МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»

Кафедра теоретической физики

СТАТИСТИЧЕСКАЯ ОБРАБОТКА ФИЗИЧЕСКОЙ ИНФОРМАЦИИ

ПРАКТИЧЕСКОЕ ПОСОБИЕ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Гомель 2005

Авторы - составители:

В.В.Андреев, доцент, кандидат физико-математических наук

Практическое пособие по курсу «Статистическая обработка физической информации» составлено в соответствии с учебными планами подготовки специалистов по специальностям:

I–31 0401-03 – «Физика (научно-педагогическая деятельность)»

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Введение

Современный инженер должен быть подготовлен к организационно-управленческой, научной, производственной и преподавательской деятельности в области экспериментального исследования физических процессов на различных уровнях структурной организации материи при различных физических условиях; теоретического анализа эффектов и явлений и предсказания новых физических закономерностей на основе современных теоретических представлений, математических и компьютерных методов; разработке приборов на основе новых материалов и физических принципов, созданию новых технологий, использующих математические методы и компьютерную технику; работе по математическому моделированию разнообразных процессов и объектов.

Одним из обязательных требований такого рода специалистов является их способность планировать, организовывать и проводить физические и радиофизические эксперименты, применяя вероятностные методы анализа и используя технические средства автоматизации эксперимента; обрабатывать и анализировать полученные результаты, создавать математические модели и программные средства, составлять отчеты и вести научно-техническую документацию.

Как известно статистическая обработка информации в физических исследованиях играет одну из важнейших ролей. Новые неизвестные явления открываются в настоящее время путем тщательной обработки результатов измерений. Современный физик обязан знать основы статистической обработке данных и с помощью программных продуктов

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

уметь проводить анализ и обработку физической информации на персональном компьютере.

Исходя из этих требований, следует что, такой специальный курс, как «Статистическая обработка физической информации» является необходимой составной частью образования современного инженера-физика.

Целью специального курса «Статистическая обработка физической информации» является овладение студентами основами математической статистики и теории вероятностей и их использование при обработке и анализе экспериментальных данных.

Задачами специального курса «Статистическая обработка физической информации» являются

– приобретение навыков и изучение способов, приёмов работы по анализу и статистической обработке информации на персональном компьютере, наиболее часто используемых в физических исследованиях (обратив внимание на смысл понятий критериев, возможности алгоритмизации),

– приобретение навыков по решению теоретических и экспериментальных задач различных физических дисциплин, таких как квантовая механика, физика ядра и атома, радиационная безопасность и другие.

– овладение новыми методами теоретических и экспериментальных исследований (решение практических задач, методы получения и обработки результатов опытов).

Материал специального курса «Статистическая обработка физической информации» базируется на ранее

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

полученных знаниях по курсу «Теория вероятностей». Полученные навыки могут быть использованы при изучении дисциплин учебного плана, таких как «Физика ядра», «Квантовая механика» и специального курса «Метод Монте-Карло в физике элементарных частиц».

Данный спецкурс формирует необходимую базу для научных исследований студентов и выполнения курсовых и дипломных работ.

В результате изучения специального курса «Статистическая обработка физической информации» студент должен

знать:

– способы, приёмы работы для процесса обработки физической информации.

уметь:

– применять системы аналитических вычислений для решения задач квантовой механики и квантовой теории поля

– использовать технические средства автоматизации эксперимента

– обрабатывать и анализировать полученные результаты, создавать математические модели и программные средства

иметь опыт:

– планирования и организации научного исследования, применения соответствующих экспериментальных и теоретических методов

– интерпретации результатов экспериментальных исследований.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Учебная программа спецкурса «Статистическая обработка физической информации» составлена в соответствии с учебным планом подготовки специалистов по специальности I-31040102 «Физика (производственная деятельность)» специализации I-310401-02 -17 «Компьютерное моделирование физических процессов».

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Лабораторная работа №1
Введение в MathCad

Цель работы:

Овладение некоторыми навыками работами в системе MathCad.

ХОД ВЫПОЛНЕНИЯ РАБОТЫ:

Упражнение1. Краткое знакомство с интерфейсом.

Ознакомьтесь, пожалуйста, с главными и необходимыми пунктами интерфейса системы MathCad.

Переднее панно системы MathCAD с активной позицией File главного меню содержит следующие пункты:

New **[Ctrl+ N]** (Создать) — открыть окно для нового документа;

Open... [Ctrl+ 0] (Открыть) — открыть существующий документ;

Save [F6] (Сохранить) — сохранить на диске текущий документ;

Save as... [Ctrl+ S] — сохранить на диске текущий документ под (Сохранить как) новым именем.

Export WorkSheets [Ctrl+ W] - экспортирование документа.

Insert [Ctrl+ I] - Вставка файла системы MathCad.

Close [Ctrl+ F4] (Закрыть) — закрыть документ.

Open Url... — установить модемную связь с Internet; (Установка Internet)

Get Notes... — установить связь с фирмой — разработчиком (Сотрудничество) системы для обеспечения совместной работы над документами;

Save Notes…-сохранить послания.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Mail... (Отправка) — отправить документ по электронной почте или по Internet.

Save Configuration [Ctrl+ V]…-сохранить текующие настройки.

Execute Configuration File [Ctrl+ E] - вызвать текущий файл настройки.

