Министерство образования и науки Российской Федерации

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ПЕТРА ВЕЛИКОГО

К. В. НИКИТИН

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ПРАКТИКУМ

Санкт-Петербург
2018

Оглавление

Правила оформления расчетных заданий

- Ни в коем случае не вставляйте в отчет код программ из Matlab (его помещайте в приложение к отчету), а также результаты работы Matlab из командного окна (их необходимо обработать и представить с учетом правильного названия переменных и экономии места).
- Большие объемы данных приводите в табличной форме маленьким шрифтом. В случае если данные в таблице не влезают на одну страницу, приводите лишь начальную и конечную часть. В целом же при больших объемах данных предпочтения всегда отдавайте графической форме представления.
- Используйте место экономно – не оставляйте лишних пустых строк.
- Нумеруйте страницы, разделы, рисунки и таблицы.
- Старайтесь оформлять отчет в едином стиле.
- Текст задания помещайте не сразу целиком вначале отчета, а частями, так чтобы после каждого описания задания следовало бы его решение.
- При выполнении расчетов вначале представляйте формулы, по которым производится расчет и уже после этого – результаты расчетов при подстановке конкретных числовых значений. Не рекомендуется сразу выписывать все формулы и затем приводить результаты расчетов по ним.
- Структура отчета должна повторять структуру задания.
- В конце работы обязательно проанализируйте все полученные результаты и напишите развернутые выводы.

Таким образом, каждый пункт задания должен иметь следующую структуру.

1. Задание.
2. Формулы или алгоритм выполнения задания.
3. Результаты расчетов после подстановки значений (таблицы, графики).
4. Анализ результатов.

Практическая работа

Отработка навыков решения практических задач

Исходные данные

- В качестве источника задач используется учебное пособие «Сборник задач по теории вероятностей, математической статистике и теории случайных функций» издательства Лань, 3 или 4 издание (доступно в электронном виде в общей папке Books файлового хранилища).
- Каждому студенту необходимо решить 27 задач из основных разделов. Задачи пронумерованы в форме x.y, где x – номер раздела, y – номер задачи. Аналогичная нумерация используется в учебнике. В личной папке файлового хранилища для каждого студента приведены номера задач в файле " zadachi_sveshnikov.txt".

Представление результатов

- Представление результатов возможно электронное (печатное) или рукописное. Титульный лист делается в печатном виде – на нем указывается группа, ФИО студента и список задач.
- Перед началом каждой задачи должен быть обязательно проставлен ее номер x.y.
- Желательно, но не обязательно соблюдать порядок решения задач.
- Недопустимым является приведение ответа без решения. В то же время сверка полученных в результате решения ответов с правильными является желательной.

Защита работы

- Следует учитывать, что при защите данной работы по выбору преподавателя студенту может быть предложено объяснить решение любой из задач.
- Рекомендуется сдавать эту работу в 2 этапа – задачи с 1 по 14 – к середине марта, задачи с 15 по 27 – к концу апреля. Необходимо учитывать фактор времени при проверке решения задач – сдача этих работ на зачетной неделе и после сильно увеличивает риск неполучения зачета (допуска к экзамену) вовремя.

Вопросы

- При возникновении вопросов по решению задач обращайтесь к учебникам, в случае же непреодолимых затруднений спрашивайте у преподавателя на упражнениях.

Расчетное задание 0

Приобретение основных навыков программирования в среде Matlab

Указания

- Данная работа предназначена для первого ознакомления со средой Matlab.
- Классический отчет по этой работе не требуется – достаточно предоставить тексты программ и скриптов для каждого из пунктов.
- Все задания выполняются полностью.
- Для заданий, у которых есть таблица с вариантами, используется свой вариант.
- Вариант указан в личной папке файлового хранилища в файле "\Tasks\Task_0.txt".
- В случае, если номер k вашего варианта для какого-то задания отсутствует, выполняйте вариант, вычислив его по простой формуле knew=mod(k,n)+1, где knew – ваш новый вариант задания, mod – функция вычисления остатка, n – количество вариантов в задании.
- В случае, если в задании нет варианта, выполняйте его с произвольными данными.

Операции с числами

Задание 1

Вычислите указанное арифметическое выражение. Укажите последовательность нажатия клавиша. Сравните полученный результат с приведенным ответом.

1. $\dfrac{\left(12\frac{1}{6}-6\frac{1}{27}-5{,}25\right)13{,}5+0{,}111}{0{,}02}.$ — 599,3

2. $\dfrac{\left(1\frac{1}{12}+2\frac{5}{32}+\frac{1}{24}\right):9{,}6+2{,}13}{0{,}0004}.$ — 6179,5

3. $\dfrac{\left(6{,}6-3\frac{3}{14}\right)5\frac{5}{6}}{(21-1{,}25):2{,}5}.$ — 2,5

4. $\dfrac{2{,}625-\frac{2}{3}\cdot 2\frac{5}{14}}{\left(3\frac{1}{12}+4{,}375\right):19\frac{8}{9}}.$ — 2,8095

5. $\dfrac{0{,}134+0{,}05}{18\frac{1}{6}-1\frac{11}{14}-\frac{2}{15}\cdot 2\frac{6}{7}}.$ — 0,0115

6. $\dfrac{\left(58\frac{4}{15}-56\frac{7}{24}\right):0{,}8+2\frac{1}{9}\cdot 0{,}225}{8{,}75\cdot 0{,}6}.$ — 0,56071

7. $\dfrac{\left(\frac{0{,}216}{0{,}15}+0{,}56\right):0{,}5}{\left(7{,}7:24{,}75+\frac{2}{15}\right)4{,}5}.$ — 2

8. $\dfrac{1\frac{4}{11}\cdot 0{,}22:0{,}3-0{,}96}{\left(0{,}2-\frac{3}{40}\right)1{,}6}.$ — 0,2

9. $\dfrac{\left(\frac{3}{5}+0{,}425-0{,}005\right):0{,}12}{30{,}5+\frac{1}{6}+3\frac{1}{3}}.$ — 0,25

10. $\dfrac{3\frac{1}{3}+2{,}5}{2{,}5-1\frac{1}{3}}\cdot\dfrac{4{,}6-2\frac{1}{3}}{4{,}6+2\frac{1}{3}}:\left(\dfrac{0{,}05}{\frac{1}{7}-0{,}125}+5{,}7\right).$ — 0,19231

11. $\dfrac{0{,}725+0{,}42(6)}{0{,}128-6{,}25-(0{,}0345:0{,}12)}\cdot 0{,}25.$ — -0,04492

12. $\dfrac{\left(4{,}5\cdot 1\dfrac{2}{3}-6{,}75\right)\cdot 0{,}(6)}{\left(3{,}333\cdot 0{,}3+0{,}222\cdot\dfrac{4}{9}\right)\ 2\dfrac{2}{3}}$. 0,17068

13. $\dfrac{\left(5\dfrac{4}{45}-4\dfrac{1}{6}\right)\ :\ 5\dfrac{8}{15}}{\left(4\dfrac{2}{3}+0{,}75\right)3\dfrac{9}{13}}\cdot 34\dfrac{2}{7}$. 0,28571

14. $\dfrac{1\dfrac{4}{11}\cdot 0{,}22\ :\ 0{,}3-0{,}96}{\left(0{,}2-\dfrac{3}{40}\right)1{,}68}$. 0,19048

15. $\dfrac{\left(40\dfrac{7}{30}-38\dfrac{5}{12}\right)\ :\ 10{,}9+\left(0{,}875-\dfrac{7}{30}\right)\cdot\dfrac{20}{11}}{0{,}008}$. 166,67

16. $\dfrac{(68{,}023-66{,}028)\ :\ 6\dfrac{1}{9}+\dfrac{7}{40}\cdot 4{,}5}{0{,}042+0{,}086}$. 8,7028

17. $\dfrac{(2{,}1-1{,}965)\ :\ (1{,}2\cdot 0{,}045)}{0{,}00325\ :\ 0{,}013}-\dfrac{4}{0{,}2\cdot 073}$. -17,397

18. $\dfrac{(1{,}88+2{,}127)\cdot 0{,}01875}{0{,}625-\dfrac{13}{18}\ :\ 3{,}13}+8{,}29$. 8,2441

19. $\dfrac{3\ :\ 0{,}4-0{,}009\ :\ (0{,}15\ :\ 2{,}5)}{0{,}32\cdot 6+0{,}033-(5{,}3-3{,}88)}$. 13,79

20. $\dfrac{(34{,}06-33{,}81)\cdot 4}{6{,}84\ :\ (28{,}57-25{,}15)}+1{,}33\ :\ \dfrac{4}{21}$. 7,4825

21. $\dfrac{8{,}8077}{20-\left(28{,}2\ :\ (13{,}333\cdot 0{,}3+0{,}0125)\right)2{,}004}$. 1,4889

22. $\dfrac{\left(1{,}75\ :\ \dfrac{2}{3}-1{,}75\cdot 1{,}125\right)\ :\ \dfrac{7}{12}}{(0{,}2012-0{,}0325)\ :\ 400}$. 2667,5

23. $\dfrac{\left(26\dfrac{1}{3}-18{,}02\cdot 0{,}75\right)\cdot 2{,}4\ :\ 0{,}88}{1{,}37-23\dfrac{2}{3}\ :\ 1{,}82}$. -3,005

24. $26\ :\ \dfrac{3\ :\ (0{,}48-0{,}27)}{2{,}52(1{,}38+2{,}45)}+1{,}27$. 18,836

25. $\left(16{,}5-13\dfrac{7}{9}\right)\dfrac{6}{11}+2{,}2\ :\ (0{,}241-0{,}91)$. -1,8036

Задание 2

Проведите вычисления по заданной формуле при заданных значениях параметров. Укажите необходимую последовательность действий.

Сравните полученный результат с приведенным ответом.

Указание. В системе MatLAB несколько последних команд запоминаются. Повторный вызов этих команд в командное окно осуществляется нажатием клавиш <↑ и <↓>. Используйте эту возможность для повторного обращения к набранной функции.

1. $3m^2 + \sqrt[3]{2n^2} : m$; а) $m = -\dfrac{14}{5}$, $n = tg\dfrac{\pi}{8}$; б) $m = 2{,}2 \cdot 10^{-2}$, $n = \dfrac{1}{3{,}1}$.

ОТВЕТ: а) 23,27; б) 26,938.

2. $\dfrac{4}{3} l^3 \sin^2 \dfrac{\alpha}{2} \sqrt{\cos\alpha}$; а) $l = 1{,}7 \cdot 10^3$, $\alpha = 18°$; б) $l = \dfrac{16}{21}$, $\alpha = \dfrac{\pi}{5}$.

ОТВЕТ: а) 1. 5633e+008; б) 5. 0651e-002.

3. $\sqrt{\dfrac{a\sqrt{b}}{\sqrt[3]{tg\alpha}}}$; а) $a = 1{,}5$, $b = 0{,}8$, $\alpha = 61°$; б) $a = 3 \cdot 10^{-2}$, $b = 0{,}71$, $\alpha = \dfrac{3}{7}\pi$.

ОТВЕТ: а) 1. 0498e+000; б) 1. 2429e-001.

4. $\dfrac{3a^2\sqrt{6{,}8 \cdot (a-b)}}{4(a+b)^3}$; а) $a = 4{,}13 \cdot 10^{-1}$, $b = \dfrac{1}{261}$;

б) $a = \sin\dfrac{5\pi}{8}$, $b = -tg12°$.

ОТВЕТ: а) 2. 9464e+000; б) 4. 9445e+000.

5. $\dfrac{c^3}{6}\cos\dfrac{\alpha}{2}\sqrt{\sin\alpha}$; а) $c = \lg 2{,}38$, $\alpha = \dfrac{\pi}{5}$; б) $c = e^{-0{,}3}$, $\alpha = 65°$.

ОТВЕТ: а) 3. 4657e-004; б) 2. 2120e-002.

6. $\sqrt{\dfrac{n^3}{16{,}3\sin\alpha\sin 2\alpha}}$; а) $n = 3{,}1516 \cdot 10^{-2}$, $\alpha = 5°$; б) $n = e^{3{,}5}$, $\alpha = \dfrac{2\pi}{13}$.

ОТВЕТ: а) 1. 1265e-002; б) 7. 6324e+001.

7. $5\sin 35° \sqrt{\dfrac{S^3\cos 36°}{\pi^3 tg\alpha}}$; а) $S = \ln 3$, $\alpha = 44°$; б) $S = \dfrac{18}{25}$, $\alpha = \dfrac{7}{12}\pi$.

ОТВЕТ: а) 5. 4283e-001; б) 8. 9703e-018+ 1. 4650e-001i.

8. $\left|\lg(1+\sin\alpha) + \ln(1-\sin\beta)\right|$; а) $\alpha = \dfrac{3\pi}{7}$, $\beta = 83°$; б) $\alpha = \dfrac{2}{3}\pi$, $\beta = 16°$.

ОТВЕТ: а) 4. 6035e+000; б) 5. 1546e-002.

9. $\sqrt[3]{\sin^2(\alpha+\beta)-\sin^2(\alpha-\beta)}$; а) $\alpha=\frac{5}{7}\pi,\ \beta=0{,}3\pi$; б) $\alpha=12^\circ,\ \beta=220^\circ$

ОТВЕТ: а) 4. 8756e-001+ 8. 4448e-001i; б) 7. 3715e-001.

10. $\left(\log_a(b+1{,}4)\right)^{-3/4}$; а) $a=3{,}56,\ b=e^{0{,}316}$; б) $a=2,\ b=2{,}1649\cdot10^{-2}$.

ОТВЕТ: а) 1. 1790e+000; б) 1. 6630e+000.

11. $3\left(p^{-2/3}+q^{-1/2}\right)\sqrt[3]{pq}$; а) $p=\ln 3,\ q=\lg 3$; б) $p=0{,}013,\ q=1{,}4\cdot10^2$.

ОТВЕТ: а) 5. 7737e+000; б) 6. 6559e+001.

12. $\frac{2}{3}m\sqrt{m\sqrt[3]{m\sqrt[4]{m}}}$; а) $m=3{,}6485\cdot10^2$; б) $m=\frac{24}{37}$.

ОТВЕТ: а) 1. 5880e+004; б) 5. 4516e-001.

13. $\frac{8}{3}S\sqrt{\frac{S}{\pi}}\sin^6\frac{\alpha}{2}$; а) $S=e^{1{,}11},\ \alpha=\frac{7}{11}\pi$; б) $S=5{,}403,\ \alpha=28^\circ$.

ОТВЕТ: а) 2. 8187e+000; б) 3. 7879e-003.

14. $2\sqrt{\frac{F}{\pi}}tg\alpha\sin^2\frac{\alpha}{2}$; а) $F=\frac{1}{0{,}03},\ \alpha=\frac{5}{7}\pi$; б) $F=\ln 7,\ \alpha=1{,}34^\circ$.

ОТВЕТ: а) -6. 6313e+000; б) 5. 0346e-006.

15. $\frac{1}{12}\cdot\frac{m^3\cos\alpha}{(\sin\alpha+\cos\alpha)^3}$; а) $m=-20{,}1,\ \alpha=20^\circ$; б) $m=\lg 13{,}6,\ \alpha=1{,}48$.

ОТВЕТ: а) -3. 0201e+002; б) 8. 5792e-003.

16. $\frac{\sqrt{3h^3}}{\cos^2\alpha}\sin\left(\alpha+30^\circ\right)\sin\left(\alpha-30^\circ\right)$;

а) $h=0{,}28,\ \alpha=41^\circ$; б) $h=e^{0{,}415},\ \alpha=237^\circ$

ОТВЕТ: а) 8. 1284e-002; б) 4. 9334e+000.

17. $\frac{\alpha}{3}\left(\lg(d+2)-tg\alpha\right)^2$;

а) $d=6{,}178,\ \alpha=20^\circ$; б) $d=-2{,}2461\cdot10^{-2},\ \alpha=1{,}146$.

ОТВЕТ: а) 3. 5028e-002; б) 1. 4003e+000.

18. $d^3ctg\alpha\sqrt{\sin^4\alpha-\cos^4\alpha}$; а) $d=10{,}6,\ \alpha=50^\circ$; б) $d=e^{2{,}3},\ \alpha=1$.

ОТВЕТ: а) 4. 1645e+002; б) 4. 1101e+002.

19. $\frac{a^2\sqrt{3}}{4}(\sec\alpha+\operatorname{cosec}\alpha)^4$;

а) $a=5{,}08,\ \alpha=25^\circ$; б) $a=\ln 1{,}37,\ \alpha=\frac{12}{25}\pi$

ОТВЕТ: а) 1. 6193e+003; б) 3. 5238e+003.

Задание 3

Выполните такие действия (см. таблицу 1.1):

а) число z1, заданное в алгебраической (экспоненциальной) форме, переведите в экспоненциальную (алгебраическую), проверьте и запишите результат;

б) число z2, заданное в экспоненциальной (алгебраической) форме, переведите в алгебраическую (экспоненциальную), проверьте и запишите езультат;

в) вычислите заданное выражение; запишите результат экспоненциальной форме, причем аргумент результата обеспечьте в границах между (-π) и +π.

Варі ант	Комплексное число				Выражение
	z_1	z_2	z_3	z_4	
1	$4+3i$	$2{,}71e^{i\pi/12}$	$1{,}82e^{-1{,}2i}$	$\sqrt{3}-2i$	$z_1^2\cdot z_2:z_3+z_4$
2	$0{,}8-2i$	$3{,}08e^{i7\pi/12}$	$8{,}01e^{2i}$	$-5+\sqrt{2}i$	$z_1^2:z_2+z_3-z_4$
3	$-0{,}7+4i$	$1{,}74e^{i0{,}3\pi}$	$3+4i$	$2{,}1e^{-2{,}3i}$	$\sqrt{z_1:z_2}\cdot z_3+z_4$
4	$-3-2i$	$3{,}21e^{15^\circ i}$	$1{,}2+3i$	$2{,}71e^{-78^\circ i}$	$\sqrt{z_1\cdot z_2}:z_3+z_4$
5	$2{,}71e^{i\pi/12}$	$-0{,}7+4i$	$1{,}31e^{-i5\pi/12}$	$-8-3i$	$\sqrt{z_1}:z_2\cdot z_3-z_4$
6	$3{,}08e^{i7\pi/12}$	$-3-2i$	$2{,}03e^{i14/13}$	$\sqrt{3}+\sqrt{2}i$	$(z_1+z_2)^4\cdot z_3:z_4$
7	$1{,}74e^{0{,}3\pi i}$	$0{,}8-2i$	$3{,}28e^{-1{,}2i}$	$\sqrt{3}-\sqrt{2}i$	$\left(\sqrt{z_1}+z_2\right)\cdot z_3:z_4$
8	$3{,}21e^{15i}$	$4+3i$	$\sqrt{3}-4i$	$1{,}23e^{111^\circ i}$	$(z_1-z_2)^3\cdot z_3+z_4$
9	$1+i\pi/2$	$1{,}2e^{107^\circ i}$	$0{,}8-2i$	$2{,}5e^{14^\circ i}$	$(z_1:z_2+z_3)^3\cdot z_4$
10	$\sqrt{5}-i$	$0{,}7e^{1{,}7i}$	$1{,}2e^{0{,}9i}$	$-3-2i$	$(z_1:z_2+z_3)^2-z_4$
11	$0{,}187-3{,}94i$	$0{,}3e^{-107^\circ i}$	$-0{,}7+4i$	$1{,}5e^{23^\circ i}$	$\sqrt[3]{z_1+z_2-z_3}\cdot z_4$
12	$-1+\sqrt{5}i$	$2{,}1e^{211^\circ i}$	$0{,}4e^{32^\circ i}$	$4+3i$	$\sqrt[3]{z_1}\cdot z_2:z_3+z_4$
13	$-\sqrt{3}-4i$	$1{,}25e^{-0{,}8i}$	$-3-2i$	$0{,}75e^{0{,}7i}$	$\left(\sqrt[3]{z_1\cdot z_2}+z_3\right):z_4$
14	$1{,}2e^{1{,}7i}$	$0{,}18-3{,}9i$	$0{,}71e^{4i}$	$0{,}8-2i$	$\left(\sqrt[3]{z_1:z_2}+z_3\right)\cdot z_4$
15	$0{,}3e^{-97^\circ i}$	$-1+\sqrt{5}i$	$-0{,}7+4i$	$5{,}2e^{71^\circ i}$	$\left(\sqrt{z_1\cdot z_2}-z_3\right):z_4$
16	$1{,}25e^{0{,}6i}$	$-\sqrt{3}-4i$	$4+3i$	$2{,}5e^{3{,}8i}$	$\left(\sqrt{z_1:z_2}-z_3\right)\cdot z_4$
17	$1{,}05e^{-0{,}4i}$	$\sqrt{5}-i$	$2{,}7e^{0{,}8i}$	$-0{,}7+4i$	$\sqrt{(z_1:z_2+z_3)\cdot z_4}$
18	$2{,}1e^{73^\circ i}$	$1+i\pi/2$	$\sqrt{2}+\sqrt{3}i$	$1{,}93e^{192^\circ i}$	$\sqrt{(z_1\cdot z_2-z_3):z_4}$
19	$2{,}7+0{,}8i$	$2e^{-\sqrt{3}i}$	$0{,}81e^{i\pi/7}$	$-\sqrt{2}-\sqrt{3}i$	$\sqrt{(z_1+z_2):z_3\cdot z_4}$
20	$-0{,}8+2{,}7i$	$-2e^{\sqrt{3}i}$	$0{,}9e^{i5\pi/7}$	$3{,}1-2{,}1i$	$\sqrt{(z_1+z_2)\cdot z_3:z_4}$
21	$-1{,}1-3{,}2i$	$0{,}33e^{-1{,}9i}$	$2e^{\sqrt{2}i}$	$2{,}08+i$	$\sqrt{z_1-z_2}\cdot z_3:z_4$
22	$2{,}1-3{,}2i$	$0{,}68e^{148^\circ i}$	$\sqrt{5}+\sqrt{2}i$	$2{,}73e^{23^\circ i}$	$\sqrt{z_1-z_2}:z_3\cdot z_4$
23	$1{,}1e^{-0{,}8i}$	$\sqrt{5}-2i$	$-1{,}7+i$	$0{,}97e^{\sqrt{2}i}$	$\left((z_1+z_2)^2-z_3\right):z_4$
24	$2{,}1e^{0{,}8i}$	$-\sqrt{5}+2i$	$1{,}7e^{\sqrt{3}i}$	$0{,}8e^{2{,}5i}$	$\left((z_1-z_2)^2+z_3\right):z_4$
25	$1{,}1e^{-2{,}1i}$	$\pi/8-2{,}1i$	$2{,}71+0{,}4i$	$1{,}71e^{-\sqrt{3}i}$	$(z_1-z_2)^3:z_3+z_4$

Задание 4

Найдите корни квадратного уравнения $ax^2 + bx + c$ при заданных значениях коэффициентов *a*, *b* и *c*.

Вариант	*a*	*b*	*c*
1	0.56	1.2e-4	4.08
2	1	0.1	100
3	4. 2e-3	8. 03e-4	1.06
4	7. 1e3	9. 4e4	8. 3e10
5	5.09	4.32	256
6	8.3	5.34	693
7	27	27	1276
8	3.08	0.2	30
9	5.3	10.6	876
10	0.45	0. 034	121
11	4.3	10.7	3. 4e3
12	13	0.8	287
13	6. 035	5.2	875
14	2.3	7.9	324
15	1	0.02	16.57
16	1.3	0.56	18.8
17	0.13	0. 056	18.8
18	17	12	956
19	0. 085	1	1. 3e3
20	1.2	0.32	15
21	7.1	6.4	256
22	0.2	0. 002	2.9
23	1. 4e-3	3.9	2. 6e2
24	0.86	3.2	5. 4e2
25	7. 3e3	8. 2e2	3. 5e8

Простейшие операции с векторами и матрицами

Задание 5

Вычислите значения функции *f(x)* на отрезке [*a*; *b*] с шагом *h*.

