Lietuvos matematikos rinkinys
Proc. of the Lithuanian Mathematical Society, Ser. A
Vol. **53**, 2012, 66–71

ISSN 0132-2818
www.mii.lt/LMR/

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 $\bar{s} := (s_1, \dots, s_n) \in \mathbb{Z}_+^n$ and $\ell(\bar{s}) := 1s_1 + \dots + ns_n$. Let $\mathcal{M}_n$ be the class of multisets σ of weight n and denote by $k_j(\sigma) \geqslant 0$ the number of components of weight j, $1 \leqslant j \leqslant n$, in $\sigma \in \mathcal{M}_n$. The vector $\bar{k}(\sigma) := \big(k_1(\sigma), \dots, k_n(\sigma)\big)$ is called the *component vector* of σ. Note that $\ell(\bar{k}(\sigma)) = n$ 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 ν_n is introduced in the set $\mathcal{M}_n$, then the distribution of component vector satisfies *the conditioning relation* $\nu_n(\bar{k}(\sigma) = \bar{s}) = P(\bar{\gamma} = \bar{s} | \ell(\bar{\gamma}) = n)$, where $\bar{\gamma} = (\gamma_1, \dots, \gamma_n)$, and $\gamma_j = NB(\pi(j), x^j)$, $1 \leqslant j \leqslant n$, 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 concerning the total variation approximations of the truncated component vectors $\bar{k}_r(\sigma) = (k_1(\sigma), \dots, k_r(\sigma))$ by appropriate vectors with independent coordinates if $r = r(n)$ and $r = o(n)$ as $n \to \infty$.

Let ρ_{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

$$\rho_{TV}\big(\mathcal{L}\big(\bar{k}_r(\sigma)\big), \mathcal{L}(\bar{\gamma}_r)\big) \ll (r/n)^{\upsilon}. \tag{1}$$

Here and afterwards $\bar{\gamma}_r = (\gamma_1, \dots, \gamma_r)$ and $\gamma_j = NB(\pi(j), q^{-j})$, $1 \leqslant j \leqslant r \leqslant n$, are mutually independent negative binomial r.vs. The positive quantity υ 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*

$$c_0 \leqslant jq^{-j}\pi(j) \leqslant c_1 \tag{2}$$

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

Theorem 1 will be proved using the analytical method proposed in 2002 by E. Manstavičius [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 $\bar{s} = (s_1, \dots, s_n) \in \mathbb{Z}_+^n$, set $\ell_{ij}(\bar{s}) := (i+1)s_{i+1} + \dots + js_j$ if $0 \leqslant i < j \leqslant n$. Moreover, let $\ell_r(\bar{s}) := \ell_{0r}(\bar{s})$, $\bar{\gamma}_r = (\gamma_1, \dots, \gamma_r)$, where $1 \leqslant r \leqslant n$ and, as previously, $\gamma_j = NB(\pi(j), q^{-j})$, $1 \leqslant j \leqslant r$, are independent. We will use the following formula (see [1]) for the total variation:

$$\rho_{TV}\big(\mathcal{L}\big(\bar{\gamma}_r | \ell(\bar{\gamma}) = n\big), \mathcal{L}(\bar{\gamma}_r)\big) = \sum_{m \in \mathbb{Z}_+} P\big(\ell_r(\bar{\gamma}) = m\big)\left(1 - \frac{P(\ell_{rn}(\bar{\gamma}) = n - m)}{P(\ell(\bar{\gamma}) = n)}\right)_+. \tag{3}$$

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

Denote

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

where $w \in \mathbb{C}$, $|w| = 1$. Then

$$\sum_{m=0}^{\infty} P\big(\ell_{rn}(\bar{\gamma}) = m\big) w^m = \frac{F(w)}{F(1)}$$

and, by Cauchy's formula,

$$P\big(\ell_{rn}(\bar{\gamma}) = m\big) = \frac{1}{2\pi i m F(1)} \int_{|w|=1} \frac{F'(w)}{w^m}\, dw. \tag{4}$$

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

$$M(w) := \exp\Bigg\{ \sum_{j=r+1}^{n} \pi(j) q^{-j} w^j \Bigg\}, \qquad H(w) := \exp\Bigg\{ \sum_{j=r+1}^{n} \sum_{k=2}^{\infty} \pi(j) q^{-jk} \frac{w^{jk}}{k} \Bigg\},$$

Moreover, set

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

$$e_r := \frac{F(1)}{D(1)} = \exp\Bigg\{ -\sum_{j=1}^{r} \sum_{k=1}^{\infty} \frac{\pi(j)}{q^{jk} k} \Bigg\}.$$

