THE BRUSS-ROBERTSON INEQUALITY: ELABORATIONS, EXTENSIONS, AND APPLICATIONS

J. MICHAEL STEELE

Abstract. The Bruss-Robertson inequality gives a bound on the maximal number of elements of a random sample whose sum is less than a specified value. The extension of that inequality which is given here neither requires the independence of the summands nor requires the equality of their marginal distributions. A review is also given of the applications of the Bruss-Robertson inequality, especially the applications to problems of combinatorial optimization such as the sequential knapsack problem and the sequential monotone subsequence selection problem.

Key Words. Order statistical inequalities, sequential knapsack problem, sequential monotone subsequence problem, sequential selection, online selection, Markov decision problems, resource dependent branching processes, Bellman equation.

Mathematics Subject Classification (2010). Primary: 60C05, 60G40, 90C40; Secondary: 60F99, 90C27, 90C39

1. Bruss-Robertson Inequality

Here, at first, we consider a finite sequence of non-negative independent random variables X_i, $i = 1, 2, \dots, n$ with a common continuous distribution function F, and, given a real value $s > 0$, we are primarily concerned with the random variable

$$M_n^*(s) = \max\{\, |A| : \sum_{i \in A} X_i \le s\}, \tag{1}$$

where, as usual, we use $|A|$ to denote the cardinality of the subset of integers $A \subset [n] = \{1, 2, \dots, n\}$. We call $M_n^*(s)$ the Bruss-Robertson maximal function, and, one should note that in terms of the traditional order statistics,

$$X_{n,1} < X_{n,2} < \cdots < X_{n,n},$$

one can also write $M_n^*(s) = \max\{k : X_{n,1} + X_{n,2} + \cdots + X_{n,k} \le s\}$.

In Bruss and Robertson (1991) it was found that the expectation of the maximal function $M_n^*(s)$ has an elegant bound in terms of the distribution function F and a natural threshold value $t(n, s)$ that one defines by the implicit relation

$$n \int_0^{t(n,s)} x \, dF(x) = s. \tag{2}$$

Specifically, one learns from Bruss and Robertson (1991, p. 622) that

$$\mathbb{E}[M_n^*(s)] \le nF(t(n, s)), \tag{3}$$

Date: This version: November 29, 2015.

J. M. Steele: Department of Statistics, The Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA, 19104. Email address: `steele@wharton.upenn.edu`.

and the main goal here is to explore this inequality with an eye toward its mastery, its extensions and its combinatorial applications.

To gain a quick appreciation of the potential of the bound (3), it is useful to take F to be the uniform distribution on $[0,1]$. By (2) we have $t(n,s) = (2s/n)^{1/2}$ provided that $(2s/n)^{1/2} \le 1$, so for $s = 1$ we find from (3) that for uniformly distributed random variables one always has

$$\mathbb{E}[\max\{|A| : \sum_{i \in A \subset [n]} X_i \le 1\}] \le \sqrt{2n}. \tag{4}$$

This tidy bound already points the way to some of the most informative combinatorial applications of the Bruss-Robertson inequality (3).

The next section elaborates on the proof of the Bruss-Robertson maximal inequality, and in Section 3 we then see how the argument of Section 2 needs only minor modifications in order to provide an inequality of unexpected generality. After illustrating this new inequality with three examples in Section 4, we turn in Section 5 to the combinatorial applications. Finally, Section 6 recalls other applications of the Bruss-Robertson maximal function, including recent applications to the theory of resource dependent branching processes and the mathematical models of societal organization.

2. An Elaboration of the Original Proof

The original proof of the Bruss-Robertson inequality (3) is not long or difficult, but by a reformulation and elaboration of that proof one does gain some concrete benefits. These benefits are explained in detail in the next section, so, for the moment, we just focus on the proof of (3).

First, by the continuity of the joint distribution of $(X_i : i \in [n])$, one finds that there is a unique set $A \subset [n]$ that attains the maximum in the definition (1) of $M_n^*(s)$. We denote this subset by $A(n,s)$, and we also introduce a second set $B(n,s) \subset [n]$ that we define by setting

$$B(n,s) = \{i : X_i \le t(n,s)\}, \tag{5}$$

where $t(n,s)$ is the threshold value determined by the implicit relation (2).

