Moments of Bernoulli Sums

R.W. Oldford and Pavel Shuldiner
University of Waterloo

October 22, 2021

Abstract

A novel multinomial theorem for commutative idempotents is shown to lead to new results about the moments, central moments, factorial moments, and their generating functions for any random variable $X = \sum_i Y_i$ expressible as a sum of Bernoulli indicator random variables Y_i. The resulting expressions are functions of the expectation of products of the Bernoulli indicator random variables. These results are used to derive novel expressions for the various moments in a large number of familiar examples and classic problems including: the binomial, hypergeometric, Poisson limit of the binomial, Poisson, Conway-Maxwell-Poisson binomial, Ideal Soliton, and Benford distributions as well as the empty urns problem and the matching problem. Other general results include expressions for the moments of an arbitrary count random variable in terms of its upper tail probabilities.

Keywords: factorial moments, central moments, binomial distribution, hypergeometric, Poisson, the empty urns problem, the matching problem, CMP-binomial, Benford distribution, ideal Soliton, Stirling numbers, Bell numbers
AMS classification: 05A15, 62E10

1 Introduction

Consider the random variable

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

which sums (not necessarily independent) Bernoulli random variables $Y_i \in \{0, 1\}$ over some countable index set $\mathcal{I}$. When Y_i is an indicator function for some event $\mathcal{A}$, the *Bernoulli sum* X counts the number of occurrences of the event in the set $\mathcal{I}$ and, as such, arises in numerous applications of probability. Of interest here is the determination of the moments, central moments, and factorial moments of any arbitrary Bernoulli sum.

We develop expressions for these moments in terms of the expectation of products of the Bernoulli Y_is. This leads to novel proofs and/or expressions for the moments in many well known problems and to novel approaches to determining such moments for any random variable expressible as a Bernoulli sum.

The paper is organized as follows. Section 2 shows that the power of a sum of idempotents is expressible in terms of the number of surjections from one finite set to another times a sum of their products. This follows as a special case of the multinomial theorem. Section 3 builds on this to develop the main general results for the moments of a Bernoulli sum. Both finite and infinite sums are considered and special attention is given to factorial moments and generating functions.

These results are then applied to develop expressions for various classic distributions and problems in Section 4. These include the binomial, poisson binomial, hypergeometric, and Conway-Maxwell-Poisson binomial distributions, the Poisson limit of a binomial by moment convergence, and the classic empty urns problem and the matching problem.

Section 5 considers the moments for any count random variable, developing expressions based on the upper tail probability of that count. This general theory is then demonstrated on the geometric, Poisson, Ideal Soliton, and Benford distributions. The paper ends with a few concluding remarks in its last section.

2 Preliminaries

First, let $\mathcal{N}_k$ denote the set $\{1, \ldots, k\}$ for any finite integer k, and $S(k, m)$ denote the number of surjections from $\mathcal{N}_k$ to $\mathcal{N}_m$. If $k < m$, no surjective function exists and $S(k, m) = 0$; otherwise, $S(k, m)$ can be written as

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

(e.g., see Wilf, 2005). The number $S(k, m)$ figures prominently in the closed form expressions which follow.

In particular, $S(k, m)$ (for all $m \leq k$) will be shown to appear in expressions for the kth moments of a Bernoulli sum. To calculate the smaller moments of importance in statistical inference (say $k \leq 4$), it will be convenient therefore to have $S(k, m)$ evaluated for a few $m \leq k$. Whenever k is at least as large as the second argument, the following values obtain: $S(k, 0) = 0$, $S(k, 1) = 1$, $S(k, 2) = 2^k - 2$, $S(k, 3) = 3^k - 3 \cdot 2^k + 3$, and $S(k, 4) = 4^k - 4 \cdot 3^k + 6 \cdot 2^k - 4$. These will appear in calculations up to the 4th moment (e.g. to determine kurtosis). Again, note that $S(k, m) = 0$ whenever $m > k$.

Moments of X are expectations of powers of X which, in the case of $X = \sum_{i=1}^{n} Y_i$, suggests beginning with a multinomial theorem, here expressed fairly generally as:

Theorem 2.1. ***The multinomial theorem.*** *Let $y_1, \ldots, y_n$ 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. E.g., see Goulden and Jackson (2004) or Wilf (2005). □

When the y_is are also idempotents, as they are in the definition of a Bernoulli sum, only those y_i with $\ell_i \geq 1$ remain and simplify to $y_i^{\ell_i} = y_i$. This leads to the following version of a multinomial theorem where now $S(k,m)$ appears.

Proposition 2.2. *Let $y_1, \ldots, y_n$ be a sequence of commutative idempotents over some ring. Then*

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

where $S(k,m)$ is the number of surjections from $\mathcal{N}_k$ onto $\mathcal{N}_m$.
(It is understood that the interior sum has $m \leq \min\{k,n\}$.)

Proof. Naively expanding $(y_1 + \cdots + y_n)^k$ gives

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

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 right-hand side 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, a unique summand of the form $y_{f(1)} y_{f(2)} \cdots y_{f(k)}$ can be assigned to each function $f \in \mathcal{F}$. That is, the naive expansion of $(y_1 + \cdots + y_n)^k$ results in n^k summands of the form $y_{f(1)} y_{f(2)} \cdots y_{f(k)}$ for some function $f : \mathcal{N}_k \to \mathcal{N}_n$.

Since $y_1, \ldots, y_n$ are commutative idempotents, each product $y_{j_1} \cdots\ y_{j_k}$ resolves to a unique $y_{i_1} \cdots\ y_{i_m}$ with indices $i_1 < \ldots < i_m$, for some $m \in \mathcal{N}_k$. Equation 1 then becomes

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

Here $a(k,m)$ counts the number terms $y_{i_1} \cdots\ y_{i_m}$ that simplify to $y_{j_1} \cdots y_{j_k}$. It remains only to show that $a(k,m)$ equals $S(k,m)$, the number of surjective maps from $\mathcal{N}_k$ onto $\mathcal{N}_m$.

To see this, first fix $\{i_1, \ldots, i_m\} \subseteq \mathcal{N}_n$ and let $F \subseteq \mathcal{F}$ denote the subset of functions for which $y_{f(1)} \cdots\ y_{f(k)}$ simplifies to $y_{i_1} \cdots y_{i_m}$. The count $a(k,m)$ is identical to $|F|$. Then consider the set, G, of all surjections $g : \mathcal{N}_k \to \{i_1, \ldots, i_m\}$, which must have size $|G| = S(k,m)$. If $F = G$, then $|F| = |G|$ and $a(k,m) = S(k,m)$, as required.

Now $F = G$ iff every $f \in F$ is also in G and every $g \in G$ is also in F. If $f \in F$, then $y_{f(1)} \cdots\ y_{f(k)} = y_{i_1} \cdots y_{i_m}$, giving $f(\mathcal{N}_k) = \{i_1, \ldots, i_m\}$, and hence $f \in G$. 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 so $g \in F$. □

Note that the inner sum $\sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} y_{i_1} \cdots\ y_{i_m}$ vanishes whenever $m > n$ and hence $(y_1 + \cdots + y_n)^k$ is expressible as a sum of at most $min(k,n)$ terms involving the coefficients $S(k,m)$.

A generalization of the result to powers of infinite sums, subject to convergence having been settled for any particular values of the y_is, is relatively straightforward.

Proposition 2.3. *Let $(y_i)_{i\geq 1}$ be a sequence of formal, commutative, idempotents over some ring. Then*

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

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

Proof. Since the $(y_i)_{i\geq 1}$ are formal, commutative idempotents over some ring, the proof follows that of Proposition 2.2. □

3 Moments of Bernoulli sums

Consider the Bernoulli sum random variable X of Section 1 with finite index set $\mathcal{I}$ of size n. Set $\mathcal{I}$ can always be re-indexed to have X appear as

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

Expressions for the moments of X can now be derived via Proposition 2.2.

