LOI DU CHI-DEUX

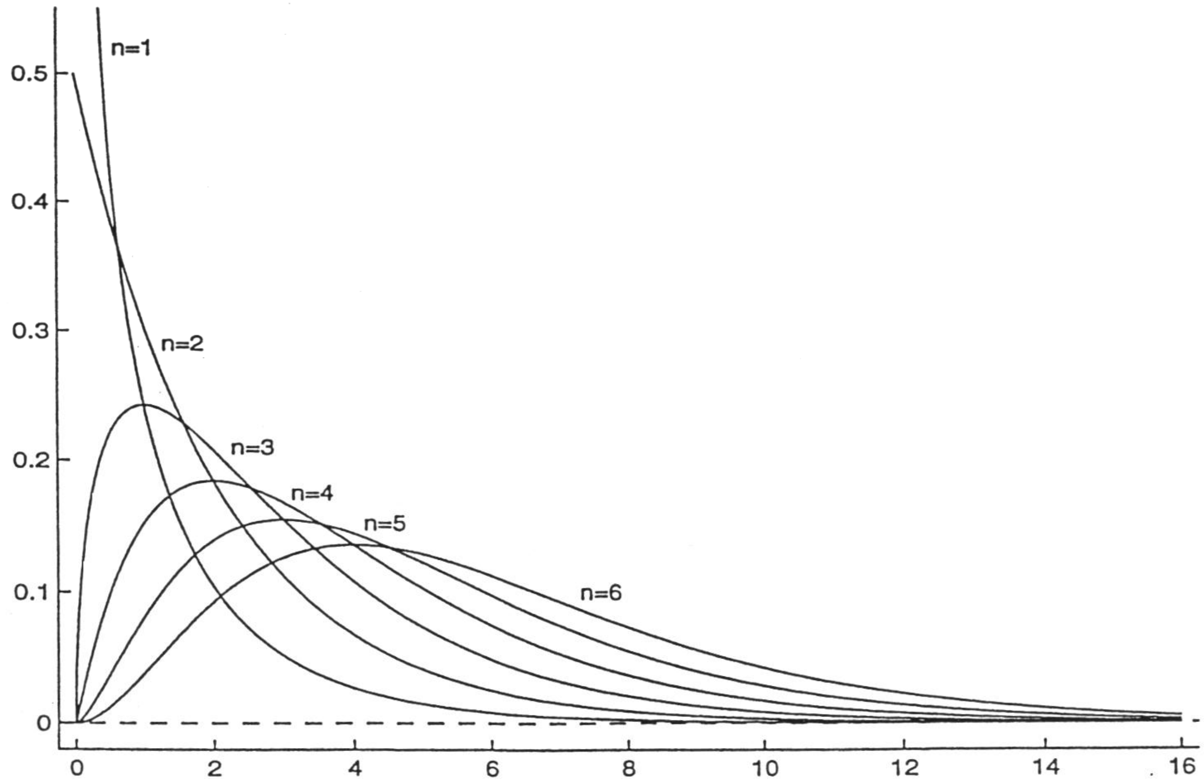
Définition : Soit $(\mathrm{X}_i)_{i=1,\dots,n}$ une suite de variables aléatoires indépendantes normales centrées réduites (c'est à dire telles que $\mathcal{L}(\mathrm{X}_i)=\mathcal{N}(0,1)$).

La somme des carrés des X_i : $\mathrm{S}=\sum_{i=1}^{n}\mathrm{X}_i^{\,2}$ est une v.a. qui suit une loi du chi-deux à *n* degrés de liberté.

On note $\mathcal{L}(\mathrm{S})=\mathcal{X}_n^2$

Propriétés : - Si $\mathcal{L}(\mathrm{S})=\mathcal{X}_n^2$ alors $\mathrm{E}(\mathrm{S})=n$ et $\mathrm{Var}(\mathrm{S})=2n$

- Si $\mathcal{L}(\mathrm{K}_1)=\mathcal{X}_{n1}^2$ et $\mathcal{L}(\mathrm{K}_2)=\mathcal{X}_{n2}^2$ et si $\mathrm{K}_1\perp\mathrm{K}_2$ alors $\mathcal{L}(\mathrm{K}_1+\mathrm{K}_2)=\mathcal{X}_{n1+n2}^2$



Fonctions densité de probabilité de la loi du $\mathcal{X}_n^2$ pour quelques valeurs de *n*

Une situation où intervient la loi du chi-deux :

Soit $(\mathrm{X}_i)_{i=1,\dots,n}$ un *n*-échantillon tel que $\mathcal{L}(\mathrm{X}_i)=\mathcal{N}(\mu,\sigma)$ avec μ et σ inconnus.

