POISSONIZATION PRINCIPLE FOR SOME CLASSES OF ADDITIVE STATISTICS

Igor Borisov, Maman Jetpisbaev

Sobolev Institute of Mathematics
Novosibirsk State University

Let $\{X_i^{(k)},\, i \geq 1\}$, $k = \overline{1, m}$, be a finite set of independent copies of a sequence of i.i.d. r.v-s in an arbitrary measurable space $(\mathfrak{X}, \mathcal{A})$ and distribution P. For any natural $n_1, \ldots, n_m$, consider m independent empirical point processes based on respective samples $X_1^{(k)}, ..., X_{n_k}^{(k)}$, $k = \overline{1, m}$:

$$V_{n_k}^{(k)}(A) := \sum_{i=1}^{n_k} I_A(X_i^{(k)}), \quad k = \overline{1, m}, \quad A \in \mathcal{A}.$$

Define also m independent accompanying Poisson point processes

$$\Pi_{n_k}^{(k)}(A) := V_{\pi_k(n_k)}(A), \quad k = \overline{1, m}, \quad A \in \mathcal{A},$$

where $\pi_k(t)$, $k = \overline{1, m}$, are independent standard Poisson processes on the positive half-line, which do not depend on all sequences $\{X_i^{(k)};\, i \geq 1\}$, $k = \overline{1, m}$.

We consider the point processes $V_{n_k}(\cdot)$ and $\Pi_{n_k}(\cdot)$ as stochastic processes with trajectories from the measurable space $(\mathbb{B}^{\mathcal{A}}, \mathcal{C})$ of all bounded functions indexed by the elements of the set $\mathcal{A}$, with the σ-algebra $\mathcal{C}$ of all cylindrical subsets of the space $\mathbb{B}^{\mathcal{A}}$. The distributions of stochastic processes $V_{n_k}(\cdot)$ and $\Pi_{n_k}(\cdot)$ on $\mathcal{C}$ are defined in a standard way.

Now, introduce the vector-valued empirical and accompanying Poisson point processes

$$\overline{V}_{\bar{n}}(A) := (V_{n_1}^{(1)}(A), ..., V_{n_m}^{(m)}(A)) \equiv \overline{V}_{\bar{n}},$$

$$\overline{\Pi}_{\bar{n}}(A) := (\Pi_{n_1}^{(1)}(A), ..., \Pi_{n_m}^{(m)}(A)) \equiv \overline{\Pi}_{\bar{n}},$$

where $\bar{n} = (n_1, n_2, ..., n_m)$. The vector-valued point processes $\overline{V}_{\bar{n}}$ and $\overline{\Pi}_{\bar{n}}$ are considered as random elements with values in the measurable space $((\mathbb{B}^{\mathcal{A}})^m, \mathcal{C}^m)$.

ADDITIVE STATISTICS

Let measurable sets $\Delta_1, \Delta_2, ...$ form a finite or countable partition of the sample space under the condition $p_i := P(\Delta_i) > 0$ for all i. Without loss of generality, we can assume that the sequence $\{p_i\}$ is monotonically nonincreasing. Denote by $\nu^{(k)}_{n_k 1}, \nu^{(k)}_{n_k 2}, \ldots$, $k = \overline{1, m}$, the corresponding group frequencies defined by the sample $X^{(k)}_1, \ldots, X^{(k)}_{n_k}$. Put

$$\bar{\nu}_{i\bar{n}} := \overline{V}_{\bar{n}}(\Delta_i) = \left(\nu^{(1)}_{n_1 i}, \ldots, \nu^{(m)}_{n_m i}\right), \;\; i = 1, 2, \ldots .$$

Consider a class of additive functionals of the form

$$\Phi_f(\overline{V}_{\bar{n}}) := \sum_{i \geq 1} f_{i\bar{n}}\left(\bar{\nu}_{i\bar{n}}\right), \tag{1}$$

where $f \equiv \{f_{i\bar{n}}\}$ is an array of arbitrary finite functions defined on $\mathbb{Z}_+^m$ under the condition

$$\sum_{i \geq 1} |f_{i\bar{n}}(0, \ldots, 0)| < \infty \ \ \forall n, \tag{2}$$

which ensures the correct definition of the functional $\Phi_f(\overline{V}_{\bar{n}})$ in the case of a countable partition of the sample space, since the sum under consideration contains only a finite set of nonzero random vectors $\bar{\nu}_{i\bar{n}}$. In the case of a finite partition and $m = 1$ additive functionals of the form (1) were considered by Yu.I. Medvedev in 1970-1977.

