MEASUREMENT: INTERDISCIPLINARY RESEARCH AND PERSPECTIVES
2022, VOL. 20, NO. 4, 215–227
https://doi.org/10.1080/15366367.2021.2005959

# Applications of Bayesian Confirmatory Factor Analysis in Behavioral Measurement: Strong Convergence of a Bayesian Parameter Estimator

Tenko Raykov[a], Philipp Doebler[b], and George A. Marcoulides[c]

[a]Measurement and Quantitative Methods, Michigan State University; [b]Department of Statistics, TU Dortmund University; [c]Gervitz Graduate School of Education, University of California

**ABSTRACT**
This article is concerned with the large-sample parameter estimatorbehavior in applications of Bayesian confirmatory factor analysis in behavioral measurement. The property of strong convergence of the popular Bayesian posterior median estimator is discussed, which states numerical convergence with probability 1 of the resulting estimates to the population parameter value as sample size increases without bound. This property is stronger than the consistency and convergence in distribution of that estimator, which have been commonly referred to in the literature. A numerical example is utilized to illustrate this almost sure convergence of a Bayesian latent correlation estimator. The paper contributes to the body of research on optimal statistical features of Bayesian estimates and concludes with a discussion of the implications of this large-sample property of the Bayesian median estimator for empirical measurement studies.

**KEYWORDS**
Almost sure convergence; Bayesian statistics; Bayesian confirmatory factor analysis; consistency; convergence in distribution; posterior median estimator; probability

Bayesian analysis and inference have been increasingly used over the past couple of decades in the behavioral, social, educational, medical, organizational and business sciences (e.g., Lynch, 2007). This statistical methodology offers researchers the opportunity to utilize available information in the form of prior distributions in the process of its integration with collected data, which has the potential of leading to posterior distributions permitting acquisition of deeper knowledge about studied phenomena. Bayesian confirmatory factor analysis (BCFA), as an essential part of Bayesian structural equation modeling (BSEM; e.g., B. Muthén & Asparouhov, 2012), has also enjoyed increased popularity in measurement applications in these and cognate disciplines, in part due to its feature of combining benefits from both factor analysis and Bayesian statistics utilizations.

As one key argument in favor of using Bayesian analysis, the large-sample behavior of resulting parameter estimators is oftentimes referred to in the literature (e.g., Gelman et al., 2013). This estimator behavior responds to the theoretically and empirically relevant question for behavioral measurement studies of what happens with the estimates resulting from a given parameter estimator as sample size increases without bound. A highly desirable feature of an estimator would be the diminishing distance with growing sample size between the estimates rendered by it and the true population value of a parameter estimated by them. Such a behavior, when the case, would correspond entirely to the intuitive notion of estimator quality as reflected in the fact that with larger and larger samples, the resulting parameter estimates become closer and closer to the pertinent true parameter. There appears to be also a widely circulated expectation among empirical researchers involved in measurement applications that parameter

**CONTACT** Tenko Raykov raykov@msu.edu Measurement and Quantitative Methods, Michigan State University, 443A Erickson Hall, East Lansing, MI 48824, USA

estimators furnished by contemporary statistical methods indeed possess this optimality property and thus show this type of convergence on the true parameter value, such as population latent correlations of interest, as sample size grows without limit.

The large-sample behavior of popular parameter estimators, like those provided by maximum likelihood (ML) and Bayesian statistics, has been the concern also of a large body of methodological and applied literature. Most of the time, however, what is meant in pertinent discussions when referring to the behavior of ML or Bayesian estimators with increasing sample size, is their consistency as well as their weak convergence, or convergence in distribution (e.g., Gelman et al., 2013). In particular, there is an impressive volume of research that deals with the large-sample behavior of ML estimators (as reviewed by Ferguson, 1996) as well as of Bayesian estimators (as reviewed, e.g., by DasGupta, 2008). However, the emphasis is placed in it on consistency and weak convergence, in part due to the fact that the latter is instrumental for obtaining respective estimator standard errors. These are two relevant asymptotic properties of Bayesian estimators, especially those resulting within the confirmatory factor analysis (CFA) and more generally the structural equation modeling (SEM) frameworks. At the same time, a frequent but generally incorrect implication from either of these convergence types seems to be the interpretation by some empirical behavioral scholars of a Bayesian point estimator approaching closely and converging numerically on the true population value of a parameter under consideration, when sample size increases without bound.

The present article is concerned with a stronger large-sample property of Bayesian point estimators than their consistency and weak convergence that are typically discussed and referred to in the literature accessible to measurement scientists. Specifically, the remaining discussion is focused on the Bayesian posterior median estimator that is often used in empirical measurement research. (This estimator is the default option in some popular Bayesian statistics and latent variable modeling software, such as for instance, Mplus; e.g., L. K. Muthén & Muthén, 2021). To this end, in what follows, we discuss and illustrate the strong (or with probability 1) convergence feature of this Bayesian median estimator as sample size grows indefinitely. (The feature is alternatively and interchangeably also referred to as strong consistency, almost sure convergence, convergence with probability 1, or convergence almost everywhere; e.g., Ferguson, 1996). This large-sample property implies (a) consistency as well as (b) convergence in distribution of the Bayesian median estimator, i.e., both (a) and (b) also hold when convergence with probability 1 does. However, the reverse is not true, i.e., (a) and/or (b) do not guarantee an estimator's almost sure convergence with increasing sample size, and hence do not imply the latter; that is, it can be shown that an estimator which is consistent and/or converges in distribution need not converge almost surely (e.g., Arnold, 1990; see also below, Raykov & Marcoulides, 2018, regarding necessary and sufficient conditions in important measurement settings). At the same time though, it is strong convergence that is in our experience what many empirical scientists tend to assume – from physics and chemistry through psychology, sociology and education – when referring to either consistency or convergence in distribution more generally of point estimators for considered parameters.

