

# The BRS-inequality and its applications

F. Thomas Bruss

*Université Libre de Bruxelles,  
Département de Mathématique,  
Campus Plaine CP 210, B-1050 Brussels, Belgium  
e-mail: [tbruss@ulb.ac.be](mailto:tbruss@ulb.ac.be)*

**Abstract:** This article is a survey of results concerning an inequality, which may be seen as a versatile tool to solve problems in the domain of Applied Probability. The inequality, which we call BRS-inequality, gives a convenient upper bound for the expected maximum number of non-negative random variables one can sum up without exceeding a given upper bound  $s > 0$ . One valuable property of the BRS-inequality is that it is valid without any hypothesis about independence of the random variables. Another welcome feature is that, once one sees that one can use it in a given problem, its application is often straightforward or not very involved.

This survey is focussed, and we hope that it is pleasant and inspiring to read. Focus is easy to achieve, given that the BRS-inequality and its most useful versions can be displayed in five Theorems and their proofs. We try to present these in an appealing way. The objective to be inspiring is harder, and the best we can think of is offering a variety of applications. Our examples include comparisons between sums of i.i.d. versus non-identically distributed and/or dependent random variables, problems of condensing point processes, awkward processes, monotone subsequence problems, knapsack problems, online algorithms, tiling policies, Borel-Cantelli type problems, up to applications in the theory of resource dependent branching processes.

Apart from our wish to present the inequality in an organised way, the motivation for this survey is the hope that interested readers may see potential of the inequality for their own problems.

**MSC2020 subject classifications:** Primary 60-01; secondary 60-02.

**Keywords and phrases:** Stopping time, random sum, Wald’s lemma, optimal selection, Knapsack problem, monotone subsequence, Poisson process, selection bias, general point process, Borel-Cantelli lemma, online selection, tiling, resource dependent branching process.

Received July 2020.

## Contents

1	Introduction . . . . .	45
2	The BRS-inequality . . . . .	46
2.1	The role of the independence hypothesis . . . . .	47
2.2	Generalised BRS-inequality . . . . .	47
2.2.1	Terminology . . . . .	50
2.3	Independence and almost-sure convergence . . . . .	50
2.4	Refining the BRS-inequality . . . . .	51
2.5	Interest of refinements . . . . .	52
2.6	Decreasing order statistics . . . . .	53

2.7	Tractability of improved inequalities . . . . .	54
2.7.1	Asymptotic improvement . . . . .	55
2.8	Interest of further refinements . . . . .	56
2.9	Evaluation of importance . . . . .	57
3	Applications . . . . .	57
3.1	Identically distributed $U[0, 1]$ variables . . . . .	58
3.1.1	Sharpening the upper bound . . . . .	58
3.2	Random variables with dependence . . . . .	58
3.3	Different distribution functions $F_k$ . . . . .	59
3.4	Point processes and selection bias . . . . .	60
3.5	Poisson processes . . . . .	61
3.6	“Awkward” point processes . . . . .	62
3.7	Sequential selection problems . . . . .	63
3.7.1	Knapsack problem . . . . .	63
3.7.2	Monotone subsequence problem . . . . .	64
3.8	Random tilings with different shapes . . . . .	66
3.8.1	Selection and online-selection . . . . .	66
3.8.2	Rectangles and ellipses . . . . .	67
3.9	Borel-Cantelli type problems . . . . .	69
3.9.1	Direct neighbours . . . . .	69
3.9.2	Spacings . . . . .	69
4	Further connections and applications . . . . .	70
4.1	Brief overview . . . . .	70
4.2	Resource dependent branching processes . . . . .	71
4.2.1	Special model assumptions . . . . .	71
4.2.2	Sub-populations . . . . .	72
	Conclusion . . . . .	73
	Acknowledgments . . . . .	73
	References . . . . .	73

## 1. Introduction

Let  $X_1, X_2, \dots$  be a sequence of non-negative random variables (possibly dependent) that are jointly continuously distributed and which all have the same absolutely continuous marginal distribution function  $F$ . For  $n \in \{1, 2, \dots\}$  and  $s \in \mathbb{R}^+$  let  $N(n, s)$  be the maximum *number* of observations among  $X_1, X_2, \dots, X_n$  that one can sum up without exceeding  $s$ . More generally, we will study as well the case where the  $X_k$  are not identically distributed, and where each  $X_k$  has an absolutely continuous marginal distribution function  $F_k$ .

The focus of this paper is to find a good upper bound for the expected value  $E(N(n, s))$ , and to show its versatility for applications. The inequality which yields such an upper bound is what we call the *BRS-inequality*. The reader may find a memory aid for the meaning of BRS by reading it as

Budget-Restrained-Sum.

Indeed, one may interpret  $X_1, X_2, \dots, X_n$  as prices for  $n$  items, all having the same utility, offered for sale, and may ask what is the maximum number of them one can expect to be able to buy with a budget limit  $s$ .

Clearly one cannot do better than looking at the complete list of prices and first select the smallest price, then add the second smallest price, and so on, as long as the current accumulated bill does not exceed  $s$ . Hence we can define  $N(s, n)$  equivalently via the increasing order statistics of  $X_1, X_2, \dots, X_n$ , denoted by  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ , namely

$$N(n, s) = \begin{cases} 0, & \text{if } X_{1,n} > s, \\ \max\{k \in \mathbb{N} : X_{1,n} + X_{2,n} + \dots + X_{k,n} \leq s\}, & \text{otherwise.} \end{cases} \quad (1)$$

Throughout this paper we will call the difference

$$R(n, s) = s - X_{1,n} - X_{2,n} - \dots - X_{N(n,s),n}$$

the *residual* of the budget, which, by definition, equals  $s$  if  $X_{1,n} > s$ . Correspondingly, we call the difference

$$\mathcal{O}(n, s) = X_{1,n} + X_{2,n} + \dots + X_{N(n,s)+1,n} - s,$$

provided that  $N(n, s) < n$ , the *overshoot* of the budget.

## 2. The BRS-inequality

The following gives a simple and useful bound on the expectation of the random variable  $N(n, s)$  defined in (1). Moreover, we add a result of almost-sure convergence for the case that the  $X_k$  are also independent:

**Theorem 1.** *Let all  $X_1, X_2, \dots, X_n$  be identically distributed non-negative random variables with absolutely continuous distribution function  $F$ . Then we have*

**Part (i)**

$$\mathbb{E}(N(n, s)) \leq nF(t), \quad (2)$$

where  $t := t(n, s)$  solves the equation

$$n \int_0^t x dF(x) = s. \quad (3)$$

**Part (ii)**

*If the random variables are moreover independent, and the budget  $s := s_n$  tends to infinity in such a way that  $\lim_{n \rightarrow \infty} n^{-1}s_n$  exists, then*

$$n^{-1}N(n, s_n)/F(t(n, s_n)) \rightarrow 1 \text{ a.s., as } n \rightarrow \infty. \quad (4)$$

The first proofs of Part (i) and Part (ii) of Theorem 1 appeared in Bruss and Robertson (1991). (From now on we abbreviate this reference by B. & R. (1991).) We will use Part (ii) of Theorem 1 later on in several places to strengthen certain results of interest we obtain from Part (i), provided that we also have independence. Our main interest in this survey is however the inequality of Part (i) and its generalisation.

As we shall see in Subsection 2.2, Part (i) was elegantly generalised in Steele (2016) by accommodating non-identically distributed random variables as well. Before we present this, we first clarify when and where an independence hypothesis is needed.

### 2.1. The role of the independence hypothesis

Clearly, the result (4) of almost sure convergence in Part (ii) requires independence of the random variables  $X_1, X_2, \dots, X_n$ , and so do all *theorems* in B. & R. (1991). Since these authors were motivated by a specific problem in the domain of branching processes in which the random variables were naturally i.i.d., they assumed that the  $X_k$  are i.i.d. random variables.

As both referees of the present survey pointed out independently, the present survey should avoid any possible confusion and make clear whether the proof of Part (i) in B. & R. (1991) is complete without the independence hypothesis.

The answer is yes, it is. This is stated and proved in Lemma 4.1 on page 622 of B. & R. (1991). This Lemma is singled out at the end of their article for this very reason, namely, no independence hypothesis is needed. The short proof links for two sets of positive real numbers, their number of elements with the respective sums of the elements in these sets and obtains a central inequality. The second step, which applies this inequality to functions of our random variables, then only uses the linearity of the expectation operator. This requires, of course, no independence, and B. & R. (1991) pointed out in line 10 of their proof of Lemma 4.1 that the inequality is always true.

As section 3 in Steele (2016) shows, Steele had overlooked Lemma 4.1. in B. & R. (1991). We will show after Theorem 2, in Remark 1, that the proof of Lemma 4.1 in B. & R. (1991) and the first part of the proof of Steele (2016), are, apart from notation, identical. This oversight has no consequences whatsoever for the present survey. Moreover, the author highly praises the important contributions of Steele (2016).

### 2.2. Generalised BRS-inequality

**Theorem 2** (Steele (2016)). *Let  $X_1, X_2, \dots, X_n$  be positive random variables that are jointly continuously distributed and such that each  $X_k$  has an absolutely continuous distribution function  $F_k$ . If  $N(n, s)$  is defined as in (1), then*

$$E(N(n, s)) \leq \sum_{k=1}^n F_k(t), \quad (5)$$

where  $t := t(n, s)$  is the unique solution of the equation

$$\sum_{k=1}^n \int_0^t x dF_k(x) = s. \quad (6)$$

Clearly, if all random variables  $X_i$ ,  $i \in \{1, 2, \dots, n\}$  have the same marginal distribution  $F$ , then (6) recaptures (3), and (5) recaptures (2). Hence Theorem 2 is a full-scale generalisation of Theorem 1 (i).

The proofs of Theorem 1 (i) and (ii) and of Theorem 2 are in two different journals. Our first goal is to offer here a single combined proof, including relevant remarks in view of possible applications.

The second objective is to present two new refined BRS-inequalities (Theorem 3 and Theorem 4), and to briefly discuss other possible directions of refinements.

Our third goal is to give several examples from different domains of Applied Probability, showing the versatility of the inequality for applications.

*Proof of Theorem 2.* We start with the proof of Part (i) of Theorem 1, and we do this in a way that leads almost directly to the proof of Theorem 2.

By the continuity of the joint distribution of  $(X_i : 1 \leq i \leq n)$ , there is clearly a unique set of selected indices  $A \subseteq \{1, 2, \dots, n\}$  of the variables that attains the maximum in the definition of  $N(n, s)$  in (1). We denote this subset by  $A(n, s)$ . Note that  $\#A(n, s) = |A(n, s)| = N(n, s)$ .

Further let

$$B(n, s) = \{i \in \{1, 2, \dots, n\} : X_i \leq t(n, s)\}, \quad (7)$$

where  $t(n, s)$  is the threshold value determined by the implicit relation (3).

