Received 09 March 2026
Revised 02 April 2026
Accepted 12 April 2026

Robustness of Bayesian coefficient alpha estimation under non-normal item distributions: Evidence from a Monte Carlo study using a normal posterior

Nguyen Ngoc Thach
Ho Chi Minh University of Banking

Abstract

Purpose – This study evaluates whether a recently proposed Bayesian estimator of coefficient alpha based on a normal posterior distribution remains reliable when applied to non-normal item responses, a common feature of behavioral and social science data.

Design/methodology/approach – A Monte Carlo simulation was conducted across 1,440 conditions defined by the number of items (5, 10, 15, and 20), correlation structure (two parallel and three congeneric models), sample size (50, 100, 150, 200, 250, and 300), item response categories (2, 3, 5, and 7), and three distribution types. Using non-informative priors and 2,000 posterior draws, the study examined three 95% Bayesian credible interval methods for coefficient alpha: percentile, normal-theory, and highest probability density intervals. Performance was assessed primarily through coverage probability, with interval width also considered.

Findings – The results show that all three credible interval methods performed similarly and generally achieved acceptable coverage across most simulation conditions. However, performance deteriorated when data involved binary items, distribution type 2, and/or a parallel model with common loadings of .705, particularly when these conditions occurred jointly. The method was more dependable when items had at least three response categories and sufficient variability.

Research limitations/implications: The findings are limited to the simulation settings examined and to the normal-posterior framework used to estimate coefficient alpha. The results suggest caution when applying this approach to binary, highly skewed, or range-restricted data, and indicate the need for future comparisons with alternative Bayesian and non-Bayesian estimators, as well as sampling-based approaches such as Gibbs sampling.

Originality/value – This study extends recent Bayesian research on coefficient alpha by explicitly testing the robustness of a computationally convenient normal-posterior estimator under non-normal item distributions. It clarifies the empirical conditions under which the method performs well and identifies situations in which its use should be treated more cautiously.

Keywords coefficient alpha, Bayesian estimation, CrIs, non-normal distributions, Monte Carlo simulation
Paper type Research article

JEL Classification — C11, C58.

© Nguyen Ngoc Thach. Publish in Asian Journal of Applied Financial Econometrics. The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest Statement: The author(s) declare(s) that there are no conflicts of interest regarding the publication of this article. The author(s) have no financial or personal relationships with other people or organizations that could inappropriately influence (bias) this work.

Data Availability Statement: The data used in this study were obtained from publicly available sources. The dataset supporting the findings of this study is available from the corresponding author Nguyen Ngoc Thach upon reasonable request.

Declaration of AI-Assisted Technologies: The author(s) declare(s) that no generative AI or AI-assisted technologies were used in the creation of this manuscript. All content was written, reviewed, and approved solely by the listed author(s).

1. Problem Formulation

Coefficient alpha is one of the most popular statistics utilized to estimate the reliability of items of a measurement instrument for a construct. When considering a measurement instrument that contains k items, $x_1, x_2, \ldots, x_k$, coefficient alpha is defined as

$$\alpha_c = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k}\sigma_{x_{ii}}}{\sigma_X^2}\right) \quad (1)$$

where $\sigma_{x_{ii}}$ is the variance of item i and σ_X^2 is the total variance of the set of items for the measurement instrument (Cronbach, 1951). Similarly, for a sample size n, the population parameters in equation 1 can be replaced by sample estimates to obtain a coefficient alpha estimate as

$$\alpha_c = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k}\hat{\sigma}_{x_{ii}}}{\hat{\sigma}_X^2}\right), \quad (2)$$

The enduring popularity of this statistic is rooted in three key qualities (Padilla, 2019). First, coefficient alpha is easy to calculate as it only requires the item covariance matrix to compute. In contrast, less popular forms of reliability are more difficult to compute and thus are not as often used. Second, coefficient alpha is viable for continuous, ordinal (i.e., Likert-type), and binary items. This allows coefficient alpha to be used with a variety of item response formats. Third, estimation of coefficient alpha only requires one administration of a measurement instrument. This is a boon when most other forms of reliability require at least two administrations to estimate. The wide use of coefficient alpha in applied research has consequently made coefficient alpha itself a subject to a great deal of research as well. Presently, the focus will be on how assumptions about the items used to estimate coefficient alpha effect the usage and performance of interval estimates about coefficient alpha and how advancements have been made to overcome limitations related to these assumptions.

2. Conceptual Framework

2.1. Coefficient Alpha in the Context of Measurement Theory

Coefficient alpha is an index of reliability, but it is only appropriate when items from a measurement instrument follow certain relationships as defined by measurement models from classical test theory (Furr & Bacharach, 2014). If the items have equal true scores and equal variability in measurement errors, items follow a parallel model. This results in items having a compound symmetric covariance matrix. If the items have equal true scores and unequal variability in measurement errors, items follow an (essentially) tau-equivalent model. This results in items having a covariance matrix that is compound symmetric with heterogeneous variances. If the items have unequal true scores and unequal variability in measurement errors, items follow a congeneric model. This results in items having a covariance matrix that is first-order heterogenous. For all measurement models, the measurement errors are assumed to be independent from the true scores and from the measurement errors of other items or separate occasions of the same item. These measurement models are important to consider as coefficient alpha is only equal to reliability (i.e., $\alpha_c = \rho_{xx'}$) when items follow a parallel or (essentially) tau-equivalent model. If the items follow a congeneric model, coefficient alpha can be shown to be only a lower bound to reliability (i.e., $\alpha_c \leq \rho_{xx'}$); Novick, 1967). As such, it is important to consider measurement model assumptions when using coefficient alpha as its interpretation changes depending on which measurement model the items follow.

