ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

ОЦЕНИВАНИЕ РАСПРЕДЕЛЕНИЙ ПО ВЫБОРКАМ СЛУЧАЙНОГО ОБЪЕМА

Тихов М.С.
Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, г. Нижний Новгород

Поступила в редакцию 24.11.2023, после переработки 30.11.2023.

Статья посвящена задаче оценивания функции распределения и предельного поведения расстояния между эмпирическим и теоретическим законами, именно, суммируемых квадратичных уклонений и статистики Смирнова и Колмогорова по выборкам случайного объема. Мы предполагаем, что этот случайный объем имеет обобщенное отрицательное биномиальное распределение и как случайная величина не зависит от исходной выборки. Найдены предельные распределения для суммируемых квадратичных уклонений ядерных оценок функции распределения по выборкам случайного объема. Показано, что для выборок случайного объема предельное распределение статистик Смирнова и Колмогорова имеет более тяжелые хвосты, чем у функции распределения Вейбулла и Колмогорова в случае выборок фиксированного объема. Мы предлагаем подход на основе асимптотического разложения, чтобы естественным образом сбалансировать асимптотическое распределение и случайный объем выборки. Рассмотрена также задача последовательного оценивания параметра сдвига равномерного распределения. Отрицательное биномиальное распределение (объем выборки ν) возникает здесь естественным образом в результате статистического эксперимента по выполнению серии независимых испытаний.

Ключевые слова: выборка случайного объема, эмпирическая функция распределения, статистика Смирнова и Колмогорова, последовательное оценивание, обобщенное отрицательное распределение.

Вестник ТвГУ. Серия: Прикладная математика. 2023. № 4. С. 5–24.
https://doi.org/10.26456/vtpmk695

Введение

В задачах математической статистики объем выборки обычно считается детерминированным, фиксированным заранее. В то же время в немалом числе

ситуаций объем выборки ν является случайной величиной. Это: последовательная проверка гипотез [1], последовательное оценивание неизвестного параметра распределения [2]-[4] – там ν является моментом остановки, т.е. зависит от наблюдаемых случайных величин. Рассматриваются также ситуации (см.[5]-[11]), когда случайный объем не зависит от наблюдаемых случайных величин. Случайность объема выборки приводит к тому, что предельные нормированные суммы (обычно предполагается, что $\mathbf{E}(\tau) \leqslant n$) могут уже не иметь нормального распределения – они могут иметь распределение Лапласа, или распределение Стьюдента [7], [9], поэтому доверительный интервал для функции распределения или параметра (при одинаковой надежности интервала) надо брать шире, чем при фиксированном объеме выборки. В пп. 1,2 мы будем рассматривать суммы случайного объема, когда τ не зависит от наблюдаемых величин. В качестве распределения для с.в. τ обычно рассматриваются либо геометрическое распределение, либо отрицательное биномиальное (**NB**). В п. 3 мы показываем, что при последовательном оценивании параметра сдвига, равномерного на интервале $(\theta - 1/2, \theta + 1/2)$, оптимальный момент остановки (см. [4]) имеет **NB**-распределение, он определяется по результатам выборки, и находим предельное распределение последовательной оценки.

1. Ядерные оценки функции распределения и суммируемые квадратичные уклонения

Пусть $T_m = T_m(X_1, ..., X_m)$, $m \in \mathbf{N}$ – статистика, построенная по выборке $\mathcal{X}^{(m)} = \{X_1, X_2, ..., X_m\}$ неслучайного объема $m \in \mathbf{N}$.

Рассмотрим последовательность дискретных случайных величин $\tau_1, \tau_2, ..., \tau_n, ...,$ зависящих от натурального n и принимающих натуральные значения. Натуральное значение $\tau_n = k \geqslant 1$ есть объем выборки $\mathcal{X}^{(k)}$. Будем предполагать, что при каждом n с.в. τ_n не зависит от $(X_1, X_2, X_3, ...)$. Определим T_{τ_n}, полагая $T_{\tau_n} = T_k$, на множестве $(\tau_n = k)$, $k \in \mathbf{N}$.

Предположения.

A1. *Существует функция распределения* $F(x)$, *такая, что для некоторого* $\gamma > 0$,

$$\sup_x |\, \mathbf{P}(m^\gamma T_m < x) - F(x) \,| \underset{m\to\infty}{\longrightarrow} 0.$$

A2. *Существуют функция распределения* $H(x)$, $H(+0) = 0$ *и последовательность чисел* $0 < g_n \uparrow \infty$, *такие, что*

$$\sup_y |\, \mathbf{P}(g_n^{-1}\tau_n < y) - H(y) \,| \underset{n\to\infty}{\longrightarrow} 0.$$

Эти условия взяты из работы [12] где показано, что имеет место следующий результат.

Теорема 1. *Пусть даны:* γ, *статистика* T_m *и случайный объем выборки* τ_n, *а также выполнены предположения* **A1** *и* **A2**. *Тогда*

$$\sup_{x\in\boldsymbol{R}} |\, \boldsymbol{P}(g_n^\gamma T_{\tau_n} < x) - G_n(x, 1/g_n) \,| \underset{n\to\infty}{\longrightarrow} 0, \ \textit{где} \quad G_n(x, 1/g_n) = \int_{1/g_n}^{\infty} F(xy^\gamma)\, dH(y).$$

В отличие от [12], где рассматривались значения $\gamma \in \{-1, -1/2, 0, 1/2, 1\}$, здесь мы рассматриваем значение $\gamma = 4/5$.

Пусть с.в. σ_n имеет геометрическое распределение $\mathbf{P}(\sigma_n \geqslant k) = q = \left(1 - \frac{1}{n^\beta}\right)^k$, $k = 1, 2, ...$, а $\tau_n = \sigma_n^{1/\beta}, \beta > 0$. Тогда $\mathbf{P}(\tau_n \geqslant k) = \mathbf{P}(\sigma_n \geqslant k^\beta) = q^{k^\beta}$, т.е. τ_n имеет дискретное распределение Вейбулла и поэтому $\mathbf{P}(\tau_n/n \geqslant y) \underset{n\to\infty}{\longrightarrow} e^{-y^\beta} = 1 - H(y)$, $y > 0$, а это есть непрерывное распределение Вейбулла.

Возьмем $\beta = 2\gamma = 8/5$. Пусть $g_n = n$. Тогда (см. [12])

$$\mathbf{P}(g_n^\gamma T_{\tau_n} \leqslant x) = \mathbf{P}(\tau_n^\gamma T_{\tau_n} \leqslant x(\tau_n/n)^\gamma) = \sum_{m=1}^{\infty} \mathbf{P}(m^\gamma T_m \leqslant x(m/n)^\gamma)\, \mathbf{P}(\tau_n = m) \approx$$

$$\approx \mathbf{E}(F(x(\tau_n/n)^\gamma) = \int_{1/g_n}^{\infty} F(xy^\gamma) d\mathbf{P}(\tau_n/n < y) \approx \int_{1/g_n}^{\infty} F(xy^\gamma)\, dH(y).$$

Пусть $F(x) = \Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} \exp(-t^2/2)\, dt$ и $H(y) = 1 - \exp(-y^{2\gamma}),\ y > 0$, а $h(y) = H'(y) = 2\gamma y^{2\gamma-1} \exp(-y^{2\gamma})$. В этом случае при указанном выборе $\mathbf{P}(g_n^\gamma T_{\tau_n} < x) \approx \int_0^\infty \Phi(xy^\gamma)\, dH(y)$ с соответствующей предельной плотностью t_2-распределения Стьюдента:

$$w_\gamma(x; s) = \frac{1}{\sqrt{2\pi}} \int_0^\infty y^\gamma e^{-x^2 y^{2\gamma}/2}\, d(1 - \exp(-y^{2\gamma})) = \frac{1}{2\sqrt{2}} \left(1 + \frac{x^2}{2}\right)^{-3/2}, \quad x \in \mathbf{R}.$$

Рассмотрим еще одно дискретное распределение случайной величины τ_n:

$$\mathbf{P}(\tau_n \leqslant k) = \exp\left(-\frac{sn}{k^\alpha}\right),\ k, n \in \mathbf{N},\ \alpha > 0,\ s > 0. \tag{1}$$

