A Gaussian alternative to using improper confidence intervals

André PLANTE

Département de Mathématiques, Université du Québec à Montréal (retraité), Montréal, Canada H3C 3P8

Key words and phrases: Dempster’s conditioning; fiducial; Fraser’s frequency interpretation; noncentrality parameter; ratio of means; slope of the major axis.

MSC 2010: Primary 62A01; secondary 62G40.

Abstract: The problem posed by exact confidence intervals (CIs) which can be either all-inclusive or empty for a nonnegligible set of sample points is known to have no solution within CI theory. Confidence belts causing improper CIs can be modified by using margins of error from the renewed theory of errors initiated by J. W. Tukey—briefly described in the article—for which an extended Fraser’s frequency interpretation is given. This approach is consistent with Kolmogorov’s axiomatization of probability, in which a probability and an error measure obey the same axioms, although the connotation of the two words is different. An algorithm capable of producing a margin of error for any parameter derived from the five parameters of the bivariate normal distribution is provided. Margins of error correcting Fieller’s CIs for a ratio of means are obtained, as are margins of error replacing Jolicoeur’s CIs for the slope of the major axis. Margins of error using Dempster’s conditioning that can correct optimal, but improper, CIs for the noncentrality parameter of a noncentral chi-square distribution are also given. The Canadian Journal of Statistics 48: 773–801; 2020 © 2020 Statistical Society of Canada

Résumé: Le problème que posent les intervalles de confiance exacts qui peuvent être soit universels soit vides pour un ensemble appréciable de points d’échantillonnage est bien connu comme étant insoluble dans la théorie des intervalles de confiance. Les ceintures de confiance causant des intervalles de confiance impropre peuvent être modifiées en utilisant des marges d’erreur issues d’une théorie renouvelée des erreurs entreprise par J. W. Tukey—décrite brièvement dans le présent article—pour laquelle l’interprétation de Fraser en termes de fréquence est étendue. Dans cette approche, qui est en accord avec l’axiomatisation des probabilités selon Kolmogorov, une probabilité et une mesure d’erreur obéissent aux mêmes axiomes, bien que les connotations de ces deux mots soient différentes. Un algorithme est fourni, capable de produire une marge d’erreur pour tout paramètre dérivé des cinq paramètres de la distribution normale à deux variables. Des marges d’erreur corrigeant les intervalles de confiance de Fieller pour un rapport de moyennes sont obtenues, de même que des marges d’erreur remplaçant les intervalles de confiance de Jolicoeur pour la pente d’un axe principal. Des marges d’erreur utilisant les erreurs conditionnelles de Dempster sont aussi fournies pour corriger les intervalles de confiance optimaux mais improprest estimant le paramètre de non-centralité d’une distribution khi-carré non-centrale. La revue canadienne de statistique 48: 773–801; 2020 © 2020 Société statistique du Canada

Additional Supporting Information may be found in the online version of this article at the publisher’s website.

* Author to whom correspondence may be addressed.

E-mail: plante.andre.2@uqam.ca

© 2020 The Authors. The Canadian Journal of Statistics/La revue canadienne de statistique published by Wiley Periodicals LLC on behalf of Statistical Society of Canada. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
1. INTRODUCTION

Using exact confidence intervals (CIs) $C(y)$ to estimate a scalar parametric function $\psi(\theta)$ when some $C(y)$’s are equal to $(-\infty, \infty)$ over a nonnegligible set of sample points $y$—although their confidence coefficient $1 - \alpha$ is strictly between zero and one—has been studied by Bennett & Neyman (1948, answering a question by W. S. Cochran), Neyman (1954), Bennett (1957), Scheffé (1970), Plante (1979b, 1991), Koschat (1987) and Gleser & Hwang (1987). Scheffé named these intervals “improper confidence sets.”

Improper CIs cannot be used. And, since it is known that $P_{\theta}\{C(y) \ni \theta \mid C(y) \neq (-\infty, \infty)\} < 1 - \alpha$ for some $\theta$, the belt $C(\cdot)$ would have to be modified over a region larger than $\{y : C(y) = (-\infty, \infty)\}$. Koschat, as well as Gleser and Hwang, show that this is impossible in important cases such as the problem of estimating the ratio of two normal means. While Scheffé used a conservative CI approach for a ratio problem, most CI users do not.

CIs that are equal to $\emptyset$ over a nonnegligible set of sample points are also called “improper” in this article—because their complementary CIs are equal to $(-\infty, \infty)$—although their effects may appear less treacherous than those of all-inclusive CIs. Improper CIs are far more frequent than is usually realized and there can be no remedy for many of them within standard CI theory.

The method given here to correct improper CIs in troublesome cases employs margins of error (MEs) and is based on the renewed (Gaussian) theory of errors initiated by Tukey (1957, now declassified). Gauss (1809, 1821, 1823, 1855) based his theory on a physical concept—the distribution of all the possible fortuitous errors associated with an observation. His model for a single observation $y_0$ can be described as $y_0 - \theta = e$, where $y_0$ is a numerical observation, $\theta$ an unknown real parameter and $e$, using probability terminology, is an error variable. The associated equation $\theta = y_0 - e$ shows $\theta$ (which can never be observed directly) seen through an “estimate” $y_0$ and an error $e$—related to a distribution of propagated errors (DPE) centred on this “estimate.” When several observations are made, their associated errors are combined to give rise to a DPE on an estimate. Gauss’s error propagation formula—called the delta method by probabilists and statisticians (Doob, 1935)—is still extensively used in the physical sciences and in engineering.

Fisher (1953) confirmed that he was working within the Gaussian theory of errors and, in Fisher (1955), suggested to J. W. Tukey that the concept of error simulation—used in this article—could be used as a means to approximately describe DPEs. Tukey (1957) adapted Gauss’s terminology to the post-Kolmogorov era and extended Gauss’s error propagation formula to obtain approximate DPEs. Fraser’s frequency interpretation (Fraser, 1961; Cox & Hinkley, 1974, p. 248) shows how to observe a DPE when the pivot is linear or based on a group of transformations. This article extends Fraser’s frequency interpretation by providing a method to observe DPEs in cases where the error model is based on a nonlinear pivot $T\{y_0, \psi(\theta)\} = e$.

Observation errors, while considered the main cause of coverage errors (the usual $\alpha$’s), are not the only cause. When causes other than observation errors distort the parameter coverage process, it is often possible to use observation errors to obtain MEs as alternative intervals by means of DPE quantiles. J. Neyman at first thought that CIs and MEs were numerically very close (Neyman, 1935; Lehmann, 1994). This is often true, so that DPEs can almost seamlessly complement many confidence distributions (Cox, 1958; Fraser, 1979; Plante, 1991, 1994; Schweder & Hjort, 2002, 2016; Singh et al., 2005; Balch, 2012; Xie & Sing, 2013; Nadarajah, Bityukov & Krasnikov, 2015; Reid & Cox, 2015). However, confidence distributions cannot estimate parameters for which the only existing exact CIs are at times improper. In these cases, CIs and MEs will differ.

This renewed theory of errors belongs to the same extended family as do the fiducial method (Fisher, 1930, 1935, 1959); the structural distribution method (Fraser, 1961, 1968); the upper and lower probability method (Dempster, 1967, 2008; Shafer, 1976) and, more recently, the generalized fiducial method (Hannig, 2009, 2013; Hannig et al., 2016). It also shares a theoretical
neighbourhood with the confidence distribution method and with the generalized confidence method (Weerahandi, 1993, 2013; Bebu & Mathew, 2008).

The additive structured model for continuous cases (Plante, 1979a, 1979b), which uses both Fraser’s (1971) and Dempster’s methods, can be used within the renewed theory of errors, but at the price of introducing an extra assumption. In the additive structured model, the upper and lower probabilities (or upper and lower error measures) coincide. This puts some distance between the additive structured model and those primarily using discrete distributions in the Dempster and Shafer tradition—such as those used by Edlefsen, Liu & Dempster (2009), Martin et al. (2010), Martin & Liu (2013, 2014b, 2016).

The article is organized as follows. Section 2 explains the problem posed by improper CIs. Section 3 discusses how R. A. Fisher and J. W. Tukey contributed to the renewed theory of errors. An outline of the renewed theory of errors is given in Section 4 where it is also shown how one can use Dempster’s conditioning within a continuous model to repair improper CIs such as the uniformly most accurate CIs estimating the noncentrality parameter of a noncentral chi-square distribution. An extended Fraser’s frequency interpretation is defined in Section 5 which also explains the relationships between the renewed theory of errors, the improper Bayesian approach, and CI theory. Section 6 describes the error simulation algorithm used to estimate any (computable) parameter derived from the five parameters of the approach, and CI theory. Applications of this algorithm to estimate a ratio of means are given in Section 7 and the slope of a major axis in Section 8. Possible objections to the renewed theory of errors such as Stein’s fiducial example, marginalization paradoxes, the Buehler and Feddersen paradox and so on, are discussed in Section 9, which also points out that marginalization is both unavoidable and manageable. Section 10 is a general discussion. In addition to the Appendix, mostly containing proofs, there is available a Software Web Appendix in the Supplementary Material containing the prototype programs used (accompanied by a README file). A Web Appendix in the Supplementary Material is also available containing data, tables, additional elementary examples, bootstrap, auxiliary Monte Carlo experiments and connections with the improper Bayesian and CI approaches.

2. THE PROBLEMS POSED BY IMPROPER CONFIDENCE INTERVALS

2.1. How Improper CIs Arise

Improper confidence sets for a parametric function $\psi(\theta)$ usually occur when generated by an improper pivot $T\{y; \psi(\theta)\} = e$. The pivot is improper when the set $T\{y; \psi(\Theta)\}$ can have a probability less than one for a nonnegligible set of sample points, where $\Theta$ is the parameter space (Plante, 1979b, 1991). If $\psi(\theta)$ is a scalar parameter, this means that the $p$-value function $p(\psi) = F(T(y_0, \psi))$ (Fraser, 2019; Fraser, Reid & Wu, 1999; Schweder & Hjort, 2002; and Fraser, Reid & Wong, 2004) is not essentially onto the error space $(0, 1)$ (modulo a uniform distribution) for a nonnegligible set of sample points, where $F(\cdot)$ is the distribution function of $T\{y; \psi(\theta)\}$ and where $y_0$ is the observed value of $y$. Given a measurable space $(U, \mathcal{G})$ and a probability space $(V, \mathcal{H}, R)$, a measurable function $f : U \rightarrow V$ is essentially onto $V$ modulo $R$ if $f^{-1}(H) = f^{-1}(H')$ implies that $R(H \triangle H') = 0$ for any $H$ and $H'$ in $\mathcal{H}$, where $\triangle$ is the symmetric-difference operator, as is explained in Plante (1991, Section 4.1) and where $f^{-1}$ is the inverse image set function of the point function $f$. (Several $p$-value functions are illustrated in the Web Appendix in the Supplementary Material.)

Note that using a $p$-value function as a means to construct CIs is introduced in Corollary 3.5.1 in Lehmann & Romano (2005), i.e., Corollary 3 of Chapter 3 in Lehmann (1959). Also, if $y$ and $\psi$ are scalars, and if $T(y, \psi)$ is the (continuous) distribution functions of $y$, then the distribution function of $T(y, \psi)$ restricted to the interval $(0, 1)$ is an identity function. This is the formulation, described with a different notation, in Fisher (1930). Plante (1976) examined a
FiguRe 1: Stein’s “natural” CIs may be improper. Probability, when using Stein’s preferred CIs, that the one-sided 95% CIs of the form \([\delta^2, \infty)\) for a noncentrality parameter \(\delta^2 \geq 0\) be equal to \([0, \infty)\) as a function of \(\delta^2\). Dotted line: \(p = 10\); dashed line: \(p = 30\); dashed and dotted line: \(p = 50\); solid line: \(p = 80\).

conjecture by Fisher (1930) concerning the monotonicity of a \(p\)-value function associated with a maximum likelihood estimator.

Using “Natural CIs,” recommended by Stein (1959) as preferable to using fiducial intervals, provides a good example of how improper CIs arise. In his example a random variable \(d^2\) has a noncentral chi-square distribution with \(p\) degrees of freedom and noncentrality parameter \(\delta^2 \geq 0\), and is associated with a pivot \(F(d^2; \delta^2, p) = e\), where \(F(d^2; \delta^2, p)\) is the distribution function of \(d^2\), and where \(e\) has a uniform density on the interval \((0, 1)\). Since \(F(d^2; \delta^2, p) < F(d^2; 0, p)\) for every \(\delta^2 > 0\), if \(d^2 > 0\), it follows that for any fixed \(d^2 > 0\), the set \(\{d^2; [0, \infty), p\}\) has a length smaller than one. Consequently, the lower confidence bound \(\delta^2 = \delta^2(d^2, 1 - \alpha) = \inf\{\delta^2 : \delta^2 \geq 0\text{ and } F(d^2; \delta^2, p) \leq 1 - \alpha\}\) for \(\delta^2\) is equal to zero—and thus improper—with probability \(F(F^{-1}(1 - \alpha); 0, p; \delta^2, p) > 0\) Figure 1 illustrates this. Note that for a fixed \(p\), the natural CIs, are the uniformly most accurate CIs to estimate \(\delta^2\) because the family of noncentral chi-square distributions is strictly ordered according to the nondecreasing likelihood ratio and because of Corollary 3.5.1 in Lehmann & Romano (2005).

It is shown in the Appendix that this lower confidence bound \(\delta^2\), although uniformly most accurate, is improper with probability \(1 - \alpha\) at the limit when \(p \to \infty\), provided \(\delta^2 = o(p^{1/2})\), and that it fails—with probability one—to underestimate \(\delta^2\) for the remaining \(\alpha\) coverage probability. Note that this solution makes sense if \(\delta^2 = 0\). (A referee pointed out that the signal to noise ratio tends to zero if \(\delta^2 = o(p^{1/2})\).)

