Критерий согласия Колмогорова

Прикладной семинар "Количественный анализ"

Софья Цай

Февраль 2025

Аннотация

Критерий согласия Колмогорова предназначен для проверки гипотезы о принадлежности выборки некоторому закону распределения. Основная цель доклада – рассмотреть статистику, используемую в критерии согласия Колмогорова, а также вывести функцию распределения Колмогорова.

Пусть имеется выборка $X_1, ..., X_n$ из произвольного непрерывного распределения $F(x)$. Рассмотрим задачу проверки статистической гипотезы:

$$H_0 : F(x) = F_0(x)$$
$$H_1 : F(x) \neq F_0(x)$$

Построим по выборке **эмпирическую функцию распределения**

$$\hat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \leq x).$$

Заметим, что случайная величина $I(X_j \leq x) \sim Bernouli(F(x))$, тогда

$$\sum_{j=1}^{n} I(X_j \leq x) = n\hat{F}_n(x) \sim Binom(n, F(x))$$

$$E(n\hat{F}_n(x)) = nF(x) \Rightarrow E(\hat{F}_n(x)) = F(x)$$

$$Var(n\hat{F}_n(x)) = nF(x)(1 - F_X(x)) \Rightarrow Var(\hat{F}_n(x)) = \frac{1}{n}F(x)(1 - F_X(x)) \to 0, n \to \infty.$$

По закону больших чисел

$$\hat{F}_n(x) \xrightarrow{\text{P}} F(x),$$

а по центральной предельной теореме

$$\frac{n\hat{F}_n(x) - nF(x)}{\sqrt{F(x)(1 - F_X(x))}} \xrightarrow{\text{d}} N(0, 1)$$

.

Теперь сформулируем теорему Колмогорова.
Теорема Колмогорова. Если $F(x)$ – непрерывная, то распределение случайной величины

$$D_n = \sup_{x \in R} |\hat{F}_n(x) - F(x)|$$

не зависит от $F(x)$. Более того,

$$\sqrt{n}D_n \xrightarrow{\text{d}} K(x),$$

где $K(x)$ – *рапределение Колмогорова* с функцией распределения $K(y) = \sum_{k \in Z} (-1)^k e^{-2k^2y^2}$.
Вся следующая часть доклада будет посвящена доказательству теоремы Колмогорова.

Лемма 1. Распределение статистики $D_n = \underset{x\in R}{sup}|\hat{F}_n(x) - F(x)|$ не зависит от $F(x)$.
Доказательство. Рассмотрим другую выборку $Y_j = F(X_j)$. Для любого $y \in (0,1)$:

$$F_Y(y) = P(Y_j \le y) = P(F(X_j) \le y) = F(x) = F(X_j \le F_X^{-1}(y)) = F(F^{-1}(y)) = y,$$

т.е. $Y_1, ..., Y_n \overset{iid}{\sim} U(0,1)$.
Далее заметим, что $I(X_j < x) = I(F(X_j) \le F(x))$.

$$D_n = \underset{x\in R}{sup}|\hat{F}_n(x) - F(x)| = \underset{x\in R}{sup}|\frac{1}{n}\sum_{j=1}^{n} I(X_j \le x) - F(x)| = \underset{x\in R}{sup}|\frac{1}{n}\sum_{j=1}^{n} I(F(X_j) \le F(x)) - F(x)| =$$

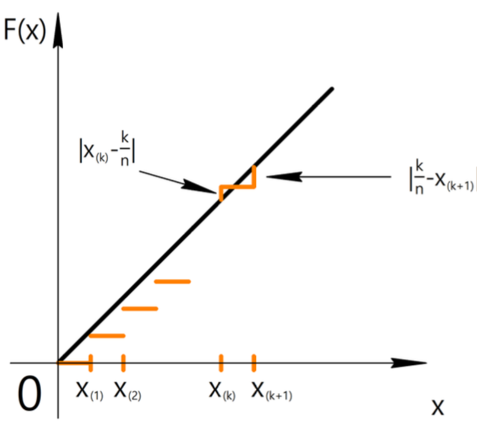
$$= \underset{x\in R}{sup}|\frac{1}{n}\sum_{j=1}^{n} I(Y_j \le F(x)) - F(x)| = \underset{y\in(0,1)}{sup}|\frac{1}{n}\sum_{j=1}^{n} I(Y_j \le y) - y|,$$

т.е. действительно, распределение последней случайной величины не зависит от $F(x)$.

