J. Appl. Prob. **51,** 483–491 (2014)
Printed in England

ON THE JOINT DISTRIBUTION OF STOPPING TIMES AND STOPPED SUMS IN MULTISTATE EXCHANGEABLE TRIALS

M. V. BOUTSIKAS * ** AND
D. L. ANTZOULAKOS,* *** *University of Piraeus*
A. C. RAKITZIS,**** *University of Cyprus*

Abstract

Let T be a stopping time associated with a sequence of independent and identically distributed 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 $X_1, \dots, X_T$, $0 \le i \le m$. In this paper 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.

Keywords: Stopping time; stopped sum; exchangeability; multistate trial; run and scan statistics; acceptance sampling; coupon collector's problem

2010 Mathematics Subject Classification: Primary 60E05; 62E15
Secondary 60G09; 60G50

1. Introduction

The study of the distributional properties of *stopping times* T and *stopped sums* of the form $S_T = \sum_{i=1}^{N} X_i$, defined over sequences of random variables (RVs) $X_1, X_2, \dots$, 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 those 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. (cf., e.g. 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, e.g. Eryilmaz (2008a), (2008b), (2010), Inoue *et al.* (2011), and 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 study these kind of problems, examined the joint distribution of (T, S_T), where T is the waiting time until the first occurrence of a success run of length k and S_T is the total number of successes until T, in a sequence of binary independent and identically distributed

Received 13 September 2012; revision received 13 March 2013.
* Postal address: Department of Statistics and Insurance Science, University of Piraeus, Piraeus 18534, Greece.
** Email address: mbouts@unipi.gr
*** Email address: dantz@unipi.gr
**** Current address: LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597, France.
Email address: athanasios.rakitzis@univ-nantes.fr

(i.i.d.) trials. Several extensions and variations of their work were subsequently examined in Balakrishnan (1997), Uchida (1998), Chadjiconstantinidis *et al.* (2000), and Inoue (2004).

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 for the case in which T is the waiting time of the rth occurrence of a pattern $\mathcal{E}$ in a sequence of binary i.i.d. trials. In this paper we generalize and extend their work by considering general stopping times defined in sequences of multistate i.i.d. 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 i.i.d. or exchangeable RVs taking values in $\{0, 1, \dots, m\}$. Finally, we present two applications related to quality control and to the coupon collector's problem.

2. Main results

2.1. Independent trials

Let $X_1, X_2, \dots$ be a sequence of RVs taking values in $\{0, 1, \dots, m\}$. Also, let 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 $\boldsymbol{X}_n = (X_1, \dots, X_n)$ (i.e. $[T = n] \in \sigma(\boldsymbol{X}_n)$). Furthermore, let $S_{n,i}$ be the stopped sum denoting the total number of outcomes of type 'i' in $\boldsymbol{X}_n$, namely,

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

where $I_A = 1$ or 0 depending on whether or not A occurs. The above are defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{z,\boldsymbol{p}})$, where the probability measure $\mathbb{P}_{z,\boldsymbol{p}}$, is such that $X_1, X_2, \dots$ are i.i.d. RVs with

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

where $\boldsymbol{z} = (z_0, z_1, \dots, z_m)$, $z_i > 0$, $\boldsymbol{p} = (p_1, \dots, p_m)$, and $p_0 = 1 - \sum_{i=1}^{m} p_i$, $p_i \in (0, 1)$. When $\boldsymbol{z} = \mathbf{1} = (1, \dots, 1)$, the RVs $X_1, X_2, \dots$ are i.i.d. with $\mathbb{P}_{\mathbf{1},\boldsymbol{p}}(X_n = i) = p_i$, $i = 0, 1, \dots, m$, $n \geq 1$. We denote by $\mathbb{E}_{z,\boldsymbol{p}}(\cdot)$ the expectation taken under the probability measure $\mathbb{P}_{z,\boldsymbol{p}}$. In the rest of this section we assume that $\mathbb{P}_{z,\boldsymbol{p}}(T < \infty) = 1$ for all positive real numbers $z_0, z_1, \dots, z_m$ in a neighborhood of 0 (cf. also Remark 2). In Proposition 1 below we express the joint PGF of $(T, \boldsymbol{S}_T) = (T, S_{T,0}, S_{T,1}, \dots, S_{T,m})$ under $\mathbb{P}_{\mathbf{1},\boldsymbol{p}}$, in terms of the PGF of T under $\mathbb{P}_{z,\boldsymbol{p}}$.

