

Le problème du char d'assaut allemand

Rapport d'arbitrage

Janvier 2019

1 Verdict

L'article peut être publié après corrections faites par l'auteur.

2 Appréciation de l'article

L'article présente un problème classique en statistique : l'estimation de la taille d'une population par prélèvement d'un échantillon aléatoire d'items pourvus de numéros de séries.

Aucun résultat nouveau n'est amené, et la manière de présenter le problème ne présente pas non plus d'innovation, mais l'intérêt de l'article est d'offrir un programme informatique permettant de valider numériquement les résultats en effectuant automatiquement de nombreuses simulations à l'aide du logiciel R.

Le problème des tanks allemands a tout à fait sa place dans un cours collégial de probabilités et statistique, car c'est une bonne façon d'illustrer la possibilité de développer différents estimateurs et de comparer leur qualité. Vu le constant influx de nouveaux enseignants au collégial, il est bon de revisiter périodiquement des problèmes permettant de développer des situations pédagogiques de qualité, même si ces problèmes sont bien connus : pour certains lecteurs, cet article pourrait être leur première rencontre avec ce problème. De plus, l'article est bien écrit. Je pense donc que l'article a sa place dans le *Bulletin*.

3 Commentaires généraux et questions

1. À la fin de la section 2.3.3, pour éviter le problème qui est mentionné (l'estimateur donne une valeur inférieure au maximum observé), il serait naturel de définir un nouvel estimateur comme étant égal à \hat{N}_1 si $X_{(n)} \leq \hat{N}_1$, et $X_{(n)}$ sinon. Je ne sais pas s'il est facile d'analyser ce nouvel estimateur corrigé, mais même si on n'en fait pas l'étude, il serait bon de le mentionner et peut-être de l'implanter en R pour le comparer aux autres.

2. Page 7, dans les résultats, vers le milieu de la page, il est écrit

Les variances de nos estimateurs diminuent lorsque i augmente, ce qui indique que \hat{N}_4 est l'estimateur ayant la variance la plus petite.

Je suppose que le i devrait être n . Une fois cette erreur corrigée, cette phrase dit deux choses mais elles ne sont pas reliées. D'une part, les variances diminuent quand n augmente. C'est vrai (et on dit alors que les estimateurs sont convergents), mais il faudrait justifier cette affirmation. D'autre part, il faut justifier que \hat{N}_4 est l'estimateur qui a la plus petite variance parmi les quatre (ce serait une occasion de parler du concept d'estimateur non biaisé à variance minimale). On pourrait aussi mentionner la différence qualitative entre \hat{N}_1 et \hat{N}_2 (dont la variance diminue comme $1/n$) et les deux autres (dont la variance diminue comme $1/n^2$). Quel est ce théorème de Lehmann-Scheffé mentionné dans le dernier résultat ? Une citation serait bienvenue.

3. Remarque esthétique : Les bordures des tableaux sont trop collées sur le texte à l'intérieur. Augmenter l'espace des bordures permettrait au texte de respirer davantage. Voir le package `booktabs` pour faire des beaux tableaux sans filets verticaux.

4. Section 3 : on voit une différence qualitative parmi les histogrammes de la figure 2, page 11 : ceux des estimateurs \hat{N}_4 et \hat{N}_5 sont asymétriques tandis que les autres sont symétriques. Il serait intéressant de le noter et aussi d'expliquer pourquoi c'est le cas. On le voit aussi à la figure 3 : les régions colorées pour \hat{N}_4 et \hat{N}_5 ont une bordure supérieure très droite mais une bordure inférieure moins nette.

La figure 4 ne me semble pas très parlante : on ne voit pas vraiment de différence entre les estimateurs \hat{N}_3 , \hat{N}_4 et \hat{N}_5 . Est-ce à cause d'un *jitter* vertical ajouté pour mieux voir les points ? J'ai du mal à croire que

l'épaisseur reste si large pour la taille de l'échantillon variant entre 100 et 500.

