ՀԱՅԿԱԿԱՆ ՍՍՌ ԳԻՏՈՒԹՅՈՒՆՆԵՐԻ ԱԿԱԴԵՄԻԱՅԻ ՏԵՂԵԿԱԳԻՐ
ИЗВЕСТИЯ АКАДЕМИИ НАУК АРМЯНСКОЙ ССР
Ֆիզիկո-մաթեմատ. գիտություններ XI, № 6, 1958 Физико-математические науки

С. Х. Туманян

О мощности критерия χ^2, прилагаемого к проблеме двух выборок, относительно „близких" альтернатив

1. Пусть производится n независимых наблюдений случайной величины ξ и проверяется гипотеза H_0, согласно которой результаты этих наблюдений образуют случайную выборку из генеральной совокупности с функцией распределения $F(x)$. Одним из наиболее распространенных методов проверки гипотезы H_0 является критерий χ^2. В этом случае совокупность всех возможных значений случайной величины разбивается на $s+1$ непресекающихся интервалов Δ_i $(i=\overline{0, s})$ и составляется статистика

$$\chi^2 = \sum_{i=0}^{s} \frac{n}{p_i}\left(\frac{\mu_i}{n} - p_i\right)^2, \tag{1.1}$$

где $p_i = \int\limits_{\Delta_i} dF(x)$, а $\frac{\mu_i}{n}$ — относительные частоты наблюдений, принадлежащих интервалам Δ_i $(i=\overline{0, s})$.

Известно, что если гипотеза H_0 справедлива, то при $n \to \infty$ статистика (1.1) имеет распределение χ^2 с s степенями свободы, т. е.

$$P\{\chi^2 < x\} \xrightarrow[n\to\infty]{} K_s(x),$$

где

$$K_s(x) = \begin{cases} \dfrac{1}{2^{\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)} \displaystyle\int_0^x x^{\frac{s}{2}-1} e^{-\frac{x}{2}} dx, & x \geqslant 0 \\ 0, & x \leqslant 0. \end{cases} \tag{1.2}$$

Предположим теперь, что гипотеза H_0 неверна, а верна альтернативная гипотеза, согласно которой вероятности принадлежности наблюдения интервалам Δ_i равны p_i' $(i=\overline{0, s})$, причем

$$p_i' - p_i = \frac{z_i}{\sqrt{n}},$$

где z_i постоянные числа.

Тогда известно, что предельное распределение статистики (1.1) при $n \to \infty$ сводится к нецентральному распределению χ^2 [1]

Критерий χ^2 приложим также к исследованию проблемы двух выборок, состоящей, как известно, в следующем.

Пусть производится m независимых наблюдений случайной величины ξ и n независимых наблюдений случайной величины η. Проверяется гипотеза, состоящая в том, что результаты этих наблюдений образуют две случайные выборки из генеральных совокупностей с одной и той же функцией распределения.

Для проверки этой гипотезы можно использовать статистику

$$\chi^2 = mn\sum_{i=0}^{s}\frac{1}{\mu_i+\nu_i}\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^2, \tag{1.3}$$

где μ_i и ν_i—числа наблюдений соответственно из первой и второй выборок, принадлежащих интервалам $\Delta_i\,(i=\overline{0, s})$.

При этом известно, что в случае если гипотеза верна, то при $m\to\infty$ и $n\to\infty$ предельное распределение статистики (1.3) также определяется соотношением (1.2). ([2], стр. 485).

Предположим теперь, что гипотеза относительно тождественности двух распределений неверна. Обозначим в этом случае вероятности, что наблюдение случайной величины ξ, соответственно η, будет принадлежать интервалу $\Delta_i\,(i=\overline{0, s})$ через p_i и p_i' и предположим, что

$$p_i'-p_i=\frac{z_i\sqrt{(m+n)p_i}}{\sqrt{mn}}, \tag{1.4}$$

где z_i — постоянные числа.

Таким образом, мы предполагаем, что альтернативная гипотеза тем ближе к проверяемой, чем более объемы первой и второй выборок.

Целью настоящей работы является установление предельного распределения статистики (1.3) при условии (1.4) и при $m\to\infty$ и $n\to\infty$.

Пусть проверяется вышеупомянутая гипотеза о тождественности двух распределений. В случае, если эта гипотеза верна, то, пользуясь распределением (1.2), при заданном уровне значимости α, мы можем определить такое значение χ_α^2, что $P\{\chi^2>\chi_\alpha^2\}=\alpha$. Если же проверяемая гипотеза неверна, а верна альтернативная гипотеза, согласно которой выполняется соотношение (1.4), то знание распределения статистики (1.3) при условии (1.4) даст возможность определить в этом случае вероятность того, что $\chi^2>\chi_\alpha^2$, т. е. вероятность отвергнуть гипотезу в случае, если она неверна при альтернативной гипотезе (1.4).

