# Computing Moments of Discrete Distributions by Finite Calculus

Sebastiano Vigna*

**Abstract**

Moments of discrete distributions have been derived several times by operator methods: summing operator series, applying differential operators, or reading recurrences off differential equations for the generating functions. The remaining approach, summation by parts (the integration by parts of finite calculus), has been used repeatedly, but never as a method in its own right; we develop it systematically. Together with the absorption identity for binomial coefficients, it computes the factorial moments of the classical discrete distributions in one or two lines each, and raw moments follow mechanically through Stirling numbers. It continues to work when the moments have no closed form uniform in the parameters, producing exact identities and fixed-parameter recurrences. Along the way we re-derive in a few lines a recurrence of Kemp for the factorial moments of Fisher's noncentral hypergeometric distribution, showing that, except in the central case, it makes every factorial moment an elementary function of the mean. We also observe that the Conway–Maxwell–binomial distribution with $\nu = 2$ is a noncentral hypergeometric distribution, and we obtain a recurrence with a drifting center for the factorial moments about the mean of the Poisson distribution, mirroring the classical recurrence for the central moments of the Gaussian distribution.

*Keywords:* Difference equations; Factorial moments; Stein's method; Summation by parts.

## 1 Introduction

Suppose you need the high-order moments of a binomially distributed random variable $S$ with parameters $n$ and $p$ in closed form. They are trivial to compute using the finite calculus popularized by *Concrete Mathematics* (Graham et al. 1994), which yields the formula

$$\mathbf{E}(S^d) \;=\; \sum_{i=0}^{d} \left\{ {d \atop i} \right\} n^{\underline{i}} p^i, \tag{1}$$

where $\left\{ {d \atop i} \right\}$ is a Stirling number of the second kind and $n^{\underline{i}} = n(n-1)\cdots(n-i+1)$ is a falling factorial power.

The formula is known, but it is surprisingly hard to locate in standard references. It appears in the equivalent "differences of zero" form (the values $\Delta^i x^d$ at $x = 0$) in a note by Link (1981), via the general Stirling-number conversion in a classroom note by Joarder & Mahmood (1997), and, proved by manipulating the derivatives of the moment generating function, in a relatively recent paper by Knoblauch (2008).

The purpose of this paper is to show that results such as (1) follow uniformly from a few elementary tools of finite calculus.

The finite calculus we use is that of *Concrete Mathematics* (Graham et al. 1994). Its classical systematic account is the book of Jordan (1950), but its ideas go back to Euler. In this calculus

*Dipartimento di Informatica, Università degli Studi di Milano, Italy
(e-mail: sebastiano.vigna@unimi.it).

the falling power $x^{\underline{m}}$ is the analogue of the monomial $x^m$, and the *factorial moments* $\mathbf{E}(X^{\underline{m}})$ are correspondingly the natural moments of a distribution on the natural numbers. Finite calculus computes them in one or two lines for essentially all classical distributions, using a few tools. The main one is summation by parts: we use it directly, together with the fundamental theorem of finite calculus (Section 3), and through a tail formula generalizing $\mathbf{E}(X) = \sum_{x\geq 0} \mathbf{P}(X > x)$ (Section 4). We also use it as a Stein-type index shift, that is, summation by parts with a vanishing boundary term: the shift covers the whole Katz family (Katz 1965) at once (Section 5) and extends beyond it (Section 7).

Absorption (Section 6) is of a different nature: it is an algebraic identity for binomial coefficients, which turns a factorial moment into a constant times the total mass of the same distribution with different parameters. Through negative falling powers, it also gives inverse moments such as $\mathbf{E}(1/(X+1))$. Raw moments are finally recovered by the change of basis $x^k = \sum_j \left\{ {k \atop j} \right\} x^{\underline{j}}$ (Section 8), which involves no summation at all. Applied to $x - \mu$, the change of basis also gives central moments from factorial moments about the mean, for which the index shift yields, in the Poisson case, a recurrence with a drifting center (Section 8.1).

The same calculus handles the two cases that Link's operator method, recalled in Section 1.1, could not manage. These are the negative binomial distribution, where "no way was found to get rid of the double summation," and the negative hypergeometric distribution, which he left as an open problem. The calculus also continues to work when the normalizing constant of the probability mass function is not an elementary function of the parameters (Section 7). We do not need generating functions or combinatorial identities beyond absorption, the binomial theorem, Vandermonde's convolution, and the Stirling change of basis.

Of course, everything below is formally equivalent to manipulating the factorial moment generating function $\mathbf{E}((1+t)^X)$, whose coefficients are the binomial moments $\mathbf{E}\binom{X}{m}$. The advantage of the finite-calculus approach is practical rather than logical: it operates directly on the sums, with explicit bookkeeping of boundary terms, and it needs neither differentiation under the expectation nor any argument at the boundary of a disk of convergence.

Section 1.1 reviews the related literature. Section 2 recalls the small amount of finite calculus we need, and it ends with a guide to the choice of the tool for a given distribution. Table 1 lists the distributions we treat, the tool and the section used for each result, and whether the result is known or, to our knowledge, new. Everything we do not mark as new is classical or recent knowledge, for which we aim to give the shortest self-contained derivations we can. Section 9 draws the analogy between continuous and discrete computations.

### 1.1 Related work

Moments of discrete distributions have been derived by operator methods before, but always through difference or differential operators, that is, on the derivative side of the calculus. Link computed raw moments of the binomial, Poisson, geometric, and hypergeometric distributions by writing $j^r = T^j 0^r$, where $T$ is the shift operator of Section 2, and summing the resulting operator series over differences of zero (Link 1981). Boullion, Seaman and Young gave an analogue with the operator $x\,d/dx$ (Boullion et al. 1992), and Kemp read factorial-moment recurrences off differential equations for the generating functions of the generalized hypergeometric family (Kemp 1968). These recurrences are *fixed-parameter*: they link factorial moments of consecutive orders at the same values of the parameters.

Summation, by contrast, has been used repeatedly but always instrumentally. Guldberg multiplied the difference equations of the classical distributions by powers and summed them to obtain moment recurrences (Guldberg 1931). Katz (1965), in his account of the family, did the same "entirely formally" to relate the parameters to the first moments, and Ord (1967) and Tripathi & Gurland (1977) did it with falling factorials, obtaining fixed-parameter factorial-moment recurrences, to estimate the parameters of their wider systems.

Link remarked that "derivation of moments for discrete distributions is essentially a problem

Table 1: The results of the paper by distribution, with the tool and the section for each. Direct summation, the tail formula, and the index shift are all forms of summation by parts; $\nu$ is the parameter of the Conway–Maxwell families, $\kappa$ a parameter of the Bessel distribution, and $\omega$ the odds ratio of the noncentral hypergeometric distribution. Sources for the known results are given in the text; for the classical closed forms of Sections 3, 4, and 6, see the general references in Section 1.1.

| Distribution | Result | Tool | § | Status |
|---|---|---|---|---|
| Uniform, geometric, logarithmic | factorial moments | direct summation | 3 | known |
| Geometric | factorial moments | tail formula | 4 | known |
| Katz family (Table 2) | factorial moments | index shift | 5 | known |
| Hypergeometric | factorial moments | absorption | 6 | known |
| Beta-binomial, neg. hypergeometric | factorial moments | absorption | 6 | known |
| Binomial, Poisson | inverse moments | absorption | 6 | known |
| Conway–Maxwell–Poisson | $\mathbf{E}(X^\nu)$, $\mathbf{E}\big((X+1)^{-\nu}\big)$ | index shift | 7.1 | known |
| $\nu = 2$ | factorial-moment recurrence | index shift | 7.1 | known |
| integer $\nu \geq 3$ | factorial-moment recurrence | index shift | 7.1 | new* |
| Bessel | identity for $\mathbf{E}\big(X(X+\kappa)\big)$ | index shift | 7.1 | known |
| Noncentral hypergeometric | recurrence (8) | index shift | 7.2 | known |
| $\omega \neq 1$ | all moments from the mean | index shift | 7.2 | new |
| Conway–Maxwell–binomial | identity for $\mathbf{E}(X^\nu)$ | index shift | 7.3 | known |
| $\nu = 2$ | is noncentral hypergeometric | — | 7.3 | new |
| Conway–Maxwell–neg. binomial | identity for $\mathbf{E}(X^\nu)$ | index shift | 7.4 | known |
| $\nu = 2$ | factorial-moment recurrence | index shift | 7.4 | new* |
| All | raw moments | change of basis | 8 | known |
| Poisson | drifting-center recurrence | index shift | 8.1 | new |

*As explicit instances; recurrences of this shape also follow, in principle, from the general theorems of Kemp (1968).

in summation, the finite difference equivalent of integration in calculus." Yet, to our knowledge, summation has never been developed as a method in its own right. Beyond the formal summations recalled above, it appears as isolated identities in the Conway–Maxwell literature (Shmueli et al. 2005, Sellers & Shmueli 2010), and systematically only in the Stein-method literature, where approximation rather than moment computation is the primary aim. The index-shift identity at the heart of Sections 5 and 7 is the characterizing identity of the birth–death Stein operators of that literature (Brown & Xia 2001, Ley et al. 2017, Upadhye et al. 2017). Moment computation has nonetheless attracted interest there recently: Stein identities have been used as estimating equations (Ebner et al. 2025) and as recurrences for exact moments, for instance of Wishart matrices (Gaunt & Ouimet 2026).