2.2. Constructing Confidence Intervals for Coefficient Alpha

Recent research has focused on the properties of interval estimation for coefficient alpha. Interval estimation is of interest because it accounts for sampling error by way of estimating a range of possible values like to contain a population parameter (Beaulieu-Prevost, 2006; Benjamin et al., 2018; Wilkinson, 1999). The most common type of interval estimate has been the confidence interval (CI). However, developing CIs for coefficient alpha has been challenging.

Much of the challenge in developing CIs for coefficient alpha has been in relation to overcoming assumptions about the estimation of coefficient alpha CIs that limit their applicability.

One such assumption to address was assuming that items follow a parallel measurement model (i.e., items having a compound symmetric covariance matrix). An initial CI for coefficient alpha was proposed by Feldt (1965) by deriving the sampling distribution of coefficient alpha assuming items were parallel and were normally distributed. Barchard & Hakstian (1997) investigated the performance of this CI via simulation and found that the CI did not perform well if items were not parallel. This was significant as requiring items to be parallel is an assumption that is not typically met in applied settings. As a result, this early coefficient alpha CI was rarely used in applied settings (Duhachek & Iacobucci, 2004). Later, van Zyl et al. (2000) showed that coefficient alpha in equation 2 is a maximum likelihood estimate (MLE) with a normal asymptotic distribution. This coefficient alpha MLE assumes items are normally distributed but does not require items to be parallel (i.e., does not require items to have a compound symmetric covariance matrix). Duhachek & Iaobucci (2004) would build off this finding to propose a normal theory (NT) CI for coefficient alpha. The significant contribution of this research was that it led to developing coefficient alpha CIs that were not limited to the parallel item assumption though still required items to be normally distributed.

This leads to the assumption of normally distributed items. Measurement instrument items used in the behavioral/social sciences tend to be Likert-type or dichotomous (e.g., yes/no, true/false, etc.), making the normality assumption of items difficult to meet in practice. As such, another assumption to address was assuming items are normally distributed. To help overcome this limitation, Yuan et al. (2003) proposed a bootstrap CI for coefficient alpha. The advantage of the bootstrap is that it makes no distributional assumptions as it uses resamples of the data (i.e., bootstrap samples) to generate an empirical sampling distribution (ESD) of a statistic from which inferences can be made. Additionally, the authors proposed a new, asymptotic distribution-free (ADF) CI for coefficient alpha, and compared the performance of the bootstrap, ADF, and NT CIs estimated from the Hopkins Symptom checklist (Derogatis et al., 1974). This comparison showed that the bootstrap CIs were most accurate. However, the ADF CI performed similarly (within three decimals) and were less computationally intensive than the bootstrap CIs, resulting in further investigation for the ADF CI (Maydeu-Olivares et al., 2007). Maydeu-Olivares et al. (2007) compared the ADF CI against the NT CI via simulation and found that the ADF CI had better coverage probability (i.e., the proportion of estimated CIs that contained the true population coefficient alpha) than the NT CI when items were non-normal.

Further refinements in coefficient alpha CI research would take place as more CIs developed. This refinement generally focused on the conditions used to determine the best performing CI. For example, Romano et al. (2010) compared eight coefficient alpha CIs via simulation with a focus on dichotomous items. Here, it was found that the Fisher z-transformation and Bonett z-transformation CIs had the best coverage probability. Likewise, Padilla et al. (2012) compared three bootstrap coefficient alpha CIs and four non-bootstrap coefficient alpha CIs via simulation with a focus on distribution types and item correlation types (i.e., measurement model types). In contrast to previous findings, the Fisher z-transformation, Bonett z-transformation, and NT CIs had the most instances of unacceptable coverage probability. Additionally, they found that the NT bootstrap had the best coverage probability across the simulation conditions.

2.3. Robustness of Alternative Bayesian Estimators for Coefficient Alpha

The previously discussed literature outlined a brief history of CI development for coefficient alpha. These developments also contributed to a greater understanding of the coefficient alpha distribution and have been leveraged to approach coefficient alpha from a Bayesian perspective. This is significant as it allows coefficient alpha estimates to be updated with prior information as well as provide a more intuitive interpretation of the corresponding Bayesian credible interval (CrI) estimates. Priors are of particular benefit in that they can be used to supplement parameter estimates when there is a small sample size. Additionally, rather than interpret a 95% CI as 95% of all similarly constructed CIs containing the parameter of interest; the 95% CrI is instead interpreted as the parameter of interest as having a 95% chance of being within the bounds of the CrI given the data.

Recently, an alternative method to obtain a Bayesian estimate of coefficient alpha was proposed (DelosReyes & Padilla, 2025). This was done by leveraging the results from van Zyl et al. (2000) wherein the MLE of coefficient alpha has a normal asymptotic distribution. From here, the posterior distribution for coefficient alpha was defined first by using conjugate priors such that the priors selected would result in a closed-form expression for the posterior. In cases without prior information, improper priors that do not integrate to 1 yet still result in posterior distributions that do integrate to 1 were selected. Given that this method is built on the van Zyl method, it is subject

to its perks and penalties. For example, it does assume items are parallel but assumes items are normally distributed.