Здесь $\mathbf{P}(\tau_n/n^{1/\alpha} \leqslant x) = \mathbf{P}(\tau_n \leqslant x/n^{1/\alpha}) \underset{m\to\infty}{\longrightarrow} e^{-s/x^\alpha} = H(y)$ и $g_n = n^{1/\alpha}$. Положим $\alpha = 2\gamma$. Тогда

$$\mathbf{P}(\tau_n^\gamma T_{\tau_n} \leqslant x((\tau_n/g_n)^\gamma \approx \int_0^\infty \Phi(xy^\gamma)\, d(e^{-s/y^{2\gamma}}),$$

с соответствующей предельной плотностью распределения Лапласа:

$$w_\gamma(x; s) = \frac{s}{\sqrt{2\pi}} \int_0^\infty y^\gamma\, e^{-x^2 y^{2\gamma}/2 - s/y^{2\gamma}}\, d\left(\frac{1}{y^{2\gamma}}\right) = \frac{s}{\sqrt{2\pi}} \int_0^\infty t^{-3/2}\, e^{-x^2 t/2 - s/t}\, dt =$$

$$= \frac{\sqrt{2s}}{2}\, e^{-\sqrt{2s}\,|\,x\,|},\ x \in \mathbf{R}.$$

Применим полученные результаты к задаче оценивания неизвестной функции распределения в зависимости «доза-эффект» по выборке случайного объема. Пусть $\mathcal{X}^{(n)} = \{X_i, 1 \leqslant i \leqslant n\}$ – последовательность независимых случайных величин с функцией распределения $F(x), x \in \mathbf{R}^1, \mathbf{P}(0 < X_i < 1) = 1$ и плотностью распределения $f(x)$, а $\{u_i \in (0,1), 1 \leqslant i \leqslant n\}$ – числовая последовательность. Мы наблюдаем выборку $\mathcal{W}^{(n)} = \{(u_i, W_i), 1 \leqslant i \leqslant n\}$, где $W_i = I(X_i < u_i)$ – индикатор события $(X_i < u_i)$. Требуется оценить неизвестную функцию распределения

$F(x)$ по выборке $\mathcal{W}^{(n)}$. В качестве оценки функции распределения $F(x)$ возьмем ядерную оценку

$$\hat{F}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} W_i K\left(\frac{x-u_i}{h}\right). \tag{2}$$

Здесь

1. $K(x) \geqslant 0$;
2. $\int_{-\infty}^{\infty} K(x)\,dx = 1$;
3. $K(x) = K(-x)$;
4. $||\,K\,||^2 = \int_{-\infty}^{\infty} K^2(x)\,dx < \infty$

и выполнены некоторые дополнительные условия (см. [15]).

Рассмотрим статистику $J_{nm} = \frac{1}{m}\sum\limits_{j=1}^{m}(F_n(x_j) - F(x_j))^2$ – суммируемое квадратичное уклонение для m заданных точек $(x_1, x_2, ..., x_m)$. Введем обозначения:

$$\sigma_1^2 = \frac{\mu_2^2(K)\|K\|^2}{4m^2}\sum_{j=1}^{m} F(x_j)(1-F(x_j))(f'(x_j))^2, \quad \sigma_3^2 = \frac{\beta_2}{m^2}\sum_{j=1}^{m} F^2(x_j)(1-F(x_j))^2,$$

$$\mu_2(K) = \int_{-1}^{1} x^2 K(x)\,dx,\ \|K\|^2 = \int_{-1}^{1} K^2(x)\,dx,\ \beta_2 = \iint_{\mathbf{R}^2} K^2(u)K^2(u+z)\,dudz.$$

Пусть $nh^5 \to \lambda \in (0,\infty)$ при $n \to \infty$. При условиях работы [15] имеем:

$$T_n = n^{4/5}(J_{nm} - \mathbf{E}(J_{nm})(\lambda^{4/5}\sigma_1^2 + \lambda^{-1/5}\sigma_3^2)^{1/2} \xrightarrow[n\to\infty]{d} \zeta \in N(0,1).$$

Значит, если объем выборки τ_n случаен и имеет дискретное распределение Вейбулла, то из предыдущих рассуждений получаем, что предельным распределением нормированных разностей оценок будет распределение Стьюдента t_2 с ф.р. $G(x) = \frac{1}{2} + \frac{x}{2\sqrt{2}}\,{}_2F_1\left(\frac{1}{2},\frac{3}{2};\frac{3}{2};-\frac{x^2}{2}\right)$, где ${}_2F_1\,(a,b;c;z)$ есть гипергеометрическая функция Гаусса, а при выборе распределения (1) в качестве предельного получим распределение Лапласа, у которых более тяжелые хвосты, чем у нормального распределения. В подтверждение этого рассмотрим плотности:

$$\omega_\gamma(x;s) = \frac{s}{\sqrt{2\pi}}\int_0^\infty y^{\gamma-2}\,e^{-(x^2y^{2\gamma}/2+s/y)}\,dy, \quad \rho_\lambda(x) = \frac{1}{\sqrt{2\pi}}\int_0^\infty y^{\gamma}\,e^{-(x^2y^{2\gamma}/2+y)}\,dy$$

и функции распределения $\Omega_\gamma(x;s) = \int_{-\infty}^{x}\omega_\gamma(z;s)\,dz$, $L(x) = \frac{1}{\sqrt{2}}\int_{-\infty}^{x} e^{-\sqrt{2}\,|z|}\,dz$, $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}\,dz$, $R_\gamma(x) = \int_{-\infty}^{x}\rho_\lambda(z)\,dz$. Тогда

$$\Omega_{2/5}(2.01424) = 0.975, \quad \Phi(1.96) = 0.975, \quad L(2.119) = 0.975, \quad R_{2/5}(3.454) = 0.975.$$

2. Эмпирическая функция распределения и статистики Смирнова и Колмогорова

Разлагая функцию $(1-\lambda)^{-a},\ 0<\lambda<1,\ a>0$ в ряд Тейлора по переменной λ в окрестности нуля, получим

$$(1-\lambda)^{-a}=1+a\lambda+\frac{a(a+1)}{2!}\lambda^2+\frac{a(a+1)(a+2)}{3!}\lambda^3+\dots\,, \tag{3}$$

поэтому определим обобщенное отрицательное биномиальное распределение (распределение Пойа) следующим образом:

$$\mathbf{P}(\nu_\lambda=1)=(1-\lambda)^a,\ \mathbf{P}(\nu_\lambda=k)=\frac{a(a+1)\cdot\ldots\cdot(a+k-2)}{(k-1)!}(1-\lambda)^a\lambda^{k-1},\ k=2,3,\dots\,.$$

Пусть $F_n(x)$ есть эмпирическая функция распределения (ф.р.), построенная по повторной выборке $\mathcal{X}^{(n)}$, $F(x)=\mathbf{P}(X_1<x)$ – теоретическая ф.р., а с.в. ν_λ не зависит от $\mathcal{X}^{(n)}$ для $n\in\mathbf{N}$, $F_{nj}=F_n(x_j)$, $p_j=F(x_j)$, $q_j=1-p_j$, $\alpha_j=nF_{nj}-np_j$, $\beta_{n,j}=\alpha_{n,j}/\sqrt{p_jq_j}$. Тогда

$$\varphi_{\alpha_{n,j}}(t)=e^{-itnp_j}(p_j\,e^{it}+q_j)^n=(p_j\,e^{itq_j}+q_j\,e^{-itp_j})^n,$$

$$\varphi_{\beta_{n,j}}(t)=\left(p_j\,e^{it\sqrt{\frac{q_j}{p_j}}}+q_j\,e^{-it\sqrt{\frac{p_j}{q_j}}}\right)^n=\psi_j^n(t)=\left(1-\frac{t^2}{2}+O(|\,t\,|^3)\right)^n,\ t\to0.$$