Gleser & Hwang (1987) showed how improper CIs can unavoidably occur due to a singularity in the parametric function \(\psi(\theta)\) being estimated. A confidence belt containing 100% improper CIs can even be uniformly most accurate—as occurs, for instance, when estimating a mixture parameter (Plante, 1991, where the mis-copied symbol “≤” on line 14 of p. 391 must be inverted). Restricted parameter problems also produce improper CIs (Quenouille, 1958, Sections 6.6 and 6.9; Mandelkern, 2002). Several authors have unsuccessfully attempted to produce improved CIs which would be proper at all levels. For instance, Casella (2002) suggested an attractive frequentist solution for a 68% upper confidence bound when \(y\) is from an \(N_1(\mu, 1)\) \((\mu \geq 0)\) distribution with a sample of size of one, but his method—which consists in gently modifying the model—still gives an empty one-sided interval at the 45% confidence level when the observation is \(y_0 = -1\). (The graph of a \(p\)-value function illustrates this statement in the Web Appendix in the Supplementary Material.)
Other improper CIs, besides the ones discussed here, exist. For example, the natural CIs to estimate the parameter in the general model of the Corollary to Theorem 1 below can be improper for the same reason that the natural CIs to estimate a noncentrality parameter and the CI for a restricted normal mean can be improper.

**Corollary 1 (Corollary to Theorem 1 in the Appendix).** Let $F(d; \delta)$ be the distribution function of a statistic $d$ taking its values in an interval $[d_0, \infty)$, where the parameter $\delta$ is in a closed interval $[\delta_0, \infty)$ ($\delta_0 > -\infty$) and where $d_0$ may be equal to $-\infty$. Suppose that $F(d; \delta)$ is continuous as a function of two variables and that the partial function $F(d; \cdot)$ is one-one for every $d$ in $(d_0, \infty)$. Then $F(d; \delta)$ is an improper pivot.

(The graphs for the various distribution functions cannot cross each other because the distribution functions are continuous and the partial function with fixed $d$ is one-one and, consequently, the family of distributions is stochastically ordered. Therefore, either none of the graphs can be above the graph of $F(\cdot; \delta_0)$ or none can be below it. In either case the family of graphs do not fill the strip $[d_0, \infty) \times (0, 1)$, which is a clear sign that the pivot is improper.)

### 2.2. The Problem with Fieller’s CIs for a Ratio of Means

Where correction is needed can be determined if one adopts the viewpoint of a user of Fieller’s method who has only his observed sample and a bivariate normality assumption. Let $\bar{x}, \bar{y}, s_x, s_y,$ and $r$ be the usual estimators of $\mu_x, \mu_y, \sigma_x, \sigma_y,$ and $\rho$. Suppose that $\bar{x} > 0$ and that $r \geq 0$. Imagine a modified mean sample point $(x', y') = (k\bar{x}, k\bar{y})$ ($k \neq 0$ an accessory variable), from possible $\mathcal{N}_2$ distributions, moving under the influence of $k$ on a given radius at an angle with the $x$ axis sufficiently less than $\pi/2$, towards a point at infinity and starting at a $0+$ distance from the origin of the plane. If only the mean sample point is varied, initially, Fieller’s 95% CIs are the interval $(-\infty, \infty)$; but they become the complement of an interval except if $k\bar{y}/k\bar{x} = (s_x/s_y)r$ and $k\bar{x} = (s_x/\sqrt{N})\nu_{d/2}$, according to the proof, in the Appendix, of Proposition 3 in Section 7. Next, they become two-sided intervals. As $k$ increases, CIs for a ratio are eventually satisfactory — when the denominator $k\bar{x}$ of the ratio $k\bar{y}/k\bar{x}$ is far from zero. However, between the limit point where CIs have an infinite length and points where CIs look quite regular, there is an interval along the given radius where the CIs are transitionally misplaced. This is essentially the region for which I can suggest a correction and can also correct CIs for some points where they have an infinite length. Fortunately, for a ratio of means, this treacherous region appears short.

In a ratio of means problem, the DPE becomes progressively flatter as the factor $k$ in $k\bar{y}/k\bar{x}$ decreases in absolute value — with the result that it becomes more and more difficult to calculate the 2.5% and 97.5% quantiles of the DPE in order to obtain a 95% ME. Thus, the number of significant figures determining a ME is important. Using a laptop computer, I found it reasonable to aim at results with $4 \pm 1$ significant figures. To simulate this required 10 million iterations. Using the DPE technique to obtain a generalized fiducial solution (Hannig, 2009; Hannig et al., 2016), i.e., approximating the coverage frequency for the entire parameter space is a technically different problem than that investigated here — which concerns a unique sample. A computing speed of greater magnitude is required to deal with sufficient points in a five dimensional space.

### 2.3. The Problem with Cox’s Revised CI Theory

After Neyman (1954) recognized that his theory of CIs needed improvement, Cox (1958) proposed a revised CI theory possibly inspired by Egon Pearson’s attitude towards the joint Neyman–Pearson theory (Lehmann, 1994). This revised theory was expanded in Cox & Hinkley (1974). It included, besides the basic concepts of alternative hypothesis and power, an emphasis on likelihood, sufficiency, minimal sufficiency, ancillary statistics, and maximum likelihood—all concepts borrowed from Fisher’s approach. Cox even included the idea of
a confidence distribution which kept some components of Fisher’s fiducial distribution. For example, he considers that if the parameter space $\Theta$ is of the form $\Theta_1 \times \Theta_2$, a confidence distribution for $\theta_1$ is a distribution $P^*_{y}(\cdot)$ over $\theta_1$-events such that $P^*_{y}(S_{1-a}(y)) = 1 - a$ for all exact 100$(1 - a)\%$ confidence sets $S_{1-a}(y)$ for $\theta_1$ in a large class and for all $a$ in the interval $(0, 1)$.

However, a corollary to Theorem 1 in Plante (1991) is that if a $\sigma$-field $\mathcal{F}$ of confidence belts is generated by a continuous pivot $T(y; \theta)$, then a confidence distribution exists over every $y$-section of $\mathcal{F}$ iff $T(y; \theta)$ is a proper pivot, i.e., iff the difficulty met by Neyman (1954) does not occur. Furthermore, confidence sets are always determined by a pivot except in esoteric cases irrelevant to statistics (Plante, 1991, Lemma 1). Therefore, Cox’s confidence distribution theory does not solve the problem raised by improper CIs.

That a confidence distribution must be induced by a proper pivot is bad news for confidence distribution users, for it implies that parameters like $100\sigma/|\mu|$, $\mu + \sigma^2/2$, and so on cannot be estimated using a natural confidence distribution based on the normal model for observations—as was shown by Plante (1991). Furthermore, being restricted to proper pivot models also implies that models that have more than one error term, as do standard variance component models, are not estimable using a natural confidence distribution. All variance component problems share the difficulty illustrated (using Stein’s natural CIs) in Section 2.1 when underestimating a noncentrality parameter. (The case of a noncentrality parameter is solved using MEs in Section 4.2.) An ad hoc conditional fiducial solution was proposed in Quenouille (1958, Section 6.9) for variance components and a systematic solution is available using an "error" or formal third space given in Cisewski & Hannig (2012) where Dempster’s conditioning (defined in Section 4.2) is possible.

2.4. A Fundamental Problem with Neyman’s Coverage Frequency Interpretation

In the physical sciences, observation errors are quantified so that estimates from one investigation can be used in other investigations and that the precision of the final results be appraised. Errors are partially due to observer’s perceptions and manipulations. Neyman added a subjective potential error-causing component—the decision which is supposed to be based on these scientific observations.

Wald (1939) criticized both the theory of the testing of hypotheses and CI theory, pointing out that they cannot always provide a solution since these theories are at their base founded on the concept of a discrete action-space {accept, reject} rather than on the concept of a continuous space—where randomization ensures that every problem has a solution as is the case in the theory of games. Wald’s approach is entirely based on a system of weights. Bartholomew (1967, 1971)—claiming to avoid known paradoxes—described several numerical applications of such a modified Neyman–Pearson theory. Plante (1987) proposed a solution to the problem of choosing between two simple hypotheses consisting of a Bayesian posterior-like everywhere-randomized procedure in which prior probabilities are replaced by losses for taking a wrong decision, as factors of the likelihood function.

3. TOWARDS A RENEWED THEORY OF ERRORS

3.1. The Unavoidable Persistence of Improper CIs

When commenting on improper CIs occurring in the restricted $\mu$ problem ($y$ has a $\mathcal{N}_1(\mu, 1)$ distribution, $\mu \geq 0$), Gleser (2002) suggested that Neyman–Pearson’s method should be restricted to the design stage of an investigation and that Fisher’s fiducial and likelihood methods should be used to analyze data.

My position, based upon counter-example logic explained in Plante (1971), is that although coverage frequency remains an important concept, the strong repeated sampling principle (Cox & Hinkley, 1974, p. 45) can no longer be an axiom used as a general validity criterion because
one counterexample is sufficient to prove a general principle wrong. Since, in the case of a ratio of means, CI theory needs to be improved (as Neyman, 1954, admitted); and, since Koschat (1987) proved, using a simplified version of the ratio of means problem, that Fieller’s exact CI solution is the only reasonably continuous one (since observation errors are the main cause of coverage errors and since observation errors are included within Gauss’s theory of errors), I suggest using observation errors in order to provide minimal, but necessary, corrections to confidence belts containing improper CIs.

3.2. From Fisher’s Fiducial Probability to Tukey’s Propagated Error Method

Fisher (1912), while a graduate student specializing in the theory of errors, deviated from Gaussian theory when introducing his concepts of likelihood function and the maximum likelihood estimation method (both not then yet named). These concepts do not conform exactly with Gauss’s concept of observation error density, nor with his maximum error density estimation method. However, Fisher’s fiducial distribution (Fisher, 1930, 1935, 1959) can be considered a major step forward within the theory of errors. Yates (1960, Chapter 7) adapted the theory of errors to sampling techniques.

Tukey (1957) warned military engineers and scientists about the technical difficulties hidden within CI theory and advised them to use Gauss’s error propagation formula instead or to use the Gaussian formula extensions which his article was providing which were also capable of yielding approximate DPEs. He demonstrated how his approach was consistent with Kolmogorov’s axiomatization of probability (Kolmogorov, 1933, 1956, p. 1) in which a probability measure and an error measure obey the same axioms, although the connotations of the two words are different. It seems evident that errors (which concern past events) and probabilities (which concern future events) should be used differently as the past and the future are not symmetrical.

Using error variables which can be simulated (Fisher, 1955) can unify various uses of formal, auxiliary, or prediction variables (often denoted $U^*$, $U'$ or $Z^*$) as used by Beran (1971), Plante (1979a, 1979b, 1994), Tsui & Weerahandi (1989), Weerahandi (1993, 2013), Hannig (2009), Hannig et al. (2016), Martin et al. (2010), Martin & Liu (2013, 2016) and accepted in several other approaches to inference centred on an estimate rather than on a parameter.

4. OUTLINE OF A RENEWED THEORY OF ERRORS

4.1. Three Spaces and a Pivot

The theory of errors founded by the astronomer Gauss is based on a physics postulate very different from the mathematical repeated sampling axiom of Neyman–Pearson theory. Gauss’s (1809, 1855, Note I, Section 1, par. 1) error postulate can be stated as follows:

**Postulate 1 (Gauss’s error postulate).** When apart from systematic errors, fortuitous observation errors are the only reason (corrected) observations and the quantities measured do not coincide, fortuitous observation errors are propagated according to the laws of probability.

The renewed theory of errors initiated by Tukey (1957) and continued below uses three spaces before the sample is drawn; a sample space, $\mathcal{Y}$, a pivot space $\mathcal{E}$, and a parameter space $\Theta_p$. They are assumed to be Borel subsets of Euclidean spaces. The symbol $\rho$ designates the dimension of the Euclidean space for $\theta$. There is a known family of absolutely continuous probability distributions $\{P_\theta\}_{\theta \in \Theta_p}$ containing an unknown unique member associated with a “unique” observed random sample $y_0$ occurring in $\mathcal{Y}$. Before sampling, the sample is a random element $y$ of $\mathcal{Y}$ governed by one and only one of the $P_\theta$’s; after sampling, $y_0$ is a known element of $\mathcal{Y}$. To apply Gauss’s postulate, it is necessary and sufficient to be willing to assert that fortuitous observation errors are the “only” reason why observations and the quantities measured do not coincide.
In this renewed Gaussian theory of errors, a DPE is the end product of the error analysis process, whereas a ME is just a summary of a DPE—useful for being expedient and able to reduce publishing costs. Since the theory of errors is a physics theory, every assumption formulated to solve a problem should have a physical basis.

To initialize a renewed theory of errors solution for a problem one needs an invertible pivot $T(y; \theta) = e$ for which the pivot space is reduced so that the partial function $T(y; \theta) : \Theta_p \rightarrow \mathcal{E}$ is one-one and onto its range space for (almost) each $y$ in $\mathcal{Y}$. Both this partial function and its inverse must be measurable. One hopes to find either boundedly complete minimal sufficient statistics to estimate $\theta$ or a Fraser structural model for this initial reduction, but one may be left with maximum likelihood estimators and ancillary statistics. This produces a reduced sample space $\hat{\Theta}_p$, a duplicate of $\Theta_p$ from which a reduced pivot space $\mathcal{E}_p$ can (theoretically) be easily obtained. One may now suppose that the $P_\theta$s are defined on $\hat{\Theta}_p$ and the pivot $T$ on $\hat{\Theta}_p \times \Theta_p$.