Таким образом, знание указанного распределения решает вопрос о предельной мощности критерия χ^2 относительно рассматриваемых альтернатив. Для решения задачи мы составим статистику

$$X^2=mn\sum_{i=0}^{s}\frac{1}{mp_i+np_i'}\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^2 \tag{1.5}$$

и установим предельное распределение X^2 при $m \to \infty$ и $n \to \infty$, после чего покажем, что предельное распределение статистики (1.3) при условии (1.4) и при $m \to \infty$ и $n \to \infty$ совпадает с предельным распределением X^2.

2. *Математическое ожидание и дисперсия* X^2

Вычислим предельные значения математического ожидания и дисперсии X^2.

Имеем

$$MX^2 = mn\sum_{i=0}^{s}\frac{1}{mp_i + np'_i}M\left(\frac{\mu_i}{m} - \frac{\nu_i}{n}\right)^2 =$$

$$= mn\sum_{i=0}^{s}\frac{1}{mp_i + np'_i}\left[\frac{M\mu_i^2}{m^2} - \frac{2M\mu_i M\nu_i}{mn} + \frac{M\nu_i^2}{n^2}\right]. \qquad (2.1)$$

Как хорошо известно

$$\begin{aligned} M\mu_i &= mp_i \\ M\mu_i^2 &= m(m-1)p_i^2 + mp_i \qquad (2.2) \\ M\nu_i &= np'_i \\ M\nu_i^2 &= n(n-1)p_i'^2 + np'_i. \end{aligned}$$

Подставляя (2.2) в (2.1), после несложных преобразований получим:

$$MX^2 = \sum_{i=0}^{s}\frac{np_i q_i + mp'_i q'_i}{mp_i + np'_i} + mn\sum_{i=0}^{s}\frac{(p'_i - p_i)^2}{mp_i + np'_i}, \qquad (2.3)$$

$$(q_i = 1 - p_i, \quad q'_i = 1 - p'_i).$$

Рассмотрим каждую из этих сумм в отдельности.

Первая сумма вследствие условия (1.4) преобразовывается следующим образом

$$\sum_{i=0}^{s}\frac{np_i q_i + mp'_i q'_i}{mp_i + np'_i} =$$

$$= \sum_{i=0}^{s}\frac{np_i q_i + m\left(p_i + \frac{z_i\sqrt{(m+n)p_i}}{\sqrt{mn}}\right)\left(q_i - \frac{z_i\sqrt{(m+n)p_i}}{\sqrt{mn}}\right)}{mp_i + n\left(p_i + \frac{z_i\sqrt{(m+n)p_i}}{\sqrt{mn}}\right)} =$$

$$= \sum_{i=0}^{s}\frac{(m+n)p_i q_i}{(m+n)p_i + \frac{z_i\sqrt{n(m+n)p_i}}{\sqrt{m}}} +$$

$$+\sum_{i=0}^{s}\frac{\dfrac{q_i z_i\sqrt{m(m+n)p_i}}{\sqrt{n}}-\dfrac{p_i z_i\sqrt{m(m+n)p_i}}{\sqrt{n}}-\dfrac{z_i^2(m+n)p_i}{n}}{(m+n)p_i+\dfrac{z_i\sqrt{n(m+n)p_i}}{\sqrt{m}}} \tag{2.4}$$

Нетрудно видеть, что в соотношении (2.4) при $m\to\infty$ и $n\to\infty$ первая сумма стремится к $\sum\limits_{i=0}^{s} q_i = s$, а вторая сумма стремится к нулю, так что

$$\sum_{i=0}^{s}\frac{np_i q_i + mp_i' q_i'}{mp_i + np_i'}\xrightarrow[m,n\to\infty]{} s. \tag{2.5}$$

Что касается второй суммы в соотношении (2.3), то, как нетрудно видеть, вследствие (1.4) имеем

$$mn\sum_{i=0}^{s}\frac{(p_i'-p_i)^2}{mp_i+np_i'}=\sum_{i=0}^{s}\frac{(m+n)p_i z_i^2}{(m+n)p_i+\dfrac{z_i\sqrt{n(m+n)p_i}}{\sqrt{m}}}=$$

$$=\sum_{i=0}^{s}\frac{z_i^2}{1+\dfrac{z_i\sqrt{n}}{\sqrt{m(m+n)p_i}}}\xrightarrow[m,n\to\infty]{}\sum_{i=0}^{s} z_i^2. \tag{2.6}$$

Из соотношений (2.3), (2.5) и (2.6) получаем, что

$$MX^2\xrightarrow[m,n\to\infty]{} s+\sum_{i=0}^{s} z_i^2. \tag{2.7}$$

Перейдем теперь к вычислению предельного значения дисперсии X^2.

