УДК 519.21

Б. В. Бондарев

О статистике Колмогорова в случае кусочно-непрерывной функции распределения

Кроме дискретных и непрерывных случайных величин в задачах практики естественным образом возникает необходимость рассматривать так называемые непрерывно-дискретные случайные величины. Подобная ситуация, например [1], имеет место при исследовании времени работы системы после первого ремонта, при прохождении сигнала через фильтр — ограничитель, при исследовании длины пересечения случайного интервала с некоторым фиксированным интервалом, а также в ряде других случаев [2, 3]. Кусочно-непрерывные функции распределения этих случайных величин допускают представление в виде смеси распределений. Действительно, в случае отсутствия у $F(x)$ сингулярной компоненты по теореме Лебега [4] имеет место разложение $F(x) = \Phi^{р}(x) + \Phi^{н}(x)$, где $\Phi^{н}(x)$ — дискретная компонента, $\Phi^{н}(x)$ — абсолютно непрерывная, $\Phi^{р}(x)$ и $\Phi^{р}(x)$ — неубывающие функции. Если, например, скачки функции $F(x)$ сосредоточены на конечном интервале, то, обозначив $F^{р}(x) = \dfrac{\Phi^{р}(x)}{\alpha}$, где $\alpha = \lim\limits_{x\to\infty}[F(x) - \Phi^{н}(x)]$, имеем $F(x) - \alpha F^{р}(x) = \Phi^{н}(x)$. Так как $\lim\limits_{x\to\infty}[F(x) - \alpha F^{р}(x)] = \lim\limits_{x\to\infty}\Phi^{н}(x) = 1-\alpha$, то, обозначив $F^{н}(x) = \dfrac{\Phi^{н}(x)}{1-\alpha}$, окончательно имеем

$$F(x) = \alpha F^{р}(x) + (1-\alpha)F^{н}(x),\quad 0<\alpha<1. \tag{1}$$

В данной статье рассматривается задача проверки гипотезы о том, что наблюдается случайная величина, распределение которой имеет представление (1). Веса $F^{н}(x)$, $F^{р}(x)$ считаются заданными, наблюдения $\{x_1, x_2, \ldots, \ldots, x_n\}$ предполагаются независимыми, одинаково распределенными. Применение для этой цели критерия χ^2 затруднено. Это касается способа разбиения оси абсцисс на интервалы и потери информации вследствие группировки данных [5]. В данной работе в качестве меры уклонения эмпирической функции распределения от гипотетической предполагается использовать статистику Колмогорова $D_n = \sqrt{n}\sup\limits_x |F_n(x) - F(x)|$, где $F_n(x)$ — эмпирическая функция распределения, построенная по $\{x_1, \ldots, x_n\}$. В случае, если смесь непрерывных распределений D_n инвариантна относительно распределения и для проверки гипотезы можно воспользоваться классическим критерием Колмогорова, либо соответствующими теоремами из [6, 7], условия которых в данной ситуации легко проверяются, так как стандартной заменой случай непрерывного распределения $F(x)$ сводится к равномерному; несмотря на то, что аналоги теоремы Колмогорова [8] имеются и для кусочно-непрерывных распределений, применение этих результатов на практике весьма затруднительно. Использование теоремы 2.1 из [6] в данной ситуации представляется автору также мало возможным из-за трудности подсчета фигурирующей там величины β_n. На основании результатов данной работы при проверке гипотезы о соответствии наблюдений распределению (1) можно воспользоваться оценкой экспоненциального типа для вероятности уклонения D_n за уровень $r > 0$, установленной для случая кусочно-непрерывного гипотетического распределения.

