistency Using Bayesian Methods

Miguel A. Padilla
Old Dominion University, mapadill@odu.edu

Guili Zhang
East Carolina University, zhangg@ecu.edu

Journal of Modern Applied Statistical Methods
May 2011, Vol. 10, No. 1, 277-286

1538 – 9472/11/$95.00

Estimating Internal Consistency Using Bayesian Methods

Miguel A. Padilla
Old Dominion University

Guili Zhang
East Carolina University

Bayesian internal consistency and its Bayesian credible interval (BCI) are developed and Bayesian internal consistency and its percentile and normal theory based BCIs were investigated in a simulation study. Results indicate that the Bayesian internal consistency is relatively unbiased under all investigated conditions and the percentile based BCIs yielded better coverage performance.

Key words: Bayesian internal consistency, coefficient alpha, confidence interval, Bayesian confidence interval, coverage probability.

Introduction

Psychological constructs are the building blocks of psychological/behavioral research. Indeed, one can easily argue that constructs are the foundation of these two sciences. A typical way of measuring a construct is through a questionnaire containing items that are purported to indirectly measure the construct of interest; thus, it becomes important that the items be consistent or reliable so that the questionnaire itself is consistent or reliable. Although there are several methods of measuring or estimating the reliability of a questionnaire, by far the most commonly used is coefficient alpha.

Coefficient alpha has remained popular since its introduction in Cronbach's (1951) article based on the work of Guttman and others in the 1940s (Guttman, 1945). Coefficient alpha is a measure of internal consistency for a group of items (i.e., questions) that are related in that they measure the same psychological/behavioral construct (Cortina, 1993). There are three main reasons for coefficient alpha's popularity. First, coefficient alpha is computationally simple. The only required quantities for its computation are the number of items, variance for each item and the total joint variance for all the items; quantities that can easily be extracted from the item covariance matrix. Second, coefficient alpha can be computed for continuous or binary data: this is a significant advantage when working with right/wrong, true/false, etc. items. Third, it only requires one test administration: Most other forms of reliability require at least two test administrations, which come at a cost of time and resources. For these reasons coefficient alpha's power to assess the psychometric property of the reliability of a measurement instrument is widely used and it has remained relatively unchanged for over 60 years.

The advent of Bayesian methodology has brought about exciting and innovating ways of thinking about statistics and analyzing data. Bayesian methods have several advantages over traditional statistics, sometimes referred as frequentist statistics (Gelman, 2004; Lee, 2004), but two advantages stand out. First, researchers can now incorporate prior knowledge or beliefs about a parameter by specifying a prior distribution for the parameter in the model; thus, the analysis is now composed of data and prior knowledge and/or beliefs. By contrast, traditional or frequentist analyses are composed only of data. Through this combination of data and prior knowledge, more can be learned about the phenomenon under study and knowledge about the phenomenon can be updated accordingly. Second, Bayesian methods provide

Miguel A. Padilla is an Assistant Professor of Quantitative Psychology in the Department of Psychology. His research interests are in applied statistics and psychometrics. Email him at: mapadill@odu.edu. Guili Zhang is an Assistant Professor of Research Methodology in the Department of Curriculum and Instruction. Her research interests are in applied statistics. Email her at: ZHANGG@ecu.edu.

credible intervals (BCIs), the Bayesian analog to confidence intervals (CIs). However, credible intervals have a different interpretation from confidence intervals. A confidence interval allows one to make statements, such as "we are 95% confident that the interval captures" the true population parameter. By contrast, a BCI allows one to say that "we are 95% confident that the true population parameter lies between the bands of the credible interval," a simpler and more powerful statement. This is, in fact, the interpretation most researchers would like to make with confidence intervals.

A Bayesian coefficient alpha retains the simplicity and power of the original coefficient alpha, but it also has the advantages of Bayesian methodology. By incorporating prior internal consistency information into the current estimation of coefficient alpha, more can be learned about the internal consistency of a measurement instrument and knowledge about the instrument can be updated accordingly. Additionally, credible intervals are generated for the Bayesian coefficient alpha. The bootstrap is the common method for generating confidence intervals for coefficient alpha; however, the bootstrap confidence interval has the same interpretation as the confidence interval from traditional statistical methods. With credible intervals direct statements can be made about where the true population coefficient alpha lies, which is a clear advantage over the standard confidence interval.