Как известно,

$$DX^2 = M(X^2)^2-(MX^2)^2. \tag{2.8}$$

Вследствие (2.7) для нахождения предельного значения DX^2 достаточно найти предельное значение $M(X^2)^2$. Имеем

$$M(X^2)^2=m^2n^2\left[\sum_{i=0}^{s}\frac{1}{mp_i+np_i'}M\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^2\right]^2=$$

$$=m^2n^2\sum_{i=0}^{s}\frac{1}{(mp_i+np_i')^2}M\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^4+$$

$$+m^2n^2\sum_{i=0}^{s}\sum_{\substack{j=0\\(i\neq j)}}^{s}\frac{1}{(mp_i+np_i')(mp_j+np_j')}M\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^2\left(\frac{\mu_j}{m}-\frac{\nu_j}{n}\right)^2. \tag{2.9}$$

Так как величины μ_i и ν_j $(i=\overline{0,s},\ j=\overline{0,s})$ независимы

$$M\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^4=\frac{M\mu_i^4}{m^4}-\frac{4M\mu_i^3\,M\nu_i}{m^3n}+$$

$$+\frac{6M\mu_i^2M\nu_i^2}{m^2n^2}-\frac{4M\mu_i\,M\nu_i^3}{mn^3}+\frac{M\nu_i^4}{n^4}, \tag{2.10}$$

$$M\left(\frac{\mu_i}{m}-\frac{\nu_i}{n}\right)^2\left(\frac{\mu_j}{m}-\frac{\nu_j}{n}\right)^2=\frac{M\mu_i^2\mu_j^2}{m^4}-\frac{2M\mu_i^2\mu_j\,M\nu_j}{m^3n}+\frac{M\mu_i^2M\nu_j^2}{m^2n^2}-$$

$$-\frac{2M\mu_i\mu_j^2M\nu_i}{m^3n}+\frac{4M\mu_i\mu_j\,M\nu_i\nu_j}{m^2n^2}-\frac{2M\mu_i\,M\nu_i\nu_j^2}{mn^3}+\frac{M\nu_i^2M\mu_j^2}{m^2n^2}$$

$$-\frac{2M\mu_j\,M\nu_i^2\nu_j}{mn^3}+\frac{M\nu_i^2\nu_j^2}{n^4}. \tag{2.11}$$

Как известно, совместной характеристической функцией величин $\mu_0,\mu_1,\cdots,\mu_s$ является ([2], стр. 455)

$$\psi(t_0,t_1,\cdots,t_s)=\left(\sum_{k=0}^{s}p_ke^{it_k}\right)^m. \tag{2.12}$$

Используя известное соотношение

$$M\mu_i^k\mu_j^l=\frac{1}{i^{k+l}}\left.\frac{\partial^{(k+l)}\psi(t_0,t_1,\cdots,t_s)}{\partial t_i^k\partial t_j^l}\right|_{t_0=0,\,t_1=0,\,\dots,\,t_s=0}\quad\begin{pmatrix}i,j=\overline{0,s}\\ k,l=\overline{0,s}\end{pmatrix},$$

получим

$$\begin{aligned}
&M\mu_i^3=m(m-1)(m-2)p_i^3+3m(m-1)p_i^2+mp_i\\
&M\mu_i^4=m(m-1)(m-2)(m-3)p_i^4+6m(m-1)(m-2)p_i^3+\\
&\qquad\qquad+7m(m-1)p_i^2+mp_i\\
&M\mu_i\mu_j=m(m-1)p_ip_j\\
&M\mu_i^2\mu_j=m(m-1)(m-2)p_i^2p_j+m(m-1)p_ip_j\\
&M\mu_i^2\mu_j^2=m(m-1)(m-2)(m-3)p_i^2p_j^2+m(m-1)(m-2)p_i^2p_j+\\
&\qquad\qquad+m(m-1)(m-2)p_ip_j^2+m(m-1)p_ip_j.
\end{aligned} \tag{2.13}$$

Аналогично

$$M\nu_i^3=n(n-1)(n-2)p_i'^3+3n(n-1)p_i'^2+np_i'$$

$$M\nu_i^4=n(n-1)(n-2)(n-3)p_i'^4+6n(n-1)(n-2)p_i'^3+$$

$$+7n(n-1)p_i'^2+np_i'$$

$$M\nu_i\nu_j = n(n-1)p'_i p'_j \tag{2.14}$$

$$M\nu_i^2\nu_j = n(n-1)(n-2)p_i'^2 p'_j + n(n-1)p'_i p'_j$$

$$M\nu_i^2\nu_j^2 = n(n-1)(n-2)(n-3)p_i'^2 p_j'^2 + n(n-1)(n-2)p_i'^2 p'_j +$$

$$+ n(n-1)(n-2)p'_i p_j'^2 + n(n-1)p'_i p'_j$$

Подставляя (2.2), (2.13) и (2.14) в (2.10) и (2.11), а затем (2.10) и (2.11) в (2.9), принимая во внимание (1.4) и переходя в (2.9) к пределу при $m \to \infty$ и $n \to \infty$, после некоторых преобразований получим:

$$\lim_{m,\, n \to \infty} M(X^2)^2 = \sum_{i=0}^{s}(3q_i^2 + 6z_i^2 q_i + z_i^4) + \sum_{\substack{i=0\\(i \neq j)}}^{s}\sum_{j=0}^{s}(3p_i p_j - p_i - p_j + 1) +$$

$$+ (q_i z_j^2 + q_j z_i^2 + z_i^2 z_j^2 - 4z_i z_j \sqrt{p_i p_j}). \tag{2.15}$$

