

# Properties of Estimators in Exponential Family Settings with Observation-based Stopping Rules

Elasma Milanzi<sup>1</sup>, Geert Molenberghs<sup>1,2\*</sup>, Ariel Alonso<sup>3</sup>, Michael GK<sup>4</sup>, Geert Verbeke<sup>1,2</sup>, Anastasios AT<sup>5</sup> and Marie Davidian<sup>5</sup>

<sup>1</sup>I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium

<sup>2</sup>I-BioStat, Katholieke Universiteit Leuven, Leuven, Belgium

<sup>3</sup>Department of Methodology and Statistics, Maastricht University, Netherlands

<sup>4</sup>Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

<sup>5</sup>Department of Statistics, North Carolina State University, Raleigh, NC, USA

## Abstract

Often, sample size is not fixed by design. A key example is a sequential trial with a stopping rule, where stopping is based on what has been observed at an interim look. While such designs are used for time and cost efficiency, and hypothesis testing theory has been well developed, estimation following a sequential trial is a challenging, still controversial problem. Progress has been made in the literature, predominantly for normal outcomes and/or for a deterministic stopping rule. Here, we place these settings in a broader context of outcomes following an exponential family distribution and, with a stochastic stopping rule that includes a deterministic rule and completely random sample size as special cases. It is shown that the estimation problem is usually simpler than often thought. In particular, it is established that the ordinary sample average is a very sensible choice, contrary to commonly encountered statements. We study (1) The so-called incompleteness property of the sufficient statistics, (2) A general class of linear estimators, and (3) Joint and conditional likelihood estimation. Apart from the general exponential family setting, normal and binary outcomes are considered as key examples. While our results hold for a general number of looks, for ease of exposition, we focus on the simple yet generic setting of two possible sample sizes,  $N=n$  or  $N=2n$ .

**Keywords:** Completely random sample size; Frequentist inference; Generalized sample average; Stochastic stopping rule; Joint modeling; Likelihood inference; Missing at random

## Introduction

It is commonly known that statistical designs where the sample size is random pose challenges beyond the fixed sample-size case and that many findings are counter-intuitive. While this has been documented for situations where the sample size depends on the data, such as in sequential trials [1,2] or incomplete data [3], it is less widespread that such counterintuitive results apply even when the sample size is completely random, Barndorff-Nielsen and Cox, in the sense that both the collected and uncollected data have no relationship to the stochastic mechanism governing the sample size. Liu and Hall [4] provided a general theory for sequential studies, where the decision to either stop or continue the study at every interim look depends deterministically on the data collected up to that point. Molenberghs et al. [5] generalized their results to the setting where the sample size may depend stochastically rather than deterministically on the observed data, a general setting that contains both sequential trials and completely random sample sizes (CRSS) as special cases. This means that the probability to stop is a function of the observed data. In other words, a coin is tossed to either stop or continue, but the probability to 'land on heads' of the coin is determined by the data observed thus far. We refer to these three settings together as a stochastic stopping rule. In practice, the deterministic stopping rule may be of highest relevance (e.g., in group sequential clinical trials). Also the CRSS may occur in practice, for example, in settings where a fixed time for the experiment is available, rather than a fixed sample size. Nevertheless, the stochastic stopping rule is technically convenient, because it allows to derive results for the deterministic case as a limit. Derivations for the stochastic rule are mathematically convenient. Molenberghs et al. [5] also discussed the related cases of incomplete longitudinal data, censored time-to-event data, joint modeling of survival and longitudinal data, and clustered data with random cluster sizes.

An important finding of Liu and Hall [4] was that the commonly used sufficient statistics in deterministic stopping designs are incomplete, a property that will be defined in the next section. Molenberghs et al. [5] generalized this to stochastic stopping rules and explore the implications of this for linear estimators based on the sample sum as well as on so-called marginal and conditional estimators. They found for stochastic stopping rules that the counterintuitive implications of a random sample size follows from two properties: (a) excluding the CRSS case, the sample size is *non-ancillary* given the sample sum; (b) the pair consisting of the accumulating sample sum and the sample size is an *incomplete* minimal sufficient statistic. These properties are defined in Section 2.

The work of Liu and Hall [4] and Molenberghs et al. [5] was confined to the special case of normally distributed outcomes. Further, Molenberghs et al. [5] illustrated their developments with a random stopping rule of probit form. These specific choices allow for insightful expressions. The latter choice is not, however, necessary for deterministic stopping rules that can be cast in the form of continuation and stopping regions or, equivalently, the boundaries between them.

While the restriction to normally distributed outcomes was a natural choice, in practice, various data types occur as well (binary,

**\*Corresponding author:** Geert Molenberghs, I-BioStat, Universiteit Hasselt, Campus Diepenbeek, Agoralaan Building D, BE 3590 Diepenbeek Diepenbeek, Belgium, Tel: 32-11-268238; E-mail: [geert.molenberghs@uhasselt.be](mailto:geert.molenberghs@uhasselt.be)

**Received** January 13, 2016 **Accepted** January 18, 2016; **Published** January 25, 2016

**Citation:** Milanzi E, Molenberghs G, Alonso A, Michael GK, Verbeke G, et al. (2016) Properties of Estimators in Exponential Family Settings with Observation-based Stopping Rules. J Biom Biostat 7: 272. doi:10.4172/2155-6180.1000272

**Copyright:** © 2016 Milanzi E, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

categorical, count, time-to-event), but one cannot simply assume that the normal-distribution-based properties simply will carry over. Extending the results in [4] presented a general deterministic stopping rule theory where the outcome follows a one-parameter exponential family, and also established incompleteness for this case. This implies, in particular, that there are infinitely many unbiased estimators, none with uniformly minimum variance. Here, we show incompleteness in the exponential family case, for a stochastic stopping rule, and derive explicit results for linear estimators as well as for marginal and conditional likelihood estimators. These general findings are then further illustrated in the normal case, making the connection to Molenberghs et al. [5], and in the case where the outcomes are binary, and hence the sample sum is binomial. In doing so, we extend the work by Molenberghs et al. [5] to most commonly encountered data types.

Our findings are essentially as follows. The classical sample average is biased in finite samples, though asymptotically unbiased for a broad classes of stopping rules. An unbiased estimator follows from the conditional likelihood, where the conditioning is on the (non-ancillary) sample size. Contrary to intuition, the conditional estimator has larger mean squared error than the ordinary sample average for sufficiently large sample size, the latter resulting from the joint likelihood, where ‘joint’ means a simultaneous model for the outcomes and the sample size. In some cases, the result holds for all sample sizes, large and small. Thus, the sample average is a valid and sensible estimator, contrary to some claims in the sequential-trial literature, for stochastic and deterministic stopping rules. The literature on sequential trials is indeed very large, with a relatively early review given by Whitehead. [Tsiatis, Rosner, and Mehta(1984)] and [Rosner and Tsiatis(1988)] address precision estimation after group sequential trials. Emerson and Fleming [2] propose estimators within an ordering paradigm. Much of this work is placed in a unifying framework by Liu and Hall [4]. A review can be found in Molenberghs et al. [5].

The finite-sample bias in the sample average disappears only in the CRSS case. Even then, it is not unique in that a whole class of so-called generalized sample average estimators can be defined, all of which are unbiased. This enables us to show that the ordinary sample average is only asymptotically optimal. Indeed there is no uniformly optimal unbiased estimator in finite samples for most exponential-family members; the exponential distribution is a noteworthy exception.

The case of two possible sample sizes,  $N=n$  and  $N=2n$  is simple yet generic, and will be adopted here, essentially without loss of generality. All developments can be generalized with ease to the setting with  $L$  possible sample sizes and accrual numbers  $n_1, \dots, n_L$ .

The remainder of this paper is organized as follows. In Section 2, the problem under investigation is formally introduced, along with key concepts. The incompleteness of the sufficient statistics is established in Section 3. Section 4 is devoted to generalized sample averages, while joint and conditional likelihood estimation is the topic of Section 5. In each of Sections 3–5, the general exponential family case is supplemented with the particular case of the normal and Bernoulli distributions.

## Notation, Basic Concepts and Problem Formulation

As stated in the introduction, we consider a simple sequential trial, where  $n$  measurements  $Y_i$  are observed, after which a stochastic stopping rule is applied and, depending on the outcome, another set of  $n$  measurements is or is not observed. Let  $Y$  be the  $(2n \times 1)$  vector of outcomes that could be collected, with the sample sum denoted by  $\kappa$ ,

and  $N$  be the realized sample size, that is,  $N=n$  or  $N=2n$ . A joint model for the stochastic outcomes is

$$f(y, N | \theta, \psi) = f(y | \theta) \cdot f(N | y, \psi) \quad (1)$$

$$= f(y | N, \theta, \psi) \cdot f(N | \theta, \psi). \quad (2)$$

The sample sum is denoted by  $\kappa$ . If necessary, a subscript will indicate over which batch the sample is calculated. Molenberghs et al.[5] noted the similarity with missing-data concepts, where (1) is a selection model factorization and (2) is a pattern-mixture factorization [3]. In all cases, it is assumed that  $f(N | y, \psi) = f(N | y^*, \psi)$  depends on observed outcomes only, and hence the sample size is determined by the first batch of observations,  $Y_1, \dots, Y_n$ . We may then write  $f(N | \kappa_n, \psi)$ . This corresponds to the frequentist concept of missingness at random [Little and Rubin(2002)]. In the limiting case of a deterministic stopping rule,  $f(N | y, \psi)$  is degenerate and  $f(N=n | y)$  equals 1 when  $K \in S \subset IR$  and 0 over its complement  $C$ , with the reverse holding for  $f(N=2n | y, \psi)$ . The CRSS case follows by assuming  $Y$  and  $N$  to be independent, meaning that both factorizations (1) and (2) trivially reduce to  $f(y | \theta) \cdot f(N | \psi)$ .

