Systematic Enumeration of Fundamental Quantities Involving Runs in Binary Strings

Félix Balado and Guénolé C.M. Silvestre
School of Computer Science
University College Dublin
Belfield Campus
Dublin 4, Ireland
felix@ucd.ie
guenole.silvestre@ucd.ie

Abstract

We give recurrences, generating functions and explicit exact expressions for the enumeration of fundamental quantities involving runs in binary strings. We first focus on enumerations concerning runs of ones, and we then analyse the same enumerations when runs of ones and runs of zeros are jointly considered. We give the connections between these two types of run enumeration, and with the problem of compositions. We also analyse the same enumerations with a Hamming weight constraint. We discuss which of the many number sequences that emerge from these problems are already known and listed in the OEIS. Additionally, we extend our main enumerative results to the probabilistic scenario in which binary strings are outcomes of independent and identically distributed Bernoulli variables.

1 Introduction

Runs in binary strings are essentially uninterrupted sequences of the same bit. In this paper we give recurrences, generating functions and explicit expressions for the enumeration of fundamental quantities involving runs in binary strings. A lot of prior work on this topic is probabilistic in nature, in which case runs of ones are typically called *success* runs [**?**] —whereas runs of zeros are called *failure* runs. The very first historical problem on runs, which was studied by de Moivre [**?**], was in fact a success runs problem. Whereas some authors focus on problems concerning success runs [**?**, **?**, **?**], others jointly consider success and failure runs [**?**, **?**, **?**, **?**, **?**].

We can see enumerative results for runs in binary strings as special cases of the above corpus of research. However, in spite of the long history of the topic, a uniform and consistent approach to enumerating runs-related quantities is currently lacking. Some authors have provided generating functions for specific settings [?, ?, ?, ?, ?, ?], while others have derived explicit (closed-form) expressions for some others [?, ?, ?, ?, ?] —either through direct combinatorial analysis or through generating functions. Other authors have given recurrence relations for different problems involving runs [?, ?, ?, ?, ?, ?, ?]. But many relevant enumeration aspects have apparently not been considered yet.

Our goal here is to try to fill this gap by systematically studying the most fundamental enumerations of quantities involving runs in binary strings within a common framework. Our strategy always involves: 1) finding one or more recurrences for each of the problems at hand; and 2) if possible, solving the recurrences to obtain exact explicit expressions by means of generating functions. This is hardly a novel approach, but one that has been frequently sidelined in favour of standard combinatorial analysis in the study of runs —which makes a number of these problems harder. The use of generating functions to study runs harks back to the earliest problem of this kind studied by de Moivre [?]. As far as we are aware, Laplace [?, p. 421] was the first author who used recurrences —or finite difference equations, as they were commonly known back then— to derive probability generating functions for problems involving runs. Later authors, prominently Feller [?, Ch. XIII], also used this approach to deal with runs, although sometimes with a definition of “run” different than the one adopted here —see Remark **??** below.

In this work, we recover many established results —often, though not always, via more streamlined derivations— while more frequently we offer novel contributions in the form of recurrences, generating functions, and explicit expressions. Although we are aware of the limitations of such explicit expressions, we do not pursue here asymptotic approximations. As we proceed, we highlight which of the enumerations that we address correspond to known number sequences —meaning that they have been documented in Sloane’s On-Line Encyclopedia of Integer Sequences (OEIS) [?].

Importantly, we also give the most basic connections between the enumerations of runs of ones, runs of ones and/or zeros, and compositions. Some of these connections have been partly identified by previous authors [?, ?, ?], but, as far as we are aware, they have not been fully investigated. Last but not least, even though our focus is on enumerations and not on probabilistic results, we show that our enumerative results (recurrences, generating functions, and explicit expressions) can be straightforwardly extended to deal with probabilistic runs, when assuming bitstrings generated by independent and identically distributed Bernoulli random variables.

1.1 Structure of the Paper

This paper is structured as follows. In Section **??** we study the number of binary strings of length n that contain prescribed quantities of runs of ones under different constraints. In Section **??** we consider the probabilistic extensions of the most relevant among the Section **??** results. We then consider in Section **??** the same problems as in Section **??**, but when the binary strings are constrained to having fixed Hamming weight. Sections **??**, **??** and **??** address the basic same problems as Sections **??**, **??** and **??**, respectively, when runs of ones and runs of zeros are jointly considered —rather than only runs of ones.

In Section **??** we study the total number of runs (under different constraints) that are found over all binary strings, or over certain subsets of these. In Section **??** we essentially address the same problems as in Section **??**, but when the goal is instead to count the number of ones contained in runs of ones, rather than the runs themselves. Finally, we draw the conclusions of this work in Section **??**.

1.2 Definitions

We start by precisely stating the definition of "run" that we use, to avoid any possible confusion with previous works.

Definition 1.1 (*Run, Length of a Run*)**.** A *run* in a binary string is an uninterrupted sequence of bits of the same kind, flanked on each side either by the opposite bit or by the start/end of the string (Mood's counting criterion [**?**]). The number of bits of the same kind in a run is its *length*.

Definition 1.2 (*n-String*)**.** To minimise excessive wording, a binary string of length n is referred to in the following as an *n-string*.

Definition 1.3 (*k-Run*)**.** A *k-run* means a run of length k, where $k \geq 0$.

Definition 1.4 (*Null Run, Nonnull Run*)**.** A *null run* is defined to be a 0-run, i.e., a zero-length run, whereas a *nonnull run* is a k-run with $k \geq 1$.

Remark. Observe that with the definition above we are just explicitly identifying the existence of zero-length runs. Although perhaps initially puzzling, null runs (0-runs) follow directly from Mood's criterion in Definition **??**. Letting $\tilde{b} = \mathrm{mod}(b+1, 2)$ with $b \in \{0, 1\}$, a null run of bs in an n-string occurs whenever a $\tilde{b}$ is immediately followed by another $\tilde{b}$, or whenever the n-string starts/ends with $\tilde{b}$. For example, $'0'010'$ contains three 0-runs of ones, $'0101$ contains one 0-run of ones, and $'0'0'0'0'$ contains five 0-runs of ones, where the apostrophes mark the positions of the null runs of ones. Even if stating the obvious, observe that a null run of bs contains no bs. The apparent lack of applications for null runs is probably the reason why they have not been considered by previous authors. However, we will see that disregarding null runs is not an option in a complete theory of runs, as they play a fundamental role in key theoretical and practical results. □

Definition 1.5 (*($\underline{k} \leq \overline{k}$)-Run, ($\geq k$)-Run, ($\leq k$)-Run*)**.** A *($\underline{k} \leq \overline{k}$)-run* is defined to be a k-run with $0 \leq \underline{k} \leq k \leq \overline{k}$.

One may verbalise a ($\underline{k} \leq \overline{k}$)-run by saying a "lower k to higher k run". A ($\geq k$)*-run* is defined to be a ($k \leq \infty$)-run, which may include null runs, whereas a ($\leq k$)*-run* is defined to be a ($1 \leq k$)-run if null runs are excluded and a ($0 \leq k$)-run otherwise. These additional definitions can be naturally verbalised as a "greater or equal than k run" and a "smaller or equal than k run," respectively. Lastly, observe that we can also express a k-run as a ($k \leq k$)-run.

Definition 1.6 (*Odd and Even Runs, p-Parity Runs*)**.** An *odd (even) run* is defined to be a run whose length is odd (even). More generally, we use the expression *odd (even) runs* when describing ensembles of runs whose lengths, which may be different, are all odd (even). For the sake of brevity in mathematical expressions, we also use the term *p-parity run*, which refers to an odd run when $p = 1$, and to an even run when $p = 0$. Throughout the paper we only consider nonnull even runs —odd runs are always nonnull.

Remark. Mood's criterion is perhaps the most common way to define a run.

However, among the authors who adhere to this criterion, some use the term "run" to denote runs formed by one single kind of bit —typically, runs of ones— whereas others jointly consider runs of ones and zeros. With the first convention, '11000111100' contains two runs of ones: a 2-run and a 4-run (or two runs of zeros: a 3-run and a 2-run), whereas with the second one it contains four runs (of ones and zeros): two 2-runs, one 3-run and one 4-run. In the first group of authors we would find most of those who deal with success runs, such as Apostol [**?**], Balakrishnan and Koutras [**?**], Makri et al. [**?**] and many others. Among authors simultaneously considering runs of ones and zeros we have Wishart and Hirschfeld [**?**], Stevens [**?**], Wald and Wolfowitz [**?**] —in their classic runs-based test—, Bloom [**?**], again Balakrishnan and Koutras [**?**], and others. The combination of Mood's criterion with two kinds of runs in binary strings is summarised in the following statement by Feller [**?**, p. 42]: *"In any ordered sequence of elements of two kinds, each maximal subsequence of elements of like kind is called a run".* It should be noted, though, that Feller also considers runs of ones (success runs) separately from runs of zeros, and that he often handles other definitions of "run"—see below.

As just mentioned, the reader must be aware that the literature contains definitions of "run" divergent from Mood's criterion. A relevant case for this paper is the definition used by Flajolet and Sedgewick [**?**, **?**] or by Nyblom [**?**], for whom a run of ones of length k in a binary string still refers to the appearance of k consecutive ones, but not necessarily terminated by a zero or by the end of the string (or started by a zero or by the start of the string). Nevertheless, when discussing n-strings devoid of such runs, as these authors do, this definition of a run of ones of length k is completely equivalent to a $(\geq k)$-run of ones according to our definition above.

Other authors consider not only definitions of "run" which do not conform to Mood's criterion, but also enumeration schemes that count overlapping runs. As a relevant example, Balakrishnan and Koutras consider Type I, II and III run enumeration schemes in their treatise on runs [**?**], of which only type II is relevant to the criterion adopted in this paper. To illustrate these schemes, consider the string '110111001111101'. With these authors' Type I scheme —originally proposed by Feller [**?**, p. 305]— a k-run does not have to end with the opposite bit or with the end of the string. Thus, according to this scheme, we may say that the aforementioned string contains four runs of ones of length 2: '$\underline{11}0\underline{11}100\underline{11}\,\underline{11}101$', which is not compatible with Mood's criterion. However, Feller's and Mood's criteria are equivalent when considering the longest run. Feller's criterion, which has often been used in runs theory, is somewhat unnatural. In fact, it was only introduced by this author to be able to study runs through renewal theory, as he considered that the classical theory of runs —based on Mood's criterion— was "messy". Next, using the Type II scheme in [**?**] we may say that the aforementioned string contains three runs of ones of length at least 2: '$\underline{11}0\underline{111}00\underline{11111}01$', which is compatible with Mood's criterion, and equivalent to us saying that it contains three (≥ 2)-runs of ones, or, equivalently, three $(2 \leq 15)$-runs of ones. On the other hand, overlapping of runs is allowed with the Type III scheme handled by Balakrishnan and Koutras (Ling's criterion [**?**]), according to which we may say, for example, that the string in question contains seven runs of ones of length 2: '$\underline{11}0\underline{1\overline{1}}\overline{1}00\underline{1\overline{1}}\,\underline{\overline{1}\overline{1}}\overline{1}01$'. Of course, this is not compatible with our criterion either.

Finally, different authors have used names other than "run" to describe runs that —usually either *avant la lettre* or unwittingly— comply with Mood's criterion. Some examples are: *sequence* [**?**], *join* [**?**], *group* [**?**], *block* [**?**, **?**], *subset of consecutive ones* [**?**], *isolated tuple* [**?**], or *clump* [**?**]. Runs with particular lengths have also been called *singles/singletons/isolated letters/isolated ones* [**?**, **?**, **?**, **?**] (i.e., 1-runs), *isolated pairs* [**?**] (i.e, 2-runs), *isolated triples* [**?**] (i.e., 3-runs), etc. The expression *maximal block/run* [**?**, **?**] has also been used to denote a run that follows Mood's criterion, but we believe that this term is best avoided in order to prevent any misunderstandings with maximum length runs. Like other previous authors [**?**, **?**], we are crediting Mood in Definition **??**, but the fact is that the use of the term "run" with the exact same meaning precedes him (see for example [**?**]), and the first explicit statement of "Mood's criterion" seems to have been subsequently done by Mosteller [**?**, p. 229]. Looser uses of the term "run" in probability are even older [**?**]. □

1.3 Notation

The binomial coefficient is defined for any two $a, b \in \mathbb{Z}$ as $\binom{a}{b} = a(a-1)\cdots(a-b+1)/b!$ if $a \geq b \geq 0$, and as $\binom{a}{b} = 0$ if $b < 0$ or $a < b$. Throughout the paper we make extensive use of Iverson's bracket notation [**?**, p. 24]: a true-or-false statement enclosed in blackboard bold straight brackets —i.e., $[\![\cdot]\!]$— takes the value 1 if the statement is true, and 0 if it is false. Finally, we follow the empty summation convention: $\sum_{i=a}^{b} = 0$ when $b < a$.

Remark. For the sake of brevity, in the remainder of this paper the unqualified term "run" always implicitly refers to a run of ones unless explicitly specified otherwise. Most of the exceptions to the aforementioned naming convention occur in Sections **??**, **??**, **??**, and **??**, where we jointly consider runs of ones and runs of zeros, and where we explicitly indicate the nature of the runs as required. Likewise, we assume that runs may be null unless explicitly indicated otherwise. □

2 Number of n-Strings that Contain Prescribed Quantities of Runs Under Different Constraints

In this section we address several fundamental enumeration problems that involve counting the number of n-strings that contain prescribed quantities of runs subject to different constraints.

2.1 Number of n-Strings that Contain Exactly m $(\underline{k} \leq \bar{k})$-Runs

We denote by $w_{\underline{k} \leq \bar{k}}(n, m)$ the number of n-strings that contain exactly m $(\underline{k} \leq \bar{k})$-runs, i.e., m runs whose lengths are within $\underline{k}$ and $\bar{k}$. Notice that such n-strings may also contain other runs longer than $\bar{k}$ or shorter than $\underline{k}$. Because of the general nature of the definition of a $(\underline{k} \leq \bar{k})$-run, the results established in this section form the foundation that allows us to address many other runs-related enumerations in a simple —and sometimes even trivial— way.

We begin by formulating a necessary condition that underpins our approach.

Necessary Condition. (Existence of n-strings containing m $(\underline{k} \leq \bar{k})$-runs)

$$w_{\underline{k} \leq \bar{k}}(n, m) > 0 \quad \Longrightarrow \quad 0 \leq m \leq \left\lfloor \frac{n+1}{\underline{k}+1} \right\rfloor. \tag{1}$$

□

The reason for the nonnegativity of m is clear, and perhaps only the case $m = 0$ merits some explanation: observe that we can have n-strings devoid of $(\underline{k} \leq \bar{k})$-runs. As for the upper bound in (**??**), if we have m runs with lengths $k_1, \ldots, k_m$ such that $\underline{k} \leq k_j \leq \bar{k}$ for $j = 1, \ldots, m$, then, because the number of zeros in an n-string that contains m runs must be at least $m - 1$, a necessary condition for these runs to fit in an n-string is

$$m\,\underline{k} + m - 1 \leq n, \tag{2}$$

which is equivalent to the upper bound in (**??**). Condition (**??**) is not sufficient, as the validity of its right-hand side does not guarantee that we can accommodate m $(\underline{k} \leq \overline{k})$-runs within an n-string. However, if the right-hand side is not true, then such accommodation is certainly impossible.

In order to streamline some formulas, it is convenient to name the difference between both sides of inequality (**??**):

$$e = n - (m\,\underline{k} + m - 1). \tag{3}$$

Of course, $e \geq 0$ is another way to express the upper bound in (**??**).

Remark. Necessary condition (**??**) does not hold if $n < -1$. However, consider the two cases

$$n = -1 \text{ and } n = 0, \tag{4}$$

for which (**??**) does not negate the possibility of existence of n-strings containing m $(\underline{k} \leq \overline{k})$-runs, as long as $m = 0$. Even if the two cases above may seem absurd —especially $n = -1$, since $n = 0$ is commonly accepted to be the degenerate case of an empty string— they play the crucial role of initialising all the recurrences that we give in this section —and also in Sections **??** and **??**. □

Finally, we also discuss a relevant enumeration that stems from the main one studied in this section: the number of n-strings that contain at least m $(\underline{k} \leq \overline{k})$-runs, which we denote by $w^{\geq m}_{\underline{k}\leq\overline{k}}(n)$, and which we may obtain from $w_{\underline{k}\leq\overline{k}}(n,m)$ using

$$w^{\geq m}_{\underline{k}\leq\overline{k}}(n) = \sum_{t=m}^{\left\lfloor \frac{n+1}{\underline{k}+1} \right\rfloor} w_{\underline{k}\leq\overline{k}}(n,t). \tag{5}$$

The case $m = 1$ can also be simply put as

$$w^{\geq 1}_{\underline{k}\leq\overline{k}} = 2^n - w_{\underline{k}\leq\overline{k}}(n,0). \tag{6}$$

The number of n-strings that contains at most m $(\underline{k} \leq \overline{k})$-runs is $w^{\leq m}_{\underline{k}\leq\overline{k}}(n) = 2^n - w^{\geq(m+1)}_{\underline{k}\leq\overline{k}}(n)$, so we do not need to discuss this complementary enumeration.

2.1.1 Recurrences

Let us first obtain a recurrence relation for $w_{\underline{k}\leq\overline{k}}(n,m)$. Consider the contribution to $w_{\underline{k}\leq\overline{k}}(n,m)$ from the ensemble of n-strings that begin

with an i-run. If $\underline{k} \leq i \leq \overline{k}$ then they contribute $w_{\underline{k}\leq\overline{k}}(n-(i+1), m-1)$ to $w_{\underline{k}\leq\overline{k}}(n,m)$. On the other hand, if $i > \overline{k}$ or $i < \underline{k}$ then they contribute $w_{\underline{k}\leq\overline{k}}(n-(i+1), m)$ to $w_{\underline{k}\leq\overline{k}}(n,m)$.

Observe that the "+1" in $(i+1)$ is there to guarantee that the first bit immediately after the i-run is a zero —i.e., to terminate the run. Thus, considering all possible lengths of a starting run we obtain the following bivariate recurrence:

$$
\begin{aligned}
w_{\underline{k}\le\bar{k}}(n,m) = & \sum_{i=\underline{k}}^{\bar{k}} w_{\underline{k}\le\bar{k}}\big(n-(i+1),m-1\big) + \sum_{i=0}^{\underline{k}-1} w_{\underline{k}\le\bar{k}}\big(n-(i+1),m\big) \\
& + \sum_{i=\bar{k}+1}^{n} w_{\underline{k}\le\bar{k}}\big(n-(i+1),m\big). \qquad (7)
\end{aligned}
$$

The natural question is: what if $i \geq n$? As we see later, as long as we take (**??**) into account, we do not have to worry about such questions with the correct initialisation of the recurrence.

An alternative way to find a recurrence for $w_{\underline{k}\le\bar{k}}(n,m)$ is based on the following observation: if we remove the first bit of all n-strings, then we can see that $w_{\underline{k}\le\bar{k}}(n,m)$ is approximately equal to $2\,w_{\underline{k}\le\bar{k}}(n-1,m)$. The overcounting or undercounting in this estimate with respect to the true enumeration only depends on the cases in which an n-string starts with a $\underline{k}$-run or with a $(\bar{k}+1)$-run, as these are the only cases in which the number of $(\underline{k}\le\bar{k})$-runs can be increased or decreased with respect to the $(n-1)$-string formed by removing the first bit. Thus, in order to make the aforementioned estimate exact we just need to make the following adjustments to it:

a) For every n-string that starts with a $\underline{k}$-run, we must add one whenever the remaining $(n-(\underline{k}+1))$-string contains $(m-1)$ $(\underline{k}\le\bar{k})$-runs, but we must subtract one whenever the remaining $(n-(\underline{k}+1))$-string contains m $(\underline{k}\le\bar{k})$-runs already.

b) For every n-string that starts with a $(\bar{k}+1)$-run, we must subtract one from the estimate whenever the remaining $(n-(\bar{k}+2))$-string contains $(m-1)$ $(\underline{k}\le\bar{k})$-runs, but we must add one whenever the remaining $(n-(\bar{k}+2))$-string contains m $(\underline{k}\le\bar{k})$-runs .

Collecting the contributions from a) and b) we get the following alternative recurrence:

$$
\begin{aligned}
w_{\underline{k}\le\bar{k}}(n,m) = 2\,w_{\underline{k}\le\bar{k}}(n-1,m) &+ w_{\underline{k}\le\bar{k}}\big(n-(\underline{k}+1),m-1\big) - w_{\underline{k}\le\bar{k}}\big(n-(\underline{k}+1),m\big) \\
&- w_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+2),m-1\big) + w_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+2),m\big). \qquad (8)
\end{aligned}
$$

Observe that recurrence (**??**) can also be derived from recurrence (**??**) using $w_{\underline{k}\le\bar{k}}(n,m) - w_{\underline{k}\le\bar{k}}(n-1,m)$. Recurrence relation (**??**) is equivalent to but simpler than (**??**): the number of recursive calls in (**??**) is n (full-history recurrence), whereas the number of recursive calls in (**??**) is five, independently of n. Also, unlike (**??**) which contains an n-dependent summation, recurrence (**??**) is directly amenable to the computation of the generating function associated to $w_{\underline{k}\le\bar{k}}(n,m)$, as we see in Section **??**. In the remainder we work with (**??**).

To initialise (**??**) we can use the case $n = 1$, in which we have by inspection that

$$w_{\underline{k}\le\bar{k}}(1,m) = \big(\llbracket \underline{k}=0\rrbracket\llbracket \bar{k}=0\rrbracket + 2\,\llbracket \underline{k}>1\rrbracket\big)\llbracket m=0\rrbracket + \big(\llbracket \underline{k}=1\rrbracket + \llbracket \underline{k}=0\rrbracket\llbracket \bar{k}\ge 1\rrbracket\big)\llbracket m=1\rrbracket \\ + \llbracket \underline{k}=0\rrbracket\llbracket m=2\rrbracket. \tag{9}$$

On the other hand, setting $n = 1$ in recurrence (**??**) we get

$$w_{\underline{k}\le\bar{k}}(1,m) = 2\,w_{\underline{k}\le\bar{k}}(0,m) + w_{\underline{k}\le\bar{k}}(-\underline{k},m-1) - w_{\underline{k}\le\bar{k}}(-\underline{k},m) \\ - w_{\underline{k}\le\bar{k}}(-(\bar{k}+1),m-1) + w_{\underline{k}\le\bar{k}}(-(\bar{k}+1),m). \tag{10}$$

We wish (**??**) to equal (**??**).

Taking (**??**) into account, we see that the desired equality is fulfilled for all m and $0 \le \underline{k} \le \bar{k}$ by choosing

$$w_{\underline{k}\le\bar{k}}(-1,m) = \llbracket m=0\rrbracket, \tag{11}$$

$$w_{\underline{k}\le\bar{k}}(0,m) = \Big\llbracket m = \llbracket \underline{k}=0\rrbracket\Big\rrbracket, \tag{12}$$

which thus constitute the initialisation of recurrence (**??**). Through the same procedure as above, the reader may verify that (**??**) and (**??**) also initialise (**??**). Even if an uncommon sight, the nested Iversonian brackets in the expression above are not a typo, and indeed we will meet again this type of expression in subsequent sections.

Finally, we can obtain a recurrence for $w^{\ge m}_{\underline{k}\le\bar{k}}(n)$ by adding recurrences (**??**) or (**??**) over the range of m in (**??**). Let us do so with the simpler recurrence (**??**). Taking into account necessary condition (**??**), we get

$$w^{\ge m}_{\underline{k}\le\bar{k}}(n) = 2\,w^{\ge m}_{\underline{k}\le\bar{k}}(n-1) + w_{\underline{k}\le\bar{k}}\big(n-(\underline{k}+1),m-1\big) - w_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+2),m-1\big). \tag{13}$$

Although the recurrence above depends on $w_{\underline{k}\le\bar{k}}(n,m)$, we show in the next section that it suffices to obtain the generating function of $w^{\ge m}_{\underline{k}\le\bar{k}}(n)$.

2.1.2 Generating Functions

Let us next obtain the bivariate ordinary generating function (ogf) associated to $w_{\underline{k}\le\bar{k}}(n,m)$, i.e.,

$$W_{\underline{k}\le\bar{k}}(x,y) = \sum_n\sum_m w_{\underline{k}\le\bar{k}}(n,m)\,x^n y^m. \tag{14}$$

Notice that we have not set summation limits in (**??**), which means that we are adding over all integers n and m. This is because, in order to streamline our task, we follow Graham et al.'s procedure in [**?**, Sec. 7.3] and apply (**??**) to a version of recurrence (**??**) valid for all values of n and m. First of all, taking necessary condition (**??**) into account, recurrence (**??**)

is valid not only for $n \geq 1$ and $m \geq 0$, but also for $n < -1$ and/or $m < 0$ —in which cases $w_{\underline{k} \leq \bar{k}}(n, m) = 0$. So let us see what happens in the remaining two cases $n = -1$ and $n = 0$, by computing $w_{\underline{k} \leq \bar{k}}(-1, m)$ and $w_{\underline{k} \leq \bar{k}}(0, m)$ using (**??**) and then comparing the results with the correct initialisation values in (**??**) and (**??**).

In the first case, from (**??**) and (**??**) we mistakenly have that $w_{\underline{k} \leq \bar{k}}(-1, m) = 0$ instead of the correct value $w_{\underline{k} \leq \bar{k}}(-1, m) = [\![m = 0]\!]$, but we can "fix" this by adding $[\![n = -1]\!][\![m = 0]\!]$ to (**??**). In the second case, according to (**??**) and using (**??**) and (**??**) we wrongly have that $w_{\underline{k} \leq \bar{k}}(0, m) = 2\,[\![m = 0]\!]$ when $\underline{k} > 0$ and $w_{\underline{k} \leq \bar{k}}(0, m) = [\![m = 0]\!] + [\![m = 1]\!]$ when $\underline{k} = 0$, rather than the correct value (**??**). Once again, we can "fix" these cases just by subtracting $[\![n = 0]\!][\![m = 0]\!]$ from (**??**). Therefore an extended version of recurrence (**??**) valid for all values of n and m when taking (**??**) into account is

$$\begin{aligned} w_{\underline{k} \leq \bar{k}}(n, m) = 2\, w_{\underline{k} \leq \bar{k}}(n-1, m) &+ w_{\underline{k} \leq \bar{k}}(n - (\underline{k}+1), m-1) - w_{\underline{k} \leq \bar{k}}(n - (\underline{k}+1), m) \\ &- w_{\underline{k} \leq \bar{k}}(n - (\bar{k}+2), m-1) + w_{\underline{k} \leq \bar{k}}(n - (\bar{k}+2), m) \\ &+ [\![m = 0]\!]\big([\![n = -1]\!] - [\![n = 0]\!]\big). \end{aligned} \tag{15}$$

This same recurrence is obtained if we first make (**??**) valid for all n and m and then obtain $w_{\underline{k} \leq \bar{k}}(n, m) - w_{\underline{k} \leq \bar{k}}(n-1, m)$ using that extended recurrence. Considering (**??**), we can now get $W_{\underline{k} \leq \bar{k}}(x, y)$ just by multiplying (**??**) on both sides by $x^n y^m$ and then adding over n and m. This yields

$$W_{\underline{k} \leq \bar{k}}(x, y) = 2x\, W_{\underline{k} \leq \bar{k}}(x, y) + x^{\underline{k}+1}(y-1)\, W_{\underline{k} \leq \bar{k}}(x, y) - x^{\bar{k}+2}(y-1)\, W_{\underline{k} \leq \bar{k}}(x, y) + x^{-1} - 1,$$

and thus the ogf sought is

$$W_{\underline{k} \leq \bar{k}}(x, y) = \frac{1 - x}{x\big(1 - 2x + (1-y)(x^{\underline{k}+1} - x^{\bar{k}+2})\big)}. \tag{16}$$

It is not difficult to identify the coefficient of $W_{\underline{k} \leq \bar{k}}(x, y)$ that corresponds to y^m, i.e., $[y^m]W_{\underline{k} \leq \bar{k}}(x, y)$. To this end we rewrite (**??**) as

$$W_{\underline{k} \leq \bar{k}}(x, y) = \frac{1 - x}{x\big(1 - 2x + x^{\underline{k}+1} - x^{\bar{k}+2}\big)} \cdot \frac{1}{(1 - cy)}$$

with

$$c = \frac{x^{\underline{k}+1} - x^{\bar{k}+2}}{1 - 2x + x^{\underline{k}+1} - x^{\bar{k}+2}}.$$

Then, because from the negative binomial theorem we have that $(1 - cy)^{-1} = \sum_{m \geq 0} (cy)^m$ it follows that

$$[y^m]W_{\underline{k} \leq \bar{k}}(x, y) = \frac{(1 - x)\big(x^{\underline{k}+1} - x^{\bar{k}+2}\big)^m}{x\big(1 - 2x + x^{\underline{k}+1} - x^{\bar{k}+2}\big)^{m+1}}, \tag{17}$$

which is the ogf enumerating the binary strings that contain exactly m ($\underline{k} \leq \bar{k}$)-runs.

Finally, we can obtain the ogf $W^{\geq m}_{\underline{k}\leq\bar{k}}(x) = \sum_x w^{\geq m}_{\underline{k}\leq\bar{k}}(n)\, x^n$ by multiplying (**??**) by x^n on both sides and then adding on n. This gives

$$\begin{aligned} W^{\geq m}_{\underline{k}\leq\bar{k}}(x) &= \frac{x^{\underline{k}+1} - x^{\bar{k}+2}}{1-2x}\,[y^{m-1}]W_{\underline{k}\leq\bar{k}}(x,y) \\ &= \frac{1-x}{(1-2x)\,x}\left(\frac{x^{\underline{k}+1} - x^{\bar{k}+2}}{1-2x+x^{\underline{k}+1} - x^{\bar{k}+2}}\right)^m, \end{aligned} \tag{18}$$

where we have used (**??**) to get the last expression. This is the ogf enumerating the binary strings that contain at least m ($\underline{k} \leq \bar{k}$)-runs—see (**??**).

2.1.3 Explicit Expressions

It is possible, in principle, to get explicit expressions for $w_{\underline{k}\leq\bar{k}}(n,m)$ from (**??**) —and for $w^m_{\underline{k}\leq\bar{k}}(n)$ from (**??**)— using the same standard method that we adopt in many subsequent sections. This method is, essentially, the repeated application of the (negative) binomial theorem, plus, in some cases, the solution of one or more simple Diophantine equations. However, the reader may verify that, in this general case, this strategy leads to overcomplicated expressions, which depend on whether $\gcd(\bar{k}-\underline{k}+1,\underline{k})$ is equal to or greater than one. Consequently, we only obtain explicit expressions for the special cases of $w_{\underline{k}\leq\bar{k}}(n,m)$ considered in Sections **??**–**??**.

2.2 Number of n-Strings that Contain Exactly m k-Runs

We denote the number of n-strings that contain exactly m k-runs by $w_k(n,m)$. Notice that such n-strings may also contain other runs of lengths different than k.

The earliest reference that we know of for this enumeration is the work of Apostol [**?**], who gave several recurrences and a generating function for $w_k(n,m)$ motivated by an electrical engineering problem. This relevant piece of work appears to have faded into obscurity, perhaps due to its idiosyncratic naming conventions. Koutras and Papastavridis [**?**] recovered some of Apostol's results, but most later authors were unaware of them. They also were unaware of the first closed-form formula for $w_k(n,m)$ —a triple-summation expression obtained as a special case of a more general computation— given by Magliveras and Wei [**?**, Thm. 2.3]. Sinha and Sinha [**?**] produced a triple-summation explicit expression for $w_k(n,m)$, drawing on a generating function for the interstitial gaps between the m k-runs.

Soon afterwards, Makri and Psillakis [**?**] gave a simpler double-summation explicit formula for $w_k(n,m)$. Their derivation exploited an existing combinatorial result of their own for the distribution of balls in urns [**?**]. More recently, Madden [**?**] has given a generating function for the equivalent problem of enumerating the n-strings *that begin with zero* and contain a prescribed number of

runs of a given length.

As far as we are concerned, this problem is a special case of the enumeration in Section **??** with $\underline{k} = \overline{k} = k$, and thus

$$w_k(n,m) = w_{k \le k}(n,m). \tag{19}$$

We also study the number of n-strings that contain at least m k-runs, which from (**??**) and (**??**) is

$$w_{\overline{k}}^{\ge m}(n) = \sum_{t=m}^{\left\lfloor \frac{n+1}{k+1} \right\rfloor} w_k(n,t). \tag{20}$$

2.2.1 Recurrences

From (**??**), (**??**) and (**??**), two recurrence relations for $w_k(n,m)$ are

$$w_k(n,m) = w_k\big(n-(k+1), m-1\big) + \sum_{\substack{i=0 \\ i \ne k}}^{n} w_k\big(n-(i+1), m\big), \tag{21}$$

and

$$\begin{aligned} w_k(n,m) = 2\, w_k(n-1,m) + w_k\big(n-(k+1), m-1\big) - w_k\big(n-(k+1), m\big) \\ - w_k\big(n-(k+2), m-1\big) + w_k\big(n-(k+2), m\big), \end{aligned} \tag{22}$$

which, from (**??**) and (**??**), are both initialised by

$$w_k(-1,m) = [\![m=0]\!], \tag{23}$$

$$w_k(0,m) = \big[\![m = [\![k=0]\!]\big]\!]. \tag{24}$$

Remark. Apostol gives a host of recurrences for this problem [**?**, Thms. 2–11]. In Apostol's notation, $A_m^{(k)} = w_k(n,m)$, but he also uses $S_m(n) = w_1(n,m)$, $P_m(n) = w_2(n,m)$ and $T_m(n) = w_3(n,m)$. This author gives several recurrences for enumerating the n-strings with prescribed numbers of *isolated singletons* (i.e., 1-runs): in particular, he gives recurrences for $w_1(n,0)$ with $n \ge 3$, for $w_1(n,1)$ with $n \ge 3$, for $w_1(n,2)$ with $n \ge 3$, and for $w_1(n,m)$ with $n \ge 3$ and $m \ge 2$. He then gives recurrences for enumerating n-strings with prescribed numbers of *isolated pairs* (i.e., 2-runs), in particular, recurrences for $w_2(n,0)$ with $n \ge 3$, for $w_2(n,1)$ with $n \ge 3$, for $w_2(n,2)$ with $n \ge 3$, and for $w_2(n,m)$ with $n \ge 3$ and $m \ge 2$. Finally, he extends these recurrences to *isolated k-tuples* (i.e., k-runs). He first gives a recurrence for $w_k(n,0)$ with $n \ge k+2$, then a recurrence for $w_k(n,1)$ with $n \ge k+2$, lastly a recurrence for $w_k(n,m)$ with $m \ge 2$, $k \ge 1$ and $n \ge mk+m-1$ [in our notation, $e \ge 0$, see (**??**)].

Any enumeration that might be obtained through Apostol's various recurrences can also be obtained through either (??) or (??) in a simpler and more general way—our recurrences above are valid for all $n \geq 1$, $m \geq 0$ and $k \geq 0$ when taking (??) into account. All this comes down to the necessity of considering the cases $n = -1$ and $n = 0$ in the initialisation —see related comments at the very end of Section ??. The importance of the initial values (??) and (??) cannot be understated: with them, all of Apostol's recurrences would become a special case of (??).

To conclude this remark, in the specific case of $m = 0$ recurrence (??) becomes

$$w_k(n,0) = 2\, w_k(n-1,0) - w_k\big(n-(k+1),0\big) + w_k\big(n-(k+2),0\big), \tag{25}$$

i.e., we recover the same recurrence given by Madden in [?, Case 4]—note that $w_k(n,0) = N(n+1,k+1)$ in this author's notation. See further comments about Madden's work in Remark ??. Of course, (??) is also one of Apostol's recurrences [?, Eq. (15)]. □

2.2.2 Generating Functions

From (??), the ogf $W_k(x,y) = \sum_{n,m} w_k(n,m)\, x^n y^m$ is

$$W_k(x,y) = \frac{1-x}{x\big(1-2x+(1-y)(x^{k+1}-x^{k+2})\big)}, \tag{26}$$

and from (??) we get

$$[y^m]W_k(x,y) = x^{m(k+1)-1}\left(\frac{1-x}{1-2x+x^{k+1}-x^{k+2}}\right)^{m+1}, \tag{27}$$

which is the ogf enumerating the binary strings that contain exactly m k-runs.

Remark. The ogf (??) was first given by Apostol [?, Eq. (25)], who determined it through his aforementioned recurrences. Koutras and Papastavridis [?, Sec. 5.c] also found (??) through the particularisation of a general result based on occupancy models (i.e., distributions of balls into cells). Finally, this ogf is also Madden's main theorem in [?].

Bearing in mind that —except for [?]— all prior work on this problem was unbeknownst to him, Madden observes regarding (??): "*This result is quite elementary, but we have not been able to find it in any other source*". Rediscoveries and reworkings like this are not uncommon in the long history of the theory of runs, and we will find more as the paper progresses. In any case, they show that fundamental problems concerning runs still merit a closer look.

□

Finally, from (??), the ogf $W_k^{\geq m}(x) = \sum_n w_k^{\geq m}(n)\, x^n$ is

$$W_k^{\geq m}(x) = \frac{1-x}{(1-2x)\,x}\left(\frac{x^{k+1}-x^{k+2}}{1-2x+x^{k+1}-x^{k+2}}\right)^{m}, \tag{28}$$

which enumerates the binary strings that contain at least m k-runs —see (**??**).

2.2.3 Explicit Expression

We find next an explicit expression for $w_k(n,m)$.

We proceed by applying successive power series expansions in order to then identify the coefficient of x^n in (**??**). Through elementary algebraic manipulations, and using e defined in (**??**), we can rewrite (**??**) as $[y^m]W_k(x,y) = x^{n-e}\left(1 - x\left((1-x)^{-1} - x^k\right)\right)^{-(m+1)}$. We now expand this expression using in succession $(1-x)^{-v} = \sum_{r\geq 0}\binom{r+v-1}{r}x^r$ (negative binomial theorem), the binomial theorem, and again the negative binomial theorem, to get

$$[y^m]W_k(x,y) = \sum_{r\geq 0}\binom{r+m}{r}\sum_{s\geq 0}\binom{r}{s}(-1)^{r-s}\sum_{t\geq 0}\binom{t+s-1}{t}x^{n-e+r(k+1)-sk+t}. \tag{29}$$

Next, we have to identify the coefficient of x^n in $[y^m]W_k(x,y)$, which, from (**??**), requires finding the ranges of indices r, s and t in that expression for which

$$r(k+1) - sk + t = e. \tag{30}$$

Equation (**??**) determines a single value of t for any given (r,s) pair. So, as long as we guarantee that $t \geq 0$, we only need to focus our attention on r and s. Because $s \leq r$ (as otherwise $\binom{r}{s} = 0$), we have from (**??**) that $t \leq e - r$.

So, the highest value of r for which t can have a nonnegative value is e. All this considered, the coefficient of x^n in (**??**), or, equivalently, $[x^n y^m]W_k(x,y)$, is

$$w_k(n,m) = \sum_{r=0}^{e}\binom{r+m}{r}\sum_{s=0}^{r}(-1)^{r-s}\binom{r}{s}\binom{e-1-r(k+1)+s(k+1)}{e-r(k+1)+sk}. \tag{31}$$

Expression (**??**) is similar to the one previously derived by Makri and Psillakis [**?**, Eq. (8)], which they obtained as a corollary to an earlier probabilistic result of theirs [**?**]. Like their expression, (**??**) is simpler than the expressions given by Magliveras and Wei [**?**, Thm. 2.3] —the earliest explicit result— or by Sinha and Sinha [**?**, (3)], because it involves two summations rather than three.

Remark. Concurring with Madden's comments in [**?**, p. 2], we also found that Sinha and Sinha's expression for $w_k(n,m)$ [**?**, Eq. (3)] (in their notation, $N_n^{m,k}$) is not always in agreement with the true counts. Their expression for $N_n^{m,k}$ is essentially correct, but it fails when $e = 0$ [in our notation, see (**??**)]. One can fix this minor oversight by starting the summations in [**?**, Eqs. (2) and (3)] at $r = 0$ rather than at $r = 1$.

□

2.2.4 OEIS

Madden [?] has pointed out that most of the number sequences that emanate from $w_k(n, m)$ are not in the OEIS. Below we give the few ones that we have found to be documented already.

- $m = 0$ (as indicated by Madden):

  $w_1(n-2, 0)$ is A005251 for $n \geq 2$.

  $w_2(n-3, 0)$ is A049856 for $n \geq 3$.

  $w_3(n-1, 0)$ is A108758 for $n \geq 1$.

- $m = 1$

  $w_1(n, 1)$ is A079662 (number of occurrences of 1 in all compositions of n without 2's).

- $w_1(n-1, m)$ is the mth column of A105114 for $n \geq 1$.

  $w_2(n-1, m)$ is the mth column of A218796 for $n \geq 1$.

- $k = 0$ (null runs)

  $w_0(n, 0)$ is A000045 (Fibonacci numbers).

  $w_0(n, 1)$ is A006367 (number of binary vectors of length $n+1$ beginning with 0 and containing just 1 singleton).

  $w_0(n+1, 2)$ is A105423 (number of compositions of $n+2$ having exactly two parts equal to 1).

- Only sequences from $w_k^{\geq m}(n)$ with $m = 1$ are in the OEIS; in this case, $w_k^{\geq 1}(n) = 2^n - w_k(n, 0)$:

  $w_1^{\geq 1}(n)$ is A384153 —see also Section **??**.

  $w_2^{\geq 1}(n)$ is A177795 (number of length n binary words that have at least one maximal run of 1's having length two).

2.3 Number of n-Strings that Contain Exactly m ($\geq k$)-Runs

Let us next address the enumeration of the n-strings that contain exactly m ($\geq k$)-runs (i.e., runs of length k or longer), which we denote by $w_{\geq k}(n, m)$. Notice that such n-strings may also contain other runs shorter than k.

Explicit solutions to this problem were previously given by Muselli [?] —in the context of probabilistic success runs— and by Makri and Psillakis [?]. In the special case $m = 0$, which is of particular relevance in Section **??**, a generating function and an explicit expression were given by Sedgewick and Flajolet [?], and a recurrence by Nyblom [?]. A probabilistic recurrence and a probability generating function were also given by Balakrishnan and Koutras [?] —see comments about these results in Remark **??** (Section **??**).

