In Bayesian Methods with Applications to Science, Policy and Official Statistics
Luxembourg: Office for Official Publications of the European Communities, 301-310.

Bayesian Predictive Procedures for Designing and Monitoring Experiments

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

Abstract

In recent years many authors have stressed the interest of the Bayesian predictive approach for designing ("how many subjects?") and monitoring ("when to stop?") experiments. The predictive distribution of a test statistic can be used to include and extend the frequentist notion of power in a way that has been termed predictive power or expected power. More generally Bayesian predictive procedures give the researcher a very appealing method to evaluate the chances that the experiment will end up showing a conclusive result, or on the contrary a non-conclusive result. The prediction can be explicitly based on either the hypotheses used to design the experiment, expressed in terms of prior distribution, or on partial available data, or on both.

Keywords: DATA MONITORING; PREDICTIVE DISTRIBUTIONS; SAMPLE SIZE DETERMINATION; STOCHASTIC CURTAILMENT; AVERAGE POWER.

1 INTRODUCTION

A major strength of the Bayesian paradigm is the ease with which one can make predictions about future observations. The predictive idea is central in experimental investigations. Furthermore Bayesian predictive probabilities are efficient tools for designing and monitoring experiments. Bayesian predictive procedures give the applied researcher a very appealing method to evaluate the chances that an experiment will end up showing a conclusive result, or on the contrary a non-conclusive result (see e.g., Berry, 1991; Lecoutre *et al.,* 1995; Dignam *et al.,* 1998). These procedures are far more intuitive and much closer to the thinking of scientists than frequentist procedures.

I shall limit my presentation to simple inferential problems, in order to emphasize conceptual and methodological rather than technical issues. However it should be clear that more complex problems could be handled with the same approach. Two examples will serve as illustration of the procedures. The first example concerns the inference about a proportion. It will serves to contrast the Bayesian approach with the frequentist procedures currently in use. The second example concerns the inference about a difference between two means.

2 INFERENCE ABOUT A PROPORTION

2.1 Example 1

As it is well known, the correct frequentist interpretation of p-values is so unnatural that it seems a losing battle to attempt to teach it. Nevertheless a frequentist colleague of mine claimed that he developed an individual teaching method that yielded a high rate of correct interpretations. We agreed that this method was effective if the error rate after teaching was less than 0.15, and that it was ineffective if the error rate was greater than 0.30.

My prior probability that the method was ineffective was so high that I thought unnecessary to collect data. Of course my colleague was unconvinced by this blatant subjectivism and designed an experiment within the traditional Neyman-Pearson framework. He considered the null hypothesis $H_0 : \varphi = \varphi_0 = 0.30$. He designed a one-sided fixed sample binomial test with specified Type I error probability $\alpha = 0.05$ and power $\beta = 0.80$ at the alternative H_1: $\varphi = \varphi_1 = 0.15$. The associated sample size was $n = 59$. For $n = 59$, the binomial test rejects H_0 at level 0.05 if the number of errors a is less than 12. Note that, due to the discreteness of the distribution, the actual error rate is only 0.035: $Pr(a < 12 \,|\, H_0 : \varphi = 0.30) = 0.035$. Similarly, the actual power is greater than 0.80: $Pr(a < 12 \,|\, H_1 : \varphi = 0.15) = 0.834$.

I convinced my colleague that it would be preferable to stop the experiment as early as possible if the method was likely to be not effective. Consequently he planned an interim analysis after 20 subjects have been included. Since the traditional Neyman-Pearson framework requires specification of all possibilities in advance, he designed a stochastically curtailed test.

The notations are summarized in the following table.