The idea behind the proof of the maximal inequality (3) is to compare the sets $A(n,s)$ and $B(n,s)$, together with their associated sums,

$$S_{A(n,s)} = \sum_{i \in A(n,s)} X_i \quad \text{and} \quad S_{B(n,s)} = \sum_{i \in B(n,s)} X_i. \tag{6}$$

Here it is useful to note that by the definitions of these sums one has the immediate relations

$$S_{A(n,s)} \le s \quad \text{and} \quad \mathbb{E}[S_{B(n,s)}] = n \int_0^{t(n,s)} x \, dF(x) = s. \tag{7}$$

Now, by its definition, $S_{A(n,s)}$ is a partial sum of order statistics. Moreover, since the summands of $S_{B(n,s)}$ consist precisely of the values $X_{n,i}$ with $X_{n,i} \le t(n,s)$, we see that $S_{B(n,s)}$ is also equal to a partial sum of order statistics of the order statistics of $\{X_1, X_2, \ldots, X_n\}$. These observations will help us with estimations that depend on the relative sizes of the two sums $S_{A(n,s)}$ and $S_{B(n,s)}$, since, in particular, one must have either $A \subset B$ or $B \supseteq A$.

If $S_{B(n,s)} \le S_{A(n,s)}$ then one has $B(n,s) \subset A(n,s)$. Moreover, the summands X_i with $i \in A(n,s) \setminus B(n,s)$ are all bounded below by $t(n,s)$, so we have the bound

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

Similarly, if $S_{A(n,s)} \le S_{B(n,s)}$ 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, we have the bound

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

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 relation

$$t(n,s)\{|A(n,s)| - |B(n,s)|\} \le S_{A(n,s)} - S_{B(n,s)}. \tag{8}$$

Here $t(n,s) > 0$ is a constant, $|A(n,s)| = M_n^*(s)$, and by (7) the right-hand side has non-positive expectation, so taking the expectations in (8) gives us

$$\mathbb{E}[M_n^*(s)] \le \mathbb{E}[|B(n,s)|] = \mathbb{E}[\sum_{i=1}^n \mathbb{1}(X_i \le t(n,s))] = nF(t(n,s)),$$

and the proof of the Bruss-Robertson inequality (3) is complete.

3. Extension of the Bruss-Robertson Inequality

The preceding argument has been organized so that it may be easily modified to give a bound that is notably more general. Specifically, one does not need independence for the Bruss-Robertson inequality (3). Moreover, after an appropriate modification of the definition of $t(n,s)$, one does not need to require that the observations have a common distribution.

Theorem 1 (Extended Bruss-Robertson Inequality)**.** *If for each $i \in [n]$ the non-negative random variable X_i has the continuous distribution F_i, and if one defines $t(n,s)$ by the implicit relation*

$$s = \sum_{i=1}^n \int_0^{t(n,s)} x \, dF_i(x), \tag{9}$$

then one has

$$\mathbb{E}[\max\big\{|A| : \sum_{i \in A \subset [n]} X_i \le s\big\}] \le \sum_{i=1}^n F_i(t(n,s)). \tag{10}$$

When the random variables X_i, $i \in [n]$, have a common distribution, then the defining condition (9) for $t(n,s)$ just recaptures the classical definition (2) of the traditional threshold value. In the same way, the upper bound in (10) also recaptures the upper bound of the original Bruss-Robertson inequality (3).

The proof of Theorem 1 requires only some light modifications of the argument of Section 2. Just as before, one defines $B(n,s)$ by (5), but now some additional care is needed with the definition of $A(n,s)$.

To keep as close as possible to the argument of Section 2, we first define a total order on the set $\{X_i : i \in [n]\}$ by writing $X_i \prec X_j$ if either one has $X_i < X_j$, or if one has both $X_i = X_j$ and $i < j$. Using this order, there is now a unique permutation $\pi : [n] \to [n]$ such that

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

and one can then take $A(n,s)$ to be largest set $A \subset [n]$ of the form

$$A = \{\pi(i) : X_{\pi(1)} + X_{\pi(2)} + \cdots + X_{\pi(k)} \le s\}. \tag{11}$$

Given these modifications, one can then proceed with the proof of key inequality (8) essentially without change. We use the same definitions (6) for the sums $S_{A(n,s)}$ and $S_{B(n,s)}$, so by the new definition (9) of $t(n,s)$, one now has

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

Since we still have $S_{A(n,s)} \le s$, the expectation on the right side of (8) is non-positive, and one can complete the proof of Theorem 1 just as one completed the proof (3) in Section 2.

