Bayesian Analysis (2013) **8**, Number 1, pp. 221–232

On the Limiting Behavior of the "Probability of Claiming Superiority" in a Bayesian Context

Morris L. Eaton*, Robb J. Muirhead† and Adina I. Soaita‡

Abstract. In the context of Bayesian sample size determination in clinical trials, a quantity of interest is the marginal probability that the posterior probability of an alternative hypothesis of interest exceeds a specified threshold. This marginal probability is the same as "average power"; that is, the average of the power function with respect to the prior distribution when using a test based on a Bayesian rejection region. We give conditions under which this marginal probability (or average power) converges to the prior probability of the alternative hypothesis as the sample size increases. This same large sample behavior also holds for the average power of a (frequentist) consistent test. We also examine the limiting behavior of "conditional average power"; that is, power averaged with respect to the prior distribution conditional on the alternative hypothesis being true.

Keywords: Bayesian design, probability of a successful trial, average power, Bayesian hypothesis testing, clinical trials, sample size determination

1 Introduction

In a recent paper, Muirhead and Soaita (2012) discuss the problem of sample size determination (SSD) in a clinical trial setting from a Bayesian viewpoint. They focus on a criterion for SSD they call the "probability of a successful trial" (PST). (Many statisticians involved in clinical trials use the informal term "successful trial" to mean that the trial data supports the hypothesis that the experimental drug is superior to a control (placebo or an existing drug).) To describe this notion, consider data $X_{(n)}$ from a parametric model $P_n(\cdot|\theta)$, with θ an element of a parameter space Θ. In all practical situations, both $X_{(n)}$ and θ will be elements of subspaces of finite dimensional Euclidean spaces.

Now, consider a proper prior distribution π on Θ and partition Θ into two disjoint subsets Θ_0 and Θ_1. The inferential interest is in concluding that $\theta \in \Theta_1$. Let $Q_n(\cdot|x_{(n)})$ denote the posterior distribution on Θ given the sample $X_{(n)} = x_{(n)}$; and let $\eta \in (0,1)$ be a pre-specifed threshold value. Muirhead and Soaita (2012) define a "successful trial" (i.e. a trial where we are able to declare superiority) as one for which

$$Q_n(\Theta_1|x_{(n)}) \geqslant \eta. \tag{1}$$

At the design stage, the left side of (1) cannot be evaluated, as the sample has not yet been observed. Thus we consider the set of all samples leading to a successful trial,

*School of Statistics, University of Minnesota, Minneapolis, MN, eaton@stat.umn.edu
†Statistical Consultant, Lyme, CT, robb.muirhead@yahoo.com
‡Pfizer Inc, Groton, CT, adina.soaita@pfizer.com

© 2013 International Society for Bayesian Analysis DOI:10.1214/13-BA809

namely

$$E_n^* = \{x_{(n)} \,:\, Q_n(\Theta_1|x_{(n)}) \geqslant \eta\} \tag{2}$$

and look at the marginal probability (under the model and the prior) of E_n^*. In other words, we calculate

$$\psi(n) = \mathcal{E}\big\{I_{E_n^*}(X_{(n)})\big\} \tag{3}$$

where the expectation $\mathcal{E}$ is under the marginal distribution of $X_{(n)}$ and $I_{E_n^*}$ is the indicator function of the set E_n^*. Muirhead and Soaita (2012) refer to $\psi(n)$ in (3) as the "probability of a successful trial" (PST) and discuss its use in Bayesian SSD.

It is reasonable to refer to the set E_n^* as a "Bayesian rejection region". We note that the (classical) power function of the test of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ based on E_n^* is just

$$\beta_n(\theta) = \mathcal{E}_\theta\{I_{E_n^*}(X_{(n)})\}, \tag{4}$$

with expectation taken with respect to the model. Then $\psi(n)$ in (3) is the same as

$$\psi(n) = \int_\Theta \beta_n(\theta)\pi(d\theta) = \mathcal{E}_\pi\{\beta_n(\theta)\}, \tag{5}$$

with expectation taken with respect to the prior π. The expressions (3) and (5) may help to explain why there is no consistent terminology in the literature. When the rejection region of a test is taken to be one that is based on a frequentist (non-Bayesian) construction, $\psi(n)$ has been called "average power", "predictive power", and "assurance" (see Whitehead et al. (2008), O'Hagan et al. (2005)). When it is based on a Bayesian rejection region (such as E_n^* in (2)), $\psi(n)$ has also been called "expected Bayesian power" (see Spiegelhalter et al. (2004)) and "predictive probability" (see Brutti et al. (2008)). In what follows, we will generally refer to $\psi(n)$, based on the rejection region E_n^*, as either the PST or simply as "average power".