Поэтому

$$\begin{aligned}\varphi_{\beta_{\nu_\lambda,j}}(t)&=\mathbf{E}(e^{it\beta_{\nu_\lambda,j}})=\sum_{k=1}^{\infty}\mathbf{E}(e^{it\beta_{\nu_\lambda,j}}|\,\nu_\lambda=k)\,\mathbf{P}(\nu_\lambda=k)=\\&=\sum_{k=1}^{\infty}\mathbf{E}(e^{it\beta_{k,j}}|\nu_\lambda=k)\,\mathbf{P}(\nu_\lambda=k)=\\&=\sum_{k=1}^{\infty}\mathbf{E}(e^{it\beta_{k,j}})\,\mathbf{P}(\nu_\lambda=k)=\sum_{k=1}^{\infty}\mathbf{E}(e^{it\beta_{k,j}})\,\mathbf{P}(\nu_\lambda=k)=\\&=\psi_j(t)\,(1-\lambda)^a\sum_{k=1}^{\infty}\frac{a(a+1)\cdot\ldots\cdot(a+k-2)}{(k-1)!}\,(1-\lambda)^a(\lambda\psi_j(t))^{k-1}=\\&=\psi_j(t)\,(1-\lambda)^a\,(1-\lambda\psi_j(t))^{-a}=\psi_j(t)\,\left(\frac{1-\lambda}{1-\lambda\psi_j(t)}\right)^a.\end{aligned}$$

Значит, при $\lambda\to1$,

$$\begin{aligned}\varphi_{\sqrt{1-\lambda}\beta_{\nu_\lambda,j}}(t)&=\varphi_{\beta_{\nu_\lambda,j}}(t\sqrt{1-\lambda})=\\&=\psi_j(t\sqrt{1-\lambda})\left(\frac{1-\lambda}{1-\lambda\psi_j(t\sqrt{1-\lambda})}\right)^a\to\left(\frac{2}{2+t^2}\right)^a=\varphi_a(t),\end{aligned}$$

а $\zeta_a/\sqrt{2}$ имеет характеристическую функцию $\varphi_a(t)$ и плотность распределения (см. [7], с.261, 2.3.(**7**))

$$f_a(x) = \frac{1}{\Gamma(a)\sqrt{\pi}} \left(\frac{|\,x\,|}{2}\right)^{a-1/2} K_{a-1/2}(|\,x\,|), \quad x \in \mathbf{R}, \tag{4}$$

где $K_\alpha(x)$ – функция Макдональда.

Рассмотрим плотность (4) при натуральных a. Если $a = 1$, то мы имеем плотность распределения Лапласа вида (с.в. ξ_1)

$$f_1(x) = \frac{1}{2}\, e^{-|\,x\,|}, \quad x \in \mathbf{R}.$$

Когда $a = 2$, эта плотность равна

$$f_2(x) = \frac{1 + |\,x\,|}{4}\, e^{-|\,x\,|}, \;\; x \in \mathbf{R},$$

(плотность распределения с.в. $\varsigma_2 = \xi_1 + \xi_2$, где ξ_1, ξ_2 – независимы), для $a = 3$ мы имеем плотность распределения суммы трех стандартных независимых показательных случайных величин $\varsigma_3 = \xi_1 + \xi_2 + \xi_3$:

$$f_3(x) = \frac{3 + 3|\,x\,| + x^2}{16}\, e^{-|\,x\,|}, \;\; x \in \mathbf{R}.$$

Из (3) имеем также:

$$\varphi_{\nu_\lambda}(t) = e^{it}(1-\lambda)^a \sum_{k=1}^{\infty} \frac{a(a+1)\cdot\ldots\cdot(a+k-2)}{(k-1)!}(\lambda e^{it})^{k-1} = e^{it}\left(\frac{1-\lambda}{1-\lambda e^{it}}\right)^a.$$

Отсюда, при $\lambda \to 1$,

$$\varphi_{(1-\lambda)\nu_\lambda}(t) = \varphi_{\nu_\lambda}((1-\lambda)t) = e^{it(1-\lambda)}\left(\frac{1-\lambda}{1-\lambda e^{it(1-\lambda)}}\right)^a \underset{\lambda\to 1}{\longrightarrow} \left(\frac{1}{1-it}\right)^a,$$

а это есть характеристическая функция гамма-распределения с плотностью $g(x) = \dfrac{x^{a-1}}{\Gamma(a)}\, e^{-x}, \; x > 0.$

Таким образом, предельной плотностью распределения величины W будет распределение Накагами:

$$h_a(x) = \frac{2x^{2a-1}}{\Gamma(a)}\, e^{-x^2}, \; x > 0, \tag{5}$$

с функцией распределения $H_a(y) = \dfrac{\gamma(a, x^2)}{\Gamma(a)}, x > 0$, где $\gamma(a, x)$ есть неполная гамма-функция.

Из приведенных выше рассуждений следует, что если $Z \in N(0,1)$, то величина $W \cdot Z$ будет иметь плотность распределения (5).

Пусть $F(x_j) = p_j$, $j = 1, 2, ..., s$, $p_1 < p_2 < ... < p_s$, $\sigma_{jk} = p_j(1 - p_k)$, $j \le k$, $\rho_{jk} = \sigma_{jk}/\sqrt{\sigma_{jj}\sigma_{kk}}$, $\Sigma = (\rho_{jk})_{s\times s}$, $b_j = \nu_q(F_n(x_j) - p_j)/\sigma_{jj}$, $\mathbf{b} = (b_1, ..., b_s)'$, $\mathbf{t} = (t_1, ..., t_s)' \in \mathbf{R}^s$. Рассматривая теперь линейные комбинации представленных величин, нетрудно получить следующую теорему.

Теорема 2. *Если $F(x)$ непрерывная функция распределения, то*

$$\eta_\lambda = \sqrt{1-\lambda}\cdot \boldsymbol{b} \xrightarrow[\lambda\to 1]{d} \zeta_a,$$

случайный вектор ζ_a имеет характеристическую функцию $\varphi_a(\boldsymbol{t}) = \left(\dfrac{2}{2+\boldsymbol{t}'\Sigma\boldsymbol{t}}\right)^a$ и эллиптически контурированную плотность распределения

$$f_a(\boldsymbol{x}) = \frac{2^{s/2}}{\Gamma(a)\,\pi^{s/2}|\,\Sigma\,|^{1/2}}\left(\frac{\sqrt{\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}}}{\sqrt{2}}\right)^{a-s/2} K_{a-s/2}(\sqrt{2\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}}),\ \boldsymbol{x}\in\boldsymbol{R}^s.$$

Далее, из [8, с. 370, Теорема 1], имеем: если $\{w(t), 0\leqslant t\leqslant T\}$ – процесс броуновского движения, то

$$\mathbf{P}\left(\max_{0\leqslant t\leqslant T} w(t) > a > 0, w(T)\in[c,d\,]\right) = \frac{1}{\sqrt{2\pi T}}\int\limits_{\max[c,a\,]}^{\max[d,a\,]} e^{-\frac{x^2}{2T}}\,dx + \\ + \frac{1}{\sqrt{2\pi T}}\int\limits_{\max[2a-c,a\,]}^{\max[2a-d,a\,]} e^{-\frac{x^2}{2T}}\,dx, \quad (6)$$

а (см. [8], с. 635)

$$\lim_{n\to\infty}\mathbf{P}\left(\sqrt{n}\sup_{-\infty<x<\infty}(F_n(x)-F(x))<\alpha\right) = \mathbf{P}\left(\sup_{0\leqslant t\leqslant 1}\beta(t)<\alpha\right) = 1-e^{-2\alpha^2},\ \alpha>0,$$

где $\{\beta(t), 0\leqslant t\leqslant 1\}$ есть *броуновский мост.*

Пусть $\Delta_n^+ = \sqrt{n}\sup\limits_{-\infty<x<\infty}(\,F_n(x)-F(x)\,)$. Так как *броуновский мост* есть *гауссовский* процесс, то предельную функцию распределения величины $\sqrt{(1-\lambda)\nu_\lambda}\,\Delta_{\nu_\lambda}^+$ при $\lambda\to 1$ найдем как следующий интеграл (см.[9],с.354,3.471 (**9**)):

$$1-\int\limits_0^\infty \frac{2w^{2a-1}}{\Gamma(a)}\exp\left(-\left(\frac{2x^2}{w^2}+w^2\right)\right)dw = 1-\int\limits_0^\infty\frac{u^{a-1}}{\Gamma(a)}\exp\left(-\left(\frac{2x^2}{u}+u\right)\right)du = \\ = 1-2\frac{(\sqrt{2}\,x)^a}{\Gamma(a)}\,K_a(2\sqrt{2}\,x),\ x>0. \quad (7)$$

