Лабораторная работа №3

Анализ закона распределения случайной величины

1. Цель работы

Изучение методов оценки интегрального закона распределения и плотности вероятности непрерывной случайной величины, анализ сходимости эмпирических законов распределения к истинным и получение навыков применения современных методов экспресс анализа распределения по выборочным значениям.

2. Необходимые теоретические сведения

Распределение непрерывной случайной величины в равной степени характеризуется функцией плотности вероятности и интегральной функцией распределения, соответственно можно выделить два классических приема анализа распределения случайной величины:

1. оценка плотности вероятности методом гистограмм;
2. построение эмпирической интегральной функции распределения.

Метод гистограмм.

По выборке можно осуществить оценку $\Pr\{\xi \in (a,b]\}$ следующим образом:

$$\Pr\{\xi \in (a,b]\} = \frac{1}{N}\sum_{k=1}^{N} I\{\xi_k \in (a,b]\},$$

где N – объем выборки, ξ_k - k-й элемент выборки, $I\{t\}$ - индикатор события t. Тогда, разбив интервал анализа плотности вероятности на множество из M неперекрывающихся подинтервалов $[a_1,b_1],(a_2,b_2],\ldots(a_M,b_M]$, можно построить следующую оценку:

$$\hat{f}_{\xi}(x)=\begin{cases}\dfrac{1}{(b_i-a_i)N}\sum\limits_{k=1}^{N} I\left\{\xi_k \in \left(a_i,b_i\right]\right\} & \text{для} \qquad x\in\left(a_i,b_i\right],\ i=\overline{1,M} \\ 0 & \text{иначе}\end{cases}.$$

(3.1)

Графическое изображение функции (3.1) называется *гистограммой* и имеет характерный ступенчатый вид.

Число интервалов разбиения рекомендуется выбирать исходя из правила Стерджеса: $M=1+\left[\log_2 N\right]$, где N – объем выборки, $[\cdot]$ – знак округления. Длины интервалы выбираются, как правило, равными, причем $a_1=\xi_{(1)}$, $b_M=\xi_{(N)}$, т.е. минимальный и максимальный элементы выборки соответственно.

Ошибку в оценке плотности вероятности можно количественно охарактеризовать, посчитав средний квадрат относительного отклонения оценки $\Pr\left\{\xi\in\left(a,b\right]\right\}$ от истинного значения $\Pr\left\{\xi\in\left(a,b\right]\right\}$:

$$e_h=\sum_{i=1}^{M}\left(1-\frac{\Pr\left\{\xi\in\left(a_i,b_i\right]\right\}}{\Pr\left\{\xi\in\left(a_i,b_i\right]\right\}}\right)^2\times\Pr\left\{\xi\in\left(a_i,b_i\right]\right\}.$$

Построение эмпирической интегральной функции.

Для оценки интегральной функции обычно исходят из определения, согласно которому $F_{\xi}(x)$ показывает, с какой вероятностью случайная величина попадает в область $\left(-\infty,x\right]$. Соответственно оценка может быть осуществлена, например, следующим способом:

$$\hat{F}_{\xi}(x)=\frac{1}{N}\sum_{k=1}^{N} I\left\{\xi_k\le x\right\}. \qquad (3.2)$$

Оценка $\hat{F}_{\xi}(x)$ называется *эмпирической интегральной функцией распределения*. Как можно видеть из выражения (3.2), $\hat{F}_{\xi}(x)$ является ступенчатой функцией, скачкообразно изменяющейся в точках

$\left\{\xi_{(1)},\xi_{(2)},\ldots,\xi_{(N)}\right\}$ на величину, равную $1/N$. Ошибку в оценивании можно определить как супремум (максимум) абсолютной разницы между $F_\xi(x)$ и ее оценкой:

$$e_c = \sup_x \left|\hat{F}_\xi^{\langle N\rangle}(x) - F_\xi(x)\right|.$$

Метод Каллена-Фрея.

Одним из способов экспресс-оценки распределения является метод диаграмм Каллена-Фрея (см. рис. 3.1). Согласно данному методу, по выборке оцениваются *коэффициент асимметрии* и *коэффициент эксцесса*, по которым можно приближенно определить класс распределений, к которому принадлежит выборка.

