Critères de Aikaide et de Bayes pour comparer des statistiques ajustés avec des données

Généralement, les modèles statistiques sont ajustés avec le principe de *vraisemblance maximale.* Cette méthode estime les paramètres du modèle, notés β, en maximisant la fonction de vraisemblance. Cette fonction, notée L(β), est le produit de la fonction de densité (ou de masse) de probabilité calculée sur les observations mais avec des paramètres inconnus β. On cherche à déterminer les valeurs de β qui maximise la fonction L(β). Au lieu de maximiser la fonction L(β), qui est un produit, il est plus pratique de minimiser de 2 fois le négatif du log de L(β) c.-à-d. – 2Log (β). Les petites valeurs de -2Log L(β) indique les meilleurs modèles. On peut aussi faire un test de différence entre un modèle complet et un modèle réduit car la différence entre la valeur de -2logL(β) est asymptotiquement distribuée selon une Khi-carrée. Le degré de liberté de la distribution est égal à la différence le nombre de paramètres du modèle complet et le nombre de paramètres du modèle réduit.

Il y a 2 critères qui sont employés dans la littérature statistique pour comparer des modèles: le critère de Aikaike, noté AICc et le Critère d'Information de Bayes, noté BIC.
Les critères sont définis par les formules (1) et (2) et les paramètres estimés de β

$$\text{AICc} = -\,2L(\hat{\beta}) + 2k + 2k\,(k+1)\,/\,(n-k-1 \qquad (1)$$

$$\text{BIC} = -\,2L(\hat{\beta}) + k\,\ln(n) \qquad (2)$$

où

- **k** nombre de paramètres estimés dans le modèle
- **n** nombre d'observations
- **SSE** somme de carrés de l'erreur dans le modèle

L'objectif est de minimiser AICc et le BIC. Le critère BIC pénalise les modèles ayant plus de paramètres que le critère AICc. Il favorise des modèles parcimonieux c.-à-d. des modèles ayant le moins de paramètres, ce qui n'est pas le cas pour AICc.

Calcul de AICc et BIC: cas de la régression par moindres carrés

Si on fait l'hypothèse que la fonction de densité est normale, alors la maximisation de la fonction de vraisemblance conduit à la minimisation de moindres carrés. Dans le cas de la régression basée sur le principe des moindres carrés, il y a une relation entre la fonction de vraisemblance $L(\hat{\beta})$ et la somme des carrés résiduels SSE.
On peut alors calculer les critères AICc et BIC avec les formules (3) et (4) qui ne demandent pas de connaitre la valeur de $L(\hat{\beta})$ qui n'est pas nécessairement disponible avec le logiciel employé mais la valeur de SSE l'est toujours. Les formules (1) et (2) deviennent alors

$$\text{AICc} = n\ln(SSE/n) + n\ln(2\pi) + n + 2k + 2k\,(k+1)\,/\,(n-k-1) \qquad (3)$$

$$\text{BIC} = n\ln(SSE/n) + n\ln(2\pi) + n + k\ln(n) \qquad (4)$$

Les 2 critères ont des éléments communs. Leur différence peut s'écrire

$$\text{AICc} - \text{BIC} = 2k + 2k\,(k+1)\,/\,(n-k-1) - k\ln(n) \qquad (5)$$