Тема 8. МЕТОД СТАТИСТИЧЕСКИХ ИСПЫТАНИЙ

. Идея его состоит в следующем: пусть нужно определить некоторое неизвестное $y=\ ?$ (может быть, это неизвестная константа, или вектор, или функция).

$$y=\ ? \qquad \xi : M\xi = y.$$

Сконструируем случайный объект ξ так, чтобы математическое ожидание $\mathbf{M}\xi$ было равно y. Затем сгенерируем ξ многократно, N раз:

$$\xi_{\mathbf{1}},\ \xi_{\mathbf{2}},\ \ldots,\ \xi_{\boldsymbol{N}}$$

и вычислим среднее значение. В силу закона больших чисел, с ростом N среднее значение сходится к $\mathbf{M}\xi = y$:

$$\hat{y} = \hat{m}_1 = \frac{1}{N}\sum_{i=1}^{N} \xi_i \xrightarrow[N\to\infty]{} M\xi = m_1 = y. \quad \textbf{(1)}$$

Количество испытаний N необходимо определить так, чтобы ошибка определения (случайная величина) была меньше заданного значения $\delta_{\mathbf{0}}$ с заданной вероятностью $\mathbf{P}_{Д}$, близкой к 1. Для размерности $\dim y = \dim m = \ = 1$ имеем:

$$\mathbf{P}\{|\hat{m}_{\mathbf{1}} - m_{\mathbf{1}}| < \delta_{\mathbf{0}}\} \ge \mathbf{P}_{Д} \approx 1.$$

Исходя из **нормальности** оценки $\hat{m}_{\mathbf{1}} = \frac{1}{N}\sum_{i=1}^{N} \xi_i$ получается условие на N:

$$N \ge \frac{\mathbf{D}\xi}{\delta_{\mathbf{0}}^2} Q^2 \equiv N_{\mathbf{0}}, \text{ где } Q = \boldsymbol{\Phi}^{-1}\left(\frac{1+\mathbf{P}_{Д}}{2}\right).$$

Однако, $\mathbf{M}\xi$ нам **неизвестно**, тем более **неизвестна дисперсия** $\mathbf{D}\xi$. Можно использовать **оценку сверху** для дисперсии. Это лишь увеличит число испытаний.

1-й вариант. Если $\mathbf{D}_{\mathbf{max}}$ — оценка сверху для $\mathbf{D}\xi$: $\mathbf{D}_{\mathbf{max}} \ge \mathbf{D}\xi$, то

$$N = \frac{Q^2}{\delta_0^2}\mathrm{D}_{\max} \ge \frac{Q^2}{\delta_0^2}\mathrm{D}\xi = N_0. \quad \textbf{(2)}$$

2-й вариант. По результатам первых 100–200 наблюдений определим **верхнюю доверительную границу** $\mathbf{D}_{\mathbf{верх}}$ для дисперсии, и тогда

$$N = \frac{Q^2}{\delta_{\mathbf{0}}^2}\mathbf{D}_{\mathbf{верх}} \ge N_{\mathbf{0}}.$$

При решении задач на **метод Монте-Карло** нужно указать
1)вычислительные формулы,
2) способ генерации нужных случайных величин и
3)число испытаний.

Вычисление интеграла: $\qquad J = \int_a^b f(x)dx$

Рассматриваем с.в. $\xi \sim R[a,b]$ и с.в. $f(\xi)$ Посмотрим ее М.О.:

$$Mf(\xi) = \int_a^b f(x)\frac{1}{b-a}dx = \frac{1}{b-a}J$$

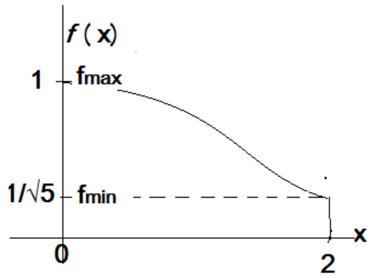
$$J = b-a \;\; Mf(\xi) \simeq b-a \;\; \frac{1}{N}\sum_{i=1}^{N} f(\xi_i) \to b-a \;\; Mf(\xi) = \mathrm{J}$$

Пример 1. Построить вычислительный процесс Монте-Карло определения интеграла

$$J = \int_0^2 \frac{dx}{\sqrt{x^2+1}}.$$

При этом погрешность вычисления с вероятностью 0,997 не должна превышать величину $\delta_0 = 10^{-3}$.

