IMBS3 - ISCID-CO, site de Dunkerque, 2015/2016

Mathématiques appliquées à la gestion - TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES

Fiche de Mathématiques 3 - Test d'ajustement du χ^2.

1 Introduction

Certains tests ont pour objet de tirer des conclusions relatives à la valeur des paramètres (moyenne, fréquence, variance) d'une ou plusieurs populations, sur la base d'informations partielles fournies par un ou plusieurs échantillons.
La même démarche peut être appliquée pour porter un " jugement" sur les caractéristiques encore plus générales de la population : la forme même de distribution du caractère étudié, la validité de sa représentation à l'aide de telle ou telle loi de probabilité particulière, les relations éventuelles entre plusieurs variables.

Concrètement, on dispose d'une distribution statistique empirique se présentant sous la forme d'une table d'effectifs ou de fréquences du caractère étudié. On désire savoir si ces effectifs ou ces fréquences sont compatibles avec une distribution théorique déterminée telle que la loi binomiale, la loi de Poisson, la loi normale ou toute autre loi de probabilité. Il s'agit en d'autres termes d'apprécier l'adéquation d'une distribution théorique particulière, en tant que représentation d'un phénomène concret observé (série empirique).

La démarche consiste donc à tester l'hypothèse selon laquelle notre échantillon serait tiré d'une population régie par une certaine loi de probabilité.
Il est évident que, même si le phénomène étudié suit effectivement la loi de probabilité dont on teste l'adéquation, les fréquences expérimentales (ou empiriques) observées sur un échantillon particulier différeront nécessairement plus ou moins des probabilités (fréquences que l'on devrait théoriquement observer selon la loi en question).

La problématique du test revient en définitive à savoir si les différences constatées entre la distribution expérimentale et la distribution théorique supposée sont explicables par l'aléa lié à la constitution de l'échantillon ou si elles sont trop importantes pour être imputables au seul hasard. En ce cas, c'est l'hypothèse de travail avancée sur la nature de la distribution qui devrait être mise en cause.

2 Test d'ajustement du χ^2 pour une loi spécifiée

Très souvent, lors de la résolution d'un problème, on rencontre des phrases du type : " Si la loi de la variable X est normale..." ou " Supposons que la loi de X soit de Bernoulli de paramètre $p = 1/2$, ..." ou en employant un langage plus courant " Supposons que deux structures différentes soient également réparties chez les bactéries ". Comment vérifier l'exactitude de ces hypothèses?
Les techniques appropriées sont appelées **les tests d'ajustement ou tests d'adéquation** (*fit tests* en anglais). Étant donné une loi de probabilité théorique, il s'agit de savoir, à partir d'un n-échantillon, c'est-à-dire de n observations indépendantes, d'une variable aléatoire X, si cette variable obéit bien à la loi spécifiée.
Le test le plus usuel est celui du χ^2 d'ajustement pour une loi multinomiale décrit au début du paragraphe suivant.

2.1 Cas d'une variable discrète

X a un nombre fini de modalités, notées $1, 2, \dots, r$ et il s'agit de tester l'hypothèse

$$H_0 : p(X = 1) = p_1,\ p(X = 2) = p_2, \dots,\ p(X = r) = p_r,$$

où $p_1, p_2, \dots, p_r$ sont des probabilités données à l'avance. Alors on considère la satistique

$$E^2 = \sum_{i=1}^{r} \frac{(N_i - np_i)^2}{np_i}$$

qui mesure l'écart relatif enter les effectifs observés N_i et les effectifs moyens np_i appelés aussi effectifs attendus (de l'anglais " expected") si H_0 est vraie. On peut démontrer que, si H_0 est vraie, et pourvu que tous les np_i soient assez grands (supérieurs à 5), E^2 suit (approximativement) une loi du χ^2 à $(r - 1)$ dgrés de liberté (notés ddl).