The idea is now to compare the sets  $A(n, s)$  and  $B(n, s)$ , together with their associated sums,

$$S_{A(n,s)} = \sum_{i \in A(n,s)} X_i \quad \text{and} \quad S_{B(n,s)} = \sum_{i \in B(n,s)} X_i. \quad (8)$$

From these definitions, one immediately gets by definition two useful relations, namely

$$S_{A(n,s)} \leq s \quad \text{and} \quad E[S_{B(n,s)}] = n \int_0^{t(n,s)} x dF(x) = s. \quad (9)$$

By its definition,  $S_{A(n,s)}$  is a partial sum of order statistics. The summands of  $S_{B(n,s)}$  consist precisely of the values  $X_{n,i}$  with  $X_{n,i} \leq t(n, s)$ . We can think of the latter as also being listed in increasing order, of course, so that  $S_{B(n,s)}$  is also equal to a partial sum of order statistics of  $X_1, X_2, \dots, X_n$ . These observations will help us with estimations of the relative sizes of the two sums  $S_{A(n,s)}$  and  $S_{B(n,s)}$ .

First, we note that we have either  $A \subset B$  or  $B \subseteq A$ . More specifically, if  $S_{B(n,s)} \leq S_{A(n,s)}$  then we have  $B(n, s) \subseteq A(n, s)$ . Moreover, the summands  $X_i$

in the difference set, that is, with  $i \in A(n, s) \setminus B(n, s)$ , these summands are all bounded below by  $t(n, s)$ . Hence we have the bound

$$S_{B(n,s)} + t(n, s) (|A(n, s)| - |B(n, s)|) \leq S_{A(n,s)}, \quad \text{if } S_{B(n,s)} \leq S_{A(n,s)}.$$

Similarly, if  $S_{A(n,s)} < S_{B(n,s)}$  then  $A(n, s) \subset B(n, s)$  and the summands  $X_i$  with  $i \in B(n, s) \setminus A(n, s)$  are all bounded above by  $t(n, s)$ ; so, in this case,

$$S_{B(n,s)} \leq S_{A(n,s)} + t(n, s) (|B(n, s)| - |A(n, s)|), \quad \text{if } S_{A(n,s)} < S_{B(n,s)}.$$

Taken together, the last two relations tell us that whatever the relative sizes of  $S_{A(n,s)}$  and  $S_{B(n,s)}$  may be, one always has the *key inequality*

$$t(n, s) (|A(n, s)| - |B(n, s)|) \leq S_{A(n,s)} - S_{B(n,s)}. \quad (10)$$

Here  $t(n, s) > 0$  is a constant. We recall that  $|A(n, s)| = N(n, s)$ , and we see from (9) that the right-hand side has non-positive expectation. Hence, taking the expectations in the key inequality (10) gives us

$$\mathbb{E}(N(n, s)) \leq \mathbb{E}(|B(n, s)|) = \mathbb{E}\left(\sum_{i=1}^n \mathbf{1}\{X_i \leq t(n, s)\}\right) = n F(t(n, s)),$$

and the proof of the upper bound (2) is complete.

Now, to prove its more general form as given in Theorem 2, we again define the set  $B(n, s)$  by (7). Additional care is here needed with the definition of  $A(n, s)$ . (As we shall see in the subsequent Remark 2, this care provides at the same time a non-negligible benefit.)

Specifically, we first define a total order on the set  $\{X_i : i \in \{1, 2, \dots, n\}\}$  by writing  $X_i \prec X_j$  if either condition (A) or condition (B) holds, where

- (A)  $X_i < X_j$ ,
- (B)  $X_i = X_j$  and  $i < j$ .

With this order, there is a unique permutation  $\pi : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  such that

$$X_{\pi(1)} \prec X_{\pi(2)} \prec \dots \prec X_{\pi(n)},$$

and we can then take  $A(n, s)$  to be largest set  $A \subset \{1, 2, \dots, n\}$  of the form

$$A = \{\pi(i) : X_{\pi(1)} + X_{\pi(2)} + \dots + X_{\pi(k)} \leq s\}. \quad (11)$$

We can now proceed with the proof of key inequality (10) in the same way.

We use the identical definitions in (7) and (8) for the sums  $S_{A(n,s)}$  and  $S_{B(n,s)}$ , because these definitions are not affected by the fact that the variables may now stem from different distributions. Hence, by the new definition of  $t(n, s)$ , we now have

$$\mathbb{E}(S_{B(n,s)}) = \sum_{i=1}^n \int_0^{t(n,s)} x dF_i(x) = s.$$

Since we still have  $S_{A(n,s)} \leq s$ , the expectation on the right side of inequality (10) is again non-positive.

But then, since all the following arguments in the proof of Part (i) of Theorem 1 are valid independently of the number of different distributions functions  $F_k$  which are involved, the proof of Theorem 2 is also complete.  $\square$

**Remark 1.** As announced in Subsection 2.1, we now show that the proof of Lemma 4.1 of B. & R. (1991) and the first part of the preceding proof of Theorem 2 are, apart from notation, identical. Indeed, inequality (4.1) of B. & R. (1991) reads after multiplication with  $t_n = t(n, s) > 0$ ,

$$t_n(N_n - \hat{N}_{t_n}) \leq \tilde{S}_{N_n} - \hat{S}_{t_n} \leq s - \hat{S}_{t_n}, \quad \forall n = 1, 2, \dots,$$

where  $N_n = N(n, s) = |A(n, s)|$ ,  $\hat{N}_{t_n} = |B(n, s)|$ ,  $\tilde{S}_{N_n} = S_{A(n,s)}$  and  $\hat{S}_{t_n} = S_{B(n,s)}$ . Therefore

$$t(n, s)(|A(n, s)| - |B(n, s)|) \leq S_{A(n,s)} - S_{B(n,s)},$$

which is the key inequality (10). It suffices to take expectations to obtain the result.

**Remark 2.** Steele's insight (Steele (2016)) that this full-scale generalisation of the Bruss-Robertson inequality should be true, was remarkable. Also, his introduction of the total order induced by (A) and (B) above was a powerful trick, since  $X_j$  and  $X_k$  with absolute continuous distributions may now coincide for  $j \neq k$  with positive probability. The inequality gains additional interest, knowing that, here again, no independence is needed.

### 2.2.1. Terminology

As we have seen, none of the required calculations require more on the joint distribution of the variables  $X_i : i \in \{1, 2, \dots, n\}$  than the joint continuity. The argument just uses point-wise bounds and the linearity of expectation. In the following we will therefore call invariably (2), respectively (5), the BRS-inequality, and (3), respectively (6), the corresponding BRS-equation. It will always be evident from the context of the problem when (6) and (5) are meant. Note also that the notions of budget-residual, respectively, budget-overshoot, carry over unchanged for Theorem 2.

As announced already, we will also present (Subsections 2.4 and 2.5) two refinements of Theorem 2. These refinements are in the same spirit as Theorem 2, and, in particular, they do not require to revise the definitions in any way. Therefore we may and will call them, without a risk of any confusion, also BRS-inequalities.

### 2.3. Independence and almost-sure convergence

The asymptotic result (4) is proved in section 2 of B. & R. (1991). Note that the general setting of Theorem 2 does not allow to include (4) directly, where

independence of the variables is needed. The convergence in (4) in Part (ii) has its interest on its own as we shall see later on in several examples. Therefore we recall here the proof, but in a slightly different way, namely in terms which make the result very intuitive:

*Proof of Theorem 1, Part (ii).* First we note from the definition of  $N(n, s)$  in (1) that, although the random time  $N(n, s)$  is not a stopping time with respect to the usual filtration generated by the random variables  $X_j$ , the random time  $N(n, s) + 1$  is a stopping time on the filtration  $(\tilde{\mathcal{F}}_k)$ , where

$$\tilde{\mathcal{F}}_k = \sigma(X_{1,n}, X_{2,n}, \dots, X_{k,n}), \quad k = 1, 2, \dots, n.$$

If we sum up all those observations which are smaller than  $t_n$ , say, then Wald's equality can be applied for the expected stopping time since, in Theorem 1, all variables follow the same distribution and thus have the same mean. Moreover, by independence, the unordered variables not exceeding  $t(n, s)$  are continuous i.i.d. random variables on the interval  $[0, t(n, s)]$ . Letting  $n \rightarrow \infty$ , the Strong Law of Large Numbers implies then that the average of these terms satisfies

$$\frac{1}{n} \sum_{k=1}^n X_k \mathbf{1}\{X_k \leq t(n, s)\} \rightarrow \int_0^{t(n, s)} x dF(x) \text{ a.s., as } n \rightarrow \infty,$$

that is, the sum of these selected variables converges almost surely to the truncated integral defined in (3).

Finally, if  $s := s_n$  grows with  $n$ , and if we let the threshold become  $t := t(n, s_n)$  then the same argument remains valid provided that  $s_n/n$  tends to a limit. Then (3) yields (4), and the proof is complete.  $\square$

#### 2.4. Refining the BRS-inequality

We now refine the BRS-inequality by exploiting more carefully the key inequality (10). This will be rather straightforward, and the reader may ask why we did not do this right away. We will shortly discuss the reason after Theorem 3 and its proof.

**Theorem 3.** *Under the conditions of Theorem 2 we have*

$$\mathbb{E}(N(n, s)) \leq \sum_{k=1}^n F_k(t(n, s)) - \frac{s - \mathbb{E}(S_{A(n, s)})}{t(n, s)}, \quad (12)$$

where  $t(n, s)$  satisfies (6), and where  $S_{A(n, s)}$  is the sum defined in (8).

*Proof.* The key inequality (10) can be written in the form

$$|B(n, s)| - |A(n, s)| \geq \frac{S_{B(n, s)} - S_{A(n, s)}}{t(n, s)}, \quad (13)$$



where  $t(n, s)$  solves the BRS-equation (6). Recall again that  $S_{A(n,s)}$  defined in (8) is the sum of the  $N(n, s)$  smallest order statistics, and thus  $|A(n, s)| = N(n, s)$ . Further, since  $S_{B(n,s)}$  is the sum of all those variables not exceeding  $t(n, s)$ , we have for the number of terms in  $B(n, s)$

$$|B(n, s)| = \sum_{k=1}^n \mathbf{1}\{X_k \leq t(n, s)\} \text{ and } E(S_{B(n,s)}) = s,$$

where the second equality was shown in (9) for identically distributed random variables. For different distribution functions it follows correspondingly from the arguments given at the end of the proof of Theorem 2.

Using these equalities we obtain from (13) by taking expectations on both sides

$$\sum_{k=1}^n F_k(t(n, s)) - E(N(n, s)) \geq \frac{s - E(S_{A(n,s)})}{t(n, s)}. \quad (14)$$

Subtracting in the last inequality the sum term on both sides, and then changing signs, yields the proof of the statement of Theorem 3.  $\square$

Now we should motivate our order of the theorems. Theorem 1 Part (i) as well as Theorem 2 are corollaries of Theorem 3. Hence, following the tradition in Mathematics, Theorem 3 should be entitled to be stated and proved first. Note however that it would have been necessary to explain already in the hypotheses for the inequality the meaning of the sets  $A(n, s)$  and  $B(n, s)$ , and also how the expected residual sum intervenes. This may distract from the main interest of the inequality, and the reader may agree with our preferred order of the theorems.

### 2.5. Interest of refinements

Theorem 3 is of interest as soon as we have additional information.

