

NOM : CORRECTION  
PRENOM :

Date :  
Groupe :

### Statistiques : Feuille de réponses du TP 3

On répondra aux questions posées aussi clairement que possible dans les espaces prévus et on remettra cette feuille de réponses en fin de TP à l'enseignant chargé du TP. Les figures obtenues seront sauvegardées au fur et à mesure dans un document d'une page que l'on enverra par mail à son enseignant en fin de séance en se logeant à l'adresse <http://www-math.unice.fr/~dehon/Scilab/formTPStat.html>

#### 1. Fluctuations d'échantillonnage :

1. Expliquer pourquoi `int(rand()+p)` donne un tirage aléatoire issu d'une loi de Bernoulli  $\mathcal{B}(1, p)$ .

$$\begin{aligned} \text{On a } \int(\text{rand}()+p)=1 &\Leftrightarrow \text{rand}() \in [1-p, 1] \\ \int(\text{rand}()+p)=0 &\Leftrightarrow \text{rand}() \in [0, 1-p]. \end{aligned}$$

Comme la loi de la variable aléa. `rand()` est uniformément distribuée sur  $[0, 1]$ , la probabilité de renvoyer 1 est longueur  $(1-p, 1] = p$  et celle de renvoyer 0 est longueur  $[0, 1-p]$

Indiquer quelques valeurs de  $f_e$  trouvées. Quelle est la valeur théorique de  $f_e$ ?  $= 1-p$   
La valeur théorique de  $f_e$  est l'espérance d'une v.a. de loi  $\mathcal{B}(1, p)$ , c'est-à-dire  
 $f_{e \text{ théo}} = p$

2. Quelle commande de Scilab utilisant `grand` simule un tirage aléatoire de loi Binomiale  $\mathcal{B}(n, p)$ ?

C'est `grand(1, 1, 'bin', n, p)`

3. Que signifie 'c' dans la définition du vecteur  $f_e$ ?

Le paramètre signifie que l'on demande le calcul de la somme des colonnes, qui est un vecteur colonne, et pas la somme de tous les termes du tableau, qui est un scalaire.

#### 2. Position et étendue de la distribution des valeurs de $f_e$ :

1. Valeurs de la moyenne et de la médiane :

Sont-elles égales? Peu différentes? Pourquoi?

Elles ne sont pas égales, mais proches. En effet, on voit sur l'histogramme que la distribution de  $f_e$  est symétrique par rapport à une valeur  $x$  proche de  $f_{e \text{ théo}} = p$ . On en déduit que la moyenne et la médiane sont proches de  $p$ .

2. Valeurs de l'étendue, de l'écart type et de l'intervalle interquartile :

Sont-elles égales? Peu différentes? Pourquoi?

Elles ne sont pas égales, ni même proches:

- L'étendue est un intervalle symétrique par rapport à  $p$  contenant toutes les valeurs
- L'intervalle interquartile est un intervalle symétrique par rapport à  $p$  contenant 50% des valeurs
- L'intervalle  $[p-\sigma, p+\sigma]$  est (dans le cas gaussien) un intervalle contenant 64% des valeurs.  
on doit donc avoir  $\text{étendue} > 2\sigma > \text{iig}$ .

3. Indiquez les 2 nombres obtenus et commentez

On obtient  $\text{étendue} = 0,13$   
 $2\sigma = 0,048$   
 $\text{iig} = 0,0325$   
On a bien  $\text{étendue} > 2\sigma > \text{iig}$

4. Expliquer pourquoi cette quantité est l'écart type théorique. Puis indiquer la valeur obtenue et la comparer avec l'écart type empirique.

$\sigma = \sqrt{\frac{p(1-p)}{n}}$  est l'écart-type de  $f_e$ , d'après le théorème central limite (cf cours)

### 3. Intervalle de confiance :

1. Que signifie `ones(100,1)` ?

Cette commande crée un vecteur colonne de dimension 100 et dont toutes les coordonnées valent 1.

2. Sur la figure, combien observez-vous d'intervalles de confiance ne contenant pas la valeur théorique? Ce résultat est-il conforme à la théorie?

On en observe 6. C'est raisonnable, compte-tenu du fait que la théorie prévoit que nous en trouvions 5% de 100, c'est-à-dire 5.

3. Quelle commande Scilab utilisez-vous pour ce calcul?

On utilise `sum(bool2s(ftheo < feinf | ftheo > fesup))`

4. Indiquez vos résultats en les commentant :

Comme l'erreur admise est plus grande (10% au lieu de 5%), nos intervalles de confiance peuvent être plus resserrés autour de valeurs expérimentales. En contrepartie, on obtient bien une dizaine d'intervalles ne contenant pas la valeur théorique de  $f_e$ .

### 4. Influence de la taille de l'échantillon :

1. Quelle valeur trouvez-vous cette fois pour  $f_e$  et quel est son intervalle de confiance?

2. Quelles différences avec l'histogramme du cas  $n = 400$ ?

L'histogramme est plus étroit, ce qui correspond bien à un écart-type plus faible.

3. Moyennes théorique et empirique :

4. Ecarts type théorique et empirique :

5. Les deux nombres trouvés sont-ils conformes à la théorie?

6. Qu'observez-vous dans le cas où  $n = 10000$ ?

Un écart-type plus faible encore.