Revue d'Economie et de Statistique Appliquée
Volume 20 number 02, April 2024
ISSN : 1112-234X EISSN : 2600-6642
Pages : 07-21

ON THE ASYMPTOTIC NORMALITY OF THE TWO-WAY INTRACLASS CORRELATION COEFFICIENT OF AGREEMENT

Abderrahmane Bourredjem[1, 2]
[1]*Inserm CIC1432, Clinical Epidemiology unit, Dijon, France ; Dijon-Bourgogne University Hospital, Centre d'investigation clinique, Module Epidémiologie Clinique/Essais cliniques, Dijon, France.,*
[2]*Institut de Mathématiques de Bourgogne, UMR 5584, CNRS, Université de Bourgogne, Dijon, France.*
abderrahmane.bourredjem@u-bourgogne.fr

Nadjia El Saadi[3]
[3]*LAMOPS, ENSSEA, Tipasa, Algeria.*

Reçu le: 11/09/2023 **Accepté le :**15/01/2024 **Publication en ligne le**: 18/04/2024

RESUME :
Le coefficient de corrélation intra-classe de concordance à deux voies (ICCa) est un indice de la reproductibilité inter-évaluateurs d'une mesure quantitative. Recommandé par les agences réglementaires (tel que la FDA et l'EMEA) pour la validation des nouvelles mesures quantitatives, il est le seul paramètre permettant la généralisation des résultats de la fiabilité à l'ensemble de la population des évaluateurs. Nous démontrons un théorème central limite (TCL) pour l'ICCa valable à la fois pour l'estimateur d'analyse de variance (ANOVA), du maximum de vraisemblance (ML) ou le ML restreinte (REML). Un intervalle de confiance asymptotique (IC) est ensuite dérivé et ses performances sont examinées par simulation par rapport à la méthode de Fleiss et Shrout (F&S) qui est la plus utilisée en pratique pour l'IC de l'ICCa. Avec des échantillons de taille grande à modérée, nous avons montré que notre méthode est plus performante que la méthode usuelle de F&S en termes de taux de recouvrement et de largeur de l'IC. Enfin, un exemple illustratif est donné à partir des données réelles d'une étude de la fiabilité inter-évaluateurs sur le handicap visuel avec 101 patients et 7 évaluateurs.

Mots clés : Coefficient de corrélation intra-classe à deux voies de concordance (ICCa), normalité asymptotique, simulations.

ABSTRACT:
The two-way intra-class correlation coefficient of agreement (ICCa) is an inter-raterreliability index recommended by the regulatory agencies for quantitative measuresvalidation and is the only parameter allowing the reliability generalization to the whole raters population. We prove a central limit theorem (CLT) for the ICCa validboth with analysis of variance (ANOVA), maximum likelihood (ML), or restrictedML (REML) estimates. An asymptotic confidence interval (CI) is then derived andits performances were examined by simulation compared to the Fleiss and Shrout method (F&S) which is the most used method in the literature. With moderate to large sample sizes, we showed that our method outperforms the historical F&S onein terms of CI recovery rate and width. An illustrative example is given based on thereal data of an inter-rater reliability study in vision disability assessment with 101 patients and 7 raters.

Key words: Two-way intra-class correlation coefficient of agreement (ICCa), asymptotic normality, simulations.

I INTRODUCTION

Intraclass correlation is a widely used concept to assess inter-rater reliability (when several raters perform a single measurementon a group of subjects). A low reliability may indicate that the raters are not well trained or that the variable to be measured is notwell defined or standardized. Hence, the reliability issue is of great importance in many fields. We consider a random sample of n subjects for which a continuous variable Y is measured independently by k raters randomly selected from a population of raters.Denote by Y_{ij} the measurement made on the i^{th} subject by the jth rater (for $i = 1 \dots n$ and $j = 1 \dots k$). Let us assume the model:

$$Y_{ij} = \mu + a_i + b_j + e_{ij} (1)$$

where μ is a constant,

a_i is the subject random effect and $a_1 \dots a_n$ are assumed independent and identically distributed (iid) with a mean of zero and a variance A,

b_j is the rater random effect and $b_1 \dots b_k$ are assumed iid with a mean of zero and a variance B,

e_{ij} is a residual component. These residuals (e_{ij}) are assumed iid and with a mean of zero and a variance E. These last three sets (a_i), (b_j), (e_{ij}) of random effects are assumed to be mutually independent.