Different pivots will produce different DPEs, unless it is proven otherwise, since calculations from each pivot, while complying to the laws of probability, may concern different fields of events.

The pivot distribution $Q$ is obtained from the sampling distribution of an estimator $\hat{\theta}$ by means of the equation $Q(A) = P_\theta(T^{-1}(A; \theta))$ in which $Q(A)$ does not depend on $\theta$ since $T$ is a pivot. Set $A$ is a pivot-event and $T^{-1}(A; \theta)$ is the inverse image set function of the partial point function $T(\cdot; \theta) : \Theta \rightarrow \mathcal{E}_p$ when $\theta$ is held fixed.

Up to now, the model concerns the future since the real time is after $y_0$ is known. Because of Postulate 1, the distribution $Q$ is then kept as an error distribution instead of a probability distribution. The space $\mathcal{E}_p$ is now an error space and a pivot-event $A$ is now called an error-event. The DPE $P_{y_0}$ over parameter-events is given by $P_{y_0}(C) = Q(T(y_0; C))$, where $C$ is a parameter-event and $T(y_0; \cdot)$ is the inverse image set function of the point function $T_{-1}(y_0; \cdot)$, which is the inverse of the partial point function $T(y_0; \cdot)$ when the first argument is held fixed.

A renewed theory of errors approach does not imply a unique answer to a problem, but promises to evaluate faithfully any error resulting from the invertible pivot used. If one uses only part of the data, this inefficiency should be indicated by a larger resulting error.

A derived scalar parameter $\psi = \psi(\theta)$ has a DPE induced from $P_{y_0}$ — as is done in an improper Bayesian approach after a posterior distribution is known over the parameter space. This process is often called “marginalization.” Section 9.4 points out that marginalization is unavoidable and that it is not difficult to choose which pivot to use in the few cases in which more than one of them is available.

4.2. Dempster’s Conditioning as an Error Propagation Method Capable of Solving Some Improper CI Problems

The noncentrality parameter ($\delta^2$) estimation problem in Section 2.1 can be solved by using what I call Dempster’s conditioning—which is to use a conditional error measure, given an error-event which is known to have happened—although revealed by one or by several statistics which may be nonancillary (Dempster, 1967, 2008). This was used by Plante (1979a, 1979b), Edlefsen, Liu & Dempster (2009), Hannig (2009), Martin et al. (2010), Martin & Liu (2013, 2014a, 2014b, 2016) and Hannig et al. (2016). The solution, which is explained in the Appendix, is

$$
\delta^2(d_0^2, 1 - \alpha) = F_{-1}(d_0^2; (1 - \alpha)F(d_0^2; 0, p), p),
$$

where $d_0^2$ is the observed value of $d^2$ and where the lower index $-1$ refers to the inverse function of the restricted partial function $F(d_0^2; \cdot, p) : [0, \infty) \rightarrow (0, F(d_0^2; 0, p)]$ when the first and third arguments are fixed. A corresponding two-sided ME would be

$$
[F_{-1}(d_0^2; (1 - \alpha/2)F(d_0^2; 0, p), p), F_{-1}(d_0^2; (\alpha/2)F(d_0^2; 0, p), p)].
$$
5. AN EXTENDED FRASER’S FREQUENCY INTERPRETATION FOR NONLINEAR PIVOTS

5.1. An Observation-Centred Frequency Interpretation

Error analysis at its elementary level uses only inequalities such as \(|y_0 - \theta| < \Delta\), where \(y_0\) is an observation, \(\theta\) the quantity measured and \(\Delta\) is the largest error (in absolute value) judged possible—thus it does not require the use of probability or measurements of uncertainty. Following the Gaussian advance from \(|y_0 - \theta| < \Delta\) to \(y_0 - \theta = e\), 19th-century metrologists did not object to \(y_0\)-centred errors. This is likely because, in Gauss’s theory, a physical law expressed by an equation is deterministic and errors are associated with the observation of a phenomenon—consequently with an estimate rather than with a parameter. However, after the introduction of CI theory in the 20th-century, errors became considered to be \(\theta\)-centred. Fraser’s frequency interpretation and its extension, provided in Section 5.3, helps bridge this gap between a Gaussian and a Neymanian viewpoint. Fraser’s frequency interpretation for linear or for group-constructed pivots, briefly described here (Fraser, 1961; Cox & Hinkley, 1974, p. 248) and its extension useful for nonlinear pivots, both show how a pre-sampling \(\theta\)-centred error random variable can produce the needed post-sampling \(y_0\)-centred DPE.

5.2. Fraser’s Frequency Interpretation

Suppose that a linear pivot \(y - \theta = e\) is available. Then, consider a large number of observations \(y_{1,0}, \ldots, y_{N,0}\) observed under apparently the same conditions and corresponding to known values \(\theta_1, \ldots, \theta_N\) of a parameter \(\theta\). Let \(y_{N+1,0}\) be a new observation for which the parameter value \(\theta_{N+1}\) is unknown and for which a reference distribution for \(\theta_{N+1}\) centred on \(y_{N+1,0}\) is needed. Fraser’s frequency interpretation can then be activated as follows. Consider the errors \(e_{i,0} = y_{i,0} - \theta_i\) \((i = 1, \ldots, N)\) and define \(\theta^{*}_{i,0}\) to be such that \(y_{N+1,0} - \theta^{*}_{i,0} = e_{i,0}\), i.e., \(\theta^{*}_{i,0} = y_{N+1,0} - e_{i,0}\). A dot diagram or a histogram of the \(\theta^{*}_{i,0}\)’s illustrates the idea of a DPE on \(y_{N+1,0}\). “This amounts to slide each \(y_{i,0}\) and \(\theta_i\) by the same amount \(y_{N+1,0} - y_{i,0}\)” (D.A.S. Fraser, communication in a seminar, 1961). The corresponding DPE is that of \(\theta_{N+1} = y_{N+1,0} - e_{N+1}\).

5.3. An Extended Fraser’s Frequency Interpretation

Fraser’s frequency interpretation can be extended to a nonlinear pivot as follows.

Suppose that an invertible nonlinear pivot \(T(\hat{\theta}, \theta) = e\) is available. Consider a large number of estimates \(\hat{\theta}_{1,0}, \ldots, \hat{\theta}_{N,0}\) observed under apparently the same conditions and corresponding to known values \(\theta_1, \ldots, \theta_N\) of a parameter \(\theta\). Let \(\hat{\theta}_{N+1,0}\) be a new estimate for which the parameter value \(\theta_{N+1}\) is unknown and for which a reference distribution for \(\theta_{N+1}\) centred on \(\hat{\theta}_{N+1,0}\) is needed. Let the errors be \(e_{i,0} = T(\hat{\theta}_{i,0}, \theta)\) \((i = 1, \ldots, N)\) and define \(\theta^{*}_{i,0}\) to be such that \(T(\hat{\theta}_{N+1,0}, \theta^{*}_{i,0}) = e_{i,0}\), i.e., \(\theta^{*}_{i,0} = T^{-1}(\hat{\theta}_{N+1,0}, e_{i,0})\), where \(T^{-1}(\hat{\theta}_{N+1,0}, \cdot)\) is the inverse of the partial function \(T(\hat{\theta}_{N+1,0}, \cdot)\) when the first argument is fixed. A dot diagram or a histogram of the \(\theta^{*}_{i,0}\)’s can illustrate the idea of a DPE on \(\hat{\theta}_{N+1,0}\). The corresponding DPE is that of \(\theta_{N+1} = T^{-1}(\hat{\theta}_{N+1,0}, e_{N+1})\).

In both Fraser’s frequency interpretation and in this extension, the \(\theta^{*}_{i,0}\)’s are an answer to the question: what would the \(\theta^{*}_{i,0}\) values have to be for all the \(\hat{\theta}_{i,0}\)’s to fall at \(\hat{\theta}_{N+1,0}\) and all the \(e_{i,0}\)’s to be unchanged?

With a nonlinear pivot it is not possible to slide each \(y_{i,0}\) and \(\theta_i\) by the “same” amount as can be done with a linear pivot. However, a transformation comparable to a “motion” in differential geometry (Matsushima, 1972, Section II.9) is achieved. The suggested frequency interpretation extension is a class of transformations \(T: \hat{\Theta}_p \times \Theta_p \longrightarrow \hat{\Theta}_p \times \Theta_p\), leaving errors invariant, with \(\hat{\theta}_{N+1,0}\) as the image, where \(\hat{\theta}_{N+1,0}\) is fixed and \(\theta\) is variable. Since these transformations

DOI: 10.1002/cjs
satisfy the equation \( F \circ F = F \), each can be called a projection into the hyperplane \( \hat{\theta} = \hat{\theta}_{N+1,0} \). If the number of points projected is large, the result is essentially a DPE. Note that the set of points corresponding to a fixed error \( e \) in \( E_p \) is \( T^{-1}(\{e\}) \), where \( T^{-1} : P(E_p) \rightarrow P(\hat{\theta}_p \times \Theta_p) \) is the inverse image set function of the point function \( T : \hat{\theta}_p \times \Theta_p \rightarrow E_p \) and where \( P \) indicates the power set of a set. Suppose that the partial function \( T(\cdot, \theta) : \hat{\Theta} \rightarrow E_p \) is also a bijection, that, for every fixed \( \theta \), every partial derivative of \( T(\hat{\theta}, \theta) \) with respect to the components of \( \hat{\theta} \) and \( \theta \) and of \( T^{-1}(e, \theta) \) and \( T^{-1}(\hat{\theta}, e) \) with respect to the components of \( e \), \( \hat{\theta} \), and \( \theta \) exists and is continuous. Thus, a projection \( F \) is comparable to a motion within the manifold \( T^{-1}(\{e\}) \), although an error is not a metric as in the differential geometry definition. If the \( e \)-space has dimension one, this “motion” is made along the curve \( T^{-1}(\{e\}) \). Given an error-event \( A \) and an estimate \( \hat{\theta}_{N+1,0} \) of \( \theta \), \( T^{-1}(\hat{\theta}_{N+1,0}, A) \) is an 100\( Q(A)\)% (generalized) ME. When examining the relationship between Gauss’s postulate and Neyman’s repeated sampling axiom/principle for a scalar parameter, it is interesting to find that the “bundle of curves” \( T^{-1}(A) \) is a 100\( Q(A)\)% (artificial) error-belt.

Gauss’s condition that “the only reason why \( \hat{\theta}_0 \neq \theta \) is the presence of error \( e \)” naturally supposes the potentially useful facts that: (i) the origin \( 0_p \) of \( E^p \) is in \( E_p \) and (ii) \( T(\hat{\theta}_0, \theta) = 0_p \) iff \( \theta = \hat{\theta}_0 \) for any observation \( \hat{\theta}_0 \). Thus, the following relation holds: \( T^{-1}(\{0_p\}) = \{(\hat{\theta}_0, \theta) : \hat{\theta}_0 = \theta \} \), which is the main diagonal of \( \hat{\theta}_p \times \Theta_p \).

5.4. Margin of Error as a Posterior Credible Set Using an Improper Prior

Fisher (1959, p. 101) called the strong repeated sampling axiom of Neyman–Pearson theory “artificial” likely because it supposes that the unknown value of the parameter is controlled. Actually, to suppose that an unknown value is controlled is illogical and does not correspond to the situation met in scientific research. Furthermore, applying the strong repeated sampling principle rigidly leads to too many improper CIs, while if problematic confidence belts are excluded, too many problems are left without an answer.

Gauss’s error postulate implies a necessary assumption concerning how parameter values occur in scientific work. This assumption, which is analogous to an improper Bayesian prior, appears in his first theorem in Gauss (1809; 1855, Note I). Proposition 1 in Section 5 of the Web Appendix in the Supplementary Material expresses this assumption in modern statistical language. To observe the right average coverage frequency experimentally for an “error belt” made up of MEs (of the same level and on estimates from the same estimator) when the pivot is linear, one must simulate an almost improper prior to produce parameter proxy values instead of controlling parameter values as is done when using the strong repeated sampling axiom. Examples of almost improper priors are given in Section 6 of the Web Appendix in the Supplementary Material.

If the pivot is nonlinear, there always exists an improper prior for which the posterior would provide the same posterior credible interval as the corresponding ME at the same level. Thus a ME cannot be improper. Unfortunately, we can only know “a posteriori” what this prior is because it depends on the observation \( y_0 \), being given by the Schweder and Hjort–Hannig–Fraser Theorem—from three different contexts—stated in Section 5 of the Web Appendix.

Dempster’s conditioning—mentioned in Section 2.1—uses an extra assumption because an invertible pivot does not exist. It is assumed, however, that a conditional error measure does exist. It is easily confirmed by simulation that in Mandelkern’s (2002) restricted mean problem, mentioned in Section 2.1, Dempster’s conditioning solution has the proper average coverage frequency if a very small uniform prior density is assumed for \( \mu \geq 0 \), for \( \mu^2 \) or for \( \mu^3 \).
6. SIMULATING THE PROPAGATION OF FIVE DIMENSIONAL ERRORS

6.1. A Solution for Any (Computable) Parameter Derived from the Five Parameters of the Bivariate Normal Distribution

Berger & Sun (2008) and Bebu & Mathews (2008) have used algorithms based on the \( \mathcal{N}_2 \) distribution to generate estimation distributions. Here, I use Fisher’s step-by-step argument (Fisher, 1959, pp. 120, 163–164, 171–175). Bebu & Mathew’s method, based on generalized pivots, has much in common with the renewed theory of errors method so that the results of both are expected to be close. This is approximately verified when examining their ratio of bivariate-lognormal-means example. Based on 5,000 iterations and expressed with the “±” notation for each of their two versions of the generalized pivot, their 95% generalized CI is 0.7857 ± 0.4650 and 0.7550 ± 0.4488 for the ratio of means of two health care costs, with \( N = 98, \bar{x} = 6.41, \bar{y} = 6.50, s^2_x = 2.73, s^2_y = 3.48, r = 0.450 \) (using the logs of the data). In comparison, the renewed theory of errors method gives 95% ME 0.733 ± 0.403 (Median DPE = 0.6248; Mean DPE ≈ 0.653). Bebu and Mathew pointed out that once their generalized pivots are established, they could estimate any (computable) parameter derived from the five parameters of the \( \mathcal{N}_2 \) distribution. Furthermore, their method provided numerical tests with larger power compared to the modified log-likelihood test in Barndorff-Nielsen (1991).