Examples.

1) Consider a finite partition $\{\Delta_i;\ i = 1, \dots, N\}$ of the sample space. Put $f_{i\bar{n}}(\bar{x}) := \frac{|\bar{x} - \bar{n}p_i|^2}{|\bar{n}p_i|}$, $i = 1, \dots, N$, where $|\cdot|$ is the standard Euclidean norm in $\mathbb{R}^m$. Then the functional

$$\Phi_{\chi^2}(\overline{V}_{\bar{n}}) := \sum_{i=1}^{N} \frac{|\bar{\nu}_{i\bar{n}} - \bar{n}p_i|^2}{|\bar{n}p_i|} \tag{3}$$

is an m-variate version of a well-known χ^2-statistic. Note that, in the present paper, we are primarily interested in the case where $N \equiv N(\bar{n}) \to \infty$ as $\bar{n} \to \infty$ (i.e., $n_k \to \infty\ \forall k \le m$).

2) Let now the sizes of all m samples be equal: $n_j = n$ $\forall j = 1, \dots, m$. In an equivalent reformulation of the original problem, we consider a sample of m-dimensional observations $\{(X_i^1, \dots, X_i^m);\ i \le n\}$ under the main hypothesis that the sample vector coordinates are independent and have the same N-atomic distribution with unknown masses $p_1, \dots, p_N$. In this case, the log-likelihood function can be represented as the additive functional

$$\Phi_{\log}(\overline{V}_{\bar{n}}) := \sum_{i=1}^{N} (\bar{\nu}_{i\bar{n}}, \bar{1}) \log p_i,$$

where $\bar{1}$ is the unit vector in $\mathbb{R}^m$ and $(\cdot, \cdot)$ is the Euclidean inner product.

3) Consider a finite or countable partition $\{\Delta_i;\ i \geq 1\}$. Let $f_{i\bar{n}}(\bar{x}) \equiv f(\bar{x}) := I_B(\bar{x})$ be the indicator function of some subset $B \subset \mathbb{Z}_+^m$. Then the functional

$$\Phi_{I_B}(\overline{V}_{\bar{n}}) := \sum_{i \geq 1} I_B(\bar{\nu}_{i\bar{n}}) \tag{4}$$

counts the number of partition elements (cells) containing any number of vector sample observations from the range B in a polynomial scheme (finite or infinite) of placing particles into cells Note that in the case of an infinite polynomial scheme in (4), it is additionally assumed that $0 \notin B$. The case of infinite polynomial scheme for $m = 1$ in (4) was studied by R.Bahadur (1960); S.Karlin (1967); D.Darling (1967); V.Kolchin, B.Sevastyanov, and V.Chistyakov (1976); A.Barbour, A.Gnedin (2009); A.Kovalevsky, M.Chebunin (2016), and others.

4) In the case $m = 1$, consider the joint distribution of the r.v-s

$$\Phi_{I_B}(V_{n_1}), \Phi_{I_B}(V_{n_1+n_2}), \ldots, \Phi_{I_B}(V_{n_1+\ldots+n_m})$$

defined in (4) by the sample $(X_1, \ldots, X_N)$, with $N = n_1 + \ldots + n_m$. It is clear that proving the multivariate CLT for this joint distribution, we study the limit behavior of the linear combinations

$$a_1\Phi_{I_B}(V_{n_1}) + a_2\Phi_{I_B}(V_{n_1+n_2}) + \ldots + a_m\Phi_{I_B}(V_{n_1+\ldots+n_m})$$

for almost all vectors $(a_1, \ldots, a_m)$. It is easy to see that

$$V_{n_1+\ldots+n_j} = V_{n_1}^{(1)} + \ldots + V_{n_j}^{(j)} \quad \forall j \leq m,$$

where the EPP $V_{n_1}^{(1)}, \ldots, V_{n_j}^{(j)}$ are defined by the independent subsamples $(X_1, \ldots, X_{n_1})$, $(X_{n_1+1}, \ldots, X_{n_1+n_2})$,…, $(X_{N-n_m+1}, \ldots, X_N)$.

