

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

Ю.С. Радченко

МЕТОДЫ ОБРАБОТКИ И ПЛАНИРОВАНИЯ
ЭКСПЕРИМЕНТА

Часть 2.

ПРОВЕРКА ГИПОТЕЗ,
АППРОКСИМАЦИЯ РАСПРЕДЕЛЕНИЙ

Учебно-методическое пособие

Воронеж
Издательский дом ВГУ
2017

ОГЛАВЛЕНИЕ

Введение.....	4
1. РАЗЛОЖЕНИЕ ПЛОТНОСТИ ВЕРОЯТНОСТИ В РЯД	4
1.1. Общие соотношения.....	4
1.2. Разложение по полиномам Эрмита.....	5
1.3. Разложение по полиномам Лагерра.....	7
2. ПРОВЕРКА ГИПОТЕЗ.....	12
2.1. Классификация гипотез	12
2.2. Общие характеристики	13
2.3. Проверка простых статистических гипотез о среднем нормального распределения (при известной дисперсии)	15
2.4. Тактическое планирование эксперимента по проверке гипотез о математическом ожидании	18
3. ПРОВЕРКА СЛОЖНЫХ ПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ.....	19
3.1. Проверка сложной параметрической гипотезы о среднем значении выборки	19
3.2. Проверка сложной гипотезы о равенстве средних значений двух выборок	21
4. ПРОВЕРКА СЛОЖНЫХ НЕПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ ОБ ОДНОРОДНОСТИ ВЫБОРОК	24
4.1. Критерий однородности Смирнова	24
4.2. Критерий однородности хи-квадрат.....	24
4.3. Критерии отбраковки аномальных значений гауссовской выборки	25
4.4. Критерии для выявления k экстремальных членов вариационного ряда, выпадающих из общей негауссовской статистики данных	26
4.5. Критерии наличия тренда и выбросов.....	29

Рекуррентное выражение для нахождения полиномов (1.6), (1.7)

$$H_{n+1}(z) - zH_n(z) + nH_{n-1}(z) = 0.$$

В выражении (1.5) использованы коэффициенты разложения вида

$$c_k = \frac{1}{\sqrt{k!}} \int_{-\infty}^{\infty} f(x) H_k(x) dx. \quad (1.8)$$

Если в (1.8) подставить (1.5), то получим

$$\begin{aligned} c_0 &= 1, \quad c_1 = 0, \quad c_2 = 0, \\ c_3 &= \frac{\kappa_3}{3!}, \quad c_4 = \frac{\kappa_4}{4!}, \quad c_5 = \frac{\kappa_5}{5!}, \\ c_6 &= \frac{\kappa_6 + 10\kappa_3^2}{6!}, \quad c_7 = \frac{\kappa_7 + 35\kappa_4\kappa_3}{7!}. \end{aligned} \quad (1.9)$$

Здесь κ_r – кумулянт порядка r распределения $f(x)$. Кумулянты получают-ся разложением в ряд Тейлора логарифма характеристической функции

$$\begin{aligned} \theta_f(u) &= \int_{-\infty}^{\infty} f(x) \exp(-jux) dx, \\ \ln(\theta(u)) &= \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} (ju)^r. \end{aligned}$$

Нетрудно получить

$$\kappa_1 = m_f, \quad \kappa_2 = \mu_{2f}, \quad \kappa_3 = \mu_{3f}, \quad \kappa_4 = \mu_{4f} - 3\mu_{2f}^2.$$

В силу нормировки и центрирования $f(x)$ получаем $\kappa_1 = 0, \kappa_2 = 1$.

Разложение (1.5) с учетом (1.7) и (1.9) называется **разложение Грамма–Шарлье**. Эджворт путем перегруппировки слагаемых с высшими кумулянтами в разложении Грамма–Шарлье получил другой ряд, названный **рядом Эджворта** (1.10).

$$\begin{aligned} f(x) &= \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \times \\ &\times \left[1 + \frac{\kappa_3}{3!} H_3(x) + \frac{\kappa_4}{4!} H_4(x) + \frac{10\kappa_3^2}{6!} H_6(x) + \frac{\kappa_5}{5!} H_5(x) + \frac{35\kappa_3\kappa_4}{5!} H_7(x) \right]. \end{aligned} \quad (1.10)$$

