Formation Nancy Février 2025

Statistique Bayésienne

Galharret Jean-Michel
ONIRIS VetAgroBio, Nantes

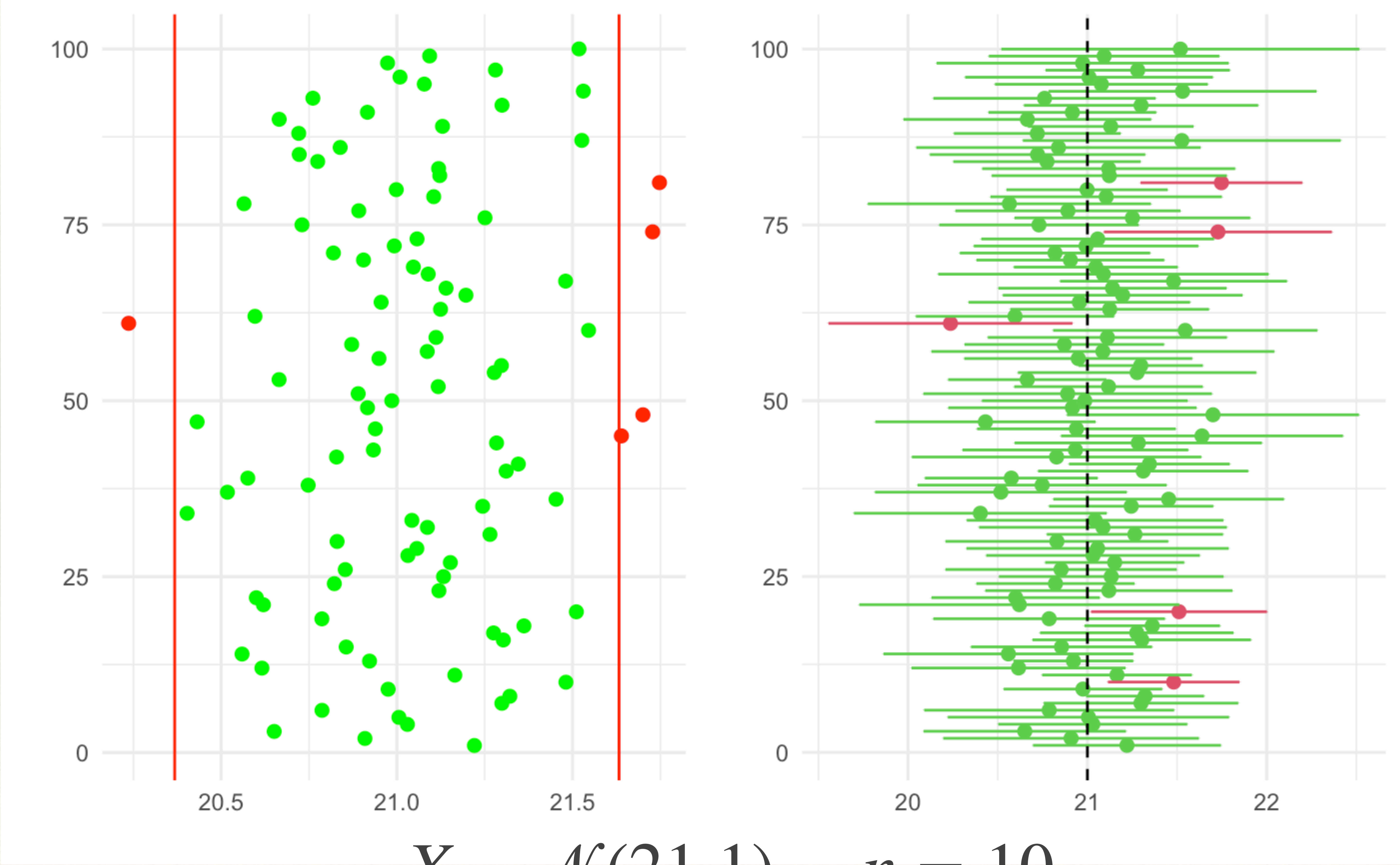
Introduction

- Les tests statistiques (vus hier) font parti de l'approche dite fréquentiste.
- Cette approche a dominé les statistiques durant tout le XX$^{\text{ième}}$ siècle. Elle a été introduite par Fisher, Neyman-Pearson.
- Cette approche a créé pour les utilisateurs de très nombreuses difficultés d'interprétation …. Résultats de tests, P-value, intervalles de confiance…

Intervalles de confiance

Moyennes de 100 échantillons (gauche) et intervalles de confiance de la moyenne de niveau de confiance 95 % (droite).

L'échantillon est **aléatoire.**



$X \sim \mathcal{N}(21{,}1), \quad n = 10$

L’approche bayésienne

- Il s’agit de remettre en cause une croyance que l’on a sur un événement à la vue des données que l’on a observées.
- Dans le cas où les données sont compatibles avec ma croyance, celle-ci est renforcée, dans le cas contraire elle peut être mise en cause.

