

Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа прикладной математики и вычислительной физики

Лабораторная работа
по дисциплине «Анализ данных с интервальной неопределенностью»
на тему **«Обработка постоянной. Применение меры совместности
к анализу данных»**

Выполнил

студент гр. 5040102/10201

Гусаров Д.А.

/_____/

Руководитель

доцент, к.ф.-м.н.

Баженов А.Н.

/_____/

Санкт-Петербург

2022

Постановка задачи

Проводится исследование из области солнечной энергетики.

Калибровка датчика ФП2 производится по эталону ФП1. Зависимость между квантовыми эффективностями датчиков предполагается постоянной для каждой пары наборов измерений

$$QE_1 = \frac{X_1}{X_2} \cdot QE_2 \quad (1)$$

QE_1 , QE_2 – эталонная эффективность эталонного и исследуемого датчика, X_1 , X_2 , или $\{x_{1i}\}_{i=1}^{200}$, $\{x_{2i}\}_{i=1}^{200}$ – измеренные мощности. Данные датчиков находятся в файлах “Канал 1_700nm_0.03.csv” и “Канал 2_700nm_0.03.csv”, или в Приложении 3.

Требуется определить параметры постоянной величины на основе двух выборок $\{x_{1i}\}_{i=1}^{200}$, $\{x_{2i}\}_{i=1}^{200}$, в частности коэффициент калибровки

$$R_{12} = \frac{X_1}{X_2} \quad (2)$$

при помощи линейной регрессии, интервальных данных и коэффициента Жаккара.

Теория и методы

Представим данные таким образом, чтобы применить понятия статистики данных с интервальной неопределенностью. Один из распространенных способов получения интервальных результатов в первичных измерениях – это «обинтерваливание» точечных значений, когда к точечному базовому значению x_{1i} , которое считывается по показаниям измерительного прибора прибавляется интервал погрешности ε

$$X_{1i} = x_{1i} + [-\varepsilon, +\varepsilon] \quad (3)$$

В конкретных измерениях $\varepsilon = 10^{-4}$ мВ. Согласно терминологии интервального анализа, рассматриваемая выборка – это вектор интервалов, или интервальный вектор $X_1 = \{X_{1i}\}_{i=1}^{200}$

Интервалы будем строить простым способом. Вначале построим линейную регрессию по известному методу наименьших квадратов в виде $L_1(n) = A_1 \cdot n + B_1$, где n – номер измерения; $L_1(n)$ – прямая, аппроксимирующая экспериментальные измерения $\{x_{1i}\}_{i=1}^{200}$. Отклонение можно вычислить как

$$\varepsilon_{1n} = |x_{1n} - L_1(n)| \quad (4)$$

Если отдельные интервалы не заключают в себе линейную регрессию, к отклонению ε_{1n} стоит растянуть, домножить на величину w_n , минимально возможную, для того, чтобы интервал коснулся линии регрессии.

Окончательно, интервальные данные представимы в виде:

$$X_{1n} = x_{1n} + [-\tilde{\varepsilon}_n, +\tilde{\varepsilon}_n] \quad (5)$$

или кратко X_1 – множество всех интервальных данных, построенных по измерениям датчика ФП1, $\tilde{\varepsilon}_n = w_n \cdot \varepsilon$, $w_n \geq 1$.

Чтобы сделать интервальную величину более константной и в дальнейшем оценить совместность двух выборок экспериментальных измерений, вычтем из интервальных данных линейную зависимость (фактически из концов интервала), получим:

$$X'_1 \leftarrow X_1 - A_1 \cdot n \quad (6)$$

Для базовых значений x_{2i} выполним аналогичные вычисления. Найдем линейную зависимость $L_2(n) = A_2 \cdot n + B_2$, интервалы X_{2i} по формуле (5) и обработанные интервалы X'_2 по формуле (6) с соответствующими индексами.