A goal of this paper is to show that for each $\eta \in (0,1)$, under conditions specified in Section 2

$$\lim_{n\to\infty} \psi(n) = \pi(\Theta_1), \tag{6}$$

i.e., the average power, as given in (3) and (5), converges to the prior probability that $\Theta \in \Theta_1$. This is Theorem 1 in Section 2. Brutti et al. (2008) derive (6) in a very special case, a single sample from a normal distribution with known variance. The limiting result in (6) plays an important role in the applied use of $\psi(n)$ in determining a sample size at the design stage. See Muirhead and Soaita (2012) for illustrative examples and further discussion.

The result (6) is established under two important assumptions. The first of these is that

$$\pi(\partial\Theta_1) = 0 \tag{7}$$

where $\partial\Theta_1$ is the topological boundary of Θ_1. We note that (7) is just the well known definition of "Θ_1 is a π-continuity set" (see Billingsley (1968)). The second assumption is that the posterior is "π-consistent". This notion, defined carefully in Section 2, has its origins in a result of Doob (1949). In essence, this result shows that in the case of an i.i.d. sample, there is a set $C_0 \subseteq \Theta$ such that $\pi(C_0) = 1$ and for each $\theta_0 \in C_0$, the posterior $Q_n(\cdot|X_{(n)})$ converges weakly to the point mass at θ_0, a.s. $P(\cdot|\theta_0)$. (Assumptions, and the nature of P, are detailed in Section 2.) Doob's theorem is discussed more completely in the Appendix, where an extension to the two sample problem is described. A careful statement and illustrative proof of Doob's theorem can be found in Ghosh and Ramamoorthi (2003).

In Section 2 we also examine the limiting behavior of "average power given that $\theta \in \Theta_1$", a quantity defined later in (18). The main results here are given in Theorems 2 and 3. Section 3 contains examples, the most important of which concerns the comparison of two normal means. The paper concludes with a discussion (Section 4).

2 Main Theorems

In this section, notation and assumptions are given, followed by statements and proofs. First, to formulate our asymptotic result, (6), consider a sequence of random vectors $X_{(n)} = (X_1, \ldots, X_n)$, $(n \geqslant 1)$, with each X_i in R^1. Thus the sample space for $X_{(n)}$ is R^n, for $n = 1, 2, \ldots$. The coordinates of $X_{(n)}$ are not assumed here to be independent, although they will be so in the examples in Section 3.

Let Θ be a parameter space which is assumed to be a Polish space. Let $P_n(\cdot|\theta)$ denote the distribution of $X_{(n)}$ on R^n, $n = 1, 2, \ldots$. We further assume that there is a probability space $(R^\infty, \mathcal{B}^\infty, P(\cdot|\theta))$ so that $P_n(\cdot|\theta)$ is the projection of $P(\cdot|\theta)$ onto R^n. That is, if B is a Borel set in R^n (so $B \times R^1 \times R^1 \times \cdots \in \mathcal{B}^\infty$), then

$$P_n(B|\theta) = P(B \times R^1 \times R^1 \times \cdots |\theta). \tag{8}$$

For $x = (x_1, x_2, \ldots) \in R^\infty$, let $x_{(n)}$ denote the vector of the first n coordinates of x. Because the $P_n(\cdot|\theta)$'s are the projections of $P(\cdot|\theta)$ onto R^n, for each integrable f defined on R^n, we have

$$\int_{R^n} f(x_{(n)}) P_n(dx_{(n)}|\theta) = \int_{R^\infty} f(x_{(n)}) P(dx|\theta) \tag{9}$$

for each $\theta \in \Theta$.

Next, let π be a proper prior distribution on Θ. Given $X_{(n)} = x_{(n)} \in R^n$, let $Q_n(\cdot|x_{(n)})$ be a version of the posterior distribution of θ.

