TP N°1 : Estimation

L'objectif de ce TP est d'illustrer certaines notions sur l'estimation à l'aide de simulations et de données réelles. Pour se faire, on utilisera le logiciel Excel.

Exercice 1 — Comparaison d'estimateurs

Le fichier CPAM.xls comptabilise le nombre de personnes par heure attentant un renseignement à un guichet dans une agence de la CPAM sur la tranche horaire 10h–12h. Il y a en tout 10 guichets dans l'agence étudiée, d'où 10 échantillons. Pour chaque guichet, 500 heures ont été observées, d'où des échantillons de taille 500.

Si on note X la variable aléatoire représentant le nombre de personnes se présentant au guichet pendant une heure, on sait que X suit une loi de Poisson de paramètre inconnu λ.

L'objectif de cet exercice est d'estimer le paramètre λ afin de pouvoir gérer au mieux le temps d'attente des clients.

On sait que

$$E(X)=\lambda=\mathrm{var}(X)$$

On peut donc envisager deux estimateurs pour λ,

la moyenne : $T_1=\overline{X}$

la variance empirique : $T_2=S^2$

1) Partie théorique :
   Afin de confirmer les observations ci-dessus, il faut étudier les propriétés des estimateurs.
   (a) Les estimateurs sont-ils sans biais ?
   (b) Sont-ils convergent ?
   (c) Peut-on comparer leur risque quadratique ?

2) Partie empirique :
   a) Pour chaque échantillon, calculer la valeur de T_1 et de T_2 [NB] pour des sous-échantillons de taille n=10, n=50, n=100, n=250 et n=500. Pour chaque taille considérée, on obtient ainsi 10 réalisations de T_1 et T_2.
   b) Sur un graphique, représenter les valeurs obtenues pour T_1 en fonction de n. Que constatez-vous ? Faire la même chose avec T_2.
   c) Sur un graphique comparer la valeur moyenne de T_1 et T_2 en fonction de n et sur un autre graphique comparer l'écart-type de T_1 et T_2 en fonction de n. Que pouvez-vous en conclure ? Y-a-t'il un estimateur meilleur que l'autre ? Si oui, est-ce vrai quelle que soit la taille de l'échantillon ?

3) Partie application :
   On sait qu'un renseignement dure en moyenne 10 min. Quelle est la probabilité d'attendre plus d'une heure ?

[NB] Excel calcule la variance sans biais

Exercice 2 — Intervalle de confiance

Soit X une variable aléatoire d'espérance μ. Excel propose une fonction pour calculer l'intervalle de confiance pour μ.

```
INTERVALLE.CONFIANCE(alpha;ecarttype;taille)
```

Cette fonction suppose que la moyenne $\overline{X}$ suit une loi normale d'espérance μ et de variance σ^2/n où n est la taille de l'échantillon et σ^2 est la variance de X.

Elle n'est donc valable que dans le cas où

- l'échantillon est de grande taille
- l'échantillon est de petite taille, gaussien et de variance connue

et uniquement pour des intervalles de confiance symétrique.

Le fichier herbicide.xls (JB Lamy, support de cours) permet d'étudier l'efficacité de trois herbicides sur trois plantes (blé, chiendent et liseron). Le fichier présente le nombre de plants dans la culture avant l'expérience et le nombre de plants survivants 10 jours après, ainsi que le taux de survivants.

1) Construire un intervalle de confiance pour la moyenne du nombre de plants survivants et pour chaque traitement. Commenter.
2) Etudier l'amplitude de l'intervalle de confiance en fonction de la valeur de α.
3) Si on se restreint au blé n'ayant subi aucun traitement, pouvez-vous construire un intervalle de confiance ?

Exercice 3 — Simulation

Soit X une variable aléatoire de loi de Bernoulli de paramètre p.

1) Générer un échantillon de n réalisations de la variable X .
2) Construire un intervalle de confiance de niveau 5% pour le paramètre p.
3) Le paramètre p est-il dans l'intervalle de confiance ?
4) Répéter plusieurs fois les points précédents et compter le nombre de fois où p n'est pas dans l'intervalle.