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

$$G_1(w) = \exp\Bigg\{ \alpha \sum_{j=r+1}^{K} \frac{\pi(j)}{q^j} w^j \Bigg\}, \qquad G_2(w) = \exp\Bigg\{ -\alpha \sum_{j=K+1}^{n} \frac{\pi(j)}{q^j} w^j \Bigg\},$$

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

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

$$q^{-m} F_m = \frac{1}{2\pi i m} \bigg(\int_{\triangle_0} + \int_{\triangle} \bigg) \frac{F'(w)(1 - G_2(w))}{w^m}\, dw + \frac{1}{2\pi i m} \int_{|w|=1} \frac{F'(w) G_2(w)}{w^m}\, dw$$
$$=:_0 + J_1 + J_2.$$

Here $\triangle_0 = \{w = e^{it}\colon |t| \leqslant T\}$, $\triangle = \{w = e^{it}\colon T < |t| \leqslant \pi\}$, 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* $n \geqslant 1$*. Moreover,*

$$\max_{w \in \triangle} \big|F(w)\big| \ll \max_{w \in \triangle} \big|M(w)\big| \ll e_r D(1) \delta^{c_0},$$

if $\delta n \geqslant 1$ *and* $0 \leqslant r \leqslant \delta n$.

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* $\delta n\geqslant 1$*. Then* $J_1\ll e_r nq^{-n}D_nK^{-1}\delta^{c_0(1-\alpha)}$ *uniformly in* $n/2\leqslant m\leqslant n$ *and* $0\leqslant r\leqslant \delta n<K<n$*. Here the constant in* $\ll$ *depends also on* α*.*

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

Lemma 3. *If* $0<\alpha<1$ *and* $1\leqslant \delta n<K<n$*, then*

$$J_2=\frac{1}{2\pi i m}\int_{|w|=1}\frac{F'(w)G_2(w)}{w^m}\,dw\ll e_rq^{-n}D_n\left(\frac{K}{n}\right)^{\alpha c_0}$$

uniformly in $n/2\leqslant m\leqslant n$*.*

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

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

$$q^{-m}F_m=J_0+O\bigl(e_rq^{-n}D_n\delta^c\bigr) \tag{5}$$

uniformly in $0\leqslant r\leqslant \delta n$ *and* $n/2\leqslant m\leqslant n$*. Moreover,*

$$q^{-n}D_n=\frac{1}{2\pi i n}\int_{\triangle_0}D'(w)\bigl(1-G_2(w)\bigr)\frac{dw}{w^n}+O\bigl(q^{-n}D_n\delta^c\bigr). \tag{6}$$

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* $0\leqslant\eta\leqslant 1/2$ *and* $1/n\leqslant\delta\leqslant 1/2$ *are arbitrary, then*

$$J_0q^n(e_rD_n)^{-1}-1\ll\eta\delta^{-1}+\delta^c+(r/n)\mathbf{1}\{r\geqslant 1\}\delta^{-1-c_3},\qquad c_3:=cc_1/c_0,$$

uniformly in $n(1-\eta)\leqslant m\leqslant n$ *and* $0\leqslant r\leqslant\delta n$*. 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 $D'(w)\bigl(1-G_2(w)\bigr)w^{-n}$ and apply (6).

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

$$q^{n-m}F_m(e_rD_n)^{-1}-1\ll\eta\delta^{-1}+\delta^c+r/n\mathbf{1}\{r\geqslant 1\}\delta^{-1-c_3}$$

uniformly when $0\leqslant r\leqslant\delta n$*,* $n(1-\eta)\leqslant m\leqslant n$*.*

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

Proof of Theorem 1. We have

$$P\bigl(\ell_{rn}(\bar\gamma)=n-m\bigr)=q^{-(n-m)}F_{n-m}/\bigl(e_rD(1)\bigr),\qquad P\bigl(\ell(\bar\gamma)=n\bigr)=q^{-n}D_n/D(1).$$

Thus,

$$P\bigl(\ell_{rn}(\bar\gamma)=n-m\bigr)/P\bigl(\ell(\bar\gamma)=n\bigr)=q^mF_{n-m}/(e_rD_n).$$

We apply Lemma 6 with $n-m$ instead of m choosing $\eta=(r/n)^{1/2}$ and $\delta=(r/n)^y$, where $0<y<\min\{1/2,1/(1+c_3)\}$ is a fixed number. So we obtain

$$\frac{P(\ell_{rn}(\bar{\gamma}) = n - m)}{P(\ell(\bar{\gamma}) = n)} - 1 \ll (r/n)^{1/2-y} + (r/n)^{cy} + (r/n)^{1-(1+c_3)y} \ll (r/n)^{\upsilon},$$

where $\upsilon = \upsilon(c_0, c_1) > 0$, uniformly in $0 \leqslant m \leqslant \sqrt{rn}$ and $1 \leqslant r \leqslant 2^{-1/y}n =: c_2 n$. The summands in (3), if $m > \sqrt{rn}$, contribute not more than

$$(rn)^{-1/2}\mathbb{E}\ell_r(\bar{\gamma}) = (rn)^{-1/2}\sum_{j\leqslant r} j\mathbb{E}\gamma_j \leqslant c_1\bigl(1-q^{-1}\bigr)^{-1}(r/n)^{1/2}.$$

