



Государственный научный центр Российской Федерации –
Институт физики высоких энергий
Национального исследовательского центра
«Курчатовский институт»

ИФВЭ 2015–1
ОЭФ

В.Б. Аникеев, В.В. Липаев

**Алгоритмы и программа для модельно- и детекторно-
независимого оценивания распределений (unfolding)**

Протвино 2015

Аннотация

Аникеев В.Б., Липаев В.В. Алгоритмы и программа для модельно- и детекторно-независимого оценивания распределений (unfolding): Препринт ИФВЭ 2015–1. – Протвино, 2015. – 16 с., 3 рис., библиогр.: 11.

Задача модельно-независимого оценивания распределения с учетом искажений, вносимых детектором (unfolding), является одной из типичных задач при обработке данных эксперимента в физике высоких энергий.

Задача решается методом математического моделирования. Выбрана математическая модель изучаемого процесса – измерение функции распределения по некоторой кинематической переменной детектором с конечным разрешением – в виде интегрального уравнения. Предложен критерий выбора биннинга для матрицы, описывающей разрешение детектора (смиринговая матрица), – это биннинг, которому соответствует минимальное число обусловленности полученной смиринговой матрицы. Для тех случаев, когда это число неприемлемо велико, предложен метод построения регуляризованного решения с оценками смещенности и ковариационной матрицы этого решения.

Разработана программа, реализующая предложенные алгоритмы в среде Linux/ROOT/LAPACK. Приведены форматы команд, управляющих работой программы, и описана процедура доступа к серверу, на котором функционирует программа.

Abstract

Anikeev V.B., Lipaev V.V. Model Free Estimation of Distributions Taking into Account Detector Imperfection (Unfolding). Algorithms and Software: IHEP Preprint 2015–1. – Protvino, 2015. – p. 16, figs. 3, refs.: 11.

Among typical problems of data processing in high energy physics there is the problem of model free estimation of distribution in some kinematical variable, taking into account the resolution in this variable – unfolding.

Mathematical modeling method is used to solve the problem. The mathematical model of the process being under the investigation, i.e. measurement of distribution in some kinematical variable with finite resolution detector, is the integral equation. The criterion to choose the optimal binning of smearing matrix (this matrix describes the resolution of the detector) is proposed – binning that provides the minimal condition number for the matrix. For the case when this number is inappropriate big the regularization for the problem solution is suggested, including estimates of bias and covariation matrix.

Software in Linux/ROOT/LAPACK environment was developed. The format of commands and the access procedure to the server where program runs are described.

Введение

Задача модельно независимого (непараметрического) оценивания распределения по некоторой кинематической переменной, учитывающая искажения (конечное разрешение), вносимые экспериментальной установкой по этой переменной x (unfolding), является одной из типичных задач при обработке данных эксперимента в физике высоких энергий. Под термином «модельно независимое (непараметрическое) оценивание» мы полагаем, что оцениваемая функция плотности вероятности (ф.п.в.) распределения $s(x)$ не описывается функцией некоторого набора параметров $\alpha - s(\alpha, x)$. Поскольку класс модельно независимых оценок включает в себя модельно зависимые оценки, выбор модели, т.е. набор параметров оцениваемой функции, можно сделать на основе экспериментальных данных, из которых «устранено» влияние детектора. Такие оценки упрощают сравнение распределений, измеренных на конкретном детекторе, с теоретическими моделями или с измеренными на других детекторах.

Задача решается методом математического моделирования. Выбрана математическая модель изучаемого процесса – измерение распределения по некоторой кинематической переменной детектором с конечным разрешением – в виде интегрального уравнения. После дискретизации модели несовершенство детектора описывается смиринговой матрицей. Предложен критерий выбора биннинга для смиринговой матрицы – такой биннинг, который приводит к минимальному числу обусловленности этой матрицы. Для тех случаев, когда это число неприемлемо велико, предложен метод регуляризации смиринговой матрицы с оценками возникающей смещенности решения и ковариационной матрицы вектора решения.

Описана программа, реализующая разработанные алгоритмы в среде Linux/ROOT/Larack. Приведены формат команд программы и процедура доступа к серверу, на котором функционирует программа.

