

Cette feuille de T.P. est consacrée à l'étude de certaines simulations avec SAS, ainsi qu'à quelques applications (ce thème n'est pas abordé dans le cours polycopié).

Préambule

La base de toute simulation est un générateur de *nombre pseudo-aléatoires*, supposés distribués selon une loi de probabilité continue, uniforme sur l'intervalle $[0, 1]$ de \mathbb{R} .

La plupart des logiciels de statistique disposent d'un tel générateur (ainsi, d'ailleurs, que la plupart des calculettes !), à partir duquel il est possible de simuler, en utilisant des transformations appropriées, de nombreuses lois de probabilité.

SAS dispose d'un générateur de nombres pseudo-aléatoires distribués selon la loi uniforme sur $[0, 1]$ ou selon diverses autres lois de probabilité :

- binomiale, Poisson et multinomiale, pour les lois discrètes ;
 - Cauchy, exponentielle, gamma, normale, triangulaire et uniforme, pour les lois continues.
- Les possibilités de SAS en matière de simulation sont en fait très étendues.

Exercice 09.1

Faire exécuter le programme suivant :

```
data uni1;
x = ranuni(5);
run;
proc print;
run;
```

La valeur 5 est appelée *seed* (c'est-à-dire semence) ; elle est nécessaire pour calculer la première valeur simulée dans une série, chaque valeur étant ensuite calculée à partir de la précédente. Relancer le même programme une seconde fois et comparer. Changer ensuite la valeur 5 et comparer.

Nota : seules les valeurs entières sont admises pour semence ; toute valeur négative entraîne l'usage de l'horloge de la machine comme semence ; désormais, **on utilisera -1 comme semence**.

Exercice 09.2

Faire exécuter le programme suivant :

```
data uni2;
  do i=1 to 25;
    x = ranuni(-1);
    output;
  end;
run;
proc print; run;
```

Enlever la commande `output` pour bien comprendre son effet. Compléter ensuite le programme ci-dessus de façon à faire apparaître la moyenne des valeurs simulées. Recommencer alors avec 100, puis 10 000 observations. Comparer vos différents résultats et commenter.

Exercice 09.3

Recommencer l'exercice précédent en remplaçant la loi uniforme par la loi binomiale.

Indications : La commande SAS permettant de simuler la loi binomiale est `ranbin(seed,n,p)` ; on fera attention de distinguer n et le nombre désiré de valeurs simulées ; prendre $n = 200$ et essayer différentes valeurs de p .

Exercice 09.4

Recommencer encore le même exercice en utilisant maintenant la loi de Poisson.

Indications : La commande SAS permettant de simuler la loi de Poisson est `ranpoi(seed,lambda)` ; prendre $\lambda = 2.7$.

Exercice 09.5

On va maintenant synthétiser les résultats de la simulation précédente, d'abord par un tableau, ensuite par 2 graphiques.

1. Reprendre le programme de l'exercice précédent en enlevant le calcul de la somme et de la moyenne des valeurs simulées, ainsi que l'impression des résultats de la simulation, et en posant $\lambda = 1$.
2. Utiliser la procédure `freq` et la commande `tables` pour obtenir une présentation synthétique des valeurs simulées.
3. Réaliser le diagramme en bâtons de ces valeurs.
4. Réaliser leur diagramme cumulatif.

Indication : utiliser la commande `retain` qui permet, dans une procédure `data`, d'initialiser une nouvelle variable à une valeur donnée. Exemple : `retain cumul 0`;

5. Recommencer cet exercice en prenant $\lambda = 3$, puis $\lambda = 15$. Commenter.

Exercice 09.6

Faire maintenant un programme de simulation de la loi multinomiale. Pour cela, utiliser la commande : `rantbl(seed,p1,...,pk)`.

Commencer par une seule valeur simulée d'une loi à 3 catégories pour en comprendre le principe ; faire ensuite simuler 10 000 valeurs ; enchaîner enfin avec la procédure `freq` pour visualiser les résultats (ne pas faire de graphique dans ce cas).

Recommencer avec une loi à 5 catégories.

Que se passe-t-il si la somme des probabilités choisies est inférieure à 1 ? Si elle est supérieure à 1 ?

Exercice 09.7

Faire maintenant une simulation de 10 000 valeurs d'une loi normale $\mathcal{N}(0,1)$ (`rannor`). En étudier les résultats au moyen de SAS-INSIGHT ⁽¹⁾ (histogramme et résumés numériques).

Recommencer avec une loi normale de moyenne 40 et de variance 25.

Exercice 09.8

On va maintenant, au moyen de diverses simulations, illustrer le théorème *central limit*. Rappeler et commenter tout d'abord ce théorème.

Faire maintenant 1 000 simulations, chacune déterminant la moyenne de 10 simulations de lois uniformes sur l'intervalle $[0,1]$ (on doit donc faire 2 boucles `do` emboîtées). Centrer et réduire les moyennes ainsi obtenues. Étudier la distribution de ces moyennes avec SAS-INSIGHT (elle doit être proche de celle d'une loi normale réduite).

Recommencer le même exercice en remplaçant successivement 10 par 5, 3, 2 et 1. Apprécier la rapidité de la convergence de la loi uniforme vers la loi normale.

(1) Pour activer SAS-INSIGHT : (a) se placer dans la fenêtre *program editor* de SAS, cliquer sur le menu déroulant **Solutions**, ensuite sur le menu **Analysis**, puis sur le sous-menu **Interactive Data Analysis** : une fenêtre de dialogue (grise) doit alors s'ouvrir ; (b) dans la partie gauche de cette fenêtre, cliquer sur **work** (ouverture des tables SAS temporaires), puis, dans sa partie droite, sur le nom de la table SAS dans laquelle a été archivée la simulation de la loi normale, enfin sur **Open** : cette table apparaît dans une nouvelle fenêtre, tandis que la précédente disparaît ; (c) cliquer alors sur la variable X (variable simulée), puis sur le menu déroulant **Analyze**, enfin sur le menu **Distribution**.