Definition 2.1: The sequence of posteriors $\{Q_n(\cdot|x_{(n)})\}$ is **consistent** at $\theta_0 \in \Theta$ if there is a set $B_0 \subseteq R^\infty$ such that $P(B_0|\theta_0) = 1$, and for each $x \in B_0$ and each neighborhood U of θ_0,

$$\lim_{n\to\infty} Q_n(U|x_{(n)}) = 1.$$

Remark 2.1: Because Θ is a Polish space, it follows from Remark 1.3.1 in Ghosh and Ramamoorthi (2003) that $\{Q_n(\cdot|x_{(n)})\}$ is consistent at θ_0 iff $Q_n(\cdot|x_{(n)})$ converges weakly to δ_{θ_0} (point mass at θ_0) a.s. $P(\cdot|\theta_0)$.

Definition 2.2: The sequence of posteriors $\{Q_n(\cdot|x_{(n)})\}$ is **π-consistent** if there is a set $C_0 \subseteq \Theta$ such that $\pi(C_0) = 1$ and for each $\theta_0 \in C_0$, the sequence of posteriors is consistent at θ_0.

In essence, the theorem of Doob (1949) states that in the i.i.d. case, when both the sample space and parameter space are Polish spaces, then the sequence of posteriors is π-consistent for each prior π.

We now proceed to formulate Theorem 1 below that will imply equation (6). To this end, let C be a Borel subset of Θ. Fix $\eta \in (0, 1)$ and let

$$E_n = \{x_{(n)} \, : \, Q_n(C|x_{(n)}) \geqslant \eta\}. \tag{10}$$

Note that the dependence of E_n on the set C is suppressed. Our general result below is formulated for an arbitrary Borel set C.

With $M_n(\cdot)$ denoting the marginal distribution of $X_{(n)}$ (under the model and the prior), our main result describes the asymptotic behavior of

$$\psi(n) = M_n(E_n). \tag{11}$$

Obviously, for each Borel subset B of R^n,

$$\begin{aligned} M_n(B) &= \int_\Theta \int_{R^n} I_B(x_{(n)}) P_n(dx_{(n)}|\theta)\pi(d\theta) \\ &= \int_\Theta \int_{R^\infty} I_B(x_{(n)}) P(dx|\theta)\pi(d\theta). \end{aligned} \tag{12}$$

Finally, we recall some standard topological notation. For a subset $D \subseteq \Theta$, D° is the interior of D, $\bar{D}$ is the closure of D, D^c is the complement of D and ∂D is the boundary of D. Note that the boundary of D is equal to the boundary of D^c.

Theorem 1: Suppose that $\{Q_n(\cdot|x_{(n)})\}$ is π-consistent and that $\pi(\partial C) = 0$. Then

$$\lim_{n\to\infty} \psi(n) = \pi(C), \tag{13}$$

where $\psi(n)$ is given by (11).

Proof: By assumption, there is a set $C_0 \subseteq \Theta$ with $\pi(C_0) = 1$ and such that the sequence of posteriors is consistent at θ_0 for each $\theta_0 \in C_0$. Because $\pi(\partial C) = 0$, we can write

$$\begin{aligned} \psi(n) &= \int_{C^\circ} \int_{R^\infty} I_{E_n}(x_{(n)}) P(dx|\theta)\pi(d\theta) + \int_{(C^c)^\circ} \int_{R^\infty} I_{E_n}(x_{(n)}) P(dx|\theta)\pi(d\theta) \\ &= \psi_1(n) + \psi_2(n). \end{aligned} \tag{14}$$

To analyze $\psi_1(n)$, fix $\theta_0 \in C^\circ \cap C_0$. Since C is a neighborhood of θ_0 and the posterior is consistent at θ_0, it follows that

$$\lim_{n\to\infty} Q_n(C|x_{(n)}) = 1 \quad \text{a.s. } P(\cdot|\theta_0).$$

Thus

$$\lim_{n\to\infty} I_{E_n}(x_{(n)}) = 1 \quad \text{a.s. } P(\cdot|\theta_0).$$

Since $\pi(C_0) = 1$, we can apply the dominated convergence theorem to see that

$$\lim_{n\to\infty} \psi_1(n) = \int_{C^0} \pi(d\theta) = \pi(C^0).$$

But $\pi(\partial C) = 0$, so $\pi(C) = \pi(C^0)$.

A similar argument shows that

$$\lim_{n\to\infty} \psi_2(n) = 0.$$

This completes the proof.

The application of Theorem 1 to the problem described in Section 1 requires the verification of two main assumptions. We need to show that $\partial\Theta_1$ has prior probability zero and that the posterior distribution is π-consistent. In the i.i.d. case, the theorem of Doob (1949) applies to show that π-consistency holds. The boundary condition on Θ_1 needs to be checked in each application.