Вариант	$f(x)$	a	b	h
1	$\dfrac{x^2}{1+0{,}25\sqrt{x}}$	1,1	3,1	0,2
2	$\dfrac{x^3-0{,}3x}{\sqrt{1+2x}}$	2,05	3,05	0,1
3	$\dfrac{2e^{-x}}{2\pi+x^3}$	0	1,6	0,16
4	$\dfrac{\cos\pi x^2}{\sqrt{1-3x}}$	-1	0	0,1
5	$\sqrt{1+4x}\sin\pi x$	0,1	0,8	0,07
6	$\dfrac{e^{x/3}}{1+x^2}$	1,4	2,4	0,1
7	$e^{-2x}+x^2-1$	0,25	2,25	0,2
8	$(e+x)\sin\left(\pi\sqrt{x-1}\right)$	1,8	2,8	0,1
9	$\sqrt{3+2x}\cdot tg\dfrac{\pi x^3}{2}$	0,1	0,9	0,08

10	$\sqrt{2+3x}\cdot\ln\left(1+3x^2\right)$	-0,1	0,9	0,1
11	$\sqrt[3]{x^2+3}\cdot\cos\frac{\pi x}{2}$	1	2,5	0,15
12	$(4+7x)\sin\left(\pi\sqrt[3]{1+x}\right)$	0	7	0,7
13	$e^{-x^2}\left(1+3x-x^2\right)$	0	2	0,2
14	$x^3-3x+\frac{8}{\sqrt{1+x^2}}$	0	1,7	0,17
15	$\sqrt{sh\sqrt{2\pi x}}$, $\left(sh\,x=\frac{e^x-e^{-x}}{2}\right)$	0	1,2	0,12
16	$\sqrt{ch\frac{x}{\sqrt{2\pi}}}$, $\left(ch\,x=\frac{e^x+e^{-x}}{2}\right)$	0,5	1,5	0,1
17	$\frac{x^3+2x}{\sqrt{1+e^x}}$	-0,2	0,8	0,1
18	$\sqrt{1+2x^2}\cdot\sin\frac{3x}{2}$	2	4	0,2
19	$\sqrt{3x^2+5}\cdot\cos\frac{\pi x}{2}$	0,5	1,5	0,1
20	$\arccos e^{-\sqrt[3]{3x}}$	0,2	0,5	0,03
21	$\arcsin e^{-x^2/5}$	8	13	0,5
22	$x+\ln\left(x+\sqrt{1+x^2}\right)$	-0,5	0,5	0,1
23	$\frac{1+e^{-x/2}}{\sqrt{3x^2+1}}$	3	5	0,2
24	$3x^3+\frac{1}{x}+e^{-2x^2}$	1,2	2,2	0,1
25	$x^{2x+1}+x^3-2x$	1	5	0,4

Функции прикладной численной математики

Задание 6

1. Введите произвольную матрицу размером (4*6). Найдите сумму наибольших элементов ее строк.

2. Введите квадратную матрицу (5*5) с одним наименьшим элементом. Найдите сумму элементов строки, в которой размещен элемент с наименьшим значением.

3. Введите матрицу (6*9), в которой есть единственные наибольший и наименьшие элементы и они расположены в разных строках. Поменяйте местами строку с наибольшим элементом и строку с наименьшим элементом.

4. Введите матрицу (5*6) с разными значениями элементов. В каждой строке выберите элемент с наименьшим значением, из полученных чисел выберите наибольшее. Найдите индексы полученных элементов.

5. Введите матрицу (5*6). Найдите вектор, элементами которого являются наибольшие элементы соответствующей строки матрицы.

6. Введите матрицу (5*6). Постройте вектор, элементами которого являются суммы наибольшего и наименьшего элементов соответствующей строки матрицы.

7. Введите матрицу (5*6). Постройте вектор, элементами которого являются средние значения элементов соответствующей строки матрицы.

8. Введите матрицу (5*6). Постройте вектор, элементами которого являются среднеквадратичные отклонения элементов соответствующей строки матрицы от их среднего значения.

9. Введите матрицу (5*6). Постройте вектор, элементами которого являются средние арифметические наибольшего и наименьшего элементов соответствующей строки матрицы.

10. Введите матрицу (6*5). Постройте вектор, элементами которого являются суммы квадратов элементов соответствующего столбца матрицы.

11. Введите матрицу (5*5). Постройте векторы, элементами которых являются суммы элементов столбцов матрицы, произведения элементов столбцов и наименьшие значения элементов столбцов.

12. Введите матрицу (5*6). Найдите среднее арифметическое наибольших и наименьших ее элементов.

13. Введите матрицу (5*5). Постройте вектор, элементами которого являются элементы главной диагонали матрицы. Найдите след матрицы.

14. Введите две матрицы (4*4). Постройте новую матрицу размером (4*8), включая в первые 4 столбца строки первой матрицы, а в другие – столбцы второй матрицы.

15. Найдите сумму всех элементов матрицы размером (4*3).

Задание 7

Вычислите векторы:

а) модуля частотной передаточной функции (ЧПФ);

б) аргумента ЧПФ;

в) действительной части ЧПФ;

г) мнимой части ЧПФ по заданным числителю и знаменателю передаточной функции (табл).

Предварительно найдите корни знаменателя передаточной функции, определите наибольшую собственную частоту ω_{max} системы. Обеспечьте вычисление ЧПФ при 100 значениях частоты ω в диапазоне от 0 до $5\omega_{max}$.

Ва-ри-ант	Числитель	Знаменатель
1	$1.82p+67.56$	$p^4+2.65p^3+3.09p^2+7.04p+34.05$
2	$4.61p^2+1.82p+67.56$	$p^4+3.65p^3+45p^2+7.04p+125$
3	$p^2+4p+23$	$p^4+2p^3+39p^2+2p+45$
4	$3p^2+1.82p+67.56$	$p^2+7.04p+34.05$
5	$p+6$	$p^2+0.7p+48$
6	$p^3+4.61p^2+1.82p$	$2.65p^3+3p^2+4p+87$
7	$p^3+4.61p^2+1.82p+67.56$	$p^4+2.65p^3+68p^2+5p+34$
8	$4.61p^2+68$	$p^4+2.65p^3+3.09p^2+7.04p+34.05$
9	7.56	$p^4+2.65p^3+3.09p^2+7.04p+34.05$
10	$p^3+1.8p+7$	$p^4+6.5p^3+39p^2+7p+45$
11	$p^3+4.61p^2+1.82p+67.56$	$p^3+3.09p^2+70p+34$
12	$p^2+1.8p+78$	$2.65p^3+3.09p^2+7.04p+34.05$
13	$p^3+1.82p+67.56$	$p^4+2.6p^3+3p^2+4p+34$
14	$p^3+4.61p^2+1.82p+67.56$	$7p^2+7p+34$
15	$4.61p^2+1.82p+67.56$	$p^2+7.04p+560$
16	$1.82p+67.56$	$3.09p^2+7.04p+34.05$
17	p^3	$3.09p^2+7.8p+125$
18	$1.82p$	$p^3+3.09p^2+7.04p+34.05$
19	$4.61p^2$	$p^2+7.04p+34.05$
20	$p^3+67.56$	$p^4+2.65p^3+3.09p^2+7.04p+34.05$
21	p^3	$p^4+2p^3+3p^2+12p+100$
22	$p^3+4.61p^{2+}1.82p+67.56$	$p^4+5p^3+30p^2+7p+305$
23	$p^2+1.82p+67.56$	$p^4+2p^3+9p^2+4p+35$
24	$p^{3+}61p^2+182p+67$	$p^4+3p^3+9p^2+0.04p+39$
25	$p^2+1.82p+67.56$	$p^4+5p^3+20p^2+7p+34$

Указание. Частотной передаточной функцией называют передаточную функцию системы при мнимых значениях аргумента ($p= j \cdot \omega$).

Собственные частоты системы - это значения модулей мнимых частей корней характеристического уравнения системы (которое получается приравниванием нулю знаменателя передаточной функции).

Задание 8

Введите произвольную матрицу размером (5*5). Найдите для этой матрицы все нижеперечисленные характеристики:

1) определитель матрицы; в случае, если определитель окажется равным нулю, или слишком малым, измените некоторые элементы матрицы и повторите вычисления;

2) обратную матрицу; проверьте правильность путем обращения обратной матрицы;

3) характеристический полином матрицы;

4) корни характеристического полинома матрицы; рассортируйте корни по комплексно-спряженным парам и в порядке возрастания величин;

5) собственные значения матрицы; сравните с ранее найденными корнями характеристического полинома;

6) LU-разложение матрицы; проверьте его правильность;

7) QR-разложение матрицы; проверьте его правильность;

8) сингулярные числа матрицы; сравните их с получаемыми при ***svd***-разложении;
9) след матрицы;
10) число обусловленности матрицы;
11) экспоненту от матрицы;
12) логарифм от экспоненты матрицы; сравните с исходной матрицей.

Построение простейших графиков

Задание 9

Постройте в графическом окне MatLAB график функции из *задания 5.*

Задание 10

Постройте в графическом окне MatLAB графики амплитудно-частотной (модуля ЧПФ) и фазочастотной (аргумента ЧПФ) характеристик функции из *задания 7.*

Операторы управления вычислительным процессом

Задание 11

1. В соответствии с таблицей ниже выполнить:

- вычисление точных (используя стандартные функции MatLAB) значений соответствующей функции в диапазоне изменения аргумента от x1 до x2 в m равноотстоящих точках этого диапазона, включая его границы;

- вычисление по указанным степенным рядам приближенных значений функции в тех же точках, ограничиваясь r первыми членами ряда;

- расчет погрешности приближенного определения функции в каждой точке, сравнивая приближенное значение с точным, и построение графика зависимости погрешности от аргумента;

- вычисление приближенных значений функции в тех же точках с относительной погрешностью не более $\varepsilon = 0.001$; построение графика полученных относительных погрешностей.

Вар.	x1	x2	m	r	f(x)	Приближенная формула
1	0.2	5	20	4	sin(x)	$\sum(-1)^k \frac{x^{2k-1}}{(2k-1)!}$
2	1	10	30	5	cos(x)	$1-\sum(-1)^k \frac{x^{2k}}{(2k)!}$
3	0.3	3	40	5	exp(x)	$1+\sum\frac{x^k}{k!}$
4	0.4	4	50	4	ln(1+x)	$\sum(-1)^{k-1}\frac{x^k}{k}$
5	0.5	5	30	3	ln(x)	$\sum\frac{a^k}{k}$, где $a=\frac{x-1}{x}$
6	0.6	6	40	4	ln(x)	$\sum(-1)^k\frac{a^k}{k}$, где $a=x-1$
7	0.7	7	50	5	ln(x)	$2\sum\frac{a^{2k-1}}{2k-1}$, где $a=\frac{x-1}{x+1}$
8	0.8	8	45	6	ln(x+a)	$\ln(a)+2\sum\frac{b^{2k-1}}{2k-1}$, где $b=\frac{x}{2a+x}$
9	1.1	11	40	3	ctg(x)	$\frac{1}{x}+2x\sum\frac{1}{x^2-k^2\pi^2}$
10	1.2	12	50	4	cosec(x)	$\sum\frac{1}{(x-k\pi)^2}$
11	1.3	13	50	5	cosec(x)	$\frac{1}{x}+2x\sum\frac{(-1)^k}{x^2+k^2\pi^2}$
12	1.4	14	60	6	arctg(x)	$\sum(-1)^{k-1}\frac{x^{2k-1}}{2k-1}$
13	1.5	15	45	5	arctg(x)	$\frac{\pi}{2}+\sum(-1)^k\frac{1}{(2k-1)x^{2k-1}}$
14	1.6	16	40	4	ln(x)	$\sum(-1)^{k-1}\frac{a^k}{r}$, где $a=x-1$
15	0.9	9	50	6	sin(x)	$x\prod(1-\frac{x^2}{(k\pi)^2})$
16	1	10	50	4	cos(x)	$\prod(1-\frac{x^2}{(2k-1)^2\pi^2})$
17	0.6	5	50	3	ln(x)	$\sum\frac{a^k}{k}$, где $a=(x-1)/x$

18	-0. 9	0.9	45	4	arcctg(x)	$\frac{\pi}{2} - \sum (-1)^k \frac{x^{2k-1}}{(2k-1)}$
19	1	20	50	5	sh(x)	$\sum \frac{x^{2k-1}}{(2k-1)!}$
20	1	20	50	5	ch(x)	$1 + \sum \frac{x^{2k}}{(2k)!}$
21	-0. 9	0.9	50	5	arcth(x)	$\sum \frac{x^{2k-1}}{2k-1}$
22	1	20	50	5	arcth(x)	$\sum \frac{1}{(2k-1)x^{2k-1}}$
23	-0. 8	0.8	50	4	arcsin(x0	$x + \sum \frac{1 \cdot 3 \cdot 5 ... (2k-1) \cdot x^{2k+1}}{2 \cdot 4 \cdot 6 .... (2k) \cdot (2k-1)}$
24	-0. 8	0.8	50	4	arccos(x 0	$\frac{\pi}{2} - \{ \ x + \sum \frac{1 \cdot 3 \cdot 5 ... (2k-1) \cdot x^{2k+1}}{2 \cdot 4 \cdot 6 .... (2k) \cdot (2k-1)} \ \}$
25	-5	5	50	6	exp(x)	$1 + \sum \frac{x^k}{k!}$

Задание 12

Вычислить значения функции из *задания 5* при значениях аргумента в диапазоне от 0.1 до 100, образующих *геометрическую прогрессию* со знаменателем, равным квадратному корню из 10, и выведите в командное окно таблицу результатов вычислений.

Задание 13

Вычислить значения модуля ЧПФ и ее аргумента (в градусах) из *задачи 7* при значениях аргумента в диапазоне от 0.1 до 100, образующих *геометрическую прогрессию* со знаменателем, равным квадратному корню из 10, и выведите в командное окно таблицу результатов вычислений.

Создание простейших файл-функций (процедур)

Задание 14

Создайте М-файл, вычисляющий значение функции из *задания 5*. Постройте график этой функции с помощью процедуры ***fplot*** в границах, заданных в задании 5. Вычислите интеграл от функции в тех же пределах, используя процедуры ***quad*** и ***quad8***. Найдите точку локального минимума и локальный минимум функции и ближайший корень (нуль).

Задание 15

Найдите точку локального минимума и локальный минимум функции двух переменных, приняв за начальную точку с заданными координатами (таблица ниже). Предварительно создайте соответствующую файл-функцию.

Вариант	x_o	y_o	$f(x,y)$
1	0	1	$e^{x+y}+(x-y)^2-2x-2y$
2	0.7	-1. 2	$(x-y)^2-\cos(x-y-1)$
3	1.5	-0. 5	$e^{x+y}-2x-2y-\cos(x-y-1)$
4	0.5	1.5	$e^{x+y}+4x^2-3x-3y$
5	0	1	$4x^2+\ln(x+y)+\dfrac{1}{x+y}$
6	1.2	0.7	$2^{x+y}-2x-2y+2(x-y)^2$
7	0	-0. 9	$e^{x-y}+2x+2y+(x+y)^2$
8	0.8	1.3	$(x-y)^2-\cos(x+y-1)$
9	1.5	0.5	$e^{x-y}-2x+2y-\cos(x+y-1)$
10	0.5	-1. 5	$e^{x-y}-3x+3y+4x^2$
11	0	-1	$4x^2+\ln(x-y)+\dfrac{1}{x-y}$
12	1.2	-0. 8	$2^{x-y}-2x+2y+2(x+y)^2$

Создание функций от функций

Задание 16

Создайте функцию (func1), которая принимает в качестве аргумента название другой функции (func2) и значения аргументов, передаваемые в функцию func2. Проделайте несколько вызовов функции func1. Реализуйте механизм передачи параметров в функцию func2 через глобальные переменные. Варианты функций func2, аргументов и параметров приведены в таблице.

Вариант	Функции func2	Аргументы	Параметры
1	y=a+x1*x2*x3	x1,x2,x3	a
2	y=a*x1+exp(-t*20)	x1,t	a
3	y=a^x1+2^x2	x1,x2	a
4	y=(t-t0)^2+3	t	t0
5	y=10*x^2+z^3+b	x,y	b
6	y=ln(1-x1*x2)*b	x1,x2	b
7	y=(x-z)^2+(x-z)+d	x,z	d
8	errordlg	Текст сообщения	Текст заголовка
9	sin(a*x), cos(b*x), tan(a*x)	x	a
10	exp(x1*x2), sin(x1*x2), a * ln(x1)	x1,x2	a
11	y=c*a^t, y = c*a*sin(t)	t	a,c
12	y=20*ln(sin(t))+d; y=1-exp(-d*t)	t	d
13	ode45,ode23	Промежуток интегрирования, начальные условия	Точность интегрирования
14	msgbox	Текст сообщения	Текст заголовка
15	questdlg	Текст сообщения	Текст заголовка

16	warndlg	Текст сообщения	Текст заголовка
17	uigetfile	Папка по умолчанию	Текст заголовка
18	uiputfile	Папка по умолчанию	Текст заголовка
19	uisetcolor	Цвет по умолчанию	Текст заголовка
20	uisetfont	Шрифт по умолчанию	Текст заголовка
21	menu	Набор пунктов меню	Заголовок меню
22	listdlg		
23	ode45,ode23	Промежуток интегрирования, начальные условия	Точность интегрирования
24	sin(a*x), cos(b*x), tan(a*x)	x	a
25	exp(x1*x2), sin(x1*x2), a * ln(x1)	x1,x2	a
27	uigetfile, uiputfile	Папка по умолчанию	Текст заголовка
28	errordlg, warndlg	Текст сообщения	Текст заголовка

Расчетное задание 1

Идентификация сообщений, передаваемых по зашумленному каналу связи.

Исходные данные

По каналу связи передаются буквы $[x_1;x_2;...;x_n]$ в двоичном коде. Последовательность переданных букв образует сообщение. Канал симметричный, вероятность искажения каждого отдельного символа (бита) равна q. В результате однократной передачи сообщения $X=[x^{(1)},x^{(2)}...,x^{(k)}]$ на приемной стороне принято сообщение $Y_1=[y_1^{(1)}y_1^{(2)}...y_1^{(k)}]$. В результате повторной передачи того же слова на приемной стороне принято слово $Y_2=[y_2^{(1)}y_2^{(2)}...y_2^{(k)}]$. В результате последней (m-й) передачи того же слова на приемной стороне принято слово $Y_m=[y_m^{(1)}y_m^{(2)}...y_m^{(k)}]$.

Варианты

Передаваемые буквы (алфавит) и их код приведен в табл.1 приложения. Для каждого студента есть свой вариант в виде текстового файла "\Tasks\Task_1.txt", находящегося в личной папке файлового хранилища. В файле задается число букв в сообщении, разрядность кода (количество бит, используемых при передаче одной буквы), шум (вероятность искажения q), число посылок m и набор из m посылок (принятых сообщений $Y^{(i)}$).

Задание

Часть 1. Последовательная передача одинаковых сообщений

1.1. Определение переданного сообщения

- вычислите априорное распределение вероятностей исходных букв алфавита p(xi), рассмотрите два случая (все дальнейшие расчеты в п. 1.1 и 1.2 необходимо будет проделать для этих двух вариантов):
    - o все символы равновероятны;
    - o вероятности букв задаются исходя из известной информации о частоте букв в русском алфавите (таблица 2);
- вычислите апостериорное распределение вероятностей после 1-й, 2-й и m-й передач для каждой s буквы сообщения - $P(x_i / y_1^{(s)})$, $P(x_i / y_1^{(s)} y_2^{(s)})$, $P(x_i / y_1^{(s)} y_2^{(s)} y_3^{(s)})$, … , $P(x_i / y_1^{(s)} y_2^{(s)} ..., y_m^{(s)})$; при расчете используются формулы (3), (4) и (6); следует учитывать, что для повторных посылок априорные вероятности будут совпадать с апостериорными для предыдущей посылки (см. формулу (6)).
- постройте график изменения апостериорного распределения вероятностей на примере любой l-ой передаваемой буквы сообщения (n передач => n графиков друг под другом, на графике по оси X – номер символа, по оси Y – вероятность)
- по максимуму апостериорной вероятности определите наиболее вероятные буквы и составьте вариант исходного переданного сообщения для 1-й, 2-й и m-й посылок;
- проанализируйте, как повторные передачи сказались на принятии решения.

1.2. Расчет энтропии и количества информации

- Выберите в посылаемом сообщении произвольную букву (под номером s), далее все вычисления будут относиться к этой букве;
- Определите апостериорные вероятности, рассматривая каждую передачу независимо от другой; схема вычислений следующая $P(x_i) \to P(y_j / x_i) \to P(y_j) \to P(x_i / y_j)$; при расчете используйте формулы (3), (4) и (5).
- Определите условные энтропии $H(X / y_j)$ на сообщения y_j по формуле (2), среднее количество информации $I(X, y_j)$ об X, содержащееся в y_j по формуле (8).
- Определите среднюю условную энтропию H(X/Y) по формуле (7) и среднюю взаимную информацию I(X,Y) по формуле (9).
- Постройте графики изменения условной энтропии $H(X / y_j)$ и количества информации $I(X, y_j)$ от номера посылки.

1.3. Сравните результаты п. 1.1 и 1.2 при различных заданиях изначальных априорных вероятностей.

Часть 2 Передача сообщения путем многократного дублирования

Рассмотрите m передач сообщений как передачу одного большого сообщения, в котором каждый символ многократно (m-кратно) дублируется

На входе $X_{new} = [x_{new}^{(1)}; x_{new}^{(2)}; ..., x_{new}^{(k)}] = [x^{(1)} x^{(1)} ... x^{(1)}, x^{(2)} x^{(2)} ... x^{(2)} ..., x^{(k)} x^{(k)} ... x^{(k)}]$,

На выходе $Y_{new} = [y_{new}^{(1)}; y_{new}^{(2)}; ...; y_{new}^{(k)}] = [y_1^{(1)} y_2^{(1)} ... y_m^{(1)} y_1^{(2)} y_2^{(2)} ... y_m^{(2)} ... y_1^{(k)} y_2^{(k)} ... y_m^{(k)}]$.

При этом новый алфавит по сути – m-кратное дублирование старого алфавита: $[x_{new1}; x_{new2}; ...; x_{newn}] = [x_1 x_1 ... x_1; x_2 x_2 ... x_2; ...; x_n x_n ... x_n]$

2.1. Определение переданного сообщения

- вычислите априорное распределение вероятностей исходных букв алфавита p(xi) – рассмотрите два случая (по аналогии с п.1. все дальнейшие расчеты в п. 2.1 и 2.2 необходимо выполнить для этих двух вариантов):
  - все символы равновероятны;
  - вероятности букв задаются исходя из известной информации о частоте букв в русском алфавите (таблица 2);
- вычислите апостериорное распределение вероятностей для каждой l буквы сообщения - $P(x_{newi} / y_{new}^{(l)})$; при расчете используются формулы (3), (4);
- постройте график апостериорного распределения вероятностей на примере l-ой передаваемой буквы сообщения
- по максимуму апостериорной вероятности определите наиболее вероятные буквы и составьте вариант исходного переданного сообщения – сравните его со случаем передачи сообщений последовательно

2.2. Расчет энтропии и количества информации

- Выберите в посылаемом сообщении ту же букву, что и использовалась в п. 1.2, далее все вычисления будут относиться к этой букве;
- Определите апостериорные вероятности; схема вычислений следующая $P(x_{new_i}) \to P(y_{new} / x_{new\,i}) \to P(y_{new}) \to P(x_{newi} / y_{new})$; при расчете используйте формулы (3), (4) и (5).
- Определите условную энтропию $H(X_{new} / y_{new})$ на сообщения y_{new} по формуле (2), среднее количество информации $I(X, y_{new})$ об X, содержащееся в y_{new} по формуле (8).
- Определите среднюю условную энтропию $H(X_{new} / Y_{new})$ по формуле (7) и среднюю взаимную информацию $I(X, Y_{new})$ по формуле (9).
- Сравните результаты (энтропия, количество информации) с п.1.2 и объясните их.

