Méthodes et résultats du cours 4M074

1. Illustrer la méthode de rejet pour la simulation des pseudo-variables aléatoires p. ex. pour la loi Gamma.
2. Illustrer la méthode de simulation d'un processus de Poisson homogène ou inhomogène par la méthode de thinning.
3. Illustration du théorème de Glivenko-Cantelli.
4. Illustrer l'approximation de la loi normale par la loi de Student. Plus précisément, on a

$$t_q \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \qquad q \to \infty,$$

   où t_q désigne la loi de Student à q degrés de liberté (cf. TD 6, Ex. 2) Ce résultat implique aussi la convergence de certaines caractéristiques de la loi de Student (comme la moyenne, variance, le coefficient d'asymétrie, le coefficient d'aplatissement (kurtosis), des quantiles) vers la moyenne/ variance/... de la loi normale standard.
5. Expliquer et illustrer le principe du bootstrap et les erreurs d'approximation associées.
6. Expliquer et illustrer la différence entre le bootstrap paramétrique et non paramétrique.
7. Illustrer des cas où le bootstrap ne marche pas et expliquer pourquoi.
8. Illustrer le résultat sur l'approximation de toute densité continue par un mélange gaussien (cf. Partie 2, Chapitre 3).

Du cours Statistique 4M015

9. Soient $X_i, \sim U[0,\theta], i = 1, 2, \dots$ i.i.d. Illustrer les différentes vitesses de convergence de l'estimateur de maximum de vraisemblance et de l'estimateur par la méthode des moments de θ (cf. cours *Statistique*, TD 3, Exercice 3 disponible sur le site `http://www.lsta.upmc.fr/guyader/statM1.html`).
10. Illustrer les performances des différents intervalles de confiance pour le paramètre de la loi de Bernoulli (IC asymptotique, par l'inégalité de Hoeffding et de Tchebychev) (Guyader (2017), Chapitre 1.3.2).
11. Illustrer la convergence des quantiles empiriques vers la quantile théorique quand la fonction de répartition F des données est strictement croissante (cf. Guyader (2017), Chapitre 2). Montrer qu'on n'a pas convergence, si F n'est pas strictement croissante.
12. Montrer que lorsque r est grand, la loi $\Gamma(r, \lambda)$ ressemble à une loi normale (Guyader (2017), p.51).
13. Expliquer la notion de la puissance d'un test. De quel facteurs dépendent la puissance d'un test ? Montrer l'allure typique de la puissance d'un test $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$ (bilatéral) ou $H_1 : \theta > \theta_0$ (unilatéral).
14. Illustrer la "compétition" du risque de première et de seconde espèce d'un test. Par exemple, pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ avec $\theta_0 < \theta_1$ considérer une région de test de la forme $\{\hat{\theta} > c\}$. Etudier comment le risque de première et de seconde espèce dépendent de la constante c.
15. Pour tester la significativité de plusieurs coefficients dans un modèle linéaire gaussien, il existe le test de Fisher et une procédure de test qui repose sur la procédure de Bonferroni (cf. Rebafka (2017), Chapitre 5). Comparer la qualité de ces deux tests (niveau et puissance des tests) dans différents scénarios. Lequel est meilleur ?
16. La théorie de l'analyse de la variance est développée sous des hypothèses fortes (loi gaussienne, homoscédasticité etc.) qui ne sont pas toujours vérifiées en pratique. Illustrer la robustesse (ou non robustesse) du test par rapport aux hypothèses du modèle (différentes variances par groupes ; des observations de loi discrète etc.)

Du cours statistique bayésienne 4M072

17. Illustrer le théorème de Bernstein-von Mises par exemple dans le cas de l'expérience de Bayes-Laplace : vraisemblance binomiale, a priori Beta(a, b) ; a posteriori beta qui s'approche d'une $\mathcal{N}(\hat{\theta}_{MV}, \frac{1}{n}I(\theta_0)^{-1})$. On peut jouer avec n et les paramètres a, b, (et possiblement θ_0) et en sortie la densité a posteriori et la distance L^1 (éventuellement approchée) entre les densités.

Modèles probabilistes avancés

18. Simuler un processus de branchement Z_t (en temps continu par exemple), dont l'espérance vaut $\mathrm{e}^{\alpha t}$ pour un certain $\alpha > 0$ (à calculer) et étudier la limite de la martingale $W_t := \mathrm{e}^{-\alpha t} Z_t$. Quand on simule plusieurs trajectoires indépendantes de Z, on voit bien que le comportement asymptotique est $\mathrm{e}^{\alpha t}$ mais avec un préfacteur différent à chaque fois (on peut passer en log aussi pour le voir encore mieux). La distribution empirique de ces préfacteurs doit converger vers la loi de W_∞…j'ai demandé à Amaury d'ajouter une réf

Méthodes statistiques avancées

19. Le k-means (Hastie et al. (2001), pp. 509) est une des méthodes de classification des observations en un petit nombre de groupes les plus utilisées en machine learning. Illustrer le k-means p. ex. sur des données bidimensionnelles d'un mélange gaussien

