NOTE SUR LES ÉTENDUES RESPECTIVES DES INTERVALLES DE CONFIANCE USUELS POUR UN MOYENNE AVEC VARIANCES CONNUES ET INCONNUES

Marc Bourdeau

École Polytechnique de Montréal,
CP-6079 Centre-ville, Montréal, Qc, Canada, H3C 3A7. Louis.Marc.Bourdeau@Gmail.com

Résumé. Beaucoup de professeurs des éléments de la statistique prennent pour acquis que, pour des lois gaussiennes, les intervalles usuels de confiance pour les moyennes sont plus étroits lorsque la variance est connue que lorsque la variance n'est pas connue. Sans jamais creuser la question tant il est évident que plus d'information apporte plus de précision. Cela est loin d'être toujours vrai. On explore la question pour des tailles échantillonnales et des confiances courantes.

Mots-clés. Intervalles de confiance des moyennes avec variances connues et variances inconnues. Interprétation fréquentielle des probabilités. Estimateurs usuels des écarts types des lois gaussiennes. Approximation de Hilford-Hilferty pour les lois χ^2.

Abstract. Many if not most professors of Statistics admit implicitely that the usual confidence intervals for means of normal random variables with known variances are more precise than the one calculated without this assumption. Simply because more information brings more precision. We show here that for usual confidence levels and usual sample sizes the probability of this being false is far from small, and tends to increase to ½ with sample sizes.

Keywords. Confidence intervals for means with known an unknown variances, frequentist interpretation of a probability, Estimation of standard deviation of normal laws. Wilson-Hilferty approximation for the χ^2 laws.

1 Introduction

Cette communication s'inscrit dans la suite des considérations de Petiot & Turlot (2012 & 2014) sur les propriétés des divers intervalles de confiance pour les proportions. On apporte un certain approfondissement aux intervalles de confiance usuels pour les moyennes.

Il est rare qu'on se penche sur les détails des choses simples, tant on les prend pour acquis. Or les apparences sont parfois trompeuses. Même dans les manuels appliquées qui connaissent des ajouts importants de matériel d'éditions en éditions (DeVore, 2011; Moore, McCabe & Craig 2012 ; Moore & Notz, 2009) ne mentionnent pas les détails qu'on expose ici. Tout au plus fait-on valoir chez certains lorsqu'on y expose les intervalles de confiance pour les moyennes, leur interprétation fréquentielle qui n'est pas simple à comprendre pour la plupart des étudiants.

C'est ainsi qu'on prend pour acquis, et c'est souvent bien avéré, que plus on connaît d'informations sur une variable aléatoire plus on peut être précis dans les estimations de ses paramètres, et que par voie de conséquence, les intervalles de confiance pour les moyennes lorsqu'on admet connaître la variance seraient plus étroits que lorsqu'on utilise l'estimateur *s* usuel de cette dernière.

Nous avons mis au point une animation pour le calcul des intervalles de confiance usuels pour les moyennes des lois gaussiennes, qu'on suppose suivre approximativement d'ailleurs elles-mêmes des lois gaussiennes même pour des lois non gaussiennes lorsque les tailles échantillonnales sont suffisantes, qui nous a fait douter de nos calculs et de notre intuition… Car, une fois nos calculs vérifiés, il a bien fallu admettre que cette intuition était fausse pour une bonne proportion des utilisations. Mais laquelle ? Cette animation était avant tout destinée à faire comprendre l'interprétation fréquentielle d'une probabilité, éternelle subtilité qui se comprend justement dans ce cas-là de manière assez difficile.

Il peut être intéressant de profiter d'une animation bien conçue pour faire voir que les intuitions peuvent tromper, en exposant les détails qu'on développe ici. Cela fait comprendre par un exemple un calcul un peu complexe sur des probabilités, ce genre de calculs fort utiles dans la pratique et qui là aussi sont souvent escamotés.