Приложение 1 Теоретические основы

Пусть X - ансамбль возможных сообщений, которые могут быть переданы по каналам связи. Априорные вероятности, с которыми на передающей стороне может появиться **i** -ое сообщение, обозначим через $p_i = p(x_i)$.

В качестве характеристики априорной неопределенности генерации того или иного сообщения на передающей стороне Шеннон предложил применить следующий функционал энтропии:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \qquad (1)$$

В этом функционале используется логарифм по основанию **2** в связи с тем, что в цифровых каналах связи информация обычно представляется в двоичных кодах, и энтропия измеряется в *битах*.

Энтропия является характеристикой неопределенности состояния ансамбля X в отличие от характеристики возможности наступления того или иного события, то есть вероятности. Энтропия не зависит от значений, которые может принимать тот или иной элемент ансамбля и от способа его представления. В частности, элементами ансамбля могут быть словесные описания или фотографические изображения. Энтропия характеризует степень хаотичности ансамбля и принимает максимальное значение, когда вероятности $p(x_i)$ одинаковы.

Инструментарий теории информации, выдвинутой К.Шенноном, эффективно используется в теории и практике передачи информации и кодирования. В этих применениях ансамбль X - ансамбль возможных сообщений. Основные результаты теории информации относятся к передаче информации в двоичном коде и к каналам, ориентированным на передачу двоичных кодов. Передача такой информации осуществляется двоичными кодовыми словами, и при каждой такой передаче искажение отдельного символа в передаваемом слове заключается в том, что вместо **1** получатель принимает **0**, или вместо нуля получатель принимает **1**. При таких условиях канал передачи двоичной информации характеризуется вероятностью искажения символа. Схематическое представление простейшего канала связи, а именно , двоичного симметричного канала представлено на рис 5.

Симметричным этот канал называется потому, что вероятности искажения символов **'0'** и **'1'** одинаковы и равны **q**, как это показано на рисунке. Вероятность **p** - вероятность неискаженной передачи символа.

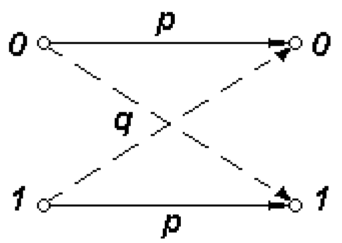


Рис.5. Двоичный симметричный канал

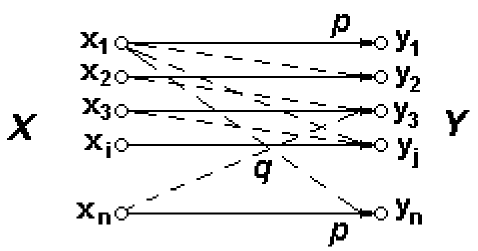


Рис.6. Передача сообщений по каналу с искажениями

Поскольку каждое сообщение $\mathbf{x}_i \in X$ передается кодовым словом, которое может состоять из нескольких символов **'0'** или **'1'**, возможностей искажения слова при передаче по такому каналу гораздо больше (см. рис. 6, где принимаемые сообщения обозначены через $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_j, \ldots \in Y$). Вероятность искажения слова в общем случае отличается от вероятности искажения отдельного символа.

Понятно, что при передаче по каналам связи сообщения (слова) $\mathbf{x}_i \in X$ и при получении сообщения $\mathbf{y}_j \in \boldsymbol{Y}$, несмотря на искажения в каналах энтропия ансамбля $\boldsymbol{X}$, несомненно, уменьшится и будет исчисляться условной энтропией на сообщение

$$H(X/y_j) = -\sum_{i=1}^{n} p(x_i/y_j) \cdot \log_2 p(x_i/y_j). \tag{2}$$

Прим. Вероятности $p(x_i/y_j)$ называются апостериорными. Для их расчета необходимо использовать формулу Байеса

$$p(x_i/y_j) = \frac{p(y_j/x_i)p(x_i)}{p(y_j)} = \frac{p(y_j/x_i)p(x_i)}{\sum_k p(y_j/x_k)p(x_k)} \tag{3}$$

$p(x_i)$ - априорные вероятности, вначале они могут быть выбраны произвольно и как правило они принимаются одинаковыми и равными 1/N, где N – общее количество слов xi.

$p(y_j/x_i)$ - условная вероятность приема y_j при условии, что было послано x_i. Данную вероятность легко найти следующим образом. Если общее количество разрядов в слове k, в t разрядах произошла ошибка (они проинвертировались), то вероятность очевидно равна

$$p(y_j/x_i) = p^{k-t}q^t \tag{4}$$

Здесь по-прежнему p – вероятность правильной передачи одного бита, q=1-p – вероятность ошибки.

$$p(y_j) = \sum_k p(y_j/x_k)p(x_k) \tag{5}$$

- вероятность приема y_j в данной задаче несет вспомогательный характер – она потребуется в том числе в дальнейших расчетах.

В случае, если при приеме одной и той же буквы несколько раз информация накапливается, для расчета апостериорных вероятностей последующих приемов в качестве априорных вероятностей необходимо использовать апостериорные вероятности после предыдущего принятого сообщения, т.е.:

$$p(x_i/y_1y_2...y_j) = \frac{p(y_j/x_i)p(x_i/y_1y_2...y_{j-1})}{p(y_j)} = \frac{p(y_j/x_i)p(x_i/y_1y_2...y_{j-1})}{\sum_k p(y_j/x_k)p(x_k/y_1y_2...y_{j-1})} \tag{6}$$

Для того, чтобы характеризовать всю систему приема - передачи, используют среднюю условную энтропию:

$$H(X/Y) = \sum_{j=1}^{n} p(y_j) \cdot H(X/y_j) = H(X,Y) - H(Y) \tag{7}$$

где $H(Y) = \sum_{j=1}^{n} p(y_j) \cdot \log_2 p(y_j)$, $H(X,Y) = -\sum_{j=1}^{n}\sum_{i=1}^{n} p(x_i,y_j) \cdot \log_2 p(x_i,y_j)$

$H(X,Y)$ - совместная энтропия двух ансамблей, равная

$H(X,Y) = H(X/Y) + H(Y) = H(Y/X) + H(X)$

Количество информации об $\mathbf{x}_i$, полученное в одном сообщении $\mathbf{y}_j$:

$$I(x_i : y_j) = \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}.$$

Среднее количество информации об $\boldsymbol{X}$, полученное в сообщении $\mathbf{y}_j$

$$I(X : y_j) = -\sum_{i=1}^{n} p(x_i/y_j) \log p(x_i) - H(X/y_j) \tag{8}$$

Количество информации измеряется в тех же единицах, что и энтропия - в битах. Средняя взаимная информация, содержащаяся в ***Y*** об ***X*** или в ***X*** об ***Y***:

$$I(X:Y)=\sum_{i=1}^{n} p(y_j) I(X:y_j) = H(X) - H(X/Y) \qquad (9)$$

то есть это количество информации численно равно количеству неопределенности, устраненной при получении одного сообщения.

Приложение 2 Вспомогательные материалы и таблицы

Используемый набор букв и символов:
'0123456789АБВГДЕЁЖЗИЙКЛМНОПРСТУФХЦЧШЩЬЫЪЭЮЯабвгдеёжзийклмнопрстуфхцчшщьы ъэюя.,!:?-_№() '

Каждая буква кодируется своим номером в наборе, представленным в двоичном коде (см. табл.). При передаче без ошибки буквы с кодом x7x6x5x4x3x2x1x0 на принимающей стороне будет получен такой же код x7x6x5x4x3x2x1x0.

Табл. 1 Символы и их коды

Символ	Код	Символ	Код	Символ	Код	Символ	Код	Символ	Код
0	0000000	Й	0010100	Э	0101000	р	0111100	?	1010000
1	0000001	К	0010101	Ю	0101001	с	0111101	-	1010001
2	0000010	Л	0010110	Я	0101010	т	0111110	_	1010010
3	0000011	М	0010111	а	0101011	у	0111111	№	1010011
4	0000100	Н	0011000	б	0101100	ф	1000000	(	1010100
5	0000101	О	0011001	в	0101101	х	1000001	)	1010101
6	0000110	П	0011010	г	0101110	ц	1000010	Пробел	1010110
7	0000111	Р	0011011	д	0101111	ч	1000011		
8	0001000	С	0011100	е	0110000	ш	1000100		
9	0001001	Т	0011101	ё	0110001	щ	1000101		
А	0001010	У	0011110	ж	0110010	ь	1000110		
Б	0001011	Ф	0011111	з	0110011	ы	1000111		
В	0001100	Х	0100000	и	0110100	ъ	1001000		
Г	0001101	Ц	0100001	й	0110101	э	1001001		
Д	0001110	Ч	0100010	к	0110110	ю	1001010		
Е	0001111	Ш	0100011	л	0110111	я	1001011		
Ё	0010000	Щ	0100100	м	0111000	.	1001100		
Ж	0010001	Ь	0100101	н	0111001	,	1001101		
З	0010010	Ы	0100110	о	0111010	!	1001110		
И	0010011	Ъ	0100111	п	0111011	:	1001111		

Табл. 2 Частота букв в русском языке

Буква	Частота	Буква	Частота	Буква	Частота
а	8.66	л	4.32	ц	0.52
б	1.51	м	3.29	ч	1.27
в	4.19	н	6.35	ш	0.77
г	1.41	о	9.28	щ	0.49
д	2.56	п	3.35	ъ	0.04
е	8.10	р	5.53	ы	2.11
ж	0.78	с	5.45	ь	1.90
з	1.81	т	6.30	э	0.17
и	7.45	у	2.90	ю	1.03
й	1.31	ф	0.40	я	2.22
к	3.47	х	0.92		

Расчетное задание 2

Статистическая обработка случайных последовательностей. Идентификация законов распределения.

Исходные данные

В результате измерений получена выборка x1, x2, …, xN из генеральной совокупности с неизвестным законом распределения. Выборочные значения расположены в файлах (для каждой группы свой каталог, для каждого варианта файл с названием Distribtuion i), где i – номер варианта (по номеру в списке преподавателя).

Варианты

Для каждого студента есть свой вариант в виде текстового файла "\Tasks\Task_2.txt", находящегося в личной папке файлового хранилища. В этом файле задано число значений N, а также сам массив выборочных значений, отделенных друг от друга пробелами. В случае дискретного распределения значения целые, в случае непрерывного – вещественные.

Справочная информация

Вся теоретическая часть по работе изложена в [1], а также в разделах помощи Matlab, в частности Statistic Toolbox.

В приложении 1 к данному заданию описаны основные распределения, даны формулы плотностей, функций, моментов и имеющихся аналитических оценок параметров по методу максимального правдоподобия. Также в приложении 1 представлены графики плотностей и функций основных распределений. Ими разумно пользоваться при подборе распределения под имеющуюся выборку путем сравнения графиков:

- относительной гистограммы и теоретической плотности распределения;
- эмпирической и теоретической функций распределения.

В приложении 2 приведены примеры вычисления оценок параметров распределений при подгонке параметров распределений к имеющейся выборке. Поэтому при использовании метода моментов и максимального правдоподобия целесообразно ознакомиться и разобраться в этих примерах.

В приложении 3 приведены теоретические основы трех основных статистических методов проверки гипотез о виде плотности распределения: хи-квадрат, Колмогорова-Смирнова и Мизеса.

Задание

1. Статистическая обработка случайных последовательностей

1.1. Считать выборку X из файла. Создать на ее основе 10 подвыборок – для этого перемешать выборку (например, командой Xperm=X(randperm(length(X))) и последовательно сформировать подвыборки (Xpodv(i) = Xperm(1+(i-1)*N/10:i*N/10))

1.2. Построить выборочную функцию распределения F(x) (она должна быть ступенчатой!!!, можно воспользоваться функцией cdfplot)

1.2. Построить абсолютную и относительную гистограммы на разных графиках (функция hist строит абсолютную гистограмму; чтобы построить относительную гистограмму выборки, нужно разделить все ее значения на ее объем). Внимательно отнеситесь к выбору количества (ширины) интервалов или столбцов - оно выбирается таким образом, чтобы самый "бедный" интервал содержал **3 ÷ 5** выборочных значений. Если у распределения есть тяжелые хвосты (несколько значений в области значений, очень далеко отстоящей от скопления основной массы данных), то желательно их отбросить. Например, если 99.9 % значений, находящихся в диапазоне [-20 20] и 0.1 % значений, находящихся в диапазоне [-20000 20000], то последние 0.1 % не позволят нормально построить гистограмму и их желательно просто не учитывать при построении гистограммы (НО помнить, что они есть и характеризуют распределение как имеющее тяжелый хвост – Коши, Парето к примеру).

1.3. Определить точечные оценки:

1.3.1. моментов

- первого начального (среднее арифметическое - mean, медиана - median, средина размаха (min + max)/2)
- центральных моментов: второго-дисперсии (var), третьего, четвертого (moment) по выборочной функции распределения

Для оценки первого начального момента использовать среднее арифметическое, выборочную медиану, средину размаха. Определить моду (максимум на графике плотности).

1.3.2. асимметрии и эксцесса (функции skewness, kurtosis);

1.3.3. границ интерквантильного промежутка Jp для P=0.95 только по полной выборке (функция quantile)

1.3.4. характеристики по пп. 1.3.1-1.3.2 по подвыборкам, сформированным в п. 1.1.

Результаты представить в таблице следующей формы.

	$\bar{x}$	x_{med}	$x_{ср}$	s^2	s	$\mathring{m}_3$	$\mathring{m}_4$	As	Ex
N									
N/10									
N/10									
…									
N/10									

Представить эти же результаты графически точками на осях с указанием масштаба на этих осях по форме:

Прим. 1 Для проверки правильности результатов нужно убедиться в близости характеристик, посчитанных по полной выборке с характеристиками, посчитанными по подвыборкам.

Прим. 2. Значения характеристик по подвыборкам не должны равномерно располагаться вокруг значений характеристики по всей выборке – это свидетельствует о том, что подвыборки брались из отсортированной выборки, что в свою очередь является ошибкой.

оценки м.о.

\+ + + + + + ◆ + + + + $\bar{x}$ (◆ - оценки по **n** = n=N)

\+ + + + + ◆ + + + + + $\tilde{x}_{med}$ (+ - оценки по **n** = n=N/10)

\+ + + + ◆ + + + + + + x_{cp}

оценки дисперсии

\+ + + + + + ◆ + + + +

оценки третьего центрального момента

\+ + + + + + ◆ + + + +

оценки четвертого центрального момента

\+ + + + + + ◆ + + + +

оценки асимметрии

\+ + + + + + ◆ + + + +

оценки эксцесса

\+ + + + + + ◆ + + + +

1.4. Определить интервальные оценки с доверительной вероятностью Q=0.8:

- первого начального и второго центрального моментов (вычисления выполнить по полной выборке и по отдельным частям, как в п. 2.1.4 - по N/10 значений в каждой частичной выборке). Прим. Значения обратных функций распределения Стьюдента и Хи-квадрат удобно вычислять с помощью функций tinv и chi2inv соответственно. Нанести на эти характеристики соответствующие значения точечных оценок (для проверки правильности доверительный интервал должен располагаться вокруг точечной оценки).
- интерквантильного промежутка J для P=0.95:
  - по всей выборке с помощью непараметрических толерантных пределов, симметричных относительно среднего арифметического и относительно нуля. Прим. Количество статистически эквивалентных блоков k, отбрасываемых от выборки при нахождении непараметрических толерантных пределов, симметричных относительно среднего арифметического определяется из неравенства: $\sum_{m=n-k}^{n} C_n^m P^m (1-P)^{n-m} \le 1-Q$ (решение данной проблемы может быть выполнено последовательным увеличением k от 0 до тех пор, пока неравенство не начнет выполняться; следует учитывать, что число сочетаний C_n^m при больших n необходимо считать с применением формулы Стирлинга). Результирующий предел будет равен $\left[X_{k/2} X_{N-k/2}\right]$ при четном k или $\left[X_{(k-1)/2} X_{N-(k-1)/2}\right]$ при нечетном k. В случае если пределы симметричны относительно нуля, то необходимо преобразовать выборку, заменив отрицательные значения на их модуль и отбросить справа (k-1) эквивалентных блоков. Результирующий предел будет равен $\left[-X_{N-k+1} X_{N-k+1}\right]$.
  - по частичным выборкам с помощью параметрических толерантных пределов, считая закон распределения генеральной совокупности нормальным. Прим. Значения толерантных множителей можно найти в [1].

Результаты представить только графически аналогично тому, как описано выше – под графическим представлением соответствующей точечной оценки, предусмотрев для каждого варианта расчета отдельную ось. Графическое представление толерантных пределов — также на отдельных осях для каждого варианта. Все оси обозначить.

Сделать выводы относительно ширины доверительных интервалов. Сравнить:

а) интерквантильные промежутки с толерантными пределами б) параметрические и непараметрические толерантные пределы, симметричные относительно среднего арифметического и относительно нуля.

2. Идентификация закона и параметров распределения

В данном задании осуществляется идентификация закона распределения исходной выборки. Для этого вначале методом проб подбирается распределение, а затем различными способами определяются параметры этого распределения. В завершении осуществляется проверка гипотез о соответствии предполагаемых законов распределения экспериментальным данным с помощью ТРЕХ критериев: "хи-квадрат", Колмогорова-Смирнова, "омега-квадрат".

Подсказка возможные распределения:

Непрерывные – арксинус, треугольное, Коши, Симпсона, Лапласа, Хи-квадрат, экспоненциальное, нормальное, равномерное, Симпсона, Стьюдента, логнормальное, гамма, Рэлея, Парето.

2.1. Начальный выбор распределения

Для начальной ориентировки в выборе закона использовать вид гистограммы, функции распределения, соотношения между моментами и полученные значения **эксцесса и асимметрии**. Удобная утилита Matlab disttool позволяет построить графики многих (но не всех!) законов (плотностей) и функций распределения, варьируя и подбирая их параметры. В результате нужно определиться с тремя основными распределениями, которые и будут идентифицироваться.

2.2. Определение параметров теоретических распределений.

Для выбранных теоретических распределений необходимо определить точные значения параметров, наиболее подходящие для описания выборки. Это необходимо сделать двумя способами:

- с помощью метода моментов, когда теоретические моменты приравниваются к выборочным и решается система уравнений по числу неизвестных параметров распределения.

- с помощью метода максимального правдоподобия – в случае, если для распределения известны аналитические ММП-оценки, можно воспользоваться ими. В общем случае необходимо найти ММП-оценки численными методами. Для этого в Matlab уже написано множество fit-функций под большое число распределений (normfit и др). В случае, если распределения нет в Matlab, его можно задать в форме встроенной функции и воспользоваться командой mle, передав туда эту функцию и начальные приближения для значений параметров (можно воспользоваться оценками метода моментов). Есть замечательная утилита Matlab – dfittool, позволяющая производить идентификацию через удобный интерфейс. Для распределений, отсутствующих в Matlab, следует использовать функцию mle (см. Приложение 1 – примеры оценки неизвестных параметров).

Сравнить оценки, полученные методом моментов и ММП. Для этого построить

- эмпирическую и теоретические функцию распределения (на 1 графике)

- гистограмму и теоретические плотности распределения (на 1 графике)

Т.о. должно быть 6 графиков (3 распределения * 2 характеристики), причем на каждом из графиков должно быть по 3-**4 зависимости** (1-эмпирическая, 2-теоретическая для оценки параметров по методу моментов, 3 и 4-теоретическая для оценки параметров по методу ММП численно и если есть, то аналитически). По графикам оценить степень сходства эмпирических и теоретических характеристик. Написать, какой метод оценки параметров дает большую точность.

2.3. Произвести проверку гипотез относительно выбранных теоретических законов распределения и их параметров (по методу ММП и моментов). Проверку провести по трем критериям - "хи-квадрат", Колмогорова-Смирнова, "омега-квадрат". Критерии можно реализовать как вручную, так и воспользоваться функциями Matlab – chi2gof – критерий Хи-квадрат, kstest – критерий Колмогорова-Смирнова. Критерий Мизеса необходимо реализовать самим и воспользоваться таблицей из [1]. Сравнить полученные статистики критериев с критическими значениями. Выбрать наиболее подходящие распределения, исходя из значениев статистики критериев.

2.4. Привести итоговую таблицу, в которой для каждого из 3 распределений приведены по 3 вида оценок (метод моментов, ММП-аналитика, ММП-численный), и для каждого из уже 9 вариантов распределений и оценок – результаты проверки гипотез по 3 критериям – статистика критерия и критическое значение.

	Распределение 1			Распределение 2			Распределение 3		
Название									
Формула плотности									
	Мет.мом.	ММП-аналит	ММП-числ	Мет.мом.	ММП-аналит	ММП-числ	Мет.мом.	ММП-аналит	ММП-числ
Пар-р 1									
Пар-р 2									
Хи-квадрат - статистика									
Хи-квадрат – критич.знач									
Хи-Квадрат - вывод									
Колм.-Смирнова - статистика									
Колм.-Смирнова – крит.значение									
Колм.-Смирнова - вывод									
Мизеса – статистика									
Мизеса – критич.значение									
Мизеса - вывод									

Прим. 1. Вначале можно воспользоваться множеством критериев для нормального распределения – ttest, ztest, vartest.

Прим. 2. В отчете отобразить все ваши пробы относительно выбора подходящего закона распределения, а не одну последнюю (наиболее подходящую).