For these reasons, being focused on this strong convergence feature, the intent of the present article is to contribute to solidifying that intuitive interpretation of the large-sample behavior of a popular Bayesian parameter estimator, such as the Bayesian posterior median, which interpretation appears to be common among behavioral measurement scholars. In addition, the paper aims at adding to the body of literature on desirable optimality properties of parameter estimators within the Bayesian statistical framework, and in particular in BCFA that is of special relevance for measurement research when conducted within the Bayesian statistics framework.

The plan of this paper is as follows. We commence with a short discussion of basic principles of BCFA that are relevant for achieving its aims. We next indicate briefly key large-sample properties of sequences of random variables, such as model parameter estimators with increasing sample size, and point out their empirical importance as well as logical relationships and differences. To this end, we also offer a concise didactic discussion of important large-sample features of parameter estimators of relevance to behavioral and social measurement researchers. Subsequently, we focus on the almost sure convergence feature of the Bayesian posterior median estimator. We then illustrate with a numerical example the relevance and utility of this large-sample property in empirical behavioral measurement. In a conclusion section, we discuss the implications of the strong consistency feature of the Bayesian median estimator for measurement related research in the behavioral, social, educational, medical and business sciences.

## Bayesian confirmatory factor analysis: A short discussion of basics

Bayesian CFA (BCFA; e.g., Lee, 2007) may be seen as an extension of 'classic' CFA that assumes parameters in a model under consideration to be unknown constants and treats collected data as individual realizations of random variables. Unlike this conventional or traditional CFA model, BCFA treats all parameters as random variables following some distribution and the data as given (cf. Bickel & Doksum, 2015). To be more concrete, for a given set of $k$ manifest measures of interest ($k > 1$), consider first a classic CFA model developed within the general linear structural relationships (LISREL) framework (e.g., Jöreskog & Sörbom, 1996):

$$\mathrm{y} = \mu + \Lambda\, \mathrm{f} + \mathrm{e}. \tag{1}$$

In Equation (1), y denotes the $k$ x 1 vector of observed variables, f is a $q$ x 1 vector of latent factors with zero means (in the single population and assessment case; $0 < q < k$), $\Lambda$ is a $k$ x $q$ factor loading matrix, $\mu$ is the $k$ x 1 vector of observed means, and e is the $k$ x 1 vector of residuals with zero means that are assumed uncorrelated with the factors. The vector $\pi$ of parameters of this model consists of variances and covariances of independent variables (common and unique factors) as well as mean structure parameters, including factor loadings (e.g., Lee, 2007). Any CFA model considered in the rest of this article is also assumed to be identified, i.e., with additional assumptions or restrictions if needed to achieve identification; furthermore, any covariance matrix of relevance or resulting when using such a model is stipulated as positive definite (e.g., Bollen, 1989).

As indicated above, in BCFA each component of the parameter vector $\pi$ is viewed as a random variable following some distribution. In this paper, we assume that the latter is absolutely continuous, as is common in Bayesian statistics applications (e.g., Gelman et al., 2013). For analyzed data on the observed variables, which data are denoted $y$ in the remainder, the behavior of an individual model parameter $\theta$ given $y$ is described by the conditional distribution $p(\theta \mid y)$, referred to as posterior distribution of $\theta$ given the data (posterior probability density function, or posterior pdf for short). In Bayesian statistics, the relationship of key relevance is the well-known Bayes rule (e.g., Lee, 2007):

$$p(\theta \mid y) = p(\theta)\, p(y \mid \theta)/p(y), \tag{2}$$

where $p(\theta)$ denotes the prior distribution (prior pdf) reflecting knowledge available before the analysis, data collection and modeling effort; $p(y)$ is the pdf of the data; and $p(y \mid \theta)$ symbolizes the pertinent likelihood (a straight-forward multidimensional extension of Equation (2) holds in case of parameter vectors with dimensionality 2 or larger; e.g., Gelman et al., 2013). Two main types of prior distributions used in Bayesian statistics, and in particular in BCFA, are informative and non-informative priors (e.g., Lynch, 2007). For the latter priors, any interval on the real line within their support is associated with the same probability as any other interval with the same length; for the former priors, this is not the case and some intervals are associated with higher probabilities than others.

The following discussion in this article is independent of the type of prior chosen in an application of BCFA, whether informative or not (as long as the posterior distributions are proper and the assumptions of Proposition 1 below are fulfilled). With this in mind, we will be concerned with the large-sample behavior or the Bayesian point estimator defined as the median of the posterior distribution for a parameter of interest. This estimator, referred to for convenience as Bayesian median estimator, (a) has a number of attractive properties, (b) is used in applications of BCFA, and (c) as mentioned is default option in popular software (such as Mplus; L. K. Muthén & Muthén, 2021). In particular, the large-sample feature of almost sure convergence for this estimator to its counterpart population parameter value (true parameter value), which property is of focal interest in the present article, holds as we will see regardless of the choice of prior type (under the assumptions of Proposition 1).

