

TPSTAT3

You ZUO

15 mars 2019

1. Ajuster une loi Bernoulli

Soit une loi Bernoulli (rbinom) avec $p = 0.7$. Simuler un échantillon i.i.d de taille $N = 10$

```
## [1] 0 1 0 1 1 1 1 1 1 0
```

1.a)

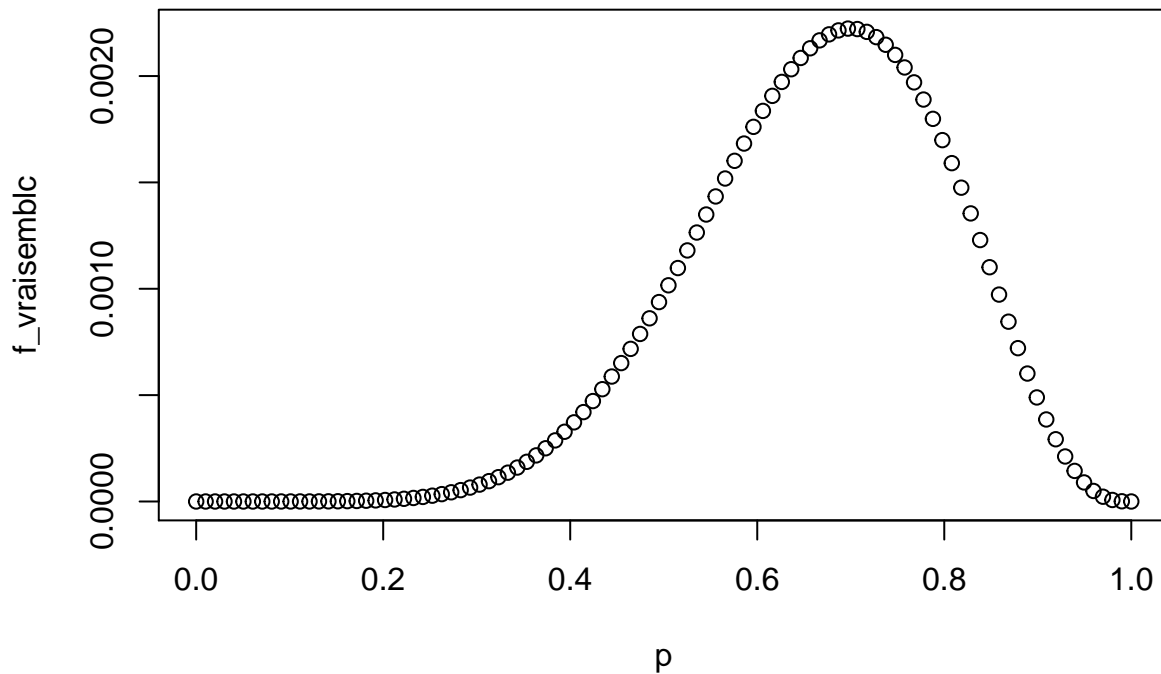
Pour estimer p , on peut utiliser la moyenne empirique:

```
## [1] 0.7
```

1.b)

On va générer une fonction de vraisemblance, nommée `L_bern`, qui donne la vraisemblance d'un échantillon de Bernoulli pour une valeur donnée de p .

1.c)