Using the prototype program included in the Software Web Appendix in the Supplementary Material, I can also produce MEs for any (computable) parameter. This is done by changing within the program the instruction \( \lambda = \mu_y/\mu_x \) (literally: “w = muy/mux”) into \( \lambda = \text{“new formula.”} \)

To exemplify this I used the data on vitamin B-12 in Section 8 and the satellite parameters: \( \hat{\theta} \) jointly sufficient estimators present—or absent—at each stage of Fisher’s procedure. The general process starts with \( p \) jointly sufficient estimators \( \hat{\theta}_1, \ldots, \hat{\theta}_p \) for parameters \( \theta_1, \ldots, \theta_p \) and assumes that the marginal density of \( \hat{\theta}_1, \ldots, \hat{\theta}_p \) for \( 1 \leq i \leq p \), is \( f(\hat{\theta}_1, \ldots, \hat{\theta}_i; \theta_1, \ldots, \theta_i) \). The marginal density of \( \hat{\theta}_1 \) is \( f(\hat{\theta}_1; \theta_1) \); the conditional density of \( \hat{\theta}_2 \) given \( \hat{\theta}_1 \) is \( f(\hat{\theta}_2 | \hat{\theta}_1; \theta_1, \theta_2) \); and the conditional density of \( \hat{\theta}_i \) given \( \hat{\theta}_1, \ldots, \hat{\theta}_{i-1} \) is \( f(\hat{\theta}_i | \hat{\theta}_1, \ldots, \hat{\theta}_{i-1}; \theta_1, \ldots, \theta_i) \). Accordingly, the step-by-step argument uses successive probability integral transformations corresponding to distribution functions \( F(\hat{\theta}_1; \theta_1), F(\hat{\theta}_2 | \hat{\theta}_1; \theta_1, \theta_2), \ldots, F(\hat{\theta}_p | \hat{\theta}_1, \ldots, \hat{\theta}_{p-1}; \theta_1, \ldots, \theta_p) \).

DOI: 10.1002/cjs

The Canadian Journal of Statistics / La revue canadienne de statistique
Let $\bar{x}, \bar{y}, s_x, s_y, r$ be estimators of $\mu_x, \mu_y, \sigma_x, \sigma_y, \rho$. The sample size is $N$ and $n = N - 1$. To use a bivariate normal distribution, first apply the transformations $\sigma_x = a^{1/2}e^{-\beta/2}$, $\sigma_y = a^{1/2}e^{-b/2}$, $s_x = a^{1/2}e^{-\beta/2}$, $s_y = a^{1/2}e^{b/2}$ (Hotelling, 1953). The reverse transformations are $a = \sigma_x \sigma_y, b = \ln(\sigma_y/\sigma_x), a = s_x s_y, b = \ln(s_y/s_x)$. Since $\{\bar{x}, \bar{y}\}$, $\{s_x, s_y, r\}$ and standardized residuals are three independent sets of random variables, and since $\{\bar{x}, \bar{y}\}$ are sufficient for $(\mu_x, \mu_y)$, two sets of parameters must be estimated—first $a$, $b$ and $\rho$ and then $\mu_x$ and $\mu_y$. The fast converging expansion of the density of $r$ in Hotelling (1953) is central to the algorithm. It is

$$f(r; \rho) = \frac{(n-1)(1-r^2)^{(n-3)/2}(1-\rho^2)^{n/2}}{\sqrt{2\pi}^3/2(1-\rho r)^{n-1/2}} \sum_{j=0}^{\infty} \frac{\Gamma^2(j+\frac{1}{2})}{\Gamma(n+j+\frac{1}{2})!} \left(1 + \rho r\right)^j.$$

(2)

**Proposition 1 (Partial five dimensional pivot).** With the notation introduced above,

$$f(b \mid r; \beta, \rho) = \frac{\sqrt{\pi/2}(1-\rho r)^{n-1/2}}{\Gamma(n) \sum_{j=0}^{\infty} \frac{\Gamma^2(j+1/2)}{\Gamma(n+j+1/2)!} \left((1+\rho r)/2\right)^j} \left\{\cosh(b-\beta) - \rho r\right\};$$

(3)

$$(a \mid b, r; \alpha, \beta, \rho) \sim \frac{\alpha(1-\rho^2)\chi^2_{2n}}{2n\{\cosh(b-\beta) - \rho r\}},$$

(4)

where $\chi^2_{2n}$ is a random variable with the chi square distribution with $2n$ degrees of freedom.

Equations (2)–(4) are sufficient to solve the simulation problem, since the distribution functions $F(r; \rho)$, $F(b \mid r; \beta, \rho)$ and $F(a \mid b, r; \alpha, \beta, \rho)$ may be calculated by integrating cubic spline polynomials representing $f(r; \rho)$, $f(b \mid r; \beta, \rho)$ and $f(a \mid b, r; \alpha, \beta, \rho)$. The results are quartic spline polynomials yielding DPEs.

Marginal and conditional error distribution functions $G(\rho; r)$, $G(\beta \mid \rho; b, r)$ and $G(\alpha \mid \beta, \rho; a, b, r)$ must, for any $(a, b, r; \alpha, \beta, \rho)$, satisfy the conditions $G(\rho; r) = 1 - F(r; \rho)$, $G(\beta \mid \rho; b, r) = 1 - F(b \mid r; \beta, \rho)$ and $G(\alpha \mid \beta, \rho; a, b, r) = 1 - F(a \mid b, r; \alpha, \beta, \rho)$. Therefore, the following corollary holds.

**Corollary 2.** Given uniform pseudo-random numbers between zero and one, $(u_1, u_2, u_3)$, simulated values for $(\rho, \beta, \alpha)$ may be obtained by solving, successively, $F(r; \rho) = 1 - u_1$ for $\rho$, $F(b \mid r; \beta, \rho) = 1 - u_2$ for $\beta$ and $F(a \mid b, r; \alpha, \beta, \rho) = 1 - u_3$ for $\alpha$.

There are two possible pivots available for obtaining a DPE for estimates of $\mu_x$ and $\mu_y$ from the bivariate normal distribution of $\bar{x}$ and $\bar{y}$ with means $\mu_x$ and $\mu_y$, standard deviations $\sigma_x/\sqrt{N}$ and $\sigma_y/\sqrt{N}$ and correlation coefficient $\rho$. Thus, if $z_1$ and $z_2$ are two independent error variables, each with the standard normal distribution, we may set

$$\mu_x = \bar{x} - (\sigma_x/\sqrt{N})z_1, \mu_y = \bar{y} - \rho(\sigma_y/\sigma_x)(\bar{x} - \mu_x) - \frac{\sqrt{1-\rho^2}\sigma_y}{\sqrt{N}}z_2 \text{ or},$$

(5)
\[
\mu_y = \bar{y} - (\sigma_y / \sqrt{N}) z_1, \quad \mu_x = \bar{x} - \rho(\sigma_x / \sigma_y)(\bar{y} - \mu_y) - \sqrt{1 - \rho^2} \sigma_x z_2. \tag{6}
\]

These two distributions appear identical when simulated, although negligible sampling variations differentiate them. However, as our main objective is to construct a ME for a ratio of means, Equation (5) is preferable to Equation (6) since the numerator is usually compared to the denominator.

Random numbers \(u_1, u_2, u_3, z_1, z_2\) — thus simulated values for \(\rho, \sigma_x, \sigma_y, \mu_x, \mu_y\) — are obtained 10 million times. A simulated value of, say, \(\lambda = \mu_x / \mu_y\) is obtained at each iteration and a histogram bin is updated. Details may be obtained from a VBA for Excel (FORTRAN like) program in the Software Web Appendix in the Supplementary Material.

It is shown in the Appendix that the five-dimensional pivot is invertible.

**Proposition 2 (The five-dimensional pivot is invertible.).** With the above notation, the transformation from the error space \(\Theta = \{(u_1, \ldots, u_5) : 0 < u_i < 1, i = 1, \ldots, 5\}\) to the parameter space \(\mathcal{E}\) defined by \(\rho = G^{-1}(u_1; r_0), \beta = G^{-1}(u_2 | r; b_0, r_0), \alpha = G^{-1}(u_3 | \beta, \rho; a_0, b_0, r_0), \sigma_x = a^{1/2}e^{-\beta/2}, \sigma_y = a^{1/2}e^{\beta/2}, (s_x, 0) = a^{1/2}e^{-b_0/2}, s_x, 0 = a^{1/2}e^{b_0/2}, \mu_x = G^{-1}(u_4 | \alpha(\sigma_x, \sigma_y), \beta(\sigma_x, \sigma_y), \rho(\bar{x}_0, s_x, 0, s_y, 0, r_0))\) and \(\mu_y = G^{-1}(u_5 | \mu_x, \alpha(\sigma_x, \sigma_y), \beta(\sigma_x, \sigma_y), \rho(\bar{x}_0, s_x, 0, s_y, 0, r_0))\) is a bijection for any estimate \(\bar{x}_0, \bar{y}_0, s_x, s_y, r_0, r_0\) where \(u_4 = \Phi(z_1)\) and \(u_5 = \Phi(z_2)\).

7. USING A RENEWED THEORY OF ERRORS TO IMPROVE FIELLER CONFIDENCE SETS

Given a sample of size \(N\) from a \(\mathcal{N}_2\) population, Fieller’s pivot for constructing Fieller’s CIs for \(\lambda\) is

\[
t = \sqrt{N} \frac{(\bar{y} - \lambda \bar{x})}{\sqrt{s_y^2 - 2rs_x\sigma_y + s_x^2\lambda^2}}, \tag{7}
\]

where the variable \(t\) in Equation (7) has the Student distribution with \(N - 1\) degrees of freedom.

For clarity, Proposition 3 describes the typology of Fieller’s confidence sets in terms of sets of “Studentized means” (\(\bar{x}^*, \bar{y}^*\)) with positive area, where \(\bar{x}^* = \sqrt{N}\bar{x}/s_x\) and \(\bar{y}^* = \sqrt{N}\bar{y}/s_y\). Confidence sets for a ratio of means are also described in Fieller (1954) and in Scheffé (1970).

**Proposition 3 (Essential typology of Fieller’s confidence sets.).** With the above notation, and writing \(t_{a/2}\) for the upper 100\(a\)%-th percentage point of the Student distribution with \(N - 1\) degrees of freedom, consider the plane \(\bar{x}^*O\bar{y}^*\). The ellipse

\[
\bar{x}^*2 - 2r\bar{x}^*\bar{y}^* + \bar{y}^*2 - (1 - r^2)t_{a/2}^2 = 0 \tag{8}
\]

has a major axis \(\bar{y}^* = \text{sign}(r)\bar{x}^*\) and is inscribed within the square \(\max\{|\bar{x}^*|, |\bar{y}^*|\} = t_{a/2}\). To determine a 100\((1 - a)\)% confidence set for \(\lambda\) based on the pivot in Equation (7), there are three regions with positive area determining the type of the resulting confidence sets:

(i) the region interior to the ellipse in Equation (8) where the confidence set is the interval \((-\infty, \infty)\);

(ii) the region \(\{(\bar{x}^*, \bar{y}^*) : |\bar{x}^*| > t_{a/2}\}\) where the confidence set is the two-sided interval

\[
\lambda = \frac{s_y}{s_x(\bar{x}^*2 - t_{a/2}^2)} \left\{ (\bar{x}^*\bar{y}^* - rt_{a/2}^2) \pm t_{a/2}\sqrt{(\bar{x}^*2 - 2r\bar{x}^*\bar{y}^* + \bar{y}^*2) - (1 - r^2)t_{a/2}^2} \right\}; \tag{9}
\]

(iii) the region within which \(|\bar{x}^*| < t_{a/2}\) but which is exterior to the ellipse in Equation (8), where the confidence set is the complement of the two-sided interval determined by Equation (9).
A solution with an upper and a lower confidence coefficient using both a “fiducial” coefficient calculated from a confidence distribution and a confidence coefficient was suggested in Plante (1994). Tsao & Hwang (1998) applied an estimated confidence approach due to Kiefer (1977) to remove the awkwardness of calling (−∞, ∞) a “95% CI.” Hannig, Wang & Iyer (2003) indicated an asymptotic approach. Liseo (2003) analyzed the ratio of means problem from different viewpoints: frequentist, likelihood and Bayesian. Moineddin, Beyene & Boyle (2003) investigated properties of various location quotient methods (“…a way of measuring the relative contribution of one specific area to the whole for a given [geographical] outcome”). Sherman, Maity & Wang (2011) compared various methods to estimate a ratio in a sample survey context.

