Examen de Statistique - M1 IM - 9 janvier 2012

Viet Chi Tran, chi.tran@math.univ-lille1.fr
Emeline Schmisser, emeline.schmisser@math.univ-lille1.fr

Calculatrice et formulaire sur copie double autorisés. Durée : 3h.

Exercice 1 (Test sur le paramètre d'une loi exponentielle)
On considère l'échantillon $(X_1, X_2, \ldots, X_n)$ de n v.a. i.i.d. de loi exponentielle de paramètre $\theta^{-1} > 0$ (la densité est $f(x) = \theta^{-1} \exp(-x/\theta)\mathbf{1}_{x>0}$).

1. Ecrire le modèle. S'agit-il d'un modèle exponentiel ? Quelle est la statistique canonique ? Est-elle exhaustive ? totale ?
2. Quel est l'estimateur du maximum de vraisemblance de $\widehat{\theta}$?
3. Est-ce que cet estimateur est efficace ?
4. Est-ce qu'il est possible de construire un estimateur sans biais avec une variance plus petite ?
5. $\widehat{\theta}$ est-il un estimateur asymptotiquement normal ? Si oui, expliciter cette convergence en loi.

Test 1 : On veut tester $H_0 \ : \ \theta = 1$ contre $H_a \ : \ \theta \neq 1$, avec un test de Wald.

6. Ecrire la statistique de test. Quel est son comportement sous H_0 ? sous H_a ?
7. Donner la région critique pour que le test soit asymptotiquement de seuil $\alpha = 0,05$.

Test 2 : On veut tester $H_0 \ : \ \theta = 1$ contre $H_a \ : \ \theta < 1$, de façon non asymptotique.

8. Le modèle est-il à rapport de vraisemblance monotone ? Donner la forme de la zone de rejet du test de Neymann-Pearson correspondant au problème.
9. Expliciter la zone de rejet pour un test de seuil $\alpha = 0,05$, en utilisant les quantiles de lois bien choisies.
10. Est-ce que ce test est uniformément plus puissant ? Si oui, pourquoi ?

Test 3 : On veut tester $H_0 \ : \ \theta = 1$ contre $H_a \ : \ \theta \neq 1$, de façon non asymptotique.

11. Donner la forme de la région critique qui pourrait être celle du test uniformément plus puissant parmi les tests sans biais.
12. Donner les équations qui permettent d'expliciter la région critique.

Exercice 2 (Tests de comparaisons de moyennes et variances)
Un premier groupe de $n = 21$ étudiants a obtenu une moyenne $\hat{\mu}_1 = 12$ au partiel avec un écart-type $\hat{\sigma}_1 = 2$ et un second groupe également de 21 étudiants, une moyenne $\hat{\mu}_2 = 13,2$ avec un écart-type $\hat{\sigma}_2 = 1,8$. $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$ sont les estimateurs sans biais corrigés des variances dans chaque groupe. On suppose que les notes sont indépendantes, de loi normale $\mathcal{N}(\mu_1, \sigma_1^2)$ pour le groupe 1, et $\mathcal{N}(\mu_2, \sigma_2^2)$ pour le groupe 2. On veut tester si les étudiants du groupe 2 ont significativement mieux réussi que ceux du groupe 1.

1. Quelles sont les lois de $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$ et $\hat{\sigma}_2^2$?

2. On va commencer par tester si les variances sont égales. On rappelle que la loi de Fisher F_{ν_1,ν_2} est la loi de

$$\frac{X/d_1}{Y/d_2} \quad \text{où} \quad X \sim \chi^2(d_1), \quad Y \sim \chi^2(d_2) \quad \text{et} \quad \text{X et Y sont indépendants.}$$

Sous H_0, quelle est la loi de $\hat{\sigma}_1^2/\hat{\sigma}_2^2$? Donner la région critique pour le test

$$\begin{cases} H_0 & \sigma_1 = \sigma_2 \\ H_1 & \sigma_1 \neq \sigma_2 \end{cases}$$

Utiliser la table de la loi de Fisher pour dire si on accepte ou on rejette le test à 5% (justifier).