2.2 Test d'ajustement du χ^2 pour une variable continue

Si l'on pose la question de savoir si une variable X suit ou non la loi normale $\mathcal{N}(0,1)$, on peut se ramener au problème précédent en discrétisant la variable : c'est-à-dire que l'on fait une partition finie de l'ensemble $\mathbb{R}$ de toutes les valeurs possibles de X formée de r intervalles successifs sans point commun :

$$]-\infty, a_1],\ ]a_1, a_2],\ \ldots,\ ]a_{r-1}, +\infty[.$$

Si l'on a observé un n-échantillon de valeurs de X soient $x_1, x_2, \ldots, x_n$, on résume ces observations en $(N_1, \ldots, N_r)$ où N_i désigne le nombre des x_i qui sont inférieurs à a_1, N_2 le nombre de ceux qui tombent entre a_1 (non compris) et a_2 (compris),...
Sous l'hypothèse

$$H_0 : X \rightsquigarrow \mathcal{N}(0,1),$$

les probabilités p_j pour que X tombe dans chacun des r intervalles $I_j =]a_{j-1}, a_j]$ peuvent être calculées :

$$p_j = \int_{a_{j-1}}^{a_j} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} dz.$$

Et on voit donc comment se ramener au problème du paragraphe précédent pour toute loi continue dont la densité est complètement spécifiée.

2.3 Rappels sur la distribution du χ^2

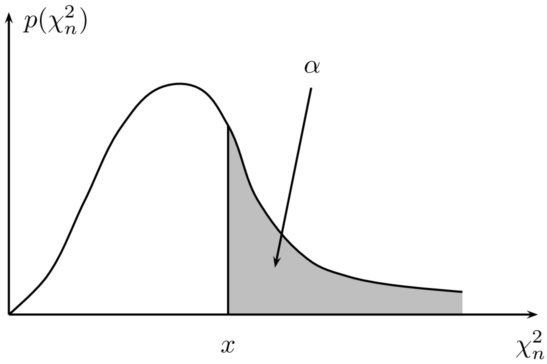
Definition 2.1 *On considère n variables indépendantes d'une loi normale centrée réduite : T_1, T_2,...,T_n. La quantité*

$$T_1^2 + T_2^2 + \ldots + T_n^2 = \sum_{i=1}^{n} T_i^2$$

est une variable aléatoire dont la distribution est celle d'un χ^2 à n degrés de liberté,

- *de moyenne $E(\chi_n^2) = n$*
- *de variance $V(\chi_n^2) = 2n$*

Graphiquement,



La variable χ^2 est tabulée en fonction du nombre n de degrés de liberté. La table (disponible à la fin de la fiche) donne pour différentes valeurs de α, la valeur de x telle que :

$$p(\chi_n^2 > x) = \alpha$$

Exemple 2.1 Calculer $p(\chi_{10}^2 > 20,5)$. On récupère à l'aide de la table, la probabilité $p(\chi_{10}^2 < 20,5) = 0,975$. Par conséquent, la probabilité recherchée $p(\chi_{10}^2 > 20,5)$ est égale à $1 - 0,975 = 0,025$.

On a la propriété suivante :

Proposition 2.1 *On a la relation*

$$\chi^2_m + \chi^2_n = \chi^2_{m+n}$$

Ce χ^2 *admet*

- *une moyenne* $E(\chi^2_{m+n}) = m + n$
- *une variance* $\sigma^2(\chi^2_{m+n}) = 2(m + n)$

et ceci par application directe du théorème sur l'addition de variables aléatoires indépendantes.

2.4 Ajustement d'une distribution observée à une distribution théorique

Étant donnée une population décrite par une variable X, un échantillon prélevé dans cette population permet de construire l'histogramme et la courbe des fréquences qui caractérisent la distribution observée de X. Ces représentations peuvent ressembler à celles d'une loi théorique, toutefois avec certains écarts.
Le test du χ^2 permet de juger si les écarts constatés entre la distribution observée et la loi théorique d'ajustement peuvent ou non être imputés au hasard.

Construction du test

1. Les hypothèses du test sont les suivantes :
   - H_0 : " X suit la loi théorique L",
   - H_1 : " X ne suit pas L".
2. La variable observée est :
   - soit discrète et prend k valeurs x_1, x_2,...,x_k
   - soit continue et classée en k classes $[a_0, a_1[$, $[a_1, a_2[$,..., $[a_{k-1}, a_k[$ de centres respectifs x_1, x_2,..., x_{k-1}, x_k.
3. Les N observations de l'échantillon sont réparties sur les k valeurs de X (si X est discrète) ou sur les k classes de X (si X est continue). On a les tableaux suivants :

x_i	n_i
x_1	n_1
x_2	n_2
$\vdots$	$\vdots$
x_k	n_k

Classes	Centres x_i	Effectifs n_i
$[a_0, a_1[$	x_1	n_1
$[a_1, a_2[$	x_2	n_2
$\vdots$	$\vdots$	$\vdots$
$[a_{k-1}, a_k[$	x_k	n_k

   avec $N = \sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k$.

