ГЛАВА VI

ПРИЛОЖЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

§ 22. Регрессионный анализ

22.1. Модели регрессии. Пусть Y и X — зависимые СВ и требуется по результатам наблюдений $(x_1, \ldots, x_n)$ за вектором X и $(y_1, \ldots, y_n)$ за вектором Y сделать обоснованное заключение о характере зависимости Y от X.

Определение 22.1. Зависимость $Y = \varphi^*(X) \triangleq \mathbf{M}[Y|X]$, обеспечивающая наилучшее (в среднем квадратическом смысле) приближение СВ Y:

$$\mathbf{M}\left[(Y - \varphi^*(X))^2\right] \leqslant \mathbf{M}\left[(Y - \varphi(X))^2\right],$$

где $\varphi(X)$ — произвольная функция, называется *функцией теоретической регрессии*. При этом X называют *регрессором*, а Y — *откликом*.

Для вычисления условного МО необходимо знать совместную функцию распределения СВ X и Y, а такой информацией исследователь, как правило, не располагает. Для преодоления этого затруднения можно, например, ограничиться рассмотрением некоторого параметрического класса функций $\varphi(x) = \varphi(x, \theta)$, где $\theta \in \Theta \subset \subset \mathbb{R}^s$ — некоторый неизвестный параметр, который требуется оценить по имеющейся выборке наблюдений $(x_1, \ldots, x_n)$ и $(y_1, \ldots, y_n)$.

Определение 22.2. Класс линейных по θ функций

$$\varphi(x, \theta) = \theta_1 \varphi_1(x) + \ldots + \theta_s \varphi_s(x),$$

где $\varphi_1(x), \ldots, \varphi_s(x)$ — известные функции, определяющие *план эксперимента*, называется *линейной регрессионной моделью*. При этом должно выполняться очевидное требование $s < n$.

Часто рассматривают модель наблюдений за СВ Y, в которой вектор X считается неслучайным и присутствуют случайные ошибки наблюдений W:

$$Y_k = \theta_1 \varphi_1(x_k) + \ldots + \theta_s \varphi_s(x_k) + W_k, \quad k = \overline{1, n}.$$

В этой модели неизвестными по-прежнему являются параметры $\theta_1, \ldots, \theta_s$, а векторы $x_1, \ldots, x_n$ и распределения СВ W_k, $k = \overline{1, n}$, считаются известными.

22.2. Схема Гаусса–Маркова. Линейную модель наблюдений, приведенную в предыдущем разделе, можно записать в более компактном виде

$$Y = A\theta + W,$$

если ввести обозначения $Y \triangleq \mathrm{col}(Y_1, \ldots, Y_n)$ для вектора наблюдений, $\theta \triangleq \mathrm{col}(\theta_1, \ldots, \theta_s)$ — для вектора неизвестных параметров, $W \triangleq \mathrm{col}(W_1, \ldots, W_n)$ — для вектора случайных ошибок наблюдения, $A \triangleq \{a_{kj}\}$ — для матрицы, составленной из элементов $a_{kj} = \varphi_j(x_k)$, $j = \overline{1, s}$, $k = \overline{1, n}$. При этом часто матрицу A называют *матрицей плана* или *регрессионной матрицей*. Без ограничения общности изложения предположим, что $\mathbf{M}[W] = 0$ и $\mathbf{M}\left[WW^T\right] = \sigma^2 I$, где I — единичная матрица размерности $n \times n$.

Определение 22.3. Оценкой вектора неизвестных параметров θ, найденной по *методу наименьших квадратов* (или МНК-*оценкой*), называют вектор

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} (Y - A\theta)^T (Y - A\theta).$$

Сформулированную задачу называют *схемой Гаусса–Маркова.*

Теорема 22.1. (*Теорема Гаусса–Маркова*). *Пусть матрица $A^T A$ невырожденная. Тогда МНК-оценка $\widehat{\theta}$ обладает следующими свойствами*:

1) *МНК-оценка $\widehat{\theta}$ существует, она единственная и определяется формулой*

$$\widehat{\theta} = \left(A^T A\right)^{-1} A^T Y;$$

2) *МНК-оценка $\widehat{\theta}$ является несмещенной и ее компоненты имеют минимальные дисперсии в классе линейных относительно Y несмещенных оценок*;

3) *ковариационная матрица МНК-оценки $\widehat{\theta}$ определяется формулой*

$$K_{\widehat{\theta}} = \sigma^2 \left(A^T A\right)^{-1}.$$

Если в схеме Гаусса–Маркова СВ W имеет нормальное распределение, то

1) МНК–оценка $\widehat{\theta}$ имеет s-мерное нормальное распределение $\mathbf{N}\left(\theta; \sigma^2\left(A^TA\right)^{-1}\right)$;

2) Квадратичная форма $(\widehat{\theta}-\theta)^TA^TA(\widehat{\theta}-\theta)\Big/\sigma^2$ имеет распределение хи-квадрат $\chi^2(s)$ с s степенями свободы.

3) МНК-оценка $\widehat{\theta}$ является также МП-оценкой.

Пример 22.1. Пусть требуется оценить зависимость

$$y = \theta_1 + \theta_2 x + \ldots + \theta_m x^{m-1}$$

по результатам наблюдений

$$Y_k = \theta_1 + \theta_2 x_k + \ldots + \theta_s x_k^{s-1} + W_k, \quad k = \overline{1,n}$$

в точках x_k, где W_k, $k=\overline{1,n}$, — случайные ошибки измерений. В данном случае элементы матрицы плана равны $a_{kj} = x_k^{j-1}$, $k=\overline{1,n}$, $j=\overline{1,s}$. Тогда искомую зависимость можно определить в виде

$$\widehat{y} = \widehat{\theta}_1 + \widehat{\theta}_2 x + \ldots + \widehat{\theta}_s x^{s-1},$$

где оценки $\widehat{\theta}_j$ параметров θ_j, $j=\overline{1,s}$, находятся по методу наименьших квадратов из условия минимума квадратической функции

$$Q(\theta_1,\ldots,\theta_s) = \sum_{k=1}^{n}\Big(y_k - \sum_{j=1}^{s}\theta_j x_k^{j-1}\Big)^2.$$

Приравнивая нулю частные производные по θ_j квадратической функции $Q(\theta_1,\ldots,\theta_s)$, можно получить систему линейных уравнений для определения оценок неизвестных параметров $\theta_1,\ldots,\theta_s$:

$$\frac{\partial Q(\theta_1,\ldots,\theta_s)}{\partial\theta_1} = \ldots = \frac{\partial Q(\theta_1,\ldots,\theta_s)}{\partial\theta_s} = 0.$$

В частном случае, когда $s=3$ приходим к задаче построения *параболической регрессии.* В данном случае после несложных преобразований можно получить следующую систему линейных уравнений, определяющих оценки $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\widehat{\theta}_3$:

$$\begin{cases} \alpha_2\widehat{\theta}_3 + \alpha_1\widehat{\theta}_2 + \alpha_0\widehat{\theta}_1 & = & \beta_0, \\ \alpha_3\widehat{\theta}_3 + \alpha_2\widehat{\theta}_2 + \alpha_1\widehat{\theta}_1 & = & \beta_1, \\ \alpha_4\widehat{\theta}_3 + \alpha_3\widehat{\theta}_2 + \alpha_2\widehat{\theta}_1 & = & \beta_2, \end{cases}$$

где коэффициенты этих уравнений расчитываются по формулам

$$\alpha_j = \frac{1}{n}\sum_{k=1}^{n} y_k^j, \quad j=\overline{0,4}, \qquad \beta_j = \frac{1}{n}\sum_{k=1}^{n} y_k x_k^j, \quad j=\overline{0,2}.$$

Решая полученную систему линейных уравнений относительно оценок $\widehat{\theta}_1$, $\widehat{\theta}_2$, $\widehat{\theta}_3$ и подставляя их значения в приведенную выше формулу, получаем искомую зависимость $\widehat{y}(x)$.

22.3. Простая линейная регрессия.

