On the moments of Bernoulli Summable Random Variables

Pavel Shuldiner
Department of Statistics & Actuarial Science, University of Waterloo

Wayne Oldford
Department of Statistics & Actuarial Science, University of Waterloo

June 12, 2020

2 Introduction

The binomial distribution is a classic example of a discrete probability distribution that students encounter. We say that the random variable X has the binomial distribution Binomial(n, p) if it can be written as a summation of n independent Bernoulli random variables where each one has success probability p. That is, if we may write $X = \sum_{i=1}^{n} Y_i$ with $Y_i \sim$ Bernoulli(p).

Poisson ([9]) was the first to consider generalizing the Binomial(n, p) distribution. In particular, he investigated the family of random variable $X = \sum_{i=1}^{n} Y_i$ where $Y_i \sim$ Bernoull(p_i). That is, he allowed the probability of success to vary between different Bernoulli random variables. Today, a random variable X with this distribution is known as a *Poisson binomial* random variable and is often denoted by $X \sim PB(p_1, \ldots, p_n)$.

In practice, the Poisson binomial distribution may apply when one is interested in counting the total number of events occurring, where each individual is independent but with possibly different underlying probability. For instance, in actuarial science ([5]), one may model the total payout for life insurance using the Poisson binomial distribution. If an individual payout for death is C and 0 otherwise and insured individual i has a probability of death p_i independently of other policyholders, the total payout is $CX = \sum_{i=1}^{n} Y_i$, where

- $Y_i \sim$ Bernoulli(p_i) is the probability of individual i passing away during their policy duration,
- $X = \sum_{i=1}^{n} Y_i$ is the total number of policyholders that have passed away at the time of the payout.

In this paper, we generalize the Poisson binomial model and allow various dependencies between the Bernoulli trials, which we call *Bernoulli summable* random variables. We exploit combinatorial ideas such as the Multinomial Theorem and the Principle of Inclusion-Exclusion to derive the moments and central moments of these random variables, which enables us to produce a Central Limit the multinomial theorem to demonstrate that if $X = \sum_{i \in \mathcal{I}} Y_i$ for some countable indexing set $\mathcal{I}$ and a collection of $Y_i \sim$ Bernoulli(p_i), then the moments of X are uniquely expressible as a linear combination of the collection of probabilities $\{P(Y_{i_1} = 1, \ldots, Y_{i_m} = 1) : \{i_1, \ldots i_m\} \subseteq \mathcal{I}\}$. We establish a connection between the moments obtained through our combinatorial link and the underlying distribution of X.

Many well-studied random variables are often characterized by a small number of parameters. For instance, the first and second moments uniquely define the distribution of a Normal random variable. In measure and probability theory, the so-called generalized problem of moments seeks to find a collection $\{f_i\}$ of functions for which the distribution of a given class of random variables X is uniquely described by the expected values of $\{E[f_i(X)]\}$.

In this paper, we examine a technique for deriving the distribution of a finite support random variable with known moments. In practice, one often has the distribution of a random variable X and hence can determine the moments

$$E(X^k) = \sum_{x \geq 0} x^k P(X = x).$$

However, for some random variables (even those with finite support), the distribution is so complex that we must resort to estimation via simulation. We present a class of finite random variables, which we call Bernoulli summable, that has a sometimes difficult-to-express distribution but with moments that can be derived using straightforward combinatorial techniques. The closed-form expressions for the moments allow us to address the problem of moments in the case of this class of random variables. That is, for a Bernoulli summable random variable X, the collection $\{f_i\}$ which addresses the problem of moments is $\{f_i(x) = x^i\}$, for a suitable collection of indices i.

The structure of the paper is as follows. First, we establish a relationship between moments of a finite support random variable and its distribution using Vandermonde matrices. Next, we propose and prove a generalization of the multinomial theorem and apply it to commutative idempotents. Finally, we use the two preceding results to describe the moments of a particular class of finitely supported random variables, thus obtaining a procedure for the exact evaluation of their distributions.

As an application, we apply our results to clique counting within homogeneous Erdos-Renyi random graphs. Thus, we obtain the moments of clique counts, prove a Lyapunov-type Central Limit Result and construct a hypothesis test based on the number of cliques.

3 Bernoulli summable random variables

Definition 3.1. We say that an infinite sequence $(Y_1, Y_2, \ldots)$ of random variables is **exchangeable** if the joint distribution of $(Y_{\sigma(1)}, Y_{\sigma(2)}, \ldots)$ is equal to that of $(Y_1, Y_2, \ldots)$ for any finite permutation σ. We say that the collection of discrete random variables $(Y_1, \ldots, Y_n)$ is $n-$**exchangeable** if

$$P(Y_1 = y_1, Y_{\sigma(2)} = y_2, \ldots, Y_n = y_n) = P(Y_{\sigma(1)} = y_1, Y_{\sigma(2)} = y_2, \ldots, Y_{\sigma(n)} = y_n),$$

for any permutation $\sigma \in S_n$ in the symmetric group on n variables.

As $n-$exchangeability applies to all permutations of length n and S_n contains an embedding of S_ℓ for all $\ell < n$, any collection that is $n-$exchangeable is also $\ell-$exchangeable for all $\ell < n$.

Definition 3.2. Let $Y = (Y_1, \ldots, Y_n)$ be a collection of discrete random variables and fix a positive integer $1 < m < n$. We call Y $m-$**symmetric** if

$$(Y_1, Y_2, \ldots, Y_m) \stackrel{D}{=} (Y_{i_1}, Y_{i_2}, \ldots, Y_{i_m}),$$

for any subset $\{i_1, i_2, \ldots, i_m\}$ from $\{1, 2, \ldots, n\}$.

The classic Binomial distribution Binomial(n, p) can be viewed as a sum of n independent, identically distributed Bernoulli(p) trials. Thus, it is an instance of both an $n-$exchangeable distribution and $m-$symmetric distribution for all $m < n$. Moreover, one can view the Binomial(n, p) distribution as a special case of the so-called Poisson Binomial distribution.

Definition 3.3. We say that a random variable X has the **Poisson Binomial distribution** $PB(p_1, \ldots, p_n)$ if X can be represented as a sum $X = \sum_{i=1}^{n} Y_i$, where $Y_i \sim$ Bernoulli(p_i) are independent.

In this paper, we generalize the Poisson Binomial distribution by investigating random variables where the assumption of independence of summands is relaxed, and under suitable conditions, we allow for the index set in the summation to be countably infinite.

Definition 3.4. A random variable X is **Bernoulli summable** if X a finite random variable that may be written as a sum of Bernoulli random variables Y_i where $i \in \mathcal{I}$ for some indexing set $\mathcal{I}$. That is, suppose that we may write

$$X = \sum_{i \in \mathcal{I}} Y_i,$$

where all of the $Y_i \sim$ Bernoulli(p_i).

We begin with a few straightforward examples of Bernoulli summable random variables.

Example 3.5 (*Degenerate random variable*)**.** For $i = 1, \ldots n$, let Y_i be the constant 1. Thus, Y_i is Bernoulli(1) and the degenerate random variable

$$X = \sum_{i=1}^{n} Y_i = n,$$

is an instance of a Bernoulli summable random variable.

Example 3.6 (*Hypergeometric distribution*)**.** Let $\mathcal{S}$ be a random sample without replacement of size n from a population with N individuals where g of them have an attribute of interest. If X denotes the number of individuals in the sample with the desired attribute, then we may write

$$X = \sum_{i=1}^{n} Y_i,$$

where Y_i is the indicator random variable recording if the $i-$th sample member has the desired attribute. We recall that in such a setting, X has the hypergeometric distribution with parameters (N, g, n).

Our initial motivation for investigating Bernoulli summable random variables was clique counting on Erdos Renyi random graphs. We say that G is a homogeneous Erdos-Renyi random graph with parameters (n, p) if G has n vertices where an edge between every pair of vertices has an equal probability p of appearing, independently of all other edges. In this paper, we write $G \sim G(n, p)$ when this is the case and we borrow the convention from combinatorics and let $\mathcal{N}_n$ denote the set of the first n natural numbers $\mathcal{N}_n := \{1, \ldots, n\}$.

Example 3.7 (*Degenerate random variable*)**.** The most trivial instance of a Bernoulli summable random variable is one where all of the indicators are 1 (or 0). Let $X_1 = \sum_{i \in \mathcal{N}_n} Y_i$ where $Y_i \sim$ Bernoulli(1), for all i. For instance, one may think of X_1 as the number of vertices in $G(n, p)$. Since we assume that we include all vertices, this is given by

$$X_1 = \sum_{i=1}^{n} Y_{i \text{ is a vertex}} = n,$$

where $Y_{i \text{ is a vertex}} \sim$ Bernoulli(1).

Example 3.8 (*Edge count on Erdos-Renyi graphs*)**.** Let G be a realization of $G(n, p)$ and denote by $\mathcal{I}$ the set of all 2-subsets of $\mathcal{N}_n$. For $i \in \mathcal{I}$, let Y_i be the indicator random variable recording if edge i is included in G. If $X := \sum_{i \in \mathcal{I}} Y_i$, then X is the Bernoulli summable random variable counting the number of edges in $G(n, p)$. Since $G(n, p)$ is an instance of a homogeneous Erdos-Renyi model for graphs, the Y_i are independent, identically distributed Bernoulli(p) random variables and thus $X \sim \text{Binomial}\left(\binom{n}{2}, p\right)$.

The random variable X in the preceding example is a simple instance of a more general class of random variables on random graphs which motivates the study of Bernoulli summable random variables.

Example 3.9 ($r-$*clique counting in* $G(n, p)$)**.** If X_r is the random variable counting the number of cliques in $G \sim G(n, p)$, then letting $\mathcal{I}$ denote the set of all $r-$subsets of $\mathcal{N}_n = \{1, 2, \ldots, n\}$ and Y_A record if the $r-$set A is a clique we have that

$$X_r = \sum_{A \subseteq \mathcal{N}_n \text{ subset of size } r} Y_A$$

and hence X_r is Bernoulli summable.

For $r = 1$ or 2, we saw that the random variable X_r has a tractable distribution. However, for $r \geq 3$ deriving the distribution of X_r is not so straightforward. Indeed, even for $r = 3$, evaluating $Pr(X_3 = t)$ for any $t \neq \binom{n}{3}$ poses a challenge that has only been addressed by asymptotic approximations in the literature ([3]). We shall derive the moments of X_r for all $r \geq 3$ and establish an expression in terms of the moments for the probabilities of the form $Pr(X_r = t)$ using Vandermonde matrices.

In the following sections, we develop the tools necessary to establish a relationship between the moments of a Bernoulli summable random variable X and its distribution. In particular, we show that $k-$th moment of X may be written as a linear combination of probabilities of the form $P(Y_{i_1} = 1, \ldots, Y_{i_m} = 1)$ for $m \leq k$.

4 The multinomial theorem

The multinomial theorem expresses the power of a series in terms of powers of summands as follows.

