A BAYESIAN PREDICTIVE PROCEDURE FOR TWO STEPS EXPERIMENTAL TRIALS.

Submited on 25/10/2016 – Accepted on 05/05/2019

Abstract

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 an inconclusive one. The prediction can be explicitly based on either the hypothesis used to monitor the experiment expressed either in terms of prior distribution, on partially available data, or on both. In this paper, we propose a Bayesian predictive methodology based on two steps which can be used to develop an adaptive design for the experimental trials. This procedure does not require intensive computation and comprehensive simulations. We have used the non-informative prior to give evidence on the objectivity of the experimental data.

Keywords: *Bayesian prediction, p-value, clinical trials, Monte-Carlo simulations, exponential models, stopping rule.*

MSC2010: *62K05*

Z. DJERIDI [1]

H. MERABET [2]

1 Mathematics Department, Jijel University, Jijel, Algéria.

2 Laboratory of Applied Mathematics and Modeling, Mathematics Department, Frères Mentouri University Constantine, Algeria.

INTRODUCTION

The Bayesian approach brings a major flexibility to the statistical methodology of the experimental trials. A major strength of the Bayesian paradigm is the ease with which one can make predictions about future observations. In particular, we are interested in using this approach aiming at prediction in the context of experimental trials. The role which the predictive probability plays in the design and monitoring trials is important in several fields (reliability of systems medicine, biology, ecology,...) ([18], [9]).

Bayesian predictive procedures have made an important contribution to inference and data analysis. Within this perspective, Bayesian predictive probabilities are a particularly useful device to communicate with the investigator. They give them a very appealing method to answer essential questions such as: "Given the current data, what is the chance that the final result will be in some sense conclusive, or on the contrary inconclusive? This question is unconditional because it requires consideration of all possible values of parameters. Whereas, traditional frequentist practice does not address these questions, predictive probabilities give them direct and natural answers. In particular, predictive procedures can be used to illustrate the effects of planning an experiment with a very small sample size, and to aid in the decision to abandon an experiment early.

For example, in phase II cancer trials, it is undesirable to stop a study early when the test drug is promising, and it is desirable to terminate the study as early as possible when the test treatment is not effective due to ethical consideration. For this purpose, a multiple stage design single arm trial is often employed to determine whether the test treatment is promising for further testing ([5]; [3]; [8]).

The methodology adapted to the context of clinical trials is characterized by many constraints and unsatisfactions and forms the subject of a deep and continuous development ([13]; [2]; [10]; [8]). One of the reasons for such interest is likely to emanate from the fact that public health authorities are responsible for the permission of putting the drugs into market and they play a primordial role in the elaboration of a rigorous methodology of clinical trials, taking into consideration the views of all the actors in this field (industries, public institutes of research, hospitals and scientific journals).

The primary goal clinical trial is to evaluate the efficiency and the tolerance of a new medical treatment. They are characterized by complex actions that cannot be readily modeled and they do not depend solely on statistical considerations (see for example [3]). In this situation, we, often, get primary experimental information in the form of step I, then we need to confirm some results ([3]; [8]; [9]). Formally, we consider the following situation: Using the data of the first sample, we can plan an experiment (a new sample) in a way to have good chances to get the intended conclusion if the experimentation is not discarded.

The main objective of this paper is to provide a hybrid Bayesian-frequentist procedure for two stages designs to test the efficacy of a new therapy. This procedure is based on the concept of the index of satisfaction which is a decreasing function of the p-value, and we envisage, given the available data, to calculate a predicted satisfaction of this index by considering the previous observations using Bayesian approach. Many authors have advocate to use Indexes in such situation as Lecoutre et al:1995, Merabet ,2004, Merabet and Labdaoui, 2015, Djeridi and Merabet, 2015, ([13][15][16][6]) because of their simplicity and flexibility in measuring the degree of satisfaction in the case of obtaining a significant result. We used this index to find a stopping rule for designing a phase II clinical trials. In this situation, we are led to a Bayesian approach butl with a frequentist test in mind.

Bayesian posterior probability, which is the probability that the parameter is contained within a meaningful region, is the best tool to answer the following question addressed by interim monitoring:" Is there convincing evidence in favor of the null alternative hypothesis?" On the other way, using stochastic curtailment methods such as, predictive probability and prediction of satisfaction, we give answers to the question: "Is the trial likely to show convincing evidence in favor of the alternative hypothesis if additional data are collected?" Because we deal, here, with the prediction of what evidence will be available at later stages in the study ([18]). If the futility is defined as a trial being unlikely to achieve its objective, then it is inherently a prediction problem and is best addressed using prediction of satisfaction.

To illustrate our procedure, we studied several exponential models choosing a non-informative prior to highlight the analysis objectivity of the experimental data. It is usual in experimental research to assume non-informative priors, as a study is expected to bring evidence by itself ([12]). Bayesians use at or otherwise improper non-informative priors in situations where prior knowledge is vague relative to the information in the likelihood, or in settings where we want data (and not the prior) to dominate the determination of the posterior ([13]). Furthermore, the Jeffrey's priors are a particular choice because it is an exact counterpart of the arbitrariness involved within the frequentist approach (section 2.3). The numeric calculations and the simulation results are presented in the form of binary outcomes for phase II clinical trials and Gaussian model.

2. STATISTIC METHODS

Bearing in mind that, that the experimental context consists of two successive experimentations, of results $\omega' \in \Omega'$ and $\omega'' \in \Omega''$, which are in general carried out independently. Their distributions built in the framework of a well established model, depend on a parameter $\theta \in \Theta$, only ω" is used to found the official conclusion of the study and to determine the user's satisfaction denoted $\phi(\omega")$ (and on the choice of the decreasing function L about which we will come back in 2.4). But, on the basis of the result ω' of first step clinical trial, it is useful to anticipate what the satisfaction will be after the second step. In our study, this prediction is carried out in a Bayesian context, i.e., based on the choice of a prior probability on Θ.

We denote:

P_Θ: Prior probability on Θ.

