Проверка гипотезы о биномиальном распределении

Пусть проведено n опытов. Каждый опыт состоит из N независимых испытаний, в каждом из которых вероятность появления некоторого события A одна и та же. Регистрируется количество появлений события A в каждом опыте. В итоге получено статистическое распределение дискретной случайной величины X, которая характеризует количество появлений события A (в первой строке приведено количество появлений события A в одном опыте, а во второй - частота n_i, т. е. количество опытов, в которых зарегистрировано x_i появлений события A):

x_i	0	1	$\cdots$	N
n_i	n_0	n_1	$\cdots$	n_N

Необходимо с помощью критерия Пирсона проверить гипотезу о биномиальном законе распределения дискретной случайной величины X.

Для того чтобы при уровне значимости α проверить гипотезу о биномиальном распределении дискретной случайной величины X, необходимо:

1) найти по формуле Бернулли

$$\mathrm{P}_N(i) = C_N^i p^i (1-p)^{N-i}$$

вероятности $\mathrm{P}_N(i)$ появления события A ровно i раз в N испытаниях ($i = 0, 1, \ldots, s$, где s - максимальное количество наблюдаемых появлений события A в одном опыте, т. е. $s <= N$);

2) определить теоретические частоты:

$$n_i' = n \cdot \mathrm{P}_N(i)$$

где n - количество опытов

3) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, положив число степеней свободы $k = s - 1$ (считается, что вероятность p появления события A задана, т. е. она не оценивалась по выборке и не объединялись малочисленные частоты).

Если вероятность p была оценена по выборке, то $k = s - 2$. Если, кроме того, были объединены малочисленные частоты, то s - количество вариант выборки, которые остались после объединения частот.

Пример 1. Над событием *A*, вероятность появления которого равняется 0,3, проведено $n = 100$ независимых испытаний, каждое из которых состояло из $N = 7$ опытов. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о биномиальном распределении случайной величины *X* (количество появлений события *A*), если получена такая выборка:

x_i	0	1	2	3	4	5	6	7
n_i	4	23	31	23	11	5	2	1

Решение. Учитывая, что $p = 0{,}3; \qquad q = 1 - p = 1 - 0{,}3 = 0{,}7,$

по формуле Бернулли $\mathbf{P}_N(i) = C_N^i p^i q^{N-i}$ вычислим вероятности $\mathrm{P}_N(i)$:

$$\mathbf{P}_7(0) = C_7^0 0{,}3^0 0{,}7^7 \approx 0{,}0824; \qquad \mathbf{P}_7(1) = C_7^1 0{,}3^1 0{,}7^6 \approx 0{,}2471;$$
$$\mathbf{P}_7(2) = C_7^2 0{,}3^2 0{,}7^5 \approx 0{,}3177; \qquad \mathbf{P}_7(3) = C_7^3 0{,}3^3 0{,}7^4 \approx 0{,}2269;$$
$$\mathbf{P}_7(4) = C_7^4 0{,}3^4 0{,}7^3 \approx 0{,}0972; \qquad \mathbf{P}_7(5) = C_7^5 0{,}3^5 0{,}7^2 \approx 0{,}0250;$$
$$\mathbf{P}_7(6) = C_7^6 0{,}3^6 0{,}7^1 \approx 0{,}0036; \qquad \mathbf{P}_7(7) = C_7^7 0{,}3^7 0{,}7^0 \approx 0{,}0002.$$

Найдём теоретические частоты по формуле $n_{0i} = n \cdot \mathrm{P}_N(i)$:

$$n'_0 = 100 \cdot \mathbf{P}_7(0) = 8{,}24 \qquad n'_1 = 100 \cdot \mathbf{P}_7(1) = 24{,}71$$
$$n'_2 = 100 \cdot \mathbf{P}_7(2) = 31{,}77 \qquad n'_3 = 100 \cdot \mathbf{P}_7(3) = 22{,}69$$
$$n'_4 = 100 \cdot \mathbf{P}_7(4) = 9{,}72 \qquad n'_5 = 100 \cdot \mathbf{P}_7(5) = 2{,}5$$
$$n'_6 = 100 \cdot \mathbf{P}_7(6) = 0{,}36 \qquad n'_7 = 100 \cdot \mathbf{P}_7(7) = 0{,}02.$$

Поскольку частоты $n_0 = 4$, $n_6 = 2$ и $n_7 = 1$ малочисленные (меньше пяти), объединим их с другими частотами, а именно:
$n_1 = 23 + 4 = 27$; $n_5 = 5 + 2 + 1 = 8$.