В различных областях анализа данных используют различные меры сходства множеств, иными словами, коэффициенты сходства. Будем использовать мультимеру Жаккара, то есть ее модификацию для интервальных данных:

$$JK = \frac{wid(\cap y_i)}{wid(\cup y_i)} \quad (7)$$

Мера Жаккара $-1 \leq JK \leq 1$ численно характеризует меру совместности интервальных данных. В качестве y_i рассматриваются интервальные данные объединенной выборки $X' = \{X'_1, RX'_2\}$. JK – число, получаемое в результате деления пересечения интервалов на их объединение. Заметим, что если при подборе калибровочного множителя R получается $JK > 0$, то выборка совместна (имеет положительную меру совместности). Поиск оптимального R_{opt} можно представить так:

$$R_{opt} = arg \left\{ \max_R JK(X') \right\} \quad (8)$$

R_{opt} – это аргумент, у которого реализуется данный функционал, максимальная оценка коэффициента калибровки R_{12} из формулы (2). Внешнюю оценку для R_{opt} можно найти разными способами, проще всего путем деления интервалов двух выборок $R = \frac{X_1}{X_2}$, в результате чего получим интервал внешней оценки $[R, \overline{R}]$ – такой интервал, в котором можно найти R_{opt} , перебирая R с некоторым шагом и вычисляя функционал (8). Интервал, в пределах которого наблюдается $JK > 0$ является внутренней оценкой коэффициента R_{opt} .

Результаты

Программный код написан на языке программирования Python с использованием библиотек Matplotlib, NumPy и Sklearn.

На рис.1 представлены экспериментальные данные, измеренные двумя датчиками, на рис.2 – те же данные, но в другом масштабе для наглядности. Можно заметить, что данные в начале и в конце диапазона измерений не достаточно хорошо описываются прямой (приблизительно первые и последние 20 точек), их можно исключить для более корректного построения линейной регрессии. На рис. 4 и 5 показаны построенные согласно описанной выше теории интервальные данные и линейная регрессия с коэффициентами $A_1 \approx 3.005 \cdot 10^{-6}$, $B_1 \approx 0.0716$, $A_2 \approx 2.284 \cdot 10^{-6}$, $B_2 \approx 0$.

