КОМБИНАТОРНЫЕ МЕТОДЫ ПОСТРОЕНИЯ ГИСТОГРАММ ФУНКЦИЙ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ СЛУЧАЙНЫХ ПОГРЕШНОСТЕЙ

В.А. Гайский, П.В. Гайский

Морской гидрофизический институт
НАН Украины
г. Севастополь, ул. Капитанская, 2
E-mail: gaysky@inbox.ru

Гипотеза о независимости сочетаний нескольких отсчетов случайной величины из множества существенно увеличивает число событий и улучшает оценки моментов.

Функция плотности распределения вероятностей (ФПРВ) случайной величины $f(x)$ является достаточной для вычисления всех ее количественных оценок [1].

Например, математического ожидания

$$\overline{X} = \int_{-\infty}^{\infty} x f(x) dx,$$

и дисперсии

$$D = \int_{-\infty}^{\infty} \left(x - \overline{X}\right)^2 f(x) dx.$$

Если получить реальные ФПРВ, то их не обязательно аппроксимировать каким-либо законом для вычисления оценок.

Известные методы построения гистограмм ФПРВ требуют наличия достаточно большого количества измерений. Например [2], согласно правила Сэвиджа для *(n–1)* столбцов гистограммы требуется n^2 измерений.

Даже в процессе градуировки трудоемкость и длительность выполнения большого числа измерений могут оказаться неприемлемыми. С другой стороны, принятие в качестве закона распределения нормального закона, как максимально энтропийного для симметричных распределений, в предложении, что хуже не будет, также трудно оправдать. Поскольку, во-первых, реальное распределение может быть несимметричным и может быть хуже и, во-вторых, зачем терять в завышении погрешностей, за снижение которых всегда приходится дорого платить.

Таким образом, желательно уметь строить гистограммы реальных ФПРВ случайных погрешностей при малых *n*.

Это возможно, если считать, что все пары, тройки и т.д. случайных погрешностей являются независимыми событиями и ограничивают множества возможных значений, вероятности попадания случайной величины в которые равны.

Рассмотрим возможные методы построения гистограмм ФПРВ с использованием гипотезы о независимости пар отсчетов.

<u>Метод равных площадей соседних пар отсчетов.</u>

На оси x размещаем вариационный ряд отсчетов случайной величины x $i=\overline{1,n-1}$ (рис. 1) , принимающей значения в соответствии с таблицей 1.

Для каждого отрезка x_1, x_{i+1} как основания строим прямоугольник площадью S (единичной) и высотой $h_{i,i+1}$, причем

$$h_{i,i+1} = \frac{S}{|x_i - x_{i+1}|}, \quad i = \overline{1,n-1}. \qquad (1)$$

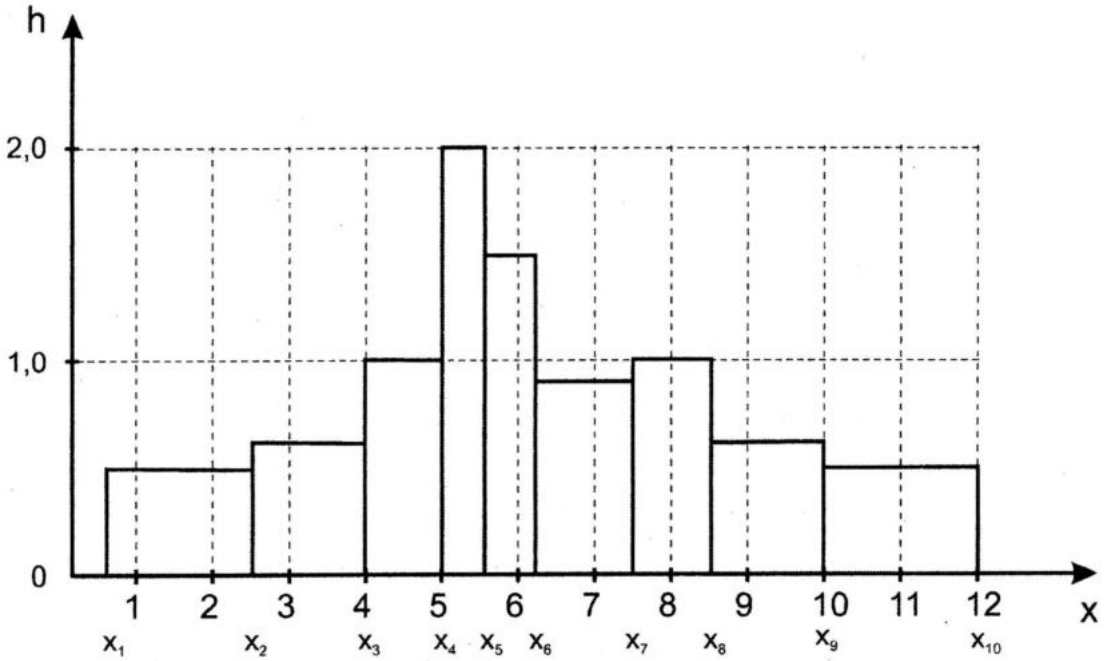
Принимаем $h_{i,i+1}$ за вероятность отсчета на отрезке $x_i\, x_{i+1}$.

