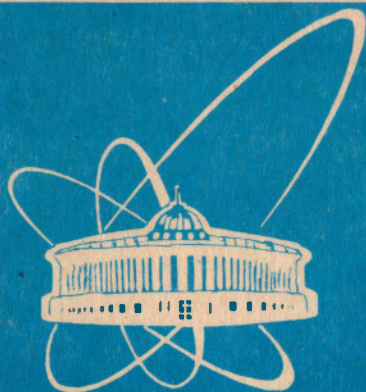


94-374



сообщения
объединенного
института
ядерных
исследований
дубна

P10-94-374

Н.Д.Гагунашвили*, А.В.Тиханин

CORDIS — ПРОГРАММА ВОССТАНОВЛЕНИЯ
ИСТИННЫХ РАСПРЕДЕЛЕНИЙ
ИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ,
ИСКАЖЕННЫХ ДЕТЕКТОРАМИ
С КОНЕЧНЫМ РАЗРЕШЕНИЕМ

*E-mail: gagunash@sunse2.jinr.dubna.su

1 Введение

Программа CORDIS (CORrection DIStortions) предназначена для восстановления истинных распределений из экспериментальных данных, искаженных детекторами с конечным разрешением. Программа может использоваться также для частичной компенсации искажений, связанных с разрешением детектора (редукции к прибору с лучшим разрешением), что может быть полезным в экспериментах, где производятся относительные измерения. В этом случае проводится учет эффективности и выравнивание функции разрешения, если она разная в разных областях изменения параметров распределений.

Для использования программы необходимо иметь данные о разрешении и эффективности детектора. Эти данные могут быть представлены как аналитически, так и в виде числового материала, получаемого с помощью программ, моделирующих детектор. Метод восстановления приведен в [1], целью данной работы является описание программы, реализующей метод. Программа поставлена на SUN SPARC station в операционной системе UNIX и на IBM PC в операционной системе MS DOS.

2 Структура и алгоритмы функционирования программы

Программа состоит из двух частей. В первой части реализована идентификация детектора, с помощью которого проводится измерение распределения. Во второй - реализована процедура восстановления истинного распределения.

Идентификация детектора или расчет элементов матрицы P , описывающей связи между истинным распределением и измеренным распределением, производится последовательно по строкам, которые соответствуют каналам измеренного распределения. Выбор элементов строки, включенных в описание связи (не нулевых элементов), производится методом пошаговой регрессии [2]. Элемент строки включается в описание связи, если при l элементах строки, уже включенных в описание,

$$\chi_l^2 - \chi_{l+1}^2 > F_{in} \cdot \frac{\chi_{l+1}^2}{k - l - 1},$$

где k - число распределений используемых для идентификации,
 χ_l^2 - хи квадрат в случае описания связи l элементами строки,
 F_{in} - константа(порог).

Один из l элементов строки, уже включенных в описание, может быть исключен, если

$$\lambda_l^2 - \lambda_{l-1}^2 < F_{out} \cdot \frac{\lambda_l^2}{k-l}.$$

Выбор ненулевых элементов строки завершается в случае, когда нет элементов строки, удовлетворяющих одному или второму неравенству. Пороги F_{in} и F_{out} выбираются пользователем. Хорошие результаты дают пороги $F_{in} = F_{out} = 3.29$, которые имеют и некоторое теоретическое обоснование [2]. Первый элемент строки в процедуре пошаговой регрессии выбирается соответствующим каналу (точке) истинного распределения, минимально удаленному от канала реконструированного распределения. Рассчитывая таким образом матрица P может быть улучшена по критерию минимума детерминанта полной матрицы ошибок восстановленного распределения. Для уменьшения детерминанта рассчитываются последовательно варианты строк, отличающиеся выбором ненулевых элементов. При этом предыдущий вариант строки матрицы P заменяется на последующий, если детерминант матрицы ошибок при этом уменьшается. Варианты строки в программе определяются выбором первого ненулевого элемента. Выбор первого элемента упорядочен по расстоянию от соответствующего ему канала (точки) истинного распределения до канала реконструированного распределения и начинается с минимального.

Вторая часть программы - это восстановление истинного распределения по методу наименьших квадратов и по методу максимального правдоподобия. Метод наименьших квадратов реализован с использованием подпрограммы решения системы линейных уравнений RSINV из библиотеки стандартных подпрограмм ЦЕРН [3]. Решение задачи по методу максимального правдоподобия проводится методом простой итерации с использованием в качестве начального приближения решения задачи по методу наименьших квадратов.