Theorem 4.1. *Let $y_1, \dots, y_n$ be a sequence of commutative elements over some ring and fix $k \geq 1$. Then*

$$(y_1 + \cdots + y_n)^k = \sum_{\substack{\ell_1 + \cdots + \ell_n = k \\ \ell_i \geq 0, \forall i}} \binom{k}{\ell_1, \cdots, \ell_n} y_1^{\ell_1} \cdots y_n^{\ell_n}.$$

Proof. May be found in [4] or [12] for example. □

When the y_i are also idempotents, that is when $y_i^2 = y_i$, then the above expression simplifies. For each term $y_1^{\ell_1} \cdots y_n^{\ell_n}$, only those y_i with $\ell_i \geq 1$ remain and become y_i. The next proposition describes the multinomial theorem when the summands are all commutative idempotents.

Proposition 4.2. *Let $y_1, \dots, y_n$ be a sequence of commutative idempotents over some ring and let $\succ$ be any total order on the $y_1, \dots, y_n$. Then*

$$(y_1 + \cdots + y_n)^k = \sum_{m=1}^{k} S(k, m) \sum_{\substack{i_1 \succ \cdots \succ i_m \\ i_1, \dots, i_m \in \mathcal{N}_n}} y_{i_1} \cdots \ y_{i_m}$$

where $S(k, m)$ is the number of surjections from $\{1, \dots, k\}$ onto $\{1, \dots, m\}$.

Proof. Prior to simplifying by commutativity, naively expanding $(y_1 + \cdots + y_n)^k$ results in

$$(y_1 + y_2 + \cdots + y_n)^k = \sum_{(j_1, j_2, \dots, j_k) \in \mathcal{N}_n^k} y_{j_1} y_{j_2} \cdots y_{j_k}. \tag{1}$$

Let $\mathcal{F}$ denote the set of all functions $f : \mathcal{N}_k \to \mathcal{N}_n$. For a product $y_{j_1} y_{j_2} \cdots y_{j_k}$ on the RHS of Equation 1, let f be the function which maps $\ell \in \mathcal{N}_k$ to j_ℓ. Since every $j_\ell \in \mathcal{N}_n$, this defines a function $f \in \mathcal{F}$. Conversely, for every function $f \in \mathcal{F}$, we can associate to it a unique summand of the form $y_{f(1)} y_{f(2)} \cdots y_{f(k)}$. Thus, we have just demonstrated that prior to simplifying by commutativity, the naive expansion of $(y_1 + \cdots + y_n)^k$ will result in n^k summands of the form $y_{f(1)} y_{f(2)} \cdots y_{f(k)}$ for some function $f : \mathcal{N}_k \to \mathcal{N}_n$.

Because $y_1, \dots, y_n$ are commutative idempotents, each product term $y_{j_1} \cdots \ y_{j_k}$ above will be equivalent, for some $m \in \mathcal{N}_k$, to a unique product $y_{i_1} \cdots \ y_{i_m}$ with ordered indices $i_1 \succ \ldots \succ i_m$. Therefore, we may rewrite Equation 1 as

$$(y_1 + y_2 + \cdots + y_n)^k = \sum_{(j_1, j_2, \dots, j_k) \in \mathcal{N}_n^k} y_{j_1} y_{j_2} \cdots y_{j_k} = \sum_{m=1}^{k} a(k, m) \sum_{\substack{i_1 \succ \ldots \succ i_m \\ i_1, \dots, i_m \in \mathcal{N}_n}} y_{i_1} \cdots \ y_{i_m}, \tag{2}$$

where $a(k, m)$ are the coefficients that record the number of ways a term of the form $y_{i_1} \cdots \ y_{i_m}$ will show up after simplifying the $y_{j_1} \cdots y_{j_k}$. We claim that $a(k, m)$ is equal to the number of surjective maps from $\mathcal{N}_k$ onto $\mathcal{N}_m$.

Fix $i_1 \neq \cdots \neq i_m \in \mathcal{N}_n$. Let $F \subseteq \mathcal{F}$ be the set of all functions $f : \mathcal{N}_k \to \mathcal{N}_n$ for which $y_{f(1)} \cdots \ y_{f(k)}$ simplifies into $y_{i_1} \cdots y_{i_m}$. Let G be the set of all surjections

$g : \mathcal{N}_k \to \{i_1, \ldots, i_m\}$. We shall show that $F = G$. If $f \in F$, then $y_{f(1)} \cdots\ y_{f(k)} = y_{i_1} \cdots y_{i_m}$ after simplification. Therefore, $f(\mathcal{N}_k) \subseteq \{i_1, \ldots, i_m\}$. Since the two summands are equal, it must be that every i_ℓ appears in $f(\mathcal{N}_k)$ and thus $f(\mathcal{N}_k) = \{i_1, \ldots, i_m\}$. Therefore, $f \in G$. If $g \in G$, then clearly

$$y_{g(1)} \cdots\ y_{g(k)} = \prod_{\ell \in g(\mathcal{N}_k)} y_\ell = \prod_{\ell \in \{i_1, \ldots, i_m\}} y_\ell,$$

and hence $g \in F$. Since $|F| = a(k, m)$ and $|G|$ is the number of surjections from a $k-$set onto an $m-$set,

$$a(k, m) = |F| = |G| = S(k, m),$$

as needed to be shown. □

Remark 4.3. *We note that the total order $\succ$ was only used for the convenience of selecting a representative from each $m-$subset of the indexing set. If one chooses another way of selecting such representative, one may simply write*

$$\sum_{i_1 \succ \ldots \succ i_m} y_{i_1} \cdots y_{i_m} = \sum_{i_1 \neq \ldots \neq i_m} y_{i_1} \cdots y_{i_m}.$$

Moreover, we note that since $S(k, m)$ is given by ([12])

$$S(k, m) = \sum_{v=0}^{k} (-1)^v \binom{m}{v} (m - v)^k,$$

we may also express this expansion as

$$(y_1 + \cdots + y_n)^k = \sum_{m=1}^{k} \sum_{\substack{i_1 \neq \cdots \neq i_m \\ i_1, \ldots, i_m \in \mathcal{N}_n}} \sum_{v=0}^{m} (-1)^v \binom{m}{v} (m - v)^k y_{i_1} \cdots y_{i_m}.$$

Next, we generalize the results above to deal with an infinite sequence of commutative variables. We begin by viewing them from the perspective of formal power series to avoid issues of convergence.

Proposition 4.4. *Let $(y_i)_{i \geq 1}$ be a sequence of formal, commutative, idempotents over some ring. Then*

$$\left(\sum_{i=1}^{\infty} y_i \right)^k = \sum_{m=1}^{k} \sum_{\substack{i_1 \neq \cdots \neq i_m \\ i_1, \ldots, i_m \in \mathbb{N}}} S(k, m) y_{i_1} \cdots y_{i_m}$$

where $S(k, m)$ is the number of surjections from $\{1, \ldots, k\}$ onto $\{1, \ldots, m\}$.

Proof. As in the proof to Proposition 4.2, it suffices to consider the coefficient of a term of the form $y_{i_1} \cdots y_{i_m}$. The proof is identical save for the fact that here we have an infinite summation on the right-hand side; however, this is not an issue when working with formal power series. □

5 Probability and the multinomial theorem

Since indicator variables over a commutative ring are commutative idempotents, we may apply Proposition 4.2 to any finite sum of indicator variables.

Proposition 5.1. *Let $Y_1, \ldots, Y_n$ be a sequence of Bernoulli random variables with parameters p_i. If X is the Bernoulli summable random variable that satisfies $X = \sum_{i=1}^{n} Y_i$, then*

$$E(X^k) = \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}).$$

Proof. Since $Y_i^2 = Y_i$ for all indicator random variables, we satisfy the hypothesis of Proposition 4.2. Thus, by Proposition 4.2, we have

$$(Y_1 + \cdots + Y_n)^k = \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m},$$

and hence by linearity of expectation

$$E(X^k) = \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}).$$

□

This allows us to express the moments of any Bernoulli summable random variable. Next, we extend Proposition 3 in order to express central moments as well.

Proposition 5.2. *Let $Y_1, \ldots, Y_n$ be a sequence of Bernoulli random variables with parameters p_i. If X is the Bernoulli summable random variable that satisfies $X = \sum_{i=1}^{n} Y_i$ and $\mu = E(X) = \sum_{i=1}^{n} p_i$, then*

$$E((X-\mu)^\ell) = \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}) (-1)^{\ell-k} \left(\sum_{i=1}^{n} p_i \right)^{\ell-k}.$$

Proof. By Proposition 5.1 and the classic Binomial theorem, we have

$$\begin{aligned}
E((X-\mu)^\ell) &= \sum_{k=0}^{\ell} \binom{\ell}{k} E(X^k)(-\mu)^{\ell-k} \\
&= \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m})(-1)^{\ell-k}(\mu)^{\ell-k} \\
&= \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m})(-1)^{\ell-k} \left(\sum_{i=1}^{n} p_i \right)^{\ell-k} \\
&= \sum_{k=0}^{\ell} \binom{\ell}{k} \left(\sum_{i=1}^{n} p_i \right)^{\ell-k} (-1)^{\ell-k} \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}).
\end{aligned}$$

□

Similarly, if we have a collection $(Y_i)_{i\geq 1}$ where $Y_i \sim \text{Bernoulli}(p_i)$ and X is the sum $X = \sum_{i\geq 1} Y_i$, then we may show that X is finite almost surely under the convergence of $\sum_{i\geq 1} p_i$. Indeed, by the Borel-Cantelli lemma, since $E(X) = \sum_{i=1}^{\infty} E(Y_i) = \mu < \infty$, it must be that the probability that infinitely many of the Y_i are 1 is 0: $P(\limsup_{n\to\infty} Y_n = 1) = 0)$.

Moreover, under sufficient assumptions on the $(Y_i)_{i\geq 1}$, we show that all of the moments of X are finite.

