Chapitre 7

Tests d'adéquation et d'indépendance

1) Test d'adéquation Dans cette section, nous présentons une méthode pour tester l'hypothèse que les fréquences observées pour les différentes catégories (classes) sont en adéquation avec une distribution donnée.

Notations

no_i représente les fréquences observés de résultats.

nt_i représente les fréquences attendu de résultats.

k représente le nombre de classes.

n représente le nombre total d'essais.

Conditions d'application

- Les données sélectionnées aléatoirement
- Pour chaque catégorie, la fréquence attendu est au moins 5 ($nt_i \lesssim 5$).

Les hypothèses à tester

H_0 : les observations suivent la distribution p_i.

H_1 : les observations ne suivent pas la distribution p_i.

Statistique de test

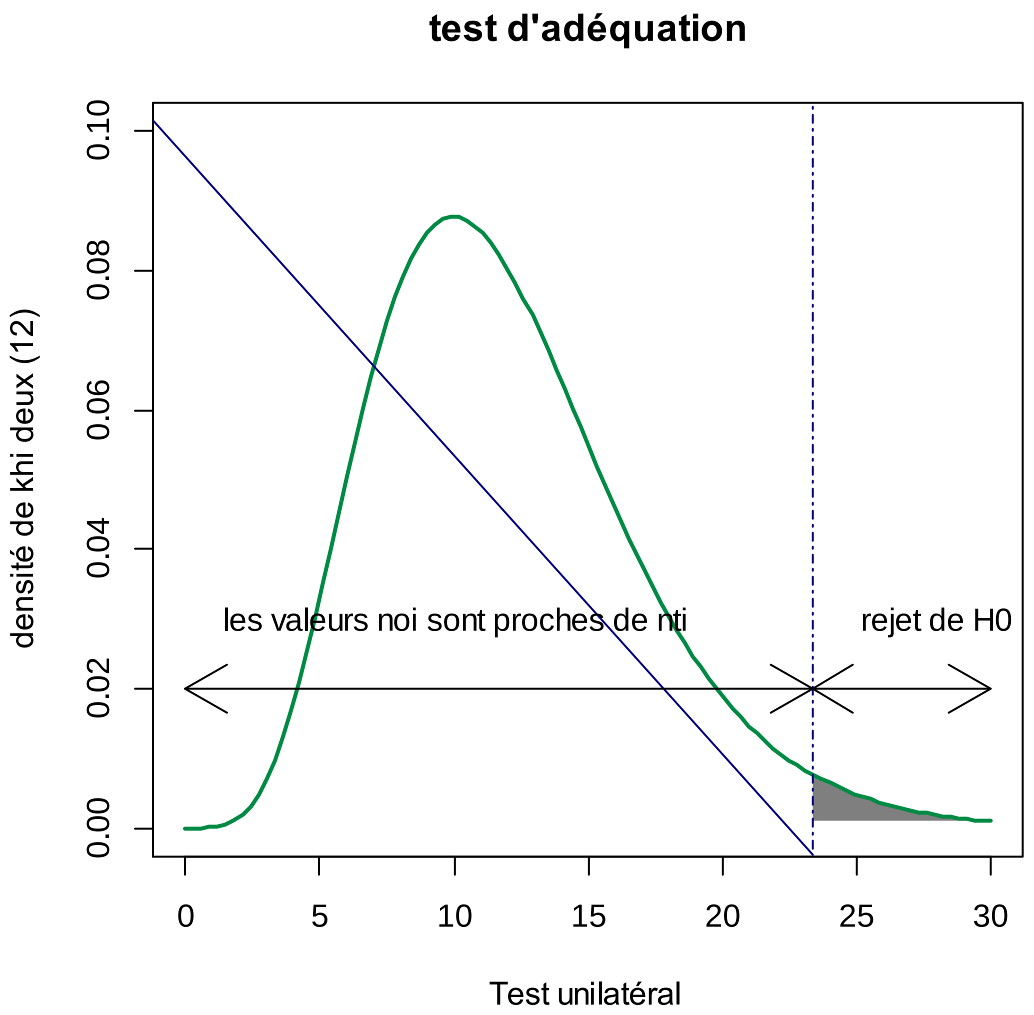
$$X^2_c = \sum \frac{(noi-nti)^2}{nti}$$

Valeurs critiques

Les valeurs critiques sont les dans a table de khi deux avec ddl=k-1

Les tests d'hypothèses d'adéquation sont toujours unilatéraux à droite.