Математическая модель измерения распределения по кинематической переменной x несовершенным детектором

Ожидаемое число событий $\bar{Y}_i$ в i -м бине по измеренной кинематической переменной x' имеет вид [1]:

$$\bar{Y}_i = N_{total} \int_{x'_i}^{x'_{i+1}} dx' \int_{-\infty}^{+\infty} E'(x') K(x, x') E(x) s(x) dx, \quad (1)$$

где x – «истинная» (true, part[icle level]) кинематическая переменная (непрерывная, дискретная, многомерная);

x' – «измеренная» (meas[ured], reco[nstructed]) кинематическая переменная;

$s(x)$ – искомое распределение по кинематической переменной x (ф.п.в.: $s(x) \geq 0$, $\int_{-\infty}^{+\infty} s(x) dx = 1$.);

$1 \geq E(x) \geq 0$. – геометрический аксептанс детектора и эффективность по x ;

$E'(x') > 0$. – аксептанс по x' : обрезания, эффективность программы восстановления;

$K(x, x')$ – ядро; для фиксированного x – функция разрешения (ф.п.в.) по x' ;

$\pm\infty$ – здесь диапазон изменения x ;

$[x'_i, x'_{i+1})$ – границы i -го бина по x' ;

N_{total} – число событий в эксперименте.

Эта модель является интегральным уравнением Фредгольма I рода относительно функции $s(x)$, т.е. малым изменениям в левой части уравнения (в гистограмме, в измерениях) могут соответствовать большие изменения в решении – задача, некорректная по Тихонову. Следует иметь в виду возможность наличия в решении компоненты ортогональной ядру, т.е. неизмеримой компоненты.

Дискретизация модели

Найти решение уравнения (1) в общем виде (не привлекая модели) – функция $s(x)$ принадлежит бесконечномерному пространству – на основании конечного числа имею-

щихся измерений невозможно, поэтому решение ищется в виде вектора $s_{N \times 1}$ ¹. В предположении, что биннинг по x и x' произведен – выбрано N и M примыкающих необязательно одинаковых интервалов по x и x' соответственно – уравнение (1) принимает вид:

$$\bar{Y}_{M \times 1} = N_{total} E'_{M \times M} \cdot K_{M \times N} \cdot E_{N \times N} \cdot s_{N \times 1}, \quad (2)$$

где вектор $s_{N \times 1}$ – искомая дискретная ф.п.в., а $N_{ev} \cdot s_{N \times 1} = X_{N \times 1}$ – гистограмма по переменной x без искажений, вносимых несовершенным детектором; $K_{M \times N}$ – матрица, описывающая миграцию событий в соседние бины, – смиринговая матрица. Если l_{mn} – число событий в бине mn двумерной гистограммы (элементы матрицы миграции событий) выборки событий со значениями «истинной» и «измеренной» кинематической переменной $\{x, x'\}$, полученной методом Монте-Карло, то элемент k_{mn} смиринговой матрицы вычисляется по формуле $k_{mn} = l_{mn} / \sum_{j=1, M} l_{jn}$ – число событий в ячейке делится на число событий в столбце. Таким образом $k_{mn} \geq 0$, а столбец матрицы есть дискретная ф.п.в., описывающая разрешение детектора в этом бине по x . Смиринговая матрица по построению является стохастической (элементы ее неотрицательны, а сумма элементов в столбце равна единице). $E_{N \times N}$ – диагональная матрица, описывающая эффективность и геометрический аксептанс детектора, $E'_{M \times M}$ – диагональная матрица, описывающая обрезания и эффективность программы восстановления событий.

На рис. 1 приведена диаграмма рассеяния выборки событий (точки), полученных методом Монте-Карло для некоторого углового распределения: x_{part} – абсцисса, «истинная» переменная x , x'_{reco} – ордината, «измеренная» переменная x' . Диагональ – линия $x_{part} = x'_{reco}$. На этой линии были бы размещены все точки выборки для идеального детектора, т.е. если бы $K(x, x') = \delta(x - x')$ и, следовательно, $K_{M \times N}$ была бы единичной матрицей. Вертикальные линии соответствуют биннингу по x (вертикальные полосы), а горизонтальные – по x' . В рассматриваемом примере биннинги по x и по x' одинаковые. Координаты перекрестий по x соответствуют среднему x -координат точек, расположенных в полосах, опирающихся на бины по x . И, соответственно, координаты перекрестий по x' соответствуют среднему x' -координат точек, расположенных в тех же полосах. Размеры плеч перекрестий по x (горизонталь) соответствуют среднеквадратичным отклонениям x -координат точек, расположенных в полосах, опирающихся на бины по x , а

¹ Для матрицы K , у которой M строк и N столбцов, используется запись $K_{M \times N}$.

размеры плеч по x' (вертикаль) соответствуют среднеквадратичным отклонениям x' -координат точек, расположенных в тех же полосах. На рисунке видно, что размеры плеч по x' больше плеч по x , что указывает на конечное разрешение детектора по кинематической переменной x .