4. Sous H_0 on note p_i la probabilité dite théorique définie par
   - $p_i = p(X = x_i / X \rightsquigarrow L)$ si X est discrète,
   - $p_i = p(X \in [a_{i-1}, a_i[/ X \rightsquigarrow L)$ si X est continue.

   $e_i = Np_i$ est l'**effectif théorique** de la i-ième classe de X.

5. L'indicateur d'écart entre les distributions observées et théoriques est

$$\chi^2_{obs} = \sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i} \tag{1}$$

   dit χ^2 **observé** ou calculé. Cet écart suit pour N suffisamment grand une loi du χ^2_ν d'où le nom du test.

Intuitivement, on comprend que cette grandeur statistique traduise l'écart entre l'échantillon et la loi conjecturée.
Si l'ajustement était parfait, cette expression du χ^2 serait nulle, les effectifs empiriques coïncidant exactement avec les effectifs théoriques.
En revanche, plus grands sont les écarts entre les effectifs observés et les effectifs théoriques $(n_i - e_i)$ et plus forte sera la valeur du χ^2.

En outre, comme la quantité (1) ne peut pas être négative, le test d'ajustement est nécessairement un test unilatéral droit.

Definition 2.2 *Le paramètre ν indiçant χ^2_ν définit le* ***nombre de degrés de liberté****. C'est le nom donné au nombre d'observations linéairement indépendantes qui apparaissent dans une somme de carrés. Autrement dit, c'est le nombre d'observations aléatoires indépendantes moins le nombre de contraintes imposées à ces observations. Le nombre ν de degrés de liberté est égal à*

- *$\nu = k - 1$ si les paramètres de la loi d'ajustement L sont donnés. Aucun paramètre n'est à estimer puisque la loi d'ajustement est totalement spécifiée. Le χ^2 est constitué de k écarts $(n_i - e_i)$ reliés par la contrainte $\sum(n_i - e_i) = \sum(n_i - Np_i) = \sum n_i - N \sum p_i = N - N = 0$ (en d'autres termes, lorsqu'on connaît la valeur de $k-1$ écarts, on peut en déduire la valeur du dernier qui n'est donc pas " libre" de varier de manière aléatoire).*
- *$\nu = k - r - 1$ si la loi d'ajustement L comporte r paramètres inconnus. On impose de ce fait autant de contraintes supplémentaires entre les observations, diminuant d'autant le nombre de degrés de liberté.*

Remarque 2.1 Le nombre d'observations par classes ne doit pas être faible, Np_i doit être supérieur à 5, $\forall i = 1, 2, \dots, k$. Dans le cas contraire, on regroupe deux ou plusieurs classes adjacentes de façon à réaliser cette condition. On tient compte de ce regroupement pour le nombre de degrés de liberté.

6. Pour un risque de première espèce α, la région critique est définie pour

$$\sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i} \geq \chi^2_{\nu,1-\alpha}$$

d'où la règle de décision :

- $\chi^2_{obs} < \chi^2_{\nu,1-\alpha}$, on décide H_0 et $X \rightsquigarrow L$,
- $\chi^2_{obs} \geq \chi^2_{\nu,1-\alpha}$, on décide H_1 et X ne suit pas la loi L.

2.5 Exemple : loi uniforme

Une statistique relative aux résultats du concours d'entrée à une grande école fait ressortir les répartitions des candidats et des admis selon la profession des parents.

Profession des candidats	Nombre de candidats	Nombre d'admis
1 Fonctionnaires et assimilés	2244	180
2 Commerce, industrie	988	89
3 Professions libérales	575	48
4 Propriétaires rentiers	423	37
5 Propriétaires agricoles	287	13
6 Artisans, petits commerçants	210	18
7 Banque, assurance	209	17
Total	4936	402

Problème : Tester l'hypothèse (risque $\alpha = 0,05$) selon laquelle la profession des parents n'a pas d'influence sur l'accès à cette grande école.