	Number of errors	Number of successes	Sample size
Current data at interim stage	a_1	$n_1 - a_1$	$n_1 = 20$
Future data	a_2	$n_2 - a_2$	$n_2 = 39$
Complete data	$a = a_1 + a_2$	$n - a$	$n = 59$

2.2 Stochastically Curtailed Testing

The idea of stochastic curtailment is a direct extension of the idea of simple curtailment (Alling, 1963) whereby an experiment could be stopped as soon as the result is inevitable. Stochastic curtailment suggests that an experiment be stopped at an interim stage when the available information determines the outcome of the experiment with high probability under either H_0 or H_1. The "conditional power" at interim analysis is defined as the probability, given φ and the available data, that the test reject H_0 at the planned termination. Lan *et al.* (1982) argue that the procedure can be used formally as a stopping rule. At interim analysis, termination occurs to reject H_0 if the conditional power at point φ_0 is high, formally if it is greater than a specified constant γ between 0.5 and 1. In our example, even if no error has been observed after 20 observations, the conditional probability of rejecting H_0 at the planned termination is less than $\frac{1}{2}$:

$$Pr(a < 12 \,|\, \varphi = 0.30 \text{ and } a_1 = 0) = Pr(a_2 < 12 \,|\, \varphi = 0.30) = 0.482 < 0.50.$$

Similarly, early termination may be allowed to accept H_0 if the conditional power at point φ_1 is less than $1-\gamma'$, where γ' is another specified constant between 0.5 and 1. For instance, if eight errors have been observed after 20 observations and if $\gamma' = 0.80$, this rule suggests stopping and accepting H_0:

$$Pr(a < 12 \,|\, \varphi = 0.15 \text{ and } a_1 = 8) = Pr(a_2 < 4 \,|\, \varphi = 0.15) = 0.143 < 0.20.$$

A criticism addressed to this procedure is that there seems little point in considering a prediction which is based on hypotheses that may be no longer fairly plausible. For instance, the stopping criterion for stopping to reject H_0 involves the probability conditional on φ_0, a value of φ which is not supported by the available data. In fact the procedure ignores the knowledge about φ accumulated by the time of the interim analysis.

2.3 The Predictive Power Approach

Many authors have advocated calculating the "predictive power", averaging conditional power over values of φ in a Bayesian calculation (see e.g., Herson, 1979; Spiegelhalter *et al.,* 1986). We are led to a Bayesian approach, but still with a frequentist test in mind. Formally, the prediction uses the posterior distribution of φ, given a prior and the data available at the interim analysis.

For the inference about a proportion, the calculations are particularly simple if we choose a conjugate *Beta* prior distribution (Lecoutre *et al.,* 1995):

Prior:	$\varphi \sim \mathit{Beta}(a_0, b_0)$
Interim posterior:	$\varphi \,|\, a_1 \sim \mathit{Beta}(a_0 + a_1, b_0 + n_1 - a_1)$
Predictive:	$a_2 \,|\, a_1 \sim \mathit{Binomial\text{-}Beta}(a_0 + a_1, b_0 + n_1 - a_1; n)$

A vague or noninformative prior is generally considered. Here, I have retained a *Beta* prior with parameters $a_0 = 1$ and $b_0 = 0$. This choice is consistent with the test procedure. I shall address this issue in more details later.

In the current example, for $a_1 = 0$, the marginal predictive probability of rejecting H_0 at the planned termination is 0.997:

$$Pr(a < 12 \,|\, a_1 = 0) = Pr(a_2 < 12 \,|\, a_1 = 0) = 0.997.$$

This probability takes into account the available data and is with no surprise largely greater than the probability conditional on $\varphi_0 = 0.30$ (0.482).

As with the conditional approach, it has been argued that the predictive power can be used formally as a stopping rule. If the same criterions are used, the rule is obviously less conservative. But, since the predictive power approach is a "hybrid" one, it's most unsatisfactory. In particular it does not give us direct Bayesian information about φ such as could be provided by a credible interval.