$P_\Theta^{\omega'}$: P*oste*rior probability on Θ, *based* on the result of the *fi*rst step.

$P_{\Omega"}^{\theta}$: Sampling distribution of the second step.

$P_{\Omega"}^{\omega'}$: Probability on ω", conditioned by the result of the first step.

2.1 General Case:

In this experimental context, we will introduce the concept of the index of satisfaction relative to a hypothesis test where the null hypothesis is of type $\theta \leq \theta_0$, in a framework where such a test may be constructed using a reasonable test function.

Let us suppose a model $(P_\theta)_{\theta\in\Theta}$_and test a null hypothesis Θ_0 versus an alternative Θ_1 , defined by an application $\Psi(:\Theta \to IR)$. We suppose that a point t_0 exists so that:

$$\boldsymbol{\theta \in \Theta_0 <=> \Psi(\theta) \leq t_0}$$

Otherwise, let us suppose that we have a real application $\boldsymbol{\xi(\Omega" \to IR)}$ such that*:*

$$\boldsymbol{\Psi(\theta_1) < \Psi(\theta_2) => \forall t, P_{\theta_1}[\xi \leq t] \geq P_{\theta_2}[\xi \leq t]}$$

Then a test of a level $\boldsymbol{\alpha}$, of $\boldsymbol{\Theta_0}$ versus $\boldsymbol{\Theta_1}(=\{\boldsymbol{\theta};\ \boldsymbol{\Psi(\theta) > t_0}\})$ is defined by rejecting the hypothesis if the experimental result, y, verifies that $\boldsymbol{\xi(y) > g(\alpha)}$, **where** $\boldsymbol{g(\alpha)}$ is the (1-α)-quantile of the distribution of $\boldsymbol{\xi}$ when $\boldsymbol{\Psi(\theta_0) = t_0}$.

Indeed, the critical region, C, of this test, is then the set of the observations y, such that $\boldsymbol{\xi(y) > g(\alpha)}$ and because of the stochastic creasing of the distributions of ξ, we have, for all θ_0 such that $\boldsymbol{\Psi(\theta_0) = t_0}$,

$$\boldsymbol{\forall \theta \in \Theta_0, P_\theta(C) \leq P_{\theta_0}(C) \leq \alpha}$$

Moreover, if $\boldsymbol{\theta_1 \in \Theta_1}$ and $\boldsymbol{\theta_2 \in \Theta_1}$ with $\boldsymbol{\Psi(\theta_1) < \Psi(\theta_2)}$,

then $\boldsymbol{P_{\theta_1}(C) \leq P_{\theta_2}(C)}$ which means that the power function increases with $\boldsymbol{\Psi(\theta)}$.

2.2 Advantage of the p-value

The *p-value* is a measure of statistical evidence that appears in virtually all experimental research papers. Its interpretation is made extraordinarily difficult because it is not part of any formal system of statistical inference. As a result, the *p-value*'s inferential meaning is widely and often wildly misconstrued (see for example [1]), a fact that has been pointed out in innumerable papers and books appearing since at least the 1940s. S. Goodman (2008) [7] reviewed a dozen of these common misinterpretations and explained why each is wrong. He, also, reviewed the possible consequences of these improper understandings or representations of its meaning.

The *p-value* is defined as the probability, under the assumption of no effect or no difference (the null hypothesis), of obtaining a result equal to or more extreme than what was actually observed (Fig. 1).

The curve represents the probability of every observed outcome under the null hypothesis. The *p-value* is the probability of the observed outcome $(\boldsymbol{x})$ plus all "more extreme" outcomes, represented by the shaded "tail area".

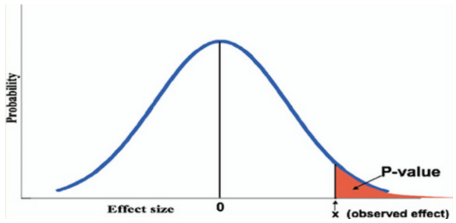

Figure 1: Graphical depiction of the definition of a (one-sided) *p-value*

In some cases particularly in the two sided testing problems, there are difficulties in defining a *p-value*. To eliminate these difficulties, we follow Fisher and define it as follows:

Definition1: The *p-value* associated with a test is the smallest significance level $\boldsymbol{\alpha}$ for which the null hypothesis is rejected. That mean, if

$$\boldsymbol{p(x) = \inf\{\alpha; x \in \Omega^{\alpha}\}}$$

Which eliminates ambiguities (as long as $\boldsymbol{\Omega^{\alpha}}$ is specified for each $\boldsymbol{\alpha}$).

Remark 2: According to Hwang et al. (1992) [11], the *p-value* is admissible in the one sided tests under some conditions, and it is a minimax rule under absolute error loss. They presented a number of examples in which the *p-value* is generalized Bayes, hence admissible under squared error loss function that it is a reasonable measure of accuracy. *P-values* are thus adaptive procedures that can be acceptable from a frequentist point of view (17).

2.3 A brief comment about the choice of the prior distribution

A prior distribution captures all of the information known about the parameters θ before we collect data.

Along with the likelihood function, it is one of the two key components of a Bayesian model.

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, for example, is conservative. But if the observed data are excluded, the test becomes liberal (12). A typical solution to overcome this problem consists in considering non-informative prior distributions ([17], [9]).

These priors cannot be expected to represent exactly total ignorance about the problem at hand, but should rather be taken as reference or default priors, upon which everyone could fall back when the prior information is missing. Then, these particular prior distributions must be derived from the sample distribution, since this is the only available information.

While there are a number of formalisms for developing non-informative prior distributions, one of the most common uses Jeffreys' rule, which results in a distribution often called a Jeffreys' prior. Define the expected Fisher information as:

$$\boldsymbol{I(\theta) = -E\left[\frac{d^2 \log(f(y|\theta))}{d\theta^2}\right]}$$

Jeffreys' rule defines a non-informative prior as

$$\boldsymbol{f(\theta) = [I(\theta)]^{1/2}}$$

More fundamentally, the choice of a prior depending on Fisher information is justified by the fact that $\boldsymbol{I(\theta)}$ is widely accepted as an indicator of the amount of information brought by the model (or the observation) about $\boldsymbol{\theta}$ ([17]).

