Chapitre 12

Tests du Chi-deux

Nous découvrons ici à la fois la notion de test et un exemple d'utilisation d'une loi très utilisée en statistique : la loi du Chi-deux. Rappelons la définition de la loi du Chi-deux.

Définition : Soit $k \geq 1$ un entier et $(U_j)_{j=1..k}$ un k v.a. i.i.d. normales, centrées, réduites : $U_j \rightsquigarrow \mathcal{N}(0,1)$. La v.a. $X := U_1^2 + \ldots + U_k^2$ suit une loi appelée *loi du Chi-deux à k degrés de liberté*, et on note $X \rightsquigarrow \chi_2(k)$. On montre que $f_X(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}$.

12.1 Exemple : tests de conformité

On tire 100 fois à pile ou face (avec une même pièce) et on obtient face 61 fois. Dans quelle mesure peut-on en déduire que la pièce est truquée ? Tel est le type de problèmes statistiques auxquels tente de répondre la théorie des tests. La réponse ici est, qu'avec un risque de se tromper inférieur à $\alpha = 5\%$, on peut affirmer que la pièce n'a pas une probabilité $p = 0.5$ de donner face : on dit qu'*on peut rejeter au seuil $\alpha = 5\%$ l'hypothèse H_0* = [la probabilté p de produire "face" est 0.5]. Voici comment on arrive à cette conclusion :

1ère étape : Formuler l'hypothèse nulle et fixer le seuil α. Il s'agit souvent de l'hypothèse qu'on souhaite réfuter (ou *rejeter*). On l'appelle l'hypothèse H_0 ou *hypothèse nulle du test ;* on fera tous les calculs sous cette hypothèse. Elle porte sur la loi suivie par l'échantillon qui sera considéré. Dans notre exemple l'hypothèse nulle est que la pièce suit une loi $\mathcal{B}(1, 0.5)$ (bien que nous soupçonnons que la pièce soit truquée : nous nous plaçons dans un démarche de réfutation). Envisageons un cas un peu plus général : les V_j de notre échantillon sont à valeurs dans un ensemble à $k+1$ valeurs : $V_j \in \{v^1, \ldots, v^i, \ldots, v^{k+1}\}$. On note $p(i) := \mathbb{P}(\{V_j = v^i\})$, la probabilité étant précisément celle correspondant à l'hypothèse nulle H_0.

2ème étape : Procéder à l'échantillonage : on note n la taille de l'échantillon. On compte et note o_i l'effectif (le nombre d'éléments) de l'ensemble des éléments de l'échantillons qui ont produit la valeur v^i ; on l'appelle l'*effectif observé* ou empirique. On détermine la valeur $t_i = n \cdot p(i)$ de l'*effectif théorique.*

3ème étape : Calcul de la statistique : calculer le nombre

$$\boxed{\chi^* := \sum_{i=1}^{k+1} \frac{(o_i - t_i)^2}{t_i}}$$

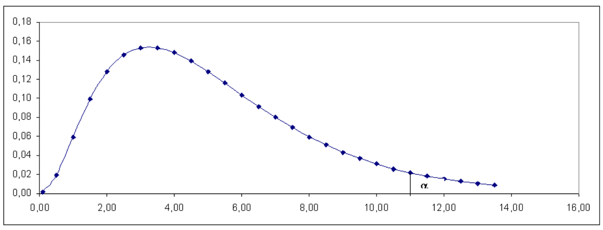

FIG. 12.1 – Densité de la lois du Chi-deux à 5 degrés de liberté $\chi_2(5)$.

4ème étape Appliquer le théorème suivant :

> **Théorème 12.1** *Soit $(V_j)_{j=1..n}$ un échantillon de v.a. à valeurs dans $\{v^1, \ldots, v^i, \ldots, v^{k+1}\}$, et soit $p(i) := \mathbb{P}(\{V_j = v^i\})$. Soit $O_i := \sum_{j=1}^n \mathbb{I}_{\{V_j=v^i\}}$ le nombre d'occurences de v^i dans l'échantillon ; soit $t_i := np(i)$. Alors la loi de $X_n := \sum_{i=1}^{k+1} \frac{(O_i-t_i)^2}{t_i}$, tend,lorsque n tend vers l'infini, vers une loi du Chi-deux à k degrés de liberté $\chi_2(k)$.*
>
> A cette fin, on détermine à l'aide de la table des $\chi_2(k)$ la valeur $\chi(k, \alpha)$ telle que $\mathbb{P}(\{X > \chi(k, \alpha)\}) = \alpha$ (pour n'importe quel $X \rightsquigarrow \chi_2(k)$). Si $\chi^* > \chi(k, \alpha)$, on *rejette l'hypothése H_0*

Exemple : Dans notre exemple on a $H_0 = [p = 0.5]$, $\alpha = 5\%$, $n = 100$, $k = 1$, $o_1 = 61$, $o_2 = 39$, $t_1 = t_2 = 0.5 \cdot 100 = 50$, et donc $\chi^* = \frac{(61-50)^2}{50} + \frac{(39-50)^2}{50} = 4.84$. Par ailleurs, dans la table, on trouve que $\chi(1, 0.05) = 3.83..$; on a donc $\chi^* = 4.84 > 3.83.. = \chi(1, 0.05)$, et on peut donc *rejeter, au seuil $\alpha = 5\%$*, l'hypothèse que la pièce n'est pas truquée. Observez qu'on n'aurait pas abouti à la même conclusion si le nombre de faces, sur 100 tirages, avait été de 59, car dans ce cas on aurait eu la statistique $\chi^* = 3.24$. En revanche, si on avait tiré 118 faces sur 200 tirages, on aurait eu $\chi^* = \frac{(118-100)^2}{100} + \frac{(92-10O)^2}{100} = 6.48 > 3.84$. En fait, au seuil $\alpha = 5\%$, 115 faces sur 200 tirages suffisent à rejeter H_0.