Il s'agit du test d'ajustement d'une distribution théorique, on pose les hypothèses

- H_0 : " la profession des parents n'a pas d'influence sur l'accès à cette grande école", la proportion des admis est constante pour toutes les professions soit $p = \dfrac{402}{4936} \simeq 0,0814$
- H_1 : " la profession des parents influe sur l'accès à cette grande école".

Sous H_0, le nombre d'admis pour la i-ième profession est $N_i p$.

i	N_i	n_i effectif observé	$N_i p$ effectif théorique	$\dfrac{(n_i - N_i p)^2}{N_i p}$
1	2244	180	$\dfrac{2244 \times 402}{4936} \simeq 182,76$	$0,0416$
2	988	89	$\dfrac{988 \times 402}{4936} \simeq 80,47$	$0,9042$
3	575	48	$\dfrac{575 \times 402}{4936} \simeq 46,83$	$0,0293$
4	423	37	$\dfrac{423 \times 402}{4936} \simeq 34,45$	$0,1887$
5	287	13	$\dfrac{287 \times 402}{4936} \simeq 23,37$	$4,6050$
6	210	18	$\dfrac{210 \times 402}{4936} \simeq 17,10$	$0,0471$
7	209	17	$\dfrac{209 \times 402}{4936} \simeq 17,02$	$\simeq 0$
Total	4936	402	402	$5,8181$

Le χ^2 observé vaut $5,8181$. Le nombre de degrés de liberté est $7 - 1 = 6$. La table fournit $\chi^2_{6;0,95} = 12,59$ donc χ^2 observé $< \chi^2_{6;0,95}$. On ne rejette pas H_0, ce qui signifie que la profession des parents n'a pas d'influence sur l'accès à cette grande école.

2.6 Exemple : loi binomiale

Supposons qu'on ait recueilli 300 boîtes contenant chacune trois ampoules. Dans chaque boîte, on compte le nombre d'ampoules défectueuses. On obtient les résultats suivants :

Nombre d'ampoules défectueuses x_i	Nombre de boîtes observées n_i
0	190
1	95
2	10
3	5
Total	300

Pour chaque ampoule testée, on peut observer deux états différents : l'ampoule est défectueuse ou non. Le nombre X d'ampoules défectueuses par boîte suit une loi binomiale de paramètres $n = 3$ et p. Déterminons p. Dans la distribution observée, le nombre d'ampoules défectueuses est de

$$0 \times 190 + 1 \times 95 + 2 \times 10 + 3 \times 5 = 130$$

soit 130 ampoules défectueuses sur un total de 900 ampoules. La proportion d'ampoules défectueuses est alors de $\dfrac{130}{900} \simeq 0,144$. Prenons $p = 0,15$.

Problème : Tester l'hypothèse (au risque $\alpha = 0,01$) selon laquelle le nombre d'ampoules défectueuses par boîte suit une loi binomiale de paramètres $n = 3$ et $p = 0,15$.

On considère donc les hypothèses suivantes :

- H_0 : " $X \rightsquigarrow \mathcal{B}(3; 0,15)$"
- H_1 : " X ne suit pas cette loi binomiale"

et on détermine ensuite les probabilités théoriques :

- $p_0 = p\{X = 0/X \rightsquigarrow \mathcal{B}\} = (0,85)^3 \simeq 0,6141$
- $p_1 = p\{X = 1/X \rightsquigarrow \mathcal{B}\} = C_3^1(0,15)(0,85)^2 \simeq 0,3251$
- $p_2 = p\{X = 2/X \rightsquigarrow \mathcal{B}\} = C_3^2(0,15)^2(0,85) \simeq 0,0574$
- $p_3 = p\{X = 3/X \rightsquigarrow \mathcal{B}\} = (0,15)^3 \simeq 0,0034$

On a le tableau (provisoire) suivant :

x_i	effectif observé n_i	p_i	effectif théorique Np_i
0	190	$0,6141$	$184,23$
1	95	$0,3251$	$97,53$
2	10	$0,0574$	$17,22$
3	5	$0,0034$	$1,02$
Total	$N = 300$	1	300

L'effectif théorique de la quatrième classe est faible, en effet $1,02 < 5$! On effectue un regroupement de classes, les classes 2 et 3.

