Technical Report No. 11/04, December 2004

THREE ISOMORPHIC VECTOR SPACES–II

K. Balasubramanian and M.I. Beg

Three isomorphic vector spaces—II

K. Balasubramanian and M.I. Beg*
Department of Mathematics and Statistics, University of Hyderabad, Hyderabad-500046, India

---

Abstract

Three isomorphic vector spaces $\mathcal{B}_N^k$, $\mathcal{C}_N^k$ and $\mathcal{D}_N^k$ are defined. The interplay of these vector spaces leads to easy proofs for multinomial identities. Using automorphism each multinomial identity is recast into $(k+1)! - 1$ more identities. Distributions arising out of some stopping rules in drawing balls of k colors from an urn with and without replacement are connected so that one can easily go from one to the other. Identities involving joint cdf's of order statistics are generalized to those coming from arbitrary multivariate distributions. Discrete distribution similar to multinomial is defined, where all the calculations can be done on usual multinomial distribution and transferred to the general.

AMS classification: 60C05; 62G30

Keywords: Isomorphic vector space; Urn model; Multinomial identity; Order statistics

---

1. **Introduction**

This paper is a direct generalization of Balasubramanian and Beg [1]. In distributions coming from urn model those that arise from sampling with replacement, like multinomial, negative multinomial etc., are comparatively easy to deal with; but the corresponding distributions whose sampling has been done without replacement like multivariate hypergeometric, multivariate negative hypergeometric are harder. We develop here a transformation which will convert a problem in without replacement to with replacement and this results in unified treatment of both. This is done essentially by associating a vector space for with replacement and another isomorphic vector space for without replacement. The isomorphism of these two vector spaces solves this problem.

Another isomorphism takes us from iid random variables to arbitrary random variables, again allowing us a uniform approach.

---

*Corresponding author. Currently on sabbatical at Concordia University, Canada.
E-mail address: m_i_beg@yahoo.com (M.I. Beg).

2. **Main results**

We take $\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! r_2! \ldots r_k! (n - r_1 - r_2 - \cdots - r_k)!}$, if $r_i \geq 0$, $\sum_{i=1}^{k} r_i \leq n$ and 0, otherwise.

If N is a positive integer, define $Z_N = \{0, 1, 2, \ldots, N\}$. Let $\underline{p} = (p_1, p_2, \ldots, p_k)$ be a vector variable taking values in $[0, 1]^k$. If $\underline{m} = (m_1, m_2, \ldots, m_k)$, where m_i's are non-negative integers, define $\underline{p}^{\underline{m}} = \prod_{i=1}^{k} p_i^{m_i}$. Let $\sigma(\underline{m}) = \sum_{i=1}^{k} m_i$. Let $\mathcal{B}_N^k$ be the vector space generated by $\{\underline{p}^{\underline{m}} | \sigma(\underline{p}) = 1, \sigma(\underline{m}) \leq N\}$ over real field.

Let M be treated as a positive integral indeterminate. Define vector non-negative integral indeterminate $\underline{R} = (R_1, R_2, \ldots, R_k)$, with $\sigma(\underline{R}) = M$. Let $\mathcal{C}_N^k$ be the vector space over the real field generated by $\{\binom{M - \sigma(\underline{m})}{\underline{R} - \underline{m}} | \sigma(\underline{m}) \leq N,\ \underline{m} \leq \underline{R}\}$, where $\binom{M}{a_1, a_2, \ldots, a_k} = \frac{M!}{a_1!, a_2!, \ldots, a_k! (M - a_1 - \cdots - a_k)!}$. We now show that $\mathcal{B}_N^k$ and $\mathcal{C}_N^k$ are isomorphic vector spaces with the isomorphism given by the mapping $T(\underline{p}^{\underline{m}}) = \binom{M - \sigma(\underline{m})}{\underline{R} - \underline{m}}$ on the generators of the vector spaces and extended linearly over all elements.

Theorem 2.1. *The vector spaces $\mathcal{B}_N^k$ and $\mathcal{C}_N^k$ are isomorphic.*

Proof. Clearly, the elements of the form $\underline{p}^{\underline{m}}$ are not all linearly independent. Suppose a linear dependence is

$$\sum_{\underline{m} \in S} K(\underline{m}) \underline{p}^{\underline{m}} = 0,$$

where S is a finite set, just to show that the sum is a finite sum, then

$$\sum_{\underline{m} \in S} K(\underline{m}) \underline{p}^{\underline{m}} (p_1 + p_2 + \ldots + p_k)^{M - \sigma(\underline{m})} = 0,\ M \geq \max_{\underline{m} \in S} \sigma(\underline{m}),$$

where M is an indeterminate taking values $\geq N$,

$$\sum_{\underline{m} \in S} K(\underline{m}) \underline{p}^{\underline{m}} \sum_{\substack{\underline{R} \geq 0 \\ \sigma(\underline{R}) = M - \sigma(\underline{m})}} \binom{M - \sigma(\underline{m})}{\underline{R}} \underline{p}^{\underline{R}} = 0$$

$$\sum_{\underline{R}} \underline{p}^{\underline{R}} \sum_{\underline{m} \in S} K(\underline{m}) \binom{M - \sigma(\underline{m})}{\underline{R} - \underline{m}} = 0.$$