Пример 22.2. Рассмотрим модель *простой линейной регрессии.* Пусть СВ Y и W связаны уравнением $Y = ax + b + W$, где СВ W может быть интерпретирована, например, как погрешность вычисления детерминированной величины $y = ax + b$. Пусть СВ W имеет нормальное распределение $\mathbf{N}(0;\sigma^2)$. Предположим, что в соотношении $y = ax + b$ параметры a и b неизвестны, а при каждом k-м наблюдении случайной величины Y детерминированная величина x принимает различные значения x_k, $k=\overline{1,n}$. Тогда СВ $Y_k = ax_k + b + W_k$, $k=\overline{1,n}$, будет иметь нормальное распределение $\mathbf{N}(ax_k + b;\sigma^2)$, так как $\mathbf{M}[Y_k] = ax_k + b$. Будем считать, что наблюдения независимы друг от друга, т. е. СВ W_k являются независимыми для разных k с одним и тем же распределением $\mathbf{N}(0;\sigma^2)$. В данном случае неоднородную выборку Z_n образуют наблюдения $Y_1, \ldots, Y_n$. Требуется по выборке Z_n и известным значениям $x_1, \ldots, x_n$ оценить неизвестные параметры a и b. Рассмотрим для реализации выборки $z_n = \mathrm{col}(y_1, \ldots, y_n)$ логарифмическую функцию правдоподобия $\ln L(z_n, a, b)$ с параметрами $\theta_1 \triangleq a$, $\theta_2 \triangleq b$, которая в данном случае имеет вид

$$\ln L(z_n, a, b) = -\frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - ax_k - b)^2 - n\ln\left(\sigma\sqrt{2\pi}\right).$$

Необходимое условие максимума этой функции:

$$\begin{cases} \dfrac{\partial}{\partial a}\ln L(z_n, a, b) = 0, \\ \dfrac{\partial}{\partial b}\ln L(z_n, a, b) = 0. \end{cases}$$

Решая полученную систему из двух уравнений с двумя неизвестными a и b для каждой реализации выборки z_n, находим оценки параметров:

$$\widehat{a}(z_n) = \frac{n\sum\limits_{k=1}^{n} x_k y_k - \sum\limits_{k=1}^{n} x_k \sum\limits_{k=1}^{n} y_k}{n\sum\limits_{k=1}^{n} x_k^2 - \left(\sum\limits_{k=1}^{n} x_k\right)^2}, \qquad \widehat{b}(z_n) = \frac{1}{n}\sum_{k=1}^{n} y_k - \frac{\widehat{a}(z_n)}{n}\sum_{k=1}^{n} x_k.$$

Затем, заменяя в полученных выражениях реализацию выборки z_n на выборку $Z_n = \mathrm{col}(Y_1, \ldots, Y_n)$, получаем точечные оценки неизвестных параметров $\widehat{a}(Z_n)$, $\widehat{b}(Z_n)$.

Пример 22.3. Рассмотрим модель простой линейной регрессии из примера 22.2, не предполагая, что ошибки W_k имеют нормальное распределение, и, кроме того, считая, что коэффициенты X_k случайны и наблюдаемы: $Y_k = aX_k + b + W_k$, $k = \overline{1,n}$. Пусть $\mathbf{M}[W_k] = 0$, $\mathbf{D}[W_k] = \sigma^2$ и неизвестна, СВ W_k, $k = \overline{1,n}$, независимы. Предположим, что СВ X_k и W_k, $k = \overline{1,n}$, независимы, причем X_k имеют одно и то же, но неизвестное распределение $F_X(x)$. По результатам наблюдений $(y_1, x_1), \ldots, (y_n, x_n)$ требуется оценить неизвестные параметры a и b в линейной регрессионной модели. Для неоднородной выборки $\overline{z}_n \triangleq \mathrm{col}(y_1, \ldots, y_n, x_1, \ldots, x_n)$ рассмотрим квадратическую функцию

$$Q(\overline{z}_n, a, b) = \frac{1}{n} \sum_{k=1}^{n} (y_k - ax_k - b)^2,$$

характеризующую среднюю по n квадратическую ошибку предсказания того, что в n наблюдениях СВ Y примет значения y_k, $k = \overline{1,n}$.

Определение 22.4. *МНК–оценками*, полученными по *методу наименьших квадратов*, неизвестных параметров a и b в линейной регрессионной модели $Y_k = aX_k + b + W_k$, $k = \overline{1,n}$, называют оценки $\widehat{a}(\overline{Z}_n)$ и $\widehat{b}(\overline{Z}_n)$, значения которых минимизируют квадратическую функцию $Q(\overline{z}_n, a, b)$, построенную по реализации выборки $\overline{z}_n$.

В данном случае видно, что функция $Q(\overline{z}_n, a, b)$ совпадает по форме с точностью до коэффициентов с логарифмической функцией правдоподобия из примера 22.2:

$$Q(\overline{z}_n, a, b) = -\frac{2\sigma^2}{n} \ln L(\overline{z}_n, a, b) - 2\sigma^2 \ln\left(\sigma\sqrt{2\pi}\right).$$

Поэтому минимум функции $Q(\overline{z}_n, a, b)$ по параметрам a и b достигается при тех же значениях $\widehat{a}$ и $\widehat{b}$, что и в методе максимального правдоподобия (минимизация функции $Q(\overline{z}_n, a, b)$ по a и b эквивалентна максимизации функции $\ln L(z_n, a, b)$) и определяется соотношениями из примера 22.2. Заменяя в этих соотношениях $\overline{z}_n$ на $\overline{Z}_n$, получаем

$$\widehat{a}(Z_n) = \widehat{r}_{XY} \sqrt{\frac{\widehat{d}_Y}{\widehat{d}_X}}, \qquad \widehat{b}(Z_n) = \widehat{m}_Y - \widehat{a}(Z_n)\widehat{m}_X.$$

Найденные по методу наименьших квадратов оценки $\widehat{a}(\overline{Z}_n)$ и $\widehat{b}(\overline{Z}_n)$ неизвестных параметров a и b построены для произвольно распределенных случайных ошибок W_k и случайных коэффициентов X_k, тогда как по методу максимального правдоподобия аналогичные оценки получены в предположении о нормальности W_k и для детерминированных значений x_k, $k = \overline{1,n}$. Иными словами, МНК–оценки оказываются более *робастными* (т. е. менее чувствительными к априорной информации о случайных коэффициентах X_k и ошибках W_k) по сравнению с ММП–оценками.

Исследуем статистические свойства найденных МНК–оценок $\widehat{a}(\overline{Z}_n)$ и $\widehat{b}(\overline{Z}_n)$. Предполагая существование моментов у СВ Y и X и переходя к пределу в соотношениях для $\widehat{a}(\overline{Z}_n)$ и $\widehat{b}(\overline{Z}_n)$ при $n \to \infty$, по закону больших чисел получаем

$$\widehat{a}(\overline{Z}_n) \xrightarrow{\text{п.н.}} a^* \triangleq \frac{\sigma_Y}{\sigma_X}\, r_{XY}, \qquad \widehat{b}(\overline{Z}_n) \xrightarrow{\text{п.н.}} b^* \triangleq m_Y - a^* m_X.$$

Пример 22.4. Пусть числовые характеристики СВ X и Y известны: $m_X, \sigma_X, m_Y, \sigma_Y, r_{XY}$, но функциональная связь между СВ X и Y неизвестна. Рассмотрим новую СВ $\widehat{Y} \triangleq aX + b$, где a и b — некоторые заданные параметры. *Задача предсказания* СВ Y заключается в том, чтобы по наблюдениям (реализациям) $x_1, \ldots, x_n$ предсказать значения наблюдаемой СВ Y на основе зависимости $\widehat{y}_k \triangleq ax_k + b$, $k = \overline{1,n}$, зная лишь числовые характеристики СВ X и Y. Ясно, что предсказываемые значения $\widehat{y}_k$ будут отличаться от реализаций y_k СВ Y, а точность предсказания определяется параметрами a и b. Зададимся целью выбрать параметры a и b так, чтобы *среднеквадратическая ошибка предсказания* была минимальна:

$$(a^*, b^*) = \arg\min_{a,b} \mathbf{M}\left[\left(Y - \widehat{Y}(X,a,b)\right)^2\right].$$

Обозначим

$$Q(a,b) \triangleq \mathbf{M}\left[\left(Y - \widehat{Y}(X,a,b)\right)^2\right] \triangleq \mathbf{M}\left[(Y - aX - b)^2\right].$$

Тогда можно установить, что

$$Q(a,b) = \sigma_Y^2(1 - r_{XY}^2) + (a\sigma_X - \sigma_Y r_{XY})^2 + (m_Y - am_X - b)^2.$$

Для того чтобы функция $Q(a,b)$ достигала минимума по a и b, достаточно выбрать

$$a^* = \frac{\sigma_Y}{\sigma_X}\, r_{XY}, \qquad b^* = m_Y - a^* m_X.$$

В этом случае

$$Q(a^*, b^*) = \sigma_Y^2(1 - r_{XY}^2).$$

Значения параметров a^* и b^* совпадают с предельными значениями оценок $\widehat{a}(\overline{Z}_n)$ и $\widehat{b}(\overline{Z}_n)$, полученными в примере 22.3.

