TD n°6 – Tests asymptotiques

stoehr@ceremade.dauphine.fr

Exercice 1. Dans le cadre de l'exercice 6 du TD n°5, on considère le problème de test

$$H_0 : \lambda = \lambda_0 \quad \text{contre} \quad H_1 : \lambda \neq \lambda_0.$$

1. Rappeler l'expression de l'estimateur du maximum de vraisemblance $\widehat{\lambda}_n$ de λ.
2. L'estimateur $\widehat{\lambda}_n$ est-il asymptotiquement normal ?
3. Proposer le test de Wald de niveau asymptotique $\alpha \in]0,1[$ permettant de tester les hypothèses H_0 et H_1.

Exercice 2. Pour une population de $n = 300$ personnes, on observe le nombre de cancers diagnostiqués sur une période de 20 ans (données fictives).

	Fumeur	Non-fumeur
Cancer diagnostiqué	33	15
Pas de cancer	117	135

Il est naturel de se demander si la différence du nombre de cancers entre fumeurs et non-fumeurs est statistiquement significative. On fait l'hypothèse des modèles statistiques suivant $\left\{\mathscr{B}(p_1) \mid p_1 \in]0,1[\right\}$ pour les fumeurs et $\left\{\mathscr{B}(p_2) \mid p_2 \in]0,1[\right\}$ pour les non-fumeurs. On note $X = (X_1, \ldots, X_n)$ le n-échantillon observé pour les fumeurs et $Y = (Y_1, \ldots, Y_n)$ le n-échantillon observé pour les non-fumeurs. On réalise le test statistique suivant

$$H_0 : p_1 = p_2 \quad \text{et} \quad H_1 : p_1 \neq p_2$$

1. Déterminer, sous H_0, la loi de la statistique de test suivante
$$h(X) = \frac{\sqrt{n}(\overline{X}_n - \overline{Y}_n)}{\sqrt{2\hat{\lambda}(1-\hat{\lambda})}} \quad \text{avec} \quad \hat{\lambda} = \frac{1}{2n}\sum_{i=1}^{n}(X_i + Y_i).$$
2. Donner la zone de rejet à choisir, pour obtenir un test de taille asymptotiquement $\alpha \in]0,1[$.
3. Déterminer l'erreur de seconde espèce et la fonction puissance de ce test de taille α.
4. Donner la conclusion, l'erreur de seconde espèce et la fonction puissance du test pour les tailles de test suivantes : 1%, 5%, 10%.
5. Calculer la p-valeur. Que pouvez-vous conclure ?

Exercice 3 *(Test d'adéquation à une loi).* On dispose de données gouvernementales américaines (fictives) qui représentent la distribution du nombre de naissance par jours sur une semaine pour toutes les naissances durant cette semaine aux USA ($n = 392960$).

Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
57 180	64 440	62 670	61 740	60 840	45 380	40 710

On souhaite mettre en place un test pour savoir si les naissances sont réparties uniformément sur les sept jours de la semaine.

1. Construire le test du χ^2 permettant de tester l'ajustement à une loi uniforme sur les septs jours de la semaine.

Exercice 4 *(Test d'ajustement à une famille paramétrée).* Dans un verger, on mène une étude pour savoir si les insectes ont un comportement grégaire ou non lorsqu'il s'agit de se nourrir d'un fruit. Pour cela, on regarde combien de fois un insecte s'est nourri sur un fruit pris au hasard. On note N ce nombre. Les spécialistes indiquent que si les insectent agissent de façon indépendante (*i.e.*, ils ne sont pas grégaires), on peut modéliser N par une variable aléatoire suivant une loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$ inconnu. Voici les résultats de l'expérience ($n = 300$)

0	1	2	3	4	5	6	7
60	105	65	47	15	4	3	1

1. Rappeler l'expression de l'estimateur du maximum de vraisemblance $\widehat{\lambda}_n$ de λ.

2. Construire un test du χ^2 d'ajustement à cette famille de loi paramétrée qui est asymptotiquement de taille $\alpha \in]0, 1[$

   (a) lorsque l'on considère les 8 classes.

   (b) lorsque l'on considère les classes en regroupant les 3 dernières.

3. Donner la p-valeur pour ces deux tests. Quelle est la conclusion pour les tailles de test usuelles ?