

On the joint distribution of stopping times and stopped sums in multistate exchangeable trials

Boutsikas M.V.¹, Antzoulakos D.L.², Rakitzis A.C.³

¹Dept. of Statistics & Insurance Science, Univ. of Piraeus, Greece (mbouts@unipi.gr).

²Dept. of Statistics & Insurance Science, Univ. of Piraeus, Greece (dantz@unipi.gr).

³Dept. of Mathematics & Statistics, Univ. of Cyprus, Cyprus (arakitz@ucy.ac.cy).

Abstract

Let T be a stopping time associated with a sequence of iid or exchangeable random variables taking values in $\{0, 1, 2, \dots, m\}$, and let $S_{T,i}$ be the stopped sum denoting the number of appearances of outcome " i " in $X_1, \dots, X_T$, $0 \leq i \leq m$. In this article we present results revealing that, if the distribution of T is known, then we can also derive the joint distribution of $(T, S_{T,0}, S_{T,1}, \dots, S_{T,m})$. Two applications, which have independent interest are offered to illustrate the applicability and the usefulness of the main results.

Key words and phrases: stopping time, stopped sums, exchangeability, multistate trials, run and scan statistics, acceptance sampling, coupon collector's problem.

AMS 2010 subject classification: Primary: 60E05, 62E15, Secondary: 60G09, 60G50

1. Introduction

The study of distributional properties of *stopping times* T and *stopped sums* of the form $S_T = \sum_{i=1}^T X_i$, defined over sequences of random variables (r.v.'s) $X_1, X_2, \dots$ has attracted much interest in the literature over the last decades because of their wide range of applicability in diverse scientific areas. Associated problems of specific interest are the ones related to the distribution theory of *runs* and *patterns*, which have direct applications in practical problems of quality control, reliability theory, actuarial science, radar astronomy, psychology, molecular biology, etc. (e.g. cf. Balakrishnan and Koutras (2002)). The distribution theory of runs and patterns has also been naturally extended in order to cover the case of *exchangeable trials* (see, for example, Eryilmaz (2008a,b, 2010), Inoue *et al.* (2011), Makri and Psillakis (2011)).

When studying a stopping time T in a sequence $X_1, X_2, \dots$, the examination of the total number $S_{T,i}$ of outcomes of type " i " in $X_1, X_2, \dots, X_T$ may provide useful information about the nature of the underlying statistical experiment. Aki and Hirano (1994), who seem to have

been the first to become attracted to these kind of problems, studied the joint distribution of (T, S_T) for T being the waiting time for the first occurrence of a success run of length k , and S_T being the total number of successes until T , in a sequence of binary independent and identically distributed (iid) trials. Several extensions and variations of their work were subsequently examined by Balakrishnan (1997), Uchida (1998), Chadjiconstantinidis *et al.* (2000) and Inoue (2004).

In a recent article, Antzoulakos and Boutsikas (2007) presented a method to obtain the joint probability generating function (pgf) of (T, S_T, F_T) directly through the pgf of T , in the case where T is the waiting time for the r -th occurrence of a pattern $\mathcal{E}$ in a sequence of binary iid trials. In this paper, we generalize and extend their work by considering general stopping times defined in sequences of multistate iid or exchangeable trials. More specifically, we derive the joint pgf of the random vector $(T, S_{T,0}, S_{T,1}, \dots, S_{T,m})$ directly via the pgf of T , when $X_1, X_2, \dots$ are iid or exchangeable r.v.'s taking values in $\{0, 1, \dots, m\}$. Finally, we present two applications related to quality control and to coupon collector's problem.

2. Main Results

2.1 Independent trials

Let $X_1, X_2, \dots$ be a sequence of r.v.'s taking values in $\{0, 1, \dots, m\}$. Let also T be a stopping time associated with $X_1, X_2, \dots$, that is, for each $n \geq 1$ the event $[T = n]$ is completely determined by the information contained in $\mathbf{X}_n = (X_1, \dots, X_n)$ (i.e., $[T = n] \in \sigma(\mathbf{X}_n)$). Furthermore, let $S_{n,i}$ be the stopped sum expressing the total number of outcomes of type " i " in $\mathbf{X}_n$, namely

$$S_{n,i} = \sum_{j=1}^n I_{[X_j=i]}, \quad i = 0, 1, \dots, m, \quad n \geq 1,$$

where $I_A = 1$ or 0 when A occurs or not. The above are defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\mathbf{z}, \mathbf{p}})$, where the probability measure $\mathbb{P}_{\mathbf{z}, \mathbf{p}}$, is such that $X_1, X_2, \dots$ are iid r.v.'s with

$$\mathbb{P}_{\mathbf{z}, \mathbf{p}}(X_n = i) = \frac{p_i z_i}{\sum_{j=0}^m p_j z_j}, \quad i = 0, 1, \dots, m, \quad n \geq 1, \quad (2.1)$$