Проанализируем минимальное значение функции

$$Q(a^*, b^*) = \mathbf{D}[Y - a^*X - b^*] = \sigma_Y^2(1 - r_{XY}^2),$$

так как $\mathbf{M}[Y - a^*X - b^*] = m_Y - a^*m_X - b^* = 0$. Таким образом, имеем

$$1 - r_{XY}^2 = \frac{\mathbf{D}\,[Y - a^*X - b^*]}{\sigma_Y^2}.$$

Отсюда следует, что коэффициент корреляции характеризует относительную (в единицах σ_Y^2) величину среднего квадратического отклонения СВ Y от линейной оценки наилучшего приближения $\widehat{Y}^* \triangleq a^*X + b^*$, т. е. r_{XY} численно характеризует степень линейной стохастической зависимости между СВ X и Y. Чем ближе $|r_{XY}|$ к единице, тем теснее будут группироваться выборочные точки (x_k, y_k) около прямой $\widehat{y} = a^*x + b^*$, называемой *прямой среднеквадратической регрессии СВ Y на СВ X*. При $r_{XY} = \pm 1$ имеем $\mathbf{D}[Y - aX - b] = 0$, поэтому, согласно свойствам МО, СВ X и Y с вероятностью 1 линейно зависимы: $Y = a^*X + b^*$.

Как отмечалось в п. 22.1, наилучшей среднеквадратической оценкой СВ Y по СВ X является условное МО $\widehat{Y}^* = \mathbf{M}[Y|X]$. В частности, если вектор $Z \triangleq \operatorname{col}(X, Y)$ — гауссовский, то из теоремы о нормальной корреляции следует, что

$$\widehat{Y}^* = \mathbf{M}[Y|X] = m_Y + \frac{\sigma_Y r_{XY}}{\sigma_X}\,(X - m_X),$$

т. е. наилучшая оценка $\widehat{Y}^*$ линейно зависит от X и совпадает с оценкой, полученной выше с помощью метода наименьших квадратов при условии, что функция $\varphi(X)$ линейна. График функции $\widehat{y}^* = \mathbf{M}[Y|x]$ называется *линией регрессии* СВ Y на СВ X. В гауссовском случае линия регрессии — прямая.

22.4. Типовые задачи.

З а д а ч а 22.1. Используя данные типовой задачи 16.1, построить линейную регрессионную модель

$$Y_k = aX_k + b + W_k, \qquad k = \overline{1, n},$$

описывающую зависимость между СВ Y (средняя температура января в г. Алатыре) и СВ X (средняя температура января в г. Саратове). Заметим, что распределение случайных ошибок W_k неизвестно.

Решение. Для оценивания неизвестных параметров a и b используем МНК. Согласно примеру 22.3, МНК–оценки имеют вид:

$$\widehat{a}(Z_n) = \widehat{r}_{XY}\sqrt{\frac{\widehat{d}_Y}{\widehat{d}_X}}, \qquad \widehat{b}(Z_n) = \widehat{m}_Y - \widehat{a}(Z_n)\widehat{m}_X.$$

В данной задаче $\widehat{r}_{XY} = 0{,}95$, $\widehat{d}_Y = 20{,}09$, $\widehat{d}_X = 22{,}14$, $\widehat{m}_X = -11{,}87$, $\widehat{m}_Y = -13{,}75$ (результат решения задачи 16.1).

Проведя необходимые вычисления, получим $\widehat{a} = 0{,}91$, $\widehat{b} = -3$.

Ответ. $\widehat{a} = 0{,}91$, $\widehat{b} = -3$.

§ 23. Метод статистических испытаний

23.1. Основные понятия.

Определение 23.1. Метод вычисления некоторой детерминированной величины a как среднего арифметического Y_n независимых одинаково распределенных СВ $X_1, \ldots, X_n$, подобранных таким образом, чтобы $Y_n \xrightarrow{\text{п.н.}} a$, называется *методом статистических испытаний* (*методом Монте–Карло*) для вычисления a.

Замечание 23.1. В некоторых учебниках метод Монте–Карло называют также *методом статистического моделирования*, что, на наш взгляд, не совсем корректно. Дело в том, что термин «*моделирование*» используется также в *теории математического моделирования* для описания процесса создания математических моделей каких–либо явлений. Подчеркнем, что в методе Монте–Карло речь идет о *статистической имитации испытаний* (опытов), а не о процессе создания ститистических моделей опытов.

23.2. Вычисление вероятности события.

Пример 23.1. Рассмотрим применение метода статистических испытаний для оценивания неизвестной вероятности $p \triangleq \mathbf{P}(A)$ некоторого события A. В соответствии с усиленным законом больших чисел частота события A сходится почти наверное к вероятности события A, т.е. $W_n(A) \triangleq M(A)/n \xrightarrow{\text{п.н.}} p$ при $n \to \infty$. Кроме того, в соответствии с теоремой Муавра–Лапласа справедливо другое предельное соотношение при $n \to \infty$:

$$\overset{*}{W}_n \triangleq \frac{M(A) - np}{\sqrt{npq}} \xrightarrow{F} U, \quad \text{где } U \sim \mathbf{N}(0;1).$$

Воспользуемся этими соотношениями для выбора *гарантирующего числа испытаний* N, при котором можно было бы сказать, что СВ $M(A)/N$ «близка» к оцениваемой вероятности p. Поступим следующим способом. Вначале зададим точность Δ оценки вероятности p: $|M/n - p| \leqslant \Delta$, где $M \triangleq M(A)$. Затем найдем надежность этой оценки, учитывая, что $\overset{*}{W}_n \approx U$,

$$\mathbf{P}\left\{\left|\frac{M}{n} - p\right| \leqslant \Delta\right\} = \mathbf{P}\left\{\frac{|M - np|}{\sqrt{npq}} \leqslant \Delta\sqrt{\frac{n}{pq}}\right\} \approx 2\Phi_0\left(\Delta\sqrt{\frac{n}{pq}}\right).$$

Задав доверительную вероятность $\delta \triangleq 1 - \alpha$, получим трансцендентное уравнение $2\Phi_0\left(\Delta\sqrt{N/pq}\right) = \delta$, из которого можно найти

$$N_\delta = \left[\frac{x^2_{(1+\delta)/2}\, p(1-p)}{\Delta^2}\right] + 1,$$

где $x_{(1+\delta)/2}$ удовлетворяет условию: $\Phi_0(x_{(1+\delta)/2}) = \delta/2$, т. е. $x_{(1+\delta)/2}$ является квантилью уровня $(1+\delta)/2$ распределения $\mathbf{N}(0;1)$. Но величины p и $q \triangleq 1 - p$ заранее неизвестны, поэтому заменим величину $p(1-p)$ на ее максимальное значение, которое, очевидно, достигается при $p = 1/2$ и равно $1/4$. Поэтому выбираем

$$N_\delta = \left[\frac{\left(x^2_{(1+\delta)/2}\right)}{(2\Delta)^2}\right] + 1.$$

В этом примере для вычисления N_δ необходимо задать уровень доверительной вероятности $\delta = 1 - \alpha$ (вторичный по отношению к p) и точность Δ вычисления p.

Замечание 23.2. При $p > 0{,}8$ обычно задают $\Delta = (1-p)/10$ и $\delta = 1 - \Delta$.