В качестве теоретических частот, которые отвечают объединённым частотам, возьмём сумму соответствующих теоретических частот:

$$n'_1 = 24{,}71 + 8{,}24 = 32{,}95 \qquad n'_5 = 2{,}5 + 0{,}36 + 0{,}02 = 2{,}88.$$

Сравним эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого составим расчётную таблицу.

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$
1	27	32,95	5,95	35,4025	1,0744
2	31	31,77	0,77	0,5929	0,0187
3	23	22,69	–0,31	0,0961	0,0042
4	11	9,72	–1,28	1,6384	0,1686
5	8	2,88	–5,12	26,2144	9,1022
$\sum$	100				$\chi^2_{\text{набл}} = 10{,}3681$

Из расчётной таблицы получаем $\chi^2_{\text{набл}} = 10{,}3681.$

По таблице критических точек распределения χ^2 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = s - 1 = 5 - 1 = 4$ находим критическую точку правосторонней критической области:

$$\chi^2_{\text{кр}}(0{,}05; 4) = 9{,}49.$$

Поскольку $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$: гипотезу о биномиальном распределении отклоняем.

Задача 1. Над событием *A*, вероятность появления которого равняется 0,4, проведено n = 200 независимых испытаний, каждое из которых состояло из N = 10 опытов. Используя критерий Пирсона, при уровне значимости 0,05 проверить, выполняется ли гипотеза о биномиальном распределении случайной величины *X* (количество появлений события *A*), если получена такая выборка:

x_i	0	1	2	3	4	5	6	7	8	9
n_i	2	12	31	42	39	35	27	5	4	2

Указание. Следует объединить малочисленные частоты: n_0 - с n_1, а n_8, n_9 и n_{10} - с n_7.

Задача 2. Проверка гипотезы о биноминальном распределении с использованием Python

1. Импортируйте необходимые библиотеки (например, numpy и scipy).
2. Задайте значения параметров биноминального распределения: n – количество испытаний, p – вероятность успеха)
3. Сгенерируйте случайную выборку X размером N из биноминального распределения с заданными параметрами n и p.
4. Посчитайте эмпирические частоты для каждого значения успеха в выборке X.
5. Используя формулу Бернулли, вычислите теоретические частоты для каждого значения удачи.
6. Сравните теоретические и эмпирические частоты для каждого значения удачи.
7. Проведите статистический тест для проверки гипотезы о том, что выборка X имеет биноминальное распределение с заданными параметрами n и p.
8. Выведите результаты теста, включая статистическую статистику и p-значение.
9. Проанализируйте результаты теста и сделайте выводы о гипотезе о биноминальном распределении на основе значений.

Проверка гипотезы о распределении Пуассона

Пусть задано точечное статистическое распределение выборки.
Необходимо с помощью критерия Пирсона проверить гипотезу о распределении генеральной совокупности по закону Пуассона.

Для того чтобы при уровне значимости *α* проверить гипотезу о том, что исследуемая случайная величина *X* распределена по закону Пуассона, необходимо:

1) найти по заданному статистическому распределению выборочное среднее $x_в$

2) взять в качестве оценки параметра *λ* распределения Пуассона выборочное среднее $λ_? = x_в$

3) найти по формуле Пуассона (или в готовой таблице) вероятности P_i появления ровно *i* событий в *n* испытаниях (*i* = 1, 2, . . . , *r*, где *r* - максимальное количество наблюдаемых событий; *n* - объём выборки);

4) определить теоретические частоты

$$n_i' = n \cdot \mathbf{P}_i$$

5) сравнить эмпирические и теоретические частоты с помощью критерия Пирсона, взяв число степеней свободы $k = s - 2$, где *s* - количество вариант выборки (если проводилось объединение мало- численных частот в одну группу, то *s* - количество вариант, которые остались после объединения частот).

Пример 1. В $n = 1000$ проверках партий товара регистрировалось количество x_i некачественной продукции, вследствие чего было получено такое статистическое распределение количества x_i брака в n_i партиях товара:

x_i	0	1	2	3	4	5	6	7
n_i	242	349	234	107	43	21	3	1

Необходимо при уровне значимости 0,05 проверить гипотезу о том, что количество бракованной продукции X распределено по закону Пуассона.