Prior Research on Coefficient Alpha CIs

As with all statistics, coefficient alpha is a population parameter and must be estimated from samples; thus, it is subject to sampling error that contributes to the variability around the true population parameter. Due to this, current statistical thinking and practice point to the need for providing confidence intervals to supplement point estimates and statistical tests (Duhachek, Coughlan & Iacobucci, 2005; Duhachek & lacobucci, 2004; Iacobucci & Duhachek, 2003; Maydeu-Olivares, Coffman & Hartmann, 2007). Many professional publications are beginning to require authors to provide CIs in addition to point estimates, standard errors and test statistics. For example, the American Psychological Association Task Force on Statistical Inference (Wilkinson, 1999) emphasizes the obligation of researchers to provide CIs for all principal outcomes; however, generating CIs for coefficient alpha has remained somewhat elusive and rarely implemented in practice.

Confidence intervals for coefficient alpha were first introduced by Kristof (1963) and Feldt (1965). These CIs assume that items are normally distributed and strictly parallel (Allen & Yen, 1979; Crocker & Algina, 1986), which implies that the item covariance matrix is compound symmetric; i.e., $\sigma_1 + \sigma^2 I\left(i = j\right)$ where σ_1 are the item variances, and σ^2 are the item covariances, and $I(.)$ is the indicator function. These confidence intervals, however, do not perform well when items are not strictly parallel (Barchard & Hakstian, 1997). Given that the strictly parallel assumption is unreasonable in applied research and that these CIs do not perform well when this assumption is violated may be the reason why coefficient alpha CIs are rarely implemented in applied research (Duhachek & lacobucci, 2004).

An improvement to the CIs proposed by Kristof (1963) and Feldt (1965) was introduced by van Zyl, Neudecker and Nel (2000) who showed that the standard method of estimating coefficient alpha is a maximum likelihood estimator (MLE) and derived its corresponding CIs. Although the coefficient alpha MLE assumes that items are normally distributed, a major advantage is that it does not require the compound symmetry assumption of the item covariance matrix. In a simulation study, Duhacket and Iacobucci (2004) compared the performance of the coefficient alpha CIs for the method proposed by Feldt (1965) and the MLE proposed by van Zyl, et al. (2000) under a non-parallel measurement model. Their results indicate that the MLE method consistently outperformed the competing methods across all simulation conditions, but because the MLE method assumes that items are normally distributed, when the assumption is violated, the results can be untrustworthy.

Normally distributed items are not a completely realistic assumption in psychological/behavioral research. Most items in measurement instruments are dichotomous (yes/no, true/false, etc.) or Likert-type items with several ordinal items: for these item types, normality is an unrealistic assumption. From this perspective, Yuan and Bentler (2002) extended the results of the coefficient alpha MLE to a wider range of distributions, pointing out that it is robust to some violations of normality. However, they point out that it is difficult to verify conditions to which the coefficient alpha MLE is robust to item non-normality. Thus, if the conditions cannot be verified theoretically then they are even more difficult to verify in applied work.

Yuan and Bentler (2003) built on this by introducing what Maydeu-Olivares and colleagues (2007) call asymptotically distribution-free (ADF) CIs for coefficient alpha. In this study the authors compared the ADF, MLE, and bootstrap coefficient alpha CIs estimated from the Hopkins Symptom Checklist (HSCL; Derogatis, Lipman, Rickels, Uhlenhuth & Covi, 1974). The results of Yuan and Bentler suggest that the ADF CIs are between the MLE and bootstrap methods in terms of their accuracy. However, they point out that the ADF CIs cannot describe the tail behavior of coefficient alpha of the HSCL due to the small sample (n = 419); they suggest that with a larger sample size the ADF CIs could better describe the distribution of coefficient alpha.

Maydeu-Olivares, et al. (2007) extended the work by Yuan, et al. (2003) by simplifying the computation of ADF CIs and investigating its performance under several simulation conditions. Of particular interest was the comparison of the ADF CIs to the MLE CIs under various conditions of skewness and kurtosis. In general, they concluded that - with approximately normal items - the MLE CIs perform well even with a sample size as small as 50. However, once the items begin to deviate from normality, the ADF CIs begin to outperform the MLE CIs. In particular, the ADF CIs outperform MLE CIs with as little a sample size of 100. When the sample size gets larger than 100 the ADF CIs perform well regardless of the skewness and kurtosis investigated by the researchers.