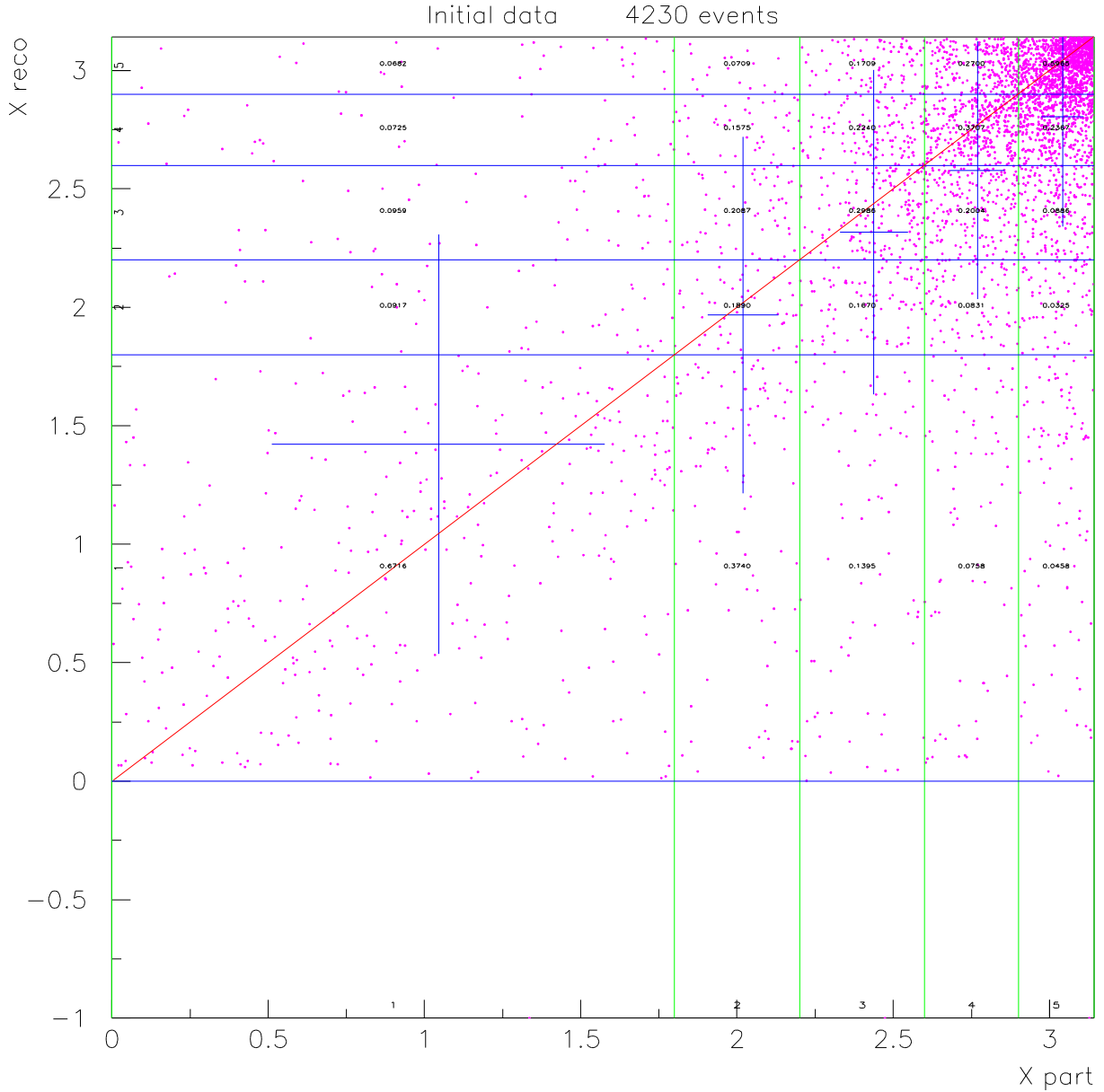


Рис. 1. Диаграмма рассеяния выборки событий (точки), описывающей несовершенство детектора (получены методом Монте-Карло). Подробности в тексте.

Решение дискретизованной задачи (2) может быть записано в виде:

$$X_{N \times 1} = (K_{M \times N})^{-1} \cdot \bar{Y}_{M \times 1}, \quad (3)$$

т.е. сведено к обращению смиринговой матрицы $K_{M \times N}$. Вообще говоря, можно обращать не только квадратные матрицы. В вычислительной практике применяется псевдо-обращение прямоугольных матриц – матрица Мура-Пенроуза [2, §5.5.4]. Далее полагается, что смиринговая матрица $K_{M \times N}$ квадратная, т.е. $N=M$.

