

## M2 Statistique et Econométrie

### Travaux pratiques du cours Techniques de sondage

#### ENNONCÉ DU TP 3 : PLANS BE ET PPS AVEC R ET SAS

A : Algorithme pour sélectionner un échantillon Bernoulli

1. Créer une fonction (en R ou SAS) qui sélectionne un échantillon de Bernoulli de paramètre  $\pi$  dans une population de taille  $N$ , en utilisant l'algorithme suivant :  
on génère des variables aléatoires  $\varepsilon_1, \dots, \varepsilon_N$  iid de loi uniforme  $\simeq \mathcal{U}_{(0,1)}$ . L'échantillon  $s$  sera composé des éléments  $k$  pour lesquels  $\varepsilon_k < \pi$ .
2. Utiliser l'énoncé de l'exercice 2.3 (point 1) de la feuille 2 pour proposer une autre façon de sélectionner un échantillon de Bernoulli. Ecrire un suite le programme en R ou SAS.

B : Estimation du nombre total de logements vacants avec un plan PPS

Nous considérons l'extrait du recensement du 1999 présenté lors du premier tp. L'objectif est d'estimer le nombre total de logements vacants or le total de la variable LOGVAC.

1. Sélectionner dans cette population un échantillon de taille  $n = 50$  selon un plan proportionnel à la taille avec remise (PPS). Les probabilités de sélection  $p_k$  sont proportionnelles aux valeurs de la variable nombre de logements dénomée LOG.
2. Calculer une estimation de  $t_{LOGVAC}$  dans cet échantillon ainsi que sa variance.
3. Calculer l'estimation de  $t_{LOGVAC}$  pour 1000 simulations. Déterminer le coefficient de variation.

- (a) **Avec R** : La fonction R du package "sampling" qui sélectionne un échantillon proportionnel à la taille (PPS) s'appelle `UPmultinomial`. Cette fonction a un seul paramètre `pik`, le vecteur de taille  $N$  des probabilités d'inclusion et retourne un vecteur de taille  $N$  contenant le nombre de fois qu'une unité a été sélectionnée.

```
> pik=c(0.2,0.7,0.8,0.5,0.4,0.4)
> s.pps=UPmultinomial(pik)
> s.pps
[1] 0 1 1 0 0 1
```

Nous avons  $0.2 + 0.7 + 0.8 + 0.5 + 0.4 + 0.4 = 3$  ce qui veut dire que 3 tirages ont été effectués dans une population de taille 6 et les unités sélectionnées sont les unités 3,4 et 6.

- (b) **Avec SAS** : la proc `surveyselect` a l'option `method=pps-wr` pour sélectionner un échantillon avec remise et proportionnel à la taille. L'instruction `SIZE` est obligatoire dans cette situation et spécifie la variable auxiliaire qui sert pour construire les probabilités de sélection  $p_k$ .