PROGRAMME DE COLLE - SEMAINE N° 20: du 17/03/2025 au 21/03/2025

Connaissances minimales attendues

Chapitre 12 - Convergences et approximations

Se reporter à un programme de colle précédent

Chapitre 13 - Estimations

- n-échantillon, réalisation d'un n-échantillon ou n-échantillon réalisé ;
- Modèle statistique, espace des paramètres. Le « contexte formel de l'estimation » ;
- Estimateur, estimation (ponctuelle) ;
- Estimateur moyenne empirique (d'ordre $n \in \mathbb{N}^*$), notation $\overline{X_n}$;
- Estimateur variance empirique (d'ordre $n \in \mathbb{N}^*$), notation S_n^2 ;
- Estimateur écart-type empirique (d'ordre $n \in \mathbb{N}^*$), notation S_n ;
- **(HP ?)** Biais d'un estimateur, notation $b_\theta(T_n)$, interprétation, estimateur sans biais, exemples de « débiaisements affines » ;
- **(HP ?)** L'estimateur moyenne empirique $\overline{X_n}$ est un estimateur sans biais de l'espérance commune des variables aléatoires de l'échantillon $(X_1, \dots, X_n)$;
- **(HP ?)** Risque quadratique d'une estimateur, notation $r_\theta(T_n)$, interprétation, exemples de calculs ;
- **(HP ?)** Décomposition du risque quadratique en $b_\theta(T_n)^2 + V_\theta(T_n)$. Pour un estimateur sans biais, le risque quadratique coïncide avec la variance ;
- « Bilan » sur l'étude qualitative d'un estimateur
- Suite d'estimateurs ;
- **(HP ?)** Estimateur asymptotiquement sans biais ;
- **(HP)** S_n^2 est un estimateur biaisé de la variance commune des variables $X_1, \dots, X_n$, mais asymptotiquement sans biais. Estimateur sans biais construit à l'aide de l'estimateur biaisé S_n^2 (estimateur variance corrigée) ;
- **(HP ?)** Estimateur convergent : définition formelle, reformulation avec le vocabulaire de la convergence en probabilité ;
- **(HP ?)** Reformulation de la loi faible des grands nombres : l'estimateur moyenne empirique est un estimateur convergent de l'espérance commune des variables aléatoires de l'échantillon $(X_1, \dots, X_n)$;
- **(HP ?)** Une condition suffisante « pratique » de convergence d'un estimateur, induite par l'inégalité de Markov :
  Si le risque quadratique d'un estimateur existe et tend vers 0 lorsque l'on fait tendre la taille de l'échantillon n vers $+\infty$, alors cet estimateur est convergent ;

- Estimateur du maximum de vraisemblance :
  - introduction, motivation et principes ;
  - fonction de vraisemblance L_n dans le cas discret, fonction de vraisemblance L_n dans le cas continu ;
  - fonction de log-vraisemblance M_n ;
  - Estimateur du Maximum de Vraisemblance (EMV) : définition ;
  - Estimateur du Maximum de Vraisemblance dans le cas du modèle statistique $\{\ \mathcal{B}(p),\ p \in]0,1[\ \}$: mise en œuvre détaillée. L'Estimateur du Maximum de Vraisemblance coïncide avec l'estimateur moyenne empirique.
  - Estimateur du Maximum de Vraisemblance dans le cas du modèle statistique $\left\{\ \mathcal{P}(\lambda),\ \lambda \in \mathbb{R}_+^*\ \right\}$: mise en œuvre détaillée. L'Estimateur du Maximum de Vraisemblance coïncide avec l'estimateur moyenne empirique.
- Intervalle de confiance au niveau de confiance $1-\alpha$ / au risque α, intervalle de confiance réalisé (fourchette) ;
- Construction détaillée d'un intervalle de confiance à un niveau de confiance $1-\alpha$ donné, à partir d'un estimateur en utilisant l'inégalité de Bienaymé-Tchebychev (si l'estimateur envisagé ne possède pas de biais) ;
- Intervalle de confiance asymptotique : définition ;
- Construction détaillée d'un intervalle de confiance asymptotique de l'espérance commune m des X_k ($k \in \mathbb{N}^*$) à l'aide du Théorème Central Limite et de l'estimateur moyenne empirique $\overline{X_n}$:
  - dans le cas où la variance est connue ;
  - dans le cas où la variance n'est pas connue (en utilisant plusieurs lemmes **(HP)** dont le lemme de Slutsky).
- Comparaison des intervalles de confiance de l'espérance commune m des X_k ($k \in \mathbb{N}^*$) obtenus à l'aide de l'inégalité de Bienaymé-Tchebychev d'une part, à l'aide du Théorème Central Limite d'autre part (l'estimateur envisagé ici étant $\overline{X_n}$ dans les deux cas).

