ON THE STATISTICS OF THE NUMBER OF FIXED-DIMENSIONAL SUBCUBES IN A RANDOM SUBSET OF THE N-DIMENSIONAL DISCRETE UNIT CUBE

SVANTE JANSON, BLAIR SEIDLER, AND DORON ZEILBERGER

Abstract. This paper consists of two independent, but related parts. In the first part we show how to use symbolic computation to derive explicit expressions for the first few moments of the number of implicants that a random Boolean function has, or equivalently the number of fixed-dimensional subcubes contained in a random subset of the n-dimensional cube. These explicit expressions suggest, but do not prove, that these random variables are always asymptotically normal.

The second part presents a full, human-generated proof, of this asymptotic normality, first proved by Urszula Konieczna.

1. Introduction

1.1. **Motivation.** Recall that an *implicant* of a Boolean function in n variables, $f(x_1, \dots, x_n)$ is a pure disjunction

$$x_{i_1}^{a_1} \wedge x_{i_2}^{a_2} \wedge \cdots \wedge x_{i_r}^{a_r}, \tag{1.1}$$

that implies it. Here $1 \leq i_1 < \dots i_r \leq n$, $a_1, \dots, a_r \in \{0, 1\}$, $z^1 = z$, and $z^0 = \bar{z}$ (the negation of z).

Fix r and let n vary. We are interested in the statistical distribution of the random variable *number of implicants of length* $n - r$ in a uniformly-at-random Boolean function of n variables. Clearly, when $r = 0$ it is nothing but the good old (fair) binomial distribution with 2^n *fair coin-tosses*, $B(\frac{1}{2}, 2^n)$.

Equivalently, for a random subset of the n-dimensional cube, we are interested in the statistical distribution of the number of r-dimensional subcubes properly contained in it.

We would like to have **explicit expressions**, in n, for the k^{th} moment of this random variable, for as many as possible r and k. This turns out to be a challenging *symbolic-computational* problem that we will address in the first part of this paper.

Date: Feb. 17, 2023.

Accompanied by Maple package SMCboole.txt, available from
`http://www.math.rutgers.edu/~zeilberg/mamarim/mamarimhtml/subcubes.html`.

In the second part we consider asymptotics as $n \to \infty$. It was proved by Urszula Konieczna [3] that for each fixed r, this distribution is *asymptotically normal*. We will reprove this by a somewhat different method which leads to a simple proof.

1.2. Our Random Variables.

The *sample space* is the set of subsets of $\{0,1\}^n$, that has cardinality 2^{2^n}.

Let's define our random variables formally. For a (uniformly-at-) random subset S, of $\{0,1\}^n$, and *fixed* r, define the **random variable**

$$X_r(S) := \text{ number of } r\text{-dimensional subcubes of } S. \tag{1.2}$$

For example, if $n=3$ and

$$S = \{000, 001, 010, 011, 100, 111\}, \tag{1.3}$$

we have

$$X_0(S) = 6, \quad X_1(S) = 6, \quad X_2(S) = 1, \quad X_3(S) = 0. \tag{1.4}$$

We would like to get, for as many pairs (k, r) as possible, **explicit** expressions in n, for the k-th moment of X_r, i.e. for

$$f_{kr}(n) := \mathbb{E}[{X_r}^k](n). \tag{1.5}$$

2. Explicit Expressions for Moments of the Number of Low-Dimensional Subcubes in a Random Subest of the n-dimensional cube

2.1. The Expectation and Variance.

The *first* moment, aka *expectation*, aka *mean*, aka *average*, is easy, using **linearity of expectation**.

For any specific subcube C of $\{0,1\}^n$, define the *atomic* random variable, X_C, on subsets, S, of $\{0,1\}^n$ as follows.

$$X_C(S) \; = \; \begin{cases} 1, & \text{if} \quad C \subset S; \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Let $\mathcal{C}(n,r)$ be the set of all $\binom{n}{r} 2^{n-r}$ r-dimensional subcubes of $\{0,1\}^n$. Then, of course,

$$X_r(S) \; = \sum_{C \in \mathcal{C}(n,r)} X_C(S). \tag{2.2}$$

Applying the *expectation* functional and using the **linearity of expectation**, we get that the average, let's call it $\mu_r(n)$, is

$$\mu_r(n) = \mathbb{E}[X_r] \; = \; \mathbb{E}\left[\sum_{C \in \mathcal{C}(n,r)} X_C \right] \; = \sum_{C \in \mathcal{C}(n,r)} \mathbb{E}[X_C]. \tag{2.3}$$