Page Setup... — установить левый и правый отступы (Параметры страницы) на странице;

Print Preview... (Просмотр)— предварительно просмотреть документ перед печатью;

Print... [Ctrl+ 0] (Печать) — распечатать документ.

Exit [Alt+ F4] (Выход) — выйти из среды MathCAD.

Ввод объектов (математических выражений, текстовых комментариев, графиков и др.) в текущее окно редактора производится по-разному.

Формы курсора

Курсор может иметь следующие формы.

\+ — крестообразный красный курсор (визир) служит для указания места для новых блоков (текстовых, формульных или графических). Курсор имеет такой вид только вне пространства блоков, т. е. на пу стом месте экрана, и может перемещаться клавишами управления курсором или устанавливаться мышью (для этого курсор мыши ставится в нужное место и нажимается ее левая клавиша), курсор в виде красной вертикальной черты (маркер ввода) служит для указания на отдельные элементы блоков, он обычно используется для ввода данных и заполнения шаблонов. В текстовых блоках используется для указания места вставки или удаления отдельных символов,

[или] — курсор в виде синих уголков разного размера, выделяющих отдельные части выражения или выражение целиком Вид курсора зависит от направления ввода Нажатие клавиши Ins или клавиш <— и —> перемещения курсора меняет направление ввода

Клавиши для выделения

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Для выполнения выделений используются указанные ниже клавиши:

вверх — превращает маркер в выделяющий уголок и расширяет его;

вниз — сужает выделяющий уголок;

-» — перемещает маркер и меняет вид уголка;

<- — перемещает маркер и меняет вид уголка;

Shift Т — выводит курсор из выражения в верхнюю часть свободного поля, делая его крестообразным;

Shift вниз — выводит курсор из выражения в нижнюю часть свободного поля;

Shift -» — выводит курсор из выражения в правую половину свободного по ля;

Shift <— — выводит курсор из выражения в левую половину свободного поля;

Space — заключает в рамку операнд, действует как несколько нажатий кла виши Т и выводит курсор из выделенного выражения;

Ins — перемещает срез рамки из правого верхнего угла в левый верх ний угол

Клавиши и их комбинации для управления редактированием

По мере освоения системы все большую помощь в редактировании оказывает знание функций управляющих клавиш и их комбинаций. Ниже дан перечень осуществляемых ими операций:

Tab — в тексте перемещает курсор на начало следующего слова, в

уравнении управляет выделением частей блока (в частности, выделяя выражения в скобках);

Shift+ Tab — в тексте перемещает курсор в начало очередного слова, в уравнении управляет выделением частей блока;

PgUp — перемещает курсор и вызывает скроллинг на пять строк вверх;

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

PgDn — перемещает курсор и вызывает скроллинг на пять строк вниз;

Ctrl+ PgUp — вызывает скроллинг на одно окно вверх;

Ctrl+ PgDn — вызывает скроллинг на одно окно вниз;

Home — устанавливает курсор в начало предшествующего блока, **Ctrl+ Home —** вызывает скроллинг с установкой курсора в начало документа;

Ctrl+ End — вызывает скроллинг с установкой курсора в конец документа.

Для демонстрации работы в системе MathCad мы ограничимся простым примером. Так, для задания переменной x=1 и вычисления значения функции sin (x) можно воспользоваться вводом с клавиатуры.

Нажимаемая клавиша	Изображение	Комментарий в окне
X	X	Ввод имени переменной
Shift + :	X: =	Ввод символа)исваивания: -
1	x: = 1	Ввод числовой константы 1
Enter		Фиксация ввода, скачок курсора
s i n(x)=	sin(x)	Ввод имени функции sin

Упражнение2. Простейшие арифметические вычисления в системе MathCad.

Используя алгоритм предыдущего примера, вычислите следующие выражения:

1. x=0, cos(x).
2. $x=30^0$, cos(x).
3. $x=45^0$, cos(x).
4. $x=60^0$, cos(x).
5. $x=270^0$,cos(x).

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

При тех же значениях аргумента, вычислите значения следующих функций: sin(x), tg(x), ctg(x).

Упражнение 3. Построение простейшей графической зависимости функции-аргумента.

Для построения простейшей графической зависимости в системе MathCad используется опция Graphics **[Ctrl+ G]** - Create X-Y Plot.

На экране появится шаблон графика, который необходимо будет подписать. После чего подтвердите свой ввод данных клавишей "Enter".

Для работы выполните следующее:

1. Введите следующее: x:=0..20 (символ ".." получается нажатием клавиши ";")
2. y(x):=sin(x)
3. Выберите опцию Graphics **[Ctrl+ G]** - Create X-Y Plot.
4. Подпишите график и нажмите "Enter".

Проделайте то же самое и для функций y(x)=cos(x), $y(x)=x^2$, y(x)=25*x-10, $y(x)=\frac{1}{x}$.

Упражнение 4 Данные файлового типа

Еще один важный тип данных системы MathCAD — файловые данные В сущности, это те же векторы и матрицы, но с элементами, которые могут записываться в виде файлов, имеющих свои имена. Файлы данных в системе представляет собою запись матриц в их естественной форме как последовательных текстовых файлов. Это простейший тип файлов, который легко обрабатывается в программах на различных языках программирования и может создаваться такими программами, благодаря чему возможен обмен данными между системой MathCAD и другими программами.