Учитывая соотношение $\sum\limits_{i=0}^{s} p_i = 1$, нетрудно показать, что

$$\sum_{i=0}^{s} 3q_i^2 + \sum_{\substack{i=0\\(i \neq j)}}^{s}\sum_{j=0}^{s}(3p_i p_j - p_i - p_j + 1) = s^2 + 2s. \tag{2.16}$$

Остается подсчитать выражение

$$A_s = \sum_{i=0}^{s}(6z_i^2 q_i + z_i^4) + \sum_{\substack{i=0\\(i \neq j)}}^{s}\sum_{j=0}^{s}(q_i z_j^2 + q_j z_i^2 + z_i^2 z_j^2 - 4z_i z_j \sqrt{p_i p_j}) =$$

$$= \left(\sum_{i=0}^{s} z_i^2\right)^2 + 6\sum_{i=0}^{s} z_i^2 - 6\sum_{i=0}^{s} z_i^2 p_i +$$

$$+ \sum_{\substack{i=0\\(i \neq j)}}^{s}\sum_{j=0}^{s}(q_i z_j^2 + q_j z_i^2 - 4z_i z_j \sqrt{p_i p_j}). \tag{2.17}$$

Так как

$$\sum_{i=0}^{s} p_i = 1 \quad \text{и} \quad \sum_{i=0}^{s} p'_i = 1,$$

то из (1.4) следует, что

$$\sum_{i=0}^{s} z_i \sqrt{p_i} = 0. \tag{2.18}$$

Далее,

$$\sum_{\substack{i=0\\(i\neq j)}}^{s}\sum_{j=0}^{s} q_i z_j^2 = \sum_{i=0}^{s}\sum_{j=0}^{s} q_i z_j^2 - \sum_{i=0}^{s} q_i z_i^2 = s\sum_{j=0}^{s} z_j^2 - \sum_{i=0}^{s} z_i^2 + \sum_{i=0}^{s} z_i^2 p_i. \tag{2.19}$$

Аналогично

$$\sum_{\substack{i=0\\(i\neq j)}}^{s}\sum_{j=0}^{s} q_i z_i^2 = s\sum_{i=0}^{s} z_i^2 - \sum_{j=0}^{s} z_j^2 + \sum_{j=0}^{s} z_i^2 p_j. \tag{2.20}$$

Учитывая (2.18) и подставляя (2.19) и (2.20) в (2.17), получим

$$A_s = \left(\sum_{i=0}^{s} z_i^2\right)^2 + 2s\sum_{i=0}^{s} z_i^2 + 4\sum_{i=0}^{s} z_i^2. \tag{2.21}$$

Из (2.15), (2.16), (2.17) и (2.21) получаем, что

$$\lim_{m,n\to\infty} M(X^2)^2 = s^2 + 2s + \left(\sum_{i=0}^{s} z_i\right)^2 + 2s\sum_{i=0}^{s} z_i^2 + 4\sum_{i=0}^{s} z_i^2. \tag{2.22}$$

Из соотношений (2.8), (2.7) и (2.22) окончательно получаем

$$\lim_{m,n\to\infty} DX^2 = 2s + 4\sum_{i=0}^{s} z_i^2. \tag{2.23}$$

3. *Совместная характеристическая функция*

величин $x_k = \sqrt{\dfrac{mn}{mp_k + np_k'}}\left(\dfrac{\mu_k}{m} - \dfrac{\nu_k}{n}\right), \ (k = \overline{0, s}).$

Найдем предельное значение совместной характеристической функции величин

$$x_k = \sqrt{\frac{mn}{mp_k + np_k'}}\left(\frac{\mu_k}{m} - \frac{\nu_k}{n}\right), \quad (k = \overline{0, s}). \tag{3.1}$$

Обозначим характеристическую функцию величин x_k $(k = \overline{0, s})$ через $\varphi_{m,n}(t_0, t_1, \cdots, t_s)$.

Полагая

$$y_k = \sqrt{\frac{mn}{mp_k + np_k'}}\frac{\mu_k}{m},$$

$$u_k = -\sqrt{\frac{mn}{mp_k + np_k'}}\frac{\nu_k}{n}, \tag{3.2}$$

так что $x_k = y_k + u_k$,

будем иметь:

$$\varphi_{m,n}(t_0, t_1, \cdots, t_s) = Me^{it_0(y_0+u_0) + it_1(y_1+u_1) + \ldots + it_s(y_s+u_s)} =$$

$$= Me^{i(t_0 y_0 + t_1 y_1 + \ldots + t_s y_s) + i(t_0 u_0 + t_1 u_1 + \ldots + t_s u_s)}. \tag{3.3}$$

Так как величины μ_k и ν_j $(k = \overline{0, s},\ j = \overline{0, s})$ независимы, то величины y_k и u_j также будут независимыми, а потому из (3.3) следует

$$\varphi_{m,n}(t_0, t_1, \cdots, t_s) = Me^{i(t_0 y_0 + t_1 y_1 + \ldots + t_s y_s)} \cdot$$
$$\cdot Me^{i(t_0 u_0 + t_1 u_1 + \ldots + t_s u_s)}. \tag{3.4}$$

Таким образом, совместная характеристическая функция величин x_k $(k = \overline{0, s})$ равна произведению совместных характеристических функций величин y_k и u_k $(k = \overline{0, s})$.