In the stopping-rule case  $\psi$  is not estimable from the data and will be assumed to be specified by design. This is different for the other settings that can also be cast in terms of (1)–(2), such as incomplete longitudinal data, clusters of random size, censored time-to-event data, joint models for longitudinal and time-to-event data, and random measurement times settings, as noted by Molenberghs et al.[5]. In these cases, a subject-specific index  $i$  needs to be introduced into (1)–(2) and  $N$  needs to be replaced by the missing data indicators, censoring indicators, and so on.

### Basic concepts

In line with Molenberghs et al. [5], we will review several fundamental concepts that are essential in what follows.

In line with [Rubin(1976)], we consider *ignorability*. For pure likelihood or Bayesian inferences, under missingness at random (MAR), inferences about  $\theta$  can be made using  $f(y_i | \theta)$  only, without the need for an explicit missing-data mechanism or, in our case, without the need for an explicit sample-size model. This is, provided the regularity condition of *separability* holds true, i. e., that the parameter space of  $(\theta', \psi')$  is the Cartesian product of their individual product spaces. In other words, this means that the sample size model does not contain information about the outcome model parameter. It implies that  $N$  could then be considered ancillary in the sense of Cox and Hinkley[6]. We will see that this is true for CRSS, but not for the other situations. Excluding MNAR, ignorability can be violated in three ways. First, even in the likelihood and Bayesian frameworks and under MAR, ignorability does not apply in a non-separable situation. Second, frequentist inferences are not necessarily ignorable under MAR. Third, assuming MAR and separability hold and we are in a likelihood or Bayesian framework, ignorability in the selection model decomposition (1) does not translate to the pattern-mixture model (2), as is clear from the presence of both  $\theta$  and  $\psi$  in both factors of (2). The latter statement is symmetric and could be made starting from a pattern-mixture view as well. The bottom line is that ignorability holds in at most one of these, except in the trivial MCAR setting, such as for CRSS.

There is a connection between ignorability and *ancillarity* [6]. They define an ancillary statistic  $T$  to be one that complements a minimally sufficient statistic  $s$  such that, given  $s$ ,  $T$  does not contain information

about the parameter of interest. Arguably the best known example is the sample size  $T=n$  when estimating a mean, provided the sample size is fixed by design or the law governing it does not depend on the mean parameter to be estimated, as with CRSS. Counterexamples are the stochastic and deterministic stopping rules.

The crucial property for Liu and Hall [4], Molenberghs et al.[5], as well as for us here is that of *completeness* [7]. A statistic  $s(y)$  of a random variable  $Y$ , with  $Y$  belonging to a family  $P_\theta$ , is complete if, for every measurable function  $g(\cdot)$ ,  $E[g(s(Y))]=0$  for all  $\theta$ , implies that  $P_\theta[g(s(Y))=0]=1$  for all  $\theta$ . The relevance of completeness for us surfaces in two ways. First, from the Lehman-Scheffé theorem [7], if a statistic is unbiased, complete, and sufficient for some parameter  $\theta$ , then it is the best mean-unbiased estimator for  $\theta$ . The lack of this property in the stopping-rule case will manifest itself when studying generalized sample averages in Section 4. Second, completeness and ancillarity are connected through Basu's theorem [7, 8]: a statistic both complete and sufficient is independent of any ancillary statistic.

### General model formulation

Assume that we collect  $n$  i. i. d. observations  $Y_1, \dots, Y_n$ , with exponential family density

$$f_\theta(y) = h(y) \exp\{\theta y - a(\theta)\}, \tag{3}$$

where  $\theta$  is the natural parameter,  $a(\theta)$  the mean generating function, and  $h(y)$  a normalizing constant. Assume a stochastic stopping rule

$$\pi(N = n | k_n) = F(k_n | \psi) = F(k_n), \tag{4}$$

with  $K_n = \sum_{i=1}^n Y_i$ . The form for (4) is left unspecified at this time. The CRSS setting follows as  $F(k_n) \equiv F$ , a constant. Likewise, when  $F(\cdot)$  is degenerate, a deterministic stopping rule ensues. When the trial is not stopped, a further  $n$  observations  $Y_{n+1}, \dots, Y_{2n}$  are collected, also with density (3). The inferential goal is to estimate  $\theta$  or a function of this, such as the population mean  $\mu$ . From the exponential-family structure, the density of  $K_n$  can be expressed

$$f_{n,\theta}(k) = h_n(k) \exp\{\theta k - na(\theta)\}. \tag{5}$$

When no ambiguity can arise, the subscript  $n$  may be dropped from  $K_n$ . Because the density integrates to 1, it trivially follows that

$$e^{na(\theta)} = \int h_n(k) e^{\theta k} dk = L\{h_n(k)\}. \tag{6}$$

While expression (6) is well known to be a Laplace transformation, it is useful to state it explicitly in preparation of the derivations in Section 3. Because the stopping rule depends on  $K_n$ , and because (4) combined with the outcome model is a pattern-mixture factorization (2),  $N$  is not ancillary to  $K$ .

When, in addition, the conditional probability of stopping is chosen to have an exponential family form, e.g.,

$$F(k_n) = F(k) = \int_{z=-\infty}^{z=A(k)} \tilde{f}_1(z) dz, \tag{7}$$

then an appealing form for the marginal stopping probability can be derived. Here  $\tilde{f}_1(z)$  can be seen as an exponential family member, underlying the stopping process. When the outcomes  $Y$  and hence  $K$  do not range over the entire real line, the lower integration limit in (7) should be adjusted accordingly, and the function  $A(K)$  should be chosen so as to obey the range restrictions. It is convenient to assume that  $\tilde{f}_1(z)$  has no free parameters; should there be the need for such, then they can be absorbed into  $A(K)$ . Hence, we can write

$$\tilde{f}_1(z) = \tilde{h}_1(z) \exp\{-\tilde{a}(0)\}. \tag{8}$$

Using (5) and (8), the marginal stopping probability becomes:

$$\begin{aligned} P(N = n) &= \int_{k=-\infty}^{k=+\infty} \int_{z=-\infty}^{z=A(k)} f_{n,\theta}(k) \tilde{f}_1(z) dz dk \\ &= \exp\{-na(\theta) - \tilde{a}(0)\} \int_{k=-\infty}^{k=+\infty} h_n(k) \left[ \int_{z=-\infty}^{z=A(k)} \tilde{h}_1(z) dz \right] e^{\theta k} dk \\ &= \exp\{-na(\theta) - \tilde{a}(0)\} L\{H_1(A(k)) \cdot h_n(k)\}, \end{aligned} \tag{9}$$

where

$$H_1(t) = \int_{z=-\infty}^{z=t} \tilde{h}_1(z) dz.$$

In the special case of a CRSS,  $A(k) \equiv A$  and (9) reduces to

$$\begin{aligned} f(N = n) &= \exp\{-na(\theta) - \tilde{a}(0)\} H_1(A) L\{h_n(k)\} \\ &= \exp\{-na(\theta) - \tilde{a}(0)\} H_1(A) e^{na(\theta)} = \int_{k=-\infty}^A \tilde{f}_1(k) dk. \end{aligned}$$

In our two special cases, (3) will be chosen as standard normal and Bernoulli, respectively. In the first of these, in concordance with Molenberghs et al. [5], (4) will be assumed to be of probit form:

$$F(k) = \Phi\left(\alpha + \beta \frac{k}{n}\right). \tag{10}$$

In the binary case, we will generally leave (4) unspecified, but for some developments it is useful to consider an explicit example, for which we will resort to the beta distribution, i. e.,

$$\tilde{f}_1(z) = \frac{z^{\alpha-1} (1-z)^{\beta-1}}{B(\alpha, \beta)}, \tag{11}$$

with  $B(\cdot, \cdot)$  the beta function. It is convenient to choose integer values, for illustrative purposes:  $\alpha = P + 1$ ,  $\beta = q + 1$  with  $P$  and  $q$  integers, changing (11) to:

$$\tilde{f}_1(z) = (p+q+1) \binom{p+q}{p} z^p (1-z)^q. \tag{12}$$

Choosing (12) leads to the conditional stopping probability:

$$F(k) = (p+q+1) \binom{p+q}{p} \sum_{\ell=0}^q \frac{(-1)^\ell}{p+\ell+1} \binom{q}{\ell} A(k)^{p+\ell+1}. \tag{13}$$

It is instructive to consider some special cases of this. When  $p=q=1$ , (12) reduces to the uniform distribution on the unit interval, and it immediately follows that  $F(k)=A(k)$ . When  $p=1$  and  $q=0$ , we find  $F(k)=A(k)^2$ . As a third and last instance, when  $p=q=1$ ,  $F(k)=3A(k)^2 - 2A(k)^3$ .

A useful function is  $A(k)=k/n$ , implying that stopping is certain when  $K=n$  and continuation is certain when  $K=0$ , while for  $0 < K < n$  stopping is probabilistic. The actual probability in these cases depends on the choice for  $p$  and  $q$ .

These choices are made to illustrate our general developments and our emphasis is not on, say, designing a particular trial. However, the class of beta-based stopping rules, for example, potentially leads to rich families of stopping rules and spending functions [9].