2.4 Reverse Stochastic Curtailing

The reverse stochastic curtailing approach has been suggested by Jennison (1992). It seems the ultimate frequentist attempt to circumvent the Bayesian predictive probabilities. Indeed, in some circumstances, parameter-free statements can be obtained by considering the sampling probability of the current data given the final data (Tan *et al.,* 1998). In our example, it can be easily verified that this probability doesn't involve φ:

$$Pr(a1 \,|\, \varphi \text{ and } a) = \tbinom{n_1}{a_1}\tbinom{n-n_1}{a-a_1}/\tbinom{n}{a}.$$

A stopping rule can be based on the following hypothetico-deductive reasoning: (1) consider the hypothesis that the final test doesn't reject H_0; (2) under this hypothesis, compute the sampling probability of at least as extreme data that the current data; (3) if this probability is small (less than $1-\gamma$), reject the hypothesis that the final test doesn't reject H_0... and reject H_0.

For instance in our example, if $a_1 = 2$, we get:

$$sup_{a \geq 12} Pr(a_1 \geq 2 \,|\, \varphi \text{ and } a) = Pr(a_1 \geq 2 \,|\, \varphi \text{ and } a = 12) = 0.141.$$

A not really surprising result is that this probability is equal to the Bayesian predictive probability that the final test doesn't reject H_0 given the current data (again for the prior $Beta(1,0)$), i.e. here:

$$Pr(a \geq 12 \,|\, a_1 = 2) = Pr(a_2 \geq 10 \,|\, a_1 = 2) = 0.141.$$

Then, for a one-sided test, the stopping rule is exactly the same as obtained in the predictive power approach.

2.5 The Bayesian Analysis

The trouble with all the above approaches is that a decision (to accept H_0 or to accept H_1) is taken at the final analysis (or eventually at an interim analysis), even if the observed proportion falls in the no-decision region $[0.15, 0.30]$. For example, if the observed proportion is 0.19 (11 out 59) H_0 is rejected, and if this proportion is 0.20 (12 out 59) H_0 is accepted. But this doesn't mean that one has proved respectively the efficacy and the inefficacy of the teaching.

What a research worker actually needs is to evaluate at any stage of the experiment the probability of some specified regions of interest and the ability of a future sample to support and corroborate findings already obtained. The Bayesian analysis addresses these issues. Bayesian methodology enables the probabilities of the prespecified regions of interest to be obtained. Such statements give straight answers to the question of effect sizes and have no frequentist counterpart. It is usual in experimental research to assume noninformative priors, as a study is expected bring evidence by itself. Indeed, in this case the posterior distribution at interim analysis and the predictive distribution for future observations are based solely on the data. Furthermore alternative choices of priors may helf refining inference.

As an example of interim analysis in our situation, consider the case where 10 errors are observed after 20 observations. Assuming the Jeffreys' noninformative prior $\varphi \sim Beta(\frac{1}{2}, \frac{1}{2})$ (see Section 2.6 for other choices), we get the interim posterior $Beta(10.5,10.5)$. In this case it is very likely that the teaching method is ineffective: given the current data, there is a 0.971 posterior probability that $\varphi > 0.30$ but only a 0.0001 probability that $\varphi < 0.15$. As a summary to help in the decision whether to terminate the experiment, it is useful to assess the predictive probability of confirming this conclusion. Here the conclusion of inefficacy ($\varphi > 0.30$) will be confirmed with a posterior probability of at least 0.95 for a total number of errors greater than or equal to 24 out 59. Given the current data, there is about 87% chance, of observing such a value.

As another example, if only two errors are observed after 20 observations, we get the interim posterior $Beta(2.5,18.5)$. In this case it is very unlikely that the teaching method is ineffective: there is a 0.018 interim posterior probability that $\varphi > 0.30$. It is much more likely that the method is effective (there is a 0.717 probability that

$\varphi < 0.15$). But no definite conclusion can be drawn at this stage. Moreover, given the current data, there is only about 30% chance of concluding to efficacy with the planned sample size.

