

Modèles additifs généralisés

Ce laboratoire doit être remis le 25 mars à 17h sur Moodle.

Données

Le fichier `portal_ot.csv` est un sous-ensemble de la base de données Portal (utilisée pour les labos 2 et 5) qui contient le nombre d'individus (n) observés de l'espèce *Onychomys torridus* (souris sauterelle du sud) dans chaque parcelle pour les années 1988 à 2002. Chaque parcelle était soumise à un traitement (`plot_type`) visant à exclure une partie ou l'ensemble des rongeurs de la parcelle.

```
portal_ot <- read.csv("../donnees/portal_ot.csv")
portal_ot$plot_type <- as.factor(portal_ot$plot_type)
portal_ot$plot_id <- as.factor(portal_ot$plot_id)
head(portal_ot)
```

```
##   species_id plot_id year n      plot_type
## 1         OT      1 1988 0 Spectab exclosure
## 2         OT      1 1989 9 Spectab exclosure
## 3         OT      1 1990 2 Spectab exclosure
## 4         OT      1 1991 3 Spectab exclosure
## 5         OT      1 1992 3 Spectab exclosure
## 6         OT      1 1993 1 Spectab exclosure
```

Note: Tel qu'indiqué dans le code ci-dessus, les variables catégorielles `plot_type` et `plot_id` doivent être converties en facteurs avant d'ajuster un GAM.

1. Estimation de la tendance démographique globale

Pour toutes les questions de cette partie, vous devez ajuster un modèle additif généralisé pour estimer la tendance démographique de l'espèce en tenant compte de l'effet des traitements: $n \sim \text{plot_type} + s(\text{year})$. Pour l'instant, nous ignorerons le groupement des mesures par placettes.

- Ajustez d'abord un GAM où les observations suivent une distribution de Poisson. De façon brève, décrivez comment le nombre d'individus varie en fonction de l'année et des traitements. En considérant la fonction de lien utilisée pour ce modèle, que signifie l'additivité des effets `plot_type` et `s(year)`?
- Est-ce que la valeur par défaut du nombre de fonctions de base k est suffisante pour représenter `s(year)` dans le modèle en (a)? Si nécessaire, réajustez le modèle avec un k plus élevé. Quelle est la valeur maximale de k que vous pouvez utiliser ici?
- Les données sont-elles surdispersées par rapport à votre modèle?
- Ajustez un nouveau GAM avec la distribution binomiale négative, spécifiée comme `family = "nb"` dans la fonction `gam`. Quel est l'estimé du paramètre θ pour ce modèle? L'ajustement semble-t-il meilleur que pour le modèle de Poisson? Y a-t-il encore des problèmes d'ajustement?

2. Ajout des effets aléatoires de placettes

- (a) À partir du modèle binomial négatif en 1(d), incluez un effet aléatoire des placettes sur l'ordonnée à l'origine. Vérifiez l'ajustement du modèle, incluant la normalité des effets aléatoires et la présence ou l'absence de surdispersion.
- (b) Ajustez maintenant un modèle avec un effet aléatoire des placettes sur la tendance démographique moyenne `s(year)`, en utilisant un terme de type `bs = "fs"` tel que vu dans le cours. Comparez ce modèle au modèle en (a) avec l'AIC. *Note:* Le package *AICcmodavg* n'est pas compatible avec les GAM, mais vous pouvez calculer l'AIC pour chaque modèle avec la fonction `AIC`.
- (c) Quelle est la fraction de la déviance expliquée par le meilleur modèle déterminé en (b)?
- (d) Finalement, illustrez la tendance démographique estimée pour cette espèce dans chaque placette. Pour ce faire, ajoutez au jeu de données `portal_ot` les valeurs prédites par le meilleur modèle en (b) ainsi qu'un intervalle de confiance à 95%, puis représentez dans un même graphique les données observées, les tendances temporelles estimées et leur intervalle de confiance pour chaque placette.

Note: Il est suggéré de diviser le graphique en une facette par placette avec la fonction `facet_wrap` de *ggplot2*.