Приводим вероятности к полной группе событий и находим значение $f(x_i\, x_{i+1})$ ФПРВ для *i*-го столбца гистограммы по формуле

$$f(x_i x_{i+1}) = \frac{h_{i,i+1}}{\sum_{i=1}^{n-1} h_{i,i+1}} = \frac{1}{|x_i - x_{i+1}| \sum_{i=1}^{n-1} |x_i - x_{i+1}|^{-1}}, \quad i = \overline{1,n-1}. \qquad (2)$$

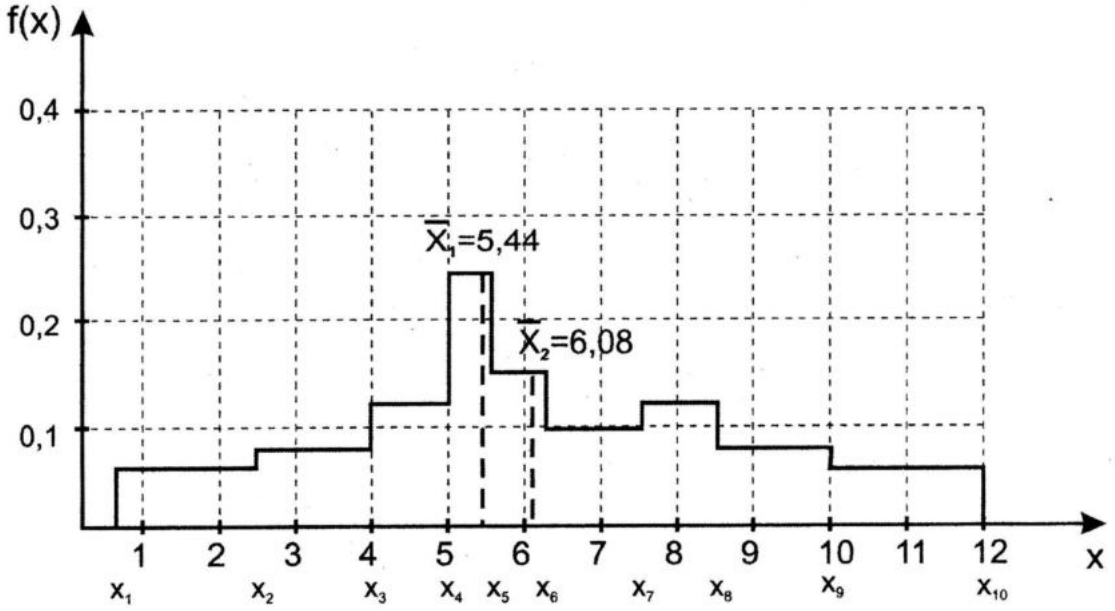
Таблица 1

N	1	2	3	4	5	6	7	8	9	10
x	0,5	2,5	4	5	5,5	6,3	7,5	8,5	10	12



Р и с. 1. Схема построения прямоугольников равных площадей для соседних пар отсчетов

Получим (n–1) столбцов гистограммы с неравными основаниями для n измерений. На рис. 2 приведена гистограмма распределения, построенная этим методом и даны оценки средних этим методом $\overline{X_2}=6{,}08$ и стандартным $\overline{X_1}=5{,}44$.