x_i	n_i	Np_i	$\frac{(n_i - Np_i)^2}{Np_i}$
0	190	$184,23$	$0,18071$
1	95	$97,53$	$0,06563$
2 ou 3	15	$18,24$	$0,57553$
Total	300	300	$0,82187$

Après le regroupement, le nombre de classes est 3, le nombre de degrés de liberté est $3 - 1 = 2$. Au risque $\alpha = 0.01$, $\chi^2_{2;0,99} = 9,21$. Donc $\chi^2_{obs} = 0,82187 < \chi^2_{2;0,99}$. On ne rejette pas H_0 au profit de H_1. On considère que le nombre d'ampoules défectueuses par boîte suit une loi binomiale de paramètre $n = 3$, $p = 0,15$ au risque $\alpha = 0,01$

2.7 Exemple : loi normale

On suppose que le rendement X (quintaux par hectares d'une parcelle de blé) suit une loi normale $\mathcal{N}(m, \sigma)$. L'observation du rendement de 1000 parcelles a donné les résultats suivants :

Rendement	Nombre de parcelles
$[0, 10[$	5
$[10, 20[$	6
$[20, 30[$	40
$[30, 40[$	168
$[40, 50[$	288
$[50, 60[$	277
$[60, 70[$	165
$[70, 80[$	49
$[80, 90[$	2
Total	1000

Afin de mettre en place un test d'ajustement, déterminons dans un premier temps la moyenne arithmétique et l'écart-type de la distribution observée :

- $\overline{x} = \dfrac{\sum_i n_i x_i}{N} = 49,76$

- $\sigma'^2 = \dfrac{\sum_i n_i x_i^2}{N} - \overline{x}^2 = 164,5424$ donc $\sigma' \simeq 12,827$

Problème : Tester l'hypothèse (risque $\alpha = 0,05$) selon laquelle l'ajustement de la distribution observée à une loi normale $\mathcal{N}(n = 50, \sigma = 13)$ est acceptable.

Les hypothèses du test du χ^2 sont les suivantes :

- H_0 : " $X \rightsquigarrow \mathcal{N}(50, 13)$"
- H_1 : " X ne suit pas $\mathcal{N}(50, 13)$"

On désigne par $[a_0, a_1[$, $[a_1, a_2[$,...,$[a_8, a_9[$ les classes et par x_1, x_2,...,x_9 les centres de ces classes. Sous H_0, $X \rightsquigarrow \mathcal{N}(50, 13)$ et $Z = \dfrac{X - 50}{13} \rightsquigarrow \mathcal{N}(0, 1)$, donc $p_i = p(X \in [a_{i-1}, a_i[) = \Pi(z_i) - \Pi(z_{i-1})$ avec $z_i = \dfrac{a_i - 50}{13}$ et $z_{i-1} = \dfrac{a_{i-1} - 50}{13}$. L'effectif théorique de la ième classe est $1000p_i$ et $\displaystyle\sum_i \frac{(n_i - Np_i)^2}{Np_i} \rightsquigarrow \chi^2_\nu$.

On a le tableau suivant :

Classe $[x_{i-1}, x_i[$	n_i	z_i	$\Pi(z_i)$	p_i	Np_i	Np_i corrigé	n_i corrigé	$\dfrac{(n_i - Np_i)^2}{Np_i}$
$[0, 10[$	5	$-3,0769$	$0,001$	$0,0009$	$0,9$	$10,4$	11	$0,0346$
$[10, 20[$	6	$-2,3077$	$0,0105$	$0,0095$	$9,5$			
$[20, 30[$	40	$-1,5385$	$0,0620$	$0,0515$	$51,5$	$51,5$	40	$2,568$
$[30, 40[$	168	$-0,7692$	$0,2209$	$0,1589$	$158,9$	$158,9$	168	$0,5211$
$[40, 50[$	288	0	$0,5$	$0,2791$	$279,1$	$279,1$	288	$0,283$
$[50, 60[$	277	$0,7692$	$0,7791$	$0,2791$	$279,1$	$279,1$	277	$0,0158$
$[60, 70[$	165	$1,5385$	$0,9380$	$0,1589$	$158,9$	$158,9$	165	$0,234$
$[70, 80[$	49	$2,3077$	$0,9895$	$0,0515$	$51,5$	$51,5$	49	$0,1214$
$[80, 90[$	2	$3,0769$	$0,9990$	$0,0095$	$9,5$	$9,5$	2	$5,9211$
Total	1000	–	–	1	1000	1000	1000	$9,7$

