Chapter 5

From Significance Test to Fiducial Bayesian Inference

Bruno Lecoutre

in **Rouanet, H., Lecoutre, M.-P., Bert, M.-C., Lecoutre, B., Bernard, J.-M., & Le Roux, B. (1998) -** *New Ways in Statistical Methodology: From Significance Tests to Bayesian Inference*, **Berne: Peter Lang, in press (first edition in french entitled** *L'Infrence Statistique dans la Dmarche du Chercheur*, **1991).**

In this chapter we shall examine how, when analyzing experimental data, the researcher can call on intuitive knowledge to understand the principles and methodological implications of two of the main statistical inference procedures, namely, the traditional significance test and fiducial Bayesian inference. The underlying general problem will be the comparison of means in experimental designs. This problem is usually considered in an analysis of variance framework. In fact, it can be amply illustrated here in the case of a simple situation of inference concerning a mean.

5.1 Some intuitive considerations

5.1.1 A basic situation: the naming and reading experiment

Suppose a psychological experiment is designed to compare the time it takes to read words designating objects and the time it takes to name those objects depicted in figures. Let us take the example of the data of four adult subjects. Each of them read and named 80 items, words designating colors and discs of those same colors respectively. The following table gives, for each subject, the difference between the naming time and the reading time (average per item) expressed in hundredth of a second (*hs*).

subject	1	2	3	4	
difference	+12	+23	+22	+05	*hs*

This sample was chosen at random by "sampling without replacing" from the population of 432 subjects of an experiment realized in the course of experimental psychology at the University of Paris V in 1976, 1977, and 1978. As usual, the distribution of the 432 observed differences for these subjects will be called the parent distribution. To simplify, we have changed one value in this distribution (-2) to another value (+40), so that the mean of the population would be a whole number. By chance, this modification made the variance of the population into simple number.

In order to better delimit the statistical inference problem, we shall consider the following situation: (i) the mean of the parent population, denoted δ (to remind that it pertains to a difference of means [1], is unknown; (ii) all of characteristics of the population apart from its mean are known. The mean of the population is thus the single unknown characteristic (*i.e.* parameter, in the mathematical sense) of the population.

The parent distribution of the 432 differences expressed as deviations from δ is shown in Figure 5.1. We observe, for instance, that 20 subjects have a value equal to δ, 16 a value equal to $\delta - 1$, 22 a value equal to $\delta + 1$, *etc.* The smallest and the largest values of the population are respectively $\delta - 47$ and $\delta + 44$. The variance of this distribution, denoted σ^2, is equal to 98.5, hence its standard deviation $\sigma = 9.9247$ hundredths of a second.

1. The solutions for the elementary problem of inference about a single mean in fact makes it possible to process more complex situations, as in the present case the inference concerning the difference between two matched sample means, or more generally, the inference about a linear combination of means from several matched samples.

5.1.2 Sampling distribution of the mean

The information a sample provides about the population mean δ is usually summarized by the mean observed in this sample. Here the sample was chosen among 1,431,118,260 possible samples. This number can be obtained by the formula:

$$\frac{432!}{(432-4)!4!} = 1,431,118,260$$

From what we know about the parent population, we can easily (of course with a computer) generate the 1,431,118,260 possible samples and compute the mean of each of them (always expressed as a deviation from δ). The distribution of these 1,431,118,260 means is the sampling distribution of the "observed mean" D statistic. It is shown in Figure 5.2. We find for instance that 31,016,451 samples (2.167%) have a mean equal to δ (which is the mode of the distribution), 30,794,705 samples (2.152%) have a mean equal to $\delta + 0.25$, and 29,580,217 samples (2.067%) have a mean equal to $\delta + 1$. The smallest observed mean is equal to $\delta - 36.25$ (for one sample), and the largest observed mean is equal to $\delta + 36.50$ (for one sample too).

The sampling distribution has the following remarkable properties.
(1) Its mean, *i.e.* the mean of all sample means, is equal to δ.
(2) Its variance, *i.e.* the variance of all sample means, denoted ϵ^2, is approximately equal to $\frac{\sigma^2}{n}$. This approximation would give here 4.9624 for the standard deviation ϵ, while the real value is $\epsilon = 4.9451$. The exact formula giving the variance for a sample of size n in a population of size N is:

$$\epsilon^2 = \left(1 - \frac{n-1}{N-1}\right) \frac{\sigma^2}{n}$$

which is indeed close to $\frac{\sigma^2}{n}$ when the ratio $\frac{n}{N}$ (the sampling rate) is small. Therefore, the standard deviation ϵ decreases as the sample size n increases. Furthermore, in so far as n is small compared to N, ϵ is scarcely affected by the size of the population.
(3) The shape of the distribution is unimodal and approximately symmetrical. Finally, in spite of the small sample size here, the sampling distribution of the mean can be well approximated by a Normal distribution with the same mean and standard deviation. This property is illustrated in the following table.

Proportions of samples with a mean greater than or equal to

$\delta - 12.5$	$\delta - 10$	$\delta - 7.5$	$\delta - 5$	$\delta - 2.5$	δ
0.99066	0.97731	0.94225	0.85338	0.69009	0.49971
	$\delta + 2.5$	$\delta + 5$	$\delta + 7.5$	$\delta + 10$	$\delta + 12.5$
	0.29716	0.15211	0.06825	0.02826	0.01078

Normal approximation

$\delta - 12.5$	$\delta - 10$	$\delta - 7.5$	$\delta - 5$	$\delta - 2.5$	δ
0.99426	0.97842	0.93532	0.84402	0.69341	0.50000
	$\delta + 2.5$	$\delta + 5$	$\delta + 7.5$	$\delta + 10$	$\delta + 12.5$
	0.30659	0.15598	0.06468	0.02158	0.00574

We assume here that the main goal of experimental data analysis is to objectively express the information provided by the data. In this framework, the inductive question can be worded as follows: what

can be said about the unknown mean of the population δ, given the following available information: (1) the data from the four subjects, summarized by their mean, denoted d_{obs} (here $d_{obs} = +15.50$ hs); and (2) the sampling distribution of the "observed mean" D statistic, which in this case is the *frame model* (see Chapter 2), essential to all statistical inference procedures?

5.1.3 The significance test

The traditional significance test is related to the sampling inference theory, *i.e.* inferences drawn from the properties of the sampling distribution [2]. Such inferences can be called "frequentist", because the only probabilities involved are frequency probabilities, where a frequency probability is defined as the proportion (or relative frequency) of an infinite sequence of repetitions of the experiment. Furthermore, these inferences can be justified and interpreted in a combinatorial framework, as shown in Chapter 4.

The significance test is based on conditional, hypothetico-deductive, reasoning. If we set δ at a given value, the problem no longer involves any unknown quantity. This defines a hypothetical reference population, and we can apply the algorithms of combinatorial inference.