Dans ce cas μ est estimé par : $\overline{\mathrm{X}}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i$ et σ par $\mathrm{S_c}^{2}=\frac{1}{n-1}\sum_{i=1}^{n}\left(\mathrm{X}_i-\overline{\mathrm{X}}\right)^2$

On a alors : $\mathcal{L}\left(\frac{(n-1)\mathrm{S_c}^{2}}{\sigma^2}\right)=\mathcal{X}_{n-1}^2$ ou encore $\mathcal{L}\left(\frac{n\mathrm{S}^2}{\sigma^2}\right)=\mathcal{X}_{n-1}^2$ avec $\mathrm{S}^2=\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{X}_i-\overline{\mathrm{X}}\right)^2$

TEST D'ADEQUATION A UNE LOI

On considère une variable X à r modalités pour laquelle on a une série statistique d'effectifs $\left(n_1, n_2, \ldots, n_r\right)$. On note $n = n_1 + n_2 + \ldots + n_r$ l'effectif total.
On se demande si ces observations correspondent à la loi de probabilité définie sur les mêmes r modalités par les probabilités $\left(p_1, p_2, \ldots, p_r\right)$

On veut tester	$\mathrm{H}_0 : \mathcal{L}(\mathrm{X}) = \left(p_1, p_2, \ldots, p_r\right)$	***contre***	$\mathrm{H}_1 : \mathcal{L}(\mathrm{X}) \neq \left(p_1, p_2, \ldots, p_r\right)$

On définit la statistique du $\mathcal{X}^2$ d'adéquation par :

$$kh = \sum_{i=1}^{r} \frac{\left(n_i - np_i\right)^2}{np_i}$$

Les n_i sont les effectifs observés, $n = n_1 + n_2 + \ldots + n_r$ est l'effectif total et les np_i sont les effectifs « théoriques » de chaque modalité.

Propriété : Si l'hypothèse H_0 est vraie alors la statistique kh est la réalisation lorsque $n \to +\infty$ et $n_i \to +\infty$ pour $i = 1, \ldots r$ d'une variable aléatoire KH telle que $\mathcal{L}(\mathrm{KH}) = \mathcal{X}^2_{r-1}$.

En pratique, nous utilisons cette loi asymptotique dès que $n \geq 50$ et $np_i \geq 5$ pour $i = 1, \ldots r$.
Si la condition $np_i \geq 5$ n'est pas remplie on regroupera les modalités voisines pour qu'elle soit remplie.

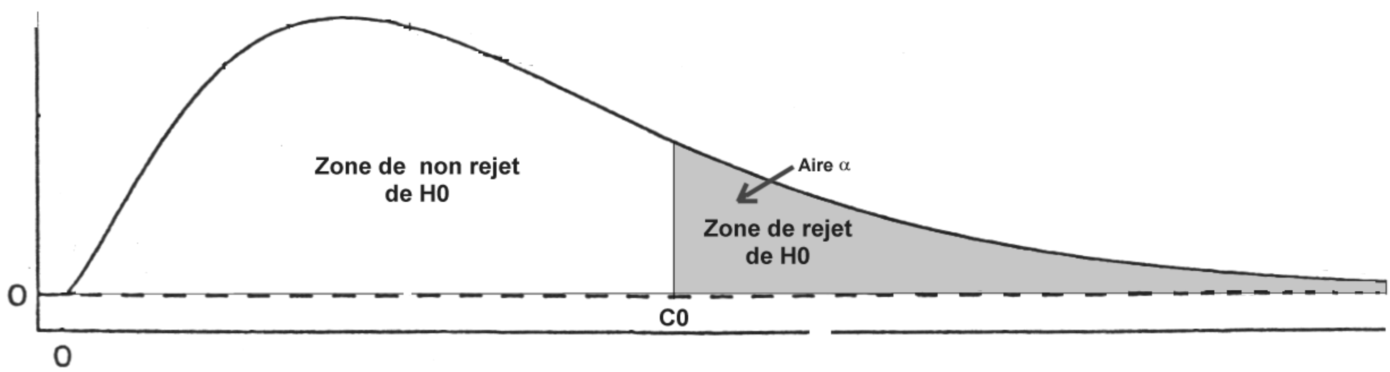
<u>**Règle de décision**</u> **:**

Pour un seuil α donné, les tables du $\mathcal{X}^2$ nous donnent une constante c_0 telle que :

$$\mathrm{P}\left(\mathcal{X}^2_{r-1} > c_0\right) = \alpha .$$

Si $kh < c_0$ j'accepte l'hypothèse H_0 au niveau $1-\alpha$

Si $kh \geq c_0$ je rejette l'hypothèse H_0 au seuil α



Exemple : souris des villes et souris des champs.