On effectue le regroupement des deux premières classes car $Np_i < 5$. Le χ^2 observé vaut $9,7$. Après le regroupement, il reste 8 classes, les deux paramètres de la loi normale sont donnés, le nombre de degrés de liberté est $\nu = 8 - 1 = 7$. À l'aide de la table, on obtient $\chi^2_{7;0,95} = 14,07$. Ainsi,

$$\chi^2_{obs} < \chi^2_{7;0,95}.$$

On ne rejette pas H_0, l'ajustement de la distribution observée à une loi normale $\mathcal{N}(50, 13)$ est acceptable.

2.8 Exemple : loi de Poisson

Souvent, lorsqu'on envisage une modèle pour un phénomène qu'on étudie, on ne spécifie pas complètement la loi qu'on considère. Supposons qu'on s'intéresse au nombre de voitures se présentant par minute à un poste de péage sur une autoroute. On peut se demander si cette variable aléatoire peut être modélisée par une loi de Poisson. On souhaite donc tester l'hypothèse fondamentale H_0 : " $X \rightsquigarrow \mathcal{P}(\lambda)$" contre l'hypothèse alternative H_1 : " X ne suit pas $\mathcal{P}(\lambda)$". On ne précise pas la valeur du paramètre λ. On peut toutefois l'estimer à partir des données disponibles mais dans ce cas, $r = 1$. Le nombre de degrés sera alors $\nu = k - r - 1 = k - 2$.

On effectue 200 comptages au péage.

x_i	0	1	2	3	4	5	6	7	8	≥ 9	Total
n_i	6	15	40	42	37	30	10	12	8	0	200
$n_i x_i$	0	15	80	126	148	150	60	84	64	0	727

où x_i et n_i désignent respectivement le nombre de voitures par minute et l'effectif correspondant lors de l'observation n°i (par exemple, $x_1 = 0$ et $n_1 = 6$ c'est-à-dire que lors de 6 observations, il y a 0 voiture). La moyenne arithmétique de cette distribution observée est

$$\overline{x} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{727}{200} = 3,635 \simeq 3,5$$

Problème : Tester l'hypothèse (au risque $\alpha = 0,01$) selon laquelle X suit une loi de Poisson de paramètre $3,5$.

On pose

- H_0 : " $X \rightsquigarrow \mathcal{P}(3,5)$"
- H_1 : " X ne suit pas $\mathcal{P}(3,5)$"

Sous H_0, $p_i = p(X = i) = e^{-3,5}\dfrac{3,5^i}{i!}$, on a donc le tableau de valeurs suivant :

x_i	n_i	p_i	Np_i	Np_i corrigé	n_i corrigé	$\dfrac{(n_i - Np_i)^2}{Np_i}$
0	6	$0,0302$	$6,04$	$6,04$	6	$0,00026$
1	15	$0,1057$	$21,14$	$21,14$	15	$1,78333$
2	40	$0,1850$	37	37	40	$0,24324$
3	42	$0,2158$	$43,16$	$43,16$	42	$0,03118$
4	37	$0,1888$	$37,76$	$37,76$	37	$0,01530$
5	30	$0,1322$	$26,44$	$26,44$	30	$0,47933$
6	10	$0,0771$	$15,42$	$15,42$	10	$1,90508$
7	12	$0,0385$	$7,7$	$7,7$	12	$2,40130$
8	8	$0,0169$	$3,38$	$5,34$	8	$1,32502$
≥ 9	0	$0,0098$	$1,96$			
Total	200	1	200	200	200	$8,18404$

On a effectué le regroupement des deux dernières classes car l'effectif théorique y est inférieur à 5.
Après ce regroupement, le nombre de classes est de 9. Le nombre de degrés de liberté est $9 - 1 - 1 = 7$. Au risque $\alpha = 0,01$, $\chi^2_{7;0,99} = 18,48$ donc $\chi^2_{obs} = 8,18404 < \chi^2_{7;0,99}$. On ne rejette pas l'hypothèse H_0 et $X \rightsquigarrow \mathcal{P}(\lambda = 3,5)$ au risque $\alpha = 0,01$.

