PROCEEDINGS OF THE YEREVAN STATE UNIVERSITY

Physical and Mathematical Sciences 2011, № 1, p. 18–22

Mathematics

STABILITY OF FREQUENCY DISTRIBUTION IN FRAME OF NATURAL PARAMETRIZATION. I

E. A. DANIELIAN, S. K. ARZUMANYAN*

Chair of Probability Theory and Mathematical Statistics, YSU

In this paper the stability problem for frequency distribution in frame of natural parameterization is formulated and discussed. The case of finite number of independent parameters is characterized. A corresponding stability problem is investigated in terms of l_p-metric.

Keywords: frequency distribution, l_p-metric, stability by parameters.

Introduction. The sequence $\{p_n\}_0^\infty$ forms a frequency distribution (FD) if $p_n>0,\ n\ge 0$ and $\sum p_n=1$. In the bioinformatics (see [1]):

$$p_n=n^{-\rho}L(n),\ \ n\ge 1,\ \ 1<\rho<+\infty,\ \lim_{n\to\infty}L(n)=L\in R^+=(0,+\infty),$$
$$\frac{L(n)}{L(n-1)}=1+o\left(\frac{1}{n}\right),\ \ n\to+\infty, \tag{1}$$

$$p_n>p_{n+1},\ \frac{p_n}{p_{n+1}}>\frac{p_{n+1}}{p_{n+2}}\ \text{ starting from some } n_0\ge 0\,. \tag{2}$$

Assume that in (1) $\rho\in(2,+\infty)$ and in (2) $n_0=0$.

The unknown FDs are approximated by various parametric distributions $\{p_n(\vec c)\}_0^\infty$ with the vector $\vec c$ of parameters, that are referred to also as FDs.

Let $\vec c=\vec c_m=(c_1,\dots,c_m)\in\Omega,\ m<+\infty$, and $K\subseteq\Omega$ be a *bounded, closed, convex* set, and μ be some metric in the set $\left\{\{p_n(\vec c_m)\}_0^\infty:\vec c_m\in\Omega\right\}$.

We say that *m*-parametric FD $\{p_n(\vec c_m)\}_0^\infty$ with independent parameters is μ-stable (with respect to the parameters) on K, if uniformly on $\vec c_m,\vec c\,'_m\in K$

$$\lim_{|\vec c_m-\vec c\,'_m|\to 0}\mu\left(\{p_n\}_0^\infty,\{p'_n\}_0^\infty\right)=0\ . \tag{3}$$

The parameters are *independent*, if not one of these is a function of others. Here $|\vec c_m-\vec c\,'_m|=\sum_{i=1}^m|c_i-c'_i|,\ \vec c_m=(c_1,\dots,c_m),\vec c\,'_m=(c'_1,\dots,c'_m),\ p_n=p_n(\vec c_m),\ \ p'_n=p_n(\vec c\,'_m),\ n\ge 0.$

* E-mail: stepan.arzumanyan@anelik.am

Similarly, $\{p_n(\vec{c}_m)\}_0^\infty$ is μ-stable, if (3) holds for any K.

Stability Problem. Introduce the metric μ and find conditions for μ-stability of FD $\{p_n(\vec{c}_m)\}_0^\infty$.

In this paper the stability problem for FD $\{p_n\}_0^\infty$ in bioinformatics is formulated in frame of *natural parameterization* (NP) in case of finite number of independent parameters and is solved in terms of l_p-metric, $p>0$.

Natural Parameterization (NP). Due to [2], $\{p_n\}_0^\infty$ is a FD, iff

$$p_n = p_0\prod_{k=1}^{n}\varepsilon_k,\ n>1,\ \varepsilon_n>0,\ \ p_0=\left(1+\sum_{n\geq 1}\prod_{k=1}^{n}\varepsilon_k\right)^{-1}, \tag{4}$$

$$\sum_{n\geq 1}\prod_{k=1}^{n}\varepsilon_k < +\infty . \tag{5}$$

Then $\varepsilon_n=(p_n/p_{n-1}),\ n\geq 1$. The coefficients $\varepsilon_1,\varepsilon_2,...$ are treated as parameters of $\{p_n\}_0^\infty$ in its NP (4)–(5). Let $T=\{1,2,...\}=\bigcup_{j=1}^{m}T_j,\ T_j=\{k_j,k_j+1,...,k_{j+1}-1\}$, $j=\overline{1,m}$, where $1=k_1<k_2<...<k_m<+\infty(=k_{m+1})$, and any parameter in the set $G_j=\{\varepsilon_k : k\in T_j\}$ is *uniquely* determined by the vector $(c_1,...,c_j)$. Here $c_i=\varepsilon_{k_i}$, $i=\overline{1,j}$. The parameters $c_1,...,c_m$ are *independent.* This is a characterization of m-parametric FD $\{p_n\}_0^\infty$ in its NP (4)–(5).

