

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное учреждение науки
«Институт ядерной физики им. Г.И. Будкера СО РАН»

УТВЕРЖДАЮ
директор ИЯФ СО РАН,

академик _____ П.В.Логачев

«_____» _____ 2018 г.

МЕТОДЫ АНАЛИЗА ФИЗИЧЕСКИХ ИЗМЕРЕНИЙ

Рабочая программа дисциплины

Направление подготовки

03.06.01 Физика и астрономия (уровень подготовки кадров высшей квалификации)

Общая трудоемкость дисциплины: 3 зачетные единицы – **108** часа.

Форма промежуточной аттестации: зачет

Виды деятельности:

Лекции	34	контактная работа обучающихся с преподавателем	68
Семинарские занятия	34		
Самостоятельная работа	36	занятия в активной и интерактивной форме	68
Консультации			
Зачеты	4	Экзамены	

Новосибирск- 2018

Программа курса «Современная физика высоких энергий» составлена в соответствии с требованиями федерального государственного стандарта высшего образования по направлению подготовки 03.06.01 Физика и астрономия (уровень подготовки кадров высшей квалификации).

Цели курса – ознакомление аспирантов с современными методами анализа результатов измерений, получаемых в физических экспериментах. Первая часть курса посвящена повторению и углублению знаний, полученных в курсах теории вероятностей и математической статистики. Вторая часть курса посвящена применению методов интеллектуального и многопараметрического анализа данных. В третьей части курса рассматриваются отдельные задачи, часто возникающие при анализе экспериментальных данных. В рамках практических занятий студенты получают возможность использовать полученные знания для решения индивидуально подобранных задач.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, семинарские занятия, самостоятельная работа студента.

Программой дисциплины предусмотрены следующие виды контроля:

Текущий контроль успеваемости: сдача заданий для самостоятельного решения.

Промежуточная аттестация: зачет по курсу.

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

Составитель:

к.ф.-м.н. Логашенко И.Б.

Рабочая программа

Содержание

1. Цели освоения дисциплины	4
2. Место дисциплины в структуре программы обучения	4
3. Компетенции обучающегося, формируемые в результате освоения дисциплины	5
4. Структура и содержание дисциплины	5
5. Образовательные технологии	9
6. Учебно-методическое обеспечение самостоятельной работы студентов и аспирантов	10
7. Фонд оценочных средств для проведения аттестации по итогам освоения дисциплины:	10
8. Учебно-методическое и информационное обеспечение дисциплины	13
9. Материально-техническое обеспечение дисциплины	14

«Методы анализа физических измерений»

Рабочая программа дисциплины

1. Цели освоения дисциплины

Целью освоения дисциплины «Методы анализа физических измерений» является практическое ознакомление студентов с современными методами анализа результатов измерений, получаемых в физических экспериментах. Для достижения поставленной цели ставятся следующие задачи:

1) изучение практического применения методов теории вероятностей, математической статистики, методов Монте-Карло для анализа данных экспериментов, оценки погрешностей, классификации данных;

2) изучение специализированного программного обеспечения Geant4 и Root.

В современных экспериментах зачастую изучаются слабые эффект или редкие процессы. Это значительно усложняет анализ данных, т.к. исследуемый эффект может сильно искажаться наличием фона, наличием корреляций и т.п. Правильная оценка величины эффекта, оценка погрешностей и достоверности результата в таких случаях требует глубокого понимания математической статистики и применения специальных подходов.

В ходе современных физических экспериментах часто возникают большие объемы информации. Характерный объем данных, накопленный в типичном эксперименте, обычно составляет 10-100 терабайт. На Большом Адронном Коллайдере каждый год планируется накапливать более 10 петабайт (10000 терабайт) в течение 10 лет. Такие объемы данных, а также их возросшая сложность, связанная с большим количеством регистрирующих систем, требуют использования автоматизированных методов анализа данных.

Обсуждаемые методы анализа данных применяются не только при анализе данных физического эксперимента, но и в других областях знаний, таких как экономика, биология и др.