Proposition 1. *The joint PGF of $(T, \boldsymbol{S}_T)$ for i.i.d. trials satisfies the relation*

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

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

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

$$E_n(\boldsymbol{s}_n) = \left\{\boldsymbol{x} \in \{0, 1, \dots, m\}^n \colon \sum_{i=1}^{n} (I_{\{x_i = 0\}}, \dots, I_{\{x_i = m\}}) = \boldsymbol{s}_n\right\},$$

https://doi.org/10.1239/jap/1402578638 Published online by Cambridge University Press

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

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

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

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{1},\boldsymbol{p}}(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}} I_{\{T<\infty\}}) &= \sum_{n=1}^{\infty} \sum_{\boldsymbol{s}_n} \mathbb{P}_{\boldsymbol{1},\boldsymbol{p}}(T = n,\ \boldsymbol{S}_n = \boldsymbol{s}_n) u^n \prod_{i=0}^m z_i^{s_{n,i}} \\
&= \sum_{n=1}^{\infty} \sum_{\boldsymbol{s}_n} a(n, \boldsymbol{s}_n) u^n \prod_{i=0}^m (p_i z_i)^{s_{n,i}} \\
&= \sum_{n=1}^{\infty} \biggl(u \sum_{i=0}^m p_i z_i\biggr)^n \sum_{\boldsymbol{s}_n} a(n, \boldsymbol{s}_n) \prod_{i=0}^m \biggl(\frac{p_i z_i}{\sum_{i=0}^m p_i z_i}\biggr)^{s_{n,i}} \\
&= \sum_{n=1}^{\infty} \biggl(u \sum_{i=0}^m p_i z_i\biggr)^n \sum_{\boldsymbol{s}_n} \mathbb{P}_{z,\boldsymbol{p}}(T = n,\ \boldsymbol{S}_n = \boldsymbol{s}_n) \\
&= \sum_{n=1}^{\infty} \biggl(u \sum_{i=0}^m p_i z_i\biggr)^n \mathbb{P}_{z,\boldsymbol{p}}(T = n) \\
&= \mathbb{E}_{z,\boldsymbol{p}}\biggl(\biggl(u \sum_{i=0}^m p_i z_i\biggr)^T I_{\{T<\infty\}}\biggr).
\end{aligned}$$

Taking into account the fact that $\mathbb{P}_{z,\boldsymbol{p}}(T < \infty) = 1$ completes the proof.

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

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{1},\boldsymbol{p}}(u^T z^{S_T}) &= \mathbb{E}_{\boldsymbol{1},\boldsymbol{p}}(u^T z^{0S_{T,0}} z^{1S_{T,1}} \cdots z^{mS_{T,m}}) \\
&= \mathbb{E}_{\tilde{z},\boldsymbol{p}}\biggl(\biggl(u \sum_{i=0}^m p_i z^i\biggr)^T\biggr) \\
&= \mathbb{E}_{\tilde{z},\boldsymbol{p}}((u\mathbb{E}_{\boldsymbol{1},\boldsymbol{p}}(z^{X_1}))^T),
\end{aligned}$$

where $\tilde{z} = (z^0, z^1, \dots, z^m)$. The above relation was proved in Antzoulakos and Boutsikas (2007) for i.i.d. $\{0, 1\}$-valued RVs (when T denotes a specific waiting time) and in Boutsikas *et al.* (2011) for real-valued i.i.d. RVs via a different methodology.