Однако в силу неустраняемой стохастической природы измерений вместо ожидаемых (средних) получают другие величины:

$$Y_{M \times 1} = \bar{Y}_{M \times 1} + \varepsilon_{M \times 1}. \quad (4)$$

Обычно $\varepsilon_{M \times 1}$ – ошибки числа входов в бины гистограммы – считают некоррелированными и распределенными по биномиальному закону ($\bar{\varepsilon}_i = 0$; $\overline{\varepsilon_i \varepsilon_i} = Y_i$; $\overline{\varepsilon_i \varepsilon_j} = 0, i \neq j$). Такое предположение близко к истине при выполнении трудно проверяемых требований: большое число бинов и большая заселенность всех бинов гистограммы. Далее предполагается, что числа входов в бины гистограммы $Y_{M \times 1}$ следуют мультиномиальному распределению [3, §4.1.2].

Итак, искомая модельно независимая (непараметрическая) оценка $\hat{X}_{M \times 1}$ вычисляется умножением матрицы, обратной смиринговой, на вектор измерений:

$$\hat{X}_{M \times 1} = (K_{M \times M})^{-1} \cdot Y_{M \times 1}. \quad (5)$$

Оптимальный выбор биннинга

Важной величиной, характеризующей свойства матрицы является её число обусловленности. Если имеется система линейных уравнений $Ax = y$, то число обусловленности матрицы $cond(A) = \|A\| \cdot \|A^{-1}\| \geq 1$ связывает приращение решения Δx с приращением правой части Δy ([4, §8.3]), является коэффициентом роста в оценке сверху относительной погрешности:

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\Delta y\|}{\|y\|}. \quad (6)$$

Матричная норма должна быть согласована с выбранной векторной нормой. Существует формула, связывающая рост погрешности решения, включая погрешности элементов матрицы, поскольку в нашей задаче элементы смиринговой матрицы, вообще говоря, случайные величины, полученные методом Монте-Карло. Далее мы считаем

смиринговую матрицу, вычисленную по выборке полученной методом Монте-Карло, заданной точно.

Естественно стремление работать со смиринговой матрицей с минимальным числом обусловленности. Его возможно уменьшить, построив другие смиринговые матрицы, варьируя положение бинов. Заметим, что задача поиска оптимального биннинга состоит из двух задач: поиск оптимального числа бинов и поиск оптимального размещения бинов [5]. Здесь рассматривается только оптимальное размещение бинов.

Задача сводится к заполнению матрицы миграции событий и вычислению смиринговой матрицы в зависимости от положения внутренних точек, определяющих размещение бинов. Элементы смиринговой матрицы по построению неотрицательны, поэтому её собственные значения (с.з.) неотрицательны, неотрицателен и её детерминант (произведение с.з.). Программа минимизации варьирует положение бинов с тем, чтобы получить смиринговую матрицу с минимальным числом обусловленности. При этом ведется контроль за тем, чтобы бины не были пустыми и не «перехлестывались» (ограничение шагов по изменению границ бинов). Поскольку в нашем случае идеальным является случай, когда смиринговая матрица единичная, смысл поиска оптимальной смиринговой матрицы (с минимальным числом обусловленности) заключается в приближении её в каком-то смысле к единичной матрице. Близкий к этому смысл имеют минимизация отношения максимального с.з. смиринговой матрицы к минимальному с.з. и максимизация её детерминанта. При реализации алгоритма заложена возможность на основании какой-то априорной информации фиксировать при минимизации положение тех или иных бинов.

Оптимизация выбора биннинга может быть полезной и при оценке параметров модели, которые можно оценить при фитировании «размытой» (смиринговая матрица) модели с параметрами к измерениям.

Оценка ковариационной матрицы модельно-независимой оценки

Исходя из принятого допущения, что гистограмма измерений $Y_{M \times 1}$ следует мультиномиальному распределению, по этому закону генерируется набор гистограмм $\{Y_{M \times 1}\}$

(метод статистических испытаний), соответственно, вычисляется набор решений (оценок) $\{\hat{X}_{M \times 1}\}$. Из этого набора вычисляется ковариационная матрица, которая считается ковариационной матрицей решения $\hat{X}_{M \times 1}$. Увеличивая набор гистограмм, сгенерированных по $Y_{M \times 1}$, можно добиться стабилизации элементов ковариационной матрицы.

