ANOTHER LOOK AT CONFIDENCE INTERVALS

FOR THE NONCENTRAL *t* DISTRIBUTION

Bruno Lecoutre

C.N.R.S. and Université de Rouen, France

Bruno Lecoutre

ERIS, Laboratoire de Mathématiques Raphaël Salem,

UMR 6085, C.N.R.S. et Université de Rouen

Avenue de l'Université, BP 12, 76801 Saint-Etienne-du-Rouvray (France)

E-mail: bruno.lecoutre@univ-rouen.fr.

Internet: http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris

June 2006

Abstract

An alternative approach to the computation of confidence intervals for the noncentrality parameter of the Noncentral *t* distribution is proposed. It involves the percent points of a statistical distribution. This conceptual improvement renders the technical process for deriving the limits more comprehensible. Accurate approximations can be derived and easily used.

Key words or phrases: Confidence intervals; Noncentral *t* distribution; Lambda-prime distribution; Bayesian inference.

Introduction

In spite of several recent presentations (see especially, Fidler & Thompson, 2001; Bird, 2002), many potential users, as well as statistical instructors, continue to think that computing or teaching confidence intervals for the noncentrality parameter of the Noncentral *t* distribution are very complex tasks. One of the pre-eminent conceptual difficulties is the lack of explicit formula. Although the considerable advances in computing techniques are supposed to render the task easy, they do not solve the conceptual difficulties.

The latter state is all the more deceptive in that when the number of degrees of freedom is large enough so that the Normal approximation holds the solution is very simple: the confidence limits are given by the percent points of a Normal distribution, as for the familiar case of an unstandardized difference between means. Thus, it can be expected that in the general case the limits would be also computed as the percent points of a statistical distribution. Unfortunately, this is not the case with the usual presentations.

Moreover, warnings about the accuracy of some computer programs of the Noncentral *t* distribution (typically, "the Noncentral *t* algorithm fails for large sample size or effect size") cast doubt on some numerical results. Consequently, there remains the need for accurate approximations, that are not currently easily available. Even when an "exact" computation is wanted, it needs an iterative algorithm, for which an accurate approximation constitutes a good starting point.

In this paper, I propose an alternative approach that results in computing the confidence limits as the percent points of a statistical distribution as in the most familiar situations. An interesting consequence of this conceptual improvement is that standard techniques to approximate statistical distributions can be used in order to find easy to use very accurate approximations. In conclusion, I will briefly examine the question of the justification and interpretation of confidence intervals.

Considerations and discussions – how and when use them? – about the use of confidence intervals for the Noncentral *t* distribution can be found elsewhere. So the present paper is not methodological. In this perspective, it will be sufficient, with no loss of generality, to consider the elementary case of the inference about a standardized difference between two means.

Computing confidence intervals from the Noncentral *t* distribution

When comparing two means, the *t* test statistic is the ratio $(\overline{Y}_1 - \overline{Y}_2)/E$ of the two statistics, $\overline{Y}_1 - \overline{Y}_2$ that is an estimate of the population difference $\mu_1 - \mu_2$ and the standard error E of that estimate (see e.g., Fidler & Thomson, 2001, p. 587). In other words, E is an estimate of the standard deviation ε of the sampling distribution for $\overline{Y}_1 - \overline{Y}_2$. For instance, in the particular case of two independent groups, assuming a common variance σ^2, we have $\varepsilon = \sigma \sqrt{1/n_1 + 1/n_2}$. The sampling distribution of the ratio $(\overline{Y}_1 - \overline{Y}_2)/E$ is a Noncentral *t* distribution with *df* degrees of freedom and a noncentrality parameter λ, equal to $(\mu_1 - \mu_2)/\varepsilon$. This distribution is usually written $t'_{df}(\lambda)$. The noncentrality parameter is termed λ, as in Algina and Keselman (2003), in order to avoid confusion with the population effect size.

