ВОСЬМАЯ "КОЛМОГОРОВСКАЯ СТУДЕНЧЕСКАЯ ОЛИМПИАДА ПО ТЕОРИИ ВЕРОЯТНОСТЕЙ"

Решения задач

Задача 1. Пусть случайная величина X имеет конечную дисперсию и не равна тождественно нулю. Доказать, что $\mathsf{P}(X=0) \le \mathsf{D}X(\mathsf{E}X^2)^{-1}$.

Решение. В силу неравенства Коши-Буняковского-Шварца

$$(\mathsf{E}X)^2 = (\mathsf{E}XI(X \neq 0))^2 \le \mathsf{E}X^2\mathsf{P}(X \neq 0) = \mathsf{E}X^2 - \mathsf{E}X^2\mathsf{P}(X=0),$$

откуда $\mathsf{E}X^2\mathsf{P}(X=0) \le \mathsf{D}X$, что и требовалось.

Задача 2. Пусть множества $A_1, \ldots, A_{2000} \subset A$ содержат по крайней мере по 6 элементов каждое, и не все они совпадают. Доказать, что существует 100 таких различных разбиений множества A на 5 попарно непересекающихся подмножеств $E_1, \ldots, E_5$, что каждое множество A_i содержаbn представителей хотя бы двух подмножеств E_i.

Решение. Каждый элемент A будем независимо от прочих красить в один из трех цветов (белый, черный, красный, синий, зеленый) с равными вероятностями. Тогда

$$\mathsf{P}(\text{ есть одноцветное } A_i) \le \sum_{i=1}^{2000} \mathsf{P}(A_i \text{ одноцветно}) = \sum_{i=1}^{2000} 5^{1-|A_i|} \le \frac{2000}{3125} < 2/3.$$

Итак, вероятность расцветки, при которой все множества A_i двух- или более- цветны, не меньше $1/3$. Элементарных исходов у нас $5^{|A|} \le 5^7$ (из-за несовпадения множеств). Поэтому элементарных исходов, устраивающих нас, по крайней мере $5^7/3$. Вероятность того, что среди множеств A_i найдется одноцветное, не превосходит $2000 \cdot 3 \cdot 3^{-7} < 1$, так что найдется расцветка, при которой все множества двух- или трехцветны. Заметим, что каждому разбиению может соответствовать не более 5! расцветок (из-за перестановки цветов). Итак, устраивающих нас разбиений по крайней мере $5^7/(3 \cdot 5!) = 3125 \cdot 5/72 > 100$.

Задача 3. Из контейнера A, в котором было 1000 зеленых и 3000 красных яблок, взяли половину яблок и перенесли в контейнер B, в котором к тому времени уже лежало 3000 зеленых и 1000 красных яблок. Затем из контейнера B извлекли одно яблоко. Найти вероятность, что оно зеленое.

Решение. Пусть G – то событие, вероятность которого надо найти, а случайная величина Z есть число перемещенных зеленых яблок (у нее гипергеометрическое распределение). По формуле полной вероятности

$$\mathsf{P}(G) = \sum_{k=0}^{1000} \frac{3000+k}{6000}\mathsf{P}(Z=k) = \frac{1}{2}\sum_{k=0}^{1000}\mathsf{P}(Z=k) + \frac{1}{6000}\sum_{k=0}^{1000} k\mathsf{P}(Z=k).$$

Первая сумма равна, очевидно, 1, а вторая равна $\mathsf{E}Z$. Чтобы его найти, представим Z в виде суммы $Z = X_1 + \ldots + X_{2000}$, где X_i есть индикатор того, что i-е переложенное яблоко зеленое. Тогда при любом i легко видеть, что $\mathsf{P}(X_i = 1) = 1/4$. Следовательно,

$$\mathsf{P}(G) = \frac{1}{2} + \frac{1}{6000} \cdot 2000 \cdot \frac{1}{4} = \frac{1}{2} + \frac{1}{12} = \frac{7}{12}.$$

Задача 4. Жители города Н. любят после работы, которая у каждого заканчивается в случайное время, уделить некоторое время рыбалке. В озере водятся караси и лещи, причем доля лещей равна p. В городе действует закон, запрещающий ловить более чем одного леща за день, а горожане исключительно законопослушны и после первой поимки леща сразу возвращаются домой. Найти долю лещей среди всей пойманной в городе рыбы.

