Практическое занятие № 2
по курсу «Статистические методы обработки данных»

Тема: Построение графиков различных распределений.
Точечные и интервальные оценки физической случайной величины.

Цель работы: Освоить методы обработки экспериментальных данных с помощью встроенных операторов MS-Excel

Для характеристики частоты появления различных значений случайной величины X (либо погрешностей приборов или результатов измерений с учетом ее систематической составляющей) теория вероятностей предлагает пользоваться указанием закона распределения вероятностей значений этой величины.

При этом различают два вида описания законов распределения:

1) интегральный
2) дифференциальный.

Интегральным законом, или функцией распределения вероятностей $F(x)$ случайной величины X, называют функцию, значения которой для каждого x есть вероятность события, состоящего в том, что случайная величина X меньше x, т.е.

$$F(x) = \mathrm{Pr}\,ob(x(k) < x) \qquad (1)$$

Очевидно, что

$F(a) \le F(b)$,при $a \le b$ (неубывающая функция) $F(-\infty) = 0, F(\infty) = 1$.

Для случайной величины с непрерывной и дифференцируемой функцией распределения F(x) можно найти дифференциальный закон распределения вероятностей. Он задается

$$p(x) = \frac{dF}{dx} \qquad (2)$$

$p(x)$ называют кривой плотности распределения.

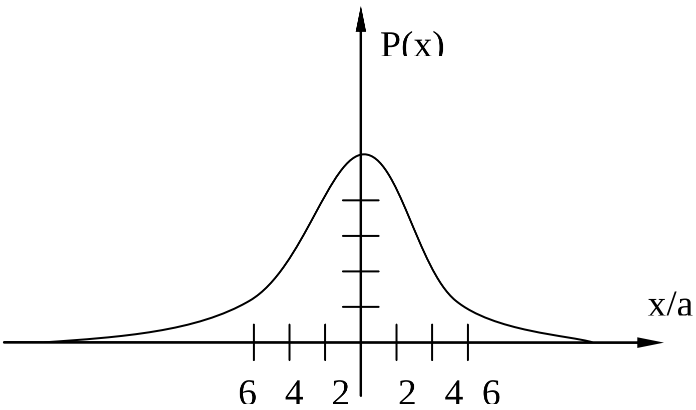
$p(x) \ge 0$; $\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$ - условие нормировки (3)

$$F(x) = \int_{-\infty}^{\text{ч}} p(\xi)\mathrm{d}\xi \qquad (4)$$

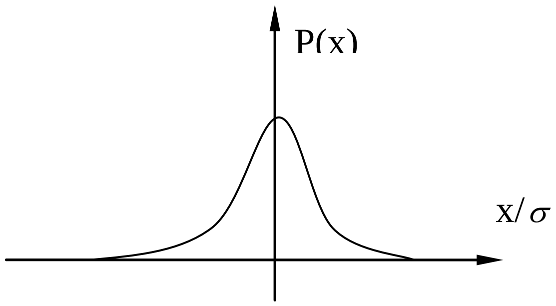
Примеры законов распределений.

А) Распределение Коши:

$$p(x) = \frac{a}{\pi(a^2 + x^2)} = \frac{1}{a\pi(1 + (x/a)^2)}$$



б) Нормальное распределение (распределение Гаусса) $P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{1}{2}(x/\sigma)^2]$



Моменты распределения.

Момент k -того порядка для непрерывной случайной величины выражается как:

$$\mu_k = \int_{-\infty}^{\infty} x^k p(x) \mathrm{d}x$$ (5) - алгебраический

$$\nu_k = \int (x-\mu)^k p(x) \mathrm{d}x$$ (б) - центральный.

где $p(x)$ – плотность распределения вероятности величины.
Первый момент называют математическим ожиданием

$$\mu_1 = M[x] = \int_{-\infty}^{\infty} x p(x) \mathrm{d}x \qquad (7)$$

и он характеризует центр распределения. Однако следует отметить, что не для всех распределений существует м.о. Например, распределение Коши. Интеграл (5) для него расходится.

