# ПРОБЛЕМЫ АЛГОРИТМИЗАЦИИ ВЫЧИСЛЕНИЯ БОЛЕЕ ВЕРОЯТНЫХ В СРЕДНЕМ ВЕЛИЧИН КРИТЕРИЕВ В ПЛП-ПОИСКЕ

***Статников И.Н., Фирсов Г.И.***

**Институт машиноведения им. А.А. Благонравова РАН, г. Москва**

***firsovgi@mail.ru***

***Рассматриваются основные идеи планируемого ЛП-поиска (ПЛП-поиска) и показано, что для реальных параметров вычислительных экспериментов можно заменить действительное значение дисперсии выборки в произвольном сечении ее оценкой. Рассматриваются проблемы, возникающие при выборе числа сочетаний при оценке более вероятных в среднем величин критериев, связанные с противоречием между желаемой точностной вероятностной оценкой и временем ее получения. Приведены результаты применения предлагаемого алгоритма и вариации в его использовании на ряде тестовых примеров.***

Общая идея построения более вероятного значения оценки математического ожидания гипотетической генеральной совокупности основывается на классическом принципе теории вероятностей, говорящем о тот, что распределение сумм случайных величин, имеющих конечные значения дисперсий, независимо от характера их исходных распределений, стремится к нормальному при неограниченном увеличении числа этих сумм [1,2]. Конечно, сочетание слов «более вероятное» сохраняет неустранимую неопределенность в величине среднего генеральной совокупности. Это обстоятельство компенсируется, в значительной мере, выбором значений двух параметров: величиной доверительного интервала и соответствующей ему величиной доверительной вероятности.

Применительно к планируемому ЛП-поиску (ПЛП-поиску) [3,4] общая идея состоит в следующем. Для каждого $k$-го критерия качества проектируемого (исследуемого) объекта имеется $\sum_{j=1}^{J} M_{kj}$ выборок $\{\Phi_{kji_j l_{ki_j}}\}$, где: $\Phi_{kji_j l_{ki_j}}$ - $l_{ki_j}$-ое значение $k$-го критерия на $i_j$-м уровне $j$-го параметра; $l_{ki_j} = \overline{1, L_{ki_j}}$, а $L_{ki_j}$ - число членов выборки на $i_j$-м уровне $j$-го параметра по $k$-му критерию; $k = \overline{1, K}$, а $K$ - число анализируемых критериев качества; $j = \overline{1, J}$, а $J$ - число варьируемых параметров; $i_j = \overline{1, M_{kj}}$, а $M_{kj}$ - число уровней $j$-го параметра по $k$-му критерию. Из множеств $\{\Phi_{kji_j l_{ki_j}}\}$ в ПЛП-поиске формируются подмножества $\{\bar{\Phi}_{kji_j}\}$ средних значений, необходимых для проведения однофакторного дисперсионного анализа. Здесь $\bar{\Phi}_{kji_j}$ - среднее значение $k$-го критерия на $i_j$-м уровне $j$-го параметра:

$$\bar{\Phi}_{kji_j} = (L_{ki_j})^{-1} \sum_{l_{ki_j}}^{L_{ki_j}} \Phi_{kji_j l_{ki_j}}.$$

График функции $\bar{\Phi}_{kji_j}(\alpha_{i_j})$ в сопоставлении со значением $\bar{\Phi}_k$ - общим средним значением $k$-го критерия, полученным по выборке из $N_0$ вычислительных экспериментов, позволял эмпирически (чаще всего, визуально) назначать новые области поиска наилучших решений по $k$-му критерию. Однако неучет того факта, что значения $\bar{\Phi}_{kji_j}$ носят выборочный характер, часто приводят к необоснованному оптимизму при выборе новых пределов варьирования параметра $\alpha_j$ и, что самое главное, к раздражительной реакции пользователя ПЛП-поиска. Хотелось бы побольше «детерминизма», хотя бы и в среднем. Можно ли было этого достичь, не увеличивая общего числа $N_0$ вычислительных экспериментов? Можно, если воспользоваться общей идеей построения более вероятного среднего.

Суть предлагаемых алгоритмов вычисления более вероятных в среднем величин критериев состоит в следующем. Имеется выборка из $l_{ki_j}$ значений $\Phi_{kji_jl_{ki_j}}$. Организуем всевозможные суммы (а, далее, средние этих сумм) из элементов этой выборки. Как известно [5], число всевозможных сочетаний $M_{ki_j}$ из элементов этой выборки равно

$$M_{ki_j} = (\sum_{m=2}^{l_{ki_j}-1} C^m_{l_{ki_j}}) = 2^{l_{ki_j}} - (l_{ki_j} + 2) \quad (1)$$

без учета среднего значения исходной выборки и

$$M_{ki_j} = 2^{l_{ki_j}} - (l_{ki_j} + 1) \quad (2)$$

с учетом этого среднего значения. Уже при $l_{ki_j} \geq 8$ число $M_{ki_j} \geq 246$. Поэтому с ростом $l_{ki_j}$ статистика $M_{ki_j}$ растет быстро и появляется возможность использовать более "качественные" суммы, т.е. организовать суммы из 10 и более элементов, а не из 2, 3 и т.д. Это обстоятельство, конечно, учитывается при программной реализации алгоритма.

Обозначим более вероятные в среднем значения *k*-го критерия на $i_j$ -ом уровне *j*-го параметра через $a_{kji_j}$. Тогда рассматривая $a_{kji_j}$ и $\sigma_{kji_j}$ как неизвестные параметры нормального распределения и решая уравнения правдоподобия [1], получим несмещенные и эффективные оценки этих параметров:

$$a_{ki_j} = (M_{ki_j})^{-1} \sum_{m_{ki_j}}^{M_{ki_j}} \bar{\Phi}_{ki_jm_{ki_j}} = \bar{\Phi}^{BB}_{ki_j} \quad (3)$$

и

$$\sigma^2_{ki_j} = \frac{1}{M_{ki_j} - 1} \sum_{m_{ki_j}}^{M_{ki_j}} (\bar{\Phi}_{ki_jm_{ki_j}} - \bar{\Phi}^{BB}_{ki_j})^2 = (S^{BB}_{ki_j})^2, \quad (4)$$

где $\bar{\Phi}_{ki_jm_{ki_j}}$ - средние значения частных сумм, составленных из *m* элементов в $l_{ki_j}$ членов ( $m = \overline{1, l_{ki_j} - 1}$ ). При этом доверительным интервалом для $a_{kji_j}$ будет интервал ( $\bar{\Phi}^{BB}_{ki_j} - S^{BB}_{ki_j} c_\alpha, \bar{\Phi}^{BB}_{ki_j} + S^{BB}_{ki_j} c_\alpha$ ), где $c_\alpha = t_\alpha \sqrt{M_{ki_j}}$ , а $t_\alpha$ - значение интеграла вероятности при данном значении доверительной вероятности *P* = 1-α. Так при α = 0,01, $M_{ki_j}$ = 1024 и $t_\alpha$ = 2,58, получим $c_\alpha$ = 2,58/32 ≈ 0,0806. Отсюда $a_{ki_j} \in (\bar{\Phi}^{BB}_{ki_j} - 0,0806 S^{BB}_{ki_j}, \bar{\Phi}^{BB}_{ki_j} + 0,0806 S^{BB}_{ki_j})$. Если же $M_{ki_j}$ =$10^4$, то при том же значении α = 0,01 находим, что $a_{ki_j} \in (\bar{\Phi}^{BB}_{ki_j} - 0,0258 S^{BB}_{ki_j}, \bar{\Phi}^{BB}_{ki_j} + 0,0258 S^{BB}_{ki_j})$.