## Large-sample behavior of parameter estimators: A brief visit

The behavior of parameter estimators as sample size increases without bound is the concern of the asymptotic distribution theory (e.g., DasGupta, 2008). Since each parameter estimator can be treated as a random variable with an individual (empirical study) realization being its value in a given sample, typically referred to as parameter estimate, this theory deals with the behavior of sequences of random variables – such as the sequence represented by the estimator when sample size grows indefinitely (e.g., Billingsley, 2011). In general, for a parameter estimator under investigation, the infinite sequence resulting when considering it a function of increasing sample size may be characterized by a type of behavior, under certain conditions, which is relevant for theoretical and empirical research utilizing it. In particular, it may well be the case then that this sequence approaches, in a certain sense, a random variable (including a constant as a special case) with rather important and beneficial empirical properties.

In applications of CFA, for instance, when using the popular maximum likelihood (ML) estimation method under normality, it is known that the resulting parameter estimator is converging in probability and in distribution to the true population parameter value with increasing sample size (for the latter convergence, once the estimator is properly normed; e.g., Bollen, 1989). These are two particular types of large-sample behavior of the resulting estimator, or asymptotic convergence types. In actual fact, under fairly general conditions the general ML estimator possesses even the property of almost sure convergence to the corresponding parameter population value (e.g., DasGupta, 2008, ch. 16; see also Ferguson, 1997, ch. 17).[1] Hence, when sample size grows indefinitely the ML estimates of any parameter of a 'classic' CFA model converge, as a numerical sequence – that is, in a calculus sense – to the true population parameter value apart from an event with 0 probability (under these general conditions and when ML is appropriately used, e.g., under normality). As an additional example of particular relevance for behavioral measurement, with normality of scale components, (a) the popular coefficient alpha estimator converges almost surely to the population scale reliability coefficient in empirical settings satisfying certain conditions; and (b) when the latter are not fulfilled, the widely used omega reliability estimator does so (for unidimensional scales with uncorrelated errors in both cases mentioned; see Raykov, 2019a, 2019b, and references therein regarding these conditions). To accomplish the goals of this article, it is fitting to explicate next the specifics of these estimator behavior types with increasing sample size, as well as their logical relationships and differences.

### *Three types of large-sample convergence of parameter estimators and their relationships*

In this subsection, we provide a concise informal discussion of three convergence types for sequences of random variables that are of particular relevance for both empirical and theoretical behavioral measurement research. As indicated above, the sequence obtained with a given estimator of a parameter of interest when sample size grows unboundedly, represents an example of such a sequence of random variables (cf., e.g., Raykov,2019a). These convergence types are of substantial importance for behavioral measurement settings, since they indicate three different kinds of large-sample behaviors exhibited by parameter estimators of special interest for measurement researchers, and thus by their resulting parameter estimates in a given empirical setup with increasing sample size. We refer to Ferguson (1996) (or Raykov, 2019a) for formal details on the following three convergence types, in particular with regard to their definitions, and provide next a conceptual discussion of them.

Perhaps the most popular type of estimator convergence with increasing sample size is that of convergence in distribution. One widely used example of such convergence is given by the central limit theorem (e.g., Raykov & Marcoulides, 2012), which states that the random sampling distribution of the mean approaches under fairly mild conditions a standard normal distribution (when the sample mean is appropriately normed). This is the weakest distribution type of all three discussed in this section, since the remaining two do not follow from convergence in distribution but rather the converse is true. Another popular estimator convergence type is that in probability, often referred to as consistency. Accordingly, as sample size increases so does also the probability of the population parameter value being very close to its estimator. Due to theoretical results demonstrated elsewhere (e.g., DasGupta, 2008), neither convergence in distribution nor consistency guarantee that as sample size increases the resulting sample parameter estimates, as real numbers, become closer and closer to that population value. The latter feature holds with probability 1 only if the estimator sequence converges almost surely to that true value. This is a very strong statement and the reason why this convergence is also called strong consistency. In actual fact, it is the strongest of all three convergence types mentioned in this section, and thus implies both convergence in distribution and consistency. This strong convergence, or convergence with probability 1, is therefore the subject of the remainder of this article, specifically in the context of Bayesian statistics and its applications in confirmatory factor analysis that is a key modeling approach in behavioral and social measurement research.

## *Why is strong convergence important to establish for confirmatory factor analysis parameter estimators?*

In applications of CFA, whether 'classic' CFA or BCFA, it is of special importance to know what happens "practically always" with the estimates of the parameters of an entertained model as sample size grows unboundedly. In particular, knowing that with the addition of data from a studied population these estimates numerically converge "practically always" on their true population counterparts (i.e., in a calculus sense as a sequence of real numbers) is highly beneficial both theoretically and empirically. This convergence does not follow, however, from (a) convergence in probability (i.e., from estimator consistency) or from (b) convergence in distribution as indicated above (e.g., Arnold, 1990) and hence needs to be ascertained in its own right.