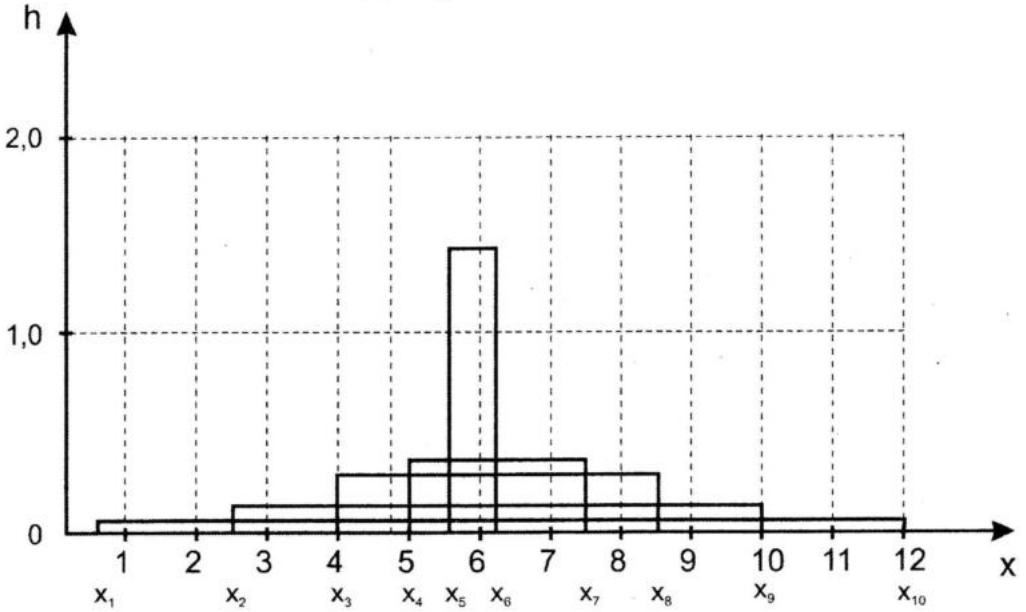
Р и с. 2. Гистограмма распределения, построенная методом равных площадей соседних пар отсчетов

Метод равных площадей пар размаха.

Предполагаем, что искомое распределение симметрично.

Строим прямоугольники равных площадей на оси x с вариационным рядом x_i для пар размаха $x_1 x_n, x_2 x_{n-1}, \ldots, x_i x_{n-1}$ (рис. 3) с высотами

$$h_{s,k} = \frac{1}{|x_s - x_k|}, \quad s = \overline{1, n/2}, \quad k = n - s. \tag{3}$$



Р и с. 3. Схема построения прямоугольников равных площадей для пар размаха

Вероятность попадания случайной величины на отрезок $x_i x_{i+1}$ в прямоугольнике с основанием $x_s x_k$ будет равна

$$p\left(\frac{i, i+1}{s, k}\right) = (x_i - x_{i+1}) h_{s,k} = \frac{x_i - x_{i+1}}{x_s - x_k}. \tag{4}$$

При этом

для $i \le \frac{n}{2}$, $1 \le s \le i$, таких прямоугольников $N_i = i$;

для $i > \frac{n}{2}$, $\frac{n}{2} \le k \le n$, таких прямоугольников $N_i = n - i$.

Всего $\max N_i = \frac{n}{2}$.

Попадание случайной величины на отрезок $x_i x_{i+1}$ есть сумма совместных событий в N_i множествах, которые определяются для всех $s \le i$.

Вероятность такого события вычисляют как сумма вероятностей совместных событий [3] по N_i множествам по формуле

$$p(i, i+1) = \sum_{s,k}^{N_i} p\left(\frac{i, i+1}{s, k}\right) - \sum_{s,k,l,v}^{C_{N_i}^2} p\left(\frac{i, i+1}{s, k}\right) p\left(\frac{i, i+1}{l, v}\right) +$$
$$+ \sum_{s,k,l,v,m,\xi}^{C_{N_i}^3} p\left(\frac{i, i+1}{s, k}\right) p\left(\frac{i, i+1}{l, v}\right) p\left(\frac{i, i+1}{m, \xi}\right) + (-1)^{N_i - 1} \prod_{s,k}^{N_i} p\left(\frac{i, i+1}{s, k}\right). \tag{5}$$

