Questions de cours

L'examen à mi-parcours porte sur les chapitres 1 à 3.

L'examen final porte sur les chapitres 2 à 5.

Chapitre 1 : rappel des probabilités

1. Définir la convergence presque sûr d'une suite de variables aléatoires et énoncer la loi forte des grands nombres.
2. Définir la convergence en loi d'une suite de variables aléatoires et énoncer le théorème central limite.
3. Enoncer le lemme de Slutsky.
4. Donner les relations [1] entre les 4 types de convergences de suites de variables aléatoires suivants : convergence presque-sûr, convergence dans $\mathbb{L}^2$, convergence en probabilité, convergence en loi.
5. Enoncer le premier théorème de continuité. Démontrer les assertions de ce théorème concernant la convergence presque-sûr.
6. Enoncer le premier théorème de continuité. Démontrer les assertions de ce théorème concernant la convergence en probabilité dans le cas où la limite est une valeur déterministe.
7. Enoncer le premier théorème de continuité. Démontrer les assertions de ce théorème concernant la convergence en loi.
8. Enoncer le deuxième théorème de continuité dans le cas multidimensionnel.
9. Démontrer la version suivante du deuxième théorème de continuité unidimensionnel. Soit I un intervalle de $\mathbb{R}$ et $g : I \to \mathbb{R}$ une fonction continûment différentiable. Si $(X_i)_{i\geqslant 1}$ est une suite de variables aléatoires à valeurs dans I telle que $\sqrt{n}(X_n - a) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \sigma^2)$ où $a \in \mathbb{R}$ et $\sigma \geqslant 0$ sont deux constantes, alors $\sqrt{n}(g(X_i) - g(a)) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, v^2)$ où $v = \sigma g'(a)$.

Chapitre 2 : Echantillonnage et méthodes empiriques

10. Définir la fonction de répartition empirique $\widehat{F}_n(t)$ et donner la loi de la variable aléatoire $n\widehat{F}_n(t)$ dans le cas où les observations sont iid.
11. Définir la fonction de répartition empirique $\widehat{F}_n(t)$ et donner les propriétés asymptotiques de la suite $(\widehat{F}_n(t))_{n\geqslant 1}$ lorsque $n \to \infty$, dans le cas où les observations sont iid.
12. Enoncer le théorème de Glivenko-Cantelli.
13. Démontrer le théorème de Glivenko-Cantelli pour les fonctions de répartition continues. (Ici, on peut se contenter de la preuve donnée dans le polycopié, qui est moins détaillée que celle donnée en amphi.)
14. Définir la moyenne et la variance empiriques des observations $X_1, \ldots, X_n$. Dans le cas où les observations sont iid, démontrer que la moyenne et la variance empiriques convergent presque sûrement lorsque $n \to \infty$. On précisera les limites et les conditions qui garantissent chacune de ces convergences.
15. Définir la moyenne $\bar{X}_n$ et la variance S_n^2 empiriques des observations $X_1, \ldots, X_n$. Dans le cas où les observations sont iid $\mathcal{N}(\mu, \sigma^2)$, que peut-on dire de la loi de $(\bar{X}_n, S_n^2)$?
    (*La réponse attendue est que* $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, $nS_n^2 \sim \sigma^2\chi^2_{n-1}$ *et* $\bar{X}_n$ *est indépendant de* S_n^2.)
16. Démontrer que si $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, alors $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
17. Démontrer que si $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, alors $nS_n^2 \sim \sigma^2\chi^2_{n-1}$.

1. Il s'agit de préciser quelle convergence implique quelle autre convergence. Par exemple, la convergence presque-sûr implique la convergence en probabilité.

18. Démontrer que si $X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, alors $\bar{X}_n$ et S_n^2 sont indépendants.

