Elect. Comm. in Probab. **7** (2002) 85–96

ELECTRONIC COMMUNICATIONS in PROBABILITY

INCLUSION–EXCLUSION *REDUX*

DAVID KESSLER
Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel
email: kessler@dave.ph.biu.ac.il

JEREMY SCHIFF
Department of Mathematics, Bar-Ilan University, Ramat Gan 52900, Israel
email: schiff@math.biu.ac.il

submitted December 18, 2001 *Final version accepted* March 11, 2002

AMS 2000 Subject classification: 60C05
Inclusion-exclusion principle, close-to-independent events

Abstract
We present a reordered version of the inclusion–exclusion principle, which is useful when computing the probability of a union of events which are close to independent. The advantages of this formulation are demonstrated in the context of 3 classic problems in combinatorics.

1 Introduction

The inclusion–exclusion principle is one of the fundamental results of combinatorics. If A is the union of the events $A_1, A_2, \ldots, A_n$ then, writing p_i for the probability of A_i, p_{ij} for the probability of $A_i \cap A_j$, p_{ijk} for the probability of $A_i \cap A_j \cap A_k$ etc, the probability of A is given by

$$\mathbf{P}(A) = \sum_i p_i - \sum_{i<j} p_{ij} + \sum_{i<j<k} p_{ijk} - \cdots + (-1)^{n-1} p_{123\ldots n} \ . \tag{1}$$

The inclusion–exclusion principle tells us that if we know the $p_i, p_{ij}, p_{ijk} \ldots$ then we can find $\mathbf{P}(A)$. In practice, though, we are unlikely to have full information on the $p_i, p_{ij}, p_{ijk} \ldots$. We are then faced with the highly nontrivial task of approximating $\mathbf{P}(A)$ taking into account whatever partial information we are given. The difficulty in this is that knowledge of any of the probabilities $p_i, p_{ij}, p_{ijk} \ldots$ places constraints on the others; these constraints have been extensively studied for about 150 years [1].

Recent work of Kahn, Linial, Nisan and Samorodnitsky [2] has shown that if n is large, and we are given all the probabilities $p_i, p_{ij}, p_{ijk} \ldots$ with up to r indices, then in general we will not be able to make any firm predictions about $\mathbf{P}(A)$ unless r is at least $O(\sqrt{n})$. On the other hand, in certain cases where the events A_i are in some sense close to being independent, then there are a number of known results bounding $\mathbf{P}(A)$, such as the Lovász local lemma [3] and Janson's inequality [4] (see [5] for an exposition), and these bounds use just the p_i and the p_{ij}. So although little can be said in general, in the case that the events A_i are close to independent we might hope that inclusion–exclusion can be used to give good estimates for $\mathbf{P}(A)$. The aim of this paper is to propose a method for this.

A first guess for approximating $\mathbf{P}(A)$, if we are given just the "low order" probabilities (the p's with few indices), might be simple truncation of the series in (1). In other words, if we know just the p_i we might take just the first term, if we know both the p_i and the p_{ij} we might take just the first two terms etc. This approach works poorly in general, but if the events A_i are close to mutually exclusive (i.e. the probabilities of the multiple intersections $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}$ drop rapidly as r increases), then such truncations do give good approximations. Of course, if the events A_i actually are mutually exclusive, then any truncation of (1) will be exact. The approach we give to approximating $\mathbf{P}(A)$ when the events A_i are close to being independent is based on a reordering of terms in the inclusion–exclusion formula, with the property that any truncation of the reordered formula is exact if the events A_i are independent.

Our reordering of the inclusion–exclusion principle is presented in section 2. In section 3 we present results from use of our approximation schemes on 3 classic problems in combinatorics, which are encouraging. In section 4 we discuss further questions arising out of our work. An appendix discusses the relationship with Janson's inequality.

As a last note in this introduction, we refer the reader to the work of Naiman and Wynn [6], who have shown that under certain circumstances there can be significant simplifications in the inclusion–exclusion principle, for example whem it is used to calculate the volume with respect to some measure of a finite union of balls in d-dimensional Euclidean space.

2 Reordering the Inclusion–Exclusion Principle

Proposition 1. *If A is the union of the events $A_1, A_2, \ldots, A_n$, and we write*