5. Section 4 : cette section a-t-elle sa place dans un article? Songer à la supprimer (ainsi que l'annexe D).

6. Le code R : Le code qui a généré les figures de l'article occupe environ cinq pages. Est-il pertinent d'inclure le code dans l'article, ou peut-on se contenter de l'hyperlien vers le serveur bitbucket ?



Si on désire inclure le code dans l'article, il faudrait utiliser une police plus petite et imprimer sur deux colonnes, afin de ne pas occuper trop d'espace (deux pages me semble un maximum).

À noter que le code est très répétitif : la page 21 consiste en 4 répétitions d'une commande où un seul paramètre varie, et idem pour la page 22.

Recommandation à l'auteur : factoriser le code pour éliminer la redondance, diminuer beaucoup sa longueur et possiblement le rendre plus clair.

4 Corrections mineures

1. Résumé, ligne 1 : seconde guerre mondiale ⇒ Seconde Guerre mondiale
2. Résumé, ligne 1 : les alliés ⇒ les Alliés
3. Résumé, ligne 4 : Britanniques et Américains ⇒ britannique et américain
4. Page 1 para 1 ligne 3 : numéros de séries ⇒ numéros de série
5. Page 1 para 1 ligne 5 : en grand détail ⇒ en détail
6. Page 1 para 1 ligne 7 : remédiés ⇒ remédié
7. Page 1 para 1 ligne 7 : pratico-pratique ⇒ pratique
8. Page 1 para 2 ligne 1 : les alliés ⇒ les Alliés
9. Page 1 para 2 ligne 3 : du nombre de nombreux ⇒ du nombre de divers
10. Page 1 para 2 ligne 3 : par exemple ⇒ comme par exemple
11. Page 1 para 2 ligne 4 : espace insécable après « tableau ».
12. Page 2 ligne 1 : Les russes ont ⇒ Les Soviétiques ont
13. Page 2 ligne 1 : production de char ⇒ production de chars
14. Page 2 ligne 2 : militaires Américains ⇒ militaires américains

15. Page 2 ligne 4 : colone \Rightarrow colonel
16. Page 2 ligne 4 : chaque chars \Rightarrow chaque char
17. Page 2 ligne 5 : voir [1] \Rightarrow [1] (et partout ailleurs, enlever le mot « voir » devant une référence)
18. Page 2 para 2 ligne 1 : non-militaires \Rightarrow non militaires
19. Page 2 para 2 ligne 2 : de Commodore 64 \Rightarrow d'ordinateurs Commodore 64
20. Page 2 para 2 ligne 3 : Le premier modèle \Rightarrow On estime que le premier modèle
21. Page 2 para 2 ligne 3 : iPhone en italique
22. Page 2 para 2 ligne 4 : The International \Rightarrow International
23. Page 2 para 3 : Dans cet article, nous nous intéresserons à l'estimation de la taille N d'une population à partir d'un échantillon aléatoire prélevé dans celle-ci, dans le cas où...
24. Section 2.1 para 1 ligne 3 : Nous pigeons \Rightarrow Nous prélevons
25. Section 2.1 para 1 ligne 4 : à partir de la population \Rightarrow dans la population
26. Page 3 para 1 Ligne 1 : devait-on mettre l'échantillon entre accolades puisque c'est un ensemble ?
27. Page 3 para 2 : préciser ce qu'est ce A qui apparaît sans prévenir. Par exemple : Pour tout entier A entre 1 et n , il est possible de calculer l'espérance...
28. Page 3 para 2 Ligne 4 : enlever « Malheureusement ».
29. Page 3 para 2 dernière ligne : nous donnerons les formules sans démonstration.
30. Page 3 para 3 : Enlever « Par contre » et garder seulement « Les annexes A et B contiennent quelques rappels... »
31. Page 3 tableau 2 : donner une référence pour ces formules.
32. Page 3 après le tableau 2 : Enlever « À l'aide des trois mesures du tableau 2 », puisque ce ne sont pas ces mesures qui nous aident à définir les estimateurs. Et pourquoi « quatre estimés » au lieu de « quatre estimateurs » ?
33. Page 3 para 4 : remplacer les guillemets anglais par des guillemets français.