Chapitre 3 : Estimation de paramètres

19. Donner la définition d'un modèle statistique paramétrique.
20. Donner la définition d'un modèle identifiable. Donner un exemple de modèle identifiable et un exemple de modèle non identifiable.
21. Donner la définition d'un modèle dominé et d'un modèle discret. Donner un exemple de modèle statistique $\{P_\theta : \theta \in [0,1]\}$ qui n'est pas discret, alors que toutes les lois P_θ sont discrètes.
22. Donner la définition d'un estimateur sans biais et d'un estimateur fortement consistant. Donner un exemple d'un estimateur qui a ces deux propriétés.
23. Donner la définition d'un estimateur sans biais et d'un estimateur fortement consistant. Donner un exemple d'un estimateur sans biais qui n'est pas fortement consistant.
24. Donner la définition d'un estimateur sans biais et d'un estimateur fortement consistant. Donner un exemple d'un estimateur fortement consistant mais biaisé.
25. Soit $\{P_\theta : \theta \in \Theta\}$ un modèle paramétrique avec $\Theta \subset \mathbb{R}^p$ un ouvert. On sait que pour une fonction $g : \mathcal{X} \to \mathbb{R}^p$, la fonction $M : \Theta \to \mathbb{R}^p$ définie par $M(\theta) = \mathbf{E}_\theta[g(X_1)]$ est continue et injective. Proposer un estimateur fortement consistant du paramètre θ. Justifier la consistance forte de cet estimateur.
26. Définir le risque quadratique d'un estimateur d'un paramètre p-dimensionnel θ. Quelle est la décomposition biais-variance de ce risque ?
27. Montrer que le risque $R(\bar{\theta}_n, \theta)$ d'un estimateur $\bar{\theta}_n$ d'un paramètre θ vérifie

$$R(\bar{\theta}_n, \theta) = \|\mathbf{E}_\theta[\bar{\theta}_n] - \theta\|^2 + \mathbf{E}_\theta\|\bar{\theta}_n - E_\theta[\bar{\theta}_n]\|^2. \tag{1}$$

Chapitre 4 : Estimateur du maximum de vraisemblance

28. Donner la définition de la vraisemblance, de la log-vraisemblance et de l'estimateur du maximum de vraisemblance (dans un modèle dominé).
29. Donner un exemple de modèle dominé et identifiable où l'EMV n'est pas unique.
30. Donner un exemple de modèle dominé où l'EMV n'existe pas.
31. Enoncer et prouver le résultat portant sur la convergence de la log-vraisemblance négative et le fait que sa limite est minimisée en θ^*.
32. Enoncer le résultat concernant la consistance de l'EMV.
33. Enoncer les hypothèses de régularité d'un modèle statistique. On pourra remplacer la formulation rigoureuse de l'hypothèse 3 par une explication informelle.
34. Définir le score d'un modèle statistique, puis l'information de Fisher. Montrer que dans un modèle régulier le score est centré.
35. Montrer que si la 3e hypothèse de régularité est vérifiée, alors l'information de Fisher peut être calculée par la formule $I(\theta) = -\mathbb{E}_\theta[\ell''(X, \theta)]$ où $\ell(x, \theta) = \log f(x, \theta)$, $f(x, \theta) = \frac{\mathrm{dP}_\theta}{\mathrm{d}\mu}(x)$ et la dérivée seconde est par rapport à θ.
36. Enoncer le théorème concernant la normalité asymptotique de l'EMV. Donner un exemple où la première hypothèse de régularité n'est pas satisfaite et la conclusion du théorème est fausse.
37. Prouver le théorème de la normalité asymptotique de l'EMV. Les passages les plus techniques peuvent être admis sans démonstration.
38. Enoncer et prouver l'inégalité de Cramér-Rao.
39. Donner la définition d'un estimateur asymptotiquement efficace (dans la famille des estimateurs asymptotiquement normaux). Énoncer le théorème portant sur l'efficacité asymptotique de l'EMV.

Chapitre 5 : Estimateur Bayesien

40. Définir le risque intégré (par rapport à une mesure a priori) et l'estimateur bayésien pour le risque quadratique.
41. Définir la mesure a posteriori et donner la formule de l'estimateur bayésien qui fait intervenir la mesure a posteriori. Donner un exemple de modèle statistique où l'estimateur bayésien est calculable de manière explicite.
42. Prouver que l'estimateur bayésien est l'espérance par rapport à la mesure a posteriori.
43. Donner la définition d'une famille de lois conjuguées pour le modèle $\{P_\theta : \theta \in \Theta\}$, en donner un exemple.
44. Enoncer le théorème portant sur la normalité asymptotique et l'efficacité asymptotique de l'estimateur bayésien.