

Математические методы в биологии

Блок 2. Случайные величины

Лекция 4

Козлова Ольга Сергеевна
89276755130, olga-sphinx@yandex.ru

Задача о счастливом билете



Какова вероятность попадания счастливого билета в общественном транспорте?

Дмитрий Мастер (1551), Вопрос решён 5 лет назад

Дополнен 5 лет назад

Вы не поняли. Я прошу решить задачу с точки зрения теории вероятностей

Нравится

Подписаться



ЛУЧШИЙ ОТВЕТ



BELLADONNA Мудрец (19097) 5 лет назад

Вероятность равна 0,055252

Посмотри здесь:

<http://www.ega-math.narod.ru/Quant/Tickets.htm#A5>

Источник: Теория вероятностей

2 Нравится

Комментировать

Пожаловаться

5 ОТВЕТОВ



Олюшка Мастер (1549) 5 лет назад

50 на 50-либо счастливый, либо нет

Нравится

Комментировать

Пожаловаться

Задача о счастливом билете



$k=9$

- Пусть номер билета шестизначный
- Критерий «счастья»: сумма первых трёх цифр равна сумме трёх последних
- Пусть номера билетов начинаются с 000 000 и заканчиваются 999 999 (их 1 млн.)

Пусть k – сумма первых (и последних) 3х цифр. $k = \overline{0,27}$ (28 знач-й)

$N(k)$ – число трёхзначных чисел с суммой цифр, равной k

Число счастливых билетов для k равно $N(k)^2$

ВОПРОС: Как найти $\sum_{k=0}^{27} N(k)^2$???

Переобозначим $N(k)$ как $N_3(k)$ и будем действовать последовательно.

$N_1(k) = 1$ для $k = \overline{0,9}$ и 0 для $k = \overline{10,27}$. (номера типа «00», «11»,..., «99», их 10 штук)

Выведем $N_n(k)$ в общем виде, зная $N_{n-1}(k)$ (зададим способ перехода).

Пусть l – первая цифра n -значного «полуномера» $0 \leq l \leq 9, l \leq k$, значит, для того, чтобы сумма цифр этого «полуномера» была равна k , остальные цифры должны в сумме давать $k - l$. Число таких «полуномеров», начинающихся на l , равно $N_{n-1}(k - l)$. Просуммируем это количество по $l = \overline{0,9}$ и получим $N_n(k) = \sum_{l=0}^9 N_{n-1}(k - l)$. Если $k < 9$, все значения $N_{n-1}(k - l)$ для $l > k$ считаем равными нулю.

Это рекуррентное соотношение для вывода $N_n(k)$ для любого n и k в общем виде!

Зная рекуррентное соотношение и $N_1(k)$ для всех k , составим таблицу для $N_n(k)$.

$N_1(k) = 1$ для $k = \overline{0,9}$ и 0 для $k = \overline{10,27}$

$$N_n(k) = \sum_{l=0}^9 N_{n-1}(k-l)$$

Например, $N_2(6) = N_1(6) + N_1(5) + N_1(4) + N_1(3) + N_1(2) + N_1(1) + N_1(0) = 7$

$N_3(5) = N_2(5) + N_2(4) + N_2(3) + N_2(2) + N_2(1) + N_2(0) = 21$

Назад к вероятности:

- Для 2х-знач. билетов $P = \frac{10}{100} = 0,1$
- Для 4х-знач. билетов $P = \frac{670}{10000} = 0,067$
- Для 6ти-знач. билетов $P = \frac{55252}{1000000} = 0,055252$
- Для 8ми-знач. билетов $P = \frac{4816030}{100000000} = 0,0481603$

Для 6ти-знач. билетов в среднем каждый 18й билет является счастливым!

k	Длина половины номера			
	n = 1	n = 2	n = 3	n = 4
0	1	1	1	1
1	1	2	3	4
2	1	3	6	10
3	1	4	10	20
4	1	5	15	35
5	1	6	21	56
6	1	7	28	84
7	1	8	36	120
8	1	9	45	165
9	1	10	55	220
10		9	63	282
11		8	69	348
12		7	73	415
13		6	75	480
14		5	75	540
15		4	73	592
16		3	69	633
17		2	63	660
18		1	55	670
19			45	660
20			36	633
21			28	592
22			21	540
23			15	480
24			10	415
25			6	348
26			3	282
27			1	220
28				165
29				120
30				84
31				56
32				35
33				20
34				10
35				4
36				1