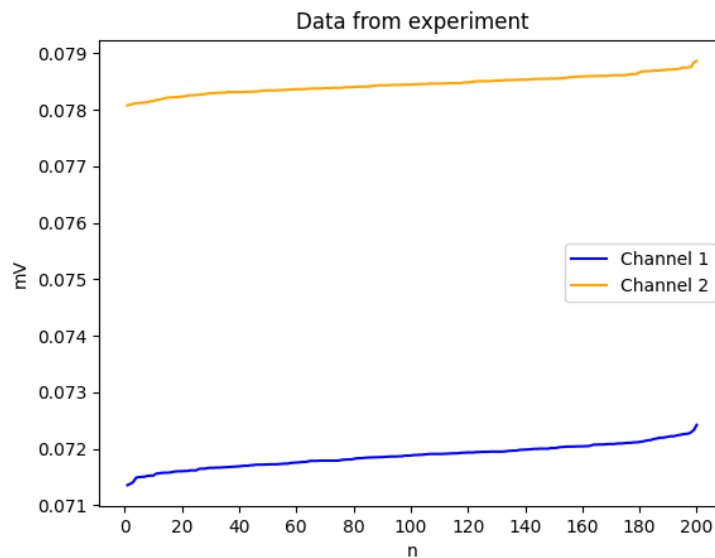


Рис. 1. Две выборки экспериментальных данных, измеренным датчиками

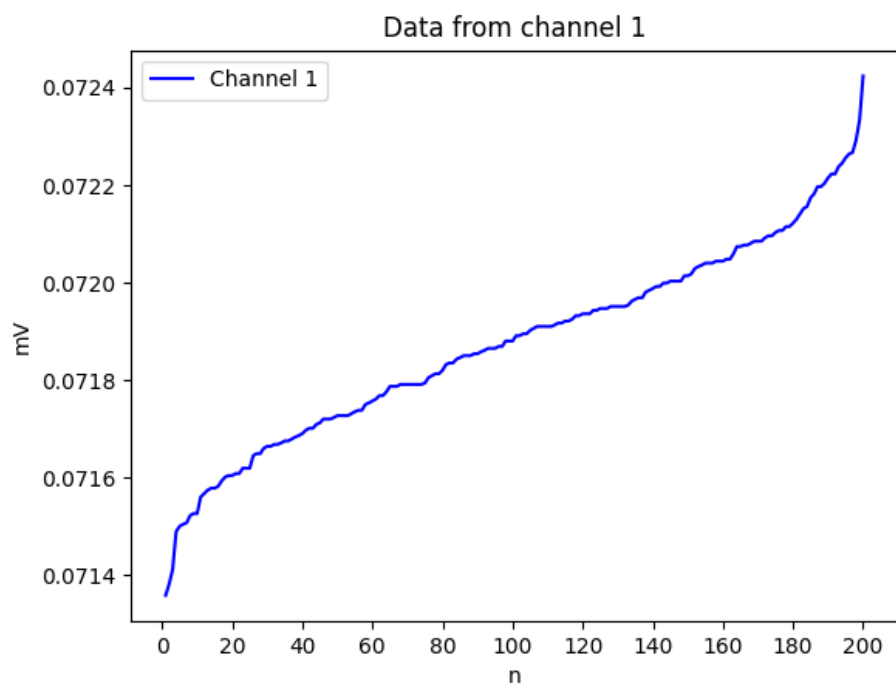


Рис. 2. Данные, измеренные датчиком ФП1

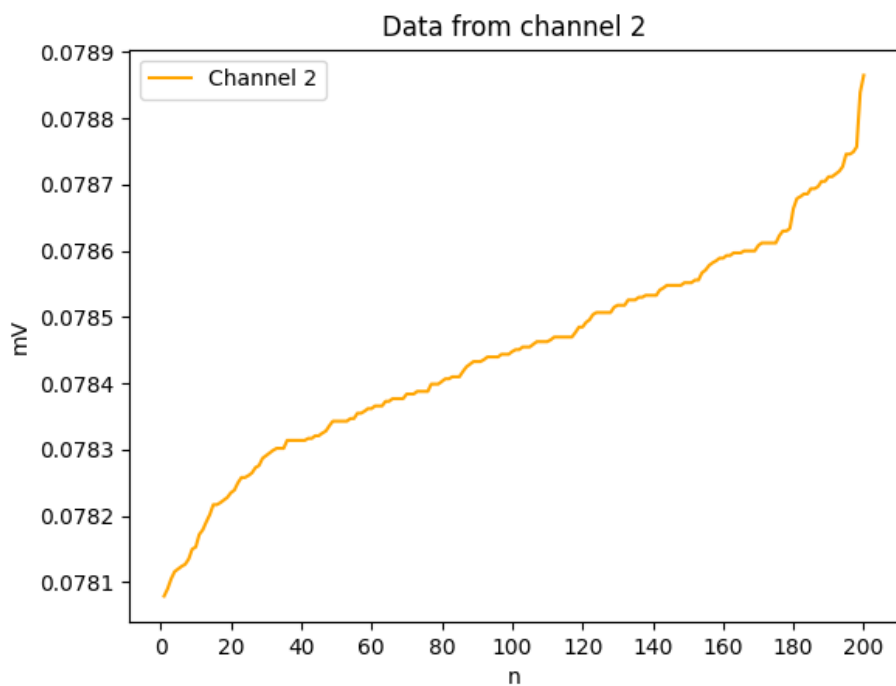


Рис. 3. Данные, измеренные датчиком ФП2

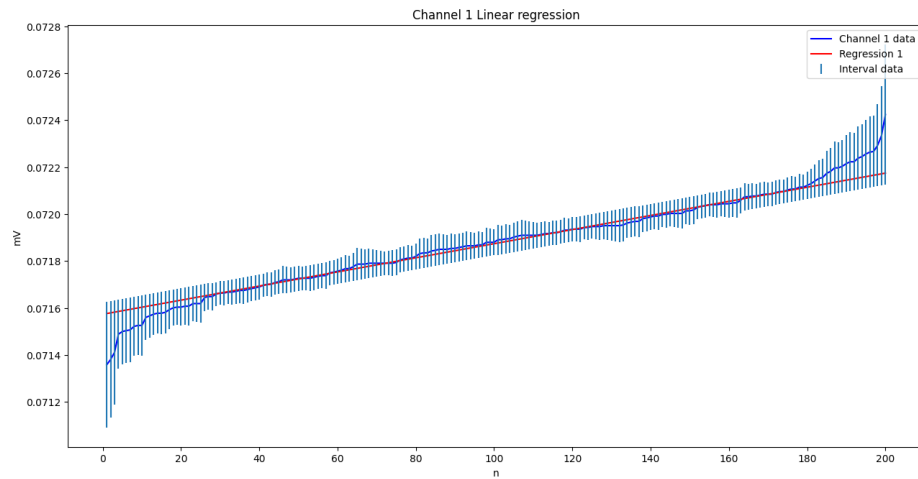


Рис. 3. Интервальные данные первой выборки и линейная регрессия