Now the probability that a random subset of $\{0,1\}^n$ contains an r-dimensional subcube C is $\left(\frac{1}{2}\right)^{2^r}$, since for each of its vertices, the chance of it belonging to S is $\frac{1}{2}$, and by *independence* the probability that all its 2^r vertices belong to S is indeed $\frac{1}{2^{2^r}}$. The probability that $X_C(S) = 0$ is of course $1 - \left(\frac{1}{2}\right)^{2^r}$, hence

$$\mathbb{E}[X_C] = 1 \cdot \left(\frac{1}{2}\right)^{2^r} + 0 \cdot \left(1 - \left(\frac{1}{2}\right)^{2^r}\right) = \left(\frac{1}{2}\right)^{2^r}. \tag{2.4}$$

Going back to equation (2.3) we have

$$\mu_r(n) = \sum_{C \in \mathcal{C}(n,r)} \mathbb{E}[X_C] = \sum_{C \in \mathcal{C}(n,r)} \frac{1}{2^{2^r}} = |\mathcal{C}(n,r)| \cdot \frac{1}{2^{2^r}} = \frac{\binom{n}{r} 2^{n-r}}{2^{2^r}}. \tag{2.5}$$

In a beautiful paper, Thanatipanonda [5] derived an explicit expression for the general second moment, for every r-dimensional cube.

Thanatipanoda's General Formula for the Second Moment:

$$\mathbb{E}[X_r^2] = \sum_{i=0}^{r} \frac{n! 2^{n-i}}{i!(r-i)!^2(n-2r+i)! 2^{2^{r+1}}} \cdot (2^{2^i} - 1) + \frac{[\binom{n}{r} 2^{n-r}]^2}{2^{2^{r+1}}}, \tag{2.6}$$

from which immediately follows, using $[\mathbb{E}(X_r - \mu_r(n))^2] = \mathbb{E}[X_r^2] - \mu_r(n)^2$, the following formula.

Thanatipanoda's General Formula for the Variance:

$$\mathrm{Var}(X_r) = \sum_{i=0}^{r} \frac{n! 2^{n-i}}{i!(r-i)!^2(n-2r+i)! 2^{2^{r+1}}} \cdot (2^{2^i} - 1). \tag{2.7}$$

Note that the variance is a polynomial in $(n, 2^n)$ of degree $2r$ in n and degree 1 in 2^n.

2.2. **Higher Moments.**

2.2.1. *Edges.* Thanatipanonda was unable to get such a general formula for higher moments, but did get $\mathbb{E}[X_1^3]$, from which he immediately deduced that the third central moment (or third-moment-about-the-mean) of X_1 is

$$\mathbb{E}[(X_1 - \mu_1(n))^3] = \frac{3n^3 2^n}{64}. \tag{2.8}$$

Using the symbolic-computational algorithms to be described in the next section, we managed to derive the following explicit formulas

$$\begin{aligned}\mathbb{E}[(X_1 - \mu_1(n))^4] = \frac{n 2^n}{1024}\big(&12n^3 2^n + 12n^2 2^n + 40n^3 \\ &+ 3n 2^n - 48n^2 + 12n - 16\big),\end{aligned} \tag{2.9}$$

$$\mathbb{E}[(X_1 - \mu_1(n))^5] = \frac{5n^3 2^n}{1024}\left(6n^2 2^n + 3n 2^n + 4n^2 - 24n + 8\right), \tag{2.10}$$

$$\mathbb{E}[(X_1-\mu_1(n))^6] = \frac{n2^n}{32768}\cdot\big(120n^5\,(2^n)^2 + 180n^4\,(2^n)^2 + 1920n^5 2^n + 90n^3\,(2^n)^2 - 840n^4 2^n - 1792n^5 + 15n^2\,(2^n)^2 - 360n^3 2^n - 5280n^4 - 300n^2 2^n + 3840n^3 - 240n2^n + 3840n^2 - 6720n + 4864\big). \quad (2.11)$$

It follows that the **scaled moments about the mean** for the third, fourth, fifth, and sixth moments, converge, as $n \to \infty$, to $0, 3, 0, 15$ respectively, the respective moments of the normal distribution, indicating that the random variable X_1 (the number of edges contained in S) is **probably** asymptotically normal. To fully prove asymptotic normality, of course, we need to prove it for all moments, not just for the first six.