Лекционная часть курса состоит из трех разделов. Первый раздел посвящен повторению и углублению знаний, полученных в курсах теории вероятностей и математической статистики. Подробно обсуждаются практические подходы к оценке параметров с помощью методов максимального правдоподобия и наименьших квадратов как в типичных случаях, так и в ситуациях, усложненных наличием фона, отсутствием достаточной экспериментальной статистики, наличием корреляций между измеряемыми величинами и т.п. Обсуждается применение байесовских методов при анализе данных физического эксперимента. Второй раздел курса посвящен применению методов интеллектуального анализа данных. Подробно обсуждаются такие методики, как нейронные сети, усиленные деревья принятия решения, линейный дискриминантный анализ и т.п., и их применение для решения задач классификации (отделения сигнала от фона), кластеризации, регрессии. В третьем разделе курса рассматриваются отдельные задачи, часто возникающие при анализе экспериментальных данных, такие, как деконволюция и анализ временных рядов.

В практической части курса студенты получают возможность на практике применить полученные знания. Первые несколько занятий посвящены ознакомлению с программным обеспечением, используемым для моделирования эксперимента и для анализа данных. В дальнейшем студенты должны выполнить несколько обязательных заданий, отражающих основные этапы анализа данных любого эксперимента: использование методов Монте-Карло, моделирование физических процессов, оценка параметров распределений, отделение сигнала от фона и т.п.

2. Место дисциплины в структуре программы обучения

Дисциплина «Методы анализа физических измерений» относится к вариативной части блока 1 «Дисциплины (модули)». Дисциплина является элективной (дисциплиной по выбору). Аспиранты, приступающие к изучению этой дисциплины, должны иметь общую базовую подготовку в рамках программы 5-6 лет обучения в ВУЗе, в том числе:

- иметь общее представление об организации физического эксперимента,
- знать основы объектно-ориентированного программирования,

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина нацелена на формирование у выпускника следующих компетенций:

- Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях (УК-1);
- Готовность участвовать в работе российских и международных исследовательских коллективов по решению научных и научно-образовательных задач (УК-3);
- Способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий (ОПК-1);
- Способность построения теоретических моделей физических явлений и процессов для решения научных и практических задач (в соответствии с направленностью подготовки) (ПК-1);
- Способность к получению, критическому осмыслению и реферативному изложению научных результатов в области физики (в соответствии с направленностью подготовки) (ПК-2);

В результате освоения дисциплины обучающийся должен:

- **Знать** теоретические положения математической статистики и теории вероятностей, лежащие в основе изучаемых методов анализа данных: методы оценки параметров распределений, методы максимального правдоподобия и наименьших квадратов, построение критериев согласия, основы теории проверки гипотез, основы теории принятия решений, байесовский подход к оценке вероятностей; знать основные алгоритмы многомерного анализа данных, в частности, методы построения функций правдоподобия, нейронных сетей, деревьев принятия решений;
- **Уметь** использовать методы Монте-Карло для моделирования эксперимента и оценки погрешностей; оценивать параметры распределений при наличии корреляций и фона; применять методы максимального правдоподобия и наименьших квадратов; применять методы многомерного анализа данных; применять комплексные алгоритмы при анализе больших массивов данных; решать типичные задачи, возникающие при анализе данных современного физического эксперимента;
- **Владеть** программным инструментарием для моделирования и анализа данных физическо-го эксперимента (пакетами ROOT и GEANT4).

4. Структура и содержание дисциплины

Дисциплина «Методы анализа экспериментальных данных» представляет собой полугодовой курс, читаемый в аспирантуре ИЯФ СО РАН. Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

№ п/п	Раздел дисциплины	Неделя семестра	Виды учебных занятий, включая самостоятельную работу студентов и трудоемкость (в часах)				
			Всего	Аудиторные часы		Сам. работа в течение семестра (не включая период сессии)	Зачет
				Лекции (кол-во часов)	Семинары		
1	Повторение основных положений теории вероятностей и математической статистики: случайные величины; дискретные и непрерывные распределения; параметры	1-2	12	4	4	4	

