Sequential Analysis, 29: 1–14, 2010

ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474940903482452

An Objective Bayesian Approach to Multistage Hypothesis Testing

Pierre Bunouf and Bruno Lecoutre
Laboratoire de Mathématique Raphaël Salem,
ERIS, CNRS–Université de Rouen, France

Abstract: A new Bayesian approach to multistage hypothesis testing is considered. Prior is derived using Jeffreys' criterion on likelihood associated with the design information. We show that the prior for sequential Bernoulli design asymptotically converges toward the Jeffreys prior in Pascal sampling model. A general rule is given for determining the design-corrected version of default priors when Jeffreys' criterion results in improper distribution. Based on the principle of design impartiality, the Bayes factor as posterior-based evidential measure can be generalized to multistage testing, so that the decision boundaries reflect equal evidence for hypotheses over stages. Effect of prior correction on design parameters and on Bayesian inference upon test termination is studied. The approach is applied to a three-stage binomial design. Last, the use of the prior as the default objective choice in multistage hypothesis testing is discussed.

Keywords: Bayes factor; Frequentist characteristics; Jeffreys' criterion; Likelihood principle; Objective prior.

Subject Classifications: 62F03; 62F15; 62L05.

1. INTRODUCTION

The supposed link between Bayes' rule and the likelihood principle has long obscured the issue of the stopping rule influence in Bayesian testing. However, the argument that the design information has no inferential value (see, e.g., Berger and Wolpert, 1988, p. 88) is not tenable for many experimenters. The so-called *unified conditional frequentist and Bayesian testing* or unified testing (see Berger et al., 1994) based on the Bayes factor offers an evocative example. The authors showed that the

Received June 18, 2009, Revised October 13, 2008 and October 28, 2009, Accepted November 1, 2009
Recommended by Adam Martinsek
Address correspondence to Pierre Bunouf, Laboratoire de Mathématique Raphaël Salem, ERIS, CNRS–Université de Rouen, Avenue de l'Université, BP 12, Saint-Etienne-du-Rouvray 76801, France; E-mail: pierre.bunouf@gmail.com

Bayesian error probabilities of hypotheses are also valid frequentist risks conditional on a partition of the outcome space. In the extension to multistage designs, Berger et al. (1999) observed that the unified testing ignores the design information, "seeming to lend frequentist support to the stopping rule principle." Nevertheless, it is well known that multiple looks at data affect the (unconditional) frequentist risks in long-run sampling context. In the Bayesian setting, Rosenbaum and Rubin (1984) studied the influence of data-dependent stopping rule on coverage probability of confidence (or credible) interval, and Spiegelhalter et al. (2004, Section 6.6.5) exhibited the impact on type 1 error in hypothesis testing.

However, based on a new formulation of Bayes' rule, de Cristofaro (2004) showed that explicit reference to the design is fully Bayesian justified and Bayesian objectivity cannot ignore such information. In this article, the unified testing is generalized to multistage designs using a design-corrected version of the Bayes factor. The approach is based on prior derived using Jeffreys' criterion on likelihood associated with the design. The characteristics of the so-called *corrected Jeffreys prior* (literally model-based Jeffreys prior corrected by the design information) and the corresponding Bayes factor are studied in one-parameter problems. Among possible candidate objective priors for multistage Bayesian analysis, the corrected Jeffreys prior satisfies the principle of design impartiality, which is based on the property of data-translated likelihood. Moreover, we show that Bayesian inference upon test termination is corrected for the stopping rule influence.

The derivation of the corrected Jeffreys prior and characteristics concerning existence and domination are presented in the next section. We show that the corrected Jeffreys prior for sequential Bernoulli design asymptotically converges toward the Jeffreys prior in Pascal sampling model. A general rule is given for determining the design-corrected version of default priors when Jeffreys' criterion results in improper distribution. The corrected Bayes factor and the multistage test are introduced in Section 3. Prior correction effect on design parameters is studied in composite hypothesis testing for continuous observations. We also highlight a risk of degeneracy phenomenon of the prior density in open design associated with infinite stopping rule. Section 4 shows an application to a three-stage binomial design. The application involves a study of the prior correction effect on the Jeffreys confidence interval obtained upon test termination. In the conclusion, we return to the role of the likelihood principle in experimental research. Then, we discuss the use of the corrected Jeffreys prior as the default objective choice in multistage hypothesis testing. Last, the extension to multiparameter problems is considered.

2. CORRECTED JEFFREYS PRIOR

We consider the K-stage design $d_{\otimes K}$ involving successive trials of n_k i.i.d. observations ($1 \leq k \leq K$) for inference on the one-dimensional parameter $\theta \in \Theta$. Let X_k be the outcome variable at stage k, we suppose that $X^{(k)} = (X_1, X_2, \ldots, X_k)$ is an independent sequence in the design $d_{\otimes K}$, with a known density function $p_k(x^{(k)} \mid \theta, d_{\otimes K})$ that satisfies minimum conditions of regularity. The sequence $X^{(k)}$ takes values in the outcome space $\mathscr{X}^{(k)}$ equipped with a σ-algebra $\mathscr{B}^{(k)}$. Let τ be a stopping rule consisting of probabilities $\tau_k(X^{(k)})$ of stopping after $x^{(k)}$ is observed. We denote the stopping stage variable by M (i.e., $\tau_k = P_\theta(M = k)$).

Most of Bayesians dealing with multistage designs are still reluctant to transgress the stopping rule principle (i.e., inference does not depend on the stopping rule that governs the experiment), in spite of explicit attempts to incorporate the design information into priors (see, e.g., Bernardo and Smith, 1994). However, the conditioning on the design is fully justified in the Bayesian approach. De Cristofaro's formulation of Bayes' rule makes explicit reference to the design $d_{\otimes K}$ as a part of the preexperimental evidence. Let e_0 contain the beliefs on the θ values before the experiment and let Π_k be a sequence of priors about θ, Bayes' rule becomes

$$\Pi_k\big(\theta \mid x^{(k)}, e_0, d_{\otimes K}\big) \propto \Pi_k(\theta \mid e_0, d_{\otimes K}) p_k\big(x^{(k)} \mid \theta, e_0, d_{\otimes K}\big). \tag{2.1}$$

Then, both the likelihood principle and its major consequence the stopping rule principle are no longer an automatic consequence of Bayes' rule. Moreover, (2.1) shows that prior ignorance cannot be characterized without reference to the design.

