ДЕВЯТАЯ КОЛМОГОРОВСКАЯ СТУДЕНЧЕСКАЯ ОЛИМПИАДА ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ

Решения задач

Задача 1. Случайная величина $X \sim Pois(\lambda)$, где $\lambda > 0$. Доказать, что
$$\mathsf{P}(X > a) \le e^{-\lambda g(a/\lambda)}$$
для каждого $a > \lambda$, здесь $g(x) = 1 - x + x \ln x$.

Решение. По неравенству Маркова для каждого $t > 0$
$$\mathsf{P}(X > a) = \mathsf{P}(e^{tX} > e^{ta}) \le e^{-ta}\mathsf{E}e^{Xt} =$$
$$= e^{-ta}\sum_{k=0}^{\infty}\frac{\lambda^k e^{tk}}{k!}e^{-\lambda} = \exp\{-ta + \lambda e^t - \lambda\}.$$
Полагаем
$$f(t) = -ta + \lambda e^t,$$
$$f'(t) = 0 \text{ при } t = \ln(a/\lambda).$$
При подстановке этого значения получаем то, что нужно.

Задача 2. Случайные величины $X_1, X_2, \ldots$ независимы, одинаково распределены и принимают только целочисленные значения. Известно, что $0 < \mathsf{P}(X_1 \text{ делится на } 3) < 1$. Пусть $S_n = X_1 + \ldots + X_n, n \in \mathbb{N}$. Найти $\lim_{n\to\infty} \mathsf{P}(S_n \text{ делится на } 3)$.

Решение. Сначала заметим, что для целочисленной случайной величины $\eta \ge 0$
$$\mathsf{P}(\eta \text{ делится на } 2) = \sum_{k=0}^{\infty}\mathsf{P}(\eta = 2k) =$$
$$= \frac{1}{2}\left(\sum_{k=0}^{\infty}\mathsf{P}(\eta = k) + \sum_{k=0}^{\infty}(-1)^k\mathsf{P}(\eta = k)\right) =$$
$$= \frac{\varphi(1) + \varphi(-1)}{2},$$
где производящая функция $\varphi(s) = \mathsf{E}s^{\eta}$. Аналогично пусть $s_1 = 1, s_2, s_3$ – корни 3 степени из 1, тогда
$$\mathsf{P}(\eta \text{ делится на } 3) = \sum_{k=0}^{\infty}\mathsf{P}(\eta = k)I\{k = 3l, l \in \mathbb{Z}\} =$$

$$= \frac{1}{3}\sum_{k=0}^{\infty} \mathsf{P}(\eta = k)\sum_{j=1}^{3} s_j^k = \frac{1}{3}\sum_{j=1}^{3}\sum_{k=0}^{\infty} \mathsf{P}(\eta = k) s_j^k = \frac{1}{3}\sum_{j=1}^{3}\varphi(s_j).$$

Можно считать сразу, что X_i принимают значения 0,1 и 2 с вероятностями p_0, p_1, p_2 соответственно. Производящая функция S_n равна $(p_0 + p_1 s + p_2 s^2)^n$. Когда $s \neq 1$, но $|s| = 1$, выражение в скобках по модулю строго меньше 1. Поэтому

$$\frac{1}{3}\sum_{j=1}^{3}(p_0 + p_1 s_j + p_2 s_j^2)^n \to \frac{1}{3}, \quad n \to \infty.$$

Для знакомых с цепями Маркова можно привести и более простое решение. Последовательность остатков от деления S_n на 3 образует цепь Маркова с переходной матрицей

$$P = \begin{pmatrix} p_0 & p_1 & p_2 \\ p_2 & p_0 & p_1 \\ p_1 & p_2 & p_0 \end{pmatrix}$$

Эта цепь Маркова удовлетворяет условиям эргодической теоремы (например, потому, что все элементы матрицы P^2 положительны). Предельное распределение задается вектором μ, являющимся решением уравнения $P^T\mu = \mu$, но такое решение имеет вид $(1/3, 1/3, 1/3)$.

Задача 3. Случайные величины X и Y имеют стандартное нормальное распределение и независимы, а функция $f : \mathbb{R} \to \mathbb{R}$ гладкая, выпуклая и неотрицательная. Доказать неравенство $cov(f(X+Y)X, f(X+Y)Y) \geq 0$ (предполагается, что данная ковариация существует).