But $\{\underline{p}^{\underline{R}} | \underline{R} \geq \underline{0},\ \sigma(\underline{R}) = M\}$ are linearly independent, a result equivalent to the completeness of the multinomial family. Hence,

$$\sum_{\underline{m} \in S} K(\underline{m}) \binom{M - \sigma(\underline{m})}{\underline{R} - \underline{m}} = 0,\ \forall \underline{R},\ \underline{R} \geq \underline{0},\ \sigma(\underline{R}) = M.$$

Also, all the steps are clearly reversible. This proves the assertion that $\mathcal{B}_N^k$ and $\mathcal{C}_N^k$ are isomorphic as vector spaces, the isomorphism given by the mapping $T(\underline{p}^{\underline{m}}) = \binom{M - \sigma(\underline{m})}{\underline{R} - \underline{m}}$. $\square$

3. Application to multinomial identities

The isomorphism of $\mathcal{B}_N^k$ and $\mathcal{C}_N^k$ affords a method of getting some multinomial identities. We now give some applications.

(i)

$$\sum_{i=1}^{k} p_i = 1 \Rightarrow \sum_{i=1}^{k} T(p_i) = T(1) \Rightarrow \sum_{i=1}^{k} \binom{M-1}{\underline{R}-e_i} = \binom{M}{\underline{R}},$$

where e_i is a k-vector with i-th component 1 and other components 0. For $k=3$, this reduces to

$$\binom{M-1}{R_1-1, R_2, R_3} + \binom{M-1}{R_1, R_2-1, R_3} + \binom{M-1}{R_1, R_2, R_3-1} = \binom{M}{R_1, R_2, R_3}.$$

(ii)

$$(p_1+p_2+\ldots+p_k)^n = 1 \Rightarrow \sum_{\underline{r},\ \sigma(\underline{r})=n} \binom{n}{\underline{r}} \underline{p}^{\underline{r}} = 1 \Rightarrow \sum_{\underline{r},\ \sigma(\underline{r})=n} \binom{n}{\underline{r}} T(\underline{p}^{\underline{r}}) = T(1)$$

$$\Rightarrow \sum_{\underline{r},\ \sigma(\underline{r})=n} \binom{n}{\underline{r}} \binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}} = \binom{M}{\underline{R}}.$$

(iii)

$$(p_1+p_2+\ldots+p_{k-1}-p_k)^n = (1-2p_k)^n \Rightarrow \sum_{\underline{r},\ \sigma(\underline{r})=n} \binom{n}{\underline{r}} \underline{p}^{\underline{r}} (-1)^{r_k} = \sum_{t=0}^{n} \binom{n}{t} (-2)^t p_k^t.$$

Applying T, we get

$$\sum_{\underline{r},\ \sigma(\underline{r})=n} (-1)^{r_k} \binom{n}{\underline{r}} \binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}} = \sum_{t=0}^{n} \binom{n}{t} (-2)^t \binom{M-t}{R_1, R_2, \ldots, R_{k-1}, R_k - t}.$$

$\mathcal{B}_N^k$and $\mathcal{C}_N^k$ can be extended to $\mathcal{B}_N^{*k}$and $\mathcal{C}_N^{*k}$ by allowing $\underline{m}$ to merely $\sigma(\underline{m}) \leq N$, but $\underline{m} \geq \underline{0}$ is dispensed with. That is, we allow even negative powers. But the isomorphism still holds between $\mathcal{B}_N^{*k}$and $\mathcal{C}_N^{*k}$. This follows from an argument similar to what was used in Balasubramanian and Beg [1]. Here again the number of terms in summation should be finite.

(iv) For $k=3$, that is, $p_1+p_2+p_3=1$,

$$\sum_{r=0}^{m} (p_2+p_3)^r = \frac{1-(p_2+p_3)^{m+1}}{1-(p_2+p_3)} = p_1^{-1} - \sum_{i=0}^{m+1} \binom{m+1}{i} (-1)^i p_1^{i-1}$$

$$\Rightarrow \sum_{r=0}^{m} \sum_{i=0}^{r} \binom{r}{i} p_2^i p_3^{r-i} = p_1^{-1} - \sum_{i=0}^{m+1} \binom{m+1}{i} (-1)^i p_1^{i-1}.$$

Applying T, we get

$$\sum_{r=0}^{m}\sum_{i=0}^{r}\binom{r}{i}\binom{M-r}{R_1, R_2-i, R_3-r+i} = \binom{M+1}{R_1+1, R_2, R_3}$$

$$-\sum_{i=0}^{m+1}\binom{m+1}{i}(-1)^i\binom{M-i+1}{R_1-i+1, R_2, R_3}.$$

4. Regular series

Suppose $\sum_{\underline{m}\in S} K(\underline{m})\underline{p}^{\underline{m}} = c\underline{p}^{\underline{m}}$, where S is no longer finite. Then

$$\sum_{\underline{m}\in S} K(\underline{m})T(\underline{p}^{\underline{m}}) = c\ T(\underline{p}^{\underline{m}}),$$

may or may not hold. If the equality holds, call the series regular. In other words, for an identity involving regular series we can apply T, term by term, to both sides. Note that a finite series is always regular.