Matrices aléatoires j'y connais pas grande chose ; le dernier m'a l'air dur ; j'ai demandé à Thierry d'ajouter des réf

19. Si on tire une grande matrice hermitienne à coefficients gaussiens, la mesure empirique de ses valeurs propres est (avec grande probabilité presque) égale à la loi du demi-cercle de Wigner ;
20. Si on tire une grande matrice unitaire uniformément (par exemple en prenant une matrice à coefficients i.i.d. gaussiens et en prenant la partie unitaire de sa décomposition polaire, ou une des deux matrices unitaires de sa décomposition en valeurs singulières), ses valeurs propres se répartissent uniformément sur le cercle unité du plan complexe ;
21. Si on tire encore une matrice unitaire uniformément, mais pas forcément très grande, l'ensemble de ses valeurs propres forme un processus ponctuel déterminantal sur le cercle unité du plan complexe : c'est un processus ponctuel dont l'intensité est la mesure uniforme sur le cercle, mais avec une répulsion entre les points, qui fait qu'ils sont beaucoup plus également répartis qu'un échantillon i.i.d. de la même taille ; la probabilité d'avoir un gros paquet de points serrés dans un coin est infime ;
22. Si on fait partir un mouvement brownien de la matrice unité dans le groupe unitaire (c'est la solution d'un système simple d'EDS linéaires, ou d'une EDS linéaire encore plus simple dans l'espace des matrices), et qu'on regarde l'évolution de ses valeurs propres, on voit des particules qui bougent sur le cercle unité du plan complexe, qui partent toutes de 1, et qui se baladent un peu dans le cercle. Quant la taille de la matrice est grande, on voit un nuage qui remplit le cercle à vitesse finie, de manière symétrique par rapport à l'axe réel, et qui remplit tout le cercle exactement au temps 4.

Bibliographie

Guyader, A. (2017). Statistique – Partie 1. Polycopié, disponible sur le site `http://www.lsta.upmc.fr/guyader/statM1.html`.

Hastie, T., R. Tibshirani, and J. Friedman (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc.

Rebafka, T. (2017). Statistique – Partie 2. Polycopié, disponible sur le site `https://www.lpsm.paris//pageperso/rebafka/index.html`.

Gilles Pagès *Vincent pourrais-tu développer un peu et ajouter des réf ? ?*

1. Simuler le problème du collectionneur de coupons avec ou sans échange.
2. La ruine du joueur à pile ou face avec ou sans calcul approché de la proba que $S_{2n} = 0 \ldots$
3. Simule un problème de pricing de billets d'avion avec surbooking (cf. dans mon vieux livre "En passant par hasard") en y ajoutant une approximation de type TCL.
4. Simuler un arbre de Galton-Watson avec extinction des noms de famille, généralement avant celle de la population elle-même.
5. La recherche de quantile par algorithme stochastique et par inversion de la répartition empirique.
6. Box-Muller versus Marsaglia (polar method) pour la gaussienne ? Faire un course. Après il faut l'illustrer...
7. Simuler des processus gaussiens à temps discret par Choleski de la matrice de covariance. Application aux marches aléatoires browniennes et browniennes fractionnaires (pas besoin de savoir ce que sont les processus en question mais ils peuvent en profiter pour apprendre un peu). But : dessiner des marches en fonction de la constante de Hurst H.

Ismael Castillo

1. méthode MCMC par exemple random walk metropolis Hastings. là je vous laisse jouer toi et Vincent, mais cela doit bien marcher j'imagine dans pas mal d'exemples simples (cf. au besoin livre de Christian Robert)

Sonia Fourati *j'ai dit à Sonia qu'il faut des réf ; le 2e sujet me semble pas approprié*

1. Un premier resultat est une simulation de la loi beta comme loi limite d'une chaine de Markov (résultat non publie). ils ne connaissent pas le resultat mais ce doit leur etre relativement accessible pour eux.
2. un deuxieme resultat, qui donnera surement plus de belles images mais dont la preuve du resultat n'est pas accessible en M1 : c'est l'illustration de la propriété trajectorielle sur les processus de levy stables obtenu dans mon papier : "inversion de l'espace et du temps des processus de Lévy stables"

Damien Simon *Vincent pourrais-tu développer un peu et ajouter des réf ? ?*

1. modèle de Wright-Fisher de populations
2. ruine du joueur
3. un Monte-Carlo sur n'importe quel modèle de méca stat (Potts en champ moyen ? En dimension 2 ?)
4. Erdös Rényi et visualisation du graphe
5. amas de percolation et transition de phase
6. un exemple de file d'attente
7. statistiques de longueurs de cycles dans une permutation aléatoire uniforme (c'est simple à simuler malgré l'intitulé)

Amaury Lambert
Tabea

1. comparer différents tests nonparamétriques d'adéquation à une loi
2. Test de Kruskal-Wallis (comparé à ANOVA)
3. convergence du spectre de la matrice de covariance...