

Федеральное государственное автономное образовательное
учреждение высшего образования «Национальный
исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Домашняя работа №2 по дисциплине
«Теория вероятностей»

Вариант № 3

Выполнил: Гаврилин Олег Сергеевич

Преподаватель: Селина Е. Г.

Группа: Р3230

Санкт-Петербург, 2024 г.

Задание 19.1

а) Располагаем значения результатов эксперимента в порядке возрастания, т.е. записываем вариационный ряд:

0.67 0.67 0.68 0.69 0.7 0.72 0.73 0.75 0.77 0.77

0.78 0.79 0.8 0.8 0.81 0.82 0.83 0.83 0.84 0.86

0.86 0.87 0.88 0.88 0.89 0.9 0.9 0.91 0.92 0.92

0.93 0.93 0.95 0.95 0.96 0.97 0.98 0.99 1.0 1.01

1.02 1.04 1.04 1.04 1.05 1.07 1.07 1.08 1.08 1.09

1.09 1.1 1.1 1.1 1.11 1.11 1.11 1.11 1.13 1.14

1.15 1.16 1.17 1.18 1.19 1.2 1.22 1.23 1.23 1.24

1.24 1.25 1.25 1.26 1.26 1.27 1.27 1.28 1.28 1.29

1.31 1.32 1.33 1.33 1.34 1.34 1.35 1.36 1.36 1.37

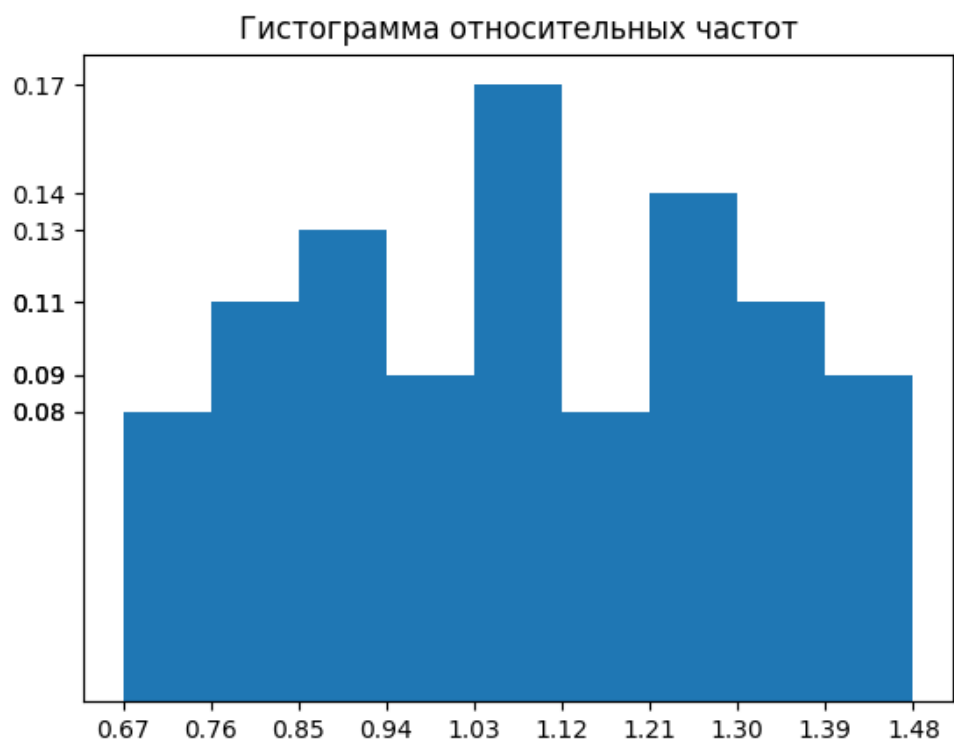
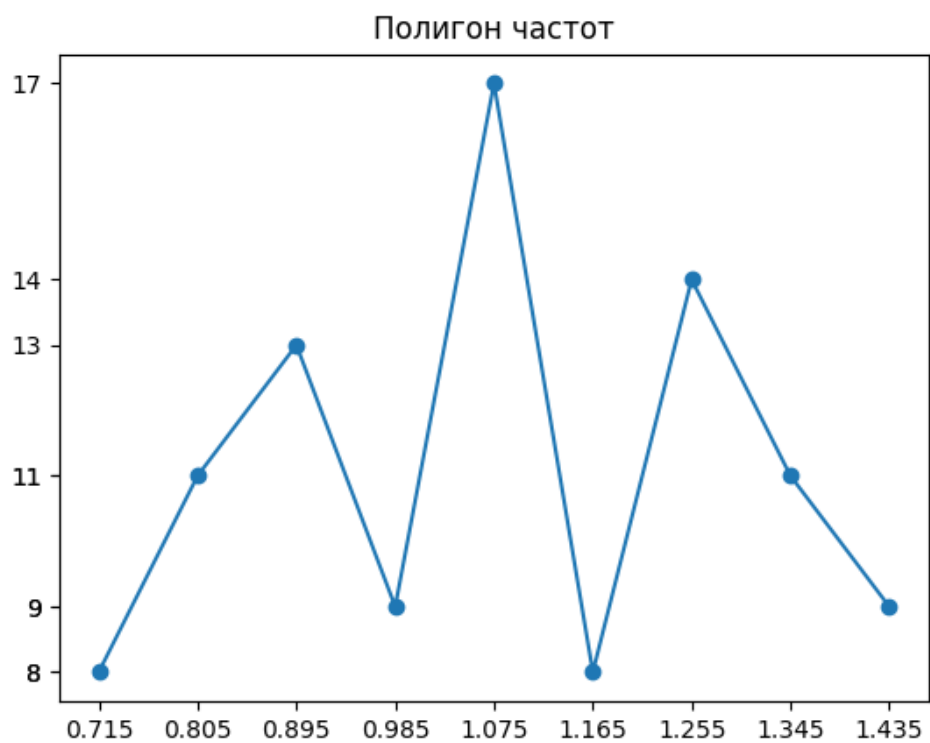
1.38 1.4 1.41 1.42 1.43 1.44 1.45 1.46 1.47 1.48