2.4 Index of satisfaction

This notion finds its origins in situations where the statistician, who carries out a test, "wishes" to detect a significant effect, *i.e.*, to reject the null hypothesis $\boldsymbol{H_0}$. Correspondingly, this statistician is, especially, more satisfied if, in function of the experimental results, this effect seems to be more significant.

2.4.1 Rudimentary index:

Being $\boldsymbol{\alpha}$ fixed, let a test of level $\boldsymbol{\alpha}$ be defined by a critical region $\boldsymbol{\Omega_1^{"(\alpha)}}$. A first index of satisfaction is defined by:

$$\boldsymbol{\phi(\omega") = 1_{\Omega_1^{"(\alpha)}}(\omega")}$$

At a fixed $\boldsymbol{\omega'}$, the prediction is:

$$\boldsymbol{\pi(\omega') = P_{\Omega"}^{\omega'}\left(\Omega_1^{"(\alpha)}\right)}$$

$$\boldsymbol{= \int_{\Omega_1^{"(\alpha)}} P_{\Omega"}^{\theta}(\Omega_1^{"(\alpha)}) P_{\Theta}^{\omega'}(d\theta)}$$

Where $\boldsymbol{P_{\Omega"}^{\theta}(\Omega_1^{"(\alpha)})}$ is the power of the test for the value of $\boldsymbol{\theta}$.

The weakness of this index is that it expresses a satisfaction in "all or nothing" fashions (significant or not significant).

2.4.2 Improved index

It is more interesting to take into account to what level will the result always appear significant. This is what the users highlight by giving, at the end of the test procedure, not only the conclusion in terms of "all nor thing" but also the smaller value of threshold for which the obtained result will be considered significant that is, from the point of view of the theory of the test, the *p-value* given in definition 1, and it is in our case

$$\boldsymbol{p = P_{\theta_0}\left(\xi > \xi(\omega'')\right)}$$

An index of satisfaction (*IS*), for the considered test of level $\boldsymbol{\alpha}$, is then defined naturally by an application from the results set in $\boldsymbol{IR_+}$ such that:

- Takes the value $\boldsymbol{0}$ if we don't reject the hypothesis, i.e., if $\boldsymbol{\xi(y) \leq g(\alpha)}$,

- And if $\boldsymbol{\xi(y) > g(\alpha)}$, is a decreasing function of $\boldsymbol{P_{\theta_0}(\xi > \xi(y))}$, which we denote

$$\boldsymbol{L(P_{\theta_0}(\xi > \xi(y)) = L(1 - F_{\theta_0}(\xi(y)))}$$

where $\boldsymbol{F_{\theta_0}}$ is the distribution of ξ at the frontier such as $\boldsymbol{\Psi(\theta_0) = t_0}$.

Remark 3: A rudimentary index is the indicator function of the critical region (which was studied by Grouin (1994)[13]) but it does not take into account of the "*p-value*".

2.5. Prediction of satisfaction

Being α fixed, let a test of level α be defined by a critical region $\Omega_1^{''(\alpha)}$. It is more interesting to take into account to what level will be the results always appear significant. We will use the new index of satisfaction (mentioned in section 2.4.2), wich was studied by Merabet H. (2004)[15], and defined for the bayesian tests, based on the same prior P_Θ, as:

$$\phi(\omega'') = \begin{cases} 0 & si \quad \omega'' \in \Omega_0^{''(\alpha)} \\ 1 - inf\left\{\beta; \omega'' \in \Omega_1^{''(\beta)}\right\} & si \quad \omega'' \in \Omega_1^{''(\alpha)} \end{cases} \quad (1)$$

A standard situation is that where an application $\psi(\Theta \to \mathbb{R})$ is such as $\Theta_0 = \{\theta;\ \psi(\theta) \le t_0\}$ and where it also exists $\xi(\Omega'' \to \mathbb{R})$ and $g(]0,1[\to \mathbb{R})$ such that

$$\Omega_1^{''(\alpha)} = \{\omega'';\ \xi(\omega'') \le g(\alpha)\}$$

Where $g(\alpha)$ is the $(1-\alpha)$ quantile of the distribution of ξ when $\psi(\theta_0) = t_0$.

Using the *p-value;*

$\boldsymbol{p(\omega'') = P_{\theta_0}(\xi > \xi(\omega''))}$,

the index of satisfaction is thus defined naturally as

$$\phi(\omega'') = \begin{cases} 0 & if \quad \xi(\omega'') \le g(\alpha) \\ L(p(\omega'')) & if \quad \xi(\omega'') > g(\alpha) \end{cases} \quad (2)$$

Where L is a decreasing function.

Let F_{θ_0} be the distribution of ξ at the frontier, *i. e.*, for any θ_0 such as $\psi(\theta_0) = t_0$, the index of satisfaction is defined by:

$$\phi(\omega'') = \begin{cases} 0 & if \quad p(\omega'') \ge 1 - \alpha \\ L\left(1 - F_{\theta_0}(\omega'')\right) & else \end{cases} \quad (2)$$

The prediction of satisfaction is then given by:

$$\boldsymbol{\pi(\omega') = \int_{\Omega_1^{''(\alpha)}} \phi(\omega'')\, P_{\Omega''}^{\omega'}(d\omega'')}$$

$$\boldsymbol{= \int_{\Theta} \left(\int_{\Omega_1^{''(\alpha)}} \phi(\omega") P_{\Omega''}^{\theta}(d\omega'') \right) P_{\Theta}^{\omega'}(d\theta)}$$

$$\boldsymbol{= \int_{\{\omega'';\xi(\omega'')>g(\alpha)\}} L\left(1 - F_{\theta_0}(\omega'')\right) P_{\Omega''}^{\omega'}(d\omega'')}$$

It is noticed that

$\int_{\Omega_1^{''(\alpha)}} \boldsymbol{\phi(\omega") P_{\Omega''}^{\theta}(d\omega'')}$ generalizes the power of the test in the logic of the index of satisfaction proposed.