Exemple pour comprendre

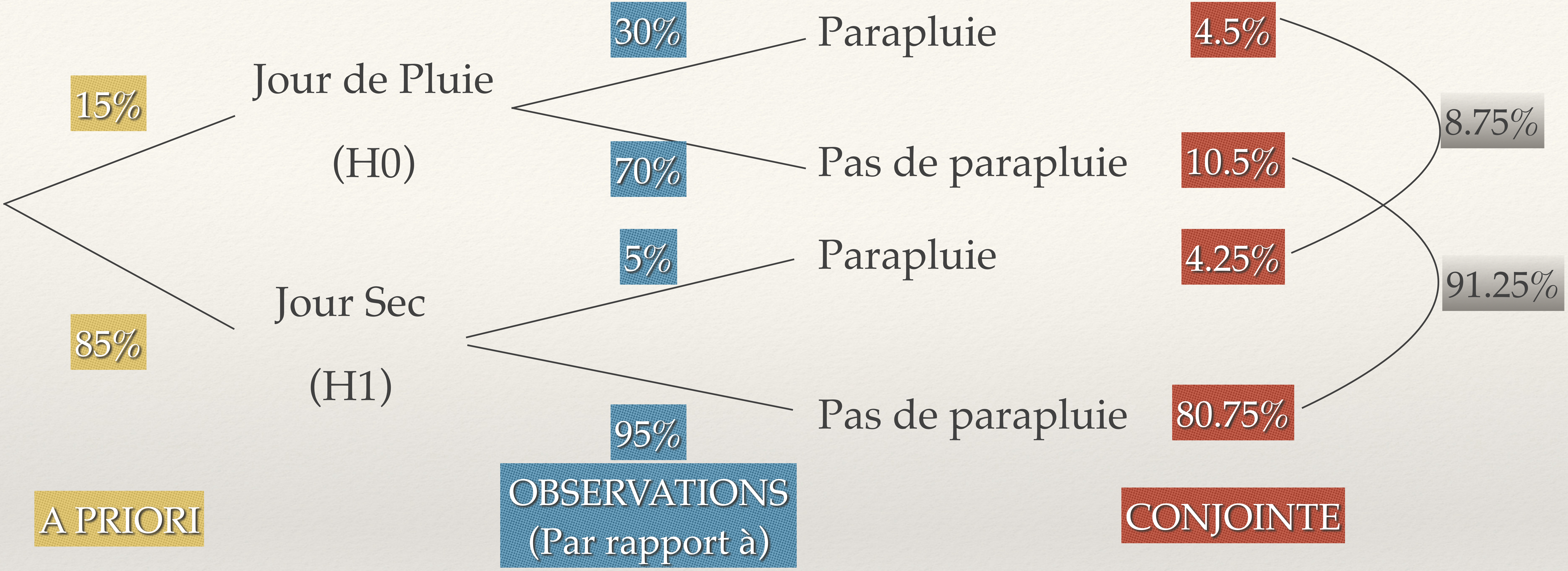
Cet exemple est issu du manuel de JAMOVI :

<u>Je prends un parapluie. Croyez-vous qu'il va pleuvoir ?</u>

On note H0 l'hypothèse il va pleuvoir (H1 l'hypothèse contraire).

- Je regarde l'historique des années antérieures et pour le mois considéré j'ai trouvé que $P(H_0) = 0.15$.

- J'ai fait aussi un recensement (le mois dernier) : j'ai pris mon parapluie alors qu'il a plu dans 30% des cas , je ne l'ai pas pris alors qu'il a plu dans 5% des cas (cf page suivante).



D'un point de vue mathématique, par exemple on a

$$P(Obs \mid H_0) = 0.3 \text{ et } P(Obs \mid H_1) = 0.05$$

$$P(H_0, Obs) = P(Obs \mid H_0) \times P(H_0) = 0.15 \times 0.3 = 0.045$$

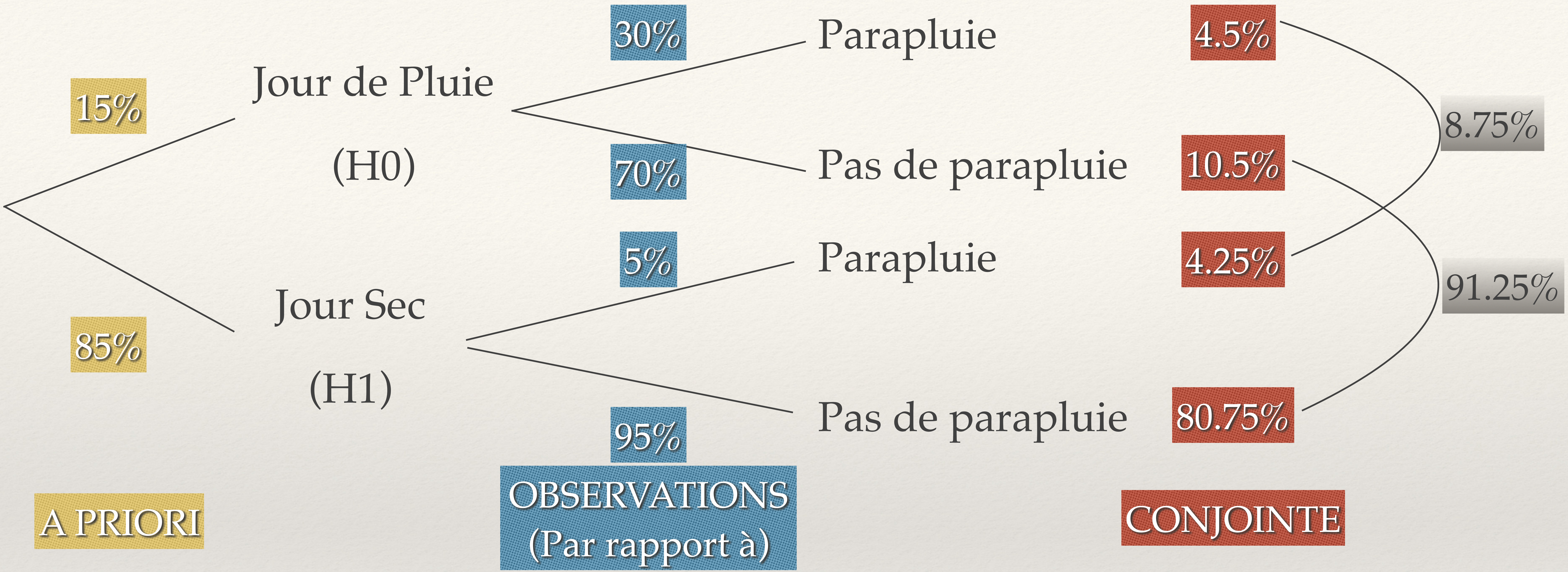
Ce qui nous intéresse (mise à jour de la croyance) :

