Rapport du Projet de Spécialité
Estimation de puissance à posteriori

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

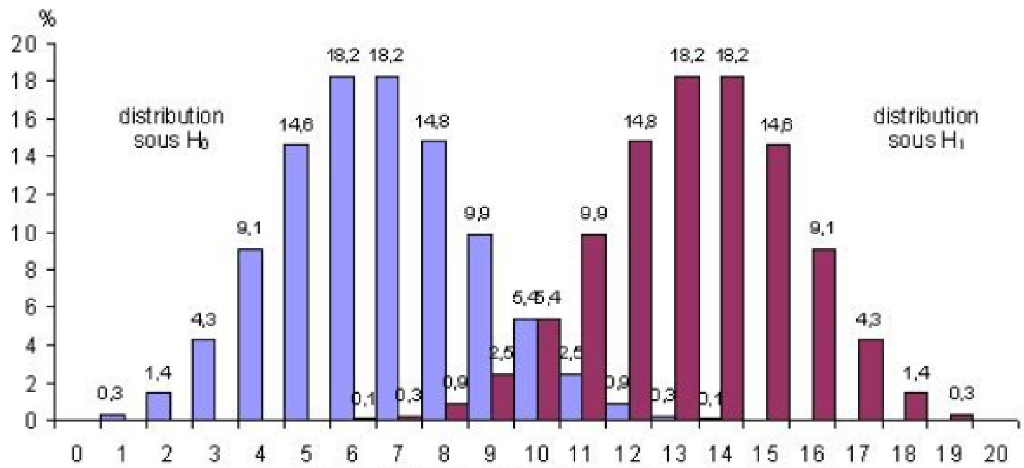
Mai - Juin 2016

1 Introduction - Sujet et problème posé

Ce rapport, rédigé par Grégoire Bonjean, Baptiste Crépin et Thomas Lair, reprend et explique tout le travail produit dans le cadre du projet de spécialité, sur le sujet de l'estimation de puissance à posteriori.

1.1 Sujet

Le problème que nous nous posons lors de ce projet est un problème issu du domaine psychologique en statistiques : il s'agit de minimiser le nombre de personnes impliquées dans une étude, afin d'observer un certain effet pour une puissance donnée(fiabilité du test).
Concrètement, on peut considérer l'exemple suivant : on inscrit le mot "bleu" dans la couleur bleue, et on demande à un groupe d'individus la couleur dans laquelle le mot est inscrit, en mesurant leur temps de réaction. De manière indépendante, on inscrit le mot "bleu" dans la couleur rouge et on pose à un groupe d'individus la même question que précédemment, en mesurant de nouveau leur temps de réaction. On nommera le premier groupe le groupe congruent, et le second le groupe incongruent. Les expériences montrent que le temps de réaction mesuré sera en moyenne plus grand pour le groupe incongruent que pour le groupe congruent.



Distribution des temps de réaction pour les groupes confruents et incongruents

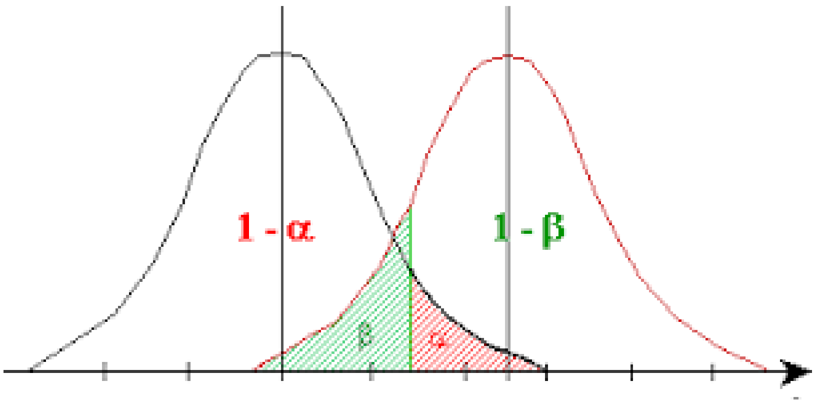
Les temps de réactions suivent, pour chaque groupe respectivement, une loi normale de moyenne le temps de réaction moyen des individus. L'effet que l'on souhaite observer ici, ou taille d'effet, serait la diférence des deux moyennes, normalisée par l'écart type des observations.

Le psychologue voudra donc savoir le nombre de personnes minimum dont il a besoin afin d'observer une différence entre les moyennes, à une puissance voulue.
En réalité, le psychologue réalisera d'abord une étude "pilote", avec un nombe de personnes réduit, qui lui donnera l'ordre de grandeur de la taille d'effet quil peut obtenir, et lui permettra d'obtenir le nombre de personnes dont il a besoin pour valider une taille d'effet.
L'objectif sera donc d'exprimer deux paramètres en fonction du dernier parmis les trois suivants : la taille d'effet, la taille d'échantillon, et la puissance.

1.2 Introduction aux concepts

1.2.1 La puissance d'un test

Un test d'hypothèse est un test entre deux hypothèses : l'hypothèse nulle H0 et l'hypothèse alternative H1. Ici, l'hypothèse nulle correspond au fait de n'observer aucune taille d'effet significative, et l'hypothèse alternative correspond au fait d'en observer une significative. La probabilité α correspond à la probabilité de rejeter l'hypothèse nulle alors qu'elle est vraie, et est appelée erreur de première espèce. La probabilité β correspond à la probabilité de retenir l'hypothèse nulle alors qu'elle est fausse. C'est l'erreur de seconde espèce. La puissance 1 - β est la probabilité de retenir l'hypothèse alternative à raison.



Exemple pour deux lois normales, groupe congruent à gauche et incongruent à droite

1.2.2 La taille d'effet

Dans le cadre de notre projet, on suppose que le praticien cherche à montrer l'existence d'une taille d'effet précise (par exemple une différence de temps de réaction supérieure à 100ms entre le cas congruent et incongruent de l'exemple précédent). Celle-ci sera donc toujours connue, et fournie par le praticien, puisque ce qu'il cherche à tester, c'est précisément l'existence de cette taille d'effet.