Under these hypotheses, the two-way intra-class correlation coefficient of agreement (abreviated in this article by ICCa) isdefined as the correlation between two independent raters (j and j'; $j \neq j'$), on the same subject (i):

$$\rho := ICCa = Corr(Y_{ij}, Y_{ij'}) = \frac{A}{A+B+E}$$

The ANOVA estimator of ICCa is denoted by ICC (2,1) in [1]. It is well known from the principles of experimental design that in the absence of replication (i.e., y_{ijh} $with\ h = 1$), the potential interaction term between subjects and raters $(ab)_{ij}$ is not estimable. Moreover, even if raters are expected to interact with subjects, in the case of inter-rater reliability there is no need to quantify the variation that is due to the subject-rater interaction since the interaction term can be blended into the error term without affecting the ICCa estimation. Furthermore, even if repeated measurements are collected for the sake of evaluating intra-rater reliability, the subject-rater interaction may be deemed unjustified by a suited design of the study (raters evaluate subjects in a random order...). McGraw and Wong (1996) [2] denote the case when interaction is absent by Case 2-A and abreviate the ICCa ANOVA estimate by ICC(A,1) (A for agreement and 1 because it is the reliability of a single rate). Several estimation methods exist for the ICCa, nevertheless, to our knowledge, no central limit theorem (CLT) was explicitly proposed concerning its inference.

II ICCA ESTIMATION METHODS

A Analysis of variance (ANOVA)

In this part, we review some facts regarding ANOVA estimates. This allows us to introduce concepts that will be useful throughout this article. Ideas are borrowed from Chapters 7 and 9 in Searle(2017) [3]. Using the well-known notations: $\overline{Y}_{..} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} Y_{ij}$, $\overline{Y}_{i.} = \frac{1}{k}\sum_{j=1}^{k} Y_{ij}$ and $\overline{Y}_{.j} = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}$ the total sum of squares (TSS) can be decomposed as follows:

$$TSS = \sum_{i=1}^{n}\sum_{j=1}^{k}(Y_{ij}-\bar{Y}_{..})^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}(Y_{ij}-\bar{Y}_{i.})^2 + k\sum_{i=1}^{n}(\bar{Y}_{i.}-\bar{Y}_{..})^2 = WSS + SSA$$

$$= n\sum_{j=1}^{k}(\bar{Y}_{.j}-\bar{Y}_{..})^2 + \sum_{i=1}^{n}\sum_{j=1}^{k}(Y_{ij}-\bar{Y}_{i.}-\bar{Y}_{.j}+\bar{Y}_{..})^2 + SSA = SSB + SSE + SSA$$

Indeed, by introducing $\overline{Y}_{i.}$ in TSS, the crossed products term is equal to zero. Then, by introducing $\overline{Y}_{.j} - \overline{Y}_{..}$ in WSS, the crossed products term is equal to zero.

Then, the expected sums of squares are:

$$\mathbb{E}(SSA) = (n-1)(kA+E)$$
$$\mathbb{E}(SSB) = (k-1)(nB+E)$$
$$\mathbb{E}(SSE) = (n-1)(k-1)E$$

It follows for the means of squares, defined by:

$$BMS = SSA/(n-1) \quad \text{(the between-subjects mean square)}$$
$$RMS = SSB/(k-1) \quad \text{(the between-raters mean-square)}$$
$$EMS = SSE/[(n-1)(k-1)] \quad \text{(the error mean-square)}$$

That their expectations are:

$$\mathbb{E}(BMS) = kA + E$$
$$\mathbb{E}(RMS) = nB + E$$
$$\mathbb{E}(EMS) = E$$

By the method of order one moments, solving the equality between expected and em pirical mean of squares, ANOVA estimators of variance components are defined by:

$$\hat{E}_{n,k} = EMS$$
$$\hat{B}_{n,k} = \frac{1}{n}(RMS - EMS)$$
$$\hat{A}_{n,k} = \frac{1}{k}(BMS - EMS)$$

These ANOVA estimators are minimum variance unbiased (MVU) and consistent estimators[3], without the data normality hypothesis.

By plugging in the ICCa formula (2), the ANOVA estimator of the ICCa under Case2-A (as well as in the Case2 withoutinteraction, see McGraw and Wong (1996)[2]) is given by:

$$\widehat{ICCa} = \frac{\hat{A}_{n,k}}{\hat{A}_{n,k} + \hat{B}_{n,k} + \hat{E}_{n,k}} = \frac{BMS - EMS}{BMS + (k-1)EMS + \frac{k}{n}(RMS - EMS)}$$

It has been shown that this ICCa estimate is consistent but biased[1].