Теорема. *Пусть $\{x_1, x_2, \ldots, x_n\}$ — независимые одинаково распределенные наблюдения над случайной величиной ε, имеющей функцию распределения $F_\xi(x) = \alpha F^{р}(x) + (1-\alpha)F^{н}(x)$, где $F^{н}(x)$ — непрерывная составляющая, $F^{р}(x)$ — чисто разрывная функция распределения. Тогда, если*

$$\int_{-\infty}^{+\infty} |x-a|^{m-2}\, dF^{р}(x) \leqslant \frac{\sigma^2 H^{m-2}}{2} m!,\quad m\geqslant 2$$

$\left(\text{здесь } a = \int\limits_{-\infty}^{+\infty} x dF^p(x)\right)$, *то справедлива оценка*

$$P\{D_n > r\} \leqslant 2\exp\left\{\frac{\delta^2}{4\alpha(1-\alpha)\sup\limits_x |F^н(x) - F^p(x)|^2}\right\} +$$

$$+ 4\left[1 - \alpha + \alpha\exp\left\{-\frac{(r-\delta)^2}{4\sigma^2\pi^2 n}\left(1 + 1{,}62\,\frac{r-\delta}{\pi\sigma^2}\sqrt{\frac{1}{2n}}\,H\right)^{-1}\right\}\right]^n +$$

$$+ 2C_1\left[\alpha + (1-\alpha)\exp\left\{-\frac{(r-\delta)^2}{n}(1-\varepsilon)^2\, 2\left(1 - \frac{C_2(r-\delta)}{n}\right)\right]^n, \quad (2)$$

если $r - \delta > C_0$. *Здесь* $C_0 = 2\mu(1-4/\mu)^{-1}$, $C_1 = \frac{11}{3}\mu$, $C_2 = \frac{4}{3}(1-\varepsilon)^{-2}\times$ $\times\,\varepsilon^{-1}(1-4/\mu)^{-1}$, μ — *наименьшее целое число, для которого* $\mu(1-4/\mu)^2\times$ $\times\,(1-1/\mu)^{-1} \geqslant 4\varepsilon^{-2}(1-\varepsilon)^2$, $0 < \varepsilon < 1$.

Замечание 1. Данный результат может иметь место в то время, как условие

$$\int\limits_{-\infty}^{+\infty} |x - a|^m dF(x) \leqslant \frac{\sigma^2 H^{m-2}}{2} m!, \quad m \geqslant 2,$$

соответствующих теорем из [6] не выполнено. Примером служит смесь некоторого дискретного распределения с распределением Коши.

Замечание 2. Оценки в правой части (2) экспоненциального типа. Это следует из элементарного неравенства

$$[p + qe^{\lambda/n}]^n \leqslant \exp\left\{-q\lambda\left(1 - \frac{\lambda}{2n}\right)\right], \quad p \geqslant 0,\ q \geqslant 0,\ p + q = 1.$$

Доказательство теоремы базируется на работах [6, 7]. Ключевым моментом доказательства является использование неравенства Эссена для оценки супремума модуля разности чисто разрывных функций, что позволило свести оценивание уклонений $F_n(x)$ и $F(x)$ в равномерной метрике к оцениванию уклонений соответствующих характеристических функций в пространстве L_2. При построении экспоненциальных оценок для вероятности уклонения статистики Колмогорова этот подход, по-видимому, применяется впервые.

Итак, пусть ν — число значений выборки $\{x_1, x_2, \ldots, x_n\}$, попавших на точки разрыва $F(x)$, $n - \nu$ — число значений выборки, попавших в интервалы непрерывности $F(x)$.

Пусть $F^p_\nu(x)$ — эмпирическая функция распределения, построенная по значениям выборки, попавшим на точки разрыва $F(x)$, $F^н_{n-\nu}(x)$ — эмпирическая функция распределения, построенная по значениям выборки, попавшим на интервалы непрерывности $F(x)$. Тогда для $0 < \delta < r$ имеем

$$P\{D_n > r\} \leqslant P\left\{\sqrt{n}\left|\frac{\nu}{n} - \alpha\right| > \frac{\delta}{\sup\limits_x |F^н(x) - F^p(x)|}\right\} +$$

$$+ \sum_{k=0}^{n} P\left\{\sqrt{k}\sup_x |F^p_k(x) - F^p(x)| > (r-\delta)\sqrt{\frac{k}{n}}\right\} P\{\nu = k\} +$$

$$+ \sum_{k=0}^{n} P\left\{\sqrt{n-k}\sup_x |F^н_{n-k}(x) - F^н(x)| > (r-\delta)\sqrt{\frac{n-k}{n}}\right\} p\{\nu = k\}.$$