Регуляризация и оценка ковариационной матрицы и смещенности регуляризованного решения

В случае, когда число обусловленности смиринговой матрицы неприемлемо или велики диагональные элементы ковариационной матрицы, обосновано применение регуляризации [6]. В нашем случае к диагонали смиринговой матрицы прибавляется неотрицательное число α , далее называемое параметром регуляризации, и перенормируются столбцы этой матрицы:

$$k_{ii}^{\alpha} = k_{ii} + \alpha, \quad k_{ij}^{\alpha} = k_{ij} / ((\sum_i k_{ij}) + \alpha) = k_{ij} / (1 + \alpha), \quad i \neq j.$$

При таком преобразовании столбцы сохраняют свойства дискретной ф.п.в. – сохраняется число событий в регуляризованном решении. Итак, при $\alpha = 0$ регуляризации нет, а при $\alpha = \infty$ смиринговая матрица становится единичной с минимально возможным числом обусловленности 1. Регуляризация приводит с одной стороны к уменьшению числа обусловленности, т.е. к уменьшению статистических ошибок в оценке решения, а с другой стороны к смещенности оценки решения, которое наряду со статистической ошибкой необходимо оценивать. Статистическая ошибка оценивается как и в случае без регуляризации (см. выше) методом статистического моделирования. Для оценки смещенности регуляризованной оценки необходимо иметь несмещенную оценку решаемой задачи. Для этого используется выборка Монте-Карло изучаемого процесса. Из этой выборки (пары x, x') строится подвыборка событий, воспроизводящая “наилучшим” образом гистограмму измерений (по x'). Используются две процедуры. В первой случайным образом (с возможными повторами) из выборки Монте-Карло выбирается число событий в бине по x' , равное числу событий в этом же бине в измеренной гистограмме. Во второй процедуре из выборки Монте-Карло последовательно исключаются события, максимизирующие расстояние Колмогорова-Смирнова [7] для кумулятивных функций вероятно-

сти измеренной гистограммы и остающейся части выборки Монте-Карло, пока это расстояние не перестанет уменьшаться. Значения x в сформированных подвыборках заполняют гистограмму несмещенной оценки нерегуляризованного решения. К сожалению этот подход («анфолдинг без анфолдинга») не позволяет удовлетворительно оценить статистические ошибки полученного решения.

Кроме того, возникает задача выбора оптимального в каком-то смысле параметра регуляризации. Обычно этот параметр выбирается таким, чтобы статистические ошибки были близки к смещенности, хотя возможны и другие мотивы, например, минимизация статистической ошибки в каких-то бинах.

Схема решения задачи модельно независимого оценивания приведена на рис. 2. Искомое решение X (вектор, гистограмма) преобразуется смиринговой матрицей K в ожидаемый вектор измерений $\bar{Y}$. Однако, в силу неустранимой статистической природы измерений (при накоплении гистограммы) мы получаем вектор Y с ковариационной матрицей компонент. Умножив этот вектор на матрицу K^{-1} , обратную смиринговой, получаем модельно независимую оценку решения обратной задачи $\hat{X}$. Методом статистических испытаний определяем ковариационную матрицу оценки. Если статистические ошибки оценки решения неприемлемо велики, то с помощью обратной регуляризованной матрицы K_{α}^{-1} для какого-то значения параметра регуляризации α ищется регуляризованная оценка решения $\hat{X}_{\alpha}$. Вычисляются смещенность и ковариационная матрица регуляризованной оценки. Параметр регуляризации подбирается вручную.

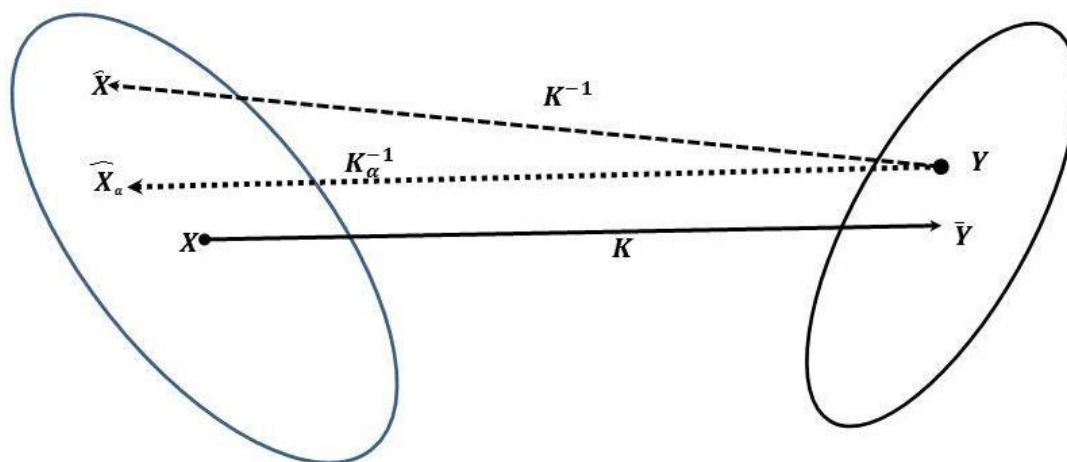


Рис. 2. Схема решения задачи модельно-независимого оценивания.

