Partiel TISD - Master Pro
Vendredi 12 novembre 2010

Durée 2 heures. Calculatrices autorisées. Documents de cours autorisés.
Le barème est donné uniquement à titre indicatif et pourra être changé lors de la correction.

Tran Viet Chi, `chi.tran@univ-lille1.fr`, bureau 316 (bâtiment M3).

Exercice 1 (Naissances, 9 pts)
On désire étudier la répartition des naissances suivant le type du jour de semaine (jours ouvrables ou week-end) et suivant le mode d'accouchement (naturel ou par césarienne). Les données proviennent du *National Vital Statistics Report* et concernent les naissances aux USA en 1997.

Naissances	Naturelles	Césarienne
Jour ouvrable	2331536	663540
W.E.	715085	135493

Table 1 – *Naissances en 1997 aux USA suivant le type de jour et la procédure d'accouchement*

On note $p_{J,N}$ (resp. $p_{W,N}$, $p_{J,C}$ et $p_{W,C}$) la probabilité qu'un bébé naisse un jour ouvrable et sans césarienne (resp. un week-end et sans césarienne, un jour ouvrable et par césarienne, un week-end et par césarienne). On note également $x_{J,N}$ (resp. $x_{W,N}$, $x_{J,C}$, $x_{W,C}$) le nombre de naissances naturelles ayant eu lieu un jour ouvrable (resp. naturelle ayant eu lieu le W.E., avec césarienne un jour ouvrable et avec césarienne le W.E.).

1. Calculer le nombre de naissances n.

2. Calculer les distributions marginales à partir de la Table 1.

3. Calculer la distribution conditionnelle du type de jour de naissance, sachant que la naissance a eu lieu par césarienne. Commenter en utilisant le résultat de la question **2**.

4. La loi de $(x_{J,N}, x_{W,N}, x_{J,C}, x_{W,C})$ est une loi usuelle. Laquelle ?

5. Ecrire la vraisemblance des observations.

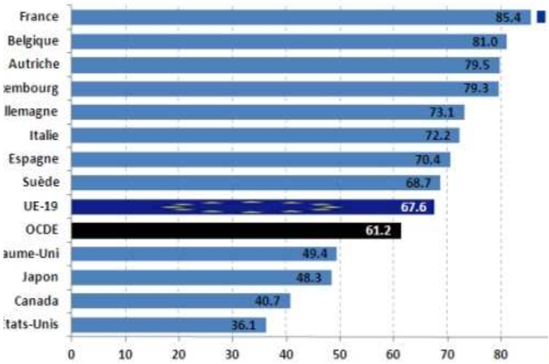
6. Calculer l'estimateur du maximum de vraisemblance de $(p_{J,N}, p_{W,N}, p_{J,C}, p_{W,C})$.

7. Donner un intervalle de confiance pour $p_{J,N}$. Faire l'application numérique.

8. A-t-on indépendance entre le type du jour de naissance (jour ouvrable ou week-end) et le mode d'accouchement (naturel ou césarienne) ? Pour répondre à cette question, on précisera l'hypothèse nulle du test, on calculera la statistique de test ξ, on précisera son comportement sous H_0 et sous H_1. Puis, on concluera sachant que **R** nous donne : `1-pchisq(,1)=0`.

Exercice 2 (Retraites, 8 points)
En coopération avec l'Union Européenne et la Banque mondiale, et en réponse à une forte demande d'indicateurs internationaux sur les retraites, l'OCDE a élaboré une base de données en ligne sur les caractéristiques actuelles des pensions et des politiques de retraite, avec des indicateurs pour tous les pays de l'OCDE/l'UE. Nous nous intéressons ici à des données de 2005 sur l'âge de départ à la retraite, à l'espérance de vie (60 pays de l'OCDE) et aux pourcentage des transferts publics dans les revenus des personnes âgées de plus de 65 ans (29 pays de l'OCDE).

(a)



(b)

	Hommes		Femmes	
	âge effectif	âge légal	âge effectif	âge légal
Autriche	58.9	65	57.9	60
Belgique	59.6	65	58.3	65
Canada	63.3	65	61.9	65
France	**58.7**	**60**	**59.5**	**60**
Allemagne	62.1	65	61.0	65
Italie	60.8	65	60.8	60
Japon	69.5	64	66.5	62
Luxembourg	59.2	65	60.3	65
Espagne	61.4	65	63.1	65
Suède	65.7	65	62.9	65
Suisse	65.2	65	64.1	64
Royaume-Uni	63.2	65	61.9	60
États-Unis	64.6	65.8	63.9	65.8
OCDE-30	63.5	64.4	62.3	63.1
UE-19	61.7	64.3	60.5	62.7

Source: Estimations de l'OCDE calculée à partir des enquêtes nationales et Européenne de la population active.

