Дорогие студенты,

данная презентация служит лишь наглядной иллюстрацией к одной из лекций по теории вероятностей для II курса факультета биоинженерии и биоинформатики.

ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Моменты

Для дискретной случайной величины ξ, со значениями

$x_1, \ x_2, \dots x_n, \dots$

которые случайная величина ξ принимает с вероятностями

$p_1, \ p_2, \dots p_n, \dots$

.

i-ый момент вычисляется по формуле

$$E\xi^i = \Sigma_k x_k^i p_k,$$

i-ый центральный момент — по формуле

$$E(\xi - E\xi)^i = \Sigma_k (x_k - E\xi)^i p_k.$$

Для непрерывной случайной величины ξ с плотностью $f_\xi(x)$ i-ым моментом назовем величину

$$a_i = E\xi^i = \int_R x^i f_\xi(x)\, dx\,,$$

i-ым центральным моментом назовем величину

$$\mu_I = E(\xi - E\xi)^i = \int_R (x - E\xi)^i f_\xi(x)\, dx\,.$$

Замечание. Для существования моментов требуется абсолютная сходимость соответствующих рядов и интегралов.

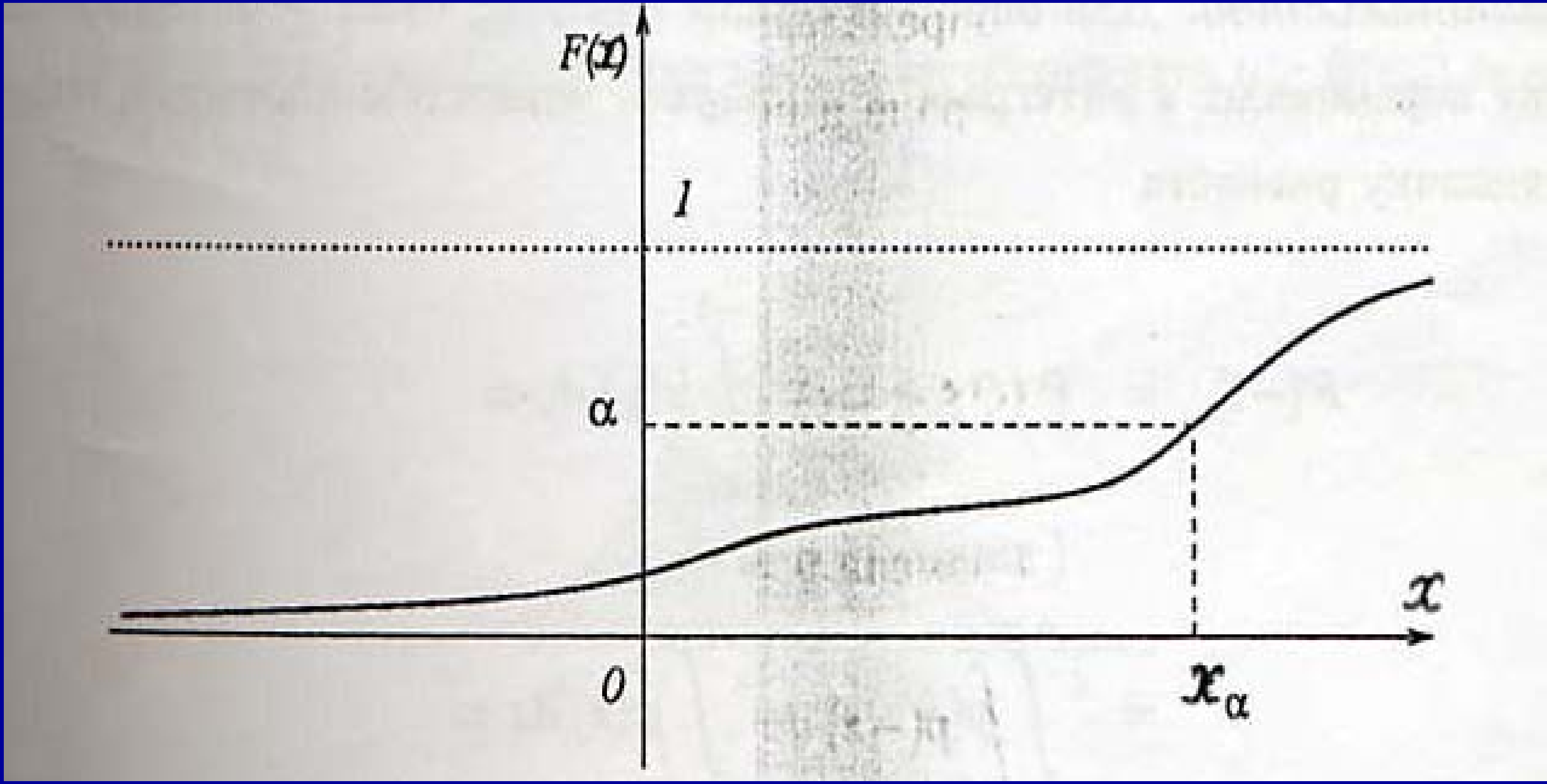
Квантили

Предположим, что строго возрастающая функция F(x) есть функция распределения некоторой непрерывной случайной величины ξ.

Квантилью уровня α для распределения, порождаемого функцией F(x), называется число x_α, являющееся решением уравнения

$$F(x_\alpha)=\alpha$$

$$F(x_\alpha)=\alpha$$



Квантиль уровня α

Медиана и квартили

Медианой называется квантиль уровня 0.5

Верхней квартилью называется квантиль уровня 0.75

Нижней квартилью называется квантиль уровня 0.25

Децили — квантили уровней 0.1, 0.2, …, 0.9

Квантили называют также процентными точками распределения

МОДА

Модой распределения называется то значение (или значения) x, при котором плотность $f_{\xi}(x)$ имеет максимум

Если плотность $f_{\xi}(x)$ имеет единственный максимум, то распределение называется *унимодальным*

