Distribution of the combinatorial multisets component vectors

Eugenijus Manstavičius, Robertas Petuchovas

Faculty of Mathematics and Informatics, Vilnius University
Naugarduko 24, LT-03225 Vilnius
E-mail: eugenijus.manstavicius@mif.vu.lt, robertas.petuchovas@mif.stud.vu.lt

Abstract. We explore a class of random combinatorial structures called weighted multisets. Their components are taken from an initial set satisfying general boundedness conditions posed on the number of elements with a given weight. The component vector of a multiset of weight \( n \) taken with equal probability has dependent coordinates, nevertheless, up to \( r = o(n) \) of them as \( n \to \infty \), we approximate by an appropriate vector comprised from independent negative binomial random variables. The main result is an estimate of the total variation distance. For illustration, we present a central limit theorem for a sequence of additive functions.

Keywords: Random combinatorial multiset, negative binomial distribution, additive function, central limit theorem.

Introduction

We examine weighted combinatorial multisets. They are comprised from components belonging to an initial class \( \mathcal{P} \) of elements having weights in \( \mathbb{N} \). The repetitions are allowed while the order is irrelevant. The weight of a multiset is the sum of weights of its components. The empty multiset has the zero weight.

Let us denote by \( \mathcal{P}_j \subset \mathcal{P} \) the subset of elements of weight \( j \in \mathbb{N} \) and let \( \pi(j) = |\mathcal{P}_j| < \infty \) be its cardinality. For an \( n \in \mathbb{N} \), set \( s := (s_1, \ldots, s_n) \in \mathbb{Z}_+^n \) and \( \ell(s) := s_1 + \cdots + s_n \). Let \( \mathcal{M}_n \) be the class of multisets \( \sigma \) of weight \( n \) and denote by \( k_j(\sigma) \geq 0 \) the number of components of weight \( j \), \( 1 \leq j \leq n \), in \( \sigma \in \mathcal{M}_n \). The vector \( k(\sigma) := (k_1(\sigma), \ldots, k_n(\sigma)) \) is called the component vector of \( \sigma \). Note that \( \ell(k(\sigma)) = n \) if \( \sigma \in \mathcal{M}_n \). All quantitative information about the introduced class of multisets lays in the following formal relation satisfied by the generating function:

\[
1 + \sum_{n=1}^{\infty} |\mathcal{M}_n| x^n = \prod_{j=1}^{\infty} \left( 1 - x^j \right)^{-\pi(j)}.
\]

If the uniform probability measure \( \nu_n \) is introduced in the set \( \mathcal{M}_n \), then the distribution of component vector satisfies the conditioning relation \( \nu_n(k(\sigma) = s) = P(\gamma = s|\ell(\gamma) = n) \), where \( \gamma = (\gamma_1, \ldots, \gamma_n) \), and \( \gamma_j = NB(\pi(j), x^j) \), \( 1 \leq j \leq n \), are mutually independent negative binomial random variables (i.r.vs) defined on some probability space \( (\Omega, \mathcal{F}, P) \) with parameters \( (\pi(j), x) \), where \( 0 < x < 1 \) is arbitrary. An extensive list of instances and the historical survey on investigations of random multisets can be found in [2] and [1]. In the present note, we discuss only the results
Distribution of the combinatorial multisets component vectors

concerning the total variation approximations of the truncated component vectors 
\( k_r(\sigma) = (k_1(\sigma), \ldots, k_r(\sigma)) \) by appropriate vectors with independent coordinates if 
\( r = r(n) \) and \( r = o(n) \) as \( n \to \infty \).

Let \( \rho_{TV} \) denote the total variation distance and \( \mathcal{L}(\cdot) \) be the distribution under the 
relevant probability measure. For brevity, we will use \( \ll \) as an analog of \( O(\cdot) \). As it 
has been proved by D. Stark [7] (see also [1]), the regularity condition \( \pi(j) \sim \theta q^j j^{-1}, \)
\( j \to \infty \), where \( \theta > 0 \) and \( q > 1 \) are constants, and some other extra technical 
requirements imply

\[
\rho_{TV}(\mathcal{L}(k_r(\sigma)), \mathcal{L}(\gamma_r)) \ll (r/n)^\nu. \tag{1}
\]

Here and afterwards \( \gamma_r = (\gamma_1, \ldots, \gamma_r) \) and \( \gamma_j = NB(\pi(j), q^{-j}), 1 \leq j \leq r \leq n \), are 
mutually independent negative binomial r.v.s. The positive quantity \( \nu \) depends on the 
constants in the conditions. A similar problem for the so-called additive arithmetical 
semigroups has been dealt with by J. Knopfmacher and W.-B. Zhang [3]. Putting 
regularity conditions on the number of semigroup elements of a given degree, they 
actually exploited some regularity of the number of prime elements. We generalize 
the estimates obtained in [1] and the most interesting part of that from [3].

