Контрольная работа по курсу «Теория вероятностей и математическая статистика»

Выполнил: ст. гр. 4030 Турбин А.
Проверил: доц. Королев В. Д.

РГРТА, 2005

Общие замечания по выполнению работы

В данной работе выполнены задания по курсу «Теория вероятностей и математическая статистика», предусмотренные в [1]. Вариант заданий №32 (по журналу группы 4030). В расчетах использованы значения $NV = 32$ (номер варианта), $NM = 35$ (общее количество студентов в группе); $KV = \frac{NV-1}{NM-1} = \frac{31}{34} \approx 0.912$ (индивидуальный коэффициент).

Текст работы подготовлен в системе LaTeX 2_ε, иллюстрации подготовлены при помощи пакета METAPOST. Проверка решений выполнена в системе математического моделирования R.

Список литературы

[1] Теория вероятностей. Математическая статистика. Случайные процессы: Методические указания; Сост. В. Д. Королев; Рязан. гос. радиотехн. акад. Рязань, 2002.

[2] Гмурман В. Е. Введение в теорию вероятностей и математическую статистику. М.: Высшая школа, 1966.

[3] Ф. А. Новиков. Дискретная математика для программистов. СПб: Питер, 2002.

[4] Общий курс высшей математики для экономистов: Учебник. Под ред. В. И. Ермакова. М.: ИНФРА-М, 2001.

[5] Доугерти К. Введение в эконометрику. М.: ИНФРА-М, 1997.

1 Игральные кости

Задача. Бросаются две игральные кости. Считая элементарными событиями упорядоченную пару (k, l), $k, l = 1, 2, 3, 4, 5, 6$, описать пространство элементарных событий, записать пространство элементарных событий в виде матрицы. Найти вероятность того, что сумма выпавших очков будет больше $3\,NV/NM$ и меньше или равна $9 + 2.5\,KV$. В четных вариантах найти ещё условную вероятность этого же события, если известно, что на первой кости выпало четное число, а в нечетных вариантах — если известно, что сумма выпавших очков нечетна.

Решение. Элементарное событие состоит в выпадении k очков на первой игральной кости и в выпадении l очков на второй игральной кости. Пространство элементарных событий можно *описать* множеством упорядоченных пар вида (k, l), $k, l = 1..6$ (всего 36 элементарных событий ω_{kl}). В соответствии с условием задачи элементарным исходом естественно считать также сумму выпавших очков $\sigma_{kl} = k + l$. Это позволяет записать пространство элементарных событий в виде матрицы (табл. 1.1).

Таблица 1.1

$k \backslash l$	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Все элементарные исходы естественно считать равновероятными: $\Pr(\omega_{kl}) = 1/36$.

Теперь необходимо определить количество элементарных исходов, которые удовлетворяют условиям задачи. Всего имеется m элементарных исходов, при которых сумма выпавших очков будет больше $3\,NV/NM = 3 \cdot 32/35 \approx 2.74$. Из табл. 1.1 легко видеть, что $m = 35$, т. к. условию $k + l > 2.74$ не удовлетворяет единственный элементарный исход $\sigma_{11} = 2$. Аналогично, всего имеется n элементарных исходов, при которых сумма выпавших очков будет меньше или равна $9 + 2.5\,KV \approx 9 + 2.5 \cdot 0.912 \approx 11.28$. Из табл. 1.1 легко видеть, что $n = 35$, т. к. условию $k + l \leqslant 11.28$ не удовлетворяет единственный элементарный исход $\sigma_{66} = 12$.

Таким образом, условиям $k + l > 2.74$ и $k + l \leqslant 11.28$ удовлетворяет 34 элементарных исхода (выделены в табл. 1.1), т. е. все исходы, кроме σ_{11} и σ_{66}. Это позволяет вычислить вероятность события, при котором данные условия выполняются: $\Pr(2.74 < k+l \leqslant 11.28) = 34/36 = 17/18 \approx 0.94$.

Далее, необходимо найти условную вероятность этого же события, если известно, что на первой кости выпало четное число очков. Обозначим через A событие, при котором выполняется главное условие задачи $(2.74 < k + l \leqslant 11.28)$; через B — выпадение четного числа очков на первой кости $(k = 2n)$. Требуется найти условную вероятность $\Pr(A/B)$.

Поскольку событие B произошло, пространство элементарных событий суживается до 18 элементов (табл. 1.2). А именно, наступление события B исключает из рассмотрения элементарные события ω_{kj}, $k = 1, 3, 5$, $j = 1..6$.

Таблица 1.2

$k \backslash l$	1	2	3	4	5	6
2	3	4	5	6	7	8
4	5	6	7	8	9	10
6	7	8	9	10	11	12

Определим количество n исходов, которые удовлетворяют условию $2.74 < k + l \leqslant 11.28$ в новом пространстве элементарных событий. Из табл. 1.2 легко видеть, что $n = 17$, т. к.

данному условию не удовлетворяет единственный элементарный исход $\sigma_{66} = 12$. Следовательно, условная вероятность $\Pr(A/B) = 17/18 \approx 0.94$.

К такому же заключению можно придти, используя *формулу Байеса*:

$$\Pr(A/B) = \frac{\Pr(A)\Pr(B/A)}{\Pr(B)}. \tag{1.1}$$

Из первой части задачи имеем $\Pr(A) = 17/18 \approx 0.94$. Вероятность выпадения четного числа очков на первой кости $\Pr(B) = 1/2$, т. к. естественно считать выпадение четного числа очков $k = 1, 3, 5$ и нечетного числа очков $k = 2, 4, 6$ равновероятными событиями.

Вычислим теперь вероятность $\Pr(B/A)$. Наступление события A означает, что пространство элементарных событий суживается до 34 элементов (табл. 1.1). При этом требуется определить вероятность выпадения четного числа очков на первой кости. Элементарные события, при которых число очков на первой кости четно, соответствуют «четным» строкам ω_{kj} таблицы 1.1, $k = 2, 4, 6$, $j = 1..6$. Таким образом, для подсчета элементарных событий, при которых число очков на первой кости четно, можно воспользоваться таблицей 1.2. А именно, таких событий всего $6+6+5 = 17$, поэтому $\Pr(B/A) = 17/34 = 1/2$. По формуле Байеса имеем

$$\Pr(A/B) = \frac{\Pr(A)\Pr(B/A)}{\Pr(B)} = \frac{17}{34} \cdot \frac{1}{2} \cdot \frac{2}{1} = \frac{17}{34} \approx 0.94.$$

Ответ: $\Pr(2.74 < k + l \leqslant 11.28) = \Pr(2.74 < k + l \leqslant 11.28/k = 2n) = 17/18 \approx 0.94$.

2 Выборочный контроль

Задача. Для реализации простейшей схемы выборочного контроля составить схему алгоритма вычисления гипергеометрических вероятностей по рекуррентной формуле. Вычислить $H(N, K, n, k)$ по рекуррентной формуле; непосредственно по формуле гипергеометрического распределения; используя формулу Стирлинга. Найти относительную погрешность последнего способа. $N = 100+r[100\,KV]$; $n = 5+r[4\,KV]$; $K = 10+r[10\,KV]$; $k = 2+r[2\,KV]$ (здесь и далее $r[z]$ — результат округления числа z до ближайшего целого).

Простейшая схема выборочного контроля состоит в следующем [1, с. 2]. Имеется партия из N изделий (объектов), среди которых K бракованных. Из N объектов извлекается выборка объемом n объектов, которые будут подвергнуты безошибочному контролю. Какова вероятность $H(N, K, n, k)$ присутствия k бракованных объектов в выборке?

Всего возможно C_N^n различных выборок. Оценим среди них количество выборок, содержащих k бракованных объектов. Бракованные объекты можно выбрать C_K^k количеством способов, оставшиеся годные $n - k$ объектов — C_{N-K}^{n-k} количеством способов. Таким образом,

$$H(N, K, n, k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}. \tag{2.1}$$

Эта формула задает *гипергеометрическое* распределение вероятностей, выраженное при помощи *числа сочетаний* (*биномиальных коэффициентов*):

$$C_n^k = \frac{n!}{k!\,(n-k)!}, \qquad C_n^0 = \frac{1}{0!} = 1. \tag{2.2}$$

Решение. Запишем рекуррентное соотношение для биномиальных коэффициентов:

$$\frac{C_n^{k+1}}{C_n^k} = \frac{n!}{(k+1)!\,(n-k-1)!} \cdot \frac{k!\,(n-k)!}{n!} = \frac{n-k}{k+1}. \tag{2.3}$$

Это позволяет записать рекуррентное соотношение для гипергеометрических вероятностей:

$$\begin{aligned}\frac{H(N,K,n,k+1)}{H(N,K,n,k)} &= \frac{C_K^{k+1}}{C_K^k}\cdot\frac{C_{N-K}^{n-k-1}}{C_{N-K}^{n-k}} = \frac{K-k}{k+1}\cdot\frac{n-k}{N-K-n+k+1} = \\ &= \frac{K-k}{k+1}\cdot\frac{1}{\dfrac{N-K+1}{n-k}-1}.\end{aligned} \tag{2.4}$$

При $k=0$ из (2.1) имеем:

$$\begin{aligned}H(N,K,n,0) &= \frac{C_{N-K}^n}{C_N^n} = \frac{(N-K)!\,(N-n)!}{N!\,(N-K-n)!} = \\ &= \frac{(N-K)(N-K-1)\ldots(N-K-n+1)}{N(N-1)\ldots(N-n+1)} = \\ &= \left(1-\frac{K}{N}\right)\left(1-\frac{K}{N-1}\right)\cdots\left(1-\frac{K}{N-n+1}\right).\end{aligned} \tag{2.5}$$

Формулы (2.4) и (2.5) позволяют сформулировать алгоритм вычисления функции H: при $k=0$ начальное значение функции H вычисляется по правой части формулы (2.5); при $k>0$ значение функции для предыдущего значения k домножается на коэффициент из правой части формулы (2.4). Заметим, что при переходе к новому значению H коэффициент из формулы (2.4) вычисляется для *предыдущего* значения k (так, для получения значения $H(k=1)$ вычисление коэффициента происходит при $k=0$).

Запишем алгоритм вычисления функции H на языке Паскаль:

```
function H(NN, KK, n, k: integer): double;
var
  v: double; { значение функции }
  i: integer; { переменная цикла }
begin
  v:=1;
  { начальное значение при k=0 }
  for i:=NN downto NN-n+1 do
    v:=v*(1-KK/i);
  { последовательное изменение значения при k=1,2,... }
  for i:=0 to k-1 do
    v:=v*(KK-i)/(i+1)/((NN-KK+1)/(n-i)-1);
  H:=v;
end;
```

Выполним предварительную проверку данного алгоритма. Для этого вычислим какое-нибудь значение функции H. Из [1, с. 3] известно, что $H(10,5,4,2)=10/21$. Запишем программу с вызовом функции H: `begin writeln(H(10,5,4,2)) end.` В результате выполнения программы получено верное значение `4.761904761904762E-001`.

Замечание. Если на каком-либо шаге выполняется соотношение $N-K+1=n-k$, то формулой (2.4) пользоваться нельзя. Так, составленный алгоритм не позволяет вычислить $H(10,7,6,4)=1/2$ [2, с. 16], т. к. на шаге $k=2$ знаменатель (2.4) обращается в 0.

Переходим к численному решению задачи. Имеем

$$\begin{aligned}N &= 100 + r[100\,KV] = 100 + r[100\cdot 0.912] = 100+91 = 191,\\ K &= 10 + r[10\,KV] = 10 + r[10\cdot 0.912] = 10+9 = 19,\\ n &= 5 + r[4\,KV] = 5 + r[4\cdot 0.912] = 5+4 = 9,\\ k &= 2 + r[2\,KV] = 2 + r[2\cdot 0.912] = 2+2 = 4.\end{aligned}$$

Вычислим начальное значение функции $H(191, 19, 9, k = 0)$ по формуле (2.5):

$$H(191, 19, 9, 0) = \left(1 - \frac{19}{191}\right)\left(1 - \frac{19}{191-1}\right)\cdots\left(1 - \frac{19}{191-9+1}\right) =$$
$$= \prod_{i=0}^{8}\left(1 - \frac{19}{191-i}\right) = \frac{16519373468}{43338469371} \approx 0.38117.$$

Замечание. Здесь и далее для получения точного результата воспользуемся возможностью символьных вычислений в системе *GNU CLISP*. Так, S-выражение для вычисления данного произведения имеет вид

```
(do ((i 0)(v 1))
    ((> i 8) v)
    (setq v (* v (- 1 (/ 19 (- 191 i)))))
    (setq i (+ i 1)))
```

Далее последовательно вычисляем значение функции H для $k = 1, 2, 3, 4$ по формуле (2.4).

$$H(191, 19, 9, k = 1) = \frac{16519373468}{43338469371} \cdot \frac{19-0}{0+1} \cdot \frac{1}{\dfrac{191-19+1}{9-0} - 1} =$$
$$= \frac{16519373468}{43338469371} \cdot \frac{171}{164} = \frac{5741489559}{14446156457} \approx 0.39744,$$

$$H(191, 19, 9, k = 2) = \frac{5741489559}{14446156457} \cdot \frac{19-1}{1+1} \cdot \frac{1}{\dfrac{191-19+1}{9-1} - 1} =$$
$$= \frac{5741489559}{14446156457} \cdot \frac{24}{55} = \frac{137795749416}{794538605135} \approx 0.17343,$$

$$H(191, 19, 9, k = 3) = \frac{137795749416}{794538605135} \cdot \frac{19-2}{2+1} \cdot \frac{1}{\dfrac{191-19+1}{9-2} - 1} =$$
$$= \frac{137795749416}{794538605135} \cdot \frac{119}{498} = \frac{32927096748}{794538605135} \approx 0.041441,$$

$$H(191, 19, 9, k = 4) = \frac{32927096748}{794538605135} \cdot \frac{19-3}{3+1} \cdot \frac{1}{\dfrac{191-19+1}{9-3} - 1} =$$
$$= \frac{32927096748}{794538605135} \cdot \frac{24}{167} = \frac{4732037856}{794538605135} \approx 0.0059557.$$

Замечание. Как видим, с возрастанием k значение функции H убывает, что в целом представляется правдоподобным: наибольшего значения функция достигает при $\frac{K}{N} \approx \frac{k}{n}$, в данном случае при $k = 1$.

Вычислим теперь $H(191, 19, 9, 4)$ при помощи ранее составленного алгоритма на языке Паскаль. Программа с вызовом функции H имеет вид: `begin writeln(H(191,19,9,4)) end`. В результате выполнения программы получено значение `5.955705393567353E-003`, которое с высокой точностью совпадает с результатами расчета.

Далее, вычислим $H(191, 19, 9, 4)$ непосредственно по формуле (2.1):

$$H(191, 19, 9, 4) = \frac{C_{19}^{4}\, C_{191-19}^{9-4}}{C_{191}^{9}} = \frac{C_{19}^{4}\, C_{172}^{5}}{C_{191}^{9}} = \frac{3876 \cdot 1183009464}{769907908375815} = \frac{4732037856}{794538605135} \approx 0.0059557.$$

Замечание. Вычисление в данном случае осложняется тем, что при использовании практически любых примитивных типов данных (`integer`, `double` и т. д.) возникает переполнение — дело в том, что число 191! содержит 355 десятичных знаков. Другими словами, факториал растет настолько быстро, что «не вмещается» ни в один «примитивный» тип. Но это ограничение можно до некоторой степени преодолеть при помощи символических вычислений, при которых используется принципиально иное представление чисел. При этом использование символических вычислений не связано с какими-либо дополнительными трудностями. Так, в данном случае для вычисления биномиальных коэффициентов использовались следующие функции:

```
(defun fact (n) (do ((v 1)) ((= n 0) v) (setq v (* v n)) (setq n (- n 1))))
(defun C (n k) (/ (fact n) (* (fact k) (fact (- n k)))))
```

Покажем, как вычислить $H(191, 19, 9, 4)$ приближенно, используя *формулу Стирлинга*

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{2.6}$$

Из (2.2) получаем

$$\begin{aligned} C_n^k = \frac{n!}{k!\,(n-k)!} &\approx \frac{\sqrt{2\pi n}\,n^n}{\sqrt{2\pi k}\,k^k\,\sqrt{2\pi(n-k)}\,(n-k)^{n-k}} = \\ &= \sqrt{\frac{n}{2\pi k(n-k)}}\cdot\frac{n^n}{k^k(n-k)^{n-k}} = \sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}. \end{aligned} \tag{2.7}$$

Тогда

$$C_{19}^4 \approx \sqrt{\frac{19}{2\pi\cdot 4\cdot 15}}\left(\frac{19}{4}\right)^4\left(\frac{19}{15}\right)^{15} \approx 0.224497\cdot 509.0664\cdot 34.6685 \approx 3962.0578,$$

$$C_{172}^5 \approx \sqrt{\frac{172}{2\pi\cdot 5\cdot 167}}\left(\frac{172}{5}\right)^5\left(\frac{172}{167}\right)^{167} \approx 0.18106\cdot 48171726.6\cdot 137.911 \approx 1202882526.75,$$

$$C_{191}^9 \approx \sqrt{\frac{191}{2\pi\cdot 9\cdot 182}}\left(\frac{191}{9}\right)^9\left(\frac{191}{182}\right)^{182} \approx 0.13623\cdot 873208028911\cdot 6532.5 \approx 777083605983290$$

и

$$H(191, 19, 9, 4) \approx \frac{3962.0578\cdot 1202882526.75}{777083605983290} \approx 0.006133.$$

Относительная погрешность приближенного вычисления составляет

$$\left|\frac{\Delta H}{H}\right| \approx \frac{0.006133 - 0.0059557}{0.0059557} \approx 0.02977 \approx 3\%.$$

Ответ: $H(191, 19, 9, 4) = \dfrac{4732037856}{794538605135} \approx 0.0059557.$

3 Предоставление кредита

Задача. Коммерсант ведет переговоры о предоставлении ему кредита с четырьмя банками. Вероятность получения кредита он оценивает соответственно как 0.2, 0.3 и 0.4. Считая, что банки принимают решения независимо друг от друга, найти вероятность того, что хотя бы в одном банке он получит кредит. Решать двумя способами.

Решение. Обозначим через A_1, A_2 и A_3 события, которые состоят в получении кредита в первом, втором и третьем банках соответственно; $p_1 = \Pr(A_1) = 0.2$, $p_2 = \Pr(A_2) = 0.3$, $p_3 = \Pr(A_3) = 0.4$. Противоположные события (отказ в предоставлении кредита) обозначим через $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$; тогда $q_1 = \Pr(\bar{A}_1) = 1 - 0.2 = 0.8$, $q_2 = \Pr(\bar{A}_2) = 1 - 0.3 = 0.7$, $q_3 = \Pr(\bar{A}_3) = 1 - 0.4 = 0.6$. Искомое событие $A = A_1 + A_2 + A_3$, противоположное ему событие $\bar{A} = \bar{A}_1\bar{A}_2\bar{A}_3$.

