МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ**
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
ЮРГИНСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ

УТВЕРЖДАЮ
Зам. директора ЮТИ ТПУ по УР

_______________ В.Л. Бибик
« __ » __________ 2016 г.

А.А. Мицель

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ В АНАЛИТИЧЕСКОЙ ИНФОРМАЦИИ

Методические указания к выполнению практических работ
по дисциплине «Системный анализ, управление и обработка информации в аналитической экономике»
для магистрантов, обучающихся по направлению 09.04.03
«Прикладная информатика»

Издательство
Юргинского технологического института (филиала)
Томского политехнического университета
2016

УДК 681
ББК 32.81
З–38

Составитель **Мицель А.А.**

З–38 Системный анализ, управление и обработка информации в аналитической экономике: методические указания к выполнению практических работ по дисциплине «Системный анализ, управление и обработка информации в аналитической экономике» для магистрантов, обучающихся по направлению 09.04.03 «Прикладная информатика» / Мицель А.А.; Юргинский технологический институт. – Юрга: Изд–во Юргинского технологического института (филиала) Томского политехнического университета, 2016. – 84 с.

УДК 681
ББК 32.81

Методические указания рассмотрены и рекомендованы
к изданию методическим семинаром кафедры
информационных систем ЮТИ ТПУ
« »__________ 2016 г.

Зав. кафедрой ИС
кандидат технических наук __________ *А.А. Захарова*

Председатель учебно–методической
комиссии ________ *Е.В. Молнина*

Рецензент

Кандидат технических наук, доцент кафедры ИС ЮТИ ТПУ
А.В. Маслов

Оглавление

Тема 1. Генерация случайных чисел с заданным законом распределения

1.1. Основные понятия и соотношения

Множество всех возможных значений случайной величины ξ, распределенной по закону F, называется ***генеральной совокупностью*** F.

Множество $\{x_1, x_2, \ldots, x_n\}$ отдельных значений случайной величины ξ, полученных в серии из n независимых экспериментов (наблюдений), называется ***выборочной совокупностью или выборкой*** объема n из генеральной совокупности.

Выборка $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$, в которой элементы упорядочены по возрастанию, называется ***вариационным рядом***.

Совокупность пар чисел $(\bar{x}_i, n_i)$, где $\bar{x}_i$, $i = \overline{1,m}$ – наблюдаемые, неповторяющиеся (для непрерывного распределения) в выборке значения, а n_i – число этих значений в выборке, называется ***статистическим рядом абсолютных частот***. Совокупность пар чисел $(\bar{x}_i, \omega_i)$, где $\omega_i = n_i / n$ называется ***статистическим рядом относительных частот***. Совокупность пар чисел $\left(\bar{x}_i, \sum_{k=1}^{i} \omega_k\right)$ называется ***статистическим рядом накопленных частот***. Статистические ряды отображают в виде таблицы:

$\bar{x}_i$	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_m$
n_i	n_1	n_2	$\cdots$	n_m
ω_i	ω_1	ω_2	$\cdots$	ω_m
$\sum_{k=1}^{i} \omega_k$	ω_1	$\omega_1 + \omega_2$	$\cdots$	1

Подобного вида статистический ряд используют обычно для описания выборки из генеральной совокупности с дискретным распределением. В этом случае статистический ряд относительных частот приближенно оценивает ряд распределения дискретной случайной величины.

Ломаная, отрезки которой соединяют точки $(\bar{x}_i, \omega_i)$, называется ***полигоном частот***. Для дискретной случайной величины полигон частот является оценкой многоугольника распределения.

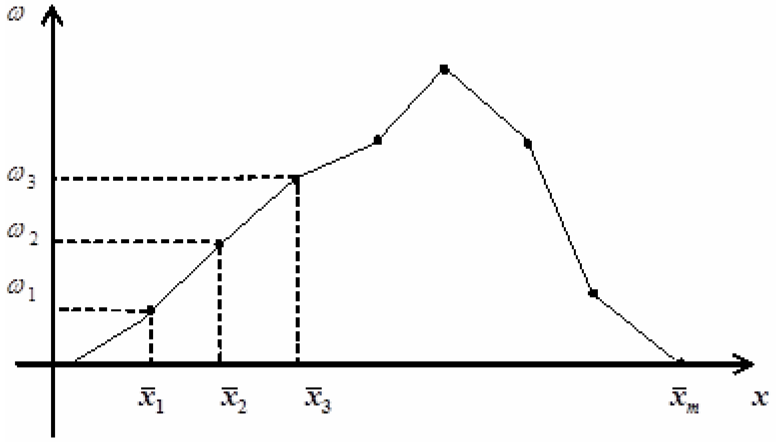

Рис. 1. Полигон частот

Для описания выборки из совокупности с непрерывным распределением используют ***сгруппированные статистические ряды***. Для этого интервал, в котором содержатся все элементы выборки, делится на m равных (или неравных) последовательных, непересекающихся интервалов $\tilde{x}_0 \div \tilde{x}_1, \tilde{x}_1 \div \tilde{x}_2, \ldots, \tilde{x}_{m-1} \div \tilde{x}_m$, и подсчитывают частоты n_i – число элементов выборки, попавших в i–ый интервал. Число интервалов группирования определяют, например, по формуле Стерджесса: $m = 1 + \left[\log_2 n\right] \approx 1 + 4 \cdot \lg n$. В результате получаем следующий статистический ряд:

$\bar{x}_i$	$\bar{x}_1$	$\bar{x}_2$	$\cdots$	$\bar{x}_m$
n_i	n_1	n_2	$\cdots$	n_m
ω_i	ω_1	ω_2	$\cdots$	ω_m
ρ_i	ρ_1	ρ_2	$\cdots$	ρ_m
$\sum \omega_i$	ω_1	$\omega_1 + \omega_2$	$\cdots$	1

Здесь $\bar{x}_i = \dfrac{\tilde{x}_{i-1} + \tilde{x}_i}{2}$ – середины интервалов группирования, $\rho_i = \dfrac{\omega_i}{\Delta x_i} = \dfrac{\omega_i}{\tilde{x}_i - \tilde{x}_{i-1}}$ – плотность частоты.

В качестве оценки кривой плотности непрерывного распределения используется ***гистограмма частот*** – ступенчатая фигура, состоящая из m прямоугольников, опирающихся на частичные интервалы (см. рисунок). Высота i–го прямоугольника полагается равной плотности частоты ρ_i. Соответственно площадь каждого прямоугольника равна $\rho_i \cdot \Delta x_i = \omega_i$ – относительной частоте.

Эмпирической функцией распределения, полученной по выборке $\{x_1, x_2, \ldots, x_n\}$, называется функция, при каждом $x \in R$ равная:

$$F_n^*(x) = \frac{\text{количество } x_i < x}{n}.$$

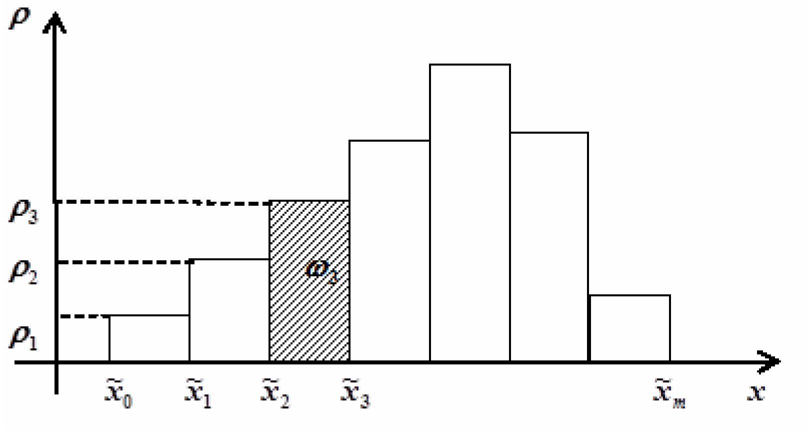

Рис. 2. Гистограмма частот

$F_n^*(x)$ есть ступенчатая функция. Эмпирическая функция распределения является оценкой теоретической функции распределения (рис 3).

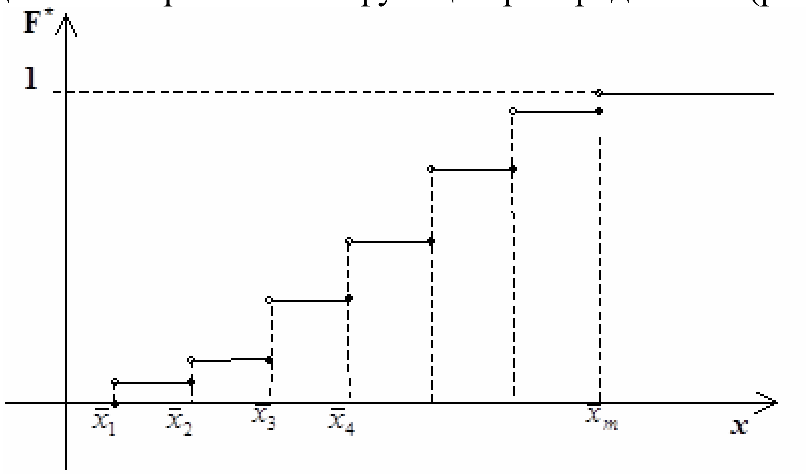

Рис. 3. Эмпирическая функция распределения

В качестве ***<u>числовых характеристик выборки</u>*** используются:

1. Выборочное среднее: $\overline{m} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

2. Выборочная дисперсия $\overline{D} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i{}^2 - \overline{X}^2$.

3. Несмещенная выборочная дисперсия $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2$.

4. Выборочные начальные и центральные моменты

$$\overline{m}_k = \overline{X^k} = \frac{1}{n}\sum_{i=1}^{n} x^k, \qquad \overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}\left(x - \overline{X}\right)^k.$$

По статистическому ряду значения этих величин могут быть найдены по формулам:

$$\overline{m} = \sum_{i=1}^{m} \overline{x}_i \omega_i, \qquad \overline{D} = \sum_{i=1}^{m} (\overline{x}_i - \overline{m})^2 \omega_i, \qquad s^2 = \frac{n}{n-1} \sum_{i=1}^{m} (\overline{x}_i - \overline{m})^2 \omega_i,$$

$$\overline{m}_k = \sum_{i=1}^{m} \overline{x}_i^k \omega_i, \quad \overline{\mu}_k = \sum_{i=1}^{m} (\overline{x}_i - \overline{m})^k \omega_i.$$

Выборочные характеристики являются приближенными значениями соответствующих числовых характеристик случайной величины ξ.

1.2. Некоторые непрерывные законы распределения случайных величин

1) ***Нормальное распределение (рис. 4)***

Определение. Случайная величина ξ имеет нормальное распределение $N(\mu, \sigma^2)$, где $\sigma > 0$, $\mu \in R$, если ее плотность распределения имеет вид:

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

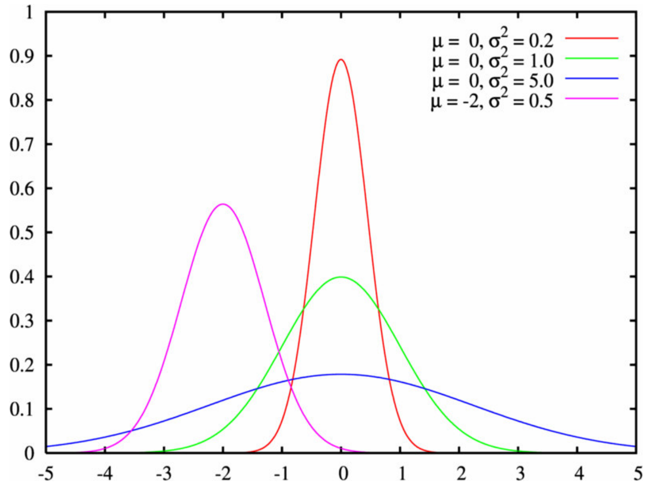

Рис. 4. Нормальное распределение

Числовые характеристики

- математическое ожидание $M(\xi) = \mu$;
- дисперсия $D(\xi) = \sigma^2$;
- коэффициент асимметрии $A = \dfrac{M(\xi-\mu)^3}{\sigma^3} = 0$;
- коэффициент эксцесса $E = \dfrac{M(\xi-\mu)^4}{\sigma^4} - 3 = 0$;

- медиана $med = F^{-1}(0,5) = \mu$;
- мода $\mathrm{mod} = \arg(\max f(x)) = \mu$.

Функция распределения (рис. 5) $F(x) = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$.

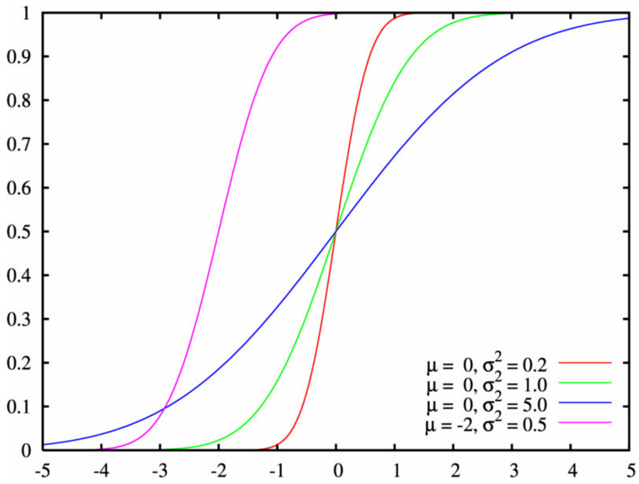

Рис. 5. Функция распределения

Применение.

Простейшие методы моделирования основываются на центральной предельной теореме. Именно, если сложить много независимых одинаково распределённых величин с конечной дисперсией, то сумма будет распределена *примерно* нормально. Например, если сложить 12 независимых базовых случайных величин, получится грубое приближение стандартного нормального распределения. Тем не менее, с увеличением слагаемых распределение суммы стремится к нормальному.

Пример 1. При сетевом планировании выполнения комплекса работ в условиях, когда продолжительности работ являются случайными величинами, то продолжительность пути при достаточно большом количестве работ может рассматриваться как нормальная случайная величина, среднее значение которой равно сумме средних значений отдельных работ, а дисперсия равна сумме дисперсий отдельных работ.

Пример 2. Нормальное распределение и связанное с ним логнормальное очень популярны в классической теории финансовых рынков. Они применяются для моделирования доходностей рисковых активов, таких как, напр., акции или валюты, распределения цены опционных контрактов и т.д. Считается, что на длинных временных интервалах – квартальных, годовых и т.д. «свечках» распределение реальных лог доходностей довольно неплохо приближается нормальным распределением.

2) ***Равномерное распределение (рис. 6)***

Определение. Говорят, что случайная величина ξ имеет непрерывное равномерное распределение $U(a,b)$ на отрезке $[a,b]$, где $a,b \in R$, если её плотность $f(x)$ имеет вид:

$$f(x)=\begin{cases}\dfrac{1}{b-a}, & \text{\textit{если}}\ \ x\in[a,b]\\ 0 & \text{\textit{если}}\ \ x\notin[a,b]\end{cases}$$

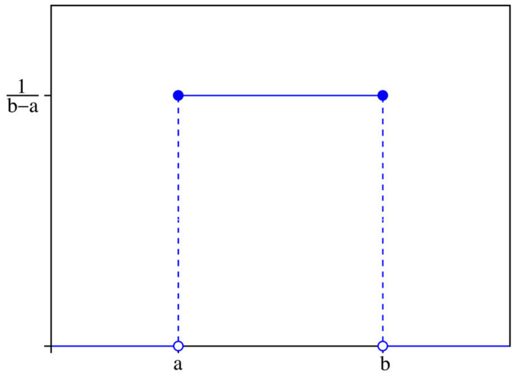

Рис. 6. Равномерное распределение

Пишут: $\xi \sim U(a,b)$, или $\xi \sim U_{a,b}$

Числовые характеристики

- математическое ожидание $M(\xi)=\dfrac{a+b}{2}$;
- дисперсия гамма–распределения $D(\xi)=\dfrac{(b-a)^2}{12}$;
- коэффициент асимметрии $A=0$;
- коэффициент эксцесса $E=-6/5$;
- медиана $med=F^{-1}(0{,}5)=\dfrac{a+b}{2}$;
- мода $\mathrm{mod}=x\in[a,b]$ – любое число из отрезка [a,b].

Функция распределения $F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\le x\le b\\ 1, & x>b\end{cases}$.

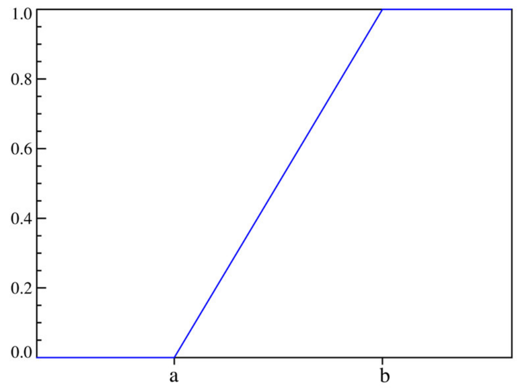

Рис. 7. Функция распределения

Применение

Равномерное распределение можно использовать при расчетах по сетевым графикам работ.

С помощью стандартного равномерного генератора можно построить генератор выборки любого непрерывного распределения с помощью метода обратной функции.

3) *Гамма распределение (рис. 8)*

Определение. Случайная величина ξ имеет гамма распределение $\Gamma(k,\theta)$, где $\theta>0$, $k>0$, если ее плотность распределения имеет вид:

$$f(x)=\begin{cases}\dfrac{x^{k-1}e^{-x/\theta}}{\theta^k\Gamma(k)}, & x>0\\ 0, & x\le 0\end{cases}$$

Здесь $\Gamma(\lambda)=\int\limits_0^\infty t^{\lambda-1}e^{-t}dt-$ гамма функция. $\Gamma(\lambda)=(\lambda-1)\Gamma(\lambda-1)$, $\Gamma(n)=(n-1)!$, $\Gamma(1/2)=\sqrt{\pi}$.

График плотности гамма–распределения имеет вид

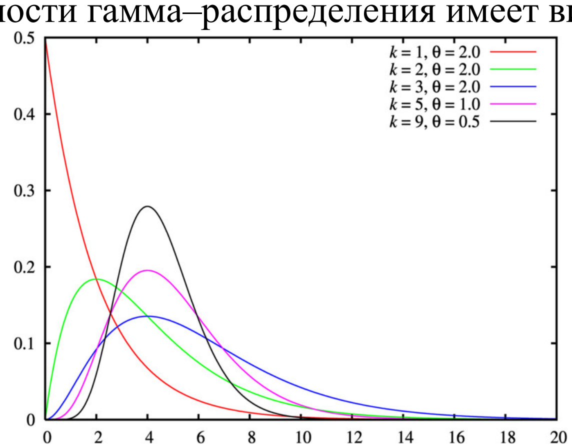

Рис. 8. Гамма распределение

Числовые характеристики

- математическое ожидание – $M(\xi) = k\theta$;
- дисперсия гамма–распределения – $D(\xi) = k\theta^2$;
- коэффициент асимметрии – $A = \dfrac{M(\xi - k\theta)^3}{(k\theta^2)^{3/2}} = \dfrac{2}{\sqrt{k}}$;
- коэффициент эксцесса – $E = \dfrac{6}{k}$.
- медиана
- мода $\text{mod} = (k-1)\theta, \quad k \ge 1$.

Функция распределения (рис. 9) $F(x) = \int\limits_0^x \dfrac{t^{k-1} e^{-t/\theta}}{\theta^k \Gamma(k)} dt$

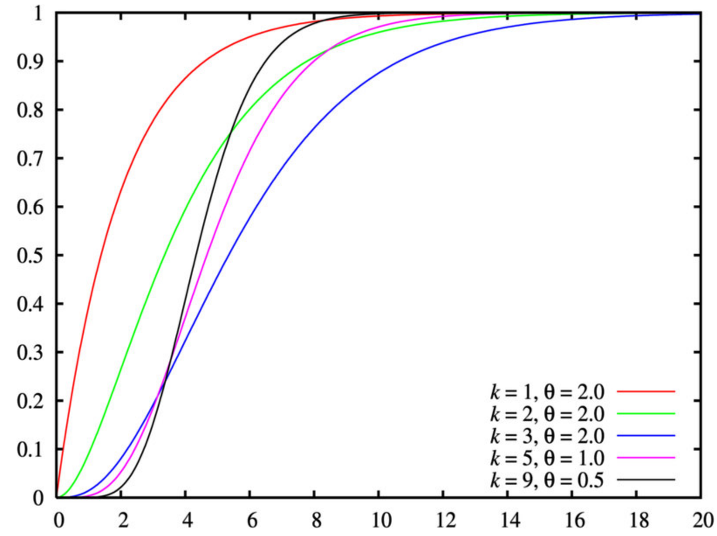

Рис. 9. Функция распределения

4) ***Распределение "хи–квадрат"***

Определение. Пусть $\xi_1, \xi_2, ..., \xi_k$ – независимые случайные величины распределенные по стандартному нормальному закону $N(0,1)$. Тогда случайная величина.

$$x = \sum_{j=1}^{k} \xi_j^2$$

имеет распределение "хи–квадрат" с k степенями свободы и обозначают χ_k^2. (Саму случайную величину также часто обозначают χ_k^2). Распределение хи–квадрат является частным случаем гамма–распределения.

Плотность распределения χ_k^2:

$$f(x)=\frac{1}{2^{k/2}\Gamma(k/2)}x^{\frac{k}{2}-1}e^{-\frac{x}{2}},\quad x>0,$$

а основные числовые характеристики равны:

- математическое ожидание $M(\chi_k^2)=k$;
- дисперсия $D(\chi_k^2)=2k$;
- коэффициент асимметрии $A=\dfrac{M(\xi-k)^3}{(2k)^{3/2}}=\sqrt{8/k}$;
- коэффициент эксцесса $E=12/k$;
- медиана $med=F^{-1}(0,5)=k-2/3$;
- мода $\mathrm{mod}=(k-2),\quad k\geq 2$.