Formally, the Noncentral *t* distribution is the noncentrality parameter λ plus the standard Normal z distribution, all divided by the square root of the usual Chi-square distribution divided by the degrees of freedom (see e.g., Fidler & Thomson, 2001, p. 589):

$$t'_{df}(\lambda) = (\lambda + z)/\sqrt{\chi^2_{df}/df} \; .$$

The traditional approach for finding the lower (for instance) limit λ_L of the noncentrality parameter λ uses the probability p_λ that $t'_{df}(\lambda)$ exceeds the value t_{CALC} observed in the data in hand:

$$p_\lambda = \Pr\!\left(t'_{df}(\lambda) > t_{CALC}\right).$$

Then we must vary the λ value in order to find, by successive approximations, the particular value λ_L such that $p_{\lambda_L} = \alpha/2$:

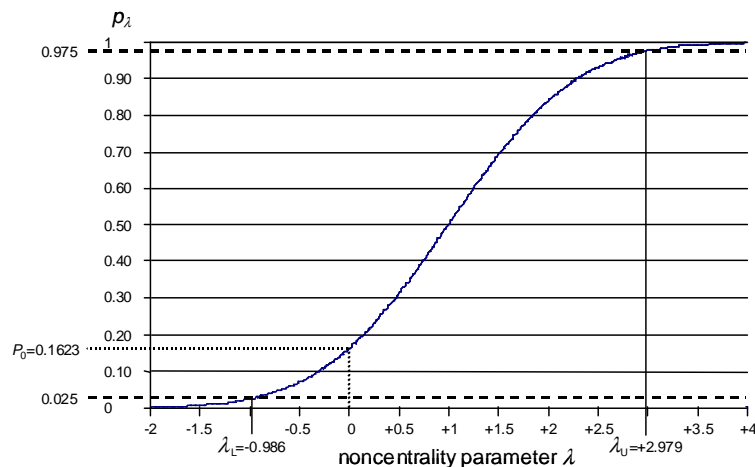$$p_\lambda = \Pr\!\left(t'_{df}(\lambda) > t_{CALC}\right) = \alpha/2. \qquad\qquad (1)$$

The conceptual difficulties come from the fact that finding the limit λ_L involves as many different distributions as considered λ values. A practical consequence is that it is a highly difficult task to derive accurate approximations.

An alternative approach: computing confidence intervals as percent points of the Lambda-prime distribution

An alternative solution consists in computing the confidence limits for the noncentrality parameter as percent points of a statistical distribution. When df is large enough so that the normal approximation holds, λ_L is simply the $100\alpha/2$ percent point of the standardized Normal distribution with mean t_{CALC}. This can be generalized by introducing an appropriate statistical distribution. Even if it has not been made explicit in the usual presentations, this distribution is in fact not unfamiliar (without speaking of the fiducial and Bayesian presentations discussed in the conclusion).

Indeed, it is usual to plot p_λ (or its complement $1-p_\lambda$) as a function of λ. An illustration is given in Figure 1 for $t_{CALC} = +1.0076$ with $df = 22$ (hence a p-value $p = 0.3246$, two-sided), which corresponds to the two-group "A" way data example given by Fidler & Thomson (2001, p. 586). The p_λ value increasingly varies from zero (when λ tends to $-\infty$) to one (when λ tends to $+\infty$), so that the corresponding curve is nothing else than the *cumulative distribution function* of a probability distribution. Such a graphical representation is commonly proposed to get a graphical solution for the confidence limits (see, for instance, Steiger and Fouladi, 1997, pp. 240), but the proponents fail to recognize that, in doing this, they implicitly define the confidence limits as the percent points of this probability distribution.

Figure 1 - Plot of p_λ as a function of λ for $t_{CALC} = +1.0076$ and $df = 22$ and graphical solution for the 95% confidence interval for λ. The curve is the cumulative distribution function of the $\Lambda'_{22}(+1.0076)$ distribution.