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

$$\mathbb{E}_{\boldsymbol{1},\boldsymbol{p}}\biggl(\biggl(\sum_{i=0}^m p_i z_i\biggr)^{-T} z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}}\biggr) = 1, \tag{2}$$

which can be considered to be an extended version of Wald's identity (see, e.g. Karlin and Taylor (1975, p. 264)). Moreover, differentiating both sides of (2) with respect to z_i and evaluating the resulting expression at $\boldsymbol{z} = \mathbf{1}$, we obtain $\mathbb{E}_{\mathbf{1},\boldsymbol{p}}(S_{T,i}) = p_i \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T)$, $i = 0, 1, \dots, m$, provided that $\mathbb{P}_{\boldsymbol{z},\boldsymbol{p}}(T < \infty) = 1$ for all $\boldsymbol{z}$ in a subinterval of $\mathbb{R}_+^m$ containing $\mathbf{1}$. Similarly, taking second-order derivatives, we obtain the variance and the covariance of the RVs $W_i = S_{T,i} - p_i T$, namely,

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

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

2.2. Exchangeable trials

In this subsection we consider the case in which $X_1, X_2, \dots$ 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 notation 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 $\boldsymbol{X}_n$. We again consider $\mathbb{P}_{\boldsymbol{z},\boldsymbol{p}}$ under which (1) holds . It is worth stressing that the RVs $X_1, X_2, \dots$ are *exchangeable* under $\mathbb{P}_e$, whereas they are *independent* under $\mathbb{P}_{\boldsymbol{z},\boldsymbol{p}}$. Consider the simplex $\mathcal{D} = \{\boldsymbol{p} \in [0, 1]^m : p_1 + p_2 + \dots + p_m \leq 1\}$. From de Finetti's representation theorem for discrete RVs (see, e.g. Mauldin *et al.* (1992)), there exists a probability measure on $\mathcal{D}$ with cumulative distribution function (CDF) G such that

$$\mathbb{P}_e(\boldsymbol{X}_n = \boldsymbol{x}) = \int_{\mathcal{D}} \prod_{i=0}^{m} p_i^{r_{n,i}(\boldsymbol{x})} \, \mathrm{d}G(\boldsymbol{p}), \qquad n \geq 1,$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$ and $r_{n,i}(\boldsymbol{x}) = \sum_{j=1}^{n} I_{\{x_j = i\}}$, $i = 0, 1, \dots, m$. 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 Proposition 2 below we express the joint PGF of $(T, \boldsymbol{S}_T)$ 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 $\mathbb{P}_{\boldsymbol{z},\boldsymbol{p}}$. We again assume that $\mathbb{P}_e(T < \infty) = \mathbb{P}_{\boldsymbol{z},\boldsymbol{p}}(T < \infty) = 1$, $\boldsymbol{p} \in \mathcal{D}$.

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

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}}) = \int_{\mathcal{D}} \mathbb{E}_{\boldsymbol{z},\boldsymbol{p}}\biggl(\biggl(u \sum_{i=0}^{m} p_i z_i\biggr)^T\biggr) \mathrm{d}G(\boldsymbol{p})$$

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

Proof. Observe first that if Y is an RV such that $Y I_{\{T=n\}}$ is a $\sigma(\boldsymbol{X}_n)$-measurable RV (i.e. there exists a measurable function $\varphi_n : \mathbb{R}^n \to \mathbb{R}$ such that $Y I_{\{T=n\}} = \varphi_n(\boldsymbol{X}_n)$), then, invoking de Finetti's representation theorem, we have

$$\begin{aligned}\mathbb{E}_e(Y I_{\{T<\infty\}}) &= \sum_{n=1}^{\infty} \mathbb{E}_e(Y I_{[T=n]}) \\ &= \sum_{n=1}^{\infty} \sum_{\boldsymbol{x}} \varphi_n(\boldsymbol{x}) \mathbb{P}_e(\boldsymbol{X}_n = \boldsymbol{x})\end{aligned}$$

