

Test paramétrique et de l'estimation de la loi géométrique et de Cauchy

TRAN Quoc Nhat Han - Andrien Wartelle

1^{er} octobre 2018

Sommaire

1	Loi de Cauchy	1
1.1	Rappel	1
1.2	Estimer les paramètres inconnus	2

Résumé

Ce rapport est à montrer l'application de certaines tests statistiques pour la loi géométrique et la loi de Cauchy. Chaque section consiste de trois étapes séparées :

1. *Estimer les paramètres inconnus*
2. *Test de l'estimation*
3. *Test d'adéquation*

1 Loi de Cauchy

1.1 Rappel

Soit f_X la fonction de densité de Cauchy de deux paramètres x_0 et a ($a > 0$), définie par :

$$f_X(x) = \frac{1}{\pi a \left(1 + \left(\frac{x-x_0}{a}\right)^2\right)} = \frac{1}{\pi} \frac{a}{(x-x_0)^2 + a} \quad (1.1)$$

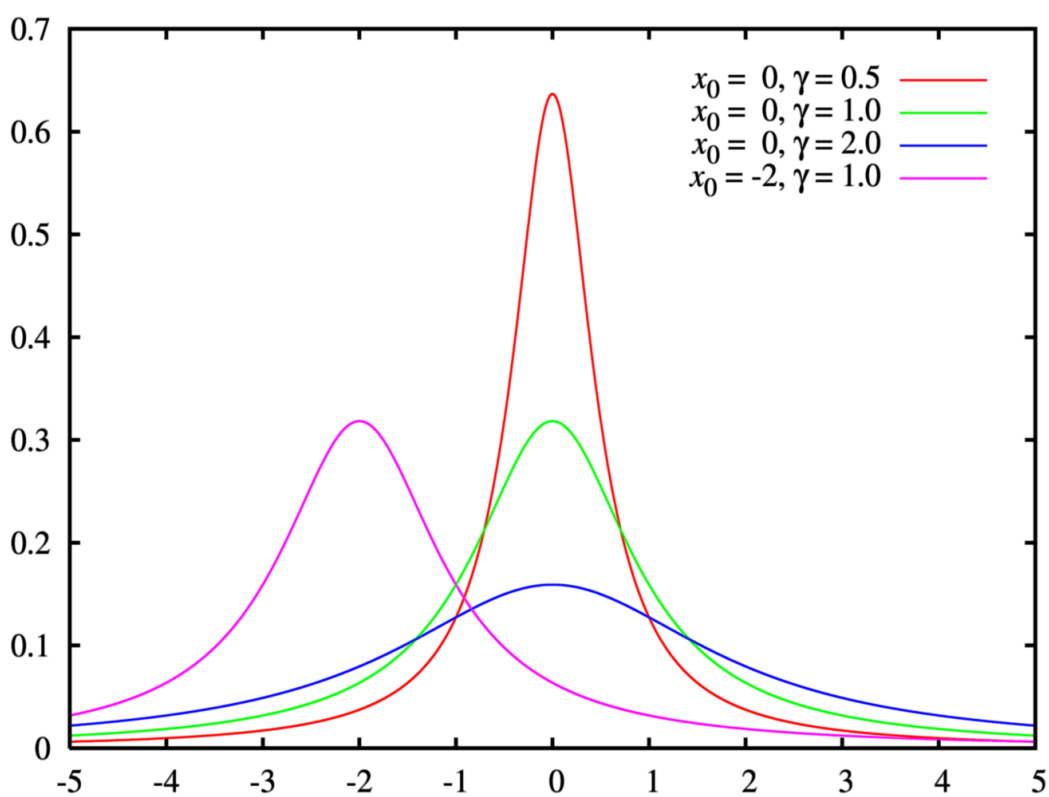


FIGURE 1 – Distribution théorique de la loi Cauchy. Source : [1]

a est dite l'échelle de la fonction, et x_0 est son médian.

La loi de Cauchy n'admet ni espérance ni écart type.