$$\begin{aligned}
q_i &= \mathbf{P}(\bar{A}_i) \\
q_{ij} &= \frac{\mathbf{P}(\bar{A}_i \cap \bar{A}_j)}{\mathbf{P}(\bar{A}_i)\mathbf{P}(\bar{A}_j)} \\
q_{ijk} &= \frac{\mathbf{P}(\bar{A}_i)\mathbf{P}(\bar{A}_j)\mathbf{P}(\bar{A}_k)\mathbf{P}(\bar{A}_i \cap \bar{A}_j \cap \bar{A}_k)}{\mathbf{P}(\bar{A}_i \cap \bar{A}_j)\mathbf{P}(\bar{A}_i \cap \bar{A}_k)\mathbf{P}(\bar{A}_j \cap \bar{A}_k)} \\
&\vdots \\
q_{12\ldots n} &= \left(\prod_i \mathbf{P}(\bar{A}_i)\right)^{(-1)^{n-1}} \left(\prod_{i<j} \mathbf{P}(\bar{A}_i \cap \bar{A}_j)\right)^{(-1)^{n-2}} \left(\prod_{i<j<k} \mathbf{P}(\bar{A}_i \cap \bar{A}_j \cap \bar{A}_k)\right)^{(-1)^{n-3}} \\
&\quad \times \cdots \times \left(\prod_i \mathbf{P}(\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_{i-1} \cap \bar{A}_{i+1} \cap \cdots \cap \bar{A}_n)\right)^{-1} \mathbf{P}(\bar{A}_1 \cap \bar{A}_2 \cap \cdots \cap \bar{A}_n)
\end{aligned}$$

then

$$\mathbf{P}(A) = 1 - \left(\prod_i q_i\right)\left(\prod_{i<j} q_{ij}\right)\left(\prod_{i<j<k} q_{ijk}\right) \cdots q_{12\ldots n} \ . \tag{2}$$

Proof. We count the number of times each factor occurs in the product of the q's in (2). For fixed r, $\mathbf{P}(\bar{A}_r)$ appears once in the product of the q_i, $\binom{n-1}{1}$ times in the denominator in the product of the q_{ij}, $\binom{n-1}{2}$ times in the numerator of the product of the q_{ijk} etc. Thus in the full product of the q's the number of factors of $\mathbf{P}(\bar{A}_r)$ is

$$1-\binom{n-1}{1}+\binom{n-1}{2}-\binom{n-1}{3}+\cdots+(-1)^{n-1}\binom{n-1}{n-1}=\begin{cases}1 & n=1\\ 0 & n>1\end{cases}.$$

Likewise for fixed r, s and $n \geq 2$, the factor $\mathbf{P}(\bar{A}_r \cap \bar{A}_s)$ appears once in the product of the q_{ij}, $\binom{n-2}{1}$ times in the denominator in the product of the q_{ijk}, $\binom{n-2}{2}$ times in the numerator of the product of the q_{ijkl} etc. Thus in the full product of the q's the number of factors of $\mathbf{P}(\bar{A}_r \cap \bar{A}_s)$ is

$$1-\binom{n-2}{1}+\binom{n-2}{2}-\binom{n-2}{3}+\ldots+(-1)^{n-2}\binom{n-2}{n-2}=\begin{cases}1 & n=2\\ 0 & n>2\end{cases}.$$

Continuing this way we see that the full product of the q's is simply

$$\begin{aligned}\mathbf{P}(\bar{A}_1 \cap \bar{A}_2 \cap \ldots \cap \bar{A}_n) &= \mathbf{P}\left(\overline{A_1 \cup A_2 \cup \ldots A_n}\right)\\ &= 1-\mathbf{P}(A_1 \cup A_2 \cup \ldots A_n)\\ &= 1-\mathbf{P}(A). \qquad \bullet\end{aligned}$$

Now

$$\begin{aligned}\mathbf{P}(\bar{A}_i) &= 1-\mathbf{P}(A_i)=1-p_i\ ,\\ \mathbf{P}(\bar{A}_i \cap \bar{A}_j) &= 1-\mathbf{P}(A_i \cup A_j)=1-p_i-p_j+p_{ij}\\ \mathbf{P}(\bar{A}_i \cap \bar{A}_j \cap \bar{A}_k) &= 1-\mathbf{P}(A_i \cup A_j \cup A_k)=1-p_i-p_j-p_k+p_{ij}+p_{jk}+p_{ki}-p_{ijk}\end{aligned}$$

etc. Thus the q_i can be written in terms of the p_i (in fact $q_i = 1-p_i$), the q_{ij} can be written in terms of the p_i and the p_{ij}, the q_{ijk} can be written in terms of the p_i, the p_{ij} and the p_{ijk} etc. With the $q_i, q_{ij}, q_{ijk}, \ldots$ written this way in terms of the $p_i, p_{ij}, p_{ijk}, \ldots$, we call (2) *the reordered inclusion–exclusion principle.* If all the products were to be multiplied out it would reduce to the standard inclusion–exclusion principle (1). But for approximation purposes, at least when the events A_i are close to independent, the form (2) turns out to be much more useful. In particular the reader will be able to verify the following result:

Proposition 2. *The A_i are independent if and only if for all $i, j, k, \ldots$, $q_{ij} = q_{ijk} = \cdots = q_{12\ldots n} = 1$.*