Решение. Сначала будем считать, что

$$\lim_{x\to\infty} e^{-x^2/2}(|f(x)| + |f'(x)| + |f''(x)|)^2 = 0.$$

Тогда по частям

$$\mathsf{E}f(X+Y)X = \frac{1}{2\pi}\iint_{\mathbb{R}^2} f(x+y)xe^{-x^2/2-y^2/2}dxdy =$$
$$= \frac{1}{2\pi}\iint_{\mathbb{R}^2} f'(x+y)e^{-x^2/2-y^2/2}dxdy.$$

Далее, снова интегрируя по частям, имеем

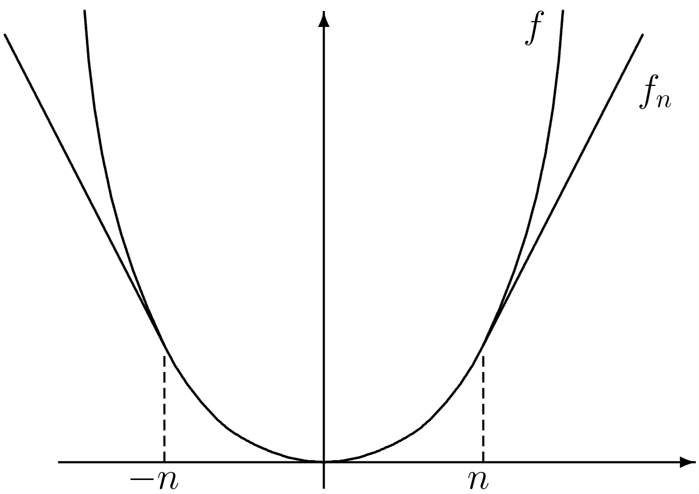
$$\mathsf{E}f^2(X+Y)XY = \frac{1}{2\pi}\iint_{\mathbb{R}^2} f^2(x+y)xye^{-x^2/2-y^2/2}dxdy =$$
$$= \frac{1}{2\pi}\iint_{\mathbb{R}^2} 2f(x+y)f'(x+y)ye^{-x^2/2-y^2/2}dydx$$
$$= \frac{1}{\pi}\iint_{\mathbb{R}^2} (f(x+y)f''(x+y) + (f'(x+y))^2)e^{-x^2/2-y^2/2}dydx.$$

Таким образом, ковариация равна

$$\frac{1}{\pi}\iint_{\mathbb{R}^2} (f(x+y)f''(x+y) + (f'(x+y))^2)e^{-x^2/2-y^2/2}dydx-$$
$$-\left(\frac{1}{2\pi}\iint_{\mathbb{R}^2} f'(x+y)e^{-x^2/2-y^2/2}dydx\right)^2 =$$
$$= \mathsf{D}f'(X+Y) + \frac{1}{\pi}\iint_{\mathbb{R}^2} f(x+y)f''(x+y)e^{-x^2/2-y^2/2}dydx+$$
$$+\frac{1}{2\pi}\iint_{\mathbb{R}^2} (f'(x+y))^2e^{-x^2/2-y^2/2}dydx \geq 0.$$

и все три слагаемых неотрицательны.

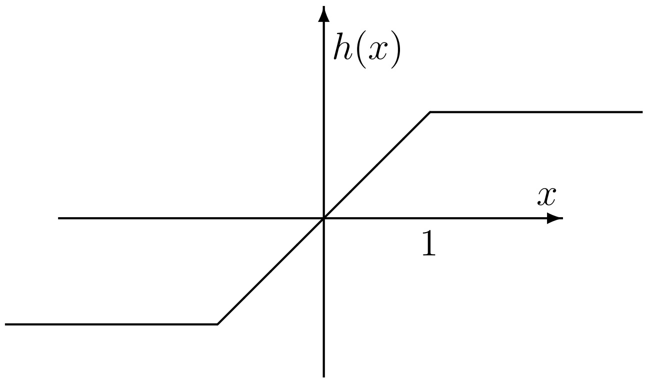
Если f слишком быстро растет на бесконечности, то определим $f_n(x)$ так:



Функция f_n в бесконечности растет не быстрее линейной, и по доказанному $cov(f_n(X+Y)X, f_n(X+Y)Y) \geq 0$. Остается применить теорему Лебега о мажорированной сходимости.