Пример 23.2. В предыдущем примере гарантирующее число испытаний было выбрано априори, т. е. до опыта, и одним и тем же для всех оцениваемых вероятностей p. Но интуитивно ясно, что N_δ должно быть различным в зависимости от значения p и от реализации числа успешных испытаний M. Например, пусть в серии из N испытаний все испытания оказались успешными, т. е. $M = N$. Найдем для этого случая зависимость $N_\delta(p)$ гарантирующего числа испытаний от доверительной вероятности $\delta = 1 - \alpha$ и оцениваемой вероятности p. С этой целью рассмотрим, как и в примере 23.1, нормированную СВ $\overset{*}{W}_N$, но потребуем выполнения другого вероятностного условия:

$$\mathbf{P}\left\{\overset{*}{W}_N \leqslant x_\delta\right\} = \delta,$$

где x_δ — квантиль уровня $\delta = 1 - \alpha$ для стандартного нормального распределения $\mathbf{N}(0;1)$. Это условие отражает наше желание добиться того, чтобы частота W_n была не меньше оцениваемой вероятности p с некоторой точностью, т. е.

$$\overset{*}{W}_N \triangleq (W_N - p)\sqrt{\frac{N}{p(1-p)}} = \left(\frac{N-K}{N} - p\right)\sqrt{\frac{N}{p(1-p)}} \leqslant x_\delta,$$

где $K \triangleq N - M$ есть случайное число неуспешных испытаний. Действительно, это неравенство эквивалентно следующему:

$$W_N \leqslant p + x_\delta\sqrt{\frac{p(1-p)}{N}}\,.$$

Если последнее неравенство выполнено для конкретно реализовавшегося значения частоты, то можно надеяться (с доверительной вероятностью δ), что полученное значение W_N лишь незначительно меньше вычисляемой вероятности p (см. более подробно о понятии доверительного интервала в пункте 19.1). Выбирая минимальное N, удовлетворяющее полученному неравенству, т. е. решая соответствующее квадратное уравнение $(N - k - Np)^2 = Nx_\delta^2 p(1-p)$, где k — реализация K, находим

$$N_\delta(p,k) = \left[\frac{2k + x_\delta^2 p + x_\delta\sqrt{4kp + p^2x_\delta^2}}{2(1-p)}\right] + 1.$$

В частности, если все испытания оказались успешными (случай $k = 0$), то

$$N_\delta(p,0) = \left[\frac{x_\delta^2 p}{1-p}\right] + 1.$$

Например, если при вычислении вероятности $p = 0{,}99$ положить, как рекомендуется в примере 23.1, $\Delta = 0{,}001$, $\delta = 1 - \alpha = 0{,}999$, то по формуле из этого примера получаем априорное значение гарантирующего числа испытаний $N_\delta = 10\ 300\ 000$, так как в этом случае $x_\delta = 3{,}209$. При тех же данных апостериорное значение гарантирующего числа непрерывно успешных испытаний оказывается намного меньше: $N_\delta(p,0) = 1030$.

Пример 23.3. Используем теперь другую предельную теорему для выбора гарантирующего числа испытаний. Предположим, что $p > 0{,}9$ и воспользуемся теоремой Пуассона. Пусть $q \triangleq 1 - p$ — вероятность неуспешных испытаний, а $W_N = K/N$ — частота неуспешных испытаний в серии из N испытаний, где K — случайное число неуспешных испытаний. В соответствии с теоремой Пуассона при больших N и малых q распределение частоты $T \triangleq W_N$ можно аппроксимировать распределением Пуассона $\mathbf{\Pi}(Nq)$, в соответствии с которым имеем

$$\mathbf{P}\left\{W_N \leqslant \frac{k}{N}\right\} \approx F_T\left(\frac{k}{N}\right) = \sum_{i=0}^{k} e^{-Nq}\frac{(Nq)^i}{i!}.$$

Воспользуемся результатом из примера 19.6, приведенного в пункте 19.3, для построения правостороннего доверительного интервала $[0, \theta_2]$, который «накрывал» бы неизвестный параметр q с доверительной вероятностью $\delta = 1 - \alpha$. Таким образом, получаем уравнение связи между параметрами N, θ_2, k, α:

$$F_T\left(\frac{k}{N}\right) = \sum_{i=0}^{k} e^{-N\theta_2}\frac{(N\theta_2)^i}{i!} = \alpha.$$

Можно показать, что функция в левой части этого уравнения монотонно убывает по k с ростом N. Поэтому корень данного уранения по N может быть легко найден, например с помощью метода дихотомии. В частном случае, когда в серии из N испытаний все испытания оказались успешными, т. е. $k = 0$, то данное уравнение принимает более простой вид $e^{-N\theta_2} = \alpha$. Отсюда находим гарантирующее число испытаний при условии, что все испытания успешны:

$$N_\delta = \left[-\frac{\ln(1-\delta)}{\theta_2}\right] + 1.$$

23.3. Вычисление определенного интеграла.

Пример 23.4. Рассмотрим определенный интеграл

$$a \triangleq \int_0^1 g(x)\,dx,$$

где $g(x)$ — ограниченная на отрезке $[0,1]$ функция ($|g(x)| \leqslant c < < \infty$ для всех $x \in [0,1]$). Рассмотрим также СВ X, равномерно распределенную на отрезке $[0,1]$, для которой $f_X(x)=1$, если $x \in \in [0,1]$, и $f_X(x)=0$, если $x \notin [0,1]$. Тогда очевидно, что

$$\mathbf{M}[g(X)] = \int\limits_{-\infty}^{+\infty} g(x) f_X(x)\,dx = \int\limits_0^1 g(x)\,dx \stackrel{\Delta}{=} a.$$

Таким образом, значение определенного интеграла можно найти как математическое ожидание СВ $\widetilde{X} \stackrel{\Delta}{=} g(X)$, где СВ X имеет равномерное распределение на отрезке $[0,1]$, $X \sim \mathbf{R}(0;1)$. Пусть теперь последовательность $\{X_n\}$, $n=1,2,\ldots$, состоит из СВ, независимых и равномерно распределенных на отрезке $[0,1]$. Можно убедиться, что в этом случае СВ $\widetilde{X}_n \stackrel{\Delta}{=} g(X_n)$, $n=1,2,\ldots$, будут также независимы, одинаково распределены с конечными МО и дисперсиями. Так как МО СВ X_n ограничены, то по усиленному закону больших чисел (теорема Колмогорова) последовательность СВ $Y_n \stackrel{\Delta}{=} \frac{1}{n}\sum\limits_{k=1}^{n} \widetilde{X}_k$ сходится почти наверное к величине $a = \mathbf{M}[g(X)]$, т.е. $Y_n \xrightarrow{\text{п.н.}} a$ при $n \to \infty$. Это значит, что, проведя большое количество испытаний, можно с высокой точностью вычислить значение интеграла a.

З а м е ч а н и е 23.3. Метод Монте–Карло имеет огромную область приложения. Наиболее сложной проблемой при его реализации является выбор необходимого числа испытаний n, такого, чтобы можно было считать, что СВ Y_n достаточно «близка» к a. Ясно, что из-за вычислительных трудностей желательно выбирать величину n с возможно меньшим гарантирующим значением. Обозначим это значение через N. Для выбора N обычно используют центральную предельную теорему, считая, что распределение нормированной суммы Z_N СВ X_k, $k=\overline{1,N}$, является стандартным нормальным распределением $\mathbf{N}(0;1)$. Рассмотрим на примерах, как выбирается величина N.

П р и м е р 23.5. Выберем гарантирующее число статистических испытаний N при вычислении значения a определенного интеграла из примера 23.4. Рассмотрим дисперсию СВ $\widetilde{X} \stackrel{\Delta}{=} g(X)$, предполагая, что $g(x) \not\equiv a$ и $g(x)$ — непрерывна,

$$\mathbf{D}[g(X)] = \int\limits_{-\infty}^{+\infty} (g(x)-a)^2 f_X(x)\,dx = \int\limits_0^1 (g(x)-a)^2\,dx \leqslant \int\limits_0^1 (|g(x)|+|a|)^2\,dx.$$

По условиям примера 23.4 имеем $|g(x)| \leqslant c$, тогда

$$|a| = \left|\int_0^1 g(x)\,dx\right| \leqslant c, \quad \mathbf{D}[g(X)] \leqslant \int_0^1 (|g(x)| + c)^2\,dx \leqslant 4c^2.$$

Учтем, что к последовательности СВ $g(X_n)$, $n = 1, 2, \ldots$, применима центральная предельная теорема, в соответствии с которой СВ

$$Z_n \triangleq \frac{1}{s_n} \sum_{k=1}^{n} [g(X_k) - m_k],$$

где $s_n^2 \triangleq \mathbf{D}\Big[\sum\limits_{k=1}^{n} g(X_k)\Big]$, $m_k \triangleq \mathbf{M}\,[g(X_k)]$, сходится по распределению к СВ $U \sim \mathbf{N}(0;1)$, т. е. $Z_n \xrightarrow{F} U$ при $n \to \infty$.