Figure 2 illustrates differences between standard CI solutions and renewed theory of errors solutions to the estimation problem of a ratio of means using the famous Cushny & Peebles data from Fisher (1925–1970, Section 26.2, since 1946) \(N = 10, \bar{x} = 2.33, \bar{y} = 0.75, s_x = 2.002, s_y = 1.789, r = 0.7952\). I am assuming normality (probably falsely), correlation between the random variables behind the samples compared for the two most peaked curves, and (falsely) no correlation for the two less peaked curves. CIs are represented by approximate confidence densities derived from Fieller’s pivot in Equation (7), while the DPE densities are derivatives of spline-smoothed empirical distribution functions derived from error simulations with 10 million iterations. Interestingly, the most peaked curve has longer and more asymmetrical tails than does the next highly peaked curve. (Details are shown in Table 1.)

Table 1 illustrates, using the Cushny and Peebles data, the suggested change to interval estimation of a ratio when the denominator is close to zero. The factor \(k\) is such that only the means are modified by the transformation \((\bar{x}, \bar{y}) \rightarrow (k\bar{x}, k\bar{y})\) so that \(k\bar{y}/k\bar{x}\) is constant. One can see on line three of the table (which is beside the region where CIs cannot be used in lines one and two) that CIs can be extremely skewed. The median (unbiased) estimates, however, are more stable than are the ends of the CIs, which are closer to a singular point. This can be

---

**Figure 2:** Error densities and confidence densities for a ratio of means. DPEs are approximations based on 10,000,000 iterations. (1) Solid line: indicates the DPE density on a ratio of means for the Cushny and Peebles data. (2) Short dashed line: indicates the approximate confidence density for a ratio of means. (3) Long dashed line: indicates the DPE density on a ratio of means—assuming equality of variances and (falsely) independence. (4) Very short dashed line: indicates the approximate confidence density assuming equality of variances and (falsely) independence.
TABLE 1: Comparison at the 95% level of Fieller’s CIs (CIs: [CI L, CI U]) and margins of error (MEs: [ME L, ME U]) for a ratio of means with large values of the coefficient of variation of the denominator estimate of mean. Symbol “M” indicates a median of a distribution. Data are the Cushny and Peebles data. Symbol k indicates a common factor of both means, other estimates being invariant. Column M is the unbiased median estimate or 0% interval. When the lower limit is larger than the upper limit, this indicates that the confidence set is the complement of the interval with the given limits. The precision achieved using 10,000,000 iterations in the simulations is indicated by the number of significant figures quoted for the MEs.

| k   | CI L | CI U | CI M  | ME L   | ME U   | ME M   |
|-----|------|------|-------|--------|--------|--------|
| 0.25| −∞   | ∞    | 0.321888 | −5.16  | 2.41   | 0.4181 |
| 0.50| 1.54750 | 1.39446a | 0.321888 | −2.51  | 1.127  | 0.3398 |
| 0.62| −44.7434 | 0.88673 | 0.321888 | −1.626 | 0.930  | 0.3294 |
| 0.70| −2.77873 | 0.80480 | 0.321888 | −1.215 | 0.847  | 0.3262 |
| 0.80| −1.21514 | 0.73875 | 0.321888 | −0.855 | 0.775  | 0.3240 |
| 1.00| −0.48485 | 0.65658 | 0.321888 | −0.439 | 0.6897 | 0.3229 |
| 1.50| −0.06621 | 0.55311 | 0.321888 | −0.067 | 0.5662 | 0.3219 |
| 3.00| 0.16423   | 0.44548 | 0.321888 | 0.1616 | 0.4505 | 0.3219 |
| 5.00| 0.23310   | 0.39874 | 0.321888 | 0.2312 | 0.4015 | 0.3219 |

*aInverted limits indicate the complement of an interval.

seen by comparing the fourth and the seventh columns with the second and fifth columns of the table. Note that the line with k = 1 corresponds to the result quoted in Fisher (1925–1970, Section 26.2, since 1946). We see clearly in Table 1 that the effect of a singular point at kµx = 0 tapers off as k increases—as Neyman had indicated. (A more extended table is given in the Web Appendix in the Supplementary Material showing the variation of estimating intervals as functions of k and r.)

If a 100(1 − α)% CI with α < 0.05 (and k = 1) is wanted, the Cushny & Peebles sample could give a CI of the form [−∞, b] ∪ [a, ∞) instead of the form [a, b] occurring at the 5% level. One could also obtain a CI of the form (−∞, ∞). Rare unacceptable CIs can also occur in repeated sampling at the 95% confidence level. Their overall effect is to decrease slightly the length of finite Fieller’s CIs.

8. ESTIMATING THE SLOPE OF THE MAJOR AXIS OF A BIVARIATE NORMAL DISTRIBUTION

8.1. Jolicoeur’s Pivot

Improper CIs are an incapacitating difficulty when estimating the slope ω = [σy − σx + ((σy − σx)2 + 4ρσyσxσy)1/2]/2ρσyσx of the major axis of a Ν2 population.

Setting CIs for the slope ω is complicated by both a singularity of this derived parameter at ρ = 0, and also, if ρ ≠ 0, because ω = ±1 when σx = σy —side-effects due to singularities are to be expected. Also, the pivotal equation giving the desired CIs provides four solutions, as was described by Jolicoeur (1973) when exploring possible complications occurring with CIs having infinite length. Furthermore, there are warnings by software providers for researchers in biology (Warton et al., 2006) concerning using CIs for the slope of the major axis using Jolicoeur’s method. Warton et al. suggest that the slope of the reduced major axis Sign(ρ)σy/σx be used instead of ω to avoid estimation problems occurring with ω.

DOI: 10.1002/cjs

The Canadian Journal of Statistics / La revue canadienne de statistique
I provide an example of a maximum likelihood estimate of the major axis of an \( \mathcal{N}_2 \) distribution and a CI based on that estimate where, in a routine data analysis, both the point estimate and the CI were found to be unacceptable—both visually and analytically. A renewed theory of errors approach appears definitely preferable to using a CI approach to estimate the slope of a major axis and related parameters—although the two approaches occasionally do give congruent answers.

The maximum likelihood estimator \( \hat{\omega} \) corrected for biased estimation of variances is obtained by substituting \( s_x, s_y, \) and \( r \) instead of \( \sigma_x, \sigma_y, \) and \( \rho \) in the formula for \( \omega \). Anderson (1963) found asymptotic CIs for \( \omega \) based on a chi-square distribution, and Jolicoeur (1968, 1973) used the related pivot,

\[
v = \frac{(N - 2)(-s_x s_y r \omega^2 + (s_x^2 - s_y^2) \omega + s_x s_y r)^2}{s_x^2 s_y^2 (1 - r^2)(\omega^2 + 1)^2}.
\] (10)

The pivot \( v \) has an \( F \) distribution with 1 and \( N - 2 \) degrees of freedom.

The pivot in Equation (10) being improper, gives confidence sets for \( \omega \) that are either an “inclusive” interval \([o_1, o_2]\); an “exclusive interval” on the extended real number system—the complement of \((o_2, o_1)\) (when \( o_1 > o_2 \)); or the interval \((-\infty, \infty)\). Jolicoeur (1973) distinguished 13 types of samples relevant for the four possible roots resulting from Equation (10).

### 8.2. The Amount of Vitamin B12 in the Blood of Mothers and Their Newborns

The routine data analysis in question shows how a standard CI method based on the maximum likelihood method and on Jolicoeur’s pivot in Equation (10) can unexpectedly provide misplaced CIs and point estimates.

The data subset of vitamin B12 birth-levels for mothers and their newborn infants, used in the next example, is extracted from a data set with several hundred variables collected from 207 births in the Baudelocque maternity clinic in Paris during a 5-month period between 1984 and 1985 for an environmental study. Medical and environmental information, as well as further statistical analysis using these data, may be found in Fréry et al. (1992). The data were kindly provided by these researchers from the Institut National de la Santé et de la Recherche Médicale de France.

I omitted 57 couples from the vitamin B12 data with either missing information or with dubious components in order to avoid being confronted with a mixed distribution not relevant for illustrating a difficulty in Jolicoeur’s technique. What remained resembles a sample from a bivariate lognormal distribution. Using the lognormality assumption, I excluded one outlier couple in which the mother had too much vitamin B12 in her blood. (A remark in a Mayo Clinic website warned against this possibility often due to vitamin supplements.) Consequently, the intermediate data summary (before reducing the number of significant figures) is \( N = 149 \), \( \bar{x} = 5.51774, \bar{y} = 6.38079, s_x = 0.44264, s_y = 0.616214, \) and \( r = 0.628373 \), where \( \bar{x} \) and \( \bar{y} \) are the natural logarithms of the amount of vitamin B12 (pg/ml) in, respectively, the mother’s blood and the newborn’s (cord) blood, while \( r \) is the sample correlation coefficient between \( x \) and \( y \), and \( s_x \) and \( s_y \) are the sample standard deviations of \( x \) and \( y \). The \((x, y)\)s in Figure 3 are assumed to resemble a random sample from a bivariate normal distribution.

In order to perform a chi-square goodness of fit test, the scatter diagram in Figure 3 was split into 20 classes. Assuming parameters equal to the corresponding estimates from the vitamin B12 data, class probabilities with a precision of at least four significant figures were calculated by simulation for simplicity. After merging classes having fewer than five expected observations with adjacent classes, but admitting one class with a frequency between 1 and 5, 14 classes remained and the chi-square test had eight degrees of freedom. The bivariate “sample” did not appear significantly different from one drawn from a bivariate normal distribution (\( P = 0.313 \)).
Figure 3: A misplaced CI for the slope of a major axis. Vitamin B12 in blood at birth data on logarithmic scales ($x$ for mother, $y$ for newborn) with lines through the point ($\bar{x}, \bar{y}$) with slopes corresponding to the 95% anomalous CI for the slope of the major axis and to the maximum likelihood estimate of the major axis.

Furthermore, both sample skewness and kurtosis coefficients for marginal distributions of log-observations are moderate. Interestingly, as shown in Figure 2 in Fréry et al. (1992), the conditional distributions of the amount of vitamin B12 in mothers’ blood, given their ages, do not generally follow simple lognormal distributions.

8.3. Misplaced Maximum Likelihood Estimate and CI for the Slope of a Major Axis

Figure 3 shows, using logarithmic scales, the scatter diagram of vitamin B12 birth-level data in mothers’ blood ($x$) and in their newborns’ cord blood ($y$), with three lines through the point ($\bar{x}, \bar{y}$); with slopes corresponding to the 95% CI ends; and, to the maximum likelihood estimate of the major axis slope. The scatter diagram major axis is within the 95% confidence cone, but the cone extends too much on the left towards the minor axis. Furthermore, the maximum likelihood
FIGURE 5: Explanation for a misplaced CI. $p$-Value function for Jolicoeur’s method on the B12 data where Jolicoeur’s method fails (solid line); and the data in Jolicoeur (1963) where Jolicoeur’s method succeeds (dotted line). According to Jolicoeur (1973), only the right side of the graphs should be considered.

major axis estimate “visually” misses the scatter major axis. Figure 4—which uses original coordinates—shows that the results obtained using the renewed theory of errors is as expected. Note that the horizontal and vertical scales are different in both figures. The 95% ME for the slope $\omega$ is $[1.3724, 1.6062]$ (DPE median = 1.48513) and the 95% CI is $[1.37523, 2.07054]$ (maximum likelihood estimate = 1.67082).

The $p$-value function for these data, given in Figure 5, explains why the CIs based on maximum likelihood estimates for this problem are unsatisfactory. A 10% CI for the slope of the major axis would be empty! A correction of corresponding side effects due to singularities affecting CIs with finite length is beyond the scope of a detailed analysis in this article. However, the main problem with using CIs for the slope of the major axis of a bivariate normal distribution is that “it is not always evident whether a particular CI is trustworthy” since it may be offset and elongated due the complex effect of several singularities.

While Jolicoeur’s technique is unsatisfactory with B12 vitamin in mothers’ and newborns’ blood data, the following example illustrates that it can also give a “good” answer. Using data in Jolicoeur (1963) concerning humerus dimensions of male *Martes Americana*, with $N = 92$, $s_x = 0.014276$, $s_y = 0.010744$, and $r = 0.593854$ (before reducing the number of significant figures due to data imprecision), the resulting 95% CI and 95% ME for $\omega$ are congruent. The 95% CI is $1.70 \pm 0.48$ with a median unbiased estimate equal to 1.60 and the 95% ME is $1.70 \pm 0.41$ with a median DPE equal to 1.60.

The suggested correction to CIs for the slope of the major axis of a $N_2$ distribution is to replace them by MEs based on a renewed theory of errors as is illustrated in Figure 4 using the original coordinates.

9. POSSIBLE OBJECTIONS TO A RENEWED THEORY OF ERRORS
9.1. Users of “Fiducial” Techniques

Manipulating error variables according to “fiducial” techniques (i.e., probability laws) was first done by Gauss (1821, 1855, Section 18) and by Fisher (1935, 1959). Fraser (1968) used hundreds of such manipulations. Formal fiducial techniques are also used in Bunke, H. (1975), Bunke, O. (1976), Barnard (1977), Plante (1979a, 1979b, 1994), and Dawid & Stone (1982). More recently, Weerahandi (1993, 2013), Iyer, Wang & Mathew (2004), Hannig (2009) and Hannig et al. (2016) performed fiducial as well as other procedures on formal random variables to obtain approximate
confidence distributions using simulation. Reid & Cox (2015) associated these approaches by Weerahandi, by Hannig, and by their collaborators with Cox’s confidence distribution approach. Hannig & Xi (2012) have indicated relationships between manipulation rules in the confidence distribution approach, in the Dempster–Shafer theory, and in generalized fiducial inference.

9.2. Objections to “Fiducial” Techniques

Mauldon (1955), Lindley (1958), Stein (1959), Buehler & Feddersen (1963) and Dawid, Stone & Zidek (1973) have provided counterexamples to fiducial techniques. As this section explains, the renewed theory of errors is not affected by these objections—although using integration to eliminate unwanted parameters.