It will be an interesting exercise to find out a necessary and sufficient condition for an infinite series to be a regular. Consider,

$$\sum_{\underline{r}\geq\underline{0}} \frac{\underline{p}^{\underline{r}}}{(r_1!, r_2!, \dots r_k!)} = e^{p_1+p_2+\dots+p_k} = ep_1^0 p_2^0 \dots p_k^0.$$

applying T, we get

$$\sum_{\underline{r}\geq\underline{0}} \binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}} \frac{1}{(r_1!, r_2!, \dots r_k!)} = e\binom{M}{\underline{R}}.$$

This can not be correct. The left hand side is actually a finite series for the summation extends only over those $\underline{r}$ for which $\underline{r} \leq \underline{R}$ and $\sigma(\underline{r}) \leq M$ and the actual equality will mean that e is a rational number. Hence the infinite series in $\underline{p}$ is not regular.

5. Automorphisms of $\mathcal{B}_N^{*k}$ and $\mathcal{C}_N^{*k}$

It is not difficult to see that $\mathcal{B}_N^{*k}$ behaves like an algebra so long as we confine to terms $\underline{p}^{\underline{r}}$ with $\sigma(\underline{r}) \leq M$. Since $\underline{p}^{\underline{r}}\ \underline{p}^{\underline{s}} = \underline{p}^{\underline{r}+\underline{s}}$ is the multiplication in $\mathcal{B}_N^{*k}$ (so long as $\sigma(\underline{r})$, $\sigma(\underline{s})$, $\sigma(\underline{r})+\sigma(\underline{s})$ all $\leq M$), we can define a multiplication in $\mathcal{C}_N^{*k}$:

$$\binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}} * \binom{M-\sigma(\underline{s})}{\underline{R}-\underline{s}} = \binom{M-\sigma(\underline{r})-\sigma(\underline{s})}{\underline{R}-\underline{r}-\underline{s}}.$$

This makes $\mathcal{C}_N^{*k}$ an algebra with the suggested restrictions on $\underline{r}$ and $\underline{s}$. Considering $\mathcal{B}_N^{*k}$ as an algebra in this spirit, the group of algebra automorphisms has $(k+1)!$ elements. These elements are

$$\sigma : (p_1, p_2, \ldots, p_k) \to (p_{\sigma(1)}, p_{\sigma(2)}, \ldots, p_{\sigma(k)}),\ \sigma \in S_k$$

where S_k is the symmetric group on $\{1, 2, \ldots, k\}$.

$$\mu_i : (p_1, p_2, \ldots, p_k) \to (-\frac{p_1}{p_i}, -\frac{p_2}{p_i}, \ldots, -\frac{p_{i-1}}{p_i}, \frac{1}{p_i}, -\frac{p_{i+1}}{p_i}, \ldots, -\frac{p_k}{p_i}),\ i = 1, 2, \ldots, k.$$

$\mu_0 : (p_1, p_2, \ldots, p_k) \to (p_1, p_2, \ldots, p_k)$, the identity transformation. It is not difficult to show that the transformations $\{\sigma\mu_i | \sigma \in S_k,\ i = 0, 1, 2, \ldots, k\}$, $(k+1)!$ in number, actually form a group of order $(k+1)!$. This group is the automorphism group of the algebra $\mathcal{B}_N^{*k}$. We can use these automorphisms to get $(k+1)!$ identities starting with one algebraic identity involving $p_1, p_2, \ldots, p_k$ (including negative powers). Usually what we get may not be very surprising or even interesting. But these automorphisms translated into $\mathcal{C}_N^{*k}$ will be very interesting. Typically when we apply

$$\begin{aligned}\mu_1 : \underline{p}^{\underline{m}} \to (\frac{1}{p_1})^{m_1} (-\frac{p_2}{p_1})^{m_2} \ldots (-\frac{p_k}{p_1})^{m_k} &= \frac{(-1)^{\sigma(\underline{m})-m_1}}{{p_1}^{\sigma(\underline{m})}} \ \frac{\underline{p}^{\underline{m}}}{p_1^{m_1}} \\ &= (-1)^{\sigma(\underline{m})-m_1} \underline{p}^{\underline{m}} \ {p_1}^{-(\sigma(\underline{m})+m_1)}.\end{aligned}$$

Hence,

$$\binom{M - \sigma(\underline{m})}{\underline{R} - \underline{m}} \to (-1)^{\sigma(\underline{m})-m_1} \binom{M - m_1}{\underline{R} - \underline{m}^*},$$

where $\underline{m}^* = (-\sigma(\underline{m}), m_2, m_3, \ldots, m_k)$. Thus in any identity involving multinomial coefficients the above transformation yields another identity. We can use any one of the $(k+1)!$ such transformations.

6. With replacement to without replacement

We will motivate this section with an example. Let $\underline{X} = (X_1, X_2, \ldots, X_k)$ have a multinomial distribution with parameters $n, p_1, p_2, \ldots, p_k$. Then $\sum P(X_1 = r_1, X_2 = r_2, \ldots, X_k = r_k) = 1$, where the summation is over all non-negative integers $r_1, r_2, \ldots, r_k$ such that $\sum_{i=1}^k r_i = n$, i.e.,

$$\sum_{\sigma(\underline{r})=n} P(\underline{X} = \underline{r}) = 1 \Rightarrow \sum_{\sigma(\underline{r})=n} \binom{n}{\underline{r}} \underline{p}^{\underline{r}} = 1.$$