Далее, по усиленному закону больших чисел,

$$Y_n \triangleq \frac{1}{n} \sum_{k=1}^{n} g(X_k) \xrightarrow{\text{п.н.}} a \triangleq \mathbf{M}[g(X)],$$

т. е. СВ Y_n сходится почти наверное к постоянной a. Это значит, что при больших n СВ Y_n «близка» к a. Предположим, что нужно вычислить a с точностью Δ: $|Y_n - a| \leqslant \Delta$. Поскольку Y_n является случайной величиной, выясним степень достоверности этой оценки:

$$\mathbf{P}\{|Y_n - a| \leqslant \Delta\} = \mathbf{P}\left\{|Y_n - a|\sqrt{\frac{n}{\mathbf{D}[g(X)]}} \leqslant \Delta\sqrt{\frac{n}{\mathbf{D}[g(X)]}}\right\}.$$

Так как $\mathbf{D}[g(X)] \leqslant 4c^2$, то

$$\mathbf{P}\{|Y_n - a| \leqslant \Delta\} \geqslant \mathbf{P}\left\{|Y_n - a|\sqrt{\frac{n}{\mathbf{D}[g(X)]}} \leqslant \Delta\frac{\sqrt{n}}{2c}\right\} = \mathbf{P}\left\{|Z_n| \leqslant \Delta\frac{\sqrt{n}}{2c}\right\}.$$

В соответствии с ЦПТ имеем $Z_n \xrightarrow{F} U$, т. е.

$$\mathbf{P}\left\{|Z_n| \leqslant \Delta\frac{\sqrt{n}}{2c}\right\} \approx 2\Phi_0\left(\Delta\frac{\sqrt{n}}{2c}\right),$$

где $\Phi_0(x) \triangleq \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_0^x e^{-t^2/2}\,dt$. Выберем некоторую доверительную вероятность $\delta = 1 - \alpha$, которая равна требуемой надежности выполнения неравенства $|Y_n - a| \leqslant \Delta$, т. е. $\mathbf{P}\{|Z_n| \leqslant \Delta\sqrt{n}/2\} = \delta$.

З а м е ч а н и е 23.4. На практике величину $\delta = 1 - \alpha$ обычно задают в пределах $0{,}95 \leqslant \delta < 1$.

Тогда, решая уравнение $2\Phi_0\left(\Delta\sqrt{N}\Big/(2c)\right)=\delta$ относительно N, можно найти гарантирующее число испытаний N_δ. Функция $\Phi_0(x)$ задается таблично, и поэтому для доверительной вероятности δ можно найти такое значение $x_{(1+\delta)/2}$, что $\Phi_0(x_{(1+\delta)/2})=\delta/2$. Число $x_{(1+\delta)/2}$ является квантилью уровня $(1+\delta)/2$ распределения $\mathbf{N}(0;1)$. Таким образом, получаем уравнение $\Delta\sqrt{N}\Big/(2c)=x_{(1+\delta)/2}$, из которого легко найти искомое значение

$$N_\delta=\left[\left(\frac{2c\,x_{(1+\delta)/2}}{\Delta}\right)^2\right]+1,$$

где символ $[\,\cdot\,]$ означает целую часть числа. Подчеркнем, что для вычисления N_δ необходимо задать точность Δ оценки величины a и уровень доверительной вероятности δ.

В данном примере рассмотрен случай вычисления $a\triangleq\mathbf{M}[g(X)]$ для равномерно распределенной СВ $X\sim\mathbf{R}(0;1)$. Аналогично можно поступить при вычислении $a\triangleq\mathbf{M}[g(X)]$ для произвольной СВ X, такой, что $|a|<\infty$.

23.4. Типовые задачи.

Задача 23.1. Предвыборный штаб кандидата X в президенты проводит социологический опрос, чтобы оценить вероятность p того, что избиратели будут голосовать на предстоящих выборах за кандидата X. Сколько избирателей надо опросить, чтобы с доверительной вероятностью 0,95 отклонение полученной оценки от истинной вероятности p было не более 0,01?

Решение. В данной задаче определен уровень доверительной вероятности $\delta=1-\alpha=0{,}95$ и точность оценки $\Delta=0{,}01$.

Нам необходимо выбрать число избирателей N_δ, такое, чтобы выполнялось соотношение

$$\mathbf{P}\left\{\left|\frac{M}{N_\delta}-p\right|<0{,}01\right\}=0{,}95.$$

Согласно примеру 23.1, имеем

$$N_\delta=\left[\frac{x^2_{(1+\delta)/2}p(1-p)}{\Delta^2}\right]+1,$$

где $x_{(1+\delta)/2}=1{,}96$ — квантиль уровня 0,975 стандартного гауссовского распределения $\mathbf{N}(0;1)$, а величина $p(1-p)$ заменяется ее максимальным значением, равным $1/4$.

Проведя необходимые вычисления, получим, что гарантирующее число опрошенных избирателей $N_\delta=9605$.

Ответ. $N_\delta=9605$.

§ 24. Задачи для самостоятельного решения

1. Антрополог собирается изучать обитателей некоторого острова. Он намерен оценить процент населения с 1-й группой крови, и будет удовлетворен, если этот процент окажется правильным в пределах $\pm 3\%$. Сколько островитян антрополог должен выбрать для изучения их группы крови, если он считает допустимым гарантировать свои выводы с доверительной вероятностью $1-\alpha = 0{,}95$.

2. Известно, что 2% населения некоторого города больны туберкулезом. Сколько жителей этого города нужно обследовать, чтобы среди обследуемых больные туберкулезом составляли $2 \pm 0{,}2\%$ с вероятностью $1-\alpha = 0{,}99$?

3. На химическом производстве получены следующие данные о зависимости выхода продукта Y [кг/час] от температуры реакции X $[{}^\circ C]$:

Таблица 24.1

X	51	32	80	73	64	45	83	44	93
Y	52,7	15,2	89,5	94,8	76,0	39,3	114,8	36,5	137,4
X	28	35	40	29	53	58	65	75	
Y	5,3	20,7	21,7	9,2	55,4	64,3	79,1	101,1	

Предполагая, что зависимость между X и Y описывается моделью простой линейной регрессии, построить МНК–оценки неизвестных параметров.

4. Результаты измерения температуры T корпуса работающего агрегата, производимые с интервалом 5 мин, представлены в табл. 24.2.

Таблица 24.2

t [мин]	5	10	15	20	25
T $[{}^\circ C]$	59,3	59,8	60,1	64,9	70,2

Считая, что зависимость между этими переменными имеет вид $T = \theta_1 + \theta_2 t + \theta_3 t^2$, найти МНК–оценки неизвестных параметров θ_1, θ_2 и θ_3.