In practice it is known that one particular value, the traditional null hypothesis $\mathcal{H}_0$: $\delta = 0$, is practically the only one considered. This avoids having to think about other values that might be chosen.

(1) Deterministic inference: The impossible sample

If δ were equal to 0, the smallest possible value for the mean D would be -36.25 and the largest would be +36.50.

(1a) Any observed mean, d_{obs}, lying outside the preceding limits (for

2. In principle, the sampling inference is only based on these properties, which constitutes the desire for objectivity. But in practice, one is in fact forced to resort to all sorts of *ad hoc* and more or less implicit presuppositions that seriously cast doubt on this objectivity.

instance $d_{obs} = -40$ or $d_{obs} = +40$) allows us therefore to definitively reject the null hypothesis (deterministic rule). Such an observed mean is strictly inconsistent with the value $\delta = 0$.

(1b) Any observed mean, d_{obs}, lying inside these limits is compatible with the value $\delta = 0$; but of course, if we observe the value $d_{obs} = +15.50$, as in the present case, this does not imply in any way that $\delta = 0$!

(2) The significance test (probabilistic inference): The rare sample

(2a) In fact, the observed value $d_{obs} = +15.50$ is clearly far from 0, but lies in the range of the "possible" values under the null hypothesis. It is therefore logically compatible with that hypothesis. Nevertheless we observe that if δ were actually equal to 0, relatively few samples would exhibit a deviation that is at least as large: 8,335,111 samples, that is, only a small proportion $p = 0.0058$ (and therefore less than 1%) would differ from 0 (in absolute value) by more than 15.5. We can thus adopt the following decision rule. If p is judged to constitute an acceptable risk, we "decide" to conclude that δ is not equal to 0. In other words, we reject the null hypothesis $\delta = 0$. We say that the result of the test was significant.

(2b) Suppose we observe a value of d_{obs} closer to 0. In this case, the proportion of samples differing from 0 (in absolute value) by more than $|d_{obs}|$ increases. For instance, if $d_{obs} = +7.5$, this proportion is $p = 0.1260$ (therefore a bit more than 10%), and is generally judged as an unacceptable risk. Consequently, the result of the test is not significant, which scrictly speaking is a statement of ignorance: we cannot conclude that δ is not equal to 0, but this does not imply in any way that $\delta = 0$.

Any methodology which complies with frequentist statistical theory should be confined to a strict rule whereby the null hypothesis is rejected when the p value is smaller than a predefined risk value α. In practice, one knows that p is assessed with respect to reference

levels (traditionally 0.001, 0.01, 0.05, and sometimes 0.10, to which are associated notations and statements formulations that abide by following regulations:

$p \leq 0.001$	***	very highly significant result
$p \leq 0.01$	**	highly significant result
$p \leq 0.05$	*	significant result
$p \leq 0.10$	(.)	nearly significant result
$p > 0.10$	NS	nonsignificant result

Furthermore, this same statistical theory makes the distinction between two-sided level and one-sided level. In concrete terms, the latter leads one to accept a smaller risk. In this case, one should consider only those samples whose mean, under the null hypothesis $\delta = 0$, is greater than +15.5 (and not those samples whose mean is less than -15.5). But with respect to the traditional use of reference levels, this looks like a simple artifice aimed at obtaining an additional *.

The researcher has to be very careful of how he or she reasons so as to avoid natural interpretation temptations that are outside the framework of the above reasoning. The decision rule is conditional to a value of δ. Consequently the p-value must not be interpreted as the "probability that the null hypothesis is true", just as 1-p is not the "probability that the null hypothesis is false". In fact p is the "conditional probability of being wrong in rejecting the null hypothesis, IF IT IS TRUE ($\delta = 0$)."

5.1.4 Towards fiducial inference

The above erroneous interpretations in probabilistic terms about hypotheses are nevertheless heard! Aren't they the natural continuation of deterministic inference? Aren't they the only direct response to the researcher's question: "given the experimental results, can I state that naming time is likely to be greater than reading time?". Consequently, as stated by Freeman (1993), "No matter how we try to explain that a p-value in no way permits this conclusion, we are fighting a losing battle."

Moreover the researcher can be tempted to justify the conclusion by using another kind of reasoning, and by seeking to probabilize the possible values of the parameter δ in an intuitive way.

(1) Deterministic inference

The possible values for D lie between $\delta - 36.25$ and $\delta + 36.50$. In other words, the deviation $D - \delta$ falls between -36.25 and +36.50. Having an observed d_{obs} equal to +15.50, we can easily deduce that δ lies between 15.50- 36.50=-21.00 (the smallest strictly compatible value corresponding to the largest positive deviation +36.50) and 15.50+36.25=+51.75 (the largest strictly compatible value corresponding to the largest negative deviation -36.25). Between these two limits the only possible values are multiples of $\frac{1}{4}$ (-21.00, -20.75, -20.50, ..., +51.50, +51.75).

Furthermore, if the deviation $d_{obs} - \delta$ is assumed known, for instance, $d_{obs} - \delta = +1$, we can deduce the value of δ in a deterministic way: in the present case, $\delta = d_{obs} - 1 = +14.5$.

(2) Probabilistic inference

The proportion of samples corresponding to a given deviation between $d_{obs} - \delta$ is given by the sampling distribution. For instance 29,580,217 samples, *i.e.* a proportion 0.0207, have a mean equal to $\delta + 1$. We can then contemplate the following reasoning: (i) if we choose a sample at random, the probability that the deviation $X = D - \delta$ (hence $D = \delta + X$), where X is a "random variable", is +1 is equal to 0.0207; (ii) given the observed value $d_{obs} = +15.5$, a deviation X=+1 corresponds to $\delta = +14.5$; (iii) it appears natural, then, to say that we have a probability of 0.0207 that δ is equal to +14.5. This is what we shall call the "fiducial" inference temptation (the word fiducial is derived from the Latin *fiducia* = confidence).