Лемма 2. Пусть $W_t, t \in [0,1]$ – Винеровский процесс. Тогда

$$\sqrt{n}D_n \xrightarrow{\text{d}} \underset{t\in[0,1]}{max}|W_t - tW_1|.$$

Доказательство. Из леммы 1 будем считать, что $X_1, ..., X_n \overset{iid}{\sim} U(0,1)$ (iid). Заметим, что $\underset{t\in[0,1]}{max}|W_t - tW_1|$ может достигаться только в точках разрыва $\hat{F}_n(x)$.



$$D_n = max(\underset{k=1,...,n}{max}|X_{(k)} - \frac{k}{n}|, \underset{k=1,...,n}{max}|X_{(k)} - \frac{k-1}{n}|).$$

Вместо двух максимумов введем один: $T_n = \underset{k=1,...,n}{max}|X_{(k)} - \frac{k}{n}|$. Тогда $|D_n - T_n| \le \frac{1}{n}$. Поэтому достаточно показать, что $\sqrt{n}T_n \xrightarrow{\text{d}} \underset{t\in[0,1]}{max}|W_t - tW_1|$.

Задача. Пусть $X_1, ..., X_n \overset{iid}{\sim} U(0,1)$, $\xi_1, ..., \xi_{n+1} \overset{iid}{\sim} Exp(1)$, $S_k = \sum_{i=1}^{k} \xi_i$. Доказать, что

$$(X_{(1)}, ..., X_{(n)}) \overset{d}{=} (\frac{S_1}{S_{n+1}}, ..., \frac{S_n}{S_{n+1}})$$

Решение 1 [2]. Для начала отметим, что из определения производной получаем

$$f_X(x) = F_X'(x) = \lim_{\varepsilon\to 0}\frac{F_X(x+\varepsilon) - F_X(x)}{\varepsilon} = \lim_{\varepsilon\to 0}\frac{P(X \in (x, x+\varepsilon))}{\varepsilon}.$$

Значит при $\varepsilon \to 0$:

$$P(X \in (x, x+\varepsilon)) = f_X(x)\varepsilon.$$

С одной стороны, заметим, что для $0 \le dx_1 \le ... \le dx_n, x_i \in [0,1], i = 1, ..., n$

$$P(X_{(1)} \in dx_1, ..., X_{(n)} \in dx_n) = \sum P(X_{k_1} \in dx_1, ..., X_{k_n} \in dx_n),$$

где суммирование производится по всем перестановкам $(k_1, ..., k_n)$ чисел $(1, ..., n)$.

$$\sum P(X_{k_1} \in dx_1, ..., X_{k_n} \in dx_n) \stackrel{iid}{=} n \cdot 1 \cdot dx_1 \cdot (n-1) \cdot 1 \cdot dx_2 \cdot ... \cdot 1 \cdot 1 \cdot dx_n = n! dx_1 ... dx_n.$$

Теперь рассмотрим совместное распределение $S_1, ..., S_n$. Если $y_i \in [0,1], i = 1, ..., n+1$

$$P(S_1 \in dy_1, S_2 \in dy_2, ..., S_{n+1} \in dy_{n+1}) = P(\xi_1 \in dy_1, \xi_2 \in dy_2 - y_1, ..., \xi_{n+1} \in dy_{n+1} - y_n) \stackrel{iid}{=}$$

$$= e^{-y_1} e^{-(y_2 - y_1)} ... e^{-(y_{n+1} - y_n)} dy_1 .... dy_{n+1} = e^{-y_{n+1}} dy_1 .... dy_{n+1}.$$

Т.к. $S_{n+1} = \xi_1 + ... + \xi_{n+1}$, где $\xi_1, ..., \xi_{n+1} \stackrel{iid}{\sim} Exp(1)$, то случайная величина S_{n+1} имеет гамма распределение:

$$P(S_{n+1} \in dy_{n+1}) = \frac{y_{n+1}^n e^{-y_{n+1}}}{n!} dy_{n+1}.$$

По определению условной вероятности

$$P(S_1 \in dy_1, ..., S_n \in dy_n | S_{n+1} \in dy_{n+1}) = \frac{P(S_1 \in dy_1, S_2 \in dy_2, ..., S_{n+1} \in dy_{n+1})}{P(S_{n+1} \in dy_{n+1})} = \frac{n!}{y_{n+1}^n} dy_1 .... dy_n.$$