So, $\{p_n\}_0^\infty=\{p_n(\vec{c}_m)\}_0^\infty$ and $\varepsilon_k=h_k(c_1,...,c_j)$ for given $j,\ j=\overline{1,m}$, and $\varepsilon_k\in G_j$. It is clear that $h_{k_j}(c_1,...,c_j)=c_j$ for $j=\overline{1,m}$. Assume that:

a) h_k for $k\neq k_j, j=\overline{1,m}$, is increased by each parameter separately;

b) partial derivatives $(\partial h_k/\partial c_i)$ for $i=\overline{1,j}$ exist and are uniformly bounded with respect to $k\geq 1$.

Now recall the form of l_p-metric: $l_p\left(\{p_n\}_0^\infty,\{p'_n\}_0^\infty\right)=\sum_{n\geq o}|p_n-p'_n|^p$, and formulate in this case the *Stability Problem*: for admissible $p>0$ prove the l_p-stability of FD $\{p_n\}_0^\infty$.

Theorem 1. In our case the FD $\{p_n\}_0^\infty$ is $l_{1/2}$-stable.

For given $K\subseteq\Omega$ and $j=1,m$ denote

$$\underline{c}_j=\{\inf c_j : \vec{c}_m\in K\},\quad \overline{c}_j=\{\sup c_j : \vec{c}_m\in K\}. \tag{6}$$

It is easy to see that for given $j,\ j=\overline{1,m}$, there is $\vec{c}_m\in K$ with $c_j=\overline{c}_j$. Indeed, if for all $\vec{c}_m\in K$ the components c_j are identical, then the statement is obvious. Assume that there are $\vec{c}_m,\vec{c}'_m\in K$ with $c_j<c'_j$ for given $j,\ j=\overline{1,m}$. Due to the

convexity of K, for any $c \in [c_j, c'_j]$ there is $\vec{c}''_m \in K$ such that $c''_j = c$. In this case, as it follows from the definition of $\overline{c}_j$ for given j, $j = \overline{1,m}$, there is a sequence $\left\{\vec{c}_m^{(k)}\right\}_1^\infty \in K$ such that $\lim\limits_{k\to+\infty} c_j^{(k)} = \overline{c}_j$. Extracting the convergent subsequence $\left\{\vec{c}_m^{(k_s)}\right\}_1^\infty \in K$ from $\left\{\vec{c}_m^{(k)}\right\}_1^\infty$, the statement is proved due to the closeness of K.

Similarly, for given j, $j = \overline{1,m}$, there is $\vec{c}_m \in K$ with $c_j = \underline{c}_j$.

Let us show that in (6)

$$\underline{c}_1 < \underline{c}_2 < \ldots < \underline{c}_m \quad \text{and} \quad \overline{c}_1 < \overline{c}_2 < \ldots < \overline{c}_m . \tag{7}$$

In order to prove the second chain of inequalities (7) assume the opposite, i.e. there are indices i and j, $i < j$, such that $\overline{c}_i \ge \overline{c}_j$. There are $\vec{c}_m, \vec{c}'_m \in K$ with $c_i = \overline{c}_i$ and $c'_j = \overline{c}_j$. Because of (2) we have $\varepsilon_s = (p_s / p_{s-1}) < (p_{s+1} / p_s)$ for $s \ge 1$, which implies that $\overline{c}_i < c_j$ and $c'_i < \overline{c}_j$ in $\vec{c}_m$ and $\vec{c}'_m$ respectively. So, we get that in $\vec{c}_m$ the component c_j exceeds the component $c'_j = \overline{c}_j$ in $\vec{c}'_m$. This contradicts to (6).

The first chain of inequalities (7) is proved similarly.

Now, given $K \subseteq \Omega$ introduce a *bounded, closed, convex* set $K^* \subseteq \Omega$ that contains K, satisfies the conditions (6), where K is replaced by K^* and

$$\vec{c}_{m*} = (\underline{c}_1, \ldots, \underline{c}_m) \in K^*, \quad \vec{c}_m^{\,*} = (\overline{c}_1, \ldots, \overline{c}_m) \in K^*. \tag{8}$$

For $K^* \subseteq \Omega$ denote: $\rho(\vec{c}_m^{\,*})$ is the parameter ρ in the presentation (1) of $\left\{\rho(\vec{c}_m^{\,*})\right\}_0^\infty$.