As for the Noncentral t, this distribution can be easily defined from the Normal and Chi-square distributions, but the result has not been so popularized. We need only to remark that $(\overline{Y}_1 - \overline{Y}_2)/E > t_{CALC}$ can be equivalently written as

$$\overline{Y}_1 - \overline{Y}_2 - t_{CALC}E > 0.$$

Consequently, p_λ is the probability that $\overline{Y}_1 - \overline{Y}_2 - t_{CALC}E$ exceeds zero.

The sampling distribution of $\overline{Y}_1 - \overline{Y}_2 - t_{CALC}E$ can be formally defined from independent standard Normal and Chi-square distributions as:

$$\varepsilon(\lambda + z - t_{CALC}\sqrt{\chi^2_{df}/df}).$$

so that

$$p_\lambda = \Pr(\overline{Y}_1 - \overline{Y}_2 - t_{CALC}E > 0) = \Pr(-z + t_{CALC}\sqrt{\chi^2_{df}/df} < \lambda) = \Pr(z + t_{CALC}\sqrt{\chi^2_{df}/df} < \lambda),$$

because the Normal distribution is symmetric around zero.

Thus, p_λ can be computed from the distribution characterized by $z + t_{CALC}\sqrt{\chi^2_{df}/df}$. This distribution, which was considered (with no name) by Fisher (1990/1973, pp. 126-127) in the fiducial framework, was called *Lambda-prime* in Lecoutre (1999). It is also a "noncentral distribution", again with *df* degrees of freedom, but with noncentrality t_{CALC}. Formally:

$$\Lambda'_{df}(t_{CALC}) = z + t_{CALC}\sqrt{\chi^2_{df}/df}.$$

Consequently, we can inverse in some sense the problem in (1) and compute p_λ as the probability that the *Lambda-prime* distribution with noncentrality t_{CALC} is smaller than λ:

$$p_\lambda = \Pr(\Lambda'_{df}(t_{CALC}) < \lambda).$$

So the curve in Figure 1 is the cumulative distribution function of the *Lambda-prime* distribution with 22 degrees of freedom and noncentrality +1.0076.

In order to find the wanted limit, we must now solve

$$p_\lambda = \Pr(\Lambda'_{df}(t_{CALC}) < \lambda_L) = \alpha/2. \qquad (2)$$

(2) is technically equivalent to (1) and requires in fact a similar iterative process, but it has a decisive conceptual advantage. Indeed, it involves a *unique* distribution, so that λ_L is the $100\alpha/2$ percent point of the $\Lambda'_{df}(t_{CALC})$ distribution. In the same way, the upper limit λ_U is its $100(1-\alpha/2)$ percent point. For instance, in Figure 1, the limits $\lambda_L = -0.986$ and $\lambda_U = +2.979$ of the 95% confidence interval are respectively the 2.5 and 97.5 percent points of the $\Lambda'_{22}(+1.0076)$ distribution.

Note again that the statistic $\overline{Y}_1 - \overline{Y}_2 - t_{CALC}E$ should not be regarded as less natural than the *t* test statistic. Indeed, it is similar to the familiar limits $\overline{Y}_1 - \overline{Y}_2 \pm t_{1-\alpha/2}E$ of the $100(1-\alpha)\%$ confidence interval for a raw difference. I will return to this analogy in the conclusion.

Approximations of the 100π percent point of the $\Lambda'_{df}(t)$ distribution

Beyond its conceptual simplification, the alternative approach allows to derive accurate approximations. In this Section, in order to simplify the notations, t_{CALC} will be written *t*.

Numerical example

Let us consider Bird's first example (Bird, 2002, p. 206), also considered in Algina and Keselman (2003). There were three independent groups of size 30 each, with means $\overline{Y}_1 = 22.467$, $\overline{Y}_2 = 24.933$, and $\overline{Y}_3 = 32.000$ and within group standard deviation 7.435. Bird reported the 97.5% confidence intervals of two standardized contrasts $(\overline{Y}_1 + \overline{Y}_2)/2 - \overline{Y}_3$ and $\overline{Y}_1 - \overline{Y}_2$. The computations for the first contrast will be detailed to illustrate the approximation methods. For this contrast, the *t* test statistic is $t = -4.9924$ (*df*=87). The exact confidence interval of λ is:

$$[-7.3766, -2.5844].$$

It can be computed by the usual method based on the Noncentral *t* distribution using the available programs (for instance the "Noncentral Distribution Calculator" of Steiger, 2004), or alternatively as the 1.25 and 98.75 percent points of the Lambda-prime distribution with 87 degrees of freedom and eccentricity -4.9924.

