SOCIETY OF ACTUARIES

Article from:

ARCH 2013.1 Proceedings

August 1- 4, 2012

James G. Bridgeman

Combinatorics for Moments of a Randomly Stopped Quadratic Variation Process

James G. Bridgeman, FSA, CERA
University of Connecticut 196 Auditorium Rd.U-3009 Storrs CT 06269-3009
bridgeman@math.uconn.edu

August 4, 2012

Abstract

A random process that includes jumps will in general have a quadratic variation that itself forms a non-trivial random process. One might be interested in moments of the quadratic variation process, for example in order to characterize it or in order to approximate it with a known process. The paper proposes a combinatoric approach to express higher moments of the quadratic variation process in terms of higher order variations of the original process and higher order autocovariations of the variations of the original process. These have lent themselves to direct calculation by Laplace transforms in the examples that gave rise to this work.

Suppose x_1 and x_2 are random variables and we want to calculate $\mathbb{E}\left[(x_1+x_2)^2\right]$. One way to proceed might be

$$\begin{aligned}\mathbb{E}\left[(x_1+x_2)^2\right] &= \mathbb{E}\left[\left(x_1^2+x_2^2\right)\right]+2\mathbb{E}\left[x_1x_2\right]\\ &= \left(\mathbb{E}\left[x_1^2\right]+\mathbb{E}\left[x_2^2\right]\right)+2\rho\left(\mathbb{E}\left[x_1\right]\mathbb{E}\left[x_2\right]\right)\end{aligned}$$

where ρ, defined as satisfying $\mathbb{E}\left[x_1x_2\right]=\rho\mathbb{E}\left[x_1\right]\mathbb{E}\left[x_2\right]$, can be called a covariation coefficient and

$$\mathbb{E}\left[(x_1+x_2)^2\right]=\left(\mathbb{E}\left[x_1^2\right]+\mathbb{E}\left[x_2^2\right]\right)+2\rho\left\{\frac{\left(\mathbb{E}\left[x_1\right]+\mathbb{E}\left[x_2\right]\right)^2}{2}-\frac{\left(\mathbb{E}\left[x_1\right]^2+\mathbb{E}\left[x_2\right]^2\right)}{2}\right\}$$

The purpose of this paper is to state and prove Theorem 1 below which generalizes this simple example to an arbitrary (possibly random) number of terms $x_1+x_2+...+x_J$ and beyond 2 to an arbitrary moment $\mathbb{E}\left[(x_1+x_2+...)^n\right]$. The problem arose in work where $\{x_j\}$ were squared increments of a randomly stopped jump process and each term on the right was summable. The theorem will apply, however, to increments of discrete random processes generally, so long as they satisfy the assumptions of the theorem to allow an application of Fubini's theorem when needed and to require covariation coefficients of all orders among the $\{x_j\}$ to satisfy a global uniformity condition.

Theorem 1 *If either* 1 *or* 2*:*

1. $x_j \geq 0$ almost always for all j, or
2. $$\mathbb{E}\left[\sum\nolimits^{II}\left|x_{j_{1,1}}\cdots x_{j_{1,i_1}}x^2_{j_{2,1}}\cdots x^2_{j_{2,i_2}}\cdots x^l_{j_{l,1}}\cdots x^l_{j_{l,i_l}}\cdots\right|\right]<\infty,$$

   for all sets of indexed non-negative integers $\left\{i_l : \sum_l l\cdot i_l = n\right\}$ where, for each such $\{i_l\}$, $\sum^{II}$ is taken over all indexed sets of permutations of sets of non-negative integers $\left\{\{j_{l,i} : 1\leq i\leq i_l\}_l\right\}$ in which no two integers $j_{l,i}$, $j_{l',i'}$ are equal,

   and if all covariation coefficients of all orders among the $\{x_j\}$ are global, not depending upon the specific subscripts j and j' for any two distinct x_j and $x_{j'}$, as specified in the statement of Lemma 7 below

then

$$\mathbb{E}\left[\left(\sum_j x_j\right)^n\right]=$$

$$=\sum\nolimits^{I}\frac{n!}{\prod\limits_l l!^{i_l}}\rho_{\{i_l\}}\sum\nolimits^{IV}\prod_m\frac{1}{j_m!}\left[(-1)^{\sum\limits_l i_{l,m}-1}\frac{\left(\sum\limits_l i_{l,m}-1\right)!}{\prod\limits_l i_{l,m}!}\sum_j\left(\prod_l\mathbb{E}\left[x^l_j\right]^{i_{l,m}}\right)\right]^{j_m}$$

where $\sum^I$ is taken over all sets of indexed non-negative integers $\left\{i_l : \sum_l l\cdot i_l = n\right\}$, for each such $\{i_l\}$ the covariation coefficient $\rho_{\{i_l\}}$ is as defined in Lemma 7 below and for each such $\{i_l\}$ the $\sum^{IV}$ is taken over all sets of indexed non-negative integers $\left\{j_m, i_{l,m} : \sum_m j_m\cdot i_{l,m} = i_l \text{ for all } l\right\}$.