В самом деле,

$$D_n \leqslant \sqrt{n}\left|\frac{\nu}{n} - \alpha\right| \sup_x |F^н(x) - F^p(x)| +$$

$$+\sqrt{\frac{\nu}{n}}\sqrt{\nu}\sup_x|F^p_\nu(x)-F^p(x)|+\sqrt{\frac{n-\nu}{n}}\sqrt{n-\nu}\sup_x|F^н_{n-\nu}(x)-F^н(x)|. \tag{3}$$

Далее, применяя неравенство $P\{|\xi|+|\eta|>r\}\leqslant P\{|\xi|>\beta r\}+P\{|\eta|>(1-\beta)r\}$ для $0<\beta<1$ и формулу полной вероятности, получаем (3).

Нетрудно заметить, что случайная величина ν имеет распределение Бернулли $P\{\nu=k\}=C^k_n\alpha^k(1-\alpha)^{n-k}$, $k=0, 1, \ldots, n$. Для оценки первого слагаемого в (3) можно воспользоваться, например, результатом из [9], в силу которого

$$P\left\{\sqrt{n}\left|\frac{\nu}{n}-\alpha\right|>\frac{\delta}{\sup_x|F^н(x)-F^p(x)|}\right\}=$$
$$=P\left\{\left|\sum_{i=1}^n(\xi_i-\alpha)\right|>\frac{\sqrt{n}\delta}{\sup_x|F^н(x)-F^p(x)|}\right\}\leqslant$$
$$\leqslant 2\exp\left\{-\frac{\delta^2}{\sup_x|F^н(x)-F^p(x)|^2\,4\alpha(1-\alpha)}\right\}, \tag{4}$$

если $0<\delta<\sqrt{n}\alpha(1-\alpha)\sup_x|F^н(x)-F^p(x)|$. В данном случае

$$|M(\xi_i-\alpha)^m|=|(1-\alpha)^m\alpha+\alpha^m(1-\alpha)|\leqslant\frac{\alpha(1-\alpha)m!}{2}$$

для $m\geqslant 2$, $B=n\alpha(1-\alpha)$, $H=1$. Из результатов работы [7] следует

$$P\{\sqrt{l}\sup|F^н_l(x)-F^н(x)|>\lambda\}\leqslant C_1\exp\left\{-\lambda^2(1-\varepsilon)^2\,2\left(1-\frac{C_2\lambda}{\sqrt{l}}\right)\right\}, \tag{5}$$

для $\lambda>C_0$, C_0, C_1, C_2 — постоянные, фигурирующие в условиях теоремы. С учетом (5) имеем

$$\sum_{k=0}^n P\left\{\sqrt{n-k}\sup_x|F^н_{n-k}(x)-F^н(x)|>(r-\delta)\sqrt{\frac{n-k}{n}}\right\}P\{\nu=k\}=$$
$$=C_1\sum_{k=0}^n C^k_n\left[\exp\left\{-\frac{(r-\delta)^2}{n}(1-\varepsilon)^2\,2\left(1-\frac{C_2(r-\delta)}{\sqrt{n}}\right\}(1-\alpha)\right]^{n-k}\alpha^k=$$
$$=C_1\left[\alpha+(1-\alpha)\exp\left\{-\frac{2(r-\delta)^2(1-\varepsilon)^2}{n}\left(1-\frac{C_2(r-\delta)}{n}\right)\right]^n. \tag{6}$$

Оценим последнее слагаемое в (3). Пусть $\Psi^p(x)$ — чисто разрывная функция распределения, имеющая такое же количество скачков, что и $F^p(x)$, причем величины скачков $\Psi^p(x)$ и $F^p(x)$ одинаковы и располагаются в таком же порядке следования. Скачки же $\Psi^p(x)$ находятся в точках $(1, 2, \ldots,)$ оси OX. Обозначим через $\Psi^p_k(x)$ эмпирическую функцию распределения для $\Psi^p(x)$, построенную по $F^p_k(x)$. Легко заметить, что $\sup_x|F^p_k(x)-F^p(x)|=$ $=\sup_x|\Psi^p_k(x)-\Psi^p(x)|$.

