# Eléments de statistique

## Répétition 4

**29 novembre 2016**

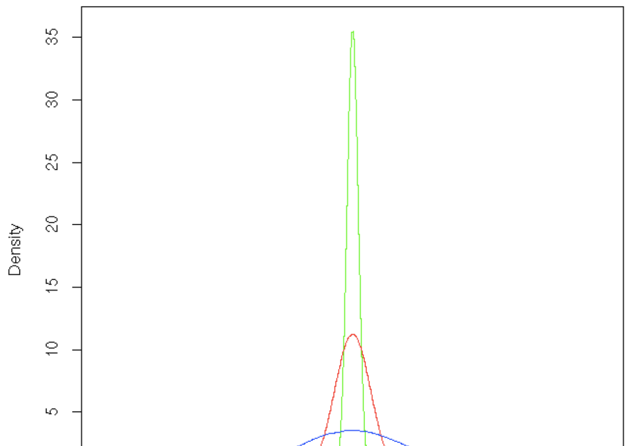
## Estimation

Le but de cette répétition est de mieux comprendre les avantages et inconvénients de la méthode du maximum de vraisemblance et de la démarche bayesienne.

## Exercices

4.1. Vous étudiez une pièce de monnaie réelle. Vous vous attendez à ce que la pièce soit quasi-équilibrée et donc que la probabilité de tomber sur pile soit proche de 50%, mais pas d'exactement $\frac{1}{2}$. Utilisez la méthode du maximum de vraisemblance d'une part et l'approche bayesienne d'autre part, pour estimer la proportion de fois où la pièce va tomber sur pile. Vous vous basez sur un échantillon i.i.d. de 20 lancés dont le résultat est de 12 piles et 8 faces. Dans le cas de l'approche bayesienne, vous supposerez que la proportion de pile suit a priori une loi $\beta$ dont vous déterminerez les paramètres qui vous semblent les plus réalistes.

Remarque : la figure suivante sera expliquée en répétition :




4.2. Construisez un intervalle de confiance à 95% de la proportion de fois où la pièce de l'exercice 4.1 va tomber sur pile, en partant du même échantillon. Utilisez à nouveau d'une part la méthode du maximum de vraisemblance et d'autre part la démarche bayesienne avec comme hypothèse a priori une loi $\beta$.

# Tests d'hypothèses

Un test d'hypothèse est une démarche scientifique consistant à évaluer une hypothèse statistique en fonction d'un jeu de données. Le fait qu'on ne dispose que d'un nombre fini d'observations implique que le résultat du test n'a jamais une garantie à 100% d'être correct. On peut en effet faire deux types d'erreurs :

— *erreur de première espèce* : rejeter l'hypothèse lorsqu'elle est vraie. On nomme la probabilité de commettre une telle erreur, le risque de 1ère espèce. On la note $\alpha$ et on parle d'un résultat *faux positif*.

— *erreur de deuxième espèce :* ne pas rejeter l'hypothèse alors qu'elle est fausse. On nomme la probabilité de commettre une telle erreur, le risque de 2ème espèce. On la note $\beta$ et on parle d'un résultat *faux négatif*.

# Exercices

4.3. Un agent d'assurance affirme que le montant moyen d'indémnisation payé aux victimes d'un incendie total de leur maison est de 135000 €. Un débiteur de police veut vérifier cette affirmation et dispose pour ce faire d'un échantillon de 16 dossiers de cas d'indémnisations de ce type. La moyenne des montants versés est de 141450 €. En supposant que $\sigma = 24000$ €, tester l'affirmation du courtier avec $\alpha = 0,05$, en supposant que les indémnisations suivent une loi normale.

4.4. Un producteur belge veut s'assurer que le volume moyen de bière contenue dans les canettes produites par sa firme est bien de 341ml. On souhaite tester si le volume moyen n'est pas significativement inférieur à 341 ml, auquel cas les clients risqueraient de se plaindre. La moyenne d'un échantillon de 36 canettes est de 332 ml. En supposant $\sigma = 6$ ml, effectuer le test d'hypothèse approprié avec $\alpha = 0,01$.

# Exercices suggérés

4.5. Répétez l'exercice 4.1 mais en utilisant un échantillon de taille 100 pour lequel il y a eu 60 fois pile et 40 fois face. Qu'observez-vous ? Interprétez. Essayez aussi de modifier l'hypothèse a priori.