В ходе создания файла система считывает значения элементов векторов и матриц поэлементно (для матриц слева направо и сверху вниз) и по ходу считывания преобразует числовые значения элементов в их символьные эквиваленты,

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

использующие ASCII-коды цифр и символы, относящиеся к заданию чисел. Эти символьные значения и записываются в виде данных файлов.

Существует семь файловых операций, рассматриваемых ниже. Создаваемые или используемые ими файлы легко просмотреть любым текстовым редактором, воспринимающим тексты в виде ASCII-кодов. При считывании файлов система обеспечивает обратное преобразование символьных представлений значений элементов в их числовые значения.

Задание

1. Ознакомиться с операторами считывания и записи в файл (READ,WRITE и т.д.) на основе примеров имеющих в справочной системе MATHCAD.
2. Считать данные для своего dat-файла (номер вашей фамилии в журнале группы соответствует номеру dat-файла).
3. Записать данные результатов расчета упражнения 3 во внешний файл.

Лабораторная работа № 2

Построение графиков различных распределений. Точечные и интервальные оценки случайной величины.

Цель работы: Освоить методы обработки экспериментальных данных с помощью встроенных операторов системы MATHCAD

Для характеристики частоты появления различных значений случайной величины X (либо погрешностей приборов или результатов измерений с учетом ее систематической составляющей) теория вероятностей

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

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.

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

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

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

$p(x)$ называют кривой плотности распределения.

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

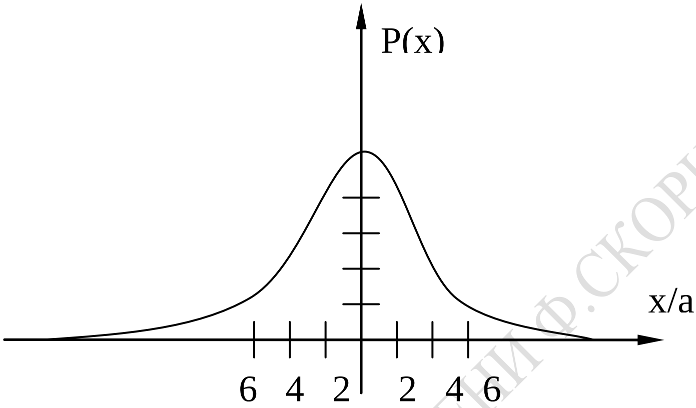
$$F(x) = \int\limits_{-\infty}^{\text{ч}} p(\xi)\mathrm{d}\xi \tag{4}$$

Примеры законов распределений.

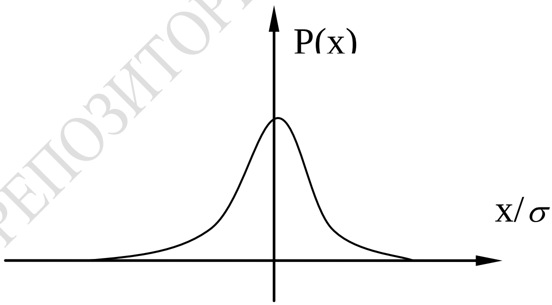
А) Распределение Коши:

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

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



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:

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

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

$$\nu_k = \int (x-\mu)^k p(x)\mathrm{d}x$$ (6) - центральный.

где $p(x)$ – плотность распределения вероятности величины.

Первый момент называют <u>математическим ожиданием</u>

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

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 <u>медиану</u>. 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

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

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Второй центральный момент называется <u>дисперсией</u> случайной величины

$$\nu_2 = D[x] = \int_{-\infty}^{\infty} p(x)(x-\mu_1)^2\,dx \qquad (9).$$

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

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

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

σ - с.к.о.

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

$$\varepsilon = \frac{\mu_4}{\sigma^4} \quad (13)$$ 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). $æ \in [0,1]$.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Задание

ЗАДАЧА 1.

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

Построить графики с помощью встроенных функций MathCad:

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;

г) плотности распределения хи-квадрат для 5 степеней свободы.

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 $\overline{x}$,

б) дисперсию S^2,

в) среднее квадратичное отклонение среднего арифметического $S_{\overline{x}}$,

г) 4-й момент и эксцесс ε,

д) коэффицент ассиметрии.

ЗАДАЧА 4. Проделать ЗАДАЧУ 3. с помощью встроенных функций MathCad.

Провести анализ и сравнение полученных данных. Сделать вывод.

ЗАДАЧА 5. С помощью встроенных функций MathCad найти интервальные оценки физической случайной величины для заданной выборки, где p=1- α, p=0.99, p=0.95, p=0.90 , т.е. найти:

а) доверительный интервал для мат. ожидания;

б) доверительный интервал для дисперсии;

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

в) сравнить доверительный интервал для мат. ожидания с интервалом, полученным с помощью $\bar{x}$ и $S_{\bar{x}}$, сделать вывод.

Лабораторная работа №3

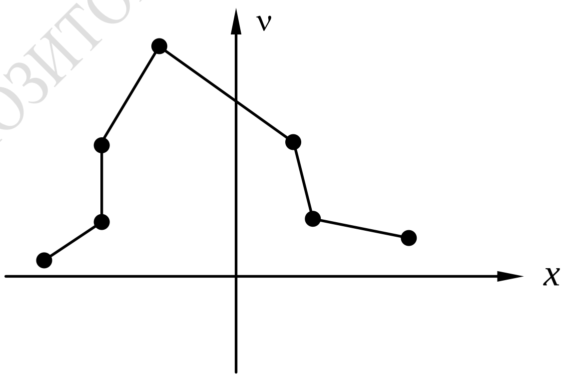
Построение гистограмм.

