УДК 519.71 (075.8)

Ю.А. ДОЛГОВ, А.Ю. ДОЛГОВ, Ю.А. СТОЛЯРЕНКО

Приднестровский государственный университет им. Т.Г. Шевченко, Тирасполь, Приднестровье, Молдова

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ МЕТОДОМ МНОГОМЕРНЫХ ТОЧЕЧНЫХ РАСПРЕДЕЛЕНИЙ

Разработан принципиально новый метод получения адекватных многомерных математических моделей на базе исходных выборок малого объёма (n=3÷15 строк). Алгоритм расчёта основан на обработке каждого фактора исходной выборки методом точечных распределений (МТР), который позволяет путём использования знаний о характере закона распределения искусственно увеличивать объём выборки. Состыковка индивидуальных факторных выборок в единую многомерную выборку большого объема (в диапазоне $n^2 \div 3n^2$) происходит по строкам с максимальным уровнем ненормированной плотности вероятности и с одновременным отсчетом всех неполных строк. Далее расчет проводится одним из методов получения адекватных математических моделей по пассивным данным.

Ключевые слова: *выборка малого объема, метод точечных распределений, математическое моделирование по пассивным данным.*

Введение

В современной промышленности существуют такие производства, которые в силу технологических ограничений не могут дать контрольные выборки достаточно большого объема, чтобы в соответствии с законами теории планирования эксперимента получить адекватную математическую модель, пригодную для управления сложным объектом контроля.

Так, например, при производстве кристаллов интегральных микросхем в силу специфики топологии пластин на ней имеются 5, редко 10 тестовых ячеек, измерения в которых должны с некоторой вероятностью отражать поведение одноименных параметров 400-5000 рабочих кристаллов. Совокупность таких измерений по нескольким десяткам параметров (факторов) на протяжении всего технологического процесса не позволяют использовать классические методы расчета выборочных параметров, которые в этом случае дают слишком большие (неэффективные) интервальные оценки, при этом на фоне ошибок их вычисления (шум эксперимента) может пропасть влияние значимого фактора. Аналогичные задачи существуют в медицине, экономике, сельском хозяйстве и других областях человеческой деятельности.

В статье предлагается метод многомерных точечных распределений (ММТР), позволяющий получать адекватные математические модели сложного объекта на основе многомерных выборок малого объема.

1. Теоретическая и экспериментальная часть

На сегодняшний день точно сформулированных методов, предназначенных исключительно для обработки выборок малого объема, не существует. Введенные Стьюдентом (У.С. Госсетом) поправки на малый объем выборок уже не удовлетворяют современное производство. Так, например, нижняя граница малой выборки объемом n=3 элемента дает ошибку в определении среднеквадратического отклонения 46,6%, что не может быть признано удовлетворительным. Наиболее серьезной теоретической работой, сформулировавшей и обосновавшей принципы специфических приемов статистической обработки выборок малого объема, явилась работа [1], развитие которой в работе [2] привело к созданию метода точечных распределений (МТР), а в работе [3] – к определению верхней границы диапазона выборок малого объема n=15 и к методу расчета парной коррекции. Алгоритмы обоих методов и примеры их применения изложены также в работе [4].

Ниже предлагается метод многомерных точечных распределений (ММТР), позволяющий получать адекватную математическую модель сложного объекта на основе многомерной выборки малого объема. Изложение метода будем вести в виде алгоритма с одновременной иллюстрацией теоретических положений конкретным производственным примером.

1. Пусть в результате производства n=5 единиц (партий, пластин) изделий получены следующие

числовые значения контрольных параметров (X_i – параметры, контролируемые в ходе технологического процесса; Y – выходной показатель качества изделия. Все наименования размерностей для простоты опущены).

2. С помощью метода точечных распределений [2 - 4] для всех X_i и Y построить таблицы расчета ненормированных плотностей вероятности в виртуальной области. В качестве примера в табл. 2 представлен один такой расчет для фактора X_1.