As the organization of Section 2 makes explicit, none of the required calculations depend on the joint distribution of $(X_i : i \in [n])$. More specifically, one just needs to note that the argument of Section 2 depends exclusively on pointwise bounds and the linearity of expectation.

4. Three Illustrative Examples

There are times when it is difficult to solve the non-linear relation (9) for $t(n,s)$. Nevertheless, there are also informative situations where this does not pose a problem, and here we consider three examples. The first example shows that one can deal quite easily with uniformly distributed random variables with multiple scales. The other two examples show that when one considers dependent random variables, some curious new phenomena can arise.

Example 1. Basic Benefits

Here, for each $i \in [n]$ we take X_i to be uniformly distributed on the real interval $[0,i]$, but we do not require that these random variables to be independent. If we also take $0 < s \le 1$ and take $n \ge 4$ (for later convenience), then the defining condition (9) tells us

$$s = \frac{1}{2} \sum_{i=1}^{n} \frac{1}{i} t^2(n,s) = \frac{1}{2} t^2(n,s) H_n \quad \text{or} \quad t(n,s) = \sqrt{2s/H_n}, \tag{12}$$

where as usual H_n denotes the n'th harmonic number. In particular, for $s = 1$ the bound (10) tells us that

$$\mathbb{E}[\max\big\{|A| : \sum_{i \in A} X_i \le 1\big\}] \le \sum_{i=1}^{n} F_i(t(n,1)) = \sum_{i=1}^{n} \frac{1}{i} (2/H_n)^{1/2} = (2H_n)^{1/2},$$

where we use $n \ge 4$ to assure that $H_n > 2$ so the formula (12) gives us an $s \in (0,1)$. This bound offers an informative complement to (4), and, here again, one may underscore that no independence is required for this inequality. The bound depends only on the marginal distributions of the X_i, $i \in [n]$.

Example 2. Extreme Dependence

Here we take X to have the uniform distribution on $[0,1]$, and we set $X_i = X$ for all $i \in [n]$. For specificity, we take $s = 1$, and from (9) we find that one has $t(n,s) = (2s/n)^{1/2}$. Thus, just as one found for a sample of n independent, uniformly distributed random variables, the upper bound provided by (10) is given by $(2n)^{1/2}$.

Nevertheless, in this case the bound is not at all sharp. To see how poorly it does, we first note that

$$M_n^*(1) = \max\{|A| : \sum_{i \in A} X_i \leq 1|] = \min\{n, \lfloor 1/X \rfloor\}. \tag{13}$$

To evaluate the expectation of $M_n^*(1)$, we first recall that there is a useful variation of the usual formula for Euler's constant $\gamma = 0.5772\ldots$ which was discovered by Pólya (1917) and which tells us that

$$\int_0^1 \left\{ \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor \right\} dx = 1 - \gamma.$$

Consequently, if we write the domain of integration as $[0, 1/n] \cup [1/n, 1]$ and note that the integrand is bounded by 1, then we have

$$\int_0^1 \left\{ \min(n, \frac{1}{x}) - \min(n, \lfloor \frac{1}{x})\rfloor) \right\} dx = 1 - \gamma + O(\frac{1}{n}).$$

The integral of the first term equals $1 + \log n$, so upon returning to (13) one finds

$$\mathbb{E}[M_n^*(1)] = \mathbb{E}[\min\{n, \lfloor 1/X \rfloor\}] = \log n + \gamma - O(\frac{1}{n}). \tag{14}$$

When we compare this to the $(2n)^{1/2}$ bound that we get from (10), we see that it falls uncomfortably far from the actual value of $\mathbb{E}[M_n^*(1)]$. This illustrates in a simple way that there is a price to be paid for the generality of Theorem 1.