The authors assessed this alternative Bayesian estimate of coefficient alpha using three types of 95% CrIs via simulation (DelosReyes & Padilla, 2025). The first was percentile-based, the second was NT-based, and the third was highest probability density (HPD) based. The simulation conditions included a) number of items ($k = 5, 10, 15, 20$), b) number of item response categories ($M = 2, 3, 5, 7$), c) correlation type, and d) sample size ($n = 50, 100, 150, 200, 250, 300$). For correlation type, five different correlation structures were investigated. The first two correlation structures were from a parallel-item one-factor model with common loadings of $\lambda_1 = 0.55$ and $\lambda_1 = 0.705$. The last three correlation structures were generated from a congeneric-item one-factor model with loadings: $\lambda_3 = 0.2, 0.3, 0.4, 0.5, 0.6$; $\lambda_4 = 0.3, 0.4, 0.5, 0.6, 0.7$; and $\lambda_5 = 0.4, 0.5, 0.6, 0.7, 0.8$. The investigated correlation structures resulted in population coefficient alphas ranging from .36 – .96. Items were binary or Likert-type and set to be (approximately) normal. For each simulation condition, 1,000 replications were obtained. Priors used for the Bayesian estimate of coefficient alpha were set to be non-informative.

The CrIs investigated generally had acceptable coverage probability across the simulation conditions but tended to have instances of unacceptable performance when a sample size of 50 was paired with certain conditions. Notably, all CrIs had unacceptable coverage when there was a congeneric model with factor loading of 0.4 – 0.8 paired with a sample size 50. In addition, the percentile and HPD CrIs had unacceptable coverage when there were 20 items paired with a sample size 50. Finally, the percentile CrI had unacceptable coverage when two item response categories were paired with a sample size 50. However, for all CrI methods, performance tended to be acceptable and stabilize to the target criteria of .95 as sample size increased.

The proposed Bayesian estimate of coefficient alpha has two noticeable features (DelosReyes & Padilla, 2025). First, it is viable if sample size was large enough (i.e.,). Second, it works when items were parallel or congeneric. However, something to note is that this method is built upon the findings by van Zyl et al. (2000) that assume normally distributed items. Furthermore, the simulation used to assess the alternative Bayesian estimation method only used items that were (approximately) normal. Therefore, it is of interest to determine if the method is viable when items are non-normal.

2.4. Parameter Estimation via Bayesian Credible Sets

The CrI estimates for the study were based on the Bayesian estimate of coefficient alpha as proposed by DelosReyes & Padilla (2025). Worth noting is when assuming normal items, van Zyl et al. (2000) showed that

$$\hat{\alpha}_c \sim \mathrm{N}\left(\alpha_c, \frac{\sigma_\alpha^2}{n}\right), \tag{3}$$

Where

$$\sigma_\alpha^2 = \left(\frac{k}{k-1}\right)^2 \left[\frac{2}{\left(1'\Sigma 1\right)^3}\right]\left[\left(1'\Sigma 1\right)\left(tr\left(\Sigma^2\right)+tr\left(\Sigma\right)^2\right)-2tr\left(\Sigma\right)\left(1'\Sigma^2 1\right)\right] \tag{4}$$

The posterior distribution, $\pi\left(\alpha_c \mid \boldsymbol{y},\ \sigma_\alpha^2\right)$, can then be summarized for inferences. For further details, see DelosReyes and Padilla (2025). For this study, three different CrIs are of interest. A percentile CrI can be obtained using the lower $\alpha/2$ and upper $1-\alpha/2$ of $\pi\left(\alpha_c \mid \boldsymbol{y},\ \sigma_\alpha^2\right)$. A NT CrI can be obtained with

$$\pi\left(\alpha_c,\ \sigma_\alpha^2 \mid \mathrm{y}\right) = \frac{p\left(\mathrm{y} \mid \alpha_c, \sigma_\alpha^2\right) p\left(\alpha_c, \sigma_\alpha^2\right)}{p\left(\mathrm{y}\right)} \tag{5}$$

The posterior distribution, $\pi\left(\alpha \mid \mathrm{y},\ \sigma\right)$, can then be summarized for inferences. For further details, see DelosReyes and Padilla (2025). For this study, three different CrIs are of interest. A percentile CrI can be obtained using the lower $\frac{\alpha}{2}$ and upper $1-\frac{\alpha}{2}$ of $\pi\left(\alpha_c \mid \mathrm{y},\ \sigma_\alpha^2\right)$. A NT CrI can be obtained with

$$\hat{\alpha}_c \pm z_{\alpha/2} SE\left(\hat{\alpha}_c\right) \tag{6}$$

where $SE(\hat{\alpha}_c)$ is the standard deviation of $\pi\left(\alpha_c \mid \mathrm{y},\ \sigma_\alpha^2\right)$. A HPD CrI can be obtained by letting a subset of the parameter space for $\hat{\alpha}_c$, $c \subset A$ be defined as

$$c = \left\{\hat{\alpha}_c :\ \pi\left(\alpha_c \mid \mathrm{y},\ \sigma_\alpha^2\right) \geq h\right\} \tag{7}$$

where h is the largest number such that

$$\int_h \pi\left(\alpha_c \mid \mathrm{y},\ \sigma_\alpha^2\right) d\alpha_c = 1 - \alpha \tag{8}$$

3. Methodology

3.1. Simulation

A Monte Carlo simulation was used to investigate the properties of the Bayesian estimate of coefficient alpha and its corresponding CrIs. The simulation conditions were 4 (number of items) × 5 (correlation type) × 6 (sample size) × 4 (item response categories) × 3 (distribution type) for a total of 1,440 conditions. All items were Likert-type (ordinal) or binary. For each simulation condition, 1,000 replications were obtained. For consistency and comparison with other CrIs and CIs in the previous literature, non-informative priors as presented in DelosReyes and Padilla (2025) were used. In addition, 2,000 posterior draws used. An outline of the Monte Carlo simulation for the study is given below:

1. Select the structure of the $k \times k$ correlation matrix **P**, where k is the number of items.
2. Select a set of thresholds ν to categorize items to a predetermined skewness and kurtosis.
3. Generate an $n \times k$ multivariate data matrix **Z** ~ N(0, **P**), where n is the sample size.
4. Categorize the generated data **Z** using the thresholds in ν to generate the dataset **X**. Each variable x in **X** is categorized by the thresholds as follows: x = m if $\nu_m < z < \nu_{m+1}$ for m = 0, 1, …, M - 1, where $\nu_0 = -\infty$ and $\nu_m = \infty$, and M is the number of categories.
5. Compute the true population coefficient alpha (α_c) according to P and the thresholds in ν. See Maydeu-Olivares et al. (2007) for details.
6. Estimate the Bayesian coefficient alpha CrIs from X as outlined in DelosReyes & Padilla,
7. 2025.
8. Determine if the CrIs contain the true population coefficient alpha (α_c).

Details of the simulation conditions are given below.

3.2. Conditions

Number of Items (k)

Previous research on coefficient alpha has investigated the number of items ranging from 3 to 40 (Barchard & Hakstian, 1997; Duhachek & Iacobucci, 2004; Maydeu-Olivares et al., 2007; Padilla et al., 2012; Padilla & Zhang, 2011; van Zyl et al., 2000). However, most of that previous research focused on items ranging from 5 to 20. Additionally, it is noted that going beyond 20 items for a measurement instrument reaches a point of diminishing returns with regards to coefficient alpha. Therefore, the following number of items were investigated: $k = 5, 10, 15, 20$.

Correlation Type (P)

Five different item correlation structures P were investigated. The first two correlation structures were from a parallel-item one factor model with common loadings $\lambda_1 = 0.55$ or $\lambda_2 = 0.705$. These two models will generate compound symmetric item correlation structures with $\rho = 0.30$ or .56, respectively. The next three correlation structures were generated from a congeneric item one-factor model with the following loadings: $\lambda_3 = 0.2, 0.3, 0.4, 0.5, 0.6$; $\lambda_4 = 0.3, 0.4, 0.5, 0.6, 0.7$; and $\lambda_5 = 0.4, 0.5, 0.6, 0.7, 0.8$. These correlation structures were chosen as they were previously investigated and allow for a greater range of conditions to explore the impact on coefficient alpha CIs/CrIs (Maydeu-Olivares et al., 2007; Padilla et al., 2012). Note that the loadings are standardized for these factor models. A summary for the population coefficient alpha these correlation structures produce is given in Table 1.

Sample Size (n)

The following sample sizes were investigated: n = 50, 100, 150, 200, 250, 300. This selection is in line with most recent research on coefficient alpha (Duhachek & Iacobucci, 2004; Maydeu-Olivares et al., 2007; Padilla et al., 2012; Padilla & Zhang, 2011). The sample size selections above 200

Table 1. Summary of Resultant Population Coefficient Alphas from Correlation Structure Types

Correlation Structure Type	Minimum α_c	Mean α_c	Maximum α_c
Parallel = 0.55	0.48	0.75	0.89
λ = 0.705	0.71	0.88	0.96
Congeneric: λ = 0.2,0.3,0.4,0.5,0.6	0.30	0.59	0.78
λ = 0.3,0.4,0.5,0.6,0.7	0.42	0.71	0.86
λ = 0.4,0.5,0.6,0.7,0.8	0.55	0.79	0.91

Note: λ = standardized factor loadings

are beyond the point of diminishing returns for coefficient alpha but were investigated as they are typically found in psychological/behavioral research.

Item Response Categories (M)

Previous research on coefficient alpha has investigated numbers of item response categories ranging from 2 to 7 (Maydeu-Olivares et al., 2007; Padilla et al., 2012; Romano et al., 2010; Yuan et al., 2003). To draw parity with previous research, the following common choices for response categories were investigated: M = 2, 3, 5, 7. For each response category, the first category was set to 0. For example, for an item with seven response categories, the categories were m = 0, 1, 2, 3, 4, 5, 6.

Distribution Type

Three different distribution types were investigated that are dependent on the number of item response categories (M). When M = 2 (i.e., items are binary), the thresholds for ν were chosen such that the distributions had the following characteristics:

2. Type 1: skewness = 0 and kurtosis = −2
3. Type 2: skewness = −1.70 and kurtosis = 0.88
4. Type 3: skewness = 0.41 and kurtosis = −1.83

The type 1 and 2 distributions for binary items were studied by Padilla et. al. (2012). The type 3 distribution for binary items was studied by Maydeu - Olivares et al. (2007). When (i.e., M > 2) the thresholds for ν were chosen such that the distribution had the following characteristics:

1. Type 1: skewness = 0 and kurtosis = 0
2. Type 2: skewness = 0 and kurtosis = 0.88
3. Type 3: skewness = 0.97 and kurtosis = -0.20

The type 1 distribution for M > 2 items was studied by Padilla et. al. (2012). The type 2 and 3 distributions for studied by Maydeu-Olivares et al. (2007). Investigating this breadth of distribution types allows for the assessment of how the alternative Bayesian estimation of coefficient alpha responds to deviations from normality and draws parity with past research. The distribution types for items with two and five categories are presented in Figure 1.

