Проверка статистических гипотез. Проверка гипотезы о виде распределения

1. Общие сведения
2. Критерий Колмогорова
3. Критерий χ^2 (Пирсона)

Пусть X - данные, $\{x\}$ - выборочное пространство, $\mathcal{F} = \{F\}$ - совокупность априори допустимых распределений X. F_X - неизвестное истинное распределение данных X, $F_X \in \{F\}$. В общем случае задача проверики гипотез ставится так: выделяется некоторое подмножество $\mathcal{F}_0 \subset \mathcal{F}$ допустимых распределений и требуется по данным X проверить, справедливо ли утверждение $H_0 : F_X \in \mathcal{F}_0$ или же оно ложно. В этом случае H_0 называется оносновной (нулевой) гипотезой. Распределения $F \in \mathcal{F}_1 = \mathcal{F} \backslash \mathcal{F}_0$ называеются альтернативными, а утверждение $H_1 : F_X \in \mathcal{F}_1$ - альтернативной гипотезой. В этом случае речь идет о проверке гипотезы H_0 против альтернативы H_1 или о задаче (H_0, H_1); иногда так же говорят, что гипотеза H_0 проверяется внутри обзей гипотезы $H : F_X \in \mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_0$. Если подмножество $\mathcal{F}_0(\mathcal{F}_1)$ состоит из одного элемента, то гипотеза H_0 (альтернатива H_1) называется простой, в противном случае сложной.

Правило, согласно которому мы, наблюдая X, принимаем решение принять гипотезу H_0 как истинную либо отклонить ее как ложную (т.е. принять альтернативную гипотезу H_1) называется **статистическим критерием** (или просто критерием).

В некоторых случаях класс $\mathcal{F}$ - это все распределения на выборочном пространстве $\{x\}$, тогда альтернативная гипотеза $H_1 = \bar{H}_0$ никак не конкретизируется, и здесь речь идет просто о согласии данных X с нулевой гипотезой H_0 (согласуются ли данные X с гипотезой H_0 или же они ее опровергают) - соответствующие критерии называются критериями согласия.
Поскольку критерий - это правило, которое для каждой реализации x выборки X должно приводить к одному из двух решений: принять гипотезу H_0 или отклонить ее (принять альтернативу H_1), то каждому критерию соответствует некоторое разбиение выборочного пространства $\mathcal{X} = \{x\}$ на два взаимно дополнительных множества $\mathcal{X}_0 \cap \mathcal{X}_1 = \emptyset, \mathcal{X}_0 \cup \mathcal{X}_1 = \mathcal{X}$. Где $\mathcal{X}_0$ состоит из тех точек x, для которых H_0 принимается, а $\mathcal{X}_1$ - из тех, для которых H_0 отвергается (принимается H_1). Таким образом, $\mathcal{X}_0$ - это область принятия гипотезы H_0, а $\mathcal{X}_1$ - область ее отклонения, которую принято называть **критической областью**.
Тем самым, любой критерий проверки гипотезы H_0 однозначно задается соответствующей критической областью $\mathcal{X}_1$ и о таком критерии часто говорят как о "критерии $\mathcal{X}_1$".

Итак, критерий $\mathcal{X}_1$ имеет вид:

$$H_0 \text{ отвергается} \Leftrightarrow X \in \mathcal{X}_1$$

Если в эксперименте наблюдается маловероятное при справедливости гипотезы H_0 событие, то считается, что гипотеза H_0 не согласуется с данными (или противоречит им), и в данном случае она отвергается (отлоняется); в противном случае считается, что данные не противоречат H_0 (или согласуется с ней), и H_0 принимается. В соответсвии с этим приницпом критическая область $\mathcal{X}_1$ должна быть выбрана так, чтобы была мала вероятность $P\{X \in \mathcal{X}_1 | H_0\}$, т.е. условная (при условии, что H_0 спрведлива) вероятность попадания значения выборки X в область $\mathcal{X}_1$. Поэтому при построении критерия задаются заранее некоторым малым числов α (например: $\alpha = 0.001; 0.01; 0.05 \text{ и т. д.}$) и налагают условие:
(1)

$$P\{X \in \mathcal{X}_1 | H_0\} \leq \alpha$$

Если выполнено (1), то говорят, что критерий $\mathcal{X}_1$ имеет уровень значимости α и подчеркивают это обозначением $\mathcal{X}_1 = \mathcal{X}_{1\alpha}$.
Ясно, что условием (1) критическая область $\mathcal{X}_{1\alpha}$ определяется неоднозначно, и чтобы утсранить эту неопределенность, надо ввести дополнительное понятие ошибок критерия.
Ошибка 1-го рода - отвергуть H_0 при условии, что она верна. Например, в биомтерических технологиях она называется FRR (false rejection rate) и означает, что человек не распознан системой.
Ошибка 2-го рода - принять H_0 при условии, что она неверна.
Например, в биометрических технологиях она называется FAR (false acceptance rate) и означает, что один человек принят системой за другого.

Введем теперь фундаментальное понятие теории проверки статистических гипотез - понятие фукнции мощности критерия. По определению, функцией мощности критерия $\mathcal{X}_{1\alpha}$ называется следующий функционал на множестве всех допустимых распределний $\mathcal{F} = \{F\}$ выборки X:
(2)

$$W(F) = W(F; \mathcal{X}_{1\alpha}) = P\{X \in \mathcal{X}_{1\alpha} | F\}, F \in \mathcal{F}$$

ДРугими словами, $W(F; \mathcal{X}_{1\alpha})$ - это вероятность попадания значения выборки X в критическую область $\mathcal{X}_{1\alpha}$, когда F - ее истинное распределение. Через функцию мощности легко выразить вероятности обоих типов ошибок, свойственных критерию $\mathcal{X}_{1\alpha}$. Именно, вероятность ошибки 1-го рода есть $W(F)$ при $F \in \mathcal{F}_0$, а второго рода $1 - W(F)$ при $F \in \mathcal{F}_1$.

$$P\{H_1 | H_0\} \ - \text{вероятность ошибки } 1 - \text{го рода}$$

$$P\{H_0 | H_1\} \ - \text{вероятность ошибки } 2 - \text{го рода}$$

Несмещенность критерия - одновременное выполнение условий:

$$W(F; \mathcal{X}_{1\alpha}) \leq \alpha, \forall F \in \mathcal{F}_0, W(F; \mathcal{X}_{1\alpha}) > \alpha, \forall F \in \mathcal{F}_1$$

Состоятельность критерия -

$$W_n(F; \mathcal{X}_{1\alpha}) \to 1, \forall F \in \mathcal{F}_1$$

Критерий Колмогорова

Этот критерий примеяют в тех случаях, когда функция $F(x)$ непрерывна. Статистика критерия определяется формулой:
(1)

$$D_n = D_n(X) = \sup_{-\infty < x < +\infty} |\hat{F}_n(x) - F(x)|$$

Где $\hat{F}(x)$ - эмпирическая функция распределения

И по теореме Колмогорова:

$$\forall t > 0 : \lim_{n \to \infty} P(\sqrt{n} D_n \leq t) = K(t) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2 t^2}$$

$$\mathcal{T}_{1\alpha} = \{t \geq t_\alpha, t_\alpha = \lambda_\alpha / \sqrt{n}\}$$

$$P\{D_n \in \mathcal{T}_{1\alpha} | H_0\} = P \sqrt{n} D_n \geq \lambda_\alpha | H_0 \approx 1 - K(\lambda_\alpha) = \alpha$$

Тем самым критерий согласия Колмогорова формулируется следюущим образом: если n довольно большое и при выбранном уровне значимости α число λ_α определено соотношением $K(\lambda_\alpha) = 1 - \alpha$, то:
(3)

$$H_0 \text{ отвергается} \Leftrightarrow \sqrt{n} D_n \geq \lambda_\alpha$$

Наконец, отметим, что для практических вычислений статистики D_n полезна эквивалентная (1) формула $D_n = max(D_n^+, D_n^-)$
, где :

$$D_n^+ = \max_{1 \leq k \leq n} (\frac{k}{n} - F(X_{(k)})), \ D_n^- = \max_{1 \leq k \leq n} (F(X_{(k)}) - \frac{k-1}{n})$$

и $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ - вариационный ряд выборки X

Критерий χ^2 (Пирсона)

Итак, пусть в эксперименте наблюдается дискретная случайная величина ξ, принимающая значения $1, 2, \ldots N$ с некоторыми вероятностями $p_1, \ldots, p_N$, $(p_1 + \ldots + p_N) = 1$. Если произведено n независимых испытаний над ξ, т.е. имеется выборка $(\xi_1, \ldots, \xi_n)$, то пусть:

$$\nu_j = \sum_{i=1}^{n} I(\xi_i = j), j = 1, \ldots, N$$

-соответствующие частоты исходов $(\nu_1 + \ldots + \nu_N = n)$. Тогда вектор $\nu = (\nu_1, \ldots, \nu_N)$ имеет полиномиальное распределение $M(n; p_1, \ldots, p_N)$.

Итак, пусть по наблюдению вектора частот $\nu = (\nu_1, \ldots, \nu_N)$ требуется проверить простую гипотезу
$H_0 : p = p^0, \text{где} p^0 = (p_1^0, \ldots, p_N^0))$ - заданный вероятностный вектор $(0 < p_j^0 < 1, j = 1, \ldots, N, p_1^0 + \ldots + p_N^0 = 1)$.
К.Пирсон в 1900г. предложил использовать в качестве меры отклонения эмпирических данных (относительных частот ν/n) от гипотетических значений p^0 меру хи-квадрат. Мера χ^2:
(4)

$$\mathring{X}_n^2 = \mathring{X}_n^2(\nu) = \sum_{j=1}^{N} \frac{(\nu_j - n p_j^0)^2}{n p_j^0} = \sum_{j=1}^{N} \frac{\nu_j^2}{n p_j^0} - n$$

На этой тестовой статитстике и основывается знаменитый критерий χ^2 К.Пирсона. В основе этого критерия лежат следующие очевидные соображения. Если гипотеза H_0 справедлива, то, поскольку относительная частота ν_j / n события $\{\xi = j\}$ является состоятельной оценкой его вероятности $p_j^0 (j = 1, \ldots, n)$, при больших n разности $|\nu_j / n - p_j^0|$ должны быть малы, следовательно, и значение статистики $\mathring{X}_n^2$ не должно быть слишком большим. Поэтому естественно задать критическую облать для гипотезы H_0 в виде $\mathcal{T}_{1\alpha} = \{\mathring{X}_n^2 > t_\alpha\}$. Где критическая граница t_α при заданном уровне значимости α должна быть выбрана из условия:
(5)

$$P\{\mathring{X}_n^2 > t_\alpha | H_0\} = \alpha$$

Теорема 1. Если $0 < p_j^0 < 1, j = 1, \ldots, N$, то при $n \to \infty$:

$$\mathcal{L}(\mathring{X}_n^2 | H_0) \to \chi^2(N-1)$$

$$H_0 \text{ отвергается} \Leftrightarrow \{\mathring{X}_n^2 > \chi^2_{1-\alpha, N-1}\}$$

Решение задач

3.1, 3.3, 3.5

№ 3.1

Для данных задачи 1.13 проверить, согласуются ли они с гипотезой H_0 о том, что монета была симметричной. Уровень значимости положите равным $\alpha_1 = 0.05, \alpha_2 = 0.1$.

$$\text{Эксперимент}: \ n = 4040, h = 2048, h - \text{выпадение герба}$$

In [50]:
```
import numpy as np
# f_obs - Observed frequencies in each category.
# f_exp - Expected frequencies in each category.
h = 2048
n = 4040
nu = np.array([h, (n-h)])
p = np.array([0.5, 0.5])
X = np.sum(nu**2 / (n*p)) - n
# degree of freedom is 1 (N = 2)
# for alpha_1 = 0.05 chi^2 is 3.841
# for alpha_2 = 0.1 chi^2 is 0.0158
print(f'X is {X} and chi2 for alpha1 = 0.05 is {3.841} -> H0 is true')
print(f'X is {X} and chi2 for alpha2 = 0.1 is {2.706} -> H0 is true')
```

```
X is 0.7762376237624267 and chi2 for alpha1 = 0.05 is 3.841 -> H0 is true
X is 0.7762376237624267 and chi2 for alpha2 = 0.1 is 2.706 -> H0 is true
```

№ 3.3

При $n = 4000$ независимых испытаний события A_1, A_2, A_3, составляющие полную группу, осуществились соответственно $1905, 1015, 1080$ раз. Проверьте, согласуются ли эти данные при уровне значимости $\alpha = 0.05$ с гипотезой $H_0 = (0.5, 0.25, 0.25)$

In [51]:
```
n = 4000
nu = np.array([1905, 1015, 1080])
p = np.array([0.5, 0.25, 0.25])
X = np.sum(nu**2 / (n*p)) - n
# degree of freedom is 2 (N = 3)
# for alpha = 0.05 chi^2 is 6.0
print(f'X is {X} and chi2 for alpha = 0.05 is {6.0} -> H0 is false')
```

```
X is 11.137500000000273 and chi2 for alpha = 0.05 is 6.0 -> H0 is false
```

№3.5

Согласуются ли данные, приведенные в задачах 1.16 и 1.17, с гипотезой о симметричности костей?

Проводились опыты с бросанием одновременно 12 игральных костей. Наблюдаемую случайную величину ξ считали равной числу костей, на которых выпало 4,5 или 6 очков. Пусть h_i - число опытов, в которых наблюдалось значение $\xi = i, i = 0, 1, \ldots, 12$. Данные для $n = 4096$ опытов приведены в следующей таблице:

$$\begin{pmatrix} i: & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\ h_i: & 0 & 7 & 60 & 198 & 430 & 731 & 948 & 847 & 536 & 257 & 71 & 11 & 0 \end{pmatrix}$$

Если верна H_0 - кости симметричны, то вероятность того, что выпало 4, 5 или 6 очков равна вероятности что выпало 1, 2 или 3 очка и равна 0.5. Тогда распределение $\xi \sim Bi(12, 0.5)$. Тогда:

$$p = (0.0002, 0.0029, 0.0161, 0.0537, 0.1208, 0.1934, 0.2256, 0.1934, 0.1208, 0.0537, 0.0161, 0.0029, 0.0002)$$

In [49]:
```
n = 4096
nu = np.array([0, 7, 60, 198, 430, 731, 948, 847, 536, 257, 71, 11, 0])
p = np.array([0.0002, 0.0029, 0.0161, 0.0537, 0.1208, 0.1934, 0.2258,
     0.1934, 0.1208, 0.0537, 0.0161, 0.0029, 0.0002])
X = np.sum(nu**2 / (n*p)) - n
# degree of freedom is 12 (N = 13)
# for alpha = 0.01 is 26.2
print(f'X is {X} and chi2 for alpha = 0.01 is {26.2} -> H0 is false')
```

```
X is 34.07416568269946 and chi2 for alpha = 0.01 is 26.2 -> H0 is false
```