В результате получаем

$$P(\frac{S_1}{S_{n+1}} \in dx_1, ..., \frac{S_n}{S_{n+1}} \in dx_n | S_{n+1} = y_{n+1}) = P(S_1 \in (S_{n+1} dx_1), ..., S_n \in (S_{n+1} dx_n) | S_{n+1} = y_{n+1}) =$$

$$= \frac{n!}{y_{n+1}^n} y_{n+1}^n dx_1 ... dx_n = n! dx_1 ... dx_n$$

Т.к. правая часть последнего выражения не зависит от y_{n+1}, то утверждение доказано.

Решение 2. Про этот способ нам рассказал Борис Демешев. Другой способ найти распределение случайного вектора – через дифференциальные формы. А именно, т.к. геометрически 2-форма $\omega_1 \wedge \omega_2(v_1, v_2)$ – это ориентированная площадь параллелограмма, порожденного векторами v_1, v_2, то вместо функции распределения $F_{X,Y}(x,y)$ можем рассмотреть $f_{X,Y}(x,y) dx \wedge dy$.
Отметим, что по свойствам дифференциальных форм

$$dx \wedge dy = -dy \wedge dx$$

$$dx \wedge dx = 0.$$

Совместная функция плотности

$$F_{\xi_1, ..., \xi_{n+1}}(t_1, ..., t_{n+1}) = e^{-t_1} ... e^{-t_{n+1}} dt_1 \wedge ... \wedge dt_{n+1}.$$

Обозначим $(\frac{S_1}{S_{n+1}}, ..., \frac{S_n}{S_{n+1}}, S_{n+1}) = (R_1, ..., R_n, S)$.

$$t_1 = SR_1$$

$$t_2 = S(R_2 - R_1)$$

$$...$$

$$t_n = S(R_n - R_{n-1})$$

$$t_{n+1} = S$$

Воспользовавшись свойствами дифференциальных форм, получим

$$dt_1 \wedge ... \wedge dt_n \wedge dS = d(SR_1) \wedge d(SR_2 - SR_1) \wedge ... \wedge d(SR_n - SR_{n-1}) \wedge dS =$$

$$= d(SR_1) \wedge d(SR_2) \wedge ... \wedge d(SR_n) \wedge dS = S^n dR_1 \wedge ... \wedge dR_n \wedge dS.$$

Функция распределения вектора $(\xi_1, ..., \xi_{n+1})$

$$e^{-t_1} ... e^{-t_{n+1}} dt_1 \wedge ... \wedge dt_{n+1} = e^{-S} S^n dR_1 \wedge ... \wedge dR_n \wedge dS.$$

Плотность $(R_1, ..., R_n)$ пропорциональна 1, значит чтобы объем под ней был равен 1, распределение вектора $(R_1, ..., R_n) = (\frac{S_1}{S_{n+1}}, ..., \frac{S_n}{S_{n+1}})$ равно $n!$.

Из результатов решенной задачи получаем

$$\sqrt{n} T_n = \sqrt{n} \max_{k=1,\dots,n} |X_{(k)} - \frac{k}{n}| \stackrel{d}{=} \sqrt{n} \max_{k=1,\dots,n} |\frac{S_k}{S_{n+1}} - \frac{k}{n}| \xrightarrow{\text{d}} \sqrt{n+1} \max_{k=1,\dots,n} |\frac{S_k}{S_{n+1}} - \frac{k}{n+1}| := \Delta_n$$

.

$$\Delta_n = \sqrt{n+1} \max_{k=1,\dots,n} |\frac{S_k}{S_{n+1}} - \frac{k}{n}| = \frac{n+1}{S_{n+1}} \max_{k=1,\dots,n} |\frac{S_k - k}{\sqrt{n+1}} - \frac{k}{n+1} \frac{S_{n+1} - (n+1)}{\sqrt{n+1}}|.$$

Принцип инвариантности Доскера-Прохорова. Пусть $\xi_1, ..., \xi_n \sim$ iid,

$$E(\xi_i) = 0, Var(\xi_i) = 1, \forall i = 1, ..., n; S_k = \sum_{i=1}^{k} \xi_i, S_0 = 0.$$