Асимметрия и эксцесс

Для случайной величины ξ число

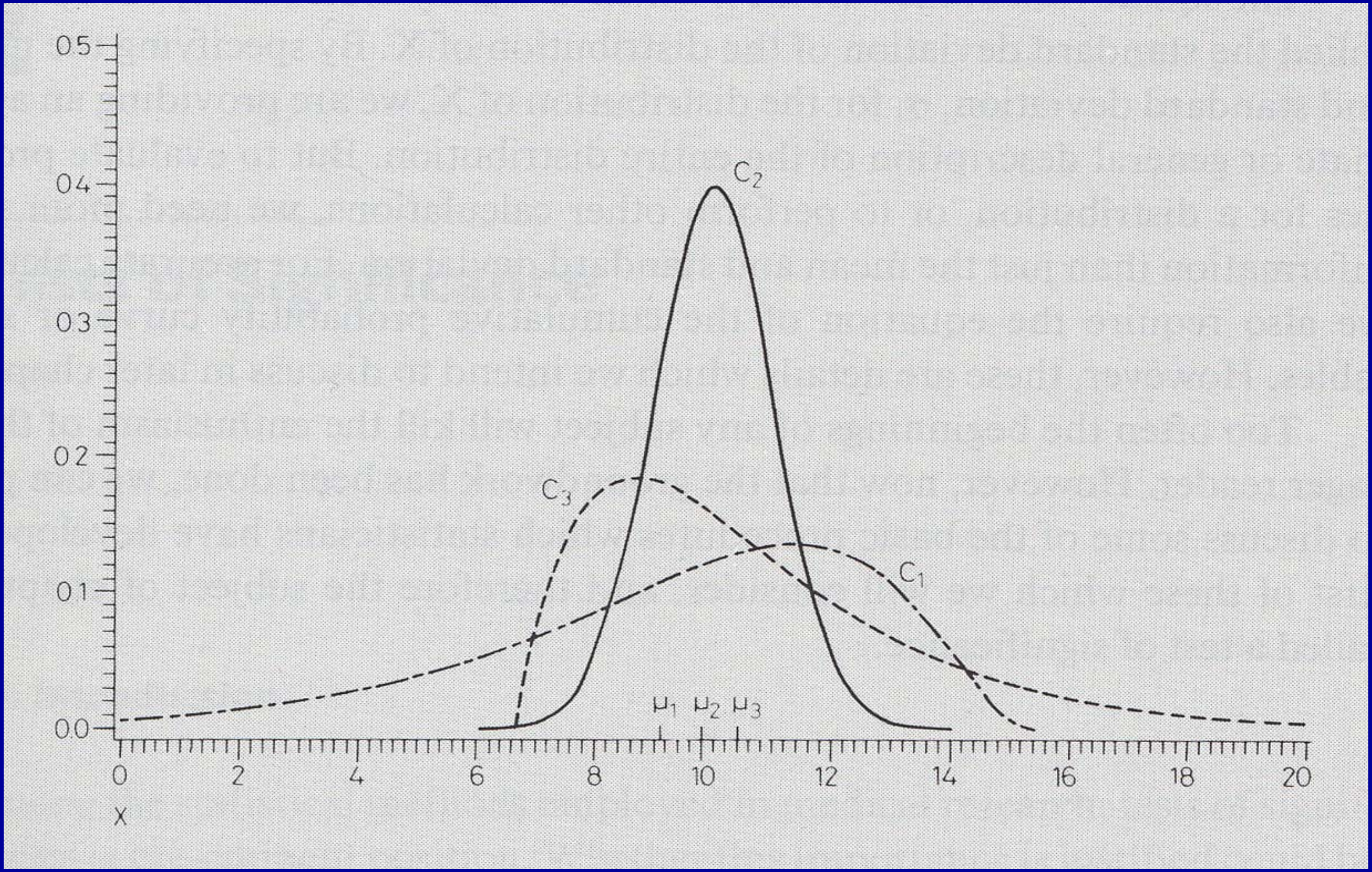
- *$(E(\xi - E\xi)^3)/(D\xi)^{3/2} = \mu_3/\sigma^3$ называется асимметрией распределения (или коэффициентом асимметрии);*
- *$(E(\xi - E\xi)^4)/(D\xi)^2 = \mu_4/\sigma^4$ называется эксцессом распределения (или коэффициентом эксцесса).*

Коэффициент асимметрии нормального распределения равен 3,

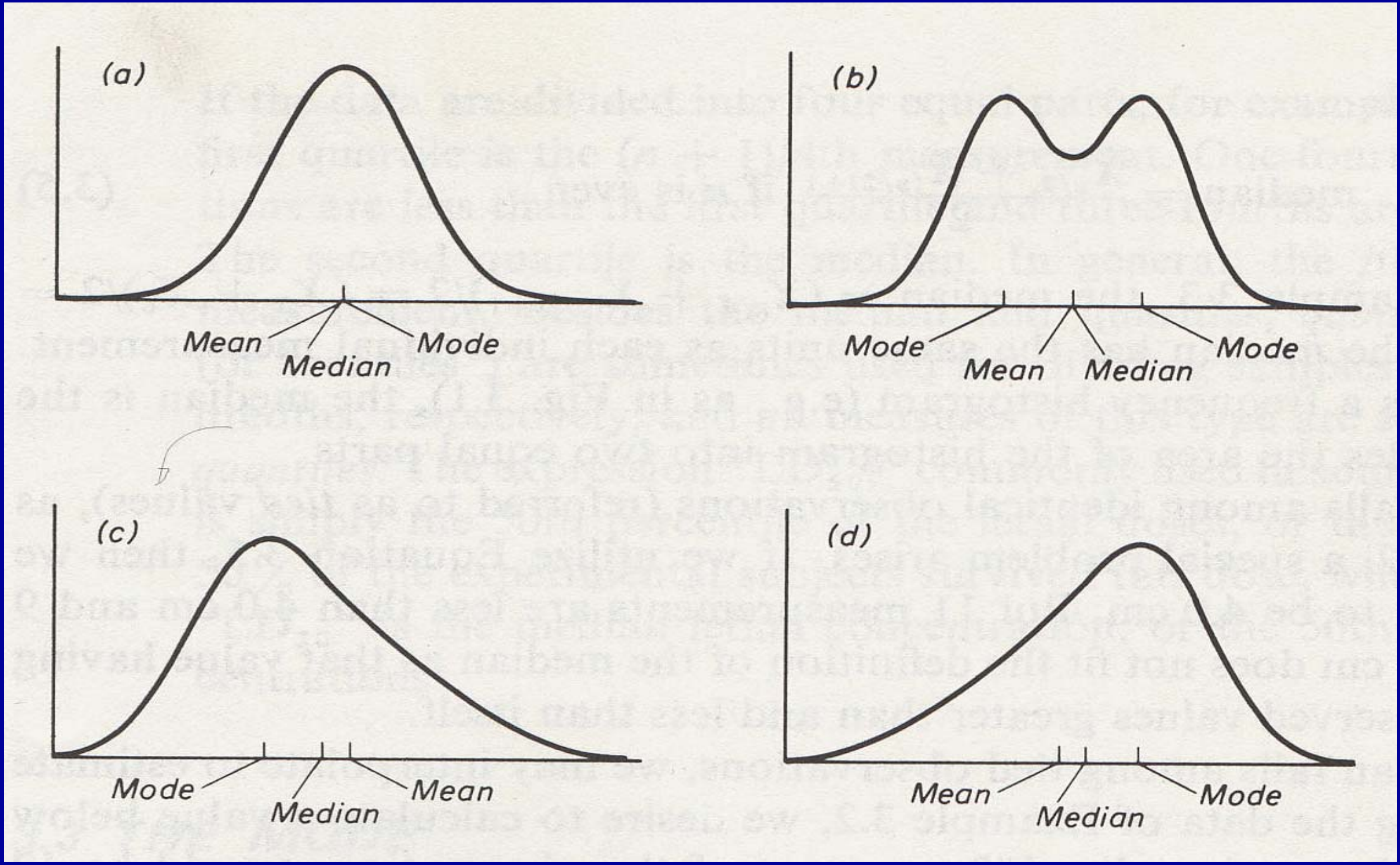
- *поэтому иногда коэффициентом эксцесса называют величину*