И для программной реализации алгоритма, и в целях научной добросовестности следует объяснить, почему при вычислении доверительного интервала $a_{kji_j}$ мы используем значения $S^{BB}_{ki_j}$, а не $\sigma^2_{ki_j}$ (как правило, неизвестное). Дело в том, что и для $\sigma^2_{ki_j}$ можно сразу записать четыре неравенства [6,7], которые выполняются одновременно с вероятностью 1-2α:

$$(\sigma_{ki_j} / \lambda) < z_1 S^{BB}_{ki_j} < \sigma_{ki_j} < z_2 S^{BB}_{ki_j} < \lambda \sigma_{ki_j}, \quad (5)$$

где λ = 1+*q*, а *q* - уровень относительной погрешности в %, устраивающий исследователя [6]:

$z_1 = \sqrt{M_{ki_j} / \chi^2_{M_{ki_j};\alpha/2}}$, $z_2 = \sqrt{M_{ki_j} / \chi^2_{M_{ki_j};1-\alpha/2}}$ ,

а

$\chi^2 = (S^{BB}_{ki_j})^2_1 + (S^{BB}_{ki_j})^2_2 + ... + (S^{BB}_{ki_j})^2_n -$

случайная величина, подчиненная $\chi^2$ -распределению с ν степенями свободы ($\nu = M_{ki_j}$, если $a_{kji_j}$ известна, и $\nu = M_{ki_j} - 1$, если этот параметр неизвестен). И если $M_{ki_j} \to \infty$, то $z_1$ и $z_2$ в (5) стремятся к 1. Реально это положение можно проиллюстрировать, выписав из [7, табл. 3.3, стр. 59] часть таблицы, где для ряда величин P = 1-2α и соответствующих чисел степеней свободы ν указаны числовые значения λ:

**Таблица 1**

| | **P = 1-2α** | | | |
|---|---|---|---|---|
| **ν** | **0,50** | **0,90** | **0,95** | **0,99** |
| 100 | 1,100 | 1,263 | 1,321 | 1,443 |
| 1000 | 1,031 | 1,076 | 1,092 | 1,122 |
| 5000 | 1,014 | 1,033 | 1,040 | 1,053 |
| 10000 | 1,010 | 1,024 | 1,028 | 1,037 |

Из этой таблицы следует, что уже при ν ≥ 100 (т.е. фактически при $M_{ki_j} \geq 10^3$) значения λ и 1/λ стремятся к 1 уже даже при большой величине *P*. И, например, при ν ≥ 10000 и *P* = 0,99 (т.е. при α = 0,005) крайние границы неравенств в (5) приобретают соответственно значения $0{,}964\,\sigma_{kji_j}$ и $1{,}037\,\sigma_{kji_j}$. И так как в реализуемом алгоритме [8] мы всегда будем иметь дело с выборками, у которых $M_{ki_j} \geq 10^3$ (т.е. $L_{ki_j} \geq 10$), то это избавляет нас от необходимости реализовывать в программе построение доверительного интервала (confidence interval) для $S^{BB}_{ki_j}$, т.е. будем молчаливо предполагать, что $(S^{BB}_{ki_j})^2 \approx \sigma^2_{ki_j}$.

Однако воспользоваться формулой (3) для получения более вероятных значений $a_{ki_j}$, вычислив все сочетания в соответствии с формулой (1) (или (2)) нельзя, так как показано, что

$$(L_{ki_j})^{-1}\sum_{l_{ki_j}}^{L_{ki_j}} \Phi_{kjl_j l_{ki_j}} = [2^{L_{ki_j}} - (L_{ki_j} + 2]^{-1}(\sum_{l_{ki_j}}^{L_{ki_j}} \Phi_{kjl_j l_{ki_j}})(\sum_{m=2}^{L_{ki_j}-1} \frac{1}{m} C^{m-1}_{L_{ki_j}-1}).$$

Конечно, при больших значениях $L_{ki_j}$ (например, при $L_{ki_j} \geq 15$), когда величины $M_{ki_j}$ достигают значений в несколько десятков тысяч единиц, воспользоваться формулой (3) можно, использовав только часть сочетаний.

Для получения более вероятных средних значений $a_{ki_j}$ был рассмотрен и другой алгоритм. Пусть величина $y_u$ принадлежит нормальному распределению ($a$, σ). Тогда можно записать, что

$$\ln y_u = \ln \frac{A}{\sigma} - \frac{(x_u - a)^2}{2\sigma^2}. \qquad (6)$$

Подставляя в (6) значения ($y_u$, $x_u$) в $u$-ой, $u$+1-ой и в $u$+2-ой точках, решая полученную таким образом систему уравнений, найдем:

$$\sigma^2 = (2\Delta)^{-1}(x_u - x_{u+2})(x_{u+1} - x_u)(x_{u+2} - x_{u+1}) \qquad (7)$$

и

$$a = (2\Delta)^{-1}[(x^2_{u+2} - x^2_{u+1}) \ln \frac{y_u}{y_{u+1}} - (x^2_{u+1} - x^2_u) \ln \frac{y_{u+1}}{y_{u+2}}], \qquad (8)$$

где $x_u < x_{u+1} < x_{u+2}$, а

$$\Delta = (x_{u+2} - x_{u+1}) \ln \frac{y_u}{y_{u+1}} - (x_{u+1} - x_u) \ln \frac{y_{u+1}}{y_{u+2}}.$$

Для Δ = const выводится условие, когда Δ < 0 (что необходимо для работоспособности формул (7) и (8). Обозначим $x_{u+2}$ - $x_{u+1}$ = $\Delta_{u+1}$ , а $x_{u+1}$ - $x_u$ =$\Delta_u$. Тогда при Δ < 0 имеем

$\ln(\frac{y_u}{y_{u+1}})^{\Delta_{u+1}} - \ln(\frac{y_{u+1}}{y_{u+2}})^{\Delta_u} < 0$, далее $\ln\frac{(y_u)^{\Delta_{u+1}}(y_{u+2})^{\Delta_u}}{(y_{u+1})^{\Delta_{u+1}+\Delta_u}} < 0$ или $(y_u)^{\Delta_{u+1}}(y_{u+2})^{\Delta_u} < (y_{u+1})^{\Delta_{u+1}+\Delta_u}$. При $\Delta_u$ =const для $\forall u$ получим:

$$y_u y_{u+2} - y_{u+1}^2 < 0. \quad (9)$$

Для уменьшения влияния выбросов на величину *a*, рассчитываемую по (8), условие (9) заменим более надежным $y_u y_{u+2} - y_{u+1}^2 \le -\varepsilon_u$ , где $0 < \varepsilon_u << 1$. Построив гистограмму распределения средних величин из выборки в $M_{ki_j}$ членов по *s* разрядам, мы можем последовательно, пользуясь формулами (7) и (8), обойти всю гистограмму, и тогда $u = \overline{1, s-2}$. А можем перебрать последовательно все разряды гистограммы, используя по три разряда в каждом вычислении. В обоих вариантах искомые оценки получаются как результат усреднения по множеству величин, полученных с помощью формул (7) и (8). Конечно, точность в определении $a_{nkJ}$ ($a_{nkJ} \equiv \tilde{a}$) зависит как от величины $M_{ki_j}$ (чем больше значение $M_{ki_j}$, тем точнее определяется величина $a_{nkJ}$), так и от числа разрядов *s* при построении гистограммы по выборке из $M_{ki_j}$ членов. Возникает проблема такого выбора *s*, чтобы построенная гистограмма достаточно хорошо аппроксимировала неизвестное нормальное распределение, тем самым способствуя более точному определению координат вершины нормального распределения, т.е. искомой величины $a_{nkJ}$. Иначе говоря, нужен критерий оптимального выбора величины *s*. Сформируем такой критерий оптимальности на основе следующей леммы из [9, стр. 411]: уклонение гистограммы случайной величины от графика ее плотности в метрике $Q^2$, когда эта плотность имеет ограниченную вторую производную, в лучшем случае имеет порядок $(M_{ki_j})^{-1/3}$ (соответственно, квадрат нормы уклонения - порядок $(M_{ki_j})^{-2/3}$), и он достигается при числе интервалов группировки $s \approx (M_{ki_j})^{-1/3}$.