Аналогично показывается, что если $\Delta_n = \sqrt{n}\sup\limits_{-\infty<x<\infty}|\,F_n(x)-F(x)\,|$, то

$$\mathbf{P}\left(\sqrt{(1-\lambda\,)\nu_\lambda}\,\Delta_{\nu_\lambda}\,<x\right)\underset{\lambda\to 1}{\longrightarrow} 1-2\frac{(\sqrt{2}\,x)^a}{\Gamma(a)}\sum_{j=1}^\infty(-1)^{j-1}j^a\,K_a(2\sqrt{2}\,jx),\ x>0.$$

Теорема 3. *При условиях теоремы* 2,

$$\text{(i)}\ \boldsymbol{P}\left(\sqrt{(1-\lambda\,)\nu_\lambda}\,\Delta_{\nu_\lambda}^+\,<x\right)\underset{\lambda\to 1}{\longrightarrow} 1-2\frac{(\sqrt{2}\,x)^a}{\Gamma(a)}\,K_a(2\sqrt{2}\,x),\ x>0, \quad (8)$$

$$\text{(ii)}\ \boldsymbol{P}\left(\sqrt{(1-\lambda\,)\nu_\lambda}\,\Delta_{\nu_\lambda}\,<x\right)\underset{\lambda\to 1}{\longrightarrow} 1-2\frac{(\sqrt{2}\,x)^a}{\Gamma(a)}\sum_{j=1}^\infty(-1)^{j-1}j^{\,a}\,K_a(2\sqrt{2}\,jx),\ x>0.$$

Рассмотрим теперь приближения для точных распределений статистики Смирнова Н.В. и статистики Колмогорова А.Н. (в основном для статистики Смирнова Н.В.), которые основаны на асимптотических разложениях. Первое уточнение предельного распределения было дано в работе [25]. Следующее уточнение (второй член асимптотического разложения) получено в работе [26], а уточнение до члена порядка $1/\sqrt{n^3}$ дано в работе [27].

Если точное распределение есть

$$\mathbf{P}(D_n^+ < \lambda) = \sum_{k=[n(1-\lambda)]+1}^{n-1} C_n^k \, \lambda \left(\frac{k}{n} + \lambda\right)^{k-1} \left(1 - \frac{k}{n} - \lambda\right)^{n-k}, \quad \text{для} \quad 0 \leqslant \lambda \leqslant 1,$$

то (см. [27], Теорема 2) асимптотическое разложение имеет вид

$$\mathbf{P}\left(D_n^+ < \frac{\lambda}{\sqrt{n}}\right) = 1 - e^{-2\lambda^2}\left(1 - \frac{2\lambda}{3\sqrt{n}} + \frac{2\lambda^2}{3n}\left(1 - \frac{2\lambda^2}{3}\right) + \right.$$
$$\left. + \frac{4\lambda}{9\sqrt{n^3}}\left(\frac{1}{5} - \frac{19\lambda^2}{15} + \frac{2\lambda^4}{3}\right) + O\left(\frac{1}{n^2}\right)\right). \quad (9)$$

В статье [28] Большев Л.Н. предложил преобразовать статистику D_n^+ и использовать статистику $\xi_n^+ = (6nD_n^+ + 1)^2/(18n)$ для проверки гипотезы согласия. Асимптотическое разложение для статистики ξ_n^+ будет иметь вид:

$$\mathbf{P}(\xi_n^+ < x) = (1 - e^{-x}) + e^{-x}\left(\frac{2x^2 - 4x - 1}{18n} + O\left(\frac{1}{n\sqrt{n}}\right)\right).$$

В результате этого преобразования аннулируется первое слагаемое разложения, т.е. коэффициент при $1/\sqrt{n}$. Если же использовать разложение (9), то асимптотическое разложение для распределения статистики ξ_n^+ будет следующим:

$$\mathbf{P}(\xi_n^+ < x) = 1 - e^{-x}\left(1 - \frac{2x^2 - 4x - 1}{18n} + \frac{2\sqrt{2x}}{27n\sqrt{n}} + O\left(\frac{1}{n^2}\right)\right).$$

Это позволяет уточнить следующую формулу из статьи [28], с.152:

$$D_n^+(\beta) = \sqrt{\frac{x_2^+(-\ln\beta)}{2n}} - \frac{1}{6n}, \quad x_2^+(y) = y - \frac{2y^2 - 4y - 1}{18n} + \frac{2\sqrt{2y}}{27n\sqrt{n}}.$$

Для приведенного в [28] примера, где $\beta = 0.05$, $\ln(1/\beta) = 2.9957$, будем иметь: $x_2^+(-\ln\beta) = 2.9957 - \frac{0.2762}{n} + \frac{0.06052446}{n\sqrt{n}}$ и для $n = 5$ получаем приближенное значение 0.5094257 вместо $D_5^+(0.05) = 0.50892675$ (в [28] округлено до 0.5090), в то время как точное значение равно 0.5094493.

Сделаем еще одно преобразование:

$$\eta_n^+ = \frac{9}{2}n + 1 - \frac{\sqrt{81n^2 - 36n\xi_n^+ + 36n + 6}}{2},$$

в результате которого в асимптотическом разложении статистики η_n^+ занулится следующее слагаемое – коэффициент при $1/n$. Поэтому, если мы будем находить

значение статистики D_n^+ по выборке случайного объема v_λ, имеющего распределение $\mathbf{P}(v_\lambda)=k)=(1-\lambda)\,\lambda^{k-1},\ k=1,2,...$, и возьмем преобразованную статистику $\eta_{v_\lambda}^+$, то получим, что ее распределение будет иметь главный член $1-e^{-x}$ и второе слагаемое $e^{-x}(\sqrt{8x}/27)(1-\lambda)\sum_{k=1}^{\infty}\lambda^{k-1}k^{-3/2}$. Но ряд $\sum_{k=1}^{\infty}k^{-3/2}$ сходится, тем более будет сходиться и ряд $\sum_{k=1}^{\infty}\lambda^{k-1}k^{-3/2}, 0<\lambda<1$. Это значит, что если объем выборки случаен, то он будет оказывать пренебрежимое влияние на распределение статистики $\eta_{v_\lambda}^+$ (а, значит, и на распределение $D_{v_\lambda}^+$) при $\lambda\to 1$. Аналогичный вывод будет иметь место и для статистики Колмогорова D_{v_λ}.

Что же касается критерия Пирсона для проверки простой гипотезы согласия по повторной выборке объема n, то ее статистика имеет вид:

$$\chi^2=\sum_{j=1}^{k+1}\frac{(m_j-np_j)^2}{np_j},$$

где $(m_1,m_2,...,m_{k+1})$ при нулевой гипотезе имеет полиномиальное распределение

$$P=\frac{n!}{m_1!m_2!\cdots m_{k+1}!}\,p_1^{m_1}p_2^{m_2}\cdots p_{k+1}^{m_{k+1}},\qquad \sum_{j=1}^{k+1}m_j=n.$$

Если для этих целей воспользоваться асимптотическим разложением, полученным в работе [29], с.164, то соответствующая предельная функция распределения записана там как

$$F_k+\frac{S_1}{n}\,(F_{k+1}-2F_{k+2}+F_k)+\frac{S_2}{n}\,(F_{k+6}-3F_{k+4}+3F_{k+2}-F_k),$$

где

$$S_1=\frac{1}{8}\left[\sum_{j=1}^{k+1}\frac{1}{p_i}-(k^2+4k+1)\right]\quad\text{и}\quad S_2=\frac{1}{24}\left[5\sum_{j=1}^{k+1}\frac{1}{p_i}-(3k^2+12k+5)\right],$$

а через F_k обозначено распределение величины χ_k^2 с k степенями свободы.

Отметим, что в асимптотическом разложении коэффициент при $1/\sqrt{n}$ равен нулю. Несмотря на то, что ряд $\sum\limits_{n=1}^{\infty}\dfrac{1}{n}$ расходится, ряд $s(\gamma)=\sum\limits_{n=1}^{\infty}\dfrac{\gamma^n}{n}$ для каждого фиксированного $0<\gamma<1$ сходится, поэтому для геометрического распределения будем иметь математическое ожидание, равное $(1-\gamma)\cdot s(\gamma)=-(1-\gamma)\cdot\ln(1-\gamma)$, и мы получаем выводы, аналогичные предыдущим.