Consequently, $\rho_{TV}(\mathcal{L}(\bar{k}_r(\sigma)), \mathcal{L}(\bar{\gamma}_r)) \ll (r/n)^{\upsilon}$ for $1 \leqslant r \leqslant c_2 n$. Seeing that the theorem claim is trivial in the case $c_2 n < r \leqslant n$, 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 $1 \leqslant j \leqslant n$, are i.r.vs. Let $a_{nj} \in \mathbb{R}$, $X_{nj} = a_{nj}\gamma_j$ if $1 \leqslant j \leqslant n$, and $X_n = X_{n1} + \cdots + X_{nn}$. Set $\Phi(x)$ for the standard normal distribution function, $u^* := \min\{|u|, 1\}\mathrm{sgn}u$, and

$$\alpha(y) := \sum_{j\leqslant y} \frac{a_{nj}^*}{j}, \quad 0 \leqslant y \leqslant n.$$

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 $\varepsilon > 0$,*

$$\sum_{j\leqslant n} \frac{1}{j}\mathbf{1}\bigl\{|a_{nj}| \geqslant \varepsilon\bigr\} = o(1), \qquad \sum_{j\leqslant n} \frac{a_{nj}^2}{j}\mathbf{1}\bigl\{|a_{nj}| < 1\bigr\} = 1 + o(1), \tag{7}$$

and

$$b_n = \alpha(n) + o(1). \tag{8}$$

Proof. The i.r.vs X_{nj}, $1 \leqslant j \leqslant n$, 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 $j \leqslant n$. Define the sequence of additive functions $h_n : \mathcal{M}_n \to \mathbb{R}$ by setting $h_n(\sigma) = \sum_{j\leqslant n} h_{nj}(k_j(\sigma))$.

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*

$$\nu_n(x) := \nu_n\bigl(h_n(\sigma) - b_n < x\bigr) = \Phi(x) + o(1) \tag{9}$$

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

$$\sum_{\delta n<j\leqslant n}\frac{a_{nj}^{*^2}}{j}=o(1) \tag{10}$$

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 $h_{nj}(k)$ satisfying the condition $h_{nj}(k)=kh_{nj}(1)=: ka_{nj}$ for $1\leqslant j\leqslant n$. Next, we split the latter into two parts: $h_n(\sigma)=(\sum_{j\leqslant r}+\sum_{r<j\leqslant n})a_{nj}k_j(\sigma)=: h_n^{(r)}(\sigma)+f_n(\sigma)$ As in [4], one can check that condition (7) yields a sequence $r=r(n)\to\infty$ such that $r=o(n)$ and

$$\nu_n\bigl(\bigl|f_n(\sigma)-\bigl(\alpha(n)-\alpha(r)\bigr)\bigr|\geqslant\varepsilon\bigr)=o(1) \tag{11}$$

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

$$\nu_n\bigl(h_n^{(r)}(\sigma)-\alpha(r)<x\bigr)=P\biggl(\sum_{j\leqslant r}X_{nj}-\alpha(r)<x\biggr)+o(1)=\Phi(x)+o(1)$$

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 combinatorical 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 multiaibių komponenčių vektorių skirstiniai

E. Manstavičius, R. Petuchovas

Nagrinėjamos atsitiktinės kombinatorinės struktūros, vadinamos svorinėmis multiaibėmis. Jos sudarytos iš komponenčių, priklausančių aibei $\mathcal{P}$, kurioje yra $\pi(j)$ elementų, o pastaroji seka tenkina aprėžtumo sąlygą. Tegul σ yra n svorio multiaibė, paimta su vienoda tikimybe, ir $k_j(\sigma)$ – svorio j komponenčių skaičius joje, čia $1\leqslant j\leqslant n$. Apibrėžę atsitiktinį vektorių $\bar{k}_r(\sigma)=(k_1(\sigma),\dots,k_r(\sigma))$, $1\leqslant r\leqslant n$, ištiriame jo skirstinio pilnosios variacijos atstumą nuo atitinkamo nepriklausomų koordinačių vektoriaus. Rezultatas panaudotas adityviųjų funkcijų centrinės ribinės teoremos įrodyme.

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