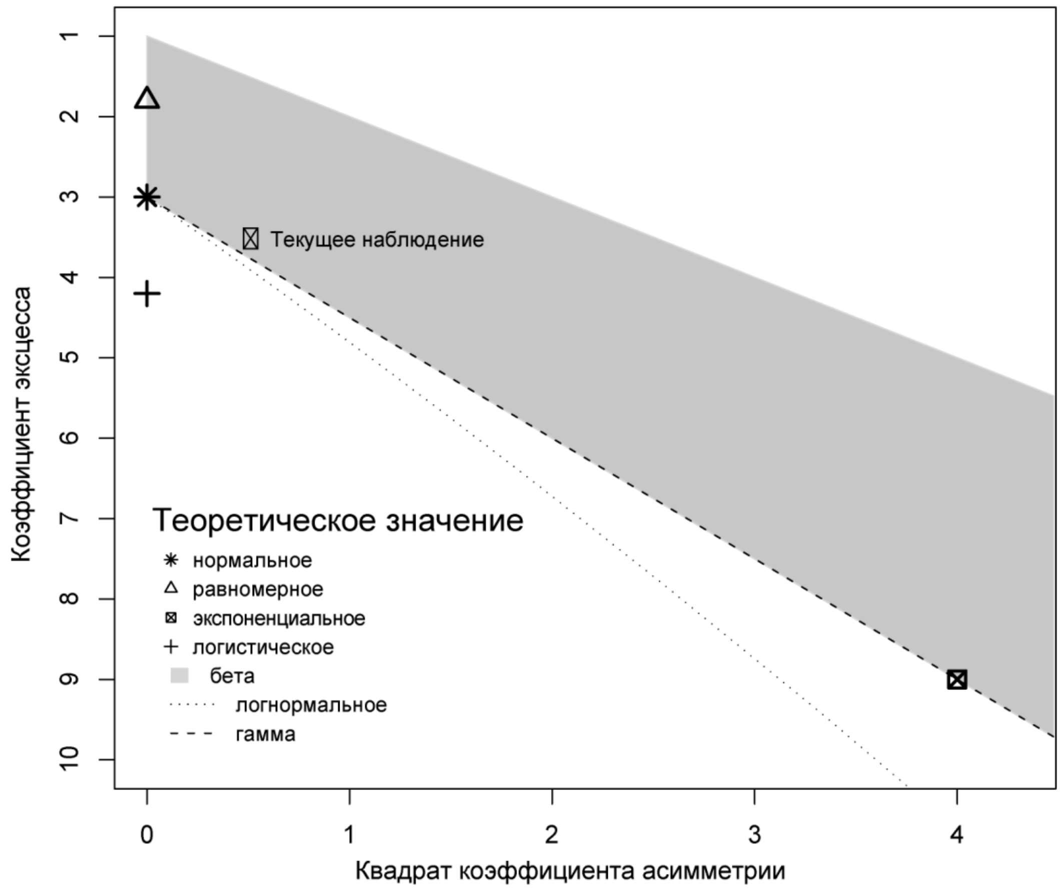

Рис. 3.1. Диаграмма Каллена-Фрея

Коэффициент асимметрии количественно характеризует степень отклонения формы плотности вероятности от симметричной функции:

$$\gamma_1 = \frac{\beta_3}{\sigma^3},$$

где β_3 — третий центральный момент, σ — среднеквадратическое отклонение.

Коэффициент эксцесса — это мера остроты пика распределения случайной величины:

$$\gamma_2 = \frac{\beta_4}{\sigma^4},$$

где β_4 — четвертый центральный момент.

Оценив по выборке соответствующие моменты (например, по методам, изученным в л/р № 1) и отложив $\hat{\gamma}_1^2$ и $\hat{\gamma}_2$ на осях диаграммы Каллена-Фрея можно приближенно оценить, к какому классу распределений принадлежит наблюдаемая выборка.

3. Порядок выполнения работы

1. Записать плотность вероятности и интегральную функцию распределения случайной величины согласно варианту (см. таб. 3.1).
2. Сформировать выборки объемом $n = 2^i$ чисел, распределенных по заданному закону (i = 5, 7, 10 и 14).
3. Для каждой из выборок построить гистограмму и эмпирическую интегральную функцию распределения. Количество подинтервалов при построении гистограммы рассчитать по формуле Стерджеса, длины подинтервалов выбирать равными.
4. По выборкам найти выборочные коэффициенты асимметрии и эксцесса. Сравнить полученные значения с истинными.
5. Построить графики зависимости величин e_h и e_c от объема выборки.

