Devoir à la maison

Convergence de la statistique χ^2.

Exercice1

Une urne contient des boules numérotées de 1 à d. la proportion des boules de numéro i est p_i , avec $p_i > 0$.

On note P le vecteur de composantes $p_i : i = 1, . . . \ d$. On tire n boules avec remise. Soient Y_n le numéro tiré au n-iéme tirage . X_n le vecteur aléatoire de composantes $X_n(i) = 1_{\{Y_n=i\}}$. Soit $S_n(i)$

Le nombre de boules de numéro i tirées en n tirages et S_n le vecteur aléatoire de de composantes $S_n(i)\ pour\ i = 1, . . ., d$.

1. Vérifier que $X_n(i) \sim \mathrm{B}(p_i)$, $S_n(i) \sim \mathrm{B}(n, p_i)$. (ce qui implique que :

$$E(X_n(i)) = p_i \ , var(X_n(i)) = p_i(1 - p_i), \qquad E(S_n(i)) = np_i\ et\ \ var(S_n(i)) = np_i(1 - p_i).$$

2. vérifier que l'espérance du vecteur X_n est P et que sa matrice de variance covariance est

$D - PP'$ où $D = diag(p_i)$. C'est-à-dire $D = \begin{pmatrix} p_1 & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & p_d \end{pmatrix}$

3. Soit $\Sigma_n = \sqrt{n}(\frac{S_n}{n} - P)$ utiliser le théorème de la limite centrale vectorielle pour vérifier que :$\Sigma_n \xrightarrow{En\ loi} G = N_d\ (0, D - PP')$.
4. Soit $T_n = n \sum_{i=1}^{n} (\frac{S_n(i)}{n} - p_i)^2 / p_i$

Nous voulons montrer que $T_n \xrightarrow{En\ loi} \chi^2(d-1)$ pour cela soit Z le vecteur aléatoire de composantes $G(i)/\sqrt{p_i} \quad i = 1, . . ., d$. Soit $Z = \begin{pmatrix} G(1)/\sqrt{p_1} \\ \\ G(d)/\sqrt{p_d} \end{pmatrix} T_n$

a) Vérifier que $Z = MG$ avec $M = \begin{pmatrix} 1/\sqrt{p_1} & \cdots & \\ \vdots & \ddots & \vdots \\ & \cdots & 1/\sqrt{p_d} \end{pmatrix} \quad et\ G = \begin{pmatrix} G(1) \\ \\ G(i) \end{pmatrix}$.

b) Soit $v_d = . \begin{pmatrix} \sqrt{p_1} \\ \\ \sqrt{p_d} \end{pmatrix}$

-Vérifier que $\Gamma_Z = M\Gamma_G M' = MDM' - MPP'M'$.

-Vérifier que $MDM' = I_d$, $MPP'M' = v_d v_d' \quad en\ déduire\ que\ \ \Gamma_Z = \ I_d - \ v_d {v_d}'$.

5. soit $Z_n = \begin{pmatrix} \Sigma_n(1)/\sqrt{p_1} \\ \\ \Sigma_n(d)/\sqrt{p_d} \end{pmatrix}$.

a) vérifier que $T_n = \sum_{i=1}^{d} \left[\sqrt{n}\ \ \left(\frac{S_n(i)}{n} - p_i\right)/p_i\right]^2 . = \sum_{i=1}^{d} [\Sigma_n(\mathrm{i})/\sqrt{p_1}]^2 = f(Z_n) = \|Z_n\|^2$

où $f(x) = \sum_{i=1}^{n} {x_i}^2$.

c) Vérifier que $Z_n = M\Sigma_n$ et utiliser 3. 4a) pour montrer que $Z_n \xrightarrow{En\ loi} Z$ en déduisant que :

$\|Z_n\|^2 \xrightarrow{En\ loi} \|Z\|^2$ soit $T_n \xrightarrow{En\ loi} \|Z\|^2$.

d) Soit $V\ une\ d \times d$ matrice orthogonale réelle dont la dernière ligne est ${v_d}'$

-donner la loi du vecteur aléatoireVZ.

-comparer $\|VZ\|^2\ et\ \|Z\|^2$ en déduire que $T_n \xrightarrow{En\ loi} \chi^2(d-1)$. qui est le résultat cherché.

On a démontré ainsi le théorème suivant :

Théorème : Soit $\{X_1, \ldots, X_n\}$ un échantillon d'une loi P à valeurs dans $\{1, \ldots, I\}$.

Posons $p_i = P(\{i\})\ et\ N_n(i) = \sum_{j=1}^{n} 1_{(X_j=i)}$ Alors :

$$T_n = \sum_{i=1}^{I} (\frac{N_n(i) - np_i}{\sqrt{np_i}})^2 \xrightarrow{En\ loi} \chi^2(I-1)$$

Exercice2

Définition : Soient $P\ et\ Q$ deux lois à valeurs dans $\{1, \ldots, I\}$. Posons $p_i = P(\{i\})$ et $q_i = Q(\{i\})$.

La quantité $\chi^2(P,Q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i}$ est appelée distance de $\chi^2\ entre\ P\ et\ Q$.

On considère deux n-échantillons $\{X_1, \ldots, X_n\}\ et\ \{Y_1, \ldots, Y_n\}$ des lois de deux variables aléatoires $X\ et\ Y$ à valeurs dans $\{1, \ldots, I\}\ et\ \{1, \ldots, J\}$ respectivement.

La loi empirique du couple (X, Y) est donnée par :$\hat{P}_n(i,j) = N_n(i,j)/n$

où $N_n(i,j) = \sum_{k=1,l=1}^{n} 1_{(X_k=i, Y_l=j)}$. On veut tester l'indépendance de $X\ et\ Y$.

Si elles sont indépendantes. la loi du couple est le produit des lois marginales.

Donc H_0: "$P_{i,j} = P_i P_j.\ 1 \le i\ \le I\ et\ 1 \le j \le J$" contre H_1: il existe (i,j) tel que: $P_{i,j} \neq P_i P_j$. avec

$P_i = \sum_{j=1}^{n} P_{i,j}\ \ etP_j = \sum_{i=1}^{n} P_{i,j.}$ Sous H_0 la loi empirique s'écrit $\overline{P}_n(i,j) = N_n(i,.)N_n(.,j)/n^2$.

Utiliser le théorème ci -dessus pour montrer que :

$\chi^2(\hat{P}_n, \overline{P}_n)$ converge vers $\chi^2((I-1)(J-1))$ sous H_0 et $+\infty\ sous\ H_1$.

Déduire de ce résultat le test d'indépendance de χ^2, qui est le suivant :

On rejette H_0 si $\chi^2(\hat{P}_n, \overline{P}_n) \ge {\chi^2}_{(I-1)(J-1),\alpha}$ quantile d'ordre $1 - \alpha\ de\ la\ loi\ \chi^2\big((I-1)(J-1)\big)$.

Exercice3

Un utilisant le tableaux contingence :

1. Vérifier que $N_n(i,j) = n_{ij}\ et\ que\ \overline{P}_n(i,j) = n_{i.}n_{.j}/n^2$,$et\ t_{ij} = n\ \overline{P}_n(i,j)$.
2. En déduire que $\chi^2(\hat{P}_n, \overline{P}_n) = \sum_{i,j}^{I \times J} (n_{ij} - t_{ij)})^2/t_{ij}$. et en déduire le résultat du cours.

3.