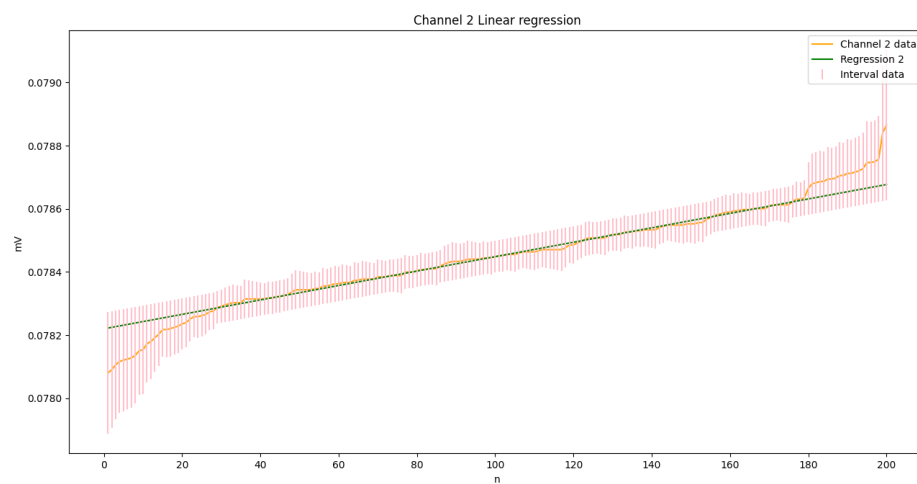


Рис. 4. Интервальные данные второй выборки и линейная регрессия

На рис. 5 визуализирован пример совместных выборок X'_1, RX'_2 , что выполняется при R , обеспечивающим $JK > 0$.

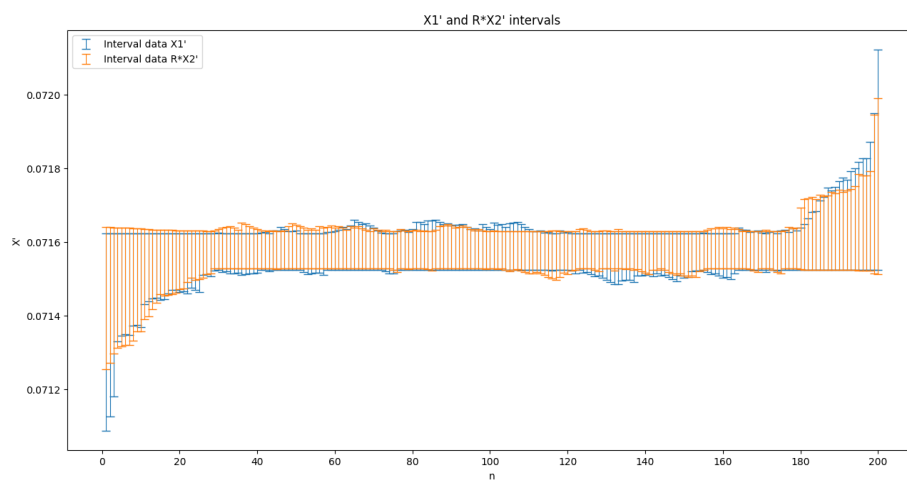


Рис. 5. Обработанные интервальные данные совместной выборки при R , обеспечивающем совместность выборок