So, in this case, we deal with a functional of the form (1) defined by m independent empirical point processes corresponding to the m independent subsamples $(X_1, \ldots, X_{n_1})$, $(X_{n_1+1}, \ldots, X_{n_1+n_2})$,. . . , $(X_{N-n_m+1}, \ldots, X_N)$, and with the array of functions of m variables

$$f_{i\bar{n}}(\bar{x}) := a_1 I_B(x_1) + a_2 I_B(x_1 + x_2) + \ldots + a_m I_B(x_1 + \ldots + x_m), \quad (5)$$

where $\bar{x} := (x_1, ..., x_m)$.

5) Consider the stochastic process $\{\Phi_{I_B}(\overline{V}_{\bar{n}});\ B \subset \mathbb{Z}_+^m\}$ indexed by all subsets of $\mathbb{Z}_+^m$. As was noted above, studying the asymptotic behavior of the joint distributions of this process can be reduced to studying the asymptotic behavior of the distributions of any linear combinations of corresponding one-dimensional projections of this process, i.e., to studying the asymptotic behavior of the distributions of functionals of the form (1) for $m = 1$ and arrays of functions

$$f_{i\bar{n}}(x) \equiv f(x) := a_1 I_{B_1}(x) + a_2 I_{B_2}(x) + \ldots + a_r I_{B_r}(x) \qquad (6)$$

for almost all vectors $(a_1, \ldots, a_r)$. For one-point sets, the asymptotic analysis of the above-mentioned joint distributions can be found, for example, by S.Karlin (1967); D.Darling (1967); V.Kolchin, B.Sevastyanov, and V.Chistyakov (1976); Barbour and Gnedin (2009).

POISSONIZATION

The Poissonian version of functional (1) under condition (2) is as follows:

$$\Phi_f(\Pi_{\bar{n}}) := \sum_{i \geq 1} f_{i\bar{n}} \left(\bar{\pi}_{i\bar{n}}\right), \tag{7}$$

where $\bar{\pi}_{i\bar{n}} = \left(\pi^{(1)}_{n_1 i}, ..., \pi^{(m)}_{n_m i}\right)$, $\pi^{(k)}_{n_k i} := \Pi_{n_k}(\Delta_i)$, $i \geq 1$, is a sequence of independent Poisson random variables with respective parameters $n_k p_i$.

In addition, one can also indicate a third class of additive functionals (under condition (2) that has the same property:

$$\Phi_f^* := \sum_{i \geq 1} f_{i\bar{n}}\left(\bar{\nu}_{i\bar{n}}^*\right),$$

where $\{\bar{\nu}_{i\bar{n}}^*,\ i \geq 1\}$ is a sequence of independent random vectors such that $\mathcal{L}(\bar{\nu}_{i\bar{n}}^*) = \mathcal{L}(\bar{\nu}_{i\bar{n}})$ for all i. The functional Φ_f^* is well defined due to the Borel–Cantelli lemma and the simple estimate $\mathbf{P}(\bar{\nu}_{i\bar{n}}^* \neq 0) = \mathbf{P}(\bar{\nu}_{i\bar{n}} \neq 0) \leq m\|\bar{n}\|p_i$, where $\|\bar{n}\| := \max_{j \leq m} n_j$.

MAIN RESULT (POISSONIZATION PRINCIPLE)

Let us agree that the symbol "$\Longrightarrow$" in what follows will denote the weak convergence of distributions. The main result of the paper is as follows.

Theorem 3. *Let $f_{i\bar{n}}(\pi_{i\bar{n}})D_{\bar{n}} \xrightarrow{p} 0$ as $\bar{n} \to \infty$ for every fixed i. Then the following three asymptotic relations are equivalent*:

$$1)\ \mathcal{L}\left(\Phi_f(\overline{V}_{\bar{n}})D_{\bar{n}} - M_{\bar{n}}\right) \Longrightarrow \mathcal{L}(\gamma) \quad as\ \ \bar{n} \to \infty,$$

$$2)\ \mathcal{L}\left(\Phi_f(\overline{\Pi}_{\bar{n}})D_{\bar{n}} - M_{\bar{n}}\right) \Longrightarrow \mathcal{L}(\gamma) \quad as\ \ \bar{n} \to \infty,$$

$$3)\ \mathcal{L}\left(\Phi_f^* D_{\bar{n}} - M_{\bar{n}}\right) \Longrightarrow \mathcal{L}(\gamma) \quad as\ \ \bar{n} \to \infty,$$

where $M_{\bar{n}}$ and $D_{\bar{n}}$ are some scalar arrays and γ is some random variable.