$$P(H_0 \mid Obs)$$

C'est ici que va intervenir la règle de Bayes :

$$P(H_0 \mid Obs) = \frac{P(H_0, Obs)}{P(Obs)} = \frac{P(Obs \mid H_0) \times P(H_0)}{P(Obs)}$$

$$\text{POSTERIOR} = \frac{\text{LIKELIHOOD} \times \text{A PRIORI}}{\text{MARGINAL}}$$

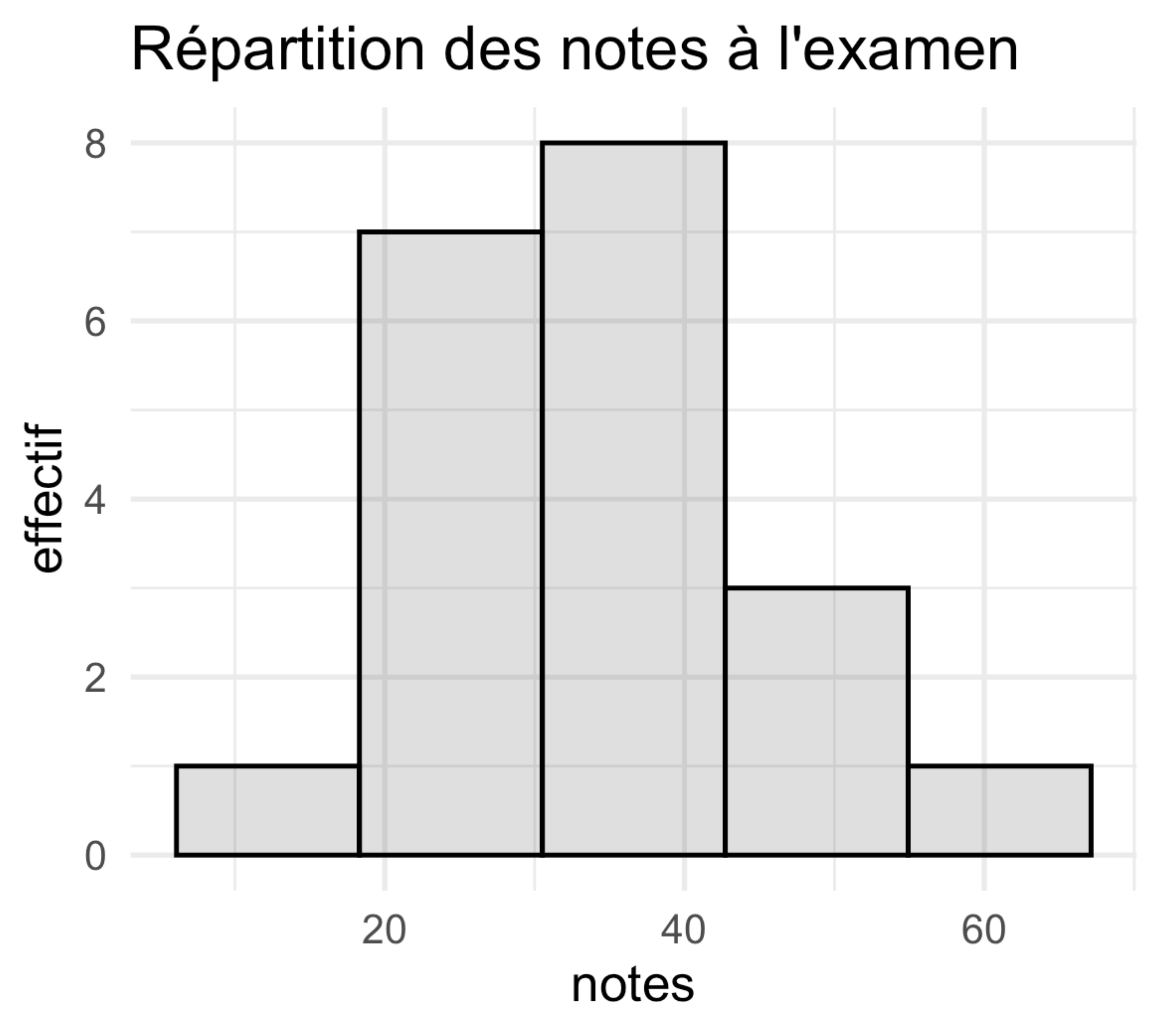