$\sum_k n^2 = 10$ (for k=1 to 3)
 $\sum_k n^2 = 670$ (for k=1 to 18)
 $\sum_k n^2 = 55252$ (for k=1 to 27)
 $\sum_k n^2 = 4816030$ (for k=1 to 36)

Числовые характеристики дискретной случайной

ВЕЛИЧИНЫ

- Математическое ожидание $M(X) = \sum_{i=1}^n X_i * P(X = X_i)$

Две д.с.в. с одинаковым мат.ожиданием:

X	-0.5	0.5
P	0.5	0.5

Y	-100	100
P	0.5	0.5

$$M(X) = M(Y) = 0$$

- Дисперсия (*dispersion - рассеяние*) – мера изменчивости случайной величины. Это математическое ожидание квадрата отклонения случайной величины от её математического ожидания:

$$D(X) = M[(X - M(X))^2]$$

ВОПРОС: Откуда квадрат??

Неслучайная (постоянная величина)

ОК, пусть без квадрата: $M(X - M(X)) = M(X) - M(M(X)) = M(X) - M(X) = 0$

- Вычисление дисперсии «в лоб»:

X	1	2	3
P	0,3	0,5	0,2

$$M(X) = 1 * 0,3 + 2 * 0,5 + 3 * 0,2 = 1,9$$

$(X-M(X))^2$	$(1-1,9)^2=(-0,9)^2=0,81$	$(2-1,9)^2=(0,1)^2=0,01$	$(3-1,9)^2=(1,1)^2=1,21$
P	0,3	0,5	0,2

$D(X)$

$$M[(X - M(X))^2] = 0,81 * 0,3 + 0,01 * 0,5 + 1,21 * 0,2 = 0,243 + 0,005 + 0,242 = 0,49$$

Формула для вычисления дисперсии

Дисперсия равна разности между математическим ожиданием квадрата случайной величины X и квадратом математического ожидания X:

$$D(X) = M(X^2) - (M(X))^2$$

Доказательство.

$$\begin{aligned} D(X) &= M[(X - M(X))^2] = M[X^2 - 2M(X)X + (M(X))^2] = \\ &M(X)^2 - M(2M(X)X) + (M(X))^2 = M(X)^2 - M(2M(X))M(X) + (M(X))^2 = \\ &= M(X)^2 - 2M(X)M(X) + (M(X))^2 = M(X)^2 - 2(M(X))^2 + (M(X))^2 = \\ &= M(X^2) - (M(X))^2 \end{aligned}$$

Note: In the original image, blue arrows point from the word 'const' to the terms $(M(X))^2$ and $M(2M(X))$ in the derivation.

Пример.

X	1	2	3
P	0,3	0,5	0,2

X²	1	4	9
P	0,3	0,5	0,2

$$M(X) = 1 * 0,3 + 2 * 0,5 + 3 * 0,2 = 1,9$$

$$M(X^2) = 1 * 0,3 + 4 * 0,5 + 9 * 0,2 = 4,1$$

$$D(X) = 4,1 - 1,9^2 = 4,1 - 3,61 = 0,49$$

Свойства дисперсии

- Дисперсия постоянной величины равна 0: $D(C) = 0$

Доказательство. $D(C) = M[(C - M(C))^2] = M(0) = 0$

- Постоянный множитель выносится за знак дисперсии возведённым в квадрат: $D(CX) = C^2 D(X)$

Доказательство. $D(CX) = M[(CX - M(CX))^2] = M[(CX - C * M(X))^2] =$
 $M[(C(X - M(X)))^2] = C^2 * M[(X - M(X))^2] = C^2 D(X)$

- Дисперсия суммы взаимно независимых случайных величин равна сумме их дисперсий

Доказательство для двух слагаемых.