Recall that  $s - S_{A(n,s)}$  is the *residual* of the available budget. If we have sufficient information to give easy bounds for its expectation  $E(s - S_{A(n,s)})$ , then Theorem 3 may give us a convenient more precise bound. This is (independently) true if  $s$  and  $n$  are such that  $P(N(n, s) < n)$  is very close to 1 so that the  $S_{A(n,s)} + X_{N(n,s)+1,n} - s$  represents the very probable positive *overshoot*. We mention that, in specific examples, such as for instance when all  $F_k$ 's have similar bounded supports, Theorem 3 help us to understand why the bounds we obtain are sometimes tight.

The latter type of arguments allows us also to defend the attitude that any strengthening of the BRS-inequality is always of interest, at least in some special cases.

Using standard arguments about conditional expectations, we obtain a version of the BRS-inequality which, depending on properties of the distributions, may have advantages. It hardly earns the name "Theorem" since it is almost

a corollary of Theorem 3, but, to be consistent on the list of BRS-theorems, it should also be called Theorem.

**Theorem 4.** Let  $N(n, s)$  be defined as in (1), and  $t(n, s)$  as in (6). Also recall the set  $A(n, s)$  defined by (8), and also that  $E(S_{A(n,s)}) \leq s$ . Denoting  $p_n = P(N(n, s) < n)$  we have

$$E(N(n, s)) \leq p_n \left( \sum_{k=1}^n F_k(t(n, s)) - \frac{s - E(S_{A(n,s)})}{t(n, s)} \right) + n(1 - p_n).$$

*Proof.* We first note the evident equality  $E(N(n, s) | N(n, s) \geq n) = n$ . Using it in expressing  $E(N(n, s))$  by conditioning on the event  $\{N(n, s) < n\}$  and its complement, we obtain

$$E(N(n, s)) = n(1 - p_n) + p_n E(N(n, s) | N(n, s) < n).$$

The statement of Theorem 4 follows now in a straightforward manner from the trivial inequality

$$E(N(n, s) | 0 \leq N(n, s) < n) \leq E(N(n, s))$$

by plugging in the upper bound for  $E(N(n, s))$  of Theorem 3, and then isolating  $E(N(n, s))$  again on the left-hand side.  $\square$

## 2.6. Decreasing order statistics

So far we have studied the sum of increasing order statistics  $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$  of the random variables  $X_1, X_2, \dots, X_n$  by thinking at the same time of the stopping time  $N(n, s) + 1$  when the sum of prices exceeds the budget  $s$ . What can we say about  $E(M(N, s))$  where

$$M(n, s) = \begin{cases} 0, & \text{if } X_{n,n} > s, \\ \max\{k \in \mathbb{N} : X_{n,n} + X_{n-1,n} + \dots + X_{n-k+1,n} \leq s\}, & \text{if } X_{n,n} \leq s, \end{cases}$$

i.e.  $E(M(n, s))$  is the maximum expected number of items one can buy with a fixed budget  $s$  by buying in decreasing order always the most expensive item?

Given that all items are supposed to have the same utility 1, the problem of finding  $E(M(n, s))$  is neither dual to finding  $E(N(n, s))$ , nor is it, a priori, a natural problem in our context. It looks like the problem of a maniac money-waster who wants to spend his or her budget as quickly as possible with minimum total utility.

As occurs often, it is the context which makes the difference. The problem makes perfect sense if we interpret *budget* and *utility* differently, as for example in the context of so-called *Resource dependent branching processes* studied in Subsection 4.2. For the latter it will be helpful to summarise Theorem 1 (ii) and to combine it with another result of B. & R. (1991) on the stopping time for decreasing order statistics.

**Theorem 5.** Suppose that  $F$  is a strictly increasing absolute continuous distribution function on  $[a, b]$  of i.i.d. random variables  $X_1, X_2, \dots, X_n$  with expected value  $E(X) < \infty$  where  $0 \leq a < b \leq \infty$  and  $F(a) = 0, F(b) = 1$ . Further let  $s = \lim_{n \rightarrow \infty} s_n/n$ , and let  $\tau$  and  $\theta$  be defined as the solutions of the respective equations

$$\int_a^\tau x dF(x) = s; \quad \int_\theta^b x dF(x) = s.$$

Then, denoting  $N_n := N(n, s_n)$ ,  $M_n := M(n, s_n)$  and  $\bar{F}(x) = 1 - F(x)$ , we have

$$(i) \quad \lim_{n \rightarrow \infty} X_{N_n} = \tau \text{ a.s.}, \quad \lim_{n \rightarrow \infty} \frac{N_n}{n} = F(\tau) \text{ a.s.}, \quad \lim_{n \rightarrow \infty} \frac{E(N_n)}{n} = F(\tau),$$

If, moreover, the  $X_1, X_2, \dots, X_n$  are bounded (i.e. if we require  $b < \infty$ ), then

$$(ii) \quad \lim_{n \rightarrow \infty} X_{n-M_n+1, n} = \theta \text{ a.s.}, \quad \lim_{n \rightarrow \infty} \frac{E(M_n)}{n} = \bar{F}(\theta), \quad \lim_{n \rightarrow \infty} \frac{M_n}{n} = \bar{F}(\theta) \text{ a.s.}$$

The proof of (i) (which collects the implications of Theorem 1 (ii)) is given in detail on pages 614-165 in B. & R. (1991), and we know that it holds without a constraint of boundedness of the  $X_k$ . The proof of (ii) goes through in an analogous way under the assumption that the  $X_k$  are bounded.

Putting (i) and (ii) in one Theorem shows us the way to quantify (asymptotically) where the sum of increasing order statistics and the sum of decreasing order statistics would meet.

## 2.7. Tractability of improved inequalities

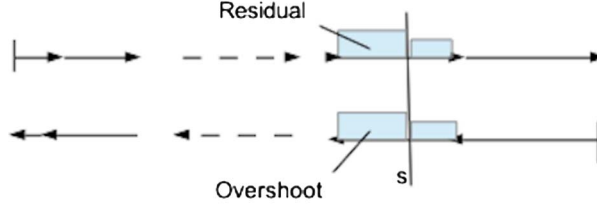
If we denote the rhs-bound of Theorem 2 by  $\alpha(n, s)$  and the rhs-bound of Theorem 3 by  $\beta(n, s)$ , then  $\delta(n, s) = \alpha(n, s) - \beta(n, s)$  is the improvement Theorem 3 gives compared to Theorem 2. According to (5) and (12) we have

$$\delta(n, s) = \frac{s - E(S_{A(n,s)})}{t(n, s)} = \frac{E(R(n, s))}{t(n, s)},$$

where the second equality follows from the definition of the residual  $R(n, s)$  in Section 1, and from the trivial equality  $E(s) = s$ . Also, we recall that  $t(n, s)$  is the solution of the BRS-equation (6). The open part of the problem is thus to find a good estimate of  $E(R(n, s))$ . This is also a part of the problem in applying Theorem 4. Since we have, a priori, for the BRS-inequalities no restriction on dependencies between the random variables  $X_1, X_2, \dots, X_n$ , the generality allows for an uncountable number of different scenarios for the joint distribution of the  $X_1, X_2, \dots, X_n$ . Therefore it seems hard to give a *generally useful* estimate for  $\delta(n, s)$ . Even having complete independence of the  $X_k$  is still of limited value, because their order statistics are no longer independent.

Now, can we find at least for i.i.d. random variables  $X_k, 1 \leq k \leq n$ , bounds of interest which are relatively easily available?

Our answer for this more modest question is yes. Indeed, suppose  $E(X) := E(X_k) < \infty$ . Denoting  $S_n = X_1 + X_2 + \dots + X_n$  we then have  $E(S_n) = nE(X) < \infty$ , and thus  $S_n < \infty$  almost surely. Let us look now at Figure 1. It might remind us of figures we have seen to visualise renewal type arguments (For an overview on renewal theory see e.g. Ross (1983)). We have no renewal process, however, and our argument must be different.



Residual and overshoot in reversed summation

FIG 1. The first line of arrows in Figure 1 displays the successive partial sums of the increasing order statistics until the interval which “covers”  $s$ . The second line starts at the end  $S_n$  and runs backwards with the partial sums of decreasing order statistics.

Suppose first that  $N(n, s) < n$ . Then the point  $s$  is almost surely covered by the interval generated by the order statistic  $X_{N(n,s)+1,n}$ . Now suppose for instance that the  $X_k$  are i.i.d. exponential with parameter 1, say. It is well known that all order statistics of exponentials are again exponential with varying parameters. Then, whatever the parameter of the density, all random variables have the memoryless property. Hence the length of the overshoot to the right of the point  $s$  as seen by an observer (he, say) of the upper line of arrows has the same distribution as  $X_{N(n,s)+1,n}$  (unconditional! on covering  $s$ ). But if another observer (she, say) standing in  $S_n$  looks into the opposite direction (lower line of arrows) then she would see the same covering order statistic after  $n - N(n, s)$  steps. For her the overshoot, now to the left of  $s$ , is the residual in the upper line. Since it is the same covering order statistic interval, his residual in the upper line must follow itself the (unconditional) distribution of  $X_{N(n,s)+1,n}$ . Hence, like in the famous waiting time paradox, since *conditional on covering*  $s$  is intrinsic in the definition of  $N(n, s)$ , we get for the expectation of the latter  $E(X_{N(n,s)+1,n}) = 2E(R(n, s))$ , which means

$$E(R(n, s)) = \frac{1}{2}E(X_{N(n,s)+1,n}).$$

2.7.1. Asymptotic improvement

Now let  $s_n$  be the available budget and suppose that  $s = \lim_{n \rightarrow \infty} s_n/n$  exists with  $0 < s < E(X)$ . Then we have  $P(N(n, s_n) < n) \rightarrow 1$  as  $n \rightarrow \infty$  so that we can confine our interest to the event  $\{N(n, s) < n\}$  we considered above. It follows from the independence hypothesis and from Theorem 5 (i), where

$\tau := t(n, s_n)$  solves the BRS-equation, that  $X_{N(n, s_n)+1, n}/t(n, s_n) \rightarrow 1$  almost surely as  $n \rightarrow \infty$ . Since almost-sure convergence implies convergence in mean, we have

$$\delta(n, s_n) = \frac{E(R(n, s_n))}{t(n, s_n)} \rightarrow \frac{1}{2}, \text{ as } n \rightarrow \infty.$$

The limiting improvement of Theorem 3 compared to Theorem 2 is thus 1/2.

This result is not exciting as an asymptotic improvement, because we have  $nF(t(n, s_n)) \rightarrow \infty$  as  $n \rightarrow \infty$ . However, as we shall see in the Applications (Subsection 3.1), we will not lose much of the improvement for smaller  $n$ , and then it is non-negligible, of course.

Note also that whenever the conditions of Theorem 5 (i) are satisfied, we have

$$0 \leq \lim_{n \rightarrow \infty} \delta(n, s_n) \leq 1$$

because the interval generated by  $X_{N(n, s_n)+1, n}$  always contains the residual, and clearly,  $X_{N(n, s_n)+1, n}/X_{N(n, s_n), n} \rightarrow 1$  a.s. Hence the limiting improvement for i.i.d. random variables is always between 0 and 1.

