

СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ

Дубна

P11-2003-143

А. В. Стадник<sup>1</sup>, Н. И. Чернов<sup>2</sup>, С. С. Шиманский<sup>3</sup>

«КОРРЕЛЯТОРЫ» И «МОМЕНТЫ»

---

<sup>1</sup>E-mai: stadnik@sunhe.jinr.ru

<sup>2</sup>University of Alabama, Birmingham, USA,

E-mail: chernov@math.uab.edu

<sup>3</sup>E-mail: shimansk@sunhe.jinr.ru



# Введение

Современные эксперименты в физике частиц дают все больший объем информации, подлежащий обработке и численному анализу, поэтому алгоритмы анализа данных, позволяющие сократить число вычислительных операций, имеют важное прикладное значение.

В столкновениях ядер на ускорителе RHIC (BNL, США) рождаются сотни и тысячи частиц. В таких столкновениях исследуется возможность проявления различного рода коллективных эффектов, которые могут проявляться в виде корреляций кинематических характеристик рожденных частиц. В настоящей работе показана возможность существенно сократить время при вычислении “корреляторов”, которые рассматриваются в работах [1, 2].

Рассмотрим произвольную статистическую выборку  $x_1, x_2, \dots, x_N$ . Обозначим математическое ожидание или среднее по выборке

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Тогда если  $y_i = x_i - \bar{x}$ , то

$$\sum_{i=1}^N y_i = 0. \quad (1)$$

Определим также величины

$$S_L = \sum_{i=1}^N (x_i - \bar{x})^L = \sum_{i=1}^N y_i^L,$$

$$D_L = \sum_{1 \leq i_1 < \dots < i_L \leq N} (x_{i_1} - \bar{x}) \cdots (x_{i_L} - \bar{x}) = \sum_{1 \leq i_1 < \dots < i_L \leq N} y_{i_1} \cdots y_{i_L}.$$

Величины  $D_L$  используются при вычислении “корреляторов”, которые определяются ниже. Простые соображения подсказывают, что  $D_L$  можно выразить через  $S_L$ . Давайте рассмотрим случай, когда у нас выборка состоит из 3-х переменных  $y_1, y_2$  и  $y_3$ .

Воспользуемся соотношением (1)

$$(y_1 + y_2 + y_3)^2 = (y_1^2 + y_2^2 + y_3^2) + 2(y_1 \cdot y_2 + y_1 \cdot y_3 + y_2 \cdot y_3) = 0.$$

Отсюда мы получаем соотношение

$$(y_1 \cdot y_2 + y_1 \cdot y_3 + y_2 \cdot y_3) = -\frac{1}{2}(y_1^2 + y_2^2 + y_3^2),$$

или, что то же самое,

$$D_2 = -\frac{1}{2}S_2.$$

Если в правой части соотношения (1) будет не 0, а константа  $K$ , то связь между  $D_L$  и  $S_L$  усложнится, но тем не менее останется. Так, для  $D_2$  мы получим соотношение

$$D_2 = \frac{1}{2}K^2 - \frac{1}{2}S_2.$$



В **Приложении 1** показано, что для любой выборки значение величины  $D_L$  можно выразить через величины  $S_i, i = 2 \dots L$ , что является следствием соотношения (1). Выполняются соотношения:

$$\begin{aligned} D_2 &= -\frac{1}{2}S_2, \\ D_3 &= \frac{1}{3}S_3, \\ D_4 &= -\frac{1}{4}S_4 + \frac{1}{8}S_2^2, \\ D_5 &= \frac{1}{5}S_5 - \frac{1}{6}S_2S_3, \\ D_6 &= -\frac{1}{6}S_6 + \frac{1}{8}S_4S_2 + \frac{1}{18}S_3^2 - \frac{1}{48}S_2^3, \\ D_7 &= \frac{1}{7}S_7 - \frac{1}{10}S_5S_2 - \frac{1}{12}S_4S_3 + \frac{1}{24}S_3S_2^2, \\ D_8 &= -\frac{1}{8}S_8 + \frac{1}{12}S_6S_2 + \frac{1}{15}S_5S_3 + \frac{1}{32}S_4^2 - \frac{1}{32}S_4S_2 - \frac{1}{36}S_3^2S_2 + \frac{1}{384}S_2^4. \end{aligned} \quad (2)$$