3 Входные данные CORDIS

Основные данные для программы вводятся с помощью FFREAD [3] следующими значениями ключей:

- NTCH - Число гистограмм, используемых для идентификации.

- NPOI - число каналов восстанавливаемого распределения.
- POIN - нижняя граница и ширины каналов восстанавливаемого распределения.
- MNPA - максимальное число ненулевых элементов строки.
- THRE - величина порога F_{in} .
- THDE - величина порога F_{out} .
- NNEW - число вариантов заполнения каждой строки при оптимизации величины детерминанта матрицы ошибок.
- NULL - максимальное число итераций для вычисления по методу максимального правдоподобия, если параметр равен 0 вычисления не производятся.
- PRUI - необходимая точность вычислений по методу максимального правдоподобия.
- FHEX - имя файла экспериментальных данных.
- FHMC - имя файла распределений, используемых для идентификации.
- FTRU - имя файла, где будет записано восстановленное распределение и полная матрица ошибок.

Файл основных данных должен иметь название FFCSO.

Структура файла экспериментальных данных

Файл экспериментальных данных содержит следующие записи:

- Число каналов гистограммы экспериментальных данных; число параметров распределения для псевдоданных, 0 для данных; нормировка.
- Параметры распределения (только для псевдоданных).
- Содержимое каналов гистограммы экспериментальных данных.

Структура файла распределений, используемых для идентификации

Файл распределений, используемых для идентификации, содержит следующие записи:

- Число каналов гистограмм реконструированных распределений, число параметров распределений, используемых для идентификации, нормировка.
- Нижние границы каналов гистограммы, верхняя граница гистограммы.
- Параметры 1-го распределения.
- Содержимое 1-ой гистограммы реконструированного распределения.
- Параметры 2-го распределения.
- Содержимое 2-ой гистограммы реконструированного распределения.
-
-
-
- Параметры n -го распределения.
- Содержимое n -ой гистограммы реконструированного распределения.

4 Подпрограммы пользователя

Пользователем должны быть сделаны подпрограммы:

- `FUNCTION FIFA(X,PARA)` проводит вычисление генерированного распределения,
X-аргумент функции распределения,
PARA-массив параметров генерированного распределения.
- `FUNCTION FIRA(X,PARA)` проводит вычисление истинного распределения (при обработке псевдоданных),
X-аргумент функции распределения,
PARA-массив параметров истинного распределения.

5 Выходные данные CORDIS

Выходные данные CORDIS представлены в виде числового материала, выдаваемого на печать, файла восстановленного распределения и файла для PAW (Physics Analysis Workstation)[3].

На печать выдается следующая информация:

- Матрица P .
- Матрица ошибок матричных элементов матрицы P .
- Матрица инцидентности к матрице P .
- χ^2 на число степеней свободы для каждого реконструированного канала при идентификации.
- Значения восстановленного распределения по методу наименьших квадратов.
- Полная матрица ошибок восстановленного распределения по методу наименьших квадратов.
- Значения восстановленного распределения по методу максимального правдоподобия.
- Полная матрица ошибок восстановленного распределения по методу максимального правдоподобия.

Структура файла восстановленного распределения

- Значения восстановленного распределения.
- Полная матрица ошибок.

Структура файла для PAW

Файл для PAW назван `hdat.dat` и содержит следующие графики и гистограммы:

- График χ^2 на степень свободы при идентификации.
- График остатков для 1-го канала при идентификации.
- График остатков для 2-го канала при идентификации.
-
-
-
- График остатков для n -го канала при идентификации.
- График вариантов 1-го канала при идентификации.
- График вариантов 2-го канала при идентификации.
-
-
-
- График вариантов n -го канала при идентификации.
- Гистограмма экспериментального распределения.
- Гистограмма восстановленного распределения по методу наименьших квадратов.
- Гистограмма восстановленного распределения по методу максимального правдоподобия.
- График остатков при восстановлении распределения по методу наименьших квадратов.
- График остатков при восстановлении распределения по методу максимального правдоподобия.
- Гистограмма остатков при восстановлении распределения по методу наименьших квадратов.

- Гистограмма остатков при восстановлении распределения по методу максимального правдоподобия.
- График остатков при сравнении распределения, восстановленного по методу наименьших квадратов, с истинным распределением (при обработке псевдоданных).
- График остатков при сравнении распределения, восстановленного по методу максимального правдоподобия, с истинным распределением (при обработке псевдоданных).