Applying T, we get $\sum_{\sigma(\underline{r})=n} \binom{n}{\underline{r}} \binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}} = \binom{M}{\underline{R}}$. This means

$$\frac{\sum_{\sigma(\underline{r})=n} \binom{n}{\underline{r}} \binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}}}{\binom{M}{\underline{R}}} = 1.$$

Hence, if $\underline{Z} = (Z_1, Z_2, \ldots, Z_k)$ and $P(\underline{Z} = \underline{r}) = [\binom{n}{\underline{r}}\binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}}]/\binom{M}{\underline{R}}$, for $\sigma(\underline{r}) = n$, $[= \binom{R_1}{r_1}\binom{R_2}{r_2}\cdots\binom{R_k}{r_k}/\binom{M}{n}]$, then we get a probability distribution which is an analogue of multinomial distribution. In fact, this distribution is the multivariate hypergeometric distribution, which arises from sampling without replacement (note that mutinomial distribution arises from sampling with replacement).

Any expectation of a function of Z, say $g(\underline{Z})$ can be got from $g(\underline{X})$ very easily and this is explained in what follows.

$$E[g(\underline{Z})] = \sum_{\sigma(\underline{r})=n} g(\underline{r})P(\underline{Z} = \underline{r}) = \frac{\sum_{\sigma(\underline{r})=n} g(\underline{r})\binom{n}{\underline{r}}\binom{M-n}{\underline{R}-\underline{r}}}{\binom{M}{\underline{R}}}$$

$$= \frac{T\{\sum_{\sigma(\underline{r})=n} g(\underline{r})\binom{n}{\underline{r}}\underline{p}^{\underline{r}}\}}{\binom{M}{\underline{R}}}.$$

Hence,

$$E[g(\underline{Z})] = \frac{T(Eg(\underline{X}))}{\binom{M}{\underline{R}}}.$$

Note that we identify $\underline{R}$ with $M\underline{p}$.

We now take an example. The mixed factorial moment of multinomial distribution is given by

$$\mu(r_1, r_2, \ldots, r_k) = E[X_1^{(r_1)}X_2^{(r_2)}\ldots X_k^{(r_k)}] = n^{(\sum_{j=1}^k r_j)}p_1^{r_1}p_2^{r_2}\ldots p_k^{r_k}.$$

Hence, the mixed factorial moment for $\underline{Z}$ will be

$$\mu_{\underline{Z}}(\underline{r}) = \frac{T\{n^{\sigma(\underline{r})}\underline{p}^{\underline{r}}\}}{\binom{M}{\underline{R}}} = \frac{n^{\sigma(\underline{r})}\binom{M-\sigma(\underline{r})}{\underline{R}-\underline{r}}}{\binom{M}{\underline{R}}}.$$

Thus, all the expectations of functions of a multivariate hypergeometric variate can be got from the corresponding expectations of a multinomial variate by a simple transformation. Truly, we may treat the transformation T as one taking us from sampling with replacement to without replacement.

We may apply a similar method, whenever we have some stopping rule for extended Bernoulli trials (an event with k outcomes with probabilities $p_1, p_2, \ldots p_k$), resulting in a random vector $\underline{X}$, whose density is a polynomial in $\underline{p}$, say $P(\underline{X} = \underline{r}) = \pi_{\underline{r}}(\underline{p})$, where $\sum_{\underline{r}} \pi_{\underline{r}}(\underline{p}) = 1$, is a polynomial in $\underline{p}$. It may happen that the range of $\underline{r}$ is infinite. Here, we need a definition. We call the series $\sum_{\underline{r}} \pi_{\underline{r}}(\underline{p}) = 1$, regular if $\sum_{\underline{r}} T[\pi_{\underline{r}}(\underline{p})] = T(1) = \binom{M}{\underline{R}}$. Not all such series are regular. A finite series is always regular.

Though $\underline{X}$ may have infinite range, $\underline{Z}$ arising out of it by the application of T has always a finite range. We give an example.

Multivariate negative multinomial and multivariate negative hypergeometric

Suppose in an experiment there are k possible outcomes $S_1, S_2, \ldots, S_k$. Let probability of getting S_i in a single trial be $p_i,\ i = 1, 2, \ldots, k$, so that $\sum_{i=1}^{k} p_i = 1$. If we perform the trials independently till we get n outcomes of the type S_1, then the distribution of $(Y_2, Y_3, \ldots, Y_k)$, where Y_i is the number of outcomes of the type S_i, has a multivariate negative multinomial distribution. This is given by

$$P(Y_2 = r_2, Y_3 = r_3, \ldots, Y_k = r_k) = \binom{n + \sum_{i=2}^{k} r_i - 1}{n-1, r_2, \ldots, r_k} p_1^n p_2^{r_2} \ldots p_k^{r_k},$$

for $r_2, r_3, \ldots, r_k \geq 0$. Applying T, we get

$$\sum_{r_2, r_3, \ldots, r_k \geq 0} \binom{n + \sum_{i=2}^{k} r_i - 1}{n-1, r_2, \ldots, r_k} \binom{M - n - r_2 - r_3 - \ldots - r_k}{\underline{R} - (n, r_2, r_3, \ldots, r_k)} = \binom{M}{\underline{R}},\ R_1 \geq n.$$