Введем последовательность случайных величин $X^{(n)} = (X_t^{(n)}, t \in [0, 1])$ как линейную интерполяцию $S_0, S_1, ..., S_n$: для $t \in [\frac{k}{n}, \frac{k+1}{n}]$

$$X^{(n)} = \frac{S_k}{\sqrt{n}}(k + 1 - nt) + \frac{S_{k+1}}{\sqrt{n}}(nt - k), k = 0, ..., n - 1.$$

Тогда

$$X^{(n)} \xrightarrow{\text{d}} W = (W_t, t \in [0, 1]).$$

Введем случайный процесс $Y^{(n)} = (Y_t^{(n)}, t \in [0, 1])$: для $t \in [\frac{k}{n+1}, \frac{k+1}{n+1}]$

$$Y_t^{(n)} = \frac{S_k - k}{\sqrt{n+1}}(k + 1 - (n + 1)t) + \frac{S_{k+1} - (k + 1)}{\sqrt{n+1}}((n + 1)t - k), k = 0, ..., n.$$

Это процесс из принципа инвариантности Донскера-Прохорова, т.к. $E(\xi_i) = 0, Var(\xi_i) = 1$. Заметим, что

$$Y_t^{(n)}(t = \frac{k}{n+1}) = \frac{S_k - k}{\sqrt{n+1}}$$

$$Y_t^{(n)}(t = 1) = \frac{S_{n+1} - (n + 1)}{\sqrt{n+1}}.$$

То есть при рассмотрении процесса Δ_n мы рассматриваем процесс $Y_t^{(n)} - tY_1^{(n)}$ и берем максимальные по модулю значения в точках вида $\frac{k}{n+1}$: $\Delta_n = \frac{n+1}{S_{n+1}} \max_{t \in [0,1]} |Y_t^{(n)} - tY_1^{(n)}|$. Из закона больших чисел $\frac{n+1}{S_{n+1}} \to 1$.

Тогда

$$\max_{t \in [0,1]} |Y_t^{(n)} - tY_1^{(n)}| \xrightarrow{\text{d}} \max_{t \in [0,1]} |W_t - tW_1|$$

$$\Delta_n \xrightarrow{\text{d}} \max_{t \in [0,1]} |W_t - tW_1|.$$

Мы показали, что статистика, которая участвует в формулировке теста Колмогорова сходится по распределению к максимуму модуля броуовского моста. Теперь остается вывести его распределение.

Обозначим броуновский мост $W_t^0 = W_t - tW_1, t \in [0,1]$. Для начала отметим, что распределение броуновского моста можно вычислить как условное распределение броуновского движения. Тогда $\forall z > 0$

$$P(\underset{t\in[0,1]}{max}|W_t^0| < z) = \lim_{\varepsilon\to 0} P(\underset{t\in[0,1]}{max}|W_t| < z | W_1 \in [-\varepsilon,\varepsilon]) = \lim_{\varepsilon\to 0} \frac{P(\underset{t\in[0,1]}{max} W_t < z, \underset{t\in[0,1]}{min} W_t > -z, W_1 \in [-\varepsilon,\varepsilon])}{P(W_1 \in [-\varepsilon,\varepsilon])}.$$

Теперь надо найти совместное распределение (m, M, W_1), где $m = \underset{t\in[0,1]}{min} W_t, M = \underset{t\in[0,1]}{max} W_t$.
Для этого сначала решим аналогичную задачу для симметричного случайного блуждания. Обозначим S_n – симметричное случайное блуждание, $m_n = \underset{1\le k\le n}{min} S_k, M = \underset{1\le k\le n}{max} S_k$.

Лемма 3. Для любых целых чисел $a \le 0 \le b, a < b, a < v < b$ выполнено

$$P(a < m_n \le M_n < b, S_n = v) = \sum_{k\in Z} P(S_n = v + 2k(b-a)) - \sum_{k\in Z} P(S_n = 2b - v + 2k(b-a)).$$