where $\mathbf{z} = (z_0, z_1, \dots, z_m)$, $z_i > 0$, $\mathbf{p} = (p_1, \dots, p_m)$, $p_0 = 1 - \sum_{i=1}^m p_i$, $p_i \in (0, 1)$. When $\mathbf{z} = \mathbf{1} = (1, \dots, 1)$, the r.v.'s $X_1, X_2, \dots$ are iid with $\mathbb{P}_{\mathbf{1}, \mathbf{p}}(X_n = i) = p_i$, $i = 0, 1, \dots, m$, $n \geq 1$. We denote by $\mathbb{E}_{\mathbf{z}, \mathbf{p}}(\cdot)$ the expectation taken under the probability measure $\mathbb{P}_{\mathbf{z}, \mathbf{p}}$. In the rest of this section we assume that $\mathbb{P}_{\mathbf{z}, \mathbf{p}}(T < \infty) = 1$ for all positive real numbers $z_0, z_1, \dots, z_m$ in a neighborhood of zero (cf. also Remark 2). In Proposition 1 we express the joint pgf of $(T, \mathbf{S}_T) = (T, S_{T,0}, S_{T,1}, \dots, S_{T,m})$ under $\mathbb{P}_{\mathbf{1}, \mathbf{p}}$, in terms of the pgf of T under $\mathbb{P}_{\mathbf{z}, \mathbf{p}}$.

Proposition 1. *The joint pgf of $(T, \mathbf{S}_T)$ for iid trials satisfies the relation*

$$\mathbb{E}_{\mathbf{1}, \mathbf{p}}(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}}) = \mathbb{E}_{\mathbf{z}, \mathbf{p}}((u \sum_{i=0}^m p_i z_i)^T),$$

for all $u, z_0, z_1, \dots, z_m \in (0, \infty)$ in a neighborhood of zero such that the above expectations exist.

Proof. Let $\mathbf{S}_n = (S_{n,0}, \dots, S_{n,m})$ and define the set

$$E_n(\mathbf{s}_n) = \{\mathbf{x} \in \{0, 1, \dots, m\}^n : \sum_{i=1}^n (I_{[x_i=0]}, \dots, I_{[x_i=m]}) = \mathbf{s}_n\},$$

where $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{s}_n = (s_{n,0}, \dots, s_{n,m})$, $\sum_{i=0}^m s_{n,i} = n$. The joint probability mass function (pmf) of $(T, \mathbf{S}_T)$ can be written in the form

$$\begin{aligned} \mathbb{P}_{\mathbf{z}, \mathbf{p}}(T = n, \mathbf{S}_T = \mathbf{s}_n) &= \sum_{\mathbf{x} \in \{0, 1, \dots, m\}^n} \mathbb{P}_{\mathbf{z}, \mathbf{p}}(T = n, \mathbf{S}_n = \mathbf{s}_n | \mathbf{X}_n = \mathbf{x}) \mathbb{P}_{\mathbf{z}, \mathbf{p}}(\mathbf{X}_n = \mathbf{x}) \\ &= \sum_{\mathbf{x} \in E_n(\mathbf{s}_n)} \mathbb{P}_{\mathbf{z}, \mathbf{p}}(T = n | \mathbf{X}_n = \mathbf{x}) \prod_{i=0}^m \left(\frac{p_i z_i}{\sum_{j=0}^m p_j z_j} \right)^{s_{n,i}} = a(n, \mathbf{s}_n) \prod_{i=0}^m \left(\frac{p_i z_i}{\sum_{j=0}^m p_j z_j} \right)^{s_{n,i}}, \end{aligned}$$

where $a(n, \mathbf{s}_n)$ is equal to the number of elements $\mathbf{x}$ in $E_n(\mathbf{s}_n)$ such that the stopping time T takes the value n when $\mathbf{X}_n = \mathbf{x}$. Thus, we get that

$$\begin{aligned} \mathbb{E}_{\mathbf{1}, \mathbf{p}}(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}} I_{[T < \infty]}) &= \sum_{n=1}^{\infty} \sum_{\mathbf{s}_n} \mathbb{P}_{\mathbf{1}, \mathbf{p}}(T = n, \mathbf{S}_n = \mathbf{s}_n) u^n \prod_{i=0}^m z_i^{s_{n,i}} \\ &= \sum_{n=1}^{\infty} \sum_{\mathbf{s}_n} a(n, \mathbf{s}_n) u^n \prod_{i=0}^m (p_i z_i)^{s_{n,i}} = \sum_{n=1}^{\infty} (u \sum_{i=0}^m p_i z_i)^n \sum_{\mathbf{s}_n} a(n, \mathbf{s}_n) \prod_{i=0}^m \left(\frac{p_i z_i}{\sum_{i=0}^m p_i z_i} \right)^{s_{n,i}} \\ &= \sum_{n=1}^{\infty} (u \sum_{i=0}^m p_i z_i)^n \sum_{\mathbf{s}_n} \mathbb{P}_{\mathbf{z}, \mathbf{p}}(T = n, \mathbf{S}_n = \mathbf{s}_n) = \sum_{n=1}^{\infty} (u \sum_{i=0}^m p_i z_i)^n \mathbb{P}_{\mathbf{z}, \mathbf{p}}(T = n) \\ &= \mathbb{E}_{\mathbf{z}, \mathbf{p}}((u \sum_{i=0}^m p_i z_i)^T I_{[T < \infty]}), \end{aligned}$$

which completes the proof of the proposition taking into account that $\mathbb{P}_{\mathbf{z}, \mathbf{p}}(T < \infty) = 1$. $\blacksquare$