In more concrete terms, for a parameter $\theta$ within a given CFA model or more generally a parametric statistical model, consider an estimator $\hat{\theta}$ of interest and denote by $\hat{\theta}_n$ its resulting estimate at sample size $n$, as well as its population value by $\theta_0$. (The discussion in this and next section applies regardless whether one is concerned with the framework of 'classic' CFA or BCFA; the population or true parameter value, $\theta_0$, is assumed existing as an unknown constant throughout this article; cf. Arnold, 1990; DasGupta, 2008). The large-sample properties of consistency and convergence in distribution of the estimator $\hat{\theta}$ do not actually entail anything specific about the behavior of these numbers $\hat{\theta}_n$ and their resulting numerical sequence as sample size increases (i.e., in a calculus sense; e.g., Apostol, 2006). That is, either of these latter two convergence types or in combination does not preclude $\hat{\theta}_n$, and their sequence, from behaving in an 'erratic' way. Specifically, convergence in distribution asserts only what happens with the sampling distribution of the estimator in question, $\hat{\theta}$, at any given sample size, as the latter increases. At the same time, the real numbers $\hat{\theta}_n$ themselves need not converge numerically to any real value, or if they do converge then their limit (i.e., that real value) need not equal the population parameter value $\theta_0$. Thereby, the difference of $\hat{\theta}_n$ from $\theta_0$ need not converge to 0, as a numerical sequence, of to any other real number, whether convergence in distribution of consistency is true for this estimator.

Quite on the contrary, the almost sure convergence of $\hat{\theta}$ with increasing sample size, which is shown in Proposition 1 below for the general Bayesian median estimator (whether within the BCFA framework or outside of it), excludes such an 'erratic' property of the limit of the associated sample estimates $\hat{\theta}_n$, as a numerical sequence, with increasing sample size (cf. Raykov, 2019a). This is achieved by 'locking' these estimates, as real numbers, to converge with probability 1 as a numerical sequence (i.e., apart from a practically ignorable event) to the population parameter value $\theta_0$ (Arnold, 1990). We refer to Ferguson (1996) and DasGupta (2008) for further and more detailed discussion of the specifics of all these convergence type relationships. Last but not least, consistency and convergence in distribution of CFA parameter estimators are not infrequently interpreted incorrectly by some empirical researchers to imply the convergence of the individual parameter estimate values of $\hat{\theta}_n$, as a numerical sequence (i.e., in a calculus sense; e.g., Apostol ,2006), to the true population value $\theta_0$ when sample size increases indefinitely. This intuitive interpretation actually represents an over-interpretation of consistency and/or weak convergence, and based on the preceding discussion (see also, e.g., Ferguson, 1996), it is justified only if strong convergence of the pertinent parameter estimator is itself shown to hold. The important reason is the fact that strong convergence does not follow logically from consistency and/or convergence in distribution, i.e., the fact that the latter two types of convergence are only necessary for but do not guarantee, singly or in tandem, strong consistency.

In the next section, we discuss more concretely strong convergence as a large-sample property of the Bayesian posterior median estimator within the framework of BCFA.

## Strong convergence of the Bayesian posterior median estimator of confirmatory factor analysis parameters

This section is concerned with the key result underlying the article, which is presented in Proposition 1 below. To state it formally, we need to introduce at this point some additional notation and assumptions (cf. DasGupta, 2008). To this end, suppose one was interested in a given CFA model in a behavioral measurement context (e.g., Raykov & Marcoulides, 2011), and specifically in its parameter $\theta$ that belongs to a parameter space $\Theta$ and is associated with an (assumed) prior distribution $p(\theta)$. In the remainder of the article, we presume that we are dealing with the empirical setting of independent and identically distributed observations (when treated as random variables prior to collecting the data $y$), which follow a distribution with a density, or pdf, $p(y \mid \theta)$. In addition, we posit the regularity conditions assuring the strong convergence of the ML estimator that were indicated above and stated in Footnote 1, as well as the following assumptions (often referred to as the assumptions of the Bernstein-von Mises theorem; e.g., Bickel & Doksum, 2015; see also DasGupta, 2008, ch. 21):

(a) The parameter space $\Theta$ is an open set of real numbers;
(b) For each $\theta$ from $\Theta$, $\int p''(y \mid \theta)\, dy = 0$, where double prime denotes second derivative with respect to $\theta$ and the integral is taken over the support of the function $p(y \mid \theta)$ (e.g., the entire real line if applicable);
(c) The information matrix $I(\theta)$ (i.e., the Fisher information function) is positive and finite for each $\theta$ from $\Theta$;
(d) The prior distribution $p(\theta)$ is continuous in some non-trivial symmetric open interval around the population value $\theta_0$ and is positive at it (i.e., $p(\theta_0) > 0$, or in other words, the true population value belongs to the support of the prior pdf);
(e) There exists a real-valued function $s(y)$ with the properties that $|p''(y \mid \theta)| < s(y)$ for each $\theta$ in the open interval mentioned in point (d) above, and its expectation with respect to the data distribution, based on the population value, is finite (i.e., $E_\theta\,[s(y)] < \infty$ for $\theta = \theta_0$).

We note that for instance, a non-informative prior for a given parameter, such as say $p(\theta) = 1$ or if need be $p(\theta) = c$ where $c$ is a sufficiently small constant, fulfills the pertinent of assumptions (a) through (e). Similarly, we point out that Proposition 1 is valid for any Bayesian median estimator, as long as its assumptions hold, which estimator may or may not arise or be used within the framework of CFA.

Under all these assumptions, the following statement is true.

Proposition 1: The Bayesian posterior median estimator converges almost surely to its population counterpart value, $\theta_0$, as sample size grows indefinitely.

This large-sample property of the Bayesian posterior median estimator, whether in a BCFA setting or outside of it, is demonstrated in Appendix A. This generally valid proposition (under its assumptions) is illustrated in the next section using a numerical example for the special case of a latent correlation estimator in a CFA context. The relevance of Proposition 1 and its utility for applications of BCFA in empirical measurement research across the behavioral, social, educational, medical, organizational, and business sciences is discussed subsequently.