По теореме сложения вероятностей совместных событий

$$\Pr(A + B) = \Pr(A) + \Pr(B) - \Pr(AB). \tag{3.1}$$

События A_1, A_2 и A_3 совместны, поэтому

$$\begin{aligned}\Pr(A_1 + A_2 + A_3) &= \Pr(A_1 + A_2) + \Pr(A_3) - \Pr((A_1 + A_2)A_3) = \\ &= \Pr(A_1) + \Pr(A_2) + \Pr(A_3) - \\ &\quad - \Pr(A_1A_2) - \Pr(A_2A_3) - \Pr(A_1A_3) + \\ &\quad + \Pr(A_1A_2A_3)\end{aligned} \tag{3.2}$$

(в общем случае теорема обобщается с помощью *принципа включения и исключения* при объединении конфигураций [3, с. 151]). По условию задачи события A_1, A_2, A_3 попарно независимы и, более того, независимы в совокупности [2, с. 33]. Это означает, что для вычисления вероятностей $\Pr(A_1A_2)$, $\Pr(A_2A_3)$, $\Pr(A_1A_3)$ и $\Pr(A_1A_2A_3)$ можно применить теорему умножения вероятностей независимых событий:

$$\Pr(A_1A_2\ldots A_n) = \Pr(A_1)\Pr(A_2)\ldots\Pr(A_n). \tag{3.3}$$

Это позволяет вычислить вероятность искомого события:

$$\begin{aligned}\Pr(A) = \Pr(A_1 + A_2 + A_3) &= p_1 + p_2 + p_3 - p_1p_2 - p_2p_3 - p_1p_3 + p_1p_2p_3 = \\ &= 0.2 + 0.3 + 0.4 - 0.2 \cdot 0.3 - 0.2 \cdot 0.4 - 0.3 \cdot 0.4 + 0.2 \cdot 0.3 \cdot 0.4 = \\ &= 0.9 - 0.06 - 0.08 - 0.12 + 0.024 = 0.664.\end{aligned}$$

С другой стороны, вероятность искомого события можно выразить через вероятность противоположного события:

$$\Pr(A) = 1 - \Pr(\bar{A}_1\bar{A}_2\bar{A}_3) = 1 - q_1q_2q_3 = 1 - 0.8 \cdot 0.7 \cdot 0.6 = 1 - 0.336 = 0.664.$$

Замечание. Эквивалентность двух решений следует из алгебраического тождества

$$p_1 + p_2 + p_3 - p_1p_2 - p_2p_3 - p_1p_3 + p_1p_2p_3 = 1 - (1 - p_1)(1 - p_2)(1 - p_3).$$

Ответ: $\Pr(A) = 0.664$.

4 Дефектные компьютеры

Задача. На стендах находятся N компьютеров, из которых K имеют скрытые дефекты. Покупатель отбирает друг за другом наугад 3 компьютера. Найти вероятность следующих событий. В четных вариантах: первые два компьютера хорошие, третий — дефектный; первый и третий из выбранных компьютеров дефектные, второй хороший. В нечетных вариантах слова «хороший» и «дефектный» следует поменять местами. $N = 15 + r[10\,KV]$, $K = 3 + r[5\,KV]$.

Решение. Обозначим события: A_1 — первый компьютер хороший, A_2 — второй компьютер хороший, A_3 — третий компьютер хороший (соответственно $\bar{A}_1$, $\bar{A}_2$, $\bar{A}_3$ — соответствующий компьютер дефектный). В данном эксперименте все эти события зависимы, т. к. очередной компьютер извлекается из числа оставшихся. Требуется найти вероятность событий $A_1A_2\bar{A}_3$ и $\bar{A}_1A_2\bar{A}_3$.

Воспользуемся теоремой умножения вероятностей зависимых событий:

$$\Pr(A_1A_2A_3\dots A_n) = \Pr(A_1)\Pr(A_2/A_1)\Pr(A_3/A_1A_2)\dots\Pr(A_n/A_1A_2\dots A_{n-1}). \tag{4.1}$$

Тогда

$$\Pr(A_1A_2\bar{A}_3) = \Pr(A_1)\Pr(A_2/A_1)\Pr(\bar{A}_3/A_1A_2),$$
$$\Pr(\bar{A}_1A_2\bar{A}_3) = \Pr(\bar{A}_1)\Pr(A_2/\bar{A}_1)\Pr(\bar{A}_3/\bar{A}_1A_2).$$

Поясним, каким образом вычисляются условные вероятности. При отборе первого компьютера всего имеется N компьютеров, из которых K имеют дефекты. Вероятность выбрать первый компьютер хорошим $\Pr(A_1) = \frac{N-K}{N}$, вероятность выбрать первый компьютер дефектным $\Pr(\bar{A}_1) = \frac{K}{N}$. Отбор второго компьютера производится из $N-1$ оставшихся компьютеров, из которых дефекты имеют либо K компьютеров (если произошло событие A_1), либо $K-1$ компьютеров (если произошло событие $\bar{A}_1$). Поэтому условные вероятности выбрать второй компьютер хорошим

$$\Pr(A_2/A_1) = \frac{(N-1)-K}{N-1} = \frac{N-K-1}{N-1}, \quad \Pr(A_2/\bar{A}_1) = \frac{(N-1)-(K-1)}{N-1} = \frac{N-K}{N-1},$$

а условные вероятности выбрать второй компьютер плохим

$$\Pr(\bar{A}_2/A_1) = \frac{K}{N-1}, \quad \Pr(\bar{A}_2/\bar{A}_1) = \frac{K-1}{N-1}.$$

Аналогичные рассуждения справедливы при отборе третьего компьютера. Следовательно,

$$\Pr(A_1A_2\bar{A}_3) = \frac{N-K}{N}\cdot\frac{N-K-1}{N-1}\cdot\frac{K}{N-2}, \quad \Pr(\bar{A}_1A_2\bar{A}_3) = \frac{K}{N}\cdot\frac{N-K}{N-1}\cdot\frac{K-1}{N-2}.$$

Имеем $N = 15 + r[10\,KV] = 15 + 9 = 24$, $K = 3 + r[5\,KV] = 3 + 5 = 8$ и

$$\Pr(A_1A_2\bar{A}_3) = \frac{16}{24}\cdot\frac{15}{23}\cdot\frac{8}{22} = \frac{40}{253} \approx 0.158, \quad \Pr(\bar{A}_1A_2\bar{A}_3) = \frac{8}{24}\cdot\frac{16}{23}\cdot\frac{7}{22} = \frac{56}{759} \approx 0.0738.$$

Ответ: $\Pr(A_1A_2\bar{A}_3) = \dfrac{40}{253} \approx 0.158$, $\Pr(\bar{A}_1A_2\bar{A}_3) = \dfrac{56}{759} \approx 0.0738$.

5 Электрическая цепь

Задача. Электрическая цепь состоит из последовательно и параллельно включенных резисторов (рис. 5.1). Вероятности их отказа равны соответственно $p_1 = p_2 = \ldots = p_m$, $p_{m+1} = p_{m+2} = \ldots = p_{m+n}$. Считая отказы независимыми, найти вероятность безотказной работы цепи (т. е. прохождения по ней тока от A к B). $m = 3 + r[3\,KV]$, $n = 5 - r[3\,KV]$; $p_1 = 0.01 + 0.09\,KV$, $p_{m+1} = 0.1 - 0.05\,KV$.

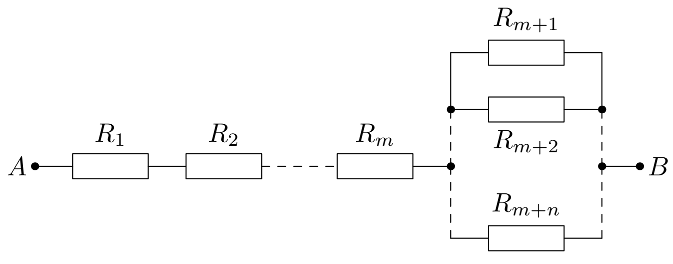

Рис. 5.1. Электрическая цепь.

Решение. Разрыв цепи возникает, если выходит из строя любой из резисторов $R_1, \dots, R_m$, либо если выходит из строя каждый из резисторов $R_{m+1}, \dots, R_{m+n}$. Обозначим через A_i выход из строя i-го резистора. Тогда разрывом цепи является событие $A_1 + A_2 + \ldots + A_m + A_{m+1}A_{m+2}\ldots A_{m+n}$, а искомым событием (безотказная работа цепи) является противоположное событие $\bar{A} = \bar{A}_1\bar{A}_2\ldots\bar{A}_m\overline{A_{m+1}A_{m+2}\ldots A_{m+n}}$. Действительно, для прохождения тока в цепи необходимо, чтобы одновременно работали (не выходили из строя) резисторы $R_1, \dots, R_m$, а также необходимо, чтобы не вышли из строя одновременно все резисторы $R_{m+1}, \dots, R_{m+n}$ (основать при помощи законов де Моргана, имея в виду теоретико-множественный характер элементарных событий).

Независимость событий $\{A_i\}$ и $\{\bar{A}_i\}$ в совокупности позволяет применить теорему умножения вероятностей независимых событий:

$$\begin{aligned}\Pr(\bar{A}) &= \Pr(\bar{A}_1)\Pr(\bar{A}_2)\ldots\Pr(\bar{A}_m)\Pr(\overline{A_{m+1}A_{m+2}\ldots A_{m+n}}) = \\ &= (1-\Pr(A_1))(1-\Pr(A_2))\ldots(1-\Pr(A_m))(1-\Pr(A_{m+1})\Pr(A_{m+2})\ldots\Pr(A_{m+n})).\end{aligned}$$

Для случая $p_1 = p_2 = \ldots = p_m$, $p_{m+1} = p_{m+2} = \ldots = p_{m+n}$

$$\Pr(\bar{A}) = (1-p_1)^m(1-p_{m+1}^n).$$

Имеем

$$\begin{gathered}m = 3 + r[3\,KV] = 3 + 3 = 6,\\ n = 5 - r[3\,KV] = 5 - 3 = 2,\\ p_1 = 0.01 + 0.09\,KV \approx 0.01 + 0.082 = 0.182,\\ p_{m+1} = 0.1 - 0.05\,KV \approx 0.1 - 0.0456 = 0.0544\end{gathered}$$

Тогда

$$\Pr(\bar{A}) = (1-0.182)^6\,(1-0.0544^2) \approx 0.2996 \cdot (1-0.002959) \approx 0.2987.$$

Ответ: вероятность безотказной работы цепи 0.2987 (т. е. примерно 30%).

6 Изготовление замков

Задача. На предприятии, изготавливающем замки, первый цех производит $P_1\%$, второй $P_2\%$, третий $P_3\%$ всех замков. Брак соответственно $Q_1\%$, $Q_2\%$, $Q_3\%$. Найти вероятность того, что случайно выбранный замок окажется годным. Если замок оказался бракованным, какова вероятность того, что он был изготовлен в k-м цехе, $k = 1, 2, 3$? $P_1 = 25 + 5\,KV$, $P_2 = 15 + 10\,KV$; $Q_1 = 3 + 4\,KV$, $Q_2 = 6 - 4\,KV$, $Q_3 = 10 - 7\,KV$.

Решение. События задачи: A — случайно выбранный замок годный (искомое событие), $\bar{A}$ — случайно выбранный замок бракованный. Гипотезы задачи: H_1 — замок изготовлен в первом цехе, H_2 — замок изготовлен во втором цехе, H_3 — замок изготовлен в третьем цехе. Гипотезы H_1, H_2, H_3 являются условиями наступления события A (или $\bar{A}$), причем все гипотезы несовместны (замок может быть изготовлен в одном и только одном из цехов), т. е. $A = AH_1 + AH_2 + AH_3$. Тогда по *формуле полной вероятности*

$$\Pr(A) = \Pr(H_1)\Pr(A/H_1) + \Pr(H_2)\Pr(A/H_2) + \Pr(H_3)\Pr(A/H_3).$$

По условию задачи

$$\begin{gathered}\Pr(H_1) = P_1, \quad \Pr(H_2) = P_2, \quad \Pr(H_3) = P_3,\\ \Pr(A/H_1) = 1 - Q_1, \quad \Pr(A/H_2) = 1 - Q_2, \quad \Pr(A/H_3) = 1 - Q_3.\end{gathered}$$

Тогда (считая переход от процентов очевидным)

$$\Pr(A) = P_1(1 - Q_1) + P_2(1 - Q_2) + P_3(1 - Q_3).$$

Имеем

$$\begin{gathered}
P_1 = 25 + 5\,KV = 25 + 5 \cdot 0.912 = 29.56\% = 0.2956, \\
P_2 = 15 + 10\,KV = 15 + 10 \cdot 0.912 = 24.12\% = 0.2412, \\
P_3 = 1 - P_1 - P_2 = 0.4632, \\
Q_1 = 3 + 4\,KV = 3 + 4 \cdot 0.912 = 6.648\% = 0.06648, \\
Q_2 = 6 - 4\,KV = 6 - 4 \cdot 0.912 = 2.352\% = 0.02352, \\
Q_3 = 10 - 7\,KV = 10 - 7 \cdot 0.912 = 3.616\% = 0.03616,
\end{gathered}$$

следовательно,

$$\begin{gathered}
\Pr(A) = 0.2956(1 - 0.06648) + 0.2412(1 - 0.02352) + 0.4632(1 - 0.03616) \approx 0.9579, \\
\Pr(\bar{A}) = 1 - 0.9579 = 0.0421.
\end{gathered}$$

Далее, при условии, что замок оказался бракованным, требуется вычислить условные вероятности $\Pr(H_1/\bar{A})$, $\Pr(H_2/\bar{A})$, $\Pr(H_3/\bar{A})$. По формуле Байеса имеем

$$\begin{aligned}
\Pr(H_1/\bar{A}) &= \frac{\Pr(H_1)\Pr(\bar{A}/H_1)}{\Pr(\bar{A})} = \frac{0.2956 \cdot 0.06648}{0.0421} \approx 0.4668, \\
\Pr(H_2/\bar{A}) &= \frac{\Pr(H_2)\Pr(\bar{A}/H_2)}{\Pr(\bar{A})} = \frac{0.2412 \cdot 0.02352}{0.0421} \approx 0.1348, \\
\Pr(H_3/\bar{A}) &= \frac{\Pr(H_3)\Pr(\bar{A}/H_3)}{\Pr(\bar{A})} = \frac{0.4632 \cdot 0.03616}{0.0421} \approx 0.3978.
\end{aligned}$$

Замечание. Наибольшие значения условных вероятностей получены для первого и третьего цехов, что в целом представляется правдоподобным: у первого цеха наибольший процент брака, а третий цех производит наибольший объем продукции.

Ответ: $\Pr(A) = 0.9579$, $\Pr(H_1/\bar{A}) = 0.4668$, $\Pr(H_2/\bar{A}) = 0.1348$, $\Pr(H_3/\bar{A}) = 0.3978$.

7 Биномиальное распределение

Задача. Найти математическое ожидание, дисперсию и с. к. о. биномиального распределения с помощью непосредственного суммирования. $n = 3 + 3\,KV$, $p = 0.1 + 0.8\,KV$. Найти также вероятность попадания соответствующей случайной величины в интервал $[0 + KV, 2 + KV]$.

Понятие *биномиального распределения* вводится следующим образом [2, с. 93]. Производится n независимых испытаний, в каждом из которых вероятность появления события A постоянна и равна p (вероятность непоявления $q = 1 - p$). В качестве дискретной случайной величины рассматривается число k появлений события А в n испытаниях, $k = 0..n$. Вероятность появления события k раз в n независимых испытаниях (*формула Бернулли*)

$$P_n(k) = C_n^k p^k q^{n-k}. \tag{7.1}$$

Решение. *Математическое ожидание* дискретной случайной величины

$$M(X) = \sum_i x_i p_i. \tag{7.2}$$

Математическое ожидание случайной величины K, распределенной по биномиальному закону (при заданном n)

$$M(K) = \sum_{k=0}^{n} kP_n(k) = \sum_{k=1}^{n} k\,C_n^k p^k q^{n-k}. \tag{7.3}$$

По условию имеем $n = 3 + 3\,KV = 3 + 3 \cdot 0.912 = 5.736 \approx 6$, $p = 0.1 + 0.8\,KV = 0.8296$, $q = 1 - p = 0.1704$. Запишем числа сочетаний «из шести»:

$$C_6^0 = C_6^6 = 1, \quad C_6^1 = C_6^5 = 6, \quad C_6^2 = C_6^4 = 15, \quad C_6^3 = 20.$$

Переходим к непосредственному суммированию:

$$\begin{aligned}\sum_{k=1}^{6} k\,C_6^k p^k q^{6-k} &= 1 \cdot 6 \cdot 0.8296 \cdot 0.1704^5 + 2 \cdot 15 \cdot 0.8296^2 \cdot 0.1704^4 + 3 \cdot 20 \cdot 0.8296^3 \cdot 0.1704^3 + \\ &\quad + 4 \cdot 15 \cdot 0.8296^4 \cdot 0.1704^2 + 5 \cdot 6 \cdot 0.8296^5 \cdot 0.1704 + 6 \cdot 1 \cdot 0.8296^6 \cdot 1 = \\ &= 1 \cdot 6 \cdot 0.00011918 + 2 \cdot 15 \cdot 0.00058025 + 3 \cdot 20 \cdot 0.0028249 + \\ &\quad + 4 \cdot 15 \cdot 0.0137535 + 5 \cdot 6 \cdot 0.06697 + 6 \cdot 1 \cdot 0.32599 = \\ &= 1 \cdot 0.0007151 + 2 \cdot 0.0087038 + 3 \cdot 0.0564995 + \\ &\quad + 4 \cdot 0.2063029 + 5 \cdot 0.401758 + 6 \cdot 0.32599 = 4.9776\end{aligned}$$

Здесь в последней строке вторыми сомножителями вычислены вероятности $P_n(k)$.

Проверка: $M(K) = np = 6 \cdot 0.8296 = 4.9776$ [2, с. 111].

Дисперсия случайной величины

$$D(X) = M[X - M(X)]^2 = M(X^2) - M^2(X). \tag{7.4}$$

Вычислим $M(K^2)$ с помощью непосредственного суммирования:

$$\begin{aligned}M(K^2) = \sum_{k=1}^{6} k^2 P_n(k) &= 1^2 \cdot 0.0007151 + 2^2 \cdot 0.0087038 + 3^2 \cdot 0.0564995 + \\ &\quad + 4^2 \cdot 0.2063029 + 5^2 \cdot 0.401758 + 6^2 \cdot 0.32599 = \\ &= 1 \cdot 0.0007151 + 4 \cdot 0.0087038 + 9 \cdot 0.0564995 + \\ &\quad + 16 \cdot 0.2063029 + 25 \cdot 0.401758 + 36 \cdot 0.32599 = 25.6246848.\end{aligned}$$

Тогда

$$D(K) = M(K^2) - M^2(K) = 25.6246848 - 4.9776^2 = 0.84818304.$$

Проверка: $D(K) = npq = 6 \cdot 0.8296 \cdot 0.1704 = 0.84818304$ [2, с. 123].

