ФИЗИЧЕСКИЙ ФАКУЛЬТЕТ
КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ

БИЛЯЛОВ Р.Ф.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лекционный курс и практические занятия

2-е издание, исправленное и дополненное

Казань – 2004

Печатается по решению Редакционно-издательского совета физического факультета.

УДК 519.21(075.8)

Билялов Р.Ф. Теория вероятностей и математическая статистика. Лекционный курс и практические занятия. 2-е изд., испр. и доп. — Казань, 2004. — 138 с.

Учебное пособие, написанное на основе книги В.П.Чистякова "Курс теории вероятностей", предназначено для оказания помощи студентам в усвоении учебного материала большого объема при прохождении его в течение одного семестра при одной лекции и при одном практическом занятии в неделю. Более простые задачи для практических занятий взяты из книги "Сборник задач по математике для для втузов (теория вероятностей и математическая статистика), под ред. А.В.Ефимова, М., Наука, 1990. Номера задач этого задачника начинаются с чисел 14 и 15.

Научный редактор:
доктор ф.-м. наук, профессор, член-корреспондент АН РТ
Аминов Л.К.
Рецензент:
заведующий кафедрой математической статистики,
доктор ф.-м. наук, профессор Володин И.Н.

1 ПРЕДМЕТ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

В результате одного опыта событие может произойти или не произойти. Но если при проведении серии из большого числа опытов относительное число случаев, когда событие произошло, колеблется около числа p, то говорят, что вероятность того, что данное событие произойдет, равна p.

Типичная задача теории вероятностей. Пусть вероятность события равна p. Проведено n опытов, какова вероятность того, что событие произойдет m раз? В теории вероятностей доказывается, что эта вероятность равна $C_n^m p^m (1-p)^{n-m}$, где C_n^m есть число сочетаний из n по m. Теория вероятностей оперирует такими понятиями, как случайная величина, случайный процесс. Так, например, случайная величина, подчиненная распределению Пуассона, есть такая величина, для которой вероятность принять неотрицательное целое значение n равна $\frac{\lambda^n}{n!}e^{-\lambda}$, где λ — некоторый параметр, имеющий смысл среднего значения этой случайной величины. Проводится большое число опытов, каждый раз записывается показанное случайной величиной значение, эти значения суммируются, затем сумма делится на число опытов. Это и есть значение параметра λ.

Распределение Пуассона использовали при изучении странного феномена [2, стр 34]. За 20 лет между 1875 и 1894 годами в 14 различных кавалерийских корпусах германской армии была собрана статистика трагических случаев, когда солдат был убит ударом копыта. Согласно 280 наблюдениям (280=14×20), 196 солдат погибли таким образом, т.е. в среднем за год $\lambda = 0.7$. Если бы число трагических исходов подчинялось распределению Пуассона с параметром $\lambda = 0.7$, то можно было бы ожидать, что при 280 наблюдениях в 139 случаях смертей нет, в 97 случаях — 1 смерть, в 34 случаях — 2 смерти и т. д. В действительности данные были соответственно 140, 91, 32 и т. д. Практика и теория оказались в столь хорошем согласии, что вряд ли можно было ожидать большего.

Распределение Пуассона находит применение в описании многих явлений. Например, если λ есть среднее число звонков в минуту на городскую телефонную станцию, то λt будет средним числом звонков за t минут, тогда вероятность того, что за время t на телефонную станцию поступило k звонков, равна $\frac{(\lambda t)^k}{k!}e^{-\lambda t}$.

В итоге предмет теории вероятностей можно определить так: *теория вероятностей занимается вычислением вероятностей сложных событий.*

Чем занимается математическая статистика? Мы считаем, что при бросании монеты вероятность p появления герба равна $\frac{1}{2}$. Но это предположение основано на теоретических рассуждениях. Бюффон проводил опыты с бросанием монеты. Монета была брошена $n = 4040$ раз, из них в 2048 случаях выпал герб. Спрашивается, какое заключение можно сделать о вероятности выпадения герба при бросании монеты? Ответ на этот вопрос математическая статистика находит, например, следующим образом. Мы имеем большое число $n = 4040$ испытаний, в каждом из которых вероятность наступления события, состоящего в том, что выпал герб, неизвестна. Обозначим ее p. Мы уже отметили, что вероятность m успехов при n испытаниях равна $P_n(m) = C_n^m p^m (1-p)^{n-m}$. Рассмотрим разность $\frac{m}{n} - p$ и представим ее в виде $\frac{m}{n} - p = x_m\sqrt{\frac{p(1-p)}{n}}$, где $x_m = \frac{m-np}{\sqrt{p(1-p)n}}$. Пусть a и b — данные числа. В математическом анализе доказывается, что при больших n приближенно имеет место:

$$\sum_{a \le x_m \le b} P_n(m) = \sum_m P(a \le x_m \le b) \approx \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{x^2}{2}}\, dx.$$

Поэтому

$$P\left(|\frac{m}{n}-p|<\epsilon\right)=P\left(|x_m|<\epsilon\sqrt{\frac{n}{p(1-p)}}\right)\approx 2\frac{1}{\sqrt{2\pi}}\int\limits_0^{\epsilon\sqrt{\frac{n}{p(1-p)}}} e^{-\frac{x^2}{2}}dx.$$

Так как $p(1-p)\leq\frac{1}{4}$ при $0<p<1$, то дальше можно писать

$$P\left(|\frac{m}{n}-p|<\epsilon\right)\geq\frac{2}{\sqrt{2\pi}}\int\limits_0^{2\epsilon\sqrt{n}} e^{-\frac{x^2}{2}}dx.$$

Потребуем, чтобы правая часть неравенства была равна 0.99, тогда

$$\frac{1}{\sqrt{2\pi}}\int\limits_0^{2\epsilon\sqrt{n}} e^{-\frac{x^2}{2}}dx=0.495.$$

Из таблицы значений интеграла находим, что верхний предел должен равняться 2.57, значит, для $n=4040$ мы находим $\epsilon=\frac{2.57}{2\sqrt{n}}\approx 0.02$. С другой стороны, неравенство $|\frac{m}{n}-p|<\epsilon$ эквивалентно неравенству $\frac{m}{n}-\epsilon<p<\frac{m}{n}+\epsilon$ или, учитывая, что $\frac{m}{n}=0.507$, неравенству $0.487<p<0.527$. Таким образом, мы приходим к следующей интерпретации опыта Бюффона. С вероятностью, не меньшей 0.99, вероятность p выпадения герба находится в интервале $(0.487, 0.527)$. Найти p, значит найти вероятностную модель. Значит, математическая статистика, опираясь на методы теории вероятностей, путем обработки данных наблюдений определяет вероятностную модель, с помощью которой можно интерпретировать явление. Задача математической статистики обратна задаче теории вероятностей. Теория вероятностей в рамках данной вероятностной модели вычисляет сложные вероятности, а математическая статистика находит эту модель.

В современном понимании задач математической статистики *математическая статистика занимается вычислением вероятности принятия неправильного решения для конкретных статистических правил или строит правила с минимальной вероятностью этой ошибки* [7].

Важна роль теории вероятностей в теоретической физике. Теория вероятностей как математическая теория лежит в основе таких разделов физики, как квантовая механика, квантовая теория поля, статистическая физика, радиофизика. Так, в квантовой механике состояние физической системы в данный момент времени может быть описано определенной функцией (вообще говоря, комплексной) $\Psi(q)$, $q = (q_1, ..., q_N)$, причем квадрат модуля этой функции $|\Psi(q)|^2$ определяет вероятность $|\Psi(q)|^2 dq$ того, что произведенное над системой измерение обнаружит значения координат в элементе $dq = dq_1 \cdots dq_N$ конфигурационного пространства координат q.

2 ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Теория вероятностей допускает аксиоматическое изложение. Аксиоматический подход предполагает введение некоторых первоначальных объектов с отдельными свойствами, которые определяются аксиомами. Затем из этих аксиом выводятся теоремы. Основное понятие теории вероятности — вероятностное пространство. Изложение теории вероятности начнём с этого понятия.

2.1 Вероятностное пространство

Вероятностное пространство — это тройка $(\Omega, \mathbf{A}, P)$ из следующих составляющих элементов: Ω — пространство элементарных событий; $\mathbf{A}$ – алгебра событий; P — вероятность. Дадим описание этих элементов.

Пространство элементарных событий Ω. Понятие элементарного события — изначальное понятие, оно определению не поддаётся. Считается, что реальный опыт сопровождается множеством взаимно исключающих друг друга элементарных событий. Элементарное событие обозначается с помощью ω. Множество Ω всех элементарных событий ω и есть пространство элементарных событий.

Пример 1. Игральная кость — это куб, грани которого пронумерованы с помощью чисел от 1 до 6. Через ω_i обозначим событие, состоящее в том, что при бросании кости грань с номером i окажется наверху. Число возможных исходов опыта конечно, равно 6.

Пример 2. Пусть опыт состоит в бросании точки на отрезок

$[0, 1]$, т. е. в выборе какой-либо точки на этом отрезке. Если в результате опыта конкретная точка отрезка выбрана, то это событие можно охарактеризовать с помощью ее координаты. Поэтому в этом опыте нужно считать, что $\Omega = [0, 1]$. Число возможных исходов бесконечно.

Алгебра событий A. Событие — это совокупность некоторого числа элементарных событий, то есть подмножество пространства Ω. A — обозначение события. Каждое событие есть подмножество в Ω — обратное не всегда верно, но об этом позднее. Событие A считается происшедшим, если произошло одно из элементарных событий ω, из которых состоит A. Пространство Ω — тоже подмножество самого себя, и считается, что Ω – событие. Оно называется достоверным событием. Можно ввести сумму $A + B$, произведение AB $(A \cap B)$, разность $A \backslash B$ событий в соответствии с операциями суммы, произведения и разности над множествами. Как известно, суммой двух множеств называется их объединение, произведением — их пересечение, разностью — совокупность элементов множества A, не принадлежащих множеству B. На языке теории вероятностей $A + B$ — событие, состоящее в том, что произошло хотя бы одно из событий: A или B. Произведение AB — событие, состоящее в том, что одновременно произошло и событие A, и событие B. Противоположное событие $\bar{A}$ — это событие $\Omega \backslash A$. На языке теории вероятностей $\bar{A}$ — это событие, состоящее в ненаступлении события A. Невозможное событие $\emptyset$ — событие, соответствующее пустому множеству или событие, не содержащее ни одного элементарного события. События A и B называются несовместными, если $AB = \emptyset$.

Алгебра событий $\mathbf{A}$ – это совокупность событий, обладающих свойствами:

1) $\Omega \in \mathbf{A}$

2) $\forall A, B \in \mathbf{A} \Rightarrow \ A + B, \ AB, \ A \backslash B, \ \bar{A} \in \mathbf{A}$.

Вероятность P. С помощью R будем обозначать множество всех вещественных чисел. Говорят, что на алгебре событий $\mathbf{A}$ задана вероятность, если задана функция P: $\mathbf{A} \to R$, $\forall A \in \mathbf{A} \to P(A) \in [0, 1] \subset R$, обладающая свойствами:

1) $P(\Omega) = 1$,

2) Для любой пары несовместных между собой событий A и B ($AB = \emptyset$), выполняется $P(A+B) = P(A) + P(B)$ (аддитивное свойство),

3) для любой последовательности $A_1 \supset A_2 \supset \cdots A_n \supset \cdots$, такой, что $\bigcap\limits_{n=1}^{\infty} A_n = \emptyset$, имеет место равенство $\lim\limits_{n\to\infty} P(A_n) = 0$ (непрерывность).

Следствие 1. $P(\bar{A}) = 1 - P(A)$.

$A+\bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$, поэтому $P(\Omega) = 1 = P(A+\bar{A}) = P(A)+P(\bar{A})$, откуда $P(\bar{A}) = 1 - P(A)$.

Следствие 2. $P(A+B) = P(A) + P(B) - P(AB)$.

$A+B = A+B\Omega = A+B(A+\bar{A}) = A+B\bar{A}$, кроме того $A(B\bar{A}) = A\bar{A}B = \emptyset$, поэтому $P(A+B\bar{A}) = P(A) + P(B\bar{A})$. С другой стороны, $BA + B\bar{A} = B$, $(BA)(B\bar{A}) = \emptyset$, поэтому $P(BA) + P(B\bar{A}) = P(B)$. В итоге $P(A+B) = P(A) + P(B) - P(AB)$.

Классическое определение вероятности. Пусть Ω — конечное множество и число его элементов $|\Omega|$ равно s: $\Omega = \{\omega_1, \omega_2, ..., \omega_s\}$. Пусть элементарные события равновероятны: $P(\omega_1) = ... = P(\omega_s) \Rightarrow P(\omega_i) = \frac{1}{s},\ i = 1, ..., s$. Положим $P(A) = \frac{|A|}{s}$. В качестве элементов для $\mathbf{A}$ возьмём все возможные подмножества из Ω. У нас есть $(\Omega, \mathbf{A}, P)$ — вероятностное пространство.

Урновые схемы. Пусть $\aleph = \{1, 2, ..., N\}$, $\omega = (i_1, i_2, ..., i_n)$ — упорядоченный набор n чисел с $i_k \in \aleph,\ k = 1, 2, ..., n$.

1) *Схема случайного выбора с возвращением.* i_k произвольны, и $\Omega = \{\omega\}$. Урновая интерпретация данной схемы такова. В урне N пронумерованных шаров. Шары тщательно перемешиваем, затем вытаскиваем шар, записываем его номер i_1, затем шар кладем обратно, все перемешиваем, вытаскиваем шар с номером i_2, и т.д. $|\Omega| = N^n$.

2) *Схема случайного выбора без возвращения.* $i_1 \neq \ss_2 \neq \cdots \neq i_n$, (шары не возвращаем). $n \leq N,\ |\Omega| = N(N-1)\cdots(N-(n-1))$.

3) *Случайные числа.* К ним придем, если применим урновую схему с возвращением для $\aleph = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$.

Геометрические вероятности.

1) Пусть $\Omega = [0,1] \in R$. Событиями назовём всевозможные множества (a,b), $[a,b)$, $(a,b]$, $[a,b]$, $0 \leq a < b \leq 1$. Пусть точки множества Ω равновероятны, это значит, что $P((a,b)) = P([a,b]) = b - a$. Тогда $P((a)) = 0$, где (a) — отдельная точка.

Теперь событиями назовём такие подмножества пространства Ω, для которых существует конечная длина. Оказывается на Ω можно указать подмножества, для которых конечная длина не существует. Поэтому их как события рассматривать нельзя [3].

2) В качестве Ω можно взять фигуру на плоскости, для которой существует площадь $S(\Omega)$. $A \subset \Omega$ — событие, если для него тоже существует конечная площадь $S(A)$ и $P(A) = \frac{S(A)}{S(\Omega)}$ и т.д.

2.1.1 1-е практическое занятие. Простейшие вероятностные схемы

Задача 14.66. В магазин поступило 30 новых цветных телевизоров, среди которых 5 имеют скрытые дефекты. Наудачу отбирается один телевизор для проверки. Какова вероятность, что он не имеет скрытых дефектов?

Решение. A={отобранный телевизор не имеет дефектов}. $|A| = 25$. $|\Omega| = 30$. $P(A) = 5/6$.

Задача 14.73. 1 сентября по расписанию запланированы три занятия по разным предметам. Студент, не успевший ознакомиться с расписанием, пытается его угадать. Какова вероятность успеха, если учесть, что на этом курсе читается 10 предметов и любое расписание из трех предметов равновероятно?

Решение. $C_{10}^3 = 120 -$ число всевозможных выборов трех предметов из десяти. Вероятность угадать три предмета в расписании равна $1/120$.

Задача 14.80. На конференцию из 3 первокурсников, 5 второкурсников и 7 третьекурсников необходимо выбрать наугад 5 человек. Найти вероятность событий: A={будут выбраны одни третьекурсники}, B={все первокурсники попадут на конференцию}, C={не будет выбрано ни одного второкурсника}.

Решение. Число всевозможных выборов 5 студентов из 15 равно

$$C_{15}^5 = 13 \cdot 11 \cdot 7 \cdot 3.\ |A| = C_7^5 = 21,\ P(A) = C_7^5 / C_{15}^5 = 1/143,$$

$$|B| = C_3^3 \cdot C_{12}^2 = 66,$$

$$P(B) = \frac{66}{C_{15}^5} = 2/91,\ P(C) = \frac{C_{10}^5}{C_{15}^5} = 12/143.$$

Задача 14.92 На пяти карточках написаны цифры от 1 до 5. Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления в ряд слева направо. Найти вероятность событий: A={появится число 123}, B={появится число, не содержащее цифры 3}.

Решение. $\omega = (i_1, i_2, i_3)$, $1 \le i_k \le 5$, $i_1 \ne i_2 \ne i_3$ — урновая схема без возвращения, $|\Omega| = 5 \cdot 4 \cdot 3 = 60$, $P(A) = 1/60$, $|B| = 4 \cdot 3 \cdot 2 = 24$, $P(B) = 24/60 = 0.4$.

Задача 2.1. Брошены две игральные кости. Предполагая, что элементарные события равновероятны, найти вероятность событий: A={на 1-й кости выпала "1"}, $\bar{A}$, B={выпала хотя бы одна "6"}, $A\bar{B}$.

Решение.$|\Omega| = 6^2 = 36$. Выпишем все элементарные события, из которых состоит событие A: $A = \{\omega : \omega = (1, i);\ i = 1, ..., 6\}$. Тогда $P(A) = |A|/|\Omega| = 6/36 = 1/6$, $P(\bar{A}) = 1 - P(A) = 5/6$. $B = \{\omega : (6, i), (i, 6), (6, 6);\ i = 1, ..., 5\}$, $P(B) = 11/36$. $A\bar{B} = \{\omega : (1, i) : i = 1, ..., 5)$, $P(A\bar{B}) = 5/36$.

Задача 2.2. На полке в случайном порядке расставлено n книг, среди которых находится двухтомник Д. Лондона. Предполагая, что различные расположения книг равновероятны, найти вероятность того, что оба тома двухтомника расположены рядом.

Решение. Число различных способов расстановки n книг равно $n!$ (число перестановок из n элементов равно $P_n = n!$). Итак, $|\Omega| = n!$. Чему равняется число элементарных событий, благоприятных событию A={ оба тома двухтомника окажутся рядом}? Рассмотрим оба тома как одну книгу. Тогда будет $(n-1)$ книга. Учтем также, что объединение двух томов имеет два варианта (1-й том, 2-й том) и (2- й том, 1-й том). Следовательно, $|A| = (n-1)! \cdot 2$. Отсюда $P(A) = 2 \cdot (n-1)!/n! = 2/n$.

Задача 2.5. Выписана последовательность из n случайных чисел. Найти вероятность событий: A={1-е число — четное}, B={среди n чисел ровно m делятся на 3}, C={среди n чисел ровно $m+2$ делятся на 3, и два из них расположены на концах последовательности}.

Решение. Имеем урновую схему с возвращением: $A = \{(i_1, ..., i_n)$: i_1 —четное, $0 \le i_k \le 9$, $k = 2, ..., 9\}$. $|A| = 5 \cdot 10^{n-1}$, $|\Omega| = 10^n$, $P(A) = 0.5$. B=$\{(..., j_1, ..., j_2, ..., j_m, ...) : j_k \in \{0, 3, 6, 9\}$, $1 \le k \le m$, остальные числа не делятся на 3}. $|B| = 4^m \cdot 6^{n-m} C_n^m$, где 4^m — мощность множества $\{(j_1, j_2, ..., j_m)\}$, C_n^m — число всевозможных вариантов расположения чисел $j_1, j_2, ..., j_m$ во всей цифровой последовательности, 6^{n-m} — мощность множества цифровых последовательностей, состоящих из чисел, не делящихся на 3. $P(B) = (0.4)^m \cdot (0.6)^{n-m} \cdot C_n^m$. Третий случай сводится ко второму. Имеем: $|C| = C_{n-2}^m \cdot 4^{m+2} \cdot 6^{n-m-2}$. $P(C) = C_{n-2}^m \cdot (0.4)^{m+2} \cdot (0.6)^{n-m-2}$.

Задача 2.9. По некоторому участку железной дороги за N интервалов времени проходит заданное количество поездов, среди которых M тяжелых. Каждый интервал времени может быть свободен или занят одним поездом. Любое расположение тяжелых поездов по интервалам времени имеет одну и ту же вероятность. Прохождение тяжелых поездов в соседних интервалах времени нежелательно. Оценить сверху вероятность появления хотя бы одной пары соседних интервалов, занятых тяжелыми поездами, если $N = 1000$, $M = 10$.

Решение. A_k={ k-й и k+1-й интервалы заняты тяжелыми поездами},

$$P(A_k) = C_{N-2}^{M-2}/C_N^M = \frac{(N-2)!}{(M-2)!(N-M)!} : \frac{N!}{M!(N-M)!} =$$

$$= \frac{M(M-1)}{N(N-1)}.$$

$A_1 + A_2 + ... + A_{N-1}$ ={ хотя бы одна пара соседних интервалов занята тяжелыми поездами},

$$P(A_1 + A_2) = P(A_1) + P(A_2) - P(A_1 A_2) \le P(A_1) + P(A_2),$$

$$P(A_1 + A_2 + ... + A_{N-1}) \le P(A_1) + ... + P(A_{N-1}) = (N-1)P(A_1) =$$

$$= \frac{M(M-1)}{N} = 0.09.$$

Задачи домашнего задания.

Задача 14.67. Автомат изготавливает однотипные детали, причем технология изготовления такова, что 5% произведенной продукции оказывается бракованной. Из большой партии взята наудачу одна деталь для контроля. Найти вероятность события ={деталь бракованная}.

Ответ: 0.05.

Задача 14.74. Зенитная батарея, состоящая из n орудий, производит залп по группе, состоящей из m самолетов. Каждое из орудий выбирает себе цель наудачу и независимо от остальных. Найти вероятность того, что все орудия выстрелят по одному самолету.

Ответ: $\frac{1}{m^{n-1}}$

Задача 14.81. Из урны, содержащей $m_1 + m_2$ шаров, из которых m_1 белых и m_2 черных, наудачу отбирают m шаров и откладывают в сторону. Найти вероятности следующих событий: A={все отложенные шары белые}, B={среди отложенных шаров ровно k белых; $k \leq m$ }.

Ответ: $P(A) = \frac{C_{m_1}^m}{C_{m_1+m_2}^m},\ P(B) = \frac{C_{m_1}^k C_{m_2}^{m-k}}{C_{m_1+m_2}^m}$.

Задача 14.93. (Продолжение задачи 14.92) При условиях предыдущей задачи найти вероятности событий: C={появится число, состоящее из последовательных цифр}, D={появится четное число}, E= {появится число, содержащее хотя бы одну из цифр: 2 или 3}.

Ответ: $P(C) = \frac{1}{20},\ P(D) = \frac{2}{5},\ P(E) = \frac{9}{10}$.

Задача 2.3. Числа 1, 2, ..., расставлены случайным образом. Предполагая, что различные разположения чисел равновероятны, найти вероятность того, что числа 1, 2, 3 расположены в порядке возрастания, но не обязательно рядом.

Ответ: $\frac{1}{6}$.

Задача 2.4. Выписаны три случайных числа. Найти вероятности событий:

A={все выписанные числа одинаковы},

B={ все выписанные числа различны},

C={среди выписанных чисел ровно два совпадения}.

Ответ: 0.01; 0.72; 0.27.

Задача 2.7. В чулане n пар ботинок. Из них случайно выбирается $2r$ ботинок $(2r < n)$. Найти вероятность того, что среди выбранных ботинок: а) нет парных; б) имеется ровно одна пара.

Ответ: а) $\frac{2^{2r}C_n^{2r}}{C_{2n}^{2r}}$ б) $\frac{n2^{2r-2}C_{n-1}^{2r-2}}{C_{2n}^{2r}}$

Задача 2.10. В купейный вагон (9 купе по 4 места) семи пассажирам продано семь билетов. Найти вероятности событий:

A={пассажиры попали в два купе},

B={пассажиры попали в три купе}.

Рассмотреть два случая:

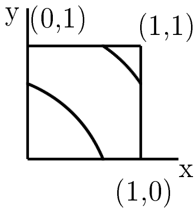
1) пассажиры покупают билеты в разное время независимо друг от друга (воспользоваться схемой случайного выбора без возвращения);

2) пассажиры едут вместе, и один покупает билеты всей группе (предположить, что номера проданных пассажиру мест идут подряд, а наименьший номер места выбирается случайно из множества номеров
{1,2,...,30}).

Ответ: 1) $P(A) = \frac{1}{28985} = 0.0000345...$, $P(B) = \frac{224}{28985} = 000777281...$;
2) $P(A) = \frac{8}{15} = 0.5333...$, $P(B) = \frac{7}{15} = 0.4666....$

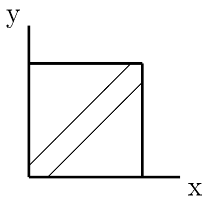
2.1.2 2-е практическое занятие. Простейшие вероятностные схемы (продолжение).

Задача 14.139. Внутри квадрата с вершинами (0,0), (1,0), (1,1), и (0,1) наудачу выбирается точка $M(x, y)$.



Найти вероятность события $A = \{(x, y) : x^2 + y^2 \le a^2, a > 0\}$.
Решение. Пусть S равна площади множества A. $P(A) = S$, $S = \frac{\pi a^2}{4}$, если $a \le 1$.
$S = \sqrt{a^2 - 1} + a^2(\frac{\pi}{4} - \arccos 1/a)$, если $1 \le a \le \sqrt{2}$.
$S = 1$, если $a > \sqrt{2}$.

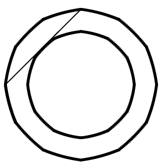
Задача 14.150. Два парохода должны подойти к одному и тому же причалу. Время прихода обоих пароходов независимо и равновозможно в течение данных суток. Определить вероятность того, что одному из пароходов придется ожидать освобождения причала, если время стоянки первого парохода — один час, а второго — два часа.



Решение. Пусть x — время прибытия 1-го парохода, y — второго, $0 \le x, y \le 24$. Если $x \le y \le x + 1$, то ждет 2-й пароход. Если $y \le x \le y + 2$, то ждет первый.

$$P(a) = \frac{24^2 - \frac{1}{2}23^2 - \frac{1}{2}22^2}{24^2} = \frac{139}{1152}.$$

Задача 14.157. Даны две концентрические окружности радиусов $r_2 > r_1$. На большей окружности наудачу ставятся две точки A и B. Какова вероятность того, что отрезок AB не пересечет малую окружность?



Решение. Искомая вероятность равна отношению длины дуги большей окружности, стягиваемой хордой к длине этой окружности, причем хорда касается меньшей окружности.

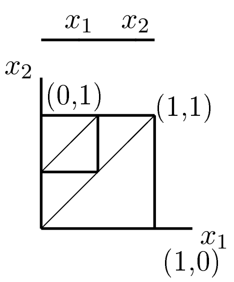
$$P = \frac{2\arcsin\sqrt{1-(r_1/r_2)^2}}{\pi} = \frac{2}{\pi}\arccos\frac{r_1}{r_2}.$$

Задача 2.15. На отрезок [0,1] наудачу брошена точка. Предположив, что ее координата ξ равномерно распределена на отрезке [0,1], найти функции $F(x) = P(\xi < x)$ и $F'(x)$ $(-\infty < x < +\infty)$.

Решение. $F(x) = P(\xi < x) = 0$, если $x \le 0$. $F(x) = |x-0|/|1-0| = x$, если $0 < x \le 1$. $F(x) = 1$, если $x > 1$. $F'(x) = 0$, если $x < 0$ или $x > 1$, $F'(x) = 1$, если $0 < x < 1$.

Задача 2.17. На отрезок [0,1] наудачу брошены две точки, разбившие его на три отрезка. Какова вероятность того, что из этих отрезков можно построить треугольник? За множество Ω принять значения пары чисел (X_1, X_2), являющихся координатами точек на отрезке [0,1]; предположить, что точка (X_1, X_2) равномерно распределена на квадрате Ω.

Решение. Пусть для определенности $X_1 < X_2$, $0 \le X_1 \le X_2$. Из трех отрезков можно построить треугольник, если сумма длин любых двух отрезков больше или равна длине третьего отрезка. Длины отрезков равны X_1, $X_2 - X_1$, $1 - X_2$. Должны выполняться неравенства



$$\begin{cases} X_1 + (X_2 - X_1) \ge 1 - X_2 \to X_2 \ge 1/2 \\ X_1 + (1 - X_2) \ge X_2 - X_1 \to X_2 - X_1 \le 1/2 \\ (X_2 - X_1) + (1 - X_2) \ge X_2 \to X_1 \le 1/2 \end{cases}.$$

Пусть A означает событие, состоящее в том, что три отрезка составляют треугольник, тогда

$$A = \{(X_1, X_2):\ 0 \le X_1 \le X_2 \le 1, X_1 \le 1/2,$$
$$X_2 \ge 1/2, X_2 - X_1 \le 1/2\},$$

в то время как

$$\Omega = \{(X_1, X_2):\ 0 \le X_1 \le X_2 \le 1\},$$

значит

$$P(A) = \text{площадь } A/\text{площадь } \Omega = 1/4.$$

Задача 2.21. В урне M белых и $N-M$ черных шаров. По схеме случайного выбора с возвращением из урны извлекается n шаров. Найти вероятность событий: 1) A={при k-том извлечении появился белый шар}; 2) B={при k-том и l-том извлечении появились белые шары); 3) C={среди n извлеченных шаров ровно m белых}.

Решение. $\Omega = \{\omega : \omega = (i_1, i_2, ..., i_k),\ i_s = 1, 2, ..., N;\ s = 1, 2, ..., n)$, $|\Omega| = N^n$.

1) $A = \{\omega = (i_1, ..., i_{k-1}, i_k, i_{k+1}, ..., i_n),\ i_k = 1, 2, ..., M;\ i_s = 1, 2, ..., N;\ s = 1, 2, ..., k-1, k+1, ..., n\}$, $|A| = MN^{n-1}$; $P(A) = M/N$.

2) $|B| = M^2N^{n-2}$; $P(B) = M^2/N^2$.

3) $|C| = C_n^m M^m (N-M)^{n-m}$; $P(C) = C_n^m (M/N)^m (1-M/N)^{n-m}$.

Задача 2.24(a). Используя таблицу случайных чисел, получить реализацию опыта, описанного в задаче 2.21, если $N = 10$, $M = 3$, $n = 20$. По полученной реализации найти частоту появления белого шара (отношение числа белых шаров в полученной последовательности к общему числу n извлеченных шаров).

Решение. Пусть белые шары нумерованы числами 0, 1, 2. В таблице случайных чисел из первой строки выписываем 20 чисел: 10097325337652013586. Среди этих чисел количество чисел, меньших или равных 2, равно 7. Значит, частота появления белого шара равна 7/20.

Задачи домашнего задания

Задача 14.140 (продолжение). В условиях предыдущей задачи найти вероятность события $B = \{(x, y) \mid xy < a,\ a > 0\}$.

$$\text{Ответ: } P(B) = \begin{cases} a(1 - \ln a), 0 < a \le 1, \\ \qquad 1, \qquad a < 1. \end{cases}$$

Задача 14.151. В случайный момент времени $x \in [0, T]$ появляется радиосигнал длительностью t_1. В случайный момент времени $y \in [0, T]$ включается приемник на время $t_2 < t_1$. Найти вероятность обнаружения сигнала, если: а) приемник настраивается мгновенно; б) время настройки приемника равно t_3 $(t_3 < t_2 < t_1)$.

Ответ: а) $1-\frac{1}{2}(1-\frac{t_1}{T})^2-\frac{1}{2}(1-\frac{t_2}{T})^2$; б) $1-\frac{1}{2}(1-\frac{t_1-t_3}{T})^2-\frac{1}{2}(1-\frac{t_2}{T})^2$.

Задача 14.158. Из отрезка $[-1,2]$ наудачу взяты два числа. Какова вероятность того, что их сумма больше единицы, а произведение меньше единицы?

Ответ: $\frac{2}{9}\ln 2+\frac{1}{6}\approx 0.321$.

Задача 2.16 . Случайная точка a с координатами (ξ_1,ξ_2) равномерно распределена в квадрате $\Omega=\{(x_1,x_2):\ 0\le x_i\le 1,\ i=1,2\}$. Найти функции $F(x)=P\{\xi_1+\xi_2<x)$, $F'(x)(-\infty<x<\infty)$.

Ответ: $F'(x)=x\ (0\le x\le 1),\ F'(x)=2-x\ (1\le x\le 2)$.

Задача 2.18. На бесконечную шахматную доску со стороной квадрата a наудачу бросается монета радиуса $r,\ 2r<a$. Найти вероятность того, что: 1) монета целиком попадет внутрь одного квадрата; 2) пересечет не более одной стороны квадрата.

Ответ: 1) $(1-\frac{2r}{a})^2,$ 2) $1-\frac{4r^2}{a^2}$.

Задача 2.20 (Парадокс Бертрана). В круге радиуса R случайно проводится хорда. Обозначим через ξ ее длину. Найти вероятность того, что длина хорды больше стороны правильного вписанного шестиугольника, если: а) середина хорды равномерно распределена в круге; б) направление хорды задано, а ее середина равномерно распределена на диаметре, перпендикулярном ее направлению; в) один конец хорды закреплен, а другой равномерно распределен на окружности.

Вероятность зависит от интерпретации слова "случайно", и ее числовые значения различны в случаях а), б), в).

Ответ: а) $\frac{3}{4}$, б) $\frac{\sqrt{3}}{2}$, в) $\frac{2}{3}$.

Задача 2.22. Решить задачу 21 в случае выбора без возвращения.

Ответ: $P(A)=\frac{M}{N};\ P(B)=\frac{M(M-1)}{N(N-1)};\ P(C)=\frac{C_M^m C_{N_M}^{n-m}}{C_N^m}$.

Задача 2.24(б). Решить задачу 2.24(а), взяв б) $N=100, M=35, n=20$.

2.2 Условная вероятность. Вероятность произведения

$P(B|A)$ — так обозначают вероятность события B при условии, что событие A произошло. Назовём его условной вероятностью события B при условии, что событие A произошло.

Определение: $\forall A, B \in \mathbf{A}, P(A) \neq 0$ *полагается* $P(B|A) = \frac{P(AB)}{P(A)}$.

Если $P(B|A) = P(B)$, то говорят, что B не зависит от A. Но если так, то $P(AB) = P(A)P(B|A) = P(A)P(B)$. Тогда $P(A|B) = P(A)$, т.е. A тоже не зависит от B. Поэтому говорят просто, что A и B независимы. Значит, если $P(AB) = P(A)P(B)$, то это есть необходимое и достаточное условие того, что A и B независимы. Формулы с условной вероятностью:

$$P(AB) = P(A)P(B|A),$$

$$P(ABC) = P(A)P(B|A)P(C|AB),$$

$$P(ABCD) = P(A)P(B|A)P(C|AB)P(D|ABC).$$

Определение. *События* $A_1, A_2, ..., A_n$ *— независимы, если для всех комбинаций индексов* $1 \leq i_1 < \cdots < i_k \leq n$ $(k = 2, ..., n)$ *имеет место соотношение*:

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}).$$

Формула полной вероятности. Формула Байеса.

Теорема. *Пусть* A *— произвольное событие, события* $B_1, ..., B_n$ *— несовместны между собой,* $(B_i) \neq 0$, $B_1 + \cdots + B_n \subset A$. *Тогда*

$$P(A) = \sum_{i=1}^{n} P(B_i) P(A|B_i) \ \text{ — \textit{формула полной вероятности.}}$$

Доказательство. $A = A \sum\limits_{i=1}^{n} B_i = \sum\limits_{i=1}^{n} AB_i$, $P(A) = P(\sum\limits_{i=1}^{n} AB_i) = \sum\limits_{i=1}^{n} P(AB_i) = \sum\limits_{i=1}^{n} P(B_i) P(A|B_i)$.

Теорема. *При выполнении условий предыдущей теоремы имеет место*:

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum\limits_{i=1}^{n} P(B_i)P(A|B_i)} \text{ – \textit{формула Байеса.}}$$

Доказательство следует из

$$P(B_i|A)P(A) = P(AB_i) = P(B_i)P(A|B_i).$$

2.2.1 3-е практическое занятие. Условные вероятности

Задача 14.164. Вероятность попасть в самолет равна 0.4, а вероятность его сбить равна 0.1. Найти вероятность того, что при попадании в самолет он будет сбит.

Решение. A ={попадание в самолет}, B = {самолет сбит при попадании}. $P(B|A) = P(AB)/P(A) = P(B)/P(A) = 1/4$.

Задача 14.168. Из множества чисел $\{1, 2, ..., N\}$ по схеме случайного выбора без возвращения выбирают три числа. Найти условную вероятность того, что третье число попадает в интервал, образованный первыми двумя, если известно, что первое число меньше второго.

Решение. A ={первое число меньше второго}, B ={третье число попадает в интервал, образуемый первыми двумя}. (123), (132), (231). $P(B|A) = 1/3$.

Задача 14.179. Из 100 студентов, находящихся в аудитории, 50 человек знают английский язык, 40 — французский и 36 — немецкий. Английский и французский языки знают 20 студентов, английский и немецкий — 8, французский и немецкий — 10. Все три языка знают 5 человек. Один из студентов вышел из аудитории. Рассмотрим следующие события: E ={вышедший знает английский язык}, F = {вышедший знает французский язык}, D = {вышедший знает немецкий язык}. Необходимо: а) указать все пары независимых событий; б) установить, являются ли события E, F и D независимыми в совокупности.