Proof. The proof will be assembled as a series of Lemmata and Remarks. ∎

Remark 2 *If $\{x_j\}$ constitute squared increments of a discrete random process, for example of a randomly stopped jump process, then they satisfy hypothesis* 1 *of Theorem 1. If $\{x_j\}$ constitute increments of a stopped discrete stochastic process and if $\{x_j\}$ have finite absolute moments of all orders then they satisfy hypothesis* 2 *of Theorem 1 since in that case there will be only a finite number of x_j.*

Remark 3 *Everything in the expression for* $\mathbb{E}\left[\left(\sum_j x_j\right)^n\right]$ *in Theorem 1 is combinatoric with the exception of all of the* $\rho_{\{i_l\}}$ *and* $\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l,m}}\right)$, *which carry the probabilistic content. In the applications from which this work arose, these probabilistic expressions are, respectively, directly calculable and directly summable for each* $\{i_l\}$ *and* $\{i_{l,m}\}$ *in the combinatorics.*

Lemma 4 *(Multinomial Theorem - slightly restated)*

$$\left(\sum_j x_j\right)^n = \sum\nolimits^I \frac{n!}{\prod_l i_l! l!^{i_l}} \sum\nolimits^{II} x_{j_{1,1}} \cdots x_{j_{1,i_1}} x_{j_{2,1}}^2 \cdots x_{j_{2,i_2}}^2 \cdots x_{j_{l,1}}^l \cdots x_{j_{l,i_l}}^l \cdots$$

where $\sum^I$ *is taken over all sets of indexed non-negative integers* $\left\{i_l : \sum_l l \cdot i_l = n\right\}$ *and, for each such set* $\{i_l\}$, $\sum^{II}$ *is taken over all indexed sets of permutations of sets of non-negative integers* $\left\{\{j_{l,i} : 1 \le i \le i_l\}_l\right\}$ *in which no two integers* $j_{l,i}$, $j_{l',i'}$ *are equal. (Compared to the usual statement of the multinomial theorem, here we treat each permutation of each* $\{j_{l,i} : 1 \le i \le i_l\}_l$ *as creating a distinct monomial in* $\sum^{II}$*.)*

Proof. The monomials defined in $\sum^I$ and $\sum^{II}$ include all (and only) the monomials that can occur in the expansion of $\left(\sum_j x_j\right)^n$. Given such a monomial $x_{j_{1,1}} \cdots x_{j_{1,i_1}} x_{j_{2,1}}^2 \cdots x_{j_{2,i_2}}^2 \cdots x_{j_{l,1}}^l \cdots x_{j_{l,i_l}}^l \cdots$, without regard to the ordering among the $x_{j_{l,1}}^l \cdots x_{j_{l,i_l}}^l$ for each l, how many times does it occur in the expansion of $\left(\sum_j x_j\right)^n$? Any $x_{j_{l,i}}$ can be chosen from any one of the n factors $\left(\sum_j x_j\right)$ of $\left(\sum_j x_j\right)^n$, but no two $x_{j_{l,i}}$ in the same monomial can be chosen from the same factor $\left(\sum_j x_j\right)$ of $\left(\sum_j x_j\right)^n$. So each such occurence of the monomial in the expansion of $\left(\sum_j x_j\right)^n$ is an assignment for all l of a unique l-element subset of the n factors in $\left(\sum_j x_j\right)^n$ to each particular $x^l_{j_{l,i}}$ in the monomial. Those are the factors which contribute that particular $x^l_{j_{l,i}}$ to the monomial. The expression $\frac{n!}{\prod_l l!^{i_l}}$ for "n-choose $..., l, l, ..., l, ...$" where l runs over all positive integers and each l occurs i_l times is the correct counting of the number of ways to make such an assignment without regard to the ordering among the $x^l_{j_{l,1}} \cdots x^l_{j_{l,i_l}}$ for each l. However, for each l, there are $i_l!$ distinct permutations of each $\{j_{l,i} : 1 \le i \le i_l\}_l$ so dividing by each $i_l!$ gives the correct count when each permutation is treated as creating a distinct monomial in $\sum^{II}$. ■

Remark 5 *The coefficient* $\frac{n!}{\prod_l i_l! l!^{i_l}}$ *in Lemma 4 is the same as appears in Faá di Bruno's formula for the chain rule for higher derivatives, and comes from the same combinatorics.*

Lemma 6 *If either* 1 *or* 2

1. $x_j \ge 0$ almost always for all j, or
2. $$\mathbb{E}\left[\sum^{II} \left| x_{j_{1,1}} \cdots x_{j_{1,i_1}} x^2_{j_{2,1}} \cdots x^2_{j_{2,i_2}} \cdots x^l_{j_{l,1}} \cdots x^l_{j_{l,i_l}} \cdots \right|\right] < \infty,$$
   for all sets of indexed non-negative integers $\left\{ i_l : \sum_l l \cdot i_l = n \right\}$ where, for each such $\{i_l\}$, $\sum^{II}$ is taken over all indexed sets of permutations of sets of non-negative integers $\left\{ \{j_{l,i} : 1 \le i \le i_l\}_l \right\}$ in which no two integers $j_{l,i}$, $j_{l',i'}$ are equal,

then

$$\mathbb{E}\left[\left(\sum_j x_j\right)^n\right] =$$

$$= \sum^I \frac{n!}{\prod_l i_l! l!^{i_l}} \sum^{II} \rho_{\{j_{l,i}\}} \mathbb{E}\left[x_{j_{1,1}}\right] \cdots \mathbb{E}\left[x_{j_{1,i_1}}\right] \mathbb{E}\left[x^2_{j_{2,1}}\right] \cdots \mathbb{E}\left[x^2_{j_{2,i_2}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,1}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,i_l}}\right] \cdots$$

where for each $\left\{ \{j_{l,i} : 1 \le i \le i_l\}_l \right\}$, in which no two integers $j_{l,i}$, $j_{l',i'}$ are equal, $\rho_{\{j_{l,i}\}}$ is the covariation coefficient defined by

$$\begin{aligned} & \mathbb{E}\left[x_{j_{1,1}} \cdots x_{j_{1,i_1}} x^2_{j_{2,1}} \cdots x^2_{j_{2,i_2}} \cdots x^l_{j_{l,1}} \cdots x^l_{j_{l,i_l}} \cdots\right] \\ = \quad & \rho_{\{j_{l,i}\}} \mathbb{E}\left[x_{j_{1,1}}\right] \cdots \mathbb{E}\left[x_{j_{1,i_1}}\right] \mathbb{E}\left[x^2_{j_{2,1}}\right] \cdots \mathbb{E}\left[x^2_{j_{2,i_2}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,1}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,i_l}}\right] \cdots \end{aligned}$$