Remark 2.2: Theorem 1 above is proved under two assumptions, namely that the posterior is π-consistent and that $\pi(\partial C) = 0$. Modifications of these are possible while maintaining the validity of (13). For example, if we assume that (i) C is open,

$$\text{(ii)} \qquad Q_n(\partial C|x_{(n)}) \to 1 \quad \text{a.s. } P(\cdot|\theta) \quad \text{for all } \theta \in \partial C,$$

and (iii) π-consistency, then (13) holds. A simple modification of the proof of Theorem 1 establishes this assertion.

An example of this is the following. Assume $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$, $\Theta = R^1$, $\Theta_0 = (-\infty, 0]$, $\Theta_1 = (0, \infty)$, so $\partial\Theta_1 = \{0\}$. Consider the prior $\pi = \frac{1}{2}\pi_0 + \frac{1}{2}\pi_1$, where π_0 is a point mass at 0 and π_1 is $N(0, 1)$. With $C = \Theta_1, \pi(\partial C) = \frac{1}{2}$, but a routine calculation shows

$$Q_n(\{0\}|x_{(n)}) \to 1 \quad \text{a.s. } P(\cdot|0)$$

and we have

$$\psi(n) \to \frac{1}{4} = \pi(\Theta_1).$$

A frequentist version of Theorem 1 is easy to formulate. Suppose that R_n is the rejection region of a test of Θ_0 versus Θ_1 and let ϕ_n be the corresponding test function (the indicator function of R_n). The power function of the test ϕ_n is just

$$\beta_n(\theta) = \mathcal{E}_\theta(\phi_n),$$

with expectation taken with respect to the model. The frequentist testing procedure defined by ϕ_n is *asymptotically consistent* for testing Θ_0 versus Θ_1 if, as $n \to \infty$, $\beta_n(\theta)$

converges to 1 on the interior of Θ_1 and converges to 0 on the interior of Θ_0. Now, if π is a proper prior distribution on Θ such that $\pi(\partial\Theta_1) = 0$, it is routine to show that

$$\lim_{n\to\infty} \int_\Theta \beta_n(\theta)\pi(d\Theta) = \pi(\Theta_1)$$

for asymptotically consistent tests. This of course is a frequentist analog of Theorem 1. Taking $R_n = E_n^*$ (as defined in (2)) establishes the Bayesian-frequentist connection.

Because of the limit result (6), and because $\psi(n)$ is increasing in n in their examples, Muirhead and Soaita (2012) suggested that it may be preferable, when choosing a sample size, to focus attention on a "normalized index" $\psi^*(n)$ given by

$$\psi^*(n) = \frac{\psi(n)}{\pi(\Theta_1)}, \tag{15}$$

for which

$$\lim_{n\to\infty} \psi^*(n) = 1, \tag{16}$$

so that $\psi^*(n)$ represents the proportion of the maximum value of the PST (average power) explained by the sample size n. We now look at an alternative proposal.

Recall from (5) that the average power $\psi(n)$ may be written as

$$\psi(n) = \int_\Theta \beta_n(\theta)\pi(d\theta), \tag{17}$$

where $\beta_n(\theta)$ is the (classical) power function of the test with rejection region E_n^* given by (2). A referee suggested that since our hope is to be able to conclude that $\theta \in \Theta_1$, it would also be of interest to consider "average Bayesian power given that $\theta \in \Theta_1$"; that is, to investigate the quantity $\tilde{\psi}(n)$ given by

$$\tilde{\psi}(n) = \frac{1}{\pi(\Theta_1)} \int_{\Theta_1} \beta_n(\theta)\pi(d\theta). \tag{18}$$

Of course, $\tilde{\psi}(n)$ is the average power with respect to the prior distribution conditional on "$\theta \in \Theta_1$". That is, let

$$\tilde{\pi}(\mathcal{A}) = \frac{\pi(\mathcal{A} \cap \Theta_1)}{\pi(\Theta_1)}, \quad \mathcal{A} \subseteq \Theta. \tag{19}$$

Then $\tilde{\pi}$ is the conditional prior obtained from π given that $\theta \in \Theta_1$. It is clear that

$$\tilde{\psi}(n) = \int_\Theta \beta_n(\theta)\tilde{\pi}(d\theta). \tag{20}$$

It is now natural to ask how $\tilde{\psi}(n)$ behaves as $n \to \infty$. An obvious analog to Theorem 1 answers this question.