APPLICATIONS

First, we note one useful property of the expectations of the functionals under consideration as functions of $\bar{n}$.

Lemma 2. *Let* $\max_{\bar{n}} \sup_{\bar{x}} |f_{i\bar{n}}(\bar{x})| \le C_i$, $\sum\limits_{i\ge 1} C_i p_i < \infty$, *and let*

$$\sum_{i\ge 1} \mathbf{E}|f_{i\bar{n}}\left(\bar{\pi}_{i\bar{n}}\right)| < \infty \ \ \forall \bar{n}. \tag{8}$$

Then the relation $\lim\limits_{\bar{n}\to\infty} |\mathbf{E}\Phi_f(\overline{V}_{\bar{n}})| = \infty$ *is equivalent to the similar relation* $\lim\limits_{\bar{n}\to\infty} |\mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}})| = \infty$. *In the case of infinite limit,*

$$\mathbf{E}\Phi_f(\overline{V}_{\bar{n}}) \sim \mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}})$$

as $\bar{n} \to \infty$.

R e m a r k 3. For functionals of the form (4) in an infinite polynomial scheme, the conditions of Lemma 2 are typical. Let $m = 1$ and $B := \{j : j > k\}$ for any $k \geq 0$. Then

$$\lim_{n\to\infty} \mathbf{E}\Phi_f(V_n) = \lim_{n\to\infty} \sum_{i\geq 1} \mathbf{P}(\nu_{in} > k) = \infty,$$

since, by virtue of the law of large numbers, $\lim\limits_{n\to\infty} \mathbf{P}(\nu_{in} > k) \to 1$ holds for every fixed i. Moreover, in the case under consideration, obviously, $\mathbf{E}\Phi_f(V_n) \leq n$. Similarly, without any restrictions on the probabilities $\{p_i\}$, the infinite limits in Lemma 2 for functionals of the form (4) also hold for the set B consisting of all odd natural numbers. Here the limit relation
$\lim\limits_{n\to\infty} \mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}}) \equiv \lim\limits_{n\to\infty} \sum\limits_{i\ \geq 1} \mathbf{P}(\pi_{in} \in B) = \infty$ follows immediately from the equality $\mathbf{P}(\pi_{in} \in B) = \frac{1}{2}(1 - e^{-2np_i})$.

It is also worth noting that for some sets B the main contribution to the limit behavior of the series $\sum\limits_{i \geq 1} \mathbf{P}(\pi_{in} \in B)$ can be made not only by their initial segments but also tails. For example, this will be the case for any one-point sets $B_k := \{k\}$ for $k > 0$ if the group probabilities are given as $p_i = Ci^{-1-b}$ or $p_i = ce^{-C_o i^\alpha}$ for some constants $c, C, C_o, b > 0$ and $\alpha \in (0, 1)$. In this case, for any subset B of natural numbers in the definition of the functionals (4), the expectation limits indicated in Lemma 2 will be infinite. On the other hand, if $p_i = ce^{-C_o i}$, then for any one-point set the expectations mentioned will be bounded uniformly in n.

Now we present one of the corollaries of Theorem 3, the law of large numbers for the additive functionals under consideration, setting in this theorem $D_{\bar{n}} := (\mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}}))^{-1}$, $M_{\bar{n}} := 0$ and $\gamma := 1$.