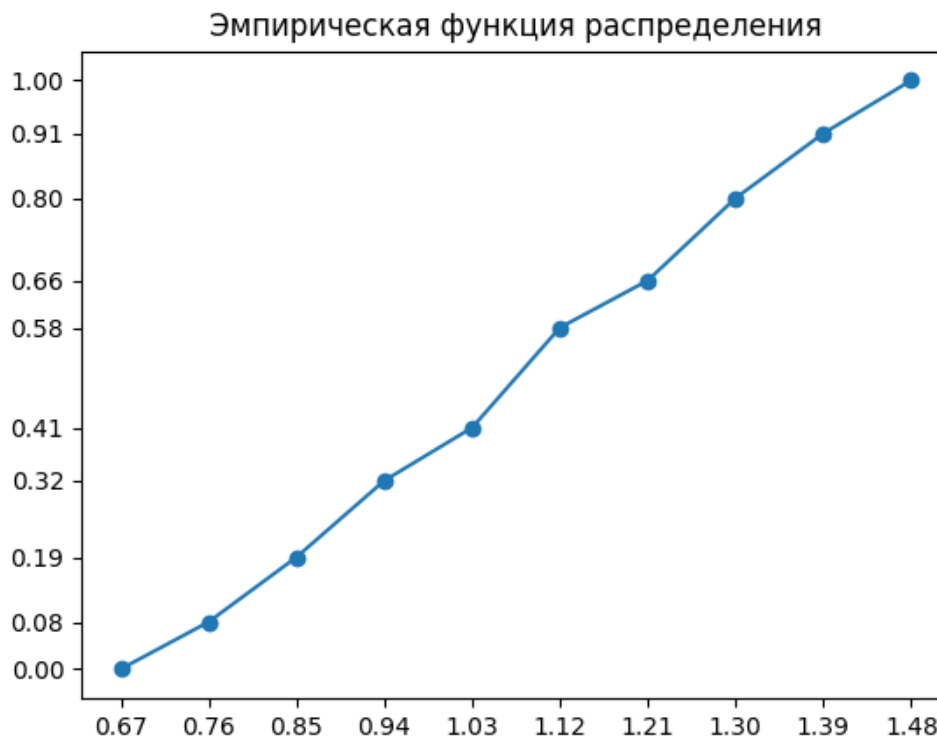
б) Находим размах варьирования: $\omega = x_{\max} - x_{\min} = 1.48 - 0.67 = 0.81$