Приложение 1 Формулы, характеристики и графики плотностей основных распределений

Распределения дискретных СВ

Распределение	Плотность вероятности	Функция распределения	Числовые характеристики	Производящая функция моментов	Оценки по ММП
Биномиальное	$P_n(k) = C_n^k p^k (1-p)^{n-k}$ $P_n(k) \approx \frac{1}{\sqrt{npq}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$F_n(y) = \sum_{k=0}^{\lfloor y \rfloor} C_n^k p^k q^{n-k}$	$M(X) = np; D(X) = npq$ $As = \frac{1-2p}{\sqrt{npq}}; ex = \frac{1-6pq}{npq}$ $\mathrm{Mod} = \lfloor (n+1)p \rfloor;$	$M_X(\nu) = (pe^{\nu} + q)^n$	$p = \frac{\sum_{i=1}^{n} x_i}{mn}$
Пуассона	$P(k) = \lambda^k e^{-\lambda} / k!$	$F(k) = \frac{\Gamma(k+1, \lambda)}{\lambda!}$	$M(X) = \lambda; D(X) = \lambda; As = \lambda^{-0.5}$ $Ex = \lambda^{-1}; \mathrm{Mod} = \lfloor \lambda \rfloor$	$M_X(\nu) = \exp(\lambda(e^{\nu} - 1)$	$\lambda = \overline{x_{\hat{a}}}$
Геометрическое	$P(k) = q^k p$		$M(X) = q/p; D(X) = q/p^2$	$M_X(\nu) = p/(1-qe^{\nu})$	
Равномерное	$P(k) = \begin{cases} \frac{1}{n}, a \le k \le b \\ 0, else \end{cases}$	$F(k) = \begin{cases} 0, k < a \\ (k-a+1)/n, a \le k \le \\ 1, k > b \end{cases}$	$M(X) = (a+b)/2;$ $D(X) = (n^2-1)/12$ $As = 0$	$M_X(\nu) = \frac{e^{a\nu} - e^{(b+1)\nu}}{n(1-e^{\nu})}$	

Распределения непрерывных СВ

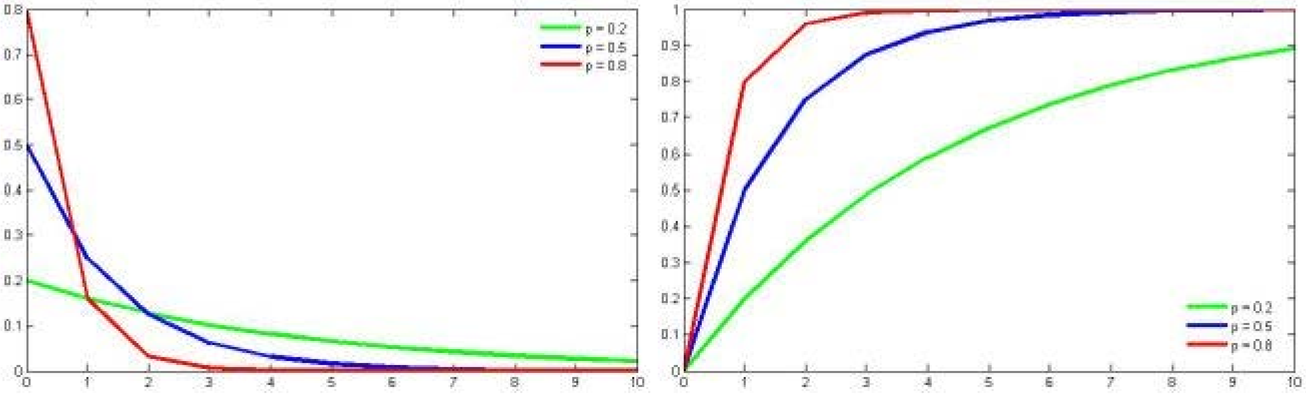
Распределение	Плотность вероятности	Функция распределения	Числовые характеристики	Производящая функция моментов (хар.функция)	Оценки по ММП
Нормальное	$p(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$	$F(x)=0.5+\Phi((x-a)/\sigma)$ $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x \exp(-t^2/2)dt$	$M(x)=a; D(x)=\sigma^2;$ $As=0, Ex=0$	$M_x(\nu)=\exp(a\nu+\frac{\sigma^2\nu^2}{2})$ $\phi_x(\nu)=\exp(ai\nu-\frac{\sigma^2\nu^2}{2})$	$a=\overline{x};$ $\sigma^2=\frac{\sum_{i=1}^{N}(x_i-\overline{x_в})^2}{N}$
Логнормальное	$p(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln(x)-a)^2}{2\sigma^2}\right)$	$F(x)=0.5+\Phi((\ln(x)-a)/\sigma)$ $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x \exp(-t^2/2)dt$	$M(X)=e^{a+\sigma^2/2}; Med=e^a;$ $D(X)=(e^{\sigma^2}-1)e^{2a+\sigma^2};$		
Коши	$p(x)=\frac{\Delta}{\pi(\Delta^2+(x-c)^2}$	$F(x)=\frac{1}{\pi}arctg(\frac{x-c}{\Delta})+0.5$ $F^{-1}(x)=c+\Delta tg(\pi(x-0.5))$	Med=Mod = c	$\phi_X(\nu)=\exp(ci\nu-\Delta\|\nu\|)$	
arcsin	$p(x)=\frac{1}{\pi\sqrt{a^2+(x-c)^2}}$	$F(x)=\begin{cases}0, x<c-a\\ \frac{1}{2}+\frac{1}{\pi}\arcsin\left(\frac{x-c}{a}\right),\\ 1, x>c+a\end{cases}$	$M(X)=Med=c;$ $D(X)=a^2/2;$ $As=0, Ex=1.5$		
Лапласа	$p(x)=\frac{\lambda}{2}\exp(-\lambda\|x-c\|)$		$M(X)=x_{0.5}=x_{\text{mod}}=c;$ $D(X)=2/\lambda^2;$ $As=0, Ex=6$	$\phi_\xi(\nu)=\frac{\lambda^2}{\lambda^2+\nu^2}e^{j\nu c}$	$\hat{c}=x_{med};$ $\hat{\lambda}=N(\sum_{i=1}^{N}\|xi-\hat{c}\|)^{-1}$
Показательное (экспоненциальное)	$p(x)=\begin{cases}\lambda e^{-\lambda x}, x>=0\\ 0, x<0\end{cases}$	$F(x)=\begin{cases}1-e^{-\lambda x}, x>=0\\ 0, x<0\end{cases}$	$\mu_1(x)=1/\lambda; D(\xi)=1/\lambda^2;$ $As=2, Ex=6, \text{Mod}=0,$ $\text{Med}=\ln(2)/\lambda$	$M_x(\nu)=\lambda/(\lambda-\nu)$ $\phi_x(\nu)=\lambda/(\lambda-i\nu)$	$\lambda=1/\overline{x}$

Гамма-распределение (Эрланга)	$p(x)=\begin{cases} x^{k-1}\dfrac{e^{-x/\theta}}{\theta^k\Gamma(k)}, x\ge 0 \\ 0, else \end{cases}$ $\Gamma(k)=(k-1)\Gamma(k-1);$ $\Gamma(0.5)=\sqrt{\pi}$		$M(X)=k\theta; D(X)=k\theta^2;$ $As=2/\sqrt{k}, Ex=6/k;$	$M_X(\nu)=(1-\theta\nu)^{-k}$ $\phi_X(\nu)=(1-\theta i\nu)^{-k}$	
Хи-квадрат	$p(x)=\dfrac{(1/2)^{n/2}}{\Gamma(n/2)}x^{n/2-1}e^{-x/2}$	$F(x)=\dfrac{\gamma(n/2, x/2}{\Gamma(n/2)})$	$M(X)=n; D(X)=2n;$ $Med\approx n-2/3$ $As=\sqrt{8/n}, Ex=12/n;$	$\phi_X(\nu)=(1-2i\nu)^{-n/2}$	
Стьюдента	$p(x)=\dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(n/2)}(1+\dfrac{y^2}{n})^{-\frac{n+1}{2}}$		$M(X)=Med=\mathrm{mod}=0;$ $D(X)=n/n-2; n>2$ $As=0, n>3,$ $Ex=(3n-6)/(n-4), n>4;$		
Равномерное	$p(x)=\begin{cases} \dfrac{1}{b-a}, a\le x\le b \\ 0, else \end{cases}$	$F(k)=\begin{cases} 0, x<a \\ (x-a)/b-a, a\le x\le b \\ 1, x>b \end{cases}$	$M(X)=Med=(a+b)/2;$ $D(X)=(b-a)^2/12$ $As=0, Ex=-1.2$	$M_X(\nu)=\dfrac{e^{\nu a}-e^{\nu b}}{\nu(b-a)}$ $\phi_X(\nu)=\dfrac{e^{\nu ia}-e^{\nu ib}}{\nu i(b-a)}$	
Треугольное	$p(x)=\begin{cases} \dfrac{2a-\lvert x-c\rvert}{4a^2}, \lvert x-c\rvert\le 2a \\ 0, else \end{cases}$		$M(X)=Med=c;$ $D(X)=2a^2/3$		
Симпсона	$p(x)=\begin{cases} \dfrac{3a^2-\lvert x-c\rvert^2}{8a^3}, \lvert x-c\rvert\le a \\ \dfrac{(3a-\lvert x-c\rvert)^2}{16a^3}, a<\lvert x-c\rvert\le 3a \\ 0, else \end{cases}$		$M(X)=Med=c;$ $D(X)=a^2$		
Рэлея	$p(x)=\dfrac{x}{\sigma^2}\exp\left(-\dfrac{x^2}{2\sigma^2}\right), x\ge 0, \sigma>0$	$F(x)=1-\exp\left(-\dfrac{x^2}{2\sigma^2}\right), x\ge 0$	$M(X)=\sqrt{\pi/2}\sigma;$ $D(X)=(2-\pi/2)\sigma^2$		

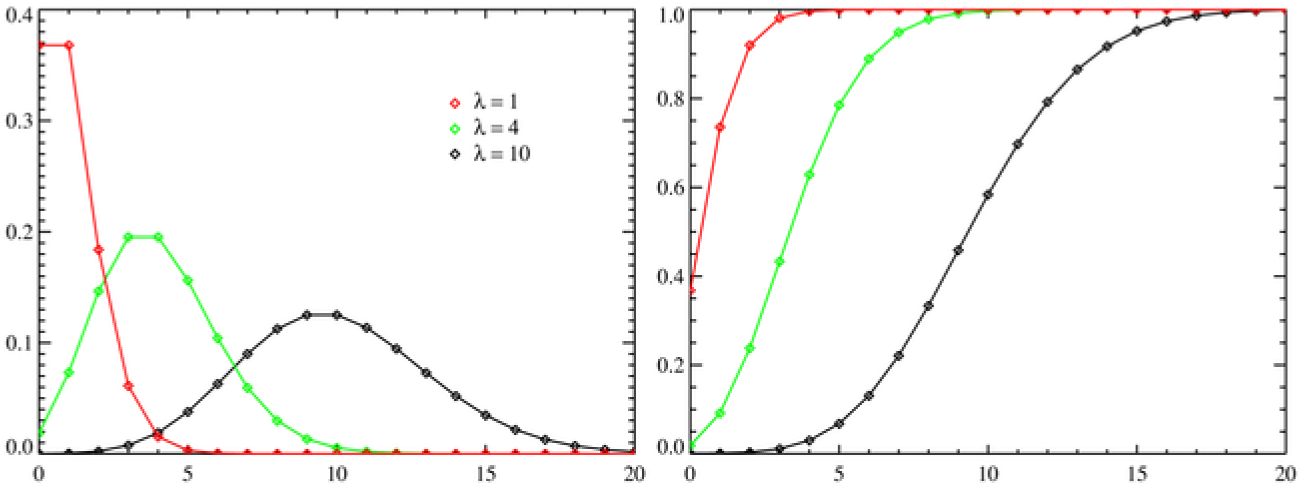
Парето	$p(x)=\frac{kx_m{}^k}{x^{k+1}}, x \ge x_m$	$F(x)=1-\left(\frac{x_m}{x}\right)^k, x \ge x_m$	$M(X^n)=kx_m{}^n/(k-n)$; $M(X)=kx_m/(k-1)$		