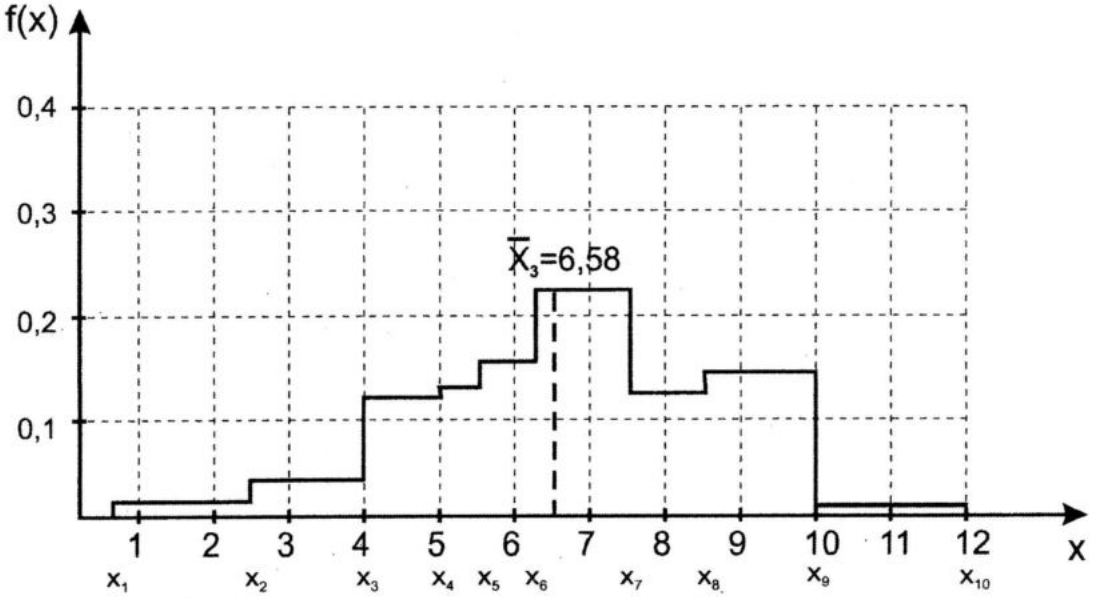
Ординату гистограммы функции распределения плотности вероятности попадания случайной величины на отрезок $x_i x_{i+1}$ получим после приведения вероятностей к полной группе событий

$$f(i,i+1)=\frac{p(i,i+1)}{\sum_{i=1}^{n-1} p(i,i+1)}. \qquad (6)$$

Учитывая выражение (4) выражение (5) можно записать в виде

$$p(i,i+1)=|x_i-x_{i+1}|\sum_{s,k}^{N_i}\frac{1}{x_s-x_k}-|x_i-x_{i+1}|^2\sum_{s,k,e,y}^{C_{N_i}^2}\frac{1}{|x_s-x_k||x_e-x_y|}+$$
$$+|x_i-x_{i+1}|^3\sum_{s,k,e,y,m,\xi}^{C_{N_i}^3}\frac{1}{|x_s-x_k||x_e-x_y||x_m-x_\xi|}+(-1)^{N_i-1}|x_i-x_{i+1}|^{N_i}\prod_{s,k}^{N_i}\frac{1}{|x_s-x_k|} \qquad (7)$$

На рис. 4 представлена гистограмма распределения, построенная этим методом и дана оценка среднего $\overline{X_3}=6{,}58$



Р и с. 4. Гистограмма распределения, построенная методом равных площадей пар размаха

<u>Метод равных площадей всех пар отсчетов (комбинаторный).</u>

Имеется вариационный ряд отсчетов x_i, $i=\overline{1,n}$. Берем все пары отсчетов и строим прямоугольники равных площадей так, как это показано для примера на рис. 5 для n=10.

Будет $N=C_n^2=\frac{n(n-1)}{2}$ разных прямоугольников. Отрезок вариационного ряда $x_i x_{i+1}$ будет входить во все прямоугольники с основаниями $x_s x_k$ при $1\le s\le i$, $i+1\le k\le n$.

Индексы s и k можно определить по таблице 2, где каждому $(i,i+1)$ отрезку соответствует столбец $k=i+1$, плюсы в котором определяют строки для значений индексов $s\le i$ и $k\ge i+1$ для примера $n=10$.

Таких прямоугольников для $(i,i+1)$-го отрезка будет $N_i=i(i-1)$. Для наглядности и примера числа N_i для $n=\overline{2,10}$ и $i=\overline{1,9}$ представлены в таблице 3.