	распределений: среднее значение, дисперсия, моменты; ковариационная матрица, коэффициент корреляции; преобразование распределения при замене переменных; основные распределения и их параметры: биномиальное, Пуассона, равномерное, нормальное, χ^2 ; центральная предельная теорема.						
2	Метод Монте-Карло, область применения. Интегрирование методом Монте-Карло. Алгоритмы генерации случайных чисел: метод Неймана, метод трансформации, комбинированный. Алгоритм генерации нормально-распределенной величины. Алгоритм Метрополиса.	3	6	2	2	2	
3	Оценка параметров распределений по ограниченной выборке. Свойства оценок: состоятельность, смещение, эффективность, робастность. Понятие информации Фишера и неравенство Рао-Крамера. Способы построения оценок, метод моментов. Способы построения несмещенной оценки, робастной оценки.	4	6	2	2	2	
4	Метод максимального правдоподобия. Оценка погрешностей и построение доверительных интервалов в методе максимального правдоподобия. Примеры использования метода максимального правдоподобия для аппроксимации гистограммы, определения времени жизни, оценки дисперсии.	5	6	2	2	2	
5	Метод наименьших квадратов. Оценка погрешностей в методе наименьших квадратов. Пример использования метода наименьших квадратов для аппроксимации гистограмм. Линейная подгонка методом	6	6	2	2	2	

	наименьших квадратов.						
6	Критерий согласия и способы его построения. Критерий χ^2 . Оценка качества аппроксимации в методе максимального правдоподобия. Другие критерии согласия: проверка последовательностей, критерий Колмогорова-Смирнова.	7	6	2	2	2	
7	Байесовский подход к оценке вероятностей. Теорема Байеса. Формулировка теоремы Байеса для непрерывных распределений. Применение теоремы Байеса для оценки погрешностей. Связь теоремы Байеса и метода максимального правдоподобия. Примеры применения теоремы Байеса: определение эффективности, оценка верхнего предела при близости измеренного значения к границе интервала возможных значений, оценка уровня сигнала при наличии фона. Понятие Байесовских сетей.	8	6	2	2	2	
8	Нейронные сети. Перцептрон. Обучение перцептрона, алгоритм обратного распространения ошибок. Глобальные методы оптимизации. Радиально-базисные сети. Задача кластеризации и сеть Кохонена. Применение нейронных сетей для классификации данных.	9-10	12	4	4	4	
9	Задача разделения сигнала и фона. Критерий разделения, мощность и значимость критерия. Методы сравнения критериев. Лемма Неймана-Пирсона.	11-12	12	4	4	4	
10	Практические методы построения критериев разделения: факторизация функции правдоподобия; линейный дискриминантный анализ Фишера; нейронные сети; деревья принятия решений; методы, основанные на подсчете числа событий.	13-14	12	4	4	4	

	Метод главных компонент.						
11	Задача обратной свертки (деконволюции). Постановка задачи. Методы получения результатов без обратной свертки. Прямое решение задачи. Регуляризация. Регуляризация Тихонова. Метод максимальной энтропии. Примеры.	15-16	12	4	4	4	
12	Фурье- анализ. Дискретное преобразование Фурье. Быстрое преобразование Фурье. Вэйвлетный анализ.	17	8	2	2	4	
17	Зачет	18	4				4
18	Итого		108	34	34	36	4

Программа курса лекций

1. Случайные величины. Дискретные и непрерывные распределения. Параметры распределений: среднее значение, дисперсия, моменты. Ковариационная матрица, коэффициент корреляции. Преобразование распределения при замене переменных. Основные распределения и их параметры: биномиальное, Пуассона, равномерное, нормальное, χ^2 . Центральная предельная теорема.
2. Метод Монте-Карло. Интегрирование методом Монте-Карло. Алгоритмы генерации случайных чисел: метод Неймана, метод трансформации, комбинированный. Алгоритм генерации нормально-распределенной величины. Алгоритм Метрополиса.
3. Оценка параметров распределений по ограниченной выборке. Точечные и интервальные оценки. Свойства оценок: состоятельность, смещение, эффективность, робастность (устойчивость). Понятие информации Фишера и неравенство Рао-Крамера. Способы построения оценок, метод моментов. Способы построения несмещенной оценки, робастной оценки.
4. Метод максимального правдоподобия. Оценка погрешностей и построение доверительных интервалов в методе максимального правдоподобия. Примеры использования метода максимального правдоподобия для аппроксимации гистограммы, определения времени жизни, оценки дисперсии.
5. Метод наименьших квадратов. Оценка погрешностей в методе наименьших квадратов. Метод наименьших квадратов в линейном приближении. Пример использования метода наименьших квадратов для аппроксимации гистограмм.
6. Критерий согласия и способы его построения. Критерий χ^2 . Оценка качества аппроксимации в методе максимального правдоподобия. Другие критерии согласия: проверка последовательностей, критерий Колмогорова-Смирнова.
7. Байесовский подход к оценке вероятностей. Теорема Байеса. Формулировка теоремы Байеса для непрерывных распределений. Применение теоремы Байеса для оценки погрешностей. Связь теоремы Байеса и метода максимального правдоподобия. Примеры применения теоремы Байеса: определение эффективности, оценка верхнего предела при близости измеренного значения к границе интервала возможных значений, оценка уровня сигнала при наличии фона. Понятие Байесовских сетей.