Решение.

а) $P(E) = 0.5$, $P(F) = 0.4$, $P(D) = 0.36$, $P(EF) = 0.2$, $P(ED) = 0.08$, $P(FD) = 0.1$, $P(EFD) = 0.06$. E и F независимы.

б) нет, так как $P(E)P(F)P(D) = 0.07 \neq 0.06$.

Задача 14.182. В ящике лежат 12 красных, 8 зеленых и 10 синих шаров. Наудачу вынимаются два шара. Найти вероятность того, что будут вынуты шары разного цвета, при условии, что не вынут синий шар.

Решение. $B = \{$не вынут синий шар$\}$, $A =\{$шары разного цвета$\}$.

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{C_{12}^1 C_2^1}{C_{20}^2} = \frac{3 \cdot 8 \cdot 2}{5 \cdot 19} = \frac{48}{95}.$$

Задача 3.1. Брошено две игральные кости. Какова вероятность того, что выпало два раза "3", если известно, что сумма выпавших очков делится на 3?

Решение. $A =\{$сумма выпавших очков делится на 3$\}= \{$(1,2), (1,5),(2,4), (3,3),(3,6),(4,5), далее те же пары, но цифры записаны в обратном порядке$\}$. $|A| = 12$, $B = \{(3,3)\}$, $|\Omega| = 36$, $P(B|A) = P(AB)/P(A) = P(B)/P(A) = 1/12$.

Задача 3.7. Случайная точка (ξ_1, ξ_2) имеет равномерное распределение в квадрате $\{(x_1, x_2) : \ 0 \le x_1 \le 1, 0 \le x_2 \le 1\}$. Пусть $C_1 = (\xi_1 \le 1/2)$, $C_2 = (\xi_2 \le 1/2)$, $C_3 = ((\xi_1 - 1/2)(\xi_2 - 1/2) < 0)$. Показать, что события C_1, C_2, C_3 попарно независимы. Являются ли события C_1, C_2, C_3 взаимно независимыми? Зависимы ли события C_1C_2 и C_3?

Решение.

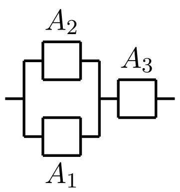
$$P(C_1) = P(C_2) = P(C_3) = 1/2,$$

$$P(C_1C_2) = P(C_1C_3) = P(C_2C_3) = 1/4,$$

$$P(C_1C_2C_3) = 0 \neq P(C_1)P(C_2)P(C_3) = 1/8.$$

Как события C_1, C_2, C_3, так и события C_1C_2, C_3 зависимы.

Задача 3.10. Элементы A_1 и A_2 электрической цепи соединены параллельно, а A_3 присоединен к ним последовательно. Вероятность выхода из строя за данный период времени элемента A_k равна p_k, $k = 1, 2, 3$. Предполагается, что элементы выходят или не выходят из строя независимо друг от друга. Найти вероятность того, что за рассматриваемый период по цепи будет проходить ток.



Решение. Пусть A_k означает событие, состоящее в том, что k-й элемент неисправен. A={проходит ток}, $A = \bar{A}_3(\bar{A}_1 + \bar{A}_2)$. $P(A) = P(\bar{A}_3)P(\bar{A}_1 + \bar{A}_2) = (1 - p_3)(1 - p_1 + 1 - p_2 - (1 - p_1)(1 - p_2)) = (1 - p_3)(1 - p_1 p_2)$.

Задача 3.16. По каналу связи может быть передана одна из трех последовательностей букв: $AAAA$, $BBBB$, $CCCC$. Известно, что вероятность каждой из последовательностей равна соответственно 0.3, 0.4, 0.3. В результате шумов буква принимается правильно с вероятностью 0.6. Вероятности приема переданной буквы за две другие равны 0.2 и 0.2. Предполагается, что буквы искажаются независимо друг от друга. Найти вероятность того, что передано $AAAA$, если на приемном устройстве получено $ABCA$.

Решение.

$$P(AAAA|ABCA) = \{P(AAAA)P(ABCA|AAAA)\}+$$

$$+\{P(AAAA)P(ABCA|AAAA) + P(BBBB)P(ABCA|BBBB)+$$

$$+P(CCCC)P(ABCA|CCCC)\} =$$

$$\frac{0.3 \cdot 0.6 \cdot 0.2 \cdot 0.2 \cdot 0.6}{0.3 \cdot 0.6 \cdot 0.2 \cdot 0.2 \cdot 0.6 + 0.4 \cdot 0.2 \cdot 0.6 \cdot 0.2 \cdot 0.2 + 0.3 \cdot 0.2 \cdot 0.2 \cdot 0.6 \cdot 0.2} =$$

$$= \frac{9}{16}.$$

Задача 3.18. На одну телефонную линию могут поступать вызовы двух типов: срочные и простые. При поступлении срочного вызова разговор по простому вызову прекращается. Вероятности поступления за время $(t, t+h)$ срочного и простого вызовов равны соответственно $\alpha_1 \cdot h + o(h)$, $\alpha_2 \cdot h + o(h)$, $h \to 0$; вероятность прекращения любого разговора за время $(t, t + h)$ равна $\beta \cdot h + o(h)$. Пусть $P_0(t)$,

$P_1(t)$, $P_2(t)$ — вероятности того, что в момент t линия свободна, занята срочным вызовом, занята простым вызовом. Написать для $P_k(t)$ дифференциальные уравнения и найти $\pi_k = \lim\limits_{t\to\infty} P_k(t)$; $k = 0, 1, 2$.

Решение. Применяя формулу полной вероятности, можем записать:

$$P_0(t+h) = P_0(t)P(\text{вызов не поступил}) +$$
$$+P_1(t)P(\text{срочный вызов прекратился}) +$$
$$+P_2(t)P(\text{простой вызов прекратился}) =$$
$$P_0(t)(1-(\alpha_1+\alpha_2)h+\cdots) + P_1(t)(\beta h+\cdots) + P_2(t)(\beta h+\cdots) =$$
$$= P_0(t) + h[-(\alpha_1+\alpha_2)P_0(t) + \beta P_1(t) + \beta P_2(t) + \cdots].$$
$$P_0'(t) = \lim_{h\to 0} \frac{P_0(t+h)-P_0(t)}{h} = -(\alpha_1+\alpha_2)P_0(t) + \beta P_1(t) + \beta P_2(t).$$
$$P_1'(t) = \alpha_1 P_0(t) - \beta P_1(t) + \alpha_1 P_2(t),$$
$$P_2'(t) = \alpha_2 P_0(t) - (\beta+\alpha_1)P_2(t).$$

В пределе $P_i(t) \to \pi_i$, $\pi_0 + \pi_1 + \pi_2 = 1$. Получаются уравнения

$$-(\alpha_1+\alpha_2)\pi_0 + \beta(\pi_1+\pi_2) = 0,\ \alpha_2\pi_0 - (\beta+\alpha_1)\pi_2 = 0,$$

откуда $\pi_0 = \frac{\beta}{\alpha_1+\alpha_2+\beta}$, $\pi_1 = \frac{\alpha_1}{\alpha_1+\beta}$, $\pi_2 = \frac{\alpha_2}{\alpha_1+\beta}\pi_0$.

Задачи домашнего задания.

Задача 14.165. Вероятность того, что прибор не откажет к моменту времени t_1, равна 0.8, а вероятность того, что он не откажет к моменту времени $t_2 (t_2 > t_1)$, равна 0.6. Найти вероятность того, что прибор, не отказавший к моменту времени t_1, не откажет и к моменту времени t_2.

Ответ: $\frac{3}{4}$.

Задача 14.169. Подбрасывают наудачу три игральные кости. Наблюдаемые события: A={на трех костях выпадут разные грани}, B ={хотя бы на одной из костей выпадет шестерка}. Вычислить $P(B|A)$ и $P(A|B)$.

Ответ: $P(B|A) = 0.5$, $P(A|B) = \frac{60}{91}$.

Задача 14.176. Из колоды в 36 карт наудачу извлекается одна карта. События: A={вынутая карта — туз}, B={вынутая карта — черной масти}, F={вынутая карта — фигура, т.е. является валетом, дамой, королем или тузом}. Установить, зависимы или независимы следующие три пары событий: A и B, A и F, F и B.

Ответ: A и B, F и B — независимы, A и F — зависимы.

Задача 14.183. На шахматную доску наудачу ставятся два слона — белый и черный. Какова вероятность того, что слоны не побьют друг друга, при условии, что белый слон попадет на одно из крайних полей доски?

Ответ: $\frac{3}{4}$.

Задача 3.2. Известно, что при бросании 10 игральных костей появилась по крайней мере одна "1". Какова вероятность того, что появилось две или более "1"?

Ответ: 0.61477....

Задача 3.5. Доказать, что события A и $\bar{B}$ независимы, если независимы события A и B.

Задача 3.9. События A_1, A_2, A_3, A_4 взаимно независимы: $P(A_k) = p_k$, $k = 1, 2, 3, 4$. Найти вероятности событий: 1) $A_1 A_3 A_4$, 2) $A_1 + A_2$, 3) $(A_1 + A_2)(A_3 + A_4)$.

Ответ: 1) $p_1(1 - p_3)p_4$, 2) $p_1 + p_2 - p_1 p_2$,
3) $(p_1 + p_2 - p_1 p_2)(p_3 + p_4 - p_3 p_4)$.

Задача 3.17. Вероятность того, что молекула, испытавшая в момент $t = 0$ столкновение с другой молекулой и не имевшая других столкновений до момента t, испытает столкновение в промежуток времени $(t, t + h)$, равна $\lambda h + o(h)$, $h \to 0$. Найти вероятность того, что время свободного пробега будет больше t.

Ответ: $e^{-\lambda t}$.

Задача 3.19. Вероятность поступления на телефонную линию одного вызова за время $(t, t+h)$ равна $\lambda h + o(h)$, $h \to 0$. Вероятность того, что ни один вызов за время $(t, t+h)$ не поступит, равна $1 - \lambda h + o(h)$. Если линия занята, то вызовы становятся в очередь. Если в момент t еще продолжается разговор, то за время $(t, t+h)$ он окончится с вероятностью $\beta h + o(h)$, $h \to 0$. Вызовы поступают независимо друг от друга. Обозначим $P_k(t)$ вероятность того, что в момент t один вызов обслуживается и $k-1$ вызовов образуют очередь ($k \geq 1$; $P_0(t)$ — вероятность того, что линия свободна). Составить для $P_k(t)$ дифференциальные уравнения. Найти $\lim\limits_{t\to\infty} P_k(t) = \pi_k$, если $\Theta = \alpha/\beta < 1$.

Ответ: $\pi_0 = 1 - \theta,\ \pi_k = (1-\theta)\theta^k,\ k \geq 1.$

2.3 Последовательность испытаний

Математическая модель последовательности из n испытаний. Пусть имеется опыт с N несовместными исходами. Эти исходы нумеруются числами $1, 2, ..., N$. Проведено n испытаний, в каждом из которых может появиться один из этих исходов. Серию из n испытаний можно описать последовательностью $\omega = (i_1, i_2, ..., i_k)$, $1 \leq i_k \leq N$, $k = 1, ..., n$. $\Omega_n = \{\omega\}$.

$$P(\omega) = P(i_1)P(i_2|i_1)P(i_3|i_1 i_2)\cdots P(i_k|i_1\cdots i_{k-1}),$$

$$\sum_{i_s=1}^{N} P(i_s|i_1\cdots i_{s-1}) = 1.$$

Если $P(A) = \sum\limits_{\omega\in A} P(\omega)$, $A \subset \Omega_n$, то мы имеем вероятностное пространство $(\Omega, \mathbf{A}, P)$ – модель последовательности из n испытаний.

Упрощающие предположения:

1) $P(i_s) = P(i_s|i_1\cdots ß_{s-1})$ — испытания независимы;

2) $P(i_s)$ не зависят от номера испытания, что интерпретируется как однородность испытаний. Обозначим $p_i = P(i)$.

Очевидно $p_1 + \cdots + p_N = 1$, $P(\omega) = p_{i_1}p_{i_2}\cdots p_{i_n}$.

Если $N = 2$, то говорят, что испытания подчиняются схеме Бернулли, $N > 2$ — полиномиальной схеме.

Схема Бернулли. $N = 2$. Два исхода, которые обозначают как: 1,2 или успех, неуспех, или выигрыш, проигрыш, или $\{A, \bar{A}\}$. Пусть m — число успехов при n испытаниях

Теорема. *Вероятность $P_n(m)$ того, что при n испытаниях произошло m успехов, равна $P_n(m) = C_n^m p^m (1-p)^{n-m}$.*

Доказательство. $P(\omega) = P(\underbrace{AA\cdots A}_{m}\underbrace{\bar{A}\bar{A}\cdots\bar{A}}_{n-m}) = p^m q^{n-m}$, и число событий ω с такой вероятностью есть C_n^m.

Полиномиальная схема. Пусть ξ_i — число случаев, когда при n испытаниях произошло событие с номером i. Тогда

$$P(\xi_1 = m_1, \xi_2 = m_2, ..., \xi_N = m_N) = \frac{n!}{m_1! m_2! \cdots m_N!} p_1^{m_1} p_2^{m_2} \cdots p_N^{m_N}.$$

Доказательство аналогично.

Теорема Пуассона. *Пусть в схеме Бернулли: $n \to \infty$, $p \to 0$, но $np \to \lambda > 0$, тогда $P_n(m) \to p_m(\lambda) = \frac{\lambda^m}{m!}e^{-\lambda}$.*

Доказательство.

$$C_n^m p^m q^{n-m} = \frac{n(n-1)\cdots(n-m+1)}{m!} p^m (1-p)^{n-m} =$$

$$= \frac{n(n-1)\cdots(n-m+1)}{n^m} \cdot \frac{1}{m!} \cdot (np)^m \cdot \left[\left(1 - \frac{np}{n}\right)^{-\frac{n}{np}}\right]^{-\frac{np}{n}(n-m)} \to$$

$$\to \frac{1}{m!} (\lambda)^m e^{-\lambda}.$$

Локальная предельная теорема Муавра- Лапласа. *Пусть в формуле Бернулли $P_n(m) = C_n^m p^m q^{n-m}$ делается допущение, что n, $m \to \infty$ так, что величина $x_m = \frac{m-np}{\sqrt{npq}}$ равномерно ограничена: существуют конечные числа a и b, такие, что $a \le x_m \le b$, тогда*

$$\frac{P_n(m)}{(2\pi npq)^{-\frac{1}{2}} e^{-\frac{x_m^2}{2}}} \to 1 \text{ при } n \to \infty.$$

Доказательство. Имеют место доказываемые в математическом анализе формула Стирлинга

$$\ln n! = \ln \sqrt{2\pi n} + n \ln n - n + O(\frac{1}{n})$$

и приближенная формула

$$\ln(1+x) \approx x - \frac{x^2}{2}.$$

Можно написать

$$m = np + x_m\sqrt{npq} = np\left(1 + x_m\sqrt{\frac{q}{np}}\right),$$

тогда

$$\ln m = \ln np + x_m\sqrt{\frac{q}{np}} - \frac{x_m^2}{2}\frac{q}{np} + \cdots .$$

Вычисляем $\ln m!$:

$$\ln m! \approx \ln\sqrt{2\pi} + \frac{1}{2}\left(\ln np + x_m\sqrt{\frac{q}{np}} - \frac{x_m^2}{2}\frac{q}{np}\right) +$$

$$+ np\left(1 + x_m\sqrt{\frac{q}{np}}\right)\left(\ln np + x_m\sqrt{\frac{q}{np}} - \frac{x_m^2}{2}\frac{q}{np}\right) -$$

$$-np - x_m\sqrt{npq} \approx$$

$$\approx \ln\sqrt{2\pi} + \frac{1}{2}\ln np + np\ln np + \frac{x_m^2 q}{2} + x_m\sqrt{npq}\ln np - np.$$

Аналогичным образом вычисляем

$$\ln(n-m)! \approx \ln\sqrt{2\pi} + \frac{1}{2}\left(\ln nq - x_m\sqrt{\frac{p}{nq}} - \frac{x_m^2}{2}\sqrt{\frac{p}{nq}}\right) +$$

$$+(nq - x_m\sqrt{npq})\left(\ln nq - x_m\sqrt{\frac{p}{nq}} - \frac{x_m^2}{2}\frac{p}{nq}\right) - nq + x_m\sqrt{npq} \approx$$

$$\approx \ln\sqrt{2\pi} + \frac{1}{2}\ln nq + nq\ln nq - x_m\sqrt{npq}\ln nq + \frac{x_m^2 p}{2} - nq.$$

Представляя $P_n(m)$ в виде $P_n(m) = \frac{n!}{m!(n-m)!}p^m q^{n-m}$, приходим к следующему результату:

$$\ln P_n(m) = -\ln\sqrt{2\pi npq} - \frac{x_m^2}{2} + \cdots ,$$

откуда следует, что

$$P_n(m) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x_m^2}{2}}.$$

Интегральная предельная теорема Муавра-Лапласа. *При обозначениях предыдущей теоремы имеет место*

$$\lim_{n\to\infty} P(a < x_m < b) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{u^2}{2}} du = \Phi(b) - \Phi(a),$$

где $\Phi(x) = \int\limits_0^x e^{-\frac{u^2}{2}} du$.

Функция $\Phi(x)$ называется функцией Лапласа, в справочниках и учебниках по теории вероятности можно найти таблицы значений этой функции.

Теорему примем без доказательства, в будущем мы получим ее как частный случай более общей центральной предельной теоремы.

Пример. В поселке A 2500 жителей. Каждый из них примерно 6 раз в месяц ездит на поезде в город B, выбирая дни поездок по случайным мотивам, независимо от остальных жителей. Какой наименьшей вместительностью должен обладать поезд, чтобы он переполнялся в среднем не чаще одного раза в 100 дней. Поезд идет раз в сутки.

Решение. Поездку жителя назовем успехом и вероятность успеха равна $p = 6/30 = 1/5$, а вероятность неуспеха равна $q = 4/5$. Число испытаний есть $n = 2500$, тогда $\sqrt{npq} = 20$. Пусть k — наименьшая вместимость поезда, m — число пассажиров. Тогда

$$P(m \geq k) = P\left(x_m \geq a = \frac{k - np}{\sqrt{npq}}\right) \approx \frac{1}{2\pi} \int\limits_a^\infty e^{-\frac{x^2}{2}} dx = \frac{1}{2} - \Phi(a) \leq 0.01,$$

откуда $\Phi(a) \geq 0.49$. Из таблицы значений функции Лапласа находим, что $a \geq 2.32$. Так как $a = \frac{k-500}{20}$, то $k \geq 546.5$.

2.3.1 4-ое практическое занятие. Последовательность испытаний

Задача 14.312. Для стрелка, выполняющего упражнение в тире, вероятность попасть в "яблочко" при одном выстреле не зависит от результатов предшествуюших выстрелов и равна $p = 1/4$. Спортсмен сделал 5 выстрелов. Найти вероятность событий: $A =$({ровно одно попадание}, $B =${ровно два попадания}.

Решение. $p = 1/4$, $q = 3/4$, $n = 5$. $P(A) = C_5^1 pq^4 = 5 \cdot \frac{3^4}{4^5} = \frac{405}{1024}$, $P(B) = C_5^2 p^2 q^3 = 10 \cdot \frac{3^3}{4^5} = \frac{135}{512}$.

Задача 14.320. В ячейку памяти ЭВМ записывается 8-разрядное двоичное число. Значения 0 и 1 в каждом разряде появляются с равной вероятностью. Случайная величина X — число единиц в записи двоичного числа. Найти вероятность событий: $A = \{X = 4\}$, $B = \{X > 4\}$.

Решение. $n = 8$. $p = q = 0.5$, $P(A) = C_8^4 p^4 q^4 = \frac{8 \cdot 7 \cdot 6 \cdot 5}{4! \cdot 2^8} = \frac{35}{128}$, $P(B) = \frac{1}{2}(1 - P(A)) = \frac{93}{256}$.

Задача 14.338. Производится стрельба из орудия по удаляющейся цели. При первом выстреле вероятность попадания равна 0.8, при каждом следующем выстреле вероятность попадания уменьшается в 2 раза. Случайная величина X — число попаданий в цель при двух выстрелах. Описать закон распределения.

Решение. Возможные значения X: 0, 1, 2. $P(X = 0) = 0.2 \cdot 0.6 = 0.12$, $P(X = 1) = 0.2 \cdot 0.4 + 0.8 \cdot 0.6 = 0.56$, $P(X = 2) = 0.8 \cdot 0.4 = 0.32$.

Задача 4.1. Два игрока, поочередно извлекают шары (без возвращения) из урны, содержащей 2 белых и 4 черных шара. Выигрывает тот, кто первым вынет белый шар. Найти вероятность выигрыша участника, начавшего игру.

Решение. Пусть Б означает событие, состоящее в том, что игрок вытащил белый шар, Ч — черный шар. Первый игрок выиграет при следующих раскладах в последовательностях испытаний: Б, ЧЧБ, ЧЧЧЧБ. Вероятность P выигрыша первым игроком равна:

$$P = \frac{2}{6} + \frac{4}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} + \frac{4}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} \cdot \frac{2}{2} = \frac{3}{5}.$$

Задача 4.5. При передаче сообщения вероятность искажения одного знака равна $1/100$. В предположении независимости искажений знаков найти вероятность того, что сообщение из 5 знаков: а) не будет искажено; б) содержит ровно одно искажение; в) содержит хотя бы два искажения.

Решение. Пусть A обозначает событие, состоящее в том, что произошло искажение. $P(A) = p = 0.01,\ q = 0.99$.

а) $P_5(0) = C_5^0 p^0 q^5 = 0.99^5 = 0.951;$
б) $P_5(1) = C_5^1 p^1 q^4 = 5 \cdot 0.01\ 0.99^4 \approx 0.048;$
в) $P_5(m \geq 2) = 1 - P_5(0) - P_5(1) \approx 0.001.$

Задача 4.19. Вероятность попадания в цель при каждом выстреле равна 0.001. Найти вероятность попадания в цель двумя и более выстрелами при залпе в 5000 выстрелов.

Решение. $p = 0.001,\ n = 5000$. Применяем предельную теорему Пуассона. $\lambda = p \cdot n = 5$.

$$P_n(0) \approx e^{-\lambda},\ P_n(1) = \lambda \cdot e^{-\lambda},\ P_n(m \geq 2) = 1 - P_n(0) - P_n(1) \approx$$

$$\approx 1 - (1+\lambda)e^{-\lambda} = 1 - 6e^{-\lambda} \approx 1 - 0.04043 = 0.95957.$$

Задача 4.27. Используя таблицу случайных чисел, выписать реализацию бросаний симметрической монеты. Длину реализации n подобрать так, чтобы частота выпадений герба отличалась от $1/2$ не более чем на $\Delta = 0.1$, с вероятностью, примерно равной 0.9. По реализации вычислить частоту выпадения герба.

Решение. $p = 0.5.\ \ |\frac{m}{n} - p| < \Delta \rightarrow |\frac{m-np}{\sqrt{npq}}| < \Delta \cdot \sqrt{\frac{n}{pq}},$

$$P(|\frac{m}{n} - p| < \Delta) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\Delta \cdot \sqrt{\frac{n}{pq}}}^{\Delta \cdot \sqrt{\frac{n}{pq}}} e^{-\frac{t^2}{2}} dt = 2\Phi(\Delta \cdot \sqrt{\frac{n}{pq}}) = 0.9.$$

Из таблицы значений функции Лапласа находим, что

$$\Delta \cdot \sqrt{\frac{n}{pq}} \approx 1.64, \text{ поэтому } n = \left(\frac{1.64}{0.1} \cdot \frac{1}{2}\right)^2 = 8.2^2 \approx 67.$$

Пусть в таблице случайных чисел числа 0, 1, 2, 3, 4 означают, что при бросании монеты выпал герб. Если последовательно идти по строкам слева направо, то число выпадений герба окажется равным 36. Значит $p = 36/67 \approx 0.54$.

Задачи домашнего задания.

Задача 14.313 (продолжение). В условиях предыдущей задачи найти вероятности событий: $C =$\{хотя бы одно попадание\}, $D =$\{не менее трех попаданий\}.

Ответ: $P(C) \approx 0.7626$, $P(D) \approx 0.1035$.

Задача 14.322. Два равносильных шахматиста договорились сыграть матч из $2n$ результативных партий. Ничья не учитывается и считается, что каждый из участников может выиграть очередную партию с вероятностью 0.5. Выигравшим матч считается тот, кто победит в большем числе партий. В каком матче больше шансов выиграть любому из участников: в матче из 8 результативных партий или из 12?

Ответ: более вероятно выиграть матч из 12 партий.

Задача 14.343. Вероятность перегорания первой, второй и третьей лампы соответственно равна 0.1, 0.2 и 0.3. Если перегорает одна лампа, то прибор выходит из строя с вероятностью 0.5, а если две или три — то прибор заведомо выйдет из строя. Найти вероятность выхода прибора из строя.

Ответ: 0.297.

Задача 4.2. Два игрока поочередно извлекают шары (без возвращения) из урны, содержащей 2 белых шара, 4 черных и 1 красный. Выигрывает тот, кто первым вынет белый шар. Если появляется красный шар, то объявляется ничья. Пусть A_1=\{выиграет игрок, начавший игру\}, A_2= \{выиграет второй участник\}, B=\{ игра закончится вничью\}. Найти $P(A_1)$, $P(A_2)$, $P(B)$.

Ответ: $P(A_1) = \frac{44}{105} = 0.4190...$, $P(A_2) = \frac{26}{105} = 0.2476...$, $P(B) = \frac{1}{3}$.

Задача 4.4. Проведено 10 независимых испытаний, каждое из которых заключается в одновременном подбрасывании трех игральных костей. Найти вероятность того, что в четырех испытаниях появятся в точности по две "6".

Ответ: $C_{10}^4(\frac{5}{72})^4(1-\frac{5}{72})^6 = 0.00317...$

Задача 4.9. Найти вероятность того, что в n испытаниях схемы Бернулли с вероятностью успеха p появится $m+l$ успехов, причем l успехов появится в последних l испытаниях.

Ответ: $C_{n-l}^m p^{m+l}(1-p)^{n-m-l}$.

Задача 4.22. Из таблицы случайных чисел отбирают числа, делящиеся на 4, до тех пор, пока не наберется 588. Найти приближенное значение вероятности того, что потребуется таблица, содержащая не меньше 2100 чисел.

У к а з а н и е. Пусть n − число чисел в требуемой таблице; μ_n — число чисел, делящихся на 4, среди n чисел таблицы ($n = 2100$). Использовать равенство $P(\nu \geq n) = P(\mu_n \leq 588)$.

Ответ: 0.0228.

Задача 4.28. Используя таблицу случайных чисел, выписать N реализаций блуждания частицы (определение случайного блуждания частицы дано на стр. 97) по целым точкам отрезка $[0, n]$ до её поглощения в 0 или в n. Величины N, n, k — начальное положение частицы, p — вероятность перехода вправо, $q = 1 - p$ выбрать равными:

а) $N = 10$, $n = 6$, $p = 1/6$, $k = 1$;

б) $N = 10$, $n = 6$, $p = 1/2$, $k = 1$.

Вычислить частоту поглощения частицы в 0 и в n; сравнить с вероятностями соответствующих поглощений.

2.4 Случайные величины

Определение. Пусть дано некоторое вероятностное пространство $(\Omega, \mathbf{A}, P)$. Действительная функция $\xi = \xi(\omega)$ на Ω называется случайной величиной, если $\forall x \in R$ множество $\{\omega: \ \xi(\omega) < x\}$ является событием из алгебры событий $\mathbf{A}$, т.е. $\{\omega: \ \xi(\omega) < x\} \in \mathbf{A}$. Вероятность $P(\xi(\omega) < x)$ называется функцией распределения случайной величины ξ и обозначается $F(x)$ или $F_\xi(x)$, если есть необходимость подчеркнуть, что речь идет о функции распределения именно случайной величины $\xi: \ F_\xi(x) = F(x) = P(\xi(\omega) < x)$.

Свойства функции распределения:

1) $F(x_1) \le F(x_2)$, если $x_1 < x_2$,

2) $F(-\infty) = 0$, $F(\infty) = 1$,

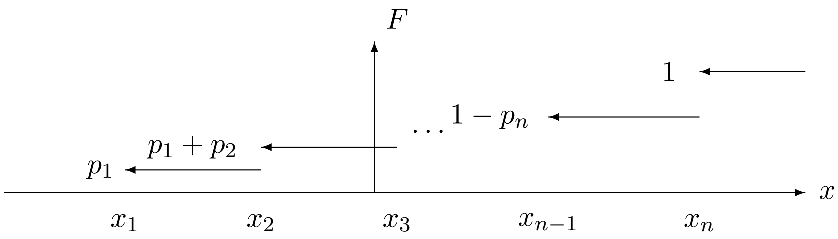
3) $\lim\limits_{x \to x_0 - 0} F(x) = F(x_0)$ (непрерывность слева). На самом деле $F(x_0) - F(x) = P(\xi < x_0) - P(\xi < x) = P(x \le \xi < x_0) \to 0$ при $x \to x_0, \ x < x_0$.

Если закон распределения случайной величины задан, то и сама случайная величина считается заданной, так как для любой функции распределения $F(x)$ со сформулированными свойствами всегда можно найти такое вероятностное пространство $(\Omega, \mathbf{A}, P)$ и функцию $\xi = \xi(\omega)$, что имеет место $P(\xi(\omega) < x) = F(x)$.

Дискретные и непрерывные случайные величины. Случайная величина ξ называется дискретной, если существуют числа $x_1, x_2, ..., x_n$ (конечное число или счетное множество), такие, что $P(\xi = x_k) = p_k \ne 0$ и $\sum\limits_k p_k = 1$. Для характеристики дискретной случайной величины используется так называемый ряд распределения:

$$\begin{array}{c|cccc} \xi & x_1 & x_2 & \cdots & x_n \\ \hline p & p_1 & p_2 & \cdots & p_n \end{array}.$$

График функции распределения дискретной случайной величины имеет ступенчатый характер:



Примеры:

1) *Биномиальное распределение*:

$$x_k = k,\ k = 0, 1, ..., n;\ p_k = C_n^k p^k (1-p)^{n-k};\ 0 < p < 1.$$

2) *Распределение Пуассона с параметром* λ:

$$x_k = k,\ k = 0, 1, ..., n, ...;\ p_k = \frac{\lambda^k}{k!} e^{-\lambda}\ \lambda > 0.$$

Случайная величина ξ называется непрерывной, если функция распределения допускает представление в виде $F(x) = \int\limits_{-\infty}^{x} p(u)du$, где $p(u) \geq 0$. Функция $p(x)$ называется плотностью распределения вероятности. Если $p(x)$ непрерывна, то $F'(x) = p(x)$. Очевидно, что $\int\limits_{-\infty}^{\infty} p(x)dx = 1$. Имеет место

$$P(a < \xi < b) = \int_a^b p(x)dx.$$

Примеры: 1) *Равномерное распределение на сегменте* $[a, b]$:

$$p(x) = \left\{ \begin{array}{l} \frac{1}{b-a},\ x \in [a, b] \\ 0,\ x \notin [a, b] \end{array} \right. .$$

2) *Нормальное (гауссовское) распределение с параметрами* (a, σ^2):

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Если параметры имеют вид (0,1), то распределение называется нормально стандартным.

3) *Показательное распределение*:

$$p(x) = \lambda e^{-\lambda x},\ x > 0,\ \lambda > 0.$$

Система случайных векторов. Случайный вектор.

а) *Дискретный случайный вектор* (ξ, η). Если в результате опыта появляются два числа (x_k, y_l), $k = 1, 2, ...; l = 1, 2, ...$, такие, что $P(\xi = x_k, \eta = y_l) = p_{kl} > 0$ и $\sum\limits_{k=1,l=1} p_{kl} = 1$, то говорят, что задан двумерный случайный вектор.

Пример. 1) Пусть в (ξ, η) ξ означает случайное число испытаний по схеме Бернулли, подчиненное распределению Пуассона с параметром λ, а η есть число успехов при $\xi = n$ испытаниях с вероятностью успеха p в отдельном испытании. Тогда

$$P(\xi = k, \eta = m) = p_{km} = P(\xi = k)P(\eta = m|\xi = k) = \frac{\lambda^k e^{-\lambda}}{k!} C_k^m p^m q^{k-m},$$

где $0 \le k,\ 0 \le m \le k$. Покажем, что сумма всех вероятностей равна единице:

$$\sum_{k=0}^{\infty} \sum_{m=0}^{k} p_{km} = \sum_{k=0}^{\infty} \sum_{m=0}^{k} \frac{\lambda^k e^{-\lambda}}{k!} C_k^m p^m q^{k-m} = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} \sum_{m=0}^{k} C_k^m p^m q^{k-m}$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1.$$

2) Пусть в (ξ, η) ξ представляет случайное число появления герба при первом бросании монеты, а η обозначает случайное число появления герба при втором бросании монеты. Если при бросании монеты появляется герб, то ξ и η будут принимать значение 1, если появляется цифра, то значения ξ и η будут равны 0. Возможные значения для (ξ, η): (0,0), (0,1), (1,0), (1,1). Во всех случаях вероятности равны $\frac{1}{4}$.

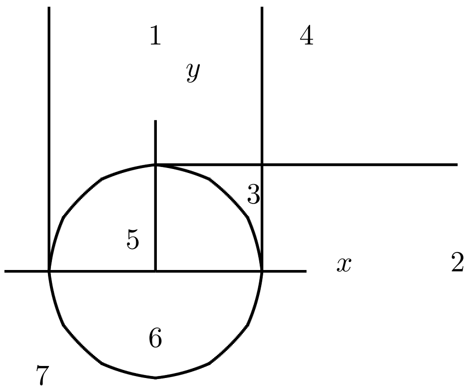
Функция распределения определяется следующим образом: $F(x,y)=P(\xi<x,\eta<y)$. Её свойства:

1) $0\le F\le 1$,
2) неубывающая функция по каждому аргументу,
3) $F(-\infty,y)=F(x,-\infty)=0,\ F(\infty,y)=F_\eta(y),\ F(x,\infty)=F_\xi(x)$.

б) *Непрерывный случайный вектор* (ξ,η). Говорят, что совокупность двух случайных векторов ξ и η образует непрерывный случайный вектор (ξ,η), если их совместная функция распределения $F(x,y)=P(\xi<x,\eta<y)$ допускает представление в виде $F(x,y)=\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y}p(u,v)dudv$. Неотрицательная подынтегральная функция $p(x,y)$ снова называется плотностью вероятности. $p(x,y)=\frac{\partial^2F}{\partial x\partial y}$.

Пример. Пусть (ξ,η) равномерно распределена по кругу $\xi^2+\eta^2\le R^2$. Найти $F(x,y),\ p(x,y)$.

Плоскость переменных x,y разбиваем на 7 частей согласно рисунку



Пусть $x^2+y^2\le R^2$. Тогда площадь области 5 $\{(u,v):\ u^2+v^2\le R^2,\ u\le x,\ v\le y,\ y\ge 0\}$ равна

$$\int\limits_{-R}^{-\sqrt{R^2-y^2}}du\int\limits_{-\sqrt{R^2-u^2}}^{\sqrt{R^2-u^2}}dv+\int\limits_{-\sqrt{R^2-y^2}}^{x}du\int\limits_{\sqrt{R^2-u^2}}^{y}dv=$$

$$= y(x + \sqrt{R^2 - y^2}) +$$

$$+ \int\limits_{-R}^{-\sqrt{R^2-y^2}} \sqrt{R^2 - u^2} du + \int\limits_{-R}^{x} \sqrt{R^2 - u^2} du =$$

$$= y(x + \sqrt{R^2 - y^2}) +$$

$$+ \frac{R^2}{2}\left(\pi + \arcsin\frac{x}{R} - \arcsin\frac{\sqrt{R^2 - y^2}}{R} + \frac{x}{R}\sqrt{1 - \frac{x^2}{R^2}} - \frac{y}{R}\sqrt{1 - \frac{y^2}{R^2}}\right) =$$

$$= xy + \frac{R^2}{2}(\pi + g(x) + g(y)),$$

где

$$g(x) = \arcsin\frac{x}{R} + \frac{x}{R}\sqrt{1 - \frac{x^2}{R^2}}.$$

В случае $y < 0$ эта площадь равна

$$y(x + \sqrt{R^2 - y^2}) +$$

$$+ \frac{R^2}{2}\left(\arcsin\frac{x}{R} + \arcsin\frac{\sqrt{R^2 - y^2}}{R} + \frac{x}{R}\sqrt{1 - \frac{x^2}{R^2}} + \frac{|y|}{R}\sqrt{1 - \frac{y^2}{R^2}}\right) =$$

$$= xy + \frac{R^2}{2}(\pi + g(x) + g(y)),$$

Для нахождения функции распределения найденные площади нужно разделить на площадь круга πR^2. Откуда $p(x,y) = F''_{xy} = \frac{1}{\pi R^2}$, если $x^2 + y^2 \leq R^2$. Для других областей площади равны: 1 — $R^2(\frac{\pi}{2} + g(x))$; 2 — $R^2(\frac{\pi}{2} + g(y))$; 3 — $R^2(g(x) + g(y))$; 4 — πR^2; 7 — 0.