Возможности дальнейшего развития алгоритмов модельно-независимого оценивания измеренных распределений

Специфика конкретного физического анализа, иными словами, объем доступной априорной информации, может потребовать соответствующей модификации алгоритмов. Наиболее легко реализуемыми являются алгоритмы:

- учитывающие неодинаковое число бинов по x и x' ;
- учитывающие тот факт, что элементы смиринговой матрицы являются случайными величинами с заметными статистическими ошибками;
- использующие в рамках описываемого проекта другие методы регуляризации, как описанные нами ранее [1] (спектральное окно, итерационный, регуляризация по Тихонову и др.), так и разработка новых;
- использующие возможность автоматического выбора параметра регуляризации.

Изложенные выше алгоритмы были применены при подготовке публикаций [8] в эксперименте DØ на ускорителе Тэватрон в лаборатории им. Ферми. На основании полученного опыта было решено организовать сервер с соответствующим программным обеспечением.

Выделенный сервер: доступ, работа с программой

Для реализации описанных выше алгоритмов была создана программа uuABL. В настоящее время она адаптирована для работы на выделенном сервере ИФВЭ **unfold.ihep.su** под OS Ubuntu 14.04 и доступна для удалённых пользователей как с компьютерами под OS Windows, так и с компьютерами под OS Linux.

Программа uuABL:

Язык: C++

OS Linux, gcc - 4.5.1

Библиотеки:

Root v5_26_00d-GCC_4_5_1 [9]

LAPACK (Linear Algebra PACKage) 3.2.2 [10]

BLAS (Basic Linear Algebra Subprograms) 3.2.2-2.1 [11].

Для того, чтобы получить доступ к серверу, необходимо послать e-mail по адресу **unfold@ihep.ru** с просьбой завести учётную запись (account) для пользователя, указав в нем имя организации и желаемое имя пользователя <username>. В ответ будет послан e-mail с текстом ключа <**private key**>, который требуется для входа нового пользователя <username> на сервер, и присоединенный файл **readme.txt**, в котором содержится инструкция для работы с сервером и программой uuABL.

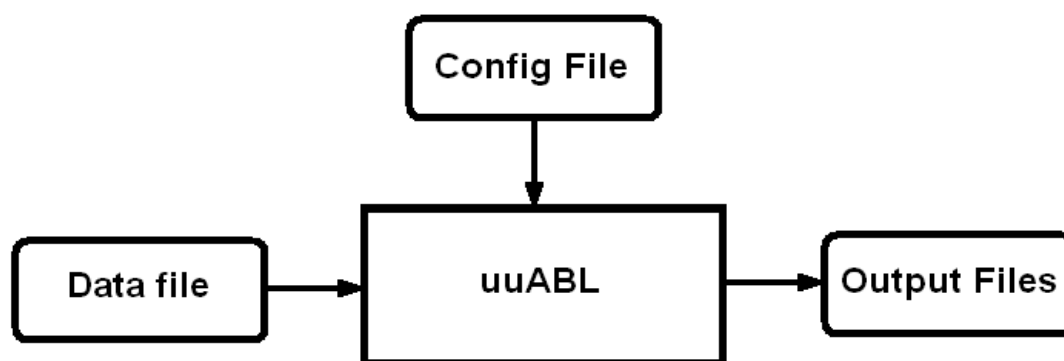


Рис. 3. Блок-схема взаимодействия программы uuABL с “внешним миром”.
Data file – файл с данными {part/true, reco/meas};
Config file – конфигурационный файл, в котором задаются параметры и ключи для программы uuABL в текстовом виде;
Output files – файлы, создаваемые в процессе работы программы uuABL.

Конфигурационный файл (Config File).

Это текстовый ASCII файл, в котором посредством команд задаются значения параметров и некоторым образом алгоритм работы программы uuABL.

Формат команды:

command_name: [parameters]

command_name располагается в начале строки и заканчивается ‘:’

Пустые строки и строки, начинающиеся с ‘#’, игнорируются.

Ниже представлены основные команды, полный список дан в файле *readme.txt*.

Все команды условно можно разбить на несколько групп.

