Ann Inst Stat Math (2014) 66:759–787
DOI 10.1007/s10463-013-0424-y

Some binary start-up demonstration tests and associated inferential methods

N. Balakrishnan · M. V. Koutras · F. S. Milienos

Received: 14 December 2012 / Revised: 25 June 2013 / Published online: 30 August 2013

Abstract During the past few decades, substantial research has been carried out on start-up demonstration tests. In this paper, we study the class of binary start-up demonstration tests under a general framework. Assuming that the outcomes of the start-up tests are described by a sequence of exchangeable random variables, we develop a general form for the exact waiting time distribution associated with the length of the test (i.e., number of start-ups required to decide on the acceptance or rejection of the equipment/unit under inspection). Approximations for the tail probabilities of this distribution are also proposed. Moreover, assuming that the probability of a successful start-up follows a beta distribution, we discuss several estimation methods for the parameters of the beta distribution, when several types of observed data have been collected from a series of start-up tests. Finally, the performance of these estimation methods and the accuracy of the suggested approximations for the tail probabilities are illustrated through numerical experimentation.

Keywords Beta binomial distribution · EM algorithm · Exchangeable random variables · Maximum likelihood estimation · Start-up demonstration test

N. Balakrishnan
Department of Mathematics and Statistics, McMaster University,
Hamilton, ON L8S 4K1, Canada
e-mail: bala@mcmaster.ca

M. V. Koutras
Department of Statistics and Insurance Science, University of Piraeus,
80 Karaoli & Dimitriou St., 185 34 Piraeus, Greece
e-mail: mkoutras@unipi.gr

F. S. Milienos (✉)
Department of Mathematics and Statistics, University of Cyprus, P.O. Box: 20537, 1678 Nicosia, Cyprus
e-mail: milienos@yahoo.com

1 Introduction

A start-up demonstration test is a mechanism by which the quality or reliability of an equipment/unit (such as batteries, lawn mowers, fire alarm systems, or power generators) is evaluated by means of successful start-ups.

Hahn and Gage (1983) studied a start-up demonstration test in which an unit under test is accepted if a number of consecutive successful start-ups is observed (CS model); the number of failures in this testing procedure is ignored. Assuming that the outcomes of the start-ups are independent and identically distributed (i.i.d.) binary random variables, a recurrence formula is derived for the probability of the waiting time until the required number of consecutive successes are observed. Viveros and Balakrishnan (1993) also studied the CS model by assuming that the outcomes of the individual start-ups are Markov-dependent binary random variables (instead of i.i.d. binary trials); the moments of the waiting time until termination and some inferential methods (based on the method of moments) for the success probability have been developed. Balakrishnan et al. (1995) derived the joint probability generating functions for various statistics associated with start-up demonstration tests, considering Markov-dependent binary outcomes. In the same article, the principle of corrective actions has been introduced in the case of independent outcomes. Subsequently, Balakrishnan et al. (1997) investigated the use of corrective actions under the setting of Markov dependence. It is worth mentioning that the corresponding waiting time problems have also been studied under the term *geometric distribution of order k* (see, e.g. Feller 1968; Philippou and Muwafi 1982; Aki et al. 1984, or the monographs by Balakrishnan and Koutras 2002 and Johnson et al. 1992). Gera (2004) also ignored the number of failures in his testing procedure and suggested accepting the unit if a number of consecutive successful start-ups or a total number of successes is met (CSTS model); his work is restricted to i.i.d. binary outcomes and a form of one-step dependence.

Balakrishnan and Chan (2000) introduced a new start-up demonstration test in which an unit is accepted if a number of consecutive successful start-ups is observed before a certain number of failures; otherwise, the unit is rejected (CSTF model). Assuming i.i.d. binary trials, they then derived the distribution of the total number of trials until termination (i.e., test length) and its conditional distributions. The underlying model in the CSTF procedure had been studied earlier in the context of multiple run sampling plans, by Vance and McDonald (1979) and Govindaraju and Lai (1999). Moreover, Smith and Griffith (2005) and Martin (2004, 2008) studied the CSTF model, by assuming Markov dependent start-ups. Chan et al. (2008) addressed the problem of estimation of the success probability (in the i.i.d. case) by using maximum likelihood estimation via the expectation-maximization (EM) algorithm and also via Bayesian estimation with a beta prior; they focused on the case when the observed data contain only the number of trials until termination. Scollnik (2010) re-examined (corrected) the Bayesian analysis presented by Chan et al. (2008) and used the Markov Chain Monte Carlo (MCMC) method for carrying out the Bayesian analysis of CSTF model. Next, Scollnik (2011) assumed the observed data to be of different type than the one considered by Chan et al. (2008) and Scollnik (2010) and carried out Bayesian estimation through the MCMC method. Recently, Eryilmaz and Chakraborti (2008) studied the CSTF model by assuming that the probability of a successful start-up is a random variable. They observed that in this case, the individual start-ups form a sequence of exchangeable binary random variables and studied the distribution of the test length, the probability of acceptance and some other stochastic characteristics of the model.

Smith and Griffith (2008) introduced three alternative start-up demonstration tests: the first one is based on total successes and total failures (TSTF model), the second on consecutive successes and consecutive failures (CSCF model), and the third takes into account the total successes and consecutive failures (TSCF model). By the use of the CSCF model, for example, the unit will be accepted if a number of consecutive successful start-ups is observed before a number of consecutive failed start-ups; otherwise, the unit will be rejected. These tests are studied in the independent and identically distributed case, by exploiting the Markov chain embedding technique (non-i.i.d. cases could be investigated by this technique, as well). Martin (2008) also studied the above three models and the CSTF model, among others, assuming Markov-dependent start-ups of a general order. Eryilmaz (2010) investigated the CSTF and TSTF models, by assuming Markov-dependent start-ups; Yalcin and Eryilmaz (2012) also studied the TSTF model, assuming that the outcome of a binary trial depends on the total number of successful start-ups that have been observed until then.

Koutras and Balakrishnan (1999) suggested using scan statistics in start-up demonstration tests by regarding the high concentration of failures as an important aspect of the test procedure. Specifically, an unit may be rejected in the first n trials in two ways: i) by observing either one failure in the first $r-1$ start-ups and one failure in the subsequent trials, or ii) by observing two failures among k start-ups and no failure in the first $r-1$ start-ups. If neither i) nor ii) occurs in the first M trials, then the unit is accepted (where n, r, k, M are properly defined positive integers). Martin (2008) further studied this model.

Antzoulakos et al. (2009) also used scan statistics criteria to propose considered a procedure which accepts the unit under test if k consecutive successes occur before observing two failures, in a moving window of length $d-2$; if two failures with at most $d-2$ distance are observed before k consecutive successes, then the unit should be rejected.

Recently, Gera (2010, 2011) studied two additional models, namely the TSCSTF model and the TSCSTFCF model. Specifically, the TSCSTF model requires either a total number of successful tests, or a specified number of consecutive successes of tests to be observed, before a total number of failures, for the acceptance of an unit under test; otherwise, the unit will be rejected. Similarly, the TSCSTFCF model requires either a total number of successful tests, or a specified number of consecutive successes of tests to be observed, before a total number of failures and a specified number of consecutive failures. Some probabilistic aspects of these models (such as the distribution of the test length, the probability of acceptance, and the estimation of the success probability) have been examined by assuming i.i.d. binary outcomes.

Apart from the binary start-up demonstration tests mentioned above, there also exist a family of multistate start-up demonstration tests (see Smith and Griffith 2011). It should also be mentioned that the study of start-up demonstration tests could be seen as a special case of the more general framework of competing patterns, studied by Aston and Martin (2005).

The rest of this paper is organized as follows: in Sect. 2, we present some basic definitions and notation to be used throughout. In Sect. 3, a general form of the exact waiting time distributions of any binary start-up demonstration test is developed, under the assumption that the outcomes form an infinitely exchangeable sequence. In Sect. 4, some approximations are developed for the waiting time distribution. In Sect. 5, several probabilistic (computation of mean and variance) and statistical (parameter estimation) results are presented for the case when the probability of success follows a beta distribution. Finally in Sect. 6, some numerical results are presented, and several comments are made about the methods of inference developed in Sects. 4 and 5.

2 Definitions and notation

Suppose the outcome of the ith start-up is described by a binary random variable X_i, $i = 1, 2, \ldots$, where

$$X_i = \begin{cases} 1, & \text{if the outcome of the } i\text{th start-up is a success} \\ 0, & \text{if the outcome of the } i\text{th start-up is a failure.} \end{cases}$$

After the first n trials, the test procedure will be terminated if the unit is either accepted or rejected; otherwise, the trials will be continued. Our study is primarily focused on the number of trials until termination (length of test), denoted by X. Hence, the number of trials until termination would be greater than n (i.e., $X > n$) if and only if we could neither reject nor accept the unit after the first n trials.

We assume here that the binary random variables X_i, $i = 1, 2, \ldots$, form a sequence of exchangeable random variables. One of the most important theorems in Bayesian statistics concerning exchangeable random variables, the celebrated de Finetti's Representation Theorem (see Jackman 2009), is going to play a crucial role in the proof of our results. But, before stating this theorem, let us introduce the notion of *exchangeable* random variables. We restrict ourselves to the theory of binary exchangeable random variables although similar properties hold true for more general cases (the continuous case, etc.).

Definition 1 The n binary random variables $X_1, X_2, \ldots, X_n$ are finitely exchangeable if their joint probability mass function $p(x_1, x_2, \ldots, x_n)$ is such that

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})$$

for all permutations π of indices $\{1, 2, \ldots, n\}$ and for every $(x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$. Moreover, an infinite sequence of random variables $X_1, X_2, \ldots$ is infinitely exchangeable if every finite subsequence is finitely exchangeable.

As can be readily seen from the above definition, exchangeable random variables possess some useful properties. For example, exchangeable random variables are identically distributed; on the other hand, independent and identically distributed random variables are exchangeable.

The classical binary version of de Finetti's Representation Theorem (de Finetti 1931) is then as follows:

Theorem 1 *If $X_1, X_2, \ldots$ is an infinitely exchangeable sequence of binary random variables with $P(X_i = 1) = p$ for every $i = 1, 2, \ldots$, then*

$$p(x_1, x_2, \ldots, x_n) = \int_0^1 \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \mathrm{d}F(p) \tag{1}$$

for any $x_i \in \{0, 1\}, i = 1, 2, \ldots$ and $n \in \{1, 2, \ldots\}$, where $F(p)$ is the limiting distribution of p, i.e.,

$$F(p) = \lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \cdots + X_n}{n} \leq p\right). \tag{2}$$

One of the most important conclusions drawn from de Finetti's Representation Theorem is that infinitely exchangeable random variables are conditionally independent and identically distributed random variables (specifically, conditionally i.i.d. given the value of p). Hence, the concept of exchangeability is equivalent to that of conditional independence with common distribution function. In addition, Theorem 1 implies that the probability $p(x_1, x_2, \ldots, x_n)$ is constant for all vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ with $\sum_{i=1}^{n} x_i = k$. From a statistical viewpoint, we only have to assign a distribution to the random variable p for calculating the joint probability mass function $p(x_1, x_2, \ldots, x_n)$ for any finite subset of size n of an infinitely exchangeable sequence.

Let us now denote by $P = \{P_j : j = 1, 2, \ldots, M_n\}$ a family of sets with the following properties:

- $P_j \subseteq \{1, 2, \ldots, n\}$ for all $j \in \{1, 2, \ldots, M_n\}$;
- if $X_i = 1$ for every $i \in P_j$, for at least one $j \in \{1, 2, \ldots, M_n\}$, then the unit will be accepted at the nth trial or before;
- there is no subset of P_j with the previous property (for all $j \in \{1, 2, \ldots, M_n\}$).

Let us also introduce the family $C = \{C_j : j = 1, 2, \ldots, N_n\}$ which possesses the following properties similar to those of family's P properties:

- $C_j \subseteq \{1, 2, \ldots, n\}$ for all $j \in \{1, 2, \ldots, N_n\}$;
- if $X_i = 0$ for every $i \in C_j$, for at least one $j \in \{1, 2, \ldots, N_n\}$, then the unit will be rejected at the nth trial or before;
- there is no subset of C_j with the previous property (for all $j \in \{1, 2, \ldots, N_n\}$).

The nature of the above families is analogous to the nature and role of the minimal path and cut sets in the statistical reliability theory (see Barlow and Proschan 1981). Before explaining the advantages of these families in a start-up demonstration test, it is useful to illustrate how can one construct them by means of an example. Note also that, although the cardinalities of the families P and C (i.e., M_n and N_n, respectively) depend on the value of n, we shall suppress the subscript n from M_n and N_n, throughout for simplicity.

Example 1 Let us consider first the CS start-up demonstration test. In this case, an unit will be accepted if and only if a number of consecutive successes, say m, will be met in the sequence $X_1, X_2, \ldots, X_n$. The members of the family $P = \{P_j : j = 1, 2, \ldots, M\}$ are as follows:

$$P_j = \{j, j+1, \ldots, j+m-1\}, \quad j = 1, 2, \ldots, M,$$

where $M = n-m+1$ while the family C is empty. In the more general CSTF start-up demonstration test, an unit will be accepted if a number of consecutive successful start-ups, say m_1, is observed before a certain total number of failures, say f_1; otherwise, the unit is rejected. In this case, we have

$$P_j = \{j, j+1, \ldots, j+m_1-1\}, j = 1, 2, \ldots, M$$

where $M = n - m_1 + 1$ and

$$C_j = \{j_1, j_2, \ldots, j_{f_1}\} \subseteq \{1, 2, \ldots, n\},$$

i.e., $|C_j| = f_1$ (note that we denote by $|A|$ the cardinality of set A) and $N = \binom{n}{f_1}$. In the TSCS start-up demonstration test, an unit will be accepted if and only if a number of consecutive successes, say m_1, or a total number (not necessarily consecutive) of successes, say m_2, will be met in the sequence $X_1, X_2, \ldots, X_n$. If $m_1 \geq m_2$, the family P consist of the sets

$$P_j = \{j_1, j_2, \ldots, j_{m_2}\}, \text{ for all } \{j_1, j_2, \ldots, j_{m_2}\} \subseteq \{1, 2, \ldots, n\} \text{ with } |P_j| = m_2,$$

i.e., $M = \binom{n}{m_2}$. If $m_1 < m_2$, the family P contains the sets

$$P_j = \{j, j+1, \ldots, j+m_1-1\}, \quad j = 1, 2, \ldots, n-m_1+1,$$

and $P_j = \{j_1, j_2, \ldots, j_{m_2}\} \subseteq \{1, 2, \ldots, n\}$, with $|P_j| = m_2$ and at most $m_1 - 1$ consecutive successful trials. Hence,

$$M = n - m_1 + 1 + N(n, m_2, m_1 - 1),$$

where (see, e.g. Balakrishnan and Koutras 2002)

$$N(n, m_2, m_1 - 1) = \sum_{j=0}^{[m_2/m_1]} (-1)^j \binom{n - m_2 + 1}{j} \binom{n - j(m_1)}{n - m_2}, \quad m_1 - 1 \leq m_2 \leq n$$

denotes the number of ways in which m_2 $1's$ can be distributed in n distinct places with at most $m_1 - 1$ consecutive $1's$ (obviously, $C_j = \emptyset$ for every j).

One may easily realize that the family $P = \{P_j : j = 1, 2, \ldots, M_n\}$ corresponds to the set of minimal path sets of a reliability system with n components whose operation is equivalent to the acceptance of the tested unit at or before the nth start-up. Therefore, one may use standard approaches in reliability analysis (such as fault tree analysis and Boolean tables) to construct the family P. Likewise, the family $C = \{C_j : j = 1, 2, \ldots, N_n\}$ corresponds to the set of minimal cut sets of a reliability system with n components whose operation is equivalent to the rejection of the tested unit at or before the nth start-up. Reliability analysis techniques can also be used to identify the members of the family C.

According to the definition of the families P and C, it is evident that an unit will be accepted if the outcomes of the trials $i, i \in P_j$ (for at least one $j \in \{1, 2, \ldots, M\}$), are successful start-ups (i.e., $X_i = 1$ for $i \in P_j$) while no other set of trials $i, i \in C_j$ (for all $j \in \{1, 2, \ldots, N\}$), has resulted in failed start-ups. In the next section, we illustrate how one can introduce binary variables X_i to describe the outcome of a start-up, and then exploit the families of sets P and C, to construct appropriate monotone binary functions, whose expected values will assist in the computation of the waiting time distributions of interest.

3 Waiting time distributions of a start-up demonstration test

Based on the families P and C introduced in the preceding section, we now introduce the binary (0-1) functions $\phi_0(\boldsymbol{x}_n)$, $\phi_1(\boldsymbol{x}_n)$ and $\phi(\boldsymbol{x}_n)$ as follows:

$$\phi_0(\boldsymbol{x_n}) = \prod_{j=1}^{N} \left(1 - \prod_{i \in C_j} (1 - x_i) \right), \quad \phi_1(\boldsymbol{x_n}) = \prod_{j=1}^{M} \left(1 - \prod_{i \in P_j} x_i \right),$$

and

$$\phi(\boldsymbol{x}_n) = \phi_0(\boldsymbol{x}_n)\phi_1(\boldsymbol{x}_n),$$

where $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$. Obviously, $\phi(\boldsymbol{X}_n) = 1$ if and only if $X > n$ and so

$$P(X > n) = E(\phi(\boldsymbol{X}_n)).$$

The next theorem offers the general form of the tail probabilities of the test length for any binary start-up demonstration test defined on an infinitely exchangeable sequence.

Theorem 2 *Let the outcomes $X_i, i = 1, 2, \ldots$, of a start-up demonstration test form a sequence of exchangeable binary random variables. Then, the tail probabilities of X are given by*

$$P(X > n) = \sum_{k=0}^{n} c_{nk} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p), \tag{3}$$

where

$$c_{nk} = \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi(\boldsymbol{x}_n), \quad k = 0, 1, \ldots, n,$$

with $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$ *and* $F(p)$ *being given by* (2).

Proof It suffices to observe that

$$P(X > n) = E(\phi(\boldsymbol{X}_n)) = \sum_{\boldsymbol{x}_n \in \{0,1\}^n} \phi(\boldsymbol{x}_n) P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n),$$

and then use Theorem 1 to express the multiple sum in the form

$$\begin{aligned} E(\phi(\boldsymbol{X}_n)) &= \sum_{k=0}^{n} \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi(\boldsymbol{x}_n) P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\ &= \sum_{k=0}^{n} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p) \left(\sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi(\boldsymbol{x}_n) \right) \\ &= \sum_{k=0}^{n} c_{nk} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p) \end{aligned}$$

with $c_{nk} = \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi(\boldsymbol{x}_n)$, $k = 0, 1, \ldots, n$. □

It is of interest to note that c_{nk} is in fact enumerating the ways in which k successes could be allocated into n positions such that $\phi(\boldsymbol{X}_n) = 1$ (hence, $c_{nk} \in \left\{0, 1, \ldots, \binom{n}{k}\right\}$). To help understand the result more clearly (i.e., the nature of the number c_{nk}), we consider the following example:

Example 2 Let us consider a TS (total successes) model where an unit will be accepted in the first n trials if and only if a number of successful start-ups, say m, have been observed. Therefore, for $m > n$, the probability $P(X > n) = 1$, while for $m \leq n$ we have

$$c_{nk} = \begin{cases} \binom{n}{k}, & \text{if } k < m \\ 0, & \text{if } k \geq m \end{cases}$$

for $k = 0, 1, \ldots, n$. Consider next the CS start-up demonstration test, described in Example 1. Then, for $m > n$, we have $P(X > n) = 1$, while for $m \leq n$ we have

$$c_{nk} = \begin{cases} \binom{n}{k}, & \text{if } k < m \\ N(n, k, m-1), & \text{if } k \geq m. \end{cases}$$

For the TSCS start-up demonstration test (see Example 1), it can be easily verified that if $n \geq m_1 \geq m_2$, then

$$c_{nk} = \begin{cases} \binom{n}{k}, & \text{if } k < m_2 \\ 0, & \text{if } k \geq m_2 \end{cases}$$

while for $m_1 < m_2 \leq n$,

$$c_{nk} = \begin{cases} N(n, k, k_1 - 1), & \text{if } k < m_2 \\ 0, & \text{if } k \geq m_2. \end{cases}$$

Also, if $m_1 > n$ or $m_2 > n$, then the TSCS model degenerates into a TS or CS model, respectively (obviously, $P(X > n) = 1$ when $m_1, m_2 > n$).

In the CSTF start-up demonstration test, an unit is accepted if a number of consecutive successful start-ups, say m_1, is observed before a total number of failures, say f_1; otherwise, the unit is rejected. Letting $n \geq \max\{m_1, f_1\}$ (we get much simpler cases if $n < \max\{m_1, f_1\}$), we have

$$c_{nk} = \begin{cases} 0, & \text{if } k \leq n - f_1 \\ N(n, k, m_1 - 1), & \text{if } k > n - f_1. \end{cases}$$

Obviously,

$$P(X = n) = P(X > n - 1) - P(X > n),$$

and making use of Theorem 2 we may readily arrive at the expression

$$P(X = n) = \sum_{k=0}^{n} v_{nk} \int_0^1 p^k (1 - p)^{n-k} \mathrm{d}F(p),$$

where

$$v_{nk} = \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} (1 - \phi(\boldsymbol{x}_n))\phi(\boldsymbol{x}_{n-1}).$$

The following theorem offers a formula for the probability mass function of the acceptance waiting time, i.e., the number of trials until an item is accepted (denoted by X_{accept}), and the respective rejection waiting time (denoted by X_{reject}). The form of these formulae are similar to (3) and result immediately from it upon exploiting the expressions

$$P(X_{\text{accept}} = n) = E\left(\phi_0(\boldsymbol{x}_n)(1 - \phi_1(\boldsymbol{x}_n))\phi_1(\boldsymbol{x}_{n-1})\right)$$

and

$$P(X_{\text{reject}} = n) = E\left(\phi_1(\boldsymbol{x}_n)(1 - \phi_0(\boldsymbol{x}_n))\phi_0(\boldsymbol{x}_{n-1})\right).$$

Theorem 3 *Let the outcomes X_i, $i = 1, 2, \ldots$, of a start-up demonstration test form a sequence of exchangeable binary random variables. Then, the probability mass function of the acceptance waiting time and the rejection waiting time are given by*

$$P(X_{\text{accept}} = n) = \sum_{k=0}^{n} b_{nk} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p)$$

and

$$P(X_{\text{reject}} = n) = \sum_{k=0}^{n} d_{nk} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p),$$

where

$$\begin{aligned} b_{nk} &= \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi_0(\boldsymbol{x}_n)(1-\phi_1(\boldsymbol{x}_n))\phi_1(\boldsymbol{x}_{n-1}), k = 0, 1, \ldots, n, \\ d_{nk} &= \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi_1(\boldsymbol{x}_n)(1-\phi_0(\boldsymbol{x}_n))\phi_0(\boldsymbol{x}_{n-1}), k = 0, 1, \ldots, n, \end{aligned}$$

and $\boldsymbol{x}_{n-1}$ stands for a vector of size $n-1$ containing the first $n-1$ components of $\boldsymbol{x}_n$.

Thus, once we assign a distribution function to the random variable p, Theorems 2 and 3 will reduce the problem of the computation of the distribution function of interest, to a combinatorial problem. Indeed, the evaluation of the quantities c_{nk}, b_{nk} and d_{nk} can be carried out through proper combinatorial arguments. It is worth noting that the quantity b_{nk} denotes the number of vectors of the space $\{0, 1\}^n$ with k $1's$ in which no rejection criterion is met (meaning that $X_i = 1$ for at least one $i \in C_j$ for all $j \in \{1, 2, \ldots, N_n\}$) while an acceptance criterion is met only with the realization of the nth trial. A similar interpretation also holds for the quantities d_{nk}.

The above results are consistent with the relevant published work, concerning specific start-up demonstration test defined on an exchangeable sequence of outcomes (see, e.g. Eryilmaz and Chakraborti 2008).

Hence, the waiting time distributions studied in this section are all of the form

$$P(W = n) = \sum_{k=0}^{n} w_{nk} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p),$$

where

$$w_{nk} = \begin{cases} v_{nk}, & \text{if } W = X \\ d_{nk}, & \text{if } W = X_{\text{reject}} \\ b_{nk}, & \text{if } W = X_{\text{accept}}. \end{cases}$$

Moreover, denoting by S_n the total number of successes in the first n trials, i.e., $S_n = \sum_{i=1}^{n} X_i$, $n = 0, 1, \ldots$, we arrive at the following theorem which gives the conditional probabilities $P(S_n = s | W = n)$.

Theorem 4 *Let the outcomes X_i, $i = 1, 2, \ldots$, of a start-up demonstration test form a sequence of exchangeable binary random variables and W be any of the random variables X, X_{reject} or X_{accept}. Then,*

$$P(S_n = s|W = n) = \frac{w_{ns} \int_0^1 p^s (1-p)^{n-s} \mathrm{d}F(p)}{\sum_{k=0}^{n} w_{nk} \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p)}$$

for $s = 0, 1, \ldots, n$.

Since the proof is almost an immediate consequence of the nature of the parameters w_{nk} (i.e., of the parameters v_{nk} or d_{nk} or b_{nk}), it is omitted.

4 Approximations of the waiting time distributions

A key aspect in the computation of the tail probabilities $P(X > n)$ is the evaluation of the quantities c_{nk}. Recall that c_{nk} denotes the number of ways in which k successes could be distributed into n positions such that $\phi(\boldsymbol{X}_n) = 1$; hence, $c_{nk} \in \left\{0, 1, \ldots, \binom{n}{k}\right\}$ for every n, k.

In this section, we present some computationally tractable bounds for the probabilities $P(X > n)$. First, let us introduce the quantities c_{nk}^1 and c_{nk}^0 as follows:

$$c_{nk}^i = \sum_{\boldsymbol{x}_n \in \{0,1\}^n : \boldsymbol{x}_n \boldsymbol{x}_n' = k} \phi_i(\boldsymbol{x}_n), i = 0, 1,$$

for $k = 0, 1, \ldots, n$. The quantity c_{nk}^1 (c_{nk}^0) denotes the number of ways in which k successes could be distributed into n positions such that the unit is not accepted (not rejected) after the first n trials.

Example 3 For the CSTF model, with parameters m_1 and f_1 (see Example 2), and $n \geq m_1$, we have

$$c_{nk}^1 = \begin{cases} \dbinom{n}{k}, & \text{if } k < m_1 \\ N(n, k, m_1 - 1), & \text{if } k \geq m_1, \end{cases}$$

while if $m_1 > n$ then $c_{nk}^1 = \dbinom{n}{k}$ for every $k = 0, 1, \ldots, n$. On the other hand, if $n \geq f_1$, we have

$$c_{nk}^0 = \begin{cases} \dbinom{n}{k}, & \text{if } k > n - f_1 \\ 0, & \text{if } k \leq n - f_1, \end{cases}$$

and if $f_1 > n$ then $c_{nk}^0 = \binom{n}{k}$ for every $k = 0, 1, \ldots, n$. Consider now the TSCSTFCF model that requires m_1 consecutive successes or a total of m_2 successful tests to be observed before f_1 consecutive failures and a total of f_2 failures, in order to accept the unit. In this case, the values of c^1_{nk} coincide with the c_{nk} of the model TSCS (see Example 2), and similarly the values of c^0_{nk} are equal to

$$c^0_{nk} = \begin{cases} \binom{n}{k}, & \text{if } n-k < f_2 \\ 0, & \text{if } n-k \geq f_2 \end{cases}$$

for $n \geq f_1 \geq f_2$; if $f_1 < f_2 \leq n$, then

$$c^0_{nk} = \begin{cases} N(n, n-k, f_1-1), & \text{if } n-k < f_2 \\ 0, & \text{if } n-k \geq f_2 \end{cases}$$

(the cases $f_1 > n$ or $f_2 > n$ lead to much simpler and similar expressions).

Note that in the definition of the quantity c^1_{nk} (c^0_{nk}), the rejection (acceptance) criteria do not play any role, a fact that makes the computation of the above quantities an easier task than the computation of the quantities c_{nk}.

In the next Theorem, we derive an upper bound for the tail probabilities $P(X > n)$. Its proof makes use of the properties of associated random variables. We recall that the random variables X_i, $i = 1, 2, \ldots$, are said to be associated if for every pair of coordinatewise nondecreasing functions $f, g : \Re^n \to \Re$ and every n, the following inequality holds true:

$$\mathrm{cov}(f(X_1, X_2, \ldots, X_n), g(X_1, X_2, \ldots, X_n)) \geq 0.$$

Two well-known properties of associated random variables are (see e.g. Barlow and Proschan 1981):

- independent random variables are associated;
- coordinatewise nondecreasing functions of associated random variables are associated.

Theorem 5 *Let the outcomes X_i, $i = 1, 2, \ldots$, of a start-up demonstration test form a sequence of exchangeable binary random variables. Then, an upper bound for the tail probability of the length of the test is given by*

$$P(X > n) \leq \sum_{k=0}^{n} \sum_{l=0}^{n} c^0_{nk} c^1_{nl} \int_0^1 p^{k+l}(1-p)^{2n-k-l} \mathrm{d}F(p) = \mathrm{UB}.$$

Proof It is clear that, given p, the binary functions $\phi_0(\boldsymbol{X}_n)$ and $1-\phi_1(\boldsymbol{X}_n)$ are increasing functions of i.i.d. random variables. In view of the aforementioned properties, given p, $\phi_0(\boldsymbol{X}_n)$ and $1-\phi_1(\boldsymbol{X}_n)$ are associated random variables and therefore, we may state that

$$\begin{aligned} &\mathrm{cov}(\phi_0(\boldsymbol{X}_n), 1-\phi_1(\boldsymbol{X}_n)|p) \geq 0 \\ &\quad \Rightarrow E(\phi_0(\boldsymbol{X}_n)\phi_1(\boldsymbol{X}_n)|p) \leq E(\phi_0(\boldsymbol{X}_n)|p)E(\phi_1(\boldsymbol{X}_n)|p). \end{aligned} \tag{4}$$

Moreover,

$$\begin{aligned} E(\phi_0(\boldsymbol{X}_n)|p) &= \sum_{\boldsymbol{x}_n \in \{0,1\}^n} \phi_0(\boldsymbol{x}_n) P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|p) \\ &= \sum_{k=0}^{n} c_{nk}^0 p^k (1-p)^{n-k}, \\ E(\phi_1(\boldsymbol{X}_n)|p) &= \sum_{\boldsymbol{x}_n \in \{0,1\}^n} \phi_1(\boldsymbol{x}_n) P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n|p) \\ &= \sum_{k=0}^{n} c_{nk}^1 p^k (1-p)^{n-k}, \end{aligned}$$

and

$$E(\phi_0(\boldsymbol{X}_n)|p) E(\phi_1(\boldsymbol{X}_n)|p) = \sum_{k=0}^{n} \sum_{l=0}^{n} c_{nk}^0 c_{nl}^1 p^{k+l} (1-p)^{2n-k-l}.$$

Thus,

$$E\left[E(\phi_0(\boldsymbol{X}_n)|p) E(\phi_1(\boldsymbol{X}_n)|p)\right] = \sum_{k=0}^{n} \sum_{l=0}^{n} c_{nk}^0 c_{nl}^1 \int_0^1 p^{k+l} (1-p)^{2n-k-l} \mathrm{d}F(p),$$

and the proof gets completed by using the fact that

$$P(X > n) = E\left[E(\phi_0(\boldsymbol{X}_n)\phi_1(\boldsymbol{X}_n)|p)\right].$$

□

In a similar manner, we may derive the following lower bound:

$$\begin{aligned} P(X > n) \geq &\sum_{k=0}^{n} c_{nk}^1 \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p) \\ &+ \sum_{k=0}^{n} c_{nk}^0 \int_0^1 p^k (1-p)^{n-k} \mathrm{d}F(p) - 1 = \mathrm{LB} \end{aligned}$$

which is an immediate consequence of the inequality

$$1 - \phi_1(\boldsymbol{X}_n) \geq \phi_0(\boldsymbol{X}_n)(1 - \phi_1(\boldsymbol{X}_n)). \tag{5}$$

As mentioned earlier, the definition of the families P and C are analogous to those of the minimal path and minimal cut sets (as used in reliability theory), respectively. Exploiting these families, we may define two distinct binary reliability systems, whose working state are described via the binary functions $\phi_0(\boldsymbol{X}_n)$ and $1-\phi_1(\boldsymbol{X}_n)$, respectively. Let us denote by R_0 and R_1 the system's reliability functions, given the success probability p, that is (see Barlow and Proschan 1981)

$$R_0 = E\left[\phi_0(\boldsymbol{X}_n)|p\right] \text{ and } R_1 = E\left[1-\phi_1(\boldsymbol{X}_n)|p\right].$$

Then, in view of the inequalities (4) and (5), the following lower and upper bounds are readily established:

$$R_0 - R_1 \leq E\left[\phi_0(\boldsymbol{X}_n)\phi_1(\boldsymbol{X}_n)|p\right] \leq R_0(1-R_1).$$

It is worth noting that, should the computation of the quantities c^1_{nk} and c^0_{nk} be computationally intractable, the above bounds may be used in conjunction with any of the well-known reliability lower and upper bounds, for a system with i.i.d. components (such as Esary–Proschan and Min–Max; see Barlow and Proschan 1981), to construct a set of alternative (but less accurate) bounds. For example, the following pairs of bounds may be used in specific applications:

$$L(p) = p^n, U(p) = 1-(1-p)^n,$$

$$L(p) = p^{\min_j |P_j|}, U(p) = 1-(1-p)^{\min_j |C_j|},$$

$$L(p) = \prod_{j=1}^{N}(1-(1-p)^{|C_j|}), U(p) = 1-\prod_{j=1}^{M}(1-p^{|P_j|}).$$

5 Statistical inference from start-up demonstration tests when the success probability follows a beta distribution

Suppose m independent start-up procedures are carried out on m different units with the success probability p among the m procedures following a beta distribution. The probability mass function of the beta distribution with parameters $\alpha > 0$, $\beta > 0$, is

$$g(p;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}, \quad p \in [0,1],$$

where $B(\alpha,\beta)$ denotes the beta function

$$B(\alpha,\beta) = \int_0^1 p^{\alpha-1}(1-p)^{\beta-1}dp.$$

Under the assumption of the beta distribution, the probability in (1) becomes

$$p(x_1, x_2, \ldots, x_n) = f(\boldsymbol{x}_n;\alpha,\beta) = \frac{B(\alpha+s, \beta+n-s)}{B(\alpha,\beta)}, \tag{6}$$

where $s = \sum_{j=1}^{n} x_j$ and $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n) \in \{0, 1\}^n$. Consequently, the waiting time distributions studied here all have a probability mass function of the form

$$P(W = n; \alpha, \beta) = \sum_{k=0}^{n} w_{nk} \frac{B(\alpha + k, \beta + n - k)}{B(\alpha, \beta)}, n = 1, 2, \ldots, \tag{7}$$

where $W = X$ or X_{reject} or X_{accept} and $w_{nk} = v_{nk}$ or d_{nk} or b_{nk}, respectively.

Clearly, there is a strong connection between the distribution of W and the Beta Binomial Distribution (BBD; see, for example, Johnson et al. (1992)). To be more specific, suppose the conditional distribution of a random variable Y is a binomial distribution with

$$P(Y = y|p) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y = 0, 1, \ldots, n.$$

Suppose further that the success probability p is a random variable with beta distribution. Then, Y follows a BBD with

$$\begin{aligned} P(Y = y; \alpha, \beta) &= \int_0^1 P(Y = y; p) g(p; \alpha, \beta) dp \\ &= \binom{n}{y} \frac{B(\alpha + y, \beta + n - y)}{B(\alpha, \beta)}, \quad y = 0, 1, \ldots, n. \end{aligned}$$

It is straightforward to see that if p is a beta distributed random variable, then the finite sum of infinite exchangeable random variables follows a BBD; note also that the beta distribution is conjugate for p with respect to the binomial distribution (e.g. Jackman (2009)). The mean and variance of Y are then given by (see Johnson et al. 1992)

$$E(Y) = n \frac{\alpha}{\alpha + \beta} \text{ and } \operatorname{var}(Y) = \frac{n(n + \alpha + \beta)\alpha\beta}{(\alpha + \beta)^2 (1 + \alpha + \beta)}, \tag{8}$$

or (see Kleinman 1973)

$$E(Y) = n\mu \text{ and } \operatorname{Var}(Y) = n\mu(1-\mu)\rho(n-1) + n\mu(1-\mu),$$

with the re-parametrization $\mu = \frac{\alpha}{\alpha+\beta}$ and $\rho = \frac{1}{1+\alpha+\beta}$ (note that $\mu, \rho \in (0, 1)$).

In this section, we discuss the estimation of the parameters α and β of the beta distribution. Since this is also explicitly connected with the estimation for the beta binomial model, there is a rich literature for the corresponding estimation procedures; see, for example, Skellam (1948), Shenton (1950), Johnson et al. (1992), Chatfield and Goodhardt (1970), Griffiths (1973), Kleinman (1973), Williams (1975), Wilcox (1979), Lee and Sabavala (1987), Lee and Lio (1999) and Everson and Bradlow (2002).

One of the popular methods for the estimation of the parameters α and β is based on maximum likelihood (see Skellam 1948; Griffiths 1973; Kleinman 1973), or by using the simpler moment estimates (see Skellam 1948; Shenton 1950; Chatfield and Goodhardt 1970; Kleinman 1973), or the "mean and zero" method (see Chatfield and Goodhardt 1970), and to treat these estimates as the true values of α and β. It is worth noting that there are no closed-forms for the MLEs and also the regularity conditions have not yet been proved. Estimation related problems have also been discussed by Tamura and Young (1986), Tamura and Young (1987), Yamamoto and Yanagimoto (1992), and Tripathi et al. (1994), among others.

There is also the full Bayesian approach in which a prior distribution is attributed to the vector (α, β) and then the posterior distribution is taken into account (see Lee and Sabavala 1987; Lee and Lio 1999; Everson and Bradlow 2002) to develop Bayesian estimators such as the mean, the median, or the mode of the marginal posterior distributions.

In this section, we study the cases when the observed data, from m independent start-up testing procedures, may be:

Case A: The whole sequence of outcomes;
Case B: The length of test and the number of successes;
Case C: The length of test only.

Though there are some discussions on the estimation problem in a start-up demonstration test framework (see Viveros and Balakrishnan 1993; Chan et al. 2008; Smith and Griffith 2008; Eryilmaz 2010; Scollnik 2010, 2011), when the problem is concerning the estimation of α and β under the beta assumption, not much work exists. For example, Chan et al. (2008) and Scollnik (2010, 2011) developed inference on p but not on the parameters α and β.

For this reason, in this section, we focus on the ML estimation and the EM approach, when the observed data belong to one of the above three forms. For the first two cases (A and B), the available results from the theory of BBD are sufficient for our purposes (as can be seen in Sect. 5.1). Hence, the new results of our study are mainly given in Sect. 5.2, for Case C.

5.1 Inference for Cases A and B

Let us assume that m independent testing procedures are carried out and the vectors $\boldsymbol{x}_i$, $i = 1, 2, \ldots, m$, describe the corresponding sequences of outcomes, i.e.,

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{n_i i}) \in \{0, 1\}^{n_i}, \quad i = 1, 2, \ldots, m.$$

Taking into account the expressions offered in (8) and assuming that $n_i = n$ for every $i = 1, 2, \ldots, m$, it can be easily seen that the estimation by the method of moments leads to the estimates (see Kleinman 1973)

$$\hat{\alpha} = \bar{p}(1 - \gamma)/\gamma, \quad \hat{\beta} = \bar{q}(1 - \gamma)/\gamma,$$

where

$$\bar{p} = \frac{1}{m}\sum_{i=1}^{m}\hat{p}_i = 1 - \bar{q}, \quad \hat{p}_i = \frac{1}{n}\sum_{j=1}^{n}x_{ij},$$

$$\gamma = \frac{n}{n-1}\frac{S}{\bar{p}\bar{q}m} - \frac{1}{n-1}, \quad S = \sum_{i=1}^{m}(\hat{p}_i - \bar{p})^2.$$

It is worth mentioning that the method cannot be applied if the computed value of γ does not belong to the interval (0, 1). If the values of n_i, $i = 1, 2, \ldots, m$, are not equal, then a weighted procedure must be followed. In this case, we may use the formulae (Kleinman (1973))

$$\bar{p} = \frac{1}{w}\sum_{i=1}^{m} w_i \hat{p}_i \tag{9}$$

and

$$\gamma = \frac{S - \bar{p}\bar{q}\sum_{i=1}^{m}(w_i/n_i)(1-w_i/n_i)}{\bar{p}\bar{q}[\sum_{i=1}^{m} w_i(1-w_i/w) - \sum_{i=1}^{m}(w_i/n_i)(1-w_i/n_i)]}, \tag{10}$$

$$S = \sum_{i=1}^{m} w_i(\hat{p}_i - \bar{p})^2,$$

with $w = \sum_{i=1}^{m} w_i$. For the weights w_i appearing in (9) and (10), Kleinman (1973) suggested three different choices:

(a) $w_i = 1$, for every $i = 1, 2, \ldots, m$;
(b) $w_i = n_i$, for every $i = 1, 2, \ldots, m$;
(c) Consider the weights given either by (a) or (b), as initial weights. Use them in formula (10) to gain an initial estimate of γ, say $\bar{\gamma}$, and then recalculate w_i as

$$w_i = \frac{n_i}{1 + \bar{\gamma}(n_i - 1)}, \quad i = 1, 2, \ldots, m.$$

Finally, use the update w_i's to reevaluate γ by (10).

On the other hand, in view of (6), the likelihood function is of the form

$$L(\alpha, \beta) = \prod_{i=1}^{m} f(\boldsymbol{x}_i; \alpha, \beta) = \prod_{i=1}^{m}\left[\frac{B(\alpha + s_i, \beta + n_i - s_i)}{B(\alpha, \beta)}\right],$$

and the likelihood equations may be solved iteratively using, for example, the Newton–Raphson method (with the moment estimators as the initial starting values; see Wilcox 1979); other direct optimization methods such as the simulated annealing method and Nelder-Mead method could also be used for this purpose.

The method of moment estimators described earlier also works well for Case B. Besides, it can be easily seen that

$$\begin{aligned} f(n_i, s_i; \alpha, \beta) &= P(X = n_i, S_{n_i} = s_i; \alpha, \beta) \\ &= v_{n_i s_i} \frac{B(\alpha + s_i, \beta + n_i - s_i)}{B(\alpha, \beta)}, \quad s_i = 0, 1, \ldots, n_i, \end{aligned}$$

and so, for Case B, the likelihood function becomes

$$L(\alpha, \beta) = \prod_{i=1}^{m} f(n_i, s_i; \alpha, \beta) = \left(\prod_{i=1}^{m} v_{n_i s_i} \right) \prod_{i=1}^{m} \left(\frac{B(\alpha + s_i, \beta + n_i - s_i)}{B(\alpha, \beta)} \right).$$

The corresponding maximization problem can be handled in a manner similar to that of Case A described above.

5.2 Inference for Case C

5.2.1 MLE approach

Consider now the case when only the lengths of the tests from m independent start-up procedures are available. The likelihood function is in this case of the form

$$L(\alpha, \beta) = \prod_{i=1}^{m} \sum_{k=0}^{n_i} v_{n_i k} \frac{B(\alpha + k, \beta + n_i - k)}{B(\alpha, \beta)}, \tag{11}$$

and the log-likelihood function is

$$l(\alpha, \beta) = \ln L(\alpha, \beta) = -m \ln B(\alpha, \beta) + \sum_{i=1}^{m} \ln \sum_{k=0}^{n_i} v_{n_i k} B(\alpha + k, \beta + n_i - k).$$

After some computations (see Appendix A), we obtain the likelihood equations as

$$\frac{\partial l(\alpha, \beta)}{\partial \alpha} = -\sum_{i=1}^{m} \sum_{r=0}^{n_i - 1} \frac{1}{\alpha + \beta + r} + \sum_{i=1}^{m} \sum_{k=0}^{n_i} \sum_{r=0}^{k-1} \frac{\pi_{ik}(\alpha, \beta)}{\alpha + r} = 0 \tag{12}$$

and

$$\frac{\partial l(\alpha, \beta)}{\partial \beta} = -\sum_{i=1}^{m} \sum_{r=0}^{n_i - 1} \frac{1}{\alpha + \beta + r} + \sum_{i=1}^{m} \sum_{k=0}^{n_i} \sum_{r=0}^{n_i - k - 1} \frac{\pi_{ik}(\alpha, \beta)}{\beta + r} = 0, \tag{13}$$

where (see also Theorem 4)

$$\pi_{ik}(\alpha, \beta) = P(S_{n_i} = k; X = n_i) = \frac{v_{n_i k} B(\alpha + k, \beta + n_i - k)}{\sum_{t=0}^{n_i} v_{n_i t} B(\alpha + t, \beta + n_i - t)}, \quad 0 \le k \le n_i.$$

Moreover (see Appendix B), we have

$$\frac{\partial \pi_{ik}(\alpha,\beta)}{\partial \alpha} = \pi_{ik}(\alpha,\beta)\left[\sum_{j=0}^{k-1}\frac{1}{\alpha+j} - \sum_{t=0}^{n_i}\sum_{j=0}^{t-1}\frac{\pi_{it}(\alpha,\beta)}{\alpha+j}\right]$$

and

$$\frac{\partial \pi_{ik}(\alpha,\beta)}{\partial b} = \pi_{ik}(\alpha,\beta)\left[\sum_{j=0}^{n_i-k-1}\frac{1}{\beta+j} - \sum_{t=0}^{n_i}\sum_{j=0}^{n_i-t-1}\frac{\pi_{it}(\alpha,\beta)}{\beta+j}\right].$$

Then the second-order partial derivatives, i.e.,

$$\frac{\partial^2 l(\alpha,\beta)}{\partial \alpha^2},\ \frac{\partial^2 l(\alpha,\beta)}{\partial \beta^2} \text{ and } \frac{\partial^2 l(\alpha,\beta)}{\partial \alpha \partial \beta},$$

can all be derived using the above relations. Here again, the likelihood equations have to be solved iteratively either by using the Newton–Raphson method or by direct optimization methods. The method of moments is not as tractable in this case, however, as in Cases A and B discussed earlier.

5.2.2 EM approach

Assume again that the only available information from the m independent start-up demonstration tests are the lengths of tests, i.e., n_i for $i = 1, 2, \ldots, m$. If the number of successes were also known, then the likelihood function would be

$$\begin{aligned} L(\alpha,\beta) &= \prod_{i=1}^{m} f(n_i, s_i; \alpha, \beta) = \left(\prod_{i=1}^{m} v_{n_i s_i}\right)\prod_{i=1}^{m}\left(\frac{B(\alpha+s_i, \beta+n_i-s_i)}{B(\alpha,\beta)}\right) \\ &= \left(\prod_{i=1}^{m} v_{n_i s_i}\right)\prod_{i=1}^{m}\left(\frac{\prod_{r=0}^{s_i-1}(\alpha+r)\prod_{r=0}^{n_i-s_i-1}(\beta+r)}{\prod_{r=0}^{n_i-1}(\alpha+\beta+r)}\right), \end{aligned}$$

with the corresponding log-likelihood function given by

$$\begin{aligned} l(\alpha,\beta) = &\sum_{i=1}^{m}\ln v_{n_i s_i} + \sum_{i=1}^{m}\sum_{r=0}^{s_i-1}\ln(\alpha+r) + \sum_{i=1}^{m}\sum_{r=0}^{n_i-s_i-1}\ln(\beta+r) \\ &- \sum_{i=1}^{m}\sum_{r=0}^{n_i-1}\ln(\alpha+\beta+r). \end{aligned}$$

Denoting by α_t and β_t the tth step estimates of α and β, we may write the mean of the log-likelihood function as follows:

$$
\begin{aligned}
Q(\alpha, \beta; \alpha_t, \beta_t) = E_{[S;X,\alpha_t,\beta_t]} l(\alpha, \beta) &= \sum_{i=1}^{m} E_{[S;X,\alpha_t,\beta_t]} \left[\ln v_{n_i s_i}\right] \\
&+ \sum_{i=1}^{m} E_{[S;X,\alpha_t,\beta_t]} \left[\sum_{r=0}^{s_i-1} \ln(\alpha + r)\right] \\
&+ \sum_{i=1}^{m} E_{[S;X,\alpha_t,\beta_t]} \left[\sum_{r=0}^{n_i-s_i-1} \ln(\beta + r)\right] - \sum_{i=1}^{m} \sum_{r=0}^{n_i-1} \ln(\alpha + \beta + r).
\end{aligned}
$$

The above three required expectations can be shown to be

$$
\begin{aligned}
E_{[S;X,\alpha_t,\beta_t]} \left[\ln v_{n_i s_i}\right] &= \sum_{j=0}^{n_i} \pi_{ij}(\alpha_t, \beta_t) \ln v_{n_i j}, \\
E_{[S;X,\alpha_t,\beta_t]} \left[\sum_{r=0}^{s-1} \ln(\alpha + r)\right] &= \sum_{s=0}^{n_i} \sum_{r=0}^{s-1} \pi_{is}(\alpha_t, \beta_t) \ln(\alpha + r), \\
E_{[S;X,\alpha_t,\beta_t]} \left[\sum_{r=0}^{n_i-s-1} \ln(\beta + r)\right] &= \sum_{s=0}^{n_i} \sum_{r=0}^{n_i-s-1} \pi_{is}(\alpha_t, \beta_t) \ln(\beta + r),
\end{aligned}
$$

and consequently, we have

$$
\begin{aligned}
Q(\alpha, \beta; \alpha_t, \beta_t) = E_{[S;X,\alpha_t,\beta_t]} l(\alpha, \beta) &= \sum_{i=1}^{m} E_{[S;X,\alpha_t,\beta_t]} \left[\ln v_{n_i s_i}\right] \\
&+ \sum_{i=1}^{m} \sum_{j=0}^{n_i} \sum_{r=0}^{j-1} \pi_{ij}(\alpha_t, \beta_t) \ln(\alpha + r) \\
&+ \sum_{i=1}^{m} \sum_{j=0}^{n_i} \sum_{r=0}^{n_i-j-1} \pi_{ij}(\alpha_t, \beta_t) \ln(\beta + r) - \sum_{i=1}^{m} \sum_{r=0}^{n_i-1} \ln(\alpha + \beta + r).
\end{aligned}
$$

Thus, the $(t + 1)$th iterate estimates of α and β will be

$$
\begin{aligned}
(\alpha_{t+1}, \beta_{t+1}) &= \operatorname{argmax}_{(\alpha,\beta)} Q(\alpha, \beta; \alpha_t, \beta_t) \\
&= \operatorname{argmax}_{(\alpha,\beta)} \left[\sum_{i=1}^{m} \sum_{j=0}^{n_i} \sum_{r=0}^{j-1} \pi_{ij}(\alpha_t, \beta_t) \ln(\alpha + r)\right. \\
&\left. + \sum_{i=1}^{m} \sum_{j=0}^{n_i} \sum_{r=0}^{n_i-j-1} \pi_{ij}(\alpha_t, \beta_t) \ln(\beta + r) - \sum_{i=1}^{m} \sum_{r=0}^{n_i-1} \ln(\alpha + \beta + r)\right].
\end{aligned}
$$

It is clear that the above maximization problem is equivalent to that when the full data were available, and its solution is secured by the use of numerical methods.

6 Some numerical results

In this section, we present some numerical results under the assumption that p follows a beta distribution. Before going into the methods of estimation of the beta parameters α and β, we first explore the accuracy of the bounds introduced in Sect. 3. Thus, in Table 1, we give some numerical results for the CSTF model (note that the applicability of the procedures practiced here is not restricted to the CSTF model alone) and various choices of the parameters α, β, m_1 and f_1; it can be easily seen that the bounds are in most cases very close to the exact value of the tail probability. Even though one cannot precisely state at what n the approximations would be warranted, Table 1 also shows that the developed approximations are accurate even when $n = 25$. This suggests that the approximations may be used for n of this size and larger.

Next, we examine the performance of estimation methods discussed in the preceding section. In Fig. 1, we focus on Case A (or equivalently, Case B) and the CSTF model, for specific values for the parameters α, β, m_1 and f_1. The method of moments, under the three different weighting methods, $w_i = 1$ (Fig. 1b) or $w_i = n_i$ (Fig. 1c) or $w_i = \frac{n_i}{1+\bar{\gamma}(n_i-1)}$ (Fig. 1d), and the MLE (Fig. 1a) are used for the estimation of the parameters α and β. The number of replications of samples was set as $r = 100$, and the computed measures are the mean absolute deviation from (α, β) defined by

$$\mathrm{MAD} = \left(\frac{1}{r} \sum_{i=1}^{r} |\hat{\alpha}_i - \alpha|, \frac{1}{r} \sum_{i=1}^{r} |\hat{\beta}_i - \beta| \right),$$

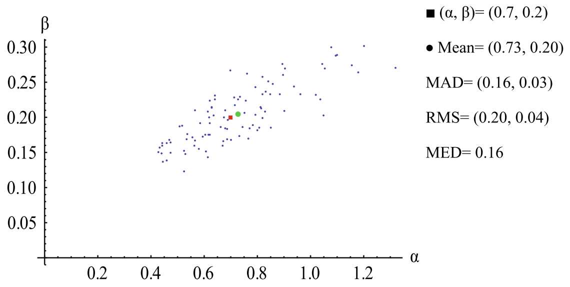
the Root Mean Square from (α, β) defined by

$$\mathrm{RMS} = \left(\sqrt{\frac{1}{r} \sum_{i=1}^{r} (\hat{\alpha}_i - \alpha)^2}, \sqrt{\frac{1}{r} \sum_{i=1}^{r} (\hat{\beta}_i - \beta)^2} \right),$$

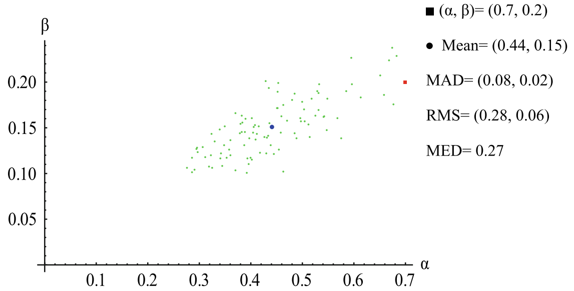
and the Mean Euclidean Distance from (α, β) defined by

Table 1 Bounds for the CSTF model

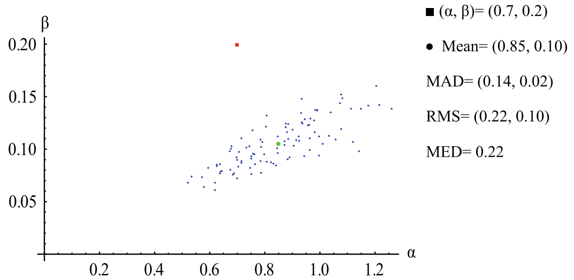
α	β	m_1	f_1	n	LB	$P(X > n)$	UB	UB−LB
9.5	1.5	10	3	10	0.461749	0.461749	0.481872	0.020124
				15	0.197201	0.210283	0.274012	0.076811
				20	0	0.049337	0.126171	0.126171
				25	0	0.007673	0.060454	0.060454
5.5	5.5	20	5	20	0.064161	0.064161	0.064403	0.000242
				25	0.028292	0.028295	0.029044	0.000752
				30	0.012699	0.012894	0.014121	0.001422
				35	0.005093	0.005914	0.007273	0.002180



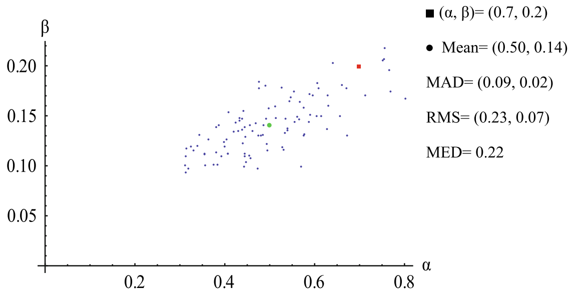
(a) MLE



(b) Method of Moments 1



(c) Method of Moments 2



(d) Method of Moments 3

Fig. 1 Estimation when the observed data are of the form of Cases A and B; the CSTF model with parameters $\alpha = 0.7$, $\beta = 0.2$, $m_1 = 10$, $f_1 = 2$, after $r = 100$ iterations and sample size $m = 150$

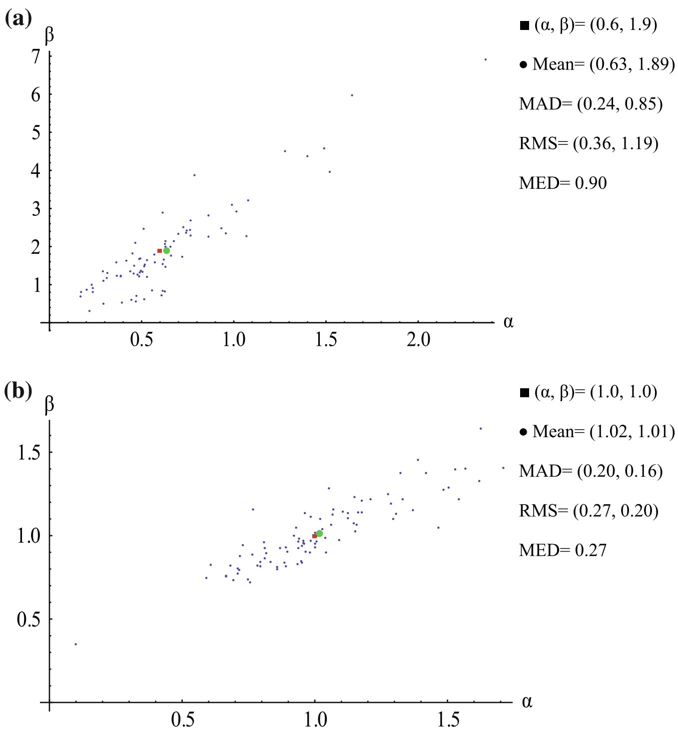


Fig. 2 MLE when the observed data are of the form of Case C

$$\text{MED} = \frac{1}{r}\sum_{i=1}^{r}\sqrt{(\hat{\alpha}_i - \alpha)^2 + (\hat{\beta}_i - \beta)^2}.$$

The quantities $\hat{\alpha}_i$ and $\hat{\beta}_i$ appearing in the formulae above are the estimates from the ith simulation.

The convergence of the Newton–Raphson method required 5–10 steps in general, and some times the algorithm failed to converge or resulted in negative values of the parameters (α, β) (the initial values were set by the method of moments). It is clear that the MLEs offer the most accurate results; the corresponding mean of 100 estimations is $(0.73, 0.20)$ (the true values were $(0.70, 0.20)$) while almost all the computed measures take the lowest values in the MLE method. For a discussion about the efficiency of these two methods (MLE and method of moments), the interested reader is also referred to Kleinman (1973).