в) *Независимость*. Если $F_{\xi\eta}(x,y) = F_\xi(x)F_\eta(y)$, то случайные величины ξ и η называются независимыми. Если (ξ, η) непрерывный случайный вектор, то условие независимости имеет вид: $p_{\xi\eta}(x,y) = p_\xi(x)p_\eta(y)$. Другое эквивалентное определение независимости: ξ и η называются независимыми, если для $\forall\, B_1, B_2 \subset R$ имеет место $P(\xi \in B_1, \eta \in B_2) = P(\xi \in B_1)P(\eta \in B_2)$.

Примеры. 1) Пусть $\Omega = \{1,2,3,4,5,6\}$ — множество элементарных событий, описывающих опыт с бросанием игральной кости.

Пусть случайная величина ξ равна нулю, если значение выпавших очков меньше или равно 4, и равна 1 в остальных случаях. Для η: 0 — если значение меньше или равно 2, 1 — в остальных случаях.

Можно записать следующую таблицу:

ξ	0	1	ряд распределения	ξ	0	1
	$\underbrace{1,2,3,4}$	$\underbrace{5,6}$		p	$2/3$	$1/3$
η	0	1	ряд распределения	η	0	1
	$\underbrace{1,2,}$	$\underbrace{3,4,5,6}$		p	$1/3$	$2/3$

Совместный ряд распределения имеет вид:

$\eta\backslash\xi$	0	1	
0	$1/3$	0	$1/3$
1	$1/3$	$1/3$	$2/3$
	$2/3$	$1/3$	

.

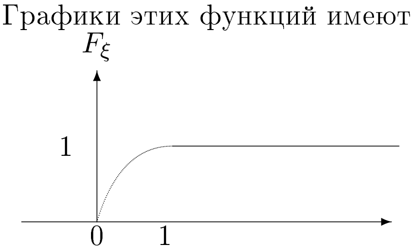
Случайные величины ξ и η зависимы.

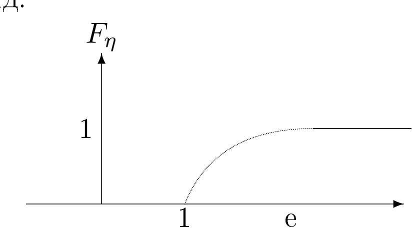
2) Пусть $\Omega = [0,1]$. Событиями назовем подмножества, имеющие длину, вероятность события считаем равной этой длине: $P(A) = |A|$. Координаты точек пространства Ω обозначим $u: \; 0 \le u \le 1$. Пусть $\xi(u) = u^2$, а $\eta(u) = e^u$. Имеем

$$F_\xi(x) = P(\xi < x) = P(u : u^2 < x) = \begin{cases} 0, \text{ если } x \le 0 \\ \sqrt{x}, \text{ если } 0 < x \le 1 \\ 1, \text{ если } 1 < x \end{cases},$$

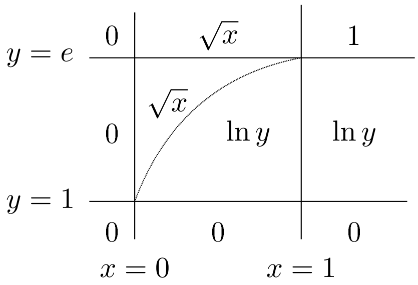
$$F_\eta(x) = P(\eta < x) = P(u : e^u < x) = \begin{cases} 0, \text{ если } x \le 1 \\ \ln x, \text{ если } 1 < x \le e \\ 1, \text{ если } 1 < x \end{cases}.$$

Графики этих функций имеют вид:





Значения совместной функции распределения $F_{\xi\eta}(x,y)$ можно представить на основе рисунка:



Так как $F_{\xi\eta}(x,y) \neq F_\xi(x)F_\eta(y)$, то случайные величины ξ и η являются зависимыми между собой случайными величинами.

3) *О вероятности* $P((\xi,\eta)\in B)$ *для непрерывного* (ξ,η). Функция распределения для двумерного непрерывного случайного вектора (ξ,η) имеет вид: $F(x,y) = P(\xi < x, \eta) < y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} p(u,v)dudv$, откуда следует

$$P(x \leq \xi < x+\Delta x, \eta < y) = P(\xi < x+\Delta x, \eta < y) - P(\xi < x, \eta < y) =$$

$$= F(x+\Delta x, y) - F(x,y) = \int\limits_{x}^{x+\Delta x}\int\limits_{-\infty}^{y} p(u,v)dudv.$$

Аналогичным образом находим

$$P(x \leq \xi < x+\Delta x, y \leq \eta < y+\Delta y) = P(x \leq \xi < x+\Delta x, \eta < y+\Delta y) -$$

$$-P(x \leq \xi < x+\Delta x, \eta < y) = \int\limits_{x}^{x+\Delta x}\int\limits_{y}^{y+\Delta y} p(u,v)dudv \approx p(x,y)\Delta x\Delta y.$$

Поэтому

$$P((\xi,\eta)\in B) \approx \sum_{i,j} p(x_i,y_j)\Delta x_i \Delta y_j \to \iint\limits_{B} p(u,v)dudv.$$

4) *Многомерное нормальное распределение.* Будем говорить, что независимые случайные величины $\xi_1, ..., \xi_r$ имеют многомерное нормальное распределение, если каждая из этих величин имеет одномерное нормальное распределение. Пусть, например, $\xi_k,\ k = 1, 2..., r$, имеют нормальное распределение с параметрами (a_k, σ_k^2). Тогда

$$p_{\xi_1,...,\xi_r}(x_1, ..., x_r) = \frac{1}{(2\pi)^{r/2}\sigma_1 \cdots \sigma_r} \exp\{-\frac{1}{2}\sum_{k=1}^{r}\frac{(x_k - a_k)^2}{\sigma_k^2}\}.$$

2.4.1 5-ое практическое занятие. Плотность и функция распределения.

Задача 14.361. Случайная величина X нормально распределена с параметрами $a = 1$, $\sigma = 2$. Выразить ее функцию распределения через функцию $\Phi(x)$.

Решение.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt,$$

$$F_X(x) = P(X < x) = \frac{1}{\sqrt{8\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(t-1)^2}{8}} dt =$$

$$= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{x-1}{2}} e^{-\frac{u^2}{2}} du = \Phi(\frac{x-1}{2}),\ u = \frac{t-1}{2}.$$

Задача 14.372. Случайная величина X подчиняется нормальному закону с параметрами $(a = 1, \sigma^2)$. Известно, что $P(X < 2) = 0.99$. Вычислить $M[X^2]$ и $P(X^2 > 2)$.

Решение. $MX = 1$, $DX = \sigma^2$, $DX = MX^2 - (MX)^2$, $MX^2 = \sigma^2 + 1$. $P(X < 2) = \Phi(\frac{2-1}{\sigma}) = \Phi(\frac{1}{\sigma}) = 0.99$, $\frac{1}{\sigma} = 2.33$, $\sigma = \frac{1}{2.33} \approx 0.43$, $MX^2 = 1.18$, $P(X^2 > 2) = P(X < -\sqrt{2}) + P(X > \sqrt{2}) = \Phi(\frac{-\sqrt{2}-1}{0.43}) + 1 - \Phi(\frac{\sqrt{2}-1}{0.43})$.

Задача 14.378. Закон распределения случайного вектора (X, Y) дискретного типа определяется следующей таблицей:

$x_i \setminus y_j$	-1	0	1
1	0.15	0.3	0.35
2	0.05	0.05	0.1

а) Найти безусловные законы распределения отдельных компонент X и Y. б) Установить, зависимы или нет компоненты X и Y? в) Вычислить вероятности $P(X = 2, Y = 0)$ и $P(X > Y)$.

Решение. а) $P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = -1) = (X = 1, Y = -1) = 0.8$, аналогичным образом подсчитываем остальные вероятности и приходим к следующим рядам распределений:

x_i	1	2
p_i	0.8	0.2

y_j	-1	0	1
p_j	0.2	0.35	0.45

б) зависимы, так как, например, $P(X = 1, Y = -1) = 0.15 \neq$ $P(X = 1)P(Y = -1) = 0.16$. в) $P(X = 2, Y = 0) = 0.05$, $P(X > Y) = 1 - 0.35 = 0.65$.

Задача 5.3. Случайная величина ξ имеет показательное распределение с плотностью распределения $p_\xi(x) = \alpha e^{-\alpha x}$ $(x > 0)$. Найти плотности распределения случайных величин: а) $\eta_1 = \sqrt{\xi}$; б) $\eta_2 = \xi^2$; в) $\eta_3 = \frac{1}{\alpha}\ln\xi$; г) $\eta_4 = 1 - e^{-\alpha\xi}$.

Решение. $F_\xi(x) = \int\limits_0^x p(t)dt = 1 - e^{-\alpha x}$.

$$F_{\eta_1}(x) = P(\sqrt{\xi} < x) = P(\xi < x^2) = F_\xi(x^2),\ p_{\eta_1}(x) = 2x\alpha e^{-\alpha x^2}, (0 < x),$$

$$F_{\eta_2}(x) = P(\xi^2 < x) = P(\xi < \sqrt{x}) = F_\xi(\sqrt{x}),\ p_{\eta_2}(x) = \frac{\alpha}{2\sqrt{x}} e^{-\alpha\sqrt{x}},$$

$$(0 < x),$$

$$F_{\eta_3}(x) = P(\frac{1}{\alpha}\ln\xi < x) = P(\xi < e^{\alpha x}) = F_\xi(e^{\alpha x}),$$

$$p_{\eta_3}(x) = \alpha^2 e^{\alpha(x - e^{\alpha x})},\ (-\infty < x < \infty),$$

$$F_{\eta_4}(x) = P(1 - e^{-\alpha\xi} < x) = P(e^{-\alpha\xi} > 1 - x) = P(\xi < -\frac{ln(1-x)}{\alpha}) =$$

$$= F_\xi(-\frac{ln(1-x)}{\alpha}),\ p_{\eta_4}(x) = 1,\ 0 \le x \le 1.$$

Задача 5.9. Случайные величины ξ_1 и ξ_2 независимы и нормально распределены с параметрами (0,1). Найти плотности распределения величин: а) $\eta_1 = \xi_1^2 + \xi_2^2$; б) $\eta_2 = \arctan \xi_1/\xi_2$; в) совместную плотность распределения (η_1, η_2).

Решение.

$$p_{\xi_1}(x) = p_{\xi_2}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$F_{\eta_1}(x) = P(\xi_1^2 + \xi_2^2 < x) = \frac{1}{2\pi} \iint\limits_{u_1^2+u_2^2<x} e^{-\frac{u_1^2+u_2^2}{2}} du_1 du_2 =$$

$$= \frac{1}{2\pi} \int\limits_0^{\sqrt{x}} d\rho \int\limits_0^{2\pi} \rho e^{-\frac{\rho^2}{2}} d\varphi =$$

$$= 1 - e^{-x/2},\ u_1 = \rho \cos\varphi,\ u_2 = \rho \sin\varphi,\ p_{\eta_1}(x) = \frac{1}{2} e^{-x/2},\ x > 0.$$

$$F_{\eta_2}(x) = P(\arctan \xi_1/\xi_2 < x) = P(\xi_1/\xi_2 < \text{tg } x) =$$

$$= \frac{1}{2\pi} \iint\limits_{u_1/u_2<\text{tg } x} e^{-\frac{u_1^2+u_2^2}{2}} du_1 du_2 = \frac{1}{2\pi} \int\limits_0^{\infty} \rho e^{-\frac{\rho^2}{2}} d\rho \int\limits_{\text{ctg } \varphi<\text{tg } x} d\varphi =$$

$$= \frac{1}{2\pi} \int\limits_{\text{tg } (\pi/2-\varphi)<\text{tg } x} d\varphi =$$

$$= \frac{x}{\pi} + \frac{1}{2},\ (-\frac{\pi}{2} < x < \frac{\pi}{2}),\ p_{\eta_2}(x) = \frac{1}{\pi}.$$

$$F_{\eta_1\eta_2}(x_1, x_2) = P(\eta_1 < x_1, \eta_2 < x_2) =$$

$$= \frac{1}{2\pi} \iint\limits_{\substack{u_1^2+u_2^2<x \\ u_1/u_2<\text{tg } x}} e^{-\frac{u_1^2+u_2^2}{2}} du_1 du_2 =$$

$$= \frac{1}{2\pi} \int\limits_0^{\sqrt{x}} \rho e^{-\frac{\rho^2}{2}} d\rho \int\limits_{\text{ctg } \varphi<\text{tg } x} d\varphi = F_{\eta_1}(x_1) F_{\eta_2}(x_2).$$

Задача 5.17. Игральная кость бросается до тех пор, пока впервые не выпадет меньше пяти очков. Обозначим через θ число очков, выпавших при последнем бросании игральной кости, а через ν − число бросаний кости. Найти совместное распределение θ и ν. Являются ли θ и ν независимыми?

Решение. Возможные значения $\nu : n = 1, 2, ...$; возможные значения $\theta : m = 1, 2, 3, 4$. $p = 4/6$, $q = 2/6$. $P(\nu = n, \theta = m) = P(\nu = n)P(\theta = m/\nu = n) = q^{n-1}p/4 = 1/(2 \cdot 3^n)$, $P(\nu = n) = q^{n-1}p$, $P(\theta = m) = 1/4$, $P(\nu = n, \theta = m) = P(\nu = n)P(\theta = m)$.

Задача 5.18. На n станках одновременно началась обработка n деталей. Предполагая, что времена обработки деталей независимы и имеют показательные распределения с параметром α, найти распределение времени: а) до получения первой обработанной детали; б)до окончания обработки всех деталей.

Решение. а) Пусть ξ — случайное время обработки детали отдельным станком, тогда $F(x) = P(\xi < x)$ — вероятность того, что деталь за время x будет сделана. $P(\xi > x) = 1 - F(x)$ — вероятность того, что деталь за время x не будет сделана. Пусть η — случайное время до получения первой обработанной детали, тогда $(\eta > x)$ представляет событие, состоящее в том, что ни одна деталь не будет сделана за время x. $P(\eta > x) = (1 - F(x))^n = e^{-n\alpha x}$, $P(\eta < x) = 1 - P(\eta > x) = 1 - e^{-n\alpha x}$, $p_\eta(x) = n\alpha e^{-n\alpha x}$.

б) Пусть ζ — случайное время до окончания обработки всех деталей. Тогда $P(\zeta < x) = (F(x))^n$, $p_\zeta(x) = nF^{n-1}(x)p_\zeta(x) = n\alpha(1 - e^{-\alpha x})^{n-1}e^{-\alpha x}$.

Задача 5.20. Машина состоит из 10 000 деталей. Каждая деталь независимо от других оказывается неисправной с вероятностью p_i, причем для $n_1 = 1000$ деталей $p_1 = 0.0003$; для $n_2 = 2000$ деталей $p_2 = 0.0005$, и для $n_3 = 7000$ деталей $p_3 = 0.0001$. Машина не работает, если в ней неисправны хотя бы две детали. Найти приближенное значение вероятности того, что машина не будет работать.

Указание. Воспользоваться теоремой Пуассона.

Решение. Пусть A_k означает событие, состоящее в том, что не исправны k деталей 1-го рода; B_k — 2-го рода; C_k — 3-го рода. Событие $A = A_0B_0C_0 + A_0B_0C_1 + A_0B_1C_0 + A_1B_0C_0$ состоит в том, что машина работает. $P(A) = e^{-(\lambda_1+\lambda_2+\lambda_3)}(1+\lambda_1+\lambda_2+\lambda_3)$, $\lambda_1 = 0.3$, $\lambda_2 = 1$, $\lambda_3 = 0.7$. $P(A) = 3e^{-2} = 0.406$. $P(\bar{A}) = 0.594$.

Задача 5.23. Используя таблицу случайных чисел, выписать 10 реализаций испытаний, описанных в задаче 5.17. Найти частоты событий: $(\theta = k)$, $k = 1, 2, 3, 4$; $(\nu = l)$, $l = 1, 2, 3, \ldots$.

Решение. В таблице случайных чисел опускаем числа 0, 7, 8, 9. Получаем следующие реализации случайного вектора (θ, ν) : (1,1), (1,3), (1,2), (2,3), (1,3), (2,2), (1, 1), (1,3), (2,3), (1,4). Частоты: $\nu = 1$ — 0.7; $\nu = 2$ — 0.3; $\theta = 1$ — 0.2; $\theta = 2$ — 0.2; $\theta = 3$ — 0.5; $\theta = 4$ — 0.1.

Задачи домашнего задания.

Задача 14.362. Случайная величина X распределена по закону $N(m, \sigma^2)$. Пользуясь таблицей функции нормального распределения, вычислить вероятность p_k того, что отклонение величины X от ее математического ожидания не превзойдет величины $k\sigma$ (ответ получить для трех значений $k = 1, 2, 3$).

Ответ: $p_1 \approx 0.683$, $p_2 \approx 0.954$, $p_3 \approx 0.997$.

Задача 14.373. Случайная величина X распределена по закону $N(m, \sigma^2)$. Вычислить $p_1 = P(X \geq x_{n2})$ и $p_2 = P(x_{n1} \leq X \leq x_{n2})$, где x_{n1} и x_{n2} — точки перегиба плотности распределения вероятностей.

Ответ: $p_1 = 0.1587$, $p_2 = 0.68$.

Задача 14.379. (продолжение **Задачи 14.378**). Построить функцию распределения $F_{X,Y}(x, y)$, оформив результат в виде таблицы, и найти m_X и m_Y.

Ответ:

$x \setminus y$	$y \leq -1$	$-1 < y \leq 0$	$0 < y \leq 1$	$1 < 0$
$x \leq 1$	0	0	0	0
$1 < x \leq 2$	0	0.15	0.45	0.8
$2 < x$	0	0.2	0.55	1

$m_X = 1.2, m_Y = 0.25$

Задача 5.4. Случайная величина ξ распределена нормально с параметрами $a = 1$, $\sigma^2 = 1$. Найти плотности распределения величин: а) $\eta_1 = \xi^2$; б) $\eta_2 = e^{\xi}$ (логарифмическое нормальное распределение).

Ответ: а) $\frac{1}{\sqrt{2\pi x}}e^{-x/2}(x>0)$; б) $\frac{1}{x\sqrt{2\pi}}e^{-\ln^2 x/2}$.

Задача 5.10. Найти совместное распределение величин $\eta_1 = \sqrt{-2\ln\xi_1}\cos(2\pi\xi_2)$, $\eta_2 = \sqrt{-2\ln\xi_1}\sin(2\pi\xi_2)$, где ξ_1 и ξ_2 независимы и имеют равномерное распределение на сегменте [0,1].

Ответ: $p_{\eta_1,\eta_2}(x_1,x_2) = \frac{1}{2\pi}e^{-(x_1^2+x_2^2)/2}$.

Задача 5.14. Обозначим τ число испытаний в схеме Бернулли до появления первого успеха включительно. Найти закон распределения τ.

Ответ: $P(\tau = k) = (1-p)^{k-1}p \ (k \geq 1)$.

Задача 5.21. Совместное распределение величин ξ_1, ξ_2 является равномерным в круге $\{(x_1,x_2): \ x_1^2+x_2^2 \leq 1\}$. Найти вероятность

$$P\left\{|\xi_1| < \frac{1}{\sqrt{2}}, |\xi_2| < \frac{1}{\sqrt{2}}\right\}.$$

Являются ли величины ξ_1, ξ_2 независимыми?

Ответ: $\frac{2}{\pi} = 0.6366197\ldots$, величины ξ_1, ξ_2 зависимы.

Задача 5.24. Используя таблицу случайных чисел, выписать реализацию 50 независимых случайных величин, равномерно распределенных на сегменте [0,1]. Числовые значения взять с точностью до 10^{-3}.

2.5 Математическое ожидание

. **Определение**. а) Пусть ξ — дискретная одномерная случайная величина. В этом случае математическое ожидание $M\xi$ или среднее значение случайной величины ξ определяется следующим образом:

$$M\xi = \sum_k x_k p_k.$$

Примеры:

1) Пусть ξ есть случайная величина, представляющая случайное число успехов при однократном производстве опыта. Тогда

$$M\xi = 0 \cdot q + 1 \cdot p = p.$$

2) *Распределение Пуассона*: $x_k = k,\ k = 0, 1, 2, ...,\ p_k = P(\xi = x_k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

$$M\xi = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} =$$

$$= e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

б) Пусть ξ распределена непрерывно на сегменте $[a, b]$, т.е. является непрерывной случайной величиной, принимающей значения в сегменте $[a, b]$. Точками $x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b$ сегмент $[a, b]$ разделим на n частей. Пусть $\Delta x_k = x_{k+1} - x_k,\ 0 \le k \le n - 1$. Тогда $P(x_k \le \xi < x_{k+1}) = \int\limits_{x_k}^{x_{k+1}} p(u)du \approx p(x_k)\Delta x_k$. Непрерывную случайную величину ξ рассмотрим приближенно как дискретную с возможными значениями x_k и соответствующими вероятностями $p(x_k)\Delta x_k$ принять эти значения, тогда за приближенное значение математического ожидания $M\xi$ можно принять $M\xi \approx \sum\limits_k x_k p(x_k)\Delta x_k$, которое при $n \to \infty,\ \Delta x_k \to 0$ будет стремиться к $\int\limits_a^b xp(xdx)$ как интегральная сумма, поэтому математическое ожидание $M\xi$ непрерывной случайной величины ξ определяется следующим образом:

$$M\xi = \int_{-\infty}^{\infty} xp(x)dx.$$

Примеры.

1) *Равномерное распределение на сегменте* $[a, b]$:

$$p(x) = \begin{cases} \frac{1}{b-a},\ x \in [a, b] \\ 0,\ x \notin [a, b] \end{cases},$$

$$M\xi=\int\limits_{-\infty}^{\infty}xp(x)dx=\int\limits_{a}^{b}x\frac{1}{b-a}dx=\frac{1}{b-a}\frac{x^2}{2}\ \Big|_{a}^{b}=\frac{a+b}{2}.$$

2) *Нормальное распределение с параметрами* (a,σ^2):

$$p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

$$M\xi=\int\limits_{-\infty}^{\infty}xp(x)dx=\frac{1}{\sqrt{2\pi\sigma^2}}\int\limits_{-\infty}^{\infty}xe^{-\frac{(x-a)^2}{2\sigma^2}}dx=$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\int\limits_{-\infty}^{\infty}(a+\sigma t)e^{-\frac{t^2}{2}}\sigma dt=\frac{a}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt+\frac{\sigma}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}te^{-\frac{t^2}{2}}dt=a.$$

в) Пусть $\xi=\varphi(\eta_1,\eta_2)$, где (η_1,η_2) — дискретный случайный вектор. Это значит, что (η_1,η_2) на плоскости (x,y) пробегает точки (x_k,y_l) и $P(\eta_1=x_k,\eta_2=y_l)=p_{kl}$, возможными значениями ξ будут $\varphi(x_k,y_l)$, а вероятность $P(\xi=\varphi(x_k,y_l))$ снова будет равна p_{kl}. Математическое ожидание функции от дискретных случайных величин определяется следующим образом:

$$M\xi=\sum_{k,l}\varphi(x_k,y_l)p_{kl}.$$

Пример. *Математическое ожидание суммы для дискретных случайных величин.*

$$M(\xi+\eta)=\sum_{k,l}(x_k+y_l)p_{kl}=\sum_{k,l}(x_k+y_l)P(\xi=x_k,\eta=y_l)=$$

$$=\sum_{k,l}x_kP(\xi=x_k,\eta=y_l)+\sum_{k,l}y_lP(\xi=x_k,\eta=y_l)=$$

$$=\sum_{k}x_k\sum_{l}P(\xi=x_k,\eta=y_l)+\sum_{l}y_l\sum_{k}P(\xi=x_k,\eta=y_l)=$$

$$=\sum_{k}x_kP(\xi=x_k)+\sum_{l}y_lP(\eta=y_l)=M\xi+M\eta.$$

г) Пусть $\xi = \varphi(\eta_1, \eta_2)$, где (η_1, η_2) — непрерывный случайный вектор и возможные значения ξ есть $\varphi(x, y)$. Имеем

$$P(x_k \le \eta_1 < x_k + \Delta x_k, y_l \le \eta_2 < y_l + \Delta y_l) \approx p(x_x, y_l)\Delta x_k \Delta y_l,$$

значит за приближенное значение $M\xi$ можно принять

$$M\xi \approx \sum_{k,l} \varphi(x_k, y_l)\Delta x_k \Delta y_l,$$

где сумма представляет интегральную сумму для

$$\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \varphi(x, y)p(x, y)dxdy.$$

Поэтому в качестве математического ожидания для функции от непрерывного случайного вектора принимается следующее число:

$$M\xi = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \varphi(x, y)p(x, y)dxdy.$$

Пример. *Математическое ожидание суммы непрерывных случайных величин.* Пусть (ξ, η) — непрерывный случайный вектор, тогда

$$M(\xi + \eta) = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x + y)p(x, y)dxdy = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} xp(x, y)dxdy +$$

$$+ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} yp(x, y)dxdy = \int\limits_{-\infty}^{\infty} x(\int\limits_{-\infty}^{\infty} p(x, y)dy)dx + \int\limits_{-\infty}^{\infty} y(\int\limits_{-\infty}^{\infty} p(x, y)dx)dy =$$

$$= \int\limits_{-\infty}^{\infty} xp_\xi(x)dx + \int\limits_{-\infty}^{\infty} yp_\eta(y)dy = M\xi + M\eta.$$

Свойства математического ожидания.

1) $MC = C$, $C -$ постоянная,
2) $MC\xi = CM\xi$,
3)$|M\xi| \leq M|\xi|$,
4) $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$,
5) $M(\xi_1\xi_2) = M\xi_1 M\xi_2$, если ξ_1, ξ_2 независимы.
Докажем последнее свойство в предположении, что (ξ_1, ξ_2) — дискретный вектор:

$$M(\xi_1\xi_2) = \sum_{k,l} x_k y_l P(\xi_1 = x_k, \xi_2 = y_l) = \sum_{k,l} x_k y_l P(\xi_1 = x_k) P(\xi_2 = y_l) =$$

$$= \sum_k x_k P(\xi_1 = x_k) \sum_l y_l P(\xi_2 = y_l) = M\xi_1 M\xi_2.$$

В случае непрерывного случайного вектора (ξ_1, ξ_2) при доказательстве свойства 5) вместо сумм нужно использовать интегралы.

Дисперсия — это среднее значение квадрата отклонения случайной величины от её среднего значения.

$$D\xi = M(\xi - M\xi)^2 = M(\xi^2 - 2(M\xi)\xi + (M\xi)^2) =$$

$$= M\xi^2 - 2(M\xi)M\xi + (M\xi)^2 = M\xi^2 - (M\xi)^2.$$

Примеры.

1) *Распределение Пуассона*: $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Мы уже получили ранее, что $\xi = \lambda$. Вычислим теперь $M\xi^2$.

$$M\xi^2 = \sum_{k=0}^{\infty} x_k^2 p_k = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} =$$

$$= e^{-\lambda} \sum_{l=0}^{\infty} \frac{(l+1)\lambda^{l+1}}{l!} = \lambda e^{-\lambda} (\sum_{l=1}^{\infty} \frac{\lambda^l}{(l-1)!} + \sum_{l=0}^{\infty} \frac{\lambda^l}{l!}) = \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) =$$

$$= \lambda^2 + \lambda.$$

$$D\xi = M\xi^2 - (M\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

2) *Равномерное распределение на сегменте* $[a,b]$. Плотность вероятности имеет вид:

$$p(x) = \begin{cases} \frac{1}{b-a}, \ x \in [a,b] \\ 0, \ x \notin [a,b] \end{cases}.$$

Находим

$$M\xi = \int\limits_{-\infty}^{\infty} xp(x)dx = \int\limits_{a}^{b} x\frac{1}{b-a}dx = \frac{1}{b-a}\frac{x^2}{2}\bigg|_a^b = \frac{a+b}{2}.$$

Аналогично

$$M\xi^2 = \int\limits_{a}^{b} \frac{x^2}{b-a}dx = \frac{b^2+ab+a^2}{3},$$

поэтому

$$D\xi = \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{(a-b)^2}{12}.$$

3) *Нормальное распределение.*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{(x-a)^2}{2\sigma^2}},\ M\xi = a,$$

$$D\xi = M(\xi - M\xi)^2 = \int\limits_{-\infty}^{\infty} (x-a)^2 \frac{1}{\sqrt{2\pi\sigma^2}}\ e^{-\frac{(x-a)^2}{2\sigma^2}} dx.$$

Введем новую переменную интегрирования $t = \frac{x-a}{\sigma}$, тогда

$$D\xi = \frac{\sigma}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} t^2\sigma^2 e^{-\frac{t^2}{2}}dt = \frac{\sigma^2}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}}dt =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\left((-t)e^{-\frac{t^2}{2}}\bigg|_{-\infty}^{\infty} + \int\limits_{-\infty}^{\infty} e^{-\frac{t^2}{2}}dt\right) = \sigma^2.$$

Свойства дисперсии.

1) $D\xi \geq 0$,

2) $DC = 0$, $C -$ постоянная,

3) $DC\xi = C^2 D\xi$,

4) $D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2$, если ξ_1 и ξ_2 независимы.

Доказательство последнего свойства:

$$D(\xi_1+\xi_2) = M(\xi_1+\xi_2-M(\xi_1+\xi_2))^2 = M((\xi_1-M\xi_1)+(\xi_2-M\xi_2))^2 =$$

$$= M((\xi_1-M\xi_1)^2+(\xi_2-M\xi_2)^2+2(\xi_1-M\xi_1)(\xi_2-M\xi_2)) = D\xi_1+D\xi_2,$$

так как $M((\xi_1 - M\xi_1)(\xi_2 - M\xi_2)) = M(\xi_1 - M\xi_1)M(\xi_2 - M\xi_2) = (M\xi_1 - M\xi_1)(M\xi_2 - M\xi_2) = 0$.

Ковариация, коэффициент корреляции.

Ковариация $cov(\xi_1, \xi_2)$ между случайными величинами ξ_1 и ξ_2 определяется следующим образом:

$$cov(\xi_1, \xi_2) = M((\xi_1 - M\xi_1)(\xi_2 - M\xi_2)) =$$

$$= M(\xi_1\xi_2 - (M\xi_1)\xi_2 - (M\xi_2)\xi_1 + M\xi_1 M\xi_2) =$$

$$= M(\xi_1\xi_2) - M\xi_1 M\xi_2 - M\xi_2 M\xi_1 + M\xi_1 M\xi_2 =$$

$$= M(\xi_1\xi_2) - M\xi_1 M\xi_2.$$

Если аргументы у ковариации совпадают, то ковариация превращается в дисперсию:

$$cov(\xi, \xi) = D\xi.$$

С помощью ковариации можно написать формулу для дисперсии суммы:

$$D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2 + 2cov(\xi_1, \xi_2).$$

Если ξ_1 и ξ_2 независимы, то $cov(\xi_1, \xi_2) = 0$.

Из положительной определенности квадратичной формы $D(c_1\xi_1 + c_2\xi_2)$ относительно c_1 и c_2:

$$D(c_1\xi_1 + c_2\xi_2) = c_1^2 D\xi_1 + c_2^2 D\xi_2 + 2c_1c_2cov(\xi_1, \xi_2) \geq 0$$

следует, что

$$\begin{vmatrix} D\xi_1 & cov(\xi_1,\xi_2) \\ cov(\xi_1,\xi_2) & D\xi_2 \end{vmatrix} \geq 0,$$

то есть $(cov(\xi_1,\xi_2))^2 \leq D\xi_1 D\xi_2$.

Коэффициент корреляции определяется как

$$\rho(\xi_1,\xi_2) = \frac{cov(\xi_1,\xi_2)}{\sqrt{D\xi_1 D\xi_2}}.$$

Свойства коэффициента корреляции.

1) $|\rho(\xi_1,\xi_2)| \leq 1$,

2) $\rho(\xi_1,\xi_2) = 0$, если ξ_1 и ξ_2 независимы,

3) Пусть $\xi_2 = A\xi_1 + B$, тогда $\rho(\xi_1,\xi_2) = \operatorname{sign} A$.

На самом деле

$$cov(\xi_1,\xi_2) = cov(\xi_1, A\xi_1 + B) = Acov(\xi_1,\xi_1) = AD\xi_1,$$

$$D\xi_2 = A^2 D\xi_1,$$

откуда

$$\rho(\xi_1,\xi_2) = \frac{AD\xi_1}{\sqrt{D\xi_1 A^2 D\xi_1}} = \frac{A}{|A|}.$$

2.5.1 6-ое практическое ззанятие. Математическое ожидание

Задача 14.389. Из урны, содержащей 6 белых и 4 черных шара, наудачу извлекают 2 шара без возвращения. Случайные величины: X — число белых шаров в выборке, Y — число черных шаров в выборке. Описать закон распределения случайного вектора (X,Y) и вычислить ρ_{XY}.

Решение. Таблица распределения для случайного вектора (X,Y) имеет вид:

$X \setminus Y$	0	1	2
0	0	0	12/90
1	0	48/90	0
2	30/90	0	0

$MX = 1 \cdot \frac{48}{90} + 2 \cdot \frac{30}{90} = \frac{6}{5}$,

$MX^2 = 1 \cdot \frac{48}{90} + 4 \cdot \frac{30}{90} = \frac{28}{15}$,

$DX = \frac{28}{15} - \frac{36}{25} = \frac{32}{75}$.

$MY = 1 \cdot \frac{48}{90} + 2 \cdot \frac{12}{90} = -\frac{4}{5}$, $MY^2 = \frac{16}{15}$, $DY = \frac{32}{75}$, $M(XY) = \frac{48}{90}$, $cov(X,Y) = \frac{48}{90} - \frac{24}{25} = -\frac{32}{75}$, $\rho_{XY} = -1$.

Задача 14.397. Число X выбирается случайным образом из множества целых чисел $\{1, 2, 3\}$. Затем из того же множества выбирается наудачу число Y, большее первого или равное ему. Описать закон распределения случайного вектора (X,Y) и определить, зависимы или независимы случайные компоненты X и Y.

Решение. Таблица распределения для случайного вектора (X,Y) имеет вид:

$X \setminus Y$	1	2	3	
1	1/9	1/9	1/9	1/3
2	0	1/6	1/6	1/3
3	0	0	1/3	1/3
	1/9	5/18	11/18	

В этой таблице последний столбец представляет ряд распределения для случайной величины X, а последняя строка — ряд распределения для случайной величины Y. Случайные величины X и Y зависимы, так как совместные вероятности не равны произведению отдельных вероятностей.

Задача 6.1. Найти математическое ожидание величины τ, где τ — число испытаний в схеме Бернулли до появления первого успеха включительно.

Решение. $P(\tau = k) = pq^{k-1}$, $q = 1 - p$, $k = 1, 2, 3, ...$,

$$M\tau = \sum_{k=1}^{\infty} kP(\tau = k) = \sum_{k=1}^{\infty} kpq^{k-1} =$$

$$= p\sum_{k=1}^{\infty} kq^{k-1} = p(\sum_{k=1}^{\infty} q^k)' = p(\frac{q}{1-q})' = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

Задача 6.8. Из 100 карточек с числами 00, 01, 02, ..., 98, 99 наудачу вынимается одна. Пусть η_1, η_2 – соответственно сумма и произведение цифр на вынутой карточке. Найти $M\eta_1$, $M\eta_2$, $D\eta_1$, $D\eta_2$.

Решение. Пусть ξ_1 – случайная величина, значения которой совпадают с первой цифрой карты, ξ_2 – то же самое для второй цифры. $P(\xi_1 = k) = P(\xi_2 = k) = 0.1$, $k = 0, 1, ..., 9$. ξ_1 и ξ_2 независимы.

$$M\xi_1 = M\xi_2 = \sum_{k=0}^{9} kP(\xi_1 = k) = 0.1 \cdot 45 = 4.5,$$

$$M\xi_1^2 = \sum_{k=0}^{9} k^2 P(\xi_1 = k) = 28.5,$$

$$D\xi_1 = D\xi_2 = 8.25, \ M(\xi_1 + \xi_2) = 9, \ M\eta_2 = M(\xi_1\xi_2) = 20.25,$$

$$D\eta_1 = D(\xi_1 + \xi_2) = 16.5,$$

$$D\eta_2 = M\eta_2^2 - (M\eta_2)^2 = 28.5^2 - 20.25^2 = 402.1875.$$

Задача 6.11. Случайные величины ξ_1, ξ_2, ξ_3, ξ_4, ξ_5 независимы. $D\xi_1 = \sigma^2$. Найти коэффициент корреляции величин: а) $\xi_1 + \xi_2$ и $\xi_3 + \xi_4 + \xi_5$; б) $\xi_1 + \xi_2 + \xi_3$ и $\xi_3 + \xi_4 + \xi_5$.

Решение. $cov(\xi_1+\xi_2, \xi_3+\xi_4+\xi_5) = 0$, $cov(\xi_1+\xi_2+\xi_3, \xi_3+\xi_4+\xi_5) = cov(\xi_3, \xi_3) = D\xi_3 = \sigma^2$, $D = (\xi_1 + \xi_2 + \xi_3) = D(\xi_3 + \xi_4 + \xi_5) = 3\sigma^2$.

$$\rho(\xi_1 + \xi_2 + \xi_3, \xi_3 + \xi_4 + \xi_5) = \frac{\sigma^2}{\sqrt{3\sigma^2 \cdot 3\sigma^2}} = \frac{1}{3}.$$

Задача 6.24. Применим ли закон больших чисел к последовательности независимых случайных величин $\xi_1, \xi_2, \dots, \xi_k, \dots$, если $P(\xi_k = \sqrt{k}) = P(\xi_k = -\sqrt{k}) = \frac{1}{2\sqrt{k}}$, $P(\xi_k = 0) = 1 - \frac{1}{\sqrt{k}}$?

