



Département de Mathématiques
Monte Carlo
2022 - 2023

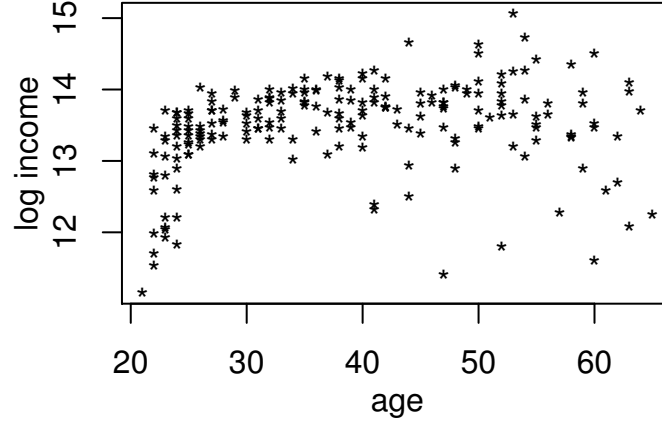
Projet ING1 - GMA

Régression par processus gaussien

Tuteur: Hassan Maatouk

1 Age and income on 205 Canadian workers

- Les données sur l'âge et le revenu réels qui comprend l'âge (en années) et le logarithme du revenu ($\log(\text{revenu})$) sur 205 travailleurs canadiens disponible à partir du package **SemiPar** de R..
- Le but de ce projet est d'estimer le revenu en fonction de l'âge.
- The observations sont données dans la figure ci-dessous:



2 Procédures

- Utiliser la régression par processus gaussien pour estimer le revenu en fonction de l'âge.
- Vous pouvez dans un premier temps utiliser plusieurs fonctions de covariance (voir [1, 2]).
- Pour comparer différentes approches, vous pouvez utiliser l'erreur de prédiction quadratique moyenne (MSPE). Par exemple, les données de taille 205 peuvent être divisées aléatoirement en jeu de données d'entraînement et en jeu de données de test:

$$\text{MSPE} = \frac{1}{n_t} \sum_{i=1}^{n_t} \left(y_i - \hat{f}(x_i) \right)^2,$$

avec n_t le nombre des observations test.

- Pour estimer les paramètres θ et σ_{noise} vous pouvez utiliser par exemple: la méthode de Maximum likelihood [2] ou bien la méthode de validation croisée (la minimisation de MSPE).

References

- [1] H. Maatouk and X. Bay. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582, 2017.
- [2] C. E. Rasmussen and C. K.I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.