Factorial moments were introduced by Steffensen (1923), and Ottestad (1939), continuing the work of Guldberg (1931), already used ratios of consecutive factorial moments as goodness-of-fit criteria, taking the closed forms from Steffensen. That these closed forms are simple is classical knowledge (Riordan 1937, Potts 1953, Johnson et al. 2005), and the monograph of Charalambides (2005) builds the theory of discrete distributions on finite differences. Yet even Di Nardo & Senato (2006), who develop the umbral calculus of factorial moments, remark that factorial moments "do not play a very prominent role in statistics."

## 2 Finite calculus in a nutshell

We recall the small amount of machinery we need, following the notation of Graham et al. (1994). The *difference operator* and the *shift operator* act on a function $g\colon \mathbf{Z} \to \mathbf{R}$ by

$$\Delta g(x) = g(x+1) - g(x), \qquad Tg(x) = g(x+1).$$

The *falling factorial power* is defined for $m \geq 0$ by

$$x^{\underline{m}} = x(x-1)\cdots(x-m+1),$$

with $x^{\underline{0}} = 1$ (an empty product), and for negative exponents by

$$x^{\underline{-m}} = \frac{1}{(x+1)(x+2)\cdots(x+m)},$$

the definition being suggested by the rule $x^{\underline{m+k}} = x^{\underline{m}}\,(x-m)^{\underline{k}}$ (take $k=-m$ and replace $x$ by $x+m$). Its case $m=1$, $x\,(x-1)^{\underline{k}} = x^{\underline{k+1}}$, is used throughout. We write $x^{\overline{m}} = x(x+1)\cdots(x+m-1)$ for rising powers, so that $x^{\underline{-m}} = 1/(x+1)^{\overline{m}}$.

The fundamental *power rule*

$$\Delta\, x^{\underline{m}} = m\, x^{\underline{m-1}} \tag{2}$$

holds for *all* integers $m$, exactly as $(x^m)' = mx^{m-1}$ does in ordinary calculus. We also use $\Delta\, c^x = (c-1)\,c^x$ and, writing $\binom{x}{m} = x^{\underline{m}}/m!$ for integers $m \geq 0$ and $\binom{x}{m} = 0$ for $m < 0$, $\Delta\binom{x}{m} = \binom{x}{m-1}$; the latter is (2) divided by $m!$, that is, Pascal's rule. Finally, the definitions give, for every constant $c$,

$$(x+c)\,x^{\underline{k}} = x^{\underline{k+1}} + (k+c)\,x^{\underline{k}}; \tag{3}$$

in particular, $x\,x^{\underline{k}} = x^{\underline{k+1}} + k\,x^{\underline{k}}$.

If $\Delta g = f$, then $g$ is an *indefinite sum* (antidifference) of $f$, and the *fundamental theorem of finite calculus* states that

$$\sum_a^b f(x)\,\delta x \;:=\; \sum_{x=a}^{b-1} f(x) \;=\; g(b) - g(a) \;=\; \big[g(x)\big]_a^b.$$

For $b=\infty$ we read the sum as $\sum_{x\geq a} f(x)$ and $[g(x)]_a^\infty$ as $\lim_{x\to\infty} g(x) - g(a)$. Applying $\Delta$ to a product, $\Delta(uv) = u\,\Delta v + (Tv)\,\Delta u$, and summing, we obtain *summation by parts*:

$$\sum_a^b u(x)\,\Delta v(x)\,\delta x = \big[u(x)v(x)\big]_a^b - \sum_a^b Tv(x)\,\Delta u(x)\,\delta x, \tag{4}$$

the shift $T$ being the only difference from integration by parts.

Throughout the paper $X$ is a random variable with values in $\mathbf{N} = \{0,1,2,\dots\}$ and probability mass function $f(x) = \mathbf{P}(X=x)$. For $m \in \mathbf{Z}$ we call $\mu_{(m)} = \mathbf{E}(X^{\underline{m}})$ the *$m$-th factorial moment* of $X$, and we occasionally use the *binomial moment* $\mathbf{E}\binom{X}{m} = \mu_{(m)}/m!$ ($m \geq 0$), which yields the cleanest closed forms; $[P]$ is the Iverson bracket, equal to 1 if $P$ holds and to 0 otherwise.

In Sections 3, 4, and 6 all sums have nonnegative summands, so no convergence subtleties arise: identities hold in $[0,\infty]$, and moments are finite exactly when the closed forms are. The manipulations of Sections 5, 7, and 8.1 do involve signed summands, but every distribution appearing there has finite support or geometric (or faster) tail decay, so all polynomial moments converge absolutely and the rearrangements are legitimate.

We close this section with a guide to the choice of the tool, for readers who want to apply the method to a distribution not treated here. The form of the probability mass function $f$ suggests the tool:

1. If $f(x)$ contains a binomial coefficient $\binom{n}{x}$, or a factorial $x!$ in the denominator, the falling power $x^{\underline{m}}$ can be absorbed into it, by (6) or by $x^{\underline{m}}/x! = 1/(x-m)!$ for $x \geq m$. The factorial moment is then the constant left in front, times the total mass of another distribution of the same family (Section 6).

2. If the ratio $f(x+1)/f(x)$ is a rational function of $x$, we write it as a difference equation $a(x)f(x+1) = b(x)f(x)$ with polynomial coefficients, valid for all $x \in \mathbf{N}$, and we apply the index shift (Lemma 7.1). In all our examples $a(x-1) = x\,h(x)$ for a polynomial $h$, and the choice $g(x) = (x-1)^{\underline{m}}$ gives $a(x-1)g(x) = h(x)\,x^{\underline{m+1}}$. Expanding both sides in falling powers with (3) and taking expectations, we obtain a linear recurrence among consecutive factorial moments.

3. If the tail $\mathbf{P}(X > x)$ is simpler than $f$, the tail formula gives the binomial moments directly (Section 4).

4. In all cases, raw moments follow from factorial moments, and central moments from factorial moments about the mean, by the change of basis of Section 8.

A recurrence linking $t$ consecutive factorial moments, whose coefficient of the highest one never vanishes, determines all of them from $\mu_{(0)} = 1$ and the next $t - 2$. For the Katz family $t = 2$, so the factorial moments have closed forms (Section 5); for the noncentral hypergeometric distribution $t = 3$, and, for $\omega \neq 1$, the mean determines all of them (Section 7.2); for the Conway–Maxwell–Poisson distribution with positive integer parameter $\nu$ we have $t = \nu + 1$ (Section 7.1).

The difference equation of item 2 must hold for all $x \in \mathbf{N}$ (Section 5). When it fails at a single point, as for a distribution truncated at zero, it is simpler to treat the untruncated distribution: the factorial moments of order $m \geq 1$ of the truncation at zero are those of the untruncated distribution divided by $1 - f(0)$.

# 3 Summing directly

We begin with three distributions whose factorial moments follow from the fundamental theorem or one summation by parts.

**Example 3.1** (Discrete uniform)**.** Let $X$ be uniform on $\{1, \dots, n\}$, where $n \geq 1$. By the fundamental theorem applied to $\Delta\binom{x}{m+1} = \binom{x}{m}$ (Pascal's rule; the resulting evaluation is the "hockey stick" identity),

$$\mathbf{E}\binom{X}{m} = \frac{1}{n}\sum\nolimits_1^{n+1}\binom{x}{m}\,\delta x = \frac{1}{n}\left[\binom{x}{m+1}\right]_1^{n+1} = \frac{1}{n}\binom{n+1}{m+1}$$

for $m \geq 1$, as then $\binom{1}{m+1} = 0$. For $m = 1, 2$ this gives $\mathbf{E}(X) = (n+1)/2$ and $\mathbf{E}(X^{\underline{2}}) = (n^2-1)/3$, whence $\mathbf{E}(X^2) = (n^2-1)/3 + (n+1)/2 = (n+1)(2n+1)/6$, the usual formula.