График плотности вероятностей для различных степеней свободы k приведен на рисунке 10.

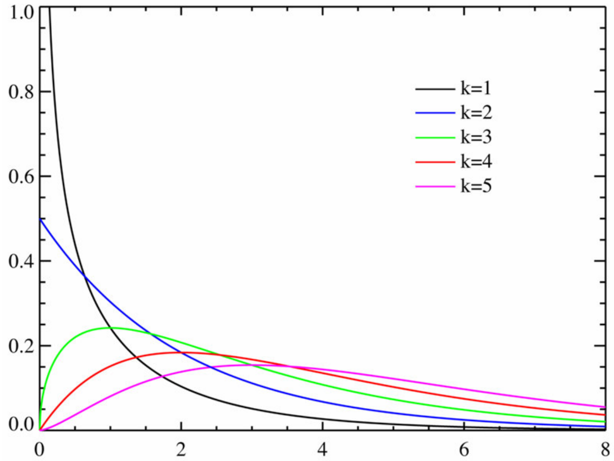

Рис. 10. Распределение "хи–квадрат"

Если случайные величины ξ и η независимы и $\xi\in\chi_k^2$, $\eta\in\chi_m^2$, то, очевидно, их сумма $\xi+\eta\in\chi_{k+m}^2$.

Функция распределения (рис.11) $F(x)=\int\limits_0^x\dfrac{t^{\frac{1}{2}-1}e^{-t/2}}{2^{k/2}\cdot\Gamma(1/2)}dt$

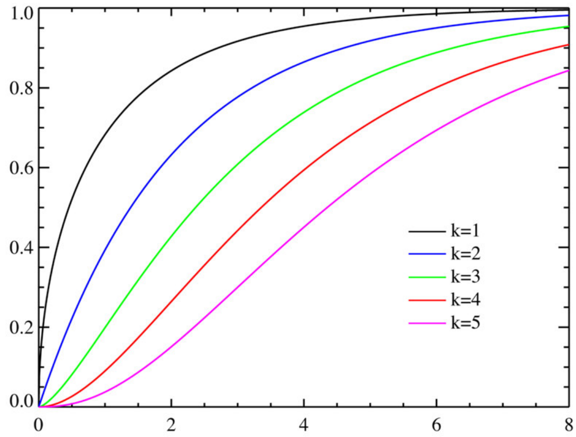

Рис. 11. Функция распределения

Применение

Критерий χ^2 (Хи–квадрат) применяется для проверки гипотезы о законе распределения. Во многих практических задачах точный закон распределения неизвестен, то есть является гипотезой, которая требует статистической проверки.

Обозначим через X исследуемую случайную величину. Пусть требуется проверить гипотезу H_0 о том, что эта случайная величина подчиняется закону распределения $F(x)$. Для проверки гипотезы произведём выборку, состоящую из n независимых наблюдений над случайной величиной X. По выборке можно построить эмпирическое распределение $F^*(x)$ исследуемой случайной величины. Сравнение эмпирического распределения $F^*(x)$ и теоретического распределения $F(x)$, соответствующего гипотезе H_0, производится с помощью критерия согласия χ^2.

5) *Распределение Стьюдента (рис. 12)*

Определение. *Пусть ξ и $\xi_1, \xi_2, ..., \xi_n$ – независимые случайные величины распределенная по закону $N_{0,1}$. Тогда распределение величины*

$$t_n = \frac{\xi}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}\xi_j^2}}$$

называется распределением Стьюдента с n степенями свободы и обозначают T_n.

Плотность распределения Стьюдента:

$$f(x)=\frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x\in R$$

Числовые характеристики

- $M(t_n)=0,\ n>1$;
- $D(t_n)=\frac{n}{n-2},\ n>2$;
- коэффициент асимметрии $A=0,\ n>3$;
- коэффициент эксцесса $E=\frac{6}{n-4},\ n>4$.
- медиана $\mathrm{med}=0$;
- мода $\mathrm{mod}=0$.

Распределение Стьюдента симметрично относительно $M(t_n)=0$.

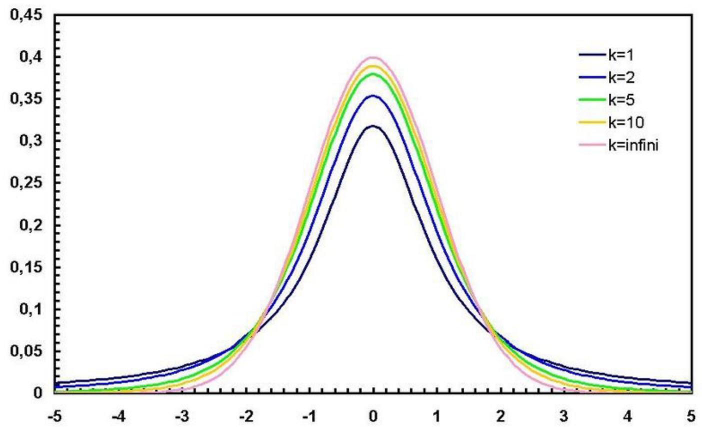

Рис. 12. Распределение Стьюдента

Так как при $n\to\infty$, *согласно закону больших чисел,* $\frac{\chi_n^2}{n}=\frac{1}{n}\sum_{i=1}^{n}\xi_i^2\xrightarrow{p}M\left(\xi^2\right)=M\left(\chi_1^2\right)=1$, *то при* $n\to\infty$ $t_n\Rightarrow N_{0,1}$.

Функция распределения (рис. 13)

$$F(x)=\int_0^x\frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}dt$$

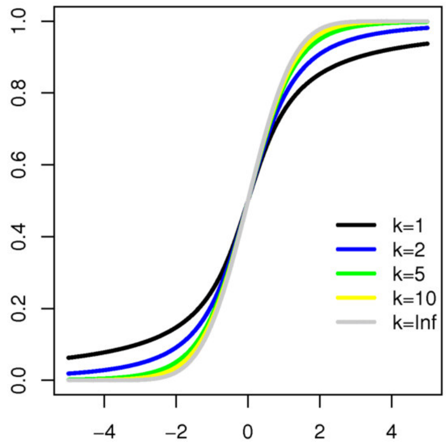

Рис. 13. Функция распределения

Применение распределения Стьюдента

Распределение Стьюдента используется в статистике для точечного оценивания, построения доверительных интервалов и тестирования гипотез, касающихся неизвестного среднего статистической выборки из нормального распределения. В частности, пусть $X_1, \ldots, X_n$ независимые случайные величины, такие что

$X_i \sim \mathrm{N}(\mu, \sigma^2),\ i = 1, \ldots, n$. Обозначим $\bar{X}$ выборочное среднее этой выборки, а S^2 её выборочную дисперсию. Тогда

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim \mathrm{t}(n-1)$$

6) Показательное распределение

Определение. Случайная величина ξ имеет показательное распределение $\Pi(\lambda)$ с параметром $\lambda > 0$ если ее плотность распределения имеет вид:

$$f(x) = \lambda e^{-\lambda x},\ x \in [0, \infty).$$

Числовые характеристики

- $M(\xi) = 1/\lambda$;
- $D(\xi) = 1/\lambda^2$;
- коэффициент асимметрии $A = \dfrac{M(\xi - 1/\lambda)^3}{(1/\lambda)^3} = 2$;
- коэффициент эксцесса $E = \dfrac{M(\xi - 1/\lambda)^4}{(1/\lambda)^4} - 3 = 6$;
- медиана $\mathrm{med} = \ln 2/\lambda$;

- мода $\text{mod} = 0$.

Плотность распределения имеет вид

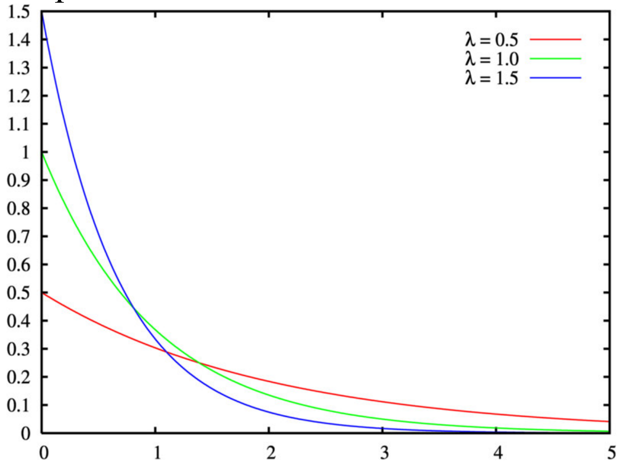

Рис. 14. Показательное распределение

Функция распределения (рис. 15) $F(x) = 1 - e^{-\lambda x}$ показана на следующем рисунке.

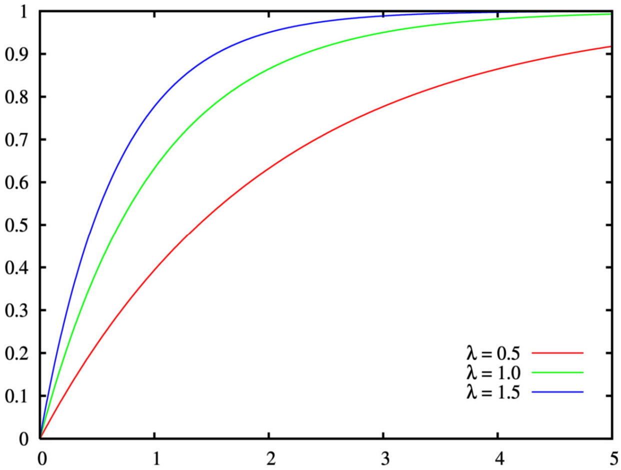

Рис. 15. Функция распределения

Применение.

Экспоненциальный закон распределения применяют для моделирования следующих явлений в системах массового обслуживания:

- времени поступления заказа на предприятие;
- посещения покупателями магазина супер–маркета;
- времени телефонных разговоров;
- срока службы деталей и узлов в компьютере.

7) Распределение Коши (рис. 16)

Определение. Для случайной величины ξ, имеющей распределение Коши $C(x_0,\gamma)$ с параметрами $x_0 \in R$, $\gamma > 0$, плотность распределения имеет вид:

$$f(x) = \frac{1}{\pi}\left[\frac{\gamma}{(x-x_0)^2+\gamma^2}\right] \quad x \in (-\infty,\infty).$$

Числовые характеристики

- математическое ожидание – не существует;
- дисперсия – не существует;
- коэффициент асимметрии – не существует;
- коэффициент эксцесса – не существует;
- медиана $\mathrm{med} = \mathrm{x}_0$;
- мода $\mathrm{mod} = \mathrm{x}_0$

Плотность распределения имеет вид

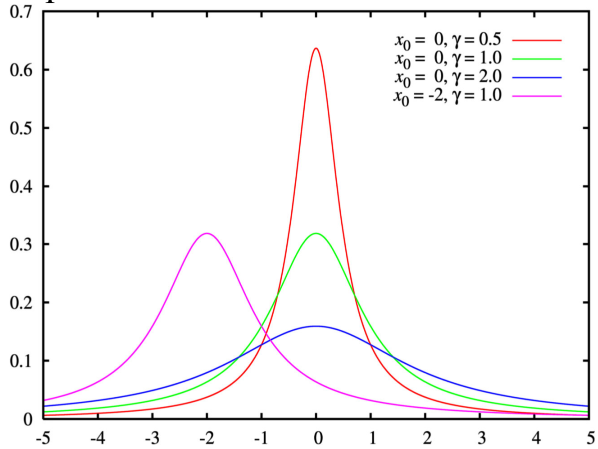

Рис. 16. Распределение Коши

Функция распределения (рис. 17) $F(x) = \frac{1}{2} + \frac{1}{\pi} arctg\left(\frac{x-x_0}{\gamma}\right)$

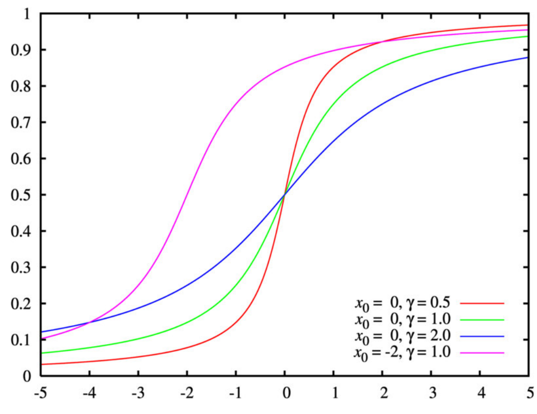

Рис. 17. Функция распределения

Применение

- Распределением Коши характеризуется длина отрезка, отсекаемого на оси абсцисс прямой, закреплённой в точке на оси ординат, если угол между прямой и осью ординат имеет равномерное распределение на интервале $(-\pi;\ \pi)$ (т.е. направление прямой изотропно на плоскости).
- В физике распределением Коши (называемым также формой Лоренца) описываются профили равномерно уширенных спектральных линий.
- Распределение Коши описывает амплитудно–частотные характеристики линейных колебательных систем в окрестности резонансных частот.

***8)* Распределение Лапласа (рис. 18)**

Определение. Распределение Лапласа – это непрерывное распределение случайной величины, при котором плотность вероятности есть

$$f(x)=\frac{\alpha}{2}e^{-\alpha|x-\beta|},\ -\infty<x<\infty,$$

где $\alpha>0$ – параметр масштаба, β – параметр сдвига ($-\infty<\beta<\infty$).

Моменты

- $m_1=\int\limits_{-\infty}^{\infty}x\frac{\alpha}{2}e^{-\alpha|x-\beta|}dx=\beta$, $m_2=\int\limits_{-\infty}^{\infty}x^2\frac{\alpha}{2}e^{-\alpha|x-\beta|}dx=\beta^2+2\alpha^2$,
- $D=m_2-m_1^2=2\alpha^2$.
- Коэффициент асимметрии $A=\dfrac{M(x-\beta)^3}{\left(\sqrt{2}\alpha\right)^3}=0$,
- коэффициент эксцесса $E=\dfrac{M(x-\beta)^4}{\left(\sqrt{2}\alpha\right)^4}-3=3$,
- медиана $\mathrm{med}=\beta$,
- мода $\mathrm{mod}=\beta$.

График плотности распределения дан на рисунке

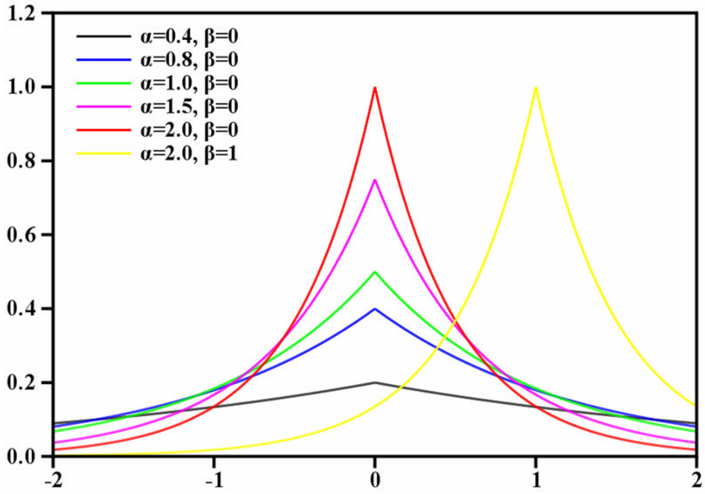

Рис. 18. Распределение Лапласа

Функция распределения имеет вид (рис. 19)

$$F(x)=\int_{-\infty}^{x}\frac{\alpha}{2}e^{-\alpha|x-\beta|}dx=\begin{cases}\dfrac{1}{2}e^{\alpha(x-\beta)}, & x\le\beta,\\ 1-\dfrac{1}{2}e^{-\alpha(x-\beta)}, & x>\beta\end{cases}$$

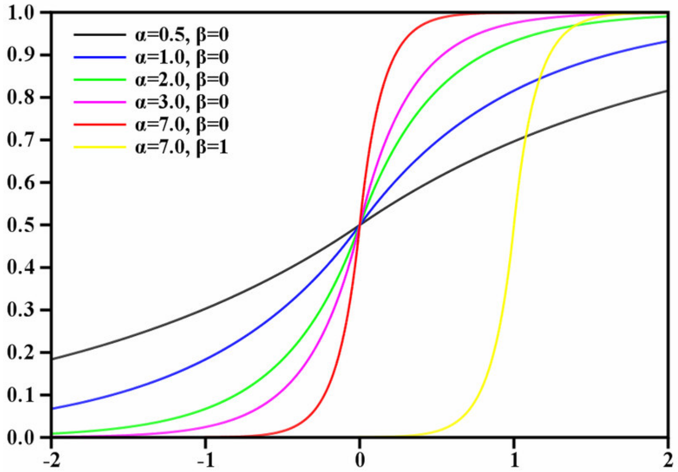

Рис. 19. Функция распределения

Применение

Распределение Лапласа используется для моделирования распределения приращений логарифмов финансовых индексов, для моделирования распределения логарифма размера частиц при дроблении, при моделировании характеристик атмосферной и плазменной турбулентности, при моделировании погрешностей измерений.

***9)* Распределение Рэлея (рис. 20)**

Определение. *Функция плотности вероятностей определяется формулой*

$$f(x)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}},\ x\ge 0,\ \sigma>0,$$

где σ *– параметр масштаба*

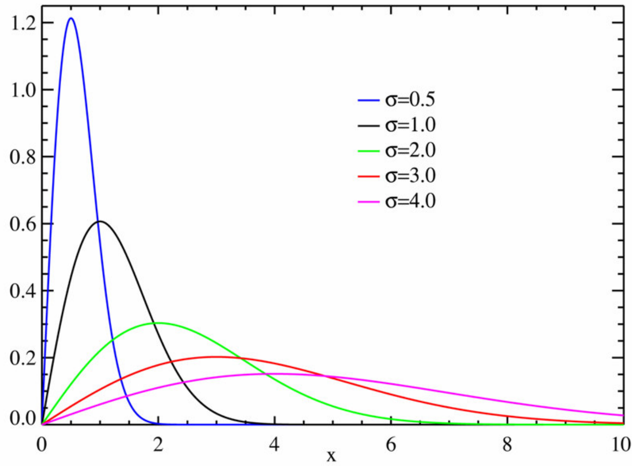

Рис. 20. Распределение Рэлея

Моменты

- $m_1=\int_0^\infty x\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}dx=\sqrt{\frac{\pi}{2}}\sigma,\ m_2=\int_{-\infty}^\infty x^2\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}dx=2\sigma^2,$
- $D=m_2-m_1^2=\left(2-\frac{\pi}{2}\right)\sigma^2.$
- Коэффициент асимметрии $A=\frac{M(x-\beta)^3}{\left(\sigma\sqrt{2-\pi/2}\right)^3}=\frac{2\sqrt{\pi}(\pi-3)}{(4-\pi)^{3/2}},$
- коэффициент эксцесса $E=\frac{M(x-\beta)^4}{\left(\sigma\sqrt{2-\pi/2}\right)^4}-3=-\frac{6\pi^2-24\pi+16}{(4-\pi)^2},$
- медиана $\mathrm{med}=\sigma\sqrt{\ln(4)}$,
- мода $\mathrm{mod}=\sigma$

Функция распределения (рис. 21) $F(x)=1-e^{-\frac{x^2}{2\sigma^2}}$

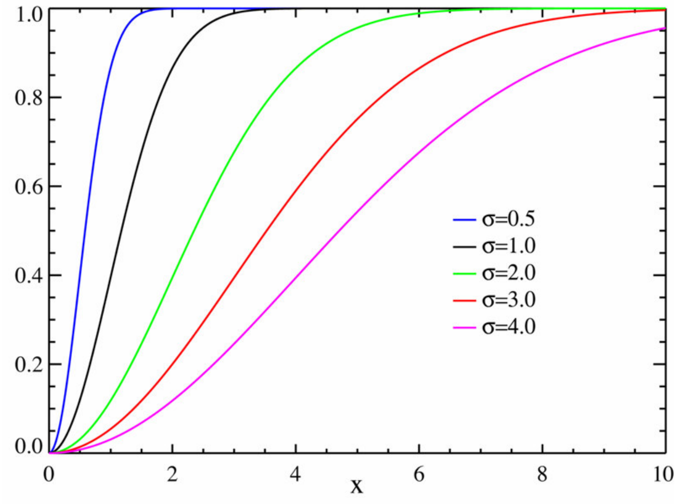

Рис. 21. Функция распределения

Применение

В задачах о пристрелке пушек. Если отклонения от цели для двух взаимно перпендикулярных направлений нормально распределены и некоррелированы, координаты цели совпадают с началом координат, то, обозначив разброс по осям как X и Y, получим выражение для величины промаха в виде $R = \sqrt{X^2 + Y^2}$. В этом случае величина имеет распределение Рэлея.

В радиотехнике для описания амплитудных флуктуаций радиосигнала.

Плотность распределения излучения абсолютно чёрного тела по частотам.

1.3. Дискретные распределения

1) *Биномиальное распределение (рис. 22)*

Биномиальное распределение — распределение количества «успехов» в последовательности из n независимых случайных экспериментов, таких что вероятность «успеха» в каждом из них постоянна и равна p.

Определение. Пусть $X_1, X_2, ..., X_n$ – конечная последовательность независимых случайных величин с распределением Бернулли, то есть

$$X_i = \begin{cases} 1 & p \\ 0 & q = 1 - p \end{cases}, \quad i = 1, 2, ..., n$$

Построим случайную величину Y:

$$Y = \sum_{i=1}^{n} X_i$$

Тогда Y – число единиц (успехов) в последовательности $X_1, X_2, ..., X_n$, имеет биномиальное распределение с n степенями свободы и вероятностью «успеха» p. Пишем: $Y \sim B(n,p)$. Распределения вероятности задаётся формулой:

$$p(k) \equiv P(Y=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}, \; k = 0,1,...,n$$

где $\binom{n}{k} = \dfrac{n!}{(n-k)!k!}$ — биномиальный коэффициент.