Corollary 2. *Let the conditions of Lemma* 2 *be fulfilled. If* $|\mathbf{E}\Phi_f(\Pi_{\bar{n}})| \to \infty$ *as* $\bar{n} \to \infty$ *then the following criterion holds*:

$$\frac{\Phi_f(\overline{V}_{\bar{n}})}{\mathbf{E}\Phi_f(\overline{V}_{\bar{n}})} \xrightarrow{p} 1 \quad \textit{iff} \quad \frac{\Phi_f(\overline{\Pi}_{\bar{n}})}{\mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}})} \xrightarrow{p} 1;$$

in this case, the normalizations $\mathbf{E}\Phi_f(\overline{V}_{\bar{n}})$ *and* $\mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}})$ *can be swapped.*

R e m a r k 4. In virtue of Chebyshev inequality, a sufficient condition for the limit relations in Corollary 2 is as follows:

$$\frac{\sum_{i\geq 1} \mathbf{Var} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})}{\left(\sum_{i\geq 1} \mathbf{E} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})\right)^2} \to 0.$$

For example, let $f_{i\bar{n}}(\cdot) \geq 0$ and $\sup_{\bar{x},i,\bar{n}} f_{i\bar{n}}(\bar{x}) \leq C_0$. Then $\mathbf{Var} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}}) \leq C_0 \mathbf{E} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})$ and

$$\frac{\sum_{i\geq 1} \mathbf{Var} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})}{\left(\sum_{i\geq 1} \mathbf{E} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})\right)^2} \leq C_0 \left| \sum_{i\geq 1} \mathbf{E} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}}) \right|^{-1} \to 0.$$

In particular, this estimate is valid in the case $f_{i\bar{n}}(\bar{x}) := I_B(\bar{x})$, where $0 \notin B$, and $\mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}}) = \sum_{i\geq 1} \mathbf{P}(\bar{\pi}_{i\bar{n}} \in B) \to \infty$.

Lemma 3. *Under the conditions* $\max_{\bar{n}} \sup_{\bar{x}} |f_{i\bar{n}}(\bar{x})| \le C_i \, \forall i$ *and* $\sum_{i \ge 1} C_i^2 p_i < \infty$ *the limit relation* $\lim_{\bar{n} \to \infty} \mathbf{Var}\Phi_f(\overline{V}_{\bar{n}}) = \infty$ *holds if and only if* $\lim_{\bar{n} \to \infty} \mathbf{Var}\Phi_f(\overline{\Pi}_{\bar{n}}) = \infty$. *In the case of infinite limit the following equivalence is valid:* $\mathbf{Var}\Phi_f(\overline{V}_{\bar{n}}) \sim \mathbf{Var}\Phi_f(\overline{\Pi}_{\bar{n}})$ *as* $\bar{n} \to \infty$.

Corollary 3. *Under the conditions of Lemma 3 and* $\mathbf{Var}\Phi_f(\overline{\Pi}_{\bar{n}}) \to \infty$ *as* $\bar{n} \to \infty$ *the limit relation*

$$\mathcal{L}\left(\frac{\Phi_f(\overline{V}_{\bar{n}}) - \mathbf{E}\Phi_f(\overline{V}_{\bar{n}})}{\mathbf{Var}^{1/2}\Phi_f(\overline{V}_{\bar{n}})}\right) \Longrightarrow N(0,1) \quad as \;\; \bar{n} \to \infty,$$

is valid if and only if

$$\mathcal{L}\left(\frac{\Phi_f(\overline{\Pi}_{\bar{n}}) - \mathbf{E}\Phi_f(\overline{\Pi}_{\bar{n}})}{\mathbf{Var}^{1/2}\Phi_f(\overline{\Pi}_{\bar{n}})}\right) \Longrightarrow \mathcal{N}(0,1) \quad as \;\; \bar{n} \to \infty,$$

where $\mathcal{N}(0,1)$ *is the standard normal distribution. In this case, the normalizing and centering sequences in these two limit relations can be respectively swapped.*

In order to prove this corollary we should put in Theorem 3 $D_{\bar{n}} := \mathbf{Var}^{-1/2}\Phi_f(\overline{\Pi}_{\bar{n}})$, $M_{\bar{n}} := \mathbf{E}\Phi_f(\overline{V}_{\bar{n}})\mathbf{Var}^{-1/2}\Phi_f(\overline{\Pi}_{\bar{n}})$, and $\mathcal{L}(\gamma) := \mathcal{N}(0,1)$.