2.9 Exercices

Exercice 1 En lançant successivement 60 fois un dé, un joueur obtient les résultats suivants :

Faces x_i	1	2	3	4	5	6
Effectifs n_i	15	7	4	11	6	17

Le dé est-il truqué?

Exercice 2 On a enregistré le nombre X de clients entrant dans un magasin en 1 minute. On a obtenu le tableau suivant :

Nombre de clients x_i	Nombre de minutes où $X = x_i$ (où il est entré x_i clients)
0	23
1	75
2	68
3	51
4	30
5	10
plus de 5	7

Peut-on admettre que les arrivées sont régies par une loi de Poisson de paramètre $m = 2$ (au seuil $\alpha = 0,05$)?

Exercice 3 Une enquête sur les chiffres d'affaires mensuels de 103 magasins de détail a donné les résultats suivants (en milliers d'euros) :

Classes de chiffres d'affaires	Centres de classes	Nombre de magasins
$5,5$ à moins de $6,5$	6	2
$6,5$ à moins de $7,5$	7	3
$7,5$ à moins de $8,5$	8	12
$8,5$ à moins de $9,5$	9	27
$9,5$ à moins de $10,5$	10	23
$10,5$ à moins de $11,5$	11	15
$11,5$ à moins de $12,5$	12	12
$12,5$ à moins de $13,5$	13	5
$13,5$ à moins de $14,5$	14	2
$14,5$ à moins de $15,5$	15	2

Peut-on considérer que l'échantillon est tiré d'une loi normale?

Exercice 4 On a étudié le nombre de garçons dans 1883 familles de 7 enfants. Les résultats sont présentés en fonction du nombre x_i de garçons, rangés de 0 à 7 :

Nombre de garçons x_i	Effectif des familles n_i
0	27
1	111
2	287
3	480
4	529
5	304
6	126
7	19

Peut-on admettre au seuil de 5% que le nombre x_i de garçons par famille obéisse à une loi binomiale? Laquelle?

Exercice 5 Une étude est menée dans une petite université sur l'absentéisme des étudiants. On aimerait savoir si certaines plages horaires sont plus propices à une absence aux cours qu'une autre. Pour cela, on a relevé, au cours d'un mois, le nombre d'absences d'étudiants aux cours d'une petite composante à différents moments de la journée. Les résultats sont les suivants :

Heures de la journée	Nombre d'étudiants absents
$[8, 10]$	25
$[10, 12]$	15
$[13, 15]$	18
$[15, 17]$	32

En considérant cet échantillon tiré au hasard, peut-on dire, au seuil de 5%, que les absences des étudiants aux cours se répartissent uniformément tout au long de la journée?

Exercice 6 Au départ d'une course de chevaux, il y a habituellement huit positions de départ, la position numéro 1 est la plus proche de la palissade. On soupçonne qu'un cheval a plus de chances de gagner quand il porte une numéro faible, c'est-à-dire lorsqu'il est plus proche de la palissade. Voici les données de 144 courses :

Numéro de départ	1	2	3	4	5	6	7	8
Nombre de victoires d'un cheval ayant ce numéro	29	19	18	25	17	10	15	11

La comparaison de la distribution observée à la distribution théorique s'effectue par un test du χ^2. Que peut-on en conclure?

Exercice 7 Le nombre d'accidents mensuels à un certain carrefour est une variable aléatoire X. On observe X durant 32 mois :

Nombre d'accidents	0	1	2	3	4	5
Nombre de mois	2	13	8	4	4	1

Peut-on admettre au seuil de 10% que X suit une loi de Poisson de paramètre 2?

Exercice 8 Une entreprise fabriquant des produits alimentaires sucrés veut élargir sa gamme de barres de céréales en en lançant une nouvelle sur le marché. Le directeur du marketing décide de faire une enquête de goût en faisant tester ce nouveau produit à 200 personnes. Le test a lieu en aveugle, et les personnes ont donc à se prononcer sur leur préférence concernant la nouvelle barre et quatre autres barres de céréales concurrentes. Les produits étant appelés A (la nouvelle barre), B, C, D et E, les résultats du test dont les suivants :

Barres	A	B	C	D	E
Nombre de préférences	40	35	55	40	30

Au seuil 5%, peut-on dire au vu des résultats d'échantillon que la nouvelle barre a meilleur goût que les autres?

