объединенный
институт
ядерных
исследований
дубна

Р10-86-547

П.В.Зрелов, В.В.Иванов

ФУНКЦИИ РАСПРЕДЕЛЕНИЯ СТАТИСТИКИ СМИРНОВА - КРАМЕРА - МИЗЕСА ДЛЯ МАЛЫХ ВЫБОРОК

Направлено в Оргкомитет
II Международной конференции по статистике,
Тампере, Финляндия, июнь 1987 г.

1986

ВВЕДЕНИЕ

Для проверки статистических гипотез о виде неизвестной функции распределения для данных, не сгруппированных в гистограмму, существует несколько критериев согласия, основанных на использовании функции правдоподобия, а также на статистиках Колмогорова и Смирнова-Крамера-Мизеса (СКМ).

Применение критериев согласия предполагает объем выборки из генеральной совокупности n достаточно большим, чтобы количественные характеристики статистик можно было считать асимптотическими, т.е. такими, которые получаются путем изучения поведения статистик при $n \to \infty$. В том случае, когда объем выборки сравнительно невелик, стараются применять критерий СКМ, поскольку соответствующая функция распределения случайной величины $n\omega_n^2$ (см. п.I) быстро сходится к асимптотической функции распределения. В литературе по математической статистике и обработке экспериментальных данных рекомендуется использовать таблицы асимптотической функции распределения в одних случаях при $n > 50$ /I/, в других - при $n > 40$ /2/. Такой подход ограничивает применение статистики СКМ областью больших n, хотя в действительности эта статистика может использоваться в качестве критерия согласия и при малых n ($n = 1,2,3,\ldots$). С другой стороны, иногда в литературе указывается, что можно использовать таблицы асимптотической функции распределения уже при $n \gg 3$ /3,4/. Это может приводить к существенным ошибкам в определении уровней значимости.

В приложениях вопрос задания определенного уровня значимости связан с пренебрежением событиями с малой вероятностью. Возможность такого пренебрежения зависит от практической важности следствий, вытекающих из наступления маловероятных событий, поэтому ошибка в определении уровня значимости может сказаться на результате исследований. Например, если изучаются события, относительный вклад которых в наблюдаемое на опыте распределение составляет от нескольких десятых до нескольких единиц процента, то изучаемый эффект может быть "потерян" из-за ошибки в определении уровня значимости, возникающей при использовании асимптотической функции распределения вместо функции распределения для конкретного n.

В книге Г.В.Мартынова /7/ упоминается работа /8/, в которой приведены некоторые процентные точки распределения статистики СКМ для выборок n = 2,3,4,5,8,10 и 20. Однако поскольку приведенная в /8/ таблица является неполной*), а часть процентных точек вычислена неточно (см. п.3), эта работа не снимает вопроса о функциях распределения при малых n.

Целью настоящей работы является вычисление значений функций распределения статистики СКМ при малых объемах выборки n = 2,3,4,5,6,7 и при n = 20,30,40,50 с использованием численного метода Монте-Карло в точках, являющихся процентилями асимптотического распределения СКМ.

I. МЕТОД ВЫЧИСЛЕНИЙ И ЕГО РЕАЛИЗАЦИЯ НА ЭВМ

Статистика СКМ имеет вид

$$n\omega_n^2 = n\int_{-\infty}^{+\infty}\left[S_n(x) - F(x)\right]^2 f(x)\,dx, \qquad (1)$$

где $F(x)$ – функция распределения случайной величины x, имеющей плотность $f(x)$, $S_n(x)$ – эмпирическая функция распределения, соответствующая выборке $x_1, x_2, \ldots, x_n$ /3/.

Асимптотическая функция распределения случайной величины $n\omega_n^2$, часто обозначаемая $a_1(Z)$, может быть представлена в виде разложения /5/:

$$a_1(Z) = \frac{1}{\pi\sqrt{Z}}\sum_{j=0}^{\infty}(-1)^j\binom{-\frac{1}{2}}{j}(4j+1)^{1/2}e^{-(4j+1)^2/16Z}K_{1/4}\left((4j+1)^2/16Z\right), \qquad (2)$$

где $Z = n\omega_n^2$, $K_{1/4}(x)$ – модифицированная функция Бесселя /6/.