**Example 3.2** (Geometric)**.** Let $\mathbf{P}(X = x) = pq^x$ on $\mathbf{N}$, where $0 < p \leq 1$ and $q = 1 - p$. Let

$$S_m = \sum\nolimits_0^{\infty} x^{\underline{m}}\, q^x\, \delta x.$$

Since $-q^x/p$ is an antidifference of $q^x$, summation by parts (4) with $u(x) = x^{\underline{m}}$ gives, for $m \geq 1$,

$$S_m = \left[-x^{\underline{m}}\,\frac{q^x}{p}\right]_0^{\infty} + \frac{mq}{p}\sum\nolimits_0^{\infty} x^{\underline{m-1}} q^x\,\delta x = \frac{mq}{p}\,S_{m-1},$$

because the boundary term vanishes at both ends ($0^{\underline{m}} = 0$ and $q^x x^{\underline{m}} \to 0$). Since $S_0 = 1/p$, unrolling the recurrence yields $S_m = m!\,q^m/p^{m+1}$ and therefore

$$\mathbf{E}(X^{\underline{m}}) = pS_m = m!\left(\frac{q}{p}\right)^m, \qquad \mathbf{E}\binom{X}{m} = \left(\frac{q}{p}\right)^m.$$

This computation parallels step by step the inductive evaluation of $\int_0^{\infty} x^m \lambda e^{-\lambda x}\,dx = m!/\lambda^m$.

**Example 3.3** (Logarithmic)**.** Let $\mathbf{P}(X = x) = q^x/(cx)$ on $\{1, 2, \dots\}$, where $0 < p < 1$, $q = 1 - p$, and $c = \ln(1/p)$. For $m \geq 1$ the falling power absorbs the denominator, as $x^{\underline{m}}/x = (x-1)^{\underline{m-1}}$, so by Example 3.2

$$\mathbf{E}(X^{\underline{m}}) = \frac{1}{c}\sum_{x\geq 1}(x-1)^{\underline{m-1}}\,q^x = \frac{q}{c}\,S_{m-1} = \frac{(m-1)!}{c}\left(\frac{q}{p}\right)^m.$$

## 4 The tail formula

The identity $\mathbf{E}(X) = \sum_{x\ge 0} \mathbf{P}(X > x)$ is classical (see, e.g., Theorem 2 in Section XI.1 of Feller 1968), and its general form

$$\mathbf{E}(g(X)) = g(0) + \sum_{x\ge 0} \Delta g(x)\, \mathbf{P}(X > x)$$

is, for nondecreasing $g$, a standard exercise in Abel summation, that is, in summation by parts. For $g(x) = \binom{x}{m}$ the formula takes a particularly clean form, and (4) gives it in a single step, with an explicit analysis of the boundary term. The tail formula is the natural tool when the tail $\mathbf{P}(X > x)$ is simpler than the probability mass function, as for the minimum of independent variables, whose tail is the product of their tails.

**Theorem 4.1.** *Let $X$ be a random variable with values in $\mathbf{N}$ and let $m \ge 1$. Then*

$$\mathbf{E}\binom{X}{m} = \sum_{x\ge 0} \binom{x}{m-1} \mathbf{P}(X > x),$$

*both sides being possibly infinite.*

*Proof.* Let $G(x) = \mathbf{P}(X > x)$ and $v(x) = -G(x-1)$, so that $\Delta v(x) = G(x-1) - G(x) = \mathbf{P}(X = x) = f(x)$. Summation by parts (4) with $u(x) = \binom{x}{m}$ gives

$$\sum\nolimits_0^L \binom{x}{m} f(x)\, \delta x = \Big[-\binom{x}{m} G(x-1)\Big]_0^L + \sum\nolimits_0^L \binom{x}{m-1} G(x)\, \delta x.$$

The boundary term vanishes at $x = 0$ because $m \ge 1$, and is nonpositive at $x = L$, so every partial sum on the left is at most the full sum on the right; this proves the claim when $\mathbf{E}\binom{X}{m} = \infty$. Otherwise monotonicity gives $\binom{L}{m} G(L-1) \le \sum_{y\ge L} \binom{y}{m} f(y) \to 0$, and the claim follows by letting $L \to \infty$. □

**Example 4.2** (Geometric, by the tail formula)**.** For the geometric distribution, $\mathbf{P}(X > x) = q^{x+1}$, and $\sum_{x\ge 0} \binom{x}{m-1} q^x = q^{m-1}/p^m$ is $S_{m-1}/(m-1)!$ from Example 3.2; Theorem 4.1 thus gives immediately

$$\mathbf{E}\binom{X}{m} = q \cdot \frac{q^{m-1}}{p^m} = \Big(\frac{q}{p}\Big)^m,$$

in agreement with that example.

## 5 Difference equations: the Katz family

The binomial, Poisson, geometric, and negative binomial distributions share a structural property: their probability mass function satisfies a first-order difference equation with coefficients linear in $x$,

$$(x+1)\, f(x+1) = (\alpha + \beta x)\, f(x) \qquad \text{for all } x \in \mathbf{N}. \tag{5}$$

The class of distributions defined by (5) is the *Katz family* (Katz 1965), known in the actuarial literature as the Panjer class (Panjer 1981, Sundt & Jewell 1981); it is the case with linear coefficients of Ord's difference-equation analogue (Ord 1967) of the Pearson system of continuous densities, whose logarithmic derivative is a rational function.

For (5) to define a nondegenerate probability distribution one needs $\alpha > 0$ and either $0 \le \beta < 1$, or $\beta < 0$ with $\alpha/(-\beta)$ a positive integer $n$. In the latter case $\alpha + \beta x$ vanishes at $x = n$, both sides of (5) are zero from that point on, and the support is $\{0, \dots, n\}$ (the binomial case of Table 2). Requiring the equation for *all* $x \in \mathbf{N}$ is what makes this work.

The main tool of this section is the following Stein-type identity, whose proof is one shift of the summation index.

Table 2: The Katz family and its factorial moments (Theorem 5.2); $\lambda > 0$, $n \geq 1$, $r > 0$, $0 < p < 1$, and $q = 1 - p$.

| Distribution | $f(x)$ | $\alpha$ | $\beta$ | $\mathbf{E}(X^{\underline{m}})$ |
|---|---|---|---|---|
| Poisson($\lambda$) | $e^{-\lambda}\lambda^x/x!$ | $\lambda$ | $0$ | $\lambda^m$ |
| Binomial($n, p$) | $\binom{n}{x}p^x q^{n-x}$ | $np/q$ | $-p/q$ | $n^{\underline{m}}p^m$ |
| Geometric($p$) | $pq^x$ | $q$ | $q$ | $m!\,(q/p)^m$ |
| Neg. binomial($r, p$) | $\binom{x+r-1}{x}p^r q^x$ | $rq$ | $q$ | $r^{\overline{m}}\,(q/p)^m$ |

**Lemma 5.1.** *Let $X$ have probability mass function satisfying* (5)*, and let $g\colon \mathbf{N} \to \mathbf{R}$ be such that $\mathbf{E}(|(\alpha + \beta X)g(X+1)|) < \infty$. Then*

$$\mathbf{E}(X\,g(X)) = \mathbf{E}((\alpha + \beta X)\,g(X+1)).$$

*Proof.*

$$\sum_{x\geq 1} x\,g(x)f(x) = \sum_{x\geq 0}(x+1)\,g(x+1)f(x+1) = \sum_{x\geq 0}(\alpha+\beta x)\,g(x+1)f(x). \qquad \square$$

Applied to falling powers, Lemma 5.1 computes all factorial moments of the family simultaneously.