Corollary 5.3. *Let $(Y_i)_{i\geq 1}$ be a sequence of independent Bernoulli(p_i) random variables with $\sum_{i\geq 1} p_i = \mu < \infty$. If X is the Bernoulli summable random variable $X = \sum_{i\geq 1} Y_i$, then for any $k \geq 1$*

$$E(X^k) = \mu \sum_{m=1}^{k} S(k,m) \left[\sum_{i_1 \neq \cdots \neq i_{m-1}} p_{i_1} \cdots p_{i_{m-1}} - (m-1) \sum_{i_1 \neq \cdots \neq i_{m-1}} p_{i_1}^2 \cdots p_{i_{m-1}} \right] < \infty.$$

Proof. Since $\sum_{i\geq 1} p_i$ is a convergent series and the product of a finite number of convergent series is a convergent series, we have that for all $m \geq 1$

$$\sum_{\substack{i_1 \neq \cdots \neq i_m \\ i_1, \ldots, i_m \in \mathbb{N}}} p_{i_1} \cdots p_{i_m} \leq \left(\sum_{i\geq 1} p_i \right)^m = \mu^m < \infty.$$

Therefore, by Proposition 4.4

$$\begin{aligned}
E(X^k) &= \sum_{m=1}^{k} \sum_{\substack{i_1 \neq \cdots \neq i_m \\ i_1, \ldots, i_m \in \mathbb{N}}} S(k,m) p_{i_1} \cdots p_{i_m} \\
&\leq \sum_{m=1}^{k} S(k,m) \mu^m \\
&< \infty.
\end{aligned}$$

Alternatively, we note that

$$\begin{aligned}
\sum_{i_1\neq\cdots\neq i_m} p_{i_1}\cdots p_{i_m} &= \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}} \sum_{i_m\notin\{i_1,\dots,i_{m-1}\}} p_{i_m} \\
&= \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}}(\mu - p_{i_1}\cdots - p_{i_{m-1}}) \\
&= \mu \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}} - \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}^2\cdots p_{i_{m-1}} - \cdots - \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}}^2 \\
&= \mu \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}} - (m-1)\sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}^2\cdots p_{i_{m-1}},
\end{aligned}$$

where the last equality follows because

$$\sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}^2 p_{i_2}\cdots p_{i_{m-1}} = \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}p_{i_2}^2\cdots p_{i_{m-1}} = \cdots = \sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}p_{i_2}\cdots p_{i_{m-1}}^2.$$

Thus,

$$\begin{aligned}
E(X^k) &= \sum_{m=1}^{k} \sum_{\substack{i_1\neq\cdots\neq i_m \\ i_1,\dots,i_m\in\mathbb{N}}} S(k,m)p_{i_1}\cdots p_{i_m} \\
&= \sum_{m=1}^{k} S(k,m)\mu\left[\sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}} - (m-1)\sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}^2\cdots p_{i_{m-1}}\right] \\
&= \mu\sum_{m=1}^{k} S(k,m)\left[\sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}\cdots p_{i_{m-1}} - (m-1)\sum_{i_1\neq\cdots\neq i_{m-1}} p_{i_1}^2\cdots p_{i_{m-1}}\right]
\end{aligned}$$

□

6 The relationship between moments and distributions

We recall an essential property for finite support random variables. That is, for a finite support random variable X, the distribution of X is uniquely determined by its moments.

Proposition 6.1. *Let X be a finite support random variable and suppose that the support of X consists of $0, 1, \dots, \ell$. Then for all $x \in \{0, 1, \dots, \ell\}$, we can write $P(X = x)$ as a linear combination*

$$P(X = x) = \sum_{k=0}^{\ell} a_{kx}E(X^k).$$

Moreover, the a_{kx} do not depend on our distribution of X nor on the values of the moments.

Proof. We know that for all $0 \leq k \leq \ell$

$$E(X^k) = \sum_{x=0}^{\ell} x^k P(X = x),$$

and hence we may write

$$\mathbf{V}\mathbf{p} = \mathbf{M},$$

where $\mathbf{V} = \mathbf{V}(\ell)$ is the Vandermonde matrix given by

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 2 & \cdots & \ell \\ 0^2 & 1^2 & 2^2 & \cdots & \ell^2 \\ \vdots & & & & \vdots \\ 0^\ell & 1^\ell & 2^\ell & \cdots & \ell^\ell \end{bmatrix},$$

$\mathbf{p}$ is the distribution vector

$$\mathbf{p} = \begin{bmatrix} P(X = 0) \\ P(X = 1) \\ P(X = 2) \\ \vdots \\ P(X = \ell) \end{bmatrix},$$

and $\mathbf{M}$ is the moments vector

$$\mathbf{M} = \begin{bmatrix} 1 \\ E(X) \\ E(X^2) \\ \vdots \\ E(X^\ell) \end{bmatrix}.$$

Now, since $\mathbf{V}$ is a Vandermonde matrix, we have that the determinant is given by ([6], [10])

$$\det(V) = \prod_{1 \leq i < j \leq \ell} (j - i) = \prod_{k=1}^{\ell} k! = \ell\$,$$

where $\ell\$$ is a so-called superfactorial. Of particular interest to us is the observation that since $\ell\$$ is nonzero, $\mathbf{V}$ is invertible and hence for some matrix $\mathbf{A}$, we have that

$$\mathbf{p} = \mathbf{A}\mathbf{M}.$$

□

Although a closed form expression for the exact entries of the Vandermonde inverse is not known, there are a few recursive relations that allow one to evaluate this. For instance, it has been shown in [1] that

$$(\ell - 1)[\mathbf{V}(\ell)]^{-1}_{i,j} = \ell[\mathbf{V}(\ell - 1)]^{-1}_{i,j} - (\ell - 1)[\mathbf{V}(\ell - 1)]^{-1}_{i,j-1} - [\mathbf{V}(\ell - 1)]^{-1}_{i-1,j}$$
$$+[\mathbf{V}(\ell - 1)]^{-1}_{i-1,j-1} + \sum_{k=1}^{\ell-2} [\mathbf{V}(\ell - 1 - k)]^{-1}_{i,j-1},$$

where we set $[\mathbf{V}(\ell)]^{-1}_{i,j}$ to 0 if i or j are greater than ℓ or less than 1.

Thus, there is a way to obtain the distribution of a finite random variable by determining finitely many moments. In particular, we need as many moments as the support of the random variable.

Proposition 6.1 is useful when one is able to obtain the moments of a finite random variable without using the probability mass function directly. By Proposition 5.1, Bernoulli summable random variables meet this criteria. According to Proposition 5.1,

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k} \sum_{\substack{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m \\ \boldsymbol{i}_1, \ldots, \boldsymbol{i}_m \in \mathcal{N}_n}} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}) \\ &= \sum_{m=1}^{k} \sum_{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m} S(k,m) P(Y_{\boldsymbol{i}_1} = 1, \cdots Y_{\boldsymbol{i}_m} = 1), \end{aligned}$$

and we see that the moments of X are uniquely determined by the family of probabilities of $\{P(Y_{\boldsymbol{i}_1} = 1, \cdots Y_{\boldsymbol{i}_m} = 1) : \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m\} \subseteq \mathcal{N}_n, \text{ for some } m \leq n\}$. We note that probabilities of the form $P(Y_j = 0, Y_k = 1 : j \in \mathcal{J}, k \in \mathcal{K})$, for some $\mathcal{J} \neq \varnothing$ do not appear in the characterization above. Thus, a natural question arises: do we have some choice in the values those probabilities might take? Moreover, does this characterization imply that there might be two distinct, exchangeable collections of Bernoulli random variables $(Y_1, \ldots, Y_n)$ and $(Z_1, \ldots, Z_n)$ for which $\sum_{i=1}^n Y_i \overset{D}{=} \sum_{i=1}^n Z_i$?

The answer to both of the questions above is no. To see this, we claim that any probability of the form $P(Y_j = 0, Y_k = 1 : j \in \mathcal{J}, k \in \mathcal{K})$ may be written uniquely an integer combination of probabilities of the form $P(Y_k = 1 : k \in \mathcal{K})$ for some collection of indexing sets $\mathcal{K}_1, \ldots, \mathcal{K}_\ell$.

Proposition 6.2. *Let $\mathcal{J}, \mathcal{K}$ be disjoint subsets of $\mathcal{N}_n$. Then any probability of the form $P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}, \forall k \in \mathcal{K})$ may be expressed as an integer combination of probabilities of the form $P(Y_k = 1 : k \in \mathcal{K}')$, where we use the convention that*

$$P(Y_k = 1 : k \in \varnothing) := 1.$$

Proof. We proceed by induction on the size of $|\mathcal{J}|$. Suppose that $\mathcal{J} = \{j\}$. If $|\mathcal{K}| = 0$, then

$$\begin{aligned} P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}, \forall k \in \mathcal{K}) &= P(Y_j = 0) \\ &= P(Y_k = 1 : k \in \varnothing) - P(Y_k = 1 : k \in \{j\}). \\ &= 1 - P(Y_j = 1) \end{aligned}$$

If $|\mathcal{K}| > 0$, then

$$\begin{aligned} P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}, \forall k \in \mathcal{K}) &= P(Y_k = 1 : k \in \mathcal{K}) - P(Y_j = 1, Y_k = 1 : \forall k \in \mathcal{K}) \\ &= P(Y_k = 1 : k \in \mathcal{K}) - P(Y_i = 1 : \forall i \in \mathcal{K} \cup \{j\}). \end{aligned}$$

Suppose that for some $\ell \leq n$, for any $\mathcal{J}, \mathcal{K} \subset \mathcal{N}_n$ disjoint with $|\mathcal{J}| \leq \ell$

$$P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}, \forall k \in \mathcal{K}) = \sum_i a_i P(Y_k = 1 : \forall k \in \mathcal{K}_i),$$

for some collection of indexing sets $(\mathcal{K}_i)_i$. Let $\mathcal{J}', \mathcal{K}' \subset \mathcal{N}_n$ disjoint with $|\mathcal{J}'| = \ell + 1$. Let x be any index in $\mathcal{J}'$ and let $\mathcal{J}_x = \mathcal{J}' \setminus \{x\}$, then

$$\begin{aligned}P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}', \forall k \in \mathcal{K}') = P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}_x, \forall k \in \mathcal{K}') \\ - P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}_x, \forall k \in \mathcal{K}' \cup \{x\}).\end{aligned}$$

By the inductive hypothesis, $P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}_x, \forall k \in \mathcal{K}')$ and $P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}_x, \forall k \in \mathcal{K}' \cup \{x\})$ are expressible as integer combinations of probabilities of the form $P(Y_k = 1 : k \in \mathcal{K}')$. Therefore, by the priniciple of induction, for any $\mathcal{J}, \mathcal{K} \subset \mathcal{N}_n$ disjoint, the probability $P(Y_j = 0, Y_k = 1 : \forall j \in \mathcal{J}, \forall k \in \mathcal{K})$ is expressible as an integer combination of probabilities of the form $P(Y_k = 1 : \forall k \in \mathcal{K})$ for some collection of $(K_i)_i$. □

Thus, not only does the joint distribution $(Y_1, \ldots, Y_n)$ determine the distribution of X, for every discrete random variable X on $\{0, 1, \ldots, n\}$, there is a unique distribution on the exchangeable Bernoulli trials $(Y_1, Y_2 \ldots Y_n)$ for which $\sum_{i=1}^n Y_i$ has the same distribution as X. For an alternative proof, we refer the reader to Proposition 2 in [7].

7 Applications

7.1 Deriving the moments of Bernoulli summable random variables

In this section, we apply Propositions 5.1 and 5.2 to derive expressions for the moments of degenerate, binomial, hypergeometric and Conway-Maxwell Binomial random variables. Lastly, we use Proposition 5.1 to expand the moment generating function for Poisson Binomial random variables in a new way.

Example 7.1 (*Degenerate random variables*)**.** In Example 3.5, we saw an instance of a simple Bernoulli summable random variable which simplifies to a degenerate random variable. Here, we use the tools established above to show that it is indeed the degenerate random variable we claimed it was. Recall that the random variable X_1 counting the number of vertices in G is given by

$$X_1 = \sum_{i=1}^n Y_{i \text{ is a vertex}} = n,$$

where $Y_{i \text{ is a vertex}} \sim \text{Bernoulli}(1)$. By Propositions 5.1 and 6.1,

$$\begin{aligned}E(X_1^k) &= \sum_{m=1}^k \sum_{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}) \\ &= \sum_{m=1}^k \sum_{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m} S(k,m) \cdot 1 \\ &= \sum_{m=1}^k \sum_{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m} S(k,m) \binom{n}{m}\end{aligned}$$

where the last equality follows from the following combinatorial argument. Every function $f : \mathcal{N}_k \to \mathcal{N}_n$ can be identified as a surjection from $f : \mathcal{N}_k \to f(\mathcal{N}_k)$. Since the right hand side is counting the number of such functions, it must be that $E(X_1^k) = n^k$ as that is the total number of functions from $\mathcal{N}_n$ into $\mathcal{N}_k$. Now, we conclude

$$E(X_1) = n, E(X_1^2) = n^2, \ldots, E(X_1^k) = n^k = E(X_1)^k$$

for all $k \in \mathbb{N}$. Since the only random variable to satisfy that $E(X^k) = E(X)^k$ is the degenerate random variable, the moments of this random variable classified the distribution and X_1 must be the constant n.

