

# Cadrage du Projet de Spécialité

## Estimation de puissance à posteriori

Étudiants : Crépin Baptiste, Bonjean Grégoire & Lair Thomas  
Encadrant : Jean-Charles Quinton

Mai - Juin 2016

## 1 Introduction - Contexte et Organisation

### 1.1 Présentation du Projet - Ressources

Ce projet, mené par un groupe de 3 étudiants, est un projet de recherche dans le domaine de la statistique inférentielle.

Celui-ci ne nécessite ni budget ni matériel particulier.

Les compétences requises pour l'appréhender correctement sont les suivantes :

- Bonnes connaissances en statistiques, et plus précisément en statistiques inférentielles.
- Simulation de Monte-Carlo et Méthode de Bootstrap.
- Utilisation de R.

### 1.2 Présentation de l'Équipe - Motivations

L'équipe projet est constituée de Crépin Baptiste, Bonjean Grégoire et Lair Thomas.

Nous sommes tout trois étudiants dans la filière Ingénierie Financière, et avons suivi le cours de Statistiques Inférentielles Avancées durant le deuxième semestre de cette année.

Ainsi nombre des notions abordées dans ce projet de recherche statistique nous sont familières, et le travail que nous serons amené à effectuer pendant ce mois de projet (notamment sur les méthodes de Bootstrap et de Monte-Carlo) nous permettra d'approfondir nos connaissances dans le domaine de la statistique inférentielle.

Ce travail sera effectué principalement en groupe, dans les locaux de l'Ensimag sur toute la durée du projet, auquel s'ajoutera un travail personnel régulier à domicile pour avancer et terminer les objectifs en cours.

L'équipe étendue comprends notre encadrant, Jean-Charles Quinton, professeur titulaire à l'Université Grenoble-Alpes et chercheur au département "Probabilités et Statistiques" du Laboratoire Jean Kuntzmann.

### 1.3 Communication

Notre équipe travaille principalement à l'Ensimag, et une grande partie de la communication sera donc orale. Nous tenons des réunions tous les matins pour mesurer régulièrement l'avancement du projet, mais des décisions et des changements de directions peuvent également être pris à d'autres moments de la journée.

- Interne : Réunions quotidiennes.
  1. Le matin : Programme de la journée, séparation du travail, objectifs du jour et long terme.
  2. Le soir : Debriefing de la journée, mise au point sur le travail qui a été fait, qu'il reste à faire, et sur les objectifs futurs.
- Externe : Communication avec l'encadrant (mail et rendez-vous).

## 1.4 Organisation du projet

Pour ce projet, nous avons organisé le travail en deux parties :

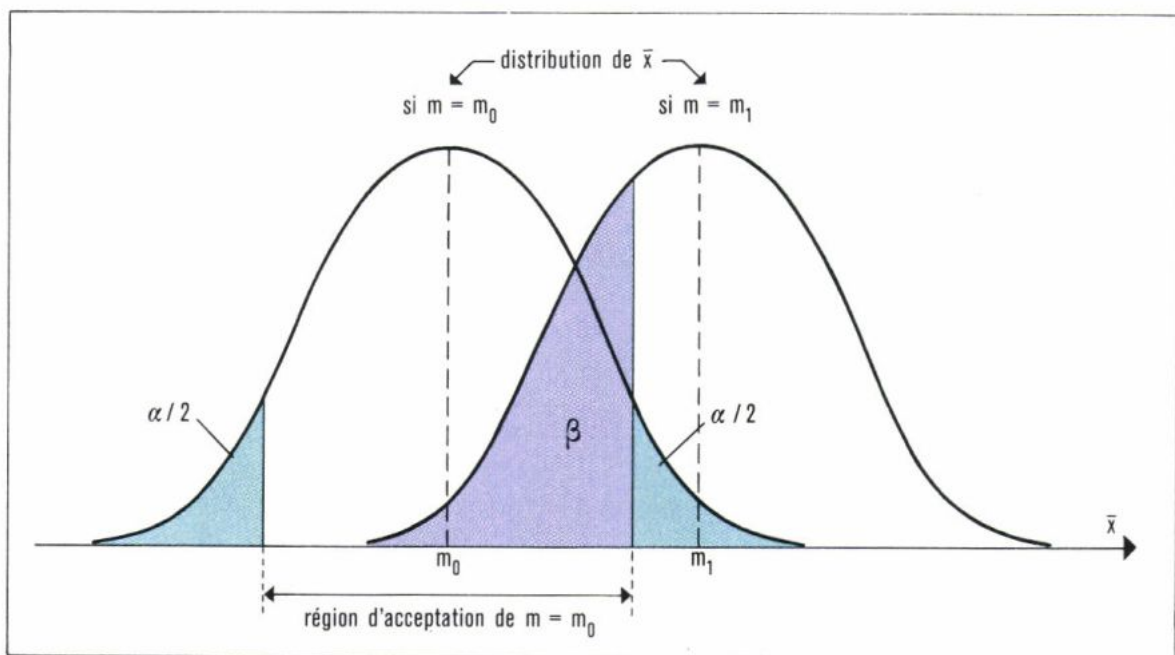
1. Travail Commun : 8h - 17h, du Lundi au Vendredi à l'ENSIMAG
2. Travail Personnel : à domicile, pour terminer les objectifs non achevés du jour par exemple.