12.2 Preuve du théorème, pour $k = 1$

Voir cours

	0,99	0,95	0,9	0,75	0,5	0,25	0,1	0,05	0,025	0,01
1	0,0001571	0,0039322	0,0157907	0,1015311	0,4549362	1,3233042	2,7055406	3,8414553	5,0239026	6,6348913
2	0,0201004	0,1025862	0,2107208	0,5753639	1,3862936	2,7725904	4,6051761	5,9914764	7,3777791	9,2103510
3	0,1148316	0,3518460	0,5843755	1,2125321	2,3659727	4,1083421	6,2513945	7,8147247	9,3484040	11,3448821
4	0,2971068	0,7107241	1,0636243	1,9225580	3,3566947	5,3852661	7,7794340	9,4877285	11,1432620	13,2766986
5	0,5542969	1,1454773	1,6103091	2,6746042	4,3514587	6,6256784	9,2363491	11,0704826	12,8324920	15,0863174
6	0,8720833	1,6353805	2,2041303	3,4545975	5,3481190	7,8408057	10,6446375	12,5915774	14,4493550	16,8118718
7	1,2390317	2,1673492	2,8331052	4,2548522	6,3458093	9,0371459	12,0170314	14,0671273	16,0127737	18,4753241
8	1,6465062	2,7326326	3,4895374	5,0706416	7,3441201	10,2188538	13,3615619	15,5073125	17,5345446	20,0901592
9	2,0878894	3,3251151	4,1681557	5,8988229	8,3428320	11,3887495	14,6836632	16,9189602	19,0227776	21,6660476
10	2,5581988	3,9402953	4,8651783	6,7371986	9,3418161	12,5488588	15,9871747	18,3070290	20,4832007	23,2092872
11	3,0534957	4,5748090	5,5777883	7,5841449	10,3409955	13,7006897	17,2750067	19,6751531	21,9200227	24,7250219
12	3,5705513	5,2260277	6,3037959	8,4384194	11,3403219	14,8453991	18,5493402	21,0260554	23,3366602	26,2169637
13	4,1068996	5,8918606	7,0414997	9,2990633	12,3397531	15,9839051	19,8119327	22,3620266	24,7355809	27,6881845
14	4,6604155	6,5706316	7,7895377	10,1653113	13,3392718	17,1169328	21,0641406	23,6847823	26,1189349	29,1411633
15	5,2293559	7,2609348	8,5467531	11,0365377	14,3388572	18,2450842	22,3071206	24,9957967	27,4883647	30,5779507
16	5,8121968	7,9616386	9,3122353	11,9122166	15,3384973	19,3688566	23,5418215	26,2962209	28,8453246	31,9998609
17	6,4077420	8,6717536	10,0851830	12,7919245	16,3381790	20,4886786	24,7690282	27,5871003	30,1909826	33,4087170
18	7,0149034	9,3904479	10,8649369	13,6752906	17,3379020	21,6048862	25,9894184	28,8693210	31,5264102	34,8052374
19	7,6326976	10,1170062	11,6509120	14,5619976	18,3376500	22,7178053	27,2035648	30,1435051	32,8523370	36,1907747
20	8,2603684	10,8507994	12,4426014	15,4517747	19,3374296	23,8276894	28,4119699	31,4104204	34,1695814	37,5662715
21	8,8971724	11,5913160	13,2395955	16,3443873	20,3372282	24,9347832	29,6150859	32,6705580	35,4788557	38,9322325
22	9,5424944	12,3380095	14,0414896	17,2396185	21,3370437	26,0392635	30,8132853	33,9244598	36,7806781	40,2894485
23	10,1956888	13,0905050	14,8479543	18,1372940	22,3368799	27,1413291	32,0068902	35,1724602	38,0756095	41,6383344
24	10,8563494	13,8484222	15,6586793	19,0372505	23,3367299	28,2411501	33,1962351	36,4150265	39,3640601	42,9797813
25	11,5239511	14,6113957	16,4734055	19,9393377	24,3365837	29,3388466	34,3815833	37,6524894	40,6464978	44,3140141
26	12,1981769	15,3791626	17,2918796	20,8434347	25,3364585	30,4345588	35,5631637	38,8851296	41,9231379	45,6416362
27	12,8784685	16,1513946	18,1138885	21,7494036	26,3363413	31,5284104	36,7412276	40,1132656	43,1945211	46,9628372
28	13,5646661	16,9278763	18,9392353	22,6571575	27,3362315	32,6204888	37,9159074	41,3371517	44,4607905	48,2781662
29	14,2564062	17,7083814	19,7677396	23,5665882	28,3361296	33,7109064	39,0874753	42,5569475	45,7222795	49,5878290
30	14,9534644	18,4926672	20,5992447	24,4776030	29,3360283	34,7997355	40,2560170	43,7729539	46,9792176	50,8921806

FIG. 12.2 – Lois du Chi-deux : Les lignes correspondent au nombre de degrés de liberté, entre 1 et 30. Chaque colonne correspond à une même probabilité d'être supérieur à la valeur indiquée dans le tableau.