Three approximation methods will be considered. The results for the two contrasts of interest are presented in Table 1. The limits for the standardized contrast $(\overline{Y}_1 + \overline{Y}_2)/2 - \overline{Y}_3$ in Table 3 are obtained by multiplying the limits for λ by the appropriate constant (0.223607) referred as SE in

Bird's table, page 208. Note that this constant can be simply computed as the ratio of the observed standardized contrast to the t test value: -1.1163/-4.9924 = 0.2236.

Table 1. Bird's example: comparison of the three approximation methods

Contrast	Value	t	97.5%CI	Exact	Approximation		
					Bird	Normal	Chi-square
$(\overline{Y}_1+\overline{Y}_2)/2-\overline{Y}_3$	-1.1163	-4.9924	LowerLimit	-1.6495	-1.6264	-1.6489	-1.6495
			UpperLimit	-0.5779	-0.6063	-0.5773	-0.5779
$\overline{Y}_1-\overline{Y}_2$	-0.3318	-1.2849	LowerLimit	-0.9123	-0.9207	-0.9123	-0.9123[*]
			UpperLimit	+0.2506	+0.2572	+0.2506	+0.2506[*]

[*]Normal Approximation

Three approximation methods

Bird's Approximation

Reconsidered in the new approach, the Bird (2002, p. 203) approximation of the 100π percent point of the Lambda-prime consists in adding t and the 100π percent point of the standard (central) t distribution with the same degrees of freedom:

$$\Lambda'_{df,\pi}(t) \approx t + t_{df,\pi}.$$

We get the approximate $100(1-\alpha)$% confidence interval for λ:

$$t \pm t_{df,1-\alpha/2},$$

hence here for $t_{87,0.9875} = +2.2809$ the approximate confidence interval:

$$-4.9924 \pm 2.2809 \rightarrow [-7.2733,-2.7116].$$

Algina and Keseleman (2003) found that the accuracy of this approximation does vary with the magnitude of the parameter, which can be verified in Table 1.

A simple normal approximation

The Lambda-prime distribution is generally asymmetric. However, when $t = 0$ it reduces to the standard Normal distribution, and when df is large it tends to the $N(t,1)$ distribution. So we can expect that a Normal approximation with the same mean and variance is appropriate, at least for small t and for large df.

The $\Lambda'_{df}(t)$ distribution has mean

$$M = kt \quad \text{where} \quad k = \sqrt{\frac{2}{df}}\,\frac{\Gamma(\frac{df+1}{2})}{\Gamma(\frac{df}{2})}\,,$$

and variance

$$V = 1+t^2(1-k^2) = 1 + t^2 - M^2,$$

so that it can be approximated by the $N(M,V)$ distribution:

$$\Lambda'_{df,\pi}(t) \approx kt + z_\pi\sqrt{1+t^2(1-k^2)}.$$

We get the approximate $100(1-\alpha)$% confidence interval for λ:

$$kt \pm z_{1-\alpha/2}\sqrt{1+t^2(1-k^2)}\,.$$

In order to find k, we can compute its logarithm:

$$\log(k) = (\log(2) - \log(df))/2 + \text{logGamma}((df+1)/2) - \text{logGamma}(df/2),$$

and then take the exponential of $\log(k)$. "logGamma(x)" is the logarithm of the Gamma function $\Gamma(x)$, that generalizes factorials to numbers beyond the integers. It is standard and for instance available in Excel.

k can also be computed using the series expansion (Johnson & Welch, 1939):

$$k = 1 - \frac{1}{4df} + \frac{1}{32df^2} + \frac{5}{128df^3} - \frac{21}{2048df^4} - \frac{399}{8192df^5} + \frac{869}{65536df^6} + \cdots.$$

Alternatively, Table 2 can be used for finding the wanted value.