4.6. Répétez l'exercice 4.2 mais en utilisant un échantillon de taille 100 pour lequel il y a eu 60 fois pile et 40 fois face. Qu'observez-vous ? Interprétez. Essayez aussi de modifier l'hypothèse a priori.

4.7. EXAMEN AOUT 2014 & JANVIER 2015
Vous étudiez un dé. Vous vous attendez à ce qu'il soit quasi-parfait et donc que la probabilité de tomber sur 6 soit proche de $\frac{1}{6}$ mais pas d'exactement $\frac{1}{6}$. Vous vous basez sur un échantillon i.i.d. de 120 lancés. Le dé est tombé 24 fois sur la valeur 6 et 96 fois sur une autre valeur.

a) Utilisez la méthode du maximum de vraisemblance pour estimer la proportion de fois où le dé va tomber sur 6. (**4 points**)

b) Utilisez la méthode du maximum de vraisemblance pour construire un intervalle de confiance à 95% de la proportion de fois où le dé va tomber sur 6. (**4 points**)

c) Utilisez l'approche bayesienne pour estimer la proportion de fois où le dé va tomber sur 6. Vous supposerez que la proportion de 6 suit a priori une loi $\beta(10, 50)$. (**6 points**)

d) Si vous deviez utiliser l'approche bayesienne pour construire un intervalle de confiance à 95% de la proportion de fois où le dé va tomber sur 6, à quel moment du raisonnement auriez-vous besoin d'une aide logicielle pour continuer ? Quelle équation poserait problème ? (**4 points**)

e) L'intervalle de confiance qui serait obtenu au point précédent suite à l'aide logicielle, serait-il forcément plus précis que l'intervalle de confiance obtenu à l'aide de la méthode du maximum de vraisemblance ? Justifier. (**2 points**)

4.8. EXAMEN JANVIER 2014 & AOUT 2015

La loi de décroissance radioactive prédit le nombre d'atomes, d'une substance radioactive donnée, non désintégrés à un temps t et sa décroissance avec le temps [1]. Un radionucléide quelconque a autant de chances de se désintégrer à un moment donné qu'un autre radionucléide de la même espèce, et la désintégration ne dépend pas des conditions physico-chimiques dans lesquelles le nucléide se trouve. On peut modéliser la probabilité qu'un atome se soit désintégré en maximum $T$ jours par

$$\boxed{\int_0^T \lambda e^{-\lambda t} dt = 1 - e^{-\lambda T}} \qquad \text{(où } \lambda \text{ dépend de la substance radioactive étudiée)}$$

Un laboratoire A reçoit une substance radiactive inconnue pour analyse. Il prépare un échantillon de 100 atomes qu'il envoie par colis sécurisé à un laboratoire B partenaire. Celui-ci reçoit le colis exactement 7 jours plus tard et constate qu'il ne reste plus que 54 atomes (46 atomes se sont désintégrés entre-temps).

a) Utilisez la méthode du maximum de vraisemblance pour estimer $\lambda$.
(**12 points**)
<u>Suggestion :</u> Poser $y = e^{-7\lambda}$

b) Utilisez la méthode du maximum de vraisemblance pour construire un intervalle de confiance à 95% de la proportion de fois où un atome de la substance étudiée se désintégre en maximum 7 jours. (**3 points**)

Après avoir envoyé le colis, le laboratoire A réalise différents tests qui l'amènent finalement à penser que la substance radioactive étudiée a de fortes chances d'être de l'iode. La période radioactive [2] de l'iode, c'est-à-dire la durée à l'issue de laquelle un atome d'iode a une chance sur deux de se désintégrer, est de 8,0207 jours.

c) Calculez $\lambda_{iode}$, la valeur théorique de $\lambda$ pour l'iode. (**1 point**)

1. dans cet exercice le temps est exprimé en jours
2. La période radioactive est aussi appellée la demi-vie

Etant donné que la substance a de fortes chances d'être de l'iode, mais que nous n'en sommes pas sûrs, nous décidons d'utiliser la démarche bayesienne sur base de l'hypothèse a priori que $\lambda$ suit une loi normale de moyenne $\lambda_{iode}$ et d'écart-type 1. Que peut-on affirmer sans faire le moindre calcul ?