2 Les deux approches

On a pensé à deux approches pour explorer la question. La première cherche à calculer la loi de l'écart entre les deux étendues, et ainsi déterminer sa moyenne et sa variance. Mais le calcul est complexe car, d'une part, on ne connaît par la loi de l'estimateur $s = \sqrt{(s^2)}$ qui sous-estime σ, et, d'autre part, les lois χ^2 ne sont connues que par approximation, la plus connue étant celle de Wilson-Hilferty (1931), et les quantiles des lois T_n n'étant aussi connus qu'approximativement (Abramovitz & Stegun, 1970).

Puisque s sous-estime σ et que les p-quantiles ($p>0,5$) sont plus grands pour les lois de Student que les quantiles équivalents des lois gaussiennes, on peut penser en fin de compte que cet écart est probablement négatif pour une certaine proportion des applications. Les calculs toutefois sont plutôt malaisés et approximatifs. Toutefois, même si on peut calculer avec de très bonnes approximations, il n'est pas utile de connaître la loi de cet écart.

À toutes fins pratiques, en effet, on n'a besoin de ne connaître que la probabilité p de l'écart entre les deux étendues, celle avec la variance inconnue soustraite de celle avec la variance connue, d'en examiner les propriétés, qui dépendent des tailles échantillonnales et des confiances désirées. Notons que toute probabilité, dans son interprétation fréquentielle, n'est que la moyenne de lois de Bernoulli itérées indéfiniment de façon indépendante.

On montrera que cette probabilité est :

$$p = \mathrm{P}[\,\chi^2_{\,n-1} < (n-1)\,K\,],$$

où $K = (\,z_{(\alpha/2)} / t_{(n-1\,;\,\alpha/2)}\,)^2$, avec les notations habituelles pour les quantiles des lois T et Z (la gaussienne). Elle varie bien sûr avec n et (1- α), la taille échantillonnale et la confiance bilatérale considérées. De plus, quelle que soit la confiance désirée, cette probabilité tend vers ½ lorsque n grandit.

On présentera l'animation (Flash) qui a inspiré ces considérations et qui met aussi en évidence l'interprétation fréquentielle des intervalles de confiance, ainsi que les détails mathématiques et les graphiques illustratifs. On pourra trouver le tout sous le lien suivant (Bourdeau, 2008).

On trouvera à la page sous-jacente une bibliographie plus complète ainsi que les dernières versions (résumé et texte plus complet) de notre présentation.

Bibliographie

Abramovitz, M., & Stegun, I. A. (Éds.). (1970). *Handbook of mathematical functions. With formulas, graphs and mathematical tables.* Récupéré sur http://people.math.sfu.ca/~cbm/aands/

Bourdeau, M. (2006). *Note on the respective widths of confidence intervals for means with known and unknown variances.* Récupéré sur http://wikistat.mgi.polymtl.ca/tiki-download_file.php?fileId=83

DeVore, J. L. (2011). *Probability and Statistics for engineers and the sciences* (éd. 8). Belmont CA: Brooks/Cole.

Moore, D. S., & Notz, W. I. (2009). *Statistics. Concepts and controversies* (éd. 7). New York NY: W. H. Freeman and Co.

Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). *Introduction to the practice of Statistics* (éd. 7). New York NY: W. H. Freeman and Co.

Petiot, J.-F., & Turlot, J.-C. (2012). *L'intervalle de confiance pour une proportion et son enseignement dans le secondaire.* Récupéré sur http://wikistat.mgi.polymtl.ca/tiki-download_file.php?fileId=81.

Petiot, J.-F., & Turlot, J.-C. (2014). *Intervalles de confiance pour une proportion: pourquoi l'intervalle "standard" de Wald est-il le plus couramment enseigné?* Récupéré sur http://papersjds14.sfds.asso.fr/submission_247.pdf

Wilson, E. B., & Hilferty, M. M. (1931). The distribution of Chi-square. *Proceedings of the National Academy of Sciences, 17*, pp. 684-688.