Remark 1. Consider the stopped sum $S_T = \sum_{i=0}^m i S_{T,i} = \sum_{j=1}^T X_j$. Letting $z_i = z^i$ for $i = 0, 1, \dots, m$, Proposition 1 yields

$$\mathbb{E}_{\mathbf{1}, \mathbf{p}}(u^T z^{S_T}) = \mathbb{E}_{\mathbf{1}, \mathbf{p}}(u^T z^{0S_{T,0}} z^{1S_{T,1}} \dots z^{mS_{T,m}}) = \mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{p}}((u \sum_{i=0}^m p_i z^i)^T) = \mathbb{E}_{\tilde{\mathbf{z}}, \mathbf{p}}((u \mathbb{E}_{\mathbf{1}, \mathbf{p}}(z^{X_1}))^T),$$

where $\tilde{\mathbf{z}} = (z^0, z^1, \dots, z^m)$. The above relation was proved by Antzoulakos and Boutsikas (2007) for iid $\{0, 1\}$ -valued r.v.'s (when T expresses a specific waiting time) and by Boutsikas *et al.* (2011) for real valued iid r.v.'s via a different methodology.

Remark 2. Under the assumptions of Proposition 1, if we take $u = (\sum_{i=0}^m p_i z_i)^{-1}$ we get

$$\mathbb{E}_{\mathbf{1}, \mathbf{p}}((\sum_{i=0}^m p_i z_i)^{-T} z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}}) = 1, \quad (2.2)$$

which can be considered as an extended version of Wald's identity (see, e.g., Karlin and Taylor (1975), p. 264). Moreover, differentiating both sides of (2.2) with respect to z_i and evaluating the resulting expression at $\mathbf{z} = \mathbf{1}$ we get $\mathbb{E}_{\mathbf{1},\mathbf{p}}(S_{T,i}) = p_i \mathbb{E}_{\mathbf{1},\mathbf{p}}(T)$, $i = 0, 1, \dots, m$, provided that $\mathbb{P}_{\mathbf{z},\mathbf{p}}(T < \infty) = 1$ for all $\mathbf{z}$ in a subinterval of R_+^m containing $\mathbf{1}$. Similarly, taking second order derivatives, we deduce the variance and the covariance of the r.v.'s $W_i = S_{T,i} - p_i T$, namely

$$\mathbb{V}_{\mathbf{1},\mathbf{p}}(W_i) = p_i(1 - p_i) \mathbb{E}_{\mathbf{1},\mathbf{p}}(T), \quad \mathbb{C}_{\mathbf{1},\mathbf{p}}(W_i, W_j) = -p_i p_j \mathbb{E}_{\mathbf{1},\mathbf{p}}(T),$$

for $i, j = 0, 1, \dots, m$, $i \neq j$. The above equations regarding $\mathbb{E}_{\mathbf{1},\mathbf{p}}(S_{T,i})$ and $\mathbb{V}_{\mathbf{1},\mathbf{p}}(S_{T,i} - p_i T)$ can be considered as versions of the well known *first* and *second Wald's equation* respectively.

2.2 Exchangeable trials

In this subsection we consider the case where $X_1, X_2, \dots$ is an infinite sequence of exchangeable multistate trials defined over a measurable space $(\Omega, \mathcal{F}, \mathbb{P}_e)$ and taking values in the set $\{0, 1, \dots, m\}$. Exchangeability implies that for all $n \geq 1$,

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

for any permutation $(\pi(1), \pi(2), \dots, \pi(n))$ of the indices $\{1, 2, \dots, n\}$. We keep all other notations and assumptions of the previous subsection. In particular, T is a stopping time associated with $X_1, X_2, \dots$, and $S_{n,i}$ denotes the total number of outcomes of type " i " in $\mathbf{X}_n$. We consider again $\mathbb{P}_{\mathbf{z},\mathbf{p}}$ under which (2.1) holds true. It is worth stressing that the r.v.'s $X_1, X_2, \dots$ are *exchangeable* under $\mathbb{P}_e$ whereas they are *independent* under $\mathbb{P}_{\mathbf{z},\mathbf{p}}$. Consider the following simplex $\mathcal{D} = \{\mathbf{p} \in [0, 1]^m : p_1 + p_2 + \dots + p_m \leq 1\}$. From de Finetti's representation theorem for discrete r.v.'s (see, e.g. Mauldin *et al.* (1992)) we have that there exists a probability measure on $\mathcal{D}$ with cumulative distribution function (cdf) G such that

$$\mathbb{P}_e(\mathbf{X}_n = \mathbf{x}) = \int_{\mathcal{D}} \prod_{i=0}^m p_i^{r_{n,i}(\mathbf{x})} dG(\mathbf{p}), \quad n \geq 1,$$

where $\mathbf{x} = (x_1, \dots, x_n)$ and $r_{n,i}(\mathbf{x}) = \sum_{j=1}^n I_{[x_j=i]}$, $i = 0, 1, \dots, m$. The measure on $\mathcal{D}$ corresponding to G is usually called the *de Finetti measure* or the *mixing measure* for the sequence $X_1, X_2, \dots$. In the Proposition 2 we express the joint pgf of $(T, \mathbf{S}_T)$ in the case of *exchangeable* trials under the measure $\mathbb{P}_e$ in terms of the pgf of T in the case of *independent* trials under the measure $\mathbb{P}_{\mathbf{z},\mathbf{p}}$. We assume again that $\mathbb{P}_e(T < \infty) = \mathbb{P}_{\mathbf{z},\mathbf{p}}(T < \infty) = 1$, $\mathbf{p} \in \mathcal{D}$.