Recent research has thus been fruitful in investigating the properties of coefficient alpha CIs; however, these CIs are based on traditional frequentist statistics. As such, they have the traditional interpretation of CIs and cannot be updated with prior information. The primary purpose of this study is to develop a Bayesian internal consistency estimate and to evaluate its performances by investigating some of its properties through simulation.

Coefficient Alpha

Consider a measurement instrument containing p items, y_1, y_2, …, y_p, that indirectly measure a single dimension, attribute, or construct. A useful and common computation in the psychological/behavioral sciences is the composite $Y = y_1 + y_2 + \ldots + y_p$. This composite is placed in statistical models such as ANOVA and regression when conducting research using the corresponding attribute as a variable. Therefore, it is important to know the reliability of the composite and hence the construct being measured.

A popular way to estimate the composite reliability is through coefficient alpha. Coefficient alpha for the population is defined as

$$\alpha_c = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p}\sigma_{ii}}{\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}}\right). \tag{1}$$

where $\sum_{i=1}^{p}\sigma_{ii}$ is the sum of all item variances and $\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}$ is the sum of all item variances and covariances. For a sample of size n, population parameters are replaced by sample estimates to obtain a coefficient alpha estimate as

$$\hat{\alpha}_c = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p}\hat{\sigma}_{ii}}{\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{\sigma}_{ij}}\right). \tag{2}$$

Note that coefficient alpha is being subscripted with c to distinguish it from the other forms that will shortly be introduced. Recall that Zyl, et al. (2000) showed that $\hat{\alpha}_c$ is the MLE for α_c.

Coefficient alpha has three interesting properties implied from the classical true score model (Allen & Yen, 1979; Crocker & Algina, 1986). First, when all items have equal true scores that relate equally to the observed scores along with equal measurement error variance, the items are said to be parallel. In this case the covariance matrix for the items has a compound symmetric structure; i.e., $\sigma_1 + \sigma^2 I(i = j)$. Second, when the measurement error variances are not equal, the items are said to be tau-equivalent. In both of these conditions coefficient alpha is equal to the reliability of a measurement instrument. Lastly, when the true scores do not relate equally to the observed scores and measurement error variances are not equal, the items are congeneric. This last condition is the more general and in this case coefficient alpha underestimates the reliability of a measurement instrument. It is from these three conditions that the conclusion $\hat{\alpha}_c \le \rho_{xx'}$ is made, where $\rho_{xx'}$ is the reliability coefficient of a measurement instrument.

Bayesian Internal Consistency

The cornerstone of Bayesian methodology is Bayes' theorem. Through Bayes' theorem all unknown parameters are considered random variables. Due to this, prior distributions must be initially defined, which is a way for researchers to express prior beliefs or available information before data are involved in the statistical analysis. Using the observed data $\boldsymbol{y}$ and prior distribution $p(\boldsymbol{\theta})$, a posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of the parameters $\boldsymbol{\theta}$ can be constructed. The posterior distribution can be fully expressed through Bayes' theorem as

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) = \frac{L(\boldsymbol{\theta} \mid \boldsymbol{y})\, p(\boldsymbol{\theta})}{\int_{\Theta} L(\boldsymbol{\theta} \mid \boldsymbol{y})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \propto L(\boldsymbol{\theta} \mid \boldsymbol{y})\, p(\boldsymbol{\theta}) \tag{3}$$

where $L(\boldsymbol{\theta} \mid \boldsymbol{y})$ is the data likelihood function and $\int_{\Theta} L(\boldsymbol{y} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ is a normalizing constant. One can directly see that the posterior is composed of both actual data and prior beliefs or knowledge about the parameter. After the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is constructed it can be summarized by the mean and SD (or SE) along with other summarizing quantities. For example, the mean and variance can be computed as

$$E(\boldsymbol{\theta} \mid \boldsymbol{y}) = \int_{\Theta} \boldsymbol{\theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta} \tag{4}$$