One could have come to a similar conclusion with estimates that are less precise than (14). Nevertheless, there is some independent benefit to seeing Euler's constant emerge from the knapsack problem. More critically, this example illustrates the reason for the more refined definition of $A(n,s)$ that was introduced in (11). Here the maximum in (13) is typically attained for many different choices of $A \subset [n]$. Nevertheless, with help from the total order $\prec$ one regains uniqueness in definition of $A(s,n)$, and, as a consequence, the logic of Section 1 serves one just as well as it did before.

Example 3. Beta Densities and a Long Monotone Sequence

Now, for each $i \in [n]$ we take X_i to have the $\texttt{Beta}(i, n-i+1)$ density, so in particular, X_i has the same marginal distribution as the i'th smallest value $U_{(i)}$ in a sample $\{U_1, U_2, \ldots, U_n\}$ of n independent random variables with the uniform distribution on $[0,1]$. Still, for the moment we make no assumption about the joint distribution of $(X_i : i \in [n])$. By the condition (9) we then have

$$\begin{aligned} s = \sum_{i=1}^n \int_0^{t(n,s)} x \, dF_i(x) &= \mathbb{E}\big[\sum_{i=1}^n U_{(i)} \mathbb{1}[U_{(i)} \in [0, t(n,s)]\big] \\ &= \mathbb{E}\big[\sum_{i=1}^n U_i \mathbb{1}[U_i \in [0, t(n,s)]\big] = \frac{1}{2} t^2(n,s) n, \end{aligned}$$

so in this case we again find $t(n, s) = (2s/n)^{1/2}$. Thus, by (10) we have the upper bound $(2n)^{1/2}$ when $s = 1$, and our bound echoes what we know from the classical inequality (4).

This inference depends only on our assumption about the marginal distributions, but we can go a bit further if we assume the equality of the joint distributions $(X_i : i \in [n])$ and $(U_{(i)} : i \in [n])$. In particular, one finds in this case that our upper bound $(2n)^{1/2}$ is essentially tight.

It is also evident in this case that one has $X_1 < X_2 < \cdots < X_n$, and this observation would be quite uninteresting, except that in the next section we will find in that in the independent case there is a remarkable link between monotone subsequences of maximal length and the Bruss-Robertson inequality. Thus, it is something of a curiosity to see how thoroughly this connection can be broken while still retaining the bound given by the general inequality of Theorem 1.

5. Sequential Subsequence Selection Problems

A basic source of interest in the Bruss-Robertson inequality (3) and its generalization (10) is that these results lead to *a priori* upper bounds for two well-studied problems in combinatorial optimization. In particular, in the classical case of independent uniformly distributed random variables, the Bruss-Robertson inequality (3) gives bounds that are that are essentially sharp for both the sequential knapsack problem and the sequential increasing subsequence selection problem.

In the sequential knapsack problem, one observes a sequence of n independent non-negative random variables $X_1, X_2, \ldots, X_n$ 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. The observations are observed sequentially, and, at time i, 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 maximize the expected *number* of items that are included in the knapsack. Since the Bruss-Robertson maximal function (1) tells one how well one could do if one knew in advance all of the values $\{X_i : i \in [n]\}$, it is evident that no strategy for making sequential choices can ever lead to more affirmative choices than $M_n(x)$.

The sequential knapsack problem is a Markov decision problem that is known to have an optimal sequential selection strategy that is given by a unique non-randomized Markovian decision rule. When one follows this optimal policy 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 $v_n(x)$. This is called the value function for the Markov decision problem, and, it can be calculated by the recursion relation

$$v_n(x) = (1 - F(x))v_{n-1}(x) + \int_0^x \max\{v_{n-1}(x), 1 + v_{n-1}(x - y)\}\, dF(y). \tag{15}$$

Specifically, one begins with the obvious relation $v_0(x) \equiv 0$, and one computes $v_n(x)$ by iteration of (15).

This is called the Bellman equation (or optimality equation) for the sequential knapsack problem, and it is easy to justify. The first term comes from the possibility that X_1 is too large to fit into the knapsack, and this event happens with probability $1-F(x)$. In this case, one cannot accept X_1, so one is left with the original capacity x and there are only $n-1$ more values to be observed. This gives one the first term of (15).