B Maximum Likelihood (ML)

Maximum likelihood (ML) estimates of μ and the variance components A, B and E can be obtained by maximizing the log- likelihood function. Under the assumption of the data normality, the likelihood of μ, A, B and E is defined by the joint density function of $y' = (Y_{11}, \dots, Y_{1k}; Y_{21}, \dots, Y_{2k}; \dots; Y_{n1}, \dots Y_{nk})$, given by[4]:

$$L(\mu, A, B, E; y') = f(y'; \mu, A, B, E)$$
$$= \frac{\exp\left[-\frac{1}{2}\left\{\frac{SSE}{E} + \frac{SSB}{nB+E} + \frac{SSA}{kA+E} + \frac{nk(\bar{y}_{..}-\mu)^2}{E+nB+kA}\right\}\right]}{\sqrt{(2\pi)^{nk}(E)^{(n-1)(k-1)}(E+nB)^{(k-1)}(E+kA)^{(n-1)}(E+nB+kA)}}$$

Thus, the log-likelihood function can be written as:

$$\ln(L) = -\frac{1}{2}\Big[(nk)\ln(2\pi) + (n-1)(k-1)\ln(E) + (k-1)\ln(Bn+E) + (n-1)\ln(Ak+E)$$
$$+ \ln(E+nB+kA) + \frac{SSE}{E} + \frac{SSB}{Bn+E} + \frac{SSA}{Ak+E} + \frac{nk(\bar{y}_{..}-\mu)^2}{E+nB+kA}\Big]$$

Equating to zero the partial derivative of $ln(L)$ with respect to μ yields $\hat{\mu} = \overline{Y}_{..}$. By injecting this solution in the derivatives of $ln(L)$ with respect to A, B and E, Sahai (1974)[4] has shown that it is not possible to obtain any closed form analytic expressionfor the ML estimators of the variance components. Thus, iterative optimization algorithms are required to solve the maximizinglikelihood equations.

C Restricted ML (REML)

The ML approach does not take into account the loss of degrees of freedom due to the fixed effects in estimating the variance components. REML remedies this by partitioning the likelihood function into two parts, with one part, free of the fixed effects, depending only on the variance components. Maximizing this part yields the REML variance components estimates.

For the ICCa model, the REML estimators of A, B and E can be obtained by maximizing that part of the likelihood functionwhich is location invariant. The restricted log-likelihood function can be written as[4] :

$$\ln(REL) = -\frac{1}{2}\left[(nk)\ln(2\pi) + (n-1)(k-1)\ln(E) + (k-1)\ln(Bn+E) + (n-1)\ln(Ak+E) + \frac{SSE}{E} + \frac{SSB}{Bn+E} + \frac{SSA}{Ak+E}\right]$$

By equating to zero the partial derivatives of $ln(REL)$ with respect to A, B and E under the constraint that the variancecomponents are non-negative, Sahai (1974)[4] derived the following formulas for the REML variance components estimates:

$$\widehat{E_{REML}} = \min\left(\frac{SSE}{(n-1)(k-1)}, \frac{SSE+SSB}{(n-1)(k-1)+k-1}, \frac{SSE+SSA}{(n-1)(k-1)+n-1}, \frac{SSE+SSB+SSA}{(n-1)(k-1)+k+n-2}\right),$$

$$\widehat{B_{REML}} = \frac{1}{n}\left\{\frac{SSB}{k-1} - min\left(\frac{SSE}{(n-1)(k-1)}, \frac{SSE+SSA}{(n-1)(k-1)+(n-1)}\right)\right\}^{+},$$

and

$$\widehat{A_{REML}} = \frac{1}{k}\left\{\frac{SSA}{n-1} - min\left(\frac{SSE}{(n-1)(k-1)}, \frac{SSE+SSB}{(n-1)(k-1)+(k-1)}\right)\right\}^{+}$$

where $n^{+} = \max(n, 0)$.

III ASYMPTOTIC JOINT NORMALITY FOR THE VARIANCE COMPONENT ESTIMATES

Amemiya (1971)[5] proposed to estimate μ by $\overline{Y}_{..}$ and derived a common asymptotic distribution for the ANOVA and ML variance component estimates under the data normality hypothesis, given as follows:

$$\begin{pmatrix} \sqrt{n}(\hat{A}_{n,k} - A) \\ \sqrt{k}(\hat{B}_{n,k} - B) \\ \sqrt{nk}(\hat{E}_{n,k} - E) \end{pmatrix} \overset{d}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, 2\begin{pmatrix} A^2 & 0 & 0 \\ 0 & B^2 & 0 \\ 0 & 0 & E^2 \end{pmatrix}\right) \text{as } n \to +\infty \text{ and } k \to +\infty \tag{4}$$

Using the Amemiya (1971) [5] approach, we prove in Appendix A that the same convergence holds for the REML variance components estimates.

We can clearly see that the effective sample size for this convergence is not the total sample size nk for all parameters, but n for A, k for B and nk for E.

In the rest of the article, to simplify the notations, we will use $\hat{A}$, $\hat{B}$, $\hat{E}$ instead of $\hat{A}_{n,k}$, $\hat{B}_{n,k}$, $\hat{E}_{n,k}$.