## Illustration on data

We exemplify here the general Proposition 1 by using as a special case a CFA setting that is of particular relevance in behavioral measurement. To this end, we utilize a sequence of simulated data sets for successive sample sizes of $n$ = 500, 1000, 2000, 5000, 10000, 20000, 50000, 100000, 200000, 500000, 1000000, 1500000 and 200000 independent cases, which were generated employing Equations (3) below for $r = 6$ observed variables $y_1, \ldots, y_6$, following a two-factor model as defined in the above Equation (1) (e.g., Mulaik, 2009). Specifically, at each of these 13 sample sizes, multi-normal data were simulated according to the following model:

$$
\begin{aligned}
y_1 &= \mu_1 + \lambda_1\, f_1 + e_1,\\
y_2 &= \mu_2 + \lambda_2\, f_1 + e_2,\\
y_3 &= \mu_3 + \lambda_3\, f_1 + e_3,\\
y_4 &= \mu_4 + \lambda_4\, f_2 + e_4,\\
y_5 &= \mu_5 + \lambda_5\, f_2 + e_5,\\
y_6 &= \mu_6 + \lambda_6\, f_2 + e_6,
\end{aligned}
\tag{3}
$$

where $f_1$ and $f_2$ were standard normal variates, like the independent error terms $e_1, \ldots, e_6$, with correlation $\rho = Corr(f_1, f_2) = .3$ (where $Corr(.)$ denotes correlation), while $\mu_1 = \mu_2 = \ldots = \mu_6 = 0$ and $\lambda_1 = \lambda_4 = 1$, $\lambda_2 = \lambda_5 = 1.5$, $\lambda_3 = \lambda_6 = 2$ were set. (The Mplus command file used for data generation, including the seed utilized thereby, is provided in Appendix B and referred to there as Code 1. Code 2 in that appendix is the Mplus command file employed to fit with BCFA the two-factor model of relevance to the so-generated data sets; see below.)

**Table 1.** Bayesian confirmatory factor analysis estimates of a latent factor correlation, $\rho$, at increasing sample sizes in used example.

| Sample size | $\rho$ | Correlation estimates ($\hat{\rho}_n$) |
|---|---|---|
| 500 | .300 | .302 |
| 1,000 | .300 | .296 |
| 2,000 | .300 | .307 |
| 5,000 | .300 | .292 |
| 10,000 | .300 | .299 |
| 20,000 | .300 | .297 |
| 50,000 | .300 | .301 |
| 100,000 | .300 | .303 |
| 200,000 | .300 | .304 |
| 500,000 | .300 | .301 |
| 1,000,000 | .300 | .300 |
| 1,500,000 | .300 | .300 |
| 2,000,000 | .300 | .300 |

The population latent correlation coefficient is $\rho$ = .300 (see Equations (3) and immediately following discussion).

The resulting latent correlation estimates for the corresponding simulated data sets, when the above two-factor model was fitted to each of them with (default) non-informative priors for all model parameters, are presented in Table 1. We similarly note that as directly verified each of the assumptions of Proposition 1 are fulfilled in this example setting, also with respect to these priors (L. K. Muthén & Muthén, 2021, ch. 11).[2]

As seen from Table 1, in this example at sample size $n$ = 1000000 the latent correlation estimates converge numerically on its true population value of .300 and remain at it thereafter (for the above stated sample sizes).[3] These findings (a) are consistent with Proposition 1 that as stated earlier holds generally (under its assumptions; see above and Appendix A), regardless of whether it is used within the BCFA framework or outside of the latter, and (b) illustrate its validity in the special case represented by this example.[4]

## Conclusion

This article discussed the almost sure (or with probability 1) convergence, or strong consistency, of the general Bayesian posterior median estimator and in particular that resulting within the framework of Bayesian CFA (e.g., Lee, 2007) that is of increasing relevance in behavioral measurement studies. Proposition 1, under its conditions explicitly mentioned above, showed this convergence for each model parameter to its counterpart true population value. This strong consistency property, which is demonstrated in Appendix A for the general case, is valid for any Bayesian median estimator, regardless of context within which Bayesian statistics is utilized (i.e., in BCFA as well as outside of BCFA), as long as the proposition's assumptions are correct. The Bayesian median estimator convergence with probability 1 is a stronger statement than (a) its convergence in distribution and (b) its convergence in probability, and in fact implies both (a) and (b). The goal of the article was to solidify what seems to be a popular interpretation among some measurement scholars of (a) and (b) for Bayesian parameter estimators, in particular in BCFA, as implying numerical convergence of the pertinent parameter estimates to the true parameter population value with increasing sample size. As elaborated in a preceding section, while this interpretation appears frequent among empirical scientists, it does not follow from (a) and/or (b). Proposition 1 accomplished that goal, under the conditions explicated in the article (see the pertinent discussion in the previous sections). Under those circumstances, therefore, applied researchers can indeed assume – e.g., based on Proposition 1 – that any Bayesian median estimator produces sample-based parameter estimates, in particular (but not only) when using BCFA in behavioral measurement settings, which with the increase of study size (sample size) converge as resulting numerical sequences practically always to their corresponding parameter population values.