2.6 A Brief Comment about the Choice of the Prior Distribution

Many potential users of Bayesian methods continue to think that they are too subjective to be scientifically acceptable. However frequentist methods are full of more or less *ad hoc* conventions. Thus the traditional p-value is based on the samples that are "more extreme" than the observed data (under the null hypothesis). But, for discrete data, it depends on whether include the observed data or not. For instance, the usual binomial test is conservative. But if the observed data are excluded, the test becomes liberal. A typical solution to overcome this problem consists in considering a *mid-p*-value (Berry and Armitage, 1995), but it has only *ad hoc* justifications.

Obviously, in this case the choice of a noninformative prior distribution cannot avoid conventions. But the particular choice of such a prior is *an exact counterpart* of the arbitrariness involved within the frequentist approach. So, in our situation, the observed significance levels of the inclusive and exclusive conventions are exactly the posterior Bayesian probabilities that φ is greater than φ_0 respectively associated with the $Beta(1,0)$ and $Beta(0,1)$ priors. The Jeffreys' prior gives an intermediate value close to the observed *mid-p*-value. Relevant references are Bernard (1996), Walley (1996), and Lecoutre and Charron (2000).

Note again that when the interim analysis suggests a given conclusion, another line of attack is to investigate the impact of skeptical or handicap prior distributions (Spiegelhalter *et al.,* 1994). In this way, the experiment will only stop if the partial data give sufficient evidence to counterbalance it.

3 INFERENCE ABOUT A DIFFERENCE BETWEEN MEANS

3.1 Example 2

Consider again an experiment designed to test the efficacy of a new teaching method. Two groups (new method *vs.* old method) of 30 subjects each are compared. The dependent variable is the numerical score to an evaluation test. The new method is considered as effective if the raw difference δ is more than +3, and ineffective otherwise. An interim analysis is conducted after 15 subjects in each group have been included.

Table 1. *The four examples of interim results.*

			Posterior probability			
Case	d_1	t_1	$\delta < -3$	$\lvert\delta\rvert < 3$	$\delta > +3$	Conclusion
1	+6.07	+3.674	< 0.001	0.037	0.963	*effective*
2	+6.07	+0.683	0.158	0.208	0.634	*no firm conclusion*
3	+1.52	+3.674	< 0.001	0.999	0.001	*ineffective*
4	+1.52	+0.683	0.026	0.718	0.256	*no firm conclusion*

Four examples of interim results are constructed by crossing the outcome of the usual two-sided t test of the null hypothesis $\delta = 0$ (significant, $t_1 = 3.674$, $p = 0.001$ *vs.* nonsignificant, $t_1 = 0.683$, $p = 0.50$) and the observed mean difference (large, $d_1 = 6.07$ *vs.* small, $d_1 = 1.52$). These data are analyzed using the usual noninformative prior distribution for comparing two normal means with equal variances (see e.g. Box and Tiao, 1973). The results are summarized in Table 1. Here again Bayesian statements bypass the shortcomings of usual null hypothesis significance tests and give straight answers to the question of the investigator. Given the interim data, cases 1 and 3 lead to the respective conclusions "effective" and "ineffective". On the contrary, cases 2 and 4 cannot lead to firm conclusions, because of the great variability observed.

3.2 Predictive Procedures

Let us consider first the question of predicting the separate results of the second part of the experiment. The notations are summarized in the following table.

	observed difference	t test statistic	lower credible limit	*df*
First part	d_1	t_1	$\underline{\ell}_1$	$q_1 = 28$
Second part	d_2	t_2	$\underline{\ell}_2$	$q_2 = 28$
Whole experiment	$d = (d_1 + d_2)/2$	t	$\underline{\ell}$	$q = 58$

Theoretical results are given in Lecoutre (1996, 1999). Given the data in hand, the predictive distribution for the difference d_2 is a *generalized* t distribution that depends only on d_1 and t_1:

$$d_2 \mid \text{ interim data} \sim t_{q_1}(d_1, 2(d_1/t_1)^2).$$

The predictive distribution for the Student's t test statistic t_2 is a distribution which I called K-prime (see Appendix):

$$t_2 \mid \text{ interim data} \sim K'_{q_1,q_2}(t_1, 2) \text{ (here } q_2 = q_1 = 28).$$