Kolmogorov (1956, pp. 14–15) proved that his system of axioms for probability is consistent, and commented that “Our system of axioms is not, however, complete, for in various problems in the theory of probability different fields of probability have to be examined” (Kolmogorov 1956, p. 3). This incompleteness has several implications. One must be aware that when applying Kolmogorov’s formalism to a statistical problem for which both a pre-data probability model and a post-data renewed theory of errors model are needed, “the same formalism is being applied to different $\sigma$-fields.” Moreover, within error analysis itself, one must not pass inadvertently from one field into another one by algebraically modifying an expression referring to the observations. Paradoxes in Mauldon (1955)—as pointed out by Fisher (1959) in the last sentence of his book—as well as marginalization paradoxes in Dawid, Stone & Zidek (1973) indicate an inadvertent crossing from one field into another field when one attempts to create a paradox against a renewed theory of errors. Thus, using this type of paradox against a renewed theory of errors would be invalid. In other words, this so-called “paradox” would exhibit a situation in which one has intuitive expectations which are not supported by Kolmogorov’s theory of probability. In fact, Fisher (1960) solved a similar alleged paradox claimed by Lindley (1958).

Problematic inconsistencies can also exist within an “alleged” Gaussian theory of errors approach. For example, Buehler & Feddersen (1963) and Brown (1967) used Buehler’s version of fiducial terminology such as “relevant subset,” in which observations would not be treated as mathematical constants in post-data algebraic manipulations (Buehler, 1959, 1980), whereas an observation $y_0$ should be so treated within a theory of errors, and within a fiducial approach. Consequently, Buehler and Feddersen’s conclusions have no relevance for criticizing a theory of errors.

9.3. Dismissal of Stein’s Fiducial Example

Stein (1959) introduced the noncentrality coefficient underestimation problem treated in Section 2 with the added condition that

$$\delta^2 = o(p^2).$$

He then showed that under his added condition in Equation (11), the coverage frequency of his alleged fiducial solution,

$$[F^{-1}(\alpha; d^2_0, p), \infty),$$

has limiting value zero as $p \to \infty$, where $d^2_0$ is the observed value of variable $\delta^2$.

However, any bounded (measurable) fixed subset of $\Theta$ such as $S_p(\eta p) = \{ (\mu_1, \ldots, \mu_p) : \delta^2 = \sum_{i=1}^p \mu_i^2 \leq \eta^2 p^2 \}$, with a fixed $p$ and—because of Equation (11)—for any (small) $\eta^2$, has a prior probability equal to zero when the prior distribution is “uniform” and improper in $R^p$. With a large enough hypersquare $K_p(k)$ with a side length $2k$, in order to have simultaneously $\delta^2 / p^2 < \eta^2$ and $V(S_p(\eta p)) / V(K_p(k)) < \epsilon$, for any $\epsilon > 0$, where $V(\cdot)$ is the hypervolume measure, it is sufficient to take $k$ larger than $\eta p / [\{ \epsilon \Gamma(p/2 + 1) \}^{1/p} (\pi/4)^{-1/2}]$ (Cramér, 1946, p. 120). Therefore, Stein’s
condition in Equation (11), has a negligible effect on 100(1 − α)% improper Bayesian credible intervals for the problem. This is true for MEs as well according to Section 5.4.

Furthermore, two Monte-Carlo experiments described in the Web Appendix in the Supplementary Material confirm that, without the condition in Equation (11), the 100(1 − α)% MEs in Equation (12) are also 100(1 − α)% improper Bayesian credible intervals—although the speed of convergence of the coverage frequency decreases as \( p \) increases, as predicted by Pinkham (1966). It is remarkable that the mean of the DPE on \( \delta^2 \) is not a consistent point estimator of \( \delta^2 \). Better point estimators are considered in Evans & Shakhatreh (2014).

When comparing the results of these Monte-Carlo experiments and Pinkham’s (1966) conclusion with Stein’s result, it is clear that the convergence to zero of the coverage of the interval in Equation (12), subject to Stein’s condition in Equation (11), is due to this condition and definitely not only to fortuitous errors—as is required in order to use Gauss’s error postulate. Therefore, Stein’s alleged fiducial interval in Equation (12) cannot be a ME as used in a renewed theory of errors.

Note that Equation (1) in Section 4.2 also provides a ME for Stein’s noncentrality coefficient underestimation problem.

9.4. Marginalization is Unavoidable but Manageable

Stein’s (1959) reckless use of marginalization in his counterexample (he has more parameters than observations!) has unfortunately given the impression that marginalization is a dangerous method that can get out of control. This is a gross exaggeration. Fieller’s pivot to estimate a ratio of means and Jolicoeur’s pivot for a major axis slope are improper; whereas, the method demonstrated here to improve corresponding confidence belts is to use a joint proper pivot for the five parameters of a \( \mathcal{N}_2 \) distribution. This is followed by a marginalization process. Marginalization is also the natural method of eliminating unwanted parameters in a Bayesian approach to a problem.

Often a proper pivot for a particular parameter does not exist—as Gleser & Hwang (1987) have shown. To be able to find a proper pivot in a space of higher dimension capable of providing a DPE for that parameter by using marginalization is then invaluable—provided one is willing to adopt the Gaussian error postulate with its extended Fraser’s frequency interpretation as a guideline. A comparison of Figures 3 and 4 emphasizes that a unique sample criterion—like that of the theory of errors—has definite advantages over the long-run criterion of CI theory.

There are rare occasions when more than one pivot exists to estimate a parameter. For example, using the Cusnhy & Peebles data when assuming normality, the pivot for the five parameters of the \( \mathcal{N}_2 \) distribution provides a ME 1.580 ± 0.9622 for \( \Delta = \mu_x - \mu_y \), which is 9% longer than the usual 95% CI (or ME) 1.580 ± 0.8800 from the pivot \( (\bar{D}_i - \Delta, 9\bar{s}^2_D/\sigma^2_D) \) for the two parameters \( \Delta \) and \( \sigma^2_D \), where \( D_i = x_i - y_i \) for \( i = 1, \ldots, 10 \). This is 3% longer per marginalized real parameter. It is not difficult to choose between these two interval estimates using an efficiency criterion. A comparable situation occurs with Behrens-Fisher MEs and approximate CIs (or MEs) based on the Welch–Satterthwaite approximate pivot. When multiplying each mean of a ratio of means by a factor of 15 using the Cusnhy–Peebles data, the 95% Fieller’s CI interval for the ratio \( \lambda = 15\mu_x/15\mu_y \) is 0.321235 ± 0.027356 while the 95% ME is 0.32128 ± 0.02813. Marginalization thus produced an almost 3% (manageable) increase of the interval’s length.

10. DISCUSSION

Fisher’s criterion of mutual logical consistency of estimation statements (Fisher, 1934, 1956; Wilkinson, 1977; Barnard, 1981; Plante, 1984, 1991, 1994) has been largely ignored by statisticians except when improper CIs occur. This article provides an alternative to using improper CIs. The technical difficulties in using these CIs (tacitly denounced by Tukey,
have been incompletely addressed. The solution presented uses Gauss’s error postulate (Section 4). While coverage frequency is recognized as important, observation errors are the main cause of coverage errors, but not their only cause. Two causes of coverage errors were shown in Section 2.1. How to correct coverage errors when not only caused by observation errors was given for a noncentrality parameter of a noncentral chi-square distribution in Section 4.2; for a ratio of means in Section 7; and for the slope of a major axis in Section 8.

Gauss introduced error analysis as one of the most fertile fields of application of mathematics to natural phenomena — although the theory of errors is presently underdeveloped when compared to its proper and improper Bayesian and Neyman–Pearsonean neighbours. A review of problems which can be treated using what is essentially Gauss’s method would be an encyclopedic endeavour.

Many current problems need attention such as:

- An inverse regression elementary example is given in the Web Appendix in the Supplementary Material. Can the technique used be extended to missing value problems?
- How far can (approximately) invertible Box–Cox transformations to normality be inverted after DPEs have been found for transformed parameters?
- Moments of almost lognormal univariate or bivariate distributions could likely be treated by using approximations based on the method of Barndorff-Nielsen & Cox (1989, Sections 5.1 and 5.2). How far could the formula $E_v \approx \exp(\mu + \sigma^2/2)$ be extended?
- In circular statistics, angles are averaged by using the formula $(\sin \theta, \cos \theta)$ which in some examples were observed readily converging to bivariate normality — restricted by using the formula $\sin^2 \theta + \cos^2 \theta \leq 1$. This method can avoid the difficulties pointed out in Gleser & Hwang (1987) and in Batschelet (1981, Section 5.3). What are the limitations of this approximate method?
- Error simulation could also accelerate the simulation process if used in certain improper Bayesian methods when the prior is noninformative and the observable random variables are normally distributed — that is, when an auxiliary error model having the same formal solution as the improper Bayesian model can be introduced. (This occurs with linear pivots as is explained in the Web Appendix in the Supplementary Material.) Comparative Monte-Carlo experiments are required.

Two significant theoretical problems are:

- Estimation of the parameters of the three-dimensional and multi-dimensional normal distributions could be useful. One could use Fisher (1962) or expand, one step at a time, Proposition 1 (in Section 6.2) or use other means.
- Medical applications of physics, for instance in the Journal of Magnetic Resonance Imaging or the IEEE Transactions on Medical Imaging, make heavy use of multi-dimensional error propagation formulas. Would Tukey’s, Fisher’s, Fraser’s, Dempster and Shafer’s (for a discrete model) or Dempster’s conditioning (for a continuous model) or Hannig’s error methods be useful?

11. CONCLUSION

Joseph Bertrand added a “prophetic” remark to his translation of Gauss’s work on the theory of errors (Gauss, 1855), when stating that Gauss was aware of critical judgements about his theory, but that Gauss firmly believed that geometers would entirely adopt his ideas once his memoirs, then quite rare, would be more publicized.
APPENDIX

Proof of a statement in Section 2.1 identifying improper CIs for a noncentrality parameter when the number of degrees of freedom tends to infinity. In Section 2.1 it is stated without proof that uniformly most accurate 100(1 − α)% CIs for the noncentrality parameter δ² of a noncentral chi-square distribution can be improper with probability 1 − α at the limit if δ² = o(p^{1/2}). The following Theorem 1 clarifies this statement.

Let d² be a random variable with a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter δ², and let F(d²; δ², p) be the distribution function of d². Suppose that δ² must satisfy the condition

\[ \delta^2 = o(p^{1/2}). \quad \text{(A1)} \]

The natural 100(1 − α)% CIs of form \([\delta^2(d^2), \infty)\) are the inverse images \(F_{-1}(d^2; (0, 1 - \alpha), p)\) of the partial point function \(F(d^2; \cdot, p)\) when the first and the third arguments are fixed.

**Theorem 1 (Limiting behaviour of natural CIs for δ²).** Under the condition in Equation (A1), (i) the probability that the natural 100(1 − α)% (0 < α < 1) CIs of the form \([\delta^2, \infty)\), for the noncentrality parameter δ² of a noncentral chi-square distribution with p degrees of freedom, be equal to \([0, \infty)\) is 1 − α at the limit when \(p \rightarrow \infty\); (ii) the conditional probability that the natural 100(1 − α)% CIs cover δ², given that they are not equal to \([0, \infty)\), is zero at the limit when \(p \rightarrow \infty\).

**Proof of Theorem 1.** Without loss of generality, suppose that \(y_i^2 > 0\) for all i.

(i) Stein (1959) pointed out that the noncentral chi-square distribution with p degrees of freedom and noncentrality parameter δ² is approximated by a \(N_1(p + \delta^2, 2p + 4\delta^2)\) distribution for large p—the approximation being uniform in δ². Expression (10) in Stein (1959) for the natural CI \([\delta^2, \infty)\), when translated into the above notation (and using 1 − α instead of Stein’s α") is \(F_{-1}(d^2; (1 - \alpha), p, \infty)\). This should be replaced by the inverse image \(F_{-1}(d^2; (0, 1 - \alpha), p)\) since the partial function \(F(d^2; \cdot, p)\) has range \((0, F(d^2; 0, p)]\) (I have assumed that each \(y_i^2\) is positive.). The probability that the 100(1 − α)% CI \([\delta^2, \infty)\) is equal to \([0, \infty)\) is \(S = F(F^{-1}(1 - \alpha; 0, p); \delta^2, p)\). If \(\Phi(z_a) = 1 - \alpha\), where \(\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi}) \exp(-v^2/2)dv\), using the Cornish–Fisher expansion (Abramowitz & Stegun, 1964, 1972, p. 935), the normal approximation gives \(S \approx F(p + z_a(2p)^{1/2}; \delta^2, p)\). Thus, because of the condition in Equation (A1), \(S \approx \Phi\{(p + z_a(2p)^{1/2} - \delta^2 - p)/(2 + 4\delta^2)^{1/2}\} = \Phi\{(z_a - \delta^2)/(2p)^{1/2}\}/(1 + 2\delta^2/p^{1/2}) \rightarrow \infty\) \(\Phi(z_a) = 1 - \alpha\), as stated.

(ii) Consider the identity: \(P_{\delta^2}: \{\delta_-^2 \leq \delta^2\} = P_{\delta^2}: \{\delta_-^2 = 0\}P_{\delta^2}: \{\delta_-^2 \leq \delta^2\} = 0\}P_{\delta^2}: \{\delta_-^2 > 0\}P_{\delta^2}: \{\delta_-^2 \leq \delta^2 \} \geq 0\} + P_{\delta^2}: \{\delta_-^2 > 0\}P_{\delta^2}: \{\delta_-^2 \leq \delta^2 \} = 0\} = 1\). Since the equation \(P_{\delta^2}: \{\delta_-^2 \leq \delta^2 \} \geq 0\} + P_{\delta^2}: \{\delta_-^2 > 0\}P_{\delta^2}: \{\delta_-^2 \leq \delta^2 \} = 0\} = 1\). Because of the condition in Equation (A1) and of part (i) of the theorem, one can conclude that \(P_{\delta^2}: \{\delta_-^2 \leq \delta^2 \} \geq 0\} - \alpha)/\{P_{\delta^2}: \{\delta_-^2 > 0\} \rightarrow \infty\) \((\alpha - \alpha)/\alpha = 0\).