Proof. Taking $\mathbb{E}$ on both sides of Lemma 4, either hypothesis 1 or hypothesis 2 of Lemma 6 allows us to move $\mathbb{E}$ inside the summation by Fubini's theorem. For hypothesis 2 this requires the observation that $\sum^{I}$ is a finite sum. ■

Lemma 7 *If for each $\{i_l\}$ in $\sum^{I}$ in Lemma 6 the $\rho_{\{j_{l,i}\}}$ in $\sum^{II}$ are all equal, that is if all covariation coefficients of all orders among the $\{x_j\}$ are global, not depending upon the specific subscripts j and j' for any two distinct x_j and $x_{j'}$, then*

$$\mathbb{E}\left[\left(\sum_j x_j\right)^n\right] =$$

$$= \sum^{I} \frac{n!}{\prod_l i_l! l!^{i_l}} \rho_{\{i_l\}} \sum^{II} \mathbb{E}\left[x_{j_{1,1}}\right] \cdots \mathbb{E}\left[x_{j_{1,i_1}}\right] \mathbb{E}\left[x^2_{j_{2,1}}\right] \cdots \mathbb{E}\left[x^2_{j_{2,i_2}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,1}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,i_l}}\right] \cdots$$

$$= \sum^{I} \frac{n!}{\prod_l i_l! l!^{i_l}} \rho_{\{i_l\}} \left\{ \prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l} - \sum^{III} \right\}$$

where $\rho_{\{i_l\}}$ = the common value of the $\rho_{\{j_{l,i}\}}$ in $\sum^{II}$ as defined in Lemma 6 and $\sum^{III}$ represents the sum of all monomial terms in $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$, with the same coefficients as in $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$, that contain two or more matching subscripts, i.e. that contain factors $\mathbb{E}\left[x^l_j\right] \mathbb{E}\left[x^{l'}_{j'}\right]$ with $j = j'$.

Proof. Factor $\rho_{\{i_l\}} = \rho_{\{j_{l,i}\}}$ out of $\sum^{II}$ in Lemma 6. Then note that the monomials in $\sum^{II}$ are identical with the monomials in the complete expansion of $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$ that contain no matching subscripts, i.e. that contain no factors $\mathbb{E}\left[x^l_j\right] \mathbb{E}\left[x^{l'}_{j'}\right]$ with $j = j'$. ■

$\sum^{III}$ can be built up using expressions similar to $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$ but containing monomials with groups of, first, two or more matching subscripts, then with groups of three or more matching subscripts, etc and finally $\sum\limits_l i_l$ matching subscripts (there cannot be any more than that matching because $\sum\limits_l l \cdot i_l = n$.)

These expressions containing monomials with groups of matching subscripts that go into $\sum^{III}$ will take the form $\prod\limits_m \left[\sum\limits_j \left(\prod\limits_l \mathbb{E}\left[x_j^l\right]^{i_{l,m}} \right) \right]^{j_m}$ where double-indexed sets of non-negative integer exponents $\{i_{l,m}\}$ define groups of matching subscripts j in sub-monomials $\prod\limits_l \mathbb{E}\left[x_j^l\right]^{i_{l,m}}$ within the monomials of $\prod\limits_m \left[\sum\limits_j \left(\prod\limits_l \mathbb{E}\left[x_j^l\right]^{i_{l,m}} \right) \right]^{j_m}$. Sets of non-negative integer outside exponents $\{j_m\}$ serve to allow us to require the sets of inside exponents $\{i_{l,m}\}$ to be unique.

For each combined set of exponents $\{j_m, i_{l,m}\}$ we require that $\sum\limits_m j_m \cdot i_{l,m} = i_l$ for each l in order to make sure that we are generating monomials in $\sum^{III}$ that correspond to monomials in $\prod\limits_l \left(\sum\limits_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$. With one exception there will be such monomials in $\sum^{III}$ for each unique set of exponents $\{j_m, i_{l,m}\}$ meeting the requirement $\sum\limits_m j_m \cdot i_{l,m} = i_l$ for each l.

The exception is the unique set of such exponents with $j_l = i_l$ and $i_{l,l} = 1$ for all l, and all other $i_{l,m} = 0$. In this case,

$$\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l,m}} \right) \right]^{j_m} = \prod_l \left(\sum_j \mathbb{E}\left[x_j^l\right] \right)^{i_l} \text{ itself.}$$

This contains some monomials where no two subscripts match, hence which are not contained in $\sum^{III}$.

Lemma 8 *If we maintain the convention that each permutation of* $\{j_{l,i} : 1 \leq i \leq i_l\}_l$ *for each* l*, where no two integers* $j_{l,i}$*,* $j_{l',i'}$ *are equal, denotes a distinct monomial* $\mathbb{E}\left[x_{j_{1,1}}\right] \cdots \mathbb{E}\left[x_{j_{1,i_1}}\right] \mathbb{E}\left[x^2_{j_{2,1}}\right] \cdots \mathbb{E}\left[x^2_{j_{2,i_2}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,1}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,i_l}}\right] \cdots$ *then the coefficient of that monomial in* $\prod\limits_l \left(\sum\limits_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$ *will be* 1*.*

Proof. There are exactly $\prod_l i_l!$ such permutations and exactly that many distinct monomials in $\prod_l \left(\sum_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$ meeting the convention. ■

Lemma 9 *For each m, the number of subscripts matched to some other subscript in each monomial of $\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right)$ is*

$$\left(\sum_l i_{l,m} - 1 \right)_+ + 1 \wedge \left(\sum_l i_{l,m} - 1 \right)_+$$

and there are j_m groups of such matching subscripts in each monomial of

$$\left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right) \right]^{j_m}$$

.