Выражения для функций распределения $\Phi_n(Z)$ величины $n\omega_n^2$ при n = 1,2 были получены в работе /4/. Для случая n = 1 функция распределения $\Phi_1(Z)$ определяется следующим образом:

*) В работе /8/ приведены процентные точки для значений функции распределения 0,005; 0,010; 0,025; 0,050; 0,100 и 0,900; 0,950; 0,990; 0,995.

$$\Phi_1(Z) = \begin{cases} 0, & Z < 1/12, \\ (4Z - 1/3)^{1/2}, & 1/12 \leq Z \leq 1/3, \\ 1, & Z > 1/3, \end{cases} \qquad (3)$$

где $Z = \omega_1^2$.

Для случая n = 2 функция распределения $\Phi_2(Z)$ имеет уже более сложный вид*):

$$\Phi_2(Z) = \begin{cases} 0, & \text{если } Z < 1/24, \\ 2\pi\left(Z - \frac{1}{24}\right), & \text{если } 1/24 \leq Z \leq 5/48, \\ \left(Z - \frac{1}{24}\right)\left[2\pi - 4\arccos\frac{1}{4}\left(Z - 1/24\right)^{-1/2} + \frac{(Z - 3/48)^{1/2}}{Z - 1/24}\right], & 5/48 < Z \leq 1/6, \\ \left(Z + \frac{1}{24}\right)\left[\frac{3\pi}{2} - 2\arccos\frac{\frac{1}{4}(1/8)^{1/2} - (Z - 1/6)^{1/2}(Z - 5/48)^{1/2}}{Z - 1/24}\right] + & \\ \quad + \frac{1}{8} + \left(\frac{1}{2}\left(Z - 1/6\right)\right)^{1/2} + \frac{1}{2}\left(Z - 5/48\right)^{1/2}, & \text{если } \frac{1}{6} < Z \leq \frac{2}{3}, \\ 1, & \text{если } Z > \frac{2}{3}. \end{cases} \qquad (4)$$

В той же работе /4/ отмечается, что можно найти вид функции распределения $\Phi_3(Z)$ и для n = 3, однако выражения в этом случае очень громоздки. Получить вид функции распределения для случая n = 4 методом, предложенным в указанной работе, чрезвычайно сложно.

Авторы работы /8/, обобщив метод /4/, получили выражение для функции распределения для любого n на первых двух интервалах возможных значений Z:

$$\Phi_n(Z) = n!\,\pi^{\frac{1}{2}n}y^n / \Gamma\left(\tfrac{1}{2}n+1\right), \qquad 1/12n \leq Z \leq (n+3)/12n^2,$$

$$\Phi_n(Z) = n!\,\pi^{\frac{1}{2}(n-1)}y^n\left\{\sqrt{\pi}/\Gamma\left(\tfrac{1}{2}n+1\right) - \left(2\int_0^{\alpha}\sin^n t\,dt\right)/\Gamma\left(\tfrac{1}{2}n+\tfrac{1}{2}\right)\right\}, \qquad (5)$$

$$(n+3)/12n^2 < Z \leq (n+6)/12n^2,$$

где $y = (Z - 1/12n)^{1/2}$, $\alpha = \arccos(1/2ny)$.

*) В работе /4/ эта формула дана с ошибкой: в выражении для функции распределения на интервале $\frac{1}{6} < Z < \frac{2}{3}$ в квадратных скобках вместо числа $\frac{5}{48}$ указано $\frac{5}{40}$.

Ниже описывается способ вычисления значений функции распределения случайной величины $n\omega_n^2$ для любых n и Z.

По определению функция распределения случайной величины $\xi = n\omega_n^2$ равна

$$\Phi_n(Z) = P\{\xi < Z\}.$$

Распределение $n\omega_n^2$ не зависит от вида распределения $f(x)$ /3/, и выражение (I) может быть переписано в виде

$$n\omega_n^2 = n\int_0^1 [S_n(y) - y]^2 dy \quad , \tag{6}$$

где y равномерно распределено на интервале (0,I). Величину $\xi = n\omega_n^2$ можно рассматривать как функцию n аргументов y_i :

$$\xi = f(y_1, y_2, \dots, y_n).$$

Элементы выборки $y_1, y_2, \dots, y_n$, рассматриваемые как координаты, определяют точку в n-мерном кубе $C_n = [0,I]^n$. Уравнение $\xi(y_1, y_2, \dots, y_n) = Z$ задает поверхность в кубе C_n. Поскольку y_i -равномерно распределенные случайные величины, то отношение объема под поверхностью

$$\xi(y_1, y_2, \dots, y_n) < Z \tag{7}$$

к объему куба C_n определяет искомую вероятность $P\{\xi < Z\} = V_Z/V$, где V_Z -объем области n-мерного пространства, отвечающий условию (7), а V - объем n-мерного куба C_n, равный I.