Исходим из того, что СВ $\xi \sim R[a,b]$. Поэтому



$$J = \int_a^b f(x)dx = (b-a)\int_a^b \frac{f(x)}{b-a}dx = (b-a)\cdot \mathrm{M}f(\xi).$$

1) Вычислительная формула.

Если сгенерированы $\xi_1, \xi_2, \ldots, \xi_N \sim R[0;2]$, то

$$\hat{J} = 2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sqrt{\xi_i^2+1}}.$$

2) Генерация СВ $\xi \sim R[0;2]$.

Используя стандартный генератор $\varepsilon_i \sim R[0;1]$, получим то, что нужно:

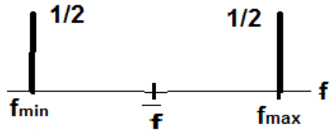
$$\xi_i = 2\varepsilon_i .$$

3) Число испытаний.

$$N_0 = \frac{Q^2}{\delta_0^2}\mathrm{D}\;(b-a)f(\xi)\; \le (b-a)^2\frac{Q^2}{\delta_0^2}\mathrm{D}_{\max} = N, \qquad \textbf{(3)}$$

где $\mathbf{D}_{\max} \ge \mathbf{D}f(\xi)$.

Оценим сверху дисперсию СВ $f(\xi)$, учитывая, что её значения находятся между значениями $f_{\min}$ и $f_{\max}$:



$$\mathrm{D}f(\xi) \le \mathrm{D}_{\max} = \left[\frac{f_{\max}-f_{\min}}{2}\right]^2 \equiv \frac{\Delta f}{2}^{\;2} \qquad \textbf{(4)}$$

√√Приведём доказательство этого неравенства. Вспомогательный факт:

$$\text{с.в. } \alpha, \text{ прогноз } c = ?, \; M(\alpha-c)^2 \to \min_c$$

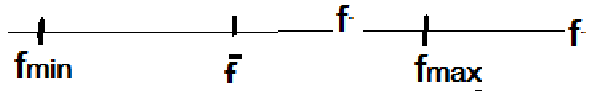
$$\textit{Дифф}: \; -2M(\alpha-c) = 0 \Rightarrow c = M\alpha$$

Имеем (с учетом обозначения $\bar{f} = f_{\max} + f_{\min} \;/2)$:

$$\mathrm{D}f(\xi) = \int\limits_{-\infty}^{+\infty} \left(f(x) - \mathrm{M}f(\xi)\right)^2 p_\xi(x)dx \le \int\limits_{-\infty}^{+\infty} \left[f(x) - \bar{f}\right]^2 p_\xi(x)dx \le$$

$$\le \int\limits_{-\infty}^{+\infty} \left(\frac{\Delta f}{2}\right)^2 p_\xi(x)dx = \left(\frac{\Delta f}{2}\right)^2 = \mathbf{D}_{\mathbf{max}},$$

поскольку $f_{\min} \le f \le f_{\max}$, $|\, f - \bar{f}\,| \le \frac{\Delta f}{2}$.