Proposition 2. *The joint pgf of $(T, \mathbf{S}_T)$ for exchangeable trials satisfies the relation*

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}}) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{z},\mathbf{p}}((u \sum_{i=0}^m p_i z_i)^T) dG(\mathbf{p}),$$

for all $u, z_0, z_1, \dots, z_m \in (0, \infty)$ in a neighborhood of zero such that the above expectations exist.

Proof. At first observe that if Y is a r.v. such that $Y \cdot I_{[T=n]}$ is a $\sigma(\mathbf{X}_n)$ - measurable r.v. (i.e. there exists a measurable function $\varphi_n : R^n \rightarrow R$ such that $Y \cdot I_{[T=n]} = \varphi_n(\mathbf{X}_n)$), then invoking de Finetti's representation theorem we have that

$$\begin{aligned}\mathbb{E}_e(Y I_{[T<\infty]}) &= \sum_{n=1}^{\infty} \mathbb{E}_e(Y I_{[T=n]}) = \sum_{n=1}^{\infty} \sum_{\mathbf{x}} \varphi_n(\mathbf{x}) \mathbb{P}_e(\mathbf{X}_n = \mathbf{x}) = \sum_{n=1}^{\infty} \sum_{\mathbf{x}} \varphi_n(\mathbf{x}) \int_{\mathcal{D}} \prod_{i=0}^m p_i^{r_{n,i}(\mathbf{x})} dG(\mathbf{p}) \\ &= \sum_{n=1}^{\infty} \int_{\mathcal{D}} \sum_{\mathbf{x}} \varphi_n(\mathbf{x}) \mathbb{P}_{\mathbf{1},\mathbf{p}}(\mathbf{X}_n = \mathbf{x}) dG(\mathbf{p}) = \sum_{n=1}^{\infty} \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(Y I_{[T=n]}) dG(\mathbf{p}) \\ &= \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(Y I_{[T<\infty]}) dG(\mathbf{p}).\end{aligned}$$

Note that, under $\mathbb{P}_{\mathbf{1},\mathbf{p}}$ the r.v.'s $X_1, X_2, \dots$ are iid with $\mathbb{P}_{\mathbf{1},\mathbf{p}}(X_j = i) = p_i, i = 0, 1, \dots, m$. If we take $Y = u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}}$, then $Y I_{[T=n]}$ is a $\sigma(\mathbf{X}_n)$ - measurable r.v., and therefore

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}} I_{[T<\infty]}) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}} I_{[T<\infty]}) dG(\mathbf{p}).$$

Applying Proposition 1 we finally deduce that

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}} I_{[T<\infty]}) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{z},\mathbf{p}}((u \sum_{i=0}^m p_i z_i)^T I_{[T<\infty]}) dG(\mathbf{p}),$$

which completes the proof taking into account that $\mathbb{P}_e(T < \infty) = \mathbb{P}_{\mathbf{z},\mathbf{p}}(T < \infty) = 1$. ■

Remark 3. As it can be seen in their proofs, Propositions 1, 2 hold true even when $\mathbb{P}_e(T < \infty) < 1$ or $\mathbb{P}_{\mathbf{z},\mathbf{p}}(T < \infty) < 1$, by inserting the indicator function $I_{[T<\infty]}$ in the expectations appearing in their statements.

Remark 4. Recall that if Y is a r.v. such that $Y I_{[T=n]}$ is $\sigma(\mathbf{X}_n)$ - measurable, and $T < \infty$ a.s., then $\mathbb{E}_e(Y) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(Y) dG(\mathbf{p})$ (see the proof of Proposition 2). Hence, on choosing $Y = S_{T,i}$ and taking into account that $\mathbb{E}_{\mathbf{1},\mathbf{p}}(S_{T,i}) = p_i \mathbb{E}_{\mathbf{1},\mathbf{p}}(T)$ (see Remark 2), we deduce that

$$\mathbb{E}_e(S_{T,i}) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(S_{T,i}) dG(\mathbf{p}) = \int_{\mathcal{D}} p_i \mathbb{E}_{\mathbf{1},\mathbf{p}}(T) dG(\mathbf{p}).$$

Analogous results can be derived for all equations stated in Remark 2, namely

$$\mathbb{V}_e(W_i) = \int_{\mathcal{D}} p_i(1 - p_i) \mathbb{E}_{\mathbf{1},\mathbf{p}}(T) dG(\mathbf{p}), \quad \mathbb{C}_e(W_i, W_j) = - \int_{\mathcal{D}} p_i p_j \mathbb{E}_{\mathbf{1},\mathbf{p}}(T) dG(\mathbf{p}).$$

where $W_i = S_{T,i} - p_i T$, $i, j = 0, 1, \dots, m$, $i \neq j$. Moreover, $\mathbb{E}_e(T) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(T) dG(\mathbf{p})$.