Задача 4. Пусть случайная величина X имеет конечную дисперсию, а функция $h(x) = \min(|x|, 1) sgn(x)$. Верно ли, что $\mathsf{D}h(X) \leq \mathsf{D}X$?

Решение. Да, верно.



Как известно,

$$\mathsf{D}\eta = \min_{a\in\mathbb{R}} \mathsf{E}(\eta - a)^2.$$

Так как для любых $x, y \in \mathbb{R}$ выполняется неравенство $|h(x) - h(y)| \leq |x - y|$, имеем

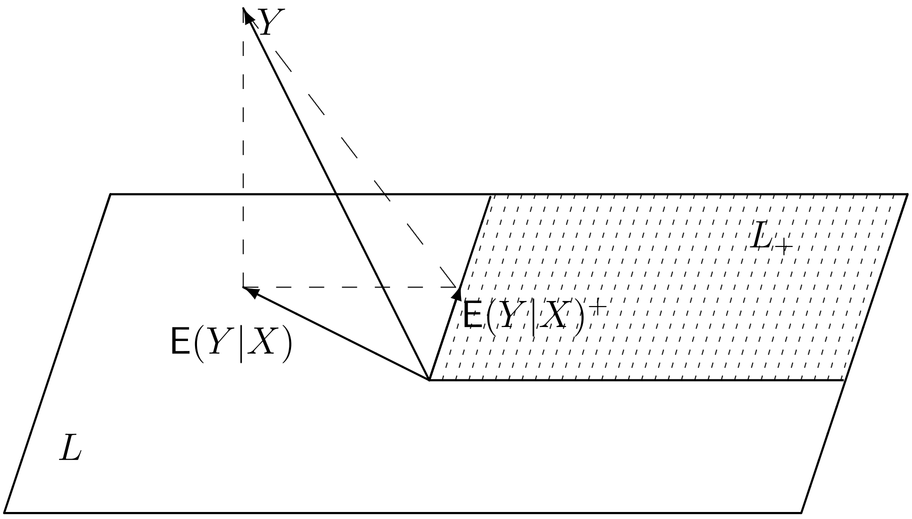
$$\mathsf{D}h(X) \leq \mathsf{E}(h(X) - h(\mathsf{E}X))^2 \leq \mathsf{E}(X - \mathsf{E}X)^2 = \mathsf{D}X.$$

Задача 5. а) **(1-2)** Пусть X и Y — квадратично-интегрируемые случайные величины на одном вероятностном пространстве, причем X принимает конечное число значений. Найти функцию $f : \mathbb{R} \to \mathbb{R}_+$, на которой достигается $\min \mathsf{E}(f(X) - Y)^2$ по всем неотрицательным f. б) **(3-5)** Тот же вопрос, но без условия конечности, а минимум берется по всем борелевским f, для которых $\mathsf{E}f^2(X) < \infty$.

Решение. Рассмотрим линейное пространство, порожденное случайной величиной Y и всеми величинами из $L = L^2(\Omega, \sigma(X), \mathsf{P})$ (в случае а) оно конечномерно), и в нем евклидову норму

$$\|\xi\|^2 = \mathsf{E}\xi^2.$$

Тогда $\min_f \mathsf{E}(Y - f(X))^2$ дается проекцией Y на $L^2(\Omega, \sigma(X), \mathsf{P})$:



Множество неотрицательных функций от X — выпуклый конус $L_+ \subset L$. Поэтому его ближайшая к Y точка находится так: сначала находим ближайшую точку в L (это $g(X) = \mathsf{E}(Y|X)$), а от нее — ближайшую в L_+. Но

$$\min_{f \geq 0} \mathsf{E}(f(X) - g(X))^2 = \mathsf{E}(f(X) - f^+(X))^2.$$

Ответ: $\mathsf{E}(Y|X)^+$.

Задача 6. В ящике лежат 1 белый и 99 черных шаров. Не глядя достанем каждой рукой по одному шару и тот шар, который оказался в левой руке, перекрасим в цвет шара в правой, а затем вернем шары назад. Найти математическое ожидание числа таких действий, после которого все шары в ящике станут одного цвета.