By the same reasoning, we find for example: a probability of 0.0217 that δ is equal to +15.5 (31,016,451 samples have a deviation $X = 0$); a probability of 0.0025 that δ is equal to +5.5 (3,572,627

samples have a deviation $X = +10$); a probability of 0.0003 that δ is equal to 0 (460,027) samples have a deviation $X = +15.5$); *etc.*

We can then construct a fiducial probability distribution expressing our uncertainty about the set of possible values for δ, which takes into account the data and the sampling distribution. This fiducial distribution is shown in Figure 5.3. Such a distribution gives a natural response to the induction problem, and a probability judgment about the magnitude of the true difference δ can be made directly from it. From this distribution, we find the probability 0.997 that the true difference δ is positive (4,236,638 samples have a deviation $X \geq = +15.5$). Moreover it is clear here that we have a high probability that δ is not only positive (like the observed value d_{obs}), but is also greater, for instance, than five hundredths of a second:

$$Prob(\delta > 0) = 0.997 \text{ and } Prob(\delta > +5) = 0.977$$

In practice, the Normal frame model, assuming a distribution $N(\delta, \epsilon^2)$, is generally used as an approximation of the sampling distribution of the observed mean D. With some precautions due to the passage from discrete to continuous distributions, the fiducial distribution

related to δ (still assuming that σ is given) can be deduced using the same reasoning. It is simply a Normal distribution, whose center is the observed mean d_{obs}, and whose standard deviation $\epsilon = \frac{\sigma}{n}$ is that of the sampling distribution. This is written, $\delta | d_{obs}, \sigma \sim N(d_{obs}, \epsilon^2)$. Hence the fiducial distribution:

$$\delta \sim N(+15.5, 4.962^2)$$

which gives:

$$Prob(\delta > 0) = 0.999 \text{ and } Prob(\delta > +5) = 0.983$$

Of course, it will again be necessary to "eliminate" the nuisance parameter σ. But for our purposes, this essentially appears as a technical amendment (the Normal distribution "is changed into" Student's distribution) and does not question the inference principle concerning the parameter of interest δ.

5.2 Bayesian inference

5.2.1 The fiducial Bayesian method

The above reasoning is obviously very intuitive and does not claim to be universal. But a formal framework that can be used to justify and derive the preceding distribution in a rigorous way is provided by Bayesian inference (see Appendix of this chapter).

It is known that, in addition to the data and the sampling model, Bayesian inference involves an external element, which is the prior (or initial) probability distribution over the parameters. Once these various ingredients have been fixed, we apply Bayes' theorem and deduce a posterior (final or revised) distribution. Thus Bayesian methods combine considerable flexibility due to the choice of the prior distribution, with mathematical rigor.

According to a certain subjective Bayesian conception, prior distributions should incorporate all known information and opinions about the parameters being studied. This view obviously gives the

method tremendous potential. But it implies completely breaking away from current practices, that in most cases fulfill a need for objectivity. This may be the very reason for the mistrust, and sometimes the hostility, that many researchers show toward Bayesian statistics.

But there is another, just as Bayesian, conception developed by Jeffreys in the thirties, following Laplace [3], wherein the prior distributions express a "state of ignorance" about the parameters (see Jeffreys, 1961, in particular). Such prior distributions are called "noninformative". From the researcher's point of view, they are vague distributions which, *a priori*, do not favor any particular value of the parameters and consequently do not introduce any information other than the data themselves. This conception has gradually become recognized as a standard. Berger (1985) tells us, "We should indeed argue that noninformative prior Bayesian analysis is the single most powerful method of statistical analysis." (page 90) and "At the very least, use of noninformative priors should be recognized as being at least as objective as any other statistical techniques." (page 110).

A name had to be given to this conception, and to the corresponding posterior distributions. We propose to call it ***fiducial Bayesian***. Indeed, from a methodological standpoint, this approach is close to the fiducial approach developed by Fisher in the thirties (see Fisher, 1990), which can be regarded as an attempt to formalize the intuitive reasoning considered above. The incentive of the fiducial approach is that it produces probability judgments which reflect only that information provided by the analyzed data ("what the data have to say"). Furthermore, in the present situation, the two approaches end in the same distribution. In short, fiducial Bayesian inference is fiducial in incentive and Bayesian in technique.

From a practical standpoint, fiducial Bayesian distributions could furnish posterior probabilities as references for public use, and could

3. We might also mention the work by Ernest Lhoste, published in the *Revue d'Artillerie* (1923).

serve as a concise and objective way of communicating results. These reference probabilities may very well differ from the personal probabilities one obtains by incorporating outside information into the prior distribution. Moreover, a perfectly feasible project for a later stage would be to go beyond fiducial Bayesian analysis, precisely by incorporating information of this type.

5.2.2 The status of Bayesian methods in analysis of variance

The current status of Bayesian statistics is puzzling. On the one hand, we find numerous recent theoretical studies on Bayesian inference in the mathematical statistics journals, most of which seem to be convinced of the superiority of this approach (see Robert, 1994). At the same time, the actual use of Bayesian methods to analyze experimental data is more and more common in applied statistics journals, especially concerning clinical trials in medicine and pharmacology: see *e.g.* Racine *et al.* (1986); Berry (1991); Spiegelhalter, Freedman and Parmar (1994).

On the other hand, in a field of application as important as analysis of variance, most of today's books on the subject do not even mention Bayesian procedures. The commonly available computer packages do not include them either, despite the continued interest they spark up. In addition, the attitude of the Bayesian proponents often looks rigid, as if the use of Bayesian methods meant abandoning the other statistical procedures now in use. Furthermore the orientation of many authors in Bayesian statistics is decision-theoretic rather than inferential. The consequence is that the contribution of Bayesian inference to experimental data analysis has often been misunderstood.

Obviously, using the Bayesian approach should not result in an abrupt changeover from the frequentist methods now being employed. Given the widespread use of significance tests, this would be highly unrealistic. As Berry (1993) says, "the steamroller of frequentism is not slowed by words." At the very least, the two methods should

co-exist for many years to come. In short, rather than replacing current practices, Bayesian procedures should incorporate, extend, and refine them.

5.2.3 Reinterpretation of frequentist procedures

A well-known feature of fiducial Bayesian inference is that it can be used to reinterpret many of the frequentist procedures (see for example Lindley, 1965; Box and Tiao, 1973; Lecoutre, 1984b; Casella and Berger, 1987). For instance, for the comparison of two means from independent groups with the usual Normal model, which assumes variance equality, the observed one-sided significance level in Student's t test can be interpreted as the fiducial Bayesian probability that the true difference and the observed difference have opposite signs (for more details, see next Section). Furthermore in this case, the Bayesian credibility interval is identical to the frequentist confidence interval.

This reinterpretation bridges the conceptual and technical gap between fiducial Bayesian inference and frequentist procedures, and for many basic analysis of variance problems, offers the researcher a smooth transition from the traditional techniques to the fiducial Bayesian method. Moreover, the fiducial-Bayesian interpretation also points out some methodological shortcomings of the currently used techniques: it is quite apparent from the above example that the significance level only makes a statement about the sign, and has nothing to say about the real magnitude (size) of the effect.

However, it must be clearly understood that, as soon as we go beyond the more basic problems, there are many cases where fiducial Bayesian inference and frequentist procedures yield irreconcilable results (at least from a theoretical point of view). A well-known example is the Behrens-Fisher problem where the means of two independent groups are compared using the Normal model with unequal variances. The source of the discrepancy is clear here: fiducial Bayesian inference is conditional upon the observed variances of the two groups, and frequentist inference is conditional upon the true

variances of the model. As soon as we no longer assume that these two true variances are equal, the discrepancy appears.