Наиболее общей характеристикой центра распределения следует считать медиану. Медиана – прямая, параллельная оси Y, проходящая через точку на оси X, слева и справа от которой вероятности появления различных значений случайной величины равны между собой и составляют $P_1 = P_2 = 0.5$.

Для дискретной величины м.о. переходит в среднее арифметическое

$$M[x] = \sum_{i=1}^{n} x_i p(x = x_i) \qquad (8).$$

Второй центральный момент называется дисперсией случайной величины

$$v_2 = D[x] = \int_{-\infty}^{\infty} p(x)(x - \mu_1)^2 dx \qquad (9).$$

Дисперсия характеризует рассеяние отдельных значений случайной величины от центра распределения.

Для дискретных значений имеем:

$$\mu_2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (10)$$

Дисперсия имеет размерность квадрата случайной величины и выражает как бы мощность рассеяния, поэтому для более наглядной характеристики используют

$$\sigma = \sqrt{D} \qquad (11),$$

которое называют среднеквадратичным отклонением (с.к.о.) и имеет размерность самой случайной величины.

Третий центральный момент μ_3 характеризует асимметрию, т.е. скошенность распределения: когда один спад – крутой, а другой – пологий. Для симметричных относительно центра распределений он равен нулю. μ_3 имеет размерность куба случайной величины, поэтому для относительной характеристики используют безразмерный коэффициент

$$S = \frac{\mu_3}{\sigma^3} \qquad (12)$$

σ - с.к.о.

Четвертый центральный момент μ_4 характеризует протяженность распределения (иногда в литературе считают, что островершинность). Значение

$$\varepsilon = \frac{\mu_4}{\sigma^4} \qquad (13)$$ называют эксцессом распределения.

$$\varepsilon \in [1, \infty[$$

Например, для нормального (кругловершинного) распределения ε =3,0.
Часто используют коэффициент эксцесса:

$\gamma_2 = \varepsilon - 3$, который для менее протяженных распределений (треугольное, равномерное) – отрицательный (от –2 до 0), а для более протяженных, чем нормальное – положителен (от 0 до ∞). Для классификации распределений часто (и удобнее) использовать другую функцию от эксцесса

$æ = \frac{1}{\sqrt{\varepsilon}}$ - контрэксцесс (9). $æ \in [0,1]$.

ЗАДАЧА 1.
а) Построить график распределения Лапласса, где плотность распределения задается формулой $p(x)=A\exp(-|x|^{\alpha})$, $A(\alpha)=\alpha/2\Gamma(1/\alpha)$, $\alpha=3$.

Построить графики с помощью встроенных функций MS-Excel:
б) плотности нормального распределения (a=0, σ =1; a=10, σ =2);
в) интегральный закон нормального распределения (a=0, σ =1; a=10, σ =2).

ЗАДАЧА 2. Загрузить из файла выборку для своего варианта *(например, если вариант 7, то извлекаются данные из файла* d7.dat *).*

ЗАДАЧА 3. С помощью формул в конспекте найти точечные оценки физической случайной величины для заданной выборки:
а) мат. ожидание $\bar{x}$,
б) дисперсию S,
в) среднее квадратичное отклонение среднего арифметического $S_{\bar{x}}$,
г) 4-й момент эксцесс ε,
д) коэффицент ассиметрии.

ЗАДАЧА 4. Проделать ЗАДАЧУ 3. с помощью встроенных функций MS-Excel. Провести анализ и сравнение полученных данных. Сделать вывод.
ЗАДАЧА 5. С помощью встроенных функций MS-Excel найти интервальные оценки физической случайной величины для заданной выборки, где p=1- α, p=0.95, p=0.99, p=0.90 , т.е. найти:
а) доверительный интервал для мат. ожидания;
б) доверительный интервал для дисперсии;
в) сравнить доверительный интервал для мат. ожидания с интервалом, полученным с помощью $\bar{x}$ и $S_{\bar{x}}$, сделать вывод.

ЗАДАЧА 6. Подготовить отчет. Включить результаты выполнения заданий из ПЗ№1 - 2.

Составители: Андреев В.В. (2014),
Бабич К.С. (2016)