**Theorem 5.2.** *Let $X$ have probability mass function satisfying* (5) *with $\beta < 1$. Then, for every $m \geq 0$,*

$$\mathbf{E}(X^{\underline{m}}) = \frac{\prod_{j=0}^{m-1}(\alpha+\beta j)}{(1-\beta)^m}.$$

*Proof.* We choose $g$ so that the left-hand side of Lemma 5.1 becomes the factorial moment we want. For $m \geq 1$, let $g(x) = (x-1)^{\underline{m-1}}$, so that $x\,g(x) = x^{\underline{m}}$ and $g(x+1) = x^{\underline{m-1}}$. Lemma 5.1 yields

$$\mu_{(m)} = \alpha\,\mu_{(m-1)} + \beta\;\mathbf{E}(X\,X^{\underline{m-1}}) = \alpha\,\mu_{(m-1)} + \beta(\mu_{(m)} + (m-1)\mu_{(m-1)}),$$

where the last step uses (3). Solving,

$$\mu_{(m)} = \frac{\alpha + \beta(m-1)}{1-\beta}\,\mu_{(m-1)},$$

and the claim follows by induction from $\mu_{(0)} = 1$. $\square$

Table 2 instantiates Theorem 5.2 on the four classical members of the family. In particular, the binomial case is the formula $\mathbf{E}(X^{\underline{m}}) = n^{\underline{m}}p^m$ underlying (1), and the negative binomial case is the one for which Link (1981) found no way to get rid of the double summation.

Moment recurrences for these distributions go back at least to Guldberg (1931) and Riordan (1937), and the closed forms are recorded in Johnson et al. (2005), both distribution by distribution and, as in Theorem 5.2, for the whole family at once (their equation (2.43); see also Potts 1953).

The index shift in the proof of Lemma 5.1 is summation by parts in disguise. Take $u = g$ and $v(x) = x\,f(x)$ in (4): the boundary term vanishes (as $v(0) = 0$ and, by (5), $g(x)v(x) = (\alpha + \beta(x-1))\,g(x)f(x-1) \to 0$), and $Tv(x) = (x+1)f(x+1) = (\alpha+\beta x)f(x)$ by (5), so that $\Delta v(x) = (\alpha+\beta x - x)f(x)$. Hence $\mathbf{E}((\alpha+\beta X - X)\,g(X)) = -\,\mathbf{E}((\alpha+\beta X)\,\Delta g(X))$, and canceling $\mathbf{E}((\alpha+\beta X)g(X))$ leaves Lemma 5.1. The shift $T$ in (4) is exactly what produces the argument $X+1$.

For the Poisson distribution ($\beta = 0$, $\alpha = \lambda$), Lemma 5.1 is the Chen–Stein characterization $\mathbf{E}(X\,g(X)) = \lambda\;\mathbf{E}(g(X+1))$ at the heart of the Poisson approximation method of Chen (1975), an adaptation of the normal approximation method of Stein (1972). The other members of the family

play the same role: the binomial case underlies binomial approximation (Ehm 1991), following Stein (1986, Lecture VII), and the negative binomial case negative binomial approximation (Brown & Phillips 1999). For the geometric distribution, the same index shift applied to the relation $f(x+1) = q\,f(x)$ gives the simpler identity $\mathbf{E}(g(X)[X \geq 1]) = q\ \mathbf{E}(g(X+1))$, used for geometric approximation (Barbour & Grübel 1995, Peköz 1996).

Lemma 5.1 is itself well known in other guises: the formal summation of Katz (1965, equation (1a)), the stationarity equation of a birth–death process with linear rates (Brown & Xia 2001), the characterizing identity of the birth–death Stein operator, a standardization of the canonical operator of the density approach (Ley et al. 2017, Section 4.4), and, centered, the source of the covariance identities of Kattumannil & Tibiletti (2012).

# 6 Absorption

Many classical distributions have probability mass functions built from binomial coefficients in $x$, and their factorial moments follow from the *absorption identity*: iterating $x\binom{n}{x} = n\binom{n-1}{x-1}$ gives

$$x^{\underline{m}}\binom{n}{x} = n^{\underline{m}}\binom{n-m}{x-m}. \tag{6}$$

A direct verification shows that, for $x \in \mathbf{N}$, (6) holds for *every* integer $m$, negative values included.

For $m \geq 0$, multiplying a probability mass function by $x^{\underline{m}}$ and absorbing produces the (shifted, rescaled) probability mass function of another distribution of the same family, whose total mass is 1. The factorial moment is the constant factor left in front by the absorption. For the binomial distribution, for instance, absorption and the binomial theorem give its row of Table 2 in one line:

$$\mathbf{E}(X^{\underline{m}}) = \sum_x x^{\underline{m}}\binom{n}{x}p^x q^{n-x} = n^{\underline{m}}p^m \sum_x \binom{n-m}{x-m}p^{x-m}q^{n-x} = n^{\underline{m}}p^m.$$

**Example 6.1** (Hypergeometric)**.** Let $\mathbf{P}(X = x) = \binom{K}{x}\binom{N-K}{n-x}/\binom{N}{n}$, where $N$, $K$, and $n$ are integers with $0 \leq K \leq N$ and $0 \leq n \leq N$; the support is $\max(0, n-N+K) \leq x \leq \min(K, n)$. Absorbing $x^{\underline{m}}$ into $\binom{K}{x}$ and applying Vandermonde's convolution $\sum_y \binom{a}{y}\binom{b}{c-y} = \binom{a+b}{c}$ (Graham et al. 1994, equation (5.22)),

$$\mathbf{E}(X^{\underline{m}}) = \frac{K^{\underline{m}}}{\binom{N}{n}} \sum_{y \geq 0} \binom{K-m}{y}\binom{N-K}{(n-m)-y} = K^{\underline{m}}\,\frac{\binom{N-m}{n-m}}{\binom{N}{n}} = \frac{K^{\underline{m}}\,n^{\underline{m}}}{N^{\underline{m}}} \qquad (m \leq N),$$

the last step being (6) again, $n^{\underline{m}}\binom{N}{n} = N^{\underline{m}}\binom{N-m}{n-m}$; or, in binomial-moment form,

$$\mathbf{E}\binom{X}{m} = \frac{\binom{K}{m}\binom{n}{m}}{\binom{N}{m}} \qquad (m \leq N).$$

Multiplied by $c!$, Vandermonde's convolution takes the falling-power form $(a+b)^{\underline{c}} = \sum_y \binom{c}{y} a^{\underline{y}}\, b^{\underline{c-y}}$, which we use below.

**Example 6.2** (Beta-binomial)**.** Let $\mathbf{P}(X = x) = \binom{n}{x}B(\alpha + x, \beta + n - x)/B(\alpha, \beta)$ on $\{0, \ldots, n\}$, where $n \in \mathbf{N}$, $\alpha, \beta > 0$, and $B$ is the beta function. Absorption gives, for $m \leq n$,

$$\mathbf{E}(X^{\underline{m}}) = \frac{n^{\underline{m}}}{B(\alpha, \beta)} \sum_{y \geq 0} \binom{n-m}{y} B(\alpha + m + y, \beta + (n-m) - y),$$

and the sum is $B(\alpha+m,\beta)$ times the total mass of the beta-binomial distribution with parameters $n-m$, $\alpha+m$, $\beta$, that is, $B(\alpha+m,\beta)$. Hence

$$\mathbf{E}(X^{\underline{m}}) = n^{\underline{m}}\,\frac{B(\alpha+m,\beta)}{B(\alpha,\beta)} = n^{\underline{m}}\,\frac{\alpha^{\overline{m}}}{(\alpha+\beta)^{\overline{m}}}.$$

For integer parameters this is the negative hypergeometric distribution, and the normalization identity used above is Vandermonde's convolution with the negative upper indices $-\alpha$ and $-\beta$, so the computation stays entirely within finite sums. This settles a question of Link (1981), who was unable to treat the negative hypergeometric distribution by finite-difference operators and asked whether it could be done. His version counts draws rather than failures, that is, $X+c$ for a constant $c$, whose factorial moments follow from those of $X$ by the falling-power form of Vandermonde's convolution.

**Example 6.3** (Inverse moments)**.** Negative falling powers make the same machinery applicable to inverse moments, which are usually computed by integrating the probability generating function, as in the classical treatment of Chao & Strawderman (1972). They remark that "negative factorial moments" are more tractable than negative power moments, "a fact we will not prove here." Negative falling powers prove it. Since $X^{\underline{-1}} = 1/(X+1)$, identity (6) with $m=-1$, that is, $\binom{n}{x}/(x+1) = \binom{n+1}{x+1}/(n+1)$, gives for the binomial distribution

$$\mathbf{E}\left(\frac{1}{X+1}\right) = \frac{1}{(n+1)p}\sum_{y\ge 1}\binom{n+1}{y}p^y q^{(n+1)-y} = \frac{1-q^{n+1}}{(n+1)p},$$

and in general, for every $m \ge 1$,

$$\mathbf{E}(X^{\underline{-m}}) = \frac{1-\sum_{j<m}\binom{n+m}{j}p^j q^{n+m-j}}{(n+1)^{\overline{m}}\,p^m};$$

the missing terms are exactly the initial segment of the binomial distribution with parameters $n+m$ and $p$ that the shift $y=x+m$ cannot reach.