Proof. For each m, if $\sum_l i_{l,m} = 0$ or 1 it doesn't create a match. If $\sum_l i_{l,m} > 1$ it creates $\sum_l i_{l,m}$ matched subscripts in the sub-monomial $\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}}$. For each m there are j_m such groups of matched subscripts in each monomial of $\left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right) \right]^{j_m}$. ■

Remark 10 *There will be larger groups of matching subscripts than the minimum set by Lemma 9 in some monomials of $\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right) \right]^{j_m}$ "by accident" in multiplying across the separate $\left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right) \right]$ factors. For the monomials of $\prod_l \left(\sum_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$ the minimum set by Lemma 9 is 0, which is consistent with Lemma 8.*

Remark 11 *Since the set of monomials involving groups of three or more matching subscripts forms a proper subset of the set of monomials involving groups of two or more matching subscripts, and so on, as we build up $\sum^{III}$ with terms to eliminate monomials with groups of, say, k matching subscripts* *we will have to adjust systematically for the presence of monomials with groups of k matching subscripts that we already put into $\sum^{III}$ "by accident" while putting in terms to eliminate monomials with groups of k' matching subscripts for each $2 \leq k' < k$.*

We now will derive a coefficient to put on each term $\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right) \right]^{j_m}$ in order to achieve in $\sum^{III}$ an exact elimination of all and only all monomials in $\prod_l \left(\sum_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$ that have any matching subscripts. This will require two separate analyses: first, a count of the number of times each pattern of matched subscripts occurs in $\prod_l \left(\sum_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$ and, second, an adjustment factor to associate with each such occurrence to account for all of the "by accident" occurrences generated by the elimination process itself as described in Remark 11. We can start, however, with the simplest term, one that requires no such adjustment factor.

Remark 12 *By Lemma 8 if we put the coefficient* 1 *on the term* $\prod_l \left(\sum_j \mathbb{E}\left[x_j^l\right] \right)^{i_l}$ *then we can rewrite the conclusion of Lemma 7 to read*

$$\mathbb{E}\left[\left(\sum_j x_j \right)^n \right] = \sum^I \frac{n!}{\prod_l i_l! l!^{i_l}} \rho_{\{i_l\}} \left\{ \sum^{IV} \right\}$$

where for each $\{i_l\}$ in $\sum^I$ we let $\sum^{IV}$ represent a sum of terms of the form $\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x_j^l\right]^{i_{l.m}} \right) \right]^{j_m}$ with a coefficient on each term chosen so that across $\sum^{IV}$ all occurences of monomials with any matching subscripts are eliminated, leaving only

$$\sum^{IV} = \sum^{II} \mathbb{E}\left[x_{j_{1,1}}\right] \cdots \mathbb{E}\left[x_{j_{1,i_1}}\right] \mathbb{E}\left[x^2_{j_{2,1}}\right] \cdots \mathbb{E}\left[x^2_{j_{2,i_2}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,1}}\right] \cdots \mathbb{E}\left[x^l_{j_{l,i_l}}\right] \cdots$$

with monomials having no matching subscripts and with separate monomials for each permutation of the the subscripts within each set $\left\{ \mathbb{E}\left[x^l_{j_{l,i}}\right] : 1 \leq i \leq i_l \right\}$.

Lemma 13 *For each set of indexed non-negative integers* $\left\{ i_l : \sum_l l \cdot i_l = n \right\}$ *and each set of indexed non-negative integers* $\left\{ j_m, i_{l,m} : \sum_m j_m \cdot i_{l,m} = i_l \text{ for all } l \right\}$, *the monomial*

$$\prod_m \prod_{k=1}^{j_m} \prod_l \mathbb{E}\left[x^l_{j_{m,k}}\right]^{i_{l,m}}$$

with no two subscripts $j_{m,k}$, $j_{m',k'}$ *equal occurs*

$$\frac{1}{\prod_m j_m!} \prod_l \frac{i_l!}{\prod_m i_{l,m}!^{j_m}}$$

times in $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$, *treating each permutation of the subscripts for a given* m *as a separate monomial. Taken over all such* $\{j_m, i_{l,m}\}$ *this will exhaust all the monomials in* $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$ *with each permutation of subscripts for each* l *treated as a separate monomial.*

Proof. In similar fashion to the proof of Lemma 4 if we were to treat permutations of the subscripts as yielding the same monomial then the proper count for each l, for each given $j_{m,k}$, would be "i_l-choose $..., i_{l,m}, ..., i_{l,m}, ...$" where each $i_{l,m}$ occurs j_m times. That gives a count of

$$\frac{i_l!}{\prod_m i_{l,m}!^{j_m}}$$

occurences in each $\left(\sum_j \mathbb{E}\left[x^l_j\right] \right)^{i_l}$. But this needs to be taken independently over all l making the count equal to

$$\prod_l \frac{i_l!}{\prod_m i_{l,m}!^{j_m}}$$

occurences. But we want to treat permutations of the subscripts for a given m as yielding different monomials. For a given m, each $\prod_{k=1}^{j_m} \prod_l \mathbb{E}\left[x^l_{j_{m,k}}\right]^{i_{l,m}}$ has the same subscript in all of its factors for each given k. The only room for permutation of subscripts is by permuting $\{j_{m,k}\}$ over k for each fixed m. There are $j_m!$ such permutations for each m since k runs from 1 to j_m. So dividing by $j_m!$ for each m gives the correct count

$$\frac{1}{\prod_m j_m!}\prod_l \frac{i_l!}{\prod_m i_{l,m}!^{j_m}}$$

for the number of occurences of the original monomial

$$\prod_m \prod_{k=1}^{j_m} \prod_l \mathbb{E}\left[x^l_{j_{m,k}}\right]^{i_{l,m}}$$

This exhausts all the monomials in $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right]\right)^{i_l}$ because for each m separately we have included all permutations of the $\{j_{m,k}\}$, which certainly includes all permutations of the subscripts for each l. ■