In the sequel, the hidden constants, if not indicated otherwise, will depend only 
on \( c_0, c_1 \) and \( q \).

**Theorem 1.** Let the class of multisets be generated by a set \( \mathcal{P} \) such that

\[
c_0 \leq j q^{-j} \pi(j) \leq c_1 \tag{2}
\]

for all \( j \geq 1 \), where \( 0 < c_0 \leq c_1 < \infty \) and \( q > 1 \) are constants. Then there is a 
positive constant \( \nu = \nu(c_0, c_1) \) such that (1) holds for \( 1 \leq r \leq n \).

Theorem 1 will be proved using the analytical method proposed in 2002 by E. Mans-
tavičius [5] and applied by him for other combinatorial structures called assemblies 
(see [4]). In Section 1, we present the main steps of the proof, the detailed exposition 
can be found in our master thesis [6]. In the last section, we prove a central limit theorem for a sequence of additive functions defined on the discussed class of multisets.

1 Sketch of the proof

For \( \bar{s} = (s_1, \ldots, s_n) \in \mathbb{Z}_+^n \), set \( \ell_{ij}(\bar{s}) := (i + 1)s_{i+1} + \cdots + js_j \) if \( 0 \leq i < j \leq n \). 
Moreover, let \( \ell_r(\bar{s}) := \ell_{0r}(\bar{s}), \gamma_r = (\gamma_1, \ldots, \gamma_r) \), where \( 1 \leq r \leq n \) and, as previously, 
\( \gamma_j = NB(\pi(j), q^{-j}), 1 \leq j \leq r \), are independent. We will use the following formula 
(see [1]) for the total variation:

\[
\rho_{TV}(\mathcal{L}(\gamma_r|\ell(\bar{s}) = n), \mathcal{L}(\gamma_r)) = \sum_{m \in \mathbb{Z}_+} P(\ell_r(\bar{s}) = m) \left( 1 - \frac{P(\ell_r(\bar{s}) = n - m)}{P(\ell_r(\bar{s}) = n)} \right). \tag{3}
\]

Here \( x_+ = \max\{0, x\} \) if \( x \in \mathbb{R} \).

Denote

\[
F(w) = \prod_{j=r+1}^{n} (1 - q^{-j}w^j)^{-\pi(j)} =: \sum_{s=0}^{\infty} q^{-s} F_s w^s,
\]

Liet. matem. rink. Proc. LMS, Ser. A, 53, 2012, 66–71.
where \( w \in \mathbb{C}, \ |w| = 1 \). Then

\[
\sum_{m=0}^{\infty} P(\ell_{rn}(\gamma) = m)w^m = \frac{F(w)}{F(1)}
\]

and, by Cauchy’s formula,

\[
P(\ell_{rn}(\gamma) = m) = \frac{1}{2\pi imF(1)} \int_{|w|=1} \frac{F'(w)}{w^m} dw.
\]

Further, let \( F(w) = M(w)H(w) \), where

\[
M(w) := \exp \left\{ \sum_{j=r+1}^{n} \pi(j)q^{-j}w^j \right\}, \quad H(w) := \exp \left\{ \sum_{j=r+1}^{\infty} \sum_{k=2}^{\infty} \pi(j)q^{-jk}w^{jk} \right\}.
\]

Moreover, set

\[
D(w) := \prod_{j=1}^{n} (1 - q^{-j}w^j)^{-\pi(j)} =: \sum_{s=0}^{\infty} q^{-s}D_s w^s,
\]

\[
e_r := \frac{F(1)}{D(1)} = \exp \left\{ - \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \pi(j)q^{-jk/k} \right\}.
\]

Let \( 0 < \alpha < 1, 0 < \delta < 1/2, \) and \( K > r \) be arbitrary parameters to be chosen later and such that \( 1 \leq \delta n < K \leq n \). We set

\[
G_1(w) = \exp \left\{ \alpha \sum_{j=r+1}^{K} \pi(j)q^{j}w^j \right\}, \quad G_2(w) = \exp \left\{ - \alpha \sum_{j=R+1}^{n} \frac{\pi(j)}{q^{j}}w^j \right\},
\]

and \( G_3(w) = M^\alpha(w) - G_1(w) \).