Графики плотностей распределений дискретных СВ

Геометрическое

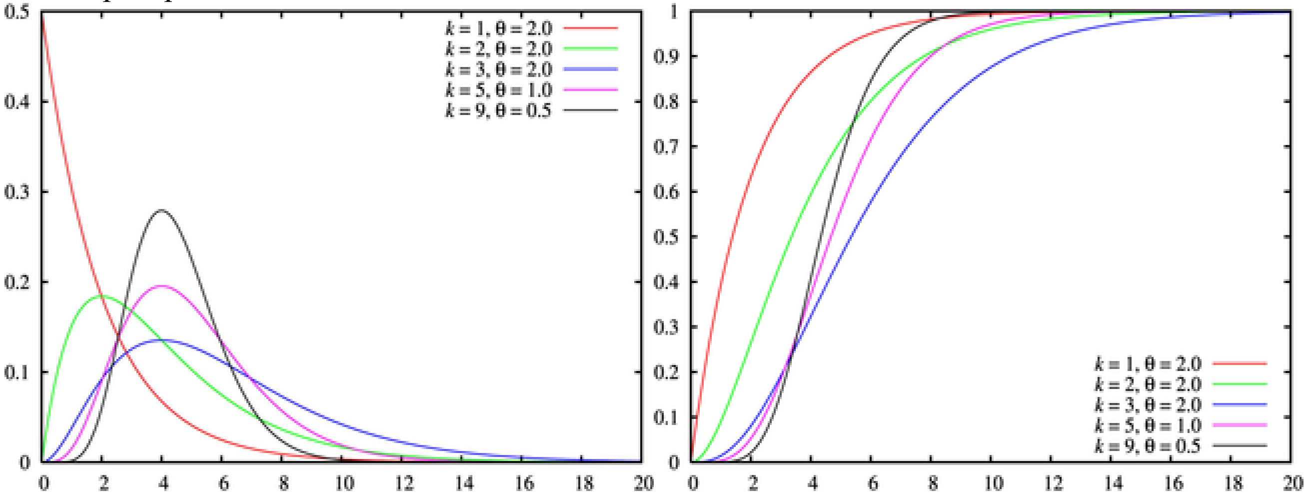


Пуассона

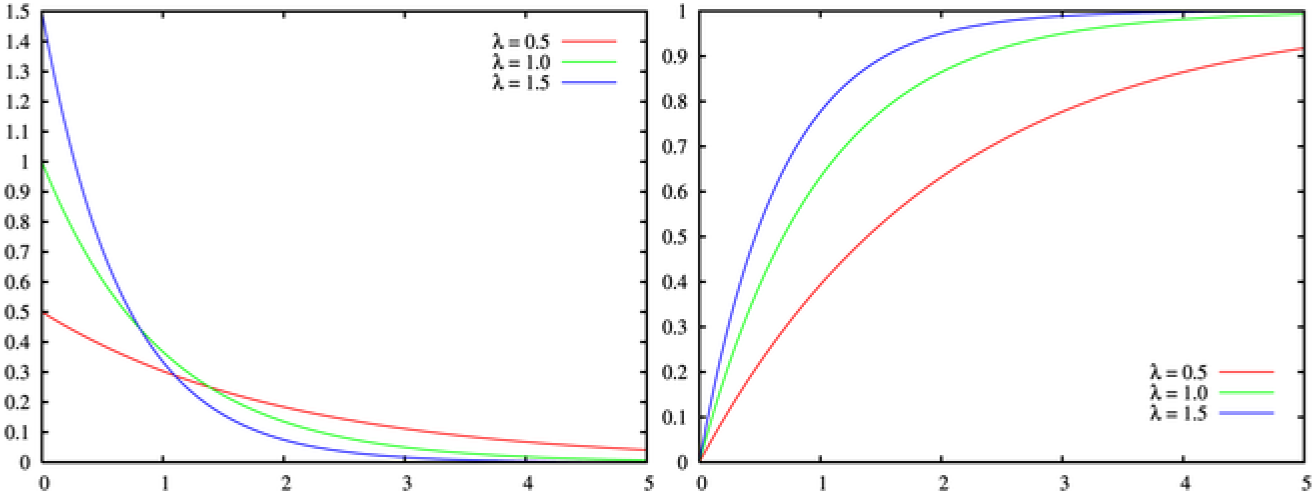


Графики плотностей распределений непрерывных СВ

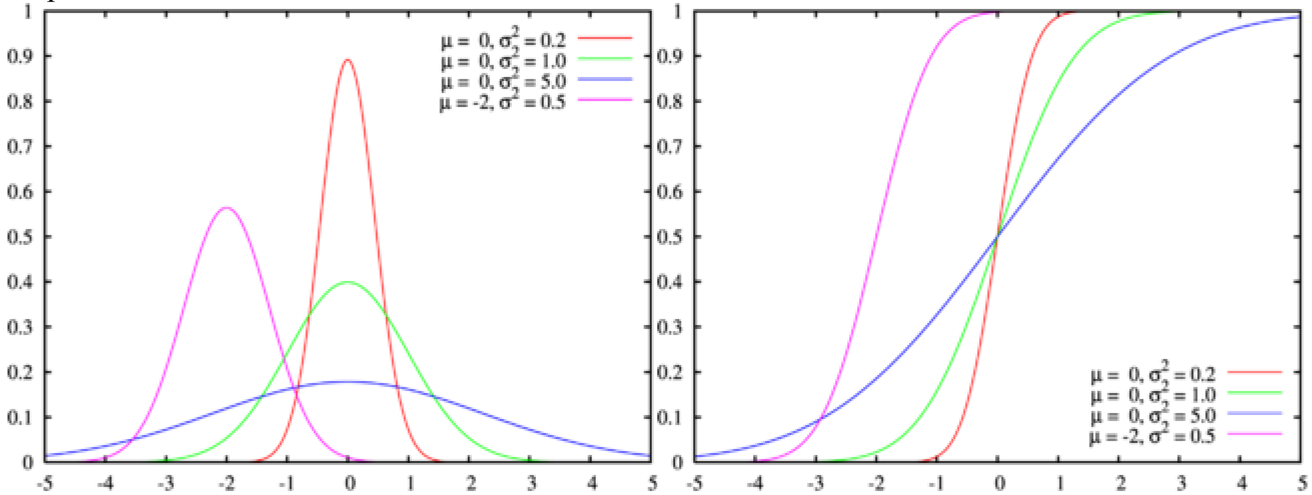
Гамма-распределение



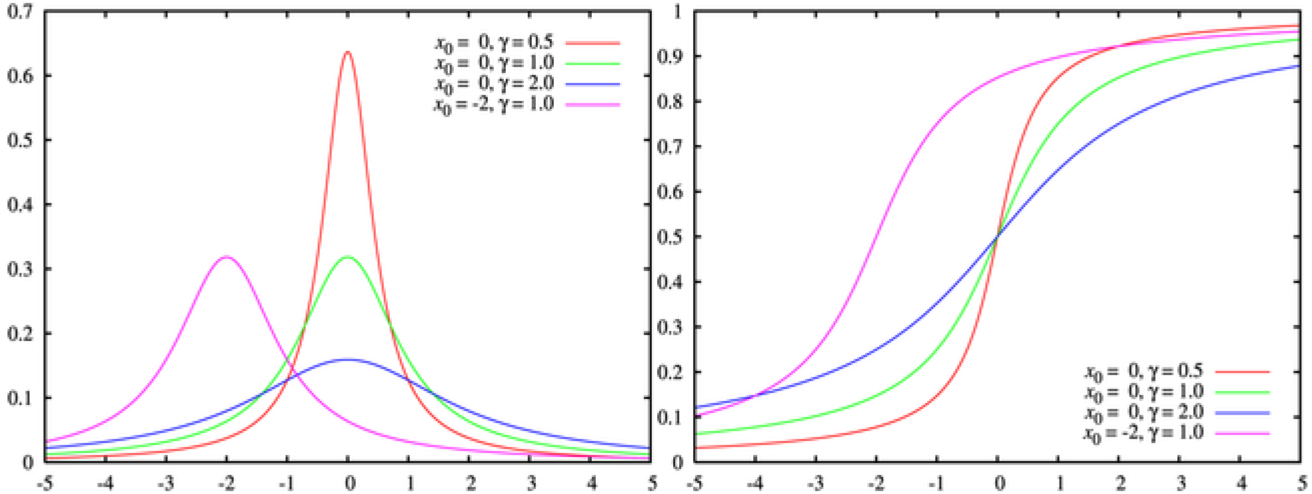
Экспоненциальное



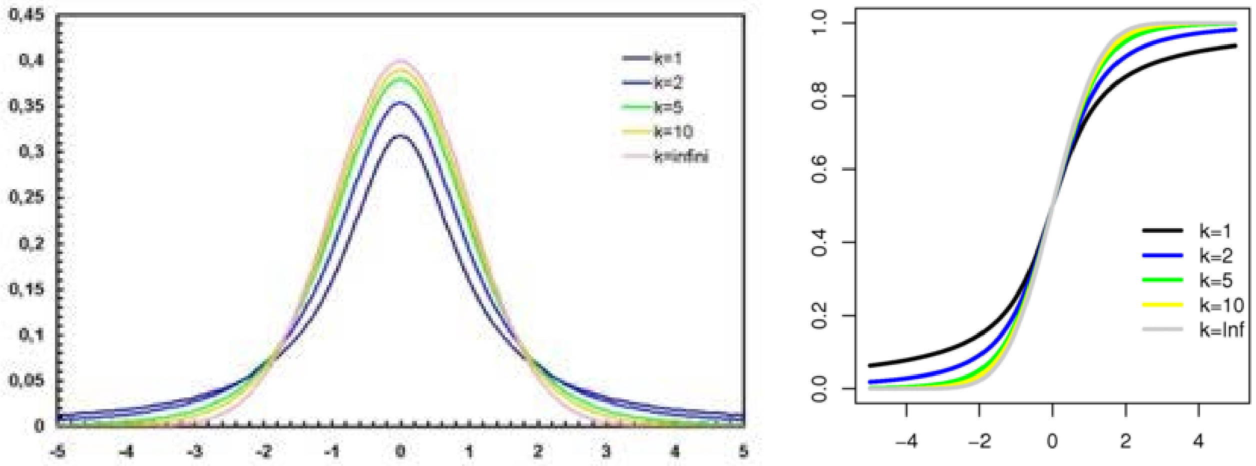
Нормальное



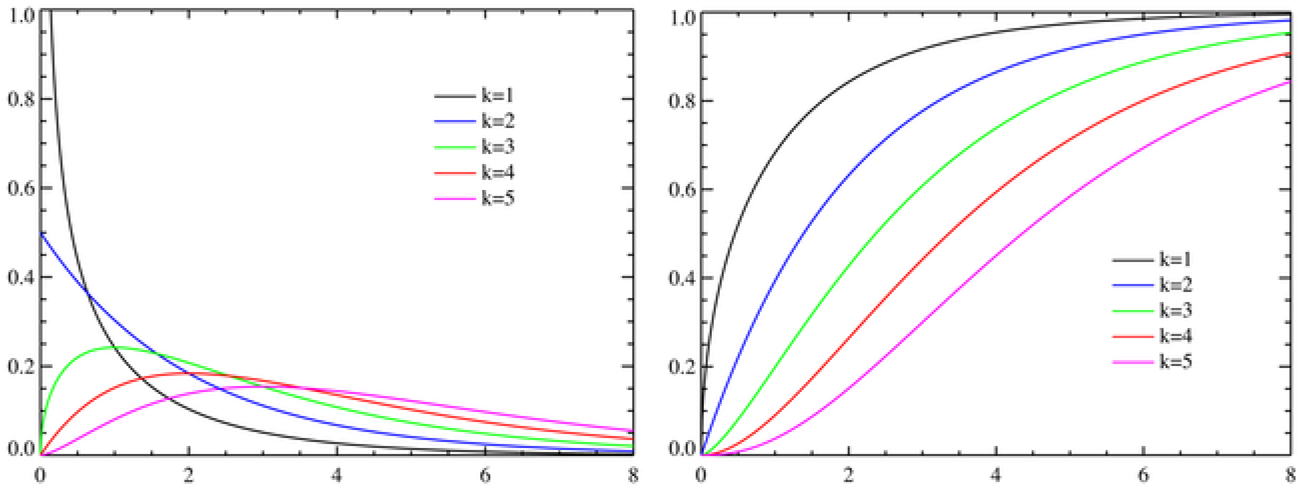
Коши



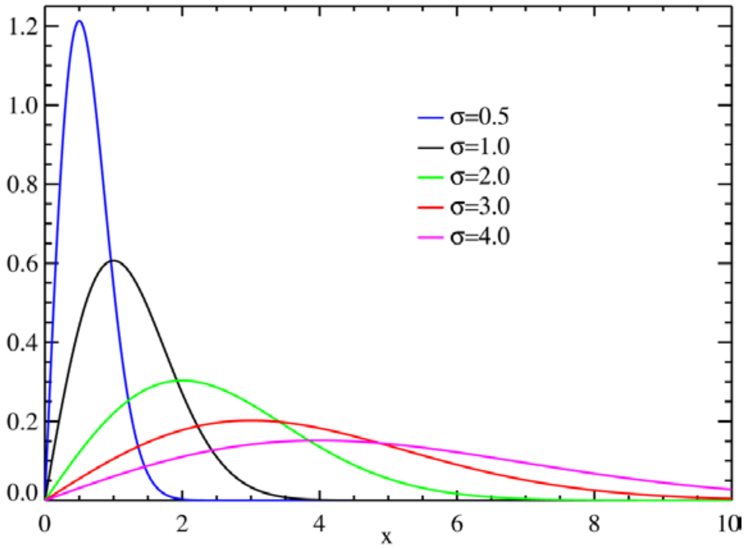
Стьюдента

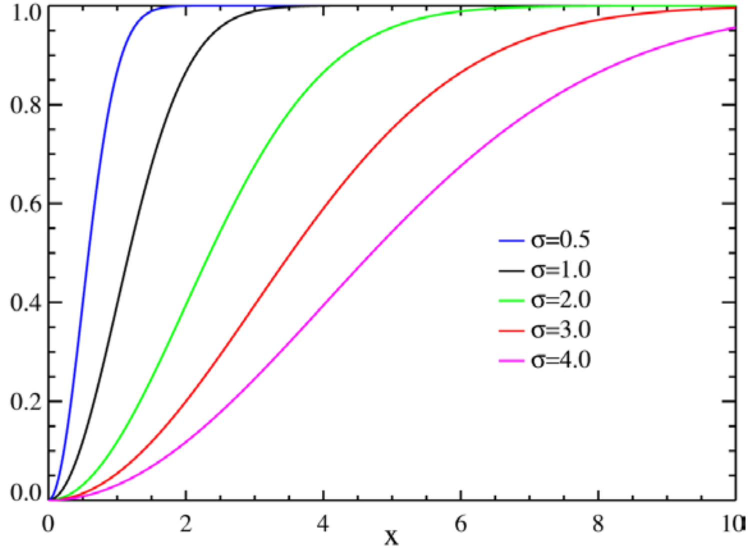


Хи-квадрат

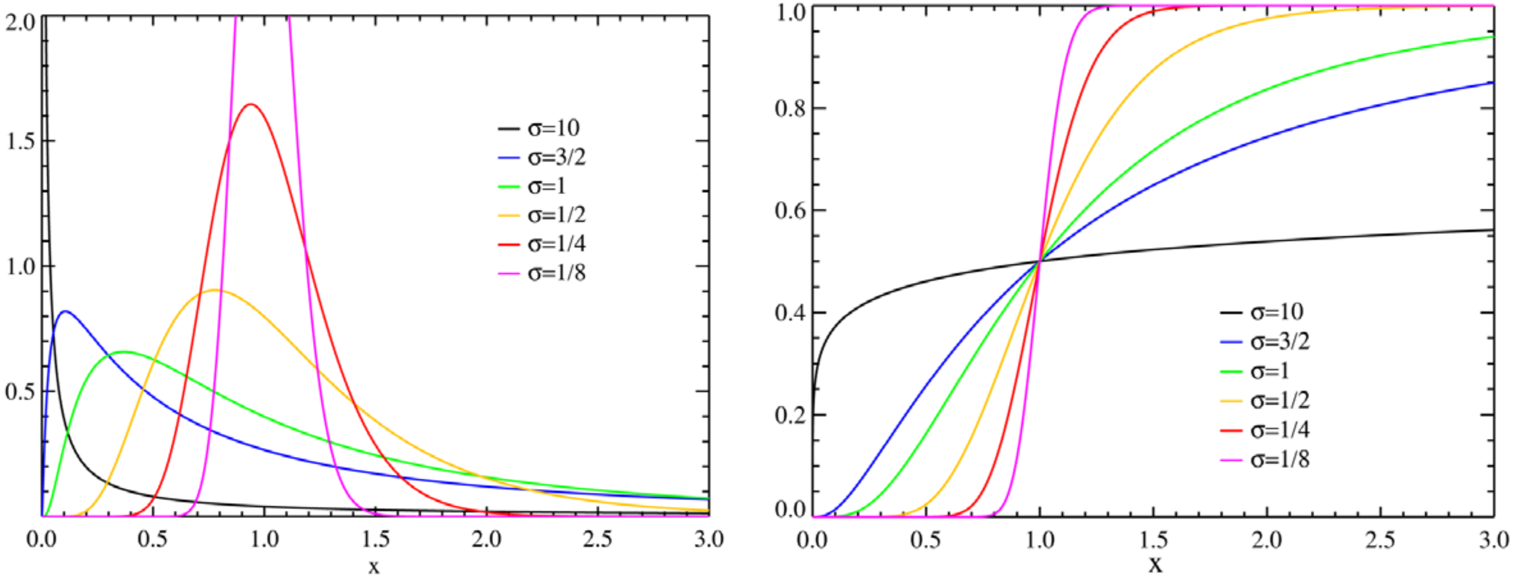


Рэлея

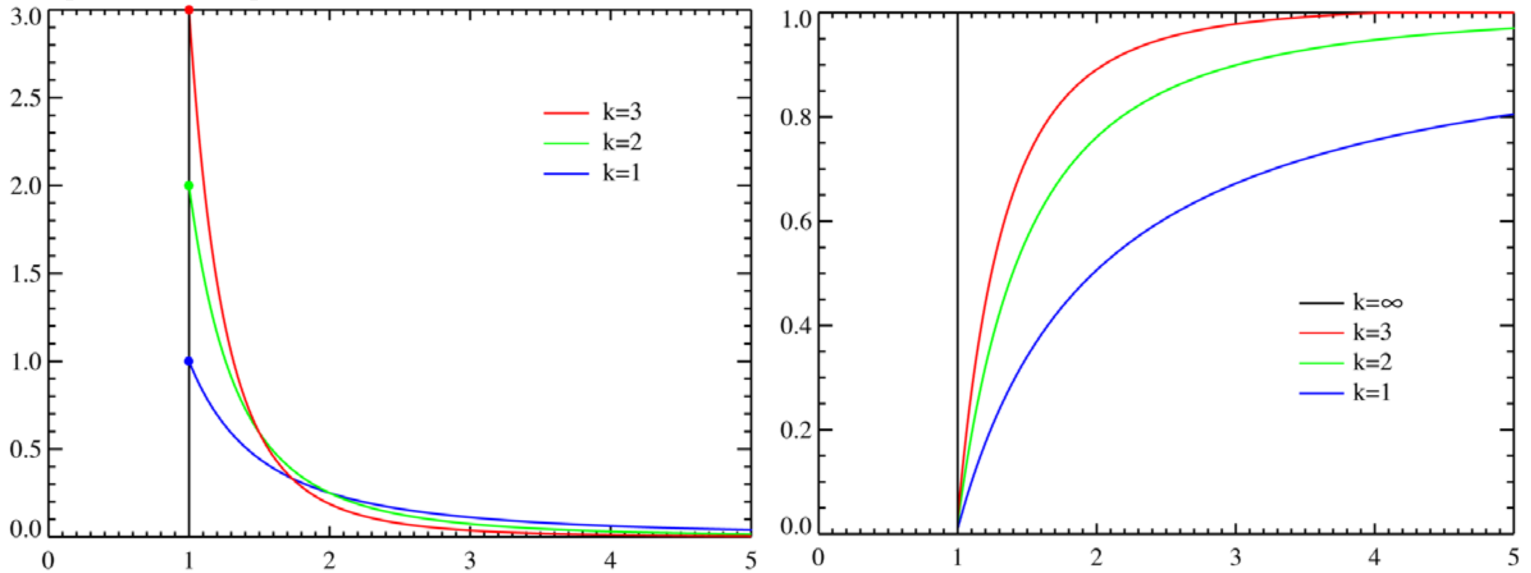




Логнормальное



Распределение Парето



Другие распределения

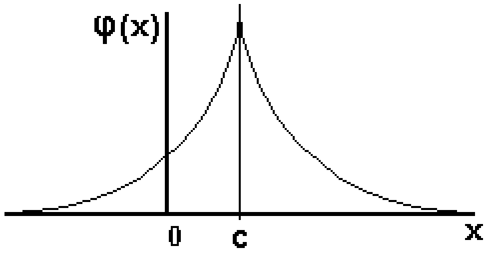


Рис. 18. Плотность распределения Лапласа

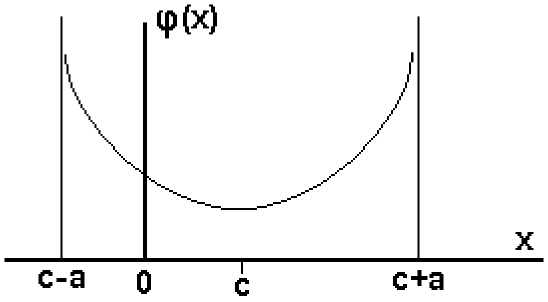


Рис. 15. Плотность распределения Arcsin

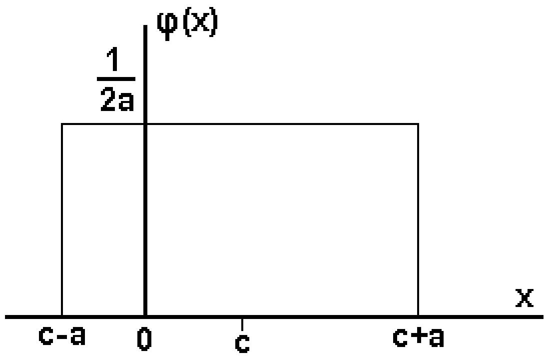


Рис. 14. Равномерная плотность распределения

Приложение 2 Оценивание параметров законов распределения

Метод моментов

Данный метод базируется на том, что для известных распределений существуют соотношения между начальными (центральными) моментами и параметрами распределений:

$$\nu_k = f_{1k}(\Theta),$$

$$\mu_k = f_{2k}(\Theta).$$

Суть метода моментов – приравнять эмпирические моменты (оценки теоретических) к самим теоретическим и по ним оценить неизвестные параметры распределения.

Примеры

1. Экспоненциальное распределение:

$\nu_1 = M[X] = 1/\lambda$, $\mu_2 = D[X] = 1/\lambda^2$

По методу моментов получаем следующие оценки параметров:

$\lambda = 1/\bar{x}$ либо $\lambda = 1/s$.

2. Равномерное распределение

$\nu_1 = M[X] = (a+b)/2$, $\mu_2 = D[X] = \sigma^2 = (b-a)^2/3$.

Решая систему получаем оценки параметров *a* и *b*:

$$\begin{cases}(a+b)/2 = \bar{x} \\ (b-a)^2/3 = s^2\end{cases} \Rightarrow \begin{cases}a+b = 2\bar{x} \\ b-a = s\sqrt{3}\end{cases} \Rightarrow \begin{cases}a = \bar{x} - s\sqrt{3}/2 \\ b = \bar{x} + s\sqrt{3}/2\end{cases}$$

3. Гамма-распределение

$\nu_1 = M[X] = k\theta$, $\mu_2 = D[X] = k\theta^2$.

По методу моментов получаем следующие оценки параметров:

$$\begin{cases}k\theta = \bar{x} \\ k\theta^2 = s^2\end{cases} \Rightarrow \begin{cases}k = \bar{x}^2/s^2 \\ \theta = s^2/\bar{x}\end{cases}$$

4. Нормальное распределение

$\nu_1 = M[X] = c$, $\mu_2 = D[X] = \sigma^2$.

По методу моментов сразу получаем оценки параметров:

$$\begin{cases}c = \bar{x} \\ \sigma^2 = s^2\end{cases}.$$

Метод максимального правдоподобия

Метод максимального правдоподобия точечной оценки неизвестных параметров заданного распределения сводится к отысканию максимума функции одного или нескольких оцениваемых параметров.

Дискретные случайные величины.

Пусть X – дискретная случайная величина, которая в результате n опытов приняла возможные значения x_1, x_2, ..., x_n. Допустим, что вид закона распределения величины X задан, но неизвестен параметр θ, которым определяется этот закон; требуется найти его точечную оценку $\acute{\theta} = \acute{\theta}(x_1, x_2, ..., x_n)$

Обозначим вероятность того, что в результате испытания величина X примет значение x_1 при параметре, равном θ, через $p(x_1/\theta)$. Данную вероятность еще иногда называют правдоподобием.

Функцией правдоподобия дискретной случайной величины X называют функцию, являющуюся произведением вероятностей $p(x_i,\theta)$ по всем выборочным значениям:

$$L(x_1,x_2,...,x_n,\theta)=p(x_1,\theta)p(x_2,\theta)...p(x_n,\theta)$$

Оценкой наибольшего правдоподобия параметра θ называют такое его значение θ^*, при котором функция правдоподобия достигает максимума.

Для удобства часто используют т.н. логарифмическую функцию правдоподобия ln$L$.

Функции L и InL достигают максимума при одном и том же значении θ, поэтому вместо отыскания максимума функции L ищут максимум функции InL. Это проще, т.к. логарифмическая функция позволяет от произведения перейти к сумме.

Для отыскания максимума функции правдоподобия ее производная или градиент приравнивается к нулю и из полученного уравнения (системы уравнений) находятся искомые оценки параметров:

$$\frac{d\ln L}{d\theta}=0 \text{ или } \nabla L_\theta=0$$

Данные уравнения называются уравнения правдоподобия.

Полученные значения оценок являются максимумом функции правдоподобия, если вторая производная отрицательна (матрица вторых производных отрицательно определена).

Найденную точку максимума θ^* принимают в качестве оценки наибольшего правдоподобия параметра θ.

Непрерывные случайные величины.

Пусть X – непрерывная случайная величина, которая в результате n испытаний приняла значения x_1, x_2, ..., x_n. Пусть вид плотности распределения – функции $f(x)$ задан, но неизвестны параметры θ, которыми определяется эта функция. Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента:

$$L(x_1,x_2,...,x_n,\theta)=f(x_1,\theta)f(x_2,\theta)...f(x_n,\theta)$$

Оценку максимального правдоподобия неизвестных параметров распределения непрерывной случайной величины ищут так же, как в случае дискретной случайной величины. Производные (градиент) от функции правдоподобия по оцениваемым параметрами приравниваются к нулю, решается уравнение (система уравнений), далее проверяется, что вторая производная меньше нуля (матрица вторых производных отрицательно определена) в точке, соответствующей решению θ^*.

Определить ММП-оценку параметра *p* биномиального распределения

$$P_N(m)=C_N^m p^m(1-p)^{N-m}$$

Составим логарифмическую функцию правдоподобия:

$$\ln L=\ln\left(\prod_{i=1}^{n}C_N^{m_i}p^{m_i}(1-p)^{N-m_i}\right)=\sum_{i=1}^{n}\ln C_N^{m_i}+\ln p\cdot\sum_{i=1}^{n}m_i+\ln(1-p)\sum_{i=1}^{n}(N-m_i)$$

Приравняем производную от $\ln L$ по p к 0 и решим данное уравнение относительно p:

$$\frac{d\ln L}{dp}=\frac{1}{p}\sum_{i=1}^{n}m_i-\frac{1}{1-p}\sum_{i=1}^{n}(N-m_i)=0\,.$$

$$(1-p)\sum_{i=1}^{n}m_i=p\sum_{i=1}^{n}(N-m_i)\Leftrightarrow\sum_{i=1}^{n}m_i=p\sum_{i=1}^{n}N\Leftrightarrow p=\frac{1}{N}\frac{\sum_{i=1}^{n}m_i}{n}=\frac{1}{N}\overline{m}\,.$$

Оценка совпадает с оценкой по методу моментов.

Найти ММП – оценку математического ожидания *c* для нормального распределения (измерения неравноточные, СКО каждого измерения равно σ_i).

Плотность вероятности нормального распределения для выборочного значения x_i равна:

$$f(x_i)=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(\frac{-(x_i-c)^2}{2\sigma_i^{\,2}}\right)$$

Составим логарифмическую функцию правдоподобия:

$$\ln L=\ln\prod_{i=1}^{n}f(x_i)=\sum_{i=1}^{n}\ln\left[\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(\frac{-(x_i-c)^2}{2\sigma_i^{\,2}}\right)\right]=\sum_{i=1}^{n}\ln\left[\frac{1}{\sqrt{2\pi}\sigma_i}\right]-\sum_{i=1}^{n}\frac{(x_i-a)^2}{2\sigma_i^{\,2}}$$

Приравняем производную от $\ln L$ по c к 0 и решим данное уравнение относительно c:

$$\frac{d\ln L}{dc}=\sum_{i=1}^{n}\frac{2(x_i-c)}{2\sigma_i^{\,2}}=0\Leftrightarrow c\sum_{k=1}^{n}\frac{1}{\sigma_k^{\,2}}=\sum_{i=1}^{n}\frac{x_i}{\sigma_i^{\,2}}\Rightarrow$$

$$c=\left(\sum_{k=1}^{n}\frac{1}{\sigma_k^{\,2}}\right)^{-1}\sum_{i=1}^{n}\frac{x_i}{\sigma_i^{\,2}}=\sum_{i=1}^{n}x_iw_i,\ w_i=\frac{1}{\sigma_i^{\,2}}\left(\sum_{k=1}^{n}\frac{1}{\sigma_k^{\,2}}\right)^{-1}$$

Найти ММП – оценку математического ожидания для нормального распределения (измерения равноточные, СКО равно σ).

Решение можно получить из предыдущего примера, положив $\sigma_i=\sigma, i=1...n$. Тогда

$$w_i=\frac{1}{\sigma^2}\left(\sum_{k=1}^{n}\frac{1}{\sigma^2}\right)^{-1}=\frac{1}{n}$$

и оценка параметра c равна

$$c=\sum_{i=1}^{n}x_iw_i=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}\,.$$

Как видно, при равноточных измерениях ММП-оценка параметра c нормального распределения совпадает со средним арифметическим и оценкой, полученной по методу моментов.

Найти ММП – оценку параметров c и σ для нормального распределения (измерения равноточные).

Для нахождения оценок параметров необходимо приравнять производные от $\ln L$ по этим параметрам к нулю:

$$\begin{cases}\dfrac{\partial\ln L}{\partial c}=0\\[2mm]\dfrac{\partial\ln L}{\partial\sigma}=0\end{cases}$$

Из первого уравнения находится оценка c (предыдущая задача): $c=\bar{x}$.

Поэтому оценку второго параметра можно найти из второго уравнения. Составим логарифмическую функцию правдоподобия для случая равноточных измерений:

$$\ln L=\ln\prod_{i=1}^{n}f(x_i)=\sum_{i=1}^{n}\ln\left[\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(x_i-c)^2}{2\sigma^2}\right)\right]=\sum_{i=1}^{n}\ln\left[\frac{1}{\sqrt{2\pi}\sigma}\right]-\sum_{i=1}^{n}\frac{(x_i-a)^2}{2\sigma^2}=$$

$$=-n\ln(\sqrt{2\pi}\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-c)^2$$

Приравняем производную от $\ln L$ по σ к 0 и решим данное уравнение относительно σ:

$$\frac{d\ln L}{d\sigma}=-\frac{N}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-c)^2=0\Leftrightarrow\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-c)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Как видно, ММП-оценка σ^2 совпадает со смещенной оценкой дисперсии и совпадает с оценкой по методу моментов.

Найти ММП – оценку параметра экспоненциального распределения

Плотность вероятности экспоненциального распределения для выборочного значения x_i равна:

$f(x_i) = \lambda \exp(-\lambda x_i)$.

Составим логарифмическую функцию правдоподобия:

$$\ln L = \ln \prod_{i=1}^{n} f(x_i) = \sum_{i=1}^{n} \ln(\lambda \exp(-\lambda x_i)) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_k$$

Приравняем производную от $\ln L$ по λ к 0 и решим данное уравнение относительно λ:

$$\frac{\partial \ln L}{\partial \lambda} = \frac{N}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Leftrightarrow \lambda = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\overline{x}}$$

Найти ММП – оценку параметров распределения Лапласа

Плотность вероятности распределения Лапласа для выборочного значения x_i равна:

$$f(x_k) = \frac{\lambda}{2} \exp(-\lambda |x_k - c|)$$

Составим логарифмическую функцию правдоподобия:

$$\ln L = \ln \prod_{i=1}^{n} f(x_i) = \sum_{i=1}^{n} \ln\left[\frac{\lambda}{2} \exp(-\lambda |x_i - c|)\right] = n \ln(\lambda / 2) - \lambda \sum_{i=1}^{n} |x_i - c|$$

ММП-оценки параметров находятся из системы уравнений:

$$\begin{cases} \dfrac{\partial \ln L}{\partial c} = 0 \\ \dfrac{\partial \ln L}{\partial \lambda} = 0 \end{cases}$$

Найдем вначале ММП-оценку математического ожидания c. Поскольку первое слагаемое не содержит c, его значение не влияет на положение максимума, его можно исключить. Останется одно отрицательное слагаемое, минимум модуля которого совпадает по положению с максимумом функции правдоподобия. Поэтому будем отыскивать ММП-оценку путем поиска минимума суммы $\sum_{i=1}^{n} | x_i - c |$. Эта сумма недифференцируема, и придется находить искомую оценку геометрически.

Пусть в результате эксперимента получено всего три выборочных значения x_1, x_2, x_3, которые разместились на вещественной оси так, как показано на рис.

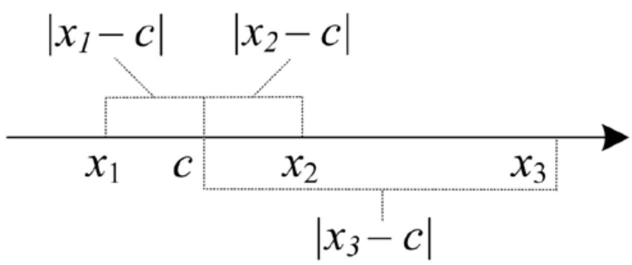


Рис. Подход к определению минимум суммы $\sum_{i=1}^{n} | x_i - c |$

Каждое слагаемое минимизируемой суммы есть расстояние от каждого выборочного значения до точки a. В ситуации, представленной на рисунке, одно из этих расстояний входит в сумму дважды. Это расстояние $|x_2 - c|$. Видер, что рассматриваемая сумма достигнет минимума только тогда, когда оценка c совместится с x_2, то есть с выборочной медианой. Добавляя к этой небольшой выборке четное количество элементов, мы придем к тому же выводу, что ММП-оценкой математического ожидания случайной величины, распределенной по Лапласу, является выборочная медиана, то есть $c = x_{med}$. И по определению ММП данная оценка эффективна. Далее найдем ММП-оценку параметра λ. Приравняем производную от ln L по λ к 0 и решим данное уравнение относительно λ:

$$\frac{d\ln L}{d\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}|x_i-c|=0\Rightarrow\lambda=\frac{n}{\sum_{i=1}^{n}|x_i-c|}=\frac{n}{\sum_{i=1}^{n}|x_i-x_{med}|}.$$

Метод минимума Хи-квадрат

Ставится задача оценки параметров θ плотности распределения $f(x, \theta)$, вид которой известен. Исходными данными являются выборочные значения, по которым строится гистограмма. Идея метода заключается в подборе таких значений искомых параметров, при которых достигается минимальное отличие кривой плотности распределения от гистограммы. В качестве меры этого отличия чаще всего используется квадратичный функционал, наиболее удобный для реализации аналитических и численных методов поиска экстремума (минимума или максимума).

В данной задаче в качестве такого функционала используется сумма, обозначенная как χ^2:

$$\chi^2=\sum_{k=1}^{K}\frac{n}{P_k}\left(P_k-\frac{n_k}{n}\right)^2=\frac{1}{n}\sum_{k=1}^{K}\frac{(n\cdot P_k-n_k)^2}{P_k},$$

В этой сумме K - общее количество интервалов, на которых построена гистограмма, n - объем выборки, n_k – количество выборочных значений, попавших в k-ый интервал гистограммы, P_k – вероятность случайной величине с плотностью $f(x, \theta)$ попасть в k-й интервал гистограммы (вероятностная мера k-го интервала гистограммы):

$$P_k=\int_{x_{k-1}}^{x_k}f(x,\theta)dx.$$

Таким образом, слагаемые этой суммы представляют собой квадраты разностей между вероятностными мерами k-ых интервалов, порожденными генеральной плотностью распределения, и частотными оценками $\tilde{P}_k=n_k/n$ этих вероятностных мер. Знаменателем каждого слагаемого является вероятность P_k, благодаря чему в процессе поиска значений параметров θ повышается вес интервалов с низкой вероятностью и тем самым обеспечивается повышенная точность подгонки в области этих интервалов. Как правило, эти интервалы находятся на удалении от центра распределения (на так называемых “хвостах” распределений).

Вероятности P_k являются функциями от искомых параметров θ, от этих же параметров зависит и величина χ^2, и процедура оценивания параметров по методу минимума χ^2 формально записывается в виде:

$$\theta^*=\underset{\theta}{\operatorname{argmin}}\,\chi^2=\underset{\theta}{\operatorname{argmin}}\left[\sum_{k=1}^{K}\frac{(n\cdot P_k(\theta)-n_k)^2}{nP_k(\theta)}\right].$$

В большинстве случаев этот минимум и оценки находятся численными методами. Доказано, что оценки, полученные методом минимума χ^2, обладают свойствами, сопоставимыми со свойствами ММП - оценок, а именно, эти оценки асимптотически эффективны.

Примеры определения оценок параметров распределений

Пример 1 Нормальное распределение, метод моментов

Предположим, что мы ищем оценки параметров нормального распределения. У нормального распределения 2 параметра – центр a и разброс (СКО) σ. По методу моментов для нормального распределения находим из таблицы распределений непрерывных СВ, что a

= M[x], σ^2=D[x], поэтому оценками параметров a и σ являются оценки математического ожидания и корень из оценки дисперсии соответственно:

$$\hat{a} = \hat{M}[X] = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}$$ - среднее арифметическое

$$\hat{\sigma} = \sqrt{\hat{D}[X]} = \sqrt{s^2}$$, где $$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$ - несмещенная выборочная оценка дисперсии.