and

$$\operatorname{var}(\theta \mid \boldsymbol{y}) = \int_{\Theta} \left[\boldsymbol{\theta} - E(\boldsymbol{\theta} \mid \boldsymbol{y})\right]^2 \pi(\boldsymbol{\theta} \mid \boldsymbol{y})\, d\boldsymbol{\theta}, \tag{5}$$

where the $SD = \sqrt{\operatorname{var}(\theta \mid \boldsymbol{y})}$ is also the SE for $E(\boldsymbol{\theta} \mid \boldsymbol{y})$. At times, the median (or 50th percentile) computed as

$$P(\boldsymbol{\theta} \mid \boldsymbol{y} \le m) = P(\boldsymbol{\theta} \mid \boldsymbol{y} \ge m) \ge \frac{1}{2} \tag{6}$$

is of interest as it is less influenced by extreme values.

For the Bayesian coefficient alpha (Balpha), first start with the multivariate normal distribution. The posterior of a multivariate normal can be described by

$$\begin{gathered}\pi(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{y}) \propto L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{y})\, p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \\ L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \boldsymbol{y})\, p(\boldsymbol{\mu} \mid \boldsymbol{\Sigma})\, p(\boldsymbol{\Sigma})\end{gathered} \; . \tag{7}$$

On the far right of (6), note that the prior for the mean is directly dependent on the prior covariance, in addition, this indicates that a different prior is specified for the covariance matrix and mean vector. By choosing the following conjugate priors for both the covariance matrix

$$\boldsymbol{\Sigma} \sim W^{-1}(d_0,\ \boldsymbol{\Lambda}) \tag{8}$$

and mean vector

$$\boldsymbol{\mu} \mid \boldsymbol{\Sigma} \sim N\left(\boldsymbol{\mu}_0, \frac{1}{n_0}\boldsymbol{\Sigma}\right), \tag{9}$$

Anderson (1984) and Schafer (1997) showed that the posterior distribution for the covariance matrix and mean vector is

$$\boldsymbol{\Sigma} \mid \boldsymbol{y} \sim W^{-1}\begin{pmatrix} n+d_0, \\ (n-1)\mathbf{S}+\boldsymbol{\Lambda}+\frac{nn_0}{n+n_0}(\bar{\boldsymbol{y}}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{y}}-\boldsymbol{\mu}_0)' \end{pmatrix} \tag{10}$$

$$\boldsymbol{\mu} \mid (\boldsymbol{\Sigma}, \boldsymbol{y}) \sim N\left(\frac{1}{n+n_0}(n\bar{\boldsymbol{y}}+n_0\boldsymbol{\mu}_0), \frac{1}{n+n_0}\boldsymbol{\Sigma}\right) \tag{11}$$

where $W^{-1}()$ denotes an inverted Wishart distribution and d_0, $\boldsymbol{\Lambda}$, $\boldsymbol{\mu}_0$, and n_0 are hyperparameters chosen by the analyst, and $\bar{\boldsymbol{y}}$ and $\mathbf{S}$ are the mean vector and covariance matrix estimated from the data. Thus, the posterior of the multivariate normal is described by two distributions which jointly are called the normal-inverse Wishart distribution. Note that a prior needs be specified for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. If no prior is available a generic noninformative prior such as $p(\boldsymbol{\theta}) \propto 1$ can be used. In this case the posterior is completely defined by the data. This parameterization fully describes the posterior and it can now be directly computed.

The coefficient alpha posterior can be difficult to obtain directly. However, by simulating $t = 1, 2, ..., T$ values from (9) and (10) as $\boldsymbol{\Sigma}^{(t)} \mid \boldsymbol{y}$ and $\boldsymbol{\mu}^{(t)} \mid (\boldsymbol{\Sigma}^{(t)}, \boldsymbol{y})$, the estimation of the coefficient alpha posterior distribution can be obtained as

$$\alpha_c^{(t)} = \frac{p}{p-1}\left(1-\frac{\sum_{i=1}^{p}\sigma_{ii}^{(t)}}{\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}^{(t)}}\right) \tag{12}$$

where $\sigma_{ii}^{(t)}$ and $\sigma_{ij}^{(t)}$ are elements of $\boldsymbol{\Sigma}^{(t)} \mid \boldsymbol{y}$. A Bayesian coefficient alpha (Balpha) can then obtained as

$$\alpha_b = E(\alpha_c \mid \boldsymbol{y}). \tag{13}$$

An alternative Bayesian coefficient alpha (BalphaM) can be obtained through

$$P(\alpha_c \mid \boldsymbol{y} \le \alpha_{b,m}) = P(\alpha_c \mid \boldsymbol{y} \ge \alpha_{b,m}) \ge 1/2.$$