Using introduced functions, we split integral in (4) to obtain

\[
q^{-m}F_m = \frac{1}{2\pi im} \left( \int_{\Delta_0} + \int_{\Delta} \right) \frac{F'(w)(1 - G_2(w))}{w^m} dw + \frac{1}{2\pi im} \int_{|w|=1} \frac{F'(w)G_2(w)}{w^m} dw
\]

\[=: J_0 + J_1 + J_2.\]

Here \( \Delta_0 = \{ w = e^{it}: |t| \leq T \}, \ \Delta = \{ w = e^{it}: T < |t| \leq \pi \} \), and \( T = (\delta n)^{-1} \).

The further steps are based upon a few estimates obtained under condition (2). We use some estimates taken from articles [5] and [4].

**Lemma 1.** We have \( D(1)n^{-1} \ll q^{-n}D_n \ll D(1)n^{-1} \) for all \( n \geq 1 \). Moreover,

\[
\max_{w \in \Delta} |F(w)| \ll \max_{w \in \Delta} |M(w)| \ll e_r D(1)\delta^n,
\]

if \( \delta n \geq 1 \) and \( 0 \leq \alpha \leq \delta n \).

**Proof.** Since \( 1 \ll H(w) \ll 1 \), we can apply Lemmas 2 and 3 in [4].
Lemma 2. Let $0 < \alpha < 1$ be arbitrary and $\bar{\delta}n \geq 1$. Then $J_1 \ll e_r n q^{\bar{\delta}} D_\eta K^{-1} \delta^\alpha (1-\alpha)$ uniformly in $n/2 \leq m \leq n$ and $0 \leq r \leq \bar{\delta}n < K < n$. Here the constant in $\ll$ depends also on $\alpha$.

Proof. Repeat the argument used in [4, Lemma 5].

Lemma 3. If $0 < \alpha < 1$ and $1 \leq \bar{\delta}n < K < n$, then

$$J_2 = \frac{1}{2\pi i m} \int_{|w|=1} \frac{F'(w) G_1(w)}{w^m} dw \ll e_r q^{-n} D_\eta \left( \frac{K}{n} \right)^{\alpha c_0}$$

uniformly in $n/2 \leq m \leq n$.

Proof. Combine $1 \ll H(w) \ll 1$ and Lemma 4 in [4].

Lemma 4. If $T = (\bar{\delta}n)^{-1} \leq 1$, then there exists a constant $c = c(c_0)$ such that

$$q^{-m} F_m = J_0 + O(e_r q^{-n} D_\eta \delta^\epsilon)$$

(5)

uniformly in $0 \leq r \leq \bar{\delta}n$ and $n/2 \leq m \leq n$. Moreover,

$$q^{-n} D_n = \frac{1}{2\pi i m} \int_{\Delta_0} D'(w) (1-G_2(w)) \frac{dw}{w^m} + O(q^{-n} D_\eta \delta^\epsilon).$$

(6)

Proof. To prove (5), use Lemmas 2 and 3. Formula (6) follows from (5) if $r = 0$.

The next claim is crucial in the applied approach. Instead of integrating the remaining integral $J_0$, we change its integrand and return to $D_n$.

Lemma 5. If $0 \leq \eta \leq 1/2$ and $1/n \leq \delta \leq 1/2$ are arbitrary, then

$$J_0 q^{m}(e_r D_n)^{-1} - 1 \ll \eta \delta^{-1} + \delta^\epsilon + (r/n) \{ r \geq 1 \} \delta^{-1-c_3}, \quad c_3 := cc_1/c_0,$$

uniformly in $n(1-\eta) \leq m \leq n$ and $0 \leq r \leq \bar{\delta}n$. Here $c = c(c_0)$ comes from Lemma 4.

Proof. As in the proof of Lemma 7 in [4] approximate the integrand of $J_0$ by $D'(w)(1-G_2(w)) w^{-n}$ and apply (6).

Lemma 6. Assume that parameters $0 \leq r \leq n$, $0 \leq \eta \leq 1/2$ and $1/n \leq \delta \leq 1/2$ are arbitrary. Then there exists positive constants $c = c(c_0)$ and $c_3 = c_3(c_0, c_1)$ such that

$$q^{m-m} F_m (e_r D_n)^{-1} - 1 \ll \eta \delta^{-1} + \delta^\epsilon + r/n \{ r \geq 1 \} \delta^{-1-c_3}$$

uniformly when $0 \leq r \leq \bar{\delta}n$, $n(1-\eta) \leq m \leq n$.

Proof. Applying Lemma 5 for relation (5), we attain lemma’s proof.