Пример 2 Гамма-распределение, метод моментов

Предположим, что мы делаем гипотезу о гамма-распределении. У него 2 параметра – k и θ. Из таблицы находим, что M[x]=kθ, D[x]=kθ^2.

Решая систему из двух уравнений с неизвестными k и θ выражаем эти параметры через моменты: k = M^2[x]/D[x], θ = D[x]/M[x]. Далее определяем оценки параметров, используя в качестве моментов M[x] и D[x] их оценки так же, как и в примере 1. Окончательно получаем

$$\hat{k} = \frac{\hat{M}^2[x]}{\hat{D}[x]} = \frac{\bar{x}^2}{s^2} \quad \hat{\theta} = \frac{\hat{D}[x]}{\hat{M}[x]} = \frac{s^2}{\bar{x}}$$

Зам. Можно было действовать и по-другому. Например, из таблицы видно, что асимметрия и эксцесс равны $As = 2/\sqrt{k}, Ex = 6/k$;

Поэтому можно сразу выразить k как 6/Ex или 4/As^2. Далее воспользовавшись оценкой эксцесса или асимметрии можно сразу найти параметр k, а затем из одного из уравнений (для мат.ожидания или дисперсии) определить второй параметр θ. Но как правило, в методе моментов используются моменты как можно меньшего порядка для нахождения оценок параметров.

Пример 3 Нормальное распределение, метод ММП, аналитический.

Из таблицы сразу находим, что для нормального распределения существует аналитическая формула для оценки параметров по ММП. Поэтому сразу находим.

$$a = \bar{x}; \sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N}}$$

Пример 4 Нормальное распределение, метод ММП, численный.

Для нормального распределения в Matlab есть функция для подгонки параметров по ММП – normfit. Ее вызов осуществляется следующим образом:

```
[muhat,sigmahat] = normfit(data)
```

где muhat – оценка центра a, sigmahat – оценка СКО σ, data – исходная выборка.

Но можно воспользоваться и более универсальной функцией подгонки по ММП – mle. Ее вызов в общем виде осуществляется следующим образом:

```
phat = mle(data,'pdf',pdf,'start',start)
```

phat – вектор оцениваемых параметров, data – выборка, pdf – указатель на функцию с плотностью распределения, start – начальное приближение для параметров

Сгенерируем выборку из 1000 точек с нормальным распределением, a = 5, σ=2:

```
data = normrnd(5, 2, [1000 1]);
```

Найдем ММП-оценку численным способом:

```
phat = mle(data,'pdf',@normpdf,'start',[1 1])
```

В качестве плотности pdf мы передали указатель на стандартную функцию плотности нормального распределения normpdf, начальное приближение мы задали равным a=1, σ=1.

Найдем аналитические оценки:

```
a_mle = mean(data)
sigma_mle = sqrt(var(data,1))
```

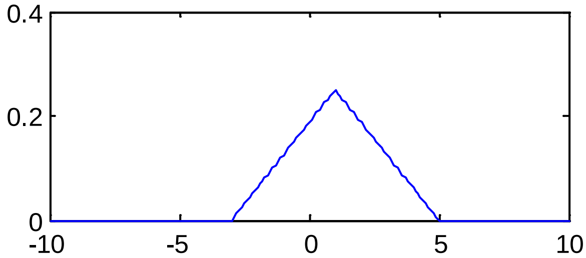
Пример 5 Распределение треугольное, метод ММП, численный.

Для начала необходимо задать плотность распределения в Matlab, поскольку ее там нет. Проще всего воспользоваться строкой-функцией:

```
pdf_tri = @(x,c,a)((abs(x-c)<=2*a).*((2*a-abs(x-c))/(4*a^2))+1e-10);
```

К функции добавляется малая константа 1e-10, чтобы плотность не обращалась в 0 – это требование для использования в дальнейшем функции mle. Убедимся, что все правильно, построим график плотности для c=1, a =2

```
x = -10:.1:10; y = pdf_tri(x,1,2); plot(x,y)
```



Интеграл от функции равен единице

```
integral(pdf_tri(,1,2),-10,10)
```

Значит все правильно и можно использовать эту функцию для нахождения плотности. Сгенерируем данные как в примере 5:

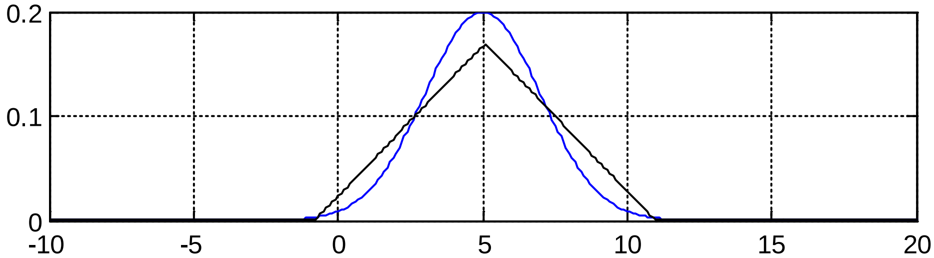
```
data = normrnd(5, 2, [1000 1]);
```

Найдем оценки для треугольного распределения по методу ММП:

```
phat = mle(data,'pdf',@ pdf_tri,'start',[1 1])
```

Функция дает ответ phat = 5.0811 2.9443. Построим на истинной нормальной плотности плотность полученного треугольного распределения:

```
x=-10:.1:20; y1 = normpdf(x,5,2); plot(x,y1); grid on; hold on;
y2 = pdf_tri(x,phat(1),phat(2)); plot(x,y2, 'k');
```



Видим, что действительно параметры треугольного распределения корректно определились под исходную выборку data.

Пример 6 Распределение Лапласа, метод ММП

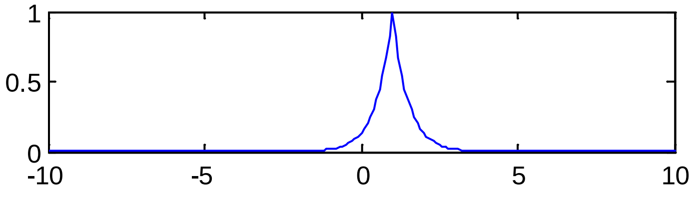
Проделаем действия по аналогии с примером 5.

Задаем плотность распределения Лапласа, учитывая, что $\lambda>0$.

```
pdf_lapl = @(x,c,l)(l/2*exp(-l*abs(x-c))*(l>0)+1e-10);
```

Строим график плотности

```
x = -10:.1:10; y = pdf_lapl(x,1,2); plot(x,y)
```



Находим интеграл и убеждаемся, что он равен 1.

```
integral(@(x)pdf_lapl(x,1,2), -10,10)
```

Генерируем нормальное распределение

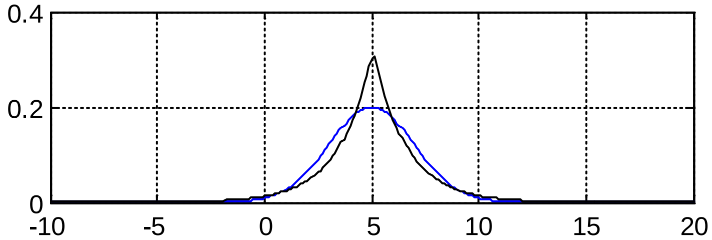
```
data = normrnd(5, 2, [1000 1]);
```

Находим оценки параметров распределения Лапласа

```
phat = mle(data,'pdf',pdf_lapl,'start',[1 1])
```

Строим исходную нормальную плотность и плотность распределения Лапласа с найденными параметрами

```
x=-10:.1:20; y1 = normpdf(x,5,2); plot(x,y1); grid on; hold on;
y2 = pdf_lapl(x,phat(1),phat(2)); plot(x,y2, 'k');
```



Из таблицы видно, что для распределения Лапласа параметры можно посчитать и аналитически:

$\hat{c} = x_{med}$;

$$\hat{\lambda} = N\left(\sum_{i=1}^{N} |xi - \hat{c}|\right)^{-1}$$

Проверяем и убеждаемся, что полученные значения совпадают с найденными функцией mle:

```
c_mle = median(data)
l_mle = length(data)/sum(abs(data-c_mle))
```

Пример 7 Программа, иллюстрирующая подгонку параметров для 4 распределений – Лапласа, треугольного, Симпсона и нормального

Чтобы изучить данную программу, необходимо создать файл с именем simp_ex.m, скопировать туда текст приведенной ниже программы, сохранить и запустить программу на выполнение. Вначале лучше выполнять программу в режиме отладки (по шагам) и контролировать изменение всех переменных.

```
function simp_ex

c = 7;
a = 3;
N = 10000;
x_range = -10:.1:20;

menu_choice = 1;
while menu_choice ~= 5
    menu_choice = menu('What to do', {'Tri', 'Sim', 'Laplace', 'Normal',
'Exit'});
    close all;
    switch menu_choice
        case 1
            tri_ex(c,a,N,x_range);
        case 2
            simpson_ex(c,a,N,x_range);
        case 3
            lapl_ex(c,a,N,x_range);
```

```
        case 4
            norm_ex(c,a,N,x_range);
    end;
end;
close all;

% Пример на треугольное распределение
function tri_ex(c,a,N,x)
% Задаем плотность треугольного распределения
tripdf = @(x,c,a)((abs(x-c)<=2*a).*((2*a-abs(x-c))/(4*a^2))+1e-10);

% Построим график плотности с параметрами c=1, a=2
y = tripdf(x,c,a);
plot(x,y)

% Убедимся, что интеграл по плотности равен 1
integral(@(x)tripdf(x,c,a),-10,10)
% Проверим, чему равны мат.ожидание и дисперсия
mean_tri = integral(@(x)tripdf(x,c,a).*x,-10,10)  % Мат.ожидание
delta1 = mean_tri - c % Должно быть равно c
var_tri = integral(@(x)tripdf(x,c,a).*(x-mean_tri).^2,-10,10) % Дисперия
delta2 = var_tri - 2/3*a^2 % Должна быть равна 2/3*a^2

% Сгенеририруем нормальное распределение N(5,2)
data = normrnd(c, a, [N 1]);

% Аппроксимируем нормальное распределение треугольным и подгоним параметры
% для треугольного
% Вначале с помощью метода ММП
phat = mle(data,'pdf',tripdf,'start',[1 1]);
c_mle = phat(1)
a_mle = phat(2)
% Затем с помощью метода моментов
c_mm = mean(data)
a_mm = sqrt(1.5*var(data))

% Далее построим на одном графике истинную плотность и две плотности
% треугольного распределения с разными
% значениями оцениваемых параметров, найденные выше
figure;
y1 = normpdf(x,c,a); % Нормальная плотность
plot(x,y1); grid on; hold on;
y2 = tripdf(x,c_mle,a_mle); % Плотность треуг. распр. с параметрами ММП
plot(x,y2, 'k');
y3 = tripdf(x,c_mm,a_mm);  % Плотность треуг. распр. с параметрами ММ
plot(x,y3, 'r');
legend({'Normal', 'Tri - MLE', 'Tri - Moment'});
% Пример на распределение Симпсона
function simpson_ex(c,a,N,x)
% Задаем плотность распределения Симпсона
simpdf = @(x,c,a) ((abs(x-c) > a) .* (abs(x-c) <= 3*a) .* (3*a-abs(x-
c)).^2/(16*a^3) + (abs(x-c) <= a).* (3*a^2 - (x-c).^2)/(8*a^3) + 1e-10);

% Построим график плотности с параметрами c=1, a=2
y = simpdf(x,c,a);
plot(x,y)

% Убедимся, что интеграл по плотности равен 1
integral(@(x)simpdf(x,c,a),-10,10)
% Проверим, чему равны мат.ожидание и дисперсия
mean_sim = integral(@(x)simpdf(x,c,a).*x,-10,10)  % Мат. ожидание
delta1 = mean_sim - c % Должно быть равна c
var_sim = integral(@(x)simpdf(x,c,a).*(x-mean_sim).^2,-10,10) % Дисперия
```

```
delta2 = var_sim - a^2 % Должна быть равна a^2

% Сгенеририруем нормальное распределение N(5,2)
data = normrnd(c, a, [N 1]);

% Аппроксимируем нормальное распределение р-м Симпсона и подгоним параметры
% для распределения Симпсона
% Вначале с помощью метода ММП
phat = mle(data,'pdf',simpdf,'start',[1 1]);
c_mle = phat(1)
a_mle = phat(2)
% Затем с помощью метода моментов
c_mm = mean(data)
a_mm = sqrt(var(data))

% Далее построим на одном графике истинную плотность и две плотности
% треугольного распределения с разными
% значениями оцениваемых параметров, найденные выше
figure;
y1 = normpdf(x,c,a); % Нормальная плотность
plot(x,y1); grid on; hold on;
y2 = simpdf(x,c_mle,a_mle); % Плотность треуг. распр. с параметрами ММП
plot(x,y2, 'k');
y3 = simpdf(x,c_mm,a_mm);  % Плотность треуг. распр. с параметрами ММ
plot(x,y3, 'r');
legend({'Normal', 'Simpson - MLE', 'Simpson - Moment'});
% Пример на распределение Лапласа
function lapl_ex(c,a,N,x)
% Задаем плотность распределения Симпсона
laplacepdf = @(x,c,a)(a/2*exp(-a*abs(x-c))*(a>0)+1e-10);

% Построим график плотности с параметрами c=1, a=2
y = laplacepdf(x,c,a);
plot(x,y)

% Убедимся, что интеграл по плотности равен 1
integral(@(x)laplacepdf(x,c,a),-10,10)
% Проверим, чему равны мат.ожидание и дисперсия
mean_laplace = integral(@(x)laplacepdf(x,c,a).*x,-10,10)  % Мат. ожидание
delta1 = mean_laplace - c % Должно быть равна с
var_laplace = integral(@(x)laplacepdf(x,c,a).*(x-mean_laplace).^2,-10,10) %
Дисперсия
delta2 = var_laplace - 2/a^2 % Должна быть равна 2/a^2

% Сгенеририруем нормальное распределение N(5,2)
data = normrnd(c, a, [N 1]);

% Аппроксимируем нормальное распределение р-м Симпсона и подгоним параметры
% для распределения Симпсона
% Вначале с помощью метода ММП
phat = mle(data,'pdf',laplacepdf,'start',[1 1]);
c_mle = phat(1)
a_mle = phat(2)
% Для распределения Лапласа известны аналитические оценки по ММП
c_mle_theory = median(data)
deltac = c_mle - c_mle_theory
a_mle_theory = length(data)/sum(abs(data-c_mle_theory))
delta = a_mle - a_mle_theory
% Затем с помощью метода моментов
c_mm = mean(data)
a_mm = sqrt(2/var(data))

% Далее построим на одном графике истинную плотность и две плотности
```

```
% треугольного распределения с разными
% значениями оцениваемых параметров, найденные выше
figure;
y1 = normpdf(x,c,a); % Нормальная плотность
plot(x,y1); grid on; hold on;
y2 = laplacepdf(x,c_mle,a_mle); % Плотность треуг. распр. с параметрами ММП
plot(x,y2, 'k');
y3 = laplacepdf(x,c_mm,a_mm);  % Плотность треуг. распр. с параметрами ММ
plot(x,y3, 'r');
legend({'Normal', 'Laplace - MLE', 'Laplace - Moment'});
% Пример на нормальное распределение
function norm_ex(c,a,N,x)

% Сгенеририруем нормальное распределение N(5,2)
data = normrnd(c, a, [N 1]);

% Аппроксимируем нормальное распределение нормальным и подгоним параметры
% Вначале с помощью метода ММП
phat = mle(data,'pdf',@normpdf,'start',[1 1]);
c_mle = phat(1)
a_mle = phat(2)
% Затем с помощью метода моментов
c_mm = mean(data)
a_mm = sqrt(var(data))

% Далее построим на одном графике истинную плотность и две плотности
% треугольного распределения с разными
% значениями оцениваемых параметров, найденные выше
figure;
y1 = normpdf(x,c,a); % Нормальная плотность
plot(x,y1); grid on; hold on;
y2 = normpdf(x,c_mle,a_mle); % Плотность треуг. распр. с параметрами ММП
plot(x,y2, 'k');
y3 = normpdf(x,c_mm,a_mm);  % Плотность треуг. распр. с параметрами ММ
plot(x,y3, 'r');
legend({'Normal', 'Normal - MLE', 'Normal - Moment'});
```

Приложение 4 Проверка гипотезы о виде плотности распределения

Критерий “хи - квадрат”

Из генеральной совокупности *X*, образованной случайной величиной *X*, извлечена выборка $x_1, x_2, ..., x_n$. Выдвигается предположение о том, что плотность распределения случайной величины есть $\varphi(\theta, x)$, где θ - вектор параметров. По выборочным данным вычисляются оценки параметров θ и проверяется сложная гипотеза

H_0: плотность распределения случайной величины *X* есть $\varphi(\theta, x)$

против альтернативы

H_1: плотность распределения случайной величины *X* не $\varphi(\theta, x)$.

Поскольку эта гипотеза сложная, задается только вероятность ошибки первого рода – уровень значимости.

Для проверки сформулированной гипотезы естественно построить оценку плотности распределения – гистограмму и сопоставить ее с предполагаемой плотностью распределения. На рис. приведен пример гистограммы и кривая предполагаемой плотности распределения $\varphi(\theta, x)$, которая построена после того, как по выборочным значениям вычислены оценки θ ее параметров .

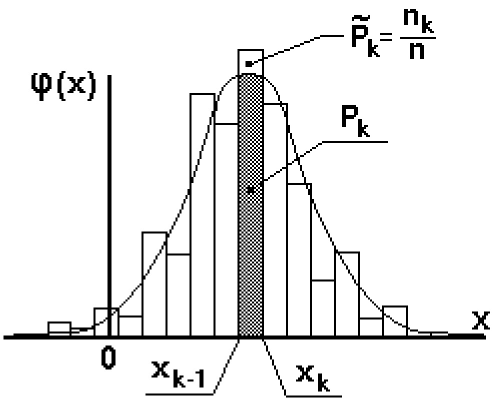


Рис. Иллюстрация к использованию критерия Хи-квадрат

Степень различия между гистограммой и предполагаемой плотностью распределения или статистика критерия выражается суммой квадратов разностей

$$\sum_{k=1}^{K}\left(P_k - \tilde{P}_k\right)^2 = \sum_{k=1}^{K}\left(\frac{nP_k - n_k}{n}\right)^2,$$

где

$$P_k = \int_{x_{k-1}}^{x_k} \varphi(\tilde{\theta}, x)dx = F(x_k) - F(x_{k-1})$$

то есть вероятность попадания значения случайной величины в интервал $(x_{k-1}, x_k]$ при условии справедливости нулевой гипотезы, $\tilde{P}_k = n_k / n$ - оценки этих вероятностей, где n_k - количество выборочных значений, попавших в интервал $(x_{k-1}, x_k]$, n - объем выборки, *K* - общее количество интервалов, на которых построена гистограмма.

При больших значениях *N* можно показать, что введенная статистика критерия принадлежит распределению Хи-квадрат с *K* – *r* степенями свободы.

$$\sum_{k=1}^{K}\frac{\left(n\cdot P_k-n_k\right)^2}{nP_k}\in\chi^2(K-r)$$

При заданной вероятности ошибки первого рода α (уровень значимости), критическое значение $z_{крит}$ назначается исходя из следующих предпосылок: при справедливости нулевой гипотезы маловероятно, чтобы статистика критерия оказалась слишком большой, поэтому достаточно задать критическое значение таким, что вероятность превышения его не больше α. Для этого достаточно вычислить квантиль $\chi^2_{1-\alpha}(K-r)$ и использовать ее в качестве значения $z_{крит}$. Вероятность ложного срабатывания, т.е. принятия гипотезы H_1 при условии, что на самом деле верна гипотеза H_0 будет равна α.

Полученный критерий называется критерием “хи - квадрат” (Пирсона) проверки гипотезы о виде плотности распределения (или закона распределения) генеральной совокупности по экспериментальным данным.

Процедура проверки гипотезы о виде плотности распределения по критерию “хи - квадрат”.

1. Задается уровень значимости α

2. По выборочным данным строится гистограмма в соответствии с указаниями.

3. Вычисляются точечные оценки моментов.

4. Из теоретических соображений, по виду гистограммы, по соотношениям между моментами, по значениям асимметрии и эксцесса, по другим соображениям выдвигается гипотеза о виде плотности распределения $\varphi(\theta,x)$.

5. Вычисляются оценки $\tilde{\theta}$ параметров предполагаемой плотности распределения, в результате получается плотность распределения $\varphi(\tilde{\theta},x)$.

6. С использованием $\varphi(\tilde{\theta},x)$ или $F(\tilde{\theta},x)$ вычисляются вероятности

$$P_k=\int_{x_{k-1}}^{x_k}\varphi(\tilde{\theta},x)dx=F(\tilde{\theta},x_k)-F(\tilde{\theta},x_{k-1}).$$

7. Вычисляется статистика критерия

$$\chi^2=\sum_{k=1}^{K}\frac{\left(n\cdot P_k-n_k\right)^2}{nP_k}.$$

8. Полученное значение сравнивается с критическим значением

$\chi^2_{1-\alpha}(K-r)$,

где r - количество оцениваемых параметров.

9. Если $\chi^2>\chi^2_{1-\alpha}(K-r)$ делается вывод о том, что экспериментальные данные не подтверждают справедливость выдвинутой гипотезы или о том, что отсутствуют достаточные основания для того, чтобы считать нулевую гипотезу справедливой. Гипотеза пересматривается, выдвигается новая нулевая гипотеза, переход на п. 4 настоящей процедуры.

10. Если $\chi^2<\chi^2_{1-\alpha}(K-r)$ делается вывод о том, что экспериментальные данные подтверждают справедливость выдвинутой гипотезы или о том, что имеются достаточные основания для того, чтобы считать нулевую гипотезу справедливой.

Зам. С уменьшением вероятности α возрастает критическое значение $\chi^2_{1-\alpha}(K-r)$, а это значит, что объективно вероятность β пропуска события (т.е. принятия гипотезы H_0 при условии, что верная гипотеза H_1). Действительно, если задать α = 0, то критическое значение $\chi^2_1(K-r)=\infty$, а это означает, что нулевая гипотеза будет всегда приниматься и ошибка второго рода β будет равна 1.

Критерий Колмогорова - Смирнова

Из генеральной совокупности, образованной случайной величиной *X*, извлечена выборка $x_1, x_2, ..., x_n$. Выдвигается предположение о том, что функция распределения случайной величины есть $F(\theta, x)$, где θ - вектор параметров. По выборочным данным вычисляются оценки параметров $\tilde{\theta}$ и проверяется сложная гипотеза

H_0: функция распределения случайной величины *X* есть $F(\tilde{\theta}, x)$

против альтернативы

H_1: функция распределения случайной величины *X* не $F(\tilde{\theta}, x)$.

Поскольку эта гипотеза сложная, задается только вероятность ошибки первого рода α - уровень значимости.

В соответствии с формулировкой гипотезы сравниваются две функции распределения: выборочная и предполагаемая, представленные на рис. Различие между ними определено, как

$$D = \sup_i \left| \tilde{F}(x_i) - F(\tilde{\theta}, x_i) \right|,$$

где $\tilde{F}(x_i)$ - значения выборочной функции распределения при $x = x_i$.

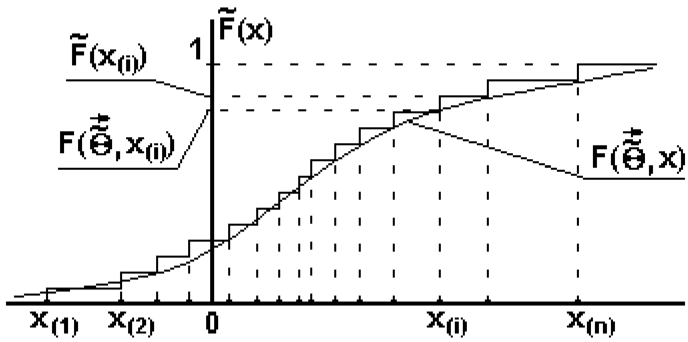


Рис. Выборочная и предполагаемая функции распределения

Статистикой критерия является величина *D*. Критические значения для различных значений *α* табулированы и приводятся практически во всех учебниках и справочниках по математической статистике. В таблице ниже приводятся некоторые часто употребляемые критические значения.