Так как совместной характеристической функцией величин μ_k $(k = \overline{0, s})$ является функция (2.12), то совместной характеристической функцией величин y_k $(k = \overline{0, s})$, определяемых соотношением (3.2), будет

$$Me^{i(t_0 y_0 + t_1 y_1 + \cdots + t_s y_s)} = \left(\sum_{k=0}^{s} p_k e^{it_k\sqrt{\frac{n}{m(mp_k + np'_k)}}}\right)^m. \tag{3.5}$$

Аналогично

$$Me^{i(t_0 u_0 + t_1 u_1 + \cdots + t_s u_s)} = \left(\sum_{k=0}^{s} p'_k e^{-it_k\sqrt{\frac{m}{n(mp_k + np'_k)}}}\right)^n. \tag{3.6}$$

Из (3.4), (3.5) и (3.6) получаем

$$\varphi_{m,n}(t_0, t_1, \cdots, t_s) = \left(\sum_{k=0}^{s} p_k e^{it_k\sqrt{\frac{n}{m(mp_k + np'_k)}}}\right)^m \times$$
$$\times \left(\sum_{k=0}^{s} p'_k e^{-it_k\sqrt{\frac{m}{n(mp_k + np'_k)}}}\right)^n. \tag{3.7}$$

Логарифмируя соотношение (3.7), получим

$$\ln \varphi_{m,n}(t_0, t_1, \cdots, t_s) = m \ln\left(\sum_{k=0}^{s} p_k e^{it_k\sqrt{\frac{n}{m(mp_k + np'_k)}}}\right) +$$
$$+ n \ln\left(\sum_{k=0}^{s} p'_k e^{-it_k\sqrt{\frac{m}{n(mp_k + np'_k)}}}\right) =$$
$$= m \ln\Bigg(1 + i\sum_{k=0}^{s} t_k p_k \sqrt{\frac{n}{m(mp_k + np'_k)}} -$$

$$-\sum_{k=0}^{s}\frac{t_k^2}{2}\cdot\frac{np_k}{m(mp_k+np'_k)}+0\left(m^{-\frac{3}{2}}\right)\Big)+$$

$$+n\ln\Big(1-i\sum_{k=0}^{s}t_kp'_k\sqrt{\frac{m}{n(mp_k+np'_k)}}-$$

$$-\sum_{k=0}^{s}\frac{t_k^2}{2}\cdot\frac{mp'_k}{n(mp_k+np'_k)}+0\left(n^{-\frac{3}{2}}\right)\Big). \tag{3.8}$$

Применяя к правой части (3.8) формулу Маклорена, будем иметь:

$$\ln\varphi_{m,n}(t_0,\ t_1,\cdots,t_s)=i\sum_{k=0}^{s}t_kp_k\sqrt{\frac{mn}{mp_k+np'_k}}-$$

$$-\sum_{k=0}^{s}\frac{t_k^2}{2}\cdot\frac{np_k}{mp_k+np'_k}+\frac{m}{2}\left(\sum_{k=0}^{s}t_kp_k\sqrt{\frac{n}{m(mp_k+np'_k)}}\right)^2-$$

$$-i\sum_{k=0}^{s}t_kp'_k\sqrt{\frac{mn}{mp_k+np'_k}}-\sum_{k=0}^{s}\frac{t_k^2}{2}\cdot\frac{np_k}{mp_k+np'_k}+$$

$$+\frac{n}{2}\left(\sum_{k=0}^{s}t_kp'_k\sqrt{\frac{m}{n(mp_k+np'_k)}}\right)^2+0\left(m^{-\frac{1}{2}}+n^{-\frac{1}{2}}\right)=$$

$$=i\sum_{k=0}^{s}t_k(p_k-p'_k)\sqrt{\frac{mn}{mp_k+np'_k}}-\sum_{k=0}^{s}\frac{t_k^2}{2}\cdot\frac{np_k+mp'_k}{mp_k+np'_k}+$$

$$+\frac{m}{2}\left(\sum_{k=0}^{s}t_kp_k\sqrt{\frac{n}{m(mp_k+np'_k)}}\right)^2+$$

$$+\frac{n}{2}\left(\sum_{k=0}^{s}t_kp_k\sqrt{\frac{m}{n(mp_k+np'_k)}}\right)^2+0\left(m^{-\frac{1}{2}}+n^{-\frac{1}{2}}\right). \tag{3.9}$$

Учитывая (1.4) и переходя в соотношении (3.9) к пределу при $m\to\infty$ и $n\to\infty$, получим

$$\lim_{m,\,n\to\infty}\ln\varphi_{m,n}(t_0,\ t_1,\cdots,\ t_s)=-i\sum_{k=0}^{s}t_kz_k-$$

$$-\sum_{k=0}^{s}\frac{t_k^2}{2}+\frac{1}{2}\left(\sum_{k=0}^{s}t_k\sqrt{p_k}\right)^2.$$

Отсюда находим предельную характеристическую функцию величин $x_k\,(k=\overline{0,\,s})$.

$$\varphi(t_0, t_1, \cdots, t_s) = \lim_{m, n \to \infty} \varphi_{m, n}(t_0, t_1, \cdots, t_s) =$$

$$= e^{-i\sum\limits_{k=0}^{s} t_k z_k - \sum\limits_{k=0}^{s} \frac{t_k^2}{2} + \frac{1}{2}\left(\sum\limits_{k=0}^{s} t_k \sqrt{p_k}\right)^2}.$$

4. Предельное распределение X^2.