Нетрудно увидеть, что

$$\frac{\kappa_3}{3!} = \gamma_1 - \text{коэффициент асимметрии,}$$

$$\frac{\kappa_4}{4!} = \gamma_2 - \text{коэффициент эксцесса.}$$

Перейдем теперь к распределению *ненормированной и нецентрированной* случайной величины ξ . Получаем преобразованную плотность вероятности

$$W_{\xi}(x) = \frac{1}{\sigma_{\xi}} f\left(\frac{x - m_{\xi}}{\sigma_{\xi}}\right). \quad (1.11)$$

Таким образом, **ряд Эджворта** для (1.11) имеет вид

$$W_{\xi}(x) = \frac{\exp(-(x - m_{\xi})^2 / 2\sigma_{\xi}^2)}{\sigma_{\xi} \sqrt{2\pi}} \times \left[1 + \frac{\kappa_3}{3!} H_3\left(\frac{x - m_{\xi}}{\sigma_{\xi}}\right) + \frac{\kappa_4}{4!} H_4\left(\frac{x - m_{\xi}}{\sigma_{\xi}}\right) + \right. \\ \left. + \frac{10\kappa_3^2}{6!} H_6\left(\frac{x - m_{\xi}}{\sigma_{\xi}}\right) + \frac{\kappa_5}{5!} H_5\left(\frac{x - m_{\xi}}{\sigma_{\xi}}\right) + \frac{35\kappa_3\kappa_4}{5!} H_7\left(\frac{x - m_{\xi}}{\sigma_{\xi}}\right) \right] \quad (1.12)$$

Первые три слагаемых ряда Грамма–Шарлье (1.5) и Эджворта (1.10) совпадают. На практике этими слагаемыми в (1.12) ограничиваются.

ПРИМЕР. Получить разложение в ряд Эджворта распределение Максвелла
Решение.

$$W_{\xi}(x) = \sqrt{\frac{2}{\pi}} x^2 \exp(-x^2 / 2).$$

m1 := 1.596

μ2 := 0.454

g1 := 0.486

g2 := 0.108

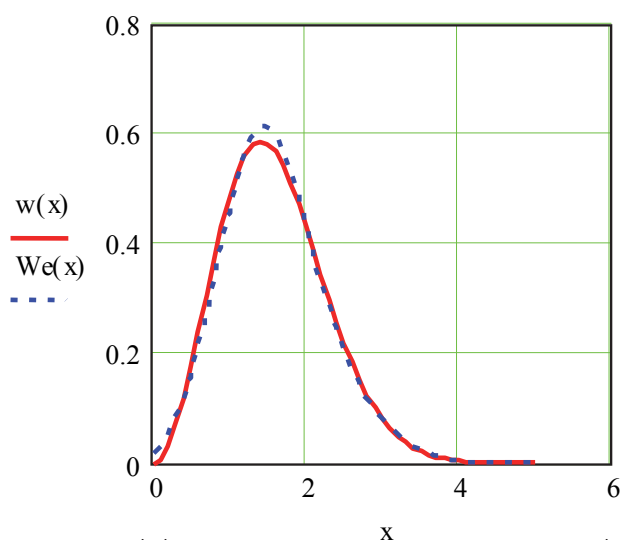


Рис. 1.2. Распределение Максвелла и его ряд Эджворта

1.3. Разложение по полиномам Лагерра

Если $\rho(x) = x^{\alpha} \exp(-x)$, $0 \leq x < \infty$, то получается ряд **Лагерра** по полиномам Лагерра. В качестве опорного распределения берется **гамма-распределение** (с точностью до нормировочной константы). Полиномы Лагерра удовлетворяют условию ортогональности

$$\int_0^{\infty} z^{\alpha} e^{-z} L_k^{(\alpha)}(z) L_m^{(\alpha)}(z) dz = \delta_{km} \cdot \frac{\Gamma(k + \alpha + 1)}{\Gamma(k + 1)}.$$

Первые четыре полинома Лагерра имеют вид:

$$\begin{aligned} L_0^\alpha(z) &= 1, \quad L_1^\alpha(z) = (\alpha + 1) - z, \\ L_2^\alpha(z) &= 0.5(\alpha + 2)(\alpha + 1) - (\alpha + 2)z + 0.5z^2 \\ L_3^\alpha(z) &= 1/6(\alpha + 3)(\alpha + 2)(\alpha + 1) - 1/2(\alpha + 3)(\alpha + 2)z + 1/2(\alpha + 3)z^2 - 1/6z^3. \end{aligned} \quad (1.13)$$