If $k > n$ then $w_{\geq k}(n, m) = 0$, and so we assume $k \leq n$. Therefore, the problem is a special case of the enumeration in Section **??** with $\underline{k} = k$ and $\overline{k} = n$, and thus

$$w_{\geq k}(n, m) = w_{k \leq n}(n, m). \tag{32}$$

We also study the number of n-strings that contain at least m ($\geq k$)-runs, which from (**??**) and (**??**), is

$$w_{\geq k}^{\geq m}(n) = \sum_{t=m}^{\left\lfloor \frac{n+1}{k+1} \right\rfloor} w_{\geq k}(n, t). \tag{33}$$

2.3.1 Recurrences

From (**??**), (**??**) and (**??**), two recurrence relations for $w_{\geq k}(n, m)$ are

$$w_{\geq k}(n, m) = \sum_{i=0}^{k-1} w_{\geq k}\big(n - (i+1), m\big) + \sum_{i=k}^{n} w_{\geq k}\big(n - (i+1), m-1\big), \tag{34}$$

and

$$w_{\geq k}(n, m) = 2\, w_{\geq k}(n-1, m) + w_{\geq k}\big(n - (k+1), m-1\big) - w_{\geq k}\big(n - (k+1), m\big). \tag{35}$$

which from (**??**) and (**??**) are both initialised by

$$w_{\geq k}(-1, m) = [\![m = 0]\!] \tag{36}$$

$$w_{\geq k}(0, m) = \big[\!\big[m = [\![k = 0]\!]\big]\!\big]. \tag{37}$$

2.3.2 Generating Functions

We next obtain the ogf $W_{\geq k}(x, y) = \sum_{n,m} w_{\geq k}(n, m)\, x^n y^m$. Although a valid ogf is directly obtained by setting $\underline{k} = k$ and $\overline{k} = n$ in (**??**), the resulting expression is n-dependent. We can derive an alternative n-independent ogf, valid for all $k \geq 0$, from (**??**) using the same procedure as in Section **??**. This yields

$$W_{\geq k}(x, y) = \frac{1 - x}{x\big(1 - 2x + (1 - y)x^{k+1}\big)}. \tag{38}$$

Through the same method used to get (??) from (??), we now obtain from (??)

$$[y^m]W_{\geq k}(x,y) = x^{m(k+1)-1}\frac{1-x}{\left(1-2x+x^{k+1}\right)^{m+1}}, \tag{39}$$

which is the ogf enumerating the binary strings that contain exactly m ($\geq k$)-runs.

Remark. We make here some comments for the case $m = 0$. In this case, recurrence (??) becomes

$$w_{\geq k}(n,0) = \sum_{i=0}^{k-1} w_{\geq k}\big(n-(i+1),0\big), \tag{40}$$

which was given by Nyblom [?, Thm. 2.1] (with validity $k \geq 2$, and with different initialisation). This author was unaware of the fact that the same recurrence had been previously given by Schilling —see (??) and (??) in Section ??. It is also interesting to compare the ogf (??) in this special case, i.e., $[y^0]W_{\geq k}(x,y) = x^{-1}(1-x)/(1-2x+x^{k+1})$, with the equivalent ogf given by Sedgewick and Flajolet [?, p. 368] for enumerating the binary strings devoid of runs of k consecutive zeros: $S_k(x) = (1-x^k)/(1-2x+x^{k+1})$. The comparison is possible because, when $m = 0$, the definition of a run of k consecutive equal bits by Sedgewick and Flajolet is equivalent to our definition of a ($\geq k$)-run —see Remark ??. Even though these two generating functions are different, they have the same coefficients in their power series expansion for $n \geq 1$. □

Finally, we may obtain $W^{\geq m}_{\geq k}(x) = \sum_n w^{\geq m}_{\geq k}(n)\, x^n$ directly from (??), but this leads to an n-dependent expression. We can derive a more general ogf, valid for all $k \geq 0$, by working from (??) specialised to this case. Doing so yields

$$\begin{aligned} W^{\geq m}_{\geq k}(x) &= \frac{x^{k+1}}{1-2x}\,[y^{m-1}]W_{\geq k}(x,y) \\ &= \frac{1-x}{(1-2x)\,x}\left(\frac{x^{k+1}}{1-2x+x^{k+1}}\right)^m, \end{aligned} \tag{41}$$

where we have used (??) in the second step above. This is the ogf enumerating the binary strings that contain at least m ($\geq k$)-runs —see (??).

2.3.3 Explicit Expression

To obtain an explicit expression for $w_{\geq k}(n,m)$ we

first apply in succession the negative binomial theorem and the binomial theorem to the denominator of (??). With this expansion we can express (??) as

$$[y^m]W_{\geq k}(x,y) = (1-x)\sum_{r\geq 0}\binom{m+r}{r}\sum_{s\geq 0}\binom{r}{s}(-1)^{(r-s)}2^s x^{n-e+(k+1)r-ks}, \tag{42}$$

where we have used (**??**) in the exponent of x. In order to determine the coefficient of x^n in the above expression we have to find the values of indices r and s that solve each of the following two Diophantine equations:

$$(k+1)r - ks = e, \tag{43}$$

$$(k+1)r - ks = e - 1. \tag{44}$$

Since $\gcd(k+1, -k) = 1$, we can solve both equations.

By inspection, a particular solution to (**??**) is $r = s = e$. Thus the general solution is of the form $r = e - k\,t$ and $s = e - (k+1)\,t$ for integer t, which fulfils $s \le r$ for nonnegative t. As we also need to guarantee $r \ge 0$ and $s \ge 0$, the valid range of t is $0 \le t \le \lfloor e/(k+1) \rfloor$. Similarly, a particular solution to (**??**) is $r = s = e - 1$, and thus the general solution is of the form $r = e - 1 - k\,t$ and $s = e - 1 - (k+1)\,t$ for integer t, and the range of t is $0 \le t \le \lfloor (e-1)/(k+1) \rfloor$.

Combining these solutions we can see that the coefficient of x^n in (**??**), or, equivalently, $[x^n y^m]W_{\ge k}(x,y)$, is

$$w_{\ge k}(n,m) = \sum_{t=0}^{\lfloor \frac{e}{k+1} \rfloor} (-1)^t \, 2^{e-(k+1)t} \left(\binom{m+e-kt}{e-kt}\binom{e-kt}{e-(k+1)t} - \frac{1}{2}\binom{m+e-1-kt}{e-1-kt}\binom{e-1-kt}{e-1-(k+1)t} \right). \tag{45}$$

Observe that we use just one summation on t, rather than the difference of two summations on t with different ranges, by taking advantage of the fact that the last binomial coefficient becomes zero for $t > \lfloor (e-1)/(k+1) \rfloor$. This single-summation formula is clearly different from the previous double-summation formulas given by Muselli [**?**, Thm. 1 with $p = q = \frac{1}{2}$ gives $w_{\ge k}(n,m)/2^n$] and Makri and Psillakis [**?**, Eq. (23)] —which closely resemble each other, and which were derived through combinatorial analysis. However Muselli also simplified his double-summation expression into a single-summation formula comparable to (**??**) [**?**, Thm. 3].

The case $m = 0$ is especially relevant, as we see in more detail in Section **??** —see also Remark **??** below. With this argument choice, expression (**??**) becomes

$$w_{\ge k}(n,0) = \sum_{t=0}^{\lfloor \frac{n+1}{k+1} \rfloor} (-1)^t \, 2^{n+1-(k+1)t} \left(\binom{n+1-kt}{n+1-(k+1)t} - \frac{1}{2}\binom{n-kt}{n-(k+1)t} \right), \tag{46}$$

which is very similar to the explicit expression given by Sedgewick and Flajolet [**?**, p. 370] —in their notation, $w_{\ge k}(n,0) = [x^n]S_k(x)$.

Remark. As pointed out by Sedgewick and Flajolet [?, p. 369],

$$w_{\geq k}(n, 0) = F^{(k)}_{n+k}, \tag{47}$$

where $F^{(j)}_i$ are the jth order Fibonacci numbers, also called Fibonacci j-step numbers, j-nacci numbers, which are defined by the recurrence

$$F^{(j)}_i = \sum_{r=1}^{j} F^{(j)}_{i-r}. \tag{48}$$

for $j \geq 1$ and $i \geq j$, initialised with $F^{(j)}_{j-1} = 1$ and $F^{(j)}_i = 0$ for $0 \leq i < j - 1$. The case $j = 2$ gives the standard Fibonacci numbers, but observe that we also allow the degenerate case $j = 1$, in which case $F^{(1)}_i = 1$ for all $i \geq 1$. The connection described by (**??**) is clear when observing recurrence (**??**) —see also [?, p. 368]— or the ogf in Remark **??**.

Considering (**??**), a consequence of (**??**) is

$$w^{\geq 1}_{\geq k}(n) = 2^n - F^{(k)}_{n+k}. \tag{49}$$

□

Remark. The special case $w_{\geq k}(n, 0)$ with $k \geq 1$ can also be used to enumerate the n-strings that feature their first consecutive appearance of k ones at index i, which we denote by $\widehat{w}_k(n, i)$. Notice that we are not only referring to the first appearance of a k-run: for the avoidance of doubt, index i marks the position of the kth one in the first uninterrupted sequence of k ones, which may be followed by a zero, a one, or the end of the string. To obtain $\widehat{w}_k(n, i)$ we just have to make the following observation: if the first k consecutive ones occur at index i, then the initial $(i - (k + 1))$-substring must be devoid of $(\geq k)$-runs. For every such start of the n-string we have 2^{n-i} possible endings. Combining these two facts, we thus have that

$$\widehat{w}_k(n, i) = w_{\geq k}(i - (k + 1), 0)\ 2^{n-i}\, [\![i \leq n]\!]. \tag{50}$$

Note that this formula works even if $i < k$, in which case $w_{\geq k}(i - (k + 1), m) = 0$ through necessary condition (**??**). The enumeration $\widehat{w}_k(n, i)$ is relevant in waiting-time type problems. Bearing in mind (**??**), the reader is referred to [?, Sec. 2.8] for a literature review on the role of jth order Fibonacci numbers in the context of waiting-time problems with runs. □

2.3.4 OEIS

We report below the number sequences stemming from $w_{\geq k}(n, m)$ that are listed the OEIS.

- $m = 0$ (see Remark **??**)

  $w_{\geq 2}(n - 2, 0)$ is A000045 (Fibonacci numbers) for $n \geq 2$.

$w_{\geq 3}(n-3,0)$ is A000073 (tribonacci numbers) for $n \geq 3$.

$w_{\geq 4}(n-4,0)$ is A000078 (tetranacci numbers) for $n \geq 4$.

$w_{\geq 5}(n-5,0)$ is A001591 (pentanacci numbers) for $n \geq 5$.

etc

- $m = 1$

  $w_{\geq 2}(n-1,1)$ is A006478 for $n \geq 1$.

- $w_{\geq 1}(n-1,m)$ is the mth column of A034839 for $n \geq 1$.

  $w_{\geq 2}(n,m)$ is the mth column of A334658.

- $k = 1$: apart from A034839, many sequences with this parameter are individually documented in the OEIS for different values of m —see (**??**) and (**??**) in Section **??**.

  $w_{\geq 1}(n,1)$ is A000217 (triangular numbers, or $\binom{n+1}{2}$).

  $w_{\geq 1}(n-1,2)$ is A000332 (binomial coefficient $\binom{n}{4}$) for $n \geq 1$.

  $w_{\geq 1}(n-1,3)$ is A000579 (binomial coefficient $\binom{n}{6}$) for $n \geq 1$.

  $w_{\geq 1}(n-1,4)$ is A000581 (binomial coefficient $\binom{n}{8}$) for $n \geq 1$.

  $w_{\geq 1}(n-1,5)$ is A001287 (binomial coefficient $\binom{n}{10}$) for $n \geq 1$.

  For $6 \leq m \leq 24$, $w_{\geq 1}(n-1,m)$ is A010965+2$(m-6)$ (binomial coefficient $\binom{n}{2m}$) for $n \geq 1$.

- $k = 0$ (number of n-strings that contain exactly m runs, including null runs): we can see from the ogf (**??**) that $w_{\geq 0}(n,m) = \binom{n}{m-1}$, and therefore this sequence features in many OEIS entries. See also Section **??**.

- Only $w_{\geq k}^{\geq m}(n)$ sequences having $m = 1$ are in the OEIS—see (**??**). Of course, from (**??**) we have $w_{\geq k}^{\geq 1}(n) = 2^n - w_{\geq k}(n,0)$.

  $w_{\geq 1}^{\geq 1}(n) = 2^n - 1$ is A000225 (sometimes called Mersenne numbers).

  $w_{\geq 2}^{\geq 1}(n)$ is A008466.

  $w_{\geq 3}^{\geq 1}(n)$ is A050231.

  $w_{\geq 4}^{\geq 1}(n)$ is A050232.

  $w_{\geq 5}^{\geq 1}(n)$ is A050233.

  $w_{\geq 6}^{\geq 1}(n)$ is A143662.

  $w_{\geq 7}^{\geq 1}(n)$ is A151975.

  $w_{\geq k}^{\geq 1}(n)$ is the kth column of A050227.

Weisstein discusses $w^{\geq 1}_{\geq k}(n)$ in [**?**] in connection with Feller's work on probabilistic success runs [**?**]. Nevertheless, we should note that $w^{\geq 1}_{\geq k}(n)$ was actually first studied by de Moivre [**?**], and the probability generating function attributed in [**?**] to Feller was first found by Laplace —see more details in Remarks **??** and **??** towards the end of Section **??**.

Finally, because of (**??**), each of the OEIS sequences above can be paired with a corresponding kth order Fibonacci sequence (mentioned at the start of this section).

2.4 Number of n-Strings that Contain Exactly m Nonnull ($\leq k$)-Runs

Let us next address what is essentially the counterpart of the problem in the previous section: the enumeration of the n-strings containing exactly m nonnull ($\leq k$)-runs (i.e., runs of length k or shorter, but strictly greater than zero), which we denote by $w_{\leq k}(n,m)$. Hence, in this section we assume $k \geq 1$. Notice that the n-strings that we enumerate may also contain any number of null runs, and/or of runs with lengths longer than k.

Austin and Guy [**?**] gave a recurrence and a semi-explicit expression for "*the number of binary sequences of length n in which the ones occur only in blocks of length at least k*" (where $k \geq 2$), that is to say, for the special case $w_{\leq(k-1)}(n,0)$ of the enumeration that we consider in this section. Other than in [**?**], $w_{\leq k}(n,m)$ seems not to have received any attention in the literature —which is somewhat surprising, given that this enumeration looks like a mere variation of $w_{\geq k}(n,m)$. A plausible reason for this is that it may be harder to address this problem through direct combinatorial analysis. This hypothesis is suggested by the explicit expression (**??**) for $w_{\leq k}(n,m)$ that we derive later —a triple-summation expression, as opposed to the single sum expression (**??**) for $w_{\geq k}(n,m)$. Another possible explanation might be the perception that the problem in this section has fewer applications. Whatever the cause, at least one good reason for studying $w_{\leq k}(n,m)$ is that the special case $w_{\leq k}(n,0)$ plays a key role in Section **??**, just like we have mentioned that $w_{\geq k}(n,0)$ plays a key role in Section **??**.

Again, the problem at hand is a special case of the enumeration in Section **??**, here using $\underline{k}=1$ and $\overline{k}=k$, and thus

$$w_{\leq k}(n,m) = w_{1 \leq k}(n,m). \tag{51}$$

We also study the number of n-strings that contain at least m ($\leq k$)-runs, which, from (**??**) and (**??**), is

$$w^{\geq m}_{\leq k}(n) = \sum_{t=m}^{\left\lfloor \frac{n+1}{2} \right\rfloor} w_{\leq k}(n,t). \tag{52}$$

2.4.1 Recurrences

From (**??**), (**??**) and (**??**), two recurrence relations for $w_{\geq k}(n,m)$ are

$$w_{\leq k}(n,m) = w_{\leq k}(n-1,m) + \sum_{i=k+1}^{n} w_{\leq k}\big(n-(i+1),m\big) + \sum_{i=1}^{k} w_{\leq k}\big(n-(i+1),m-1\big), \tag{53}$$

and

$$\begin{aligned} w_{\leq k}(n,m) = 2\,w_{\leq k}(n-1,m) + w_{\leq k}(n-2,m-1) - w_{\leq k}(n-2,m) \\ - w_{\leq k}\big(n-(k+2),m-1\big) + w_{\leq k}\big(n-(k+2),m\big), \end{aligned} \tag{54}$$

both of which, from (**??**), (**??**) and (**??**), are initialised by

$$w_{\leq k}(-1,m) = w_{\leq k}(0,m) = [\![m=0]\!]. \tag{55}$$

Remark. In the special case $m=0$, recurrence (**??**) becomes

$$w_{\leq k}(n,0) = 2\,w_{\leq k}(n-1,0) - w_{\leq k}(n-2,0) + w_{\leq k}\big(n-(k+2),0\big), \tag{56}$$

which was given by Austin and Guy for $a_n^{(k)} = w_{\leq(k-1)}(n,0)$ [**?**, Eq. (1')]. Importantly, these authors used the very same initialisation strategy as above, i.e., (**??**).

□

2.4.2 Generating Functions

We next obtain $W_{\leq k}(x,y) = \sum_{n,m} w_{\leq k}(n,m)\,x^n y^m$. Setting $\underline{k}=1$ and $\overline{k}=k$ in (**??**) we have

$$W_{\leq k}(x,y) = \frac{1-x}{x\big(1-2x+(1-y)(x^2-x^{k+2})\big)}, \tag{57}$$

whereas (**??**) yields in turn

$$[y^m]W_{\leq k}(x,y) = \frac{x^{2m-1}(1-x)}{1-x^k}\left(\frac{1-x^k}{1-2x+x^2-x^{k+2}}\right)^{m+1}. \tag{58}$$

This is the ogf enumerating the binary strings that contain exactly m nonnull $(\leq k)$-runs.
Finally, from (**??**), $W_{\leq k}^{\geq m}(x) = \sum_n w_{\leq k}^{\geq m}(n)\,x^n$ is

$$W_{\leq k}^{\geq m}(x) = \frac{1-x}{(1-2x)\,x}\left(\frac{x^2-x^{k+2}}{1-2x+x^2-x^{k+2}}\right)^{m}, \tag{59}$$

which is the ogf enumerating the binary strings that contain at least m nonnull $(\leq k)$-runs —see (**??**).

2.4.3 Explicit Expression

We now derive an explicit expression for $w_{\leq k}(n,m)$ by finding $[x^n y^n]W_{\leq k}(x,y)$, or, equivalently, the coefficient of x^n in the power series expansion of (**??**). To do so, we rewrite this expression as $[y^m]W_{\leq k}(x,y) = x^{2m-1}(1-x^k)^m(1-x)^{-(2m+1)}\big(1-x^{k+2}/(1-x)^2\big)^{-(m+1)}$ and we develop each of the last three factors in it into power series. Using the negative binomial series in the first of these factors, the binomial theorem in the second one, and again the negative binomial theorem (twice) in the third one, we have that

$$\begin{aligned}
\left(1-x^k\right)^m &= \sum_{r\geq 0}\binom{m}{r}(-1)^r x^{k\,r},\\
(1-x)^{-(2m+1)} &= \sum_{s\geq 0}\binom{s+2m}{s}x^s,\\
\left(1-\frac{x^{k+2}}{(1-x)^2}\right)^{-(m+1)} &= \sum_{t\geq 0}\binom{t+m}{t}x^{(k+2)\,t}\sum_{p\geq 0}\binom{p+2t-1}{p}x^p.
\end{aligned}$$

Collecting the exponents of x, we see that we have to determine next the ranges of the summation indices r,s,t and p that fulfil $2m-1+k\,r+s+(k+2)\,t+p=n$, or, using (**??**), $k\,r+s+(k+2)\,t+p=e$. Let us define at this point the auxiliary variable

$$d = e - k\,r.$$

As all indices are nonnegative, $s+(k+2)\,t+p\geq 0$. So we must guarantee $d\geq 0$ in order to have solutions to our problem, which implies $r\leq\lfloor e/k\rfloor$. We next have to solve $d=s+(k+2)\,t+p$ for integers s,t and p, where $s,t,p\geq 0$. The maximum of t happens when $s=p=0$, and thus $t\leq\lfloor d/(k+2)\rfloor$. Given t, the maximum of s happens when $p=0$, and thus $s\leq d-(k+2)\,t$. Finally, given t and s, we have that $p=d-(k+2)\,t-s$.

Collecting all these solutions, we can see that the coefficient of x^n in (**??**) is

$$w_{\leq k}(n,m) = \sum_{r=0}^{\lfloor e/k\rfloor}(-1)^r\binom{m}{r}\sum_{t=0}^{\lfloor d/(k+2)\rfloor}\binom{t+m}{t}\sum_{s=0}^{d-(k+2)t}\binom{s+2m}{s}\binom{d-kt-s-1}{d-(k+2)t-s}. \tag{60}$$

In the case $m=0$, which is especially relevant in Section **??**, the triple-summation expression above simplifies considerably:

$$w_{\leq k}(n,0) = 1+\sum_{t=1}^{\lfloor\frac{n+1}{k}\rfloor}\frac{n+1-kt}{2t}\binom{n-kt}{n+1-(k+2)t}. \tag{61}$$

This expression also constitutes an alternative to the semi-explicit expression $w_{\leq(k-1)}(n,0) = \mathrm{round}(c_k\gamma_k^n)$ given by Austin and Guy [**?**, p. 85], which is based on asymptotic considerations and requires determining a special constant c_k and finding the real root γ_k of a kth order polynomial.

2.4.4 OEIS

As in previous sections, we report the few sequences emanating from $w_{\leq k}(n,m)$ that we have been able to find in the OEIS.

- $m = 0$

  $w_{\leq 1}(n-2,0) = w_1(n-2,0)$ is A005251 for $n \geq 2$.

  $w_{\leq 2}(n-1,0)$ is A005252 for $n \geq 1$.

  $w_{\leq 3}(n,0)$ is A005253 (number of binary words of length n in which the ones occur only in blocks of length at least 4).

  $w_{\leq 4}(n-2,0)$ is A005689 (number of Twopins positions) for $n \geq 6$.

  $w_{\leq 5}(n-1,0)$ is A098574 for $n \geq 1$.

  $w_{\leq 6}(n+6,0)$ is A217838 (number of n element 0..1 arrays with each element the minimum of 7 adjacent elements of a random 0..1 array of $n+6$ elements).

  These OEIS sequences actually led us to finding reference [?].

- $m = 1$

  $w_{\leq 1}(n,1) = w_1(n,1)$ is A079662 (number of occurrences of 1 in all compositions of n without 2's).

- Sequences from $w_{\leq k}^{\geq m}(n)$ —see (??):

  $w_{\leq 1}^{\geq 1}(n) = w_1^{\geq 1}(n)$ is A384153 —see also Section ??.

2.5 Number of n-Strings that Contain Exactly m Nonnull Runs

In this section we study the number of n-strings that contain exactly m nonnull runs (of arbitrary lengths, all strictly greater than zero), which we denote by $w(n,m)$. As may be expected, this problem is also a special case of the enumeration in Section ?? with $\underline{k} = 1$ and $\overline{k} = n$, and thus

$$w(n,m) = w_{1 \leq n}(n,m). \tag{62}$$

Alternatively, using the enumerations in Sections ?? and ?? we may also write

$$w(n,m) = w_{\geq 1}(n,m) = w_{\leq n}(n,m). \tag{63}$$

We also study the number of n-strings that contain at least m runs, which from (??) and (??) is

$$w^{\geq m}(n) = \sum_{t=m}^{\left\lfloor \frac{n+1}{2} \right\rfloor} w(n,t). \tag{64}$$

2.5.1 Recurrences

From (**??**), (**??**) and (**??**) two recurrences for $w(n,m)$ are:

$$w(n,m) = w(n-1,m) + \sum_{i=1}^{n} w(n-(i+1), m-1),$$

and

$$w(n,m) = 2\,w(n-1,m) + w(n-2,m-1) - w(n-2,m),$$

both of which, from (**??**), (**??**) and (**??**), have initial values

$$w(-1,m) = w(0,m) = [\![m=0]\!].$$

2.5.2 Generating Functions

From (**??**) and (**??**), we may get the ogf $W(x,y) = \sum_{n,m} w(n,m)\,x^n y^m$ in two different ways: by setting $k=n$ in (**??**) —which is the same as setting $\underline{k}=1$ and $\overline{k}=n$ in (**??**)— or else by setting $k=1$ in (**??**). The first option gives an n-dependent ogf, but the second option gives a simpler more general version:

$$W(x,y) = \frac{1-x}{x\bigl(1-2x+(1-y)x^2\bigr)}. \tag{65}$$

On the other hand, setting $k=1$ in (**??**), yields

$$[y^m]W(x,y) = \frac{x^{2m-1}}{(1-x)^{2m+1}}, \tag{66}$$

which is the ogf enumerating the binary strings that contain exactly m nonnull runs.

Finally, setting $k=1$ in (**??**) we have that $W^{\geq m}(x) = \sum_n w^{\geq m}(n)\,x^n$ is

$$W^{\geq m}(x) = \frac{1}{1-2x}\left(\frac{x}{1-x}\right)^{2m-1}, \tag{67}$$

which is the ogf enumerating the binary strings that contain at least m nonnull runs — see (**??**).

2.5.3 Explicit Expressions

Using (**??**), we can get explicit expressions for $w(n,m)$ by setting $k=1$ in (**??**) or by setting $k=n$ in (**??**), but the resulting formulas are not obviously simplifiable —especially the second one. We may obtain a much simpler expression by relying on (**??**), which, by applying the negative binomial theorem, can be expanded as

$$[y^m]W(x,y) = \sum_{t\geq 0} \binom{t+2m}{2m} x^{2m+t-1}. \tag{68}$$

To determine $w(n,m) = [x^n y^m]W(x,y)$ we have to find the value of t for which $n = 2m+t-1$, from which we get

$$w(n,m) = \binom{n+1}{2m}. \tag{69}$$

Remark. Goulden and Jackson [**?**, Sec. 2.4.4] found (**??**) using the following alternative ogf —cf. (**??**):

$$W(x,y) = \frac{1+(y-1)x}{1-2x+(1-y)x^2}.$$

At any rate, expression (**??**) suggests that $w(n,m)$ can also be found through basic combinatorial reasoning. One such combinatorial explanation is as follows: assume that the number of ones in an n-string that contains exactly m nonnull runs is $t = m, \ldots, n-(m-1)$ —as the minimum is m and there must be at least $m-1$ zeros. For a given t, there are $\binom{t-1}{t-m}$ ways to choose the lengths of the m runs, and $\binom{n-t+1}{n-t-m+1}$ ways to place the $n-t$ zeros around them. Using $\binom{a}{b} = \binom{a}{a-b}$, we can thus write

$$w(n,m) = \sum_{t=m}^{n-m+1} \binom{t-1}{m-1}\binom{n-t+1}{m}. \tag{70}$$

We can also put this summation as $w(n,m) = \sum_{t=1}^{n+1} \binom{t-1}{m-1}\binom{n-t+1}{m}$, because the first binomial coefficient in (**??**) is zero for $t < m$ and the second one is zero for $t > n-m+1$. Hence, $w(n,m) = \sum_{t=0}^{n} \binom{t}{m-1}\binom{n-t}{m}$, which is a variation of Vandermonde's convolution that adds up to (**??**) —see [**?**, Eq. (5.26)].

□

2.5.4 OEIS

We report here the OEIS sequences relevant to the enumerations in this section.

- Many $w(n,m)$ sequences are in the OEIS for different values of m —see sequences with $k=1$ in Section **??**.
- As for $w^{\geq m}(n)$ —see (**??**)—, it yields the following OEIS sequences:

  $w^{\geq 1}(n) = 2^n - 1$ is A000225.

$w^{\geq 2}(n)$ is A002662.

$w^{\geq 3}(n)$ is A002664.

$w^{\geq 4}(n)$ is A035039.

$w^{\geq 5}(n)$ is A035041.

These connections are due to the fact that, using $\binom{n+1}{2t} = \binom{n}{2t} + \binom{n}{2t-1}$, from (**??**) and (**??**) we can write

$$w^{\geq m}(n) = \sum_{t=m}^{\lfloor (n+1)/2 \rfloor} \left(\binom{n}{2t} + \binom{n}{2t-1} \right) = \sum_{\substack{t=2m \\ t \text{ even}}}^{n} \binom{n}{t} + \sum_{\substack{t=2m-1 \\ t \text{ odd}}}^{n} \binom{n}{t} = \sum_{t=2m-1}^{n} \binom{n}{t}$$
$$= 2^n - \sum_{t=0}^{2(m-1)} \binom{n}{t},$$

where in the second equality we have taken advantage of the fact that $\binom{n}{t} = 0$ for $t > n$ to set n as the upper limit in both summations. The last expression is the common definition to all five OEIS sequences above.

2.6 Number of n-Strings Whose Longest Run Is a k-Run or a $(\leq k)$-Run

We denote the number of n-strings whose longest run is a k-run by $\overline{w}_k(n)$, whereas $\overline{w}_{\leq k}(n)$ represents the number of n-strings whose longest run is a $(\leq k)$-run. A recurrence for $\overline{w}_{\leq k}(n)$ was previously given by Schilling [**?**], and, in the context of probabilistic success runs, explicit expressions were given by Godbole [**?**] and Muselli [**?**]. Finally, Flajolet and Sedgewick [**?**] gave a generating function for $\overline{w}_{\leq k}(n)$ constructed through the symbolic method, and Prodinger gave its asymptotic behaviour [**?**]. As for previous work on $\overline{w}_k(n)$, probabilistic versions of this enumeration were given by Muselli [**?**] and by Makri et al. [**?**], and Schilling provided asymptotic estimates of its distribution [**?**].

For all $k \geq 0$, the two enumerations that we are interested in are related as follows

$$\overline{w}_{\leq k}(n) = \sum_{j=0}^{k} \overline{w}_j(n). \tag{71}$$

Conversely, for $k \geq 1$

$$\overline{w}_k(n) = \overline{w}_{\leq k}(n) - \overline{w}_{\leq (k-1)}(n), \tag{72}$$

whereas $\overline{w}_0(n) = \overline{w}_{\leq 0}(n)$.

Both enumerations follow directly from the results in Section **??** because of the following fact:

$$\overline{w}_{\leq k}(n) = w_{\geq (k+1)}(n, 0). \tag{73}$$

Thus, $\overline{w}_0(n) = w_{\geq 1}(n, 0) = 1$ for $n \geq 1$, as the longest run is a null run only in the all-zeros n-string.

Of course, considering (**??**) and (**??**), $\overline{w}_{\leq k}(n)$ can also be expressed as a special case of $w_{\underline{k} \leq \bar{k}}(n, m)$ in Section **??**, but using (**??**) we are able to simplify our presentation and directly obtain explicit expressions.

2.6.1 Recurrences

We start by finding recurrences for $\overline{w}_{\leq k}(n)$. Schilling's recurrence [**?**, Eq. (1)] is recovered by directly applying (**??**) to (**??**):

$$\overline{w}_{\leq k}(n) = \sum_{i=0}^{k} \overline{w}_{\leq k}(n - (i + 1)). \tag{74}$$

On the other hand, an alternative recurrence is obtained by directly applying (**??**) to (**??**), which gives

$$\overline{w}_{\leq k}(n) = 2\, \overline{w}_{\leq k}(n - 1) - \overline{w}_{\leq k}(n - (k + 2)). \tag{75}$$

From (**??**), (**??**) and (**??**), we can see that the initialisation of both (**??**) and (**??**) is

$$\overline{w}_{\leq k}(-1) = \overline{w}_{\leq k}(0) = 1. \tag{76}$$

Remark. Schilling initialised (**??**) in [**?**, Eq. (1)] using

$$\overline{w}_{\leq k}(n) = 2^n \quad \text{if } n \leq k, \tag{77}$$

but there is no need to resort to (**??**) when initialising (**??**) with (**??**). Also, Bloom gave recurrence (**??**) for $k = 4$ [**?**, Eq. (9)]. □

We may also obtain recurrences for $\overline{w}_k(n)$ for $k \geq 1$ by relying on (**??**) and (**??**). Using (**??**) in (**??**), after some elementary algebraic manipulations we get

$$\overline{w}_k(n) = \sum_{i=0}^{k-1} \overline{w}_k(n - (i + 1)) + \sum_{j=0}^{k} \overline{w}_j(n - (k + 1)), \tag{78}$$

whereas if we instead use (**??**) in (**??**) we get the following alternative recurrence:

$$\overline{w}_k(n) = 2\, \overline{w}_k(n - 1) + \sum_{j=0}^{k-1} \overline{w}_j(n - (k + 1)) - \sum_{j=0}^{k} \overline{w}_j(n - (k + 2)). \tag{79}$$

From (**??**) and (**??**), it follows that $\overline{w}_k(-1) = \overline{w}_k(0) = 0$ for $k \geq 1$. As seen from (**??**) and (**??**), we also need initialisation for $k = 0$. As $\overline{w}_0(n) = w_{\geq 1}(n, 0)$, we have from (**??**) and (**??**) that $\overline{w}_0(-1) = \overline{w}_0(0) = 1$. Thus, the initialisation of recurrences (**??**) and (**??**) is

$$\overline{w}_k(-1) = \overline{w}_k(0) = [\![k = 0]\!].$$

The ease with which these recurrences are initialised shows that null runs are inherent to this problem —although the stronger reason for taking null runs into account will be seen in Section ??.

2.6.2 Generating Functions

We may obtain generating functions for $\overline{w}_{\le k}(n)$ and $\overline{w}_k(n)$ by using the recurrences in the previous section, although the reader may verify that this is easier for $\overline{w}_{\le k}(n)$ than for $\overline{w}_k(n)$. Let us follow instead the path of least resistance by exploiting (??) and (??). The ogf $\overline{W}_{\le k}(x) = \sum_n \overline{w}_{\le k}(x)\, x^n$ follows directly from (??) and (??):

$$\overline{W}_{\le k}(x) = \frac{1-x}{x\big(1-2x+x^{k+2}\big)}. \tag{80}$$

Likewise, for $k \ge 1$ the ogf $\overline{W}_k(x) = \sum_n \overline{w}_k(x)\, x^n$ is directly obtained using (??) in (??), which yields

$$\begin{aligned}\overline{W}_k(x) &= \overline{W}_{\le k}(x) - \overline{W}_{\le (k-1)}(x)\\ &= \frac{x^k(1-x)^2}{(1-2x+x^{k+1})(1-2x+x^{k+2})}.\end{aligned} \tag{81}$$

Remark. If we use (??) and (??) to recover $\overline{W}_{\le k}(x)$ from $\overline{W}_k(x)$, instead of (??) we get

$$\overline{W}_{\le k}(x) = \frac{1-x^{k+1}}{1-2x+x^{k+2}}, \tag{82}$$

which was given by Flajolet and Sedgewick in [?, p. 51]. Although (??) is different from (??), it has the same coefficients in its power series expansion for $n \ge 1$. Prodinger gives an asymptotic analysis of $\overline{w}_{\le k}(n)$ using this ogf [?]. □

2.6.3 Explicit Expressions

Explicit expressions for $\overline{w}_k(n)$ and $\overline{w}_{\le k}(n)$ are available through (??), (??) and (??). The resulting formula for $\overline{w}_{\le k}(n)$ is comparable to the simplified single-summation combinatorial expression given by Muselli [?, Cor. 1 using $p = q = 1/2$ gives $\overline{w}_{\le (k-1)}(n)/2^n$]. In contrast, our single-summation expression for $\overline{w}_k(n)$ is much simpler than the triple-summation expression given by Makri et al. [?, Thm. 2.1.4 using $p = q = 1/2$ gives $\overline{w}_k(n)/2^n$].

Let us comment on some further aspects of our results in this section. Through (??) and (??) we have that

$$\overline{w}_{\le k}(n) = F^{(k+1)}_{n+k+1}, \tag{83}$$

where $F^{(j)}_i$ are the jth order Fibonacci numbers —see (??). This is also clear from recurrence (??). Regarding (??), Schilling observed already the special case $\overline{w}_{\le 1}(n) = F^{(2)}_{n+2}$, when he mentioned that "(...) *the number* (...) *of sequences of length* n *that contain no two consecutive heads is the* $(n+2)$*nd Fibonacci number*" [**?**]. See as well Nyblom's comments regarding this case [**?**, Cor. 2.1].

Additionally, because from (**??**) it follows that $\overline{w}_1(n) = \overline{w}_{\leq 1}(n) - 1$, then we also have that

$$\overline{w}_1(n) = F^{(2)}_{n+2} - 1. \tag{84}$$

Remark. For n-strings drawn uniformly at random, the expected length of the longest run is $\bar{\ell}_n = \left(\sum_{k=0}^{n} k\, \overline{w}_k(n)\right)/2^n$. Thus, by writing $\overline{w}_k(n) = [x^n]\overline{W}_k(x)$, we have from (**??**) that

$$\bar{\ell}_n = \frac{1}{2^n}[x^n]\sum_{k=1}^{n} k\, \frac{x^k(1-x)^2}{(1-2x+x^{k+1})(1-2x+x^{k+2})}. \tag{85}$$

This expression is equivalent to the one given by Sedgewick and Flajolet for the same quantity [**?**, p. 426], which in fact contains a small oversight: their summation must start at $k=1$ rather than at $k=0$. Of course, we can also get $\bar{\ell}_n$ as a double-summation explicit expression through (**??**), (**??**) and (**??**).

Using (**??**) and a computer algebra system, we may also check the elegant asymptotic estimate $\hat{\bar{\ell}}_n = \gamma \ln 2 + \log_2(n/2)$ given by Schilling in [**?**] against the exact value $\bar{\ell}_n$ for some moderate values of n. For instance, $\hat{\bar{\ell}}_{1000} = 9.3658$ and $\hat{\bar{\ell}}_{3000} = 10.9508$, whereas $\bar{\ell}_{1000} = 9.3000$ and $\bar{\ell}_{3000} = 10.8839$.

□

2.6.4 OEIS

We report here the enumerations in this section that yield OEIS sequences.

- Sequences emanating from $\overline{w}_k(n)$:

  $\overline{w}_1(n-2)$ is A000071 (Fibonacci numbers minus one) for $n \geq 2$ —cf. (**??**).

  $\overline{w}_j(n-1)$ for $j = 2, 3$ and 4 are A000100, A000102 and A006979, respectively (number of compositions of n in which the maximal part is $j+1$) for $n \geq 1$.

  $\overline{w}_5(n-1)$ is A006980 for $n \geq 6$.

  $\overline{w}_k(n)$ is the kth column of A048004.

  Additionally, $2^n \bar{\ell}_n$ is A119706 —see (**??**).

- Sequences emanating from $\overline{w}_{\leq k}(n)$: see $m = 0$ in Section **??** —recall (**??**).

2.7 Number of n-Strings Whose Shortest Nonnull Run Is a k-Run or a $(\geq k)$-Run

Let us call $\underline{w}_k(n)$ the number of n-strings whose shortest nonnull run is a k-run, and let $\underline{w}_{\geq k}(n)$ be the number of n-strings whose shortest run is a $(\geq k)$-run. We therefore assume $k \geq 1$ in this section. These problems have been studied by Makri et al. [?] in the context of probabilistic success runs.

Both quantities are related as follows:

$$\underline{w}_{\geq k}(n) = \sum_{j=k}^{n} \underline{w}_j(n),$$

and so we also have

$$\underline{w}_k(n) = \underline{w}_{\geq k}(n) - \underline{w}_{\geq (k+1)}(n). \tag{86}$$

The enumeration of $\underline{w}_{\geq k}(n)$ and that of $\underline{w}_k(n)$ follow directly from the results in Section **??**, because for $k > 1$ we have the following relation:

$$\underline{w}_{\geq k}(n) = w_{\leq (k-1)}(n, 0) - 1, \tag{87}$$

whereas

$$\underline{w}_{\geq 1}(n) = 2^n - 1. \tag{88}$$

We subtract one both in (**??**) and in (**??**) to discount the all-zeros n-string. Of course, considering (**??**), we have that the connection in (**??**) can also be expressed using the general enumeration $w_{\underline{k} \leq \bar{k}}(n, m)$ studied in Section **??**, but (**??**) allows us to simplify our presentation and to obtain explicit expressions.

2.7.1 Recurrences

From (**??**) and (**??**), a recurrence for $\underline{w}_{\geq k}(n)$ when $k > 1$ is

$$\underline{w}_{\geq k}(n) = \underline{w}_{\geq k}(n-1) + \sum_{i=k}^{n} \underline{w}_{\geq k}(n-(i+1)) + \max(n-k+1, 0), \tag{89}$$

and from (**??**) and (**??**) an alternative recurrence for $k > 1$ is

$$\underline{w}_{\geq k}(n) = \; 2\,\underline{w}_{\geq k}(n-1) - \underline{w}_{\geq k}(n-2) + \underline{w}_{\geq k}(n-(k+1)) + [\![k \leq n]\!]. \tag{90}$$

Considering (**??**) and (**??**), the initialisation of these two recurrences is

$$\underline{w}_{\geq k}(-1) = \underline{w}_{\geq k}(0) = 0. \tag{91}$$

From (**??**) and (**??**) we may obtain recurrences for $\underline{w}_k(n)$ when $k > 1$. Using (**??**) in (**??**) the first recurrence is

$$\underline{w}_k(n) = \underline{w}_k(n-1) + \sum_{i=k+1}^{n} \underline{w}_k(n-(i+1)) + \sum_{j=k}^{n} \underline{w}_j(n-(k+1)) + [\![k \le n]\!], \tag{92}$$

where we have used $\max(n-k+1,0) - \max(n-k,0) = [\![k \le n]\!]$. Using (**??**) in (**??**) we get the alternative recurrence

$$\underline{w}_k(n) = 2\,\underline{w}_k(n-1) - \underline{w}_k(n-2) + \sum_{j=k}^{n} \underline{w}_j(n-(k+1)) - \sum_{j=k+1}^{n} \underline{w}_j(n-(k+2)) + [\![k = n]\!], \tag{93}$$

where we have used $[\![k \le n]\!] - [\![k+1 \le n]\!] = [\![k = n]\!]$. Inputting (**??**) in (**??**), we see that both (**??**) and (**??**) are initialised by

$$\underline{w}_k(-1) = \underline{w}_k(0) = 0.$$

Finally, we address the $k = 1$ case. From (**??**) and (**??**) we obtain $\underline{w}_1(n) = 2^n - 1 - \underline{w}_{\ge 2}(n)$, which can be evaluated using recurrences (**??**) or (**??**).