2.2.2. *Squares.* We only managed to get explicit expressions for the third and fourth moments for X_2.

$$\mathbb{E}[(X_2-\mu_2(n))^3] = \frac{2^n n\,(n-1)}{32768}\left(9n^4 + 6n^3 + 21n^2 - 16n - 34\right), \quad (2.12)$$

$$\mathbb{E}[(X_2-\mu_2(n))^4] = \frac{2^n n\,(n-1)}{4194304}\cdot\big(12n^6 2^n + 12n^5 2^n + 520n^6 + 24n^4 2^n - 24n^5 - 12n^3 2^n + 1272n^4 - 9n^2 2^n - 840n^3 - 27n2^n - 5232n^2 - 2768n + 240\big). \quad (2.13)$$

2.2.3. *3-dimensional cubes.* We only managed to get an explicit expression for the third moment for X_3.

$$\mathbb{E}[(X_3-\mu_3(n))^3] = \frac{2^n n\,(n-1)\,(n-2)}{2415919104}\big(14n^6 + 24n^5 + 479n^4 + 2046n^3 + 6779n^2 + 15444n - 23112\big). \quad (2.14)$$

2.3. **Our Method.** Obviously we did not derive these formulas by hand. We had to teach our computer how to find them. It also uses *linearity of expectation*, but with higher moments things get very complicated. Recall that

$$X_r(S) = \sum_{C\in\mathcal{C}(n,r)} X_C(S). \quad (2.15)$$

Hence, the k-th moment is

$$\mathbb{E}\left[(X_r)^k\right] = \mathbb{E}\left[\Big(\sum_{C\in\mathcal{C}(n,r)} X_C(S)\Big)^k\right]$$

$$= \sum_{[C_1,\ldots,C_k]\in\mathcal{C}(n,r)^k} \mathbb{E}[X_{C_1}X_{C_2}\cdots X_{C_k}]. \tag{2.16}$$

So we sum over all $(\binom{n}{r}2^{n-r})^k$ members of $\mathcal{C}(n,r)^k$. Since the product $X_{C_1}(S)X_{C_2}(S)\cdots X_{C_k}(S) = 1$ if **each** of $C_1, C_2, \ldots, C_k$ is properly included in S (i.e. if each vertex in $C_1 \cup C_2 \cup \ldots C_k$ belongs to S), and 0 otherwise, the contribution, or weight, due to each such term is

$$\text{Weight}([C_1, C_2, \ldots, C_k]) = \mathbb{E}\left[X_{C_1}X_{C_2}\cdots X_{C_k}\right] = \frac{1}{2^{|C_1\cup C_2\cup\ldots C_k|}}. \tag{2.17}$$

2.3.1. *Data Structure.* Every r-dimensional subcube of $\{0,1\}^n$ has the form

$$C = \{(x_1,\ldots,x_n) \in \{0,1\}^n \,|\, x_{i_1} = \alpha_{i_1}, \ldots, x_{i_{n-r}} = \alpha_{i_{n-r}}\}, \tag{2.18}$$

for some $1 \le i_1 < i_2 < \cdots < i_{n-r} \le n$ and $(\alpha_{i_1}, \ldots \alpha_{i_{n-r}}) \in \{0,1\}^{n-r}$. A good way to represent it on a computer is as a row-vector of length n, in the *alphabet* $\{0,1,*\}$, where the entries corresponding to $i_1, i_2, \ldots, i_{n-r}$ have $\alpha_{i_1}, \ldots . \alpha_{i_{n-r}}$ respectively and the remaining r entries are filled with **wild cards**, denoted by $*$.

For example, if $n = 7$ and $r = 3$, the 3-dimensional cube

$$\{(x_1,\ldots,x_7) \in \{0,1\}^7 \,|\, x_2 = 1, x_4 = 1, x_5 = 0, x_7 = 1\}, \tag{2.19}$$

is represented by

$$*1*10*1. \tag{2.20}$$

We are trying to find a weighted count of **ordered** k-tuples of r-dimensional subcubes. The natural data structure for these is the set of k by n matrices in the 'alphabet' $\{0,1,*\}$ where every row has exactly r 'wildcards', *.