Proposition 3.1. *When X is expressible as a finite Bernoulli sum $X = \sum_{i=1}^{n} Y_i$, the kth **moment** of X is expressible as*

$$E(X^k) = \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subseteq\mathcal{N}_n} E(Y_{i_1}\cdots Y_{i_m}).$$

Proof. Since $Y_i^2 = Y_i \;\forall\; i$, it follows from Proposition 2.2 that

$$X^k = (Y_1 + \cdots + Y_n)^k = \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subseteq\mathcal{N}_n} Y_{i_1}\cdots Y_{i_m}.$$

The result follows by applying expectation $E(\cdot)$ (or, more generally, any linear) operator to each side. □

Since

$$E(Y_{i_1}\cdots Y_{i_m}) = Pr(Y_{i_1} = 1, \cdots, Y_{i_m} = 1)$$

Proposition 3.1 shows that the moments of *any* finite Bernoulli sum random variable can be investigated via the joint distribution of those Bernoulli random variables used to construct it – indeed, Proposition 3.1 could be rewritten in terms of this probability.

The *central moments* are generally of more statistical interest and a similar result is found for them by applying Proposition 3.1. In this case, let $p_i = Pr(Y_i = 1) = E(Y_i)$ denote the ith marginal mean in the sum and $\mu = E(X) = \sum_{i=1}^{n} p_i$ the mean of X. A similar expression for the kth central moment is given in Proposition 3.2.

Proposition 3.2. *When X is expressible as a finite Bernoulli sum $X = \sum_{i=1}^{n} Y_i$, with $p_i = Pr(Y_i = 1) = E(Y_i)$, then the kth* ***central moment*** *of X is expressible as*

$$E((X-\mu)^k) = (-\mu)^k + \sum_{\ell=1}^{k} \binom{k}{\ell} (-\mu)^{k-\ell} \sum_{m=1}^{\ell} S(\ell, m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} E(Y_{i_1} \cdots Y_{i_m})$$

where $\mu = E(X) = \sum_{i=1}^{n} p_i$.

Proof. Applying the binomial expansion, then Proposition 3.1, yields

$$\begin{aligned} E((X-\mu)^k) &= \sum_{\ell=0}^{k} \binom{k}{\ell} E(X^\ell)(-\mu)^{k-\ell} \\ &= \binom{k}{0} \cdot 1 \cdot (-\mu)^k + \sum_{\ell=1}^{k} \binom{k}{\ell} (-\mu)^{k-\ell} E(X^\ell) \\ &= (-\mu)^k + \sum_{\ell=1}^{k} \binom{k}{\ell} (-\mu)^{k-\ell} \sum_{m=1}^{\ell} S(\ell, m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} E(Y_{i_1} \cdots Y_{i_m}) \end{aligned}$$

□

Of course, whenever the Y_is are also **independently distributed**, the above moment expressions (and those which follow) simplify by replacing $E(Y_{i_1} \cdots Y_{i_m})$ by $p_{i_1} \cdots p_{i_m}$, where each $Y_i \sim \text{Bernoulli}(p_i)$.

3.1 Moments of an infinite sequence

Consider now an infinite sequence

$$(Y_i)_{i \geq 1} = Y_1, Y_2, \ldots$$

of Bernoulli random variables and their sum

$$X = \sum_{i=1}^{\infty} Y_i$$

being such that $Pr(X < \infty) = 1$. (This condition is satisfied, for example, whenever the first moment of X is bounded, that is, whenever $E(X) = \sum_{i=1}^{\infty} E(Y_i) = \sum_{i \geq 1} p_i = \mu < \infty$. From this it follows (e.g., by the Borel-Cantelli lemma) that $Pr(\limsup_{n \to \infty} Y_n = 1) = 0$, and, so, that the probability is zero that infinitely many of the Y_is will be 1.)

In this case, Proposition 2.3 gives the kth moment for this sum of countably infinite Bernoulli random variables (whenever all relevant sums converge).

Proposition 3.3. *Let $X = \sum_{i=1}^{\infty} Y_i$ be the sum of the sequence of Bernoulli random variables $(Y_i)_{i\geq 1}$, with $p_i = Pr(Y_i = 1) = 1 - Pr(Y_i = 0)$, then we may write*

$$E(X^k) = \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subset\mathbb{N}} E(Y_{i_1}\cdots Y_{i_m}).$$

In the special case where the Y_is are also independent, then Proposition 3.3 allows us to draw the interesting conclusion that *a bounded first moment of X implies that all higher order moments are also bounded*. This result is formally given in Proposition 3.4:

Proposition 3.4. *Let $(Y_i)_{i\geq 1}$ be a sequence of **independent** Bernoulli(p_i) random variables with $\sum_{i\geq 1} p_i = \mu < \infty$. For the Bernoulli (infinite) sum random variable $X = \sum_{i\geq 1} Y_i$, and for any $k \geq 1$,*

$$E(X^k) < \infty.$$

Proof. By Proposition 3.3 and independence of the Y_i,

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subset\mathbb{N}} p_{i_1}\cdots p_{i_m} \\ &\leq \sum_{m=1}^{k} S(k,m) \left(\sum_{i\geq 1} p_i\right)^m \\ &= \sum_{m=1}^{k} S(k,m)\ \mu^m \\ &< \infty. \end{aligned}$$

□

In this special case of an infinite sequence of independent Bernoulli random variables, an expression for the moments involving only the first moments of the Y_is and of X can be easily had as well.

Proposition 3.5. *Let $(Y_i)_{i\geq 1}$ be a sequence of **independent** Bernoulli(p_i) random variables with $\sum_{i\geq 1} p_i = \mu < \infty$. For the Bernoulli (infinite) sum random variable $X = \sum_{i\geq 1} Y_i$, and for any $k \geq 2$, the kth **moment** of X is*

$$E(X^k) = \sum_{m=1}^{k} S(k,m) \left[\mu^m - \sum_{s=0}^{k-2} \mu^s (m-1-s) \sum_{\{i_1\ldots,i_{m-1-s}\}\subset\mathbb{N}} p_{i_1}^2 (p_{i_2}\cdots p_{i_{m-1-s}})\right].$$

Proof. Fix an integer $r \geq 2$ and note that

$$\begin{aligned}\sum_{\{i_1,\ldots,i_r\}\subset\mathbb{N}} p_{i_1}\cdots p_{i_r} &= \sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}\cdots p_{i_{r-1}}\left(\sum_{i\notin\{i_1,\ldots,i_{r-1}\}} p_i\right)\\ &= \sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}\cdots p_{i_{r-1}}\left(\mu - \sum_{i\in\{i_1,\ldots,i_{r-1}\}} p_i\right)\\ &= \mu \sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}\cdots p_{i_{r-1}} - (r-2)\sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}^2\cdots p_{i_{r-1}},\end{aligned}$$

where the last equality follows from the fact that

$$\sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}^2 p_{i_2}\cdots p_{i_{r-1}} = \sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}p_{i_2}^2\cdots p_{i_{r-1}} = \cdots = \sum_{\{i_1,\ldots,i_{r-1}\}\subset\mathbb{N}} p_{i_1}p_{i_2}\cdots p_{i_{r-1}}^2.$$

Recursively rewriting $\sum_{\{i_1,\ldots,i_r\}\subset\mathbb{N}} p_{i_1}\cdots p_{i_r}$ in terms of sums over one fewer index (viz., $r-1$ indices) each time gives the desired result via Proposition 3.1. □

3.2 Factorial moments

The kth falling factorial of x is the kth degree polynomial in x

$$[x]_k := x(x-1)(x-2)\cdots(x-(k-1)) = \prod_{m=0}^{k-1}(x-m),$$

where $k \in \mathbb{N}$ and $x \in \mathbb{R}$. Replacing x by a random variable X, the corresponding kth **factorial moment** is defined to be $E([X]_k)$. Like $E(X^k)$ this is the expected value of the product of k terms. Note this is different from $E(X!)$, the **expected factorial** of X, where the number of products in $X!$ is itself be a random variable (viz., X).