Из соотношений (1.5) и (3.1) следует, что

$$X^2 = \sum_{k=0}^{s} x_k^2. \tag{4.1}$$

В случае, когда $p_k' - p_k = 0$ $(k = \overline{0, s})$, совместной характеристической функцией величин

$$x_k = \sqrt{\frac{mn}{(m+n) p_k}} \left(\frac{\mu_k}{m} - \frac{\nu_k}{n}\right)$$

является функция

$$\varphi(t_0, t_1, \cdots, t_s) = e^{-\sum\limits_{k=0}^{s} \frac{t_k^2}{2} + \frac{1}{2}\left(\sum\limits_{k=0}^{s} t_k \sqrt{p_k}\right)^2},$$

а тогда, как известно, совместной функцией распределения величин x_k $(k = \overline{0, s})$ в пределе является многомерная нормальная функция распределения с математическими ожиданиями $Mx_k = 0$ $(k = \overline{0, s})$ и с матрицей вторых моментов $\Lambda = J - \bar{p}\bar{p}'$, ранга s, где J — единичная матрица, $\bar{p}' = (\sqrt{p_0}, \cdots, \sqrt{p_s})$ — вектор-строка, а $\bar{p}$ — соответствующий вектор-столбец.

Следовательно, распределением статистики X^2 в этом случае будет обычное распределение χ^2 с s степенями свободы, определяемое соотношением (1.2).

Если же существуют такие значения k, для которых $p_k' - p_k \neq 0$, то предельной характеристической функцией величин x_k, определяемых соотношением (3.1), является функция (3.4), а совместным распределением этих величин в пределе является многомерное нормальное распределение ранга s с математическими ожиданиями $Mx_k = -z_k$ и с той же матрицей вторых моментов $\Lambda = J - \bar{p}\bar{p}'$.

При этом предельным распределением статистики X^2 будет нецентральное распределение χ^2 с параметром, равным $\sum\limits_{k=0}^{s} z_k^2$.

Таким образом,

$$\lim_{m,n\to\infty} P(X^2 < x) = \frac{e^{-\frac{1}{2}\sum\limits_{k=0}^{s} z_k^2}}{2\left(\sum\limits_{k=0}^{s} z_k^2\right)^{\frac{s-1}{4}}} \int_0^x t^{\frac{s-1}{4}} I_{\frac{s-1}{2}}\left(t\sqrt{\sum_{k=0}^{s} z_k^2}\right) dt, \qquad (4.2)$$

где $I_s(t)$ — видиозмененная функция Бесселя первого рода. ($I_s(t) = t^{-s} J_s(it)$, где $J_s(t)$ — функция Бесселя первого рада).

5. Сходимость $\chi^2 - X^2$ по вероятности к нулю при $m \to \infty$ и $n \to \infty$

В этом пункте будет показано, что величина $\chi^2 - X^2$ сходится по вероятности к нулю, если $m \to \infty$ и $n \to \infty$.

Согласно (1.3) и (1.5) имеем

$$\chi^2 - X^2 = mn \sum_{k=0}^{s} \left(\frac{1}{\mu_k + \nu_k} - \frac{1}{mp_k + np_k'}\right)\left(\frac{\mu_k}{m} - \frac{\nu_k}{n}\right)^2. \qquad (5.1)$$

Если к функции $\frac{1}{x}$ применить теорему о среднем, то можно написать

$$\frac{1}{x} = \frac{1}{c} - \frac{x-c}{[c+\theta(x-c)]^2}, \quad \text{где} \quad 0 < \theta < 1.$$

Полагая $x = \mu_k + \nu_k$, а $c = mp_k + np_k'$, будем иметь

$$\frac{1}{\mu_k + \nu_k} = \frac{1}{mp_k + np_k'} - \frac{[\mu_k + \nu_k - (mp_k + np_k')]}{[mp_k + np_k' + \theta_k(\mu_k + \nu_k - mp_k - np_k')]^2}, \qquad (5.2)$$

$0 < \theta_k < 1, \quad k = \overline{0, s}.$

Подставляя (5.2) в (5.1), получаем

$$\chi^2 - X^2 = -mn \sum_{k=0}^{s} \frac{(\mu_k + \nu_k - mp_k - np_k')\left(\frac{\mu_k}{m} - \frac{\nu_k}{n}\right)^2}{[mp_k + np_k' + \theta_k(\mu_k + \nu_k - mp_k - np_k')]^2},$$

откуда

$$|\chi^2 - X^2| = mn \sum_{k=0}^{s} \frac{|\mu_k + \nu_k - mp_k - np_k'|\left(\frac{\mu_k}{m} - \frac{\nu_k}{n}\right)^2}{[mp_k + np_k' + \theta_k(\mu_k + \nu_k - mp_k - np_k')]^2}. \qquad (5.3)$$

Оценим математическое ожидание $|\chi^2 - X^2|$.