Recall that in Example 3.9, we introduced an instance of Bernoulli summable random variables X_r that count the number of $r-$cliques on $G(n, p)$. In the special case that $r = 2$, we obtain a Binomial$\left(\binom{n}{2}, p\right)$ random variable. In this section, we write down a closed-form expression for the raw and central moments of a Binomial$(N, 2)$ random variable, for $N \in \mathbb{N}$.

Example 7.2 (*Moments of binomial random variables*)**.** When $r = 2$, the Bernoulli summable random variable X_2 can be written as

$$X_2 = \sum_{\boldsymbol{i}} Y_{\boldsymbol{i}},$$

where $\boldsymbol{i}$ is a 2-subset of $\mathcal{N}_n$ and $Y_{\boldsymbol{i}} \sim$ Bernoulli(p), independently. By Proposition 5.1, the moments of X_2 are

$$\begin{aligned}
E(X_2^k) &= \sum_{\boldsymbol{i_1}} S(k,1)\, E(Y_{\boldsymbol{i_1}}) + \sum_{\boldsymbol{i}_1 \neq \boldsymbol{i}_2} S(k,2)\, E(Y_{\boldsymbol{i_1}} Y_{\boldsymbol{i_2}}) + \cdots + \sum_{\boldsymbol{i}_1 \neq \boldsymbol{i}_2 \neq \cdots \neq \boldsymbol{i}_k} S(k,k)\, E(Y_{\boldsymbol{i_1}} \cdots Y_{\boldsymbol{i_k}}) \\
&= \sum_{\boldsymbol{i_1}} S(k,1)\, p + \sum_{\boldsymbol{i}_1 \neq \boldsymbol{i}_2} S(k,2)\, p^2 + \cdots + \sum_{\boldsymbol{i}_1 \neq \boldsymbol{i}_2 \neq \cdots \neq \boldsymbol{i}_k} S(k,k)\, p^k \\
&= \binom{N}{1} S(k,1)\, p + \binom{N}{2} S(k,2)\, p^2 + \cdots + \binom{N}{k} S(k,k)\, p^k \\
&= \sum_{i=0}^{k} \binom{N}{i} S(k,i) p^i,
\end{aligned}$$

where $N = \binom{n}{2}$. In fact, Proposition 5.1 implies that in general, if X is Binomial(n, p) then

$$E(X^k) = \sum_{i=0}^{k} \binom{n}{i} S(k,i)\, p^i, \tag{3}$$

and by simplifying, one can confirm that this is equivalent to the formulation of Binomial moments in [8]. The advantage that comes with using Proposition 4.2 is that it does not rely on recursion which would need to be rederived for other families of random variables and provides explicit expressions for moments in terms of the joint probabilities

$$P(Y_{\boldsymbol{i_1}} = 1, \ldots, Y_{\boldsymbol{i_k}} = 1) = E(Y_{\boldsymbol{i_1}} \cdots Y_{\boldsymbol{i_k}}).$$

Moreover, we can obtain an explicit expression for the central moments using Corollary 5.2. In particular, one can show that for our edge count random variable X_2

$$E((X_2-\mu)^\ell) = \binom{n}{2}^\ell \sum_{k=0}^{\ell}\sum_{m=1}^{k}(-1)^{\ell-k}\binom{\ell}{k}\binom{\binom{n}{2}}{m}S(k,m)p^{m+\ell-k},$$

and more generally,

$$E((X-\mu)^\ell) = n^\ell \sum_{k=0}^{\ell}\sum_{m=1}^{k}(-1)^{\ell-k}\binom{\ell}{k}\binom{n}{m}S(k,m)p^{m+\ell-k}$$

when $X \sim \text{Binomial}(n,p)$.

As an aside, we note that formulation 3 implies that for $k \geq n$, $E(X_2^k)$ consists is a sum consists of exactly n terms since $S(n,i)$ vanishes for $i \geq n$.

Example 7.3 (*Moments of hypergeometric random variables*)**.** Let $\mathcal{S}$ be a sample of size n from a population consisting of N individuals where g of the individuals have a desired trait. We saw in Example 3.6 that the random variable X counting the number of individuals with the desired trait is a Bernoulli summable random variable with

$$X = \sum_{i=1}^{n} Y_i,$$

where $Y_i \sim \text{Bernoulli}\left(\frac{g}{N}\right)$. By Proposition 5.1, for $k \leq g$

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\neq\cdots\neq\boldsymbol{i}_m} S(k,m)E(Y_{\boldsymbol{i}_1}\cdots Y_{\boldsymbol{i}_m}) \\ &= \sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\neq\cdots\neq\boldsymbol{i}_m} S(k,m)\frac{g(g-1)\cdots(g-m+1)}{N(N-1)\cdots(N-m+1)} \\ &= \sum_{m=1}^{k}\binom{n}{m}S(k,m)\frac{g(g-1)\cdots(g-m+1)}{N(N-1)\cdots(N-m+1)}, \end{aligned}$$

where the last equality followed from $m-$symmetry. By Proposition 5.2, for $\ell \leq g$

$$E((X-\mu)^\ell) = \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\binom{n}{m}S(k,m)\frac{g(g-1)\cdots(g-m+1)}{N(N-1)\cdots(N-m+1)}(-1)^{\ell-k}\left(n\frac{g}{N}\right)^{\ell-k}.$$

Example 7.4. Suppose we have ℓ indistinguishable balls which we assign uniformly at random into n distinguishable urns. Let Y_i be the indicator random variable recording if urn i is empty and let X be the corresponding Bernoulli summable random variable counting the total number of empty urns. That is, $X = \sum_{i=1}^{n} Y_i$.

Through a straightforward counting argument, one can show that there $\binom{\ell+n-1}{n}$ ways to distribute ℓ indistinguishable balls into n distinguishable urns and therefore

$$\begin{aligned} P(Y_i = 1) &= \frac{\#\text{ ways to distribute } m \text{ balls into } n-1 \text{ urns}}{\#\text{ ways to distribute } m \text{ balls into } n \text{ urns}} \\ &= \frac{\binom{n+\ell-2}{\ell}}{\binom{n+\ell-1}{\ell}} \\ &= \frac{n-1}{\ell+n-1}. \end{aligned}$$

By the same argument, for a subset $\{i_1, \ldots, i_m\}$ of $\mathcal{N}_n$,

$$\begin{aligned} P(Y_{i_1} = 1, Y_{i_2} = 1, \ldots, Y_{i_m} = 1) &= \frac{\#\text{ ways to distribute } \ell \text{ balls into } n-m \text{ urns}}{\#\text{ ways to distribute } \ell \text{ balls into } n \text{ urns}} \\ &= \frac{\binom{n+\ell-(m+1)}{\ell}}{\binom{n+\ell-1}{\ell}} \\ &= \frac{(n-1)(n-2)\cdots(n-m)}{(\ell+n-1)(\ell+n-2)\cdots(\ell+(n-m))}. \end{aligned}$$

Thus, we may write the moments of X as

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k} \sum_{i_1 \neq \cdots \neq i_m} S(k,m) E(Y_{i_1} \cdots Y_{i_m}) \\ &= \sum_{m=1}^{k} \sum_{i_1 \neq \cdots \neq i_m} S(k,m) \frac{(n-1)(n-2)\cdots(n-m)}{(\ell+n-1)(\ell+n-2)\cdots(\ell+(n-m))} \\ &= \sum_{m=1}^{k} \binom{n}{m} S(k,m) \frac{(n-1)(n-2)\cdots(n-m)}{(\ell+n-1)(\ell+n-2)\cdots(\ell+(n-m))}, \end{aligned}$$

where the last equality followed from $m-$symmetry of the distribution.

Similarly, we can deduce the central moments by Proposition 5.2:

$$\begin{aligned} E((X-\mu)^\ell) &= \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} \sum_{i_1 \neq \cdots \neq i_m} S(k,m) P(Y_{i_1} = 1, \ldots, Y_{i_m} = 1)(-1)^{\ell-k} \left(\sum_{i=1}^{n} p_i \right)^{\ell-k} \\ &= \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} (-1)^{\ell-k} \binom{n}{m} S(k,m) \frac{(n-1)(n-2)\cdots(n-m)}{(\ell+n-1)(\ell+n-2)\cdots(\ell+(n-m))} \times \\ &\qquad \left(\frac{n(n-1)}{\ell+n-1} \right)^{\ell-k}, \end{aligned}$$

where $p_i = \dfrac{n-1}{\ell+n-1}$ and $\mu := n\dfrac{n-1}{\ell+n-1}$.

Example 7.5 (*Conway-Maxwell Binomial distribution*)**.** For $n \in \mathbb{N}, p \in [0,1], \nu \in \mathbb{R},$, we say that X has the Conway-Maxwell Binomial distribution with parameters (n, p, ν) if the distribution of X is given by

$$P(X = j) = \frac{1}{C_{n,p,\nu}} \binom{n}{j}^{\nu} p^j (1-p)^{n-j},$$

where $C_{n,p,\nu}$ is the normalizing constant

$$C_{n,p,\nu} = \sum_{j=0}^{n} \binom{n}{j}^{\nu} p^j (1-p)^{n-j}.$$

This distribution was first investigated in [11], where the authors remarked that the random variable X can be viewed as a sum of exchangeable, non-independent Bernoulli random variables Y_i with joint distribution given by

$$P(Y_1 = y_1, \ldots, Y_n = y_n) = \frac{1}{C_{n,p,\nu}} \binom{n}{\sum_{i=1}^n y_i}^{\nu-1} p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i},$$

where $\nu > 1$ in the case of negatively correlated trials and $\nu < 1$ for positively correlated trials. We may use this observation to write an expression for the moments of X using Proposition 5.1.