4. Варианты задания

Таб. 3.1. Вариант заданий к лабораторной работе №3

Номер варианта	Распределение	Параметры распределения
1	Экспоненциальное	$\lambda = 2$

5. Содержание отчета

1. Цель работы.
2. График плотности вероятности и интегральной функции распределения выбранного закона распределения.
3. Описание разработанной программы: список использованных переменных, блок-схема, листинг программы.
4. Гистограммы и эмпирические интегральные функции распределения, построенные по пункту 3 порядка выполнения лабораторной работы. На рисунках изобразить истинные интегральные и дифференциальные функции распределения.
5. Диаграмма Каллена-Фрея с отмеченными на ней выборочными коэффициентами асимметрии и эксцесса (по пункту 4 порядка выполнения лабораторной работы).
6. Графики, построенные по пункту 5 порядка выполнения лабораторной работы.
7. Количественные и качественные выводы.

ПРИЛОЖЕНИЕ 1

Основные функции и операторы Octave/Matlab

Функция/оператор	Описание
`x = zeros(M,N);`	Создать массив, состоящий из нулей, размером [M x N].
`x = ones(M,N);`	Создать массив, состоящий из единиц, размером [M x N].
`x = eye(N);`	Создать единичную матрицу размером [N x N].
`z = x + y; z = x - y;`	Поэлементное сложение/вычитание массивов.
`z = x.*y; z = x./y;`	Поэлементное перемножение/деление массивов.
`z = x.^N;`	Поэлементное возведение элементов массива x в степень N.
`z = x*y;`	Матричное перемножение матриц x и y.
`z = x';`	Транспонирование матрицы.
`z = inv(x);`	Обращение матрицы.
`L = length(x);`	Длина вектора x.
`[M,N] = size(x);`	Размер матрицы x.
`Y = mean(x)`	Нахождение среднего арифметического элементов массива.
`Y = sort(x)`	Сортировка элементов массива по возрастанию.
`Y = ceil(X)`	Округление до ближайшего целого $\geq X$.
`Y = fix(X)`	Усечение дробной части числа.
`Y = floor(X)`	Округление до ближайшего целого $\leq X$.
`Y = round(X)`	Округление до ближайшего целого.
`z = X(n)`	Вывод n-го элемента массива.
`z = min(x)`	Нахождение минимального элемента массива.
`z = max(x)`	Нахождение максимального элемента массива.
`mdx = median(X)`	Возвращает значение срединного элемента массива.
`sin(x), cos(x), tan(x)`	Тригонометрические функции.
`pi`	Число Пи.
`plot(x,y)`	Построение обычной функции в линейном масштабе, когда одномерный массив x соответствует значениям аргумента, а одномерный массив y - значениям функции.

`semilogx(x,y)`
`semilogy(x,y)`
`loglog(x,y)`
	Построение графика в полулогарифмическом (по одной из осей) и логарифмическом масштабах (по основанию 10).

ПРИЛОЖЕНИЕ 2

Формирование случайных величин в пакетах Octave/Matlab

Закон распределения	Способ формирования[1]
Нормальный	`X = random('Normal', mu, sigma, [M,N]);`
Непрерывный равномерный	`X = random('Uniform', a, b, [M,N]);`
Экспоненциальный	`X = random('Exponential', 1/lambda, [M,N])`

Симпсон	`X1 = random('Uniform', 0, (b-a)/2, [M,N]);`
	`X2 = random('Uniform', 0, (b-a)/2, [M,N]);`
	`X = (a+b)/2 + (X1 - X2);`

Коши	`X = mu + random('T', 1, [M,N]);`
Стьюдент	`X = mu + random('T', n, [M,N]);`
Арксинусный	`X = mu + sin( random('Uniform', -pi, +pi, [M,N]) );`
Логистический	`X1 = random('Uniform', 0, 1, [M,N]);` `X = mu + s*log( X1./(1-X1) );`

Лаплас	`X1 = random('Discrete Uniform', 2, [M,N]);`
	`X2 = random ('Exponential', 1/lambda, [M,N]);`
	`X = mu + ( (-1).^ X1 ).*X2;`

Релей	`X = random('Rayleigh', sigma, [M,N]);`
Биномиальный	`X = random('Binomial', N, p, [M,N]);`
Дискретный равномерный	`X = a-1 + random('Discrete Uniform', b-a+1, [M,N]);`

[1] M, N – размерность формируемого массива случайных величин, X – сформированный массив