3. Последовательное оценивание

Рассмотрим равномерное на интервале $(\theta-1/2,\theta+1/2)$ распределение и оценку Питмена (см. [29]) параметра сдвига θ по выборке $(x_n^{(r+1)},x_n^{(r+2)},...,x_n^{(n-s)})$, которая есть

$$t_n=\frac{(s+1)\,x_n^{(r+1)}+(r+1)\,x_n^{(n-s)}}{r+s+2}+\frac{s-r}{2(r+s+2)},$$

математическое ожидание и дисперсия которой равны, соответственно

$$\mathbf{E}_\theta(t_n) = \theta \quad \text{и} \quad \mathbf{D}(t_n) = \frac{(r+1)\,(s+1)}{(n+1)\,(n+2)\,(r+s+2)}\,.$$

В работе [21] для оценки параметра сдвига равномерного распределения была предложена статистика (эта же оценка приведена в книге [22], с. 718):

$$m = \frac{(n-2s-1)\,x_n^{(r+1)} + (n-2r-1)\,x_n^{(n-s)}}{2\,(n-r-s-1)},$$

дисперсия которой равна

$$\mathbf{D}(m) = \frac{(n-2s-1)\,(r+1) + (n-2r-1)\,(s+1)}{4\,(n+1)\,(n+2)\,(n-r-s-1)},$$

при этом

$$\mathbf{D}(m) - \mathbf{D}(t_n) = \frac{(r-s)^2}{4\,(n-r-s-1)\,(r+s+2)\,(n+2)} \geqslant 0\,,$$

т.е. дисперсия оценки t_n меньше, чем дисперсия оценки m, если $r \neq s$, и они совпадают, если $r = s$.

В дальнейшем, чтобы получить простые формулы, возьмем $r = s$. Тогда эти оценки равны $t_n = \dfrac{x_n^{(r+1)} + x_n^{(n-r)}}{2}$, а $\mathbf{D}(t_n) = \dfrac{(r+1)}{2\,(n+1)\,(n+2)}$.

Рассмотрим теперь последовательное оценивание параметра θ равномерного на $(\theta-1/2, \theta+1/2)$ распределения в случае $r = s$ по выборке $(x_n^{(r+1)}, x_n^{(r+2)}, ..., x_n^{(n-r)})$.

Обозначим $\lambda_n = 1 - x_n^{(n-r)} + x_n^{(r+1)}$. Найдем момент порядка $a = 2m,\ m \in \mathbf{N}$ в два шага:

$$\mathbf{E}_\theta((t_n-\theta)^\alpha) = \mathbf{E}_\theta(\mathbf{E}_\theta((t_n-\theta)^\alpha |\, \mathcal{F}_n)), \quad \text{где} \quad \mathcal{F}_n = \sigma\{x_2-x_1, x_3-x_2, ..., x_n-x_{n-1}\}\,.$$

Имеем

$$\mathbf{E}_\theta((t_n-\theta)^\alpha |\, \mathcal{F}_n) = \mathbf{E}_0(t_n^\alpha |\, \mathcal{F}_n) =$$
$$= \frac{\int\limits_{x_n^{(n-r)}-1/2}^{x_n^{(r+1)}+1/2} (t_n-y)^\alpha\,(y-x_n^{(r+1)}-1/2)^r\,(1/2-x_n^{(n-r)}-y)^r\,dy}{\int\limits_{x_n^{(n-r)}-1/2}^{x_n^{(r+1)}+1/2} (y-x_n^{(r+1)}-1/2)^r\,(1/2-x_n^{(n-r)}-y)^r\,dy} = \frac{I_2}{I_1}.$$

Вычислим интеграл I_2:

$$I_2 = \int\limits_0^{\lambda_n} (\lambda_n/2-x)^\alpha(\lambda_n-x)^r x^r\,dt = 2\int\limits_0^{\lambda_n/2} z^\alpha(\lambda_n^2/4-z^2)^r\,dz =$$
$$= 2\,(\lambda_n/2)^{\alpha+2r+1}\int\limits_0^1 u^\alpha(1-u^2)^r\,du = \left(\frac{\lambda_n}{2}\right)^{\alpha+2r+1}\frac{\Gamma((\alpha+1)/2)\Gamma(r+1)}{\Gamma((\alpha+1)/2+r+1)}\,.$$

Аналогично, $I_1 = \lambda_n^{2r+1} \cdot \dfrac{\Gamma^2(r+1)}{\Gamma(2r+2)}$, поэтому

$$\mathbf{E}_\theta((t_n-\theta)^\alpha|\,\mathcal{F}_n) = \frac{\lambda_n^\alpha}{2^{2r+\alpha+1}}\,\frac{\Gamma((\alpha+1)/2)\,\Gamma(2r+2)}{\Gamma(r+1)\,\Gamma(r+1+(\alpha+1)/2)}\,.$$

Используя плотность распределения разности $x_n^{(n-r)} - x_n^{(r+1)}$ для равномерного на $(0,1)$ распределения (см. [22], p.14, (2.3.4)), получаем

$$\mathbf{E}(\lambda_n^\alpha) = \int_0^1 (1-w)^\alpha \cdot \frac{\Gamma(n+1)}{\Gamma(2r+2)\Gamma(n-2r-1)} w^{n-2r-2}\,(1-w)^{2r+1}\,dw = \\ = \frac{\Gamma(2r+\alpha)\,\Gamma(n+1)}{\Gamma(2r+2)\,\Gamma(n+\alpha+1)}.$$

Значит,

$$\mathbf{E}_\theta((t_n-\theta)^\alpha) = \frac{\Gamma(2r+2+\alpha)\,\Gamma(n+1)\,\Gamma((\alpha+1)/2)}{2^{2r+\alpha+1}\,\Gamma(n+\alpha+1)\,\Gamma(r+1)\,\Gamma(r+1+(\alpha+1)/2)} = \\ = \frac{\Gamma(2r+1+\alpha)\,\Gamma(n+1)\,\Gamma((\alpha+1)/2)}{2^{2r+\alpha}\,\Gamma(n+\alpha+1)\,\Gamma(r+1)\,\Gamma(r+(\alpha+1)/2)}.$$

Учитывая формулу Лежандра удвоения гамма-функции (см. [24], с.19, (15)) $\sqrt{\pi}\,\Gamma(2z) = \Gamma(z+1/2)\,\Gamma(z)\cdot 2^{2z-1}$, получим

$$\mathbf{E}_\theta((t_n-\theta)^\alpha) = \frac{(2r+\alpha)\,\Gamma(2r+\alpha/2)\,\Gamma(n+1)\,\Gamma(\alpha)}{2^\alpha\,\Gamma(n+\alpha+1)\,\Gamma(r+1)\,\Gamma(\alpha/2)} = \frac{(2r+\alpha)\,\Gamma(2r+m)\,\Gamma(n+1)\,\Gamma(\alpha)}{2^\alpha\,\Gamma(n+\alpha+1)\,\Gamma(r+1)\,\Gamma(m)}\,.$$

Если $r = s = 0$, то (см. [3], Теорема 2) *оптимальный инвариантный последовательный план* $\Delta_0 = \Delta(\tau_0, \hat\theta_{\tau_0})$ *оценивания параметра сдвига равномерного на* $(\theta-1/2, \theta+1/2)$ *распределения имеет вид*:

$$\tau_0 = \min\{k \geqslant 2 : x_k^{(k)} - x_k^{(1)} \geqslant 1 - 2/n\} \quad \hat\theta_{\tau_0} = (x_{\tau_0}^{(1)} + x_{\tau_0}^{(\tau_0)})/2\,,$$

для которого $\mathbf{E}_\theta(\tau_0) = n$, *и*

$$\mathbf{E}_\theta((\hat\theta_n-\theta)^\alpha) = \frac{\Gamma(n+1)\Gamma(\alpha+1)}{2^\alpha\Gamma(n+\alpha+1)} \sim \frac{\Gamma(\alpha+1)}{2^\alpha n^\alpha} \quad \textit{при} \quad n\to\infty.$$

Моменты нечетного порядка равны нулю, то есть если $\alpha = 2m+1$, то $\mathbf{E}_\theta((\hat\theta_n-\theta)^\alpha) = 0$.