https://doi.org/10.1239/jap/1402578638 Published online by Cambridge University Press

$$= \sum_{n=1}^{\infty} \sum_{\boldsymbol{x}} \varphi_n(\boldsymbol{x}) \int_{\mathcal{D}} \prod_{i=0}^{m} p_i^{r_{n,i}(\boldsymbol{x})} \, \mathrm{d}G(\boldsymbol{p})$$

$$= \sum_{n=1}^{\infty} \int_{\mathcal{D}} \sum_{\boldsymbol{x}} \varphi_n(\boldsymbol{x}) \mathbb{P}_{\mathbf{1},\boldsymbol{p}}(\boldsymbol{X}_n = \boldsymbol{x}) \, \mathrm{d}G(\boldsymbol{p})$$

$$= \sum_{n=1}^{\infty} \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(Y I_{\{T=n\}}) \, \mathrm{d}G(\boldsymbol{p})$$

$$= \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(Y I_{\{T<\infty\}}) \, \mathrm{d}G(\boldsymbol{p}).$$

Note that, under $\mathbb{P}_{\mathbf{1},\boldsymbol{p}}$, the RVs $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{P}_{\mathbf{1},\boldsymbol{p}}(X_j = i) = p_i$, $i = 0, 1, \ldots, m$. If we take $Y = u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}}$ then $Y I_{\{T=n\}}$ is a $\sigma(\boldsymbol{X}_n)$-measurable RV, and, therefore,

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}} I_{\{T<\infty\}}) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}} I_{\{T<\infty\}}) \, \mathrm{d}G(\boldsymbol{p}).$$

Applying Proposition 1 we finally deduce that

$$\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}} I_{\{T<\infty\}}) = \int_{\mathcal{D}} \mathbb{E}_{z,\boldsymbol{p}}\left(\left(u \sum_{i=0}^{m} p_i z_i\right)^T I_{\{T<\infty\}}\right) \mathrm{d}G(\boldsymbol{p}).$$

Taking into account the fact that $\mathbb{P}_e(T < \infty) = \mathbb{P}_{z,\boldsymbol{p}}(T < \infty) = 1$ completes the proof.

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

Remark 4. Recall that if Y is an RV such that $Y I_{\{T=n\}}$ is $\sigma(\boldsymbol{X}_n)$-measurable, and $T < \infty$ almost surely, then $\mathbb{E}_e(Y) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(Y) \, \mathrm{d}G(\boldsymbol{p})$ (see the proof of Proposition 2). Hence, choosing $Y = S_{T,i}$ and taking into account the fact that $\mathbb{E}_{\mathbf{1},\boldsymbol{p}}(S_{T,i}) = p_i \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T)$ (see Remark 2), we deduce that

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

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

$$\mathbb{V}_e(W_i) = \int_{\mathcal{D}} p_i(1-p_i) \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T) \, \mathrm{d}G(\boldsymbol{p}), \qquad \mathbb{C}_e(W_i, W_j) = -\int_{\mathcal{D}} p_i p_j \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T) \, \mathrm{d}G(\boldsymbol{p}),$$

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

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

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

where $S_T = \sum_{j=1}^{T} X_j$, $\tilde{z} = (z^0, z^1, \ldots, z^m)$. Moreover,

$$\mathbb{E}_e(S_T) = \int_{\mathcal{D}} \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(X_1) \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T) \, \mathrm{d}G(\boldsymbol{p}).$$

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 RV) a random sample is taken from each lot to assess its quality. The sampling procedure usually starts with a '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 to a 'tightened' or 'reduced' inspection level is set according to whether the quality of the products has deteriorated or improved, respectively. Traditionally, each inspected lot is classified as either 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, $1 \le t \le \nu$, be the statistic of interest of the tth 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 ith sample, $i = 1, 2, \dots$. The ith lot is classified at level 0, 1, or 2 according to whether $L < Z_i < U$, $Z_i \ge U$, or $Z_i \le L$, respectively.