Going back over the theoretical background of statistical inference, a provocative attitude could be to assert that the only real justification for using frequentist significance tests and confidence intervals is Bayesian. One can recognize at the very least that the Bayesian interpretation is far more intuitive and much closer to the thinking of researchers.

5.3 The usual significance test revisited in the light of fiducial Bayesian inference: does it allow for a nave methodology for determining the magnitude of an effect?

The researcher constantly faces the gap between the types of induction questions that can legitimately be asked – for instance, is the effect of this factor large (notable) or on the contrary small (negligible)? – and the brutal verdict provided by the usual significance test, significant or nonsignificant. Then it is tempting (if not unavoidable) to proceed to do what we call a "nave fiducial Bayesian analysis", *i.e.* to search for an intuitive assessment of the real magnitude of the effect based on the two available elements, the observed effect and the significance level We shall see that the fiducial Bayesian reinterpretation of significance levels – observed level or fixed (reference) level – clarifies this practice, while pointing out its pitfalls.

It will be sufficient here to consider the same elementary inference problem about the difference of means δ under the Normal model with a known standard deviation σ. Suppose the observed difference d_{obs} is positive, and that the inference aims to show: (i) either that the true difference δ is greater than one relevant positive value x ($\delta > x$), which we shall call the *notable effect* problem, (ii) or that the true difference δ is, in absolute value, smaller than one relevant positive value y ($-y < \delta < y$), which we shall call the *negligible effect*

problem.

Of course we shall first assume that a descriptive conclusion – notable (greater than x) or negligible (smaller in absolute value than y) d_{obs} effect – has been obtained for the sample. We then try to extend (generalize) this conclusion to the population. Let us introduce the following notations:

α	two-sided fixed level
p	two-sided observed level
d_α	critical two-sided (positive) value at the fixed level α (value of D from which the result of the test is declared significant)

Let us recall the previously stated results: (i) the sampling distribution of the observed difference D statistic under the null hypothesis $\mathcal{H}_0$: $\delta = 0$ is the Normal distribution $N(0, \epsilon^2)$; (ii) given the observed difference d_{obs}, the fiducial Bayesian distribution relative to the true difference δ is the Normal distribution $N(d_{obs}, \epsilon^2)$.

5.3.1 Reinterpretation of the observed level

For obvious reasons of symmetry, we can easily deduce (see Figure 5.4) the following fiducial Bayesian statements:

$$Prob(\delta < 0) = \frac{p}{2} \quad \text{or} \quad Prob(\delta > 0) = 1 - \frac{p}{2}$$
$$Prob(\delta < 0 \text{ or } \delta > 2d_{obs}) = p \quad \text{or} \quad Prob(0 < \delta < 2d_{obs}) = 1 - p$$

These statements are the fiducial Bayesian reinterpretation of the one-sided ($\frac{p}{2}$) and two-sided (p) observed levels, respectively.

(1) "Significant" result (p is "small")

In this case, the fiducial Bayesian probability that δ is positive ($1-\frac{p}{2}$) is high. In other words, it is well established that δ has the same sign as d_{obs}. In the reinterpretation of the one-sided observed level, the verdict is based solely on the sign of δ. Even if d_{obs} is notable, the preceding statements do not authorize the conclusion that there is

a notable effect. In fact, the only possibly relevant statement about the magnitude of the effect is $Prob(0 < \delta < 2d_{obs}) = 1 - p$, which leads to the conclusion that there is a negligible effect (and positive besides that) in the case where $2d_{obs}$ is negligible (with a significant result...).

(2) "Nonsignificant" result (p is not "small")

In this case, the usual (conventional) practice amounts to considering that $1 - p$ (or even $1 - \frac{p}{2}$) is not a sufficient guarantee to conclude that an effect exists. Therefore, among the probabilities associated with the different statements, $\frac{p}{2}$ ($< \frac{1}{2}$), $1 - \frac{p}{2}$, p, and $1 - p$, only p, if it is high, could provide a satisfying guarantee, but the corresponding conclusion ($\delta < 0$ or $\delta > 2d_{obs}$) is hardly worth anything... [4]

The reinterpretation of the observed level itself is therefore slightly informative, as far as the magnitude of the effect is concerned.

5.3.2 Reinterpretation of the fixed level

For the fixed level α, the following fiducial Bayesian statements can again be deduced for reasons of symmetry (see Figure 5.5):

$$Prob(\delta < d_{obs} - d_\alpha) = \frac{\alpha}{2} \quad \text{or} \quad Prob(\delta > d_{obs} - d_\alpha) = 1 - \frac{\alpha}{2}$$

$$Prob(\delta < d_{obs} - d_\alpha \text{ or } \delta > d_{obs} + d_\alpha) = \alpha$$
$$\text{or} \quad Prob(d_{obs} - d_\alpha < \delta < d_{obs} + d_\alpha) = 1 - \alpha$$

These statements are the fiducial Bayesian reinterpretation of the one-sided ($\frac{\alpha}{2}$) and two-sided (α) fixed levels, respectively.

4. In the more favorable case where $p = 1$, *i.e.* $d_{obs} = 0$, it reduces to the trivial statement $Prob(\delta < 0 \text{ or } \delta > 0) = 1$.

Note that the statement $Prob(d_{obs} - d_\alpha < \delta < d_{obs} + d_\alpha) = 1 - \alpha$ is obviously the fiducial Bayesian interpretation of the usual $1 - \alpha$ confidence interval (centered around d_{obs}) for δ. In the Bayesian framework, the interval $[d_{obs} - d_\alpha, d_{obs} + d_\alpha]$ is usually called a *credibility interval*. This name distinguishes it from the frequentist confidence interval and reminds us that it is correct in this case to say "the probability that δ lies between $d_{obs} - d_\alpha$ and $d_{obs} + d_\alpha$ is equal to $1 - \alpha$."

According to the usual conventions, α is chosen to be "small", so that $1 - \alpha$ (or $1 - \frac{\alpha}{2}$) can be considered as a sufficient guarantee to allow for an inductive conclusion, regardless of whether the outcome of the test is significant or nonsignificant.

The limits in the preceding statements are $d_{obs} - d_\alpha$ and $d_{obs} + d_\alpha$.
(i) If the limit $d_{obs} - d_\alpha$ is positive and notable, the statement $Prob(\delta > d_{obs} - d_\alpha) = 1 - \frac{\alpha}{2}$ allows for the conclusion that there is a notable effect. This assumes that the result is "clearly" significant relative to the fixed level α (d_{obs} notably higher than d_α, therefore p clearly smaller than α).
(ii) If the two limits, $d_{obs} - d_\alpha$ and $d_{obs} + d_\alpha$, are negligible, the statement $Prob(d_{obs} - d_\alpha < \delta < d_{obs} + d_\alpha) = 1 - \alpha$ allows for the conclusion that there is a negligible effect (but this is not a symmetrical interval around zero [5]. This is independent of the significance of the result: we can have $d_{obs} > d_\alpha$ or $d_{obs} < d_\alpha$.