## Incomplete Sufficient Statistics

### The general case

We now consider the role of completeness in this setting, building upon the work of Liu and Hall [4], Liu et al. and Molenberghs et al. [5]. A sufficient statistic for this setting is  $(K, N)$ . In line with the developments in the above papers, the joint distribution for  $(K, N)$  is:

$$p(k, n) = f_n(k) \cdot F(k), \tag{14}$$

$$p(k, 2n) = f_{2n}(k) - \int f_n(z) f_n(k-z) F(z) dz. \tag{15}$$

When the stopping rule leads to range restrictions in the sense of Lehman and Stein [10], it is known that the sufficient statistic is complete. Hence, for the rest of this section, we assume their necessary and sufficient conditions do not hold. It is known that these conditions do not hold for the normal distribution, in contrast to classes of stopping rules for the Poisson and binomial distributions, for example.

Assume now that a function  $g(K, N)$  exists such that its expectation is zero for all values of the parameter and further that integrands are not zero almost everywhere over their integration ranges. Such a function must satisfy:

$$\int g(k, n) f_n(k) F(k) dk + \int g(k, 2n) f_{2n}(k) dk - \int \int g(k, 2n) f_n(z) f_n(k-z) F(z) dz dk = 0. \quad (16)$$

Substituting the general exponential form (5) into (16), and using (6), leads to

$$\int h_n(k) e^{ok} dk \cdot \int g(k, n) h_n(k) F(k) e^{ok} dk - \int g(k, 2n) h_{2n}(k) e^{ok} dk - \int g(k, 2n) \left[ \int h_n(z) h_n(k-z) F(z) dz \right] e^{ok} dk. \quad (17)$$

Because the left hand side of (17) is a convolution, and using the uniqueness of the Laplace transform, we find:

$$g(k, 2n) = - \frac{\int g(z, n) h_n(z) h_n(k-z) F(z) dz}{h_{2n}(k) - \int h_n(z) h_n(k-z) F(z) dz}. \quad (18)$$

Hence, when  $g(k, n)$  is chosen arbitrarily, (18) prescribes the choice for  $g(k, 2n)$  which leads to a counterexample to completeness, hence establishing incompleteness.

For the CRSS case, when  $F(k)=F$ , a constant, and also choosing  $g(k, n)=c$ , a constant, it follows that

$$g(k, 2n) = \frac{-F}{1-F} \cdot c.$$

In the limiting case of a deterministic stopping rule,  $F(Z)=1$  over the stopping region  $S$  and 0 over its set complement  $C$ . It then follows that (14)–(15) reduce to:

$$p(k, n) = f_n(k) \cdot I(k \in S), \quad (19)$$

$$p(k, 2n) = \int_C f_n(z) f_n(k-z) dz. \quad (20)$$

For the deterministic case, (18) becomes:

$$g(k, 2n) = - \frac{\int_S g(z, n) h_n(z) h_n(k-z) dz}{h_{2n}(k) - \int_S h_n(z) h_n(k-z) dz} = - \frac{\int_S g(z, n) h_n(z) h_n(k-z) dz}{\int_C h_n(z) h_n(k-z) dz}. \quad (21)$$

Expression (21) follows from the fact that, in the deterministic case,  $F(k)=1$  over the stopping region  $S$  and 0 elsewhere. The transition from one denominator to the other follows from observing that the convolution of  $f_n(k)$  with itself produces  $f_{2n}(k)$ , and then replacing all of these by their explicit exponential-family form (5). Alternatively, it is easy to show that (21) follows immediately from the definition of a function  $G(K, N)$  and (19)–(20).

Evidently, these results agree with Liu et al. for the limiting case of a deterministic stopping rule. These authors establish incompleteness for the exponential-family setting, under a regularity condition regarding the support spaces, which is sufficient but not necessary. Hence, when this condition is not satisfied, there may be exceptions to this result [11]. This issue falls outside the scope of this paper.

The implication of these findings is that whenever they hold, the Lehmann-Scheffé theorem cannot be applied (Section 2). It follows

that a best mean-unbiased estimator does not necessarily exist for the average. In the next section, it will be shown that this is indeed the case for many, but not all outcome distributions and stopping rules, given that, for example, the exponential distribution does admit a uniform optimum. It will be shown that no optimum exists for the normal case, in line with Molenberghs et al. [5], and neither for the Bernoulli and Poisson cases, for a wide class of stopping rules.

### The normal case

In this section, we summarize the arguments in Molenberghs et al. [5]. The same is true for Sections 4. 2 and 5. 2. Consider the outcome to be standard normal with mean  $\mu$  and let stopping be governed by (10). They derived from first principles that the marginal probability of stopping is:

$$P(N = n) = \Phi \left( \frac{\alpha + \beta \mu}{\sqrt{1 + \beta^2 / n}} \right). \quad (22)$$

This expression also follows as a special case of (9) by choosing (10) as the stopping rule, i. e.,  $\tilde{f}_1(z)$  as the standard normal density and  $A(k) = \alpha + \beta k/n$ , and further  $f_{n, \theta = \mu} = \varphi_{\mu, s}(k)$ , where  $\varphi_{\mu, s}(k)$  is the normal density with mean  $\mu$  and variance  $s$ . Details of this derivation are provided in Appendix 1.

Clearly, (22) depends on  $\mu$ , implying that this pattern-mixture formulation is non-separable. In contrast, although the observed data are present in the conditional stopping probability,  $\mu$  is not, implying separability in the selection model formulation.

In this case (14)–(15) takes the form

$$p_\mu(N, k) = p_0(N, k) \cdot \exp \left( k \mu - \frac{1}{2} n \mu^2 \right) \quad (23)$$

with

$$p_0(n, k) = \phi_n(k) \cdot \Phi \left( \alpha + \frac{\beta}{n} k \right), \quad (24)$$

$$p_0(2n, k) = \phi_{2n}(k) \cdot \left[ 1 - \Phi \left( \frac{\alpha + \frac{\beta k}{2n}}{\sqrt{\frac{2n + \beta^2}{2n}}} \right) \right]. \quad (25)$$

Here,  $\phi_s(k)$  is the normal density with mean 0 and variance  $s$ . Expression (25) is more explicit than (15), making use of the fact that the outcome densities are normal and the stopping probability is written as a normal cumulative distribution function. The derivation can be found in Molenberghs et al. [5]. Based on the fact that integrating the joint densities specified by (23)–(25) over  $k$  and summing over  $N$  should be equal to one, leads to the identity:

$$\int \varphi_{\mu, n}(k) \cdot \Phi \left( \alpha + \frac{\beta}{n} k \right) dk = \int \varphi_{\mu, 2n}(k) \cdot \Phi \left( \frac{\alpha + \frac{\beta k}{2n}}{\sqrt{\frac{2n + \beta^2}{2n}}} \right) dk. \quad (26)$$

In Section 4. 1, (26) will be derived in general.

The specific form of condition (18) is:

$$g(k, 2n) \cdot p_0(2n, k) = - \int \phi_n(k-z) \cdot g(z, n) \cdot \phi_n(z) \cdot \Phi \left( \alpha + \frac{\beta}{n} z \right) dz. \quad (27)$$

In the CRSS case, (24)–(25) reduce to:

$$p_0(n, k) = \phi_n(k) \cdot \Phi, \quad (28)$$

$$p_0(2n, k) = \phi_{2n}(k) \cdot (1 - \Phi), \quad (29)$$

where  $\Phi \equiv \Phi(\alpha)$ . Then here, as in the general case, (27) simplifies and leads to an explicit solution for a number of cases, especially when  $g(k, n)$  is chosen to be a constant.

In addition, for this case, other explicit examples can be constructed, even when  $\beta \neq 0$ . We reproduce the two examples of Molenberghs et al. [5] (Appendix 2).

### The binary case

While the binary case follows from the general considerations given in Section 3.1, it is insightful to examine this outcome type in some detail; here, integration is replaced by summation. Let the Bernoulli probability be  $\pi$ . The sample sum  $k$  then follows a *Bin* ( $\pi, N$ ) distribution and

$$f_{N,\pi}(k) = \binom{N}{k} \pi^k (1 - \pi)^{N-k}. \quad (30)$$

For now, as in the general case, we leave  $F(k)$  unspecified. The joint distribution of  $(K, N)$  now takes the form

$$p(k, n) = \binom{n}{k} \pi^k (1 - \pi)^{n-k} F(k), \quad (31)$$

$$p(k, 2n) = \pi^k (1 - \pi)^{2n-k} \left[ \binom{2n}{k} - H(k) \right] = \pi^k (1 - \pi)^{2n-k} \widetilde{H}(k), \quad (32)$$

where

$$H(k) = \sum_{z=0 \vee (k-n)}^{k \wedge n} \binom{n}{z} \binom{n}{k-z} F(z), \quad (33)$$

the meaning of  $\widetilde{H}(k)$  is obvious, a  $\vee b = \max(a, b)$ , and a  $\wedge b = \min(a, b)$ .

When stopping rule (13) is chosen, (31) becomes:

$$p(k, n) = \binom{n}{k} \pi^k (1 - \pi)^{n-k} (p + q + 1) \binom{p+q}{p} \sum_{\ell=0}^q \frac{(-1)^\ell}{p + \ell + 1} \binom{q}{\ell} A(k)^{p+\ell+1}. \quad (34)$$

The marginal stopping probability can be derived by summing (34) over  $k$  but is generally unwieldy. In the particular case that  $p=q=0$  and  $A(k)=k/n$ , we find

$$p(k, n) = \frac{k}{n} \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad (35)$$

$$p(k, 2n) = \frac{2n-k}{2n} \binom{2n}{k} \pi^k (1 - \pi)^{2n-k}. \quad (36)$$

While the derivation of (35) is obvious, that of (36) is less straightforward and details are given in Appendix 3-5. From (35), we deduce immediately that

$$P(N = n) = \frac{1}{n} \sum_{k=0}^n k \binom{n}{k} \pi^k (1 - \pi)^{n-k} = \pi.$$

In other words, this particular choice of conditional stopping rule produces essentially the simplest possible marginal stopping probability that depends on the parameter  $\pi$  that governs the outcomes.