8. Нейронные сети. Однослойный и многослойный перцептрон. Обучение перцептрона, алгоритм обратного распространения ошибок. Глобальные методы оптимизации. Радиально-базисные сети. Задача кластеризации и сеть Кохонена. Применение нейронных сетей для классификации данных.
9. Задача разделения сигнала и фона (задача проверки гипотез). Критерий разделения, мощность и значимость критерия. Методы сравнения критериев. Простые гипотезы, лемма Неймана-Пирсона и наилучший критерий разделения. Критерий разделения в случае сложных гипотез. Практические методы построения критериев разделения: факторизация функции правдоподобия; линейный дискриминантный анализ Фишера; нейронные сети; усиленные деревья принятия решений; методы, основанные на подсчете числа событий. Метод главных компонент.
10. Задача обратной свертки (unfolding). Постановка задачи. Методы получения результатов без обратной свертки. Прямое решение задачи. Регуляризация. Регуляризация Тихонова. Метод максимальной энтропии.
11. Фурье- анализ. Дискретное преобразование Фурье. Быстрое преобразование Фурье. Вэйвлетный анализ.

План практических занятий

1. Знакомство с программным пакетом ROOT (ПО для обработки данных)
2. Выполнение задания №1: методы Монте-Карло и параметры распределений.
3. Знакомство с программным пакетом GEANT (ПО для моделирования взаимодействия частиц с веществом)
4. Выполнение задания №2: моделирование отклика простого детектора элементарных частиц
5. Выполнение задания №3: анализ результатов, полученных в ходе выполнения задания №2, с использованием метода максимального правдоподобия
6. Знакомство с пакетом TMVA (ПО для многомерного статистического анализа данных)
7. Выполнение задания №4: анализ результатов, полученных в ходе выполнения задания №2, с использованием нейронных сетей

Выполнение задания №5: анализ результатов, полученных в ходе выполнения задания №2, с использованием различных алгоритмов многомерного анализа данных.

5. Образовательные технологии

Теоретический материал курса освещается в ходе лекций. В лекциях обсуждается как необходимый математический аппарат и теоретические аспекты алгоритмов, так и реальные примеры использования обсуждаемых методов из практики наиболее известных экспериментов в мировой науке. В ходе лекций поощряются вопросы слушателей, часть тем обсуждается в форме дискуссий. Материал всех лекций доступен в электронном виде. В ходе лекций широко используются компьютерные демонстрации.

Практические занятия построены в форме семинаров, в ходе которых слушатели знакомятся с необходимыми программными продуктами. Все семинарские занятия проводятся в интерактивной форме. В ходе занятий слушатели выполняют задания самостоятельно или в группах из двух человек. Каждое задание отражает определенный этап анализа данных и определенный класс методов анализа. Пройдя все задания, слушатели получают практический опыт полноценного анализа результатов эксперимента и опыт использования широкого класса алгоритмов. Для сдачи задания слушатели должны не только успешно решить поставленную

задачу, но и продемонстрировать понимание соответствующего теоретического материала. В процессе выполнения и сдачи задания слушатели постоянно взаимодействуют с преподавателем.

В качестве текущего контроля успеваемости используется контроль выполнения практических заданий.

6. Учебно-методическое обеспечение самостоятельной работы студентов и аспирантов

Самостоятельная работа студентов поддерживается следующими методическими пособиями:

А.Д. Букин, С.И. Эйдельман. ЭВМ в планировании и обработке эксперимента. Учебное пособие. Изд. 2-е, испр. и доп. Ред.-изд. отдел НГУ, Новосибирск, 2002, 114 стр.

Лотов В.И. Теория вероятностей и математическая статистика. Новосибирск: НГУ, 2006.

Логашенко И.Б. Методы анализа экспериментальных данных. Электронный лекционный курс / Новосибирск: НГУ, 2013.