Bayesian credible intervals can then obtained by the lower $\alpha/2$ and upper $1-\alpha/2$ percentiles of the sample, where α is the type I error rate. One can also obtain a normal theory based credible interval as $\alpha_b \pm Z_{\alpha/2}SD$. Other summary measures can also be computed as indicated above.

Methodology

Simulation

A 4 × 3 × 6 Monte Carlo simulation design was utilized to investigate the properties of Bayesian coefficient alphas. First, the number of items was investigated: 5, 10, 15 and 20 and it was found that coefficient alpha increases as a function of the number of items, however, it is constrained to one. Although it is possible for tests and/or surveys to have more than 20 items, going beyond 20 items reaches a point of diminishing returns in terms of investigating coefficient alpha.

Second, the mean item correlation $(\bar{r})$ was investigated: 0.173, 0.223, and 0.314. The mean item correlation is defined as

$$\bar{r} = \frac{2\sum_{i<j}^{p} r_{ij}}{p(p-1)} \tag{14}$$

These mean items correlations were investigated because they generate coefficient alphas that range from 0.50 to 0.90, a sufficient range to investigate the properties of the Balpha.

Third, sample size was also explored: 50, 100, 150, 200, 250 and 300. As is the case for the number of items, going beyond a sample size of 200 reaches a point of diminishing returns in terms of investigating coefficient alpha (Duhachek & lacobucci, 2004); however, these are sample sizes typically found in psychological/behavioral research. Table 1 presents coefficient alpha as a function of mean item correlation and number of items and shows a reasonable range of coefficient alpha that may be found in psychological/behavioral research.

Table 1: Population Coefficient Alpha for Items by Mean Item Correlations

Items	Mean Item Correlations		
	.1667	.2208	.3103
5	.5001	.5862	.6923
10	.6667	.7392	.8182
15	.7500	.8095	.8709
20	.8000	.8500	.9000

Multivariate normal data were generated with mean vector zero and correlation matrix **R** of dimensions defined by the number of items in the simulation. **R** was chosen to have homogenous off-diagonal elements that generated the corresponding mean item correlation.

For each condition of the simulation study 1,000 replications were obtained. In each replication, Balpha was computed along with the SE and 95% BCIs. Relative bias for Balpha was computed as:

$$\hat{\alpha}_b = \frac{\bar{\hat{\alpha}}_b - \alpha}{\alpha}. \tag{15}$$

The average of the estimated SE was computed as

$$\overline{SE} = \frac{\sum_{i=1}^{B} SE\left(\hat{\alpha}_{b,i}\right)}{B} \tag{16}$$

where B is the number of replications. Lastly, two forms of BCI intervals were computed. The first BCI was obtained by the lower $\alpha/2$ and upper $1-\alpha/2$ percentiles of the sample. The second BCI was obtained as $\hat{\alpha}_b \pm Z_{\alpha/2} SE\left(\hat{\alpha}_b\right)$. The coverage probability of the 95% BCIs were computed as the proportion of times the BCI contains the population parameter α_c.

Coverage can be judged by forming confidence intervals around the coverage. Coverage should not fall approximately two standard errors (SEs) outside the nominal coverage probabilities (p) (Burton, Altman, Royston & Holder, 2006). The standard error is defined as

$$SE(p) = \sqrt{\frac{p(1-p)}{B}} \tag{17}$$

where B is the number of simulations in the study. For the current study, $p = .95$ with $SE(p) = .006892$ and the CI is $[.936,\ .964]$. Thus, coverage that falls outside this CI is considered unacceptable.

For this study, Balpha and corresponding 95% BCIs were estimated from a total of 1,000 simulations from the posterior distribution. In addition, the prior for Balpha was set to be noninformative. A noninformative prior essentially lets the data essentially speak for themselves.