For $i = 1, 2, \dots$, let $X_i = 1$, 0, or 2 when $Z_i \ge U$, $L < Z_i < U$, or $Z_i \le L$, respectively. As usual, denote by F_{V_1,V_2} and f_{V_1,V_2} the joint CDF and the joint probability density function (PDF) of the RVs V_1 and V_2, respectively, and by F_V and f_V the CDF and the PDF of an RV V. It is now easy to check that, for every $n \ge 1$ and $(x_1, \dots, x_n) \in \{0, 1, 2\}^n$,

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

where $c_1 = \sum_{i=1}^n I_{\{x_i=1\}}$, $c_2 = \sum_{i=1}^n I_{\{x_i=2\}}$, and $G(\boldsymbol{p}) = G(p_1, p_2) = F_{U,L}(F_Z^{-1}(1 - p_1), F_Z^{-1}(p_2))$. Hence, the RVs $X_1, X_2, \dots$ form an infinite sequence of trinary exchangeable RVs with mixing measure $G(\boldsymbol{p})$. Typical random thresholds L and U used in applications are the rth and the sth $(1 \le r < s \le \nu)$ smallest order statistics of the reference sample, 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 $y_1, y_2 \in R$, $y_1 > y_2$. Assuming that $F_Z = F_Y$, it can be verified that

$$
\begin{aligned}
\mathrm{d}G(p_1, p_2) &= \mathrm{d}(F_{U,L}(F_Y^{-1}(1 - p_1), F_Y^{-1}(p_2))) \\
&= \frac{\nu! \, p_1^{\nu-s} p_2^{r-1} (1 - p_1 - p_2)^{s-r-1}}{(r-1)! \, (s-r-1)! \, (\nu-s)!} \, \mathrm{d}p_1 \, \mathrm{d}p_2,
\end{aligned}
$$

which implies that 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).

https://doi.org/10.1239/jap/1402578638 Published online by Cambridge University Press

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

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

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 ($m = 2$)

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

where $\boldsymbol{z} = (1, z_1, z_2)$ and $\mathcal{D} = \{\boldsymbol{p} \in [0, 1]^2 : p_1 + p_2 \leq 1\}$. Note that, under $\mathbb{P}_{z,\boldsymbol{p}}$, the X_i are i.i.d. trinary RVs ($m = 2$) with PDF given by (1).

For a sequence of i.i.d. $\{0, 1, 2\}$-valued RVs with respective probabilities ρ_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):

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

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 obtain

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

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

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

The PGFs of $S_{T,1}$ and $S_{T,2}$ readily follow from the above expression by setting $u = 1$ with $z_2 = 1$ and $z_1 = 1$, respectively. 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

$$\begin{aligned} \mathbb{P}_e(S_{T,1} = i) &= \frac{1}{i!} \frac{\mathrm{d}^i}{\mathrm{d}z_1^i} \mathbb{E}_e(z_1^{S_{T,1}}) \bigg|_{z_1=0} \\ &= \frac{\nu! \int_0^1 \int_0^{1-p_1} h^{(i)}_{\boldsymbol{p};1}(0, 1) p_1^{\nu-s} p_2^{r-1} (1 - p_1 - p_2)^{s-r-1}\, \mathrm{d}p_2\, \mathrm{d}p_1}{(r-1)!\,(s-r-1)!\,(\nu-s)!\, i!}, \end{aligned}$$

where $h^{(i)}_{\boldsymbol{p};1}(0, 1)$ is $\mathrm{d}^i h_{\boldsymbol{p};1}(z_1, 1)/\mathrm{d}z_1^i|_{z_1=0}$. As an illustration, in Figure 1 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 $h^{(i)}_{\boldsymbol{p};1}(0, 1)/i!$ were easily computed using the `SeriesCoefficient` function from the MATHEMATICA® software package. The results were verified using Monte Carlo simulation.