Цель работы: Овладеть способами построения гистограмм с помощью различных критериев выбора числа столбцов.

Краткая теория

Определения гистограммы, ее назначение.

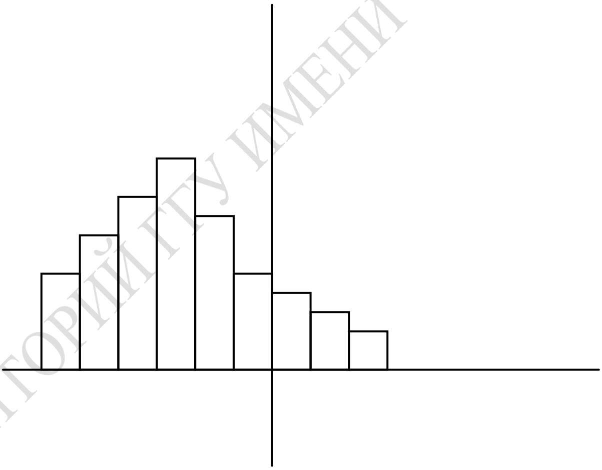
Полученные экспериментальные данные представляют, как правило, в виде таблиц. Полученные таблицы удобно представить графически. Используя частность n (отношение числа событий, попавших в данный интервал, к общему числу событий) и положение середины интервала можно построить полигон частот.



РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Это же распределение можно представить в виде гистограммы.

Для построения гистограммы необходимо над каждым отрезком оси абсцисс, соответствующим интервалу значений измеряемой величины, построить прямоугольник, площадь которого пропорциональна плотности или частоте попадания в этот интервал. Обычно выбирают интервалы одинаковой ширины, поэтому высота прямоугольников различна.



Цель обработки данных заключается в выявлении вида распределений случайных величин и оценки параметров установленного распределения.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Для определения оценок математического ожидания, с.к.о., эксцесса не требуется какого-либо группирования данных.

Для определения медианы, сгибов, использования критерия согласия Колмогорова-Смирнова или для обнаружения промахов, экспериментальные данные необходимо расположить в порядке возрастания, т.е. построить вариационный ряд (упорядоченную выборку).

Для определения формы распределения, для использования критериев согласия Пирсона и др., для сопоставления гипотез о форме распределения и т.д. простого упорядочения выборки уже недостаточно, а выборка должна быть представлена в виде гистограммы, состоящей из m столбцов с определенной протяженностью d соответствующих им интервалов.

Оптимальное число интервалов для получения гистограммы экспериментальных данных. (Или как построить гистограмму). Как выбрать m и d.

Общепринято делать интервалы одинаковыми. Хотя в дальнейшем увидим, что это условие необязательно.

Условие равновеликости интервалов удобно с практической точки зрения.

Во-первых, очевидно, что существует оптимальное число интервалов группирования, когда ступенчатая огибающая гистограммы наиболее близка к плавной кривой распределения.

К примеру, при группировании данных в большое число меньших интервалов, некоторые из них окажутся пустыми или малозначительными. Гистограмма будет отличаться от плавной кривой распределения вследствие изрезанности многими всплесками и провалами.

Т.е. 1-е требование:

Размер интервала (ячейки) должен быть достаточно широким для обеспечения хороших статистических свойств будущей гистограммы (достаточно большая пуассоновская

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

статистика, минимальные корреляции (связи) с соседними ячейками).

При слишком малом числе m интервалов, гистограмма отличается от действительной кривой распределения вследствие слишком крупной ступенчатости. Из-за чего будут потеряны характерные особенности. Например, если взять m=1, т.е. d равно размаху экспериментальных данных, то любое распределение сводится к равномерному, а если m=3, то любое колоколообразное распределение сведется к треугольному.

В примере для обработки линейчатых спектров большая ячейка может привести к потере спектральной линии

Т.о. <u>2-е требование</u>:

Размер ячейки должен быть достаточно узким для того, чтобы прорисовывалась «тонкая структура» исследуемой величины.

Как видим, требования являются противоречивыми.

Укрупнение интервалов группирования является методом «фильтрации различных случайных выбросов и провалов», но слишком протяженные интервалы сглаживают особенности искомого закона распределения.

Таким образом, задача выбора оптимального числа интервалов при построении гистограммы – это задача оптимальной фильтрации, а оптимальным числом m интервалов является максимальное возможное сглаживание случайных флуктуаций данных, которое сочетается с минимальным искажением от сглаживания самой кривой искомого распределения.

<u>Рекомендации по выбору m.</u>

I группа: эвристические критерии (без доказательства).

1) Формула Старджеса

$$m = \log_2 n + 1 \approx 3.3 \lg n + 1 \qquad (1)$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

2) Формула Брукса и Каррузера

$$m = 5\lg n \qquad (2)$$

3)
$$m = \sqrt{n} \qquad (3)$$

Эти три формулы являются часто встречающимися в литературе по математической статистике.

II группа: с использованием χ^2.

В ней используется рассмотрение интервалов не с равной длинной, а с равной вероятностью в соответствии с принимаемой моделью (т.е. предположением о законе распределения).

Число интервалов с равной вероятностью, которые мы обозначили как K, отличаются от числа m с равной длиной d (в несколько раз).