Représentation graphique



Interprétation

Le non rejet de H_0 signifie que les valeurs noi et nti sont proches.

Le rejet de H_0 signifie que les valeurs noi et nti sont éloignées.

Exemple

A partir du génotype des parents, on s'attend à ce que les enfants aient des génotypes répartis comme suit : 25% de génotype AA, 50% de génotype Aa et 25% de génotype aa. Pour une maladie particulière, AA représente un enfant sain, Aa un enfant porteur et aa un enfant malade.

Le tableau suivant donne les fréquences des génotypes pour 90 malades sélectionnés aléatoirement.

génotype	AA	Aa	aa
Fréquences observées (noi)	22	55	13

Tester au niveau de significativité $\alpha = 0.01$ l'hypothèse que ces fréquences observées correspondent aux fréquences attendues.

Solution

Les données ont été sélectionnées aléatoirement.

Il reste à vérifier que les fréquences attendues sont toutes d'au moins 5 et cela en calculant les nti.

H_0 : la distribution des génotypes des enfants est p1=0.25(AA), p2=0.5(Aa), p3= 0.25 (aa).

H_1 : au moins une des proportions ci-dessus est différente des valeurs supposées.

Calcul des nti.

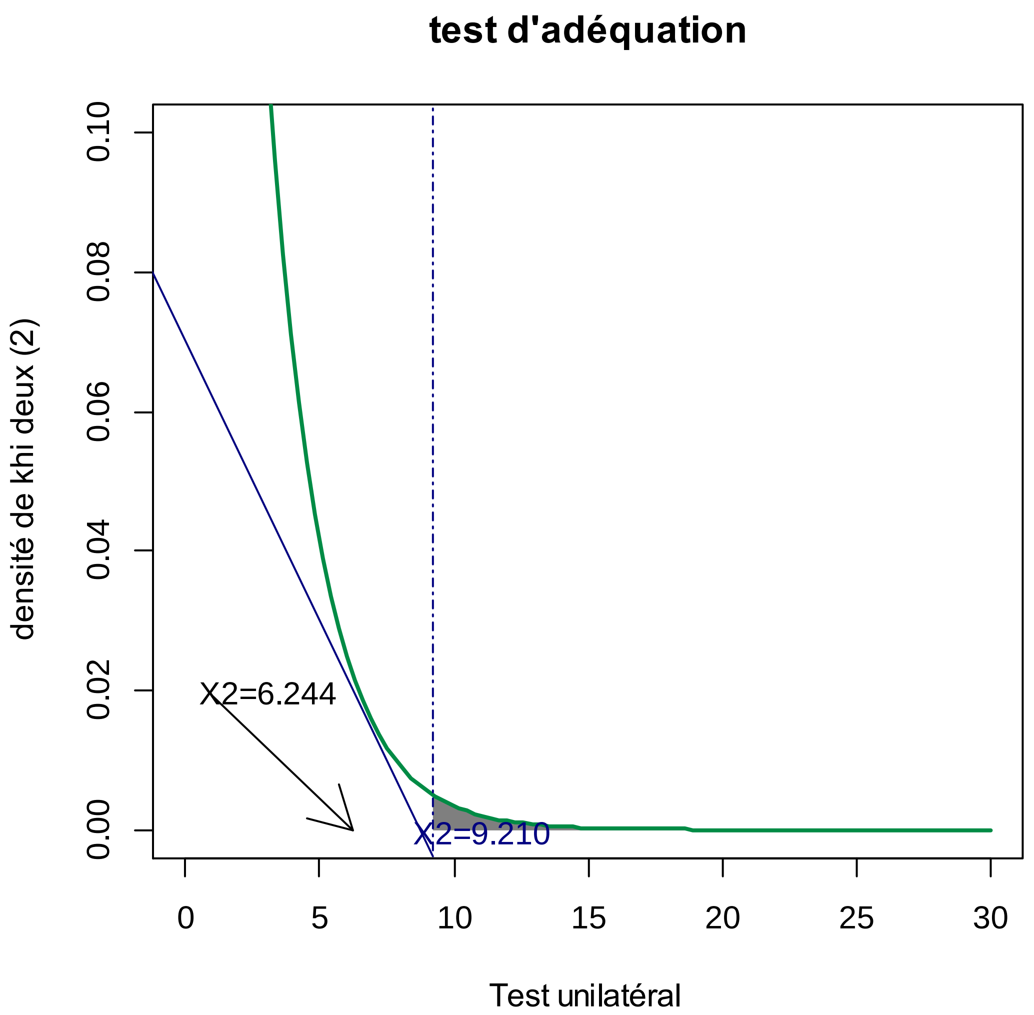
génotype	AA	Aa	aa
Fréquences observées (noi)	22	55	13
nti = n*pi	90*0.25=22.5	90*0.5=45	90*0.25=22.5
noi-nti	-0.5	10	-9.5
$(noi-nti)^2$	0.25	100	90.25
$\frac{(noi-nti)^2}{nti}$	0.0111	2.2222	4.0111