The condition for the existence of a non-trivial function  $g(K, N)$  with expectation zero for all  $\pi$  is a discrete version of (16) and reads:

$$\sum_{k=0}^n g(k, n) F(k) \binom{n}{k} \pi^k (1 - \pi)^{n-k} + \sum_{k=0}^{2n} g(k, 2n) \widetilde{H}(k) \pi^k (1 - \pi)^{2n-k} = 0. \quad (37)$$

Writing  $\gamma = \pi / (1 - \pi)$ , (37) becomes

$$\sum_{k=0}^n g(k, n) F(k) \binom{n}{k} \gamma^k + \sum_{k=0}^{2n} g(k, 2n) \widetilde{H}(k) (1 - \pi)^n \gamma^k = 0. \quad (38)$$

Using the discrete-data version of (6), i. e.,

$$(1 - \pi)^{-n} = \sum_{k=0}^n \binom{n}{k} \gamma^k,$$

it follows that

$$-\sum_{k=0}^{2n} \left[ \sum_{z=0 \vee (k-n)}^{k \wedge n} \binom{n}{z} \binom{n}{k-z} g(z, n) F(z) \right] \theta^k = \sum_{k=0}^{2n} g(k, 2n) \widetilde{H}(k) \theta^k.$$

Owing to equality of polynomial coefficients, we find:

$$g(k, 2n) = - \frac{\sum_{z=0 \vee (k-n)}^{k \wedge n} \binom{n}{z} \binom{n}{k-z} g(z, n) F(z)}{\widetilde{H}(k)},$$

the discrete-data version of (18). In other words, an example has been constructed that establishes incompleteness.

## Generalized Sample Averages

### The general case

To underscore the impact of incompleteness of the statistics  $(K, N)$ , Molenberghs et al. [5] generalized the sample average (3) to

$$\bar{\mu} = \frac{K}{N} \cdot [c \cdot I(N = n) + d \cdot I(N = 2n)] = K \cdot \left[ \frac{c \cdot I(N = n)}{n} + \frac{d \cdot I(N = 2n)}{2n} \right], \quad (39)$$

for some constants  $c$  and  $d$ . We will refer to it as the *generalized sample average* (GSA). The ordinary sample average follows as  $c=d=1$ . In this section, (39) will be considered from a general exponential-family perspective. Sections 4. 2 and 4. 3 bring out some further specifics for the normal and Bernoulli cases, respectively.

From (5), the mean follows as  $\mu = \partial \alpha(\theta) / \partial \theta$ . The expectation is:

$$E(\bar{\mu}) = \frac{c}{n} \int k f_n(k) F(k) dk + \frac{d}{2n} \int k f_{2n}(k) dk - \frac{d}{2n} \iint k f_n(z) f_n(k-z) F(z) dk dz. \quad (40)$$

This form can be simplified. We will derive two identities that are useful here and in what follows. Because integrating (14)-(15) over  $K$  and summing over  $N$  should lead to unity, it follows that

$$\int f_n(k) F(k) dk = \iint f_n(z) f_n(k-z) F(z) dk dz. \quad (41)$$

This equation obviously also follows from first principles. Likewise, we have that

$$\begin{aligned} \iint k f_n(z) f_n(k-z) F(z) dk dz &= \int z f_n(z) F(z) \left[ \int_k k f_n(k-z) dk \right] dz \\ &= \int z f_n(z) F(z) [n\mu + z] dz \\ &= n\mu A_n(\mu) + B_n(\mu), \end{aligned} \quad (42)$$

where

$$A_n(\mu) = \int f_n(k) F(k) dk, \quad B_n(\mu) = \int k f_n(k) F(k) dk. \quad (43)$$

Using (42), we can rewrite (40) as

$$E(\bar{\mu}) = d \cdot \mu + \frac{2c-d}{2n} B_n(\mu) - \frac{d \cdot \mu}{2} A_n(\mu) \quad (44)$$

$$= d \cdot \mu + \frac{1}{2} \left[ (2c-d) E \left( \frac{K}{N} \mid N = n \right) - d \cdot \mu \right] \cdot P(N = n). \quad (45)$$

While obvious, it is useful to spell out (44)-(45) for the ordinary sample average:

$$E(\bar{\mu}) = \mu + \frac{1}{2n} B_n(\mu) - \frac{\mu}{2} A_n(\mu) \quad (46)$$

$$= \mu + \frac{1}{2} \left[ E \left( \frac{K}{N} \mid N = n \right) - \mu \right] \cdot P(N = n). \quad (47)$$

It is very intuitive that the bias in the sample average is a simple function of the difference between conditional and marginal expectation of  $K/N$  on the one hand, and the probability of stopping on the other.

The specific form of (40) will depend on both the exponential family member considered and the form of the stopping rule. In general, the expectation may be a non-linear function of  $\mu$  and hence there may be no constants  $c$  and  $d$  for which the expectation is  $\mu$ . Hence, in many situations, all linear estimators of the form (39) may be biased. Examples are given in Sections 4. 2 and 4. 3.

We now turn to the asymptotic behavior of the GSA, i. e., the case where  $n \rightarrow +\infty$ . Because the sample sum  $K$  converges to a  $N(n\mu, n\sigma^2)$  variable, and using a first-order Taylor series expansion  $F(k) \approx F(n\mu) + F'(n\mu)(k - n\mu)$ , we find from first principles:

$$A_n(\mu) \approx F(n\mu), \tag{48}$$

$$B_n(\mu) \approx n\mu F(n\mu) + n\sigma^2 F'(n\mu). \tag{49}$$

Using (48) and (49), (44) converges to:

$$E(\bar{\mu}) \xrightarrow{n \rightarrow +\infty} d \cdot \mu + (c-d) \cdot \mu \lim_{n \rightarrow +\infty} F(n\mu) + \frac{2c-d}{2} \sigma^2 \lim_{n \rightarrow +\infty} F'(n\mu). \tag{50}$$

In particular, for the ordinary sample average:

$$E(\bar{\mu}) \xrightarrow{n \rightarrow +\infty} \mu + \frac{1}{2} \sigma^2 \lim_{n \rightarrow +\infty} F'(n\mu). \tag{51}$$

In Section 4. 2, we will see that (50) is finite and, moreover, (51) equals  $\mu$ . Sufficient conditions for this to hold in general can be given. Assume that  $F(\cdot)$  is a continuously differentiable function that depends on  $k$  as a function of  $k/n$ . To emphasize this, write

$$F(k) = F(\eta(k/n)). \tag{52}$$

Then  $F(n\mu) = F(\eta(\mu))$ , independent of  $n$ , and  $F'(n\mu) = \eta'(\mu)F'(\eta(\mu))/n$ , which depends on  $n$ , only through the factor  $n$ .

$n^{-1}$  and hence converges to zero. More generally, a stopping rule that satisfies  $F'(n\mu) \xrightarrow{n \rightarrow +\infty} 0$  ensures that the sample average is asymptotically unbiased.

For a GSA to be asymptotically unbiased, (50) should equal  $\mu$ . Assume that the third term on the right hand side of (50) is zero and  $F(n\mu) = \tilde{F}(\mu)$  does not depend on  $n$ . The GSA is unbiased if  $d + (c-d)\tilde{F}(\mu) = 1$  for all values of  $\mu$  (note that, when  $\mu=0$ , the limit is trivially equal to zero). This equation can be satisfied if  $\tilde{F}(\mu)$  is constant, i. e., in the CRSS case to be discussed next. Otherwise, the equation can be satisfied only for  $c=d=1$ , i. e., the ordinary sample average.

For the GSA to be unbiased in the finite-sample case, (44) needs to equal  $\mu$ , leading to the requirement:

$$d = \frac{2\mu - 2c\tilde{\mu}P(N=n)}{2\mu - (\tilde{\mu} + \mu)P(N=n)}, \tag{53}$$

with  $\tilde{\mu} = E(K/N | N=n)$ . Evidently, this is a function of  $\mu$  in the non-CRSS case and hence no uniformly unbiased estimator exists. Further, unless in the CRSS case, the ordinary sample average never satisfies (53) because this would imply that  $\tilde{\mu} = \mu$  and hence the stopping probability would be independent of  $\mu$ .

In the specific case of a CRSS, the constant  $F$  is taken out of the integrals on the right hand side of (44) and we easily find:

$$E_{CRSS}(\bar{\mu}) = [cF + d(1-F)]\mu, \tag{54}$$

which is unbiased if and only if

$$d = \frac{1-cF}{1-F}. \tag{55}$$

An obvious solution is  $c=d=1$ , the sample average, next to an infinite number of unbiased linear estimators of the type (39). Note that (55) follows from (53) upon observing that in the CRSS case  $\mu = \tilde{\mu}$  and  $P(N=n)=F$ .

In addition to studying the overall expectation of the GSA, it is of interest to consider the conditional expectations. These are:

$$E(\bar{\mu} | N=n) = \frac{c}{n} \cdot \frac{B_n(\mu)}{A_n(\mu)}, \tag{56}$$

$$E(\bar{\mu} | N=2n) = \frac{d}{2n} \cdot \frac{2n\mu - n\mu A_n(\mu) - B_n(\mu)}{1 - A_n(\mu)}. \tag{57}$$

The ordinary sample average versions follow by setting  $c=d=1$  in (56)–(57).