In the same way the predictive distribution for the $100(1-\alpha)\%$ Bayesian lower credible limit involves a K-prime distribution:

$$\underline{\ell}_2 = d_2 - \tfrac{d_2}{t_2} t_{q_2,1-\alpha} \mid \text{ interim data} \sim d_1 - \tfrac{d_1}{t_1} K'_{q_2,q_1}(t_{q_2,1-\alpha}, 2)$$

where $t_{q_2,1-\alpha}$ is the $(1-\alpha)$-percentile of the standard Student's t distribution with q_2 degrees of freedom. Note that the predictive distributions depend only on d_1 and t_1. This property illustrates the great generality of these formulae. All these results can be easily generalized to future data with a different sample size.

The final results for the whole data can be predicted as well. The predictive distribution for the final difference d (given the interim data) is deduced from the predictive distribution for d_2 (since $d = (d_1+d_2)/2$). The predictive distributions for the test statistic t and the credible limit $\underline{\ell}$ evaluate the chances that the experiment will end up showing a conclusive result, in the light of the observations already made. They are not available in closed form but they can be easily simulated.

For each of the four cases we can compute the predictive probabilities of obtaining, respectively for the separate data of the second part of the experiment and for the whole data: (1) an observed difference (resp. d_2 and d) greater than +3; (2) a posterior probability of a positive difference greater than 0.95, that is equivalently a significant t test at one-sided level $\alpha = 0.05$ (resp. $t_2 > +1.701$, 28 *df* and $t > +1.672$, 58 *df*); (3) a 95% lower credible limit (resp. $\underline{\ell}_2$ and $\underline{\ell}$) greater than +3. These probabilities are given in Table 2.

Table 2. *Predictive probabilities for the four examples of interim results*

Future data only

Case	d_1	t_1	$Pr(d_2 > +3)$	$Pr(t_2 > +1.701)$	$Pr(\underline{\ell}_2 > +3)$
1	+6.07	+3.674	0.900	0.902	0.543
2	+6.07	+0.683	0.596	0.241	0.174
3	+1.52	+3.674	0.009	0.902	< 0.001
4	+1.52	+0.683	0.321	0.241	0.051

Whole data

Case	d_1	t_1	$Pr(d > +3)$	$Pr(t > +1.672)$	$Pr(\underline{\ell} > +3)$
1	+6.07	+3.674	0.993	0.998	0.819
2	+6.07	+0.683	0.686	0.244	0.120
3	+1.52	+3.674	< 0.001	0.998	< 0.001
4	+1.52	+0.683	0.177	0.244	0.005

The more impressive finding is that, for Cases 3 and 4, it is very unlikely that the conclusion of efficacy could be asserted at the end of the study. In particular for Case 3 this might reinforce the decision to stop the experiment suggested by the conclusion of inefficacy obtained at interim analysis. It is enlightening to contrast this result with the very high probability of a significant result at interim analysis being confirmed with additional data.

3.3 Evaluating the Sample Size

The predictive approach can also be used to evaluate, at the time of planning experiment, if a given sample size is appropriate for demonstrating a given conclusion. When prior information is available, in particular from a pilot study, the predictive probabilities give a useful summary to help in the choice of the sample size.

Bayesian approaches to sample size determination are especially discussed in Adcock (1997) and Joseph and Bélisle (1997). Joseph and Bélisle distinguish between a fully Bayesian approach and a mixed Bayesian-likelihood approach. The former utilizes the prior information for both the derivation of the predictive distribution of the data and the posterior inference. The latter only uses the prior information in order to evaluate the sample size while retaining a noninformative prior for the data analysis itself. The above procedures for predicting the separate results of the second part of the experiment and the final results for the whole data respectively apply to these two issues.