Среднее квадратическое отклонение случайной величины

$$\sigma(X) = \sqrt{D(X)}. \tag{7.5}$$

Имеем $\sigma(K) = \sqrt{D(K)} = \sqrt{0.84818304} \approx 0.92097$.

Вероятность попадания дискретной случайной величины в интервал $[0.912, 2.912]$ в данном случае равна вероятности того, что случайная величина принимает значение $k = 1$ или $k = 2$. Следовательно,

$$\begin{aligned}\Pr(0.912 < K < 2.912) &= \Pr(K = 1) + \Pr(K = 2) = \\ &= P_n(1) + P_n(2) = 0.0007151 + 0.0087038 = 0.0094189\end{aligned}$$

Ответ: при $n = 6$, $p = 0.8296$ математическое ожидание 4.9776, дисперсия 0.84818304, с. к. о. 0.92097; вероятность попадания в интервал $[0.912, 2.912]$ 0.0094189.

8 Экспоненциальное распределение

Задача. Найти математическое ожидание, дисперсию, с. к. о., медиану и моду экспоненциального распределения; $F_\xi(x) = 1 - e^{-x/b}$ при $x \geqslant 0$, $F_\xi(x) = 0$ при $x < 0$. $b = 0.1 + 10\,KV$.

Решение. Введем обозначение $\lambda = 1/b$. Экспоненциальное распределение в условии задано при помощи (интегральной) функции распределения

$$F(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-\lambda x}, & \text{если } x \geqslant 0. \end{cases} \tag{8.1}$$

Соответствующая ей плотность вероятности

$$f(x) = \frac{dF}{dx} = \begin{cases} 0, & \text{если } x < 0, \\ \lambda e^{-\lambda x}, & \text{если } x \geqslant 0. \end{cases} \tag{8.2}$$

Математическое ожидание случайной величины X, имеющей экспоненциальное распределение

$$M(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{\infty} x\, \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} x e^{-\lambda x} dx. \tag{8.3}$$

Выполним интегрирование по частям $\int u\, dv = uv - \int v\, du$, полагая $u = x$, $dv = e^{-\lambda x} dx$; тогда $du = dx$, $v = -\dfrac{e^{-\lambda x}}{\lambda}$ и

$$\int x e^{-\lambda x} dx = -\frac{x e^{-\lambda x}}{\lambda} + \int \frac{e^{-\lambda x} dx}{\lambda} = -\frac{x e^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2} = -\frac{(\lambda x + 1)\, e^{-\lambda x}}{\lambda^2},$$

$$\int_0^{\infty} x e^{-\lambda x} dx = -\frac{(\lambda x + 1)\, e^{-\lambda x}}{\lambda^2}\bigg|_0^{\infty} = -\lim_{x\to\infty} (x e^{-x}) + \frac{e^0}{\lambda^2} = 0 + \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Следовательно,

$$M(X) = \lambda \cdot \frac{1}{\lambda^2} = \frac{1}{\lambda} = b. \tag{8.4}$$

Замечание. Справедливость аналитических преобразований проверена в системе символьных вычислений Maxima. Так, проверка неопределенного и определенного интегралов для вычисления математического ожидание выглядит следующим образом:

```
Maxima 5.9.2 http://maxima.sourceforge.net
Using Lisp CLISP 2.35 (2005-08-29)
Distributed under the GNU Public License. See the file COPYING.
(%i1) integrate(x*lambda*%e^(-lambda*x), x);
                                             - x lambda
                           (x lambda + 1) %e
(%o1)                    - --------------------------
                                       lambda
(%i2) integrate(x*lambda*%e^(-lambda*x), x, 0, inf);
Is  lambda  positive, negative, or zero?
positive;
                                         1
(%o2)                                 ------
                                      lambda
(%i3)
```

Здесь введены два выражения для интегрирования: `integrate(`*выражение, переменная*`)` и `integrate(`*выражение, переменная, нижний предел, верхний предел*`)`; запись `%e` обозначает символьную константу $e \approx 2.71828$, что отличает ее от имени переменной `e`. В дальнейшем также дан ответ `positive` на уточняющий вопрос (значение λ влияет на сходимость интеграла). Система вычисляет неопределенный и определенный интегралы в общем виде, то есть, вообще говоря, не производит численных вычислений (однако система производит упрощение выражений без потери точности).

Дисперсия случайной величины X, имеющей экспоненциальное распределение

$$D(X) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - M^2(X) = \lambda \int_0^{\infty} x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2}. \tag{8.5}$$

Выполним интегрирование по частям, полагая $u = x^2$, $dv = e^{-\lambda x} dx$; тогда $du = 2x\, dx$, $v = -\dfrac{e^{-\lambda x}}{\lambda}$ и (учитывая предыдущий результат)

$$\int x^2 e^{-\lambda x} dx = -\frac{x^2 e^{-\lambda x}}{\lambda} + \frac{2}{\lambda}\int x e^{-\lambda x} dx = -\frac{(\lambda^2 x^2 + 2\lambda x + 2) e^{-\lambda x}}{\lambda^3},$$

$$\int_0^{\infty} x^2 e^{-\lambda x} dx = -\frac{(\lambda^2 x^2 + 2\lambda x + 2) e^{-\lambda x}}{\lambda^3}\bigg|_0^{\infty} = 0 + \frac{2}{\lambda^3} = \frac{2}{\lambda^3}.$$

Следовательно,

$$D(X) = \frac{2\lambda}{\lambda^3} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} = b^2. \tag{8.6}$$

Медиана $Me(X)$ непрерывной случайной величины X определяется соотношениями

$$\Pr(X < Me(X)) = \Pr(X \geqslant Me(X)) = F(Me(X)) = \frac{1}{2}. \tag{8.7}$$

Другими словами, для нахождения медианы $Me(X)$ нужно решить уравнение $F(x) = 1 - e^{-\lambda x} = 0.5$; используя элементарные преобразования и логарифмирование, получаем

$$e^{-\lambda x} = \frac{1}{2}, \quad \ln e^{-\lambda x} = -\lambda x = \ln\frac{1}{2} = -\ln 2, \quad x = \frac{\ln 2}{\lambda}.$$

Следовательно, медиана случайной величины X, имеющей экспоненциальное распределение

$$Me(X) = \frac{\ln 2}{\lambda} = b \ln 2. \tag{8.8}$$

Модой $Mo(X)$ непрерывной случайной величины X называется такое ее значение, которому соответствует максимальное значение плотности ее вероятности. Нетрудно видеть, что плотность вероятности $f(x)$ случайной величины, имеющей экспоненциального распределения, достигает максимального значения в точке $x = 0$ (при этом $f(0) = \lambda$).

По условию задачи $b = 0.1 + 10\,KV = 0.1 + 10 \cdot 0.912 = 9.22$, тогда $M(X) = b = 9.22$, $D(X) = b^2 = 85.0084$, $\sigma(X) = \sqrt{D(X)} = b = 9.22$, $Me(X) = b \ln 2 \approx 9.22 \cdot 0.69315 \approx 6.3908$.

Ответ: при $b = 9.22$ математическое ожидание 9.22, дисперсия 85.0084, с. к. о. 9.22, медиана 6.3908, мода 0.

9 Нормальное распределение

Задача. Случайная величина ξ имеет нормальное распределение с математическим ожиданием $\mu = 10 + 20\,KV$ и с. к. о. $\sigma = 0.5 + |\mu|(0.5 + 2\,KV)$. Найти вероятности попадания этой случайной величины в интервалы $(-\infty, KV)$, $[KV, KV+1]$, $[2\,KV, +\infty]$. Найти также квантили порядков $0.1(0.1 + KV)$, $0.9(1 + 0.05\,KV)$. Построить график функции и плотности распределения. При использовании таблиц можно в случае необходимости ограничиться линейной интерполяцией.

Решение. Плотность вероятности и интегральная функция нормального распределения

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \tag{9.1}$$

Вероятность попадания нормально распределенной случайной величины ξ в заданный интервал (a, b) (или $[a, b]$ — вероятность попадания в точки a, b равна нулю)

$$\Pr(a < \xi < b) = \Phi\Big(\frac{b-\mu}{\sigma}\Big) - \Phi\Big(\frac{a-\mu}{\sigma}\Big), \tag{9.2}$$

где $\Phi(x)$ — (нормированная) *функция Лапласа*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{x^2}{2}}\, dx. \tag{9.3}$$

Свойства функции Лапласа:

$$\Phi(0) = 0, \quad \lim_{x\to+\infty} \Phi(x) = \frac{1}{2}, \quad \Phi(-x) = -\Phi(x). \tag{9.4}$$

Будем считать, что вычисление значений функции Лапласа не представляет сложности. Заметим однако, что в современных системах имеется тенденция к использованию «функции ошибок» $\operatorname{erf}(x) = 2\Phi(\sqrt{2}x)$ в качестве основной неэлементарной функции такого рода.

Переходим к численному решению задачи. По условию имеем $\mu = 10 + 20 \cdot 0.912 = 28.24$, $\sigma = 0.5 + 28.24\,(0.5 + 2 \cdot 0.912) = 66.12976$. Требуется найти вероятность попадания в интервалы $(-\infty, 0.912)$, $[0.912, 1.912]$, $[1.824, +\infty]$. Из (9.2) и (9.4) имеем

$$\begin{aligned}
\Pr(-\infty < \xi < 0.912) &= \Phi\left(\frac{0.912 - 28.24}{66.12976}\right) - \lim_{x\to-\infty} \Phi\left(\frac{x - 28.24}{66.12976}\right) = \\
&= \Phi(-0.4132481) + \frac{1}{2} = -0.1602876 + \frac{1}{2} = 0.3397124, \\
\Pr(0.912 < \xi < 1.912) &= \Phi\left(\frac{1.912 - 28.24}{66.12976}\right) - \Phi\left(\frac{0.912 - 28.24}{66.12976}\right) = \\
&= \Phi(-0.3981263) + 0.1602876 = -0.1547315 + 0.1602876 = 0.0055561, \\
\Pr(1.824 < \xi < +\infty) &= \lim_{x\to+\infty} \Phi\left(\frac{x - 28.24}{66.12976}\right) - \Phi\left(\frac{1.824 - 28.24}{66.12976}\right) = \\
&= \frac{1}{2} - \Phi(-0.3994571) = \frac{1}{2} + 0.1552218 = 0.6552218.
\end{aligned}$$

Вычислим также вероятность попадания в интервал $[1.824, 1.912]$:

$$\Pr(1.824 < \xi < 1.912) = -0.1547315 + 0.1552218 = 0.0004903.$$

Теперь несложно проверить условие нормировки:

$$\begin{aligned}
&\Pr(-\infty < \xi < 0.912) + \Pr(0.912 < \xi < 1.912) + \\
&\qquad + \Pr(1.824 < \xi < +\infty) - \Pr(1.824 < \xi < 1.912) = \\
&\qquad\qquad = 0.3397124 + 0.0055561 + 0.6552218 - 0.0004903 = 1.
\end{aligned}$$

Квантилем (или *квантилью*, англ. *quantile*) непрерывной случайной величины ξ называется ее значение x_q, отвечающее заданному уровню вероятности q (порядку квантиля), такое, что выполняется равенство

$$\Pr(\xi < x_q) = F_\xi(x_q) = q. \tag{9.5}$$

Для нормального распределения в силу (9.2) и (9.4) получаем

$$\Pr(\xi < x_q) = \Pr(-\infty < \xi < x_q) = \Phi\Big(\frac{x_q - \mu}{\sigma}\Big) + \frac{1}{2} = q, \quad \Phi\Big(\frac{x_q - \mu}{\sigma}\Big) = q - \frac{1}{2}. \tag{9.6}$$

Для вычисления квантилей в общем виде требуется указать функцию, обратную интегральной функции распределения. Тогда $x_q = F_\xi^{-1}(q)$. Для нормального распределения задача вычисления квантилей в общем виде сводится к нахождению функции, обратной функции Лапласа; однако в данном случае, как и в общем случае, нахождение обратной функции едва ли возможно. На практике квантили вычисляются с помощью последовательных приближений методом Ньютона (см. файл `qnorm.c` в исходных кодах системы R).

По условию задачи требуется найти квантили порядков

$$q_1 = 0.1(0.1 + KV) = 0.1(0.1 + 0.912) = 0.1012, \quad q_2 = 0.9(1 + 0.05\,KV) = 0.94104.$$

С учетом (9.4) и (9.6) по таблице значений функции Лапласа [2, с. 368–369] находим

$$\Phi(1.27) = 0.3980 \approx -\Big(q_1 - \frac{1}{2}\Big) = 0.3988, \quad \Phi(1.56) = 0.4406 \approx \Big(q_2 - \frac{1}{2}\Big) = 0.44104,$$
$$\frac{x_{q_1} - 28.24}{66.12976} = -1.27, \quad x_{q_1} = -1.27 \cdot 66.12976 + 28.24 = -55.7447952,$$
$$\frac{x_{q_2} - 28.24}{66.12976} = 1.56, \quad x_{q_2} = 1.56 \cdot 66.12976 + 28.24 = 131.4024256.$$

Проверка: вычислим квантили x_{q_1} и x_{q_2} в системе математического моделирования R.

```
R : Copyright 2004, The R Foundation for Statistical Computing
Version 1.9.1  (2004-06-21), ISBN 3-900051-00-3

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R in publications.

> qnorm(0.1012, mean=28.24, sd=66.12976);
[1] -56.05849
> qnorm(0.94104, mean=28.24, sd=66.12976);
[1] 131.6381
>
```

Здесь используется функция `qnorm(`q`, mean=`μ`, sd=`σ`)`, которая вычисляет квантиль нормального распределения при заданных параметрах методом Ньютона. Считая полученные значения достаточно точными, видим, что относительная погрешность вычисления квантилей «по таблице» не превышает 1%.

Графики функций f_ξ и F_ξ нормального распределения для $\mu = 28.24$, $\sigma = 66.12976$ на интервале $(\mu - 4\sigma, \mu + 4\sigma)$ приведены на рис. 9.1 и рис. 9.2.

Ответ: численные значения искомых величин приведены в решении.

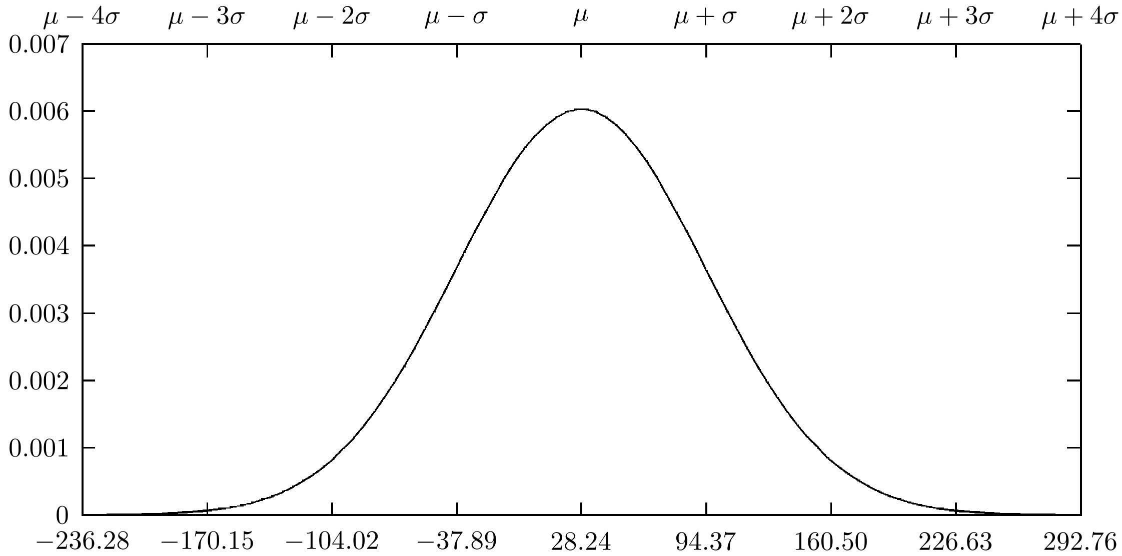

Рис. 9.1. Плотность вероятности нормального распределения.

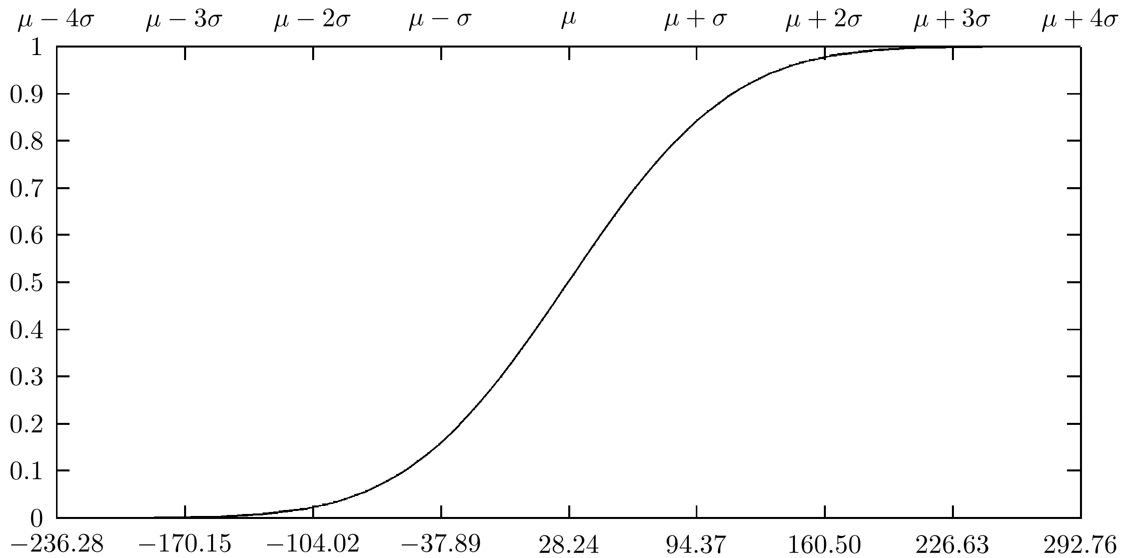

Рис. 9.2. Интегральная функция нормального распределения.