$(\mu_4/\sigma^4)-3$

Пример. Плотности распределений C_1, C_2 и C_3 имеют одинаковую медиану, равную 10, но различные средние значения μ_1, μ_2 и μ_3



Типы плотностей непрерывных распределений



БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

$\Omega = \{ \omega = (\theta_1, \theta_2, \dots, \theta_n), \theta_l = 0,1\},$

$\xi(\omega) = \Sigma_i \theta_l$

Случайную величину $\xi(\omega)$ называют **числом успехов** в последовательности n независимых испытаний Бернулли.

$\xi(\omega)$	0	1	...	k	... n
P_ξ	$(1-p)^n$	$np(1-p)^n$	...	$C_n^k p^k (1-p)^{n-k}$	... p^n

Распределение $\xi(\omega)$ называется **биномиальным.**
$M\xi = np$, $D\,\xi = np(1-p)$,
асимметрия биномиального распределения : $(1-2p)(np(1-p))^{-1/2}$

Параметры распределения **n** и **p**

РАСПРЕДЕЛЕНИЕ ПУАССОНА

$\xi(\omega)$	0	1	...	n
P_{ξ}	$e^{-\lambda}$	$\lambda e^{-\lambda}$	...	$(\lambda^{n} e^{-\lambda})/n!$

M $\xi = \lambda$

$D\ \xi = \lambda$

асимметрия распределения : $(\lambda)^{-1/2}$

эксцесс распределения : $(\lambda\ )^{-1}$

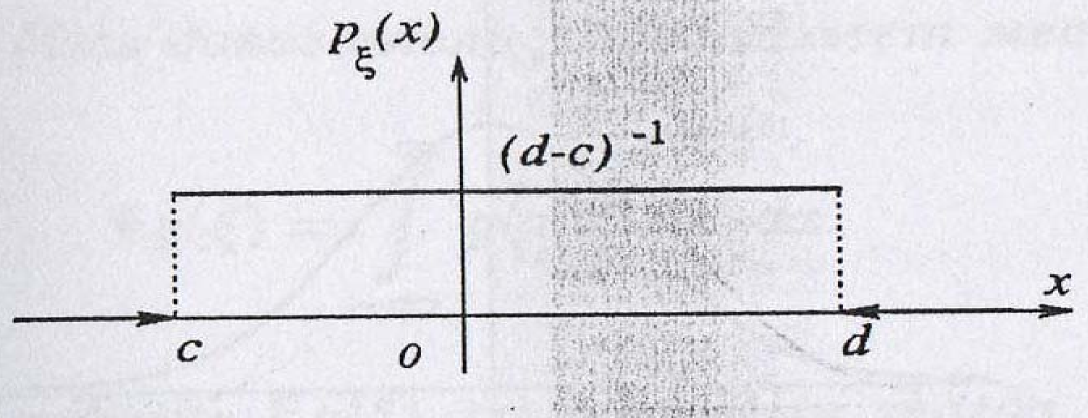
(при определении эксцесса: $(\mu 4\ /\sigma 4)-3)$

Параметр распределения λ

РАВНОМЕРНОЕ РАСПРЕДЕЛЕНИЕ

Равномерное распределение в отрезке $[c, d]$

$$p_\xi(x) = \begin{cases} \dfrac{1}{d-c}, & x \in [c, d], \\ 0, & x \notin [c, d]. \end{cases}$$



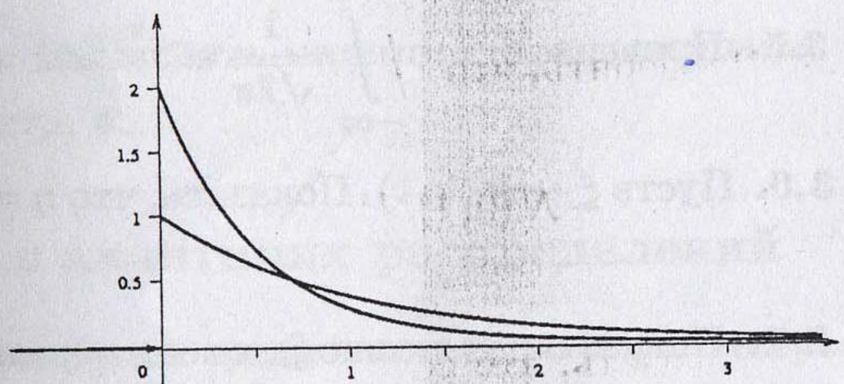
$M\ \xi = (c+d)/2$

$D\ \xi = ((d\text{-}c)^2)/12$

Полностью определяется заданием отрезка [c, d]

Экспоненциальное (показательное) распределение с параметром $\lambda > 0$

$$p_\xi(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0, \end{cases}$$



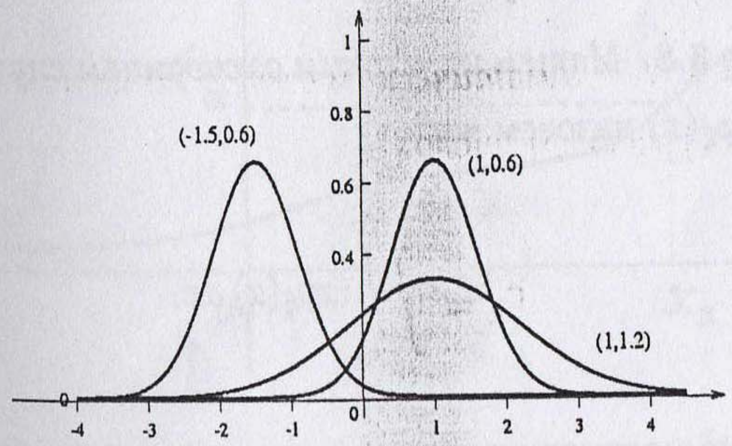
$M\,\xi = 1/\lambda$

$D\,\xi = 1/\lambda^2$

Нормальное (или гауссовское) распределение с параметрами a и σ^2

$a \in R,\ \sigma > 0$

$$p_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$



Стандартное нормальное распределение имеет параметры a = 0 и σ^2 = 1

Полезные замечания

- Около 68% индивидуумов в популяции имеют значения ξ в пределах а $\pm\sigma$
- Около 95% индивидуумов в популяции имеют значения ξ в пределах а $\pm 2\sigma$
- Около 99,865% индивидуумов в популяции имеют значения ξ в пределах а $\pm 3\sigma$ (правило “3-х сигм”)

- Асимметрия равна нулю, эксцесс равен 3 (или 0 в зависимости от определения коэффициента эксцесса)
- а определяет положение центра распределения
- σ определяет форму распределения

Распределения, связанные с нормальным

Распределение χ^2

Пусть $\xi_1, \xi_2, \ldots \xi_n$ — независимые случайные величины, имеющие стандартное нормальное распределение с параметрами $N(0,1)$.