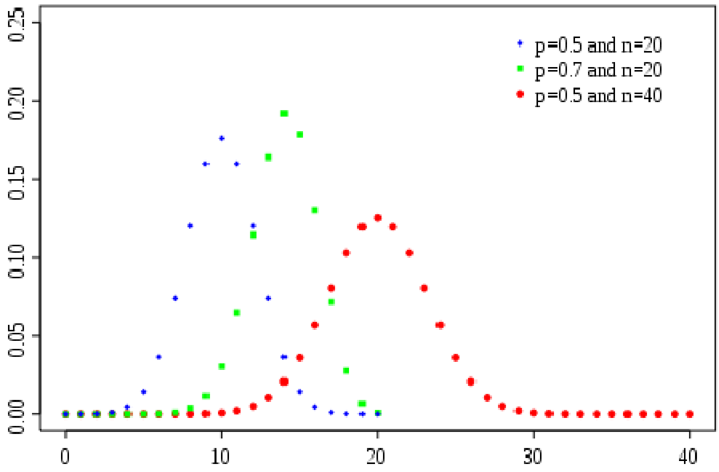

Рис. 22. Биномиальное распределение

Функция распределения

Функция распределения биномиального распределения может быть записана в виде суммы:

$$F(y) \equiv P(Y \le y) = \sum_{k=0}^{[y]} \binom{n}{k} \cdot p^k \cdot q^{n-k}, \; y \in R,$$

где $[y]$ обозначает наибольшее целое, не превосходящее число y.

Моменты

Производящая функция моментов биномиального распределения имеет вид:

$M_Y(t) = E(e^{tY}) = (pe^t + q)^n$, откуда

$E[Y] = np,$

$E[Y^2] = np(q+np)$,

а дисперсия случайной величины $D[Y] = npq$.

Коэффициент асимметрии $A = \dfrac{q-p}{\sqrt{npq}}$, коэффициент эксцесса $E = \dfrac{1-6pq}{npq}$,

медиана $med = \begin{cases} [np]-1 \\ [np] \\ [np]+1 \end{cases}$,

мода $\mathrm{mod} = [(n+1)p]$.

Справка. $E[Y] = \dfrac{d}{dt} M_Y(t) \Big|_{t=0}$, $E[Y^2] = \dfrac{d^2}{dt^2} M_Y(t) \Big|_{t=0}$

Свойства биномиального распределения

Пусть $Y_1 \sim B(n,p)$ и $Y_2 \sim B(n,1-p)$. Тогда $p_{Y_1}(k) = p_{Y_2}(n-k)$.

Пусть $Y_1 \sim B(n_1,p)$ и $Y_2 \sim B(n_2,p)$. Тогда $Y_1 + Y_2 \sim B(n_1+n_2,p)$.

Связь с другими распределениями

Если $n=1$, то получаем распределение Бернулли

Если n большое, то в силу центральной предельной теоремы $B(n,p) \sim N(np,npq)$, где $N(np,npq)$ – нормальное распределение с математическим ожиданием np и дисперсией npq.

Если n большое, а λ – фиксированное число, то $B(n,\lambda/n) \simeq P(\lambda)$, где $P(\lambda)$ – распределение Пуассона с параметром λ.

2) *Распределение Пуассона*

Распределение Пуассона моделирует случайную величину, представляющую собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью и независимо друг от друга.

Определение. Говорят, что случайная дискретная величина Y имеет распределение Пуассона $P(\lambda)$ с параметром λ, если функция вероятности задается следующей формулой

$$p(k) = P(Y=k) = \frac{\lambda^k}{k!} e^{-\lambda},\ \lambda > 0;\ k = 0,1,2,\ldots.$$

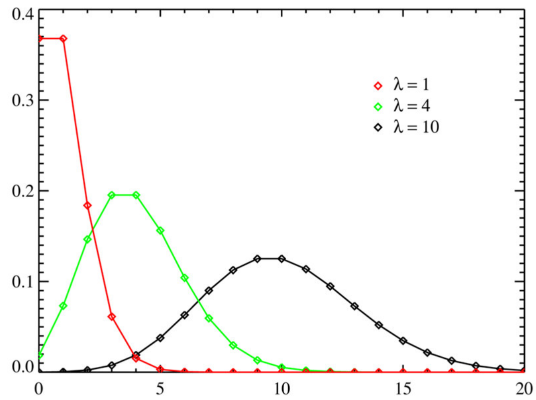

Рис. 23. Распределение Пуассона

Функция распределения (рис. 24)

Функция распределения биномиального распределения может быть записана в виде суммы:

$$F(y) \equiv P(Y \leq y) = \frac{\Gamma(k+1,\lambda)}{k!},$$

где $[y]$ обозначает наибольшее целое, не превосходящее число y.

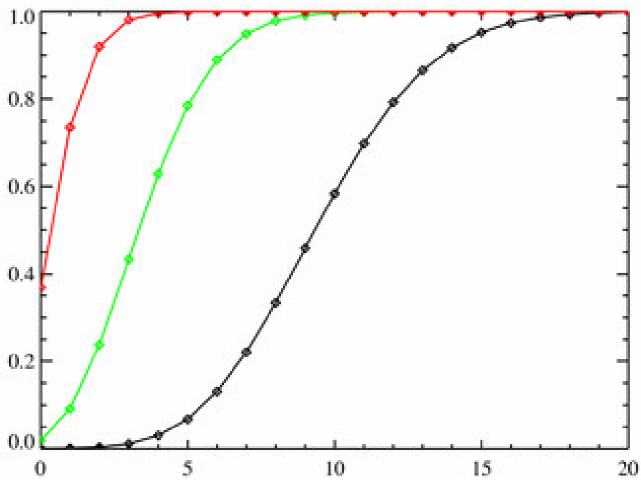

Рис. 24. Функция распределения

Моменты

Производящая функция моментов биномиального распределения имеет вид:

$$M_Y(t) = E(e^{tY}) = e^{\lambda(e^t - 1)},$$

откуда

$$E[Y] = \lambda,$$

$$E[Y^2] = \lambda^2 + \lambda ,$$

а дисперсия случайной величины $D[Y] = \lambda$. Коэффициент асимметрии $A = \lambda^{-1/2}$, коэффициент эксцесса $E = \lambda^{-1}$, медиана $med =$.

Применение

Изначально распределение Пуассона было предложено для моделирования потока входящих телефонных звонков на коммутатор. Примеры других ситуаций, которые можно смоделировать, применив это распределение: поломки оборудования, длительность исполнения ремонтных работ стабильно работающим сотрудником, ошибка печати и др.

3) ***Геометрическое распределение (рис. 25)***

Геометрическое распределение – это распределение дискретной случайной величины, равной количеству испытаний случайного эксперимента до наблюдения первого «успеха».

Определение. Пусть $X_1, X_2, ..., X_n$ – конечная последовательность независимых случайных величин с распределением Бернулли, то есть

$$X_i = \begin{cases} 1 & p \\ 0 & q = 1 - p \end{cases}, \quad i = 1, 2, ..., n$$

Построим случайную величину Y:

$Y = \min\{i \mid X_i = 1\} - 1$ – количество «неудач» до первого «успеха». Распределение случайной величины Y называется геометрическим с вероятностью «успеха» p и обозначается следующим образом: $Y \sim G(p)$.

Числовые характеристики

- *математическое ожидание* $M(Y) = \dfrac{1-p}{p}$;
- *дисперсия* $D(Y) = \dfrac{1-p}{p^2}$;
- коэффициент асимметрии $A = \dfrac{2-p}{\sqrt{1-p}}$;
- коэффициент эксцесса $E = 6 + \dfrac{p^2}{1-p}$

Функция вероятности случайной величины имеет вид:

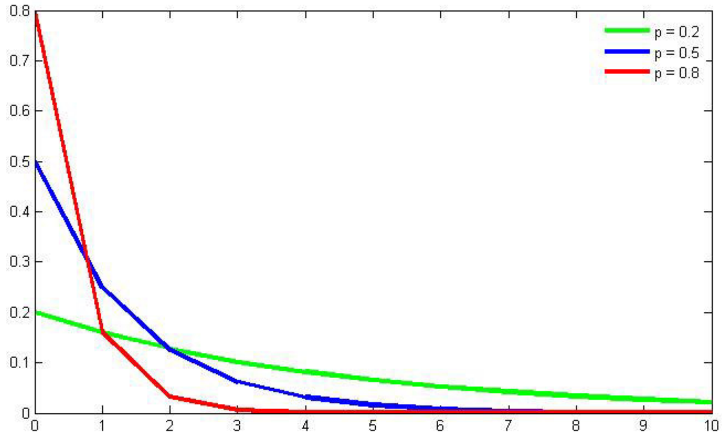

Рис. 25. Геометрическое распределение

Функция распределения (рис. 26)

Функция распределения вероятностей случайной величины Y имеет вид

$$P(Y = n) = q^n p,\ n = 0,1,2,...,$$

Функция распределения вероятностей случайной величины приведена на рисунке

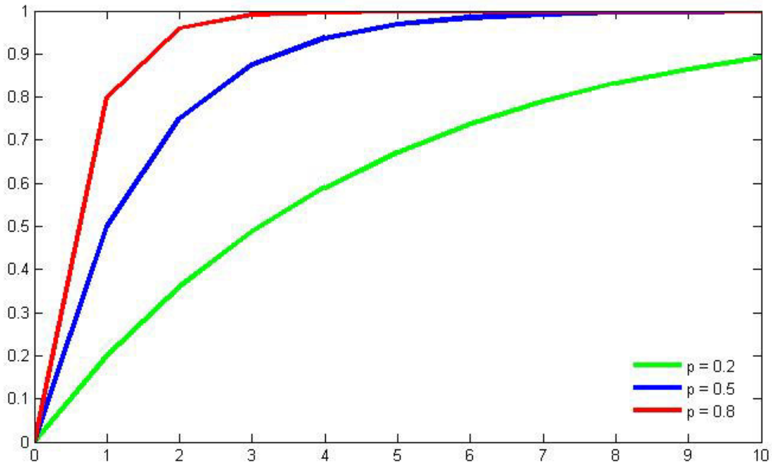

Рис. 26. Функция распределения

Тема 2. Генерирование непрерывных случайных чисел

2.1. Метод обратной функции

Для генерации случайных чисел распределенных по законам, которые отсутствуют в генераторе случайных чисел пакета EXCEL, можно использовать следующие приемы.

1. Воспользуемся следующей теоремой из курса “Теория вероятностей”: Если случайная величина ξ имеет непрерывную и строго моно-

тонную функцию распределения $F_\xi(x)$, а величина $z \in U_{0,1}$, то $\xi = F_\xi^{-1}(z)$, где $F_\xi^{-1}(x)$ – функция обратная к $F_\xi(x)$.

Итак, если $y = F(x)$ – функция распределения непрерывной случайной величины, для которой можно найти обратную $x = F^{-1}(y)$, то генерируя случайную величину z, равномерно распределенную на $[0, 1]$, получим величину $\xi = F^{-1}(z)$ с требуемой функцией распределения $F(x)$. Данный прием можно использовать для генерации случайных величин, распределенных по показательному закону, по закону Коши и др.

Пример 1. Распределение Коши имеет плотность $f(x) = \dfrac{1}{\pi \cdot \theta} \cdot \dfrac{1}{1 + ((x-a)/\theta)^2}$, $x \in R$.

Функция распределения Коши:

$$F(x) = \int_0^x \frac{1}{\pi \cdot \theta} \cdot \frac{1}{1 + ((x-a)/\theta)^2} dx = \frac{1}{\pi} arctg\left(\frac{x-a}{\theta}\right) + \frac{1}{2}.$$

Обратная функция $F^{-1}(x) = a + \theta \cdot tg\left(\pi \cdot (x - 1/2)\right)$. Генерируя случайную величину z, равномерно распределенную на $(0, 1)$, получим случайную величину $x = a + \theta \cdot tg\left(\pi \cdot (z - 1/2)\right)$ с плотностью распределения $f(x) = \dfrac{1}{\pi \cdot \theta} \cdot \dfrac{1}{1 + ((x-a)/\theta)^2}$.

Пример 2. Задан показательный закон с плотностью $f(x) = \lambda e^{-\lambda x}$. Функция распределения $F(x) = \int\limits_{-\infty}^{x} f(x)dx = \int\limits_{-\infty}^{x} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$. Обратная функция $F^{-1}(x) = -\dfrac{1}{\lambda}\ln(1-x)$. Генерируем равномерную случайную величину $z \in U_{0,1}$ и вычисляем искомый результат $x = -\dfrac{1}{\lambda}\ln(1-z)$.

Пример 3. Задан закон распределения Лапласа с плотностью $f(x) = \dfrac{1}{\sigma\sqrt{2}} e^{-\frac{|x-a|\sqrt{2}}{\sigma}}$, $x \in R$.

Функция распределения Лапласа имеет вид

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2}} e^{-\frac{|x-a|\sqrt{2}}{\sigma}} dx = \begin{cases} \dfrac{1}{2} e^{\frac{\sqrt{2}}{\sigma}(x-a)}, & x \le a, \\ 1 - \dfrac{1}{2} e^{-\frac{\sqrt{2}}{\sigma}(x-a)}, & x > a \end{cases}.$$

Моделирование случайной величины, распределенной по закону Лапласа проводим по формулам.

$$x = \begin{cases} a + \dfrac{\sigma}{\sqrt{2}} \ln(2z), & \text{если} \;\; z \le 0,5 \\ a - \dfrac{\sigma}{\sqrt{2}} \ln(2(1-z)), & \text{если} \;\; z > 0,5 \end{cases},$$

где z – равномерная случайная величина из интервала (0,1).

Пример 4. Задан закон распределения Релея с плотностью $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$, $x > 0$.

Функция распределения Рэлея $F(x) = \int\limits_0^x \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx = 1 - e^{-\frac{x^2}{2\sigma^2}}$. Моделирование случайной величины, распределенной по закону Релея проводим по формуле.

$$x = \sqrt{-2\sigma^2 \ln(1-z)}.$$

Пример 5. Задан закон распределение с плотностью $f(x) = \frac{2a}{(1+ax)^3}$, $x \in [0, \infty)$. Функция распределения $F(x) = \int\limits_0^x \frac{2a}{(1+ax)^3} dx = 1 - \frac{1}{(1+ax)^2}$. Случайную величину моделируем по формуле $x = \frac{1}{a}\left(1 - \frac{1}{\sqrt{1-z}}\right)$.

2.2. Метод композиции случайных величин

Если случайная величина есть композиция других случайных величин, то генерируем эти величины и строим из них искомую величину. Данный прием можно использовать, например, для получения случайных величин распределенных по законам χ^2, Стьюдента. Так, случайной величиной, имеющей распределение "хи–квадрат" с k степенями свободы называют величину равную сумме квадратов k независимых стандартных нормальных случайных величин, т.е. $\chi^2 = \sum\limits_{n=1}^{k} \xi_n^2$, $\xi \in N_{0,1}$. Случайной величиной t, имеющей распределение Стьюдента с k степенями свободы называют величину равную $t = \frac{\xi}{\sqrt{\chi^2/k}}$, где ξ – случайная величина распределенная по закону $N_{0,1}$, а χ^2 – независимая от нее случайная величина распределенная по закону хи–квадрат с k степенями свободы.

Тема 3. Критерии проверки гипотезы о законе распределения выборочных данных

Цель работы:

Оценка закона распределения генеральной совокупности на основе выборочных данных.

3.1. Критерий χ^2 (Пирсона) для простой гипотезы

Пусть $\{X_1, X_2, \dots, X_n\}$ выборка из генеральной совокупности F. Проверяется гипотеза $H_0 : F = F_1$ против альтернативы $H_1 : F \neq F_1$.

Представим выборку в виде группированного ряда, разбив предполагаемую область значений случайной величины на m интервалов. Пусть n_i – число элементов выборки попавших в i–ый интервал, а p_i – теоретическая вероятность попадания в этот интервал при условии истинности H_0. Составим статистику $\rho(X) = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i}$, которая характеризует сумму квадратов отклонения наблюдаемых значений n_i от ожидаемых np_i по всем интервалам группирования.

Теорема Пирсона. Если H_0 верна, то при фиксированном m и $n \to \infty$

$$\rho(X) = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i} \Rightarrow \chi^2_{m-1}. \qquad (1)$$

Таким образом, статистику $\rho(\vec{X})$ можно использовать в качестве статистики критерия согласия для проверки гипотезы о виде закона распределения, который будет иметь вид:

$$F(X) = \begin{cases} H_0, & \rho(X) < \tau_{1-\alpha} \\ H_1, & \rho(X) \geq \tau_{1-\alpha} \end{cases}, \text{ где } \rho(X) = \sum_{i=1}^{m} \frac{\left(n_i - np_i\right)^2}{np_i}, \qquad (2)$$

где $\tau_{1-\alpha}$–квантиль распределения χ^2_{m-1}.

Данный критерий называется критерием χ^2 или критерием согласия Пирсона.

Замечание. Критерий не состоятелен для альтернатив, для которых $\tilde{p}_i = p_i$ для всех $i \in \{1, 2, \dots, m\}$. Поэтому, следует стремиться к как можно большему числу интервалов группирования. Однако, с другой стороны, сходимость к χ^2 величины $\frac{(n_i - np_i)^2}{np_i}$ обеспечивается ЦПТ, то есть ожидаемое значение np_i для каждой ячейки не должно быть слишком мало. Поэтому обычно число интервалов выбирают таким образом, чтобы $np_i \geq 5$.

3.2. Критерий χ^2 (Пирсона) для сложной гипотезы

Пусть $\{X_1, X_2, \ldots, X_n\}$ выборка из генеральной совокупности F. Проверяется сложная гипотеза $H_0 : F = F_\theta$, где θ – неизвестный параметр распределения F (или вектор параметров), против альтернативы $H_1 : F \neq F_\theta$.

Пусть выборка по прежнему представлена в виде группированного ряда и n_i – число элементов выборки попавших в i–ый интервал, $i \in \{1, 2, \ldots, m\}$. Статистику (1) мы не можем в этом случае использовать для построения критерия Пирсона, так как не можем вычислить теоретические значения вероятностей p_i, которые зависят от неизвестного параметра θ. Пусть θ^* – оценка параметра θ, а $p_i^*(\theta^*)$ – соответствующие ей оценки вероятностей p_i. Составим статистику

$$\rho(X) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*}.$$

Теорема Пирсона. Если H_0 верна, и l – число компонент вектора θ (число неизвестных параметров распределения), то при фиксированном m и $n \to \infty$

$$\rho(X) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*} \Rightarrow \chi^2_{m-l-1}. \qquad (3)$$

Таким образом, критерий Пирсона для параметрической гипотезы будет иметь вид:

$$F(X) = \begin{cases} H_0, & \rho(X) < \tau_{1-\alpha} \\ H_1, & \rho(X) \geq \tau_{1-\alpha} \end{cases}, \qquad \rho(X) = \sum_{i=1}^{m} \frac{\left(n_i - np_i^*\right)^2}{np_i^*}, \quad (4)$$

где $\tau_{1-\alpha}$ – квантиль распределения χ^2_{m-l-1}.

Замечание. Вообще говоря, оценки, используемые для построения статистики критерия хи–квадрат, должны быть определены из условия минимума статистики $\rho(X)$. Поэтому желательно уточнить оценки, найденные другим способом (методом максимального правдоподобия или методом моментов) путем минимизации $\rho(X)$.

3.3. Метод моментов оценки параметров распределения

Идея этого метода заключается в приравнивании теоретических и эмпирических моментов.

Пусть $X = (x_1, x_2, ..., x_n)$ – независимая выборка из распределения F_θ, зависящего от неизвестного параметра $\theta = (\theta_1, \theta_2, ..., \theta_k) \in \Theta \subset R^k$. Моментом i –го порядка называется функция

$$\mu_i(\theta_1, ..., \theta_k) = E[x^i] = \begin{cases} \int x^i f(x, \theta_1, ..., \theta_k) dx, & if \ \ x \ continuous \ \ value \\ \sum\limits_j x_j^i p(x_j, \theta_1, ..., \theta_k), & if \ x \ discrete \ value \end{cases}$$

где $f(x,\theta)$ – плотность распределения непрерывной случайной величины x, $p(x_j,\theta)$ – вероятность дискретной случайной величины. Теоретический момент является функцией неизвестных параметров $\theta = (\theta_1, \theta_2, ..., \theta_k)$.

Выборочным (эмпирическим) моментом i –го порядка называется величина

$$m_i(x_1, x_2, ..., x_n) = \frac{1}{n} \sum_{j=1}^{n} x_j^i .$$

Отметим, что по своему определению эмпирические моменты являются функциями от выборки.

Для нахождения неизвестных параметров (будем обозначать их $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_k)$) составим систему уравнений.

$$\mu_1(\hat{\theta}_1, ..., \hat{\theta}_k) = m_1,$$

$$\mu_2(\hat{\theta}_1, ..., \hat{\theta}_k) = m_2,$$

....

$$\mu_k(\hat{\theta}_1, ..., \hat{\theta}_k) = m_k.$$

Далее решаем систему относительно параметров $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_k)$. В результате получим

$$\hat\theta_1 = \hat\theta_1(x_1,...,x_n),$$
$$\hat\theta_2 = \hat\theta_2(x_1,...,x_n),$$
$$\ldots\ldots\ldots\ldots\ldots\ldots,$$
$$\hat\theta_k = \hat\theta_k(x_1,...,x_n).$$

Найденные параметры зависят от выборки $X=(x_1,x_2,...,x_n)$.

Пример 1. Пусть $X \in \Pi_\alpha$, где Π_α – показательный закон распределения с параметром α. Найти оценку параметра $\hat\alpha$.

Решение.

$\mu_1(\alpha)=E[x]=\int_0^\infty x\cdot\alpha e^{-\alpha x}dx=\frac{1}{\alpha}$. Приравниваем к $m_1=\frac{1}{n}\sum_{j=1}^{n}x_j$. Отсюда получим: $\hat\alpha=\frac{1}{m_1}=\frac{1}{\frac{1}{n}\sum_{j=1}^{n}x_j}$.