Proof of Theorem 1. We have

$$P(\ell_n(\bar{\gamma}) = n - m) = q^{-m} F_{n-m}/(e_r D(1)), \quad P(\ell(\bar{\gamma}) = n) = q^{-n} D_n/D(1).$$

Thus,

$$P(\ell_n(\bar{\gamma}) = n - m) / P(\ell(\bar{\gamma}) = n) = q^m F_{n-m}/(e_r D_n).$$

We apply Lemma 6 with $n - m$ instead of $m$ choosing $\eta = (r/n)^{1/2}$ and $\delta = (r/n)^\varphi$, where $0 < \varphi < \min \{1/2, 1/(1 + c_3) \}$ is a fixed number. So we obtain
Theorem 2. Let the class of multisets \( \mathcal{M}_n \) satisfy condition (2). Assume that \( h_{nj}(k) = \text{o}(1) \) for every fixed \( j, k \in \mathbb{N} \). If conditions (7) and (8) are satisfied for \( a_{nj} := h_{nj}(1) \), then

\[
\nu_n(x) := \nu_n(h_n(\sigma) - b_n < x) = \Phi(x) + \text{o}(1)
\]

uniformly in \( x \in \mathbb{R} \). Conversely, if

\[
P(\ell_{rn}(\bar{\gamma}) = n - m) - 1 \ll (r/n)^{1/2-\nu} + (r/n)^{c_0} + (r/n)^{1-(1+c_3)} \ll (r/n)^{\nu},
\]

where \( \nu = \nu(c_0, c_1) > 0 \), uniformly in \( 0 \leq m \leq \sqrt{r/m} \) and \( 1 \leq r \leq 2^{-1/\nu} : = c_2n \). The summands in (3), if \( m > \sqrt{r/m} \), contribute not more than

\[
(rn)^{-1/2} \mathbb{E} \ell_r(\bar{\gamma}) = (rn)^{-1/2} \sum_{j \leq r} j \mathbb{E} \gamma_j \leq c_1 (1-q^{-1})^{-1} (r/n)^{1/2}.
\]

Consequently, \( \rho_{TV}(\mathcal{L}(h_{\bar{\gamma}}), \mathcal{L}(\bar{\gamma})) \ll (r/n)^{\nu} \) for \( 1 \leq r \leq c_2n \). Seeing that the theorem claim is trivial in the case \( c_2n < r \leq n \), we finish the proof.

2 Central limit theorem

As an application of Theorem 1, we now present an analog of the well-known Feller–Lindeberg theorem. As previously, let condition (2) be satisfied and \( \gamma_j = NB(\pi(j), q^{-j}) \), where \( q > 1 \) and \( 1 \leq j \leq n \), are i.r.vs. Let \( a_{nj} \in \mathbb{R} \), \( X_n = a_{nj} \gamma_j \) if \( 1 \leq j \leq n \), and \( X_n = X_{n1} + \cdots + X_{nn} \). Set \( \Phi(x) \) for the standard normal distribution function, \( u^* := \min \{|u|, 1\} \sgn u \), and

\[
\alpha(y) := \sum_{j \leq y} a_{nj}^2 j, \quad 0 \leq y \leq n.
\]

Assume that \( n \to \infty \) in the limit relations.

Lemma 7. In the notation above, let \( a_{nj} = o(1) \) for each fixed \( j \in \mathbb{N} \). The relation

\[
P(X_n - b_n < x) = \Phi(x) + o(1)
\]

with some \( b_n \in \mathbb{R} \) uniformly in \( x \in \mathbb{R} \) holds if and only if, for every \( \varepsilon > 0 \),

\[
\sum_{j \leq n} \frac{1}{j} \{ |a_{nj}| \geq \varepsilon \} = o(1), \quad \sum_{j \leq n} \frac{a_{nj}^2}{j} \{ |a_{nj}| < 1 \} = 1 + o(1),
\]

and

\[
b_n = \alpha(n) + o(1).
\]

Proof. The i.r.vs \( X_{nj} \), \( 1 \leq j \leq n \), are infinitesimal. Hence the claim is just a special case of the mentioned Feller–Lindeberg theorem.

Let \( h_{nj}(k) \) be a three-dimensional real sequence such that \( h_{nj}(0) \equiv 0 \) for \( j \leq n \). Define the sequence of additive functions \( h_n : \mathbb{M}_n \to \mathbb{R} \) by setting \( h_n(\sigma) = \sum_{j \leq n} h_{nj}(k_{j}(\sigma)) \).