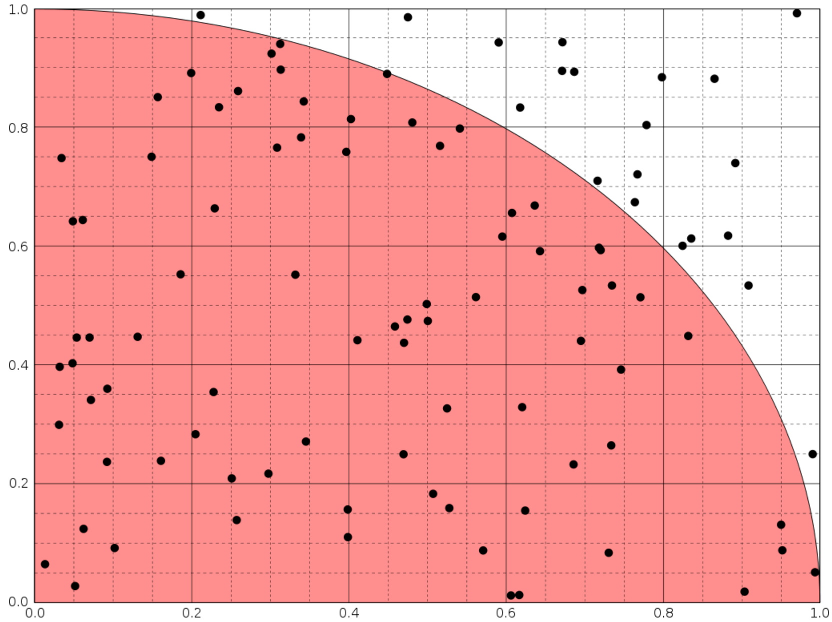
1.2.3 Méthodes de Montecarlo et de Bootstrap

Pour estimer la puissance des tests futurs à réaliser, et valider la signification de nos résultats, nous employons des méthodes de ré-échantillonnage.
Ces méthodes se basent sur des procédés aléatoires qui exploitent les données disponibles, qui sont dans notre cas, les données de l'étude pilote.
Les méthodes de ré-échantillonnage utilisées dans notre projet sont les méthodes de Bootstrap et Monte Carlo.
Pour chacune des 2 méthodes, nous présentons dans un premier temps en quoi elle consiste, puis comment nous l'avons utilisée pour répondre à notre problème.

Monte Carlo

Principe :

Une bonne illustration de la méthode de Monte Carlo est son application dans le calcul du nombre π. Prenons un quart disque de centre 0 et de rayon 1.
On tire N points au hasard de coordonnées comprises entre 0 et 1, comme illustré sur la figure ci-dessous intitulée "Monte Carlo". Pour chacun des points de coordonnées (x,y), on peut facilement déterminer s'il appartient au disque ou non, grâce à la fonction "distance euclidienne" d : $(x, y) \mapsto \sqrt{x^2 + y^2}$.
Si $d(x, y) < 1$ alors (x,y) appartient au disque.
La probabilité qu'un point appartienne au disque est de $\frac{\pi}{4}$ d'après la formule de la surface d'un disque Ainsi, en tirant un grand nombre de points, on obtient une bonne estimation de la quantité $\frac{\pi}{4}$, (et a fortiori de π) en effectuant le rapport du nombre de points dans le disque au nombre total de points tirés.



Monte Carlo

Utilisation :

Au lieu de générer des points aléatoires comme dans l'illustration précédente, nous générons des échantillons aléatoires (dans chacun des sous-cas traités) un nombre N de fois.

Alors que les points avaient pour contrainte d'être compris entre 0 et 1, ici les échantillons ont pour contrainte de suivre la loi de probabilité estimée à partir de l'étude pilote. On peut facilement déterminer si l'on rejette l'hypothèse nulle H0 par l'emploi de la fonction "statistique du test" correspondante. Ainsi la puissance $1 - \beta$ étant la probabilité qu'on rejette H0, en tirant un grand nombre d'échantillons aléatoires, on obtient une bonne approximation de la puissance.
Nous utilisons donc cette méthode pour estimer la puissance de tests. Exemple :
Pour l'exemple de la comparaison de moyenne à 0 d'un unique échantillon X, en posant k la taille de chacun des échantillons générés par la méthode de Monte Carlo, la fonction statistique du test est T : $X \mapsto \frac{\hat{X}}{\frac{\hat{\sigma}}{k}}$
Si $T(X) < qt(\frac{\alpha}{2}, n-1)$ alors on rejette H0 On obtient alors que la puissance est le rapport du nombre de fois où l'on rejette H0 à partir d'un tirage sur le nombre de tirages total.

Bootstrap

Principe :

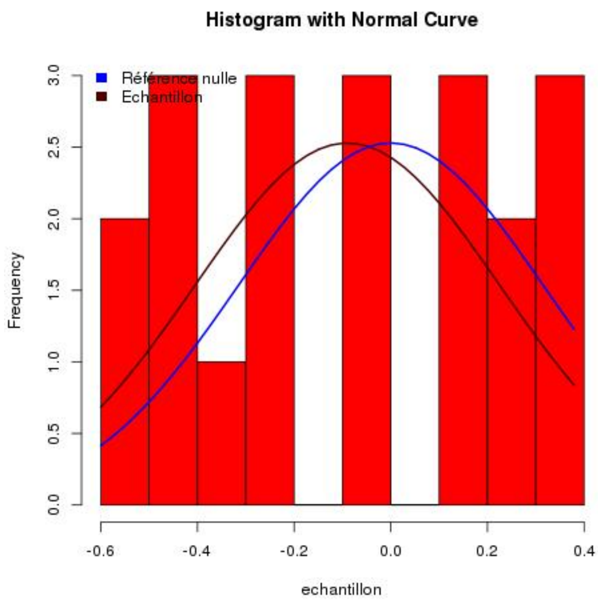
Le Bootstrapping est une méthode de statistique inférentielle, dont le principe est de créer des échantillons que par tirage dans l'ancien, avec remise à partir de l'échantillon initial.
Il s'agit de choisir la taille des sous-échantillons à tirer parmi l'échantillon initial, et de tirer un grand nombre N de sous-échantillons.

Utilisation :
Ici, l'échantillon initial est le pilote. Nous l'utilisons pour estimer les bornes de l'intervalle de confiance (avec alpha = 0.05) de l'écart-type du pilote. Les bornes pour celui de la moyenne sont caculées par un autre moyen.
Cela nous permettra de définir "un intervalle de confiance" pour la puissance estimée, plutôt qu'une valeur unique en fonction des paramètres empiriques (Cf section démarche - estimation avec intervalle de confiance).
Nous utilisons la fonction boot du package "boot".