Chapitre 14 - Fonctions de deux variables réelles à valeurs réelles

- Ensemble $\mathbb{R}^2$;
- Distance euclidienne sur $\mathbb{R}^2$, notation $d((x_A, y_A), (x_B, y_B))$
- Propriétés vérifiées par la distance euclidienne d sur $\mathbb{R}^2$: d est positive, symétrique, vérifie l'inégalité dite triangulaire et la propriété dite de séparation ;
- Boule ouverte de $\mathbb{R}^2$, notation $B_o(\Omega, r)$ (boule ouverte de centre Ω et r) ;
- Boule fermée de $\mathbb{R}^2$, notation $B_f(\Omega, r)$ (boule fermée de centre Ω et r) ;
- Partie ouverte (ouvert) de $\mathbb{R}^2$, partie fermée (fermé) de $\mathbb{R}^2$, partie bornée (borné) de $\mathbb{R}^2$;
- Tout produit cartésien de deux intervalles de $\mathbb{R}$ ouverts / fermés / bornés est un ouvert / fermé / borné de $\mathbb{R}^2$;
- Fonction de deux variables réelles à valeurs réelles ;

- Domaine de définition d'une fonction de deux variables réelles à valeurs réelles ;
- Fonctions partielles associées à une fonction de deux variables réelles à valeurs réelles ;
- Applications coordonnées ;
- Fonction polynomiale de deux variables réelles à valeurs réelles ;
- Graphe d'une fonction de deux variables réelles à valeurs réelles : notion de surface/nappe. Notation Γ_f pour désigner la nappe représentant f ;
- Ligne de niveau k associée à une fonction de deux variables réelles à valeurs réelles f : définition, exemple, interprétations ;
- Fonction de deux variables réelles à valeurs réelles définie au voisinage de $(x_0, y_0) \in \mathbb{R}^2$;
- Limite (finie) d'une fonction de deux variables réelles en un point de $\mathbb{R}^2$, unicité de la limite ;
- Continuité d'une fonction de deux variables réelles à valeurs réelles en un point de $\mathbb{R}^2$;
- Continuité d'une fonction de deux variables réelles à valeurs réelles sur une partie incluse dans son domaine de définition ;
- Continuité sur $\mathbb{R}^2$ des fonctions coordonnées ;
- Opérations usuelles sur les fonctions continues de deux variables réelles : somme, produit, inverse ;
- Continuité des fonctions polynomiales de deux variables réelles ;
- Continuité de la composée $\varphi \circ f$ avec $f : A \subset \mathbb{R}^2 \longrightarrow I$ (fonction de <u>deux</u> variables réelles à valeurs réelles) continue et $\varphi : I \longrightarrow \mathbb{R}$ (fonction d'<u>une</u> variable réelle à valeurs réelles) continue

Savoir-faire et Méthodes à savoir appliquer

« Les incontournables »

- Les Savoir-Faire du programme précédent sont toujours valables.
- Préciser la nature topologique d'une partie de $\mathbb{R}^2$: s'agit-il d'un ouvert de $\mathbb{R}^2$? d'un fermé de $\mathbb{R}^2$? ni l'un ni l'autre ? s'agit d'une partie bornée de $\mathbb{R}^2$?
- Déterminer le domaine de définition d'une fonction de deux variables réelles à valeurs réelles et représenter ce domaine de définition (en faisant apparaître les éventuelles droites/courbes remarquables « à la frontière » du domaine) ;
- Justifier qu'une fonction de deux variables réelles à valeurs réelles est de classe continue sur son domaine de définition ;