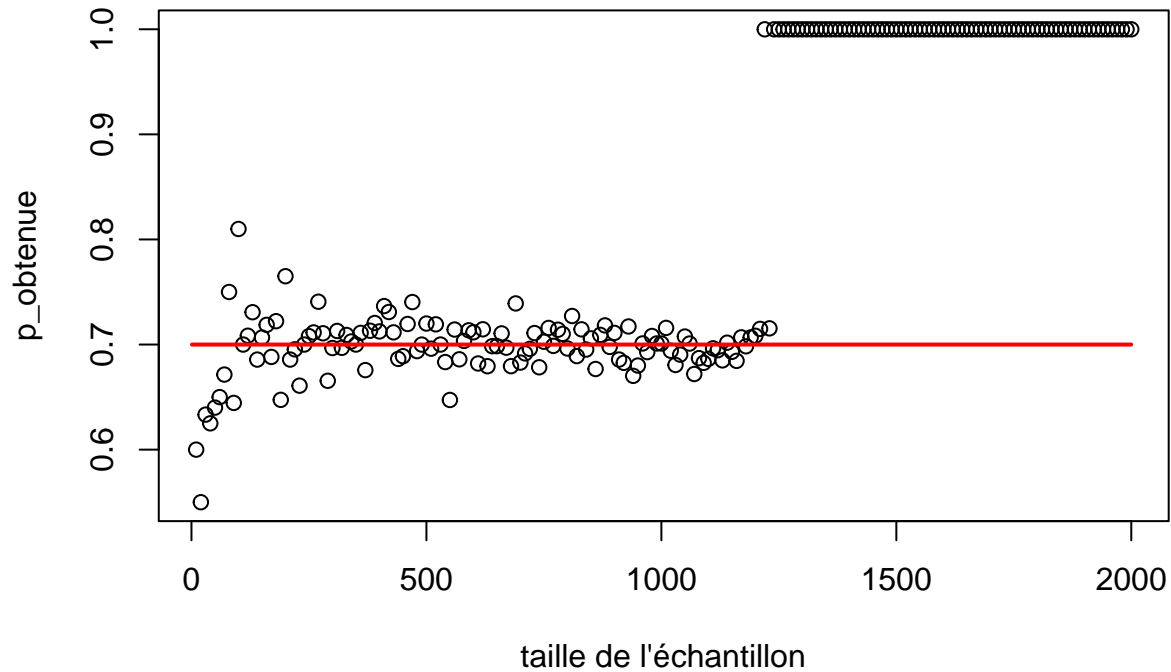
Selon la courbe, on a trouvé que la fonction de vraisemblance atteint un maximum lorsque la probabilité est proche de sa valeur réelle.

1.d)

```
## $maximum
## [1] 0.6999843
##
## $objective
## [1] 0.002223566
```

1.e)

On va tester avec des échantillons de taille allant de $N = 10$ à $N = 2000$



On a remarqué que quand la taille de l'échantillon devient grand, les résultats de la fonction optimize se rapproche de p , la probabilité théorique. Mais quand N dépasse un certain nombre très grand, les résultats sont faux : les résultats sont tous égaux à 1. On peut combattre l'instabilité des calculs en passant au log-vraisemblance.

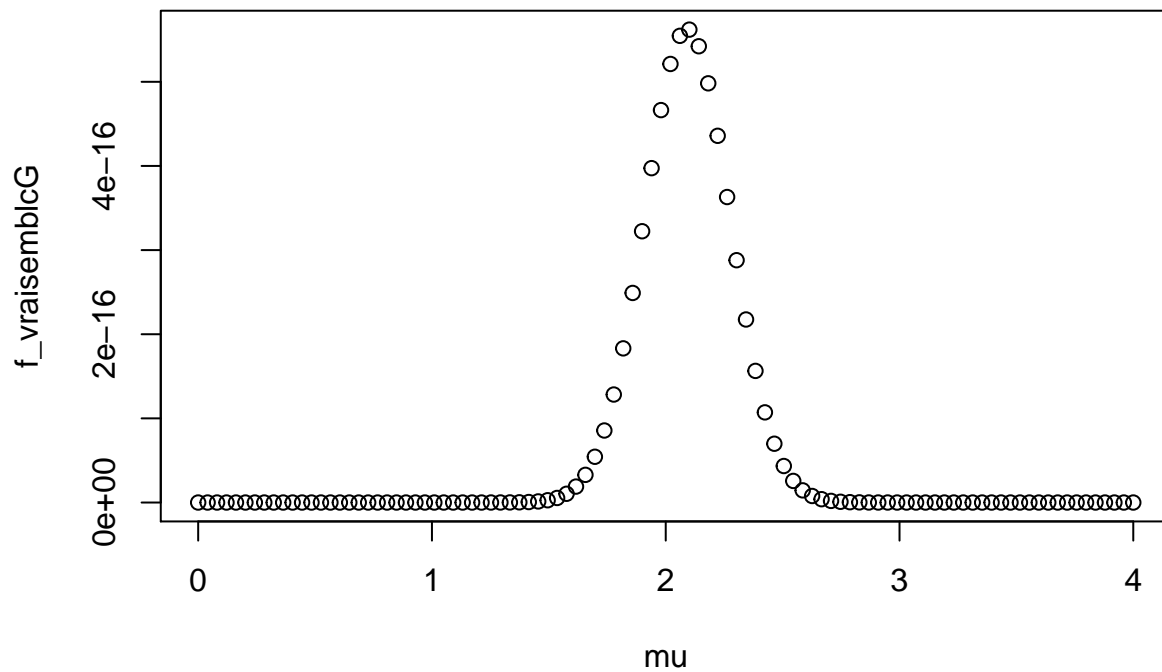
1.f)

```
## $maximum
## [1] 0.5800008
##
## $objective
## [1] -68.0292
```

On considère que c'est une loi Bernoulli, parce que l'échantillon ne contient que 2 types de valeurs distinctes: 0 ou 1.