$$\begin{aligned}
E(X^k) &= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_m} S(k,m) E(Y_{i_1}, \ldots, Y_{i_m}) \\
&= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_n} S(k,m) P(Y_{i_1} = 1, \ldots, Y_{i_m} = 1) \\
&= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_m} S(k,m) \sum_{\substack{y_j \in \{0,1\} \\ \forall j \notin \{i_1,\ldots,i_m\}}} P(Y_{i_1} = 1, \ldots, Y_{i_m} = 1, Y_j = y_j : j \notin \{i_1, \ldots, i_m\}) \\
&= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_m} S(k,m) \sum_{\substack{y_j \in \{0,1\} \\ \forall j \notin \{i_1,\ldots,i_m\}}} P(Y_{i_1} = 1, \ldots, Y_{i_m} = 1, Y_j = y_j : j \notin \{i_1, \ldots, i_m\}) \\
&= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_m} S(k,m) \sum_{\substack{y_j \in \{0,1\} \\ \forall j \notin \{i_1,\ldots,i_m\}}} \frac{1}{C_{n,p,\nu}} \binom{n}{m + \sum_{j \notin \{i_1,\ldots,i_m\}} y_j}^{\nu-1} \\
&\qquad p^{m+\sum_{j \notin \{i_1,\ldots,i_m\}} y_j} (1-p)^{n-\left(m+\sum_{j \notin \{i_1,\ldots,i_m\}} y_j\right)} \\
&= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_m} S(k,m) \sum_{\ell=0}^{n-m} \binom{n-m}{\ell} \binom{n}{m+\ell}^{\nu-1} p^{m+\ell}(1-p)^{n-(m+\ell)} \\
&= \sum_{m=1}^{k} \sum_{i_1 \neq \ldots \neq i_m} S(k,m) \sum_{s=m}^{n} \binom{n-m}{s-m} \binom{n}{s}^{\nu-1} p^s (1-p)^{n-s}.
\end{aligned}$$

Moreover, if we suppose that $Y_i \sim \text{Bernoulli}(p_i)$ for $i = 1, \ldots, n$ and we let $\mu := \sum_{i=1}^n p_i = E(X)$, by Proposition 5.2 the central moments are given by

$$
\begin{aligned}
E((X-\mu)^\ell) &= \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\sum_{i_1\neq\cdots\neq i_m} S(k,m)E(Y_{i_1}\cdots Y_{i_m})(-1)^{\ell-k}\left(\sum_{i=1}^{n}p_i\right)^{\ell-k}\\
&= \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\sum_{i_1\neq\cdots\neq i_m} S(k,m)P(Y_{i_1}=1,\ldots,Y_{i_m}=1)(-1)^{\ell-k}\left(\sum_{i=1}^{n}p_i\right)^{\ell-k}\\
&= \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\sum_{i_1\neq\cdots\neq i_m} S(k,m)(-1)^{\ell-k}\left(\sum_{i=1}^{n}p_i\right)^{\ell-k}\times\\
&\qquad \sum_{\substack{y_j\in\{0,1\}\\ \forall j\notin\{i_1,\ldots,i_m\}}} P(Y_{i_1}=1,\ldots,Y_{i_m}=1,Y_j=y_j : j\notin\{i_1,\ldots,i_m\})\\
&= \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\sum_{i_1\neq\cdots\neq i_m} S(k,m)(-1)^{\ell-k}\left(\sum_{i=1}^{n}p_i\right)^{\ell-k}\times\\
&\qquad \sum_{s=m}^{n}\binom{n-m}{s-m}\binom{n}{s}^{\nu-1}p^s(1-p)^{n-s}.
\end{aligned}
$$

Example 7.6 (*The moment generating function of Poisson Binomial random variables*)**.** Suppose that for $1\le i\le n$, $Y_i\sim$ Bernoulli(p_i) independently. The random variable X given by the sum $X=\sum_{i=1}^n Y_i$ is thus a Poisson Binomial random variable with parameters $(p_1,\ldots,p_n)$.

As the Y_i are independent, computing the moment generating function of X is straightforward:

$$
\begin{aligned}
M_X(t) &= E(e^{t\sum_{i=1}^n Y_i})\\
&= \prod_{i=1}^{n}(1-p_i+p_ie^t).
\end{aligned}
$$

As a consequence of Proposition 5.1, we may simplify this product.

$$
\begin{aligned}
M_X(t) &= E(e^{tX})\\
&= \sum_{k\ge 0}\frac{t^k}{k!}E(X^k)\\
&= 1+\sum_{k\ge 1}\frac{t^k}{k!}\sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\neq\cdots\neq\boldsymbol{i}_m} S(k,m)E(Y_{\boldsymbol{i}_1}\cdots Y_{\boldsymbol{i}_m})\\
&= 1+\sum_{k\ge 1}\frac{t^k}{k!}\sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\neq\cdots\neq\boldsymbol{i}_m} S(k,m)p_{i_1}\cdots p_{i_m}.
\end{aligned}
$$

7.2 The degree distribution of $G(n,p)$

In this section, we derive the joint distribution of the degrees of vertices in $G(n,p)$. We use this in conjunction with our previous results to count the number of vertices in $G(n,p)$ of a particular, fixed degree.

The following is a well-known result on the degrees of vertices in $G(n, p)$.

Proposition 7.7. *Let $i \in \mathcal{N}_n$ and $G \sim G(n,p)$. If D_i is the degree of vertex i in G, then*

$$P(D_i = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

That is, $D_i \sim$ Binomial$(n-1, p)$.

Proof. Let E_{rs} be the indicator random variable recording if the edge between r and s is present in G. Then

$$D_i = \sum_{j \neq i} E_{ij},$$

and since the E_{ij} are independent, Bernoulli(p) for all $i \neq j$, $D_i \sim$ Binomial$(n-1, p)$. □

In our quest to derive the joint distribution of the degrees, we begin by studying the more straightforward case where we are only interested in the degrees of two vertices. First, we demonstrate that the degree of one vertex is independent of the degree of another, given their adjacency.

Proposition 7.8. *Suppose that $n \geq 2$ and $G = G(n,p)$. Then*

$$P(D_1 | E_{12}, D_2) = P(D_1 | E_{12}).$$

Proof.

$$P(D_1 = d_1 | D_2 = d_2, E_{12} = e) = \frac{P(D_1 = d_1, E_{12} = e, D_2 = d_2)}{P(D_2 = d_2, E_{12} = e,)}$$

$$= \frac{P(\sum_{j \neq 1,2} E_{j1} = d_1 - e, \quad \sum_{j \neq 1,2} E_{j2} = d_2 - e, E_{12} = e)}{P(\sum_{j \neq 1,2} E_{j2} = d_2 - e, E_{12} = e)}$$

$$= \frac{\binom{n-2}{d_1-e} p^{d_1-e}(1-p)^{n-2-(d_1-e)} \binom{n-2}{d_2-e} p^{d_2-e}(1-p)^{n-2-(d_2-e)} \times p^e}{\binom{n-2}{d_2-e} p^{d_2-e}(1-p)^{n-2-(d_2-e)} \times p^e}$$

$$= \binom{n-2}{d_1-e} p^{d_1-e}(1-p)^{n-2-(d_1-e)}$$

$$= \frac{P(\sum_{j \neq 1,2} E_{j1} = d_1 - e, E_{12} = e)}{P(E_{12} = e)}$$

$$= P(D_1 = d_1 | E_{12} = e).$$

□

Using Proposition 7.8, we can derive the bivariate degree distribution of an Erdos Renyi graph.

Proposition 7.9. *The joint distribution* (D_1, D_2) *of* $G = G(n,p)$ *is given by*

$$\left(\left[\frac{n-1-d_1}{n-1}\right]\binom{n-2}{d_2}p^{d_2}(1-p)^{n-2-d_2} + \left[\frac{d_1}{n-1}\right]\binom{n-2}{d_2-1}p^{d_2-1}(1-p)^{n-2-d_2+1}\right) \times \binom{n-1}{d_1}p^{d_1}(1-p)^{n-1-d_1}.$$

Proof.

$$\begin{aligned} P(D_2 = d_2|D_1 = d_1) &= \sum_{e\{0,1\}} P(D_2 = d_2, E_{12} = e|D_1 = d_1) \\ &= \sum_{e\in\{0,1\}} \frac{P(D_2 = d_2, E_{12} = e, D_1 = d_1)}{P(D_1 = d_1)} \\ &= \sum_{e\in\{0,1\}} \frac{P(D_2 = d_2|E_{12} = e, D_1 = d_1)P(E_{12} = e, D_1 = d_1)}{P(D_1 = d_1)} \\ &= \sum_{e\in\{0,1\}} \frac{P(D_2 = d_2|E_{12} = e)P(E_{12} = e, D_1 = d_1)}{P(D_1 = d_1)}, \end{aligned}$$

by Proposition 7.8. Now,

$$\begin{aligned} \frac{P(E_{12} = e, D_1 = d_1)}{P(D_1 = d_1)} &= \frac{p^e(1-p)^{1-e}\binom{n-2}{d_1-e}p^{d_1-e}(1-p)^{n-2-(d_1-e)}}{\binom{n-1}{d_1}p^{d_1}(1-p)^{n-1-d_1}} \\ &= \frac{\binom{n-2}{d_1-e}p^{d_1}(1-p)^{n-1-d_1}}{\binom{n-1}{d_1}p^{d_1}(1-p)^{n-1-d_1}} \\ &= I_{[e=0]}\left[\frac{n-1-d_1}{n-1}\right] + I_{[e=1]}\left[\frac{d_1}{n-1}\right], \end{aligned}$$

we can evaluate the conditional distribution

$$\begin{aligned} P(D_2 = d_2|D_1 = d_1) &= \sum_{e\in\{0,1\}} \binom{n-2}{d_2-e}p^{d_2-e}(1-p)^{n-2-(d_2-e)}\left(I_{[e=0]}\left[\frac{n-1-d_1}{n-1}\right] + \left[I_{[e=1]}\frac{d_1}{n-1}\right]\right) \\ &= \left[\frac{n-1-d_1}{n-1}\right]\binom{n-2}{d_2}p^{d_2}(1-p)^{n-2-d_2} \\ &\qquad + \left[\frac{d_1}{n-1}\right]\binom{n-2}{d_2-1}p^{d_2-1}(1-p)^{n-2-d_2+1}. \end{aligned}$$

Therefore, the joint is given by

$$\left(\left[\frac{n-1-d_1}{n-1}\right]\binom{n-2}{d_2}p^{d_2}(1-p)^{n-2-d_2} + \left[\frac{d_1}{n-1}\right]\binom{n-2}{d_2-1}p^{d_2-1}(1-p)^{n-2-d_2+1}\right) \times\binom{n-1}{d_1}p^{d_1}(1-p)^{n-1-d_1}.$$

□

Next, we generalize Proposition 7.8 to the case of more than two vertices.

Proposition 7.10. *Let $\mathcal{I}$ be a collection of vertices in $G(n,p)$ and let $\mathcal{D} := \{D_i : i \in \mathcal{I}\}$. If $j \notin \mathcal{I}$ and $\mathbf{E}_j$ is the random vector recording the edges between j and $\mathcal{I}$ then*

$$P(D_j|\mathbf{E}_j, \mathcal{D}) = P(D_j|\mathbf{E}_j).$$

Proof.

$$\begin{aligned} P(D_j = d_j|\mathbf{E}_j = \mathbf{e}_j, \mathcal{D} = \mathbf{d}) &= \frac{P(D_j = d_j, \mathbf{E}_j = \mathbf{e}_j, \mathcal{D} = \mathbf{d})}{P(\mathbf{E}_j = \mathbf{e}_j, \mathcal{D} = \mathbf{d})} \\ &= \frac{P\left(\sum_{\ell \neq j} E_{\ell j} = d_j - \sum_{i \in I} e_{ij}, \mathbf{E}_j = \mathbf{e}_j, \left[\sum_{k \neq i} E_{ki} = d_i - e_{ij} : i \in \mathcal{I}\right]\right)}{P\left(\mathbf{E}_j = \mathbf{e}_j, \left[\sum_{k \neq i} E_{ki} = d_i - e_{ij} : i \in \mathcal{I}\right]\right)} \\ &= P\left(\sum_{\ell \neq j} E_{\ell j} = d_j - \sum_{i \in I} e_{ij}\right) \frac{P\left(\mathbf{E}_j = \mathbf{e}_j, \left[\sum_{k \neq i} E_{ki} = d_i - e_{ij} : i \in \mathcal{I}\right]\right)}{P\left(\mathbf{E}_j = \mathbf{e}_j, \left[\sum_{k \neq i} E_{ki} = d_i - e_{ij} : i \in \mathcal{I}\right]\right)} \\ &= P(D_j|\mathbf{E}_j). \end{aligned}$$

□

Finally, we can state the main result regarding joint degree distributions in $G(n,p)$.