Let's call this set of matrices, that correspond to $\mathcal{C}(n,r)^k$, $\mathcal{C}(n,r,k)$.

For any **specific**, numeric n, there are 'only' $(2^{n-r}\binom{n}{r})^k$ of these matrices, and for each and every one of them one can find the cardinality of the union of the corresponding subcubes, let's call it v, and add to the running sum $\frac{1}{2^v}$. But we want to do it for **symbolic** n, i.e. for 'all' n. We will soon see how, for each *specific* (numeric) r and k this can be done *in principle*, but only for relatively small r and k *in practice.* But let's try and push it as far we can. An interesting consequence of our algorithm is the precise degree in n and 2^n of the expression for $\mathbb{E}[X_r^k](n)$.

2.3.2. *The Kernel.* A key object in our approach is the **kernel**. Given a $k \times n$ matrix M in the alphabet $\{0,1,*\}$ let's call a column **active** if it contains at least one '$*$'. Note that the matrix has exactly $k \cdot r$ '$*$'s, hence the number of **active columns**, let's call it a, is between r and $k \cdot r$. [1] The **kernel** of M is the submatrix of M consisting of its active columns.

[1]More generally, if we want to find an expression for the *mixed moment* $\mathbb{E}[X_{r_1}\cdots X_{r_k}]$ the number of active columns is between $\max(r_1,\ldots,r_k)$ and $r_1 + \cdots + r_k$.

Let $\mathcal{C}_a(n,r,k)$ be the subset of $\mathcal{C}(n,r,k)$ matrices with exactly a active columns. We will call such a matrix in **canonical form** if the active columns are occupied by the a leftmost columns (i.e. its kernel is contiguous starting in the first column). Let's denote by $\overline{\mathcal{C}}_a(n,r,k)$ the set of such matrices in canonical form. Obviously, there are $\binom{n}{a}$ ways to choose which of the n columns are active and hence

$$\begin{aligned}\text{Weight}(\mathcal{C}(n,r,k)) &= \sum_{a=r}^{rk} \text{Weight}(\mathcal{C}_a(n,r,k)) \\ &= \sum_{a=r}^{rk} \binom{n}{a} \text{Weight}(\overline{\mathcal{C}}_a(n,r,k)). \qquad (2.21)\end{aligned}$$

For any set, S, Weight(S) is the sum of the weights of its members. Note that this has degree rk in n.

It remains to do a *weighted-count*, where every matrix gets 'credit' $1/2^v$, where v is the cardinality of the union of the corresponding subcubes represented by the k rows, for the set $\overline{\mathcal{C}}_a(n,r,k)$, of matrices in canonical form. Note that there are only **finitely many** choices for the a leftmost columns, i.e. the set of $k \times a$ matrices in the alphabet $\{0,1,*\}$ with the property that every column has at least one '$*$', and every row has exactly r '$*$'s. These can be divided into *equivalence classes* obtained by permuting rows and columns and transposing 0 and 1 in any given column. Once these are sorted into equivalence classes, one needs only examine one representative, and then multiply the weight by the cardinality of the class.

But what about the $n-a$ rightmost columns? There are $2^{k(n-a)}$ possible submatrices; the alphabet here is $\{0,1\}$. Almost all of these have distinct rows, more precisely,

$$\binom{2^{n-a}}{k} k! \qquad (2.22)$$

of them, and these will produce the smallest possible weight in conjunction with any kernel. The other extreme is that all the rows of the submatrix consisting of the $n-a$ rightmost columns are identical, and then there are only 2^{n-a} choices to fill them in.

In general, every such member of $\overline{\mathcal{C}}_a(n,r,k)$, determines a **set partition** of the set of rows $\{1,\dots,k\}$, where two rows are *roomates* if they have the same last $n-a$ entries. If that set-partition has m members $1 \le m \le k$, then the number of choices of assigning **different** $0-1$ vectors of length $n-a$ to each of the parts of the set-partition is

$$\binom{2^{n-a}}{m} m!. \qquad (2.23)$$

Now for each a and for each set-partition, we let the computer generate the **finite** set of $k \times a$ matrices in the alphabet $\{0,1,*\}$. Each of the members of the set partition has its own submatrix, and we ask our computer to kindly find the number of vertices in the corresponding union of subcubes corresponding to each member of the examined set partition. Since they are disjoint, we add them up, getting v for that particular pair (matrix, set-partition), giving credit $1/2^v$.