Таблица 1

Экспериментальные данные контрольных операций

Номер изделия, f	Факторы, X_i				Выходная величина, Y_f
	X_{1f}	X_{2f}	X_{3f}	X_{4f}	
1	0,70	1,91	13,9	1,28	57,2
2	0,61	1,94	16,1	1,15	61,6
3	0,60	2,17	15,6	142	63,5
4	0,52	2,10	16,7	1,01	66,5
5	0,58	2,16	13,0	1,38	82,1

Таблица 2

Расчет ненормированных плотностей вероятностей для каждого измерения фактора X_1

j	X_{1j}	X_{1f}				
		0,52	0,58	0,60	0,61	0,70
1	0,451	0,26	0,01			
2	0,461	0,37	0,02			
3	0,472	0,52	0,04	0,01		
4	0,482	0,66	0,06	0,02	0,01	
5	0,493	0,81	0,11	0,04	0,02	
6	0,503	0,92	0,18	0,07	0,04	
7	0,513	099	0,28	0,11	0,07	
8	0524	1,00	0,41	0,19	0,12	
9	0,534	0,95	055	0,29	0,19	
10	0545	0,84	0,70	0,42	0,30	
11	0,555	0,70	0,84	0,56	0,42	
12	0,565	0,56	0,94	0,70	0,56	0,01
13	0,576	0,41	1,00	0,85	0,72	0,02
14	0,586	0,29	0,99	0,95	0,85	0,03
15	0,597	0,18	0,92	1,00	0,95	0,05
16	0,607	0,11	0,81	0,99	1,00	0,08
17	0,617	0,07	0,68	0,92	0,99	0,14
18	0,628	0,04	0,52	0,80	0,91	0,23
19	0,638	0,02	0,38	0,66	0,80	0,33
20	0,649	0,01	0,26	0,50	0,65	0,47
21	0,659		0,17	0,37	0,50	0,62
22	0,669		0,10	0,26	0,37	0,76
23	0,680		0,06	0,16	0,25	0,89
24	0,690		0,04	0,10	0,16	0,97
25	0,701		0,02	0,05	0,09	1,00
26	0,711		0,01	0,03	0,05	0,97
27	0,721			0,02	0,03	0,89
28	0,732			001	0,02	0,75
29	0,742				0,01	0,60
30	0,753					0,45

3. Для каждой строки f табл. 1 исходных экспериментальных данных построить таблицы данных МТР, в которые вносить одновременно величины двух столбцов X_{ij} из соответствующей таблицы ненормированных плотностей вероятностей (подобных табл. 2) и столбца X_{if}. Выравнивание (состыковка) столбцов X_{ij} и X_{if} (а также Y_j и Y_f) должно происходить по уровню максимальной ненормированной плотности вероятности. В нашем случае таких таблиц должно быть 5. Пример таблицы данных МТР вместе с их ненормированными плотностями вероятности для случая строки f=1 табл. 1 приведен в табл. 3.