6 Сообщения, выдаваемые программой во время ее работы

В случае нормального завершения работы программы выдается сообщение:

STOP: CORDIS NORMAL EXIT.

В случае, если параметр THRE меньше THDE, что приводит к закликиванию программы, выдается сообщение:

STOP: THRE SHOULD BE GREATER THEN THDE.

В случае, если не может быть подобрана матрица P, у которой все столбцы не нулевые, выдается сообщение:

STOP: CAN NOT FIND P MATRIX.

В случае, если не может быть проведено восстановление истинного распределения по методу наименьших квадратов, выдается сообщение:

STOP: MATRIX NON-POSITIVE IN CORR SUBROUTINE.

В случае, если не может быть проведено восстановление истинного распределения по методу максимального правдоподобия, выдается сообщение:

STOP: MATRIX NON-POSITIVE IN LCORR SUBROUTINE.

7 Пользовательский интерфейс программы CORDIS

При работе с программой CORDIS рекомендуется использовать CMZ (A Source Code Management System)[4]. Программа CORDIS — это CMZ-файл cor1.cmz. Вызов программы в UNIX осуществляется с помощью последовательности команд:

```
\rm -fr tmp
mkdir tmp
cmz -b cor1
cd tmp
f77 -c -g -Nx600 -N1300 *.f
f77 -temp=, -g -Nx600 -N130 -o ../cor1.e *.o \
/cern/93b/lib/libpacklib.a
\rm *.o
cd ..
nice cor1.e>cor1.out &
```

Данный вызов использует файл cor1.knshrc, в котором содержатся инструкции для CMZ-системы:

```
set_file 'tmp/*.f' -F
file cor1 -r
sequence dkeep
buffer .
buffer //cor1/cor1/*
buffer
ctof -b
use cor1.cra
cfor $USE
rel cor1
exit
```

Файл, в котором будут содержаться коррекции пользователя, должен называться cor1.cra.

8 Пример входных данных для программы CORDIS

Входные данные программы содержатся в трех файлах:

Файл основных данных FFCO

```
LIST
HOLL          $
C
C            NUMBER OF HISTOGRAMS FOR THE IDENTIFICATION
C
NTCH          100
C
C            NUMBER OF POINTS FOR THE UNFOLDED DISTRIBUTION
c
NPOI          12
C
C            POINTS OF THE UNFOLDED DISTRIBUTION
C
POIN          4. 12*1.
C
C            MAXIMAL NUMBER NON ZERO ELEMENTS OF A STRING
C
MNPA          12
C
C            THE THRESHOLD FOR F- STATISTICS
C
THRE          3.29
C
C            THE SECOND THRESHOLD FOR F- STATISTICS
C
THDE          3.29
C
C            NUMBER OF VERSIONS OF THE FIRST STEP
C
NNEW          12
C
C            MAXIMAL NUMBER OF ITERATION FOR THE LIKELIHOOD METHOD,
```

```

C                               IF ZERO, SKIP
C
NULI                30
C
C          PRECISION OF THE UNFOLDED FUNCTION BY LIKELIHOOD METHOD
C
PRLI                0.001
C
C          FILE NAME FOR THE EXPERIMENTAL HISTOGRAM
C
FHEX                $FHEX$
C
C          FILE NAME FOR MONTE-CARLO HISTOGRAMS
C
FHMC                $FHMC$
C
C          FILE NAME FOR THE UNFOLDED DISTRIBUTION
C          AND THE COMPLETE MATRIX OF ERRORS
C
FTRU                $FTRU$
KEYS
STOP

```

Файл экспериментальных данных FHEX

90 6 8.20087E-04

2.00000	1.00000	10.00000	14.0000	1.00000	1.00000	
34.0000	32.0000	26.0000	26.0000	42.0000	38.0000	22.0000
38.0000	33.0000	22.0000	36.0000	52.0000	32.0000	54.0000
56.0000	57.0000	52.0000	67.0000	88.0000	70.0000	87.0000
102.000	98.0000	127.000	114.000	135.000	122.000	121.000
148.000	135.000	152.000	133.000	145.000	126.000	164.000
173.000	184.000	160.000	182.000	163.000	173.000	170.000
156.000	162.000	153.000	164.000	150.000	163.000	140.000
152.000	121.000	142.000	131.000	120.000	127.000	125.000
114.000	89.0000	106.000	106.000	105.000	78.0000	90.0000
96.0000	88.0000	89.0000	66.0000	70.0000	84.0000	61.0000