Theorem 1 follows from the next

Theorem 2. In our case the FD $\{p_n\}_0^\infty$ is l_p-stable on K^* with any

$$p > (1/\rho(\vec{c}_m^{\,*})). \tag{9}$$

Indeed, the l_p-stability on K^* implies the l_p-stability on K, which generates K^* with the same *p* (see (9)). So, in particular, $\{p_n\}_0^\infty$ is $l_{1/2}$-stable on *any* $K \subseteq \Omega$.

Auxiliary Statements. Rewrite (4) in the form

$$p_n(\vec{c}_m) = \frac{g_n(\vec{c}_m)}{g(\vec{c}_m)}, \quad n \ge 0, \quad g(\vec{c}_m) > 0, \tag{10}$$

$$g_n(\vec{c}_m) = \prod_{k=1}^{n} \varepsilon_k, \; n \ge 0, \quad g(\vec{c}_m) = \frac{1}{p_0(\vec{c}_m)} = 1 + \sum_{n\ge1} \prod_{k=1}^{n} \varepsilon_k \quad \left(\prod_{k=0}^{1} \equiv 1\right). \tag{11}$$

Let K^* be generated by K. Since $\vec{c}_{m*} = (\underline{c}_1, \ldots, \underline{c}_m) \in K^*$, $\vec{c}_m^{\,*} = (\overline{c}_1, \ldots, \overline{c}_m) \in K^*$, where $\underline{c}_j$ and $\overline{c}_j$ for $j = \overline{1,m}$ are defined in (6) and $\underline{c}_1 < \underline{c}_2 < \ldots < \underline{c}_m$, $\overline{c}_1 < \overline{c}_2 < \ldots < \overline{c}_m$ (see (7)), therefore, due to condition (a),

$$g_n(\vec{c}_m^{\,*}) = \max_{\vec{c}_m \in K} g_n(\vec{c}_m) \quad \text{for all} \quad n \ge 0, \tag{12}$$

$$g_n(\vec{c}_{m*}) = \min_{\vec{c}_m \in K^*} g(\vec{c}_m). \tag{13}$$

Lemma 1. The function $g(\vec{c}_m)$ is continuous on K^*.

Proof. The conditions (a) and (b) on K^* imply the existence of constant $A \in [1,+\infty)$ (depending only on K^*) such that for any $j=\overline{1,m}$, $k \in T_j$ and $\vec{c}_m \in K^*$

$$0 \le \frac{\partial h_k(c_1,...,c_j)}{\partial c_i} \le A, \quad i=\overline{1,j}\,. \tag{14}$$

With the help of (14) and *Mean Value Theorem* we obtain: for given $j=\overline{1,m}$, $k \in T_j$ and any $\vec{c}_m, \vec{c}\,'_m \in K$

$$\begin{gathered}\left|\varepsilon_k - \varepsilon'_k\right| = \left|h_k(c_1,...,c_j) - h_k(c'_1,...,c'j)\right| \le \\ \le \sum_{i=1}^{j}\left|h_k(c_1,...,c_i,c'_{i+1},...,c'_j) - h_k(c_1,...,c_{i-1},c'_i,...,c'_j)\right| \le A\sum_{i=1}^{j}\left|c_i - c'_i\right| \le A\left|\vec{c}_m - \vec{c}\,'_m\right|,\end{gathered} \tag{15}$$

where $\varepsilon_k = h_k(c_1,...,c_j)$, $\varepsilon'_k = h_k(c'_1,...,c'_j)$. For given $j=\overline{1,m}$ and $k \in T_j$ denote $r_k(\overline{c}_1,...\overline{c}_j) = \left(\prod_{i=1}^{j-1}\prod_{s \in T_i} h_s(\overline{c}_1,...,\overline{c}_i)\right)\prod_{s=k_j}^{k} h_s(\overline{c}_1,...,\overline{c}_j)$, and continue the estimations using (15)

$$\begin{gathered}\left|\prod_{s=1}^{k}\varepsilon_s - \prod_{s=1}^{k}\varepsilon'_s\right| \le \sum_{i=1}^{k}\left|\prod_{s=1}^{i}\varepsilon_s\prod_{s=i=1}^{k}\varepsilon'_s - \prod_{s=1}^{i-1}\varepsilon_s\prod_{s=i}^{k}\varepsilon'_s\right| \le \\ \le \sum_{i=1}^{k}\left(\prod_{s=1}^{i-1}\varepsilon_s\prod_{s=i+1}^{k}\varepsilon'_s\right)\left|\varepsilon_i - \varepsilon'_i\right| \le \frac{r_k(\vec{c}_1,...,\vec{c}_j)}{\underline{c}_1}\sum_{i=1}^{k}\left|\varepsilon_i - \varepsilon'_i\right| \le \frac{r_k(\vec{c}_1,...,\vec{c}_j)}{\underline{c}_1}A\left|\vec{c}_m - \vec{c}\,'_m\right|k\,.\end{gathered}$$