Bayesian prior distribution allows recovering a part of the information implicitly contained in the design and lost in the likelihood. The solution proposed in this article is based on Jeffreys' criterion, which agrees with the principle of design impartiality (de Cristofaro, 2004): a design is impartial with respect to θ if the property of data-translated likelihood (i.e., the information on θ is contained in the likelihood location only) introduced in Box and Tiao (1992) is satisfied or approximately satisfied. The use of Jeffreys' criterion on likelihood associated with the design yields a prior proportional to the naive (i.e., design-unrelated) Jeffreys prior times $E_\theta^{1/2}(M)$ (see Govindarajulu, 1981).

Govindarajulu derived the prior from the design-associated likelihood

$$L^A\big(\theta; x^{(m)}, d_{\otimes K}\big) = [L(\theta; x_1)]^{1_{m=1}} \times \cdots \times \big[L\big(\theta; x^{(K)}\big)\big]^{1_{m=K}}.$$

Let $I(\theta \mid x^{(m)})$ be the Fisher information about θ contained in $x^{(m)}$ based on the naive likelihood $L(\theta; x^{(m)})$. The Fisher information derived from the design-associated likelihood is

$$\begin{aligned} I\big(\theta \mid x^{(m)}, d_{\otimes K}\big) &= -E_\theta\left[\frac{\partial^2}{\theta^2}\log L^A\big(\theta; x^{(m)}, d_{\otimes K}\big)\right] \\ &= I(\theta \mid x_1)P_\theta(M=1) + \cdots + I\big(\theta \mid x^{(K)}\big)P_\theta(M=K) \\ &= I(\theta \mid x_1)\left[1 + \frac{n_2}{n_1}P_\theta(M \geq 2) + \cdots + \frac{n_K}{n_1}P_\theta(M=K)\right] \\ &= I(\theta \mid x_1)E_\theta(M). \end{aligned} \tag{2.2}$$

The density of the corrected Jeffreys prior is proportional to $I(\theta \mid x^{(m)}, d_{\otimes K})^{1/2}$. The corrected Jeffreys prior reflects the degree of certainty associated with the projected design $d_{\otimes K}$ by overweighing the probabilities about θ values more likely leading to late termination. Greater is the certainty about a value of θ, higher is its initial probability. Consequently, posterior-based inference on θ is corrected for the stopping rule influence.

2.1. Existence and Domination

The existence of the corrected Jeffreys prior $\Pi^{CJ}(\theta \mid d_{\otimes K})$ requires the expectation of M to be bounded. Then, if the density of the naive Jeffreys prior $\Pi^J(\theta)$ is integrable over Θ, the corrected version is proper (i.e., $\int_\Theta d(\Pi^{CJ}(\theta \mid d_{\otimes K})) < \infty$). However, improper Jeffreys priors can be asymptotically approached by proper corrected Jeffreys priors using truncation method.

We illustrate such truncation method in the Pascal (or inverse binomial) sampling model associated with the design d_{Pas}. The stopping rule in the design d_{Pas} is infinite (i.e., $P_\theta(M < \infty) \neq 1$ a.s. when $\theta \to 0$) and Jeffreys' criterion results in the improper $Be\left(0, \frac{1}{2}\right)$ prior distribution.

Theorem 2.1. *Let us consider the K-stage Bernoulli design $d_{Ber^{\otimes K}}$ for an experiment based on successive Bernoulli trials $Y_k = 0, 1$ $(k = 1, \ldots, K)$ with early stopping if the outcome is observed (i.e., $Y_k = 1$). The Pascal sampling model describes the distribution of the outcome occurrence in $d_{Ber^{\otimes K}}$ when $K \to \infty$.*

Proof. The corrected Jeffreys prior for the design $d_{Ber^{\otimes K}}$ is

$$\begin{aligned}\Pi^{CJ}(\theta \mid d_{Ber^{\otimes K}}) &\propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}(1 + (1-\theta) + \cdots + (1-\theta)^{K-1})^{\frac{1}{2}} \\ &= \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}\left(\frac{1-(1-\theta)^K}{\theta}\right)^{\frac{1}{2}}. \end{aligned} \tag{2.3}$$

When $K \to \infty$, the proper density of the corrected Jeffreys prior for $d_{Ber^{\otimes K}}$ tends to the improper density of the Jeffreys prior for d_{Pas}, i.e.,

$$\lim_{K\to\infty} \Pi^{CJ}(\theta \mid d_{Ber^{\otimes K}}) \to \Pi^{J}(\theta \mid d_{Pas}) \sim Be\left(0, \frac{1}{2}\right).$$

Formally, the stopping stage $M' = \inf\{k : Y_k = 1 \text{ or } k = K\}$ in $d_{Ber^{\otimes K}}$ is a truncation of the stopping stage in d_{Pas}. □

Compared to the symmetric density of the naive Jeffreys prior $Be\left(\frac{1}{2}, \frac{1}{2}\right)$, the unnormalized density in (2.3) assigns higher probabilities to the low values of θ as K increases. The corrected Jeffreys prior compensates the positive bias induced by the stopping rule in the design $d_{Ber^{\otimes K}}$ on the maximum likelihood estimator (*MLE*), which is $\hat{\theta}^{ML} = 1/M$.