χ_n^2 – распределением с n степенями свободы называется распределение случайной величины

$$\chi_n^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_n^2 .$$

Это распределение сосредоточенно на положительной полуоси и имеет плотность

$$p_{\chi_n^2}(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

где $\Gamma(\alpha) = \int\limits_0^{+\infty} t^{\alpha-1} e^{-t} dt$ — гамма-функция.

$M\chi_n^2 = n$, $D\chi_n^2 = 2n$, *асимметрия:* $2^{3/2} n^{-1/2}$

Параметр распределения n

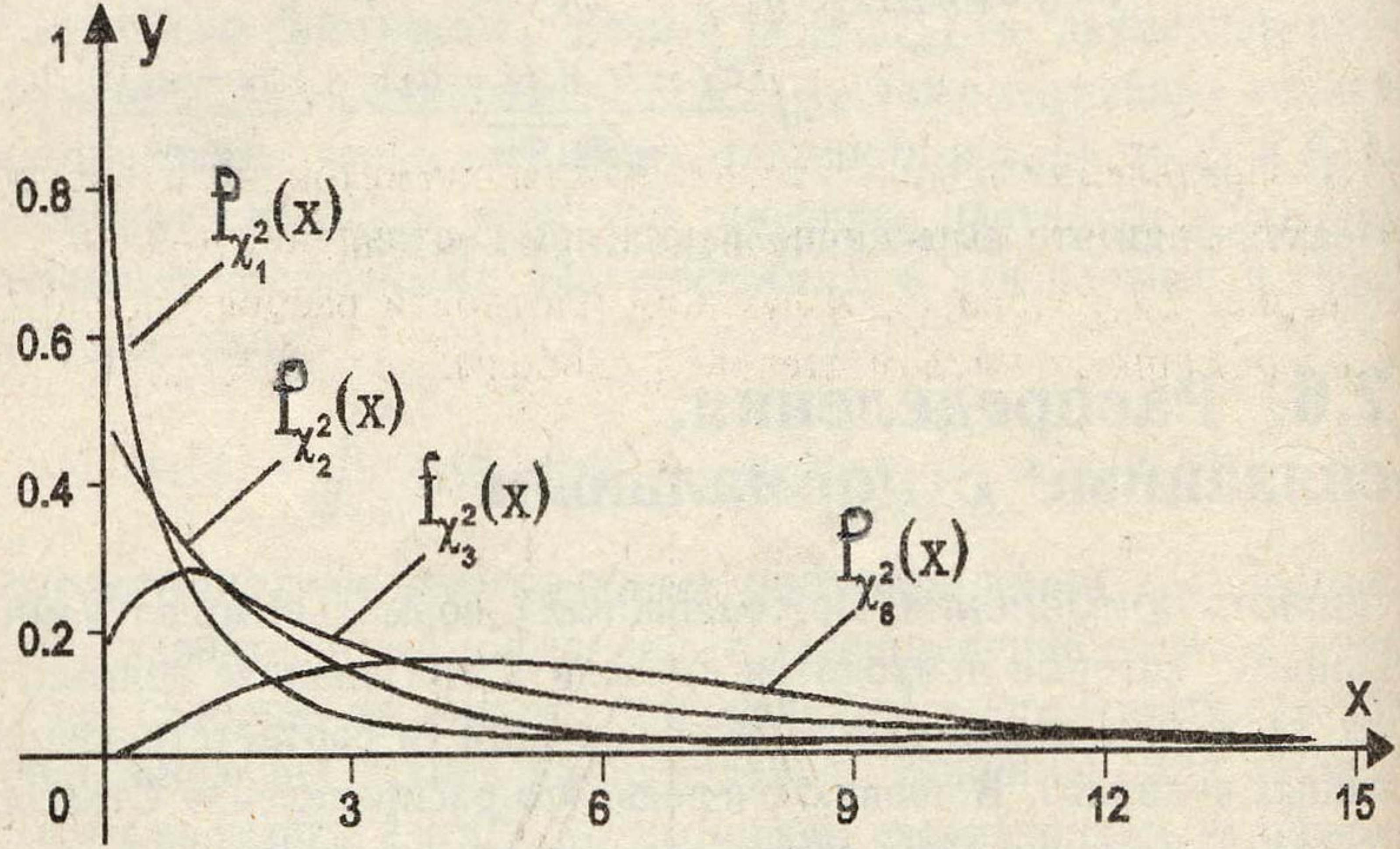


Рис. 2.6. Функции плотности распределения хи-квадрат с различным числом степеней свободы

Распределение Стьюдента

Распределением Стьюдента с n степенями свободы называется распределение случайной величины t_n, где ξ_0 — случайная величина, имеющая стандартное нормальное распределение с параметрами $(0,1)$, а χ_n^2 подчиняется распределению χ^2 с n степенями свободы

$$t_n = \xi_0 \left(\chi_n^2/n\right)^{-1/2} .$$

$$Mt_n=0,\ Dt_n=n/(n-2)$$

Плотность этого распределения представляет собой симметричную функцию, задаваемую формулой

$$p_{t_n}(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}.$$

По форме график функции $p_{t_n}(x)$ напоминает график плотности стандартного нормального закона, но с более медленным убыванием "хвостов".