The Poisson analogue, obtained by absorbing $\lambda^x/x!\cdot x^{\underline{-m}} = \lambda^{-m}\lambda^{x+m}/(x+m)!$, is

$$\mathbf{E}(X^{\underline{-m}}) = \frac{1-e^{-\lambda}\sum_{j<m}\lambda^j/j!}{\lambda^m}, \quad \text{in particular} \quad \mathbf{E}\left(\frac{1}{X+1}\right) = \frac{1-e^{-\lambda}}{\lambda}.$$

# 7 Beyond the Katz family

All distributions treated so far have factorial moments with known elementary closed forms. The index shift of Section 5, however, uses only the fact that the probability mass function satisfies a difference equation $a(x)f(x+1) = b(x)f(x)$, and it remains useful when the normalizing constant is not an elementary function of the parameters. Closed forms for the moments are then not to be expected, and the results to aim for are exact identities and linear recurrences between moments.

The main tool is the following generalization of Lemma 5.1.

**Lemma 7.1** (Generalized Stein identity)**.** *Let $X$ have probability mass function $f$ satisfying $a(x)f(x+1) = b(x)f(x)$ for all $x \in \mathbf{N}$. Then, for every $g$ such that the expectations below converge absolutely,*

$$\mathbf{E}\big(a(X-1)\,g(X)\,[X\ge 1]\big) = \mathbf{E}\big(b(X)\,g(X+1)\big).$$

*Proof.* Shift the summation index, as in Lemma 5.1, which is the case $a(x)=x+1$, $b(x)=\alpha+\beta x$; there $a(x-1)=x$ vanishes at zero and makes the bracket redundant. □

The following subsections use Lemma 7.1 along the lines of the guide at the end of Section 2: we write the difference equation satisfied by the probability mass function, take $g = 1$ to obtain an exact identity, and, when the coefficients are polynomials, take $g(x) = (x-1)^{\underline{m}}$ to obtain a recurrence for the factorial moments. Identities of this kind are scattered through the literature; most notably, Daly & Gaunt (2016) develop Stein characterizations of the Conway–Maxwell–Poisson and Conway–Maxwell–binomial distributions and derive moment identities from them.

When the ratio $f(x+1)/f(x)$ is a rational function of $x$, the normalizing constant is a value of a classical special function (a generalized hypergeometric function). These functions satisfy contiguity relations, that is, linear relations connecting their values at parameters that differ by integers. Each recurrence below is such a relation, because absorbing $x^{\underline{m}}$ into $f$ writes $\mu_{(m)}$ as a value of the same function, at integer-shifted parameters, divided by the normalizing constant.

All distributions of this section with rational ratio $f(x+1)/f(x)$ belong to Kemp's family of generalized hypergeometric probability distributions, whose probability generating functions have the form $F(\lambda z)/F(\lambda)$ for a generalized hypergeometric function $F$ (Johnson et al. 2005, Section 2.4.1); for the Conway–Maxwell families defined below, this requires their parameter $\nu$ to be an integer. For the members of this family, Kemp's differential-equation method yields fixed-parameter factorial-moment recurrences in principle (Kemp 1968). Except where noted, however, the recurrences below seem never to have been written out, several of the distributions having entered the literature only after Kemp's paper. The index shift obtains all of them in a few lines, without using special functions or differentiation.

### 7.1 The Conway–Maxwell–Poisson distribution

The Conway–Maxwell–Poisson distribution (Conway & Maxwell 1962, Shmueli et al. 2005), a standard model for over- and underdispersed count data, has probability mass function

$$f(x) = \frac{1}{Z(\lambda,\nu)}\,\frac{\lambda^x}{(x!)^\nu}, \qquad Z(\lambda,\nu) = \sum_{x\ge 0}\frac{\lambda^x}{(x!)^\nu},$$

where $\lambda > 0$ and $\nu \ge 0$, with $\lambda < 1$ if $\nu = 0$ so that the series converges. For $\nu$ other than 0 and 1 no elementary closed form for $Z(\lambda,\nu)$ is known, nor is one known for the mean, for which asymptotic approximations are used in practice.

The probability mass function satisfies $(x+1)^\nu f(x+1) = \lambda f(x)$, so for $\nu > 0$ Lemma 7.1 with $a(x-1) = x^\nu$ (which vanishes at zero) gives

$$\mathbf{E}\big(X^\nu g(X)\big) = \lambda\,\mathbf{E}\big(g(X+1)\big), \tag{7}$$

the Stein characterization exploited by Daly & Gaunt (2016). Even without a closed form for $Z(\lambda,\nu)$, three choices of $g$ yield exact results:

- $g = 1$ gives $\mathbf{E}(X^\nu) = \lambda$, for every real $\nu > 0$, an identity of Sellers & Shmueli (2010) generalized by Daly & Gaunt (2016);
- $g(x) = [x \ge 1]\,x^{-\nu}$ gives $\mathbf{E}\big((X+1)^{-\nu}\big) = (1 - f(0))/\lambda$, generalizing the Poisson case of Example 6.3; in series form, it appears in Daly & Gaunt (2016, proof of Lemma 3.7);
- for integer $\nu$, falling powers give recurrences. Let $\nu = 2$ and $g(x) = (x-1)^{\underline{m}}$; since $x^2\,(x-1)^{\underline{m}} = x\,x^{\underline{m+1}} = x^{\underline{m+2}} + (m+1)x^{\underline{m+1}}$, identity (7) becomes

$$\mu_{(m+2)} = \lambda\,\mu_{(m)} - (m+1)\,\mu_{(m+1)}.$$

The recurrence determines every factorial moment of the Conway–Maxwell–Poisson distribution with $\nu = 2$ as a polynomial in $\lambda$ and the mean $\mu = \mathbf{E}(X)$: for instance, $m = 0$ gives $\mathbf{E}\big(X(X-1)\big) =$

$\lambda - \mu$, and $\operatorname{var}(X) = \lambda - \mu^2$ follows from $\mathbf{E}(X^2) = \lambda$. For $\nu = 2$ the factorial moments also admit closed forms in terms of modified Bessel functions, $\mathbf{E}(X^{\underline{m}}) = \lambda^{m/2} I_m(2\sqrt{\lambda})/I_0(2\sqrt{\lambda})$ (Yuan & Kalbfleisch 2000, Section 2; Daly & Gaunt 2016, Section 2.4), and the recurrence above is precisely the three-term contiguity relation of the functions $I_m$, obtained without mentioning them.

For the Conway–Maxwell–Poisson distribution with larger integer $\nu$, iterating (3) expands $x^{\nu-1}x^{\underline{m+1}}$ into falling powers and yields a $(\nu+1)$-term recurrence; for $\nu = 3$,

$$\mu_{(m+3)} + (2m+3)\,\mu_{(m+2)} + (m+1)^2 \mu_{(m+1)} = \lambda\,\mu_{(m)}.$$

The factorial moments $\mu_{(1)}, \dots, \mu_{(\nu-1)}$ have no known elementary expression, but all the following ones are determined by them exactly. Nonelementary closed forms for the raw moments, in terms of hyper-Bessel functions, have been given recently by Popov (2026, Theorem 8.3).

Lemma 7.1 applies in the same way to the Bessel distribution of Yuan & Kalbfleisch (2000), whose parameters we write as $\theta > 0$ and $\kappa > -1$ ($\theta = a^2/4$ and $\kappa = \nu$ in their notation). Its probability mass function is proportional to $\theta^x/(x!\,\Gamma(x+\kappa+1))$; for $\kappa = 0$ this is the Conway–Maxwell–Poisson distribution with $\nu = 2$ and $\lambda = \theta$. The probability mass function satisfies $(x+1)(x+\kappa+1)f(x+1) = \theta f(x)$, so that $g = 1$ gives the exact identity $\mathbf{E}\big(X(X+\kappa)\big) = \theta$, again free of Bessel functions. The identity also follows, through the three-term contiguity relation of the modified Bessel functions, from the expressions for $\mathbf{E}(X)$ and $\mathbf{E}(X^2)$ given by Yuan & Kalbfleisch (2000, equation (2.1)).