Таблица 2

Схема определения индексов s и k пересекающихся множеств событий для отрезка $x_i x_{i+1}$ при $k = i+1$

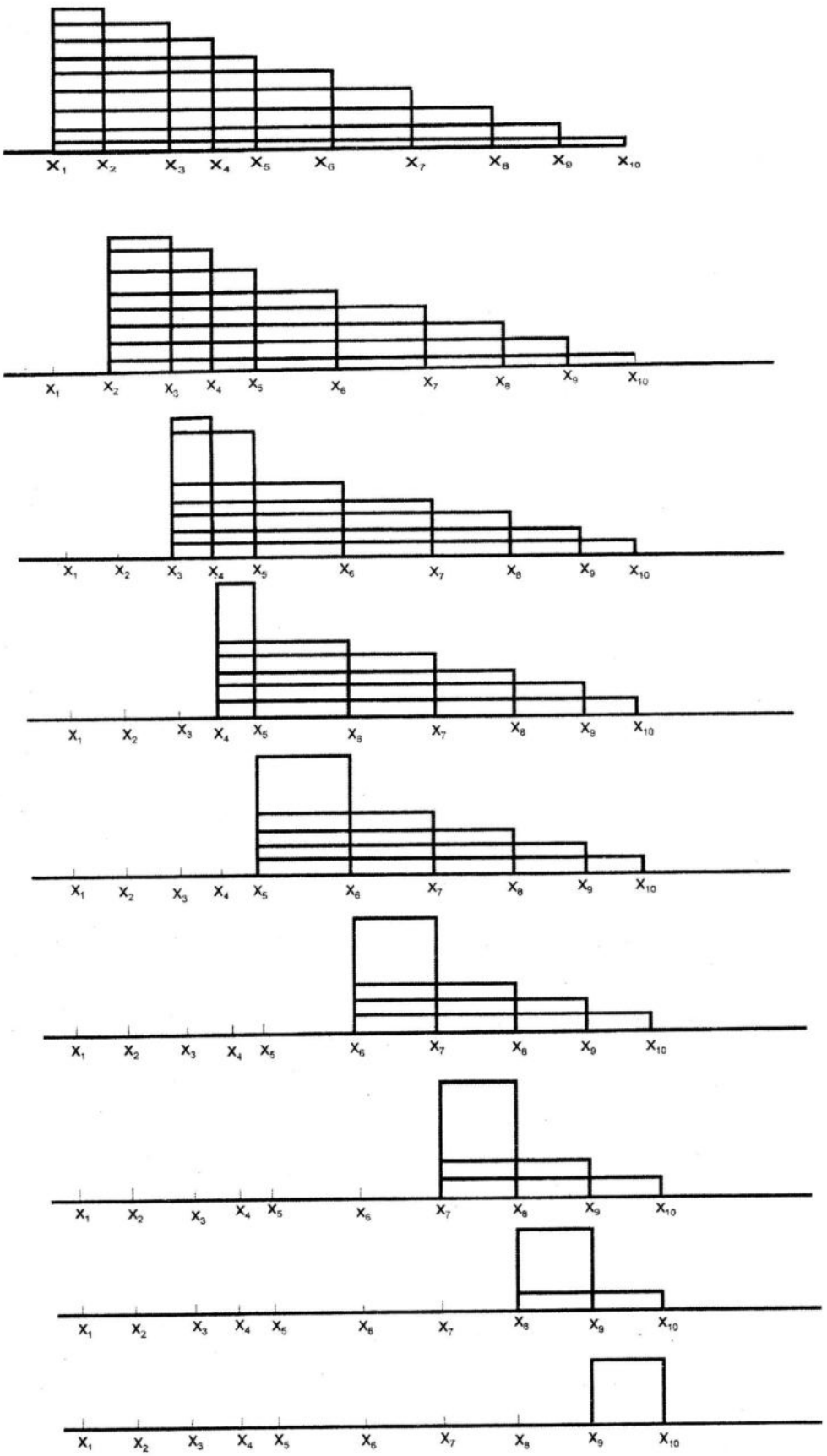
s	k								
	2	3	4	5	6	7	8	9	10
1	+	+	+	+	+	+	+	+	+
2		+	+	+	+	+	+	+	+
3			+	+	+	+	+	+	+
4				+	+	+	+	+	+
5					+	+	+	+	+
6						+	+	+	+
7							+	+	+
8								+	+
9									+

Таблица 3

Значение N_i количества множеств совместимых событий, в которые входит отрезок $x_i x_{i+1}$ для $n = \overline{2,10}$, $i = \overline{1,9}$

i	n								
	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9
2		2	4	6	8	10	12	14	16
3			3	6	9	12	15	18	21
4				4	8	12	16	20	24
5					5	10	15	20	25
6						6	12	18	24
7							7	14	21
8								8	16
9									9

Для вычисления вероятностей попадания случайной величины на отрезок $x_i x_{i+1}$ в конкретном прямоугольнике с основанием $x_s x_k$ $p\left(\frac{i,i+1}{s,k}\right)$ справедлива формула (4), аналогично для вероятности суммы совместных событий $p(i,i+1)$ справедливы формулы (5) и (7), а для ординат гистограммы функции РПВ $f(i,i+1)$ формула (6).