Theorem 2: Suppose that $\{Q_n(\cdot|x_{(n)})\}$ is π-consistent and that $\pi(\partial\Theta_1) = 0$. Then

$$\lim_{n\to\infty} \tilde{\psi}(n) = 1.$$

Proof: The proof is a minor modification of the proof of Theorem 1. The details are omitted.

A useful alternative form of Theorem 2 is available when the alternative Θ_1 is actually an open set, as in the case of many interesting examples.

Theorem 3: Suppose that $\{Q_n(\cdot|x_{(n)})\}$ is π-consistent and that Θ_1 is a non-empty open set. Then

$$\lim_{n\to\infty} \tilde{\psi}(n) = 1.$$

Proof: Since $\{Q_n(\cdot|x_{(n)})\}$ is π-consistent, the test with rejection region E_n^* is (frequentist) asymptotically consistent. The openness of Θ_1 implies that $\beta_n(\theta) \to 1$ for each $\theta \in \Theta_1$. The bounded convergence theorem gives the desired conclusion.

3 Examples

The simplest examples of Theorem 1 concern i.i.d. samples, since Doob's theorem implies π-consistency. For example, with data $X_{(n)} = (X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, consider testing $\Theta_0 = \{\mu, \sigma^2 |\, \mu \leqslant 0\}$ versus $\Theta_1 = \{\mu, \sigma^2 |\, \mu > 0\}$. For any prior π on $\Theta = \Theta_0 \cup \Theta_1$ that is absolutely continuous with respect to Lebesgue measure $d\mu d\sigma$, we have $\pi(\partial\Theta_1) = 0$, so Theorem 1 implies that $\psi(n)$ converges to $\pi(\Theta_1)$.

The main example of this section involves data from two independent normal samples.

Example 3.1: Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent, with the X's i.i.d. $N(\mu_1, \sigma^2)$ and the Y's i.i.d. $N(\mu_2, \sigma^2)$. We consider the case where the three parameters are unknown (the case of known σ^2 is a little easier). In this example,

$$\Theta = \{\mu_1, \mu_2, \sigma^2 |\, \mu_i \in R^1, i = 1, 2 \text{ and } \sigma > 0\},$$

$$\Theta_0 = \{\mu_1, \mu_2, \sigma^2 |\, \mu_1 \leqslant \mu_2\}, \quad \text{and} \quad \Theta_1 = \{\mu_1, \mu_2, \sigma^2 |\, \mu_1 > \mu_2\}.$$

Clearly $\partial\Theta_1 = \{\mu_1, \mu_2, \sigma^2 |\, \mu_1 = \mu_2\}$ and $\partial\Theta_1$ has Lebesgue measure zero when Θ is considered a subset of R^3.

Let π be a prior on Θ that is absolutely continuous with respect to Lebesgue measure. Given this prior for $\theta = (\mu_1, \mu_2, \sigma)$, let $Q_{m,n}(d\theta|X_{(m)}, Y_{(n)})$ denote the posterior. With $p = \min\{m, n\}$, the argument in the Appendix shows that when $p \to \infty$, the posterior is π-consistent. Since $\pi(\partial\Theta_1) = 0$, Theorem 1 implies that

$$\lim_{p\to\infty} \psi(p) = \pi(\Theta_1). \tag{21}$$

The use of this result in determining sample sizes is discussed in Muirhead and Soaita (2012).

Example 3.2: This example is intended to illustrate the two measures $\psi^*(n)$ given by (15) and $\tilde{\psi}(n)$ given by (18). It is based on Example 3.1.1 in Muirhead and Soaita (2012) where the subjects in the clinical trial have "Restless Legs Syndrome". Let U be the difference between the disease score sample means of two goups (experimental drug and placebo), with equal sample sizes n_1 in each group, so that the total sample size is $n = 2n_1$. Under standard normality assumptions, $U \sim N(\delta, 4\sigma^2/n)$, where σ^2 is the common variance in each group, and where $\delta > 0$ favors the experimental drug. The variance σ^2 is assumed known (from an earlier study), with $\sigma = 8$. For a 1-sided 0.025-level z-test to have 80% power at the "clinically meaningful difference"of 4 (drug versus placebo), 64 subjects in each group are required, for a total sample size of 128.