### 7.2 Fisher's noncentral hypergeometric distribution

Let $X_1$, $X_2$ be independent binomial random variables with $K$ and $N-K$ trials and success probabilities $p_1, p_2 \in (0,1)$; the conditional law of $X_1$ given $X_1 + X_2 = n$ is Fisher's noncentral hypergeometric distribution, also called the extended hypergeometric distribution (Harkness 1965). It has probability mass function

$$f(x) = \frac{1}{C}\binom{K}{x}\binom{N-K}{n-x}\omega^x,$$

where $\omega = (p_1/(1-p_1))/(p_2/(1-p_2))$ is the odds ratio and $C$ is the normalizing constant; $N$, $K$, $n$, and the support are as in Example 6.1.

The distribution is the central object of conditional inference on odds ratios and of biased urn models. Neither $C$ (a Gauss hypergeometric polynomial) nor any moment has a closed form uniform in the parameters: in practice the mean and variance are obtained by numerical evaluation of the distribution (Liao & Rosen 2001, Fog 2008), or by approximations.

Writing $M = N - K - n$, the probability mass function satisfies

$$(x+1)(M+x+1)\,f(x+1) = \omega\,(K-x)(n-x)\,f(x)$$

for all $x \in \mathbf{N}$. Indeed, the factor $(K-x)(n-x)$ vanishes at the upper end of the support; when the lower end is positive (that is, when $M < 0$), the factor $M + x + 1$ vanishes just below it, so the relation extends beyond the support with both sides zero. Since $a(x-1) = x(M+x)$ vanishes at zero, Lemma 7.1 gives

$$\mathbf{E}\big(X(M+X)\,g(X)\big) = \omega\;\mathbf{E}\big((K-X)(n-X)\,g(X+1)\big).$$

For $g = 1$ this is the exact version of the classical mean approximation, which solves $\mu(M+\mu) = \omega(K-\mu)(n-\mu)$ for the mean $\mu$ as if the expectation factored.

Now take $g(x) = (x-1)^{\underline{m}}$. On the left,

$$x(M+x)\,(x-1)^{\underline{m}} = (M+x)\,x^{\underline{m+1}} = x^{\underline{m+2}} + (M+m+1)\,x^{\underline{m+1}};$$

on the right, writing $(K-x)(n-x) = (x-K)(x-n)$ and applying (3) with $c=-n$, then with $c=-K$ to both resulting terms,

$$(K-x)(n-x)\,x^{\underline{m}} = (K-m)(n-m)\,x^{\underline{m}} - (K+n-2m-1)\,x^{\underline{m+1}} + x^{\underline{m+2}}.$$

Taking expectations and collecting the terms in $\mu_{(m+2)}$ yields

$$(1-\omega)\mu_{(m+2)} = \omega(K-m)(n-m)\mu_{(m)} - \big(M+m+1+\omega(K+n-2m-1)\big)\mu_{(m+1)}. \quad (8)$$

Recurrence (8) itself is known: Kemp (1968) worked out the present distribution as an example of her method, and the two-line derivation above independently confirms her result.

Harkness (1965) used a contiguity relation of the Gauss hypergeometric function to express the variance through the mean (the case $m=0$ of (8)), remarking that "recurrence relations for higher moments may be obtained similarly, but appear to be of little value." We think otherwise: for $\omega \neq 1$, (8) makes every factorial moment an elementary function of the parameters and the mean $\mu$. The observation is immediate from the recurrence, but we have not found it stated anywhere. Example 7.2, in Section 7.3, applies it to a concrete case. (Dividing by $1-\omega$ makes the recurrence numerically ill-conditioned near $\omega = 1$, but its symbolic content is unaffected.)

At $\omega = 1$ the left-hand side vanishes and the recurrence degenerates to $(N-m)\mu_{(m+1)} = (K-m)(n-m)\,\mu_{(m)}$, recovering the factorial moments of the central hypergeometric distribution obtained by absorption in Section 6.

### 7.3 The Conway–Maxwell–binomial distribution

The distribution with probability mass function proportional to $\binom{n}{x}^{\nu} p^x q^{n-x}$ on $\{0,\dots,n\}$ ($n \in \mathbf{N}$, $0<p<1$, $q=1-p$, $\nu>0$) was introduced by Shmueli et al. (2005), and studied further by Borges et al. (2014) (see also the erratum, Borges et al. 2023) and, as a model for sums of possibly dependent Bernoulli variables, by Kadane (2016). Again, the normalizing constant and the moments have no closed form uniform in the parameters. The probability mass function satisfies $(x+1)^{\nu} f(x+1) = (p/q)(n-x)^{\nu} f(x)$ for all $x \in \mathbf{N}$, reading $(n-x)^{\nu}$ as 0 for $x > n$, so Lemma 7.1 gives

$$\mathbf{E}\big(X^{\nu} g(X)\big) = \frac{p}{q}\,\mathbf{E}\big((n-X)^{\nu} g(X+1)\big),$$

the Stein identity of Daly & Gaunt (2016, Lemma 2.9), whose $g=1$ case is the exact identity

$$\mathbf{E}(X^{\nu}) = \frac{p}{q}\,\mathbf{E}\big((n-X)^{\nu}\big)$$

for every real $\nu > 0$.

Moreover (and this observation appears to be new), since

$$\binom{n}{x}^2 \omega^x = \binom{n}{x}\binom{n}{n-x}\omega^x,$$

the case $\nu = 2$ *is* Fisher's noncentral hypergeometric distribution with parameters $N=2n$, $K=n$, sample size $n$, and odds ratio $\omega = p/q$. Recurrence (8) therefore applies verbatim and, for $p \neq 1/2$, determines all its factorial moments from the mean.

**Example 7.2** (Exact moments from the mean)**.** For the Conway–Maxwell–binomial distribution with $n=4$, $p=1/3$, and $\nu=2$, the identification above gives $\omega = 1/2$ and $M=0$. The normalizing constant $C$ is the finite sum $\sum_x \binom{4}{x}^2 2^{-x} = 321/16$, and a similar sum gives the mean $\mu = 172/107$. Recurrence (8) then yields, with no further summation, $\mu_{(2)} = 164/107$, $\mu_{(3)} = 72/107$, and $\mu_{(4)} = 8/107$, exactly.

### 7.4 The Conway–Maxwell–negative binomial distribution

The same index shift applies to the Conway–Maxwell–negative binomial distribution of Zhang et al. (2018), with probability mass function proportional to $\binom{x+r-1}{x}^{\nu} q^x$ ($r, \nu > 0$, $0 < q < 1$). A closely related generalization appears in the work of Chakraborty & Ong (2016). The probability mass function satisfies $(x+1)^{\nu} f(x+1) = q\,(x+r)^{\nu} f(x)$, so Lemma 7.1 gives the Stein characterization of Zhang et al. (2018)

$$\mathbf{E}(X^{\nu} g(X)) = q\ \mathbf{E}((X+r)^{\nu} g(X+1)).$$

For $g = 1$ it gives the identity $\mathbf{E}(X^{\nu}) = q\ \mathbf{E}((X+r)^{\nu})$ for every real $\nu > 0$; at $\nu = 1$ this reduces to $\mathbf{E}(X) = qr/(1-q)$, the mean of the negative binomial distribution. For $\nu = 2$, take $g(x) = (x-1)^{\underline{m}}$. The left-hand side expands as in Section 7.1, and applying (3) twice with $c = r$ gives $(x+r)^2 x^{\underline{m}} = x^{\underline{m+2}} + (2r+2m+1)x^{\underline{m+1}} + (r+m)^2 x^{\underline{m}}$; hence

$$(1-q)\,\mu_{(m+2)} = q(r+m)^2 \mu_{(m)} + \big(q(2r+2m+1) - (m+1)\big)\mu_{(m+1)}.$$

Since $1 - q > 0$, this three-term recurrence determines every factorial moment as an elementary function of $q$, $r$, and the mean, as (8) does for the noncentral hypergeometric distribution.