Results

Relative bias for Balpha and corresponding standard errors are reported in Table 2. First, Balpha and BalphaM always tend to slightly underestimate the population coefficient alpha; however, both Balpha and BalphaM are relatively unbiased under all investigated conditions. Second, Balpha and BalphaM estimates get closer to the population coefficient alpha as sample size increases. Third, Balpha and BalphaM estimates get closer to the population coefficient alpha as the number of items increases. In addition, Balpha and BalphaM estimates get better as the mean item correlation increases. Lastly, BalphaM is consistently closer to the population coefficient than Balpha although the difference is nominal.

In terms of the standard error (SE), a few things should be pointed out. First, the SEs are smaller as the mean item correlation increases. Second, standard errors improve as sample size increases as should be expected. For samples sizes from 100 to 300, the SE difference is nominal when the number of items is between 5 and 10. When the number of items is between 15 and 20, the SE difference is nominal regardless of the sample size. Third, the SEs improve as the mean item correlation increase although the difference can be considered nominal; in most of these conditions, increasing the number of posterior samples should improve the estimation of the SEs.

Table 2: Balpha and BlaphaM Relative Bias with Standard Errors*

Number of Items	Sample Size	Mean Item Correlation		
		.1667	.2208	.3103
5	50	-.0639, -.0458 (.1189)	-.0592, -.0434 (.1034)	-.0437, -.0328 (.0792)
	100	-.0384, -.0278 (.0835)	-.0304, -.0226 (.0699)	-.0173, -.0124 (.0518)
	150	-.0278, -.0205 (.0673)	-.0191, -.0141 (.0557)	-.0098, -.0066 (.0412)
	200	-.0198, -.0145 (.0577)	-.0128, -.0092 (.0477)	-.0109, -.0085 (.0357)
	250	-.0121, -.0079 (.0510)	-.0104, -.0074 (.0424)	-.0055, -.0037 (.0314)
	300	-.0167, -.0133 (.0467)	-.0103, -.0078 (.0387)	-.0070, -.0012 (.0285)
10	50	-.0537, -.0421 (.0860)	-.0325, -.0243 (.0666)	-.0228, -.0176 (.0468)
	100	-.0250, -.0203 (.0546)	-.0137, -.0103 (.0423)	-.0084, -.0062 (.0295)
	150	-.0100, -.0071 (.0426)	-.0090, -.0068 (.0335)	-.0062, -.0049 (.0235)
	200	-.0145, -.0122 (.0370)	-.0074, -.0059 (.0287)	-.0040, -.0030 (.0199)
	250	-.0073, -.0055 (.0324)	-.0075, -.0063 (.0256)	-.0034, -.0026 (.0177)
	300	-.0075, -.0061 (.0295)	-.0054, -.0044 (.0231)	-.0027, -.0020 (.0161)
15	50	-.0328, -.0245 (.0672)	-.0252, -.0192 (.0516)	-.0152, -.0115 (.0348)
	100	-.0136, -.0105 (.0410)	-.0104, -.0082 (.0314)	-.0052, -.0038 (.0211)
	150	-.0083, -.0064 (.0321)	-.0068, -.0055 (.0246)	-.0042, -.0034 (.0167)
	200	-.0082, -.0068 (.0275)	-.0056, -.0046 (.0209)	-.0036, -.0030 (.0142)
	250	-.0089, -.0078 (.0245)	-.0033, -.0025 (.0184)	-.0023, -.0018 (.0125)
	300	-.0048, -.0039 (.0219)	-.0037, -.0031 (.0167)	-.0017, -.0013 (.0113)
20	50	-.0276, -.0205 (.0580)	-.0213, -.0162 (.0440)	-.0129, -.0097 (.0292)
	100	-.0100, -.0076 (.0335)	-.0090, -.0073 (.0254)	-.0049, -.0037 (.0168)
	150	-.0065, -.0050 (.0260)	-.0041, -.0030 (.0194)	-.0033, -.0026 (.0130)
	200	-.0047, -.0036 (.0219)	-.0039, -.0032 (.0165)	-.0027, -.0022 (.0110)
	250	-.0050, -.0042 (.0194)	-.0028, -.0022 (.0145)	-.0013, -.0009 (.0096)
	300	-.0040, -.0033 (.0175)	-.0024, -.0019 (.0132)	-.0014, -.0011 (.0088)

*Note: The first number is Balpha followed by BalphaM. Numbers in parentheses are standard errors.