Remark. The same strategy in (**??**) was used by Makri et al. [**?**, Thm. 2.1.2] to get, in their notation, $P(M_n = k)$ from an explicit computation for $P(M_n \ge k)$. These two probabilities correspond to the enumerative quantities $\underline{w}_k(n)$ and $\underline{w}_{\ge k}(n)$. Also, recurrence (**??**) is the exact enumerative counterpart of the probabilistic recurrence given by Makri et al. for $P(M_n \ge k)$ [**?**, Thm. 2.1.3]. These authors derived their recurrence from their explicit expression, rather than the other way around, and they did not provide initialisation values.

□

2.7.2 Generating Functions

We may derive generating functions from the recurrences in the previous section, but the simplest way to obtain them is by exploiting (**??**) and (**??**). From (**??**) and (**??**), the ogf $\underline{W}_{\ge k}(x) = \sum_n \underline{w}_{\ge k}(n)\, x^n$ for $k > 1$ is

$$\underline{W}_{\ge k}(x) = \frac{1-x}{x(1-2x+x^2-x^{k+1})} - \frac{1}{1-x}, \tag{94}$$

whereas from (**??**) we have

$$\underline{W}_{\ge 1}(x) = \frac{1}{1-2x} - \frac{1}{1-x}. \tag{95}$$

Finally, from (**??**) and (**??**), for $k > 1$ the ogf $\underline{W}_k(x) = \sum_n \underline{w}_k(n)\, x^n$ is

$$\underline{W}_k(x) = \underline{W}_{\ge k}(x) - \underline{W}_{\ge (k+1)}(x)$$

$$= \frac{x^k(1-x)^2}{((1-x)^2 - x^{k+1})((1-x)^2 - x^{k+2})}, \tag{96}$$

whereas from (**??**), (**??**) and (**??**) we have that

$$\underline{W}_1(x) = \frac{1}{1-2x} - \frac{1-x}{x(1-2x+x^2-x^3)}. \tag{97}$$

To conclude, we observe that both (**??**) and (**??**) happen to be valid not only for $k > 1$ but also for $k = 1$. That is to say, for $k = 1$ and $n \geq 1$ (**??**) and (**??**) have the same coefficients of x^n as (**??**) and (**??**), respectively.

2.7.3 Explicit Expressions

Explicit expressions for $\underline{w}_{\geq k}(n)$ and $\underline{w}_k(n)$ are available through (**??**), (**??**) and (**??**). These single-summation expressions are simpler than the double-summation combinatorial expressions found by Makri et al [**?**, Thms. 2.1.1 and 2.2.2 with $p = q = 1/2$ and multiplied by 2^n].

2.7.4 OEIS

We report next any sequences related to $\underline{w}_k(n)$ and $\underline{w}_{\geq k}(n)$ found in the OEIS.

- Sequences emanating from $\underline{w}_k(n)$: none of the $\underline{w}_k(n)$ sequences, for $k \geq 1$, were in the OEIS previous to this work; below are the currently listed sequences:

  $\underline{w}_1(n) = w_1^{\geq 1}(n) = w_{\leq 1}^{\geq 1}(n)$ is A384153.

  $\underline{w}_2(n)$ is A384154.

  $\underline{w}_3(n)$ is A384155.

  $\underline{w}_k(n)$ is the kth column of A388718.

- The sequences emanating from $\underline{w}_{\geq k}(n)$ in the OEIS are:

  $\underline{w}_{\geq 1}(n) = 2^n - 1$ is A000225.

  $\underline{w}_{\geq 2}(n+2)$ is A077855.

  $\underline{w}_{\geq 3}(n)$ is A130578 (Number of different possible rows —or columns— in an $n \times n$ crossword puzzle).

  $\underline{w}_{\geq 4}(n)$ is A209231 (Number of binary words of length n such that there is at least one 0 and every run of consecutive 0's is of length ≥ 4).

  $\underline{w}_{\geq k}(n)$ is the kth column of A388547.

2.8 Number of n-Strings that Contain Exactly m Nonnull p-Parity Runs

Let $w_{[p]}(n,m)$ represent the number of n-strings that contain exactly m nonnull p-parity runs, i.e., exactly m runs whose lengths $k_1,\dots,k_m$ are strictly greater than zero and have parity $\mathrm{mod}(k_i,2)=p$ for $i=1,\dots,m$. The n-strings that we enumerate may have more than m nonnull runs, as long as the lengths of all additional runs have parity opposite to p, and they may also have any number of null runs. For the first time in this paper we address an enumeration unrelated to the results in Section **??**. To the best of our knowledge, the only authors that have addressed a similar problem are Grimaldi and Heubach [**?**], who studied the number of n-strings devoid of odd runs —a special case of $w_{[p]}(n,m)$ with $p=1$ and $m=0$.

First of all, we state a necessary condition similar to (**??**).

Necessary Condition. (Existence of n-strings containing m nonnull p-parity runs)

$$w_{[p]}(n,m)>0 \quad\Longrightarrow\quad 0\le m\le \left\lfloor \frac{n+1}{2+[\![p=0]\!]}\right\rfloor. \tag{98}$$

□

This is really the same condition as (**??**) in Section **??**, just noting that the lengths of all nonnull p-parity runs are lower bounded by $\underline{k}=1+[\![p=0]\!]$.

We also study the number of n-strings that contain at least m nonnull p-parity runs, which we denote by $w_{[p]}^{m}(n)$. This enumeration can be obtained from $w_{[p]}(n,m)$ as

$$w_{[p]}^{\ge m}(n)=\sum_{t=m}^{\left\lfloor \frac{n+1}{2+[\![p=0]\!]}\right\rfloor} w_{[p]}(n,t). \tag{99}$$

2.8.1 Recurrences

We can produce a recurrence to enumerate $w_{[p]}(n,m)$ with a similar strategy as in Section **??**. Since we are not counting null runs, we can split the quantity $w_{[p]}(n,m)$ into two contributions:

a) The n-strings that begin with 0 contribute $w_{[p]}(n-1,m)$ to $w_{[p]}(n,m)$.

b) As for the n-strings that begin with 1, those that start with an odd i-run contribute $w_{[p]}(n-(i+1),m-[\![p=1]\!])$ to $w_{[p]}(n,m)$, whereas those that start with an even nonnull i-run contribute $w_{[p]}(n-(i+1),m-[\![p=0]\!])$. Equivalently, the n-strings that start with a nonnull i-run contribute $w_{[p]}(n-(i+1),m-[\![p=\mathrm{mod}(i,2)]\!])$ to $w_{[p]}(n,m)$.

Collecting these two contributions we get the following recurrence:

$$w_{[p]}(n,m) = w_{[p]}(n-1,m) + \sum_{i=1}^{n} w_{[p]}\big(n-(i+1), m - [\![p = \mathrm{mod}(i,2)]\!]\big). \tag{100}$$

To find initial values for recurrence (**??**) we consider the case $n = 1$, in which we know by inspection that

$$w_{[p]}(1,m) = 2\,[\![p=0]\!][\![m=0]\!] + [\![p=1]\!][\![m=1]\!]. \tag{101}$$

On the other hand, setting $n = 1$ in (**??**) and taking into account (**??**) yields

$$w_{[p]}(1,m) = w_{[p]}(0,m) + w_{[p]}(-1, m - [\![p=1]\!]). \tag{102}$$

We wish (**??**) to equal (**??**).

Taking into account (**??**), equality between the two expressions is

achieved by choosing

$$w_{[p]}(-1,m) = w_{[p]}(0,m) = [\![m=0]\!], \tag{103}$$

which are therefore the initial values of (**??**).

We produce next a recurrence for $w_{[p]}^{\geq m}(n)$ by applying (**??**) to (**??**), which yields

$$w_{[p]}^{\geq m}(n) = w_{[p]}^{\geq m}(n-1) + \sum_{i=1}^{n}\left(w_{[p]}^{\geq m}(n-(i+1)) + [\![p = \mathrm{mod}(i,2)]\!]\, w_{[p]}(n-(i+1), m-1)\right). \tag{104}$$

Even if this recurrence depends on $w_{[p]}(n,m)$, we see in the next section that it suffices to obtain the ogf of $w_{[p]}^{\geq m}(n)$.

2.8.2 Generating Functions

Let us next obtain the ogf $W_{[p]}(x,y) = \sum_{n,m} w_{[p]}(n,m)\, x^n y^m$. First of all, we obtain a recurrence valid for all n and m. Setting $n = -1$ in (**??**) we get $w_{[p]}(-1,m) = 0$ instead of the correct value $w_{[p]}(-1,m) = [\![m=0]\!]$ given by (**??**). We can can "fix" this by adding $[\![n=-1]\!][\![m=0]\!]$ to (**??**). On the other hand, setting $n = 0$ in (**??**) after this change yields $w_{[p]}(0,m) = [\![m=0]\!]$, which is correct. We thus have an extended recurrence valid for all n and m, which we use subsequently.

Next, we need a recurrence without a full-history summation to be able to deduce the ogf. We can obtain one by calculating $w_{[p]}(n,m) - w_{[p]}(n-2,m)$ using the extended recurrence. This yields

$$\begin{aligned} w_{[p]}(n,m) - w_{[p]}(n-2,m) &= w_{[p]}(n-1,m) - w_{[p]}(n-3,m) \\ &\quad + w_{[p]}\big(n-2, m-[\![p=1]\!]\big) + w_{[p]}\big(n-3, m-[\![p=0]\!]\big) \end{aligned}$$

$$+\big(\llbracket n=-1\rrbracket-\llbracket n=1\rrbracket\big)\llbracket m=0\rrbracket. \tag{105}$$

Remark. Unlike in Sections **??** and **??**, obtaining the above difference before making (**??**) valid for all n and m does not render a valid recurrence. In Section **??** we were able to first separately reason each recurrence and then verify that the second one could be obtained from the first, but this is not the case here. In general, we can freely operate with a recurrence to obtain a new equivalent recursive relation as long as the original recurrence is valid for all integer values of its arguments. If this is not true, then the manipulation is not guaranteed to render a valid relation. □

By multiplying next (**??**) on both sides by $x^n y^m$ and then adding on n and m we find

$$\begin{aligned}W_{[p]}(x,y)-x^2W_{[p]}(x,y)&=x\,W_{[p]}(x,y)-x^3W_{[p]}(x,y)+x^2y^{\llbracket p=1\rrbracket}W_{[p]}(x,y)\\&\quad+x^3y^{\llbracket p=0\rrbracket}W_{[p]}(x,y)+x^{-1}+x,\end{aligned}$$

which yields the ogf

$$W_{[p]}(x,y)=\frac{1-x^2}{x\big(1-x-x^2\big(1+y^{\llbracket p=1\rrbracket}\big)+x^3\big(1-y^{\llbracket p=0\rrbracket}\big)\big)}. \tag{106}$$

Just like in Section **??**, obtaining $[y^m]W_{[p]}(x,y)$ is straightforward by putting (**??**) as a function of $(1-cy)^{-1}$. Doing so we get

$$[y^m]W_{[p]}(x,y)=\frac{(1-x^2)\,x^{(2+\llbracket p=0\rrbracket)m}}{x\left(1-x-(1+\llbracket p=0\rrbracket)\,x^2+\llbracket p=0\rrbracket\,x^3\right)^{m+1}}, \tag{107}$$

which is the ogf enumerating the binary strings that contain exactly m nonnull p-parity runs.

Remark. In the special case $p=1$ and $m=0$ (i.e., number of n-strings devoid of odd runs) the ogf (**??**) is $[y^0]W_{[1]}(x,y)=(1-x^2)/\big(x(1-x-x^2)\big)$, which, as indicated by Grimaldi and Heubach [**?**, Thm. 2], corresponds to

$$w_{[1]}(n,0)=F^{(2)}_{n+1}, \tag{108}$$

i.e., the Fibonacci sequence shifted one position. □

To conclude this section, we derive the ogf $W^{\geq m}_{[p]}(x)=\sum_n w^{\geq m}_{[p]}(n)\,x^n$. We first obtain a recurrence free from n-dependent summations by subtracting $w^{\geq m}_{[p]}(n-2)$ from $w^{\geq m}_{[p]}(n)$ using (**??**), which gives

$$w^{\geq m}_{[p]}(n)-w^{\geq m}_{[p]}(n-2)=w^{\geq m}_{[p]}(n-1)-w^{\geq m}_{[p]}(n-3)$$

$$+ [\![p=1]\!] w_{[p]}(n-2,m-1) + [\![p=0]\!] w_{[p]}(n-3,m-1). \quad (109)$$

As usual, by multiplying (**??**) on both sides by x^n and then adding over n we get

$$\begin{aligned} W_{[p]}^{\geq m}(x) &= \frac{[\![p=1]\!]\, x^2 + [\![p=0]\!]\, x^3}{1-x-2x^2}\, [y^{m-1}] W_{[p]}(x,y) \\ &= \frac{(1-x^2)\big([\![p=1]\!]\, x + [\![p=0]\!]\, x^2\big)\left(x^{2+[\![p=0]\!]}\right)^{m-1}}{(1-x-2x^2)\big(1-x-(1+[\![p=0]\!])\, x^2 + [\![p=0]\!]\, x^3\big)^m}, \end{aligned} \quad (110)$$

where we have used (**??**) in the last step. This is the ogf enumerating the binary strings that contain at least m nonnull p-parity runs—see (**??**).

2.8.3 Explicit Expressions

In this section we find explicit expressions for the two cases of $w_{[p]}(n,m)$. We first deal with the case $p=1$. Rewriting (**??**) as $[y^m]W_{[1]}(x,y) = x^{2m-1}(1-x^2)^{-m}(1-x/(1-x^2))^{-m-1}$ and applying the negative binomial theorem twice, we can expand this ogf as

$$[y^m]W_{[1]}(x,y) = \sum_{s\geq 0}\sum_{t\geq 0} \binom{s+m}{m}\binom{t+s+m-1}{t} x^{s+2t+2m-1}. \quad (111)$$

We now find the coefficient of x^n in this expression by finding the nonnegative indices s and t such that $s+2t+2m-1=n$. The minimum of t happens for $s=0$, and therefore $t \leq \lfloor (n-2m+1)/2 \rfloor$. Given a value of t, we have $s = n-2m-t+1$. Considering these solutions and (**??**) we thus have that

$$w_{[1]}(n,m) = \sum_{t=0}^{\lfloor \frac{n-2m+1}{2} \rfloor} \binom{n-m-2t+1}{m}\binom{n-m-t}{t}. \quad (112)$$

As for the case $p=0$, we now rewrite (**??**) as $[y^m]W_{[0]}(x,y) = x^{3m-1}(1-x^2)^{-m}(1-x-x^2/(1-x^2))^{-m-1}$ before applying in succession the negative binomial theorem, the binomial theorem, and again the negative binomial theorem. This yields the expansion

$$[y^m]W_{[0]}(x,y) = \sum_{s\geq 0}\sum_{t\geq 0}\sum_{r\geq 0} \binom{s+m}{m}\binom{s}{t}\binom{r+t+m-1}{r} x^{s+t+2r+3m-1}. \quad (113)$$

To extract the coefficient of x^n we have to find the indices that fulfil $s+t+2r+3m-1=n$. By setting $s=t=0$ we have the upper bound $r \leq \lfloor (n-3m+1)/2 \rfloor$. Likewise, by setting $t=0$ for a fixed value of r we have $t \leq n-2r-3m+1$. Finally, t is determined by the equation above given s and t. All this considered, we have from (**??**) that

$$w_{[0]}(n,m) = \sum_{r=0}^{\lfloor \frac{n-3m+1}{2} \rfloor} \sum_{s=0}^{n-2r-3m+1} \binom{s+m}{s}\binom{s}{n-2r-s-3m+1}\binom{n-r-s-2m}{r}. \quad (114)$$

2.8.4 OEIS

The sequences in the OEIS emanating from the enumeration studied in this section are:

- $m = 0$

  $w_{[1]}(n-1, 0)$ is A000045 (Fibonacci numbers) —see Remark **??**.

  $w_{[0]}(n-1, 0)$ is A028495.

- $m = 1$

  $w_{[1]}(n, 1)$ is A029907.

  $w_{[0]}(n, 1)$ is A384497.

- $w_{[1]}(n, m)$ is the mth column of A119473.

  $w_{[0]}(n, m)$ is the mth column of A391669.

- Sequences from $w^{\geq m}_{[p]}(n)$ —see (**??**).

  $w^{\geq 1}_{[1]}(n)$ is A027934.

  $w^{\geq 1}_{[0]}(n)$ is A387332.

3 Extensions to Probabilistic Runs

At several points in Section **??** we have specialised probabilistic results about success runs from other authors (e.g., [?, ?]) in order to compare them to our enumerations. Thus, one might conclude that the enumerative results that we have given so far are less general than their probabilistic counterparts. But this is a two-way street: as we see in this section, every enumerative recurrence in Section **??** also has a direct probabilistic translation into the case in which the n-strings are outcomes from n independent and identically distributed (iid) Bernoulli random variables with parameter q, i.e., when each bit in the n-string is independently drawn with probability $0 < q < 1$ of getting a 1.

Since probabilistic results are not our main goal in this paper, we mainly examine how to extend our most relevant enumerative results in Sections **??** and **??** to the probabilistic scenario described above. Observe that this also includes the probabilistic extensions of our results in Sections **??**–**??**, which are, essentially, special cases or consequences of Section **??**. In fact, we also look into the probabilistic extension of the enumerative results in Sections **??**, **??** and **??**, since these are connected with several prior findings in the literature.

In order not to overload notation, it is understood that all recurrences, probability generating functions, explicit expressions, and moments in this section implicitly depend on q.

3.1 Probability that an n-String Contains Exactly m $(\underline{k} \leq \bar{k})$-Runs

We call $\pi_{\underline{k}\leq\bar{k}}(n,m)$ the probability that an n-string contains exactly m $(\underline{k} \leq \bar{k})$-runs. The obvious case is $q = 1/2$, in which $\pi_{\underline{k}\leq\bar{k}}(n,m) = w_{\underline{k}\leq\bar{k}}(n,m)/2^n$. Likewise, the probability that an n-string contains at least m $(\underline{k} \leq \bar{k})$-runs is $\pi^{\geq m}_{\underline{k}\leq\bar{k}}(n) = w^{\geq m}_{\underline{k}\leq\bar{k}}(n)/2^n$ in this case. We examine next how to get these quantities for arbitrary q.

3.1.1 Recurrences

We may directly write two equivalent probability recurrences for $\pi_{\underline{k}\leq\bar{k}}(n,m)$ by using the two enumerative recurrences in Section **??** in conjunction with the law of total probabilities. From (**??**) we have the recurrence

$$\begin{aligned}\pi_{\underline{k}\leq\bar{k}}(n,m) = (1-q)\Bigg(\sum_{i=\underline{k}}^{\bar{k}} q^i \pi_{\underline{k}\leq\bar{k}}\big(n-(i+1),m-1\big) + \sum_{i=0}^{\underline{k}-1} q^i \pi_{\underline{k}\leq\bar{k}}\big(n-(i+1),m\big)\\ + \sum_{i=\bar{k}+1}^{n} q^i \pi_{\underline{k}\leq\bar{k}}\big(n-(i+1),m\big)\Bigg).\end{aligned} \tag{115}$$

whereas from (**??**) we have the alternative recurrence

$$\begin{aligned}\pi_{\underline{k}\leq\bar{k}}(n,m) = \pi_{\underline{k}\leq\bar{k}}(n-1,m) + q^{\underline{k}}(1-q)\Big(\pi_{\underline{k}\leq\bar{k}}\big(n-(\underline{k}+1),m-1\big) - \pi_{\underline{k}\leq\bar{k}}\big(n-(\underline{k}+1),m\big)\Big)\\ - q^{\bar{k}+1}(1-q)\Big(\pi_{\underline{k}\leq\bar{k}}\big(n-(\bar{k}+2),m-1\big) - \pi_{\underline{k}\leq\bar{k}}\big(n-(\bar{k}+2),m\big)\Big).\end{aligned} \tag{116}$$

Notice that the 2 factor in the first term of (**??**) becomes 1 in the equivalent term of (**??**). This is because our probabilistic estimate for the value of $\pi_{\underline{k}\leq\bar{k}}(n,m)$ from $\pi_{\underline{k}\leq\bar{k}}(n-1,m)$ is $q\,\pi_{\underline{k}\leq\bar{k}}(n-1,m) + (1-q)\,\pi_{\underline{k}\leq\bar{k}}(n-1,m)$. Of course, we can also get (**??**) from (**??**), using $\pi_{\underline{k}\leq\bar{k}}(n,m) - q\,\pi_{\underline{k}\leq\bar{k}}(n-1,m)$. In the following we work with the simpler recurrence (**??**), as it is also directly amenable to obtaining the probability generating function (pgf) of $\pi_{\underline{k}\leq\bar{k}}(n,m)$. As we have done before, to find initial values we use the probability for $n = 1$, which, by inspection, is

$$\begin{aligned}\pi_{\underline{k}\leq\bar{k}}(1,m) = \big(q\,[\![\underline{k}=0]\!][\![\bar{k}=0]\!] + (1-q)[\![\underline{k}=1]\!] + [\![\underline{k}>1]\!]\big)[\![m=0]\!]\\ + q\left([\![\underline{k}=0]\!][\![\bar{k}\geq 1]\!] + [\![\underline{k}=1]\!]\right)[\![m=1]\!] + (1-q)\,[\![\underline{k}=0]\!][\![m=2]\!].\end{aligned} \tag{117}$$

Now, setting $n = 1$ in (**??**) yields

$$\pi_{\underline{k}\leq\bar{k}}(1,m) = \pi_{\underline{k}\leq\bar{k}}(0,m) + q^{\underline{k}}(1-q)\Big(\pi_{\underline{k}\leq\bar{k}}(-\underline{k},m-1) - \pi_{\underline{k}\leq\bar{k}}(-\underline{k},m)\Big)$$

$$-\,q^{\bar{k}+1}(1-q)\Big(\pi_{\underline{k}\le\bar{k}}(-(\bar{k}+1),m-1)-\pi_{\underline{k}\le\bar{k}}(-(\bar{k}+1),m)\Big),\tag{118}$$

We want (??) to equal (??).

Taking necessary condition (??) into account, we can achieve this equality for all m and $0\le\underline{k}\le\bar{k}$ by letting

$$\pi_{\underline{k}\le\bar{k}}(-1,m)=\frac{1}{1-q}[\![m=0]\!],\tag{119}$$

$$\pi_{\underline{k}\le\bar{k}}(0,m)=\big[\!\big[m=[\![\underline{k}=0]\!]\big]\!\big],\tag{120}$$

which therefore constitute the initialisation of (??). These values initialise (??) as well.

Remark. As if a binary string of length -1 were not weird enough, take a moment to ponder that $1/(1-q)$ is greater than one in the initialisation (??) of the probability recurrence just given. □

As for $\pi^{\ge m}_{\underline{k}\le\bar{k}}(n)$, we have from (??) that $\pi^{\ge m}_{\underline{k}\le\bar{k}}(n)=\sum_{t=m}^{\lfloor(n+1)/(\underline{k}+1)\rfloor}\pi_{\underline{k}\le\bar{k}}(n,t)$. Applying this expression to (??) we get the recurrence

$$\pi^{\ge m}_{\underline{k}\le\bar{k}}(n)=\pi^{\ge m}_{\underline{k}\le\bar{k}}(n-1)+q^{\underline{k}}(1-q)\,\pi_{\underline{k}\le\bar{k}}(n-(\underline{k}+1),m-1)-q^{\bar{k}+1}(1-q)\,\pi_{\underline{k}\le\bar{k}}(n-(\bar{k}+2),m-1),\tag{121}$$

which, as we see in the next section, suffices to get the pgf of $\pi^{\ge m}_{\underline{k}\le\bar{k}}(n)$.

3.1.2 Probability Generating Functions

We can now obtain the bivariate pgf $\Pi_{\underline{k}\le\bar{k}}(x,y)=\sum_{n,m}\pi_{\underline{k}\le\bar{k}}(n,m)\,x^n y^m$ following the same steps as in Section **??**. In this case we need to add $\big([\![n=-1]\!]-q[\![n=0]\!]\big)[\![m=0]\!]/(1-q)$ to (??) to render it valid for all n and m. Using this extended recurrence, we directly get

$$\Pi_{\underline{k}\le\bar{k}}(x,y)=\frac{1-q\,x}{(1-q)\,x\Big(1-x-(y-1)(1-q)(q^{\underline{k}}x^{\underline{k}+1}-q^{\bar{k}+1}x^{\bar{k}+2})\Big)}.\tag{122}$$

We can extract the coefficient of y^m in this pgf following the same steps as in Section **??**, which in this case lead us to obtain

$$[y^m]\Pi_{\underline{k}\le\bar{k}}(x,y)=\frac{(1-q)^{m-1}(1-q\,x)\big(q^{\underline{k}}x^{\underline{k}+1}-q^{\bar{k}+1}x^{\bar{k}+2}\big)^m}{x\Big(1-x+(1-q)\big(q^{\underline{k}}x^{\underline{k}+1}-q^{\bar{k}+1}x^{\bar{k}+2}\big)\Big)^{m+1}}.\tag{123}$$

This is the pgf giving the probability that a binary string contains exactly m ($\underline{k}\le\bar{k}$)-runs.

The pgf $\Pi^{\geq m}_{\underline{k}\leq\bar{k}}(x)=\sum_n \pi^{\geq m}_{\underline{k}\leq\bar{k}}(n)\,x^n$ is obtained by multiplying (**??**) on both sides by x^n and then adding over n, which yields

$$\begin{aligned}\Pi^{\geq m}_{\underline{k}\leq\bar{k}}(x)&=\frac{1-q}{1-x}\big(q^{\underline{k}}x^{\underline{k}+1}-q^{\bar{k}+1}x^{\bar{k}+2}\big)[y^{m-1}]\Pi_{\underline{k}\leq\bar{k}}(x,y)\\&=\frac{(1-q\,x)}{(1-q)\,x\,(1-x)}\left(\frac{(1-q)\big(q^{\underline{k}}x^{\underline{k}+1}-q^{\bar{k}+1}x^{\bar{k}+2}\big)}{1-x+(1-q)\big(q^{\underline{k}}x^{\underline{k}+1}-q^{\bar{k}+1}x^{\bar{k}+2}\big)}\right)^m,\end{aligned}\tag{124}$$

where we have used (**??**). This is the pgf giving the probability that a binary string contains at least m ($\underline{k}\leq\bar{k}$)-runs.

Remark. As a rule, pgf (**??**) allows us to pursue explicit expressions for $\pi_{\underline{k}\leq\bar{k}}(n,m)$ in particular cases of $\underline{k}$ and $\bar{k}$, in the same way that we have sought explicit expressions for $w_{\underline{k}\leq\bar{k}}(n,m)$ in particular cases of $\underline{k}$ and $\bar{k}$ in Sections **??**–**??**. If we are interested instead in the computation of moments for a given n, we can do so through (**??**) but sometimes the alternative route through $[x^n]\Pi_{\underline{k}\leq\bar{k}}(x,y)$ is cleaner. Although getting the coefficient of x^n is not as straightforward as getting the coefficient of y^m (it cannot usually be done in closed form), in special cases one can still get manageable expressions for $[x^n]\Pi_{\underline{k}\leq\bar{k}}(x,y)$ through the binomial theorem. The extra work involved sometimes actually pays off, as neater explicit expressions can sometimes be determined through $[x^n]\Pi_{\underline{k}\leq\bar{k}}(x,y)$. □

3.2 Probability that an n-String Contains Exactly m k-Runs

We denote by $\pi_k(n,m)$ the probability that an n-string contains exactly m k-runs. This is a special case of the probability in Section **??** with $\underline{k}=\bar{k}=k$, and thus

$$\pi_k(n,m)=\pi_{k\leq k}(n,m).\tag{125}$$

We discuss this special case in some detail because Makri and Psillakis [**?**] previously studied this probability. Marbe [**?**, **?**] and Cochran [**?**] also gave an expectation connected to it.

3.2.1 Recurrences

Using (**??**), recurrence (**??**) becomes

$$\pi_k(n,m)=(1-q)\left(q^k\pi_k\big(n-(k+1),m-1\big)+\sum_{\substack{i=0\\i\neq k}}^{n}q^i\pi_k\big(n-(i+1),m\big)\right)\tag{126}$$

whereas from (**??**) we have the alternative recurrence

$$\pi_k(n,m)=\pi_k(n-1,m)+q^k(1-q)\Big(\pi_k\big(n-(k+1),m-1\big)-\pi_k\big(n-(k+1),m\big)\Big)$$

$$
- q^{k+1}(1-q)\Big(\pi_k\big(n-(k+2),m-1\big) - \pi_k\big(n-(k+2),m\big)\Big), \tag{127}
$$

which from (**??**) and (**??**) are both initialised by

$$
\begin{aligned}
\pi_k(-1,m) &= \frac{1}{1-q}[\![m=0]\!],\\
\pi_k(0,m) &= \big[\![m=[\![k=0]\!]\big]\!].
\end{aligned}
$$

3.2.2 Probability Generating Functions

By setting $\underline{k} = \overline{k} = k$ in (**??**) we get the pgf $\Pi_k(x,y) = \sum_{n,m} \pi_k(n,m)\, x^n y^m$:

$$
\Pi_k(x,y) = \frac{1-q\,x}{(1-q)\,x\,\Big(1-x-(y-1)(1-q)(q^k x^{k+1} - q^{k+1}x^{k+2})\Big)}. \tag{128}
$$

Similarly, from (**??**) we have

$$
[y^m]\Pi_k(x,y) = \frac{(1-q)^{m-1}(1-q\,x)\big(q^k x^{k+1} - q^{k+1}x^{k+2}\big)^m}{x\Big(1-x+(1-q)\big(q^k x^{k+1} - q^{k+1}x^{k+2}\big)\Big)^{m+1}}, \tag{129}
$$

which is the pgf giving the probability that a binary string contains exactly m k-runs, whereas from (**??**) we get

$$
\Pi_k^{\geq m}(x) = \frac{(1-q\,x)}{(1-q)\,x\,(1-x)}\left(\frac{(1-q)\big(q^k x^{k+1} - q^{k+1}x^{k+2}\big)}{1-x+(1-q)\big(q^k x^{k+1} - q^{k+1}x^{k+2}\big)}\right)^m. \tag{130}
$$

This is the pgf giving the probability that a binary string contains at least m k-runs.

3.2.3 Explicit Expression

In this section we obtain an explicit expression for $\pi_k(n,m)$ by extracting the coefficient of x^n in (**??**). Applying the negative binomial theorem and the binomial theorem twice, we can see that (**??**) can be expanded as

$$
[y^m]\Pi_k(x,y) = \sum_{s\geq 0}\sum_{t\geq 0}\sum_{r\geq 0}\binom{s+m}{m}\binom{s}{t}\binom{m+t+1}{r}(-1)^{t+r}(1-q)^{t+m-1}q^{k(t+m)+r}x^{(k+1)m-1+s+kt+r}. \tag{131}
$$

The coefficient of x^n above corresponds to the nonnegative solutions of indices s,t and r in $(k+1)m-1+s+kt+r=n$, or, equivalently, in $s+kt+r=e$ using (**??**) with $\underline{k}=k$. For

$k > 1$, the maximum of t happens for $s = r = 0$, and thus $t \leq \lfloor e/k \rfloor$. Similarly, $r = 0$ let us see that $s \leq e - kt$, whereas r is determined by the previous equation for any two values of t and s. If $k = 0$ then $t \leq e$, as the second binomial is zero otherwise. Collecting these solutions, we have from (**??**) that

$$\pi_k(n, m) = \sum_{t=0}^{\min(\lfloor \frac{e}{k} \rfloor, e)} \sum_{s=0}^{e-kt} \binom{s+m}{m} \binom{s}{t} \binom{m+t+1}{e-kt-s} (-1)^{e-(k-1)t-s} (1-q)^{t+m-1} q^{km+e-s}. \tag{132}$$

This is equivalent to the explicit expression for $\pi_k(n, m)$ given by Makri and Psillakis [**?**, Thm. 2.1] —in their notation $P(E_{n,k} = m) = \pi_k(n, m)$.

3.2.4 Moments

The pgf (**??**) also allows us to determine the factorial moments of the random variable (rv) modelling the number of k-runs in an n-string drawn at random, which we denote by $M_{k,n}$. With standard probabilistic notation we have that

$$\Pr\big(M_{k,n} = m\big) = \pi_k(n, m). \tag{133}$$

The first factorial moment (expectation) can be obtained as $\mathrm{E}(M_{k,n}) = [x^n](\partial/\partial y)\Pi_k(x, y)|_{y=1}$. Thus, we have determine next the coefficient of x^n in

$$\left.\frac{\partial \Pi_k(x, y)}{\partial y}\right|_{y=1} = \frac{(1-qx)^2 q^k x^k}{(1-x)^2}. \tag{134}$$

By applying the negative binomial theorem and the binomial theorem we have that

$$\left.\frac{\partial \Pi_k(x, y)}{\partial y}\right|_{y=1} = \sum_{t \geq 0} \sum_{s \geq 0} \binom{t+1}{t} \binom{2}{s} (-1)^s q^{s+k} x^{k+s+t}. \tag{135}$$

From $n = k + s + t$ we have that $t \leq n - k$ and $s = n - k - t$, and therefore

$$\mathrm{E}(M_{k,n}) = \sum_{t=0}^{n-k} (t+1) \binom{2}{n-k-t} (-1)^{n-k-t} q^{n-t}. \tag{136}$$

There are three cases in this expression, depending on $n - k = 0, 1$ or $n - k \geq 2$, but the last two cases have the same solution. Evaluating them one sees that

$$\mathrm{E}(M_{k,n}) = \Big(\big((n-k-1)(1-q) + 2\big)(1-q)\, [\![k < n]\!] + [\![k = n]\!]\Big)\, q^k. \tag{137}$$

The second factorial moment $\mathrm{E}(M_{k,n}(M_{k,n} - 1))$ is the coefficient of x^n in

$$\left.\frac{\partial^2 \Pi_k(x, y)}{\partial y^2}\right|_{y=1} = \frac{2(1-q)(1-qx)^3 q^{2k} x^{2k+1}}{(1-x)^3}. \tag{138}$$

As before, this can be expanded as

$$\left.\frac{\partial^2 \Pi_k(x,y)}{\partial y^2}\right|_{y=1} = 2(1-q)\sum_{t\geq 0}\sum_{s\geq 0}\binom{t+2}{t}\binom{3}{s}(-1)^s q^{2k+s}x^{2k+1+t+s}, \tag{139}$$

and we now need to solve $2k+1+t+s=n$. Observing that $t \leq n-2k-1$ and $s=n-2k-1-t$, we have that

$$\mathrm{E}\big(M_{k,n}(M_{k,n}-1)\big) = (1-q)\sum_{t=0}^{n-2k-1}(t+2)(t+1)\binom{3}{n-2k-1-t}(-1)^{n-2k-1-t}q^{n-t-1}. \tag{140}$$

There are four nonzero cases in (**??**), which, letting $\xi = n-2k-1$, correspond to $\xi = 0,1,2$ and $\xi \geq 3$. The last three cases have the same solution, and after some algebra we can see that

$$\begin{aligned}\mathrm{E}\big(M_{k,n}(M_{k,n}-1)\big) = (1-q)q^{n-1}\Big((1-q)q^{-\xi}\big(2\,(1+q+q^2)+3\,(1-q^2)\xi+(1-q)^2\xi^2\big)[\![\xi>0]\!]\\ +2\,[\![\xi=0]\!]\Big)\end{aligned} \tag{141}$$

With (**??**) and (**??**) we can obtain the variance of $M_{k,n}$ using $\mathrm{Var}(M_{k,n}) = \mathrm{E}(M_{k,n}(M_{k,n}-1)) + \mathrm{E}(M_{k,n}) - \mathrm{E}^2(M_{k,n})$. For example, a trivial case is $\mathrm{E}(M_{n,n}) = q^n$ and $\mathrm{Var}(M_{n,n}) = q^n(1-q^n)$, which can also be seen by observing that $M_{n,n}$ is a Bernoulli rv with parameter q^n.

The expectation (**??**) was originally given by Marbe [**?**, **?**] and then put on a sounder footing by Cochran [**?**, Eq. (5)] —see more details in Section **??**, Remark **??**. The expectation and variance of $M_{0,n}$ were implicitly given by Bloom [**?**] —see discussion towards the end of Remark **??** in Section **??**.

3.3 Probability that an n-String Contains Exactly m $(\geq k)$-Runs

We denote by $\pi_{\geq k}(n,m)$ the probability that an n-string contains exactly m $(\geq k)$-runs. This is a special case of the probability in Section **??** with $\underline{k}=k$ and $\overline{k}=n$, and thus

$$\pi_{\geq k}(n,m) = \pi_{k\leq n}(n,m). \tag{142}$$

We discuss this special case in some detail because several authors have previously studied this probability [**?**, **?**, **?**, **?**, **?**].

3.3.1 Recurrences

Using (**??**), recurrence (**??**) becomes

$$\pi_{\geq k}(n,m) = (1-q)\left(\sum_{i=k}^{n} q^i \pi_{\geq k}(n-(i+1),m-1) + \sum_{i=0}^{k-1} q^i \pi_{\geq k}(n-(i+1),m)\right), \tag{143}$$

whereas the second recurrence (**??**) is now

$$\pi_{\geq k}(n,m) = \pi_{\geq k}(n-1,m) + q^k(1-q)\Big(\pi_{\geq k}(n-(k+1),m-1) - \pi_{\geq k}(n-(k+1),m)\Big). \tag{144}$$

From (**??**) and (**??**) both recurrences are initialised by

$$\pi_{\geq k}(-1,m) = \frac{1}{1-q}[\![m=0]\!], \tag{145}$$

$$\pi_{\geq k}(0,m) = \big[\!\!\big[m = [\![k=0]\!]\big]\!\!\big]. \tag{146}$$

Lastly, (**??**) becomes

$$\pi_{\geq k}^{\geq m}(n) = \pi_{\geq k}^{\geq m}(n-1) + q^k(1-q)\,\pi_{\geq k}(n-(k+1),m-1). \tag{147}$$

3.3.2 Probability Generating Functions

Although we can get valid pgfs by specialising (**??**), (**??**) and (**??**) with $\underline{k}=k$ and $\overline{k}=n$, the resulting expressions depend on n and require $k \leq n$. We obtain next simpler pgfs valid for all $k \geq 0$ by working directly with (**??**) and (**??**). The pgf $\Pi_{\geq k}(x,y) = \sum_n \pi_{\geq k}(n,m)\,x^n y^m$ can be obtained by making recurrence (**??**) valid for all n and m, multiplying it on both sides by $x^n y^m$, and then adding over n and m. This yields

$$\Pi_{\geq k}(x,y) = \frac{1-q\,x}{(1-q)\,x\,\big(1-x-(y-1)(1-q)q^k x^{k+1}\big)}. \tag{148}$$

Through the same method as before, we can extract the coefficient of y^m from from (**??**):

$$[y^m]\Pi_{\geq k}(x,y) = \frac{(1-q\,x)\big((1-q)q^k x^{k+1}\big)^m}{(1-q)\,x\,\big(1-x+(1-q)q^k x^{k+1}\big)^{m+1}}, \tag{149}$$

which is the pgf giving the probability that a binary string contains exactly m ($\geq k$)-runs.

Also, multiplying (**??**) on both sides by x^n and then adding over n yields

$$\Pi_{\geq k}^{\geq m}(x) = \frac{1-q}{1-x}\,q^k x^{k+1}[y^{m-1}]\Pi_{\geq k}(x,y) \tag{150}$$

$$= \frac{(1-q\,x)}{(1-q)\,x\,(1-x)}\left(\frac{(1-q)q^k x^{k+1}}{1-x+(1-q)q^k x^{k+1}}\right)^m \tag{151}$$

where in the second step above we have used (**??**). Expression (**??**) is the pgf giving the probability that a binary string contains at least m ($\geq k$)-runs.

We finally determine the coefficient of x^n in (**??**), as in this case this allows us to get cleaner expressions later. By applying the negative binomial theorem and then the binomial theorem, we can express (**??**) as

$$\Pi_{\geq k}(x,y) = (x^{-1} - q) \sum_{s\geq 0} \sum_{t=0}^{s} \binom{s}{t} (1-q)^{t-1} q^{kt} (y-1)^t x^{s+kt}. \tag{152}$$

So to get $[x^n]\Pi_{\geq k}(x,y)$ we just need to solve $s + kt - 1 = n$ and $s + kt = n$ for nonnegative indices s and t. If $k > 1$, in both cases the maximum of t happens when $s = 0$, and corresponds to $\lfloor (n+1)/k \rfloor$ and $\lfloor n/k \rfloor$, respectively. Thus, from (**??**) we can write

$$[x^n]\Pi_{\geq k}(x,y) = \sum_{t=0}^{\min\left(\lfloor \frac{n+1}{k} \rfloor, n+1\right)} \left(\binom{n-kt+1}{t} - q\binom{n-kt}{t} \right) (1-q)^{t-1} q^{kt} (y-1)^t. \tag{153}$$

Observe that the summation only goes up to $n+1$ when $k = 0$, because in this case of the first binomial coefficient becomes zero for higher values of t.

3.3.3 Explicit Expression

We now produce an explicit expression for $\pi_{\geq k}(n,m)$ by extracting the coefficient of y^m from (**??**). By applying the binomial theorem to $(y-1)^t$ we immediately see that $\pi_{\geq k}(n,m) = [x^n y^m]\Pi_{\geq k}(x,y)$ is

$$\pi_{\geq k}(n,m) = \sum_{t=0}^{\min\left(\lfloor \frac{n+1}{k} \rfloor, n+1\right)} (-1)^{t-m} \binom{t}{m} \left(\binom{n-kt+1}{t} - q\binom{n-kt}{t} \right) (1-q)^{t-1} q^{kt}. \tag{154}$$

which is very close to the simplest explicit expression for $\pi_{\geq k}(n,m)$ given by Muselli [**?**, Thm. 3] —in his notation, $P(M_n^{(k)} = m) = \pi_{\geq k}(n,m)$.