Remark 5. Similar to Remark 1, we can show that, under the assumptions of Proposition 2,

$$\mathbb{E}_e(u^T z^{S_T}) = \int_{\mathcal{D}} \mathbb{E}_{\tilde{\mathbf{z}},\mathbf{p}}((u \mathbb{E}_{\mathbf{1},\mathbf{p}}(z^{X_1}))^T) dG(\mathbf{p}),$$

where $S_T = \sum_{j=1}^T X_j$, $\tilde{\mathbf{z}} = (z^0, z^1, \dots, z^m)$. Moreover, $\mathbb{E}_e(S_T) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\mathbf{p}}(X_1) \mathbb{E}_{\mathbf{1},\mathbf{p}}(T) dG(\mathbf{p})$.

3. Applications

3.1 On the quality of lots until switching the inspection level in acceptance sampling

In lot-by-lot acceptance sampling systems for variables (the quality characteristic of interest is described by a continuous r.v.) a random sample is taken from each lot to assess its quality. The sampling procedure usually starts with “normal” inspection that is used as long as products are produced near an acceptable quality level. However, when there are indications that the quality of the products has changed, a switch in a “tightened” or “reduced” inspection level is instituted according to whether the quality of the products has been, respectively, deteriorated or improved. Traditionally, each inspected lot is classified either as accepted or rejected. However, for more general applications, we may assume that each lot is classified into one of the three levels 0, 1, and 2 indicating intermediate quality, highest quality and lowest quality, respectively (see, e.g., Lou and Fu (2009)). In such cases the switching rules are naturally described by specific $\{0, 1, 2\}$ -patterns formed by the successively inspected lots.

Suppose that ν independent reference samples are available from the process producing products at the intermediate level and let Y_t , $1 \leq t \leq \nu$, be the statistic of interest of the t -th sample, such as the sample mean or the sample median. Let $U = g_1(Y_1, Y_2, \dots, Y_\nu)$ and $L = g_2(Y_1, Y_2, \dots, Y_\nu)$ be two random thresholds determining the quality of future lots as follows: From each (future) lot take a random sample of size n and let Z_i be the statistic of interest for the i -th sample, $i = 1, 2, \dots$. The i -th lot is classified at level 0, 1 and 2 if $L < Z_i < U$, $Z_i \geq U$ and $Z_i \leq L$, respectively.

For $i = 1, 2, \dots$, let $X_i = 1, 0$ or 2 when $Z_i \geq U$, $L < Z_i < U$ or $Z_i \leq L$ respectively. As usual, we denote by F_{V_1, V_2} and f_{V_1, V_2} the joint cdf and the joint probability density function (pdf) of r.v.'s V_1, V_2 respectively, and by F_V , f_V the cdf and the pdf of a r.v. V . It is now easy to check that, for every $n \geq 1$, and $(x_1, \dots, x_n) \in \{0, 1, 2\}^n$,

$$\begin{aligned} \mathbb{P}_e(X_1 = x_1, \dots, X_n = x_n) &= \int_{R^2} (1 - F_Z(u))^{c_1} (F_Z(l))^{c_2} (F_Z(u) - F_Z(l))^{n-c_1-c_2} dF_{U,L}(u, l) \\ &= \int_{\mathcal{D}} p_1^{c_1} p_2^{c_2} (1 - p_1 - p_2)^{n-c_1-c_2} dG(\mathbf{p}), \end{aligned}$$

where $c_1 = \sum_{i=1}^n I_{[x_i=1]}$, $c_2 = \sum_{i=1}^n I_{[x_i=2]}$ and $G(\mathbf{p}) = G(p_1, p_2) = F_{U,L}(F_Z^{-1}(1-p_1), F_Z^{-1}(p_2))$. Hence, the r.v.'s $X_1, X_2, \dots$ form an infinite sequence of trinary exchangeable r.v.'s with mixing measure $G(\mathbf{p})$. Typical random thresholds L, U used in applications are the r -th and the s -th ($1 \leq r < s \leq \nu$) smallest order statistics of the reference sample to be denoted by $Y_{r:\nu}$ and $Y_{s:\nu}$, respectively. The joint pdf $f_{U,L} = f_{Y_{s:\nu}, Y_{r:\nu}}$ is given by

$$f_{U,L}(y_1, y_2) = \nu! \frac{(F_Y(y_2))^{r-1} (F_Y(y_1) - F_Y(y_2))^{s-r-1} (1 - F_Y(y_1))^{\nu-s}}{(r-1)!(s-r-1)!(\nu-s)!} f_Y(y_1) f_Y(y_2),$$

for $y_1, y_2 \in R, y_1 > y_2$. By assuming that $F_Z = F_Y$ it can be verified that

$$dG(p_1, p_2) = d(F_{U,L}(F_Y^{-1}(1-p_1), F_Y^{-1}(p_2))) = \frac{\nu! p_1^{\nu-s} p_2^{r-1} (1-p_1-p_2)^{s-r-1}}{(r-1)!(s-r-1)!(\nu-s)!} dp_1 dp_2,$$

which implies that the de Finetti's measure is the Dirichlet distribution with parameters $s-r$, $\nu-s+1$ and r , and it does not depend on the particular distribution of Y (or Z).

