

TP3 : Evaluation

1 Tests du χ^2 d'indépendance

Exercice 1 Soient $(X_i)_{1 \leq i \leq n}$ et $(Y_i)_{1 \leq i \leq n}$ deux échantillons indépendants de même loi binomiale $\mathcal{B}(2, 0.5)$.

1. Simuler pour $n = 100$ ces échantillons, et faire un test d'indépendance du χ^2 de niveau 0.05 (il existe une commande pour le test du χ^2 sous R aller voir **chisq.test** dans l'aide).
2. On pose ensuite que $Z_i = \epsilon_i X_i + (1 - \epsilon_i) Y_i$ où $(\epsilon_i)_{1 \leq i \leq n}$ est un échantillon de loi $\mathcal{B}(p)$. Pour chacune des valeurs de n et de p suivantes, simuler ces échantillons et faire un test d'indépendance entre X et Z de niveau 0.05.
 - (a) $n = 100$ et $p = 0.5$,
 - (b) $n = 100$ et $p = 0.01$,
 - (c) $n = 10000$ et $p = 0.01$.

Exercice 2 Soient $(X_i)_{1 \leq i \leq n}$ et $(Y_i)_{1 \leq i \leq n}$ deux échantillons indépendants de même loi normale centrée réduite $\mathcal{N}(0, 1)$.

1. Simuler deux échantillons pour différentes valeurs de n allant de 10 à 1000 et faire (pour chaque n) un test d'indépendance du χ^2 de niveau 0.05. Représenter sur un graphique les résultats du test.
2. Soit $\theta \in [0, 2\pi]$. On pose $A_i^\theta = X_i \cos(\theta) + Y_i \sin(\theta)$ et $B_i^\theta = -X_i \sin(\theta) + Y_i \cos(\theta)$. Simuler ces échantillons pour différents θ (en choisir 3) et représenter les p-value issues des différents tests sous la forme d'un graphique.
3. Si les test précédent étaient réalisés au niveau 0.01, qu'elle serait la conclusion? Pourrait-on s'y attendre? (Indication : Deux variables aléatoires gaussiennes sont indépendantes si leur covariance est égale à 0)

2 Tests du χ^2 d'adéquation

Considérons un tableau de valeurs A issu d'une même loi P . Le logiciel R dispose de certaines commandes de base donnant une première illustration des données. Ainsi, on peut citer les commandes **mean**, **var** et **sd** pour la moyenne, la variance et l'écart type. Pour les quantiles, on utilisera **boxplot** et **summary**. Enfin, pour les tracés de distributions, les commandes **hist** et **ecdf** permettent d'obtenir des informations sur la densité et la fonction de répartition de P .

Exercice 3 Considérons l'échantillon statistique issu de la même loi P_X donné par le tableau suivant

3	2	1	1	1	4	1	1	4	5	2	4	1	2	1
1	1	1	0	2	1	0	2	1	2	1	5	1	3	2

1. Donner la moyenne, la variance, l'écart type ainsi que les quantiles de cette série.

2. Réaliser un histogramme ainsi qu'une étude de la fonction de répartition empirique. Qu'en déduisez-vous ?

3. Mettre en place des tests d'adéquation pour vérifier votre conjecture (on pourra consulter l'aide sur la commande `chisq.test`).

Exercice 4 Considérons l'échantillon statistique issu de la même loi P_Y donné par le tableau suivant

1.37	2.72	0.42	0.04	1.04	0.47	4.58	0.71	0.89	0.09
1.27	0.30	1.12	1.21	0.48	2.10	0.44	0.79	1.34	0.78
2.73	0.12	0.29	0.11	0.07	0.35	1.89	0.68	0.80	0.38
0.88	0.16	0.06	0.24	2.76	1.41	1.98	0.02	0.67	1.28

1. Donner la moyenne, la variance, l'écart type ainsi que les quantiles de cette série.

2. Réaliser un histogramme ainsi qu'une étude de la fonction de répartition empirique. Qu'en déduisez-vous ?

3. Mettre en place des tests d'adéquation pour vérifier votre conjecture.

4. Un de vos collègues pense que cet échantillon suit une loi $\mathcal{N}(\mu, 1)$, qu'en pensez-vous ?

5. (Bonus) A l'aide de la commande `ks.test`, réaliser un test de Kolmogorov Smirnov sur votre jeu de donnée. Vérifier son adéquation à une loi Expo, Normale puis Gamma. Que constatez-vous ?

6. (Bonus) Tirer vos propres jeux de données continus et réaliser les tests du Khi-deux et de K.Smirnov. Que constatez-vous ?

3 BONUS pour ceux qui n'ont pas rendu le premier TP : Graphiques bi-variés

`plot` : c'est la fonction de base pour afficher des points avec deux coordonnées. Elle re-crée un graphique à chaque appel, et place des points selon deux axes.

`legend` : cette fonction permet d'ajouter une légende.

Exemple 1 `plot(T[,c(1,3)], pch = ifelse(T$SEX == 1,25,26), col = ifelse(T$SEX == 1,"blue","red"))`

`legend("bottomleft", c("H","F"), col = c("blue","red"), pch = c(25,26))`

Le premier paramètre de la fonction `plot` – si c'est une table ayant au moins deux colonnes – ou les deux premiers paramètres – si ce sont des vecteurs – donnent les coordonnées des points à dessiner. Autres paramètres :

`pch` un (vecteur d')entier(s) indiquant le(s) symbole(s) à utiliser ;

`col` un (vecteur de) couleur(s) indiquant le(s) couleur(s) à utiliser ("black" par défaut, on peut utiliser "blue", "red", ..., la commande `colors()` retourne la liste des couleurs possibles) ;

`xlim, ylim` les limites des axes horizontaux et verticaux (par défaut, les limites sont calculées de manière à faire rentrer exactement tous les points) ;

`xlab, ylab` les labels des axes horizontaux et verticaux (par défaut, les noms des variables) ;

`main` le titre du graphique (en gras, centré au-dessus du graphique) ;

sub le sous-titre du graphique (centré au-dessous du graphique).

Exercice 5 Créer le graphique ci-contre, avec des couleurs distinctes pour distinguer les hommes des femmes, ainsi que leur statut marital. (Il faut utiliser les fonctions `text`, pour d'ajouter du texte à un graphique, et `arrows`, pour ajouter des flèches (la fonction `lines` permet d'ajouter des lignes)).

Créer des graphiques dans des fichiers C'est possible à l'aide des fonctions `pdf`, `tiff`, `png`, `jpeg` entre autres. Par exemple : `pdf(mon_fichier.pdf)`
... instructions graphiques ...
`def.off()`

Pour aller plus loin Le manuel d'introduction à R (cran.r-project.org/doc/manuals/R-intro.pdf) contient une liste de fonctions graphiques, ainsi qu'une liste de paramètres que peuvent accepter ces fonctions (page 68).

Enquête ONU 1967 :
Temps passé en act. prof. et pour le ménage, par cat. de personne