A ICCa central limit theorem (CLT)

Let $\rho = \text{ICCa} = \frac{A}{A+B+E}$ and $\hat{\rho} = \frac{\hat{A}}{\hat{A}+\hat{B}+\hat{E}}$ its ANOVA, ML or REM estimator.

We will prove the following CLT under the condition that $\frac{n}{k} \xrightarrow{n,k\to\infty} c$ and $0 < c < \infty$

$$\boxed{\sqrt{n}(\hat{\rho} - \rho) \overset{n,k\to\infty}{\rightsquigarrow} N(0, 2\rho^4[(\frac{1}{\rho} - 1)^2 + cu^2])\text{... ICCa CLT}}$$

Where $u = \frac{B}{A}$

Ideas and notations are borrowed from Chapters 2 and 3 in van der Vaart (1998) [6]. Note first that $\rho = f(A, B, E)$ where the nonlinear application f is differentiable at $\theta_0 = (A, B, E)$, with continuous partial derivatives. This means that for $h = (h_1, h_2, h_3) \in R^3$ tending to 0, we have

$$f(\theta_0 + h) - \rho = \langle \nabla f(\theta_0), h \rangle + o(\|h\|)$$

Where $\nabla f(\theta_0)$ is the gradient vector of f evaluated at θ_0 and $\langle .,.$ is the Euclidean inner product in R^3. Consider now $\hat{\theta} = (\hat{A}, \hat{B}, \hat{E})$ and the random deviation $\hat{h} = \hat{\theta} - \theta_0$. We have with Lemma 2.12 in [6] that

$$f(\hat{\theta}) - \rho = \langle \nabla f(\theta_0), \hat{\theta} - \theta_0 \rangle + o_p(\|\hat{\theta} - \theta_0\|)$$

and

$$a_{n,k}(f(\hat{\theta}) - \rho) - \langle \nabla f(\theta_0), a_{n,k}(\hat{\theta} - \theta_0) \rangle = o_p(a_{n,k}\|\hat{\theta} - \theta_0\|). \tag{5}$$

To be sure that $a_{n,k}(f(\hat{\theta}) - \rho)$ converges weakly to a non-degenerate distribution, we must ensure that $a_{n,k}|\hat{\theta} - \theta_0|$ is $O_p(1)$. All the norms being equivalent in R^3, we can consider the norm defined by $|\mathrm{h}| = \max(|\mathrm{h}_1|, |\mathrm{h}_2|, |\mathrm{h}_3|)$ and look for the component in $\hat{\mathrm{h}}$ with the highest rate of convergence towards zero. We deduce from (4) that $\sqrt{\mathrm{n}}(\widehat{\mathrm{E}} - \mathrm{E}) = \mathrm{o_p}(1)$ and $\sqrt{\mathrm{k}}(\widehat{\mathrm{E}} - \mathrm{E}) = \mathrm{o_p}(1)$ since $\sqrt{\mathrm{nk}}(\widehat{\mathrm{E}} - \mathrm{E}) \rightsquigarrow \mathcal{N}(0{,}2\mathrm{E}^2)$ and $\max(\sqrt{\mathrm{n}}, \sqrt{\mathrm{k}}) = \mathrm{o}(\sqrt{\mathrm{nk}})$ if both k and n tend to infinity. Consequently the term involving $\mathrm{a_{n,k}}(\widehat{\mathrm{E}} - \mathrm{E})$ in the differential $\langle \nabla \mathrm{f}(\theta_0), a_{n,k}(\hat{\theta} - \theta_0)\rangle$ will be necessary negligible compared to the two other terms to keep a rest term negligible in probability when multiplying by a dilatation factor $\mathrm{a_{n,k}}$ the linear part of the Taylor expansion as in (5).Then, we can distinguish three configurations according to the asymptotic behavior of k and n.

• k is negligible compared to n, meaning that n/k tends to infinity. In that case, the rest term tends in probability to zero and the linear part to a non-degenerate distribution if $\mathrm{a_{n,k}}$ and $\sqrt{\mathrm{k}}$ are asymptotically equivalent. We then get $\sqrt{\mathrm{k}}(\widehat{\mathrm{A}} - \mathrm{A})$ tends to zero in probability and

$$\sqrt{\mathrm{k}}(\mathrm{f}(\hat{\theta}) - \rho) \rightsquigarrow \mathrm{N}\left(0{,}2\left(\frac{\partial \mathrm{f}}{\partial \mathrm{B}}\right)^2 \mathrm{B}^2\right).$$

• n is negligible compared to k, meaning that n/k tends to zero. Then $a_{n,k}$ and $\sqrt{n}$ should be asymptotically equivalent. We have $\sqrt{n}(\hat{B} - B)$ tends to zero in probability and

$$\sqrt{n}(f(\hat{\theta}) - \rho) \rightsquigarrow N\left(0{,}2\left(\frac{\partial f}{\partial A}\right)^2 A^2\right).$$

• n and k tend to infinity at the same rate, that is to say $\lim n/c \to c > 0$.