Theorem 7.11. *Let $G = G(n,p)$ and fix $m \leq n$. If*

$$\mathcal{E}_n = \{\mathbf{e} : \mathbf{e} = (e_{ij})_{\{i,j\} \subset \mathcal{N}_n}, e_{ij} \in \{0,1\}\},$$

the joint distribution of $(D_1, D_2, \dots D_m) = (d_1, \dots, d_m)$ is

$$\sum_{\mathbf{e} \in \mathcal{E}_m} p^{|\mathbf{e}|}(1-p)^{\binom{m}{2}-|\mathbf{e}|} \prod_{i=1}^{m} \binom{n-m}{d_i - \sum_{j \in \mathcal{N}_n \setminus \{i\} e_{ij}}} p^{d_i - \sum_{j \in \mathcal{N}_n \setminus \{i\}} e_{ij}} (1-p)^{n-m-d_i + \sum_{j \in \mathcal{N}_n \setminus i} e_{ij}},$$

where

$$|\mathbf{e}| = \sum_{\{i,j\} \subset \mathcal{N}_n} e_{i,j}.$$

Proof. We will proceed inductively on m. Base case: $m = 2$: We saw in Proposition 7.9 that $P(D_1 = d_1, D_2 = d_2)$ is given by

$$\left(\left[\frac{n-1-d_1}{n-1}\right]\binom{n-2}{d_2}p^{d_2}(1-p)^{n-2-d_2} + \left[\frac{d_1}{n-1}\right]\binom{n-2}{d_2-1}p^{d_2-1}(1-p)^{n-2-d_2+1}\right) \times \binom{n-1}{d_1}p^{d_1}(1-p)^{n-1-d_1}. \quad (4)$$

Our goal is to demonstrate that this distribution is equal to

$$\begin{aligned} &(1-p)\left[\binom{n-2}{d_1}p^{d_1}(1-p)^{n-2-d_1}\binom{n-2}{d_2}p^{d_2}(1-p)^{n-2-d_2}\right] \\ &+ p\left[\binom{n-2}{d_1-1}p^{d_1-1}(1-p)^{n-1-d_1}\binom{n-2}{d_2-1}p^{d_2-1}(1-p)^{n-1-d_2}\right]. \end{aligned} \quad (5)$$

Dividing Equation 4 by $P(D_1 = d_1) = \binom{n-1}{d_1} p^{d_1}(1-p)^{n-1-d_1}$, gives

$$\begin{aligned}
(*) &= \frac{\binom{n-2}{d_1}}{\binom{n-1}{d_1}} p^{d_1-d_1}(1-p)^{n-1-d_1-(n-1-d_1)} P(D_2 = d_2, E_{12} = 0) \\
&\quad + \frac{\binom{n-2}{d_1-1}}{\binom{n-1}{d_1}} p^{d_1-d_1}(1-p)^{n-1-d_1-(n-1-d_1)} P(D_2 = d_2, E_{12} = 1) \\
&= \frac{n-1-d_1}{n-1}\binom{n-2}{d_2} p^{d_2}(1-p)^{n-2-d_2} + \frac{d_1}{n-1}\binom{n-2}{d_2-1} p^{d_2-1}(1-p)^{n-1-d_2},
\end{aligned}$$

which agrees with Proposition 7.9 after dividing by $P(D_1 = d_1)$. Thus we have just demonstrated our base case holds.

Now, suppose that the claim holds for some $m \geq 2$. Recall that Proposition 7.10 states that

$$P(D_j | \mathbf{E}_j, \mathcal{D}) = P(D_j | \mathbf{E}_j),$$

where j is a vertex and $\mathcal{D}$ is the random variable recording the degrees of all vertices i in some set of vertices $\mathcal{I}$. Thus,

$$\begin{aligned}
P([D_i = d_i, 1 \leq i \leq m+1]) &= \sum_{\mathbf{e}_{m+1}} P([D_i = d_i, 1 \leq i \leq m+1], \mathbf{E}_{m+1} = \mathbf{e}_{m+1}) \\
&= \sum_{\mathbf{e}_{m+1}} P(D_{m+1} = d_{m+1} | \mathbf{E}_{m+1} = \mathbf{e}_{m+1}, [D_i = d_i, 1 \leq i \leq m]) \\
&\qquad \times P(\mathbf{E}_{m+1} = \mathbf{e}_{m+1}, [D_i = d_i, 1 \leq i \leq m])) \\
&= \sum_{\mathbf{e}_{m+1}} P(D_{m+1} = d_{m+1} | \mathbf{E}_{m+1} = \mathbf{e}_{m+1}) \\
&\qquad \times P(\mathbf{E}_{m+1} = \mathbf{e}_{m+1}, [D_i = d_i, 1 \leq i \leq m])).
\end{aligned}$$

Simplifying terms we have

$$P(D_{m+1} = d_{m+1} | \mathbf{E}_{m+1} = \mathbf{e}_{m+1}) = \binom{n-(m+1)}{d_{m+1} - |\mathbf{e}_{m+1}|} p^{d_{m+1} - |\mathbf{e}_{m+1}|}(1-p)^{n-(m+1)-(d_{m+1} - |\mathbf{e}_{m+1}|)}.$$

Next, we can use the inductive hypothesis as follows

$$\begin{aligned}
P(\mathbf{E}_{m+1} = \mathbf{e}_{m+1}, [D_i = d_i, 1 \leq i \leq m])) &= P(\mathbf{E}_{m+1} = \mathbf{e}_{m+1}, [D_i = d_i - e_{i,m+1}, 1 \leq i \leq m] \\
&\qquad\qquad \text{in } G \setminus \{m+1\}) \\
&= p^{|\mathbf{e}_{m+1}|}(1-p)^{m+1-|\mathbf{e}_{m+1}|} \\
&\quad \times \sum_{\mathbf{e} \in \mathcal{E}_m} p^{|\mathbf{e}|}(1-p)^{\binom{m}{2} - |\mathbf{e}|} \prod_{i=1}^{m} \binom{n-1-m}{d_i - e_{i,m+1} - \sum_{j \neq \{i,m+1\}} e_{ij}} \\
&\quad \times p^{d_i - e_{i,m+1} - \sum_{j \neq \{i,m+1\}} e_{ij}}(1-p)^{n-m-1-d_i+e_{i,m+1}+\sum_{j \neq \{i,m+1\}} e_{ij}}.
\end{aligned}$$

Thus, we see that

$$P([D_i = d_i, 1 \le i \le m+1]) = \sum_{\mathbf{e}_{m+1}} \binom{n-(m+1)}{d_{m+1} - |\mathbf{e}_{m+1}|} p^{d_{m+1}-|\mathbf{e}_{m+1}|}(1-p)^{n-(m+1)-(d_{m+1}-|\mathbf{e}_{m+1}|)}$$
$$\times p^{|\mathbf{e}_{m+1}|}(1-p)^{m+1-|\mathbf{e}_{m+1}|} \sum_{\mathbf{e}\in\mathcal{E}_m} p^{|\mathbf{e}|}(1-p)^{\binom{m}{2}-|\mathbf{e}|} \prod_{i=1}^{m} \binom{n-1-m}{d_i - e_{i,m+1} - \sum_{j\neq\{i,m+1\}} e_{ij}}$$
$$\times p^{d_i - e_{i,m+1} - \sum_{j\neq\{i,m+1\}} e_{ij}}(1-p)^{n-m-1-d_i+e_{i,m+1}+\sum_{j\neq\{i,m+1\}} e_{ij}}$$

which can be further simplified into

$$= \sum_{\mathbf{e}_{m+1},\mathbf{e}} p^{|\mathbf{e}|+|\mathbf{e}_{m+1}|}(1-p)^{m+1+\binom{m}{2}-|\mathbf{e}_{m+1}|-|\mathbf{e}|} \binom{n-m-1}{d_{m+1} - \sum_{i\neq m+1} e_{i,m+1}} p^{d_{m+1}-\sum_{i\neq m+1} e_{i,m+1}}$$
$$\times (1-p)^{n-(m+1)-(d_{m+1}-\sum_{i\neq m+1} e_{i,m+1})} \prod_{i=1}^{m} \binom{n-1-m}{d_i - e_{i,m+1} - \sum_{j\neq\{i,m+1\}} e_{ij}}$$
$$\times p^{d_i - e_{i,m+1} - \sum_{j\neq\{i,m+1\}} e_{ij}}(1-p)^{n-m-1-d_i+e_{i,m+1}+\sum_{j\neq\{i,m+1\}} e_{ij}}$$

$$= \sum_{\mathbf{f}=(\mathbf{e},\mathbf{e}_{m+1})\in\mathcal{E}_{m+1}} p^{|\mathbf{f}|}(1-p)^{\binom{m+1}{2}-|\mathbf{f}|} \prod_{i=1}^{m+1} \binom{n-1-m}{d_i - \sum_{j\in\mathcal{N}_n\setminus\{i\}} f_{ij}}$$
$$\times p^{d_i - \sum_{j\in\mathcal{N}_n\setminus\{i\}} f_{ij}}(1-p)^{n-(m+1)-d_i+\sum_{j\in\mathcal{N}_n\setminus i} f_{ij}},$$

as claimed. Therefore, by induction, the joint distribution of $(D_1, \ldots, D_m)$ is given by

$$\sum_{\mathbf{e}\in\mathcal{E}_m} p^{|\mathbf{e}|}(1-p)^{\binom{m}{2}-|\mathbf{e}|} \prod_{i=1}^{m} \binom{n-m}{d_i - \sum_{j\in\mathcal{N}_n\setminus\{i\}} e_{ij}} p^{d_i-\sum_{j\in\mathcal{N}_n\setminus\{i\}} e_{ij}}(1-p)^{n-m-d_i+\sum_{j\in\mathcal{N}_n\setminus i} e_{ij}},$$

□

Suppose interest lies in counting the number of vertices of degree ℓ in $G(n,p)$. Let Z_ℓ denote the total count and $Y_i^{(\ell)}$ be the indicator random variable recording if vertex i has degree ℓ. Therefore, $Y_i^{(\ell)}$ is Bernoulli$(\binom{n-1}{k} p^\ell (1-p)^{n-1-\ell})$ and Z_ℓ is Bernoulli summable.

