Математическая статистика

Домашняя работа № 4

Проверка статистических гипотез

Попов Юрий, СКБ-172

ОГЛАВЛЕНИЕ

Предисловие

Немного теории

Определение 1 *Статистическая гипотеза* - это некоторое предположение о виде или параметрах распределений

Определение 2 *Статистический критерий* - это правило, по которому каждой реализации выборки ставится в соответствие решение: принимаем гипотезу H_0 или отвергаем ее(то есть принимаем гипотезу H_1)

Определение 3 *Уровень значимости статистического теста* - допустимая для данной задачи вероятность ошибки первого рода, то есть вероятность отклонить нулевую гипотезу, когда на самом деле она верна

Определение 4 В случае, когда H_0 и H_1 - простые гипотезы, $P(x \in \mathcal{F}|H_0)) = \lambda$ - *ошибка 1 рода*
$\beta = P(H_0|H_1)$ - *ошибка второго рода*

Определение 5 Если H_0 состоит из одного определения, то говорят, что H_0 - *простая гипотеза*, иначе H_0 *сложная гипотеза*

Определение 6 Если H_1 состоит из одного определения, то говорят, что H_1 - *простая гипотеза*, иначе H_1 *сложная гипотеза*

Определение 7 *Функция мощности критерия* - это функционал на множестве допустимых распределений $\mathcal{F}$ и выборке X

Перейдем к практике

4.1 Проверка гипотезо виде распределений

Критерий согласия хи квадрат

Свойства данного критерия:

Положительные стороны:

Возможность применения как для дискретных, так и для непрерывных случайных величин, а так же отсутствие необходимости учитывать точные значения наблюдений (бывают случаи, когда исходные данные имеют не числовой характер), простота и наглядность.

Недостатки:

Недостатком этого критерия является тот факт, что в случае применения его для непрерывных распределений используется группировка наблюдений, приводящая к рассмотрению дискретной схемы, что приводит к некоторой потере информации, а также к необходимости рассмотрения вопроса о виде и числе интервалов.

Критерий согласия Колмогорова - Смирнова

Свойства данного критерия:

Положительные стороны:

Сильной стороной этого критерия является то, что, имея таблицы значений функции Колмогорова $K(t)$, мы можем рассчитывать критерий для проверки гипотезы относительно произвольной непрерывной функции распределения $F(x)$

Недостатки:

Недостатком является то, что он работает только для непрерывных распределений, по своей сути он является асимптотическим и , при его применении учитывается лишь наибольшее расхождение между эмпирической и теоретической функциями распределения, т.е. используется не вся информация. Оценка согласия по одной точке, особенно при небольшой длине выборки может плохо отражать соответствие эмпирических данных теоретическому закону распределения

4.1.1 Геометрическое распределение

Критерий согласия хи-квадрат для простой гипотезы

Проверим с помощью χ^2 - критерия Пирсона нулевую гипотезу$H_0 =$ (Наша выборка распределена по геометрическому закону), то есть $p_k = P(X = k) = p(1-p)^k, k = 0, 1, \ldots,$ при соответствующем уровне значимости

Для начала вычисляем выборочное среднее. Так как в этом случае у нас простоя гипотеза, то мы просто подставляем значение параметра. Я взял $p = 0.2$

Затем, я посчитал теоретические вероятности и теоретические частоты по соответствующим формулам:

$$p_k = 0.2(1-0.2)^k$$

и

$$n'_k = np_k$$

Затем я рассчитал меру отклонения Пирсона для каждого значения и составил итоговую таблицу:

```
Параметр равен:  0.2
```

	k	n_k	p_k	n_k_2	Мера
0	0	18	0.200	20.0	0.200
1	1	19	0.160	16.0	0.562
2	2	19	0.128	12.8	3.003
3	3	10	0.102	10.2	0.004
4	4	10	0.082	8.2	0.395
5	5	3	0.066	6.6	1.964
6	6	6	0.052	5.2	0.123
7	7	1	0.042	4.2	2.438
8	8	1	0.034	3.4	1.694
9	9	4	0.027	2.7	0.626
10	10	4	0.021	2.1	1.719
11	11	1	0.017	1.7	0.288
12	12	1	0.014	1.4	0.114
13	13	1	0.011	1.1	0.009
14	15	1	0.007	0.7	0.129
15	20	1	0.002	0.2	3.200
Сумма	126	100	0.965	96.5	0.127

n = 100

Из расчетной таблицы видно значение критерия Пирсона χ^2(правый нижний угол) $=$ 0.127.

Уровень значимости равен 0.1

Критическая точка для уровня значимости 0.1 при количестве степеней свободы $k = 15(16-1)$ равна $22,30712$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по геометрическому закону с параметром $p = 0.2$.

Уровень значимости равен 0.05

Критическая точка для уровня значимости 0.05 при количестве степеней свободы $k = 15(16-1)$ равна $24,99579014$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по геометрическому закону с параметром $p = 0.2$.