Пример 2. Пусть $X \in U_{\alpha,\beta}$, где $U_{\alpha,\beta}$ – равномерный закон распределения с параметрами α,β. Найти оценки параметров $\hat\alpha$ и $\hat\beta$.

Решение.

$$\mu_1(\alpha,\beta)=E[x]=\frac{1}{\beta-\alpha}\int_\alpha^\beta x dx=\frac{\alpha+\beta}{2},$$

$$\mu_2(\alpha,\beta)=E[x^2]=\frac{1}{\beta-\alpha}\int_\alpha^\beta x^2 dx=\frac{\beta^2+\alpha\beta+\alpha^2}{3}.$$

Получим систему

$$\frac{\alpha+\beta}{2}=m_1,$$
$$\frac{\beta^2+\alpha\beta+\alpha^2}{3}=m_2.$$

Решение системы: $\hat\beta=m_1+\sqrt{3}\sqrt{m_2-m_1^2}=m_1+\sqrt{3}\sigma$, $\hat\alpha=m_1-\sqrt{3}\sigma$. Здесь $\sigma^2=m_2-m_1^2$ – дисперсия выборочного распределения.

3.4. Метод максимального правдоподобия

Пусть $X=(x_1,x_2,...,x_n)$ – независимая выборка из распределения F_θ, зависящего от неизвестного параметра $\theta=(\theta_1,\theta_2,...,\theta_k)\in\Theta\subset R^k$.

Функцией правдоподобия $L(\theta,x)=L(\theta_1,...,\theta_k;x_1,...,x_n)$ называют функцию.

$$L(\theta,x)=\begin{cases}\prod\limits_{j=1}^{n} f(x_j,\theta_1,...,\theta_k), & if\ \ x\ continuous\ \ value\\ \prod\limits_{j=1}^{n} p(x_j,\theta_1,...,\theta_k), & if\ \ x\ discrete\ \ value\end{cases}$$

В качестве оценки параметров $\hat{\theta}=(\hat{\theta}_1,...,\hat{\theta}_k)$ примем значения этих параметров, при которых функция правдоподобия принимает максимальное значение, т.е. $\hat{\theta}=(\hat{\theta}_1,...,\hat{\theta}_k)=\arg\left(\max\limits_{\theta} L(\theta,x)\right)$. Если функция $L(\theta,x)=L(\theta_1,...,\theta_k;x_1,...,x_n)$ является дифференцируемой по переменным $\theta_1,\theta_2,...,\theta_k$, оценки параметров удовлетворяют системе уравнений:

$$\frac{dL(\theta_1,...,\theta_k;x)}{d\theta_i}=0,\ i=1,2,...,k.$$

Пример 3. Пусть $X\in\Pi_\alpha$, где Π_α – показательный закон распределения с параметром α. Найти оценку параметра $\widehat{\alpha}$ методом максимального правдоподобия.

Решение. Запишем функцию правдоподобия.

$L(\alpha,x)=\prod\limits_{j=1}^{n}\alpha e^{-\alpha x_j}=\alpha^n e^{-\alpha\sum\limits_{j=1}^{n}x_j}$. Из условия максимума $\dfrac{dL(\alpha;x)}{d\alpha}=0$ получим следующее уравнение: $n\alpha^{n-1}e^{-\alpha\sum\limits_{j=1}^{n}x_j}-\alpha^n\sum\limits_{j=1}^{n}x_j e^{-\alpha\sum\limits_{j=1}^{n}x_j}=0$. Отсюда следует $\widehat{\alpha}=\dfrac{1}{\frac{1}{n}\sum\limits_{j=1}^{n}x_j}$. Таким образом, эта оценка совпала с оценкой, полученной методом моментов (см. пример 1).

Пример 4. Пусть $X\in U_{\alpha,\beta}$, где $U_{\alpha,\beta}$ – равномерный закон распределения с параметрами α,β. Найти оценки параметров $\widehat{\alpha}$ и $\hat{\beta}$ методом максимального правдоподобия.

Решение.

Функция правдоподобия равна

$$L(\alpha,\beta)=\begin{cases}\prod\limits_{j=1}^{n}\dfrac{1}{(\beta-\alpha)}, & \alpha\le\tilde{x}_1\le,\tilde{x}_2\le\ldots\le\tilde{x}_n\le\beta\\ 0, & \widetilde{X}\in[\alpha,\beta]\end{cases}$$ или

$$L(\alpha,\beta)=\begin{cases}\dfrac{1}{(\beta-\alpha)^n}, & \alpha\le\tilde{x}_1\le,\tilde{x}_2\le\ldots\le\tilde{x}_n\le\beta\\ 0, & \widetilde{X}\in[\alpha,\beta]\end{cases},$$ где $\widetilde{X}=(\tilde{x}_1,...,\tilde{x}_n)$ – упорядоченная выборка (вариационный ряд).

Проанализируем неравенства $\alpha\le\tilde{x}_1\le,\tilde{x}_2\le\ldots\le\tilde{x}_n\le\theta$. Если взять $\beta<\tilde{x}_n$, т.е. меньше максимального значения выборки, то $L(\alpha,\beta)$ обратится в ноль, и только при $\beta=\tilde{x}_n$ функция $L(\alpha,\beta)$ будет отлична от нуля и равна $\dfrac{1}{(\beta-\alpha)^n}$. Этот результат будет тем более верен, если $\beta<\tilde{x}_i,\ i<n$. Поэтому получим $\hat{\beta}=\tilde{x}_n$. Проводя аналогичные рассуждения относительно левой границы интервала, получим $\hat{\alpha}=\tilde{x}_1$.

Таким образом, оценки параметров $\hat{\alpha},\hat{\beta}$, полученные методом максимального правдоподобия, отличаются от оценок этих параметров, полученных методом моментов (см. пример 2).

3.5. Метод наименьших квадратов

Пусть дана табличная функция $y(x)$

x	x_1	x_2	...	x_n
y	y_1	y_2	...	y_n

Необходимо аппроксимировать эти данные некоторой параметрической функцией $f(\theta_1,...,\theta_k;x)$, т.е. заменить функцию $y(x)$ функцией $f(\theta_1,...,\theta_k;x)$:

$$y(x)\approx f(\theta_1,...,\theta_k;x).$$

Параметры $\theta_1,...,\theta_k$ будем подбирать таким образом, чтобы расхождение табличной функции с функцией $f(\theta_1,...,\theta_k;x)$ было минимальным. Для этого построим функционал

$F(\theta_1,...,\theta_k)=\sum_{j=1}^{n}\left(y_i-f(\theta_1,...,\theta_k;x_j)\right)^2$ и найдем его минимум. Необходимое условие минимума имеет вид:

$$\frac{dF(\theta_1,...,\theta_k)}{d\theta_i}=0,\ i=1,...,k.$$

Решаем эту систему уравнений и получаем значения параметров $\hat{\theta}=(\hat{\theta}_1,...,\hat{\theta}_k)$.

Пример 5. Пусть дана независимая выборка $X=(x_1,x_2,...,x_n)$ из распределения F_θ. Разобьем весь диапазон данных $[x_{\min},x_{\max}]$ на m интервалов и построим гистограмму. Обозначим середины интервалов $\bar{x}_i=\frac{\tilde{x}_{i-1}+\tilde{x}_i}{2}$, $i=1,2,...,m$. Тогда плотность частоты $\rho_i=\frac{\omega_i}{\Delta x_i}=\frac{\omega_i}{\tilde{x}_i-\tilde{x}_{i-1}}$ (высоты столбиков гистограммы) будут значениями табличной функции $y_i,\ i=1,2,...,m$. Здесь $\omega_i=n_i/n$ – относительные частоты.

Таким образом, мы получили табличную функцию

$\tilde{x}$	$\tilde{x}_1$	$\tilde{x}_2$	…	$\tilde{x}_n$
ρ	ρ_1	ρ_2	…	ρ_n

Поставим следующую задачу. Подобрать параметры известного закона непрерывного распределения $f(\theta_1,...,\theta_k;x)$ так, чтобы расхождение между гистограммой и функцией $f(\theta,x)$ было минимально. В результате мы приходим к методу наименьших квадратов.

3.7. Пример выполнения задания

Дана выборка объемом $n=100$ из неизвестного распределения F:

12,35	3,72	2,96	4,57	1,01	7,08	10,24	1,66	3,87	20,32
0,89	3,81	0,95	7,62	23,47	2,56	19,15	0,00	8,38	0,67
15,56	0,30	4,83	10,32	13,91	12,25	11,34	13,68	9,42	3,12
11,49	4,11	41,99	15,87	18,78	6,99	17,25	4,27	16,06	5,78
15,69	10,35	11,22	4,24	36,67	22,37	0,22	8,61	16,05	18,06
0,21	20,82	12,36	10,04	2,36	0,44	7,36	10,99	6,36	22,54
28,67	18,77	5,95	17,91	2,59	7,76	1,92	21,88	5,54	11,59
0,71	8,41	2,18	2,50	0,65	4,67	11,17	4,76	0,60	3,39
3,88	39,28	5,70	1,46	3,25	3,57	17,85	1,18	1,34	9,14
25,64	4,07	8,95	25,71	4,94	15,65	1,25	3,01	18,10	6,52

Построим статистический ряд, осуществив группировку данных. Находим $x_{\min} = 0{,}04$, $x_{\max} = 26{,}52$. Число интервалов группирования m определяем по формуле Стерджесса: $m = 1 + [\log_2 n] = 7$. Для удобства возьмем в качестве нижней границы первого интервала значение $\tilde{x}_0 = 0$, а в качестве верхней границы последнего интервала значение $\tilde{x}_7 = 42$, тогда длина каждого интервала группирования будет равна $\Delta x = 42/7 = 6$. Подсчитывая частоты, получаем следующий ряд:

Интервал $\tilde{x}_{i-1} - \tilde{x}_i$	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30	30 – 36	36 – 42
Частота n_i	46	23	14	11	3	0	3

Видим, что частоты распределены по интервалам крайне неравномерно, поэтому делаем перегруппировку данных, добиваясь более равномерного распределения частот по интервалам. В результате получаем следующий ряд:

Интервал $\tilde{x}_{i-1} - \tilde{x}_i$	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 42
Середина $\bar{x}_i$	2	6	10	12	18	22	33
Частота n_i	33	20	16	9	10	6	6
Относительная частота ω_i	0,338	0,20	0,16	0,09	0,10	0,06	0,06
Плотность частоты ρ_i	0,0825	0,0500	0,0400	0,0225	0,025	0,0150	0,00333

Соответствующая гистограмма приведена на рисунке 27.

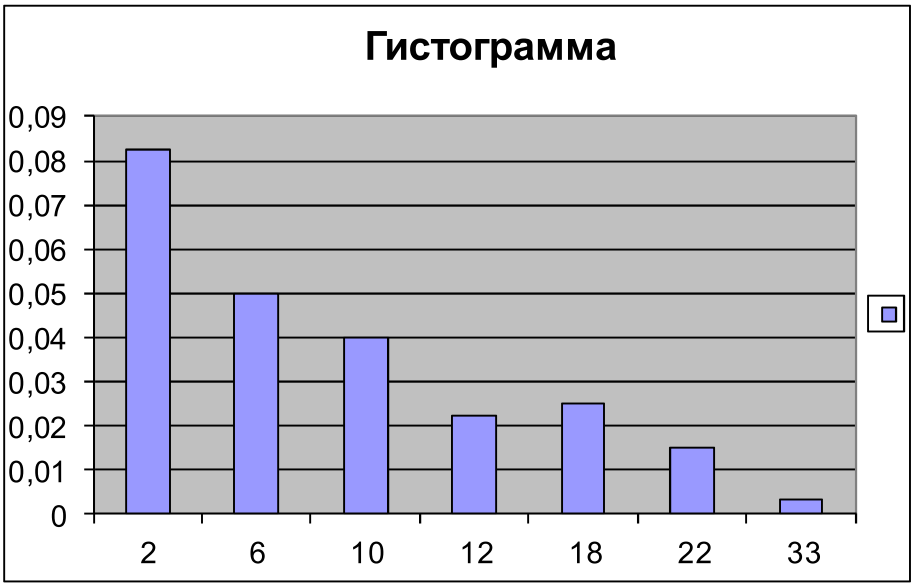

Рис. 27. Гистограмма распределения

Анализируя гистограмму, видим, что распределение экспериментальных данных похоже на показательное распределение. Таким образом, выдвигаем гипотезу о том, что выборочные данные имеют показательное распределение. В качестве оценки неизвестного параметра α этого распределения возьмем оценку, полученную по методу моментов (через первый момент). Так как $m_1 = 1/\lambda$, то $\lambda^* = 1/\overline{X} = 1/9{,}72 \approx 0{,}103$. Вычислим значения плотности показательного распределения $f(x) = \lambda^* e^{-\lambda^* x}$ в точках, соответствующих серединам интервалов группирования, и сравним гистограмму с графиком плотности (рис. 28):

Интервал $\tilde{x}_{i-1} - \tilde{x}_i$	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 42
Середина $\bar{x}_i$	2	6	10	12	18	22	33
Плотность частоты ρ_i	0,0825	0,0500	0,0400	0,0225	0,025	0,0150	0,00333
Теоретическая плотность	0,08377	0,0555	0,03677	0,02993	0,01614	0,01069	0,0034

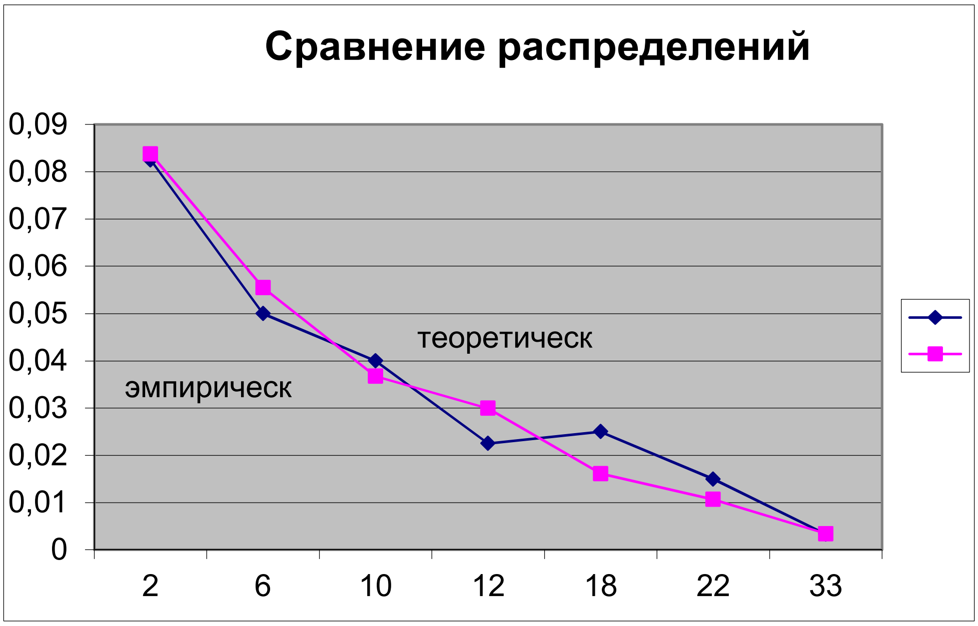

Рис. 28. Сравнение распределений

Найдем также оценки коэффициента асимметрии и эксцесса распределения и сравним их с коэффициентом асимметрии и эксцессом показательного распределения.

Выборочный коэффициент асимметрии $\overline{A} = \dfrac{1}{(n-1)s^3}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^3 \approx 1{,}4$

Выборочный эксцесс $\overline{E} = \dfrac{1}{(n-1)s^4}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4 - 3 \approx 2{,}15$.

Коэффициент асимметрии для показательного распределения $A = 2$, эксцесс $E = 6$. Хотя различие есть, тем не менее можно утверждать, что данные выборочные характеристики имеют смещение (относительно нуля) в сторону характеристик показательного закона.

Применим критерий Пирсона для проверки нашей гипотезы о законе распределения выборочных данных. Подсчитаем вероятности p_i^* попадания в каждый интервал при условии, что генеральная совокупность имеет показательное распределение с параметром $\lambda^* = 0{,}103$: $p_i^* = F(\tilde{x}_i) - F(\tilde{x}_{i-1}),\quad i = \overline{1,7}$, $F(x) = 1 - e^{-\lambda^* x}$ – функция распределения показательного закона. Причем для последнего интервала, полагаем $\tilde{x}_7 = \infty$ и, соответственно, $F(\tilde{x}_7) = 1$, поскольку теоретически для показательного распределения плотность отлична от нуля на интервале $(0, \infty)$. Далее находим ожидаемые значения – np_i^* и нормированные квадраты отклонений $(np_i^* - n_i)^2 / np_i^*$ по всем интервалам. Результаты оформляем в виде таблицы:

Интервал	0 – 4	4 – 8	8 – 12	12 – 16	16 – 20	20 – 24	24 – 42
Частота n_i	33	20	16	9	10	6	6
Вероятность p_i^*	0,3374	0,2236	0,1481	0,0981	0,0650	0,0431	0,0133
Ожидаемые значения np_i^*	33,74	22,36	14,81	9,81	6,502	4,31	8,46
$(n_i - np_i^*)^2 / np_i^*$	0,0164	0,248	0,0951	0,06758	1,8811	0,6643	0,7147

Находим наблюдаемое значение статистики критерия Пирсона: $\rho_{набл} = \sum_{i=1}^{7}\dfrac{\left(n_i - np_i^*\right)^2}{np_i^*} \approx 3{,}15$.

Зададим уровень значимости $\alpha = 0{,}05$. Для заданного уровня значимости и числа степеней свободы $k = m - l - 1 = 7 - 1 - 1 = 5$ ($l = 1$ – так как один параметр распределения λ мы оценивали по выборке) найдем критическое значение статистики, как критическую точку распределения χ_5^2 уровня $\alpha = 0{,}05$ (или что тоже самое – квантиль уровня 0,95): $\rho_{кр} = 11{,}07$. (Критическую точку заданного уровня можно получить, например, используя функцию пакета EXCEL – ХИ2ОБР).

Так как $\rho_{набл} < \rho_{кр}$, то гипотеза о распределении данных по показательному закону принимается.

Уточним значение λ^*, минимизируя наблюдаемое значение статистики $\rho_{набл}$. Используя последовательные итерации, можно получить оценку $\lambda_1^* = 0{,}101$, при которой наблюдаемое значение $\rho_{набл} \approx 3$. Видим, что в данном случае эта оценка практически не отличается от оценки метода моментов.

Тема 4. Линейная регрессия

Цель работы:

Оценка уравнения линейной регрессии на основе выборочных данных

4.1. Построение модели регрессии

Рассмотрим линейную по коэффициентам модель регрессии:

$$y = f\left(x;\beta\right) + \varepsilon = \beta_0 + \beta_1 f_1(x) + \beta_2 f_2(x) + \ldots + \beta_k f_k(x) + \varepsilon, \qquad (1)$$

где ε – случайная величина с математическим ожиданием равным нулю и дисперсией σ^2; $f_j(x),\ j = 1,...,k$ – некоторые заданные функции.

Полагая, $x_j = f_j(x),\quad j = \overline{1,k}$ перейдем к модели множественной линейной регрессии:

$$y = f\left(x;\beta\right) + \varepsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon. \qquad (2)$$

Пусть для оценки неизвестных параметров $\beta_j,\ j = \overline{0,k}$ уравнения регрессии (2) взята выборка объемом n из значений величин $(Y, X_1, X_2, \ldots, X_k)$. Тогда

$Y = X\beta + \varepsilon$,

где $Y = \left(y_1, y_2 ..., y_n\right)^T$ – вектор значений переменной y;

$\beta = \left(\beta_0, \beta_1, \ldots, \beta_k\right)^T$ – вектор параметров модели;

$\varepsilon = \left(\varepsilon_1, \ldots, \varepsilon_n\right)^T$ – вектор ошибок, где $\varepsilon_i \in N\left(0, \sigma^2\right)$ и независимы;

X – матрица исходных данных переменных X_j размерами $n \times (k+1)$. Первый столбец матрицы X содержит единицы (значения фиктивной переменной x_0), остальные столбцы значения переменных $x_1, x_2, ..., x_k$:

$$X = \begin{pmatrix} 1 & x_1^1 & \cdots & x_k^1 \\ 1 & x_1^2 & \cdots & x_k^2 \\ & & \cdots & \\ 1 & x_1^n & \cdots & x_k^n \end{pmatrix}$$

Для нахождения оценки β^* вектора параметров $\beta = (\beta_0, \beta_1, \ldots, \beta_k)^T$ используем метод наименьших квадратов, согласно которому в качестве оценок $\beta_0^*, \beta_1^*, \ldots, \beta_k^*$ берутся такие, которые минимизируют сумму квадратов Q отклонений значений y_i от $f(\vec{x}_i)$:

$$Q = \sum_{i=1}^{n} \left(y_i - f(\vec{x}_i)\right)^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = \left(Y - X\beta\right)^T \left(Y - X\beta\right). \qquad (3)$$

Оценка β^* метода наименьших квадратов имеет вид:

$$\beta^* = \left(X^T X\right)^{-1} X^T Y. \qquad (4)$$

4.2. Оценка погрешности регрессии

Качество регрессионной модели можно оценить, используя оценку s^2 дисперсии предсказания σ^2:

$$s^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} e^2 = \frac{1}{n-k-1} e^T e, \qquad \text{где}$$

$\hat{y}_i = \beta_0^* + \beta_1^* x_i + \ldots \beta_k^* x_k$. Качество модели также можно оценить с использованием оценки коэффициента детерминации: $R^2 = \dfrac{\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$.

Чем ближе значения R^2 к 1, тем большую долю дисперсии величины Y объясняет модель регрессии.

4.3. Оценка адекватности модели

Для оценки адекватности модели вычисляется статистика

$$F=\frac{\frac{1}{k}\sum_{i=1}^{n}(\hat{y}_i-\overline{y_i})^2}{\frac{1}{n-k-1}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}.$$ Если $F>F_{\alpha}(k;n-k-1)$, то модель признается адекватной. Здесь α – уровень значимости.