Решение.

$$M\xi_k = (-\sqrt{k}) \cdot \frac{1}{2\sqrt{k}} + 0 \cdot (1 - \frac{1}{\sqrt{k}}) + \sqrt{k} \cdot \frac{1}{2\sqrt{k}} = 0,$$

$$M\xi_k^2 = k \cdot \frac{1}{2\sqrt{k}} + 0 \cdot (1 - \frac{1}{\sqrt{k}}) + k \cdot \frac{1}{2\sqrt{k}} = \sqrt{k}, \ D\xi_k = \sqrt{k}.$$

$$\frac{\sum\limits_{k=1}^{n} D\xi_k}{n^2} = \frac{\sum\limits_{k=1}^{n} \sqrt{k}}{n^2} < \frac{n\sqrt{n}}{n^2} = \frac{1}{\sqrt{n}} \to 0 \text{ при } n \to \infty.$$

Закон больших чисел применим.

Задача 6.26. Доказать, что к последовательности $\xi_1, \xi_2, \dots, \xi_k, \dots$ применим закон больших чисел, если $|cov(\xi_k, \xi_l)| \le C$ для всех $k, l = 1, 2, \dots$ и $cov(\xi_k, \xi_l) \to 0$ при $|k - l| \to \infty$.

Решение. $\forall\ \epsilon > 0\ \exists\ K \to |cov(\xi_i, \xi_j)| < \epsilon$ если $|i - j| > K$.

$$\frac{1}{n^2} D(\xi_1 + \xi_2 + \dots + \xi_n) =$$

$$= \frac{1}{n^2} \left(\sum_{i=1}^{n} D\xi_i + 2 \sum_{\substack{i<j \\ j-i<K}} cov(\xi_i, \xi_j) + 2 \sum_{\substack{i<j \\ j-i \ge K}} cov(\xi_i, \xi_j) \right) \le$$

$\le \dfrac{1}{n^2}(C \cdot (n^2 - (n-K)^2) + \epsilon(n-K)^2) = (1 - \dfrac{K}{n})^2 \epsilon + C(\dfrac{2K}{n} - \dfrac{K^2}{n^2}) \to \epsilon$ при $n \to \infty$.

Задачи домашнего задания

Задача 14.390. Производится два выстрела по мишени в неизменных условиях. Вероятность попадания в мишень при одном выстреле равна p. Случайные величины: X — число выстрелов до первого попадания (включительно), Y — число промахов. Описать закон распределения случайного вектора (X, Y). Найти центр рассеяния данного распределения и значения σ_X^2 и σ_Y^2.

Ответ:

$x_i \setminus y_j$	0	1	2
1	p^2	pq	0
2	0	pq	q^2

$(m_X, m_Y) = (1 + q, 2q)$
$\sigma_X^2 = pq, \sigma_Y^2 = 2pq.$

Задача 14.398. В условиях задачи 14.397 найти коэффициент корреляции ρ_{XY}.

Ответ: $\rho_{XY} = \sqrt{6/17} \approx 0.594.$

Задача 6.2. У человека в кармане n ключей, из которых только один подходит к его двери. Ключи последовательно извлекаются (без возвращения) до тех пор, пока не появится нужный ключ. Обозначим через ξ номер испытания, в котором появился нужный ключ. Найти $M\xi$.

Ответ: $(n+1)/2$.

Задача 6.9. Совместное распределение случайных величин ξ_1 и ξ_2 задано таблицей:

$\xi_1 \setminus \xi_2$	-1	0	1
-1	$1/8$	$1/12$	$7/24$
1	$5/24$	$1/6$	$1/8$

Найти $M\xi_1$, $M\xi_2$, $D\xi_1$, $D\xi_2$, $cov(\xi_1, \xi_2)$.

Ответ: $M\xi_1 = 0$, $M\xi_2 = 1/12$, $D\xi_1 = 1$, $D\xi_2 = 107/144$,
$cov(\xi_1, \xi_2) = -1/4$.

Задача 6.12. Пусть $(\xi_1, \xi_2, ..., \xi_N)$ — дискретный случайный вектор с полиномиальным распределением

$$P(\xi_1 = m_1, \xi_2 = m_2, ..., \xi_N = m_N) = \frac{n!}{m_1! m_2! \cdots m_N!} p_1^{m_1} \cdots p_N^{m_N}.$$

Найти $M\xi_k$, $cov(\xi_k, \xi_l)$, $k, l = 1, 2, ..., N$.

Ответ: $M\xi_k = np_k$, $D\xi_k = np_k(1-p_k)$, $cov(\xi_k, \xi_l) = -np_k p_l (k \neq l)$.

Задача 6.25. Применим ли закон больших чисел к последовательности независимых случайных величин $\xi_1, \xi_2, ..., \xi_k, ...$, если они нормально распределены с $M\xi_k = 0$, $D\xi_k = ck^\alpha$; $c > 0$, $\alpha > 0$ — некоторые постоянные?

Ответ: При $\alpha < 1$ применим; при $\alpha \geq 1$ не применим.

Задача 6.31. Предполагается провести 10 измерений $x_1, x_2, ..., x_{10}$ неизвестной величины a. Считая $x_1, ..., x_{10}$ независимыми нормально распределенными случайными величинами с $Mx_k = a$, $Dx_k = 0.01$, подобрать Δ так, чтобы

$$P\left(\left|\frac{x_1 + x_2 + \cdots + x_{10}}{10} - a\right| < \Delta\right) = 0.99.$$

Ответ: 0.08145...

2.6 Закон больших чисел

Лемма. *Пусть случайная величина* ξ *неотрицательна, т.е.* $P(\xi \geq 0) = 1$, *тогда* $\forall \epsilon > 0$ $P(\xi \geq \epsilon) \leq \frac{M\xi}{\epsilon}$.

Доказательство. Пусть сначала ξ распределено непрерывно. Тогда

$$M\xi = \int_{-\infty}^{\infty} xp(x)dx = \int_{0}^{\infty} xp(x)dx \geq \int_{\epsilon}^{\infty} xp(x)dx \geq \epsilon \int_{\epsilon}^{\infty} p(x)dx = \epsilon P(\xi \geq \epsilon).$$

Если ξ распределено дискретно, тогда

$$M\xi = \sum_k x_k p_k = \sum_{x_k \geq 0} x_k p_k \geq \epsilon \sum_{x_k \geq \epsilon} p_k = \epsilon P(\xi \geq \epsilon).$$

Неравенство Чебышева. $P(|\xi - M\xi| \geq \epsilon) \leq \frac{D\xi}{\epsilon^2}$.

Доказательство.

$$P(|\xi - M\xi| \geq \epsilon) = P((\xi - M\xi)^2 \geq \epsilon^2) \leq \frac{M(\xi - M\xi)^2}{\epsilon^2} = \frac{D\xi}{\epsilon^2}.$$

Теорема. *Если случайные величины* $\xi_1, \xi_2, ..., \xi_n$ *независимы и*

$$\lim_{n\to\infty} \frac{\sum\limits_{k=1}^{n} D\xi_k}{n^2} = 0, \text{ *то* } \lim_{n\to\infty} P\left(\left|\frac{\sum\limits_{k=1}^{n} \xi_k}{n} - \frac{\sum\limits_{k=1}^{n} M\xi_k}{n}\right| < \epsilon\right) = 1.$$

Доказательство.

$$P\left(\left|\frac{\sum\limits_{k=1}^{n} \xi_k}{n} - \frac{\sum\limits_{k=1}^{n} M\xi_k}{n}\right| \geq \epsilon\right) = P\left(\left|\frac{\sum\limits_{k=1}^{n} \left(\xi_k - M\xi_k\right)}{n}\right| \geq \epsilon\right) \leq$$

$$\le \frac{D\left(\frac{\sum\limits_{k=1}^{n}\left(\xi_k - M\xi_k\right)}{n}\right)}{\epsilon^2} = \frac{D\sum\limits_{k=1}^{n}\xi_k}{n^2\epsilon^2} \to 0.$$

Отсюда следует утверждение теоремы.

Сходимость по вероятности. *Говорят, что последовательность* ξ_n *стремится по вероятности к* ξ (что обозначается так: $\xi_n \xrightarrow[n\to\infty]{P} \xi$), *если* $\forall\ \epsilon > 0$ $\lim\limits_{n\to\infty} P\left(\left|\xi_n - \xi\right| < \epsilon\right) = 1$.

Доказанная выше теорема формулируется так:

$$\frac{\sum\limits_{k=1}^{n}\xi_k}{n} - \frac{\sum\limits_{k=1}^{n}M\xi_k}{n} \xrightarrow[n\to\infty]{P} 0.$$

Теорема Чебышева: *Если случайные величины* $\xi_1, \xi_2, ..., \xi_n$ *попарно независимы и* $D\xi_i \le C$, *то для* $\forall\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\sum\limits_{k=1}^{n}\xi_k}{n} - \frac{\sum\limits_{k=1}^{n}M\xi_k}{n}\right| < \epsilon\right) = 1.$$

Достаточно доказать, что $\lim\limits_{n\to\infty}\frac{\sum\limits_{k=1}^{n}D\xi_k}{n^2} = 0$. Это очевидно, так как $\frac{\sum\limits_{k=1}^{n}D\xi_k}{n^2} \le \frac{\sum\limits_{k=1}^{n}C}{n^2} = \frac{C}{n} \to 0$ при $n \to \infty$.

Следствие. *Если случайные величины* $\xi_1, \xi_2, ..., \xi_n$ *попарно независимы и одинаково распределены, имеют конечную дисперсию, то для* $\forall\epsilon > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{\sum\limits_{k=1}^{n}\xi_k}{n} - a\right| < \epsilon\right) = 1,$$

где $a = M\xi_k$.

Теорема Бернулли. *Пусть* μ_n *— число успехов в испытаниях Бернулли с вероятностью* p *появления успеха в отдельном испытании. Тогда для* $\forall\epsilon > 0$ $\lim\limits_{n\to\infty} P\left(\left|\frac{\mu_n}{n} - p\right| < \epsilon\right) = 1$.

Доказательство. Пусть ξ_k описывает результат k-го испытания и имеет ряд распределения

$$\frac{x_k| \ 0 \ | \ 1}{p_k| \ q \ | \ p},$$

тогда $M\xi_k = 0 \cdot q + 1 \cdot p = p$, $M\xi^2 = 0^2 \cdot q + 1^2 \cdot p = p$, $D\xi = m\xi_i^2 - (M\xi_i)^2 = p - p^2 = pq$, $\xi_1 + \xi_2 + \cdots \xi_n = \mu_n$. После этого теорема Бернулли представляет частный случай следствия теоремы Чебышева, так как $\frac{\sum\limits_{k=1}^{n} \xi_k}{n} = \frac{\mu_n}{n}$.

2.7 Предельные теоремы

Характеристические функции. Пусть ξ – случайная величина. Характеристическая функция по определению есть $f(t) = f_\xi(t) = Me^{i\xi t} = M(\cos(\xi t) + i \sin(\xi t))$.

Примеры:

1) *Распределение Пуассона.*

$$f = \sum_k e^{ix_k t} p_k = \sum_{k=0}^{\infty} e^{ix_k t} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda + \lambda e^{it}}.$$

2) *Равномерное распределение на сегменте* $[a, b]$.

$$f(t) = M\exp(i\xi t) = \int\limits_{-\infty}^{\infty} e^{ixt} p(x) dx = \int\limits_a^b \frac{\exp(ixt)}{b-a} dx = \frac{e^{ibt} - e^{iat}}{it(b-a)}.$$

3) *Нормальное распределение.*

$$f(t) = \int\limits_{-\infty}^{\infty} e^{ixt} p(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} e^{ixt - \frac{(x-a^2)}{2\sigma^2}} dx.$$

Сделаем замену: $\frac{x-a}{\sigma} = u$. Тогда

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} e^{i(a+\sigma u)t - \frac{u^2}{2}} \sigma du =$$

$$= \frac{e^{iat}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\cos\sigma ut + i \sin\sigma ut) e^{-\frac{u^2}{2}} du = \frac{e^{iat}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(\sigma ut)\ e^{-\frac{u^2}{2}} du.$$

Пусть

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(\sigma ut)\ e^{-\frac{u^2}{2}} du.$$

Тогда

$$g'(t) = -\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u \sin(\sigma ut)\ e^{-\frac{u^2}{2}} du = -\sigma^2 t g(t).$$

Искомая функция $g(t)$ удовлетворяет дифференциальному уравнению $g'(t) = -\sigma^2 t g(t)$ с начальным условием $g(0) = 1$. Общее решение такого уравнения — $g(t) = Ce^{-\sigma^2 t^2 2}$. Из условия $g(0) = 1$ следует, что $C = 1$. Значит,

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{iat - \sigma^2 t^2/2}.$$

Свойства характеристической функции:

1) $f^{(k)}(t) = \frac{d^k}{dt^k} Me^{i\xi t} = M(i\xi)^k e^{i\xi t},\ f^{(k)}(0) = i^k M\xi^k$,

2) $f_{a\xi+b}(t) = Me^{i(a\xi+b)t} = e^{ibt} Me^{i\xi(at)} = e^{ibt} f_\xi(at)$,

3) пусть $\xi_1,\ \xi_2$ — независимые случайные величины, тогда

$$f_{\xi_1+\xi_2}(t) = Me^{i(\xi_1+\xi_2)t} = Me^{i\xi_1 t} e^{i\xi_2 t} = Me^{i\xi_1 t} Me^{i\xi_2 t} = f_{\xi_1}(t) f_{\xi_2}(t),$$

4) пусть ξ — непрерывная случайная величина, тогда $f_\xi(t) = \int\limits_{-\infty}^{\infty} e^{ixt} p(x) dx$ — это преобразование Фурье; если $p(x)$ – непрерывно дифференцируема, то $p(x) = \frac{1}{2\pi} \int\limits_{\infty}^{\infty} e^{-ixt} f(t) dt$, но $p(x)$ однозначно определяет функцию распределения $F(x) = \int\limits_{-\infty}^{x} p(u) du$. Таким образом, если ξ – непрерывная случайная величина, то между $f_\xi(t)$ и $F_\xi(x)$ существует непрерывное взаимно однозначное соответствие. Оказывается, это верно для любых случайных величин.

Центральная предельная теорема.

1)(Упрощенная формулировка.) *Пусть величины* $\xi_1, \xi_2, ..., \xi_n$ *независимы, одинаково распределены, имеют конечную дисперсию* σ^2, $M\xi_k = a$, $\eta_n = \frac{\xi_1+\xi_2+\dots+\xi_n-na}{\sigma\sqrt{n}}$. *Тогда*

$$F_{\eta_n}(x) = P\left(\frac{\xi_1+\xi_2+\dots+\xi_n-na}{\sigma\sqrt{n}} < x\right) \to \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du$$

при $n \to \infty$.

Рассмотрим

$$f_{\eta_n}(t) = f_{\sum\limits_{k=1}^{n}\frac{\xi_k-a}{\sigma\sqrt{n}}}(t) = \prod_{k=1}^{n} f_{\frac{\xi_k-a}{\sigma\sqrt{n}}}(t) = \left[f_{\frac{\xi_1-a}{\sigma\sqrt{n}}}(t)\right]^n =$$

$$= \left[f_{\xi_1-a}(\frac{t}{\sigma\sqrt{n}})\right]^n = \left[1 + it\frac{M(\xi_1-a)}{\sigma\sqrt{n}} + i^2t^2\frac{M(\xi_1-a)^2}{2\sigma^2 n} + \cdots\right]^n =$$

$$= \left(1 - \frac{t^2\sigma^2}{2\sigma^2 n} + \cdots\right)^n \to e^{-\frac{t^2}{2}}$$

при $n \to \infty$. Таким образом, предел есть характеристическая функция нормально-стандартного распределения. Значит

$$F_{\frac{\xi_1+\xi_2+\dots+\xi_n-na}{\sigma\sqrt{n}}}(x) \to \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du.$$

2)(Формулировка Ляпунова.) *Пусть* $\xi_1, \xi_2, ..., \xi_n$ — *независимые случайные величины, обладающие конечным третьим центральным моментом* $c_3^3 = M|\xi - a_k|^3$, $a_k = M\xi_k$. *Положим* $A_n = \sum\limits_{k=1}^{n} a_k$, $B_n^2 = \sum\limits_{k=1}^{n} D\xi_k$, $C_n^3 = \sum\limits_{k=1}^{n} c_k^3$. *Если* $\frac{C_n}{B_n} \to 0$ *при* $n \to \infty$, *то*

$$P\left(\frac{\xi_1+\xi_2+\dots+\xi_n-A_n}{B_n} < x\right) \to \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du$$

при $n \to \infty$.

Теорему примем без доказательства.

3) **Интегральная предельная теорема Муавра – Лапласа.** *Пусть p — вероятность успеха в испытаниях Бернулли, n — число испытаний, m — число успехов в этих испытаниях, $x_m = \frac{m-np}{\sqrt{npq}}$, тогда для $\forall a, b$*

$$\lim_{n\to\infty} P(a < x_m < b) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{u^2}{2}} du.$$

Доказательство. k-му испытанию поставим в соответствие случайную величину ξ_k с рядом распределения

$$\frac{x_k | \ 0 \ | \ 1}{p_k | \ q \ | \ p},$$

$q = 1 - p$. Ввиду независимости испытаний члены последовательности $\xi_1, \xi_2, ..., \xi_n$ независимы и одинаково распределены. $M\xi_k = p$, $D\xi_k = pq$, $\xi_1+\xi_2+\cdots+\xi_n = m$, $A_n = np$, $B_n^2 = npq$, значит, $x_m = \frac{m-np}{\sqrt{npq}}$ представляет собой левую часть неравенства под знаком вероятности в центральной предельной теореме, поэтому $\lim\limits_{n\to\infty} P(x_m < x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$, откуда $\lim\limits_{n\to\infty} P(a < x_m < b) = \frac{1}{\sqrt{2\pi}} \int\limits_a^b e^{-\frac{u^2}{2}} du$.

2.7.1 7-ое практическое занятие. Предельные теоремы

Задача 14.556. Проводятся последовательные испытания по схеме Бернулли. Вероятность осуществления события A в одном испытании $p = 0.6$. Считая применимыми предельные теоремы Муавра – Лапласа, вычислить вероятности следующих событий: $B=$ {событие A произойдет в большинстве из 60 испытаний}, $C=${число успешных осуществлений события A в 60 испытаниях будет заключено между 30 и 42}.

Решение. $P(B) = P(m \geq 31) = P(\frac{m-np}{\sqrt{npq}} \geq \frac{31-np}{\sqrt{npq}}) =$
$= P(\frac{m-36}{\sqrt{60\cdot 0.6\cdot 0.4}} \geq \frac{31-36}{12}\sqrt{10}) = 0.5 + \Phi(\frac{5\sqrt{10}}{12}) = 0.5 + \Phi(1.2) = 0.9032$,

$P(C) = P(30 \leq m \leq 42) = P(-\frac{6\sqrt{10}}{12} \leq x_m \leq \frac{6\sqrt{10}}{12}) = 2\Phi(1.2) = 0.8046$.

Задача 14.568. В опыте Бюффона монета была подброшена 4040 раз, причем герб выпал 2048 раз. С какой вероятностью можно при повторении опыта получить такое же или еще большее отклонение относительной частоты успехов от вероятности успеха в одном опыте?

Решение. n=4040, m=2048, $P(m \geq 2048$ или $m \leq 1992) =$
$= 2P(\frac{m-2020}{\sqrt{4040 \cdot 0.5 \cdot 0.5}} \geq \frac{28}{\sqrt{1010}}) = 2(x_m \geq 0.9) = 2(0.5 - 0.32) = 1 - 0.64 =$ 0.36.

Задача 14.574. Среднее число вызовов на АТС за 1 минуту равно $\lambda = 20 = m_X$. Найти вероятности следующих событий: $A = \{X \geq 20\}$; $B = \{10 < X < 30\}$.

Решение. Пусть ξ распределена по Пуассону с параметром λ. Если λ велика, то $\frac{\xi-\lambda}{\sqrt{\lambda}}$ подчиняется нормально-стандартному распределению. $P(A) = P(X \geq 20) = P(\frac{X-20}{\sqrt{20}} \geq \frac{20-20}{\sqrt{20}} \approx \frac{1}{\sqrt{2\pi}} \int\limits_0^\infty e^{-\frac{t^2}{2}} dt = 0.5$, $P(B) = P(10 \leq X \leq 30) = P(\left|\frac{X-20}{\sqrt{20}}\right| < \frac{10}{\sqrt{20}}) \approx 2 \cdot 0.49 = 0.98$.

Задача 7.6(в). Вычислить характеристические функции следующих законов распределения: в) показательного.

Решение. в) $p(x) = \alpha e^{-\alpha x}$, $x > 0$. $f(t) = Me^{i\xi t} = \int\limits_0^\infty \alpha e^{-\alpha x} e^{ixt} dx =$ $\frac{\alpha}{\alpha - it}$.

Задача 7.7(а). Найти закон распределения, соответствующий характеристической функции $\cos t$.

Решение. $f(t) = \cos t = \frac{e^{it}+e^{-it}}{2} = \frac{1}{2} \cdot e^{i(-1)t} + \frac{1}{2} \cdot e^{i \cdot 1 \cdot t}$, поэтому ряд рапределения имеет вид:

x_k	-1	1
p_k	0.5	0.5

Задача 7.11. Складывается 10^4 чисел, каждое из которых округлено с точностью до 10^{-m}. Предполагая, что ошибки от округления независимы и в интервале $(-0.5 \cdot 10^{-m}; 0.5 \cdot 10^{-m})$ распределены равномерно, найти пределы, в которых с вероятностью, не меньшей 0.99, будет лежать суммарная ошибка.

Решение. $n = 10^4$. ξ_k распределена равномерно на $[-\alpha, \alpha]$, $\alpha = 0.5 \cdot 10^{-m}$, $k = 1, 2, ..., n$. $\eta = \sum\limits_{k=1}^{n} \xi_k$ — суммарная ошибка. $M\xi_k = 0$.

$M\xi_k^2 = D\xi_k = \int\limits_{-\alpha}^{\alpha} x^2 \frac{1}{2\alpha} dx = \frac{\alpha^2}{3}$. $M\eta = 0$, $D\eta = \frac{n\alpha^2}{3}$, $\zeta = \frac{\eta}{\sqrt{D\eta}} = \frac{\eta\sqrt{3}}{100\alpha}$ распределена нормально-стандартно. $(|\zeta| < A) = 2\Phi(A) = 0.99$, $A = \Phi^{-1}(0.495) = 2.58$. $|\eta| < \frac{100\alpha \cdot 2.58}{\sqrt{3}}$.

Задача 7.14. Пусть случайные величины $\xi_1, \xi_2, ..., \xi_n$ независимы; $P(\xi_k = 1) = 1 - P(\xi_k = 0) = p_k(n)$; $k = 1, 2, ...$ Найти $\lim\limits_{n\to\infty} P(\xi_1 + \xi_2 + \cdots + \xi_n = m)$, если $p_1(n) + \cdots + p_n(n) \to \lambda$ и $\max\limits_{1\le k\le n} p_k(n) \to 0$ при $n \to \infty$.

Решение. $f_{\xi_k}(t) = e^{it\xi_k} = e^{it} p_k(n) + 1 - p_k(n) = 1 + (e^{it} - 1)p_k(n)$, $f_{\sum \xi_k}(t) = \prod\limits_{k=1}^{n} \Bigg(1 + (e^{it} - 1)p_k(n)\Bigg)$, $\ln f_{\sum \xi_k}(t) = \sum\limits_{k=1}^{n} \ln\Bigg(1 + (e^{it} - 1)p_k(n)\Bigg) \approx \sum\limits_{k=1}^{n} \Bigg((e^{it} - 1)p_k(n) + \cdots\Bigg) \to (e^{it} - 1)\lambda$, $f_{\sum \xi_k}(t) \to e^{\lambda(e^{it} - 1)}$, т.е. распределение суммы $\xi_1 + \xi_2 + \cdots + \xi_n$ стремится к распределению Пуассона.

Задачи домашнего задания.

Задача 14.557. При условиях задачи 14.556 найти вероятность события D={событие A осуществляется 36 раз в 60 испытаниях}.

Ответ: $P(D) \approx 0.1051$.

Задачи 14.569, 14.570. Сколько раз нужно подбросить монету, чтобы с вероятностью, не меньшей 0.975, утверждать, что частота выпадения герба попадет в интервал (0.4, 0.6)? Получить оценку указанного числа: а) используя второе неравенство Чебышева; б) применяя интегральную теорему Муавра – Лапласа.

Ответ: а) Не менее 1000 раз, б) не менее 127 раз.

Задача 7.6(а, г). Вычислить характеристические функции распределений: а) биномиального; г) равномерного на отрезке $[-1, 1]$.

Ответ: а) $(pe^{it} + q)^n$; г) $\sin t/t$.

Задача 7.7(б). Найти законы распределения, соответствующие характеристическим функциям: б) $\cos^2 t$.

Ответ: $P(\xi = 2) = P(\xi = -2) = 1/4$, $P(\xi = 0) = 1/2$.

Задача 7.9. Величины ξ_1, ξ_2, ξ_3 независимы и нормально распределены с параметрами $(1,1)$, $(0,4)$, $(-1,1)$. Найти: а) $P(\xi_1+\xi_2+\xi_3 < 0)$; б) $P(|2\xi_1 - \xi_2 + \xi_3| < 3)$.

Ответ: а) 0.5, б) 0.6826.

Задача 7.15. Плотность распределения $p_\xi(x)$ случайной величины ξ непрерывна и ограничена на отрезке $[a,b]$ и равна 0 вне $[a,b]$. Положим $\eta_n = \{n\xi\}$, где $\{x\}$ — дробная часть числа x. Найти $\lim\limits_{n\to\infty} P(\eta_n < x)$, $0 < x < 1$.

Ответ: $x(0 < x < 1)$.

Задача 7.19. Пусть τ_2 — число испытаний в схеме Бернулли, прошедшее от начала испытаний до появления второго успеха. Получить 10 независимых реализаций величины τ_2, если вероятность успеха в отдельном испытании: а) $p = 0.5$; б) $p = 0.75$. Найти средние арифметические и сравнить с $M\tau_2$.

3 ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

3.1 Понятие выборки. Выборочное распределение

Пусть ξ — некоторая случайная величина, и проводится эксперимент, в котором осуществляются повторные независимые наблюдения над этой случайной величиной. Пусть этих наблюдений (испытаний) n. Появляющиеся при испытаниях числа обозначим $x_1, x_2, ..., x_n$. Удобно считать, что k-ое испытание описывается случайной величиной X_k, подчиненной тому же закону распределения, что и ξ. Вектор $X = (X_1, X_2, \ldots, X_n)$ из независимых, одинаково распределенных по тому же закону, что и наблюдаемая случайная величина ξ, случайных величин $X_1, X_2, ..., X_n$ называется случайной выборкой, X_k – элементом выборки, n — объемом выборки, $x = (x_1, x_2, ..., x_n)$ — реализацией выборки.

Рассмотрим событие $(X_k < x)$, заключающееся в том, что во время k-го испытания случайная величина ξ примет значение, меньшее x. Число таких случаев обозначим $\mu_n(x)$. Значит, $\mu_n = |k : X_k < x|$ — число элементов выборки, меньших x. Обозначим $F_n^*(x) = \frac{\mu_n(x)}{n}$. Это случайная величина, представляющая частоту события $(\xi < x)$, называется выборочным распределением или эмпирической функцией распределения.

Теорема:

$$F_n^*(x) \xrightarrow{P} F(x), \text{ если } n \to \infty.$$

Доказательство. Пусть для $\forall x \in (-\infty, \infty)$ $p = P(\xi < x) = P(X_k <$

x), т.е. $(X_k < x)$ — "успех", а p — вероятность "успеха", $\mu_n(x)$ — число "успехов", тогда $F_n^*(x)$ – частота "успеха", а по теореме Бернулли,

$$\frac{\mu_n(x)}{n} \xrightarrow{P} p, \text{ если } n \to \infty,$$

поэтому наша теорема доказана.

При задании реализации $x = (x_1, x_2, ..., x_n)$ выборки $X = (X_1, X_2, ..., X_n)$ эмпирическая функция распределения $F_n^*(x)$ становится обычной функцией. Для построения этой функции элементы реализации $x = (x_1, x_2, ..., x_n)$ расположим в порядке возрастания: $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$. Эта последовательность называется вариационным рядом. Если члены вариационного ряда различны, то реализация $F_n^*(x)$ имеет вид:

$$F_n^*(x) = \begin{cases} 0, \ x \le x_{(1)} \\ \frac{1}{n}, \ x_{(1)} < x \le x_{(2)} \\ \frac{2}{n}, \ x_{(2)} < x \le x_{(3)} \\ \cdots \\ \frac{n-1}{n}, \ x_{(n-1)} < x \le x_{(n)} \\ 1, \ x_{(n)} < x. \end{cases}$$

Если некоторые члены вариационного ряда совпадают, то различные значения этих членов обозначим $\tilde{x}_1 = x_{(1)}$, $\tilde{x}_2$,...,$\tilde{x}_m = x_{(n)}$, и пусть эти значения в вариационном ряду повторяются k_1, k_2,...,k_m раз соответственно $(k_1 + k_2 + \cdots + k_m = n)$, тогда реализация $F_n^*(x)$ имеет вид:

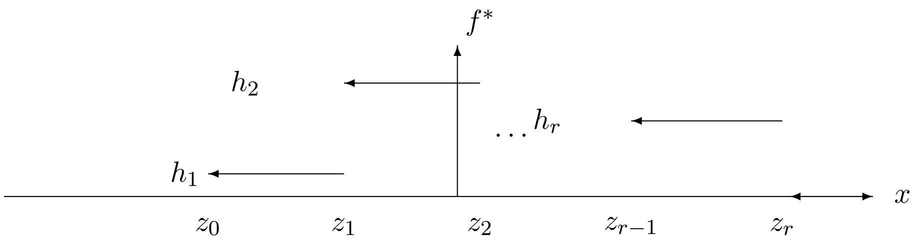
$$F_n^*(x) = \begin{cases} 0, \ x \le \tilde{x}_1 \\ \frac{k_1}{n}, \ \tilde{x}_1 < x \le \tilde{x}_2 \\ \frac{k_1+k_2}{n}, \ \tilde{x}_2 \le x < \tilde{x}_3 \\ \cdots \\ \frac{n-k_m}{n}, \ \tilde{x}_{(m-1)} \le x < \tilde{x}_{(m)} \\ 1, \ \tilde{x}_{(m)} \le x. \end{cases}$$

Пример. Опыт Бюффона. Бюффон $n = 4040$ раз бросал монету, при этом "герб" выпал $m = 2048$ раз. В реализации

$x = (x_1, x_2, ..., x_{4040}) = (1, ..., 0, ..., 1, ..., 0, ...1)$ число единиц будет равно $m = 2048$, поэтому

$$F^*_{4040}(x) = \begin{cases} 0, \ x < 0 \\ \frac{1992}{4040}, \ 0 \le x < 1 \\ 1, \ 1 \le x. \end{cases}$$

Гистограмма. Сегмент $[x_{(1)}, x_{(n)}]$ делят на некоторое число равных частей точками $z_0 = x_{(1)} < z_1 < \cdots < z_r = x_{(n)}$. Вычисляют $p^*_k = F^*_n(z_{k+1}) - F^*_n(z_k) = \frac{m_k}{n}$, $m_k -$ число значений случайной величины ξ, попавших в $k -$ ый интервал. Полагаем $h_k = \frac{p^*_k}{\frac{x_{(n)} - x(1)}{r}}$. Получаем ступенчатую функцию, имеющую график:



Она называется гистограммой реализации.

3.2 Точечные оценки параметров

В теории вероятностей наиболее часто встречающиеся законы распределения содержат параметры. Например, распределение Пуассона $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, ...$, $\lambda > 0$ содержит параметр λ, нормальное распределение $N(a, \sigma^2)$ с плотностью вероятности $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-a)^2}{2\sigma^2})$ — параметры a, σ^2. Если в некоторой статистической задаче вероятностная модель описывается некоторым распределением с неизвестным параметром, то возникает проблема определения этого параметра.

Эмпирическая формула $\theta^*_n = \theta^*_n(X_1, ..., X_n)$ для вычисления какого - либо параметра θ, возникающего при описании случайной величины ξ, называется оценкой параметра θ.

Требования на оценку:

1) Несмещенность. Если $M\theta_n^*(X_1, ..., X_n) = \theta$, то оценка называется несмещенной. Требование несмещенности оценки означает, что по крайней мере "в среднем" используемая оценка приводит к желаемому результату.

2) Если $\theta_n^* \xrightarrow[n\to\infty]{P} \theta$, то оценка называется состоятельной. Если оценка состоятельна, то при больших объемах n выборки распределение значений параметра должно концентрироваться около истинного значения параметра.

3) Среди всех оценок нужно выбирать ту, для которой дисперсия наименьшая.

Иногда построенные оценки обладают свойством асимптотической нормальности. Если функции распределения последовательности случайных величин $(\eta_n - A_n)/B_n$ сходятся при $n \to \infty$ к функции $\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp(-\frac{u^2}{2})du$, то говорят, что случайная величина η_n при $n \to \infty$ асимптотически нормальна с параметрами (A_n, B_n^2).

Метод наибольшего правдоподобия для нахождения оценок параметров. Если плотность вероятности непрерывной случайной величины ξ содержит параметр θ, т.е. она имеет вид $p(x, \theta)$, то для выборки $X = (X_1, X_2, ..., X_n)$ вводят так называемую функцию правдоподобия L :

$$L = L(X_1, X_2, ..., X_n, \theta) = p(X_1, \theta)p(X_2, \theta) \cdots p(X_n, \theta).$$

Если ξ — дискретная случайная величина и $p_x(\theta) = P(\xi = x)$, то L определяется как

$$L = p_{X_1}(\theta)p_{X_2}(\theta) \cdots p_{X_n}(\theta).$$

В качестве оценки θ^* параметра θ берется то его значение, на котором достигается максимальное значение функции правдоподобия $L(X_1, X_2, ..., X_n, \theta)$. Таким образом, оценка θ^* должна удовлетворять уравнению

$$\frac{\partial L}{\partial \theta} = 0 \text{ или } \frac{\partial \ln L}{\partial \theta} = 0.$$

Оказывается, при достаточно общих условиях на случайную величину ξ уравнение $\frac{\partial \ln L}{\partial \theta} = 0$ имеет решение θ^*, которое представляет состоятельную, асимптотически нормальную оценку параметра θ.

Пример. Используя метод наибольшего правдоподобия, найти по выборке $X_1, ..., X_n$, где $(X_k = m) = \frac{\lambda^m}{m!} e^{-\lambda}$, $m = 0, 1, \ldots$, оценку λ^* параметра λ. Будет ли эта оценка несмещенной и состоятельной? Найти $M\lambda^*$, $D\lambda^*$.

$$L = \frac{\lambda^{X_1}}{X_1!} e^{-\lambda} \frac{\lambda^{X_2}}{X_2!} e^{-\lambda} \cdots \frac{\lambda^{X_n}}{X_n!} e^{-\lambda} = \frac{\lambda^{\sum\limits_{i=1}^{n} X_i}}{\prod\limits_{k=1}^{n} X_k!} e^{-n\lambda},$$

$$\ln L = -n\lambda + \sum_{i=1}^{n} X_i \ln \lambda - \sum_{i=1}^{n} \ln(X_i!), \;\; \frac{\partial \ln L}{\partial \lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^{n} X_i = 0,$$

$$\lambda^* = \frac{\sum\limits_{i=1}^{n} X_i}{n}.$$

$M\lambda^* = \frac{\sum\limits_{i=1}^{n} MX_i}{n} = \lambda$, λ^* — несмещенная оценка. $D\lambda^* = \frac{1}{n^2} nD\xi = \frac{\lambda}{n}$. $P(|\lambda^* - \lambda| \geq \epsilon) \leq \frac{D\lambda^*}{\epsilon^2} \to 0$ при $n \to \infty$, поэтому λ^* — состоятельная оценка.

Пример.Опыт Бюффона.$L = p^{x_1} q^{1-x_1} \cdots p^{x_n} q^{1-x_n} = \sum\limits_{k=1}^{n} p^{x_k} q^{1-x_k}$. $\ln L = \sum x_k \ln p + (n - \sum x_k) \ln(1-p)$. $\frac{\sum x_k}{p} - \frac{n - \sum x_k}{1-p} = 0$, $\sum x_k - np = 0$. $p^* = \frac{\sum x_k}{n} \to \frac{2048}{4040}$.

3.3 Выборочные моменты

У случайной величины ξ можно определить так называемые теоретические моменты: $\alpha_\nu = M\xi^\nu$ – моменты порядка ν, $\mu_\nu = M(\xi - M\xi)^\nu$ – центральные моменты порядка ν. В частности $\alpha_1 = M\xi$ – математическое ожидание, $\mu_1 = 0$, $\mu_2 = D\xi$. Теоретическим моментам сопоставляются выборочные моменты $a_\nu,\ m_\nu$. $a_\nu = \frac{\sum\limits_{k=1}^{n} X_k^\nu}{n}$ называется выборочным моментом порядка ν, в частности $a_1 = \frac{\sum\limits_{k=1}^{n} X_k^1}{n} = \bar{X}$ называется выборочным средним, $m_\nu = \frac{\sum\limits_{k=1}^{n} (X_k - \bar{X})^\nu}{n}$ — центральный выборочный момент порядка ν. Выборочные моменты являются точечными оценками соответствующих теоретических моментов.