2.3.3. *Implementation.* All this is implemented in the Maple package `SMCboole.txt`, available from:
`https://sites.math.rutgers.edu/~zeilberg/tokhniot/SMCboole.txt`

In particular '`Moms(A,n);`', for any list of non-negative integers $A = [r_1, \ldots r_k]$ gives you the mixed moment $\mathbb{E}[X_{r_1} \cdots X_{r_k}]$. For example, to get the third moment of the number of edges (i.e. 1-dimensional subcubes) type `Moms([1,1,1],n);` , which very quickly returns:

$$\frac{2^n n^2 \left((2^n)^2 n + 12\, 2^n n + 6\, 2^n + 24n \right)}{512}. \tag{2.24}$$

To get the third moment of the number of squares (i.e. 2-dimensional subcubes), type `Moms([2,2,2],n);` , getting

$$\begin{aligned} \frac{2^n n\,(n-1)}{2097152} \cdot \Big(& (2^n)^2 n^4 - 2\,(2^n)^2 n^3 + 48\, 2^n n^4 + (2^n)^2 n^2 \\ & + 576 n^4 + 24\, 2^n n^2 + 384 n^3 - 72\, 2^n n \\ & + 1344 n^2 - 1024 n - 2176 \Big). \end{aligned} \tag{2.25}$$

The third moment of the number of 3-dimensional cubes takes a bit longer, and we were unable to compute the fourth moment of the number of 3-dimensional cubes; it took too much time and too much space.

More informative for *statistical purposes* are the **central moments**, $\mathbb{E}[(X_r - \mu_r(n))^k]$, that Maple easily derives, using *linearity of expectation* from the pure moments. The function call for this is `MOMrk(r,k,n);` where `r` and `k` are *numeric* but n is a **symbol** denoting the dimension of the ambient cube.

To get the explicit expression given above for the third through sixth moments for the number of edges, the third and fourth moments for the number of squares, and the third moment for the number of 3-dimensional subcube (all *about the mean*) we typed:
`MOMrk(1,3,n);` , `MOMrk(1,4,n);` , `MOMrk(1,5,n);` , `MOMrk(1,6,n);` ,
`MOMrk(2,3,n);` , `MOMrk(2,4,n);` ,
`MOMrk(3,3,n);`
respectively. To our chagrin, '`MOMrk(3,4,n);`' took too long.

2.3.4. *Consequence of the algorithm.* : The k-th moment of X_r is a bivariate polynomial in $(n, 2^n)$ of degree $k\,r$ in n and degree k in 2^n .

This raises the *theoretical* possibility (in God's computer) of finding these expressions by **pure brute force**. The generic polynomial in $(n, 2^n)$ of degree kr in n and degree k in 2^n has $(1+kr)(1+k)$ 'degrees of freedom'. So using *undetermined coefficients* we need to generate a table of $\mathbb{E}[X_r^k](n)$

for $1 \leq n \leq (1+kr)\cdot(1+k)$. After gathering the data, we use linear algebra to solve a system of $(1+kr)\cdot(1+k)$ equations with that many unknowns. For each specific $n = n_1$ there are 'only' $2^{2^{n_1}}$ subsets, and for each of them we can ask how many r-dimensional subcubes do they contain, raise it to the k-th power and take the average. Alas, 2^{2^5} is already big enough, so only God's computer, with practically infinite time and space, can carry this brute force approach.

3. Asymptotic Normality

As said in the introduction, asymptotic normality of X_r was proved by Urszula Konieczna [3]. The random variable X_r depends on n, and for emphasis, we denote it in the sequel by $X_{n,r}$. We may then state the result as follows, where $N(0,1)$ denotes the standard normal distribution.

Theorem 3.1 ([3])**.** *For every fixed $r \geq 0$, we have as $n \to \infty$*

$$\frac{X_{n,r} - \mu_r(n)}{\sqrt{\operatorname{Var}(X_{n,r})}} \xrightarrow{\mathrm{d}} N(0,1) \tag{3.1}$$

with convergence in distribution and of all moments.