Имеем

$$M|\chi^2 - X^2| = mn\sum_{k=0}^{s} M\frac{|\mu_k+\nu_k-mp_k-np_k'|\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2}{[mp_k+np_k'+\theta_k(\mu_k+\nu_k-mp_k-np_k')]^2}. \quad (5.4)$$

Так как $\mu_k \geqslant 0$ и $\nu_k \geqslant 0$, получаем

$$mp_k+np_k'+\theta_k(\mu_k+\nu_k-mp_k-np_k') \geqslant (1-\theta_k)(mp_k+np_k'). \quad (5.5)$$

Принимая во внимание, что в (5.4) под знаком математического ожидания находится положительная величина, а также учитывая неравенство (5.5), будем иметь

$$M\frac{|\mu_k+\nu_k-mp_k-np_k'|\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2}{[mp_k+np_k'+\theta_k(\mu_k+\nu_k-mp_k-np_k')]^2} \leqslant$$

$$\leqslant \frac{M|\mu_k+\nu_k-mp_k-np_k'|\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2}{(1-\theta_k)^2(mp_k+np_k')^2}$$

и, следовательно,

$$M|\chi^2 - X^2| \leqslant mn\sum_{k=0}^{s}\frac{M|\mu_k+\nu_k-mp_k-np_k'|\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2}{(1-\theta_k)^2(mp_k+np_k')^2}. \quad (5.6)$$

Согласно неравенству Коши-Буняковского имеем:

$$M|\mu_k+\nu_k-mp_k-np_k'|\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2 \leqslant$$

$$\leqslant \sqrt{M(\mu_k+\nu_k-mp_k-np_k')^2}\sqrt{M\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^4}. \quad (5.7)$$

Легко видеть, что

$$M(\mu_k+\nu_k-mp_k-np_k')^2 = M(\mu_k-mp_k)^2 + M(\nu_k-np_k')^2 =$$

$$= mp_kq_k + np_k'q_k'.$$

Подставляя в полученное выражение вместо p_k' его значение по формуле (1.4), а также вместо q_k' его значение

$$q_k' = q_k - \frac{z_k\sqrt{(m+n)p_k}}{\sqrt{mn}},$$

получим

$$M(\mu_k+\nu_k-mp_k-np_k')^2 = (m+n)p_kq_k + O(\sqrt{m}+\sqrt{n}). \quad (5.8)$$

Остается подсчитать $M\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^4$.

Подставляя (2.2), (2.13) и (2.14) в (2.10) после некоторых преобразований, получим

$$M\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^4=\frac{3(np_kq_k+mp'_kq'_k)^2}{m^2n^2}+$$

$$+\frac{6z_k^2(m+n)}{m^2n^2}(np_kq_k+mp'_kq'_k)+\frac{z_k^2(m+n)^2}{m^2n^2}-$$

$$-\frac{4p_kq_k(q_k-p_k)}{m^2}\cdot\frac{z_k\sqrt{m+n}}{\sqrt{mn}}-\frac{4p'_kq'_k(q'_k-p'_k)}{n^2}\cdot\frac{z_k\sqrt{m+n}}{\sqrt{mn}}+$$

$$+\frac{p_kq_k(1-6p_kq_k)}{m^3}+\frac{p'_kq'_k(1-6p'_kq'_k)}{n^3}. \tag{5.9}$$

Формула (5.9) показывает, что

$$M\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^4=0\left[\frac{(m+n)^2}{m^2n^2}\right]. \tag{5.10}$$

Согласно (5.8) и (5.10) из (5.7) следует, что

$$M\,|\mu_k+\nu_k-mp_k-np'_k|\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2=0\left[\frac{(m+n)^{\frac{3}{2}}}{(mn)^{\frac{1}{2}}}\right],$$

так что согласно (5.6)

$$M\,|\chi^2-X^2|=O\left(\frac{1}{\sqrt{m+n}}\right)$$

и поэтому

$$\lim_{m,\,n\to\infty}M\,|\chi^2-X^2|=0. \tag{5.11}$$

Применим к случайной величине $|\chi^2-X^2|$ неравенство Чебышева. Тогда, каково бы ни было $\varepsilon>0$, будем иметь

$$P\{|\chi^2-X^2|>\varepsilon\}<\frac{M\,|\chi^2-X^2|}{\varepsilon}.$$

Исходя из (5.11) заключаем, что для любого $\varepsilon>0$

$$P\{|\chi^2-X^2|>\varepsilon\}\to 0 \quad \text{при} \quad m\to\infty \quad \text{и} \quad n\to\infty,$$

а это и означает, что статистика χ^2-X^2 сходится по вероятности к нулю при $m\to\infty$ и $n\to\infty$.

6. Предельное распределение χ^2.

В этом пункте будет показано, что предельное распределение статистики χ^2, определяемой соотношением (1.3), при условии (1.4) совпадает с предельным распределением X^2 при $m\to\infty$ и $n\to\infty$.