Несмещенность оценок теоретических моментов определяется с помощью выборочных моментов:

$$Ma_\nu = M\frac{\sum\limits_{k=1}^{n} X_k^\nu}{n} = \frac{1}{n}\sum_{k=1}^{n} MX_k^\nu = \frac{1}{n}\sum_{k=1}^{n} \alpha_k = \alpha_\nu.$$

Состоятельность оценок теоретических моментов в виде выборочных моментов:

$$Da_\nu = D\frac{\sum\limits_{k=1}^{n} X_k^\nu}{n} = \frac{1}{n^2}\sum_{k=1}^{n} DX_k^\nu = \frac{1}{n^2}\sum_{k=1}^{n} D\xi^\nu = \frac{1}{n}(M\xi^{2\nu} - (M\xi)^2) =$$

$$= \frac{\alpha_{2\nu} - \alpha_\nu^2}{n}.$$

Откуда

$$P(|a_\nu - \alpha_\nu| \geq \epsilon) \leq \frac{M(a_\nu - \alpha_\nu)^2}{\epsilon^2} = \frac{Da_\nu}{\epsilon^2} = \frac{\alpha_{2\nu} - \alpha_\nu^2}{n\epsilon^2} \to 0$$

при $n \to \infty$. Поэтому $a_\nu \xrightarrow[n\to\infty]{P} \alpha_\nu$.

Выборочные моменты обладают ещё одним свойством: они *асимптотически нормальны с параметрами* $(\alpha_\nu, \frac{\alpha_{2\nu} - \alpha_\nu^2}{n})$. Для доказательства этого свойства рассмотрим

$$\eta_\nu = \frac{a_\nu - \alpha_\nu}{\sqrt{Da_\nu}} = \frac{\sum\limits_{k=1}^{n} X_k^\nu - n\alpha_\nu}{\mu_\nu\sqrt{n}},$$

откуда видно, что η_ν обладает структурой случайной величины в аргументе центральной предельной теоремы, поэтому она в пределе $n \to \infty$ асимптотически стандартно нормальна, значит a_ν асимптотически нормален.

Пример. Опыт Бюффона. Для случайной величины ξ, представляющей случайное число успехов при однократном производстве опыта, имеем $M\xi = 0 \cdot q + 1 \cdot p = p = \alpha_1$, $a_1 = \bar{X} = \frac{\sum X_k}{n} \to \frac{2048}{4040}$.

Математическое ожидание и дисперсия центрального момента 2-го порядка По определению $m_2 = \frac{\sum\limits_{k=1}^{n}(X_k - \bar{X})^2}{n}$. Вычислим Mm_2. Положим $Y_k = X_k - MX_k = X_k - M\xi$, очевидно $MY_k = MX_k - MMX_k = MX_k - MX_k = 0$, $DY_k = DX_k = D\xi = \mu_2$. Определим $\bar{Y} = \frac{\sum\limits_{k=1}^{n} Y_k}{n}$. Так как $X_k - \bar{X} = Y_k + M\xi - \frac{\sum\limits_{k=1}^{n}(Y_k + M\xi)}{n} = Y_k - \bar{Y}$, то

$$m_2 = \frac{\sum\limits_{k=1}^{n}(Y_k - \bar{Y})^2}{n} = \frac{1}{n}\sum_{k=1}^{n}(Y_k^2 - 2\bar{Y}Y_k + \bar{Y}^2) = \frac{1}{n}\sum_{k=1}^{n}Y_k^2 - 2\bar{Y}^2 + \bar{Y}^2 =$$

$$= \frac{1}{n}\sum_{k=1}^{n}Y_k^2 - \bar{Y}^2.$$

Вычислим $M\bar{Y}^2$:

$$M\bar{Y}^2 = M\left(\frac{\sum\limits_{k=1}^{n} Y_k}{n}\right)^2 = \frac{1}{n^2}\sum_{k,l=1}^{n} M(Y_kY_l) = \frac{1}{n^2}\sum_{k=1}^{n} M(Y_kY_k) = \frac{\mu_2}{n},$$

так как Y_k и Y_l независимы, поэтому $M(Y_kY_l) = MY_kMY_l = 0$, если $k \neq l$. Теперь можно перейти к вычислению Mm_2.

$$Mm_2 = \frac{1}{n}\sum_{k=1}^{n} MY_k^2 - \frac{\mu_2}{n} = \mu_2(1 - \frac{1}{n}).$$

Несколько более громоздкие вычисления приводят к следующему значению Dm_2:

$$Dm_2 = \frac{\mu_4 - \mu_2^2}{n} - 2\frac{\mu_4 - 2\mu_2^2}{n^2} + \frac{\mu_4 - 3\mu_2^2}{n^3}.$$

Замечаем, что $Mm_2 \neq \mu_2$, поэтому m_2 — смещенная оценка для μ_2. Но если определить $s^2 = \frac{nm_2}{n-1}$, то $s^2 = \mu_2$. s^2 называется приведенной дисперсией. $Mm_2 = \mu_2(1-\frac{1}{n})$ можно записать как $(m_2-\mu_2) = (1/n)$, аналогичным образом можно доказать, что $(m_2 - \mu_2)^2 = (1/n)$, поэтому $(|m_2 - \mu_2| > \epsilon) \leq M(m_2 - \mu_2)^2/\epsilon^2 \to 0$ при $n \to \infty$, это значит $m_2 \xrightarrow[n\to\infty]{P} \mu_2$, то есть m_2 является состоятельной оценкой для μ_2. Аналогично показывается, что и $s^2 \xrightarrow[n\to\infty]{P} \mu_2$.

Пример. Опыт Бюффона.

$$m_2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}^2 = \frac{2048}{4040} - \frac{2048^2}{4040^2} = \frac{2048\cdot 1992}{4040^2},\ s^2 = \frac{2028\cdot 1992}{4040\cdot 4039}.$$

3.3.1 8-ое практическое занятие. Выборочные моменты

Задача 9.2. Используя таблицу нормальных случайных чисел, получить реализацию выборки $X_1, X_2, ..., X_n$, где X_k нормально распределены с $MX_k = 0.5$; $DX_k = 1$; $n = 22$. Найти вариационный ряд $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$; эмпирическую функцию распределения; $\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$ (сравнить с MX_k); $s^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(x_k - \bar{x})^2$ (сравнить с DX_k).

Решение. $n = 22$. Числа из таблицы нормальных чисел (числа в скобках означают порядковый номер чисел, если их записать в порядке возрастания):

0.464(15)	0.137(12)	2.455(22)	-0.323(7)	-0.068(10)
0.296(14)	-0.288(8)	1.298(19)	0.060(11)	-2.256(1)
-0.531 (5)	-0.194(9)	0.543(16)	-1.558(2)	0.187(13)
-1.190(3)	1.486(21)	-0.354(6)	-0.634(4)	0.697(17)
0.926(18)	1.375(20)			

Для получения выборки, требуемой условием задачи, к написанным числам нужно прибавить 0.5. Эмпирическая функция распределения — ступенчатая функция. В точках, описываемых членами вариационного ряда, она терпит скачок, равный $1/22$.

$$\bar{x} = \frac{\sum x_i}{n} = 0.6149, \;\; s^2 = 1.1186.$$

Задача 9.3. Пусть (x_1, y_1),...,(x_n, y_n) — независимые одинаково распределенные двумерные величины. Положим $m_{11} = \frac{1}{n}\sum\limits_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y})$, где $\bar{x} = (x_1 + ... + x_n)/n, \bar{y} = (y_1 + \cdots + y_n)/n$. Найти Mm_{11} и показать, что $Dm_{11} = O(\frac{1}{n})$ при $n \to \infty$.

Решение. Положим $Mx_k = a$, $My_k = b$, $X_k = x_k - a$, $Y_k = y_k - b$. Тогда $x_k = X_k + a$, $y_k = Y_k + b$, $\bar{x} = \bar{X} + a$, $\bar{y} = \bar{Y} + b$. Находим

$$m_{11} = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})(Y_k - \bar{Y}) = \frac{1}{n}\sum_{k=1}^{n} X_k Y_k - \bar{X}\bar{Y}.$$

Так как $M(X_k Y_k) = cov(x_1, y_1)$, $MX_k = MY_k = 0$, то

$$M(\bar{X}\bar{Y}) = \frac{1}{n^2}\sum_{k,l} M(X_k Y_l) = \frac{1}{n^2}\sum_{k \neq l} MX_k MY_k + \frac{1}{n^2}\sum_{k=l} M(X_k Y_l) =$$

$$= \frac{1}{n^2}\sum_{k=1}^{n} M(X_k Y_k) = \frac{1}{n} cov(x_1, y_1).$$

$$Mm_{11} = \frac{1}{n}\sum cov(x_1, y_1) - \frac{1}{n} cov(x_1, y_1) = (1 - \frac{1}{n}) cov(x_1, y_1).$$

$$m_{11} = (\frac{1}{n} - \frac{1}{n^2})\sum_{k=l} X_k Y_k - \frac{1}{n^2}\sum_{\substack{k,l \\ k \neq l}} X_k Y_l.$$

$$Dm_{11} = cov(m_{11}, m_{11}) =$$

$$= cov\left(\frac{n-1}{n^2}\sum_{k=l} X_k Y_k - \frac{1}{n^2}\sum_{\substack{k,l \\ k \neq l}} X_k Y_l \;,\; \frac{n-1}{n^2}\sum_{s=l} X_s Y_s - \frac{1}{n^2}\sum_{\substack{s,t \\ s \neq t}} X_s Y_t\right)$$

$$= \frac{(n-1)^2}{n^4} \sum_{k,s} cov(X_kY_k, X_sY_s) - 2\frac{n-1}{n^4} \sum_{\substack{k,s,t \\ s\neq t}} cov(X_kY_k, X_sY_t) +$$

$$+ \frac{1}{n^4} \sum_{\substack{k,l,k\neq l \\ s,t,s\neq t}} cov(X_kY_l, X_sY_t) = \frac{(n-1)^2}{n^4} n\ cov(X_1Y_1, X_1Y_1) -$$

$$-2\frac{n-1}{n^4} \sum_{\substack{s,t \\ s\neq t}} (cov(X_sY_s, X_sY_t) + cov(X_tY_t, X_sY_t)) +$$

$$+ \frac{1}{n^4} \sum_{\substack{k,l,k\neq l \\ s,t,s\neq t}} M(X_kY_lX_sY_t) = \frac{(n-1)^2}{n^3}(\mu_{2,2} - \mu_{1,1}^2) + \frac{1}{n^4} \sum_{\substack{k,l \\ k\neq l}} M(X_k^2Y_l^2) =$$

$$= \frac{(n-1)^2}{n^3}(\mu_{2,2} - \mu_{1,1}^2) + \frac{n-11}{n^3}\mu_{0,2}\mu_{2,0},$$

где $\mu_{k,l} = M(X_1^kY_1^l)$.

Задача 15.7. Построить график эмпирической функции распределения, гистограмму и полигон частот для выборки, представленной статистическим рядом:

z_i	15	16	17	18	19
n_i	1	4	5	4	2

Решение. Эмпирическая функция распределения:

$$F_n^*(x) = \frac{1}{n} \sum_{z_i<x} n_i = \frac{1}{16} \sum_{z_i<x} n_i = \begin{cases} 0, x \leq 15 \\ \frac{1}{16}, 15 < x \leq 16 \\ \frac{5}{16}, 16 < x \leq 17 \\ \frac{10}{16}, 17 < x \leq 18 \\ \frac{14}{16}, 18 < x \leq 19 \\ 1, 19 < x \end{cases}.$$

Гистограмма:

$$p_n^*(x) = \begin{cases} \frac{1}{16}, 14.5 < x \leq 15.5 \\ \frac{4}{16}, 15.5 < x \leq 16.5 \\ \frac{5}{16}, 16.5 < x \leq 17.5 \\ \frac{4}{16}, 17.5 < x \leq 18.5 \\ \frac{2}{16}, 18.5 < x \leq 19.5 \end{cases}.$$

Задача 15.13. Методом моделирования получить реализации выборок объема $n = 10$ для случайной величины с показательным законом распределения $Ex(\lambda)$ с $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$.

Решение. Функция распределения для показательного распределения с параметром λ имеет вид: $F(x) = 1 - e^{-\lambda x}$. Значения y случайной величины Y, подчиненной показательному распределению с параметром λ, из значений x случайной величины X, подчиненной равномерному распределению на [0,1], получаются по формуле $y = -\ln(1-x)/\lambda$. В нижеприведенных таблицах в первой строке указаны значения x, во второй — y.

	0.100	0.376	0.863	0.876	0.480	0.742	0.524	0.226	0.645	0.320
$\lambda = 1$	0.105	0.472	1.990	2.090	0.654	1.350	0.742	0.256	1.040	0.386
	0.973	0.520	0.467	0.375	0.564	0.962	0.037	0.895	0.093	0.902
$\lambda = 2$	1.810	0.367	0.315	0.235	0.415	1.640	0.019	1.130	0.048	1.160
	0.253	0.135	0.354	0.420	0.894	0.480	0.084	0.319	0.320	0.560
$\lambda = 3$	0.097	0.048	0.146	0.182	0.748	0.218	0.029	0.128	0.129	0.274

Задача 15.14. Автомобили подъезжают к автозаправочной станции последовательно, причем время между прибытием двух автомобилей имеет показательное распределение с параметром $\lambda_1 = 1$. Если очереди нет, автомобиль заправляется сразу, в противном случае он становится в очередь. Время заправки автомобиля имеет показательное распределение с параметром $\lambda_2 = 2$. Используя выборки, полученные в задаче 15.13, составить таблицу, содержащую время подъезда для каждого из пяти последовательно прибывающих автомобилей, время начала и конца заправки, продолжительность ожидания в очереди, общее время на ожидание и обслуживание.

Решение.

Время подъезда	начало заправки	конец заправки	продолжит. ожид. в очер.	общее время
0.0	0.0	1.810	0.0	1.81
.105	1.810	2.177	1.705	2.072
.577	2.177	2.492	1.600	1.915
2.567	2.567	2.802	0.0	0.235
4.657	4.657	5.072	0.0	0.415

Задачи домашнего задания.

Задача 9.1. По выборке $x_1, ..., x_n$, полученной в задаче 5.24, найти: а) вариационный ряд $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$;

б) эмпирическую функцию распределения (построить ее график и график теоретической функции распределения); в) $\bar{x} = (x_1 + ... + x_n)/n$ (сравнить с Mx_k); г) $s^2 = \frac{1}{n-1}\sum\limits_{k=1}^{n}(x_k - \bar{x})^2$ (сравнить с Dx_k).

Задача 9.4. Показать, что величина m_{11}, определенная в задаче 9.3, при $n \to \infty$ асимптотически нормальна.

Задача 15.8. Выполнить программу задачи 15.7 по отношению к выборке

z_i	2	3	4	5	6	7	8
n_i	1	3	4	6	5	2	1

Задача 15.15. Пусть t_i — время наработки на отказ i-го элемента схемы. Известно, что t_i распределено по закону $Ex(\lambda_i)$, $i = 1, 2, 3$. Используя выборки, полученные в задаче 15.13, построить эмпирическую функцию распределения времени выработки на отказ для схемы.

3.4 Точные выборочные распределения

Распределение хи-квадрат. Пусть $\xi_0, \xi_1, ..., \xi_n$ независимы, распределены нормально-стандартно с параметрами (0, 1). Строим новую случайную величину $\chi_n^2 = \xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$ и ищем её функцию распределения:

$$F_{\chi_n^2}(x) = P(\chi_n^2 < x) = P(\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2 < x) =$$

$$= \int\limits_{x_1^2+x_2^2+\dots+x_n^2<x} \cdots \int \frac{1}{(\sqrt{2\pi})^n}\, e^{-\frac{x_1^2+x_2^2+\dots+x_n^2}{2}} dx_1\cdots dx_n.$$

В n-мерном пространстве переменных $x_1, x_2, ..., x_n$ перейдем к обобщенным сферическим координатам $r, \varphi, \theta, \theta_1, ..., \theta_{n-3}$:

$$\begin{array}{llllllll}
x_1 & = r & \cos\varphi & \sin\theta & \sin\theta_1 & \sin\theta_2 & \cdots & \sin\theta_{n-3}, \\
x_2 & = r & \sin\varphi & \sin\theta & \sin\theta_1 & \sin\theta_2 & \cdots & \sin\theta_{n-3}, \\
x_3 & = r & \cos\theta & & \sin\theta_1 & \sin\theta_2 & \cdots & \sin\theta_{n-3}, \\
x_4 & = r & \cos\theta_1 & & & \sin\theta_2 & \cdots & \sin\theta_{n-3}, \\
. & . & . & . & . & . & \cdots & . \\
x_n & = r & \cos\theta_{n-3} & & & & & \sin\theta_{n-3}.
\end{array}$$

Как известно, для сферических координат в трехмерном пространстве якобиан перехода равен $J = r^2 \sin\theta$, в общем случае можно показать, что $J = r^{n-1} f(\theta, \theta_1, ..., \theta_{n-3})$, где f есть некоторая функция своих аргументов. Поэтому

$$F_{\chi_n^2}(x) = \frac{1}{(\sqrt{2\pi})^n} \int \cdots \int f(\theta, ...\theta_{n-3}) d\varphi d\theta d\theta_1 \cdots d\theta_{n-3} \int\limits_0^{\sqrt{x}} r^{n-1} e^{-\frac{r^2}{2}} dr$$

$$= C \int\limits_0^{\sqrt{x}} r^{n-1} e^{-\frac{r^2}{2}} dr.$$

Очевидно, что $F_{\chi_n^2}(\infty) =$

$$= 1 = C \int\limits_0^{\infty} r^{n-1} e^{-\frac{r^2}{2}} dr = 2^{n/2-1} C \int\limits_0^{\infty} t^{n/2-1} e^{-t} dt = 2^{n/2-1} C\Gamma(n/2).$$

При вычислении интеграла была сделана замена: $r^2/2 = t$. Γ – гамма-функция Эйлера. Поэтому $C^{-1} = 2^{n/2-1}\Gamma(n/2)$. Случайная величина с такой функцией распределения называется случайной величиной, подчиняющейся распределению хи-квадрат с n степенями свободы. Значения функции распределения этой случайной величины табулированы.

Распределение Стьюдента с n степенями свободы. Так называется распределение случайной величины $\tau_n = \frac{\xi_0}{\sqrt{\chi_n^2/n}}$. Функция распределения вычисляется аналогично предыдущему случаю. Её значения также табулированы.

3.5 Теоремы о распределениях, связанных с выборочными моментами для нормального распределения

Теорема 1. *Если элементы выборки $X = (X_1, ..., X_n)$ распределены нормально с параметрами (a, σ^2) и независимы, то независимы $\bar{X}$ и m_2, причем $\bar{X}$ распределено нормально с параметрами $(a, \frac{\sigma^2}{n})$, а $\frac{nm_2}{\sigma^2}$ подчинено распределению χ^2 с $n-1$ степенями свободы.*

Доказательство. Введем величины $Y_k = \frac{X_k - a}{\sigma}$. Y_k распределены нормально-стандартно. Пусть

$$Z = \big(Z_1, ..., Z_n\big) = YC = \big(Y_1, ..., Y_n\big) \begin{pmatrix} 1/\sqrt{n} & \cdots \\ \cdot & \cdots \\ 1/\sqrt{n} & \cdots \end{pmatrix},$$

где C — ортогональная матрица ($C^T C = 1$), у которой элементы первого столбца одинаковы и равны $1/\sqrt{n}$, а остальные элементы произвольны. Имеем:

$$Z_1 = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} = \sqrt{n}\bar{Y},\ \ \bar{Y} = \frac{\sum_{k=1}^n Y_k}{n},$$

$$\bar{X} = \frac{\sum_{k=1}^n X_k}{n} = \frac{\sum_{k=1}^n (\sigma Y_k + a)}{n} = \sigma\bar{Y} + a = a + \frac{\sigma Z_1}{\sqrt{n}},$$

$$\frac{nm_2}{\sigma^2} = \frac{n}{\sigma^2}\frac{\sum_{k=1}^n (X_k - \bar{X})^2}{n} = \sum_{k=1}^n (Y_k - \bar{Y})^2 = \sum_{k=1}^n (Y_k^2 - 2Y_k\bar{Y} + \bar{Y}^2) =$$

$$= \sum_{k=1}^n Y_k^2 - n\bar{Y}^2 = \sum_{k=1}^n Z_k^2 - Z_1^2 = \sum_{k=2}^n Z_k^2 = \chi_{n-1}^2.$$

В ходе вычислений мы воспользовались тем, что при ортогональном преобразовании $Z = YC$ имеет место $\sum\limits_{k=1}^{n} Y_k^2 = \sum\limits_{k=1}^{n} Z_k^2$.

Теорема 2. *Величина* $\frac{\bar{X}-a}{\sqrt{m_2}}\sqrt{n-1}$ *подчиняется распределению Стьюдента с* $(n-1)$ *степенями свободы.*

Доказательство.

$$\frac{\bar{X}-a}{\sqrt{m_2}}\sqrt{n-1} = \frac{\sigma Z_1/\sqrt{n}}{\sqrt{\sigma^2\chi^2_{n-1}/n}} = \frac{Z_1}{\sqrt{\chi^2_{n-1}/n}} = \tau_{n-1}.$$

3.5.1 9-ое практическое занятие. Точечные оценки

Задача 9.7. По выборке $(X_1, X_2, ..., X_n)$, где X_k, $k = 1, ..., n$, независимы, $MX_k = a$, $DX_k = \sigma_k^2$ (σ_k известны), найти несмещенную линейную относительно $X_1, X_2, ..., X_n$ оценку a^* параметра a с наименьшей дисперсией.

Решение. Пусть $a^* = \alpha_1 X_1 + \dots + \alpha_n X_n$, тогда $Ma^* = a(\alpha_1 + \dots + \alpha_n)$, $Da^* = \alpha_1^2\sigma_1^2 + \dots + \alpha_n^2\sigma_n^2$. Так как $Ma^* = a$, то $\alpha_1 + \dots + \alpha_n = 1$. Значит, нужно найти минимум функции $\alpha_1^2\sigma_1^2 + \dots + \alpha_n^2\sigma_n^2$ при условии $\alpha_1 + \dots + \alpha_n = 1$. Применяем метод неопределенных множителей Лагранжа, получаем:

$$\frac{\partial}{\partial\alpha_i}\{\alpha_1^2\sigma_1^2 + \dots + \alpha_n^2\sigma_n^2 + \lambda(\alpha_1 + \dots + \alpha_n - 1)\} = 0,$$

$$2\sigma_i^2\alpha_i + \lambda = 0,\ \alpha_i = -\frac{\lambda}{2\sigma_i^2}.$$

$$\sum \alpha_i = -\frac{\lambda}{2}\sum\frac{1}{\sigma_i^2} = 1,\ \lambda = -\frac{2}{\sum\limits_{i=1}^{n}\frac{1}{\sigma_i^2}},\ a^* = \sum_{k=1}^{n}\frac{X_k}{\sigma_i^2} \Big/ \sum_{i=1}^{n}\frac{1}{\sigma_i^2}.$$

Задача 9.9(а). Пусть $x_1, ..., x_n$ — независимые нормально распределенные случайные величины с параметрами (0, 1). Положим $\bar{X}_k = \frac{1}{k}\sum\limits_{i=1}^{k} x_i$, $S_k = \sum\limits_{i=1}^{k}(x_i - \bar{X}_k)^2$, $k = 1, 2, ..., n$.
Доказать, что а) величины $\bar{X}_{k+1} - \bar{X}_k$ и $\bar{X}_{l+1} - \bar{X}_l$ независимы; б)

величина S_n имеет распределение хи-квадрат с $n-1$ степенями свободы.

Решение. а) Линейные комбинации независимых случайных величин независимы, если их ковариация равна нулю. Пусть $l < k$.

$$cov(\bar{X}_{k+1} - \bar{X}_k, \bar{X}_{l+1} - \bar{X}_l) = cov(\bar{X}_{k+1}, \bar{X}_{l+1}) - cov(\bar{X}_{k+1}, \bar{X}_l) -$$

$$-cov(\bar{X}_k, \bar{X}_{l+1}) + cov(\bar{X}_k, \bar{X}_l) = \frac{1}{(k+1)(l+1)} \sum_{i=1}^{l+1} DX_i + \frac{1}{kl} \sum_{i=1}^{l} DX_i -$$

$$-\frac{1}{(k+1)l} \sum_{i=1}^{k} DX_i - \frac{1}{k(l+1)} \sum_{i=1}^{l+1} DX_i = 0.$$

б) $S_{k+1} = \sum\limits_{i=1}^{k+1} (x_i - \bar{X}_{k+1})^2 = \sum\limits_{i=1}^{k} (x_i - \bar{X}_k + \bar{X}_k - \bar{X}_{k+1})^2 + (x_{k+1} - \bar{X}_{k+1})^2$

$$= \sum_{i=1}^{k} \left((x_i - \bar{X}_k) + 2(x_i - \bar{X}_k)(\bar{X}_k - \bar{X}_{k+1}) + (\bar{X}_k - \bar{X}_{k+1})^2 \right) +$$

$$+ \left((k+1)\bar{X}_{k+1} - k\bar{X}_k - \bar{X}_{k+1} \right)^2 = S_k + k(k+1)(\bar{X}_k - \bar{X}_{k+1})^2.$$

$$D(\sqrt{k(k+1)}(\bar{X}_k - \bar{X}_{k+1})^2) = D\frac{kx_{k+1} - x_1 - \cdots - x_k}{\sqrt{k(k+1)}} = \frac{k^2+k}{k(k+1)} = 1.$$

$$S_n = \sum_{k=1}^{n-1} k(k+1)(\bar{X}_k - \bar{X}_{k+1})^2.$$

Задача 9.10. Используя метод наибольшего правдоподобия, найти по выборке $X_1, ..., X_n$, где $P(X_k = m) = \frac{\lambda^m}{m!} e^{-\lambda}$, $m = 0, 1, ...$, оценку λ^* параметра λ. Будет ли эта оценка несмещенной и состоятельной? Найти $M\lambda^*$, $D\lambda^*$.

Решение.

$$L = \frac{\lambda^{X_1}}{X_1!} e^{-\lambda} \frac{\lambda^{X_2}}{X_2!} e^{-\lambda} \cdots \frac{\lambda^{X_n}}{X_n!} e^{-\lambda} = \frac{\lambda^{\sum\limits_{i=1}^{n} X_i}}{\prod\limits_{k=1}^{n} X_k!} e^{-n\lambda},$$

$$\ln L = -n\lambda + \sum_{i=1}^{n} X_i \ln\lambda - \sum_{i=1}^{n} \ln(X_i!),\ \ \frac{\partial \ln L}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i = 0,$$

$$\lambda^* = \frac{\sum\limits_{i=1}^{n} X_i}{n}.$$

$M\lambda^* = \frac{\sum\limits_{i=1}^{n} MX_i}{n} = \lambda$, λ^* — несмещенная оценка. $P(|\lambda^* - \lambda| \geq \epsilon) \leq \frac{D\lambda^*}{\epsilon^2} \to 0$ при $n \to \infty$, поэтому λ^* — состоятельная оценка.

Задача 15.116. В качестве оценки математического ожидания $m = M\xi$ по выборке $(X_1, X_2, ..., X_n)$ предлагается взять статистику $m^* = X_1$. Проверить несмещенность и состоятельность этой оценки.

Решение. Несмещенность: $Mm^* = MX_1 = m$. Состоятельность: $P(|m^* - m| \geq \epsilon) = P(|X_1 - m| \geq \epsilon) \leq \frac{DX_1}{\epsilon^2}$. Отсюда не следует, что эта вероятность стремится к нулю. С другой стороны, заданная случайная величина X_1 не может стремиться по вероятности к постоянной величине. Поэтому m^* — несостоятельная оценка.

Задача 15.118. Случайная величина X имеет распределение с плотностью $p(x)$, равной e^{a-x} при $x \geq a$ и нулю при $x < a$. Для оценки неизвестного параметра a по выборке $X_1, X_2, ..., X_n$ для случайной величины X предлагается выбрать статистику $a^* = X_{(1)} = \min\limits_{1\leq i\leq n} X_i$. Проверить несмещенность и состоятельность этой оценки.

Решение. $F_X(x) = \int\limits_a^x e^{a-u}du = 1 - e^{a-x}$, $x \geq a$.

$$F_{X_{(1)}} = P(X_{(1)} < x) = 1 - P(X_{(1)} \geq x) = 1 - \prod_{i=1}^{n} P(X_i > x) =$$

$$= 1 - (1 - F_X(x))^n = 1 - e^{n(a-x)},\ x \geq a.$$

$MX_{(1)} = \int\limits_a^\infty xne^{n(a-x)}dx = a + \frac{1}{n} \neq a$. Оценка смещенная.

$$P(|X_{(1)} - a| \geq \epsilon) \leq \frac{M(X_{(1)} - a)^2}{\epsilon^2},$$

$$M(X_{(1)}-a)^2=\int\limits_a^\infty (x-a)^2 ne^{n(a-x)}dx=\frac{2}{n^2}\to 0$$

при $n\to\infty$. Оценка состоятельная.

Задачи домашнего задания.

Задача 9.8. Пусть $x_{in_1},...,x_{in_i}$, $i=1,...,I$, — независимые нормально распределенные величины с параметрами (a,σ^2);

$$\bar{x}_i=\frac{1}{n_i}\sum_{k=1}^{n_i}x_{ik},\ s_i^2=\frac{1}{n_i-1}\sum_{k=1}^{n_i}(x_{ik}-\bar{x}_i)^2.$$

Показать, что величина $\bar{x}=\sum\limits_{i=1}^{I}c_i\bar{x}_i$, где $c_i=(1/s_i^2)/(1/s_1^2+...+1/s_I^2)$, является несмещенной оценкой a.

Задача 15.117. Рассмотрим две выборки объемов n_1 и n_2 из одной генеральной совокупности со средним m и дисперсией σ^2. Пусть $\bar{X}_1$, $\bar{X}_2$, S_1^2 и S_2^2 — несмещенные оценки средних и дисперсий, определенные по этим выборкам. Показать, что объединенные оценки, вычисляемые по формулам

$$\bar{X}=\frac{n_1\bar{X}_1+n_2\bar{X}_2}{n_1+n_2},\ S^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2},$$

будут несмещенными и состоятельными оценками m и σ^2.

Задача 15.119. Пусть $x_1,x_2,...,x_n$ — выборка из генеральной совокупности, имеющей равномерное распределение $R(0,1)$. Показать, что статистика

$$\tilde{m}=\frac{1}{2}(x^{(1)}+x^{(n)}),$$

где $x^{(1)}$ и $x^{(n)}$ — соответственно наименьший и наибольший элементы выборки, является несмещенной и состоятельной оценкой математического ожидания m.

3.6 Интервальные оценки

Пусть дана выборка $X = (X_1, ..., X_n)$ с $P(X_i < x) = F(x, \Theta)$, где Θ — неизвестный параметр. Пусть удается найти такие функции $\underline{\Theta}(X_1, ..., X_n)$, $\overline{\Theta}(X_1, ..., X_n)$, что для всех $X = (X_1, ..., X_n)$ $\underline{\Theta}(X) < \overline{\Theta}(X)$ и $P(\underline{\Theta}(X) < \Theta < \overline{\Theta}(X)) = 1 - 2\alpha$, тогда $(\underline{\Theta}(X), \overline{\Theta}(X))$ — доверительный интервал для Θ, $1 - 2\alpha$ — доверительная вероятность, 2α — уровень значимости.

Пример. *Доверительные интервалы для параметров нормального распределения.* Пусть $X = (X_1, ..., X_n)$ — выборка, элементы которой независимы и распределены нормально с параметрами (a, σ^2).

1) *Доверительный интервал для a, если известно σ^2.* X подчиняется нормальному распределению $N(a, \sigma^2/n)$, тогда $\frac{\bar{X}-a}{\sigma/\sqrt{n}}$ — подчиняется $N(0,1)$. Пусть $\frac{1}{\sqrt{2\pi}}\int\limits_{u_\alpha}^{\infty} e^{-\frac{x^2}{2}}dx = \alpha$. Тогда $P\left(\left|\frac{(\bar{X}-a)\sqrt{n}}{\sigma}\right| < u_\alpha\right) = 1 - 2\alpha$. Поэтому доверительный интервал имеет вид:

$$\bar{X} - \frac{u_\alpha \sigma}{\sqrt{n}} < a < \bar{X} + \frac{u_\alpha \sigma}{\sqrt{n}}.$$

2) *Доверительный интервал для a, если неизвестно σ^2.* $\frac{\bar{X}-a}{\sqrt{m_2}}\sqrt{n-1}$ подчиняется распределению Стьюдента τ_{n-1}. Если $P(|\tau_{n-1}| < t_{\alpha,n-1}) = 1 - 2\alpha$, то доверительный интервал имеет вид:

$$\bar{X} - \frac{m_2}{\sqrt{n-1}}t_{\alpha,n-1} < a < \bar{X} + \frac{m_2}{\sqrt{n-1}}t_{\alpha,n-1}.$$

3) *Доверительный интервал для σ^2, если известно a.* Так как X подчиняется $N(a, \sigma^2)$, то $\frac{X_k - a}{\sigma}$ — подчиняется $N(0,1)$. Если ввести $S^2 = \sum\limits_{k=1}^{n}(X_k - a)^2$, то S^2/σ^2 будет подчиняться распределению χ^2_n. Положим $P(\chi^2_n > u_{n,\alpha}) = \alpha$, тогда также $P(\chi^2_n > u_{n,1-\alpha}) = 1 - \alpha$, и имеем

$$P(u_{n,1-\alpha} < \chi^2_n < u_{n,\alpha}) = P(\chi^2_n < u_{n,\alpha}) - P(\chi^2_n < u_{n,1-\alpha}) =$$

$$= 1 - \alpha - (1 - (1 - \alpha)) = 1 - 2\alpha.$$

Неравенство $u_{n,1-\alpha} < S^2/\sigma^2 < u_{n,\alpha}$ можно переписать в виде $S^2/u_{n,\alpha} < \sigma^2 < S^2/u_{n,1-\alpha}$, поэтому

$$P(S^2/u_{n,\alpha} < \sigma^2 < S^2/u_{n,1-\alpha}) = 1 - 2\alpha.$$

4) *Доверительный интервал для σ^2, если неизвестен a.* $\frac{nm_2}{\sigma^2}$ подчиняется распределению χ^2_{n-1}, и если обозначить $P(\chi^2_{n-1} > u_{n-1,\alpha}) = \alpha$, то

$$P(nm_2/u_{n-1,\alpha} < \sigma^2 < nm_2/u_{n-1,1-\alpha}) = 1 - 2\alpha.$$

3.7 Статистическая проверка гипотез

При обработке данных наблюдений можно выдвигать разные предположения (гипотезы) о свойствах искомой вероятностной модели. Возникает проблема проверки правильности этих гипотез. Мы сначала изучим один из критериев по проверке одной определенной гипотезы, а затем перейдем к рассмотрению общих принципов статистической проверки гипотез.

3.7.1 Критерий согласия хи-квадрат Пирсона

По реализации выборки $X = (X_1, ..., X_n)$ нужно решить, является ли *заданная функция $F(x)$ функцией распределения случайной величины ξ*. Разобьем числовую ось точками $z_1 < z_2 < \cdots < z_r$ на $(r+1)$ частей:

$$(-\infty, z_1), [z_1, z_2), ..., [z_{r-1}, z_r), [z_r, \infty).$$

Пусть

$$p_1 = P(\xi \in (-\infty, z_1)), ..., p_{i+1} = P(\xi \in [z_i, z_{i+1})), ...,$$

$$p_{r+1} = P(\xi \in (z_r, \infty)),$$

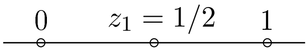
$x = (x_1, ..., x_n)$, m_k – число значений случайной величины, попавших в интервал $[z_{k-1}, z_k)$, $\sum\limits_{k=1}^{r+1} m_k = n$. Если наш выбор правилен (в смысле взятия $F(x)$ в качестве функции распределения), то

$$\frac{m_k}{n} \xrightarrow[n\to\infty]{P} p_k$$

по теореме Бернулли. По интегральной теореме Муавра – Лапласа величина $\frac{m_k - np_k}{\sqrt{np_k q_k}}$ $(q_k = 1 - p_k)$ — асимптотически нормальна с параметрами (0,1), а величина $\eta_k = \frac{m_k - np_k}{\sqrt{np_k}}$ — асимптотически нормальна с параметрами $(0, q_k)$. Но $\sum\limits_{k=1}^{r+1} \sqrt{p_k}\eta_k = \frac{1}{\sqrt{n}}(\sum\limits_{k=1}^{r+1} m_k - n\sum\limits_{k=1}^{r+1} p_k) = \frac{1}{\sqrt{n}}(n-n) = 0$. Значит, независимых величин — r штук. Можно показать, что при больших n сумма $\sum\limits_{k=1}^{r+1} \eta_k^2$ подчиняется распределению χ_r^2. Значит, при больших n сумма $\sum\limits_{k=1}^{r+1} \frac{(m_k - np_k)^2}{np_k}$ подчиняется распределению χ^2 с r степенями свободы. В итоге получаем

Критерий χ^2: *По таблицам по α (например, $\alpha = 0.05$) находим $\chi^2_{r,\alpha}$ такое, что $P(\chi^2_r > \chi^2_{r,\alpha}) = \alpha$. Вычисляем $\chi^2_{r,выч} = \sum\limits_{k=1}^{r+1} \frac{(m_k - np_k)^2}{np_k}$. Если p_k, а значит, и $F(x)$, угаданы правильно, то $\chi^2_{r,выч}$ должно быть небольшим. Если $\chi^2_{r,выч} < \chi^2_{r,\alpha}$, то считается, что гипотеза о $F(x)$ не противоречит опыту. Если же случилось маловероятное $\chi^2_{r,выч} > \chi^2_{r,\alpha}$, то считается, что гипотеза о $F(x)$ неправильна.*

Новый анализ опыта Бюффона. В опыте Бюффона монета была брошена $n = 4040$ раз, число выпадений герба равно $m_1 = 2048$, $m_2 = 1992$ – число выпадений решетки. Гипотеза: вероятности выпадений герба и решетки равны между собой и имеют значения $p = q = \frac{1}{2}$. Опыту с бросанием монеты сопоставляем дискретную случайную величину ξ с функцией распределения: $F(x) = 0$, если $x \le 0$; $F(x) = 1/2$, если $0 < x \le 1$; $F(x) = 1$, если $1 < x$. Числовую прямую делим на две части с помощью точки $z_1 = 1/2$.



