ECG2B
Lycée Marcelin Berthelot

Année scolaire 2024-2025
Mathématiques

Programme de colle n°19
Semaine du 17/03

Probabilités : convergences et estimations

Généralités
Tous les résultats sur les variables discrètes et à densité sont à connaître.

Inégalités

- Inégalité de Markov pour une variable positive admettant une espérance.
- Inégalité de Bienaymé-Tchebychev pour une variable admettant une variance.

(démonstrations non exigibles).

Loi faible des grands nombres

- Suite de variables aléatoires indépendantes identiquement distribuées (notation : iid).
- Moyenne empirique : si $(X_1, \dots, X_n)$ sont n variables aléatoires, leur moyenne empirique est $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Espérance et variance de la moyenne empirique de n variables iid (dans le cas d'existence de ces quantités). Ceci doit être redémontré.
- Loi faible des grands nombres :
$$\forall \varepsilon > 0,\ \lim_{n\to+\infty} \mathbb{P}\left(|\overline{X_n} - \mathbb{E}(X)| \geqslant \varepsilon\right) = 0$$
(la terminologie de convergence en probabilité n'est pas au programme).

Convergence en loi

- Définition : (X_n) converge en loi vers X (notation : $X_n \xrightarrow{\mathscr{L}} X$) si, <u>en tout point x où la fonction de répartition F_X est continue</u>, on a :
$$\lim_{n\to+\infty} F_{X_n}(x) = F_X(x)$$
- Dans le cas où (X_n) et X sont à valeurs dans $\mathbb{Z}$, équivalence avec :
$$\forall x \in X(\Omega),\ \lim_{n\to+\infty} \mathbb{P}(X_n = x) = \mathbb{P}(X = x)$$
- NB : en vue de démontrer certaines convergences en loi, les étudiants doivent connaître sans hésitation la fonction de répartition d'une variable aléatoire constante égale à $a \in \mathbb{R}$.

Théorème Central Limite

- Définition de la moyenne centrée réduite de n variables iid d'espérance m et de variance $V = \sigma^2$:
$$\overline{X_n}^* = \frac{\overline{X_n} - \mathbb{E}\left(\overline{X_n}\right)}{\sqrt{V\left(\overline{X_n}\right)}} = \sqrt{n}\left(\frac{\overline{X_n} - m}{\sigma}\right)$$
- TCL : si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables iid admettant une variance, la suite $\left(\overline{X_n}^*\right)$ converge en loi vers une variable suivant $\mathscr{N}(0,1)$.

- Conséquences :

  - à « n grand », approximation de la loi de $\overline{X_n}$ par une loi normale.
  - Sous certaines conditions sur les paramètres, on a les approximations suivantes :
    - $\mathscr{B}(n,p) \simeq \mathscr{N}(np, npq)$
    - $\mathscr{P}(\lambda) \simeq \mathscr{N}(\lambda, \lambda)$.

  L'énoncé devra faire remarquer : « on admet qu'on peut approximer la loi ... par une loi normale » .

Estimateurs

Attention, le programme a beaucoup changé sur ce point. Tout le vocabulaire est devenu hors-programme, à l'exception de ce qui est ci-dessous.

- Définitions :
  Soit X une variable aléatoire suivant une loi $\mathscr{L}(\theta)$ de paramètre θ inconnu (à estimer) :

  - Un échantillon de la loi de X est une famille $(X_k)_{1\leqslant k\leqslant n}$ de variables aléatoires indépendantes, identiquement distribuées suivant la loi de X.
  - Une réalisation de cet échantillon est un n-uplet $(x_k)_{1\leqslant k\leqslant n}$ de réels, où $x_k = X_k(\omega)$ est la valeur prise par X_k sur l'expérience considérée.
    Ceci s'interprète comme les valeurs de X observées.
  - Si θ est le paramètre à estimer : estimateur de θ, de $g(\theta)$ (une variable aléatoire fonction quelconque de l'échantillon, indépendante de θ).
  - Pour bien se mettre d'accord sur la quantité à estimer, on pourra dire : « on considère T_n comme un estimateur de $g(\theta)$ » .

- Exemple fondamental : la moyenne empirique. Si X a pour espérance m et pour variance v, les étudiants doivent savoir redémontrer que $\mathbb{E}\left(\overline{X_n}\right) = m$ et $V\left(\overline{X_n}\right) = \dfrac{v}{n}$.

À partir d'ici : si T_n est un estimateur de $g(\theta)$, il s'agit surtout de contrôler $\mathbb{P}(|T_n - g(\theta)| > \varepsilon)$ (ou $\mathbb{P}(|T_n - g(\theta)| \leqslant \varepsilon)$) à l'aide de Bienaymé-Tchebychev ou de valeurs de Φ si, en conséquence du TCL, on approxime la loi de $\overline{X_n}^*$ par une loi normale.
Biais et risque quadratique ne sont plus au programme mais on peut aborder des problématiques de rapidité de convergence de l'estimateur, notamment en observant la vitesse à laquelle $V(T_n) \to 0$.

À partir de vendredi : intervalles de confiance

Intervalle de confiance pour le paramètre θ au niveau de confiance $1-\alpha$; intervalle de confiance asymptotique.
Exemples d'utilisation de Bienaymé-Tchebychev ou du TCL pour obtenir de tels intervalles.