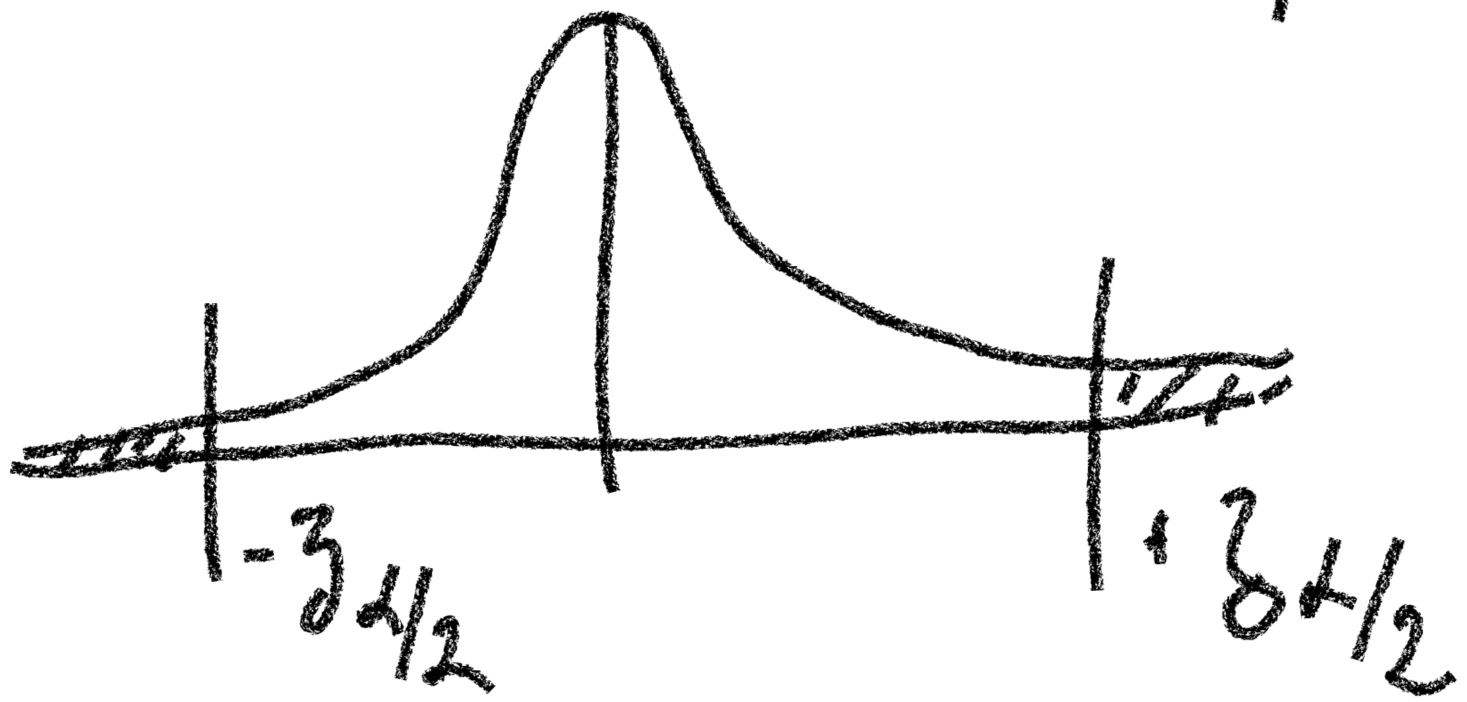
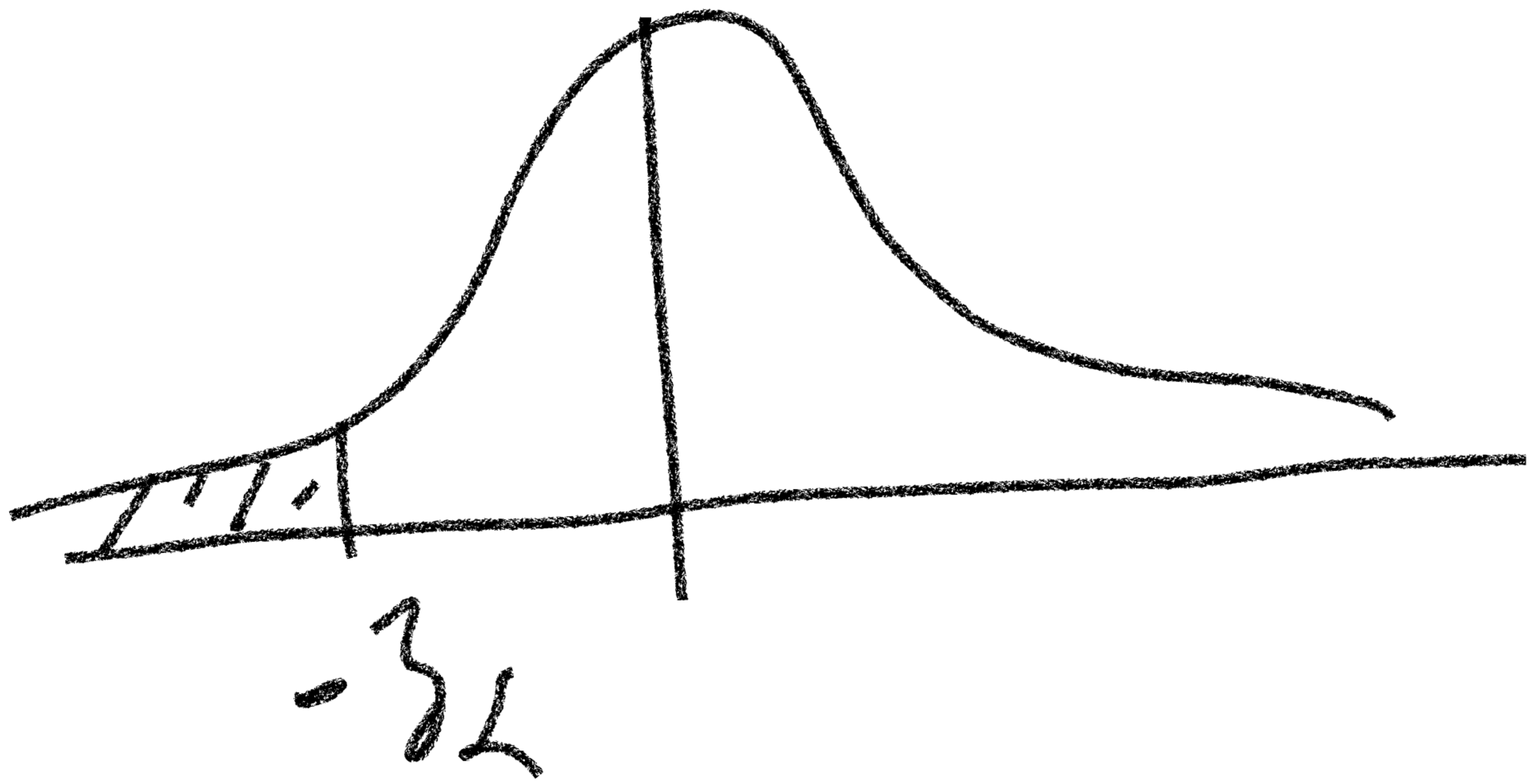
$$\boxed{\begin{array}{r}\text{One Sample } Z\text{-Test} \\ T\text{-Test}\end{array}}$$