Величина V_Z вычисляется геометрическим методом Монте-Карло как отношение числа случайных точек N_Z, удовлетворяющих условию (7), к общему числу точек N, умноженное на V :

$$V_Z = \frac{N_Z}{N} V.$$

Отсюда следует, что искомая вероятность равна $P\{\xi < Z\} = \frac{N_Z}{N}$ с точностью, определяемой геометрическим методом Монте-Карло.

Процедура вычисления функции распределения состоит в следующем: генерируется n равномерно распределенных случайных чисел y_i в интервале (0,I). Эти числа располагаются в виде вариационного ряда и

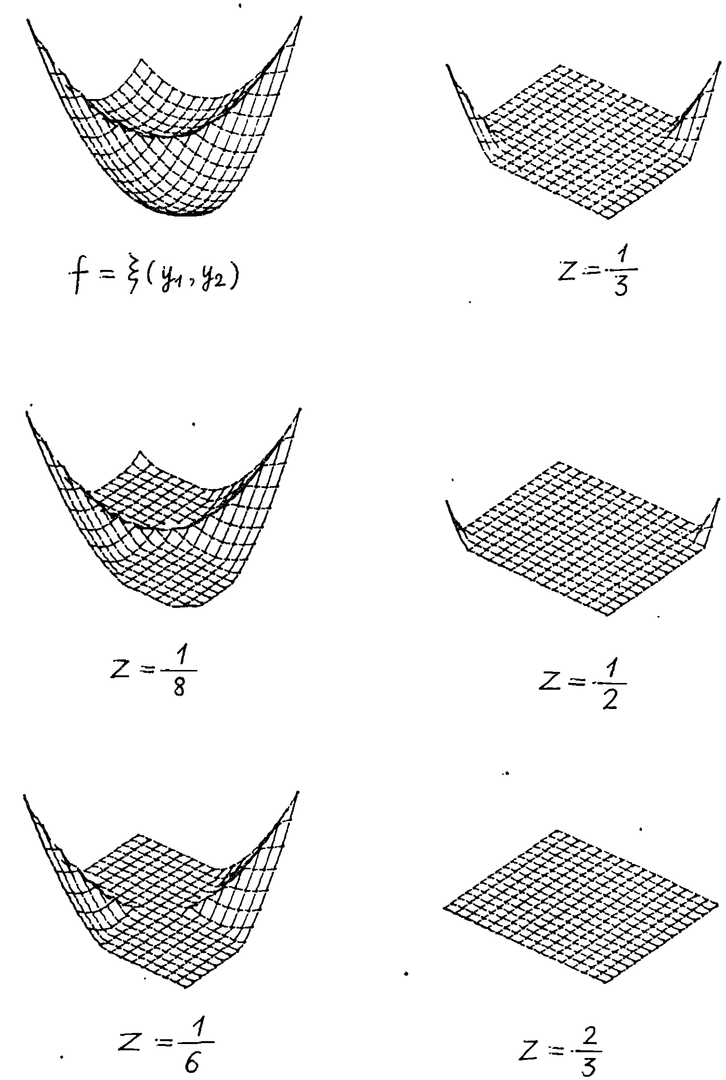

Рис.I. Представлена поверхность $f = \xi(y_1, y_2)$, а также поверхности, определяемые условием $f = \max\{Z, \xi(y_1, y_2)\}$ для $Z = \frac{1}{8}$; $\frac{1}{6}$; $\frac{1}{3}$; $\frac{1}{2}$; $\frac{2}{3}$.

используются для вычисления величины $\xi = n\omega_n^2$ по формуле

$$n\omega_n^2 = \frac{1}{12n} + \sum_{k=1}^{n}\left[\frac{2k-1}{2n} - y_k\right]^2,$$

которая эквивалентна формуле (6) при условии $y_1 < y_2 < \dots < y_n$. Величина ξ затем сравнивается с фиксированным значением Z. Таким образом, для каждого Z определяется N_Z, а величина $\phi_n(Z)$ оценивается как $\frac{N_Z}{N}$. N задается таким образом, чтобы обеспечить необходимую точность вычислений (см. п.2).