In other words, for independent events A_i we can "truncate" the product in (2) after as many terms as we wish, and the result will be exact. The $q_{ij}, q_{ijk}, \ldots$ provide a measure of the dependence of the events. In addition to the formulae given above for the $q_{ij}, q_{ijk}, \ldots$ we note

$$\begin{aligned}q_{ij} &= \frac{\mathbf{P}(\bar{A}_i|\bar{A}_j)}{\mathbf{P}(\bar{A}_i)}\ ,\\ q_{ijk} &= \frac{\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_j \cap \bar{A}_k)}{\mathbf{P}(\bar{A}_i)}}{\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_j)}{\mathbf{P}(\bar{A}_i)}\,\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_k)}{\mathbf{P}(\bar{A}_i)}}\ ,\\ q_{ijkl} &= \frac{\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_j \cap \bar{A}_k \cap \bar{A}_l)}{\mathbf{P}(\bar{A}_i)}\,\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_j)}{\mathbf{P}(\bar{A}_i)}\,\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_k)}{\mathbf{P}(\bar{A}_i)}\,\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_l)}{\mathbf{P}(\bar{A}_i)}}{\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_j \cap \bar{A}_k)}{\mathbf{P}(\bar{A}_i)}\,\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_j \cap \bar{A}_l)}{\mathbf{P}(\bar{A}_i)}\,\dfrac{\mathbf{P}(\bar{A}_i|\bar{A}_k \cap \bar{A}_l)}{\mathbf{P}(\bar{A}_i)}}\ .\end{aligned}$$

From these formulae (and their generalizations with more indices) we see that if the event A_i is independent of the set of events $A_{j_1}, \dots, A_{j_r}$ (in the sense that any information on $A_{j_1}, \dots, A_{j_r}$ does not affect A_i), then $q_{ij_1} = q_{ij_1 j_2} = \cdots = q_{ij_1 j_2 \dots j_r}$.
What we have written up to here is probably sufficient to justify using truncations of (2), with the $q_i, q_{ij}, q_{ijk}, \dots$ written out in terms of the $p_i, p_{ij}, p_{ijk}, \dots$, as a method for approximation of probabilities for a union of events that are close to independent. But before moving to examples, we mention another property that reinforces the link between this and the standard inclusion–exclusion principle.

Definition. We say a function f of the $p_i, p_{ij}, p_{ijk}, \dots$ is *homogeneous of order n* if

$$f(\lambda p_i, \lambda^2 p_{ij}, \lambda^3 p_{ijk}, \dots) = \lambda^n f(p_i, p_{ij}, p_{ijk}, \dots) \ ,$$

and, in greater generality, *of order n* if

$$f(\lambda p_i, \lambda^2 p_{ij}, \lambda^3 p_{ijk}, \dots) = O(\lambda^n) \ .$$

With this definition of order, the standard inclusion–exclusion principle writes $\mathbf{P}(A)$ as a sum of terms which are, respectively, homogeneous of order 1,2,3,... The reordered inclusion–exclusion principle has a similar property:

Proposition 3. *As functions of the $p_i, p_{ij}, p_{ijk}, \dots$, $q_i - 1$ is of order 1, $q_{ij} - 1$ is of order 2, $q_{ijk} - 1$ is of order 3 etc.*

Before we explain the proof of this we demonstrate it explicitly in the first few cases. For the one index case it is obvious since

$$q_i = 1 - p_i \ . \tag{3}$$

For the two index case we have

$$q_{ij} = \frac{1 - p_i - p_j + p_{ij}}{(1-p_i)(1-p_j)} = 1 + \frac{p_{ij} - p_i p_j}{(1-p_i)(1-p_j)} \ . \tag{4}$$

For the three index case, it can be checked that

$$q_{ijk} = \frac{(1 - p_i - p_j - p_k + p_{ij} + p_{jk} + p_{ik} - p_{ijk})(1-p_i)(1-p_j)(1-p_k)}{(1-p_i-p_j+p_{ij})(1-p_j-p_k+p_{jk})(1-p_i-p_k+p_{ik})} = 1 + \beta \tag{5}$$

$$\text{where} \quad \beta = \frac{\begin{pmatrix} -p_i p_j p_k (2 - p_i - p_j - p_k) - p_{ijk}(1-p_i)(1-p_j)(1-p_k) - p_{ij} p_{ik} p_{jk} \\ +p_i p_{jk}(1 - p_i - p_j p_k) + p_j p_{ik}(1 - p_j - p_i p_k) + p_k p_{ij}(1 - p_k - p_i p_j) \\ -p_{ij} p_{ik}(1 - p_j - p_k) - p_{jk} p_{ij}(1 - p_i - p_k) - p_{ik} p_{jk}(1 - p_i - p_j) \end{pmatrix}}{(1-p_i-p_j+p_{ij})(1-p_j-p_k+p_{jk})(1-p_i-p_k+p_{ik})} \ .$$

β is clearly of order 3.