We can generalize this procedure to a family of limited indexes defined by

$\boldsymbol{L(p) = (1-p)^l}$, where $\boldsymbol{l \ge 0}$.

It is preferable to choose limited indexes because they are easy to interpret.

Remark 4:

1. If $\boldsymbol{l = 1, 1 - \phi(\omega'')}$ is the p-value.

2. In the case where $\boldsymbol{l = 0}$, one finds the indicator function of the critical region which is the rudimentary index of Grouin ([13]).

3. APPLICATION

Djeridi and Merabet (2016) [6] proposed to calculate explicitly or numerically the prediction of satisfaction in several exponential models: binomial, poisson, gamma and gaussian for $L(p) = (1-p)$ in the case of a test of threshold, α where the null hypothesis is

$H_0: \theta \le \theta_0$. We are led to a Bayesian approach, (but still with a frequentist test in mind) when the prior distribution of the unknown parameter θ is non-informative. In this case, we choose the *Jeffreys' prior* (section 2.3).

3.1 Binomial distribution

Let us suppose that all random variables are independent $X'_i\ (1 \le i \le k)$ and $X''_j\ (1 \le j \le n)$ has a bernoulli distribution $B(\theta)$, where θ is unknown.

Then $\omega' = \sum_{i=1}^{k} X'_i$ has a binomial distribution $B(k, \theta)$ and $\omega'' = \sum_{j=1}^{n} X''_j$ has a binomial distribution $B(n, \theta)$.

By choosing the non-informative prior of Jeffrey's for θ:

$$f(\theta) = \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \propto Beta\left(\frac{1}{2}, \frac{1}{2}\right)$$

The posterior density of θ given ω' is a beta distribution $Beta\left(\omega' + \frac{1}{2}, k - \omega' + \frac{1}{2}\right)$ ([10]), and the predictive of ω'' given ω' is given by:

$$\nu(\omega''|\omega') = \frac{C_n^{\omega''}\beta\left(\omega'' + \omega' + \frac{1}{2}, n + k - (\omega' + \omega'') + \frac{1}{2}\right)}{\beta\left(\omega' + \frac{1}{2}, k - \omega' + \frac{1}{2}\right)}$$

Where $\beta(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$

$and\ C_x^y = \frac{x!}{y!(x-y)!}$.

Then, the index of satisfaction *(IS) i*s:

$$\phi(\omega'') = \begin{cases} 0 & if \quad \omega'' < q_0 \\ \sum_{t=0}^{\omega''-1} C_N^t {\theta_0}^t (1-\theta_0)^{N-t} & if\ \omega''is\ integer\ and\ \ \omega'' \geq q_0 \end{cases}$$

Where

$$q_0 = inf\left\{u; \sum_{t=u}^{N} C_N^t {\theta_0}^t (1-\theta_0)^{N-t} \leq \alpha\right\}$$

and, the prediction of satisfaction *(PIS)* is:

$$\pi(\omega') = \sum_{\omega''=q_0}^{N}\left(\sum_{t=0}^{\omega''-1} C_N^t {\theta_0}^t (1-\theta_0)^{N-t}\right)\nu(\omega''|\omega')$$

$$= \sum_{\omega''=q_0}^{N}\left(\sum_{t=0}^{\omega''-1} C_N^t {\theta_0}^t (1-\theta_0)^{N-t}\right) \times \frac{C_N^{\omega''}\beta\left(\omega'' + \omega' + \frac{1}{2}, N + K - (\omega' + \omega'') + \frac{1}{2}\right)}{\beta\left(\omega' + \frac{1}{2}, K - \omega' + \frac{1}{2}\right)}$$

3.2 Poisson Sampling

Let us suppose that $X'_i\ (1 \leq i \leq n)$ and $X''_j\ (1 \leq j \leq k)$ are *i.i.d.* real random variables of Poisson distribution $\mathcal{P}(\theta)$, where θ is unknown.

Then $\omega' = \sum_{i=1}^{n} X'_i$ have a Poison distribution $\mathcal{P}(n\theta)$ and $\omega'' = \sum_{j=1}^{k} X''_j$ have a Poisson distribution $B(k\theta)$. If θ has a non informative prior $f(\theta) = \theta^{-1}$

Then the posterior density of θ given ω' will be

$$f(\theta|\omega') \propto Gamma(\omega', n)$$

And the predictive of ω'' given ω' is

$$\nu(\omega''|\omega') = \frac{\Gamma(\omega' + \omega'')}{\Gamma(\omega')\omega''!}\left(\frac{n}{n+k}\right)^{\omega'}\left(\frac{k}{n+k}\right)^{\omega''}$$

The index of satisfaction is then expressed as:

$$\phi(\omega'') = \begin{cases} 0 & if \quad \omega'' < q_0 \\ \sum_{s=0}^{\omega''-1} e^{-k\theta_0}\frac{(k\theta_0)}{s!} & if \quad \omega'' \geq q_0 \end{cases}$$

Where

$$q_0 = inf\left\{s; \sum_{s=0}^{u-1} e^{-k\theta_0}\frac{(k\theta_0)}{s!} \geq 1 - \alpha\right\}$$

And the prediction of satisfaction is given by:

$$\pi(\omega') = \sum_{\omega''=q_0}^{\infty}\left(\sum_{s=0}^{\omega''-1} e^{-k\theta_0}\frac{(k\theta_0)}{s!}\right)\nu(\omega''|\omega')$$

$$= \sum_{\omega''=q_0}^{\infty}\left(\sum_{s=0}^{\omega''-1} e^{-k\theta_0}\frac{(k\theta_0)}{s!}\right)\frac{\Gamma(\omega' + \omega'')}{\Gamma(\omega')\omega''!}\left(\frac{n}{n+k}\right)^{\omega'}\left(\frac{k}{n+k}\right)^{\omega''}$$

3.3 Gamma distribution

Let us suppose that $X'_i\ (1 \leq i \leq k)$ and $X''_j\ (1 \leq j \leq n)$ are *i.i.d.* real random variables of Gamma distribution $G(p,\theta)$ where θ is unknown and p is known. Then, $\omega' = \sum_{i=1}^{k} X'_i$ have a Gamma distribution $G(kp,\theta)$ and $\omega'' = \sum_{j=1}^{n} X''_j$ have a Gamma distribution$B(np,\theta)$. Let be $K = kp$ and $N = np$.