The bias of the *MLE* in $d_{Ber^{\otimes K}}$ is

$$E_{d_{Ber^{\otimes K}},\theta}\left(\frac{1}{M}\right) - \theta = \sum_{k=1}^{K}(1-\theta)^{k-1}\theta\frac{1}{k} - \theta = \sum_{k=2}^{K}(1-\theta)^{k-1}\theta\frac{1}{k} > 0. \tag{2.4}$$

Maxima of both the bias of the *MLE* (2.4) and the prior correction effect in (2.3) are reached when $K \to \infty$. Then, the stopping stage M follows a geometric distribution and the bias of the *MLE* is deduced from

$$E_{d_{Pas},\theta}\left(\frac{1}{M}\right) = \frac{\theta}{1-\theta}\log\frac{1}{\theta}.$$

In the regular case, naive Jeffreys prior pertains to a class of continuous and positive densities that have polynomial majorants and benefit of good properties for the derivation of proper posteriors (despite there is no general statement). However, the naive Jeffreys prior is often improper when the parameter space is unbounded. In that case, the corrected Jeffreys prior is also improper. The proof is straightforward from (2.2). The corrective term $E_\theta(M)^{1/2}$ is bounded and admits the majorant function $E_\theta(M)$. Then, one easily derives a polynomial approximation of $E_\theta(M)^{1/2}$, which is function of the terms $P_\theta(M \geq k)$ $(k = 2, \ldots, K)$ and also contains a constant term. Consequently, the corrected Jeffreys prior admits a mixture density including the improper component $d(\Pi^J(\theta))$.

Various alternatives have been suggested when Jeffreys' criterion results in the improper uniform distribution such as in the normal case (see, e.g., Jeffreys, 1961). These alternatives are often proper 'diffuse' priors reflecting a status of objectivity. Approximate design-corrected version of such default priors can be obtained using the *correction transposition rule*, which consists in transposing the corrective term from the improper Jeffreys prior to default priors. Unnormalized densities are obtained by multiplying default prior densities by $E_\theta(M)^{1/2}$ borrowed from the corrected Jeffreys prior (see an illustration in the next section). Jeffreys' criterion imposes a condition on the parameter so that the likelihood locally and approximately undergoes a translation for different observations. This condition is maintained using the correction transposition rule if default prior densities are sufficiently spread-out, so that their design-corrected versions satisfy the principle of design impartiality.

The domination of the likelihood by the prior is another important characteristic. In objective Bayesian analysis, the influence of naive priors is usually low and disappears as the observed sample size increases. Conversely, the correction effect of the corrected Jeffreys prior depends on the variation in θ of the likelihood relative to integral forms of $p_k(x^{(k)} \mid \theta, d_{\otimes K})$ $(k = 1, \ldots, K-1)$. The proof is straightforward from (2.2). The corrective term $E_\theta(M)^{1/2}$ depends on $P_\theta(M \geq k) = \int_{J^{\otimes k-1}} p_{k-1}(x^{(k-1)} \mid \theta, d_{\otimes K}) dx^{(k-1)}$ $(k = 2, \ldots, K)$ where $J^{\otimes k-1} = J_1 \times \cdots \times J_{k-1}$ is the $k-1$-dimensional support of the outcome sequences.

3. CORRECTED BAYES FACTOR TEST

The recours to objective priors in hypothesis testing is limited as the division of the parameter space in two disjoint subsets contradicts this concept (Robert, 2001). However, stopping rule favors one of the hypotheses if the parameter subspace contains the θ values that are the most associated with early termination. Consequently, prior objectivity in the sense of ensuring equal support to hypotheses shouldn't ignore the design information.

The stopping rule is often based on the decision rule concerning hypotheses such as in the familiar sequential probability ratio test (SPRT) introduced in Wald (1947). Formally, the decision rule D takes values D_A and D_R such that the events $\{D = D_A \cap M = m\}$ and $\{D = D_R \cap M = m\}$ are determined by $x^{(m)}$ for each m. The density $p_m(x^{(m)} \mid \theta, d_{\otimes K})$ in the design $d_{\otimes K}$ is then the restriction of the unique probability measure defined on the smallest sigma algebra containing all the σ-algebra $\mathcal{B}^{(k)}$ $(k = 1, \ldots, K)$ to the one associated with a termination at stage m.

The objectivist Bayesians prefer using the Bayes factor, noted B_k, which is irrespective of the relative prior weights of hypotheses. The multistage experiment stops when B_k provides enough evidence for decision-making. For the set of composite hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1, \quad (\Theta_0 \cap \Theta_1 = \emptyset),$$

the stopping stage is

$$M = \min\{k \geq 1 : B_k \notin (R, A) \text{ or } k = K\},$$

where the H_0 rejection region is such that $\{D = D_R \cap M = m\} = \{B_m \leq R\}$ and the H_0 acceptance region is such that $\{D = D_A \cap M = m\} = \{B_m \geq A\}$ $(R \leq 1 \leq A)$.

The prior predictive distribution based on the corrected Jeffreys prior under H_i $(i = 0, 1)$, $\Gamma^{CJ}_{H_i,k} : \mathscr{B}^{(k)} \to [0, 1]$, describes an expectation concerning $x^{(k)}$ associated with the data generation process and the design, i.e.,

$$\Gamma^{CJ}_{H_i,k}\big(x^{(k)} \mid d_{\otimes K}\big) = \int_{\Theta_i} p_k(x^{(k)} \mid \theta, d_{\otimes K})\Pi^{CJ}(\theta \mid d_{\otimes K})d\theta. \tag{3.1}$$

Subsequently, we define the corrected Bayes factor B^{CJ}_k as

$$B^{CJ}_k = \Gamma^{CJ}_{H_0,k}\big(x^{(k)} \mid d_{\otimes K}\big)/\Gamma^{CJ}_{H_1,k}\big(x^{(k)} \mid d_{\otimes K}\big).$$

The parameters of the design $d_{\otimes K}$ are determined by the test based on the corrected Bayes factor in Definition 3.1. The reported errors are the posterior probabilities of hypotheses.

Definition 3.1. Corrected Bayes factor test (CBFT)
If $B^{CJ}_k \leq R$, stop, reject H_0 and report the error $\alpha(x^{(k)} \mid d_{\otimes K}) = B^{CJ}_k/(1 + B^{CJ}_k)$,
if $B^{CJ}_k \geq A$, stop, accept H_0 and report the error $\beta(x^{(k)} \mid d_{\otimes K}) = 1/(1 + B^{CJ}_k)$.
Otherwise, if $k \leq K$ continue to stage $k + 1$, or if $k = K$ make no decision.