In this case:

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow[nk\to\infty]{} N\left(0{,}2\rho^4\left[\left(\frac{1}{\rho} - 1\right)^2 + cu^2\right]\right) \text{ ... ICCa CLT}$$

where $u = \frac{B}{A}$.

In practice, for the asymptotic variance expression, c, ρ and u are unknown and should be replaced by $\hat{c} = \frac{n}{k}$, $\hat{\rho}$ and $\hat{u} = \frac{\hat{B}}{\hat{A}}$, respectively. Since $\hat{\rho}$, $\hat{A}$ and $\hat{B}$ are consistent and $\frac{n}{k} \xrightarrow{n,k\to\infty} c$ (with $0 < c < +\infty$), this can be obtained properly using the Slutsky theorem [6], leading to the following asymptotic pivotal function for the ICCa:

$$\sqrt{n}\frac{(\hat{\rho} - \rho)}{\sqrt{2\widehat{\rho^4}\left[\left(\frac{1}{\hat{\rho}} - 1\right)^2 + \frac{n}{k}\widehat{u^2}\right]}} \rightsquigarrow N(0{,}1) \text{ ... ICCa CLT reduced form}$$

B ICCa CLT confidence interval

From the last result, we can derive the following convergence:

$$P\left(\rho \in \left[\hat{\rho} \pm Z_{\alpha/2}\frac{\sigma_{\hat{\rho},\hat{u},\hat{c}}}{\sqrt{n}}\right]\right) \xrightarrow{n,k\to\infty} 1-\alpha$$

$$\text{where } \sigma_{\hat{\rho},\hat{u},\hat{c}} = \sqrt{2\hat{\rho}^4\left[\left(\frac{1}{\hat{\rho}}-1\right)^2+\frac{n}{k}\hat{u}^2\right]}$$

$Z_{\alpha/2}$ is the $1-\alpha/2$ order quantile of the Gaussian distribution and $\hat{\rho}$ and $\hat{u}$ are the estimated ρ and u.

Hence, an asymptotic $(1-\alpha)$ level, bilateral, confidence interval could be defined by:

$$\left[\hat{\rho} - Z_{\alpha/2}\frac{\sigma_{\hat{\rho},\hat{u},\hat{c}}}{\sqrt{n}}, \hat{\rho} + Z_{\alpha/2}\frac{\sigma_{\hat{\rho},\hat{u},\hat{c}}}{\sqrt{n}}\right] \text{... ICCa CLT CI}$$

IV FLEISS & SHROUT CONFIDENCE INTERVAL METHOD

Fleiss and Shrout (F&S) proposed a large sample CI based on a Satterthwaite's two-moment approximation to a linear combination of independent chi-square random variables[7]. They considered the following linear combination W of RMS and EMS:

$$W = \frac{nICCa}{k(1-ICCa)}RMS + \left\{1+\frac{n(k-1)ICCa}{k(1-ICCa)}\right\}EMS$$

and they determined a parameter ν such that the first two moments of $\nu W/E[W]$ coincide with those of a chi-square distribution with ν degrees of freedom. An upper (U_{FS}) and lower (L_{FS}) one-sided 100 (1 - α)% confidence limits were given as:

$$U_{FS} = \frac{F_*(BMS) - EMS}{F_*(BMS) + d_2(RMS) + d3(EMS)}$$

$$L_{FS} = \frac{BMS - F^*(EMS)}{BMS + F^*[+d_2(RMS) + d3(EMS)]}$$

where $d_2 = k/n,\ d_3 = k-1-k/n$

$F_* = F_{1-\alpha;\hat{\nu},n-1}$ is the upper 100 (1 - α) percentile of an F-distribution with degrees of freedom $\hat{\nu}$ in the numerator and $n-1$ in the denominator.

$F^* = F_{1-\alpha;n-1,\hat{\nu}}$ is the upper 100 (1 - α) percentile of an F-distribution with degrees of freedom $n-1$ in the numerator and $\hat{\nu}$ in the denominator.

ν is the Satterthwaite's approximate degrees of freedom given as:

$$\nu = \frac{(k-1)(n-1)(k * ICCa * F_r + n(1+(k-1)ICCa) - k * ICCa)^2}{(n-1)k^2 * ICCa^2 * F_r^2 + (n(1+(k-1)ICCa) - k * ICCa)^2}$$

and $F_r = RMS/EMS$. As ICCa true value is unknowk, it will be replaced by $\widehat{ICCa}$ giving $\hat{\nu}$.