The kth power of x can be expressed (Stanley, 2011, p. 74) in terms of falling factorials as

$$x^k = \sum_{m=1}^{k} S_2(k,m)[x]_m,$$

where $S_2(k,m)$ is the Stirling number of the second kind, typically defined as the number of ways to partition a set of k labelled objects into m nonempty unlabelled subsets. It follows, then, that these are directly related to the number of surjections from a $k-$set onto an $m-$set as

$$S(k,m) = m!\ S_2(k,m)$$

and hence that

$$x^k = \sum_{m=1}^{k} \frac{S(k,m)}{m!}[x]_m = \sum_{m=1}^{k} S(k,m)\binom{x}{m}.$$

Similarly, the falling factorial is written as a sum of powers as

$$[x]_k = \sum_{m=1}^{k} S_1(k,m) x^m$$

where $S_1(k,m)$ is the Stirling number of the first kind. Similar expressions may now be found involving a Bernoulli sum X in place of x.

First, we relate $\binom{X}{m}$ to the Bernoullis that define X.

Proposition 3.6. *If X is a Bernoulli sum $X = \sum_{i \in I} Y_i$ for some countable indexing set $\mathcal{I}$, then for $m \geq 1$,*

$$\binom{X}{m} = \sum_{\{i_1,\ldots,i_m\} \subseteq \ \mathcal{I}} Y_{i_1} \cdots Y_{i_m}.$$

Proof. Let $\mathcal{J} = \{i \in \mathcal{I} : Y_i = 1\}$ denote the subset of the indices in $\mathcal{I}$ for which $Y_i = 1$. Then

$$\begin{aligned}
\sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{I}} Y_{i_1} \cdots Y_{i_m} &= \sum_{\{i_1,\ldots,i_m\} \subseteq \ \mathcal{J}} Y_{i_1} \cdots Y_{i_m} \\
&= \sum_{\{i_1,\ldots,i_m\} \subseteq \ \mathcal{J}} 1 \\
&= \binom{|\mathcal{J}|}{m} \\
&= \binom{X}{m}.
\end{aligned}$$

□

An earlier, inductive, proof of this result for the case of finite $\mathcal{I}$ is given by Iyer (1958).

A similar approach yields a general result relating $X!$ to its Bernoulli constituents.

Proposition 3.7. *Let $X = \sum_{i \in \mathcal{I}} Y_i$ be a Bernoulli sum, where $\mathcal{I}$ is a countable indexing set. Then we may write $X!$ in terms of the (Y_i) as follows*

$$X! = \sum_{\mathcal{H} \subseteq \mathcal{I}} |\mathcal{H}|! \left(\prod_{i \in \mathcal{H}} Y_i \prod_{i \notin \mathcal{H}} (1 - Y_i) \right).$$

Proof. Consider the set $\mathcal{J} := \{i \in \mathcal{I} : Y_i = 1\}$. In this case, $|\mathcal{J}| = X$ and $|\mathcal{J}|! = X!$. For any other set $\mathcal{H} \subseteq \mathcal{I}$, either $\mathcal{H} = \mathcal{J}$, or $\mathcal{H} \neq \mathcal{J}$.

If $\mathcal{H} = \mathcal{J}$, then

$$\prod_{i \in H} Y_i \prod_{i \notin \mathcal{H}} (1 - Y_i) = \prod_{i \in \mathcal{H}} 1 \prod_{i \notin \mathcal{H}} (1 - 0) = 1$$

If $\mathcal{H} \neq \mathcal{J}$, then there exists j for which $j \in \mathcal{J}$ but $j \notin \mathcal{H}$. Then,

$$\prod_{i \in \mathcal{H}} Y_i \prod_{i \notin \mathcal{H}} 1 - Y_i) = \prod_{i \in \mathcal{H}} Y_i \times 0 = 0.$$

Together these give

$$\sum_{\mathcal{H}\subseteq\mathcal{I}} |\mathcal{H}|! \left(\prod_{i\in\mathcal{H}} Y_i \prod_{i\notin\mathcal{H}} (1-Y_i)\right) = |\mathcal{J}|! \prod_{i\in\mathcal{J}} Y_i \prod_{i\notin\mathcal{J}} (1-Y_i) = |\mathcal{J}|! = X!$$

□

Taking expectations yields the following expressions for a Bernoulli sum $X = \sum_{i\in\mathcal{I}} Y_i$:

- the kth factorial moment in terms of the Bernoulli random variables

$$E([X]_k) = k! \sum_{\{i_1,\ldots,i_k\}\subseteq\ \mathcal{I}} E(Y_{i_1}\cdots Y_{i_k})$$

  or, in terms of the moments of X as

$$E([X]_k) = \sum_{m=1}^{k} S_1(k,m) E(X^m)$$

- the kth moment in terms of the factorial moments of X

$$E(X^k) = \sum_{m=1}^{k} S_2(k,m)\ E([X]_m)$$

- the kth central moment

$$E((X-\mu)^k) = (-\mu)^k + \sum_{j=1}^{k} \left(\sum_{m=j}^{k} S_2(m,j)\binom{k}{m}(-\mu)^{k-m}\right) E([X]_j)$$

- and the expected factorial in terms of the Bernoulli random variables

$$E(X!) = \sum_{\mathcal{H}\subseteq\mathcal{I}} |\mathcal{H}|!\ \times E\left[\prod_{i\in\mathcal{H}} Y_i \prod_{i\notin\mathcal{H}} (1-Y_i)\right].$$

Central moments for small k can always be written in terms of the moments or in terms of the factorial moments. When $k = 2$, a nice symmetry appears in either expression for the variance of X:

$$Var(X) = E(X^2) - (E(X))^2 = E([X]_2) - [E(X)]_2.$$

3.3 A statistical interpretation

Central moments are statistically meaningful for any random variable X where available. However, when X is a Bernoulli sum a few more meaningful interpretations are available.

Imagine a collection of individuals $i \in \mathcal{I}$, from which a random number X provides a population $\mathcal{J}$ of size X. Samples of fixed size k are to be drawn from the resulting population $\mathcal{J}$. Here, $X = \sum_{i\in\mathcal{I}} Y_i$ and $\mathcal{J} = \{i \in \mathcal{I} : Y_i = 1\}$ with (possibly dependent) random variables $Y_i \sim Bernoulli(p_i)$ (indicating inclusion in the population $\mathcal{J}$ when $Y_i = 1$ and exclusion when $Y_i = 0$).

In this case, the *expected number of samples of size k*

- is the kth factorial moment $E([X]_k)$ when sampling *without replacement* and
- is the kth moment $E(X^k)$ when sampling *with replacement*.

The expected factorial $E(X!)$ is the *expected number of permutations* one would have in the indices found by forming a population in this way.