Les week end sont à priori laissés libres, sauf en cas de retard du projet. Ces horaires relativement légers nous permettent de travailler en dehors sur les différentes tâches du projet, qui sont pour la plupart indépendantes les unes des autres. En effet, notre projet se compose de plusieurs petites tâches à implémenter plutôt que d'un objectif long terme.

## 2 Développement

### 2.1 Objectifs et Enjeux

L'objectif de ce projet est de calculer, pour une étude statistique, un paramètre en fonction de deux autres. Le but est en effet d'observer une taille d'effet (concrètement, la différence de taille moyenne entre deux populations d'âges différents par exemple), pour une certaine puissance de test (probabilité d'observer l'effet lorsqu'il existe, voir figure ci-dessous) donnée et pour une certaine taille d'échantillon. Le coût réel d'une étude statistique étant proportionnel au nombre de personnes qu'elle implique, estimer la taille minimum de l'échantillon nécessaire pour observer l'effet recherché à la puissance voulue a un intérêt évident : celui de minimiser le coût de l'étude.



Test bilatéral de l'hypothèse  $m = m_0$ , au risque  $\alpha$ .  $1 - \beta =$   
puissance du test si  $m = m_1$ .

Notre projet se situe dans un cadre d'études statistiques liées à la psychologie. L'objectif est tout d'abord d'obtenir une estimation de puissance pour les cas classiques, puis une estimation de taille d'échantillon. Ensuite, la gestion de cas plus complexes (régression multilinéaire et approximation de cas réels) doit être implémentée. Enfin l'interface graphique doit être mise en place, et nous avons fait le choix de commencer sa création assez tôt dans le projet pour pouvoir l'utiliser tout au long de celui-ci. Les objectifs d'estimation sont clairement mesurables, car certaines fonctions de R permettent de vérifier nos résultats.

## 2.2 Périmètre

Notre travail dans son état final consistera en : une interface graphique référençant tous les cas classiques d'estimation de puissance, dont les paramètres sont choisis par l'utilisateur, et qui comprendra la possibilité de joindre un cas réel (lecture de données d'observations dans un fichier). Nous avons choisis de nous intéresser aux estimations de puissance et de taille d'échantillon plutôt qu'à l'estimation de taille d'effet (qui reste un élément central de la détermination des deux autres grandeurs).

La taille d'effet sera donc toujours supposée connue, et nous ne chercherons pas à l'estimer dans ce projet. En effet, dans le cadre des études psychologiques dans lequel nous nous situons (notre encadrant de projet travaille spécifiquement avec des psychologues), nous aurons toujours un *a priori* sur la taille d'effet et c'est donc la recherche de la puissance et de la taille d'échantillon entraînant cette taille d'effet qui nous intéressera.

## 2.3 Macro-Planning

Ce projet d'estimation de puissance est un projet de recherche, pour lequel nous serons amené tout d'abord comme dit plus haut à considérer des cas simples avant de se pencher sur des cas plus complexes.

Nous avons donc assez instinctivement fait le choix d'une approche incrémentale, en se fixant des paliers dans l'avancement du projet.

Les dates clés sont les suivantes :

- 16 Mai : Début du travail de recherche et de prise en main.
- 20 Mai : Établissement du Cahier des Charges et du Cadrage.
- Du 23 Mai au 1er Juin : Cas Simples (calcul de puissance, étude pilote, intervalles de confiance et estimation de nombre d'observations).
- Du 2 Juin au 8 Juin : Cas de Régression et Cas Réel.
- 8 Juin : Fin de Projet
- 9 ou 10 Juin : Soutenance

## 2.4 Risques et Dépendances

Les risques auxquels nous serons confrontés durant le déroulement de ce projet, et que nous surveillerons donc tout particulièrement, sont les suivant :

- Interpretation :
  1. Risque : Mauvaise interprétation du sujet, départ sur une idée qui ne résout pas le problème initial posé par le sujet.
  2. Solution : Travail approfondi de prise en main pour cerner au mieux possible les notions et les enjeux.
- Performances :
  1. Risque : Avoir de mauvaises performances sur nos simulations, ou des simulations fausses.
  2. Solution : Comparaison avec les fonctions classiques de R, analyse graphique des résultats.
- Temps :
  1. Risque : Ne pas remplir la totalité de nos objectifs dans le temps imparti.
  2. Solution : Diagramme de Gantt et suivi quotidien de l'avancement de projet.

Ce projet de recherche est globalement très libre et nous pouvons avancer dessus sans dépendances particulières.

## 2.5 Indicateurs

Les indicateurs que nous avons choisis pour piloter notre projet sont les suivant :

- Avancement des livrables :  
Comme ce projet de recherche s'articule autour d'une approche incrémentale, les indicateurs d'avancement des livrables seront l'atteinte des différents incréments établis en début de projet (estimation fonctionnelle pour les cas simples, puis pour les cas de régression, etc...).
- Délais :  
Diagramme de Gantt
- Motivation de l'équipe :  
Ponctualité des membres, présence ou non de conflits au sein de l'équipe.

Signature des membres de l'équipe projet :