На рис. 6 показана зависимость коэффициента Жаккара от коэффициента калибровки R . Согласно внешней оценке оптимальное значение оптимального R_{opt} осуществлялся в диапазоне $[R, \bar{R}] \approx [0.90943, 0.92264]$. Как интервал можно представить $R_{12} \approx [0.91385, 0.91623]$. В нашем эксперименте, максимум коэффициента Жаккара имеет значение 0.088, чуть меньше 0.1, но совершенно не близкое к 1. Это с наличием различных погрешностей, которые на практике невозможно устранить, но несмотря на их присутствие, поведение коэффициента Жаккара позволило найти оптимальный калибровочный коэффициент $R_{opt} \approx 0.91504$. Знак коэффициента Жаккара может свидетельствовать о том, является ли минимум по включению правильным интервалом, и в данном случае говорит о совместности двух выборок. Таким образом, можно сказать, что область, где $JK(R_{12}) \geq 0$ является оценкой искомой величины R_{12} .

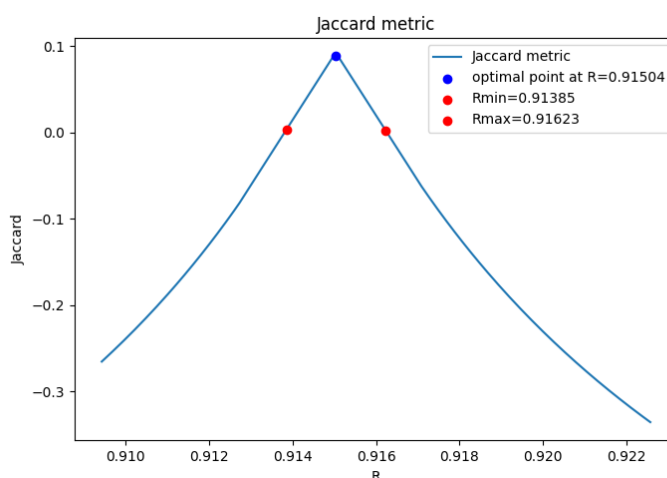


Рис. 6. Значения коэффициента Жаккара от коэффициента калибровки

Ссылка на GitHub с реализацией

https://github.com/dimerf99/interval_analysis/tree/main/lab_1

Файлы данных:

Канал 1_700nm_0.03.csv

Канал 2_700nm_0.03.csv

Коэффициенты линейной регрессии

№ выборки	A_i	B_i
1	3.005e-06	0.0715736
2	2.284e-06	0.0782195

Экспериментальные данные

Экспериментальные данные (точечные) и значения w_n для коррекции погрешности $\varepsilon = 10^{-4}$

№ изм.	Выборка 1		Выборка 2	
	X_{1n} , мВ	w_n	X_{2n} , мВ	w_n
1	7.1358E-02	2.19	7.8079E-02	1.43
2	7.1381E-02	1.99	7.8090E-02	1.34
3	7.1411E-02	1.72	7.8105E-02	1.21
4	7.1489E-02	1.0	7.8116E-02	1.13
5	7.1500E-02	1.0	7.8120E-02	1.11
6	7.1504E-02	1.0	7.8124E-02	1.09
7	7.1507E-02	1.0	7.8127E-02	1.09
8	7.1522E-02	1.0	7.8135E-02	1.03
9	7.1526E-02	1.0	7.8150E-02	1.0
10	7.1526E-02	1.0	7.8153E-02	1.0
11	7.1560E-02	1.0	7.8172E-02	1.0
12	7.1567E-02	1.0	7.8179E-02	1.0
13	7.1574E-02	1.0	7.8191E-02	1.0
14	7.1578E-02	1.0	7.8202E-02	1.0
15	7.1578E-02	1.0	7.8217E-02	1.0
16	7.1582E-02	1.0	7.8217E-02	1.0
17	7.1593E-02	1.0	7.8220E-02	1.0
18	7.1601E-02	1.0	7.8224E-02	1.0
19	7.1604E-02	1.0	7.8228E-02	1.0