3.4 Generating functions

Various generating functions for a Bernoulli sum X are now easily had by substitution of

- $E(X^k)$ in the *moment generating function*

$$M_X(s) = E(e^{sX}) = 1 + \sum_{k=1}^{\infty} E(X^k)\frac{s^k}{k!}$$

- $E([X]_k)$ in the *factorial moment generating function* (e.g., see Johnson, Kemp, & Kotz, 2005, p. 59)

$$H_X(s) = 1 + \sum_{k=1}^{\infty} E([X]_k)\frac{s^k}{k!}$$

- and, from Fréchet (1943),

$$Pr(X = x) = \sum_{j \geq x} (-1)^{x+j} \binom{j}{x} \frac{E([X]_j)}{j!},$$

  or, after substitution for the factorial moments,

$$Pr(X = x) = \sum_{j \geq x} (-1)^{x+j} \binom{j}{x} \sum_{\{i_1,\ldots,i_j\} \subseteq \mathcal{I}} E(Y_{i_1} \cdots Y_{i_j}),$$

  the probability $Pr(X = x)$ into the *probability generating function*

$$G_X(s) = E(s^X) = \sum_{k=0}^{\infty} s^k Pr(X = k).$$

The factorial moment generating function, $H_X(s)$, can be related (again, see Johnson et al., 2005, p. 59) to the probability generating function, $G_X(s)$, as

$$H_X(s) = G_X(1 + s) = E((1 + s)^X).$$

It follows that whenever factorial moments are such that $H_X(s)$ has a tidy closed form, the probability generating function of X might be easily obtained through the reverse relation

$$G_X(s) = H_X(s - 1). \tag{3}$$

This approach will be illustrated for the binomial distribution in Section 4.1, and for the classic matching problem of Section 4.5, to determine expressions for the probability generating function of X in each of these classic cases.

4 Classic examples

Bernoulli sums naturally arise in many classic problems and lead to well known distributions. In this section, the results of Section 3 are applied to several of these where the Bernoulli sum is over a finite index set (of size n), namely

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

where $Y_i \sim Bernoulli(p_i)$ with $p_i = Pr(Y_i = 1) = 1 - Pr(Y_i = 0)$ for $i = 1, \ldots, n$.

4.1 Binomial X

The simplest case where the Y_is are independent and identically distributed (i.i.d.) with $p_i = p \; \forall \; i$, $X \sim binomial(n, p)$. The kth moment of X can be written as

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} p^m \\ &= \sum_{m=1}^{k} S(k,m) \binom{n}{m} p^m. \end{aligned}$$

Knoblauch (2008) found an equivalent expression through a recursive argument. Here the result is easily had from the simple application of the more general Proposition 3.1. Central moments follow from Proposition 3.2:

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

The kth factorial moment has a appealingly simple expression $E([X]_k) = [n]_k \; p^k$ derived as

$$E([X]_k) = k! \sum_{\{i_1,\ldots,i_k\} \subseteq \mathcal{N}_n} p_{i_1} \cdots p_{i_k} = k! \binom{n}{k} p^k = [n]_k \; p^k,$$

the familiar $E(X) = np$ being the special case when $k = 1$.

Note that whenever $k \geq n$, many terms disappear in the above moment expressions since $S(n,m)$ vanishes whenever $m > n$ and the sum $\sum_{\{i_1,\ldots,i_k\} \subseteq \mathcal{N}_n}$ is over the empty set.

The moment generating function of a binomial X also has a new expression following application of Equation 4,namely

$$M_X(t) = 1 + \sum_{k \geq 1} \frac{t^k}{k!} \sum_{m=1}^{k} S(k,m) \binom{n}{m} p^m \tag{4}$$

compared to $M_X(t) = (1 - p + pe^t)^n$.

Recall that the probability generating function of $X \sim binomial(n, p)$ is

$$G_X(s) = ((1-p) + ps)^n .$$

By Equation 3, we find that the factorial moment generating function for X is

$$\begin{aligned} H_X(s) &= ((1-p) + p(1+s))^n \\ &= \sum_{m=0}^{n} s^m \left(\sum_{r=m}^{n} \sum_{\ell=0}^{n-r} \binom{n}{r} \binom{r}{m} (-1)^\ell p^{\ell+r} \right), \end{aligned}$$

by applying the binomial theorem and changing the order of summation. This provides us an additional expression for the $k-$th factorial moment of X:

$$[n]_k p^k = \sum_{r=k}^{n} \sum_{\ell=0}^{n-r} \binom{n}{r} \binom{r}{k} (-1)^\ell p^{\ell+r}.$$

4.1.1 Poisson binomial X

If $Y_i \sim Bernoulli(p_i)$ independently for all i but $p_i \neq p_j$ for (at least one) $i \neq j$, the distribution of X is called a Poisson binomial distribution (e.g., see Shah, 1973). The various moments of X are exactly as given by the relevant results of Section 3 with $E(Y_{i_1} \cdots Y_{i_m})$ everywhere replaced by $p_{i_1} \cdots p_{i_m}$. So too for its moment generating function.

4.2 Hypergeometric X

Consider a sample of size n randomly drawn without replacement from a population of N individuals where g of them have some trait which is absent from the remaining $N - g$. The indicator random variable, Y_i, records if the ith individual selected has the desired trait ($Y_i = 1$) or not ($Y_i = 0$) and $X = \sum_{i=1}^{n} Y_i$ counts the number in the sample having the trait.

The ith draw will be a Bernoulli random variable Y_i with probability

$$p_i = \frac{g - \ell}{N - (i-1)}$$

where ℓ is the number of previous $(i-1)$ draws having the trait. A sample of m of these Bernoullis will have

$$E(Y_{i_1} \cdots Y_{i_m}) = \frac{g(g-1)\cdots(g-m+1)}{N(N-1)\cdots(N-m+1)} = \frac{[g]_m}{[N]_m}$$

provided $m \leq g$ and will be zero whenever $m > g$ (since at least one Y_i must be zero).

The kth moment of X, following Proposition 3.1, is now

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{\min k,g} S(k,m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} \frac{[g]_m}{[N]_m} \\ &= \sum_{m=1}^{\min k,g} S(k,m) \binom{n}{m} \frac{[g]_m}{[N]_m}, \end{aligned}$$

where the last equality followed from $m-$symmetry. Similarly, the central moments are

$$E((X-\mu)^k) = \left(-n\frac{g}{N}\right)^k + \sum_{\ell=1}^{M}\binom{k}{\ell}\sum_{m=1}^{\ell} S(\ell,m)\binom{n}{m}\frac{[g]_m}{[N]_m}\left(-n\frac{g}{N}\right)^{k-\ell}$$

where $M = \min k, g$.

The factorial moments again have a pleasingly simple expression when $k \leq g$ (zero whenever $k > g$), namely,

$$E([X]_k) = \; k!\binom{n}{k}\frac{[g]_k}{[N]_k} \;=\; [n]_k\,\frac{[g]_k}{[N]_k}.$$

Where the binomial $E([X]_k) = [n]_k p^k$, the hypergeometric now has $\frac{[g]_k}{[N]_k}$ in place of p^k, as one might expect.

4.3 CMP-binomial X

For $n \in \mathbb{N}, p \in [0,1], \nu \in \mathbb{R}$, a random variable X has a *Conway-Maxwell-Poisson (CMP) binomial* distribution with parameters (n,p,ν) if its probability mass at $X = j$ ($j \in \mathcal{N}_n$) is given by

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

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

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

The distribution is formed from a Conway-Maxwell-Poisson (CMP) random variable conditional on the sum of that variable and another one independently generated from a different CMP-distribution.

Just as the CMP-distribution generalizes a Poisson random variable to model count data having variability larger (over dispersed) or smaller (under dispersed) than that of a Poisson, the *CMP-binomial* generalizes the binomial distribution. A CMP-binomial distribution is binomial when $\nu = 1$ and has larger (smaller) variance than a binomial when $\nu < 1$ ($\nu > 1$). When $\nu = 0$, the most extreme values of 0 and n are favoured; when $\nu \to \infty$ the count X achieves the middle value of $n/2$ when n is even and $(n \pm 1)/2$ when n is odd. See Shmueli, Minka, Kadane, Borle, and Boatwright (2005) for details.