## 8 From factorial to raw moments

Raw moments now follow immediately. The change of basis between ordinary and falling powers,

$$x^k = \sum_{j=0}^{k} \left\{ {k \atop j} \right\} x^{\underline{j}},$$

is one of the defining properties (Graham et al. 1994) of the Stirling numbers of the second kind $\left\{ {k \atop j} \right\}$. They can be computed by the recurrence $\left\{ {k+1 \atop j} \right\} = j \left\{ {k \atop j} \right\} + \left\{ {k \atop j-1} \right\}$, with $\left\{ {0 \atop 0} \right\} = 1$ and $\left\{ {k \atop 0} \right\} = \left\{ {0 \atop j} \right\} = 0$ for $k, j > 0$. Since expectation is linear, we obtain:

**Theorem 8.1.** *For every random variable $X$ with values in $\mathbf{N}$ and every $k \geq 0$,*

$$\mathbf{E}(X^k) = \sum_{j=0}^{k} \left\{ {k \atop j} \right\} \mathbf{E}(X^{\underline{j}}),$$

*both sides being possibly infinite. In particular, for the Katz family* (5)*,*

$$\mathbf{E}(X^k) = \sum_{j=0}^{k} \left\{ {k \atop j} \right\} \frac{\prod_{i=0}^{j-1} (\alpha + \beta i)}{(1-\beta)^j}.$$

For the first moments, $\left\{ {2 \atop 1} \right\} = \left\{ {2 \atop 2} \right\} = \left\{ {3 \atop 1} \right\} = \left\{ {3 \atop 3} \right\} = 1$ and $\left\{ {3 \atop 2} \right\} = 3$ give, writing $\mu = \mu_{(1)} = \mathbf{E}(X)$,

$$\mathbf{E}(X^2) = \mu_{(2)} + \mu, \qquad \operatorname{var}(X) = \mu_{(2)} + \mu - \mu^2, \qquad \mathbf{E}(X^3) = \mu_{(3)} + 3\mu_{(2)} + \mu.$$

For the binomial distribution, Theorem 8.1 is exactly (1). For the Poisson distribution it is the classical evaluation $\mathbf{E}(X^k) = \sum_j \left\{ {k \atop j} \right\} \lambda^j$, the Touchard polynomial of $\lambda$, which at $\lambda = 1$ returns Dobiński's formula for the Bell numbers.

### 8.1 Central moments

Central moments can of course be expanded binomially over raw moments, but they are genuinely less natural objects in the discrete world. The falling powers of $x - \mu$ do not interact with the probability mass functions as neatly as those of $x$, which is one way to understand why "handy"

closed forms for the central moments of the binomial distribution require the finer analysis of Knoblauch (2008) and Skorski (2025).

Central moments can nonetheless be obtained from the *factorial moments about the mean* exactly as raw moments are obtained from factorial moments: the change of basis applied to $x-\mu$ gives $\mathbf{E}((X-\mu)^k) = \sum_j \left\{ {k \atop j} \right\} \mathbf{E}((X-\mu)^{\underline{j}})$. For the Poisson distribution, the Stein identity produces a recurrence for these factorial moments about the mean that mirrors the recurrence $\mathbf{E}((Y-\mu)^m) = (m-1)\sigma^2\,\mathbf{E}((Y-\mu)^{m-2})$ for a Gaussian variable $Y$ with mean $\mu$ and variance $\sigma^2$. The price is a center that drifts by one.

**Proposition 8.2.** *Let $X$ be Poisson with parameter $\lambda$ and, for $c \in \mathbf{R}$, let $\tau_m(c) = \mathbf{E}((X-c)^{\underline{m}})$. Then, for $m \geq 2$,*

$$\tau_m(\lambda) = (m-1)\,\lambda\,\tau_{m-2}(\lambda+1).$$

*Proof.* Let $g(x) = (x-1-\lambda)^{\underline{m-1}}$, so that $(x-\lambda)g(x) = (x-\lambda)^{\underline{m}}$ and $g(x+1)-g(x) = \Delta g(x) = (m-1)(x-1-\lambda)^{\underline{m-2}}$ by the power rule (2), which is insensitive to the shift by $\lambda$. By Lemma 5.1 in its Poisson form $\mathbf{E}(X\,g(X)) = \lambda\;\mathbf{E}(g(X+1))$,

$$\begin{aligned}\tau_m(\lambda) &= \mathbf{E}(X\,g(X)) - \lambda\,\mathbf{E}(g(X)) = \lambda\;\mathbf{E}(g(X+1)-g(X)) \\ &= (m-1)\,\lambda\;\mathbf{E}((X-(\lambda+1))^{\underline{m-2}}).\end{aligned}$$

□

The proposition gives, for instance, $\tau_2(\lambda) = \lambda$, $\tau_3(\lambda) = -2\lambda$, and $\tau_4(\lambda) = 3\lambda(\lambda+2)$. The off-center values it requires, such as $\tau_2(\lambda+1) = \lambda+2$, follow from those at $\lambda$ by the falling-power form of Vandermonde's convolution (Example 6.1) with $a = X-\lambda$ and $b=-1$, that is, $\tau_k(\lambda+1) = \sum_j \binom{k}{j}(-1)^{k-j}(k-j)!\,\tau_j(\lambda)$. Since $\tau_1(\lambda) = 0$, the change of basis, with the Stirling numbers $\left\{ {3 \atop 2} \right\} = 3$, $\left\{ {4 \atop 2} \right\} = 7$, and $\left\{ {4 \atop 3} \right\} = 6$, turns these values into the central moments $\mathbf{E}((X-\lambda)^3) = 3\tau_2(\lambda) + \tau_3(\lambda) = \lambda$ and $\mathbf{E}((X-\lambda)^4) = 7\tau_2(\lambda) + 6\tau_3(\lambda) + \tau_4(\lambda) = \lambda(1+3\lambda)$. The leading term $3\lambda^2$ of $\tau_4(\lambda)$ is the Gaussian $3\sigma^4$, as expected from the central limit theorem; the drift of the center is a purely discrete phenomenon, invisible in the continuous limit.

# 9 Conclusions

None of the distributions we treated required more than a few lines, and the lines were always of one of three kinds: a summation by parts, an absorption, a change of basis.

Every computation above translates, term by term, an argument of the continuous calculus. The difference $\Delta$ stands for the derivative, the falling power $x^{\underline{m}}$ for the monomial $x^m$, the geometric distribution for the exponential, and summation by parts (4) for integration by parts. The Katz difference equation (5) stands for the Pearson system, the Stein-type identity of Lemma 5.1 for the Stein identities of the Gaussian and gamma laws, and the absorption identity (6) for absorption into the gamma and beta integrals. Finally, the drifting-center recurrence of Proposition 8.2 stands for the Gaussian recurrence $\mathbf{E}((Y-\mu)^m) = (m-1)\sigma^2\,\mathbf{E}((Y-\mu)^{m-2})$. The Stirling change of basis is the only step whose continuous counterpart is trivial.

The operator methods of Link, of Boullion, Seaman and Young, and of Kemp work on the derivative side of the calculus; the method developed here works on the integral side. This choice has pedagogical value: factorial moments become computations one can do at the blackboard, rather than exercises in differentiating generating functions.

The choice has practical value as well. The method handles the two cases that resisted Link's operators, one of which he left as an open problem. It also keeps producing exact identities and recurrences for distributions whose normalizing constants are not elementary, since difference equations with coefficients of any degree are amenable to the same index-shift argument (Section 7). Negative falling powers, moreover, give inverse moments uniformly (Example 6.3).

We thus hope to have shown that finite calculus, with summation by parts as its main device, is the natural tool for the computation of the moments of discrete distributions.

## References


Barbour, A. D. & Grübel, R. (1995), 'The first divisible sum', *Journal of Theoretical Probability* **8**(1), 39–47. DOI: 10.1007/BF02213453.

Borges, P., Rodrigues, J., Balakrishnan, N. & Bazán, J. (2014), 'A COM–Poisson type generalization of the binomial distribution and its properties and applications', *Statistics & Probability Letters* **87**, 158–166. DOI: 10.1016/j.spl.2014.01.019.

Borges, P., Rodrigues, J., Balakrishnan, N. & Bazán, J. (2023), 'Erratum to "A COM–Poisson type generalization of the binomial distribution and its properties and applications" [Statist. Probab. Lett. 87 (2014) 158–166]', *Statistics & Probability Letters* **201**, 109883. DOI: 10.1016/j.spl.2023.109883.

Boullion, T. L., Seaman, Jr., J. W. & Young, D. M. (1992), 'Moments of discrete probability distributions derived using a differential operator', *The American Statistician* **46**(1), 22–24. DOI: 10.1080/00031305.1992.10475840.

Brown, T. C. & Phillips, M. J. (1999), 'Negative binomial approximation with Stein's method', *Methodology and Computing in Applied Probability* **1**(4), 407–421. DOI: 10.1023/A:1010094221471.

Brown, T. C. & Xia, A. (2001), 'Stein's method and birth–death processes', *The Annals of Probability* **29**(3), 1373–1403. DOI: 10.1214/aop/1015345606.

Chakraborty, S. & Ong, S. H. (2016), 'A COM–Poisson-type generalization of the negative binomial distribution', *Communications in Statistics—Theory and Methods* **45**(14), 4117–4135. DOI: 10.1080/03610926.2014.917184.