The asymptotic behavior of (56)–(57), follows from applying (48) and (49):

$$E(\bar{\mu} | N=n) \xrightarrow{n \rightarrow +\infty} c \left( \mu + \sigma^2 \lim_{n \rightarrow +\infty} \frac{F'(n\mu)}{F(n\mu)} \right), \tag{58}$$

$$E(\bar{\mu} | N=2n) \xrightarrow{n \rightarrow +\infty} d \left( \mu - \frac{1}{2} \sigma^2 \lim_{n \rightarrow +\infty} \frac{F'(n\mu)}{1 - F(n\mu)} \right). \tag{59}$$

For the ordinary sample average, when  $F'(n\mu)$  converges to zero, the conditional expectations converge to  $\mu$ . In case the limits in (58) and (59) differ from zero, there is a choice for  $c$  and  $d$  that produces conditional expectations equal to  $\mu$ :  $c = \mu / [\mu + \sigma^2 Q_1(\mu)]$  and  $d = \mu / [\mu - 0.5\sigma^2 Q_2(\mu)]$ , with obvious notation. Evidently, these are not uniform and therefore not useful in practice. These values for  $c$  and  $d$  lie at different sides of unity. We will return to the implications of limiting expressions (51) and (58)–(59) in Section 4. 2.

A natural follow-up question is whether there is a, perhaps a uniform, optimal estimator in the CRSS case. From straightforward algebra we find that

$$\text{var}(\bar{\mu}) = \mu^2(1-c)^2 \left( \frac{F}{1-F} \right) + \frac{\sigma^2}{n} \frac{1-2Fc+2Fc^2-F^2c^2}{2(1-F)}, \tag{60}$$

which is minimal for

$$c_{opt} = \frac{2\mu^2n + \sigma^2}{2\mu^2n + \sigma^2(2-F)}, \quad d_{opt} = \frac{2\mu^2n + 2\sigma^2}{2\mu^2n + \sigma^2(2-F)}. \tag{61}$$

In (60) and (61),  $\sigma^2$  is the variance. It follows as either the first derivative of the mean function or, in the slightly more general case where there is an overdispersion parameter, as the first derivative of the mean multiplied with the overdispersion parameter.

Whereas constraint (55) on the pair (c, d) does not depend on the particular exponential family considered, rather only on the constant probability of stopping, this is not true for the optimality condition (61). Because of its dependence on  $\mu$  and  $\sigma^2$ , (61) will not generally allow for a uniform optimum, except in specific examples. A few examples are given in Table 1. As Molenberghs et al. [5] observed for the normal case, most solutions indeed indicate that there is no uniform minimum, even though all coefficients converge to 1 if the sample size increases. A noteworthy exception is the exponential family distribution, for which there is a uniform solution common to all values of the mean parameter and different from 1, for every value of the sample size  $n$  (Table 1).

In all cases, when  $F=0$  then  $d=1$  and  $c$  is irrelevant, while for  $F=1$ , the reverse is true.

| Exp. fam. member | C  | D   |
|------------------|--|---|
| Normal           | $\frac{2\mu^2n + \sigma^2}{2\mu^2n + \sigma^2(2-F)}$ | $\frac{2\mu^2n + 2\sigma^2}{2\mu^2n + \sigma^2(2-F)}$ |
| Bernoulli        | $\frac{2\pi n + (1-\pi)}{2\pi n + (1-\pi)(2-F)}$     | $\frac{2\pi n + 2(1-\pi)}{2\pi n + (1-\pi)(2-F)}$     |
| Poisson          | $\frac{2\lambda n + 1}{2\lambda n + 2 - F}$          | $\frac{2\lambda n + 2}{2\lambda n + 2 - F}$           |
| Exponential      | $\frac{2n + 1}{2n + 2 - F}$                          | $\frac{2n + 2}{2n + 2 - F}$                           |

**Table 1:** Coefficients for optimum unbiased generalized sample average estimators, in the case of a completely random sample size.

We have seen above that, even for CRSS, the sample average is not optimal, and that there is no uniform optimal solution, even though the sample average approximately is. The exponential case is an exception to this, as we saw above. However, the sample average is optimal in the restricted class of estimators that is invariant to future decisions. Indeed, if stopping occurs, then the choice of the coefficient  $c$  leads to an unbiased estimator, provided the appropriate  $d$  is chosen. However, this  $d$  will never be used as it pertains to ‘future’ observations. This can be avoided only by setting both coefficients to be equal, from which the conventional sample average emerges.

The asymptotic behavior for a deterministic stopping rule is completely captured by the normal case, described in Section 4. 2, because the stopping rule  $F(k)$  has the effect of restricting the integrals over the stopping and continuation regions S and C, respectively. This, together with the fact that  $f_n(k)$  approaches a normal density with mean  $n\mu$  and variance  $n\sigma^2$  establishes this fact. As a result, we can restrict considerations regarding the deterministic case to the finite-sample situation. But also this one is very straightforward. Given that the joint distribution (14)–(15) becomes (19)–(20), the functions  $A_n(\mu)$  and  $B_n(\mu)$  in (43) take the form:

$$A_n(\mu) = \int_S f_n(k) dk, \quad B_n(\mu) = \int_C k f_n(k) dk. \quad (62)$$

and all results, such as marginal and conditional expectations of the GSA, carry over.

### The normal case

Molenberghs et al. [5] showed that expectation (40) of generalized sample average (39) becomes, for the normal case with probit stopping probability:

$$E(\bar{\mu}) = d\mu + (c-d)\mu\Phi(v) + \frac{2c-d}{2n} \frac{\beta}{\sqrt{1+\beta^2/n}} \cdot \phi(v) \quad (63)$$

with  $v = (\alpha + \beta\mu) / \sqrt{1 + \beta^2/n}$ .

The specific case of a CRSS, here corresponding with  $\beta = 0$ , has been considered in Section 4. 1.

When  $\beta \neq 0$ , expression (63) does not in general simplify. It is easy to see here that there cannot be a uniformly unbiased estimator, i. e., that there cannot exist  $c$  and  $d$  such that (63) reduces to  $\mu$ , for all  $\mu$ , and in particular for  $\mu = 0$ . For this special case

$$0 = \frac{2c-d}{2n} \cdot \frac{\beta}{\sqrt{1+\beta^2/n}} \cdot \phi(v_0),$$

where  $v_0 = (\alpha) / \sqrt{1 + \beta^2/n}$ . Given that  $\beta \neq 0$ , this expression leads to the condition  $2c=d$ . Substituting this back into (63), which should be  $\mu$  for every value of  $\mu$ , and not just for  $\mu = 0$ , produces  $E(\bar{\mu}) = c\mu[2 - \Phi(v)]$ , which equals  $\mu$  only if  $c = [2 - \Phi(v)]^{-2}$ . Based on this, given that  $\Phi(v)$

is not constant but rather depends on  $\mu$ , unless  $\beta = 0$ , we see that there can be no uniformly unbiased estimator for the generalized sample average type. In other words, a simple average estimator, that merely uses the observed measurements in a least-squares fashion, can never be unbiased unless  $\beta = 0$ .

Molenberghs et al. [5] quantified the asymptotic bias. In Section 4. 1 this was done in general for CRSS. Turning to the case of  $\beta \neq 0$ , Molenberghs et al. [5] began with the ordinary sample average  $c=d=1$ , which leads to expectation:

$$E(\hat{\mu}) = \mu + \frac{1}{2n} \frac{\beta}{\sqrt{1+\beta^2/n}} \cdot \phi(v) \xrightarrow{n \rightarrow +\infty} \mu + \frac{1}{2n} \beta \cdot \phi(\alpha + \beta\mu) \xrightarrow{n \rightarrow +\infty} \mu. \quad (64)$$

In particular, when  $\beta \rightarrow +\infty$ , we see that

$$E(\hat{\mu}) = \mu + \frac{1}{2\sqrt{n}} \cdot \phi(\sqrt{n}\mu) \xrightarrow{n \rightarrow +\infty} \mu. \quad (65)$$

There exist other choices that also lead to asymptotically unbiased generalized sample averages. For  $\beta \neq 0$  but finite, the expectation becomes

$$E(\bar{\mu}) \xrightarrow{n \rightarrow +\infty} d\mu + (c-d)\mu\Phi(\alpha + \beta\mu), \quad (66)$$

which equals  $\mu$  if and only if:

$$d = \frac{1 - c\Phi(\alpha + \beta\mu)}{1 - \Phi(\alpha + \beta\mu)}. \quad (67)$$

While (67) and (55) are similar, there is a crucial difference between these: the latter is independent of  $\mu$ , while the former is not, except when  $c=d=1$ . In other words, there is no uniformly asymptotically unbiased generalized sample average for finite, non-zero  $\beta$ , except for the ordinary sample average itself.

The above limits also follow from (50) and (51), because now  $\eta(k/n) = \alpha + \beta k/n$  and the derivative therefore is  $F'(n\mu) = \phi(\alpha + \beta\mu) \cdot \beta/n$ , which leads to (64).

Molenberghs et al. [5] also studied the deterministic stopping rule case, following from  $\beta \rightarrow \infty$ , because then (66) becomes

$$E(\bar{\mu}) = d\mu + (c-d)\mu\Phi(\sqrt{n}\mu) + \frac{2c-d}{2\sqrt{n}} \phi(\sqrt{n}\mu) \xrightarrow{n \rightarrow +\infty} \begin{cases} c\mu & \text{if } \mu > 0, \\ d\mu & \text{if } \mu < 0, \\ 0 & \text{if } \mu = 0. \end{cases} \quad (68)$$

This provides us with the interesting situation that, for positive  $\mu$ ,  $c=1$  yields an asymptotically unbiased estimator, regardless of  $d$ , with the reverse holding for negative  $\mu$ . In the special case that  $\mu=0$ , both coefficients are immaterial. In addition, we see here as well that the only uniform solution is obtained by requiring that the bias asymptotically vanishes for all values of  $\mu$ , that is  $c=d=1$ .