Proposition 7.12. *The $k-$th moment of Z_ℓ is*

$$E(Z_\ell^k) = \sum_{m=1}^{k} \sum_{i_1\neq\cdots\neq i_m} S(k,m) \sum_{\mathbf{e}\in\mathcal{E}_m} p^{|\mathbf{e}|}(1-p)^{\binom{m}{2}-|\mathbf{e}|} \times$$
$$\prod_{i=1}^{m} \binom{n-m}{\ell - \sum_{j\in\mathcal{N}_n\setminus\{i\}} e_{ij}} p^{\ell-\sum_{j\in\mathcal{N}_n\setminus\{i\}} e_{ij}}(1-p)^{n-m-\ell+\sum_{j\in\mathcal{N}_n\setminus i} e_{ij}}.$$

Proof. By Proposition 5.1,

$$E(Z_\ell^k) = \sum_{m=1}^{k} \sum_{i_1\neq\cdots\neq i_m} S(k,m) P(Y_{i_1}^{(\ell)} = 1, \cdots, Y_{i_m}^{(\ell)} = 1)$$

Since $P(Y_{i_1}^{(\ell)} = 1, \cdots, Y_{i_m}^{(\ell)} = 1) = P(D_{i_1} = \ell, \ldots, D_{i_m} = \ell)$,

$$E(Z_\ell^k) = \sum_{m=1}^{k} \sum_{i_1 \neq \cdots \neq i_m} S(k,m) \sum_{\mathbf{e} \in \mathcal{E}_m} p^{|\mathbf{e}|}(1-p)^{\binom{m}{2}-|\mathbf{e}|} \times$$
$$\prod_{i=1}^{m} \binom{n-m}{\ell - \sum_{j \in \mathcal{N}_n \setminus \{i\} e_{ij}}} p^{\ell - \sum_{j \in \mathcal{N}_n \setminus \{i\}} e_{ij}} (1-p)^{n-m-\ell+\sum_{j \in \mathcal{N}_n \setminus i} e_{ij}},$$

by Theorem 7.11. □

7.3 Bernoulli summable random variables on random graphs

The following lemma is necessary for expressing the probability of seeing a particular collection of cliques in a homogeneous Erdos-Renyi graph.

Lemma 7.13. *Let G be a realization of $G(n,p)$. Fix $r \leq s$ and suppose that $A = \{\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m\}$ is a collection of m $s-$cliques. The total number of $r-$cliques that are contained within A is*

$$\sum_{\varnothing \neq J \subseteq \{1,\ldots,m\}} (-1)^{|J|+1} \binom{I_J}{r},$$

where $I_J := |\bigcap_{j \in J} \boldsymbol{i}_j|$.

Proof. Our goal is to count the number of K_r that are induced by at least one of the $s-$cliques in A. Let $\binom{\boldsymbol{i}_j}{r} := \{\{\alpha_1, \ldots, \alpha_r\} \subseteq \boldsymbol{i}_j : \alpha_1 \neq \cdots \neq \alpha_r\}$ denote the set of $r-$cliques induced by the clique $\boldsymbol{i}_j$. Next, we prove that for any $\varnothing \neq J \subseteq \{1, \ldots, m\}$,

$$\left| \bigcap_{j \in J} \binom{\boldsymbol{i}_j}{r} \right| = \binom{I_J}{r},$$

by showing that $\bigcap_{j \in J} \binom{\boldsymbol{i}_j}{r} = \binom{\bigcap_{j \in J} \boldsymbol{i}_j}{r}$.

Fix $\{\alpha_1, \ldots, \alpha_r\} \in \bigcap_{j \in J} A_j$. Then $\{\alpha_1, \ldots, \alpha_r\} \subset \boldsymbol{i}_j$ for all $j \in J$ and therefore

$$\{\alpha_1, \ldots, \alpha_r\} \in \binom{\bigcap_{j \in J} \boldsymbol{i}_j}{r}.$$

Conversely, if $\{\alpha_1, \ldots, \alpha_r\} \in \binom{\bigcap_{j \in J} \boldsymbol{i}_j}{r}$ then $\{\alpha_1, \ldots, \alpha_r\} \subset \boldsymbol{i}_j$ for all $j \in J$ and thus

$$\{\alpha_1, \ldots, \alpha_r\} \in A_j,$$

for all $j \in J$ and the claim follows.

Therefore, the total number of $r-$cliques within A is $\left|\bigcup_{j \in \{1,\ldots,m\}} \binom{\boldsymbol{i}_j}{r}\right|$. By Principle of Inclusion Exclusion ([12]), we know that

$$\left|\bigcup_{j\in J}\binom{\boldsymbol{i}_j}{r}\right| = \sum_{\varnothing\neq J\subseteq\{1,\dots,m\}}(-1)^{|J|+1}\left|\bigcap_{j\in J}\binom{\boldsymbol{i}_j}{r}\right| = \sum_{\varnothing\neq J\subseteq\{1,\dots,m\}}(-1)^{|J|+1}\binom{I_J}{r},$$

as needed to be shown. □

Our initial motivation for this theorem was to find the moments of clique counts on random graphs and we can now evaluate these moments. Our next result achieves this objective.

Corollary 7.14. *Let G be a realization of $G(n,p)$. Let X_r denote the number of cliques of size r in G and let $Y_{\boldsymbol{i}}$ be the indicator variable recording if $\boldsymbol{i}$ forms an $r-$clique, where $\boldsymbol{i}$ is an $r-$subset of the vertices $\{1,\dots,n\}$. Then $X_r = \sum_{\boldsymbol{i}} Y_{\boldsymbol{i}}$ satisfies*

$$E(X_r^k) = \sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\succ\cdots\succ\boldsymbol{i}_m}\sum_{v=0}^{m}(-1)^v\binom{m}{v}(m-v)^k p^{f(\boldsymbol{i}_1,\dots,\boldsymbol{i}_m)},$$

and

$$E((X_r-\mu)^\ell) = \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\succ\cdots\succ\boldsymbol{i}_m}\sum_{v=0}^{m}(-1)^v\binom{m}{v}(m-v)^k(-1)^{\ell-k}p^{f(\boldsymbol{i}_1,\dots,\boldsymbol{i}_m)}\left(\binom{n}{r}p^{\binom{r}{2}}\right)^{\ell-k},$$

where $\boldsymbol{i}_1,\dots,\boldsymbol{i}_m$ are distinct, lexicographically ordered r-subsets of $\{1,\dots,n\}$ and

$$f(\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m) = \sum_{\varnothing\neq J\subseteq\{1,\dots,m\}}(-1)^{|J|+1}\binom{I_J}{2}.$$

Proof. Since $X_r = \sum_{\boldsymbol{i}} Y_i$ is a sum of finitely many, distinct indicator variables, we may apply Theorem 4.2 and thus

$$\begin{aligned} X_r^k &= \left(\sum_{\boldsymbol{i}} Y_{\boldsymbol{i}}\right)^k \\ &= \sum_{m=1}^{k}\sum_{i_1\succ\cdots\succ i_m}\sum_{v=0}^{m}(-1)^v\binom{m}{v}(m-v)^k Y_{\boldsymbol{i}_1}\cdots Y_{\boldsymbol{i}_m}. \end{aligned}$$

Thus, in order to evaluate the expected value $E(X_r^k)$, we need to evaluate

$$E(Y_{\boldsymbol{i}_1}\cdots Y_{\boldsymbol{i}_m}) = P(Y_{\boldsymbol{i}_1}=1,\cdots,Y_{\boldsymbol{i}_m}=1).$$

By applying Lemma 7.13 to $r=2$, we know that this is given by $p^{f(\boldsymbol{i}_1,\dots,\boldsymbol{i}_m)}$, where

$$f(\boldsymbol{i}_1,\cdots,\boldsymbol{i}_m) = \sum_{\varnothing\neq J\subseteq\{1,\dots,m\}}(-1)^{|J|+1}\binom{I_J}{2}.$$

Since we can evaluate the joint probabilities $P(Y_{\boldsymbol{i}_1}=1,\cdots,Y_{\boldsymbol{i}_m}=1)$, we may apply Proposition 5.2 and conclude

$$E((X_r-\mu)^\ell) = \sum_{k=0}^{\ell}\binom{\ell}{k}\sum_{m=1}^{k}\sum_{\boldsymbol{i}_1\succ\cdots\succ\boldsymbol{i}_m}\sum_{v=0}^{m}(-1)^v\binom{m}{v}(m-v)^k(-1)^{\ell-k}p^{f(\boldsymbol{i}_1,\dots,\boldsymbol{i}_m)}\left(\binom{n}{r}p^{\binom{r}{2}}\right)^{\ell-k}.$$

□

In the literature, identifying the expected value of the number of $r-$cliques in an Erdos-Renyi graph $G \sim G(n,p)$ is straightforward: For $\boldsymbol{i}$ an $r-$subset of $\{1,\dots,n\}$, let $Y_{\boldsymbol{i}}$ be the indicator random variable recording if $\boldsymbol{i}$ forms an $r-$clique in G. Then X_r, the number of $r-$cliques in $G(n,p)$ is given by

$$X_r = \sum_{\boldsymbol{i}:\ r\text{-subset of } \{1,\dots,n\}} Y_{\boldsymbol{i}},$$

and by linearity of expectation, we have

$$E(X_r) = \sum_{\boldsymbol{i}:\ r\text{-subset of } \{1,\dots,n\}} E(Y_{\boldsymbol{i}}) = \sum_{\boldsymbol{i}:\ r\text{-subset of } \{1,\dots,n\}} p^{\binom{r}{2}} = \binom{n}{r} p^{\binom{r}{2}},$$

since $E(Y_{\boldsymbol{i}}) = P(Y_{\boldsymbol{i}} = 1) = P(\text{The } \binom{r}{2} \text{ edges between the } r \text{ vertices in } \boldsymbol{i} \text{ are in } G) = p^{\binom{r}{2}}$. Here, we show how Proposition 4.2 can be used to express the variance of the count of $r-$cliques in a more painless fashion than in [2].