№ изм.	Выборка 1		Выборка 2	
	X_{1n} , мВ	w_n	X_{2n} , мВ	w_n
20	7.1604E-02	1.0	7.8235E-02	1.0
21	7.1608E-02	1.0	7.8239E-02	1.0
22	7.1608E-02	1.0	7.8250E-02	1.0
23	7.1619E-02	1.0	7.8258E-02	1.0
24	7.1619E-02	1.0	7.8258E-02	1.0
25	7.1619E-02	1.0	7.8261E-02	1.0
26	7.1645E-02	1.0	7.8265E-02	1.0
27	7.1649E-02	1.0	7.8273E-02	1.0
28	7.1649E-02	1.0	7.8276E-02	1.0
29	7.1660E-02	1.0	7.8287E-02	1.0
30	7.1664E-02	1.0	7.8291E-02	1.0
31	7.1664E-02	1.0	7.8295E-02	1.0
32	7.1668E-02	1.0	7.8299E-02	1.0
33	7.1668E-02	1.0	7.8302E-02	1.0
34	7.1671E-02	1.0	7.8302E-02	1.0
35	7.1675E-02	1.0	7.8302E-02	1.0
36	7.1675E-02	1.0	7.8314E-02	1.0
37	7.1679E-02	1.0	7.8314E-02	1.0
38	7.1683E-02	1.0	7.8314E-02	1.0
39	7.1686E-02	1.0	7.8314E-02	1.0
40	7.1690E-02	1.0	7.8314E-02	1.0
41	7.1697E-02	1.0	7.8314E-02	1.0
42	7.1701E-02	1.0	7.8317E-02	1.0
43	7.1701E-02	1.0	7.8317E-02	1.0
44	7.1709E-02	1.0	7.8321E-02	1.0
45	7.1712E-02	1.0	7.8321E-02	1.0
46	7.1720E-02	1.0	7.8325E-02	1.0
47	7.1720E-02	1.0	7.8328E-02	1.0
48	7.1720E-02	1.0	7.8336E-02	1.0
49	7.1723E-02	1.0	7.8343E-02	1.0
50	7.1727E-02	1.0	7.8343E-02	1.0
51	7.1727E-02	1.0	7.8343E-02	1.0
52	7.1727E-02	1.0	7.8343E-02	1.0
53	7.1727E-02	1.0	7.8343E-02	1.0
54	7.1731E-02	1.0	7.8347E-02	1.0
55	7.1735E-02	1.0	7.8347E-02	1.0
56	7.1738E-02	1.0	7.8355E-02	1.0
57	7.1738E-02	1.0	7.8355E-02	1.0
58	7.1750E-02	1.0	7.8358E-02	1.0
59	7.1753E-02	1.0	7.8362E-02	1.0
60	7.1757E-02	1.0	7.8362E-02	1.0
61	7.1761E-02	1.0	7.8366E-02	1.0
62	7.1768E-02	1.0	7.8366E-02	1.0
63	7.1768E-02	1.0	7.8366E-02	1.0
64	7.1776E-02	1.0	7.8373E-02	1.0
65	7.1787E-02	1.0	7.8373E-02	1.0
66	7.1787E-02	1.0	7.8377E-02	1.0
67	7.1787E-02	1.0	7.8377E-02	1.0