If we replace in our example the i.i.d. exponentials by i.i.d. uniform  $X_k$  then the limiting improvement is also 1/2. However, the author does not know under which general conditions on the  $X_k$  one can right away prove the existence of a limiting improvement. And even then, any result for i.i.d. random variables is still modest compared to what one would really like to have, namely a general assessment of the added value of Theorem 3 compared to Theorem 2, which both allow for the random variables dependence and arbitrarily many different distribution functions.

## 2.8. Interest of further refinements

It is now no surprise that, in some special cases, other refinements are possible. We want to exemplify another one of such cases, but we confine ourselves to describe things verbally, because it suffices to present ideas which come along quite naturally. There are simply too many special cases which come to our mind.

Consider for instance the case of identically distributed random variables  $X_j \sim X$  (as in Theorem 1). Suppose moreover that  $s := s_n$  is small compared to  $nE(X)$ . Then, for large  $n$ , the probability of not exceeding the budget  $s_n$  becomes small, and so  $p_n = P(N(n, s) < n)$  becomes large as  $n$  grows. Consequently, it will become very probable that the budget will be exceeded, and we know

$$P(X_{N(n, s)+1, n} \in \{X_1, X_2, \dots, X_n\}) \rightarrow 1 \text{ as } n \rightarrow \infty.$$

This guarantees that the sum  $S_{A(n, s_n)}$  behaves essentially like  $s_n$  as  $n$  becomes large. Now recall that, conditioned on  $N(n, s) < 1$ , the random time  $N(n, s) + 1$  is a stopping time with respect to the filtration

$$\tilde{\mathcal{F}}_k = \sigma(X_{1, n}, X_{2, n}, \dots, X_{k, n}), k = 1, 2, \dots, n.$$

Recall also that all  $X_j \sim X$  have the same mean and that the generalised version of Wald's equation works with weaker conditions than independence. Thus, even though we started with a special case and, compared to independence, arguably partially weaker condition that  $s_n$  is small compared to  $nE(X)$ , this may still allow for a residual-type argument as exemplified in Subsection 2.7.

The author thinks that in this survey it would be hardly rewarding to go further with discussions of more specific cases, and that it is preferable to view results which are, arguably, of sufficient general interest. If a researcher would like to apply a result for a given problem, needing slightly more precise estimates, and searching ways to obtain new refinements, then she or he, knowing their own problem best, would probably know best in which corner of the set of options to look for.

### 2.9. Evaluation of importance

As so often in Mathematics, it is not always the very strongest form of a result which draws attention but rather the most convenient form of a sufficiently strong result, which does so. Seen under this angle of view, our analysis of relative importance is as follows.

Theorem 1 (i) and its generalisation Theorem 2 give a convenient upper bound of  $E(N(n, s))$ . They involve those functions and parameters which intervene naturally, namely, apart from the budget  $s$ , the number of random variables and their distributions. They neither speak about specific properties of the latter nor would they refer indirectly to such properties. It is this modesty in the hypotheses that makes, as we think, Theorems 1 (i) and 2 particularly appealing.

Theorems 2 and its special case Theorem 1 (i) are in that sense likely to be the most interesting results. The author thinks that Theorem 3 is broad and therefore of true interest; Theorem 4 may already be close to the borderline. Theorem 1 (i) has a number of citations to its credit. So far it stands out, and the likely reasons are easy to understand. The univariate case is typically more frequent in studied problems than cases involving different distributions. Also, and in particular, Theorem 1 (i) (B. & R. (1991)) has been around much longer than Theorem 2 (Steele (2016)).

The author thinks that, given its generality, Theorem 2, i.e. Steele's version, is very promising from the point of view of future applications. So are then in particular the generalisations Theorem 3 and Theorem 4. However, it is sometimes hard to substantiate their benefit compared with Theorem 2.

## 3. Applications

The goal of this section is to convince that the BRS-inequality is an interesting result, and both flexible and versatile for applications.

We begin our examples with identically distributed uniform random variables. In this example we also compare the improvement of Theorem 3 compared to Theorem 2 (i.e. to Theorem 1 (i)), because i.i.d. uniform random variables are relevant in many contexts.

### 3.1. Identically distributed $U[0, 1]$ variables

Let  $X_1, X_2, \dots, X_n$  be uniform random variables on  $[0, 1]$ , that is  $F(x) = x$  on  $[0, 1]$ . To compute an upper bound for the maximum number of  $X_k$ 's we can sum up without exceeding  $s$  with  $0 < s < n/2$  we solve the BRS-equation (3), that is

$$n \int_0^{t(n,s)} x dF(x) = n \int_0^{t(n,s)} x dx = s \iff t(n,s) = \sqrt{2s/n}. \quad (15)$$

Hence, from the BRS-inequality (2)

$$E(N(n,s)) \leq nF(\sqrt{2s/n}) = \sqrt{2sn}. \quad (16)$$

This bound becomes trivial for  $s \geq n/2 = nE(X)$ , namely  $E(N(n,s)) \leq n$ . We can say slightly more, i.e.  $E(N(n, n/2)) < n$ , because  $n$  items need in expectation the budget  $n/2$ . However, they need a larger budget with a strictly positive probability.

If the  $X_k$ 's are moreover independent, then this bound is essentially tight according to (4).

#### 3.1.1. Sharpening the upper bound

The limiting improvement of  $1/2$  we have seen in Subsection 2.7 for exponential random variables, can also be proved (with different arguments) to hold for i.i.d.  $X_k \sim U[0, 1]$ .

What comes as a nice surprise is to see how quickly the improvement of Theorem 3 compared to Theorem 1, i.e.  $\delta(n, s_n)$  defined in Subsection 2.7, tends to  $1/2$ . This is worth noting since it is also the relative improvement which counts. Even with  $n$  as small as  $n = 7$  and  $s = 0.1$ , i.e.  $s_n = s_7 = 0.7$ , the average improvement in 10 000 simulations (Mathematica, version 5.2; double-precision) turned out  $\hat{\delta}_{10\,000} \approx 0.4664$ . Increasing  $n$  or  $s$  (or both) shifts the improvement rather quickly to  $1/2$ . For example, for  $n = 70$  and  $s_{70} = 7$ , the average improvement for the same number of simulations turned out  $\hat{\delta}_{10\,000} \approx 0.4991$ .

### 3.2. Random variables with dependence

Recall that no independence assumption about the  $X_k$ 's is needed. So let us compare the extreme case of total independence with the case of full dependence such as  $X_1 = X_2 = \dots = X_n$ , say. Here, by the way, we see that the total order defined in the proof before (11) through (A) and (B) is essential. In the case of uniform random variables on  $[0, 1]$  we know from Subsection 3.1 that  $E(N(n,s)) \leq nF(t(n,s)) = \sqrt{2sn}$ . Moreover in the case of i.i.d.  $U[0, 1]$  random variables we have from (4) also the asymptotic relationship  $N(n, s_n) \sim \sqrt{2s_n}$  so that the upper bound computed in (15) and (16) is asymptotically tight.

In the case  $X_1 = X_2 = \dots = X_n$  however, there is trivially only one order statistic, namely  $X_1$ , and our bound should be worse. Indeed, it is easy to compute this bound, namely

$$\begin{aligned} E(N(n, s)) &= \sum_{k=1}^n P(N(n, s) \geq k) = \sum_{k=1}^n P(X_1 \leq s/k) \\ &= \sum_{k=1}^n (\mathbf{1}\{k < s\} + s\mathbf{1}\{k \geq s\}/k) \sim [s] + s \log(n), \end{aligned}$$

where  $[s]$  denotes the largest integer smaller than  $s$ . This upper bound grows more slowly in  $n$  than  $\sqrt{2sn}$ , and it is clear why: independent draws bear naturally the potential of producing small order statistics, whereas these independent draws are now wiped out by the total dependence.

### 3.3. Different distribution functions $F_k$ .

For each  $k \in \{1, 2, \dots, n\}$  let  $X_k$  to be uniform on the real interval  $[0, k]$ . For convenience we take  $0 < s \leq 1$  and  $n \geq 4$ . The BRS-equation (6) tells us

$$s = \frac{1}{2} \sum_{i=1}^n \frac{1}{i} t^2(n, s) = \frac{1}{2} t^2(n, s) H_n \quad \text{or} \quad t(n, s) = \sqrt{2s/H_n}, \quad (17)$$

where  $H_n$  denotes the  $n$ 'th harmonic number. In particular, for  $s = 1$ , we get from (5) using (17)

$$E(N(n, 1)) \leq \sum_{i=1}^n F_i(t(n, 1)) = \sum_{i=1}^n \frac{1}{i} (2/H_n)^{1/2} = (2H_n)^{1/2},$$

where we use  $n \geq 4$  to assure that  $H_n > 2$ . This bound grows also much more slowly than the bound  $\sqrt{2ns}$  we had obtained in 3.1. The reason is now that small order statistics become rarer with the increasing length of the supports  $[0, k]$  of the uniform  $F_k$ 's.

If, in contrast, we have decreasing lengths of supports as  $k$  grows, the upper bound for  $E(N(n, s))$  should go up. We would expect this upper bound to become much closer to  $n$  and in the order of  $n$  if the supports of the  $F_k$  are shrinking more or less in the same order or quicker as/than the expected order statistics of  $n$  variables would do on the support of  $F_1$ . To exemplify this on  $[0, 1]$ , let

$$F_k(x) = \mathbf{1}\{0 \leq x \leq 1/k\} kx + \mathbf{1}\{x > 1/k\}, \quad k = 1, 2, \dots, n,$$

and let again  $s = 1$ , say. Putting  $h(t) = [1/t]$  and  $t = t(n, 1)$ , the BRS-equation (6) becomes

$$\sum_{k=1}^n \int_0^t x dF_k(x) = \sum_{k=1}^{h(t)} kt^2/2 + \sum_{h(t)+1}^n 1/(2k) = 1.$$



Clearly we must have  $t(n, 1) \rightarrow 0$  as  $n \rightarrow \infty$ , and elementary asymptotic analysis shows that  $t(n, 1) \sim e^{3/2}/n$ . Plugging  $t(n, 1)$  into the BRS-inequality (5) yields then (using Mathematica) that, for  $n$  sufficiently large, the upper bound for  $E(N(n, 1))$  is about  $0.8884n$ .

### 3.4. Point processes and selection bias

Let  $T_0 = 0$  and, recursively,  $T_k = T_{k-1} + X_k, k = 1, 2, \dots$  where the  $X_k$ 's are positive random variables with absolute continuous laws  $F_k$ . The  $X_k$ 's are now seen as the inter-arrival times of a point process  $(T_k)_{k=1,2,\dots}$  on the positive half-line.

Suppose now that we fix  $n$  and wait for the  $n$ th arrival, occurring at time  $T_n = T$ , say. Given  $s$  with  $0 < s < T$  we ask what is the maximum density of arrival points we can expect on a set  $\mathcal{S} \in U[T]$  with Lebesgue measure  $s$ , where  $U[T]$  is the set of all subsets of  $\{[T_{k-1}, T_k]: 1 \leq k \leq n\}$ ? (Note that if we did not confine our interest to  $U(T)$ , then the answer would be trivially  $\infty$ , because we could, for instance, collect open balls  $\mathcal{B}(T_j, \epsilon)$  and then let  $\epsilon \rightarrow 0+$ .)

