

# Résumé réunion 20 Février 2019

---

## Préparation des données: les données manquantes

Nous avons devoué les processus gaussiens, qui sont des modèles probabilistes pouvant nous permettre à la fois de prédire les valeurs manquantes dans nos séries temporelles et à faire de la classification.

Nous avons exploré deux hypothèses de prédictions des distmod manquants. L'étape de compréhension des données nous a permis de voir que les attributs distmod et hostgal\_photoz étaient corrélées et que la fonction  $f(\text{hostgal\_photoz}) = \text{distmod}$  avait l'allure d'une fonction racine carrée. Nous avons donc construit un modèle de régression linéaire

$$\text{distmod} = w_1 \times \text{hostgal\_photoz} + w_2 \times \sqrt{\text{hostgal\_photoz}} + w_3 \times \text{hostgal\_photoz\_err} + w_4$$

La seconde piste explorée est un processus gaussien sur les variables hostgal\_photoz et hostgal\_photoz\_err. Cette piste semble être beaucoup plus prometteuse, surtout qu'on a l'intervalle de confiance associée à chaque prédiction.

## Conclusion

Durant cette semaine, nous avons découvert les processus gaussiens et avons construit deux modèles pour prédire les distmods manquants. La prochaine séance portera toujours sur la préparation de données et plus particulièrement sur la prédiction des flux manquants.