When the stopping rule is finite, the Bayesian error probabilities of the *CBFT* are also valid risks in the conditional frequentist approach (see, e.g., Berger et al., 1997, Dass and Berger, 2003, for the extension to composite hypothesis testing). The (ancillary) conditioning statistic is a one-one transformation of m that yields a partitioning of the outcome sequences support in two subsets characterizing the same evidence for H_0 and H_1. The principle of combining Bayesian-frequentist approaches in the unified testing was emphasized in Bayarri and Berger (2004).

However, the experimental design influences posterior-based evidential measures such as the Bayes factor because early stopping happens only when outcome sequence is sufficiently indicative of one hypothesis. Despite the stopping rule, strict application of the likelihood principle imposes the use of naive priors. Relaxing this principle, the corrected Jeffreys prior assigns higher density mass to θ values associated with later expected stopping stage relative to the naive Jeffreys prior. Prior predictive distributions carry the prior correction to the Bayes factor. Based on the principle of design impartiality, the corrected Bayes factor is a valid evidential measure, so that the decision boundaries of the *CBFT* reflect equal evidence for hypotheses over stages. The prior correction effect on design parameters radically differs from the unconditional frequentist approach, which aims at preserving nominal risks in long-run sampling context. The *CBFT* generalizes the unified testing to multistage designs using appropriate priors.

As for any test based on the Bayes factor, a major issue with the *CBFT* arises when prior is improper as the prior predictive distributions under hypotheses in (3.1) cannot be derived. As mentioned in Section 2.1, if the naive Jeffreys prior is improper, the corrected Jeffreys prior is also improper. Such situation can be overcome with the use of default prior with flat or diffuse density. Then, its design-corrected version is derived using the correction transposition rule.

This case is illustrated for a two-stage experiment involving two sets of 10 i.i.d. $N(\mu, 1)$ observations to test the point null hypothesis $H_0 : \{\mu = 0\}$ versus the composite alternative $H_1 : \{\mu > 0\}$. The naive Jeffreys prior under alternative is the improper uniform distribution. The default prior is the half normal $HN(0, 2)$ distribution, which is proportional to the normal $N(0, 2)$ for positive values (see arguments in Berger and Sellke, 1987). We set the values $A = R^{-1} = 5$ and assign equal prior probabilities to H_0 and H_1. Let Z_1 be the mean at stage 1, Z_2 the mean accrued until stage 2, and Φ the cumulative distribution function of the standard normal law. The density of the corrected Jeffreys prior is proportional to $E_\theta^{1/2}(M) = (1 + n_2/n_1 \Phi(\sqrt{n_1}(Z_1 - \mu) \in J_1)))^{1/2}$, where J_1 is the interval for $\sqrt{n_1} Z_1$ such that $B_1^{CJ} \in (R, A)$. According to the correction transposition rule, the density of the design-corrected $HN(0, 2)$ prior is proportional to $\exp(-\mu^2/2) E_\theta^{1/2}(M)$. Its derivation requires an iterative procedure as the stopping rule is part of the prior. The curves of prior and prior predictive densities under H_1 are displayed in Figure 1. F1

The prior correction causes an increase of prior predictive density mass for z_k $(k = 1, 2)$ generated by μ values more associated with expected termination at stage 2. Let z_k^A and z_k^R $(k = 1, 2)$ be the boundaries of Z_k for acceptance and rejection of the null hypothesis, respectively. We obtain $(z_1^A, z_1^R) = (-0.30, 0.66)$ and $(z_2^A, z_2^R) = (-0.08, 0.69)$ in the corrected approach instead of $(-0.20, 0.67)$ and $(-0.03, 0.69)$ in the naive approach.

Beyond the decision to 'accept' or 'reject' H_0, experimenter is concerned with the magnitude of the parameter irrespective of whether the test declares statistical significance. In the long-run frequentist context, the departure of coverage function from the nominal level is indicative of the stopping rule influence on confidence (or credible) intervals. Let $[\hat{\theta}^{low}, +\infty)$ and $(-\infty, \hat{\theta}^{upp}]$ be the one-sided confidence intervals and consider the sufficient bivariate statistic (M, Y_m) where Y_m is the outcome accrued until stage m. The coverage functions of both intervals are

$$C^{low}(\theta; d_{\otimes K}) = P_\theta[\theta \geq \hat{\theta}^{low}(M, Y_m)] \quad \text{and} \quad C^{upp}(\theta; d_{\otimes K}) = P_\theta[\theta \leq \hat{\theta}^{upp}(M, Y_m)]. \tag{3.2}$$

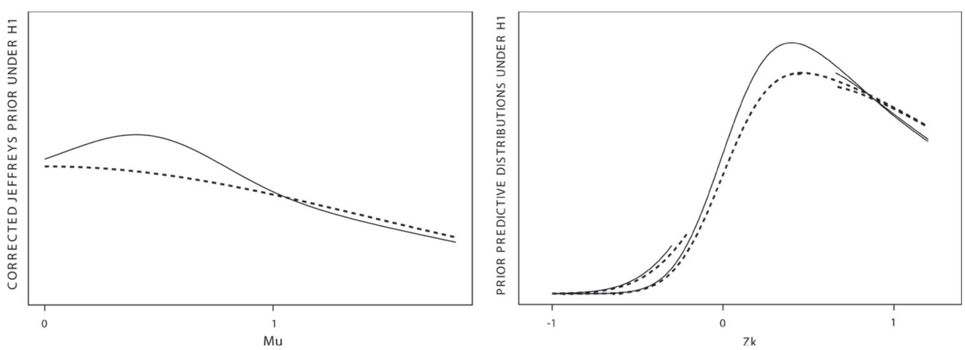

Figure 1. Naive (- - -) and design-corrected (—) $HN(0, 2)$ prior densities under $H_1 : \{\mu > 0\}$ (left) and prior predictive densities in the support of z_k under stopping at stage $k = 1, 2$ (right) for the two-stage design with $A = R^{-1} = 5$ and $n_1 = n_2 = 10$.

In (3.2), the prior correction corrects coverage function of the one-sided Jeffreys confidence intervals for the stopping rule influence whatever $\theta \in \Theta$ if, for any couple of possible pairs (m, y_m) and $(m', y'_{m'})$, the ordering of the confidence limits is the same using the naive and the corrected Jeffreys priors. This condition is satisfied in many multistage designs for lattice data (see application to the binomial case in Section 4).