We can deduce the following general rules that a nave fiducial Bayesian analysis should have to obey.
(i) If d_{obs} is "clearly notable" and the result is "clearly significant", then, to the extent that these conditions correspond to a notable $d_{obs} - d_\alpha$ value, the conclusion that there is a notable effect is often reasonable. This is quite obviously a situation that puts the researcher at his (or her) ease, since there is convergence between the

5. The usual confidence interval is centered around the observed effect and can therefore only provide an imperfect answer to the negligible effect problem. The construction of confidence intervals centered around zero has a long story in statistics. This problem is of particular interest, since it brings a clear cut between the frequentist and Bayesian approaches (see Appendix).

result of the test and the descriptive conclusion. But the conclusion nevertheless remains rather impressionist.
(ii) If d_{obs} is notable and the result is nonsignificant, then no inductive conclusion can be reached: we cannot conclude that an effect exists, and it is obviously out of the question to conclude that there is a non-notable effect. This is an often cumbersome situation for the researcher, who cannot generalize the descriptive conclusion (one way to get through this is often to claim that "there is a trend"). But this situation is not contradictory in realityexcep, since it simply indicates that the experimental information is insufficient to reach a conclusion.
(iii) If d_{obs} is "clearly" negligible and the result is "clearly significant", then, in so far as these conditions correspond to negligible $d_{obs} - d_\alpha$ and $d_{obs} + d_\alpha$ values (they imply at least that $d_{obs} - d_\alpha$ is "clearly negligible"), a negligible effect conclusion is generally reasonable. This situation generally appears contradictory, or at least cumbersome, to the researcher. There is no paradox however, except for the fact that this can only happen if the experimental accuracy is "very good", which means that ϵ is very small. In such a case, according to the significance test, a small observed difference can be significant (the test is said to be *powerful*), while there will be very little dispersion around d_{obs} in the fiducial Bayesian distribution. This shows that we are in fact dealing here with a privileged situation!
(iv) If d_{obs} is "clearly negligible" and the result is nonsignificant, then these conditions only imply that d_{obs} is less than d_α. But they can correspond to negligible as well as non-negligible $d_{obs} - d_\alpha$ and $d_{obs} + d_\alpha$ values. In this situation, the significance level by itself does not bring in any useful information: no conclusion can be reached. Nevertheless, like the first one, this situation is often regarded as favorable by the researcher (in spite of warnings about nonsignificant results), since there is apparent convergence between the descriptive conclusion and the result of the test.

In summary a nave fiducial Bayesian practice appears at the very least to be tricky. It is in fact possible, in a formal way, to deduce a

fiducial Bayesian statement about the magnitude of the effect, from the observed effect and the significance level. But the link between them is at the very least far from obvious: in particular, the value of the test statistic has to be recomputed (Lecoutre, 1985).

5.4 Illustrations of the fiducial Bayesian procedures: conflictual situations

The *conflictual situations* presented in Chapter 3 are considered here in the light of fiducial Bayesian inference. This will serve as a general illustration of the fiducial Bayesian procedures, and will show how they can immediately be implemented from usual statistical outcomes (table of means and significance tests). Taking into account the relative simplicity of the situations considered, all the of results presented can be calculated with the help of a detailed table of Student's t distribution. Alternatively, these results can be obtained from the **PAC** (Program for the Analysis of Comparisons) computer software by Lecoutre and Poitevineau (1992).

5.4.1 Interaction situation

In this situation (see Sections 3.2.2 and 3.4.1), we have the following table of means:

	Factor B		
Factor A	$b1$	$b2$	
$a1$	80.4	66.1	73.3
$a2$	66.5	62.1	64.3
	73.5	64.1	*seconds*

From these means, we define the main observed effects d_{obs} for the two experimental factors, and their interaction (differences of the differences):

Factor A $\quad d_{obs} = 73.3 - 64.3 = +9.0$
Factor B $\quad d_{obs} = 73.5 - 64.1 = +9.4$
Interaction A.B $\quad d_{obs} = (80.4 - 66.1) - (66.5 - 62.1) = +9.9$

The corresponding F ratios (with 1 and 60 degrees of freedom) have the observed values:

Factor A	$F = 4.0$	$p < 0.05$
Factor B	$F = 4.3$	$p < 0.05$
Interaction A.B	$F = 1.2$	NS

At the descriptive level, an effect associated with each of the two factors A and B, is found to exist, as well as a notable interaction effect. But at the inferential level, the outcome of the significance test is nonsignificant for the interaction effect, while the main effects are significant, hence the conflict.

A very simple and general result is that the fiducial Bayesian distribution for the true effect δ is a generalized Student's t distribution, whose center is the observed effect d_{obs} (instead of 0 for the usual elementary t), and whose scale factor is $e = \frac{|d_{obs}|}{\sqrt{F}}$ (instead of 1). The distribution has q degrees of freedom, according to the denominator of the F ratio. This is written: $\delta \sim t_q(d_{obs}, e^2)$. If q is high, this is approximately a Normal distribution with center (mean) d_{obs} and standard deviation e: $\delta \sim N(d_{obs}, e^2)$ (the exact standard deviation is $e\sqrt{\frac{q}{q-2}}$).

This result is in fact applicable to all inferences about a linear combination of means for which we know a significance test using Student's t distribution with q degrees of freedom, or Fisher-Snedecor's F distribution with 1 and q degrees of freedom (which is the square of the t distribution with q degrees of freedom). It therefore establishes an immediate technical link between the fiducial Bayesian inference and the significance test.

Here, the following distributions can be immediately deduced (see Figure 5.6):

Factor A	$\delta \sim t_{60}(+9.0, 4.5^2)$	where $e = \frac{9.0}{\sqrt{4.0}} = 4.5$
Factor B	$\delta \sim t_{60}(+9.4, 4.5^2)$	where $e = \frac{9.4}{\sqrt{4.3}} = 4.5$
Interaction A.B	$\delta \sim t_{60}(+9.9, 9.0^2)$	where $e = \frac{9.9}{\sqrt{1.2}} = 9.0$

With factors A and B, we can indeed extend the descriptive conclusion that there is a notable effect. With a fiducial Bayesian guarantee of 0.90, we can state that the true difference between $a1$ and $a2$ is greater than 3.2 seconds and that the true difference between $b1$ and $b2$ is greater than 3.6 seconds:

$$\begin{array}{ll} \text{Factor A} & Prob(\delta > +3.2) = 0.90 \\ \text{Factor B} & Prob(\delta > +3.6) = 0.90 \end{array}$$

With the interaction, on no account can the nonsignificant result be interpreted in favor of a negligible effect:

$$\text{Interaction A.B} \quad Prob(|\delta| < 21.6) = 0.90$$

The fiducial Bayesian inference does not conflict with the descriptive conclusion of a strong interaction effect, but it clearly shows that the information available in the data is insufficient to generalize this conclusion: more data or external information is needed.