Доказательство. Будем доказывать методом мат. индукции.
1) База: $n = 0$.
1. $v = 0, a < 0 < b$
LHS $= P(a < m_n \le M_n < b, S_n = v) = P(a < 0 \le 0 < b, S_n = 0) = 1$
RHS: $\sum_{k\in Z} P(S_n = v + 2k(b-a)) = 1$, $\sum_{k\in Z} P(S_n = 2b - v + 2k(b-a)) = 0$.
2. $v \ne 0, a < 0 < b$
LHS $= P(a < m_n \le M_n < b, S_n = v) = 0$, т.к. $S_0 = 0$
3. $a = 0$ или $b = 0$
LHS $= P(a < m_n \le M_n < b, S_n = v) = 0$
RHS $= \sum_{k\in Z} P(S_n = v + 2k(b-a)) - \sum_{k\in Z} P(S_n = 2b - v + 2k(b-a))$ т.к. суммы совпадают.
2) Пусть верно для $n-1$, тогда докажем что то же будет верно для n.
1. $a = 0$ или $b = 0$
LHS $= P(a < m_n \le M_n < b, S_n = v) = 0$
RHS $= \sum_{k\in Z} P(S_n = v + 2k(b-a)) - \sum_{k\in Z} P(S_n = 2b - v + 2k(b-a))$ т.к. суммы совпадают.
2. $a < 0 < b$
Для начала обозначим $LHS = P(a < m_n \le M_n < b, S_n = v) := p_n(a,b,v)$, а $p_n(j) := P(S_n = j)$.
Во-первых, заметим, что

$$p_n(j) = \frac{1}{2}p_{n-1}(j-1) + \frac{1}{2}p_{n-1}(j+1) \tag{1}$$

Действительно, заметим, что $S_n = X_1 + ... + X_n = S_{n-1} + X_n$. Чтобы оказаться в точке $S_n = j$ на шаге n, на шаге $n-1$ мы находимся либо в $S_{n-1} = j-1$, либо в $S_{n-1} = j+1$. Если $S_{n-1} = j-1$, то чтобы оказаться в $S_n = j$, $X_n = 1$, а $P(X_n = 1) = \frac{1}{2}$, т.к. мы рассматриваем симметричное случайное блуждание. Аналогично в случае $S_{n-1} = j+1$. По формуле полной вероятности получаем ответ.

Во-вторых, заметим, что

$$p_n(a,b,v) = \frac{1}{2}p_{n-1}(a-1,b-1,v-1) + \frac{1}{2}p_{n-1}(a+1,b+1,v+1) \tag{2}$$

Опять же, чтобы $S_n = v$, необходимо, чтобы $S_{n-1} = v-1$ или $S_{n-1} = v+1$. Если $S_n = v$ и $S_{n-1} = v-1$, то $X_n = 1$, тогда чтобы $m_n > a, M_n < b$ должно выполняться $m_{n-1} > a-1, M_{n-1} < b-1$. Аналогично, если $S_n = v$ и $S_{n-1} = v+1$, то $X_n = -1$, тогда чтобы $m_n > a, M_n < b$ должно выполняться $m_{n-1} > a+1, M_{n-1} < b+1$. С учетом $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$,из формулы полной вероятности получаем $p_n(a,b,v) = \frac{1}{2}p_{n-1}(a-1,b-1,v-1) + \frac{1}{2}p_{n-1}(a+1,b+1,v+1)$.

Тогда из (2) получим:

$$p_n(a,b,v) = \frac{1}{2}p_{n-1}(a-1,b-1,v-1) + \frac{1}{2}p_{n-1}(a+1,b+1,v+1) =$$

[по предположению индукции]

$$= \frac{1}{2}[\sum_{k\in Z} p_{n-1}(v-1+2k(b-a)) - \sum_{k\in Z} p_{n-1}(2b-v-1+2k(b-a))] + \frac{1}{2}[\sum_{k\in Z} p_{n-1}(v+1+2k(b-a)) - \sum_{k\in Z} p_{n-1}(2b-v+1+2k(b-a))] =$$

$$= \frac{1}{2}[\sum_{k\in Z} p_{n-1}(v-1+2k(b-a)) + \sum_{k\in Z} p_{n-1}(v+1+2k(b-a))] - \frac{1}{2}[\sum_{k\in Z} p_{n-1}(2b-v-1+2k(b-a)) + \sum_{k\in Z} p_{n-1}(2b-v+1+2k(b-a))] =$$

$$= \sum_{k\in Z}[\frac{1}{2}p_{n-1}(v-1+2k(b-a)) + \frac{1}{2}p_{n-1}(v+1+2k(b-a))] - \sum_{k\in Z}[\frac{1}{2}p_{n-1}(2b-v-1+2k(b-a)) + \frac{1}{2}p_{n-1}(2b-v+1+2k(b-a))] =$$

$$= \sum_{k\in Z} p_n(v+2k(b-a)) - \sum_{k\in Z} p_n(2b-v+2k(b-a)).$$

Утверждение доказано.