4 CONCLUSION

Time's up to come to a positive agreement for procedures of experimental data analysis that bypass the common misuses of null hypothesis significance testing, while filling up its role of "an aid to judgement" which "should not be confused with automatic acceptance tests, or 'decision functions' " (Fisher, 1990/1925, page 128). This agreement should meet scientists' requirements, in particular the need for objective statements and the need for procedures about effect sizes. Undoubtedly, there is an increasing acceptance that Bayesian inference can be ideally suited for this purpose.

But, unfortunately, the contribution of Bayesian inference to experimental data analysis and scientific reporting has been obscured by the insistence of many authors for pointing out the merits of the subjective decision theoretic Bayesian conception. "But the primary aim of a scientific experiment is not to precipitate decisions, but to make an appropriate adjustment in the degree to which one accepts, or believes, the hypothesis or hypotheses being tested" (Rozeboom, 1960).

It must be emphasized that Bayesian procedures have also an important contribution to inference and data analysis. "A widely accepted objective Bayes theory, which fiducial inference was intended to be, would be of immense theoretical and practical importance. A successful objective Bayes theory would have to provide good frequentist properties in familiar situations, for instance, reasonable coverage probabilities for whatever replaces confidence intervals" (Efron 1998, page 106).

Within this perspective, Bayesian predictive probabilities are a particularly useful device to communicate with the investigators. They give them a very appealing method to answer essential questions such as: "how big should be the experiment to have a reasonable chance of demonstrating a given conclusion?"; "given the current data, what is the chance that the final result will be in some sense conclusive, or on the contrary inconclusive?" These questions are unconditional in that they require consideration of all possible value of parameters. Whereas traditional frequentist practice doesn't address these questions, predictive probabilities give them direct and natural answer. In particular predictive procedures can be used to illustrate the effects of planning an experiment with a too small sample size, and to aid the decision to abandon an experiment if the predictive probability appears poor.

A fundamental property is that predictive probability statements are conditional on the current data, and then valid independently of the stopping rule. In more complex situations these statements require heavy computations which have long been an impediment to their use, but are now easily affordable. Predictive procedures are suitable for many issues, for instance: planning an experiment from a pilot study whose data are or are not included in the final analysis; conducting interim analyses (planned or unplanned); continuing an experiment beyond its initially planned term. According to the situation the prediction can be explicitly based on either the hypotheses used to design the experiment, expressed in terms of the prior distribution, or on partial available data, or on both.

REFERENCES

Adcock, C.J. (1997). Sample size determination: A review. *The Statistician* **46**, 261–283.

Alling D.W. (1963). Early decision in the Wilcoxon two-sample test. *J. Amer. Statist. Assoc.* **5**8, 713–720.

Bernard, J.-M. (1996). Bayesian interpretation of frequentist procedures for a Bernoulli process. *Amer. Statist.* **5**0, 7–13.

Berry, D.A. (1991). Experimental design for drug development: a Bayesian approach. *J. Biopharmaceutical Statist.* **1**, 81–101.

Berry, G. and Armitage, P. (1995). Mid-P confidence intervals: a brief review. *The Statistician* **4**4, 417–423.

Box, G.E.P. and Tiao, J.W. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

Dignam, J.J., Bryant, J., Wieand, H.S., Fisher, B. and Wolmark, N. (1998). Early stopping of a clinical trial when there is evidence of no treatment benefit: protocol B-14 of the National Surgical Adjuvant Breast and Bowel Project. *Controlled Clin. Trials* **19**, 575–588.

Efron, B. (1998). R.A. Fisher in the 21st century (with discussion). *Statist. Sci.* **1**3, 95–122.

Fisher, R. A. (1990). *Statistical Methods, Experimental Design and Scientific Inference* (re-issue edited by J.H. Bennet with a foreword by F. Yates). Oxford: University Press.