Решение. Пусть τ_k — математическое ожидание времени достижения этого момента, если в начальный момент в ящике k белых и $100 - k$ черных шаров. Тогда $\tau_0 = \tau_{100} = 0$. Если $0 < k < 100$, то на первом шаге перекраска в белый цвет произойдет с вероятностью

$$p_k = \frac{k(100 - k)}{99 \cdot 100},$$

с такой же вероятностью — перекраска в черный. Итак,

$$\tau_k = 1 + p_k(\tau_{k-1} + \tau_{k+1}) + (1 - 2p_k)\tau_k,$$

$$2p_k\tau_k = 1 + p_k(\tau_{k-1} + \tau_{k+1}), \quad k = 1, \ldots, 99,$$

$$\tau_k = \frac{1}{2p_k} + \frac{1}{2}(\tau_{k-1} + \tau_{k+1}), \quad k = 1, \ldots, 99.$$

Возьмем сумму по всем k от 1 до 99 :

$$\sum_{k=1}^{99} \tau_k = \sum_{k=1}^{99} \frac{1}{2p_k} + \frac{1}{2}\left(\sum_{k=1}^{98} \tau_k + \sum_{k=2}^{99} \tau_k\right).$$

После сокращения остается

$$\tau_1 + \tau_{99} = \sum_{k=1}^{99} \frac{1}{p_k}.$$

В силу симметрии между белыми и черными шарами

$$\tau_1 = \sum_{k=1}^{99} \frac{1}{2p_k} = 99 \cdot 50 \sum_{k=1}^{99} \frac{1}{100}\left(\frac{1}{k} + \frac{1}{100-k}\right).$$

Ответ:$99\sum_{k=1}^{99} k^{-1}$.

Задача 7. Случайная величина X неотрицательна и не зависит от величины $Y \sim N(0,1)$, а их произведение распределено по Лапласу (плотность $\lambda e^{-\lambda|x|}/2$, где $\lambda > 0$ — известный параметр). Найти плотность X.

Решение. Найдем характеристическую функцию XY:

$$\mathsf{E}e^{itXY} = \frac{\lambda^2}{\lambda^2 + t^2} = \mathsf{E}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{itXy} e^{-y^2/2} dy = \mathsf{E}e^{-t^2X^2/2}$$

в силу теоремы Фубини. Итак, для $s > 0$

$$\mathsf{E}e^{-sX^2} = \frac{\lambda^2/2}{\lambda^2/2 + s}.$$

Рассуждая так же с таким $t \in \mathbb{C}$, что $t^2/2 = is$, получаем

$$\mathsf{E}e^{isX^2} = \frac{\lambda^2/2}{\lambda^2/2 + is},$$

это характеристическая функция показательной случайной величины с параметром $\lambda^2/2$. Знакомые с понятием преобразования Лапласа и теоремой единственности для голоморфных функций могут также заметить, что характеристическая функция неотрицательной случайной величины получается из преобразования Лапласа заменой аргумента s на $-is$. Зная функцию распределения X^2, находим плотность самой X:

$$-\frac{d\mathsf{P}(X \geq t)}{dt} = -\frac{d\mathsf{P}(X^2 \geq t^2)}{dt} = -\frac{de^{-\lambda^2 t^2/2}}{dt} = \frac{\lambda^2 t}{2} e^{-\lambda^2 t^2/2}.$$

Ответ: $\lambda^2 t e^{-\lambda^2 t^2/2}/2$.

Задача 8. Пусть на плоскости задано пуассоновское поле точек интенсивности $\lambda > 0$ (т.е. случайно разбросаны точки так, что если ограниченные борелевские множества $B_1, \ldots, B_n$ попарно не пересекаются, то числа точек в них независимы и имеют распределение Пуассона с параметрами $\lambda|B_1|, \ldots, \lambda|B_n|$ соответственно; $|A|$ обозначает меру Лебега A). Каждую точку $x \in \mathbb{Z}^2$ соединим отрезком с ближайшей к ней точкой пуассоновского поля. Найти математическое ожидание суммарной длины всех частей таких отрезков, находящихся внутри квадрата $[0,1]^2$.

Решение.