1.3 Choix et interprétations du sujet

Le sujet de l'estimation de puissance est un sujet très vaste, tant les méthodes de calculs dépendent des hypothèses du test, nous avons donc du décider les aspects que nous voulions développer en priorité, et ceux que nous préférions laisser de côté.
https ://preview.overleaf.com/public/yfrgwxbknfhw/images/46b1fec04f767ba94a6195dbdcc6638f4cc651cd.jpeg
Nous avons gardé à l'esprit l'objectif primaire : minimiser le coût de l'étude statistique du psychologue, et donc minimiser la taille de l'échantillon. L'estimation de la taille d'échantillon pour une puissance fixée et une certaine taille d'effet se devait donc d'être dans nos objectifs de calcul.
Également, le psychologue doit avoir besoin de calculer la puissance de son étude statistique, afin de valider son étude sur des critères statistiques, et de donner du crédit à l'effet observé.
Le paramètre, parmis les trois paramètres que nous prenons en compte dans ce sujet, que nous avons décidé de ne pas calculer, est la taille d'effet : en effet, le psychologue a toujours un apriori sur la taille d'effet (qu'il aura pu précalculer sur une première étude), et nous n'avons pas jugé nécessaire de nous attarder sur son implémentation.
Notre délimitation du travail est la suivante : nous sommes partis dans une optique de documentation de calculs sur les cas d'école d'estimation de puissance, dans lesquels l'utilisateur devra se placer plutôt que de fournir un cas réel.
Ces cas d'école sont les suivants :

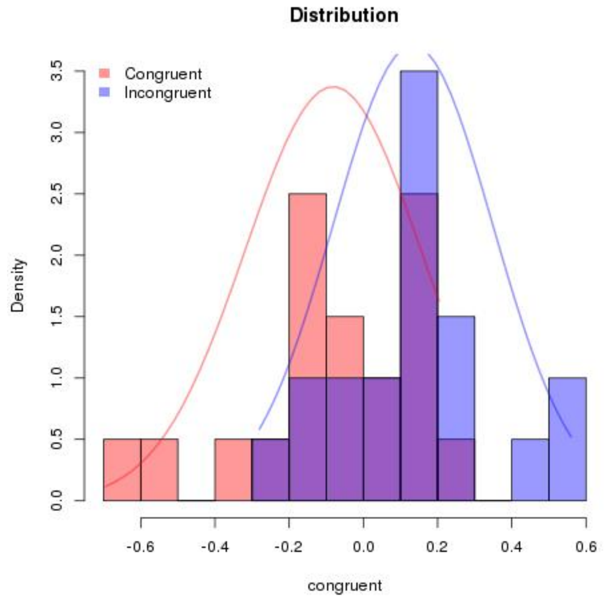
1. Test à un échantillon
   Ici, on compare un échantillon , de moyenne nulle. Le but du test d'hypothèse sera ici de montrer que la moyenne de la première loi normale est différente de zéro. (Hypothèse nulle : la loi est de moyenne nulle, hypothèse alternative : la loi normale est de moyenne non nulle).



Distribution pilote pour une loi normale à un échantillon

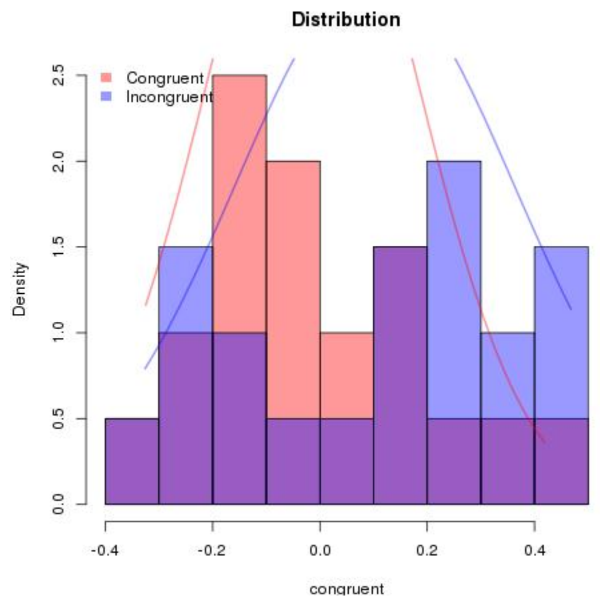
2. Test à deux échantillons indépendants

Ici,on compare deux échantillons, l'un correspondant au test congruent, et l'autre au test incongruent. Les tests sont considérés comme passés dans des structures indépendantes, et les éléments des deux groupes sont non corrélés.



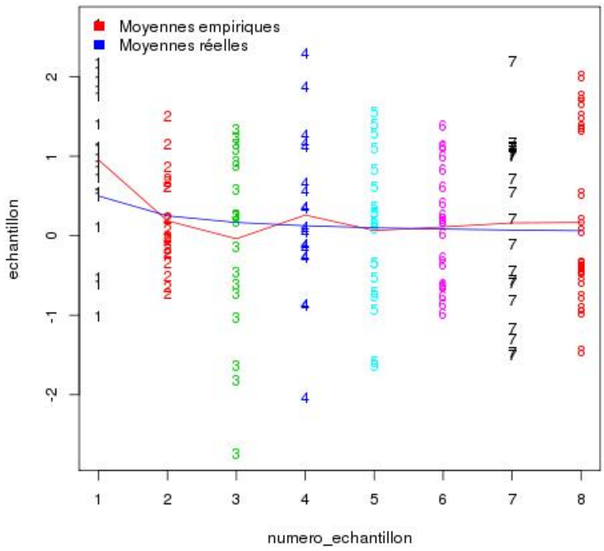
Distribution pilote pour deux lois normales indépendantes

3. Test à deux échantillons appareillés
   Ici,on compare deux échantillons congruents et incogruents. Les tests ici sont dépendants et les éléments des deux groupes sont corrélés. Concrètement, pour notre exemple d'étude psychologique, ce cas serait adapté si les personnes passant le test congruent et le test incongruent sont les mêmes, ou si certaines personnes passent les deux tests. En effet, une personne ayant un temps de réaction plus long que les autres, aura tendance à faire baisser la moyenne des temps de réaction dans les deux tests, rendant les tests corrélés.



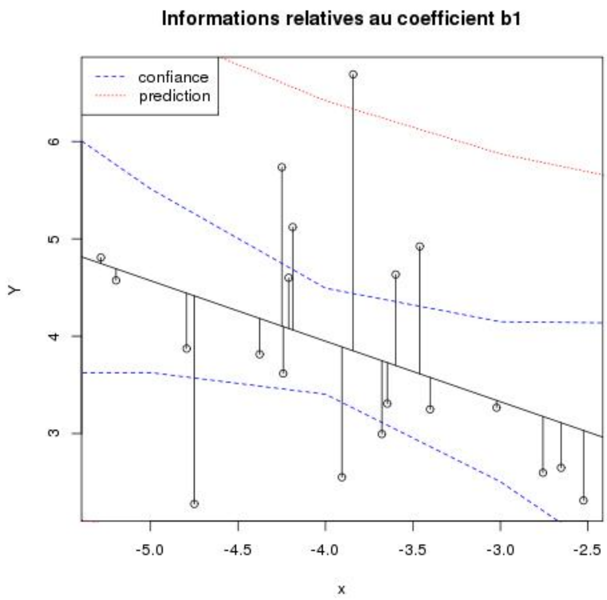
Distribution pilote pour deux lois normales appareillées

4. Test d'analyses de variances
   Dans ce test, on considère k populations, au lieu de 2 pour les deux précédents tests, en les considérant de même écart type et de même taille. Pour le test d'hypothèse, l'hypothèse nulle correspondra au cas où toutes les moyennes sont égales, et l'hypothèse alternative correspondra au cas où au moins une des moyennes est significativement différente des autres.