Следствие. Для любых целых чисел $a \le 0 \le b, a < b, a \le u_1 < u_2 \le b$ выполнено

$$P(a < m_n \le M_n < b, u_1 < S_n < u_2) = \sum_{k\in Z} P(u_1 + 2k(b-a) < S_n < u_2 + 2k(b-a)) -$$

$$- \sum_{k\in Z} P(2b - u_2 + 2k(b-a) < S_n < 2b - u_1 + 2k(b-a)).$$

Доказательство. Получается суммированием по значениям $v \in (u_1, u_2)$ из леммы 3.

Теперь сформулируем лемму о совместном распределении (m, M, W_1).
Лемма 4. Для любых вещественных $a < 0 < b, a < u < v < b$ выполнено

$$P(a < m \le M < b, u < W_1 < v) = \sum_{k\in Z} P(u+2k(b-a) < \xi < v+2k(b-a)) - \sum_{k\in Z} P(2b-v+2k(b-a) < \xi < 2b-u+2k(b-a)),$$

где $\xi \sim N(0,1)$.
Доказательство. В силу принципа инвариантности и наследования сходимости получаем, что

$$(\frac{m_n}{\sqrt{n}}, \frac{M_n}{\sqrt{n}}, \frac{S_n}{\sqrt{n}}) \xrightarrow{\text{d}} (m, M, W_1).$$

Тогда

$$P(a < m \le M < b, u < W_1 < v) = \lim_{n\to\infty} P(a < \frac{m_n}{\sqrt{n}} \le \frac{M_n}{\sqrt{n}} < b, u < \frac{S_n}{\sqrt{n}} < v).$$

Сделаем границы целыми: обозначим $a' = [a\sqrt{n}], b' = [b\sqrt{n}], u' = [u\sqrt{n}], v' = [v\sqrt{n}]$. Тогда из следствия получаем

$$P(a < m \le M < b, u < W_1 < v) = \lim_{n\to\infty} P(a < \frac{m_n}{\sqrt{n}} \le \frac{M_n}{\sqrt{n}} < b, u < \frac{S_n}{\sqrt{n}} < v) =$$

$$= \lim_{n\to\infty}(\sum_{k\in Z} P(u'+2k(b'-a') < S_n < v'+2k(b'-a')) - \sum_{k\in Z} P(2b'-v'+2k(b'-a') < S_n < 2b'-u'+2k(b'-a'))).$$

При фиксированном k в силу ЦПТ выполнено

$$\lim_{n\to\infty} P(u' + 2k(b'-a') < S_n < v' + 2k(b'-a')) = \lim_{n\to\infty} P(\frac{u' + 2k(b'-a')}{\sqrt{n}} < \frac{S_n}{\sqrt{n}} < \frac{v' + 2k(b'-a')}{\sqrt{n}}) =$$

$$= P(u + 2k(b-a) < \xi < v + 2k(b-a)).$$

Аналогично для второй суммы.
Можно показать что ряды

$$\sum_{k\in Z} P(\frac{u' + 2k(b'-a')}{\sqrt{n}} < \frac{S_n}{\sqrt{n}} < \frac{v' + 2k(b'-a')}{\sqrt{n}})$$

и

$$P(\frac{2b'-v'+2k(b'-a')}{\sqrt{n}} < \frac{S_n}{\sqrt{n}} < \frac{2b'-u'+2k(b'-a')}{\sqrt{n}})$$

сходятся равномерно. Значит можем переставить местами сумму по k и предел по n [3]. Таким образом, получаем исходное равенство.