Пусть $f(z)=\sum_{l=-\infty}^{+\infty}e^{ilz}p_l$, где p_l — величина скачка $\Psi^p(x)$ в l-й точке разрыва. Пусть $f_k(z)=\frac{1}{k}\sum_{j=1}^k e^{izl_j}$ — эмпирическая характеристическая функция, построенная по значениям выборки, попавшим на точки разрыва. Используя результат И. П. Цареградского [9], имеем

$$\sqrt{k}\sup_x|F_k^p(x)-F^p(x)|\leqslant\frac{1}{4}\int_{-\pi}^{\pi}\frac{\sqrt{k}\,|f_k(z)-f(z)|}{|z|}\,dz.$$

Таким образом,

$$P\{\sqrt{k}\sup_x|F_k^p(x)-F^p(x)|>\lambda\}\leqslant P\left\{\int_{-\pi}^{\pi}\frac{\sqrt{k}\,|f_k(z)-f(z)|}{|z|}\,dz>4\lambda\right\}\leqslant$$

$$\leqslant P\left\{\int_{-\pi}^{\pi}\frac{k\,|f_k(z)-f(z)|^2}{|z|^2}\,dz>\frac{8\lambda^2}{\pi}\right\}.$$

Так как $K|f_k(z)-f(z)|\,|z|^{-2}=[X_k^2(z)+Y_k^2(z)]\,k^{-1}$, где

$$X_k^2(z)=\left(\sum_{i=1}^{k}\frac{[\cos l_iz-M\cos l_jz}{z}\right)^2,\quad Y_k^2(z)=\left(\sum_{i=1}^{k}\frac{[\sin l_iz-M\sin l_iz}{z}\right),$$

то

$$P\left\{\sqrt{k}\sup_x|F_k^p(x)-F^p(x)|>\sqrt{\frac{k}{n}}\,(r-\delta)\right\}\leqslant$$

$$\leqslant P\left\{\left[\int_{-\pi}^{\pi}X_k^2(z)\,dz\right]^{1/2}>2\,\frac{k(r-\delta)}{\sqrt{\pi n}}\right\}+P\left\{\left[\int_{-\pi}^{\pi}Y_k^2(z)\,dz\right]^{1/2}>2\,\frac{k(r-\delta)}{\sqrt{\pi n}}\right\}.$$

Для того чтобы воспользоваться результатом работы [6], вычислим оценки моментов $M\left|\dfrac{\cos l_1z-M\cos l_1z}{z}\right|^m$, $M\left|\dfrac{\sin l_1z-M\sin l_1z}{z}\right|^m$ при $m\geqslant 2$.

Пусть $a=\sum\limits_{k=-\infty}^{+\infty}kp_k=\int\limits_{+\infty}^{\infty}xdF^p(x)$, $\sum\limits_{k=-\infty}^{+\infty}|k-a|^m\,p_k\leqslant\dfrac{\sigma^2H^{m-2}}{2}\,m!$, $m=2,3,\ldots$. Нетрудно убедиться в том, что

$$M\left|\frac{\cos l_1z-M\cos l_1z}{z}\right|^m\leqslant\sum_k\sum_j\left|\frac{\cos kz-\cos jz}{z}\right|^m p_kp_j\leqslant$$

$$\leqslant\sum_k\sum_j|k-j|^m\,p_kp_j\leqslant 2^m\sum|k-a|^m\,p_k\leqslant$$

$$\leqslant\frac{2^m\sigma^2H^{m-2}}{2}\,m!=\frac{(2\sigma)^2(2H)^{m-2}}{2}\,m!$$