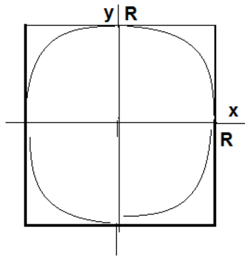


Вычисляем $\mathbf{D}_{\mathbf{max}}$, и затем N по формуле (3):

$$\mathbf{D}_{\mathbf{max}} = \left[\frac{1 - 1/\sqrt{5}}{2}\right]^2 \approx \frac{1}{16}, \qquad Q = 3, \qquad N = \frac{2^2 \cdot 9}{(10^{-3})^2} \cdot \frac{1}{16} \approx 2 \cdot 10^6.$$

Пример 2. Построить процесс вычисления площади $S(D_R)$ области D_R, ограниченной линией

$$x^4 + y^4 = R^4, \ R = 3.$$



Погрешность не должна превышать $\delta_0 = 0{,}01$ с вероятностью 0,997.

1) Вычислительная формула

Сначала упростим задачу, чтобы было меньше арифметических операций. Если сожмем в R раз по одной оси и по другой (т.е. разделим уравнение на R^4), то получим $x^4 + y^4 = 1$. А потом рассмотрим только область D_0 в 1-м квадранте

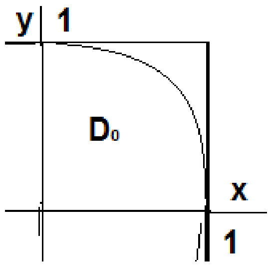
Заметим, что справедливо равенство

$$S(D_R) = R^2 \cdot S(D_1) = R^2 \cdot 4 \cdot S(D_0),$$

где D_0 — область, ограниченная линией

$$x^4 + y^4 = 1$$

в первом квадранте. Как ее вычислить?



Бросаем в квадрат $K = [0;1] \times [0;1]$ случайную точку (ξ, η). Ясно, что

$$p_0 = \mathrm{P}\left((\xi, \eta) \in D_0\right) = S(D_0).$$

Получаем:

$$S(D_R) = 4R^2 \cdot \mathrm{P}\left((\xi, \eta) \in D_0\right) = 4R^2 p_0.$$

Генерируем N раз случайную точку (ξ, η) в квадрате K: (ξ_i, η_i), $i = \overline{1, N}$, $\xi_i \sim R[0;1]$, $\eta_i \sim R[0;1]$ и подсчитываем относительную частоту попадания этой точки в область D_0:

$$\hat{p}_0 = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i \to \mathbf{M}\varepsilon_i = p_0, \text{ где } \varepsilon_i = \begin{cases} 1, \text{ если } \xi_i^4 + \eta_i^4 \le 1, \\ 0, \text{ иначе;} \end{cases} \qquad \mathbf{(5)}$$

Получаем оценку искомой площади:

$$\hat{S}(D_R) = 4R^2 \hat{p}_0 = 4R^2 \sum_{i=1}^{N} \varepsilon_i \Big/ N \to 4R^2 p_0 = S(D_R).$$

2)Генерация: $\xi_i \sim R[0;1]$, $\eta_i \sim R[0;1]$, i=1,2,…N

3) Число N испытаний:

$$N_0 = \frac{Q^2}{\delta_0^2} D\ 4R^2\varepsilon_i\ = \frac{Q^2}{\delta_0^2}16R^4 D\ \varepsilon_i\ \le \frac{Q^2}{\delta_0^2}4R^4 = N,$$

где учтено, что дисперсия $\mathbf{D}\ \varepsilon_i\ = p_0(1-p_0) \le 1/4$. Получаем

$$N = \frac{Q^2}{\delta_0^2}4R^4 = \frac{3^2}{(0{,}01)^2}4\cdot 3^4 = 9\cdot 100^2\cdot 324 \approx 290\cdot 10^3,$$

т.е. внушительное число триста тысяч испытаний. На каждое испытание в соответствии с (**5**) требуется 6 арифметических операций.

Замечания.