Régressions linéaires

5. Régressions simples et complexes.



Régressions linéaires

Pour chacun de ces cas, 5 objectifs étaient à traiter :

1. Calcul analytique (usage des fonctions de R)
2. Estimation de puissance avec génération de pilote
3. Estimation de puissance avec intervalle de confiance et estimation de taille d'échantillon
4. Tracé de graphe
5. Rajout du cas sur la plateforme (interface graphique)

1.4 Objectifs fixés

L'objectif est de fournir au psychologue statisticien un véritable outil de calcul, qui lui permettra, en fonction du cas d'école dans lequel il se trouve, de calculer la puissance de son étude en fonction de la taille de son échantillon, et d'estimer la taille d'échantillon dont il a besoin pour valider son test à la puissance voulue.
Paramètres à entrer : ?

2 Démarche

2.1 Documentation

Nous allons expliquer de manière générale la démarche pour estimer la puissance et la taille d'échantillon. La démarche était relativement similaire dans tous les cas classiques, malgré quelques spécificités.

Documentation et identification du test statistique :
Avant de commencer toute implémentation, il a tout d'abord fallu se documenter sur ces cas. Une partie de ce travail de documentation consistait à un travail de compréhension, et permettait d'identifier la statistique de test à utiliser (le test de Student pour la loi normale à un échantillon, à deux échantillons appareillés, à deux échantillons indépendants, et le test de Fisher pour le cas de l'anova)

2.2 Etude statistique sur R

Calcul analytique :
Les calculs analytiques apparaissent dans notre code et dans nos résultats par la variable p5package, et utilisent en fait des fonctions de test de R, dont nous avons du installer le package, du type power.t.test, pwr.test.anova...
Le calcul analytique necessite donc d'identifier la loi que suivra la statistique (loi de Fisher et loi de Student), et à identifier les paramètres de ces fonctions : par exemple, pour le test à un échantillon, la fonction de test s'écrit : $p5package = power.t.test(n = n, d = meand/sd, sig.level = alpha, type = "one.sample", alternative = "two.sided")$
Ici, meand représente la moyenne de l'échantillon à simuler, alpha représente l'erreur de première espèce, et le reste des paramètres correspond au type du test. Ces fonctions de R nous serviront de référence pour vérifier la fiabilité et la cohérence de nos estimateurs, et apparaissent dans nos tableaux de résultats.

Estimation de puissance avec génération du pilote :
Cette partie de la démarche s'intéresse à une incertitude sur la taille d'effet. Le but ici est de calculer nous même la puissance, en générant un pilote et en calculant la statistique de test. Le pilote correspond à une petite étude, qui permettrait au chercheur d'obtenir des informations préliminaires. Dans le cas de notre simulation, il s'agit d'une loi normale de petite taille, à partir duquel on dupliquera les informations grace aux méthodes de Montecarlo. Montecarlo sera appliqué le nombre voulu de fois, à une certaine taille, et c'est sur les élèments que l'on obtient grac à cette méthode que l'on calculera la statistique de test.
Une fois cette statistique de test obtenue, on peut la comparer au quantile de la loi correspondante. Le nombre de fois où cette statistique sera supérieure au quantile, divisé par le nombre total de boucles de Montecarlo, nous donnera la puissance du test. Dans un premier temps nous avons stocké nos résultats dans un tableau, afin de vérifier sa fiabilité en comparaison des estimateurs de R.

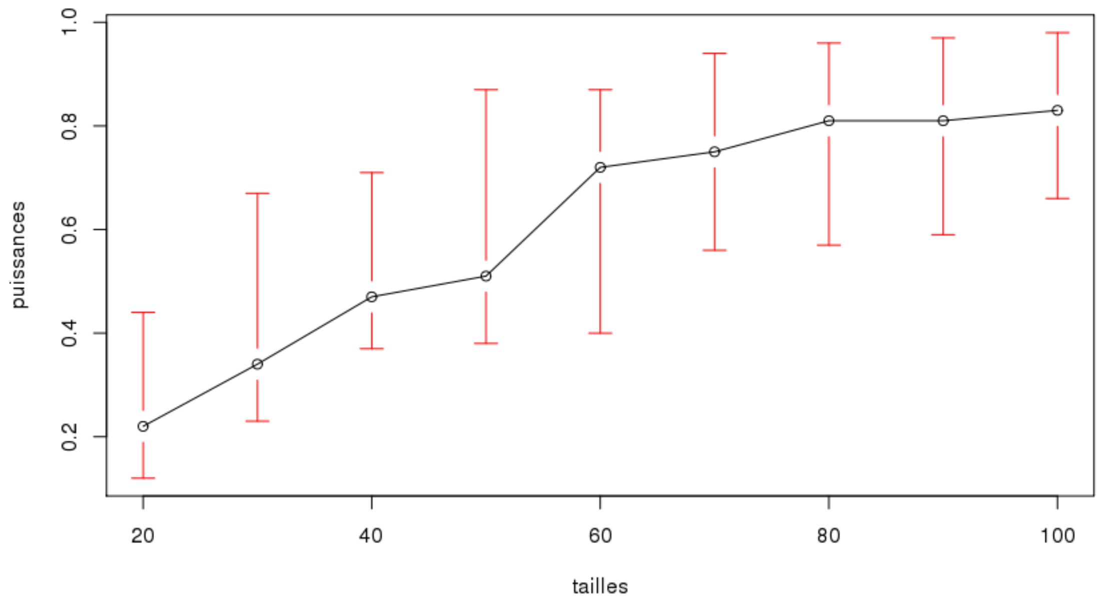
n	sd	runs	p5_hand	p5_package	fval_mean	thr
2000	1	1000	0.143	0.1444108	1.256376	1.884576

Tableau obtenu pour le test à un échantillon

Estimation de puissance et de taille d'échantillon avec intervalle de confiance :
Dans cette partie, le but est de générer un graphe de la puissance en fonction de la taille d'échantillon (taille des prélèvements de Montecarlo), avec un intervalle de confiance sur la puissance.La démarche est simple : pour toutes les tailles de l'intervalle choisi (tailles de Montecarlo), on réitère l'étape expliquée

précédemment, tout en calculant les valeurs inférieures, moyennes et supérieures pour la puissance, calculées à partir d'un intervalle de confiance pour les moyennes généré par R (fonction confint) et par bootstrapping pour les écarts-types (comme expliqué plus haut). Par interpolation, on récupère la valeur minimale de la taille d'échantillon dont on a besoin à une puissance donnée.