R e m a r k 5. The validity of the central limit theorem for the sequence $\Phi_f(\overline{\Pi}_{\bar{n}})$ in Theorem 3 will be justified if, say, the third-order Lyapunov condition is met:

$$\frac{\sum_{i\geq 1} \mathbf{E}|f_{i\bar{n}}(\bar{\pi}_{i\bar{n}}) - \mathbf{E}f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})|^3}{\left(\sum_{i\geq 1} \mathbf{Var} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})\right)^{3/2}} \to 0 \quad \text{as} \quad \bar{n} \to \infty.$$

For example, let $\sup_{\bar{x},i,\bar{n}} |f_{i\bar{n}}(\bar{x})| \leq C_0$. Then it is easy to see that

$$\sum_{i\geq 1} \mathbf{E}|f_{i\bar{n}}(\bar{\pi}_{i\bar{n}}) - \mathbf{E}f_{i\bar{n}}(\bar{\pi}_{i\bar{n}})|^3 \leq 2C_0 \sum_{i\geq 1} \mathbf{Var} f_{i\bar{n}}(\bar{\pi}_{i\bar{n}}).$$

Examples of asymptotic behavior of the mean and variance.
1) Let $m = 1$, $B_k := \{i : i > k\}$ for any $k \in \mathbb{Z}_+$ and let $p_i := C\, i^{-1-b}$, where $b > 0$, $i = 1, 2, \ldots$. Then

$$\mathbf{E} \sum_{i \geq 1} I(\pi_{in} > k) = \sum_{i \geq 1} \mathbf{P}(\pi_{in} > k) = \sum_{i \geq 1} \gamma_{k+1,1}(np_i)$$

$$\sim (Cn)^{\frac{1}{1+b}} \int_0^\infty \gamma_{k+1,1}(y^{-1-b}) dy = \frac{(Cn)^{\frac{1}{1+b}}}{k!} \Gamma\left(k + \frac{b}{1+b}\right), \quad (9)$$

where $\gamma_{k+1,1}(z) := \int_0^z \frac{t^k}{k!} e^{-t} dt$, $\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} dt$, $z > 0$. For an arbitrary subset B of naturals one has

$$\mathbf{E} \sum_{i \geq 1} I(\pi_{in} \in B) \sim \frac{(Cn)^{\frac{1}{1+b}}}{(1+b)} \sum_{k \in B} \frac{1}{k!} \Gamma\left(k - \frac{1}{1+b}\right). \quad (10)$$

$\Phi_f(\Pi_{\bar{n}}) := \sum_{i\geq 1}\sum_{s\leq r} a_s I(\pi_{in} = k_s)$, where k_j are pairwise different.

$$\mathbf{Var}\sum_{i\geq 1}\sum_{s\leq m} a_s I(\pi_{in} = k_s) \sim \frac{(Cn)^{\frac{1}{1+b}}}{b+1}\left[\sum_{s=1}^{r}\frac{a_s^2}{k_s!}\Gamma\left(k_s - \frac{1}{b+1}\right)\right.$$

$$\left. - \sum_{s\neq j=1}^{r}\frac{2^{\frac{1}{b+1}-k_s-k_j}a_s a_j}{k_s!k_j!}\Gamma\left(k_s + k_j - \frac{1}{b+1}\right)\right] \equiv n^{\frac{1}{1+b}}\sum_{s,j=1}^{r} B_{s,j}a_s a_j$$

if only $\sum_{s,j=1}^{r} B_{s,j}a_s a_j \neq 0$. Note that $\{B_{s,j}\}$ is the limit covariance matrix of the sequence of random vectors

$$\left(n^{-\frac{1}{2(1+b)}}\sum_{i\geq 1} I(\pi_{in} = k_1), \ldots, n^{-\frac{1}{2(1+b)}}\sum_{i\geq 1} I(\pi_{in} = k_m)\right).$$

$$\Phi_f(\Pi_{\bar{n}}) := \sum_{i\geq 1}\sum_{s\leq r} a_s I(\pi_{in} > k_s) \ \forall k_1 \leq \ldots \leq k_r.$$

$$\mathbf{Var}\Phi_f(\Pi_n) \sim (Cn)^{\frac{1}{1+b}} \left[\sum_{s=1}^{r} a_s^2 \int_0^\infty \Gamma_{k_s+1,1}(v^{-1-b})dv \right.$$
$$\left. - \sum_{s\neq j=1}^{r} a_s a_j \int_0^\infty \Gamma_{k_s+1,1}(v^{-1-b})\Gamma_{k_j+1,1}(v^{-1-b})dv \right].$$