3.3. Analysis

To evaluate the performance of the CrIs, coverage probability was used. Coverage probability is defined as the proportion of estimated CrIs that contain the true population coefficient alpha. Coverage probability was assessed using Bradley's (1978) liberal criterion defined by

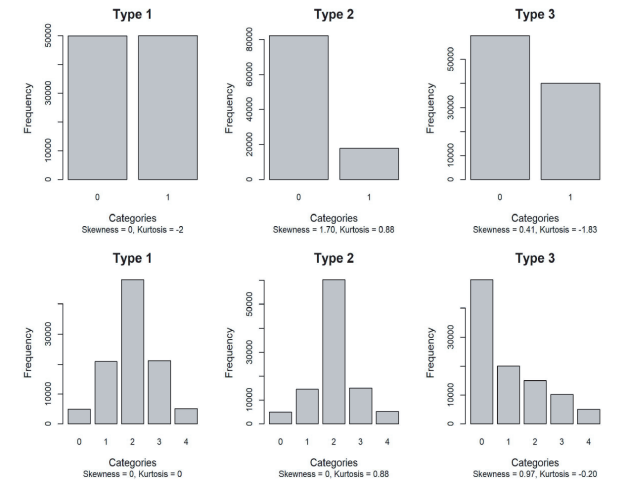
$$1 - 1.5\alpha \leq 1 - \alpha^* \leq 1 - 0.5\alpha, \tag{9}$$

where α^* is the true type I error probability. Here the $100(1-\alpha)\%$ CrIs were estimated with $\alpha = .05$. Therefore, acceptable coverage probability is given by [.925, .975].

CrI width was also considered as a criterion for assessing CrI performance. CrI width gives information about the precision of a CrI estimate (Anderson, 2019). It is possible for two CrI methods to have the same coverage but with different interval widths (i.e., levels of precision). As such, CrI width is relevant if the CrI methods performed similarly in coverage probability.

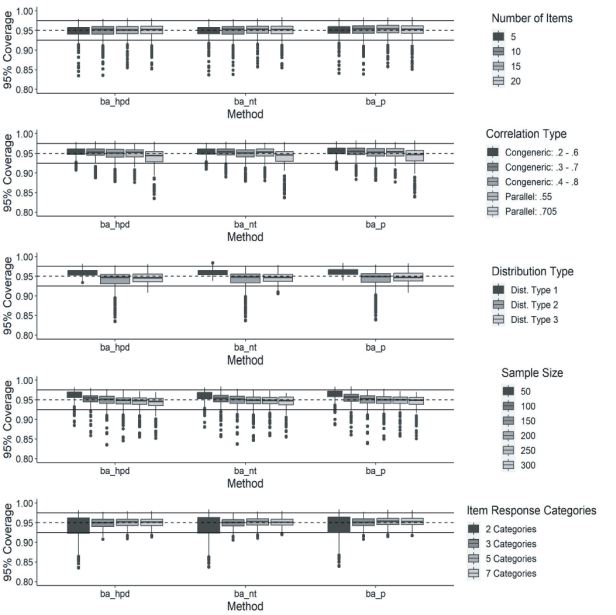
4. Results

Figure 2 shows the coverage probability performance for each CrI across the simulation main



Source(s): Figure by the authors

Figure 1. Bar Graphs of the Dichotomous and 5-point Likert Items Utilized in Simulation



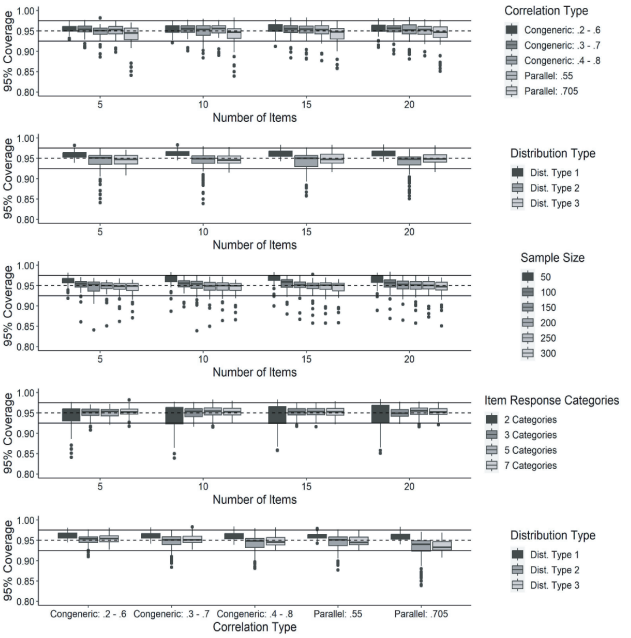
Source(s): Figure by the authors

Figure 2. Distribution of the 95% Credible Intervals for all Main Effect Simulation Conditions

effects (i.e., the simulation of conditions of , and Distribution Type individually). It was found that all CrI methods performed similarly to one another and generally had acceptable coverage probability across all simulation conditions. However, there are some key patterns to notice across all CrI methods investigated. For the item response categories condition, there is more variability in coverage probability when there are only 2 categories compared to all other categories. In addition, for the distribution type condition, there is more variability in coverage probability in distribution type 2 compared to all other conditions. Lastly, for the sample size condition, coverage probability tended to be near the upper bound of the coverage probability criteria (i.e., .975) when sample size was 50 stabilized to .95 as sample size increased.