3.3.4 Moments

The pgf (**??**) can also be used to obtain moments of the rv modelling the number of ($\geq k$)-runs in an n-string drawn at random, which we denote by $M_{\geq k,n}$. With standard probabilistic notation,

$$\Pr\left(M_{\geq k,n} = m\right) = \pi_{\geq k}(n,m). \tag{155}$$

The expectation of this random variable is $\mathrm{E}(M_{\geq k,n}) = (d/dy)[x^n]\Pi_{\geq k}(x,y)|_{y=1}$. After differentiating (**??**) with respect to y, setting $y = 1$ lets us see that the only nonzero term corresponds to $t = 1$, as long as $k \leq n$. Therefore,

$$\mathrm{E}(M_{\geq k,n}) = \big((n-k)(1-q) + 1\big)\, q^k \, [\![k \leq n]\!]. \tag{156}$$

The second factorial moment is $\mathrm{E}(M_{\geq k,n}(M_{\geq k,n}-1)) = (d^2/dy^2)[x^n]\Pi_{\geq k}(x,y)|_{y=1}$. After differentiating (??) twice with respect to y and then setting $y=1$, we see that the only nonzero term corresponds to $t=2$. Thus, the second factorial moment is zero if $k>0$ and $\lfloor (n+1)/k \rfloor = 1$. Therefore we have that

$$\mathrm{E}\big(M_{\geq k,n}(M_{\geq k,n}-1)\big) = \big((n-2k)(1-q)+1+q\big)(n-2k)(1-q)q^{2k}\left(\llbracket k>0\rrbracket\llbracket k<n\rrbracket+\llbracket k=0\rrbracket\right). \tag{157}$$

The variance of $M_{\geq k,n}$ follows from the two moments above —see Section ??. The expectation (??) was apparently first given by Goldstein [?, Eq. (5)], without using a pgf. Aki and Hirano also obtained (??) and its corresponding variance [?, pp. 317–318] through a pgf similar to (??) [?, Eq. (5)].

Remark. Krishnan Nair [?, p. 84] gave a pgf which should be equivalent to (??), but which appears to be incorrect. Also, Balakrishnan and Koutras gave a recurrence for $\pi_{\geq k}(n,m)$ very similar to (??) [?, p. 144] but with different initialisation. These authors also gave a bivariate pgf for this case [?, Eq. (5.11)], which, however, they did not get from their aforementioned recurrence. In our notation, the pgf given by Balakrishnan and Koutras is

$$\Pi_{\geq k}(x,y) = \frac{1+(y-1)(q\,x)^k}{1-x-(y-1)(1-q)q^k x^{k+1}}. \tag{158}$$

It is easier to obtain $[y^m]\Pi_{\geq k}(x,y)$ in closed-form from (??) than from (??) —(??) is witness to this. A consequence is that is simpler obtaining (??) using (??) than using (??). □

Remark. The results in this section also allow us to recover the oldest historical result on success runs by de Moivre [?, Prob. LXXXVIII, pp. 243–248]: "*To find the Probability of throwing a Chance assigned a given number of times without intermission, in any given number of Trials*," or, in terms of our definitions, to find $\pi^{\geq 1}_{\geq k}(n) = 1-\pi_{\geq k}(n,0)$. Using the same numerical values originally employed by de Moivre and (??), for $n=10$, $k=3$ and $q=1/2$, we may see that $\pi^{\geq 1}_{\geq 3}(10) = 0.5078 = 65/128$, which we can also verify using $w^{\geq 1}_{\geq 3}(10)/2^{10}$ —see (??)—, whereas if $q=2/3$ then $\pi^{\geq 1}_{\geq 3}(10) = 0.8121 = 592/729$. The truncated decimal values are computed using (??), whereas the fractions can be obtained, for instance, by determining $[x^{10}]\pi^{\geq 1}_{\geq 3}(x)$ using (??) and a computer algebra system. Both probabilities are correctly given in [?], even though there is a known mistake in the procedure followed therein [?, p. 418]. De Moivre did not give a recurrence nor did he provide a proof for his solution, which was based on adding a given number of terms in the power series expansion of a generating function. According to Hald's account [?, p. 420], the first solution to de Moivre's problem based on a finite difference equation (i.e., a recurrence) was given by Simpson. Simpson also gave an explicit solution as an infinite series. However, as also indicated by Hald, Laplace was the first author who essentially derived pgf (??) for $m=1$ [?, p. 252] —see Remark ?? below.

Finally, Uspensky gave a semi-explicit closed-form expression for the probability $\pi_{\geq k}(n,0) = 1 - \pi^{\geq 1}_{\geq k}(n)$ [**?**, Eq. (3)].

$\square$

Remark. We can also use $\pi_{\geq k}(n,0)$ to obtain the probability $\widehat{\pi}_k(n,i)$ that we have to "wait" i indices in order to observe the first $k \geq 1$ consecutive ones in a randomly drawn n-string. Reasoning like in Remark **??**, this probability is

$$\widehat{\pi}_k(n,i) = \pi_{\geq k}\big(i-(k+1),0\big)\,(1-q)q^k\,[\![i \leq n]\!]. \tag{159}$$

When $q = 1/2$, (**??**) follows directly from (**??**) by using $\widehat{\pi}_k(n,i) = \widehat{w}_k(n,i)/2^n$.

Laplace studied the case $\widehat{\pi}_k(n,n)$ in [**?**, Liv. II, Ch. II, N°12], which we denote in the following by $\widehat{\pi}_k(n)$. From (**??**) and (**??**) we have that the pgf $\widehat{\Pi}_k(x) = \sum_n \widehat{\pi}_k(n)\,x^n$ is

$$\widehat{\Pi}_k(x) = (1-q)q^k x^{k+1}[y^0]\Pi_{\geq k}(x,y) \tag{160}$$

$$= \frac{(1-q\,x)q^k x^k}{1-x+(1-q)q^k x^{k+1}}, \tag{161}$$

which was given by Laplace [**?**, p. 252], and which, from (**??**) and (**??**), can also be put as

$$\widehat{\Pi}_k(x) = (1-x)\,\Pi^{\geq 1}_{\geq k}(x).$$

Feller gave (**??**) in [**?**, Ch. XIII, Eq. (7.6)], but he was unaware of Laplace's earlier computation —even though he references other results from [**?**]. This has led a number of authors to believe that (**??**) was originally given by Feller. Observe that we have not used Feller's criterion to produce (**??**), and, in fact, Laplace did not use renewal theory either to derive (**??**) —see again discussion in Remark **??**.

$\square$

3.4 Probability that the Longest Run of an n-String Is a $(\leq k)$-Run

We denote the probability that the longest run of an n-string is a $(\leq k)$-run by $\overline{\pi}_{\leq k}(n)$. Of course, this is a special case of the results in the previous section because

$$\overline{\pi}_{\leq k}(n) = \pi_{\geq (k+1)}(n,0). \tag{162}$$

We treat this special case separately because it was previously dealt with by Muselli [**?**] and Schilling [**?**].

3.4.1 Recurrences

Using (**??**) in recurrences (**??**) and (**??**) we have, respectively, the following two recurrences:

$$\overline{\pi}_{\leq k}(n) = (1-q)\sum_{i=0}^{k} q^i\, \overline{\pi}_{\leq k}\big(n-(i+1)\big), \tag{163}$$

and

$$\overline{\pi}_{\leq k}(n) = \overline{\pi}_{\leq k}(n-1) - q^{k+1}(1-q)\,\overline{\pi}_{\leq k}\big(n-(k+2)\big). \tag{164}$$

From (**??**), (**??**) and (**??**), they are both initialised using $\overline{\pi}_{\leq k}(-1) = 1/(1-q)$ and $\overline{\pi}_{\leq k}(0) = 1$.

Remark. Although (**??**) is the direct probabilistic counterpart of (**??**), Schilling did not provide this recurrence. This author proposed instead to compute $\overline{\pi}_{\leq k}(n)$ as follows [**?**, Eq. (3)]:

$$\overline{\pi}_{\leq k}(n) = \sum_{r=0}^{n} \overline{w}_{\leq k}(n,r)\, q^r (1-q)^{n-r}, \tag{165}$$

where $\overline{w}_{\leq k}(n,r)$ —studied later in Section **??**— is the number of n-strings of Hamming weight r whose longest run is a $(\leq k)$-run, which Schilling showed how to obtain recursively using (**??**) [**?**, Eq. (4)]. □

3.4.2 Probability Generating Function

From (**??**) and (**??**), the pgf $\overline{\Pi}_{\leq k}(x) = \sum_n \overline{\pi}_{\leq k}(n)\, x^n$, is

$$\overline{\Pi}_{\leq k}(x) = \frac{1-q\,x}{(1-q)\,x\,\big(1-x+(1-q)q^{k+1}x^{k+2}\big)}. \tag{166}$$

3.4.3 Explicit Expressions

As $\overline{w}_{\leq k}(n,r)$ has the single-summation closed form (**??**), Schilling would probably be pleased to see that (**??**) is actually a double-summation explicit expression for $\overline{\pi}_{\leq k}(n)$ —although his asymptotic results are a more powerful proposition indeed. In any case, from (**??**) and (**??**) we get a single-summation explicit expression for $\overline{\pi}_{\leq k}(n)$, similar to the one given by Muselli [**?**, Eq. (16)]. Equivalent formulas were previously found in reliability problems by Hwang [**?**, Thm. 3] and by Lambiris and Papastavridis [**?**, Eq. (1)].

3.5 Probability that an n-String Contains Exactly m Nonnull p-Parity Runs

Let $\pi_{[p]}(n,m)$ be the probability that an n-string contains exactly m nonnull p-parity runs. Like in Section **??**, if $q = 1/2$ then we simply have that $\pi_{[p]}(n,m) = w_{[p]}(n,m)/2^n$. We examine the case for arbitrary q next.

3.5.1 Recurrence

We may write a probability recurrence for $\pi_{[p]}(n,m)$ by using the enumerative recurrence (**??**) in conjunction with the law of total probabilities:

$$\pi_{[p]}(n,m) = (1-q)\left(\pi_{[p]}(n-1,m) + \sum_{i=1}^{n} q^i \pi_{[p]}\big(n-(i+1), m - [\![p = \operatorname{mod}(i,2)]\!]\big)\right). \tag{167}$$

As usual, we find initial values for recurrence (**??**) by using the case $n = 1$, which we know by inspection to be

$$\pi_{[p]}(1,m) = \big(q\,[\![p=0]\!] + 1 - q\big)[\![m=0]\!] + q\,[\![p=1]\!][\![m=1]\!]. \tag{168}$$

On the other hand, setting $n = 1$ in (**??**) yields

$$\pi_{[p]}(1,m) = (1-q)\left(\pi_{[p]}(0,m) + q\,\pi_{[p]}(-1, m - [\![p=1]\!])\right). \tag{169}$$

We wish (**??**) to equal (**??**).

We can achieve this equality by using

$$\pi_{[p]}(-1,m) = \frac{1}{1-q}[\![m=0]\!],$$
$$\pi_{[p]}(0,m) = [\![m=0]\!],$$

which are therefore the initial values of (**??**).

From (**??**) we also have have $\pi_{[p]}^{\geq m}(n) = \sum_{t=m}^{\lfloor (n+1)/(2+[\![p=0]\!])\rfloor} \pi_{[p]}(n,t)$. Applying this expression to recurrence (**??**) we get

$$\pi_{[p]}^{\geq m}(n) = (1-q)\Big(\pi_{[p]}^{\geq m}(n-1) + \sum_{i=1}^{n} q^i \big(\pi_{[p]}^{\geq m}(n-(i+1)) + [\![p = \operatorname{mod}(i,2)]\!]\,\pi_{[p]}(n-(i+1), m-1)\big)\Big). \tag{170}$$

3.5.2 Probability Generating Functions

In order to obtain the pgf $\Pi_{[p]}(x,y) = \sum_{n,m} \pi_{[p]}(n,m)\, x^n y^m$ we first make recurrence (**??**) valid for all n and m by adding $[\![n=-1]\!][\![m=0]\!]/(1-q)$ to (**??**). We then get a recurrence free from n-dependent summations by obtaining $\pi_{[p]}(n,m) - q^2\, \pi_{[p]}(n-2,m)$ using the extended recurrence, which yields

$$\begin{aligned}\pi_{[p]}(n,m) = q^2\, \pi_{[p]}(n-2,m) + (1-q)\Big(&\pi_{[p]}(n-1,m) - q^2\, \pi_{[p]}(n-3,m)\\ &+ q\, \pi_{[p]}(n-2, m-[\![p=1]\!]) + q^2\, \pi_{[p]}(n-3, m-[\![p=0]\!])\Big)\\ &+ \big([\![n=-1]\!] - q^2[\![n=1]\!]\big)[\![m=0]\!]\frac{1}{1-q}.\end{aligned} \tag{171}$$

We can now directly obtain $\Pi_{[p]}(x,y)$ from (**??**) by the usual method of multiplying across by $x^n y^m$ and adding over the range of n and m:

$$\Pi_{[p]}(x,y) = \frac{1-q^2x^2}{(1-q)\,x\,\big(1-q^2\,x^2-(1-q)\big(x+qx^2y^{[\![p=1]\!]}-q^2x^3(1-y^{[\![p=0]\!]})\big)\big)}. \tag{172}$$

To conclude, we extract the coefficient of y^m from (**??**) using the same approach as in Section **??**, to get

$$[y^m]\Pi_{[p]}(x,y) = \frac{(1-q^2x^2)\big((1-q)q^{1+[\![p=0]\!]}x^{2+[\![p=0]\!]}\big)^m}{(1-q)\,x\,\big(1-(1-q)x-q^{1+[\![p=1]\!]}\,x^2+[\![p=0]\!](1-q)q^2x^3\big)^{m+1}}, \tag{173}$$

which is the pgf giving the probability that a binary string contains exactly m nonnull p-parity runs.

The related pgf $\Pi_{[p]}^{\geq m}(x) = \sum_n \pi_{[p]}^{\geq m}(n)\, x^n$ can be obtained from (**??**) in a similar way as (**??**) in the previous section. We first obtain a recurrence free from n-dependent summations by subtracting $q^2\pi_{[p]}^{\geq m}(n-2)$ from $\pi_{[p]}^{\geq m}(n)$ using (**??**), which yields the following recurrence:

$$\begin{aligned}\pi_{[p]}^{\geq m}(n) = q^2\pi_{[p]}^{\geq m}(n-2) + (1-q)\Big(&\pi_{[p]}^{\geq m}(n-1) + q\,\pi_{[p]}^{\geq m}(n-2)\\ &+ [\![p=1]\!]q\, w_{[p]}(n-2,m-1) + [\![p=0]\!]q^2\, w_{[p]}(n-3,m-1)\Big).\end{aligned} \tag{174}$$

As usual, by multiplying (**??**) on both sides by x^n and then adding over n we get

$$\begin{aligned}\Pi_{[p]}^{\geq m}(x) &= \frac{(1-q)\big([\![p=1]\!]\,q\,x^2 + [\![p=0]\!]\,q^2\,x^3\big)}{1-(1-q)x-qx^2}\,[y^{m-1}]\Pi_{[p]}(x,y)\\ &= \frac{qx\big([\![p=1]\!]+[\![p=0]\!]\,qx\big)(1-q^2x^2)\big((1-q)q^{1+[\![p=0]\!]}x^{2+[\![p=0]\!]}\big)^{m-1}}{\big(1-(1-q)x-qx^2\big)\big(1-(1-q)x-q^{1+[\![p=1]\!]}\,x^2+[\![p=0]\!](1-q)q^2x^3\big)^m},\end{aligned} \tag{175}$$

where we have used (**??**). This pgf gives the probability that a binary string contains at least m nonnull p-parity runs.

3.5.3 Explicit Expressions

We next determine explicit expressions for $\pi_{[p]}(n,m)$ by extracting the coefficient of x^n in (**??**). We begin with the case $p=1$, where we first express (**??**) as $[y^m]\Pi_{[1]}(x,y) = (1-q)^{m-1}q^m x^{2m-1}/((1-q^2x^2)^m(1-(1-q)x/(1-q^2x^2))^{m+1})$. Applying the negative binomial theorem twice to this expression we get

$$[y^m]\Pi_{[1]}(x,y) = \sum_{t\geq 0}\sum_{s\geq 0}\binom{t+m}{m}\binom{s+t+m-1}{s}(1-q)^{t+m-1}q^{2s+m}x^{2m-1+2s+t}. \tag{176}$$

The coefficient of x^n is found through the nonnegative indices t and s that satisfy $2m-1+2s+t=n$. The maximum of s happens when $t=0$, and thus $s \leq \lfloor (n-2m+1)/2 \rfloor$, whereas t is determined by the previous equation for any given value of s. From these considerations and (**??**) we have that

$$\pi_{[1]}(n,m) = \sum_{s=0}^{\lfloor \frac{n-2m+1}{2} \rfloor}\binom{n-m+1-2s}{m}\binom{n-m-s}{s}(1-q)^{n-m-2s}q^{2s+m}. \tag{177}$$

We now deal with the case $p=0$, which is slightly more involved as it involves a double summation rather than a single one —cf. Section **??**. In this case (**??**) can be expanded using the negative binomial theorem and the binomial theorem (twice) as follows:

$$[y^m]\Pi_{[0]}(x,y) = \sum_{t\geq 0}\sum_{s\geq 0}\sum_{r\geq 0}\binom{t+m}{m}\binom{t}{s}\binom{t-s+1}{r}(1-q)^{t-s+m-1}q^{2m+s+2r}(-1)^r x^{3m-1+t+s+2r}. \tag{178}$$

To determine the coefficient of x^n we just determine the nonnegative indices t, s and r that fulfil $n = 3m-1+t+s+2r$. From $t=s=0$ we have $s \leq \lfloor (n-3m+1)/2 \rfloor$, and from $s=0$ we have that $t \leq n-3m+1-2r$. Finally $s = n-3m+1-2r-t$. Thus we have from (**??**) that

$$\pi_{[0]}(n,m) = \sum_{r=0}^{\lfloor n-3m+1 \rfloor 2}\sum_{t=0}^{n-3m+1-2r}\binom{t+m}{m}\binom{t}{s}\binom{t-s+1}{r}(1-q)^{t-s+m-1}q^{2m+s+2r}(-1)^r, \tag{179}$$

where s depends on r and t as just indicated.

3.5.4 Moments

We denote by $M_{[p],n}$ the random variable that models the number of p-parity runs in an n-string drawn at random, i.e.,

$$\Pr(M_{[p],n} = m) = \pi_{[p]}(n,m). \tag{180}$$

Using (**??**), we may obtain the moments of $M_{[p],n}$. First, the expectation is $\mathrm{E}(M_{[p],n}) = [x^n](\partial/\partial y)\Pi_{[p]}(x,y)|_{y=1}$. Since we have that

$$\left.\frac{\partial \Pi_{[p]}(x,y)}{\partial y}\right|_{y=1} = \frac{qx(1-qx)\big(qx[\![p=0]\!] + [\![p=1]\!]\big)}{(1-x)^2(1+qx)}, \tag{181}$$

by applying the negative binomial theorem twice we get

$$\left.\frac{\partial \Pi_{[p]}(x,y)}{\partial y}\right|_{y=1} = qx(1-qx)\big(qx[\![p=0]\!] + [\![p=1]\!]\big)\sum_{t\geq 0}\sum_{r\geq 0}\binom{t+1}{t}(-1)^r q^{r+1}x^{t+r}. \tag{182}$$

Focussing on the $p=1$ case, the coefficient of x^n is found using the nonnegative indices t that solve $n = t+r+1$ and $t+r+2$. The first equation implies that $t \leq n-1$, whereas the second one implies $t \leq n-2$. We thus have from (**??**) that

$$\begin{aligned}\mathrm{E}(M_{[1],n}) &= \sum_{t=0}^{n-1}(t+1))(-1)^{n-1-t}q^{n-t} - q\sum_{t=0}^{n-2}(t+1))(-1)^{n-2-t}q^{n-t-1}\\ &= \frac{2q^2\big(1-n(1+q)-(-1)^n q^n\big)}{(1+q)^2} + nq.\end{aligned} \tag{183}$$

For large n, $\mathrm{E}(M_{[1],n}) \approx (1-2q/(1+q))\,nq$. The case $p=0$ can be determined using (**??**) and (**??**):

$$\mathrm{E}(M_{[0],n}) = \mathrm{E}(M_{\geq 1,n}) - \mathrm{E}(M_{[1],n}). \tag{184}$$

The second factorial moment may be computed in a similar fashion.

3.6 Discussion

It should be clear from our exposition in this section that it is relatively straightforward to obtain probabilistic counterparts of every aspect of the enumeration problems that we have studied in Section **??**, and that this procedure is generally simpler and more versatile than addressing success runs from a combinatorial analysis viewpoint. In particular, pgfs are easy to obtain through recurrences, and they are generally preferable to complicated explicit expressions when it comes to obtaining moments. If desired, explicit expressions can be also obtained from pgfs, and, importantly, they are also a conduit to asymptotic probabilistic results.

Last but not least, we see later in Section **??** that all the probabilistic results in this section can also be approximated through deterministic enumerations via the law of large numbers.

4 Number of n-Strings of Hamming Weight r that Contain Prescribed Quantities of Runs Under Different Constraints

In this section we address the same basic problems as in Section **??** but when the n-strings are restricted to having Hamming weight r —i.e., to containing exactly r ones. In the terminology of Goulden and Jackson [**?**, Sec. 2.4.7] these are enumerations *with type restriction*. Many early works in the theory of runs assume this constraint, particularly those that use runs for hypothesis tests. The results in this section are complemented by other results given later in Section **??**, where runs of ones and zeros are jointly considered.

We cover the counterparts of the problems in Sections **??** and **??** in Sections **??** and **??**, respectively. The former enumeration also solves the same problems as in Sections **??**–**??** when the fixed Hamming weight constraint is observed. We actually discuss two of these special cases in more detail in Sections **??** and **??**, since they have been studied by a considerable number of previous authors [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**].

Whereas the enumerative recurrences in Sections **??** and **??** are bivariate, solving fixed Hamming weight versions of these enumerations involves establishing trivariate recurrence relations. This means that the corresponding ogfs are also trivariate, which in some scenarios may hinder the determination of compact explicit expressions. Nevertheless, this is possible in a number of special cases.

4.1 Number of n-Strings of Hamming Weight r that Contain Exactly m $(\underline{k} \leq \bar{k})$-Runs

Let $w_{\underline{k}\leq\bar{k}}(n,m,r)$ represent the number of n-strings of Hamming weight r that contain exactly m $(\underline{k} \leq \bar{k})$-runs. As usual, the n-strings that we enumerate may have other runs, as long as they are longer than $\bar{k}$ or shorter than $\underline{k}$. Observe that this is a conditional version of the general enumeration in Section **??** in which we only consider the $\binom{n}{r}$ n-strings that have Hamming weight r.

Next, we need to tighten necessary condition (**??**) in order to take the Hamming weight constraint into account.

Necessary Condition. (Existence of n-strings of Hamming weight r containing m $(\underline{k} \leq \bar{k})$-runs)

$$w_{\underline{k}\leq\bar{k}}(n,m,r) > 0 \quad \Longrightarrow \quad 0 \leq m \leq \min\left(\left\lfloor \frac{n+1}{\underline{k}+1} \right\rfloor, \left\lfloor \frac{r}{\underline{k}} \right\rfloor\right). \tag{185}$$

Observe that if $r > 0$ and $\underline{k} = 0$ then the upper bound on m is $n+1$. When $r = 0$ and $\underline{k} = 0$ the ratio $r/\underline{k}$ is undefined, but the upper bound is still $n+1$, as this is the number of null runs in the all-zeros n-string. $\square$

4.1.1 Recurrences

We can resort to a nearly identical reasoning as in Section **??** to get a recurrence for this enumeration: the quantity $w_{\underline{k}\le\bar{k}}(n,m,r)$ can be broken down into different contributions corresponding to the ensemble of n-strings that start with an i-run. If $i > \bar{k}$ or $i < \underline{k}$ then the contribution is $w_{\underline{k}\le\bar{k}}(n-(i+1),m,r-i)$, but if $\underline{k} \le i \le \bar{k}$ then the contribution is $w_{\underline{k}\le\bar{k}}(n-(i+1),m-1,r-i)$. Thus, considering all possible lengths of the starting run we get the following trivariate recurrence:

$$\begin{aligned} w_{\underline{k}\le\bar{k}}(n,m,r) = &\sum_{i=\underline{k}}^{\bar{k}} w_{\underline{k}\le\bar{k}}\big(n-(i+1),m-1,r-i\big) + \sum_{i=0}^{\underline{k}-1} w_{\underline{k}\le\bar{k}}\big(n-(i+1),m,r-i\big) \\ &+ \sum_{i=\bar{k}+1}^{n} w_{\underline{k}\le\bar{k}}\big(n-(i+1),m,r-i\big). \end{aligned} \tag{186}$$

Observe that the only difference with respect to recurrence (**??**) are the updates of the Hamming weights in the recursive invocations on the right hand side. Following the same steps as in Section **??**, we can argue an alternative recurrence equivalent to (**??**):

$$\begin{aligned} w_{\underline{k}\le\bar{k}}(n,m,r) = &\; w_{\underline{k}\le\bar{k}}(n-1,m,r) + w_{\underline{k}\le\bar{k}}(n-1,m,r-1) \\ &+ w_{\underline{k}\le\bar{k}}\big(n-(\underline{k}+1),m-1,r-\underline{k}\big) - w_{\underline{k}\le\bar{k}}\big(n-(\underline{k}+1),m,r-\underline{k}\big) \\ &- w_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+2),m-1,r-(\bar{k}+1)\big) + w_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+2),m,r-(\bar{k}+1)\big), \end{aligned} \tag{187}$$

which parallels recurrence (**??**), and which can also be obtained by calculating $w_{\underline{k}\le\bar{k}}(n,m,r) - w_{\underline{k}\le\bar{k}}(n-1,m,r-1)$ using (**??**).

In the following we work with the simpler recurrence (**??**), as it is free from n-dependent summations and thus directly amenable to determining the ogf. In order to find initialisation values we rely on the case $n=1$, in which we know by inspection that

$$\begin{aligned} w_{\underline{k}\le\bar{k}}(1,m,r) = &\; [\![r=0]\!]\big([\![m=0]\!][\![\underline{k}\ge 1]\!] + [\![m=2]\!][\![\underline{k}=0]\!]\big) \\ &+ [\![r=1]\!]\Big([\![m=0]\!]\big([\![\underline{k}>1]\!] + [\![\underline{k}=0]\!][\![\bar{k}=0]\!]\big) + [\![m=1]\!][\![\underline{k}=1]\!]\Big), \end{aligned} \tag{188}$$

Setting next $n=1$ in (**??**) yields

$$w_{\underline{k}\le\bar{k}}(1,m,r) = w_{\underline{k}\le\bar{k}}(0,m,r-1) + w_{\underline{k}\le\bar{k}}(0,m,r)$$

$$+ w_{\underline{k}\le\bar{k}}(-\underline{k}, m-1, r-\underline{k}) - w_{\underline{k}\le\bar{k}}(-\underline{k}, m, r-\underline{k})$$
$$- w_{\underline{k}\le\bar{k}}\big(-(\bar{k}+1), m-1, r-(\bar{k}+1)\big) + w_{\underline{k}\le\bar{k}}\big(-(\bar{k}+1), m, r-(\bar{k}+1)\big), \tag{189}$$

Taking (??) into account, we may verify that (??) equals (??) for all m, r and $0 \le \underline{k} \le \bar{k}$ if we choose the values

$$w_{\underline{k}\le\bar{k}}(-1, m, r) = [\![m = 0]\!][\![r = 0]\!], \tag{190}$$
$$w_{\underline{k}\le\bar{k}}(0, m, r) = \big[\!\big[m = [\![\underline{k} = 0]\!]\big]\!\big][\![r = 0]\!], \tag{191}$$

which therefore constitute the initialisation of (??). The reader may check that these values also initialise the equivalent recurrence (??).

4.1.2 Generating Function

We obtain next the trivariate ogf $W_{\underline{k}\le\bar{k}}(x, y, z) = \sum_{n,m,r} w_{\underline{k}\le\bar{k}}(n, m, r)\, x^n y^m z^r$ using recurrence (??). We first need to make this recurrence valid for all values of n, m and r, which, like in Section ??, simply involves extending (??) so that it works for $n = -1$ and $n = 0$. Through the same procedure as in that section, the reader may verify that we now have to add $[\![n = -1]\!][\![m = 0]\!][\![r = 0]\!] - [\![n = 0]\!][\![m = 0]\!][\![r = 1]\!]$ to (??). We then multiply both sides of the extended recurrence by $x^n y^m z^r$ and sum on n, m, and r, from which we readily get

$$W_{\underline{k}\le\bar{k}}(x, y, z) = \frac{1 - xz}{x\Big(1 - x(z+1) + (1-y)\big(x^{\underline{k}+1} z^{\underline{k}} - x^{\bar{k}+2} z^{\bar{k}+1}\big)\Big)}. \tag{192}$$

Like in Section ??, if we write this ogf as a function of $(1 - cy)^{-1}$

we can see that the coefficient of y^m in (??) is

$$[y^m] W_{\underline{k}\le\bar{k}}(x, y, z) = \frac{(1 - xz)\Big(x^{\underline{k}+1} z^{\underline{k}} - x^{\bar{k}+2} z^{\bar{k}+1}\Big)^m}{x\Big(1 - x(z+1) + x^{\underline{k}+1} z^{\underline{k}} - x^{\bar{k}+2} z^{\bar{k}+1}\Big)^{m+1}}. \tag{193}$$

4.2 Number of n-Strings of Hamming Weight r that Contain Exactly m Nonnull Runs

We denote by $w(n, m, r)$ the number of n-strings of Hamming weight r that contain exactly m nonnull runs. This enumeration —in fact, the probability of drawing m nonnull runs when n-strings of Hamming weight r are drawn uniformly at random, which we denote by $\pi(n, m, r) = w(n, m, r)/\binom{n}{r}$— was explicitly solved by a number of authors [?, ?, ?, ?].

This is a particular case of the analysis in Section **??** with $\underline{k} = 1$ and $\overline{k} = n$, and therefore

$$w(n, m, r) = w_{1 \le n}(n, m, r). \tag{194}$$

4.2.1 Recurrences

Specialising recurrences (**??**) and (**??**) to this case yields the following two recursive relations:

$$w(n, m, r) = w(n-1, m, r) + \sum_{i=1}^{n} w(n - (i+1), m-1, r-i), \tag{195}$$

and

$$\begin{aligned} w(n, m, r) = {} & w(n-1, m, r-1) + w(n-1, m, r) \\ & + w(n-2, m-1, r-1) - w(n-2, m, r-1), \end{aligned} \tag{196}$$

both of which from (**??**) and (**??**) are initialised by

$$w(-1, m, r) = w(0, m, r) = [\![m = 0]\!][\![r = 0]\!]. \tag{197}$$

4.2.2 Generating Function

We obtain next the ogf $W(x, y, z) = \sum_{n,m,r} w(n, m, r)\, x^n y^m z^r$. Although we may indeed specialise (**??**) to this case with $\underline{k} = 1$ and $\overline{k} = n$, this gives an n-dependent ogf. A simpler, more general ogf is obtained by proceeding from recurrence (**??**) just like in we did to get (**??**) from (**??**). This yields

$$W(x, y, z) = \frac{1 - xz}{x(1 - x - xz + x^2 z - x^2 yz)}, \tag{198}$$

from which we have

$$[y^m]W(x, y, z) = \frac{x^{2m-1} z^m}{(1 - xz)^m (1 - x)^{m+1}}. \tag{199}$$

4.2.3 Explicit Expression

Next, we get an explicit expression for $w(n, m, r)$. Applying the negative binomial theorem twice, we can express (**??**) as

$$[y^m]W(x, y, z) = \sum_{s \ge 0} \sum_{t \ge 0} \binom{s + m - 1}{m - 1} \binom{t + m}{m} x^{s+t+2m-1} z^{s+m}. \tag{200}$$

Thus the coefficient of $x^n z^r$ corresponds to the indices that fulfil $n = s + t + 2m - 1$ and $r = s + m$, i.e., $s = r - m$ and $t = n - s - 2m + 1 = n - r - m + 1$. Therefore, from (**??**) we have that $w(n, m, r) = [x^n y^m z^r]W(x, y, z)$ is

$$w(n, m, r) = \binom{r-1}{m-1}\binom{n-r+1}{m}. \tag{201}$$

Either in this enumerative form or in its probabilistic version $\pi(n, m, r) = w(n, m, r)/\binom{n}{r}$, expression (**??**) was previously given by Stevens [**?**, Eq. (3.05)], Mood [**?**, Eq. (2.11)], Charalambides [**?**, p. 97], Gibbons and Chakraborti [**?**, Cor. 3.2.1] Schuster [**?**, Cor. 3.5] and others. Schuster explicitly notes that $\pi(n, m, r)$ is a hypergeometric distribution.

4.3 Number of n-Strings of Hamming Weight r Whose Longest Run Is a k-Run or a $(\leq k)$-Run

We consider here the enumeration of the n-strings of Hamming weight r whose longest run is a k-run or a $(\leq k)$-run, which we respectively denote by $\overline{w}_k(n, r)$ and $\overline{w}_{\leq k}(n, r)$. These are the Hamming weight constrained versions of the enumerations in Section **??**.

Many authors have studied these problems, both from enumerative and probabilistic viewpoints. In the latter case we refer to the situation in which the n-strings of Hamming weight r are drawn uniformly at random, where we are interested in the probabilities $\overline{\pi}_k(n, r) = \overline{w}_k(n, r)/\binom{n}{r}$ and $\overline{\pi}_{\leq k}(n, r) = \overline{w}_{\leq k}(n, r)/\binom{n}{r}$ —i.e., the cumulative probability. The earliest study is by Mosteller, who gave $\overline{\pi}_{\geq k}(n, r) = 1 - \overline{\pi}_{\leq (k-1)}(n, r)$ for the case $n = 2r$ [**?**, Eq. (4)]. Bateman [**?**] first considered an arbitrary ratio r/n in the same problem and gave an explicit formula for $\overline{\pi}_{\geq k}(n, r)$. David and Barton [**?**], Burr and Cane [**?**], Philippou and Makri [**?**], Gibbons and Chakraborti [**?**], and Schuster [**?**] gave exact formulas for $\overline{\pi}_{\leq k}(n, r)$. Recurrences for $\overline{w}_{\leq k}(n, r)$ and $\overline{\pi}_{\leq k}(n, r)$ where given by Schilling [**?**] and Schuster [**?**], respectively. Finally, and Nej and Reddy [**?**] gave a recurrence and a closed-form explicit expression for $\overline{w}_k(n, r)$.

We address this enumeration using the exact same approach as in Section **??**. Like in that section, we first state the connection between $\overline{w}_k(n, r)$ and $\overline{w}_{\leq k}(n, r)$. For all $k \geq 0$ we have that

$$\overline{w}_{\leq k}(n, r) = \sum_{j=0}^{k} \overline{w}_j(n, r), \tag{202}$$

and, conversely, for all $k \geq 1$ it holds that

$$\overline{w}_k(n, r) = \overline{w}_{\leq k}(n, r) - \overline{w}_{\leq (k-1)}(n, r), \tag{203}$$

whereas $\overline{w}_0(n, r) = \overline{w}_{\leq 0}(n, r)$. The expression that allows us to get these two quantities through the general enumeration $w_{\underline{k} \leq \overline{k}}(n, m, r)$ is

$$\overline{w}_{\leq k}(n, r) = w_{\geq (k+1)}(n, 0, r), \tag{204}$$

where $w_{\geq k}(n,m,r) = w_{k\leq n}(n,m,r)$. Thus $\overline{w}_0(n,r) = w_{\geq 1}(n,0,r) = [\![r=0]\!]$ for $n \geq 1$, as the longest run is a null run only in the all-zeros n-string, which can only happen if $r=0$.

4.3.1 Recurrences

We first find recurrences for $\overline{w}_{\leq k}(n,r)$. By applying (??) to (??) we get

$$\overline{w}_{\leq k}(n,r) = \sum_{i=0}^{k} \overline{w}_{\leq k}(n-(i+1), r-i), \tag{205}$$

which parallels (??). Also, by applying (??) to (??) we get the following alternative recurrence:

$$\overline{w}_{\leq k}(n,r) = \overline{w}_{\leq k}(n-1,r) + \overline{w}_{\leq k}(n-1,r-1) - \overline{w}_{\leq k}\big(n-(k+2), r-(k+1)\big), \tag{206}$$

which parallels (??). From (??), (??) and (??), both (??) and (??) are initialised using

$$\overline{w}_{\leq k}(-1,r) = \overline{w}_{\leq k}(0,r) = [\![r=0]\!]. \tag{207}$$

We can also get recurrences for $\overline{w}_k(n,r)$ when $k \geq 1$. By inputting (??) in (??) we find the recurrence

$$\overline{w}_k(n,r) = \sum_{i=0}^{k-1} \overline{w}_k(n-(i+1), r-i) + \sum_{j=0}^{k} \overline{w}_j(n-(k+1), r-k), \tag{208}$$

whereas if we input (??) in (??) we get the alternative recurrence

$$\overline{w}_k(n,r) = \overline{w}_k(n-1,r) + \overline{w}_k(n-1,r-1) + \sum_{j=0}^{k-1} \overline{w}_j(n-(k+1), r-k) - \sum_{j=0}^{k} \overline{w}_j(n-(k+2), r-(k+1)). \tag{209}$$

From (??) and (??), the initialisation of both recurrences is $\overline{w}_k(-1,r) = \overline{w}_k(0,r) = 0$ for $k \geq 1$, but we can see that we also need initialisation for the case $k=0$. Since $\overline{w}_0(n,r) = w_{\geq 1}(n,0,r)$, from (??) and (??) we have that $\overline{w}_0(-1,r) = \overline{w}_0(0,r) = [\![r=0]\!]$. Thus recurrences (??) and (??) are both initialised by

$$\overline{w}_k(-1,r) = \overline{w}_k(0,r) = [\![k=0]\!][\![r=0]\!].$$

Again, (??) and (??) parallel their unconstrained counterparts (??) and (??), respectively.

Remark. Recurrence (??) was given by Schilling in [?, Eq. (4)]. This author examines in some detail the case $k=3$ of (??), but is nevertheless somewhat vague about its general initialisation. Schuster also gave a recurrence for $\overline{\pi}_{\leq k}(n,r) = \overline{w}_{\leq k}(n,r)/\binom{n}{r}$ based on Pascal's triangle of order $k+1$ [?, Cor. 4.3], which is essentially different from either (??) or (??). In any case, both recurrences above are simpler. Also, recurrence (**??**) was previously given by Nej and Reddy [**?**, Thm. 3.2] —in their notation, $\overline{w}_k(n,r) = F_n(r,k)$. These authors state that their recurrence is valid for $n \geq 3$, $1 \leq r \leq n-2$ and $\lfloor n/(r+1) \rfloor \leq k \leq n$.

Regarding the piecemeal validity of the recurrences given by previous authors mentioned in this remark: notice that all the recurrences given here are valid for all $n \geq 1$ and $r, k \geq 0$ with the initialisations given, just taking necessary condition (**??**) into account. As in other such fragmentary recurrences we have met before (see Section **??**), this shows the necessity of considering the cases $n = -1$ and $n = 0$ in the initialisation in order to obtain the simplest and most general recurrences. □

4.3.2 Generating Functions

Considering (**??**), we may obtain the ogf $\overline{W}_{\leq k}(x,z) = \sum_{n,r} \overline{w}_{\leq k}(n,r)\, x^n z^r$ by setting $\underline{k} = k+1$, $\overline{k} = n$ and $m = 0$ in (**??**), but this leads to an n-dependent ogf. A simpler, more general ogf is obtained from recurrence (**??**) as we show next. Adding $[\![n=-1]\!][\![r=0]\!] - [\![n=0]\!][\![r=1]\!]$ to (**??**) to make it valid for all values of the parameters, and then multiplying it on both sides by $x^n z^r$ and summing over n and r we obtain

$$\overline{W}_{\leq k}(x,z) = \frac{1-xz}{x\big(1-x-xz+x^{k+2}z^{k+1}\big)}. \tag{210}$$

It is possible to derive the ogf $\overline{W}_k(x,z) = \sum_{n,r} \overline{w}_k(n,r)\, x^n z^r$ from any of the two recurrences (**??**) or (**??**), but this is not straightforward due to the summations on j leading to recursive relations on the generating function itself. Nevertheless it is a simple matter to get this ogf directly from (**??**) and (**??**):

$$\overline{W}_k(x,z) = \frac{1-xz}{x}\left(\frac{1}{1-x-xz+x^{k+2}z^{k+1}} - \frac{1}{1-x-xz+x^{k+1}z^{k}}\right). \tag{211}$$

4.3.3 Explicit Expressions

We find next an explicit expression for $\overline{w}_{\leq k}(n,r)$, for which we rewrite (**??**) first as $\overline{W}_{\leq k}(x,z) = x^{-1}\big(1-(x-x^{k+2}z^{k+1})/(1-xz)\big)^{-1}$. Applying the negative binomial theorem twice and then the binomial theorem to this expression, we can expand the ogf as

$$\overline{W}_{\leq k}(x,z) = \sum_{t\geq 0}\sum_{s\geq 0}\sum_{p\geq 0} \binom{s+t-1}{t-1}\binom{t}{p}(-1)^p (xz)^{p(k+1)+s} x^{t-1}. \tag{212}$$

To extract the coefficient of $x^n z^r$ we have to find nonnegative indices t, s and p that fulfil

$$n = p(k+1) + s + t - 1,$$

$$r = p(k+1) + s.$$

From the second equation, the maximum of p happens when $s = 0$, and thus $p \le \lfloor r/(k+1) \rfloor$. One value of p determines $s = r - p(k+1)$ and $t = n + 1 - p(k+1) - s$. We thus have from (**??**) that $\overline{w}_{\le k}(n,r) = [x^n z^r]\overline{W}_{\le k}(x,z)$ is

$$\overline{w}_{\le k}(n,r) = \sum_{p=0}^{\lfloor \frac{r}{k+1} \rfloor} (-1)^p \binom{n - p(k+1)}{n-r} \binom{n-r+1}{p}. \tag{213}$$

Bateman was the first author who gave an expression parallel to (**??**) [**?**, p. 101]. Her formula gives the probability

$\overline{\pi}_{\ge k}(n,r)$, and it is based on a generating function for compositions restricted to a maximum part —see start of Section **??** for the connection between runs and compositions. David and Barton [**?**, p. 230] and then Schuster [**?**, Cor. 4.2] gave alternative derivations of (**??**) in the form $\overline{\pi}_{\le k}(n,r)$. The former authors used combinatorial arguments and characteristic (Bernoulli) random variables, whereas the latter author used the theory of exchangeability of random variables and generating functions. Schuster was driven by the computational impracticality of some earlier explicit expressions for $\overline{\pi}_{\le k}(n,r)$ [**?**, **?**], but he must have missed the expression by Bateman that we mentioned above.