In one of the settings in Muirhead and Soaita (2012), the prior distribution for the treatment effect δ is $N(\Delta, 64)$, and the threshold in (1) used to define a "successful trial" is $\eta = 0.975$. Graphs of the normalized index $\psi^*(n)$ versus n are given in Figure 1 of Muirhead and Soaita (2012) for various values of the prior mean Δ. It is pointed out (for example) that, when $\Delta = 4$ (the "clinically meaningful" effect size), $\psi^*(100) = 0.79$, and there is not much to be gained (in terms of the PST (average power)) by taking a larger total sample size than $n = 100$. What is interesting is that, for the range of Δ's considered, there turns out to be *no discernible difference* between the graphs of $\psi^*(n)$ versus n and $\tilde{\psi}(n)$ versus n. The primary reason for this here is the fact that the threshold $\eta = 0.975$ is set very high, as it would need to be for regulatory drug approval.

A lower value of the the threshold η would be appropriate in exploratory and early phase trials, and this is where we see a difference between $\psi^*(n)$ and $\tilde{\psi}(n)$. Figure 1 gives the graphs of both $\psi^*(n)$ and $\tilde{\psi}(n)$ when $\eta = 0.8$ (and $\Delta = 4$). (Both functions converge to one as $n \to \infty$.) As is to be expected, $\tilde{\psi}(n) < \psi^*(n)$ for each n. For example, if the total sample size is $n = 20$, we have $\psi^*(20) = 0.82$ and $\tilde{\psi}(20) = 0.79$. For further discussion, see Section 4.

4 Discussion and summary

We began this paper with a focus on (unconditional) average power given in (3) and (5), namely

$$\psi(n) = P(X_{(n)} \in E_n^*) = P(\text{reject } H_0),$$

where the set E_n^* given in (2) is a Bayesian rejection region, and where the probability involved is computed using the marginal distribution of $X_{(n)}$. We established conditions (Theorem 1) under which, as $n \to \infty$,

$$\psi(n) \to \pi(\Theta_1) \tag{22}$$

the prior probability of H_1. (Under these conditions, the "normalized average power" $\psi^*(n)$ given by (15) converges to one.) We also noted that the limiting result (22) also holds if E_n^* is replaced by the rejection region of a test that is asymptotically consistent (in the usual classical sense).

We further considered the "conditional power" $\tilde{\psi}(n)$ given by (18), and established

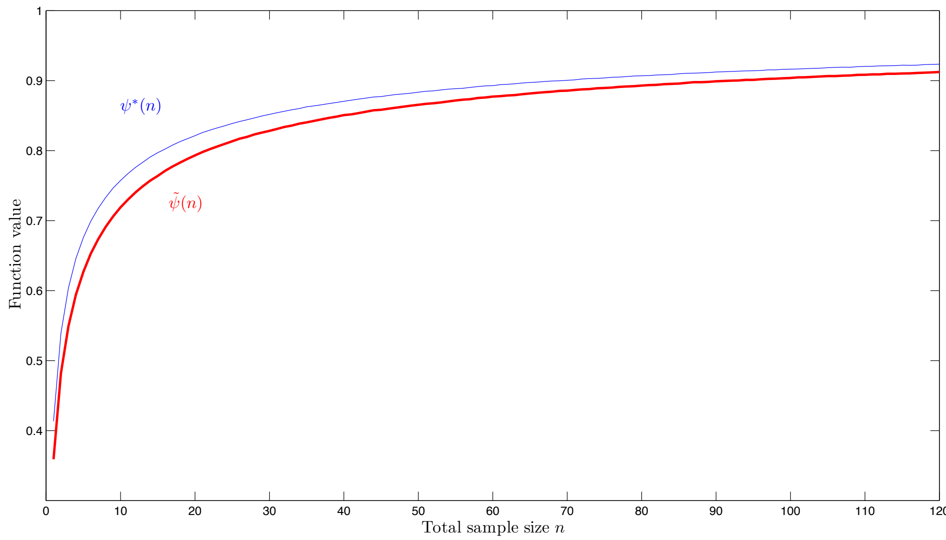

Figure 1: Graphs of $\psi^*(n)$ and $\tilde{\psi}(n)$ when the threshold is η=0.8 (and Δ=4).

conditions (Theorems 2 and 3) under which, as $n \to \infty$,

$$\tilde{\psi}(n) \to 1.$$

The use of the two quantities $\psi^*(n)$ and $\tilde{\psi}(n)$ in a sample sizing context was illustrated in Example 3.2.