ОТВЕТЫ

Глава № 1

1. *а*) $\dfrac{k-1}{17}$ при $1 \leqslant k \leqslant 18$ и 1 при $k \geqslant 18$; *б*) $\dfrac{1}{136}$; *в*) $\dfrac{1}{100}$. **2.** $\dfrac{1}{720}$. **3.** $\mathbf{P}(A) = \dfrac{1}{45}$; $\mathbf{P}(B) = \dfrac{7}{45}$; $\mathbf{P}(C) = \dfrac{1}{15}$. **4.** $\mathbf{P}(A) = \dfrac{C_4^2 C_{48}^{24}}{C_{52}^{26}}$; $\mathbf{P}(B) = \dfrac{C_{48}^{26} + C_{48}^{22}}{C_{52}^{26}}$; $\mathbf{P}(C) = \dfrac{C_4^1 C_{48}^{25} + C_4^3 C_{48}^{23}}{C_{52}^{26}}$. **5.** $\dfrac{365!}{(365-r)!365^r}$. **6.** $\dfrac{1}{60}$. **7.** $\dfrac{2639}{3750} \approx 0{,}7037$. **8.** $0{,}082$. **9.** $0{,}405$. **10.** $\dfrac{3}{248} \approx 0{,}012$. **11.** *а*) $\dfrac{1}{210}$; *б*) $\dfrac{3}{7}$; *в*) $\dfrac{209}{210}$. **12.** Указание: исследовать равновероятность предложенных комбинаций. **13.** *а*) $0{,}009$; *б*) $0{,}38$. **14.** $0{,}25$. **15.** *а*) $0{,}753$; *б*) $\dfrac{2^{50}}{C_{100}^{50}} \approx 10^{-14}$. **16.** $\left(\dfrac{n-2}{n-1}\right)^{r-2}$. **17.** $0{,}0000189$. **18.** $0{,}001$. **19.** Зависимы. **20.** Верно. **21.** $\dfrac{13}{16}$. **22.** Не менее 298. **23.** $0{,}962$. **24.** $\dfrac{2}{5}$. **25.** 0. **26.** $0{,}187$. **27.** $0{,}2304$. **28.** Нет. **29.** Можно. **30.** Верно. **31.** 1. **32.** Верно. **33.** В случае несовместных A и B. **34.** $\dfrac{2}{3}$. **35.** Совместны. **36.** 1. **37.** Зависимы, несовместны. **38.** Независимы. **39.** Зависимы. **40.** $0{,}704$. **41.** $0{,}1$. **42.** $0{,}0028$. **43.** *а*) $0{,}782$; *б*) $0{,}16356$. **44.** $2p - p^2$. **45.** $\dfrac{2}{9}$. **46.** $\mathbf{P}(B|A) \approx 0{,}3876$; $\mathbf{P}(\overline{B}|A) \approx 0{,}6124$; $\mathbf{P}(B|\overline{A}) \approx$ $\approx 0{,}1022$; $\mathbf{P}(\overline{B}|\overline{A}) \approx 0{,}8978$. **47.** $\dfrac{2}{3}$ и $\dfrac{1}{3}$. **48.** $1 - \dfrac{1}{2!} + \dfrac{1}{3!} + \ldots$ $\ldots + \dfrac{(-1)^{n-1}}{n!}$. **49.** $C_{2n-k}^{n-k}\left(\dfrac{1}{2}\right)^{2n-k}$. **50.** $0{,}5$. **51.** *а*) $0{,}1468$; *б*) $0{,}9437$. **52.** $0{,}0307$. **53.** $0{,}6647$. **54.** $0{,}7748$. **55.** *а*) $0{,}3281$; *б*) $0{,}4095$. **56.** Гипотезами являются наборы событий 1 и 4. **57.** Безразлично, вероятность сдачи экзамена равна $\dfrac{1}{6}$. **58.** 0,455. **59.** 0,6405. **60.** Нет. **61.** $\dfrac{1}{5}$. **62.** Может. **63.** $\dfrac{2}{3}$. **64.** $0{,}2941$. **65.** В 3-ю фирму. **66.** Может. **67.** Может. **68.** Нет. **69.** Могут. **70.** $\mathbf{P}(H_1|H_1) = 1$, $\mathbf{P}(H_2|H_1) =$ $= \mathbf{P}(H_3|H_1) = 0$. **72.** $\dfrac{p_1}{p_1+p_2}$; $\dfrac{p_2}{p_1+p_2}$. **73.** $0{,}8596$. **74.** $0{,}8(9)$. **75.** $0{,}5375$. **76.** *а*) $0{,}019$. *б*) $0{,}3158$; **77.** $0{,}2$. **78.** $0{,}4286$. **79.** $\dfrac{1}{4}$. **80.** *а*) $0{,}5787$; *б*) $0{,}0017$. **81.** *а*) $0{,}1154$; *б*) $0{,}3293$. **82.** $0{,}3288$. **83.** *а*) $\dfrac{1}{30}$; *б*) $0{,}064$. **84.** $0{,}995$. **85.** $0{,}4$.

Глава № 2

1. Не является. **2.** Нет. **3.** Неверно. **4.** *а*) 3,5; *б*) −0,575; *в*) 1,495. **5.** Наиболее вероятное значение равно 1, наименее вероятное значение равно 4. **6.** $\mathbf{M}[X]\cdot\mathbf{D}[Y] > \mathbf{M}[Y]\cdot\mathbf{D}[X]$. **7.** 1. **8.** Для любых x. **9.** $F_X(1/2) = F_X(-1/2) = 1/2$. **10.** $\mathbf{Bi}(1; 0{,}8)$. **11.** Равны. **12.** 0. **13.** *а*) Да; *б*) Да. **14.** 0,488. **15.** 0 и 1. **16.** 1. **17.** Равны. **18.** Является. **19.** $\mathbf{P}\{X < \mathbf{M}[X]\} = \mathbf{P}\{[X] > \mathbf{M}[X]\}$; $\mathbf{P}\{|X - \mathbf{M}[X]| < \sqrt{\mathbf{D}[X]}\} = \dfrac{1}{\sqrt{3}}$. **20.** Они равны. **21.** Они равны. **22.** $\mathbf{M}[X] = \mathbf{D}[X] = \mathbf{M}[Y] = \mathbf{D}[Y] = 1$. **23.** $\mathbf{P}\{0 < X < 1\} < \mathbf{P}\{0 < Y < 1\}$. **24.** $X \sim \mathbf{Bi}(15; p)$, $p = \dfrac{5}{72}$; **25.** $\mathbf{M}[T] = 4 - \dfrac{4}{e}$. **26.** *а*) 100 тыс. руб.; *б*) 30; *в*) 100 тыс. руб. **27.** $\mathbf{M}[X] = 2{,}106$, $\mathbf{D}[X] = 1{,}9588$. **28.** *а*) $\mathbf{M}[X] = 60$, $\mathbf{D}[X] = 42$; *б*) $C_{200}^{60}(0{,}3)^{60}(0{,}7)^{140}$. **29.** 1. **30.** 16, 3,2. **31.** Средний выигрыш — 79 рублей. **32.** $\mathbf{M}[X] = 2000$, $\sigma_X = 40$. **33.** $h = 1/4$,

$$F(x) = \begin{cases} 0, & x \in (-\infty, -1], \\ 3(x+1)/4, & x \in (-1, 0], \\ 3/4, & x \in (0, 1], \\ 1/2 + x/4, & x \in (1, 2], \\ 1, & x \in (2, \infty], \end{cases} \qquad \mathbf{M}[(2 - X)(X - 3)] = -\frac{41}{6},$$

$\mathbf{D}[2 - 3X] = \dfrac{15}{2}$. **34.** 0,1168. **35.** $\mathbf{M}[X] = 10$, $\mathbf{D}[X] = 100$, $\mathbf{M}[X^2] = 200$. **36.** 0,082. **37.** $a(1-p) + pb$. **38.** Будет. **39.** $F_X(f_Y(\mathbf{M}[X])) > F_Y(f_X(\mathbf{M}[Y]))$. **40.** $1 - e^{-1}$. **41.** $F_X(\mathbf{M}[X]) > f_X(\mathbf{M}[X])$, $F_X(-\mathbf{M}[X]) = f_X(-\mathbf{M}[X])$. **42.** 0. **43.** $\mathbf{P}\{X < \mathbf{M}[X]\} < \mathbf{P}\{X > \mathbf{D}[X]\}$. **44.** $\dfrac{1}{2}$. **45.** −16, $1 - e^{-1}$, 1. **46.** 2, $\dfrac{2}{3}$, $\dfrac{1}{18}$, $\dfrac{3}{4}$. **47.** Для X вероятность больше. **48.** 0,5, 153, 81. **49.** e^{-10}. **50.** 45,99. **51.** 2. **52.** 3. **53.** $-\dfrac{\ln(1-p)}{\lambda}$. **54.** *а*) $a < 0$, $b = -\dfrac{1}{\pi}$, (или $a > 0$, $b = \dfrac{1}{\pi}$), $c = \dfrac{1}{2}$; *б*) $f(x) = \dfrac{a}{\pi(x^2 + a^2)}$.

Глава № 3

1. Вообще говоря, нет. **2.** 0,84. **3.** $\mathbf{M}[X^3 - Y^2] = -2$. **4.** $\mathbf{M}[X + Y] = 1$; $(X + Y) \sim \mathbf{Bi}\left(3; \dfrac{1}{3}\right)$. **5.** *б*) $\mathbf{M}[X] = 0$, $\mathbf{M}[Y] = \dfrac{11}{6}$, $\mathbf{M}[X + Y] = \dfrac{11}{6}$, $\mathbf{D}[X + Y] \approx 0{,}8055$, равенства справедливы. **6.** 0. **7.** 0,669. **8.** *а*) $a = \dfrac{1}{\pi^2}$; *б*) Независимы. **9.** СВ X и Y независимы и некоррелированны. **10.** Нет. **11.** Является. **12.** Будет. **13.** $k_{XX} = 1$. **14.** $k_{XY} = 0$. **15.** *а*) $r_{XY} = 1$; *б*) $r_{XY} = -1$. **16.** $k_{XY} = 0$. **17.** 3,35. **18.** Наиболее вероятным претендентом от первого блока является первый претендент, а от второго — второй претендент, $r_{XY} \approx 0{,}27$. **19.** Вероятность встретить дефект $B \approx 0{,}0118$.