Можно показать, что $F=\left(\frac{n-k-1}{k}\right)\frac{R^2}{1-R^2}$, где R^2 – квадрат коэффициента детерминации (множественный коэффициент корреляции)

4.4 Анализ регрессионных остатков

Определенную информацию об адекватности уравнения регрессии дает исследование остатков вида $e_i=\widehat{y}_i-y_i$. Если выборочная регрессия $\widehat{y}$ удовлетворительно описывает истинную зависимость между y и x, остатки e_i должны быть независимыми нормально распределенными случайными величинами с нулевым средним и в значениях ε_i должен отсутствовать тренд. Нормальность распределения остатков e_i может быть установлена одним из критериев согласия (см. раздел 3.3).

Гипотезу о равенстве $M(e)=0$ можно проверить любым параметрическим или непараметрическим критерием сравнения среднего с заданным значением (в нашем случае с нулем, см. раздел 4.4).

Независимость в последовательности значений e_i $(i=1,...,n)$ может быть проверена с помощью сериального коэффициента корреляции Дарбина–Ватсона. Статистика сериального коэффициента корреляции имеет вид:

$$D=\frac{\sum_{i=2}^{n}\left(e_i-e_{i-1}\right)^2}{\sum_{i=1}^{n}e_i^2}.$$

Если $D>D_1(\alpha)$ или $D>4-D_1(\alpha)$, то с достоверностью α принимается гипотеза о наличии соответственно отрицательной или положительной корреляции остатков.

Если $D_2(\alpha)>D>D_1(\alpha)$ или $4-D_1(\alpha)>D>4-D_2(\alpha)$, то критерий не позволяет принять решение по гипотезе о наличии или отсутствии корреляции остатков. Если $D_2(\alpha)<D<4-D_2(\alpha)$, то гипотеза корреляции остатков отклоняется. Критические значения $D_1(\alpha)$ и $D_2(\alpha)$ для различных α и числа k коэффициентов в регрессии приведены в табл. 11 (см. статистические таблицы).

4.5. Оценка дисперсии коэффициентов регрессии и доверительных интервалов

Оценка дисперсии коэффициента β_j находится по формуле: $s_j^2 = S^2\left[(X^T X)^{-1}\right]_{jj}$, где $\left[(X^T X)^{-1}\right]_{jj}$ соответствующий диагональный элемент матрицы $(X^T X)^{-1}$.

Доверительные интервал для σ^2 находится с использованием статистики $\chi^2 = (n-k-1)S^2/\sigma^2$, которая при нормальном распределении ε_i имеет распределение хи–квадрат с $(n-k-1)$ степенью свободы.

Для проверки значимости коэффициентов уравнения регрессии используем статистику $t_j = \dfrac{\beta_j^*}{\sqrt{S^2\left[(X^T X)^{-1}\right]_{jj}}}$, которая при истинности гипотезы $H_0:\beta_j = 0$, имеет распределение Стьюдента с $(n-k-1)$ степенью свободы. Если для заданного уровня значимости α значение $|t_j|$ больше критического $t_{крит} = t_{1-\alpha/2}(n-k-1)$, то нулевая гипотеза отвергается и коэффициент признается значимым. В противном случае коэффициент признается незначимым, и соответствующее слагаемое исключается из модели.

В пакете Excel рассчитывается также уровень значимости α статистики $|t_j|$, т.е. вероятность $P\left(x > \left|t_j\right|\right)$. Степень значимости параметров распределения качественно определяется по уровню значимости: не значимые ($\alpha \geq 0{,}100$), слабо значимые ($0{,}100 > \alpha \geq 0{,}050$), статистически значимые ($0{,}050 > \alpha \geq 0{,}010$), сильно значимые ($0{,}010 > \alpha \geq 0{,}001$), высоко значимые ($0{,}001 > \alpha$).

Доверительный интервал для значений y в точке $X_0 = (1, x_1^0, \ldots, x_k^0)^T$ будет иметь вид:

$\left[\left(X_0^T\beta^*\right) \pm t_{1-\alpha/2} S\left(1+\sqrt{X_0^T\left[(X^T X)^{-1}\right]X_0}\right)\right]$, так как погрешность модели $y = f(x)+\varepsilon$ будет определяться двумя источниками: погрешностью $(\Delta f)^2 = S^2\left(X_0^T\left[(X^T X)^{-1}\right]X_0\right)$, связанной с погрешностями параметров модели, и погрешностью собственно модели $\varepsilon^2 = S^2$.

Здесь t_α – квантиль распределения Стьюдента с $n-k-1$ степенью свободы.

4.6. Пример построения уравнения регрессии

Имеется выборка значений совместно наблюдаемых величин X и Y:

X	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
Y	2,96	0,61	4,63	2,44	2,23	4,89	4,98	3,89	6,74	8,07
X	5,5	6	6,5	7	7,5	8	8,5	9	9,5	10
Y	8,34	9,56	9,30	12,35	11,46	11,09	7,91	8,16	6,54	7,88

Требуется подобрать подходящую модель регрессии, характеризующую зависимость Y от X, если известно, что ошибка $\sigma^2 = 1{,}3$.

Нанесем точки (X,Y) на координатную плоскость – построим корреляционное поле, соответствующее нашей выборке (рис. 29)

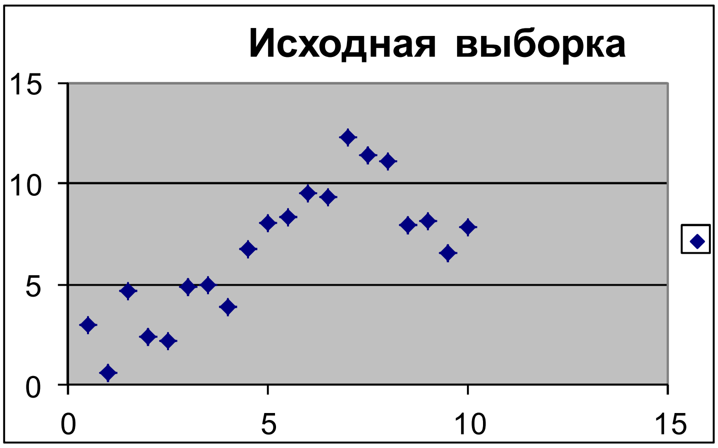

Рис. 29. Корреляционное поле

Видим, что существует зависимость, между значениями X и Y, причем зависимость явно нелинейная. Попробуем аппроксимировать эту зависимость для начала полиномами различных порядков. Возьмем в качестве уравнения регрессии квадратное уравнение:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

Чтобы воспользоваться МНК для оценки коэффициентов, проведем линеаризацию модели, положив $x_1 = x$, $x_2 = x^2$, получим

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Тогда оценку вектора параметров, согласно МНК, найдем как

$$B^* = \left(X^T X\right)^{-1} X^T Y$$

Здесь X – матрица, первый столбец которой содержит единицы, а второй и последующий значения x_1 и x_2.

Для облегчения подбора модели можно воспользоваться встроенными функциями пакета EXCEL (для выбранной модели все равно потом потребуется провести все вычисления вручную, чтобы построить доверительные интервалы). В пакете анализа необходимо выбрать функцию “регрессия”, задать столбец значений Y и матрицу, соответствующую X (единичный столбец в этом случае задавать не надо). Если выбрать вывод остатков, то помимо регрессионной статистики, будут выведены и предсказанные значения Y, т.е.

$$y^* = \beta_0^* + \beta_1^* x + \beta_2^* x^2$$

Для нашей модели регрессионная статистика, полученная пакетом Fxcel будет иметь следующий вид:

Таблица 1.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,852379622
R–квадрат	0,72655102
Нормированный R–квадрат	0,694380552
Стандартная ошибка	1,820336831
Наблюдения	20

Таблица 2.

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	149,6702188	74,83510939	22,58750495	1,63336E–05
Остаток	17	56,32303623	3,313119778		
Итого	19	205,993255	10,841750	0,305589	

Таблица 3.

	Коэффициенты	*Стандартная ошибка*	*t–статистика*	*P–Значение*
Y–пересечение	–0,963028418	1,354297293	–0,711090853	0,486670901
Переменная X1	2,604940094	0,594036759	4,385149663	0,000403854
Переменная X2	–0,167559332	0,054953956	–3,049085893	0,007253372

(продолжение таблицы 3)

Нижние 95%	*Верхние 95%*

–3,82010793	1,894090386
1,351577249	3,858001699
–0,283477711	–0,051610008

Здесь в первой таблице:

- Множественный R – корень квадратный из коэффициента детерминации $\sqrt{R^2}$;
- R–квадрат – коэффициент детерминации $R^2 = \dfrac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$;
- Нормированный R–квадрат – это скорректированная величина коэффициента детерминации, вычисляемая по формуле $\hat{R}^2 = 1 - (1 - R^2)\dfrac{n-1}{n-k-1}$;
- Стандартная ошибка – значение $S = \sqrt{S^2}$, где $S^2 = \dfrac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – оценка дисперсии предсказания σ^2;
- Наблюдения – объем выборки n.

Во второй таблице:

- df – степени свободы v;
- SS – сумма квадратов разностей:

1) между модельными значениями и средним

$$SS_{регрессия} = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 = k \cdot S_R^2;$$

2) остатки $SS_{остаток} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = S^2 \cdot (n - k - 1)$;

3) между исходными данными и средним

$$SS_{итого} = \sum_{i=1}^{n}(y_i - \overline{y})^2 = (n-1) \cdot S_y^2;$$

- $MS = SS / df$;
- F – статистика $F_{Excel} = \dfrac{\frac{1}{k}\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\frac{1}{n-k-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \dfrac{S_R^2}{S^2}$.

Если рассчитать $F = \dfrac{S^2}{S_y^2}$, то получим $F = \dfrac{3,313119}{10,841750} = 0,305589$.

- Значимость F – значение вероятности $F(x,k,n-k-1)$ при $x = \dfrac{S_R^2}{S^2}$, т.е. это уровень значимости принятия нулевой гипотезы H_0.

В третьей таблице:

- Коэффициенты – значения оценок коэффициентов $\beta_0^*, \beta_1^*, \beta_2^*$;
- Стандартная ошибка – значения оценок среднеквадратичных отклонений коэффициентов $s_j = \sqrt{S^2 \cdot \left[(X^T X)^{-1}\right]_{jj}}$;
- t–статистика – наблюдаемые значения статистик критерия проверки значимости коэффициентов соответственно $t_j = \dfrac{\beta_j^*}{\sqrt{S^2\left[(X^T X)^{-1}\right]_{jj}}}$;
- P–значения – достигнутые значения уровня значимости $P\left(x > |t_j|\right)$.
- Нижние и верхние границы 95%–го доверительного интервала $\beta_j^* \pm t_{0,05}(v)\sqrt{S^2\left[(X^T X)^{-1}\right]_{jj}}$.

Соответствующий график предсказанных значений в сравнении с исходными данными имеет вид (рис. 30):

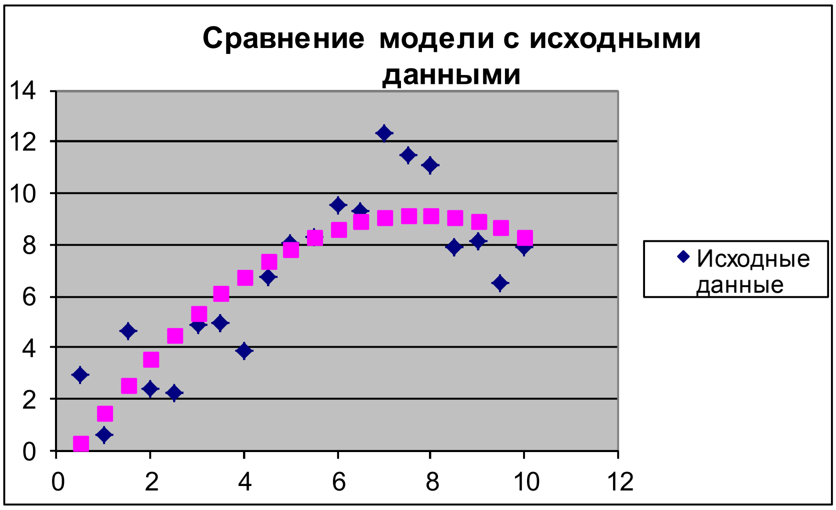

Рис. 30. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 x^2$ *с исходными данными*

Отметим, что полученная оценка значения $\sigma = \sqrt{1,3} = 1,14$ велика: $S = 1,82$. Что касается коэффициентов модели, то, кроме β_0, все они значимо отличаются от нуля (достигнутый уровень значимости достаточно мал, поэтому можно отвергнуть гипотезу о равенстве коэффициентов нулю).

Попробуем улучшить модель, увеличим порядок полинома, пусть

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Проводим линеаризацию, полагая $x_1 = x$, $x_2 = x^2$, $x_3 = x^3$, и оцениваем коэффициенты новой модели.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,925507816
R–квадрат	0,856564717
Нормированный R–квадрат	0,829670602
Стандартная ошибка	1,358958106
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t–статистика*	*P–Значение*
Y–пересечение	3,176148907	1,484433212	2,13963746	0,048142193
Переменная X 1	–1,619155418	1,194561911	–1,355438679	0,194104204
Переменная X 2	0,814063749	0,261005991	3,118946605	0,006612096
Переменная X 3	–0,062325275	0,016365816	–3,808259635	0,001545564

График показан на рис. 31

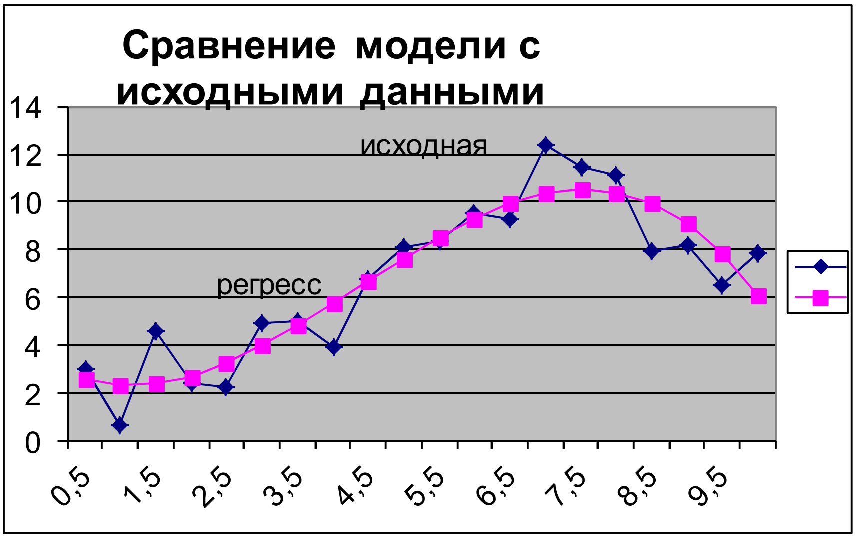

Рис. 31. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ *с исходными данными*

Заметим, что коэффициент детерминации увеличился, а оценка σ (стандартная ошибка) уменьшилась, что говорит о лучшем качестве модели по сравнению с предыдущей. Причем значение этой оценки близко к значениям $\sigma = 1,3$, указанному в задании. Из коэффициентов можно считать, что β_1 не значимо отличается от нуля (достигнутый уровень значимости $\alpha = 0{,}194$, говорит о том, что при истинности гипотезы $H_0 : \beta_1 = 0$, такое или большее значение t–статистики критерия могло наблюдаться с вероятностью 0,194). Поэтому, можно положить $\beta_1 = 0$ и модель соответственно примет вид:

$$y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$$

Результаты регрессионного анализа для этой модели:

Регрессионная статистика	
Множественный R	0,916566765
R–квадрат	0,840094635
Нормированный R–квадрат	0,82128224
Стандартная ошибка	1,39201886
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t–статистика*	*P–Значение*
Y–пересечение	1,356030259	0,648133062	2,092209668	0,051733578
Переменная X 1	0,469599546	0,060941974	7,70568321	6,05722E–07
Переменная X 2	–0,041727702	0,006223514	–6,70484584	3,6971E–06

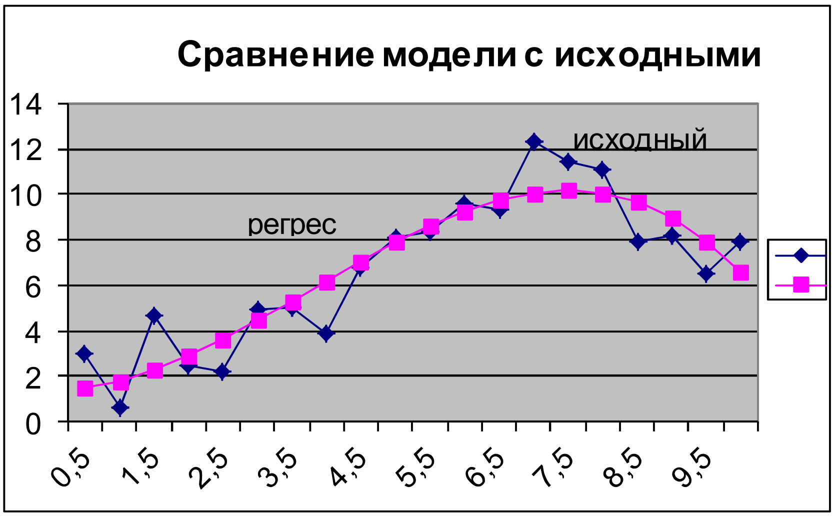

Рис. 32. Сравнение модели $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$ *с исходными данными*

И хотя, параметры модели немного ухудшились (сравните R–квадрат и стандартную ошибку!), тем не менее все коэффициенты получились значимыми, поэтому данная модель, предпочтительнее предыдущей.

Можно ли повысить еще качество модели? В классе полиномов это сделать не удастся. Повышение порядка полинома (можно проверить!), уже больше не понижает стандартную ошибку. Следовательно, улучшение нужно искать, используя иные классы функций. Например, можно предположить, что существует периодическая зависимость значений Y от X, тогда надо добавить в модель гармонические составляющие вида $\beta_1 cos(\omega x) + \beta_2 sin(\omega x)$. Частоты этих составляющих ω придется подбирать отдельно. Можно предположить, анализируя зависимость, что у нас присутствует периодическая составляющая с частотой порядка 1 (одно колебание за весь интервал изменений X). Построим модель вида

$$y = \beta_0 + \beta_1 x + \beta_2 cos(2\pi x/10) + \beta_3 sin(2\pi x/10).$$

Получим оценки для этой модели:

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,934315451
R–квадрат	0,872945362
Нормированный R–квадрат	0,849122618
Стандартная ошибка	1,279008211
Наблюдения	20

	Коэффициенты	*Стандартная ошибка*	*t–статистика*	*P–Значение*

Y–пересечение	5,080728355	0,88616922	5,733361352	3,08053E–05
Переменная X 1	0,308759559	0,159762044	1,932621493	0,0711836
Переменная X 2	–1,299656278	0,412270758	–3,152433814	0,006163641
Переменная X 3	–3,032825609	0,646493647	–4,691191664	0,000245237

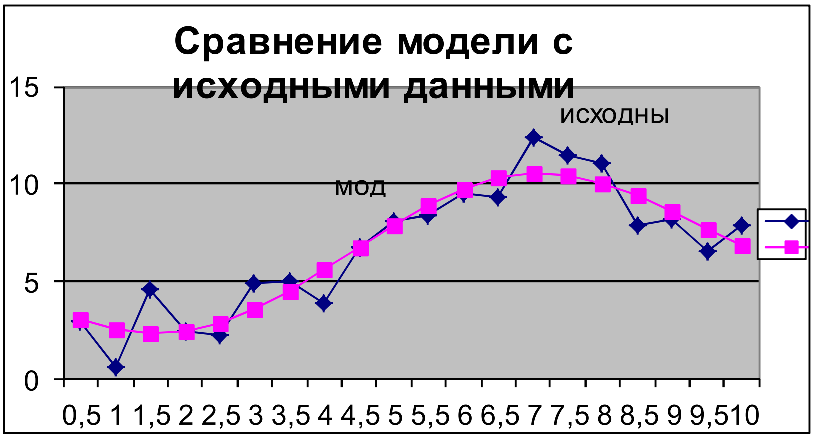

Рис. 33. Сравнение модели $y = \beta_0 + \beta_1 x + \beta_2 cos(2\pi x/10) + \beta_3 sin(2\pi x/10)$ *с исходными данными*

Заметим, что мы получили значение стандартной ошибки меньше заданой величины c. Это говорит о том, что дальше улучшать модель бессмысленно. Если мы продолжим, то будем по сути аппроксимировать случайные ошибки, а не реальную существующую зависимость Y от X. Качество модели выше, чем у модели вида $y = \beta_0 + \beta_2 x^2 + \beta_3 x^3$. Однако, данная модель содержит на один коэффициент больше, и кроме того содержит параметр, значение которого по сути определяется вручную. Поскольку модели принципиально различные, то предпочесть следует ту, которая более соответствует физике явления (если она известна). Мы остановимся на последней модели $y = \beta_0 + \beta_1 x + \beta_2 Cos(2\pi x/10) + \beta_3 Cos(2\pi x/10)$.

Найдем доверительный интервал для σ^2 соответствующий уровню 0,95. Находим квантили распределения хи–квадрат уровней 0,025 и 0,975 соответственно для числа степеней свободы $\nu = n - k - 1 = 20 - 3 - 1 = 16$ (k –число коэффициентов модели, не считая β_0): $\tau_{0,025} = 6,91$, $\tau_{0,975} = 28,85$. Тогда $\frac{(n-k-1)s^2}{\tau_{0,975}} < \sigma^2 < \frac{(n-k-1)s^2}{\tau_{0,025}} \Rightarrow$ $0,91 < \sigma^2 < 3,79 \Rightarrow 0,95 < \sigma < 1,95$.

Доверительные интервалы для коэффициентов уравнения регрессии можно найти в Итоговой статистике.