La fonction de répartition :

$$F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x-x_0}{a}\right) + \frac{1}{2} \quad (1.2)$$

1.2 Estimer les paramètres inconnus

Calcul théorique

Soient n réalisations x_1, x_2, \dots, x_n . Assumons que ces données suivent la loi de Cauchy (1.1).

Nous allons utiliser la méthode de maximum de rapport de vraisemblance.

En l'absence d'information de la distribution, nous assumons que ces mesures sont indépendants. Posons une variable aléatoire X_i correspondante à chaque réalisation x_i pour $i = \overline{1, n}$.

On écrit la loi conjointe de ces n variables :

$$\begin{aligned} L(X) &= L(X_1, X_2, \dots, X_n) = \prod_{i=1}^n f_X(X_i) \\ &= \prod_{i=1}^n \frac{1}{\pi} \frac{a}{(x_i - x_0)^2 + a} \\ &= \pi^{-n} \prod_{i=1}^n \frac{a}{(x_i - x_0)^2 + a} \end{aligned}$$

On cherche l'optimum maximale.

$$\begin{aligned} \frac{\partial L}{\partial x_0}(X) &= \pi^{-n} \sum_{j=1}^n \left(-\frac{a(2x_j - 2x_0)}{[(x_0 - x_j)^2 + a]^2} \prod_{i=1; i \neq j}^n \frac{a}{(x_i - x_0)^2 + a} \right) \\ &= \pi^{-n} \left(\prod_{i=1}^n \frac{a}{(x_i - x_0)^2 + a} \right) \left(-\sum_{j=1}^n \frac{2x_0 - 2x_j}{(x_0 - x_j)^2 + a} \right) \\ \frac{\partial L}{\partial x_0}(X) &= 0 \Leftrightarrow \sum_{j=1}^n \frac{x_0 - x_j}{(x_0 - x_j)^2 + a} = 0 : \text{insolvable par la main} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial a}(X) &= \pi^{-n} \sum_{j=1}^n \left(\frac{(x_j - x_0)^2 + a - a}{(x_j - x_0)^2 + a} \prod_{i=1; i \neq j}^n \frac{a}{(x_i - x_0)^2 + a} \right) \\ &= \pi^{-n} a^{n-1} \frac{\sum_{j=1}^n (x_0 - x_j)^2}{\prod_{i=1}^n ((x_i - x_0)^2 + a)} > 0 \forall a > 0 \end{aligned}$$

Résolution par R

On génère une échantillon avec a et x_0 au choix. Puis, utiliser la fonction *mle* (Maximum Likelihood Estimator) de librairie *stats4* avec 2 valeurs initiales \hat{a} et \hat{x}_0 pour estimer a et x_0 .

Nous essayons d'estimer x_0 et a directement. L'algorithme d'approximation implémenté dans R a besoin un bon point de départ, sinon le résultat obtenu variera grossièrement.

- Etant donné que la médiane est théoriquement aussi x_0 , choisissons x_0 comme la médiane de l'échantillon.
 - Nous avons $f_X(x_0) = \frac{1}{\pi a}$. En plus, $f_X(x_0)$ est la valeur maximale d de densité. Prenons alors $\hat{a} = \frac{1}{\pi d}$.
- Testons avec $x_0 = 13$ et $a = 0.5$.

```
1 # fixer le germe et paramètres réels
2 set.seed(2018)
3 real_x0 = 13
4 real_a = 0.5
5 nbreaks = 20
6
7 # générer échantillons
8 N = 1000
9 x = rcauchy(n = N,
10            location = real_x0,
11            scale = real_a)
12
13 # définir fonction log négative pour MLE
14 ll = function(location, scale) {
15     dist = suppressWarnings(dcauchy(x,
16                                     location,
17                                     scale,
18                                     log = TRUE))
19     -sum(dist)
20 }
21
22 # MLE (Maximum Likelihood Estimator)
23 library(stats4)
24 hat_x0 = median(x)
25 d = hist(x,
26          breaks = nbreaks,
27          plot = FALSE) # histogram
28 hat_a = 1 / (pi * max(d[["density"]]))
29 result = mle(ll,
30             start = list(location = hat_x0, scale = hat_a))
31
32 # illustrer
33 hist(x,
34      freq = F,
35      breaks = nbreaks,
36      col = "green",
37      xlab = "x",
38      ylab = "Densité",
39      ylim = c(0, 0.2),
40      main = "Loi de Cauchy - réalité vs théorique") # histogramme
41 curve(dcauchy(x, real_x0, real_a),
42       add = TRUE,
```

```

43     col = "black") # théorique
44 curve(dcauchy(x, result@coef[1], result@coef[2]),
45       add = TRUE,
46       col = "red") # réalité
47 legend("topright",
48       legend = c("histogramme", "théorique", "réalité"),
49       fill = c("green", "black", "red"))