Table 2 - k values for df ranking from 1 to 100. For $k > 100$ the approximation $k \approx 1-1/(4df)$ gives an error less than 10^{-5}.

df	k	df	k	df	k	df	k	df	k
1	0.797885	2	0.886227	3	0.921318	4	0.939986	5	0.951533
6	0.959369	7	0.965030	8	0.969311	9	0.972659	10	0.975350
11	0.977559	12	0.979406	13	0.980971	14	0.982316	15	0.983484
16	0.984506	17	0.985410	18	0.986214	19	0.986934	20	0.987583
21	0.988170	22	0.988705	23	0.989193	24	0.989640	25	0.990052
26	0.990433	27	0.990786	28	0.991113	29	0.991418	30	0.991703
31	0.991969	32	0.992219	33	0.992454	34	0.992675	35	0.992884
36	0.993080	37	0.993267	38	0.993443	39	0.993611	40	0.993770
41	0.993922	42	0.994066	43	0.994203	44	0.994335	45	0.994460
46	0.994580	47	0.994695	48	0.994806	49	0.994911	50	0.995013
51	0.995110	52	0.995204	53	0.995294	54	0.995381	55	0.995465
56	0.995546	57	0.995624	58	0.995699	59	0.995772	60	0.995842
61	0.995910	62	0.995976	63	0.996040	64	0.996102	65	0.996161
66	0.996219	67	0.996276	68	0.996330	69	0.996383	70	0.996435
71	0.996485	72	0.996534	73	0.996581	74	0.996627	75	0.996672
76	0.996716	77	0.996759	78	0.996800	79	0.996841	80	0.996880
81	0.996918	82	0.996956	83	0.996993	84	0.997028	85	0.997063
86	0.997097	87	0.997131	88	0.997163	89	0.997195	90	0.997226
91	0.997257	92	0.997286	93	0.997315	94	0.997344	95	0.997372
96	0.997399	97	0.997426	98	0.997452	99	0.997478	100	0.997503

In Bird's example, we find $k = 0.997131$, $M = -4.9781$, $V = 1.1428$, hence for $z_{0.9875} = +2.2414$ the approximate confidence interval for λ:

$$-4.9781 \pm 2.2414 \sqrt{1.1428} \rightarrow [-7.3742, -2.5820]$$

that is close to the exact interval.

A Chi-square approximation

For large t values, one can find again better approximations that take into account the asymetry of the distribution. This needs to consider the third central moment, that can be deduced from the mean:

$$W = 2k^2 - \frac{2df-1}{df} kt^3 = 2M^3 - \frac{2df-1}{df} t^2 M.$$

We can then compute the *skewness* of the distribution as the ratio of W to the third power of the square root of the variance V (i.e. $W/V^{(3/2)}$). The skewness is a measure of the degree of asymmetry of the distribution. When it is small, we can use the Normal approximation $N(M,V)$ above. For practical applications, I found empirically that a more sophisticated approximation is not necessary when the skewness is smaller than 0.001.

Otherwise, we can use the following Chi-square approximation that fits also the skewness. It involves again reasonably simple computations.

Let

$$c = \frac{W}{4V} , q = \frac{V}{2c^2} \text{ and } a = M\text{-}qc .$$

Then the approximation is given by percent points of the Chi-square distribution with q degrees of freedom:

$$\Lambda'_{df,\pi}(t) \approx a + c \chi^2_{q,\pi} \quad \text{if } c > 0,$$

$$\Lambda'_{df,\pi}(t) \approx a + c \chi^2_{q,1-\pi} \quad \text{if } c < 0.$$

If $t > 0$ (which is equivalent to $c > 0$), we get the approximate $100(1-\alpha)\%$ confidence interval for λ:

$$[a + c \chi^2_{q,\alpha/2} , a + c \chi^2_{q,1-\alpha/2}].$$

If $t < 0$, the limits are exchanged.