Решение. Пусть A_i — событие, состоящее в том, что случайно выбранная рыба на своей рыбалке была i-й пойманной по счету. Тогда условная вероятность того, что это лещ, при условии A_i, равна p. Поэтому и безусловная вероятность того же события равна p.

Задача 5. Последовательность случайных величин $(X_n)_{n\in\mathbb{N}}$ сходится по вероятности к случайной величине X, причем при каждом $n \in \mathbb{N}$ величины X_n и X независимы. Верно ли, что $X = const.$ п.н.?

Решение. Да, верно. В самом деле, пусть случайный вектор $Y_n = (X_n, X)$, $n \in \mathbb{N}$. Тогда $Y_n \to (X, X)$ по вероятности, $n \to \infty$ (в силу покомпонентной сходимости). Так как компоненты исходных векторов независимы, то компоненты предельного вектора тоже независимы (потому что из сходимости по вероятности вытекает, что характеристические функции сходятся). Итак, X не зависит от X, так что $X = const$.

Если не знать ХФ (2 курс не знает), то можно так: пусть X не константа, тогда найдутся $a, b \in \mathbb{R}$ и $\delta > 0$, такие, что $a < b$, $\mathsf{P}(X \le a) > \delta$, $\mathsf{P}(X \ge b) > \delta$. Возьмем настолько большое $N \in \mathbb{N}$, чтобы $\mathsf{P}(|X_n - X| > (b-a)/3) < \delta^2/2$, когда $n > N$. Пусть точка $x \in (a; a+(b-a)/3)$ является точкой непрерывности функции распрделения сл. в. X. Тогда при $n > N$

$$\mathsf{P}(X_n \le x, X \ge b) \le \mathsf{P}(|X_n - X| > (b-a)/3) < \delta^2/2,$$

но в силу независимости

$$\liminf_{n\to\infty} \mathsf{P}(X_n \le x, X \ge b) = \liminf_{n\to\infty} \mathsf{P}(X_n \le x)\mathsf{P}(X \ge b) \ge \delta \lim_{n\to\infty} \mathsf{P}(X_n \le x) \ge \delta^2,$$

получили противоречие.

Задача 6. Пусть $X_1, X_2, \ldots$ — последовательность пуассоновских случайных величин с параметром 1. Доказать, что $\mathsf{E}\max\{X_1, X_2, \ldots, X_n\} = O(\ln n)$, $n \to \infty$.

Решение. Функция $x \mapsto e^x$ выпукла и $\mathsf{E}e^{X_1} < \infty$. По неравенству Иенсена

$$\exp\left(\mathsf{E}\max\{X_1, X_2, \ldots, X_n\}\right) \le \mathsf{E}\exp\left(\max\{X_1, X_2, \ldots, X_n\}\right) =$$

$$= \mathsf{E}\max\{e^{X_1}, \ldots, e^{X_n}\} \le \mathsf{E}(e^{X_1} + \ldots + e^{X_n}) = n\mathsf{E}e^{X_1}.$$

Второе решение:

$$\exp\left(\mathsf{E}\max\{X_1, X_2, \ldots, X_n\}\right) = \sum_{k=1}^{\infty} \mathsf{P}(\max\{X_1, X_2, \ldots, X_n\} \ge k) = [\ln n] + \sum_{k \le [\ln n]+1}^{\infty} n\mathsf{P}(X_1 \ge k),$$

далее оценка второго слагаемого (которое ограничено по n).