Полиномы Лагерра связаны рекуррентным соотношением, по которому можно рассчитать любой полином

$$(m + 1)L_{m+1}^\alpha(z) + (z - 2m - 1 - \alpha)L_m^\alpha(z) + (\alpha + m)L_{m-1}^\alpha(z) = 0.$$

Разложение нормированной случайной величины с плотностью вероятности $f(z)$, $z \geq 0$ можно записать как

$$f(z) = z^\alpha e^{-z} \sum_{k=0}^{\infty} c_k L_k^{(\alpha)}(z), \quad z \geq 0 \quad (1.14)$$

На рис.1.3 приведены графики некоторых функций Лагерра

$$Lag_k(\alpha, x) = x^\alpha e^{-x} L_k^{(\alpha)}(x).$$

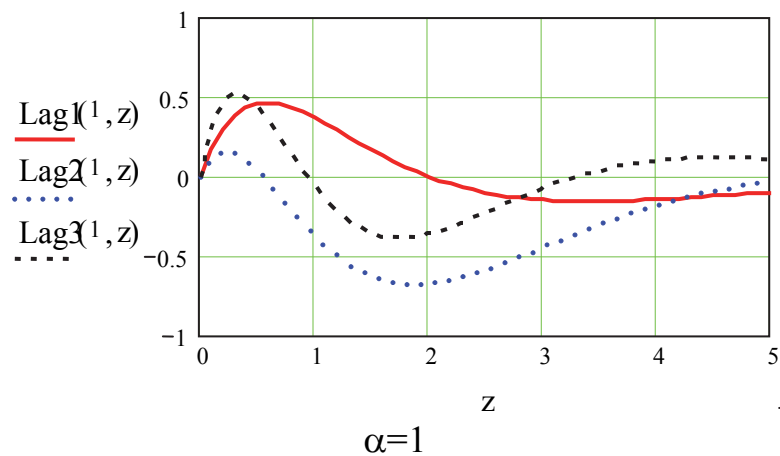
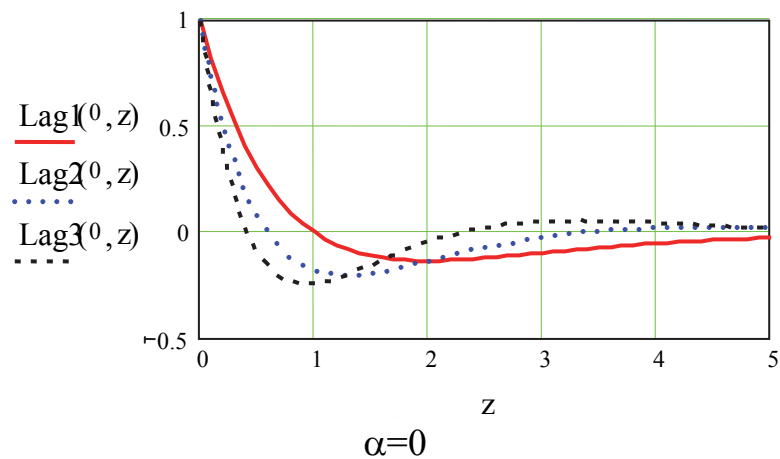


Рис. 1.3. Графики функций Лагерра первых порядков

Коэффициенты разложения в (1.14) считаются по формуле

$$c_k = \frac{k!}{\Gamma(k + \alpha + 1)} \int_0^\infty f(z) L_k^{(\alpha)}(z) dz.$$

В теории вероятностей двухпараметрическое гамма-распределение имеет вид

$$W(x) = \frac{1}{\beta \Gamma(\alpha + 1)} \left(\frac{x}{\beta} \right)^\alpha \exp\left(-\frac{x}{\beta}\right), \quad x \geq 0, \alpha > -1, \beta > 0.$$

Для расчетов целесообразно ввести нормировку $z = x / \beta$. Тогда

$$W(x) = \frac{1}{\beta} f\left(\frac{x}{\beta}\right).$$

Поэтому

$$W_\xi(x) = \frac{1}{\beta} \left(\frac{x}{\beta} \right)^\alpha \exp\left(-\frac{x}{\beta}\right) \sum_{k=0}^{\infty} d_k L_k^{(\alpha)}\left(\frac{x}{\beta}\right), \quad (1.15)$$

