Corrigé du TD 7
Mars 2011

Exercice 1. On souhaite vérifier la qualité du générateur de nombres aléatoires d'une calculatrice scientifique. Pour cela, on procède à 250 tirages dans l'ensemble $\{0, \dots, 9\}$ et on obtient les résultats suivants :

x	0	1	2	3	4	5	6	7	8	9
$N(x)$	28	32	23	26	23	31	18	19	19	31

A l'aide du test du χ^2, vérifier si le générateur produit des entiers indépendants et uniformément répartis sur $\{0, \dots, 9\}$.

On applique le cours. On doit ici faire un test du χ^2 pour vérifier si l'échantillon $X_1, \dots, X_n$ suit la loi uniforme sur $\{0, \dots, 9\}$. Il y a quatre étapes à suivre :

1. On calcule la statistique de test qui mesure la distance entre la loi réalisée et la loi théorique.

$$D_n = n \sum_{i=0}^{9} \frac{(\hat{p}_i - p_i)^2}{p_i}$$

où $\hat{p}_i$ représentent les fréquences empiriques observées et $p_i = \frac{1}{10}$ les fréquences théoriques.

2. D'après cours, la statistique de test D_n convergent en loi vers un χ^2 de degrés de liberté $10 - 1 = 9$. On peut assimiler D_n à v.a. suivant une loi du χ^2 si $np_i > 5,\ \forall i\ (np_i = 25 > 5)$ et si $n \geqslant 50$.

3. Pour un niveau α, la région de rejet est :

$$W = \{D_n > F^{-1}_{\chi^2_9}(1 - \alpha)\}$$

où $F^{-1}_{\chi^2_9}$ est l'inverse de la fonction de répartition de la loi du χ^2_9.

Applications Numériques :

x	0	1	2	3	4	5	6	7	8	9
$N(x)$	28	32	23	26	23	31	18	19	19	31
$\hat{p}_i$	$\frac{28}{250}$	$\frac{32}{250}$	$\frac{23}{250}$	$\frac{26}{250}$	$\frac{23}{250}$	$\frac{31}{250}$	$\frac{18}{250}$	$\frac{19}{250}$	$\frac{19}{250}$	$\frac{31}{250}$
p_i	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$	$0,1$
$n\frac{(\hat{p}_i - p_i)^2}{p_i}$	$0,36$	$1,96$	$0,16$	$0,04$	$0,16$	$1,44$	$1,96$	$1,44$	$1,44$	$1,44$

La somme de la dernière ligne donne $D_{250} = 10,4$
D'après le tableau des fractiles de la loi du χ^2, $F^{-1}_{\chi^2_9}(1 - \alpha) = 16,92$. Comme $D_{250} < F^{-1}_{\chi^2_9}(1 - \alpha)$, on accepte l'hypothèse H_0.

Exercice 2. Pour déterminer si les merles vivent en communauté ou en solitaire, on procède à l'expérience suivante : on dispose un filet dans la zone d'habitat des merles, et on vient relever le nombre de captures pendant 89 jours. On obtient les résultats suivants :

Nombre de captures	0	1	2	3	4	5	6
Nombre de jours	56	22	9	1	0	1	0

1. On suppose qu'une loi de Poisson est représentative de l'expérience. Construire un estimateur du paramètre de cette loi.

La vraisemblance du modèle de Poisson est :

$$\begin{aligned}
L(x,\lambda) &= e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^n x_i}}{\Pi_{i=1}^n x_i!} \\
\ln L(x,\lambda) &= -n\lambda + \ln(\lambda)\sum_{i=1}^n x_i + \ln(\Pi_{i=1}^n x_i!) \\
\frac{\partial \ln L(x,\lambda)}{\partial \lambda} &= -n + \frac{\sum_{i=1}^n x_i}{\lambda}
\end{aligned}$$

qui s'annule quand $\lambda = \frac{1}{n}\sum_{i=1}^n x_i$. On vérifie que la dérivée seconde est négative en cette valeur.

On obtient l'EMV : $\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^n X_i$

2. Vérifier à l'aide d'un test du χ^2 l'adéquation du modèle aux données. Faire l'application numérique au niveau $\alpha = 5\%$.

On calcule la statistique de test :

$$D_n = n\sum_{i=0}^{6}\frac{(\hat{p}_i - p_i(\hat{\lambda}_n))^2}{p_i}$$

où $\hat{p}_i$ représentent les fréquences empiriques observées et $p_i(\hat{\lambda}_n) = e^{-\hat{\lambda}_n}\frac{\hat{\lambda}_n^j}{j!}$ les fréquences théoriques. D_n convergent en loi vers un χ^2_{7-1-1}. Le test est défini par la région critique :

$$W = \{D_n > F^{-1}_{\chi^2_5}(1-\alpha)\}$$

où $F^{-1}_{\chi^2_5}$ est l'inverse de la fonction de répartition de la loi du χ^2_5.

A.N. On obtient $\hat{\lambda}_n = 0,539$ et $D_n = 50,9$ et $F^{-1}_{\chi^2_5}(0,95) = 11,07 < D_n$. Donc, on rejette H_0.

Remarque : Cette conclusion est erronnée car la condition requise $np_i > 5$ pour assimiler la loi de D_n à un χ^2 n'est pas remplie. Cela est du au choix inadéquate de classes.

3. Reprendre l'exercice en groupant les catégories Nombre de captures $= 2,3,4,5$ et 6 en Nombre de captures ≥ 2.

On refait le travail avec cette fois 3 classes.

Nombre de captures	0	1	≥ 2
Nombre de jours	56	22	11

La statistique de test :

$$D_n = n \sum_{i=0}^{2} \frac{(\hat{p}_i - p_i(\hat{\lambda}_n))^2}{p_i}$$

On vérifie que $np_i > 5$. D_n suit la loi du χ^2 (à un degré de liberté). La région de rejet devient :

$$W = \{D_n > F_{\chi^2}^{-1}(1-\alpha)\}$$

A.N. $D_n = 2,00 < F_{\chi^2}^{-1}(0,95) = 3,84$. On accepte donc H_0.