5.4.2 Replication situation

In this situation (see Sections 3.2.2 and 3.4.2) the conflict comes from the apparent divergence of results between the experiment and its replication. We have the following results:

Experiment 1

$$d_{obs} = +3.75 \quad \text{and} \quad t = +2.35\ [39\ df] \quad p < 0.025$$
$$\text{hence } \delta \sim t_{39}(+3.75, 1.60^2) \text{ where } e = \tfrac{3.75}{2.35} = 1.60$$

Experiment 2 (version A)

$$d_{obs} = +1.63 \quad \text{and} \quad t = +0.96\ [39\ df] \quad p > 0.30$$
$$\text{hence } \delta \sim t_{39}(+1.63, 1.70^2) \text{ where } e = \tfrac{1.63}{0.96} = 1.70$$

Experiment 2 (version B)

$$d_{obs} = +0.05 \quad \text{and} \quad t = +0.03\ [39\ df] \quad p > 0.95$$
$$\text{hence } \delta \sim t_{39}(+0.05, 1.67^2) \text{ where } e = \tfrac{0.05}{0.03} = 1.67$$

By pooling the results of the two experiments, we obtain, for each of the two versions:

Version A $d_{obs} = +2.69$ and $t = +2.31\ [79\ df]$ $p < 0.025$
hence $\delta \sim t_{79}(+2.69, 1.16^2)$

Version B $d_{obs} = +1.90$ and $t = +1.62\ [79\ df]$ $p > 0.10$
hence $\delta \sim t_{79}(+1.90, 1.17^2)$

The fiducial Bayesian distribution obtained for the pooled data (see Figure 5.7) shows that the results of the two experiments, without being convergent, are nevertheless entirely compatible (especially for version A). After pooling, we obtain the statements:

Version A $Prob(+0.76 < \delta < +4.62) = 0.90$
Version B $Prob(-0.05 < \delta < +3.85) = 0.90$

Here again, additional information appears to be necessary in order to specify the real magnitude of the effect.

5.5 An overview on predictive fiducial Bayesian procedures

As shown in Chapter 3, an important aspect of statistical induction is making predictions. In this case we want to express our uncertainty about the value of a statistic – typically here, the difference d – that we would observe for new data. Once again, the fiducial Bayesian inference offers a direct and very intuitive solution. Let us consider the "statistical prediction" situation presented in Chapter 3 (see Sections 3.3.2 and 3.4.4). In the experiment conducted, we have observed the difference d_{obs} and the value of Student's t statistic. Then we want to make a prediction about a replication of this experiment (with the same sample size). Three sets of data were considered:

Situation 1 $d_{obs} = +1.82$ and $t = +2.093$ [19 df] $p = 0.05$
hence $\delta \sim t_{19}(+1.82, 0.87^2)$
Situation 2 $d_{obs} = +0.92$ and $t = +1.058$ [19 df] $p = 0.30$
hence $\delta \sim t_{19}(+0.92, 0.87^2)$
Situation 3 $d_{obs} = +0.22$ and $t = +0.253$ [19 df] $p = 0.80$
hence $\delta \sim t_{19}(+0.22, 0.87^2)$

The predictive distribution for the observed difference d' found in a future sample will naturally be more scattered than the distribution of δ relative to the population (this is all the more true since the size of the new sample will be smaller). In fact, the uncertainty about the results of the replication is added to the uncertainty about δ after the performed experiment.

In the basic situation of making an inference about a mean under the Normal model with known variance, the fiducial Bayesian predictive distribution for d', given the mean d_{obs} observed in the first experiment, is simply a Normal distribution, whose center is d_{obs}, and whose variance is equal to the sum of the variances of the sampling distributions of the means for each of the two samples, ϵ^2 and ϵ'^2: $d' \sim N(d_{obs}, \epsilon^2 + \epsilon'^2)$.

This result can be generalized to the case of an unknown variance σ^2 (Lecoutre, 1984a, 1996), and we obtain the predictive distributions (see figure 5.8):

Situation 1 $d' \sim t_{19}(+1.82, 2 \times 0.87^2) \sim t_{19}(+1.82, 1.23^2)$
Situation 2 $d' \sim t_{19}(+0.92, 2 \times 0.87^2) \sim t_{19}(+0.92, 1.23^2)$
Situation 3 $d' \sim t_{19}(+0.22, 2 \times 0.87^2) \sim t_{19}(+0.22, 1.23^2)$

The first question pertains to the sign of the difference d' in the replication. As a general rule, it is clear that the fiducial Bayesian probability that d' is positive lies between $\frac{1}{2}$ (the probability that d' is greater than d_{obs}) and $1 - \frac{p}{2}$ (the probability that δ is positive, here, for each of the three situations, 0.975, 0.85, and 0.60). More precisely, we get:

$$\begin{array}{ll} \text{Situation 1} & Prob(d' > 0) = 0.92 \\ \text{Situation 2} & Prob(d' > 0) = 0.77 \\ \text{Situation 3} & Prob(d' > 0) = 0.57 \end{array}$$

The second question is about the significance test statistic in the replication. Assuming again that the variance σ^2 is known, this question simply amounts to a question about d'.

Situation 1: "What, for you, is the probability that in the second experiment the observed difference, d', will have the same sign as d_{obs}, and that the result of Student's t test will be at least as significant as in the first experiment?"
For σ^2 given, this can be reduced simply to $d' > d_{obs}$, and hence:

$$Prob(d' > d_{obs}) = 0.50$$

Situations 2 *and* 3: "What, for you, is the probability that in the second experiment the observed difference, d', will have the same sign as d_{obs}, and that the result of Student's t test will be at least as nonsignificant as in the first experiment?"
For σ^2 given, this can be reduced simply to $0 < d' \leq d_{obs}$, and hence:

$$\begin{aligned} Prob(0 < d' \leq d_{obs}) &= Prob(d' > 0) - Prob(d' > d_{obs}) \\ &= Prob(d' > 0) - 0.50 \end{aligned}$$

which is therefore smaller than $\frac{1-p}{2}$. This reasoning gives the approximate solutions:

$$\begin{array}{lll} \text{Situation 1} & Prob(t' > +2.093) & = Prob(d' > +1.82) \\ & & = 0.50 \\ \text{Situation 2} & Prob(0 < t' < +1.058) & = Prob(0 < d' < +0.92) \\ & & = 0.77 - 0.50 = 0.27 \end{array}$$

$$\text{Situation 3} \quad \begin{aligned} Prob(0 < t' < +0.253) &= Prob(0 < d' < +0.22) \\ &= 0.57 - 0.50 = 0.07 \end{aligned}$$

In fact, in the case where σ^2 is unknown, the additional uncertainty about the standard deviation observed in the replication must be taken into account. But, with two decimal places, the exact solution (see Lecoutre, 1996) gives the same values as the above approximate solution.

