

TP R 2: statistiques bayésiennes

BINET Thomas

Statistiques biomédicales

Table des matières

Tests Bayésiens	1
Echantillon de Bernoulli	1
Echantillon de loi normale	2

Important : Ceci est un document R notebook. En cliquant sur knit vous aurez le choix entre un document html, pdf ou word. Le TP doit être rendu sous la forme d'un fichier NOM_prenom_Bayes.html. Il est à remettre dans moodle. CTRL+Alt+i permet d'ouvrir une cellule de code compilable.

Tests Bayésiens

Echantillon de Bernoulli

Cette partie a pour but d'étudier les erreurs de type I de plusieurs tests, dans le cas standard d'une comparaison d'efficacité d'un nouvel agent avec un médicament de référence. On suppose que l'efficacité du médicament de référence est $p_0 = 0.7$ et les hypothèses antagonistes sont $H_0 : p \leq p_0$ et $H_1 : p > p_0$.

1. La fonction suivante renvoie la moyenne, la p-valeur, et la borne supérieur de l'intervalle de fluctuation sous l'hypothèse H_0 , ci-dessus.

```
test.p = function(p0, vector, conf.level) {  
  if (length(vector)==0) { cat("Erreur ! Le vecteur ", substitute(vector), "est vide.\n")}  
  else {  
    n = length(vector)-sum(is.na(vector))  
    moyenne = mean(vector, na.rm=T)  
    p_val = 1- pnorm(moyenne, p0, sqrt(p0*(1-p0)/n))  
    borne_sup = qnorm(conf.level, p0, sqrt(p0*(1-p0)/n))  
    return(list(moyenne=moyenne, p_val=p_val, borne_sup=borne_sup))  
  }  
  
x = rbinom(100,1,0.7) ; test.p(p0=0.7,vector=x,conf.level = 0.95)
```

```
## $moyenne  
## [1] 0.78  
##  
## $p_val  
## [1] 0.0404278  
##  
## $borne_sup  
## [1] 0.7753767
```

Créer une fonction permettant d'évaluer par simulation numérique le pourcentage d'erreur de type 1 commises. Cette fonction prendra en paramètre d'entrée : `N`, `n`, `p0`, `conf.level` où `N` est le nombre de tests simulés, `n` la taille de l'échantillon et `p` la vraie probabilité de succès du nouvel agent.

2. Nous allons maintenant effectuer des tests bayésiens. On choisit un a priori conjugué avec les données (loi Beta) et la fonction de perte considérée est la fonction "0-1" :

$$l(\theta, T) = \alpha_0 \mathbb{I}_{\{T=1, \theta \in \Theta_0\}} + \alpha_1 \mathbb{I}_{\{T=0, \theta \in \Theta_1\}}.$$

Le tests de Bayes T pour cette fonction de perte et un a priori Π est :

$$T(X_1^n) = \mathbb{I}_{\left\{B_n \leq \frac{\alpha_1 \Pi(\Theta_1)}{\alpha_0 \Pi(\Theta_0)}\right\}}, \text{ avec } B_n = \frac{\Pi_n(\Theta_0) \Pi(\Theta_1)}{\Pi_n(\Theta_1) \Pi(\Theta_0)}.$$

Créer une fonction permettant d'effectuer un test bayésien. Cette fonction prendra en paramètre d'entrée : `p0`, `vector`, `a`, `b` (les paramètres de l'a priori), `alpha0`, `alpha1`.

3. Créer une fonction, `err_I_b`, permettant d'étudier l'erreur de type 1 dans le sens fréquentiste (données générées sous une unique loi de paramètre `p`) . **4.** Tester la fonction pour les différentes circonstances suivantes : `n=10, 50, 100` et $\alpha_0/\alpha_1 = \Pi([0, p_0])/\Pi([p_0, 1])$ ou $\alpha_0/\alpha_1 = 19/1$. On choisira un a priori uniforme.

Echantillon de loi normale

Les médecins souhaitent qu'un traitement amène le biomarqueur b_0 à se situer en moyenne dans l'intervalle $I = [45, 65]$. Chacune des 5 doses testées ont été soumises à 25 patients ; les patients ne sont traités qu'à une dose (125 volontaires). Pour chacune des doses, on souhaite calculer la probabilité que le biomarqueur soit dans l'intervalle. Le modèle bayésien suivant est utilisé :

$$X \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2),$$

avec une loi conjuguée pour les paramètres θ et σ (conjugate prior).

```
data=read.csv("biomarker_dose.csv")
```

1. Dans un premier temps, on suppose la variance connue est identique dans chaque groupe : $\sigma^2 = 6$. Quelle est la dose la plus adaptée

2. Maintenant, la variance n'est plus connue. Les probabilités sont à évaluer par méthode de Monte Carlo en utilisant comme a priori la loi normale inverse-gamma : $\theta, \sigma^2 \sim \mathcal{N}\Gamma^{-1}(\mu, \lambda, \alpha, \beta)$. Pour générer sous l'a posteriori, on a :

$$x \mid y, \mu, \lambda \sim \mathcal{N}(\mu, \sigma^2/\lambda) \quad \text{et} \quad y \mid \alpha, \beta \sim \Gamma^{-1}(\alpha, \beta) \Rightarrow (x, y) \sim \mathcal{N}\Gamma^{-1}(\mu, \lambda, \alpha, \beta)$$