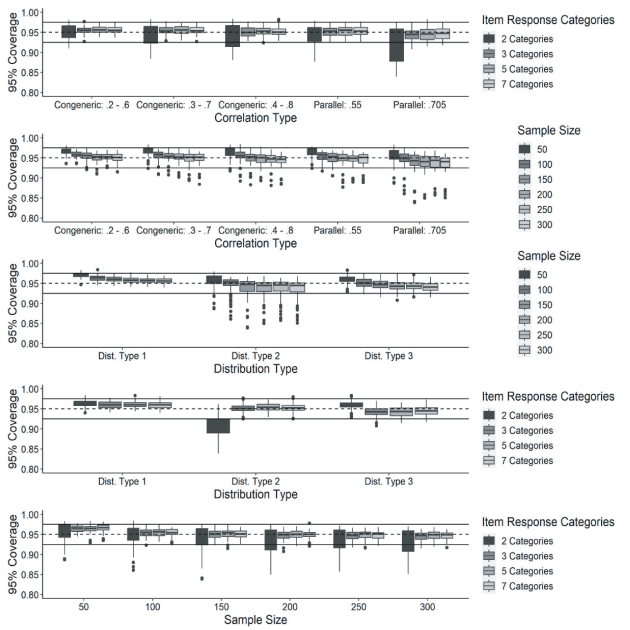
Figures 3 – 5 show the coverage probability performance of each CrI method based on the pairwise simulation conditions (i.e., any pair of combinations of the simulation conditions

of k, P, n, M, and Distribution Type). Again, all CrIs performed similarly to one another and generally had acceptable coverage probability across all pairwise simulation conditions. However, there were some performance discrepancies amongst the CrI methods depending on the specific combination of pairwise simulation conditions.



Note: Bayesian alpha Percentile credible interval based on 2,000 posterior draws; Dashed line at .95 and solid lines at acceptable coverage of [.925, .975].

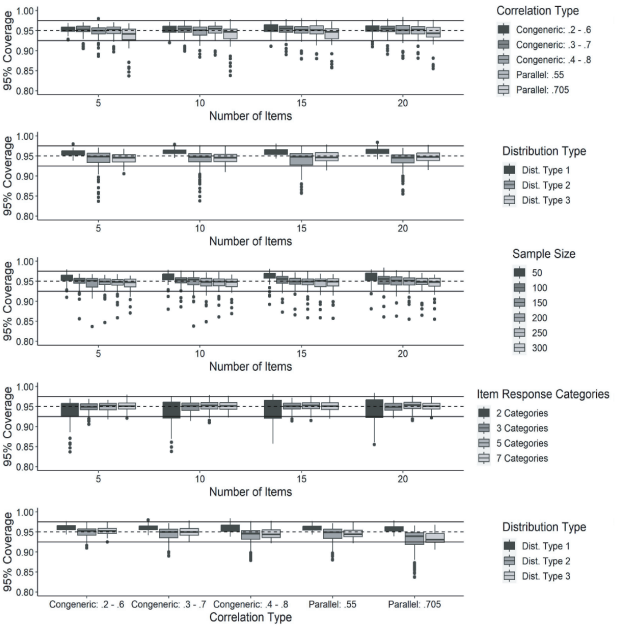
Figure 3. Percentile 95% Credible Interval Coverage for all Pairwise Simulation Conditions



Note: Bayesian alpha Percentile credible interval based on 2,000 posterior draws; Dashed line at .95 and solid lines at acceptable coverage of [.925, .975].

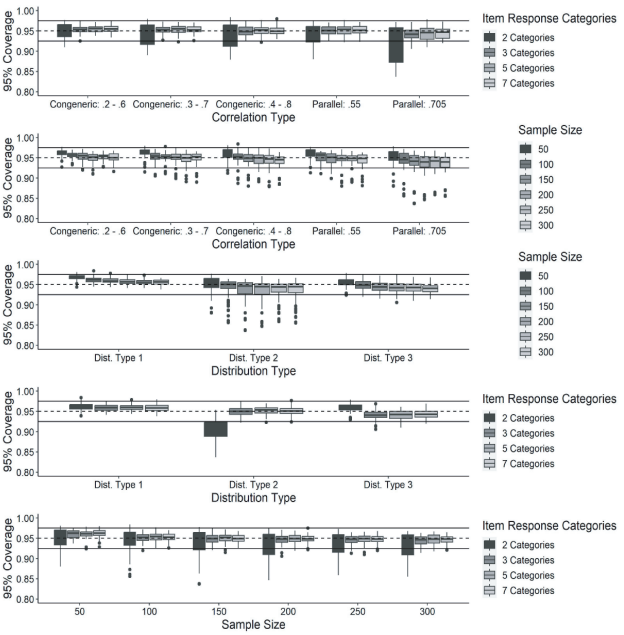
Figure 3 Continued. Percentile 95% Credible Interval Coverage for all Pairwise Simulation Conditions

For the percentile CrI (Figure 3), the simulation conditions that involved binary items (M = 2), distribution type 2, and/or a parallel model with common loadings of .705 tended to have unacceptable coverage. For binary items, performance was impacted across all number of items; for all correlation types except a congeneric model with factor loadings of 0.2 – 0.6; or



Note: Bayesian alpha normal theory credible interval based on 2,000 posterior draws; Dashed line at .95 and solid lines at acceptable coverage of [.925, .975].

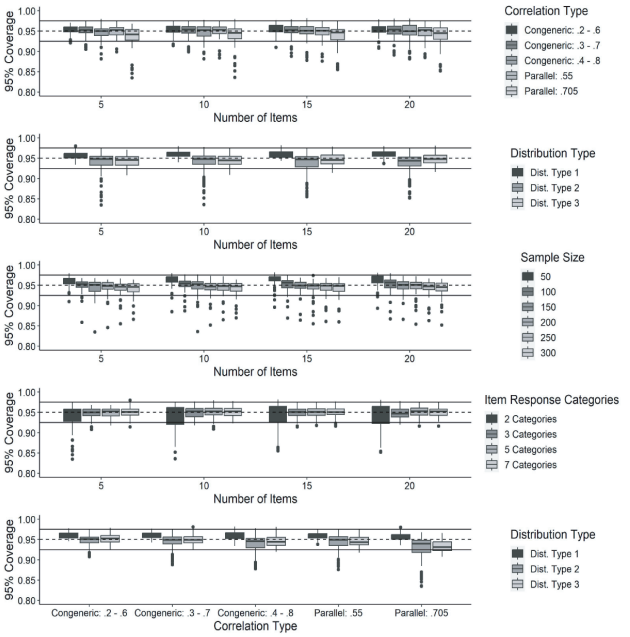
Figure 4. Normal Theory 95% Credible Interval Coverage for all Pairwise Simulation Conditions



Note: Bayesian alpha Percentile credible interval based on 2,000 posterior draws; Dashed line at .95 and solid lines at acceptable coverage of [.925, .975].