Критерий согласия хи-квадрат для сложной гипотезы

Проверим с помощью χ^2 - критерия Пирсона нулевую гипотезу$H_0 =$ (Наша выборка распределена по геометрическому закону), то есть $p_k = P(X = k) = p(1-p)^k, k = 0, 1, \ldots,$ при соответствующем уровне значимости

Для начала вычисляем выборочное среднее. В этом случае у нас сложная гипотеза, поэтому найдем оценку нашего параметра. В моем случае это $p = \dfrac{1}{\overline{x}}$. Оно получилась равна 0.104

Затем, я посчитал теоретические вероятности и теоретические частоты по соответствующим формулам:

$$p_k = 0.104(1 - 0.104)^k$$

и

$$n'_k = np_k$$

Затем я рассчитал меру отклонения Пирсона для каждого значения и составил итоговую таблицу:

	k	n_k	p_k	n_k_2	Мера
0	0	14	0.109	10.9	0.882
1	1	16	0.097	9.7	4.092
2	2	13	0.087	8.7	2.125
3	3	11	0.077	7.7	1.414
4	4	10	0.069	6.9	1.393
5	5	11	0.061	6.1	3.936
6	6	2	0.055	5.5	2.227
7	7	4	0.049	4.9	0.165
8	8	2	0.043	4.3	1.230
9	9	4	0.039	3.9	0.003
10	10	1	0.034	3.4	1.694
11	11	2	0.031	3.1	0.390
12	12	2	0.027	2.7	0.181
13	13	4	0.024	2.4	1.067
14	15	1	0.019	1.9	0.426
15	17	1	0.015	1.5	0.167
16	18	1	0.014	1.4	0.114
17	24	1	0.007	0.7	0.129
Сумма	165	100	0.857	85.7	2.386

n = 100

Из расчетной таблицы видно значение критерия Пирсона χ^2(правый нижний угол) = 2.386.

Уровень значимости равен 0.1

Критическая точка для уровня значимости 0.1 при количестве степеней свободы $k = 16(18-1-1)$ равна $23,54182892$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по геометрическому закону с оценкой параметра, равной $p = 0.104$

Уровень значимости равен 0.05

Критическая точка для уровня значимости 0.05 при количестве степеней свободы $k = 16(18 - 1 - 1)$ равна $26,2962276$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по геометрическому закону с оценкой параметра, равной $p = 0.104$.

4.1.2 Экспоненциальное распределение

Критерий согласия хи-квадрат для простой гипотезы

Проверим с помощью χ^2 - критерия Пирсона нулевую гипотезу$H_0 =$ (Наша выборка распределена по экспоненциальному закону), то есть $p_k = P(X = k) = \lambda e^{-\lambda k}, k = 0, 1, \ldots$, при соответствующем уровне значимости

Мне было необходимо разбить мой интервал на промежутки, и посчитать сколько значений моей выборки попадает в каждый из промежутков. Я решил разбивать на интервалы так, чтобы в каждый из них попадало одинаковое количество чисел из моей выборки.

Соответственно, я разбил на интервалы и посчитал количество вхождений.

Затем вычислил выборочное среднее и, так как в этом случае у нас простая гипотеза, то взял значение параметра $\lambda = 0.8$

Затем, я посчитал теоритические вероятности попадания в интервалы и теоретические частоты по соответствующим формулам:

$$P_i = P(x_i < X < x_{i+1}) = e^{-\lambda x_i} - e^{-\lambda x_{i+1}} = e^{-0.8 x_i} - e^{-0.8 x_{i+1}}$$

и

$$n'_k = np_k$$

Затем я рассчитал меру отклонения Пирсона для каждого значения и составил итоговую таблицу:

```
Параметер равен:  0.8
```

	x_k	x_k+1	n_k	P_k	n_k_2	Мера
0	0.006	0.047	5.0	0.032	3.2	1.012
1	0.047	0.157	5.0	0.081	8.1	1.186
2	0.157	0.219	5.0	0.043	4.3	0.114
3	0.219	0.302	5.0	0.054	5.4	0.030
4	0.302	0.406	5.0	0.063	6.3	0.268
5	0.406	0.497	5.0	0.051	5.1	0.002
6	0.497	0.668	5.0	0.086	8.6	1.507
7	0.668	0.704	5.0	0.017	1.7	6.406
8	0.704	0.850	5.0	0.063	6.3	0.268
9	0.850	0.993	5.0	0.055	5.5	0.045
10	0.993	1.203	5.0	0.070	7.0	0.571
11	1.203	1.386	5.0	0.052	5.2	0.008
12	1.386	1.563	5.0	0.044	4.4	0.082
13	1.563	1.661	5.0	0.022	2.2	3.564
14	1.661	1.981	5.0	0.060	6.0	0.167
15	1.981	2.146	5.0	0.025	2.5	2.500
16	2.146	2.442	5.0	0.038	3.8	0.379
17	2.442	2.617	5.0	0.019	1.9	5.058
18	2.617	3.355	5.0	0.055	5.5	0.045
19	3.355	5.274	5.0	0.054	5.4	0.030
Сумма	23.203	28.471	100.0	0.984	98.4	23.242

n = 100

Из расчетной таблицы видно значение критерия Пирсона χ^2(правый нижний угол) = 23.242.