As already noted at the very beginning of this paper, the case $r = 0$ is well known, since then the variables X_C are independent and this becomes an instance of the central limit theorem for binomial variables, known since a pamphlet by de Moivre in 1733. (This case can be used as a sanity check in the arguments below.)

The proof in [3] is based on the method of moments. We will in this section survey how a variation of this method, based on **cumulants** and **dependency graphs** leads to a quick proof with almost no calculations.

Remark 3.2. Urszula Konieczna [3] considered a more general case where each subcube appears with probability $p \in (0,1)$, where $p = p(n)$ may depend on n, and showed asymptotic normality for a large range of $p(n)$. A special case, not including the case $p = \frac{1}{2}$ treated here, was earlier shown by Karl Weber [6]. This extension can also be treated by the method below, but for simplicity we continue to consider only the case $p = \frac{1}{2}$ discussed in Section 2.

The aim is thus to show that the leading terms in the expressions for the central moments $\mathbb{E}[(X_{n,r} - \mu_r(n))^k]$ are what we would expect from a normal distribution, or, more formally, that the scaled central moments

$$\mathbb{E}\left[\left(\frac{X_{n,r} - \mu_r(n)}{\sqrt{\operatorname{Var}(X_{n,r})}}\right)^k\right] \tag{3.2}$$

converge as $n \to \infty$ to the corresponding moment $\mathbb{E}[Z^k]$ of a standard normal random variable Z, for every fixed $r \geq 0$ and $k \geq 1$. Then, the random variables $X_{n,r}$ are asymptotically normal by the method of moments.

We thus want to find the leading term of the moment (3.2). In Section 2, we computed the central moments from the pure moments. This involves massive cancellation of high-order terms, and is less suitable for a (human) proof for general k.

As a first step, we may modify (2.16) and instead expand the central moment as

$$\mathbb{E}\big[(X_{n,r} - \mu_r(n))^k\big] = \sum_{[C_1,\dots,C_k]\in\mathcal{C}(n,r)^k} \mathbb{E}\big[(X_{C_1} - \mu)\cdots(X_{C_k} - \mu)\big], \quad (3.3)$$

where $\mu = \mathbb{E}[X_C]$ is given by (2.4). Note that two variables X_{C_1} and X_{C_2} are independent if $C_1 \cap C_2 = \emptyset$. In particular, for $k = 2$, all such terms in the sum in (3.3) vanish; this leaves only $\Theta(n^{2r}2^n)$ terms of the $\Theta\big(n^{2r}2^{2n}\big)$ terms in the full sum, or in (2.16), which is reflected in the leading terms in (2.6) and (2.7).

For higher moments, we can obtain a further reduction by considering **cumulants** (also called **semiinvariants**) instead of moments. We describe this briefly, referring to e.g. [4] or [2, pp. 145–149] for details. In general, if Y is a random variable with moments $m_k := \mathbb{E}[Y^k]$ and moment generating function $\sum_{k=0}^{\infty} m_k t^k/k!$, then the cumulants $\varkappa_k = \varkappa_k(Y)$ are the coefficients of the generating function

$$\sum_{k=1}^{\infty} \varkappa_k \frac{t^k}{k!} := \log\left(\sum_{k=0}^{\infty} m_k \frac{t^k}{k!}\right). \quad (3.4)$$

(The generating functions can be regarded as formal power series.) Moreover, this generalizes to mixed cumulants $\varkappa_k(Y_1,\dots,Y_k)$ of several random variables. This means that there are algebraic relations: each cumulant $\varkappa_k$ is a polynomial in the moments m_ℓ of order $\ell \le k$, and conversely, and this generalizes to mixed cumulants and moments. For example, $\varkappa_1 = m_1 = \mathbb{E}\,Y$, and $\varkappa_2 = m_2 - m_1^2 = \operatorname{Var}(Y)$, and the mixed cumulant $\varkappa(Y_1,Y_2) = \mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\,\mathbb{E}[Y_2]$, the covariance of Y_1 and Y_2.

A convenient property of cumulants is that a normal distribution $N(\mu,\sigma^2)$ has moment generating function $\exp\big(\mu t + \frac{\sigma^2}{2}t^2\big)$, and thus by (3.4) cumulants $\varkappa_k = 0$ for $k \ge 3$. Consequently, to show that a normalized sequence of random variables is asymptotically normal, it suffice to show that each cumulant $\varkappa_k$ with $k \ge 3$ converges to 0.