Remark 14 *Lemma 13 tells us how many times the term*

$$-\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x^l_j\right]^{i_{l.m}}\right)\right]^{j_m}$$

would have to occur in $\sum^{IV}$ *in Remark 12 to eliminate matching subscripts in* $\prod_l \left(\sum_j \mathbb{E}\left[x^l_j\right]\right)^{i_l}$*, if only we could ignore "by accident" terms as described in Remark 11*

It remains, finally, to determine a factor, other than -1 perhaps, to put on each such occurrence of each term

$$\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x^l_j\right]^{i_{l.m}}\right)\right]^{j_m}$$

in $\sum^{IV}$ to adjust for the "by accident" terms as described in Remark 11 so that each occurrence of each monomial within any of the

$$\prod_m \left[\sum_j \left(\prod_l \mathbb{E}\left[x^l_j\right]^{i_{l.m}}\right)\right]^{j_m}$$

that contains any matching subscripts is exactly eliminated within $\sum^{IV}$ in Remark 12. To do so requires yet more notation.

Notation 15 *Let the set of non-negative integers* $\{f_k\}$ *indexed by* $k \geq 2$ *represent any monomial which for each* $k \geq 2$ *has exactly* f_k *groups of* k *subscripts matching each other. To be clear, within each such group of* k *the subscripts match each other, but they do not match any other subscripts in the monomial, not even the subscripts in the other groups of* k *subscripts if* f_k *happens to be bigger than* 1.

Example 16 *Using Lemma 9, for the monomial in Lemma 13 for each* $k \geq 2$, $f_k = \sum^{V} j_m$ *where* $\sum^{V}$ *runs over all* m *such that* $\sum_l i_{l,m} = k$.

Lemma 17 *In a sum over successive monomials with groups of increasing numbers of matching subscripts to eliminate all matching subscripts, as described in Remark 11, a monomial whose subscript matching is represented by* $\{f_k\}$ *should be given an adjustment factor*

$$\prod_k \left[(-1)^{(k-1)}(k-1)!\right]^{f_k}$$

Proof. Proceed by induction on $\sum_k f_k \cdot (k-1)$ to show both that the adjustment factor for a monomial represented by $\{f_k\}$ must be

$$\text{Factor}(\{f_k\}) = \prod_k \left[(-1)^{(k-1)}(k-1)!\right]^{f_k} \qquad \text{Equation (1)}$$

and that

$$\prod_k \left(1 + \sum^{VI} \frac{k!}{\left(k - \sum_{l=2}^{k} l \cdot i_l\right)! \prod_{l=2}^{k} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}\right)^{f_k} = 0$$

Equation (2)

where for each k the $\sum^{VI}$ is taken over all sets of non-negative integers indexed by $l \geq 2$

$$\left\{i_l : 2 \leq \sum_{l=2}^{k} l \cdot i_l \leq k\right\}.$$

For $\sum_k f_k \cdot (k-1) = 1$, $f_2 = 1$ must be the only non-zero f_k so the required adjustment factor is -1 because there is exactly $f_2 = 1$ match that requires elimination and there are no "by accident" occurrences of that match stemming from eliminating monomials with fewer matches (there are no fewer matches than this.)

But in this case

$$\prod_k \left[(-1)^{(k-1)} (k-1)!\right]^{f_k} = \left[(-1)^1 \, 1!\right]^1 = -1$$

and $\sum^{VI}$ is over the single element $i_2 = 1$ so

$$\prod_k \left(1 + \sum^{VI} \frac{k!}{\left(k - \sum_{l=2}^{k} l \cdot i_l\right)! \prod_{l=2}^{k} i_l! l!^{i_l}} \prod_l (-1)^{i_l (l-1)} (l-1)!^{i_l} \right)^{f_k} = \left(1 + (-1)^1 \, 1!\right)^1 = 0$$

verifying Equations (1) and (2) when $\sum_k f_k \cdot (k-1) = 1$.

Now assume by induction that Equations (1) and (2) are correct for all $\{f_k\}$ with $\sum_k f_k \cdot (k-1)$ smaller than the current $\sum_k f_k \cdot (k-1)$.

Then Factor($\{f_k\}$) for the current $\{f_k\}$ in the successive elimination of all matches must be a sum containing four terms:

First, -1 to eliminate the original copy of $\{f_k\}$ itself

Second, $-\prod_k \left(1 + \sum^{VI} \frac{k!}{\left(k - \sum_{l=2}^{k} l \cdot i_l\right)! \prod_{l=2}^{k} i_l! l!^{i_l}} \prod_l (-1)^{i_l (l-1)} (l-1)!^{i_l} \right)^{f_k}$

to eliminate copies of $\{f_k\}$ introduced "by accident" (as explained in Remark 11) at earlier stages, with $\sum_k f_k \cdot (k-1)$ smaller than the current value of $\sum_k f_k \cdot (k-1)$. Each term in an additive expansion of $\prod_k (\sim)^{f_k}$ represents a matching pattern in a monomial at such an earlier stage whose elimination contributes copies of $\{f_k\}$ "by accident". The coefficients count the number of times each such monomial occurred in the earlier stages. For each block of k the count for each $\left\{ i_l : \sum_{l=2}^{k} l \cdot i_l \leq k \right\}$ is "k-choose $..., l, ..., l, ...$ with each l occurring i_l times" divided by $\prod_l i_l!$ permutations because we treat each permutation of subscripts as a distinct monomial.

By induction, the $\prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}$ are the factors for each occurrence of each such earlier matching pattern, hence also the number of copies of $\{f_k\}$ introduced "by accident" for each such occurence.

Third, $+1$ because $\prod_k 1^{f_k}$ should not have been subtracted. Each 1 is just a place-holder allowing $\prod_k (\sim)^{f_k}$ to select all matching patterns at earlier stages that introduced copies of $\{f_k\}$ "by accident."