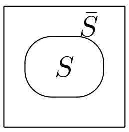
$r = 1$, $\alpha = 0.05$, $\chi^2_{1,0.05} = 3.84$. Вычисляем

$$\chi^2_{1,\text{выч}} = \sum_{k=1}^{r+1} \frac{(m_k - np_k)^2}{np_k} = \frac{(2048 - 2020)^2}{4040 \cdot 1/2} + \frac{(1992 - 2020)^2}{4040 \cdot 1/2} = 0.78.$$

Видим, что $\chi^2_{1,\text{выч}} < \chi^2_{1,0.05}$, поэтому можно считать, что гипотеза не противоречит эксперименту.

3.7.2 Общие принципы статистической прверки гипотез

Пусть $X = (X_1, ..., X_n)$ — некоторая выборка. Нужно проверить, что гипотеза H_0 верна или не верна. Пусть S – *критическая область для проверки гипотезы* H_0*: если* $X \in S$*, то будем считать, что гипотеза не верна, если же* $X \in \bar{S}$*, то будем считать, что гипотеза верна.* Вероятность $P_{H_0}(S) = \alpha$ называется уровнем значимости. Но может быть, что гипотеза H_0 верна, но $X \in S$, т.е. гипотезу H_0 нужно считать неверной. Мы совершаем *ошибку первого рода: отвергается верная гипотеза.* $\alpha = P_{H_0}(S)$ – вероятность ошибки первого рода. Но может быть, что гипотеза H_0 верна, но $X \in S$, т.е. гипотезу H_0 нужно считать неверной. Мы совершаем *ошибку первого рода: отвергается верная гипотеза.* $\alpha = P_{H_0}(S)$ – вероятность ошибки первого рода.



Но может быть, что гипотеза H_0 не верна, но $X \in \bar{S}$, тогда мы совершаем *ошибку 2-го рода, когда неверную гипотезу принимаем за верную.* Чтобы оценить вероятность ошибки 2-го рода, допустим, что имеется конкурирующая гипотеза H_1.

Тогда $P_{H_1}(\bar{S}) = \beta$ – вероятность ошибки 2-го рода. β должна иметь наименьшее значение. Вероятность $1 - \beta = P_{H_1}(S)$ опровергнуть неверную гипотезу называется мощностью критерия. S нужно подобрать так, чтобы при заданном α величина $1 - \beta$ имела наибольшее значение.

Пусть $L_{H_0}(X_1, ..., X_n)$ – функция правдоподобия для выборки X при гипотезе H_0, $L_{H_1}(X_1, ..., X_n)$ – то же самое при гипотезе H_1. Положим $S_C = \{X = (X_1, ..., X_n) : \ L_{H_1}/L_{H_0} \geq C\}$. Пусть для любого α найдется такое C, что $P_{H_0}(S_C) = \alpha$.

Теорема Неймана – Пирсона. *Среди всех критериев, различающих гипотезы* H_0 *и* H_1 *с заданной ошибкой первого рода* α*, наиболее мощным является критерий, определяемый критическим множеством* S_C.

Доказательство. Пусть S – некоторое критическое множество $P_{H_0}(S)=\alpha$. Имеем

$$P_{H_0}(S_C\backslash S_CS)=\alpha-P_{H_0}(S_CS)=P_{H_0}(S)-P_{H_0}(S_CS)=P_{H_0}(S\backslash S_CS),$$

поэтому

$$P_{H_1}(S_C\backslash S_CS)=\int\limits_{S_C\backslash S_CS}L_{H_1}dx\geq C\int\limits_{S_C\backslash S_CS}L_{H_0}dx=CP_{H_0}(S_C\backslash S_CS)=$$

$$=CP_{H_0}(S\backslash S_CS)=\int\limits_{S\backslash S_CS}L_{H_0}dx\geq\int\limits_{S\backslash S_CS}L_{H_1}dx=P_{H_1}(S_C\backslash S_CS).$$

В итоге получаем

$$P_{H_1}(S)=P_{H_1}(S\backslash S_CS)+P_{H_1}(S_CS)\leq$$

$$\leq P_{H_1}(S_C\backslash S_CS)+P_{H_1}(S_CS)=P_{H_1}(S_C).$$

Пример. *Бернуллиевская модель*. Пусть $H_0:\ p=p_0$, $H_1:\ p=p_1$ и $p_0<p_1$. Так как $L_{H_0}=p_0^m(1-p_0)^{n-m}$ и $L_{H_1}=p_1^m(1-p_1)^{n-m}$, то критическое множество S_C задается с помощью неравенства $L_{H_1}/L_{H_0}=\left(\left(\frac{p_1}{1-p_1}\right)/\left(\frac{p_0}{1-p_0}\right)\right)^m\left(\frac{1-p_1}{1-p_0}\right)^n\geq C$, откуда $m\geq\left(\ln C+n\ln\frac{1-p_0}{1-p_1}\right)/\ln\frac{p_1(1-p_0)}{p_0(1-p_1)}=m_\alpha$, значит $S_C=\{m\geq m_\alpha\}$. Вычисляем

$$P_{H_0}(S)=\sum_{m\geq m_\alpha}C_n^mp_0^m(1-p_0)^{n-m}\approx\frac{1}{\sqrt{2\pi}}\int\limits_{\frac{m_\alpha-np_0}{\sqrt{np_0q_0}}=u_\alpha}^{\infty}e^{-\frac{u^2}{2}}du=\alpha,$$

откуда $m_\alpha=np_0+\sqrt{np_0q_0}u_\alpha$.

Приложение к опыту Бюффона. H_0: $p=\frac{1}{2}$, H_1: $p>\frac{1}{2}$, $\alpha=0.05$. Из таблиц для нормально стандартного распределения находим $u_\alpha=1.65$, значит $m_\alpha=2020+1.65\sqrt{\frac{4040}{4}}\approx 2072$. Но $m=2048\in\bar{S}$, поэтому гипотеза H_0 принимается, а гипотеза H_1 отвергается.

3.7.3 10-ое практическое занятие. Доверительный интервал. Статистическая проверка гипотез

Задача 9.13. По выборке, полученной в задаче 9.2, построить доверительный интервал для a с доверительной вероятностью 0.95 и для σ^2 с доверительной вероятностью 0.94.

Решение. a и σ^2 неизвестны. Для построения доверительного интервала для a воспользуемся случайной величиной $\tau_{n-1} = \frac{(\bar{X}-a)\sqrt{n-1}}{\sqrt{m_2}}$, подчиненной распределению Стьюдента с $n-1$ степенями свободы, $n = 22$, $\bar{X} = 0.6149$, $m_2 = 1.0678$. $\tau_{0.025,21} = 2.08$, $\sqrt{m_2/(n-1)} = 0.2256$, $\sqrt{m_2/(n-1)}\tau_{\alpha,n-1} = 0.469$. Доверительный интервал: $0.146 < a < 1.084$.

Для построения доверительного интервала для σ^2 используем случайную величину $m_2 n/\sigma^2$, подчиненную распределению χ^2_{n-1}. Из таблицы находим: $\chi^2_{0.99,21} = 8.9$, $\chi^2_{0.05,21} = 32.7$. $22m_2/\chi^2_{0.05,21} = 0.7181$, $22m_2\chi^2_{0.99,21} = 2.6395$. $P(0.7181 < \sigma^2 < 2.6395) = 0.94$.

Задача 9.15. Найти статистику наиболее мощного критерия, различающего по выборке $X_1, X_2, ..., X_n$ гипотезы

$$H_0: \ P(X_k = i) = p_i^0 > 0; \ \ H_1: \ P(X_k = i) = p_i^1 > 0; \ i = 1, ..., N;$$

$$\sum_{i=1}^{n} p_i^{(0)} = \sum_{i=1}^{n} p_i^{(1)} = 1.$$

Решение. $L_{H_0}(X_1, X_2, ..., X_n) = {p_1^{(0)}}^{m_1} {p_2^{(0)}}^{m_2} \cdots {p_N^{(0)}}^{m_N}$, m_k — число случаев, когда $X = k$. $L_{H_1}(X_1, X_2, ..., X_n) = {p_1^{(1)}}^{m_1} {p_2^{(2)}}^{m_2} \cdots {p_N^{(1)}}^{m_N}$. Статистика наиболее мощного критерия:

$$\ln(L_{H_1}/L_{H_0}) = m_1 \ln(p_1^{(1)}/p_1^{(0)}) + m_2 \ln(p_2^{(1)}/p_2^{(0)}) + \cdots m_N \ln(p_N^{(1)}/p_N^{(0)}).$$

Задача 9.20. Пусть $X_1, X_2, ..., X_{15}$ — выборка, для которой $P(X_k = l) = p_l$, $l = 1, 2, 3$. Получить 10 реализаций этой выборки с $p_l = p_l^{(0)} = 1/3$, $l = 1, 2, 3$, и 10 реализаций с $p_1 = p_1^{(1)} = 0,40$, $p_2 = p_2^{(1)} = 0,42$, $p_3 = p_3^{(1)} = 0,18$. Для каждой выборки, используя наиболее мощный критерий, найденный в задаче 9.15, выбрать одну из двух гипотез. Для вычисления ошибок 1-го и 2-го рода α

и β использовать нормальное приближение. Выбрать $\alpha = \beta$. Найти частоты ошибок 1-го и 2-го рода.

Решение. Для построения десяти реализаций выборки для гипотезы H_0 в каждой из десяти первых строк таблицы случайных чисел берем 15 первых отличных от нуля чисел, среди них отбираем m_1 чисел, находящихся между 1 и 2, m_2 – между 4 и 6, m_3 – между 7 и 9. Дня построения реализаций выборки для гипотезы H_1 будем считать, что m_1 — число двузначных чисел между 0 и 39, m_2 — между 40 и 81, m_3 — между 82 и 99. Получаем

H_0	m_1	8	2	4	5	5	5	7	6	3	6
	m_2	4	8	6	2	2	2	6	6	6	2
	m_3	3	5	5	8	8	8	2	3	6	7
H_1	m_1	8	5	7	11	4	6	11	9	4	5
	m_2	6	8	5	2	10	6	3	3	9	8
	m_3	1	2	3	2	1	3	1	3	2	2

.

$$ln(L_{H_1}/L_{H_0}) = 0.182(15 - m_2 - m_3) + 0.231m_2 - 0.616m_3 =$$
$$= 0.049m_2 - 0.798m_2 + 2.73.$$

При гипотезе H_0 величины m_2 и m_3 считаем подчиняющимися приближенно нормальному распределению с параметрами $(15p_1^{(0)}, 15p_1^{(0)} \cdot (1 - p_1^{(0)})) = (5, 1.825^2)$. Рассмотрим $P_{H_0}(m_2 + m_3 > 15)$. Положим $\xi_i = (m_i - 5)/1.825$, $i = 2, 3$. ξ_i распределены нормально-стандартно. Тогда $P_{H_0}(m_2 + m_3 > 15) = P((\xi_1 + \xi_2)/\sqrt{2} \geq 1.94) = 0.026$. Поэтому в определении критического множества

$$S_C = \{(m_2, m_3) : \ 0.049m_2 - 0.798m_3 > C; m_2 + m_3 \leq 15\}$$

условие $m_2 + m_3 \leq 15$ можно считать достоверным событием, значит

$$S_C = \{(m_2, m_3) : \ 0.049m_2 - 0.798m_3 > C\} =$$
$$= \{(\xi_2, \xi_3) : \ 0.061\xi_2 - 0.998\xi_3 > \frac{C + 1.377}{1.469}\}.$$

Ошибка первого рода

$$\alpha = P_{H_0}(S_C) = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{C+1.377}{1.469}}^{\infty} e^{-\frac{t^2}{2}} dt.$$

Для гипотезы $H_1 : m_2 \sim N(15p_2^{(1)}, 15p_2^{(1)}q_2^{(1)}) = N(6.3, 1.912^2)$, $m_3 \sim N(15p_3^{(1)}, 15p_3^{(1)}q_3^{(1)}) = N(2.7, 1.488^2)$. Положим $\eta_2 = \frac{m_2-6.3}{1.912}$, $\eta_2 = \frac{m_3-2.7}{1.488}$. Тогда $P(m_2+m_3 > 15) = P(0.789\eta_2 + 0.614\eta_3 > 2.248) = 0.007$. Этой вероятностью тоже пренебрегаем.

$$\beta = P_{H_1}(\bar{S}_C) = P(0.049(1.912\eta_2 + 6.3) - 0.798(1.488\eta_3 + 2.7) \leq C) =$$

$$= (0.079\eta_2 - 0.997\eta_3 \leq \frac{C - 1.846}{1.191}) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\frac{C-1.846}{1.191}}^{\infty} e^{-\frac{t^2}{2}} dt.$$

Из $\alpha = \beta$ следует $\frac{C+1.377}{1.469} = -\frac{C-1.846}{1.191}$, откуда $C = -1.63$. Значения $0.049m_2 - 0.798m_3$:

H_0	-2.2	-3.6	-3.70	-6.3	-6.30	-6.3	-1.30	-2.10	-4.50	-5.6
H_1	-0.5	-1.2	-2.15	-1.5	-0.31	-2.1	-0.65	-2.25	-1.15	-1.20

Частота ошибки первого рода: 0.1, второго рода — 0.3.

Задача 15.170. В результате $n = 16$ наблюдений для емкости конденсатора найдено $\bar{X} = 20$ мкф. Найти 90%-ные и 99%-ные доверительные интервалы для математического ожидания емкости. Среднеквадратичное отклонение $\sigma = 4$ мкф.

Решение. Так как среднеквадратичное отклонение известно, $\sigma = 4$ мкф, то нужно использовать статистику $\frac{(\bar{X}-m)\sqrt{n}}{\sigma}$, имеющую нормальное стандартное распределение. Рассматриваем два случая: $\alpha = 0.05$ и $\alpha = 0.005$. В таблице решений уравнения $\frac{1}{\sqrt{2\pi}} \int\limits_{u_\alpha}^{\infty} e^{-\frac{u^2}{2}} du = \alpha$ находим соответственно значения $u_{0.05} = 1.65$ и $u_{0.005} = 2.58$. Тогда $P(20 - 1.65 < m < 20 + 1.65) = 0.9$ и $P(20 - 2.26 < m < 20 + 2.26) = 0.99$.

Задача 15.196. Из большой партии транзисторов одного типа были случайным образом отобраны и проверены 100 штук. У

36 транзиоторов коэффициент усиления оказался меньше 10. Найти 95% -ный доверительный интервал для доли таких транзисторов по всей партии.

Решение. Пусть p — вероятность того, что у транзистора коэффициент усиления меньше 10. $P(|\frac{m-np}{\sqrt{npq}}| < u_\alpha) = 1 - 2\alpha = 0.95$. $\alpha = 0.025$, $u_\alpha = 1.96$. Имеем $|\frac{m-np}{n}| < \frac{u_\alpha\sqrt{pq}}{\sqrt{n}} \leq \frac{u_\alpha}{2\sqrt{n}}$, так как $\sqrt{pq} \leq 0.5$. $|\frac{m}{n} - p| < \frac{1.96}{2 \cdot 10} = 0.098 \to 0.36 - 0.1 < p < 0.36 + 0.1$.

Задача 15.218. Большая партия изделий может содержать некоторую долю дефектных. Поставщик утверждает, что эта доля составляет 5%; покупатель предполагает, что доля дефектных изделий равна 10%. Условия поставки: из партии случайным образом отбирается и проверяется 10 изделий; партия принимается и проверяется на условиях поставщика, если при проверке обнаружено не более одного дефектного изделия; в противном случае партия принимается на условиях покупателя. Сформулировать эту задачу в терминах теории проверки статистических гипотез и ответить на следующие вопросы:

а) Каковы статистика критерия, область ее значений, критическая область?

б) Какое распределение имеет статистика критерия?

в) В чем состоят проверяемая и альтернативная гипотезы?

г) В чем состоят ошибки первого и второго рода и каковы их вероятности?

Решение. Гипотеза поставщика $H_0 : p = p_0 = 0.05$. Гипотеза покупателя: $H_1 : p = p_1 = 0.1$. Для определения критерия рассматривается случайное число m дефектных изделий среди 10. Критическая область: $S = \{m > 1\}$. Ошибка первого рода: вероятность дефектных изделий равна 0.05, но $m > 1$ и принимается гипотеза H_1. $\alpha = P_{H_0}(S) = 1 - q_0^{10} - C_{10}^1 q_0^9 p_0 = 0.086$. Ошибка второго рода: H_0 не верна, т.е. p=0.1, но принимается H_0. $\beta = P_{H_1} = q_1^{10} + C_{10}^1 q_1^9 p_1 = 0.736$.

Задачи домашнего задания.

Задача 9.14. Используя критерий χ^2, проверить гипотезу о том, что выборка, полученная в задаче 5.24, соответствует равномерному распределению на отрезке [0, 1]. Уровень значимости $\alpha = 0.05$.

Задача 15.171. Найти 90%-ные и 99%-ные доверительные интервалы для математического ожидания времени безотказной работы электронной лампы, если $\bar{x} = 500$, $n = 100$, среднеквадратичное отклонение известно и равно 10 часам.

Ответ: (498.35, 501.64), (497.42, 502.58).

Задача 15.197. С автоматической линии, производящей подшипники, было отобрано 400 штук, причем 10 оказалось бракованными. Найти 90%-ный доверительный интервал для вероятности появления бракованного подшипника. Сколько подшипников надо проверить, чтобы с вероятностью 0.9973 можно было утверждать, что вероятность появления бракованного подшипника не отличается от частоты более чем на 5%?

Ответ: (0.012, 0.038), $n \geq 88$.

Задача 15.219. Из продукции автомата, обрабатывающего болты с номинальным значением контролируемого размера $m_0 = 40$ мм, была взята выборка болтов объема $n = 36$. Выборочное среднее контролируемого размера $\bar{x} = 40.2$ мм. Результаты предыдущих измерений дают основание предполагать, что действительные размеры болтов образуют нормально распределенную совокупность с дисперсией $\sigma^2 = 1$ мм^2. Можно ли по результатам проведенного выборочного обследования утверждать, что контролируемый размер в продукции автомата не имеет положительного смещения по отношению к номинальному размеру? Принять $\alpha = 0.01$. Какова критическая область в этом случае?

Ответ: Да. $V_k = \{\bar{X} > 40.39\}$.

3.8 Линейная регрессия и метод наименьших квадратов

1) **Постановка задачи**. Пусть y есть функция от x: $y = \sum\limits_{l=1}^{k-1} \beta_l x^l$. Коэффициенты β_l неизвестны. Проведены n измерений величины y

при заданных $x = x_i$: $y_i = \beta_0 + \beta_1 x_i + \dots + \beta_l x_i^{k-1} + \delta_i$, $i = 1, ..., n$. δ_i — ошибки измерений, которые независимы и распределены нормально с параметрами (a, σ^2). Удобно использовать матричные обозначения:

$$Y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_1^{k-1} & x_2^{k-1} & \cdots & x_n^{k-1} \end{pmatrix},$$

$$\delta = \begin{pmatrix} \delta_1 \\ \cdot \\ \cdot \\ \cdot \\ \delta_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \cdot \\ \cdot \\ \cdot \\ \beta_{k-1} \end{pmatrix}.$$

тогда $Y = X^T\beta + \delta$, $\delta = Y - X^T\beta$. Нужно найти β.

2) **Метод наименьших квадратов**. Рассмотрим

$$Q = \sum_{i=1}^{n} \delta_i^2 = \delta^T\delta = (Y - X^T\beta)^T(Y - X^T\beta) =$$

$$= Y^TY - (Y^TX^T\beta + \beta^TXY) + \beta^TXX^T\beta = Y^TY - 2\beta^TXY + \beta^TXX^T\beta.$$

Требуем минимума этой функции Q: $\frac{\partial Q}{\partial \beta_l} = 0$, $l = 0, 1, ..., k-1$. Обозначим

$$Z = XY = \begin{pmatrix} z_0 \\ \cdot \\ \cdot \\ \cdot \\ z_{k-1} \end{pmatrix},$$

тогда $\beta^TXY = \sum\limits_{i=1}^{k-1} \beta_i z_i$ и $\frac{\partial(\beta^TXY)}{\partial \beta_l} = \frac{\partial}{\partial \beta_l} \sum\limits_{i=1}^{k-1} \beta_i z_i = z_l = (XY)_l$. Обозначим $S = XX^T = (s_{ij})$, $i, j = 0, ..., k-1$. Тогда

$$\frac{\partial}{\partial \beta_l}(\beta^TXX^T\beta) = \frac{\partial}{\partial \beta_l}(\sum_{i,j=0}^{k-1} s_{ij}\beta_i\beta_j) = 2\sum_{j=0}^{k-1} s_{lj}\beta_j = 2(S\beta)_l.$$

В итоге получаем, что система $\frac{\partial Q}{\partial \beta_l} = 0$ эквивалентна системе $-2z_i + 2\sum\limits_{j=0}^{k-1} s_{ij}\beta_j = 0$, которая в матричной записи имеет вид $S\beta = XY$. Решение этой системы обозначим β^*. β^* — оценка параметра β. $\beta^* = S^{-1}(XY) = (S^{-1}X)Y = S^{-1}X(X^T + \delta) = \beta + S^{-1}X\delta$.

3) **Свойства** β^* .

а) $M\beta^* = M(\beta + S^{-1}X\delta) = M\beta + S^{-1}XM\delta = M\beta = \beta$, β^* — несмещенная оценка для β.

б)

$$D\beta^* \equiv cov(\beta^*, \beta^{*T}) = \Big(cov(\beta_i^*, \beta_j^*)\Big) = cov(\beta + S^{-1}X\delta, \beta^T + \delta^T X^T S^{-1}) =$$

$$= cov(S^{-1}X\delta, \delta^T X^T S^{-1}) = S^{-1}X cov(\delta, \delta^T) X^T S^{-1} = S^{-1}X\sigma^2 I_n X^T S^{-1}$$
$$= \sigma^2 S^{-1},$$

так как

$$cov(\delta_i, \delta_j) = \begin{pmatrix} 0, i \neq j \\ \sigma^2, i = j \end{pmatrix}, \text{ и } D\delta = cov(\delta, \delta^T) = \sigma^2 I_n.$$

4) **Невязка**. Определяется невязка e , которая вычисляется следующим образом:

$$e = Y - X^T\beta^* = X^T\beta + \delta - X^T(\beta + S^{-1}X\delta) = (I_n - X^T S^{-1}X)\delta.$$

Обозначим $\alpha = I_n - X^T S^{-1} X$, тогда $e = \alpha\delta$. $\alpha^T = \alpha$, то-есть α — симметрическая матрица, так как $(X^T S^{-1} X)^T = X^T S^{-1} X^{T^T} = X^T S^{-1} X$. Пусть $Q_0 = e^T e = \delta^T \alpha^T \alpha\delta = \delta^T \alpha\delta$, так как $\alpha\alpha = (I_n - X^T S^{-1} X)(I_n - X^T S^{-1} X) = I_n - 2X^T S^{-1} X + X^T S^{-1} X X^T S^{-1} X = I_n - X^T S^{-1} X = \alpha$. Рассмотрим $Q|_{\beta=\beta^*} = (Y^T - \beta^{*T} X)(Y - X^T\beta^*) = e^T e = Q_0$, то-есть Q_0 есть минимум Q. Q_0 называется остаточной суммой квадратов. Раз α – симметрическая матрица, то существует ортогональное преобразование O ($O^T O = I_n$), что имеет место $\alpha = O^T \alpha_{\text{diag}} O$, где матрица α_{diag} — диагональная матрица. Но $\alpha^2 = O^T \alpha_{\text{diag}} O O^T \alpha_{\text{diag}} O = O^T \alpha_{\text{diag}} \alpha_{\text{diag}} O = \alpha = O^T \alpha_{\text{diag}} O$. Поэтому, диагональные элементы матрицы α_{diag} равны 0 или 1. С другой стороны, сумма диагональных элементов матрицы α, обозначаемая след(α), равна: след(α) = след$(I_n - X^T S^{-1} X) = n -$ след$(X^T S^{-1} X) = n -$ след$(X X^T S^{-1}) = n -$ след$(S S^{-1}) = n -$ след$(I_k) = n - k$. В итоге ранг матрицы α равен её следу $n - k$.

Вычислим

$$MQ_0 = M\delta^T \alpha \delta = M \sum_{i,j=1}^{n} \alpha_{ij} \delta_i \delta_j = \sum_{i,j=1}^{n} \alpha_{ij} M \delta_i \delta_j = \sigma^2 \sum_{i=1}^{n} \alpha_{ii} =$$

$$= \sigma^2 \text{след}(\alpha) = \sigma^2 (n - k).$$

Отсюда, если

$$s^2 = \frac{Q_0}{n-k},$$

то $Ms^2 = \sigma^2$ и s^2 есть несмещенная оценка для σ^2. С другой стороны, так как $Q_0 = e^T e = (O\delta)^T \alpha_{\text{diag}} (O\delta)$, то $\frac{Q_0}{\sigma^2} \propto \chi^2_{n-k}$. Рассмотрим

$$D(\beta^* - \beta) = cov(\beta^* - \beta, \beta^{*T} - \beta^T) = cov(S^{-1} X \delta, \delta^T X^T S^{-1}) =$$

$$= S^{-1} X cov(\delta, \delta^T) X^T S^{-1} = \sigma^2 S^{-1} X X^T S^{-1} = \sigma^2 S^{-1},$$

значит $D(\beta^* - \beta)_i = \sigma^2 (S^{-1})_{ii}$. В итоге имеем

$$\frac{(\beta^* - \beta)_i}{\sqrt{\sigma^2 (S^{-1})_{ii}}} \propto N(0,1), \ \frac{Q_0}{\sigma^2 (n-k)} = \frac{s^2}{\sigma^2} \propto \frac{\chi^2_{n-k}}{n-k}$$

и

$$\frac{(\beta^* - \beta)_i}{\sqrt{\sigma^2 (S^{-1})_{ii}}} : \sqrt{\frac{s^2}{\sigma^2}} = \frac{(\beta^* - \beta)_i}{\sqrt{s^2 (S^{-1})_{ii}}} \propto \tau_{n-k}.$$

Последние соотношения — источники для получения интервальных оценок для σ^2 и β_i.

5) **Линейные гипотезы**. Можно предположить, что коэффициенты β_i должны удовлетворять условиям: $t_{l0}\beta_0 + t_{l1}\beta_1 + \dots + t_{l(k-1)}\beta_{k-1} = t_l, \ \ l = 1, ..., m$. В матричной записи мы имеем:

$$T\beta = t, \ T = \begin{pmatrix} t_{10} & \cdots & t_{1(k-1)} \\ \cdot & \cdots & \cdots \\ \cdot & \cdots & \cdots \\ \cdot & \cdots & \cdots \\ t_{m0} & \cdots & t_{m(k-1)} \end{pmatrix}, t = \begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ \cdot \\ t_m \end{pmatrix}.$$

Ищется минимум $Q = (Y^T - \beta^T X)(Y - X^T\beta)$ при условии $T\beta = t$. Оказывается, точка условного минимума $\tilde{\beta}$ связана с точкой β^* абсолютного минимума следующим образом:

$$\tilde{\beta} = \beta^* - S^{-1}T^T D^{-1}(T\beta^* - t),\ D = TS^{-1}T^T.$$

Относительный минимум $Q(\tilde{\beta})$ и абсолютный минимум $Q_0 = Q(\beta^*)$ связаны с помощью соотношения

$$Q(\tilde{\beta}) = Q_0 + Q_T,\ Q_T = (T\beta^* - t)^T D^{-1}(T\beta^* - t).$$

Оказывается $\frac{Q_T}{\sigma^2} \propto \chi^2_m$, поэтому $\frac{Q_T}{m} : \frac{Q_0}{n-k} = \frac{\chi^2_m}{m} : \frac{\chi^2_{n-k}}{n-k} = F(m, n-k)$ и $F(m, n-k)$ называется распределением Фишера.

Пусть гипотеза $T\beta = t$ верна, тогда Q_T близка к нулю, тогда близка к нулю и $F(m, n-k)$. Пусть определяется $F_\alpha(m, n-k)$ из условия $P(F(m, n-k) > F_\alpha(m, n-k)) = \alpha$. . Если α мало, то $(F(m, n-k) < F_\alpha(m, n-k))$ близко к достоверному событию. Поэтому, если вычисленное значение $F(m, n-k)$ удовлетворяет неравенству $F(m, n-k) < F_\alpha(m, n-k))$, , то гипотеза принимается, если не удовлетворяет, то гипотеза отвергается.

3.8.1 11-ое практическое занятие. Регрессионный анализ

Задача 9.16. Для величины $\sigma^{2*} = \frac{1}{n}\sum\limits_{i=1}^{n}(y_i - a^*x_i - b^*)^2$, где

$$a^* = a + \frac{\sum\limits_{i=1}^{n} x_i\delta_i}{\sum\limits_{i=1}^{n} x_i^2} \text{ и } b^* = b + \frac{\sum\limits_{i=1}^{n}\delta_i}{n}$$

— оценки для a и b в задаче для линейной регрессии с $y_i = ax_i + b + \delta_i$, $\sum\limits_{i=1}^{n} x_i = 0$, найти σ^{2*}.

Решение. $\sigma^{2*} = \frac{1}{n}\sum\limits_{i=1}^{n}(y_i - a^*x_i - b^*)^2$.

$$y_i - a^*x_i - b^* = \delta_i - \frac{1}{n}\sum_{k=1}^{n}\delta_k - \frac{x_i}{\sum\limits_{k=1}^{n} x_k^2}\sum_{j=1}^{n} x_j\delta_j,$$

$$(y_i - a^*x_i - b^*)^2 = \delta_i^2 + \frac{1}{n^2}(\sum_{k=1}^{n}\delta_k)^2 + \frac{x_i^2}{(\sum\limits_{k=1}^{n} x_k^2)^2}(\sum_{j=1}^{n} x_j\delta_j)^2 -$$

$$-\frac{2\delta_i}{n}\sum_{k=1}^{n}\delta_k - \frac{2x_i\delta_i}{\sum\limits_{k=1}^{n} x_k^2}\sum_{j=1}^{n} x_j\delta_j + \frac{2}{n}\sum_{l=1}^{n}\delta_l\frac{x_i}{\sum\limits_{k=1}^{n} x_k^2}\sum_{j=1}^{n} x_j\delta_j,$$

$$\sigma^{2*} = \frac{1}{n}\Bigg[\sum_{k=1}^{n}\delta_k^2 + \frac{(\sum\limits_{k=1}^{n}\delta_k)^2}{n} + \frac{\big(\sum\limits_{j=1}^{n} x_j\delta_j\big)^2}{\sum\limits_{k=1}^{n} x_k^2} - \frac{2}{n}\Bigg((\sum_{k=1}^{n}\delta_k)^2 + n\frac{(\sum\limits_{j=1}^{n} x_j\delta_j)^2}{\sum\limits_{k=1}^{n} x_k^2}\Bigg)\Bigg] =$$

$$= \frac{1}{n}\Bigg[\sum_{k=1}^{n}\delta_k^2 - \frac{1}{n}\Bigg(\sum_{k=1}^{n}\delta_k\Bigg)^2 - \frac{\Big(\sum\limits_{k=1}^{n} x_k\delta_k\Big)^2}{(\sum\limits_{k=1}^{n} x_k)^2}\Bigg].$$

$$M\delta_k^2 = \sigma^2.\ M\sigma^{2*} = \frac{1}{n}(n\sigma^2 + \sigma^2 + \sigma^2 - 2\sigma^2 - 2\sigma^2) = \frac{n-2}{n}\sigma^2.$$

Задача 9.18. Пусть y_i — значения функции $f(x) = ax^2 + bx + c$, измеренные в точках $x_i = 0.2 + 0.5(i-1)$, $i = 1, ..., 10$. При $a = 1, b = 2, c = -1$ найти реализацию выборки $y_i = x^2 + 2x - 1 + \delta_i$, $i = 1, 2, ..., 10$, где δ_i независимы и нормально распределены с $M\delta_i = 0$, $D\delta_i = 0.06$. Методом наименьших квадратов получить оценки a^*, b^*, c^* параметров a, b, c. Сравнить эти оценки с известными истинными значениями.

Решение. Если $\sigma^2 = 0.06$, то $\sigma \approx 0.25$. Из таблицы нормально распределенных случайных чисел с $\sigma = 1$ берем 10 чисел, передвигаясь по строкам слева направо. Эти числа делим на четыре, к ним прибавляем $x_i^2 + 2x_i - 1$. В итоге имеем:

x	0.200	0.700	1.200	1.700	2.200	2.700	3.2000	3.7000	4.200	4.700
y	-.446	0.923	3.441	5.210	8.223	11.76	15.569	20.369	25.05	29.94

По методу наименьших квадратов в качестве оценки параметров a, b, c принимаются значения a^*, b^*, c^*, дающие минимум функции

$$\Theta(a,b,c) = \sum_{i=1}^{10}(y_i - (c + bx_i + ax_i^2))^2.$$

Из необходимых условий минимума функции $\Theta(a,b,c)$, т.е. из $\frac{\partial\Theta}{\partial a} = 0$, $\frac{\partial\Theta}{\partial b} = 0$, $\frac{\partial\Theta}{\partial c} = 0$ получим систему уравнений для определения a, b, c :

$$\left\{ \begin{array}{l} \sum\limits_{i=1}^{n}(y_i - (c + bx_i + ax_i^2)) = 0 \\ \sum\limits_{i=1}^{n}(y_i - (c + bx_i + ax_i^2))x_i = 0 \\ \sum\limits_{i=1}^{n}(y_i - (c + bx_i + ax_i^2))x_i^2 = 0 \end{array} \right. ,$$

которую можно переписать в виде:

$$\left\{ \begin{array}{l} cn + b\sum\limits_{i=1}^{n} x_i + a\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} y_i \\ c\sum\limits_{i=1}^{n} x_i + b\sum\limits_{i=1}^{n} x_i^2 + a\sum\limits_{i=1}^{n} x_i^3 = \sum\limits_{i=1}^{n} x_i y_i \\ c\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i^3 + a\sum\limits_{i=1}^{n} x_i^4 = \sum\limits_{i=1}^{n} x_i^2 y_i \end{array} \right. .$$

Решения этих уравнений обозначаются a^*, b^* и c^*. Эту систему уравнений удобно записать в матричных обозначениях. Введем следующие матрицы:

$$\beta = \begin{bmatrix} c \\ b \\ a \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_{10} \end{bmatrix}, X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{10} \\ x_1^2 & x_2^2 & \cdots & x_{10}^2 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ .2 & .7 & \cdots & 4.7 \\ .04 & .49 & \cdots & 22.09 \end{bmatrix}.$$

В этих обозначениях система для a, b, c примет вид: $(XX^T)\beta = XY$. Обозначим

$$S = XX^T = \begin{bmatrix} n & \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 \\ \sum\limits_{i=1}^{n} x_i & \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i^3 \\ \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i^3 & \sum\limits_{i=1}^{n} x_i \end{bmatrix},$$

и разрешая это матричное уравнение относительно β, и подставляя сюда числовые значения матричных элементов, получим:

$$\begin{bmatrix} c^* \\ b^* \\ a^* \end{bmatrix} = \begin{bmatrix} -0.927 \\ 2.148 \\ 0.981 \end{bmatrix}.$$

Задача 15.361. По выборкам наблюдений

X	-3	-2	-1	0	1	2	3
Y	6	0	-1	-1	2	5	12

требуется

а) найти оценки параметров модели $= \beta_0 + \beta_1 x + \beta_2 x^2$;

б) проверить значимость модели;

в) найти оценки дисперсии ошибок наблюдений и ковариационной матрицы;

г) определить доверительные интервалы для параметров и дисперсии ошибок наблюдений при заданном уровне значимости α ($\alpha = 0.05$) . Предполагается, что ошибки наблюдений не коррелированы и имеют нормальное распределение $N(0, \sigma^2)$.