The pleasing asymptotic behavior of the sample average is connected to the choice of the stopping rule, in view of limiting expressions (51), (58), and (59). In this case,  $\lim_{n \rightarrow +\infty} F(n\mu) = \Phi(\alpha + \beta\mu)$ , a constant in  $]0, 1[$ , while  $\lim_{n \rightarrow +\infty} F'(n\mu) = (\beta/n)\phi(\alpha + \beta\mu) = 0$ . Hence, the limits of  $F'(n\mu)$ ,  $F'(n)/F(n\mu)$ , and  $F'(n\mu)/[1 - F(n\mu)]$  are zero. The essence is that the stopping rule is a cumulative density function based transformation of a linear predictor in  $k/n$ . It is therefore of interest to examine the consequences of switching to a different class of stopping rule. Therefore, we change the stopping rule to  $\Phi(\alpha + \beta k)$ . Then  $F'(n\mu) = \beta\phi(\alpha + \beta n\mu)$  which again tends to zero. However, depending on the sign of  $\beta$  and  $\mu$ ,  $\Phi(\alpha + \beta n\mu)$  tends to either zero or one. Applying de l'Hôpital's rule to the case where  $F(n\mu)$  tends to zero as well, produces  $-\beta(\alpha + \beta n\mu)$  which tends to infinity, and hence the regularity condition (58) appears *not* to be satisfied. This requires careful qualification,

because not only does  $F(n\mu)$  appear in (58), it is also the probability with which  $N=n$ , which then equally well tends to zero. Thus, for this case, in the limit,  $E(\bar{\mu} | N=2n) = E(\bar{\mu})$  and unbiasedness still applies. Evidently, when  $1 - F(n\mu)$  tends to zero rather than  $F(n\mu)$ , we are in the mirror image of the above situation, and the result is the same. This result applies more generally. If  $F(k) = \Phi(\alpha + \beta kn^m)$ , with  $m$  any real number, then  $F(n\mu)$  converges to zero whatever  $m$  is. Further,  $F(n\mu)$  converges to  $\Phi(\alpha + \beta\mu)$  for  $m=-1$ ,  $\Phi(\alpha)$  for  $m<-1$ , and  $\Phi(\pm\infty)$  (i. e., 0 or 1) for  $m > -1$ . This means that the sample average is asymptotically unbiased in all cases, and even conditionally asymptotically unbiased, based on the same logic as before.

**The binary case**

An explicit form for the expectation of the generalized sample average in the Bernoulli case is

$$E(\bar{\pi}) = d\pi + \frac{c}{n} \sum_{k=0}^n k \binom{n}{k} \pi^k (1-\pi)^{n-k} F(k) - \frac{d}{2n} \sum_{k=0}^{2n} k \pi^k (1-\pi)^{2n-k} H(k), \tag{69}$$

with  $H(k)$  as in (33).

The CRSS has been covered in Section 4. 1, and the coefficients for optimal estimators listed in Table 1. As an example, when stopping rule (13) is chosen, with  $p=q=0$  and  $A(k)=k/n$ , we have that  $F(k)=A(k)=k/n$  and

$$H(k) = k / 2n \cdot \binom{2n}{k}.$$

Hence, (69) becomes

$$\begin{aligned} E(\bar{\pi}) &= d\pi + \frac{c}{n^2} E_n(K^2) - \frac{d}{(2n)^2} E_{2n}(K^2) \\ &= \pi \left[ d + \frac{c}{n} (1-\pi + n\pi) - \frac{d}{2n} (1-\pi + 2n\pi) \right]. \end{aligned} \tag{70}$$

Clearly, the estimator is unbiased if and only if

$$d = \frac{1 - \frac{c}{n} (1-\pi + n\pi)}{1 - \frac{1}{2n} (1-\pi + 2n\pi)}.$$

Hence, there is no uniform solution, neither in  $\pi$  nor in  $n$ . When  $n \rightarrow +\infty$ ,

$$d \xrightarrow{n \rightarrow +\infty} \frac{1-c\pi}{1-\pi}. \tag{71}$$

Note that the ordinary sample average, i. e.,  $c=d=1$ , is a solution to (71), as it should.

Turning to the case of a deterministic stopping rule, assume that the stopping region  $S$  is defined by  $(k \leq k_0)$ , i. e.,  $F(k)=1$  if  $k \leq k_0$  and 0 otherwise. Functions  $A_n(\pi)$  and  $B_n(\pi)$  as in (62) are here:

$$A_n(\pi) = \sum_{k=0}^{k_0} n \pi^k (1-\pi)^{n-k} = I(k_0, n, \pi), \tag{72}$$

$$B_n(\pi) = \sum_{k=0}^{k_0} kn \pi^k (1-\pi)^{n-k} = n\pi I(k_0 - 1, n - 1, \pi). \tag{73}$$

$I(k, n, \pi)$ , the binomial cumulative distribution function, is actually defined by (72). Various alternative formulations exist, but none is of direct use to us here. The expectation of the GSA becomes:

$$E(\bar{\pi}) = \pi \left[ d + \frac{2c-d}{2} I(k_0 - 1, n - 1, \pi) - \frac{d}{2} I(k_0, n, \pi) \right]. \tag{74}$$

For the ordinary sample average, (74) reduces to

$$E(\bar{\pi}) = \pi \left\{ 1 + \frac{1}{2} [I(k_0 - 1, n - 1, \pi) - I(k_0, n, \pi)] \right\}.$$

**Likelihood Estimators**

**The general case**

For notational convenience, we introduce the indicator variable  $Z=I(N=n)$ .

The joint likelihood for the observed data and stopping occurrence is:

$$L(\mu) = h_N(k) \exp\{\theta k - Na(\theta)\} \cdot F(k)^z \cdot [1 - F(k)]^{1-z}. \tag{75}$$

Likelihood decomposition (75) is of a selection model type. The factors pertaining to stopping are free of the mean parameter  $\mu$ . This simplifies the kernel of the log-likelihood  $l(\mu)$ , score function  $S(\mu)$ , and Hessian  $H(\mu)$ :

$$l(\mu) = \ln h_N(k) + \theta k - Na(\theta), \tag{76}$$

$$S(\mu) = k - Na'(\theta) = k - N\mu, \tag{77}$$

$$H(\mu) = -Na''(\theta) = -N\mu'. \tag{78}$$

The simplicity of this estimator is a direct consequence of ignorability. Based on (14)–(15), the conditional probability for the sample sum  $K$ , given the sample size  $N$ , can be derived. For the case that  $N=n$ , the likelihood function is:

$$L_n(\mu) = \frac{F(k)h_n(k)e^{\theta k}}{\int F(k)h_n(k)e^{\theta k} dk}, \tag{79}$$

leading to the following expressions for the log-likelihood, score, and Hessian:

$$l_n(\mu) = \ln F(k) + \ln h_n(k) + \theta k - \ln \int F(k)h_n(k)e^{\theta k} dk, \tag{80}$$

$$S_n(\mu) = k - \frac{B_n(\mu)}{A_n(\mu)} = k - E(K | N = n), \tag{81}$$

$$\begin{aligned} H_n(\mu) &= - \left[ \frac{C_n(\mu)}{A_n(\mu)} - \left\{ \frac{B_n(\mu)}{A_n(\mu)} \right\}^2 \right] \\ &= - [E(K^2 | N = n) - E(K | N = n)^2] = -var(K | N = n). \end{aligned} \tag{82}$$

Here  $A_n(\mu)$  and  $B_n(\mu)$  are as defined in (43), and

$$C_n(\mu) = \int k^2 f_n(k) F(k) dk.$$

When  $N=2n$ , the likelihood takes the form:

$$L_{2n}(\mu) = \frac{D_n(\mu)}{1 - A_n(\mu)}, \tag{83}$$

with

$$D_n(\mu) = \exp\{\theta k - 2na(\theta)\} \left[ h_{2n}(k) - \int h_n(z)h_n(k-z)F(z)dz \right].$$

Then, the counterparts to (80)–(82) are:

$$\begin{aligned} l_{2n}(\mu) &= \theta k - 2na(\theta) + \ln \left[ h_{2n}(k) - \int h_n(z)h_n(k-z)F(z)dz \right] \\ &\quad - \ln \{1 - A_n(\mu)\}, \end{aligned} \tag{84}$$

$$\begin{aligned} S_{2n}(\mu) &= k - \frac{2n\mu - n\mu A_n(\mu) - B_n(\mu)}{1 - A_n(\mu)} \\ &= k - E(K | N = 2n), \end{aligned} \tag{85}$$

$$H_{2n}(\mu) = - \frac{2n\sigma^2 - n\sigma^2 A_n(\mu) + (n\mu)^2 A_n(\mu) - C_n(\mu)}{1 - A_n(\mu)}$$



$$+ \frac{[B_n(\mu) - n\mu A_n(\mu)][2n\mu - n\mu A_n(\mu) - B_n(\mu)]}{[1 - A_n(\mu)]^2}$$

$$= -[E(K^2 | N = 2n) - E(K | N = 2n)^2] = -\text{var}(K | N = 2n). \quad (86)$$

From the form of (81) and (85), it is immediately clear that the conditional expectations of the conditional scores are equal to zero and therefore also the marginal expectation.

The expectation of the joint likelihood based estimator, which is the ordinary sample average, was presented in Section 4. 1. Even though there is small-sample bias in most cases different from CRSS, wide classes of stopping rules are asymptotically unbiased. The bias expressions in the conditional expectation of the sample average, which of course are also the bias expressions for the joint likelihood estimator, are of the form  $E(K/N|N) - \mu$ . These expressions coincide with the correction in conditional score equations (81) and (85) relative to (77), which follows immediately upon rewriting the former as  $S_n(\mu) = k - N\mu + \{N\mu - E(K | N)\}$ .