Tracé de graphe :
Ensuite, nous traçons la courbe de la puissance en fonction de la taille des échantillons de Montecarlo, avec intervalle de confiance. Ce graphe sera intégré à l'interface graphique, et sera un des produits des fonctionnalités de cas simple. Également, nous traçons les histogrammes des échantillons, comme présentés dans les figures en début de rapport, ainsi que les courbes des lois normales associées.



Puissance avec intervalle de confiance pour les deux normales indépendantes

Création d'un package :
Avant de pouvoir passer à l'implémentation du cas dans l'interface graphique, nous avons fait le choix de créer des packages pour nos fonctions R, documentant ainsi ces fonctions et permettant un appel plus simple depuis l'interface.
Les documentations des packages sont fournies et permettent d'accèder aux implémentations des fonctions en R ainsi qu'à leurs descriptions.
Le package "regression" comprends la génération d'observation par regression (simple, multiple, avec effets mixtes, et avec intéractions), le package "cas.simples" comprends les 3 cas de t-test de Student et le package "anova.regression" comprends le cas de l'Anova et des regressions simple et multiple.

2.3 Interface graphique

Implémentation de l'interface :
Avant toute chose, notre volonté en implémentant l'interface graphique était de la rendre la plus ergonomique et simple d'utilisation possible, et de rendre la navigation claire entre les différents cas d'utilisation. On retrouve par exemple une barre de navigation en haut de chaque fenêtre, des boutons suivant et précédent, etc...
Nous avons choisi de la développer dans le langage Java, très fourni et pratique pour l'implémentation des interfaces graphiques, et avec lequel nous avons déjà eu l'occasion d'en implémenter.
Finalement la jonction entre les langages R et Java se fait avec le package rJava de R, et l'instanciation dans Java d'une Rengine appelant les packages R que nous avons créé La structure globale du code Java que nous avons produit pour cette interface se retrouve dans les diagrammes de classes qui suivent.

Toutes les classes de l'interface sont les suivantes :

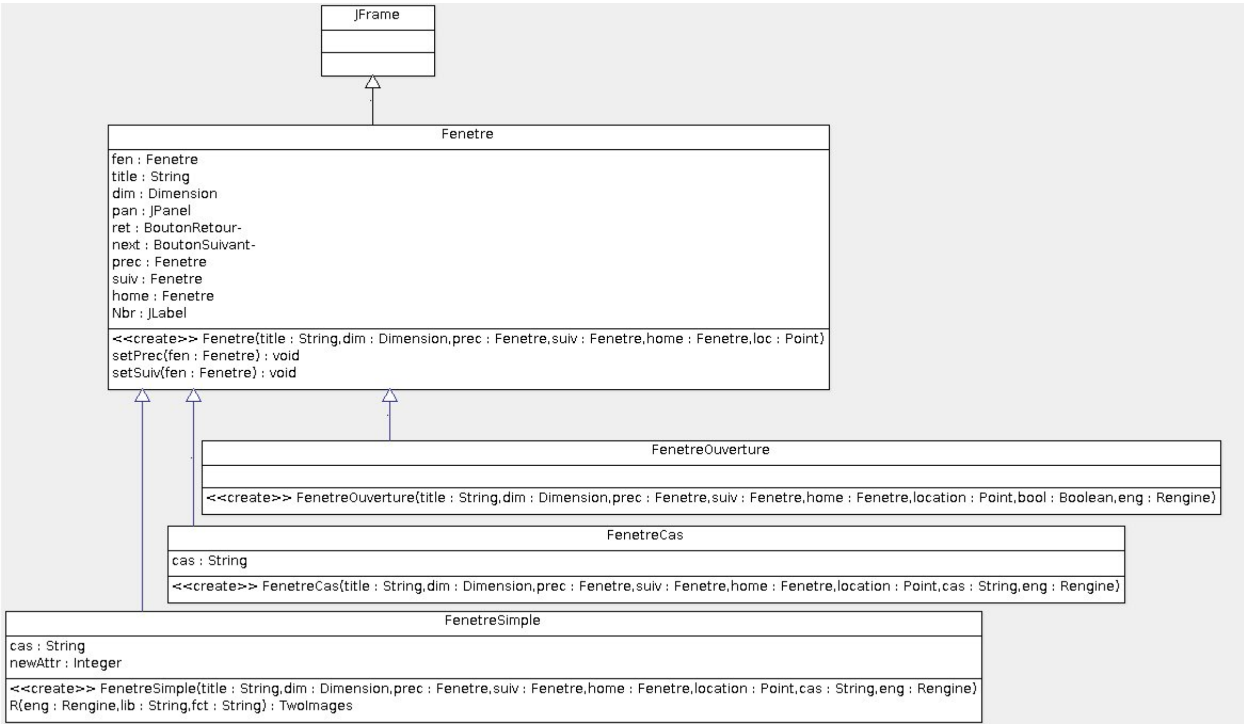


Diagramme des classes des différentes fenêtres qui composent notre interface graphique, toutes héritées de la classe JFrame.

Toutes les fenêtres héritent de la super-classe Fenetre, qui contient les caractéristiques communes de celles-ci, comme par exemple la barre de navigation.