Задача 7. Пусть $(S_n)_{n\in\mathbb{N}}$ — простейшее случайное блуждание, т.е. $S_0 = 0$ и $S_n = \xi_1 + \ldots + \xi_n$, где $(\xi_j)_{j\in\mathbb{N}}$ независимы и принимают значения 1 и -1 с вероятностями $1/2$. Обозначим $\tau = \inf\{n \in \mathbb{N} : S_n = 0\}$ момент первого возвращения в нуль. Для каждого $a \in \mathbb{N}$ найти $\mathsf{E}N_a$, где $N_a = |\{j < \tau : S_j = a\}|$ — время, проведенное блужданием в точке a, до момента τ.

Решение. Введем процессы $S_n^{(x)} = S_n + x$ — случайные блуждания, выходящие из точки $x \in \mathbb{N}$, и случайные величины $N_{a,x}$ — времена, проведенные такими блужданиями в точке a до их первого попадания в нуль (включая момент старта, если $a = x$). Тогда

$$\mathsf{E}N_a = \mathsf{E}N_a I(S_1 = 1) + \mathsf{E}N_a I(S_1 = -1) = \mathsf{E}N_a I(S_1 = 1) = \frac{1}{2}\mathsf{E}N_{a,1},$$

потому что процесс $(S_n, n \ge 1 | S_1 = 1)$ распределен как $(S_n^{(1)}, n \ge 0)$. Далее будем вычислять $\mathsf{E}N_{a,x}$.

Возьмем натуральное $m >> \max\{a, x\}$ и положим

$$\tau_{x,m} = \inf\{t \ge 0 : S_t^{(x)} \in \{0, m\}\ \} \quad \text{и} \quad N_{a,x,m} = \left|\left\{j \le \tau_{x,m} : S_j = a\right\}\right|$$

(проведем кроме нижней границы в нуле еще и верхнюю). По теореме о монотонной сходимости имеем $\mathsf{E}N_{a,x,m} \to N_{a,x}$ при $m \to \infty$. Обозначим $u(x) = \mathsf{E}N_{a,x,m}$, $x = 0, 1, \ldots, m$.

Если $x \ne a$, и при этом $0 < x < m$, то по формуле полной вероятности

$$u(x) = \frac{1}{2}(u(x-1) + u(x+1))$$

(либо на первом шаге ушли вниз, либо вверх, а далее все идет независимо от этого выбора). Это уравнение задает арифметическую прогрессию. Итак, на промежутках $[0, a]$ и $[a, m]$ график u — это прямая. Также ясно, что $u(0) = u(m) = 0$. По той же формуле полной вероятности при $x = a$ (выход из точки a) имеем

$$u(a) = 1 + \frac{1}{2}(u(a-1) + u(a+1)). \tag{1}$$

Пусть слева от a уравнение прямой имеет вид $u(x) = \alpha x$, а справа $u(x) = \gamma - \beta x$. Тогда имеем систему трех уравнений:

$$\gamma - \beta m = 0,$$

$$\alpha a - \beta a + \gamma = 0,$$

$$\alpha a = 1 + \frac{1}{2}(\alpha(a-1) + \gamma - \beta(a+1)) \Leftrightarrow \alpha(a+1) + \beta(a+1) - \gamma = 2$$

(первое в точке m, второе в точке a, последнее – из (1)).

Нам нужно только число α, так как $\alpha = u(1)$. Имеем

$$\alpha = \frac{\begin{vmatrix} 0 & -m & 1 \\ 0 & a & -1 \\ 2 & a+1 & -1 \end{vmatrix}}{\begin{vmatrix} 0 & -m & 1 \\ a & a & -1 \\ a+1 & a+1 & -1 \end{vmatrix}} = \frac{2\begin{vmatrix} -m & 1 \\ a & -1 \end{vmatrix}}{\begin{vmatrix} 0 & -m & 1 \\ a & a & -1 \\ 1 & 1 & 0 \end{vmatrix}} = \frac{2(m-a)}{m} \to 2,\ m \to \infty.$$

Итак, $\mathsf{E}N_a = 1$.