Для $n = 2$ указанный метод имеет наглядную геометрическую интерпретацию. В этом случае условие $\xi(y_1, y_2) < Z$ определяет область в единичном квадрате $C_2 = [0,1]^2$. Площадь этой области определяет вероятность $P\{\xi < Z\}$. На рис.I представлена поверхность $f = \xi(y_1, y_2)$, а также поверхности, определяемые условием $f = max\{Z, \xi\}$ для $Z = \frac{1}{8}$; $\frac{1}{6}$; $\frac{1}{3}$; $\frac{1}{2}$; $\frac{2}{3}$. Площади горизонтальных участков этих поверхностей пропорциональны значениям функции распределения $\phi_2(Z)$ в соответствующих точках Z.

2. ОШИБКИ ВЫЧИСЛЕНИЙ

В используемом для решения данной задачи методе Монте-Карло оценкой $\phi_n(Z)$ служит величина

$$\theta_N = \frac{1}{N}\sum_{i=1}^{N} Z_i, \qquad (8)$$

где

$$Z_i = \begin{cases} 1, & \text{если } \xi = f(y_1, y_2, \dots, y_n) < Z, \\ 0, & \text{если } \xi \geqslant Z. \end{cases} \qquad (9)$$

Обозначим $I = \phi_n(Z)$. Рассмотрим дисперсию случайной величины Z:

$$DZ = MZ^2 - (MZ)^2.$$

В силу (9) $MZ^2 = MZ$, т.о.,

$$DZ = I - I^2. \qquad (10)$$

При достаточно больших N имеет место неравенство /I/:

$$|\tilde{I} - I| \leqslant x_\beta \sqrt{\frac{DZ}{N}},$$

Таблица I

Функции распределения случайной величины $n\omega_n^2$:

$\Phi_n(z) = Pr(n\omega_n^2 < z)$

N%	z	$\Phi_1(z)$	$\Phi_2(z)$	$\Phi_3(z)$	$\Phi_4(z)$	$\Phi_5(z)$	$\Phi_6(z)$	$\Phi_7(z)$	$a_1(z)$
1	.02480	.000	.000	.000	.002	.004	.005	.006	.010
2	.02878	.000	.000	.001	.007	.010	.012	.013	.020
3	.03177	.000	.000	.006	.014	.017	.020	.021	.030
4	.03430	.000	.000	.013	.021	.025	.028	.030	.040
5	.03656	.000	.000	.021	.029	.034	.037	.039	.050
6	.03865	.000	.000	.028	.038	.042	.046	.048	.060
7	.04061	.000	.000	.037	.046	.051	.055	.057	.070
8	.04247	.000	.005	.045	.055	.060	.064	.066	.080
9	.04427	.000	.016	.053	.063	.070	.073	.075	.090
10	.04601	.000	.027	.062	.072	.079	.083	.085	.100
11	.04772	.000	.038	.071	.082	.089	.092	.094	.110
12	.04939	.000	.049	.080	.091	.098	.102	.104	.120
13	.05103	.000	.059	.089	.101	.108	.111	.113	.130
14	.05265	.000	.069	.099	.111	.118	.121	.123	.140
15	.05426	.000	.079	.108	.121	.127	.130	.133	.150
16	.05586	.000	.089	.118	.131	.137	.140	.143	.160
17	.05746	.000	.099	.128	.141	.147	.150	.153	.170
18	.05904	.000	.109	.138	.151	.157	.160	.162	.180
19	.06063	.000	.119	.148	.161	.167	.169	.172	.190
20	.06222	.000	.129	.158	.171	.176	.179	.182	.200
21	.06381	.000	.139	.168	.181	.186	.189	.192	.210
22	.06541	.000	.149	.178	.192	.196	.199	.202	.220
23	.06702	.000	.159	.188	.202	.206	.209	.213	.230
24	.06863	.000	.169	.199	.212	.216	.220	.223	.240
25	.07025	.000	.180	.209	.222	.226	.230	.233	.250
26	.07189	.000	.190	.219	.232	.236	.240	.243	.260
27	.07354	.000	.200	.229	.242	.246	.250	.253	.270
28	.07521	.000	.211	.240	.252	.256	.260	.263	.280
29	.07690	.000	.221	.250	.262	.266	.271	.274	.290
30	.07860	.000	.232	.261	.272	.276	.281	.284	.300
31	.08032	.000	.243	.272	.282	.286	.291	.294	.310
32	.08206	.000	.254	.283	.292	.297	.301	.304	.320
33	.08383	.045	.265	.294	.303	.307	.312	.314	.330
34	.08562	.096	.276	.305	.313	.317	.322	.324	.340
35	.08744	.128	.288	.316	.323	.328	.333	.335	.350
36	.08928	.154	.299	.327	.333	.338	.343	.345	.360
37	.09115	.177	.311	.338	.343	.348	.353	.355	.370
38	.09306	.197	.323	.348	.353	.359	.364	.365	.380
39	.09499	.216	.335	.359	.363	.369	.374	.376	.390
40	.09696	.233	.347	.370	.373	.380	.384	.386	.400
41	.09896	.250	.360	.381	.384	.390	.394	.396	.410
42	.10100	.266	.373	.391	.394	.401	.405	.406	.420
43	.10308	.281	.386	.402	.404	.411	.415	.417	.430
44	.10520	.296	.399	.412	.415	.422	.425	.427	.440
45	.10736	.310	.411	.423	.425	.432	.435	.437	.450
46	.10956	.324	.422	.433	.436	.443	.446	.447	.460
47	.11182	.338	.434	.443	.446	.453	.456	.458	.470
48	.11412	.351	.445	.453	.457	.464	.466	.468	.480
49	.11647	.364	.456	.464	.467	.474	.477	.478	.490
50	.11888	.377	.467	.474	.478	.485	.487	.488	.500