3.1. Prior Correction Effect on Design Parameters

The case of open designs ($K \to \infty$) raises the question of the finiteness of the stopping rule. This characteristic has been explored for the *SPRT* for a long time. Stein (1946) showed that the stopping stage for testing point hypotheses is exponentially bounded (i.e., satisfies $P_\theta(M > n) < c\rho^n$ for some $c < \infty$ and $0 < \rho < 1$) except if the log probability ratio is degenerate at 0. In composite hypothesis testing, Wald suggested a reduction to point hypothesis by means of weight function. If a group of invariance transformations exists for such reduction, Wijsman (1971) gave sufficient conditions on observation distribution for the stopping rule to be finite. In this section, the effect of prior correction on parameters of K-stage *CBFT*-based designs is studied as K increases. Then, we highlight a risk of degeneracy phenomenon of the corrected Jeffreys prior in open design.

Theorem 3.1. *The increase of K in K-stage CBFT-based symmetric design for composite hypotheses of the type $H_0 : \{\theta \geq \theta_0\}$ versus $H_1 : \{\theta < \theta_0\}$ for continuous outcome yields more conservative decision boundaries (i.e., wider non decision region).*

Proof. Let $d^{Sym}_{\otimes K}$ be a K-stage symmetric *CBFT*-based design for continuous outcomes X_k with fixed values of A and R, such that $A = R^{-1}$. To ease the reading, the corrected Bayes factor in $d^{Sym}_{\otimes K}$ is noted B^K_k in this section, and $\mathcal{S}^K$ denotes the support of $X^{(k)}$ $(k = 1, \ldots, K)$ such that $A < B^K_k < R$ $(k = 1, \ldots, K-1)$. We assume that naive Jeffreys prior under hypothesis is not degenerate. In the parameter space, ω is the common boundary of the closures of Θ_0 and Θ_1. We note by $M(\omega, \epsilon)$ the ϵ-neighborhood of ω defined as the set of all $\theta \in \Theta$ such that $\|\omega - \theta\| < \epsilon$. Based on a fixed positive scalar λ, we also introduce ϵ^K in $M(\omega, \epsilon^K)$, which is the maximum neighborhood width such that $d(\Pi^{CJ}(\theta \mid d^{Sym}_{\otimes K})) \geq \lambda$ whatever $\theta \in M(\omega, \epsilon^K) \cap \Theta_i$ $(i = 0, 1)$. In the $K+1$-stage design $d^{Sym}_{\otimes K+1}$, the related quantities are B^{K+1}_k, $\mathcal{S}^{K+1}$, which is the support of $X^{(k)}$ $(k = 1, \ldots, K+1)$ such that $A < B^{K+1}_k < R$ $(k = 1, \ldots, K)$, and ϵ^{K+1}. We also define $\mathcal{S}^{K+1*}$ as the K-dimensional restriction of $\mathcal{S}^{K+1}$ for $X^{(k)}$ $(k = 1, \ldots, K)$ in the design $d^{Sym}_{\otimes K+1}$.

Relative to the design $d^{Sym}_{\otimes K}$, the additional stage $K+1$ in $d^{Sym}_{\otimes K+1}$ causes an increase of $E_\theta(M)$ around $\theta = \omega$. The density mass of the corrected Jeffreys prior concentrates so that if a sufficiently narrow neighborhood of ω is considered, we have the relation $\epsilon^{K+1} \leq \epsilon^K$ whatever $\lambda > 0$. Consequently, the density mass of both prior predictive distributions under H_0 and H_1 increases for the set of $X^{(k)}$ that provides the poorest evidence for hypotheses. This yields smaller amplitude of the corrected Bayes factor (i.e., $|B^{K+1}_k - 1| < |B^K_k - 1|$, $k = 1, \ldots, K$) and extension of the support of $X^{(k)}$ $(k = 1, \ldots, K)$ (i.e., $\mathcal{S}^K \in \mathcal{S}^{K+1*}$). □

Corollary 3.1. *The corrected Jeffreys prior associated with CBFT-based symmetric design can degenerate in open design.*

Proof. The proof follows the proof of Theorem 3.1. As K increases, the prior correction assigns weight on narrowing neighborhood of ω. If $E_\theta(M)$ does not converge toward a finite function when $K \to \infty$, a degeneracy phenomenon of the prior density occurs at $\theta = \omega$ (i.e., $\epsilon^\infty \to 0$ whatever $\lambda > 0$). From (2.2), the asymptotic behavior of $E_\theta(M)$ results from two opposite contributions when going from the design $d_{\otimes K}^{Sym}$ to $d_{\otimes K+1}^{Sym}$: the term $\{n_{K+1}/n_1 P_\theta(M = K+1)\}$ generates a 'concentration effect' around $\theta = \omega$ whereas the other term $\{1 + n_2/n_1 P_\theta(M \geq 2) + \cdots + n_K/n_1 P_\theta(M \geq K)\}$ generates a 'flattening effect' caused by the extension of $\mathcal{S}^{K+1*}$ relative to $\mathcal{S}^K$. Convergence occurs if the flattening effect annihilates the concentration effect as K increases. Degeneracy phenomenon of the corrected Jeffreys prior in open design $d_{\otimes\infty}^{Sym}$ is associated with infinite stopping rule. Consequently, we have $B_k^\infty \to 1$ $(k = 1, \ldots)$ and infinite extension of the support S^∞. □

4. APPLICATION TO THE BINOMIAL CASE

Let us note $d_{Bin^{\otimes K}}$ the K-stage binomial design involving sequences of independent outcomes $X_k \sim Bin(\theta, n_k = 10)$. The testing hypotheses are $H_0 : \{\theta \leq 0.3\}$ versus $H_1 : \{\theta > 0.3\}$. The design parameters are based on the values $A = 19$ and $R = 1/19$ associated with the nominal level $\alpha^* = \beta^* = 0.05$ for the type 1 and 2 error probabilities.