Вычисление  $D_L$  по формулам (2) имеет по крайней мере три **существенных преимущества** перед прямым вычислением:

### 1. Меньшее количество операций.

Так как при вычислении  $S_L$  суммируются всегда  $N$  членов, а при вычислении  $D_L$  суммируются  $C_N^L = \frac{N!}{L!(N-L)!}$ , то при больших  $N$  и  $L$  прямое вычисление  $D_L$  требует неоправданно много вычислительного времени в сравнении с вычислениями через  $S_L$ . Так, например, прямое вычисление  $D_3$  при  $N = 1000$  потребует суммирования  $C_{1000}^3 = 166167000$  членов. В то же время расчет через  $S_3$  требует суммирования всего 1000 членов, что потребует на 5 порядков меньше процессорного времени. Число операций при фиксированном  $L$  и большом  $N$  будет порядка  $N$  при использовании формул (2), и порядка  $N^L$  операций при прямом вычислении в соответствии с определением. Таким образом, прямое вычисление  $D_L$  при  $N > 1000$  и  $L > 4$  становится практически невозможным и предложенный подход является единственно возможным.

### 2. Меньшие ошибки округления.

Накопленные ошибки округления при выполнении операций с плавающей точкой пропорциональны их количеству, поэтому при вычислении  $D_L$  по формулам (2) накопленные ошибки существенно меньше.

### 3. "Правильная" асимптотика.

Формулы (2) дают "правильную" асимптотику роста "корреляторов" в пределе при  $N \rightarrow \infty$ . Исходя из них получается, что  $D_L$  растет как  $N^{(\frac{L}{2})}$  (т.е. как  $N$  в степени целая часть от  $\frac{L}{2}$ ), в то время как по прямой формуле может показаться, что  $D_L$  растет как  $C_N^L$  (число сочетаний из  $N$  по  $L$ ), что асимптотически эквивалентно  $N^L$ .

В работе [1] предлагается исследовать асимптотическое поведение величин

$$f_L = \frac{|K_L|^{\frac{1}{L}}}{|K_2|}, \quad (3)$$

где  $L \geq 3$ .

Рассмотрим случай, когда величины  $K_L$  выражаются через  $D_L$  следующим образом:

$$K_L = \bar{D}_L = \frac{1}{C_N^L} D_L \quad (4)$$

— коррелятор порядка  $L$ , где

$$C_N^L = \frac{N!}{L!(N-L)!}$$

— число слагаемых при вычислении  $D_L$ .

В данной работе мы в качестве примера применения соотношений (2) провели численное исследование поведения величин (3) для простого распределения плотности вероятности.

## Численное моделирование

Для численного моделирования и оценки асимптотического поведения величин (3) было выбрано распределение

$$f(x) = \frac{1}{\eta} \sqrt{\frac{2}{\pi}} \cdot e^{-\frac{x^2}{2\eta^2}}.$$

Как следует из формул (2), для исследования асимптотического поведения при больших  $N$  величин  $f_L$  необходимо и достаточно знания асимптотики величин  $S_i$ . Через  $S_L$  определяются **центральные моменты**

$$\bar{S}_L = \frac{1}{N} S_L.$$

На рис. 1-4 построены **центральные моменты**  $\bar{S}_3, \bar{S}_4, \bar{S}_5, \bar{S}_6$  модельной выборки в зависимости от её длины  $N$ . Видно, что для выбранного распределения величины  $S_3, S_4, S_5, S_6$  при больших  $N$  пропорциональны длине выборки, так как **центральные моменты** при больших  $N$  выходят на константы.