Решение. С учетом случайных флуктуаций переменной Y результаты наблюдений, по предположению, являются реализациями случайных величин

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i,$$

где ϵ_i – случайные ошибки наблюдений.

а) По методу наименьших квадратов минимизируя функцию

$$\Theta(\beta_0,\beta_1,\beta_2)=\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1x_i+\beta_2x_i^2))^2,\quad (n=7)$$

получим уравнения для определения параметров β_0, β_1 и β_2. Получение решения уравнений описано в задаче 9.18. Решение имеет вид: $\beta = S^{-1}XY$, где использованы обозначения

$$\beta=\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix}, Y=\begin{bmatrix}y_1\\ y_2\\ \cdot\\ \cdot\\ \cdot\\ y_n\end{bmatrix}, X=\begin{bmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_n\\ x_1^2 & x_2^2 & \cdots & x_n^2\end{bmatrix}=$$

$$=\begin{bmatrix}1 & 1 & \cdots & 1\\ -3 & -2 & \cdots & 3\\ 9 & 4 & \cdots & 9\end{bmatrix}, S=XX^T.$$

Подставляя числовые значения матричных элементов, обозначая найденные значения параметров β_i^*, получаем $\beta^{*T}=(\beta_0^*,\beta_1^*,\beta_2^*)=$
$=(-1.333, 1.071, 1.119)$.

б) Модель называется статистически значимой, если отвергается гипотеза $H_0:\ \beta_1=0, \beta_2=0$.

Для проверки значимости гипотезы принимается статистика

$$F=\frac{Q_R/(k-1)}{Q_0/(n-k)}=\frac{Q_R}{(k-1)s^2},$$

где $s^2=\frac{Q_0}{n-k}$, $Q_0=\sum\limits_{i=1}^{n}(y_i-(\beta_0^*+\beta_1^*x_i+\beta_2^*x_i^2))^2$ — остаточная сумма квадратов, $Q_R=\beta^{*T}XY-n\bar{y}^2$ – сумма квадратов, обусловленная регрессией.

Если гипотеза H_0 верна, то статистика F имеет распределение Фишера с $k-1$ и $n-k$ степенями свободы. Вычисляя величины, входящие в статистику F, получим $Q_R=137.33$, $s^2=0.381$, $F=$

180.25. Если F больше, чем критическое значение $F_\alpha(k-1, n-k)$ при заданном уровне значимости α, то гипотеза H_0 отвергается. Для $\alpha = 0.05$, $n-k = 7-3 = 4$, $k-1 = 2$ $F_{0.05}(2,4) = 6.94$ и $F > F_{0.05}(2,4)$, что означает, что модель статистически значима.

в) Оценка дисперсии ошибок наблюдений равна s^2 и вычислена в пункте б). Оценка ковариационной матрицы

$$K = s^2 S^{-1} = 0.381 \begin{bmatrix} 0.381 & 0 & -0.048 \\ 0 & 0.036 & 0 \\ -0.048 & 0 & 0.012 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.127 & 0 & -0.018 \\ 0 & 0.014 & 0 \\ -0.018 & 0 & 0.005 \end{bmatrix}.$$

Диагональные элементы этой матрицы дают оценки дисперсии параметров β_0^*, β_1^* и β_2^*.

г) Доверительный интервал для β_i определяется выражением $(\beta_i^* - \tau_\alpha\sqrt{K_{ii}}, \beta_i^* + \tau_\alpha\sqrt{K_{ii}})$, где K_{ii} – диагональные элементы ковариационной матрицы, τ_α – квантиль распределения Стьюдента с $n-k$ степенями свободы и уровнем значимости α. Для $\alpha = 0.05$ $\tau_\alpha(4) = 2.132$ и доверительные интервалы для β_0, β_1 и β_2 соответственно равны $(-2.04, -0.57)$; $(0.82, 1.32)$ и $(0.97, 1.26)$.

Доверительный интервал дисперсии ошибок наблюдения согласно критерию χ^2 с $n-k$ степенями свободы определяется выражением

$$\frac{(n-k)s^2}{\chi^2_\alpha(n-k)} < \sigma^2 < \frac{(n-k)s^2}{\chi^2_{1-\alpha}(n-k)}$$

и при $\chi^2_{0.95} = 9.49$ и $\chi^2_{0.05} = 0.711$ равен $(0.161, 2.134)$.

Задачи домашнего задания.

Задача 9.17. В случае произвольных $a_1, a_2, ..., a_r$ найти MQ_1 и MQ_2, где $Q_1 = \sum\limits_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2$, $Q_2 = \sum\limits_{i=1}^{r}\sum\limits_{j=1}^{n_i}(x_{ij} - \bar{x}.)^2$, $\bar{x}_i = \frac{\sum_{k=1}^{n_i} x_{ik}}{n_i}$, $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{n_r} x_{ij}$, $x_{ij} \sim N(a_i, \sigma^2)$.

Задача 9.19. Заменить в задаче 9.18 реализацию $\delta_1, ..., \delta_{10}$ на реализацию этих величин, соответствующую равномерному распределению с $M\delta_i = 0$, $D\delta_i = 0.06$. Выполнить для новой реализации задание задачи 9.18.

Задача 15.362. Повторить задание задачи 15.361 по отношению к выборке

X	-2	-1	0	1	2
Y	3	0	3	6	9

, $\alpha = 0.05$

4 ЭЛЕМЕНТЫ ТЕОРИИ СЛУЧАЙНЫХ ПРОЦЕССОВ

4.1 Цепи Маркова

Пусть дана совокупность случайных величин ξ_t , где $t = 0, 1, 2, ...,$ (t истолковывается как время). Значения ξ_t предполагаются принадлежащими множеству целых чисел $\aleph = \{1, 2, ..., N\}$ и интерпретируются как номера состояний некоторой системы. Последовательность случайных величин называется цепью Маркова, если для любых

1) $0 \leq t_1 < t_2 < \cdots < t_n < s < t$,

2) $i, j \in \aleph$,

3) $B_1, ..., B_n \subset \aleph$

имеет место

$$P(\xi_t = j | \xi_1 \in B_1, ..., \xi_n \in B_n, \xi_s = i) = P(\xi_t = i | \xi_s = i).$$

Коротко цепь Маркова можно охарактеризовать так: будущее при заданном настоящем от прошлого не зависит. Величины

$$p_{ij}|_t(1) = P(\xi_{t+1} = j | \xi_t = i)$$

представляют вероятности перехода за один шаг в момент t. Если $p_{ij}|_t(1) = \text{const}$, то цепь Маркова называется однородной. Мы будем рассматривать только однородные цепи Маркова. Матрица

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ p_{21} & \cdots & p_{2N} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ p_{N1} & \cdots & p_{NN} \end{pmatrix}, \sum_{j=1}^{N} p_{ij} = 1,$$

называется матрицей вероятностей перехода. Сосчитаем $p_{ij}|_t(2) = P(\xi_{t+2} = j|\xi_t = i)$:

$$P(\xi_{t+2} = j|\xi_t = i) =$$

$$= \sum_{k=1}^{N} P(\xi_{t+1} = k|\xi_t = i)P(\xi_{t+2} = j|\xi_t = i, \xi_{t+1} = k) =$$

$$= \sum_{k=1}^{N} P(\xi_{t+1} = k|\xi_t = i)P(\xi_{t+2} = j|\xi_{t+1} = k) = \sum_{k=1}^{N} p_{ik}p_{kj} = (P^2)_{ij}.$$

Аналогично $p_{ij}|_t(s) = P(\xi_{t+s} = j|\xi_t = i) = (P^s)_{ij}$, поэтому, если ввести вектор начальных вероятностей $q = (q_1, q_2, ..., q_N)$, $q_i = P(\xi_0 = i)$, то

$$P(\xi_t = i) = \sum_{k=1}^{N} P(\xi_0 = k)P(\xi_t = i|\xi_0 = k) = \sum_{k=1}^{N} q_k p_{ki} = (qP^t)_i.$$

Теорема. *Пусть $\exists\ s$, такое, что у P^s все элементы больше нуля, тогда*

$$\lim_{t\to\infty} P^t = \begin{pmatrix} q_1^* & \cdots & q_N^* \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ q_1^* & \cdots & q_N^* \end{pmatrix},$$

где $q^* = (q_1^*, q_2^*, ..., q_N^*)$ — так называемый вектор стационарных состояний. Вектор стационарных состояний определяется следующими условиями: $q^*P = q^*$, $\sum\limits_{i=1}^{N} q_i^* = 1$, $q_i^* \geq 0$. Теорему примем без доказательства.

Следствие: *При любом векторе q начальных вероятностей*

$$\lim_{t\to}(qP^t) = q^*.$$

Следствие интерпретируется так: если цепь Маркова допускает вектор стационарных состояний, то независимо от начального распределения вероятностей система в бесконечности приходит к стационарному состоянию.

Задача о случайном блуждании. Пусть на прямой рассматриваются точки с целочисленными координатами $0, 1, ..., n$. Предполагаем, что материальная точка может двигаться по этим точкам на прямой, и случайная величина ξ_t описывает положение материальной точки на прямой в момент времени t . В следующий момент времени частица может перейти на соседние точки с вероятностями: $P(\xi_{t+1} = \xi_t + 1) = p$, $P(\xi_{t+1} = \xi_t - 1) = q$, $\xi_t \neq 0, n$; если $\xi_t = 0$ или $\xi_t = n$, то $\xi_{t+1} = \xi_t$ (движение частицы между поглощающими стенками). Задачу о случайном блуждании частицы можно интерпретировать и как задачу о разорении игрока. Пусть n — сумма денег у двух игроков, а ξ_t — число денег у первого игрока. Пусть проводятся независимые испытания, в которых событие может произойти с вероятностью , тогда, если выпало событие , то первый игрок от второго получает рубль, если не происходит, то наоборот. Введем $\pi_k(t) = P(\xi_t = n|\xi_0 = k)$. Тогда

$$\begin{gathered}\pi_k(t+1) = P(\xi_{t+1} = n|\xi_0 = k) = \\ = P(\xi_1 = k-1|\xi_0 = k)P(\xi_{t+1} = n|\xi_1 = k-1) + \\ + P(\xi_1 = k+1|\xi_0 = k)P(\xi_{t+1} = n|\xi_0 = k+1) = \\ = \pi_{k-1}(t)\cdot q + \pi_{k+1}(t)\cdot p.\end{gathered}$$

Предполагая, что $t \to \infty$, и обозначая $\lim\limits_{t\to\infty} \pi_k(t) = \pi_k$, получаем

$$\pi_k = q\pi_{k-1} + p\pi_{k+1}.$$

Очевидно $\pi_0 = 0$, $\pi_n = 1$. Будем искать решения π_k в виде $\pi_k = \lambda^k$. Получаем уравнение

$$\lambda^k = q\lambda^{k-1} + p\lambda^{k+1} \text{ или } p\lambda^2 - \lambda + q = 0,$$

откуда

$$\lambda = \frac{1 \pm \sqrt{1-4pq}}{2p} = \frac{1 \pm (p-q)}{2p}.$$

Корни $\lambda_1 = 1,\ \lambda_2 = q/p$, значит

$$\pi_k = C_1\lambda_1^k + C_2\lambda_2^k = C_1 + C_2(\frac{q}{p})^k.$$

Граничные условия приводят к уравнениям

$$C_1 + C_2 = 0,\ C_1 + C_2(\frac{q}{p})^n = 1.$$

Откуда

$$C_2 = -C_1,\ C_1 = \frac{1}{1-(q/p)^2} \text{ и } \pi_k = \frac{1-(q/p)^k}{1-(q/p)^n}.$$

Если $p \to q$, то $\pi_k \to k/n$.

4.1.1 12-ое практическое занятие. Цепи Маркова

Задача 8.1. Матрица вероятностей перехода цепи Маркова имеет вид:

$$P = \begin{bmatrix} .1 & .5 & .4 \\ .6 & .2 & .2 \\ .3 & .4 & .3 \end{bmatrix}$$

(точка перед числом означает нулевую целую часть: .7 =0.7). Распределение по состояниям в момент $t = 0$ определяется вектором $q = (.7, .2, .1)$. Найти: 1) распределение по состояниям в момент $t = 2$; 2) вероятность того, что в моменты $t = 0, 1, 2, 3$ состояниями цепи будут соответственно 1,3,3,2; 3) стационарное распределение.

Решение.

1) $P = \begin{bmatrix} .1 & .5 & .4 \\ .6 & .2 & .2 \\ .3 & .4 & .3 \end{bmatrix}$, $P^2 = \frac{1}{100}\begin{bmatrix} 43 & 31 & 26 \\ 24 & 42 & 34 \\ 36 & 35 & 29 \end{bmatrix}$,

$$(.7, .2, .1) \cdot P^2 = \frac{1}{1000}(385, 336, 279).$$

2) $P(\xi_0=1,\xi_1=3,\xi_2=3,\xi_3=2)=.7\cdot.4\cdot.3\cdot.4=\frac{336}{10000}$.
3)

$$\left\{\begin{array}{l}(q_1,q_2,q_3)P=(q_1,q_2,q_3)\\ q_1+q_2+q_3=1\end{array}\right. \to \left\{\begin{array}{l}q_1+6q_2+3q_3=10q_1\\ 5q_1+2q_2+4q_3=10q_2\\ q_1+q_2+q_3=1\end{array}\right. \to$$

$$\left\{\begin{array}{l}q_1=16/47\\ q_2=17/47\\ q_3=14/47\end{array}\right. .$$

Задача 8.4. Пусть $\xi_t,\ t=1,2,...,$ — независимые одинаково распределенные случайные величины. $P(\xi_t=1)=p,\ P(\xi_t=-1)=1-p$. Являются ли цепью Маркова величины: а) $\eta_t=\xi_t\xi_{t+1}$; б) $\eta_t=\xi_1\xi_2\cdots\xi_t$; в) $\eta_t=\varphi(\xi_t,\xi_{t+1})$, где $\varphi(-1,-1)=1$, $\varphi(-1,1)=2$, $\varphi(1,-1)=3$, $\varphi(1,1)=4$? Для цепей Маркова найти матрицы вероятностей перехода за один шаг.

Решение.

а) Рассмотрим две частные условные вероятности.

$$P(\eta_{t+1}=1|\eta_{t-1}=\eta_t=1)=\frac{P(\eta_{t-1}=\eta_t=\eta_{t+1}=1)}{P(\eta_{t-1}=\eta_t=1)}=$$

$$=\frac{P(\xi_{t-1}=\xi_t=\xi_{t+1}=\xi_{t+2}=\pm1)}{P(\xi_{t-1}=\xi_t=\xi_{t+1}=\pm1)}=\frac{p^4+q^4}{p^2+q^2},$$

$$P(\eta_{t+1}=1|\eta_t=1)=\frac{p^3+q^3}{p^2+q^2}.$$

Из равенства двух условных вероятностей следует, что p=q=0.5. Но если $p=q$, то для любых $\epsilon_1,\epsilon_2,\epsilon_3$, равных ±1, имеем:

$$P(\eta_{t+1}=\epsilon_1|\eta_{t-1}=\epsilon_2,\eta_t=\epsilon_3)=P(\eta_{t+1}=\epsilon_1|\eta_t=\epsilon_3).$$

т.е. η_t образуют цепь Маркова.

$$P(\eta_{t+1}=1|\eta_t=1)=0.5, P(\eta_{t+1}=-1|\eta_t=-1)=\frac{p^2q+pq^2}{2pq}=0.5,$$

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

б)

$$P(\eta_{t+2} = k|\eta_t = i, \eta_{t+1} = j) =$$
$$= P(\eta_t = i, \eta_{t+1} = j, \eta_{t+2} = k)/P(\eta_t = i, \eta_{t+1} = j) =$$
$$P(\eta_t = i, \xi_{t+1} = j/i, \xi_{t+2} = k/j)/P(\eta_t = i, \xi_{t+1} = j/i) =$$
$$= P(\xi_{t+2} = k/j) = P(\eta_{t+2} = k|\eta_{t+1} = j).$$

Цепь Маркова.

в) Так как φ – функция двух переменных, то φ^{-1} – двухкомпонентная функция: $\varphi^{-1} = (\varphi_1^{-1}, \varphi_2^{-1})$:

$$P(\eta_{t+2} = k|\eta_t = i, \eta_{t+1} = j) =$$
$$= P(\eta_t = i, \eta_{t+1} = j, \eta_{t+2} = k)/P(\eta_t = i, \eta_{t+1} = j) =$$
$$= P(\xi_t = \varphi_1^{-1}(i), \xi_{t+1} = \varphi_2^{-1}(i), \xi_{t+1} = \varphi_1^{-1}(j), \xi_{t+2} = \varphi_2^{-1}(j),$$
$$\xi_{t+2} = \varphi_1^{-1}(k), \xi_{t+3} = \varphi_2^{-1}(k))/$$
$$P(\xi_t = \varphi_1^{-1}(i), \xi_{t+1} = \varphi_2^{-1}(i), \xi_{t+1} = \varphi_1^{-1}(j), \xi_{t+2} = \varphi_2^{-1}(j)) =$$
$$= \delta_{\varphi_2^{-1}(j), \varphi_1^{-1}(j)} \delta_{\varphi_2^{-1}(j), \varphi_1^{-1}(k)} P(\xi_{t+3} = \varphi_2^{-1}(k)) = P(\eta_{t+2} = k|\eta_{t+1} = j),$$

где $\delta_{m,n}$ – символ Кронекера.

$$p_{11} = P(\eta_{t+1} = 1/\eta_t = 1) = \delta_{\varphi_2^{-1}(1), \varphi_1^{-1}(1)} P(\xi_{t+1} = \varphi_1^{-1}(1)) = q, \text{ и т.д.}$$

$$P = \begin{bmatrix} q & p & 0 & 0 \\ 0 & 0 & q & p \\ q & p & 0 & 0 \\ 0 & 0 & q & p \end{bmatrix}.$$

Задача 8.6. В N ячейках последовательно независимо друг от друга равновероятно размещают частицы. Пусть $\mu_0(n)$ — число ячеек, оставшихся пустыми после размещения n частиц. Показать, что последовательность $\mu_0(n)$, $n = 0, 1, 2, ...$ является цепью Маркова. Найти вероятности перехода.

Решение. Является, так как случайная величина $\mu_0(n)$ описывает только число пустых ячеек, но не историю заполнения этих ячеек. Пусть $p_{i,j} = P(\mu_0(n+1) = j|\mu_0(n) = i)$.

$p_{i,j} = 0$, если $j < i-1$, так как при добавлении одной частицы число пустых ячеек может уменьшиться только на единицу.

$p_{i,j} = 1/N$, если $j = i-1$. Частица попала в пустую ячейку.

$p_{i,j} = (N-i)/N$, если $j = i$. Частица попала в занятую ячейку.

$p_{i,j} = 0$, если $j > i$. При появлении новой частицы число пустых ячеек не увеличивается .

Задача 8.8. Игральная кость все время перекладывается случайных образом с одной грани равновероятно на любую из четырех соседних граней независимо от предыдущего. К какому пределу стремится при $t \to \infty$ вероятность того, что в момент времени t кость лежит на грани "6", если в момент $t = 0$ она находилась в этом же положении $(t = 0, 1, 2, ...)$?

Решение. Попарно противоположные грани обозначим числами 1, 2; 3,4; 5,6, которые можно считать номерами состояний. Тогда

$$P = \frac{1}{4}\begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix},\ qP = q \to q = \frac{1}{6}(1,1,1,1,1,1).$$

Задача 8.11. Матрица вероятностей перехода $P = (p_{ij})$ цепи Маркова ξ_t определяется формулами $p_{11} = 1-\alpha$, $p_{12} = \alpha$, $p_{21} = \beta$, $p_{22} = 1-\beta$. Доказать, что

$$\begin{bmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{bmatrix} = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^t}{\alpha+\beta}\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}.$$

Найти стационарные вероятности.

Решение. Применяем метод математической индукции:

$$1)\ P(1) = \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{1-\alpha-\beta}{\alpha+\beta}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix},$$

2) $P(t+1) = P(t)P = \left(\frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^t}{\alpha+\beta}\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}\right)\cdot$

$$\cdot\left(\frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{1-\alpha-\beta}{\alpha+\beta}\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}\right) =$$

$$= \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix} + \frac{(1-\alpha-\beta)^{t+1}}{\alpha+\beta}\begin{bmatrix} \alpha & -\alpha \\ -\beta & \beta \end{bmatrix}.$$

Так как $0 < \alpha < 1$, $0 < \beta < 1$, то $-1 < 1-\alpha-\beta < 1$, значит $(1-\alpha-\beta)^t \to 0$ при $t \to \infty$, поэтому

$$P(t) \to \frac{1}{\alpha+\beta}\begin{bmatrix} \beta & \alpha \\ \beta & \alpha \end{bmatrix}.$$

Задачи домашнего задания.

Задача 8.2. Пусть ξ_t, $t = 0, 1, 2, ...,$ — цепь Маркова с матрицей вероятностей перехода $P = (p_{ij})$ и начальным распределением по состояниям $q = (q_1, ..., q_N)$. Вычислить следующие вероятности $P(\xi_s = i)$, $P(\xi_s = i, \xi_{t+s} = j)$.

Ответ: $P(\xi_s = i) = \sum\limits_{i_0,...,i_{s-1}} q_{i_0} p_{i_0 i_1} \cdots p_{i_{s-1} i}$,

$$P(\xi_s = i, \xi_{t+s} = j) = P(\xi_s = i) \sum_{i_1,...,i_{t-1}} p_{i i_1} \cdots p_{i_{t-1} j}.$$

Задача 8.3. (Продолжение задачи 8.2). Доказать, что

$$P(\xi_{t+s} = j | \xi_s = i) = P(\xi_t = j | \xi_0 = i).$$

Задача 8.7. Пусть $\eta_1, \eta_2, ...$ — последовательность независимых одинаково распределенных случайных величин и $f(x, y)$ — функция, принимающая значения в множестве $\{1, ..., N\}$; $x = 1, 2, ..., N$, множество значений y совпадает с множеством значений η_1. Является ли последовательность случайных величин

$$\xi_0 (P(\xi_0 = k) = p_k^{(0)}),\ k = 0, 1, 2, ..., N,$$

$$\xi_{t+1} = f(\xi_t, \eta_{t+1}),\ t = 0, 1, 2, ...,$$

цепью Маркова.

Ответ: является.

Задача 8.9. По двум урнам разложено N черных и N белых шаров так, что каждая урна содержит N шаров. Число черных шаров в первой урне определяет состояние системы. В каждый момент времени выбирают случайно по одному шару из урн и выбранные шары меняют местами. Найти вероятности перехода и показать, что стационарные вероятности $q_k^* = C_N^k C_N^{N-k}/C_{2N}^N$, $k = 1, ..., N$.

Задача 8.12. Для цепи Маркова, определенной в задаче 8.11, обозначим τ_t число попаданий в состояние 1 за время t. Доказать, что

$$\lim_{t\to\infty} P\left(\left|\frac{\tau_t}{t} - \frac{\beta}{\beta+\alpha}\right| < \epsilon\right) = 1 \text{ при любом } \epsilon > 0.$$

4.2 Случайные процессы

Определение. Процесс с независимыми приращениями. Случайным процессом называется совокупность случайных величин ξ_t, где $t = 0, 1, 2, ...$ (случайный процесс с дискретным временем) или $t \in [0, T]$ (случайный процесс с непрерывным временем). При фиксированном ω $\xi_t(\omega)$ как функция от t называется траекторией процесса. Случайный процесс ξ_t называется процессом с независимыми приращениями, если для $\forall t_1 < t_2 < \cdots < t_n$ случайные величины $\xi_{t_1}, \xi_{t_2} - \xi_{t_1}, ..., \xi_{t_n} - \xi_{t_{n-1}}$ независимы.

4.2.1 Пуассоновский и винеровский процессы

Случайный процесс ξ_t с непрерывным временем называется пуассоновским, если выполнены следующие условия:

1) ξ_t — процесс с независимыми приращениями;

2) для $\forall s, t_1, t_2$ случайные величины $\xi_{t_2} - \xi_{t_1}, \xi_{t_2+s} - \xi_{t_1+s}$ одинаково распределены (однородность);

3) $\xi_0 = 0$;

4) при $h \to 0$ имеет место

$$\begin{cases} P(\xi_h = 0) = 1 - \lambda h + o(h), \\ P(\xi_h = 1) = \lambda h + o(h), \\ P(\xi_h \geq 2) = o(h). \end{cases}$$

При работе со случайными величинами ξ с целочисленными неотрицательными значениями полезна так называемая производящая функция $\varphi_\xi(x) = Mx^\xi$. Для распределения Пуассона с $P(\xi = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ имеем

$$\varphi_\xi(x) = \sum_{k=0}^{\infty} x^k \frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(x\lambda)^k}{k!}e^{-\lambda} = e^{\lambda(x-1)}.$$

Пусть $\varphi_t(x) = Mx^{\xi_t}$. Очевидно, что $\varphi_0 = 1$. Далее имеем

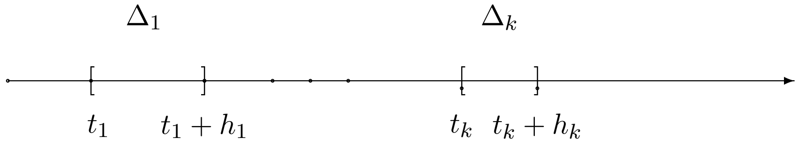
$$\varphi_{t+h}(x) = Mx^{\xi_{t+h}} = Mx^{\xi_t + (\xi_{t+h} - \xi_t)} = Mx^{\xi_t}Mx^{\xi_{t+h}-\xi_t} = \varphi_t Mx^{\xi_h} =$$
$$= \varphi_t(x^0(1 - \lambda h + o(h)) + x^1(\lambda h + o(h)) + \cdots) =$$
$$= \varphi_t(1 + \lambda h(x-1) + o(h)).$$

Отсюда
$$\frac{\varphi_{t+h} - \varphi_t}{h} = \varphi_t\lambda(x-1) + \frac{o(h)}{h},$$
поэтому $\frac{\partial}{\partial t}(\varphi_t) = \lambda(x-1)\varphi_t$. Решая это уравнение, находим $\varphi_t(x) = Ce^{\lambda(x-1)t}$, где $\ = 1$, так как $\varphi_0 = 1$. Таким образом $P(\xi_t = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$, то есть ξ_t подчинена распределению Пуассона с параметром λt.

Рассмотрим случайную величину $\xi_{t+h} - \xi_t$, которая ввиду однородности процесса распределена как $\xi_h - \xi_0 = \xi_h$, поэтому ξ_{t+h} распределена как $\xi_t + \xi_h$, у ξ_h наиболее вероятное значение 0, следующее по вероятности значение 1. Значит, $\xi_t(\omega)$ с ростом t может увеличиться, увеличиться лишь на 1, то есть терпеть скачок. Поэтому $\xi_t(\omega)$ есть число скачков траектории за время $[0, t]$. Введем случайную величину $\Theta_k = \Theta_k(\omega)$ — момент k — го скачка траектории процесса. Тогда $\tau_k = \Theta_k - \Theta_{k-1}$ — промежуток времени между двумя последовательными скачками.

Теорема. *Случайные величины $\tau_1, \tau_2, ..., \tau_n$ при любых n независимы, каждая из них имеет показательное распределение с параметром λ.*

Доказательство. Возьмем последовательность моментов $0 = t_0 < t_1 < \cdots < t_n$ и положим $\Delta_k = [t_k, t_k + h_k]$, $h_k > 0$, $k = 1, ..., n$, где h_k такие, что Δ_k не пересекаются.



Тогда можно провести следующие вычисления

$$P(\Theta_1 \in \Delta_1, \Theta_2 \in \Delta_2, ..., \Theta_n \in \Delta_n) =$$

$$= P\left(\bigcap_{k=1}^{n}\left\{(\xi_{t_k} - \xi_{t_{k-1}} = 0) \cap (\xi_{t_k+h_k} - \xi_{t_k} = 1)\right\}\right) = \prod_{k=1}^{n} e^{-\lambda(t_k - t_{k-1})}.$$

$$\cdot \lambda h_k e^{-\lambda h_k} = h_1 \cdots h_n \lambda^n e^{-(t_1 - t_0 + t_2 - t_1 + \cdots + t_n - t_{n-1} + h_1 + \cdots + h_n)} =$$

$$= \lambda^n h_1 \cdots h_n e^{-\lambda t_n} e^{-\lambda(h_1 + \cdots + h_n)} = \int\limits_{\Delta_1 \times \cdots \times \Delta_n} \cdots \int p_{\Theta_1, ..., \Theta_n}(t_1, ..., t_n) dt_1 \cdots dt_n,$$

откуда следует, что $p_{\Theta_1, ..., \Theta_n}(t_1, ..., t_n) = \lambda^n e^{-\lambda t_n}$. Теперь найдем распределение случайных величин $\tau_k = \Theta_k - \Theta_{k-1}$. Для этого вычислим их совместную функцию распределения:

$$F_{\tau_1, \tau_2, ..., \tau_n}(s_1, s_2, ..., s_n) = P(\tau_1 < s_1, \tau_2 < s_2, ..., \tau_n < s_n) =$$

$$= P(\Theta_1 < s_1, \Theta_2 - \Theta_1 < s_2, ..., \Theta_n - \Theta_{n-1} < s_n) =$$

$$= \int\limits_{-\infty}^{s_1} du_1 \int\limits_{-\infty}^{s_2 + u_1} du_2 \cdots \int\limits_{-\infty}^{s_n + u_{n-1}} du_n \lambda^n e^{-\lambda u_n},$$

откуда, для совместной плотности вероятности получаем

$$p_{\tau_1, \tau_2, ..., \tau_n}(s_1, s_2, ..., s_n) = \frac{\partial^n}{\partial s_1 \cdots \partial s_n} F_{\tau_1, \tau_2, ..., \tau_n}(s_1, s_2, ..., s_n) =$$

$$= \frac{\partial^{n-1}}{\partial s_2 \cdots \partial s_n} \int\limits_{-\infty}^{s_2+s_1} du_2 \cdots \int\limits_{-\infty}^{s_n+u_{n-1}} du_n \lambda^n e^{-\lambda u_n} =$$

$$= \lambda^n e^{-\lambda(s_1+\cdots+s_n)} = \prod_{k=1}^{n} p_{\tau_k}(s_k),$$

где $p_{\tau_k}(s_k) = \lambda e^{-\lambda s_k}$. Теорема доказана.

Интерпретация пуассоновского процесса с помощью модели радиоактивного распада. Как известно, атом радия может самопроизвольно распасться на атом радона и альфа- частицу. Пусть имеется бесконечное число атомов радия. В начальный момент система находится в состоянии $\xi_0 = 0$, в котором нет альфа-частиц. В этом состоянии система находится случайное время τ_1. Затем происходит один распад, система переходит в состояние $\xi_{\Theta_1} = 1$ $(\Theta_1 = \tau_1)$, в котором находится случайное время τ_2, затем переходит в состояние $\xi_{\Theta_2} = 2$ $(\Theta_2 = \tau_1 + \tau_2)$ и т.д. Если T_0 – период полураспада, т.е. $e^{-\lambda T_0} = \frac{1}{2}$, то $\lambda = \ln 2/T_0$.

Винеровский процесс. По определению, ξ_t есть винеровский процесс, если выполнены следующие условия:

1) ξ_t есть случайный процесс с независимыми приращениями;
2) ξ_t есть однородный случайный процесс;
3) $\xi_0 = 0$;
4) при $h \to 0$ имеет место

$$\begin{cases} M\xi_h = ah + o(h), \\ M\xi_h^2 = bh + o(h), \\ M\xi_h^3 = o(h). \end{cases}$$

Так же, как в случае пуассоновского процесса, можно показать, но уже применяя аппарат характеристических функций, что ξ_t подчиняется нормальному распределению с параметрами (at, bt).

4.2.2 Ветвящийся процесс с дискретным временем

Пусть дано множество независимых, одинаково распределенных случайных величин $\xi_k(t)$ с целочисленными значениями, для которых

$k = 1, 2, ..., t = 0, 1, 2, ...$ Случайная величина $\xi_k(t)$ интерпретируется как случайное число потомков частицы с номером k, существовавшей в момент t. Пусть в начальный момент имеется одна частица: $\mu_0 = 1$, тогда число частиц в следующий момент времени равно: $\mu_1 = \xi_t(0)$. Для последующих моментов имеем

$$\mu_2 = \begin{cases} 0, \ \mu_1 = 0, \\ \xi_1(1) + \xi_2(1) + \cdots + \xi_{\mu_1}(1), \ \mu_1 \neq 0 \end{cases}$$

и

$$\mu_{t+1} = \begin{cases} 0, \ \mu_t = 0, \\ \xi_1(t) + \xi_2(t) + \cdots + \xi_{\mu_t}(t), \ \mu_t \neq 0 \end{cases}.$$

Вычислим $A_t = M\mu_t$ − среднее число частиц в момент t. Для этого рассмотрим производящие функции $\Phi_t = Mx^{\mu_t}$ и $\varphi(x) = Mx^{\xi_k(t)} = \sum\limits_{m=0}^{\infty} x^m p_m$, где $p_m = P(\xi_k(t) = m)$. Положим $a = \varphi'(x)|_{x=1} = \sum m p_m = M\xi_k(t)$. Очевидно, что $\Phi_1 = M\xi^{\mu_1} = Mx^{\xi_1(0)} = \varphi(x)$, далее рассматриваем

$$\Phi_{t+1} = Mx^{\mu_{t+1}} = \sum_{m=0}^{\infty} x^m P(\mu_{t+1} = m) =$$

$$= \sum_{m=0}^{\infty} x^m \sum_{k=0}^{\infty} P(\mu_t = k) P(\xi_1 + \cdots + \xi_k = m) =$$

$$= \sum_{k=0}^{\infty} P(\mu_t = k) \sum_{m=0}^{\infty} x^m P(\xi_1 + \cdots + \xi_k = m) = \sum_{k=0}^{\infty} P(\mu_t = k) \cdot$$

$$\cdot Mx^{\xi_1 + \cdots + \xi_k} == \sum_{k=0}^{\infty} P(\mu_t = k) \varphi(x)^k = \Phi_t(\varphi(x)),$$

Найденную связь между Φ_{t+1}, Φ_t и φ используем для вычисления $A_t = M\mu_t = (\Phi_t)'_x|_{x=1}$ следующим образом:

$$A_{t+1} = (\Phi_{t+1})'_x|_{x=1}(\Phi_t(\varphi(x)))'_x|_{x=1} = \frac{d\Phi_t(\varphi)}{d\varphi} \varphi'_x|_{x=1} = A_t a,$$

откуда $A_t = a^t$. В пределе $t \to \infty$ находим:

$a < 1, \ A_t \to 0$ — докритический процесс;

$a = 1, \ A_t = 1$ — критический процесс;

$a > 1, \ A_t \to \infty$ — надкритический процесс.

4.2.3 Марковский процесс. Процесс гибели и размножения

Случайный процесс ξ_t называется марковским, если для $\forall\ n,\ t_1 < t_2 < \cdots < t_n < t_{n+1}$, для любых событий $\xi_{t_1} \in B_1, ..., \xi_{t_{n-1}} \in B_{n-1}$, $\xi_{t_n} = x$ выполняется равенство

$$P(\xi_{t_{n+1}} \in B_{n+1} | \xi_{t_1} \in B_1, ..., \xi_{t_{n-1}} \in B_{n-1}, \xi_{t_n} = x) =$$

$$= P(\xi_{t_{n+1}} \in B_{n+1} | \xi_{t_n} = x).$$

То есть состояние системы в будущем при заданном настоящем не зависит от прошлого. Марковский процесс ξ_t называется процессом гибели и размножения, если при $h \to 0$ имеет место:

$$\begin{array}{l} P(\xi_{t+h} = k+1 | \xi_t = k) = \lambda_k h + o(h), \\ P(\xi_{t+h} = k-1 | \xi_t = k) = \mu_k h + o(h), \\ P(\xi_{t+h} = k | \xi_t = k) = 1 - (\lambda_k + \mu_k) h + o(h), \end{array}$$

где относительно постоянных предполагается $\lambda_0, \lambda_k, \mu_k > 0$ $(k > 0)$, $\mu_k = 0$ $(k \leq 0)$. Введем $P_k(t) = P(\xi_t = k)$. Для $k > 0$ имеем

$$P_k(t+h) = P(\xi_{t+k} = k) = \sum_{l=0}^{\infty} P(\xi_t = l) P(\xi_{t+h} = k | \xi_t = l) =$$

$$P_{k+1}(t)\mu_{k+1}h + P_k(t)(1 - (\lambda_k + \mu_k)h) + P_{k-1}(t)\lambda_{k-1}h + o(h),$$

откуда получаем дифференциальные уравнения:

$$P_k' = \lambda_{k-1} P_{k-1} - (\lambda_k + \mu_k) P_k + \mu_{k+1} P_{k+1},\ k > 0 \ \text{ и } P_0' = -\lambda P_0 + \mu_1 P_1.$$

Если все $\mu_k = 0$, то процесс называется процессом чистого размножения. В этом случае система дифференциальных уравнений имеет вид:

$$\begin{array}{l} P_k' = \lambda_{k-1} P_{k-1} - \lambda_k P k,\ k > 0, \\ P_0' = -\lambda_0 P_0. \end{array}$$

Если все λ в процессе чистого размножения одинаковы, то этот процесс есть пуассоновский процесс.

