

Colles série 5 : Les théorèmes limites et l'adéquation à la loi normale

Sujet 1:

Soit X une variable de loi $\mathcal{N}(\mu, \sigma^2)$ avec $\mu = 1$ et $\sigma^2 = 4$.

1. Simuler un échantillon de taille $n = 1000$ et calculer pour chaque échantillon des k -premiers valeurs la moyenne, la moyenne des carrés et la variance empirique.
2. Représenter la suite des n valeurs obtenues (s_k^2 pour $k = 1, \dots, n$) par un nuage de points en imposant les limites 1 et 7 sur l'axe des ordonnées. Observe-t-on une convergence lorsque $k \rightarrow \infty$? Si oui vers quelle limite (la faire figurer en rouge sur le graphique) ?

Sujet 2:

Soit X une variable de loi $\mathcal{B}(p)$ avec $p = 0.5$.

1. Tirer N (avec N assez grand, au moins 10^3 ...) réalisations d'un échantillon de taille n (pour commencer n petit) et les ranger dans une matrice à N lignes et n colonnes. Calculer les moyennes de chacune des N réalisations du n -échantillon avec la fonction `rowMeans()` et les affecter à `xbarn`.
2. Calculer la moyenne empirique et la variance empirique corrigée (s'^2) de l'échantillon `xbarn` et représenter sa répartition avec un histogramme. Y superposer ensuite la densité d'une loi normale convenablement choisie. Que se passe-t-il pour $n = 5$? pour $n = 500$? Quel résultat vient d'être illustré ici ?

Préambule sujets 3 à 6 :

Ces sujets portent sur les données du fichier `donneesSerie4.csv` qui sera téléchargé et envoyé (sans être ouvert) dans le répertoire qui sera ensuite choisi dans R studio comme répertoire de travail (via le menu `session/set working directory /...`). Ensuite pour charger les données sous R il faut exécuter :

```
read.table("donneesSerie4.csv", header=T, sep=";", dec=".") -> dat
```

Le data.frame nommé `dat` ainsi créé se manipule comme `mtcars`. Il est constitué de six échantillons de taille $n = 100$ tirés sous des lois différentes dont trois sont des lois normales, une est uniforme, une de Student et une du Chi-deux. Les paramètres de ces lois sont inconnus.

Sujet 3:

Dans cet exercice nous n'utiliserons que les échantillons `e1` et `e2`.

1. Faire un résumé numérique de chacun des deux échantillons et les comparer. Calculer leurs écart-type empiriques corrigés (s') et les comparer. Ces comparaisons permettent-elles de dire que ces échantillons sont tirés sous la même loi ou le contraire ?
2. Faire le QQ-plot de l'échantillon `e1` en fonction des quantiles de la loi normale centrée réduite (avec `qqnorm()`). Y ajouter, en rouge, la droite de Henry dont l'ordonnée à l'origine et la pente sont respectivement la moyenne empirique et l'écart-type empirique corrigé s' (avec `abline()`). Faire de

même avec **e2** et afin de comparer les deux graphiques les représenter sur une même fenêtre coupée en deux parties (1 ligne x 2 colonnes) en imposant les mêmes limites en ordonnées pour chacun des deux **qqnorm()**. Que conclure (on pourra noter l'intervalle contenant tout l'échantillon **e2**) ? Sachant les lois utilisées, laquelle proposer pour chacun des deux échantillons ?

Sujet 4:

Dans cet exercice nous n'utiliserons que les échantillons **e3** et **e4**.

1. Faire un résumé numérique de chacun des deux échantillons et les comparer. Calculer leurs écart-type empiriques corrigés (s') et les comparer. Ces comparaisons permettent-elles de dire que ces échantillons sont tirés sous la même loi ou le contraire ?
2. Faire le QQ-plot de l'échantillon **e3** en fonction des quantiles de la loi normale centrée réduite (avec **qqnorm()**). Y ajouter, en rouge, la droite de Henry dont l'ordonnée à l'origine et la pente sont respectivement la moyenne empirique et l'écart-type empirique corrigé s' (avec **abline()**). Que conclure de ce nuage de points et en particulier du signe des quantiles empiriques ? Faire de même avec **e4** et afin de comparer les deux graphiques les représenter sur une même fenêtre coupée en deux parties (1 ligne x 2 colonnes) en imposant les mêmes limites en ordonnées pour chacun des deux **qqnorm()**. Sachant les lois utilisées, laquelle proposer pour chacun des deux échantillons ?

Sujet 5:

Dans cet exercice nous n'utiliserons que les échantillons **e5** et **e6**.

1. Faire un résumé numérique de chacun des deux échantillons et les comparer. Calculer leurs écart-type empiriques corrigés (s') et les comparer. Ces comparaisons permettent-elles de dire que ces échantillons sont tirés sous la même loi ou le contraire ?
2. Faire le QQ-plot de l'échantillon **e5** en fonction des quantiles de la loi normale centrée réduite (avec **qqnorm()**). Y ajouter, en rouge, la droite de Henry dont l'ordonnée à l'origine et la pente sont respectivement la moyenne empirique et l'écart-type empirique corrigé s' (avec **abline()**). Que conclure de ce nuage de points et en particulier de la symétrie des quantiles empiriques ? Faire de même avec **e6** et afin de comparer les deux graphiques les représenter sur une même fenêtre coupée en deux parties (1 ligne x 2 colonnes) en imposant les mêmes limites en ordonnées pour chacun des deux **qqnorm()**. Sachant les lois utilisées, laquelle proposer pour chacun des deux échantillons ?

Sujet 6:

Dans cet exercice nous n'utiliserons tous les échantillons du data.frame **dat**.

1. Partitionner la fenêtre graphique en deux lignes et six colonnes. Faire les histogrammes (en densité) de chacun des six échantillons sur la première ligne en ajoutant à chaque histogramme une verticale rouge passant par la valeur moyenne de l'échantillon. Sur la seconde ligne les six QQ-plot selon la loi normale (avec **qqnorm()**) auquel seront ajoutées les droites de Henry respectives (d'ordonnée à l'origine la moyenne de l'échantillon et pente l'écart-type corrigé s').
2. Trois échantillons sont obtenus avec des lois normales, lesquels ? Précisez-en approximativement les paramètres. Pour les trois autres échantillons, leur associer une des trois autres lois proposées (chacune n'étant utilisée que pour un échantillon).