Define now the following switching rule: If k_1 (resp. k_2) consecutive sample statistics Z_i exceed U (resp. are below L) then the inspection level switches from normal to reduced (resp. to tightened). Thus, the stopping time until a switch in the inspection level occurs is $T = \min\{T_1, T_2\}$, where

$$T_1 = \min\{j : X_{j-k_1+1} = \dots = X_j = 1\}, \quad T_2 = \min\{j : X_{j-k_2+1} = \dots = X_j = 2\}.$$

The stopped sums $S_{T,i} = \sum_{j=1}^T I_{[X_j=i]}, i = 1, 2$, express the number of level 1 and level 2 lots, respectively, until T , and their distributions may provide crucial information about the vendor's manufacturing process. It follows from Proposition 2 that ($m = 2$)

$$\mathbb{E}_e(u^T z_1^{S_{T,1}} z_2^{S_{T,2}}) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{z}, \mathbf{p}}((u((1-p_1-p_2) + p_1 z_1 + p_2 z_2))^T) dG(\mathbf{p}),$$

where $\mathbf{z} = (1, z_1, z_2)$ and $\mathcal{D} = \{\mathbf{p} \in [0, 1]^2 : p_1 + p_2 \leq 1\}$. Note that, under $\mathbb{P}_{\mathbf{z}, \mathbf{p}}$ the X_i 's are iid trinary r.v.'s ($m = 2$) with pdf given by (2.1).

For a sequence of iid $\{0, 1, 2\}$ -valued r.v.'s with respective probabilities $\rho_j, j = 0, 1, 2$, the pgf $\mathbb{E}_{\mathbf{1}, \boldsymbol{\rho}}(w^T), \boldsymbol{\rho} = (1 - \rho_1 - \rho_2, \rho_1, \rho_2)$, can be found in Balakrishnan and Koutras (2002), p. 236, and it is given by

$$\mathbb{E}_{\mathbf{1}, \boldsymbol{\rho}}(w^T) = \frac{(\rho_1 w)^{k_1} (1-\rho_1 w)(1-(\rho_2 w)^{k_2}) + (\rho_2 w)^{k_2} (1-\rho_2 w)(1-(\rho_1 w)^{k_1})}{(1-w)(1-(\rho_1 w)^{k_1})(1-(\rho_2 w)^{k_2}) + (\rho_1 w)^{k_1} (1-\rho_1 w)(1-(\rho_2 w)^{k_2}) + (\rho_2 w)^{k_2} (1-\rho_2 w)(1-(\rho_1 w)^{k_1})}.$$

By replacing ρ_1, ρ_2 and w by $(p_1 z_1) / \sum_{j=0}^m p_j z_j, (p_2 z_2) / \sum_{j=0}^m p_j z_j$ and $u \sum_{j=0}^m p_j z_j$, respectively, in the above formula we get

$$h_{\mathbf{p}; u}(\mathbf{z}) = \mathbb{E}_{\mathbf{z}, \mathbf{p}}((u \sum_{j=0}^m p_j z_j)^T) = \frac{a_1^{k_1} (1-a_1)(1-a_2^{k_2}) + a_2^{k_2} (1-a_2)(1-a_1^{k_1})}{(1-u \sum_{j=0}^m p_j z_j)(1-a_1^{k_1})(1-a_2^{k_2}) + a_1^{k_1} (1-a_1)(1-a_2^{k_2}) + a_2^{k_2} (1-a_2)(1-a_1^{k_1})}$$

where $a_1 = p_1 z_1 u, a_2 = p_2 z_2 u, p_0 = 1 - p_1 - p_2$. Therefore,

$$\mathbb{E}_e(u^T z_1^{S_{T,1}} z_2^{S_{T,2}}) = \frac{\nu! \int_0^1 \int_0^{1-p_1} h_{\mathbf{p}; u}(\mathbf{z}) p_1^{\nu-s} p_2^{r-1} (1-p_1-p_2)^{s-r-1} dp_2 dp_1}{(r-1)!(s-r-1)!(\nu-s)!}.$$

The pgf of $S_{T,1}$ (resp. $S_{T,2}$) follows readily from the above expression by setting $u = 1$ and $z_2 = 1$ (resp. $z_1 = 1$). For example, the distribution of $S_{T,1}$ can be numerically evaluated for specific values of the parameters k_1, k_2, ν, r and s with the aid of the relation

$$\mathbb{P}_e(S_{T,1} = i) = \frac{1}{i!} \frac{d^i}{dz_1^i} \mathbb{E}_e(z_1^{S_{T,1}}) \Big|_{z_1=0} = \frac{\nu! \int_0^1 \int_0^{1-p_1} h_{\mathbf{p}; 1}^{(i)}(0, 1) p_1^{\nu-s} p_2^{r-1} (1-p_1-p_2)^{s-r-1} dp_2 dp_1}{(r-1)!(s-r-1)!(\nu-s)! i!},$$

where $h_{\mathbf{p};1}^{(i)}(0,1)$ is $\left. \frac{d^i}{dz_1^i} h_{\mathbf{p};1}(z_1,1) \right|_{z_1=0}$. As an illustration, in Figures 1,2, the pmf of $S_{T,1}$ is given for two sets of values of the parameters ν, s, r, k_1, k_2 . The values of $h_{\mathbf{p};1}^{(i)}(0,1)/i!$ were easily computed using the function `SeriesCoefficient` of the *Wolfram Mathematica* software. The results were verified using Monte Carlo simulation.