<u>Je prends un parapluie. Croyez-vous qu'il va pleuvoir ?</u>

$$P(H_0 | Obs) = \frac{4.5}{8.75} = 51.4\,\%$$

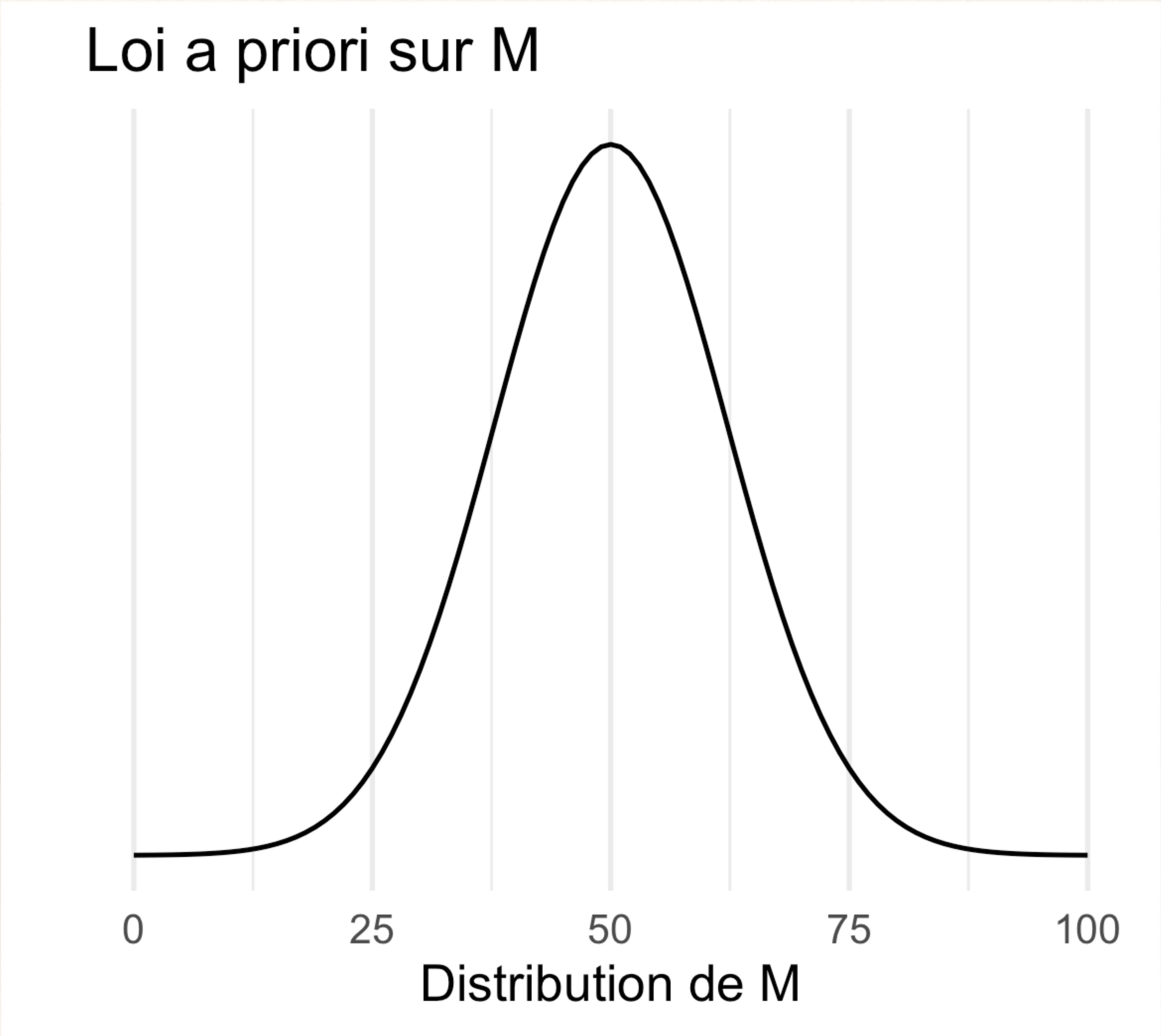
Autre exemple

- On prépare un QCM/100 pour avoir une moyenne de 50 avec 90% des étudiants entre 30 et 70.
- On l'a testé sur 20 étudiants .
- Les résultats obtenus sont donnés ci-contre :