2. Ajuster d'une loi normale d'écart type connu

2.b)

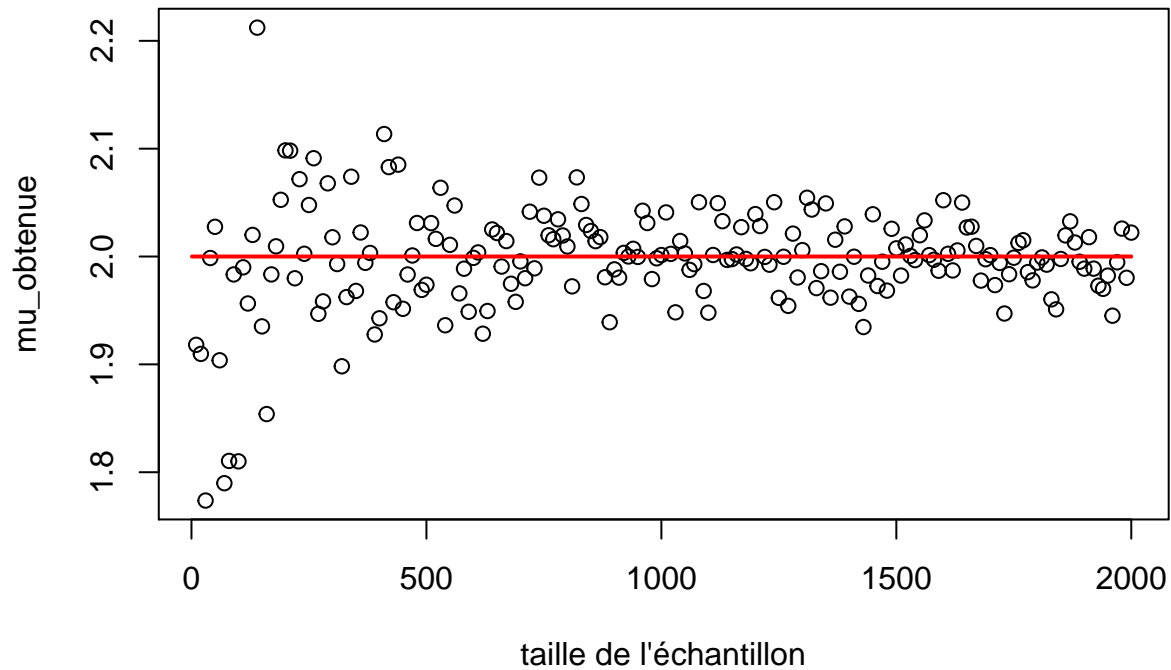


On a remarqué que la fonction de vraisemblance pour 100 lois normales atteint un maximum lorsque la valeur de μ est proche de sa valeur réelle 2.

2.c)

```
## $maximum
## [1] 2.091696
##
## $objective
## [1] 5.627342e-16
```

2.d)