The discussion in this paper is associated with several limitations that need to be pointed out. First, the article is not concerned with speed of convergence per se. In particular, Proposition 1 does not have implications that would indicate sample size at which the Bayesian estimator of concern will yield an estimate that is sufficiently close, for pre-specified empirical purposes, to the parameter population value or become identical to it. We conjecture that the sample size at which a parameter estimator, in particular within BCFA, is close 'enough' for practically treating it identical to the corresponding true population value, depends on multiple factors. These will likely include number of observed variables, number of latent variables, reliability of individual manifest measures, the parameter in question, related characteristics of the used model, prior(s) used (while satisfying the conditions of Proposition 1 indicated earlier), and subject-matter specifics, to mention a few possible factors. In addition, all else being the same, certain priors may require higher (or smaller) sample sizes to achieve such estimate proximity to the corresponding parameter population value. Given the complexity of the issue, we encourage future studies on speed of convergence, including corresponding simulation studies evaluating the effects of these and other possible factors, in particular the type of prior and its specifics, which go beyond the confines of this article.

Second, Proposition 1 is only concerned with the Bayesian posterior median estimator and may or may not hold for the Bayesian posterior mean or mode estimators. Additional research, which lies beyond the frame of this paper, is encouraged that addresses the latter estimators' large-sample behavior (as well as possible conditions under which it holds). It should be also emphasized that this article carries no implications about relative speed of convergence of the posterior median, mean or mode estimators, and in fact in some settings it may be either of the latter two that converge faster to the corresponding parameter population value, or achieves practically satisfactory proximity to it, than that median estimator. Future research is therefore recommended that examines also this relative convergence speed issue.

Third, Proposition 1 does not imply that once reaching a particular (pre-specified) 'minimal' proximity to the population value, the parameter estimates after the associated sample size will all remain within that vicinity (or stay essentially identical to the population parameter value) for any larger sample size. This is because the proposition is only of asymptotic nature that precludes more concrete finite-sample interpretations of this kind. In actual fact, the statement of the proposition is consistent with the parameter estimate 'stepping out' of, and then 'returning back' into, a pre-specified vicinity (even of zero mass or probability on the real axis) of the population parameter value a finite number of times with increasing sample size.

Fourth, Proposition 1 is not to be interpreted as ensuring that in every empirical study the Bayesian posterior median estimates will converge as a numerical sequence on the true parameter population value. This is because its statement is only about almost sure convergence and thus holds apart from an event with probability 0. In some empirical cases, therefore, it is not impossible that this zero-probability event occurs, which itself may be a reason to look at (some of) them as potentially being associated with a discovery of interest per se in the substantive domain of application. Last but not least, there is a set of regularity conditions that were stated earlier, which only in their simultaneous validity guarantee that the statement of the proposition holds. Hence none of them alone, nor any (incomplete) combination of them, ensures the Bayesian median estimator's strong convergence or contains sufficient information that would allow a scholar to determine a minimal sample size beyond which the resulting Bayesian parameter estimates, in particular BCFA estimates, will fall within a pre-specified neighborhood of the true population value of a parameter in a model of interest.

In conclusion, this article discussed the almost sure convergence of the general Bayesian posterior median estimator, and especially the Bayesian median estimator in Bayesian confirmatory factor analysis that is a widely utilized methodology of high relevance in behavioral and social measurement, as well as aimed in addition to disseminate among measurement scholars this theoretically and empirically important result. The paper in particular justified the interpretation of this estimator's large-sample behavior as reaching practically always the true parameter population value (under the conditions of Proposition 1), which behavior tends to be assumed by empirical scientists but is not implied per se by estimator consistency and convergence in distribution known to hold for Bayesian estimators (under respective conditions; e.g., Gelman et al., 2013).

## Notes

1. These fairly general conditions include the requirement of the data distribution belonging to the exponential family, the true parameter value lying in the interior of the parameter space $\Theta$, and the information matrix being positive definite (and with finite elements) for each value in that interior. The conditions are satisfied for instance, by the normal distribution (of relevance in the illustration section). The one-parameter exponential family includes in addition to the normal also the binomial, Bernoulli, Poisson and exponential distributions, which together cover at least the majority of empirical settings of relevance for instance in psychology and the behavioral and social sciences currently. The proof that these conditions guarantee the strong consistence of the ML estimator, is provided for instance, in DasGupta (2008, pp. 241-242; see also proof of Theorem 16.1; Ferguson, 1996).

2. All 13 simulated data sets were associated with no error messages during the pertinent data generation process. For all 13 BCFA models fitted to them thereafter (as two-factor models with 3 indicators per factor), the potential scale reduction index went below 1.1 early in the iteration process and did not return above 1.1 for the remainder of the requested Bayesian iterations (see Code 2 in Appendix A; L. K. Muthén & Muthén, 2021).
3. Since Proposition 1 is generally valid (under its conditions that hold in this section, as mentioned in the main text), the choice of rounding-off to third decimal digit in this example and Table 1 is not essential for the illustration aims of the special case used in the present section. The reason is that strong convergence of the Bayesian posterior median estimates stated by Proposition 1, implies their stabilization within (or being covered by) any pre-specified neighborhood of the population latent correlation value, here of .300, at a sufficiently large-sample size and above it (e.g., Apostol, 2006; hence, inclusion of higher sample sizes in Table 1 is not necessary).
4. We stress that the purpose of this example is merely to illustrate the generally valid Proposition 1 by way of a concrete special case of it (as defined by Equations (3) and their surrounding discussion in the illustration section), rather than demonstrate this proposition's validity; the demonstration of the latter is given in Appendix A for the general case (under the conditions pointed out previously in this paper and stated in Proposition 1).