Интерес представляют и более вероятные в среднем значения *k*-го критерия в каждой из *n*-х *J*-мерных точек ($n = \overline{1, N_0}$) на предмет выявления потенциально экстремальных точек (min или max). В этом случае в каждой *n*-ой точке мы имеем совокупность средних значений $\{a_{ki_j}\}$, также являющихся выборкой из нормальной совокупности с параметрами $a_{nkJ}$ и $\sigma_{nkJ}$, где $n = \overline{1, N_0}$. Оценки этих параметров вычислим аналогично (3) и (4):

$$a_{nkJ} = (J)^{-1}\sum_{j=1}^{J} a_{ki_j} = (JM_{ki_j})^{-1}\sum_{j=1}^{J}\sum_{m_{ki_j}=1}^{M_{ki_j}} \bar{\Phi}_{nkJ}^{BB},$$

и

$$\sigma_{nkJ}^2 = (JM_{ki_j} - 1)^{-1}[\sum_{j=1}^{J}\sum_{m_{ki_j}=1}^{M_{ki_j}} \bar{\Phi}_{ki_j m_{ki_j}}^2 - (JM_{ki_j})(\bar{\Phi}_{nkJ}^{BB})^2] = (S_{nkJ}^{BB})^2,$$

Конечно, при реальных значениях $M_{ki_j}$ (≥ 1000) и J ≥ 2 оценка

$$(S_{nkJ}^{BB})^2 \approx (JM_{ki_j})^{-1}(\sum_{j=1}^{J}\sum_{mki_j}^{M_{ki_j}} \tilde{\Phi}_{ki_j m_{ki_j}}^2) - (JM_{ki_j})(\bar{\Phi}_{nkJ}^{BB})^2$$

уже является приемлемой по точности.

При практической реализации алгоритма построения более вероятных в среднем величин функции чувствительности *k* –го критерия по *j*-–му параметру и более вероятных в среднем значений *k* –го критерия в *J* –мерной точке в ПЛП-поиске [3,4], понятно, что для различных значений числа уровней *j*-го параметра по *k*-му критерию. $M_{ki_j}$ следует учитывать различные количества сочетаний, тем более, что с ростом числа членов выборки на $i_j$ -м уровне *j*-го

параметра по *k*-му критерию $L_{ki_j}$ значение $M_{ki_j}$ растет $M_{ki_j} = (\sum_{m=2}^{L_{ki_j}-1} C^m_{L_{ki_j}}) = 2^{L_{ki_j}} - (L_{ki_j} + 2)$ довольно быстро (табл. 2):

**Таблица 2**

| $L_{ki_j}$ | 8 | 10 | 12 | 15 | 16 | 20 |
|---|---|---|---|---|---|---|
| $M_{ki_j}$ | 246 | 1012 | 4082 | 32751 | 65508 | 1048554 |

Из табл. 2 следует, что если получение более вероятного среднего значения гипотетической генеральной совокупности основывается на данных одной выборки из этой совокупности, то скупиться на число вычислений не следует, и ограничение накладывается только возможностями ЭВМ, в основном, временными. Но при использовании описываемого алгоритма в ПЛП-поиске нужно идти на компромиссы при выборе числа сочетаний между желаемой точностной вероятностной оценкой и временем ее получения. Выбор такого компромиссного числа сочетаний при длине выборки в $L_{ki_j}$ членов остается за пользователем ПЛП-поиска. Пользователь такой выбор производит в диалоге с ЭВМ, где ему будет показано, что при таком-то значении *m* длина рассчитываемой выборки составит $M_{ki_j}$ членов. Очевидно, что если $L_{ki_j} \le 15$ (табл. 2), то для расчета по формулам

$$a_{ki_j} = (M_{ki_j})^{-1} \sum_{m_{ki_j}}^{M_{ki_j}} \bar{\Phi}_{ki_j m_{ki_j}} = \bar{\Phi}^{BB}_{ki_j} \qquad (10)$$

и $\sigma^2_{ki_j} = \frac{1}{M_{ki_j} - 1} \sum_{m_{ki_j}}^{M_{ki_j}} (\bar{\Phi}_{ki_j m_{ki_j}} - \bar{\Phi}^{BB}_{ki_j})^2 = (S^{BB}_{ki_j})^2$, где $\bar{\Phi}_{ki_j m_{ki_j}}$ - средние значения частных сумм, составленных из *m* элементов в $L_{ki_j}$ членов ($m = \overline{1, L_{ki_j} - 1}$) следует придать *m* все значения от 2 до ($L_{ki_j}$ - 1). С другой стороны, если $L_{ki_j}$ все же велико (например, $L_{ki_j} \ge 30$), то при уверенности в качественности «случайных» членов выборки, бывает достаточным для получения более вероятного значения среднего (и среднеквадратичного отклонения) гипотетической генеральной совокупности ограничиться значениями *m* = 2;3. Рассмотрим результаты применения предлагаемого алгоритма и вариации в его использовании на ряде тестовых примеров, взятых из различных источников [10-12].

Пример 1. В книге [10] описывается эксперимент, заимствованный из [11]. Суть эксперимента состоит в случайном извлечении карточек, пронумерованных от 1 до 10, из десяти пачек, каждая из которых состояла из десяти карточек. Эксперимент проводился для иллюстрации действия закона больших чисел. В табл. 2.1 [10 стр. 21] приведена таблица распределения выборочных сумм. Средняя всей исходной («генеральной») совокупности экспериментов известна наперед: $a_т$ = 55. Результаты экспериментов безусловно подтвердили суть закона больших чисел: от таблички к табличке в табл. 2.1 (т.е. с увеличением объема выборок) размах колебаний разности между наименьшей и наибольшей случайной величинами постоянно уменьшается. При работе с предлагаемым алгоритмом была использована первая табличка табл. 2.1 ($L_{ki_j}$ = 30). Средняя величина это выборки равнялась $\tilde{a}_1$ = 55,2, что по сравнению с $a_т$ дает относительную ошибку $\delta_1$ = 0,364%. Сформированная из первой таблички новая выборка ($M_{ki_j}$ = 435), состоявшая из средних значений всех парных сумм элементов ($C^2_{30}$ = 435), дала среднее значение $\tilde{a}_2$ = 55,216 с относительной

ошибкой $\delta_2$ = 0,393%. Определив размах выборки $\Delta\Phi = \Phi_{max} - \Phi_{min}$ = 20,92, делили его на различное число разрядов *s*, применяя затем формулы

$$\sigma^2 = (2\Delta)^{-1}(x_u - x_{u+2})(x_{u+1} - x_u)(x_{u+2} - x_{u+1}) \qquad (11)$$

и

$$a = (2\Delta)^{-1}[(x_{u+2}^2 - x_{u+1}^2)\ln\frac{y_u}{y_{u+1}} - (x_{u+1}^2 - x_u^2)\ln\frac{y_{u+1}}{y_{u+2}}], \qquad (12)$$

где величина $y_u$ принадлежит нормальному распределению ($a$, $\sigma$), $x_u < x_{u+1} < x_{u+2}$, а $\Delta = (x_{u+2} - x_{u+1})\ln\frac{y_u}{y_{u+1}} - (x_{u+1} - x_u)\ln\frac{y_{u+1}}{y_{u+2}}$. В итоге получили табл. 3. Далее, пользуясь леммой из [9], выбрали $\delta(\Delta\Phi) = \Phi_{max} - \Phi_{min}$ = 2,761 и, начиная от $\Phi = \Phi_{min}$, покрыли весь интервал $\Delta\Phi$ кусочками $\delta(\Delta\Phi)$. При этом $s = s_{opt}$ = 8, а расчеты по формулам (11) и (12) дали $\tilde{a}$ = 55,463 и $\delta$ = 0,84%.

**Таблица 3**

| **s** | **4** | **5** | **8** | **10** |
|---|---|---|---|---|
| **Относительная ошибка δ, %** | 1,38 | 1,77 | 1,76 | 0,78 |

Пример 2. В этом примере [10], заимствованном из [11], представлено эмпирическое распределение 511 выборочных дисперсий. Каждая выборка состояла из 10 элементов. В табл. 2.2 [10, стр. 25] по существу представлена гистограмма этого распределения *s* = 13). При этом точное значение средней величины дисперсии $a_т$ = 100. Трудности использования предлагаемого алгоритма в рассматриваемом примере состояли в том, что не было исходных данных, а в качестве таковых пользовались данными гистограммы [10, табл. 2.2], чем ввели в расчеты неизвестную некую систематическую ошибку.