Joseph, L. and Bélisle, P. (1997). Bayesian sample size determination for normal means and differences between normal means. *The Statistician* **46**, 209–226.

Lan, K.K.G., Simon, R. and Halperin, M. (1982). Stochastically curtailed tests in long-term clinical trials. *Comm. Statist. Sequential Analysis* **C1**, 207–219.

Lecoutre, B. (1996). *Traitement Statistique des Donnes Exprimentales: Des Pratiques Traditionnelles aux Pratiques Baysiennes* (with Bayesian Windows programs by B. Lecoutre and J. Poitevineau, freely available at Internet address http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/pac.htm or mail to bruno.lecoutre@univ-rouen.fr). Montreuil, FR: CISIA.

Lecoutre, B. (1999). Two useful distributions for Bayesian predictive procedures under normal models. *J. Statist. Planning and Inference* **77**, 93–105.

Lecoutre, B. and Charron, C. (2000). Bayesian procedures for prediction analysis of implication hypotheses in 2×2 contingency tables. *J. Educational & Behavioral Statist.* **25**, 185–201.

Lecoutre, B., Derzko, G. and Grouin, J.-M. (1995). Bayesian predictive approach for inference about proportions. *Statist. in Medicine* **14**, 1057–1063.

Rozeboom, W.W. (1960). The fallacy of the hypothetico-deductive significance test. *Psychological Bul.* **57**, 416–428.

Spiegelhalter, D.J., Freedman, L.S. and Parmar, M.K.B. (1994). Bayesian approaches to randomized trials. *J. Roy. Statist. Soc. A* **157**, 357–416.

Walley, P. (1996). Inferences from multinomial data: learning about a bag of marbles (with discussion). *J. Roy. Statist. Soc. B* **5**8, 3–57.

APPENDIX: THE K' DISTRIBUTION

If the three real variables x, g^2 and h^2, with g^2 and h^2 independent, are such that

$$x|g^2, h^2 \sim N\left(a\frac{g}{h}, b^2h^2\right) \qquad g^2 \sim \frac{\chi_q^2}{q} \qquad h^2 \sim \frac{\chi_r^2}{r} \qquad (g > 0, h > 0)$$

then x has the K' distribution, with q and r degrees of freedom and with parameters a and b^2: $x \sim K'_{q,r}(a, b^2)$. The K' distribution includes as a particular case the noncentral t distribution (for $q = \infty$).

b appears as a scale factor: $K'_{q,r}(a, b^2) = bK'_{q,r}\left(\frac{a}{b}, 1\right)$. Hence it is sufficient to study the particular case $b^2 = 1$. If $x \sim K'_{q,r}(a, 1)$ its probability density function is defined for q and r positive real numbers and can be expressed as:

$$p(x) = \frac{1}{\sqrt{\pi r}} \frac{1}{\Gamma(\frac{q}{2})\Gamma(\frac{r}{2})} \left(\frac{q}{q+a^2}\right)^{\frac{q}{2}} \left(\frac{r}{r+x^2}\right)^{\frac{r+1}{2}} \times \sum_{j=0}^{+\infty} \frac{1}{j!} \Gamma\left(\frac{q+j}{2}\right) \Gamma\left(\frac{r+j+1}{2}\right) \left(\frac{4}{(q+a^2)(r+x^2)}\right)^{\frac{j}{2}} (ax)^j$$

From the *pdf* the cumulative distribution function can be expressed in terms of infinite series of multiples of incomplete beta function ratios, leading to simple and efficient algorithms for numerical computations. More details are given in Lecoutre (1999).

Bruno Lecoutre
ERIS, Laboratoire de Mathématiques Raphaël Salem,
UMR 6085 C.N.R.S. et Université de Rouen, Mathématiques, Site Colbert,
76821 Mont-Saint-Aignan Cedex, France.
E-mail: bruno.lecoutre@univ-rouen.fr.
Internet: http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/Eris.htm