На рис. 5-8 представлены зависимости от  $N$  отношений  $f_3, f_4, f_5, f_6$ . Фитирование гистограмм рис. 1-4 проводилось степенной зависимостью вида

$$F(N) = p_1 N^{p_0} + p_2.$$

Полученные численные значения показателя степени  $p_0$  можно сравнить с "аналитическими оценками", которые следуют из формул (2), если  $\bar{S}_i \sim const$  при  $N \gg 1$ .

Таблица, приведенная ниже, демонстрирует согласие "аналитических" и численных оценок асимптотики  $f_L$ .

$f_L$	"аналитическая оценка"	численная оценка
$f_3$	$f_3 \sim N^{-\frac{1}{3}}, p_0 = 0.33$	$p_0 = -0.338 \pm 0.03$
$f_4$	$f_4 \sim const, p_0 = 0$	$p_0 = 5 \cdot 10^{-5} \pm 3 \cdot 10^{-6}$
$f_5$	$f_5 \sim N^{-\frac{1}{2}}, p_0 = 0.20$	$p_0 = -0.202 \pm 0.005$
$f_6$	$f_6 \sim const, p_0 = 0$	$p_0 = 1 \cdot 10^{-4} \pm 6 \cdot 10^{-6}$



## Заключение

В данной работе мы обсудили алгоритм, позволяющий существенно сократить время вычислений так называемых “корреляторов”. Показано, как провести оценку асимптотического поведения “корреляторов”, исследуя поведение **центральных моментов**.

## Приложение 1

Пусть  $x_1, \dots, x_N$  – действительные числа (статистическая выборка) и

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

обозначим среднее (математическое ожидание). Введем для удобства новые переменные  $y_i = x_i - \bar{x}$ . Для каждого  $L \geq 1$  рассмотрим величины

$$S_L = \sum_{i=1}^N y_i^L$$

и

$$D_L = \sum_{1 \leq i_1 < \dots < i_L \leq N} y_{i_1} \dots y_{i_L},$$

где сумма берется по всем комбинациям  $L$  различных индексов, причем каждая комбинация учитывается только однажды.

Важно также заметить, что

$$y_1 + \dots + y_N = 0. \quad (5)$$

Возведение в квадрат этого уравнения дает

$$\sum_{i=1}^2 y_i^2 + 2 \sum_{i < j} y_i y_j = 0.$$

Отсюда получаем соотношение

$$S_2 + 2D_2 = 0,$$

которое и есть искомая зависимость для  $L = 2$ .

Теперь рассмотрим случай  $L \geq 3$ . Прежде всего заметим, что

$$D_L = \frac{1}{L!} \sum_{i_1 \neq \dots \neq i_L} y_{i_1} \dots y_{i_L}, \quad (6)$$

где сумма взята по всем комбинациям с  $L$  различными индексами  $i_1, \dots, i_L$  (потому что число перестановок  $L$  различных индексов равно  $L!$ ). Суммирование по  $i_L$  дает

$$\sum_{i_L \notin \{i_1, \dots, i_{L-1}\}} y_{i_L} = -y_{i_1} - \dots - y_{i_{L-1}}$$

в силу уравнения (5). Это позволяет нам сократить суммирование в (6). Это и есть основной принцип вычислений. Для  $L = 3$  мы получаем:

$$\begin{aligned} D_3 &= \frac{1}{3!} \sum_{i \neq j \neq k} y_i y_j y_k = \\ &= \frac{1}{3!} \sum_{i \neq j} y_i y_j (-y_i - y_j) = \\ &= -\frac{1}{3!} \sum_{i \neq j} y_i^2 y_j - \frac{1}{3!} \sum_{i \neq j} y_i y_j^2. \end{aligned}$$

Теперь мы используем (5) снова, чтобы вычислить

$$\sum_{j \neq i} y_j = -y_i$$

и

$$\sum_{i \neq j} y_i = -y_j.$$

Таким образом,

$$D_3 = \frac{1}{6} \sum_{i=1}^N y_i^3 + \frac{1}{6} \sum_{j=1}^N y_j^3,$$

что дает

$$D_3 = \frac{1}{3} S_3.$$

Вычисления для  $L = 4$  и более немного длиннее, но очень похожи.