The quantities ψ, ψ^*, and $\tilde{\psi}$ are, of course, related. We now detail this, suppressing the dependence on n. By definition, $\psi^* = \psi/\pi(\Theta_1)$. Conditioning on whether H_0 or H_1 is true gives

$$\psi = P(\text{reject } H_0 | H_1 \text{ true})\pi(\Theta_1) + P(\text{reject } H_0 | H_0 \text{ true})\pi(\Theta_0). \tag{23}$$

Rejecting H_0 when it is true represents a false positive (FP), and (23) may be written

$$\psi = \tilde{\psi} \cdot \pi(\Theta_1) + P(\text{FP})\pi(\Theta_0)$$

so that

$$\psi^* = \tilde{\psi} + P(\text{FP})\frac{\pi(\Theta_0)}{\pi(\Theta_1)}. \tag{24}$$

As $n \to \infty$ (and for any fixed $\eta \in (0, 1)$),

$$\psi^* \to 1, \quad \tilde{\psi} \to 1, \quad \text{and} \quad P(\text{FP}) \to 0.$$

Remarks:

(a) We noted in Example 3.2 that $\tilde{\psi} < \psi^*$. Equation (24) shows that ψ^* is "inflated" by the probability of a false positive. This latter probability decreases as the threshold η increases. (We noted in Example 3.2 that there was virtually no difference, for any n, between ψ^* and $\tilde{\psi}$ when $\eta = 0.975$, a high threshold. As η is decreased, a difference appears in the graphs; this difference, of course, decreases as n increases.)

(b) In a sample sizing problem, it is natural to ask: If the alternative is true, what is the probability that (using the Bayesian rejection region (2)) we *conclude* that this is so. This probability is the "conditional power" $\tilde{\psi}(n)$ (see (18) and (20)). We would then select a sample size n^* as the smallest value of n such that $\tilde{\psi}(n) \geqslant \gamma$, where $\gamma \in (0, 1)$ is a pre-specified number. In other words, n^* is the smallest value of n such that the conditional probability of deciding in favor of the alternative, given the alternative is true, is at least γ. In our view, the use of $\tilde{\psi}$ (rather than ψ or ψ^*) appears to be the most plausible of the three "average power functions" in explaining the application to sample sizing.

(c) Finally, there may be a technical advantage to preferring $\tilde{\psi}$ in some situations. Theorem 3 shows that the conditional power converges to 1, even when the boundary $\partial\Theta_1$ of Θ_1 has positive probability under the prior.

References

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. 222

Blackwell, D. and Dubins, L. (1962). "Merging of opinions with increasing information." *Annals of Mathematical Statistics*, 33: 882–886. 232

Brutti, P., De Santis, F., and Gubiotti, S. (2008). "Robust Bayesian sample size determination in clinical trials." *Statistics in Medicine*, 27: 2290–2306. 222

Doob, J. L. (1949). "Application of the theory of martingales." In *Le Calcul des Probabilites et ses Applications*, 23–27. Centre National de la Recherche Scientifique, Paris. Colloques Internationalaux du Centre National de le Recherche Scientifique, no 13. 222, 224, 225

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. New York: Springer Verlag. 223, 231

Muirhead, R. J. and Soaita, A. I. (2012). "On an approach to Bayesian sample sizing in clinical trials." In Jones, G. and Shen, X. (eds.), *Multivariate Statistics in Modern Statistical Analysis: A Festschrift in honor of Morris L. Eaton*. Institute of Mathematical Statistics, Beachwood, Ohio. To appear. Available online at arXiv:1204.4460v1. 221, 222, 226, 227, 228

O'Hagan, A., Steven, J. W., and Campbell, M. J. (2005). "Assurance in clinical trial design." *Pharmaceutical Statistics*, 4: 187–201. 222

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health Care Evaluation*. New York: Wiley. 222

Whitehead, J., Valdes-Marquez, E., Johnson, P., and Graham, G. (2008). "Bayesian sample size for exploratory clinical trials incorporating historical data." *Statistics in Medicine*, 27: 2307–2327. 222

Appendix

The purpose of this appendix is to argue that in the two sample case of Example 3.1, the posterior is π-consistent. Of course, assumptions are needed to make this claim correct. These will be detailed in the discussion below.