For the more interesting second term of (15), we consider the case where one has $X_1 = y \le x$, so one has the option either to accept or to reject X_1. If we reject X_1, we have no increment to our knapsack count and we have the value $v_{n-1}(x)$ for the expected number of selections from the remaining values. On the other hand, if we accept X_1, we have added 1 to our knapsack count. We also have a remaining capacity of $x - y$, and we have $n-1$ observations to be seen. One takes the better of these two values, and this gives us the second term of (15).

Now we consider the problem of sequential selection of a monotone decreasing subsequence. Specifically, we observe sequentially n independent random variables $X_1, X_2, \dots, X_n$ with the common continuous distribution F, and we make monotonically decreasing choices

$$X_{i_1} > X_{i_2} > \cdots > X_{i_k}.$$

Our goal here is to maximize the expected number of choices that we make. Again we have a Markov decision problem with an unique optimal non-randomized Markov decision policy. 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 $\widetilde{v}_n(x)$ for the expected number of selections made under the optimal policy when the state variable is x and where there are n observations that remain to be observed. In this case the Bellman equation given by Samuels and Steele (1981) can be written as

$$\widetilde{v}_n(x) = (1-F(x))\widetilde{v}_{n-1}(x) + \int_0^x \max\{\widetilde{v}_{n-1}(x), 1+\widetilde{v}_{n-1}(y)\}\, dF(y), \tag{16}$$

where again one has the obvious relation $\widetilde{v}_n(x) \equiv 0$ for the initial value. In (16) the decision to select $X_1 = y$ would move the state variable to y, so here we have $1+\widetilde{v}_{n-1}(y)$ where earlier we had the term $1+v_{n-1}(x-y)$ in the knapsack Bellman equation (15). In knapsack problem the state variable moves from x to $x-y$ when $X_1 = y$ is selected.

In general, the solutions of (15) and (16) are distinct. Nevertheless, Coffman, Flatto and Weber (1987) observed that $v_n(x)$ and $\widetilde{v}_n(x)$ are equal when the observations are uniformly distributed. This can be proved formally by an inductive argument that uses the two Bellman equations (15) and (16).

Proposition 2. *If $F(x) = x$ for $0 \le x \le 1$, then one has*

$$v_n(x) = \widetilde{v}_n(x) \quad \textit{for all } n \ge 0 \textit{ and } 0 \le x \le 1.$$

Proof. For $n = 0$ we have $v_0(x) = \widetilde{v}_0(x) = 0$ for all $x \in [0,1]$, and this gives us the base case for an induction. To make the inductive step from $n-1$ to n, we first use the Bellman equation (15) and then use the induction hypothesis to get

$$\begin{aligned} v_n(x) &= (1-x)v_{n-1}(x) + \int_0^x \max\{v_{n-1}(x), 1+v_{n-1}(x-y)\}\, dy \\ &= (1-x)\widetilde{v}_{n-1}(x) + \int_0^x \max\{\widetilde{v}_{n-1}(x), 1+\widetilde{v}_{n-1}(x-y)\}\, dy \\ &= (1-x)\widetilde{v}_{n-1}(x) + \int_0^x \max\{\widetilde{v}_{n-1}(x), 1+\widetilde{v}_{n-1}(y)\}\, dy = \widetilde{v}_n(x), \end{aligned}$$

where in passing to the last line one uses the symmetry of the uniform measure on $[0,x] \subset [0,1]$. Naturally, for the last equality one just needs to use the second Bellman equation (16). □

Despite the equality of the value functions established by this proposition, no one has yet found any direct choice-by-choice coupling between the sequential knapsack problem and the sequential monotone subsequence selection problem. Nevertheless, one can create a detailed linkage between these two problems that does yield more than just the equality of the associated expected values.

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 one still finds useful distributional relationships.