**Таблица 4**

| **Условия эксперимента (типы используемых сочетаний)** | $C_{13}^2$ | $C_{13}^2 + C_{13}^3$ | $C_{13}^2 + C_{13}^3 + C_{13}^4$ | $C_{13}^2 + C_{13}^3 + C_{13}^4$ + часть элементов из $\{C_{13}^5\}$ |
|---|---|---|---|---|
| **Длина выборки** $M_{ki_j}$ | 78 | 364 | 1079 | 1578 |
| **Относительная ошибка δ, %** | 45 | 42 | 41 | 36 |

В табл. 4 представлены результаты расчетов более вероятных средних значений дисперсии по формуле (10) и относительные ошибки для совокупностей, составленных из частных средних, когда брались суммы в сочетаниях из двух, трех, четырех элементов и, частично, из пяти элементов. Можно наблюдать весьма малую сходимость к точному значению с ростом $M_{ki_j}$, что, конечно, объясняется как малым числом элементов, используемых при составлении частных сумм, так и «плохой» случайностью используемых данных (например, правый «хвост» в табл. 2.2 придает сильно асимметричный характер гистограмме). Другой подход в получении более вероятного среднего значения дисперсии по исходным данным (табл. 2.2 в [10]) состоит в использовании формул (2) и (3). В дальнейшем такой подход будем условно именовать МЕТОДОМ «ТРЕХ ТОЧЕК». В этом случае происходит последовательный перебор всех разрядов гистограммы, используя по три в каждом вычислении. В этом случае число полученных значений $a$ будет равно $M = 1 + (s-3)(s^2+6)/6$. И в этом случае при умеренных значениях *s*, рекомендуемых в различных руководствах по математической статистике, получаются не очень большие значения *M* (табл. 5), но достаточные для получения из множества значений $\{a\}$ ее средней величины $\tilde{a}$, принимаемой за искомую.

Для выборки, составленной только из парных сочетаний, первоначально строилась гистограмма для $s$ = 10 разрядов. При последовательном обходе этой гистограммы методом «трех точек» получили подмножество значений $\{a\}$, средняя величина которых равна 145,82, что дает относительную ошибку δ ≈ 46%. При попарном последовательном объединении разрядов этой гистограммы ($s$ = 5) и полном переборе методом «трех точек» ($M$ = 10) получили $\tilde{a}$ = 124 и δ = 24%. Характерным для этой гистограммы ($s$ = 5) было то обстоятельство, что отношения количеств элементов, накопленных в каждом разряде, к максимальному количеству элементов, накопленному в центральном разряде, были больше величины 0,2 (10/25, 21/25, 25/25, 15/25 и 7/25). При выборе величины разряда $\delta(\Delta\Phi) = (\Phi_{max} - \Phi_{min}) / \sqrt[3]{78}$ получилось также 5 разрядов ($s$ = 5), и при $M$ = 10 величина $\tilde{a}$ = 139 и δ = 39%. В этой гистограмме (с шагом δ(ΔΦ)) в 5-ый разряд попало всего одно значение Φ, что при делении на максимальное число элементов во 2-ом разряде гистограммы (равное 28) дает величину 1/28 < 0,04. При объединении последнего разряда с предыдущим получили $s$ = 4 и при полном переборе методом «трех точек» ($M$ = 4) достигли $\tilde{a}$ = 135,25 и δ ≈ 35,3%.

**Таблица 5**

| **s** | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|
| **M** | 4 | 10 | 20 | 35 | 56 | 81 | 120 | 165 | 220 | 286 |

Далее анализировалась выборка длиной в $M_{ki_j} = C_{13}^2 + C_{13}^3 = 364$ элемента. При равномерном разбиении интервала ($\Phi_{max}$ - $\Phi_{min}$) на десять разрядов ($s$ = 10) получили $\tilde{a}$ = 157,27 и δ ≈ 57,3%. Когда в этой гистограмме верхний и нижний разряды объединили соответственно с последующим и предыдущим разрядами ($s$ = 8), то получили $\tilde{a}$ = 137,83 и δ ≈ 38%. Когда в этой же гистограмме ($s$ = 10) последовательно объединили по два разряда (стало $s$ = 5), то при полном переборе полученной гистограммы методом «трех точек» ($M$ = 10) получили результаты $\tilde{a}$ = 142,94 и δ ≈ 43%. При выборе длины разряда, равной $\delta(\Delta\Phi) = (\Phi_{max} - \Phi_{min}) / \sqrt[3]{364}$ получили $s$ = 7 и при последовательном обходе методом «трех точек» ($M = s$ - 2) нашли: $\tilde{a}$ = 141,2 и δ ≈ 41,2%. При объединении последнего разряда этой же гистограммы с предыдущим ($s$ = 6), тем же способом получили: $\tilde{a}$ = 138 и δ ≈ 38%.

Анализировалась выборка длиной $M_{ki_j} = C_{13}^2 + C_{13}^3 + C_{13}^4 = 1079$ элементов. Вначале весь интервал изменения значений Φ был разбит на 13 разрядов, в результате чего после последовательного обхода гистограммы методом «трех точек» получили: $\tilde{a}$ = 71,62 и δ ≈ 28%. Так как отношения количеств элементов в самом верхнем и самом нижнем разрядах гистограммы (равное 209) были соответственно 3/209 < 0,02 и 2/209 < 0,01, то произвели объединение первых двух верхних разрядов и последних нижних двух разрядов (получили $s$ = 11 и соответствующие отношения: 14/209 > 0,05 и 8/209 > 0,03). При последовательном обходе этой гистограммы методом «трех точек» достигли таких результатов: $\tilde{a}$ = 95,81 и δ ≈ 4,2%. При выборе длины разряда гистограммы по формуле $\delta(\Delta\Phi) = (280 - 20) / \sqrt[3]{M_{ki_j}}$ ≈ 25,524 весь интервал разбился на $s$ = 10 разрядов, и при последовательном обходе этой гистограммы методом «трех точек» получили $\tilde{a}$ = 128,68 и δ ≈ 28,7%.

В дальнейшем к анализируемой выборке была добавлена еще порция элементов, составленная из средних значений части сумм, образованных сочетаниями элементов из тринадцати по пять, так что $M_{ki_j}$ = 1560. Сначала интервал ΔΦ = $\Phi_{max}$ - $\Phi_{min}$ равномерно разделили на $s$ = 12 разрядов. При последовательном обходе гистограммы методом «трех точек» получили $\tilde{a}$ = 151,24 и δ ≈ 51,2%. Затем в этой гистограмме объединили два верхних и два нижних разряда, получив $s$ = 10. В этом случае при последовательном обходе

гистограммы методом «трех точек» получили $\tilde{a}$ = 115 и δ = 15%. При полном обсчете исходно построенной гистограммы ($s$ = 12) методом «трех точек» ($M$ = 120) получили $\tilde{a}$ = 118,62 и δ ≈ 18,6%. Далее, исходя из оптимальной оценки числа разрядов была вычислена величина $\delta(\Delta\Phi) = \Delta\Phi / \sqrt[3]{1560} \approx 20{,}9523$ и построена новая гистограмма. В этой гистограмме было также получено $s$ = 12 разрядов и при последовательном обходе гистограммы методом «трех точек» получили $\tilde{a}$ = 140,79 и δ ≈ 41%. При слиянии в этой гистограмме первых двух последних разрядов ($s$ = 10) и при последовательном обходе вновь полученной гистограммы методом «трех точек» было найдено, что $\tilde{a}$ = 106,3 и δ ≈ 6,3%.

Наконец, методом «трех точек» результаты были получены и на исходной гистограмме [10, табл. 2.2], у которой был отброшен последний 13-й разряд, т.е. $s$ = 12. При полном обсчете этой гистограммы методом «трех точек» ($M$ = 220) было получено $\tilde{a}$ = 72,75 и δ ≈ 27,3%. При последовательном обходе этой гистограммы методом «трех точек» ($M = s - 2$ = 10) было получено $\tilde{a}$ = 115,66 и δ ≈ 16%. Когда у этой гистограммы два последних нижних разряда объединили в один, то при последовательном обходе гистограммы методом «трех точек» получили $\tilde{a}$ = 97,91 и δ ≈ 2,1% ($M = s - 2 = 9$). Такие неплохие результаты, полученные методом «трех точек», легко объясняются тем, что исходные выборки, по которым рассчитывали выборочные дисперсии, содержали по 10 элементов [10, 11], что, естественно, придало выборочным оценкам более представительный («качественный») характер.