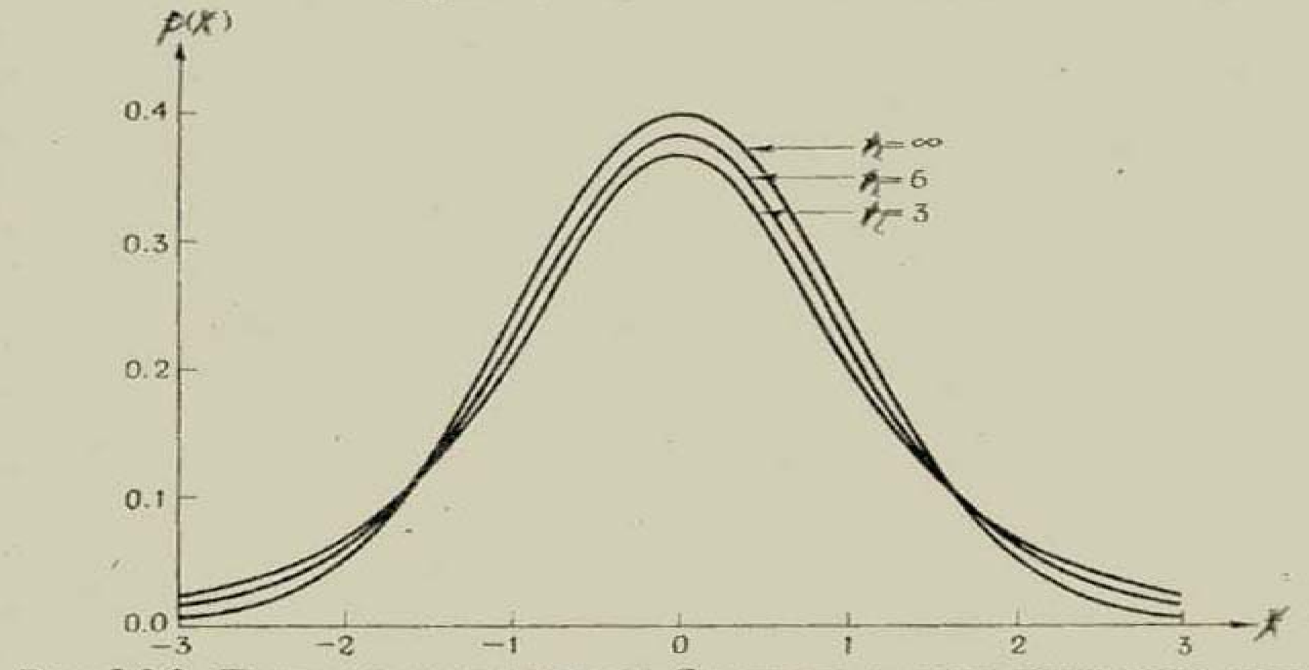


Рис. I.2.3. Плотность распределения Стьюдента с тремя вариантами значений числа степеней свободы n.

F - распределение

Пусть независимые случайные величины ξ и η имеют распределения $\chi^2(\nu_1)$ и $\chi^2(\nu_2)$.

F-распределением с ν_1 и ν_2 степенями свободы называется распределение случайной величины

$$F(\nu_1, \nu_2) = (\chi^2(\nu_1)/\nu_1)/(\chi^2(\nu_2)/\nu_2),$$

Параметры распределения ν_1, ν_2

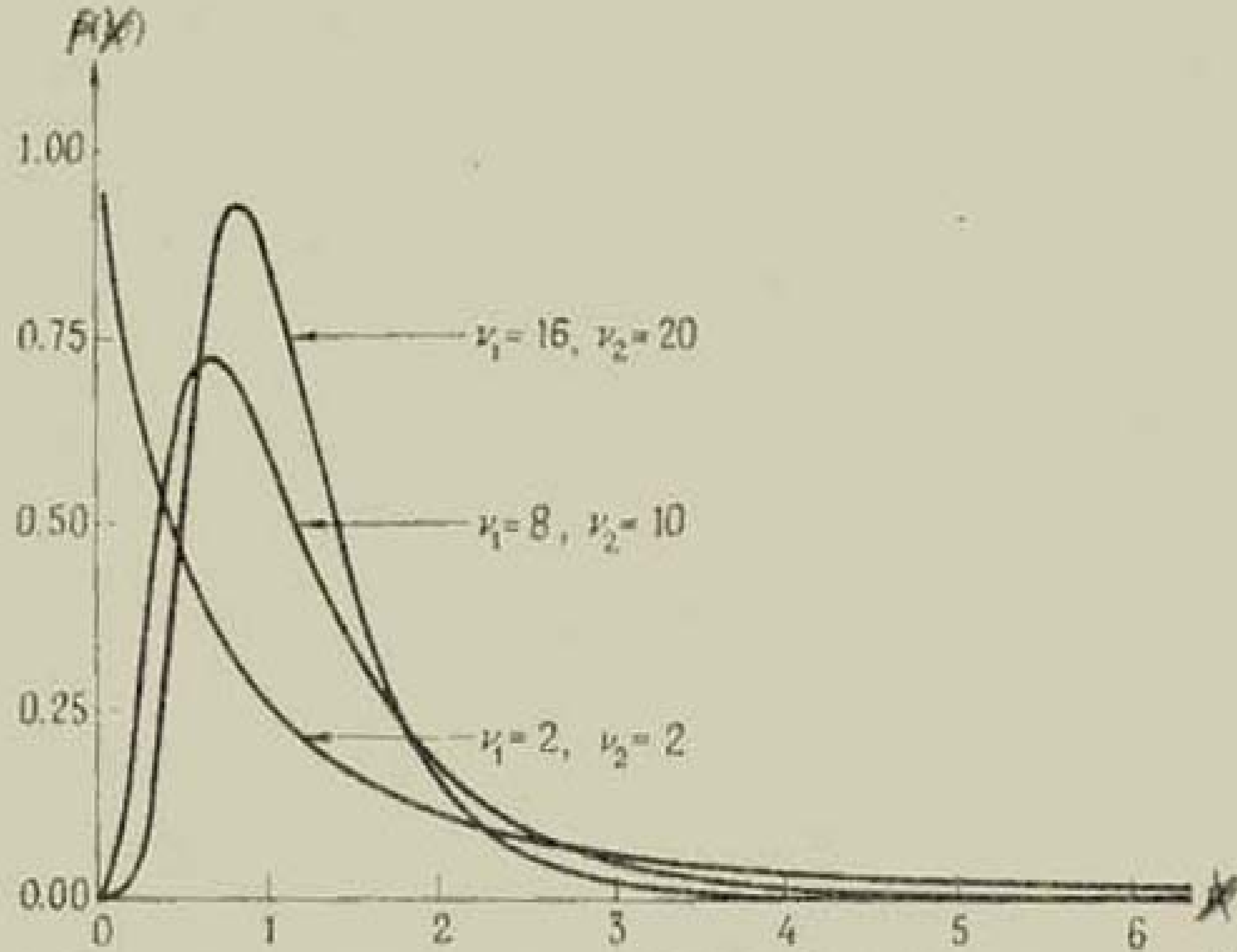


Рис. 1.2.4. Плотность F-распределения с тремя вариантами значений чисел степеней свободы ν_1, ν_2.

Логарифмически-нормальное распределение

Пусть ξ — случайная величина. Если существует такое число θ, что величина $ln(\xi - \theta)$ распределена нормально, то говорят, что ξ имеет логарифмически-нормальное распределение ($\xi > \theta$).