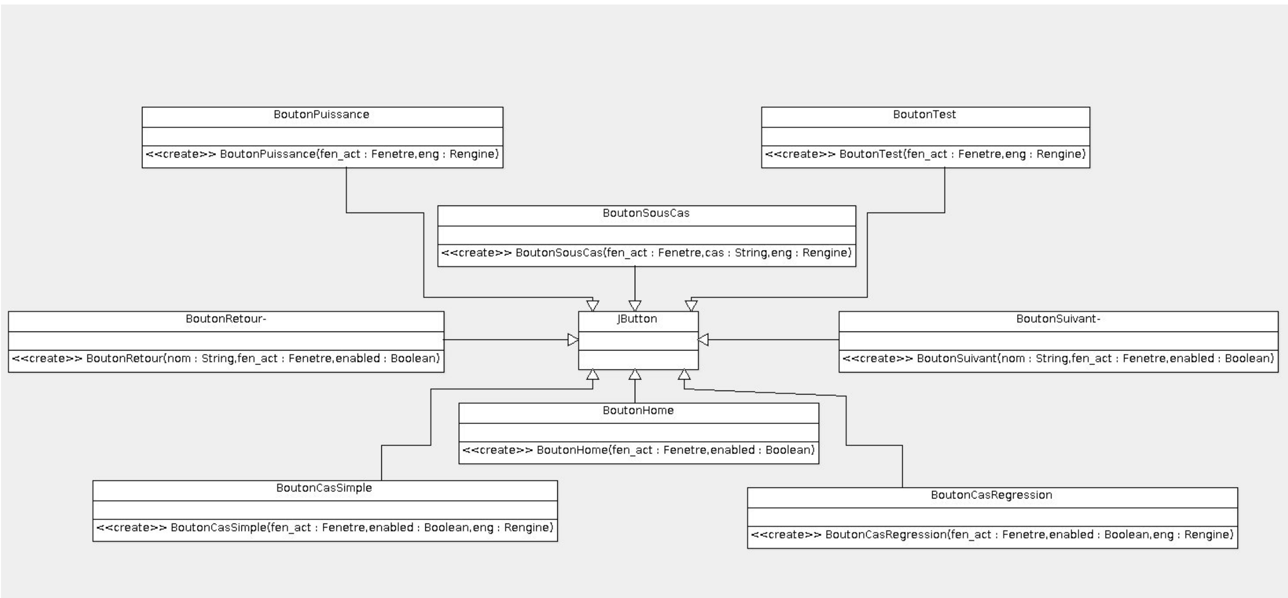


Diagramme des classes des différents boutons qui composent notre interface graphique, toutes héritées de la classe JButton.

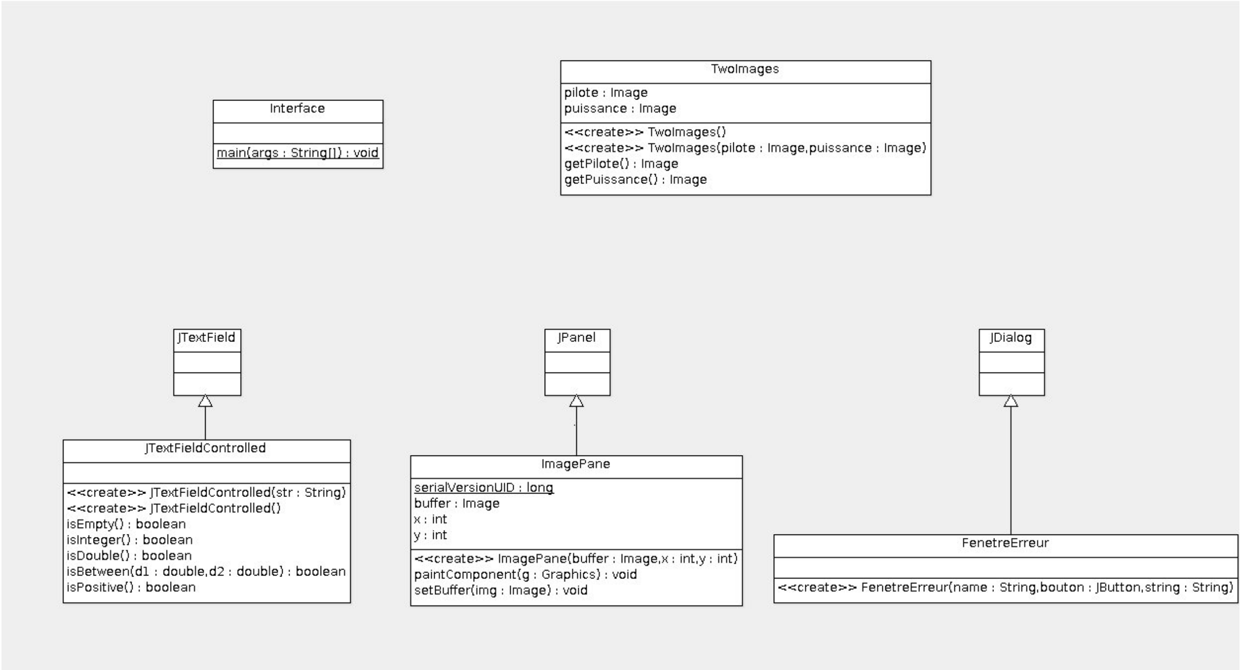


Diagramme des autres classes.

Ces trois premières figures présentent toutes les classes et les différents héritages de notre implémentation. Les figures suivantes donnent deux diagrammes de classes à différents niveaux de précision pour comprendre comment ces classes intéragissent entre elles.

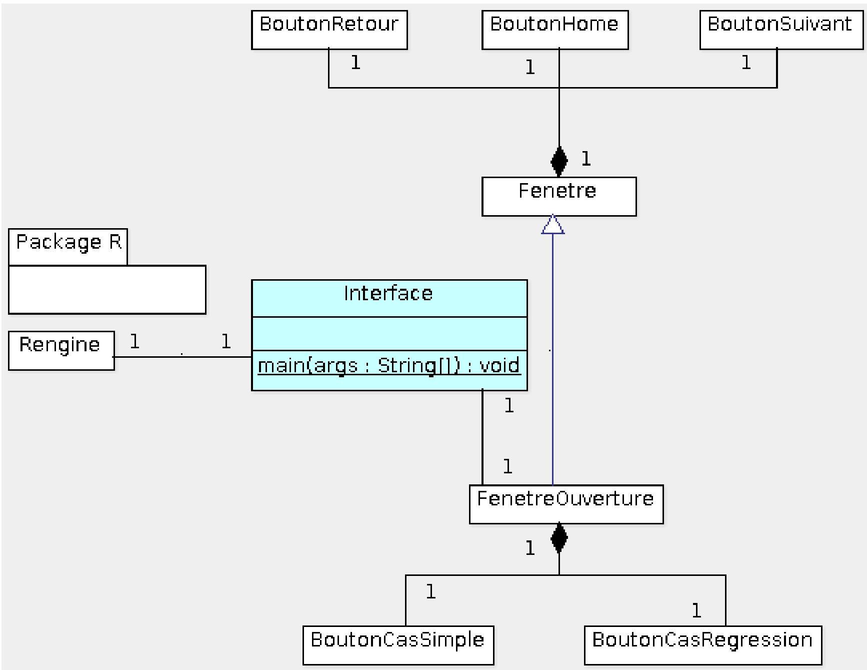


Diagramme des classes à l'éxecution du main avec l'instanciation d'une FenetreOuverture.

Lorsque l'on lance notre application, une connexion entre R et Java se créait avec l'instanciation d'une Rengine. Celle ci charge les différents packages que nous avons créer pour permettre l'appel des fonctions de calcul de puissance. Une FenetreOuverture s'ouvre. Héritant de la classe Fenetre, elle se compose d'un BoutonHome, d'un BoutonRetour et d'un BoutonSuivant et en tant que FenetreOuverture elle se compose aussi d'un BoutonCasSimple et d'un BoutonCasRegression permettant de choisir le cas que l'on veut utiliser.