Пример 3. Рассматривались две выборки из последовательности равномерно распределенных по вероятности чисел на интервале (0,1). Математическое ожидание у этой последовательности $a_т$ = 0,5. Первая выборка ($L_{ki_j}$ = 5) {0,8125; 0,3125; 0,5625; 0,0625; 0,4375} дает среднее значение $\tilde{a}$ = 0,4375 и δ = 12,5%, а вторая выборка ($L_{ki_j}$ = 6) {0,8125; 0,3125; 0,5625; 0,0625; 0,4375; 0,9375} дает среднее значение $\tilde{a}$ = 0,5208 и δ ≈ 4,16%. Из этих обеих выборок были сформированы новые выборки, состоящие из всевозможных частных сумм (средних), длина которых по формуле $M_{ki_j} = 2^{L_{ki_j}} - (L_{ki_j} + 1)$, где $M_{ki_j}$ - число уровней $j$-го параметра по $k$-му критерию, $L_{ki_j}$ - число членов выборки на $i_j$ -м уровне $j$-го параметра по $k$-му критерию, равнялась соответственно $M_{ki_j}$ =26 и $M_{ki_j}$ = 57. Результаты анализа этих выборок методом «трех точек» при произвольном назначении числа разрядов $s$ в каждой выборке отражены в табл. 6. Из этой таблицы следует, что при больших значениях $L_{ki_j}$, а, значит, и $M_{ki_j}$, результаты лучше. Выбирая для первой выборки ($M_{ki_j}$ = 26) шаг гистограммы $\delta(\Delta\Phi) = (0{,}6875 - 0{,}1875) / \sqrt[3]{26} \approx 0{,}16878$, получили $s$ =3 и $\tilde{a}$ = 0,4313 при δ ≈ 13,73%. Изменяя в этой гистограмме шаг δ(ΔΦ) на ±0,01δ(ΔΦ), получили практически те же результаты, что и вышеприведенные. Для второй выборки $\delta(\Delta\Phi) = (0{,}875 - 0{,}1875) / \sqrt[3]{57} \approx 0{,}17864$. Получилась гистограмма с $s$ = 4 и при обсчете ее методом «трех точек» нашли: $\tilde{a}$ = 0,5317 и δ ≈ 6,34%. При уменьшении шага этой гистограммы на 0,01δ(ΔΦ) получили снова $s$ = 4 и при этом: $\tilde{a}$ = 0,5109 и δ ≈ 2,18%. При увеличении шага гистограммы на 0,01δ(ΔΦ) снова получили $s$ = 4 и при этом: $\tilde{a}$ = 0,5269 и δ ≈ 5,38%. Если усреднить все три результата (исходной гистограммы при δ(ΔΦ) =0,17864, гистограммы с уменьшенным шагом и гистограммы при увеличенным шаге), то получим: $\tilde{a}$ = 0,5232 и δ ≈ 4,64%. Мы получили лучший результат, чем в табл. 4, когда при $M_{ki_j}$ = 57 весь интервал изменения ΔΦ делился ровно на $s$ = 4 разряда.

**Таблица 6**

| $M_{ki_j}$ | **26** | **57** |
|---|---|---|

| s | 3 | 4 | 5 | 6 | 4 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| $\tilde{a}$ | 0,4370 | 0,4375 | 0,4376 | 0,4169 | 0,4468 | 0,4747 | 0,4712 |
| δ% | 12,60 | 12,50 | 12,48 | 16,62 | 10,68 | 5,07 | 5,57 |

Пример 4. В этом примере [12, табл. 1], почерпнутом из [13], приведены результаты серии испытаний, когда монета подбрасывалась в общей сложности 10000 раз. При этом отдельно рассматривались серии по n = 100 испытаний и в каждой серии регистрировалось соответствующее количество n(Г) выпадений герба. Конечно, результаты этой таблицы [12, табл. 1], прекрасно иллюстрируют действие закона больших чисел, когда вероятности выпадения герба и в малых сериях (n = 100) и в больших (n = 1000) колеблются вокруг теоретического значения ат = 0,5. Анализировались первая и вторая строки (соответственно, первая и вторая выборки) этой таблицы. В обеих выборках по нашей терминологии $L_{ki_j}$ =10, но в первой выборке $\tilde{a}$ = 0,51 и δ = 2%, а во второй выборке - $\tilde{a}$ = 0,485 и δ = 3%. Из первой выборки была сформирована новая выборка, состоящая из средних значений сумм парных сочетаний элементов исходной выборки ( $M_{ki_j}$ = 45). При равномерном разбиении полученного интервала ΔФ на s = 4 и s = 5 разрядов, методом «трех точек» были достигнуты результаты, отраженные в табл. 7. При вычислении шага $\delta(\Delta\Phi) = \Delta\Phi / \sqrt[3]{45} = 14 / \sqrt[3]{45} \approx 3{,}936$ и использовании этого шага при построении гистограммы, снова оказалось s = 4, а оценки были получены такие: $\tilde{a}$ = 0,51535 и δ ≈ 3,07%. Из элементов второй выборки генерировались выборки, состоящие только из парных сочетаний элементов ( $M_{ki_j}$ = 45), и из парных сочетаний элементов и сочетаний типа $C_{10}^3$ ( $M_{ki_j}$ = 45 + 120). Для случая учета только парных сочетаний и равномерного разбиения интервала ΔФ на разряды результаты расчета методом «трех точек» представлены в табл. 8. Для этой же выборки ( $M_{ki_j}$ = 45) при выборе шага $\delta(\Delta\Phi) = \Delta\Phi / \sqrt[3]{45}$ получено $\tilde{a}$ = 0,5198 и δ ≈ 3,96%. В этой же гистограмме при изменении шага δ(ΔФ) на ±0,01δ(ΔФ) были получены соответственно: $\tilde{a}$ = 0,5208 и δ ≈ 4,15% и $\tilde{a}$ = 0,5188 и δ ≈ 3,76%. При усреднении всех трех результатов получили: $\tilde{a}$ = 0,5198 и δ ≈ 3,96%. При $M_{ki_j}$ = 165 и равномерном разбиении интервала ΔФ при различных значениях s результаты приведены в табл. 9. При выборе $\delta(\Delta\Phi) = \Delta\Phi / \sqrt[3]{165}$ оказалось s = 6 и получены такие результаты: $\tilde{a}$ = 0,4898 и δ ≈ 2,04%.

**Таблица 7**

| s | $\tilde{a}$ | δ% |
|---|---|---|
| 4 | 0,5238 | 4,76 |
| 5 | 0,5275 | 5,50 |

**Таблица 8**

| s | $\tilde{a}$ | δ% |
|---|---|---|
| 4 | 0,4928 | 1,44 |
| 5 | 0,5540 | 10,80 |

**Таблица 9**

| S | $\tilde{a}$ | δ% |
|---|---|---|
| 5 | 0,4863 | 2,74 |
| 8 | 0,4838 | 3,23 |
| 9 | 0,5314 | 6,28 |

Рассмотренные модельные (тестовые) примеры и известные теоретические положения в математической статистике позволяют сделать следующие выводы и соответствующие этим выводам рекомендации для программной (машинной) реализации алгоритма:

1) для получения достаточно достоверных оценок по формулам

$$a_{ki_j} = (M_{ki_j})^{-1} \sum_{m_{ki_j}}^{M_{ki_j}} \bar{\Phi}_{ki_j m_{ki_j}} = \bar{\Phi}_{ki_j}^{BB},$$

$$\sigma^2_{ki_j} = \frac{1}{M_{ki_j} - 1} \sum_{m_{ki_j}}^{M_{ki_j}} (\bar{\Phi}_{ki_j m_{ki_j}} - \bar{\Phi}^{BB}_{ki_j})^2 = (S^{BB}_{ki_j})^2,$$

$$(\sigma_{ki_j} / \lambda) < z_1 S^{BB}_{ki_j} < \sigma_{ki_j} < z_2 S^{BB}_{ki_j} < \lambda \sigma_{ki_j},$$

$$\ln y_u = \ln \frac{A}{\sigma} - \frac{(x_u - a)^2}{2\sigma^2},$$

где $\bar{\Phi}_{ki_j m_{ki_j}}$ - средние значения частных сумм, составленных из m элементов в $L_{ki_j}$ членов ($m = \overline{1, L_{ki_j} - 1}$), $a_{kji_j}$ и $\sigma_{kji_j}$ - неизвестные параметры нормального распределения, желательно, чтобы $M_{ki_j} \geq 106$; в этом случае мы получим, что при $\alpha = 0{,}01$ $a_{ki_j} \in (\bar{\Phi}^{БВ}_{ki_j} - 0{,}00258 S^{БВ}_{ki_j}, \bar{\Phi}^{БВ}_{ki_j} + 0{,}00258 S^{БВ}_{ki_j})$; очевидно, что при полном переборе всех сочетаний элементов исходной выборки число $M_{ki_j} \geq 106$ обеспечивается числом $L_{ki_j} \geq 20$ и, значит, в этом случае можно не пользоваться формулами $\sigma^2 = (2\Delta)^{-1}(x_u - x_{u+2})(x_{u+1} - x_u)(x_{u+2} - x_{u+1})$ и $a = (2\Delta)^{-1}[(x^2_{u+2} - x^2_{u+1}) \ln \frac{y_u}{y_{u+1}} - (x^2_{u+1} - x^2_u) \ln \frac{y_{u+1}}{y_{u+2}}]$, где величина yu принадлежит нормальному распределению (a, σ), $x_u < x_{u+1} < x_{u+2}$, а $\Delta = (x_{u+2} - x_{u+1}) \ln \frac{y_u}{y_{u+1}} - (x_{u+1} - x_u) \ln \frac{y_{u+1}}{y_{u+2}}$;