Логашенко И.Б. Практикум по методам анализа экспериментальных данных. Методическое пособие / Новосибирск: НГУ, 2013. 48 с.

Система контроля включает текущий (по ходу семестра) контроль освоения практического материала.

Текущий контроль по практике: осуществляется в ходе семестра путем приема заданий

7. Фонд оценочных средств для проведения аттестации по итогам освоения дисциплины:

Текущий контроль по дисциплине и промежуточная аттестация осуществляется по посещаемости и по результатам выполнения практических заданий.

Зачет по курсу выставляется по результатам сдачи практических заданий

Критерии оценки:

- «зачтено» выставляется аспиранту, который посещает большую часть занятий, в ходе работы на лекциях и практических занятиях показал, что знает основной материал, при этом возможно, что не усвоил его деталей, затрудняется в выполнении практических задач; выполнил не менее 3 заданий;
- «не зачтено» выставляется аспиранту, который не знает значительной части программного материала, допускает существенные ошибки, выполнил менее 3 заданий.

Примеры практических заданий:

1.1 Toy Monte-Carlo

Постановка задачи

Часто требуется оценить статистическую и систематическую ошибки выбранной процедуры анализа данных. В простых случаях, например, когда все распределения являются нормальными и используются хорошо известные статистические приемы, существуют аналитические формулы для подобных оценок. Однако в более сложных случаях необходимо использовать моделирование методом Монте-Карло.

Одна из наиболее распространенных процедур такого моделирования – toy Monte-Carlo, «простое» моделирование. В рамках этой процедуры исходные распределения для сырых данных считаются известными. По заданным распределениям генерируется массив сырых данных, которые затем проходят выбранную процедуру анализа и ее результаты сравниваются с параметрами, заложенными в моделирование. Как правило, генерируется множество независимых массивов данных, и из распределения полученных результатов извлекается информация о статистической ошибке и смещении процедуры.

Более сложными процедурами являются полное или параметризованное моделирование методом Монте-Карло, в котором сырые данные генерируются с использованием математических моделей, учитывающих лежащие в основе физические процессы. Пакет Geant4 предназначен для моделирования именно такого типа. Как правило, такие процедуры требуют гораздо больше вычислительных ресурсов, и они используются для оценки влияния параметров детектора на результаты анализа. Для оценки статистических свойств анализа достаточно toy Monte-Carlo.

В задании требуется провести небольшое исследование с помощью простого моделирования методом Монте-Карло.

1. Необходимо написать генератор случайных чисел, распределенных согласно плотности, заданной преподавателем. Следует использовать метод обратной трансформации или комбинированный метод
2. Продемонстрировать свойства двух различных оценок положения распределения: выборочного среднего и выборочной медианы. Пусть в одном эксперименте генерируется n_{data} (типичное значение 100) величин, распределенных согласно заданному распределению. По полученной выборке определяются значения выборочного среднего и выборочной медианы. Для того, чтобы изучить свойства этих двух оценок, повторим процедуру для n_{exp} (типичное значение 1000) экспериментов. Из распределения полученных значений оценок для различных экспериментов необходимо извлечь следующую информацию.
 - 1) Оценить мат.ожидание заданного распределения, определить точность этой оценки и сравнить полученное значение с расчетным.
 - 2) Оценить асимптотическую эффективность выборочной медианы в сравнении с выборочным средним. Для этого нужно провести моделирование для разных значений n_{data} и определить, чему асимптотически равно отношение дисперсий двух оценок.
 - 3) Продемонстрировать робастность выборочной медианы. Для этого модифицируйте генератор случайных чисел, добавив в него небольшую вероятность «выбросов». Покажите, как смещение и дисперсия двух оценок зависят от доли выбросов.
3. Продемонстрируйте предсказание центральной предельной теоремы. Для этого сгенерируйте распределение нормированной суммы 2, 5 и 100 случайных величин, подчиняющихся заданному распределению. Сходится ли полученное распределение к нормальному?