Соотношения для  $D_L$  более высоких порядков можно получить, используя программу на языке символьных вычислений REDUCE [3, 4], приведенную в приложении 2.

В качестве примера результата подобных вычислений можно привести соотношение для  $D_{13}$ :

$$\begin{aligned} D_{13} &= \frac{1}{13} S_{13} - \frac{1}{22} S_{11} S_2 - \frac{1}{30} S_{10} S_3 - \frac{1}{36} S_9 S_4 + \frac{1}{72} S_9 S_2^2 - \frac{1}{40} S_8 S_5 + \frac{1}{48} S_8 S_3 S_2 - \frac{1}{42} S_7 S_6 + \\ &+ \frac{1}{56} S_7 S_4 S_2 + \frac{1}{126} S_7 S_3^2 - \frac{1}{336} S_7 S_2^3 + \frac{1}{60} S_6 S_5 S_2 + \frac{1}{72} S_6 S_4 S_3 - \frac{1}{144} S_6 S_3 S_2^2 + \\ &+ \frac{1}{150} S_5^2 S_3 + \frac{1}{160} S_5 S_4^2 - \frac{1}{160} S_5 S_4 S_2^2 - \frac{1}{180} S_5 S_3^2 S_2 + \frac{1}{1920} S_5 S_2^4 - \frac{1}{192} S_4^2 S_3 S_2 - \\ &- \frac{1}{648} S_4 S_3^3 + \frac{1}{576} S_4 S_3 S_2^3 + \frac{1}{1296} S_3^3 S_2^2 - \frac{1}{11520} S_3 S_2^5. \end{aligned}$$



## Приложение 2

Здесь мы приводим программу на языке REDUCE [3, 4], которая вычисляет соотношения типа (2) для любого порядка. Для этого определена процедура  $D(n)$ . Например, для того чтобы получить соотношение для  $D_{13}$ , приведившееся выше, достаточно после загрузки в REDUCE файла с текстом приводимой ниже программы выполнить команду  $D(13)$ ;

Ниже следует текст программы:

```
input_case nil$
operator S;
S(1):=0$

procedure p3(x,l,r);
if r={} then s(reverse(x.l))
else if first(r)>x then s(append(reverse(x.l),r))
else p3(x,first(r).l,rest(r))$

procedure p2(x,l,r,z);
if r={} then z
else p2(x,first(r).l,rest(r),z-p3(x+first(r),l,rest(r)))$

procedure s(l);
if rest(l)={} then S(first(l))
else p2(first(l),{,rest(l),if first(l)=1 then 0 else S(first(l))*s(rest(l))}$

procedure D(n);
s(for i:=1:n collect 1)/factorial(n);

end;
```

Программа на языке REDUCE, реализующая алгоритм вычисления  $D_L$ , приведенный в приложении 1, написана А.Г.Грозиным (БИЯФ СО РАН, Новосибирск).