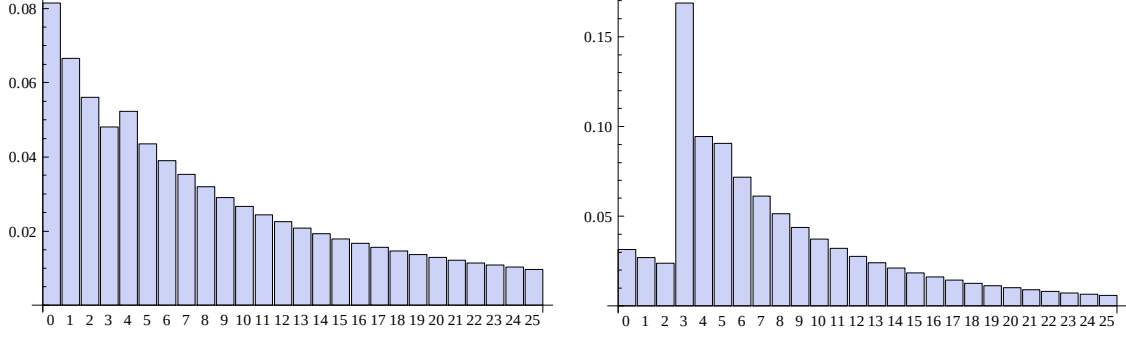


Figure1. $\nu = 20, s = 17, r = 5, k_1 = 4; k_2 = 3$ Figure 2. $\nu = 10, s = 7, r = 3, k_1 = 3; k_2 = 4$
Furthermore, for the above two sets of values of the parameters we have that $\mathbb{E}_e(S_{T,1}) = \int_{\mathcal{D}} p_1 \mathbb{E}_{\mathbf{1},\mathbf{p}}(T) dG(\mathbf{p}) \simeq 23.8989$ and $\mathbb{E}_e(S_{T,1}) \simeq 10.2634$ respectively.

3.2 On the number of coupons of type "i" in a coupon collector's type problem

An urn contains $c_0 + c_1 + \dots + c_m = c$ balls (or coupons) of which c_i bear the number i (are of type "i"), $i = 0, 1, \dots, m$. A ball (or coupon) is drawn at random from the urn, its number (type) is recorded, and it is replaced into the urn together with s ($s = 0, 1, \dots$) balls bearing the same number. This procedure is repeated until each of the $m+1$ numbers appear at least once (i.e. all types of coupons are collected) and denote by T the corresponding stopping time. Let X_n ($n \geq 1$) be a r.v. denoting the type of coupon recorded at the n -th drawing. When $s = 0$, the r.v.'s $X_1, X_2, \dots$ are iid with $p_i = \mathbb{P}_{\mathbf{1},\mathbf{p}}(X_n = i) = c_i/c$, $i = 0, 1, \dots, m$, and the pgf of T is given by

$$\mathbb{E}_{\mathbf{1},\mathbf{p}}(w^T) = 1 + (-1)^{m+1}(1-w) + \sum_{j=0}^{m-1} \left(\sum_{0 \leq i_0 < i_1 < \dots < i_j \leq m} \frac{(-1)^{m-j}(1-w)}{1 - (p_{i_0} + \dots + p_{i_j})w} \right),$$

(see, Inoue and Aki (2007)). However, when $s > 0$, $X_1, X_2, \dots$ becomes a sequence of exchangeable r.v.'s. In this case, it is known that the corresponding de Finetti's measure with cdf G is the Dirichlet distribution (see Eryilmaz (2008a)), with

$$dG(\mathbf{p}) = \frac{\Gamma(\sum_{i=0}^m a_i)}{\prod_{i=0}^m \Gamma(a_i)} \prod_{i=0}^m p_i^{a_i-1} d\mathbf{p},$$

where $a_i = c_i/s$, $i = 0, 1, \dots, m$. It is worth mentioning that under exchangeability ($s > 0$) the probability $\mathbb{P}_e(X_n = i)$ remains the same as in the iid case ($s = 0$), that is $\mathbb{P}_e(X_n = i) = c_i/c = a_i/(a_0 + a_1 + \dots + a_m)$, $i = 0, 1, \dots, m$.

Let $S_{T,i}$, $i = 0, 1, \dots, m$, be the number of coupons of type "i" drawn until all types of coupons are collected (i.e. until T). Besides the distribution of T , it would also be of practical importance to have knowledge about the distribution of $S_{T,i}$ in the iid, as well as, in the exchangeable case. This can be accomplished by employing Propositions 1 and 2. More specifically, since $\mathbb{E}_{\mathbf{1},\mathbf{p}}(T) < \infty$ when every $p_i \notin \{0, 1\}$, we have in the exchangeable model ($s > 0$) that,

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \dots z_m^{S_{T,m}}) = \frac{\Gamma(\sum_{i=0}^m a_i)}{\prod_{i=0}^m \Gamma(a_i)} \int_0^1 \dots \int_0^1 \mathbb{E}_{\mathbf{z},\mathbf{p}}((u(\sum_{i=0}^m p_i z_i))^T) \prod_{i=0}^m p_i^{a_i-1} dp_1 \dots dp_m$$

where $p_0 = 1 - \sum_{i=1}^m p_i$, and $\mathbb{E}_{\mathbf{z},\mathbf{p}}((u(\sum_{i=0}^m p_i z_i))^T)$ is equal to $\mathbb{E}_{\mathbf{1},\mathbf{p}}(w^T)$ given above, after replacing p_i with $p_i z_i / \sum_{i=0}^m p_i z_i$, $i = 0, 1, \dots, m$ and w with $u(\sum_{i=0}^m p_i z_i)$.