## Acknowledgments

We are grateful to A. DasGupta, S. Penev, T. Asparouhov, and D. J. Hand for valuable discussions on asymptotic theory, Bayesian statistics, and their software implementation. Thanks are also due to the Editor and several anonymous Referees for their critical comments on an earlier version of the paper, which have contributed substantially to its improvement.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## References

Apostol, T. (2006). *Calculus*. Wiley.
Arnold, S. F. (1990). *Mathematical statistics*. Prentice-Hall.
Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics* (Vol. 1). Taylor & Francis.
Billingsley, P. (2011). *Probability and measure*. Wiley.
Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.
DasGupta, A. (2008). *Asymptotic theory of statistics and probability*. Springer.
Ferguson, T. S. (1996). *A course in large sample theory*. Chapman & Hall.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (Third ed.). Taylor & Francis.
Jöreskog, K. G., & Sörbom, D. (1996). *LISREL user's guide*. Scientific Software International.
Lee, S.-Y. (2007). *Structural equation modeling. A Bayesian approach*. Wiley.
Lynch, S. M. (2007). *An introduction to applied Bayesian statistics and estimation for social scientists*. Springer.
Mulaik, S. A. (2009). *Factor analysis*. Chapman-Hall.
Muthén, B., & Asparouhov, T. (2012). Bayesian SEM: A more flexible representation of substantive theory. *Psychological Methods*, *17*(3), 313–335. https://doi.org/10.1037/a0026802
Muthén, B., Muthén, L., & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Muthén & Muthén.
Muthén, L. K., & Muthén, B. O. (2021). *Mplus user's guide*. Muthén & Muthén.
Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.
Raykov, T., & Marcoulides, G. A. (2012). *Basic statistics: An introduction with R*. Rowman & Littlefield.
Raykov, T., & Marcoulides, G. A. (2018). *A course in item response theory and modeling with Stata*. Stata Press.
Raykov, T. (2019a). Strong convergence of the coefficient alpha estimator for reliability of multiple-component measuring instruments. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(3), 430–436. https://doi.org/10.1080/10705511.2018.1515019
Raykov, T. (2019b). Strong consistency of reliability estimators for multiple-component measuring instruments. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 750–756. https://doi.org/10.1080/10705511.2018.1559737

# Appendix A

## Demonstration of Proposition 1

To accomplish the aims of this Appendix, we will need the following two lemmas.

Lemma 1: If $\{a_n\}$ $(n \geq 1)$ is a sequence of real numbers, such that

$$\sqrt{n}\,|a_n| \to 0 \text{ as } n \to \infty, \tag{A.1}$$

where $|.|$ denotes absolute value and "$\to$" convergence (of real numbers; e.g., Apostol, 2006), then also $|a_n| \to 0$ as $n \to \infty$ (and $a_n \to 0$ as $n \to \infty$) holds.

Indeed, for any pre-specified $\varepsilon > 0$, from (A.1) follows that there exists an integer $N$ such that for any integer $n > N$ also $|\sqrt{n}\ |a_n| - 0| = |\sqrt{n}\ |a_n|| = \sqrt{n}\ |a_n| < \varepsilon$ holds. The last inequality means that for such an $n$ also

$$|a_n| = |a_n - 0| < \varepsilon/\sqrt{n} < \varepsilon \tag{A.2}$$

is true. From the last pair of inequalities and the definition of limit of a sequence of real numbers (e.g., Apostol, 2006), $|a_n| \to 0$ as $n \to \infty$ follows, and similarly $a_n \to 0$ as $n \to \infty$ is implied.

Corollary: If the convergence in (A.1) above is almost sure to begin with, where $\{a_n\}$ $(n \geq 1)$ is a sequence of real-valued random variables that are defined on a common probability space, then $\to_{\text{a.s.}} 0$ and $|a_n| \to_{\text{a.s.}} 0$ as $n \to \infty$ is true, with "$\to_{\text{a.s}}$" denoting almost sure convergence.

Lemma 2: Suppose $\{A_n\}$ $(n \geq 1)$ and $\{B_n\}$ $(n \geq 1)$ are 2 sequences of real-valued random variables defined on a common probability space, with the property that each of them almost surely converges to the real numbers $A$ and $B$, respectively. Then their sum, $A_n + B_n$, also almost surely converges to $A + B$.

This statement follows directly from the definition of almost sure convergence (see introduction section) and the additive property of limits of sequences of real numbers (e.g., Apostol, 2006).

We will also make use below of the following key result that has been proved elsewhere (see citations given next).

Theorem 1: (DasGupta, 2008, p. 302; Bickel & Doksum, 2015, p. 343): Suppose that given the parameter $\theta$ the random variables $X_1, X_2, \ldots, X_n$ are independently and identically distributed random variables following the probability measure $P_\theta$, and that the conditions of the Bernstein-von Mises theorem hold (e.g., pp. 291–293 in DasGupta, 2008; these are the same conditions (a) through (e) in the main text of this article). Let $\theta^*_{\text{B,m},n}$ denote the Bayesian posterior median estimator and $\theta^*_{\text{ML},n}$ symbolize the maximum likelihood (ML) estimator, both at sample size $n$. Then

$$\sqrt{n}\ |\ \theta^*_{\text{B,m},n} - \theta^*_{\text{ML},n}\ | \to_{\text{a.s.}} 0 \tag{A.3}$$

for any true value $\theta_0$ of $\theta$ in $\Theta$.

For the proof of Theorem 1, see DasGupta (2008, pp. 302–303) and Bickel and Doksum (2015, pp. 343–345).