10 Случайный вектор

Задача. Закон распределения дискретного случайного вектора (ξ, μ) с реализациями (x_i, y_i) задан следующим образом: $\Pr(\xi = x_i, \mu = y_i) = \frac{i+j}{40(KV+2)+2j-i}$ для $j = 1, 2, 3, 4$, $i = 0, 1, 2, 3$ кроме пары $i = 3$, $j = 4$. $\Pr(\xi = x_3, \mu = y_4) = 1 - \sum \Pr(\xi = x_i, \mu = y_i)$, где суммирование ведется по всем парам (i, j) кроме $i = 3$, $j = 4$. $x_i = 10i + 20\,KV$, $y_j = 5j + 10\,KV$. Записать ковариационную и корреляционную матрицы; построить график теоретической функции регрессии: в нечетных вариантах ξ на μ, в четных вариантах μ на ξ; вероятности попаданий случайной точки в следующую область: $-4.5 + 10\,KV \leqslant x < 15\,KV + 2.2$; $-2.7 + 5\,KV < y < 7KV + 2.3$.

Решение. Будем также обозначать случайную величину ξ через X и случайную величину μ через Y. Для расчета исходных данных задачи (закон распределения случайного вектора) составим программу на языке Паскаль.

```
const
  NN=3; MM=4;
  KV=0.912;
var
  i, j: integer; { переменные циклов }
  sum: double; { для накопления сумм }
  xi: array[1..NN] of double; { значения с.в. X }
  yj: array[1..MM] of double; { значения с.в. Y }
  pij: array[1..NN,1..MM] of double; { вероятности (x[i],y[j]) }
begin
  for i:=1 to NN do { вычисление значений с.в. X }
    xi[i]:=10*i+20*KV;
  for j:=1 to MM do { вычисление значений с.в. Y }
    yj[j]:=5*j+10*KV;
  for i:=1 to NN do { вычисление вероятностей (x[i],y[j]) }
    for j:=1 to MM do
      pij[i,j]:=(i+j)/(40*(KV+2)+2*j-i);
  pij[NN,MM]:=0; { обнуление вероятности (x[3],y[4]) }
  sum:=0; { вычисление суммы вероятностей }
  for i:=1 to NN do
    for j:=1 to MM do
      sum:=sum+pij[i,j];
  writeln('Сумма вероятностей ', sum:1:6);
  pij[NN,MM]:=1-sum; { вычисление вероятности (x[3],y[4]) }
  writeln('Закон распределения случайного вектора (X=x[i],Y=y[j])');
  for j:=1 to MM do begin { вывод значений y[j] }
    if j=1 then write(' ':13);
    write('y', j, '=', yj[j]:1:6, ' ');
    if j=MM then writeln;
  end;
  for i:=1 to NN do { вывод значений x[i] и p[ij] }
    for j:=1 to MM do begin
      if j=1 then write('x', i, '=', xi[i]:1:6, ' ');
      write('p', i, j, '=', pij[i,j]:1:6, ' ');
      if j=MM then writeln;
    end;
end.
```

Вывод программы:

```
Сумма вероятностей 0.392973
Закон распределения случайного вектора (X=x[i],Y=y[j])
             y1=14.120000 y2=19.120000 y3=24.120000 y4=29.120000
x1=28.240000 p11=0.017024 p12=0.025109 p13=0.032927 p14=0.040492
x2=38.240000 p21=0.025755 p22=0.033761 p23=0.041501 p24=0.048988
x3=48.240000 p31=0.034638 p32=0.042560 p33=0.050218 p34=0.607027
```

Несложная текстовая обработка полученных результатов позволяет непосредственно записать закон распределения случайного вектора (ξ, μ) в виде таблицы (табл. 10.1).

Табл. 10.1. Закон распределения случайного вектора (ξ, μ).

	$y_1 = 14.12$	$y_2 = 19.12$	$y_3 = 24.12$	$y_4 = 29.12$
$x_1 = 28.24$	$p_{11} = 0.017024$	$p_{12} = 0.025109$	$p_{13} = 0.032927$	$p_{14} = 0.040492$
$x_2 = 38.24$	$p_{21} = 0.025755$	$p_{22} = 0.033761$	$p_{23} = 0.041501$	$p_{24} = 0.048988$
$x_3 = 48.24$	$p_{31} = 0.034638$	$p_{32} = 0.042560$	$p_{33} = 0.050218$	$p_{34} = 0.607027$

Для дальнейшего решения задачи требуется вычислить условные математические ожидания. *Условное математическое ожидание* дискретной случайной величины Y при $X = x_0$ (соответственно, случайной величины X при $Y = y_0$)

$$M(Y/X = x_0) = \sum_{j=1}^{m} y_j p(y_j/x_0), \quad M(X/Y = y_0) = \sum_{i=1}^{n} x_i p(x_i/y_0). \tag{10.1}$$

Условные вероятности составляющих

$$p(x_i/y_j) = \frac{p(x_i, y_j)}{p(y_j)}, \quad p(y_j/x_i) = \frac{p(x_i, y_j)}{p(x_i)}, \tag{10.2}$$

при этом (безусловные) вероятности составляющих

$$p(x_i) = \sum_{j=1}^{m} p(x_i, y_j), \quad p(y_j) = \sum_{i=1}^{n} p(x_i, y_j). \tag{10.3}$$

Условное математическое ожидание $M(Y/X = x_0)$ можно рассматривать как функцию от x, которая каждому возможному значению x_0 случайной величины X ставит в соответствие значение условного математического ожидания случайной величины Y. Условное математическое ожидание в таком случае называют *функцией регрессии* Y на X:

$$M(Y/X = x) = M(Y/x) = \varphi(x). \tag{10.4}$$

Дополним ранее составленную программу:

```
var
  { ... }
  pi: array[1..NN] of double; { вероятности x[i] составляющей X }
  pj: array[1..MM] of double; { вероятности y[j] составляющей Y }
  cond_pij: array[1..NN,1..MM] of double; { условные вероятности p(x[i]/y[j]) }
  cond_pji: array[1..MM,1..NN] of double; { условные вероятности p(y[j]/x[i]) }
begin
  { ... }
  writeln('Вероятности составляющей X');
  for i:=1 to NN do begin { вычисление вероятности p(x[i]) }
    sum:=0; { суммирование p(x[i],y[j]) по j }
    for j:=1 to MM do
      sum:=sum+pij[i,j];
    pi[i]:=sum;
    writeln('p(x', i, ')=', pi[i]:1:6);
  end;
  writeln('Вероятности составляющей Y');
  for j:=1 to MM do begin { вычисление вероятности p(y[j]) }
    sum:=0; { суммирование p(x[i],y[j]) по i }
    for i:=1 to NN do
      sum:=sum+pij[i,j];
    pj[j]:=sum;
    writeln('p(y', j, ')=', pj[j]:1:6);
  end;
  { вычисление условных вероятностей p(x[i]/y[j]) и p(y[j]/x[i]) }
  for i:=1 to NN do
    for j:=1 to MM do begin
      cond_pij[i,j]:=pij[i,j]/pj[j];
```

```
      cond_pji[j,i]:=pij[i,j]/pi[i];
    end;
  writeln('Условные мат. ожидания M(Y/X=x[i])');
  for i:=1 to NN do begin
    sum:=0;
    for j:=1 to MM do
      sum:=sum+yj[j]*cond_pji[j,i];
    writeln('M(Y/X=x', i, ')=', sum:1:6);
  end;
  writeln('Условные мат. ожидания M(X/Y=y[j])');
  for j:=1 to MM do begin
    sum:=0;
    for i:=1 to NN do
      sum:=sum+xi[i]*cond_pij[i,j];
    writeln('M(X/Y=y', j, ')=', sum:1:6);
  end;
end.
```

Дополнительный вывод программы:

```
Вероятности составляющей X
p(x1)=0.115553
p(x2)=0.150005
p(x3)=0.734443
Вероятности составляющей Y
p(y1)=0.077418
p(y2)=0.101430
p(y3)=0.124646
p(y4)=0.696507
Условные мат. ожидания M(Y/X=x[i])
M(Y/X=x1)=23.312369
M(Y/X=x2)=22.910559
M(Y/X=x3)=27.491196
Условные мат. ожидания M(X/Y=y[j])
M(X/Y=y1)=40.515172
M(X/Y=y2)=39.960555
M(X/Y=y3)=39.627164
M(X/Y=y4)=46.373939
```

Полученные результаты позволяют построить график функции регрессии Y на X (рис. 10.1), а также график функции регрессии X на Y (рис. 10.2).

Ковариацией (*корреляционным моментом*) случайных величин X и Y называют смешанный центральный момент второго порядка:

$$\operatorname{cov}(X,Y) = \mu_{xy} = M[(X-M(X))(Y-M(Y))] = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - M(X))(y_j - M(Y))p(x_i, y_j), \tag{10.5}$$

при этом $M(X)$ и $M(Y)$ с учетом (10.3) вычисляются как обычно.

Коэффициент корреляции случайных величин X и Y выражает ковариацию в безразмерных (относительных) единицах:

$$\rho_{xy} = \frac{\mu_{xy}}{\sigma_x\sigma_y}, \quad -1 \leqslant \rho_{xy} \leqslant 1. \tag{10.6}$$

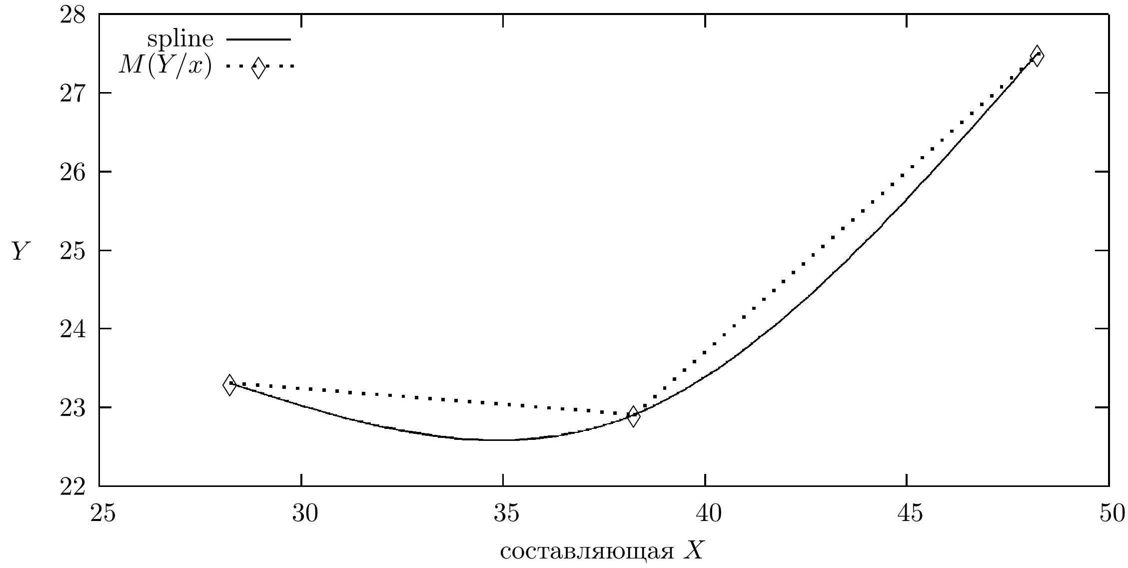

Рис. 10.1. Кривая регрессии Y на X.

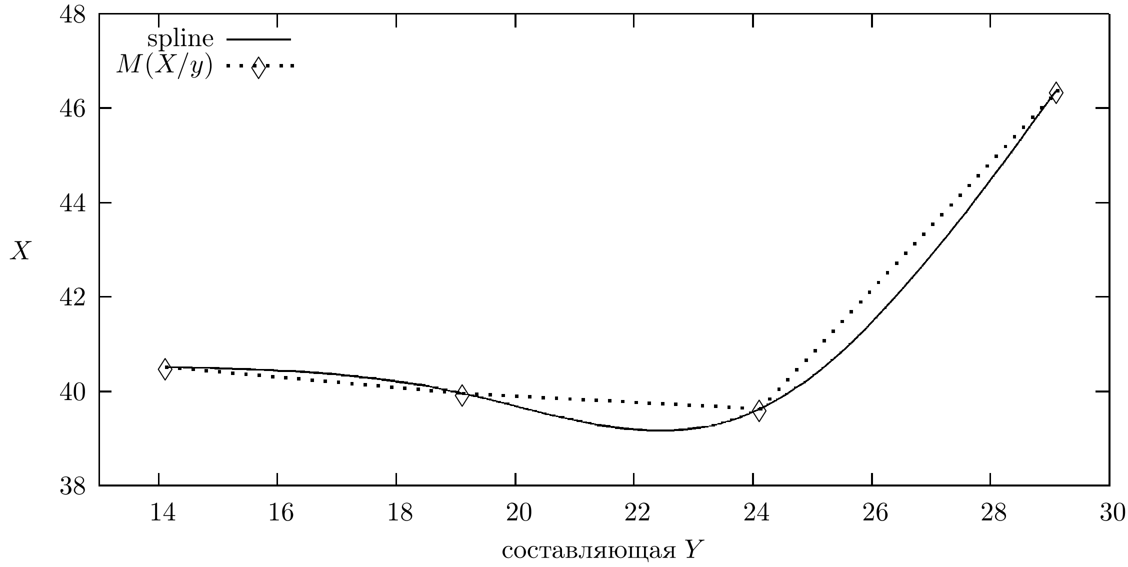

Рис. 10.2. Кривая регрессии X на Y.

Дополним программу ещё раз:

```
var
  { ... }
  Mx, My, Dx, Dy, sdX, sdY: double; { мат. ожидания, дисперсии и с.к.о. }
  cov, rho: double; { ковариация и коэффициент корреляции (rho - греч. буква) }
begin
  { ... }
  sum:=0; { вычисление M(X) }
  for i:=1 to NN do
    sum:=sum+xi[i]*pi[i];
  Mx:=sum;
  writeln('M(X)=', Mx:1:6);
  sum:=0; { вычисление M(Y) }
  for j:=1 to MM do
    sum:=sum+yj[j]*pj[j];
  My:=sum;
```

```
  writeln('M(Y)=', My:1:6);
  sum:=0; { вычисление D(X) }
  for i:=1 to NN do
    sum:=sum+xi[i]*xi[i]*pi[i];
  Dx:=sum-Mx*Mx;
  sdX:=sqrt(Dx);
  writeln('D(X)=', Dx:1:6, ' sdX=', sdX:1:6);
  sum:=0; { вычисление D(Y) }
  for j:=1 to MM do
    sum:=sum+yj[j]*yj[j]*pj[j];
  Dy:=sum-My*My;
  sdY:=sqrt(Dy);
  writeln('D(Y)=', Dy:1:6, ' sdY=', sdY:1:6);
  sum:=0; { вычисление cov(X,Y) }
  for i:=1 to NN do
    for j:=1 to MM do
      sum:=sum+(xi[i]-Mx)*(yj[j]-My)*pij[i,j];
  cov:=sum;
  writeln('cov(X,Y)=', cov:1:6);
  rho:=cov/sdX/sdY;
  writeln('rho(X,Y)=', rho:1:6);
end.
```

Дополнительный вывод программы:

```
M(X)=44.428901
M(Y)=26.321205
D(X)=46.697034 sdX=6.833523
D(Y)=22.844887 sdY=4.779633
cov(X,Y)=12.069705
rho(X,Y)=0.369537
```

Результаты расчетов позволяют записать *ковариационную* и *корреляционную матрицы*:

$$C(X,Y) = \begin{pmatrix} D(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(X,Y) & D(X) \end{pmatrix}, \quad Corr = \begin{pmatrix} 1 & \rho_{xy} \\ \rho_{xy} & 1 \end{pmatrix}. \tag{10.7}$$

Имеем $D(X) = 46.697034$, $\mathrm{cov}(X,Y) = 12.069705$, $\rho_{xy} = 0.369537$, поэтому

$$C(X,Y) = \begin{pmatrix} 46.697034 & 12.069705 \\ 12.069705 & 46.697034 \end{pmatrix}, \quad Corr = \begin{pmatrix} 1 & 0.369537 \\ 0.369537 & 1 \end{pmatrix}.$$

Ответ: численные значения приведены в решении.

11 Оценка параметров выборки

Задача. В таблице 11.1 даны 10 чисел, являющихся реализациями случайной величины, равномерно распределенной на $[0,1]$ (нуль целых опущен).

Табл. 10.1.

i	1	2	3	4	5	6	7	8	9	10
R_i	5427	3815	2912	4018	8434	2351	21511	0790	9569	9786

На основе каждой пары чисел R_i, R_{i+1} следует получить выборку $\{x_i\}$, порожденную нормально распределенной случайной величиной с параметрами μ, σ, следующим образом: $x_i = \mu + \sigma\sqrt{-2\ln R_i}\cos(2\pi R_{i+1})$ $x_{i+1} = \mu + \sigma\sqrt{-2\ln R_i}\sin(2\pi R_{i+1})$, $i = 1, 3, 5, 7, 9$ (объем нормальной выборки равен количеству чисел R_i: из пары R_1, R_2 получаются x_1, x_2; из пары R_3, R_4 — x_3, x_4 и т. д.). Значения μ и σ такие же, как в задаче 9. На основе полученной выборки найти оценки математического ожидания, дисперсии, с. к. о.; построить доверительные интервалы для этих характеристик (для μ как при неизвестном σ, так и при известном σ, взятом из задачи 9); проверить гипотезу H_0: $\mu = 0$ против H_1: $\mu \neq 0$, взяв то же значение σ и $\alpha = 0.01 + 0.09\,KV$; произвести проверку нормальности выборки по критерию Колмогорова; $\gamma = 1 - \alpha$.

Решение. Если условие задачи понято нами правильно, на основе чисел R_i требуется получить выборку $\{x_i\}$ из 10 элементов, в которой нечетные и четные элементы вычисляются различно (с использованием «косинуса» и «синуса» соответственно). В дальнейшем необходимо исследовать некоторые свойства полученной таким образом выборки.

В таком случае решение задачи предполагает значительный объем однотипных вычислений (расчетов), справедливость которых не поддается элементарной проверке; поэтому целесообразно производить расчеты на ЭВМ.