Задача 8. Пусть $W = (W_t)_{t \ge 0}$ — винеровский процесс. а) Найти математическое ожидание времени, которое его график проведет выше прямой $y = t$; б) Записать дисперсию этого времени в виде интеграла от элементарных функций; в) Вычислить ее.

Решение. а) По теореме Фубини

$$\mathsf{E}Leb\{t>0: W_t>t\} = \mathsf{E}\int_0^\infty I(W_t>t)dt = \int_0^\infty \mathsf{E}I(W_t>t)dt =$$

$$= \int_0^\infty \mathsf{P}(W_t>t)dt = \int_0^\infty \mathsf{P}\left(\frac{W_t}{\sqrt{t}}>\sqrt{t}\right)dt = \int_0^\infty \mathsf{P}\left(Z>\sqrt{t}\right)dt,$$

где $Z \sim N(0,1)$. Интеграл можно вычислить, перейдя к двойному, но проще еще раз применить теорему Фубини:

$$\int_0^\infty \mathsf{P}\left(Z>\sqrt{t}\right)dt = \left\|\frac{x=\sqrt{t}, t=x^2}{dt=2xdx}\right\| = \int_0^\infty \mathsf{P}\left(Z>x\right)2xdx =$$

$$= \mathsf{E}\int_0^\infty I\left(Z>x\right)2xdx = \mathsf{E}\int_0^Z 2xdx = \mathsf{E}(Z^+)^2 = \frac{1}{2}\mathsf{E}Z^2 = \frac{1}{2}$$

(здесь и далее при $a>b$ считается, что $\int_a^b := 0$).

б) Матожидание уже посчитано, значит, нужно только матожидание квадрата. По теореме Фубини

$$\mathsf{E}\int_0^\infty I(W_t>t)dt\int_0^\infty I(W_s>s)ds = \int_0^\infty\int_0^\infty \mathsf{P}(W_t>t, W_s>s)dsdt =$$

$$= \int_0^\infty\int_0^\infty \mathsf{P}(W_t>t, W_s>s)dtds = 2\int_0^\infty\int_0^t \mathsf{P}(W_t-W_s+W_s>t, W_s>s)dsdt =$$

$$= 2\int_0^\infty\int_0^t \mathsf{P}(Z\sqrt{t-s}+U\sqrt{s}>t, U>\sqrt{s})dsdt =: I.$$

Здесь $Z, U \sim N(0,1)$ и независимы. В терминах плотностей это

$$\frac{1}{\pi}\int_0^\infty\int_0^t\int_{\sqrt{s}}^\infty e^{-u^2/2}\int_{(t-u\sqrt{s})/\sqrt{t-s}}^\infty e^{-z^2/2}dzdudsdt.$$

в) Чтобы вычислить интеграл из пункта б), снова используем теорему Фубини. В этом пункте мы видели, что нужно найти

$$I = 2\int_0^\infty\int_0^t \mathsf{P}(Z\sqrt{t-s}+U\sqrt{s}>t, U>\sqrt{s})dsdt.$$

Замена $x=\sqrt{s}$, $y=\sqrt{t-s}$. Обратное отображение: $s=x^2$, $t=x^2+y^2$, якобиан равен $4xy$, область значений $\{x>0, y>0\}$.

$$I = 8\int_0^\infty\int_0^\infty \mathsf{P}(U>x, x^2+y^2<Ux+Zy)xydydx = 8\mathsf{E}\int_0^\infty\int_0^\infty I(x<U, Ux+Zy>x^2+y^2)xydydx.$$