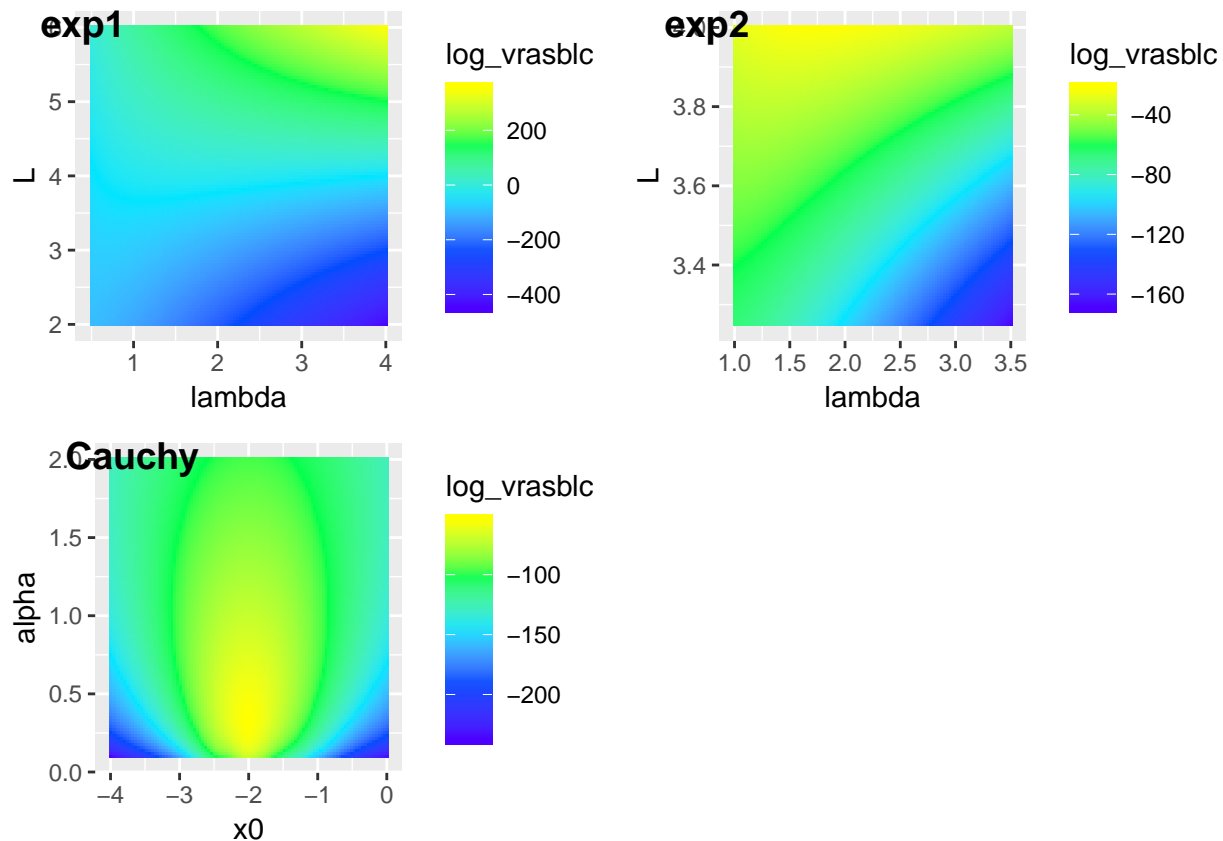
On a remarqué que quand la taille de l'échantillon devient grand, les résultats de la fonction optimize se rapproche de $\mu = 2$, la moyenne théorique.

3. Ajuster d'une loi à plusieurs paramètres

Soit $N = 50$

On calcule la log-vraisemblance des échantillons qui ont été générés, en faisant varier un paramètre à la fois.

Afin de voir l'influence du changement de deux paramètres sur la log-vraisemblance, j'ai choisi d'utiliser les coordonnées x et y pour représenter les valeurs de mes deux paramètres. Pour représenter la valeur de la log-vraisemblance, j'ai utilisé un dégradé de couleur. Plus la couleur est claire, plus la valeur de la log-vraisemblance est grande.



D'après les résultats des figures ci-dessus, on peut voir que les résultats de l'estimation de la distribution de Cauchy avec le maximum de vraisemblance sont fondamentalement cohérents avec la théorie, mais que les résultats de la distribution exponentielle sont assez différents des vraies valeurs des paramètres.

Cependant, après d'avoir recherché sur la log-vraisemblance pour la loi exponentielle:

$$f(x) = \lambda e^{-\lambda(x-L)} \quad (x > L)$$

on a trouvé que quand on a $L \geq 4$ (L en théorie), le terme $-\lambda(x-L)$ devient positif et augmente selon les augmentations de L et λ , et la log-vraisemblance sera beaucoup élevée dû à cela, donc il faut limiter le champ d'action de L aux valeurs positives inférieures à 4. Puis on peut voir selon le graphe en droite que le résultat devient plus rationnel.

On compare la valeur obtenue de cette façon à la valeur qu'on obtient en appliquant le solveur mle

```
##
## Call:
## mle(minuslogl = Loi_exp, start = list(lambda = 0, L = 0), method = "L-BFGS-B",
##     lower = c(0.01, 2), upper = c(4, 4))
##
## Coefficients:
##     lambda      L
## 1.752967 4.000000
##
## Call:
## mle(minuslogl = Loi_cauchy, start = list(x0 = -5, alpha = 0),
```

```
##      method = "L-BFGS-B", lower = c(-5, 0.1), upper = c(5, 10))
##
## Coefficients:
##          x0          alpha
## -2.0116108  0.3260133
```