Preuves exigibles

Propositions

1. Approximation de la loi binomiale par la loi de Poisson **[C]** ;
2. $\overline{X_n}^{\,*} = S_n^*$ **[C]** ;
3. Justification à l'aide du Théorème Central Limite de l'approximation de la loi binomiale par la loi normale **[C]** ;
4. Justification à l'aide du Théorème Central Limite de l'approximation de la loi de Poisson par la loi normale **[A]** ;
5. **(HP ?)** $\overline{X_n}$ est un estimateur sans biais de l'espérance **[C]** ;
6. **(HP ?)** Décomposition du risque quadratique en biais^2+variance **[C]** ;
7. **(HP)** S_n^2 est un estimateur biaisé de la variance commune des variables $X_1, \dots, X_n$, mais asymptotiquement sans biais. Estimateur sans biais construit à l'aide de l'estimateur biaisé S_n^2 (estimateur variance corrigée) **[C]** ;
8. **(HP ?)** Condition suffisante de convergence d'un estimateur impliquant le risque quadratique (sous réserve d'existence) **[C]** ;
9. **(HP ?)** $\overline{X_n}$ est un estimateur convergent de l'espérance **[C]** ;
10. **(HP)** S_n^2 est un estimateur convergent de la variance (à noter : tous les lemmes **(HP)** dont le lemme de Slutsky sont admis, mais devront être correctement énoncés lors de la restitution) **[A]** ;
11. **(HP)** S_n est un estimateur convergent de l'écart-type (à noter : tous les lemmes **(HP)** dont le lemme de Slutsky sont admis, mais devront être correctement énoncés lors de la restitution) **[A]** ;
12. Dans le cadre du modèle statistique $\{ \mathcal{B}(p),\ p \in]0,1[\}$, l'Estimateur du Maximum de Vraisemblance existe, est unique, et coïncide avec l'estimateur moyenne empirique $\overline{X_n}$ **[C]** ;
13. Dans le cadre du modèle statistique $\left\{ \mathcal{P}(\lambda),\ \lambda \in \mathbb{R}_+^* \right\}$, l'Estimateur du Maximum de Vraisemblance existe, est unique, et coïncide avec l'estimateur moyenne empirique $\overline{X_n}$ **[A]** ;
14. Construction d'un intervalle de confiance avec l'inégalité de Bienaymé-Tchebychev : le contexte et les grandes étapes **[Poly]** ;
15. Construction d'un intervalle de confiance asymptotique avec le Théorème Central Limite dans le cas où la variance est connue **[Poly]** ;
16. Construction d'un intervalle de confiance asymptotique avec le Théorème Central Limite dans le cas où la variance n'est pas connue (à noter : tous les lemmes **(HP)** dont le lemme de Slutsky sont admis, mais devront être correctement énoncés lors de la restitution). **[A]**.