N%	z	$\Phi_1(z)$	$\Phi_2(z)$	$\Phi_3(z)$	$\Phi_4(z)$	$\Phi_5(z)$	$\Phi_6(z)$	$\Phi_7(z)$	$a_1(z)$
51	.12134	.390	.478	.484	.489	.495	.497	.499	.510
52	.12387	.403	.489	.494	.499	.506	.508	.509	.520
53	.12646	.415	.499	.504	.510	.516	.518	.519	.530
54	.12911	.428	.510	.515	.521	.527	.528	.530	.540
55	.13183	.440	.521	.525	.532	.537	.539	.540	.550
56	.13463	.453	.532	.535	.542	.548	.549	.550	.560
57	.13751	.466	.543	.546	.553	.558	.559	.561	.570
58	.14046	.478	.554	.556	.564	.568	.570	.571	.580
59	.14350	.491	.565	.566	.575	.579	.580	.581	.590
60	.14663	.503	.576	.577	.586	.589	.591	.592	.600
61	.14986	.516	.587	.587	.597	.600	.601	.602	.610
62	.15319	.529	.599	.598	.607	.610	.611	.612	.620
63	.15663	.541	.610	.608	.618	.620	.622	.623	.630
64	.16018	.554	.622	.619	.629	.631	.632	.633	.640
65	.16385	.568	.634	.630	.640	.641	.642	.643	.650
66	.16765	.581	.646	.641	.651	.652	.653	.654	.660
67	.17159	.594	.657	.652	.661	.662	.663	.664	.670
68	.17568	.608	.667	.663	.672	.673	.674	.675	.680
69	.17992	.622	.678	.674	.683	.683	.684	.685	.690
70	.18433	.636	.688	.685	.693	.693	.694	.695	.700
71	.18892	.650	.698	.697	.704	.704	.705	.706	.710
72	.19371	.664	.709	.708	.714	.714	.715	.716	.720
73	.19870	.679	.719	.719	.725	.725	.726	.726	.730
74	.20392	.695	.729	.730	.735	.735	.736	.737	.740
75	.20939	.710	.739	.742	.745	.746	.746	.747	.750
76	.21512	.726	.750	.753	.756	.756	.757	.758	.760
77	.22114	.742	.760	.765	.766	.767	.767	.768	.770
78	.22748	.759	.770	.777	.777	.777	.778	.778	.780
79	.23417	.777	.781	.788	.787	.788	.788	.789	.790
80	.24124	.795	.791	.799	.798	.798	.799	.799	.800
81	.24874	.813	.802	.811	.809	.809	.809	.809	.810
82	.25670	.833	.812	.822	.819	.819	.820	.820	.820
83	.26520	.853	.823	.833	.830	.830	.830	.830	.830
84	.27429	.874	.834	.844	.840	.840	.840	.840	.840
85	.28406	.896	.845	.855	.851	.851	.851	.851	.850
86	.29460	.919	.856	.866	.861	.861	.861	.861	.860
87	.30603	.944	.868	.876	.872	.872	.871	.871	.870
88	.31849	.970	.879	.886	.883	.882	.882	.882	.880
89	.33217	.998	.891	.896	.893	.893	.892	.892	.890
90	.34730	1.000	.903	.906	.904	.903	.902	.902	.900
91	.36421	1.000	.915	.916	.914	.913	.913	.913	.910
92	.38331	1.000	.928	.926	.925	.924	.923	.923	.920
93	.40520	1.000	.940	.937	.935	.934	.933	.933	.930
94	.43077	1.000	.953	.947	.946	.944	.944	.943	.940
95	.46136	1.000	.965	.957	.956	.954	.954	.953	.950
96	.49929	1.000	.978	.968	.966	.965	.964	.963	.960
97	.54885	1.000	.990	.978	.976	.975	.974	.973	.970
98	.61981	1.000	.998	.988	.985	.984	.983	.983	.980
99	.74346	1.000	1.000	.997	.994	.993	.993	.992	.990
100	1.16786	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999