В случае $r = s \neq 0$ оптимальный инвариантный последовательный план $\Delta_0 = \Delta(\tau_0, \hat\theta_{\tau_0})$ оценивания параметра сдвига равномерного на $(\theta-1/2, \theta+1/2)$ распределения имеет вид:

$$\tau_0 = \min\{k \geqslant 2r+2 : x_k^{(k)} - x_k^{(1)} \geqslant 1 - (2r+2)/n\} \quad \hat\theta_{\tau_0} = (x_{\tau_0}^{(r)} + x_{\tau_0}^{(\tau_0-r)})/2\,,$$

для которого также $E_\theta(\tau_0) = n$. Следовательно

$$\mathbf{E}_\theta((\hat\theta_n-\theta)^\alpha) = \frac{\Gamma(n+1)\Gamma(\alpha+1)}{2^\alpha\Gamma(n+\alpha+1)} \sim \frac{\Gamma(\alpha+1)}{2^\alpha n^\alpha} \quad \text{при} \quad n\to\infty.$$

Вернемся к последовательному плану оценивания по полной выборке и рассмотрим случайную величину τ_0, определенную следующим образом:
$\tau_0 = \min\{k \geqslant 2 : x_k^{(k)} - x_k^{(1)} \geqslant 1-\varepsilon\}$, $\varepsilon = 2/n$ и событие $A_k = \{x_k^{(k)} - x_k^{(1)} < 1-\varepsilon\}$. Тогда $(\tau_0 = k) = A_{k-1} \cap \overline{A_k}$ и $\mathbf{P}(\tau_0 = k) = \mathbf{P}(A_{k-1}) - \mathbf{P}(A_k)$, так как $A_k \subset A_{k-1}$. Используя плотность распределения размаха $w = x_k^{(k)} - x_k^{(1)}$ для равномерного на $(0,1)$ распределения, получаем

$$\mathbf{P}(A_k) = \int\limits_0^{1-\varepsilon} k(1-k)w^{k-2}(1-w)\,dw = k(1-\varepsilon)^{k-1} - (k-1)(1-\varepsilon)^k,$$

поэтому

$$\mathbf{P}(\tau_0 = k) = (k-1)(1-\varepsilon)^{k-2}\varepsilon^2, \qquad k = q+1, q+2, ...,$$

т.е. отрицательное биномиальное распределение с параметрами $(2, \varepsilon)$.

Найдем характеристическую функцию $\varphi_1(t)$ величины $n\,(\hat{\theta}_n - \theta)$ при $n \to \infty$). Имеем:

$$\varphi_1(t) = 1 + \sum_{m=1}^{\infty} \frac{\Gamma(2k+1)}{2^{\,2k}} \cdot \frac{(-t^2)^k}{(2k)!} = \sum_{k=0}^{\infty} \left(-\frac{t^2}{2}\right)^k = \frac{2}{2+t^2},$$

что является характеристической функцией распределения Лапласа с плотностью $\lambda(x) = \frac{1}{\sqrt{2}}\, e^{-|\,x\,|\sqrt{2}},\ x \in \mathbf{R}$.

Пусть $L_1(\tau, \delta_\tau, \theta)$ и $L_2(\tau, \delta_\tau, \theta)$ – неотрицательные измеримые функции потерь (необязательно выпуклые) от принятия последовательного плана $\Delta = (\tau, \delta_\tau)$, когда истинное значение параметра равно θ. Пусть

$$0 < \mathbf{E}_\theta(L_1(\tau, \delta_\tau, \theta)) < \infty, \quad 0 < \mathbf{E}_\theta(L_2(\tau, \delta_\tau, \theta)) < \infty.$$

Лемма 1. *Пусть существует при каждом θ такой план Δ_c, что*

$$\boldsymbol{E}_\theta(L_1(\Delta_c, \theta) + cL_2(\Delta_c, \theta)) = \inf_{\Delta} \boldsymbol{E}_\theta(L_1(\Delta, \theta) + cL_2(\Delta, \theta)), \tag{10}$$

и пусть функции $q_\theta(c) = \boldsymbol{E}_\theta(L_2(\Delta_c, \theta))$, $Q_\theta(c) = \boldsymbol{E}_\theta(L_1(\Delta_c, \theta))$ – функции ограниченной вариации, причем существуют обратные функции $q_\theta^{-1}(c)$ и $Q_\theta^{-1}(c)$. Тогда

$$Q_\theta(c) = -\int\limits_0^c x\, dq_\theta(x) = \int\limits_{\boldsymbol{E}_\theta(L_2(\Delta_c,\theta))}^{\infty} q_\theta^{-1}(y)\, dy, \tag{11}$$

$$q_\theta(c) = -\int\limits_0^c x^{-1}\, dQ_\theta(x) = \int\limits_{\boldsymbol{E}_\theta(L1(\Delta_c,\theta))}^{\infty} (Q_\theta^{-1}(y))^{-1}\, dy. \tag{12}$$

Доказательство. Докажем соотношение (11), соотношение (12) доказывается аналогично.

Если $c_1 \neq c$, то из (10) следует, что

$$\begin{cases} Q_\theta(c) + c\, q_\theta(c) \leqslant Q_\theta(c_1) + c\, q_\theta(c_1), \\ Q_\theta(c_1) + c_1 q_\theta(c_1) \leqslant Q_\theta(c) + c_1\, q_\theta(c), \end{cases}$$

откуда имеем

$$\begin{cases} Q_\theta(c) - Q_\theta(c_1) + c\,(\,q_\theta(c) - q_\theta(c_1)\,) \leqslant 0, \\ Q_\theta(c) - Q_\theta(c_1) + c_1\,(\,q_\theta(c) - q_\theta(c_1)\,) \geqslant 0. \end{cases}$$

Разобьем отрезок $[0, c]$ точками $0 \equiv c_0 < c_1 \ldots < c_k < c \equiv c_{k+1}$. Учитывая, что $Q_\theta(0) = 0$, складывая последовательно предыдущие неравенства, мы получаем

$$\begin{cases} Q_\theta(c) + \sum\limits_{i:\,0 \leqslant c_i \leqslant c} c_i\,(\,q_\theta(c_i) - q_\theta(c_{i+1})\,) \leqslant 0, \\ Q_\theta(c) + \sum\limits_{i:\,0 \leqslant c_i \leqslant c} c_i\,(\,q_\theta(c_i) - q_\theta(c_{i+1})\,) \geqslant 0. \end{cases}$$

Так как функция $g(x) = x$ – непрерывна, а $q_\theta(c)$ есть функция ограниченной вариации, то

$$\begin{cases} Q_\theta(c) + \int\limits_0^c x\,dq_\theta(x) \leqslant 0, \\ Q_\theta(c) + \int\limits_0^c x\,dq_\theta(x) \geqslant 0. \end{cases}$$

Последние неравенства совместимы, только если

$$Q_\theta(c) = -\int_0^c x\,dq_\theta(x),$$

что составляет утверждение леммы. □

Пусть $\tau(c)$ момент остановки, для которого

$$\mathbf{E}_\theta\,(\,|\,t_{\tau(c)} - \theta\,|^{\,\alpha}) + c\,\mathbf{E}_\theta(\tau(c)) = \min_\tau\{\mathbf{E}_\theta\,(\,|\,t_\tau - \theta\,|^{\,\alpha}) + c\,\mathbf{E}_\theta(\tau)\},$$

и пусть

$$Q(c) = \mathbf{E}_\theta\,(\,|\,t_{\tau(c)} - \theta\,|^{\,\alpha}), \quad q(c) = \mathbf{E}_\theta(\tau(c)).$$

В нашем случае $q^{\,-1}(x) = \mu_\alpha\,\dfrac{(2r+2)^\alpha}{(2r+\alpha+2)\,n^\alpha}\,\alpha\,x^{\,\alpha-1}$, где

$$\mu_\alpha = \frac{\Gamma(2r+2)}{\Gamma^2(r+1)}\int_0^1 |\,1/2 - y\,|^\alpha y^r(1-y)^r\,dy,$$

поэтому

$$\mathbf{E}_\theta(\,|\,t_{\tau_o} - \theta\,|^{\,\alpha}\,) = \mu_\alpha\,\frac{\alpha(2r+2)^{1+\alpha}}{(2r+\alpha+2)\,n^\alpha}\int_0^1 t^{\,\alpha-1}\,dt = \mu_\alpha\,\frac{(2r+2)^{1+\alpha}}{(2r+\alpha+2)\,n^\alpha}.$$

Заметим, что если $r = s = 0$, то $\mu_\alpha = \dfrac{1}{(\alpha+1)\,2^\alpha}$.