Dans une étude portant sur l'orientation spatiale chez les souris, les animaux de l'expérience ont été placés un par un au centre d'un labyrinthe radiaire comportant 8 allées orientées dans les huit directions de la rose des vents. Chaque animal s'est échappé par l'une de ces allées. Les expériences ont porté d'une part sur des souris de laboratoire et d'autre part sur ***des souris sauvages récemment capturées en un lieu situé au nord-est du laboratoire***. Les répartitions des directions de fuite sont données dans le tableau ci-dessous.

Directions	Souris de laboratoire	Souris sauvages
N	17	26
NO	25	17
O	13	9
SO	28	2
S	19	3
SE	20	16
E	22	33
NE	16	54

On désire tester, dans le cas des souris de laboratoire d'une part puis dans le cas des souris sauvages l'hypothèse que le choix de la direction de fuite se fait au hasard ?

Si la réponse est OUI alors l''effectif total (160) ainsi que les effectifs théoriques des 8 modalités (20) vérifient les conditions d'application **du test d'adéquation du** $\mathcal{X}^2$

Directions	Souris de laboratoire			Souris sauvages		
	ni	npi	(ni-npi)²/npi	ni	npi	(ni-npi)²/npi
N	17	20	0,45	26	20	1,80
NO	25	20	1,25	17	20	0,45
O	13	20	2,45	9	20	6,05
SO	28	20	3,20	2	20	16,20
S	19	20	0,05	3	20	14,45
SE	20	20	0,00	16	20	0,80
E	22	20	0,20	33	20	8,45
NE	16	20	0,80	54	20	57,80
	160		8,40	160		106,00

Choisissons le seuil de 0.05. Le nombre de degrés de liberté est 8-1=7
La lecture dans la table du $\mathcal{X}^2$ (voir la table en page 4 de ce document) donne $c_0 = 14.07$

- ➔ Dans le cas des souris de laboratoire : $\boldsymbol{kh \approx 8.4}$ est inférieur à $\boldsymbol{c_0 = 14.07}$. Les écarts entre les valeurs observés et les valeurs théoriques ne sont pas significatifs et on accepte l'hypothèse H_0 au niveau $1-\alpha$

- ➔ Dans le cas des souris sauvages : $\boldsymbol{kh \approx 106}$ est largement supérieur à $\boldsymbol{c_0 = 14.07}$. Les écarts entre les valeurs observés et les valeurs théoriques sont significatifs et on rejette l'hypothèse H_0 au seuil α.

Annexe

TABLE DE LA FONCTION DE REPARTITION DU $\mathcal{X}^2$

Degré	Probabilité									
liberté	**0,005**	**0,010**	**0,025**	**0,050**	**0,100**	**0,900**	0,950	**0,975**	**0,990**	**0,995**
1	0,00	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,17	2,83	12,02	**14,07**	16,01	18,48	20,28
8	1,34	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,73	26,76
12	3,07	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	49,65
28	12,46	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	53,67
31	14,46	15,66	17,54	19,28	21,43	41,42	44,99	48,23	52,19	55,00
32	15,13	16,36	18,29	20,07	22,27	42,58	46,19	49,48	53,49	56,33
33	15,82	17,07	19,05	20,87	23,11	43,75	47,40	50,73	54,78	57,65
34	16,50	17,79	19,81	21,66	23,95	44,90	48,60	51,97	56,06	58,96
35	17,19	18,51	20,57	22,47	24,80	46,06	49,80	53,20	57,34	60,27
36	17,89	19,23	21,34	23,27	25,64	47,21	51,00	54,44	58,62	61,58
37	18,59	19,96	22,11	24,07	26,49	48,36	52,19	55,67	59,89	62,88
38	19,29	20,69	22,88	24,88	27,34	49,51	53,38	56,90	61,16	64,18
39	20,00	21,43	23,65	25,70	28,20	50,66	54,57	58,12	62,43	65,48
40	20,71	22,16	24,43	26,51	29,05	51,81	55,76	59,34	63,69	66,77