Лемма (транспортный принцип): если функция $f : \mathbb{Z}^d \times \mathbb{Z}^d \to \mathbb{R}_+$ такова, что $f(x+h, y+h) = f(x,y)$ при всех $x, y, h \in \mathbb{Z}^d$, то

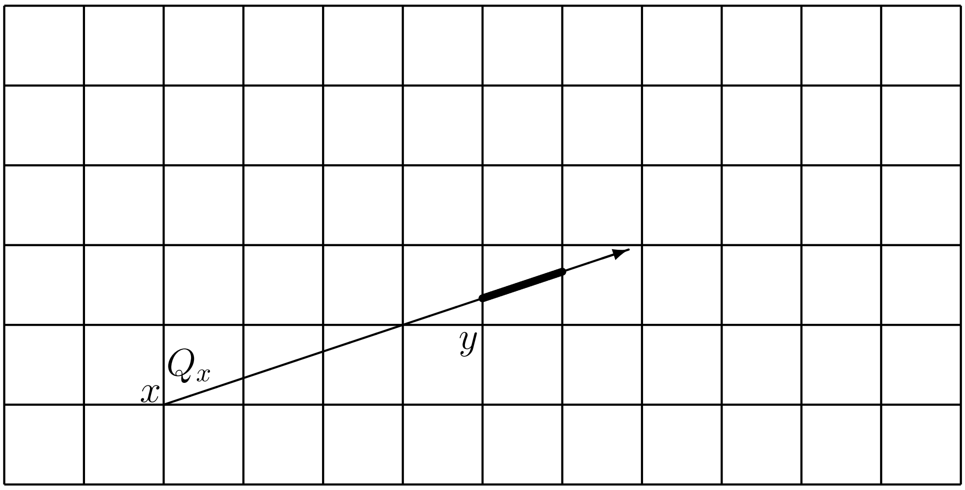
$$\sum_{y \in \mathbb{Z}^d} f(0,y) = \sum_{x \in \mathbb{Z}^d} f(x,0).$$

Доказательство леммы.

$$\sum_{y \in \mathbb{Z}^d} f(0,y) = \sum_{y \in \mathbb{Z}^d} f(-y,0) = \sum_{x \in \mathbb{Z}^d} f(x,0). \quad \square$$

Пусть $M(x, B)$ — длина той части отрезка, выходящего из $x \in \mathbb{Z}^2$, которая лежит внутри множества B. Обозначим квадрат $Q_x = [0,1)^2 + x$ (для $x \in \mathbb{Z}^2$) и применим лемму к

$$f(x,y) = \mathsf{E} M(x, Q_y).$$



$f(x,y)$ — математическое ожидание длины жирного отрезка

Тогда по лемме

$$\mathsf{E}\sum_{x\in\mathbb{Z}^d} M(x,Q_0) = \mathsf{E}M(x,\mathbb{R}^2).$$

Итак, нужно найти ожидание длины X того из проведенных отрезков, который отложен от нуля. т.е.

$$\int_0^\infty \mathsf{P}(X>t)dt = \int_0^\infty e^{-\lambda\pi t^2}dt = \frac{1}{\sqrt{2\lambda\pi}}\sqrt{\pi/2} = \frac{1}{2\sqrt{\lambda}}.$$

Ответ: $1/2\sqrt{\lambda}$.

Задача 9. Пусть случайная величина X, имеющая непрерывную плотность, сосредоточена на отрезке $[0,\pi]$ и такова, что распределения $\cos X$ и $\cos 2X$ совпадают. Доказать, что распределение X — равномерное на $[0,\pi]$.

Решение. Прежде всего заметим, что $\cos X$ и $\cos 2^n X$ распределены одинаково при каждом $n>1$, так как

$$\cos 2^{n+1}X \overset{d}{=} 2\cos^2 2^n X - 1 \overset{d}{=} 2\cos^2 2^{n-1}X - 1 \overset{d}{=} \cos 2^n X.$$

Пусть множество $A\subset[0,2\pi]$ борелевское и p — плотность X. Зададим $\varepsilon>0$ и найдем настолько большое n, чтобы на каждом отрезке длины не больше $\pi/2^n$, содержащемся в $[0,\pi]$, колебание функции p было меньше