Fourth, $+\prod_k \left[(-1)^{(k-1)} (k-1)!\right]^{f_k}$ because $\prod_k \left(\text{the } i_k = 1 \text{ term in } \sum^{VI}\right)^{f_k}$ should not have been subtracted. It describes the same matching pattern as $\{f_k\}$, not some earlier stage.

Putting these four terms together, noting that the first and third cancel each other,

$$\text{Factor}(\{f_k\}) = \prod_k \left[(-1)^{(k-1)} (k-1)!\right]^{f_k}$$

$$-\prod_k \left(1 + \sum^{VI} \frac{k!}{\left(k - \sum_{l=2}^{k} l \cdot i_l\right)! \prod_{l=2}^{k} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}\right)^{f_k}$$

Equation (3)

There now are two cases: (1) $\sum_k f_k > 1$ or (2) $\sum_k f_k = 1$.

Case (1): If $\sum_k f_k > 1$ then for each k

$$\left(1 + \sum^{VI} \frac{k!}{\left(k - \sum_{l=2}^{k} l \cdot i_l\right)! \prod_{l=2}^{k} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}\right) = 0$$

by induction since each represents Equation (2) for a set $\{i_l\}$ with $\sum_l i_l \cdot (l-1) < \sum_k f_k \cdot (k-1)$. This establishes Equation (2) for $\{f_k\}$ which in turn by Equation (3) establishes Equation (1) for $\{f_k\}$.

Case (2): If $\sum_k f_k = 1$ then let k be the lone index for which $f_k = 1$, all the others being $= 0$. Then Equation (3) becomes

$$\text{Factor}\left(\{f_k\}\right) =$$

$$= -\left(1 + \sum\nolimits^{VII} \frac{k!}{\left(k - \sum_{l=2}^{k-1} l \cdot i_l\right)! \prod_{l=2}^{k-1} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}\right)$$

where $\sum^{VII}$ is taken over all sets of non-negative integers indexed by $l \geq 2$

$$\left\{i_l : 2 \leq \sum_{l=2}^{k-1} l \cdot i_l \leq k\right\}.$$

Note the difference to $\sum^{VI}$ where the indices l run up to k, not $k-1$, because the $i_k = 1$ term is cancelled for $\sum^{VII}$ by the $(-1)^k (k-1)!$ term at the beginning of Equation (3).

For each such set $\{i_l\}$ use a partition of unity

$$\frac{1}{k}\left(k - \sum_{m=2}^{k-1} m \cdot i_m\right) + \frac{1}{k} \sum_{m=2}^{k-1} m \cdot i_m = 1$$

to write

$$\text{Factor}\left(\{f_k\}\right) =$$

$$= -\left(\begin{array}{c} 1 + \sum\nolimits^{VII} \frac{k!}{\left(k - \sum_{l=2}^{k-1} l \cdot i_l\right)! \prod_{l=2}^{k-1} i_l! l!^{i_l}} \cdot \\ \left(\frac{1}{k}\left(k - \sum_{m=2}^{k-1} m \cdot i_m\right) + \frac{1}{k} \sum_{m=2}^{k-1} m \cdot i_m\right) \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l} \end{array}\right)$$

$$= -\left(1 + \sum\nolimits^{VIII} \frac{(k-1)!}{\left(k - 1 - \sum_{l=2}^{k-1} l \cdot i_l\right)! \prod_{l=2}^{k-1} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}\right)$$

$$- \sum_{m=2}^{k-1} \frac{m}{k} \sum\nolimits^{VII} i_m \frac{k!}{\left(k - \sum_{l=2}^{k-1} l \cdot i_l\right)! \prod_{l=2}^{k-1} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}$$

where the $\sum^{VIII}$ in the first term is taken over all sets of non-negative integers indexed by $l \geq 2$

$$\left\{ i_l : 2 \leq \sum_{l=2}^{k-1} l \cdot i_l \leq k-1 \right\}$$

because when $\sum_{m=2}^{k-1} m \cdot i_m = k$ the partition of unity becomes $0+1$, and where the changed order of summation in the second term is justified by the fact that $i_m = 0$ precisely when it would otherwise be illegitimate.

Now

$$\left(1 + \sum^{VIII} \frac{(k-1)!}{\left(k-1-\sum_{l=2}^{k-1} l \cdot i_l \right)! \prod_{l=2}^{k-1} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l} \right) = 0$$

by induction because it is just Equation (2) for the case $f_{k-1} = 1$, all other $= 0$.

That leaves

$$\text{Factor}\left(\{f_k\}\right) =$$

$$= -\sum_{m=2}^{k-1} \frac{m}{k} \sum^{VII} i_m \frac{k!}{\left(k-\sum_{l=2}^{k-1} l \cdot i_l \right)! \prod_{l=2}^{k-1} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}$$

$$= -\sum_{m=2}^{k-1} \frac{(k-1)!}{(m-1)!\,(k-m)!} (-1)^{m-1} (m-1)! \cdot$$

$$\left(1 + \sum^{IX} \frac{(k-m)!}{\left(k-m-\sum_{l=2}^{k-m} l \cdot i_l \right)! \prod_{l=2}^{k-m} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l} \right)$$

where for each m the $\sum^{IX}$ is taken over all sets of non-negative integers

$$\left\{ i_l : 2 \leq \sum_{l=2}^{k-m} l \cdot i_l \leq k-m \right\},$$

where we have factored out $\frac{k!}{m!(k-m)!} (-1)^{m-1} (m-1)!$ from each term of $\sum^{VII}$, lowered the corresponding i_m by 1, fortunately the i_m in each term of $\sum^{VII}$ has the effect of dividing the corresponding $i_m!$ in the denominator down to

$(i_m-1)!$ which is just what's needed in lowering the i_m by 1, the upper limit of the index sums for $\sum^{IX}$ are reduced from k in $\sum^{VII}$ to $k-m$ as i_m is lowered by 1, l can run up to no more than $k-m$ because anything higher would violate the $k-m$ upper limit on the index sum, and the 1 term picks up the only $i_{k-1}=1$ term from $\sum^{VII}$ so that $2 \leq \sum_{l=2}^{k-m} l \cdot i_l \leq k-m$ in the definition of $\sum^{IX}$ makes sense in all cases.