Turning to precision and information, first note that for CRSS,  $H_n(\mu) = -n\sigma^2$  and  $H_{2n}(\mu) = -2n\sigma^2$ ; hence the marginal and conditional information in this case reduces to  $I(\mu) = I_c(\mu) = n\sigma^2(2 - F)$ .

In the general case, the marginal and conditional information are

$$I(\mu) = n\sigma^2[2 - A_n(\mu)], \quad (87)$$

$$I_c(\mu) = n\sigma^2[2 - A_n(\mu)] - \frac{[n\mu A_n(\mu) - B_n(\mu)]^2}{A_n(\mu)[1 - A_n(\mu)]}. \quad (88)$$

Using information expressions (87)–(88), the bias for the marginal likelihood estimator, and the fact that the conditional likelihood estimator is unbiased, the mean squared error expressions are:

$$MSE_n(\hat{\mu}) = \frac{1}{n\sigma^2[2 - A_n(\mu)]} + \frac{1}{4n^2}[n\mu A_n(\mu) - B_n(\mu)]^2, \quad (89)$$

$$MSE_n(\hat{\mu}_c) = \frac{1}{n\sigma^2[2 - A_n(\mu)]} \quad (90)$$

$$+ \frac{[n\mu A_n(\mu) - B_n(\mu)]^2}{A_n(\mu)[1 - A_n(\mu)]\{n\sigma^2[2 - A_n(\mu)]^2 - n\sigma^2[2 - A_n(\mu)][n\mu A_n(\mu) - B_n(\mu)]^2\}}.$$

Recall that for CRSS  $B_n(\mu) = n\mu A_n(\mu)$  and both MSE expressions coincide. In the asymptotic case, (89)–(90) can be approximated, using (48)–(49), as:

$$MSE_{n \rightarrow \infty}(\hat{\mu}) \simeq \frac{1}{n\sigma^2[2 - F(n\mu)]} + \frac{\sigma^4}{4} F'(n\mu)^2, \quad (91)$$

$$MSE_{n \rightarrow \infty}(\hat{\mu}_c) \simeq \frac{1}{n\sigma^2[2 - F(n\mu)]}$$

$$\frac{F'(n)}{[2 - F(n\mu)]\{F(n\mu)[1 - F(n\mu)][2 - F(n\mu)] - n\sigma^2 F'(n\mu)^2\}} \quad (92)$$

Returning to the exact expressions (89)–(90), it is relatively straightforward to show that (89) is smaller than (90) if and only if  $\sigma^2 A_n(\mu)[1 - A_n(\mu)][2 - A_n(\mu)] \geq 4$ . Requiring that this inequality is satisfied for all values of  $A_n(\mu)$  in the unit interval comes down to requiring that  $\sigma^2 = 2.54$ . Hence, the MSE is smaller in the marginal case if the variance is sufficiently small. For binary data, for example this is always satisfied given that the variance takes the form  $\pi(1-\pi)$ . Also, asymptotically,  $A_n(\mu)$  typically tends to either 0 or 1, and the above requirement is then also satisfied. In case  $F'(n)$  tends to zero as  $n$  tends to infinity, both MSE expressions tend to the same limit.

## The normal case

Molenberghs et al. [5] studied this case in detail. Because of the relatively simple expressions for the normal density and the probit stopping rule (22), additional insight can be gained. We summarize their arguments in Appendix 2.

## The binary case

Joint-likelihood expressions for the binary case, in the probability parameter  $\pi$  are:

$$L(\pi) = \binom{N}{k} \pi^k (1-\pi)^{N-k} \cdot F(k)^\pi \cdot [1 - F(k)]^{1-\pi}, \quad (93)$$

$$\ell(\pi) \propto k \ln \pi + (N-k) \ln(1-\pi), \quad (94)$$

$$S(\pi) = \frac{k}{\pi} - \frac{N-k}{1-\pi}, \quad (95)$$

$$H(\pi) = \frac{-k + 2\pi k - \pi^2 N}{\pi^2(1-\pi)^2}. \quad (96)$$

The expected Hessian, for fixed sample size, is well known to be  $-N/[\pi(1-\pi)]$ . However, with our stopping rule  $F(k) = k/n$ , it can be shown to be

$$E[H(\pi)] = -\frac{(2-\pi)n}{\pi(1-\pi)}. \quad (97)$$

Likewise, given that the solution to  $S(\pi)$  is the sample average, the bias is

$$\text{Bias}(\hat{\pi}) = \frac{\pi(1-\pi)}{2n}, \quad (98)$$

which implies that the bias must be less than  $1/(8n)$ . We will return to this in what follows.

Turning to the conditional expressions, for  $N=n$ , (79)–(82) become:

$$L_n(\pi) = \frac{\binom{n}{k} \pi^k (1-\pi)^{n-k} F(k)}{A_n(\pi)}, \quad (99)$$

with

$$A_n(\pi) = \sum_{\ell=0}^n \binom{n}{\ell} \pi^\ell (1-\pi)^{n-\ell} F(\ell),$$

leading to:

$$\ell_n(\pi) \propto k \ln \pi + (n-k) \ln(1-\pi) - \ln A_n(\pi), \quad (100)$$

$$S_n(\pi) = \frac{k}{\pi} - \frac{n-k}{1-\pi} - \frac{n\pi A_n(\pi) - B_n(\mu)}{A_n(\pi)}. \quad (101)$$

For the case where  $N=2n$  we obtain:

$$L_{2n}(\pi) = \frac{\binom{2n}{k} \pi^k (1-\pi)^{2n-k} [1 - F(k)]}{1 - A_n(\pi)}, \quad (102)$$

$$\ell_{2n}(\pi) \propto k \ln \pi + (2n-k) \ln(1-\pi) - \ln[1 - A_n(\pi)], \quad (103)$$

$$S_{2n}(\pi) = \frac{k}{\pi} - \frac{2n-k}{1-\pi} + \frac{n\pi A_n(\pi) - B_n(\mu)}{1 - A_n(\pi)}. \quad (104)$$

The fact that  $E[S_n(\pi)|N] = 0$  follows from the derivations in Section 5. 1, as well as from first principles.

It is clear that the above expressions are slightly different than the general expressions (75)–(78), because  $\pi$  is not the natural parameter.

This does not prohibit further derivations but makes them cumbersome from an algebraic standpoint. Therefore, we switch to the logit form, i.e.,  $\alpha = \ln[\pi/(1 - \pi)]$  will be used. Furthermore, we restrict attention to the particular stopping rule used in previous sections,  $F(k) = k/n$ . Then, (93)-(96) become:

$$L(\alpha) = \binom{N}{k} \frac{e^{\alpha k}}{(1 + e^\alpha)^N} \cdot \left(\frac{k}{n}\right)^z \cdot \left(\frac{n-k}{n}\right)^{1-z}, \quad (105)$$

$$\ell(\alpha) \propto \alpha k - N \ln(1 + e^\alpha), \quad (106)$$

$$S(\alpha) = k - N\pi, \quad (107)$$

$$H(\alpha) = -N\pi(1 - \pi). \quad (108)$$

The use of  $\pi$  on the right hand sides of (107) and (108) rather than  $\alpha$  is for convenience only. The expected Hessian is straightforward to derive, given that  $E(N) = n(2 - \pi)$ :

$$E[H(\alpha)] = -n(2 - \pi)\pi(1 - \pi). \quad (109)$$

In fact, this calculation is considerably easier than the derivation of (97), even though they are equivalent. Indeed, (97) follows from (109) by applying the delta method. Because  $\pi = \text{expit}(\alpha)$ , the derivative is  $\Delta = \partial\alpha / \partial\pi = [\pi(1 - \pi)]^{-1}$ , and  $E[H(\pi)] = \Delta^2 \cdot E[H(\alpha)]$ , as it should.

The forms for (100)–(104), supplemented with the Hessians, are:

$$\ell_n(\alpha) \propto \alpha k - n \ln(1 + e^\alpha) - \alpha + \ln(1 + e^\alpha), \quad (110)$$

$$S_n(\alpha) = (k - 1) - (n - 1)\pi, \quad (111)$$

$$H_n(\alpha) = -(n - 1)\pi(1 - \pi), \quad (112)$$

$$\ell_{2n}(\alpha) \propto \alpha k - 2n \ln(1 + e^\alpha) + \ln(1 + e^\alpha), \quad (113)$$

$$S_{2n}(\alpha) = k - (2n - 1)\pi, \quad (114)$$

$$H_{2n}(\alpha) = -(2n - 1)\pi(1 - \pi). \quad (115)$$

Note that the conditional Hessians are in line with what one would expect from conditioning upon the sample size: one ‘degree of freedom’ is removed for mean parameter estimation. Such an operation though, is standard only when the sample size is fixed. The counterintuitive effect on the efficiency was seen in general in Section 5. 1 and very explicitly for the normal data setting in Section 5. 2. Straightforward algebra then establishes:

$$E[H_N(\alpha)] = -\pi(1 - \pi)[(2 - \pi)n - 1] = E[H(\alpha)] + \pi(1 - \pi). \quad (116)$$

Thus, the conditional information is expected to take one subject less into account than the marginal expectation, precisely the opposite of what one would expect in the fixed sample-size case. The bias in the estimators is easy to quantify, given that the estimators are  $\hat{\pi} = k / N$  in the marginal case and  $\hat{\pi}_c = (k - 1) / (n - 1)$  when  $N = n$  and  $\hat{\pi}_c = k / (2n - 1)$  when  $N = 2n$ . The biases are  $(n - k) / [(n - 1)]$  and  $-k / [2n(2n - 1)]$ , respectively. This follows from the difference between the marginal and conditional estimators, given that the latter is unbiased. For this stopping rule,  $E(K|N = n) = n\pi + 1 - \pi$  and  $E(K|N = 2n) = \pi(2n - 1)$ , and so the average bias is (98), as we expect.