Г. Манн и А. Вальд установили, что при $n \to \infty$ оптимальное число K равновероятных интервалов задается соотношением

$$K = 4\sqrt[5]{2}\left(\frac{n}{Z_\alpha}\right)^{0.4} \qquad (4)$$

Z_α - квантиль нормального распределения, соответствующий вероятности $P = 1 - \alpha$, где α - принятый уровень значимости.

$$Z_\alpha = (2\pi)^{-1}\int_{-\infty}^{z_\alpha} e^{-\frac{z^2}{2}}dz = 1 - \alpha$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

На практике часто берут $\alpha = 0.1$, тогда

$$K \approx 1.9n^{0.4} \tag{5}$$

III группа.

Поскольку для K интервалы получаются не равной длины, то это приводит к ряду неудобств при построении гистограмм, но зато при этом мы неявно закладываем при использовании χ^2 выбор K в зависимости от формы распределения.

III группа рекомендаций «устраняет» недостаток II группы возвращаясь к интервалам m с равной длиной d, но при этом и учитывает, в отличие от I группы, форму распределения (форма характеризуется эксцессом ε и контрэксцессом æ.

Примером является соотношение, полученное в работе Алексеевой:

$$m = \frac{4}{\kappa} \lg \frac{n}{10} \tag{6.6}$$

Трудность использования III группы состоит в том, что число интервалов часто приходиться выбирать прежде, чем будут найдены оценки κ , $\overline{x}$ и т.д.

Эту трудность обходят следующим образом: наиболее часто встречаются распределения с ε от 1.8 до 6 (от равномерного до Лапласа, включая нормальное ε =3). Для этих граничных точек имеем:

$m_{min}=0.55n^{0.4}$ и $m_{max}=1.25n^{0.4}$ (6.9)

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Искомое m сложно выбрать близким к этому интервалу, при этом m лучше выбрать нечетным, т.к. при четном m для островершинных распределений в центре гистограммы оказывается два столбца равных по высоте и середина распределения принудительно утолщается.

<u>«Практические» рекомендации:</u>

1) Для практического определения числа интервалов воспользоваться формулой (6.9), выбрав при этом m нечетным.
2) Так как крайние точки могут располагаться несимметрично, то ширина d столбца гистограммы определяется по отклонению от центра ΔX_m наиболее удаленной точки.

$$d = \frac{2\Delta X_m}{m}$$

3) При этом полученное значение d необходимо округлять в большую сторону, чтобы крайняя точка не оказалась за пределами крайнего столбца.
4) Величину d при этом удобно выбирать так, чтобы она делилась на 2 так, чтобы потом центральный столбец можно было бы поделить пополам для уточнения
центра распределения.

Задание

ЗАДАЧА 1. Построить полигон частот для данных выборок. Число интервалов m=10.

ЗАДАЧА 2. Построить соответствующие задаче 1 гистограммы.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Примечание. Построение гистограмм провести с помощью встроенных функций MathCAD 2 различными способами (не меньше).

ЗАДАЧА 3. Построить гистограммы, используя эвристические критерии и критерии с учетом формы распределения. Сравнить полученные гистограммы с гистограммами, полученными в задаче 2 и сделать выводы.

Лабораторная работа № 4

Промахи и методы их исключения.

Цель работы: Овладеть методами исключения промахов из значений случайной величины.

Краткая теория
Промахи и методы их исключения.

Одним из условий правомерности статистической выборки является требование ее однородности, т.е. принадлежности всех ее членов к одной и той же генеральной совокупности.

Однако на практике это требование очень часто нарушается. И, если скажем, при обработке вручную еще можно вспомнить как (при каких условиях) были получены «подозрительные» данные, то при автоматической обработке данных необходимы методы исключения «чужих» для данной выборки результатов.

Отсчеты, резко отклоняющиеся по своим значениям от большинства других отсчетов принято называть промахами и исключать их из выборки.

Но особую неприятность доставляют отсчеты, которые и не входят в компактную группу отсчетов, но и не удалены от

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

нее на значительное расстояние. Такой отсчет называют предполагаемым промахом.

В экспериментальной практике исследователи просто отбрасывали крайние, «слишком удаленные от центра наблюдения». Это получило название цензурирования выборки.

Однако для принятия решения необходимы какие-либо формальные критерии.

Простейший метод заключается в использовании «правила 3σ», когда по выборке с удаленными отчетами (предполагаемыми промахами) вычисляется оценка σ и граница $|X_{гр}|=3\sigma$, а все $|x_i| \geq 3\sigma$.

«Правило 3σ» обосновано на неравенстве Чебышева:

$$P\{|x_i - \overline{x}| \geq a\} \leq \frac{\sigma^2}{a^2} \quad (a>0)$$

Если все $x_i \geq 0$, то

$$P(x \geq a) \leq \frac{\overline{x}}{a} \qquad (a>0)$$

Если x имеет одномодальное распределение (непрерывное), то справедлива более сильная оценка

$$P_{пов}\{|x_i - \overline{x}| \geq a\} \leq \frac{4}{9} \frac{1+S^2}{\left(\frac{a}{\sigma} - |S|\right)^2},$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

где S - пирсоновская мера асимметрии (для распределений симметричных относительно моды S=0).

$$S = \frac{\bar{x} - \xi}{\sigma},$$

где ξ есть максимальная вероятность.