Now, for $2 \leq m \leq k-2$

$$\left(1+\sum^{IX} \frac{(k-m)!}{\left(k-m-\sum_{l=2}^{k-m} l \cdot i_l\right)! \prod_{l=2}^{k-m} i_l! l!^{i_l}} \prod_l (-1)^{i_l(l-1)} (l-1)!^{i_l}\right)=0$$

by induction since it is just Equation (2) for the case $f_{k-m}=1$, all other $=0$, so finally only the $m=k-1$ term remains

$$\begin{aligned}\text{Factor}\left(\{f_k\}\right) &= -\frac{(k-1)!}{(k-2)!\,(1)!}(-1)^{k-2}(k-2)! \\ &= (-1)^{k-1}(k-1)!\end{aligned}$$

which establishes Equation (1) in Case (2). By Equation (3), this establishes Equation (2) for Case (2) and the induction for Lemma 17 is complete. ■

The proof of Theorem 1 is now complete, combining Lemma 6, Lemma 7, Remark 12, Lemma 13, Remark 14, Lemma 9 (see Example 16) and Lemma 17. Note that $\prod_l i_l!$ factors in numerator and denominator cancel out when Lemma 13 is combined with Remark 12 or Lemma 7, so that an exponent of j_m can be pulled out of everything except the $j_m!$.

Example 18 *What does Theorem 1 tell us to do in the original case $n=2$? Use a tabular format*

$$\begin{array}{cccccccccc} \mathbf{l}= & \mathbf{2} & \mathbf{1} & & & \mathbf{2} & \mathbf{1} & \mathbf{j}_m & m & \\ i_l= & 1 & 0 & i_{l,m}= & [& 1 & 0 & 1 & 1 &] \\ & 0 & 2 & & [& 0 & 2 & 1 & 1 &] \\ & & & & [& 0 & 1 & 2 & 1 &] \end{array}$$

and read off the terms from the groupings on the right, using both the left and the right for coefficients in Theorem 1 as needed. The result as expected is

$$\mathbb{E}\left[\left(\sum_j x_j\right)^2\right]=\left(\sum_j \mathbb{E}\left[x_j^2\right]\right)+2\rho_{\{0,2\}}\left\{\left(-\frac{1}{2}\sum_j \mathbb{E}\left[x_j\right]^2\right)+\frac{1}{2}\left(\sum_j \mathbb{E}\left[x_j\right]\right)^2\right\}$$

Example 19 *What about $n = 3$?*

$$
\begin{array}{rccccrccccccc}
\mathbf{l} = & \mathbf{3} & \mathbf{2} & \mathbf{1} & & & & \mathbf{3} & \mathbf{2} & \mathbf{1} & \mathbf{j}_m & m & \\
i_l = & 1 & 0 & 0 & i_{l,m} = & [& & 1 & 0 & 0 & 1 & 1 &] \\
& 0 & 1 & 1 & & [& & 0 & 1 & 1 & 1 & 1 &] \\
& & & & & \lceil & & 0 & 1 & 0 & 1 & 1 & \rceil \\
& & & & & \lfloor & & 0 & 0 & 1 & 1 & 2 & \rfloor \\
& 0 & 0 & 3 & & [& & 0 & 0 & 3 & 1 & 1 &] \\
& & & & & \lceil & & 0 & 0 & 2 & 1 & 1 & \rceil \\
& & & & & \lfloor & & 0 & 0 & 1 & 1 & 2 & \rfloor \\
& & & & & [& & 0 & 0 & 1 & 3 & 1 &]
\end{array}
$$

, already much more complex with six groupings of $i_{l,m}$. The resulting six term expression is

$$
\mathbb{E}\left[\left(\sum_j x_j\right)^3\right] =
$$

$$
\begin{aligned}
= \quad & \left(\sum_j \mathbb{E}\left[x_j^3\right]\right) \\
& +3\rho_{\{0,1,1\}}\left[\left(-\sum_j \mathbb{E}\left[x_j^2\right]\mathbb{E}\left[x_j\right]\right) + \left(\sum_j \mathbb{E}\left[x_j^2\right]\right)\left(\sum_j \mathbb{E}\left[x_j\right]\right)\right] \\
& +6\rho_{\{0,0,3\}}\left[\begin{array}{c} \left(\frac{1}{3}\sum_j \mathbb{E}\left[x_j\right]^3\right) + \left(-\frac{1}{2}\sum_j \mathbb{E}\left[x_j\right]^2\right)\left(\sum_j \mathbb{E}\left[x_j\right]\right) \\ +\frac{1}{6}\left(\sum_j \mathbb{E}\left[x_j\right]\right)^3 \end{array}\right]
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{E}\left[x_j^2 x_k\right] &= \rho_{\{0,1,1\}}\mathbb{E}\left[x_j^2\right]\mathbb{E}\left[x_k\right] \text{ for all } j \neq k \\
\mathbb{E}\left[x_i x_j x_k\right] &= \rho_{\{0,0,3\}}\mathbb{E}\left[x_i\right]\mathbb{E}\left[x_j\right]\mathbb{E}\left[x_k\right] \text{ for all i} \neq j \neq k
\end{aligned}
$$

With care, this result can be verified by algebraic calculations independent of the tabular display.

