Liste de questions auxquelles il faut savoir répondre

1. Quelle est la densité de la loi normale ?
2. Quelle est la définition d'une loi de Student ?
3. Quelle est la définition d'une loi de Fisher ?
4. Quelle est la définition d'une loi du Chi-deux ? Donner sa densité.
5. Soit T_n une loi de Student à n paramètres. Quelle est la limite en loi de T_n ? Donner une approximation des quantiles de T_n dans ce cadre.
6. Quelle est la limite en loi de X_n/n où $X_n \sim \chi^2(n)$?
7. Donner un intervalle de confiance pour la moyenne d'un échantillon iid gaussien (variance connue).
8. Même question lorsque la variance n'est pas connue.
9. Démontrer que $\hat{\sigma}_n^2 = (n-1)^2 \sum_{i=1}^n (X_i - \bar{X})^2$ n'est pas biaisé.
10. Dans un modèle exponentiel, qu'est-ce que la fonction de partition du modèle ? Exprimer son gradient et sa matrice hessienne.
11. Lorsque les X_i sont gaussiens, quelle est la loi de $\hat{\sigma^2}_n$?
12. On observe $(X_1, \ldots, X_n)$ de loi iid uniforme sur $[0, b]$. Donner l'estimateur du maximum de vraisemblance pour b. Calculer son risque quadratique.
13. Énoncer le lemme de Slutsky.
14. Donner un intervalle de confiance asymptotique pour la moyenne d'un échantillon iid $(X_1, \ldots, X_n)$, dont on ne connaît pas la variance (on sait qu'elle existe).
15. On observe $(X_1, \ldots, X_n)$ sur un ensemble fini $\mathcal{X}$. Proposer un test d'adéquation de leur loi à la loi uniforme sur $\mathcal{X}$.
16. On observe $(X_1, \ldots, X_n)$ dans un modèle gaussien $\mathcal{N}(\mu, 1)$. On veut tester $H_0 : \{\mu = 0\}$ contre $H_1 : \{\mu = 1\}$. Formuler le test de la moyenne et calculer sa puissance.
17. En 1980, une proportion $p_1 = 30.6\%$ des $n_1 = 55$ millions de ressortissants français avaient moins de 20 ans. En 2000, cette proportion est $p_2 = 25,6\%$ et la France compte alors $n_2 = 61$ millions citoyens français avaient moins de 20 ans. Tester l'hypothèse selon laquelle la proportion des moins de 20 ans est restée inchangée.
18. Décrire la méthode des moments dans un modèle exponentiel.
19. Donner un exemple de test de niveau α et de puissance nulle.

20. Quelle est l'espérance d'une loi $\chi^2(1)$? sa variance ?

21. Que dire d'un estimateur dont le risque quadratique tend vers zéro ?

22. Énoncer le lemme de Neyman-Pearson. Est-il toujours vrai pour les tests d'hypothèses composites ?

23. Donner un exemple d'estimateur du maximum de vraisemblance qui a de mauvaises propriétés.

24. Donner la définition et l'interprétation du r^2 d'une régression linéaire.

25. Écrire un modèle de Bernoulli sous forme exponentielle.

26. On observe $X \sim \text{Bin}(N, 1/2)$. Estimer N par la méthode des moments.

27. Qu'est-ce qu'un quantile ? Pour $X \sim \mathscr{N}(0,1)$, montrer que $\mathbf{P}(X > t) \leqslant (t\sqrt{2\pi})^{-1}e^{-t^2/2}$. Donner une approximation du quantile d'ordre α quand α est très petit.

28. Énoncer l'inégalité de Hoeffding pour des variables de Bernoulli.

29. Définir la puissance d'un test d'hypothèses composites.

30. Écrire la formule de l'estimateur des MCO d'un modèle linéaire gaussien de type $Y = \beta X + \varepsilon$ avec $\varepsilon_i \sim \mathscr{N}(0, \sigma^2)$, iid. Donner sa loi.

31. Donner l'expression et la loi de l'estimateur $\hat{\sigma}_n^2$ de la variance du bruit dans un modèle linéaire gaussien $Y = X\beta + \varepsilon$. Expliquer pourquoi il est indépendant de $\hat{\beta}$.

32. Donner un exemple de modèle statistique non identifiable.

33. Tester la significativité d'un coefficient dans une régression linéaire.

34. Qu'est-ce qu'une p-valeur ?

35. Donner la fonction caractéristique et tous les moments de la loi normale centrée réduite.

36. Quelle est la loi de la p-valeur sous l'hypothèse nulle (lorsqu'elle est simple) ?

37. À quoi sert la delta-méthode ? Donner un exemple simple d'utilisation.

38. Dans le cadre du modèle linéaire $Y = X\beta + \varepsilon$, on demande souvent à $X^\top X$ d'être inversible. Pourquoi ? Que dire, si ce n'est pas le cas ? Et que faire ?

39. Donner deux exemples d'inégalités de concentration et donner les utiliser pour obtenir deux intervalles de confiance non-asymptotiques sur p, ayant observé un échantillon iid $X_1, \ldots, X_n$ de lois $\text{Ber}(p)$.

40. Comparer la longueur des intervalles de confiance pour la moyenne d'un échantillon iid $\text{Ber}(p)$ obtenus avec l'inégalité de Bienaymé-Chebychev, l'inégalité de Hoeffding, le TCL.

41. On observe deux échantillons $(X_1, \ldots, X_n)$ et $(Y_1, \ldots, Y_m)$ de lois $\mathscr{N}(\mu_1, \sigma_1^2)$ et $\mathscr{N}(\mu_2, \sigma_2^2)$, tous les paramètres étant inconnus. Tester $H_0 : \{\sigma_1 = \sigma_2\}$.

42. Qu'est-ce que l'information de Fisher ? Donner une interprétation dans le cadre des modèles exponentiels.

43. Énoncer le théorème de Cochran et donner un exemple d'utilisation.

44. Énoncer la borne de Cramér-Rao.

45. À quoi sert la statistique de Kolmogorov-Smirnov ?

46. On possède un échantillon iid d'une loi diffuse. Donner un estimateur consistant de sa fonction de répartition.

47. Expliquer le lien entre la méthode d'estimation par EMV et la divergence de Kullback-Leibler.

48. Écrire l'EMV du paramètre d'un échantillon de lois de Poisson.

49. Écrire l'EMV du paramètre de position d'un échantillon de lois de Laplace.

50. Dans une régression $Y = X\beta + \varepsilon$, tester $\beta \in V$ pour un sous-espace vectoriel V.

51. Montrer pourquoi l'EMV dans un modèle exponentiel[1] est asymptotiquement normal.

52. Formuler un test ANOVA d'égalité des moyennes entre deux sous-échantillons.

53. On observe $X_1, \ldots, X_n$ iid de loi $\mathscr{N}(\mu, \sigma^2)$. Quelle est la loi de $\bar{X}_n$? Quelle est la loi de $n^{-1} \sum_{i=1}^n (X_i - \mu)^2$? Quelle est la loi de $(n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$? Pourquoi cette différence ? À quoi sert-elle ?

54. Donner la formule pour $\hat{a}_0, \hat{a}_1$ dans la régression linéaire $y_i = a_0 + a_1 x_i + \varepsilon_i$. On observe une nouvelle variable explicative x; donner un intervalle de confiance pour y.

[1] Sous les bonnes hypothèses !