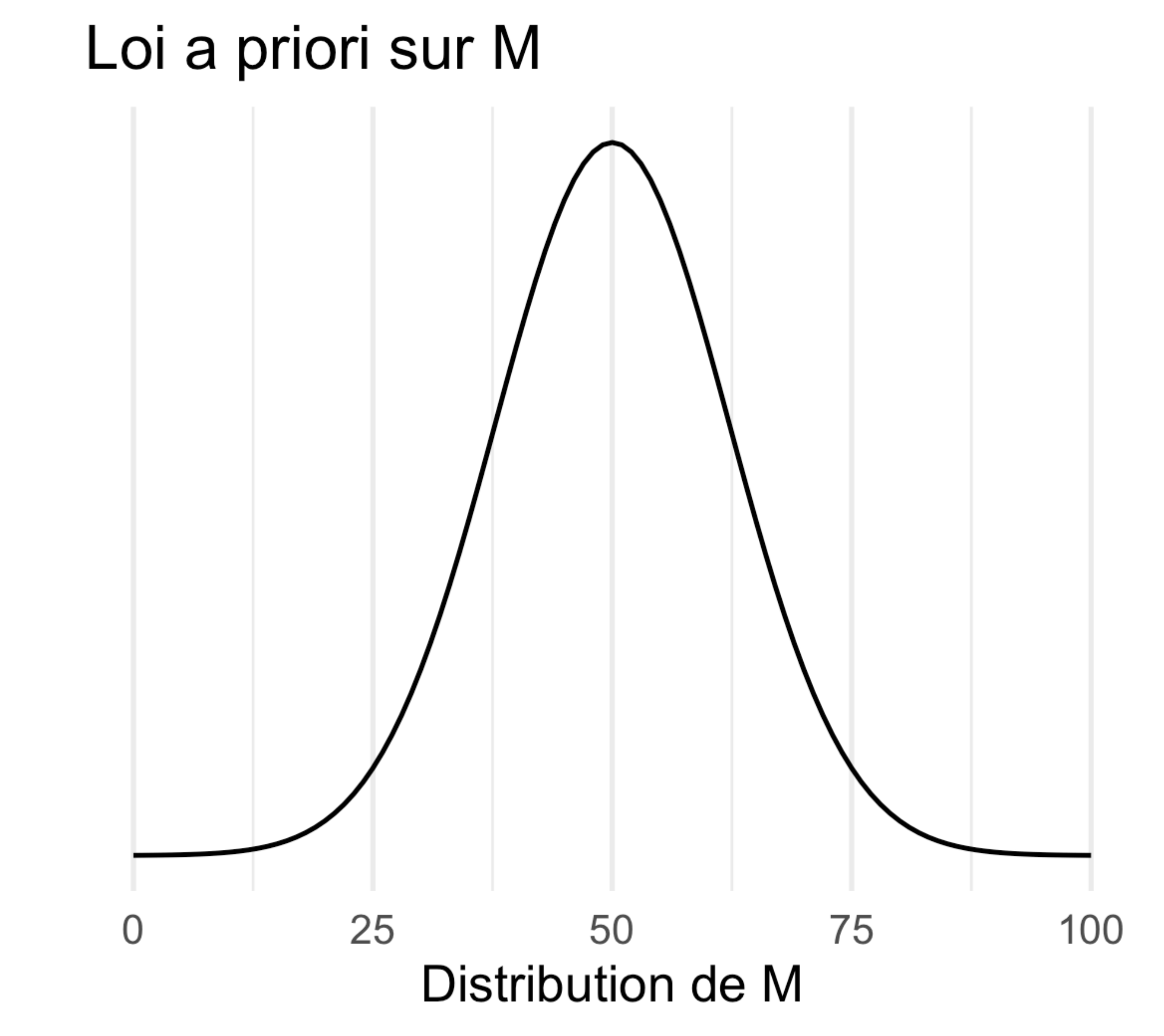


Traduction de l' a priori

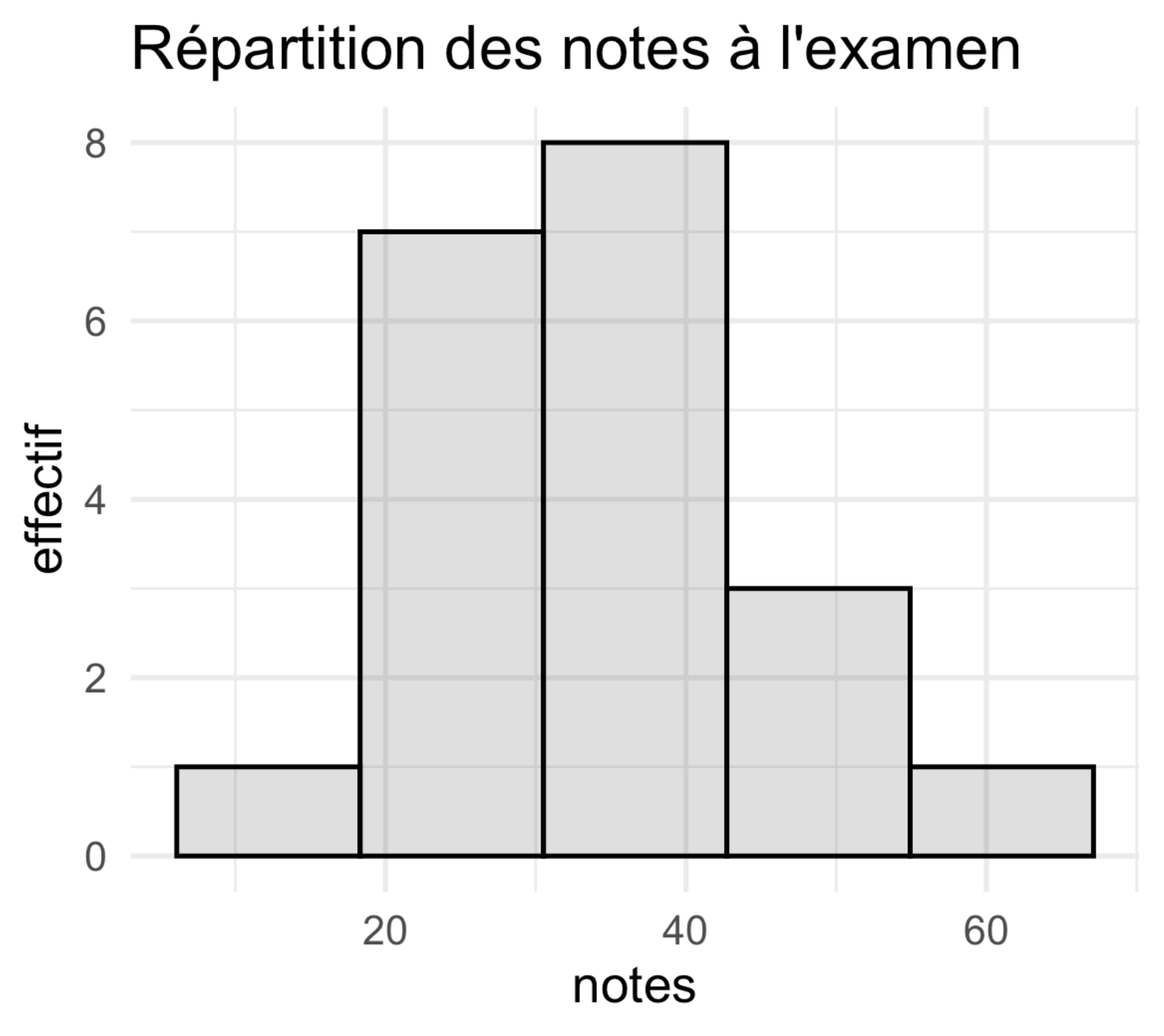
- On appelle M la moyenne recherchée. Compte tenu du contexte on peut prendre 50 avec 90% des étudiants entre 30 et 70. On va prendre une loi normale de moyenne 50 et un écart type pour coller au 90 % c'est à dire $\frac{20}{1.65} = 12.2.$
- En modélisation bayésienne le paramètre M est considéré comme **aléatoire** c'est à dire a une loi de probabilité (ici on prend une loi normale