The mixed cumulants are multilinear, and thus we have an analog of (2.16) and (3.3):

$$\varkappa_k\big(X_{n,r}\big) = \varkappa\big(X_{n,r},\dots,X_{n,r}\big) \;=\; \sum_{[C_1,\dots,C_k]\in\mathcal{C}(n,r)^k} \varkappa\big(X_{C_1},\dots,X_{C_k}\big). \quad (3.5)$$

We can here reduce the number of terms in the sum by a general property of cumulants: If there is a partition $\{1,\dots,k\} = I \cup J$ into two non-empty sets I and J such the two families of random variables $(Y_i \mid i \in I)$ and $(Y_j \mid j \in J)$ are independent of each other, then the mixed cumulant $\varkappa(Y_1,\dots,Y_k) = 0$. (Informally, this means that the kth cumulant $\varkappa_k(X_{n,r})$ can be seen as some kind of measure of kth order dependencies in the sum (2.2), with lower order dependencies removed.)

As noticed above, each of our variables X_C is independent of most of the others. We need to keep track of not just pairwise independence, but also independence between families of such variables; it is then convenient to use the notion of **dependency graphs**.

Definition. Let $(Y_\alpha)_{\alpha\in\mathcal{A}}$ be a family of random variables, with some arbitrary index set $\mathcal{A}$. A **dependency graph** for this family is a graph Γ with vertex set $\mathcal{A}$, such that if I and J are two disjoint subsets of $\mathcal{A}$ such that there is no edge in Γ with one endpoint in I and the other in J, then the two families $(Y_\alpha \mid \alpha \in I)$ and $(Y_\beta \mid \beta \in J)$ are independent of each other.

Lemma 3.3. *Let $\Gamma(n,r)$ be the graph with vertex set $\mathcal{C}(n,r)$ and an edge between two distinct cubes $C, C' \in \mathcal{C}(n,r)$ iff $C \cap C' \neq \emptyset$. Then $\Gamma(n,r)$ is a dependency graph for the family $(X_C \mid C \in \mathcal{C}(n,r))$.*

Proof. Let I and J be disjoint subsets of $\mathcal{C}(n,r)$ with no edge between I and J. Then $V_I := \bigcup_{C\in I} C$ and $V_J := \bigcup_{C\in J} C$ are two disjoint subsets of $\{0,1\}^n$. Thus the two random sets $S\cap V_I$ and $S\cap V_J$ are independent. Since the random variables $(X_C \mid C \in I)$ only depend on $S \cap V_I$, and similarly for J, it follows that the families $(X_C \mid C \in I)$ and $(X_C \mid C \in J)$ are independent. ∎

We can now use the dependency graphs $\Gamma(n,r)$ and the machinery above to estimate cumulants of $X_{n,r}$ and show that $\varkappa_k\big(X_{n,r}\big) = o\big((\operatorname{Var} X_{n,r})^{k/2}\big)$ as $n \to \infty$ for every fixed $r \geq 0$ and $k \geq 3$. However, we do not have to do these calculations, since they already have been done under general hypotheses, leading to convergence theorems that are easy to apply in situations like ours where there is a rather sparse dependency graph. We use here the following theorem, taken (with minor changes in notation) from [1, Theorem 2, with moment convergence by its proof and Theorem 1]. See e.g. [2, Theorems 6.18 and 6.20] for some related theorems (proved by the same cumulant method) that also can be used to show Theorem 3.1, and [2, Theorem 6.33] for another related theorem (proved by a different method) that also yields the asymptotic normality in Theorem 3.1 (but not immediately moment convergence).

Theorem 3.4 ([1])**.** *Suppose that, for each n, $(Y_{n,\alpha} \mid \alpha \in \mathcal{A}_n)$ is a family of bounded random variables: $|Y_{n,\alpha}| \leq A_n$. Let $N_n := |\mathcal{A}_n|$. Suppose further that Γ_n is a dependency graph for this family, and let M_n be the maximal degree of Γ_n (assuming this is not 0, in which case we let $M_n := 1$). Let $W_n = \sum_{i\in\mathcal{A}_i} Y_{n,i}$ and $\sigma_n^2 := \operatorname{Var}(W_n)$. If there exists a positive integer m such that*

$$(N_n/M_n)^{1/m} M_n A_n/\sigma_n \to 0 \qquad \text{as } n\to\infty, \tag{3.6}$$

then

$$\big(W_n - \mathbb{E}\, W_n\big)/\sigma_n \xrightarrow{\mathrm{d}} N(0,1) \qquad \text{as } n \to \infty, \tag{3.7}$$

with convergence in distribution and with all moments.