Thus, due to (11), for any $k \ge 1$ on K^* we have

$$\left|g_k(\vec{c}_m) - g_k(\vec{c}\,'_m)\right| = \frac{A}{\underline{c}_1}kg_k(\vec{c}^{\,*}_m)\left|\vec{c}_m - \vec{c}\,'_m\right|\,. \tag{16}$$

Since (1) holds with $\rho \in (2,+\infty)$, therefore,

$$B = \sum_{k \ge 1} k\, g_k(\vec{c}^{\,*}_m) < +\infty\,. \tag{17}$$

Taking into account (11), (16) and (17), for $\vec{c}_m, \vec{c}\,'_m \in K^*$ we obtain an inequality

$$\left|g(\vec{c}_m) - g(\vec{c}\,'_m)\right| = \left|\sum_{k \ge 1}\left(g_k(\vec{c}_m) - g_k(\vec{c}\,'_m)\right)\right| \le \frac{A}{\underline{c}_1}\left|\vec{c}_m - \vec{c}\,'_m\right|\sum_{k \ge 1} k\, g_k(\vec{c}_m) = D\left|\vec{c}_m - \vec{c}\,'_m\right|, \tag{18}$$

where the constant $D = (AB/\underline{c}_1) \in R^+$ depends only on K^*.

With the help of inequality (18) the continuity of $g(\vec{c}_m)$ on K^* is proved.

During the proof of Lemma 1 the following statement was established (see (16)).

Lemma 2. The functions $g(\vec{c}_m)$ for all $n \ge 0$ are continuous on K^*.

Stability Criterion. Let the FD $\{p_n(\vec{c}_m)\}_0^\infty$, where $\vec{c}_m = (c_1,...,c_m) \in \Omega$, is a vector of independent parameters that satisfies conditions (1), (2), and has the form (10). In [3] under the following additional conditions on $K \in \Omega$:

1. There is $\vec{c}_m^{\,+} \in K$ such that $g_n(\vec{c}_m^{\,+}) = \max\limits_{\vec{c}_m \in K} g_n(\vec{c}_m)$ for all $n \geq 0$;

2. $g(\vec{c}_m)$ is continuous with respect to $\vec{c}_m \in K$,

it was established

Criterion. FD $\{p_n(\vec{c}_m)\}_0^\infty$ is l_p-stable on K, iff uniformly on $\vec{c}_m, \vec{c}\,'_m \in K$

$$\lim_{|\vec{c}_m - \vec{c}\,'_m| \to 0} |g_n(\vec{c}_m) - g_n(\vec{c}\,'_m)| = 0 \quad \text{separately for} \quad n \geq 0. \tag{19}$$

Here $p > \left(1/\rho(\vec{c}_m^{\,+})\right)$ and $\rho(\vec{c}_m^{\,+})$ is the parameter ρ in (1) for $\left\{p_n(\vec{c}_m^{\,+})\right\}_0^\infty$.

In our case, when $g_n(\vec{c}_m)$, $n \geq 1$, and $g(\vec{c}_m)$ have the form (11), the Condition 1 is fulfilled on K^* with $\vec{c}_m^{\,+} = \vec{c}_m^{\,*}$, the fulfillment Condition 2 on K^* follows from Lemma 1. Since, due to Lemma 2, $g_n(\vec{c}_m)$ for $n \geq 1$ are continuous on K^*, therefore, they are *uniformly continuous* on K^*, which implies (19) on K^* in our case. Thus, applying the Criterion in our case, one obtains Theorem 2.

Received 10.01.2011

REFERENCES

1. **Astola J., Danielian E.** Frequency Distributions in Biomolecular Systems and Growing Networks. Tampere: TICSP, Series 31, 2006.
2. **Danielian E.A., Avagyan G.P.** Doklady NAN Armenii, 2008, v. 4, № 4, p. 17–23 (in Russian).
3. **Jakovlev S.P.** Modelirovanie, Optimizatsia, Upravlenie. Yer.: SEUA, 2008, v. 11, № 1, p. 139–144 (in Russian).