Рассмотрим последовательный план оценивания когда $r = s = 0$, определяемый моментом остановки τ_0, для которого $\mathbf{E}(\theta(\tau_0)) = n$, с терминальным решением $\hat\theta_{\tau_0}$. Для $\hat\theta_{\tau_0}$ моменты нечетного порядка равны 0, а моменты четного порядка равны

$$\mathbf{E}(\tilde\theta^{\,\alpha}_{\tau_0}) = \frac{2}{n^\alpha(\alpha+1)(\alpha+2)}.$$

Если взять треугольное распределение величины $\zeta = \xi/n$, где ξ имеет плотность $\rho(x) = 1 - |\,x\,|, |\,x\,| \leqslant 1$, то моменты четного порядка будут равны $\mathbf{E}(\zeta^\alpha) = \dfrac{2}{(\alpha+1)(\alpha+2)n^\alpha}$ и выполнено условие Карлемана. Это означает, что для $n \geqslant 2$, распределение величины $\tau_0(\tilde{\theta}_{\tau_0} - \theta)$ является треугольным, имеющим ограниченный носитель, т.е. легкие хвосты.

Заключение

В работе рассмотрены задачи статистического оценивания распределений по выборкам случайного объема, когда объем и исходная выборка независимы (п.1 и 2) и задача статистического оценивания параметра сдвига равномерного распределения (п.3), в которой случайный объем выборки является моментом остановки. Если предельные распределения статистик в первом случае имеют более тяжелые хвосты, нежели для выборок фиксированного объема, то в последнем случае объем выборки, являясь случайной величиной, определяется по исходной выборке, а распределение последовательной оценки имеет треугольное распределение, у которого легкие хвосты.

Список литературы

[1] Wald A. Sequential Test of Statistical Hypotheses // The Annals of Mathematical Statistics. 1945. Vol. 16, № 2. Pp. 117–186.

[2] Linnik Yu.V., Romanovsky I.V. Some new results in sequential estimation theory // Berkeley Symposium on Mathematical Statistics and Probability. 1972. Pp. 85–96.

[3] Тихов М.С. Об оптимальных планах последовательного оценивания при неквадратичных потерях // Теория вероятностей и ее применения. 1978. Т. 23, № 1. С. 137–143.

[4] Тихов М.С. Последовательное оценивание параметра сдвига равномерного распределения по цензурированным типа II выборкам // Записки научных семинаров ЛОМИ. 1984. Т. 136. С. 183–192.

[5] Renyi A. On the central limit theorem for the sum of a random number of independent random variables // Acta Mathematica Academiae Scientiarum Hungaricae. 1963. Vol. 11. Pp. 97–102. https://doi.org/10.1007/BF02020627

[6] Blum J.R., Hanson D.L., Rosenblatt J.I. On the central limit theorem for the sum of a random number of independent random variables // Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete. 1963. Vol. 1. Pp. 389–393. https://doi.org/10.1007/BF00533414

[7] Гнеденко Б.В., Стоматович С., Шукри А. О распределении медианы // Вестник Московского университета. Серия 1: Математика. Механика. 1984. № 2. С. 59–63.

[8] Бенинг В.Е. О мощности критериев в случае выборок случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 5–25. https://doi.org/10.26456/vtpmk514

[9] Бенинг В.Е., Королев В.Ю. Об использовании распределения Стьюдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения. 2004. Т. 49, № 3. С. 417–435. https://doi.org/10.1137/S0040585X97981159

[10] Слепов Н.А. Скорость сходимости распределений геометрических сумм к закону Лапласа // Теория вероятностей и ее применения. 2021. Т. 66, № 1. С. 149–174. https://doi.org/10.4213/tvp5363

[11] Toda A.A. Weak limit of the geometric sum of independent but not identically distributed random variables. 2011. https://doi.org/10.48550/arXiv.1111.1786

[12] Christoph G., Ulyanov V. Second Order Chebyshev-Edgeworth-Type Approximations for Statistics Based on Random Size Samples // Mathematics. 2023. Vol. 11. ID 1848. https://doi.org/10.3390/math11081848

[13] Прудников А.П., Брычков Ю.А., Маричев О.И. Интегралы и ряды. Т. 1: Элементарные функции. М.: ФИЗМАТЛИТ, 2002. 632 с.

[14] Tikhov M.S. Statistical estimation based on interval censored data // Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life. Eds. by N. Balakrishnan, M.S. Nikulin, M. Mesbah, N. Limnios. Series: Statistics for Industry and Technology. Boston, MA: Birkhauser, 2004. Pp. 211–218. https://doi.org/10.1007/978-0-8176-8206-4_14

[15] Тихов М.С. Асимптотические распределения суммируемых квадратичных уклонений оценок функции распределения в зависимости «доза-эффект» // Обозрение прикладной и промышленной математики. 2009. Т. 16, № 5. С. 772–786.

[16] Okumura H., Naito K. Non-parametric kernel regression for multinomial data // Journal Multivariate Analysis. 2006. Vol. 97. Pp. 2009–2022.

[17] Тихов М.С. Отрицательная λ-биномиальная регрессия в зависимости доза-эффект // Вестник ТвГУ. Серия: Прикладная математика. 2022. № 4. С. 53–75. https://doi.org/10.26456/vtpmk649

[18] Градштейн И.С., Рыжик И.М. Таблицы интегралов, сумм, рядов и произведений. М.: Физматлит, 1963. 1100 с.

[19] Колмогоров А.Н. Об эмпирическом определении закона распределения // Теория вероятностей и математическая статистика. М.: Наука, 1986. С. 134–141.

[20] Гихман И.И., Скороход А.В. Введение в теорию случайных процессов. М.: Наука, 1965. 356 с.

[21] Sarhan A., Greenberg B. Estimation of location and scale parameter for the rectangular population fron censored samples // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1959. Vol. 21, № 2. Pp. 356–363.

[22] Кендалл М., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973. 900 с.

[23] David H., Nagaraja H. Order Statistics. Wiley, 2003. 475 p.

[24] Бейтман Г., Эрдейи А. Высшие трансцендентные функции. М.: Наука, 1973. 296 с.

[25] Смирнов Н.В. Приближение законов распределения случайных величин по эмпирическим данным // Успехи математических наук. 1944. Т. 1, № 10. С. 179–206.

[26] Королюк В.С. Асимптотические разложения для критериев согласия А. Н. Колмогорова и Н. В. Смирнова // Известия АН СССР. Серия математическая. 1955. Т. 19, № 2. С. 103–124.

[27] Ли-Цянь Ч. О точном распределении статистики Н.В.Смирнова и его асимптотическом разложении // Математика. 1960. Т. 4, № 2. С. 121–134.

[28] Большев Л.Н. Асимптотически пирсоновские преобразования // Теория вероятностей и ее применения. 1963. Т. 8, № 2. С. 129–155. https://doi.org/10.1137/1108012

[29] Hoel P.G. On the chi-square distribution for small samples // The Annals of Mathematical Statistics. 1938. Vol. 9, № 3. Pp. 158–165.

[30] Каган А.М. Теория оценивания для семейств с параметрами сдвига, масштаба и экспонентных // Труды Математического института имени В. А. Стеклова. 1968. Т. 104. С. 19–87.

Образец цитирования

Тихов М.С. Оценивание распределений по выборкам случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2023. №4. С. 5–24. https://doi.org/10.26456/vtpmk695

Сведения об авторах

1. **Тихов Михаил Семенович**
профессор кафедры теории вероятностей и анализа данных института информационных технологий, математики и механики Нижегородского государственного университета им. Н.И. Лобачевского.