$$\begin{aligned} D(X + Y) &= M((X + Y)^2) - (M(X + Y))^2 = \\ &= M(X^2 + 2XY + Y^2) - (M(X) + M(Y))^2 = M(X^2) + M(2XY) + M(Y^2) - \\ &= M(X^2) + 2M(X)M(Y) - (M(X))^2 - (M(Y))^2 = \\ &= M(X^2) + M(Y^2) - (M(X))^2 - (M(Y))^2 = D(X) + D(Y) \end{aligned}$$

- $D(X - Y) = D(X) + D(Y)$ Доказательство:

$$D(X + (-Y)) = D(X) + (-1)^2 D(Y)$$

Дисперсия числа появлений событий в независимых испытаниях

- n независимых испытаний
- Событие A появляется в каждом из них с вероятностью p
- Дискретная случайная величина X – число появления события A в этих испытаниях

ВОПРОС: Чему равна дисперсия случайной величины X - числа появлений события A в испытаниях?

ОТВЕТ: Дисперсия числа появлений события A в n испытаниях равно произведению n на p на $(1-p)$: $D(X) = n * p * (1-p)$

Доказательство: Пусть X_1 – число появления события A в первом испытании, X_2 – во втором и т.д., X_n – в n -ом. Всего событие A появилось $X_1 + X_2 + \dots + X_n$ раз. По свойству дисперсии суммы, $D(X) = D(X_1) + D(X_2) + \dots + D(X_n)$.

Распишем $D(X_1), D(X_2), \dots, D(X_n)$: $D(X_1) = M(X_1^2) - (M(X_1))^2$

Дискретная случайная величина X_1 (как и X_n) принимает значение 1 с вероятностью p (событие случилось) и значение 0 с вероятностью $(1-p)$ (событие не случилось). Поэтому $D(X_1) = 1 * p + 0 * (1-p) - (1 * p + 0 * (1-p))^2 = p - p^2 = p(1-p)$. И так для каждого n , поэтому $D(X) = n * p * (1-p)$.

Данная случайная величина X распределена по биномиальному закону, поэтому можно сказать, что

ДИСПЕРСИЯ БИНОМИАЛЬНОГО РАСПРЕДЕЛЕНИЯ С ПАРАМЕТРАМИ n И p РАВНА ПРОИЗВЕДЕНИЮ $n * p * (1-p)$.

Среднее квадратическое отклонение

- Определение. Среднее квадратическое отклонение случайной величины – это квадратный корень из её дисперсии.

СИГМ $\rightarrow \sigma(X) = \sqrt{D(X)}$

Среднее квадратическое отклонение позволяет сохранить размерность случайной величины

Типовая задача на вычисление мат.ожидания, дисперсии и среднего квадратического отклонения дискретной случайной величины.

Пусть случайная величина X имеет следующий закон распределения:

X	2	3	10
P	0,1	0,4	0,5

Найти $M(X)$, $D(X)$ и $\sigma(X)$

$$M(X) = 2 * 0,1 + 3 * 0,4 + 10 * 0,5 = 6,4$$

X²	4	9	100
P	0,1	0,4	0,5

$$M(X^2) = 4 * 0,1 + 9 * 0,4 + 100 * 0,5 = 54$$

$$\Rightarrow D(X) = 54 - 6,4^2 = 13,04$$

$$\sigma(X) = \sqrt{13,04} \approx 3,61$$

Ответ. Данная случайная величина X имеет мат.ожидание 6,4, дисперсию 13,04 и среднее квадратическое отклонение 3,61.

От дискретности – к непрерывности!

- Закон распределения дискретной с.в. можно задать таблично, перечислив все её значения. Закон распределения непрерывной с.в., принимающей любые значения на промежутке $[a,b]$, так задать нельзя!

Что же делать?

- Введём понятие функции распределения вероятностей случайной величины.

Определение. Функция распределения $F(x)$, x – любое действительное число – такая функция, которая определяет вероятность того, что случайная величина X примет значение, меньшее x :

$$F(x) = P(X < x)$$

Значения такой функции лежат на отрезке $[0,1]$, а сама функция неубывающая.