Proof of Proposition 3. We consider the following perturbative approach to the reordered inclusion–exclusion principle. For independent events the first truncation

$$\mathbf{P}(A) \approx 1 - \prod_i (1 - p_i) \tag{6}$$

is exact. In general, (6) is not exact, but it is still correct to first order, that is if we ignore terms of order greater than 1 on each side. Suppose now that we try to modify (6) to make it correct to second order by looking at formulae of the form

$$\mathbf{P}(A) \approx 1 - \prod_i (1 - p_i) \prod_{i<j} (1 + \alpha_{ij}^{(2)}) \ , \tag{7}$$

where $\alpha_{ij}^{(2)}$ is homogeneous of order 2. A brief calculation shows this can be done if (and only if) we choose $\alpha_{ij}^{(2)} = p_{ij} - p_i p_j$. With this choice made, we can proceed to try to modify (7) to make it correct to order 3. This requires adding in two extra terms, both homogeneous of order 3,

$$\begin{aligned} \mathbf{P}(A) &\approx 1 - \prod_i (1 - p_i) \prod_{i<j} (1 + \alpha_{ij}^{(2)} + \alpha_{ij}^{(3)}) \prod_{i<j<k} (1 + \alpha_{ijk}^{(3)}) \ , \\ \text{where} \quad \alpha_{ij}^{(3)} &= (p_i + p_j)(p_{ij} - p_i p_j) \\ \alpha_{ijk}^{(3)} &= p_i p_{jk} + p_j p_{ik} + p_k p_{ij} - 2 p_i p_j p_k - p_{ijk} \ . \end{aligned}$$

We can continue to modify in this way to make the formula correct to arbitrary order; furthermore in the second product all correction terms will only depend on p_i, p_{ij}, in the third product all correction terms will only depend on p_i, p_{ij}, p_{ijk} etc. In this manner we build up the reordered inclusion–exclusion principle order-by-order, and in particular we deduce proposition 3. •

3 Examples

In the three examples below we approximate the probability of events using the first, second and third truncations of (2), i.e. the approximations

$$\mathbf{P}(A) \approx 1 - \left(\prod_i q_i \right) \ , \tag{8}$$

$$\mathbf{P}(A) \approx 1 - \left(\prod_i q_i \right) \left(\prod_{i<j} q_{ij} \right) \ , \tag{9}$$

$$\mathbf{P}(A) \approx 1 - \left(\prod_i q_i \right) \left(\prod_{i<j} q_{ij} \right) \left(\prod_{i<j<k} q_{ijk} \right) \ , \tag{10}$$

where q_i, q_{ij}, q_{ijk} are given by (3)-(5).

3a. The Derangement Problem. Suppose that in any pack of cards there are k copies of n different cards (for conventional cards $k = 4$, $n = 13$). Two players each take a pack of cards, and draw a card at random. We say a "match" occurs if they draw cards of the same kind. The players continue to draw cards at random and compare until their packs are used up. What is the probability that in the process there will be at least one match?

No closed form solution is known for this problem, but it is possible to derive an expansion for the probability in powers of $\frac{1}{n}$, the first few terms of which are

$$1-e^{-k}\left(1-\frac{k-1}{2n}+\frac{3k^3-14k^2+15k-4}{24kn^2}+\dots\right), \tag{11}$$

see for example [7].

To apply our methods, let A_i, $i=1,\dots,nk$, be the event that the players draw the same card on draw number i. Clearly $p_i=\frac{1}{n}$ for every i, so using the first reordered approximation (8) we obtain

$$\mathbf{P}(A)\approx 1-\left(1-\frac{1}{n}\right)^{nk}. \tag{12}$$

In the limit of large n we recover the correct limit $1-e^{-k}$. We emphasize that this comes from using *just the first reordered approximation*, i.e. just using the p_i; in contrast, if we were to take just the first term of the standard inclusion–exclusion formula we would obtain the absurd answer $\mathbf{P}(A)\approx k$. Note that expanding (12) in powers of $\frac{1}{n}$ gives

$$\mathbf{P}(A)\approx 1-e^{-k}\left(1-\frac{k}{2n}+\frac{(3k-8)k}{24n^2}+\dots\right),$$

and we see that the first reordered approximation in fact gives more than just the leading order term in $\frac{1}{n}$: For large k we obtain the correct dominant contribution to the coefficients of both the $O\left(\frac{1}{n}\right)$ and $O\left(\frac{1}{n^2}\right)$ terms. This is a first hint that our methods work well not just in the large n limit.

Moving to the second reordered approximation, there are a total of $\frac{1}{2}nk(nk-1)$ events $A_i\cap A_j$. In $\frac{1}{2}nk(k-1)$ of these player 1 draws identical cards on draws i and j, and we have $p_{ij}=\frac{k-1}{n(nk-1)}$. In the remaining $\frac{1}{2}n(n-1)k^2$, player 1 draws distinct cards on draws i and j, and we have $p_{ij}=\frac{k}{n(nk-1)}$. Thus (9) gives

$$\mathbf{P}(A)\approx 1-\left(1-\frac{1}{n}\right)^{nk}\left(\frac{1-\frac{2}{n}+\frac{k-1}{n(nk-1)}}{\left(1-\frac{1}{n}\right)^2}\right)^{\frac{1}{2}nk(k-1)}\left(\frac{1-\frac{2}{n}+\frac{k}{n(nk-1)}}{\left(1-\frac{1}{n}\right)^2}\right)^{\frac{1}{2}nk^2(n-1)}$$

Expanding this in powers of $\frac{1}{n}$ gives

$$\mathbf{P}(A)\approx 1-e^{-k}\left(1-\frac{k-1}{2n}+\frac{3k^3-14k^2+15k+12}{24kn^2}+\dots\right),$$

where now the coefficient of $\frac{1}{n}$ is exact, and the coefficient of $\frac{1}{n^2}$ is improving (at least for large k).