Дополнительные вопросы

1. При каком уровне выбросов стоит использовать выборочную медиану вместо выборочного среднего?
2. Сколько нужно сложить случайных величин, чтобы 95% доверительный интервал, определенный в предположении нормального распределения, был правильным с заданной точностью?
3. Замените свое распределение распределением Коши. Работает ли в этом случае центральная предельная теорема?
4. Вычислите доверительный интервал для параметров Mean (выборочное среднее) и RMS (выборочная дисперсия) для распределения h_3 на Рисунке 22, зная свойства исходного распределения и постановку эксперимента. Попали ли значения этих параметров, полученные в вашем численном эксперименте, в полученные интервалы?

1.2 Моделирование детектора

Постановка задачи

Требуется смоделировать отклик детектора элементарных частиц. Детектор представляет собой простой многослойный калориметр, либо куб, либо цилиндр, разделенный на N слоев ($N = 10$).

В качестве активного вещества используются широко распространенные в калориметрии материалы: кристаллы NaI, CsI, BGO, LSO, сжиженные благородные газы Ar, Kr, Xe. Каждый студент использует индивидуальный активный материал, предложенный преподавателем.

Регистрируемые частицы влетают в калориметр вдоль оси детектора. Возможны 3 типа частиц: электроны e , мюоны μ и гамма-кванты γ . Начальная энергия всех частиц одинакова, тип частицы должен выбираться случайным образом для каждого события.

Программа моделирования должна быть написана таким образом, что следующие параметры могут быть легко изменены без значительного переписывания кода: размеры детектора, количество слоев, активное вещество, энергия влетающих частиц.

В качестве результата необходимо провести моделирование 10000 событий. Для каждого события необходимо сохранить: тип частицы, полное энергосодержание в калориметре, энергосодержание в каждом слое. Результаты моделирования в форме дерева ROOT необходимо сохранить, т.к. эти данные будут использованы в последующих заданиях.

Основные и дополнительные вопросы

1. Объясните на качественном уровне отличия между полученными распределениями для разных типов частиц.
2. Попробуйте изменить список возможных взаимодействий для какого либо типа частиц в `ExN03PhysicsList` и посмотрите, как изменятся полученные распределения. Объясните результаты на качественном уровне.

1.3 Метод максимального правдоподобия

Постановка задачи

Пусть детектор регистрирует частицы двух типов и для каждой частицы измеряется только полное энерговыделение. Необходимо с помощью метода максимального правдоподобия оценить число событий каждого типа, анализируя измеренное распределение полного энерговыделения.

Из массива экспериментальных данных, сгенерированного в рамках задания 1.2, необходимо сформировать гистограмму полного энерговыделения в калориметре для событий двух типов (например, e и μ). Данная гистограмма представляет собой модель экспериментальных данных.

В экспериментальных данных перемешаны энерговыделения двух типов событий. Для того, чтобы оценить число событий каждого типа, необходимо подогнать данную гистограмму функцией

$$f(x) = n_e f_e(x) + n_\mu f_\mu(x)$$

где x – энерговыделение, n_k – число событий k -того типа, $f_k(x)$ – плотность распределения энерговыделения для событий k -того типа. Подгонку необходимо реализовать методом максимального правдоподобия. По результатам подгонки необходимо получить оценку числа событий каждого типа и точность этой оценки.

Основные и дополнительные вопросы

1. Сравните полученные оценки с их истинным значением (поскольку данные получены с помощью моделирования, точно известно, сколько событий какого типа). Согласуются ли полученные оценки с точным значением?
2. Как можно оценить смещение полученной оценки?
3. Какие выводы можно сделать на основе формы доверительного эллипса?
4. Пусть в результате эксперимента нужно получить не число событий каждого типа n_e и n_μ , а их сумму ($n_e + n_\mu$) (или разность ($n_e - n_\mu$), отношение n_e/n_μ и т.п.). Как вычислить значение суммы (отношения, разности и т.п.) и оценить ее ошибку, получив оценку n_e и n_μ с помощью метода максимального правдоподобия? Указание: необходимо учесть корреляцию между n_e и n_μ .
5. Получите оценку суммы (разности, отношения и т.п.) из вопроса 4 напрямую с помощью метода максимального правдоподобия. Совпала ли полученная оценка с расчетным значением, полученным в вопросе 4? Указание: для этого необходимо переписать функцию правдоподобия так, чтобы необходимая величина (сумма, разность, отношение и т.п.) являлась свободным параметром при минимизации.

1.4 Нейронные сети

Постановка задачи

Пусть стоит задача идентификации типа частицы, зарегистрированной детектором. В отличие от предыдущего задания, в котором было необходимо определить полное число событий каждого типа, теперь требуется идентифицировать каждое событие.