With the preceding in this Appendix in mind, under the conditions of Theorem 1 (which, again, are the same as (a) through (e) in the main text,) and the conditions for almost sure convergence of the ML estimator mentioned in the main text (e.g., Footnote 1), the validity of Proposition 1 now follows, i.e.,

$$\theta^*_{\text{B,m},n} \to_{\text{a.s.}} \theta_0, \tag{A.4}$$

for any true value $\theta_0$ of $\theta$. (The almost sure convergence in (A.4) is with respect to the relevant probability measure $P_\theta$ for $\theta = \theta_0$, and for any prior fulfilling the conditions mentioned above.)

Indeed, first let us observe that

$$|\ \theta^*_{\text{B,m},n} - \theta_0\ | = |\ (\theta^*_{\text{B,m},n} - \theta^*_{\text{ML},n}) + (\theta^*_{\text{ML},n} - \theta_0)\ | \leq |\ \theta^*_{\text{B,m},n} - \theta^*_{\text{ML},n}\ | + |\theta^*_{\text{ML},n} - \theta_0\ |, \tag{A.5}$$

where the respective property of absolute value is used for the sum of two or more real numbers (e.g., Apostol, 2006). Next, due to (i) Theorem 1 in this Appendix, and specifically Equation (A.3), and (ii) Lemma 1 above (as well as the definition of almost sure convergence; see earlier section), it follows that

$$|\ \theta^*_{\text{B,m},n} - \theta^*_{\text{ML},n}\ | \to_{\text{a.s.}} 0. \tag{A.6}$$

Then, due to Theorem 16.1 in DasGupta (2008), i.e., the almost sure convergence of the ML estimator to the true value (under the pertinent conditions, as found there and in Footnote 1, as well as pointed out above in this Appendix), also

$$| \theta^*_{\mathrm{ML},n} - \theta_0 | \to_{\mathrm{a.s.}} 0 \tag{A.7}$$

holds. Finally, due to Lemma 2, from Equations (A.5) through (A.7) follows that

$$| \theta^*_{\mathrm{B,m},n} - \theta_0 | \to_{\mathrm{a.s.}} 0,$$

which is equivalent to the Bayesian posterior median estimator almost sure convergence to the true parameter population value.

In short, the general Bayesian posterior median estimator, and in particular the Bayesian posterior median estimator within the framework of Bayesian CFA, always converges with increasing sample size and regardless of the used prior (as long as the conditions of Proposition 1 hold) to the true population parameter value, apart from an event with 0 probability. In other words, in any empirical study in an empirical science, apart from such studies that in total have 0 probability (and hence each of them having zero probability), the Bayesian posterior median estimator converges to the true parameter value, i.e., the resulting parameter estimates converge as a numeric sentence (that is, in a calculus sense) to the true parameter (population) value as sample size increases without bound.

## Appendix B

### Mplus source code for data simulation and model fitting used in the illustration section

```
TITLE:CODE 1. COMMAND FILE FOR SIMULATION OF THE 13 DATA SETS USED IN THE
ILLUSTRATION SECTION.
MONTECARLO:
    NAMES = Y1-Y6;
    NOBSERVATIONS = 500; ! STATE HERE SAMPLE SIZE.
    NREPS = 1;
    SEED = 3112487;
    SAVE = BA_SC_500.DAT; ! STATE HERE NAME OF GENERATED RAW DATA FILE.
MODEL POPULATION:
    [Y1-Y6@0];
    Y1-Y6@1;
    F1 BY Y1*1 Y2*1.5 Y3*2;
    F2 BY Y4*1 Y5*1.5 Y6*2;
    F1 WITH F2*.3;
    F1-F2@1;
    [F1-F2@0];
MODEL:
    [Y1-Y6@0];
    Y1-Y6@1;
    F1 BY Y1*1 Y2*1.5 Y3*2;
    F2 BY Y4*1 Y5*1.5 Y6*2;
    F1 WITH F2*.3;
    F1-F2@1; [F1-F2@0];
OUTPUT: TECH1 TECH9;
```

*Note*. Exclamation mark is used to include annotating comments (see L. K. Muthén & Muthén, 2021, ch. 11, for additional explanations of input commands, subcommands, and keywords utilized).

```
TITLE:     CODE 2. COMMAND FILE FOR FITTING A TWO-FACTOR MODEL VIA BCFA.
           (SEE Table 1 FOR RESULTING CORRELATION ESTIMATES WITH INCREASING
           SAMPLE SIZE.)
DATA:      FILE = BA_SC_500.DAT; ! STATE NAME OF USED RAW DATA FILE.
VARIABLE: NAMES = Y1-Y6;
ANALYSIS: ESTIMATOR = BAYES;
```

```
          PROCESSORS = 2;
          CHAINS = 2;
          FBITERATIONS = 20000;
MODEL:    F1 BY Y1* Y2-Y3; ! CFA MODEL FOR LATENT CORRELATION ESTIMATION.
          F2 BY Y4* Y5-Y6;
          F1-F2@1;
OUTPUT:   TECH1 TECH8;
```

For each of the 13 simulated data sets it is recommendable to fit subsequently the same model, as described in this code, using the latter with several times as many requested iterations (cf., e.g., L. K. Muthén & Muthén, 2021).
*Note*. See Note to Code 1 regarding input commands, subcommands, and keywords utilized, as well as Muthén et al. (2016, ch. 9). (The above source code indicates its use with the simulated data at sample size $n$ = 500.)