The maximum obtainable density is  $N(n, s)/s$  with  $N(n, s)$  defined in (1) based on the inter-arrival times  $X_1, X_2, \dots, X_n$ . Thus from the BRS-inequality we know that the maximum expected density is bounded by

$$d_{\max}(s) := \frac{E(N(n, s))}{s} \leq \frac{1}{s} \sum_{k=1}^n F_k(t(n, s)), \quad (18)$$

where  $t(n, s)$  solves the BRS-equation (3) if the  $X_k$  are identically distributed, and otherwise solves (6). We also note that in the case of independence (where we have the stronger result (4)), the problem of finding the set with *minimum* expected density is then *asymptotically* dual, because it suffices to maximise the number of points on the complementary set with Lebesgue measure  $T_n - s$ .

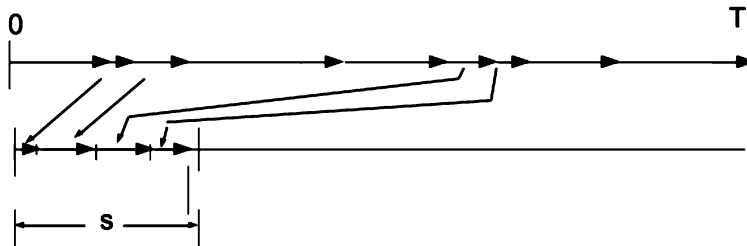


FIG 2. Figure 2 shows for a given “budget time”  $s$  the selection of the 4 smallest inter-arrival times until arrival time  $T = T_9$ . As  $n$  becomes larger, i.e. as  $T = T_n$  moves to the right, the number of arrival points in a selected set of fixed Lebesgue measure  $s$  cannot but increase.

### 3.5. Poisson processes

As a specific example, consider a homogeneous Poisson process of rate 1. Hence here all the  $F_k$  are the same. Suppose that a dishonest statistician would like to make an observer (client) believe that the rate is higher than 1 by claiming that the missing data are due to the fact that the counting process could not be recorded during certain sub-periods. How far can this type of dishonesty bias the perceived rate?

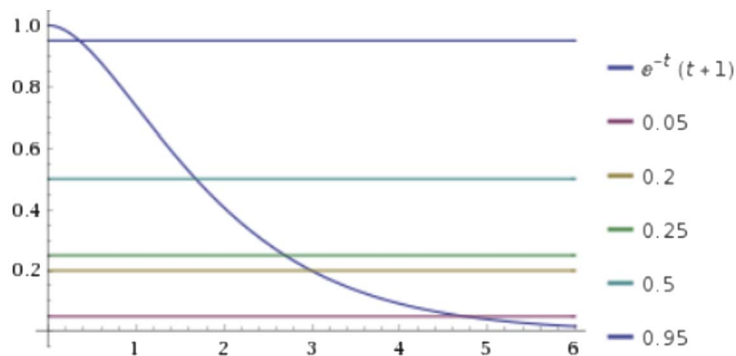


FIG 3. Figure 3 displays for a Poisson process of rate 1 the lhs of (19) for  $0 \leq t \leq T = n$ , and the rhs of (19) for a selection of values of  $s/n$  ranging between 5% and 95% (horizontal lines  $\equiv$  fraction of the observation time.)

The inter-arrival times of the homogeneous Poisson process (rate 1) are i.i.d. exponential random variables with distribution function  $F(x) = 1 - e^{-x}$ . No conditional bias intervenes here since we look at the random horizon  $T = T_n$ , where  $T_n$  denotes the  $n$ th arrival time. The BRS-equation (3) yields then

$$\int_0^t x dF(x) = \int_0^t x e^{-x} dx = \frac{s}{n},$$

or equivalently, the implicit equation

$$e^{-t}(t+1) = 1 - \frac{s}{n}. \quad (19)$$

Thus from (18) we obtain with  $t := t(n, s)$  solving (19)

$$d_{\max}(s) \leq \frac{1}{s} n F(t(n, s)) = \frac{1}{s} n \left(1 - e^{-t(n, s)}\right). \quad (20)$$

In the Poisson process case the random variables  $X_k$  are i.i.d., and so (4) implies that the obtained bounds obtained from the inequality (20) are essentially tight (for  $s/T$  bounded away from 0). The 50% line for instance (see Figure 3) cuts the curve in  $t \approx 1.67835$ , and (19) and (20) yield a maximum density  $\approx 1.62664$ , that is about 63 percent more than the true rate. Clearly, the less time we observe, the more the observed rate can grow. With only 5% observation time we obtain from (19) and (20) that the dishonest statistician can cheat (a naive client) by a factor almost 6.

### 3.6. “Awkward” point processes

Sometimes one comes across random structures which are complicated, but have components which are nevertheless well understood, and/or tractable. Staying in the class of problems we have considered in Subsections 3.4. and 3.5, the following awkward counting process (completely invented) exemplifies such an instance.

Let  $(X_k)_{k=1,2,\dots}$  be a random sequence of inter-arrival times of a point process.  $X_1$  is uniform on  $[0, 1]$ , say,  $X_2 = 1 - X_1$ ,  $X_3 = 1 - X_2 = X_1$ , ... and this alternating pattern goes on until some random time  $C_1$ .  $X_{C_1+1}$  is then again  $U[0, 1]$ -distributed independently, and the following variables play the same alternating game as before until some random time  $C_2$ . Then again a new  $U[0, 1]$  random variable  $X_{C_2+1}$  is drawn ... and so on. Suppose we are stopped after a large number of  $n$  observations (in total). Can we say anything nontrivial about the expected number of variables among  $X_1, X_2, \dots, X_n$  we can sum up without exceeding a given  $s$ ?

At first glance it may seem that we must know something about the mechanism producing the random times  $C_1, C_2, \dots$ .

However, we simply concentrate on the essential information which is given: whatever the law of  $X_1, X_2, \dots$ , whatever their dependence structure, all the *marginals* of the  $X_k$  are the same, namely  $F_k(t) = F(t) = t$  on  $[0, 1]$ . Hence, as seen in Subsection 3.1, (16), we always have the convenient bound

$$E(N(n, s)) \leq \sqrt{2sn}. \quad (21)$$

If we happen moreover to know one of the expected change times,  $E(C_n)$  say, then we have also a nice upper bound for  $E(N(C_n, s))$ , namely by the tower property of conditional expectations,

$$E(N(C_n, s)) = E[E(N(C_n, s)|C_n)] \leq E\left(\sqrt{2sC_n}\right) \leq \sqrt{2sE(C_n)},$$

where the first inequality follows from (21), and the second one from the concavity of the square-root function and Jensen’s inequality.

With more information one can sometimes say more. To stay with our example, suppose all block lengths  $C_{i+1} - C_i$  of strictly dependent random variables are geometric random variables with parameter  $p$ , say. Instead of looking at the horizon of  $n$  variables we now look at  $n$  blocks, that is altogether  $C_n$  variables with budget  $s := s_{C_n}$ . The expected length of blocks consisting of strongly dependent variables is thus  $1/p$ .

On the one hand, the larger  $p$  gets, the higher becomes the fraction of independent  $U[0, 1]$  random variables so that upper bound for  $C_n$  variables

$$\sqrt{2s_{C_n}C_n} \sim \sqrt{2s_{C_n}n/p}$$

must become better and better. On the other hand, if  $p$  becomes small, the expected block lengths become large so that most blocks contain with each  $X$  also the value  $1 - X$ . This opens then a new option, namely we can recompute a new threshold by taking the variables two by two.

**Heuristic arguments:** All what we say here is that in point processes defined by a mixture of both i.i.d. inter-arrival times and strongly dependent ones, it can be worth checking whether the fraction of i.i.d. random variables is non-negligible, knowing that the BRS-bound is essentially tight for these. For instance, if these variables are  $U([0, 1])$  random variables, then  $N(\cdot, s)$  is in the order of square-root of their number. The contribution of the others will be somewhere between a logarithmic order (see Subsection 3.2) and an order of square-root, and may contribute to the total  $E(N(n, s))$  just something asymptotically negligible.

### 3.7. Sequential selection problems

As has been noticed by several authors before, as e.g. Gnedin (1999) and Arlotto et al. (2015a; 2015b), the BRS-inequality leads to *a priori* upper bounds for two well-studied problems in combinatorial optimisation. In particular, in the classical case of independent uniformly distributed random variables, the BRS-inequality (Theorem 1 (i)) gives bounds that are essentially sharp for both the sequential knapsack problem and the sequential (increasing) subsequence selection problem (Steele (2016)).

#### 3.7.1. Knapsack problem

In this problem, one observes a sequence of  $n$  independent non-negative random variables  $X_1, X_2, \dots, X_n$  with a fixed, known distribution  $F$ . One is also given a real value  $x \in [0, \infty)$  that one regards as the capacity of a knapsack into which selected items are placed. Observations are sequential, and when  $X_i$  is first observed, one either selects  $X_i$  for inclusion in the knapsack or else  $X_i$  is rejected from any future consideration. The goal is to maximise the expected number of items,  $\tilde{N}(n, x)$ , say of items that can be sequentially packed without recall into the knapsack with initial capacity  $x$ . Since the BRS-inequality tells us how well we could do if we knew in advance all of the values  $\{X_i : 1 \leq i \leq n\}$ , it is evident that no strategy for making sequential choices can ever lead in expectation to more affirmative choices than  $E(N(n, x))$ , and thus from (2) we have the bounds

$$E(\tilde{N}(n, x)) \leq E(N(n, x)) \leq nF(t(n, x)), \quad (22)$$

where  $t := t(n, x)$  solves the associated BRS-equation (3).

Now, packing an item into the knapsack will not change the distribution of the items to come. The only effect of packing an item of size  $y$  say is reducing the current remaining capacity by  $y$ . Hence the problem is a Markov decision problem so that the optimal sequential selection strategy given by a unique non-randomised Markovian decision rule. Beginning with  $n$  values to be observed and with an initial knapsack capacity of  $x$ , the expected number of selections that

one makes is denoted by  $v_n(x)$ . It is easy to see that the value function for this Markov decision problem can be calculated by the recursion relation

$$v_n(x) = (1 - F(x))v_{n-1}(x) + \int_0^x \max\{v_{n-1}(x), 1 + v_{n-1}(x - y)\} dF(y). \quad (23)$$

Specifically, one begins with the obvious relation  $v_0(x) \equiv 0$ , and one computes  $v_n(x)$  by iteration. This is the Bellman equation (optimality equation) for the sequential knapsack problem.