I. Команды задания имен директорий и файлов для входных/выходных данных

Имена директорий указываются либо как */full_path/dirname*, либо относительно той, в которой запускается программа *uuABL* – *./dirname*.

input_dir: *dirname*

dirname – директория, где находится файл с исходными данными.

output_dir: *dirname*

dirname – директория, куда будут заноситься все файлы, создаваемые во время работы *uuABL*.

data_file: *datafilename*

datafilename – имя файла с данными Монте-Карло, описывающими искажение изучаемой кинематической переменной детектором/экспериментальной установкой. Файл является ASCII-файлом, каждая строка которого содержит три действительных числа – вес и значения переменной на уровнях *part/true* и *reco/meas*. Таким образом, число строк в этом файле определяет объем выборки Монте-Карло.

log_file: *filename*

filename – имя *logfile* (*протокола*), куда будет занесена в текстовом виде информация о работе программы *uuABL*. Если не задано, по умолчанию это будет файл *log.txt*.

II. Команды задания бининга

Формат задания значений – свободный, что подразумевает, что они могут занимать произвольное число строк и начинаться в строке с командой сразу после неё. Разделителем между значениями являются пробел, табуляция или конец строки. В настоящее время реализован вариант программы, где числа бинов на уровнях *part/true* и *reco/meas* одинаковы.

part_level_binning: *V₁ V₂ ...*

$V_1 V_2 \dots$ – границы бинов для распределения по изучаемой кинематической переменной на уровне *part/true*. Число значений должно быть $N_{bin}+1$, где N_{bin} – число бин.

reco_level_binning: $V_1 V_2 \dots$

$V_1 V_2 \dots$ – границы бинов для распределения по изучаемой кинематической переменной на уровне *reco/meas*.

По выборке Монте-Карло и заданным биннингам для кинематической переменной на уровнях *part/true* и *reco/meas* строится матрица миграции событий. Двумерное изображение этой матрицы сохраняется в файле *datafilename.eps*. Нормировкой на единицу числа событий, принадлежащих каждому бину на уровне *part/true* по бинам на уровне *reco/meas* матрица миграций преобразуется в матрицу размытия (*smearing matrix*). Эта матрица является стохастической матрицей поскольку ее столбцы являются дискретной функцией плотности вероятности – элементы неотрицательны и в сумме равны единице. Умножая гистограмму экспериментальных данных на матрицу, обратную смиринговой, получаем оценку распределения, в которой устранено размывающее влияние детектора. Важным является корректное оценивание ковариационной матрицы (статистические ошибки и корреляции) значений гистограммы восстановленного/скорректированного распределения. Число обусловленности смиринговой матрицы является оценкой сверху фактора увеличения статистических ошибок экспериментальных данных. Таким образом, варьируя границы бинов, можно попытаться уменьшить этот фактор – число обусловленности смиринговой матрицы или близкое ему по смыслу число.

III. Команды указания типа минимизируемого функционала для Minuit

minimize_smearing_matrix_condition_number:

Ищется минимальное значение числа обусловленности смиринговой матрицы.

minimize_max/min_smearing_matrix_eigenvalues_ratio:

Ищется минимальное значение отношения модулей максимального собственного значения (*eigenvalue*) смиринговой матрицы к минимальному с.з.

maximize_determinant:

Ищется минимум $1/D$, где $D(>0.)$ – детерминант смиринговой матрицы, т.е. произведение с.з. смиринговой матрицы.

IV. Команды задания параметров для розыгрыша гистограммы

экспериментальных данных по мультиномиальному распределению

exp_data: $V_1 V_2 \dots$

$V_1 V_2 \dots$ содержимое бинов (числа событий) в гистограмме экспериментальных данных.

multinomial_numb_prob: $numb$

numb – число розыгрышей гистограммы экспериментальных данных по мультиномиальному распределению. Величина определяется стабилизацией контролируемых величин на приемлемом уровне точности.

Задание бинов, для которых считается суммарная ошибка для каждого бина (квадратично сложенные оценочные дисперсия и смещённость); по умолчанию включает все бины. Имеется возможность либо указать номера конкретных бинов командой

critical_bins: $Bin_1 Bin_2 \dots$

либо для всех, кроме перечисленных в команде

noncritical_bins: $Bin_1 Bin_2 \dots$

alpha_min_max_step: $min\ max\ step$

Задание набора значений параметра регуляризации смиринговой матрицы α : $\alpha_1=min$, $\alpha_2=\alpha_1+step$, $\alpha_3=\alpha_2+step, \dots, \alpha_{last}=max$

Значения параметра регуляризации смиринговой матрицы α также могут быть заданы перечислением значений в команде

alpha_list: $\alpha_1\ \alpha_2\ \dots$

V. Команда запуска обработки входных данных

start:

Пример конфигурационного файла

```
# Задание директории для файлов, создаваемых во время работы
# программы uuABL
output dir: ./OUT
# Задание директории, в которой находится файл с данными
# Монте-Карло
input dir:  ./INP
# Имя файла с данными Монте-Карло
data file: testdatafile
# Задание биннинга для part/true
part level binning:
30. 40. 50. 60. 70. 80. 90. 100. 125. 150. 200. 350. 500.
# Задание биннинга для reco/meas
reco level binning:
30. 40. 50. 60. 70. 80. 90. 100. 125. 150. 200. 350. 500.
# тип минимизируемого функционала для Minuit - число обуслов-
# ленности смиринговой матрицы.
minimize_smearing_matrix_condition_number:
# Задание значений в бинах (числа событий) в гистограмме экс-
# периментальных данных
exp_data: 3029 6874 3707 2103 1244 767 506 650 242 224 149 14
# число розыгрышей гистограммы экспериментальных данных по
# мультиномиальному распределению
multinomial_numb_prob: 1000000
# Значения параметра регуляризации смиринговой матрицы Alpha
alpha_list: 0.0
# Запуск программы uuABL на выполнение с учетом заданных пара-
# метров
start:
```

Заключение

Авторы благодарят сотрудничество ДО в лице Денисова Д., Денисова С.П. и Бандурина Д. за предоставленную возможность применить свои предложения при анализе

данных, полученных на установке DØ, Разумова И.А. за помощь в настройке выделенного сервера, Садовского С.А. и Качаева И.А. за ценные замечания и Бритвича Г.И. за постоянный интерес к работе.

Список литературы

- [1] Anykeyev V.B., Spiridonov A.A. and Zhigunov V.P., Comparative investigation of unfolding methods, Nucl. Instrum. Methods A303 (1991) 350.
Anikeyev V.B. and Zhigunov V.P., Phys. Part. Nucl. 24, 424 (1993) (in English).
- [2] Голуб Дж., Ван Лоун Ч. Матричные вычисления. – М.: Мир, 1999.
- [3] Идье В. и др. Статистические методы в экспериментальной физике. – М.: Атомиздат, 1976.
- [4] Вержбицкий В.М. Вычислительная линейная алгебра. – М.: Высшая школа, 2009.
- [5] Zlokazov V. On optimum binning for spectral distributions. Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment. 01/1989; 275(2):392-396.
- [6] Тихонов А.Н., Гончарский А.В., Степанов В.В., Ягола А.Г. Численные методы решения некорректных задач. – М.: Наука, 1990.
- [7] Боровков А.А. Математическая статистика. – М.: Наука, 1994.
- [8] Azimuthal decorrelations and multiple parton interactions in $\gamma\gamma+2$ jet and $\gamma\gamma+3$ jet events in $p\bar{p}$ collisions at $\sqrt{s}=1.96$ TeV DØ Collaboration (V.M. Abazov (Dubna, JINR) *et al.*). Jan 2011. Published in **Phys.Rev. D83 (2011) 052008**.
Measurement of the differential cross sections for isolated direct photon pair production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV DØ Collaboration (V.M. Abazov (Dubna, JINR) *et al.*). Jan 2013. 10 pp. Published in **Phys.Lett. B725 (2013) 6-14**.
Measurement of the differential cross section of photon plus jet production in $p\bar{p}$ collisions at $\sqrt{s}=1.96$ TeV DØ Collaboration (V.M. Abazov (Dubna, JINR) *et al.*). Aug 12, 2013. 23 pp. Published in **Phys.Rev. D88 (2013) 072008**.
- [9] <http://root.cern.ch/> — A Data Analysis Framework.

- [10] www.netlib.org/blas/ — Basic Linear Algebra Subroutines (BLAS).
www.netlib.org/lapack/ — Linear Algebra PACKage (LAPACK).
- [11] <http://root.cern.ch/root/html/TMinuit.html>,
<http://wwwinfo.cern.ch/asdoc/minuit/minmain.html>.

Рукопись поступила 15 января 2015 г.

В.Б. Аникеев, В.В. Липаев

Алгоритмы и программа для модельно- и детекторно-независимого оценивания распределений (unfolding).

Препринт отпечатан с оригинала-макета, подготовленного авторами.

Подписано к печати 22.01.2015	Формат 60 × 84/16.	Цифровая печать.
Печ.л. 1,25. Уч.– изд.л. 1,73.	Тираж 80. Заказ 2.	Индекс 3649.

ФГБУ ГНЦ ИФВЭ НИЦ «Курчатовский институт»
142281, Московская область, г. Протвино, пл. Науки, 1

www.ihep.ru; библиотека <http://web.ihep.su/library/pubs/all-w.htm>

Индекс 3649

П Р Е П Р И Н Т 2015-1,
ФГБУ ГНЦ ИФВЭ НИЦ «Курчатовский институт», 2015