FIGURE 1 – *(a) Pourcentage des transferts publics dans les revenus des personnes âgées de plus de 65 ans. (b) Ages de départ à la retraite (observés) et légaux, par sexe.*

La variable correspondant à l'âge de départ à la retraite pour les hommes (resp. pour les femmes) est `sortie_h` (resp. `sortie_f`). L'espérance de vie des hommes est `espvie_h`. Le pourcentage des transferts publics dans les revenus des plus de 65 ans est la variable `pub`. Les sorties des analyses statistiques réalisées avec **SAS** se trouvent en annexe : Annexe 1 - étude de `sortie_h` ; Annexe 2 - étude de `sortie_f` ; Annexe 3 - étude de `y` (Question 4) ; Annexe 4 - corrélation de `sortie_h` et `espvie_h` ; Annexe 5 - étude de `pub`.

1. Donner l'âge moyen et la variance de l'âge de départ à la retraite, par sexe. Lire sur la Table 1 l'âge de départ à la retraite en France.

2. Les hommes partent-ils à la retraite plus tard que les femmes ? Faire une analyse de la variance. On donne pour simplifier le même poids à chaque pays.

3. A partir des tests réalisés par SAS, pensez-vous que la loi log-normale peut bien modéliser l'âge de départ à la retraite pour les hommes ? et la loi gamma ?

4. On sait que la `proc univariate` de SAS teste si la moyenne de la variable peut-être considérée comme nulle ou non. Pour tester si l'âge de départ à la retraite des hommes en France est plus bas ou non que pour les autres pays, on réalise donc une `proc univariate` sur la variable `y=sortie_h-60`. Expliquer pourquoi et interpréter les résultats de SAS.

5. On se demande si l'âge de départ à la retraite est lié à l'espérance de vie. Commenter les résultats de SAS, en particulier le test de l'Annexe 4.

6. Donner le pourcentage moyen de transferts publics dans les revenus des plus de 65 ans. La distribution de ce pourcentage est-elle symétrique ? plus étalée qu'une normale ?

7. D'après la Table 1, ce pourcentage est de 85,4373% pour la France. Quelle proportion de pays investissant moins dans les retraites des plus de 65 ans.

Exercice 3 (Histogramme, 5 points)
On considère un échantillon de v.a. i.i.d. $X_1, \ldots, X_n$ ayant une densité f sur $[0,1]$. Soit une partition

$$\left(C_j = \left[\frac{j-1}{m}, \frac{j}{m}\right[\right)_{j \in \{1,\ldots,m\}}$$

de $[0,1]$ et on approche f par l'histogramme associé à cette partition :

$$\widehat{f}(x) = m \sum_{j=1}^{m} f_j \mathbf{1}_{x \in C_j}, \qquad \text{où} \quad f_j = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{X_i \in C_j}.$$

On introduit aussi :

$$p_j = \mathbb{P}(X_1 \in C_j).$$

1. Vérifier que $\widehat{f}$ est bien une densité de probabilité.

2. Vers quoi converge $\widehat{f}$ lorsque $n \to +\infty$?

3. Pour mesurer le risque de $\widehat{f}$ on introduit le MISE (Mean Integrated Square Error) :

$$MISE(m) = \int_0^1 \mathbb{E}\left[\left(\widehat{f}(x) - f(x)\right)^2\right] dx.$$

3.1. Montrer que :

$$\mathbb{E}\left[\left(\widehat{f}(x) - f(x)\right)^2\right] = \left(\mathbb{E}\left[\widehat{f}(x)\right] - f(x)\right)^2 + \mathrm{Var}[\widehat{f}(x)]. \tag{1}$$

3.2. Pour x fixé, quelle est la loi de $\widehat{f}(x)$? En déduire son espérance et sa variance.

3.3. On considère l'intégrale du premier terme du membre de droite de (1) :

$$\int_0^1 \left(\mathbb{E}\left[\widehat{f}(x)\right] - f(x)\right)^2 dx$$

Exprimer ce terme en fonction de l'intégrale de f^2 et des p_j.

3.4. En déduire que :

$$MISE(m) = \int_0^1 f^2(x)dx + \frac{m}{n} - m\frac{n+1}{n} \sum_{j=1}^{m} p_j^2.$$

Commenter.

3.5. Si f est continue, justifier que si l'on fait n vers $+\infty$, puis m vers $+\infty$, alors le MISE tend vers 0.