Уровень значимости равен 0.1

Критическая точка для уровня значимости 0.1 при количестве степеней свободы $k = 19(20-1)$ равна $27,20357103$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по экспоненциальному закону с параметром $\lambda = 0.8$.

Уровень значимости равен 0.05

Критическая точка для уровня значимости 0.05 при количестве степеней свободы $k = 19(20 - 1)$ равна $30,14352721$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по экспоненциальному закону с параметром $\lambda = 0.8$.

Критерий согласия хи-квадрат для сложной гипотезы

Проверим с помощью χ^2 - критерия Пирсона нулевую гипотезу$H_0 =$ (Наша выборка распределена по экспоненциальному закону), то есть $p_k = P(X = k) = \lambda e^{-\lambda k}, k = 0, 1, \ldots$, при соответствующем уровне значимости

Мне было необходимо разбить мой интервал на промежутки, и посчитать сколько значений моей выборки попадает в каждый из промежутков. Я решил разбивать на интервалы так, чтобы в каждый из них попадало одинаковое количество чисел из моей выборки.

Соответственно, я разбил на интервалы и посчитал количество вхождений.

Затем вычислил выборочное среднее и нашел оценку нашего параметра. В моем случае это $\lambda = \dfrac{1}{\overline{x}}$. Оно получилась равна 0.732

Затем, я посчитал теоритические вероятности попадания в интервалы и теоретические частоты по соответствующим формулам:

$$P_i = P(x_i < X < x_{i+1}) = e^{-\lambda x_i} - e^{-\lambda x_{i+1}} = e^{-0.732 x_i} - e^{-0.732 x_{i+1}}$$

и

$$n'_k = np_k$$

Затем я рассчитал меру отклонения Пирсона для каждого значения и составил итоговую таблицу:

```
Оценка параметра равна:  0.732
```

	x_k	x_k+1	n_k	P_k	n_k_2	Мера
0	0.010	0.139	5.0	0.089	8.9	1.709
1	0.139	0.192	5.0	0.034	3.4	0.753
2	0.192	0.391	5.0	0.118	11.8	3.919
3	0.391	0.515	5.0	0.065	6.5	0.346
4	0.515	0.575	5.0	0.029	2.9	1.521
5	0.575	0.692	5.0	0.054	5.4	0.030
6	0.692	0.794	5.0	0.043	4.3	0.114
7	0.794	0.882	5.0	0.035	3.5	0.643
8	0.882	0.983	5.0	0.037	3.7	0.457
9	0.983	1.133	5.0	0.051	5.1	0.002
10	1.133	1.321	5.0	0.056	5.6	0.064
11	1.321	1.464	5.0	0.038	3.8	0.379
12	1.464	1.573	5.0	0.026	2.6	2.215
13	1.573	1.702	5.0	0.028	2.8	1.729
14	1.702	1.805	5.0	0.021	2.1	4.005
15	1.805	2.024	5.0	0.040	4.0	0.250
16	2.024	2.308	5.0	0.043	4.3	0.114
17	2.308	2.710	5.0	0.047	4.7	0.019
18	2.710	3.402	5.0	0.055	5.5	0.045
19	3.402	4.898	5.0	0.055	5.5	0.045
Сумма	24.615	29.503	100.0	0.964	96.4	18.359

n = 100

Из расчетной таблицы видно значение критерия Пирсона χ^2(правый нижний угол) = 18.359.

Уровень значимости равен 0.1

Критическая точка для уровня значимости 0.1 при количестве степеней свободы $k = 18(20 - 1 - 1)$ равна $25,98942308$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по экспоненциальному закону с оценкой параметра, равной $\lambda = 0.732$.

Уровень значимости равен 0.05

Критическая точка для уровня значимости 0.05 при количестве степеней свободы $k = 18(20-1-1)$ равна $28,86929943$

Так как наблюдаемое значение критерия меньше меньше критического, то нет основания отвергнуть нулевую гипотезу о распределении нашей выборки по экспоненциальному закону с оценкой параметра, равной $\lambda = 0.732$.

Критерий согласия Колмогорова - Смирнова для простой гипотезы

Пусть дана выборка $X = (X_1 \dots X_n)$ из распределения $\mathscr{L}(\xi)$ и F_ξ - неизвестное распределение.

- $H_0 : F_\xi = F(x)$ - простая гипотеза
- H_1 : не $F(x)$

Критерий Колмогорова основан на теореме Колмогорова:

$$D_n = D_n(x) = \sup_{x \in R} |\hat{F}_n(x) - F(x)|$$

где D_n - это отклонение эмпирической функции распределения от теоретической функции распределения.
$\hat{F}_n$ - оптимальная несмещенная состоятельная оценка для $F(x)$

Критерий согласия Колмогорова - Смирнова для сложной гипотезы

4.2 Задание для рассматриваемых распределений

4.2.1 Геометрическое распределение

4.2.2 Экспоненциальное распределение

Литература

[1]

[2] ссылка1

[3] ссылка2

[4] // ссылка3

[5] // ссылка4