The proof of the statement in Section 2.1 is now complete.

**Proof of Equation (1) in Section 4.2.** Assume that \(Y = \{d^2\} = (0, \infty)\) (\(d^2 > 0\), without loss of generality), \(\Theta = \{\delta^2\} = (0, \infty)\), \(E = \{x\} = (0, 1)\), and that the distribution \(Q(\cdot)\) has a uniform density over \(E\)-events. The model

\[ d^2 = F^{-1}(e; \delta^2, p), \quad \text{(A2)} \]

where \(F^{-1}(\cdot; \delta^2, p)\) is the inverse of \(F(\cdot; \delta^2, p)\), is a “structured model” (Fraser, 1971). It is known that the family of distributions of \(d^2\) is strictly ordered according to the nondecreasing

---

The Canadian Journal of Statistics / La revue canadienne de statistique

DOI: 10.1002/cjs
likelihood ratio ordering, thus is strictly stochastically ordered. This implies that the model in Equation (A2) is additive (Plante, 1979a), i.e., \( \forall e, F^{-1}(\{e\}; \delta^2, p) \cap F^{-1}(\{e\}; \delta^2', p) = \emptyset \) if \( \delta^2 \neq \delta^2' \), where now \( F^{-1}(\cdot; \delta^2, p) \) is the inverse image set function of \( F(\cdot; \delta^2, p) \). Therefore Dempster’s upper and lower error measures (or probabilities) are equal. The “set of possible antecedents,” after observing \( d^2 = d^2_0 \neq 0 \), is \( E_{d_0^2} = F_{-1}(d^2_0; [0, \infty), p) = (0, F(d^2_0); 0, p)] \), where \( F_{-1}(d^2_0; \cdot, p) \colon \mathcal{P}([0, \infty)) \to \mathcal{P}((0, 1)) \) is the inverse image set function of the point function \( F(d^2_0; \cdot, p) \) when the first and third arguments are fixed, and where \( \mathcal{P}(\cdot) \) designates the class of all subsets of the argument. This set of possible antecedents has a positive length because of the assumption that \( d^2_0 > 0 \). The unique “cofunction” (Plante, 1979a) is the point function \( F_{-1}(d^2_0; \cdot, p) \colon E_{d_0^2} \to [0, \infty), \) where now \( F_{-1}(d^2_0; \cdot, p) \) is the inverse point function of the restricted function \( F(d^2_0; \cdot, p) \colon [0, \infty) \to E_{d_0^2} \). The following statement holds for any \( \mathcal{E} \)-event \( A \) (Plante, 1979b, Theorem 1):

\[
\{F^{-1}(e; \delta^2; p) = d^2_0 \} \iff \{\{\delta^2 \in F_{-1}(d^2_0; A \cap E_{d_0^2}, p)\} \iff \{e \in A \cap E_{d_0^2}\}.\]

By definition, the “structured measure,” i.e., the conditional error distribution, given \( E_{d_0^2} \), is \( Q(\cdot | E_{d_0^2}) \). This conditional error distribution has a uniform density on the error-interval \( (0, F(d^2_0); 0, p)] = E_{d_0^2} \). Since \( F(d^2; \delta^2, p) \) is decreasing in its second argument, when the first and third arguments are fixed, it follows that the lower 100\((1 - \alpha)\)% error bound is \( F_{-1}(d^2_0; (1 - \alpha)F(d^2_0; 0, p), p) \) as stated, where \( F_{-1}(d^2_0; \cdot, p) \) is the inverse of the function \( F(d^2_0; \cdot, p) \) defined on \( E_{d_0^2} \) when the first and the third arguments are fixed.

To check that observation errors and coverage errors are not far apart when using Equation (1) in the main article, I made an elementary Monte-Carlo experiment at the 95% level based on 100,000 iterations, using the “null” distribution \( F(d^2; 20, 18) \). The resulting coverage frequency was: 0.949.

**Proof of Proposition 1.** The joint density of \((s_x, s_y, r)\) is well known from the Wishart distribution and can be obtained from Anderson (1984), Section 4.2.2 or from Hotelling (1953). The marginal distribution of \( r \) is also well known (Fisher, 1915, Hotelling, 1953).

Using expression (31) in Andersen (1984), p. 111, obtained from the Wishart density, and using the duplication formula \( \sqrt{\pi} \Gamma(2c)/2^{2c-1} = \Gamma(c)\Gamma(c + 1/2) \) for the gamma function, the joint density of \( s_x, s_y, \) and \( r \) may be expressed as

\[
f(s_x, s_y, r; \sigma_x, \sigma_y, \rho) = \frac{n^n(1 - r^2)^{(n-3)/2}s_x^{-1}s_y^{-1}}{\pi(n-2)!(1 - \rho^2)^{n/2}\sigma_x^\alpha\sigma_y^\alpha} \times \exp \left\{ -\frac{n}{2(1 - \rho^2)} \left( \frac{s_x^2}{\sigma_x} - 2\rho r \frac{s_x s_y}{\sigma_x \sigma_y} + \frac{s_y^2}{\sigma_y} \right) \right\}. \tag{A3}
\]

Since the Jacobian is 1/2 when transforming density (A3) from \((s_x, s_y, r; \sigma_x, \sigma_y, \rho)\) to \((a, b, r; \alpha, \beta, \rho)\), the joint density of \( a, b, \) and \( r \) is,

\[
f(a, b, r; \alpha, \beta, \rho) = \frac{n^n(1 - r^2)^{(n-3)/2}a^{n-1}}{2\pi(n-2)!(1 - \rho^2)^{n/2}a^\alpha} \times \exp \left[ -\frac{n}{1 - \rho^2} \frac{a(\cosh(b - \beta) - \rho r)}{\alpha} \right]. \tag{A4}
\]
The joint distribution of $b$ and $r$ may be obtained as follows. If we set $a = [(1 - \rho^2)\alpha/n\{\cosh(b - \beta) - pr\}]w$, and then take $(w, b, r)$ as variables, the only factor of $w$ in the joint density of $(w, b, r)$, according to (A4), is $w^{n-1}\exp\{-w\}/\Gamma(n)$. The joint density of $b$ and $r$, by integrating with respect to $w$, is

$$f(b, r; \beta, \rho) = \frac{(n-1)(1-r^2)^{(n-3)/2}(1 - \rho^2)^{n/2}}{2\pi\{\cosh(b - \beta) - pr\}^n}.$$  

(A5)

Equation (3) of the main article follows from Equation (2) of the main article and from Equation (A5) above. Equation (4) in the main article follows from Equations (A4) and (A5) above.

Proof of Proposition 2. With the notation of Section 6, it is sufficient to prove that $F(r; \rho)$, $F(b \mid r; \beta, \rho)$, $F(a \mid b, r; \alpha, \beta, \rho)$, $F(\bar{x} \mid a, b, r; \mu_x, \alpha, \beta, \rho)$, and $F(\bar{y} \mid \bar{x}, a, b, r; \mu_y, \mu_x, \alpha, \beta, \rho)$ are bijections as functions of the first argument after the symbol “;” when other variables and parameters are fixed. However, from Equations (3)–(6) of the article, it follows that the distributions used for $b, a, \bar{x},$ and $\bar{y}$ are either unrestricted location parameter models or transformable to unrestricted location parameter models. Since an unrestricted location parameter model with a positive density on $R^1$ defines a proper pivot—the distribution function is strictly increasing and the graph of the family of distribution functions fills the strip $R^1 \times (0, 1)$—it is sufficient to prove that Fisher’s (1930) pivot $F(r; \rho) = e$ is proper.

Assume that $n = N - 1 \geq 3$. This ensures that the left tail of the density increases strictly and that the right tail decreases strictly if $-1 < \rho < 1$ (Cramér, 1946, Section 29.7). The nonnegative function $f(r; \rho)$ defined on $-1 \leq r \leq 1$ and $-1 - e \leq \rho < 1$ is continuous and bounded, where $e$ is an arbitrary small positive number. (In fact, $f(r; \rho)$ is continuous on $[0, 1] \times [0, 1]$ except at the two points for which $\rho = 1$.) Therefore $F(r; \rho)$ is continuous on $(-1 + e, 1 - e)$ as a function of $\rho$ for a fixed $r$. It is known that the family of distributions of $r$ has a strictly monotone (increasing) likelihood ratio (Lehmann & Romano, 2005, p. 231). Thus the family is strictly stochastically ordered. This implies that $F_{-1}(r; \cdot)$ exists if $\|r\| < 1$, is one-one, is strictly decreasing as a function defined on $[F(r; 1 - e), F(r; -1 + e)]$, and is continuous (thus integrable) since $F(r; \cdot)$ is continuous and monotone on a finite interval, where the lower index $-1$ indicates the inverse with respect to the second argument when the first argument is fixed. Let us show that $\lim_{\rho \downarrow 1} F(r; \rho) = 0$ if $r$ is fixed in $(-1, 1)$. With this condition on $r$, one may now assume that $-1 < \rho < 1$. But, $\lim_{\rho \downarrow 1} f(r; \rho) = 0$ and $f(r_1; \rho) < f(r; \rho)\forall r_1 \in (-1, r)$ hold if $\rho$ is sufficiently close to 1. It follows that $\lim_{\rho \downarrow 1} F(r; \rho) = 0$ since the interval $[-1, r]$ has a finite length. Using a similar argument, it also follows that $\lim_{\rho \uparrow 1}(1 - F(r; \rho)) = 0$. Thus $F(r; \rho)$ as a function of $\rho$—$r$ being fixed—takes all the values in the interval $(0, 1)$ so that $F_{-1}(r; \cdot) : (0, 1) \rightarrow (-1, 1)$ is a bijection if $-1 < r < 1$.

Proof of Proposition 3. The condition $t_{a/2} \geq N^{1/2}[\gamma - \lambda\bar{x}] / (s_y^2 - 2s_x s_y + s_x^2\lambda^2)^{1/2}$ is equivalent to the condition $t_{a/2}^2 = 2s_x s_y + t_{a/2}^2 \lambda + s_x^2 \lambda^2 \geq s_y^2 \gamma^2 + s_y^2 \bar{x}^2 \lambda^2 - 2s_x s_y \bar{x}^2 \gamma \lambda$ ($\bar{x} = s_x \bar{x}^2 / N^{1/2}, \bar{y} = s_y \bar{y}^2 / N^{1/2}$). Re-arranging this expression in terms of powers of $\lambda$ gives

$$\gamma = s_y^2 (\bar{x}^2 - t_{a/2}^2) \lambda^2 - 2s_x s_y (\bar{x}^2 \gamma^2 - t_{a/2}^2) \lambda + s_y^2 (\gamma^2 - t_{a/2}^2) \lambda^2 \leq 0,$$

(A6)

where $\gamma = \gamma(\lambda)$ is a parabola in the $\lambda O \gamma$ plane. The discriminant corresponding to this parabola is negative iff

$$\bar{x}^2 - 2r \bar{x} \bar{y} \gamma + \gamma^2 - (1 - r^2) t_{a/2}^2 < 0.$$  

(A7)
If \( r \neq 0 \), the points \((\bar{x}^*, \bar{y}^*)\) corresponding to Equation (A7) lie inside an ellipse with major axis \(\bar{y}^* = \text{sign}(r)\bar{x}^*\), inscribed within the square \(\max\{|\bar{x}^*|, |\bar{y}^*|\} = t_{a/2}\), and touching the square at the points \((t_{a/2}, rt_{a/2}), (−t_{a/2}, −rt_{a/2}), (rt_{a/2}, t_{a/2})\), and \(−rt_{a/2}, −t_{a/2}\). If \( r = 0 \) \((= \pm rt_{a/2}\)), then the ellipse is a circle centred at the origin. Therefore, the condition in Equation (A7) implies that \(\bar{x}^* < t_{a/2}^2\) as the parabola \(\gamma = \gamma(\lambda)\) lies entirely below the \(λ\)-axis. The inequality in Equation (A7) is the case when all values of \(λ\) are a solution of inequality in Equation (A6). Furthermore, the solution is real when \(\bar{x}^* \geq t_{a/2}^2\) — but let us consider only the case \(\bar{x}^* > t_{a/2}^2\) since the case \(\bar{x}^* = t_{a/2}^2\) occurs with probability zero. The solution of \(\gamma = 0\), when it exists, and if \(\bar{x}^* \neq t_{a/2}^2\), where \(γ\) is defined in Equation (A6), is determined by the values

\[
\lambda = \frac{s_y}{s_x(\bar{x}^* - t_{a/2}^2)} \left\{ \frac{\bar{x}^*\bar{y}^* - rt_{a/2}^2}{t_{a/2}} \pm t_{a/2} \sqrt{\bar{x}^2 - 2\bar{x}^*\bar{y}^* r + \bar{y}^2 - (1 - r^2)t_{a/2}^2} \right\}. \tag{A8}
\]

When the divisor \(s_x(\bar{x}^* - t_{a/2}^2)\) changes from being positive to negative as \(\bar{x}^*\) decreases, the set of \(\lambda\)s satisfying Equation (A8) changes from an interval to become the complement of an interval—except for the two points for which \(\bar{y}^* = \pm rt_{a/2}\) where it becomes \((−\infty, \infty)\). It is not necessary to consider points on the lines \(\bar{x}^* = \pm t_{a/2}\) and points on the boundary of the ellipse in Equation (A7) since they occur with probability zero.