The variances are equal to the negative inverses of the expected Hessians. These, combined with bias (98), readily leads to the MSE. Of course, (112) and (115) are for  $\hat{\alpha}_c$  and hence the delta method needs to be applied to obtain the variances for  $\hat{\pi}_c$ . Note that the variance for  $\hat{\pi}$  was already derived in (97), but applying the delta method to (108) gives the exact same result. The additional expressions are

$$\text{var}(\hat{\pi}_c | N = n) = \frac{\pi(1 - \pi)}{n - 1}, \quad (117)$$

$$\text{var}(\hat{\pi}_c | N = 2n) = \frac{\pi(1 - \pi)}{2n - 1}, \quad (118)$$

with the expected conditional Hessians the inverses of these quantities:

$$\text{var}(\hat{\pi}_c) = \text{var}(\hat{\pi}_c | N = n) \cdot \pi + \text{var}(\hat{\pi}_c | N = 2n) \cdot (1 - \pi) = \frac{(n\pi + n - 1)}{(n - 1)(2n - 1)} \pi(1 - \pi). \quad (119)$$

Note that the derivation of overall variance (119) involves the expectation of the conditional variances only, while the variance of the conditional expectations is zero, because both conditional estimators are unbiased. Finally,

$$\begin{aligned} \text{MSE}(\hat{\pi}) &= \text{var}(\hat{\pi}) + \text{Bias}(\hat{\pi})^2 = \frac{\pi(1 - \pi)}{n(2 - \pi)} + \frac{\pi^2(1 - \pi)^2}{(2n)^2} \\ &= \frac{1}{n} \cdot \frac{\pi(1 - \pi)}{2 - \pi} + O(n^{-2}), \end{aligned} \quad (120)$$

$$\begin{aligned} \text{MSE}(\hat{\pi}_c) &= \text{var}(\hat{\pi}_c) = \frac{(n\pi + n - 1)\pi(1 - \pi)}{(n - 1)(2n - 1)} \\ &= \frac{1}{n} \cdot \frac{\pi(1 - \pi)(\pi + 1)}{2} + O(n^{-2}). \end{aligned} \quad (121)$$

Calculating the difference between (121) and (120), we find

$$\text{MSE}(\hat{\pi}_c) - \text{MSE}(\hat{\pi}) = \frac{1}{n} \frac{\pi^2(1 - \pi)^2}{2(2 - \pi)} + O(n^{-3}).$$

Hence, like in the normal case, the joint estimator is more efficient than the marginal one. Of course, the MSE increase when moving from the joint to the conditional estimator is modest, with  $\text{MSE}(\hat{\pi}) \leq \text{MSE}(\hat{\pi}_c) \leq 1.125 \text{MSE}(\hat{\pi})$ , the maximum discrepancy reached for  $\pi = 0.5$ , and equality for  $\pi = 0$  or  $\pi = 1$ .

## Concluding Remarks

We have considered the consequences for statistical inference of a random sample size. Our setting is that of univariate random variables from the exponential family that are subject to a stopping rule such that the sample size is either  $N = n$  or  $N = 2n$ , with  $n$  specified by design. The stopping rule is stochastic and is allowed to depend on the sample sum  $\kappa$  over the first  $n$  observations. The rule is generic in the sense that its limiting cases are a deterministic stopping rule, such as in a sequential trial, and a completely random sample size, independent of the data. This setting extends those of both Liu et al. and Molenberghs et al. [5]; the former restrict attention to a deterministic stopping rule, although they do so for an arbitrary number of interim looks. The latter confined attention to normally distributed outcomes only.

We have focused on three important inferential aspects. First, we have shown that the sufficient statistic  $(K, N)$  is incomplete. Second, we have examined the consequences of this for the sample average, as well as for linear generalizations thereof. We have shown that there is small-sample bias, except for the CRSS case. Even then, there is no optimal estimator, except for the exponential distribution, for which the optimum differs from the ordinary sample average. Third, we have studied maximum likelihood estimation in both a joint as well as a conditional framework. The joint likelihood is for the exponential-family parameter and the stopping rule simultaneously. The conditional likelihood starts from the conditional distribution of the outcomes, given the sample size. Also here, counterintuitive results are derived. The joint likelihood produces the sample average as maximum likelihood estimator, which is biased in finite samples but is asymptotically unbiased, provided a regularity condition on the

stopping rule applies. The conditional likelihood estimator is unbiased, even in small samples. This notwithstanding, the sample average has smaller MSE than the conditional estimator in many important cases, such as the normal and binary examples considered, as well as when the variance of the outcomes is sufficiently small. Under regularity conditions, both estimators are asymptotically equivalent, with the difference between both being  $O(n^{-1})$ . The regularity condition is not very restrictive; it essentially comes down to requiring that  $F'(k=n\mu)$  approaches zero where  $F$  is the stopping rule. For broad classes of parametric functions, this condition is satisfied. We have shown that the corresponding conditional expectations are unbiased.

Hence, when the regularity conditions are satisfied, the sample average remains an attractive and sensible choice for sequential trials. Thus, while some familiar inferential properties no longer hold, estimation after sequential trials will often be more straightforward than commonly considered. In other words, often, there will be no need for modified estimators as have been proposed in the literature [2,11] or for our conditional estimator. Of course, if finite-sample unbiasedness is of overriding importance, such estimators may be preferred.

Note that there are several situations possible where sample size is or appears to be random, yet distinct from our setting. One example is when studies are stopped for futility reasons and estimation takes place only when the trial runs its full course [12-15].

Molenberghs et al. [5] considered several ramifications of their developments. They commented on the situation of an arbitrary number of looks in a sequential trial, and considered in detail the CRSS case for more than two possible sample sizes. All of this was done for normally distributed outcomes. They also commented on the connection between their derivations and longitudinal outcomes subject to dropout of an MAR type, where dropout depends on observed but not further on unobserved outcomes. While similar, there are subtle differences because now the randomness in the sample size pertains to the number of measurements per subject, rather than to the number of subjects. The difference lies in the fact that measurements within a subject are not independent. Our results extend to these settings as well for the exponential family. Furthermore, connections can be made with a variety of other settings with random sample sizes, such as clustered data with informative cluster sizes, time-to-event data subject to censoring, jointly observed longitudinal and time-to-event data, and random observation times. These settings are currently scrutinized further, and will be reported in a separate manuscript. It would also be of interest to extend our results to the case of semi-parametric models, e.g., generalized estimating equations [Liang and Zeger(1986)]. This is outside of the scope of the current manuscript [16-18].

In our illustrations, we have focused on the important, but still particular, case of binary outcomes, juxtaposed to the normal outcomes case of Molenberghs et al. [5]. Evidently, other data types, such as counts, time-to-event data, and ordinal outcomes may be handled in the same way, as long as exponential families are used. Technically, the ordinal case is a little more involved, because ordinal outcomes are frequently modeled using a sequence of dummy variables, thus requiring a multivariate version of our developments. Fortunately, this poses no insurmountable problems.

## Acknowledgments

Geert Molenberghs, Mike Kenward, Marc Aerts, and Geert Verbeke gratefully acknowledge support from IAP research Network P6/03 of the Belgian Government (Belgian Science Policy). The work of Anastasios Tsiatis and Marie Davidian was supported in part by NIH grants P01 CA142538, R37 AI031789, R01 CA051962, and R01 CA085848.

## References

1. Hughes MD, Pocock SJ (1988) Stopping rules and estimation problems in clinical trials. *Stat Med* 7: 1231-1242.
2. Emerson SS, Fleming TR (1990) Parameter estimation following group sequential hypothesis testing. *Biometrika* 77: 875-892.
3. Little RJA, Rubin DB (2002) *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.
4. Liu A, Hall WJ (1999) Unbiased estimation following a group sequential test. *Biometrika* 86: 71-78.
5. Molenberghs G, Kenward MG, Aerts M, Verbeke G, Tsiatis AA, et al. (2012) On random sample size, ignorability, ancillarity, completeness, separability, and degeneracy: sequential trials, random sample sizes, and missing data. *Stat Methods Med Res* 23: 11-41.
6. Cox DR, Hinkley DV (1974) *Theoretical Statistics*. Chapman & Hall, London.
7. Casella G, Berger RL (2001) *Statistical Inference*. Duxbury Press, Pacific Grove.
8. Basu D (1955) On statistics independent of a complete sufficient statistic. *Sankhya* 15: 377-380.
9. Jennison C, Turnbull BW (2000) *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, London.
10. Lehmann EL, Stein C (1950) Completeness in the sequential case. *Annals of Mathematical Statistics* 21: 376-385.
11. Jung SH, Kim KM (2004) On the estimation of the binomial probability in multistage clinical trials. *Stat Med* 23: 881-896.
12. Cohen A, Sackrowitz HB (1989) Two stage conditionally unbiased estimators of the selected mean. *Statistics & Probability Letters* 8: 273-278.
13. Armitage P (1975) *Sequential Medical Trials*. Blackwell, Oxford.
14. Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47: 663-685.
15. Kenward MG, Molenberghs G (1998) Likelihood based frequentist inference when data are missing at random. *Statistical Science* 13: 236-247.
16. Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73: 13-22.
17. Little RJA (1993) Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88: 125-134.
18. Little RJA (1994) A class of pattern-mixture models for normal incomplete data. *Biometrika* 81: 471-483.