Let $Y_k = \sum_{i=1}^k X_i$ be the cumulated number of successes until stage k $(k = 1, \ldots, K)$, the boundaries of Y_k for acceptance and rejection of H_0 are noted y_k^A and y_k^R, respectively. The stopping rule is determined by $P_\theta(M \geq k)$ $(k = 2, \ldots, K)$, which is the sum of probabilities

$$p\big(x^{(i)} \mid \theta\big) = \binom{n_1}{x_1} \cdots \binom{n_i}{x_i} \theta^{y_i}(1-\theta)^{n_1+\cdots+n_i-y_i}$$

for $x^{(i)}$ in the $k-1$-dimensional support restriction

$$\mathcal{S}_k^{Bin^{\otimes K}} = \big\{x^{(i)} : y_i^A < y_i < y_i^R;\ i = 1, \ldots, k-1\big\}.$$

Table 1 shows the design boundaries of the naive test and the *CBFT* for the 3-stage design $d_{Bin^{\otimes 3}}$ and 5-stage design $d_{Bin^{\otimes 5}}$. T1

Table 1. Decision boundaries of the naive test and the *CBFT* for the designs $d_{Bin^{\otimes 3}}$ and $d_{Bin^{\otimes 5}}$

	(y_1^A, y_1^R)	(y_2^A, y_2^R)	(y_3^A, y_3^R)	(y_4^A, y_4^R)	(y_5^A, y_5^R)
Naive test	(1, 6)	(3, 10)	(5, 14)	(8, 18)	(10, 22)
CBFT in $d_{Bin^{\otimes 3}}$	(0, 6)	(3, 11)	(5, 14)		
CBFT in $d_{Bin^{\otimes 5}}$	(0, 7)	(2, 11)	(5, 14)	(7, 18)	(10, 22)

Although the scope of Theorem 3.1 limits to continuous outcome, the increase of K value in the K-stage *CBFT*-based $d_{Bin^{\otimes K}}$ design has the same influence on the decision boundaries with an increase of the non decision region. The determination of the set of the *CBFT* boundaries for the design $d_{Bin^{\otimes 5}}$ reveals a practical issue in the iterative process. The implementation in the prior of the design information with $(y_1^A, y_1^R) = (0, 6)$ for the first stage results in the boundaries $(y_1^A, y_1^R) = (0, 7)$, even though the implementation of $(y_1^A, y_1^R) = (0, 7)$ in the prior results in the boundaries $(y_1^A, y_1^R) = (0, 6)$. The boundaries $(y_1^A, y_1^R) = (0, 7)$ are kept as the first situation appears to be the less contradictory.

Table 2. Bayes factor, test decision, and limits of the one-sided 95% Jeffreys confidence intervals using approaches based on the naive and the corrected Jeffreys priors for all pairs (m, y_m) in $d_{Bin^{\otimes 3}}$ design

	Naive approach			Corrected approach		
(m, y_m)	B_m	H_0	95% CI	B_m	H_0	95% CI
(1, 0)	248	Acc	(0.0002, 0.171)	159	Acc	(0.0002, 0.193)
(1, 1)	21.5	Acc	(0.018, 0.331)	–	–	–
(2, 1)	–	–	–	414	Acc	(0.010, 0.191)
(2, 2)	95.6	Acc	(0.029, 0.250)	71.4	Acc	(0.032, 0.260)
(2, 3)	24.8	Acc	(0.056, 0.314)	19.9	Acc	(0.061, 0.319)
(3, 4)	96.9	Acc	(0.057, 0.259)	75.8	Acc	(0.061, 0.265)
(3, 5)	32.9	Acc	(0.079, 0.299)	26.8	Acc	(0.083, 0.304)
(3, 6)	13.4	ND	(0.103, 0.338)	11.3	ND	(0.107, 0.340)
(3, 7)	6.21	ND	(0.127, 0.375)	5.37	ND	(0.132, 0.376)
(3, 8)	3.11	ND	(0.153, 0.412)	2.75	ND	(0.157, 0.410)
(3, 9)	1.64	ND	(0.180, 0.448)	1.47	ND	(0.182, 0.444)
(3, 10)	0.870	ND	(0.207, 0.482)	0.794	ND	(0.208, 0.477)
(3, 11)	0.455	ND	(0.235, 0.516)	0.422	ND	(0.236, 0.509)
(3, 12)	0.228	ND	(0.264, 0.550)	0.216	ND	(0.263, 0.541)
(3, 13)	0.107	ND	(0.323, 0.582)	0.104	ND	(0.291, 0.573)
(3, 14)	0.046	Rej	(0.324, 0.614)	0.047	Rej	(0.320, 0.604)
(3, 15)	0.018	Rej	(0.354, 0.645)	0.019	Rej	(0.349, 0.635)
(3, 16)	0.006	Rej	(0.386, 0.676)	0.007	Rej	(0.379, 0.667)
(3, 17)	0.002	Rej	(0.418, 0.706)	0.002	Rej	(0.409, 0.698)
(3, 18)	5.5×10^{-4}	Rej	(0.450, 0.736)	6.5×10^{-4}	Rej	(0.440, 0.728)
(3, 19)	1.3×10^{-4}	Rej	(0.484, 0.765)	1.7×10^{-4}	Rej	(0.473, 0.759)
(3, 20)	–	–	–	3.7×10^{-5}	Rej	(0.506, 0.788)
(2, 10)	0.052	Rej	(0.324, 0.676)	–	–	–
(2, 11)	0.017	Rej	(0.370, 0.720)	0.018	Rej	(0.360, 0.708)
(2, 12)	0.004	Rej	(0.417, 0.762)	0.005	Rej	(0.405, 0.753)
(2, 13)	0.001	Rej	(0.467, 0.803)	0.001	Rej	(0.452, 0.796)
(2, 14)	1.9×10^{-4}	Rej	(0.518, 0.842)	2.4×10^{-4}	Rej	(0.502, 0.837)
(2, 15)	2.7×10^{-5}	Rej	(0.571, 0.878)	3.7×10^{-5}	Rej	(0.557, 0.876)
(1, 6)	0.041	Rej	(0.347, 0.815)	0.047	Rej	(0.331, 0.802)
(1, 7)	0.007	Rej	(0.442, 0.883)	0.009	Rej	(0.418, 0.876)
(1, 8)	8.8×10^{-4}	Rej	(0.547, 0.940)	0.001	Rej	(0.522, 0.938)
(1, 9)	5.7×10^{-5}	Rej	(0.669, 0.982)	8.2×10^{-5}	Rej	(0.653, 0.982)
(1, 10)	1.1×10^{-6}	Rej	(0.829, 0.999)	1.7×10^{-6}	Rej	(0.826, 0.999)

Acc = accept; ND = no decision; and Rej = reject.