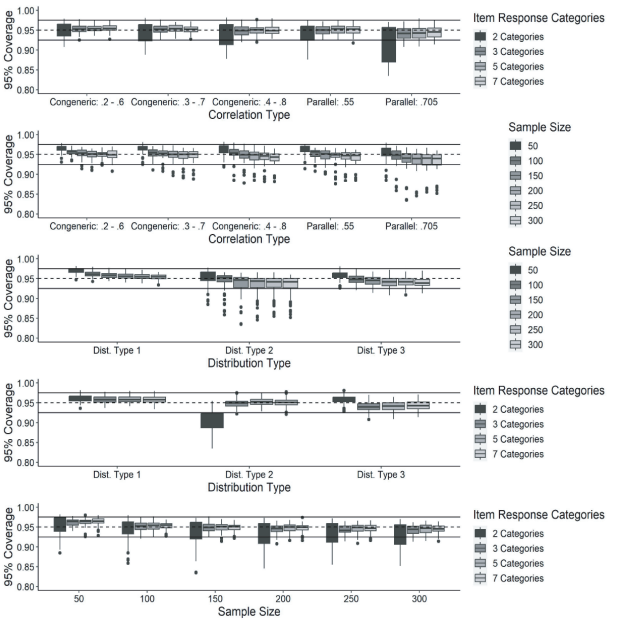
Figure 4 Continued. Normal Theory 95% Credible Interval Coverage for all Pairwise Simulation Conditions

with a sample size of 150 – 300. For distribution type 2, performance was impacted across all number of items; or with sample size of 100 – 300. For a parallel model with common loadings of .705, performance was impacted across all number of items; or with a sample size of 100 – 300. Unsurprisingly, the pairwise combinations of binary items with distribution type 2; distribution



Note: Bayesian alpha highest probability density credible interval based on 2,000 posterior draws; Dashed line at .95 and solid lines at acceptable coverage of [.925, .975].

Figure 5. Highest Probability Density 95% Credible Interval Coverage for all Pairwise Simulation Conditions



Note: Bayesian alpha Percentile credible interval based on 2,000 posterior draws; Dashed line at .95 and solid lines at acceptable coverage of [.925, .975].

Figure 5 Continued. Highest Probability Density 95% Credible Interval Coverage for all Pairwise Simulation Conditions

type 2 with a parallel model with common loadings of .705; or binary items with a parallel model with common loadings of .705 had the largest performance impact of unacceptable coverage.

For the NT CrI (Figure 4), the simulation conditions that involved binary items (M =2), distribution type 2, and/or a parallel model with common loadings of .705 tended to have unacceptable coverage. For binary items, performance was impacted across all number of items; for all correlation types except a congeneric model with factor loading of 0.2 – 0.6 ; or with a sample size of 150 – 300. For distribution type 2, performance was impacted across all number of items; or with sample size of 100 - 300. For a parallel model with common loadings of .705, performance was impacted across all number of items; or sample size of 100 – 300. Like the percentile CrI, the pairwise combinations of binary items with distribution type 2; distribution type 2 with a parallel model with common loadings of .705; or binary items with a parallel model with common loadings of .705 had the largest performance impact of unacceptable coverage.

For the HPD CrI (Figure 5), the simulation conditions that involved binary items (M = 2), distribution type 2, and/or a parallel model with common loadings of .705 tended to have unacceptable coverage. For binary items, performance was impacted across all number of items; for all correlation types except a congeneric model with factor loadings of 0.2 – 0.6; or with a sample size of 150 – 300. For distribution type 2, performance was impacted across all number of items; or with a sample size of 100 – 300. For a parallel model with common loadings of .705, performance was impacted across all number of items; or with sample size of 100 – 300. As with the percentile and NT CrIs, the pairwise combinations of binary items with distribution type 2; distribution type 2 with a parallel model with common loadings of .705; or binary items with a parallel model with common loadings of .705 had the largest performance impact of unacceptable coverage.

In summary, the CrIs investigated were impacted by the same simulation conditions and had similar performance. Specifically, simulation conditions that involved binary items (M = 2), distribution type 2, and/or a parallel model with common loadings of .705 tended to have unacceptable coverage. For binary items, performance was impacted across all number of items; for all correlation types except a congeneric model with factor loadings of 0.2 – 0.6; or with a sample size of 150 – 300. For distribution type 2, performance was impacted across all number of items; or with a sample size of 100 – 300. For a parallel model with common loadings of .705, performance was impacted across all number of items; or with sample size of 100 – 300.

Additionally, the pairwise combinations of binary items with distribution type 2; distribution type 2 with a parallel model with common loadings of .705; or binary items with a parallel model with common loadings of .705 had the largest performance impact with unacceptable coverage.

5. Conclusion and Discussion

An alternative method to obtain a Bayesian estimate of coefficient alpha based on a normal posterior was recently introduced and assessed (DelosReyes & Padilla, 2025). That initial proposal showed promising results as the CrIs investigated showed generally good results even when items were not parallel (i.e., compound symmetric). However, the method assumes items are normally distributed. Therefore, it is of interest to assess how the alternative method performs when using non-normal items.