From (**??**) we also get a single-summation explicit expression for $\overline{w}_k(n,r)$ through (**??**). Regarding this case, Nej and Reddy found a more compact closed-form expression (without summations) for $\overline{w}_k(n,r)$ through combinatorial analysis, although restricted to the special case $r < n-1$ and $r < 2k$ [**?**, Thm. 3.1]. This suggests that it might be possible to evaluate (**??**) in closed form, perhaps using the "snake oil" method [**?**], although we have not succeeded in doing so.

4.4 Number of n-Strings of Hamming Weight r that Contain Exactly m Nonnull p-Parity Runs

In this section we consider the counterpart of Section **??** under a Hamming weight constraint. We denote by $w_{[p]}(n,m,r)$ the number of n-strings of Hamming weight r that contain exactly m nonnull p-parity runs.

As usual, we need to state first a necessary condition similar to (**??**).

Necessary Condition. (Existence of n-strings of Hamming weight r containing m nonnull p-parity runs)

$$w_{[p]}(n,m,r) > 0 \implies 0 \le m \le \min\left(\left\lfloor \frac{n+1}{2 + [\![p=0]\!]} \right\rfloor, \left\lfloor \frac{r}{1 + [\![p=0]\!]} \right\rfloor \right). \tag{214}$$

□

4.4.1 Recurrence

We obtain a recurrence with a rationale similar to that in Section **??**: n-strings that start with 0 contribute $w_{[p]}(n-1,m,r)$ to $w_{[p]}(n,m,r)$; on the other hand, n-strings that start with a nonnull i-run contribute $w_{[p]}(n-(i+1),m,r-i)$ if $p=\operatorname{mod}(i,2)$ but $w_{[p]}(n-(i+1),m-1,r-i)$ if $p\neq\operatorname{mod}(i,2)$. These two contributions lead to the following trivariate recurrence:

$$w_{[p]}(n,m,r)=w_{[p]}(n-1,m,r)+\sum_{i=1}^{n}w_{[p]}(n-(i+1),m-[\![p=\operatorname{mod}(i,2)]\!],r-i).\quad(215)$$

Again, the only difference with respect to (**??**) are the Hamming weight updates. We initialise (**??**) by first determining its correct value for $n=1$. By inspection, the value is

$$w_{[p]}(1,m,r)=[\![r=0]\!][\![m=0]\!]+[\![r=1]\!]\Big([\![m=0]\!][\![p=0]\!]+[\![m=1]\!][\![p=1]\!]\Big).\quad(216)$$

Setting $n=1$ in (**??**) now yields

$$w_{[p]}(1,m,r)=w_{[p]}(0,m,r)+w_{[p]}(-1,m-[\![p=1]\!],r-1).\quad(217)$$

It is readily verified that (**??**) equals (**??**) for the following initialisation values:

$$w_{[p]}(-1,m,r)=w_{[p]}(0,m,r)=[\![m=0]\!][\![r=0]\!].\quad(218)$$

4.4.2 Generating Function

We obtain next the ogf $W_{[p]}(x,y,z)=\sum_{n,m,r}w_{[p]}(n,m,r)\,x^ny^mz^r$. To this end, we first add $[\![n=-1]\!][\![m=0]\!][\![r=0]\!]$ to (**??**) to make it valid for all n, m and r. The next step is obtaining a recurrence without n-dependent summations, for which we compute $w_{[p]}(n,m,r)-w_{[p]}(n-2,m,r-2)$ using the extended recurrence. Doing so gives

$$\begin{aligned}w_{[p]}(n,m,r)=\;&w_{[p]}(n-2,m,r-2)+w_{[p]}(n-1,m,r)-w_{[p]}(n-3,m,r-2)\\&+w_{[p]}(n-2,m-[\![p=1]\!],r-1)+w_{[p]}(n-3,m-[\![p=0]\!],r-2)\\&+[\![m=0]\!]\big([\![n=-1]\!][\![r=0]\!]-[\![n=1]\!][\![r=2]\!]\big).\end{aligned}\quad(219)$$

We can now determine $W_{[p]}(x,y,z)$ by multiplying both sides of (**??**) by $x^ny^mz^r$ and then adding on $n,m,$ and r. This yields

$$W_{[p]}(x,y,z)=\frac{1-x^2z^2}{x\Big(1-x-x^2z\big(z+y^{[\![p=1]\!]}\big)+x^3z^2\big(1-y^{[\![p=0]\!]}\big)\Big)}.\quad(220)$$

With the usual strategy, we can see that the coefficient of y^m in (**??**) is

$$[y^m]W_{[p]}(x,y,z)=\frac{(1-x^2z^2)(x^{2+[\![p=0]\!]}z^{1+[\![p=0]\!]})^m}{x\Big(1-x-x^2z^2-[\![p=0]\!]\,x^2z\,(1-xz)\Big)^{m+1}}.\quad(221)$$

4.5 Probabilistic Connections

If the probability of drawing a one is q, then an n-string drawn at random will roughly contain nq ones for large n with high probability —by the law of large numbers. Thus, for large n, all the probabilistic results in Section **??** can be approximated through their deterministic counterparts in this section, by assuming that the Hamming weight of an n-string drawn at random is always $\lfloor nq \rfloor$. With this assumption we have that, for large n,

$$\pi_{\underline{k} \leq \bar{k}}(n, m) \approx w_{\underline{k} \leq \bar{k}}(n, m, \lfloor nq \rfloor) \binom{n}{\lfloor nq \rfloor}^{-1}, \tag{222}$$

and

$$\pi_{[p]}(n, m) \approx w_{[p]}(n, m, \lfloor nq \rfloor) \binom{n}{\lfloor nq \rfloor}^{-1}. \tag{223}$$

Of course, these connections may also be used in reverse, that is to say, to approximate Hamming-constrained enumerations using probabilistic expressions.

5 Number of n-Strings that Contain Prescribed Quantities of Nonnull Runs of Ones and/or Zeros Under Different Constraints

In Section **??** we studied enumerations of n-strings containing prescribed numbers of runs of ones in binary strings. In this section we see how analogous techniques allow us to enumerate the n-strings that contain prescribed quantities of runs of ones *and/or zeros* —i.e., for the sake of clarity, the case where each run may indistinctly be a run of ones or a run of zeros— in several similar scenarios. We focus on extending the main enumerations in Sections **??** and **??** to the same scenarios for runs of ones and zeros, but we also examine the extension of the special enumerations in Sections **??**, **??** and **??**, as these problems have been addressed by other authors [**?**, **?**, **?**, **?**]. The main change in our approach in this section with respect to previous sections is that we always establish two mutual recurrences, rather than a single one —as we have always done up to this point. These mutual recurrences correspond, respectively, to enumerations of n-strings that start with a zero or with a one. Importantly, we only consider nonnull runs of ones or zeros.

As mentioned in the introduction, a number of authors have considered runs of ones and runs of zeros jointly [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**].

On first impression, this setting is not so relevant to the study of success runs, but, in contrast, it can lead to better statistics in runs-based hypothesis tests (as more information is taken into account when considering both kinds of runs). Importantly, it also has clear direct implications for the problem of compositions (ordered partitions) of n, which are the different ways in which we can partition the sum $1 + 1 + \cdots + 1 = n$ into nonzero ordered parts. The intimate connection between runs of ones and zeros in n-strings and compositions is easy to understand: we can visualise the parts of a composition of n as the ordered lengths of an alternating series of runs of ones and zeros in an n-string, and in its ones' complement. For example, consider the following composition of $n = 9$ into five parts: $1+3+2+2+1 = 9$. We can represent this composition using the sequence of lengths of five alternating runs of ones and zeros in two related 9-strings: '011100110', and its ones' complement, '100011001' —see also Figure **??**, where the top row can be interpreted as a partition of $1 + 1 + \cdots + 1$ into nonzero parts. Two other consequences of this observation are: 1) all enumerations in this section must be even valued, and 2) the aforementioned mutual recurrences are always symmetric. To the best of our knowledge, the earliest author who saw a connection between compositions and runs was Bateman [**?**].

The relationship between runs *of ones* and compositions is perhaps less transparent. However, the results in Section **??** also hint at a close association between runs of ones and compositions, as shown by the many OEIS sequences cited that are simultaneously related to both. Grimaldi and Heubach [**?**] have in fact explicitly described one case of this relationship. The reason for these connections is the intrinsic link that exists in certain scenarios between the enumerations of $(n-1)$-strings containing prescribed quantities of runs of ones and the enumerations of n-strings containing prescribed quantities of runs of ones and/or zeros —which, as we have discussed in the previous paragraph, are themselves directly connected to compositions of n. We explicitly give the simplest of such links in Theorems **??** and **??** in Sections **??** and **??**, respectively.

Remark **??** at end of this section overviews the main consequences of our results for the problem of compositions.

5.1 Number of n-Strings that Contain Exactly m Nonnull ($\underline{k} \leq \bar{k}$)-Runs of Ones and/or Zeros

We denote by $s_{\underline{k}\leq\bar{k}}(n, m)$ the number of n-strings that contain exactly m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros, which may also contain other runs with lengths not in the prescribed range. If we denote the number of compositions of n that contain exactly m parts between $\underline{k}$ and $\bar{k}$ by $c_{\underline{k}\leq\bar{k}}(n, m)$, then, from the observations in the previous section, we also have that

$$c_{\underline{k}\leq\bar{k}}(n, m) = \frac{1}{2}\, s_{\underline{k}\leq\bar{k}}(n, m). \tag{224}$$

As in Section **??**, we establish first a necessary condition that guides our enumeration.

Necessary Condition. (Existence of n-strings containing m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros)

$$s_{\underline{k}\leq\bar{k}}(n,m) > 0 \quad \Longrightarrow \quad 0 \leq m \leq \left\lfloor \frac{n}{\underline{k}} \right\rfloor. \tag{225}$$

□

The upper bound in (??) is due to the fact that it must always hold that $m\,\underline{k} \leq n$ for an n-string to be potentially able to hold m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros. Thus, unlike in the enumerations of runs of ones, we now necessarily have that the enumerations must be zero when $n = -1$, which implies that only the case $n = 0$ is involved in the initialisations.

5.1.1 Recurrence

We may obtain a recurrence for $s_{\underline{k}\leq\bar{k}}(n,m)$ by defining first two mutual recurrences for the number of n-strings that start with bit b and contain exactly m ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros, which we denote by $s^b_{\underline{k}\leq\bar{k}}(n,m)$ with $b \in \{0,1\}$. Considering all the n-strings that start with an i-run of each kind, we can see through the usual strategy —see Section **??**— that

$$s^b_{\underline{k}\leq\bar{k}}(n,m) = \sum_{i=\underline{k}}^{\bar{k}} s^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-i,m-1) + \sum_{i=1}^{\underline{k}-1} s^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-i,m) + \sum_{i=\bar{k}+1}^{n} s^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-i,m), \tag{226}$$

where $b \in \{0,1\}$ and $\tilde{b} = \mathrm{mod}(b+1,2)$. Of course, these recurrences closely parallel (**??**). To initialise these two mutual recurrences we can use the case $n = 1$, in which we known by inspection that

$$s^b_{\underline{k}\leq\bar{k}}(1,m) = [\![m=0]\!][\![\underline{k}>1]\!] + [\![m=1]\!][\![\underline{k}=1]\!]. \tag{227}$$

Setting $n = 1$ in (**??**) we get

$$s^b_{\underline{k}\leq\bar{k}}(1,m) = s^{\tilde{b}}_{\underline{k}\leq\bar{k}}(0,m-[\![\underline{k}=1]\!]). \tag{228}$$

In can be easily verified that (**??**) and (**??**) are equal when

$$s^b_{\underline{k}\leq\bar{k}}(0,m) = [\![m=0]\!], \tag{229}$$

which thus initialises the two recurrences. Since the enumeration we are interested in can be put as

$$s_{\underline{k}\leq\bar{k}}(n,m) = s^0_{\underline{k}\leq\bar{k}}(n,m) + s^1_{\underline{k}\leq\bar{k}}(n,m), \tag{230}$$

from this expression and (??) we have the recurrence

$$s_{\underline{k}\le\bar{k}}(n,m) = \sum_{i=\underline{k}}^{\bar{k}} s_{\underline{k}\le\bar{k}}(n-i,m-1) + \sum_{i=1}^{\underline{k}-1} s_{\underline{k}\le\bar{k}}(n-i,m) + \sum_{i=\bar{k}+1}^{n} s_{\underline{k}\le\bar{k}}(n-i,m), \tag{231}$$

which from (??) and (??) is initialised by

$$s_{\underline{k}\le\bar{k}}(0,m) = 2\,[\![m=0]\!]. \tag{232}$$

5.1.2 Generating Function

To get the generating function of $s_{\underline{k}\le\bar{k}}(n,m)$ we make first (??) valid for all n and m. Since setting $n=0$ in (??) gives zero, we achieve our goal by adding $2[\![n=0]\!][\![m=0]\!]$ to it. We can now subtract $s_{\underline{k}\le\bar{k}}(n-1,m)$ from $s_{\underline{k}\le\bar{k}}(n,m)$ using this extended recurrence to get a recurrence free from n-dependent summations:

$$\begin{aligned} s_{\underline{k}\le\bar{k}}(n,m) = {} & 2\,s_{\underline{k}\le\bar{k}}(n-1,m) + s_{\underline{k}\le\bar{k}}(n-\underline{k},m-1) - s_{\underline{k}\le\bar{k}}(n-\underline{k},m) \\ & - s_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+1),m-1\big) + s_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+1),m\big) \\ & + 2[\![m=0]\!]\big([\![n=0]\!]-[\![n=1]\!]\big). \end{aligned} \tag{233}$$

We get next the ogf $S_{\underline{k}\le\bar{k}}(x,y) = \sum_{n,m} s_{\underline{k}\le\bar{k}}(n,m)\,x^n y^m$ through the usual procedure of multiplying (??) on both sides by $x^n y^m$ and then summing over n and m, which yields

$$S_{\underline{k}\le\bar{k}}(x,y) = \frac{2(1-x)}{1-2x+(1-y)(x^{\underline{k}}-x^{\bar{k}+1})}. \tag{234}$$

We can now state a simple but relevant theorem.

Theorem 5.1 (Fundamental link between the enumeration of n-strings containing prescribed quantities of ($\underline{k}\le\bar{k}$)-runs of ones and/or zeros and its counterpart for runs of ones)**.**

$$S_{\underline{k}\le\bar{k}}(x,y) = 2x\,W_{(\underline{k}-1)\le(\bar{k}-1)}(x,y). \tag{235}$$

Proof. Substitute (??) *into* (??). □

We can also produce this theorem by comparing recurrences (??) and (??) together with their initialisations, but any possible hesitation about the relationship between these two families of enumerations vanishes after comparing their respective ogfs. As far as we know, the earliest sign of Theorem **??** appears in a result by Schilling —see start of Section **??**

Theorem **??** is the main reason why we have restricted our analysis in this section (Section **??**) to nonnull runs of ones and zeros. Since null runs of ones are well defined (see again Definition **??** and Remark **??** in the introduction), results for nonnull runs of ones and zeros can always be translated into results for runs of ones through (??) when $\underline{k}\ge 1$.

Remark. A *Smirnov word* is a string of symbols from an alphabet in which no two adjacent symbols are equal —equivalently, a string without *levels* [**?**, Secs. 2.4.13, 2.4.14]. The number of binary Smirnov words of length n is trivially two for $n \geq 1$. We verify next that this can indeed be seen through (**??**). Using this expression and (**??**) we can write

$$[y^m]S_{\geq k}(x, y) = [y^m]\, 2\, x\, W_{\geq (k-1)}(x, y) = 2\, x^{mk} \frac{1 - x}{\left(1 - 2x + x^k\right)^{m+1}}. \tag{236}$$

So $[y^0]S_{\geq 2}(x, y) = 2/(1 - x)$, and thus $s_{\geq 2}(n, 0) = 2$.

□

5.2 Number of n-Strings that Contain Exactly m Nonnull k-Runs of Ones and/or Zeros

We denote by $s_k(n, m)$ the number of n-strings that contain exactly m nonnull k-runs of ones and/or zeros. This enumeration is a special case of $s_{\underline{k} \leq \bar{k}}(n, m)$, because

$$s_k(n, m) = s_{k \leq k}(n, m). \tag{237}$$

We consider this special enumeration separately because the case $k = 1$ was previously studied by Bloom [**?**].

5.2.1 Recurrences

From (**??**) we have the recurrence

$$s_k(n, m) = s_k(n - k, m - 1) + \sum_{\substack{i=0 \\ i \neq k}}^{n} s_k(n - i, m), \tag{238}$$

which, from (**??**), is initialised by $s_k(0, m) = 2\, [\![m = 0]\!]$. We can also specialise (**??**) to get the following alternative recurrence:

$$\begin{aligned} s_k(n, m) = 2\, s_k(n - 1, m) + s_k(n - k, m - 1) - s_k(n - k, m) - s_k(n - (k + 1), m - 1) \\ + s_k(n - (k + 1), m) + 2[\![m = 0]\!]\big([\![n = 0]\!] - [\![n = 1]\!]\big), \end{aligned} \tag{239}$$

which does not require initialisation as it is valid for all values of n and m —when taking (**??**) into account.

5.2.2 Generating Functions

Considering (**??**), the coefficient of y^m in $S_k(x,y) = \sum_{n,m} s_k(n,m)\, x^n y^m$ can be obtained through Theorem **??** and (**??**):

$$[y^m]S_k(x,y) = 2\, x^{mk}\left(\frac{1-x}{1-2x+x^k-x^{k+1}}\right)^{m+1}. \tag{240}$$

This ogf enumerates the n-strings that contain exactly m k-runs of ones and/or zeros.

5.2.3 Explicit Expression

From (**??**) we have $s_k(n,m) = 2\, w_{k-1}(n-1,m)$, which gives us a double-summation explicit expression through (**??**). This expression does of course work for $k=1$, where we have

$$s_1(n,m) = 2\, w_0(n-1,m). \tag{241}$$

Observe that this mapping makes explicit use of an enumeration of null runs of ones. In any case, we see in the next remark that a simpler single-summation expression is possible when $k=1$.

Remark. Bloom used the term *single* to mean a 1-run of ones or zeros —a name reminiscent of Apostol's *isolated singleton* for a 1-run of ones [**?**]— and obtained a recurrence for the number of n-strings containing exactly m 1-runs of ones and/or zeros [**?**]. Setting $k=1$ in (**??**) we get

$$\begin{aligned} s_1(n,m) = {} & s_1(n-1,m) + s_1(n-1,m-1) + s_1(n-2,m) - s_1(n-2,m-1) \\ & + 2\, [\![m=0]\!]\big([\![n=0]\!] - [\![n=1]\!]\big). \end{aligned} \tag{242}$$

This is essentially Bloom's recurrence [**?**, Eq. (1)] but in an even more general form, since (**??**) is valid for all n and m thanks to the inhomogeneous term —absent in Bloom's expression. Due to this, Bloom's recursion does not work whenever both arguments of $s_1(\cdot,\cdot)$ equal zero in any of its five instances in (**??**), which the author deals with through the initialisation procedure. Bloom also observes that $s_1(n,0) = 2\, F^{(2)}_{n-1}$, which, for example, can also be seen by taking into account (**??**) and the comment about $w_0(n,0)$ in Section **??**. Lastly, Bloom obtains, through counting arguments, a single-summation explicit expression for $s_1(n,m)$. We can find the same expression from the ogf (**??**), which in this case takes the form

$$[y^m]S_1(x,y) = 2\, x^m\left(\frac{1-x}{1-x-x^2}\right)^{m+1}. \tag{243}$$

Applying the negative binomial theorem twice, we can express (**??**) as

$$[y^m]S_1(x,y) = 2\sum_{t\geq 0}\sum_{l\geq 0}\binom{t+m}{m}\binom{l+t-1}{t-1}x^{m+2t+l}. \tag{244}$$

To get the coefficient of x^n in (**??**) we find the nonnegative indices t and l such that $n = m + 2t + l$. From this relation, the maximum of t happens when $l = 0$, which implies that $t \le \lfloor (n-m)/2 \rfloor$. For a given value of t, we have that $l = n - m - 2t$. Therefore, from (**??**) we have that

$$s_1(n,m) = 2 \sum_{t=0}^{\lfloor \frac{n-m}{2} \rfloor} \binom{t+m}{m} \binom{n-m-t-1}{t-1}, \tag{245}$$

which is the same as [**?**, Eq. (7)], just noting that for $t = 0$ the second binomial coefficient is zero —which means that the summation may start at $t = 1$.

To conclude this remark, we define a rv modelling the number of 1-runs of ones and/or zeros in an n-string drawn uniformly at random, which we call $\widetilde{M}_{1,n}$. In standard probability notation, $\Pr(\widetilde{M}_{1,n} = m) = s_1(n,m)/2^n$. Bloom gave the expectation and variance of $\widetilde{M}_{1,n}$ in [**?**, Eqs. (3) and (4)]. From (**??**), we get the first and second factorial moments of $\widetilde{M}_{1,n}$ by using $n-1$, $k = 0$, and $q = 1/2$ in (**??**) and (**??**), respectively —see Section **??**. The expectation is

$$\mathrm{E}\big(\widetilde{M}_{1,n}\big) = \left(\frac{n+2}{4}\right) [\![n > 1]\!] + [\![n = 1]\!], \tag{246}$$

whereas the second factorial moment is

$$\mathrm{E}\big(\widetilde{M}_{1,n}(\widetilde{M}_{1,n} - 1)\big) = \left(\frac{n(n+5)}{16}\right) [\![n > 2]\!] + \frac{1}{2} [\![n = 2]\!]. \tag{247}$$

Thus $\mathrm{Var}(\widetilde{M}_{1,n}) = (5n+4)/16$ for $n \ge 2$, as Bloom proves by induction. Of course, (**??**) and (**??**) can be readily extended to $\widetilde{M}_{k,n}$ in the same scenario. However, with unequal bit probabilities we must follow a different procedure —see Section **??**. □

5.3 Number of n-Strings that Contain Exactly m Nonnull Runs of Ones and/or Zeros

We denote by $s(n,m)$ the number of n-strings that contain exactly m nonnull runs (of any lengths, all strictly greater than zero) of ones and/or zeros. This enumeration is a special case of $s_{\underline{k} \le \bar{k}}(n,m)$, as

$$s(n,m) = s_{1 \le n}(n,m). \tag{248}$$

We consider this special enumeration separately because it was previously studied by Goulden and Jackson [**?**, p. 76].

5.3.1 Recurrences

From (**??**) and (**??**) we have the recurrence

$$s(n,m) = \sum_{i=1}^{n} s(n-i, m-1), \tag{249}$$

$$\begin{array}{c}|{*}{*}{*}|{*}{*}{*}{*}|{*}{*}|{*}{*}{*}{*}{*}|\cdots|{*}|{*}{*}| \\ \hline 111\ 0000\ 11\ 00000 \cdots\ 0\ 11 \\ 000\ 1111\ 00\ 11111 \cdots\ 1\ 00\end{array}$$

Figure 1: Partitioning of an n-string into nonempty substrings versus runs.

which, from (**??**), is initialised by $s(0, m) = 2\,[\![m = 0]\!]$. An alternative recurrence is found by specialising (**??**):

$$s(n, m) = s(n-1, m) + s(n-1, m-1) + 2[\![m = 0]\!]\big([\![n = 0]\!] - [\![n = 1]\!]\big). \tag{250}$$

This recurrence does not require initialisation and it is valid for all m and n.

5.3.2 Generating Function

The coefficient of y^m in $S(x, y) = \sum_{n,m} s(n, m)\, x^n y^m$ can be obtained by setting $k = 1$ in (**??**):

$$[y^m]S(x, y) = \frac{2x^m}{(1-x)^m}. \tag{251}$$

5.3.3 Explicit Expression

To get $s(n, m) = [x^n y^m]S(x, y)$ we apply the negative binomial theorem to (**??**), which allows us to express it as

$$[y^m]S(x, y) = 2\sum_{t\geq 0}\binom{t+m-1}{m-1}x^{t+m}. \tag{252}$$

Thus, from $n = t + m$, the coefficient of x^n in (**??**) is

$$s(n, m) = 2\binom{n-1}{m-1}, \tag{253}$$

which was given by Goulden and Jackson [**?**, p. 76].

Remark. Expression (**??**) can alternatively be obtained through simple combinatorial reasoning. Consider the partitioning of an n-string into m nonempty substrings, where $1 \leq m \leq n$. As illustrated in the "stars and bars" example in Figure **??**, in which the n-string is represented by n asterisks and the m partitions by $m + 1$ vertical bars, we can put this partition into a bijection with $2m$ runs of ones and zeros —corresponding to an n-string and to its ones' complement. The number of ways in which we can partition an n-string into m nonempty substrings is $\binom{n-1}{m-1}$, from which (**??**) follows.

Finally, note that we can use (**??**) and Theorem **??** to write $s(n,m) = 2\,w_{0\leq n-1}(n-1,m)$. Thus, from (**??**), the number of n-strings that contain exactly m runs *of ones*, including null runs, is $\binom{n}{m-1}$ —cf. (**??**), which does not count null runs. □

5.4 Number of n-Strings Whose Longest Nonnull Run Is a $(\leq k)$-Run of Ones or Zeros

We denote by $\bar{s}_{\leq k}(n)$ the number of n-strings whose longest nonnull run is a $(\leq k)$-run of ones or zeros. Thus, we assume $k \geq 1$. We can study this problem as a special case of the results in Section **??**, because

$$\bar{s}_{\leq k}(n) = s_{(k+1)\leq n}(n,0). \tag{254}$$

Again, we look at this special case in some detail because previous authors studied this problem. Schilling deduced through simple reasoning that [**?**, Eq. (2)]

$$\bar{s}_{\leq k}(n) = 2\,\overline{w}_{\leq(k-1)}(n-1), \tag{255}$$

which is Theorem **??** in action in this special case. Also, Suman studied the number of b-ary strings of length n whose longest run is a $(\leq k)$-run of any of the b-ary symbols [**?**], giving a recurrence, an ogf, and three explicit expressions (one of them asymptotic). We only consider here Suman's nonasymptotic results in the binary case, in which his r and k correspond in our setting to k and 2, respectively. In Suman's notation, $\bar{s}_{\leq k}(n) = c_n$. Finally, Bloom also considered the asymptotics of $\bar{s}_{\leq k}(n)$ in a particular case [**?**].

5.4.1 Recurrences

By applying (**??**) to recurrence (**??**) we get

$$\bar{s}_{\leq k}(n) = \sum_{i=1}^{k} \bar{s}_{\leq k}(n-i), \tag{256}$$

initialised by $\bar{s}_{\leq k}(0) = 2$. This recurrence was given by Suman [**?**, p. 121]. Of course, (**??**) is also Schilling's recurrence (**??**) transformed by (**??**).

An alternative recurrence is obtained by applying (**??**) to (**??**), which yields

$$\bar{s}_{\leq k}(n) = 2\,\bar{s}_{\leq k}(n-1) - \bar{s}_{\leq k}\big(n-(k+2)\big) + 2\big([\![n=0]\!] - [\![n=1]\!]\big), \tag{257}$$

and does not require initialisation —other than taking (**??**) into account. Bloom discussed the asymptotics of $\bar{s}_{\leq 5}(n)$ using the recurrence for runs of ones equivalent to (**??**) —i.e., considering (**??**), recurrence (**??**) with $k = 4$, relying on its characteristic equation [**?**, p. 126].

5.4.2 Generating Functions

Through (**??**) and (**??**) we have that the ogf $\overline{S}_{\leq k}(x) = \sum_n \bar{s}_{\leq k}(n)\, x^n$ is

$$\overline{S}_{\leq k}(x) = \frac{2(1-x)}{1-2x+x^{k+1}}. \tag{258}$$

Suman [**?**, p. 121] gave the following alternative equivalent ogf:

$$\overline{S}_{\leq k}(x) = \frac{1-x^{k+1}}{1-2x+x^{k+1}}, \tag{259}$$

which is, in Suman's notation, $A(x)$ in the binary case with some straightforward algebraic simplification.

5.4.3 Explicit Expression

Suman extracted a double-summation explicit expression for $\bar{s}_{\leq k}(n)$ from (**??**) [**?**, Thm. 1], and he then simplified it into a single-summation version [**?**, Thm. 2]. Regarding our own results, we can see that using (**??**), (**??**) and (**??**) we can directly obtain a single-summation explicit expression.

5.5 Number of n-Strings that Contain Exactly m Nonnull p-Parity Runs of Ones and/or Zeros

We denote by $s_{[p]}(n, m)$ the number of n-strings that contain exactly m nonnull p-parity runs of ones and/or zeros, i.e., exactly m runs of ones and/or zeros whose lengths $k_1, \ldots, k_m$ have parity $\mathrm{mod}(k_i, 2) = p$ for $i = 1, \ldots, m$ and are strictly greater than zero. The n-strings that we enumerate may have more than m nonnull runs of ones and/or zeros, as long as the lengths of all these additional runs have parity opposite to p, and any number of null runs (whose parity is even). For the same reason as in Section **??**, if we call $c_{[p]}(n, m)$ the number of compositions of n that contain exactly m "p-parity parts" —i.e., parts whose parity is p— then we also have that

$$c_{[p]}(n, m) = \frac{1}{2}\, s_{[p]}(n, m). \tag{260}$$

As far as we are aware, Goulden and Jackson are the only authors who have addressed enumeration problems related to the one considered in this section. These authors have enumerated the n-strings in which an odd run of zeros is never followed by an odd run of ones [**?**, Sec. 2.4.6], and the n-strings in which all runs of ones are even and all runs of zeros are odd [**?**, Ex. 2.4.3]. Clearly, the results in this section do not encompass these two enumerations.

Before continuing, we state a necessary condition similar to (**??**).

Necessary Condition. (Existence of n-strings containing m nonnull p-parity runs of ones and/or zeros)

$$s_{[p]}(n,m) > 0 \quad \Longrightarrow \quad 0 \leq m \leq \left\lfloor \frac{n}{1+[\![p=0]\!]} \right\rfloor. \tag{261}$$

□

This is really the same necessary condition as (**??**), just noting that the lengths of all nonnull p-parity runs of ones and/or zeros are lower bounded by $\underline{k} = 1 + [\![p=0]\!]$.

5.5.1 Recurrence

We can produce a recurrence to enumerate $s_{[p]}(n,m)$ with the same strategy used in Section **??**: we first produce two mutual recurrences for the number of n-strings that start with bit b and contain exactly m p-parity runs of ones and/or zeros, denoted by $s^b_{[p]}(n,m)$, where $b \in \{0,1\}$. Considering all the n-strings that start with a nonnull i-run of each length, we can readily see through a similar reasoning as in Section **??** that the mutual recurrences sought are given by

$$s^b_{[p]}(n,m) = \sum_{i=1}^{n} s^{\tilde{b}}_{[p]}(n-i, m-[\![p=\mathrm{mod}(i,2)]\!]), \tag{262}$$

for $b \in \{0,1\}$, and where $\tilde{b} = \mathrm{mod}(b+1,2)$. To find initial values for the mutual recurrences in (**??**) we use the case $n=1$, in which we know by inspection that

$$s^b_{[p]}(1,m) = [\![m=0]\!][\![p=0]\!] + [\![m=1]\!][\![p=1]\!]. \tag{263}$$

On the other hand, setting $n=1$ in (**??**) yields

$$s^b_{[p]}(1,m) = s^{\tilde{b}}_{[p]}(0, m-[\![p=1]\!]). \tag{264}$$

We may verify that (**??**) equals (**??**) when

$$s^b_{[p]}(0,m) = [\![m=0]\!]. \tag{265}$$

which are therefore the initial values of the mutual recurrences in (**??**). Since the enumeration we are interested in can be written as

$$s_{[p]}(n,m) = s^0_{[p]}(n,m) + s^1_{[p]}(n,m), \tag{266}$$

from this expression and (**??**) we have the recurrence

$$s_{[p]}(n,m) = \sum_{i=1}^{n} s_{[p]}(n-i, m-[\![p=\mathrm{mod}(i,2)]\!]), \tag{267}$$

which from (**??**) and (**??**) is initialised by

$$s_{[p]}(0,m) = 2[\![m=0]\!]. \tag{268}$$

5.5.2 Generating Functions

We now obtain the ogf $S_{[p]}(x,y) = \sum_{n,m} s_{[p]}(n,m)\, x^n y^m$. First of all, we make $s_{[p]}(n,m)$ valid for all values of n and m. Setting $n=0$ in (**??**) we get $s_{[p]}(0,m)=0$ instead of the correct value $2[\![m=0]\!]$. Thus we just need to add $2[\![m=0]\!][\![n=0]\!]$ to (**??**) to obtain an extended recurrence valid for all n and m, which we use subsequently.

Next, we obtain a recurrence without an n-dependent summation by determining $s_{[p]}(n,m) - s_{[p]}(n-2,m)$ using the extended recurrence. This yields

$$\begin{aligned} s_{[p]}(n,m) = s_{[p]}(n-2,m) + s_{[p]}(n-1,m-[\![p=1]\!]) + s_{[p]}(n-2,m-[\![p=0]\!]) \\ + 2[\![m=0]\!]\big([\![n=0]\!]-[\![n=2]\!]\big). \end{aligned} \tag{269}$$

By multiplying next (**??**) on both sides by $x^n y^m$ and then adding on n and m we find that

$$S_{[p]}(x,y) = \frac{2\,(1-x^2)}{1 - x\, y^{[\![p=1]\!]} - x^2\big(1+y^{[\![p=0]\!]}\big)}. \tag{270}$$

In general, this ogf cannot always be connected to its counterpart (**??**) that counts only runs of ones, unlike $S_{\underline{k}\le\bar{k}}(x,y)$ in the previous section —see Theorem **??**. But not all hope is lost, as the next simple theorem shows.

Theorem 5.2. *Fundamental link between enumerations of n-strings that contain prescribed quantities of nonnull even runs of ones and/or zeros and enumerations of n-strings containing prescribed quantities of odd runs of ones.*

$$S_{[0]}(x,y) = 2x\, W_{[1]}(x,y). \tag{271}$$

Proof. Substitute (**??**) *into* (**??**). □

Remark. There is no similar converse mapping between $S_{[1]}(x,y)$ and $W_{[0]}(x,y)$, as one might expect. This asymmetry is rather unsatisfying, but it has a satisfying explanation: we have only considered *nonnull* p-parity runs of ones to get $W_{[p]}(x,y)$. This assumption is immaterial when $p=1$, as odd runs of ones are always nonnull. However, when $p=0$ we are disregarding all null runs of ones, which have even parity. Thus, the symmetry of the setting with respect to runs of ones and zeros is broken. We will not delve further into this issue here, but it is not difficult to see that if we modify $W_{[p]}(x,y)$ to include null runs of ones, then symmetry is restored and (**??**) becomes $S_{[p]}(x,y) = 2x\, W_{[\tilde{p}]}(x,y)$ with $p\in\{0,1\}$ and $\tilde{p} = \mathrm{mod}(p+1,2)$. This is yet another example of the relevant role played by null runs. □

Extracting the coefficient of y^m from (**??**) using the procedure that we have repeatedly used throughout the paper yields

$$[y^m]S_{[p]}(x,y) = \frac{2\,(1-x^2)\, x^{(1+[\![p=0]\!])\,m}}{\Big(1-[\![p=0]\!]x - \big(1+[\![p=1]\!]\big)x^2\Big)^{m+1}}. \tag{272}$$

This is the ogf enumerating the binary strings that contain exactly m nonnull p-parity runs of ones and/or zeros.

Remark. In this remark we briefly discuss the consequences of our results in Section **??** for compositions. We start with the results in Section **??**. Putting together Theorem **??** and the connection (**??**) between runs of ones and zeros and compositions, we find that the ogf $C_{\underline{k}\leq\bar{k}}(x,y) = \sum_{n,m} c_{\underline{k}\leq\bar{k}}(n,m)\, x^n y^m$ is

$$C_{\underline{k}\leq\bar{k}}(x,y) = \frac{1}{2} S_{\underline{k}\leq\bar{k}}(x,y) = x\, W_{\underline{k}-1\leq\bar{k}-1}(x,y). \tag{273}$$

Thus, not only all results in Sections **??**–**??** —which, we remind, only concern runs of ones— can be directly applied to the corresponding enumerations of n-strings with prescribed quantities of nonnull runs of ones and/or zeros, but they can also be directly applied to the corresponding enumerations of compositions of n with prescribed quantities of parts. We give next some examples that recover known results due to other authors. We mainly focus on generating functions, but of course recurrences and explicit results also follow from the connections given.

- From (**??**), by dividing (**??**) and (**??**) by 2 we get the ogf and the explicit expression for compositions with exactly m parts (of any sizes), i.e., $[y^n]C(x,y) = x^m/(1-x)^m$ and $c(n,m) = \binom{n-1}{m-1}$, both of which were given by Riordan [**?**, p. 124]. Goulden and Jackson [**?**, p. 53] gave as well $c(n,m)$ and $C(x,y) = (1-x)/(1-(1+y)x)$, which we can also obtain using $\underline{k}=1$ from (**??**) and (**??**). Observe that Goulden and Jackson also found (**??**), but apparently they did not make the connection between runs (in their nomenclature, maximal blocks) and compositions.

- Using (**??**) and (**??**), the ogf for the number of compositions with no part greater than k is
$$\overline{C}_{\leq k}(x) = x\,\overline{W}_{\leq(k-1)}(x) = \frac{x - x^{k+1}}{1-2x+x^{k+1}}, \tag{274}$$
which was also given by Riordan [**?**, Eq. (40)]. An alternative ogf for (**??**) was also given by Heubach and Mansour [**?**, Ex. 2.7].

- From (**??**) and (**??**), the number of compositions with no part greater than k and at least one part equal to k has ogf
$$\overline{C}_k(x) = x\,\overline{W}_{(k-1)}(x) = \frac{x^k(1-x)^2}{(1-2x+x^k)(1-2x+x^{k+1})}, \tag{275}$$
which was given by Riordan as well [**?**, p. 155].

  Expression (**??**) can be read as a bijection between the compositions of $n+1$ with largest part equal to k and the n-strings with longest run (of ones) a $(k-1)$-run. Nej and Reddy [?] say that there is a bijection between the compositions of $n+1$ with largest part k and the n-strings with Hamming weight r whose longest run is a $(k-1)$-run, but, given (??), this assertion requires considering all possible Hamming weights $0 \leq r \leq n$.

- Finally, from (??) and (??) we have that the number of compositions with no part equal to k has ogf

$$[y^0]C_k(x,y) = [y^0]x\,W_{(k-1)}(x,y) = \frac{1-x}{1-2x+x^k-x^{k+1}}. \tag{276}$$

  This ogf was given by Chinn and Heubach [?, Thm. 1]. The recurrences given by these authors for this enumeration [?, Thm. 1] can also be found by using $c_k(n,m) = w_{k-1}(n-1,m)$ with $m=0$ in (??) or in (??).‘

We consider subsequently the consequences for compositions of the results in Section ??. Letting $C_{[p]}(x,y) = \sum_{n,m} c_{[p]}(n,m)\,x^n y^m$, from (??) we have that Theorem ?? carries over to the enumeration of compositions having prescribed quantities of even parts as follows:

$$C_{[0]}(x,y) = \frac{1}{2}S_{[0]}(x,y) = x\,W_{[1]}(x,y). \tag{277}$$

Thus, from these equalities and (??) we have

$$C_{[0]}(x,y) = \frac{1-x^2}{1-x-x^2(1+y)}, \tag{278}$$

which was previously given by Goulden and Jackson [?, p. 54].

Grimaldi and Heubach identified the bijection between the compositions of $n+1$ with only odd parts and the n-strings without odd runs of zeros [?, Sec. 5]. Now, from (??) we can write

$$[y^m]C_{[0]}(x,y) = [y^m]x\,W_{[1]}(x,y), \tag{279}$$

which allows us to make a stronger statement: there is a bijection between the compositions of $n+1$ with exactly m even parts and the n-strings that contain exactly m odd runs of ones (or, indeed, exactly m odd runs of zeros), of which Grimaldi and Heubach's observation is the special case $m=0$. Finally, in this case we have from (??) and (??) that

$$[y^0]C_{[0]}(x,y) = \frac{1-x^2}{1-x-x^2}, \tag{280}$$

i.e., the ogf enumerating compositions with only odd parts, which was given by Heubach and Mansour [?, Ex. 2.9].

Last but not least, Heubach and Mansour obtained many of their results for compositions as a specialisation of a general theorem of theirs that gives the ogf of the number of compositions with m parts in a general subset $A \subseteq \mathbb{N}$ [?, Thm. 2.1]. Through Theorems ?? and ??, most ogfs for runs in the current paper can in principle be recovered by relying on Heubach and Mansour's theorem, and also new results for runs be obtained.

Some further comments about compositions and runs are given in Remark ??.

□

5.5.3 OEIS

The sequences connected to $s_{[p]}(n, m)$ in the OEIS are:

- $m = 0$

  $s_{[1]}(n, 0)$ is A077957 for $n \geq 1$.

  $s_{[0]}(n-1, 0)$ is A006355 for $n \geq 2$.