2) из п.1 следует также, что при $L_{ki_j} > 20$ можно использовать не все сочетания элементов исходной выборки, а только часть их, учитывающих в одном сочетании более двух, трех или еще более элементов, но так, чтобы получить $M_{ki_j} \geq 10^6$;

3) при $L_{ki_j} < 20$ предлагается для получения более вероятных значений $a_{ki_j}$ пользоваться методом «трех точек». При этом, как показывает анализ, на точность получаемых таким способом оценок $\tilde{a}_{ki_j}$ влияет качество гистограммы, определяемое рядом факторов: степенью симметричности построенной гистограммы, «хвостами» гистограммы и, разумеется, правильным выбором числа *s* разрядов гистограммы;

4) опыт показывает, что при выборе шага гистограммы по формуле $\delta(\Delta\Phi) = \Delta\Phi / \sqrt[3]{M_{ki_j}}$ оценки $\tilde{a}_{ki_j}$ по точности получаются не хуже, а чаще всего, и лучше у всех анализировавшихся тестовых примеров, чем при равномерном разбиении;

5) данные табл. 10 показывают, что при $L_{ki_j} \leq 10$ необходимо для организации выборки в $M_{ki_j}$ членов перебрать все сочетания элементов исходной выборки $\left(m = \overline{2, L_{ki_j} - 1}\right)$; при $10 < L_{ki_j} \leq 15$ можно брать $m = \overline{4, L_{ki_j} - 1}$, что обеспечивает значения $M_{ki_j}$ в интервале [1810, 32751]; при $15 < L_{ki_j} < 20$ можно менять значения *m* от 6 до ($L_{ki_j}$ - 1), что обеспечивает значения $M_{ki_j}$ в интервале [58640, 524267]; при $L_{ki_j} \geq 20$ можно брать $m = \overline{10, L_{ki_j} - 1}$, что даст $M_{ki_j} \geq 458913$;

Таблица 10

| $L_{ki_j}$ | 8 | 10 | 12 | 15 | 16 | 20 |
|---|---|---|---|---|---|---|
| $M_{ki_j}$ | 246 | 1012 | 4082 | 32751 | 65508 | 1048554 |

6) повышает точность получаемых оценок методом «трех точек» и процедура проверки соотношения числа элементов выборки, попавших в *s*-тый разряд гистограммы, к максимальному числу элементов выборки, оказавшихся в каком-то разряде гистограммы; если это отношение оказывается меньше наперед заданной положительной величины (например, 0,05), то проверяемый разряд следует объединить с последующим;

7) при построении гистограммы с шагом δ(ΔΦ), вычисленным по формуле $\delta(\Delta\Phi) = \Delta\Phi / \sqrt[3]{M_{ki_j}}$, полезным оказывается еще и построение двух гистограмм с шагами $\delta(\Delta\Phi)(1 \pm \varepsilon_s)$, где $0 < \varepsilon_s << 1$, с дальнейшим усреднение всех трех оценок (по трем гистограммам) $a_{ki_j}$;

8) и, наконец, следует сказать следующее. Предложенный алгоритм получения более вероятных средних значений статистических характеристик гипотетической выборки, конечно, сродни по духу методам, предлагаемым авторами в [14, 15]. Еще в 70-е годы XX столетия (да и навеян ими). Но если в [14, 15] решается задача получения достоверных оценок статистических характеристик при отказе от предположения о нормальности распределения исходных данных (случайных величин), то описываемый алгоритм с самого начала базируется на законе нормального распределения сумм случайных величин при неограниченном увеличении числа этих сумм; поэтому в первом случае все алгоритмы строятся на основе генерирования очень большого числа выборок рандомизированным способом из одной гипотетической совокупности и усреднения выборочных оценок; в данном же алгоритме ищутся, по существу, координаты вершины эмпирического нормального распределения, близкого к неизвестному нормальному распределению, и координаты такой вершины и дают оценку более вероятного значения искомой статистической характеристики.

Как следует из вышеприведенных выводов и рекомендаций, при $L_{ki_j} \le 20$ проблемы программной реализации алгоритма достаточно просты. Но как только $L_{ki_j} > 20$, возникает вопрос, как быть при росте $L_{ki_j}$, имея в виду способ формирования сумм (средних величин) различных сочетаний элементов из выборки длиной $L_{ki_j}$. Дело в том, что используемый в алгоритме способ формирования этих сумм требует оперирования с прямоугольными матрицами размерности $(L_{ki_j} \times C^m_{L_{ki_j}}) > 10^4$, где $L_{ki_j}$ - число строк этой матрицы, а $C^m_{L_{ki_j}}$ - число ее столбцов. Но так как $C^m_{L_{ki_j}} = C^{L_{ki_j}-m}_{L_{ki_j}}$, то для ситуаций $20 < L_{ki_j} \le 1450$ можно пользоваться уже упомянутыми прямоугольными матрицами, ориентируясь не на значения *m*, а на величины ($L_{ki_j}$ - *m*), ибо в этом случае размерность матриц не меняется, а в формировании выборки из $C_{ki_j}$ членов будут участвовать более «качественные» суммы. Здесь лишь нужно ограничиться каким-то значением $C_{ki_j}$. Убедительной представляется величина $C_{ki_j} \ge 10^6$

И вновь сталкиваемся с проблемой, когда $L_{ki_j} > 1450$ (тогда уже $C^2_{L_{ki_j}} \ge 10^6$, т.е. в этом случае уже при $m \ge 2$ можно получить выборку средних величин длиной $C^2_{L_{ki_j}} \ge 10^6$. Ответ представляется следующим: если возникает такая гипотетическая ситуация в математических экспериментах (например, при использовании ПЛП-поиска), когда в $i_j$ -том сечении *j*-го параметра по *k*-му критерию накапливается более 1450 значений $\Phi_{ki_j} l_{ki_j}$ ($l_{ki_j}$ =1, …, 1450, ), то можно легко организовать подвыборки из ($L_{ki_j}$ - 1) элементов исходной выборки и далее строить гистограмму из $C_{ki_j}$ = $L_{ki_j}$ средних величин, используя в

дальнейшем метод «трех точек». Но и в этом случае нужно где-то остановиться. В качестве пороговой величины в этом случае выбираем значение $C_{ki_j} = \text{L}_{ki_j} = 10^4$. Столь небольшие значения $C_{ki_j}$ (от 1450 до $10^4$) безусловно компенсируются «высоким качеством» средних величин, вычисляемых по подвыборкам, содержащих более тысячи элементов. Если же $\text{L}_{ki_j} > 10^4$, то в этом случае будем полагать среднее значение такой выборки приближенно равным искомой оценке. Разумеется, если решается задача, результаты которой должны обеспечить высокую (высочайшую) надежность решения, то порог для $C_{ki_j}$ должен быть поднят, и здесь остановка будет только за техническими возможностями ЭВМ (быстродействие и оперативная память). Все сказанное относится и к физическим экспериментам, когда варьируется один входной параметр системы, и при каждом варьировании производится $\text{L}_{ki_j} \geq 1450$ дублирующих экспериментов.