Exercice 9 Un responsable qualité d'une entreprise fabriquant de l'appareillage électronique a mesuré la durée de vie de 60 dispositifs électroniques d'un même type. Il a obtenu les résultats suivants :

Durée de vie (en heures)	Nombre de dispositifs
$[250; 270[$	3
$[270; 290[$	5
$[290; 310[$	15
$[310; 330[$	22
$[330; 350[$	13
$[350; 370[$	2

Les donnée permettent-elles, au seuil de 5%, de penser que la durée de vie d'un dispositif électronique de ce type est distribué selon une loi normale?

Exercice 10 On a observé le nombre de défauts de pièces de tissu traitées par un teinturier. Les résultats de 50 observations sont reproduits dans le tableau ci-dessous (par exemple, 8 des 50 pièces présentaient 3 défauts). Des études antérieures avaient permis de faire l'hypothèse que le nombre de défauts par pièce pouvait être considéré comme une variable aléatoire X obéissant à une loi de Poisson. Les observations permettent-elles de confirmer cette hypothèse, au seuil de 5%?

x_i	0	1	2	3	4	5
n_i	6	14	16	8	4	2

ANNEXE - Quantiles de la loi du χ^2_ν

La table donne les valeurs (quantiles) $\chi^2_{\nu,1-\alpha}$ telles que $P(\chi^2_\nu < \chi^2_{\nu,1-\alpha}) = 1-\alpha$.

ν	$1-\alpha$									
	$0,005$	$0,010$	$0,025$	$0,050$	$0,100$	$0,900$	$0,950$	$0,975$	$0,990$	$0,995$
1	$0,0000393$	$0,000157$	$0,000982$	$0,00393$	$0,0158$	$2,71$	$3,84$	$5,02$	$6,63$	$7,88$
2	$0,0100$	$0,0201$	$0,0506$	$0,103$	$0,211$	$4,61$	$5,99$	$7,38$	$9,21$	$10,60$
3	$0,072$	$0,115$	$0,216$	$0,352$	$0,584$	$6,25$	$7,81$	$9,35$	$11,34$	$12,84$
4	$0,207$	$0,297$	$0,484$	$0,711$	$1,064$	$7,78$	$9,49$	$11,14$	$13,28$	$14,86$
5	$0,412$	$0,554$	$0,831$	$1,145$	$1,61$	$9,24$	$11,07$	$12,83$	$15,09$	$16,75$
6	$0,676$	$0,872$	$1,24$	$1,64$	$2,20$	$10,64$	$12,59$	$14,45$	$16,81$	$18,55$
7	$0,989$	$1,24$	$1,69$	$2,17$	$2,83$	$12,02$	$14,07$	$16,01$	$18,48$	$20,28$
8	$1,34$	$1,65$	$2,18$	$2,73$	$3,49$	$13,36$	$15,51$	$17,53$	$20,09$	$21,96$
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	$11,81$	$12,88$	$14,57$	$16,15$	$18,11$	$36,74$	$40,11$	$43,19$	$46,96$	$49,64$
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	$13,79$	$14,95$	$16,79$	$18,49$	$20,60$	$40,26$	$43,77$	$46,98$	$50,89$	$53,67$
40	$20,71$	$22,16$	$24,43$	$26,51$	$29,05$	$51,81$	$55,76$	$59,34$	$63,69$	$66,77$
50	$27,99$	$29,71$	$32,36$	$34,76$	$37,69$	$63,17$	$67,50$	$71,42$	$76,15$	$79,49$
60	$35,53$	$37,48$	$40,48$	$43,19$	$46,46$	$74,40$	$79,08$	$83,30$	$88,38$	$91,95$
70	$43,28$	$45,44$	$48,76$	$51,74$	$55,33$	$85,53$	$90,53$	$95,02$	$100,4$	$104,2$
80	$51,17$	$53,54$	$57,15$	$60,39$	$64,28$	$96,58$	$101,9$	$106,6$	$112,3$	$116,3$
90	$59,20$	$61,75$	$65,65$	$69,13$	$73,29$	$107,6$	$113,1$	$118,1$	$124,1$	$128,3$
100	$67,33$	$70,06$	$74,22$	$77,93$	$82,36$	$118,5$	$124,3$	$129,6$	$135,8$	$140,2$