- $m = 1$

  $s_{[1]}(2n-3, 1)$ is A057711 for $n \geq 2$.

  $s_{[0]}(n+1, 1)$ is A320947 for $n \geq 2$.

6 Extensions to Probabilistic Runs of Ones and/or Zeros

Just as in Section ??, where we discussed the probabilistic extension of the enumerative results in Section ??, we can extend the enumerative results in Section ?? to the same probabilistic scenario: the case where the n-strings are outcomes from n iid Bernoulli random variables with parameter q, where $0 < q < 1$ is the probability of drawing a 1. As it may be intuited, in this scenario there is no easy mapping between the results for runs of ones and the results for runs of ones and zeros, i.e., there are no probabilistic analogues of Theorems ?? and ??, and the expressions can get rather involved.

We mainly deal in this section with the extension of the results in Section ?? to the probabilistic scenario. We also examine the extension of the special case in Section ??, as this was previously analysed in detail by Wishart and Hirschfeld [?]. The same approach can obviously be used to extend the results in Section ??, although we omit it here.

Like in Section ??, it is understood that all expressions in this section implicitly depend on q.

6.1 Probability that an n-String Contains Exactly m Nonnull ($\underline{k} \leq \bar{k}$)-Runs of Ones and/or Zeros

We call $\lambda_{\underline{k}\leq\bar{k}}(n,m)$ the probability that an n-string contains exactly m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros. The obvious case is $q = 1/2$, in which $\lambda_{\underline{k}\leq\bar{k}}(n,m) = s_{\underline{k}\leq\bar{k}}(n,m)/2^n$. This was in fact the approach followed in special cases of this expression by Bloom [?] and by Suman [?], who produced $\lambda_1(n,m) = s_1(n,m)/2^n$ and $\overline{\lambda}_{\leq k}(n) = \bar{s}_{\leq k}(n)/2^n$, respectively, using explicit expressions —Suman's results actually apply to the longest run of any b-ary symbol in b-ary strings of length n. We examine next how to get $\lambda_{\underline{k}\leq\bar{k}}(n,m)$ for arbitrary q.

6.1.1 Recurrence

As per our notation conventions, we denote the joint probability that an n-string starts with b and contains exactly m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros by $\lambda^b_{\underline{k}\leq\bar{k}}(n,m)$. By the law of total probabilities, the probability that we wish to determine can be expressed in terms of these joint probabilities as

$$\lambda_{\underline{k}\leq\bar{k}}(n,m) = \lambda^0_{\underline{k}\leq\bar{k}}(n,m) + \lambda^1_{\underline{k}\leq\bar{k}}(n,m). \tag{281}$$

Letting

$$q_b = (1-q)[\![b=0]\!] + q[\![b=1]\!] \tag{282}$$

and invoking again the law of total probabilities, we have that the probabilistic version of the mutual recurrences in (**??**) is

$$\lambda^b_{\underline{k}\leq\bar{k}}(n,m) = \sum_{i=\underline{k}}^{\bar{k}} q_b^i\,\lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-i,m-1) + \sum_{i=1}^{\underline{k}-1} q_b^i\,\lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-i,m) + \sum_{i=\bar{k}+1}^{n} q_b^i\,\lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-i,m), \tag{283}$$

where $b \in \{0,1\}$ and $\tilde{b} = \mathrm{mod}(b+1,2)$. By inspection, the value of $\lambda^b_{\underline{k}\leq\bar{k}}(n,m)$ when $n=1$ is

$$\lambda^b_{\underline{k}\leq\bar{k}}(1,m) = q_b\left([\![m=0]\!][\![\underline{k}>1]\!] + [\![m=1]\!][\![\underline{k}=1]\!]\right). \tag{284}$$

On the other hand, setting $n=1$ in (**??**) gives

$$\lambda^b_{\underline{k}\leq\bar{k}}(1,m) = q_b\,\lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(0, m-[\![\underline{k}=1]\!]), \tag{285}$$

and we can see that (**??**) equals (**??**) for

$$\lambda^b_{\underline{k}\leq\bar{k}}(0,m) = [\![m=0]\!]. \tag{286}$$

This is therefore the initialisation of (**??**). Although we can get $\lambda_{\underline{k}\leq\bar{k}}(n,m)$ using (**??**) and (**??**), the asymmetry of the mutual recurrences in (**??**) prevents us from obtaining a recurrence for $\lambda_{\underline{k}\leq\bar{k}}(n,m)$ itself. This theme —asymmetry between mutual recurrences— will resurface in all the enumerations in Section **??**, which can be, in fact, related asymptotically to the results in this section through the law of large numbers.

Remark. Schilling mentioned that a recurrence for $\overline{\lambda}_{\leq k}(n) = \lambda_{(k+1)\leq n}(n,0)$ can be obtained [**?**, p. 200] in a similar way as (**??**). While he did not furnish the recurrence, this author indicated that the approximation $\overline{\lambda}_{\leq k}(n) \approx \overline{\pi}_{\leq k}(n)$ —see Section **??**— works well for very large n when using $\check{q} = \max(q, 1-q)$ to compute $\overline{\pi}_{\leq k}(n)$. Empirically, for fixed n, the accuracy of this approximation increases as k increases.

□

6.1.2 Probability Generating Function

In spite of not having a recurrence for $\lambda_{\underline{k}\leq\bar{k}}(n,m)$, it is still possible to get the pgf $\Lambda^b_{\underline{k}\leq\bar{k}}(x,y) = \sum_{n,m} \lambda^b_{\underline{k}\leq\bar{k}}(n,m)\, x^n y^m$, and then, through (**??**), obtain $\Lambda_{\underline{k}\leq\bar{k}}(x,y) = \sum_{n,m} \lambda_{\underline{k}\leq\bar{k}}(n,m)\, x^n y^m$ using

$$\Lambda_{\underline{k}\leq\bar{k}}(x,y) = \Lambda^0_{\underline{k}\leq\bar{k}}(x,y) + \Lambda^1_{\underline{k}\leq\bar{k}}(x,y). \tag{287}$$

As usual, we first make (**??**) valid for all values of n and m. Setting $n = 0$ in (**??**) gives $\lambda^b_{\underline{k}\leq\bar{k}}(0,m) = 0$ instead of (**??**), so we just need to add $[\![n=0]\!][\![m=0]\!]$ to (**??**) to achieve our goal. Using this extended recurrence, we compute next $\lambda^b_{\underline{k}\leq\bar{k}}(n,m) - q_b\, \lambda^b_{\underline{k}\leq\bar{k}}(n-1,m)$ to produce a mutual recurrence free from n-dependent summations. This yields

$$\begin{aligned}
\lambda^b_{\underline{k}\leq\bar{k}}(n,m) - q_b\, \lambda^b_{\underline{k}\leq\bar{k}}(n-1,m) = {} & q_b{}^{\underline{k}}\, \lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-\underline{k}, m-1) - q_b{}^{\bar{k}+1}\, \lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-(\bar{k}+1), m-1) \\
& + q_b\, \lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-1,m) - q_b{}^{\underline{k}}\, \lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-\underline{k}), m) \\
& + q_b{}^{\bar{k}+1}\, \lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(n-(\bar{k}+1), m) \\
& + \big([\![n=0]\!] - q_b\,[\![n=1]\!]\big)[\![m=0]\!].
\end{aligned} \tag{288}$$

Multiplying now (**??**) on both sides by $x^n y^m$ and adding over n and m we obtain

$$\Lambda^b_{\underline{k}\leq\bar{k}}(x,y)(1-q_b\,x) = \Lambda^{\tilde{b}}_{\underline{k}\leq\bar{k}}(x,y)\Big((y-1)\big(q_b{}^{\underline{k}}x^{\underline{k}} - q_b{}^{\bar{k}+1}x^{\bar{k}+1}\big) + q_b\,x\Big) + 1 - q_b\,x. \tag{289}$$

This is a system of two equations with two unknowns, i.e., the pgfs $\Lambda^0_{\underline{k}\leq\bar{k}}(x,y)$ and $\Lambda^1_{\underline{k}\leq\bar{k}}(x,y)$. In order to streamline the upcoming expressions we now define

$$x_b = q_b\,x. \tag{290}$$

Solving the system in (**??**) for $\Lambda^1_{\underline{k}\le\bar{k}}(x,y)$ we get

$$\Lambda^1_{\underline{k}\le\bar{k}}(x,y)=\frac{(1-x_1)(1-x_0)+\Big((y-1)\big(x_1^{\underline{k}}-x_1^{\bar{k}+1}\big)+x_1\Big)(1-x_0)}{(1-x_1)(1-x_0)-\Big((y-1)\big(x_1^{\underline{k}}-x_1^{\bar{k}+1}\big)+x_1\Big)\Big((y-1)\big(x_0^{\underline{k}}-x_0^{\bar{k}+1}\big)+x_0\Big)}. \tag{291}$$

Inputting (**??**) in (**??**) we get $\Lambda^0_{\underline{k}\le\bar{k}}(x,y)$, and using then (**??**) we arrive after some algebra at

$$\Lambda_{\underline{k}\le\bar{k}}(x,y)=\frac{2-x_1-x_0+(y-1)\Big(\big(x_1^{\underline{k}}-x_1^{\bar{k}+1}\big)(1-x_0)+\big(x_0^{\underline{k}}-x_0^{\bar{k}+1}\big)(1-x_1)\Big)}{(1-x_1)(1-x_0)-\Big((y-1)\big(x_1^{\underline{k}}-x_1^{\bar{k}+1}\big)+x_1\Big)\Big((y-1)\big(x_0^{\underline{k}}-x_0^{\bar{k}+1}\big)+x_0\Big)}. \tag{292}$$

This pgf allows us to obtain $\lambda_{\underline{k}\le\bar{k}}(n,m)=[x^ny^m]\Lambda_{\underline{k}\le\bar{k}}(x,y)$ using a computer algebra system, in a more efficient manner than through recurrences (**??**) or (**??**). We must surely give up hope of finding a reasonably simple closed-form general expression here, but this is not the case in special scenarios as we see next.

6.2 Probability that an n-String Contains Exactly m Nonnull Runs of Ones and/or Zeros

Let $\lambda(n,m)$ be the probability that an n-string contains exactly m nonnull runs of ones and/or zeros. This is a special case of the probability studied in the previous section with $\underline{k}=1$ and $\bar{k}=n$, and thus $\lambda(n,m)=\lambda_{1\le n}(n,m)$. Wishart and Hirschfeld previously studied this problem in some detail [**?**].

6.2.1 Recurrences

Letting $\lambda^b(n,m)=\lambda^b_{1\le n}(n,m)$, the mutual recurrences (**??**) become

$$\lambda^b(n,m)=\sum_{i=1}^{n} q_b^i\,\lambda^{\tilde{b}}(n-i,m-1), \tag{293}$$

which, from (**??**), are initialised by $\lambda^b(0,m)=[\![m=0]\!]$. On the other hand, the alternative mutual recurrences (**??**) now become

$$\lambda^b(n,m)-q_b\,\lambda^b(n-1,m)=q_b\,\lambda^{\tilde{b}}(n-1,m-1)+\big([\![n=0]\!]-q_b\,[\![n=1]\!]\big)[\![m=0]\!], \tag{294}$$

which do not need initialisation thanks to the inhomogeneous term. The two mutual recurrences in (**??**) were given by Wishart and Hirschfeld [**?**, Eqs. (1) and (2)] —in their notation, $\lambda(n,m)=P_n(m-1)$— but without the inhomogeneous term, which in fact makes (**??**) more general, as it renders it valid for all $n\ge1$ and $m\ge0$ taking (**??**) into account.

6.2.2 Probability Generating Function

Although we may particularise (**??**) using $\underline{k} = 1$ and $\overline{k} = n$, as in other similar cases this approach leads to an n-dependent pgf. A simpler, more general pgf is possible in this case. Working from (**??**) and following the same steps as in Section **??** (i.e., first obtaining a system of two equations with the two unknowns $\Lambda^b(x,y) = \sum_{n,m} \lambda^b(x,y)\, x^n y^m$, then solving it, and finally adding the two solutions) is not difficult to see that the pgf $\Lambda(x,y) = \sum_{n,m} \lambda(n,m)\, x^n y^m$ is

$$\Lambda(x,y) = \frac{2(1-q_0x)(1-q_1x) + \big(q_0(1-q_1x) + q_1(1-q_0x)\big)\, xy}{(1-q_0x)(1-q_1x) - q_0q_1x^2y^2}. \tag{295}$$

As usual, we may extract the coefficient of y^m from (**??**) using the negative binomial theorem. For m even we have

$$[y^m]\Lambda(x,y) = 2\left(\frac{q_0q_1\, x^2}{1-x+q_0q_1x^2}\right)^{\frac{m}{2}}, \tag{296}$$

while for m odd the expression is

$$[y^m]\Lambda(x,y) = \left(\frac{1}{q_0q_1x} - 2\right)\left(\frac{q_0q_1\, x^2}{1-x+q_0q_1x^2}\right)^{\frac{m+1}{2}}. \tag{297}$$

6.2.3 Explicit Expressions

We next determine the coefficient of x^n in pgfs (**??**) and (**??**) in order to get single-summation explicit expressions for $\lambda(n,m) = [x^n y^m]\Lambda(x,y)$. We start by developing the common term in (**??**) and (**??**) into a power series using the negative binomial theorem followed by the binomial theorem:

$$\left(\frac{q_0q_1\, x^2}{1-x+q_0q_1x^2}\right)^{v} = \sum_{t\geq 0}\sum_{p\geq 0}\binom{t+v-1}{t}\binom{t}{p}(-1)^p(q_0q_1)^{p-v}x^{2v+t+p}. \tag{298}$$

The coefficient of x^n in (**??**) is found by solving $2v+t+p=n$ for the nonnegative indices t and p. The maximum of t happens when $p=0$, and thus $t \leq n-2v$. Therefore, letting $\omega(n,v) = [x^n]\big(q_0q_1\, x^2/(1-x+q_0q_1x^2)\big)^v$ and using (**??**) in the following, we have from (**??**) that

$$\omega(n,v) = \sum_{t=0}^{n-2v}\binom{t+v-1}{t}\binom{t}{n-2v-t}(-1)^{n-2v-t}((1-q)q)^{n-v-t}. \tag{299}$$

Hence, from (**??**) and (**??**) we have that for m even

$$\lambda(n,m) = 2\,\omega\Big(n, \frac{m}{2}\Big), \tag{300}$$

whereas from (**??**) and (**??**) the m odd case is

$$\lambda(n,m) = \frac{1}{q(1-q)}\,\omega\Big(n+1, \frac{m+1}{2}\Big) - 2\,\omega\Big(n, \frac{m+1}{2}\Big). \tag{301}$$

These two single-summation expressions for $\lambda(n,m)$ are similar to the ones found by Wishart and Hirschfeld [**?**, Eqs. (22) and (23)]. However, the analysis by these authors —which relies on the moment generating function (mgf) centered about the mean, rather than the pgf— also allows them to study the asymptotics of $\lambda(n,m)$ in a more direct manner.

When $q = 1/2$ we are back to a symmetric scenario, where we have $\lambda(n,m) = s(n,m)/2^n$. Therefore, using (**??**) we can see that

$$\lambda(n,m) = \frac{\binom{n-1}{m-1}}{2^{n-1}}, \tag{302}$$

which is the binomial distribution with parameters $n-1$ and $1/2$, as observed by Wishart and Hirschfeld [**?**, p. 231].

6.2.4 Moments

We now let $\widetilde{M}_n$ be the rv modelling the number of nonnull runs of ones and/or zeros in an n-string drawn at random, whose distribution can be expressed as $\Pr(\widetilde{M}_n = m) = \lambda(n,m)$. Using (**??**), the first moment of $\widetilde{M}_n$ is the coefficient of x^n in

$$\left.\frac{\partial\Lambda(x,y)}{\partial y}\right|_{y=1} = \frac{x\big(1-x+2(1-q)qx\big)}{(1-x)^2}. \tag{303}$$

Expanding this expression using the negative binomial coefficient we have that

$$\left.\frac{\partial\Lambda(x,y)}{\partial y}\right|_{y=1} = \frac{x}{1-x} + 2(1-q)\,qx^2\sum_{t\geq 0}\binom{t+1}{t}x^t, \tag{304}$$

and, hence, the coefficient of x^n in this expression is

$$\mathrm{E}\big(\widetilde{M}_n\big) = 1 + 2(1-q)q(n-1). \tag{305}$$

Another way to obtain this expectation is through (**??**). Explicitly denoting by $M^{(q)}_{\geq k,n}$ the rv in Section **??** with parameter q, we have that $\mathrm{E}(\widetilde{M}_n) = \mathrm{E}(M^{(q)}_{\geq 1,n} + M^{(1-q)}_{\geq 1,n}) = \mathrm{E}(M^{(q)}_{\geq 1,n}) + \mathrm{E}(M^{(1-q)}_{\geq 1,n})$, even if $M^{(q)}_{\geq 1,n}$ and $M^{(1-q)}_{\geq 1,n}$ are not independent. This strategy was used by Cochran [**?**, Eq. (6)] to obtain the expected number of k-runs of ones and/or zeros in an n-string drawn at random from $\mathrm{E}(M_{k,n})$ —see (**??**).

For the second factorial moment we follow the same procedure as above but with the second derivative of (**??**) evaluated at $y = 1$:

$$\left.\frac{\partial^2 \Lambda(x,y)}{\partial y^2}\right|_{y=1} = \frac{2(1-q)qx^2\big(2 - x - (1-2q)^2x^2\big)}{(1-x)^3}. \tag{306}$$

Applying the negative binomial theorem, the coefficient of x^n in this expression is

$$\mathrm{E}\big(\widetilde{M}_n(\widetilde{M}_n - 1)\big) = (1-q)q\big(2n(n-1) - (n-1)(n-2) - (1-2q)^2(n-2)(n-3)\big), \tag{307}$$

from which we may obtain $\mathrm{Var}(\widetilde{M}_n)$ in conjunction with (**??**). Wishart and Hirschfeld obtained (**??**) directly from their version of recurrence (**??**) —recall that in their notation $\lambda(n,m) = P_n(m-1)$— and then used the mgf to get the second, third and fourth *semi-invariants* (i.e., cumulants) of the distribution [**?**, Eqs. (13) and (14)].

7 Number of n-Strings of Hamming Weight r that Contain Prescribed Quantities of Nonnull Runs of Ones and/or Zeros Under Different Constraints

In this section we address the enumeration of the n-strings of Hamming weight r that contain prescribed quantities of nonnull runs of ones and/or zeros under different constraints. This section is to Section **??** what Section **??** is to Section **??**. One important implication of the Hamming weight constraint is that, unlike in Section **??**, no enumeration in this section is related to a counterpart enumeration concerning the compositions of n. This is because an n-string and its ones' complement do not have the same Hamming weight in general. For the same reason, we do not have analogues of Theorems **??** and **??** in this setting: the results in this section are not related to the results in Section **??**. Like in Sections **??** and **??**, we only consider nonnull runs of ones or zeros, and we approach the enumerations through mutual recurrences. Importantly, the Hamming weight constraint creates a fundamental asymmetry between these mutual recurrences like the one we observed in Section **??**, which, in general, makes it harder to obtain overall recurrences and generating functions. Due to this difficulty, we only get explicit expressions in two particular cases.

Regarding previous work on this topic, Stevens [**?**] and Wald and Wolfowitz [**?**] first addressed the enumeration of the n-strings that have a prescribed number m of nonnull runs or ones and/or zeros under the Hamming weight constraint. Their results were later rederived by other authors [**?**, **?**]. Also, Bateman [**?**], Schuster [**?**], Bloom [**?**], and Jackson (see [**?**]) tackled the enumeration of the n-strings of Hamming weight r whose longest nonnull run is a ($\leq k$)-run of ones or zeros.

7.1 Number of n-Strings of Hamming Weight r that Contain Exactly m Nonnull ($\underline{k} \leq \bar{k}$)-Runs of Ones and/or Zeros

In this section we enumerate the n-strings of Hamming weight r that contain exactly m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros, which we denote by $s_{\underline{k}\leq\bar{k}}(n, m, r)$. Thus, we assume $\underline{k} \geq 1$.

As in Section **??**, we first narrow down necessary condition (**??**) to take into account the Hamming weight constraint.

Necessary Condition. (Existence of n-strings of Hamming weight r containing m nonnull ($\underline{k} \leq \bar{k}$)-runs of ones and/or zeros)

$$s_{\underline{k}\leq\bar{k}}(n, m, r) > 0 \quad \Longrightarrow \quad 0 \leq m \leq \min\left(\left\lfloor \frac{n}{\underline{k}} \right\rfloor, \left\lfloor \frac{r}{\underline{k}} \right\rfloor + \left\lfloor \frac{n-r}{\underline{k}} \right\rfloor\right). \tag{308}$$

□

7.1.1 Recurrence

We can enumerate $s_{\underline{k}\leq\bar{k}}(n, m, r)$ with the same strategy used in Section **??**: we first produce two mutual recurrences for the number of n-strings of Hamming weight r that start with bit b and contain exactly m $\underline{k} \leq \bar{k}$-runs of ones and/or zeros, denoted by $s^b_{\underline{k}\leq\bar{k}}(n, m, r)$, where $b \in \{0, 1\}$. The two trivariate mutual recurrences sought are just like in (**??**) but updating the Hamming weight constraints:

$$\begin{aligned} s^b_{\underline{k}\leq\bar{k}}(n, m, r) = {} & \sum_{i=\underline{k}}^{\bar{k}} s^{\tilde{b}}_{\underline{k}\leq\bar{k}}\big(n-i, m-1, r - [\![b=1]\!]\, i\big) + \sum_{i=1}^{\underline{k}-1} s^{\tilde{b}}_{\underline{k}\leq\bar{k}}\big(n-i, m, r - [\![b=1]\!]\, i\big) \\ & + \sum_{i=\bar{k}+1}^{n} s^{\tilde{b}}_{\underline{k}\leq\bar{k}}\big(n-i, m, r - [\![b=1]\!]\, i\big), \end{aligned} \tag{309}$$

where $b \in \{0, 1\}$ and $\tilde{b} = \mathrm{mod}(b+1, 2)$. To initialise these two mutual recurrences we use the case $n = 1$, in which we know by inspection that

$$s^b_{\underline{k}\leq\bar{k}}(1, m, r) = [\![r=b]\!]\Big([\![m=0]\!][\![\underline{k}>1]\!] + [\![m=1]\!][\![\underline{k}=1]\!]\Big). \tag{310}$$

Setting $n = 1$ in (**??**) we have

$$s^b_{\underline{k}\leq\bar{k}}(1, m, r) = s^b_{\underline{k}\leq\bar{k}}\big(0, m - [\![\underline{k}=1]\!], r - [\![b=1]\!]\big). \tag{311}$$

We see that (**??**) equals (**??**) if we choose the following initialisation values:

$$s^b_{\underline{k}\leq\bar{k}}(0, m, r) = [\![m=0]\!][\![r=0]\!]. \tag{312}$$

The main enumeration in this section can now be expressed as

$$s_{\underline{k}\le\bar{k}}(n,m,r) = s^0_{\underline{k}\le\bar{k}}(n,m,r) + s^1_{\underline{k}\le\bar{k}}(n,m,r). \tag{313}$$

Importantly, unlike in Section **??** the mutual recurrences in (**??**) for the n-strings that start with 0 or with 1 are not symmetric. This is analogous to what we saw in Section **??**. Because of this, we cannot simply input (**??**) into (**??**) and then directly work out a recurrence for $s_{\underline{k}\le\bar{k}}(n,m,r)$. Therefore we do not attempt to obtain such a general recurrence, but we see later that this asymmetry can be wrestled with in special cases.

7.1.2 Generating Functions

While we have not been able to produce a general recurrence for $s_{\underline{k}\le\bar{k}}(n,m,r)$, we can still determine its ogf $S_{\underline{k}\le\bar{k}}(x,y,z) = \sum_{n,m,r} s_{\underline{k}\le\bar{k}}(n,m,r)\, x^n y^m z^r$ through the mutual recurrences for $s^b_{\underline{k}\le\bar{k}}(n,m,r)$. First of all, setting $n=0$ in (**??**) we get $s^b_{\underline{k}\le\bar{k}}(n,m,r) = 0$ instead of (**??**). Thus, to get recurrences valid for all n, m, and r we add $[\![n=0]\!][\![m=0]\!][\![r=0]\!]$ to (**??**). We next get recurrences without n-dependent summations by calculating $s^b_{\underline{k}\le\bar{k}}(n,m,r) - s^b_{\underline{k}\le\bar{k}}\big(n-1,m,r-[\![b=1]\!]\big)$ using the extended recurrences. This yields

$$\begin{aligned} s^b_{\underline{k}\le\bar{k}}(n,m,r) = {} & s^b_{\underline{k}\le\bar{k}}\big(n-1,m,r-[\![b=1]\!]\big) + s^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(n-\underline{k},m-1,r-[\![b=1]\!]\,\underline{k}\big) \\ & - s^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+1),m-1,r-[\![b=1]\!]\,(\bar{k}+1)\big) + s^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(n-1,m,r-[\![b=1]\!]\big) \\ & - s^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(n-\underline{k},m,r-[\![b=1]\!]\,\underline{k}\big) + s^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(n-(\bar{k}+1),m,r-[\![b=1]\!]\,(\bar{k}+1)\big) \\ & + [\![m=0]\!]\big([\![r=0]\!][\![n=0]\!] - [\![r=b]\!][\![n=1]\!]\big). \end{aligned} \tag{314}$$

We are now ready to obtain the trivariate ogfs $S^b_{\underline{k}\le\bar{k}}(x,y,z) = \sum_{n,m,r} s^b_{\underline{k}\le\bar{k}}(n,m,r)\, x^n y^m z^r$ for $b \in \{0,1\}$. Multiplying (**??**) on both sides by $x^n y^m z^r$ and adding over n, m, and r yields the following system of equations:

$$S^0_{\underline{k}\le\bar{k}}(x,y,z)(1-x) = S^1_{\underline{k}\le\bar{k}}(x,y,z)\Big(\big(x^{\underline{k}} - x^{\bar{k}+1}\big)(y-1) + x\Big) + 1 - x, \tag{315}$$

$$S^1_{\underline{k}\le\bar{k}}(x,y,z)(1-xz) = S^0_{\underline{k}\le\bar{k}}(x,y,z)\Big(\big(x^{\underline{k}} z^{\underline{k}} - x^{\bar{k}+1} z^{\bar{k}+1}\big)(y-1) + xz\Big) + 1 - xz. \tag{316}$$

Solving the system we get

$$S^1_{\underline{k}\le\bar{k}}(x,y,z) = \frac{(1-x)\Big(\big(x^{\underline{k}} z^{\underline{k}} - x^{\bar{k}+1} z^{\bar{k}+1}\big)(y-1) + 1\Big)}{(1-x)(1-xz) - \Big(\big(x^{\underline{k}} z^{\underline{k}} - x^{\bar{k}+1} z^{\bar{k}+1}\big)(y-1) + xz\Big)\Big(\big(x^{\underline{k}} - x^{\bar{k}+1}\big)(y-1) + x\Big)}, \tag{317}$$

whereas $S^0_{\underline{k}\leq\overline{k}}(x,y,z)$ is obtained by substituting (**??**) in (**??**), which yields

$$S^0_{\underline{k}\leq\overline{k}}(x,y,z) = \frac{\Big(\big(x^{\underline{k}}z^{\underline{k}} - x^{\overline{k}+1}z^{\overline{k}+1}\big)(y-1)+1\Big)\Big(\big(x^{\underline{k}} - x^{\overline{k}+1}\big)(y-1)+x\Big)}{(1-x)(1-xz) - \Big(\big(x^{\underline{k}}z^{\underline{k}} - x^{\overline{k}+1}z^{\overline{k}+1}\big)(y-1)+xz\Big)\Big(\big(x^{\underline{k}} - x^{\overline{k}+1}\big)(y-1)+x\Big)} + 1. \tag{318}$$

Thus, from (**??**), (**??**) and (**??**) the desired ogf is

$$S_{\underline{k}\leq\overline{k}}(x,y,z) = \frac{(1-x)(1-xz) + (1-x)\Big(\big(x^{\underline{k}}z^{\underline{k}} - x^{\overline{k}+1}z^{\overline{k}+1}\big)(y-1)+1\Big) + (1-xz)\Big(\big(x^{\underline{k}} - x^{\overline{k}+1}\big)(y-1)+x\Big)}{(1-x)(1-xz) - \Big(\big(x^{\underline{k}}z^{\underline{k}} - x^{\overline{k}+1}z^{\overline{k}+1}\big)(y-1)+xz\Big)\Big(\big(x^{\underline{k}} - x^{\overline{k}+1}\big)(y-1)+x\Big)}. \tag{319}$$

Although finding an explicit expression is impracticable, one can use a computer algebra system to obtain $s_{\underline{k}\leq\overline{k}}(n,m,r) = [x^n y^m z^r] S_{\underline{k}\leq\overline{k}}(x,y,z)$ from (**??**).

7.2 Number of n-Strings of Hamming Weight r that Contain Exactly m Nonnull Runs of Ones and/or Zeros

We call $s(n,m,r)$ the number of n-strings of Hamming weight r that contain exactly m nonnull runs of ones and/or zeros (of arbitrary lengths, all strictly greater than zero). This enumeration —in fact, the probability $\lambda(n,m,r) = s(n,m,r)/\binom{n}{r}$— was first given by Stevens [**?**]. However, its relevance in the theory of runs is mainly due to Wald and Wolfowitz, who independently derived the same result one year later [**?**] to use it as the basis of their runs-based hypothesis test for the identity between the distributions from which two samples are drawn.

The probability $\lambda(n,m,r)$ was also rederived later in alternative ways by several other authors, such as Gibbons [**?**] and Schuster [**?**]. The latter author, who denotes the probability $\lambda(n,m,r)$ by R, comments [**?**, Rem. 2]): "(...) *much of the difficulty in studying the theoretical properties of R is due to the fact that R is not symmetrically defined* (...)". This difficulty, which was in fact the stumbling block preventing us from producing a general recurrence in Section **??**, also rears its ugly head in our approach to this particular enumeration, although in a less severe manner.

As in previous cases, $s(n,m,r)$ is a special case of our general analysis in Section **??** with $\underline{k}=1$ and $\overline{k}=n$, and therefore

$$s(n,m,r) = s_{1\leq n}(n,m,r). \tag{320}$$

7.2.1 Recurrence

In order to delve deeper into this special enumeration, we start by specialising recurrence (**??**) to get

$$s^b(n,m,r) = \sum_{i=1}^{n} s^{\tilde{b}}\big(n-i, m-1, r - [\![b=1]\!]\, i\big). \tag{321}$$

We certainly face the same issue here as in the general case: the two mutual recurrences defined by (**??**) are not symmetric. However, in this particular case we may symmetrise them, which allows us to produce a recurrence for $s(n,m,r)$. To this end, we fist make (**??**) valid for all values of the parameters by adding $[\![n=0]\!][\![m=0]\!][\![r=0]\!]$ to it. We then apply this extended recurrence to itself, which yields

$$\begin{aligned} s^b(n,m,r) = \sum_{i=1}^{n} \Bigg(& \sum_{j=1}^{n} s^b\big(n-(i+j), m-2, r - [\![b=1]\!]\, i - [\![b=0]\!]\, j\big) \\ & + [\![n=i]\!][\![m=1]\!]\big[\!\big[r = [\![b=1]\!]\, i\big]\!\big] \Bigg) \\ & + [\![n=0]\!][\![m=0]\!][\![r=0]\!], \end{aligned} \tag{322}$$

where we have extended the summation on j from $n-i$ to n to make the symmetry due to the summations on i and j clear. This does not alter the recurrence, as from necessary condition (**??**) we have that $s^b(n,m,r) = 0$ for $n < 0$.

We can now directly obtain $s(n,m,r) = s^0(n,m,r) + s^1(n,m,r)$ using (**??**), which yields the following recurrence:

$$\begin{aligned} s(n,m,r) = & \sum_{i=1}^{n} \sum_{j=1}^{n-i} s\big(n-(i+j), m-2, r-i\big) \\ & + [\![n>0]\!][\![m=1]\!]\big([\![r=n]\!] + [\![r=0]\!]\big) + 2[\![n=0]\!][\![m=0]\!][\![r=0]\!]. \end{aligned} \tag{323}$$

Observe that, after the addition of the two mutual recurrences, we have set the upper limit of the summation on j back to the more efficient $n-i$.

7.2.2 Generating Function

One way to work out the ogf $S(x,y,z) = \sum_{n,m,r} s(n,m,r)\, x^n y^m z^r$ is simply to specialise (**??**), but as in other similar cases this gives an n-dependent ogf and is anyway unwieldy. As we see next, we can get a much simpler and general expression. While we have the option of finding an ogf through a version of recurrence (**??**) without n-dependent summations, a gentler approach is to produce instead ogfs from the recurrences (**??**) specialised to this case, i.e., $S^b(x,y,z) = \sum_{n,m,r} s^b(n,m,r)\, x^n y^m z^r$, and then obtain $S(x,y,z) = S^0(x,y,z) + S^1(x,y,z)$ like we did in Section **??**. The specialisation of (**??**) yields

$$\begin{aligned} s^b(n,m,r) &= s^b\big(n-1,m,r-[\![b=1]\!]\big) + s^{\tilde{b}}\big(n-1,m-1,r-[\![b=1]\!]\big) \\ &\quad + [\![m=0]\!]\big([\![r=0]\!][\![n=0]\!] - [\![r=b]\!][\![n=1]\!]\big). \end{aligned} \tag{324}$$

This recurrence is valid already for all values of the parameters. Thus, multiplying it on both sides by $x^n y^m z^r$ and then adding over n, m and r we get

$$S^b(x,y,z)(1 - xz^{[\![b=1]\!]}) = S^{\tilde{b}}(x,y,z)\, xyz^{[\![b=1]\!]} + (1 - xz^b). \tag{325}$$

Again, we have a linear system of two equations with two unknowns, i.e., $S^0(x,y,z)$ and $S^1(x,y,z)$. Solving it we find that

$$S(x,y,z) = \frac{2 - 2x + xy - 2xz + 2x^2z + xyz - 2x^2yz}{1 - x - xz + x^2z - x^2y^2z}. \tag{326}$$

7.2.3 Explicit Expressions

We obtain next $s(n,m,r) = [x^n y^m z^r]S(x,y,z)$ from (**??**). To simplify the procedure it is convenient to rewrite (**??**) as follows:

$$S(x,y,z) = \frac{2 + \big(-1 + (1-x^2z)(1-xz)^{-1}(1-x)^{-1}\big)y}{1 - x^2z(1-xz)^{-1}(1-x)^{-1}y^2}. \tag{327}$$

By applying the negative binomial theorem, we can express (**??**) as

$$S(x,y,z) = \left(2 + \left(-1 + \frac{1-x^2z}{(1-xz)(1-x)}\right)y\right) \sum_{p\geq 0} \left(\frac{x^2z}{(1-xz)(1-x)}\right)^p y^{2p}. \tag{328}$$

Thus, defining

$$\sigma_{\{a,b\}}(x,z) = \frac{(x^2z)^a}{\Big((1-xz)(1-x)\Big)^b}, \tag{329}$$

we have that the coefficient of y^m for m even is

$$[y^m]S(x,y,z) = 2\,\sigma_{\left\{\frac{m}{2},\frac{m}{2}\right\}}(x,z), \tag{330}$$

whereas for m odd

$$[y^m]S(x,y,z) = -\sigma_{\left\{\frac{m-1}{2},\frac{m-1}{2}\right\}}(x,z) + \sigma_{\left\{\frac{m-1}{2},\frac{m+1}{2}\right\}}(x,z) - \sigma_{\left\{\frac{m+1}{2},\frac{m+1}{2}\right\}}(x,z). \tag{331}$$

On the other hand, we can expand (**??**) by applying the negative binomial theorem twice to get

$$\sigma_{\{a,b\}}(x,z) = \sum_{s\geq 0}\sum_{t\geq 0}\binom{s+b-1}{b-1}\binom{t+b-1}{b-1}x^{2a+s+t}z^{a+s}. \tag{332}$$

In order to determine $[x^n z^r]\sigma_{\{a,b\}}(x,z)$ we need to find the indices s and t that fulfil $n = 2a+s+t$ and $r = a+s$. From the second equation we see that $s = r-a$, and thus $t = n-r-a$. Therefore, from these solutions and (**??**) we have that

$$[x^n z^r]\sigma_{\{a,b\}}(x,z) = \binom{r-a+b-1}{b-1}\binom{n-r-a+b-1}{b-1}. \tag{333}$$

So, from (**??**) and (**??**) we obtain the following expression for m even:

$$s(n,m,r) = 2\binom{r-1}{\frac{m}{2}-1}\binom{n-r-1}{\frac{m}{2}-1}. \tag{334}$$

From (**??**) and (**??**) we have that, for m odd, the expression is

$$s(n,m,r) = -\binom{r-1}{\frac{m-1}{2}-1}\binom{n-r-1}{\frac{m-1}{2}-1} + \binom{r}{\frac{m-1}{2}}\binom{n-r}{\frac{m-1}{2}} - \binom{r-1}{\frac{m-1}{2}}\binom{n-r-1}{\frac{m-1}{2}}. \tag{335}$$

Applying $\binom{a}{b} = \binom{a-1}{b} + \binom{a-1}{b-1}$ to both binomial coefficients in the positive term in (**??**) we see that the negative terms cancel out, which leads to a simpler expression for m odd:

$$s(n,m,r) = \binom{r-1}{\frac{m-1}{2}-1}\binom{n-r-1}{\frac{m-1}{2}} + \binom{r-1}{\frac{m-1}{2}}\binom{n-r-1}{\frac{m-1}{2}-1}. \tag{336}$$

In the form $\lambda(n,m,r) = s(n,m,r)/\binom{n}{r}$, the explicit expressions (**??**) and (**??**) are well-known since the works of Stevens [**?**, Eqs. (3.31) and (3.32)] and Wald and Wolfowitz [**?**, Eqs. (7) and (8)]. See also the rederivations by Gibbons [**?**, Thm. 3.2.2] and Schuster [**?**, Cor. 3.7], and the alternative moment computations by Guenther [**?**].

Remark. Like in Section **??**, we can invoke the law of large numbers to write $\lambda(n,m) \approx \lambda(n,m,\lfloor nq\rfloor) = s(n,m,\lfloor nq\rfloor)/\binom{n}{\lfloor nq\rfloor}$ for large n, where $\lambda(n,m)$ is the probability discussed in Section **??**. Observe that both in $s(n,m,r)$ and in $\lambda(n,m)$ there is a dichotomy between two different explicit expressions depending on the parity of m. For example, if we let $\widetilde{M}_{n,r}$ be the rv with distribution $\Pr(\widetilde{M}_{n,r} = m) = s(n,m,r)/\binom{n}{r}$, then we can approximate its expectation using (**??**) and $q = r/n$ to see that, for large n, $\mathrm{E}(\widetilde{M}_{n,r}) \approx 1 + 2(n-r)r(n-1)/n^2 \approx 1 + 2(n-r)r/n$, an expression that is well known [**?**, Eq. (12)].

This connection between $\lambda(n,m)$ and $\lambda(n,m,r)$ implies that Wishart and Hirschfeld's work [**?**] somehow predated that of Wald and Wolfowitz [**?**]: Wald and Wolfowitz's test could, in principle, have been based on the earlier results in [**?**] for large n. □

7.3 Number of n-Strings of Hamming Weight r Whose Longest Nonnull Run Is a $(\leq k)$-Run of Ones or Zeros

We denote by $\bar{s}_{\leq k}(n,r)$ the number of n-strings of Hamming weight r whose longest nonnull run is a $(\leq k)$-run of ones or zeros, where $k \geq 1$. Bateman [?] gave an explicit expression for the complementary probability of $\overline{\lambda}_{\leq k}(n,r) = \bar{s}_{\leq k}(n,r)/\binom{n}{r}$, and tables for $\overline{\lambda}_k(n,r) = \overline{\lambda}_{\leq k}(n,r) - \overline{\lambda}_{\leq (k-1)}(n,r)$ for small values of n, r and k. Schuster [?], Bloom [?], and Jackson (see [?]) gave recurrences for $\bar{s}_{\leq k}(n,r)$. Schuster also extended and corrected Bateman's tables.

We can study this problem as a special case of the results in Section **??**, because

$$\bar{s}_{\leq k}(n,r) = s_{(k+1)\leq n}(n,0,r). \tag{337}$$

7.3.1 Recurrence

Let us see how we can deduce a recurrence for $\bar{s}_{\leq k}(n,r)$ from the general expressions. Calling $\bar{s}^{\,b}_{\leq k}(n,r) = s^{b}_{(k+1)\leq n}(n,0,r)$, we can express the quantity for which we want a recurrence as

$$\bar{s}_{\leq k}(n,r) = \bar{s}^{\,0}_{\leq k}(n,r) + \bar{s}^{\,1}_{\leq k}(n,r). \tag{338}$$

With the parameters of this special case (**??**) becomes

$$\bar{s}^{\,b}_{\leq k}(n,r) = \sum_{i=1}^{k} \bar{s}^{\,\tilde{b}}_{\leq k}\big(n-i, r - [\![b=1]\!]\, i\big). \tag{339}$$

Again, the main obstacle to finding a recurrence for $\bar{s}_{\leq k}(n,r)$ is the asymmetry between the two mutual recurrences in (**??**). Nevertheless, like in Section **??**, in this case it is also possible to symmetrise them in a straightforward manner. We start by making (**??**) valid for all n and r by adding $[\![n=0]\!][\![r=0]\!]$ to it. We then apply this extended recurrence to itself, which gives

$$\begin{aligned}\bar{s}^{\,b}_{\leq k}(n,r) = {}& \sum_{i=1}^{k}\left(\sum_{j=1}^{k} \bar{s}^{\,b}_{\leq k}\big(n-(i+j), r-[\![b=1]\!]\, i - [\![b=0]\!]\, j\big) + [\![n=i]\!]\,[\![r=[\![b=1]\!]\, i]\!]\right)\\ &+ [\![n=0]\!][\![r=0]\!].\end{aligned} \tag{340}$$

We can now input (**??**) into (**??**) to obtain the following recursive relation:

$$\begin{aligned}\bar{s}_{\leq k}(n,r) = {}& \sum_{i=1}^{k}\left(\sum_{j=1}^{k} \bar{s}_{\leq k}\big(n-(i+j), r-i\big) + [\![n=i]\!]\big([\![r=i]\!] + [\![r=0]\!]\big)\right)\\ &+ 2[\![n=0]\!][\![r=0]\!].\end{aligned} \tag{341}$$

Taking (**??**) into account, (**??**) is valid for all n and r, and so it does not need initialisation.