X^2_c = 0.111+0.2222+4.0111= 6.2444.

La valeur critique est lu dans la table de khi deux ddl=3-1=2 et α=0.01 on a alors 9.210.

Décision : X^2_c ne tombe pas dans la région critique, il n'y a pas suffisamment de preuves pour garantir le rejet de H_0.

On ne peut pas rejeter l'hypothèse que la distribution est p1=0.25, p2=0.5, p3=0.25



2) Test d'indépendance

Tableaux de contingence : indépendance et homogénéité

Un tableau de contingence est un tableau dans lequel les fréquences correspondent à deux variables : une variable est utilisée en ligne et l'autre en colonne.

Un test d'indépendance teste l'hypothèse nulle qu'il n'y a pas d'association entre la variable en ligne et celle en colonne dans un tableau de contingence.

Conditions requises.

- Les données d'échantillon ont été sélectionnées aléatoirement.
- H_0 est l'affirmation que les variables de ligne et de colonne sont indépendantes, H_1 est l'affirmation que les variables de ligne et de colonne sont dépendantes.
- Pour chaque case du tableau de contingence, la fréquence attendu n_{ti} est au moins 5.

Statistique de test

$$\chi_c^2 = \sum \frac{(noij - ntij)^2}{ntij}$$

Valeurs critiques

la valeur est lue dans la table de khi deux et le degrés de liberté ddl= (r-1)(k-1)

où r est le nombre de lignes et k le nombre de colonnes.

Dans un test d'indépendance la région critique est située à droite de la valeur critique.

La fréquence attendue pour un tableau de contingence

$$nt_{ij} = \frac{(some\ de\ la\ ligne\ i)(somme\ de\ la\ colonne\ j)}{n},$$

fréquence attendue pour la case ij

Exemple : « test de l'efficacité de vaccin de Salk »

Dans une expérience, on a donné à des enfants le vaccin de Salk contre la polyométrie et à d'autres enfants un placebo. Les résultats de l'expérience sont résumés dans le tableau suivant

	Oui	non
Groupe traitement vaccin	33	200712
Groupe placebo	115	201114

Utiliser un niveau de significativité de 0.05 pour tester l'hypothèse d'indépendance entre les groupes et le traitement.

Solution

Les données sont des comptages de fréquences indépendants.

Il reste à vérifier la 2ème condition ; $nt_{ij} \geq 5$.

On teste les hypothèses suivantes :

H_0 : recevoir le vaccin Salk est indépendant d'avoir la polio

H_1 : recevoir le vaccin Salk et avoir la polio sont dépendants.