The necessary information to apply the third reordered approximation (10), is as follows:

Type of events $A_i\cap A_j\cap A_k$	Number of such events	Probability p_{ijk}
Player 1 draws 3 distinct cards	$\frac{n(n-1)(n-2)k^3}{6}$	$\frac{k^3}{nk(nk-1)(nk-2)}$
Player 1 draws 2 cards of one type and 1 of another type	$\frac{n(n-1)k^2(k-1)}{2}$	$\frac{k^2(k-1)}{nk(nk-1)(nk-2)}$
Player 1 draws 3 identical cards	$\frac{nk(k-1)(k-2)}{6}$	$\frac{k(k-1)(k-2)}{nk(nk-1)(nk-2)}$

Due to its length we do not write down explicitly the formula obtained by putting this information into (10). As might by now be expected, if we expand the answer in powers of

$\frac{1}{n}$ we obtain exact agreement with all three coefficients given in (11). In practice it seems the third reordered approximation outperforms the first three terms of the large n expansion (11). For conventional cards, $n = 13$ and $k = 4$, the third reordered approximation gives $1 - \mathbf{P}(A) \approx 0.01623287$, which has an error of less than 5% that of (11), which gives $1 - \mathbf{P}(A) \approx 0.01622939$. (we compare against the reported value $1 - \mathbf{P}(A) \approx 0.01623273$ given in [8]). For the case $n = 3$ and $k = 8$, the second and third reordered approximations give $1 - \mathbf{P}(A) \approx 7.47 \times 10^{-5}$ and $1 - \mathbf{P}(A) \approx 7.83 \times 10^{-5}$ respectively, the latter being in good accord with the result of a computer simulation of 10^9 trials, in which we observed 78119 cases of "no matches". In this case, with $k > n$, we do not expect (11) to do well, and it gives $1 - \mathbf{P}(A) \approx 9.09 \times 10^{-5}$.

In the case $k = 1$ the general derangement problem reduces to the famous "hatcheck" problem. In numerous probability textbooks, see for example [9], the full inclusion–exclusion principle is used to show that as $n \to \infty$ the probability of a match tends to $1 - e^{-1}$. We emphasize again that we have obtained this from just the first reordered approximation, using just the values of the p_i.

3b. Success Runs in Coin Flips. In a series of x coin flips, what is the probability that there will be at least one run of n successive "heads"?

Again no closed form solution is known to this problem, but if we denote the probability P_n^x then by conditioning on the possible outcomes of the first n events we obtain the recursion

$$P_n^x = \frac{1}{2}P_n^{x-1} + \frac{1}{4}P_n^{x-2} + \ldots + \frac{1}{2^n}P_n^{x-n} + \frac{1}{2^n} \ .$$

Using this, along with the starting values $P_n^x = 0$, $x = 0, 1, \ldots, n-1$, it is easy (for given n) to generate the probabilities numerically. Alternatively, writing $P_n^x = 1 - \frac{R_n^x}{2^x}$ we obtain the n-Fibonacci recursion [10]

$$R_n^x = R_n^{x-1} + R_n^{x-2} + \ldots + R_n^{x-n} \ ,$$

with starting values $R_n^x = 2^x$, $x = 0, 1, \ldots, n-1$. For fixed n one can solve numerically for the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic equation

$$\lambda^n = \lambda^{n-1} + \lambda^{n-2} + \cdots + 1 \tag{13}$$

of the recursion, and also determine constants $C_1, C_2, \ldots, C_n$ such that $R_n^x = C_1\lambda_1^x + C_2\lambda_2^x + \cdots + C_n\lambda_n^x$. It is known [10] that (13) has one root, say λ_1, close to 2 and all other roots inside the unit circle, which for large x will give very small contributions to the P_n^x. Thus when x is large

$$P_n^x \approx 1 - C_1\left(\frac{\lambda_1}{2}\right)^x \tag{14}$$

where λ_1 and C_1 depend on n alone and λ_1 is close to 2. It is possible, with substantial effort, to obtain expansions of λ_1 and C_1 in powers of 2^{-n}. As we will see, our methods produce such results with ease.