Доверительный интервал для значений y для тех же значений X, что приведены в выборке, рассчитываем по формуле $y_j = \left[\left(X^T B^* \right)_j \pm t_\alpha s \left(1 + \sqrt{\left(X(X^T X)^{-1} X^T \right)_{jj}} \right) \right]$. Доверительный интервал для значений y, полученный по этим формулам, отображен на рис 34, где t_α квантиль распределение Стьюдента с $n-k-1$ степенью свободы (доверительный уровень возьмем 0,67, тогда $\alpha = 0{,}33$ и $t_\alpha = 1{,}0047$)

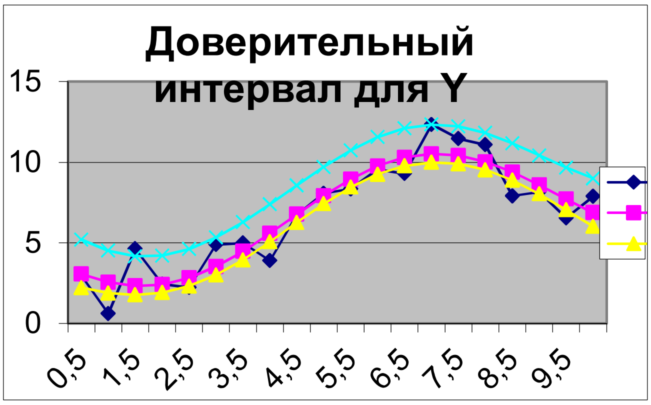

Рис. 34. Доверительный интервал для Y

Тема 5. Анализ риска банкротства предприятия на основе нечетких множеств

Введение

В условиях рыночных отношений всегда важно иметь представление о финансовом состоянии организаций, с которыми ведется ежедневный бизнес, и знать, как собственное предприятие выглядит в их глазах. Зачастую весьма полезно обладание информацией о том, насколько привлекательным предприятие предстает с точки зрения потенциального делового партнера, либо кредитора, в случаях, когда возникает необходимость привлечения дополнительных средств для развития, либо поддержания бизнеса. Особенно важно проведение такого анализа в условиях напряженной экономической обстановки, либо в условиях кризиса, когда конкуренция растет, а цена ошибки становится все более весома.

5.1. Комплексный коэффициентный метод финансового анализа. Степень риска банкротства

В настоящее время в мировой учетно–аналитической практике известны десятки показателей, используемых для оценки имущественного и финансового состояния компаний. Классифицируя эти показатели, выделяют обычно шесть групп [1,2], описывающих:

- имущественное положение компании;
- ликвидность;
- финансовую устойчивость;
- деловую активность;
- рентабельность;
- положение на рынке ценных бумаг.

Наиболее распространенные показатели **имущественного положения** компании следующие:

- сумма хозяйственных средств, находящихся в собственности и распоряжении компании;
- доля активной части основных средств;
- коэффициент износа;
- коэффициент обновления;
- коэффициент выбытия.

Показатели **ликвидности и платежеспособности** компании следующие:

- величина собственных оборотных средств;
- коэффициент текущей ликвидности;
- коэффициент быстрой ликвидности;
- коэффициент обеспеченности текущей деятельности собственными оборотными средствами;
- коэффициент покрытия запасов.

Показатели **финансовой устойчивости** компании:

- коэффициент финансовой автономии;
- коэффициент маневренности собственного капитала;
- коэффициент структуры долгосрочных источников финансирования;
- коэффициент структуры привлеченных средств;
- коэффициент структуры заемных средств;
- коэффициент обеспеченности процентов к уплате;
- коэффициент покрытия постоянных финансовых расходов.

Показатели **деловой активности** компании:

- коэффициент устойчивости экономического роста;
- коэффициент фондоотдачи;
- коэффициент оборачиваемости средств в активах.

Показатели **рентабельности** компании следующие:

- рентабельность совокупного капитала;

- рентабельность собственного капитала;
- рентабельность инвестиций;
- валовая рентабельность реализованной продукции (валовая маржа).

Показатели **положения компании на рынке ценных бумаг**:

- доход на акцию;
- ценность акции (price–to–earnings ratio);
- дивидендная доходность акции;
- коэффициент котировки акции (price–to–book ratio).

По ряду показателей известны некие нормативы, характеризующие их значение положительно или отрицательно. Например, когда собственные средства предприятия превышают половину всех пассивов, соответствующий этой пропорции коэффициент автономии больше 0.5, и это его значение считается "хорошим" (соответственно, когда оно меньше 0.5 – "плохим"). Но в большинстве случаев показатели, оцениваемые при анализе, однозначно нормировать невозможно. Это связано со спецификой отраслей экономики, с текущими особенностями действующих предприятий, с состоянием экономической среды, в которой они работают.

Тем не менее, любое заинтересованное положением предприятия лицо (руководитель, инвестор, кредитор, аудитор и т.д.), далее именуемое лицом, принимающим решения (ЛПР), не довольствуется простой количественной оценкой показателей. Для ЛПР важно знать, приемлемы ли полученные значения, хороши ли они, и в какой степени. Кроме того, ЛПР стремится установить логическую связь количественных значений показателей выделенной группы с риском банкротства. То есть ЛПР не может быть удовлетворено бинарной оценкой "хорошо – плохо", его интересуют оттенки ситуации и экономическая интерпретация этих оттеночных значений. Задача осложняется тем, что показателей много, изменяются они зачастую разнонаправленно, и поэтому ЛПР стремится "свернуть" набор всех исследуемых частных финансовых показателей в один комплексный, по значению которого и судить о степени благополучия ("живучести") фирмы и о том, насколько далеко или близко предприятие отстоит от банкротства.

Таким комплексным коэффициентам, характеризующим положение хозяйствующего субъекта в целом, является степень риска банкротства.

Степень риска банкротства – это комплексный показатель, характеризующий как финансовое положение предприятия, так и качество управления им, которое, в конечном счете, получает свое выражение в финансовом эквиваленте, но не исчерпывается одними лишь финансовыми последствиями [3–5].

5.2. Анализ риска банкротства предприятия с использованием теории нечетких множеств

Задача определения степени риска банкротства является актуальной как для собственников предприятия, так и для его кредиторов.

Излагаемый далее подход к анализу риска банкротства позволяет учитывать количественные (финансовые) и качественные (индикаторные) показатели в анализе, позволяет анализировать риск банкротства, настраиваясь не только на страну, период времени, отрасль, но и на само предприятие, на его экономическую и управленческую специфику. Предлагается своего рода конструктор, который может быть использован (собран) любым экспертом по своему усмотрению.

Метод носит матричный характер, где по строкам матрицы располагаются отдельные количественные показатели, характеризующие различные стороны финансовой деятельности компании, а по столбцам располагаются качественные уровни данных показателей, выраженные на естественном языке (например, «низкий», «средний», «высокий»). На пресечении столбцов и строк располагается степень текущего количественного уровня фактора качественному подмножеству, измеренная определенным образом. Тогда результирующий финансовый показатель получается как двойная свертка компонент построенной матрицы с предопределенными весами.

Теория нечетких множеств изложена в работах [6,7] и приводится в приложении А.

5.2.1. Нечетко–множественная модель финансового состояния компании

В предлагаемой модели компания описывается набором количественных и качественных факторов финансового анализа общим числом N. При этом все факторы являются измеримыми, т.е. имеют носитель со своей областью определения на вещественной оси.

Нечеткие описания в структуре метода анализа риска появляются в связи с **неуверенностью** эксперта, которая возникает в ходе классификации уровня факторов. Например, если эксперт не может четко разграничить понятия «высокой» и «максимальной» вероятности. Или когда надо провести границу между средним и низким уровнем значения параметра. Нечеткое описание параметра включает в себя следующие этапы:

1. Эксперт фиксирует показатель (фактор) и его количественный носитель.

2. На выбранном носителе эксперт строит лингвистическую переменную со своим терм–множеством значений. Например: переменная **«Уровень показателя X»** может обладать терм –множеством значений «*Очень низкий, Низкий, Средний, Высокий, Очень высокий*».

3. Чтобы конструктивно описать лингвистическую переменную, эксперт выбирает соответствующий ей *количественный признак* – например, сконструированный специальным образом показатель уровня X, который принимает значения от нуля до единицы.

4. Далее эксперт каждому значению лингвистической переменной (которое, по своему построению, является **нечетким подмножеством** значений интервала (0,1) – области значений показателя уровня X) сопоставляет **функцию принадлежности** уровня X тому или иному нечеткому подмножеству. Общеупотребительными функциями в этом случае являются **трапециевидные** функции принадлежности (см. рис. 35). Верхнее основание трапеции соответствует полной уверенности эксперта в правильности своей классификации, а нижнее – уверенности в том, что никакие другие значения интервала (0,1) не попадают в выбранное нечеткое подмножество.

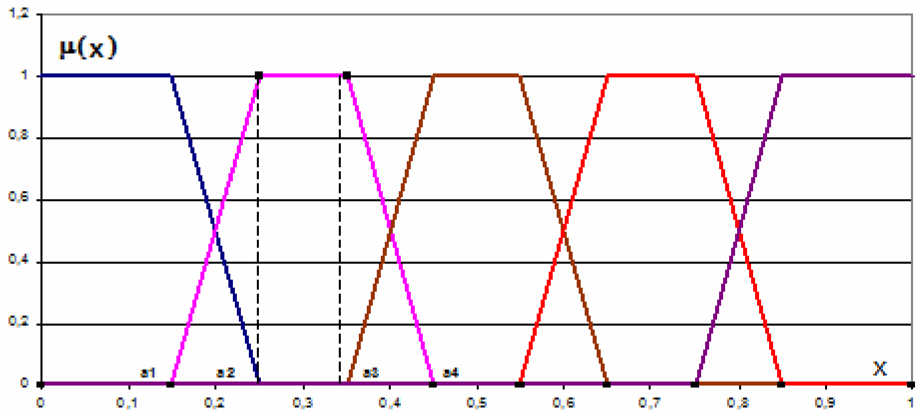

Рис. 35 Качественный вид функций принадлежности

Для целей компактного описания трапециевидные функции принадлежности $\mu(x)$ удобно описывать трапециевидными числами вида

$$\beta(a_1, a_2, a_3, a_4), \qquad (1)$$

где a_1 и a_4 – абсциссы нижнего основания, а a_2 и a_3 – абсциссы верхнего основания трапеции (рис. 1), задающей $\mu(x)$ в области с ненулевой принадлежностью **носителя** x соответствующему нечеткому подмножеству.

Таким образом, построен классификатор параметра по качественному уровню. И аналитик может употреблять его как математический объект в соответствующих операциях и методах.

5.3. Описание метода решения

Продемонстрируем, как применяется полученный классификатор при поэтапном конструировании нечетко–множественной модели компании.

Этап 1 (Лингвистические переменные и нечеткие подмножества).

Пусть

А) Лингвистическая переменная **E «Состояние предприятия»** имеет пять значений:

$\mathbf{E_1}$ – нечеткое подмножество состояний "предельного неблагополучия";

$\mathbf{E_2}$ – нечеткое подмножество состояний "неблагополучия";

$\mathbf{E_3}$ – нечеткое подмножество состояний "среднего качества";

$\mathbf{E_4}$ – нечеткое подмножество состояний "относительного благополучия";

$\mathbf{E_5}$ – нечеткое подмножество состояний "предельного благополучия".

Б) Соответствующая переменной **E** лингвистическая переменная **G** «**Риск банкротства»** также имеет пять значений:

$\mathbf{G_1}$ – нечеткое подмножество "запредельная степень риска банкротства",

$\mathbf{G_2}$ – нечеткое подмножество "степень риска банкротства высокая (опасная)**",**

$\mathbf{G_3}$ – нечеткое подмножество " степень риска банкротства средняя (пограничная)",

$\mathbf{G_4}$ – нечеткое подмножество " низкая степень риска банкротства (приемлемая) ",

$\mathbf{G_5}$ – нечеткое подмножество "риск банкротства незначителен".

Носитель множества G – показатель степени риска банкротства g – принимает значения от нуля до единицы по определению.

В) Для произвольного отдельного финансового или управленческого показателя $\mathbf{X_i}$ задаем лингвистическую переменную $\mathbf{B_i}$ **«Уровень показателя $\mathbf{X_i}$»** на нижеследующем терм–множестве значений:

$\mathbf{B_{i1}}$ – подмножество "очень низкий уровень показателя X_i",

$\mathbf{B_{i2}}$ – подмножество "низкий уровень показателя X_i**",**

$\mathbf{B_{i3}}$ – подмножество "средний уровень показателя X_i",

$\mathbf{B_{i4}}$ – подмножество "высокий уровень показателя X_i",

$\mathbf{B_{i5}}$ – подмножество "очень высокий уровень показателя X_i".

Причем здесь и далее по умолчанию предполагаем:

1. Рост отдельного показателя X_i сопряжен со снижением степени риска банкротства, с улучшением самочувствия рассматриваемого предприятия. Если для данного показателя наблюдается противоположная тенденция, то в анализе его следует заменить сопряженным. Например, показатель доли заемных средств в активах предприятия разумно заменить показателем доли собственных средств в активах.

2. Выполняется дополнительное условие соответствия множеств **B, E** и **G** следующего вида: если все показатели в ходе анализа обладают, в соответствии с классификацией, уровнем подмножества $\mathbf{B_{ij}}$, то состояние предприятия квалифицируется как $\mathbf{E_i}$, а степень риска банкротства – как $\mathbf{G_j}$**.** Выполнение этого условие влияет, с одной стороны, на правильную количественную классификацию уровней показателей (см. далее этап 5 метода) и на правильное определение уровня значимости показателя в системе оценки (см. далее этап 3 метода).

Этап 2 (Показатели). Построим набор отдельных показателей $\mathbf{X}=\{X_i\}$ общим числом N, которые, по мнению эксперта–аналитика, с одной стороны, влияют на оценку риска банкротства предприятия, а, с другой стороны, оценивают различные по природе стороны деловой и финансовой жизни предприятия.

Этап 3 (Значимость). Сопоставим каждому показателю X_i уровень его **значимости** для анализа r_i. Чтобы оценить этот уровень, нужно расположить все показатели по порядку убывания значимости так, чтобы выполнялось правило

$$r_1 \geq r_2 \geq ...r_N . \qquad (2)$$

Если система показателей проранжирована в порядке убывания их значимости, то значимость i–го показателя r_i следует определять по правилу Фишберна [5]:

$$r_i = \frac{2(N - i + 1)}{(N+1)N} . \qquad (3)$$

Правило Фишберна отражает тот факт, что об уровне значимости показателей неизвестно ничего кроме (2). Тогда оценка (3) отвечает максимуму энтропии имеющейся информационной неопределенности об объекте исследования.

Если же все показатели обладают равной значимостью (равнопредпочтительны или системы предпочтений нет), тогда

$$r_i = 1/N. \qquad (4)$$

Этап 4 (Классификация степени риска). Построим **классификацию** текущего значения g показателя степени риска как критерий разбиения этого множества на нечеткие подмножества (таблица 5.1):

Таблица 5.1.

Классификация степени риска банкротства

Интервал значений g	Классификация уровня параметра	Степень оценочной уверенности (функция принадлежности)
$0 \le g \le 0.15$	G_5	1
$0.15 < g < 0.25$	G_5	$\mu_5 = 10 \times (0.25 - g)$
	G_4	$1 - \mu_5 = \mu_4$
$0.25 \le g \le 0.35$	G_4	1
$0.35 < g < 0.45$	G_4	$\mu_4 = 10 \times (0.45 - g)$
	G_3	$1 - \mu_4 = \mu_3$
$0.45 \le g \le 0.55$	G_3	1
$0.55 < g < 0.65$	G_3	$\mu_3 = 10 \times (0.65 - g)$
	G_2	$1 - \mu_3 = \mu_2$
$0.65 \le g \le 0.75$	G_2	1
$0.75 < g < 0.85$	G_2	$\mu_2 = 10 \times (0.85 - g)$
	G_1	$1 - \mu_2 = \mu_1$
$0.85 \le g \le 1.0$	G_1	1

Этап 5 (Классификация значений показателей). Построим классификацию текущих значений **x** показателей **X** как критерий разбиения полного множества их значений на нечеткие подмножества вида **B.**

Таблица 5.2.

Классификация значений показателей

Наименование показателя	Критерий разбиения по подмножествам				
	B_{i1}	B_{i2}	B_{i3}	B_{i4}	B_{i5}
X_1	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}
…	…	…	…	…	…
X_i	β_{i1}	β_{i2}	β_{i3}	β_{i4}	β_{i5}
…	…	…	…	…	…
X_N	β_{N1}	β_{N2}	β_{N3}	β_{N4}	β_{N5}

Этап 6 (Оценка уровня показателей). Произведем оценку текущего уровня показателей и сведем полученные результаты в таблицу 5.3.

Таблица 5.3.

Оценка уровня показателей

Наименование показателя	Текущее значение

X_1	x_1
…	…
X_i	x_i
…	…
X_N	x_N

Этап 7 (Классификация уровня показателей). Проведем классификацию текущих значений **x** по критерию. Результатом проведенной классификации является таблица 5.4, где λ_{ij} – уровень принадлежности носителя x_i нечеткому подмножеству $\mathbf{B_{ij}}$.

Таблица 5.4.

Классификация уровня показателей

Наименование показателя	**Результат классификации по подмножествам**				
	B_{i1}	**B_{i2}**	**B_{i3}**	**B_{i4}**	**B_{i5}**
X_1	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}
…	…	…	…	…	…
X_i	λ_{i1}	λ_{i2}	λ_{i3}	λ_{i4}	λ_{i5}
…	…	…	…	…	…
X_N	λ_{N1}	λ_{N2}	λ_{N3}	λ_{N4}	λ_{N5}

Этап 8. Вычисление комплексного показателя финансового состояния компании

Комплексный показатель степени риска банкротства g вычисляется путем двойной свертки данных таблицы 5.4:

$$g = \sum_{j=1}^{5} g_j \sum_{i=1}^{N} r_i \lambda_{ij}, \quad (5)$$

где

$$g_j = 0.9 - 0.2(j-1), \quad (6)$$

λ_{ij} определяется по таблице 5.4, а r_i – по формуле (3) или (4).

Существо формул (5) и (6) состоит в следующем. Внутреннее суммирование в (5) производится по значимостям показателя, а внешнее суммирование – по узловым точкам пятиуровневого классификатора степени риска. Таким образом, результирующая оценка риска определяется как средневзвешенное по всем участвующим в оценке показателям, с одной стороны, и по всем качественным уровням этих показателей, с другой стороны.

Распознаем полученное значение степени риска на базе классификатора таблицы 5.1. Результатом классификации является лингвистическое описание степени риска банкротства компании и степень уверенности эксперта в таком результате распознавания.

5.4. Пример оценки риска банкротства компании

5.4.1. Постановка задачи

Рассмотрим алгоритм использования нечетко–множественной методики оценки банкротства Недосекина [5] на примере ОАО «СИСТЕМА» [8] за период 2009 – 2011 г.г..

Его практическая реализация предусматривает выполнение следующих шагов:

1. Идентификация лингвистических переменных «Состояние предприятия» E_j (например, «предельное неблагополучие», «неблагополучие», «среднего качества», «относительное благополучие» и «предельное благополучие»), «Риск банкротства» G_j (например, «запредельный», «высокий», «пограничный», «приемлемый» и «незначительный»), а также установление однозначного соответствия между введенными переменными. Элементами носителя (множества G) является показатель степени банкротства g, принимающий значения от 0 до 1.

2. Выбор финансовых показателей X_i, которые наилучшим образом характеризуют отдельные аспекты деятельности предприятия и влияют на оценку риска банкротства (см. табл. 5.5).

Таблица 5.5.

Характеристика финансовых показателей

Шифр показателя X_i	**Наименование показателя X_i**	**Описание**
X_1	Коэффициент автономии	Собственный капитал / Валюта баланса
X_2	Коэффициент промежуточной ликвидности	[Краткосрочная дебиторская задолженность + Краткосрочные финансовые вложения +Денежные средства] / Краткосрочные обязательства
X_3	Коэффициент абсолютной ликвидности	[Денежные средства + Краткосрочные финансовые вложения] / Краткосрочные обязательства
X_4	Оборачиваемость активов	Выручка за период / Средняя стоимость активов
X_5	Рентабельность основной деятельности	Прибыль от реализации / Затраты на производство продукции
X_6	Рентабельность активов	Чистая прибыль / Средняя балансовая стоимость активов

- X_1 – коэффициент автономии (коэффициент концентрации собственных средств, который показывает долю собственных средств в стоимости имущества предприятия),
- X_2 – коэффициент промежуточной ликвидности (отражает способность компании погашать свои текущие обязательства в случае возникновения сложностей с реализацией продукции),
- X_3 – коэффициент абсолютной ликвидности (характеризует способность компании погашать текущие обязательства за счёт денежных средств, средств на расчетный счетах и краткосрочных финансовых вложений; наиболее важен для поставщиков товарно–материальных ресурсов и для банков, кредитующих предприятие),
- X_4 – оборачиваемость активов (отражает скорость оборота денежных средств вложенных в имущество),
- X_5 – рентабельность основной деятельности (определяет, сколько чистой прибыли получено с 1 рубля затрат на производство),
- X_6 – рентабельность активов (показывает сколько предприятие имеет чистой прибыли с рубля вложенного в капитал).

5.4.2. Оценка финансового состояния ОАО «СИСТЕМА»

Финансовое состояние предприятия ОАО «СИСТЕМА» характеризуется следующими финансовыми показателями (табл. 5.6)

Таблица 5.6.

Значения финансовых показателей предприятия ОАО «СИСТЕМА» [8]

Шифр показателя X_i	**Наименование показателя X_i**	**Значение X_i в период I (2008 г) ($x_{I,i}$)**	**Значение X_i в период II (2009 г) ($x_{II,i}$)**	**Значение X_i в период III (2010 г) ($x_{III,i}$)**
X_1	Коэффициент автономии	0.92	0,877	0,876
X_2	Коэффициент промежуточной ликвидности	5,013	2,685	3,911
X_3	Коэффициент абсолютной ликвидности	1,406	1,09	1,045
X_4	Оборачиваемость активов	0,134	0,125	0,137
X_5	Рентабельность основной деятельности	0,081	0,229	0,359
X_6	Рентабельность активов	0,09	–0,088	0,072

Затем для каждого показателя вводится лингвистическая переменная «Уровень показателя X_i», например, подмножество «очень низкий», «низкий», «средний», «высокий», «очень высокий».