### 3.7.2. Monotone subsequence problem

Now suppose that we observe sequentially  $n$  independent random variables  $X_1, X_2, \dots, X_n$  with the common continuous distribution  $F$ . Our goal is to make monotonically (decreasing, say) choices

$$X_{i_1} > X_{i_2} > \dots > X_{i_k},$$

trying to maximise the expected number of choices. Here, prior to making any selection, we take the state variable  $x$  to be the supremum of the support of  $F$ , which may be infinity. After we have made at least one selection, we take the state variable  $x$  to be the value of the *last* selection that was made. Now we write  $\tilde{v}_n(x)$  for the expected number of selections made under the optimal policy when the state variable is  $x$  and where there are  $n$  observations to come. In this case the Bellman equation given by Samuels and Steele (1981) can be written as

$$\tilde{v}_n(x) = (1 - F(x))\tilde{v}_{n-1}(x) + \int_0^x \max\{\tilde{v}_{n-1}(x), 1 + \tilde{v}_{n-1}(y)\} dF(y), \quad (24)$$

where again one has the obvious relation  $\tilde{v}_n(x) \equiv 0$  for the initial value. In (20) the decision to select  $X_1 = y$  would move the state variable to  $y$ , so here we have  $1 + \tilde{v}_{n-1}(y)$  where earlier we had the term  $1 + v_{n-1}(x - y)$  in the knapsack Bellman equation. In the knapsack problem the state variable moves from  $x$  to  $x - y$  when  $X_1 = y$  is selected.

In general, the solutions of (23) and (24) are distinct, but Coffman et al. (1987) made the essential observation that  $v_n(x)$  and  $\tilde{v}_n(x)$  are equal when the observations are uniformly distributed. This allows to create an interesting detailed linkage between these two problems: The first step is to note that the equality of the value functions permits one to construct optimal selection rules that can be applied simultaneously to the same sequence of observations. The selections that are made will be different in the two problems, but we see a useful distributional relationship. The essential observation is that the second term of the Bellman equation leads one almost immediately to the construction of an optimal selection strategy for the monotone subsequence problem. These strategies lead one in turn to a more detailed understanding of the number of values that one actually selects.

First, one notes that it is easy to show (Samuels and Steele (1981)) that there is a unique  $y \in [0, 1]$  that solves the “equation of indifference”:

$$\tilde{v}_{n-1}(x) = 1 + \tilde{v}_{n-1}(y).$$

If we denote this solution by  $\alpha_n(x)$ , we can use its values to determine the rule for making the sequential selections. At the moment just before  $X_i$  is presented, we face the problem of selecting a monotone sequence from among the  $n - i + 1$  values  $X_i, X_{i+1}, \dots, X_n$ , and if we let  $\tilde{S}_{i-1}$  denote the last of the values  $X_1, X_2, \dots, X_{i-1}$  that has been selected so far, then we can only select  $X_i$  if it is not greater than the most recently selected value  $\tilde{S}_{i-1}$ . In fact, one would choose to select  $X_i$  if and only if it falls in the interval  $[\tilde{S}_{i-1}, \tilde{S}_{i-1} - \alpha_{n-i+1}(\tilde{S}_{i-1})]$ . Thus, the actual number of values selected out of the original  $n$  is the random variable given by

$$\tilde{V}_n \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbf{1} \{X_i \in [\tilde{S}_{i-1}, \tilde{S}_{i-1} - \alpha_{n-i+1}(\tilde{S}_{i-1})]\}.$$

By the same logic, one finds that in the sequential knapsack problem the number of values that are selected by the optimal selection rule can be written as

$$V_n \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbf{1} \{X_i \in [0, \alpha_{n-i+1}(S_{i-1})]\}$$

where now  $S_{i-1}$  denotes the capacity that remains after all of the knapsack selections have been made from the set of values  $X_1, X_2, \dots, X_{i-1}$  that have already been observed. By this parallel construction we have

$$\mathbb{E}[V_n] = v_n(1) = \tilde{v}_n(1) = \mathbb{E}[\tilde{V}_n].$$

Moreover, one has  $S_0 = 1 = \tilde{S}_0$ , and then see the equality of the joint distributions of the vectors

$$(S_0, S_1, \dots, S_{n-1}) \quad \text{and} \quad (\tilde{S}_0, \tilde{S}_1, \dots, \tilde{S}_{n-1}),$$

since the two processes  $\{S_i : 0 \leq i \leq n\}$  and  $\{\tilde{S}_i : 0 \leq i \leq n\}$  are (temporally non-homonomous) Markov chains that have the same transition kernel at each time epoch.

Theorem 1 (i) tells us now that  $\mathbb{E}[V_n] \leq \sqrt{2n}$ . By the distributional identity of  $V_n$  and  $\tilde{V}_n$  we find indirectly that

$$\mathbb{E}[\tilde{V}_n] \leq \sqrt{2n} \quad \text{for all } n \geq 1. \quad (25)$$

It turns out that (25) can be proved by a remarkable variety of methods. In particular, Gnedin(1999) gave a direct proof where one can even accommodate a random sample size  $N$  and where the upper bound is now replaced with the natural proxy  $(2\mathbb{E}[N])^{1/2}$ . More recently, Arlotto et al. (2015a) gave two

further proofs of (25) as consequences of bounds that were developed for the *quickest selection problem*, a sequential decision problem that provides a kind of combinatorial dual to the classical sequential selection problem.

The distributional identity can also be used to make some notable inferences about the knapsack problem from what has been discovered in the theory of sequential monotone selections. For example, by building on the work of Bruss and Delbaen ((2004)), Arlotto et al. (2015b) found that

$$\text{Var}[\tilde{V}_n] \sim \frac{1}{3}\sqrt{2n} \quad \text{and} \quad \frac{\tilde{V}_n - \sqrt{2n}}{3^{-1/2} (2n)^{1/4}} \Rightarrow \mathcal{N}(0, 1), \quad (26)$$

where  $\Rightarrow$  stands for convergence in law. Thus, as a consequence of the distributional identity  $V_n \sim \tilde{V}_n$  one has the same results for the knapsack variable  $V_n$  for i.i.d.  $U([0, 1])$  random variables.

It is also interesting to recall here the *non-sequential* (or clairvoyant) selection problem where one studies the random variable

$$L_n = \max\{k : X_{i_1} < X_{i_2} < \dots < X_{i_k}, 1 \leq i_1 < i_2 < \dots < i_k \leq n\}.$$

This classic problem has a long history, beautifully told in both Romik (2014) and in Aldous and Diaconis (1999). Here the most relevant part of that story is that Baik et al. (1999) found the asymptotic distribution of  $L_n$ , and, in particular, they found that one has the asymptotic relation

$$\text{E}[L_n] = 2\sqrt{n} - \alpha n^{1/6} + o(n^{1/6}) \quad \text{where } \alpha = 1.77108\dots \quad (27)$$

For the dominant part  $2\sqrt{n}$  we see, as already observed in Samuels and Steele (1981), the advantage of clairvoyance is essentially the factor  $\sqrt{2}$ .

### 3.8. Random tilings with different shapes

We have given so far already several applications of Theorem 1. Now we show that the generalisation provided in Theorem 2 has much to offer for other applications. For example, look at the following “multi-type” online problem:

#### 3.8.1. Selection and online-selection

We are given a connected  $d$ -dimensional subset  $S$  of  $\mathbb{R}^d$  with Lebesgue measure  $s$ , and a sequence of  $d$ -dimensional random shapes which we would like to fit into  $S$  without overlappings. If the sequence contains  $n_k$  shapes of type  $k$ , and there are  $\sigma$  types of shapes, what is the maximum number of non-overlapping shapes we can hope to fit into  $S$ ? Also, what would be a good strategy to fit the shapes online, i.e., sequentially and without recall?

First, neglecting the online question, we have a convenient upper bound by the BRS-equality. We define  $n = n_1 + n_2 + \dots + n_\sigma$ , and let  $V_1, V_2, \dots, V_n$  be the volumes of the  $n$  random shapes. If the randomization algorithm is known

for the  $\sigma$  types of shapes and the same *within* each class of shapes then we can compute the distribution functions  $F_k, k = 1, 2, \dots, \sigma$  of their volumes. We suppose that the latter are absolute continuous random variables  $X_k$ . Hence from (5) and (6)

$$\mathbb{E}(N(n, s)) \leq \sum_{k=1}^{\sigma} n_k F_k(t(n, s)), \quad n = n_1 + n_2 + \dots + n_{\sigma}, \quad (28)$$

where  $t(n, s)$  solves the corresponding BRS-equation

$$\sum_{k=1}^{\sigma} n_k \int_0^{t(n, s)} v dF_k(v) = s. \quad (29)$$

Now, coming back to online-selection, how well could we do with a skilfull online selection if we select, independently of the type of shapes all those objects which have a random volume  $V \leq t(n, s)$ , and fit these best possible into  $S$ ?

As an example we study a specific tiling problem involving two types of shapes.

### 3.8.2. Rectangles and ellipses

Suppose  $S$  is the unit square  $S = [0, 1] \times [0, 1]$  and that we have two types of shapes, namely rectangles (type 1) and ellipses (type 2). Let  $X, Y$  be independent  $U[0, 1]$  random variables, and  $A, B$  be independent  $U[0, 1/2]$  random variables. Here  $X$  and  $Y$  denote the random length of the sides of a rectangle, and  $A$  and  $B$  the random height and width of an ellipse. Further let  $X_j, Y_j, j = 1, 2, \dots, n$  respectively  $A_j, B_j, j = 1, 2, \dots, n$  be independent versions of these.

Then we get for the surface area distribution function  $F_1(u)$  (rectangle) for  $u \in [0, 1]$ ,

$$\begin{aligned} F_1(u) &= P(XY \leq u) = \int_0^1 P\left(Y \leq \frac{u}{X} \mid X = x\right) dx \\ &= \int_0^u dx + \int_u^1 \frac{u}{x} dx = u(1 - \log(u)), \end{aligned} \quad (30)$$

and similarly for the distribution of the surface area of a random ellipse

$$\begin{aligned} F_2(u) &= P\left(AB \leq \frac{u}{\pi}\right) = \int_0^{1/2} P\left(B \leq \frac{u}{A\pi} \mid A = a\right) da \\ &= \mathbf{1}\{0 \leq u \leq \pi/4\} \frac{4u}{\pi} \left(1 - \log\left(\frac{4u}{\pi}\right)\right) + \mathbf{1}\{\pi/4 \leq u \leq 1\}. \end{aligned} \quad (31)$$

We have  $s = 1$  and  $\sigma = 2$ . For our example (see Figure 4) let us choose  $n_1 = 300$  and  $n_2 = 150$ , say. Putting  $t := t(450, 1)$  we must solve according to the BRS-equation (29)

$$300 \int_0^t u dF_1(u) + 150 \int_0^t u dF_2(u) = 1, \quad (32)$$



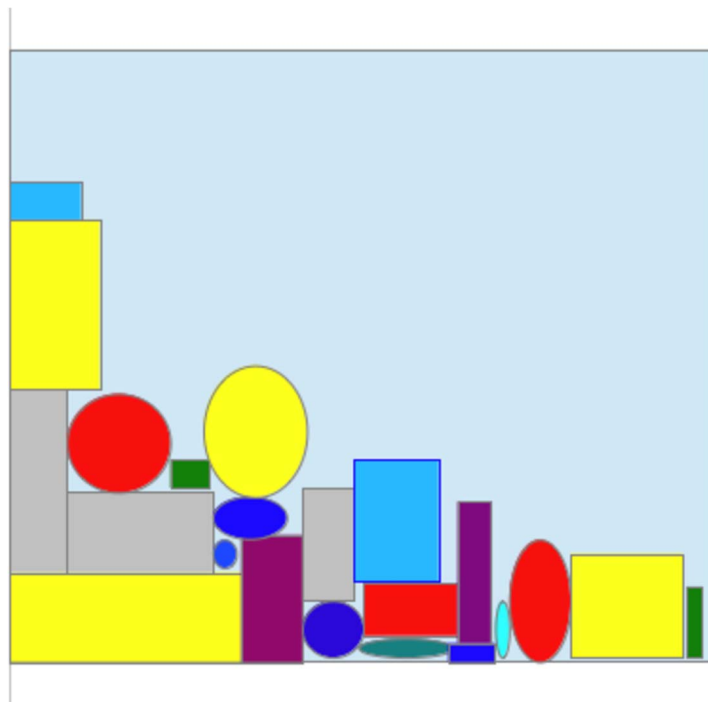


FIG 4. Figure 4 displays online tiling with rectangles and ellipses. The square has surface 1, and  $n_1 = 300$  and  $n_2 = 150$ . Solving the corresponding BRS-equation (32) yields (using Mathematica)  $t = 0.0326\dots$  as upper surface threshold (limit) for the rectangles and ellipses we select.