Полученные выше оценки математических ожиданий и дисперсий более вероятных в среднем значений *k*-го критерия в *J*-мерной точке в ПЛП-поиске [3,4], нужны для выявления (автоматического) многомерных точек $\bar{\alpha}$ (узлов), потенциальных кандидатов на точки глобального или одного из локальных экстремумов рассматриваемого критерия $\Phi_k(\bar{\alpha})$. И нужны, конечно, эти точки не сами по себе, а их координаты $\alpha_{nki_j}$ ($n = \overline{1, N_0}$ – номер узла; $k$ - номер вычисляемого критерия; $j$ - номер варьируемого параметра; $i_j = \overline{1, M_{kj}}$ ). Однако, «лобовой» обзор всех $N_0$ узлов с целью выявления таких потенциальных точек какой-то рекуррентной процедурой невозможен даже на супер-ЭВМ, поскольку с ростом значений $J$ и $M_j$ число $N_0$ стремительно растет (например, при $J = 5$ и для $\forall j$ $M_j$ получим $N_0 = 10^5 = 100000$.

Поэтому для вычисления оценок $a_{nkj}$ во всех $N_0$ узлах сформированной решетки предложен другой путь, нашедший свое воплощение в алгоритме *CMPANP* (calculation of more probability average in nodal point). Предварительно все значения $\bar{\Phi}_{i_j}^{\text{БВ}}$, полученные тем или иным способом для данного *k*–го критерия записываются в одномерном упорядоченном массиве $\Phi^{\text{БВ}}(\bullet)$ размерности $\left(1 \times \sum_{j=1}^{J} M_{kj}\right)$. При этом важен сам принцип упорядочения, который заключается в следующем: все величины $\bar{\Phi}_{i_j}^{\text{БВ}}$ располагаются так в одномерном массиве, что соответствующие им значения одного параметра $\left\{\alpha_{ki_j}\right\}$ образуют упорядоченную числовую подпоследовательность вида $\left\{\alpha_{k1_j} < \alpha_{k2_j} < \ldots < \alpha_{kM_j}\right\}$ , а эти подпоследовательности соответствуют такой числовой последовательности

$$\left\{\alpha_{k1_j} < \ldots < \alpha_{kM_j}, \alpha_{k1_{j+1}} < \ldots < \alpha_{kM_{j+1}}, \ldots, \alpha_{k1_J} < \ldots < \alpha_{kM_J}\right\},$$

причем, порядок расположения этих последовательностей в смысле значений $j$ не имеет значения. Сформированный таким образом массив величин $\{\bar{\Phi}_{i_j}^{\text{БВ}}\}$ позволяет реализовать перебор всех $N_0$ узловых точек с фиксацией значений координат этих точек.

Идея перебора заключается в том [16], что любое число из натурального ряда чисел $N > 1$ может быть представлено суммой фиксированного числа слагаемых из натурального ряда, причем, два представления будут отличаться друг от друга местами, занимаемыми в них одинаковыми слагаемыми. Для реализации нашей идеи фиксированное число слагаемых одного представления для каждого $N$ берем равным $J$, и тогда количество таких представлений $\varphi(N; J)$ натурального числа $N$ будет равно $\varphi(N; J) = C_{N-1}^{J-1}$. Например, если $N =$

7 и $J = 4$, то $\varphi(N) = C_6^3 = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} = 20,$ и можно записать, что 7 = (1+2+2+2; 2+1+2+2; 2+2+1+2; 2+2+2+1; 1+1+2+3; 1+2+1+3; 2+1+1+3; 1+3+2+1; 1+2+3+1; 3+2+1+1; 3+1+2+1; 3+1+1+2; 2+3+1+1; 1+1+1+4; 1+1+4+1; 1+4+1+1; 4+1+1+1; 1+1+3+2; 2+1+3+1; 1+3+1+2). Теперь рассматривая каждое слагаемое в суммах как номер сечения (уровня) *j*-го параметра (по порядку следования величин *j*), получим например, что первая сумма в скобках позволяет получить точку $\bar{\alpha}_{nk}$ с такими координатами:

$$\bar{\alpha}_{nk} = \left(\alpha_{k1_1}, \alpha_{k2_2}, \alpha_{k2_3}, \alpha_{k2_4}\right) \tag{13}$$

А формула (13) позволяет легко вычислить в данной точке значения $a_{nkj}$, пользуясь элементами массива $\bar{\Phi}^{\text{БВ}}_{i_j}$:

$$a_{nkJ} = \frac{1}{J}(\Phi^{\text{БВ}}_{\text{k1}_1} + \Phi^{\text{БВ}}_{\text{k2}_2} + \Phi^{\text{БВ}}_{\text{k2}_3} + \Phi^{\text{БВ}}_{\text{k2}_4}) \tag{14}$$

Очевидно, что в формулах (13) и (14) индекс *n* не может быть использован, поскольку его значения в ходе перебора узлов решетки не будут составлять какую-либо упорядоченную числовую последовательность.

Теперь рассмотрим принципы программной реализации предлагаемого алгоритма. Были установлены следующие факты:

1. Для $J = 2$ и любого $N \geq J$ способ генерирования сумм, равных *N* и состоящих из двух слагаемых, состоит в построении матрицы размерности (2×(*N* - 1)), имеющей вид

$$\begin{pmatrix} A_{\max} & 1 & 2 & 3 & \ldots & A_{\max}-1 \\ 1 & A_{\max} & A_{\max}-1 & A_{\max}-2 & \ldots & J \end{pmatrix},$$

где $A_{\max} = N - J + 1$ для любых значений *J* и $N \geq J$. При $N = J$ получаем единственную сумму из *J* единиц, т.е. получаем многомерную точку $\bar{\alpha}_k = (\alpha_{k1_1}, \alpha_{k1_2}, ..., \alpha_{k1_J})$. Для $J = 2$ имеем $A_{\max} = N - 1$.

2. При любом $J > 2$ и $N - J = 1$ (т.е. $A_{\max} = 2$) всевозможные разложения числа *N* на суммы из *J* слагаемых определяются $\varphi(N, J)$ столбцами матрицы размером  такого вида, где $\varphi(N; J) = C_{N-1}^{J-1} = C_J^{J-1} = J$:

$$\begin{pmatrix} A_{\max} & 1 & \ldots & 1 \\ 1 & A_{\max} & \ldots & 1 \\ \ldots & \ldots & \ldots & \ldots \\ 1 & \ldots & \ldots & A_{\max} \end{pmatrix} \tag{15}$$

3. При любом $J > 2$ и $N - J = 2$ (т.е. $A_{\max} = 3$) всю матрицу размерности ($J \times \varphi(N)$) можно представить состоящей из двух подматриц размерности ($J \times J$) и ($J \times \varphi_1(N, J)$), где ($J + \varphi_1(N, J) = \varphi(N)$. Первая подматрица имеет вид (15), а вторая подматрица состоит из столбцов, элементы которых суть 2 или 1, причем, число двоек $n_1 = N - J = A_{\max} - 1$, а число единиц $n_2 = 2J - N$ (естественно, $n_1 + n_2 = J$). Отсюда $\varphi_1(N, J) = \frac{J!}{n_1! n_2!}$ и с учетом значений $n_1$ и $n_2$ имеем окончательно $\varphi_1(N, J) = J(J - 1) / 2$.

Алгоритм построения подматрицы размерности ($J \times \varphi_1(N, J)$) такой: строится матрица размерности ($J \times \varphi_1(N, J)$), столбцы которой содержат ($N - J$) единиц и ($2J - N$) нулей. Эти столбцы образуются (выбираются) из последовательности двоичных записей чисел натурального ряда, которые лежат в интервале $2^0 \leq N < 2^J$. Причем, имеются в виду фиксированные двоичные записи, т.е. такие, у которых нули будут стоять и перед единицами: важно лишь, чтобы длина записи (число двоичных разрядов) строго равнялась *J*. Теперь ко всем элементам столбцов построенной таким образом матрицы прибавим по 1 и

получим искомую подматрицу ($J \times \varphi_1(N)$). Рассмотрим такой пример. Пусть $J = 5$ и $N = 7$. Значит, $N - J = 2$ и $A_{max} = N - J + 1 = 3$. Подматрица (15) имеет следующий вид

$$\begin{pmatrix} 3 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 \\ 1 & 1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 1 & 3 \end{pmatrix}.$$

Заметим, что $\varphi_1(N, J) = \frac{5!}{2!3!} = 10$, а $\varphi(N, J) = 5 + 10 = 15$. Теперь строим подматрицу подматрицу ($J \times \varphi_1(N)$). Матрицу фиксированных двоичных записей из нулей и единиц из интервала $2^0 \le N < 2^5$ запишем в виде таблицы 11.