Напишем тождество

$$\chi^2 = X^2 + (\chi^2 - X^2). \tag{6.1}$$

Как было показано, предельное распределение статистики X^2 есть нецентральное распределение χ^2, определяемое формулой (4.2).

Что касается величины $\chi^2 - X^2$, то, как было показано в предыдущем пункте, при $m \to \infty$ и $n \to \infty$ она сходится по вероятности к нулю.

На основании известной теоремы ([2], стр. 281) в этом случае статистика χ^2 будет иметь предельное распределение, совпадающее с предельным распределением статистики X^2.

Итак,

$$\lim_{m,\, n \to \infty} P\{\chi^2 < x\} = \frac{e^{-\frac{1}{2}\sum\limits_{k=0}^{s} z_k^2}}{2\left(\sum\limits_{k=0}^{s} z_k^2\right)^{\frac{s-1}{4}}} \int\limits_0^x t^{\frac{s-1}{4}} e^{-\frac{t}{2}} I_{\frac{s-1}{2}}\left(\sqrt{t\sum_{k=0}^{s} z_k^2}\right) dt.$$

Таким образом, предельным распределением статистики (1.3) при условии (1.4) и при $m \to \infty$ и $n \to \infty$ является нецентральное распределение χ^2, являющееся бесселевым распределением.

Институт математики и механики
АН Армянской ССР

Поступило 10 III 1958

Ս. Խ. Թումանյան

ԵՐԿՈՒ ԸՆՏՐՈՒԹՅՈՒՆՆԵՐԻ ՊՐՈԲԼԵՄԻՆ ԿԻՐԱՌՎՈՂ «ՄՈՏ» ԱԼՏԵՐՆԱՏԻՎՆԵՐԻ ՆԿԱՏՄԱՄԲ χ^2 ՀԱՅՏԱՆԻՇԻ ՀԶՈՐՈՒԹՅԱՆ ՄԱՍԻՆ

Ա Մ Փ Ո Փ Ո Ւ Մ

Դիցուք կատարվում են ξ պատահական մեծության m անկախ դիտումներ և η պատահական մեծության n անկախ դիտումներ։ Արվում է հիպոթեզ, որի համաձայն այս դիտումների արդյունքները կազմում են միևնույն բաշխման ֆունկցիա ունեցող հիմնական համախմբություններից երկու պատահական ընտրություններ։ ξ և η պատահական մեծությունների բոլոր հնարավոր արժեքների բազմությունը տրոհվում է $s+1$ չհատվող $\Delta_k (k=\overline{0, s})$ ինտերվալների և կազմվում է $\chi^2 = mn\sum\limits_{k=0}^{s}\frac{1}{\mu_k+\nu_k}\left(\frac{\mu_k}{m}-\frac{\nu_k}{n}\right)^2$ մեծությունը, որտեղ μ_k և ν_k $\Delta_k (k=\overline{0, s})$ ինտերվալներին պատկանող դիտումների թվերն են՝ համապատասխանաբար առաջին և երկրորդ ընտրություններից։

Դիտարկվում է χ^2 մեծության սահմանային բաշխումը գտնելու խնդիրը երբ $m \to \infty$ և $n \to \infty$, այն ենթադրությամբ, որ ընդունված հիպոթեզը ճիշտ չէ, այլ ճիշտ է ընդունվածին «մոտ» ալտերնատիվ հիպոթեզը, որի համաձայն

$$p'_k - p_k = \frac{z_k\sqrt{(m+n)\,p_k}}{\sqrt{mn}},$$

որտեղ p'_k և p_k — Δ_k ինտերվալին պատկանող դիտման համապատասխանաբար իսկական և հիպոթետիկ հավանականություններն են, իսկ $z_k\ (k=\overline{0, s})$ հաստատուն թվեր են։

Ցույց է տրված, որ

$$\lim_{m,\, n\to\infty} P\{\chi^2 < x\} = \frac{e^{-\frac{1}{2}\sum\limits_{k=0}^{s} z_k^2}}{2\left(\sum\limits_{k=0}^{s} z_k^2\right)^{\frac{s-1}{4}}} \int\limits_0^x t^{\frac{s-1}{4}} e^{-\frac{t}{2}} I_{\frac{s-1}{4}}\left(\sqrt{t\sum_{k=0}^{s} z_k^2}\right) dt$$

որտեղ $I_s(t)$ — առաջին կարգի ձևափոխված Բեսսելի ֆունկցիան է ($I_s(t) = i^{-s}J_s(it)$ ուր $J_s(it)$ առաջին կարգի Բեսսելի ֆունկցիա է)։

Այս բաշխման որոշումը χ^2 մեծության համար լուծում է երկու ընտրությունների պրոբլեմին կիրառվող «մոտ» ալտերնատիվների նկատմամբ χ^2 հայտանիշի հզորության հարցը։

ЛИТЕРАТУРА

1. *Cochran. W. G.* The χ^2 Test of Goodness of Fit." Annals of Math. Stat., Vol. 23, № 3, 1952.

2. *Крамер. Г.* Математические методы статистики. Гос. издат. иностран. литературы, Москва, 1948.