Аналогично,

$$M\left|\frac{\sin l_1z-M\sin l_1z}{z}\right|^m\leqslant\frac{(2\sigma)^2(2H)^{m-2}}{2}\,m!$$

Тогда

$$M\left[\int_{-\pi}^{\pi}\left[\frac{\cos l_1z-M\cos l_1z}{z}\right]dz\right]^{m/2}\leqslant\frac{(\sqrt{2\pi}2\sigma)^2(2\sqrt{2\pi}H)^{m-2}}{2}\,m!$$

Аналогично,

$$M\left[\int_{-\pi}^{\pi}\left(\frac{\sin l_1z-M\sin l_1z}{z}\right)^2dz\right]^{m/2}\leqslant\frac{(\sqrt{2\pi}2\sigma)^2(2\sqrt{2\pi}H)^{m-2}}{2}\,m!$$

и

$$P\left\{\sqrt{k}\sup_x|F_k^p(x)-F^p(x)|>\sqrt{\frac{k}{n}}\,(r-\delta)\right\}\leqslant$$

$$\leqslant P\left\{\left[\int_{-\pi}^{\pi}X_k^2(z)\,dz\right]^{1/2}>(\sqrt{2k}\pi 2\sigma)\frac{r-\delta}{\sigma\pi}\sqrt{\frac{k}{2n}}\right\}+$$

$$+P\left\{\left[\int_{-\pi}^{\pi} Y_k^2(z)\,dz\right]^{1/2} > (\sqrt{2k}\pi 2\sigma)\frac{r-\delta}{\sigma\pi}\sqrt{\frac{k}{2n}}\right\} \leqslant$$

$$\leqslant 4\exp\left\{-\frac{(r-\delta)^2}{4\sigma^2\pi^2}\frac{k}{n}\left[1+1{,}62\frac{r-\delta}{\sigma^2\pi}\frac{H}{\sqrt{2n}}\right]^{-1}\right\}$$

и

$$\sum_{k=0}^{n} P\left\{\sqrt{k}\sup_x |F_k^p(x)-F^p(x)| > (r-\delta)\sqrt{\frac{k}{n}}\right\} P\{\nu=k\} \leqslant$$

$$\leqslant 4\left[1-\alpha+\alpha\exp\left\{-\frac{(r-\delta)^2}{4\sigma^2\pi^2}\left[1+1{,}62\frac{r-\delta}{\sigma^2\pi}\frac{H}{\sqrt{2n}}\right]^{-1}\right\}\right]. \qquad (7)$$

Из (7), (6) и (4) следует (2).

1. *Пугачев В. С.* Теория вероятностей и математическая статистика.— М. : Наука, 1979.— 496 с.
2. *Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д.* Математические методы в теории надежности.— М. : Наука, 1969.— 524 с.
3. *Симонова Г. И.* Устойчивые оценки параметра экспоненциального распределения // Изв. АН СССР. Техн. кибернетика.— 1981.— № 1.— С. 123—129.
4. *Гихман И. И., Скороход А. В.* Теория вероятностей и математическая статистика.— Киев : Вища шк., 1979.— 408 с.
5. *Кендалл М., Стьюарт Л.* Статистические выводы и связи.— М. : Наука, 1973.— 909 с.
6. *Yurinskii V. V.* Exponential inequalities for sums of random vectors // J. Multivar. Anal.— 1976.— N 6.— P. 473—499.
7. *Юринский В. В.* О неравенствах больших уклонений некоторых статистик // Теория вероятностей и ее применения.— 1971.— **14**, № 2.— С. 386—389.
8. *Schmid P.* On the Kolmogorov and Smirnov theorems for discontinuous distribution functions // Ann. Math., Statist.— 1958.— **29**.— P. 1011—1027.
9. *Петров В. В.* Суммы независимых случайных величин.— М. : Наука, 1972.— 416 с.

Донец. ун-т

Получено 17.12.85,
после доработки — 26.05.86