Shmueli et al. (2005) remark that the random variable X can also be viewed as a sum of exchangeable, Bernoulli random variables Y_i with joint probability

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

where $\nu > 1$ in the case of negatively correlated trials and $\nu < 1$ for positively correlated trials. This observation allows an expression to be written for the moments of X from an expression $Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1)$ for an arbitrary $m-$set $\{i_1, \ldots, i_m\} \subseteq \mathcal{N}_n$.

$$\begin{aligned}
Pr(Y_{i_1}, \ldots, Y_{i_m}) &= \sum_{\substack{y_j \in \{0,1\} \\ \forall j \notin \{i_1,\ldots,i_m\}}} Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1, \text{ and } Y_j = y_j, \forall j \notin \{i_1, \ldots, i_m\}) \\
&= \sum_{\substack{y_j \in \{0,1\} \\ \forall j \notin \{i_1,\ldots,i_m\}}} \frac{1}{C_{n,p,\nu}} \binom{n}{m + \sum_{j \notin \{i_1,\ldots,i_m\}} y_j}^{\nu-1} \\
&\qquad \times p^{m+\sum_{j \notin \{i_1,\ldots,i_m\}} y_j} (1-p)^{n-(m+\sum_{j \notin \{i_1,\ldots,i_m\}} y_j)} \\
&= \frac{1}{C_{n,p,\nu}} \sum_{s=0}^{n-m} \sum_{\substack{y_j \in \{0,1\} \\ \forall j \notin \{i_1,\ldots,i_m\} \\ \sum_{j \notin \{i_1,\ldots,i_m\}} y_j = s}} \binom{n}{m+s}^{\nu-1} \times p^{m+s}(1-p)^{n-(m+s)} \\
&= \frac{1}{C_{n,p,\nu}} \sum_{s=0}^{n-m} \binom{n-m}{s} \binom{n}{m+s}^{\nu-1} p^{m+s}(1-p)^{n-(m+s)} \\
&= \frac{1}{C_{n,p,\nu}} \sum_{\ell=m}^{n} \binom{n-m}{\ell-m} \binom{n}{\ell}^{\nu-1} p^{\ell}(1-p)^{n-\ell}.
\end{aligned}$$

The kth moment of $X \sim CMP - binomial(n, p, \nu)$ is then

$$\begin{aligned}
E(X^k) &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} E(Y_{i_1} \cdots Y_{i_m}) \\
&= \frac{1}{C_{n,p,\nu}} \sum_{m=1}^{k} S(k,m) \binom{n}{m} \sum_{\ell=m}^{\min n,k} \binom{n-m}{\ell-m} \binom{n}{\ell}^{\nu-1} p^{\ell}(1-p)^{n-\ell} \\
&= \frac{1}{C_{n,p,\nu}} \sum_{m=1}^{k} S(k,m) \sum_{\ell=m}^{\min n,k} \binom{\ell}{m} \binom{n}{\ell}^{\nu} p^{\ell}(1-p)^{n-\ell}.
\end{aligned}$$

Similarly, the kth central moment is

$$(-np)^k + \frac{1}{C_{n,p,\nu}} \sum_{\ell=1}^{k} \binom{k}{\ell} (-np)^{k-\ell} \sum_{m=1}^{\ell} S(\ell,m) \sum_{\ell=m}^{\min n,k} \binom{\ell}{m} \binom{n}{\ell}^{\nu} p^{\ell}(1-p)^{n-\ell}$$

and the kth factorial moment

$$E([X]_k)\frac{k!}{C_{n,p,\nu}}\sum_{\ell=k}^{n}\binom{\ell}{k}\binom{n}{\ell}^{\nu}p^{\ell}(1-p)^{n-\ell}.$$

4.4 The empty urns problem

Consider the problem of assigning ℓ indistinguishable balls uniformly at random into n distinguishable urns. Let Y_i be 1 if urn i is empty and 0 otherwise, and $X=\sum_{i=1}^{n}Y_i$ be the Bernoulli sum counting the total number of empty urns.

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

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

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

$$Pr(Y_{i_1}=1,Y_{i_2}=1,\ldots,Y_{i_m}=1)=\frac{(n-1)(n-2)\cdots(n-m)}{(\ell+n-1)(\ell+n-2)\cdots(\ell+n-m)}=\frac{[n-1]_m}{[\ell+n-1]_m}.$$

The kth moment of X is

$$\begin{aligned}E(X^k)&=\sum_{m=1}^{k}\sum_{\{i_1,\ldots,i_m\}\subseteq\mathcal{N}_n}S(k,m)E(Y_{i_1}\cdots Y_{i_m})\\&=\sum_{m=1}^{k}S(k,m)\binom{n}{m}\frac{[n-1]_m}{[\ell+n-1]_m},\end{aligned}$$

the kth central moment (from Proposition 3.2)

$$E((X-\mu)^k)=(-\mu)^k+\sum_{\ell=1}^{k}\binom{k}{\ell}(-\mu)^{k-\ell}\sum_{m=1}^{\ell}S(\ell,m)\binom{n}{m}\frac{[n-1]_m}{[\ell+n-1]_m},$$

where $\mu=n\frac{n-1}{\ell+n-1}$. The kth factorial moment has a particularly simple representation as

$$E([X]_k)=\frac{[n]_k[n-1]_k}{[\ell+n-1]_k}.$$

Matching moments shows X of the urn problem to have a Hypergeometric distribution with parameters $(n,n-1,\ell+n-1)$.

4.5 The matching problem

The matching problem dates back to Montmort (de Montmort, 1713) and is the problem of taking n paired elements, randomly permuting the first elements over all pairs, then letting X be the number of correctly matched pairs after the random permutation. Examples are n letters matched correctly to n envelopes, couples separated at a dance and dance partners formed by randomly assigning one of each sex to the pair, and so on. The random variable X can be expressed as a sum of Bernoulli random variables taking value 1 when a correct match occurs and zero otherwise.

More abstractly, let $f : \mathcal{N}_n \to \mathcal{N}_n$ be a permutation on $\mathcal{N}_n$ and let X denote the number of fixed points of f (i.e., the number of $i \in \mathcal{N}_n$ for which $f(i) = i$). If f is picked uniformly at random from the set of all permutations on $\mathcal{N}_n$, denoted $Sym(n)$, then the distribution of X can be shown to tend to $Poisson(1)$ as $n \to \infty$. We do that by expressing X as a sum of Bernoullis and then examining and comparing moments.

Let Y_i be the Bernoulli random variable recording if $f(i) = i$. As f is chosen uniformly at random from $Sym(n)$,

$$Pr(Y_i = 1) = \frac{1}{n}$$

for all $i = 1, \ldots, n$.

Fix $m \leq n$ and consider an $m-$subset $\{i_1, \ldots, i_m\} \subseteq \mathcal{N}_n$. The probability that all $\{i_1, \ldots, i_m\}$ are fixed points of f is $(n-m)!/n!$. This is because if $f(i_j) = i_j$ for all $j \in \{1, \ldots, m\}$, one must only consider how to assign the other $(n-m)$ points so that f is a bijection. There are $(n-m)!$ ways to do this as $|Sym(n-m)| = (n-m)!$. Therefore,

$$Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1) = \frac{(n-m)!}{n!}.$$

By Proposition 3.1, for $k \leq n$

$$\begin{aligned}
E(X^k) &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1) \\
&= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\} \subseteq \mathcal{N}_n} \frac{(n-m)!}{n!} \\
&= \sum_{m=1}^{k} S(k,m) \binom{n}{m} \frac{(n-m)!}{n!} \\
&= \sum_{m=1}^{k} \frac{S(k,m)}{m!} \\
&= \sum_{m=1}^{k} S_2(k,m) \\
&= B_k,
\end{aligned}$$

where B_k is the $k-$th *Bell number*, the number of ways to partition a set of size k into a family of nonempty, unlabelled, pairwise disjoint subsets. On the other hand, for $k > n$, the

inner sum vanishes whenever $m > n$ and hence

$$\begin{aligned}
E(X^k) &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subseteq\mathcal{N}_n} Pr(Y_{i_1}=1,\ldots,Y_{i_m}=1) \\
&= \sum_{m=1}^{n} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subseteq\mathcal{N}_n} Pr(Y_{i_1}=1,\ldots,Y_{i_m}=1) \\
&= \sum_{m=1}^{n} \frac{S(k,m)}{m!} \\
&= \sum_{m=1}^{n} S_2(k,m) \\
&= B_k - \sum_{m=n+1}^{k} S_2(k,m),
\end{aligned}$$

which can be interpreted as the number of ways to partition a set of size k into at most n classes.