1. Вопрос: зачем задача была сведена к вычислению области D_0? Можно было кидать случайную точку в квадрат $[-1;1]\times[-1;1]$. Для этого нужно генерировать случайную точку α_i, β_i с координатами:

$$\alpha_i = 2\xi_i - 1,\ \beta_i = 2\eta_i - 1,\ \xi_i \sim R[0;1],\ \eta_i \sim R[0;1],$$

на что требуется еще 4 операции (два сложения и два вычитания). В результате время выполнения увеличивается более, чем в полтора раза.

2. Можно было бы уменьшить количество испытаний, вычислив точно существенную часть площади D_0, а именно вложенную в D_0 четверть круга, площадь которой $\pi/4$, а методом статистических испытаний вычислять площадь области ΔD_0 между кривыми $x^4 + y^4 = 1$ и $x^2 + y^2 = 1$.

В этом случае

$$S(D_0) = \pi/4 + S(\Delta D_0) = \pi/4 + \mathbf{P}\ \ \xi,\eta\ \in \Delta D_0\ = \pi/4 + p_\Delta,$$

и вероятность $\mathbf{P}\ \ \xi,\eta\ \in \Delta D_0\ = p_\Delta$ оценивается следующим образом:

$$\hat{p}_\Delta = \frac{1}{N_\Delta}\sum_{i=1}^{N_\Delta}\varepsilon_i \to p_\Delta,\ \text{где}\ \varepsilon_i = \begin{cases} 1, \text{если } \xi_i^4 + \eta_i^4 \le 1 \text{ и } \xi_i^2 + \eta_i^2 > 1, \\ 0, \text{иначе.} \end{cases} \qquad \textbf{(6)}$$

Пример 3. Определить функцию распределения $F_\eta(y)$ случайной величины

$$\eta = f(\xi_1, \xi_2, \xi_3) = \xi_1\cdot e^{-\xi_2} + \xi_3^2,$$

где $\xi_1 \sim N(a=1, \sigma^2=4)$, $\xi_2 \sim E(3)$ (показательное распределение) и $\xi_3 \sim R[1,3]$. Описать вычислительный процесс определения методом статистических испытаний. Погрешность при любом значении y не должна превышать $\delta_0 = 0{,}01$ с вероятностью $P = 0{,}999$.

$$F_\eta(y) = P\ \ \eta = \xi_1 + e^{-\xi_2} + \xi_3^2 < y\ = \iiint\limits_{(x_1,x_2,x_3):\ x_1\cdot e^{-x_2}+x_3^2<y} p_\xi(x_1,x_2,x_3)dx_1dx_2dx_3,$$

Процесс вычисления методом Монте-Карло прост.

1) N раз сгенерируем тройку $(\xi_1, \xi_2, \xi_3)_1$, $(\xi_1, \xi_2, \xi_3)_2$, …, $(\xi_1, \xi_2, \xi_3)_N$.

2) N раз вычислим значение СВ η: $\eta_1, \eta_2, ..., \eta_N$, и составим вариационный ряд:

$$\eta_{(1)} \le \eta_{(2)} \le \ldots \le \eta_{(N)}.$$

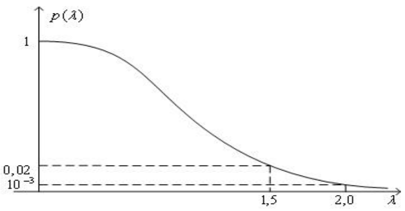
3) По вариационному ряду вычислим функцию эмпирического распределения $F^*_{\eta N}(y)$:

$$F^*_{\eta N}(y) \xrightarrow[N\to\infty]{} F_\eta(y).$$

4) Найдем число испытаний N, при котором

$$\mathbf{P}\left\{\sup_{y} |F^*_{\eta N}(y) - F_\eta(y)| \le \delta_0\right\} \equiv \mathbf{P}\{D_N \le \delta_0\} = 0{,}999, \qquad \textbf{(7)}$$

где D_N — статистика Колмогорова, для которой известна асимптотика:



$$\mathrm{P}\{D_N\sqrt{N} > \lambda\} \xrightarrow[N\to\infty]{} \mathrm{P}(\lambda). \qquad \textbf{(8)}$$