If θ has a non informative prior $f(\theta) = \theta^{-1}$. Then the posterior density of θ given ω' will be:

$$f(\theta|\omega') \propto Gamma(K, \omega')$$

And the predictive of ω'' given ω' is:

$$\nu(\omega''|\omega') = \frac{1}{\beta(N,K)}\frac{(\omega')^{N-1}(\omega')^{K}}{(\omega'' + \omega')^{N+K}}$$

The index of satisfaction is then expressed as:

$$\phi(\omega'') = \begin{cases} 0 & if \quad \omega'' < q_0 \\ F(\omega'') = \int_0^{\omega''}\frac{(\theta_0)^N}{\Gamma(N)}t^{N-1}e^{-t\theta_0}dt & if \quad \omega'' \geq q_0 \end{cases}$$

Where $F(q_0) = 1 - \alpha$.

And the prediction of satisfaction is given by:

$$\pi(\omega') = \int_{q_0}^{\infty}\left(\int_0^{\omega''}\frac{(\theta_0)^N}{\Gamma(N)}t^{N-1}e^{-t\theta_0}dt\right)\nu(\omega''|\omega'')\mathrm{d}\omega''$$

$$= \int_{q_0}^{\infty} \left(\int_0^{\omega''} \frac{(\theta_0)^N}{\Gamma(N)} t^{N-1} e^{-t\theta_0} dt \right) \frac{1}{\beta(N,K)} \frac{(\omega'')^{N-1}(\omega')^K}{(\omega''+\omega')^{N+K}} \, d\omega''$$

This can be estimated numerically.

3.4. Gaussian model

Relying on the Central Limit Theorem, statisticians in the first half of the nineteenth century were almost always referring to the normal distribution. There are obviously many phenomena for which a normal model is not applicable, but it is still extensively used, in particular, in econometrics and in fields where the Central Limit Theorem approximation can be justified (particle reliability, etc.). In fact, the normal approximation is often justified for asymptotic reasons (see [17]). Therefore, it is of interest to study in detail this particular distribution from a Bayesian viewpoint. The corresponding calculations of the prediction being realizable by the Monte-Carlo methods (section 4).

We perform independent observations of same normal random variable $\mathcal{N}(\theta, \sigma^2)$. In all that follows, Φ and φ (resp. T_{n-1} and t_{n-1}) indicates the cumulative distribution function and the density of the distribution $\mathcal{N}(0,1)$ respectively (resp. of the student distribution $\mathcal{T}_1(n-1,0,1)$) .

The first result; $\underline{x} = (x_1, x_2, \dots, x_n)$, is a series of n observations and the second result is a series; $\underline{y} = (y_1, y_2, \dots, y_k)$.

For obvious reasons of exhaustiveness, we will base all calculations on $x = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $y = \frac{1}{k}\sum_{j=1}^{k} y_j$, of distributions $\mathcal{N}(\theta, {\sigma_1}^2)$ and $\mathcal{N}(\theta, {\sigma_2}^2)$, respectively, where ${\sigma_1}^2 = \frac{\sigma^2}{n}$ and ${\sigma_2}^2 = \frac{\sigma^2}{k}$.

We suppose here that σ^2 is unknown (so ${\sigma_1}^2$ and ${\sigma_2}^2$) (The situation where σ^2 is known was studied in [15]). We choose as a priori distribution for (θ, σ^2) the non-informative distribution $\pi(\theta, \sigma^2) = \frac{1}{\sigma}$ ([17]). We wish to test a null assumption of type $\theta \leq \theta_0$.

We use here a usual test ranging on $\underline{y}$, whose critical region is $]q_0, +\infty[$, where $q_0 = \theta_0 + {S'}_2 u_\alpha^+$, u_α^+ indicating the upper α quantile of the standard normal distribution $\mathcal{N}(0,1)$:
$\Phi(u_\alpha^+) = 1 - \alpha$ and ${S'}_2 = \frac{S_2}{\sqrt{k}}$. The posterior density associated to the prior $\pi(\theta, \sigma^2) = \frac{1}{\sigma}$ and applied to the second phase $\underline{y} = (y_1, y_2, \dots, y_k)$ is then:

$$\theta | \sigma, y, {S_2}^2 \sim \mathcal{N}\left(y, \frac{\sigma^2}{k}\right) \text{ and}$$

$$\sigma^2 | y, {S_2}^2 \sim \mathcal{IG}\left(\frac{k-1}{2}, \frac{{S_2}^2}{2}\right)$$

Where $${S_2}^2 = \frac{1}{k}\sum_{j=1}^{k}(y_j - y)^2.$$

And the predictive density of y given x is given by:

$$f_x(y) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\frac{S_1}{\sqrt{nk}}\sqrt{n+k}} \left(\frac{\frac{(y-x)^2}{S_1^2}}{\frac{kn}{(n+k)}} + 1 \right)^{-\frac{n}{2}}$$

Where ${S_1}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - x)^2$.

We identify a student distribution $\mathcal{T}_1\left(n-1, x, \frac{S_1}{\sqrt{nk}\sqrt{n+k}}\right)$

Finally the prevision of satisfaction is:

$$\pi(x) = \int_{q_0}^{+\infty} \Phi\left(\frac{y-\theta_0}{\frac{S_2}{\sqrt{k}}}\right) \times \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} \frac{1}{\frac{S_1}{\sqrt{nk}}\sqrt{n+k}} \left(\frac{\frac{(y-x)^2}{S_1^2}}{\frac{kn}{(n+k)}} + 1 \right)^{-\frac{n}{2}} dy$$

4. EXAMPLES

4.1 Application for binary outcomes:

We consider the design for real data, taken from the study of predictive probability approach ([14]; [3]; [20]) .