**Таблица 11**

| **φ1(N)** | **N** | **Подходит (♦) или не подходит (•) двоичная запись** | **Разряды записи** | | | | |
|---|---|---|---|---|---|---|---|
| | | | 5 | 4 | 3 | 2 | 1 |
| | **1** | • | 0 | 0 | 0 | 0 | 1 |
| | **2** | • | 0 | 0 | 0 | 1 | 0 |
| **1** | **3** | ♦ | 0 | 0 | 0 | 1 | 1 |
| | **4** | • | 0 | 0 | 1 | 0 | 0 |
| **2** | **5** | ♦ | 0 | 0 | 1 | 0 | 1 |
| **3** | **6** | ♦ | 0 | 0 | 1 | 1 | 0 |
| | **7** | • | 0 | 0 | 1 | 1 | 1 |
| | **8** | • | 0 | 1 | 0 | 0 | 0 |
| **4** | **9** | ♦ | 0 | 1 | 0 | 0 | 1 |
| **5** | **10** | ♦ | 0 | 1 | 0 | 1 | 0 |
| | **11** | • | 0 | 1 | 0 | 1 | 1 |
| **6** | **12** | ♦ | 0 | 1 | 1 | 0 | 0 |
| | **13** | • | 0 | 1 | 1 | 0 | 1 |
| | **14** | • | 0 | 1 | 1 | 1 | 0 |
| | **15** | • | 0 | 1 | 1 | 1 | 1 |
| | **16** | • | 1 | 0 | 0 | 0 | 0 |
| **7** | **17** | ♦ | 1 | 0 | 0 | 0 | 1 |
| **8** | **18** | ♦ | 1 | 0 | 0 | 1 | 0 |
| | **19** | • | 1 | 0 | 0 | 1 | 1 |
| **9** | **20** | ♦ | 1 | 0 | 1 | 0 | 0 |
| | **21** | • | 1 | 0 | 1 | 0 | 1 |
| | **22** | • | 1 | 0 | 1 | 1 | 0 |
| | **23** | • | 1 | 0 | 1 | 1 | 1 |
| **10** | **24** | ♦ | 1 | 1 | 0 | 0 | 0 |

Теперь, прибавив ко всем элементам строк таблицы, помеченных знаком ♦, по единице, получим искомую матрицу для $J = 5$ и $N = 7$

$$\begin{pmatrix} 3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 1 & 2 \\ 1 & 1 & 3 & 1 & 1 & 1 & 2 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 3 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 & 3 & 2 & 2 & 1 & 2 & 1 & 1 & 2 & 1 & 1 & 1 \end{pmatrix} \quad (16)$$

у которой $\varphi(N, J) = 15$ и сумма элементов каждого столбца равна 7.

4. И, наконец, при $J > 2$ и $N - J > 2$ (общий случай), когда пп. 2 и 3 уже реализованы, процедура построения матрицы разложений для числа $N + 1$ приобретает следующий рекуррентный характер:

а) ко всем элементам первой строки матрицы (16) добавляется единица, и вновь созданная матрица образует первую порцию столбцов в матрице разложений для $N + 1$;

б) затем во всех столбцах матрицы (16), у которых элементы первой строки равны 1, к элементам второй строки добавляется по единице, и созданная совокупность столбцов образует новую порцию в матрице разложений для $N + 1$;

в) далее во всех столбцах матрицы (16), у которых элементы первых двух строк подряд равны 1, прибавляем к элементам третьей строки по 1, и это дает нам новую порцию столбцов в матрице для $N + 1$;

г) такая процедура продолжается до тех пор, пока полностью не заполним матрицу разложений для $N + 1$, у которой число столбцов равно $\varphi(N+1; J) = C_N^{J-1}$.

Пример реализации описанной процедуры для $J = 4$ в работе не приводится ввиду громоздкости, однако из него очевидна неэффективность прямого (буквального) использования описанного алгоритма. Действительно, например, если $J = 6$ и все $M_j = \text{const} = 8$, то число узловых точек построенной решетки $NNPL = 8^6 = 262144$ (*NNPL* - number of nodal points of the lattice - число узлов решетки). При этом, чтобы не пропустить ни одной точки, мы должны в описываемом алгоритме менять $N$ от $J$ до $\sum_{j=1}^{6} M_j = 48$, и если бы мы так и поступили, то всего просмотрели бы $(C_5^5 + C_6^5 + ... + C_{47}^5) = 12460547$ узлов, из которых нам нужны всего 262144 узла. Поэтому на каждом шаге работы программы, начиная с $N - J > 2$, мы в соответствии с нашим алгоритмом будем при переходе от матрицы для $N$ к матрице для $N + 1$ рассматривать только те столбцы в матрице для $N + 1$, у которых все элементы меньше соответствующих значений $M_j$. Для упрощения в программе мы будем сразу фиксировать столбцы для $N = J$. Для программной реализации *CMPANP* необходимо знать, каких максимальных размеров нужно заказывать матрицу разложений числа $N$ для данного $J$. Так как $N \in \left[ J, \sum_{j=1}^{J} M_j \right]$, то $N_{\max}$, для которого $\varphi(N_{\max}, J) = \max \varphi(N, J)$, принадлежит также этому интервалу. Для четных значений $J$

$$N_{\max} = 0,5\left( J + \sum_{j=1}^{J} M_j \right),$$

а для нечетных значений $J$

$$N_{\max} = 0,5\left( J + \sum_{j=1}^{J} M_j - 1 \right).$$

Программная реализация предусматривает и случай $J = 2$, хотя для ПЛП-поиска [3,4,8] рекомендуются лишь значения $J \geq 4$.

## Литература

1. *Ван дер Варден Б.Л.* Математическая статистика. - М.: Изд-во иностр. лит-ры, 1965. - 435 с.
2. *Михог Г., Урсяну В.* Выборочный метод и статистическое оценивание. - М.: Финансы и статистика, 1985. - 245 с.
3. *Статников И.Н., Андреенков Е.В.* ПЛП-поиск – эвристический метод решения задач математического программирования. – М.: ИИЦ МГУДТ, 2006г. – 140 с.
4. *Статников И.Н., Фирсов Г.И.* О некоторых возможностях ПЛП-поиска в решении задач моделирования и исследования динамических систем машин // Южно-Сибирский научный вестник. – 2012. - № 1. - С.92-96.
5. *Абезгауз Г.Г., Тронь А.П., Копенкин Ю.Н., Коровина И.А.* Справочник по вероятностным расчетам. - М.: Воениздат, 1966. - 408 с.
6. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. - М.: Наука, 1965. - 464 с.
7. *Оуэн Д.Б.* Сборник математических таблиц. - М.: ВЦ АН СССР, 1973. - 586 с.
8. *Статников И.Н., Фирсов Г.И.* ПЛП-поиск и его реализация в среде MATLAB // Проектирование инженерных и научных приложений в среде MATLAB. - М.: ИПУ РАН, 2004. - С.398-411.
9. *Ченцов Н.Н.* Статистические решающие правила и оптимальные выводы. - М.: Наука, ГРФМЛ, 1972. - 520 с.
10. *Дружинин Н.К.* Выборочное наблюдение и эксперимент. – М.: Статистика, 1977. – 176 с.
11. *Tippelt L.H.G.* Statistics. – Oxford: Oxford Universiti Press, 1958. – P.94-97.
12. *Прохоров Ю.В., Розанов Ю.А.* Теория вероятностей. Основные понятия. Предельные теоремы. – М.: Наука, ГРФМЛ, 1967. – 496 с.
13. *Феллер В.* Введение в теорию вероятностей и ее приложения. Т. 1 и 2. – М.: Мир, 1984.
14. *Диаконис П., Эфрон Б.* Статистические методы с интенсивным использованием ЭВМ В мире науки. – 1983. – № 7. – С.60-73.
15. *Эфрон Б.* Нетрадиционные методы многомерного статистического анализа. – М.: Финансы и статистика, 1988. – 263 с.
16. *Виленкин Н.Я.* Комбинаторика. – М.: Наука, ГРФМЛ, 1969. – 328 с.