Определение системы весов показателей.

Для характеристики показателей воспользуемся экспертным методом.

В качестве членов экспертной группы было привлечено 8 специалистов одного из филиалов ОАО "СИСТЕМА".

На основе полученных данных о значимости показателей (табл. 5.7) в результате экспертного опроса была дана оценка согласованности мнений экспертов с помощью коэффициента конкордации (W).

Таблица 5.7.

Оценки экспертов

	Э$_1$	**Э$_2$**	**Э$_3$**	**Э$_4$**	**Э$_5$**	**Э$_6$**	**Э$_7$**	**Э$_8$**
X_1	6	6	6	6	5	6	6	5
X_2	2	2	2	2	2	1	2	1
X_3	1	1	1	1	1	2	1	2
X_4	5	5	5	5	6	5	5	6
X_5	4	4	4	3	3	4	4	4
X_6	3	3	3	4	4	3	3	3

В общем виде формула расчета коэффициента конкордации имеет следующий вид:

$$W = \frac{12 \sum_{j=1}^{m} d_j^2}{n^2(m^3 - m)},$$

где n – количество экспертов ($n = 8$),

m – количество параметров (оцениваемых объектов ($m = 6$),

d_j^2 – отклонение суммы рангов по j–ому параметру от среднего значения рангов:

$$d_j^2 = \left(\sum_{i=1}^{n} R_{ji} - \frac{n(m+1)}{2} \right)^2,$$

где $R_{ji} \in \{1, 2, ..., m\}$ – ранг j–го параметра i – го эксперта ($i = 1, 2, ..., n$).

Результат обработки экспертных оценок значимости приведен в табл. 5.8.

Таблица 5.8.

Оценка согласованности мнений экспертов

Коэффициент	Сумма рангов, R_{ji}	d_j^2	$1/R_{ji}$	Оценка важности R_{ji}/n
X_1	46	324	0,022	6
X_2	14	196	0,071	2
X_3	10	324	0,100	1
X_4	42	196	0,024	5
X_5	30	4	0,033	4
X_6	26	4	0,038	3
Итого	168	1048	0,289	
W=0.94				

Таким образом, следует отметить, что степень согласованности мнений экспертов составляет 0,94 (что больше критического значения 0,75), что является достаточной согласованностью.

Установление соответствия между значениями показателя степени риска (g) и нечеткими подмножествами множества G (см. табл. 5.9).

Таблица 5.9.

Классификация степени риска банкротства [3,4]

Интервал значений g	**Классификация уровня параметра**	**Функция принадлежности**
$0 \leq g \leq 0.15$	риск банкротства незначителен	1
$0.15 < g < 0.25$	риск банкротства незначителен	$\mu_5 = 10 \times (0.25 - g)$
	приемлемый риск банкротства	$1 - \mu_5 = \mu_4$
$0.25 \leq g \leq 0.35$	приемлемый риск банкротства	1
$0.35 < g < 0.45$	приемлемый риск банкротства	$\mu_4 = 10 \times (0.45 - g)$
	риск банкротства пограничный	$1 - \mu_4 = \mu_3$
$0.45 \leq g \leq 0.55$	риск банкротства пограничный	1
$0.55 < g < 0.65$	риск банкротства пограничный	$\mu_3 = 10 \times (0.65 - g)$
	риск банкротства высокий	$1 - \mu_3 = \mu_2$
$0.65 \leq g \leq 0.75$	риск банкротства высокий	1
$0.75 < g < 0.85$	риск банкротства высокий	$\mu_2 = 10 \times (0.85 - g)$
	запредельный риск банкротства	$1 - \mu_2 = \mu_1$
$0.85 \leq g \leq 1.0$	запредельный риск банкротства	1

Оценка границ интервалов значений показателей. Построение функций принадлежности.

Пусть экспертному сообществу в качестве исходной информации предлагается интервал по каждому показателю. Необходимо в рамках этого интервала выполнить внутреннее разбиение на 5 уровней (например, «очень низкий», «низкий», «средний», «высокий», «очень высокий»).

Подробно рассмотрим определение граничных значений для показателя X_5 – рентабельность деятельности, значение которого изменяется в интервале [–1;1]. Данные, полученные в результате экспертного опроса о показателе рентабельности деятельности представлены в табл. 5.10. Данные экспертного опроса по другим показателям приведены в приложении.

Таблица 5.10.

Данные опроса экспертов по показателю X_5

	$Э_1$	$Э_2$	$Э_3$	$Э_4$	$Э_5$	$Э_6$	$Э_7$	$Э_8$
«очень низкий»	–1; –0,06	–1; –0,08	–1; –0,03	–1; –0,03	–1; –0,02	–1; –0,02	–1; –0,01	–1; –0,02
«низкий»	–0,06; 0,04	–0,08; 0,01	–0,03; 0,05	–0,03; 0,04	–0,02; 0,07	–0,02; 0,07	–0,01; 0,08	–0,02; 0,08
«средний»	0,04; 0,1	0,01; 0,1	0,05; 0,15	0,04; 0,14	0,07; 0,17	0,07; 0,15	0,08; 0,18	0,08; 0,18
«высокий»	0,1; 0,24	0,1; 0,3	0,15; 0,4	0,14; 0,4	0,17; 0,4	0,15; 0,45	0,18; 0,45	0,18; 0,48
«очень высокий»	0,24;1	0,3;1	0,4;1	0,4;1	0,4;1	0,45;1	0,45;1	0,48;1

Результатом опроса является 8 интервалов вещественной оси $[a_i;b_i],\ i=1,..,8$.

Определим $A=\min_i\{a_i\}$, $B=\min\{\max_i(a_i),\min_i(b_i)\}$ $C=\max\{\max_i(a_i),\min_i(b_i)\}$, $D=\max_i\{b_i\}$.

Тогда четыре пары чисел – *(A;0), (B;1), (C;1), (D;0)* – являются множеством вершин трапециевидной функции принадлежности. Например, для значения «очень низкий» получим точки: *(–1;0), (–1;1), (–*

0,01;1), (–0,01;0); для значения *«низкий» – (–0,08;0),(–0,01;1); (0,01;1); (0,08;1).*

В результате мы получим опорные точки трапециевидных функций принадлежности, которые представлены в табл. 5.11.

Таблица 5.11.

Опорные точки трапециевидных функций принадлежности для показателя X_5

Опорные точки	**Нечеткое подмножество**				
	«очень низкий»	**«низкий»**	**«средний»**	**«высокий»**	**«очень высокий»**
a_1	–1	–0,08	0,01	0,1	0,24
a_2	–1	–0,01	0,08	0,18	0,48
a_3	–0,08	0,01	0,1	0,24	1
a_4	–0,01	0,08	0,18	0,48	1

Для всех остальных показателей оценка границ интервалов значений и построение функции принадлежности происходит аналогичным образом.

Таким образом, выбранные показатели на основании предварительного экспертного анализа получили следующую классификацию (табл. 5.12):

Таблица 5.12.

Результаты классификации параметров $X_1 – X_6$

Шифр показателя	**Т–числа {γ} для значений лингвистической переменной "Величина параметра"**				
	"очень низкий"	**"низкий"**	**"средний"**	**"высокий"**	**"очень высокий"**
X_1	(0; 0; 0,1; 0,2)	(0,1; 0,2; 0,3; 0,4)	(0,3; 0,4; 0,45; 0,5)	(0,45; 0,5; 0,6; 0,8)	(0,6; 0,8; 1; 1)
X_2	(0; 0; 0,2; 0,4)	(0,2; 0,4; 0,6; 0,7)	(0,6; 0,7; 0,8; 1)	(0,8; 1; 1,5; 1,8)	(1,5; 1,8; ∞; ∞)
X_3	(0; 0; 0,05; 0,08)	(0,05; 0,08; 0,1; 0,2)	(0,1; 0,2; 0,3; 0,4)	(0,3; 0,4; 0,5; 1)	(0,5; 1; ∞; ∞)
X_4	(0; 0; 0,12; 0,15)	(0,12; 0,15; 0,2; 0,25)	(0,2; 0,25; 0,35; 0,45)	(0,35; 0,45; 0,6; 1)	(0,6; 1; ∞; ∞)
X_5	(–1; –1; –0,08; –0,01)	(–0,08; –0,01; 0,01; 0,08)	(0,01; 0,08; 0,1; 0,18)	(0,1; 0,18; 0,24; 0,48)	(0,24; 0,48; 1; 1)
X_6	(–1; –1; 0; 0)	(0; 0; 0,006; 0,015)	(0,006; 0,015; 0,06; 0,1)	(0,06; 0,1; 0,24; 0,4)	(0,24; 0,4; ∞; ∞)

В табл. 5.12 приведена классификация показателей с использованием трапециевидных чисел вида (a_1, a_2, a_3, a_4), где a_1 и a_4 – абсциссы нижнего основания, а a_2 и a_3 – абсциссы верхнего основания трапеции. Верхнее основание трапеции соответствует уверенности эксперта в правильности своей классификации, а ее ребра задают интервал неуверенности.

Распознавание уровня показателей.

Если значение фактора точно попадают в рассматриваемый интервал, то значение равно единице для данного качественного уровня и нулю для всех остальных уровней. Если значение лежит в зоне неуверенности, то для двух смежных классов формируются значения, сумма которых равна единице. Вычисление значений идет по правилу вычисления ординаты наклонного ребра трапециевидной функции принадлежности по заданной абсциссе точки на нижнем основании трапеции.

На основе табл. 5.12 построим правила распознавания уровней показателей $X_1 – X_6$.

Итак, проведем вычисления на примере X_6 – показателя рентабельности активов за 2010 год. Значение показателя равно 0,072. Воспользуемся табл. 8 «Результаты классификации параметров $X_1 – X_6$» для определения степени оценочной уверенности принадлежности к этому уровню.

Графическая интерпретация представлена на рис. 36, где отображен переход от уровня «средний» к уровню «высокий».

Рис. 36. Переход от уровня средний» к уровню «высокий»

Так как фактическое значение, равное 0,072 попадает в переходные интервал, то принадлежность уровню и степени оценочной уверенности этой принадлежности можно определить из уравнения:

$$СОУ^{средний} = \frac{-x+0,1}{0,04} = \frac{-0.072+0,1}{0,04} = 0,7$$

$$СОУ^{высокий} = \frac{x-0,06}{0,04} = \frac{0.072-6}{0,04} = 0,3$$

где $СОУ^{средний}$ – степень оценочной уверенности принадлежности к «среднему» уровню на основании линейной функции принадлежности;

$СОУ^{высокий}$ – степень оценочной уверенности принадлежности к «высокому» уровню на основании линейной функции принадлежности

Таким образом, $СОУ^{средний} = 0,7$; $СОУ^{высокий} = 0,3$.

Аналогичным образом происходит распознавание уровней остальных показателей по критерию табл. 5.12. Результатом проведенной классификации является табл. 5.13:

Таблица 5.13.

Классификация уровня показателей X_1– X_6 за 2008–2010гг.

год	**Показатель X_i**	**Значение $\{\lambda_{ij}\}$**				
		λ_{i1}	λ_{i2}	λ_{i3}	λ_{i4}	λ_{i5}
2008	X_1	0	0	0	0	1
	X_2	0	0	0	0	1
	X_3	0	0	0	0	1
	X_4	0,533	0,467	0	0	0
	X_5	0	0	1	0	0
	X_6	0	0	0,250	0,750	0
2009	X_1	0	0	0	0	1
	X_2	0	0	0	0	1
	X_3	0	0	0	0	1
	X_4	0,833	0,167	0	0	0
	X_5	0	0	0	1	0
	X_6	1	0	0	0	0
2010	X_1	0	0	0	0	1
	X_2	0	0	0	0	1

	X_3	0	0	0	0	1
	X_4	0,433	0,567	0	0	0
	X_5	0	0	0	0,504	0,496
	X_6	0	0	0,7	0,3	0

Анализ указывает на то, что количественное изменение коэффициента автономии, коэффициента промежуточной ликвидности и коэффициента абсолютной ликвидности не сопровождаются качественным ростом. Однако во втором и третьем периодах наблюдается качественный рост рентабельности деятельности. Также во втором и третьем периодах наблюдается разнонаправленное качественное изменение оборачиваемости активов и рентабельности активов.

Анализ результатов. Определение степени риска банкротства.

В таблице 5.14 приведены величины $v_j = \sum_{i=1}^{N} r_i \lambda_{ij}, \ j = 1,...,5$ для трех периодов. Здесь $r_i = 1/N, \ N = 6$ – веса показателей.

Таблица 5.14.

Усредненные значения терм–множества {□}

Год	Усредненные значения {□}				
	"очень низкий"	**"низкий"**	**"средний"**	**"высокий"**	**"очень высокий"**
2008	0,089	0,078	0,208	0,125	0,5
2009	0,305	0,028	0	0,167	0,5
2010	0,072	0,095	0,117	0,134	0,583

В соответствии с результатом распознавания количественное значение степени риска банкротства предприятия g определяется по формуле двойной свертки Недосекина [5] (5):

$$g = \sum_{j=1}^{5} g_j v_j = \sum_{j=1}^{5} g_j \sum_{i=1}^{N} r_i \lambda_{ij}, \quad \text{где } g_j = 0.9 - 0.2(j-1)$$

С учетом вышеизложенного, оценка степени риска банкротства по формуле (5) дает $g_I = 0,326$, $g_{II} = 0,394$, $g_{III} = 0,288$, т.е. наблюдается сокращение риска банкротства.

Лингвистическое распознавание значений g по данным для первого периода определяет с абсолютной степенью уверенности (функция принадлежности равна 1) степень риска банкротства предприятия ОАО «СИСТЕМА» как приемлемая. Степень риска банкротства предприятия во втором периоде распознается примерно с одинаковой степенью соответствия как приемлемая (0,56) и как пограничная (0,44). В третьем периоде степень риска банкротства предприятия распознается уверенно как приемлемая (функция принадлежности равна 1).

СОДЕРЖАНИЕ ОТЧЕТА ПО ПРАКТИЧЕСКИМ РАБОТАМ

1. Титульный лист (Приложение А).
2. Название и цель лабораторной работы.
3. Вариант задания.
4. Детальное описание выполненных действий.
5. Обязательное представление скриншотов из программы, выполняющей анализ.
6. Выводы по работе.

СПИСОК ЛИТЕРАТУРЫ

1. Силич В.А., Силич И.П. Теория систем и системный анализ: Учебное пособие. – Томск: Томский политехнический университет, 2010. – 281 с.

2. Маслов А.В. Теория систем и системный анализ / А.В. Маслов. – Юрга: Изд–во ЮТИ ТПУ, 2011. – 238 с.

3. Черников Ю.Г. Системный анализ и исследование операций: Учебное пособие для вузов. – М.: Издательство Московского государственного горного университета, 2006. – 370 с. – [Электронный ресурс] – Режим доступа: http://ezproxy.ha.tpu.ru:2117/view/book/3512/.

4. Анфилатов В.С. и др. Системный анализ в управлении: Учебное пособие / В.С. Анфилатов, А.А. Емельянов, А.А. Кукушкин; Под ред. А.А. Емельянова. – М.: Финансы и статистика, 2009. – 368 с.

5. Баканов М.И., Шеремет А.Д. Теория экономического анализа. М.: Финансы и статистика, 1997.

6. Шеремет А.Д., Негашев Е.В. Методика финансового анализа. – М.: ИНФРА –М 1999.

7. Зайченко Ю., Рогоза С., Столбунов В. Сравнительный анализ методов оценки риска банкротства – [Электронный ресурс]. – режим доступа: http://www.foibg.com/ibs_isc/ibs–07/IBS–07–p15.pdf – свободный.

8. Недосекин А.О. Применение теории нечетких множеств к задачам управления финансами – [Электронный ресурс]. – режим доступа: http://www.cfin.ru/press/afa/2000–2/08–1.shtml – свободный.

9. Недосекин А.О., Максимов О.Б. Новый комплексный показатель оценки финансового состояния предприятия – [Электронный ресурс]. – режим доступа: http://www.vmgroup.ru/publications/public4.htm – свободный.

10. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М.: Мир, 1976.

11. Кофман А. Введение в теорию нечетких множеств: Пер. с франц. – М.: Радио и связь, 1982.

12. Официальный сайт Федеральной сетевой компании Единой энергетической системы [Электронный ресурс]. – Режим доступа: http://www.fsk–ees.ru/shareholders_and_investors/disclosure_of_information/quarterly_reports/ – свободный.

13. Михальчук А.А., Шинкеев М.Л. Статистический анализ экономических данных. Учебное методическое пособие. – Томск: Томск: Изд. ТПУ,2007.- 65с.
14. Мицель А.А., Кабалин А.А. Модели риска и прогнозирования банкротства предприятия//Управление риском, 2013, №1. – С. 44-52
15. Википедия. [Электронный ресурс]. – Режим доступа: http://ru.wikipedia.org/– свободный.

Приложение 1

Бухгалтерский баланс ОАО «ФСК ЕЭС» 2009г.

Приложение
к Приказу Минфина РФ
от 22.07.2003 № 67н

БУХГАЛТЕРСКИЙ БАЛАНС
на 31 декабря 200 9 г.

		КОДЫ	
	Форма № 1 по ОКУД	0710001	
	Дата (год, месяц, число)	2009 12 31	
Организация **Открытое акционерное общество "Федеральная сетевая компания Единой энергетической системы"**	по ОКПО	56947007	
Идентификационный номер налогоплательщика	ИНН	**4716016979**	
Вид деятельности **передача электроэнергии**	по ОКВЭД	40.10.2	
Организационно-правовая форма / форма собственности **открытое акционерное общество / смешанная российская собственность с долей федеральной собственности**	по ОКОПФ/ОКФС	47	41
Единица измерения: **тыс. руб.**	по ОКЕИ	384/385	

Местонахождение (адрес) **117630, г. Москва, ул. Ак. Челомея, д. 5А**

Дата утверждения	
Дата отправки (принятия)	30.03.2010

АКТИВ	Код показателя	На начало отчетного года	На конец отчетного периода
1	2	3	4
I. ВНЕОБОРОТНЫЕ АКТИВЫ			
Нематериальные активы, в том числе:	110	1 535 909	1 396 257
права на патенты, программы, товарные знаки и иные аналогичные	111	899 364	977 021
другие виды нематериальных активов	112	636 545	419 236
Основные средства, в том числе:	120	157 129 616	151 603 670
земельные участки и объекты природопользования	121	39 273	156 495
здания, машины и оборудование, сооружения	122	155 676 946	149 628 614
другие виды основных средств	123	1 413 397	1 818 561
Незавершенное строительство, в том числе:	130	150 373 965	216 529 585
оборудование к установке	131	13 418 743	18 484 815
вложения во внеоборотные активы	132	136 955 222	198 044 770
Доходные вложения в материальные ценности	135	-	-
Долгосрочные финансовые вложения	140	207 778 567	66 970 387
Отложенные налоговые активы	145	-	-
Прочие внеоборотные активы	150	1 653 608	1 415 088
ИТОГО по разделу I	190	518 471 665	437 914 987
II. ОБОРОТНЫЕ АКТИВЫ			
Запасы, в том числе:	210	3 305 661	2 427 514
сырье, материалы и другие аналогичные ценности	211	1 728 321	2 262 155
затраты в незавершенном производстве	213	-	-
готовая продукция и товары для перепродажи	214	84 863	29 993
расходы будущих периодов	216	1 492 477	135 366
прочие запасы и затраты	217	-	-
Налог на добавленную стоимость по приобретенным ценностям	220	1 961 283	2 070 794
Дебиторская задолженность (платежи по которой ожидаются более чем через 12 месяцев после отчетной даты), в том числе:	230	10 871 826	20 492 819
покупатели и заказчики	231	-	185 910
авансы выданные	234	14 413	36
прочие дебиторы	235	10 857 413	20 306 873
Дебиторская задолженность (платежи по которой ожидаются в течение 12 месяцев после отчетной даты), в том числе:	240	107 347 626	117 170 891
покупатели и заказчики	241	5 053 894	8 949 413
задолженность участников (учредителей) по взносам в уставный капитал	242	-	-
авансы выданные	243	68 970 457	67 036 337
прочие дебиторы	244	33 323 275	41 185 141
Краткосрочные финансовые вложения	250	49 390 019	69 127 725
Денежные средства, в том числе:	260	5 640 064	11 312 141
касса	261	2 617	2 439
расчетные счета	262	5 635 181	11 305 731
валютные счета	263	-	-
денежные документы	264	13	9
прочие денежные средства	265	2 253	3 962
Прочие оборотные активы	270	-	-
ИТОГО по разделу II	290	178 516 479	222 601 884
БАЛАНС	300	696 988 144	660 516 871

Форма 0710001 с. 2

ПАССИВ	Код показателя	На начало отчетного года	На конец отчетного периода
1	2	3	4
III. КАПИТАЛ И РЕЗЕРВЫ			
Уставный капитал	410	576 757 098	576 757 098
Собственные акции, выкупленные у акционеров	411	-	-
Добавочный капитал	420	59 502 413	59 386 652
Резервный капитал	430	9 910 770	10 134 044
Непокрытый убыток прошлых лет	460	(6 841 528)	(6 944 397)
Нераспределенная прибыль прошлых лет	465		
Нераспределенная прибыль отчетного года	470	-	-
Нераспределенный убыток отчетного года	475	-	(59 865 994)
ИТОГО по разделу III	490	639 328 753	579 467 403
IV. ДОЛГОСРОЧНЫЕ ОБЯЗАТЕЛЬСТВА			
Займы и кредиты	510	18 000 000	6 000 000
Отложенные налоговые обязательства	515	517 859	1 435 064
Прочие долгосрочные обязательства	520	130	5 098
ИТОГО по разделу IV	590	18 517 989	7 440 162
V. КРАТКОСРОЧНЫЕ ОБЯЗАТЕЛЬСТВА			
Займы и кредиты	610	16 161 487	7 481 469
Кредиторская задолженность, в том числе:	620	22 601 947	65 802 620
поставщики и подрядчики	621	10 175 866	11 018 708
задолженность перед персоналом организации	622	76 713	134 473
задолженность перед государственными внебюджетными фондами	623	15 102	18 906
задолженность по налогам и сборам	624	295 113	653 884
авансы полученные	627	7 096 192	7 114 653
прочие кредиторы	625	4 942 961	46 861 996
Задолженность перед участниками (учредителями) по выплате доходов	630	84 474	46 898
Доходы будущих периодов	640	293 494	278 319
Резервы предстоящих расходов	650	-	-
Прочие краткосрочные обязательства	660	-	-
ИТОГО по разделу V	690	39 141 402	73 609 306
БАЛАНС	700	696 988 144	660 516 871

Справка о наличии ценностей, учитываемых на забалансовых счетах			
Арендованные основные средства	910	3 272 196	6 189 371
в том числе по лизингу	911	-	-
Товарно-материальные ценности, принятые на ответственное хранение	920	812 468	633 348
Товары, принятые на комиссию	930	-	-
Списанная в убыток задолженность неплатежеспособных дебиторов	940	-	414 819
Обеспечения обязательств и платежей полученные	950	58 437 451	70 943 984
Обеспечения обязательств и платежей выданные	960	1 175 189	30 815
Износ жилищного фонда	970	-	-
Износ объектов внешнего благоустройства и других аналогичных объектов	980	-	-
Нематериальные активы, полученные в пользование	990	-	-
Спецодежда	991	-	-
Бланки строгой отчетности	1000	7	11

Руководитель ________ (подпись) Д.А. Трошенков (расшифровка подписи)   Главный бухгалтер ________ (подпись) В.В. Щукин (расшифровка подписи)

" 30 " марта 20 10 г.