№ изм.	Выборка 1		Выборка 2	
	X_{1n} , мВ	w_n	X_{2n} , мВ	w_n
68	7.1791E-02	1.0	7.8377E-02	1.0
69	7.1791E-02	1.0	7.8377E-02	1.0
70	7.1791E-02	1.0	7.8384E-02	1.0
71	7.1791E-02	1.0	7.8384E-02	1.0
72	7.1791E-02	1.0	7.8384E-02	1.0
73	7.1791E-02	1.0	7.8388E-02	1.0
74	7.1791E-02	1.0	7.8388E-02	1.0
75	7.1794E-02	1.0	7.8388E-02	1.0
76	7.1805E-02	1.0	7.8388E-02	1.0
77	7.1809E-02	1.0	7.8399E-02	1.0
78	7.1813E-02	1.0	7.8399E-02	1.0
79	7.1813E-02	1.0	7.8399E-02	1.0
80	7.1820E-02	1.0	7.8403E-02	1.0
81	7.1832E-02	1.0	7.8407E-02	1.0
82	7.1835E-02	1.0	7.8407E-02	1.0
83	7.1835E-02	1.0	7.8410E-02	1.0
84	7.1843E-02	1.0	7.8410E-02	1.0
85	7.1846E-02	1.0	7.8410E-02	1.0
86	7.1850E-02	1.0	7.8418E-02	1.0
87	7.1850E-02	1.0	7.8425E-02	1.0
88	7.1850E-02	1.0	7.8429E-02	1.0
89	7.1854E-02	1.0	7.8433E-02	1.0
90	7.1854E-02	1.0	7.8433E-02	1.0
91	7.1858E-02	1.0	7.8433E-02	1.0
92	7.1861E-02	1.0	7.8436E-02	1.0
93	7.1865E-02	1.0	7.8440E-02	1.0
94	7.1865E-02	1.0	7.8440E-02	1.0
95	7.1865E-02	1.0	7.8440E-02	1.0
96	7.1869E-02	1.0	7.8440E-02	1.0
97	7.1869E-02	1.0	7.8444E-02	1.0
98	7.1880E-02	1.0	7.8444E-02	1.0
99	7.1880E-02	1.0	7.8444E-02	1.0
100	7.1880E-02	1.0	7.8448E-02	1.0
101	7.1891E-02	1.0	7.8451E-02	1.0
102	7.1891E-02	1.0	7.8451E-02	1.0
103	7.1895E-02	1.0	7.8455E-02	1.0
104	7.1895E-02	1.0	7.8455E-02	1.0
105	7.1902E-02	1.0	7.8455E-02	1.0
106	7.1906E-02	1.0	7.8459E-02	1.0
107	7.1910E-02	1.0	7.8463E-02	1.0
108	7.1910E-02	1.0	7.8463E-02	1.0
109	7.1910E-02	1.0	7.8463E-02	1.0
110	7.1910E-02	1.0	7.8463E-02	1.0
111	7.1910E-02	1.0	7.8466E-02	1.0
112	7.1913E-02	1.0	7.8470E-02	1.0
113	7.1917E-02	1.0	7.8470E-02	1.0
114	7.1917E-02	1.0	7.8470E-02	1.0
115	7.1921E-02	1.0	7.8470E-02	1.0

№ изм.	Выборка 1		Выборка 2	
	X_{1n} , мВ	w_n	X_{2n} , мВ	w_n
116	7.1921E-02	1.0	7.8470E-02	1.0
117	7.1925E-02	1.0	7.8470E-02	1.0
118	7.1932E-02	1.0	7.8477E-02	1.0
119	7.1932E-02	1.0	7.8485E-02	1.0
120	7.1936E-02	1.0	7.8485E-02	1.0
121	7.1936E-02	1.0	7.8492E-02	1.0
122	7.1936E-02	1.0	7.8496E-02	1.0
123	7.1943E-02	1.0	7.8504E-02	1.0
124	7.1943E-02	1.0	7.8507E-02	1.0
125	7.1947E-02	1.0	7.8507E-02	1.0
126	7.1947E-02	1.0	7.8507E-02	1.0
127	7.1947E-02	1.0	7.8507E-02	1.0
128	7.1951E-02	1.0	7.8507E-02	1.0
129	7.1951E-02	1.0	7.8515E-02	1.0
130	7.1951E-02	1.0	7.8518E-02	1.0
131	7.1951E-02	1.0	7.8518E-02	1.0
132	7.1951E-02	1.0	7.8518E-02	1.0
133	7.1954E-02	1.0	7.8526E-02	1.0
134	7.1962E-02	1.0	7.8526E-02	1.0
135	7.1966E-02	1.0	7.8526E-02	1.0
136	7.1969E-02	1.0	7.8530E-02	1.0
137	7.1969E-02	1.0	7.8530E-02	1.0
138	7.1981E-02	1.0	7.8533E-02	1.0
139	7.1984E-02	1.0	7.8533E-02	1.0
140	7.1988E-02	1.0	7.8533E-02	1.0
141	7.1992E-02	1.0	7.8533E-02	1.0
142	7.1992E-02	1.0	7.8541E-02	1.0
143	7.1999E-02	1.0	7.8544E-02	1.0
144	7.1999E-02	1.0	7.8548E-02	1.0
145	7.2003E-02	1.0	7.8548E-02	1.0
146	7.2003E-02	1.0	7.8548E-02	1.0
147	7.2003E-02	1.0	7.8548E-02	1.0
148	7.2003E-02	1.0	7.8548E-02	1.0
149	7.2014E-02	1.0	7.8552E-02	1.0
150	7.2014E-02	1.0	7.8552E-02	1.0
151	7.2018E-02	1.0	7.8552E-02	1.0
152	7.2029E-02	1.0	7.8556E-02	1.0
153	7.2033E-02	1.0	7.8556E-02	1.0
154	7.2036E-02	1.0	7.8567E-02	1.0
155	7.2040E-02	1.0	7.8571E-02	1.0
156	7.2040E-02	1.0	7.8578E-02	1.0
157	7.2040E-02	1.0	7.8582E-02	1.0
158	7.2044E-02	1.0	7.8585E-02	1.0
159	7.2044E-02	1.0	7.8589E-02	1.0
160	7.2044E-02	1.0	7.8589E-02	1.0
161	7.2048E-02	1.0	7.8593E-02	1.0
162	7.2048E-02	1.0	7.8593E-02	1.0
163	7.2059E-02	1.0	7.8597E-02	1.0