To apply reordered inclusion–exclusion to this problem, define A_i to be the event that there is a run of n heads starting on the ith coin flip, $i = 1, 2, \ldots, x-n+1$. Here by "starting" we mean strictly starting, i.e. except when $i = 1$ there is a tail on the $(i-1)$th flip. Thus

$$p_1 = \frac{1}{2^n} \ , \qquad p_i = \frac{1}{2^{n+1}} \ , \quad i = 2, \ldots, x-n+1 \ .$$

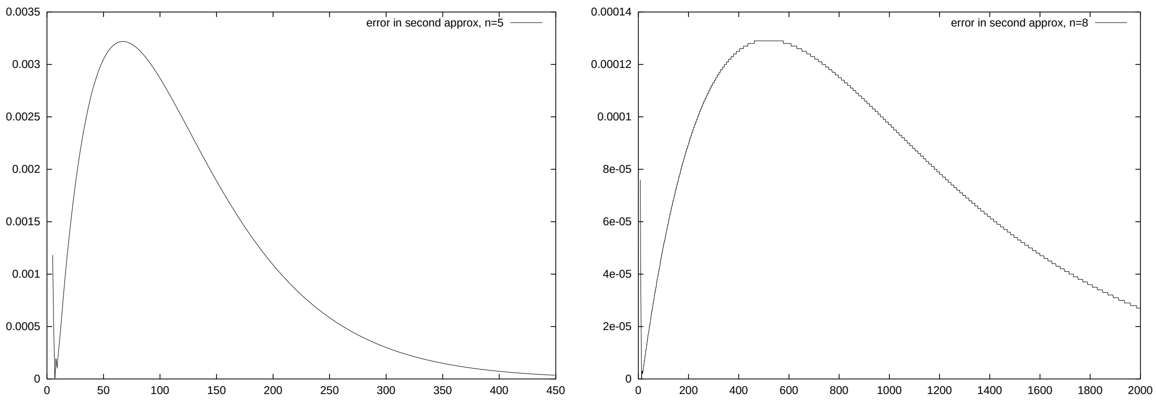

Figure 1: Absolute errors in the second reordered approximation for the success run problem, $n = 5$ and $n = 8$.

Applying (8) we have

$$P_n^x \approx 1 - \left(1 - \frac{1}{2^{n+1}}\right)^{x-n} \left(1 - \frac{1}{2^n}\right) .$$

This approximation is of the form (14), with

$$\frac{\lambda_1}{2} = 1 - \frac{1}{2^{n+1}} , \qquad C_1 = \left(1 - \frac{1}{2^{n+1}}\right)^{-n} \left(1 - \frac{1}{2^n}\right) \approx 1 + \frac{n-2}{2^{n+1}} .$$

Here in the last line we have indicated the start of the expansion of C_1 in powers of 2^{-n}.
To implement the second reordered approximation we need to look at the events $A_i \cap A_j$. Because we have defined A_i to be the event that there is a success run starting strictly on the ith flip, A_i and A_j will be mutually exclusive if $|i - j| \le n$. But for $|i - j| > n$ A_i and A_j are independent. As explained in the introduction, when the events A_i are close to being mutually exclusive, we expect the standard form of the inclusion–exclusion principle to reliable approximations, and when the A_i are close to independent we expect the reordered forms to be better. In the success run problem, for $x \le O(n)$ most pairs of events are mutually exclusive, but for $x \gg n$ most pairs of events are independent. So we expect truncation of the standard inclusion–exclusion principle to work well for low x and the reordered approximations to work well for high x. This prediction is born out, and in fact improved upon, in practice. Here, however, we just report some results using the second reordered approximation. This is

$$P_n^x \approx 1 - \left(1 - \frac{1}{2^{n+1}}\right)^{x-n} \left(1 - \frac{1}{2^n}\right) \left(\frac{1 - \frac{1}{2^{n+1}} - \frac{1}{2^{n+1}}}{\left(1 - \frac{1}{2^{n+1}}\right)^2} \right)^{n(x - \frac{3}{2}n - \frac{1}{2})} \left(\frac{1 - \frac{1}{2^n} - \frac{1}{2^{n+1}}}{\left(1 - \frac{1}{2^n}\right)\left(1 - \frac{1}{2^{n+1}}\right)} \right)^n ,$$

which is of the form (14), with

$$\frac{\lambda_1}{2} = \left(1 - \frac{1}{2^n}\right)^n \left(1 - \frac{1}{2^{n+1}}\right)^{1-2n} \approx 1 - \frac{1}{2^{n+1}} - \frac{n}{2^{2n+2}} + \dots ,$$

and

$$\begin{aligned} C_1 &= \left(1-\frac{1}{2^n}\right)^{1-\frac{3}{2}n-\frac{3}{2}n^2}\left(1-\frac{1}{2^{n+1}}\right)^{n(3n-1)}\left(1-\frac{3}{2^{n+1}}\right)^{n} \\ &\approx 1+\frac{n-2}{2^{n+1}}+\frac{n(2n-3)}{2^{2n+2}}+\dots\,. \end{aligned}$$

In Figure 1 we show the (absolute) errors in the second reordered approximation for $n = 5$ and $n = 8$. Even the low x errors are reasonably small. At the peak of error P_n^x is about 0.7 in the case $n = 5$ and 0.6 in the case $n = 8$.