d) que la valeur de $\lambda$ estimée sera plus grande qu'au point a) ? (**2 points**)

e) que l'intervalle de confiance à 95% sera plus précis qu'au point b) ? (**2 points**)

4.9. EXAMEN SEPTEMBRE 2016

L'assistant du cours de statistique du professeur Wehenkel s'intéresse à la motivation des étudiants inscrits au cours. Il se demande notamment quelle est la proportion d'étudiants qui ne présentent pas l'examen de première session [3]. Il se base sur un échantillon i.i.d. de 164 étudiants et constate qu'exactement 12 d'entre-eux n'avaient pas présenté l'examen de première session.

a) Utilisez la méthode du maximum de vraisemblance pour estimer la proportion d'étudiants qui ne présentent pas l'examen de première session du cours de statistique. (**4 points**)

b) Utilisez la méthode du maximum de vraisemblance pour construire un intervalle de confiance à 95% de la proportion d'étudiants qui ne présentent pas l'examen de première session du cours de statistique. (**4 points**)

c) Utilisez l'approche bayesienne pour estimer la proportion d'étudiants qui ne présentent pas l'examen de première session du cours de statistique. Vous supposerez que la proportion demandée suit a priori une loi $\beta(10, 100)$. (**6 points**)

d) Si vous deviez utiliser l'approche bayesienne pour construire un intervalle de confiance à 95% de la la proportion d'étudiants qui ne présentent pas l'examen de première session du cours de statistique, à quel moment du raisonnement auriez-vous besoin d'une aide logicielle pour continuer ? Quelle équation poserait problème ? (**4 points**)

e) L'intervalle de confiance qui serait obtenu au point précédent suite à l'aide logicielle, serait-il forcément plus précis que l'intervalle de confiance obtenu à l'aide de la méthode du maximum de vraisemblance ? Justifier. (**2 points**)

4.10. Un producteur de remorques prétend que son modèle de base supporte un poids de 800 kg avec un écart-type de 50 kg. Au seuil de 1%, tester l'hypothèse $\mu = 800kg$ versus l'hypothèse alternative $\mu \neq 800kg$, en sachant qu'un échantillon de 50 remorques a été testé et que le poids moyen supporté fût de 780 kg.

---

3. En d'autres mots, la proportion d'étudiants qui signent l'examen ou ne viennent pas du tout.

## Solutions des exercices suggérés

4.5. Max. de vraisemblance : 60% Approche bayesienne : 53,36%

4.6. Max. de vraisemblance : [0,504 ; 0,696] Approche bayesienne : [0,477 ; 0,590]

4.7. a) $p = \frac{1}{5}$ b) [0,1284 ; 0,2716] c) $p = \frac{33}{178} = 18,54\%$

d) Trouver $\epsilon$ tel que $\boxed{\int_{\frac{33}{178}-\epsilon}^{\frac{33}{178}+\epsilon} \frac{\Gamma(180)}{\Gamma(34)\Gamma(146)}\, p^{33}(1-p)^{145}dp = 0,95}$ via logiciel.

e) L'approche bayésienne permet d'obtenir un intervalle de confiance plus précis au pris d'une hypothèse a priori (qui semble toutefois très raisonnable).

4.8. a) $\lambda = 0,088$ b) [0,3623 ; 0,5577] c) $\lambda_{iode} = 0,0864$

d) On obtiendrait un $\lambda$ plus petit étant donné que l'approche bayésienne consiste à trouver un compris entre la valeur obtenue à l'aide de la méthode du maximum de vraissemblance (0,088) et l'hypothèse a priori (0,0864).

e) On trouverait un intervalle de confiance plus précis, au prix d'une hypothèse a priori qui n'est pas nécessairement correcte.

4.9. a) $p = \frac{3}{41}$ b) [3,33% ; 11,3%] c) $p = \frac{21}{272}$.

d) Trouver $\epsilon$ tel que $\boxed{\int_{\frac{21}{272}-\epsilon}^{\frac{21}{272}+\epsilon} \frac{\Gamma(274)}{\Gamma(22)\Gamma(252)}\, p^{21}(1-p)^{251}dp = 0,95}$ via logiciel.

e) L'approche bayésienne permet d'obtenir un intervalle de confiance plus précis au pris d'une hypothèse a priori (qui semble toutefois très raisonnable).

4.10. 780 n'est pas dans l'intervalle d'acceptation [781,83 ; 818,71], on rejette $H_0$.