In practice q is generally very large and the Wilson and Hilferty (1931) approximation can be used (this is needed if your computer program does not work for high degrees of freedom values):

$$\chi^2_{q,\pi} \approx q \left(\sqrt{\frac{2}{9q}} z_\pi + 1 - \frac{2}{9q} \right)^3$$

Some programs for the Chi-square distribution accepts only integer degrees of freedom. In this case, tthe 100π percent point of the Gamma distribution with parameter $q/2$ can be used alternatively:

$$\chi^2_{q,\pi} = 2\text{Gamma}_\pi(q/2).$$

In Bird's example, we find $W = -0.0041$, $c = -0.0009016$, $q = 702948.01$ and $a = 628.7998$, hence the approximate confidence interval (computations have been performed with the maximum number of decimals for intermediate values):

[628.7998 -0.0009016×705608.34 , 628.7998 -0.0009016×700293.06] → [-7.3766,-2.5844],

where

$$\chi^2_{702948.01,0.0125} = 700293.06 \text{ and } \chi^2_{702948.01,0.0975} = 705608.34$$

are computed using the Wilson-Hilferty approximation.

This interval coincides with the exact interval with four decimal place accuracy.

A comparison of the three methods

Table 3 gives a more systematic comparison of the three approximation methods. The exact probability levels associated with the different approximations of the 100π percent point of the $\Lambda'_{df}(t)$ distribution are reported for $100\pi = 2.5$ and $100\pi = 97.5$ (which gives the limits of the 95% confidence interval), and for $100\pi = 0.5$ and $100\pi = 99.5$ (which gives the limits of the 99% confidence interval). In the two cases, results are given for 10 and 50 degrees of freedom. They are reported only for positive values of t; the results for negative values can be deduced by symmetry.

Bird's approximation is very inaccurate for small df or large t and can hardly be recommended. By contrast, the simple Normal approximation works very well. The Chi-square approximation is quasi exact for most practical applications.

Table 3. Exact probability levels associated with the three approximations of the 100π percent point of the $\Lambda'_{df}(t)$ distribution

2.50% and 97.50% percent points

	Bird's approximation				Normal approximation				Chi square approximation			
	df = 10		df = 50		df = 10		df = 50		df = 10		df = 50	
t	2.50	97.50	2.50	97.50	2.50	97.50	2.50	97.50	2.50	97.50	2.50	97.50
0	1.29	98.71	2.23	97.77	2.50	97.50	2.50	97.50	2.50*	97.50*	2.50*	97.50*
0.25	1.33	98.72	2.24	97.77	2.50	97.50	2.50	97.50	2.50*	97.50*	2.50*	97.50*
0.50	1.38	98.70	2.26	97.77	2.50	97.50	2.50	97.50	2.50*	97.50*	2.50*	97.50*
0.75	1.46	98.66	2.28	97.76	2.50	97.50	2.50	97.50	2.50	97.50	2.50*	97.50*
1.00	1.57	98.60	2.31	97.74	2.49	97.49	2.50	97.50	2.50	97.50	2.50*	97.50*
1.50	1.86	98.41	2.39	97.69	2.48	97.48	2.50	97.50	2.50	97.50	2.50*	97.50*
2.00	2.27	98.10	2.50	97.61	2.46	97.46	2.50	97.50	2.50	97.50	2.50*	97.50*
2.50	2.81	97.68	2.63	97.50	2.42	97.43	2.50	97.50	2.50	97.50	2.50	97.50
3.00	3.50	97.14	2.80	97.37	2.39	97.39	2.49	97.49	2.50	97.50	2.50	97.50
4.00	5.31	95.75	3.22	97.01	2.30	97.31	2.49	97.49	2.50	97.50	2.50	97.50
5.00	7.63	94.01	3.77	96.53	2.22	97.24	2.47	97.48	2.49	97.50	2.50	97.50
10.00	21.04	84.57	8.19	92.66	1.95	97.04	2.40	97.40	2.47	97.51	2.50	97.50
15.00	30.54	77.72	14.10	87.58	1.84	96.97	2.33	97.34	2.46	97.52	2.50	97.50
20.00	36.36	73.27	19.72	82.78	1.80	96.94	2.29	97.31	2.45	97.52	2.50	97.50
25.00	40.12	70.25	24.39	78.77	1.78	96.93	2.27	97.28	2.45	97.52	2.50	97.50