Выборку разобьём на 9 интервалов. Длина частичного интервала: $h = \frac{\omega}{l} = \frac{0.81}{9} = 0.09$

Интервал	Середина интервала x_i	Частота f_i	Относительная частота $w_i = \frac{f_i}{n}$	Плотность относительной частоты $\frac{w_i}{h}$	w_i^{HAK}	$x_i f_i$	$x_i^2 f_i$
0.67 - 0.76	0.715	8	0.08	0.89	0.08	5.72	4.0898
0.76 - 0.85	0.805	11	0.11	1.22	0.19	8.855	7.128275
0.85 - 0.94	0.895	13	0.13	1.44	0.32	11.635	10.413325
0.94 - 1.03	0.985	9	0.09	1	0.41	8.865	8.732025
1.03 - 1.12	1.075	17	0.17	1.89	0.58	18.275	19.645625
1.12 - 1.21	1.165	8	0.08	0.89	0.66	9.32	10.8578
1.21 - 1.3	1.255	14	0.14	1.56	0.8	17.57	22.05035
1.3 - 1.39	1.345	11	0.11	1.22	0.91	14.795	19.899275
1.39 - 1.48	1.435	9	0.09	1	1.0	12.915	18.533025
	Σ	$n = 100$	1	-	-	107.95	121.3495

в) Строим полигон частот, гистограмму относительных частот и график эмпирической функции распределения.





г) Находим выборочное среднее и выборочную дисперсию:

$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{107.95}{100} = 1.0795 \approx 1.08$$

$$D = \frac{\sum x_i^2 f_i}{n} - \bar{x}^2 = \frac{121.3495}{100} - 1.0795^2 \approx 0.048$$

$$\sigma = \sqrt{D} = 0.22$$

Интервал	Середина интервала x_i	Частота f_i	$x_i f_i$	$x_i^2 f_i$
0.67 - 0.76	0.715	8	5.72	4.0898
0.76 - 0.85	0.805	11	8.855	7.128275
0.85 - 0.94	0.895	13	11.635	10.413325
0.94 - 1.03	0.985	9	8.865	8.732025
1.03 - 1.12	1.075	17	18.275	19.645625
1.12 - 1.21	1.165	8	9.32	10.8578
1.21 - 1.3	1.255	14	17.57	22.05035
1.3 - 1.39	1.345	11	14.795	19.899275
1.39 - 1.48	1.435	9	12.915	18.533025
	Σ	$n = 100$	107.95	121.3495

Выборочная дисперсия является *смещенной оценкой* генеральной дисперсии, а исправленная дисперсия – *несмещенной оценкой*:

$$\tilde{D}_B = \frac{n}{(n-1)} D_B = \frac{100}{99} * 0.048 = 0.049$$

$$\tilde{\sigma}_B = \sqrt{\tilde{D}_B} = 0.22$$

д) Согласно критерию Пирсона, необходимо сравнить эмпирические и теоретические частоты. Эмпирические частоты даны. Найдем теоретические частоты. Для этого пронумеруем X, т. е. перейдем к СВ $z = (x - \bar{x})/\sigma_B$ и вычислим концы интервалов z_i и z_{i+1} , причем наименьшее значение z , т.е. z_1 , положим стремящимся к $-\infty$, а наибольшее, т. е. z_{m+1} к $+\infty$. Результаты занесем в таблицу. Все $n_i > 5$, значит объединять интервалы не требуется.

i	Границы интервала $x_i; x_{i+1}$		$x_i - \bar{x}$	$x_{i+1} - \bar{x}$	Границы интервала $z_i; z_{i+1}$	
	x_i	x_{i+1}			$z_i = \frac{x_i - \bar{x}}{\sigma_B}$	$z_{i+1} = \frac{x_{i+1} - \bar{x}}{\sigma_B}$
1	0.67	0.76	-	-0,32	-	-1,45
2	0.76	0.85	-0,32	-0,23	-1,45	-1,05
3	0.85	0.94	-0,23	-0,14	-1,05	-0,64
4	0.94	1.03	-0,14	-0,05	-0,64	-0,23
5	1.03	1.12	-0,05	0,04	-0,23	0,18
6	1.12	1.21	0,04	0,13	0,18	0,59
7	1.21	1.3	0,13	0,22	0,59	1
8	1.3	1.39	0,22	0,31	1	1,41
9	1.39	1.48	0,31	-	1,41	-