Перейдем к повторному интегралу (используя то же соглашение об отрезке отрицательной длины, что и в п. а)):

$$I = 8\int_0^U x\int_0^{\sqrt{\frac{U^2+Z^2}{4}-\left(x-\frac{U}{2}\right)^2}+\frac{Z}{2}} y\,dy\,dx = 4\mathsf{E}\int_0^U x\left(\sqrt{\frac{U^2+Z^2}{4}-\left(x-\frac{U}{2}\right)^2}+\frac{Z}{2}\right)^2dx =$$

$$= 4\mathsf{E}\int_0^U x\left(\frac{U^2+Z^2}{4}-\left(x-\frac{U}{2}\right)^2+Z\sqrt{\frac{U^2+Z^2}{4}-\left(x-\frac{U}{2}\right)^2}+\frac{Z^2}{4}\right)dx.$$

Ожидание интеграла с корнем равно нулю (потому что это нечетная функция от симметричной величины $Z\sim N(0,1)$). Поэтому можно рассматривать только остальные слагаемые:

$$I = 4\mathsf{E}\int_0^U x\left(\frac{U^2+Z^2}{4}-\left(x-\frac{U}{2}\right)^2+\frac{Z^2}{4}\right)dx =$$

$$= 4\mathsf{E}\int_0^U x\left(Ux-x^2+\frac{Z^2}{2}\right)dx = 2\mathsf{E}Z^2\int_0^U x\,dx+4\mathsf{E}U\int_0^U x^2\,dx-4\int_0^U x^3\,dx =$$

$$= \mathsf{E}(U^+)^2+4\frac{\mathsf{E}(U^+)^4}{3}-\mathsf{E}(U^+)^4=\frac{1}{2}+\frac{1}{3}\frac{\mathsf{E}U^4}{2}=1 \Rightarrow \mathsf{D}Leb\{t>0: W_t>t\}=1-\frac{1}{4}=\frac{3}{4}.$$

Задача 9. Делается одно наблюдение X нормальной случайной величины с неизвестным средним $\mu\in\mathbb{R}$ и дисперсией 1. Для оценки μ используют выражение $f(X)$, где функция f непрерывна и $\mathsf{E}f(X)^2<\infty$ при любом μ. Доказать, что минимум выражения $\sup_{\mu\in\mathbb{R}}\mathsf{E}(f(X)-\mu)^2$ по всем таким f достигается при $f(x)=x$.

Решение. Заметим, что при $f(x)=x$ имеем $\sup_{\mu\in\mathbb{R}}\mathsf{E}(f(X)-\mu)^2=1$, так что надо доказать отсутствие функции, для которой эта верхняя грань меньше 1. Идея в том, чтобы μ сделать случайной. Пусть θ — квадратично-интегрируемая случайная величина с положительной плотностью q, а $Z=Y+\theta$, где $Y\sim N(0,1)$ не зависит от θ. Прежде всего заметим, что

$$\sup_{\mu\in\mathbb{R}}\mathsf{E}(f(X)-\mu)^2=\sup_{\mu\in\mathbb{R}}\mathsf{E}(f(Y+\mu)-\mu)^2=\sup_{\mu\in\mathbb{R}}\int_{\mathbb{R}}(f(y+\mu)-\mu)^2\frac{e^{-y^2/2}}{\sqrt{2\pi}}dy\geq$$

$$\geq\int_{\mathbb{R}}\int_{\mathbb{R}}\mathsf{E}(f(y+t)-t)^2\frac{e^{-y^2/2}}{\sqrt{2\pi}}q(t)dydt=\mathsf{E}(f(Z)-\theta)^2.$$

Сделаем замену переменной $x=y+t$ (переход к плотности вектора (Z,θ)):

$$\mathsf{E}(f(Z)-\theta)^2=\int_{\mathbb{R}}\int_{\mathbb{R}}\mathsf{E}(f(x)-t)^2\frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}q(t)dtdx.$$

Здесь написан квадрат расстояния в $L^2(\Omega,\mathcal{F},\mathsf{P})$-норме от случайной величины θ до $f(Z)$. Величины, измеримые относительно Z, образуют замкнутое линейное пространство в L^2 и представляются в виде $g(Z)$ (в частности, для непрерывных g). Следовательно, минимум этих расстояний достигается на проекции θ на $L^2(\Omega,\sigma(Z),\mathsf{P})$.