Плотность вероятности случайной величины ξ имеет вид

$$p(x) = \frac{1}{(x-\theta)\,\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\{\ln(x-\theta) - a\}^2/\sigma^2\right].$$

Параметры распределения θ, a, σ^2

$M\xi = \theta + exp(a + (\sigma^2/2))$, $D\xi = exp(2a + \sigma^2)(exp\sigma^2 - 1)$

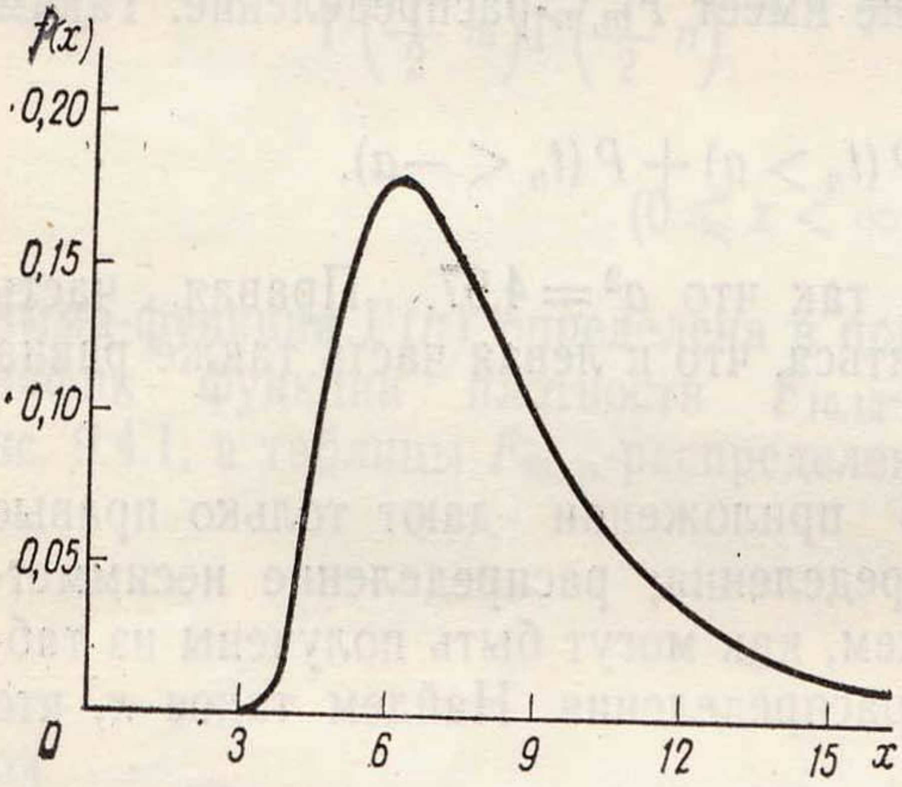


Рис. 9.5.1. Функция плотности вероятностей логарифмически-нормального распределения с параметрами $\theta=3$, $\mu=1{,}5$ и $\sigma^2=0{,}36$. Среднее равно 8,366, дисперсия равна 12,475

ЗАМЕЧАНИЕ

Во многих приложениях параметр θ равен нулю, и ξ имеет двупараметрическое распределение. Оно дает более реалистическое описание совокупностей таких положительно-определенных случайных величин, как напр.рост, вес и т. п., чем нормальное распределение, приписывающее положительные вероятности и отрицательным значениям переменной.

ПРИМЕР (Задание 3.1) Найти вероятность того, что случайная величина t (имеющая t-распределение Стьюдента) с десятью степенями свободы лежит в пределах от 1,5 до 2,0.

ПРИМЕР (Задание 3.2). Известно, что в некоторой стране рост взрослых мужчин приблизительно описывается нормальным распределением со средним значением 175,6 см и стандартным отклонением 7,63 см. Выбирается случайным образом взрослый мужчина. Какова вероятность того, что его рост ниже 185 см?

МНОГОМЕРНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Совместное распределение — термин, относящийся к распределению нескольких случайных величин, заданных на одном и том же вероятностном пространстве

Понятие о совместном (многомерном) распределении

Совместную реализацию случайных величин $\xi_1, \xi_2, \dots \xi_k$ можно представить как вектор $(x_1, x_2, \dots, x_k)$. Выборочным пространством S $(S=R^n)$ является множество всех возможных векторов $(x_1, x_2, \dots, x_k)$.

Любые подмножества вида $\{\xi_1 \le b_1, \xi_2 \le b_2, \dots, \xi_k \le b_k\}$, *где* $b_1, b_2, \dots, b_k$ *— числа из R, являются событиями.*

Рассмотрим событие A из выборочного пространства S.

Понятие о совместном распределении

Пусть $\omega \in \Omega$.

Необходимо измерить k характеристик элемента ω.

Введем случайные величины $\xi_1, \xi_2, \ldots \xi_k$ $(1<k<\infty)$, каждая из которых определена на Ω и принимает значения из R. Конкретное значение $\xi_i(\omega)=x_i$ $(i=1,2,\ldots k)$ есть реализация случайной величины ξ_i, а вектор $(x_1, x_2, \ldots, x_k)$ — совместная реализация случайных величин $\xi_1, \xi_2, \ldots \xi_k$ на элементе ω.

Смысл вероятности P(A) события A

P(A) есть доля элементов в Ω., наборы значений $(x_1, x_2, \ldots, x_k)$ *которых принадлежат событию A.*

Численность возрастных групп супружеских пар для некоторой гипотетической совокупности

Возраст жены, лет	Возраст мужа, лет			Всего
	15—29	30—44	45+	
15—29	21 643 (0,193 79)	5 304 (0,047 49)	976 (0,008 74)	27 923 (0,250 02)
30—44	3 905 (0,034 97)	35 219 (0,315 35)	6 742 (0,060 37)	45 866 (0,410 69)
45+	391 (0,003 50)	4 596 (0,041 15)	32 906 (0,294 64)	37 893 (0,339 29)
Всего	25 939 (0,232 26)	45 119 (0,403 99)	40 624 (0,363 75)	111 682 (1,000 00)

Пример. Пусть из данной совокупности случайным образом выбрана некоторая супружеская пара. Обозначим возраст мужа буквой m, а возраст жены буквой f. Найти следующие вероятности:

$P(m \geq 45;\quad 15 \leq f \leq 29) = 0{,}00874;$

$P(30 \leq f \leq 44) = 0{,}41069;$

$P(m \geq 45) = 0{,}36375;$

$P(15 \leq f \leq 29 \mid m \geq 45) = 0{,}00874 / 0{,}36375 = 0{,}02403;$

$P(m \geq 45 \mid 15 \leq f \leq 29) = 0{,}00874 / 0{,}25002 = 0{,}03496.$

Совместная функция распределения случайных величин

Совместной функцией распределения k случайных величин $\xi_1, \xi_2, \dots \xi_k$ *называется функция*
$F_{\xi_1, \xi_2, \dots \xi_k}(x_1, x_2, \dots, x_k)$ *такая, что*

$$F_{\xi_1, \xi_2, \dots \xi_k}(x_1, x_2, \dots, x_k) = P(\xi_1 \le x_1, \xi_2 \le x_2, \dots \xi_k \le x_k).$$

Полиномиальное (мультиномиальное) распределение

Полиномиальное распределение есть многомерное обобщение биномиального распределения