Consider sequences $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ of real valued random variables and assume that the X's are i.i.d. $P_1(\cdot|\theta)$, the Y's are i.i.d. $P_2(\cdot|\theta)$, and the X's and Y's are independent. The parameter θ is an element of a Polish space Θ. Given a positive integer k, let $Z_{(k)} = (Z_1, \ldots, Z_k)$ where $Z_i = (X_i, Y_i), i = 1, 2, \ldots$. Then the Z_i's are i.i.d. on R^2 with distribution $P_3(\cdot|\theta) \equiv P_1(\cdot|\theta) \times P_2(\cdot|\theta)$. Let $\mathcal{F}_k$ be the σ-field generated by $Z_{(k)}$ and let π be a prior distribution on Θ. Under the assumption that the map $\theta \longleftrightarrow P_3(\cdot|\theta)$ is $1-1$, Doob's Theorem (see Theorem 1.3.2 on page 22 of Ghosh and Ramamoorthi (2003)) implies that the posterior distribution $Q_k(\cdot|Z_{(k)})$ is π-consistent. Of course, in standard σ-field notation,

$$Q_k(\cdot|Z_{(k)}) = Q_k(\cdot|\mathcal{F}_k).$$

A basic step in the proof of Doob's Theorem uses the Martingale Convergence Theorem to conclude that for each Borel set $C \subseteq \Theta$,

$$\lim_{k\to\infty} Q_k(C|Z_{(k)}) = \lim_{k\to\infty} \mathcal{E}(I_C|\mathcal{F}_k) = \mathcal{E}(I_C|\mathcal{F}_\infty),$$

where $\mathcal{F}_\infty$ is the limit of increasing σ-fields $\mathcal{F}_k$. One then shows there is a set $\Theta_0 \subseteq \Theta$ such that $\pi(\Theta_0) = 1$ and for $\theta \in \Theta_0 \cap C$,

$$\mathcal{E}(I_C|\mathcal{F}_\infty) = 1 \quad \text{a.e. } P^\infty(\cdot|\theta), \tag{25}$$

where $P^\infty(\cdot|\theta)$ is the infinite product measure $P_3(\cdot|\theta) \times P_3(\cdot|\theta) \times \cdots$. That this establishes the π-consistency is argued on pages 23-24 of Ghosh and Ramamoorthi (2003).

To apply the above to the two sample problem of Example 3.1, consider sequences $m_1, m_2, \ldots$ and $n_1, n_2, \ldots$ of non-decreasing positive integers both converging to infinity. Let $(X_1, \ldots, X_{m_p}) = X_{(m_p)}$ and $(Y_1, \ldots, Y_{n_p}) = Y_{(n_p)}$ be a sample of X's and Y's. Let $\mathcal{G}_p$ be the σ-field generated by $\{X_{(m_p)}, Y_{(n_p)}\}$ and let $Q_p^*(\cdot|\mathcal{G}_p)$ be a version of the conditional distribution of θ given $\{X_{(m_p)}, Y_{(n_p)}\}$. Thus, for a Borel set $C \subseteq \Theta$,

$$Q_p^*(C|\mathcal{G}_p) = \mathcal{E}(I_C|\mathcal{G}_p)$$

for $p = 1, 2, \ldots$. With $k = \min\{n_p, m_p\}$, note that

$$\mathcal{F}_k \subseteq \mathcal{G}_p \subseteq \mathcal{F}_\infty.$$

Since $\mathcal{F}_k$ converges to $\mathcal{F}_\infty$, it follows that $\mathcal{G}_p$ converges to $\mathcal{F}_\infty$. This implies (see Theorem 2 in Blackwell and Dubins (1962)) that

$$\lim_{p\to\infty} \mathcal{E}(I_C|\mathcal{G}_p) = \mathcal{E}(I_C|\mathcal{F}_\infty).$$

However, this is characterized by (25) which in turn shows that $Q_p^*(\cdot|\mathcal{G}_p)$ is a π-consistent sequence of posteriors.

It is clear that the above argument can be extended to the r-sample case to establish the π-consistency of a posterior. Note that π-consistency can fail rather dramatically when observations are not independent. For example, consider $X_{(n)} \in R^n$ which is $N_n(\theta, \Sigma_0)$, where the vector of means $\theta \in R^n$ has all its coordinates equal to an unknown parameter μ, and where the $n \times n$ covariance matrix Σ_0 has diagonal elements equal to 1 and off-diagonal elements equal to a known value $\rho \in (0, 1)$. If we take a $N(0, 1)$ prior for μ, the posterior can be calculated explicitly and is not π-consistent.

Acknowledgments

The authors would like to thank a referee for many thoughtful and insightful comments.