Таблица 3

Пример состыковки таблиц виртуальных данных для случая строки f=1 табл. 1

X_1	P_1	X_2	P_2	X_3	P_3	X_4	P_4	Y	P_Y
0,565	0,01					0,936	0,01		
0,576	0,02					0,963	0,02		
0,586	0,03					0,991	0,04		
0,597	0,05			11,45	0,05	1,018	0,06		
0,607	0,08			11,70	0,09	1,045	0,10	44,10	0,10
0,617	0,14			11,95	0,15	1,073	0,17	45,62	0,16
0,628	0,23	1,77	0,21	12,20	0,23	1,100	0,26	47,14	0,26
0,638	0,33	1,79	0,31	12,45	0,35	1,127	0,38	48,66	0,37
0,649	0,47	1,81	0,45	12,70	0,48	1,155	0,52	50,18	0,51
0,659	0,62	1,83	0,60	12,95	0,63	1,182	0,67	51,70	0,67
0,669	0,76	1,85	0,75	13,20	0,78	1,209	0,81	53,22	0,81
0,680	0,89	1,87	0,88	13,45	0,90	1,237	0,93	54,74	0,92
0,690	0,97	1,89	0,97	13,70	0,91	1,264	0,98	56,26	0,98
0,701	1,00	1,91	1,00	13,95	1,00	1,291	0,99	57,78	0,99
0,711	0,97	1,93	0,97	14,20	0,96	1,319	0,94	59,30	0,94
0,721	0,89	1,95	0,88	14,45	0,86	1,346	0,83	60,82	0,84
0,732	0,75	1,97	0,75	14,70	0,72	1,373	0,70	62,34	0,70
0,742	0,60	1,99	0,60	14,95	0,57	1,401	0,54	63,86	0,55
0,753	0,45	2,01	0,45	15,20	0,43	1,428	0,40	65,38	0,40
		2,03	0,31	15,45	0,30	1,455	0,28	66,90	0,28
		2,05	0,21	15,70	0,19	1,483	0,18	68,42	0,18
		2,07	0,13	15,95	0,12	1,510	0,11	69,94	0,11
		2,09	0,07	16,20	0,07	1,537	0,06	71,46	0,06
		2,11	0,04	16,45	0,04	1,565	0,03	7298	0,03
		2,13	0,02	16,70	0,02	1,592	0,01	74,50	0,02
		2,15	0,01	16,95	0,01			76,02	0,01

Таблица 4

Полная выборка ММТР

K	X_{1k}	X_{2k}	X_{3k}	X_{4k}	Y_k	K	X_{1k}	X_{2k}	X_{3k}	X_{4k}	Y_k
1	0,628	1,77	12,20	1,100	47,14	40	0,534	2,05	14,20	1,264	53,22
2	0,638	1,79	12,45	1,127	48,66	41	0,545	2,07	14,45	1,291	54,74
3	0,649	1,81	12,70	1,155	50,18	42	0,555	2,09	14,70	1,319	56,26
4	0,659	1,83	12,95	1,182	51,70	43	0,565	2,11	1495	1,346	57,78
5	0,669	1,85	13,20	1,209	53,22	44	0,576	2,13	15,20	1,373	59,30
6	0,680	1,87	13,45	1,237	54,74	45	0,586	2,15	15,45	1,401	60,82
7	0,690	1,89	13,70	1,264	56,26	46	0,597	2,17	15,70	1,428	62,34
8	0,701	1,91	13,95	1,291	57,78	47	0,607	2,19	15,95	1,455	6386
9	0,711	1,93	14,20	1,319	59,30	48	0,617	2,21	16,20	1,483	65,38
10	0,721	1,95	14,45	1,346	60,82	49	0,628	2,23	16,45	1,510	66,90
11	0,732	1,97	14,70	1,373	62,34	50	0,638	2,25	16,70	1,537	68,42
12	0,742	1,99	14,95	1,401	63,86	51	0,649	2,27	16,95	1,565	69,94
13	0,753	2,01	15,20	1,428	65,38	52	0,659	2,29	17,20	1,592	71,46
14	0,513	1,77	13,95	0,909	48,66	53	0,669	2,31	17,45	1,619	72,98
15	0,524	1,79	14,20	0,936	50,18	54	0,680	2,33	17,70	1,647	74,50
16	0,534	1,81	14,45	0,963	51,70	55	0,461	1,99	15,20	0,854	57,78
17	0,545	1,83	14,70	0,991	53,22	56	0,472	2,01	15,45	0,881	59,30
18	0,555	1,85	14,95	1,018	54,74	57	0,482	2,03	15,70	0,909	60,82
19	0,565	1,87	15,20	1,045	56,26	58	0,493	2,05	15,95	0,936	62,34
20	0,576	1,89	15,45	1,079	57,78	59	0,503	2,07	16,20	0,963	63,86
21	0,586	1,91	15,70	1,100	59,30	60	0,513	2,09	16,45	0,991	65,38
22	0,597	1,93	15,95	1,127	60,82	61	0,524	2,11	16,70	1,018	66,90
23	0,607	1,95	16,20	1,155	62,34	62	0,534	2,13	16,95	1,045	68,42
24	0,617	1,97	16,45	1,182	63,86	63	0,546	2,15	17,20	1,073	69,94
25	0,628	1,99	16,70	1,209	65,38	64	0,555	2,17	17,45	1,100	71,46
26	0,638	2,01	16,95	1,237	66,90	65	0,565	2,19	17,70	1,127	72,98
27	0,649	2,03	17,20	1,264	68,42	66	0,576	2,21	17,95	1,155	74,50