**ACKNOWLEDGEMENTS**

Several people contributed to this article: the nonmathematical parts of the article were written in collaboration with my wife Michael Plante. The main reviewer’s comments led to the inclusion of several ideas not included in my original manuscript; the associate editor’s remarks encouraged me to stress important points and to drop less essentially relevant ones; the main reviewer’s report of a previous rejected submission was helpful in preparing a new submission; a suggestion of Victoria Slonosky was pivotal to the research project; Michaela Plante provided technical support; conversations with Michaela Plante, Alexander Somma, and Alexandre Plante were helpful. CJS editor Grace Yi and CJS assistant Julie Falkner were gently patient with my successive revisions of the article and a CJS copyeditor made impressively detailed comments.

**BIBLIOGRAPHY**

Abramowitz, M. & Stegun, I. (1964, 1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. John Wiley & Sons, New York.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, 34, 122–148.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley & Sons, New York.

Balch, M. S. (2012). Mathematical foundations for a theory of confidence structures. *International Journal of Approximate Reasoning*, 53, 1003–1019.

Barnard, G. A. (1977). Pivotal inference and the Bayesian controversy. *Bulletin of the International Statistical Institute*, 47, 543–551.

Barnard, G. A. (1981). *A Coherent View of Statistical Inference*. Department of Statistics & Actuarial Science, University of Waterloo, Canada. Technical Report Series.

Barndorff-Nielsen, O. (1991). An ancillary invariant modification of the signed log-likelihood ratio. *Scandinavian Journal of Statistics*, 18, 341–352.

Barndorff-Nielsen, O. & Cox, D. R. (1989). *Asymptotic Techniques*. Chapman and Hall, London and New York.

Bartholomew, D. J. (1967). Hypothesis testing when the sample size is treated as a random variable. *The Journal of the Royal Statistical Society Series B (Methodology)*, 29, 53–82.
Bartholomew, D. J. (1971). Comment on A. Plante’s “Counter-Examples and Likelihood”. In V. P. Godambe & D. A. Sprott (Eds.) Foundations of Statistical Inference, Holt, Rinehart and Winston of Canada Ltd., Montreal and Toronto, 367–368.

Batschelet, E. (1981). Circular Statistics in Biology. Academic Press, London.

Bebu, I. & Mathew, T. (2008). Comparing the means and variances of a bivariate lognormal distribution. Statistics in Medicine, 27, 2684–2696.

Bennett, B. M. (1957). On the performance characteristic of certain methods of determining confidence limits. Sankhyā: The Indian Journal of Statistics (1933–1960), 15, 1–12.

Bennett, B. M. & Neyman, J. (1948). The performance characteristic of certain methods for obtaining confidence intervals (Abstract). The Annals of Mathematical Statistics, 19, 116.

Beran, R. J. (1971). On distribution-free statistical inference with upper and lower probabilities. The Annals of Mathematical Statistics, 31, 1208–1211.

Berger, J. O. & Sun, D. (2008). Objective priors for the bivariate normal model. The Annals of Statistics, 36, 963–982.

Brown, L. (1967). The conditional level of Student’s $t$ test. The Annals of Mathematical Statistics, 38, 1068–1071.

Buehler, R. J. (1959). Validity criteria for statistical inference. The Annals of Mathematical Statistics, 30, 845–863.

Buehler, R. J. (1980). Fiducial inference. In S. E. Fienberg & D. V. Hinkley (Eds.) R. A. Fisher: An Appreciation, Springer–Verlag, New York.

Buehler, R. J. & Feddersen, A. P. (1963). Note on a conditional property of Student’s $t$. The Annals of Mathematical Statistics, 34, 1098–1100.

Bunke, H. (1975). Statistical inference: Fiducial and structural vs. likelihood. Mathematische Operationsforschung und Statistik, 6, 667–676.

Bunke, O. (1976). Conditional probability in incompletely specified stochastic equations and statistical inference. Mathematische Operationsforschung und Statistik, 7, 673–678.

Casella, G. (2002). Comment on M. Mandelkern’s “Setting Confidence Intervals for Bounded Parameters”. Statistical Science, 17, 158–160.

Ciesiulewicz, J. & Hannig, J. (2012). Generalized fiducial inference for normal linear mixed model. The Annals of Statistics, 40, 2102–2127.

Cox, D. R. (1958). Some problems connected with statistical inference. The Annals of Mathematical Statistics, 29, 357–372.

Cox, D. R. & Hinkley, D. V. (1974). Theoretical Statistics. Chapman and Hall, London.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ.

Dawid, A. P. & Stone, M. (1982). The functional-model basis of fiducial inference. The Annals of Statistics, 10, 1054–1067.

Dawid, A. P., Stone, M., & Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. Journal of the Royal Statistical Society Series B (Methodology), 35, 189–233.

Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics, 36, 325–339.

Dempster, A. P. (2008). The Dempster–Shafer calculus for statisticians. International Journal of Approximate Reasoning, 48, 365–377.

Doob, J. L. (1935). The limiting distributions of certain statistics. The Annals of Mathematical Statistics, 6, 160–169.

Edlefsen, P. T., Liu, C., & Dempster, A. P. (2009). Estimating limits from Poisson counting data using Dempster–Shafer analysis. The Annals of Applied Statistics, 3, 764–790.

Evans, M. & Shakhatreh, M. (2014). Consistency of Bayesian estimates for the sum of squared normal means with a normal prior. Sankhyā: The Indian Journal of Statistics, 79-A, 25–47. (Pt 1).

Feiller, E. C. (1954). Some problems in interval estimation. Journal of the Royal Statistical Society Series B (Methodology), 16, 175–185.

Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. Messenger of Mathematics, 41, 155–160.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an infinitely large population. Biometrika, 10, 507–521.

Fisher, R. A. (1925). Statistical Methods for Research Workers, 1st to 14th ed., Oliver and Boyd, Edinburgh.
Fisher, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society*, 26, 528–535.

Fisher, R. A. (1934). Contribution to the discussion of Neyman’s “On the two Different Aspects of the representative Method: The Method of Stratified Sampling and the Method of Purposive Selection”. *Journal of the Royal Statistical Society*, 97, 614–619.

Fisher, R. A. (1935). Dispersion on a sphere. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 217, 295–305.

Fisher, R. A. (1953). The fiducial argument in statistical inference. *Annals of Eugenics*, 61, 391–398.

Fisher, R. A. (1955). Fisher to John Tukey: July 29, 1955. In J. H. Bennett (Ed.) *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*, University Press, Oxford.

Fisher, R. A. (1956). On a test of significance in Pearson’s *Biometrika Tables* (no. 11). *Journal of the Royal Statistical Society Series B* (Methodology), 18, 56–60.

Fisher, R. A. (1959). *Statistical Methods and Scientific Inference*, 2nd ed. (1st ed., 1956; 3rd ed., 1973, posthumous), Oliver and Boyd, Edinburgh.

Fisher, R. A. (1960). On some extension of Bayesian inference proposed by Mr Lindley. *Journal of the Royal Statistical Society Series B* (Methodology), 22, 299–301.

Fisher, R. A. (1962). The simultaneous distribution of correlation coefficients. *Sankhyā: The Indian Journal of Statistics* (1933–1960), 24, 1–8.

Fraser, D. A. S. (1961). The fiducial method and invariance. *Biometrika*, 48, 261–280.

Fraser, D. A. S. (1968). *The Structure of Inference*. John Wiley & Sons, New York.

Fraser, D. A. S. (1971). Events, information processing, and the structured model (with discussion). In V. P. Godambe & D. A. Sprott (Eds.) *Foundations of Statistical Inference*, Holt, Rinehart and Winston of Canada Ltd., Toronto and Montreal, 32–55.

Fraser, D. A. S. (1979). *Inference & Linear Models*. McGraw-Hill, New York.

Fraser, D. A. S. (1991). Statistical inference: Likelihood to significance. *Journal of the American Statistical Association*, 86, 258–265.

Fraser, D. A. S. (2004). Inference for bounded parameters. *Physical Review D*, 69, 033002.

Fraser, D. A. S., Reid, N., & Wong, A. (2004). Regression analysis, nonlinear or nonnormal: Simple and accurate *p*-values from likelihood analysis. *Journal of the American Statistical Association*, 34, 1286–1294.

Fréry, N., Huel, G., Leroy, M., Moreau, T., Savard, R., Blot, P., & Lellouch, J. (1992). Vitamin B12 among parturients and their newborns and its relationship with birthweight. *European Journal Obstetrics & Gynecology and Reproductive Biology*, 45, 155–163.

Gauss, C. F. (1809). *Theoria Motus Corporum Celestium*. Perthes et Besser, Hamburg (Google Book).

Gauss, C. F. (1821). *Theoria combinationis observationum erroribus minimis obnoxiae*: Pars prior. *Göttingische gelehrte Anzeigen*, 33, 321–327.

Gauss, C. F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*: Pars posterior. *Göttingische gelehrte Anzeigen*, 32, 313–318.

Gauss, C. F. (1855). *Méthode des Moindres Carrés: Mémoires sur la Combinaison des Observations, (traduits par Joseph Bertrand)*. Mallet-Bachelier, Paris (Google Book).

Gleser, L. J. & Hwang, J. T. (1987). The nonexistence of 100(1 − α)% confidence sets of finite expected diameter in error-in-variables and related models. *The Annals of Statistics*, 15, 1351–1362.

Gleser, L. J. & Hwang, J. T. (1987). The nonexistence of 100(1 − α)% confidence sets of finite expected diameter in error-in-variables and related models. *The Annals of Statistics*, 15, 1351–1362.

Hannig, J. (2002). Comment on M. Mandelkern’s “Setting Confidence Intervals for Bounded Parameters”. *Statistical Science*, 17, 161–163.

Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, 19, 491–544.

Hannig, J. (2013). Generalized fiducial inference via discretization. *Statistica Sinica*, 23, 489–514.

Hannig, J., Iyer, H., Lai, R. C. S., & Lee, T. C. M. (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association*, 111, 1346–1361.

Hannig, J., Wang, C. M., & Iyer, H. K. (2003). Uncertainty calculation for the ratio of dependent measurements. *Metrologia*, 40, 177–183.

Hannig, J. & Xi, M. (2012). A note on Dempster–Schafer recombination of confidence distributions. *Electronic Journal of Statistics*, 6, 1943–1966.
Plante, A. (1984). A reexamination of Stein’s antifiducial example. *The Canadian Journal of Statistics*, 12, 135–141.

Plante, A. (1987). A decision-likelihood solution to the problem of comparing two simple hypotheses. In B. MacNeill & G. Umphrey (Eds.) *Advances in the Statistical Sciences: Foundations of Statistical Inference*, Springer, New York, 51–61.

Plante, A. (1991). An inclusion-consistent solution to the problem of absurd confidence statements: 1. Consistent exact confidence-interval estimation. *The Canadian Journal of Statistics*, 19, 389–397.

Plante, A. (1994). An inclusion-consistent solution to the problem of absurd confidence statements: 2. Consistent non-exact confidence-interval estimation. *The Canadian Journal of Statistics*, 22, 1–13.

Quenouille, M. H. (1958). *Fundamentals of Statistical Reasoning*. Charles Griffin, London.

Reid, N. & Cox, D. R. (2015). On some principles of statistical inference. *International Statistical Review*, 83, 293–308.

Scheffé, H. (1970). Multiple testing versus multiple estimation. Improper confidence sets. Estimation of directions and ratios. *The Annals of Mathematical Statistics*, 41, 1–29.

Schweder, T. & Hjort, N. L. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics*, 29, 309–332.

Schweder, T. & Hjort, N. L. (2016). *Confidence, Likelihood, Probability*. Cambridge University Press, Cambridge, UK.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton and London.

Sherman, M., Maity, A., & Wang, S. (2011). Inferences for the ratio: Fieller’s interval, log ratio, and large sample based confidence intervals. *Advances in Statistical Analysis*, 95, 313–323.

Singh, K., Xie, M., & Strawderman, W. E. (2005). Combining Information from Independent Sources through Confidence Distributions. *The Annals of Statistics*, 33, 159–183.

Stein, C. M. (1959). An example of wide discrepancy between confidence intervals and fiducial intervals. *The Annals of Mathematical Statistics*, 30, 877–881.

Tsao, A. & Hwang, J. T. (1998). Improved confidence estimators for Fieller’s confidence sets. *The Canadian Journal of Statistics*, 26, 299–310.

Tsui, K. W. & Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84, 602–607.

Tukey, J. W. (1957). *The Propagation of Errors, Fluctuations and Tolerances: Basic Generalized Formulas*, Technical Report No. 10 Part 1-10, Statistical Techniques Research Group, Princeton (US ARMY Unclassified AD 0155082, Google Scholar).

Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. *The Annals of Mathematical Statistics*, 10, 299–326.

Warton, D. I., Wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line fitting methods for allometry. *Biological Reviews*, 81, 259–291.

Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88, 899–905.

Weerahandi, S. (2013). *Exact Statistical Methods for Data Analysis*. Springer Science + Business Media Inc., New York.

Wilkinson, G. N. (1977). On resolving the controversy in statistical inference (with discussion). *The Journal of the Royal Statistical Society Series B (Methodology)*, 39, 119–171.

Xie, M. & Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *International Statistical Review*, 81, 3–39.

Yates, F. (1960). *Sampling Methods for Censuses & Surveys*. Charles Griffin & Company Limited, London.

Received 21 December 2018
Accepted 21 May 2020

DOI: 10.1002/cjs