For $k \leq n$, the $k-$th factorial moments of X is given by

$$\begin{aligned}
E([X]_k) &= k! \sum_{\{i_1,\ldots,i_k\}\subseteq\mathcal{N}_n} Pr(Y_{i_1}=1,\ldots,Y_{i_k}=1) \\
&= k!\frac{(n-k)!}{n!}\binom{n}{k} \\
&= 1.
\end{aligned}$$

Therefore, the factorial moment generating function of X is given by

$$H_X(s) = \sum_{k=0}^{n} \frac{s^k}{k!},$$

which by Equation 3 gives us that the probability generating function of X is

$$\begin{aligned}
G_X(s) &= H_X(s-1) \\
&= \sum_{k=0}^{n} \frac{(s-1)^k}{k!} \\
&= \sum_{\ell=0}^{n} s^\ell \sum_{k=\ell}^{n} \binom{k}{\ell} \frac{(-1)^{k-\ell}}{k!},
\end{aligned}$$

which provides us with another method for deriving the probability distribution of X *without* using Principle of Inclusion/Exclusion or counting derangements in permutations.

Now, the exponential generating function for B_k given by (e.g., see Stanley, 2011, p. 74)

$$\sum_{k\geq 0} B_k \frac{t^k}{k!} = e^{e^t-1}$$

is a special case (viz., $\lambda = 1$) of the moment generating function for a Poisson(λ) random variable W

$$M_W(t) = \sum_{k \geq 0} E(W^k) \frac{t^k}{k!} = e^{\lambda(e^t - 1)}.$$

For $k \leq n$, the moments of X match those of $W \sim$ Poisson(1); as $n \to \infty$, the moment generating functions agree and X converges to a Poisson(1) random variable.

4.6 The Poisson limit of a binomial

Proposition 3.1 can also be used to provide a novel proof that a binomial(n, p) random variable approaches a Poisson(λ) with $np \to \lambda$ as $n \to \infty$.

For $X = \sum_{i=1}^{n} Y_i$, each Y_i is independent, identically distributed Bernoulli(p) random variable and, as seen earlier, the kth moment

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k} S(k,m) \binom{n}{m} p^m \\ &= \sum_{m=1}^{k} S(k,m) \binom{n}{m} \left(\frac{np}{n}\right)^m \\ &= \sum_{m=1}^{k} \frac{S(k,m)}{m!} \left(\frac{[n]_m}{n^m}\right) (np)^m . \end{aligned}$$

As $n \to \infty$, the ratio $\left(\frac{[n]_m}{n^m}\right) \to 1$, $np \to \lambda$, and

$$E(X^k) \to \sum_{m=1}^{k} \frac{S(k,m)}{m!} \lambda^m = \sum_{m=1}^{k} S_2(k,m) \lambda^m$$

which is the kth moment of a Poisson(λ) random variable expressed as a Touchard polynomial in λ (e.g., see Riordan, 1937). It follows that as $n \to \infty$, $X \sim binomial(n, p)$ converges to a Poisson(λ) with $\lambda = \lim_{n \to \infty} np$.

When $p = \frac{1}{n}$, then binomial X converges to Poisson(1) and its kth moment is the Bell number B_k, as in the matching problem of Section 4.5.

5 Counts more generally

Focus has been on Bernoulli sums that rise naturally as counts of events. In this section, we consider any "count random variable" N to be that having support on *any* subset of the extended natural numbers $\mathbb{N}_0$ (e.g., counts n would be impossible whenever $Pr(N = n) = 0$). Previous results are extended to N by matching it to a Bernoulli sum X constructed from the upper tail probabilities of N.

Proposition 5.1. *Let N denote any discrete random variable on $\mathbb{N}_0$. Consider the Bernoulli random variable Y_i indicating whether $N \geq i$ or not; that is*

$$Y_i = \begin{cases} 1 & \text{if} \quad N \geq i, \\ \\ 0 & \text{otherwise.} \end{cases}$$

Then the Bernoulli sum $X = \sum_{i=1}^{\infty} Y_i$ has the same distribution as N.

Proof.

$$\begin{aligned} Pr(X = x) &= Pr\left(\sum_{i=1}^{\infty} Y_i = x\right) \\ &= Pr([Y_i = 1, \forall i : i \leq x] \cap [Y_i = 0, \forall i : i > x]) \\ &= Pr(N = x). \end{aligned}$$

□

The moments of N are identified with those of X and the Bernoullis Y_i defined above. The results for the moments and factorial moments now follow.

Proposition 5.2. *For any discrete random variable N on $\mathbb{N}_0$, the kth moment of N is*

$$E(N^k) = \sum_{m=1}^{k} S(k, m) \sum_{M \geq m} \binom{M-1}{m-1} Pr(N \geq M),$$

and the kth factorial moments is

$$E([N]_k) = k! \sum_{M \geq k} \binom{M-1}{k-1} Pr(N \geq M).$$

Proof. From Proposition 3.1,

$$E(N^k) = \sum_{m=1}^{k} S(k, m) \sum_{\{i_1, \ldots, i_m\} \subset \mathbb{N}} Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1)$$

Now,

$$\begin{aligned} \sum_{\{i_1, \ldots, i_m\} \subset \mathbb{N}} Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1) &= \sum_{\{i_1, \ldots, i_m\} \subset \mathbb{N}} Pr(\max(i_1, \ldots, i_m) \leq N) \\ &= \sum_{M=1}^{\infty} \sum_{\{i_1, \ldots, i_{m-1}\} \subseteq M-1} Pr(N \geq M) \\ &= \sum_{M=1}^{\infty} \binom{M-1}{m-1} Pr(N \geq M). \end{aligned}$$

Therefore,

$$E(N^k) = \sum_{m=1}^{k} S(k,m) \sum_{M=1}^{\infty} \binom{M-1}{m-1} Pr(N \geq M).$$

For the factorial moment,

$$\begin{aligned}
m! \sum_{M=m}^{\infty} \binom{M-1}{m-1} Pr(N \geq M) &= m! \sum_{M \geq m} \sum_{\ell \geq M} Pr(N = \ell) \binom{M-1}{m-1} \\
&= m! \sum_{\ell \geq m} Pr(N = \ell) \sum_{M=1}^{\ell} \binom{M-1}{m-1} \\
&= m! \sum_{\ell \geq m} Pr(N = \ell) \binom{\ell}{m} \\
&= m! \sum_{\ell \geq m} \frac{[\ell]_m}{m!} Pr(N = \ell) \\
&= \sum_{\ell \geq m} [\ell]_m Pr(N = \ell) \\
&= \sum_{\ell \geq 0} [\ell]_m Pr(N = \ell) \\
&= E([N]_m) \quad \text{by definition.}
\end{aligned}$$

□

Note that an expression for $E(N^k)$ has also recently been derived by Chakraborti, Jardim, and Epprecht (2019, eq. (10)), namely

$$E(N^k) = \sum_{i=0}^{\infty} ((i+1)^k - i^k) Pr(N > i).$$

Chakraborti et al. (2019) claim their formulation to be the first for $E(N^k)$ expressed in terms of the upper tail probability of N; if so, then Proposition 5.2 may provide the second for $E(N^k)$ and the first for $E([N]_k)$. These results are best appreciated whenever the cumulative distribution function of N has form allowing simplification, especially when multiplied by binomial coefficients.