Таблица 4

Критические значения критерия Колмогорова-Смирнова

α \ n	25	50	80	100
0.2	0.208	0.148	0.118	0.106
0.1	0.238	0.169	0.135	0.121
0.05	0.264	0.188	0.150	0.134

Если *n* > 10, для расчета критических значений можно пользоваться приближенной формулой

$$D_\alpha = \sqrt{-\frac{\ln(0.5 \cdot \alpha)}{2 \cdot n}} - \frac{1}{6n}.$$

Процедура проверки гипотезы о виде функции распределения по критерию Колмогорова - Смирнова.

1. Задается уровень значимости α.
2. По выборочным данным строится выборочная функция распределения.

3. Вычисляются точечные оценки моментов.

4. Из теоретических и практических соображений (вид выборочной функции распределения, гистограммы, соотношения между моментами, значениям асимметрии и эксцесса) выдвигается гипотеза о виде функции распределения $F(\theta, x)$ и тем самым - о виде плотности распределения $\varphi(\theta, x)$.

5. Оценивается *r* параметров $\hat{\theta}$ предполагаемой функции распределения и ее значения $F(\hat{\theta}, x_i)$ при $x = x_i$.

6. Вычисляется статистика критерия $D = \sup_i \left| \tilde{F}(x_i) - F(\tilde{\theta}, x_i) \right|$

7. Полученное значение сравнивается с критическим значением D_α.

8. Если $D > D_\alpha$, делается вывод о том, что экспериментальные данные не подтверждают справедливость выдвинутой гипотезы или о том, что отсутствуют достаточные основания для того, чтобы считать нулевую гипотезу справедливой. Гипотеза пересматривается, выдвигается новая нулевая гипотеза, переход на п. 4 настоящей процедуры.

9. Если $D \le D_\alpha$ делается вывод о том, что экспериментальные данные подтверждают справедливость выдвинутой гипотезы или о том, что имеются достаточные основания для того, чтобы считать нулевую гипотезу справедливой.

Зам. Для корректного применения критерия Колмогорова - Смирнова выборку $x_1, x_2, ..., x_n$. следует разделить на две части и по одной из них оценить параметры $\tilde{\theta}$, а по другой построить выборочную функцию распределения и вычислить статистику критерия *D*. Это позволяет избавиться от необходимости учета зависимости между выборочными значениями, которая появляется в результате вычисления параметров предполагаемой плотности распределения, как это было в случае применения критерия χ^2.

Критерий w² Мизеса

Из генеральной совокупности, образованной случайной величиной *X*, извлечена выборка $x_1, x_2, ..., x_n$. Выдвигается предположение о том, что функция распределения случайной величины есть $F(\theta, x)$, где θ - вектор параметров. По выборочным данным вычисляются оценки параметров $\tilde{\theta}$ и проверяется сложная гипотеза

H_0: функция распределения случайной величины *X* есть $F(\tilde{\theta}, x)$

против альтернативы

H_1: функция распределения случайной величины *X* не $F(\tilde{\theta}, x)$.

Поскольку эта гипотеза сложная, задается только вероятность ошибки первого рода α, которая в подобных случаях именуется уровнем значимости.

В соответствии с формулировкой гипотезы сравниваются две функции распределения: выборочная (п. 2.2) и предполагаемая. Различие между ними определено, как

$$\omega^2 = \int_{-\infty}^{\infty} \left| \tilde{F}(x) - F(\tilde{\theta}, x) \right|^2 \varphi(\tilde{\theta}, x) \cdot dx,$$

где $\varphi(\tilde{\theta}, x)$ - предполагаемая плотность распределения.

Этот интеграл вычисляется, как сумма интегралов по интервалам между соседними членами вариационного ряда. Если на этих интервалах предполагаемая функция распределения интерполируется прямой линией, то этот интеграл выражается суммой

$$\omega^2 = \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left| \tilde{F}(x) - F(\tilde{\theta}, x) \right|^2 \varphi(\tilde{\theta}, x) \cdot dx = \frac{1}{12n^2} + \frac{1}{n} \cdot \sum_{i=1}^{n} \left[F(\tilde{\theta}, x_i) - \frac{2i-1}{2n} \right]^2.$$

В качестве статистики критерия используется

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F\left(\tilde{\theta}, x_i\right) - \frac{2i-1}{2n}\right]^2$$

Критические значения $(nw^2)_\alpha$, табулированные в таблицах математической статистики, в таблице ниже приводятся некоторые часто употребляемые критические значения.

Таблица 5

Критические значения критерия ω^2 Мизеса

α	0.03	0.05	0.1	0.2
$(nw^2)_\alpha$	0.55	0.4614	0.3473	0.2415

Процедура проверки гипотезы о виде функции распределения по критерию ω^2 Мизеса.

1. Задается уровень значимости α
2. По выборочным данным строится выборочная функция распределения.
3. Вычисляются точечные оценки моментов.
4. Из теоретических соображений выдвигается гипотеза о виде функции распределения $F(\theta, x)$ и тем самым - о виде плотности распределения $\varphi(\theta, x)$.
5. Оценивается *r* параметров $\vec{\theta}$ предполагаемой функции распределения и ее значения $F(\hat{\theta}, x_i)$ при $x = x_i$.
6. Вычисляется статистика критерия

$$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[F\left(\tilde{\vec{\Theta}}, x_i\right) - \frac{2i-1}{2n}\right]^2$$

7. Полученное значение сравнивается с критическим значением $(nw^2)_\alpha$.
8. Если $n\omega^2 > n\omega^2{}_\alpha$ делается вывод о том, что экспериментальные данные не подтверждают справедливость выдвинутой гипотезы или о том, что отсутствуют достаточые основания для того, чтобы считать нулевую гипотезу справедливой. Гипотеза пересматривается, выдвигается новая нулевая гипотеза, переход на п. 4 настоящей процедуры.
9. Если $n\omega^2 \le n\omega^2{}_\alpha$ делается вывод о том, что экспериментальные данные подтверждают справедливость выдвинутой гипотезы или о том, что имеются достаточые основания для того, чтобы считать нулевую гипотезу справедливой.

Критерий ω^2 Мизеса - равномерно наиболее мощный критерий проверки гипотезы о виде функции распределения.

Расчетное задание 3

Аппроксимация результатов измерений зависимых переменных.

Исходные данные

В результате измерений при значениях независимой переменной

$x_1, x_2, ... x_k$

получены следующие данные:

$y11, y21, ..., yk1$

$y12, y22, ..., yk2$

....

$y1n, y2n, ..., ykn$

Варианты

Для каждого студента есть свой вариант в виде текстового файла "\Tasks\Task_3.txt", находящегося в личной папке файлового хранилища. В этом файле содержатся выборочные значения в следующем формате:

```
nx = ##
ny = ##
x= ## ## ## … ##
y[##] = ## ## ## … ##
……………………
y[##] = ## ## ## … ##
```

где «nx» – число точек, в которых проводились измерения, «ny» – количество измерений в каждой точке, «x» – сами точки, «y[##]» – все измерения (n штук) в соответствующей точке.

Вся теоретическая часть по работе изложена в [1], а также в разделах помощи Matlab, в частности Statistic Toolbox.

Задание

1. Вычислить в каждой точке $\boldsymbol{xi}$ средние арифметические значения $\overline{\boldsymbol{yi}}$, оценки дисперсий $\boldsymbol{s}_i^2$, параметрические толерантные пределы для погрешностей, доверительные интервалы для математических ожиданий, проверить гипотезу о равенстве дисперсий в этих точках по критерию Кочрена (см. приложение 3);

2. Произвести последовательную полиномиальную аппроксимацию Прим. В качестве значений y при аппроксимации необходимо использовать средние арифметические значения $\overline{\boldsymbol{yi}}$ (см. приложение 3).

2.1. Начать с нулевой степени полинома

2.2. вычислить оценки коэффициентов $\hat{a}$ полинома МНК или МНД (в зависимости от исхода проверки гипотез о равенстве дисперсий) для заданной степени полинома.

Зам. При определении $\hat{a}$ требуется обращать матрицу S_ε. Если n>k, то можно посчитать обратную матрицу S_ε^{-1} по исходной матрице S_ε, поскольку определитель S_ε не равен 0. В противном случае в качестве S_ε используется диагональная матрица с дисперсиями в каждой из k точек на диагонали.

2.3. Проверить гипотезу о степени q полинома, и если она не будет отвергнута, оценить дисперсии s_{ai}^2 и ковариационную матрицу оценок коэффициентов S_a, в противном случае увеличить степень полинома. Для проверки гипотезы используется критерий Фишера. Если число измерений n больше величины $k-q-1$, то статистикой критерия является выражение $F = \dfrac{(n-k+q+1)}{(n-q-1)(n-1)}R^2$, в противном случае $F = \dfrac{R^2}{(k-q-1)}$.

Зам. Размерность S_a равна (q+1)*(q+1), в то время как размерность S_ε равна k*k.

2.4. Вычислить корреляционную матрицу R_a и коэффициенты корреляции $r_a(i,j)$ между оценками коэффициентов по матрице ковариации

2.5. Пусть была получена степень q полинома, прошедшая гипотезу о степени полинома. Произвести все те же действия для полинома степени, равной k-1 (вычислить коэффициенты и корреляцию между ними). Сравнить результаты для степени q и k-1 (качество аппроксимации, корреляционная матрица коэффициентов, матрица ковариации исходных данных S_ε и ее обусловленность). См указания в приложении.

3. Произвести аппроксимацию исходной зависимости другими способами. Представить полученные графики аппроксимации (полученная аппроксимирующая кривая одним цветом, точки, по которым проводилась аппроксимация маркерами одного типа и все исходные точки маркерами другого типа). Проанализировать и сравнить полученные результаты.

3.1 Произвести аппроксимацию зависимости прямой линией с помощью функций regress (использует метод R-Square), robustfit (робастная регрессия), polyfit (полиномиальная регрессия с n=1), ridge (ридж-регрессия с регуляризацией). Проанализировать полученные результаты.

3.2. Произвести полиномиальную аппроксимацию с помощью функций polyfit (polyval). Можно воспользоваться утилитой polytool, являющейся графическим интерфейсом к polyfit. Подобрать степень полинома, наилучшим способом аппроксимирующую исходную зависимость.

3.3. Произвести кусочную полиномиальную аппроксимацию с помощью функций interp1 (линейная, кубическая), pchip (полиномами Эрмита), spline (сплайны). Сравнить качество аппроксимации с предыдущими результатами.

3.4 Произвести нелинейную аппроксимацию с помощью функции nlinfit. В качестве нелинейной функции использовать произведение полинома на гармоническую функцию:

$$y(x) = (\sin \alpha x + \beta)(a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0)$$

4. В выводах детально сравнить все использованные способы аппроксимации зависимостей, выделить преимущества и недостатки каждого из методов в смысле качества аппроксимации, трудоемкости вычислений и других факторов.

Приложение 1

Указания к полиномиальной аппроксимации

- **Внимание!** *Степень полинома a-priori не известна*, поэтому необходимо начинать попытки аппроксимации с наименьшей степени и проверять гипотезу о степени полинома. При отклонении гипотезы увеличить степень аппроксимирующего полинома на единицу и проделать указанную процедуру вновь.
- Приводить все промежуточные результаты, получаемые при каждой попытке.
- Пусть гипотеза оказалась не отвергнутой при степени **q**. Тогда следует задать степень полинома равной **k-1** и вновь вычислить оценки коэффициентов этого полинома и ковариационную матрицу оценок коэффициентов. Сопоставить с предыдущими результатами и прокомментировать возможное увеличение дисперсий оценок коэффициентов и ухудшение обусловленности задачи.

Представление результатов

- числовое - результаты, полученные на каждой итерации при каждом значении степени полинома q, в том числе, значения статистики критерия проверки степени полинома и критические значения критерия,
- графическое - изобразить все точки, толерантные пределы и границы доверительных интервалов, средние значения и полученную аппроксимирующую полиномиальную функцию, причем эту функцию, границы доверительных интервалов и толерантные пределы соединить плавной линией; там же нанести функцию, полученную при безызбыточной аппроксимации полиномом степени k-1.

Приложение 2 Оценивание коэффициентов аппроксимирующих полиномов

Измерения однократные

В соответствии с формулировкой задачи в результате эксперимента при фиксированных значениях $x_1, x_2, ..., x_k$ мы получаем значения $y_{\varepsilon_1}, y_{\varepsilon_2}, ..., y_{\varepsilon_k}$:

$$\begin{cases} y_{\varepsilon_1} = y_1 + \varepsilon_1 = a_0 + a_1 x_1 + a_2 x_1^2 + ... + a_q x_1^q + \varepsilon_1 \\ y_{\varepsilon_2} = y_2 + \varepsilon_2 = a_0 + a_1 x_2 + a_2 x_2^2 + ... + a_q x_2^q + \varepsilon_2 \\ \vdots \\ y_{\varepsilon_k} = y_k + \varepsilon_k = a_0 + a_1 x_k + a_2 x_k^2 + ... + a_q x_k^q + \varepsilon_k \end{cases}$$

Таким образом к каждому значению y_k, полученному в результате полиномиальной аппроксимации в точке x_k подмешивается случайная составляющая ε_k. Эти значения представлены на рис. точками, лежащими вне кривых

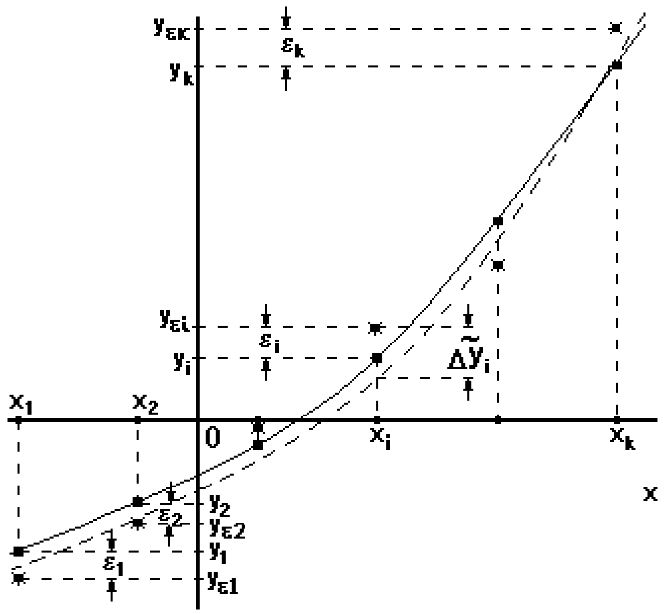


Рис. Иллюстрация к задаче полиномиальной аппроксимации

Приведенная система равенств есть система k уравнений, из которой необходимо получить оценки $q + 1$ коэффициентов $a_0, a_1, ..., a_q$.

Для того, чтобы эту систему записать в матричном виде, введем матрицу

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^q \\ 1 & x_2 & x_2^2 & \cdots & x_2^q \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_k & x_k^2 & \cdots & x_k^q \end{pmatrix}.$$

Тогда система уравнений записывается в виде

$$y_\varepsilon = Xa + \varepsilon, \ \varepsilon \in N_k(0, \Sigma_\varepsilon), \ y_\varepsilon \in N_k(Xa, \Sigma_\varepsilon).$$

В результате применения метода максимального правдоподобия к выборке СВ y_ε можно получить ММП-оценки коэффициентов аппроксимирующего полинома:

$\hat{a} = \left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1} X^T \Sigma_\varepsilon^{-1} y_\varepsilon$. – оценка a,

$\Sigma_{\hat{a}} = \left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1}$ – ковариационная матрица оценки а,

$\hat{a} \in N_{q+1}\left(a, \left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1}\right)$ – распределение оценки а.

Минимальное значение, которое принимает минимизируемая квадратичная форма $(Xa - y_\varepsilon)^T \Sigma_\varepsilon^{-1}(Xa - y_\varepsilon)$ обозначается через R^2 и равно

$$R^2 = \left(X\hat{a} - y_\varepsilon\right)^T \Sigma_\varepsilon^{-1}\left(X\hat{a} - y_\varepsilon\right).$$

R^2 является случайной величиной и имеет распределение Хи-квадрат:

$$R^2 \in \chi^2(k - q - 1).$$

Равноточные измерения

Если измерения независимые, то ковариационная матрица Σ_ε диагональная. Если при этом измерения равноточные, т.е при всех i = 1, 2, ..., k $\sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2$, то

$$\Sigma_\varepsilon = \sigma_\varepsilon^2 \cdot E\,,\ \Sigma_\varepsilon^{-1} = \sigma_\varepsilon^{-2} \cdot E$$

где E - единичная матрица, а коэффициенты аппроксимирующего полинома имеют следующие средние значения

$$\hat{a} = \left(X^T X\right)^{-1} X^T y_\varepsilon$$

ковариационную матрицу

$$\Sigma_{\hat{a}} = \sigma_\varepsilon^2 \left(X^T X\right)^{-1}$$

и принадлежат нормальному распределению

$$a \in N_{q+1}\left(\hat{a}, \sigma_\varepsilon^2 \left(X^T X\right)^{-1}\right).$$

Измерения многократные, характеристики погрешностей измерений известны

Если измерения многократные, то каждое значение y_i, соответствующее x_i, измеряется несколько раз. В результате можно вычислить средние значения:

$$\bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}\,,$$

Погрешности измерений известны и задаются ковариационной матрицей Σ_ε. Тогда ковариационная матрица усредненных значений равна Σ_ε / n Как и в предыдущем случае, возможны два варианта – измерения равноточные или неравноточные.

а) Если измерения равноточные, то $\Sigma_\varepsilon^{-1} = \sigma_\varepsilon^{-2} E$ метод называется МНК, и в результате его применения получаются следующие параметры оцениваемых коэффициентов:

$$\hat{a} = \left(X^T X\right)^{-1} X^T \bar{y}\,,$$

$$\Sigma_{\hat{a}} = \frac{1}{n}\sigma_\varepsilon^2 \left(X^T X\right)^{-1}$$

принадлежащие нормальному распределению

$$a \in N_{q+1}\left(\hat{a}, \frac{1}{n}\sigma_\varepsilon^2 \left(X^T X\right)^{-1}\right)$$

Минимальное значение квадратичной формы равно

$$R^2 = n \cdot \sigma_\varepsilon^{-2} \cdot \left(\bar{y} - X\hat{a}\right)^T \left(\bar{y} - X\hat{a}\right)$$

и имеет распределение Хи-квадрат:

$$R^2 \in \chi^2(k - q - 1).$$

б) Если измерения неравноточные, то применяется т.н. обобщенный МНК (ОМНК), в котором используется полная ковариационная матрица. В результате получаются следующие оценки коэффициентов полинома

$$\hat{a} = \left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1} X^T \Sigma_\varepsilon^{-1} \bar{y},$$

$$\Sigma_{\hat{a}} = \frac{1}{n}\left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1},$$

принадлежащие нормальному распределению

$$\vec{\hat{a}} \in N_{q+1}\left(\vec{a}, \frac{1}{n}\left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1}\right).$$

Минимальное значение квадратичной формы равно

$$R^2 = n \cdot \left(\vec{y} - X\vec{a}\right)^T \Sigma_\varepsilon^{-1} \left(\vec{y} - X\vec{a}\right),$$

и имеет распределение Хи-квадрат:

$$R^2 \in \chi^2(k - q - 1).$$

Измерения многократные, характеристики погрешностей измерений неизвестны.

В данном случае в отличие от предыдущего ковариационная матрица погрешностей измерений не известна и оценивается по самим измерениям. Введем следующие обозначения:

q – количество коэффициентов полинома

k – количество различных точек, в которых измеряются данные

n – количество измерений в каждой из k точек

Элементы ковариационной матрицы можно приближенно посчитать по известной формуле:

$$s_{\varepsilon i,j} = \frac{1}{n-1} \cdot \sum_{k=1}^{n} \left(y_{ik} - \bar{y}_i\right)\left(y_{jk} - \bar{y}_j\right)$$

Тогда получаются следующие оценки коэффициентов полинома

$$\hat{a} = \left(X^T S_\varepsilon^{-1} X\right)^{-1} X^T S_\varepsilon^{-1} \bar{y},$$

$$S_{\hat{a}} = \frac{1}{n}\left(X^T S_\varepsilon^{-1} X\right)^{-1},$$

принадлежащие нормальному распределению

$$a \in N_{q+1}\left(\hat{a}, \frac{1}{n}\left(X^T S_\varepsilon^{-1} X\right)^{-1}\right),$$

Минимальное значение квадратичной формы равно

$$R^2 = n \cdot \left(\bar{y} - Xa\right)^T S_\varepsilon^{-1} \left(\bar{y} - Xa\right)$$

и при домножении на константу имеет распределение Фишера:

$$\frac{(n - k + q + 1)}{(k - q - 1)(n - 1)} \cdot R^2 \in F(k - q - 1, n - k + q + 1) \ .$$

Упрощенный случай (n<k-q-1)

Когда количество измерений невелико оцениваются только дисперсии $\sigma_{\varepsilon i}^2$ при каждом измерении:

$$s_{\varepsilon_i}^2 = \frac{1}{n-1} \cdot \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2$$

Затем формируется диагональная ковариационная матрица S_ε, на диагонали которой стоят найденные дисперсии:

$$S_{\varepsilon}=\begin{pmatrix} s_{\varepsilon_1}^2 & 0 & \cdots & 0 \\ 0 & s_{\varepsilon_2}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & s_{\varepsilon_k}^2 \end{pmatrix},$$

и применяется при вычислении оценок коэффициентов.

В этой ситуации F-распределению Фишера подчиняется случайная величина

$$\frac{R^2}{(k-q-1)} \in F(k-q-1, n-1)$$

с числом степеней свободы $k - q - 1$ и $n - 1$.

Равноточные измерения

В частном случае, когда по результатам проверки по критерию Кочрена гипотезы о равенстве дисперсий $\sigma_{\varepsilon,i}^2$ принимается решение о применении МНК, тогда вычисляется средняя оценка дисперсии по k дисперсиям в каждой точке:

$$s_{\varepsilon}^2=\frac{1}{k}\sum_{i=1}^{k}s_{\varepsilon,i}^2 ,$$

которая подставляется вместо σ_{ε}^2 во всех соответствующих формулах. В результате получаются следующие оценки коэффициентов полинома

$$\hat{a}=\left(X^T X\right)^{-1} X^T \bar{y},$$

$$S_{\hat{a}}=\frac{1}{n}s_{\varepsilon}^2\left(X^T X\right)^{-1},$$

принадлежащие нормальному распределению.

$$a \in N_{q+1}\left(\hat{a},\frac{1}{n}s_{\varepsilon}^2\left(X^T X\right)^{-1}\right).$$

Минимальное значение квадратичной формы равно

$$R^2 = n\cdot s_{\varepsilon}^{-2}\cdot(\bar{y}-X\hat{a})^T(\bar{y}-X\hat{a})$$

и в этом случае случайная величина

$$\frac{R^2}{(k-q-1)} \in F(k-q-1, n-1) \ .$$

распределена в соответствии с F-распределением Фишера с числом степеней свободы $k - q - 1$ и $n - 1$.

Особенности вычислений при реализации МНК и ОМНК

В процессе оценивания коэффициентов аппроксимирующих полиномов приходится обращать матрицы $X^T X$ и $X^T\Sigma_{\varepsilon}^{-1}X$. Устойчивость результатов подобных действий в сильной степени зависит от обусловленности обращаемых матриц. Что касается матриц Σ_{ε} и S_{ε}, то они вполне могут оказаться особенными.

Обусловленность матриц характеризуется числом обусловленности, которое есть не что иное, как коэффициент “отношения” погрешностей экспериментальных данных и погрешностей округления к погрешностям результатов вычислений.