Функцию распределения вероятностей можно задать не только для непрерывной, но и для дискретной с.в.!

Пример. Дискретная с.в. распределена по закону:

X	1	4	8
P	0,3	0,1	0,6

Определим функцию её распределения.

$F(x)=0$ для $x \leq 1$ (X не принимает значения, меньше 1)

$F(x)=0,3$ для $1 < x \leq 4$ (X может принять значение 1 с вероятностью 0,3)

$F(x)=0,4$ для $4 < x \leq 8$ (X может принять значение 1 с вероятностью 0,3 или значение 4 с вероятностью 0,1)

$F(x)=1$ для $x > 8$ (все значения д.с.в. лежат левее 8 на числовой оси)

Плотность распределения вероятностей с.в. (только для непрерывных!)

- Плотность распределения вероятностей – первая производная от функции распределения (функция распределения – первообразная плотности)

$$f(x) = F'(x)$$

↑
Плотность

←
Функция
распределения

Плотность распределения – это вероятность того, что непрерывная с.в. примет значение на интервале (a,b), равна интегралу от плотности распределения, взятому в пределах от a до b:

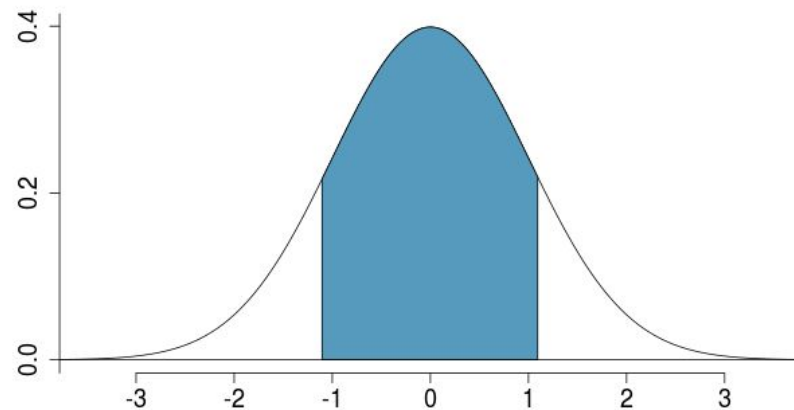
$$P(a < X < b) = \int_a^b f(x) dx$$

Или, геометрически: вероятность того, что н.с.в. примет значение на интервале (a,b), равна площади криволинейной трапеции, ограниченной:

- Кривой плотности распределения сверху
- Осью абсцисс – снизу
- Прямыми $x=a$ и $x=b$ – слева и справа

На картинке $a=-1,11$ и $b=1,1$

Вероятность того, что н.с.в. примет знач. на $(-1,11, 1,11)$, равна 0,731.



$$P(-1.11 < X < 1.1) = 0.731$$

Законы распределения н.с.в.

1. Равномерный закон распределения вероятностей

Распределение вероятностей называют **равномерным**, если на отрезке $[a,b]$, которому принадлежат все возможные значения случайной величины, плотность распределения вероятностей сохраняет постоянное значение, равное $1/(b-a)$

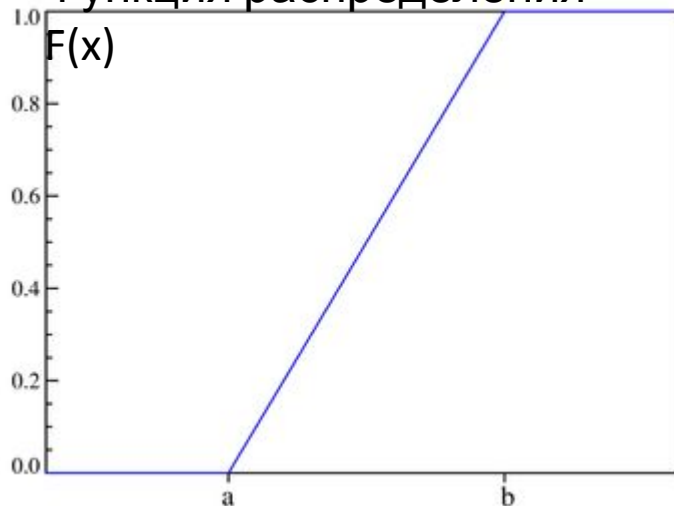
Пример. Пусть с.в. X принимает с одинаковой вероятностью любое значение на отрезке $[a,b]$. Тогда её функция распределения $F(x)$ выглядит так:

$$F(x)=0, \text{ если } x \leq a,$$

$$F(x)=1, \text{ если } x > b$$

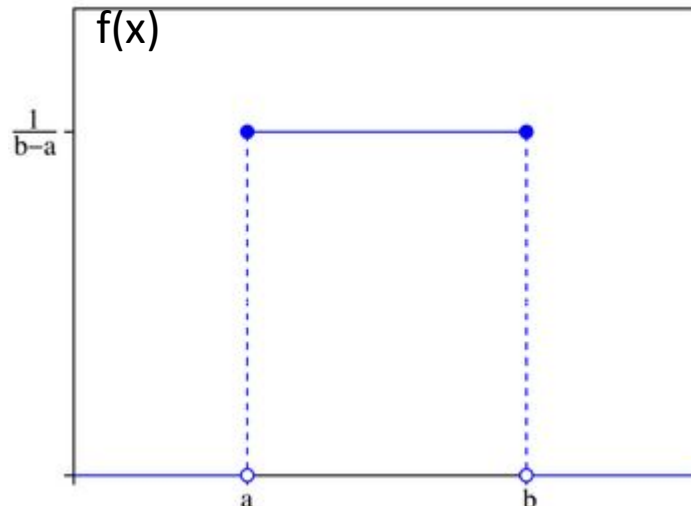
$$F(x)=\int_a^b \frac{1}{b-a} dx = \frac{x-a}{b-a}, \text{ если } a < x \leq b$$

Функция распределения



*Первая производная
линейной функции –
константа!*

Плотность распределения



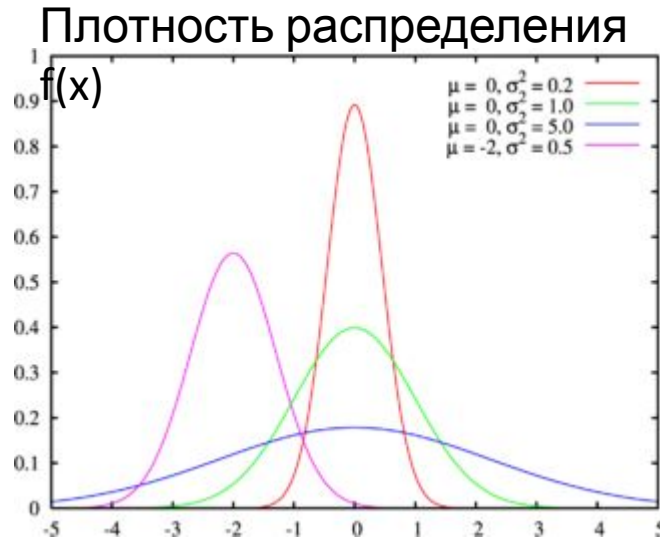
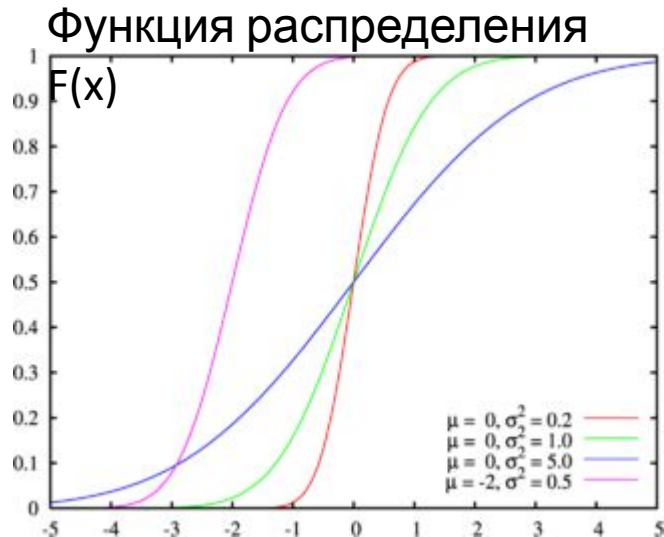
Законы распределения н.с.в.