34. Page 4 para 4 : bizarre d'utiliser l'environnement `quote` pour la phrase qui dit qu'on arrondit les estimateurs.
35. Page 4 dernier para, ligne 3 : dépendamment du fait ... impair \Rightarrow selon la parité du nombre de données
36. Page 4 ligne -2 : non-biaisé \Rightarrow non biaisé (ou sans biais)
37. Page 5 après le tableau 3 : dire explicitement que l'estimateur est sans biais, et peut-être ajouter que la variance diminue avec n (elle tend vers 0 quand n tend vers l'infini, donc l'estimateur est convergent, mais ce n'est pas approprié ici parce que l'échantillon étant prélevé sans remise, n est borné supérieurement par N).
38. Page 6 para 2 ligne 3 : deuxième phrase \Rightarrow Par symétrie, il est raisonnable de s'attendre à ce qu'il y ait autant de numéros de série non observés supérieurs à $X_{(n)}$ que de numéros de série non observés inférieurs à $X_{(1)}$.
39. Page 6 : non-observés \Rightarrow non observés partout (non devant un adjectif ne prend pas de trait d'union, sauf exception)
40. Page 8 ligne 3 : numéros de séries \Rightarrow numéros de série
41. Page 8 ligne 6 : que nous avons trouvé \Rightarrow tel que calculé
42. Page 8 ligne 8 : Ce résultat reste vrai : de quel résultat parle-t-on ? Veut-on dire : « le calcul qui suit ne dépend pas de s » ?
43. Page 9 ligne 3 : nous pouvons donc supposer \Rightarrow nous pouvons en conclure (mais ce serait le moment de parler de convergence d'un estimateur !)
44. Page 12 dernier para : commencer le paragraphe en indiquant que nous faisons un autre type de simulation.
45. Page 12 ligne -3 : l'estimation obtenu \Rightarrow l'estimation obtenue
46. Page 13 section 4 ligne 2 : utilisaient \Rightarrow utilisait
47. Page 13 section 4 ligne 3 : étaient \Rightarrow était
48. page 13 ligne -3 : la présence de numéros entre crochets au début de chaque ligne risque de confondre le lecteur.
49. Page 14 conclusion, dernier para : Davantage d'informations est par contre disponible \Rightarrow Le lecteur intéressé peut trouver plus d'information
50. Page 14 annexe A ligne 6 : confusion possible entre le k de la définition 1 et le k de la propriété $E(k) = k$. De plus, écrire $E(k)$ est dangereux : k est un réel, par une variable aléatoire.

51. Page 15 Annexe B ligne 1 : ce qui suit le « c'est-à-dire » n'apporte rien de nouveau.
52. Page 15 ligne -6 : Reformuler pour dire ce qu'on veut dire : si $i < A$ ou si $i > N - n + A$, alors $P(X_{(A)} = i) = 0$. Pour le calcul qui suit, préciser qu'on suppose que $A \leq i \leq N - n + A$.
53. Page 16, au milieu de la page : démontrer l'espérance \Rightarrow calculer l'espérance
54. Page 16 en bas : Les démonstrations \Rightarrow Les calculs
55. Page 25 : Les items 8, 9, 11 de la bibliographie ont des espaces beaucoup trop grands entre les mots.
56. Page 25 : Les items 8, 9, 10, 11 ont un point de trop après le titre de l'article.
57. Page 25 : Item 11, il n'est pas recommandé de mettre un lien bit.ly (de nature fragile) dans une bibliographie. Inclure le lien final.