Threshold Strategies from Value Functions

The essential observation is that the second term of the Bellman equation (16) 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 (cf. Samuels and Steele (1981)) that there is a unique $y \in [0,1]$ that solves the "equation of indifference":

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

We denote this solution by $\alpha_n(x)$, and we 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 $X_i, X_{i+1}, \ldots, X_n$, and if we let $\widetilde{S}_{i-1}$ denote the last of the values $X_1, X_2, \ldots, X_{i-1}$ that has been selected so far, then we can only select X_i if it is not greater than the most recently selected value $\widetilde{S}_{i-1}$. In fact, one would choose to select X_i if and only if it falls in the interval $[S_{i-1}, S_{i-1} - \alpha_{n-i+1}(\widetilde{S}_{i-1})]$. Thus, the actual number of values selected out of the original n is the random variable given by

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

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

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

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

By this parallel construction and by Proposition 2, we have

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

but considerably more is true. Initially, one has $S_0 = 1 = \widetilde{S}_0$, so, one further finds the equality of the joint distributions of the vectors

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

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

This equivalence tells us in turn that the partial sums

$$V_{n,k} \stackrel{\text{def}}{=} \sum_{i=1}^{k} \mathbb{1}(X_i \in [0, \alpha_{n-i+1}(S_{i-1})]) \quad \text{and}$$

$$\widetilde{V}_{n,k} \stackrel{\text{def}}{=} \sum_{i=1}^{k} \mathbb{1}(X_i \in [\widetilde{S}_{i-1},\, \widetilde{S}_{i-1} - \alpha_{n-i+1}(\widetilde{S}_{i-1})]),$$

satisfy the distributional identity

$$V_{n,k} \stackrel{\text{d}}{=} \widetilde{V}_{n,k} \quad \text{for all } 1 \le k \le n. \tag{17}$$

Nevertheless, the two processes $\{V_{n,k} : 1 \le k \le n\}$ and $\{\widetilde{V}_{n,k} : 1 \le k \le n\}$ are not equivalent as processes. Despite the equality of the marginal distributions, the joint distributions are wildly different.

Classical Consequences

The Bruss-Robertson inequality (4) tells us directly that $\mathbb{E}[V_n] \le \sqrt{2n}$, so, by the distributional identity (17), we find indirectly that

$$\mathbb{E}[\widetilde{V}_n] \le \sqrt{2n} \quad \text{for all } n \ge 1. \tag{18}$$

It turns out that (18) can be proved by a remarkable variety of methods. In particular, Gnedin (1999) gave a direct proof of (18) where one can even accommodate a random sample size N and where the upper bound of (18) is replaced with the natural proxy $(2\mathbb{E}[N])^{1/2}$. More recently, Arlotto, Mossel and Steele (2015) gave two further proofs of (18) 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 (17) 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 (2001) and Bruss and Delbaen (2004), it was found in Arlotto, Nguyen and Steele (2015) that one has

$$\text{Var}[\widetilde{V}_n] \sim \frac{1}{3}\sqrt{2n} \quad \text{and} \quad \frac{\widetilde{V}_n - \sqrt{2n}}{3^{-1/2}(2n)^{1/4}} \Rightarrow \text{N}(0,1). \tag{19}$$

Thus, as a consequence of the distributional identity (17) one has the same results for the knapsack variable V_n for independent observations with the uniform distribution on $[0,1]$.

It seems quite reasonable to conjecture that there are results that are analogous to (19) for the sequential knapsack problem where the driving distribution F is something other than the uniform. Nevertheless, proofs of such results are unlikely to be easy since the proofs of the relations (19) required a sustained investigation of the Bellman equation (16).

Finally, one should also recall the *non-sequential* (or clairvoyant) selection problem where one studies the random variable

$$L_n = \max\{k : X_{i_1} < X_{i_2} < \ldots < X_{i_k},\ 1 \le i_1 < i_2 < \cdots < i_k \le n\}.$$

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

$$\mathbb{E}[L_n] = 2\sqrt{n} - \alpha n^{1/6} + o(n^{1/6}) \quad \text{where } \alpha = 1.77108... \tag{20}$$

Ironically, it is still not known if the map $n \mapsto \mathbb{E}[L_n]$ is concave, even though this seems like an exceptionally compelling conjecture. The estimate (20) certainly suggests this, and, moreover, we already know from Arlotto, Nguyen and Steele (2015, p. 3604) that for the analogous sequential problems the corresponding map $n \mapsto \mathbb{E}[\widetilde{V}_n] = \mathbb{E}[V_n]$ is indeed concave.