2. Нормальный (гауссовский) закон распределения вероятностей

- Случайная величина X определена на всей оси абсцисс
- Два параметра: μ (мат. ожидание) и σ (среднее квадратическое отклонение)

Нормальное распределение – то распределение, плотность которого задаётся формулой Гаусса:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Изменение μ приводит к сдвигу кривой вдоль оси x ; изменение σ приводит к сжатию/растяжению кривой

Нормированное (стандартное) норм.распр. – норм.распр. с параметрами $\mu=0$ и $\sigma=1$

Ещё о нормальном распределении

- Правило одной сигмы:

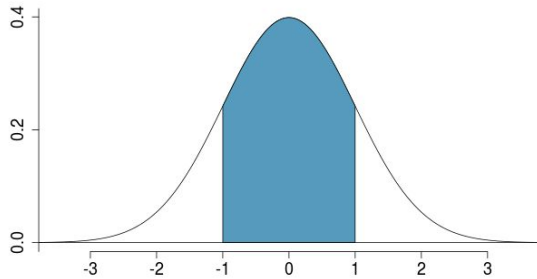
Непрерывная случайная величина, распределённая по нормальному закону, попадает в интервал $\mu \pm \sigma$ с вероятностью 0,68

- Правило двух сигм:

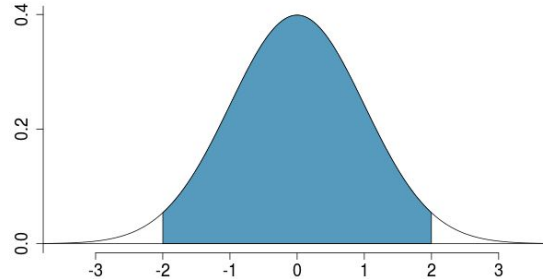
Непрерывная случайная величина, распределённая по нормальному закону, попадает в интервал $\mu \pm 2\sigma$ с вероятностью 0,95

- Правило трёх сигм:

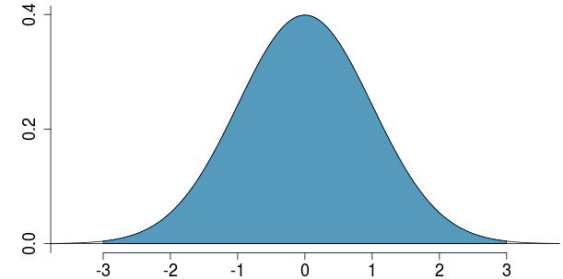
Непрерывная случайная величина, распределённая по нормальному закону, попадает в интервал $\mu \pm 3\sigma$ с вероятностью почти 1



$$P(-1 < X < 1) = 0.683$$



$$P(-2 < X < 2) = 0.954$$



$$P(-3 < X < 3) = 0.997$$

ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА: если случайная величина X представляет собой сумму очень большого числа взаимно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.

Законы распределения н.с.в.

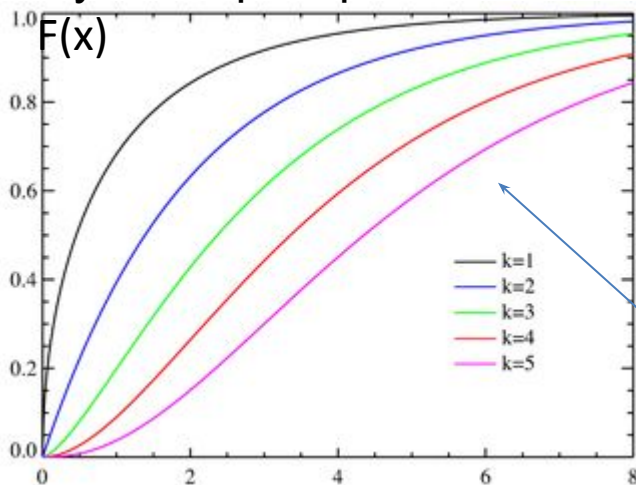
3. Распределение «хи-квадрат» (χ^2)

Распределение «хи-квадрат» – распределение случайной величины, представляющей собой сумму квадратов k независимых стандартных нормальных величин:

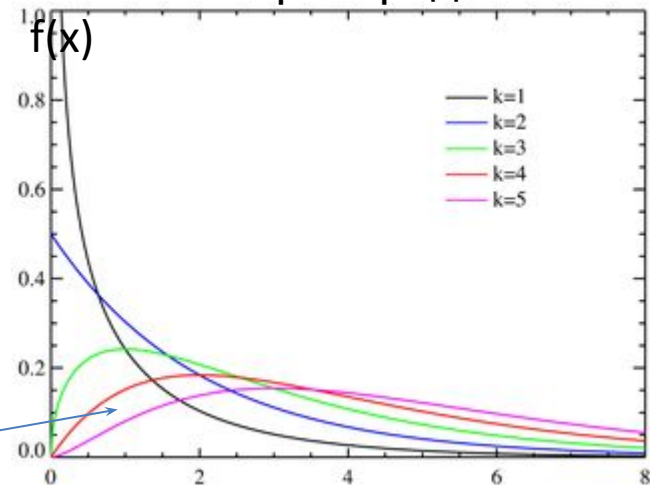
$$\chi^2 = \sum_{i=1}^k X_i^2$$

С увеличением k распределение «хи-квадрат» приближается к нормальному.

Функция распределения



Плотность распределения



Уже для $k=5$ видны зачатки «нормальности»

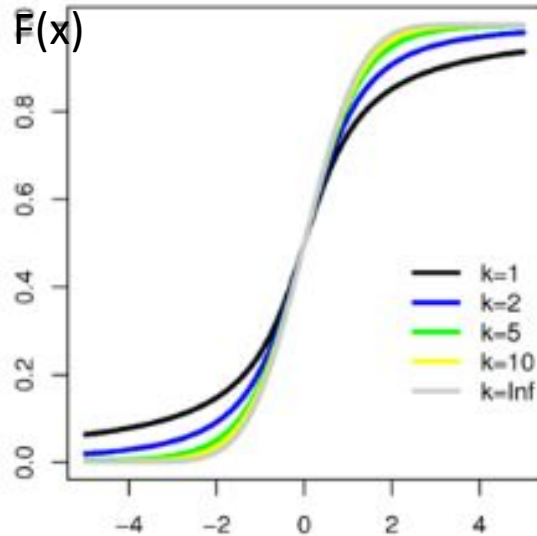
Законы распределения н.с.в.

4. Распределение Стьюдента

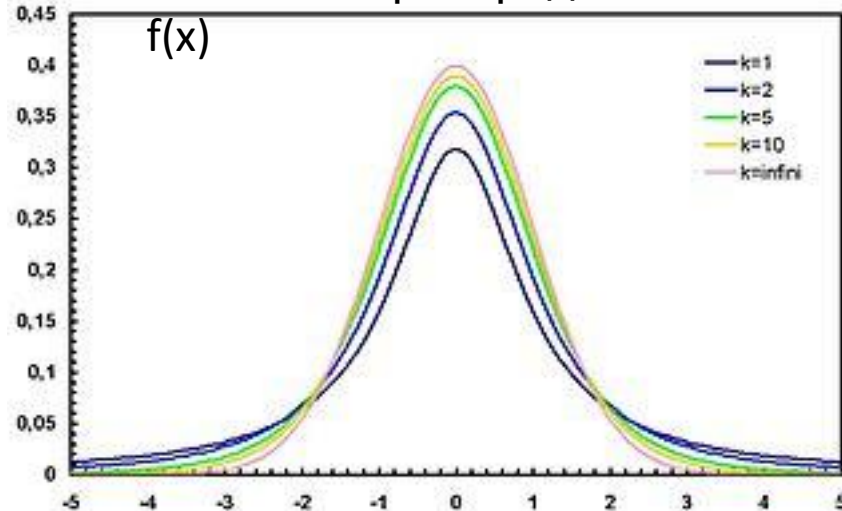
Пусть z – нормальная стандартная случайная величина, а V – независимая от неё величина, имеющая распределение χ^2 с использованием k независимых нормальных стандартных величин (число k ещё называют числом степеней свободы). Тогда случайная величина $T = \frac{Z}{\sqrt{V/k}}$ имеет распределение Стьюдента с k степенями свободы (его ещё называют t -распределением).