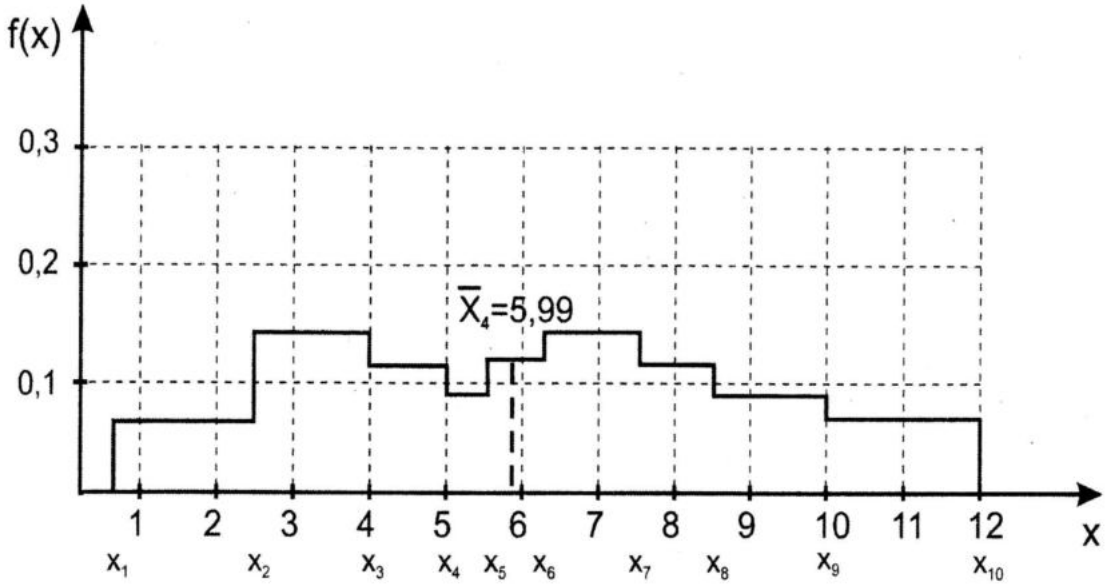


Р и с. 5. Схема построения прямоугольников равных площадей для всех пар отсчетов

Отличие двух методов состоит в правиле формирования пар отсчетов и количестве N_i множеств событий, в которые входит отрезок $x_i x_{i+1}$ За счет использования пар отсчетов мы имеем $C_n^2 = \frac{n(n-1)}{2} \approx \frac{n^2}{2}$ событий, вместо n (по числу отсчетов) в традиционном анализе.

Следовательно выигрыш в уменьшении дисперсии средней величины составит в $n/2$ раз, а среднеквадратичная погрешность уменьшится не в $\sqrt{n}$, а в $n/\sqrt{2}$ раз.

На рис. 6 представлена гистограмма распределения, построенная этим методом с точностью первой итерации по формуле (5) и дана оценка среднего $\overline{X_4} = 5{,}99$.



Р и с. 6. Гистограмма распределения, построенная методом равных площадей всех пар отсчетов

Возможное расширение комбинаторного метода на независимые тройки, четверки, пятерки и ρ отсчетов дает увеличение числа событий N в соответствии с формулой

$$N = c_n^{\rho} = \frac{n(n-1)\ldots(n-\rho)}{\rho!}.$$

Это будут комбинаторные методы равных ρ-гиперобъемов сочетаний ρ-отсчетов. Разработка этих методов представляется перспективной.

СПИСОК ЛИТЕРАТУРЫ

1. *Новицкий П.В., Зограф И.А.* Оценка погрешностей результатов измерений. 2-е изд., перераб. и доп. – Л.: Энергоатомиздат. Ленинградское отделение. – 1991. – 304 с.
2. *Абезгауз Г.Г., Тронь А.П., Копенкин Ю.Н., Коровина И.А.* Справочник по вероятностным расчетам. – М.: Военное изд-во МО СССР. – 1996. – 407 с.