Exemples/ exercices

1. A l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que l'intégrale $\int_0^{+\infty} (1 - \Phi(t))\, dt$ converge **[TD]** ;
2. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes suivant toutes la loi $\mathcal{U}([0,1])$. Alors, $(\min(X_1,\ldots,X_n))_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire certaine égale à 0 **[TD]** ;
3. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes suivant toutes la loi $\mathcal{E}(1)$. Alors, $(\max(X_1,\ldots,X_n) - \ln(n))_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire certaine égale à Z dont la fonction de répartition est $x \mapsto e^{-e^{-x}}$ (loi de Gumbel) **[TD]** ;
4. Si X est une variable aléatoire à densité, alors la suite $\left(Xe^{1/n}\right)_{n\in\mathbb{N}^*}$ converge en loi vers X **[C]**
5. Si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires telle que, pour tout $n \in \mathbb{N}^*$, X_n suit la loi uniforme à densité sur $\left[-\frac{1}{n}, \frac{1}{n}\right]$, alors $(X_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire certaine égale à 0 **[C]** ;
6. Si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires telle que, pour tout $n \in \mathbb{N}^*$, X_n suit la loi uniforme discrète sur $[\![0,n]\!]$, alors $\left(\frac{1}{n}X_n\right)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi uniforme à densité sur $[0,1]$ **[AC]** ;
7. Si, pour tout $n \in \mathbb{N}^*$, S_n suit la loi $\mathcal{B}(n,p)$, alors, $\lim\limits_{n\to+\infty} \mathbb{P}(S_n \leqslant np) = \frac{1}{2}$ **[C]** ;
8. « Théorème des 12 uniformes » : si, pour tout $k \in [\![1,12]\!]$, X_k est une variable aléatoire de loi $\mathcal{U}\left(\left[-\frac{1}{2}\ ,\ \frac{1}{2}\right]\right)$ et si les X_k ($k \in [\![1,12]\!]$) sont mutuellement indépendantes, alors $\sum\limits_{k=1}^{12} X_k$ suit approximativement la loi $\mathcal{N}(0,1)$ **[C]** ;
9. Approximation (à l'aide du Théorème Central Limite) de $\mathbb{P}(X \in [405,495])$ dans le cas où X suit la loi binomiale de paramètres $(900, 0.5)$ **[C]** ;
10. A l'aide du Théorème Central Limite, montrer que :
$$\sum_{i=0}^{n} \frac{n^i}{i!} \underset{n\to+\infty}{\sim} \frac{e^n}{2} \qquad \textbf{[TD]}$$
11. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes suivant la loi $\mathcal{B}(p)$, avec p paramètre inconnu que l'on cherche à estimer.
Alors, $\overline{X_n}$ est un estimateur convergent de p et,
de plus, pour $n \in \mathbb{N}^*$, pour tout $\alpha \in]0,1[$, $\left[\overline{X_n} - \frac{1}{2\sqrt{\alpha n}}\ ;\ \overline{X_n} + \frac{1}{2\sqrt{\alpha n}}\right]$ est un intervalle de confiance de p au niveau de confiance $1-\alpha$ **[C]** ;
12. Modèle statistique $\left\{\ \mathcal{U}([0,\theta]),\ \theta \in \mathbb{R}_+^*\ \right\}$, et estimateur $W_n = \max(X_1,\ldots,X_n)$.
    * W_n est un estimateur biaisé de θ. Construction d'un estimateur Z_n sans biais de θ à l'aide de W_n **[C]** ;
    * Z_n est un estimateur convergent de θ **[C]** ;
    * Construction, pour tout $\alpha \in]0,1[$, d'un intervalle de confiance de θ au niveau de confiance de $1-\alpha$ à l'aide de Z_n et de l'inégalité de Bienaymé-Tchebychev **[C]**.

13. Toute boule ouverte est une partie ouverte et bornée de $\mathbb{R}^2$ **[Ac]** ;

[A] : Annexe **Preuves** ;
[C] : Preuve traitée au tableau **en cours** ;
[Poly] : Preuve traitée dans le polycopié ;
[Ac] : Annexe **Corrections** ;
[TD] : Travaux Dirigés

Informatique

Tout le contenu des polycopiés :