It is not difficult to show that this equality is true showing thereby that the series

$$\sum_{r_2, r_3, \ldots, r_k \geq 0} \binom{n + \sum_{i=2}^{k} r_i - 1}{n-1, r_2, \ldots, r_k} p_1^n p_2^{r_2} \ldots p_k^{r_k} = 1,$$

is regular. This gives rise to random variables $(Z_2, Z_3, \ldots, Z_k)$ having distribution

$$P(Z_2 = r_2, Z_3 = r_3, \ldots, Z_k = r_k) = \frac{\binom{n + \sum_{i=2}^{k} r_i - 1}{n-1, r_2, \ldots, r_k} \binom{M - n - r_2 - r_3 - \ldots - r_k}{\underline{R} - (n, r_2, r_3, \ldots, r_k)}}{\binom{M}{\underline{R}}}.$$

This distribution may be called the multivariate negative hypergeometric distribution. This arises as follows. Suppose an urn contains R_i balls of colors $C_i,\ i = 1, 2, \ldots, k$ with $\sum_{i=1}^{k} R_i = M,\ R_1 \geq n$. Balls are drawn one by one without replacement till n balls of color C_1 are obtained. Then Z_i is the number of balls of color $C_i,\ i = 2, 3, \ldots, k$ got in this sequence of trials. In other words, this is the without replacement version of multivariate negative multinomial distribution if we assume $R_i = Mp_i,\ i = 1, 2, \ldots, k$ and the balls are drawn with the same stopping rule with replacement.

If $g(Z_2, Z_3, \ldots, Z_k)$ is any function of $Z_2, Z_3, \ldots, Z_k$ then we have the identity

$$E[g(Z_2, Z_3, \ldots, Z_k)] = \frac{T\{E[g(Y_2, Y_3, \ldots, Y_k)]\}}{\binom{M}{\underline{R}}}.$$

This makes every calculation in the 'without replacement' case to be performed in 'with replacement' followed by an application of T.

'With replacement' calculations are much easier compared to 'without replacement'. So, the use of T is indeed of for reaching significance. The scope of this method is as follows.

Suppose we have an urn containing R_i balls of color C_i, $i = 1, 2, \ldots, k$ with $\sum_{i=1}^k R_i = M$. We define $p_i = (R_i/M)$, $i = 1, 2, \ldots, k$. In one case we draw balls from the urn one by one with replacement and stop using some stopping rule. When we stop, let Y_i be the number of balls of color C_i obtained, for $i = 1, 2, \ldots, k$. Let $(Y_1, Y_2, \ldots, Y_k)$ have a proper distribution with $P(Y_1 = r_1, Y_2 = r_2, \ldots, Y_k = r_k) = \pi_{\underline{r}}(\underline{p})$, $\sum_{\underline{r}} \pi_{\underline{r}}(\underline{p}) = 1$, where the summation is over all relevant values of $\underline{r}$. We also assume that this series, $\sum_{\underline{r}} \pi_{\underline{r}}(\underline{p}) = 1$, if infinite, is regular, then we can take balls one by one without replacement and use the same stopping rule and get random variables $(Z_1, Z_2, \ldots, Z_k)$ where Z_i is the number of balls of color C_i obtained at the time of stopping, $i = 1, 2, \ldots, k$. In this case for any function $g(Z_1, Z_2, \ldots, Z_k)$ we can get expectation with the simple formula

$$E[g(Z_1, Z_2, \ldots, Z_k)] = \frac{T\{E[g(Y_1, Y_2, \ldots, Y_k)]\}}{\binom{M}{\underline{R}}}.$$

7. The vector space $\mathcal{D}_N^k$

Suppose $\underline{p}_i = (p_{i1}, p_{i2}, \ldots, p_{ik})$, with $\sum_{j=1}^k p_{ij} = 1$, $i = 1, 2, \ldots, N$. Let $S_{a_1}, S_{a_2}, \ldots, S_{a_k}$ be disjoint subsets of Z_N of cardinalities $a_1, a_2, \ldots, a_k$, respectively. Let

$$p_j(S_a) = \prod_{i \in S_a} p_{ij}$$

and

$$[a_1, a_2, \ldots, a_k] = \binom{N}{a_1, a_2, \ldots, a_k}^{-1} \sum p_1(S_{a_1}) p_2(S_{a_2}) \ldots p_k(S_{a_k}),$$

where the summation is over all disjoint subsets $S_{a_1}, S_{a_2}, \ldots, S_{a_k}$ of Z_N.
or

$$[\underline{a}] = \binom{N}{\underline{a}}^{-1} \sum \prod_{j=1}^k p_j(S_{a_j}),$$

where $\binom{N}{\underline{a}} = \frac{N!}{a_1! a_2! \ldots a_k! (N - \sigma(\underline{a}))!}$. Let $\mathcal{D}_N^k$ be the vector space generated by $\{[\underline{a}],\ \underline{a} = (a_1, a_2, \ldots, a_k)$ with $\sigma(\underline{a}) \leq N\}$ over the real field.

Theorem 7.1. The vector spaces $\mathcal{D}_N^k$, $\mathcal{C}_N^k$ and $\mathcal{B}_N^k$ are isomorphic.