3c. The Birthday Problem. Assume there are D days in a year, and there is equal probability of being born on any particular day of the year. What is the probability P that in a group of N people ($0 \le N \le D$) there are (at least) two with the same birthday?
This problem has closed form solution

$$P = 1-\frac{D!}{(D-N)!D^N} = 1-\left(1-\frac{1}{D}\right)\left(1-\frac{2}{D}\right)\dots\left(1-\frac{N-1}{D}\right)$$

but it is nevertheless interesting to see what can be done with the reordered approximations to the inclusion–exclusion principle. Let A_i, $i = 1,\dots,\frac{1}{2}N(N-1)$, be the event that pair i share the same birthday. Clearly $p_i = \frac{1}{D}$. Thus the first reordered approximation gives

$$P \approx 1-\left(1-\frac{1}{D}\right)^{\frac{1}{2}N(N-1)}\,.$$

The events A_i are pairwise independent, so the second reordered approximation is the same as the first. There are, however, correlations between triples of events A_i: if persons 1 and 2 share a birthday and so do persons 2 and 3 then necessarily so do persons 1 and 3. There are $\frac{1}{6}N(N-1)(N-2)$ triples of events correlated in this way, and thus the third reordered approximation reads

$$P \approx 1-\left(1-\frac{1}{D}\right)^{\frac{1}{2}N(N-1)}\left(\frac{1-\frac{3}{D}+\frac{2}{D^2}}{\left(1-\frac{1}{D}\right)^3}\right)^{\frac{1}{6}N(N-1)(N-2)}\,.$$

In Figure 2 we show the error in the first and third reordered approximations for $D = 365$.

4 Comments and Further Directions

In this article we have given an introduction to the reordered inclusion–exclusion principle and shown it can be useful in approximating probabilities. We are hopeful that the set of applications we have presented here will be enlarged, and that also the necessary theoretical developments will appear to justify the approximation scheme. Central to this is a better understanding of the quantities $q_{ijk}, q_{ijkl}, \dots$. We suspect that the condition $q_{i_1\dots i_r} = 1$ might be interpretable as a criterion that all dependence between the r events $A_{i_1},\dots,A_{i_r}$ is determined through the dependence of subsets of $r-1$ events. In [11] Savit and Green considered an ordered sequence of dependent events $A_{i_1},\dots,A_{i_r}$ and proposed the condition $\mathbf{P}(A_{i_1}|A_{i_2}\cap\dots\cap A_{i_r}) = \mathbf{P}(A_{i_1}|A_{i_2}\cap\dots\cap A_{i_{r-1}})$ as a suitable definition of "$(r-2)$-lag dependence" of the sequence, in the sense that event A_{i_1} depends on the $(r-1)$ events before

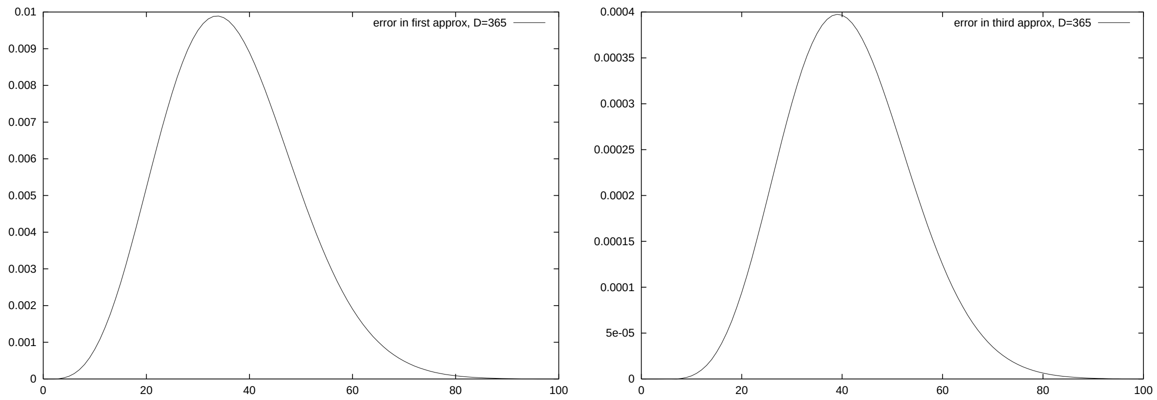

Figure 2: Absolute errors in the first and third reordered approximations for the birthday problem, $D = 365$.

it "only through" the the $(r-2)$ events before it. The condition $q_{i_1 \ldots i_r} = 1$ seems to be a symmetric version of this condition.

In addition to understanding the meaning of the $q_{ijk}, q_{ijkl}, \ldots$, it is necessary to compute the constraints on the "higher order" q's induced by knowledge of the "lower order" q's. This should allow estimation of the error in truncation of the reordered inclusion–exclusion principle. In suitable circumstances we expect the results of truncation of the reordered inclusion–exclusion principle to give bounds, not just approximations, for probabilities (see the appendix on the relationship with Janson's inequality). It would be extremely interesting if the Penrice inequalities for the derangement problem, as presented in [8], could be shown to come from a general result in probability.