5.6 Conclusion: the methodological contributions of Bayesian inference

The use of the fiducial Bayesian method (and the Bayesian method in general) in experimental data analysis appears fully justified: not only does it provide an easy and natural interpretation for the procedures, directly in terms of probabilities on parameters, its solutions can be backed theoretically in a much more satisfying manner.

Many books have indeed pointed out the advantages of Bayesian inference. For an introduction, the reader may wish to consult Phillips (1973), Novick and Jackson (1974), and Lee (1997). But, rather than theoretical considerations, the reader may instead want to know more about the methodological contributions of Bayesian procedures to experimental data analysis. The features already discussed will be summarized here, and some further attractive features will be briefly outlined.

5.6.1 Conclusions about the magnitude of effects

First and foremost, Bayesian procedures are ideally suited to drawing conclusions about the magnitude of the investigated effects. They provide direct answers to the real questions raised in virtually all applications. In pharmacology, for example, one may want to find out whether a certain dose of a new drug is notably more effective that some other dose. Another example is when a placebo effect is observed (*significant* when enough patients are included), in which case

what needs to be shown is that the effect is limited (if not negligible) compared to the effect of the drug. These questions are naturally worded as follows: (i) in the first case, what is the probability that the difference between the two means is large? (ii) in the second case, what is the probability that the difference (in absolute value) is small?

The problem of the magnitude of effects can no longer be avoided, as is implicitly done in most publications. Assessing the magnitude of an effect cannot be a problem handed over to the statistician. By nature, this question precedes the statistical inference. In a given field, the researcher, taking into account the current state of knowledge, must make the effort to specify what effects are negligible or notable, at the very least by assessing the relative magnitudes of effects [6].

Let us recall again that the usual significance test obviously does not answer these questions: a "significant" result only means that the hypothesis of a null effect can be rejected, and a "nonsignificant" result is nothing more than a statement of ignorance. On the contrary, Bayesian inference provides direct responses.

Our work on analysis of variance shows that standard Bayesian procedures can be implemented as easily as the traditional F ratios: see Lecoutre, 1981a, 1983, 1984a, 1996; Rouanet and Lecoutre, 1983; Rouanet, 1996. For complex experimental designs, the construction of these procedures is based on the specific inference principle (see Rouanet and Lecoutre, 1983; Lecoutre, 1996). In short, this principle consists of considering the comparisons of interest separately, and making each inference from specifically relevant derived data, as illustrated above in the elementary example of the naming and reading experiment.

---

6. In pharmacological studies, for example, the negligible effect problem is explicitly formulated in terms of "equivalence". Thus two drugs A and B are said to be (absolutely) equivalent for a certain variable if the difference between the means μ_A and μ_B is smaller in absolute value than a given value. This quantity is often defined as a percentage of variation (for instance $|\mu_A\text{-}\mu_B|$ smaller than 25% of the observed mean for drug A).

The technical problems linked to the use of Bayesian distributions are now easily solved by computers. The fiducial Bayesian method has been applied many times to real data: see in particular Rouanet, Lpine, and Pelnard-Considre (1975), Rouanet and Lpine (1977), Rouanet, Lpine, and Holender (1978), Lecoutre (1981b), Hoc (1983), Denhire and Lecoutre (1983), Ciancia *et al.* (1988), Lecoutre (1992), Clment and Richard (1997). Its uses range from testing "sharp models" (Rouanet, 1986; Lecoutre, Rouanet, and Denhire, 1988), to searching by means of an exploratory process for the "significant features" in the data (Rouanet, Lecoutre, and Bernard, 1987).

5.6.2 Inferences about individual effects

Another highlighting feature of the Bayesian method is that inferences can be made about individual effects. For example, in the naming and reading experiment, the naming time was found to be greater than the reading time for means. The next step is to find out whether or not it is also greater in most cases. If we assume that the observed differences between these two times come from a parent population with a $N(\delta, \sigma^2)$ distribution, the problem is to determine whether the proportion of notable differences in this population is sufficiently large.

Let us define for any number x the proportion φ_x of the population greater than x. We have symbolically written:

$$\varphi_x = Prob\left(N(\delta, \sigma^2) > x\right) = Prob\left(N(0,1) < \frac{\delta - x}{\sigma}\right)$$

Bayesian inference provides a direct answer to the problem: we simply compute the posterior probability that φ_x is greater than π. This probability is:

$$Prob(\varphi_x > \pi) = Prob\left(\frac{\delta - x}{\sigma} > \overline{z}_\pi\right)$$

which, if σ is given, reduces to:

$$Prob(\varphi_x > \pi) = Prob(\delta > x + \overline{z}_\pi \sigma)$$

where $\overline{z}_\pi$ is the upper point of the $N(0,1)$ distribution such that $Prob(N(0,1) < \overline{z}_\pi) = \pi$.

In the example of the data for four subjects considered in Section 5.1.1 (the naming and reading experiment), all four observed differences are greater than $x = 2$. For $\pi = 0.80$ ($\overline{z}_\pi = +0.8416$), and assuming again $\sigma^2 = 98.5$ known, we get:

$$\begin{aligned} Prob(\varphi_x > 0.80) &= Prob(\delta > 2 + 0.8416 \times 9.9247) \\ &= Prob(\delta > 10.353) = 0.85 \end{aligned}$$

which is directly deduced from the fiducial Bayesian distribution obtained in Section 5.1.4 under the Normal model, $\delta \sim N(+15.5, 4.962^2)$. Hence we have a guarantee of 0.85 that the proportion of population differences greater than 2 is larger than 80%.

Alternatively, for any given guarantee γ and any given proportion π, there exits a value x such that the proportion φ_x of population differences greater than x is at least π with probability γ. The higher the proportion π and the larger the value x, the more conclusive is the experiment with regard to the notable greater length of the naming time "in most cases". For instance here, we have a guarantee of 0.90 that the proportion of population differences greater than 0.788 is larger than 80%.

This result is generalized to the case of an unknown parent variance σ^2 (Lecoutre, 1996, Chapter 1). Its methodological utility has been illustrated for validating models in experimental psychology (Rouanet, Lpine, and Holender, 1978), and for assessing individual equivalence in pharmacology (Lecoutre and Derzko, 1997).