Подставим нужные параметры из леммы 4 в полученное ранее представление распределения броуновского моста через условное распределение броуновского движения. Пусть

$$u=-\varepsilon, v=\varepsilon, a=-z, b=z.$$

Тогда получим,

$$P(\max_{t\in[0,1]}|W_t^0|<z)=\lim_{\varepsilon\to 0}P(\max_{t\in[0,1]}|W_t|<z|W_1\in[-\varepsilon,\varepsilon])=\lim_{\varepsilon\to 0}\frac{P(\max_{t\in[0,1]}W_t<z,\ \min_{t\in[0,1]}W_t>z, W_1\in[-\varepsilon,\varepsilon])}{P(W_1\in[-\varepsilon,\varepsilon])}=$$

$$=\lim_{\varepsilon\to 0}[\frac{1}{P(|W_1|\le\varepsilon)}\sum_{k\in Z}P(4kz-\varepsilon<\xi<4kz+\varepsilon)-\frac{1}{P(|W_1|\le\varepsilon)}\sum_{k\in Z}P(2z-\varepsilon+4kz<\xi<2z+\varepsilon+4kz)].$$

Заметим, что для фиксированного k при $\varepsilon\to 0$:

$$P(4kz-\varepsilon<\xi<4kz+\varepsilon)=\int_{-\varepsilon}^{\varepsilon}\frac{1}{\sqrt{2\pi}}e^{-\frac{(y+4kz)^2}{2}}\,dy \sim [\text{по формуле средних}] \sim 2\varepsilon\frac{1}{\sqrt{2\pi}}e^{-\frac{(4kz)^2}{2}};$$
$$P(2z-\varepsilon+4kz<\xi<2z+\varepsilon+4kz)=\int_{-\varepsilon}^{\varepsilon}\frac{1}{\sqrt{2\pi}}e^{-\frac{(y+(4k+2)z)^2}{2}}\,dy\sim 2\varepsilon\frac{1}{\sqrt{2\pi}}e^{-\frac{((4k+2)z)^2}{2}};$$
$$P(-\varepsilon<W_1<\varepsilon)=\int_{-\varepsilon}^{\varepsilon}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\,dy\sim 2\varepsilon\frac{1}{\sqrt{2\pi}}.$$

Опять же, можно показать что ряды сходятся равномерно, значит можем переставить местами суммирование по k и взятие предела по ε. Получаем,

$$P(\max_{t\in[0,1]}|W_t^0|<z)=\lim_{\varepsilon\to 0}[\frac{1}{P(|W_1|\le\varepsilon)}\sum_{k\in Z}P(4kz-\varepsilon<\xi<4kz+\varepsilon)-\frac{1}{P(|W_1|\le\varepsilon)}\sum_{k\in Z}P(2z-\varepsilon+4kz<\xi<2z+\varepsilon+4kz)]=$$

$$=\lim_{\varepsilon\to 0}[\sum_{k\in Z}\frac{1}{P(|W_1|\le\varepsilon)}P(4kz-\varepsilon<\xi<4kz+\varepsilon)-\sum_{k\in Z}\frac{1}{P(|W_1|\le\varepsilon)}P(2z-\varepsilon+4kz<\xi<2z+\varepsilon+4kz)]=$$

$$=\sum_{k\in Z}\lim_{\varepsilon\to 0}\frac{1}{P(|W_1|\le\varepsilon)}P(4kz-\varepsilon<\xi<4kz+\varepsilon)-\sum_{k\in Z}\lim_{\varepsilon\to 0}\frac{1}{P(|W_1|\le\varepsilon)}P(2z-\varepsilon+4kz<\xi<2z+\varepsilon+4kz)=$$

$$=\sum_{k\in Z}e^{-\frac{(4kz)^2}{2}}-\sum_{k\in Z}e^{-\frac{((4k+2)z)^2}{2}}=\sum_{k\in Z}e^{-2(2k)^2z^2}-\sum_{k\in Z}e^{-2(2k+1)^2z^2}=\sum_{k\in Z}(-1)^k e^{-2k^2z^2}:=K(z).$$

Если $D_n>k_{1-\alpha}$, где k_p – квантиль распределения Колмогорова, то нулевая гипотеза отвергается. Квантили распределения Колмогорова указаны в таблице:

α	0,5	0,15	0,1	0,05	0,025	0,01	0,001
$x_{1-\alpha}$	0,83	1,14	1,23	1,36	1,48	1,63	1,95

Материалы

[1] Шабанов Д. А. Математическая статистика. URL: ссылка на видеолекции.
[2] Ширяев А. Н. Вероятность – 1. – Москва: Издательство МЦНМО, 2004.
[3] Ширяев А. Н. Броуновское движение и винеровская мера. Теория, применения, аналитические методы: В 2 т. Т. 1. – М.: МЦНМО, 2023.
[4] Зорич В. А. Математический анализ. Часть 2. Москва: Издательство МЦНМО, 2019.