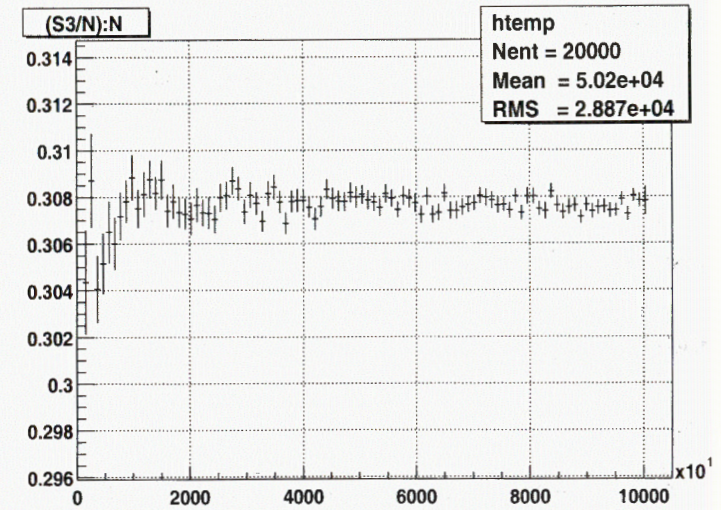


Рис. 1:  $\bar{S}_3$  vs.  $N$

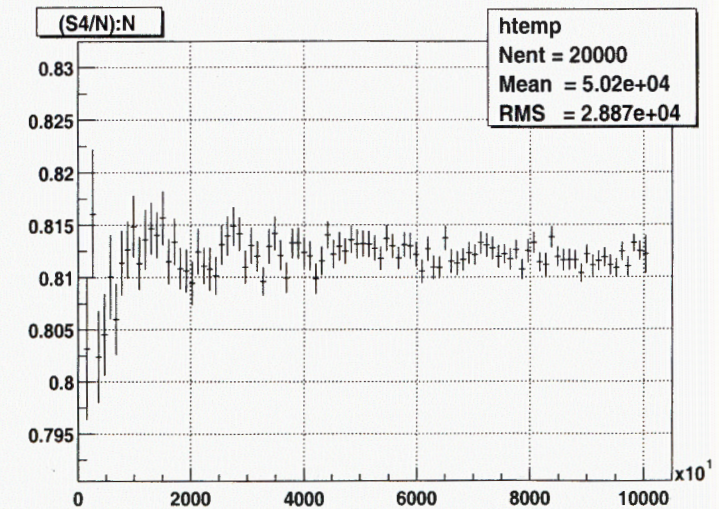


Рис. 2:  $\bar{S}_4$  vs.  $N$



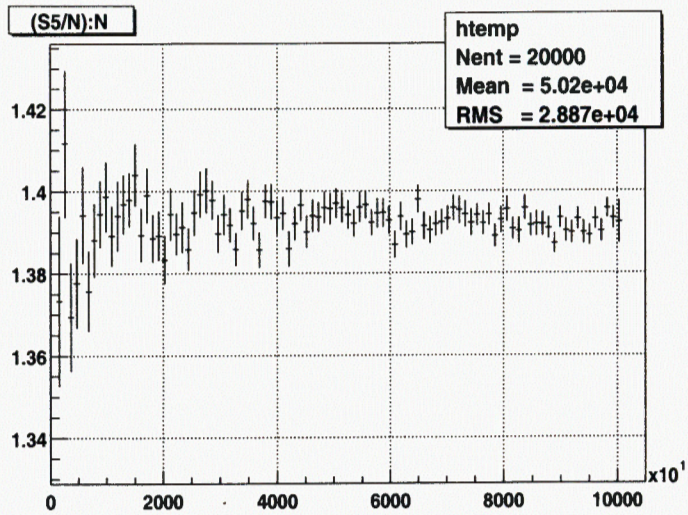


Рис. 3:  $\bar{S}_5$  vs.  $N$

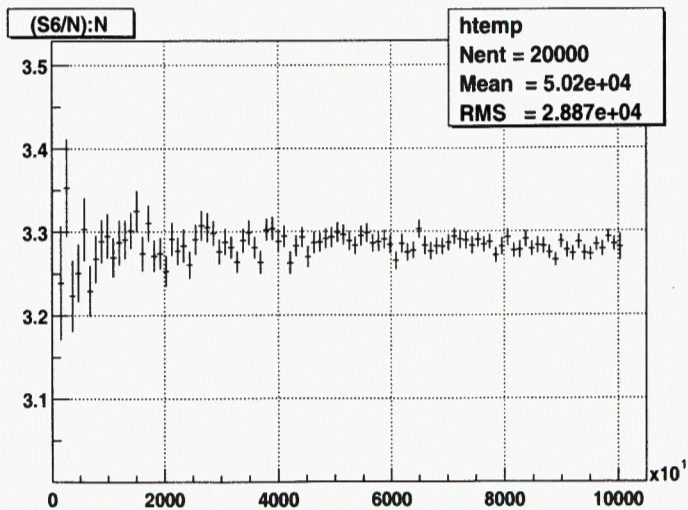


Рис. 4:  $\bar{S}_6$  vs.  $N$

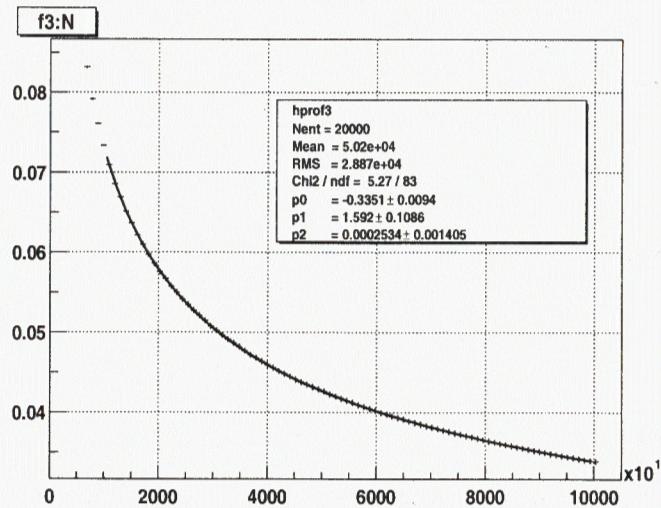


Рис. 5:  $f_3$  vs.  $N$

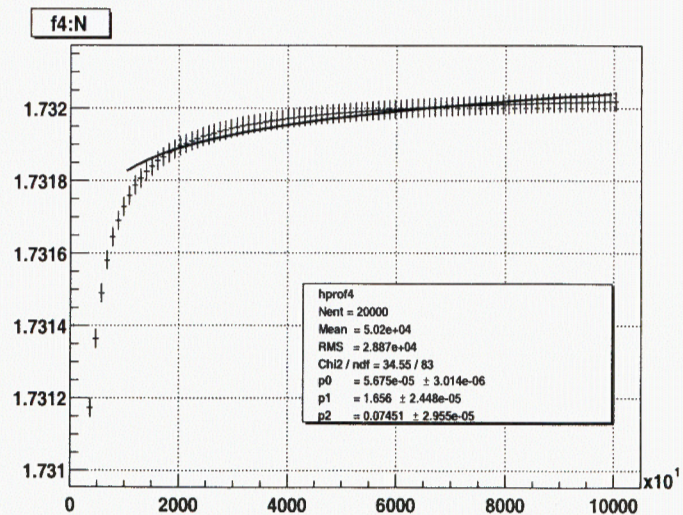


Рис. 6:  $f_4$  vs.  $N$



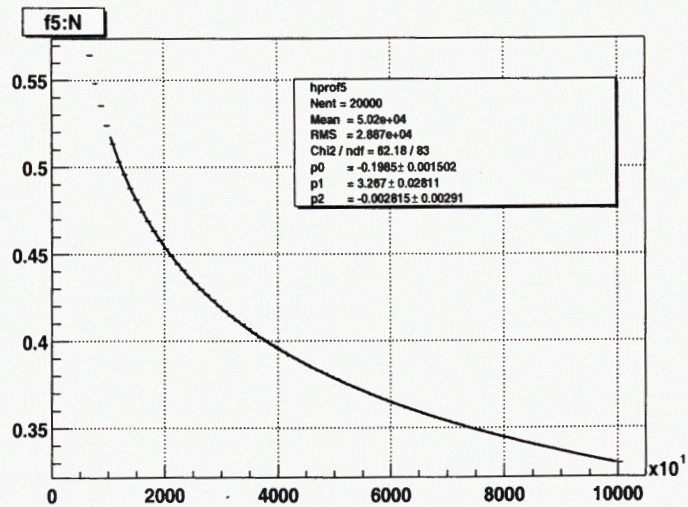


Рис. 7:  $f_5$  vs.  $N$

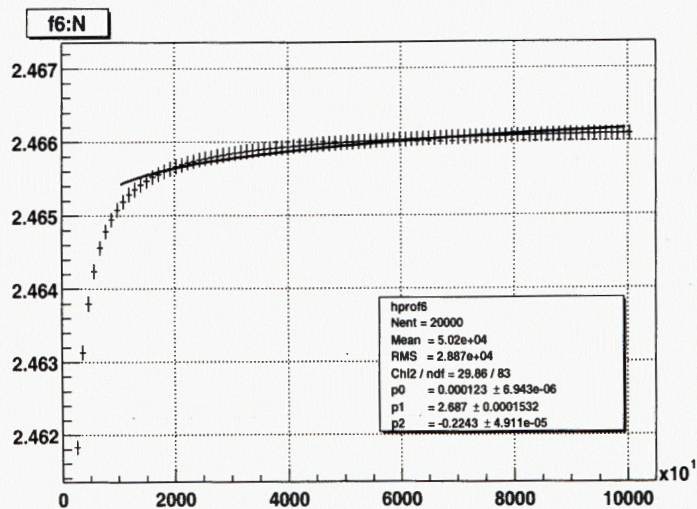


Рис. 8:  $f_6$  vs.  $N$

## Список литературы

- [1] А.Н. Сисакян, *Явление термализации в физике адронов*, УФН, том 173, вып. 3, 328–332 (2003).
- [2] A. Bialas, V. Koch, *Event-by-event fluctuations and inclusive distributions*, Phys. Lett. B 456 (1999) 1.
- [3] A. G. Grozin, *Using REDUCE in High Energy Physics*, Cambridge University Press, 2001.
- [4] В.Ф. Еднерал, А.П. Крюков, А.Я. Родионов, *Язык аналитических вычислений REDUCE*, МГУ, 1989.