6. Further Connections and Applications

Here the whole goal has been to explain, extend, and explore the upper bound given by the Bruss-Robertson inequality. Still, there is a second side to the Bruss-Robertson maximal function, and both lower bounds and asymptotic relations have been developed in investigations by Coffman, Flatto and Weber (1987), Bruss and Robertson (1991), and Rhee and Talagrand (1991). Furthermore, Boshuizen and Kertz (1999) have even established the joint convergence in distribution of the (suitably normalized) Bruss-Robertson maximal function and a sequence of approximate solutions to the sequential knapsack problem, although this result does not seem to speak directly to the problem of proving a wider analog of (19).

The applications that have been considered here were all taken from combinatorial optimization. Nevertheless, there are several other areas where the Bruss-Robertson maximal function (1) has a natural place. For example, Gribonval, Cevher and Davies (2012) use bounds from Bruss and Robertson (1991) in their study of compressibility of high dimensional distributions.

Finally, one should note that the bounds of Bruss and Robertson (1991) have a natural role in the theory of resource dependent branching processes, or RDBPs. This is a rich theory that in turn takes a special place in the recent work of Bruss and Duerinckx (2015) on new mathematical models of societal organization. These models have been further explained for a broader (but still mathematical) audience in Bruss (2014), where the theory of Bruss and Duerinckx (2015) is also applied to two contemporary European public policy issues.

Acknowledgements

The author is pleased to thank Alessandro Arlotto, Peichao Peng, Paul Shaman and an anonymous reader for many useful comments.

References

Arlotto, A., Mossel, E. and Steele, J. M. (2015), 'Quickest online selection of an increasing subsequence of specified size', to appear in *Random Structures and Algorithms (eprint arXiv:1412.7985)* .

Arlotto, A., Nguyen, V. V. and Steele, J. M. (2015), 'Optimal online selection of a monotone subsequence: a central limit theorem', *Stochastic Process. Appl.* **125**(9), 3596–3622.

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.* **12**(4), 1119–1178.

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.* **31**(1), 178–198.

Bruss, F. T. (2014), 'Grenzen einer jeden Gesellschaft', *Jahresber. Dtsch. Math.-Ver.* **116**(3), 137–152.

Bruss, F. T. and Delbaen, F. (2001), 'Optimal rules for the sequential selection of monotone subsequences of maximum expected length', *Stochastic Process. Appl.* **96**(2), 313–342.

Bruss, F. T. and Delbaen, F. (2004), 'A central limit theorem for the optimal selection process for monotone subsequences of maximum expected length', *Stochastic Process. Appl.* **114**(2), 287–311.

Bruss, F. T. and Duerinckx, M. (2015), 'Resource dependent branching processes and the envelope of societies', *Ann. Appl. Probab.* **25**(1), 324–372.

Bruss, F. T. and Robertson, J. B. (1991), '"Wald's lemma" for sums of order statistics of i.i.d. random variables', *Adv. in Appl. Probab.* **23**(3), 612–623.

Coffman, Jr., E. G., Flatto, L. and Weber, R. R. (1987), 'Optimal selection of stochastic intervals under a sum constraint', *Adv. in Appl. Probab.* **19**(2), 454–473.

Gnedin, A. V. (1999), 'Sequential selection of an increasing subsequence from a sample of random size', *J. Appl. Probab.* **36**(4), 1074–1085.

Gribonval, R., Cevher, V. and Davies, M. E. (2012), 'Compressible distributions for high-dimensional statistics', *IEEE Trans. Inform. Theory* **58**(8), 5016–5034.

Pólya, G. (1917), 'Über eine neue Weise bestimmte Integrale in der analytischen Zahlentheorie zu gebrauchen', *Nachr. Ges. Wiss. Göttingen* pp. 149–159.

Rhee, W. and Talagrand, M. (1991), 'A note on the selection of random variables under a sum constraint', *J. Appl. Probab.* **28**(4), 919–923.

Romik, D. (2014), *The Surprising Mathematics of Longest Increasing Subsequences*, Cambridge University Press, Cambridge.

Samuels, S. M. and Steele, J. M. (1981), 'Optimal sequential selection of a monotone sequence from a random sample', *Ann. Probab.* **9**(6), 937–947.