This method is implemented in the most popular statistical softwares:

1. R functions: psych::ICC(), Irr::icc() and irrICC::icc2.nointer.fn().

2. Stata command: icc.

3. SPSS module: Reliability analysis (two-Way random model, absolute agreement, single measure).

4. SAS: no official routine for the ICCa F&S CI, only the %IntraCC() macro for the ICCa ANOVA estimation.

V SIMULATION STUDY OF THE ICCA CLT CI

A simulation study has been conducted to evaluate the ICCa CLT CI recovery rate and mean width. Our CLT based method was compared to the F&S7. Five designs were studied: small sample size with n=30,k=5, moderate sample size with n=40,k=10 then n=60,k=10 and large sample size with n=115,k=15 then n=150,k=15. Indeed, in practice, the number of repetition k rarely exceeds 15 as the used devices are often costly and raters are usually practitioners who have limited time for research activities. Based on the Shrout (1998)8 reliability thresholds, ICCa values of interest were 0.55 for fair agreement, 0.65 then 0.75 for moderate agreement and 0.85 for substantial agreement. Without loss of generality, the total variance was set to 20=A+B+E so that values of A=11,13,15 and 17 yields ICCa=0.55,0.65,0.75 and 0.85 respectively. For each case, 1000 data sets are simulated under the normal distribution for (a_i), (b_j) and (e_ij). The simulations were conducted using the R software (version 4.1.0). The F&S method was implemented using the R psych::ICC() function (psych package version 2.1.6). The CLT-based method was implemented easily from the two-way crossed random effect model variance components ANOVA estimates using the anova(lm()) R base function. For each simulation, a 95% bilateral ICCa CI was derived using the three methods (F&S and CLT). Overall the 1000 simulations, the recovery rate (or coverage probability, estimated by the percentage of simulations when the CI contain the true ICCa value) and average interval width (upper bound - lower bound; mean over the 1000 simulations) were calculated.

Results are shown in Table1. The F&S method recovery rate is around 90% for all designs including when n=150,k=15 or n=k=50. Our CLT method performs better than the F&S in all cases, except in the small size design (n=30,k=5) where the two methods were close with a coverage rate about 90%. Our CLT method is not recommended for a small sample size (where $n \leq 30$ or $k \leq 5$). For the moderate sample size, the coverage rate of our CLT method is about 93%. The expected coverage of 95% is closely approximated for n=115,k=15 and achieved for n=150,k=15.

VI ILLUSTRATIVE EXAMPLE

We illustrate the use of the proposed ICCa CLT CI on the open data from the Baskaran et al. study (PLoS ONE, 2019) in sensory disabilities assessment9 In this inter-rater reliability study, seven raters assessed the reading performance of 101 patients with low vision, using the MNREAD test. One parameter of interest was the critical print size (CPS). The estimated ICCa reported for CPS was of 0.77 (95% CI [0.69,0.83]) but this estimation was poorer among the four less experienced raters (0.70; 95% CI [0.57,0.80]) when compared to the three experienced ones (0.82; 95% CI [0.76,0.88]). The confidence intervals were calculated using the F&S method.

The CLT 95% CIs that we computed are [0.703,0.827] for all raters, [0.609,0.794] for the less-experienced raters and [0.767,0.877] for experienced raters. This is very close to the F&S CI, except for the less-experienced raters where F&S interval is about 20% wider than the CLT in this case, which is not a minor difference. The F&S interval lower bound is shifted to the left, relative to the CLT interval, corresponding to lower estimates of the ICCa, more far more distant from that of the expert raters. Results are detailed in Table 2.

VII CONCLUSION

The current paper proves the asymptotic equivalence between the ANOVA, ML and REML estimation methods for the ICCa and gives an explicit common CLT for its inference. Our simulation results shows that the ICCa CLT CI has a better coverage than the popular method of F&S and is more accurate (provide a shorter CI). This can be explained by the fact that F&S involves two stage approximation7 . Moreover, our method is simpler to implement.