Россия, 603950, г. Нижний Новгород, пр. Гагарина, д. 23, ННГУ им. Н.И. Лобачевского. E-mail: tikhovm@mail.ru

ESTIMATING DISTRIBUTIONS FROM SAMPLES WITH RANDOM SIZE

Tikhov M.S.
Lobachevsky State University of Nizhniy Novgorod, Nizhniy Novgorod

Received 24.11.2023, revised 30.11.2023.

The article is concerned with the estimating problem of a distribution function and the limiting behavior of the distance between the empirical and theoretical laws, namely, integrated square errors and Smirnov and Kolmogorov statistics by the samples with random size. We suppose that this random size and the initial sample are independent random variables and this random variable has the generalized negative binomial distribution. We find limiting distributions for integrated square errors of kernel distribution function estimators by the samples with random size. It is shown that for samples with random size the limiting distribution of the Smirnov and Kolmogorov statistics has more heavier tails than the Weibull and Kolmogorov distribution function for samples with the fixed size. We propose an asymptotic expansion approach to naturally balance the asymptotic distribution and random sample size. The problem of sequential estimation of the shift parameter of the uniform distribution is considered. The negative binomial distribution (samples size ν) arises naturally here from a statistical experiment of performing a series of independent trials.

Keywords: sample with random size, empirical distribution function, Kolmogorov statistics, sequential estimation, generalized negative binomial distribution.

Citation

Tikhov M.S., "Estimating distributions from samples with random size", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2023, № 4, 5–24 (in Russian). https://doi.org/10.26456/vtpmk695

References

[1] Wald A., "Sequential Test of Statistical Hypotheses", *The Annals of Mathematical Statistics*, **16**:2 (1945), 117–186.

[2] Linnik Yu.V., Romanovsky I.V., "Some new results in sequential estimation theory", *Berkeley Symposium on Mathematical Statistics and Probability*, 1972, 85–96.

[3] Tihov M.S., "On optimal sequential estimation procedures for non-quadratic loss functions", *Theory of Probability and its Applications*, **23**:1 (1978), 132–138.

[4] Tikhov M.S., "Sequential estimation of the uniform distribution shift parameter for censored type II samples", *Notes of LOMI scientific seminars*, **136** (1984), 183–192 (in Russian).

[5] Renyi A., "On the central limit theorem for the sum of a random number of independent random variables", *Acta Mathematica Academiae Scientiarum Hungaricae*, **11** (1963), 97–102, https://doi.org/10.1007/BF02020627.

[6] Blum J.R., Hanson D.L., Rosenblatt J.I., "On the central limit theorem for the sum of a random number of independent random variables", *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **1** (1963), 389–393, https://doi.org/10.1007/BF00533414.

[7] Gnedenko B.V., Stomatovich S., Shukri A., "About the distribution of the median", *Vestnik Moskovskogo universiteta. Seriya 1: Matematika. Mekhanika [Bulletin of the Moscow University. Series 1: Mathematics. Mechanics]*, 1984, №2, 59–63 (in Russian).

[8] Bening V.E., "On the power of criteria in the case of samples of random size", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2018, №4, 5–25 (in Russian), https://doi.org/10.26456/vtpmk514.

[9] Bening V.E., Korolev V.Y., "On an application of the Student distribution in the theory of probability and mathematical statistics", *Theory of Probability and its Applications*, **49**:3 (2005), 377–391, https://doi.org/10.1137/S0040585X97981159.

[10] Slepov N.A., "Convergence rate of random geometric sum distributions to the Laplace law", *Theory of Probability and its Applications*, **66**:1 (2021), 121–141, https://doi.org/10.4213/tvp5363.

[11] Toda A.A., *Weak limit of the geometric sum of independent but not identically distributed random variables*, 2011, https://doi.org/10.48550/arXiv.1111.1786.

[12] Christoph G., Ulyanov V., "Second Order Chebyshev-Edgeworth-Type Approximations for Statistics Based on Random Size Samples", *Mathematics*, **11** (2023), 1848, https://doi.org/10.3390/math11081848.

[13] Prudnikov A.P., Brychkov Yu.A., Marichev O.I., *Integraly i ryady [Integrals and series. Elementary functions]*, Vol. 1: Elementarnye funktsii, Fizmatlit Publ., Moscow, 2002 (in Russian), 632 pp.

[14] Tikhov M.S., "Statistical estimation based on interval censored data", *Parametric and Semiparametric Models with Applications to Reliability, Survival Analysis, and Quality of Life*, Statistics for Industry and Technology, eds. N. Balakrishnan, M.S. Nikulin, M. Mesbah, N. Limnios, Birkhauser, Boston, MA, 2004, 211–218, https://doi.org/10.1007/978-0-8176-8206-4_14.

[15] Tikhov M.S., "Asimptoticheskie raspredeleniya summiruemykh kvadratichnykh uklonenij otsenok funktsii raspredeleniya v zavisimosti ¡¿", *Obozrenie prikladnoj i promyshlennoj matematiki [Review of Applied and Industrial Mathematics journal]*, **16**:5 (2009), 772–786 (in Russian).

[16] Okumura H., Naito K., "Non-parametric kernel regression for multinomial data", *Journal Multivariate Analysis*, **97** (2006), 2009–2022.

[17] Tikhov M.S., "Negative λ-binomial regression in dose-effect relationship", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2022, №4, 53–75 (in Russian), https://doi.org/10.26456/vtpmk649.

[18] Gradshtejn I.S., Ryzhik I.M., *Tablitsy integralov, summ, ryadov i proizvedenij [Table of Integrals, Series, and Products]*, Fizmatlit Publ., Moscow, 1963 (in Russian), 1100 pp.

[19] Kolmogorov A.N., "On the empirical definition of the law of distribution", *Teoriya veroyatnostej i matematicheskaya statistika [Probability theory and mathematical statistics]*, Nauka Publ., Moscow, 1986, 134–141 (in Russian).

[20] Gikhman I.I., Skorokhod A.V., *Vvedenie v teoriyu sluchajnykh protsessov [Introduction to the theory of random processes]*, Nauka Publ., Moscow, 1965 (in Russian), 356 pp.

[21] Sarhan A., Greenberg B., "Estimation of location and scale parameter for the rectangular population fron censored samples", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **21**:2 (1959), 356–363.

[22] Kendall M., Styuart A., *Statisticheskie vyvody i svyazi [Statistical conclusions and connections]*, Nauka Publ., Moscow, 1973 (in Russian), 900 pp.

[23] David H., Nagaraja H., *Order Statistics*, Wiley, 2003, 475 pp.

[24] Bejtman G., Erdeji A., *Vysshie transtsendentnye funktsii*, Nauka Publ., Moscow, 1973 (in Russian), 296 pp.

[25] Smirnov N.V., "Approximate laws of distribution of random variables from empirical data", *Uspekhi Matematicheskikh Nauk [Achievements of mathematical sciences]*, **1**:10 (1944), 179–206 (in Russian).

[26] Korolyuk V.S., "Asymptotic expansions for the criteria of agreement by A.N.Kolmogorov and N.V.Smirnov", *Izvestiya AN SSSR. Seriya matematicheskaya [Izvestia of the USSR Academy of Sciences. Mathematical series]*, **19**:2 (1955), 103–124 (in Russian).

[27] Li-Tsyan Ch., "On the exact distribution of N.V.Smirnov's statistics and its asymptotic decomposition", *Matematika [Mathematics]*, **4**:2 (1960), 121–134 (in Russian).

[28] Bol'shev L.N., "Asymptotically Pearson's Transformations", *Theory of Probability and its Applications*, **8**:2 (1963), 121–146, https://doi.org/10.1137/1108012.

[29] Hoel P.G., "On the chi-square distribution for small samples", *The Annals of Mathematical Statistics*, **9**:3 (1938), 158–165.

[30] Kagan A.M., "Theory of estimation for families with shift, scale and exponential parameters", *Trudy Matematicheskogo Instituta im. V. A. Steklova [Proceedings of the Steklov Mathematical Institute]*, **104** (1968), 19–87 (in Russian).

Author Info

1. **Tikhov Mikhail Semenovich**
Professor at Probability Theory and Data Analysis department,
Lobachevsky State University of Nizhny Novgorod.

Russia, 603950, Nyzhniy Novgorod, 23 Gagarin av., UNN. E-mail: tikhovm@mail.ru