Furthermore, for the two sets of parameter values considered in Figure 1, we have $\mathbb{E}_e(S_{T,1}) = \int_{\mathcal{D}} p_1 \mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T)\, \mathrm{d}G(\boldsymbol{p}) \simeq 23.8989$ and $\mathbb{E}_e(S_{T,1}) \simeq 10.2634$, respectively.

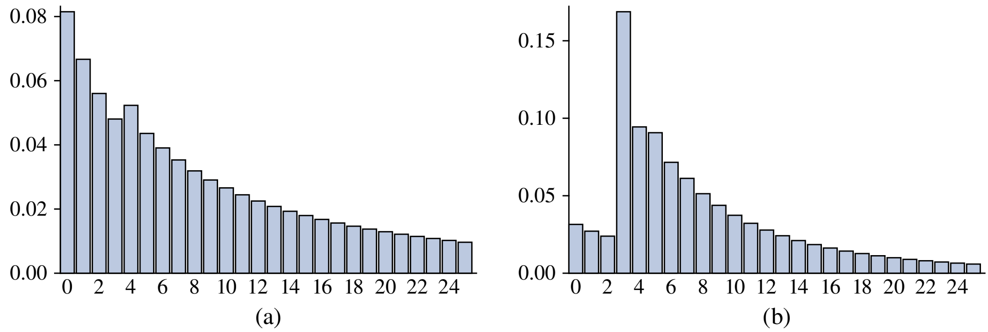

FIGURE 1: The PMF of $S_{T,1}$ for (a) $\nu = 20$, $s = 17$, $r = 5$, $k_1 = 4$, $k_2 = 3$ and (b) $\nu = 10$, $s = 7$, $r = 3$, $k_1 = 3$, $k_2 = 4$.

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

An urn contains $c_0 + c_1 + \cdots + c_m = c$ balls (or coupons) of which c_i bear the number i (are of type 'i'), $i = 0, 1, \ldots, m$. 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 $s\,(s = 0, 1, \ldots)$ 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 coupon are collected), denoting by T the corresponding stopping time. Let X_n ($n \geq 1$) be an RV that denotes the type of coupon recorded at the nth drawing. When $s = 0$, the RVs $X_1, X_2, \ldots$ are i.i.d. with $p_i = \mathbb{P}_{\mathbf{1},\boldsymbol{p}}(X_n = i) = c_i/c$, $i = 0, 1, \ldots, m$, and the PGF of T is given by

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

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

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

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

Let $S_{T,i}$, $i = 0, 1, \ldots, m$, be the number of coupons of type 'i' drawn until all types of coupon 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 i.i.d. case, as well as the exchangeable case. This can be accomplished by employing Propositions 1 and 2. More specifically, since $\mathbb{E}_{\mathbf{1},\boldsymbol{p}}(T) < \infty$ when every $p_i \notin \{0, 1\}$, in the exchangeable model ($s > 0$) we have

$$\begin{aligned}&\mathbb{E}_e(u^T z_0^{S_{T,0}} z_1^{S_{T,1}} \cdots z_m^{S_{T,m}}) \\ &\quad = \frac{\Gamma(\sum_{i=0}^m a_i)}{\prod_{i=0}^m \Gamma(a_i)} \int_0^1 \cdots \int_0^1 \mathbb{E}_{\boldsymbol{z},\boldsymbol{p}}\Bigg(\Bigg(u\Bigg(\sum_{i=0}^m p_i z_i\Bigg)\Bigg)^T\Bigg) \prod_{i=0}^m p_i^{a_i - 1}\,\mathrm{d}p_1 \cdots \mathrm{d}p_m,\end{aligned}$$