Запишем алгоритм получения выборки $\{x_i\}$ и вычисления точечных оценок выборки на языке Паскаль.

```
const
  NN = 10;
  R: array[1..NN] of double = { таблица 11.1 }
    (0.5427, 0.3815, 0.2912, 0.4018, 0.8434,
     0.2351, 0.21511, 0.0790, 0.9569, 0.9786);
  mu = 28.24; sigma = 66.12976; { параметры задачи 9 }
var
  i, j: integer; { переменные циклов }
  sum: double; { для суммирования }
  x: array[1..NN] of double; { выборка x[i] }
  Mx, Dx, sdX: double; { точечные оценки м.о., дисперсии и с.к.о }
begin
  i:=1; { вычисление x[i] }
  while i<NN do begin
    x[i  ]:=mu+sigma*sqrt(-2*ln(R[i]))*cos(2*pi*R[i+1]);
    x[i+1]:=mu+sigma*sqrt(-2*ln(R[i]))*sin(2*pi*R[i+1]);
    writeln('x', i  , '=', x[i  ]:1:6);
    writeln('x', i+1, '=', x[i+1]:1:6);
    i:=i+2;
  end;
  sum:=0; { среднее выборочное }
  for i:=1 to NN do
    sum:=sum+x[i];
  Mx:=sum/NN;
  writeln('Среднее выборочное ', Mx:1:6);
  sum:=0; { выборочная "исправленная" дисперсия }
  for i:=1 to NN do
    sum:=sum+(x[i]-Mx)*(x[i]-Mx);
  Dx:=sum/(NN-1);
  writeln('Выборочная дисперсия ', Dx:1:6);
  sdX:=sqrt(Dx);
  writeln('Выборочное с.к.о. ', sdX:1:6);
```

```
end.
```

Вывод программы:

```
x1=-25.527467
x2=77.785735
x3=-56.484434
x4=88.343759
x5=31.848017
x6=66.666519
x7=130.178515
x8=83.449956
x9=47.692641
x10=25.608517
Среднее выборочное 46.956176
Выборочная дисперсия 3105.333859
Выборочное с.к.о. 55.725523
```

Доверительный интервал для оценки математического ожидания нормального распределения при неизвестном σ с вероятностью γ

$$M(X) \in \left[\bar{X} \pm \frac{S \cdot t_{n-1;0.5(1+\gamma)}}{\sqrt{n}}\right], \tag{11.1}$$

где $\bar{X}$ — среднее выборочное, S — «исправленное» выборочное с. к. о., $t_{n-1;0.5(1+\gamma)}$ — квантиль порядка $0.5(1+\gamma)$ распределения Стьюдента с $n-1$ степенями свободы.

По условию задачи $\alpha = 0.01 + 0.09 \cdot 0.912 = 0.09208$, $\gamma = 1 - \alpha = 0.90792$, поэтому находим $t_{9;0.95396} = 1.884922$; следовательно, доверительный интервал для $\bar{X} = 46.956176$ при $S = 55.725523$ с вероятностью примерно 91%

$$46.956176 \pm \frac{55.725523 \cdot 1.884922}{\sqrt{10}},$$

т. е. примерно 46.956 ± 33.216.

Доверительный интервал для оценки математического ожидания нормального распределения при известном σ

$$M(X) \in \left[\bar{X} \pm \frac{\sigma \cdot \psi_{0.5(1+\gamma)}}{\sqrt{n}}\right], \tag{11.2}$$

где ψ_q — квантиль порядка q стандартного нормального распределения.

Имеем $\sigma = 66.12976$, $\psi_{0.95396} = 1.684526$, поэтому доверительный интервал

$$46.956176 \pm \frac{55.725523 \cdot 1.684526}{\sqrt{10}},$$

т. е. примерно 46.956 ± 29.685.

Доверительный интервал для оценки дисперсии нормального распределения

$$D(X) \in \left[\frac{(n-1)S^2}{\chi^2_{n-1;0.5(1+\gamma)}}, \frac{(n-1)S^2}{\chi^2_{n-1;0.5(1-\gamma)}}\right], \tag{11.3}$$

где $\chi^2_{n-1;q}$ — квантиль порядка q распределения χ^2 с $n-1$ степенями свободы.

Находим $\chi^2_{9;0.95396} = 17.17534$, $\chi^2_{9;0.04604} = 3.241189$, поэтому доверительный интервал

$$\left[\frac{9 \cdot 3105.333859}{17.17534}, \frac{9 \cdot 3105.333859}{3.241189}\right],$$

т. е. примерно $[1627.217, 8622.763]$. Следовательно, доверительный интервал для с. к. о. примерно $[40.339, 92.859]$.

Проверим гипотезу H_0: $\mu = 0$. Значение $\mu = 0$ не попадает в доверительный интервал для оценки математического ожидания (как при неизвестном, так и при известном σ), поэтому гипотезу H_0 нужно отвергнуть, а принять гипотезу H_1: $\mu \neq 0$.

Проверим нормальность выборки по критерию Колмогорова. Часть вычислений целесообразно провести в системе R (система позволяет работать с наборами значений). Далее приведен протокол работы с системой (с комментариями).

```
> # исходные данные
> data=c(-25.527467, 77.785735, -56.484434, 88.343759, 31.848017,
+ 66.666519, 130.178515, 83.449956, 47.692641, 25.608517);
> # упорядочиваем выборку по возрастанию
> data=sort(data)
> data
 [1] -56.48443 -25.52747  25.60852  31.84802  47.69264  66.66652  77.78574
 [8]  83.44996  88.34376 130.17852
> # оценка mu и sigma
> mu=mean(data)
> mu
[1] 46.95618
> sigma=sqrt(var(data))
> sigma
[1] 55.72552
> # значения функции распределения для упорядоченной выборки
> Fx=pnorm(data, mean=mu, sd=sigma)
> Fx
 [1] 0.03170883 0.09667611 0.35082806 0.39315037 0.50527224 0.63821966
 [7] 0.70995028 0.74372894 0.77116963 0.93233807
> # последовательность i=1..10
> i=1:10
> # покажем последовательности i/10 и (i-1)/10
> i/10; (i-1)/10
 [1] 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
 [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
> # теперь можно вычислить значения |F(X)-(i-1)/n|
> D1=abs(Fx-(i-1)/10)
> D1
 [1] 0.031708827 0.003323891 0.150828058 0.093150370 0.105272244 0.138219658
 [7] 0.109950283 0.043728935 0.028830374 0.032338071
> # а также значения |F(X)-i/n|
> D2=abs(Fx-i/10)
> D2
 [1] 0.068291173 0.103323891 0.050828058 0.006849630 0.005272244 0.038219658
 [7] 0.009950283 0.056271065 0.128830374 0.067661929
> # из D1 и D2 нужно выбрать максимальные значения
> D=pmax(D1,D2)
> D
 [1] 0.06829117 0.10332389 0.15082806 0.09315037 0.10527224 0.13821966
 [7] 0.10995028 0.05627106 0.12883037 0.06766193
> # статистика Колмогорова
> max(D)
[1] 0.1508281
```

```
> # критическое значение статистики
> alpha=0.09208
> c=2*pi/(pi-2)
> c
[1] 5.503877
> d=sqrt(log(2*sqrt(c)/alpha)/c)
> d
[1] 0.8451145
> d/(sqrt(10)-0.01+0.85/sqrt(10))
[1] 0.2470321
>
```

Таким образом, в соответствии с критерием согласия Колмогорова для данной выборки получена статистика Колмогорова $D_n^* = 0.1508281$ и критическое значение статистики $D_{n,\alpha}^{(\mu,\sigma)} = 0.2470321$. Имеем $D_n^* < D_{n,\alpha}^{(\mu,\sigma)}$, поэтому распределение можно считать нормальным.

Покажем, как вычислить статистику Колмогорова для данной выборки `data` с использованием встроенной функции `ks.test` (Kolmogorov-Smirnov test).

```
> ks.test(data, "pnorm", mean=mu, sd=sigma)
One-sample Kolmogorov-Smirnov test
data:  data
D = 0.1508, p-value = 0.9768
>
```

Данная проверка показывает, что статистика Колмогорова вычислена, по-видимому, верно. Заметим однако, что критерий Колмогорова в данном случае применен при использовании *оценок* параметров распределения по самой выборке (как это указано в [1, с. 12]). Поскольку выборка была порождена при иных параметрах нормального распределения на основе таблицы (см. условие задачи), то представляется интересным произвести проверку нормальности выборки, имея в виду изначальные параметры распределения.

```
> mu=28.24
> sigma=66.12976
> ks.test(data, "pnorm", mean=mu, sd=sigma)
One-sample Kolmogorov-Smirnov test
data:  data
D = 0.2841, p-value = 0.3948
>
```

Как видим, вопрос о нормальности выборки в таком случае стоит более остро: выборку уже нельзя считать нормальной, хотя статистика Колмогорова все же близка к критическому значению (т. е. сохраняется высокая неопределенность).

Ответ: численные значения приведены в решении.

12 Доверительный интервал вероятности брака

Задача. В выборке объема n из N изделий оказалось m бракованных. Построить доверительный интервал для вероятности брака p, считая распределение числа обнаруженных бракованных изделий биномиальным. Величину γ взять из задачи 11; $n = r[100 + 200\,KV]$; $m = r[11 + 60\,KV]$.

Решение. Точечная оценка вероятности брака $p^* = \frac{m}{n}$. В грубом приближении границы доверительного интервала

$$p^* \pm \psi_{0.5(1+\gamma)} \sqrt{\frac{p^*(1-p^*)}{n}}. \tag{12.1}$$

Более точным является доверительный интервал по Муавру–Лапласу:

$$\frac{p^* + \dfrac{\psi_q^2}{2n} \pm \psi_q \sqrt{\dfrac{p^*(1-p^*)}{n} + \dfrac{\psi_q^2}{4n^2}}}{1 + \dfrac{\psi_q^2}{n}}, \quad q = 0.5(1+\gamma). \tag{12.2}$$

Вычислим границы этих интервалов в системе R. Имеем $n = r[100 + 200 \cdot 0.912] = r[282.4] = 282$, $m = r[11 + 60 \cdot 0.912] = r[65.72] = 66$, $\gamma = 0.90792$.

```
> n=282
> m=66
> p=m/n
> p
[1] 0.2340426
> gamma=0.90792
> q=0.5*(1+gamma)
> q
[1] 0.95396
> psi=qnorm(q)
> psi
[1] 1.684526
> plusminus=c(-1,1)
> # грубое приближение
> p+plusminus*psi*sqrt(p*(1-p)/n)
[1] 0.1915705 0.2765146
> # более точное приближение
> (p+psi^2/2/n+plusminus*psi*sqrt(p*(1-p)/n+psi^2/4/n^2))/(1+psi^2/n)
[1] 0.1943492 0.2790350
>
```

Получены доверительные интервалы $[0.1915705, 0.2765146]$ и $[0.1943492, 0.2790350]$ в грубом и более точном приближениях соответственно.

Подумаем, каким образом можно вычислить доверительный интервал более точно. Требуется найти δ из соотношения $\Pr\left(\left|\frac{m}{n} - p\right| < \delta\right) = \gamma$. Домножим выражение в скобках на n: $\Pr\left(|m - np| < \delta n\right) = \gamma$. Обозначим $l = \delta n$, $k = np$, тогда получим условие $\Pr(|m-k| < l) = \gamma$ или $\Pr(m - l < k < m + l) = \gamma$, которое выражает вероятность попадания случайной величины K, распределенной по биномиальному закону, в интервал $[m - l, m + l]$. Другими словами, для отыскания l (и, следовательно, δ) требуется решить уравнение

$$f(l) = \sum_{k=m-l}^{m+l} P_n(k) = \gamma, \tag{12.3}$$

вернее, подобрать такое значение l, при котором равенство (12.3) выполняется наиболее точно. Заметим, что K является дискретной случайной величиной, поэтому равенство (12.3), вообще говоря, не может выполняться точно. Будем искать такое значение l, что $f(l-1) < \gamma$ и $f(l) \geqslant \gamma$.

```
> f <- function(l) sum(dbinom(seq(m-l,m+l), n, p))
> gamma
[1] 0.90792
> f(10)
[1] 0.8606352
```

```
> f(11)
[1] 0.894611
> f(12)
[1] 0.9216622
> 11/n
[1] 0.03900709
> 12/n
[1] 0.04255319
> p+plusminus*11/n
[1] 0.1950355 0.2730496
> p+plusminus*12/n
[1] 0.1914894 0.2765957
>
```

Таким образом, найдено значение $l = 12$, $\delta = 0.04255319$ и доверительный интервал $[0.1914894, 0.2765957]$ с вероятностью, которая несколько больше γ.

Ответ: доверительный интервал $[0.1915, 0.2766]$ с вероятностью примерно 0.9217.

13 Проверка гипотез при статистическом контроле качества

Задача. A. Пусть $k = KV$. Для $\alpha = 0.01 + 0.14k$, $p_0 = 0.3 - 0.2k$ найти $n_{\min}$ при статистическом контроле качества. Для $n = 2n_{\min} + 1$ сформулировать правило проверки гипотезы $H_0 : p \geqslant p_0$ против H_1: $p < p_0$.

B. Сформулировать правило проверки тех же гипотез при тех же α и p_0, если n не задано, а даны $\beta = 2$ и $p_1 = 0.75p_0$.

C. Для заданий из предыдущих пунктов построить график функции мощности критерия по 6 точкам: $p = 0$, $p = 0.5p_1$, $p = p_1$, $p = p_0$, $p = 0.5(1 + p_0)$, $p = 1$.

Поясним условие задачи [1, с. 12–13]. Имеется партия из N изделий, из которых K бракованных. Производится случайная выборка объемом n изделий, из которых вследствие тщательной проверки установлено k бракованных. Необходимо сделать некоторые выводы относительно вероятности брака $p = K/N$ во всей партии.

Покупатель устанавливает допустимую долю брака p_0, с которой (или меньшей) он готов примириться. Выдвигаются гипотезы H_0: $p \geqslant p_0$ (основная гипотеза) и H_1: $p < p_0$. Требуется установить критическое число m бракованных деталей, так что при $k \leqslant m$ принимается гипотеза H_1, а при $k > m$ — гипотеза H_0.

Ошибка первого рода совершается, когда $k \leqslant m$ и принимается гипотеза H_1, а на самом деле истинна H_0. Считая k реализацией некоторой случайной величины η, ограничим ошибку первого рода:

$$\Pr(\eta \leqslant m;\, p \geqslant p_0) \leqslant \alpha. \tag{13.1}$$

Если $n \leqslant 0.1N$, то можно с хорошим приближением полагать, что величина η имеет биномиальное распределение

$$\Pr(\eta = k) = C_n^k p^k (1-p)^{n-k}. \tag{13.2}$$

Тогда фактическая вероятность ошибки первого рода $\alpha_f(p) = \Pr(\eta \leqslant m;\, p \geqslant p_0) = \sum_{k=0}^{m} Pr(\eta = k)$ и ограничение (13.1) принимает вид

$$\sum_{k=0}^{m} C_n^k p^k (1-p)^{n-k} \leqslant \alpha, \quad p \geqslant p_0. \tag{13.3}$$

Функция $\alpha_f(p)$ монотонно убывает; максимум достигается при $p = p_0$: $\max \alpha_f(p) = \alpha_f(p_0)$. Следовательно, для ограничения ошибки первого рода получаем неравенство

$$\sum_{k=0}^{m} C_n^k p_0^k (1-p_0)^{n-k} \leqslant \alpha. \tag{13.4}$$

При $(1-p_0)^n > \alpha$ это неравенство не выполняется. Существует минимальное значение

$$n_{\min} = \frac{\ln\alpha}{\ln(1-p_0)}, \tag{13.5}$$

такое, что лишь при $n \geqslant n_{\min}$ возможна проверка гипотез с ошибкой первого рода, не большей α. Для нахождения критического значения m получаем систему, состоящую из (13.4) и неравенства

$$\sum_{k=0}^{m+1} C_n^k p_0^k (1-p_0)^{n-k} > \alpha. \tag{13.6}$$

Ошибка второго рода совершается, когда $k > m$ и принимается гипотеза H_0, а на самом деле истинна H_1. Фактическая вероятность ошибки второго рода

$$\beta_f(p) = \Pr(\eta > m;\, p < p_0) = \sum_{k=m+1}^{n} C_n^k p^k (1-p)^{n-k} = 1 - \sum_{k=0}^{m} C_n^k p^k (1-p)^{n-k}. \tag{13.7}$$

Функция $\beta_f(p)$ монотонно возрастает и $\lim_{p\to p_0}\beta_f(p) = 1-\alpha_f(p)$. Поскольку $\alpha_f(p_0)$ — величина малая, то верхняя грань $\beta_f(p)$ всегда близка к 1 и ограничить ее невозможно. На практике вводится интервал (p_1, p_0) — зона безразличия к ошибке второго рода, $p_1 = (0.75\ldots0.9)p_0$. Ограничение вероятности ошибки второго рода производится в точке p_1, что приводит к неравенству

$$1 - \sum_{k=0}^{m} C_n^k p_1^k (1-p_1)^{n-k} \leqslant \beta. \tag{13.8}$$

Решение. По условию имеем $\alpha = 0.01 + 0.14\cdot0.912 = 0.22768$, $p_0 = 0.3 - 0.2\cdot0.912 = 0.1176$.

А. Требуется найти минимальный объем выборки $n_{\min}$, при котором возможна проверка основной гипотезы с допустимой долей брака $p_0 = 0.1176$ и вероятностью ошибки первого рода не более $\alpha = 0.22768$. В силу (13.5) получаем

$$n_{\min} = \left\lceil \frac{\ln 0.22768}{\ln(1-0.1176)} \approx \frac{-1.4798}{-0.12511} \approx 11.828 \right\rceil = 12.$$

Далее, для $n = 2n_{\min}+1 = 25$ требуется сформулировать правило проверки основной гипотезы, т. е. указать критическое число m бракованных деталей, такое что при $k > m$ принимается гипотеза H_0 (высокая доля брака).

В силу (13.4) и (13.6) требуется подобрать такое значение m, что при заданных n, p_0 и α соответствующие неравенства выполняются. Другими словами, будем рассматривать функцию биномиального распределения как функцию вероятности от числа m бракованных деталей при заданных параметрах n и p_0: $f(m) = \sum_{k=0}^{m}\Pr(\eta = k)$. Тогда требуется подобрать такое m, что $f(m) \leqslant \alpha$ и $f(m+1) > \alpha$.

Выполним расчеты в системе R.

```
> n=25
> p=0.1176
> f <- function(m) sum(dbinom(seq(0,m), n, p))
> f(0)
[1] 0.04381648
> f(1)
[1] 0.1898052
> f(2)
[1] 0.4232813
```

Таким образом, найдено значение $m = 1$. Это позволяет сформулировать правило проверки гипотез: если в случайной выборке объемом $n = 25$ изделий количество бракованных изделий $m > 1$, тогда следует принять гипотезу H_0 (высокая доля брака); если же $m = 0$ или $m = 1$, тогда следует принять гипотезу H_1 (приемлемая доля брака).

B. Требуется сформулировать правила для гипотез при $\alpha = 0.22768$, $p_0 = 0.1176$, $\beta = 2\alpha = 0.45536$ и $p_1 = 0.75 \cdot 0.1176 = 0.0882$. Другими словами, требуется указать такие значения n (объем выборки) и m (критическое число бракованных деталей), что проверка гипотез возможна при заданных ограничениях на вероятность обишки первого рода α и второго рода β.

В силу (13.4), (13.6) и (13.8) необходимо подобрать минимальное значение n, для которого существует значение m, при котором соответствующие неравенства выполняются.