Окончание табл. 4

K	X_{1k}	X_{2k}	X_{3k}	X_{4k}	Y_k	K	X_{1k}	X_{2k}	X_{3k}	X_{4k}	Y_k
28	0,659	2,05	17,45	1,291	69,94	67	0,586	2,23	18,20	1,182	76,02
29	0,669	2,07	17,70	1,319	71,46	68	0,513	2,05	11,45	1,209	72,98
30	0,680	2,09	17,95	1,346	72,98	69	0,524	2,07	11,70	1,237	74,50
31	0,690	2,11	18,20	1,373	74,50	70	0,534	2,09	11,95	1,264	76,02
32	0,701	2,13	18,45	1,401	76,02	71	0,545	2,11	12,20	1,291	77,54
33	0,711	2,15	18,70	1,428	77,54	72	0,555	2,13	12,45	1,319	79,06
34	0,472	1,93	12,70	1,100	44,10	73	0,565	2,15	12,70	1,346	80,58
35	0,482	1,95	12,95	1,127	45,62	74	0,576	2,17	12,95	1,373	82,10
36	0,493	1,97	13,20	1,155	47,14	75	0,586	2,19	13,20	1,401	83,62
37	0,503	1,99	13,45	1,182	48,66	76	0,597	2,21	13,45	1,428	85,14
38	0,513	2,01	13,70	1,209	50,18	77	0,607	2,23	13,70	1,455	86,66
39	0,524	2,03	13,95	1,237	51,70	78	0,617	2,25	13,95	1,483	88,18

Таблица 5

Коэффициенты парной корреляции

	X_1	X_2	X_3	X_4	Y
X_1	1	0,059	0,327	0,631	0,184
X_2	0,059	1	0,276	0,631	0,729
X_3	0,327	0,276	1	0,323	0,613
X_4	0,631	0,631	0,323	1	0,494
Y	0,480	0,729	0,613	0,494	1

Таблица 6

Переход числовых значений факторов в коды

X_i	Области		
	x_i=-1	x_i=0	x_i=+1
X_1	$X_{1j}\le 0{,}575$	$0{,}575< X_{1j} <0{,}612$	$X_{1j} \ge 0{,}612$
X_2	$X_{2j}\le 2{,}01$	$2{,}01< X_{2j} <2{,}08$	$X_{2j} \ge 2{,}08$
X_3	$X_{3j}\le 14{,}37$	$14{,}37< X_{3j} <15{,}59$	$X_{3j} \ge 15{,}59$
X_4	$X_{4j}\le 1{,}193$	$1{,}193< X_{4j} <1{,}287$	$X_{4j} \ge 1{,}287$

5. По таблице полной выборки ММТР определяем коэффициенты корреляции всех факторов и выходной величины по принципу «каждый с каждым» методами, изложенными в работе [5]. Результаты заносятся в табл. 5. При необходимости детального анализа таблицы коэффициентов парной корреляции рекомендуется воспользоваться методом корреляционных плеяд [6] в сочетании с экспертным методом весовых коэффициентов важности [7].

6. Так как согласно теории планирования эксперимента [8] моделированию подлежат только независимые (слабокоррелированные) факторы, то анализ корреляционной табл. 5 дает основание считать, что представленная многомерная виртуальная выборка табл. 4 содержит в себе три модели:

$Y = f(X_1, X_3)$, $Y = f(X_2, X_3)$ и $Y = f(X_3, X_4)$.

7. Найдем эти модели одним из методов, позволяющих получить адекватные математические модели по пассивным данным – модифицированным методом случайного баланса (ММСБ) [7]. Переход от числовых величин X_i к кодам x_i происходит по формуле $x_i >< \overline{X}_i \pm 0{,}25S_i$, числовые значения границ областей представлены в табл. 6.