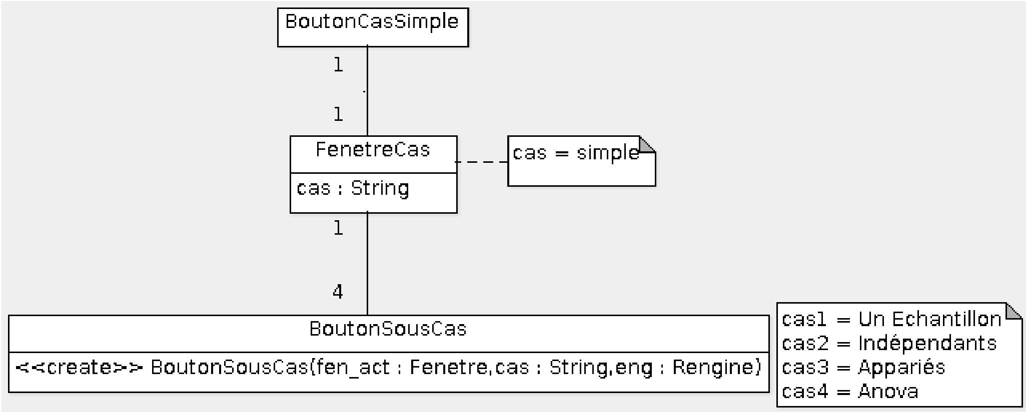


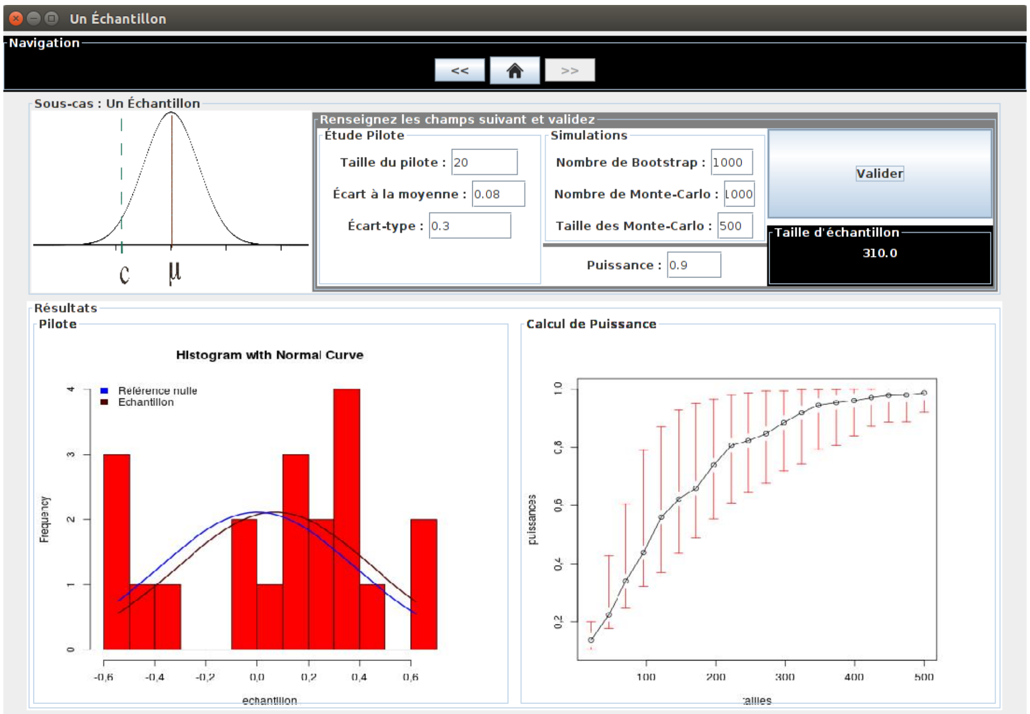
Diagramme des classes à l'éxecution du main avec l'instanciation d'une FenetreOuverture.

Lorsque l'on clique sur le BoutonCasSimple, une FenetreCas s'ouvre avec comme attribut cas = simple. Celle-ci se compose de 4 BoutonSousCas laissant le choix entre différents t-test ou une anova.
On continue alors encore à descendre dans l'arborescence de l'application jusqu'à choisir le sous-cas qui nous intéresse, et alors l'on rentre les paramètres désirés dans l'interface qui s'occupera d'appeler la fonction adéquates des packages avec les arguments correspondants.

3 Application finale

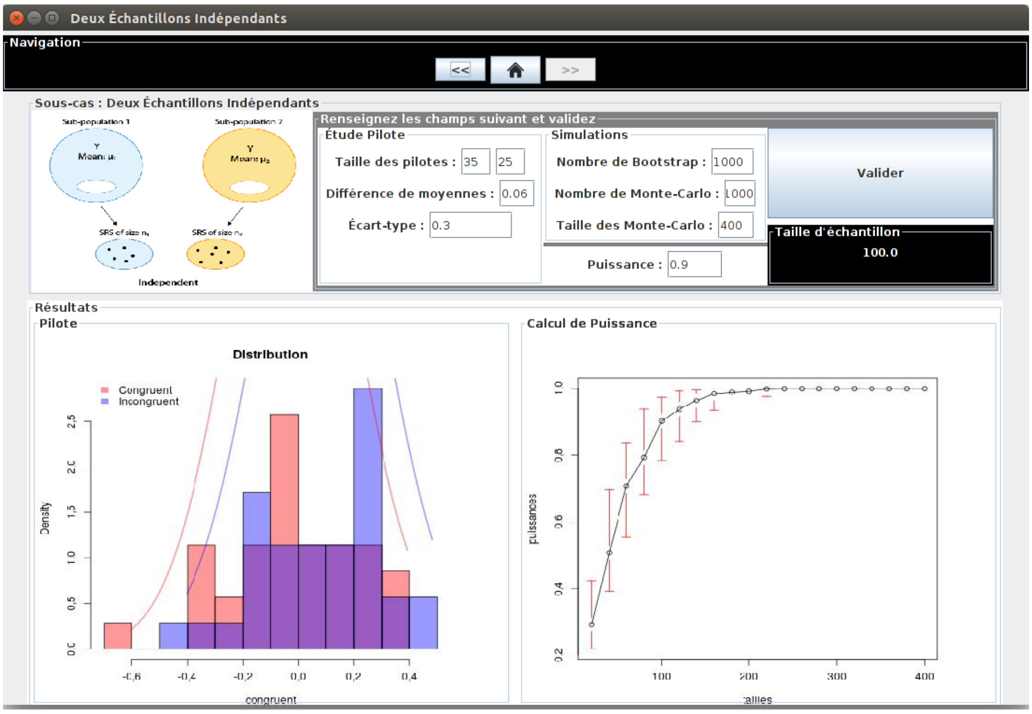
3.1 Fonctionnalités

Une fois tous les champs correctement renseignés (cf. Manuel Utilisateur pour le détail sur les restrictions des champs), le programme exécute les calculs avec les fonctions R, l'interface se bloque en attente du résultat et retourne les graphes de distributions du pilote et de calcul de puissance. Elle fournit alors aussi la taille d'échantillon recherché par le statisticient (ou le psychologue) qui se sert de l'interface.