Вычислим вероятность ошибки второго рода для ранее построенной конфигурации ($n = 25$, $m = 1$).

```
> p1=0.0882
> m=1
> g <- function(m) 1-sum(dbinom(seq(0,m), n, p1))
> g(m)
[1] 0.942305
```

Т. е. получено значение $\beta_f(p_1) = 0.942305$, которое превышает допустимую вероятность ошибки второго рода. Чтобы уменьшить β_f, нужно увеличить m, но тогда будет превышена вероятность ошибки первого рода α_f. Следовательно, необходимо составить новую конфигурацию с бóльшим объемом выборки.

Рациональным представляется увеличить объем выборки вдвое: $n = 50$ (в дальнейшем при уменьшении объема выборки можно будет воспользоваться методом «половинного деления»).

```
> n=50
> f(1)
[1] 0.01471331
> f(2)
[1] 0.05648622
> f(3)
[1] 0.1455614
> f(4)
[1] 0.2850491
> # для n=50 найдено значение m=3
> m=3
> g(m)
[1] 0.653634
```

Т. е. получена новая конфигурация $n = 50$, $m = 3$, $\alpha_f \approx 0.145$, $\beta_f \approx 0.654$. Увеличим объем выборки до $n = 100$:

```
> n=100
> f(5)
[1] 0.01801509
> f(6)
[1] 0.04263555
> f(7)
[1] 0.08669792
> f(8)
[1] 0.1549636
> f(9)
[1] 0.2479651
> m=8
> g(m)
[1] 0.5255059
```

Т. е. получена новая конфигурация $n = 100$, $m = 8$, $\alpha_f \approx 0.155$, $\beta_f \approx 0.525$. Увеличим объем выборки до $n = 200$:

```
> n=200
> f(17)
[1] 0.08903707
> f(18)
[1] 0.1335288
> f(19)
[1] 0.1903275
> f(20)
[1] 0.2588335
> m=19
> g(m)
[1] 0.3119002
```

Т. е. получена новая конфигурация $n = 200$, $m = 19$, $\alpha_f \approx 0.190$, $\beta_f \approx 0.311$. Она удовлетворяет всем ограничениям задачи; однако, вероятно, можно указать конфигурацию с меньшим объемом выборки, которая также удовлетворяет ограничениям задачи (и поэтому практически предпочтительнее). Используя метод «половинного деления», положим $n = (100 + 200)/2 = 150$.

```
> n=150
> f(13)
[1] 0.1457985
> f(14)
[1] 0.2164836
> f(15)
[1] 0.3018953
> m=14
> g(m)
[1] 0.3446922
```

Т. е. получена новая конфигурация $n = 150$, $m = 14$, $\alpha_f \approx 0.216$, $\beta_f \approx 0.345$. Значение β_f, по-видимому, может быть увеличено, поэтому следует составить конфигурацию с объемом выборки $n = 125$ и т. д.

Короче, назрела необходимость в составлении алгоритма, который по заданным ограничениям $\langle \alpha, \beta, p_0, p_1 \rangle$ методом «половинного деления» подбирает оптимальную конфигурацию $\langle n, m \rangle$, которую можно рекомендовать для осуществления выборочного контроля на практике.

```
# вычисление <n,m> при заданных <alpha,beta,p0,p1>
nm <- function (alpha, beta, p0, p1) {
  # начальные объемы выборки
  n_min <- 0
  n_max <- as.integer(log(alpha)/log(1-p0))*100+2
  # специально для проверки расчетов задачи 13
  n_min <- 0
  n_max <- 200
  # вероятность ошибки первого рода
  f <- function(m,n)   sum(dbinom(seq(0,m), n, p0))
  # вероятность ошибки второго рода
  g <- function(m,n) 1-sum(dbinom(seq(0,m), n, p1))
  while ((n_max-n_min)>1) {
```

```
    # половинное деление
    n <- as.integer((n_max+n_min)/2)
    # подбор значения m
    for (i in 0:n) {
      # ошибка первого рода
      v <- f(i,n)
      if (v <= alpha) {
        af <- v
        m <- i
      } else {
        break
      }
    }
    # ошибка второго рода
    bf <- g(m,n)
    # является ли полученная конфигурация допустимой
    ok <- (bf <= beta)
    # отладочная печать
    cat("ok=", ok, "n_max=", n_max, "n_min=", n_min, "\t",
      "n=", n, "m=", m, "af=", af, "bf=", bf, "\n")
    if (ok) {
      n_max <- n
      n_ok <- n
      m_ok <- m
    } else {
      n_min <- n
    }
  }
  # возвращаемое значение
  c(n_ok, m_ok)
}
```

В результате запуска алгоритма получаем:

```
> nm(0.22768, 0.45536, 0.1176, 0.0882)
ok=FALSE  n_max=200 n_min=0    n=100 m=8  af=0.1549636 bf=0.5255059
ok=TRUE   n_max=200 n_min=100  n=150 m=14 af=0.2164836 bf=0.3446922
ok=TRUE   n_max=150 n_min=100  n=125 m=11 af=0.1889490 bf=0.423774
ok=TRUE   n_max=125 n_min=100  n=112 m=10 af=0.2209246 bf=0.4009423
ok=FALSE  n_max=112 n_min=100  n=106 m=9  af=0.1874455 bf=0.460568
ok=FALSE  n_max=112 n_min=106  n=109 m=9  af=0.1616800 bf=0.4965753
ok=FALSE  n_max=112 n_min=109  n=110 m=9  af=0.1537297 bf=0.508444
ok=FALSE  n_max=112 n_min=110  n=111 m=9  af=0.1460893 bf=0.5202302
[1] 112  10
```

Т. е. найдена оптимальная конфигурация $n = 112$, $m = 10$, $\alpha_f = 0.2209246$, $\beta_f = 0.4009423$.

Однако вопрос об оптимальной конфигурации представляется более тонким. Если удалить из нашего алгоритма строки с комментарием «специально для проверки расчетов…», то результат будет уже другим:

```
ok=TRUE   n_max=1102 n_min=0   n=551 m=58 af=0.2038955  bf=0.07141509
ok=TRUE   n_max=551  n_min=0   n=275 m=27 af=0.1834767  bf=0.2402702
ok=TRUE   n_max=275  n_min=0   n=137 m=12 af=0.1694830  bf=0.433942
```

```
ok=FALSE  n_max=137  n_min=0   n=68  m=5  af=0.17456    bf=0.5613751
ok=TRUE   n_max=137  n_min=68  n=102 m=9  af=0.2264332  bf=0.4120014
ok=FALSE  n_max=102  n_min=68  n=85  m=7  af=0.2037736  bf=0.4782612
ok=TRUE   n_max=102  n_min=85  n=93  m=8  af=0.2210861  bf=0.4369451
ok=FALSE  n_max=93   n_min=85  n=89  m=7  af=0.1644714  bf=0.5314812
ok=FALSE  n_max=93   n_min=89  n=91  m=7  af=0.1471885  bf=0.5573228
ok=FALSE  n_max=93   n_min=91  n=92  m=7  af=0.1391102  bf=0.5700133
[1] 93  8
```

Т. е. найдена «более оптимальная» конфигурация $n = 93$, $m = 8$, $\alpha_f = 0.2210861$, $\beta_f = 0.4369451$, в которой α_f и β_f ближе к заданным параметрам α и β. Это говорит о том, что метод «половинного деления» в данном случае применять нельзя.

Тогда вместо подбора n методом «половинного деления» проще всего реализовать полный перебор: $n = 1, 2, 3, \ldots$ После несложной модификации алгоритма получена следующая «более оптимальная» конфигурация: $n = 83$, $m = 7$, $\alpha_f = 0.2259005$, $\beta_f = 0.4510817$, которую окончательно следует признать оптимальной. (Проверка: при $n = 82$ имеем либо $m = 7$, $\alpha_f = 0.1381169$, $\beta_f = 0.5925415 > \beta$, либо $m = 7$, $\alpha_f = 0.2375997 > \alpha$, $\beta_f = 0.437398$.)

C. Требуется построить график функции мощности критерия

$$M(p) = \sum_{k=0}^{m} C_n^k p^k (1-p)^{n-k}, \quad 0 \leqslant p \leqslant 1 \tag{13.9}$$

при $n = 83$, $m = 7$, по шести точкам точкам $p = 0$, $p = 0.5 \cdot 0.0882 = 0.0441$, $p = 0.0882$, $p = 0.1176$, $p = 0.5 \cdot (1 + 0.1176) = 0.5588$, $p = 1$. Имеем $M(0) = 1$, $M(0.0441) = 0.9699053$, $M(0.0882) = 0.5489183$, $M(0.1176) = 0.2259005$, $M(0.5588) =$ `7.469659e-20`, $M(1) = 0$. График функции мощности приведен на рис. 13.1. Заметим, что $M(0.2) = 0.003462594$.

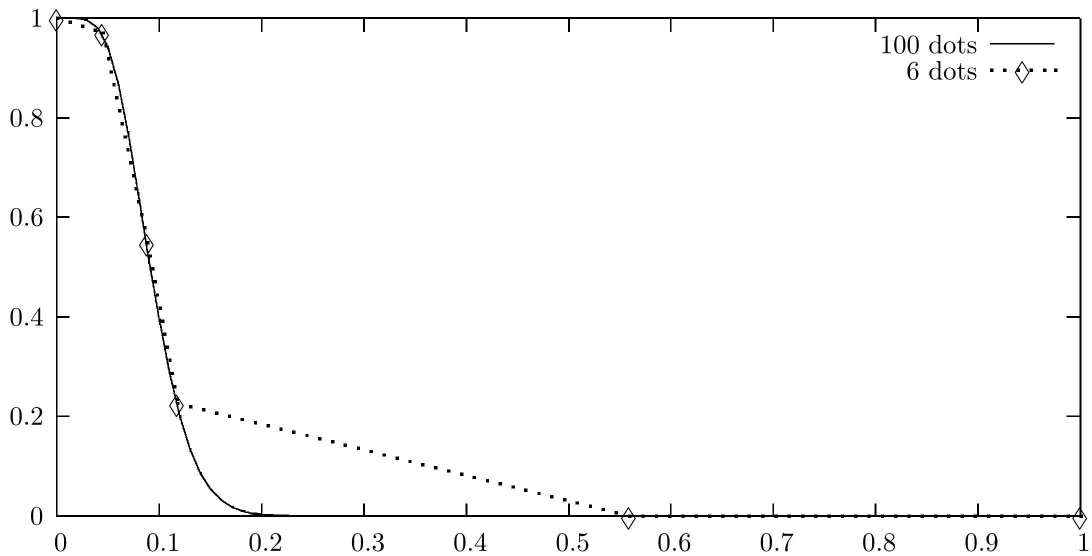

Рис. 13.1. Функция мощности критерия.

Ответ: Минимальный объем выборки $n = 12$; при заданных ограничениях на вероятность ошибок первого и второго рода следует делать контрольную выборку объемом $n = 83$ детали, при наличии в выборке $m > 7$ бракованных деталей считать партию бракованной.

14 Исследование зависимостей

Задача. На основе таблицы из задачи 11 получить выборку $\varepsilon_i = 0.1\sigma\sqrt{-2\ln R_i}\cos(2\pi R_{i+1})$, $\varepsilon_{i+1} = 0.2\sigma\sqrt{-2\ln R_i}\sin(2\pi R_{i+1})$, $i = 3, 5, 7, 9$. Составить выборку $y_i = (2\,KV + 1)(x_i + 2 + \varepsilon_{i+2})$, $i = 1, 2, \ldots$ (значений y_i получится на 2 меньше, чем x_i). Перенумеровать значения x_i, положив $x_i = x_{i+2}$, $i = 1, 2, \ldots, 8$. Полученное множество пар (x_i, y_i), $i = 1, 2, \ldots, 8$ считать реализациями двумерного случайного вектора (ξ, η). Построить диаграмму рассеяния, проверить гипотезу некоррелированности; с помощью метода наименьших квадратов построить простейшую регрессионную модель, оценить ее качество, проверить гипотезу об отсутствии линейной зависимости, нанести линию регрессии на диаграмму рассеяния. Проделать то же самое, построив квадратичную функцию регрессии $y^* = b_0 + b_1 x + b_2 x^2$ (исключая проверку гипотезы $b_1 = 0$). Выполнить то же самое, изменив y_i следующим образом: $y_i = y_i + (1 + 10\,KV)i + (1 + 4\,KV)i^2$. Для этой вновь полученной парной выборки построить также квадратичную функцию регрессии $y^* = b_0 + b_1 x + b_2 x^2$.

Решение. Расчеты целесообразно провести в системе R.

```
> # параметры задачи 11
> R
 [1] 0.54270 0.38150 0.29120 0.40180 0.84340 0.23510 0.21511 0.07900 0.95690
[10] 0.97860
> # параметры задачи 9
> mu
[1] 28.24
> sigma
[1] 66.12976
> # создание выборки "эпсилон"
> eps=array(0,10)
> eps
 [1] 0 0 0 0 0 0 0 0 0 0
> i=c(3,5,7,9)
> eps[i]=0.1*sigma*sqrt(-2*log(R[i]))*cos(2*pi*R[i+1])
> # выборка eps заполнена частично
> eps
 [1]  0.0000000  0.0000000 -8.4724434  0.0000000  0.3608017  0.0000000
 [7] 10.1938515  0.0000000  1.9452641  0.0000000
> # работа с наборами значений в системе R хорошо продумана
> eps[i+1]=0.2*sigma*sqrt(-2*log(R[i]))*sin(2*pi*R[i+1])
> # выборка eps заполнена полностью
> eps
 [1]  0.0000000  0.0000000 -8.4724434 12.0207518  0.3608017  7.6853038
 [7] 10.1938515 11.0419913  1.9452641 -0.5262966
> # повторно построим выборку x[i] из задачи 11
> i=c(1,3,5,7,9)
> x=array(0,10)
> x[i]=mu+sigma*sqrt(-2*log(R[i]))*cos(2*pi*R[i+1])
> # выборка x построена частично
> x
 [1] -25.52747   0.00000 -56.48443   0.00000  31.84802   0.00000 130.17851
 [8]   0.00000  47.69264   0.00000
> x[i+1]=mu+sigma*sqrt(-2*log(R[i]))*sin(2*pi*R[i+1])
> # выборка x построена полностью
> x
```

```
 [1] -25.52747  77.78574 -56.48443  88.34376  31.84802  66.66652 130.17851
 [8]  83.44996  47.69264  25.60852
> # (результаты совпадают с расчетами на Паскале, см. задачу 11)
> # построим теперь выборку y
> i=1:8
> y=array(0,8)
> KV=0.912
> y[i]=(2*KV+1)*(x[i]+2+eps[i+2])
> y
[1]  -90.36775  259.26152 -152.84514  276.83407  124.37424  225.09683  378.76555
[8]  239.82441
> # перенумеруем элементы x
> x=array(x[i+2], 8)
> x
[1] -56.48443  88.34376  31.84802  66.66652 130.17851  83.44996  47.69264
[8]  25.60852
> # случайный вектор (x,y)
> xy=cbind(x,y)
> xy
             x          y
[1,] -56.48443  -90.36775
[2,]  88.34376  259.26152
[3,]  31.84802 -152.84514
[4,]  66.66652  276.83407
[5,] 130.17851  124.37424
[6,]  83.44996  225.09683
[7,]  47.69264  378.76555
[8,]  25.60852  239.82441
```

Таким образом, получено множество пар (x_i, y_i): $i = 1$: $(-56.48443, -90.36775)$; $i = 2$: $(88.34376, 259.26152)$; $\ldots$; $i = 8$: $(25.60852, 239.82441)$.

Диаграмма рассеяния построена на рис. 14.1 — точки (x_i, y_i) нанесены на координатную плоскость.

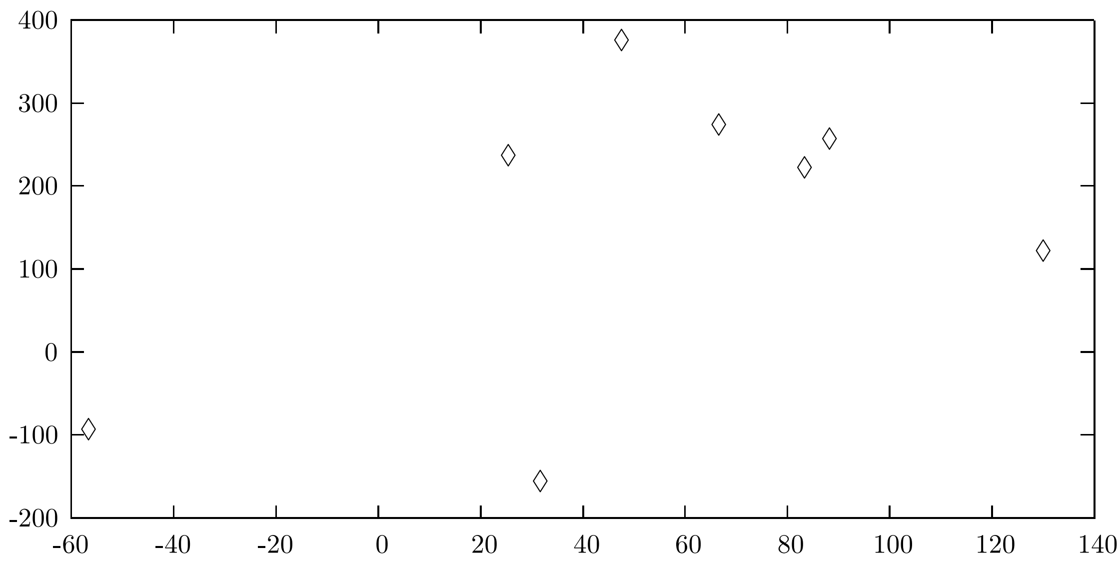

Рис. 14.1. Диаграмма рассеяния.

Проверим гипотезу некоррелированности. Для этого вычислим выборочный коэффициент корреляции

$$r_{xy} = \frac{\sum_{i=1}^{n} \frac{x_i y_i}{n} - \bar{x}\bar{y}}{s_x s_y}, \qquad (14.1)$$

где $\bar{x}$, $\bar{y}$ — выборочные средние, s_x, s_y — выборочные с. к. о. величин ξ, η.

```
> (sum(x*y/8)-mean(x)*mean(y))/sqrt(var(x)*var(y))
[1] 0.4228127
> corr(xy)
[1] 0.4832145
```

Получено значение $r_{xy} = 0.4228127$. *Замечание.* Здесь `var(x)` вычисляет «исправленную» выборочную дисперсию, которая, как известно, несколько больше «неисправленной». Однако встроенная функция вычисления коэффициента корреляции `corr` использует «неисправленную» дисперсию. Тогда встает вопрос, подлежит ли «исправлению» выборочная ковариация в числителе [5, с. 48]. Другими словами, имеется некоторая путаница в обозначениях.