Находим теоретические вероятности P_i и теоретические частоты $n'_i = nP_i = 100P_i$. Составляем расчетную таблицу.

i	Границы интервала $z_i; z_{i+1}$		$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i = \Phi(z_{i+1}) - \Phi(z_i)$	$n'_i = 100P_i$
	z_i	z_{i+1}				
1	-	-1,45	-0,5	-0,4265	0,0735	7,35
2	-1,45	-1,05	-0,4265	-0,3531	0,0734	7,34
3	-1,05	-0,64	-0,3531	-0,2389	0,1142	11,42
4	-0,64	-0,23	-0,2389	-0,091	0,1479	14,79
5	-0,23	0,18	-0,091	0,0714	0,1624	16,24
6	0,18	0,59	0,0714	0,2224	0,151	15,1
7	0,59	1	0,2224	0,3413	0,1189	11,89
8	1	1,41	0,3413	0,4207	0,0794	7,94
9	1,41	-	0,4207	0,5	0,0793	7,93
\sum_i	-	-	-	-	1	100

Вычислим наблюдаемое значение критерия Пирсона. Для этого составим расчетную таблицу. Последние два столбца служат для контроля вычисления по формуле:

$$\chi^2_{\text{набл}} = \frac{1}{n} \sum_{i=1}^k n_i^2 - n$$

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$(n_i - n'_i)^2 / n'_i$	n_i^2	n_i^2 / n'_i
1	8	7,35	0,65	0,4225	0,057482993	64	8,707482993
2	11	7,34	3,66	13,3956	1,825013624	121	16,48501362
3	13	11,42	1,58	2,4964	0,218598949	169	14,79859895
4	9	14,79	-5,79	33,5241	2,266673428	81	5,476673428

5	17	16,24	0,76	0,5776	0,035566502	289	17,7955665
6	8	15,1	-7,1	50,41	3,338410596	64	4,238410596
7	14	11,89	2,11	4,4521	0,374440706	196	16,48444071
8	11	7,94	3,06	9,3636	1,17929471	121	15,23929471
9	9	7,93	1,07	1,1449	0,144375788	81	10,21437579
\sum_i	100	100	-	-	$\chi^2_{\text{набл}} = 9,4398573$	-	109,4398573

Контроль: $\frac{\sum n_i^2}{n} - n = \frac{\sum (n_i - n'_i)^2}{n} = 109,4398573 - 100 = 9,4398573$

По таблице критических точек распределения χ^2 , уровню значимости $\alpha = 0,025$ и числу степеней свободы $k = l - 3 = 9 - 3 = 6$ находим: $\chi^2_{\text{кр}} = 14,4$

Так как $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, то гипотеза H_0 о нормальном распределении генеральной совокупности принимается.

е) Если СВ X генеральной совокупности распределена нормально, то с надежностью $\gamma = 0,9$ можно утверждать, что математическое ожидание α СВ X покрывается доверительным интервалом

$$\bar{x} - \frac{\tilde{\sigma}_B}{\sqrt{n}} t_\gamma < \tilde{x} < \bar{x} + \frac{\tilde{\sigma}_B}{\sqrt{n}} t_\gamma$$

Выберем уровень надежности $\gamma = 0.9$. По таблице нормального распределения находим $t_\gamma = 1.645$.

Вычислим точность оценки:

$$\frac{t_\gamma \sigma}{\sqrt{n}} \approx 0.0362$$

Таким образом доверительным интервалом для α будет:

$$1.0438 < \tilde{x} < 1.1162$$

Доверительный интервал, покрывающий среднее квадратичное отклонение σ с заданной надежностью γ ,

$$\tilde{\sigma}_B(1 - q) < \tilde{\sigma} < \tilde{\sigma}_B(1 + q)$$

При $\gamma = 0,9$ и $n = 100$ имеем: $q = 0,102$. Доверительным интервалом для σ будет

$$0.198 < \tilde{\sigma} < 0.242$$

Задание 19.2

Для подсчета числовых характеристик (выборочных средних \bar{x} и \bar{y} , выборочных средних квадратичных отклонений s_x и s_y и выборочного корреляционного момента s_{xy}) составляем расчетную таблицу. При заполнении таблицы осуществляем контроль по строкам и столбцам:

$$\sum_{i=1}^6 m_{x_i} = \sum_{j=1}^8 m_{y_j} = n = 100$$

$$\sum_{i=1}^6 \sum_{j=1}^8 m_{ij} x_i = \sum_{i=1}^6 m_{x_i} x_i = 124000$$

$$\sum_{i=1}^6 \sum_{j=1}^8 m_{ij} y_j = \sum_{j=1}^8 m_{y_j} y_j = 776500$$

$$\sum_{i=1}^6 \left(x_i \sum_{j=1}^8 m_{ij} y_j \right) = \sum_{j=1}^8 \left(y_j \sum_{i=1}^6 m_{ij} x_i \right) = 1119680000$$

