УДК 519.2

Э.М.Кудлаев
(Московский университет)

РАСПРЕДЕЛЕНИЕ СТАТИСТИКИ УИТВОРТА ПРИ СБЛИЖАЮЩИХСЯ АЛЬТЕРНАТИВАХ

Пусть $U_n(1),\dots,U_n(n)$ – вариационный ряд (в.р.), построенный по простой случайной выборке $U_1,\dots,U_n$ из равномерного на $[0,1]$ распределения, а $\varphi_n(u)$, $n=1,2,\dots,$ – последовательность функций распределения, определенных при $u\in[0,1]$.
Рассмотрим случайную величину

$$\xi_n = \max_{j=1,\dots,n+1} |\varphi_n(U_n(j)) - \varphi_n(U_n(j-1)) - (n+1)^{-1}|, \qquad (1)$$

где $U_n(0)=0$, $U_n(n+1)=1$.

Если, в частности, $\varphi_n(u)\equiv u$ при всех n, то можно показать (см., например, [3], с. 77), что

$$\lim_{n\to\infty} P\{(n+1)\xi_n - \ln(n+1) + 1 < x\} = \exp\{-e^{-x}\}. \qquad (2)$$

Отметим, что статистики

$$\max_{j=1,\dots,n+1}[U_n(j)-U_n(j-1)] \quad \text{и} \quad \min_{j=1,\dots,n+1}[U_n(j)-U_n(j-1)]$$

изучались еще Уитвортом [2]; их совместное асимптотическое $(n\to\infty)$ распределение выписано Дарлингом ([1] с. 252). Если $\mathcal{J}_n$ – класс случайных множеств $A_n(j)=(U_n(j-1),U_n(j)]$, $j=1,\dots,n+1$, то на нем можно ввести две вероятностные меры: истинную P_n, определяемую соотношениями $P_n(A_n(j))=\varphi_n(U_n(j))-\varphi_n(U_n(j-1))$, и эмпирическую P_n^*, определяемую соотношениями $P_n^*(A_n(j))=(n+1)^{-1}$, $j=1,\dots,n+1$. При этом величина ξ_n совпадает с расстоянием по вариации между мерами P_n и P_n^* на классе $\mathcal{J}_n$.

В настоящей работе получено асимптотическое распределение величины ξ_n, если

$$\varphi_n(u) = u + \frac{1}{\ln(n+1)}\cdot\int_0^u b_n(x)\,dx, \qquad (3)$$

а интеграл $\int_0^u b_n(x)dx$ есть величина порядка $O(1)$. Полученный результат используется для построения критерия согласия и вычисления его мощности при специальных сближающихся альтернативах. Пусть

$$b_n(u) = b(u) + \varepsilon_n(u), \tag{4}$$

где функции $b(u)$, $\varepsilon_n(u)$ непрерывны на $[0,1]$ и, кроме того, $\varepsilon_n(u) \to 0$ при $n \to \infty$ равномерно по всем $u \in [0,1]$.

Т е о р е м а. Если выполняется условие (4) и

$$\xi_n^0 = (n+1)\xi_n - \ln(n+1) + 1, \tag{5}$$

то

$$\lim_{n\to\infty} P\left\{\xi_n^0 < x\right\} = exp\left\{-\mu \cdot e^{-x}\right\}, \tag{6}$$

$$\mu = \int_0^1 exp\left\{b(u)\right\}du. \tag{7}$$

Д о к а з а т е л ь с т в о. Воспользуемся обозначениями:

$$b_j = b\left(\frac{j}{n+1}\right), \quad \eta_n = \max_{j=1,\dots,n+1}\left|\left[1 + \frac{b_j}{\ln(n+1)}\right]\left[U_n(j) - U_n(j-1)\right] - (n+1)^{-1}\right|,$$

$$\eta_n^0 = (n+1)\eta_n - \ln(n+1) + 1,$$

$$h_j(r;\lambda) = \begin{cases} 1, & \text{если } \left|\left[1 + \dfrac{b_j}{\ln(n+1)}\right]\cdot r - (n+1)^{-1}\right| < \lambda, \\ 0, & \text{если } \left|\left[1 + \dfrac{b_j}{\ln(n+1)}\right]\cdot r - (n+1)^{-1}\right| \geqslant \lambda, \end{cases}$$

$$j = 1, \dots, n+1, \quad \lambda > 0.$$

На основании теоремы о среднем значении

$$\int_{U_n(j-1)}^{U_n(j)} b_n(u)du = \left[U_n(j) - U_n(j-1)\right]\cdot b_n(\xi_n^*(j)),$$

где

$$\xi_n^*(j) = U_n(j-1) + \theta_n^*(j)\cdot\left[U_n(j) - U_n(j-1)\right], \quad |\theta_n^*(j)| < 1.$$

Так как равномерно по всем j

$$U_n(j-1) = \frac{j}{n+1} + O_p\left(\frac{1}{\sqrt{n}}\right),$$

то

$$\zeta_n^*(j) = \frac{j}{n+1} + O_p\left(\frac{1}{\sqrt{n}}\right).$$

Поэтому в силу равномерной непрерывности на $[0,1]$ функций $b(u)$, $\mathcal{E}_n(u)$ имеем

$$b_n(\zeta_n^*(j)) = b_j + \mathcal{E}_n,$$

где последовательность величин $\mathcal{E}_n$ стремится к нулю равномерно по всем $j = 1,\dots,n+1$, что влечет равенство

$$\xi_n^0 = \eta_n^0 + \frac{\mathcal{E}_n}{\ln(n+1)} \cdot \max_{j=1,\dots,n+1} [U_n(j) - U_n(j-1)].$$

На основании формулы (2) последнее соотношение перепишется в следующем виде:

$$\xi_n^0 = \eta_n^0 + O_p(1). \tag{8}$$

Теперь, следуя методу Дарлинга, выпишем предельное распределение для η_n^0. Имеем

$$\begin{aligned} p_n(\lambda) &= P\{\eta_n < \lambda\} = M\left[\prod_{j=1}^{n+1} h_j(U_n(j) - U_n(j-1); \lambda)\right] = \\ &= \frac{n!}{2\pi i} \cdot \int_{c-i\infty}^{c+i\infty} e^z \cdot \prod_{j=1}^{n+1} H_j(z)dz, \quad Re\, c > 0; \end{aligned} \tag{9}$$

здесь

$$H_j(z) = \int_0^{+\infty} h_j(r;\lambda) \cdot e^{-rz} \cdot dr = z^{-1}(e^{-\beta_j z} - e^{-\alpha_j z}),$$

$$\alpha_j = \left[1 + \frac{b_j}{\ln(n+1)}\right]^{-1} \cdot \left(\lambda + \frac{1}{n+1}\right), \quad \beta_j = \max\left\{0, \left[1 + \frac{b_j}{\ln(n+1)}\right]^{-1}\left(\lambda + \frac{1}{n+1}\right)\right\}.$$

Последнее равенство в соотношении (9) написано на основании теоремы 2.1 [1]. Переходя в интеграле правой части соотношения (9) к новой переменной $z = (n+1)w$, получим

$$p_n(\lambda) = \frac{1}{2\pi i} \cdot \frac{(n+1)!}{(n+1)^{n+1}} \cdot \int_{c'-i\infty}^{c'+i\infty} exp\{(n+1)(w - \ln w)\} \times \tag{10}$$

$$\times \prod_{j=1}^{n+1} [(n+1)w \cdot H_j((n+1)w)]dw, \quad Re c' > 0.$$

Функция $\psi(w) = w - \ln w$ в полуплоскости $Re\, w > 0$ имеет единственную седловую точку $w = 1$, причем линия наибольшего ската, проходящая через эту точку, является прямой, параллельной мнимой оси. Выберем $c' = 1$ и введем новые переменные

$$w = 1 + iy(n+1)^{-1/2}, \quad \lambda = (n+1)^{-1} \cdot [x + \ln(n+1) - 1];$$

при этом элемент $exp\{(n+1)(w - \ln w)\}dw$ заменится элементом

$$\frac{i \cdot e^{n+1}}{\sqrt{n+1}} \cdot e^{-y^2/2} [1 + \rho_n(y)]dy \qquad (11)$$

и согласно формуле Стирлинга

$$\frac{(n+1)!\, e^{n+1}}{(n+1)^{n+3/2}} \cdot \sqrt{2\pi}\, (1 + o(1)); \qquad (12)$$

здесь и в дальнейшем через $\rho_n(x_1, \dots, x_r)$ будем обозначать величину, стремящуюся к нулю (при $r \to \infty$) равномерно по всем точкам $(x_1, \dots, x_r)$ из каждого ограниченного r -мерного параллелепипеда. Далее, применяя стандартные разложения, вместо $\alpha_j z$ можно написать выражение

$$[\ln(n+1) + x][1 - \frac{b_j}{\ln(n+1)} + o(\frac{1}{\ln^2 n})][1 + \frac{iy}{\sqrt{n+1}}] =$$

$$= \ln(n+1) - b_j + x + \rho_n(x, y);$$

кроме того, начиная с некоторого номера n_0, величина

$$[1 + \frac{b_j}{\ln(n+1)}]^{-1} \cdot [-\frac{x}{n+1} - \frac{\ln(n+1)}{n+1} + \frac{2}{n+1}]$$