The Bayesian credible intervals are displayed in Table 3 and are more interesting. In general, most of the credible intervals fall within the acceptable range of [.936, .964] based on 1,000 replications. In addition, the percentile based BCIs are consistently closer to 0.95 than the normal theory based BCIs. With 5 items, only two BCIs were not within the acceptable range. When the number of items shifts to 10, seven BCIs were not within the acceptable range, but most of the unacceptable BCIs are normal theory based.

As the number of items increases, more BCIs begin to fall outside the acceptable range, but once again, most of the unacceptable BCIs are normal theory based. However, the unacceptable BCIs occur when the numbers of items are between15 to 20 and are paired with the smaller sample sizes. Specifically, when the numbers of items are 15, unacceptable BCIs occur at a sample size of 50. Also, when the numbers of items are 20, unacceptable BCIs occur at sample sizes of 50 to 100. In both cases, more normal theory BCIs become unacceptable as the item mean correlation increases. However, the percentile based BCIs tend to remain more stable and closer to 0.95.

Conclusion

The building blocks of psychological/behavioral research are psychological constructs, which are typically indirectly measured through items on questionnaires. It is crucial to have items that are consistent or reliable in order for research results to be trustworthy and useful. A popular method for estimating a form of reliability is internal consistency via coefficient alpha (Cronbach, 1951; Guttman, 1945). However, coefficient alpha has remained unchanged for over 60 years. Many professional publications are encouraging and/or mandating researchers to supplement their parameter estimates with CIs. Although CIs for coefficient alpha have recently enjoyed fruitful research (Barchard & Hakstian, 1997; Duhachek & lacobucci, 2004; Feldt, 1965; Kristof, 1963; Maydeu-Olivares, et al., 2007; van Zyl, et al., 2000; Yuan & Bentler, 2002; Yuan, et al., 2003), they are rarely implemented in applied research. In addition, all current coefficient alpha CIs are frequentist based and, as such, they have the traditional, less desirable CI interpretation and cannot use prior information to stabilize inferences or update information.

This study developed a Bayesian coefficient alpha (Balpha or BalphaM) and its corresponding BCIs. The results from the Monte Carlo investigations indicate that Balpha and BalphaM are relatively unbiased under all investigated conditions of the simulation. However, Balpha and BalphaM have the added advantage of having the BCIs, which have the interpretation researchers really want to make with CIs. Again, BCIs allow one to make the following simpler and more powerful statement: results show 95% confidence that the true population parameter lies between the bands of the credible interval.

In terms of coverage, the percentile based BCIs performed better than the normal theory based BCIs. In particular, the normal theory BCIs begin to perform poorly when the mean item correlation is $\overline{r} = .3101$, and the condition worsens as the number of items increases. However, increasing the sample size offsets these conditions. In fact, having a sample size of 250 or more appears to provide protection against this breakdown of the normal theory BCIs. Conversely, the percentile based BCIs remain more consistent, but begin to become unacceptable with the smaller sample sizes and when the number of items is between 15 and 20. However, they remain acceptable as long as the sample size is at least 100. Thus, percentile based BCIs are recommended over the normal theory BCIs.

In general, this suggests that as the number of items increases a larger sample size is required to provide stable inferences. This is not a surprising result. In traditional frequentist statistics, this would be the only option. However, in Bayesian methodology there are two potential additional options to stabilize inferences. First, the number of posterior samples can be increased. This would increase the precession of the estimates. Second, a prior can be specified, which will stabilize inferences that, in turn, will provide better coverage.