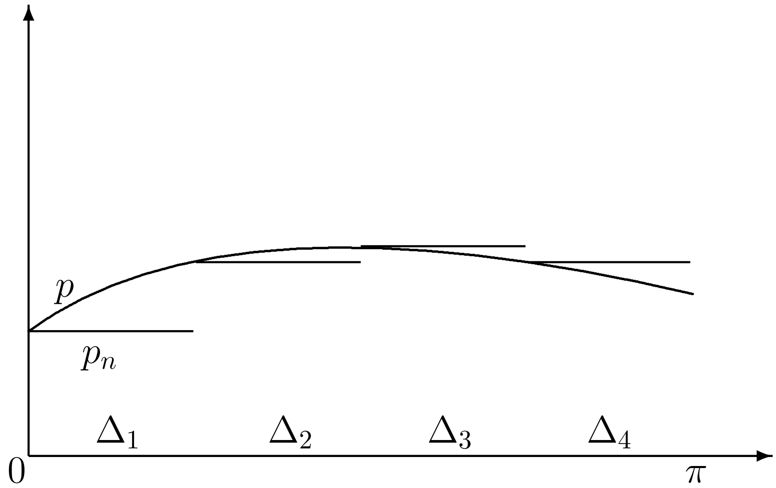
ε (пользуясь равномерной непрерывностью). Имеем

$$\mathsf{P}(X \in A) = \int_A p(x)dx = \mathsf{P}(\cos X \in \cos A) = \mathsf{P}(\cos 2^n X \in \cos A) =$$

$$= \mathsf{P}(2^n X \in A + 2\pi k,\ k \in \mathbb{Z}) + \mathsf{P}(2^n X \in -A + 2\pi k,\ k \in \mathbb{Z}) =$$

$$= \mathsf{P}(X \in \pm 2^{-n} A + 2^{1-n}\pi k,\ k \in \mathbb{Z}).$$

Пусть $\Delta_1, \ldots, \Delta_{2^n}$ — интервалы длины $\pi/2^n$, делящие $[0, \pi]$, и пусть p_n — функция, постоянная на каждом из них и на каждом из интервалов Δ_i равная значению p в левой крайней точке Δ_i :



Так как $|p(x) - p_n(x)| \leq \varepsilon$, когда $0 < x < \pi$, то можно продолжить равенство:

$$\int_A p(x)dx = \sum_{j=1}^{2^{n-1}} \int_{2^{-n}A + 2^{1-n}\pi j} p(x)dx + \sum_{j=1}^{2^{n-1}} \int_{-2^{-n}A + 2^{1-n}\pi j} p(x)dx \leq$$

$$\leq \pi\varepsilon + \sum_{j=1}^{2^{n-1}} \int_{2^{-n}A + 2^{1-n}\pi j} p_n(x)dx + \sum_{j=1}^{2^{n-1}} \int_{-2^{-n}A + 2^{1-n}\pi j} p_n(x)dx =$$

$$= \pi\varepsilon + \frac{|A|}{\pi} \int_0^\pi p_n(x)dx \leq \pi\varepsilon + \frac{|A|}{\pi}(1 + \pi\varepsilon).$$

В силу произвольности $\varepsilon > 0$ получаем, что $\int_A p(x)dx \leq |A|/\pi$. Аналогичные оценки приводят к противоположному неравенству.

Задача 10. (3-5) Случайные величины $X_1, X_2, \ldots$ независимы и одинаково распределены, а F — некоторая функция распределения. Построить какой-нибудь критерий для проверки гипотезы о том, что величины X_i имеют функцию распределения F (против альтернативы, что их функция распределения не такая), мощность которого при любой альтернативе стремится к 1 (при $n \to \infty$).

Решение. Пусть $Z_1, Z_2, \ldots$ — независимые $N(0,1)$ случайные величины. Будем наблюдать величины $X_i + Z_i$ и проверять гипотезу о том, что их распределение есть свертка F с $N(0,1)$. У такой суммы функция распределения непрерывна. Если F_1 и F_2 — различные функции распределения, то их свертки с $N(0,1)$ также неодинаково распределены. Действительно,

$$\mathsf{E}e^{it(X_1+Z_1)} = e^{-t^2/2}\mathsf{E}e^{itX_1}.$$

Итак, мы проверяем гипотезу, что данные случайные величины имеют непрерывную функцию распределения $F * \Phi$, против альтернативы, что у них другая (но непрерывная) функция распределения. Эту гипотезу можно проверять с помощью критерия Колмогорова.