Corollary 7.15. *The variance $V(X_r)$ of the number of cliques of size $r \geq 3$ in $G(n,p)$ is given by*

$$V(X_r) = \binom{n}{r} p^{\binom{r}{2}} + \sum_{s=0}^{r-1} \binom{n}{s}\binom{n-s}{r-s}\binom{n-r}{r-s} p^{2\binom{r}{2}-\binom{s}{2}} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2.$$

Proof. From Theorem 4.2, we have

$$\begin{aligned}
V(X_r) &= E(X_r^2) - E(X_r)^2 \\
&= \sum_{m=1}^{2} \sum_{\boldsymbol{i_1}\neq\cdots\neq\boldsymbol{i_m}} \sum_{v=0}^{m} (-1)^v \binom{m}{v} (m-v)^k p^{f(\boldsymbol{i}_1,\dots,\boldsymbol{i}_m)} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2 \\
&= \sum_{\boldsymbol{i}} (-1)^0 \binom{1}{0} (1-0)^2 p^{\binom{r}{2}} + \sum_{\boldsymbol{i_1}\neq\boldsymbol{i_2}} \sum_{v=0}^{2} (-1)^v \binom{2}{v} (2-v)^2 p^{2\binom{r}{2}-\binom{|\boldsymbol{i}_1\cap\boldsymbol{i}_2|}{2}} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2 \\
&= \sum_{\boldsymbol{i}} p^{\binom{r}{2}} + \sum_{\boldsymbol{i_1}\neq\boldsymbol{i_2}} p^{2\binom{r}{2}-\binom{|\boldsymbol{i}_1\cap\boldsymbol{i}_2|}{2}} \sum_{v=0}^{2} (-1)^v \binom{2}{v} (2-v)^2 - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2 \\
&= \sum_{\boldsymbol{i}} p^{\binom{r}{2}} + \sum_{\boldsymbol{i_1}\neq\boldsymbol{i_2}} 2p^{2\binom{r}{2}-\binom{|\boldsymbol{i}_1\cap\boldsymbol{i}_2|}{2}} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2 \\
&= \sum_{\boldsymbol{i}} p^{\binom{r}{2}} + \sum_{(\boldsymbol{i_1},\boldsymbol{i_2}):\boldsymbol{i_1}\neq\boldsymbol{i_2}} p^{2\binom{r}{2}-\binom{|\boldsymbol{i}_1\cap\boldsymbol{i}_2|}{2}} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2 \\
&= \binom{n}{r} p^{\binom{r}{2}} + \sum_{s=0}^{r-1} \binom{n}{s}\binom{n-s}{r-s}\binom{n-r}{r-s} p^{2\binom{r}{2}-\binom{s}{2}} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2,
\end{aligned}$$

where we note that $\binom{n}{s}\binom{n-s}{r-s}\binom{n-r}{r-s}$ is the number of ways to construct a pair $(\boldsymbol{i}_1, \boldsymbol{i}_2)$ of distinct $r-$sets for which the intersection is of size s. □

Proposition 7.16. *Let $G_1, \dots, G_N, \dots$ be a sequence of independent, Erdos Renyi random graphs $G(n,p)$. Let $X_r(G_i)$ be the random variable counting the number of $r-$cliques on G_i and denote by $S_m := \sum_{i=1}^m \frac{X_r(G_i)}{m}$ the average number of $r-$cliques in the graphs $G_1, \dots, G_m$. Then as n approaches infinity,*

$$\sqrt{n}(S_n - \mu) \to N(0, \sigma^2),$$

where

$$\mu = \binom{n}{r} p^{\binom{r}{2}}$$

and

$$\sigma^2 = \binom{n}{r} p^{\binom{r}{2}} + \sum_{s=0}^{r-1} \binom{n}{s}\binom{n-s}{r-s}\binom{n-r}{r-s} p^{2\binom{r}{2}-\binom{s}{2}} - \left[\binom{n}{r} p^{\binom{r}{2}}\right]^2.$$

Proof. This is a straightforward application of the Lindeberg-Levy Central Limit Theorem. □

7.4 Inhomogeneous Erdos Renyi graphs

In this subsection, we specialize our results on Bernoulli summable random variables to inhomogeneous Erdos Renyi graphs where the edge inclusions are independent but need not have the same inclusion probability.

Example 7.17. Although our focus above was on homogeneous Erdos Renyi graphs, our results are easily extended to inhomogeneous Erdos Renyi graphs by considering appropriate Bernoulli summable random variables. For instance, consider the random graph G on 3 vertices with independent edge probabilities (p_1, p_2, p_3). If our goal is to obtain the distribution X_2 of the number of edges in G, we can proceed as we would in the homogeneous case: identify the moments using Proposition 5.1 and multiply by the inverse of a correctly chosen Vandermonde matrix.

Let Y_i be the indicator recording if edge i is included in G and let $\alpha_2 := \sum_{i_1 \neq i_2} p_{i_1} p_{i_2}$, $\alpha_3 := p_1 p_2 p_3$, then the moments of X_2 are

$$\begin{aligned} E(X_2) &= \sum_{i=1}^{3} p_i := \mu, \\ E(X_2^2) &= \sum_{m=1}^{2} \sum_{\boldsymbol{i_1} \neq \cdots \neq \boldsymbol{i_m}} S(2,m) E(Y_{\boldsymbol{i_1}} \cdots Y_{\boldsymbol{i_m}}) \\ &= \mu + 2(p_1p_2 + p_1p_2 + p_1p_3) = \mu + 2\alpha_2, \end{aligned}$$

and

$$\begin{aligned} E(X_2^3) &= \sum_{m=1}^{3} \sum_{\boldsymbol{i_1} \neq \cdots \neq \boldsymbol{i_m}} S(3,m) E(Y_{\boldsymbol{i_1}} \cdots Y_{\boldsymbol{i_m}}) \\ &= \mu + 6(p_1p_2 + p_1p_2 + p_1p_3) + 6(p_1p_2p_3) = \mu + 6\alpha_2 + 6\alpha_3. \end{aligned}$$

If $\mathbf{M}$ denotes the matrix

$$\mathbf{M} = \begin{bmatrix} 1 \\ \mu \\ \mu + 2\alpha_2 \\ \mu + 6\alpha_2 + 6\alpha_3 \end{bmatrix},$$

and $\mathbf{V}$ is the Vandermonde matrix

$$\mathbf{V} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 4 & 9 \\ 0 & 1 & 8 & 27 \end{bmatrix},$$

then the product $\mathbf{V}^{-1}\mathbf{M}$ yields the probability distribution of X_2:

$$\begin{bmatrix} P(X_2 = 0) \\ P(X_2 = 1) \\ P(X_2 = 2) \\ P(X_2 = 3) \end{bmatrix} = \begin{bmatrix} 1.000 & -1.833 & 1.000 & -0.166 \\ 0.000 & 3.000 & -2.500 & 0.500 \\ 0.000 & -1.500 & 2.000 & -0.500 \\ 0.000 & 0.333 & -0.500 & 0.166 \end{bmatrix} \begin{bmatrix} 1 \\ \mu \\ \mu + 2\alpha_2 \\ \mu + 6\alpha_2 + 6\alpha_3 \end{bmatrix}$$
$$= \begin{bmatrix} 1 - 1.833\mu + (\mu + 2\alpha_2) - 0.166(\mu + 6\alpha_2 + 6\alpha_3) \\ 3\mu - 2.5(\mu + 2\alpha_2) + 0.5(\mu + 6\alpha_2 + 6\alpha_3) \\ -1.5\mu + 2(\mu + 2\alpha_2) - 0.5(\mu + 6\alpha_2 + 6\alpha_3) \\ 0.333\mu - 0.5(\mu + 2\alpha_2) + 0.166(\mu + 6\alpha_2 + 6\alpha_3) \end{bmatrix}.$$

Although our work in the previous sections assumed that the random graph G is a realization of a homogeneous Erdos-Renyi graph, our results on clique counting and the degree distribution can be generalized to inhomogeneous Erdos-Renyi random graphs, where independence still holds but the probability of edge inclusion p might differ, because of the versatility of Theorem 4.2. In particular, all of the results generalize by replacing p by the appropriate edge probability p_{ij} in each of the claims we made above.

For instance, the following proposition gives the moments of clique counts in the inhomogeneous case.

Proposition 7.18. *Let G be a realization of an inhomogeneous Erdos-Renyi on the vertices $\{1, \ldots, n\}$, where an edge e is included with probability p_e independently of all other edges. Let X_r denote the number of cliques of size r in G and let $Y_{\boldsymbol{i}}$ be the indicator variable recording if $\boldsymbol{i}$ forms an $r-$clique, where $\boldsymbol{i}$ is an $r-$subset of the vertices $\{1, \ldots, n\}$. Then $X_r = \sum_{\boldsymbol{i}} Y_{\boldsymbol{i}}$ satisfies*

$$E(X_r^k) = \sum_{m=1}^{k} \sum_{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m} S(k,m) \prod_{e \in E(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m)} p_e$$

and

$$E((X_r - \mu)^\ell) = \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} \sum_{\boldsymbol{i}_1 \neq \cdots \neq \boldsymbol{i}_m} S(k,m) \left[\prod_{e \in E(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m)} p_e \right] \left(-\sum_{\boldsymbol{i}} \prod_{e \in E(\boldsymbol{i})} p_e \right)^{\ell - k}$$

where $E(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m)$ is the union of the edges of the cliques on $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m$.

Proof. Since $X_r = \sum_{\boldsymbol{i}} Y_i$ is a sum of finitely many, distinct indicator variables, we may apply Theorem 4.2 and thus

$$E(X_r^k) = \left(\sum_{\boldsymbol{i}} Y_{\boldsymbol{i}}\right)^k$$
$$= \sum_{m=1}^{k} \sum_{i_1 \neq \cdots \neq i_m} S(k,m) E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}).$$

By assumption of independence,

$$E(Y_{\boldsymbol{i}_1} \cdots Y_{\boldsymbol{i}_m}) = \prod_{e \in E(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m)} p_e,$$

as needed to be shown. Note that the mean of X_r is $\mu := E(X_r) = \sum_{\boldsymbol{i}} E(Y_{\boldsymbol{i}}) = \sum_{\boldsymbol{i}} \prod_{e \in E(\boldsymbol{i})} p_e$. Therefore, the $\ell-$th central moment is

$$E((X_r - \mu)^\ell) = \sum_{k=0}^{\ell} \binom{\ell}{k} E(X_r^k) \left(-\sum_{\boldsymbol{i}} \prod_{e \in E(\boldsymbol{i})} p_e\right)^{\ell-k}$$
$$= \sum_{k=0}^{\ell} \binom{\ell}{k} \sum_{m=1}^{k} \sum_{\boldsymbol{i_1} \neq \cdots \neq \boldsymbol{i_m}} S(k,m) \left[\prod_{e \in E(\boldsymbol{i}_1, \ldots, \boldsymbol{i}_m)} p_e\right] \left(-\sum_{\boldsymbol{i}} \prod_{e \in E(\boldsymbol{i})} p_e\right)^{\ell-k}$$

□

8 Conclusion

References

[1] C. M. Bender, D. C. Brody, and B. K. Meister. Inverse of a vandermonde matrix. *preprint*, 2002.

[2] B. Bollobás. Degree sequences of random graphs. *Discrete Mathematics*, 33(1):1–19, 1981.

[3] J. Gilmer and S. Kopparty. A local central limit theorem for triangles in a random graph. *Random Structures & Algorithms*, 48(4):732–750, 2016.

[4] I. P. Goulden and D. M. Jackson. *Combinatorial enumeration*. Courier Corporation, 2004.

[5] Y. Hong. On computing the distribution function for the poisson binomial distribution. *Computational Statistics & Data Analysis*, 59:41–51, 2013.

[6] R. A. Horn, R. A. Horn, and C. R. Johnson. *Topics in matrix analysis*. Cambridge university press, 1994.

[7] J. B. Kadane et al. Sums of possibly associated bernoulli variables: The conway–maxwell-binomial distribution. *Bayesian Analysis*, 11(2):403–420, 2016.

[8] A. Knoblauch. Closed-form expressions for the moments of the binomial probability distribution. *SIAM Journal on Applied Mathematics*, 69(1):197–204, 2008.

[9] S. 'e. o. D. Poisson. *Research on the probability of judgments in criminal matters and in civil matters*. Bachelor, 1837.

[10] D. Sharpe. *Rings and factorization*. CUP Archive, 1987.

[11] G. Shmueli, T. P. Minka, J. B. Kadane, S. Borle, and P. Boatwright. A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142, 2005.

[12] H. S. Wilf. *generatingfunctionology*. AK Peters/CRC Press, 2005.