В результате расчетов получились следующие адекватные математические модели:

$$\hat{Y} = 63{,}74 + 3{,}56x_3 + 3{,}56x_1x_3\,; \tag{1}$$

$$\hat{Y} = 64{,}04 + 7{,}78x_2 + 3{,}52x_3 - 6{,}23x_2x_3\,; \tag{2}$$

$$\hat{Y} = 64{,}09 + 3{,}18x_3 + 5{,}67x_4 - 4{,}87x_3x_4\,; \tag{3}$$

(при числовых значениях кодов $-1 \le x_i \le +1$).

Все три модели работоспособны, однако наилучшей из них является модель (2) как имеющая максимальную информационную емкость [7] и минимальную дисперсию адекватности.

2. Анализ полученных результатов

В отличие от классических методов обработки статистических данных, основанных, как правило, на свертывании этих данных (например, в виде гистограммы) и получающейся при этом потере информации, метод точечных распределений, опираясь на знание закона плотности вероятности генеральной совокупности, из которой взята выборка, «развертывает» выборочные данные и получает дополнительную информацию. Вопрос о величине объема выборки, эквивалентной в смысле полученной информации той, с которой мы имели бы дело при классических методах определения средних арифметических и эмпирических дисперсий, решен в работе [1] через D – статистику Колмогорова и аналогичную ей D_n – статистику, вычисленную в условиях малых выборок. Считая, что количество информации в малой и эквивалентной выборках одинаково, можно написать равенство для одной и той же доверительной вероятности

$$D\sqrt{n} = D_n\sqrt{n_э}\,, \tag{4}$$

откуда

$$n_э = n\left(\frac{D}{D_n}\right)^2, \tag{5}$$

где $n_э$ – объем эквивалентной информации выборки при условии нахождения оценок ее параметров по классическим формулам математической статистики.

Результаты расчетов для некоторых объемов (с округлением до целых) представлены в табл. 7.

Таблица 7

Реальные n и эквивалентные $n_э$ объемы выборок

n	3	5	8	10	15
D/D_n	1,683	1,816	1,681	1,636	1,601
$n_э$	10	16	23	27	38
$\frac{\sigma(S)}{\sigma}$,%	46,6	34,1	26,2	23,2	18,7
$\frac{\sigma(S_э)}{\sigma}$,%	23,2	18,1	15,0	13,8	12,6

В табл. 7 $\frac{\sigma(S)}{\sigma}$ и $\frac{\sigma(S_э)}{\sigma}$ – ошибка вычисления СКО.Если же пользоваться специально разработанными для МТР методами расчета параметров выборки, то общее увеличение эквивалентного объема достигает в среднем величины n^2. Поскольку максимальные величины ненормированных плотностей являются плавающими, то при состыковке МТР – выборок образуются неполные строки, что приводит к несколько заниженному объему конечной многофакторной выборки в диапазоне от n^2 до $3n^2$ строк.

Выводы

1. Найден принципиально новый метод построения адекватных многомерных моделей по выборкам малого объема.

2. Вероятность построения нескольких адекватных моделей на виртуальном пространстве большей выборки увеличивается с ростом числа факторов.

3. Разделение факторов на независимые (слабокоррелированные) группы необходимо производить с помощью корреляционных плеяд.

4. Из нескольких адекватных математических моделей, полученных из одной и той же многомерной виртуальной совокупности, наилучшей признается та, у которой

а) максимальная информационная емкость;

б) минимальная дисперсия адекватности;

в) минимальное число членов регрессии.