63.0000	68.0000	70.0000	64.0000	58.0000	46.0000	41.0000
40.0000	47.0000	41.0000	32.0000	25.0000	27.0000	23.0000
15.0000	20.0000	30.0000	28.0000	27.0000	39.0000	

Файл распределений, используемых для идентификации, ФНМС

90 6 8.20087E-04

2.00000	4.12500	4.87500	5.25000	5.50000	5.75000	6.00000
6.12500	6.25000	6.37500	6.50000	6.62500	6.75000	6.87500
7.00000	7.12500	7.25000	7.37500	7.50000	7.62500	7.75000
7.87500	8.00000	8.12500	8.25000	8.37500	8.50000	8.62500
8.75000	8.87500	9.00000	9.12500	9.25000	9.37500	9.50000
9.62500	9.75000	9.87500	10.0000	10.1250	10.2500	10.3750
10.5000	10.6250	10.7500	10.8750	11.0000	11.1250	11.2500
11.3750	11.5000	11.6250	11.7500	11.8750	12.0000	12.1250
12.2500	12.3750	12.5000	12.6250	12.7500	12.8750	13.0000
13.1250	13.2500	13.3750	13.5000	13.6250	13.7500	13.8750
14.0000	14.1250	14.2500	14.3750	14.5000	14.6250	14.7500
14.8750	15.0000	15.1250	15.2500	15.3750	15.5000	15.6250
15.7500	15.8750	16.0000	16.2500	16.5000	16.8750	18.0000

1.90958	1.46661	8.34898	10.07019	1.48268	1.48529	
195.414	156.423	145.895	142.889	171.802	220.511	110.875
133.019	116.939	139.783	137.467	192.794	161.018	192.062
240.560	258.473	191.296	201.391	232.162	177.543	236.718
260.909	257.525	264.166	263.223	251.016	269.738	284.609
283.083	272.513	257.971	303.352	263.092	294.659	284.350
266.332	248.333	265.451	278.963	229.631	318.870	253.117
234.885	268.178	223.211	197.026	198.542	200.395	199.109
196.662	187.296	155.586	156.219	137.882	154.679	111.105
129.207	101.940	115.412	103.399	94.3534	91.7555	92.7384
64.3799	61.3163	58.0044	60.9055	59.2235	46.8758	35.1630
43.3854	39.7006	34.2925	37.5878	29.2221	21.7178	25.9055
21.3792	18.6340	19.6790	8.17865	17.1754	11.3701	12.8194
9.45767	12.2393	14.3598	8.63772	11.2387	17.7944	

1021.69	830.164	743.396	755.399	892.974	1102.27	538.680
651.920	587.837	683.777	636.478	926.912	785.031	904.728
1063.39	1212.95	777.964	752.192	977.091	645.074	919.188
957.438	930.227	877.769	874.667	697.198	866.135	819.444
884.536	761.985	665.085	863.209	610.724	668.569	657.847
659.182	544.399	546.857	611.583	458.848	643.433	509.100
408.284	542.133	378.281	360.539	310.980	327.486	346.000
307.374	280.025	248.522	207.997	182.356	218.073	139.975
152.924	131.188	176.241	119.408	136.104	114.669	102.237
71.0515	67.1303	64.3379	65.7806	60.4089	43.5784	29.8007
41.2316	39.3164	29.3249	31.6256	23.1734	16.6368	20.0356
16.4340	9.24530	15.1608	6.17172	10.1245	9.12375	8.20964
5.94140	8.40324	7.17158	4.33660	7.60699	9.81899	

Далее следуют еще 99 записей вида последних трех блоков для других значений параметров распределений.

Авторы выражают благодарность И.А.Савину и В.Г.Кривохижину за поддержку и постоянный интерес к работе, А.П.Нагайцеву за замечания высказанные им после прочтения рукописи публикации.

Литература

- [1] N.D.Gagunashvili, Nucl. Instr. and Meth. A 343 (1994) 606.
- [2] G.A.F. Seber, Linear Regression Analysis. John Wiley and Sons, New York, London, Sydney, Toronto, 1977.
- [3] CERN Program library, CERN Geneva, Switzerland, 1993.
- [4] CMZ. User's Guide & Reference Manual (1.41), 1993.

• Рукопись поступила в издательский отдел
22 сентября 1994 года.