*Normal approximation (exact for $t = 0$)

0.50% and 99.50% percent points

	Bird's approximation				Normal approximation				Chi square approximation			
	df = 10		df = 50		df = 10		df = 50		df = 10		df = 50	
t	0.50	99.50	0.50	99.50	0.50	99.50	0.50	99.50	0.50	99.50	0.50	99.50
0	0.08	99.92	0.37	99.63	0.50	99.50	0.50	99.50	0.50*	99.50*	0.50*	99.50*
0.25	0.08	99.92	0.37	99.63	0.50	99.50	0.50	99.50	0.50*	99.50*	0.50*	99.50*
0.50	0.08	99.92	0.38	99.63	0.50	99.50	0.50	99.50	0.50*	99.50*	0.50*	99.50*
0.75	0.09	99.92	0.38	99.63	0.50	99.50	0.50	99.50	0.50	99.50	0.50*	99.50*
1.00	0.11	99.91	0.39	99.62	0.50	99.50	0.50	99.50	0.50	99.50	0.50*	99.50*
1.50	0.14	99.88	0.41	99.60	0.49	99.49	0.50	99.50	0.50	99.50	0.50*	99.50*
2.00	0.20	99.83	0.44	99.58	0.48	99.48	0.50	99.50	0.50	99.50	0.50*	99.50*
2.50	0.30	99.74	0.48	99.55	0.46	99.46	0.50	99.50	0.50	99.50	0.50	99.50
3.00	0.44	99.61	0.53	99.50	0.45	99.45	0.50	99.50	0.50	99.50	0.50	99.50
4.00	0.92	99.17	0.67	99.38	0.41	99.41	0.49	99.49	0.50	99.50	0.50	99.50
5.00	1.77	98.44	0.86	99.21	0.37	99.37	0.49	99.49	0.49	99.51	0.50	99.50
10.00	11.13	91.64	3.03	97.23	0.25	99.27	0.45	99.45	0.47	99.52	0.50	99.50
15.00	21.38	84.59	7.25	93.53	0.21	99.23	0.42	99.42	0.45	99.53	0.50	99.50
20.00	28.67	79.33	12.32	89.21	0.19	99.22	0.40	99.40	0.44	99.53	0.49	99.50
25.00	33.67	75.54	17.15	85.12	0.18	99.21	0.39	99.39	0.44	99.53	0.49	99.51

*Normal approximation (exact for $t = 0$)

Conclusion

Let us return to the analogy between the statistic $\overline{Y}_1 - \overline{Y}_2 - t_{CALC}E$ and the familiar limits $\overline{Y}_1 - \overline{Y}_2 \pm t_{1-\alpha/2}E$ of the 100(1-α)% confidence interval for a raw difference.

It can be tempting to consider the interval $\overline{Y}_1 - \overline{Y}_2 \pm t_{CALC}E$. Let us assume for instance that t_{CALC} is positive. Then we can remark that t_{CALC} is the 100(1-p/2) percent point $t_{1-p/2}$ of the t distribution, where p is the two-sided p-value of the usual t test. Thus, the analogy seems again more compelling. By the definition of t_{CALC}, for the data in hand the bounds of this interval are zero, the traditional "null hypothesis" value, and two times the observed difference, what Rosnow and Rosenthal (1996) called the "counternull value".

In their methodological paper, Rosnow and Rosenthal (page 336) considered such an interval. Taking the example of an observed difference between two means +0.266 and a p-value p =0.23, they interpreted the specific null-conternull interval [0,+0.532] as "a 77% confidence interval", that is as a 100(1-p)% confidence interval. This interpretation reveals a typical confusion between frequentist and Bayesian probabilities. In the frequentist conception of confidence intervals, the confidence level is "the proportion of *repeated intervals* that contain the (fixed) parameter;" it is usually termed the *coverage* probability. The procedure, and in particular the confidence level, must be determined before knowing the data.