where  $F_1(u)$ , respectively  $F_2(u)$  are given by (30), respectively (31). The numerical solution of equation (32) (by Mathematica) yields  $t = 0.0326\dots$ . Hence from inequality (28)

$$E(N(450, 1)) \leq 300F_1(t) + 150F_2(t) = 69.325\dots$$

Instead of storing 300 rectangles and 150 ellipses we randomised (for convenience) the shape  $P(\text{type 1}) = 2/3, P(\text{type 2}) = 1/3$ , so that we obtain in 450 trials in expectation 300 rectangles and 150 ellipses. Then we randomised their respective parameters, packing those rectangles or ellipses with surface area  $\leq t$  independently of their type, online, and just on eyesight, in the square.

In Figure 4 the upper bound  $E(N(450, 1)) \leq 69.325\dots$  seems surprisingly feasible. Admittedly, we were lucky with a few very small ellipses. Recall the bounds in (22) and note that our upper bound is a full-information upper bound, and not only an upper bound for an online strategy. With other types of shapes the relative loss through unusable space may be higher, and for certain non-convex shapes it is likely to be much higher. Figure 4 looks unfinished, and this was on purpose. We wanted to stop half-way i.e. at 225 simulations, in order to

see how it looks so far. Not on purpose was that we had to stop earlier due to a legally imposed interruption, and forgotten to take note of the exact last step number (at around 150?)

The only theory we applied here was thus computing the BRS-threshold  $t(n, s) = t(450, 1)$ . It is of great help to have just a single decision function, namely the area threshold  $t(n, s)$ . We asked the computer to serve us sequentially only those random shapes with area below  $t(n, s)$ . Then each offer meant “Place it!”, and that’s what we did (without rotations of the shapes), just as they came. With the right computed threshold it works well, and is just fun to observe.

### 3.9. Borel-Cantelli type problems

For Borel-Cantelli type problems, the BRS-inequality can also be of interest. It sometimes allow us to isolate on a set of dependent events condensed subsets on which sub-events become independent. In our first example we use an argument in the spirit of the key-inequality (10).

#### 3.9.1. Direct neighbours

Let  $X_1, X_2, \dots, X_n$  be i.i.d. uniform random variables on  $[0, 1]$ . Further let  $A_n$  be the event that the  $(n + 1)$ st draw  $X_{n+1}$  becomes a direct neighbour of  $X_n$ , that is, the interval  $[\min\{X_n, X_{n+1}\}, \max\{X_n, X_{n+1}\}]$  contains no  $X_j$  with  $1 \leq j \leq n - 1$ . First, we are interested to know  $P(A_n \text{ i.o.})$  as  $n \rightarrow \infty$ , where “i.o.” stands for infinitely-often. Clearly  $A_n$  depends on where  $X_n$  has landed so that the  $A_n$  are not independent. Hence  $\sum_n P(A_n) = \infty$  does not imply from the Borel-Cantelli Lemma that  $P(A_n \text{ i.o.}) = 1$ .

However, fix now a constant  $c$  with  $0 < c < 1$  and then think of all spacings with length  $\leq c/n$  united in one set  $A$ . This takes less space than Lebesgue measure  $c$  so that the difference set  $B \subset [0, 1]$  with  $A \cup B = [0, 1]$  has Lebesgue measure at least  $1 - c$ . This is already the simple idea: at least the Lebesgue measure  $1 - c$  is available for  $X_{n+1}$  *independently* of  $n$ , and within the set having at least this measure, we know that the minimal distance between neighbored points is larger than  $c/n$ .

To finish the question, let now  $B_n := \{X_{2n} \text{ and } X_{2n+1} \text{ become neighbours}\}$ . Clearly, the  $B_n$  are independent. Also,  $P(B_n) > (1 - c)c/(4n)$ , because if  $X_{2n} \in B$ , which occurs with probability at least  $1 - c$ , more than one half of  $c/(2n)$  of space is free for the next point  $X_{2n+1}$  to fall to the left or right of it and thus become its direct neighbour. Since  $\{B_n \text{ i.o.}\} \subset \{A_n \text{ i.o.}\}$ , and the harmonic sum diverges, we have  $P(A_n \text{ i.o.}) = 1$ .

#### 3.9.2. Spacings

Second, let us now look at all spacings generated sequentially by the draws  $X_n$ . Fix a function  $k := k(n)$  with  $1 \leq k \leq n + 1$  and look at the events

$\mathcal{K}_n = \{X_{n+1} \text{ falls in the } k(n)\text{th smallest spacing}\}$ . Bruss et al. (1990) found criteria on  $k(n)$  for  $P(\mathcal{K}_n \text{ i.o.}) = 0$ , respectively  $P(\mathcal{K}_n \text{ i.o.}) = 1$ . In a nice more elaborate approach, Alsmeyer (1991) improved these results by giving the complete if-and-only-if answer.

Now, what about the event of hitting infinitely often *one* of the  $k(n)$  smallest spacings? Does it suffice to replace Alsmeyer’s sum criterion by the corresponding double sum for all  $j = 1, 2, \dots, k(n)$ ? (Since Alsmeyer refers in his paper to decreasing order statistics we must read his criterion by replacing his  $k(n)$  by  $n - k(n) + 1$ .) This seems intuitive, and a similar argument as above shows that this must be true. However, here the BRS-inequality cannot give as much as Alsmeyer’s approach, which uses the strong property of uniform spacings behaving, normalised, like independent exponential variables.

Under weaker conditions, the BRS-inequality may however give results where other approaches cannot help. For example if the BRS-sum  $S_{A(n, s_n)}$  (see (8) to (10)) “accommodates” the sum of the  $k(n)$  smallest spacings for  $n$  sufficiently large, and if  $E(N(n, s_n))$  stays bounded then we *always* have  $P(\mathcal{K}_n \text{ i.o.}) = 0$ . Moreover, if a similar independent set can be found as in the first question above, then we are also able to find, with  $\sum_{j=1}^n s_j$  tending to infinity as  $n \rightarrow \infty$ , the desired converse.

## 4. Further connections and applications

### 4.1. Brief overview

Here the whole goal has been to explain, extend, and explore the upper bound given by the BRS-inequality. The only work we know in which Theorem 2 was applied is Steele (2016). Since, as said before, Theorem 1 has been around since 1991 already, we heard in talks occasionally speakers referring to it. We also know references in the literature where Theorem 1 was used explicitly or referred to. In an attempt to be somewhat complete in this survey, here is a list in alphabetic order of the written references we know:

- Arlotto, A. and Steele, J. M. (2011),
- Arlotto A., Mossel E. and Steele J. M. (2015a), (2015b)
- Arlotto A., Wei Y., and Xie (2018)
- Boshuizen F. A. and Kertz R. P. (1995), (1999)
- Bruss F. T. (2014)
- Caballero J. A., Lepora N. F. and Gumey K.N. (2015)
- Cain N. and Shea-Brown (2013)
- Gnedin A. V. (1999)
- Gnedin A. V. and Seksenbayev A. (2019)
- Gribonval R., Cevher V. and Davies M. E. (2012)
- Jacquemain A. (2017)
- Kaluszka M., Okolevski A. and Szymanska K. (2005)
- Kiwi M. and Soto A. J (2015)
- Lasner Z. and DeMille D. (2018)

Li L., Zhou H., Xiong S. X., Yang J. and Y. Mao (2019)  
 Peng P. and Steele J.M. (2016)  
 Pohl O. (2016)  
 Rhee and Talagrand (1991)  
 Stanke M. (2004)  
 Steele J.M. (2016)  
 Truong L V, Tan V. Y. F. (2018)  
 Tunc S., Donmez M.A. and Kozat S.S. (2013)  
 Zhang J.J.J. (2019)

Theorem 1 also plays a central role in the theory of so-called *resource dependent branching processes*, or RDBPs, which we will treat as our last example of applications. The work of B. & R. (1991), instigated by the article of Coffman et al. (1987), was motivated in particular by questions arising in the context of RDBPs.

#### 4.2. Resource dependent branching processes

A resource dependent branching process (RDPB) is a branching process trying to model the development of human populations in a way which seems as realistic as possible. In these RDPB's particles (individuals) have to work in order to survive and to reproduce. Moreover, they live in a society and are supposed to observe the rules their society imposes on them. Individuals receive resources from a global resource space left by their ancestors, and they also create new resources. These are used as resources for their own consumption, but also to add to the resource space for the next generation. All these quantities are modelled by random variables.

##### 4.2.1. Special model assumptions

It is the society who is supposed to decide how resources are distributed among the individuals. Decisions are taken according to two guiding objectives. On the one hand the society wants to survive, and on the other hand, the average individual would like to have a certain standard of living. These objectives are typically in competition to each other. It is this feature which makes a RDPB very different from other branching process models, and the methods to study them have to be different.

In a RDPB individuals and society must also take into account changing environments. In contrast to what one usually calls a process with random environment, the changing environment is now also the result of policies exercised by the society and not only the result of changing parameters, as for instance changing reproduction rates for the individuals. Decisions are reviewed in every generation. (For branching processes with random environments in the more classical sense see in particular the remarkable generality of models enabled by the unifying approach of Kersting (2019).)

Returning to the notion of claims (consumption) in RDBPs, it is the interplay of claims and productivity which will, together with the society's policy to distribute resources, be decisive for the possible survival of the process. Namely, those individuals, who do not receive their minimum claims, are supposed to refuse to replicate within the population. This is for individuals a tool of defence or a means to have a say in policy. We recall that the first priority of the society is to survive. The second one is to do so with a comfortable or at least acceptable standard of living which is modeled for the individuals as the average size of accepted claims.

In the so-called *weakest-first society*, the definition (1) plays a central role as we explain first for a single population. If in a given generation there are  $n$  individuals claiming  $X_1, X_2, \dots, X_n$  under a currently available resource space (budget)  $s$ , and if the weakest (smallest claims) are served first (w.f.-society) then  $E(N(n, s))$  of the  $n$  can be expected to reproduce (having children) for the next generation. (The fact that human reproduction is not asexual, and that thus men and woman and their mating behavior intervene, is, fortunately, here irrelevant. It can be neglected by confining the interest to the so-called *average reproduction mean of mating units*, as shown in Bruss (1984).)