Proof.

$$\prod_{i=1}^N (1 + \sum_{j=1}^k u_j p_{ij}) = \sum_{\substack{\underline{\alpha} \\ \sigma(\underline{\alpha}) \leq N}} \underline{u}^{\underline{\alpha}} [\underline{\alpha}] \binom{N}{\underline{\alpha}},$$

where $\underline{u} = (u_1, u_2, \ldots, u_k)$ and $\underline{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$. But

$$\prod_{i=1}^{N}(1 + \sum_{j=1}^{k} u_j p_{ij}) = \prod_{i=1}^{N}(\sum_{j=1}^{k}(1 + u_j)p_{ij}) = \sum_{\substack{\underline{r} \\ \sigma(\underline{r})=N}} (\underline{u} + \underline{1})^{\underline{r}} \, [\underline{r}] \binom{N}{\underline{r}},$$

where $\underline{1} = (1, 1, \ldots, 1)$, a k-vector. Equating coefficients of $\underline{u}^{\underline{\alpha}}$ in both these expansions, we get

$$[\underline{\alpha}] \binom{N}{\underline{\alpha}} = \sum_{\substack{\underline{r} \\ \sigma(\underline{r})=N}} \binom{\underline{r}}{\underline{\alpha}} [\underline{r}] \binom{N}{\underline{r}}, \quad \binom{\underline{r}}{\underline{\alpha}} = \prod_{i=1}^{k} \binom{r_i}{\alpha_i}.$$

Simplifying this, we get

$$[\underline{\alpha}] = \sum_{\substack{\underline{r} \\ \sigma(\underline{r})=N}} \binom{N - \sigma(\underline{\alpha})}{\underline{r} - \underline{\alpha}} [\underline{r}].$$

Suppose

$$\sum_{\underline{\alpha} \in S} K(\underline{\alpha})[\underline{\alpha}] = \sum_{\underline{\alpha} \in S} K(\underline{\alpha}) \sum_{\underline{r}} \binom{N - \sigma(\underline{\alpha})}{\underline{r} - \underline{\alpha}} [\underline{r}] = 0$$

is a linear dependence of $[\underline{\alpha}]$'s. This can be written as

$$\sum_{\substack{\underline{r} \\ \sigma(\underline{r})=N}} [\underline{r}] \sum_{\underline{\alpha} \in S} K(\underline{\alpha}) \binom{N - \sigma(\underline{\alpha})}{\underline{r} - \underline{\alpha}} = 0.$$

But $\{[\underline{r}],\ \sigma(\underline{r}) = N\}$ are linearly independent. In fact, even for the special case $\underline{p_1} = \underline{p_2} = \ldots = \underline{p_N} = \underline{p},\ [\underline{r}] = \underline{p}^{\underline{r}}$, all these are linearly independent from the completeness of multinomial family. Hence,

$$\sum_{\underline{\alpha} \in S} K(\underline{\alpha}) \binom{N - \sigma(\underline{\alpha})}{\underline{r} - \underline{\alpha}} = 0, \ \forall\ \underline{r},\ \sigma(\underline{r}) = N.$$

Clearly, all these steps are reversible. This shows that the vector spaces $\mathcal{D}_N^k$ and $\mathcal{C}_N^k$ are isomorphic. We already know that $\mathcal{C}_N^k$, $\mathcal{B}_N^k$ are isomorphic. $\square$

8. Applications to identities involving cdf's of order statistics

Suppose $X_1, X_2, \ldots$ are iid random variables with cdf $F(x)$. Let $F_{\underline{m}:n}(\underline{x})$, $\underline{m} = (m_1, m_2, \ldots, m_{k-1})$, $\underline{x} = (x_1, x_2, \ldots, x_{k-1})$, be the joint cdf of of $X_{m_1:n}, X_{m_2:n}, \ldots, X_{m_{k-1}:n}$, where $X_{r:n}$ is the rth order statistic from a sample of size n.

Suppose an identity of the form

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n)F_{\underline{m}:n}(\underline{x}) = 0$$

holds for some finite set S. Let us also consider random variables $Y_1, Y_2, \cdots, Y_N$ which may have any arbitrary N-variate distribution. Assume

$$N \geq \max_{(\underline{m},n)\in S} \{n\}.$$

For any $n \leq N$ and $1 \leq m_1 < m_2 < \ldots < m_{k-1} \leq n$, we define $\overline{F}_{\underline{m}:n}(\underline{x})$ as the average of the cdf's of $X_{m_1:n}, X_{m_2:n}, \ldots, X_{m_{k-1}:n}$, from all the $\binom{N}{n}$ samples of size n from $Y_1, Y_2, \cdots Y_N$. We will prove that if

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n)F_{\underline{m}:n}(\underline{x}) = 0$$

holds for all iid variables, then

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n)\overline{F}_{\underline{m}:n}(\underline{x}) = 0$$

holds for arbitrary $Y_1, Y_2, \cdots, Y_N$.