It should be noted that the purpose of this study was to demonstrate how a Bayesian internal consistency can be estimated under the basic assumptions made of reliability (Allen &

Table 3: Balpha and BalpahM Bayesian Credible Interval Coverage

Number of Items	Sample Size	Mean Item Correlation		
		.1667	.2208	.3103
5	50	**.974**, .963	.945, .951	.949, **.968**
	100	.949, .953	.959, .958	.942, .950
	150	.944, .956	.952, .960	.951, .953
	200	.947, .955	.946, .953	.942, .949
	250	.938, .948	.944, .948	.950, .954
	300	.845, .949	.959, .960	.948, .948
10	50	.961, **.971**	.952, **.968**	**.966**, **.980**
	100	.949, .961	.961, **.965**	**.968**, **.976**
	150	.951, .951	.960, .961	.954, .959
	200	.937, .942	.953, .958	.959, .963
	250	.951, .958	.945, .951	.963, .965
	300	.949, .953	.947, .954	.947, .949
15	50	**.979**, **.992**	**.977**, **.988**	**.970**, **.981**
	100	**.968**, **.969**	**.969**, **.970**	.962, **.969**
	150	.955, .955	.958, .964	.965, **.966**
	200	.954, .966	.963, .964	.941, .947
	250	.941, .944	.946, .956	.952, .959
	300	.957, .957	.945, .948	.955, .953
20	50	**.978**, **.991**	**.979**, **.997**	**.984**, **.989**
	100	**.967**, **.967**	.958, **.967**	**.973**, **.974**
	150	.960, **.966**	.962, .958	.962, **.968**
	200	.953, .963	.955, .960	.957, **.968**
	250	.950, .953	.960, .964	.962, .960
	300	.964, .968	.953, .958	.955, .968

*Note: The first number is the percentile BCIs followed by the normal theory based BCIs. Unacceptable coverage is bolded; acceptable coverage is within [.936, .964].

Yen, 1979; Crocker & Algina, 1986), thus, study provides a springboard from where future research on Bayesian coefficient alpha can be conducted. However, like any simulation study, this research is limited by the type and number of conditions investigated. In this study, only homogenous items were investigated.

Additionally, items were continuous and normally distributed. Further research is required to investigate the robustness of a Bayesian coefficient alpha to violations of the basic reliability assumptions and to establish its properties under binary or ordinal items.

As noted by Duhachek and Iacobucci (2004) and Maydeu-Olivares, et al. (2007), reporting only a point estimate of coefficient alpha is no longer sufficient. With inferential techniques reporting the SE and CIs provide more information as to the size and stability of the point estimate; in this case the point estimate is coefficient alpha. Within this context, a Bayesian internal consistency estimate may provide an attractive alternative to current coefficient alpha CIs because it provides researchers with BCIs that can be interpreted in a way researchers want and can make use of prior information to stabilize inferences.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks-Cole Publishing Company.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (*2nd Ed.*). New York: Wiley.

Barchard, K. A., & Hakstian, A. R. (1997). The robustness of confidence intervals for coefficient alpha under violation of the assumption of essential parallelism. *Multivariate Behavioral Research*, *32*(*2*), 169-191.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(*24*), 4279-4292.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(*1*), 98-104.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(*3*), 297-334.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science*, *19*(*1*), 1-15.

Duhachek, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the standard error of the coefficient alpha index of reliability. *Marketing Science*, *24*(*2*), 294-301.

Duhachek, A., & lacobucci, D. (2004). Alpha's atandard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology*, *89*(*5*), 792-808.

Feldt, L. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, *30*(*3*), 357-370.

Gelman, A. (2004). *Bayesian data analysis* (*2nd Ed.*). Boca Raton, Fla.: Chapman & Hall/CRC.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(*4*), 255-282.

Iacobucci, D., & Duhachek, A. (2003). Advancing Alpha: Measuring reliability with confidence. *Journal of Consumer Psychology*, *13*(*4*), 478-487.

Kristof, W. (1963). The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika*, *28*(*3*), 221-238.

Lee, P. M. (2004). *Bayesian statistics : An introduction* (*3rd Ed.*). London: Arnold.

Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, *12*(*2*), 157-176.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall/CRC.

van Zyl, J., Neudecker, H., & Nel, D. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, *65*(*3*), 271-280.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*(*8*), 594-604.

Yuan, K.-H., & Bentler, P. (2002). On robusiness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, *67*(*2*), 251-259.

Yuan, K.-H., Guarnaccia, C. A., & Hayslip, B. (2003). A study of the distribution of sample coefficient alpha with The Hopkins Symptom Checklist: Bootstrap versus asymptotics. *Educational and Psychological Measurement*, *63*(*1*), 5-23.