$$S = \frac{\gamma_1(\gamma_2 + 6)}{2(5\gamma_2 - 6\gamma_1^2 + 6)}$$

$$\gamma_1 = \frac{m_3}{m_2^{3/2}} \qquad \gamma_2 = \frac{m_4}{m_2^2} - 3$$

$$m_j = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^j$$

Если положить а=3σ, то имеем

$$P_{пов}\{|x_i - x| \geq 3\sigma\} \leq \frac{1}{9} \approx 0.11$$

или для одномодальных симметричных:

$$P_{пов}\{|x_i - x| \geq 3\sigma\} \leq \frac{4}{9}\frac{1}{9} = \frac{4}{81} \approx 0{,}05,$$

т.е. для произвольного распределения вероятность, что $x_i \geq \bar{x}$ на 3σ, составляет 11%, а для одномодальных симметричных 5%.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Однако правило «3σ» хорошо для нормального распределения.

Действительно, при n=100 появление $|x_i| \geq 3\sigma$ можно считать промахом, то для равномерного промахом можно считать уже $|x_i| = 1.8\sigma$, а для распределения Лапласа $|x_i| = 3\sigma$ есть отсчет, принадлежащий данной выборке.

На этом примере видно, что граница цезурирования зависит не только от объема выборки n, но также и от формы распределения.

Зависимость от n полуколичественно можно оценить из условия, границы цензурирования должны осекать в среднем менее одной точки, тогда назначение границ с уровнем значимости $\alpha = 1 - P$, где P=n/n+1 дает зависимость от n.

Однако $P = f(t)$, где $t\sigma = X_{гр}$ различна для разных законов распределения.

$$P = f(\varepsilon, t)$$

Данные зависимости

$$t = t(\bar{\varepsilon}, n) \qquad\qquad t\sigma = X_{гр}$$

постройте для некоторых классов распределений.

В заключение отметим, что все оценки $\bar{X}$, σ и т.д. должны пересчитываться после цензурирования выборки.

Задание

Задача 1. Сгенерировать равномерное (a=5, b=№ своего варианта) и нормальное (a=1, σ=№ своего варианта)

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

распределения при помощи встроенных функций ДЛЯ n=100, 200.

Задача 2. С помощью правила **«3σ»** провести операцию цензурирования полученных выборок.

Задача 3. Провести операцию цензурирования полученных выборок с помощью критериев, учитывающих форму распределения и объем выборки.

Примечание. Отобразить графически операцию цензурирования выборок для полигона частот (отобразить состояние до и после цензурирования).

Задача 4. Построить гистограммы до цензурирования и после. Сравнить результаты. Сделать выводы.

Лабораторная работа № 5

Идентификация форм распределения экспериментальных данных.

Цель работы: Овладеть методами идентификация форм распределения

Краткая теория

Идентификация формы распределения экспериментальных данных.

Критерии согласия.

Наиболее распространенным является критерий согласия Пирсона (χ^2 - критерий).

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Суть критерия состоит в вычислении величины

$$\chi^2_{эк} = \sum_{j=1}^{m} \frac{\left(\nu_j - p_j\right)^2}{p_j} \qquad (1),$$

где m – число интервалов, ν_j - частота попадания в j-й интервал, p_j - вероятности в том же j-м столбце, рассчитанном по заданной модели.

Вычисленная $\chi^2_{экспер}$ сравнивается с $\chi^2(\kappa,\alpha)$ с числом степеней свободы $\kappa = m - 1 - r$, где r – число определяемых по статистике параметров, необходимых для совмещения модели и гистограммы (как правило, r=2, $\overline{X}$ и σ).

Если $\chi^2_{экс} \le \chi^2_{\alpha,\kappa}$, то гипотеза о распределении при данном уровне значимости α не противоречит модели, если $\chi^2 > \chi^2_{\alpha,\kappa}$, то гипотеза отклоняется.

Как правило, χ^2 дает лишь отрицательный ответ, однако на практике используется и как для принятия положительного решения.

Критерий Колмогорова-Смирнова.

В этом критерии используется максимальное значение модуля разности между экспериментальным и теоретическим распределением интегральных функций распределений. Математически это означает:

$$D = \max | F_{экспер}(x) - F_{теор}(x) | / n$$

и сравнивается с граничным значением вероятности

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

$$P \geq 2\exp\left(-2nD^2\right) = 2\exp(-2\Delta^2 / n)$$

$$\Delta = | F_{экспер} - F_{теор} |$$

Здесь P – искомая вероятность того, что искомая выборка может соответствовать полученной модели.

Можно построить и доверительный интервал следующего типа:

$$P_{дов}[F_{экспер} - d_\alpha \leq F_{теор} < F_{экспер} + d_\alpha] = 1 - \alpha,$$

где критические значения $d\alpha$

$$d_\alpha \approx \sqrt{-0.5\ln\left(\frac{\alpha}{2}\right)/n} \qquad \text{(n>35)}$$

Здесь α - уровень значимости.

Доверительный интервал представляет собой полосу с шириной $\pm d_\alpha$ около выбранного нами (измеренного) $F_{экспер}$ и с вероятностью $1-\alpha$ истинная функция $F_{теор}(x)$ лежит внутри полосы.

Преимуществом перед χ^2 состоит в том, что не требуется группирование данных в интервалы.