Proof. We know that $F_{\underline{m}:n}(\underline{x})$ for iid case is a polynomial in $F(x_1), F(x_2) - F(x_1), F(x_3) - F(x_2), \ldots, F(x_{k-1}) - F_(x_{k-2}), 1 - F(x_{k-1})$, say $\sum_{\underline{r}} C_{\underline{m}:n}(\underline{r})\underline{p}^{\underline{r}}$, where $p_1 = F(x_1), p_2 = F(x_2) - F(x_1), \ldots, p_{k-1} = F(x_{k-1}) - F(x_{k-2}), p_k = 1 - F(x_{k-1})$. We then get,

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n) \sum_{\underline{r}} C_{\underline{m}:n}(\underline{r})\underline{p}^{\underline{r}} = 0.$$

By isomorphism, we should have

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n) \sum_{\underline{r}} C_{\underline{m}:n}(\underline{r})[\underline{r}] = 0.$$

If we take $p_{ij} = I_{(x_{j-1} < Y_i \leq x_j)}$, where $x_0 = -\infty$ and $x_k = \infty$, then taking expectation, we get

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n) \sum_{\underline{r}} C_{\underline{m}:n}(\underline{r})E[\underline{r}] = 0$$

$$\sum_{(\underline{m},n)\in S} K(\underline{m},n)\overline{F}_{\underline{m}:n}(\underline{x}) = 0. \qquad \square$$

9. Application: Generalization of some discrete distributions

We will illustrate the principle by a univariate example first. Consider the binomial distribution $B(n,p)$. Here, the binomial random variable X has the property $P(X=r)=\binom{n}{r}p^r q^{n-r}$. Define another random variable Y, with the same range as X, $P(Y=r)=\binom{n}{r}[r,n-r]$. By isomorphism, $\sum_{r=0}^{n}\binom{n}{r}[r,n-r]=1$, showing that Y has indeed a genuine distribution. Y has parameters $(n,\pi_1,\pi_2,\ldots,\pi_N)$ which may be written as $(n,\underline{\pi})$. Here, $\underline{\pi}=(\pi_1,\pi_2,\ldots,\pi_N)$ is the vector with which we define $[r,n-r]$. Here again, if $E[g(x)]=\sum_{r=0}^{n}C_r p^r q^{n-r}$, then $E[g(Y)]=\sum_{r=0}^{n}C_r[r,n-r]$ and no separate calculation for Y is needed. Y may be called a multiparameter binomial variate.

This method can be used only when the random variable X defined by a stopping rule in Bernoulli trial is bounded. In general, if we have a random variable X taking values $0,1,2,\ldots,n$ where $P(X=r)=\pi_r(p,q)$, a polynomial in p and q, say $\pi_r(p,q)=\sum_{(a,b)\in S_r}C_{a,b}(r)p^a q^b$. Then we may define a random variable Y taking values $0,1,2,\ldots,n$ with $P(Y=r)=\sum_{(a,b)\in S_r}C_{a,b}(r)[a,b]$. Y will have parameters $(n,\pi_1,\pi_2,\ldots,\pi_N)$. If $E[g(X)]=\sum_{r=0}^{n}\sum_{(a,b)\in S_r}d_{a,b}(r)p^a q^b$, then $E[g(Y)]=\sum_{r=0}^{n}\sum_{(a,b)\in S_r}d_{a,b}(r)[a,b]$. Now we can generalize this to multivariate discrete distributions. If we have $\underline{X}=(X_1,X_2,\ldots,X_m)$ taking values in $M=(0,1,2,\ldots,n)^m$, where $P(\underline{X}=\underline{r})=\pi_{\underline{r}}(\underline{p})$, a polynomial in $\underline{p}$ $(=(p_1,p_2,\ldots,p_k))$ $(\sum_{i=1}^{k}p_i=1)$. Writing $\pi_{\underline{r}}(\underline{p})=\sum_{\underline{i}}C_{\underline{r}}(\underline{i})\underline{p}^{\underline{i}}$, we can define another random variable $\underline{Y}=(Y_1,Y_2,\ldots,Y_m)$ taking values in $M=(0,1,2,\ldots,n)^m$, where $P(\underline{Y}=\underline{r})=\sum_{\underline{i}}C_{\underline{r}}(\underline{i})[\underline{i}]$, $\underline{r}\in M$. If $E[g(X)]=\sum_{\underline{i}}d_{\underline{r}}(\underline{i})\underline{p}^{\underline{i}}$, then $E[g(Y)]=\sum_{\underline{i}}d_{\underline{r}}(\underline{i})[\underline{i}]$, by isomorphism.

10. Restricted multivariate exponential type distributions

Suppose $\underline{X}$ has a k-variate distribution over a subset Ω of Z^{+k}, where $Z^+=\{0,1,2,\ldots\}$ with

$$P(\underline{X}=\underline{x})=h(\underline{p})g(\underline{x})\underline{p}^{\underline{m}(\underline{x})},\ \underline{x}\in\Omega,$$

where $\underline{p}=(p_1,p_2,\ldots,p_n)$, $p_i\geq 0$ such that $\sum_{i=1}^{n}p_i\leq 1$ and $\underline{m}(\underline{x})=(m_1(\underline{x}),m_2(\underline{x}),\ldots,m_n(\underline{x}))$ and $h(\underline{p})$ of the form $\sum_{\underline{r}\in S}C(\underline{r})\underline{p}^{\underline{r}}$, where $\underline{r}$ is an n-vector of integers (not necessarily non-negative). We call the distribution a restricted multivariate exponential type distribution.