In the case of the Rosnow and Rosenthal interval [0,+0.532], we can envisage two possibilities to define the procedure and thus to compute the coverage probability.

On the one hand, the procedure can consist in computing the interval $\overline{Y}_1 - \overline{Y}_2 \pm t_{CALC}E$, with the *data dependent* value t_{CALC}. For each repeated sample the bounds of this interval are zero and the particular counternull value for this sample. Of course, the coverage probability of this interval varies with the parameters and *it is not equal to* 0.77 (except for two particular values of the ratio (μ_1-μ_2)/ε symmetrical around zero).

On the other hand, the procedure can consist in computing the interval $\overline{Y}_1 - \overline{Y}_2 \pm t_{1-p/2}E$, with the *fixed* value p = 0.23 for each repeated sample. The coverage probability of this interval is 0.77, but *this is not a frequentist procedure*, because 0.77 has been determined by the data in hand. Clearly, 0.77 is here a *data dependent* probability, which needs a Bayesian approach to be correctly interpreted. The Bayesian inference associates to the interval [0,+0.532] the "posterior probability that *this interval* contains the parameter, given the data."

Although confidence intervals refers to a frequentist justification, they are often (mis)interpreted in Bayesian terms. The distinction between the frequentist coverage probability and the Bayesian posterior probability is all the more subtle in the present situation that it turns out that it is correct from

a Bayesian viewpoint to say that "there is a 77% chance that the interval [0,+0.532] contains $(\mu_1-\mu_2)/\varepsilon$", or again in the example in Figure 1 to say that "there is a 95% chance that the interval [-0.986,+2.979] contains the noncentrality parameter λ". This simply assumes a prior distribution that does not favor any particular value of the parameters, what Bayesian called a "noninformative" prior. This distribution is revised by the data and the corresponding posterior distribution for the noncentrality parameter λ is just the $\Lambda'_{df}(t_{CALC})$ distribution. Consequently, the Lambda-prime distribution, in addition to its status of sampling distribution gains the status of a probability distribution that expresses our uncertainty about the unknown parameter λ. It results that we can use the confidence interval for λ with the benefits of both the frequentist and Bayesian interpretations and without worrying about the "correct justification" (not to speak of Fisher's fiducial argument).

The fact that even experts in statistics are not immune to conceptual confusions and interpret frequentist intervals in Bayesian terms should not be regarded as an "error". Rather this means that, as most statistical inference users, they are also, perhaps without knowing it, interested in the Bayesian probability. This should invite us not to radicalize the opposition between the Bayesian and frequentist inferences but rather to consider their interplay. This is a difficult challenge, but it is already well advanced in the statistical literature (see Bayarri & Berger, 2004).

References

Algina, J., & Keselman, H.J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement*, *63*, 537-553.

Bayarri, M.J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 19, 58–80

Bird, K. D. (2002). Confidence intervals for effect sizes in analysis of variance. *Educational and Psychological Measurement*, *62*, 197-226.

Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed and random-effects effect sizes. *Educational and Psychological Measurement*, *61*, 575-605.

Fisher, R. A. (1990/1973). *Statistical methods and scientific inference* (3rd edition 1973 reprinted). Oxford: Oxford University Press.

Johnson, N. L., & Welch, B. L. (1939). On the calculation of the cumulants of the χ distribution. *Biometrika*, *31*, 216-218.

Lecoutre, B. (1999). Two useful distributions for Bayesian predictive procedures under normal models. *Journal of Statistical Planning and Inference*, *77*, 93-105.

Rosnow, R.L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, *1*, 331-340.

Steiger, J. H. (2004). Beyond the *F*-test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods, 9,* 164-182.

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah, NJ: Lawrence Erlbaum.

Wilson, E. B., & Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences of the United States of America*, *17*, 684-688.