где $\tilde{I} = \theta_N$, x_β – корень уравнения $\Phi(x) = \beta$, $\Phi(x)$ – функция ошибок, а β – коэффициент доверия. Учитывая (10), для оценки ошибки вычислений получаем

$$|\theta_N - I| \leq x_\beta \sqrt{\frac{I - I^2}{N}} .$$

Оценка ошибки вычислений производилась с коэффициентом доверия $\beta = 0{,}997$, $x_\beta = 3$. N выбиралось таким, чтобы ошибка вычислений была меньше 0,0009 для выборок $n = 2,3,4,5,6,7$ и меньше 0,009 для выборок $n = 20,30,40,50$.

3. АНАЛИЗ РЕЗУЛЬТАТОВ ВЫЧИСЛЕНИЙ

В таблице I приведены значения функции распределения $\Phi_n(Z)$ в точках, являющихся процентилями асимптотического распределения для $n = 2,3,4,5,6,7$ (значения Z взяты из работы /5/). Точность значений $\Phi_n(Z)$ равна единице последнего знака.

В таблицу включены значения асимптотической функции распределения $a_1(Z)$, которые вычислялись по формуле (2), а также значения $\Phi_1(Z)$, вычисленные по формуле (3).

Значения $\Phi_2(Z)$ сравнивались со значениями, вычисленными по формулам (4). В пределах указанной точности эти величины совпадают.

Вычисления, проведенные для $n = 20,30,40,50$ с точностью 0,01, показали, что значения функций распределения для этих выборок совпадают в пределах ошибки вычислений с $a_1(Z)$, поэтому они не были включены в таблицу.

Величины максимальных отклонений значений функций распределения $\Phi_i(Z)$ $(i = 2,3,\ldots,7)$ от значений асимптотической функции распределения с учетом возможной ошибки вычислений приведены в таблице 2.

Таблица 2. Максимальные значения разности $|\Phi_n(Z) - a_1(Z)|$ для малых выборок

Объем выборки	2	3	4	5	6	7
$\lvert\Phi_n(Z) - a_1(Z)\rvert$	0,034	0,027	0,023	0,016	0,014	0,013

Значения функции распределения для $n = 2,3,4,5$, указанные в таблице I, сопоставлялись с табличными данными работы /8/. Для этого данные табл. I аппроксимировались полиномиальной зависимостью 5-й степени. Значение полинома в рассматриваемой процентной точке сравнивалось со значением, приведенным в /8/. Для выборки n = 2 таблицы хорошо согласуются. Для распределений с n = 3,4,5 значения процентных точек, полученные авторами /8/ обратной интерполяцией формул (5), согласуются с данными таблицы I, а значения, полученные с использованием аппроксимации распределения ω^2 распределением χ^2 или с использованием метода Монте-Карло, заметно отличаются от приведенных в таблице I. Величины максимальных отклонений для выборок n = 3,4,5 составляют с учетом возможной ошибки вычислений соответственно 0,007, 0,004 и 0,009.