The remainder of this section explores application of Proposition 5.2 to several familiar cases.

5.1 Geometric distribution

Suppose interest lay in the number X of tosses of a coin at which the first head occurs; X is a geometric(p) distribution with p being the probability of heads ($p = 0.5$ for a fair coin). Surprisingly, the random variable X can be written as a Bernoulli sum $X = \sum_{i=1}^{\infty} Y_i$ for suitably defined Bernoulli Y_is, allowing the previous results to be applied.

The representation is as follows. Take $(Z_i)_{i\geq 1}$ to be the sequence of independent Bernoulli(p) random variables representing the sequence of potential coin tosses ($Z_i = 1$ for heads and zero otherwise). Let N be the index of the first $Z_i = 1$ in the sequence, that is

$$N = \min i : Z_i = 1.$$

Consider now the Bernoulli random variables formed from the upper tail of the distribution of the index N:

$$Y_i = \begin{cases} 1 & \text{if } N \geq i \\ 0 & \text{otherwise} \end{cases}$$

The Bernoulli sum $X = \sum_{i=1}^{\infty} Y_i$ is the number of coin tosses (Z_is) that have occurred when the first head ($Z_i = 1$) appears in the sequence. While $X = N$, each provides a different way of looking at the same random variable.

Given the Bernoulli sequence $(Y_i)_{i\geq 1}$ and any $m-$set of indices $\{i_1, \ldots, i_m\} \subseteq \mathcal{N}_n$, the joint probability of m Y_is can be written as

$$\begin{aligned} Pr(Y_{i_1} = 1, \ldots, Y_{i_m} = 1) &= Pr(N \geq i_1, \ldots, N \geq i_m) \\ &= Pr(N \geq \max i_1, \ldots, i_m). \end{aligned}$$

Then by Proposition 3.3

$$\begin{aligned} E(X^k) &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subset\mathbb{N}} E(Y_{i_1} \cdots Y_{i_m}) \\ &= \sum_{m=1}^{k} S(k,m) \sum_{\{i_1,\ldots,i_m\}\subset\mathbb{N}} Pr(N \geq \max i_1, \ldots, i_m) \\ &= \sum_{m=1}^{k} S(k,m) \sum_{M\geq 1} \sum_{\{i_1,\ldots,i_{m-1}\}\subseteq\mathcal{N}_{M-1}} Pr(N \geq M) \quad (M \text{ being the max index}) \\ &= \sum_{m=1}^{k} S(k,m) \sum_{M\geq 1} \binom{M-1}{m-1} [Pr(N > M) + Pr(N = M)] \\ &= \sum_{m=1}^{k} S(k,m) \sum_{M\geq 1} \binom{M-1}{m-1} [(1-p)^M + (1-p)^{M-1}p] \\ &= \sum_{m=1}^{k} S(k,m) \sum_{M\geq 1} \binom{M-1}{m-1} (1-p)^{M-1}. \end{aligned}$$

For an indeterminate y and $k \in \mathbb{N}$,

$$\sum_{n\geq 0} \binom{n}{k} y^n = \frac{y^k}{(1-y)^{k+1}}$$

from which it follows that

$$\sum_{M\geq 1}\binom{M-1}{m-1}(1-p)^{M-1}=\frac{(1-p)^{m-1}}{p^m}$$

giving the expression for the kth moment to be

$$E(X^k)=\sum_{m=1}^{k}S(k,m)\frac{(1-p)^{m-1}}{p^m}.$$

Similarly, the kth factorial moment has the simpler expression

$$E([X]_k)=\frac{(1-p)^{k-1}}{p^k}.$$

We note that since the probability generating function of X has the closed form expression

$$G_X(s)=\frac{p}{1-s(1-p)},$$

by Equation 3, the factorial moment generating function is given by

$$\begin{aligned}H_X(s)&=\frac{p}{1-(1+s)(1-p)}\\&=p\sum_{\ell\geq 0}[(1+s)(1-p)]^{\ell}.\end{aligned}$$

5.2 Poisson distribution

In Section 4.6 the tidy expression

$$E(N^k)=\sum_{m=1}^{k}S_2(k,m)\lambda^m$$

for the kth moment of $N\sim Poisson(\lambda)$ appeared. The proof of this result given by Riordan (1937) is recursive. It can also be proved by direct application of Proposition 5.2 as follows:

$$\begin{aligned}E(N^k)&=\sum_{m=1}^{k}S(k,m)\sum_{M=m}^{\infty}\binom{M-1}{m-1}Pr(N\geq M)\\&=\sum_{m=1}^{k}S(k,m)\sum_{M=m}^{\infty}\binom{M-1}{m-1}\sum_{\ell=M}^{\infty}e^{-\lambda}\frac{\lambda^{\ell}}{\ell!}\\&=e^{-\lambda}\sum_{m=1}^{k}S(k,m)\sum_{\ell=m}^{\infty}\frac{\lambda^{\ell}}{\ell!}\sum_{M=1}^{\ell}\binom{M-1}{m-1}\quad\text{by reorganizing the sums}\\&=e^{-\lambda}\sum_{m=1}^{k}S_2(k,m)m!\sum_{\ell=m}^{\infty}\frac{\lambda^{\ell}}{\ell!}\binom{\ell}{m}\end{aligned}$$

as $\sum_{M=1}^{\ell} \binom{M-1}{m-1} = \binom{\ell}{m}$,

$$
\begin{aligned}
&= e^{-\lambda} \sum_{m=1}^{k} S_2(k,m)\lambda^m \sum_{\ell=m}^{\infty} \frac{\lambda^{(\ell-m)}}{(\ell-m)!} \\
&= e^{-\lambda} \sum_{m=1}^{k} S_2(k,m)\lambda^m e^{\lambda} \\
&= \sum_{m=1}^{k} S_2(k,m)\lambda^m.
\end{aligned}
$$

Following the same route as for $E(N^k)$, the kth factorial moment of $N \sim Poisson(\lambda)$ has the even simpler expression:

$$E([N]_k) = \lambda^k.$$

5.3 Ideal soliton distribution

For an integer r with $r \geq 2$, we say that N follows the ideal soliton distribution, $soliton(r)$, when

$$
\begin{aligned}
Pr(N = 1) &= \frac{1}{r}, \\
Pr(N = i) &= \frac{1}{i(i-1)},
\end{aligned}
$$

for $i \in \{2, \ldots, r\}$ and is zero otherwise. It can be shown by induction that

$$\sum_{i=2}^{\ell} \frac{1}{i(i-1)} = \frac{\ell - 1}{\ell}.$$

From this it follows that

$$Pr(N \leq \ell) = \frac{1}{r} + \frac{\ell - 1}{\ell} \quad \text{for } \ell = 1, 2, \ldots r$$

and

$$Pr(N \geq \ell) = \begin{cases} 1 & \text{when } \ell = 1 \\ \\ \frac{r-1}{r} - \frac{\ell-2}{\ell-1} & \ell = 2, 3, \ldots, r. \end{cases}$$

To derive the moments and factorial moments, we need the following lemma.

Lemma 5.3. *For any $m, r \in \mathbb{N}$,*

$$\sum_{M=0}^{r} M[M]_m = (m+1)!\binom{r+1}{r-m-1} + m \cdot m!\binom{r+1}{r-m}.$$

Proof. First, an expression for the ordinary generating function $f(x) = \sum_{i\geq 0} i[i]_{m-2}\ x^j$ corresponding to the sequence $(i[i]_{m-2})_{i\geq 0}$ is determined. Multiplying by $\frac{1}{1-x}$ then gives the generating series for $(\sum_{i=0}^{r} i[i]_{m-2})_{r\geq 0}$.