Постройте нейронную сеть, которая будет решать поставленную задачу. Используйте массив экспериментальных данных, сгенерированный в рамках задания 1.2, для обучения сети. В качестве входов сети используйте энерговыделения в каждом слое, в качестве выхода – тип частицы. В

качестве сигнала используйте события с гамма-квантами, в качестве фона – события с электронами.

Результатами выполнения задания являются структура сети, распределения ответа сети для событий сигнала и фона, и кривая $\alpha - \beta$ (ROC).

Основные и дополнительные вопросы

1. Нарисуйте кривые $\alpha - \beta$ для различных конфигураций сети.
2. Используйте подмножество слоев в качестве входов сети. Нарисуйте соответствующие кривые $\alpha - \beta$ для различных наборов входов и сделайте вывод о том, какие слои предоставляют больше информации о типе частицы.
3. Покажите влияние разрешения детектора на результат работы сети. Для этого модифицируйте энергосыделения в слоях, добавив к ним шум, и постройте семейство кривых $\alpha - \beta$. Сеть должна быть обучена на исходных (не модифицированных) данных.
4. Покажите влияние калибровки детектора на результат работы сети. Для этого модифицируйте энергосыделения в слоях, умножив их на масштабный коэффициент. Масштабный коэффициент должен быть выбран случайно для каждого слоя, но один и тот же коэффициент должен быть использован для всех событий. Такой подход позволяет смоделировать именно ошибку калибровки, когда существует неизвестный масштабный фактор, который не меняется от события к событию. Необходимо провести несколько раундов такого моделирования, и для каждого раунда построить кривую $\alpha - \beta$. Сеть должна быть обучена один раз на исходных (не модифицированных) данных. Ширина семейства кривых покажет влияние ошибки калибровки.

1.5 Многомерная классификация данных

Постановка задачи

В этом задании требуется решить ту же задачу, что и в предыдущем задании - необходимо идентифицировать тип частицы, зарегистрированной детектором. Однако теперь эту задачу надо решить не с помощью нейронной сети, а с помощью других алгоритмов многомерной классификации.

Используйте массив экспериментальных данных, сгенерированный в рамках задания 1.2, для обучения алгоритмов. В качестве входных данных используйте энергосыделения в каждом слое, в качестве результата работы алгоритма – тип частицы. В качестве сигнала используйте события с гамма-квантами, в качестве фона – события с электронами.

Необходимо сравнить следующие алгоритмы:

- 1) отношение правдоподобий в предположении независимости входных данных;
- 2) подсчет числа событий в тренировочном множестве в окрестности текущего события;
- 3) усиленные деревья принятия решений.

Требуется реализовать эти алгоритмы, построить отклик классификатора для событий сигнала и фона и кривую $\alpha - \beta$ (ROC) для каждого алгоритма, и сделать вывод, какой алгоритм лучше всего работает в данном случае.

8. Учебно-методическое и информационное обеспечение дисциплины

Список основной и дополнительной литературы:

Основная литература

1. Логашенко И.Б. Методы анализа экспериментальных данных. Электронный лекционный курс.
2. Логашенко И.Б. Практикум по методам анализа экспериментальных данных. Методическое.
3. А.Д. Букин, С.И. Эйдельман. ЭВМ в планировании и обработке эксперимента. Учебное пособие. Изд. 2-е, испр. и доп. Ред.-изд. отдел НГУ, Новосибирск, 2002, 114 стр.

Открытые интернет-ресурсы

1. б) программное обеспечение и Интернет-ресурсы:
- 2.
3. Описание пакета ROOT. <http://root.cern.ch/drupal/content/users-guide>
4. Описание пакета TMVA. <http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf>
5. Описание пакета Geant4. <http://geant4.cern.ch/support/userdocuments.shtml>
- 6.
7. .

9. Материально-техническое обеспечение дисциплины

Лекционные занятия проводятся в аудитории, оборудованной мультимедийным проектором и компьютером, необходимых для презентации электронного варианта лекций и проведения компьютерных демонстраций.

Практические занятия проводятся в терминальном классе. На каждом компьютере доступны операционная среда UNIX и программные пакеты GEANT4 и ROOT, необходимые для выполнения заданий.