Литература

1. *Гаскаров Д.В. Малая выборка / Д.В. Гаскаров, В.И. Шаповалов. – М.: Статистика, 1978. – 248 с.*

2. *Долгов А.Ю. Повышение эффективности статистических методов контроля и управления технологическими процессами изготовления микросхем. – Дисс. на соиск. уч. степ. канд. техн. наук. – М.: МГАПИ, 2000. – 218 с.*

3. *Столяренко Ю.А. Контроль кристаллов интегральных схем на основе статистического моделирования методом точечных распределений. – Дисс. на соиск. уч. степ. канд. техн. наук. – М.: ГУП «Спурт», 2006. – 192 с.*

4. *Долгов А.Ю. Моделирование: Учебное пособие / А.Ю. Долгов, Ю.А. Столяренко. – Тирасполь: Изд-во Приднестровского ун-та, 2006. – 96 с.*

5. Митропольский А.К. Техника статистических вычислений / А.К. Митропольский. – 2-е изд., перераб.и доп. – М.: Наука, 1971. – 576 с.

6. Дружинин Г.В. Методы и оценки прогнозирования качества/ Г.В. Дружинин. – М.: Радио и связь. 1982. – 160 с.

7. Долгов Ю.А. Статистическое моделирование: Учебник для ВУЗов / Ю.А. Долгов. – Тирасполь: РИО ПГУ, 2002. – 280 с.

8. Адлер Ю.П. Планирование эксперимента при поиске оптимальных условий / Ю.П. Адлер, Е.В. Маркова, Ю.В. Грановский. – 2-е изд., перераб. и доп. – М.: Наука, 1976. – 279 с.

Поступила в редакцию 14.09. 2009

Рецензент: д-р техн. наук, проф., зав. кафедрой компьютерных систем и сетей В.С. Харченко, Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Харьков, Украина.

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ МЕТОДОМ БАГАТОМІРНИХ ТОЧКОВИХ РОЗПОВСЮДЖЕНЬ

Ю.О. Долгов, О.Ю. Долгов, Ю.О. Столяренко

Розроблено принципово новий метод отримання адекватних багатовимірних математичних моделей на базі істотних вибірок малого обсягу (n = 3÷15 рядків). Алгоритм розрахунку оснований на обробці кожного фактору істотньої вибірки методом точкових розповсюджень, котрий дозволяє шляхом використання знань шодо характеру закону розповсюджень штучно збільшувати обсяг вибірки. З'єднання індивідуальних факторних вибірок у єдину багатомірну вибірку великого обсягу (у діапазоні $n^2 \div 3n^2$) відбувається по стрічкам з максимальним рівнем ненормованої щільності ймовірності та з рівночасним відсіченням усіх неповних рядків. Подалі розрахунок виконується одним з методів моделювання за пасивними даними.

Ключові слова: вибірка малого обсягу, метод точкових розповсюджень, математичне моделювання за пасивними даними.

MULTIPLE-POINT-DISTRIBUTION SIMULATION

Y.A. Dolgov, A.Y. Dolgov, Y.A. Stolyarenko

A principle new method of adequate multiple mathematical model (simulation) on a base of small size sample (n=3÷15 lines) is worked up. The algorithm is based on treatment of every factor of small size sample by point-distribution method, which allow to raise of sample size from knowledge of distribution low. The joinder of individual factor samples in united multiple big size sample (in value from n^2 to $3n^2$) is happened on lines with maximum level of non-standartizing density function and with out off all incomplete lines simultaneously. The farther calculation is executed by one of passive data simulation method.

Kay words: small size sample, point-distribution method, passive data simulation.

Долгов Юрий Александрович – д-р техн. наук, профессор, зав. кафедрой информационных технологий и автоматизированного управления производственными процессами, Приднестровский государственный университет им. Т.Г. Шевченко, Тирасполь, Приднестровье, Молдова, e-mail: dolax@mail333.com.

Долгов Алексей Юрьевич – канд. техн. наук, доцент, зав. кафедрой автоматизированных производственных средств и систем Приднестровского государственного университета им. Т.Г. Шевченко, Тирасполь, Приднестровье, Молдова, e-mail: dolgov@ spsu.ru.

Столяренко Юлия Александровна – канд. техн. наук, доцент, доцент кафедры информационных технологий и автоматизированного управления производственными процессами Приднестровского государственного университета им. Т.Г. Шевченко, Тирасполь, Приднестровье, Молдова.