Для квадратных симметричных матриц, каковыми являются матрицы, перечисленные выше, число обусловленности определено равенствами

$$cond\left(X^T X\right)=\sqrt{\frac{\lambda_{\max}\left(X^T X\right)}{\lambda_{\min}\left(X^T X\right)}},\ cond\left(X^T\Sigma_{\varepsilon}^{-1}X\right)=\sqrt{\frac{\lambda_{\max}\left(X^T\Sigma_{\varepsilon}^{-1}X\right)}{\lambda_{\min}\left(X^T\Sigma_{\varepsilon}^{-1}X\right)}}$$

где $\lambda_{\max}(\cdot)$, $\lambda_{\min}(\cdot)$ - наибольшее и наименьшее собственные числа соответствующей матрицы.

Число обусловленности матриц, используемых в МНК или в ОМНК, может достигать значений $10^6 \div 10^{12}$ и выше.

Известно, что число обусловленности указанных матриц монотонно возрастает с увеличением количества столбцов матрицы X, то есть с увеличением порядка q или, что то же самое, с увеличением числа коэффициентов полинома. Максимального значения число обусловленности достигает при $q + 1 = k$. Особенно опасной оказывается ситуация, когда количество оцениваемых коэффициентов превышает их фактическое количество, то есть при завышении степени полинома.

Можно рекомендовать три способа повышения устойчивости оценок коэффициентов МНК и ОМНК.

1. Не стремиться к излишне высокому порядку аппроксимирующего полинома, использовать априорную информацию о гладкости аппроксимируемой функции.

2. При необходимости аппроксимации функции $y = f(x)$ полиномом высокого порядка вплоть до $q = k - 1$ использовать метод регуляризации А.Н.Тихонова.

3. Третий способ заключается в таком размещении значений аргумента x, при котором число обусловленности матрицы $\left(X^T X\right)$ или $\left(X^T \Sigma_{\varepsilon}^{-1} X\right)$ было минимальным – оптимальное планирование.

Регуляризация

Исходное уравнение вычисления коэффициентов преднамеренно искажается таким образом, чтобы это искажение заведомо улучшало обусловленность. Таким регуляризирующим искажением является, по Тихонову, увеличение диагональных элементов матрицы системы уравнений:

$$\hat{a}_{\alpha} = (X^T \Sigma_{\varepsilon}^{-1} X + \alpha E)^{-1} X^T \Sigma_{\varepsilon}^{-1} \vec{y}_{\varepsilon}, \ \alpha > 0 .$$

Число α называется параметром регуляризации. Оценка $\hat{a}_{\alpha}$ в литературе называется ридж-оценкой. Эта оценка смещена. Ее смещение примерно равно

$$\Delta a \approx \alpha \left(X^T \Sigma_{\varepsilon}^{-1} X\right)^{-1} X^T \Sigma_{\varepsilon}^{-1} y_{\varepsilon} .$$

Проблема состоит выборе такого значения параметра регуляризации, при котором результирующая погрешность, вызванная смещением оценки и плохой обусловленностью матрицы системы, была минимальной.

На рис. показана принципиальная возможность такого выбора. Однако, универсального практического рецепта выбора оптимального значения параметра регуляризации пока не существует.

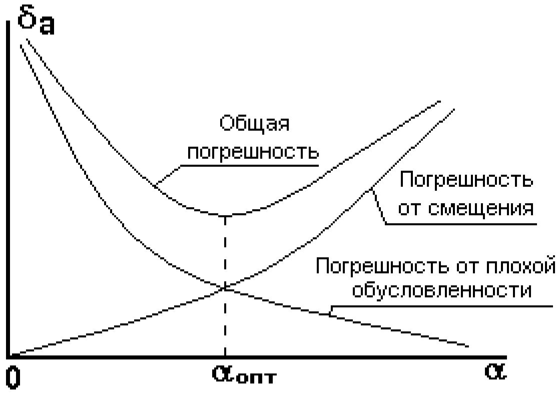


Рис. Зависимость погрешности оценки от выбора параметра регуляризации

Оптимальное планирование

Из конструкции матрицы X видно, что ее элементы изменяют свои значения в зависимости от значений аргумента полинома: $x_1, x_2, ..., x_k$. Поэтому, возможно, существует такой план расстановки этих значений, при котором погрешности в каком-либо смысле будут минимальными.

Такие планы, действительно, существуют:

A - оптимальный план эксперимента - план, при котором достигается минимум следа матрицы $X^T X$, то есть минимум суммы ее диагональных элементов.

D - оптимальный план эксперимента - план, при котором достигается минимальное значение определителя матрицы $X^T X$.

C - оптимальный план эксперимента - план, при котором достигается минимальное значение числа обусловленности матрицы $X^T X$. C - оптимальный план эквивалентен D - оптимальному плану.

Расширение класса аппроксимирующих полиномов

При аппроксимации возможно использование не только обычных, но и т.н. обобщенных аппроксимирующих полиномов, когда вместо степенных используются какие-либо другие произвольные функции:

$$y(x) = a_0 + a_1\varphi_1(x) + a_2\varphi_2(x) + ... + a_q\varphi_q(x),$$

где $\{\varphi_i(x)\}, i = 0,1,2,...,q$ - система базисных функций.

Если эти функции ортогональны и соответствуют характеру аппроксимируемой зависимости лучше, чем степени *x*, то для достижения необходимой точности аппроксимации может понадобиться меньше членов, чем в случае аппроксимации степенным полиномом. А это обстоятельство способствует улучшению обусловленности задачи и является четвертым средством повышения устойчивости оценок МНК и ОМНК.

Матрица X в этом случае будет иметь следующую конструкцию:

$$X = \begin{pmatrix} 1 & \varphi_1(x_1) & \cdots & \varphi_q(x_1) \\ 1 & \varphi_1(x_2) & \cdots & \varphi_q(x_2) \\ \vdots & \vdots & \cdots & \vdots \\ 1 & \varphi_1(x_k) & \cdots & \varphi_q(x_k) \end{pmatrix}.$$

Это единственное отличие, а все остальные формулы, замечания и рекомендации остаются в силе без каких-либо изменений.

Приложение 3 Проверка гипотез при полиномиальной аппроксимации

Критерий Кочрена проверки гипотезы о равенстве дисперсий

Данный критерий проверяет равноточность измерений (равенство дисперсий) в различных точках. В случае равноточности используется простой метод МНК, в противном случай – более сложный ОМНК.

Суть критерий сводится к следующему.

Имеются *k* выборок $y_{i1}, y_{i2}, ..., y_{in}, i = 1,2,...,k$ объемом n каждая. Выборки изъяты из нормальных генеральных совокупностей, дисперсии которых $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$. По выборочным значениям вычислены оценки математических ожиданий и дисперсий

$$\bar{y}_i = \frac{1}{n} \cdot \sum_{j=1}^{n} y_{ij}, \; s_i^2 = \frac{1}{n-1} \cdot \sum_{j=1}^{n} \left(y_{ij} - \bar{y}_i\right)^2.$$

Выдвигается гипотеза

H_0: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ против альтернативы $H_1 : \overline{H}_0$.

Статистика критерия Кочрена

$$g = \frac{s_{\max}^2}{\sum_{i=1}^{k} s_i^2}.$$

Критическое значение $g_\alpha(n-1, k)$ выбирается из таблиц критических значений критерия Кочрена.

Рассчитанное значение статистики критерия сравнивается с критическим значением. Если $g \le g_\alpha(n-1, k)$, делается вывод о том, что нулевая гипотеза не противоречит экспериментальным данным, и для аппроксимации применяется МНК.

В противном случае, если $g > g_\alpha(n-1, k)$, делается противоположный вывод, и для аппроксимации применяется ОМНК.

Проверка гипотезы о степени аппроксимирующего полинома, характеристики погрешностей измерений известны.

Измерения однократные

Выше методом ММП были получены результаты оценивания коэффициентов аппроксимирующих полиномов путем минимизации квадратичных форм:

- $(Xa - y_\varepsilon)^T (Xa - y_\varepsilon)$ - в случае равноточных измерений и применения МНК,
- $(Xa - y_\varepsilon)^T \Sigma_\varepsilon^{-1} (Xa - y_\varepsilon)$ - в случае неравноточных измерений и применения ОМНК.

Из-за случайности погрешностей измерений оценки $\hat{a}$ также случайны. Поэтому и минимальные значений квадратичных форм также случайны:

- $R^2 = \sigma_\varepsilon^{-2} (X\hat{a} - \vec{y}_\varepsilon)^T (X\hat{a} - y_\varepsilon)$ - в случае применения МНК
- $R^2 = (X\hat{a} - y_\varepsilon)^T \Sigma_\varepsilon^{-1} (X\hat{a} - y_\varepsilon)$ - в случае применения ОМНК.

Найденные оценки коэффициентов несмещены:

$$M[X\hat{a} - y_\varepsilon] = 0$$

Плотности распределения обеих квадратичных форм одинаковы: это плотность распределения “хи - квадрат” с (k - q - 1) степенями свободы, то есть

$$R^2 \in \chi^2(k - q - 1),$$

где q + 1 – количество оцениваемых коэффициентов.

На практике модель практически никогда не бывает известной. Тогда, при ошибочном назначении степени р аппроксимирующего полинома, меньшей, чем истинная степень q, оценки коэффициентов оказываются смещенными, поэтому

$$M[X\hat{a} - y_\varepsilon] \ne 0,$$

значения R^2 существенно возрастают, и плотность распределения изменяется.

Далее рассмотрим случай $p < q$, т.к в противном случае ухудшается обусловленность обращаемых матриц и теряется вычислительная устойчивость МНК и ОМНК.

Пусть в приведенных условиях при истинной степени аппроксимирующего полинома q была назначена степень $p < q$, и с помощью МНК или ОМНК получены оценки p + 1 коэффициента. Обозначим вектор найденных таким образом оценок через $\hat{a}_p$, а квадратичные формы, вычисленные при этих значениях оценок - через R_p^2.

Сформулируем гипотезу.

H_0: степень аппроксимирующего полинома $p = q$,

H_1: степень аппроксимирующего полинома $p < q$.

Статистикой критерия проверки этой гипотезы является квадратичная форма R_p^2, которая при $p = q$, то есть при справедливости нулевой гипотезы равна R^2 и распределена, как $\chi^2(k-q-1)$. Поэтому в качестве критического значения при заданной ошибке первого рода α квантиль $\chi^2_{1-\alpha}(k-q-1)$.

Процедура проверки гипотезы в условиях, когда заданы вид полинома и вероятность α следующая.

1. Оцениваются коэффициенты полинома

$\hat{a}_p = \left(X^T X\right)^{-1} X^T y_\varepsilon$ (МНК)

$\hat{a}_p = \left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1} X^T \Sigma_\varepsilon^{-1} y_\varepsilon$ (ОМНК).

2. Вычисляется статистика критерия

$R_p^2 = \sigma_\varepsilon^{-2}\left(X\hat{a}_p - y_\varepsilon\right)^T\left(X\hat{a}_p - y_\varepsilon\right)$

$R_p^2 = \left(X\hat{a}_p - y_\varepsilon\right)^T \Sigma_\varepsilon^{-1}\left(X\hat{a}_p - y_\varepsilon\right)$.

3. Значение статистики R_p^2 сравнивается с критическим значением:

- если $R_p^2 \le \chi^2_{1-\alpha}(k-q-1)$, делается вывод о том, что нет достаточных оснований для того, чтобы отвергнуть нулевую гипотезу,

- в противном случае, если $R_p^2 > \chi^2_{1-\alpha}(k-q-1)$, делается вывод о том, что нет достаточных оснований для того, чтобы считать нулевую гипотезу справедливой.

Измерения многократные

Этот случай отличается от предыдущего тем, что в качестве исходных данных для полиномиальной аппроксимации используется не вектор результатов однократных измерений $\vec{y}_\varepsilon$, а вектор средних арифметических значений $\vec{\bar{y}}$ результатов многократных измерений.

Тогда, если для аппроксимации назначена степень полинома *p*, и $\hat{a}_p$ - вектор оценок (р + 1) коэффициента этого полинома, то статистикой критерия проверки гипотезы о степени полинома будет

$R_p^2 = n\cdot\left(X\hat{a}_p - \bar{y}\right)^T \Sigma_\varepsilon^{-1}\left(X\hat{a}_p - \bar{y}\right)$.

В случае, когда измерения равноточные и применяется МНК,

$R_p^2 = n\cdot\sigma_\varepsilon^{-2}\cdot\left(X\hat{a}_p - \bar{y}\right)^T\left(X\hat{a}_p - \bar{y}\right)$.

Как и ранее, при условии справедливости нулевой гипотезы, то есть при $p = q$ $R_p^2 = R_q^2$, и $R_q^2 \in \chi^2(k-q-1)$. Критическое значение при заданном значении вероятности α есть $\chi^2_{1-\alpha}(k-q-1)$.

Гипотеза формулируется по аналогии с формулировкой, приведенной в предыдущем п. а).

H_0: степень аппроксимирующего полинома $p = q$,

H_1: степень аппроксимирующего полинома $p < q$.

Процедура проверки гипотезы в условиях, когда заданы вид полинома и вероятность α.

1. Оцениваются коэффициенты полинома

$\hat{a}_p = \left(X^T X\right)^{-1} X^T \bar{y}$ (МНК)

$\hat{\vec{a}}_p = \left(X^T \Sigma_{\varepsilon}^{-1} X\right)^{-1} X^T \Sigma_{\varepsilon}^{-1} \vec{\bar{y}}$ (ОМНК).

2. Вычисляется статистика критерия

$R_p^2 = n \cdot \left(X\hat{a}_p - \bar{y}\right)^T \Sigma_{\varepsilon}^{-1} \left(X\hat{a}_p - \bar{y}\right)$ (МНК)

$R_p^2 = n \cdot \sigma_{\varepsilon}^{-2} \cdot \left(X\hat{a}_p - \bar{y}\right)^T \left(X\hat{a}_p - \bar{y}\right)$ (ОМНК)

3. Значение статистики R_p^2 сравнивается с критическим значением:

- если $R_p^2 \le \chi_{1-\alpha}^2(k-q-1)$, делается вывод о том, что нет достаточных оснований для того, чтобы отвергнуть нулевую гипотезу,

- в противном случае, если $R_p^2 > \chi_{1-\alpha}^2(k-q-1)$, делается вывод о том, что нет достаточных оснований для того, чтобы считать нулевую гипотезу справедливой.

Заметим, что в соответствии с центральной предельной теоремой плотность распределения средних арифметических асимптотически нормальна. Поэтому, начиная с n = 15 ÷ 20, требования к нормальности погрешностей измерений могут быть значительно смягчены.

Проверка гипотезы о степени аппроксимирующего полинома, характеристики погрешностей измерений неизвестны.

Если характеристики погрешностей измерения значений аппроксимируемой функции неизвестны, то неизбежно приходится выполнять многократные измерения и по результатам этих измерений оценивать характеристики погрешностей: дисперсии $\sigma_{\varepsilon i}^2$ или ковариационную матрицу S_{ε}. При вынужденном пренебрежении коррелированностью между измерениями по причине матрица S_{ε} оказывается диагональной, и это обстоятельство несколько снижает эффективность оценок коэффициентов аппроксимирующего полинома, но оценки коэффициентов остаются несмещенными, если, конечно, назначенная степень полинома *p* равной истинной степени *q*.

Здесь, как и в предыдущем подразделе в качестве исходных данных для полиномиальной аппроксимации используется не вектор результатов однократных измерений $\vec{y}_{\varepsilon}$, а вектор средних арифметических значений $\bar{y}$ результатов многократных измерений.

Статистикой критерия проверки гипотезы о степени полинома будет

- в случае, когда по результатам проверки гипотезы о равенстве дисперсий принято решение о применении МНК,

$$F_p = \frac{1}{k-q-1} \cdot R_p^2,$$

где $R_p^2 = n \cdot s_{\varepsilon}^{-2} \cdot \left(X\hat{a}_p - \bar{y}\right)^T \left(X\hat{a}_p - \bar{y}\right)$,

- в противном случае, при применении ОМНК

$$F_p = \frac{n-k+q+1}{(k-q-1)(n-1)} \cdot R_p^2,$$

где

$$R_p^2 = n \cdot \left(X\hat{a}_p - \bar{y}\right)^T S_{\varepsilon}^{-1} \left(X\hat{a}_p - \bar{y}\right).$$

При условии справедливости нулевой гипотезы, то есть при *p* = *q*, статистика $F_q = F_p$ в обоих случаях распределена по закону распределения Фишера:

- при применении МНК $F \in F(k-q-1, n-1)$ с количеством степеней свободы (k - q - 1) и (n - 1),

- при применении ОМНК $F \in F(k-q-1, n-k+q+1)$ с количеством степеней свободы (k - q - 1) и (n - k + q + 1).

Гипотеза о степени полинома формулируется как и ранее:

H_0: степень аппроксимирующего полинома $p = q$,

H_1: степень аппроксимирующего полинома $p < q$.

Процедура проверки гипотезы в условиях, когда заданы вид полинома и вероятность α, а также вычислены оценки $\bar{y}$ и s_ε^2 или S_ε.

1. Оцениваются коэффициенты полинома

$\hat{a}_p = \left(X^T X\right)^{-1} X^T \bar{y}$ (МНК)

$\hat{\vec{a}}_p = \left(X^T \Sigma_\varepsilon^{-1} X\right)^{-1} X^T \Sigma_\varepsilon^{-1} \vec{\bar{y}}$ (ОМНК).

2. Вычисляется статистика критерия

- $F_p = \dfrac{n}{k-q-1} n \cdot s_\varepsilon^{-2} \cdot \left(X\hat{a}_p - \bar{y}\right)^T \left(X\hat{a}_p - \bar{y}\right)$ (МНК)

- $F_p = \dfrac{n(n-k+q+1)}{(k-q-1)(n-1)} n \cdot \left(X\hat{a}_p - \bar{y}\right)^T S_\varepsilon^{-1} \left(X\hat{a}_p - \bar{y}\right)$.

3. Значение статистики F_p сравнивается с критическим значением распределения Фишера:

- $F_p \le F_{1-\alpha}(k-q-1, n-1)$ (МНК)

- $F_p \le F_{1-\alpha}(k-q-1, n-k+q+1)$ (ОМНК),

делается вывод о том, что нет достаточных оснований для того, чтобы отвергнуть нулевую гипотезу,

- в противном случае делается вывод о том, что нет достаточных оснований для того, чтобы считать нулевую гипотезу справедливой.

Последовательная полиномиальная аппроксимация с проверкой гипотезы о степени полинома

На практике при экспериментальном исследовании зависимостей порядок степенного или обобщенного полинома, как правило, неизвестен, за исключением случаев, когда этот порядок задан заранее. Поэтому практически важной задачей является выбор такого порядка (степени) полинома, при котором обеспечиваются:

- заданная точность аппроксимации исследуемой функции,

- вычислительная устойчивость операций, выполняемых для нахождения оценок коэффициентов аппроксимирующего полинома.

Эти два требования противоположны.

С одной стороны, точность аппроксимации гладких функций степенным или обобщенным полиномами повышается с увеличением степени полинома.

С другой стороны, с увеличением степени полинома, увеличивается количество оцениваемых коэффициентов и в связи с этим ухудшается обусловленность матриц, участвующих в выполняемых операциях, что приводит к утрате вычислительной устойчивости этих операций. Неустойчивость вычислений доходит до экстремально высокого уровня, когда заданная при аппроксимации степень полинома настолько высока, что приходится находить оценки несуществующих или пренебрежимо малых коэффициентов.

Можно повысить вычислительную устойчивость полиномиальной аппроксимации за счет оптимального планирования и регуляризации, но это не является панацеей.

Практически реализуемой является стратегия полиномиальной аппроксимации, основанная на следующих принципах:

- не стремиться к точности аппроксимации, превышающей точность исходных данных, которая определяется характеристиками погрешностей измерений и количеством независимых повторных измерений *n*,

- подбирать подходящую степень (порядок) полинома путем направленного перебора от наименьшей в сторону возрастания до того значения, при котором характеристики погрешности аппроксимации окажутся согласованными с погрешностями исходных данных.

При реализации такой стратегии погрешность аппроксимации монотонно уменьшается, а погрешность, порождаемая вычислительной неустойчивостью, монотонно повышается. М.А.Красносельским доказано, что подобная стратегия является регуляризирующей стратегией решения плохообусловленных и некорректных задач.

Может оказаться, что этот процесс наращивания степени полинома остановится до достижения заданной заранее точности аппроксимации. Это значит, что при имеющейся погрешности исходных данных данная точность недостижима. В этом случае необходимо повысить точность измерений или увеличить количество независимых многократных измерений.

В условиях, когда погрешности исходных данных случайны, остановка процесса монотонного наращивания степени аппроксимирующего полинома может осуществляться только путем проверки гипотезы о степени полинома в соответствии с одним из критериев.

Алгоритм последовательной полиномиальной аппроксимации функций с проверкой гипотезы о степени полинома:

1. Выполняются эксперименты при значениях $x_i \quad i = 1,2,...,k$, где *k* должно быть больше максимально возможного предполагаемого количества коэффициентов искомого полинома.

2. Задается уровень значимости $\alpha \in [0.05, 0.2]$.

3. При необходимости в зависимости от ситуации вычисляются оценки $\bar{\bar{y}}, s_{\varepsilon i}^2, S_\varepsilon$, проверяется гипотеза о равенстве дисперсий $\sigma_{\varepsilon i}^2$ и принимается решение о применении МНК или ОМНК.

4. Задается начальная степень полинома *p*.

5. Формируется матрица *X*.

6. В зависимости от ситуации вычисляются оценки коэффициентов аппроксимирующего полинома порядка (степени) *p*.

7. Вычисляется статистика R_p^2 или F_p (в зависимости от ситуации) критерия проверки гипотезы о степени полинома, значение этой статистики сравнивается с соответствующим критическим значением.

8. Делается вывод о справедливости нулевой гипотезы:

- если принято решение об отклонении нулевой гипотезы, *p* = *p* + 1, переход на шаг алгоритма 5.

- в противном случае переход на шаг алгоритма 9.

9. По одной из подходящих формул вычисляется ковариационная матрица $\Sigma_{\hat{a}}$ погрешностей определения коэффициентов аппроксимирующего полинома.

Литература

Теория

1. Вентцель Е.С. Теория вероятностей. – 5-е изд., стер. – М.: Высшая школа, 1998. – 576 с.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и её инженерные приложения. – 2-е изд., стер. – М: Высшая школа. – 480 с.
3. Гнеденко Б.В. Курс теории вероятностей. – 6-е изд., пер. и доп. – М.: Гл.ред. ФМЛ изд-ва «Наука», 1988. – 448 с.
4. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2003. – 479 с.
5. Феллер В. Введение в теорию вероятностей и ее приложения. Том 1. – 2-е изд. – М.: Мир, 1967. – 500 с.

Задачи

Основная

1. Сборник задач по теории вероятностей, математической статистике. / под общ. ред. Свешникова А.А. 4-е изд., стер. – СПб.: Лань, 2008. – 448 с.

Дополнительная

2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. Задачи и упражнения. – М.: Гл.ред. ФМЛ изд-ва «Наука», 1969. – 368 с.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – 3-е изд., пер. и доп. – М.: Высшая школа, 1979. – 400 с.
4. Мостеллер Ф. Пятьдесят занимательных вероятностных задач с решениями. – 2-е изд., пер. с англ. – М.: Гл.ред. ФМЛ изд-ва «Наука», 1975. – 112 с.
5. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / под ред. Свешникова А.А. – М.: Гл.ред. ФМЛ изд-ва «Наука», 1970. – 656 с.

Matlab

1. Лазарев Начала программирования в среде Matlab
2. Мэтьюс Финк Численные методы Использование Matlab
3. Потемкин Система Инженерных И Научных Расчетов Matlab
4. Чен, Джиблин, Ирвинг Matlab в математических исследованиях.djvu – есть упражнения с комментариями

Расчетные задания

1. Солопченко Г.Н. Теория вероятностей и математическая статистика : учебное пособие. – изд-е 3-е, пер. и доп. – СПб, изд-во СПбГПУ, 2015. –214 с.