$H0: \mu = \mu_0 \qquad \alpha = 5\% \,/\, 1\%$

$H1: \mu \neq \mu_0$ (two-tailed / two-sided)



$\mu < \mu_0$ (one-sided / one-tailed)



$$z = \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{N}}} \sim \mathcal{N}(0, 1)$$

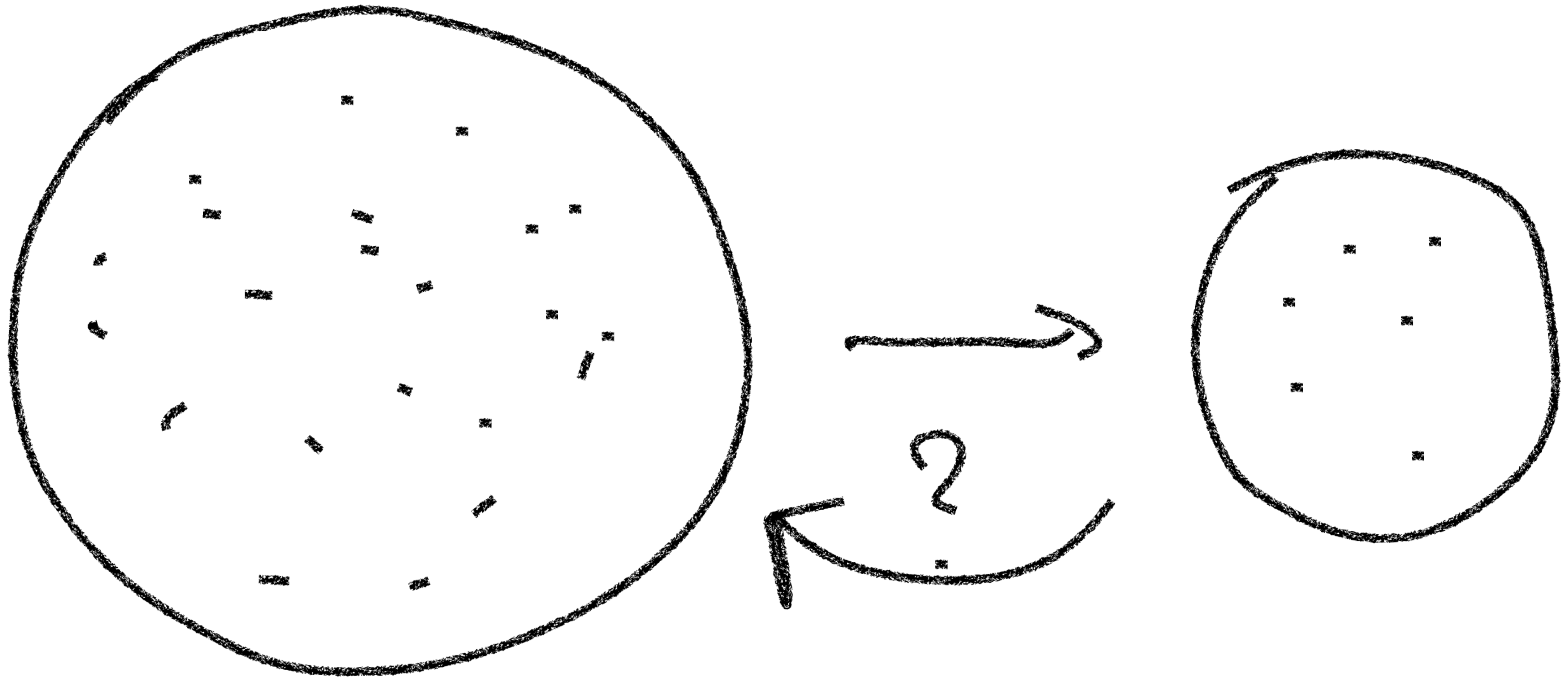
$$t = \frac{\mu - \mu_0}{\frac{\sigma_{emp}}{\sqrt{N-1}}} \sim \text{Loi de Student à } N-1 \text{ degrés de liberté}$$

$N \gg 1$: on peut faire un Z-test

N "pas très grand" < 100 :

Si on connaît σ_0, on utilise le z-test

Sinon, le t-test



Two - Sample tests : A / B testing

1) Paired t - test :

A suivie avant et après une expérience
(ex : administrat° d'un traitement médical)

On regarde une certaine variable avant et après (ex : tx de charge virale via PCR)

Paired t-test

One Sample t-test

H0 : $\mu_1 = \mu_2$ $\Longleftrightarrow$ H0: $\mu_1 - \mu_2 = 0$

H1 : $\mu_1 \neq \mu_2$

H1: $\mu_1 - \mu_2 \neq 0$

ttest1samp

ttestrel

(Scipy)

Independant Two Sample T-test

A / B $\quad H_0 : \mu_A = \mu_B \quad H_1 : \mu_A \neq \mu_B$

$\sigma_A = \sigma_B$

ttest_ind(..., equal_var = True/False)

$\sigma_A \neq \sigma_B$

z-test , t-test , ANOVA , test s/ la variance (χ_2)

↳ variables numériques

quel test(s) utiliser dans le cas de variables catégorielles ?

1) À quel point une distribution d'un échantillon observé correspond à une distribution de référence connue ?

Ex : Un dé à 6 faces. Vous le lancez plusieurs fois. Est-ce que la distribution obtenue correspond à une loi uniforme ?

2) Étant donné 2 variables catégorielles observées, peut-on rejeter l'hypothèse d'indépendance ?

<u>Ex :</u> Variable A : "animaux" {"chien", "chat", "hamster"}

Variable B : "ville" {"Paris", "Rennes", "Marseille"}

	animaux	ville
id=...	chien	Marseille
	chat	Lyon
	⋮	⋮

Goodness of Fit :

$$G = \frac{(\text{observations} - \text{attendus})^2}{\text{attendus}}$$

Distribution observée a loi uniforme

Face	Observations	Attendus
1	5	10
2	8	10
3	9	10
4	8	10
5	10	10
6	20	10
Total	60	

$$G = \frac{(O-E)^2}{E} = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i}$$

$$G = \frac{1}{10}\left((-5)^2 + (-2)^2 + (-1)^2 + (-2)^2 + 0^2 + 10^2\right) = 13{,}4$$

Animaux \ Ville	Paris	Rennes	Mars.	Total
Chat	26	48	12	86
Chien	52	97	23	172
Hamster	12	25	6	43
Total	90	170	41	301

H0 : Indép. entre Animaux et Ville

H1 : Pas d'ind.