```

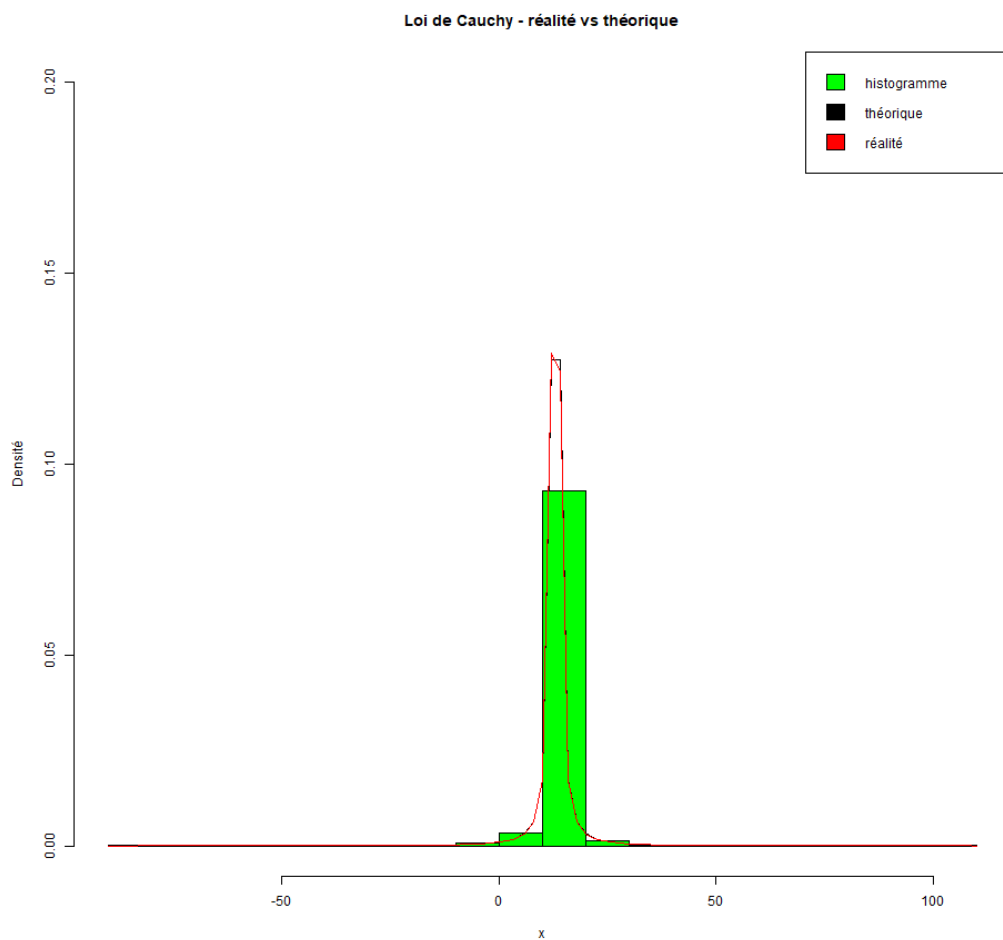


FIGURE 2 – Représentation de l’histogramme et de la courbe de densité

Les valeurs trouvées par R, $\hat{x}_0 = 12.9877139$ et $\hat{a} = 0.4951073$, sont vraiment approchées à celles théoriques.

Références

[1] Wikipedia,

https://en.wikipedia.org/wiki/Cauchy_distribution