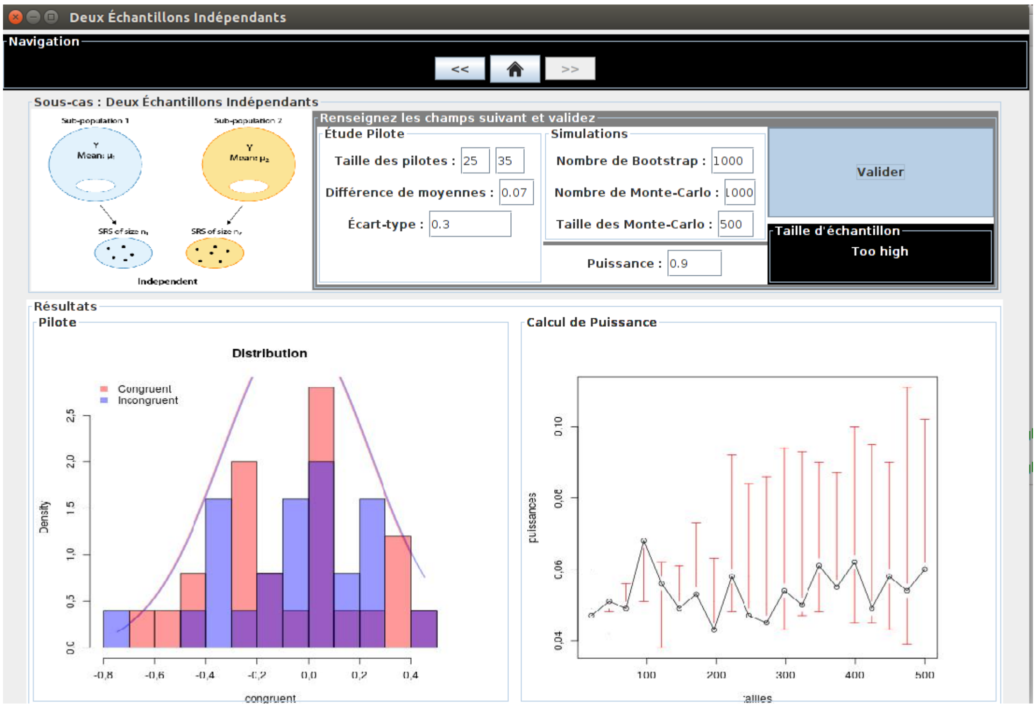
Etude statistique sur pilote dans le cas d'un échantillon unique.

Etude statistique sur pilote dans le cas de deux échantillons appariés.



Etude statistique sur pilote dans le cas de deux échantillons indépendants.

Les trois figurent précédentes montrent des cas où la présence d'un effet est démontrable. Cela vient de l'écart existant bel et bien entre les moyennes des deux normales de nos échantillons.
Si le pilote est moins bon et que les moyennes de ces normales sont plus resserées, alors la puissance peinera à augmenter avec la taille d'échantillon et on peut même dans de rares cas où les moyennes sont quasiment identiques obtenir des résultats démontrant l'absence d'effet à observer, comme dans le cas suivant :



Etude statistique sur deux échantillons indépendants dont les moyennes sont quasiment égales et donc où la taille d'effet est quasiment nulle. La puissance ne dépasse pas ici 0.07.

3.2 Performances et Fiabilité

Durant toute la durée de notre projet, nous avons eu besoin de vérifier la fiabilité de nos estimateurs. Pour faire cela, nous comparions nos estimateurs (de puissance et de taille d'échantillon), aux estimateurs disponibles dans les packages de R (power, pwr,...). Également, il nous a fallu vérifier la cohérence de nos estimateurs par rapport à ces estimateurs repères : lorsque le nombre de boucles de Montecarlo devenait très grand, nos estimateurs devaient tendre vers les estimateurs des packages. Également, lorsque de modifications des tailles du pilote, des écarts types ou des moyennes de référence, les variations devaient être les mêmes.
De plus, on a pu vérifier la cohérence de nos estimateurs en les comparant entre eux. Par exemple, le cas d'analyse de variance et le cas de deux échantillons indépendants, revient au même, et en fixant toutes les autres variables à l'identique, on devait obtenir le même résultat.

3.3 Un projet de recherche : ressenti général

La particularité de ce sujet réside dans sa différence par rapport aux projets de développement auquel nous sommes habitués. En effet, le sujet est vaste et ne prédispose pas d'un rendu spécifique. Toute la difficulté était donc de décider de l'objectif, et de la manière dont nous allions atteindre cet objectif. Une partie importante du travail fut donc de se documenter sur le sujet, et de se renseigner sur ce qui avait déjà été fait dans le domaine.
De même nous avons été très libres dans le choix de direction pour ce projet, le développement d'une interface graphique est le choix que nous avons fait mais nous aurions pu partir sur tout autre chose. C'est donc un projet que nous avons porté avec nous, que nous nous sommes approprié et sur lequel nous avons beaucoup aimé travailler.

4 Conclusion

Pour conclure, nous pouvons dire que, à travers une démarche de documentation et de recherche, nous avons élaboré une application qui permet au psychologue statisticien d'estimer la puissance et la taille d'échantillon de ses tests, visant à observer une certaine taille d'effet, dans un ensemble de cas usuels. Pour cela, nous avons procédé par incrémentation : nous avons tout d'abord généré une

puissance à partir d'une étude pilote en usant des méthodes de Montecarlo, puis nous avons rajouté une incertitude et mis en place des intervalles de confiance avec des méthodes de Bootstrap. Enfin, nous avons connecté notre code R avec notre interface Java grace à rJava.
Ce travail pourrait être étendu aux études de cas réèls : le psychologue fournirait les résultats de son étude à l'application (son ou ses échantillons), et après calcul des paramètres de base (moyenne, écart type), en fonction du cas dans lequel on se situait, l'application fournirait au psychologue la puissance et la taille d'effet correspondant à son étude réelle.

5 Références

1. Documentation : **Stéphane**, **CHAMPELY**. *Tests statistiques paramétriques : Puissance, taille d'effet et taille d'échantillon (sous R)*