Proof of Theorem 3.1. Fix $r \geq 0$. We let c_1, c_2, c_3 denote unimportant positive constants that may depend on r, but not on n.

We use Theorem 3.4, with $Y_{n,C} := X_C$ for $C \in \mathcal{A}_n := \mathcal{C}(n,r)$, so $W_n = X_{n,r}$. We use Lemma 3.3 and take the dependency graph $\Gamma_n = \Gamma(n,r)$. Then

$$N_n = |\mathcal{C}(n,r)| = \binom{n}{r} 2^{n-r} \leq n^r 2^n, \tag{3.8}$$

and, by (2.7) (where $i = 0$ yields the leading term),

$$\sigma_n^2 := \operatorname{Var}(X_{n,r}) \geq c_1 n^{2r} 2^n. \tag{3.9}$$

Furthermore, we may take

$$A_n := 1. \tag{3.10}$$

Moreover, from the definition of the graph $\Gamma(n,r)$ follows that it is regular, with degree

$$M_n \leq 2^r \binom{n}{r} \leq c_2 n^r. \tag{3.11}$$

We choose $m = 3$ in Theorem 3.4. (This is the standard case, which corresponds to the method of moment without extra arguments, see the proof in [1].) Then (3.8)–(3.11) yield

$$(N_n/M_n)^{1/m} M_n A_n/\sigma_n = N_n^{1/3} M_n^{2/3} A_n/\sigma_n \leq c_3 \frac{n^r 2^{n/3}}{n^r 2^{n/2}} \to 0 \tag{3.12}$$

as $n \to \infty$, so (3.6) holds. Consequently, Theorem 3.4 applies and shows Theorem 3.1. □

4. Conclusion

In the first part, we demonstrated the power of computers to *automatically* derive complicated explicit expressions, all polynomials in n and 2^n, for low moments of the random variable *number of fixed-size implicants* of a Boolean function of n variable, or equivalently, the number of fixed-dimensional subcubes of a random subset of the n-dimensional unit cube. These implied that the scaled first few moments for low-dimensional subcubes tend, as n goes to infinity, to those of the Normal Distribution, suggesting, but by no means proving, that for each fixed dimension, this random variable is *asymptotically normal*.

In the second part, we demonstrated the power of *human ingenuity*, and the traditional lore of probability and statistics, by presenting a simple proof of asymptotic normality, first proved in [3].

References

[1] Svante Janson, Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. *Annals of Probability* **16** (1988), 305–312.

[2] Svante Janson, Tomasz Łuczak and Andrzej Ruciński, *Random Graphs.* Wiley, New York, 2000.

[3] Urszula Konieczna, Asymptotic normality of subcubes in random subgraphs of the n-cube. *Discrete Mathematics* **121** (1993), 117–122.

[4] V. P. Leonov and Albert Nikolaevich Shiryaev, On a method of semi-invariants. *Teor. Veroyatnost. i Primenen* **4** (1959), 342–355. (Russian.) English transl: *Theor. Probability Appl.* **4** (1959), 319–329.

[5] Thotsaporn Aek Thanatipanonda, *Reviews of Symbolic Moment Calculus*, arXiv:2003.11749. `https://arxiv.org/abs/2003.11749`

[6] . Karl Weber, Poisson convergence in the n-cube. *Math. Nachr.* **131** (1987), 49–57.

Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson [At] math [Dot] uu [Dot] se`

URL: `http://www.math.uu.se/~svante/`

Department of Mathematics, Rutgers University (New Brunswick), Piscataway NJ 08854, USA

E-mail address: `blair [At] math [Dot] rutgers [Dot] edu`

Department of Mathematics, Rutgers University (New Brunswick), Piscataway, NJ 08854, USA

E-mail address: `zeilberg [At] math [Dot] rutgers [Dot] edu`