<u>Критерий ω^2 (критерий Мизеса).</u>

Этот критерий также использует не сгруппированные данные. Рассчитывается величина

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

$$\omega^2 = \int_{-\infty}^{\infty} \left[F_{экспер} - F_{теор} \right]^2 dF_{теор}$$

$$dF_{теор} = f_{теор}(x)dx$$

В итоге для

$$F_{экспер} = \begin{cases} 0, & x<x_1 \\ k/n, & x_k \le x < x_{k+1}, \; k=1\ldots n-1 \\ 1, & x>x_n \end{cases}$$

Получаем

$$\omega^2 = \frac{1}{n}\{\frac{1}{12n} + \sum_{k=1}^{n} \left[F(x_k) - \frac{2k-1}{2n} \right]^2$$

Затем сравнивают $n\omega^2$ с $\left(n\omega^2\right)_{крит}$ с заданным уровнем значимости α.

α	$\left(n\omega\right)_{крит}$
0.5	0.1184
0.4	0.1467
0.3	0.1843

α	$\left(n\omega\right)_{крит}$
0.05	0.4614
0.03	0.5489
0.02	0.6198

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

0.2	0.2412
0.1	0.3473

0.01	0.7435
0.001	1.1679

(Эта таблица для n>40).

Здесь более полно используется информация о выборке.

<u>Замечания для критериев согласия</u>.

При использовании критериев согласия обычно оговаривается, что их применение корректно лишь при достаточно «больших» выборках и, как правило, указывается n>200.

Располагая соотношением для выбора m, увидим это более наглядно

$$m = \frac{\varepsilon + 1.5}{6} n^{0.4} \Rightarrow n = \left[\frac{6m}{\varepsilon + 1.5} \right]^{2.5}$$

При использовании χ^2 желательно иметь число отклонений (1-3, как обычно), а хотя бы 7-9-11.

Тогда для нормального распределения ($\varepsilon = 3$) при m=7-11 имеем n=170-800 отсчетов; для равномерного $(\varepsilon = 1.8)$ при m=7-11 необходимо уже n=600-2700 отсчетов. Аналогично оценим объем выборки с использованием критерия Колмогорова. Так для

$$\alpha = 0.05 \qquad d_{\alpha=0.05} = \frac{0.61}{\sqrt{n}}$$

$$n = 100 \qquad d_{\alpha=0.05} = 0.061$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Но объем 500-2500 для экспериментаторов практически не достижим.

Задание

Задача 1. С помощью критерия χ^2 определить является ли случайная физическая величина нормальным распределением т.е. определить с какой вероятностью данная выборка будет нормальным распределением.

Задача 2. Проделать задачу 1 с помощью критерия Колмогорова-Смирнова.

Задача 3. Построить графики и сделать визуальное сравнение выбранной модели и экспериментальной кривой. Сделать выводы.

Лабораторная работа № 6

Преобразование Фурье.

Цель работы: Овладеть методами построения частотных характеристик случайных величин с помощью быстрого преобразования Фурье.

Краткая теория

Преобразование Фурье.

Пусть M – число точек гистограммы h(i), тогда дискретное преобразование Фурье для нее равно:

$$g(k)=\frac{1}{M}\sum_{i=1}^{M}h(i)\left[\exp\left(-2\pi(i-1)j\frac{k}{M}\right)\right]$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

$k=0,..M\text{-}1 \qquad j^2=-1$

Обратное распределение Фурье выглядит следующим образом:

$$h(i)=\sum_{k=0}^{M-1} g(k)\exp\left(j2\pi ib/M\right), \qquad i=1..M$$

Как и в непрерывном случае преобразование Фурье показывает из каких частот «построена» гистограмма h(i).

Число частот для гистограмм ограничено (в силу периодичности $\exp(jn)$ и оно тем меньше, чем меньше число каналов (интервалов), представляющих спектр, т.е. чем шире ячейка гистограммы.

Численное осуществление преобразования Фурье и обратного преобразования Фурье проводится с помощью алгоритмов быстрого преобразования Фурье.

Другой аспект использования преобразования Фурье – поиск периодичностей в гистограмме *h(i)*. Если *h(i)* есть положение периодических функций с периодами T_l, $l=1..N$, то *g(k)* будет иметь *L* пиков в точках, соответствующих частотам периодических функций.

Рассмотрим идею быстрого преобразования Фурье.

Стандартное преобразование Фурье можно записать в виде

$$g(k)=\sum_{i=0}^{M-1}\frac{h(i)}{M}W(k_i),\; k=0..M-1,\; i=1..M$$

$$W(k_i)=\exp\left[-j2\pi\frac{ki}{M}\right]$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Для расчета этого ряда необходимо M^2 операций умножения и сложения комплексных чисел (одна операция (комплексная) эквивалентна четырем операциям сложения и умножения действительных чисел).

Идея быстрого преобразования Фурье основывается на представлении величины M в виде произведения (отличных от единицы) сомножителей и в выполнении обычного преобразования Фурье для более коротких последовательностей, число членов в которых определяется соответствующими сомножителями.

Если M можно представить в виде произведения p целых и больших единицы чисел:

$$M = \prod_{i-1}^{p} r_i = r_1 r_2 .. r_p ,$$

последовательность $g(k)$ может быть найдена путем расчета суммы p слагаемых.

$\frac{N}{r_1}$ - преобразований Фурье ($4r_1^2$ операций с действительными числами).

$\frac{N}{r_2}$ - преобразований Фурье ($4r_2^2$ операций с действительными числами).