отрицательна при всех x, принадлежащих ограниченному интервалу. Поэтому, начиная с n_0, имеем

$$exp\{-\beta_j(n+1+iy\sqrt{n+1})\} - exp\{-\alpha_j(n+1+iy\sqrt{n+1})\} =$$

$$= 1 - exp\{-\ln(n+1) + b_j - x + \rho_n(x,y)\} = 1 - (n+1)^{-1} \cdot exp\{-x +$$

$$+b_j\} + (n+1)^{-1}\cdot \rho_n(x,y).$$

Так как

$$(n+1)^{-1}\sum_{j=1}^{n+1} e^{b_j} = \mu + o(1),$$

то произведение $\prod_{j=1}^{n+1} zH_j(z)$ перепишется в виде

$$exp\Big\{\sum_{j=1}^{n+1} ln\big[1-(n+1)^{-1}exp\{-x+b_j\} + (n+1)^{-1}\cdot\rho_n(x,y)\big]\Big\} =$$

$$= exp\{-\mu\cdot e^{-x}\}\cdot[1+\rho_n(x,y)].$$

Отсюда, замечая, что в нашем случае возможен предельный переход под знаком интеграла, и из соотношений (8), (10)-(12) вытекает утверждение теоремы.

Применим полученный результат к построению критерия согласия. Пусть $X_n(1),\dots,X_n(n)$ - в.р., построенный по простой случайной выборке размера n из совокупности с функцией распределения $F_n(x)$, и пусть $\widetilde{F}_n(x)$ - некоторая функция распределения, которая проверяется на согласие с $F_n(x)$.

Рассмотрим статистику

$$\widetilde{\xi}_n = \max_{j=1,\dots,n+1} |\widetilde{F}_n(X_n(j)) - \widetilde{F}_n(X_n(j-1)) - (n+1)^{-1}|, \qquad (13)$$

$$\widetilde{F}_n(X_n(0)) = 0,\ \widetilde{F}_n(X_n(n+1)) = 1.$$

Так как последовательность $F_n(X_n(1)),\dots,F_n(X_n(n))$ образует в.р. $U_1(1),\dots,U_n(n)$, то равенство (13) можно переписать в виде

$$\widetilde{\xi}_n = \max_{j=1,\dots,n+1} |\varphi_n(U_n(j)) - \varphi_n(U_n(j-1)) - (n+1)^{-1}|, \qquad (13')$$

здесь

$$\varphi_n(u) = \widetilde{F}_n[F_n^{-1}(u)],\ F_n^{-1}(u) = \sup\{x : F_n(x) \le u\}.$$

В обозначениях $f_n(x) = F_n'(x)$, $\tilde{f}_n(x) = \tilde{F}_n'(x)$ имеем для нашего случая

$$\frac{b_n(u)}{\ln(n+1)} = \varphi_n'(u) - 1 = \frac{\tilde{f}_n[F_n^{-1}(u)]}{f_n[F_n^{-1}(u)]} - 1 \cdot \qquad (I4)$$

Если $\tilde{\xi}_n^0 = (n+1)\tilde{\xi}_n - \ln(n+1) + 1$, то для проверки гипотезы $\tilde{F}_n(x) = F_n(x)$ против альтернатив $\tilde{F}_n(x) \neq F_n(x)$ определим критерий с критической областью

$$\mathfrak{D}: \tilde{\xi}_n^0 < -C_{\alpha_1} \quad \text{или} \quad \tilde{\xi}_n^0 > C_{\alpha_2}, \qquad (I5)$$

где константы C_{α_1}, C_{α_2} подбираются по заданным числам α_1, α_2; $\alpha_1, \alpha_2 > 0, \alpha_1 + \alpha_2 < 1$, из соотношений

$$C_{\alpha_1} = \ln\ln\alpha_1^{-1}, \quad C_{\alpha_2} = -\ln\ln(1-\alpha_2)^{-1}. \qquad (I6)$$

Воспользовавшись теоремой, получим утверждение.

Следствие. Если функция $b_n(u)$, определяемая из уравнения (I4), удовлетворяет условию (4), то критерий с критической областью $\mathfrak{D}$ имеет при $n \to \infty$ мощность

$$\beta_\infty(\alpha_1, \alpha_2) = 1 + exp\{-\mu \cdot e^{C_{\alpha_1}}\} - exp\{-\mu \cdot e^{C_{\alpha_2}}\}. \qquad (I7)$$

Замечание. При каждом $n = 1, 2, \dots$

$$\int_0^1 b_n(u)du = \ln(n+1) \cdot \int_0^1 [\varphi_n'(u) - 1]du = 0,$$

а, следовательно,

$$\int_0^1 b(u)du = 0,$$

т.е. функция $b(u)$ обязательно принимает отрицательные значения. Поэтому может случиться, что $\mu = 1$, и в этом случае указанный критерий асимптотически не различает рассматриваемых здесь альтернатив.

Литература

I. Darling D.A. On a Class of Problems Related to the Random Division of an Interval.-Ann. Math. Stat., v.24, 2, 1953, 239-253.

2. Whitworth W.A. Choice and Chance.-Cambridge University Press, 1897.

3. Кудлаев Э.М. О непараметрических статистиках, построенных по отрезкам вариационного ряда. - Труды Сиб. физ.-техн. ин-та при Томском ун-те, 1973, вып. 63, 69-81.