где

$$d_k = \frac{k!}{\Gamma(k + \alpha + 1)} \int_0^\infty W_\xi(x) L_k^{(\alpha)}\left(\frac{x}{\beta}\right) dx. \quad (1.16)$$

Константы α, β в (1.15), (1.16) целесообразно выразить через моменты

$$\alpha = \frac{\kappa_1^2}{\kappa_2} - 1 = \frac{m_\xi^2}{\sigma_\xi^2} - 1,$$

$$\beta = \frac{\kappa_2}{\kappa_1} = \frac{\sigma_\xi^2}{m_\xi}.$$

Тогда

$$d_0 = \frac{1}{\Gamma(\alpha + 1)}, \quad d_1 = d_2 = 0, \quad d_3 = \frac{\alpha + 1}{\Gamma(\alpha + 4)} \left(2 - \frac{\kappa_1 \kappa_3}{\kappa_2^2} \right), \quad (1.17)$$

$$d_4 = \frac{\alpha + 1}{\Gamma(\alpha + 5)} \left(18 - \frac{12 \kappa_3 \kappa_1}{\kappa_2^2} + \frac{\kappa_4 \kappa_1^2}{\kappa_2^3} \right),$$

$$d_5 = \frac{\alpha + 1}{\Gamma(\alpha + 6)} \left(144 - \frac{120 \kappa_3 \kappa_1}{\kappa_2^2} + \frac{20 \kappa_4 \kappa_1^2}{\kappa_2^3} - \frac{\kappa_5 \kappa_1^3}{\kappa_2^4} \right).$$

ПРИМЕР Получить разложение в ряд Лагерра распределения Максвелла.

Решение: $W_{\xi}(x) = \sqrt{\frac{2}{\pi}} x^2 \exp(-x^2 / 2)$.

$$m_1 = 1.596, \quad m_2 = 3, \quad m_3 = 6.383,$$

$$\mu_2 = 0.453, \quad \alpha = 4.626, \quad \beta = 0.284.$$

Из выражений (1.17) следует:

$$d_0 = \frac{1}{\Gamma(\alpha + 1)} = 0.016;$$

$$d_3 = \frac{1}{\Gamma(\alpha + 4)} \left[\frac{m_2}{\beta^2} (\alpha + 3) - \frac{m_3}{\beta^3} \right] = 2.57 \cdot 10^{-4}.$$

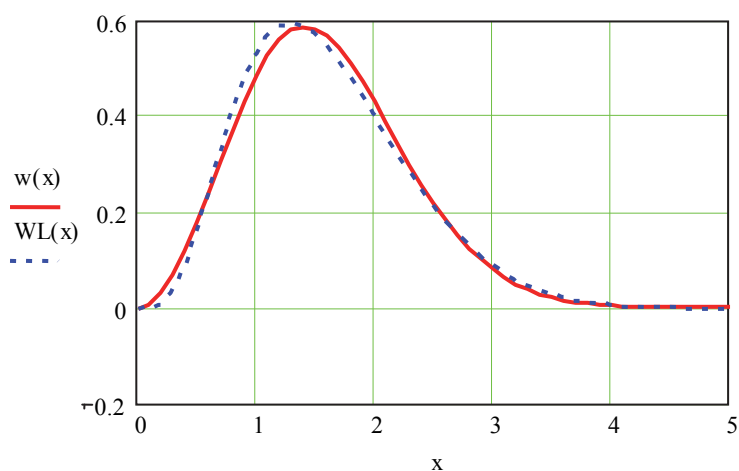


Рис.1.4. Распределение Максвелла и его ряд Лагерра

Пример. Анализ отличий двух кадров изображений

Для определения отличий изображений можно применять решающие правила,

$$D_0 = \frac{\sum_{k,m} (X_{k,m} - C_{k,m}^{(1)})^2}{(C_{0,0}^{(1)})^2} > h_0 \quad \text{или} \quad D_E = \frac{\sum_{k,m} (X_{k,m} - C_{k,m}^{(1)})^2}{\sum_{k,m} (C_{k,m}^{(1)})^2} > h_E. \quad (1.18)$$

Здесь $C_{k,m}^{(1)}$ – спектр опорного изображения, $X_{k,m}$ – спектр текущего кадра.

Решающие статистики (1.18) в общем случае будут подчиняться распределению, которое можно аппроксимировать рядом Лагерра.