A final open problem is whether the reordered inclusion–exclusion can be generalized, like the standard one [9], to give expressions for the probability of m amongst n events occurring.

Acknowledgment

J.S. wishes to thank Nati Linial and Alex Samorodnitsky for useful discussions.

Appendix: The Relationship with Janson's Inequality

Our second approximation formula is

$$\mathbf{P}(A) \approx 1 - \prod_i (1-p_i) \prod_{i<j} \left(1 + \frac{p_{ij} - p_i p_j}{(1-p_i)(1-p_j)}\right) . \tag{15}$$

Suppose that any two events A_i, A_j are either independent or positively correlated, in the sense $\mathbf{P}(A_i|A_j) > \mathbf{P}(A_i) \;\Leftrightarrow\; \mathbf{P}(A_j|A_i) > \mathbf{P}(A_j) \;\Leftrightarrow\; p_{ij} > p_i p_j$. Suppose furthermore that $p_i \leq \epsilon$ for all i, and write $\Delta = \sum_{i<j}(p_{ij} - p_i p_j) \geq 0$. Clearly all terms in the second product in (15) are greater than 1 (if the events A_i, A_j are dependent) or equal to 1 (if the events A_i, A_j are independent). Thus the product is not less than 1. To bound the product from above we

use the arithmetic-geometric inequality to get

$$\prod_{i<j}\left(1+\frac{p_{ij}-p_ip_j}{(1-p_i)(1-p_j)}\right) \le \left(1+\frac{1}{N}\sum_{i<j}\frac{p_{ij}-p_ip_j}{(1-p_i)(1-p_j)}\right)^N ,$$

where N denotes the number of pairs. Since $\frac{1}{1-p_i} \le \frac{1}{1-\epsilon}$ for all i, and $p_{ij} \ge p_ip_j$ for all i,j, it follows that

$$\prod_{i<j}\left(1+\frac{p_{ij}-p_ip_j}{(1-p_i)(1-p_j)}\right) \le \left(1+\frac{1}{N}\frac{\Delta}{(1-\epsilon)^2}\right)^N < \exp\left(\frac{\Delta}{(1-\epsilon)^2}\right) .$$

Thus we see that *our second approximation formula obeys a Janson-type inequality*, namely the approximation is $\mathbf{P}(A) \approx 1 - P\prod_i(1-p_i)$ where $1 \le P < \exp\left(\frac{\Delta}{(1-\epsilon)^2}\right)$ (c.f. [5]).

References

[1] G.Boole, *An Investigation of The Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*, Macmillan (1854), reprinted Dover (1958).

[2] N.Linial and N.Nisan, *Approximate Inclusion–Exclusion*, Combinatorica **10** (1990) 349-365; J.Kahn, N.Linial and A.Samorodnitsky, *Inclusion–Exclusion: Exact and Approximate*, Combinatorica **16** (1996) 465-477; A.Samorodnitsky, *Approximate Inclusion–Exclusion and Orthogonal Polynomials*, preprint.

[3] S.Janson, *Poisson Approximation for Large Deviations*, Random Structures and Algorithms **1** (1990) 221-229.

[4] P.Erdös and L.Lovász, *Problems and Results on 3-chromatic Hypergraphs and Some Related Questions*, in Infinite and Finite Sets, ed. A.Hajnal et al., North Holland (1975).

[5] N.Alon and J.Spencer, *The Probablistic Method*, Wiley (1992).

[6] D.Q.Naiman and H.P.Wynn, *Inclusion–Exclusion–Bonferroni Identities and Inequalities for Discrete Tube-Like Problems via Euler Characteristics*, Ann.Stat. **20** (1992) 43-76; *Abstract Tubes, Improved Inclusion–Exclusion Identities and Inequalities and Importance Sampling*, Ann.Stat. **25** (1997) 1954-1983; *Improved Inclusion–Exclusion Inequalities for Simplex and Orthant Arrangements*, JIPAM J.Inequal.Pure Appl. Math **2** (2001) article 18.

[7] J.Riordan, *An Introduction to Combinatorial Analysis*, Wiley (1958).

[8] F.F.Knudsen and and I. Skar, *On the Asymptotic Solution of a Card-Matching Problem*, Math. Mag. **69** (1996) 190-197.

[9] W.Feller, *An Introduction to Probability Theory and its Applications*, third edition, Wiley (1968).

[10] E.P.Miles, Jr., *Generalized Fibonacci Numbers and Associated Matrices*, *Amer. Math. Monthly* **67** (1960) 745-752.

[11] R. Savit and M.L.Green, *Time Series and Dependent Variables*, *Physica D* **50** (1991) 95-116 (correction: *Physica D* **55** (1992) 234); M.L.Green and R.Savit, *Dependent Variables in Broad Band Continuous Time Series*, *Physica D* **50** (1991) 521-544.