Приложение 2

Отчет о прибылях и убытках ОАО «ФСК ЕЭС» за 2009г.

Приложение
к Приказу Минфина РФ
от 22.07.2003 № 67н

ОТЧЕТ О ПРИБЫЛЯХ И УБЫТКАХ
за 2009 г.

		КОДЫ		
	Форма № 2 по ОКУД	0710002		
	Дата (год, месяц, число)	2009	12	31
Организация **Открытое акционерное общество "Федеральная сетевая компания Единой энергетической системы"**	по ОКПО	56947007		
Идентификационный номер налогоплательщика	ИНН	**4716016979**		
Вид деятельности **передача электроэнергии**	по ОКВЭД	40.10.2		
Организационно-правовая форма / форма собственности **открытое акционерное общество / смешанная российская собственность с долей федеральной собственности**	по ОКОПФ/ОКФС	47	41	
Единица измерения: **тыс. руб.**	по ОКЕИ	384/385		

Показатель наименование	код	За отчетный период	За аналогичный период предыдущего года
1	2	3	4
Доходы и расходы по обычным видам деятельности			
Выручка (нетто) от продажи товаров, продукции, работ, услуг (за минусом налога на добавленную стоимость, акцизов и аналогичных обязательных платежей), в том числе:	010	85 077 809	68 485 030
услуги по передаче электроэнергии	011	80 173 317	66 128 765
прочая деятельность	012	4 904 492	2 356 265
Себестоимость проданных товаров, продукции, работ, услуг, в том числе:	020	(64 079 927)	(58 977 340)
услуги по передаче электроэнергии	021	(62 732 093)	(57 107 753)
прочая деятельность	022	(1 347 834)	(1 869 587)
Валовая прибыль (010 + 020)	029	20 997 882	9 507 690
Коммерческие расходы	030	-	-
Управленческие расходы	040	(5 128 305)	(4 351 940)
Прибыль (убыток) от продаж (029 + 030 + 040)	050	15 869 577	5 155 750
Прочие доходы и расходы			
Проценты к получению	060	7 291 952	6 806 385
Проценты к уплате	070	(1 717 506)	(2 385 645)
Доходы от участия в других организациях	080	717 256	223 272
Прочие доходы	090	105 760 531	31 347 105
Прочие расходы	100	(181 970 591)	(34 970 048)
Прибыль (убыток) до налогообложения (050 + 060 + 070 + 080 + 090 + 100)	140	(54 048 781)	6 176 819
Условный расход (доход) по налогу на прибыль (140 x 24% 2008г., х 20% 2009г)	143	10 809 756	(1 482 437)
Постоянные налоговые обязательства	200	(16 588 331)	(1 952 589)
Отложенные налоговые активы	141	(180 217)	6 767
Отложенные налоговые обязательства	142	(722 009)	(216 856)
Текущий налог на прибыль (143+200-141+142)	150	(4 876 349)	(3 224 937)
Иные аналогичные обязательные платежи	151	(4 642)	461 546
Корректировка налога на прибыль за прошлые периоды	152	(33 996)	1 262 136
Чистая прибыль (убыток) отчетного периода (140 + 143 + 200 + 151+152) или (140 + 141+142 + 150 + 151+152)	190	(59 865 994)	4 465 475
СПРАВОЧНО			
Базовая прибыль (убыток) на 100 000 000 акций	201	(5 190)	575
Разводненная прибыль (убыток) на 100 000 000 акций	202	-	-

Форма 0710002 с. 2

РАСШИФРОВКА ОТДЕЛЬНЫХ ПРИБЫЛЕЙ И УБЫТКОВ

Показатель		За отчетный период		За аналогичный период предыдущего года	
наименование	код	прибыль	убыток	прибыль	убыток
1	2	3	4	5	6
Штрафы, пени и неустойки, признанные или по которым получены решения суда (арбитражного суда) об их взыскании		397 323	40 390	152 738	1 861
Прибыль (убыток) прошлых лет		412 431	521 681	115 403	288 008
Возмещение убытков, причиненных неисполнением или ненадлежащим исполнением обязательств		-	-	-	-
Курсовые разницы по операциям в иностранной валюте		26 455	15 949	8 065	15 619
Отчисления в оценочные резервы		х	9 404 732	х	5 028 828
Списание дебиторских и кредиторских задолженностей, по которым истек срок исковой давности		6 556	9 122	25 169	17 230

Руководитель ____________ Д.А. Трошенков
(подпись) (расшифровка подписи)

Главный бухгалтер ____________ В.В. Щукин
(подпись)

" 30 " марта 2010 г.

Приложение 3

Бухгалтерский баланс ОАО «ФСК ЕЭС» 2010г.

МИ ФНС РФ ... № 4
ПОЛУЧЕНО
30 МАР 2011
ГОСНАЛОГИНСПЕКТОР
НИКУЛИНА Т.Д.
Подпись

Приложение
к Приказу Минфина РФ
от 22.07.2003 № 67н

БУХГАЛТЕРСКИЙ БАЛАНС

на 31 декабря 2010 г.

		КОДЫ	
Форма № 1 по ОКУД		0710001	
Дата (год, месяц, число)		2010 12 31	
Организация	Открытое акционерное общество "Федеральная сетевая компания Единой энергетической системы"	по ОКПО	56947007
Идентификационный номер налогоплательщика		ИНН	4716016979
Вид деятельности	передача электроэнергии	по ОКВЭД	40.10.2
Организационно-правовая форма / форма собственности	открытое акционерное общество / смешанная российская собственность с долей федеральной собственности	по ОКОПФ/ОКФС	47 / 41
Единица измерения:	тыс. руб.	по ОКЕИ	384/385
Местонахождение (адрес)	117630, г. Москва, ул. Ак. Челомея, д. 5А		
		Дата утверждения	
		Дата отправки (принятия)	21.03.2011

АКТИВ	Код показателя	На начало отчетного года	На конец отчетного периода
1	2	3	4
I. ВНЕОБОРОТНЫЕ АКТИВЫ			
Нематериальные активы, в том числе:	110	1 396 257	917 625
права на патенты, программы, товарные знаки и иные аналогичные	111	977 021	696 044
другие виды нематериальных активов	112	419 236	221 581
Основные средства, в том числе:	120	237 753 751	236 193 167
земельные участки и объекты природопользования	121	156 495	827 221
здания, машины и оборудование, сооружения	122	235 778 695	232 958 699
другие виды основных средств	123	1 818 561	2 407 247
Незавершенное строительство, в том числе:	130	216 529 585	298 644 138
оборудование к установке	131	18 484 815	17 905 969
вложения во внеоборотные активы	132	198 044 770	280 738 169
Доходные вложения в материальные ценности	135	-	-
Долгосрочные финансовые вложения	140	66 970 387	104 137 547
Отложенные налоговые активы	145	-	-
Прочие внеоборотные активы	150	1 415 088	894 579
ИТОГО по разделу I	190	524 065 068	640 787 056
II. ОБОРОТНЫЕ АКТИВЫ			
Запасы, в том числе:	210	2 427 514	4 632 226
сырье, материалы и другие аналогичные ценности	211	2 262 155	4 407 467
затраты в незавершенном производстве	213	-	-
готовая продукция и товары для перепродажи	214	29 993	30 011
расходы будущих периодов	216	135 366	194 748
прочие запасы и затраты	217	-	-
Налог на добавленную стоимость по приобретенным ценностям	220	2 070 794	2 295 467
Дебиторская задолженность (платежи по которой ожидаются более чем через 12 месяцев после отчетной даты), в том числе:	230	20 492 819	8 696 249
покупатели и заказчики	231	185 910	68 106
авансы выданные	234	36	-
прочие дебиторы	235	20 306 873	8 628 143
Дебиторская задолженность (платежи по которой ожидаются в течение 12 месяцев после отчетной даты), в том числе:	240	117 170 891	157 647 614
покупатели и заказчики	241	8 949 413	8 669 641
задолженность участников (учредителей) по взносам в уставный капитал	242	-	-
авансы выданные	243	67 036 337	97 636 854
прочие дебиторы	244	41 185 141	51 341 119
Краткосрочные финансовые вложения	250	69 127 725	46 244 024
Денежные средства, в том числе:	260	11 312 141	11 243 302
касса	261	2 439	3 664
расчетные счета	262	11 305 731	11 233 055
валютные счета	263	-	-
денежные документы	264	9	1
прочие денежные средства	265	[illegible]	6 582
Прочие оборотные активы	270	-	-
ИТОГО по разделу II	290	222 601 884	230 758 882
БАЛАНС	300	746 666 952	871 545 938

ПРИЛОЖЕНИЕ
к аудиторскому заключению
ЗАО "ПрайсвотерхаусКуперс Аудит"
28 МАР 2011

Аудитор

Приложение 4

Отчет о прибылях и убытках ОАО «ФСК ЭЭС» за 2010г.

МИ ФНС РФ по КН № 4
ПОЛУЧЕНО
2011
ГОСНАЛОГИНСПЕКТОР
НИКУЛИНА Т.Д.
Подпись

Приложение
к Приказу Минфина РФ
от 22.07.2003 № 67н

ОТЧЕТ О ПРИБЫЛЯХ И УБЫТКАХ
за 2010 г.

		КОДЫ
	Форма № 2 по ОКУД	0710002
	Дата (год, месяц, число)	2010 | 12 | 31
Организация **Открытое акционерное общество "Федеральная сетевая компания Единой энергетической системы"**	по ОКПО	56947007
Идентификационный номер налогоплательщика	ИНН	**4716016979**
Вид деятельности **передача электроэнергии**	по ОКВЭД	40.10.2
Организационно-правовая форма / форма собственности **открытое акционерное общество / смешанная российская собственность с долей федеральной собственности**	по ОКОПФ/ОКФС	47 | 41
Единица измерения: **тыс. руб.**	по ОКЕИ	384/385

Показатель		За отчетный период	За аналогичный период предыдущего года
наименование	код		
1	2	3	4
Доходы и расходы по обычным видам деятельности			
Выручка (нетто) от продажи товаров, продукции, работ, услуг (за минусом налога на добавленную стоимость, акцизов и аналогичных обязательных платежей), в том числе:	010	111 084 675	85 077 809
услуги по передаче электроэнергии	011	109 510 275	80 173 317
прочая деятельность	012	1 574 400	4 904 492
Себестоимость проданных товаров, продукции, работ, услуг, в том числе:	020	(75 518 397)	(64 079 927)
услуги по передаче электроэнергии	021	(74 694 570)	(62 732 093)
прочая деятельность	022	(823 827)	(1 347 834)
Валовая прибыль (010 + 020)	029	35 566 278	20 997 882
Коммерческие расходы	030	-	-
Управленческие расходы	040	(6 209 146)	(5 128 305)
Прибыль (убыток) от продаж (029 + 030 + 040)	050	29 357 132	15 869 577
Прочие доходы и расходы			
Проценты к получению	060	5 436 238	7 291 952
Проценты к уплате	070	(273 751)	(1 717 506)
Доходы от участия в других организациях	080	422 310	717 256
Прочие доходы	090	142 534 195	105 760 531
Прочие расходы	100	(109 157 601)	(181 970 591)
Прибыль (убыток) до налогообложения (050 + 060 + 070 + 080 + 090 + 100)	140	68 318 523	(54 048 781)
Условный расход по налогу на прибыль (140 х 20%)	143	(13 663 705)	10 809 756
Постоянные налоговые обязательства	200	3 184 752	(16 588 331)
Отложенные налоговые активы	141	(33 442)	(180 217)
Отложенные налоговые обязательства	142	(1 181 205)	(722 009)
Текущий налог на прибыль (143+200-141-142)	150	(9 264 306)	(4 876 349)
Иные аналогичные обязательные платежи	151	43 226	(4 642)
Корректировка налога на прибыль за прошлые периоды	152	205 592	(33 996)
Чистая прибыль (убыток) отчетного периода (140 + 143 + 200 + 151+152) или (140 + 141+142 + 150 + 151+152)	190	58 088 388	(59 865 994)
СПРАВОЧНО			
Базовая прибыль (убыток) на 100 000 000 акций	201	4 760	-5 190
Разводненная прибыль (убыток) на 100 000 000 акций	202	-	-

ПРИЛОЖЕНИЕ
к аудиторскому заключению
ЗАО "ПрайсвотерхаусКуперс Аудит"
29 МАР 2011

Аудитор

Форма 0710002 с. 2

РАСШИФРОВКА ОТДЕЛЬНЫХ ПРИБЫЛЕЙ И УБЫТКОВ

Показатель		За отчетный период		За аналогичный период предыдущего года	
наименование	код	прибыль	убыток	прибыль	убыток
1	2	3	4	5	6
Штрафы, пени и неустойки, признанные или по которым получены решения суда (арбитражного суда) об их взыскании	230	495 902	31 922	397 323	40 390
Прибыль (убыток) прошлых лет	240	320 195	229 331	412 431	521 681
Возмещение убытков, причиненных неисполнением или ненадлежащим исполнением обязательств	250	-	-	-	-
Курсовые разницы по операциям в иностранной валюте	260	10 970	10 017	26 455	15 949
Отчисления в оценочные резервы	270	x	18 611 054	x	9 404 732
Списание дебиторских и кредиторских задолженностей, по которым истек срок исковой давности	280	185	142 486	6 556	9 122

Руководитель ________ (подпись) А.А. Трошенков (расшифровка подписи)

Главный бухгалтер ________ (подпись) В.В. Щукин

" 29 " марта 2011 г.

ПРИЛОЖЕНИЕ
к аудиторскому заключению
ЗАО "ПрайсвотерхаусКуперс Аудит"

29 МАР 2011

Аудитор ________

7

Приложение 5

Данные опроса экспертной группы

Таблица 6.1 – Данные опроса экспертов по показателю X_1

	$Э_1$	$Э_2$	$Э_3$	$Э_4$	$Э_5$	$Э_6$	$Э_7$	$Э_8$
«очень низкий»	0; 0,1	0; 0,1	0; 0,2	0; 0,15	0; 0,1	0; 0,1	0; 0,1	0; 0,1
«низкий»	0,1; 0,3	0,1; 0,3	0,2; 0,4	0,15; 0,4	0,1; 0,3	0,1; 0,3	0,1; 0,3	0,1; 0,3
«средний»	0,3; 0,5	0,3; 0,5	0,4; 0,5	0,4; 0,45	0,3; 0,5	0,3; 0,5	0,3; 0,5	0,3; 0,45
«высокий»	0,5; 0,8	0,5; 0,8	0,5; 0,7	0,45; 0,6	0,5; 0,7	0,5; 0,6	0,5; 0,6	0,45; 0,75
«очень высокий»	0,8; 1	0,8; 1	0,7; 1	0,6; 1	0,7; 1	0,6; 1	0,6; 1	0,75; 1

Таблица 6.2 – Данные опроса экспертов по показателю X_2

	$Э_1$	$Э_2$	$Э_3$	$Э_4$	$Э_5$	$Э_6$	$Э_7$	$Э_8$
«очень низкий»	0; 0,2	0; 0,2	0; 0,4	0; 0,35	0; 0,4	0; 0,3	0; 0,3	0; 0,2
«низкий»	0,2; 0,6	0,2; 0,6	0,4; 0,7	0,35; 0,7	0,4; 0,65	0,3; 0,7	0,3; 0,6	0,2; 0,6
«средний»	0,6; 1	0,6; 1	0,7; 0,9	0,7; 0,8	0,65; 1	0,7; 0,9	0,6; 0,8	0,6; 0,8
«высокий»	1; 1,5	1; 1,8	0,9; 1,5	0,8; 1,7	1; 1,6	0,9; 1,5	0,8; 1,5	0,8; 1,8
«очень высокий»	1,5; ∞	1,8; ∞	1,5; ∞	1,7; ∞	1,6; ∞	1,5; ∞	1,5; ∞	1,8; ∞

Таблица 6.3 – Данные опроса экспертов по показателю X_3

	$Э_1$	$Э_2$	$Э_3$	$Э_4$	$Э_5$	$Э_6$	$Э_7$	$Э_8$
«очень низкий»	0; 0,08	0; 0,05	0; 0,06	0; 0,05	0; 0,05	0; 0,06	0; 0,08	0; 0,05
«низкий»	0,08; 0,2	0,05; 0,2	0,06; 0,12	0,05; 0,1	0,05; 0,15	0,06; 0,1	0,08; 0,18	0,05; 0,2
«средний»	0,2; 0,4	0,2; 0,4	0,12; 0,32	0,1; 0,3	0,15; 0,3	0,1; 0,3	0,18; 0,4	0,2; 0,4
«высокий»	0,4; 0,8	0,4; 1	0,32; 0,5	0,3; 0,6	0,3; 0,6	0,3; 0,5	0,4; 0,8	0,4; 0,8
«очень высокий»	0,8; ∞	1; ∞	0,5; ∞	0,6; ∞	0,6; ∞	0,5; ∞	0,8; ∞	0,8; ∞

Таблица 6.4 – Данные опроса экспертов по показателю X_4

	$Э_1$	$Э_2$	$Э_3$	$Э_4$	$Э_5$	$Э_6$	$Э_7$	$Э_8$
«очень низкий»	0; 0,12	0; 0,15	0; 0,12	0; 0,12	0; 0,12	0; 0,15	0; 0,15	0; 0,15
«низкий»	0,12; 0,25	0,15; 0,25	0,12; 0,2	0,12; 0,2	0,12; 0,2	0,15; 0,25	0,15; 0,25	0,15; 0,25
«средний»	0,25; 0,35	0,25; 0,45	0,2; 0,4	0,2; 0,4	0,2; 0,45	0,25; 0,35	0,25; 0,4	0,25; 0,35
«высокий»	0,35; 0,85	0,45; 1	0,4; 0,8	0,4; 0,8	0,45; 0,6	0,35; 0,65	0,4; 0,8	0,35; 0,75
«очень высокий»	0,85; ∞	1; ∞	0,8; ∞	0,8; ∞	0,6; ∞	0,65; ∞	0,8; ∞	0,75; ∞

Таблица 6.5 – Данные опроса экспертов по показателю X_6

	$Э_1$	$Э_2$	$Э_3$	$Э_4$	$Э_5$	$Э_6$	$Э_7$	$Э_8$
«очень низкий»	–1; 0	–1; 0	–1; 0	–1; 0	–1; 0	–1; 0	–1; 0	–1; 0
«низкий»	0; 0,01	0; 0,015	0; 0,006	0; 0,01	0; 0,01	0; 0,01	0; 0,012	0; 0,008
«средний»	0,01; 0,1	0,015; 0,1	0,006; 0,06	0,01; 0,06	0,01; 0,06	0,01; 0,1	0,012; 0,1	0,008; 0,08
«высокий»	0,1; 0,3	0,1; 0,4	0,06; 0,24	0,06; 0,3	0,06; 0,3	0,1; 0,4	0,1; 0,4	0,08; 0,24
«очень высокий»	0,3; ∞	0,4; ∞	0,24; ∞	0,3; ∞	0,3; ∞	0,4; ∞	0,4; ∞	0,24; ∞

Приложение А

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

ЮРГИНСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ (ФИЛИАЛ)
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ «НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Направление 230700 Прикладная информатика
Уровень: магистрант

Кафедра: информационных систем

Генерация случайных чисел с заданным законом распределения

Дисциплина: Системный анализ, управление и обработка информации в аналитической экономике

Выполнил: Студент гр. 17В20	________________ (Подпись)	Т.В. Внуковский
	________________ (Дата)	
Проверил: Преподаватель	________________ (Подпись)	И.О. Фамилия
	________________ (Дата)	

Юрга – 2016

Учебное издание

Составитель МИЦЕЛЬ Артур Александрович

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ В АНАЛИТИЧЕСКОЙ ЭКОНОМИКЕ

Методические указания к выполнению практических работ
по дисциплине «Системный анализ, управление и обработка информации в аналитической экономике»
для магистрантов, обучающихся по направлению 230700
«Прикладная информатика»

Печатается в редакции автора–составителя

Отпечатано в Издательстве ЮТИ ТПУ в полном соответствии с качеством предоставленного оригинал–макета