- **TP1 - Cours (rappels)** et **TP1 - Exercices**
- **TP2 - Algorithmique de listes (rappels)** [calcul du (second) minimum (et indices associés), (second) maximum (et indices associés), des valeurs les plus proches (et indices associés), recherche dichotomique dans une liste triée, algorithmes gloutons (« plus grand nombre », « déplacement d'une grenouille », rendu de monnaie, problème des conférenciers), algorithmes de tri **(HP)** (tri-bulles et tri-par-insertion)].
- **TP3 - Calculs numériques** [algorithme de dichotomie, méthode de point fixe, méthode de Newton-Raphson **(HP)**, méthode des rectangles permettant le calcul approché d'une intégrale sur un segment**(HP)**, application au calcul des valeurs approchées des valeurs prises par la fonction de répartition de toute variable aléatoire de loi $\mathcal{N}(0,1)$ **(HP)**].
- **TP4 - Graphes (rappels)** [révision du vocabulaire et des théorèmes du programme, représentation d'un graphe avec Python (matrice numpy d'adjacence, liste des listes d'adjacence, dictionnaire), fonctions de conversion d'une représentation à une autre, divers exercices de manipulation de graphes (ordre, nombre d'arêtes, caractère adjacent de deux sommets, voisinage, degré, caractère isolé d'un sommet, caractère simple, complet, eulérien d'un graphe, calcul du nombre de chaînes reliant deux sommets donnés, calcul des distances combinatoires à un sommet dans un graphe simple, calcul de l'excentricité, du diamètre d'un graphe, algorithme de Dijkstra (calcul d'une chaîne de poids minimal)].
- **TP5 - Simulation d'expériences aléatoires - le cas discret (rappels et compléments)** [diagramme en bâtons (`bar`), la fonction `unique` (**(HP)**, à savoir donc implémenter à la main), matrices de booléens (pour compter le nombre de coefficients d'un tableau `numpy` vérifiant une contrainte donnée), les fonctions `random`, `randint`, `binomial`, `geometric`, `poisson` de la sous-bibliothèque `random` de la bibliothèque `numpy`, méthode de Monte-Carlo (estimation d'une espérance (moments, espérance de transformées), d'une probabilité, ...), divers exercices (tirages avec remise, sans remise, selon un protocole plus « exotique », marches aléatoires, graphes aléatoires d'Erdös-Renyi (estimation du nombre d'arêtes, du nombre de sommets isolés, de la probabilité qu'il soit connexe, etc))]
- **TP6 - Bases de données et langage SQL** [entité, table, base de données, attribut, descripteur, domaine, type, enregistrement, champ, schéma de table, schéma relationnel d'une base de données, clé primaire, clé étrangère, requête SQL de création de table (`CREATE TABLE ...`) (avec éventuellement déclaration des clés primaires et étrangères (`PRIMARY KEY(...)`, `FOREIGN KEY(...) REFERENCES ...`), insertion des données dans une table (`INSERT INTO ... VALUES`), suppression des données dans une table (`DELETE FROM ...`, `DELETE FROM ... WHERE`), suppression d'une table (`DROP TABLE ...`) **(HP)**, modification des données d'une table (`UPDATE ... SET ... WHERE ...`), projection sans contrainte (`SELECT ... FROM ...`), projection sans doublon (`SELECT DISTINCT ... FROM ...`), projection avec données triées dans l'ordre croissant/décroissant (`SELECT ... FROM ... ORDER BY ... ASC/DESC`), projection avec données tronquées (`SELECT ... FROM ... LIMIT ... OFFSET ...`), projection avec renommage (`AS`), restriction (projection avec contrainte) (`SELECT ... FROM ... WHERE ...`), produit cartésien de deux tables, jointure de deux tables (`SELECT ... FROM ... INNER JOIN ... ON ... = ....`), opérateurs ensemblistes (`UNION`, `INTERSECT`, `EXCEPT`), fonction d'agrégation (`COUNT(...)`, `SUM(...)`, `AVG(...)`, `MIN(...)` et `MAX(...)`), utilisation des fonctions d'agrégation sur des « groupes » (`GROUP BY .... HAVING ...`) (**(HP)**)]
- **TP7 - Chaînes de Markov** [simulation d'une trajectoire et représentations graphiques, calcul « exact » de la loi de X_n à l'aide des puissances de la matrice de transition (en utilisant les fonctions