Examples

1. *Multinomial Distribution*

$$P(\underline{X}=\underline{x})=\binom{n}{\underline{x}}\underline{p}^{\underline{x}},\ \Omega=\{\underline{x}\geq\underline{0},\ \sigma(\underline{x})=n\},$$

where

$$h(\underline{p}) \equiv 1,\ g(\underline{x}) = \binom{n}{\underline{x}},\ \underline{m}(\underline{x}) = \underline{x}.$$

2. *Multivariate Geometric Distribution*

$$P(\underline{X} = \underline{x}) = \binom{\sigma(\underline{x})}{\underline{x}} \underline{p}^{\underline{x}}(1 - \sigma(\underline{p})),\ \Omega = Z^{+k},$$

where

$$h(\underline{p}) = (1 - \sigma(\underline{p})),\ g(\underline{x}) = \binom{\sigma(\underline{x})}{\underline{x}},\ \underline{m}(\underline{x}) = \underline{x}.$$

3. *Negative Multinomial Distribution*

$$P(\underline{X} = \underline{x}) = \binom{n + \sum_{i=1}^{k} x_i - 1}{\underline{x}, n-1} (1 - \sigma(\underline{p}))^n \underline{p}^{\underline{x}},\ \underline{x} \geq \underline{0}.$$

Many more examples can be given. In all these cases we can define the corresponding 'without replacement' versions as follows. Define $\underline{Y}$ by

$$P(\underline{Y} = \underline{y}) = g(\underline{y}) \binom{N - \sigma(\underline{m}(\underline{x}))}{\underline{R} - \underline{m}(\underline{x})} * \sum_{\underline{r} \in S} C(\underline{r}) \binom{N - \sigma(\underline{r})}{\underline{R} - \underline{r}}$$

$$= g(y) \sum_{\underline{r} \in S} C(\underline{r}) \binom{N - \sigma(\underline{m}(\underline{x})) - \sigma(\underline{r})}{\underline{R} - \underline{m}(\underline{x}) - \underline{r}}$$

where $\underline{x}$ takes appropriate values in Ω, i.e., a subset of Ω, say Ω^*, in which $\sigma(\underline{m}(\underline{x})) - \sigma(\underline{r}) \leq N$ and $\underline{m}(\underline{x}) + \underline{r} \leq \underline{R}$. Essentially this subset Ω^* is a finite subset. All the properties of $\underline{Y}$ that can be got through expectation of functions of $\underline{Y}$, can be easily got from expectation of the same function of $\underline{X}$, which is expected to be easier to calculate. This transition from 'with replacement' to 'without replacement' is achieved through the versatile transformation T.

Reference

[1] K. Balasubramanian and M. I. Beg, Three isomorphic vector spaces —applications to binomial identities, urn models and order statististics, Linear Algebra Appl. 388 (2004) 79-89.

List of Recent Technical Reports

46. B. N. Dimitrov, Z. Khalil, M. E. Ghitany and V. V. Rykov, *Likelihood Ratio Test for Almost Lack of Memory Distributions*, November 2001

47. Yogendra P. Chaubey and Anthony Crisalli, *The Generalized Smoothing Estimator*, April 2002

48. Yogendra P. Chaubey and Pranab K. Sen, *Smooth Isotonic Estimation of Density, Hazard and MRL Functions*, April 2002

49. Pablo Olivares, *Maximum Likelihood Estimators for a Branching-Diffusion Process*, August 2002

50. Shuanming Li and José Garrido, *On Ruin for the Erlang*(n) *Risk Process*, June 2003

51. G. Jogesh Babu and Yogendra P. Chaubey, *Smooth Estimation of a Distribution and Density function on a Hypercube Using Bernstein Polynomials for Dependent Random Vectors* , August 2003

52. Shuanming Li and José Garrido, *On the Time Value of Ruin for a Sparre Anderson Risk Process Perturbed by Diffusion*, November 2003

53. Yogendra P. Chaubey, Cynthia M. DeSouza and Fassil Nebebe, *Bayesian Inference for Small Area Estimation under the Inverse Gaussian Model via Cibbs Sampling*, December 2003

54. Alexander Melnikov and Victoria Skornyakova, *Pricing of Equity–Linked Life Insurance Contracts with Flexible Guarantees*, May 2004

55. Yi Lu and José Garrido, *Regime–Switching Periodic Models for Claim Counts*, June 2004.

56. I. Urrutia-Romaní, R. Rodríguez-Ramos, J. Bravo-Castillero and R. Guinovart-Díaz, *Asymptotic Homogenization Method Applied to Linear Viscoelastic Composites. Examples*, August 2004.

57. Yi Lu and José Garrido, *Double Periodic Non-Homogeneous Poisson Models for Hurricanes Data*, September 2004.

58. M.I. Beg and M. Ahsanullah, *On Characterizing Distributions by Conditional Expectations of Functions of Generalized Order Statistics*, September, 2004.

59. M.I. Beg and M. Ahsanullah, *Concomitants of Generalized Order Statistics from Farlie-Gumbel-Morgenstern Distributions*, September, 2004.

60. Yogendra P. Chaubey and Debaraj Sen, *An investigation into properties of an estimator of mean of an inverse Gaussian population*, September, 2004.

61. Steven N. Evans and Xiaowen Zhou, *Balls-in-boxes duality for coalescing random walks and coalescing Brownian motions*, September, 2004.

62. Qihe Tang, *Asymptotic ruin probabilities of the renewal model with constant interest force and regular variation*, November, 2004.

63. Xiaowen Zhou, *On a classical risk model with a constant dividend barrier*, November, 2004.

64. K. Balasubramanian and M.I. Beg, *Three isomorphic vector spaces–II*, December, 2004.

Copies of technical reports can be requested from:

Prof. Xiaowen Zhou
Department of Mathematics and Statistics
Concordia University
7141, Sherbrooke Street West
Montréal (QC) H4B 1R6 CANADA