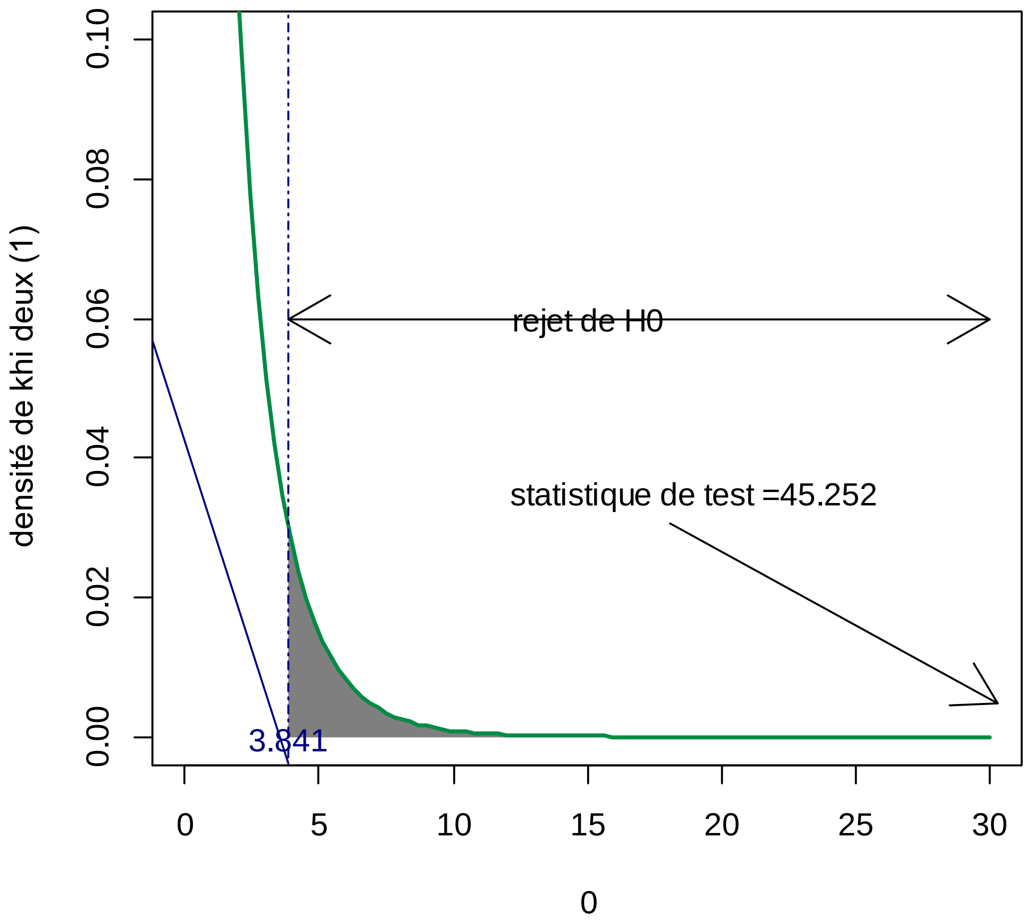
Calcul des ntij

	Oui	non
Groupe traitement vaccin	33 (73.911)	200712(200671.089)
Groupe placebo	115(74.089)	201114 (201154.911)

$$\chi_c^2 = \frac{(33-73.911)^2}{73.911} + \frac{(200712-200671.089)^2}{200671.089} + \frac{(115-74.089)^2}{74.089} + \frac{(201114-201154.911)^2}{201154.911}$$

$$= 22.645 + 0.008 + 22.591 + 0.008 = 45.252$$

La valeur critique est χ^2 = 3.841 trouvée dans la table de khi deux (ligne ddl=(r-1)(k-1)=(2-1)(2-1)=1 et colonne 0.05)



Interprétation comme la statistique de test tombe dans la région critique, on rejette H_0 qui est « recevoir le vaccin Salk est indépendant d'avoir la polio ». il apparaît que les variables sont dépendantes.