Bruss and Duerinckx (2015) proved that no society can possibly survive unless the w.f.-society survives forever with a strictly positive probability. This result does not surprise because saving resources is helpful for survival if all other parameters stay unchanged. However, the logical converse that wasting resources by greedy individuals is bad for survival is harder to prove. To see this, recall the random variable  $M(n, s)$  defined in Subsection 2.6, namely

$$M(n, s) = \begin{cases} 0, & \text{if } X_{n,n} > s, \\ \max\{k \in \mathbb{N} : X_{n,n} + \dots + X_{n-k+1,n} \leq s\}, & \text{if } X_{n,n} \leq s, \end{cases} \quad (33)$$

where the  $X_{k,n}$  now denote the order statistics of  $n$  claims. If the (strongest) individuals with the largest claims are served first, then less individuals can reproduce. However, with a smaller population in the next generation, their largest claims will go down in expectation (as seen in (33)). According to the model of an RDBP this will be favourable for future available resources, and thus for individuals to replicate.

This is in contrast to what we see for the monotone behaviour of the smallest order statistic, which will go down in size when the sample size increases, so that the trend is maintained. This leads to the study of  $N(n, s)$  and  $M(n, s)$ . Theorem 1 (ii) and Theorem 5 (ii) together allowed to prove the so-called Theorem of Envelopment (Bruss and Duerinckx (2015)). This theorem states that any such RDBP will (in its number of effectives) fluctuate between the bounds which are necessary for the survival of the w.f.-society, respectively, the s.f.-society.

#### 4.2.2. Sub-populations

Theorem 2 allows for dealing with populations which split into several sub-populations with their own parameters. As indicated in the more elementary ex-

amples seen before, some dependencies may be fully compatible with the essence of conclusions. The final goal is to see under which conditions sub-populations may arrive at a certain equilibrium. For a long-term equilibrium between sub-populations to exist, it is a necessary condition for them to survive, and we now understand these conditions. We see that in political decision-making, which is (too often) online, necessary conditions typically play a more important role than sufficient conditions. Also, depending on the degree of independence assumptions for the behaviour of individuals within the same sub-population, these can turn into sufficient conditions.

Immigration of sub-populations and its effect on equilibria can now, due to Theorem 2, be treated in one single framework (Bruss (2018)). Understanding the *necessary conditions* for the survival of sub-populations within a given population is a first step to see whether an equilibrium between sub-populations is possible. For such questions, independence assumptions become more and more difficult to defend, if one wants to stay within realistic assumptions. This is why the author believes that, if RDBPs, as imperfect as they might be, will attract interest as reasonable models for interacting human sub-populations, then Steele's contribution will be very important for their understanding.

## Conclusion

As the style of the present survey may suggest, it was written with the objective in mind to convince: the BRS-inequality is a versatile tool. The author would be delighted to see readers profit from it for problems, or sub-problems, in their own research.

## Acknowledgments

The author would like to thank cordially J. Michael Steele, University of Pennsylvania, Wharton School, for his stimulating article of 2016, and also for so many interesting and greatly enjoyed discussions we had together since then. His sincere appreciation and thanks go also to the Editor and to the referees for their work and helpful comments.

## References

- Aldous, D. and Diaconis, P. (1999), *Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem*, Bull. Amer. Math. Soc. (N.S.), 36, 413-432. [MR1694204](#)
- Alsmeyer G. (1991), *Complete answer to an interval splitting problem*, Statistics & Probability Letters, Vol. 12, 285-287. [MR1131050](#)
- Arlotto A., Mossel E. and Steele J. M. (2015a), *Quickest online selection of an increasing subsequence of specified size*, Random Structures and Algorithms (eprint arXiv:1412.7985). [MR3536538](#)

- Arlotto A., Nguyen V. V. and Steele J. M. (2015b), *Optimal online selection of a monotone subsequence: a central limit theorem*, Stochastic Process. & Appl., Vol. 125 (9), 3596-3622. [MR3357621](#)
- Arlotto, A. and Steele, J. M. (2011), *Optimal sequential selection of a unimodal subsequence of a random sequence*, Combin. Prob. Comput., 20, 799-814. [MR2847268](#)
- Arlotto A., Wei Y. and Xie X. (2018) *An adaptive  $O(\log n)$ -optimal policy for the online selection of a monotone subsequence from a random sample*, Random Structures & Algorithms, 2018 – Wiley Online Library. [MR3731611](#)
- Baik J., Deift P. and Johansson K. (1999), *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. Vol. 12 (4), 1119-1178. [MR1682248](#)
- Boshuizen F. A. and Kertz R. P. (1995), *Largest-fit selection of random sizes under a sum constraint: comparisons by weak convergence*, in Festschrift in honour of V.S. Korolyuk (A.V. Skorokhod and Y.V. Borovskikh, eds.), 55-78. [MR1713992](#)
- Boshuizen F. A. and Kertz R. P. (1999), *Smallest-fit selection of random sizes under a sum constraint: weak convergence and moment comparisons*, Adv. in Appl. Probab. Vol. 31 (1), 178-198. [MR1699667](#)
- Bruss F. T. (1984), *A note on extinction criteria for bisexual Galton-Watson processes*, J. Appl. Probab., 915-919. [MR0766827](#)
- Bruss F. T. (2014), *Grenzen einer jeden Gesellschaft*, Jahresbericht Deutsch. Math.-Verein. Vol. 116 (3), 137-152. [MR3233985](#)
- Bruss F. T. (2018) *Equilibrium Equations for Human Populations with Immigration*, arXiv:[1805.01395](#).
- Bruss F. T., Jammalamadaka S. R. and Zhou X. (1990), *On an interval splitting problem*, Statistics & Probability Letters, Vol. 10, 321-324. [MR1069909](#)
- Bruss F. T. and Delbaen F. (2004), *A central limit theorem for the optimal selection process for monotone subsequences of maximum expected length*, Stochastic Proc. & Applic. Vol. 114 (2), 287-311. [MR2101246](#)
- Bruss F. T. and Duerinckx M. (2015), *Resource dependent branching processes and the envelope of societies*, Ann. of Appl. Probab., Vol. 25 (1), 324-372. [MR3297775](#)
- Bruss F. T. and Robertson J. B. (1991) *'Wald's Lemma' for Sums of Order Statistics of i.i.d. Random Variables*, Adv. Appl. Probab., Vol. 23, 612-623. [MR1122878](#)
- Caballero J. A., Lepora N. F. and Gumez K.N. (2015), *Probabilistic Decision Making with Spikes: From ISI distributions to behavior via information gain*, PLoS One, Vol. 10 (4); avail. at <https://doi.org/10.1371/journal.pone.0124787>.
- Cain N. and Shea-Brown (2013), *Impact of Correlated Neural Activity on Decision-Making Performance*, Neural Computation, Vol. 25 (2), 289-327. [MR3058498](#)
- Coffman Jr. E. G., Flatto L. and Weber R. R. (1987), *Optimal selection of stochastic intervals under a sum constraint*, Adv. Appl. Probab. Vol. 19 (2), 454-473. [MR0889945](#)

- Gnedin A. V. (1999), *Sequential selection of an increasing subsequence from a sample of random size*, J. Appl. Probab. Vol. 36 (4), 1074-1085. [MR1742151](#)
- Gnedin A. V. and Seksenbayev A. (2019), *Asymptotics and Renewal Approximation in the Online Selection of Increasing Subsequence*, arxiv:[1904.11213v2](#).
- Gribonval R., Cevher V. and Davies M. E. (2012), *Compressible distributions for highdimensional statistics*, IEEE Trans. Inform. Th., Vol. 58 (8), 5016-5034. [MR2957417](#)
- Jacquemain A. (2017) *Lorenz curves interpretations of the Bruss-Duerinckx theorem for resource dependent branching processes*, arXiv:[1708.01085](#).
- Kaluszka M., Okolevski A. and Szymanska K. (2005), *Sharp bounds for L-statistics from dependent samples of random length*, Journ. Stat. Planning and Inference, Vol. 127, Issue 1-2, 71-89. [MR2102705](#)
- Kersting G. (2019) *A unifying approach to branching processes in varying environments*, arXiv:[1703.01960v7](#). [MR4094390](#)
- Kiwi M. and Soto A. J. (2015). *Longest Increasing Subsequences of Randomly Chosen Multi-Row Arrays*, Combinatorics, Probability and Computing, Vol. 24, pp. 254-293. [MR3318046](#)
- Lasner Z. and DeMille D. (2018) *Statistical sensitivity of phase measurements via laser-induced fluorescence with optical cycling detection*, Phys. Rev. A 98, 053823; Published online 14 Nov. 2018
- Li L., Zhou H., Xiong S. X., Yang J. and Y. Mao (2019) *Compound Model of Task Arrivals and Load-Aware Offloading for Vehicular Mobile Edge Computing Networks*, in IEEE Access, vol. 7, pp. 26631-26640.
- Maghsudi S. and Hossain E. (2017). *Distributed use association in energy harvesting small networks: An exchange economy with uncertainty*, IEEE Xplore, Vol. 1 (3), 294-308.
- Peng P. and Steele J.M. (2016) *Sequential selection of a monotone subsequence from a random permutation*, Proceed. Amer. Math. Society, Vol. 144, 4973-4982. [MR3544544](#)
- Pohl O. (2016) *Chemotaxis of sephoretic active particles and bacteria*, Doctorate thesis, Technische Universität Berlin, pp. 1-126.
- Rhee W. and Talagrand M. (1991), *A note on the selection of random variables under a sum constraint*, J. Appl. Probab. Vol. 28 (4), 919-923. [MR1133802](#)
- Romik D. (2014), *The Surprising Mathematics of Longest Increasing Subsequences*, Cambridge University Press, Cambridge. [MR3468738](#)
- Ross S. M. (1983), *Stochastic processes*, Wiley and sons. [MR0683455](#)
- Samuels S. M. and Steele J. M. (1981), *Optimal sequential selection of a monotone sequence from a random sample*, Ann. Probab. Vol. 9 (6), 937-947. [MR0632967](#)
- Stanke M. (2004), *Sequential selection of random vectors und a sum constraint*, J. Appl. Prob., Vol. 41 (1), 131-146. [MR2036277](#)
- Steele J. M. (2016), *The Bruss-Robertson Inequality: Elaborations, Extensions, and Applications*, Math. Applicanda, Vol. 44 (1), 3-16. [MR3557088](#)
- Truong L. V. and Tan V. Y. F. (2018) *On Gaussian MACs with Variable-Length Feedback and Non-Vanishing Error Probabilities*, arxiv:[1609.000594v3](#). [MR3782257](#)



- Tunc S., Donmez M. A. and Kozat S. S. (2013) *Optimal investment under transaction cost: A threshold-rebalanced portfolio approach*, IEEE Transactions on Signal Processing, Vol. 61 (12), 3129-3142.
- Zhang J. J. J. (2019) *Online Resource Allocation with Stochastic Resource Consumption*, Department of Technology, Stern School of Business, New York Univ., DOI: [10.13140/RG.2.2.27542.09287](https://doi.org/10.13140/RG.2.2.27542.09287).