PSI* 2024/25 DL 5

Métaphore de la cantine

Pour vendredi 20 octobre.

La "métaphore de la cantine" a été utilisée pour proposer un modèle pour le problème suivant "dans un échantillon de n insectes (par exemple) pris au hasard, quelle est la loi du nombre d'espèces présentes ?". On souhaite proposer un modèle simple, décrit par un unique paramètre θ.

Soit θ un réel strictement positif. Des individus numérotés $1, 2, ..., n, ..$ arrivent successivement dans une salle de restaurant contenant une infinité de tables infiniment longues. Le premier individu s'assied.

Pour tout entier $k \geq 1$, lorsque l'individu $k+1$ arrive,

- soit il occupe une nouvelle table avec la probabilité $\frac{\theta}{k+\theta}$,
- ou sinon, il choisit au hasard un des k convives déjà attablés et s'assied à la même table,

Les choix successifs ne sont donc déterminés que par l'occupation des tables.

L'entier K_n désigne le nombre de tables occupées lorsque n convives sont installés et pour $1 \leq i \leq n$, on note :

$$q_{n,i} = \mathbb{P}(K_n = i)$$

1. Etude générale du modèle

a. Montrer, pour tout entier $n \geq 1$, que

$$q_{n+1,1} = \frac{n!}{(n+\theta)(n-1+\theta)...(1+\theta)}$$

b. Pour $2 \leq i \leq n+1$, trouver une relation entre $q_{n+1,i}, q_{n,i}$ et $q_{n,i-1}$.

c. Quelle est la probabilité que l'ensemble (infini) de tous les convives se loge sur une seule table ?

2. Polynôme associé

a. On définit le polynôme P_n donné par $P_n(x) = \sum\limits_{i=1}^{n} q_{n,i} x^i$. Donner une relation de récurrence vérifiée par (P_n)

b. En déduire que : $P_n(x) = \dfrac{L_n(\theta x)}{L_n(\theta)}$ où $L_n(x) = \prod\limits_{i=0}^{n-1}(x+i)$.

c. Exprimer l'espérance de K_n à l'aide de P_n, puis sous forme d'une somme qu'on ne cherchera pas à calculer.

d. Pour chaque i, indiquer comment exprimer $q_{n,i}$ à partir de P_n.

e. On cherche à retrouver plus directement la valeur de l'espérance. Montrer, à partir du modèle, que : $K_n = \sum\limits_{i=1}^{n} \varepsilon_i$ où les ε_i $i = 1, ..., n$ sont des variables aléatoires de Bernoulli dont on précisera les paramètres.

Retrouver par cette méthode la valeur de $\mathbb{E}[K_n]$.

f. En utilisant chacune des deux méthodes précédentes, calculer de deux façons la valeur de la variance de K_n.

Bon à savoir : si les v.a. sont indépendantes, la variance de la somme est la somme des variances ; ce sera (re)-prouvé un peu plus tard, avec une hypothèse un peu moins forte d'ailleurs.

3. Estimées

a. Montrer que

$$\int_0^n \frac{\theta}{\theta + x}\,\mathrm{d}x \leq \mathbb{E}\left[K_n\right] \leq 1 + \int_0^{n-1} \frac{\theta}{\theta + x}\,\mathrm{d}x$$

b. En déduire un équivalent de $\mathbb{E}\left[K_n\right]$ quand n tend vers $+\infty$.

c. Étudier la différence $\mathbb{V}\left[K_n\right] - \mathbb{E}\left[K_n\right]$ et en déduire un équivalent de $\mathbb{V}\left[K_n\right]$ quand n tend vers $+\infty$.

d. Montrer que pour $\varepsilon > 0$, $\lim\limits_{n \to +\infty} \mathbb{P}(|\frac{K_n}{\ln n} - \theta| \geq \varepsilon) = 0$.