`np.dot` et `al.matrix_power`), estimation de la loi de X_n à l'aide de la méthode de Monte-Carlo, mise en évidence de la convergence (sous réserve d'existence) de la chaîne de Markov, mise en évidence d'une loi stationnaire (ou bien grâce à la « loi limite » mise en évidence au niveau du point précédent, ou bien en donnant un vecteur propre de ${}^tM$ associé à la valeur propre 1, à coefficients positifs et dont la somme vaut 1 (utilisation des fonctions `al.eig` et `np.sum`)), simulation du premier l'instant où la chaîne de Markov est passée pour la kème fois par l'état initial, simulation du premier instant où la chaîne de Markov est passée au moins une fois par tous ses états, étude approfondie des exemples suivants : lancers de deux pièces suivant un protocole particulier, un modèle épidémiologique, un modèle simplifié de « bonus-malus » utilisé par les assurances automobiles, l'algorithme de *PageRank* de Google]

- **(Annexe Python - Chapitre 11)** [Utilisation de la fonction `odeint` de la sous-bibliothèque `integrate` de la bibliothèque `scipy` pour représenter des trajectoires d'une équation différentielle linéaire d'ordre 1, d'un système différentiel linéaire homogène d'ordre 1 et d'une équation différentielle linéaire homogène d'ordre 2]

- **TP8 - Simulations de variables aléatoires à densité** [histogramme non normalisé, histogramme normalisé (définitions formelles), la fonction `plt.hist` : utilisation « basique » et quelques options, lien (justifié) entre l'enveloppe d'un histogramme normalisé et une densité de probabilité, fonction de répartition empirique : définition, résultats de convergence (dont le théorème de Glivenko-Cantelli **(HP)**), simulation de réalisations à l'aide de Python, simulation d'échantillons de variables aléatoires indépendantes et identiquement distribuées selon une loi à densité usuelle : fonction `random`, `uniform` **(HP)**, `exponential` et `normal` de la sous-biblothèque `random` de la bibliothèque `numpy`, applications à des conjectures de lois, méthode d'inversion de simulation d'une variable aléatoire à densité : principe et résultat fondamental ($F_X(X)$ suit la loi $\mathcal{U}([0,1])$) et applications à la simulation des lois exponentielle, de Rayleigh **(HP)**, de Gumbel **(HP)** et de Pareto **(HP)**, méthode de Monte-Carlo d'estimation d'une intégrale : principe et exemples, illustrations du Théorème Central Limite]

- **TP9 - Estimations** [simulation et mise en évidence de la convergence des estimateurs usuels ($\overline{X_n}$, S_n^2, S_n), comparaison de deux estimateurs du même paramètre (exemple du modèle statistique $\left\{ \mathcal{U}([0,\theta]),\ \theta \in \mathbb{R}_+^* \right\}$ et des deux estimateurs $V_n = 2\overline{X_n}$ et $Z_n = \dfrac{n+1}{n} \max(X_1, \dots, X_n)$), simulation ((exemple du modèle statistique $\{ \mathcal{B}(p),\ p \in [0,1] \}$)) des intervalles de confiance naturellement induit par l'inégalité de Bienaymé-Tchebychev, par le Théorème Central Limite, comparaison de la qualité des intervalles de confiance]

sont au programme de cette semaine (y compris toutes les extensions hors-programme **(HP)**).
On pourra proposer aux étudiants des questions de cours d'informatique et/ou des exercices, en s'inspirant très fortement des exercices présents dans les polycopiés susmentionnés.

Quelques remarques destinées aux colleurs

- La colle commencera par quelques questions de cours (restitution d'une définition, d'une proposition/théorème, d'une ou des méthodes de base) ;
- On pourra demander aux élèves de prouver un des théorèmes (de manière complète ou incomplète) répertoriés dans la rubrique **Preuves exigibles** et tester la compréhension des élèves à propos de ce qu'ils écrivent ;
- On évitera autant que possible de consacrer plus d'une demi-heure aux questions de cours ;
- Les exercices proposés devront être de niveau progressif ;
- Peu d'exercices/exemples relatifs aux fonctions de deux variables réelles ont été abordés en cours ;
- On accordera un soin tout particulier à la rédaction et à la rigueur ;
- Toute erreur répertoriée dans le document Erreurs graves sera lourdement sanctionnée ;
- Si l'élève interrogé ne répond pas correctement aux questions de cours, on attribuera une note strictement inférieure à la moyenne, et ce indépendamment de la suite de l'interrogation orale.