ACKNOWLEDGMENTS

We are grateful to Aurélie Calabrèse (Aix Marseille Université - Laboratoire de psychologie cognitive) for sharing their data and answering our questions about their analysis details.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

APPENDIX

A. PROOF OF THE ASYMPTOTIC JOINT NORMALITY FOR THE REML VARIANCECOMPONENTS ESTIMATES

Let us prove that the joint convergence of the REML variance components estimates is the same as for the ANOVA or ML cases (see section III). First of all, the asymptotic behavior of REML estimators has been studied in the general context of Gaussian linear models by [10]. Under the data normality assumption, the two-way random effects model is a particular case. For our model, the rank p of the fixed effects design matrix X is fixed ($p = 1$), so that, the REML estimates are asymptotically normal with zero mean (i.e. consistent) and a variance-covariance matrix equal to the inverse of the restricted information matrix. Specifically, remember that the restricted log-likelihood function (see section II.C) can be written as:

$$\ln(REL) = -\frac{1}{2}\left[(nk)\ln(2\pi) + (n-1)(k-1)\ln(E) + (k-1)\ln(Bn+E) + (n-1)\ln(Ak+E) + \frac{SSE}{E} + \frac{SSB}{Bn+E} + \frac{SSA}{Ak+E}\right]$$

To prove that the variance-covariance matrix is identical to the ANOVA and ML one, we proceed as in Amemiya (1971) [5], page 9. Using $Ln(REL)$ we have the following derivatives:

$$\frac{\partial \ln REL}{\partial A} = -\frac{1}{2}\left(\frac{(n-1)k}{Ak+E} - \frac{SSAk}{(Ak+E)^2}\right)$$

$$\frac{\partial \ln REL}{\partial B} = -\frac{1}{2}\left(\frac{(k-1)n}{Bn+E} - \frac{SSBn}{(Bn+E)^2}\right)$$

$$\frac{\partial \ln REL}{\partial E} = -\frac{1}{2}\left(\frac{(n-1)(k-1)}{E} + \frac{k-1}{Bn+E} + \frac{n-1}{Ak+E} - \frac{SSE}{E^2} - \frac{SSB}{(Bn+E)^2} - \frac{SSA}{(Ak+E)^2}\right)$$

For the second derivatives, we have:

$$\frac{\partial^2 \ln REL}{\partial A\, \partial B} = \frac{\partial^2 \ln REL}{\partial B\, \partial A} = 0$$

$$\frac{\partial^2 \ln REL}{\partial B^2} = -\frac{1}{2}\left(-\frac{(k-1)n^2}{(Bn+E)^2} + \frac{2SSBn^2}{(Bn+E)^3}\right)$$

$$\frac{\partial^2 \ln REL}{\partial A^2} = \frac{1}{2}\frac{(n-1)k^2}{(Ak+E)^2} - \frac{SSAk^2}{(Ak+E)^3}$$

$$\frac{\partial^2 \ln REL}{\partial E\, \partial A} = \frac{1}{2}\frac{(n-1)k}{(Ak+E)^2} - \frac{SSAk}{(Ak+E)^3}$$

$$\frac{\partial^2 \ln R\,EL}{\partial E\,\partial B} = \frac{1}{2}\frac{(k-1)n}{(Bn+E)^2} - \frac{SSBn}{(Bn+E)^3}$$

$$\frac{\partial^2 \ln R\,EL}{\partial E^2} = \frac{1}{2}\left(\frac{(n-1)(k-1)}{E^2} + \frac{k-1}{(Bn+E)^2} + \frac{n-1}{(Ak+E)^2} - \frac{2SSE}{E^3} - \frac{2SSB}{(Bn+E)^3} - \frac{2SSA}{(Ak+E)^3}\right)$$

Next, let us calculate the expectations of the second derivatives. As $E(SSA) = (n-1)(kA+E)$, $E(SSB) = (k-1)(nB+E)$ and $E(SSE) = (n-1)(k-1)E$ (see section II.A), then:

$$E\left(\frac{\partial^2 \ln R\,EL}{\partial A^2}\right) = -\frac{1}{2}\frac{(n-1)k^2}{(Ak+E)^2}$$

$$E\left(\frac{\partial^2 \ln R\,EL}{\partial B^2}\right) = -\frac{1}{2}\frac{(k-1)n^2}{(Bn+E)^2}$$

$$E\left(\frac{\partial^2 \ln R\,EL}{\partial E^2}\right) = -\frac{1}{2}\frac{(n-1)(k-1)}{E^2} - \frac{1}{2}\frac{k-1}{(Bn+E)^2} - \frac{1}{2}\frac{n-1}{(Ak+E)^2}$$

$$E\left(\frac{\partial^2 \ln R\,EL}{\partial E\,\partial A}\right) = -\frac{1}{2}\frac{(n-1)k}{(Ak+E)^2}$$

$$E\left(\frac{\partial^2 \ln R\,EL}{\partial E\,\partial B}\right) = -\frac{1}{2}\frac{(k-1)n}{(Bn+E)^2}$$

$$E\left(\frac{\partial^2 \ln R\,EL}{\partial B\,\partial A}\right) = 0$$

Then, the inverse asymptotic variance-covariance matrix is obtained by:

$$-\frac{1}{n}E\left(\frac{\partial^2 \ln R\,EL}{\partial A^2}\right) = \frac{1}{2}\frac{(n-1)k^2}{n(Ak+E)^2} \xrightarrow{n,k\to\infty} \frac{1}{2A^2}$$

$$-\frac{1}{k}E\left(\frac{\partial^2 \ln R\,EL}{\partial B^2}\right) = \frac{1}{2}\frac{(k-1)n^2}{k(Bn+E)^2} \xrightarrow{n,k\to\infty} \frac{1}{2B^2}$$

$$-\frac{1}{nk}E\left(\frac{\partial^2 \ln R\,EL}{\partial E^2}\right) = -\frac{-\frac{1}{2}\frac{(n-1)(k-1)}{E^2} - \frac{1}{2}\frac{k-1}{(Bn+E)^2} - \frac{1}{2}\frac{n-1}{(Ak+E)^2}}{nk} \xrightarrow{n,k\to\infty} \frac{1}{2E^2}$$

$$-\frac{1}{\sqrt{n\sqrt{nk}}}E\left(\frac{\partial^2 \ln R\,EL}{\partial E\,\partial A}\right) = \frac{1}{2}\frac{(n-1)k}{\sqrt{n\sqrt{nk}}(Ak+E)^2} \xrightarrow{n,k\to\infty} 0$$

$$-\frac{1}{\sqrt{k\sqrt{nk}}}E\left(\frac{\partial^2 \ln R\,EL}{\partial E\,\partial B}\right)=\frac{1}{2}\frac{(k-1)n}{\sqrt{k\sqrt{nk}}(Bn+E)^2}\xrightarrow{n,k\to\infty}0$$

Which leads to the same variance-covariance matrix as in the ANOVA or ML case.

ICCa	CI	n=30	k=5	n=40	k=10	n=60	k=10	n=115	k=15	n=150	k=15
	Method	Recovery %	Width	Recovery %	Width	Recovery %	Width	Recovery %	Width	Recovery %	Width
0.55	F&S	90.9	0.289	91.7	0.219	90.1	0.174	90.2	0.134	92.4	0.122
0.55	CLT	88.3	0.281	92.3	0.234	93	0.200	94.2	0.150	94.9	0.137
0.65	F&S	91.4	0.249	92.0	0.191	90.6	0.152	90.3	0.113	91.9	0.101
0.65	CLT	89.9	0.244	93.1	0.207	93	0.173	93.5	0.127	94.6	0.114
0.75	F&S	91.1	0.212	91.7	0.162	90.2	0.129	90.5	0.099	92.6	0.096
0.75	CLT	90.6	0.214	94.0	0.178	94.4	0.151	94.8	0.110	95.4	0.109
0.85	F&S	90.3	0.169	91.1	0.123	89.9	0.098	89.8	0.080	90.8	0.075
0.85	CLT	90.5	0.17	93.1	0.136	92.5	0.119	94.0	0.090	94.6	0.085

Table 1 Coverage probabilities and average widths for two-sided 95% ICCa confidence intervals.

	$\hat{ICCa}$	F&S 95% CI	width	CLT 95% CI	width
Overall (n=101, k=7)	0.77	[0.689 , 0.827]	0.138	[0.703 , 0.827]	0.124
Experienced raters (n=101, k=3)	0.82	[0.756 , 0.879]	0.123	[0.767 , 0.877]	0.11
Less experienced raters (n=101, k=4)	0.7	[0.571 , 0.796]	0.225	[0.609 , 0.794]	0.185

Table 2 CPS ICCa results for Baskaran et al. study (PLoS ONE, 2019)

References

[1] Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin 1979; 86(2): 420-428.

[2] McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1996;1(1): 30-46.

[3] Searle S, Gruber MH. Linear Models, second edition. Wiley: New York . 2017.

[4] Sahai H. Non-negative maximum likelihood and restricted maximum likelihood estimators of variance components in two simple linear models. Utilitas Math. 1974; 5: 151-160.

[5] Amemiya T. The Estimation of the Variances in a Variance-Components Model. International Economic Review 1971; 12:1-13.

[6] Vaart v. dAW. Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics Cambridge University Press, Cambridge. 1998

[7] Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass coefficient. Psychometrika 1978; 43: 259-262

[8] Shrout P. Measurement reliability and agreement in psychiatry. Stat Methods Med Res 1998; 7: 301-317.

[9] Baskaran K, Macedo AF, He Y, et al. Scoring reading parameters: An interrater reliability study using the MNREAD chart. *PLoS ONE* 2019; 14: 1-14.

[10] . Cressie N, Lahiri S. The Asymptotic Distribution of REML Estimators. Journal of Multivariate Analysis 1993; 45(2):217-233.