Remark. At first sight, (**??**) may look unrelated to the recurrence for the same enumeration first cleverly surmised and then proved by Bloom [**?**, Eq. (9)] —in his notation, $\bar{s}_{\le k}(n,r) = C_{k+1}(n-r,r)$. However, if we obtain $\bar{s}_{\le k}(n,r) - \bar{s}_{\le k}(n-1,r)$ using (**??**), then we essentially recover Bloom's recurrence:

$$\begin{aligned}\bar{s}_{\le k}(n,r) = &\sum_{i=0}^{k} \bar{s}_{\le k}\big(n-(i+1), r-i\big) - \sum_{i=1}^{k} \bar{s}_{\le k}\big(n-(i+k+1), r-i\big)\\ &+ \sum_{i=1}^{k} \big([\![n=i]\!] - [\![n=i+1]\!]\big)\big([\![r=i]\!] + [\![r=0]\!]\big)\\ &+ 2\big([\![n=0]\!] - [\![n=1]\!]\big)[\![r=0]\!]. \end{aligned} \tag{342}$$

In Bloom's recurrence, index i only affects the second argument of $C_t(\cdot,\cdot)$ whereas above it affects both arguments of $\bar{s}_{\le k}(\cdot,\cdot)$. This just a notational artifact: in Bloom's notation the addition of the two arguments yields the length of the binary string, and thus the homogeneous part of both recurrences is exactly the same. The inhomogeneous term in Bloom's recurrence does not always equal the inhomogeneous term in (**??**), which simply indicates a different initialisation strategy:

both recurrences return the same values.

Finally, if we obtain $\bar{s}_{\le k}(n,r) - \bar{s}_{\le k}(n-1,r-1)$ using (**??**), we also essentially recover Jackson's recurrence [**?**, Eq. (13)]:

$$\begin{aligned}\bar{s}_{\le k}(n,r) = &\,\bar{s}_{\le k}\big(n-1,r-1\big) + \bar{s}_{\le k}\big(n-1,r\big) - \bar{s}_{\ge k}\big(n-(k+2), r-(k+1)\big)\\ &- \bar{s}_{\le k}\big(n-(k+2), r-1\big) + \bar{s}_{\le k}\big(n-2(k+1), r-(k+1)\big)\\ &+ \sum_{i=1}^{k} \big([\![n=i]\!] - [\![n=i+1]\!]\big)\big([\![r=i]\!] + [\![r=0]\!]\big)\\ &- \sum_{i=1}^{k} \big([\![n=i+1]\!] - [\![n=i+2]\!]\big)\big([\![r=i+1]\!] + [\![r=1]\!]\big)\\ &+ 2\big([\![n=0]\!] - [\![n=1]\!]\big)[\![r=0]\!] - 2\big([\![n=1]\!] - [\![n=2]\!]\big)[\![r=1]\!]. \end{aligned} \tag{343}$$

Again the homogeneous parts of (**??**) and [**?**, Eq. (13)] are identical, but the inhomogeneous parts are different. At any rate, both recurrences deliver the same numerical values. The number of recurrent calls made by (**??**), (**??**) and (**??**) is k^2, $2k+1$ and 5, respectively. Thus, (**??**) is the most efficient recurrence when $k > 2$, only improved by (**??**) when $k = 1$ or $k = 2$.

To conclude, we should mention that Schuster was the first author who found a recurrence for $\bar{s}_{\le k}(n,r)$ —in fact, for $\overline{\lambda}_{\le k}(n,r) = \bar{s}_{\le k}(n,r)/\binom{n}{r}$ [**?**, Cor. 5.4]. His recurrence depends on an involved ancillary function, and it does not seem to bear a close relationship to the recurrences discussed above. □

7.3.2 Generating Functions

We now obtain $\overline{S}_{\le k}(x,z)=\sum_{n,r}\bar{s}_{\le k}(n,r)\,x^n z^r$. We can do so using (**??**), (**??**) or (**??**) without further ado, since none of these equivalent recurrences contains n-dependent summations and all of them are valid for all arguments. Let us use (**??**), as it involves the least amount of algebra. Multiplying this recurrence on both sides by $x^n z^r$ and then adding over n and r we can see that

$$\overline{S}_{\le k}(x,z)=\frac{2-x(1+z)+x^{k+2}(z+z^{k+1})-x^{k+1}(1+z^{k+1})}{1-x(1+z)+x^{k+2}(z+z^{k+1})-x^{2(k+1)}z^{k+1}}. \tag{344}$$

The reader may verify that the same ogf is obtained from recurrences (**??**) or (**??**).

Using instead Jackson's original recurrence [**?**, Eq. (13)] we get a somewhat simpler ogf:

$$\overline{S}_{\le k}(x,z)=\frac{1-x^{k+1}(1+z^{k+1})+x^{2(k+1)}z^{k+1}}{1-x(1+z)+x^{k+2}(z+z^{k+1})-x^{2(k+1)}z^{k+1}}. \tag{345}$$

Using any of these two ogfs, we may obtain $\bar{s}_{\le k}(n,r)=[x^n z^r]\overline{S}_{\le k}(x,z)$ using a computer algebra system.

7.3.3 Explicit Expression

In order to extract the coefficient of $x^n z^r$ from $\overline{S}_{\le k}(x,z)$ we define

$$a^{(u,v,w)}(x,z)=\frac{x^v z^w\left(1-x-xz+x^{k+2}z+x^{k+2}z^{k+1}\right)^{-u}}{1-\left(x^{2(k+1)}z^{k+1}\right)\left(1-x-xz+x^{k+2}z+x^{k+2}z^{k+1}\right)^{-1}}. \tag{346}$$

which we use to rewrite (**??**) as

$$\overline{S}_{\le k}(x,z)=a^{(0,0,0)}(x,z)+a^{(1,0,0)}(x,z)-a^{(1,k+1,0)}(x,z)-a^{(1,k+1,k+1)}(x,z). \tag{347}$$

Applying the negative binomial theorem (twice) and then the binomial theorem (three times), we can expand (**??**) as

$$a^{(u,v,w)}(x,z)=\sum_{t,p,i,j,l\ge 0}(-1)^i\binom{p+t+u-1}{p}\binom{p}{i}\binom{p-i}{j}\binom{i}{l}x^{2(k+1)t+p+(k+1)i+v}\,z^{(k+1)t+i+j+lk+w}. \tag{348}$$

Thus, the coefficient of $x^n z^r$ in this expression is obtained by determining the nonnegative indices t,p,i,j, and l that fulfil $n=2(k+1)t+p+(k+1)i+v$ and $r=(k+1)t+i+j+lk+w$. This implies that $t\le\min(\lfloor(n-v)/(2(k+1))\rfloor,\lfloor(r-w)/(k+1)\rfloor)$, $l\le\lfloor(r-w-(k+1)t)/k\rfloor$ and $i\le\lfloor(n-2(k+1)t-v)/(k+1)\rfloor$, whereas, for any (t,l,i) triple, j is determined by the second equation given and p by the first one. Hence, we have

$$[x^n z^r]a^{(u,v,w)}(x,z)=\sum_{t,l,i\ge 0}(-1)^i\binom{p+t+u-1}{p}\binom{p}{i}\binom{p-i}{j}\binom{i}{l}, \tag{349}$$

where the upper limits of the summations on t, l, and i are given by the three aforementioned inequalities, and p and j are determined using the two equations above. We do not need to check for nonnegativity of p and j, as if that were the case the corresponding terms in (**??**) would cancel out. Using (**??**) and (**??**), we have an explicit triple-summation expression for $\bar{s}_{\leq k}(n,r) = [x^n z^r]\overline{S}_{\leq k}(x,y)$, comparable to the triple-summation expression originally given by Bateman for $\overline{\lambda}_{\geq k}(n,r)$ [**?**, p. 100].

Remark. The ogfs (**??**) or (**??**), or the explicit expression in the previous section, allow us to address the same problem that motivated Bloom's work in [**?**]. Bloom's research was spurred by his desire to determine the accuracy of the following assertion by M. Gardner in [**?**, p. 124]: "*(...) a shuffled deck of cards will contain coincidences. For instance, almost always there will be a clump of six or seven cards of the same color.*" Gardner, like Bloom, uses the word "clump" to mean a run.

More specifically, what Bloom calls a "k-clump" is a run of length k or longer of either colour of the deck, and he wishes to determine the likelihood that a shuffled deck will contain a 6-clump.

If we just look at the colour of the cards, then there is a bijection between the shuffles of a standard deck and the binary strings of length 52 and Hamming weight 26 (where 0 means "red" and 1 means "black", or vice versa). The number of n-strings devoid of ($\geq$ 6)-runs of ones or zeros is $\bar{s}_{\leq 5}(52,26) = 265,692,662,174,100$, and, thus, the probability that a shuffled deck of cards drawn uniformly at random contains at least one 6-clump is $1 - \bar{s}_{\leq 5}(52,26)/\binom{52}{26} = 0.464241$.

This certainly coincides with Bloom's analysis, and it confirms this author's finding that Gardner's claim is hardly supported by the numbers. Bloom also used the expectation (**??**) in Section **??** as an indirect way to support his findings. However, as seems to have happened so many times in the history of the theory of runs, this author missed Bateman's results. Had he not missed them, he would simply have implemented the explicit expression given by Bateman [**?**, p. 100] to answer his question. But had he done so, we would have missed on his insightful paper.

Lastly, the tables for $\overline{\lambda}_{\leq k}(n,r) = \bar{s}_{\leq k}(n,r)/\binom{n}{r}$ given by Schuster [**?**, pp. 112–116] —in his notation, $\overline{\lambda}_{\leq k}(n,r) = F_M(k, n-r, r)$— can also be easily reproduced through Bateman's explicit expression (or, of course, through our own explicit expression or through the ogfs above). □

7.3.4 OEIS

Integer sequence $\binom{52}{26} - \bar{s}_{(k-1)}(52,26)$ is A086438. On the other hand, the numerators and denominators of the (simplified) fraction sequence $1 - \bar{s}_{(k-1)}(52,26)/\binom{52}{26}$ are, respectively, A086439 and A086440.

7.4 Number of n-Strings of Hamming Weight r that Contain Exactly m Nonnull p-Parity Runs of Ones and/or Zeros

To conclude this section, we enumerate the n-strings of Hamming weight r that exactly contain m nonnull p-parity runs of ones and/or zeros, which we denote by $s_{[p]}(n,m,r)$. The necessary condition that we require in this case, which is just the appropriate variation of (**??**), is given next.

Necessary Condition. (Existence of n-strings of Hamming weight r containing m nonnull p-parity runs of ones and/or zeros)

$$s_{[p]}(n,m,r) > 0 \quad \Longrightarrow \quad 0 \le m \le \min\left(\left\lfloor \frac{n}{1+[\![p=0]\!]} \right\rfloor, \left\lfloor \frac{r}{1+[\![p=0]\!]} \right\rfloor + \left\lfloor \frac{n-r}{1+[\![p=0]\!]} \right\rfloor\right). \tag{350}$$

□

7.4.1 Recurrence

As in all previous cases in Section **??**, we start by finding two mutual recurrences for this enumeration corresponding to the n-strings that start with bit $b \in \{0,1\}$, which we denote by $s^b_{[p]}(n,m,r)$. Through the standard procedure, these two mutual recurrences are given by

$$s^b_{[p]}(n,m,r) = \sum_{i=1}^{n} s^{\tilde{b}}_{[p]}\big(n-i, m-[\![p=\mathrm{mod}(i,2)]\!], r-[\![b=1]\!]\,i\big), \tag{351}$$

where the only difference with respect to (**??**) is the Hamming weight update when $b=1$.

When $n=1$ we know by inspection that

$$s^b_{[p]}(1,m,r) = [\![r=b]\!]\big([\![m=0]\!][\![p=0]\!] + [\![m=1]\!][\![p=1]\!]\big), \tag{352}$$

whereas setting $n=1$ in (**??**) yields

$$s^b_{[p]}(1,m,r) = s^{\tilde{b}}_{[p]}\big(0, m-[\![p=1]\!], r-[\![b=1]\!]\big). \tag{353}$$

We can verify that (**??**) equals (**??**) for

$$s^b_{[p]}(0,m,r) = [\![m=0]\!][\![r=0]\!], \tag{354}$$

which therefore constitute the initialisation of (**??**). Using these recurrences we can now compute

$$s_{[p]}(n,m,r) = s^0_{[p]}(n,m,r) + s^1_{[p]}(n,m,r). \tag{355}$$

Like in Sections **??** and **??**, it is also possible to obtain a recurrence on $s_{[p]}(n,m,r)$ by symmetrising (**??**). After adding $[\![n=0]\!][\![m=0]\!][\![r=0]\!]$ to (**??**) to make it valid for all values of the parameters, we may apply this recurrence to itself to get

$$\begin{aligned}s^b_{[p]}(n,m,r) = \sum_{i=1}^{n}\Bigg(&\sum_{j=1}^{n} s^b_{[p]}\big(n-(i+j), m-[\![p=\operatorname{mod}(i,2)]\!]-[\![p=\operatorname{mod}(j,2)]\!], r-[\![b=1]\!]i-[\![b=0]\!]j\big)\\ &+[\![n=i]\!]\,\big[\!\big[m=[\![p=\operatorname{mod}(i,2)]\!]\big]\!\big]\,\big[\!\big[r=[\![b=1]\!]\,i\big]\!\big]\Bigg)\\ &+[\![n=0]\!][\![m=0]\!][\![r=0]\!], \end{aligned} \tag{356}$$

where we have extended the summation on j from $n-i$ to n to highlight the symmetries of the expression with respect to i and j. This is possible because of necessary condition (**??**). Using next (**??**) in (**??**) we get the recurrence

$$\begin{aligned}s_{[p]}(n,m,r) =& \sum_{i=1}^{n}\sum_{j=1}^{n-i} s_{[p]}\big(n-(i+j), m-[\![p=\operatorname{mod}(i,2)]\!]-[\![p=\operatorname{mod}(j,2)]\!], r-i\big)\\ &+[\![n>0]\!]\,\big[\!\big[m=[\![p=\operatorname{mod}(n,2)]\!]\big]\!\big]\,\big([\![r=0]\!]+[\![r=n]\!]\big)+2[\![n=0]\!][\![m=0]\!][\![r=0]\!]. \end{aligned} \tag{357}$$

Like we did in (**??**), in the recurrence above we have again set the upper limit of the summation on j to $n-i$.

7.4.2 Generating Function

The simplest way to obtain the ogf $S_{[p]}(x,y,z)=\sum_{n,m,r} s_{[p]}(n,m,r)\,x^n y^m z^r$ is to proceed as in Sections **??** and **??**. That is to say, we first get the ogfs $S^b_{[p]}(x,y,z)=\sum_{n,m,r} s^b_{[p]}(n,m,r)$ in order to then obtain

$$S_{[p]}(x,y,z)=S^0_{[p]}(x,y,z)+S^1_{[p]}(x,y,z). \tag{358}$$

First, we obtain versions of (**??**) without n-dependent summations. We start by adding $[\![n=0]\!][\![m=0]\!][\![r=0]\!]$ to (**??**) to make it valid for all values of the parameters. We can then obtain $s^b(n,m,r)-s^b(n-2,m,r-2[\![b=1]\!])$ using this extended recurrence, which yields

$$\begin{aligned}s^b_{[p]}(n,m,r) =& s^b_{[p]}\big(n-2,m,r-2[\![b=1]\!]\big)+s^{\tilde b}_{[p]}\big(n-1,m-[\![p=1]\!],r-[\![b=1]\!]\big)\\ &+s^{\tilde b}_{[p]}\big(n-2,m-[\![p=0]\!],r-2[\![b=1]\!]\big)\\ &+[\![n=0]\!][\![m=0]\!][\![r=0]\!]-[\![n=2]\!][\![m=0]\!]\,\big[\!\big[r=2[\![b=1]\!]\big]\!\big]. \end{aligned} \tag{359}$$

Thus, multiplying this recurrence on both sides by $x^n y^m z^r$ and adding over n,m, and r we get the following system of two equations with two unknowns, i.e., $S^0_{[p]}(x,y,z)$ and $S^1_{[p]}(x,y,z)$:

$$S^b_{[p]}(x,y,z)\big(1-x^2z^{2[\![b=1]\!]}\big)=S^{\tilde b}_{[p]}(x,y,z)\big(xy^{[\![p=1]\!]}z^{[\![b=1]\!]}+x^2y^{[\![p=0]\!]}z^{2[\![b=1]\!]}\big)+1-x^2z^{2[\![b=1]\!]}. \tag{360}$$

Solving the system and using (??), it can be seen that

$$S_{[p]}(x,y,z) = \frac{2x^4z^2\big(y^{[\![p=0]\!]}-1\big)+x^3z(z+1)y^{[\![p=1]\!]}-x^2(z^2+1)\big(y^{[\![p=0]\!]}-2\big)-x(z+1)y^{[\![p=1]\!]}-2}{x^4z^2(y^{2[\![p=0]\!]}-1)+x^3z(z+1)y^{[\![p=0]\!]-[\![p=1]\!]}+x^2(zy^{2[\![p=1]\!]}+z^2+1)-1}. \tag{361}$$

8 Number of Runs Under Different Constraints Over All n-Strings and Over Restricted Subsets

Our main goal in this section is enumerating how many runs under different constraints are found over all n-strings. The most basic results are the enumerations of the $(\underline{k} \le \overline{k})$-runs over all n-strings in Section **??** and over all n-strings with Hamming weight r in Section **??**, which we then specialise to several particular cases. In Section **??** we study the number of runs over the n-strings enumerated in Sections **??** and **??**. These same problems are revisited in Section **??**, but with the goal of determining the number of runs of ones *and zeros* instead. Several of the questions addressed in this section, or variations of them, were previously studied by Marbe [**?**], Gold [**?**], Cochran [**?**], Bloom [**?**], Sinha and Sinha [**?**, **?**], Makri et al. [**?**, **?**], Nyblom [**?**], and Grimaldi and Heubach [**?**] —and perhaps by other authors we do not know of, since these are fairly common problems. In fact, many of the authors just mentioned were not aware of previous results.

Since runs of zeros appear at some points throughout this section, the reader is again reminded that, unless explicitly said otherwise, the term "run" without any qualifier refers to a run of ones.

8.1 Number of $(\underline{k} \le \overline{k})$-Runs Over All n-Strings

We denote the number of $(\underline{k} \le \overline{k})$-runs over all n-strings by $\rho_{\underline{k}\le\overline{k}}(n)$. Considering (**??**), it is possible to compute $\rho_{\underline{k}\le\overline{k}}(n)$ by relying on the enumeration $w_{\underline{k}\le\overline{k}}(n,m)$ studied in Section **??** as follows:

$$\rho_{\underline{k}\le\overline{k}}(n) = \sum_{m=1}^{\left\lfloor \frac{n+1}{\underline{k}+1} \right\rfloor} m\, w_{\underline{k}\le\overline{k}}(n,m). \tag{362}$$

This summation requires an explicit expression for $w_{\underline{k}\le\overline{k}}(n,m)$, which we have not provided. However, it is also possible to obtain $\rho_{\underline{k}\le\overline{k}}(n)$ directly from the ogf (**??**):

$$\rho_{\underline{k}\le\overline{k}}(n) = [x^n]\left.\frac{\partial W_{\underline{k}\le\overline{k}}(x,y)}{\partial y}\right|_{y=1}. \tag{363}$$

See that this is essentially the same as the computation of the first moment with a pgf —cf. Section **??**. From (**??**) we have that

$$\left.\frac{\partial W_{\underline{k}\leq\overline{k}}(x,y)}{\partial y}\right|_{y=1} = \frac{(1-x)\big(x^{\underline{k}} - x^{\overline{k}+1}\big)}{(1-2x)^2}. \tag{364}$$

By applying the negative binomial theorem, we can expand (**??**) as follows:

$$\left.\frac{\partial W_{\underline{k}\leq\overline{k}}(x,y)}{\partial y}\right|_{y=1} = \big(x^{\underline{k}} - x^{\underline{k}+1} - x^{\overline{k}+1} + x^{\overline{k}+2}\big)\sum_{t\geq 0}\binom{t+1}{t}2^t x^t. \tag{365}$$

To extract the coefficient of x^n we have to solve four equations for the nonnegative index t: a) $n = \underline{k} + t$, b) $n = \underline{k} + 1 + t$, c) $n = \overline{k} + 1 + t$, and d) $n = \overline{k} + 2 + t$. When each of these equations has a solution, from (**??**) their respective contributions to the coefficient of x^n are:

a) $(n - \underline{k} + 1)\, 2^{n-\underline{k}}$

b) $-(n - \underline{k})\, 2^{n-(\underline{k}+1)}$

c) $-(n - \overline{k})\, 2^{n-(\overline{k}+1)}$

d) $(n - \overline{k} - 1)\, 2^{n-(\overline{k}+2)}$

- If $\overline{k} + 2 \leq n$, as $\underline{k} \leq \overline{k}$ there is a solution in each of the four cases, so all four contributions must be added.
- If $\overline{k} + 1 = n$ only the first three equations have a solution, but all four contributions can still be added because the fourth one is zero anyway.
- If $\overline{k} \geq n$ then the last two equations do not have a solution, and we have the following cases for the first two:
  - If $\underline{k} > n$ no equation has a solution.
  - If $\underline{k} < n$ the first two equations have a solution, so the first two contributions must be added.
  - If $\underline{k} = n$ only the first equation has a solution, but the first two contributions can still be added because the second one is zero anyway.

Collecting all these contributions we finally can see that

$$\rho_{\underline{k}\leq\overline{k}}(n) = (n - \underline{k} + 2)\, 2^{n-(\underline{k}+1)}\, [\![\underline{k} \leq n]\!] - (n - \overline{k} + 1)\, 2^{n-(\overline{k}+2)}\, [\![\overline{k} < n]\!]. \tag{366}$$

Observe that (**??**) is valid for all $0 \leq \underline{k} \leq \bar{k}$ and $n \geq 1$ —for example, it can enumerate the null runs over all n-strings if desired.

Although we will not show it here, (**??**) can also be obtained by establishing a recurrence for $\rho_{\underline{k} \leq \bar{k}}(n)$ from first principles (i.e, without resorting to any of the results in previous sections), and then solving it directly by unrolling it. In any case, the approach that we have given above is quicker and cleaner.

8.1.1 Special Cases

In this section we discuss several specialisations of $\rho_{\underline{k} \leq \bar{k}}(n)$.

We denote the number of k-runs over all n-strings by $\rho_k(n)$, and the number of nonnull runs (of arbitrary lengths, all strictly greater than zero) over all n-strings by $\rho(n)$. These enumerations are special cases of our analysis in the previous section, as we can write $\rho_k(n) = \rho_{k \leq k}(n)$ and $\rho(n) = \rho_{1 \leq n}(n)$. Thus, using (**??**) we have that

$$\rho_k(n) = (n - k + 3)\, 2^{n-(k+2)}\, [\![k < n]\!] + [\![k = n]\!], \tag{367}$$

and

$$\rho(n) = (n + 1)\, 2^{n-2}. \tag{368}$$

Next, we denote the number of $(\geq k)$-runs and of nonnull $(\leq k)$-runs over all n-strings by $\rho_{\geq k}(n)$ and by $\rho_{\leq k}(n)$, respectively. Once again, these are special cases of our analysis in the previous section, because we can write $\rho_{\geq k}(n) = \rho_{k \leq n}(n)$ and $\rho_{\leq k}(n) = \rho_{1 \leq k}(n)$. Therefore, from (**??**) we have

$$\rho_{\geq k}(n) = (n - k + 2)\, 2^{n-(k+1)}\, [\![k \leq n]\!] \tag{369}$$

and

$$\rho_{\leq k}(n) = \rho(n) - (n - k + 1)\, 2^{n-(k+2)}\, [\![k < n]\!], \tag{370}$$

where $\rho(n)$ is given by (**??**).

Remark. Leaving aside Marbe's formula for a moment —see below— the first author who produced (**??**) was Gold [**?**, **?**], in his analysis of the predictability of two-state meteorological series. This author gave $\rho_k(n)/2^n$ as the expected number of k-runs in an n-string drawn uniformly at random —he actually obtained twice the value of this expectation, as he addressed runs of ones and zeros jointly. Gold arrived at his result by combining $\sum_{k=1}^{n} k\, \rho_k(n) = n2^{n-1}$ (see Remark **??**) and $\rho_{k-1}(n-1) = \rho_k(n)$, an identity which can easily be argued. Gold's expectation was then extended by Cochran [**?**, Eq. (5)] to the case where the bits are not equally likely —Cochran, who was studying the spread of diseases in plants arranged in rows, acknowledges Marbe's priority regarding his result [**?**], but he indicates that the original proof [**?**, p. 9] omits essential steps, and gives his own proof by induction. Cochran's expectation in [**?**, Eq. (5)] is in fact (**??**) in Section **??**, which simplifies to (**??**) using $q = 1/2$ and multiplying by 2^n.

We briefly discuss other approaches for the derivation of some of these enumerations, as we retrace the work of other authors who previously dealt with them. These authors were unaware of the aforementioned results. As indicated by Sinha and Sinha [?] and by Makri and Psillakis [?, Eq. (9)], one way to obtain $\rho_k(n)$ is

$$\rho_k(n) = \sum_{m=1}^{\left\lfloor \frac{n+1}{k+1} \right\rfloor} m\, w_k(n,m), \tag{371}$$

which is the specialisation of (??). However, this is not an easy approach —consider using (??) in (??). Tellingly, Makri et al. were only able to evaluate an expression similar to (??) to obtain (??) [?, Eq. (7), $a = e$] by using an earlier probabilistic result of theirs, but not by using their own explicit expression for $w_k(n,m)$ [?, Eq. (8)] in (??).

As mentioned at the end of the previous section, it is possible to establish a recurrence for $\rho_{\underline{k}\le\bar{k}}(n)$ *ab initio*. In the special case $\underline{k} = \bar{k} = k$, this recurrence is

$$\rho_k(n) = 2\,\rho_k(n-1) + 2^{n-(k+2)} \tag{372}$$

for $k+2 \le n$, whereas $\rho_{n-1}(n) = 2$ and $\rho_n(n) = 1$. Recurrence (??) and explicit expression (??) were worked out from first principles by Sinha and Sinha in [?] —according to [?], an earlier appearance of these authors' results regarding $\rho_k(n)$ is in K. Sinha's PhD thesis [?]. Sinha and Sinha cite [?] in passing in [?], without realising that Cochran's article already contains the closed-form solution (??) they find. Makri et al. [?, Eq. (9), $a = e$] also noticed recurrence (??) after finding the explicit expression (??) for $\rho_k(n)$. We would also like to mention that we gave two other *ab initio* approaches to the problem of finding $\rho_k(n)$ in [?].

We should mention that, after determining the closed-form expression (??) for $\rho_k(n)$ through any of the methods just mentioned, one can then obtain $\rho_{\underline{k}\le\bar{k}}(n) = \sum_{k=\underline{k}}^{\bar{k}} \rho_k(n)$ —which only involves summations of the form $\sum_k a^k$ and $\sum_k k\, a^k$. While this makes the approach in the previous section unnecessary, we believe that the method that we have adopted is interesting because it connects with the enumerations in Section ??. It also allows for a consistent methodology throughout the paper —cf. Sections ??, ??, ??, ??, ??, ?? and ??.

As regards $\rho(n)$, this enumeration may be obtained in a number of different ways as well. One possibility is to use $\rho_k(n)$ to compute $\rho(n) = \sum_{k=1}^{n} \rho_k(n)$, as pointed out by Sinha and Sinha [?]. Using (??) we can thus write

$$\rho(n) = 1 + \sum_{k=1}^{n-1} (n-k+3)\, 2^{n-(k+2)}, \tag{373}$$

which, of course, evaluates to (??). But we may also get $\rho(n)$ without resorting to $\rho_k(n)$. One way is by specialising (??) with $\underline{k} = 1$ and $\bar{k} = n$ to get $\rho(n) = \sum_{m=1}^{\lfloor (n+1)/2 \rfloor} m\, w(n,m)$, which allows us to obtain this enumeration by using (??):

$$\rho(n) = \sum_{m=1}^{\lfloor \frac{n+1}{2} \rfloor} m \binom{n+1}{2m}.$$

This summation yields (??), a fact that one may verify using generating functions. Similarly, we can get $\rho(n)$ from the number of n-strings with exactly m nonnull runs of ones and/or zeros, $s(n,m)$. By symmetry, $\rho(n) = (1/2)\sum_{m=1}^{n} m\, s(n,m)$. Inputting (??) in this expression we get

$$\rho(n) = \sum_{m=1}^{n} m \binom{n-1}{m-1}, \tag{374}$$

and we may verify that this summation also gives (??). Lastly, Nyblom [?, Lem. 3.1] also found a recurrence for $\rho(n)$ which, of course, can also be derived from (??), and solved it to get (??) —in actual fact, Nyblom's results are for $2\,\rho(n)$, as this author counts both runs of ones and of zeros.

Finally, we make a few comments about $\rho_{\geq k}(n)$. Expression (??) can also be obtained from expectation (??) in Section ??, by setting $q = 1/2$ and then multiplying by 2^n. Bloom gave (??) divided by 2^{n-1} —i.e., the expected number of ($\geq k$)-runs of ones and/or zeros in an n-string drawn uniformly at random [?, Eq. (12)]. He most likely also used a probabilistic approach, but he gives no details. Another way to obtain $\rho_{\geq k}(n)$ is by specialising (??) with $\underline{k} = k$ and $\overline{k} = n$, which yields $\rho_{\geq k}(n) = \sum_{m=1}^{\lfloor (n+1)(k+1) \rfloor} m\, w_{\geq k}(n,m)$. Although it is again very difficult to handle this summation using an explicit expression for $w_{\geq k}(n,m)$, such as (??), a probabilistic version of this expression was used by Makri et al. to obtain (??) [?, Eq. (7), $a = g$]. These authors also deduced a recurrence working backwards from the explicit expression [?, Eq. (9), $a = g$].

□

Remark. There are some immediate connections between the results in this section and compositions. For example, the total number of parts in all compositions of n is the same as the total number of nonnull runs over all n-strings, $\rho(n)$, and thus given by (??). In similar fashion, the total number of parts equal to $k \geq 1$ in all compositions of n equals the total number of runs of length k over all n-strings, $\rho_k(n)$, which is given by (??). In the special case $k = 1$, both recurrence (??) and closed-form expression (??) were given by Chin et al. [?, Eqs. (1) and (2)]. These authors studied a problem concerning Cuisenaire rods (number rods) which is equivalent to enumerating all parts equal to 1 over all compositions of n. □

8.1.2 OEIS

We report next sequences connected to the enumerations in this section that are found in the OEIS:

- All sequences emanating from $\rho_k(n)$, $\rho_{\geq k}(n)$, and $\rho(n)$ are in the OEIS:
  $\rho_k(n+k)$ is A045623 (number of 1's in all compositions of $n+1$).
  $\rho_{\geq k}(n+k)$ and $\rho(n+1)$ are A001792.
- Sequences emanating from $\rho_{\leq k}(n)$:
  $\rho_{\leq 1}(n+1)$ is A045623 —just like $\rho_k(n+k)$.
  $\rho_{\leq 2}(n+1)$ is A106472.
  $\rho_{\leq 3}(n)$ is A386878.

8.2 Number of ($\underline{k} \leq \overline{k}$)-Runs Over All n-Strings of Hamming Weight r

In this section we obtain the total number of $(\underline{k} \leq \overline{k})$-runs over all n-strings of Hamming weight r, which we denote by $\rho_{\underline{k}\leq\overline{k}}(n,r)$. We can obtain this quantity directly from (**??**) as follows:

$$\rho_{\underline{k}\leq\overline{k}}(n,r) = [x^n z^r] \left. \frac{\partial W_{\underline{k}\leq\overline{k}}(x,y,z)}{\partial y} \right|_{y=1}. \tag{375}$$

From (**??**) we have that

$$\left. \frac{\partial W_{\underline{k}\leq\overline{k}}(x,y,z)}{\partial y} \right|_{y=1} = \frac{(1-xz)\big((xz)^{\underline{k}} - (xz)^{\overline{k}+1}\big)}{(1-x-xz)^2}. \tag{376}$$

By rewriting this expression as $((xz)^{\underline{k}} - (xz)^{\overline{k}+1})/\big((1-xz)(1-x/(1-xz))^2\big)$ and then applying the negative binomial theorem twice, we can expand it as follows:

$$\left. \frac{\partial W_{\underline{k}\leq\overline{k}}(x,y,z)}{\partial y} \right|_{y=1} = \big((xz)^{\underline{k}} - (xz)^{\overline{k}+1}\big) \sum_{t\geq 0}\sum_{s\geq 0} \binom{t+1}{t}\binom{s+t}{t} x^t (xz)^s. \tag{377}$$

To extract the coefficient of $x^n z^r$ we just have to find the nonnegative indices t and s that fulfil two sets of equations: on the one hand, $n = \underline{k}+t+s$ and $r = \underline{k}+s$; on the other hand, $n = \overline{k}+1+t+s$ and $r = \overline{k}+1+s$. The solutions of the first set are $t = n-r$ and $s = r-\underline{k}$, whereas the solutions of the second one are $t = n-r$ and $s = r-\overline{k}-1$. For a solution to exist, both sets require $n \geq r$, whereas $r \geq \underline{k}$ and $r \geq \overline{k}+1$ are required in the first and second set, respectively. From these considerations and (**??**) we thus have that

$$\rho_{\underline{k}\leq\overline{k}}(n,r) = (n-r+1)\left(\binom{n-\underline{k}}{r-\underline{k}} - \binom{n-\overline{k}-1}{r-\overline{k}-1}\right) [\![n \geq r]\!]. \tag{378}$$

Observe that the constraints on r are taken care of by the binomial coefficients. Like (**??**), this expression is valid for all $0 \leq \underline{k} \leq \overline{k}$ and $n \geq 1$, apart from all r.

8.2.1 Special Cases

We consider next several specialisations of $\rho_{\underline{k}\leq\bar{k}}(n,r)$. We denote the number of $(\geq k)$-runs and the total number of nonnull runs over all n-strings of Hamming weight r by $\rho_{\geq k}(n,r)$ and $\rho(n,r)$, respectively. These enumerations are special cases of (**??**), as $\rho_{\geq k}(n,r) = \rho_{k\leq n}(n,r)$ and $\rho(n,r) = \rho_{1\leq n}(n,r) = \rho_{\geq 1}(n,r)$. Thus, from (**??**) we see that

$$\rho_{\geq k}(n,r) = (n-r+1)\binom{n-k}{r-k}, \tag{379}$$

whereas

$$\rho(n,r) = (n-r+1)\binom{n-1}{r-1}. \tag{380}$$

Notice that we do not need Iverson brackets in these two expressions. Also, denote the number of nonnull $(\leq k)$-runs and the number of k-runs over all n-strings of Hamming weight r by $\rho_{\leq k}(n,r)$ and $\rho_k(n,r)$, respectively. As above, these enumerations are special cases of (**??**), because $\rho_{\leq k}(n,r) = \rho_{1\leq k}(n,r)$ and $\rho_k(n,r) = \rho_{k\leq k}(n,r) = \rho_{\geq 1}(n,r)$. Thus, from (**??**) we have that

$$\rho_{\leq k}(n,r) = (n-r+1)\left(\binom{n-1}{r-1} - \binom{n-k-1}{r-k-1}\right)[\![n \geq r]\!], \tag{381}$$

and

$$\rho_k(n,r) = (n-r+1)\binom{n-k-1}{r-k}[\![n \geq r]\!]. \tag{382}$$

Remark. If we denote by $M_{\geq k,n,r}$ the rv that models the number of $(\geq k)$-runs in an n-string of Hamming weight r drawn uniformly at random, then its expectation is $\mathrm{E}(M_{\geq k,n,r}) = \rho_{\geq k}(n,r)/\binom{n}{r}$. Similarly, if $\widetilde{M}_{\geq k,n,r}$ models the number of $(\geq k)$-runs of ones and/or zeros in an n-string of Hamming weight r drawn uniformly at random then

$$\begin{aligned}\mathrm{E}\big(\widetilde{M}_{\geq k,n,r}\big) &= \frac{1}{\binom{n}{r}}\big(\rho_{\geq k}(n,r) + \rho_{\geq k}(n,n-r)\big)\\ &= \frac{1}{\binom{n}{r}}\left((n-r+1)\binom{n-k}{n-r} + (r+1)\binom{n-k}{r}\right).\end{aligned} \tag{383}$$

Bloom used a probabilistic rationale to give the following expression for $\mathrm{E}\big(\widetilde{M}_{\geq k,n,r}\big)$ [**?**, p. 370]:

$$\mathrm{E}\big(\widetilde{M}_{\geq k,n,r}\big) = \frac{1}{\binom{n}{k}}\left((n-r+1)\binom{r}{k} + (r+1)\binom{n-r}{k}\right), \tag{384}$$

where we have divided by $k!$ both the numerator and the denominator of Bloom's original expression, for ease of comparison with (**??**). As $\binom{n-k}{n-r}\binom{n}{k} = \binom{r}{k}\binom{n}{r}$ and $\binom{n-k}{r}\binom{n}{k} = \binom{n-r}{k}\binom{n}{r}$ —which is verified simply by developing the factorials in all these binomial coefficients— then we can see that (**??**) is the same as Bloom's expression (**??**). □

8.3 Number of Nonnull Runs Over All n-Strings that Contain Exactly m Nonnull Runs Under Different Constraints

Next, we deal with the problems of enumerating the number of nonnull runs over all n-strings that contain exactly m nonnull $(\underline{k} \leq \bar{k})$-runs, or exactly m nonnull p-parity runs. In keeping with our notation conventions, we call these quantities $\rho_{\underline{k}\leq\bar{k}}(n,m)$ and $\rho_{[p]}(n,m)$, respectively.

These problems are not as elementary as the ones we have dealt with in the previous two sections. Relatively simple explicit expressions do not seem possible in general, although they are achievable in particular cases —see Section **??**. However, we can obtain recurrence relations and generating functions for these enumerations by relying on their counterparts in Sections **??** and **??**, i.e., $w_{\underline{k}\leq\bar{k}}(n,m)$ and $w_{[p]}(n,m)$, respectively.

8.3.1 Number of Nonnull Runs Over All n-Strings that Contain Exactly m Nonnull $(\underline{k} \leq \bar{k})$-Runs

We obtain first a recurrence for $\rho_{\underline{k}\leq\bar{k}}(n,m)$, by splitting the contributions due n-strings with different initial bit. As indicated, we assume $\underline{k} \geq 1$. The n-strings that start with 0 contribute $\rho_{\underline{k}\leq\bar{k}}(n-1,m)$, to $\rho_{\underline{k}\leq\bar{k}}(n,m)$. On the other hand, the n-strings that start with a nonnull i-run with $\underline{k} \leq i \leq \bar{k}$ contribute $\rho_{\underline{k}\leq\bar{k}}(n-(i+1),m-1)$ runs plus the $w_{\underline{k}\leq\bar{k}}(n-(i+1),m-1)$ initial runs themselves. If they start with a nonnull i-run with $i < \underline{k}$ or $i > \bar{k}$, then they contribute $\rho_{\underline{k}\leq\bar{k}}(n-(i+1),m)$ runs plus the $w_{\underline{k}\leq\bar{k}}(n-(i+1),m)$ initial runs. This yields the recurrence

$$\begin{aligned}\rho_{\underline{k}\leq\bar{k}}(n,m) = \rho_{\underline{k}\leq\bar{k}}(n-1,m) &+ \sum_{i=\underline{k}}^{\bar{k}} \Big(\rho_{\underline{k}\leq\bar{k}}(n-(i+1),m-1) + w_{\underline{k}\leq\bar{k}}(n-(i+1),m-1)\Big)\\ &+ \sum_{i=1}^{\underline{k}-1} \Big(\rho_{\underline{k}\leq\bar{k}}(n-(i+1),m) + w_{\underline{k}\leq\bar{k}}(n-(i+1),m)\Big)\\ &+ \sum_{i=\bar{k}+1}^{n} \Big(\rho_{\underline{k}\leq\bar{k}}(n-(i+1),m) + w_{\underline{k}\leq\bar{k}}(n-(i+1),m)\Big). \end{aligned} \tag{385}$$

To initialise the recurrence we use the case $n = 1$, in which we can see by inspection that

$$\rho_{\underline{k}\leq\bar{k}}(1,m) = [\![m=0]\!][\![\underline{k}>1]\!] + [\![m=1]\!][\![\underline{k}=1]\!] \tag{386}$$

On the other hand, specialising recurrence (**??**) for $n = 1$ and considering (**??**) we get

$$\rho_{\underline{k}\leq\bar{k}}(1,m) = \rho_{\underline{k}\leq\bar{k}}\big(0,m\big) + \rho_{\underline{k}\leq\bar{k}}\big(-1,m-[\![\underline{k}=1]\!]\big) + w_{\underline{k}\leq\bar{k}}\big(-1,m-[\![\underline{k}=1]\!]\big). \tag{387}$$

We wish (**??**) to equal (**??**). Taking (**??**) and (**??**) into account, equality between these two expressions is fulfilled for all m and $1 \le \underline{k} \le \overline{k}$ by choosing

$$\rho_{\underline{k}\le\bar{k}}(-1,m) = \rho_{\underline{k}\le\bar{k}}(0,m) = 0, \tag{388}$$

which are therefore the initial values of (**??**).