Since $\sum_{i\geq 0} x^i = \frac{1}{1-x}$, differentiating with respect to x and then multiplying by x gives

$$\sum_{i\geq 0} ix^i = \frac{x}{(1-x)^2}.$$

Differentiating with respect to x, $(m-2)$ times, the left hand side becomes

$$\sum_{i\geq 0} i \times i(i-1)\cdots(i-m-1)\ x^{i-m-2} = \sum_{i\geq 0} i[i]_{m-2}\ x^{i-m-2}.$$

Applying the general Leibniz rule, the same derivative of right hand side is

$$\frac{x(m-1)!}{(1-x)^m} + (m-2)\frac{(m-2)!}{(1-x)^{m-1}}.$$

Multiplying both sides by $\dfrac{x^{m+2}}{(1-x)}$ gives

$$\sum_{r\geq 0}\sum_{i=0}^{r} i[i]_{m-2}x^r = \frac{x^{m+3}(m-1)!}{(1-x)^{m+1}} + (m-2)\frac{x^{m+2}(m-2)!}{(1-x)^m}.$$

Since

$$\frac{1}{(1-x)^k} = \sum_{n\geq 0}\binom{n+k-1}{n}x^n,$$

the claim immediately follows. □

Proposition 5.4. *Let N follow the soliton(r) distribution with $r \geq 2$. For $k \geq 1$, the moments of N are given by*

$$E(N^k) = H_r + \sum_{m=2}^{k} S(k,m)\left[\frac{r-1}{r}\binom{r}{m} - \binom{r-3}{m} - (m-2)^2\binom{r-3}{m-1}\right]$$

For $k \geq 2$, the $k-$th factorial moment of N is

$$E([N]_k) = k!\left[\frac{r-1}{r}\binom{r}{k} - \binom{r-3}{k} - (k-2)^2\binom{r-3}{k-1}\right].$$

Proof. Once more, by Proposition 3.1 and Proposition 3.6, we need to evaluate sums of the form

$$\sum_{M=m}^{r}\binom{M-1}{m-1}Pr(N \geq M).$$

If $m = 1$, this becomes

$$\begin{aligned}\sum_{M=m}^{r} \binom{M-1}{m-1} Pr(N \geq M) &= \sum_{M=1}^{r} Pr(N \geq 1) \\ &= \sum_{M=1}^{r} \sum_{\ell=M}^{r} Pr(N = \ell) \\ &= \sum_{\ell=1}^{r} \sum_{M=1}^{\ell} Pr(N = \ell) \\ &= \sum_{\ell=1}^{r} \ell \; Pr(N = \ell) \\ &= \frac{1}{r} + \sum_{j=2}^{r} \frac{j}{j(j-1)} = H_r.\end{aligned}$$

Otherwise, for $m \geq 2$,

$$\begin{aligned}\sum_{M=m}^{r} \binom{M-1}{m-1} Pr(X \geq M) &= \sum_{M=m}^{r} \binom{M-1}{m-1} \left[\frac{r-1}{r} - \frac{M-2}{M-1}\right] \\ &= \frac{r-1}{r} \sum_{M=m}^{r} \binom{M-1}{m-1} - \sum_{M=m}^{r} \binom{M-1}{m-1} \frac{M-2}{M-1}.\end{aligned}$$

By straightforward algebraic manipulation,

$$\sum_{M=m}^{r} \binom{M-1}{m-1} \frac{M-2}{M-1} = \frac{1}{(m-1)!} \sum_{M=m}^{r} (M-2)[M-2]_{m-2}.$$

By replacing m and r by $m-2$ and $r-2$, respectively, in Lemma 5.3 and dividing by $(m-1)!$ it follows that

$$\begin{aligned}\frac{1}{(m-1)!} \sum_{M=m}^{r} (M-2)[M-2]_{m-2} &= \frac{1}{(m-1)!} \sum_{M=m}^{r-2} (M)[M]_{m-2} \\ &= \binom{r-1}{r-m-1} + (m-2)^2 \binom{r-1}{r-m}.\end{aligned}$$

The result now follows from a straightforward application of Propositions 3.6 and 3.1. □

5.4 Benford distribution

Benford's distribution encapsulates the notion that in many real world settings, the leading digits in a numerical data set are more likely to be small. In particular, we say that D follows Benford's distribution if for a digit $d \in \{1, \ldots, 9\}$,

$$Pr(D = d) = \log_{10}(d+1) - \log_{10}(d).$$

Due to its telescoping nature, the complementary CDF of D is given by

$$Pr(D \geq d) = 1 - \log_{10}(d).$$

Therefore, by Proposition 5.2, the $k-$th moment of D is

$$\begin{aligned} E(D^k) &= \sum_{m=1}^{k} S(k,m) \sum_{M=m}^{9} \binom{M-1}{m-1} (1 - \log_{10}(M)) \\ &= \sum_{m=1}^{k} S(k,m) \left[\binom{9}{m} - \sum_{M=m}^{9} \binom{M-1}{m-1} \log_{10}(M) \right], \end{aligned}$$

and factorial moments with the form

$$E([D]_k) = k! \left[\binom{9}{k} - \sum_{M=k}^{9} \binom{M-1}{k-1} \log_{10}(M) \right],$$

for $k \leq 9$. Of course, the results above extend to any general base b by noting that $D_b \sim$ Benford(b) satisfies

$$\sum_{M=m}^{b-1} \binom{M-1}{m-1} Pr(D_b \geq M) = \left[\binom{b-1}{m} - \sum_{M=k}^{b-1} \binom{M-1}{k-1} \log_b(M) \right].$$

6 Concluding remarks

A novel multinomial theorem for commutative idempotents (Proposition 2.2) led to new general expressions for the moments (including central and factorial) of a Bernoulli sum (e.g., Propositions 3.1 to 3.3) as well as corresponding generating functions. The general expressions depend on the determination of the expected product of subsets of the Bernoulli random variables. By evaluating these in particular cases a number of new expressions for moments and generating functions of many common distributions and classic problems.

The success of the approach in these examples mark it as potentially fruitful in more novel distributions and problems where this expectation might be more readily available. To that end, the representation of $\binom{X}{m}$ for random count X and fixed m (Proposition 3.6), and of $X!$ (Proposition 3.7), as the product of Bernoullis may also be more generally useful.

In other instances, the general representation of the various moments for a count variable N expressed in terms of the upper probability of that N (Proposition 5.2) may be valuable in yet other problems, as shown in the examples of Section 5.

The framework of Bernoulli sum random variables appears a viable tool for problems involving count data.

References

Chakraborti, S., Jardim, F., & Epprecht, E. (2019). Higher-order moments using the survival function: The alternative expectation formula. *The American Statistician*, *73*(2), 191-194. doi: 10.1080/00031305.2017.1356374

de Montmort, P. R. (1713). *Essay d'analyse sur les jeux de hazard...* J. Quillau.

Fréchet, M. (1943). Sur l'extension de certaines évaluations statistiques au cas de petits échantillons. *Revue de l'Institut International de Statistique*, 182–205.

Goulden, I. P., & Jackson, D. M. (2004). *Combinatorial Enumeration*. Dover Publications.

Iyer, P. K. (1958). A theorem on factorial moments and its applications. *The Annals of Mathematical Statistics*, *29*(1), 254–261.

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate Discrete Distributions* (Vol. 444). John Wiley & Sons.

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

Riordan, J. (1937). Moment recurrence relations for binomial, poisson and hypergeometric frequency distributions. *The Annals of Mathematical Statistics*, *8*(2), 103–111.

Shah, B. (1973). Distribution of sum of independent integer valued random-variables. *The American Statistician*, *27*(3), 123–124.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. (2005). 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.

Stanley, R. P. (2011). *Enumerative Combinatorics, Volume I* (2nd ed.).

Wilf, H. S. (2005). *generatingfunctionology*. A K Peters/CRC Press.