3. On admet maintenant que $\sigma_1^2 = \sigma_2^2 = \sigma^2$. On veut tester

$$\begin{cases} H_0 & \mu_1 = \mu_2 = \mu \\ H_1 & \mu_1 < \mu_2 \end{cases}$$

Sous H_0, quelle loi suit $\sqrt{n}(\hat{\mu}_1 - \hat{\mu}_2)$? $(n-1)(\hat{\sigma}_1^2 + \hat{\sigma}_2^2)$?

$$T = \sqrt{n}\frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} ?$$

4. Donner la forme de la zone de rejet.
5. Utiliser les tables statistiques pour dire si on accepte ou on rejette le test (à 5%). Justifier.

Exercice 3 (Eléphants (d'après Dussaix et Grosbras))
Un directeur de cirque possède $N = 100$ éléphants. Il souhaite estimer le poids total de son troupeau afin de les emmener en tournée en bateau. On s'intéresse à la variable poids Y et on notera Y_i le poids du $i^{\text{ème}}$ éléphant ($i \in \{1, \dots, 100\}$). Les éléphants sont regroupés en deux classes : les mâles ($h = 1$) et les femelles ($h = 2$).

L'année précédente, le directeur avait déjà fait peser ses éléphants et trouvait les résultats suivants :

Classe h	Effectif N_h	Poids moyen $\bar{Y}_h$	Variance corrigée $\bar{S}_h^2$
Mâles : $h = 1$	60	6	4
Femelles : $h = 2$	40	4	2.25

Partie 1 : sondage aléatoire simple (SAS). Le directeur suppose que les variances de Y par classe restent celles de l'année dernière. Pour estimer le poids total $P = \sum_{i=1}^N Y_i$, il procède à un SAS de n éléphants tirés sans remise. Il constitue ainsi un échantillon s.

1. Recalculer les probabilités d'inclusion π_i.
2. Rappeler l'expression de l'estimateur d'Horvitz-Thompson $\widetilde{P}$ de P. Quel est l'estimateur associé $\widetilde{Y}$ de $\mathbb{E}(Y)$?
3. On rappelle que la variance de $\widetilde{P}$ est

$$\text{Var}(\widetilde{P}) = N^2\Big(1 - \frac{n}{N}\Big)\frac{\bar{S}^2}{n} \tag{1}$$

où $\bar{S}^2$ est la variance corrigée de Y (attention, il faut prendre en compte les variances inter et intra classes dans le calcul). Calculer numériquement $\text{Var}(\widetilde{Y})$ si l'on tire 10 éléphants pour le SAS.

Partie 2 : sondage stratifié par sexe. Le directeur choisit de faire un sondage stratifié. Soient n_1 et n_2 le nombre d'éléphants mâles et femelles à tirer. On pose $n = n_1 + n_2$.

1. Quelles sont les probabilités d'inclusion ?
2. Donner la définition de l'estimateur de Horvitz-Thompson $\widehat{P}$ dans ce cas. Quel est l'estimateur associé $\widehat{Y}$ pour $\mathbb{E}(Y)$?
3. En adaptant l'équation (1), donner $\mathrm{Var}(\widehat{P})$ en fonction de n_1, n_2, N_1, N_2, $\bar{S}_1^2$ et $\bar{S}_2^2$?
4. On considère un sondage stratifié proportionnel : le taux de sondage $f = 1/10$ est le même dans chacune des strates. Calculer n_1 et n_2. Comparer $\widetilde{Y}$ et $\widehat{Y}$.
5. En utilisant la question 3, montrer que
$$\mathrm{Var}(\widehat{Y}) = \frac{1-f}{n} \frac{N_1 \bar{S}_1^2 + N_2 \bar{S}_2^2}{N}.$$
Calculer numériquement cette variance. Cela vous semble-t-il paradoxal ?
6. Justifier pourquoi on a :
$$\mathrm{Var}(\widetilde{Y}) = \frac{1-f}{n} \left(\frac{N_1 \big(\bar{Y}_1 - \mathbb{E}(Y)\big)^2 + N_2 \big(\bar{Y}_2 - \mathbb{E}(Y)\big)^2}{N-1} \right) + \frac{N}{N-1} \mathrm{Var}(\widehat{Y}) - \frac{1-f}{n} \frac{\bar{S}_1^2 + \bar{S}_2^2}{N-1}.$$
Commenter ; comprenez-vous le paradoxe ?

7. Calculer numériquement la part de réduction de la variance par rapport au SAS, c'est-à-dire $1 - \mathrm{Var}(\widehat{Y})/\mathrm{Var}(\widetilde{Y})$. Commenter.