Полиномиальное распределение применяется при статистической обработке выборок из больших совокупностей , элементы которых разделяются более чем на две категории

Полиномиальное распределение

Полиномиальное распределение является обобщением биномиального распределения. Опишем его в терминах подбрасывания k-гранной кости. Пусть вероятность выпадения грани i при одном бросании равна p_i $(i = 1, 2, \cdots, k)$ и совершается n независимых бросаний, то вероятность события, которое заключается в том, что первая грань выпадет n_1 раз, вторая грань выпадет n_2 раз и.т.д. равна:

$$P(n_1, n_2, \cdots, n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

где

$$\sum_{i=1}^{k} p_i = 1,$$

$$\sum_{i=1}^{k} n_i = n.$$

При каждом бросании исходом является либо выпадение i-той грани с вероятностью p_i, либо невыпадение с вероятностью $1 - p_i$. Из этого вытекает, что среднее число выпадений i-той грани равно np_i и дисперсия равна $np_i(1 - p_i)$.

Results and Conclusions of the Prospective Cardiovaskular Münster Study

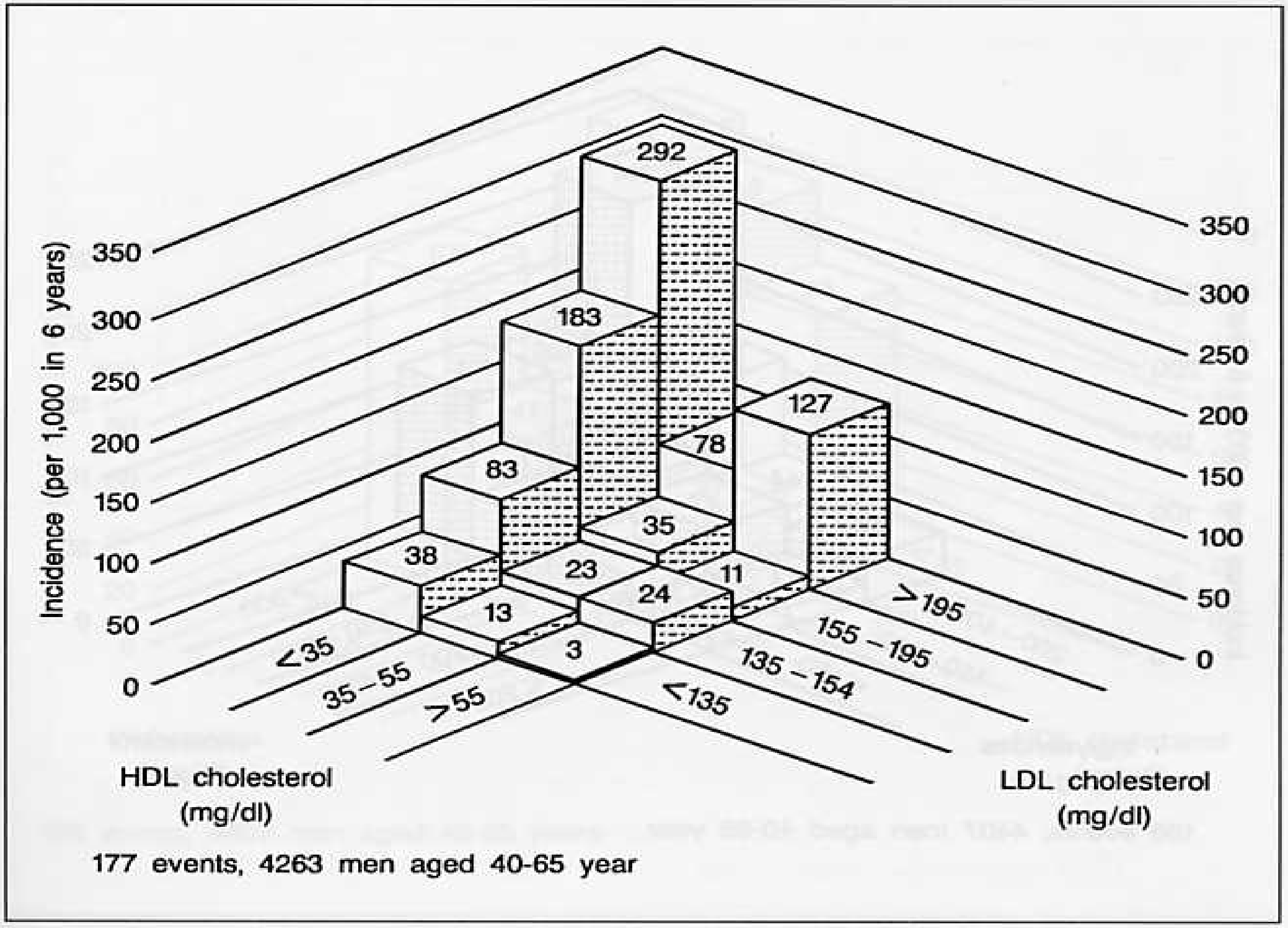


Fig. 22: CHD risk according to LDL cholesterol and HDL cholesterol.

Совместная плотность распределения непрерывных случайных величин

1) $p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) \geq 0$,

2) $$\int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} p_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1,$$

$\forall x_1, \ldots, x_n \in \mathrm{R}$

$$\int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_n} p_{\xi_1 \ldots \xi_n}(y_1, \ldots, y_n)\, dy_1 \cdots dy_n = F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n).$$

Многомерное нормальное распределение

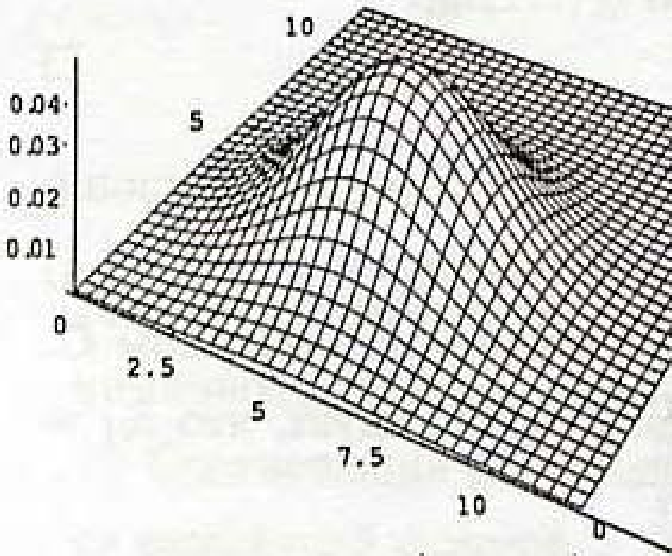
Говорят, что две случайные величины X_1 и X_2 распределены по двумерному нормальному закону, если все линейные комбинации вида $a_1X_1+a_2X_2$ являются нормально распределенными случайными величинами. Функция плотности совместного распределения величин X_1 и X_2 имеет вид

$$f(x_1,\ x_2)=\frac{1}{2\pi\sigma_1\sigma_2\,(1-\rho^2)^{\frac{1}{2}}}\exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2-\right.\right.$$