In a phase II trials, an investigator plans to enroll a maximum of $N_{max} = 40$ patients into the study. At a given time $\omega' = 16$ responses are observed in $k = 23$ patients. In the light of this result should the investigator continue the trial or stop it using the index of satisfaction *IS* and its prediction *PIS* if he enrolls all patients?

Assuming a prior distribution of θ as $Beta(1/2,1/2)$ and with the number of responses in future $n = 17$ patients, ω'' follows a *beta-binomial* distribution $(17, 16.5, 7.5)$. At each possible value of $\omega'' = i$, the posterior probability of θ follows a *beta* distribution $\theta|\omega', \omega'' = i \sim Beta(\frac{1}{2} + \omega' + i, \frac{1}{2} + N_{max} - \omega'' - i)$

For this example

$$\theta|\omega' = 16, \omega'' = i \sim Beta(16.5 + i, 24.5 - i)$$

In order to use the index of satisfaction *IS* for $\theta_0 = 60\%$, we have to find $q_0 = 13$, for level of significance $\alpha =$

0.05. So the index of satisfaction *IS* and its prediction *PIS* will be **Table 1**

ω''	$\phi(\omega'')$	ω'	$\pi(\omega')$
		13	0.13
		14	0.21
		15	0.29
< 13	0	16	0.40
13	0.9536	17	0.52
14	0.9877	18	0.64
15	0.9979	19	0.76
16	0.9998	20	0.86
17	1.0000	21	0.93
		22	0.98
		23	0.997

Table 1: IS and PIS for different values of ω' and ω''.

In this case, the prediction of satisfaction will be $\pi(\omega' = 16) = 0.40$ and we will reject H_0. This result is the same as the Simon's design ([20]) and the Predictive Probability design introduced by Lee and Liu (2008) [14] but with little satisfaction. On the other hand, if we take $\omega' = 20$, then $\pi(\omega' = 20) = 0.85$ and we should reject the null hypothesis with a great satisfaction, so we are more satisfied about the efficacy of the treatment.

4.1.1 PIS-procedure as a stopping rule:

This procedure can be used formally as a stopping rule for clinical trials. At interim analysis, termination occurs to reject H_0 if the prediction of satisfaction *PIS* at point $\theta_0$is high, formally, if it is greater than a specified constant γbetween 0.5 and 1.

The specific stopping criteria are typically unique to each trial and include ethical and business considerations, such as risk/ benefit considerations, available resources, opportunity cost, and overall statistical power. In the context of interim monitoring for futility, prediction of satisfaction is naturally appealing because it directly addresses the relevant question, that is, whether a trial is likely to reach its objective if continued to the planned maximum sample size.

In our example, even if we take $\gamma \geq 0.5$, than $\pi(\omega' = 17) = 0.52$ (table 1). In this case, our satisfaction will be great and we are satisfied that our treatment is promising and we should collect more information about it. Furthermore, this non-informative prior and cut-off reserve type I error which is:

$IP(\omega' > 17|\theta_0 = 0.60) = 0.05$. In this case, the actual power of this design, for an alternative of θ_1=0.80, is 0.84.

Tables 2 (a) and (b) gives the values of the prediction of satisfaction *PIS* for different values of θ_0, k, n, ω', for $\alpha = 0.05$.

		$\pi(\omega')\ for$			
k, n	ω'	$\theta_0 = 0.2$	$\theta_0 = 0.4$	$\theta_0 = 0.6$	$\theta_0 = 0.8$
23, 17	1	0.0078	0.0001	0.0000	0.0000
	2	0.0282	0.0004	0.0000	0.0000
	3	0.0688	0.0017	0.0000	0.0000
	4	0.1332	0.0051	0.0001	0.0000
	5	0.2203	0.0127	0.0004	0.0000
	6	0.3250	0.0271	0.0012	0.0000
	7	0.4392	0.0517	0.0030	0.0000
	8	0.5539	0.0898	0.0068	0.0001
	9	0.6609	0.1439	0.0141	0.0002
	10	0.7543	0.2152	0.0271	0.0006
	11	0.8308	0.3025	0.0485	0.0013
	12	0.8897	0.4027	0.0818	0.0029
	13	0.9321	0.5101	0.1303	0.0062
	14	0.9609	0.6179	0.1968	0.0125
	15	0.9790	0.7191	0.2827	0.0240
	16	0.9896	0.8073	0.3867	0.0444
	17	0.9953	0.8783	0.5045	0.0788
	18	0.9981	0.9305	0.6282	0.1343
	19	0.9993	0.9651	0.7473	0.2193
	20	0.9998	0.9852	0.8504	0.3428
	21	1.0000	0.9951	0.9279	0.5096
	22	1.0000	0.9989	0.9757	0.7132
	23	1.0000	0.9999	0.9967	0.91991

Table 2 (a): *PIS* for different values of θ_0.

		$\pi(\omega')\ for$		
k, n	ω'	θ_0 =0.2	θ_0 =0.5	θ_0 =0.7
15, 20	1	0.0274	0.0001	0.0000
	2	0.0919	0.0005	0.0000
	3	0.2038	0.0022	0.0001
	4	0.3533	0.0076	0.0003
	5	0.5178	0.0216	0.0013
	6	0.6722	0.0515	0.0043
	7	0.7984	0.1067	0.0122
	8	0.8887	0.1951	0.0304
	9	0.9456	0.3197	0.0680
	10	0.9769	0.4735	0.1373
	11	0.9917	0.6387	0.2508
	12	0.9976	0.7901	0.4144
	13	0.9995	0.9045	0.6167
	14	0.9999	0.9711	0.8197
	15	1.0000	0.9967	0.9649

Table 2 (b): *PIS* for different values of θ_0, k, n, ω', for $\alpha = 0.05$.