ЗАКЛЮЧЕНИЕ

В работе предложен способ вычисления функций распределения статистики Смирнова-Крамера-Мизеса для малых выборок. С его помощью вычислены значения функций распределения для n = 2,3,4,5,6,7 с точностью 0,00I и для n = 20,30,40,50 с точностью 0,0I в точках, являющихся процентилями асимптотического распределения указанной статистики. В отличие от работы /8/ значения функций распределения вычислены во всей области их изменения, в существенно большем числе точек и с высокой точностью. Кроме того, впервые получены данные для n = 6 и 7.

Сравнение полученных табличных данных с соответствующими значениями функции $a_1(Z)$ позволяет сделать вывод, что в случае малых выборок асимптотический предел достигается не так быстро, как это отмечалось в работах /3,4/, указываются возможные ошибки в определении значений функций распределения при использовании $a_1(Z)$ вместо $\Phi_n(Z)$ для n = 2,3,4,5,6,7. Использование данных таблицы I дает возможность корректно применять критерий Смирнова-Крамера-Мизеса в случае малых выборок. В частности, для n = 2,3,4,5,6,7 доверительная вероятность определяется с точностью 0,I%.

Авторы считают приятным долгом поблагодарить Г.А.Ососкова и Н.И.Чернова за полезные обсуждения и поддержку.

ЛИТЕРАТУРА

I. Соболь И.М. Численные методы Монте-Карло. "Наука", М., I973.
2. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики. "Наука", М., I965.

3. Статистические методы в экспериментальной физике. Пер. с англ. Под ред. А.А.Тяпкина. Атомиздат, М., I976.
4. Marshall A.W. The Small Sample Distribution nw_n^2, Ann. Math. Statist., 1958, v.29, p.307.
5. Anderson T.W., Darling D.A. Asimptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. Ann. Math. Statist., 1952, v.23, p.193.
6. Янке Е., Эмде Ф., Леш Ф. Специальные функции. "Наука", М., I977.
7. Мартынов Г.В. Критерии омега-квадрат. "Наука", М., I978.
8. Stephens M., Maag U.R. Futher Percentage Points for W^2, Biometrica, 1968, v.55, p.428.

Рукопись поступила в издательский отдел
8 августа I986 года.

НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

Вы можете получить по почте перечисленные ниже книги, если они не были заказаны ранее.

Зрелов П.В., Иванов В.В.

P10-86-547

Функции распределения статистики Смирнова - Крамера - Мизеса для малых выборок

Предложен способ вычисления функций распределения статистики Смирнова - Крамера - Мизеса. С его помощью вычислены значения функций распределния $\Phi_n(z)$ для выборок размером $n = 2,3,4,5,6,7$ с точностью 0,001 и для $n = 20,30,40,50$ с точностью 0,01 в точках, являющихся процентилями асимптотического распределения $a_1(z)$ указанной статистики. Значения функций распределения вычислены во всей области их изменения, в большом числе точек и с высокой точностью. Кроме того, впервые получены данные для $n = 6$ и 7. Сравнение полученных табличных данных с соответствующими значениями функции $a_1(z)$ позволяет сделать вывод о том, что в случае малых выборок асимптотический предел достигается не так бысто, как это отмечалось в некоторых работах. Использование табличных данных работы позволяет корректно применять критерий соответствия Смирнова - Крамера - Мизеса в случае малых выборок.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна 1986

Перевод О.С.Виноградовой

Zrelov P.V., Ivanov V.V.

P10-86-547

The Small Sample Cumulative Distributions of the Smirnov - Cramer - Mises Statistics

The method for computing the small sample cumulative distributions of the Smirnov - Cramer - Mises statistic is presented. By means of this method the values of the cumulative functions $\Phi_n(z)$ for the sample size $n = 2,3,4,5,6,7$ with the accuracy 0.001 and for $n = 20,30,40,50$ with the accuracy 0.01 at the percentage points of the asymptotic distribution $a_1(z)$ are calculated. The values of $\Phi_n(z)$ are calculated in the whole range of z, in the large amount of points and with high accuracy. The values of $\Phi_n(z)$ for $n = 6$ and 7 are presented for the first time. The comparison of the obtained results with the function $a_1(z)$ at selected points leads to the conclusion that the convergence to the asymptotic distribution is not so extremely rapid as it was mentioned in some papers. The results of our calculations enables the correct application of the Smirnov - Cramer - Mises test in the small sample cases.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna 1986