Пусть $\theta\sim N(0,\rho)$. Указанная проекция — условное математическое ожидание $\mathsf{E}(\theta|Z)$. Тогда в силу свойств гауссовских векторов

$$\mathsf{E}(\theta|Z)=\mathsf{E}(\theta-aZ+aZ|Z)=aZ,$$

если a выбрано из условия

$$cov(\theta-aZ,Z)=cov(\theta,Z)-a\mathsf{D}Z=cov(\theta,Y+\theta)-a\mathsf{D}(Y+\theta)=\rho-a(\rho+1)=0\Rightarrow a=\frac{\rho}{\rho+1}.$$

Итак, минимум квадратического отклонения (при данном распределении θ) достигается на функции $f_\rho(x) = \rho x/(\rho+1)$. При этом

$$\mathsf{E}(f_\rho(Z)-\theta)^2 = \frac{\rho^2}{(\rho+1)^2}\mathsf{E}Z^2 + \mathsf{E}\theta^2 - 2\frac{\rho}{\rho+1}\mathsf{E}Z\theta = \frac{\rho^2}{\rho+1} + \rho - 2\frac{\rho^2}{\rho+1} = \frac{\rho^2+\rho-\rho^2}{\rho+1} \to 1$$

при $\rho \to \infty$. Итак, $\inf_f \mathsf{E}(f(Z)-\theta)^2 \geq 1$.

Задача 10. Пусть $p > 0$, а $X_1, X_2, \ldots$ — такие случайные величины, что при каждом $\varepsilon > 0$ сходится ряд $\sum_{n=1}^\infty n^{-1}\mathsf{P}(\max_{k=1,\ldots,n}|X_k| > \varepsilon n^{1/p})$. Доказать, что $X_n/n^{1/p} \to 0$ п.н., когда $n \to \infty$.

Решение.Зафиксируем $\varepsilon > 0$. Имеем последовательность оценок

$$\sum_{k=1}^\infty \mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{(2^k+1)^{1/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{(k+1)/p}}\right\} > \varepsilon\right) \leq \sum_{k=1}^\infty \mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{2^{k/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{k/p}}\right\} > \varepsilon\right) =$$

$$= \sum_{k=1}^\infty \mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{2^{(k+1)/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{(k+1)/p}}\right\} > 2^{-1/p}\varepsilon\right) =$$

$$= \sum_{k=1}^\infty \sum_{n=2^k+1}^{2^{k+1}} 2^{-k}\mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{2^{(k+1)/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{(k+1)/p}}\right\} > 2^{-1/p}\varepsilon\right) \leq$$

$$\leq \sum_{k=1}^\infty \sum_{n=2^k+1}^{2^{k+1}} 2^{-k}\mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{(2^k+1)^{1/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{(k+1)/p}}\right\} > 2^{-1/p}\varepsilon\right) \leq$$

$$\leq 2\sum_{k=1}^\infty \sum_{n=2^k+1}^{2^{k+1}} n^{-1}\mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{(2^k+1)^{1/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{(k+1)/p}}\right\} > 2^{-1/p}\varepsilon\right) =$$

$$= 2\sum_{n=1}^\infty n^{-1}\mathsf{P}\left(\max\left\{\frac{X_{2^k+1}}{(2^k+1)^{1/p}}, \ldots, \frac{X_{2^{k+1}}}{2^{(k+1)/p}}\right\} > 2^{-1/p}\varepsilon\right) \leq 2\sum_{n=1}^\infty n^{-1}\mathsf{P}\left(\max_{j=1,\ldots,n}\frac{X_n}{n^{1/p}} > 2^{-1/p}\varepsilon\right) < \infty.$$

Следовательно, первый ряд сходится и по лемме Бореля-Кантелли $\limsup_{n\to\infty} X_n/n^{1/p} \leq \varepsilon$. Остается рассмотреть последовательность $(-X_n, n \in \mathbb{N})$.