№ изм.	Выборка 1		Выборка 2	
	X_{1n} , мВ	w_n	X_{2n} , мВ	w_n
164	7.2074E-02	1.0	7.8597E-02	1.0
165	7.2074E-02	1.0	7.8597E-02	1.0
166	7.2077E-02	1.0	7.8600E-02	1.0
167	7.2077E-02	1.0	7.8600E-02	1.0
168	7.2081E-02	1.0	7.8600E-02	1.0
169	7.2085E-02	1.0	7.8600E-02	1.0
170	7.2085E-02	1.0	7.8608E-02	1.0
171	7.2085E-02	1.0	7.8612E-02	1.0
172	7.2092E-02	1.0	7.8612E-02	1.0
173	7.2096E-02	1.0	7.8612E-02	1.0
174	7.2096E-02	1.0	7.8612E-02	1.0
175	7.2103E-02	1.0	7.8612E-02	1.0
176	7.2107E-02	1.0	7.8623E-02	1.0
177	7.2107E-02	1.0	7.8630E-02	1.0
178	7.2115E-02	1.0	7.8630E-02	1.0
179	7.2115E-02	1.0	7.8634E-02	1.0
180	7.2122E-02	1.0	7.8664E-02	1.0
181	7.2130E-02	1.0	7.8679E-02	1.0
182	7.2141E-02	1.0	7.8682E-02	1.0
183	7.2152E-02	1.0	7.8686E-02	1.0
184	7.2156E-02	1.0	7.8686E-02	1.0
185	7.2174E-02	1.0	7.8694E-02	1.0
186	7.2182E-02	1.0	7.8694E-02	1.0
187	7.2197E-02	1.0	7.8697E-02	1.0
188	7.2197E-02	1.0	7.8705E-02	1.0
189	7.2204E-02	1.0	7.8705E-02	1.0
190	7.2215E-02	1.0	7.8712E-02	1.0
191	7.2223E-02	1.0	7.8712E-02	1.0
192	7.2223E-02	1.0	7.8716E-02	1.0
193	7.2238E-02	1.0	7.8720E-02	1.0
194	7.2245E-02	1.0	7.8727E-02	1.0
195	7.2256E-02	1.0	7.8746E-02	1.0
196	7.2264E-02	1.01	7.8746E-02	1.0
197	7.2267E-02	1.01	7.8749E-02	1.0
198	7.2293E-02	1.24	7.8757E-02	1.0
199	7.2334E-02	1.62	7.8839E-02	1.65
200	7.2424E-02	2.49	7.8865E-02	1.89