Проверка гипотезы некоррелированности H_0: $\rho_{\xi\eta} = 0$ против гипотезы H_1: $\rho_{\xi\eta} \neq 0$ производится с помощью неравенства

$$|r_{xy}| > \frac{t_{n-2;1-0.5\alpha}}{\sqrt{n-2+t^2_{n-2;1-0.5\alpha}}}, \qquad (14.2)$$

где $t_{m,q}$ — квантиль распределения Стьюдента порядка q с m степенями свободы. Если это неравенство выполняется, принимается гипотеза H_1, в противном случае H_0.

Возьмем значение $\alpha = 0.09208$ из задачи 11.

```
> alpha=0.09208
> n=8
> t=qt(1-0.5*alpha,n-2)
> t
[1] 2.002709
> t/sqrt(n-2+t*t)
[1] 0.6329691
```

Таким образом, $0.4228127 \not> 0.6329691$, поэтому следует принять гипотезу H_0: величины некоррелированы. *Замечание 1.* Высокое критическое значение в данном случае обусловлено малым объемом выборки; можно показать, что при $n \to \infty$ правая часть (14.2) стремится к нулю. *Замечание 2.* С другой стороны, коэффициент корреляции является мерой *линейной* корреляционной связи; т. е. по (14.2) нельзя судить о возможной нелинейной зависимости между случайными величинами. Для оценки произвольной корреляционной связи используется *корреляционное отношение* [2, с. 353].

Переходим к построению простейшей (линейной) регрессионной модели. Требуется определить коэффициенты b_0, b_1 уравнения регрессии

$$y^* = b_0 + b_1 x. \qquad (14.3)$$

Регрессия по методу наименьших квадратов дает следующие значения (оценки) параметров:

$$b_1 = \frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2}, \quad b_0 = \bar{y} - b_1\bar{x}. \qquad (14.4)$$

```
> b1=(sum(x*y)/n-mean(x)*mean(y))/(sum(x*x)/n-mean(x)^2)
> b1
[1] 1.626011
> b0=mean(y)-b1*mean(x)
> b0
[1] 72.80046
```

Получены значения $b_0 = 72.80046$, $b_1 = 1.626011$.

Проверка. Воспользуемся встроенной функцией `lm` для вычисления параметров линейной модели:

```
> lm(y ~ x)
(Intercept)             x
     72.800         1.626
>
```

Линия регрессии изображена на рис. 14.2.

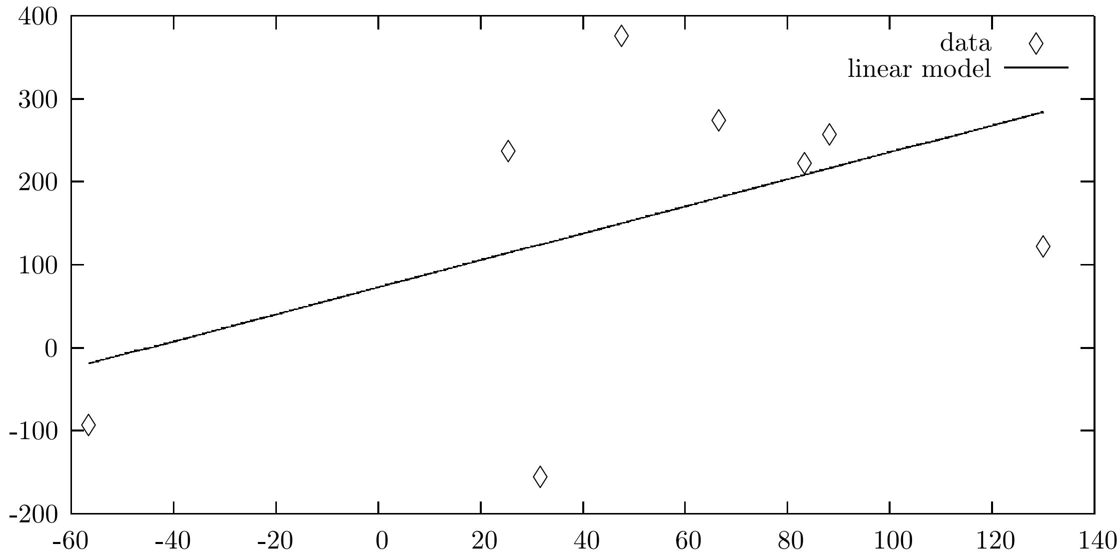

Рис. 14.2. Линия регрессии.

Для оценка качества аппроксимации экспериментальных данных выборочной функцией регрессии вычисляется *коэффициент детерминации*

$$R^2 = 1 - \frac{\sum(y_i - y_i^*)^2}{\sum(y_i - \bar{y})^2}. \tag{14.5}$$

Здесь $y_i^* = y^*(x_i)$. При $0.5 \leqslant R^2 < 1$ качество аппроксимации считается хорошим, при $0.35 < R^2 < 0.5$ — удовлетворительным, при $R^2 < 0.35$ — плохим.

```
> fy <- function(x) b1*x+b0
> R2=1-sum((y-fy(x))^2)/sum((y-mean(y))^2)
> R2
[1] 0.2334962
```

Получено значение $R^2 = 0.2334962$, при котором аппроксимацию следует считать неудовлетворительной.

Проверим гипотезу об отсутствии линейной зависимости H_0: $b_1 = 0$ против гипотезы о наличии некоторой зависимости H_1: $b_1 \neq 0$. Гипотеза H_0 принимается, если доверительный интервал

$$b_1 \pm t_{n-2;1-0.5\alpha} \frac{s}{\sqrt{Q_x}}, \quad \text{где} \;\; Q_x = \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s^2 = \sum_{i=1}^{n} \frac{(y_i - y_i^*)^2}{n-2} \tag{14.6}$$

накрывает 0.

```
> alpha=0.09208
> t=qt(1-0.5*alpha,n-2)
> t
[1] 2.002709
> Qx=sum((x-mean(x))^2)
> Qx
[1] 21526.78
> s2=sum((y-fy(x))^2)/(n-2)
> s2
[1] 31139.32
> sqrt(s2)/sqrt(Qx)
[1] 1.202721
> plusminus=c(-1,1)
> b1+plusminus*t*sqrt(s2)/sqrt(Qx)
[1] -0.7826903  4.0347121
```

Т. к. данный интервал накрывает 0, то следует считать справедливой гипотезу H_0: $b_1 = 0$.

Переходим к построению квадратичной регрессионной модели. Требуется определить коэффициенты b_0, b_1, b_2 уравнения регрессии

$$y^* = b_0 + b_1 x + b_2 x^2. \tag{14.7}$$

Система нормальных уравнений регрессии по методу наименьших квадратов принимает вид

$$\begin{cases} b_0 \sum x_i^0 + b_1 \sum x_i^1 + b_2 \sum x_i^2 = \sum x_i^0 y_i, \\ b_0 \sum x_i^1 + b_1 \sum x_i^2 + b_2 \sum x_i^3 = \sum x_i^1 y_i, \\ b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 = \sum x_i^2 y_i, \end{cases} \tag{14.8}$$

в данном случае

$$\begin{cases} 8\, b_0 + 417.3035\, b_1 + 43294.56\, b_2 = 1260.944, \\ 417.3035\, b_0 + 43294.56\, b_1 + 3750349\, b_2 = 100777.3, \\ 43294.56\, b_0 + 3750349\, b_1 + 433154678\, b_2 = 7504522. \end{cases}$$

Решение системы: $b_0 = 105.65$, $b_1 = 2.893$, $b_2 = -0.01828$ (все значащие цифры верные).

Проверка. Воспользуемся встроенной функцией `nlm(f,p)` для вычисления параметров нелинейной модели; здесь `f` — модель (функция от параметров модели, которую нужно минимизировать), `p` — вектор начальных приближений параметров модели.

```
> fn <- function(p) sum((y-p[3]*x^2-p[2]*x-p[1])^2)
> nlm(fn,p=c(1,1,1))
$minimum
[1] 154916.5
$estimate
[1] 105.65771419   2.89361367  -0.01828945
$gradient
[1] -3.029993e-06  6.034772e-05  9.313226e-04
$iterations
[1] 29
```

Проверим качество квадратичной модели.

```
> b0=105.65771419
> b1=2.89361367
```

```
> b2=-0.01828945
> fy <- function(x) b0+b1*x+b2*x^2
> R2=1-sum((y-fy(x))^2)/sum((y-mean(y))^2)
> R2
[1] 0.3644474
```

Получено значение $R^2 = 0.3644474$, при котором аппроксимацию можно считать условно удовлетворительной.

Кривая квадратичной регрессии изображена на рис. 14.3.

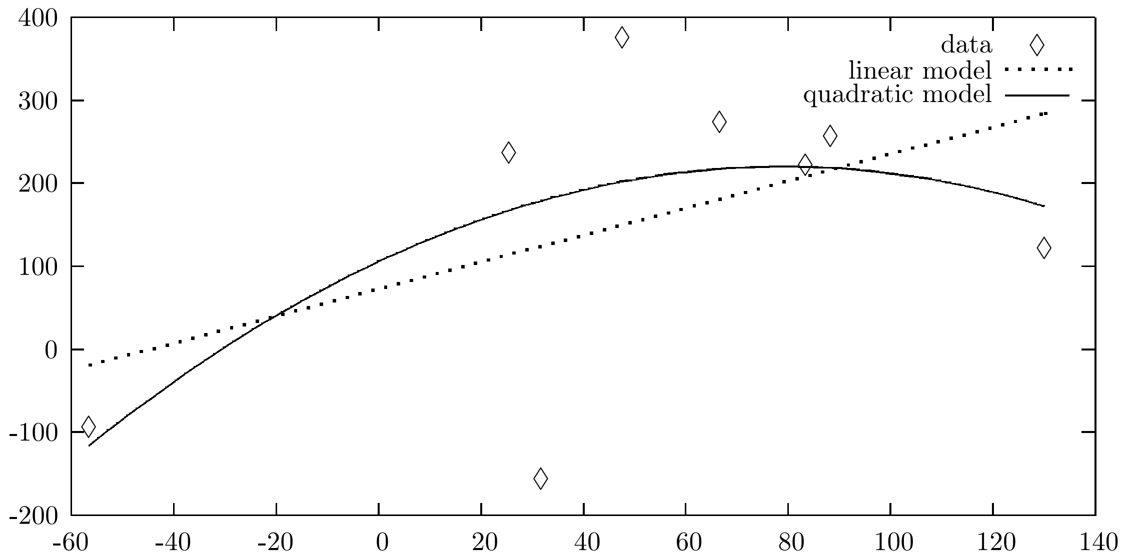

Рис. 14.3. Кривая квадратичной регрессии.

Теперь требуется построить новую выборку (случайный вектор), изменив y_i следующим образом: $y_i = y_i + (1 + 10\,KV)i + (1 + 4\,KV)i^2$.

```
> # текущие значения y[i]
> y
[1]  -90.36775  259.26152 -152.84514  276.83407  124.37424  225.09683  378.76555
[8]  239.82441
> i=1:8
> y[i]=y[i]+(1+10*KV)*i+(1+4*KV)*i^2
> # новые значения y[i]
> y
[1] -75.59975 298.09352 -80.65314 391.68207 291.17424 453.14483 677.35755
[8] 618.25641
> # новый случайный вектор
> xy=cbind(x,y)
> xy
             x         y
[1,] -56.48443 -75.59975
[2,]  88.34376 298.09352
[3,]  31.84802 -80.65314
[4,]  66.66652 391.68207
[5,] 130.17851 291.17424
[6,]  83.44996 453.14483
[7,]  47.69264 677.35755
[8,]  25.60852 618.25641
```

Диаграмма рассеяния изображена на рис. 14.4.

Вычислим коэффициент корреляции.

```
> (sum(x*y/8)-mean(x)*mean(y))/sqrt(var(x)*var(y))
[1] 0.3494671
> corr(xy)
[1] 0.399391
```

Коэффициент корреляции уменьшился.

Построим линейную регрессию и вычислим коэффициент детерминации.

```
> lm(y~x)
(Intercept)            x
    215.629        2.033
> fy <- function(x) 2.033*x+215.629
> R2=1-sum((y-fy(x))^2)/sum((y-mean(y))^2)
> R2
[1] 0.1595132
```

Получена линейная регрессия $y^* = 2.033x + 215.629$, коэффициент детерминации $R^2 = 0.1595132$ (качество аппроксимации ухудшилось). Линия регрессии нанесена на рис. 14.4.

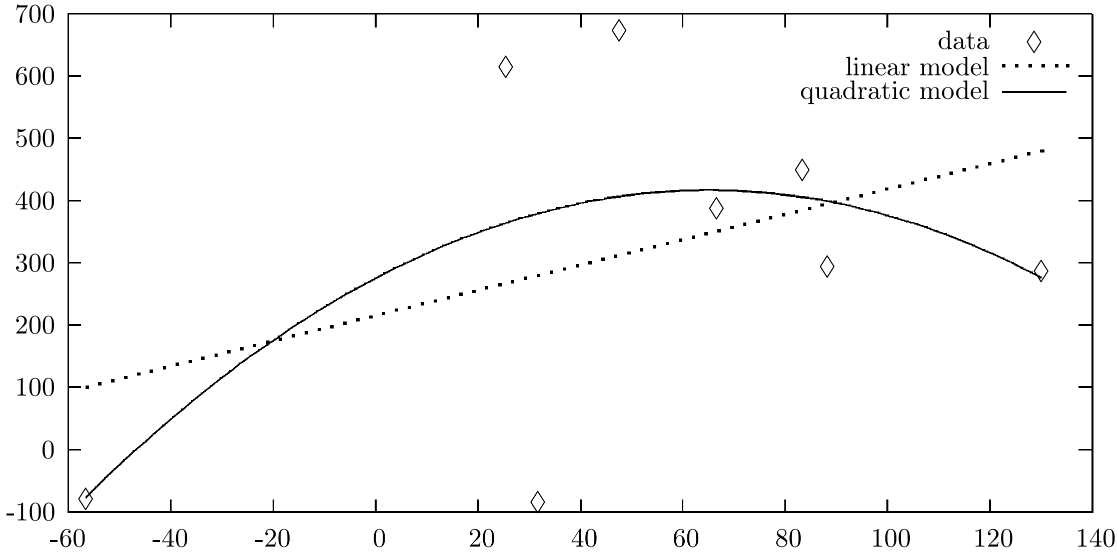

Рис. 14.4. Диаграмма рассеяния, линейная и нелинейная регрессии.

Построим теперь квадратичную регрессию и вычислим коэффициент детерминации.

```
> fn <- function(p) sum((y-p[3]*x^2-p[2]*x-p[1])^2)
> nlm(fn,p=c(1,1,1))
$minimum
[1] 362909.1
$estimate
[1] 275.48246232   4.34223359  -0.03331673
$gradient
[1] 9.112033e-06 2.959825e-04 3.585941e-02
$iterations
[1] 31
> fy <- function(x) 275.48246232+4.34223359*x-0.03331673*x^2
> R2=1-sum((y-fy(x))^2)/sum((y-mean(y))^2)
> R2
[1] 0.3494358
```

Получена квадратичная регрессия $y^* = 275.48 + 4.3422x - 0.033317x^2$, коэффициент детерминации остался приблизительно на прежнем уровне. Кривая регрессии нанесена на рис. 14.4.

Ответ: численные значения приведены в решении.

15 Группировка данных

Задача. После группировки данных получены следующие результаты: $x_{\min} = -2.3\sigma$, $x_{\max} = 2\sigma$, σ взять из задачи 9, объем выборки n из задачи 12. Найти

$$p_j^{\#} = \left(\oslash\left(\frac{u_j - \mu^{\#}}{\sigma}\right) - \oslash\left(\frac{u_{j-1} - \mu^{\#}}{\sigma}\right)\right)(1 + 0.1R_j), \quad \mu^{\#} = \frac{x_{\max} + x_{\min}}{2},$$

R_j — из задачи 11. При $r > 10$ положить $R_{10+j} = R_j$, $j \leqslant 10$. Найти n_j, округлив $np_j^{\#}$ до целого числа, $\sum n_j = n$. После этого пересчитать $p_j^{\#} = n_j/n$, составить таблицу группировки, найти числовые характеристики выборки по сгруппированным данным, включая поправку Шеппарда и сравнение ее с найденной выборочной дисперсией. Построить гистограмму, кривую, грубо оценивающую плотность распределения. В той же системе координат построить пунктиром график плотности нормального распределения с $\mu = \mu^{\#}$ и использованной выше σ.

Замечание. Обозначение $\oslash$ ранее не использовалось. Однако интуиция подсказывает, что $\oslash(x) \equiv \Phi(x)$, где $\Phi(x)$ — интегральная функция стандартного нормального распределения.

Решение. Выполним расчеты в системе R.

```
> # sigma из задачи 9
> sigma
[1] 66.12976
> xmin=-2.3*sigma
> xmax=2*sigma
> xmin; xmax
[1] -152.0984
[1] 132.2595
> mu=(xmin+xmax)/2
> mu
[1] -9.919464
> # n из задачи 12
> n=282
> # количество классов
> n^0.4
[1] 9.552117
> r=10
> # шаг группировки
> h=(xmax-xmin)/r
> h
[1] 28.43580
> # верхние границы классов
> j=1:r
> u=xmin+j*b
> u
 [1] -123.662651  -95.226854  -66.791058  -38.355261   -9.919464   18.516333
 [7]   46.952130   75.387926  103.823723  132.259520
```

```
> # R[j] из задачи 11
> R
 [1] 0.54270 0.38150 0.29120 0.40180 0.84340 0.23510 0.21511 0.07900 0.95690
 [10] 0.97860
> # найдем p[j]
> p=(pnorm((u-mu)/sigma)-pnorm((u-b-mu)/sigma))*(1+0.1*R)
> p
 [1] 0.02840057 0.05793823 0.09917546 0.14427640 0.18043654 0.17031429
 [7] 0.14168695 0.09713051 0.06114948 0.02957483
> sum(p)
[1] 1.010083
> # найдем n[j]
> nj=n*p
> nj
 [1]  8.008961 16.338580 27.967481 40.685944 50.883104 48.028631 39.955719
 [8] 27.390803 17.244154  8.340101
> sum(nj)
[1] 284.8435
> # округлим значения n[j]
> sum(round(nj))
[1] 284
> # потребуется немного уменьшить значения n[j], чтобы их сумма была n=282
> sum(round(nj*0.99))
[1] 282
> nj=round(nj*0.99)
> nj
 [1]  8 16 28 40 50 48 40 27 17  8
> # пересчитаем значения p[j]
> p=nj/n
> p
 [1] 0.02836879 0.05673759 0.09929078 0.14184397 0.17730496 0.17021277
 [7] 0.14184397 0.09574468 0.06028369 0.02836879
> sum(p)
[1] 1
> # середины (центры) классов
> X=(u+u-b)/2
> X
 [1] -137.880550 -109.444753  -81.008956  -52.573159  -24.137362    4.298434
 [7]   32.734231   61.170028   89.605825  118.041622
> # высота прямоугольников гистограммы
> e=p/h
> e
 [1] 0.0009976437 0.0019952875 0.0034917530 0.0049882186 0.0062352733
 [6] 0.0059858624 0.0049882186 0.0033670476 0.0021199929 0.0009976437
```

Результаты расчетов сведены в таблицу 15.1.