χ^2-STATISTICS

Finally, here is another consequence of theorem 3, relating to the asymptotic behavior of χ^2-statistics in (3) for $m = 1$ and $N \equiv N(n) \to \infty$. First of all, note that

$$\mathbf{E}\Phi_{\chi^2}(\Pi_n) = N,$$

$$D_n := \mathbf{Var}\Phi_{\chi^2}(\Pi_n) = 2N + \sum_{i=1}^{N} \frac{1}{np_i}.$$

Corollary 4. *Let $N \equiv N(n) \to \infty$ as $n \to \infty$. Then the following two asymptotic relations are equivalent*:

$$\mathcal{L}\left(\frac{\Phi_{\chi^2}(V_n) - N}{{D_n}^{1/2}}\right) \Longrightarrow \mathcal{N}(0,1), \tag{11}$$

$$\mathcal{L}\left(\frac{\Phi_{\chi^2}(\Pi_n) - N}{{D_n}^{1/2}}\right) \Longrightarrow \mathcal{N}(0,1). \tag{12}$$

Note that the centering sequence E_n can be replaced with its equivalent sequence $\mathbf{E}\Phi_{\chi^2}(V_n) = N - 1$. Replacing in the normalization in (11) the variance D_n with the variance of the χ^2-statistic itself, i.e., by the term

$$\mathbf{Var}\Phi_{\chi^2}(V_n) = 2N + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{np_i} - \frac{3N-2}{n},$$

is possible only if these two variances are equivalent. For example, this would be the case if $\min_{i \le N} np_i \to \infty$. This means that the growth rate of the sequence $N \equiv N(n)$ is subject to appropriate constraints, which is not the case in the above consequence.
So, in Corollary 4, we can talk about a double limit when $n, N \to \infty$.

The formulated criterion allows us to establish a fairly general sufficient condition for the asymptotic normality of χ^2-statistics with an increasing number of groups.

Theorem 4. *Let $N \equiv N(n) \to \infty$ as $n \to \infty$. Then the asymptotic relation* (11) *is valid if*

$$\frac{\sum_{i=1}^{N}(np_i)^{-2}}{\left(N + \sum_{i=1}^{N}(np_i)^{-1}\right)^{3/2}} \longrightarrow 0 \qquad (13)$$

as $n \to \infty$.

The problem of finding more or less broad sufficient conditions for asymptotic normality χ^2-statistics with a growing number of groups were studied by many authors in the second half of the last century (for example, see S.Tumanyan, Y.Medvedev, V.Kruglov).

Note that all known sufficient conditions for the above weak convergence provide execution of (13). For example, the condition $\min_{i \leq N} np_i \to \infty$ along with $N \to \infty$ (see G.Steek, S.Tumanyan), obviously immediately entails the limit relation (13). It is equally obvious that the requirement of the so-called regularity of polynomial models (see Y.Medvedev, V.Kruglov), i.e.,

$$0 < c_1 \leq \min_{i \leq N} Np_i, \quad \max_{i \leq N} Np_i < c_2 < \infty,$$

where the constants c_1 and c_2 do not depend on N, also implies (13).

On the other hand, it is easy to construct examples when the regularity requirement of the polynomial model is violated, the relation (13) is valid.

For example, let $p_i := C_N i^{-1-b}$, $i = 1, \dots, N$, where $b > 0$ and $C_N := \left(\sum_{i \le N} i^{-1-b}\right)^{-1}$. It is easy to see that, as $N \to \infty$, the sums $\sum_{i=1}^{N} p_i^{-2}$ and $\sum_{i=1}^{N} p_i^{-1}$ increase as N^{3+2b} and N^{2+b}, respectively. Therefore, as $n, N \to \infty$, the limit relation (13) is equivalent to

$$\frac{N^{3+2b}}{\sqrt{n}(N^{2+b})^{3/2}} = \frac{N^{b/2}}{\sqrt{n}} \to 0, \quad \text{i.e., } N = o(n^{\frac{1}{b}}).$$

The conditions by Tumanyan will be satisfied if $N = o(n^{\frac{1}{1+b}})$.