$\frac{N}{r_p}$ - преобразований Фурье, которые требуют $4r_p^2$ операций.

Всего

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

$$4\left(\frac{N}{r_p}4r_p^2+\frac{N}{r_1}4r_1^2\right)=4N\sum_{i=1}^{p}r_1 \ .$$

В результате по сравнению с обычным коэффициентом ускорения вычислений (к.у.в.)

$$к.у.в.=\frac{M^2}{4M\sum_{i=1}^{p}r_i}=\frac{M}{4\sum_{i=1}^{p}r_i}$$

Если

$$M=2^p=r_1..r_2\Rightarrow 2\log_2 M=2p=\sum_{i=1}^{p}r_i=2p$$

$$\sum_{i=1}^{p}r_i=2p$$

$$к.у.в.=\frac{M^2}{4M2p}=\frac{M}{8p}$$

При $M=2^p$ все значения $W(k_i)$ равны 1 либо –1, то есть все операции умножения заменяются сложением или вычитанием. Т.о.

$$к.у.в.=\frac{N}{4p}$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

Таким образом, быстрое преобразование Фурье сводится к следующим операциям.

Вычисляется сумма элементов, которую обозначим как

$$\mathrm{A}_{\nu}\left(k_0, k_1..k_{\nu-1}; i_0, i_1..i_{p-\nu-1}\right)$$

1) $$\mathrm{A}_1\left(л_0, n_0, n_1..n_{p-\nu-1}\right) = \sum_{n_{p-1}=0}^{1} \frac{h(i)}{M} \exp\left(-j\pi k_0 i_{p-1}\right)$$

2) $$\mathrm{A}_{\nu}\left(k_0..k_{\nu-1}, i_0..i_{p-\nu-1}\right) = \sum_{n_{p-\nu}=0}^{1} \mathrm{A}_{\nu-1}\left(k_0, k_1..k_{\nu-2}, n_0..n_{p-\nu}\right) \mathrm{T}\left(k_0, k_1..k_{p-2}\right) \exp\left(-j\pi k_{\nu-1} \frac{i_{p-\nu}}{M}\right)$$

Наконец,

$$g(k) = \mathrm{A}_p\left(k_0..k_{p-1}\right) = \sum_{n_0=0}^{1} \mathrm{A}_{p-1}\left(k_0, k_1..k_{p-2}, n_0\right) \mathrm{T}\left(k_0, k_1..k_{p-2}\right) \exp\left(-j\pi k_{p-1}, \frac{i_0}{M}\right)$$

Применим быстрое преобразование Фурье к действительным последовательностям

а) Если $x(n)$ и $y(n)$, *n=0,1..M-1* действительной последовательности $z(n) = x(n) + jy(n)$. Фурье-образ $z(n)$ Z(k)

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

$$Z(k)=\sum_{n=0}^{M-1} z(n)\exp\left[-j\frac{2\pi kn}{M}\right] \quad k=0,1..M-1$$

можно рассчитать по БПФ. Тогда Фурье-образ X(k) и Y(k) рассчитывается

$$X(k)=\frac{Z(k)+Z^*(N-k)}{2}$$

$$Y(k)=\frac{Z(k)-Z^*(N-k)}{2j}$$

$$Z^*(N-k)=\sum_{n=0}^{N-1}[x(n)-jy(n)]\exp\left[-j\frac{2\pi nk}{M}\right]$$

Можно и таким приемом: последовательность X(n), где n=0..2N-1, разбивают на две последовательности $\upsilon(n)$ и $V(2n+1)$, где n=0,1,2..N-1. Затем находим

$$Z(n)=\upsilon(2n)+jV(2n+1)$$

Затем необходимы $\upsilon(k)$ и $V(k)$, а фурье-образ

$$\mathrm{X}(k)=\upsilon(k)+V(k)\exp\left[-j\frac{\pi}{N}\right]$$

$$\mathrm{X}(k+N)=\upsilon(k)-V(k)\exp\left[-j\frac{\pi}{N}\right],$$

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ

где $k=0,1,2..N-1$.

Задание

Задача 1. С помощью встроенных функций для прямого преобразования Фурье построить «частотное» распределения для заданной физической величины. (Выборки брать из лаб.р.№5).

Задача 2. По полученному частотному распределению построить обратное преобразование Фурье («временное»). Построить гистограмму.

Задача 3. Построить исходную гистограмму. Сравнить с гистограммой из задачи 2. Оформить лабораторную работу. Сделать выводы.

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф. СКОРИНЫ

Андреев Виктор Васильевич

СТАТИСТИЧЕСКАЯ ОБРАБОТКА
ФИЗИЧЕСКОЙ ИНФОРМАЦИИ

ПРАКТИЧЕСКОЕ ПОСОБИЕ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

В авторской редакции

Подписано в печать .2005 г. () Формат 60x84 1/16. Бумага писчая No 1. Отпечатано на ризографе. Гарнитура «Таймс».
Усл.п.л. 3, 4. Уч.-изд. л. 2, 1 .
Тираж 25 экз.

Учреждение образования
«Гомельский государственный университет
имени Франциска Скорины»
246019, г. Гомель, ул. Советская, 104

Отпечатано в учреждении образования
«Гомельский государственный университет
имени Франциска Скорины»
246019, г. Гомель, ул. Советская, 104

РЕПОЗИТОРИЙ ГГУ ИМЕНИ Ф.СКОРИНЫ