Табл. 15.1. Сгруппированные данные.

j	1	2	3	4	5
$[u_{j-1}, u_j)$	$[-152, -124)$	$[-124, -95.2)$	$[-95.2, -66.8)$	$[-66.8, -38.4)$	$[-38.4, -9.92)$
X_j	-137.88	-109.45	-81.01	-52.57	-24.14
n_j	8	16	28	40	50
$p_j^{\#}$	0.02837	0.05674	0.09929	0.14184	0.1773
e_j	0.0009976	0.001995	0.003491	0.004988	0.006235

j	6	7	8	9	10
$[u_{j-1}, u_j)$	$[-9.92, 18.5)$	$[18.5, 46.9)$	$[46.9, 75.4)$	$[75.4, 103.8)$	$[103.8, 132.3]$
X_j	4.298	32.73	61.17	89.61	118.04
n_j	48	40	27	17	8
$p_j^{\#}$	0.17021	0.14184	0.09574	0.06028	0.02837
e_j	0.005986	0.004988	0.003367	0.00212	0.0009976

Для вычисления выборочных характеристик по сгруппированным данным считается, что все данные, попавшие в j-й интервал, равны X_j, откуда

$$x_G^{\#} = \frac{1}{n}\sum_{j=1}^{r} n_j X_j = \sum_{j=1}^{r} p_j^{\#} X_j, \quad D_G^{\#} = \frac{1}{n-1}\sum_{j=1}^{r} n_j X_j^2 = \frac{1}{n-1}\sum_{j=1}^{r}(X_j - x_G^{\#})^2, \quad s_G = \sqrt{D_G^{\#}}. \tag{15.1}$$

```
> xG=sum(p*X)
> xG
[1] -9.919464
> DG=sum(nj*X^2)/(n-1)
> DG
[1] 3771.952
> sG=sqrt(DG)
> sG
[1] 61.41622
```

Таким образом, получены значения $x_G^{\#} = -9.919464$, $D_G^{\#} = 3771.952$, $s_G = 61.41622$.

Поправка Шеппарда для выборочной дисперсии

$$\Delta_{\text{Sh}} = \frac{h^2}{12}, \quad D_{\text{Sh}}^{\#} = D_G^{\#} - \Delta_{\text{Sh}}, \quad s_{\text{Sh}} = \sqrt{D_{\text{Sh}}^{\#}}. \tag{15.2}$$

Имеем $\Delta_{\text{Sh}} = \frac{28.43580^2}{12} = 67.38288$, $D_{\text{Sh}}^{\#} = 3771.952 - 67.38288 = 3704.569$, $s_{\text{Sh}} = \sqrt{3704.569} = 60.86517$, $\Delta_{\text{Sh}}/D_G^{\#} = 67.38288/3771.952 = 0.01786419$ (поправка Шеппарда относительно мала).

На рис. 15.1 изображена гистограмма, кривая плотности распределения, а также плотность нормального распределения при $\mu = -9.919464$, $\sigma = 66.12976$. Кривая распределения построена при помощи сплайн-интерполяции по точкам (X_j, e_j), дополнительно использованы точки $(x_{\min} - \frac{h}{2}, 0)$, $(x_{\max} + \frac{h}{2}, 0)$.

Ответ: численные значения приведены в решении.

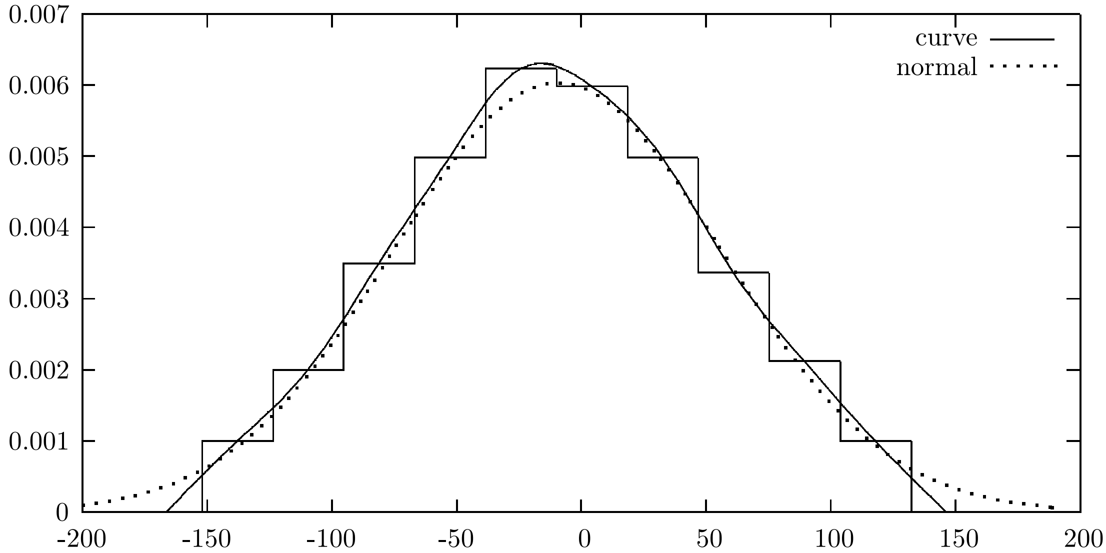

Рис. 15.1. Гистограмма, кривая плотности распределения и плотность нормального распределения.

16 Марковская цепь

Задача. Для однородной марковской цепи с тремя состояниями и дискретным временем найти двухшаговую и трехшаговую матрицы вероятностей перехода; вектор вероятностей состояний на первом, втором и третьем шагах; стационарное распределение вероятностей. Вероятности перехода p_{ij} взять равными произведениям $10\,\Pr(\xi = x_i, \eta = y_i)$ (из задачи 10), считая $i = 0, 1, 2$, $j = 0, 1$. Координаты вектора начальных вероятностей равны 1/3; вероятности третьего столбца матрицы P найти из условия нормировки по строкам. Изменить координаты вектора начальных вероятностей так, чтобы марковская цепь стала стационарной, для этой цепи найти коэффициент корреляции между соседними сечениями.

Решение. Будем также использовать в качестве начального индекса 1. Одношаговая матрица вероятностей перехода (на основе данных из табл. 10.1)

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 0.17024 & 0.25109 & p_{13} \\ 0.25755 & 0.33761 & p_{23} \\ 0.34638 & 0.42560 & p_{33} \end{pmatrix} = \begin{pmatrix} 0.17024 & 0.25109 & 0.57867 \\ 0.25755 & 0.33761 & 0.40484 \\ 0.34638 & 0.42560 & 0.22802 \end{pmatrix}.$$

Вычислим двухшаговую $(P^T)^2$ и трехшаговую $(P^T)^3$ матрицы переходов.

```
> P=rbind(c(0.17024,0.25109,0.57867),
+ c(0.25755,0.33761,0.40484),
+ c(0.34638,0.42560,0.22802))
> P
        [,1]    [,2]    [,3]
[1,] 0.17024 0.25109 0.57867
[2,] 0.25755 0.33761 0.40484
[3,] 0.34638 0.42560 0.22802
> # проверка условий нормировки
> sum(P)
[1] 3
> # операция транспонирования
> t(P)
```

```
        [,1]    [,2]    [,3]
[1,] 0.17024 0.25755 0.34638
[2,] 0.25109 0.33761 0.42560
[3,] 0.57867 0.40484 0.22802
> # умножение матриц обозначается %*%
> # двухшаговая матрица
> P2 = t(P) %*% t(P)
> P2
          [,1]      [,2]      [,3]
[1,] 0.2940896 0.2710252 0.2475626
[2,] 0.3737980 0.3509486 0.3277047
[3,] 0.3321124 0.3780261 0.4247327
> # трехшаговая матрица
> P3=t(P) %*% t(P) %*% t(P)
> P3
          [,1]      [,2]      [,3]
[1,] 0.2613746 0.2674668 0.2736643
[2,] 0.3413879 0.3474234 0.3535631
[3,] 0.3972375 0.3851097 0.3727726
```

Найдем векторы вероятности состояний на первом, втором и третьем шагах.

```
> # вектор начальных вероятностей
> p0=c(1/3,1/3,1/3)
> # вектор вероятностей на первом шаге
> p1=t(P) %*% p0
> p1
          [,1]
[1,] 0.2580567
[2,] 0.3381000
[3,] 0.4038433
> # вектор вероятностей на втором шаге
> p2=t(P) %*% p1
> p2
          [,1]
[1,] 0.2708925
[2,] 0.3508171
[3,] 0.3782904
> # проверка
> P2 %*% p0
          [,1]
[1,] 0.2708925
[2,] 0.3508171
[3,] 0.3782904
> # вектор вероятностей на третьем шаге
> p3=t(P) %*% p2
> p3
          [,1]
[1,] 0.2675019
[2,] 0.3474582
[3,] 0.3850399
```

Вектор стационарных состояний p_{st} определяется из соотношений

$$P^T p_{\mathrm{st}} = p_{\mathrm{st}}, \quad \sum p_{\mathrm{st}} = 1 \tag{16.1}$$

(изучить теорему о неподвижной точке [4, с. 414]), в данном случае из системы уравнений

$$\begin{cases} x_1 = p_{11}x_1 + p_{21}x_2 + p_{31}x_3, \\ x_2 = p_{12}x_1 + p_{22}x_2 + p_{32}x_3, \\ \qquad\qquad x_1 + x_2 + x_3 = 1, \end{cases}$$

или, в матричной форме,

$$\begin{pmatrix} p_{11}-1 & p_{12} & \dots & p_{1n} \\ p_{12} & p_{22}-1 & \dots & p_{2n} \\ \multicolumn{4}{c}{\dots\dots\dots\dots\dots} \\ 1 & 1 & \dots & 1 \end{pmatrix} p_{\mathrm{st}} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 1 \end{pmatrix}$$

Другими словами, нужно транспонировать матрицу P, вычесть из диагональных элементов 1, а последнюю строку заполнить единицами. Тогда можно воспользоваться стандартной процедурой решения системы линейных уравнений.

```
> A=t(P)
> A[1,1]=A[1,1]-1
> A[2,2]=A[2,2]-1
> A[3,1]=A[3,2]=A[3,3]=1
> A
         [,1]     [,2]    [,3]
[1,] -0.82976  0.25755 0.34638
[2,]  0.25109 -0.66239 0.42560
[3,]  1.00000  1.00000 1.00000
> b=c(0,0,1)
> pst=solve(A,b)
> pst
[1] 0.2682104 0.3481600 0.3836296
```

Проверим выполнение условий (16.1).

```
> t(P) %*% pst
          [,1]
[1,] 0.2682104
[2,] 0.3481600
[3,] 0.3836296
> sum(pst)
[1] 1
```

В стационарной марковской цепи вектор вероятностей на i-м шаге равен вектору вероятности на $i+1$-м шаге. Рассмотрим случайную величину $\xi(t_i)$ как сечение процесса в момент t_i, а случайную величину $\xi(t_{i+1})$ как сечение процесса в момент t_{i+1} (реализацией случайной величины является переход системы в одно из состояний). Эти случайные величины принимают свои значения с одинаковыми вероятностями. Отсюда нужно сделать вывод, что коэффициент корреляции между соседними сечениями равен 1.

Ответ: численные значения приведены в решении.

17 Характеристики случайного процесса

Задача. Найти математическое ожидание, дисперсию, автоковариационную и корреляционную функции, а также одномерную и двумерную плотности распределения случайного процесса $\xi(t) = Ut^c$, где U — случайная величина, имеющая экспоненциальное распределение $p_u(u) = \lambda e^{-\lambda u}$ при $u \geqslant 0$, $p_u(u) = 0$ при $u < 0$; $c = 0.1 + 0.9\,KV$. Является ли этот процесс стационарным?

Решение. Математическое ожидание и дисперсия (см. задачу 8)

$$E[\xi(t)] = E(Ut^c) = t^c E(U) = \frac{t^c}{\lambda}, \quad D[\xi(t)] = D(Ut^c) = t^{2c} D(U) = \frac{t^{2c}}{\lambda^2}.$$

Ковариационная функция

$$K(s,t) = \operatorname{cov}[\xi(s), \xi(t)] = E[\xi(s)\xi(t)] - E[\xi(s)]E[\xi(t)] = E[Us^c Ut^c] - \frac{s^c t^c}{\lambda^2} =$$
$$= s^c t^c E[U^2] - \frac{s^c t^c}{\lambda^2} = s^c t^c \int_0^\infty u^2 p_u(u)\,du - \frac{s^c t^c}{\lambda^2} = s^c t^c \cdot \frac{2}{\lambda^2} - \frac{s^c t^c}{\lambda^2} = \frac{s^c t^c}{\lambda^2}.$$

Корреляционная функция

$$\rho(s,t) = \frac{K(s,t)}{\sigma(s)\,\sigma(t)} = \frac{s^c t^c}{\lambda^2} \Big/ \frac{s^c}{\lambda} \Big/ \frac{t^c}{\lambda} = 1.$$

Одномерная и двумерная плотности распределения случайного процесса $\xi(t) = Uf(t)$, если известна плотность распределения $p_u(u)$ и $f(t) \neq 0$, определяются следующим образом [1, с. 20]:

$$p_\xi(x,t) = \frac{p_u[x/f(t)]}{|f(t)|}, \quad p_\xi(x_1, s; x_2, t) = \frac{p_u[x_1/f(s)]\, p_u[x_2/f(t)]}{|f(s)f(t)|}. \tag{17.1}$$

В данном случае имеем

$$p_\xi(x,t) = \frac{\lambda e^{-\frac{\lambda x}{t^c}}}{t^c}, \quad p_\xi(x_1, s; x_2, t) = \frac{\lambda^2 e^{-\frac{\lambda x_1}{s^c} - \frac{\lambda x_2}{t^c}}}{s^c t^c}.$$

Случайный процесс не является стационарным, т. к. $E[\xi(t)] \neq E[\xi(s)]$.

Ответ: численные значения приведены в решении.

Дополнение к задаче 11

По результатам проверки работы от 17.01.2006 необходимо сделать несколько пояснений к задаче 11.

О целесообразности проведения расчетов на ЭВМ

Решение задачи предполагает значительный объем однотипных вычислений (расчетов), справедливость которых не поддается элементарной проверке. Имелось в виду следующее. При выполнении арифметических операций часто можно приближенно оценить результат, не прибегая к точным расчетам. Проведя такую проверку «в уме», можно говорить о «правдоподобном» или «неправдоподобном» значении. Однако по условию задачи 11 требуется получить выборку с использованием таблицы случайных чисел и, с другой стороны, с использованием тригонометрических функций. В таком случае любая ошибка в расчетах, проведенных «вручную», может остаться незамеченной. При этом расчеты сами по себе не являются сложными, но их требуется повторить несколько раз (для каждого элемента выборки). Это говорит о целесообразности проведения расчетов на ЭВМ. При этом не используются какие-либо специальные алгоритмы, т. е. ЭВМ в данном случае можно рассматривать как «программируемый калькулятор».

Тестирование программы

Вычислим значения элементов выборки x_1 и x_2 вручную. Из условия

$$x_i = \mu + \sigma\sqrt{-2\ln R_i}\cos(2\pi R_{i+1}), \quad x_{i+1} = \mu + \sigma\sqrt{-2\ln R_i}\sin(2\pi R_{i+1})$$

вычислим

$$x_1 = \mu + \sigma\sqrt{-2\ln R_1}\cos(2\pi R_2), \quad x_2 = \mu + \sigma\sqrt{-2\ln R_1}\sin(2\pi R_2).$$

Из таблицы 11.1 $R_1 = 0.5427$, $R_2 = 0.3815$. Тогда

$$\begin{gathered}
\ln R_1 = \ln 0.5427 = -0.611198597912778,\\
\sqrt{-2\ln R_1} = \sqrt{1.22239719582556} = 1.10562072874271,\\
2\pi R_2 = 2\cdot 3.14159265358979\cdot 0.3815 = 2.39703519468901,\\
\sin(2\pi R_2) = \sin 2.39703519468901 = 0.677646437470104,\\
\cos(2\pi R_2) = \cos 2.39703519468901 = -0.735387860781014.
\end{gathered}$$

Значения $\mu = 28.24$, $\sigma = 66.12976$ из задачи 9. Тогда

$$\begin{gathered}
x_1 = 28.24 + 66.12976\cdot 1.10562072874271\cdot -0.735387860781014 = -25.5274668017022,\\
x_2 = 28.24 + 66.12976\cdot 1.10562072874271\cdot 0.677646437470104 = 77.7857353501452.
\end{gathered}$$

Результаты расчетов совпадают с выводом программы (см. с. 23).

О вычислении квантилей

Требуется обосновать нахождение квантилей

$$t_{9;0.95396} = 1.884922, \quad \psi_{0.95396} = 1.684526, \quad \chi^2_{9;0.95396} = 17.17534, \quad \chi^2_{9;0.04604} = 3.241189.$$

Эти значения получены в системе математического моделирования R.

```
> qt(0.95396,9)
[1] 1.884922
> qnorm(0.95396)
[1] 1.684526
> qchisq(0.95396,9)
[1] 17.17534
> qchisq(0.04604,9)
[1] 3.241189
```

Здесь все функции вычисления квантилей начинаются с буквы `q`; далее, `t` означает распределение Стьюдента (т. н. t-статистика), `norm` означает нормальное распределение (по умолчанию — стандартное нормальное распределение), `chisq` означает распределение χ^2.

Трудность вычисления квантилей в общем виде связана с отысканием обратной функции распределения (см. замечание на с. 15).

Квантиль нормального распределения легко проверить по таблице. Из таблицы нормированной функции Лапаласа [2, с. 368] $\Phi(1.68) = 0.4535$, $\Phi(1.69) = 0.4545$. Следовательно, $\psi_{0.5+0.4540} = \psi_{0.954} \approx (1.68 + 1.69)/2 = 1.685$.

Далее, известно, что квантиль распределения Стьюдента при $n > 30$ приближенно равен квантилю стандартного нормального распределения, а в данном случае при $n = 9$ несколько больше его.

Далее, по таблице «критические точки распределения χ^2» [4, с. 512] для $n = 9$ находим значение 16.9 при уровне значимости $\alpha = 0.05$ и значение 3.33 при $\alpha = 0.95$.

Заметим, что нахождение квантилей по таблицам не более, а *менее* надежно, чем использование вычислительных процедур (но, быть может, более наглядно).