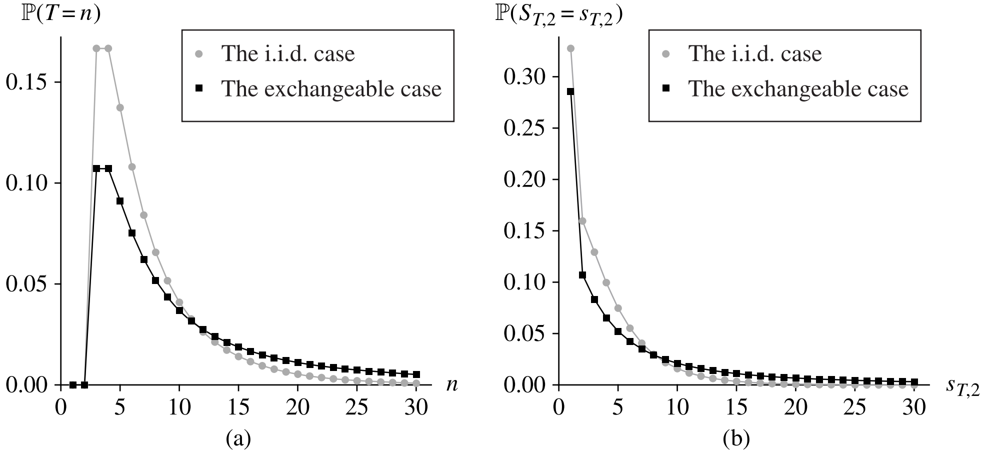

FIGURE 2: (a) The PMF of T. (b) The PMF of $S_{T,2}$.

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

For illustrative purposes, we consider the exchangeable model with $m = 2$, $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, and $s = 1$, and the corresponding i.i.d. model with $s = 0$ (i.e. $p_0 = \frac{1}{6}$, $p_1 = \frac{1}{3}$, and $p_2 = \frac{1}{2}$). In Figure 2 we plot the PMFs of T and $S_{T,2}$ for both models.

References

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.

ANTZOULAKOS, D. L. AND BOUTSIKAS, M. V. (2007). A direct method to obtain the joint distribution of successes, failures and patterns in enumeration problems. *Statist. Prob. Lett.* **77,** 32–39.

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.

BALAKRISHNAN, N. AND KOUTRAS, M. V. (2002). *Runs and Scans with Applications*. John Wiley, New York.

BOUTSIKAS, M. V., RAKITZIS, A. C. AND ANTZOULAKOS, D. L. (2011). On the relation between the distributions of stopping time and stopped sum with applications. Preprint. Available at http://arxiv.org/abs/1008.0116v2.

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.

ERYILMAZ, S. (2008a). Distribution of runs in a sequence of exchangeable multi-state trials. *Statist. Prob. Lett.* **78,** 1505–1513.

ERYILMAZ, S. (2008b). Run statistics defined on the multicolor urn model. *J. Appl. Prob.* **45,** 1007–1023.

ERYILMAZ, S. (2010). Discrete scan statistics generated by exchangeable binary trials. *J. Appl. Prob.* **47,** 1084–1092.

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.

INOUE, K. AND AKI, S. (2007). On generating functions of waiting times and numbers of occurrences of compound patterns in a sequence of multistate trials. *J. Appl. Prob.* **44,** 71–81.

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.

KARLIN, S. AND TAYLOR, H. M. (1975). *A First Course in Stochastic Processes*, 2nd edn. Academic Press, New York.

LOU, W. Y. W. AND FU, J. C. (2009). On probabilities for complex switching rules in sampling inspection. In *Scan Statistics*, eds J. Glaz, V. Pozdnyakov and S. Wallenstein, Birkhäuser, Boston, MA, pp. 203–219.

MAKRI, F. S. AND PSILLAKIS, Z. M. (2011). On success runs of length exceeded a threshold. *Methodology Comput. Appl. Prob.* **13,** 269–305.

MAULDIN, R. D., SUDDERTH, W. D. AND WILLIAMS, S. C. (1992). Pólya trees and random distributions. *Ann. Statist.* **20,** 1203–1221.

UCHIDA, M. (1998). On generating functions of waiting time problems for sequence patterns of discrete random variables. *Ann. Inst. Statist. Math.* **50,** 655–671.