Theorem 2. Let the class of multisets \( \mathcal{M}_n \) satisfy condition (2). Assume that \( h_{nj}(k) = o(1) \) for every fixed \( j, k \in \mathbb{N} \). If conditions (7) and (8) are satisfied for \( a_{nj} := h_{nj}(1) \), then

\[
\nu_n(x) := \nu_n(h_n(\sigma) - b_n < x) = \Phi(x) + o(1)
\]

uniformly in \( x \in \mathbb{R} \). Conversely, if
Distribution of the combinatorial multisets component vectors

\[ \sum_{\delta n < j \leq n} \frac{a_{n,j}^2}{j} = o(1) \]  \hspace{1cm} (10)

for every \( 0 < \delta < 1 \), then convergence (9) with some \( b_n \) implies relations (7) and (8).

**Proof.** We indicate the main steps only. First, we verify that convergence (9) can hold only simultaneously with that for the sequence of functions \( h_n(\sigma) \) defined via \( h_{nj}(k) = kh_{nj}(1) =: ka_{nj} \) for \( 1 \leq j \leq n \). Next, we split the latter into two parts: \( h_{nj}(\sigma) = (\sum_{j \leq r \leq n} a_{nj} h_{nj}(r)) =: h_{nj}^{(r)}(\sigma) + f_n(\sigma) \)

As in [4], one can check that condition (7) yields a sequence \( r = r(n) \to \infty \) such that \( r = o(n) \) and

\[ \nu_n(\lfloor f_n(\sigma) - (\alpha(n) - \alpha(r)) \rfloor \geq \varepsilon) = o(1) \]  \hspace{1cm} (11)

for every \( \varepsilon > 0 \). Moreover, by Theorem 1 and Lemma 7,

\[ \nu_n(h_n^{(r)}(\sigma) - \alpha(r) < x) = P\left( \sum_{j \leq r} X_{nj} - \alpha(r) < x \right) + o(1) = \Phi(x) + o(1) \]

uniformly in \( x \in \mathbb{R} \). The last two relations furnish the proof of the sufficiency part.

In the necessity part, we can again use Theorem 1 and Lemma 7 because of condition (10) also implies (11). So we arrive at the last relation. Consequently, the necessity in Theorem 2 is assured by that in Lemma 7. The theorem is proved.

**References**

[1] R. Arratia, A.D. Barbour and S. Tavaré. *Logarithmic Combinatorial Structures: a Probabilistic Approach*. EMS Monographs in Mathematics, 2003. ISBN 3-03719-000-0.

[2] R. Arratia and S. Tavaré. Independent process approximations for random combinatorial structures. *Adv. Math.*, 104(1):90–154, 1994.

[3] J. Knopfmacher and W.-B. Zhang. *Number Theory Arising from Finite Fields*. Marcel Dekker, 2001. ISBN 0-8247-0577-7.

[4] E. Manstavičius. Total variation approximation for random assemblies and a functional limit theorem. *Monatsh. Math.*, 161:313–334, 2009.

[5] E. Manstavičius. Mappings on decomposable combinatorial structures: analytic approach. *Comb. Probab. Comput.*, 11:61–78, 2002.

[6] R. Petuchovas. Distribution of combinatorial component vectors. Vilnius University database of final theses, 2012.

[7] D. Stark. Total variation asymptotics for independent process approximations of logarithmic multisets and selections. *Rand. Struct. Alg.*, 11:51–80, 1997.

**REZIUMĖ**

**Kombinatorinių multiabilių komponenčių vektorių skirstiniai**

E. Manstavičius, R. Petuchovas

Naigrinėjamos atsitiktinės kombinatorinės struktūros, vadinamos svorinėmis multiabiliomis. Jos sudarytos iš komponenčių, priklausančių aiške \( \mathcal{P} \), kurioje yra \( \pi(j) \) elementų, o pastaroji seka tenkina aprėkutumo sąlygą. Tegu \( \sigma \) yra \( n \) svorio multiiba, paimta su vienoda tikimybė, ir \( k_j(\sigma) - \) svorio \( j \) komponenčių skaičius joje, \( 1 \leq j \leq n \). Apibrėžkite atsitiktinį vektorių \( \mathbf{k}(\sigma) = (k_1(\sigma), \ldots, k_r(\sigma)), 1 \leq r \leq n \), išstiriate jo skirstinio pilnosios variacijos atstumą nuo atitinkamo nepriklausomo koordinacinių vektoriaus. Rezultatas panaudojus adityvųjų funkcijų centrinės ribinės teoremos įrodymo.

**Raktiniai žodžiai:** atsitiktinės kombinatorinės struktūros, svorinės multiabilės, neigiamasis binominis skirstinys.

Liet. matem. rink. *Proc. LMS, Ser. A*, 53, 2012, 66–71.