Université de Nice
Département de Mathématiques

Année 2005-2006
Licence MI/MP 2e année

Statistiques : Enoncé du TP5
Test d'équirépartition

Un casino souhaite déterminer si les joueurs trichent à l'aide de dés pipés ou non. On met donc en œuvre une expérience destinée à établir si un dé est pipé. Un dé sera déclaré pipé si les probabilités d'apparition de chacune de ses faces ne sont pas égales. On s'accorde une marge d'erreur de $\alpha = 10\%$.
On commence par recueillir des données sur un certain nombre de dés dans le casino : chaque dé est jeté 300 fois et l'on note dans un tableau le nombre d'apparitions de chacune des 6 faces.

1. Simulation des lancés de dés recueillis

Entrez dans Scilab la commande `nblances = 300`.

a. Les commandes suivantes permettent de simuler les lancés de quatre dés du casino :

- `de1 = int(rand(1, nblances)*5 + 1)`
- `de2 = int(rand(1, nblances)*6 + 1)`
- `de3 = int(rand(1, nblances)*6 - .5) + 1`
- `de4 = int(rand(1, nblances)*5 - .5) + 1`

b. Utiliser la commande
`histplot([.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5], deX)`
pour essayer de déterminer graphiquement lesquels de ces dés sont pipés et lesquels ne le sont pas.

2. Établissement de la loi de répartition pour un dé normal

Pour établir comment un dé normal se comporte, on va réaliser 100 fois l'expérience des 300 tirages. Cela nous permettra d'analyser les fréquences d'apparition de chacune des faces, d'une expérience à l'autre.

a. Sachant que la commande `int(rand()*6 + 1)` renvoie chacun des entiers entre 1 et 6 avec la même probabilité, construire le tableau `tirages` des 100×300 lancés du dé.

b. Grâce à la commande
`for i = 1:6, totaux(:, i) = sum(bool2s(tirages == i), 'c'), end,`
on obtient un tableau `totaux` récapitulant, par expérience, le nombre d'apparitions de chacune des faces.
À partir du tableau `totaux`, qui contient les nombres d'apparitions, construire un tableau `distfreq` qui contient les différences entre les fréquences d'apparition observées (obtenues dans le tableau `totaux/nblances`) et les fréquences d'apparitions attendues (1/6 pour chaque face).

c. Sachant que la commande `.*` permet de multiplier terme-à-terme deux tableaux de mêmes dimensions, construire le tableau `distfreqcarre` des différences au carré entre fréquences d'apparition observées et fréquences d'apparitions attendues.

d. La commande `d2exp = sum(distfreqcarre, 'c')` permet alors d'obtenir un vecteur contenant, pour chaque expérience, la somme des carrés des écarts aux fréquences attendues. Tracer un histogramme à 20 classes de ces valeurs.

3. Analyse de la loi de répartition pour un dé normal

On souhaite maintenant étudier la répartition des valeurs du tableau `d2exp`. Pour chaque seuil s, on peut déterminer la proportion $\pi(s)$ des valeurs de `d2exp` qui sont plus grandes que s. En particulier, on appelle *neuvième décile* le seuil d_9 tel que $\pi(d_9) = alpha = 0,1$.

a. Sachant que la commande `sort(d2exp)` renvoie un tableau contenant les valeurs du tableau `d2exp` triées dans l'ordre décroissant, calculer le neuvième décile `d9` de `d2exp`.

4. Détection des dés pipés

a. Comme pour le dé normal, on peut calculer pour chaque dé `X` à évaluer le nombre d'apparitions de chaque face par la commande
`for i = 1:6, totalX(i) = sum(bool2s(depipe == i)), end`
Comme précédemment, calculer pour chacun des quatre dés les différences au carré entre les fréquences observées et les fréquences attendues (1/6 pour chaque face), puis leur somme `d2deX` (remplacer `X` par le numéro du dé).

b. Comparer chacune des valeurs `d2de1`, `d2de2`, `d2de3` et `d2de4` au neuvième décile `d9` obtenu pour le dé normal et conclure.

5. Comment se passer des tests sur le dé normal ?
On peut éviter de calculer expérimentalement le neuvième décile pour le dé normal, comme nous l'avons fait. En effet, pour un dé normal, on peut montrer que si l'on note f_i la fréquence observée d'apparition de la face i, g_i la fréquence attendue et σ_i son écart-type, la loi de probabilité de la somme

$$\sum_{i=1}^{6}\left(\frac{f_i - g_i}{\sigma_i}\right)^2$$

a une densité qu'on peut approcher par celle de la loi notée χ_5^2.
Dans le cas des fréquences d'apparition des faces, tous les σ_i^2 sont égaux à $p(1-p)n$, où $p = 1/6$ est la probabilité d'apparition d'une face et où $n = 300$ est le nombre de tirages.
On souhaite utiliser cette propriété pour obtenir mathématiquement une bonne valeur de d_9. La commande

```
d9math = cdfchi("X", 5, .9, .1)
```

permet d'obtenir le seuil `d9math` au delà duquel se trouvent 10% des valeurs d'une variable aléatoire de loi χ_5^2.

a. En utilisant le fait que nous avons obtenu `d9` en considérant la répartition des valeurs de

$$\sum_{i=1}^{6}(f_i - g_i)^2$$

et que nous avons obtenu `d9math` en simulant plutôt

$$\sum_{i=1}^{6}\left(\frac{f_i - g_i}{\sigma_i}\right)^2,$$

établir le lien entre `d9math` et `d9`. Le vérifier dans Scilab.

b. Dans la question **2.d.**, nous avons tracé un histogramme de la loi de répartition obtenue expérimentalement. Superposer à cet histogramme la loi du χ_5^2 obtenue mathématiquement par les commandes

```
xx = .0002 :  .0002 :  .0106; yy = xx*36*300/5;
for i=2:53, zz(i) = cdfchi("PQ", yy(i), 5) - cdfchi("PQ", yy(i-1), 5), end;
plot2d(xx, zz)
```

pour comparer la loi expérimentale à une approximation théorique.