Коэффициент корреляции $\approx 0{,}066$. **20.** $U \sim \mathbf{Bi}(1; 0{,}67)$.
21. $f(x,y) = \begin{cases} \lambda\mu e^{-(\lambda x+\mu y)}, & x \geqslant 0,\ y \geqslant 0, \\ 0 - & \text{в остальных случаях,} \end{cases}$
значение искомой вероятности равно $(1 - e^{-\lambda})(1 - e^{-\mu})$.
22. $0{,}1077$. **23.** $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. **24.** *а*) $\dfrac{3}{28}$; *б*) $\dfrac{3}{14}$.
25. *а*) $f(x,y) = \begin{cases} (\ln 2)^2 2^{-(x+y)}, & x \geqslant 0,\ y \geqslant 0, \\ 0 - & \text{в остальных случаях;} \end{cases}$ *б*) $0{,}0208$.

Глава № 4

1. $\dfrac{3}{4}$. **2.** Измененная верхняя граница: 70; оценка, полученная с помощью неравенства Чебышева: $0{,}88125$; уточненая оценка: $0{,}99626$. **3.** 325 562. **4.** Применим. **5.** Оценка, полученная с помощью неравенства Чебышева: $n \geqslant 5469$; уточненая оценка: $n \geqslant 1792$. **6.** $(26\,205,\ 33\,795)$. **7.** $\dfrac{3}{8}$. **8.** $p \leqslant 0{,}245$. **9.** $p \geqslant 0{,}6094$. **10.** Применим. **11.** $p \leqslant \dfrac{1}{5}$. **12.** $x \leqslant 4{,}16$. **13.** Оценка, полученная с помощью неравенства Чебышева: $p \geqslant 0$; уточненая оценка: $p \geqslant 0{,}64762$. **14.** Оценка, полученная с помощью неравенства Чебышева: $n \geqslant 12\,000$; уточненая оценка: $n \geqslant 2305$. **15.** *а*) $0{,}0291$; *б*) $0{,}9715$. **16.** $\dfrac{1}{2}$. **17.** Не менее 24 438 денежных единиц. **18.** $n = 975$.

Глава № 5

2. $\widehat{\theta} = 0{,}93$. **3.** $\widehat{m}_X = 165{,}77$, $\widehat{d}_X = 34{,}24$. **4.** $[-17{,}22, -10{,}28]$. **5.** $[10{,}35, 54{,}79]$. **6.** H_0 принимается. **7.** H_0 принимается. **8.** H_0 принимается. **9.** Результат действия лекарства не зависит от способа его применения. **10.** H_0 принимается. **11.** Гипотеза о независимости отвергается.

Глава № 6

1. 1068. **2.** 32617. **3.** $y = -48{,}59 + 1{,}94x$. **4.** $T = 61{,}84 - 0{,}67t + 0{,}04t^2$.

ТАБЛИЦЫ

1. Функция Лапласа $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-t^2/2}\, dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	,0039	,0079	,0119	,0159	,0199	,0239	,0279	,0318	,0358
0,1	,0398	,0438	,0477	,0517	,0556	,0596	,0635	,0674	,0714	,0753
0,2	,0792	,0831	,0870	,0909	,0948	,0987	,1025	,1064	,1102	,1140
0,3	,1179	,1217	,1255	,1293	,1330	,1368	,1405	,1443	,1480	,1517
0,4	,1554	,1591	,1627	,1664	,1700	,1736	,1772	,1808	,1843	,1879
0,5	,1914	,1949	,1984	,2019	,2054	,2088	,2122	,2156	,2190	,2224
0,6	,2257	,2290	,2323	,2356	,2389	,2421	,2453	,2485	,2517	,2549
0,7	,2580	,2611	,2642	,2673	,2703	,2733	,2763	,2793	,2823	,2852
0,8	,2881	,2910	,2938	,2967	,2995	,3023	,3051	,3078	,3105	,3132
0,9	,3159	,3185	,3212	,3228	,3263	,3289	,3314	,3339	,3364	,3389
1,0	,3413	,3437	,3461	,3485	,3508	,3531	,3554	,3576	,3599	,3621
1,1	,3643	,3665	,3686	,3707	,3728	,3749	,3769	,3790	,3810	,3829
1,2	,3849	,3868	,3887	,3906	,3925	,3943	,3961	,3979	,3997	,4014
1,3	,4032	,4049	,4065	,4082	,4098	,4114	,4130	,4146	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4250	,4264	,4278	,4292	,4305	,4318
1,5	,4331	,4344	,4357	,4369	,4382	,4394	,4406	,4417	,4429	,4440
1,6	,4452	,4463	,4473	,4484	,4495	,4505	,4515	,4525	,4535	,4544
1,7	,4554	,4563	,4572	,4581	,4590	,4599	,4608	,4616	,4624	,4632
1,8	,4640	,4648	,4656	,4663	,4671	,4678	,4685	,4692	,4699	,4706
1,9	,4712	,4719	,4725	,4732	,4738	,4744	,4750	,4755	,4761	,4767
2,0	,4772	,4777	,4783	,4788	,4793	,4798	,4803	,4807	,4812	,4816
2,1	,4821	,4825	,4830	,4834	,4838	,4842	,4846	,4850	,4853	,4857
2,2	,4861	,4864	,4867	,4871	,4874	,4877	,4880	,4884	,4887	,4889
2,3	,4892	,4895	,4898	,4901	,4903	,4906	,4908	,4911	,4913	,4915
2,4	,4918	,4920	,4922	,4924	,4926	,4928	,4930	,4932	,4934	,4936
2,5	,4937	,4939	,4941	,4943	,4944	,4946	,4947	,4949	,4950	,4952
2,6	,4953	,4954	,4956	,4957	,4958	,4959	,4960	,4962	,4963	,4964
2,7	,4965	,4966	,4967	,4968	,4969	,4970	,4971	,4972	,4972	,4973
2,8	,4974	,4975	,4976	,4976	,4977	,4978	,4978	,4979	,4980	,4980
2,9	,4981	,4981	,4982	,4983	,4983	,4984	,4984	,4985	,4985	,4986
3,0	,49865									
3,5	,499767									
4,0	,4999683									
4,5	,4999966									
5,0	,49999971									

2. Квантили $x_\alpha(n)$ уровня α распределения $\chi^2(n)$:

$$\mathbf{P}\{X \leqslant x_\alpha(n)\} = \alpha$$

$n \diagdown \alpha$	0,05	0,1	0,2	0,3	0,5	0,7	0,8	0,9	0,95
1	0,00	0,02	0,06	0,15	0,46	1,07	1,64	2,71	3,84
2	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,60	5,99
3	0,35	0,58	1,01	1,42	2,37	3,66	4,64	6,25	7,82
4	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49
5	1,15	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,07
6	1,65	2,20	3,07	3,83	5,35	7,23	8,56	10,64	12,59
7	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,02	14,07
8	2,73	3,49	4,59	5,53	7,34	9,52	11,03	13,36	15,51
9	3,32	4,17	5,38	6,39	8,34	10,66	12,24	14,68	16,92
10	3,94	4,86	6,18	7,27	9,34	11,78	13,44	15,99	18,31
11	4,58	5,58	6,99	8,15	10,34	12,90	14,63	17,28	19,68
12	5,23	6,30	7,81	9,03	11,34	14,01	15,81	18,55	21,03
13	5,89	7,04	8,63	9,93	12,34	15,12	16,98	19,81	22,36
14	6,57	7,79	9,47	10,82	13,34	16,22	18,15	21,06	23,69
15	7,26	8,55	10,31	11,72	14,34	17,32	19,31	22,31	25,00
16	7,96	9,31	11,15	12,62	15,34	18,42	20,47	23,54	26,29
17	8,67	10,08	12,00	13,53	16,34	19,51	21,62	24,78	27,60
18	9,39	10,86	12,86	14,44	17,34	20,60	22,76	25,59	28,87
19	10,11	11,65	13,72	15,35	18,34	21,70	23,90	27,20	30,14
20	10,85	12,44	14,58	16,27	19,34	22,80	25,04	28,41	31,41
21	11,59	13,24	15,44	17,18	20,30	23,90	26,17	29,61	32,67
22	12,34	14,04	16,31	18,10	21,30	24,90	27,30	30,81	33,92
23	13,09	14,85	17,19	19,02	22,30	26,00	28,43	32,01	35,17
24	13,85	15,66	18,06	19,94	23,30	27,10	29,55	33,20	36,42
25	14,61	16,47	18,94	20,90	24,30	28,20	30,78	34,38	37,65
26	15,38	17,29	19,82	21,80	25,30	29,20	31,80	35,56	38,89
27	16,15	18,11	20,70	22,70	26,30	30,30	32,91	36,74	40,11
28	16,93	18,94	21,60	23,60	27,30	31,40	34,03	37,92	41,34
29	17,71	19,77	22,50	24,60	28,30	32,50	35,14	39,09	42,56
30	18,49	20,60	23,40	25,50	29,30	33,50	36,25	40,26	43,77