С ростом k распределение Стьюдента стремится к нормальному.

Функция распределения



Плотность распределения



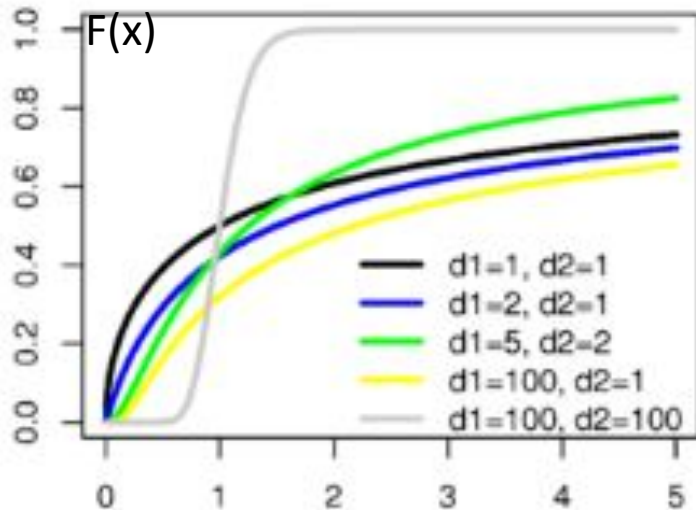
Бледно-фиолетовая линия – обыкновенное нормальное распределение

Законы распределения н.с.в.

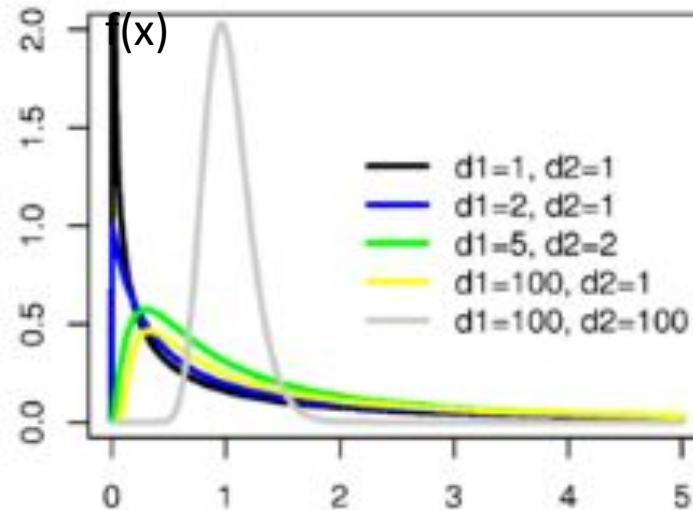
5. Распределение Фишера

Пусть U и V – две независимые случайные величины, имеющие распределение χ^2 с k_1 и k_2 степенями свободы, соответственно. Тогда случайная величина $F = \frac{U/k_1}{V/k_2}$ имеет распределение Фишера (или Фишера-Снедекора) со степенями свободы k_1 и k_2 . Также его называют F-распределением.

Функция распределения



Плотность распределения



Резюме

- $D(X) = M[(X - M(X))^2]$ - определение дисперсии д.с.в.
- $D(X) = M(X^2) - (M(X))^2$ - удобная формула для вычисления дисперсии д.с.в.
- $D(X) = n * p * (1-p)$ – дисперсия биномиального распределения с пар-рами n и p
- $\sigma(X) = \sqrt{D(X)}$ - среднее квадратическое отклонение
- $F(x) = P(X < x)$ – функция распределения
- $f(x) = F'(x)$ – плотность распределения