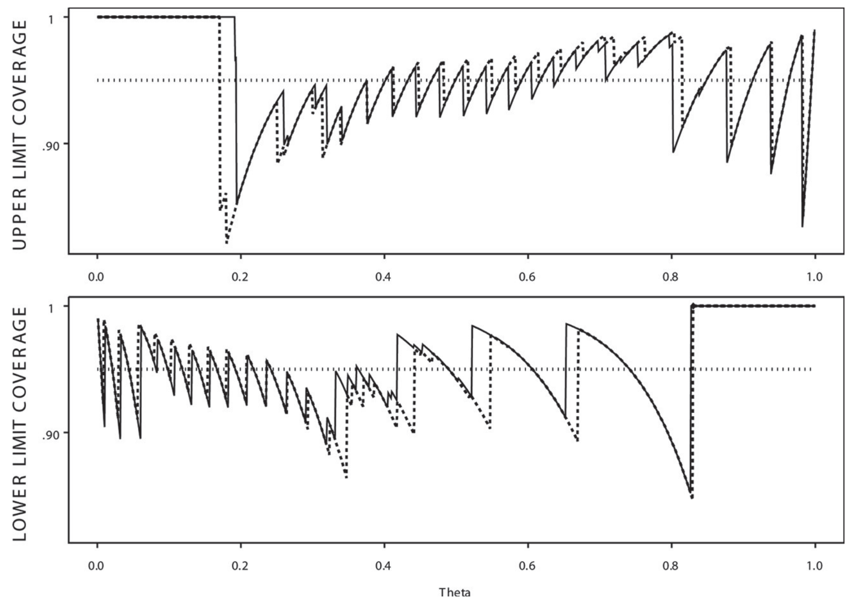

Figure 2. Coverage functions of the upper limit (top) and the lower limit (bottom) one-sided 95% Jeffreys confidence intervals obtained using the naive (- - -) and the corrected (—) Jeffreys priors in the *CBTF*-based $d_{Bin^{\otimes 3}}$ design.

Table 2 presents global results using approaches based on the naive and the corrected Jeffreys priors for all pairs (m, y_m) in the 3-stage design $d_{Bin^{\otimes 3}}$. Bayes factor, test decision, and limits of the one-sided 95% Jeffreys confidence intervals are given. T2

Coverage function of one-sided Jeffreys confidence interval for binomial fixed sample was approached in Cai (2005). Figure 2 displays the coverage curves of the one-sided 90% Jeffreys confidence intervals obtained using the naive and the corrected Jeffreys priors in the *CBFT*-based $d_{Bin^{\otimes 3}}$ design. The curves present discontinuities at the confidence limits of all pairs (m, y_m). The stopping rule influence on the coverage function of the upper limit confidence interval results in under- and overestimation of the nominal level for increasing values of θ, and the inverse for the lower limit confidence interval. This influence is more marked in the neighborhood of the confidence limits of the stopping boundary pairs (k, y_k^A) or (k, y_k^R) $(k = 1, 2)$. From Table 2, the ordering of the confidence limits of all pairs (m, y_m) is the same using the naive and the corrected Jeffreys priors (Note: confidence limits of the pairs (2, 1) and (3, 20) obtained using the naive Jeffreys prior in the *CBFT*-based $d_{Bin^{\otimes 3}}$ design are (0.0089, 0.180) and (0.518, 0.793), respectively). Based on arguments developed in Section 3, the prior correction effect corrects coverage functions for the stopping rule influence whatever θ. F2

The use of the corrected Jeffreys prior for point estimation is coherent if this prior is already used for hypothesis testing and interval estimation. However, the posterior mean estimator based on the corrected version of Haldane's prior offers an interesting alternative in terms of frequentist characteristics (see Bunouf and Lecoutre, 2008). This prior is also derived using the Fisher information of design-associated likelihood but the density is proportional to $I(\theta \mid x^{(m)}, d_{\otimes K})$ instead of $I(\theta \mid x^{(m)}, d_{\otimes K})^{1/2}$ for the corrected Jeffreys prior.

5. CONCLUDING REMARKS

Bayesian approach has never provided a satisfactory answer to the issue of the stopping rule influence in multistage design. The supposed link between Bayes' rule and the likelihood principle (or its major consequence the stopping rule principle) has long been a pill hard to swallow for experimenters willing to adopt Bayesian methods. One may interpret that Bayesian designs are open to unscrupulous manipulation as the experimenter is allowed to choose the stopping stage without formal rule. As underlined by Spiegelhalter (2006), the controversy is illustrated in a recent Food and Drug Administration (FDA) draft guidance (FDA, 2006), which advocates that "the design of a Bayesian clinical trial involves pre-specification of (and agreement on) both the prior information and the model. (…) A change (…) at a later stage of the trial may imperil the scientific validity of the trial results."

Based on de Cristofaro's formulation of Bayes' rule, objective Bayesian analysis cannot depart from design considerations (de Cristofaro, 2004). Moreover, any candidate prior should satisfy the principle of design impartiality and yield posterior credible sets that have good frequentist coverage properties (de Cristofaro, 2008). As mentioned in Kass and Wasserman (1996), assignments of prior probabilities by formal rules cannot be expected to represent exactly total ignorance. However, in this article we show that the corrected Jeffreys prior has the required properties to be one of the default priors reflecting objectivity upon which everyone could fall back when the design information is available prior to the experiment. A large diffusion of this prior in multistage hypothesis testing will require further results concerning the prior characteristics in open design and the prior correction effect on design parameters for several common data distributions.