Chao, M.-T. & Strawderman, W. E. (1972), 'Negative moments of positive random variables', *Journal of the American Statistical Association* **67**(338), 429–431. DOI: 10.1080/01621459.1972.10482404.

Charalambides, C. A. (2005), *Combinatorial Methods in Discrete Distributions*, Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ. DOI: 10.1002/0471733180.

Chen, L. H. Y. (1975), 'Poisson approximation for dependent trials', *The Annals of Probability* **3**(3), 534–545. DOI: 10.1214/aop/1176996359.

Conway, R. W. & Maxwell, W. L. (1962), 'A queuing model with state dependent service rates', *Journal of Industrial Engineering* **12**, 132–136.

Daly, F. & Gaunt, R. E. (2016), 'The Conway–Maxwell–Poisson distribution: Distributional theory and approximation', *ALEA. Latin American Journal of Probability and Mathematical Statistics* **13**(2), 635–658. DOI: 10.30757/ALEA.v13-25.

Di Nardo, E. & Senato, D. (2006), 'An umbral setting for cumulants and factorial moments', *European Journal of Combinatorics* **27**(3), 394–413. DOI: 10.1016/j.ejc.2004.12.001.

Ebner, B., Fischer, A., Gaunt, R. E., Picker, B. & Swan, Y. (2025), 'Stein's method of moments', *Scandinavian Journal of Statistics* **52**(4), 1594–1624. DOI: 10.1111/sjos.70003.

Ehm, W. (1991), 'Binomial approximation to the Poisson binomial distribution', *Statistics & Probability Letters* **11**(1), 7–16. DOI: 10.1016/0167-7152(91)90170-V.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Vol. I, third edn, John Wiley & Sons, New York.

Fog, A. (2008), 'Sampling methods for Wallenius' and Fisher's noncentral hypergeometric distributions', *Communications in Statistics—Simulation and Computation* **37**(2), 241–257. DOI: 10.1080/03610910701790236.

Gaunt, R. E. & Ouimet, F. (2026), 'On Wilks' problem: Exact recursive formulas via Stein's method for the joint moments of disjoint principal minors of Wishart random matrices', arXiv preprint arXiv:2607.03984.

Graham, R. L., Knuth, D. E. & Patashnik, O. (1994), *Concrete Mathematics: A Foundation for Computer Science*, second edn, Addison-Wesley, Reading, MA.

Guldberg, A. (1931), 'On discontinuous frequency-functions and statistical series', *Scandinavian Actuarial Journal* **1931**(3), 167–187. DOI: 10.1080/03461238.1931.10405860.

Harkness, W. L. (1965), 'Properties of the extended hypergeometric distribution', *The Annals of Mathematical Statistics* **36**(3), 938–945. DOI: 10.1214/aoms/1177700066.

Joarder, A. H. & Mahmood, M. (1997), 'Classroom note: An inductive derivation of Stirling numbers of the second kind and their applications in statistics', *Journal of Applied Mathematics & Decision Sciences* **1**(2), 151–157. DOI: 10.1155/S1173912697000138.

Johnson, N. L., Kemp, A. W. & Kotz, S. (2005), *Univariate Discrete Distributions*, third edn, John Wiley & Sons, Hoboken, NJ. DOI: 10.1002/0471715816.

Jordan, C. (1950), *Calculus of Finite Differences*, second edn, Chelsea Publishing Company, New York.

Kadane, J. B. (2016), 'Sums of possibly associated Bernoulli variables: The Conway–Maxwell-Binomial distribution', *Bayesian Analysis* **11**(2), 403–420. DOI: 10.1214/15-BA955.

Kattumannil, S. K. & Tibiletti, L. (2012), 'Moment identity for discrete random variable and its applications', *Statistics* **46**(6), 767–775. DOI: 10.1080/02331888.2011.555548.

Katz, L. (1965), Unified treatment of a broad class of discrete probability distributions, *in* G. P. Patil, ed., 'Classical and Contagious Discrete Distributions', Statistical Publishing Society, Calcutta, pp. 175–182.

Kemp, A. W. (1968), 'A wide class of discrete distributions and the associated differential equations', *Sankhyā, Series A* **30**(4), 401–410.

Knoblauch, A. (2008), 'Closed-form expressions for the moments of the binomial probability distribution', *SIAM Journal on Applied Mathematics* **69**(1), 197–204. DOI: 10.1137/070700024.

Ley, C., Reinert, G. & Swan, Y. (2017), 'Stein's method for comparison of univariate distributions', *Probability Surveys* **14**, 1–52. DOI: 10.1214/16-PS278.

Liao, J. G. & Rosen, O. (2001), 'Fast and stable algorithms for computing and sampling from the noncentral hypergeometric distribution', *The American Statistician* **55**(4), 366–369. DOI: 10.1198/000313001753272547.

Link, R. F. (1981), 'Moments of discrete probability distributions derived using finite difference operators', *The American Statistician* **35**(1), 44–46. DOI: 10.1080/00031305.1981.10479303.

Ord, J. K. (1967), 'On a system of discrete distributions', *Biometrika* **54**(3–4), 649–656. DOI: 10.1093/biomet/54.3-4.649.

Ottestad, P. (1939), 'On the use of the factorial moments in the study of discontinuous frequency distributions', *Scandinavian Actuarial Journal* **1939**(1), 22–31. DOI: 10.1080/03461238.1939.10419248.

Panjer, H. H. (1981), 'Recursive evaluation of a family of compound distributions', *ASTIN Bulletin* **12**(1), 22–26. DOI: 10.1017/S0515036100006796.

Peköz, E. A. (1996), 'Stein's method for geometric approximation', *Journal of Applied Probability* **33**(3), 707–713. DOI: 10.2307/3215352.

Popov, V. (2026), '$m$-Bell and $m$-Stirling numbers: Iterated binomial transforms, hyper-Bessel functions, and moments of the Conway–Maxwell–Poisson distribution', arXiv preprint arXiv:2608.12011.

Potts, R. B. (1953), 'Note on the factorial moments of standard distributions', *Australian Journal of Physics* **6**(4), 498–499. DOI: 10.1071/PH530498.

Riordan, J. (1937), 'Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions', *The Annals of Mathematical Statistics* **8**(2), 103–111. DOI: 10.1214/aoms/1177732430.

Sellers, K. F. & Shmueli, G. (2010), 'A flexible regression model for count data', *The Annals of Applied Statistics* **4**(2), 943–961. DOI: 10.1214/09-AOAS306.

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. DOI: 10.1111/j.1467-9876.2005.00474.x.

Skorski, M. (2025), 'Handy formulas for binomial moments', *Modern Stochastics: Theory and Applications* **12**(1), 27–41. DOI: 10.15559/24-VMSTA260.

Steffensen, J. F. (1923), 'Factorial moments and discontinuous frequency-functions', *Scandinavian Actuarial Journal* **1923**(2), 73–89. DOI: 10.1080/03461238.1923.10410768.

Stein, C. (1972), A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *in* 'Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability', Vol. 2, University of California Press, Berkeley, CA, pp. 583–602. DOI: 10.1525/9780520423671-036.

Stein, C. (1986), *Approximate Computation of Expectations*, Vol. 7 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*, Institute of Mathematical Statistics, Hayward, CA. DOI: 10.1214/lnms/1215466568.

Sundt, B. & Jewell, W. S. (1981), 'Further results on recursive evaluation of compound distributions', *ASTIN Bulletin* **12**(1), 27–39. DOI: 10.1017/S0515036100006802.

Tripathi, R. C. & Gurland, J. (1977), 'A general family of discrete distributions with hypergeometric probabilities', *Journal of the Royal Statistical Society, Series B (Methodological)* **39**(3), 349–356. DOI: 10.1111/j.2517-6161.1977.tb01634.x.

Upadhye, N. S., Čekanavičius, V. & Vellaisamy, P. (2017), 'On Stein operators for discrete approximations', *Bernoulli* **23**(4A), 2828–2859. DOI: 10.3150/16-BEJ829.

Yuan, L. & Kalbfleisch, J. D. (2000), 'On the Bessel distribution and related problems', *Annals of the Institute of Statistical Mathematics* **52**(3), 438–447. DOI: 10.1023/A:1004152916478.

Zhang, H., Tan, K. & Li, B. (2018), 'COM-negative binomial distribution: Modeling overdispersion and ultrahigh zero-inflated count data', *Frontiers of Mathematics in China* **13**(4), 967–998. DOI: 10.1007/s11464-018-0714-z.