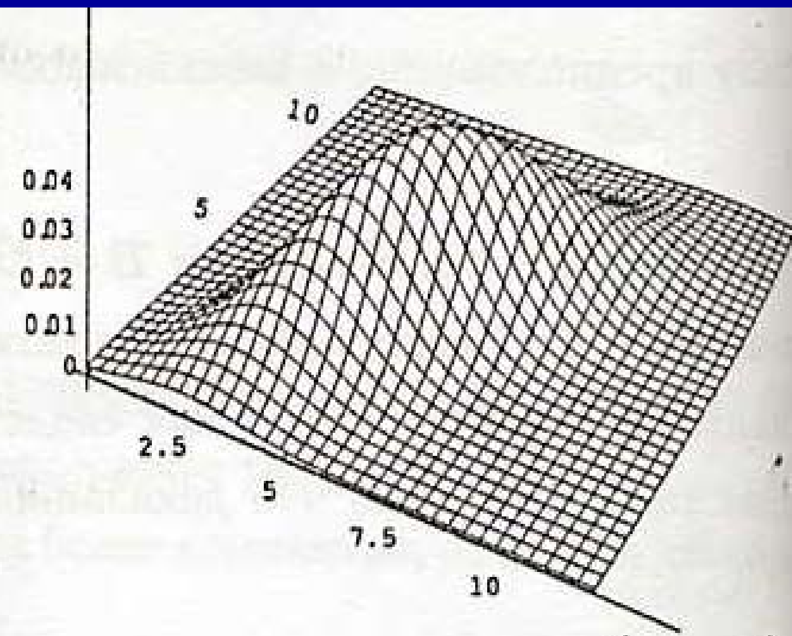
$$\left.\left.-\,2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right]. \qquad (9.6.1)$$

Параметры μ_1 и μ_2 равны математическим ожиданиям величин X_1 и X_2, σ^2_1 и σ^2_2 — дисперсии X_1 и X_2, а ρ — коэффициент корреляции.

Многомерное нормальное распределение



$$a = (5.5, 5),\ B = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$



$$a = (5.5, 6),\ B = \begin{pmatrix} 4 & 4 \\ 4 & 8 \end{pmatrix}$$

Многомерное нормальное распределение

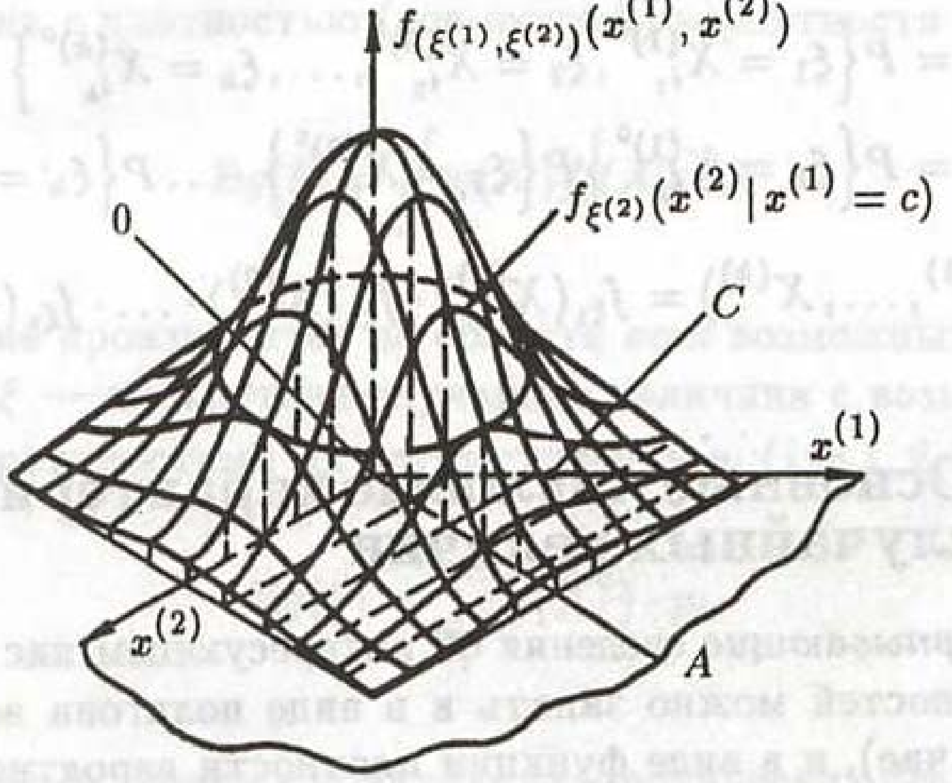


Рис. 2.6. Поверхность двумерной плотности вероятности (нормальный закон с «отсеченными» краями)

Использование моделей законов распределения вероятностей в теории и практике

- Моделирование поведения исследуемых признаков. При этом способ вычисления вероятности того, что признак (в рамках модели) примет то или иное значение (дискретные модели) или будет принимать значения в заданном интервале (непрерывные модели) задается аналитически. Этот способ называют также *параметрической* формой задания закона распределения вероятностей.

- Использование в качестве вспомогательного технического средства при реализации различных методов статистического анализа данных — при построении разного рода статистических оценок и статистических критериев.

В частности, построение точечных и интервальных оценок для неизвестных значений параметров анализируемой модели, статистическая проверка гипотез, связанных со структурой и параметрами рассматриваемой модели невозможна без знания моделей χ^2 , стьюдентовского и F-распределений.

- При подборе модели следует представлять общую схему (механизм) формирования значений той или иной “модельной ” случайной величины

- Прикладные возможности моделей многомерных моделей законов распределения вероятностей значительно скромнее.

  Наибольшее распространение получили ***многомерный нормальный закон*** (для непрерывной многомерной случайной величины) и ***полиномиальный закон распределения вероятностей*** (для дискретной многомерной случайной величины).

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

I. *Доасимптотические*, позволяющие анализировать основные закономерности в поведении случайной величины по ее главным числовым характеристикам — среднему значению, дисперсии и.т.п., без знания общего вида закона распределения (*Неравенство Чебышева*).

II. *Асимптотические*, позволяющие анализировать основные закономерности в поведении сумм большого числа случайных слагаемых (т. е. их сходимость к некоторым постоянным значениям по мере роста числа слагаемых, или описывать асимптотический вид закона распределения этих сумм) без точного знания законов распределения отдельных слагаемых (*Закон больших чисел*, *Центральная предельная теорема*)

III. *Теория преобразований случайных величин*, позволяющие находить закон распределения функций от набора случайных величин, совместное распределение которых нам известно.