Proof of Theorem 4.

$$\xi_{in} := \frac{(\pi_{in} - np_i)^2}{np_i} - 1, \quad i = 1, \dots, N(n), \;\; n \geq 1.$$

The Lyapunov condition of third order, which guarantees the fulfillment of the central limit theorem (12), is as follows:

$$D_n^{-3/2} \sum_{i=1}^{N(n)} \mathbf{E}|\xi_{in}|^3 \to 0 \;\text{ as } n \to \infty. \tag{14}$$

To estimate the absolute third moment in (14), we need the well-known recurrence relation

$$\mathbf{E}(\pi_\lambda - \lambda)^n = \lambda \sum_{k=0}^{n-2} C_{n-1}^k \mathbf{E}(\pi_\lambda - \lambda)^k, \quad n \geq 2,$$

where π_λ is a Poisson random variable with parameter λ. From here it follows that

$$\mathbf{E}(\pi_\lambda - \lambda)^6 = 15\lambda^3 + 25\lambda^2 + \lambda.$$

Putting $\lambda = np_i$ and using the elementary estimate $|a^2 - 1|^3 \leq 4(a^6 + 1)$, we obtain

$$\mathbf{E}|\xi_{in}|^3 \leq \frac{4}{(np_i)^3}\left(15(np_i)^3 + 25(np_i)^2 + np_i\right) + 4 = 64 + \frac{100}{np_i} + \frac{4}{(np_i)^2}.$$

Thus, the Lyapunov ratio (14) is estimated by the value

$$\frac{64N + 100\sum_{i=1}^{N}\frac{1}{np_i} + 4\sum_{i=1}^{N}\frac{1}{(np_i)^2}}{\left(2N + \sum_{i=1}^{N}\frac{1}{np_i}\right)^{3/2}}$$

$$\leq 100\left(2N + \sum_{i=1}^{N}\frac{1}{np_i}\right)^{-1/2} + \frac{4\sum_{i=1}^{N}\frac{1}{(np_i)^2}}{\left(N + \sum_{i=1}^{N}\frac{1}{np_i}\right)^{3/2}} \to 0,$$

that is true in virtue of (13). $\square$

GENERALIZATION

We can reformulate the above-mentioned Poissonization duality theorem for more general type of additive statistics

$$V_n^{(m)}(\nu_n) := \sum_{i_1 \leq \ldots \leq i_m} f_{n,i_1,\ldots,i_m}(\nu_{n,i_1}, \ldots, \nu_{n,i_m}),$$

where $\{f_{n,i_1,\ldots,i_m}(\cdot)\}$ is an array of finite functions defined on Z_+^m and satisfying only the restriction

$$\sum_{i_1 \leq \ldots \leq i_m} |f_{n,i_1,\ldots,i_m}(0, \ldots, 0)| < \infty \quad \forall n.$$

Theorem. *As $n \to \infty$, the following three limit relations are equivalent*:

$$1)\ \mathcal{L}\left(V_n^{(m)}(\nu_n)D_n - M_n\right) \Longrightarrow \mathcal{L}(\gamma),$$

$$2)\ \mathcal{L}\left(V_n^{(m)}(\pi_n)D_n - M_n\right) \Longrightarrow \mathcal{L}(\gamma),$$

$$3)\ \mathcal{L}\left(V_n^{(m)}(\nu_n^*)D_n - M_n\right) \Longrightarrow \mathcal{L}(\gamma)$$

provided that $f_{n,i_1,\ldots,i_m}(\pi_{n,i_1},\ldots,\pi_{n,i_m})D_n \xrightarrow{p} 0$ for all multiindices $(i_1,\ldots,i_m)$, where M_n and D_n are some sequences, γ is some random variable, and the symbol «$\Longrightarrow$» denotes the weak convergence of distributions.

For example, one can study in this setting the functional

$$\tilde{\Phi}_I^{(m+2)}(V_n) := \sum_{i\geq 1} I_{\bar{A}}(\nu_{i-1,n}) I_A(\nu_{i,n}) \cdots I_A(\nu_{i+m-1,n}) I_{\bar{A}}(\nu_{i+m,n}),$$

where $0 \notin A$, $\bar{A}$ is the complement of A in the set $\mathbb{Z}_+$, and $\nu_{0n} = 0$. This functional is the number of success chains of length m in the Bernoulli trials $\{I_A(\nu_{i,n});\ i \geq 1\}$.