The extension of the corrected Jeffreys prior to multiparameter problems requires further considerations. Jeffreys' criterion for a p-dimensional vector Θ yields a prior density proportional to $E_\Theta^{p/2}(M)\Pi^J(\Theta)$ where $\Pi^J(\Theta)$ is the naive Jeffreys prior of X_1 (Govindarajulu, 1981). Box and Tiao (1992, p. 53) showed that the property of data-translated likelihood remains approximately valid, so that the principle of design impartiality can be extended to multiparameter problems. However, the issue of separation between parameters of interest and nuisance parameters has raised controversies initiated by Jeffreys himself, which he answered by suggesting a collection of ad hoc rules (Jeffreys, 1961). The importance of this issue is amplified in the corrected Jeffreys prior due to the dependency of the corrective term on the dimension of the whole parameter space. Several authors have developed alternative priors, such as the reference prior based on the maximum-entropy property (see, e.g., Bernardo and Smith, 1994). Design-corrected version can be derived from the design-associated likelihood. Suppose that $\Theta = (\Theta_{(1)}, \ldots, \Theta_{(q)})$ is a q-ordered group where the dimension of component $\Theta_{(i)}$ is p_i for $1 \leq i \leq q$ and assume that the stopping rule depends only on $\Theta_{(1)}$. The rule based on the maximum-entropy property yields a prior density proportional to $E_\Theta^{p_1/2}(M)\Pi^R(\Theta)$, where $\Pi^R(\Theta)$ is the naive reference prior of X_1 (Ye, 1993). Reference priors for some common multiparameter multistage problems are given in Sun and Berger (2008). The dependency of the prior correction on the dimension of $\Theta_{(1)}$ provides a sound argument for using this prior in hypothesis testing, given that it coincides with the corrected Jeffreys prior in one-parameter problems. However, such a perspective requires an extension of the principle of design impartiality and further research to assess the prior correction effect on testing design parameters.

ACKNOWLEDGMENTS

The authors are indebted to the anonymous referee, the Associate Editor, and the Editor for their helpful comments.

REFERENCES

Bayarri, M. J. and Berger, J. O. (2004). The Interplay of Bayesian and Frequentist Analysis, *Statistical Science* 19: 58–80.

Berger, J. O. and Sellke, T. (1987). Testing a Point Null Hypothesis: Irreconcilability of *P* Values and Evidence. With Comments and a Rejoinder by the Authors, *Journal of the American Statistical Association* 82: 112–139.

Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle*, second edition, Hayward, CA: Institute of Mathematical Statistics Monograph Series.

Berger, J. O., Brown, L. D., and Wolpert, R. L. (1994). A Unified Conditional Frequentist and Bayesian Test for Fixed and Sequential Hypothesis Testing, *Annals of Statistics* 22: 317–352.

Berger, J. O., Boukai, B., and Wang, Y. (1997). Unified Frequentist and Bayesian Testing of a Precise Hypothesis. With Comments and a Rejoinder by the Authors, *Statistical Science* 12: 133–160.

Berger, J. O., Boukai, B., and Wang, Y. (1999). Simultaneous Bayesian-Frequentist Sequential Testing of Nested Hypotheses, *Biometrika* 86: 79–92.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*, Chichester: Wiley.

Box, G. E., and Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*, New York: Wiley.

Bunouf, P. and Lecoutre, B. (2008). On Bayesian Estimators in Multistage Binomial Designs, *Journal of Statistical Planning and Inference* 138: 3915–3926.

Cai, T. T. (2005). One-Sided Confidence Intervals in Discrete Distributions, *Journal of Statistical Planning and Inference* 131: 63–88.

de Cristofaro, R. (2004). On the Foundations of Likelihood Principle, *Journal of Statistical Planning and Inference* 126: 401–411.

de Cristofaro, R. (2008). A New Formulation of the Principle of Indifference, *Synthèse, Special Issue: A Selection of Papers Presented to the First Symposium on Philosophy, History and Methodology of ERROR, Virginia Tech.*, 329–339.

Dass, S. (2001). Unified Bayesian and Conditional Frequentist Testing for Discrete Distributions, *Sankhya, Series B* 63: 251–269. AQ1

Dass, S. and Berger, J. O. (2003). Unified Conditional Frequentist and Bayesian Testing of Composite Hypotheses, *Scandinavian Journal of Statistics* 30: 193–210.

FDA (2006). *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*, Rockville, MD: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health.

Govindarajulu, Z. (1981). *The Statistical Analysis of Hypothesis Testing, Point and Interval Estimation, and Decision Theory*, Columbus, OH: American Sciences Press.

Jeffreys, H. (1961). *Theory of Probability*, Oxford: Oxford University Press.

Kass, R. E. and Wasserman, L. (1996). The Selection of Prior Distributions by Formal Rules, *Journal of the American Statistical Association* 91: 1343–1370.

Robert, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Motivations to Computational Implementation*, second edition, New York: Springer-Verlag.

Rosenbaum, P. R. and Rubin, D. B. (1984). Sensitivity of Bayes Inference with Data-Dependent Stopping Rules, *The American Statistician* 38: 106–109.

Spiegelhalter, D. (2006). Comments on Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials, Royal Statistical Society, August 20, 2006.

Spiegelhalter, D. J., Abrams K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Chichester: John Wiley & Sons.

Stein, C. (1946). A Note on Cumulative Sums, *Annals of Mathematical Statistics* 17: 498–499.

Sun, D. and Berger, J. (2008). Objective Bayesian Analysis Under Sequential Experimentation, *IMS Collections, Pushing The Limits of Contemporary Statistics: Contributions in Honour of Jayanta K. Ghosh* 3: 19–32.

Wald, A. (1947). *Sequential Analysis*, New York: John Wiley & Sons.

Wijsman, R. A. (1971). Exponentially Bounded Stopping Time of Sequential Probability Ratio Tests for Composite Hypotheses, *Annals of Mathematical Statistics* 42: 1859–1869.

Ye, K. (1993). Reference Priors when the Stopping Rule Depends on the Parameter of Interest, *Journal of the American Statistical Association* 88: 360–363.