Larger values of θ_0, for example, 80% have a slower rate of convergence to 1 than the smaller values, for example, 20% (see figure 2 (a)), because we need more arguments to reject the null hypothesis. Also, for type I error, the rejection of the null hypothesis will be hard if we increase the level of significance (figure 2 (b)).

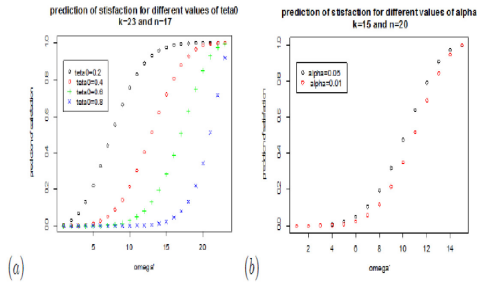

Figure 2: PIS augments slowly if θ increases (a) and if the level of significance is greater (b).

A criticism addressed to this procedure, that it does not give us direct Bayesian information about θ such as could provided by a credible interval. Also, to prove the efficacy of the treatment we should have a big probability of success. In the example above, the trial will stop for futility if less than 17 successes/23 (74%) are observed.

3.4.1. Monte Carlo's Method :

In order to carry out the calculation of $\pi(x)$ using a Monte Carlo method, and by change of variable, we rewrite it in the following form:

$$\pi(x) = [1 - T_{n-1}(a + \gamma u_\alpha^+)] \int_{-\infty}^{+\infty} \Phi\left(\frac{z - a}{\gamma}\right) \times \frac{t_{n-1}(z)}{1 - T_{n-1}(a + \gamma u_\alpha^+)} \mathbb{I}_{[a+\gamma u_\alpha^+, +\infty[} dz$$

Where $a = \frac{\sqrt{n-1}}{S'_1}(\theta_0 - x)$, $\gamma = \sqrt{n-1}\frac{S'_2}{S'_1}$

With $S'_1 = \frac{S_1}{\sqrt{\frac{nk}{n+k}}}$, $S'_2 = \frac{S_2}{\sqrt{k}}$, and $\frac{t_{n-1}(z)}{1-T_{n-1}(a+\gamma u_\alpha^+)} \mathbb{I}_{[a+\gamma u_\alpha^+, +\infty[}$

is the probability density Q deduced from the cumulative distribution function of the student distributioned by the event $[a+, +\infty[$.

The Monte Carlo method then consists in approaching $\pi(x)$by:

$$[1 - T_{n-1}(a + \gamma u_\alpha^+)] \left[\frac{1}{N}\sum_{i=1}^{N} \Phi\left(\frac{Z_i - a}{\gamma}\right)\right]$$

Where the Z_i are N realisations of the probability Q. The pulling of the Z_i proceeds in the following way:

- U_i is drawn according to the uniform distribution on[0,1].
- $V_i = T_{n-1}(a + \gamma u_\alpha^+) + \left(1 - T_{n-1}(a + \gamma u_\alpha^+)\right)U_i$; *i.e.,* that V_i follows the uniform distribution on $[T_{n-1}(a + \gamma u_\alpha^+), 1]$.

$Z_i = {T^{-1}}_{n-1}(V_i)$, *i.e.*, that Z_i follows the distribution Q.

4.2. Result's representation and discussion

We will find below the representative curves of π as a function of the observation $x = \frac{1}{n}\sum_{i=1}^{n} x_i$. We have considered only the case where $\theta_0 = 0$ and in the first as in the second sample, the observations are of the same unit variance σ^2 but where the numbers can vary, considering that a modification of θ_0 and σ^2 will only result in a translation effect. We have considered, the two cases: $\alpha = 0.05$ and $\alpha = 0.01$. On the other hand, we have taken $k = 10$, 20 or 30 for $n = 10$ (Figure 3 and Figure 4).

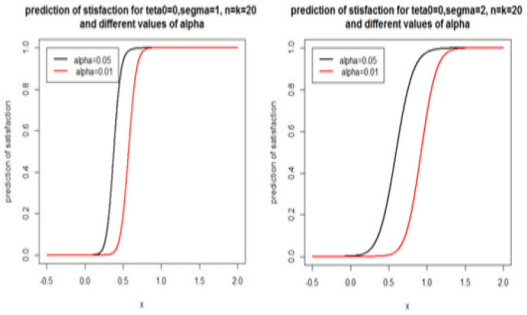

Figure 3: Prediction of satisfaction based on 5000 iterations for $\alpha = 0.05$, $\sigma^2 = 1$ and 4. Graphs with a step 0.001 for x.

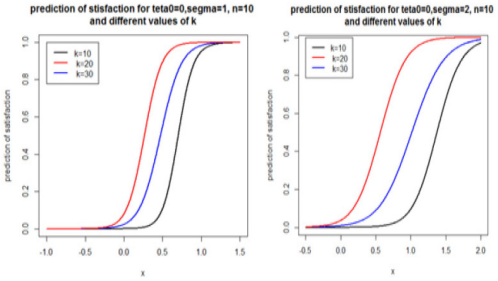

Figure 4: Prediction of satisfaction based on 5000 iterations for $\alpha = 0.01$, $\sigma^2 = 1$ and 4. Graphs with a step 0.001 for x.

Graphs (3(a)- 3(b)) represent the prediction of satisfaction when $\alpha = 0.05$. The first one is for $\sigma = 1$ and the second is for $\sigma = 2$. We see clearly that the satisfaction rises with n and the convergence becomes slower in the second one and it is clear that when k augments the satisfaction increases fastly.

Furthermore, graphs (4(a)- 4(b)) represent the prediction of satisfaction when $\alpha = 0.01$ and $\alpha = 0.05$ for $n = k = 10$ and $\sigma = 1$ or $\sigma = 2$ respectively. We can make the same remark but the values augment moving away from 0 fastly in the second graph than in the first one because we need more arguments to reject the null hypothesis. This conveys well the interest of the consideration of the p-value in the index of satisfaction that the reject region is more informative since x is larger, which gives importance to the indicated index.