Решение. Сначала найдём выборочное среднее:

$$\overline{x}_{\text{в}} = \frac{0\cdot 242 + 1\cdot 349 + 2\cdot 234 + 3\cdot 107 + 4\cdot 43 + 5\cdot 21 + 6\cdot 3 + 7\cdot 1}{1000} = 1{,}44.$$

Возьмем в качестве оценки параметра λ распределения Пуассона выборочное среднее:

$$\lambda^\star = \overline{x}_{\text{в}} = 1{,}44.$$

Предполагаемый закон Пуассона имеет вид

$$\mathbf{P}_{1000}(i) = 1{,}44^i \cdot \frac{e^{-1{,}44}}{i!}.$$

Положив $i = 0, 1, \dots, 7$, вычислим вероятности $P_i = P_{1000}(i)$:

$$\mathbf{P}_0 = 1{,}44^0 \cdot \frac{e^{-1{,}44}}{0!} \approx 0{,}2369; \qquad \mathbf{P}_1 = 1{,}44^1 \cdot \frac{e^{-1{,}44}}{1!} \approx 0{,}3412;$$

$$\mathbf{P}_2 = 1{,}44^2 \cdot \frac{e^{-1{,}44}}{2!} \approx 0{,}2456; \qquad \mathbf{P}_3 = 1{,}44^3 \cdot \frac{e^{-1{,}44}}{3!} \approx 0{,}1179;$$

$$\mathbf{P}_4 = 1{,}44^4 \cdot \frac{e^{-1{,}44}}{4!} \approx 0{,}0424; \qquad \mathbf{P}_5 = 1{,}44^5 \cdot \frac{e^{-1{,}44}}{5!} \approx 0{,}0122;$$

$$\mathbf{P}_6 = 1{,}44^6 \cdot \frac{e^{-1{,}44}}{6!} \approx 0{,}0029; \qquad \mathbf{P}_7 = 1{,}44^7 \cdot \frac{e^{-1{,}44}}{7!} \approx 0{,}0006.$$

Найдём теоретические частоты $n_{0i} = n \cdot P_i$:

$$n'_0 = 1000 \cdot \mathbf{P}_0 = 236{,}9; \qquad n'_1 = 1000 \cdot \mathbf{P}_1 = 341{,}2;$$

$$n'_2 = 1000 \cdot \mathbf{P}_2 = 245{,}6; \qquad n'_3 = 1000 \cdot \mathbf{P}_3 = 117{,}9;$$

$$n'_4 = 1000 \cdot \mathbf{P}_4 = 42{,}4; \qquad n'_5 = 1000 \cdot \mathbf{P}_5 = 12{,}2;$$

$$n'_6 = 1000 \cdot \mathbf{P}_6 = 2{,}9; \qquad n'_7 = 1000 \cdot \mathbf{P}_7 = 0{,}6.$$

Поскольку частоты $n_6 = 3$ и $n_7 = 1$ малочисленные (меньше пяти), объединим их с частотой n_5, а именно $n_5 = 21 + 3 + 1 = 25.$

В качестве теоретической частоты, которая отвечает объединённой частоте, возьмём сумму соответствующих теоретических частот:

$$n'_5 = 12{,}2 + 2{,}9 + 0{,}6 = 15{,}7.$$

Сравним эмпирические и теоретические частоты с помощью критерия Пирсона. Для этого составим расчётную таблицу

i	n_i	n'_i	$n_i - n'_i$	$(n_i - n'_i)^2$	$\dfrac{(n_i - n'_i)^2}{n'_i}$
0	242	236,9	–5,1	26,01	0,1098
1	349	341,2	–7,8	60,84	0,1783
2	234	245,6	11,6	134,56	0,5479
3	107	117,9	10,9	118,81	1,0077
4	43	42,4	–0,6	0,36	0,0085
5	25	15,7	–9,3	86,49	5,5089
$\sum$	1000				$\chi^2_{\text{набл}} = 7{,}3611$

Из расчётной таблицы получаем

$$\chi^2_{\text{набл}} = 7{,}3611$$

По таблице критических точек распределения χ_2 при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = s - 1 = 6 - 2 = 4$

находим критическую точку правосторонней критической области: $\chi^2_{\text{кр}}(0{,}05; 4) = 9{,}49.$

Поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$, нет оснований отклонять гипотезу о распределении случайной величины X (количество бракованного товара в партии) по закону Пуассона.

Задача 2. В $n = 1000$ проверках партий товара регистрировалось количество x_i некачественной продукции, вследствие чего было получено такое статистическое распределение количества x_i брака в n_i партиях товара:

x_i	0	1	2	3	4	5	6	7
n_i	427	363	154	41	9	3	2	1

Необходимо при уровне значимости 0,05 проверить гипотезу о том, что количество бракованной продукции X распределено по закону Пуассона.

Указание. Следует объединить малочисленные частоты n_6 и n_7 с n_5.