In Fig. 2, we deal with the case when the observed data contain only the lengths of tests (Case C). For the evaluation of the initial values for the numerical method, we decided to follow the following procedure: we simulate m discrete, independent and uniformly distributed random numbers $u_1, u_2, \ldots, u_m$, with

$$u_i \in \{0, 1, \ldots, \min\{f_1, n_i\}\},\ i = 1, 2, \ldots, m,$$

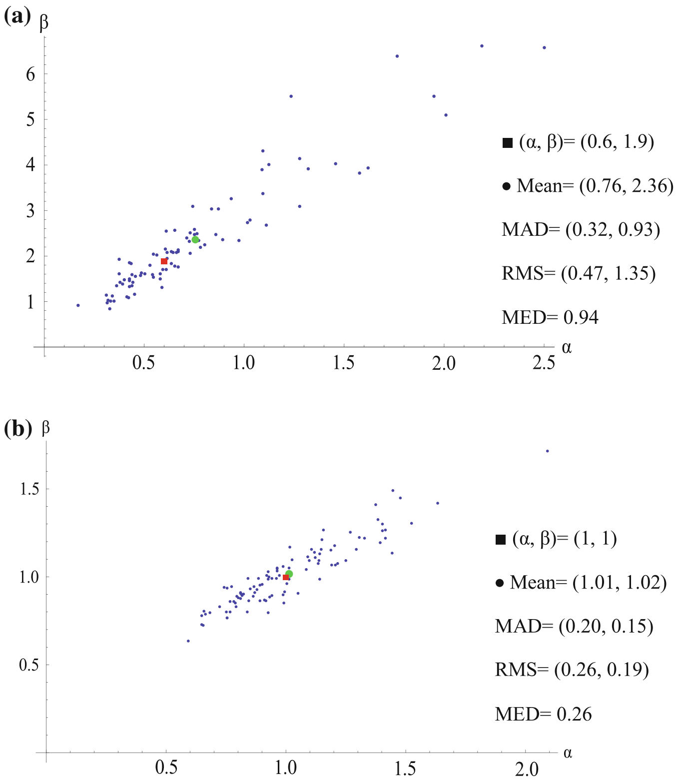


Fig. 3 EM algorithm estimation when the observed data are of the form of Case C

where $n_i, i = 1, 2, \ldots, m$, are the observed lengths of tests. These random numbers $u_1, u_2, \ldots, u_m$ play the role of the unobserved number of failures, and the method of moments (see Sect. 5.1) is then used to set the initial values. It is evident that in both of these models (for $\{\alpha = 0.6, \beta = 1.9, m_1 = 5, f_1 = 2\}$ in Fig. 2a and $\{\alpha = 1.0, \beta = 1.0, m_1 = 8, f_1 = 1\}$ in Fig. 2b), the MLE method offers accurate and promising results (especially for the second model).

In Fig. 3, we present the results obtained by the use of EM algorithm for the cases studied in Fig. 2. Although the observed outcomes are now less accurate (compare the three distances measures displayed in the corresponding graphs), we see that the fit is still quite good.

From these simulation results, it is observed that the maximum likelihood estimates are more accurate than the method of moments. Even though the estimates obtained by the method of moments were not accurate and sometimes even outside the range of the parameter, it provided convenient starting values for the numerical optimization procedure required in determination of MLEs. Moreover, this resulted in reducing the number of iterations in the optimization step, thus reducing the computational time. The scatter plots in Figs. 1, 2 and 3 also revealed that the true parameter value is located close to the center of the values of the MLEs, but not so for the method of moments estimates. Moreover, the contours being close to elliptical suggest that approximate normalized confidence intervals using the MLEs would result in good interval estimation of the model parameters. A more detailed study on this interval estimation method and the parametric bootstrap method as well as hypothesis testing methods will naturally be of great interest.

Appendix A

Before proceeding to the partial derivatives of the log-likelihood function from (11), it is necessary to mention that

$$\frac{\partial B(x,y)}{\partial x} = B(x,y)(\Psi(x)-\Psi(x+y)) \text{ and } \frac{\partial B(x,y)}{\partial y} = B(x,y)(\Psi(y)-\Psi(x+y)),$$

where $\Psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function. The following recurrence relation is also very useful in the sequel:

$$\Psi(x+1) = \Psi(x) + \frac{1}{x}, \text{ for every } x.$$

Thus, the partial derivatives of the log-likelihood function for Case C (see (11)) are as follows:

$$\begin{aligned}
\frac{\partial l(\alpha,\beta)}{\partial \alpha} &= -m(\Psi(\alpha)-\Psi(\alpha+\beta)) \\
&\quad + \sum_{i=1}^{m} \frac{\sum_{k=0}^{n_i} v_{n_i k} B(\alpha+k,\beta+n_i-k)(\Psi(\alpha+k)-\Psi(\alpha+\beta+n_i))}{\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t,\beta+n_i-t)} \\
&= -m(\Psi(\alpha)-\Psi(\alpha+\beta)) - \sum_{i=1}^{m} \Psi(\alpha+\beta+n_i) \\
&\quad + \sum_{i=1}^{m} \frac{\sum_{k=0}^{n_i} v_{n_i k} B(\alpha+k,\beta+n_i-k)\Psi(\alpha+k)}{\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t,\beta+n_i-t)} \\
&= -m(\Psi(\alpha)-\Psi(\alpha+\beta)) - \sum_{i=1}^{m}\left(\Psi(\alpha+\beta) + \sum_{r=1}^{n_i} \frac{1}{\alpha+\beta+n_i-r}\right) \\
&\quad + \sum_{i=1}^{m} \frac{\sum_{k=0}^{n_i} v_{n_i k} B(\alpha+k,\beta+n_i-k)(\Psi(\alpha) + \sum_{r=0}^{k-1} \frac{1}{\alpha+r})}{\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t,\beta+n_i-t)} \\
&= -m(\Psi(\alpha)-\Psi(\alpha+\beta)) - m\Psi(\alpha+\beta) - \sum_{i=1}^{m}\sum_{r=1}^{n_i} \frac{1}{\alpha+\beta+n_i-r} \\
&\quad + \sum_{i=1}^{m} \frac{\sum_{k=0}^{n_i} v_{n_i k} B(\alpha+k,\beta+n_i-k)\Psi(\alpha) + \sum_{k=0}^{n_i} v_{n_i k} B(\alpha+k,\beta+n_i-k) \sum_{r=0}^{k-1} \frac{1}{\alpha+r}}{\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t,\beta+n_i-t)}
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{i=1}^{m}\sum_{r=1}^{n_i}\frac{1}{\alpha+\beta+n_i-r}+\sum_{i=1}^{m}\frac{\sum_{k=0}^{n_i}\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\sum_{r=0}^{k-1}\frac{1}{\alpha+r}}{\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}\\
&= -\sum_{i=1}^{m}\sum_{r=0}^{n_i-1}\frac{1}{\alpha+\beta+r}+\sum_{i=1}^{m}\sum_{k=0}^{n_i}\sum_{r=0}^{k-1}\frac{\pi_{ik}(\alpha,\beta)}{\alpha+r}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial l(\alpha,\beta)}{\partial\beta} &= -m(\Psi(\beta)-\Psi(\alpha+\beta))\\
&\quad+\sum_{i=1}^{m}\frac{\sum_{k=0}^{n_i}\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)(\Psi(\beta+n_i-k)-\Psi(\alpha+\beta+n_i))}{\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}\\
&= -m(\Psi(\beta)-\Psi(\alpha+\beta))-\sum_{i=1}^{m}\Psi(\alpha+\beta+n_i)\\
&\quad+\sum_{i=1}^{m}\frac{\sum_{k=0}^{n_i}\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\Psi(\beta+n_i-k)}{\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}\\
&= -\sum_{i=1}^{m}\sum_{r=0}^{n_i-1}\frac{1}{\alpha+\beta+r}+\sum_{i=1}^{m}\frac{\sum_{k=0}^{n_i}\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\sum_{r=0}^{n_i-k-1}\frac{1}{\beta+r}}{\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}\\
&= -\sum_{i=1}^{m}\sum_{r=0}^{n_i-1}\frac{1}{\alpha+\beta+r}+\sum_{i=1}^{m}\sum_{k=0}^{n_i}\sum_{r=0}^{n_i-k-1}\frac{\pi_{ik}(\alpha,\beta)}{\beta+r}.
\end{aligned}
$$

Appendix B

The partial derivatives of $\pi_{ik}(\alpha,\beta)$ are as follows:

$$
\begin{aligned}
&\frac{\partial\pi_{ik}(\alpha,\beta)}{\partial\alpha} = \frac{\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)}{\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}\\
&= \frac{\upsilon_{n_ik}B'(\alpha+k,\beta+n_i-k)\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}{\left(\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)\right)^2}\\
&\quad-\frac{\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\sum_{t=0}^{n_i}\upsilon_{n_it}B'(\alpha+t,\beta+n_i-t)}{\left(\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)\right)^2}\\
&= \frac{\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\,(\Psi(\alpha+k)-\Psi(\alpha+\beta+n_i))\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}{\left(\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)\right)^2}\\
&\quad-\frac{\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)\,(\Psi(\alpha+t)-\Psi(\alpha+\beta+n_i))}{\left(\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)\right)^2}\\
&= \frac{\upsilon_{n_ik}B(\alpha+k,\beta+n_i-k)\Psi(\alpha+k)\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)}{\left(\sum_{t=0}^{n_i}\upsilon_{n_it}B(\alpha+t,\beta+n_i-t)\right)^2}
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{v_{n_i k} B(\alpha+k, \beta+n_i-k) \sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t) \Psi(\alpha+t)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&= \Psi(\alpha+k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \sum_{t=0}^{n_i} \pi_{it}(\alpha, \beta) \Psi(\alpha+t) \\
&= \Psi(\alpha+k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \sum_{t=0}^{n_i} \pi_{it}(\alpha, \beta) \left(\Psi(\alpha) + \sum_{j=0}^{t-1} \frac{1}{\alpha+j} \right) \\
&= \Psi(\alpha+k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \left(\Psi(\alpha) + \sum_{t=0}^{n_i} \sum_{j=0}^{t-1} \frac{\pi_{it}(\alpha, \beta)}{\alpha+j} \right) \\
&= \Psi(\alpha+k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \Psi(\alpha) - \pi_{ik}(\alpha, \beta) \sum_{t=0}^{n_i} \sum_{j=0}^{t-1} \frac{\pi_{it}(\alpha, \beta)}{\alpha+j} \\
&= \pi_{ik}(\alpha, \beta) \left[\sum_{j=0}^{k-1} \frac{1}{\alpha+j} - \sum_{t=0}^{n_i} \sum_{j=0}^{t-1} \frac{\pi_{it}(\alpha, \beta)}{\alpha+j} \right]
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \pi_{ik}(\alpha, \beta)}{\partial \beta} &= \frac{v_{n_i k} B(\alpha+k, \beta+n_i-k)}{\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)} \\
&= \frac{v_{n_i k} B'(\alpha+k, \beta+n_i-k) \sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&\quad -\frac{v_{n_i k} B(\alpha+k, \beta+n_i-k) \sum_{t=0}^{n_i} v_{n_i t} B'(\alpha+t, \beta+n_i-t)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&= \frac{v_{n_i k} B(\alpha{+}k, \beta{+}n_i-k) \left(\Psi(\beta{+}n_i-k) - \Psi(\alpha{+}\beta{+}n_i)\right) \sum_{t=0}^{n_i} v_{n_i t} B(\alpha{+}t, \beta{+}n_i-t)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&\quad -\frac{v_{n_i k} B(\alpha{+}k, \beta{+}n_i-k) \sum_{t=0}^{n_i} v_{n_i t} B(\alpha{+}t, \beta{+}n_i-t) \left(\Psi(\beta{+}n_i-t) - \Psi(\alpha{+}\beta{+}n_i)\right)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&= \frac{v_{n_i k} B(\alpha+k, \beta+n_i-k) \Psi(\beta+n_i-k) \sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&\quad -\frac{v_{n_i k} B(\alpha+k, \beta+n_i-k) \sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t) \Psi(\beta+n_i-t)}{\left(\sum_{t=0}^{n_i} v_{n_i t} B(\alpha+t, \beta+n_i-t)\right)^2} \\
&= \Psi(\beta+n_i-k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \sum_{t=0}^{n_i} \pi_{it}(\alpha, \beta) \Psi(\beta+n_i-t) \\
&= \Psi(\beta+n_i-k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \sum_{t=0}^{n_i} \pi_{it}(\alpha, \beta) \left(\Psi(\beta) + \sum_{j=0}^{n_i-t-1} \frac{1}{\beta+j} \right) \\
&= \Psi(\beta+n_i-k) \pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta) \left(\Psi(\beta) + \sum_{t=0}^{n_i} \sum_{j=0}^{n_i-t-1} \frac{\pi_{it}(\alpha, \beta)}{\beta+j} \right)
\end{aligned}
$$

$$= \Psi(\beta + n_i - k)\pi_{ik}(\alpha, \beta) - \pi_{ik}(\alpha, \beta)\Psi(\beta) - \pi_{ik}(\alpha, \beta) \sum_{t=0}^{n_i} \sum_{j=0}^{n_i - t - 1} \frac{\pi_{it}(\alpha, \beta)}{\beta + j}$$

$$= \pi_{ik}(\alpha, \beta) \left[\sum_{j=0}^{n_i - k - 1} \frac{1}{\beta + j} - \sum_{t=0}^{n_i} \sum_{j=0}^{n_i - t - 1} \frac{\pi_{it}(\alpha, \beta)}{\beta + j} \right].$$

References

Aki, S., Kuboki, H., Hirano, K. (1984). On discrete distributions of order k. *Annals of the Institute of Statistical Mathematics*, *36*, 431–440.

Antzoulakos, D. L., Koutras, M. V., Rakitzis, A. C. (2009). Start-up demonstration tests based on run and scan statistics. *Journal of Quality Technology*, *41*, 1–12.

Aston, J. A. D., Martin, D. E. K. (2005). Waiting time distributions of competing patterns in higher-order Markovian sequences. *Journal of Applied Probability*, *42*, 977–988.

Balakrishnan, N., Chan, P. S. (2000). Start-up demonstration tests with rejection of units upon observing d failures. *Annals of the Institute of Statistical Mathematics*, *52*, 184–196.

Balakrishnan, N., Koutras, M. V. (2002). *Runs and Scans with Applications*. New York: Wiley.

Balakrishnan, N., Balasubramanian, K., Viveros, R. (1995). Start-up demonstration tests under correlation and corrective action. *Naval Research Logistics*, *42*, 1271–1276.

Balakrishnan, N., Mohanty, S. G., Aki, S. (1997). Start-up demonstration tests under Markov dependence model with corrective actions. *Annals of the Institute of Statistical Mathematics*, *49*, 155–169.

Barlow, R., Proschan, F. (1981). *Statistical Theory of Reliability and Life Testing*. New York:Holt, Reinhart and Winston.

Chan, P. S., Ng, H. K. T., Balakrishnan, N. (2008). Statistical inference for start-up demonstration tests with rejection of units upon observing d failures. *Journal of Applied Statistics*, *35*, 867–878.

Chatfield, C., Goodhardt, G. J. (1970). The beta-binomial model for consumer purchasing behavior. *Journal of the Royal Statistical Society, Series C*, *19*, 240–250.

de Finetti, B. (1931). Funcione caratteristica di un fenomeno aleatorio. *Atti della Reale Accademia Nazionale dei Lincii*, *4*, 251–299.

Eryilmaz, S. (2010). Start-up demonstration tests under Markov dependence. *Pakistan Journal of Statistics*, *26*, 637–647.

Eryilmaz, S., Chakraborti, S. (2008). On start-up demonstration tests under exchangeability. *IEEE Transactions on Reliability*, *57*, 627–632.

Everson, P. J., Bradlow, E. T. (2002). Bayesian inference for the beta-binomial distribution via polynomial expansions. *Journal of Computational and Graphical Statistics*, *11*, 202–207.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications* (3rd ed., Vol. I). New York: Wiley.

Gera, A. E. (2004). Combined k-out-of-n:G, and consecutive k_C-out-of-n:G systems. *IEEE Transactions on Reliability*, *53*, 523–531.

Gera, A. E. (2010). A new start-up demonstration test. *IEEE Transactions on Reliability*, *59*, 128–131.

Gera, A. E. (2011). A general model for start-up demonstration tests. *IEEE Transactions on Reliability*, *60*, 295–304.

Govindaraju, K., Lai, C. D. (1999). Design of multiple run sampling plan. *Communications in Statistics - Simulation and Computation*, *28*, 1–11.

Griffiths, D. A. (1973). Maximum likelihood estimation for the beta binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics*, *29*, 637–648.

Hahn, G. J., Gage, J. B. (1983). Evaluation of a start-up demonstration test. *Journal of Quality Technology*, *15*, 103–106.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. New Jersey: Wiley.

Johnson, N. L., Kotz, S., Kemp, A. W. (1992). *Univariate Discrete Distributions* (2nd ed.). New York: Wiley.

Kleinman, J. C. (1973). Proportions with extraneous variance: single and independent samples. *Journal of the American Statistical Association*, *68*, 46–54.

Koutras, M. V., Balakrishnan, N. (1999). A start-up demonstration test using a simple scan-based statistic. In J. Glaz, N. Balakrishnan (Eds.), *Scan Statistics and Applications* (pp. 251–267). Boston: Birkhauser.

Lee, J. C., Lio, Y. L. (1999). A note on Bayesian estimation and prediction for the beta-binomial model. *Journal of Statistical Computation and Simulation*, *63*, 73–91.

Lee, J. C., Sabavala, D. J. (1987). Bayesian estimation and prediction for the beta-binomial model. *Journal of Business and Economic Statistics*, *5*, 357–367.

Martin, D. E. K. (2004). Markovian start-up demonstration tests with rejection of units upon observing d failures. *European Journal of Operational Research*, *155*, 474–486.

Martin, D. E. K. (2008). Application of auxiliary Markov chains to start-up demonstration tests. *European Journal of Operational Research*, *184*, 574–582.

Philippou, A. N., Muwafi, A. A. (1982). Waiting for the k-th consecutive success and the fibonacci sequence of order k. *The Fibonacci Quarterly*, *20*, 28–32.

Scollnik, D. P. M. (2010). Bayesian statistical inference for start-up demonstration tests with rejection of units upon observing d failures. *Journal of Applied Statistics*, *37*, 1113–1121.

Scollnik, D. P. M. (2011). Bayesian inference for a class of start-up demonstration tests with rejection of units upon the observation of d failures. *Communications in Statistics-Theory and Methods*, *40*, 2528–2538.

Shenton, L. R. (1950). Maximum likelihood and the efficiency of the method of moments. *Biometrika*, *37*, 111–116.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society, Series B*, *10*, 257–261.

Smith, M. L., Griffith, W. S. (2005). Start-up demonstration tests based on consecutive successes and total failures. *Journal of Quality Technology*, *37*, 186–198.

Smith, M. L., Griffith, W. S. (2008). The analysis and comparison of start-up demonstration tests. *European Journal of Operational Research*, *186*, 1029–1045.

Smith, M. L., Griffith, W. S. (2011). Multistate start-up demonstration tests. *International Journal of Reliability, Quality and Safety Engineering*, *18*, 99–117.

Tamura, R. N., Young, S. S. (1986). The incorporation of historical control information in tests of proportions: Simulation study of Tarone's procedure. *Biometrics*, *42*, 343–349.

Tamura, R. N., Young, S. S. (1987). A stabilized moment estimator for the beta-binomial distribution. *Biometrics*, *43*, 813–824.

Tripathi, R. C., Gupta, R. C., Gurland, J. (1994). Estimation of parameters in the beta binomial model. *Annals of the Institute of Statistical Mathematics*, *46*, 317–331.

Vance, L. C., McDonald, G. C. (1979). A class of multiple run sampling plans. *Technometrics*, *21*, 141–146.

Viveros, R., Balakrishnan, N. (1993). Statistical inference from start-up demonstration test data. *Journal of Quality Technology*, *25*, 119–130.

Wilcox, R. R. (1979). Estimating the parameters of the beta binomial distribution. *Educational and Psychological Measurement*, *39*, 527–535.

Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics*, *31*, 949–952.

Yalcin, F., Eryilmaz, S. (2012). Start-up demonstration test based on total successes and total failures with dependent start-ups. *IEEE Transactions on Reliability*, *61*, 227–230.

Yamamoto, E., Yanagimoto, T. (1992). Moment estimators for the beta-binomial distribution. *Journal of Applied Statistics*, *19*, 273–283.