3. Квантили $t_{\frac{\alpha+1}{2}}(n)$ уровня $\dfrac{\alpha+1}{2}$ распределения $\mathbf{S}\,(n)$:

$$\mathbf{P}\left\{T \leqslant t_{\frac{\alpha+1}{2}}(n)\right\} = \frac{\alpha+1}{2}$$

$n \setminus \alpha$	0,2	0,6	0,8	0,9	0,95	0,98	0,99	0,999
1	0,325	1,376	3,078	6,314	12,706	31,821	63,657	636,619
2	0,289	1,061	1,886	2,920	4,303	6,965	9,925	31,598
3	0,277	0,978	1,638	2,353	3,182	4,541	5,841	12,941
4	0,271	0,941	1,533	2,132	2,776	3,747	4,604	8,610
5	0,267	0,920	1,476	2,015	2,571	3,365	4,032	6,859
6	0,265	0,906	1,440	1,943	2,447	3,143	3,707	5,959
7	0,263	0,896	1,415	1,895	2,365	2,998	3,499	5,405
8	0,262	0,889	1,397	1,860	2,306	2,896	3,355	5,041
9	0,261	0,883	1,383	1,833	2,262	2,821	3,250	4,781
10	0,260	0,879	1,372	1,812	2,228	2,764	3,169	4,587
11	0,260	0,876	1,363	1,796	2,201	2,718	3,106	4,437
12	0,259	0,873	1,356	1,782	2,179	2,681	3,055	4,318
13	0,259	0,870	1,350	1,771	2,160	2,650	3,012	4,221
14	0,258	0,868	1,345	1,761	2,145	2,624	2,977	4,140
15	0,258	0,866	1,341	1,753	2,131	2,602	2,947	4,073
16	0,258	0,865	1,337	1,746	2,120	2,583	2,921	4,015
17	0,257	0,863	1,333	1,740	2,110	2,567	2,898	3,965
18	0,257	0,862	1,330	1,734	2,101	2,552	2,878	3,922
19	0,257	0,861	1,328	1,729	2,093	2,539	2,861	3,883
20	0,257	0,860	1,325	1,725	2,086	2,528	2,845	3,850
21	0,257	0,859	1,323	1,721	2,080	2,518	2,831	3,819
22	0,256	0,858	1,321	1,717	2,074	2,508	2,819	3,792
23	0,256	0,858	1,319	1,714	2,069	2,500	2,807	3,767
24	0,256	0,857	1,318	1,711	2,064	2,492	2,797	3,745
25	0,256	0,856	1,316	1,708	2,060	2,485	2,787	3,725
26	0,256	0,856	1,315	1,706	2,056	2,479	2,779	3,707
27	0,256	0,855	1,314	1,703	2,052	2,473	2,771	3,690
28	0,256	0,855	1,313	1,701	2,048	2,467	2,763	3,674
29	0,256	0,854	1,311	1,699	2,045	2,462	2,756	3,659
30	0,256	0,854	1,310	1,697	2,042	2,457	2,750	3,646
40	0,255	0,851	1,303	1,684	2,021	2,423	2,704	3,551
60	0,254	0,848	1,296	1,671	2,000	2,390	2,660	3,460
120	0,254	0,845	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,253	0,842	1,282	1,645	1,960	2,326	2,576	3,291

СПИСОК ЛИТЕРАТУРЫ

1. *Агапов Г.И.* Задачник по теории вероятностей. — М.: Высшая школа, 1986.

2. *Асриев А.В., Кибзун А.И.* Практикум по статистическому моделированию на ЭВМ. — М.: Изд-во МАИ, 1989.

3. *Вентцель Е.С.* Теория вероятностей. — М.: Наука, 1969.

4. *Вентцель Е.С., Овчаров Л.А.* Прикладные задачи теории вероятностей. — М.: Радио и Связь, 1983.

5. *Войтенко М.А.* Руководство к решению задач по теории вероятностей: Учебное пособие. / Под ред. А.И. Карасева. — М.: Изд-во ВЗФЭИ, 1998.

6. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1998.

7. *Мешалкин Л.Д.* Сборник задач по теории вероятностей. — М.: Изд-во МГУ, 1963.

8. *Горяинова Е.Р., Наумов А.В., Сиротин А.Н.* Практические занятия по курсу теории вероятностей. — М.: МАИ, 1999.

9. *Горяинова Е.Р., Наумов А.В., Сиротин А.Н.* Решение задач по теории вероятностей. — М.: МАИ, 2001.

10. *Емельянов Г.В., Скитович В.П.* Задачник по теории вероятностей и математической статистике. — Л.: Изд-во ЛГУ, 1967.

11. *Зубков А.М., Севастьянов Б.А., Чистяков В.П.* Сборник задач по теории вероятностей. — М.: Наука, 1989.

12. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. — М.: Высшая школа, 1984.

13. *Кан Ю.С.* Методические указания к практическим занятиям по теории вероятностей на базе компьютерного курса. — М.: Изд-во МАИ, 1996.

14. *Кибзун А.И., Панков А.Р., Сиротин А.Н.* Учебное пособие по теории вероятностей. — М.: Изд-во МАИ, 1993.

15. *Кибзун А.И., Наумов А.В.* Лекции по теории вероятностей. — М.: МАИ, 2000.

16. *Колемаев В.А., Староверов О.В., Турундаевский В.Б.* Теория вероятностей и математическая статистика. — М.: Высшая школа, 1991.

17. *Кожевников Ю.В.* Введение в математическую статистику. — Казань: Изд-во КГТУ, 1996.

18. *Королюк В.С. и др.* Справочник по теории вероятностей и математической статистике. — М.: Наука, 1985.

19. *Крамер Г.* Математические методы статистики. — М.: Мир, 1975.

20. *Лавренченко А.С.* Лекции по математической статистике и теории случайных процессов. — М.: МАИ, 1974.

21. *Прохоров А.В., Ушаков В.Г., Ушаков Н.Г.* Задачи по теории вероятностей. — М.: Наука, 1986.

22. *Пугачев В.С.* Введение в теорию вероятностей. — М.: Наука, 1968.

23. *Пугачев В.С.* Теория вероятностей и математическая статистика. — М.: Наука, 1979.

24. *Путко Б.А., Фридман М.Н., Исаева Р.Н., Медведева Л.А., Шатунова М.Н.* Математика (Теория вероятностей и математическая статистика) / Под ред. Н.Ш. Кремера. — М.: Экономическое образование, 1998.

25. *Ю.А. Розанов Ю.А.* Лекции по теории вероятностей. — М.: Наука, 1968.

26. *Секей Г.* Парадоксы в теории вероятностей и математической статистике. — М.: Мир, 1990.

27. *Чистяков В.П.* Курс теории вероятностей. — М.: Наука, 1987.

28. Сборник задач по математике для втузов. Часть 3. Специальные курсы. Под ред. А.В. Ефимова. — М.: Наука, 1984.

29. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. Под ред. *А.А. Свешникова.* — М.: Наука, 1970.

30. Теория вероятностей. Под. ред. *В.С.Зарубина, А.П.Крищенко.* — М.: Изд-во МВТУ им. Н.Э. Баумана, 1999.

31. *Hiller F.S., Lieberman G.J.* Introduction to Stochastic Models in Operations Research. McGraw–Hill, N.Y., 1990.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