Example 20 *For $n = 4$ the combinatorics jump to fourteen groupings of $i_{l,m}$ and fourteen terms in the expression for* $\mathbb{E}\left[\left(\sum_j x_j\right)^4\right]$.

$$
\begin{array}{lccccclcccccccc}
\mathbf{l} = & \mathbf{4} & \mathbf{3} & \mathbf{2} & \mathbf{1} & & & & \mathbf{4} & \mathbf{3} & \mathbf{2} & \mathbf{1} & \mathbf{j}_m & \mathbf{m} & \\
i_l = & 1 & 0 & 0 & 0 & i_{l,m} = & & [& 1 & 0 & 0 & 0 & 1 & 1 &] \\
 & 0 & 1 & 0 & 1 & & & [& 0 & 1 & 0 & 1 & 1 & 1 &] \\
 & & & & & & & \lceil & 0 & 1 & 0 & 0 & 1 & 1 & \rceil \\
 & & & & & & & \lfloor & 0 & 0 & 0 & 1 & 1 & 2 & \rfloor \\
 & 0 & 0 & 2 & 0 & & & [& 0 & 0 & 2 & 0 & 1 & 1 &] \\
 & & & & & & & [& 0 & 0 & 1 & 0 & 2 & 1 &] \\
 & 0 & 0 & 1 & 2 & & & [& 0 & 0 & 1 & 2 & 1 & 1 &] \\
 & & & & & & & \lceil & 0 & 0 & 1 & 1 & 1 & 1 & \rceil \\
 & & & & & & & \lfloor & 0 & 0 & 0 & 1 & 1 & 2 & \rfloor \\
 & & & & & & & \lceil & 0 & 0 & 1 & 0 & 1 & 1 & \rceil \\
 & & & & & & & \lfloor & 0 & 0 & 0 & 2 & 1 & 2 & \rfloor \\
 & & & & & & & \lceil & 0 & 0 & 1 & 0 & 1 & 1 & \rceil \\
 & & & & & & & \lfloor & 0 & 0 & 0 & 1 & 2 & 2 & \rfloor \\
 & 0 & 0 & 0 & 4 & & & [& 0 & 0 & 0 & 4 & 1 & 1 &] \\
 & & & & & & & \lceil & 0 & 0 & 0 & 3 & 1 & 1 & \rceil \\
 & & & & & & & \lfloor & 0 & 0 & 0 & 1 & 1 & 2 & \rfloor \\
 & & & & & & & [& 0 & 0 & 0 & 2 & 2 & 1 &] \\
 & & & & & & & \lceil & 0 & 0 & 0 & 2 & 1 & 1 & \rceil \\
 & & & & & & & \lfloor & 0 & 0 & 0 & 1 & 2 & 2 & \rfloor \\
 & & & & & & & [& 0 & 0 & 0 & 1 & 4 & 1 &]
\end{array}
$$

$$\mathbb{E}\left[\left(\sum_j x_j\right)^4\right] =$$

$$\begin{aligned}
= \quad & \left(\sum_j \mathbb{E}\left[x_j^4\right]\right) \\
& +4\rho_{\{0,1,0,1\}}\left[\left(-\sum_j \mathbb{E}\left[x_j^3\right]\mathbb{E}\left[x_j\right]\right) + \left(\sum_j \mathbb{E}\left[x_j^3\right]\right)\left(\sum_j \mathbb{E}\left[x_j\right]\right)\right] \\
& +6\rho_{\{0,0,2,0\}}\left[\left(-\frac{1}{2}\sum_j \mathbb{E}\left[x_j^2\right]^2\right) + \frac{1}{2}\left(\sum_j \mathbb{E}\left[x_j^2\right]\right)^2\right] \\
& +12\rho_{\{0,0,1,2\}}\left[\begin{array}{c} \left(\sum_j \mathbb{E}\left[x_j^2\right]\mathbb{E}\left[x_j\right]^2\right) + \left(-\sum_j \mathbb{E}\left[x_j^2\right]\mathbb{E}\left[x_j\right]\right)\left(\sum_j \mathbb{E}\left[x_j\right]\right) \\ + \left(\sum_j \mathbb{E}\left[x_j^2\right]\right)\left(-\frac{1}{2}\sum_j \mathbb{E}\left[x_j\right]^2\right) + \left(\sum_j \mathbb{E}\left[x_j^2\right]\right)\frac{1}{2}\left(\sum_j \mathbb{E}\left[x_j\right]\right)^2 \end{array}\right] \\
& +24\rho_{\{0,0,0,4\}}\left[\begin{array}{c} \left(-\frac{1}{4}\sum_j \mathbb{E}\left[x_j\right]^4\right) + \left(\frac{1}{3}\sum_j \mathbb{E}\left[x_j\right]^3\right)\left(\sum_j \mathbb{E}\left[x_j\right]\right) \\ +\frac{1}{2}\left(-\frac{1}{2}\sum_j \mathbb{E}\left[x_j\right]^2\right)^2 + \left(-\frac{1}{2}\sum_j \mathbb{E}\left[x_j\right]^2\right)\frac{1}{2}\left(\sum_j \mathbb{E}\left[x_j\right]\right)^2 \\ +\frac{1}{24}\left(\sum_j \mathbb{E}\left[x_j\right]\right)^4 \end{array}\right]
\end{aligned}$$

It is quite tedious and difficult to verify this result by algebraic calculations independent of the tabular display. Clearly it gets out of hand to try to write out the expression for $\mathbb{E}\left[\left(\sum_j x_j\right)^n\right]$ as n increases. The logic for the tabular displays for i_l and $i_{l,m}$, however, can with some care be programmed for general n, so Theorem 1 provides an algorithm to compute $\mathbb{E}\left[\left(\sum_j x_j\right)^n\right]$ whenever the $\rho_{\{i_l\}}$ are computable and the $\sum_j \prod_l \mathbb{E}\left[x_j^l\right]^{i_{l,m}}$ summable.