Recurrence (**??**) also allows us to find the ogf $R_{\underline{k}\le\bar{k}}(x,y) = \sum_{n,m} \rho_{\underline{k}\le\bar{k}}(n,m)\, x^n y^m$ in terms of the ogf $W_{\underline{k}\le\bar{k}}(x,y)$ in (**??**). Recurrence (**??**) is valid already for all values of the parameters, so we just obtain the difference $\rho_{\underline{k}\le\bar{k}}(n,m) - \rho_{\underline{k}\le\bar{k}}(n-1,m)$ using (**??**) which yields

$$\begin{aligned}
\rho_{\underline{k}\le\bar{k}}(n,m) = {} & 2\,\rho_{\underline{k}\le\bar{k}}(n-1,m) + w_{\underline{k}\le\bar{k}}(n-2,m) \\
& + \rho_{\underline{k}\le\bar{k}}(n-(\underline{k}+1),m-1) + w_{\underline{k}\le\bar{k}}(n-(\underline{k}+1),m-1) \\
& - \rho_{\underline{k}\le\bar{k}}(n-(\underline{k}+1),m) - w_{\underline{k}\le\bar{k}}(n-(\underline{k}+1),m) \\
& - \rho_{\underline{k}\le\bar{k}}(n-(\bar{k}+2),m-1) - w_{\underline{k}\le\bar{k}}(n-(\bar{k}+2),m-1) \\
& + \rho_{\underline{k}\le\bar{k}}(n-(\bar{k}+2),m) + w_{\underline{k}\le\bar{k}}(n-(\bar{k}+2),m).
\end{aligned}$$

By multiplying this expression on both sides by $x^n y^m$ and then adding on n and m, we get

$$R_{\underline{k}\le\bar{k}}(x,y) = \frac{x^2 - (1-y)(x^{\underline{k}+1} - x^{\bar{k}+2})}{1 - 2x + (1-y)(x^{\underline{k}+1} - x^{\bar{k}+2})} W_{\underline{k}\le\bar{k}}(x,y), \tag{389}$$

where $W_{\underline{k}\le\bar{k}}(x,y)$ is given by (**??**).

8.3.2 Number of Nonnull Runs Over All n-Strings that Contain Exactly m Nonnull p-Parity Runs

We now give a recurrence relation for $\rho_{[p]}(n,m)$. Following the same strategy as in the previous section, the n-strings that start with 0 contribute $\rho_{[p]}(n-1,m)$ runs to $\rho_{[p]}(n,m)$. As for the n-strings that start with 1, if they start with an odd i-run then they contribute $\rho_{[p]}(n-(i+1),m-[\![p=1]\!])$ runs plus the $w_{[p]}(n-(i+1),m-[\![p=1]\!])$ initial runs themselves. If they start with an even i-run then they contribute $\rho_{[p]}(n-(i+1),m-[\![p=0]\!])$ runs plus the $w_{[p]}(n-(i+1),m-[\![p=0]\!])$ initial runs. This yields the recurrence

$$\begin{aligned}
\rho_{[p]}(n,m) = {} & \rho_{[p]}(n-1,m) + \sum_{i=1}^{n} \rho_{[p]}(n-(i+1),m-[\![p=\mathrm{mod}(i,2)]\!]) \\
& + \sum_{i=1}^{n} w_{[p]}(n-(i+1),m-[\![p=\mathrm{mod}(i,2)]\!]).
\end{aligned} \tag{390}$$

To initialise (**??**) we use the value at $n = 1$, which, by inspection, is

$$\rho_{[p]}(1, m) = [\![p = 0]\!][\![m = 0]\!] + [\![p = 1]\!][\![m = 1]\!]. \tag{391}$$

On the other hand, setting $n = 1$ in (**??**) we have that

$$\rho_{[p]}(1, m) = \rho_{[p]}(0, m) + \rho_{[p]}(-1, m - [\![p = 1]\!]) + w_{[p]}(-1, m - [\![p = 1]\!]). \tag{392}$$

We want (**??**) to equal (**??**).

Taking (**??**) and (**??**) into account, we see that equality between the two expressions is achieved for all m by choosing

$$\rho_{[p]}(-1, m) = \rho_{[p]}(0, m) = 0,$$

which are therefore the initial values of (**??**).

We may next obtain the ogf $R_{[p]}(x, y) = \sum_{n,m} \rho_{[p]}(n, m)\, x^n y^m$ in terms of the ogf $W_{[p]}(x, y)$ in (**??**). We first obtain the following difference using (**??**):

$$\begin{aligned}\rho_{[p]}(n, m) - \rho_{[p]}(n - 2, m) = \rho_{[p]}&(n - 1, m) - \rho_{[p]}(n - 3, m)\\ &+ \rho_{[p]}\big(n - 2, m - [\![p = 1]\!]\big) + \rho_{[p]}\big(n - 3, m - [\![p = 0]\!]\big)\\ &+ w_{[p]}\big(n - 2, m - [\![p = 1]\!]\big) + w_{[p]}\big(n - 3, m - [\![p = 0]\!]\big). \end{aligned} \tag{393}$$

As usual, by multiplying by $x^n y^m$ on both sides of this expression and then adding on n and m and using (**??**) we obtain

$$R_{[p]}(x, y) = \frac{x^2 y^{[\![p=1]\!]} + x^3 y^{[\![p=0]\!]}}{1 - x - x^2\big(1 + y^{[\![p=1]\!]}\big) + x^3\big(1 - y^{[\![p=0]\!]}\big)} W_{[p]}(x, y).$$

Remark. Grimaldi and Heubach [**?**, Thm. 4] gave a recurrence, an explicit expression and a generating function for the total number of runs of ones *and zeros* in the $w_{[1]}(n, 0)$ n-strings devoid of odd runs of zeros—in their notation, t_n. Notice that $\rho_{[1]}(n, 0)$ solely gives the total number of runs of ones in n-strings devoid of odd runs of ones. In any case, the corresponding generating functions are very similar. In the specific case of $m = 0$ and $p = 1$, (**??**) becomes

$$\rho_{[1]}(n, 0) - \rho_{[1]}(n - 2, 0) = \rho_{[1]}(n - 1, 0) + w_{[1]}\big(n - 3, 0\big).$$

From here, we have that

$$[y^0]R_{[1]}(x, y) = \frac{x^3}{1 - x - x^2}\, [y^0]W_{[1]}(x, y) \tag{394}$$

$$= \frac{x^2 - x^4}{(1 - x - x^2)^2}, \tag{395}$$

where we have used (**??**) in the second step. For comparison's sake, the related ogf given in [**?**] is $G_{t_n}(x) = (x - x^4)/(1 - x - x^2)^2$. See also Remark **??**. □

8.3.3 OEIS

The only sequences arising from the enumerations in Section **??** that we have been able to find in the OEIS are:

$\rho_1(n-1,0)$ is A136444 $\left(\sum_{k=0}^{n} k\binom{n-k}{2k}\right)$ for $n \geq 1$.

$\rho_{\geq 1}(n,1) = w_{\geq 1}(n,1)$ is A000217.

$\rho_{\geq 1}(n-1,2)$ is A034827: $2\binom{n}{4}$, for $n \geq 1$.

$\rho_{\geq 2}(n-1,0)$ is A001629 (Self-convolution of Fibonacci numbers), for $n \geq 1$.

$\rho_{\leq 1}(n-1,0) = \rho_1(n-1,0)$ is A136444 for $n \geq 1$.

$\rho_{[1]}(n-1,0)$ is A029907 for $n \geq 1$, or $\rho_{[1]}(n+1,0) = w_{[1]}(n,1)$ for $n \geq 0$.

$\rho_{[0]}(n-1,0) = w_{[0]}(n,1)$ is A384497 for $n \geq 1$.

8.4 Number of Nonnull Runs of Ones and/or Zeros Over All n-Strings that Contain Exactly m Nonnull Runs of Ones Under Different Constraints

In this section we address the problems of enumerating the number of nonnull runs of ones and zeros over all n-strings that contain exactly m nonnull $(\underline{k} \leq \overline{k})$-runs of ones or exactly m nonnull p-parity runs of ones. We call these two quantities $\beta_{\underline{k}\leq\overline{k}}(n,m)$ and $\beta_{[p]}(n,m)$, respectively. Special cases of both problems were previously studied by Nyblom [**?**] and by Grimaldi and Heubach [**?**].

These two enumerations can be obtained in a similar way as the equivalent enumerations in the previous section. The main difference is that, rather than being respectively based on recurrences for $w_{\underline{k}\leq\overline{k}}(n,m)$ and $w_{[p]}(n,m)$ like in Section **??**, the two recurrences are now based on mutual recurrences for the same two quantities constrained to the initial bit of the n-strings —just like the ones used in Section **??**, but for runs of ones only. The asymmetry of such mutual recurrences implies that deriving ogfs, although certainly feasible, is rather laborious —cf. (**??**).

Therefore we content ourselves with only providing recurrences for $\beta_{\underline{k}\leq\overline{k}}(n,m)$ and $\beta_{[p]}(n,m)$.

8.4.1 Number of Nonnull Runs of Ones and/or Zeros Over All n-Strings that Contain Exactly m Nonnull $(\underline{k} \leq \overline{k})$-Runs of Ones

As indicated we assume $\underline{k} \geq 1$. We begin by finding mutual recurrences for the number of n-strings that start with $b \in \{0,1\}$ and contain exactly m $(\underline{k} \leq \overline{k})$-runs of ones, which we denote by $w^b_{\underline{k}\leq\overline{k}}(n,m)$. If an n-string starts with an i-run of zeros then it contributes

$w^1_{\underline{k}\le\bar{k}}(n-i,m)$ to $w^0_{\underline{k}\le\bar{k}}(n,m)$. On the other hand if an n-string starts with an i-run of ones, if $\underline{k} \le i \le \bar{k}$ then it contributes $w^0_{\underline{k}\le\bar{k}}(n-i,m-1)$ to $w^1_{\underline{k}\le\bar{k}}(n,m)$, but otherwise it contributes $w^0_{\underline{k}\le\bar{k}}(n-i,m)$. Therefore

$$w^b_{\underline{k}\le\bar{k}}(n,m) = \sum_{i=\underline{k}}^{\bar{k}} w^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(n-i, m-[\![b=1]\!]\big) + \sum_{i=1}^{\underline{k}-1} w^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m) + \sum_{i=\bar{k}+1}^{n} w^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m), \tag{396}$$

where $b \in \{0,1\}$ and $\tilde{b} = \mathrm{mod}(b+1,2)$. When $n=1$ we know by inspection that

$$w^b_{\underline{k}\le\bar{k}}(1,m) = [\![b=0]\!][\![m=0]\!] + [\![b=1]\!]\big([\![m=0]\!][\![\underline{k}>1]\!] + [\![m=1]\!][\![\underline{k}=1]\!]\big). \tag{397}$$

On the other hand, in the specific case of $n=1$ (??) with becomes

$$w^b_{\underline{k}\le\bar{k}}(1,m) = w^{\tilde{b}}_{\underline{k}\le\bar{k}}\big(0, m-[\![b=1]\!][\![\underline{k}=1]\!]\big). \tag{398}$$

We may verify that (??) equals (??) when the following initial values are used:

$$w^b_{\underline{k}\le\bar{k}}(0,m) = [\![m=0]\!]. \tag{399}$$

Of course, $w_{\underline{k}\le\bar{k}}(n,m) = w^0_{\underline{k}\le\bar{k}}(n,m) + w^1_{\underline{k}\le\bar{k}}(n,m)$, which allows us to recover the enumeration in Section ?? for $\underline{k} \ge 1$ in an alternative way. However, it is not immediately obvious how to obtain a recurrence for $w_{\underline{k}\le\bar{k}}(n,m)$ through this approach, due to the asymmetry between the two mutual recurrences in (??) —similarly to the enumeration in Section ??. Moreover, this alternative approach also makes the derivation of the ogf (??) less simple.

After these preliminaries we are ready to tackle the enumeration of $\beta_{\underline{k}\le\bar{k}}(n,m)$. We begin by obtaining mutual recurrences for the number of nonnull runs of ones and zeros in n-strings that start with b and contain exactly m nonnull ($\underline{k} \le \bar{k}$)-runs (of ones), which we denote by $\beta^b_{\underline{k}\le\bar{k}}(n,m)$. These two recurrences can be reasoned as follows: the n-strings that start with an i-run of zeros contribute $\beta^1_{\underline{k}\le\bar{k}}(n-i,m)$ plus $w^1_{\underline{k}\le\bar{k}}(n,m)$ initial runs of zeros to $\beta^0_{\underline{k}\le\bar{k}}(n,m)$. As for the n-strings that start with an i-run of ones, if $\underline{k} \le i \le \bar{k}$ then they contribute $\beta^0_{\underline{k}\le\bar{k}}(n-i,m-1)$ plus $w^0_{\underline{k}\le\bar{k}}(n-i,m-1)$ initial runs of ones to $\beta^1_{\underline{k}\le\bar{k}}(n,m)$. If they start with an i-run of ones with $i < \underline{k}$ or $i > \bar{k}$ then they contribute $\beta^0_{\underline{k}\le\bar{k}}(n-i,m)$ plus $w^0_{\underline{k}\le\bar{k}}(n-i,m)$ initial runs of ones. This yields

$$\beta^b_{\underline{k}\le\bar{k}}(n,m) = \sum_{i=\underline{k}}^{\bar{k}} \Big(\beta^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m-[\![b=1]\!]) + w^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m-[\![b=1]\!])\Big)$$

$$+\sum_{i=1}^{\underline{k}-1}\Big(\beta^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m)+w^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m)\Big)+\sum_{i=\bar{k}+1}^{n}\Big(\beta^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m)+w^{\tilde{b}}_{\underline{k}\le\bar{k}}(n-i,m)\Big). \tag{400}$$

When $n=1$ we have by inspection that $\beta^b_{\underline{k}\le\bar{k}}(1,m)=w^b_{\underline{k}\le\bar{k}}(1,m)$ —see (**??**). Thus, the initialisation of (**??**) is $\beta^b_{\underline{k}\le\bar{k}}(0,m)=0$. Finally, the enumeration that we are interested in is

$$\beta_{\underline{k}\le\bar{k}}(n,m)=\beta^0_{\underline{k}\le\bar{k}}(n,m)+\beta^1_{\underline{k}\le\bar{k}}(n,m). \tag{401}$$

Remark. Nyblom [**?**, Thm. 3.2] gave a recurrence based on the generalised Fibonacci numbers (**??**) to compute the total number of runs of ones and zeros over all the $w_{\ge k}(n,0)$ n-strings that contain no runs of ones of length k or longer for $k\ge 3$, which he denotes by $\rho_k(n)$. Observe that (**??**) allows us to obtain $\rho_k(n)=\beta_{k\le n}(n,0)$ for all $1\le k\le n$.

□

8.4.2 Number of Nonnull Runs of Ones and/or Zeros Over All n-Strings that Contain Exactly m Nonnull p-Parity Runs of Ones

As in the previous section, we find first mutual recurrences for the number of n-strings that start with $b\in\{0,1\}$ and contain exactly m p-parity runs of ones, which we denote by $w^b_{[p]}(n,m)$. If an n-string starts with an i-run of zeros then it contributes $w^1_{[p]}(n-i,m)$ to $w^0_{[p]}(n,m)$. On the other hand if an n-string starts with an i-run of ones, it contributes $w^0_{[p]}(n-i,m-1)$ to $w^1_{[p]}(n,m)$ if $p=\mathrm{mod}(i,2)$; otherwise, it contributes $w^0_{[p]}(n-i,m)$. Therefore

$$w^b_{[p]}(n,m)=\sum_{i=1}^{n}w^{\tilde{b}}_{[p]}\big(n-i,m-[\![b=1]\!][\![p=\mathrm{mod}(i,2)]\!]\big), \tag{402}$$

where $b\in\{0,1\}$ and $\tilde{b}=\mathrm{mod}(b+1,2)$. In the case $n=1$ we know by inspection that

$$w^b_{[p]}(1,m)=[\![b=0]\!][\![m=0]\!]+[\![b=1]\!]\big([\![m=0]\!][\![p=0]\!]+[\![m=1]\!][\![p=1]\!]\big), \tag{403}$$

whereas in the specific case of $n=1$, expression (**??**) yields

$$w^b_{[p]}(1,m)=w^{\tilde{b}}_{[p]}\big(0,m-[\![b=1]\!][\![p=1]\!]\big). \tag{404}$$

We wish (**??**) to equal (**??**), which happens the following initial values are used:

$$w^b_{[p]}(0,m)=[\![m=0]\!]. \tag{405}$$

Thus the enumeration in Section **??** can alternatively be obtained using (**??**) as $w_{[p]}(n,m)=w^0_{[p]}(n,m)+w^1_{[p]}(n,m)$. But just like in the previous section, the asymmetry between the mutual recurrences makes it more difficult to get a recurrence for $w_{[p]}(n,m)$, or its ogf, through this approach.

In any case, we only use the recurrences in (**??**) to help us get $\beta_{[p]}(n,m)$. We can do so by obtaining, in turn, mutual recurrences for the number of nonnull runs of ones and zeros in n-strings that start with b and contain exactly m nonnull p-parity runs (of ones), which we denote by $\beta^b_{[p]}(n,m)$. These two mutual recurrences can be reasoned in a similar way as in previous section: the n-strings that start with an i-run of zeros contribute $\beta^1_{[p]}(n-i,m)$ plus $w^1_{[p]}(n,m)$ initial runs of zeros to $\beta^0_{[p]}(n,m)$. As for the n-strings that start with an i-run of ones, if $p = \mathrm{mod}(i,2)$ then they contribute $\beta^0_{[p]}(n-i,m-1)$ plus $w^0_{[p]}(n-i,m-1)$ initial runs of ones to $\beta^1_{[p]}(n,m)$; otherwise they contribute $\beta^0_{[p]}(n-i,m)$ plus $w^0_{[p]}(n-i,m)$ initial runs of ones. This yields

$$\beta^b_{[p]}(n,m) = \sum_{i=1}^{n}\Big(\beta^{\tilde{b}}_{[p]}\big(n-i, m-b\,[\![p=\mathrm{mod}(i,2)]\!]\big) + w^{\tilde{b}}_{[p]}\big(n-i, m-b\,[\![p=\mathrm{mod}(i,2)]\!]\big)\Big) \tag{406}$$

When $n=1$ we have that $\beta^b_{[p]}(1,m) = w^b_{[p]}(1,m)$ by inspection —see (**??**). Thus, the initialisation of (**??**) is $\beta^b_{[p]}(0,m) = 0$. Finally, the enumeration that we are interested in is

$$\beta_{[p]}(n,m) = \beta^0_{[p]}(n,m) + \beta^1_{[p]}(n,m). \tag{407}$$

Remark. Grimaldi and Heubach's enumeration mentioned in Remark **??** [**?**, Thm. 4], can be obtained as a special case of (**??**): $t_n = \beta_{[1]}(n,0)$. □

9 Number of Ones in Runs Under Different Constraints Over All n-Strings and Over Restricted Subsets

Our main goal in this section is enumerating how many ones are found in runs under different constraints over all n-strings. The most basic result, in Section **??**, is the enumeration of the ones in $(\underline{k} \le \overline{k})$-runs over all n-strings, of which we consider several relevant special cases. We also study in Section **??** the number of ones in the n-strings enumerated in Sections **??** and **??**. Several of the enumerations considered here were previously studied by Makri et al. [**?**], Nyblom [**?**], and Grimaldi and Heubach [**?**].

9.1 Number of Ones in $(\underline{k} \le \overline{k})$-Runs Over All n-Strings

We denote by $t_{\underline{k}\le\overline{k}}(n)$ the number of ones in $(\underline{k} \le \overline{k})$-runs over all n-strings. First of all, it is not possible to establish a simple formula for $t_{\underline{k}\le\overline{k}}(n)$ by directly exploiting the main results in Section **??** —i.e., using formulas parallel to (**??**) or to (**??**). It is possible, though, to obtain $t_{\underline{k}\le\bar{k}}(n)$ in a more roundabout way by using the enumeration in Section **??** as follows:

$$t_{\underline{k}\le\bar{k}}(n) = \sum_{k=\underline{k}}^{\bar{k}} k \sum_{m=1}^{\left\lfloor \frac{n+1}{k+1} \right\rfloor} m\, w_k(n,m). \tag{408}$$

Nevertheless, if our intention is to evaluate this expression through an explicit formula for $w_k(n,m)$ such as (**??**), then this expression just adds another layer of misery with respect to the already hard evaluation of (**??**). A kinder option is to directly replace the second summation in (**??**) with $\rho_k(n)$. Denoting by $t_k(n)$ the number of ones in k-runs over all n-strings, we trivially have that

$$t_k(n) = k\rho_k(n), \tag{409}$$

and so we can write (**??**) as

$$t_{\underline{k}\le\bar{k}}(n) = \sum_{k=\underline{k}}^{\bar{k}} t_k(n). \tag{410}$$

As from (**??**) and (**??**) we have that

$$t_k(n) = k\,(n-k+3)\,2^{n-(k+2)}[\![k<n]\!] + n\,[\![k=n]\!], \tag{411}$$

we can see that the evaluation of (**??**) only involves summations of the forms $\sum_k ka^k$ and $\sum_k k^2a^k$. Thus, after some algebra, we can put the desired enumeration as

$$t_{\underline{k}\le\bar{k}}(n) = \big(n(\underline{k}+1) - \underline{k}(\underline{k}-1)\big)\,2^{n-(\underline{k}+1)}\,[\![\underline{k}\le n]\!] - \big(n\,(\bar{k}+2) - \bar{k}\,(\bar{k}+1)\big)\,2^{n-(\bar{k}+2)}\,[\![\bar{k}<n]\!] \tag{412}$$

As usual, this expression is valid for $0 \le \underline{k} \le \bar{k}$ and $n \ge 1$ —even though, of course, null runs contain no ones and thus do not contribute to $t_{\underline{k}\le\bar{k}}(n)$.

We will not show it here, but, like the main result in Section **??**, (**??**) can also be obtained by establishing a recurrence for $t_{\underline{k}\le\bar{k}}(n)$ from first principles (i.e, without resorting to any of the results in previous sections), and then solving it directly by unrolling it.

9.1.1 Special Cases

We denote the number of ones in ($\ge k$)-runs, the number of ones in ($\le k$)-runs, and the total number of ones over all n-strings by $t_{\ge k}(n)$, $t_{\le k}(n)$ and $t(n)$, respectively. These are special cases of the result in the previous section, since we can write $t_{\ge k}(n) = t_{k\le n}(n)$, $t_{\le k}(n) = t_{1\le k}(n)$, and $t(n) = t_{1\le n}(n)$. Thus, using (**??**) we have that

$$t_{\ge k}(n) = \big(n(k+1) - k(k-1)\big)\,2^{n-(k+1)}\,[\![k\le n]\!], \tag{413}$$

Because $t(n) = t_{\geq 1}(n)$, we have

$$t(n) = n\,2^{n-1}. \tag{414}$$

Finally,

$$t_{\leq k}(n) = t(n) - \big(n\,(k+2) - k\,(k+1)\big)\,2^{n-(k+2)}\,[\![k < n]\!]. \tag{415}$$

Remark. In the context of runs in binary strings, the expression for $t(n)$, i.e., (**??**), was given by Nyblom [**?**, Lem. 3.1].

Of course, one can find (**??**) without using runs at all. For example, $\sum_{j=1}^{n} j\binom{n}{j} = n\,2^{n-1}$. This quantity was also used by Gold [**?**] in his computation of $\rho_k(n)$ —see Remark **??**.

Makri and Psillakis sketched a procedure in [**?**, Eq. (25)] to obtain $t_{\geq k}(n)$ —in their notation, $R_{n,k}^{(s)}$. At that point they did not complete the calculation, probably due to their approach making it rather hard to obtain a closed-form expression. At the very end of the same paper they stated: "*A simple explicit form of* $R_{n,k}^{(s)}$ *remains an open issue*". However, soon afterwards, Makri et al. were able to give the explicit closed-form expression (**??**) for $t_{\geq k}(n)$ [**?**, Eq. (7), $a = s$], by exploiting a previous probabilistic result of theirs about success runs statistics [**?**]. From (**??**), they also deduced a recurrence for $t_{\geq k}(n)$ [**?**, Eq. (9), $a = s$] —see also the last paragraph in Section **??**. □

9.1.2 OEIS

Below are the OEIS sequences that we have been able to find in connection to the enumerations in this section.

- Only four sequences emanating from $t_k(n)$ appear to be documented in the OEIS:

  $t_1(n+1)$ is A045623 —cf. $\rho_1(n+1)$.

  $t_2(n+1)$ is A087447 for $n \geq 2$.

  $t_3(n+2)$ is A084860 for $n \geq 2$.

  $t_4(n+3)$ is A001792 for $n \geq 2$. Incidentally, this is the same sequence as $\rho(n+1) = \rho_{\geq k}(n+k)$, but the connection between both is not immediately obvious.

- $t(n)$ is A001787.

- Sequences emanating from $t_{\geq k}(n)$:

  $t_{\geq 1}(n)$ is A001787 —cf. $t(n)$.

  $t_{\geq 2}(n)$ is A066373.

  $t_{\geq 3}(n+1)$ is A128135 for $n \geq 2$.

  $t_{\geq 4}(n)$ is A386250.

  $t_{\geq 5}(n+3)$ is A053220 for $n \geq 2$.

- Sequences emanating from $t_{\leq k}(n)$:

  $t_{\leq 1}(n)$ is A045623 —cf. $t_1(n+1)$.

  $t_{\leq 2}(n)$ is A386270.

9.2 Number of Ones Over All n-Strings that Contain Exactly m Nonnull Runs Under Different Constraints

This section is the equivalent of Section **??**. Here we deal with the problems of enumerating how many ones are found over all n-strings that contain exactly m nonnull $(\underline{k} \leq \overline{k})$-runs or exactly m nonnull p-parity runs. In keeping with our notation conventions, we call these quantities $t_{\underline{k}\leq\bar{k}}(n,m)$ and $t_{[p]}(n,m)$, respectively. Special cases of these problems were previously studied by Nyblom [**?**] and by Grimaldi and Heubach [**?**].

Again, while relatively simple explicit expressions do not seem possible in general, we give recurrence relations and generating functions for these enumerations by relying on their counterparts in Sections **??** and **??**, i.e., $w_{\underline{k}\leq\bar{k}}(n,m)$ and $w_{[p]}(n,m)$, respectively. We must mention though that explicit expressions are possible in some particular cases —see Remark **??** and Section **??**.

9.2.1 Number of Ones Over All n-Strings That Contain Exactly m Nonnull $(\underline{k} \leq \bar{k})$-Runs

We obtain first a recurrence for $t_{\underline{k}\leq\bar{k}}(n,m)$, the number of ones over all n-strings that contain exactly m nonnull $(\underline{k} \leq \bar{k})$-runs. As indicated, we assume $\underline{k} \geq 1$. Following the customary strategy, the n-strings that start with 0 contribute $t_{\underline{k}\leq\bar{k}}(n-1,m)$ runs to $t_{\underline{k}\leq\bar{k}}(n,m)$. As for the n-strings that start with 1, if they start with a nonnull i-run with $\underline{k} \leq i \leq \bar{k}$ then they contribute $t_{\underline{k}\leq\bar{k}}(n-(i+1),m-1)$ ones plus the $i\,w_{\underline{k}\leq\bar{k}}(n-(i+1),m-1)$ initial ones themselves. If they start with a nonnull i-run with $i < \underline{k}$ or $i > \bar{k}$ then they contribute $t_{\underline{k}\leq\bar{k}}(n-(i+1),m)$ ones plus the $i\,w_{\underline{k}\leq\bar{k}}(n-(i+1),m)$ initial ones. This yields the recurrence

$$
\begin{aligned}
t_{\underline{k}\leq\bar{k}}(n,m) = t_{\underline{k}\leq\bar{k}}(n-1,m) &+ \sum_{i=\underline{k}}^{\bar{k}} \Big(t_{\underline{k}\leq\bar{k}}(n-(i+1),m-1) + i\,w_{\underline{k}\leq\bar{k}}(n-(i+1),m-1)\Big)\\
&+ \sum_{i=1}^{\underline{k}-1} \Big(t_{\underline{k}\leq\bar{k}}(n-(i+1),m) + i\,w_{\underline{k}\leq\bar{k}}(n-(i+1),m)\Big)\\
&+ \sum_{i=\bar{k}+1}^{n} \Big(t_{\underline{k}\leq\bar{k}}(n-(i+1),m) + i\,w_{\underline{k}\leq\bar{k}}(n-(i+1),m)\Big), \quad (416)
\end{aligned}
$$

which is naturally very similar to (**??**). To initialise the recurrence we use the case $n = 1$, which leads to the very same equations as in Section **??**, and thus, like in (**??**), the initial values of (**??**) are

$$t_{\underline{k}\le\bar{k}}(-1, m) = t_{\underline{k}\le\bar{k}}(0, m) = 0.$$

In order to obtain the ogf $T_{\underline{k}\le\bar{k}}(x, y) = \sum_{n,m} t_{\underline{k}\le\bar{k}}(n, m)\, x^n y^m$ we need a recurrence without an n-dependent summation. We can achieve this by first obtaining the difference $t_{\underline{k}\le\bar{k}}(n, m) - t_{\underline{k}\le\bar{k}}(n - 1, m)$ using (**??**), and then obtaining the same difference again but using the recurrence resulting from the first step. This is a straightforward but tedious computation and so, instead of listing the intermediate steps, we directly give the final recurrence —which is of course equivalent to (**??**):

$$\begin{aligned}
t_{\underline{k}\le\bar{k}}(n, m) = {} & 3\, t_{\underline{k}\le\bar{k}}(n - 1, m) + t_{\underline{k}\le\bar{k}}(n - (\underline{k} + 1), m - 1) - t_{\underline{k}\le\bar{k}}(n - (\bar{k} + 2), m - 1) \\
& - t_{\underline{k}\le\bar{k}}(n - (\underline{k} + 1), m) + t_{\underline{k}\le\bar{k}}(n - (\bar{k} + 2), m) - 2\, t_{\underline{k}\le\bar{k}}(n - 2, m) \\
& - t_{\underline{k}\le\bar{k}}(n - (\underline{k} + 2), m - 1) + t_{\underline{k}\le\bar{k}}(n - (\bar{k} + 3), m - 1) \\
& + t_{\underline{k}\le\bar{k}}(n - (\underline{k} + 2), m) - t_{\underline{k}\le\bar{k}}(n - (\bar{k} + 3), m) \\
& - (\underline{k} - 1)\, w_{\underline{k}\le\bar{k}}(n - (\underline{k} + 2), m - 1) - (\bar{k} + 1)\, w_{\underline{k}\le\bar{k}}(n - (\bar{k} + 2), m - 1) \\
& + \bar{k}\, w_{\underline{k}\le\bar{k}}(n - (\bar{k} + 3), m - 1) + \underline{k}\, w_{\underline{k}\le\bar{k}}(n - (\underline{k} + 1), m - 1) \\
& + w_{\underline{k}\le\bar{k}}(n - 2, m) + (\underline{k} - 1)\, w_{\underline{k}\le\bar{k}}(n - (\underline{k} + 2), m) - \underline{k}\, w_{\underline{k}\le\bar{k}}(n - (\underline{k} + 1), m) \\
& - \bar{k}\, w_{\underline{k}\le\bar{k}}(n - (\bar{k} + 3), m) + (\bar{k} + 1)\, w_{\underline{k}\le\bar{k}}(n - (\bar{k} + 2), m).
\end{aligned} \tag{417}$$

Now, multiplying (**??**) on both sides by $x^n y^m$ and summing over n and m we finally obtain the desired ogf in terms of (**??**):

$$T_{\underline{k}\le\bar{k}}(x, y) = \frac{x^2 - (1 - y)\Big(\underline{k}x^{\underline{k}+1} - (\underline{k} - 1)x^{\underline{k}+2} - (\bar{k} + 1)x^{\bar{k}+2} + \bar{k}x^{\bar{k}+3}\Big)}{1 - 3x + 2x^2 + (1 - y)\big(x^{\underline{k}+1} - x^{\underline{k}+2} - x^{\bar{k}+2} + x^{\bar{k}+3}\big)} W_{\underline{k}\le\bar{k}}(x, y). \tag{418}$$

Remark. Nyblom [**?**, Thm. 3.1] gave a recurrence to compute the number of *zeros* in n-strings devoid of $(\ge k)$-runs, which he denotes by $\rho_k(n)$. We can also obtain this enumeration as a special case of our results in this section by using $t_{\ge k}(n, 0) = t_{k\le n}(n, 0)$, which allows us to write $\rho_k(n) = n\, w_{\ge k}(n, 0) - t_{\ge k}(n, 0)$. Also, this expression is valid for all $k \ge 1$ whereas the recurrence in [**?**] requires $k \ge 3$.

□

9.2.2 Number of Ones Over All n-Strings That Contain Exactly m Nonnull p-Parity Runs

Finally, we find a recurrence relation for $t_{[p]}(n, m)$. The n-strings that start with 0 contribute $t_{[p]}(n-1, m)$ ones to $t_{[p]}(n, m)$. As for the n-strings that start with 1, if they start with an odd

i-run then they contribute $t_{[p]}(n-(i+1), m-[\![p=1]\!])$ ones plus the $i\, w_{[p]}(n-(i+1), m-[\![p=1]\!])$ initial ones themselves. If they start with an even nonnull i-run then they contribute $t_{[p]}(n-(i+1), m-[\![p=0]\!])$ runs plus the $i\, w_{[p]}(n-(i+1), m-[\![p=0]\!])$ initial runs. This yields the recurrence

$$\begin{aligned} t_{[p]}(n,m) = t_{[p]}(n-1,m) &+ \sum_{i=1}^{n} t_{[p]}(n-(i+1), m-[\![p=\mathrm{mod}(i,2)]\!]) \\ &+ \sum_{i=1}^{n} i\, w_{[p]}(n-(i+1), m-[\![p=\mathrm{mod}(i,2)]\!]). \end{aligned} \tag{419}$$

To initialise (**??**) we use special case $n=1$, which leads to the very same equations as in Section **??** —i.e., (**??**) and (**??**). Thus the initial values of (**??**) are

$$t_{[p]}(-1,m) = t_{[p]}(0,m) = 0.$$

In order to obtain the ogf $T_{[p]}(x,y) = \sum_{n,m} t_{[p]}(n,m)\, x^n y^m$ we need a recurrence without an n-dependent summation. We can achieve this by first obtaining the difference $t_{[p]}(n,m) - t_{[p]}(n-2,m)$ using (**??**), and then obtaining again the same difference using the recurrence resulting from the first step. We directly give the final recurrence —which is of course equivalent to (**??**):

$$\begin{aligned} t_{[p]}(n,m) = t_{[p]}(n-1,m) &+ 2\, t_{[p]}(n-2,m) - 2\, t_{[p]}(n-3,m) + t_{[p]}(n-2, m-[\![p=1]\!]) \\ &+ t_{[p]}(n-3, m-[\![p=0]\!]) - t_{[p]}(n-4,m) + t_{[p]}(n-5,m) \\ &- t_{[p]}(n-4, m-[\![p=1]\!]) - t_{[p]}(n-5, m-[\![p=0]\!]) \\ &+ w_{[p]}(n-4, m-[\![p=1]\!]) + w_{[p]}(n-2, m-[\![p=1]\!]) \\ &+ 2\, w_{[p]}(n-3, m-[\![p=0]\!]). \end{aligned} \tag{420}$$

Lastly, multiplying (**??**) on both sides by $x^n y^m$ and summing over n and m we obtain the desired ogf in terms of (**??**):

$$T_{[p]}(x,y) = \frac{2x^3 y^{[\![p=0]\!]} + (x^2+x^4) y^{[\![p=1]\!]}}{(x^5-x^3) y^{[\![p=0]\!]} + (x^4-x^2) y^{[\![p=1]\!]} + 1 - x - 2x^2 + 2x^3 + x^4 - x^5}\, W_{[p]}(x,y). \tag{421}$$

Remark. Grimaldi and Heubach [**?**, Thm. 3] gave a recurrence, an explicit expression and a generating function for the number of zeros in n-strings devoid of odd runs of zeros—in their notation, z_n. Notice that $t_{[1]}(n,0)$ gives the number of ones in n-strings devoid of odd runs of ones, and thus, by symmetry, $z_n = t_{[1]}(n,0)$. Let us verify this fact by comparing the corresponding ogfs. In the specific case of $m=0$ and $p=1$, recurrence (**??**) becomes

$$t_{[1]}(n,0) = t_{[1]}(n-1,0) + 2\, t_{[1]}(n-2,0) - t_{[1]}(n-3,0) - t_{[1]}(n-4,0) + 2\, w_{[1]}\big(n-3,0\big).$$

From here, we have that

$$[y^0]T_{[1]}(x,y) = \frac{2x^3}{1-x-2x^2+x^3+x^4}\,[y^0]W_{[1]}(x,y) \tag{422}$$

$$= \frac{2x^2}{(1-x-x^2)^2}, \tag{423}$$

where we have used (**??**) to get (**??**) from (**??**). As expected, the ogf (**??**) is the same as $G_{z_n}(x)$ in [**?**, Thm. 3]. Grimaldi and Heubach also studied the number of ones in n-strings devoid of odd runs of zeros, which they denoted by w_n. Using our own results, we can write $w_n = n\,w_{[1]}(n,0) - t_{[1]}(n,0)$. □

9.2.3 OEIS

The only sequences resulting from the enumerations in Section **??** that we have been able to find in the OEIS are:

$t_1(n,0)$ is A259966 (Total binary weight of all A005251(n) binary sequences of length n not containing any isolated 1's) —cf. $w_{\le 1}(n-1,0) = w_1(n-1,0)$.

$t_{\ge 1}(n,1)$ is A000292 (Tetrahedral —or triangular pyramidal— numbers).

$t_{\ge 2}(n,0)$ is A001629 —cf. $\rho_{\ge 2}(n,0)$.

$t_{\le 1}(n,0)$ is A259966 —cf. $t_1(n,0)$.

10 Conclusions

We have studied fundamental enumeration problems concerning runs in binary strings, where the runs conform to Mood's criterion [**?**]. While many of these problems had been solved over the years by different authors through a variety of methods, we felt that there were gaps in the literature, and that a uniform, systematic treatment would be beneficial to bring together and connect the many contributions to the topic. We believe that the notation conventions that we have chosen, even if necessarily unorthodox when it comes to probabilities (in order to accommodate recurrences and pgfs), are effective when it comes to displaying the relationships between enumerations concerning runs of ones, runs of ones and/or zeros, and compositions, their probabilistic extensions, and their associated generating functions. We consider that a relevant contribution of this paper is the identification of the existence of null runs of ones (0-runs). Probably due to their lack of practical use, null runs seem to have escaped the attention of previous authors.

Nevertheless, we have seen that runs of ones of zero length are essential to connect enumerations of runs of ones and zeros and enumerations of runs of ones (Theorems **??** and **??**).

Null runs also naturally emerge in problems involving the longest run of ones (Sections **??** and **??**).

Hopefully, we have been able to show that approaching runs-related problems by means of recurrences and generating functions is, for the most part, more straightforward than doing so by means of direct combinatorial analysis methods. The exception to this rule may be some enumeration problems with a Hamming weight constraint. We are not reinventing the wheel here, since, as discussed in the introduction, this is an old and venerable approach —even when it comes to studying runs. However, this powerful technique has probably not been exploited to its full advantage in this area of research.

It would be naive to think that we have exhausted all interesting runs-related problems that one can address through the strategy adopted in this paper. Extensions to probabilistic scenarios —briefly discussed in Sections **??** and **??** and at some other points— can be investigated much further, following in the steps of previous authors who went beyond the iid case to study Markov dependencies [**?**, **?**, **?**, **?**]. Many purely enumerative problems not addressed here also merit consideration, such as for example joint enumerations —see [**?**, Sec. 2.4.9].

Of course, explicit expressions based on binomial coefficients have numerical limitations for large n. However, the generating functions that we have provided can definitely be helpful in further studies about the asymptotics the enumerations studied here. Austin and Guy [**?**], Sedgewick and Flajolet [**?**], Bloom [**?**], Suman [**?**], Prodinger [**?**], and other authors, previously followed this approach in some special scenarios. Another interesting route in terms of asymptotics could be using moment generating functions instead of probability generating functions, as done by Wishart and Hirschfeld [**?**] in one particular problem.

We should also remark that the mutual recurrences strategy on which the results in Sections **??**, **??**, and **??** hinge —which was first used by Wishart and Hirschfeld in a probabilistic context [**?**]— can be readily extended to the enumeration of runs of different kinds in b-ary strings with $b \geq 2$. Some of these b-ary generalisations were already given by previous authors [**?**, **?**], although not through the procedure we just mentioned. Any of the research avenues sketched above may prove fruitful: in Schilling's words, "*the variety of potential applications of runs theory is virtually boundless*" [**?**].

The problems addressed in this paper contain a very rich structure in terms of the many distinct number sequences that they encompass —many of them listed in Sloane's On-Line Encyclopedia of Integer Sequences, but many others yet unexplored. The omissions may be as relevant as the inclusions: observe that few sequences corresponding to enumerations with $m \geq 2$ are in the OEIS. The many connections unearthed thanks to the inestimable help of the OEIS may open up new uses for the diverse formulas and results given here.

To conclude, the recurrences for runs of ones that we have produced compel us to say that, in the same sense that $\sqrt{-1}$ exists as a mathematical object, a binary string with length -1 does also have some kind of fleeting existence. Intriguingly, we are justified in stating that the number of binary strings of length -1 that are devoid of runs of ones of length $k \geq 0$ is exactly one —in the same sense that we accept that there is a single binary string of length 0 (empty binary string). We have also seen that, when binary strings are drawn uniformly at random with probability $0 < q < 1$ of drawing a 1, the probability of a (or rather, "the") binary string of length -1 without runs of ones of length $k \geq 0$ must be $1/(1-q) > 1$. The majority of recurrences for runs of ones that we have given would not be correct *for all* $n \geq 1$ *and for all valid values of the parameters* without similar puzzling assertions being true. Had we ignored or tried to skirt around these strange-looking facts, then most of those recurrences would have been less general and more piecemeal —i.e., only valid for certain ranges of the parameters and/or requiring ad-hoc initialisation. This fragmentary character is in fact present in many recurrences concerning runs (or compositions) previously given by other authors [**?**, **?**, **?**, **?**, **?**, **?**]. We are not the first ones to observe this phenomenon: the necessity of considering a binary string of length -1 in recurrences involving runs of ones was implicit already in the work of of Austin and Guy [**?**].

2020 *Mathematics Subject Classification*: Primary 05A15

Keywords: run, success run, recurrence, ordinary generating function, probability generating function.