Expanding on that previous research, the effect non-normal items had on the alternative Bayesian estimate of coefficient was assessed via simulations. Generally, the CrIs investigated performed well and similarly to one another. However, it was found that simulation conditions that involved binary items, distribution type 2, and/or a parallel model with common loadings of .705 tended to have negatively impacted performance. This effect was further pronounced if these simulation conditions were paired with one another. Even so, the results are limited to the conditions investigated in the current study.

A common feature of these conditions is that they have some form of range restriction. First, the alternative method CrIs are based on making direct draws from a normal posterior distribution. However, binary items cannot achieve normality and are thus at odds with draws from the normal posterior distribution. This can be seen as binary items having a range restriction that cannot accurately reflect the variability in a normal distribution. Second, items with skewed distributions tend to have range restrictions because of floor/ceiling effects. This can be seen for the items with skewness greater or equal to 0.97 from figure 1. Lastly, a parallel model with a common loading of .705 has a corresponding compound symmetric item correlation structure with . This created the strongest coefficient alphas with an average coefficient alpha of .88 (see table 1). This created a

range restriction at the upper limit of 1 for coefficient alpha. Any one of these conditions negatively impacts the CrIs, but any combination of them exacerbated that impact. Range restrictions attenuate variance and accurate variance is necessary to have accurate CrIs.

There are a couple directions to expand on the current study. First, it would be of interest to directly compare the performance of the alternative method for obtaining a Bayesian estimate of coefficient alpha by DelosReyes and Padilla (2025) to other methods. These comparisons were not done here as the focus was to evaluate the viability of the method on non-normal data. Second, given how the CrIs were impacted by binary items, it may be interest to explore using a Gibbs sampler when making posterior draws (Hoff, 2009). Unlike the MCMC used here that directly makes draws for a normal posterior, an advantage of the Gibbs sample is that it cycles through data at each draw (i.e., iteration). This cycling though the data allows the data to have more impact on the posterior, which in turn can help improve the situation with binary items. Incidentally, cycling through the data at each draw may also help with skewed items (i.e., non-normal items).

All things considered, the simulation demonstrated that the alternative method to obtain a Bayesian estimate of coefficient alpha was found to be generally viable for non-normal items. There was a negative performance impact if items were binary, came from a distribution type 2, and/or a parallel model with common loadings of .705. Conversely, it was shown that the method is viable if items have at least three item response categories, have a distribution less kurtotic than distribution type 2 (i.e., kurtosos < 0.88), and/or have more variability than a parallel model with common loadings of .705 would result in. Essentially, the method works for non-normal items if the items are allowed to have a healthy amount of variability.

References

Anderson, A. A. (2019). Assessing statistical results: Magnitude, precision, and model uncertainty. *The American Statistician*, 73, 118-121. doi:10.1080/00031305.2018.1537889.

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. doi:10.1207/s15327906mbr3202_4.

Beaulieu-Prevost, D. (2006). Confidence intervals: From tests of statistical significance to confidence intervals, range hypotheses, and substantial effects. *Tutorials in Quantitativve Methods for Psychology*, 2(1), 11-19. doi:10.20982/tqmp.02.1.p011.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., . . . Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behavior*, 2, 6-10. doi:10.1038/s41562-017-0189-z.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2), 144-152. doi:10.1111/j.2044-8317.1978.tb00581.x.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi:10.1007/BF02310555.

DelosReyes, J. M. V., & Padilla, M. A. (2025). Obtaining a Bayesian estimate of coefficient alpha using a posterior normal distribution. *Educational and Psychological Measurement*, 0(0). doi:10.1177/00131644241311877.

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. doi:10.1002/bs.3830190102.

Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): an accurate and precise confidence interval estimate. *The Journal of Applied Psychology*, 89(5). doi:10.1037/0021-9010.89.5.792.

Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, 30(3), 357-370. doi:10.1007/BF02289499.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods.* London; New York: Springer.

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. doi:10.1037/1082-989X.12.2.157.

Novick, M. R. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32(1), 1-13. doi:10.1007/BF02289400.

Padilla, M. A. (2019). A primer on reliability via coefficient alpha and omega. *Archives of Psychology*, 3(8). doi:10.31296/aop.v3i8.125.

Padilla, M. A., Divers, J., & Newton, M. (2012). Coefficient alpha bootstrap confidence interval under nonnormality. *Applied Psychological Measurement*, 36(5), 331-348. doi:10.1177/0146621612445470.

Padilla, M. A., & Zhang, G. (2011). Estimating internal consistency using Bayesian methods. *Journal of Modern Applied Statistical Methods*, 10(1), 277-286. doi:10.22237/jmasm/1304223840.

Romano, J. L., Kromrey, J. D., & Hibbard, S. T. (2010). A Monte Carlo study of eight confidence interval methods for coefficient alpha. *Educational and Psychological Measurement*, 70(3), 376-393. doi:10.1177/0013164409355690.

van Zyl, J. M., Neudecker, H., & Nel, D. G. (2000). On the distribution of the maximum likelihood estimator of Cronbach's alpha. *Psychometrika*, 65(3), 271-280. doi:10.1007/bf02296146.

Wilkinson, L. (1999). Statistical methods in psychology journals. *American Psychologist*, 54(8), 594-604. doi:10.1037/0003-066X.54.8.594.

Yuan, K. H., Guarnaccia, C. A., & Hayslip, B., Jr. (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. doi:10.1177/0013164402239314.

Corresponding author
Nguyen Ngoc Thach can be contacted at: thachnn@hub.edu.vn