Здесь $\mathbf{P}(\lambda)$ — табулированная функция Колмогорова. Из **(7)** переходим к противоположному событию:

$$10^{-3} = \mathrm{P}\{\sqrt{N}D_N > \sqrt{N}\delta_0 = \lambda\} \approx \mathrm{P}(\lambda), \quad \mathrm{N}>20$$

Значению 10^{-3} соответствует $\lambda = 2{,}0$, откуда

$$\sqrt{N}\delta_0 = \lambda = 2{,}0 \quad \text{и}$$

$$N = \left(\frac{2}{\delta_0}\right)^2 = \left(\frac{2}{0{,}01}\right)^2 = 40 \cdot 10^3.$$

ДОМАШНЕЕ ЗАДАНИЕ

8.1. Построить процедуру вычисления определённого интеграла

$$\boldsymbol{J} = \int_{-1}^{1} e^{-x^2-2x} \cos x \, dx$$

методом Монте-Карло. Определить число испытаний таким образом, чтобы погрешность не превосходила $\varepsilon = 10^{-3}$ с вероятностью $I = 0{,}999$.

Ответ: $\mathrm{J} = \sqrt{\pi} e \int_{-1}^{1} \frac{e^{-\frac{(x+1)^2}{2\cdot(1/2)}}}{\sqrt{2\pi\cdot\frac{1}{2}}} \cdot \cos x dx = e\sqrt{\pi}\int_{-1}^{1} p_\xi(x; m, \sigma) \cos x dx$,

$$p_\xi(x; m, \sigma) \sim N\left(m = -1, \sigma = \frac{1}{\sqrt{2}}\right)$$

$$\hat{J} = e\sqrt{\pi}\,\frac{1}{N}\sum_{i=1}^{N} f(\xi_i), \quad f(\xi_i) = \begin{cases} \cos \xi_i, \text{ если } |\xi_i| \le 1, \\ 0, \quad \text{если } |\xi_i| > 1; \end{cases} \quad \xi_i \sim N\left(m = -1, \sigma = \frac{1}{\sqrt{2}}\right).$$

$$\mathrm{D}f(\xi) < \left(\frac{\Delta f}{2}\right)^2 = \left(\frac{1-\cos 1}{2}\right)^2 = \left(\sin^2(1/2)\right)^2 < \frac{1}{16};$$

$$N_0 = \frac{Q^2}{\delta_0^2}\mathrm{D}\left[e\sqrt{\pi}\, f(\xi)\right] \le e^2\pi\frac{Q^2}{\delta_0^2}\mathrm{D}f(\xi) < \frac{e^2\pi}{16}\frac{Q^2}{\delta_0^2} = N.$$

8.2. Построить по методу Монте-Карло вычислительный процесс определения функции распределения СВ

$$\eta = \xi_1^2 + e^{\xi_2^2} + \sqrt{\xi_3}, \text{ где } \xi_1 \sim R[0;2],\ \xi_2 \sim N(0;4),\ \xi_3 \sim E(a=2).$$

Точность определения для любого значения аргумента должна быть не хуже $\delta_0 = 10^{-2}$ с вероятностью $P_g = 0,99$. Определить число испытаний.

Ответ: см. пример 3.

8.3. Смоделировать $n = 3$ реализаций случайной величины ξ, распределенной по закону

$$\mathbf{P}\{\xi = k\} = \frac{1}{2}C_m^k p^k (1-p)^{m-k} + \frac{1}{2}C_m^k q^k (1-q)^{m-k},$$

где $m = 4$, $p = 0,2$, $q = 0,7$.

Указание: это СВ, которая, в зависимости от случая, с вероятностью 0,5 распределена по закону $Bi(m, p)$, а с вероятностью 0,5 распределена по закону $Bi(m, q)$.