CSQ

$E_{Chien, Rennes} = N \cdot p_{Chien} \cdot p_{Rennes}$

$p_{Chien} = \frac{172}{301}$

$p_{Rennes} = \frac{170}{301}$

$N = 301$

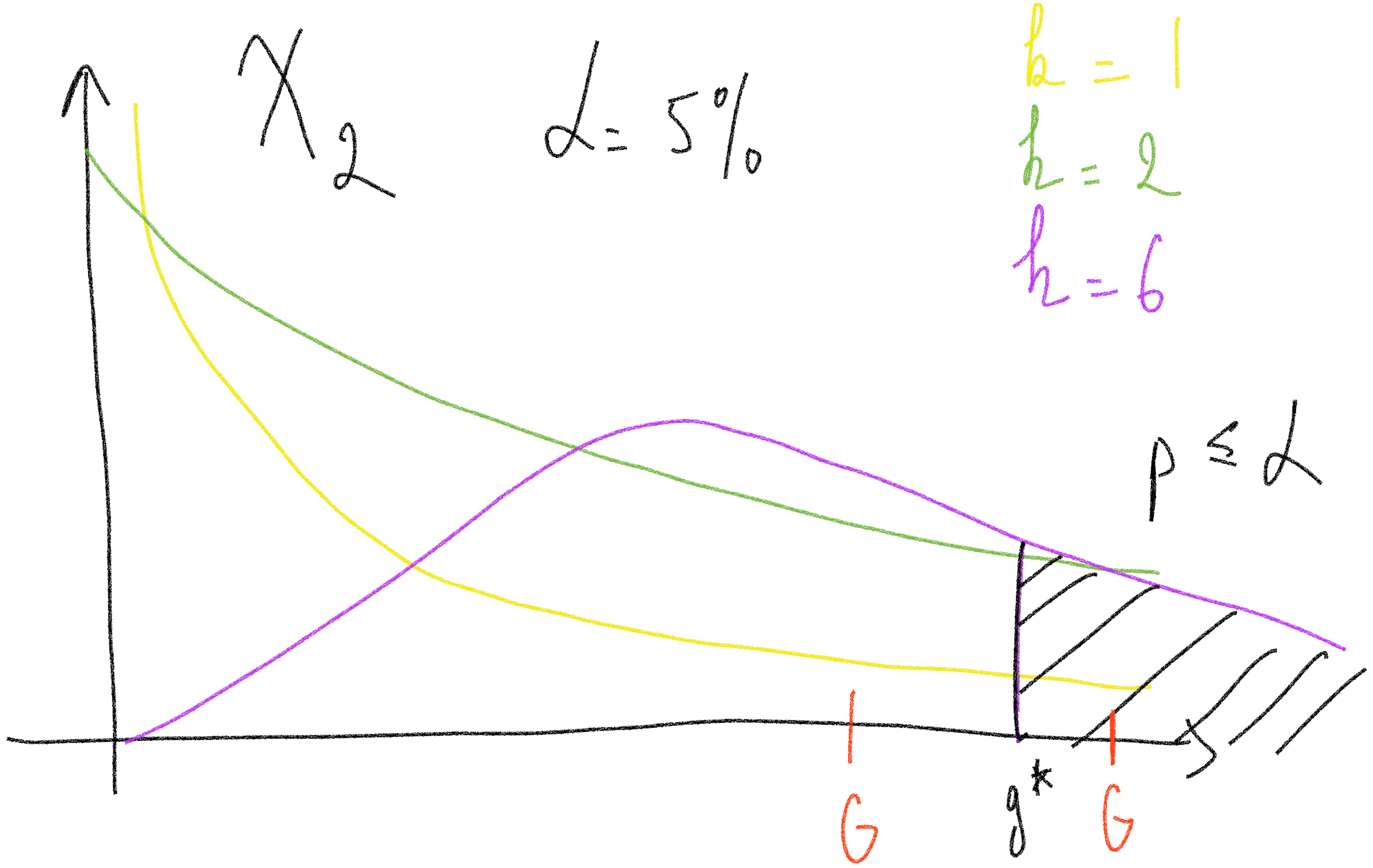
$E_{Chien, Rennes} = 301 \cdot \left(\frac{172}{301}\right) \cdot \left(\frac{170}{301}\right) \approx 97{,}14$

$$G = \sum_{\substack{i=1 \\ \text{lignes}}}^{3} \sum_{\substack{j=1 \\ \text{colonnes}}}^{3} \frac{(O_{i,j} - E_{i,j})^2}{E_{ij}}$$

$G \sim \chi_2$ (k degrés de liberté)

1) $k = 6 - 1 = 5$

2) $k = (\text{colonnes} - 1) \times (\text{lignes} - 1)$
$= (3-1)(3-1) = 4$


χ_2
$\alpha = 5\%$
$k = 1$
$k = 2$
$k = 6$
$p \leq \alpha$
G
g^*
G

$k = 5$ $\qquad G = 13{,}4$

$g^{*}_{\alpha = 5\%} = 11{,}070$ $\qquad G > g^{*}_{5\%} \Rightarrow$ On rejette $\mathcal{H}_0$ (le dé est pipé)