Получено 18 июля 2003 г.



**ТЕМАТИЧЕСКИЕ КАТЕГОРИИ ПУБЛИКАЦИЙ  
ОБЪЕДИНЕННОГО ИНСТИТУТА  
ЯДЕРНЫХ ИССЛЕДОВАНИЙ**

Индекс	Тематика
1.	Экспериментальная физика высоких энергий
2.	Теоретическая физика высоких энергий
3.	Экспериментальная нейтронная физика
4.	Теоретическая физика низких энергий
5.	Математика
6.	Ядерная спектроскопия и радиохимия
7.	Физика тяжелых ионов
8.	Криогеника
9.	Ускорители
10.	Автоматизация обработки экспериментальных данных
11.	Вычислительная математика и техника
12.	Химия
13.	Техника физического эксперимента
14.	Исследования твердых тел и жидкостей ядерными методами
15.	Экспериментальная физика ядерных реакций при низких энергиях
16.	Дозиметрия и физика защиты
17.	Теория конденсированного состояния
18.	Использование результатов и методов фундаментальных физических исследований в смежных областях науки и техники
19.	Биофизика

Стадник А. В., Чернов Н. И., Шиманский С. С.  
«Корреляторы» и «моменты»

P11-2003-143

Представлен новый метод вычисления «корреляторов» через центральные моменты, который позволил не только существенно сократить время вычислений, но и сделать правильную оценку асимптотического поведения «корреляторов».

Проведено численное моделирование с простой функцией плотности распределения вероятности, подтверждающее аналитические оценки.

Работа выполнена в Лаборатории высоких энергий им. В. И. Векслера и А. М. Балдина ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна, 2003

Перевод авторов

Stadnik A. V., Chernov N. I., Shimanskiy S. S.  
«Correlators» and «Moments»

P11-2003-143

In this paper we present a new method for computing «correlators» through central moments. This not only allows us to substantially reduce the time of computation, but also leads to correct asymptotic estimates for the «correlators».

A computer experiment involving a simple density function is in good agreement with our analytical results.

The investigation has been performed at the Veksler–Baldin Laboratory of High Energies, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna, 2003