5. CONCLUSION

Bayesian predictive procedures have an important contribution to inference and data analysis. Within this perspective, Bayesian predictive probabilities can be used for interim monitoring of experimental trials to estimate the probability of observing a statistically significant result if the trials are to continue to its predefined maximum sample size.

The main objective of this paper is to present an answer to the question:" How to evaluate, if a given experiment will be conclusive about a hypothesis before it is performed?". The answer is given by the proposal hybrid Bayesian-frequentist procedure to evaluate whether a *p*-

value-based hypothesis test will yield a conclusive result in the context of clinical trials. The proposed design is based on a family of limited indexes of satisfaction which was the generalization of the "rudimentary" index of satisfaction considered in [13].

The methodology is useful in two-steps testing procedures; the result of the first step is used to decide if the experiment will be continued. Given the posterior distribution derived from the available data, the prediction of satisfaction is defined as the predictive expectation of the index of satisfaction for the future sample. We consider different cases of the application of the proposed procedure with a non-informative prior.

Furthermore, we can use this procedure to develop an adaptive design for experimental trials especially in the sequential analyses to monitor trials very well by choosing any cohort size for the steps. Examples in section 4, for both binary outcomes and it's approximation to the Gaussian model, gives the characteristics of this procedure.

For example, in the context of interim monitoring for futility for single–arm clinical trials, prediction of satisfaction is naturally appealing because it directly addresses the relevant question, that is, whether a trial is likely to reach its objective if continued to the planned maximum sample size. In this situation, the index of satisfaction can be used formally as a stopping rule. At interim analysis, termination occurs to reject H_0 if the prediction of satisfaction *PIS*, at a point θ_0, is high, formally, if it is greater than a specified constant between 0.5 and 1. The specific stopping criteria are typically unique to each trial and include ethical and business considerations, such as risk/ benefit considerations, available resources, opportunity cost, and overall statistical power.

Bayesian experimental trial simulation is a generic tool that can compute the predictive satisfaction for any trial result, whether that is based on a Bayesian analysis of the data, frequentist significance tests or a formal decision analysis such as a decision by a health care provider to put a drug in the market. In our paper, we have taken an inferential problem related to the binary outcomes and the Gaussian model usig this methodology and this stopping rule for the trial. The simulations results have perfectly illustrated the procedure which ensures the neutrality, the objectivity and esspecially the ethical considerations. The numeric calculus were similars to those obtained by [8] and [20].

AKNOWLEDGEMENTS

We absolutely would like to thank Mr. Zougab Nabil Professor at the University of Tizi-Ouzou and LAMOS of Bejaia, for his help and advice on the R simulation in this work. we thank too the anonymous referees for several suggestions that have improved the presentation of this paper.

REFRENCES:

[1] Berger, J.O. and Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence.. Journal of the American Statistical Association, Vol. 82, No. 397, 112-122.

[2] Chow, S. C. (2005). Statistical consideration of adaptive methods in clinical development. Journal Biopharm. Stat., 15, 575-91.

[3] Chow, S. C., Berry, S.M., Carlin, B.P., Lee, J.J., Müller, P. (2011). Bayesian adaptive methods for clinical trials. Chapman &Hall/CRC Press, New York.

[4] Chow, S. C. Corey, R., Lin, M., (2012). On the independence of data monitoring committee in adaptive design clinical trials. Journal of biopharmaceutical statistics, 22(4): 853-867.

[5] Chow, S. C. and Chang M., (2007). Adaptive design in clinical trials. Chapman & Hall/CRC. Biostatistics Series, USA.

[6] Djeridi, Z. Merabet, H., (2016). Choice of the prior in the Bayesian design for the clinical trials . Sciences et Technologie: A , N44, 27-31.

[7] Goodman S., (2008). A Dirty Dozen: Twelve P-Value Misconceptions. In Siminar in Hematologie, 45: 135-140.

[8] Guosheng, Y. Chen, C., Lee, J.J. (2012). Phase II trial design with Bayesian adaptive randomization and predictive probability. statistical society: Series C (Applied statistics), 61, 219-235.

[9] Hamada, M. S., Wilson, A.G., Reese, C. S., Martz, H.F., (2008). Bayesian reliability. Springer Series in statistics.

[10] Hoff, P.D., (2009). A first course in Bayesian statistical Methods. Springer science + Business media, New York.

[11] Hwang, J.T., Casela, G., Robert, C., Wells, M.T., Farrell, R.H., (1992). Estimation of accuracy in testing. The annals of statistics, Vol. 20, No. 1; 490-509.

[12] Lecoutre, B, (2001). A Bayesian predictive procedures for designing and monitoring experiments. In Bayesian methods with applications to science, Policy and official statistics Luxemburg. office for official publications of the european communities, 301-310.

[13] Lecoutre, B., Derzko, G. and Grouin, J. M., (1995). Bayesian predictive approach for inference about proportions. Statistics in Medecine, 14: 1057- 1063.

[14] Lee, J.J. and Liu, D.D., (2008). A predictive probability design for phase II cancer clinical trials. Clinical Trials, 5: 93-106.

[15] Merabet, H., (2004). Index and prevision of satisfaction in exponential models for clinical trials. Statistica, anno LXIV, n3, 441-453.

[16] Merabet, H., Labdaoui, A., (2015). Bayesian prediction using two stages design in expiremental trials. Nature & Technologie A-Sciences fondamentals et engineering, n 12: 11-23.

[17] Robert, C.P., (2007). Bayesian choice.From Decision-Theoretic Foundations to Computational Implementation. Springer science + Business media, New York.

[18] Saville, B. R. , Connor, J. T. , Ayers, G. D. and Alvarez, J., (2014). The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. Clin Trials,DOI: 10.1177/1740774514531352.

[19] Shao,Y., Mukhi, V., and Goldberg, J. D., (2008). A hybrid Bayesian-frequentist approach to evaluate clinical trial designs for tests of superiority and non-inferiority. Statist. Med., 27:504-519.

[20] Simon, R., (1989). Optimal two-stage design for phase II clinical trials. Controlled clinical trials, 10: 1-10.