

# Exploration Numérique 1

24/08/2020

Bienvenue dans cette première exploration numérique du cours de statistiques MAP433. Ces quelques questions vous permettront de réaliser des implémentations rapides de notions vues en amphi ou en PC.

**Modalités et notation :** Les énoncés des EN seront mis en ligne sur le Moodle du cours le vendredi soir. Les réponses doivent être déposées sur le Moodle au plus tard le mercredi de la semaine suivante, à 20h. Il y aura une EN par semaine de cours, hors les 3 semaines correspondant aux 3 semaines de remise des DM et la dernière semaine de cours (se reporter au calendrier en ligne pour le détail).

Le langage utilisé pour répondre aux questions sera, au choix, Python ou R. **Vous devrez insérer votre code et vos réponses dans un Notebook Jupyter (fichier .ipynb) à déposer sur le Moodle.** Chaque semaine, des trinômes seront tirés au sort et leur rendu sera évalué. Chaque trinôme sera noté exactement 1 fois. Tout non-rendu d'EN entraînera des points de malus. Les notes compteront pour la note finale du module selon le barème détaillé sur le Moodle.

## 1 Estimation paramétrique

Pour  $\lambda > 0$ , on rappelle l'expression de la loi de Poisson de paramètre  $\lambda$ , de densité  $\{p_\lambda(k), k \in \mathbb{N}\}$  par rapport à la mesure de comptage  $\mu$  sur  $\mathbb{N}$  :

$$p_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}$$

On cherche à modéliser le nombre de buts marqués par une équipe de football. Pour cela, nous reprenons les données obtenues sur les matches de *premier league*, pour 1900 matches. On rappelle les données obtenues pour l'équipe visiteuse en Figure 1. Pour simplifier, les scores observés supérieurs à 5 buts ont été regroupés.

1. Créer un tableau de données contenant les valeurs de la figure 1.
2. Calculer, à partir du tableau de données précédent, l'estimateur du maximum de vraisemblance  $\hat{\lambda}$  de l'intensité.
3. Déterminer l'estimateur  $\tilde{\lambda}$  de l'intensité  $\lambda$  basée sur le moment d'ordre 2.
4. Comparer sur un barplot, les observations, les prédictions  $p_{\hat{\lambda}}(k)$ ,  $k = 1, \dots, 5$  et  $p_{\tilde{\lambda}}(k)$ ,  $k = 1, \dots, 5$ .

Nombre de buts marqués	Nombre d'observations
0	692
1	680
2	335
3	131
4	51
5	11
Total	1900

FIGURE 1 – Nombre de buts marqués par l'équipe visiteuse.

Proposons un nouveau modèle pour expliquer le nombre de buts marqués par une équipe de football. On suppose que, lors du match, une équipe joue  $n$  actions, chacune ayant une probabilité  $p \in ]0, 1[$  de se concrétiser en but. Les hypothèses fortes prises dans ce modèle sont que :

- Les valeurs de  $n$  et  $p$  sont identiques pour chaque équipe et dans le temps,
- Les résultats de chaque action sont indépendantes.

Selon ce modèle, le nombre  $k$  de buts marqués par une équipe obéit à une loi binomiale de paramètres  $(n, p)$ , de densité  $\{p_{n,p}(k), k \in \mathbb{N}\}$  par rapport à la mesure de comptage  $\mu$  sur  $\mathbb{N}$  :

$$p_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \mathbb{N}$$

5. Supposons que le nombre d'occasions par match est connu,  $n = 30$ . Calculer le maximum de vraisemblance  $\hat{p}$  de  $p$  au regard des données obtenues.
6. Quelle relation remarquez-vous entre  $\hat{\lambda}$ ,  $\hat{p}$  et  $n$  ? Quelle est l'explication de cette relation ?