4.2.4 Система массового обслуживания с потерями

Имеется n обслуживающих устройств. Отдельное устройство обрабатывает поступившее требование за случайное время, подчиняющееся показательному распределению с параметром μ ($\mu > 0$), это время не зависит от работы других устройств, от поступающих новых требований. Пусть требования, поступающие на пункт обслуживающих устройств, образуют пуассоновский процесс с параметром λ. Если все устройства заняты, то вновь поступающее требование теряется. Пусть ξ_t — число требований в системе в момент t. Возможные значения $0, 1, ..., n$. Рассмотрим вероятность $P(\xi_{t+h} = k|\xi_t = k)$, которая в простейшей интерпретации означает вероятность того, что ни одно устройство не закончило обработку требований за время h и не поступило ни одно новое требование. Поэтому

$$P(\xi_{t+h} = k|\xi_t = k) = \left(1 - \int\limits_0^h \mu e^{-\mu x} dx\right)^k \left(1 - \int\limits_0^h \lambda e^{-\lambda x} dx\right) =$$

$$= (1 - \mu h + o(x))^k (1 - \lambda h + o(x)) = 1 - (k\mu + \lambda)h + o(x).$$

Событие, состоящее в том, что поступило одно новое требование и ни одно обслуживающее устройство не закончило работу, будет характеризоваться вероятностью перехода

$$P(\xi_{t+h} = k + 1|/\xi_t = k) = \left(1 - \int\limits_0^h \mu e^{-\mu x} dx\right)^k (\lambda h + o(x)) = \lambda h + o(x).$$

Наконец, если одно устройство выполнило требование и поступления новых требований не было, то нужно вводить вероятность перехода

$$P(\xi_{t+h} = k - 1|\xi_t = k) = C_k^1 \mu h (1 - \mu h)^{k-1} \left(1 - \int\limits_0^h \lambda e^{-\lambda x} dx\right) =$$

$$= k\mu h + o(h).$$

Снова введем обозначение: $P_k(t) = P(\xi_t = k)$, $k = 0, 1, ..., n$, тогда легко получить следующие дифференциальные уравнения:

$$\begin{aligned}
&P_0' = -\lambda P_0 + \mu P_1,\\
&P_k' = \lambda P_{k-1} - (\lambda + k\mu)P_k + (k+1)\mu P_{k+1},\ 1 \leq k \leq n-1,\\
&P_n' = \lambda P_{k-1} - n\mu P_n.
\end{aligned}$$

Стационарное решение. Пусть $P_k' = 0$. Кроме того, имеем $\sum\limits_{k=0}^{n} P_k = 1$. Из первого уравнения предыдущей системы находим

$$P_1 = \frac{\lambda}{\mu} P_0 = \Theta P_0,$$

из второго уравнения $2\mu P_2 - (\lambda + \mu)P_1 + \lambda_0 = 0$:

$$P_2 = \frac{(1 + \Theta)\Theta - \Theta}{2} P_0 = \frac{\Theta^2}{2!} P_0.$$

И т.д.

$$P_k = \frac{\Theta^k}{k!} P_0.$$

Сумма всех вероятностей должна быть равна единице, поэтому

$$P_0 = \left(\sum_{k=0}^{n} \frac{\Theta^k}{k!} \right)^{-1}.$$

P_n — вероятность потери требования.

4.2.5 Ветвящийся процесс с непрерывным временем

Пусть каждая частица независимо от своего прошлого и от поведения других частиц за время h: погибает с вероятностью $\mu h + o(h)$, делится на две частицы с вероятностью $\lambda h + o(h)$, не изменяется с вероятностью $1 - (\lambda + \mu)h + o(h)$. Пусть ξ_t — число частиц в момент t. Тогда имеем следующие вероятности перехода:

$$\begin{aligned}
&P(\xi_{t+h} = k + 1 | \xi_t = k) = \lambda k h + o(h),\\
&P(\xi_{t+h} = k - 1 | \xi_t = k) = \mu k h + o(h),\\
&P(\xi_{t+h} = k | \xi_t = k) = 1 - (\lambda + \mu) k h + o(h).
\end{aligned}$$

Соответствующие дифференциальные уравнения для $P_k(t) = P(\xi_t = k)$ имеют вид:

$$\begin{aligned}
&P_0' = \mu P_1,\\
&P_1' = -(\lambda+\mu)P_1 + 2\mu P_2,\\
&P_2 = \lambda P_1 - 2(\lambda+\mu)P_2 + 3\mu P_3,\\
&\dots\dots\dots\dots\dots\dots
\end{aligned}$$

4.2.6 13-ое и 14-ое практические занятия. Элементы теории случайных процессов

Задача 10.1. Найти $cov(\xi_t, \xi_{t+s})$, если: а) ξ_t — пуассоновский процесс с параметром λ; б) ξ_t — винеровский процесс с $a = 0$.

Решение. ξ_t, $\xi_{t+s} - \xi_t$ независимы, поэтому

$$cov(\xi_t, \xi_{t+s}) = cov(\xi_t, \xi_{t+s} - \xi_t + \xi_t) = cov(\xi_t, \xi_t) =$$

$$= D\xi_t = \begin{cases} \lambda t \text{ в случае а)}, \\ bt \text{ в случае б)}. \end{cases}$$

Задача 10.3. Пусть ν, η_1, $\eta_2, \dots$ — независимые случайные величины: $P(\nu = k) = (\lambda T)^k e^{-\lambda T}/k!$; величины $\eta_l (1 = 1, 2, \dots)$ равномерно распределены на отрезке [0,T]. Обозначим ξ_t число величин η_l, удовлетворяющих неравенству $\eta_l < t,\ l = 1, 2, \dots, \nu$, если $\nu > 0$, и $\xi_t = 0$ при $\nu = 0$. Найти вероятность

$$P(\xi_{t_1} = k_1, \xi_{t_2} - \xi_{t_1} = k_2 - k_1, \xi_T - \xi_{t_2} = k_3 - k_2).$$

Решение. ξ_t − функция случайных величин ν, η_1, $\eta_2, \dots$.

$$P(\xi_{t_1} = k_1, \xi_{t_2} - \xi_{t_1} = k_2 - k_1, \xi_T - \xi_{t_2} = k_3 - k_2) =$$

$$= P(\xi_{t_1} = k_1, \xi_{t_2} = k_2, \xi_T = k_3) = \sum_{l=k_3}^{\infty} P(\nu = l) \cdot P(\xi_{t_1} = k_1 | \nu = l) \cdot$$

$$\cdot P(\xi_{t_2} = k_2 | \nu = l, \xi_{t_1} = k_1) \cdot P(\xi_T = k_3 | \nu = l, \xi_{t_1} = k_1, \xi_{t_2} = k_2) =$$

$$\sum_{l=k_3}^{\infty} \frac{(\lambda T)^l}{l!} e^{-\lambda T} \cdot C_l^{k_1} \left(\frac{t_1}{T}\right)^{k_1} \left(\frac{T-t_1}{T}\right)^{l-k_1} \cdot C_{l-k_1}^{k_2-k_1} \left(\frac{t_2-t_1}{T-t_1}\right)^{k_2-k_1} \cdot$$

$$\left(\frac{T-t_1-(t_2-t_1)}{T-t_1}\right)^{1-k_1-(k_2-k_1)} C_{l-k_2}^{k_3-k_2}\left(\frac{T-t_2}{T-t_2}\right)^{k_3-k_2} 0^{(l-k_2)-(k_3-k_2)} =$$

$$= \frac{(\lambda T)^{k_3}}{k_3!}e^{-\lambda T}\cdot C_l^{k_1}\left(\frac{t_1}{T}\right)^{k_1}\left(\frac{T-t_1}{T}\right)^{k_3-k_1}\cdot C_{k_3-k_1}^{k_2-k_1}\left(\frac{t_2-t_1}{T-t_1}\right)^{k_2-k_1}\cdot$$

$$\cdot\left(\frac{T-t_2}{T-t_1}\right)^{k_3-k_2} = \lambda^{k_3}e^{-\lambda T}\frac{t_1^{k_1}(t_2-t_1)^{k_2-k_1}(T-t_2)^{k_3-k_2}}{k_1!(k_2-k_1)!(k_3-k_2)!}.$$

Задача 10.5. Пусть μ_t – ветвящийся процесс с непрерывным временем. Найти: а) дифференциальное уравнение для $F(t,x) = M(x^{\mu_t}|\mu_0 = 1)$; б) явное выражение для $F(t,x)$; в) $A(t) = M(\mu_t|\mu_0 = 1)$.

Решение. а) μ_t – число частиц в момент t. Если в начальный момент было k частиц, то каждая i —ая частица ($i = 1,..,k$) в момент t даст случайное число потоков μ_t^i, распределенное как $(\mu_t|\mu_0 = 1)$, поэтому

$$M(x^{\sum\limits_{i=1}^{k}\mu_t^i}|\mu_0^1 = \cdots = \mu_0^k = 1) = \prod_{i=1}^{k} M(x^{\mu_t^i}|\mu_0^i = 1) = F(t,x)^k.$$

$$F(t+h,x) = M(x^{\mu_{t+h}}|\mu_0 = 1) =$$

$$= \sum_{k=0}^{\infty} P(\mu_h = k|\mu_0 = 1)M(x^{\mu_{t+h}}|\mu_0 = 1, \mu_h = k) =$$

$$= \sum_{k=0}^{\infty} P(\mu_h = k|\mu_0 = 1)M(x^{\mu_{t+h}}|\mu_h = k) =$$

(свойство марковости)

$$= \sum_{k=0}^{\infty} P(\mu_h = k|\mu_0 = 1)M(x^{\mu_t}|\mu_0 = k) = \sum_{k=0}^{\infty} P(\mu_h = k|\mu_0 = 1)\cdot$$

(свойство однородности)

$$\cdot F(t,x)^k = \mu h + (1-(\lambda+\mu)h)F + \lambda h F^2 + \cdots .$$

$$\frac{dF}{dt} = \lim_{h\to 0}\frac{F(t+h,x)-F(t,x)}{h} = \mu - (\lambda+\mu)F + \lambda F^2.$$

б) Решение последнего дифференциального уравнения с начальным условием $F(0,x) = x$ находим в виде:

$$F = \begin{cases} 1 + \frac{(\mu-\lambda)(x-1)e^{(\lambda-\mu)t}}{(\mu-\lambda)+(x-1)(\lambda e^{(\lambda-\mu)t}-1)}, \ \mu \neq \lambda \\ 1 + \frac{x-1}{1+\lambda(1-x)t}, \ \mu = \lambda \end{cases}.$$

$$F(t,x) = M(x^{\mu_t}|\mu_0 = 1),$$

$$F'_x\Big|_{x=1} = M(\mu_t x^{\mu_t}|\mu_0 = 1)\Big|_{x=1} = M(\mu_t|\mu_0 = 1) = e^{(\lambda-\mu)t}.$$

Задача 10.7. Пусть ξ_t — процесс гибели и размножения. Положим $m_k = M(\tau|\xi_0 = k)$, $0 \leq k \leq n$, где τ — время до первого достижения состояния n. Найти: 1) математическое ожидание времени пребывания в состоянии k до первого выхода из него; 2) вероятность того, что первым переходом из k будет переход в $k+1$ $(k-1)$. Составить систему уравнений для m_k.

Решение. 1) Пусть ζ_k означает время до первого выхода из состояния k. $(\zeta_k > x)$ – событие, состоящее в том, что процесс находится в состоянии k в течение времени $(0,x)$. $P(\zeta_k > x + h|\zeta_k > \ ) = 1 - h(\mu_k + \lambda_k) + \cdots$. Обозначим $G(x) = P(\zeta_k > x)$.

$$G(x+h) = (\zeta_k > x+h) = P(\zeta_k > x) \cdot P(\zeta_k > x+h|\zeta_k > x) =$$

$$= G(x) \cdot (1 - h(\mu_k + \lambda_k) + \cdots). \ G' = -(\mu_k + \lambda_k) \cdot G, \ G - e^{-(\mu_k+\lambda_k)x}.$$

$$F_{\zeta_k}(x) = P(\zeta_k < x) = 1 - e^{-(\mu_k+\lambda_k)x}, p_{\zeta_k} = (\mu_k + \lambda_k)e^{-(\mu_k+\lambda_k)x},$$

$$M\zeta_k = \int\limits_0^\infty x p_{\zeta_k}(x)dx = \frac{1}{\mu_k + \lambda_k}.$$

2) Вероятность нахождения в состоянии k в течение времени $(0,x)$ и затем перехода в состояние $k+1$ из состояния k в промежутке $(x, x+dx)$ равна

$$\lambda_k e^{-(\mu_k+\lambda_k)x}dx.$$

Вероятность перехода в состояние $k+1$ из состояния k за бесконечное время равна:

$$\int\limits_0^\infty \lambda_k e^{-(\mu_k+\lambda_k)x}dx = \frac{\lambda_k}{\mu_k + \lambda_k}.$$

Аналогично, $\frac{\mu_k}{\mu_k+\lambda_k}$ — время перехода из k в $k-1$.

3) Пусть τ —время, за которое система из состояния k переходит в состояние n, τ' — время до первого выхода из состояния k, τ'' — время прихода впервые в состояние n после выхода из состояния k.

$$m_k = M\tau = M\tau' + M\tau'' =$$

$$= M\tau' + P(k \to k-1)M(\tau''|k \to k-1)) + P(k \to k+1)M(\tau''|k \to k+1) =$$

$$= \frac{1}{\mu_k+\lambda_k} + \frac{\mu_k}{\mu_k+\lambda_k} m_{k-1} + \frac{\lambda_k}{\mu_k+\lambda_k} m_{k+1}.$$

Задача 10.9. Пусть ξ_t – цепь Маркова с непрерывным временем с двумя состояниями (1 и 2) и при $h \to 0$

$$P(\xi_h = 2|\xi_0 = 1) = \alpha h + o(h),\ P(\xi_h = 1|\xi_0 = 2) = \beta h + o(h).$$

Пусть $\tau_k(t)$, $k = 1, 2$, — суммарная длительность пребывания в состоянии k за время t цепи Маркова. Составить дифференциальное уравнение для

$$f_k(t, s) = M(e^{is(\tau_1(t)-\tau_2(t))}|\xi_0 = k).$$

Решение.

$$f_1(t+h) = M(e^{is(\tau_1(t+h)-\tau_2(t+h))}|\xi_0 = 1) =$$

$$= M(e^{is(\tau_1(h)-\tau_2(h)+\tau_1(t+h)-\tau_1(h)-(\tau_2(t+h)-\tau_2(h)))}|\xi_0 = 1) =$$

$$= \sum_k P(\xi_h = k|\xi_0 = 1)$$

$$\cdot M(e^{is(\tau_1(h)-\tau_2(h))} \cdot e^{is(\tau_1(t+h)-\tau_1(h)-(\tau_2(t+h)-\tau_2(h))}|\xi_0 = 1, \xi_h = k) =$$

$$\sum_k P(\xi_h = k|\xi_0 = 1) \cdot M(e^{is(\tau_1(h)-\tau_2(h))}|\xi_0 = 1, \xi_h = k) \cdot f_k(t, s) =$$

$$(1 - \alpha h + \cdots)e^{ish} f_1(t, s) + (\alpha h + \cdots)(1 + \cdots) f_2(t, s).$$

$$\frac{\partial}{\partial t} f_1(t, s) = (is - \alpha) f_1(t, s) + \alpha f_2(t, s).$$

Аналогичным образом находим

$$\frac{\partial}{\partial t} f_2(t,s) = \beta f_1(t,s) - (\beta + is) f_2(t,s).$$

Задача 10.11. Движение точки по прямой управляется цепью Маркова ξ_t, определенной в задаче 9. Если $\xi_t = 1$, то точка движется вправо со скоростью v_1, а если $\xi_t = 2$, то влево со скоростью v_2. Пусть η_t — координата точки в момент t. Найти $\lim\limits_{t\to\infty} M_k(t,x)$, где

$$M_k(t,x) = M(\eta_t | \eta_0 = x, \xi_0 = k),\ k = 1, 2.$$

Решение. Положим $m_1(t) = M(\eta_t | \xi_0 = 1)$, $m_2 = M(\eta_t | \xi_0 = 2)$.

$$m_1(t+h) = M(\eta_{t+h} | \xi_0 = 1) = M(\eta_{t+h} - \eta_h + \eta_h | \xi_0 = 1) =$$

$$= m_1(h) + M(\eta_{t+h} - \eta_h | \xi_0 = 1) = m_1(h) + \sum_{k=1}^{2} P(\xi_h = k | \xi_0 = 1) \cdot$$

$$\cdot M(\eta_{t+h} - \eta_h | \xi_h = k) = v_1 h + (1 - \alpha h) m_1(t) + \alpha h m_2(t).$$

$$m_1' = v_1 + \alpha(m_2 - m_1),\ m_1(0) = 0,$$

$$m_2' = -v_2 - \beta(m_2 - m_1),\ m_2(0) = 0.$$

$$\lim_{t\to\infty} m_k(t)/t = (\beta v_1 - \alpha v_2)/(\alpha + \beta).$$

Задача 10.13. Движение точки по прямой управляется цепью Маркова, определенной в задаче 9. Если $\xi_0 = 1$, то точка движется вправо со скоростью v, а при $\xi_t = 2$ точка движется влево с постоянным ускорением a (при начале движения с ускорением начальная скорость считается равной 0). Пусть η_t – координата точки в момент t. Является ли процесс η_t марковским? Составить интегральное уравнение для $M_k(t,x) = M(\eta_t | \eta_0 = x, \xi_0 = k)$, $k = 1, 2$. Найти $\lim\limits_{t\to\infty} M_k(t,x)/t$.

Решение. Процесс η_t не является марковским, так как координата η_t в момент t после момента t_0 зависит от состояния до момента t_0. Если, например, $\xi_{t_0-\delta} = 1$, $\xi_s = 2$ при $t_0 - \delta < s < t_0 + \delta = t$, то $\eta_{t_0} = \eta_{t_0-\delta} - a \cdot \delta^2/2$, $\eta_t = \eta_{t_0-\delta} - 2a\delta^2$. Если же $\xi_{t_0-\delta} = 1$, $\xi_s = 2$ при

$t_0-\delta<s<t_0$ и при $t_0<s<t_0+\delta$ и $\xi_{t_0}=1$, то $\eta_{t_0}=\eta_{t_0-\delta}-a\cdot\delta^2/2$, $\eta_t=\eta_{t_0}-a\delta^2/2=\eta_{t_0-\delta}-a\delta^2$.

Пусть τ – случайное время до первого выхода из состояния 1 при условии, что $\xi_0=1$. Как в задаче 7 находим, что $p_\xi()=\alpha^{-\alpha x}$. $P(\tau>t)=e^{-\alpha t}$.

$$M_1(t,x)=M(\eta_t/\eta_0=x,\xi_0=1)=$$

$$=P(\tau>t)M(\eta_t|\eta_0=x,\xi_s=1,\ 0\le s\le t)+$$

$$+\int\limits_0^t p_\tau(u)duM(\eta_t|\eta_0=x,\xi_s=1 \text{ при } 0\le s\le u,\xi_0=2).$$

$$M_1(t,x)=P(\tau>t)(x+vt)+\int\limits_0^t p_\tau(u)M_2(t-u,x+vu)du=$$

$$=e^{-\alpha t}(x+vt)+\int\limits_0^t \alpha e^{-\alpha u}M_2(t-u,x+vu)du.$$

Пусть теперь τ —случайное время до первого выхода из состояния 2 при условии, что $\xi_0=2$. $p_\tau(x)=\beta e^{-\beta x}$. Если $\tau=u<t$, то в промежутке $[0,u]$ частица будет двигаться с ускорением a влево и займет в конце промежутка положение $x-au^2/2$. В итоге, как и в случае $M_1(t,x)$, получаем

$$M_2(t,x)=e^{-\beta t}(x-at^2/2)+\int\limits_0^t \beta e^{-\beta u}M_1(t-u,x-at^2/2)du.$$

Так как $M_k(t,x)=x+M_k(t,0)$, то при $=0$ имеем

$$\begin{cases} M_1(t,0)=v(1-e^{-\alpha t})/\alpha+\int\limits_0^t \alpha e^{-\alpha u}M_2(t-u,0)du, \\ M_2(t,0)=a[(\beta t+1)e^{-\beta t}-1]/\beta^2+\int\limits_0^t \beta e^{-\beta u}M_1(t-u)du \end{cases}.$$

Используя операционное исчисление, находим изображения

$$m_1(p) = \frac{v(p+\beta)^2 - \beta\alpha}{p^2(p+\beta)(p+\alpha+\beta)},\ m_2(p) = \frac{v\beta(p+\beta) - a(p+\alpha)}{p^2(p+\beta)(p+\alpha+\beta)},$$

откуда

$$\lim_{t\to\infty} \frac{M_k(t)}{t} = m_k(p)|_{p=0} = \frac{v\beta^2 - a\alpha}{\beta(\alpha+\beta)}.$$

Задача 10.15. Пусть $\xi_t(t = ...-2,-1,0,1,2,...)$ — последовательность независимых случайных величин с $\xi_t = a$, $D\xi_t = \sigma^2$. Положим $\eta_t = c_0\xi_1 + c_1\xi_{t-1} + c_2\xi_{t-2}$, $c_0 + c_1 + c_2 = 1$ и $a_t^* = (\eta_1 + \eta_2 + \cdots \eta_t)/t$. Найти a_t^*, Da_t^*. Является ли a_t^* несмещенной и состоятельной оценкой a?

Решение. $M\eta = a$. $Ma^* = a$. $a*$ — несмещенная оценка. $D\eta_t = (c_0^2 + c_1^2 + c_2^2)\sigma^2$.

$$cov(\eta_{t_1}, \eta_{t_2}) = \sigma^2 \begin{cases} 0, & t_1 + 2 < t_2 \\ c_0c_2, & t_1 + 2 = t_2 \\ c_0c_1 + c_1c_2, & t_1 + 2 > t_2 \end{cases}$$

$Da_t^* = t^{-2}\sigma^2(t(c_0^2 + c_1^2 + c_2^2) + 2(t-1)(c_0c_1 + c_1c_2) + 2(t-2)c_0c_2) \to 0, t \to \infty$. a_t^* — состоятельная оценка.

Задачи домашнего задания.

Задача 10.2. Доказать, что для любого промежутка времени $\tau_k = \theta_k - \theta_{k-1}$, где θ_k — момент k–го скачка процесса Пуассона, выполняется равенство $P(\tau_k > t + s|\tau > s) = P(\tau_k > t)$.

Задача 10.4. Пусть μ_t — ветвящийся процесс с непрерывным временем; $\varphi(x) = Mm^{\mu_t} = px^2 + 1 - p$. Найти: a) $\lambda = \lim\limits_{t\to\infty} P(\mu_t = 0)$, $p > 1/2$; б) $B_t = M\mu_t(\mu_t - 1)$ и $D\mu_t$ при $p = 1/2$.

Ответ: а) $\lambda = (1 - p/p)$; б) $D\mu_t = B_t = t,\ M\mu_t = 1$.

Задача 10.6. Случайные величины $\xi_t(t = 1,2,...)$ независимы и $P(\xi_t = 1) = p$, $P(\xi_t = -1) = 1 - p = q$. Положим

$$\eta_{t+1} = \begin{cases} \eta_t + \xi_{t+1}, & \text{если } \eta_t \neq 0, \\ a, & \text{если } \eta_t = 0, t = 0,1,2,...; \end{cases}$$

$\eta_0 = k$. Найти: 1) вероятность того, что процесс η_t когда-либо попадет в состояние 0, если $a = 0$; 2) стационарное распределение вероятностей $\pi_l = \lim\limits_{t\to\infty} P(\eta_l)$, $a = 1$, $q > p$.

Ответ: 1) 1, если $q \geq p$; $(q/p)^k$, если $q < p$; 2) $\pi_0 = \frac{q-p}{1+q-p}$; $\pi_k = \frac{\pi_0}{p}(p/q)^k,\ k \geq 1$.

Задача 10.8. Марковский случайный процесс ξ_t с непрерывным временем называют цепью Маркова, если множество его состояний конечно или счетно. Пусть ξ_t — цепь Маркова с двумя состояними (1 и 2) и при $h \to 0$ $P(\xi_h = 2|\xi_0 = 1) = \alpha h + o(h)$, $P(\xi_h = 1|\xi_0 = 2) = \beta h + o(h)$. Найти $P_{ij}(t) = P(\xi_t = j|\xi_0 = i)$, $i, j = 1, 2$.

Ответ: $\|P_{ij}(t)\| = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{e^{-(\alpha+\beta)t}}{\alpha+\beta}\begin{matrix} \beta & \alpha \\ \beta & \alpha \end{matrix}\Bigg)$.

Задача 10.10. Найти $\lim\limits_{t\to\infty} \frac{1}{t} M(\tau_1(t)|\xi_0 = k)$, где $\tau_1(t)$ определено в задаче 10.9.

Указание. Использовать формулу $\tau_1(t) = \int\limits_0^t \delta_{1,\xi_u} du$, где $\delta_{1,1} = 1$, $\delta_{1,2} = 0$.

Ответ: $\beta/(\alpha+\beta)$

Задача 10.12. Решить задачу 10.11, используя равенство $\eta_t = v_1\tau_1(t) - v_2\tau_2(t) + x$, где $\tau_k(t)$ определены в задаче 10.9, и решение задачи 10.10.

Задача 10.14. Пусть $\xi_t (t = \ldots - 2, -1, 0, 1, 2, \ldots)$ — последовательность независимых случайных величин с $M\xi_t = a$, $D\xi_t = \sigma^2$. Положим $\eta_t = c_0\xi_t + c_1\xi_{t-1} + c_2\xi_{t-2}$, $c_0 + c_1 + c_2 = 1$. Найти $M\eta_t$, $D\eta_t$, $cov(\eta_{t_1}, \eta_{t_2})$. Проверить, что $cov(\eta_{t_1}, \eta_{t_2}) = R(|t_1 - t_2|)$.

Ответ: $M\eta_t = a,\ D\eta_t = (c_0^2 + c_1^2 + c_2^2)\sigma^2 = R(0)$, $R(1) = c_1(c_0 + c_2)\sigma^2$, $R(2) = c_0c_2\sigma^2$, $R(k) = 0,\ k \geq 0$.

Задача 10.16. Пусть η_t — процесс, определенный в задаче 10.14. Показать, что $R_t^*(k) = \frac{1}{t}\sum\limits_{s=1}^{t}(\eta_s - a_t^*)(\eta_{s+k} - a_t^*)$, $k = 0, 1, 2$, является асимптотически несмещенной оценкой $R(k)$.

5 ТАБЛИЦЫ

5.1 Случайные числа

10	09	73	25	33	76	52	01	35	86	34	67	35	48	76
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37
08	42	26	89	53	19	64	50	93	03	23	20	90	25	60
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65
12	80	79	99	70	80	15	73	61	47	64	03	23	66	53
80	95	90	91	17	39	29	27	49	45	66	06	57	47	17
20	63	61	04	02	00	82	29	16	65	31	06	01	08	05
15	95	33	47	64	35	08	03	36	06	85	26	97	76	02
88	67	67	43	97	04	43	62	76	59	63	57	33	21	35
98	95	11	68	77	12	17	17	68	33	73	79	64	57	53
34	07	27	68	50	36	69	73	61	70	65	81	33	98	85
45	57	18	24	06	35	30	34	26	14	86	79	90	74	39
02	05	16	56	92	68	66	57	48	18	73	05	38	52	47
05	32	54	70	48	90	55	35	75	48	28	46	82	87	09
03	52	96	47	78	35	80	83	42	82	60	93	52	03	44
11	19	92	91	70	98	52	01	77	67	14	90	56	86	07
23	40	30	97	32	11	80	50	54	31	39	80	82	77	32
18	62	38	85	79	83	45	29	96	34	06	28	89	80	83
83	49	12	56	24	88	68	54	02	00	86	50	75	84	01
35	27	38	84	35	99	59	46	73	48	87	51	76	49	69

5.2 Нормальное распределение и значения функции u_α

Значения функции $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{u^2}{2}} du.$

В первой строке таблицы на стр.88 после x указаны сотые доли значений x, меньшие 5. В первой строке таблицы на стр. 89 после x указаны сотые доли значений x, большие 4.

x	0	1	2	3	4
0.0	0.0000	0.0040	0.0080	0.0120	0.0160
0.1	398	438	478	517	557
0.2	793	832	871	910	948
0.3	0.1179	0.1217	0.1255	0.1293	0.1331
0.4	554	591	628	664	700
0.5	915	950	985	0.2019	0.2054
0.6	0.2257	0.2291	0.2324	357	389
0.7	580	611	642	673	703
0.8	881	910	939	967	995
0.9	0.3159	0.3186	0.3212	0.3238	0.3264
1.0	413	437	461	485	508
1.1	643	665	686	708	729
1.2	849	869	888	907	925
1.3	0.4032	0.4049	0.4066	0.4082	0.4099
1.4	192	207	222	236	251
1.5	332	345	357	370	382
1.6	452	463	474	484	495
1.7	554	564	573	582	591
1.8	641	649	656	664	671
1.9	713	719	726	732	738
2.0	772	778	783	788	793
2.1	821	826	830	834	838
2.2	861	864	868	871	875
2.3	893	896	898	901	904
2.4	918	920	922	925	927
2.5	938	940	941	943	945
2.6	953	955	956	957	959
2.7	965	966	967	968	969
2.8	974	975	976	977	977
2.9	981	982	982	983	984
3.0	987	987	987	988	988

x	5	6	7	8	9
0.0	0.0200	0.0239	0.0279	0.0319	0.0359
0.1	596	636	675	714	753
0.2	987	0.1026	0.1064	0.1103	0.1141
0.3	0.1368	406	443	480	517
0.4	736	772	808	844	879
0.5	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	422	454	486	517	549
0.7	734	764	794	823	852
0.8	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	289	315	340	365	389
1.0	531	554	577	599	621
1.1	749	770	790	810	830
1.2	944	962	980	997	0.4015
1.3	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	265	279	292	306	319
1.5	394	406	418	429	441
1.6	505	515	525	535	545
1.7	599	608	616	625	633
1.8	678	686	693	699	706
1.9	744	750	756	761	767
2.0	798	803	808	812	817
2.1	842	846	850	854	857
2.2	878	881	884	887	890
2.3	906	909	911	913	916
2.4	929	931	932	934	936
2.5	946	948	949	951	952
2.6	960	961	962	963	964
2.7	970	971	972	973	974
2.8	978	979	979	980	981
2.9	984	985	985	985	986
3.0	989	989	989	990	990

Значения функции u_α определяются равенством
$\alpha = \frac{1}{\sqrt{2\pi}} \int\limits_{u_\alpha}^{\infty} e^{-\frac{x^2}{2}} dx.$

α	0.001	0.005	0.010	0.015	0.020	0.025
u_α	3.0902	2.5758	2.3263	2.1701	2.0537	1.9600
α	0.030	0.035	0.040	0.045	0.050	
u_α	1.8808	1.8119	1.7507	1.6954	1.6449	

5.3 Распределение Стьюдента

Значения функции $t_{\alpha,n}$.

Функция $t_{\alpha,n}$ определяется равенством $P(|\tau_n| < t_{\alpha,n}) = 1 - 2\alpha$, где случайная величина τ_n имеет распределение Стьюдента с n степенями свободы. Плотность распределения τ_n равна

$$p_{\tau_n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \cdot (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}.$$

$n \setminus 2\alpha$	0.1	0.05	0.02	0.01
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
12	1.782	2.179	2.681	3.055
14	1.761	2.145	2.624	2.977
16	1.746	2.120	2.583	2.921
18	1.734	2.101	2.552	2.878
20	1.725	2.086	2.528	2.845
22	1.717	2.074	2.508	2.819
30	1.697	2.042	2.457	2.750
∞	1.645	1.960	2.326	2.576

5.4 χ^2-распределение и F-распределение

Значения функции $\chi^2_{\alpha,m}$.

Функция $\chi^2_{\alpha,m}$ определяется равенством $P(\chi^2_m > \chi^2_{\alpha,m}) = \alpha$, где случайная величина χ^2_m имеет χ^2-распределение с m степенями свободы. Плотность распределения χ^2_m равна

$$p_{\chi^2_m}(x) = \frac{1}{\Gamma(\frac{n}{2})2^{n/2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \ x > 0.$$

$m \setminus \alpha$	0.9	0.10	0.05	0.02	0.01	0.005
1	0.00016	2.7	3.8	5.4	6.6	7.9
2	0.020	4.6	6.0	7.8	9.2	11.6
3	0.115	6.3	7.8	9.8	11.3	12.8
4	0.30	7.8	9.5	11.7	13.3	14.9
5	0.55	9.2	11.1	13.4	15.1	16.3
6	0.87	10.6	12.6	15.0	16.8	18.6
7	1.24	12.0	14.1	16.6	18.5	20.3
8	1.65	13.4	15.5	18.2	20.1	21.9
9	2.09	14.7	16.9	19.7	21.7	23.6
10	2.56	16.0	18.3	21.2	23.2	25.2
11	3.1	17.3	19.7	22.6	24.7	26.8
12	3.6	18.5	21.0	24.1	26.2	28.3
13	4.1	19.8	22.4	25.5	27.7	29.8
14	4.7	21.1	23.7	26.9	29.1	31.0
15	5.2	22.3	25.0	28.3	30.6	32.5
16	5.8	23.5	26.3	29.6	32.0	34.0
17	6.4	24.8	27.6	31.0	33.4	35.5
18	7.0	26.0	28.9	32.3	34.8	37.0
19	7.6	27.2	30.1	33.7	36.2	38.5
20	8.3	28.4	31.4	35.0	37.6	40.0
21	8.9	29.6	32.7	36.3	38.9	41.5
22	9.5	30.8	33.9	37.7	40.3	42.5
23	10.2	32.0	35.2	39.0	41.6	44.0
24	10.9	33.2	36.4	40.3	43.0	45.5
25	11.5	34.4	37.7	41.6	44.3	47.0

Значения функции $F_{\alpha;n_1,n_2}$.

Функция $F_{\alpha;n_1,n_2}$ определяется равенством $P(F_{n_1,n_2} > F_{\alpha;n_1,n_2} = \alpha$, где случайная величина F_{n_1,n_2} имеет F-распределение с n_1 и n_2 степенями свободы. Плотность распределения F_{n_1,n_2} равна

$$p_{F_{n_1,n_2}}(x) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}(\frac{n_1}{n_2})^{\frac{n_1}{2}}\frac{x^{\frac{n_1}{2}-1}}{(\frac{n_1}{n_2}x+1)^{\frac{n_1+n_2}{2}}},\ x > 0.$$

$\alpha = 0.05$

$n_2 \setminus n_1$	10	200	30	40	50	100	∞
10	2.97	2.77	2.70	2.67	2.64	2.59	2.54
15	2.55	2.33	2.25	2.21	2.18	2.12	2.07
20	2.35	2.12	2.04	1.99	1.96	1.90	1.84
30	2.16	1.93	1.84	1.79	1.76	1.69	1.62
40	2.07	1.84	1.74	1.69	1.66	1.59	1.51
50	2.02	1.78	1.69	1.63	1.60	1.52	1.44
100	1.92	1.68	1.57	1.51	1.48	1.39	1.28
∞	1.83	1.57	1.46	1.40	1.35	1.24	1.00

5.5 Нормально распределенные случайные числа

Приведенные в таблице числа можно рассматривать как реализации независимых нормально распределенных величин с параметрами $a = 0, \sigma^2 = 1$.

0.464	0.137	2.455	−0.323	−0.068	0.296	−0.288	1.298
0.060	−2.256	−0.531	−0.194	0.543	−1.558	0.187	−1.190
1.486	−0.354	−0.634	0.697	0.926	1.375	0.785	−0.963
1.022	−0.472	1.279	3.521	0.571	−1.851	0.194	1.192
1.394	−0.555	0.046	0.321	2.945	1.974	−0.258	0.412
0.906	−0.513	−0.525	0.595	0.881	−0.934	1.579	0.161
1.179	−1.055	0.007	0.769	0.971	0.712	1.090	−0.631
−1.501	−0.488	−0.162	−0.136	1.033	0.303	0.448	0.748
−0.690	0.756	−1.618	−0.345	−0.511	−2.051	−0.457	−0.218
1.372	0.225	0.378	0.761	0.181	−0.736	0.960	−1.530
−0.482	1.678	−0.057	−1.229	−0.486	0.856	−0.491	−1.983
−1.376	−0.150	1.356	−0.561	−0.256	−0.212	0.219	0.779
−1.010	0.598	−0.918	1.598	1.065	0.415	−0.169	0.313
−0.005	−0.899	0.012	−0.725	0.147	−0.121	1.096	0.181
1.393	−1.163	−0.911	1.231	−0.199	−0.246	1.239	−2.574

6 ЛИТЕРАТУРА

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

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

3. *Натансон И.П.* Теория функций вещественной переменной. — М.:ГИТТЛ, 1957.

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

5. *Боровков А.А.* Теория вероятностей. — М.: Наука, 1976.

6. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика. — М.: Высшая школа, 1992.

7. *Володин И.Н.* Математическая статистика. — Казань: Изд-во КГУ, 2000.

Содержание

Билялов Ранат Фаизович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Лекционный курс и практические занятия

2-е издание, исправленное и дополненное

Подписано в печать 12.07.2004. Форм. 60× 84 1/16. Гарнитура "Таймс".
Печать офсетная. Печ. л. 8.63. Тираж 250. Заказ 201
Лаборатория оперативной полиграфии КГУ
420045, Казань, Кр. Позиция, 2а
Тел. 72-22-54