For illustrative purposes we consider the exchangeable model with $m = 2, c_1 = 1, c_2 = 2, c_3 = 3, s = 1$ and the corresponding iid model with $s = 0$ (i.e. $p_0 = \frac{1}{6}, p_1 = \frac{1}{3}, p_2 = \frac{1}{2}$). Figures 1,2 depict the pmfs of $T, S_{T,2}$ for both models (dots: iid, boxes: exchangeable case).

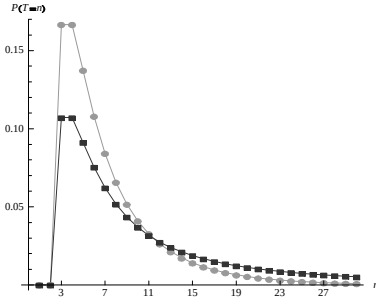


Figure 1 The pmf of T

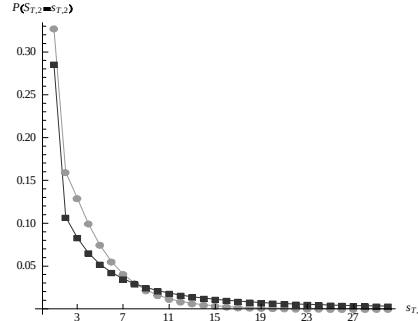


Figure 2 The pmf of $S_{T,2}$

References

- [1] AKI, S. AND HIRANO, K. (1994). Distributions of numbers of failures and successes until the first consecutive k successes, *Ann. Inst. Statist. Math.* **46**, 193-202.
- [2] ANTZOULAKOS, D. L. AND BOUTSIKAS, M. (2007). A direct method to obtain the joint distribution of successes, failures and patterns in enumeration problems. *Statist. Prob. Lett.* **77**, 32-39.
- [3] BALAKRISHNAN, N. (1997). Joint distributions of numbers of success-runs and failures until the first consecutive k successes in a binary sequence. *Ann. Inst. Statist. Math.* **49**, 519-529.
- [4] BALAKRISHNAN, N. AND KOUTRAS, M. V. (2002). *Runs, Scans and Applications.*, John Wiley & Sons.

- [5] BOUTSIKAS, M., RAKITZIS, A. C. AND ANTZOULAKOS, D. L. (2011). On the relation between the distributions of stopping time and stopped sum with applications, *arXiv:1008.0116v2*.
- [6] CHADJICONSTANTINIDIS, S., ANTZOULAKOS, D. L. AND KOUTRAS, M. V. (2000). Joint distributions of successes, failures and patterns in enumeration problems. *Adv. Appl. Prob.* **32**, 866-884.
- [7] ERYILMAZ, S. (2008a). Distribution of runs in a sequence of exchangeable multistate trials. *Statist. Prob. Lett.* **78**, 1505-1513.
- [8] ERYILMAZ, S. (2008b). Run statistics defined on the multicolor urn model. *J. Appl. Prob.* **45**, 1007-1023.
- [9] ERYILMAZ, S. (2010). Discrete scan statistics generated by exchangeable binary trials. *J. Appl. Prob.* **47**, 1084-1092.
- [10] INOUE, K. (2004). Joint distributions associated with patterns, successes and failures in a sequence of multi-state trials, *Ann. Inst. Statist. Math.* **56**, 143-168.
- [11] INOUE, K. AND AKI, S. (2007). On generating functions of waiting times and number of occurrences of compound patterns in a sequence of multistate trials. *J. Appl. Prob.* **44**, 71-81.
- [12] INOUE, K., AKI, S. AND HIRANO, K. (2011). Distributions of simple patterns in some kinds of exchangeable sequences. *J. Statist. Planning Infer.* **141**, 2532-2544.
- [13] KARLIN, S. AND TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*, (2nd Ed), Academic Press.
- [14] LOU, W. Y. W. and FU, J. C. (2009). On probabilities for complex switching rules in sampling inspection. In *Scan Statistics: Methods and Applications*. eds. J. Glaz, V. Pozdnyakov and S. Wallenstein. Birkhauser, Boston, 203-219.
- [15] MAKRI, F. S. AND PSILLAKIS, Z. M. (2011). On success runs of length exceeded a threshold. *Methodology Comput. Appl. Prob.* **13**, 269-305.
- [16] MAULDIN, R., SUDDERTH, W. D. AND WILLIAMS, S. C. (1992). Polya trees and random distributions. *Ann. Statist.* **20**, 1203-1221.
- [17] UCHIDA, M. (1998). On generating functions of waiting time problems for sequence patterns of discrete random variables. *Ann. Inst. Statist. Math.* **50**, 655-671.