5.6.3 Greater flexibility for analyzing and monitoring experiments

Clearly, the Bayesian approach offers more flexibility to experimental data analysis. In addition to the necessary objective statements for reporting results based on fiducial Bayesian procedures, it provides an efficient tool for personal decisions and for designing ("How many subjects?") and monitoring ("When to stop?") experiments.

On the one hand, various prior distributions expressing results from other experiments or subjective opinions of well-informed specific individuals, whether *skeptical* or *enthusiastic*, can be investigated to assess the robustness of the conclusions: see in particular Bayesian methodology for clinical trials exposed by Spiegelhalter *et al.* (1994), and Lecoutre (1996, Chapter 3). Technically, the posterior distribution corresponding to a given prior distribution can be derived directly from the fiducial Bayesian distribution. The property of the latter to convey the information contained in the data is highlighted here.

On the other hand, Bayesian predictive probabilities can be used especially for choosing a sample size and for conducting interim analyses. They enable the researcher to evaluate the real chances of a given conclusion to be obtained with possible future observations, on the basis either of a "pilot" study or of partial results of a current experiment: see e.g. Choi and Pepple (1989); Berry (1991); Lecoutre, Derzko, and Grouin (1995); Grouin and Lecoutre (1996); Lecoutre (1996, chapter 8).

5.7 Appendix

5.7.1 Bayesian inference concerning the mean δ under the Normal model (σ^2 known)

The inference is based on the sampling distribution of the observed mean d (conditional to the true mean δ):

$$d|\delta \sim N\left(\delta, \frac{\sigma^2}{n}\right)$$

where n is the sample size and the parent variance σ^2 is assumed known. In this situation, d is a sufficient statistic for δ and therefore summarizes all information provided by the sample.

(1) For a prior distribution of δ with probability density function $p(\delta)$, the Bayes formula gives the density function of the posterior Bayesian distribution (conditional to d):

$$p(\delta|d) = \frac{p(d|\delta)p(\delta)}{\int p(d|\delta)p(\delta)\, d\delta}$$

As a function of δ and d, the posterior density function is simply proportional to the product $p(d|\delta) \times p(\delta)$, hence:

$$p(\delta|d) \propto \exp\left(\frac{(d-\delta)^2}{2\frac{\sigma^2}{n}}\right) \times p(\delta)$$

(2) If we choose the noninformative prior locally uniform distribution for δ, *i.e.* a constant density function on an arbitrarily large interval, the preceding expression remains unchanged. The classic result follows:

$$\delta|d \sim N\left(d, \frac{\sigma^2}{n}\right)$$

In this solution we go from the sampling distribution $d|\delta \sim N(\delta, \frac{\sigma^2}{n})$ to the posterior distribution $\delta|d \sim N(d, \frac{\sigma^2}{n})$, which corresponds to an intuitive "pivot" and can as such be justified by Fisher's fiducial argument. This is why we call this solution the standard Bayesian solution or the fiducial Bayesian solution.

(3) If we choose the Normal (*conjugate*) prior distribution for δ:

$$\delta \sim N\left(d_0, \frac{\sigma^2}{n_0}\right)$$

the resulting posterior and predictive distributions are again Normal, and are respectively:

$$\begin{aligned} \delta|d &\sim N\left(d_1, \frac{\sigma^2}{n_1}\right) \text{ where } d_1 = \frac{n_0 d_0 + nd}{n_0 + n} \text{ and } n_1 = n_0 + n \\ d &\sim N\left(d_0, \left(\frac{1}{n_0} + \frac{1}{n}\right)\sigma^2\right) \end{aligned}$$

and we again obtain the fiducial Bayesian distribution $\delta|d \sim N(d, \frac{\sigma^2}{n})$ as a borderline case, when $n_0 \to 0$.

5.7.2 Confidence interval centered around zero

Unfortunately the uniformly most powerful test of the null hypothesis $|\delta| > x$ against the alternative $|\delta| \leq x$ has highly undesirable properties (see Schervish, 1995, page 252). As a consequence no satisfactory exact

frequentist confidence interval centered around zero can be obtained. The generally adopted solution is to construct a "more than $1-\alpha$" confidence interval as the set of positive x such that the two one-sided tests $H_0 : \delta = x$ against $H_1 : \delta > x$ and $H_0 : \delta = -x$ against $H_1 : \delta < -x$ are simultaneously nonsignificant at level α (see *e.g.*, Schuirmann, 1987). Practically we simply compute the usual $1-2\alpha$ (and not $1-\alpha$) confidence interval for δ and consider the largest in absolute value of its two bounds (Deheuvels, 1984).

The confidence interval obtained by this procedure is shorter than the $1-\alpha$ fiducial Bayesian credibility interval, revealing an irreconcilable discrepancy between the two solutions. From the fiducial Bayesian viewpoint, the confidence interval has a too weak posterior probability (less than $1-\alpha$), while from the frequentist viewpoint the Bayesian solution must be discarded since it gives a too much larger interval. But it must be recalled that, in the frequentist framework, the particular type of interval used must be specified before collecting data, while it is not a prerequisite of Bayesian methods.

5.7.3 The PAC ("Program for the Analysis of Comparisons") software

The technical problems involved in the use of Bayesian distributions are now easily solved by computers.

PAC (Lecoutre and Poitevineau, 1992) is a general univariate and multivariate analysis of variance program. It includes the traditional analysis of variance significance tests, but offers additional capabilities for searching for conclusions about the magnitude of effects and investigating assumptions about variances and covariances.

Effect size measures, both for raw effects and for standardized effects (generalizing the Cohen's d and f indexes), are systematically computed. Corresponding fiducial Bayesian credibility intervals (using noninformative prior distributions), as well as alternative frequentist confidence intervals, are routinely available for asserting the importance of effects. For one degree of freedom comparisons, *conjugate* prior distributions (which are in same family as the fiducial Bayesian distribution) can be used to incorporate outside information. A "Bayesian module" displays and prints Bayesian probability distributions and calculates the corresponding probability statements, in interaction with the user.

Furthermore procedures involving no assumptions about variances and

covariances are provided for most usual situations. These procedures are direct extensions of the Behrens-Fisher solution to the basic problem of comparing two means with variances not assumed to be equal.

All of the procedures are applicable to general experimental designs (in particular, repeated measures designs), balanced or not balanced, with univariate or multivariate data, and covariables. A powerful request language allows the user to easily perform specific analyses for all comparisons of interest: main effects, partial effects, interaction effects, conditional effects, component effects in polynomial regression, *etc.*

A Windows limited version of **PAC** and other Bayesian programs are freely available on the Internet at the following address:

http://epeire.univ-rouen.fr/labos/eris/pac.html