Вычисляем выборочные средние \bar{x} и \bar{y} , $i = \overline{1, 6}$; $j = \overline{1, 8}$:

$$\bar{x} = \frac{\sum \sum m_{ij} x_i}{n} = \frac{\sum m_{x_i} x_i}{n} = \frac{124000}{100} = 1240$$

$$\bar{y} = \frac{\sum m_{y_j} y_j}{n} = \frac{776500}{100} = 7765$$

Выборочные дисперсии находим по формулам:

$$s_x^2 = \frac{1}{n-1} \left(\sum m_{x_i} x_i^2 - \frac{1}{n} \left(\sum m_{x_i} x_i \right)^2 \right) = \frac{1}{99} \left(204920000 - \frac{1}{100} (124000)^2 \right) = 516767,6767$$

$$s_y^2 = \frac{1}{n-1} \left(\sum m_{y_j} y_j^2 - \frac{1}{n} \left(\sum m_{y_j} y_j \right)^2 \right) = \frac{1}{99} \left(6713350000 - \frac{1}{100} (776500)^2 \right) = 6907348,4848$$

	j	1	2	3	4	5	6	7	8	9	10	11	12	13
i	$\begin{matrix} Y \\ X \end{matrix}$	1200	2700	4200	6700	8200	9700	11200	12700	m_x	$m_x x_i$	$\sum_{j=1}^k m_{ij} y_j$	$m_{x_i} x_i^2$	$x_i \sum_{j=1}^k m_{ij} y_j$
1	20	4	2	5						11	220	31200	4400	624000
2	520			7	5	2				14	7280	79300	3785600	41236000
3	1020				9	14	6			29	29580	233300	30171600	237966000
4	1520				7	8	6			21	31920	170700	48518400	259464000
5	2020					4	5	7		16	32320	159700	65286400	322594000
6	2520						3	2	4	9	22680	102300	57153600	257796000
7	m_y	4	2	12	21	28	20	9	4	100	124000	776500	204920000	1119680000
8	$m_y y_j$	4800	5400	50400	140700	229600	194000	100800	50800	776500				
9	$\sum_{i=1}^m m_{x_i} x_i$	80	40	3740	22420	35560	32900	19180	10080	124000				
10	$m_{ij} y_j^2$	5760000	14580000	21168000	94269000	18827200	18818000	11289600	64516000	67133500				
				0	0	00	00	00	0	00				
11	$y_j \sum_{i=1}^m m_{x_i} x_i$	96000	108000	15708000	15021400	29159200	31913000	21481600	12801600	11196800				
					0	0	0	0	0	00				

Гаврилин Олег Сергеевич, 408409, Р3230

Корреляционный момент вычисляем по формуле:

$$S_{xy} = \frac{1}{n-1} \left(\sum \sum m_{ij} x_i y_j - \frac{1}{n} (\sum m_{xi} x_i) (\sum m_{yj} y_j) \right) \\ = \frac{1}{99} \left(1119680000 - \frac{1}{100} (124000 * 776500) \right) \approx 1584040,404$$

Оценкой теоретической линии регрессии является эмпирическая линия регрессии, уравнение которой имеет вид

$$y = \bar{y} + r_{xy} \frac{S_y}{S_x} (x - \bar{x}),$$

где $S_x = \sqrt{516767,6767} \approx 718,866$; $S_y = \sqrt{6907348,4848} \approx 2628,183$;

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{1584040,404}{718,866 * 2628,183} = \frac{1584040,404}{1889311,400478} \approx 0,838422$$

Составляем уравнение эмпирической линии регрессии y на x:

$$y = 7765 + 0,838422 * \frac{2628,183}{718,866} (x - 1240)$$

$$y = 3,06529x + 3964,046$$

График изображен ниже:

