

## Parameter estimation

### From parameter estimation to hypothesis testing

Suppose that we have sample  $X$ , that follow the distribution with density  $p(\cdot)$ , parametrized by (unknown) parameters  $\theta$  that we'd like to estimate from sample:

$$p(X|\theta) \rightarrow \max_{\theta}$$

Now let's turn it inside out. In the given setting, let's have a guess  $\theta_0$  of the parameters. Now, we can try to estimate the likelihood of observing our sample given these parameters. If the likelihood is high enough, we may assume that our guess is correct. This is the essence of hypothesis testing.

## Hypothesis testing

### Hypotheses

A hypothesis is any statement about sample:

$$H_0 : X \sim p(\cdot)$$

- 1. Hypothesis  $H_0$  is called null hypothesis.

- 2. This is a simple hypothesis. There exist composite hypothesis, when the distribution is defined up to a family:

$$H_0 : X \sim p(\cdot) \in \mathcal{F} = \{p_\theta(\cdot), \theta \in \Theta\}$$

- Example:  $H_0 : X \sim \mathcal{N}(\cdot, \cdot)$

- 1. We always have alternative hypothesis  $H_1$  (one-sided or two-sided), which is the negation of null hypothesis.

## Hypothesis testing

Now, how do we define whether our guess is correct? We may go for the likelihood, but it does not give us an answer, just a number. We would like to have an instrument that will simply tell us, whether the guess is correct or not. This instrument is called hypothesis testing. We will introduce a decision (rejection) rule in the following form:

$$\varphi(X) = \text{Ind}\{T(X) > c\}$$

where  $T$  is a test statistic (replacement for likelihood) and  $c$  is the statistic's critical value.

Since the decision is a random variable itself, it will never be certain. There may be two types of errors:

- 1. Reject null hypothesis when it's true (Type I)
- 2. Accept null hypothesis when it's false (Type II)

We'd like to limit the probability of Type I error with a (small)  $\alpha$  called confidence level.

$$P(H_1|H_0) \leq \alpha$$

The probability of not making Type II error is denoted  $\beta$  and is called test power.

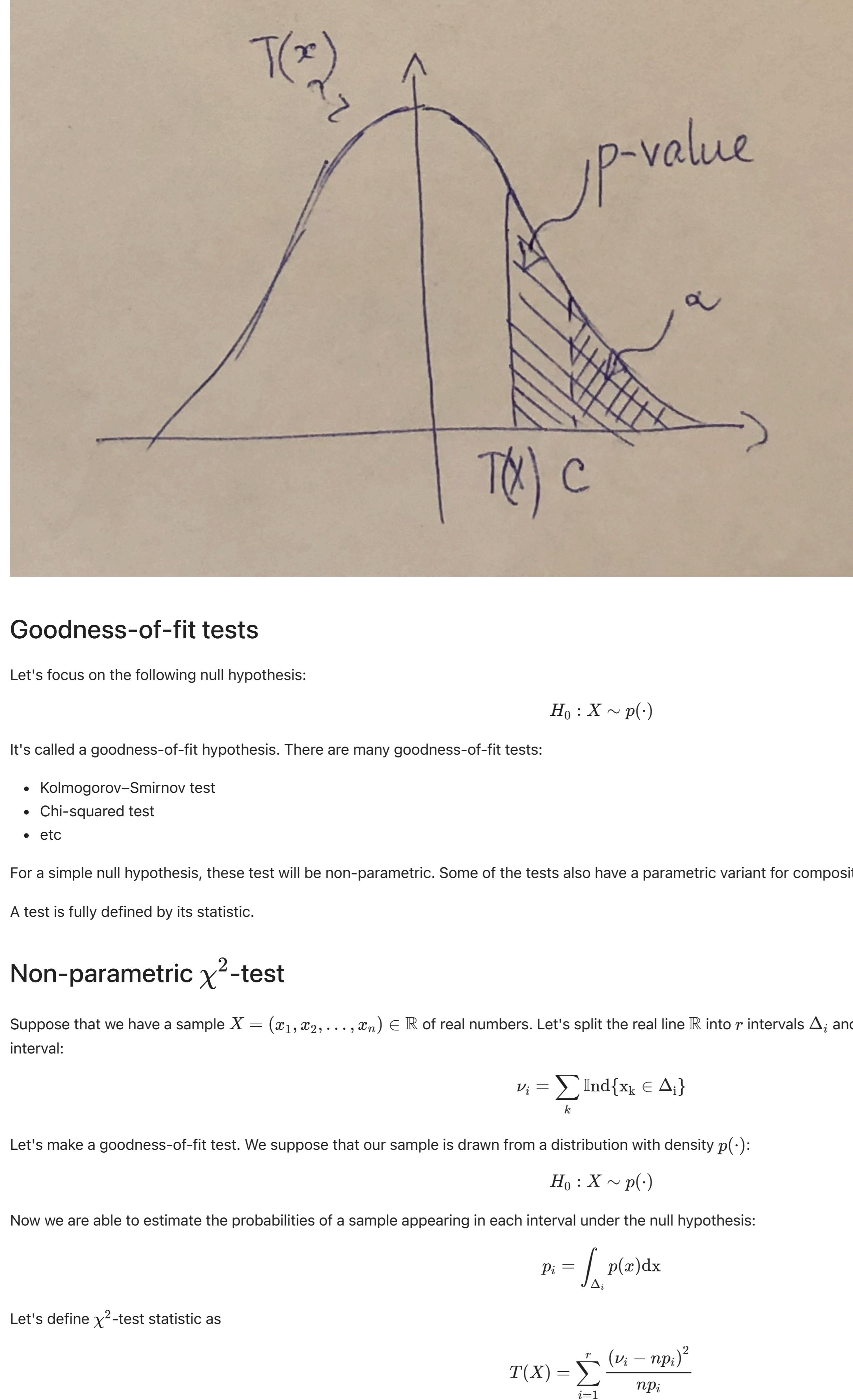
$$\beta = P(H_0|H_1) = 1 - P(H_1|H_0)$$

## Hypothesis testing

Decision (rejection) rule:

$$\varphi(X) = \text{Ind}\{T(X) > c\}$$

where  $T$  is a test statistic (replacement for likelihood) and  $c$  is the statistic's critical value.

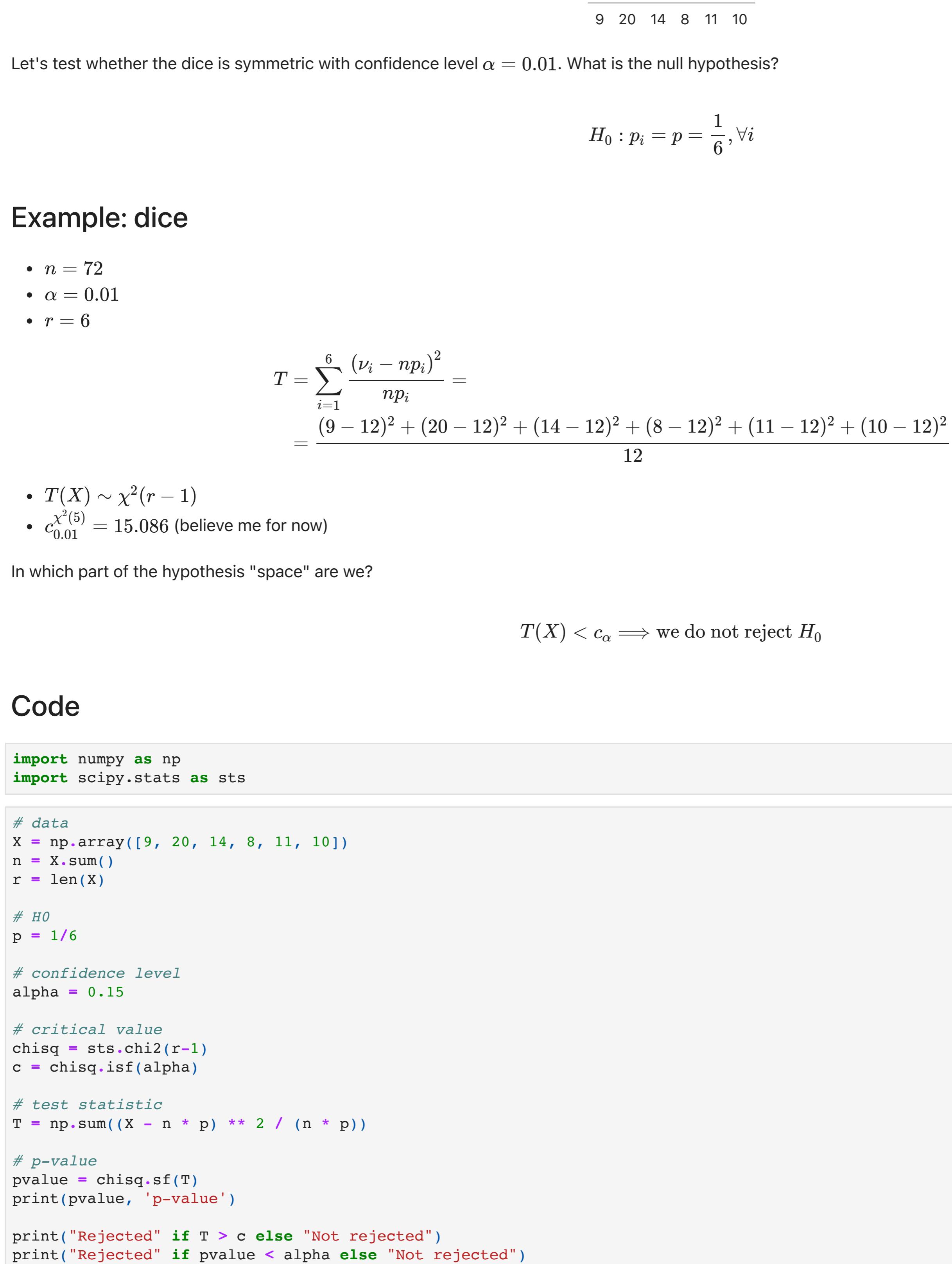


## Hypothesis testing

Decision (rejection) rule:

$$\varphi(X) = \text{Ind}\{T(X) > c\}$$

where  $T$  is a test statistic (replacement for likelihood) and  $c$  is the statistic's critical value.



## Goodness-of-fit tests

Let's focus on the following null hypothesis:

$$H_0 : X \sim p(\cdot)$$

It's called a goodness-of-fit hypothesis. There are many goodness-of-fit tests:

- Kolmogorov-Smirnov test
- Chi-squared test
- etc

For a simple null hypothesis, these test will be non-parametric. Some of the tests also have a parametric variant for composite hypothesis.

A test is fully defined by its statistic.

### Non-parametric $\chi^2$ -test

Suppose that we have a sample  $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}$  of real numbers. Let's split the real line  $\mathbb{R}$  into  $r$  intervals  $\Delta_i$ , and let's count how many samples reside in each interval:

$$\nu_i = \sum_k \text{Ind}\{x_k \in \Delta_i\}$$

Let's make a goodness-of-fit test. We suppose that our sample is drawn from a distribution with density  $p(\cdot)$ :

$$H_0 : X \sim p(\cdot)$$

Now we are able to estimate the probabilities of a sample appearing in each interval under the null hypothesis:

$$p_i = \int_{\Delta_i} p(x) dx$$

Let's define  $\chi^2$ -test statistic as

$$T(X) = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i}$$

The claim is

$$T(X) \xrightarrow{d} \chi^2(r-1)$$

### Example: dice

Let's assume that we rolled the dice  $n = 72$  times and obtained the following results  $\nu_i$ :

$$\begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 9 & 20 & 14 & 8 & 11 & 10 \end{array}$$

Let's test whether the dice is symmetric with confidence level  $\alpha = 0.01$ . What is the null hypothesis?

$$H_0 : p_i = \frac{1}{6}, \forall i$$

### Example: dice

- $n = 72$
- $\alpha = 0.01$
- $r = 6$

$$T = \sum_{i=1}^6 \frac{(\nu_i - np_i)^2}{np_i} = \frac{(9-12)^2 + (20-12)^2 + (14-12)^2 + (8-12)^2 + (11-12)^2 + (10-12)^2}{12} = \frac{98}{12} = 8.16$$

- $T(X) \sim \chi^2(r-1)$
- $c_{0.01}^{(15)} = 15.086$  (believe me for now)

In which part of the hypothesis "space" are we?

$$T(X) < c_{0.01} \implies \text{we do not reject } H_0$$

### Code

```
In [1]: import numpy as np
import scipy.stats as sts
```

```
In [5]: # data
x = np.array([19, 20, 14, 8, 11, 10])
n = x.sum()
r = len(x)
```

```
# H0
p = 1/6
```

```
# confidence level
alpha = 0.15
```

```
# critical value
chisq = sts.chisq(r-1)
```

```
c = chisq.isf(alpha)
```

```
# test statistic
T = np.sum((x - n * p) ** 2 / (n * p))
```

```
# p-value
pvalue = chisq.sf(T)
```

```
print(pvalue, "p-value")
```

```
0.1472863513001779 p-value
```

```
Rejected
```

```
Rejected
```

```
result = sts.chisquare(x)
```

```
print("Rejected" if result.pvalue < alpha else "Not rejected")
```

```
Not rejected
```

### Parametric $\chi^2$ -test

Same formulation: Suppose that we have a sample  $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}$  of real numbers. Let's split the real line  $\mathbb{R}$  into  $r$  intervals  $\Delta_i$ , and let's count how many samples reside in each interval:

$$\nu_i = \sum_k \text{Ind}\{x_k \in \Delta_i\}$$

Let's make a goodness-of-fit test. We suppose that our sample is drawn from a distribution with density  $p(\cdot)$ :

$$H_0 : X \sim p(\cdot)$$

Now we are able to estimate the probabilities of a sample appearing in each interval under the null hypothesis:

$$p_i = \int_{\Delta_i} p(x) dx$$

Let's define  $\chi^2$ -test statistic as

$$T(X) = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i}$$

The claim is

$$T(X) \xrightarrow{d} \chi^2(r-1)$$

But where do we get  $\theta$  from?

We can take MLE. This will result (prove!) in:

$$\sum_{i=1}^r \frac{\nu_i - np_i(\theta)}{np_i(\theta)} = 0$$

The p-value is

$$\sum_{i=1}^r \frac{\nu_i - np_i(\theta)}{np_i(\theta)} = \chi^2(r-1)$$

Example: dice

- $n = 72$
- $\alpha = 0.01$
- $r = 6$

$$T = \sum_{i=1}^6 \frac{(\nu_i - np_i)^2}{np_i} = \frac{(9-12)^2 + (20-12)^2 + (14-12)^2 + (8-12)^2 + (11-12)^2 + (10-12)^2}{12} = \frac{98}{12} = 8.16$$

- $T(X) \sim \chi^2(r-1)$

- $c_{0.01}^{(15)} = 15.086$  (believe me for now)

In which part of the hypothesis "space" are we?

$$T(X) < c_{0.01} \implies \text{we do not reject } H_0$$

### Code

```
In [1]: import numpy as np
import scipy.stats as sts
```

```
In [5]: # data
x = np.array([19, 20, 14, 8, 11, 10])
n = x.sum()
r = len(x)
```

```
# H0
p = 1/6
```

```
# confidence level
alpha = 0.15
```

```
# critical value
chisq = sts.chisq(r-1)
```

```
c = chisq.isf(alpha)
```

```
# test statistic
T = np.sum((x - n * p) ** 2 / (n * p))
```

```
# p-value
pvalue = chisq.sf(T)
```

```
print(pvalue, "p-value")
```

```
0.1472863513001779 p-value
```

```
Rejected
```

```
Rejected
```

```
result = sts.chisquare(x)
```

```
print("Rejected" if result.pvalue < alpha else "Not rejected")
```

```
Not rejected
```

### Parametric $\chi^2$ -test

The test statistic is defined as

$$\varphi(X) = \text{Ind}\{T(X) > c\}$$

The claim is

$$T(X) \xrightarrow{d} \chi^2(r-1)$$

But where do we get  $\theta$  from?

We can take MLE. This will result (prove!) in:

$$\sum_{i=1}^r \frac{\nu_i - np_i(\theta)}{np_i(\theta)} = 0$$

The p-value is

$$\sum_{i=1}^r \frac{\nu_i - np_i(\theta)}{np_i(\theta)} = \chi^2(r-1)$$

Example: dice

- $n = 72$
- $\alpha = 0.01$
- $r = 6$

$$T = \sum_{i=1}^6 \frac{(\nu_i - np_i)^2}{np_i} = \frac{(9-12)^2 + (20-12)^2 + (14-12)^2 + (8-12)^2 + (11-12)^2 + (10-12)^2}{12} = \frac{98}{12} = 8.16$$

- $T(X) \sim \chi^2(r-1)$

- $c_{0.01}^{(15)} = 15.086$  (believe me for now)

In which part of the hypothesis "space" are we?

$$T(X) < c_{0.01} \implies \text{we do not reject } H_0$$