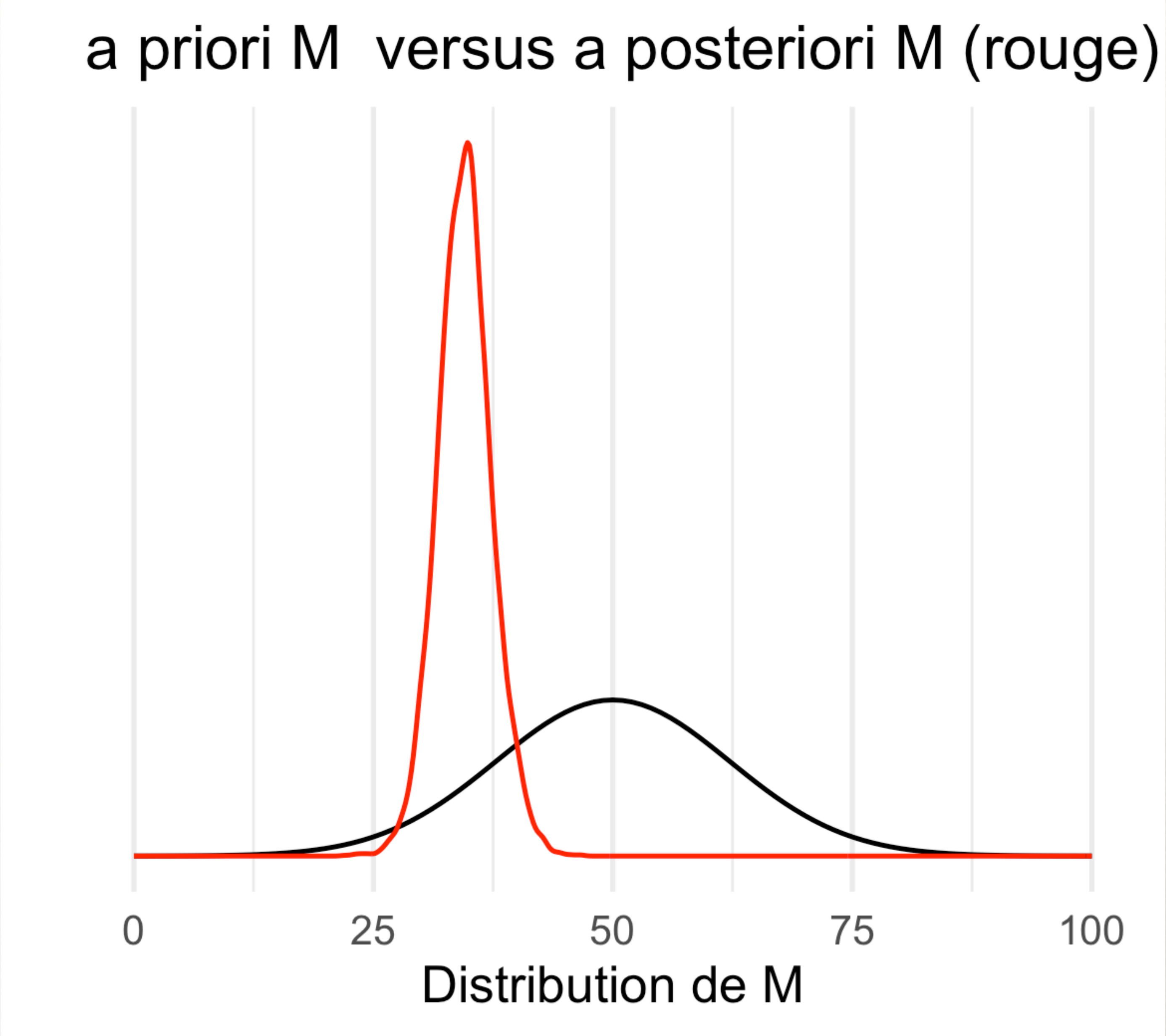
Loi a posteriori (mise à jour)



+



=



Simulation de la loi a posteriori

- Les lois a posteriori sont simulées selon la méthode MCMC (Monte Carlo Markov Chain).
- Le logiciel Stan dans R permet de simuler ces lois :

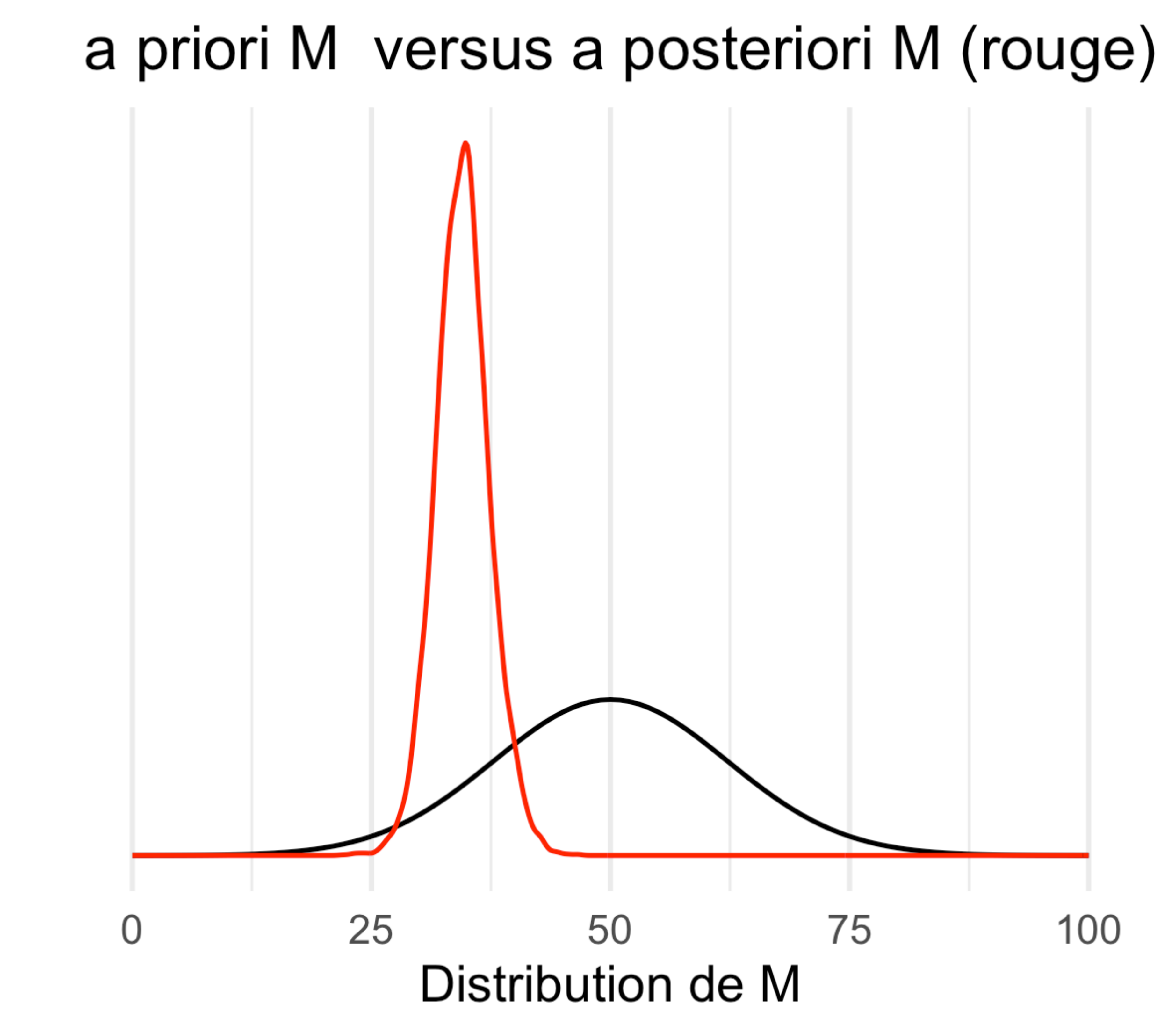
```
model1<-'
data {
int<lower=1> n;
vector[n] X;
}

parameters {
real<lower=0,upper=100> M;
real<lower=0> sig;
}

model{
X~normal(M,sig);
M~normal(50,12.2);
}
'
```

Que peut-on en déduire

- On obtient une distribution de M mise à jour par rapport aux données observées.
- Le QCM n'est pas bien calibré !
- Si on veut une estimation de M en général on prend la moyenne de la loi a posteriori de M ici c'est 34.5. En fait on trouve la moyenne des observations (qui est l'estimateur de M par le maximum de vraisemblance !)



Tout ça pour ça ?

- Finalement on trouve une estimation de M qui est égale à l'estimation de M par la méthode fréquentiste habituelle !

- Pourtant le gain de cette approche est **très grand**. On va le voir à travers l'exemple de l'intervalle de crédibilité / intervalle de confiance de la moyenne.

- L'intervalle de crédibilité à 95% est défini par $P(M \in \mathcal{I} \mid Obs) = .95$ il se déduit donc directement de la loi a posteriori de M ici on a $\mathcal{I} = [29.3,40.5]$.

- Ici l'intervalle de confiance au sens fréquentiste est $IC = [30,39.1]$ (donc on ne trouve pas le même résultat, mais c'est surtout l'interprétation qui est totalement différente

Différence (philosophique)entre les deux approches

<u>Approche fréquentiste :</u>

- Paramètre fixe.
- Intervalle de confiance aléatoire
- Dissymétrie entre l'hypothèse nulle et son alternative

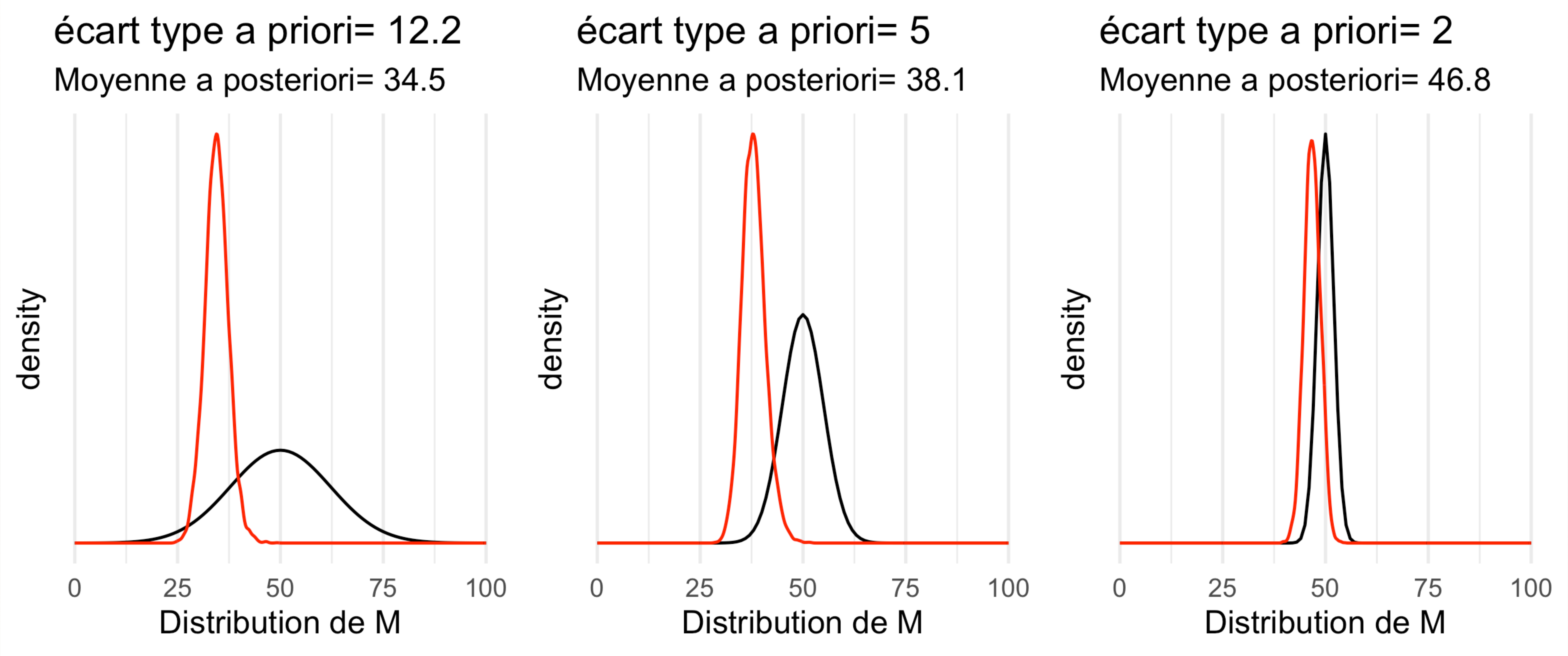
Nécessite des répétitions

<u>Approche bayésienne :</u>

- Paramètre aléatoire.
- Intervalle de crédibilité fixe.
- Symétrie entre les hypothèses

Facile à interpréter

Attention aux a priori



Tests Bayésiens

Introduction

- Du point de vue Bayésien prendre une décision entre deux hypothèses H_1, H_2 revient à comparer deux modèles M_1, M_2.
- On veut comparer $P(M_1 | Obs)$ et $P(M_2 | Obs)$. Pour ce faire on calcule le ratio

$$\underbrace{\frac{P(M_1 | D)}{P(M_2 | D)}}_{\text{Posterior odds}} = \underbrace{\frac{P(D | M_1)}{P(D | M_2)}}_{\text{Bayes Factor}} \times \underbrace{\frac{P(M_1)}{P(M_2)}}_{\text{Prior odds}}$$

En général on ne privilégie pas une hypothèse par rapport à l'autre donc $P(M_1) = P(M_2)$. Le facteur de Bayes de H1 contre H2 est noté BF_{12}

Des points délicats

- Le facteur de Bayes est influencé par les choix de loi a priori sur les paramètres du modèle.
- Le calcul du facteur de Bayes requiert dans de nombreuses applications une grande puissance de calcul ce qui a été longtemps un problème.
- Il faut de nouveau définir des valeurs seuils qui comme la p-value vont être arbitraires

Table 7.1 Evidence categories for the Bayes factor BF_{12} (Jeffreys, 1961).

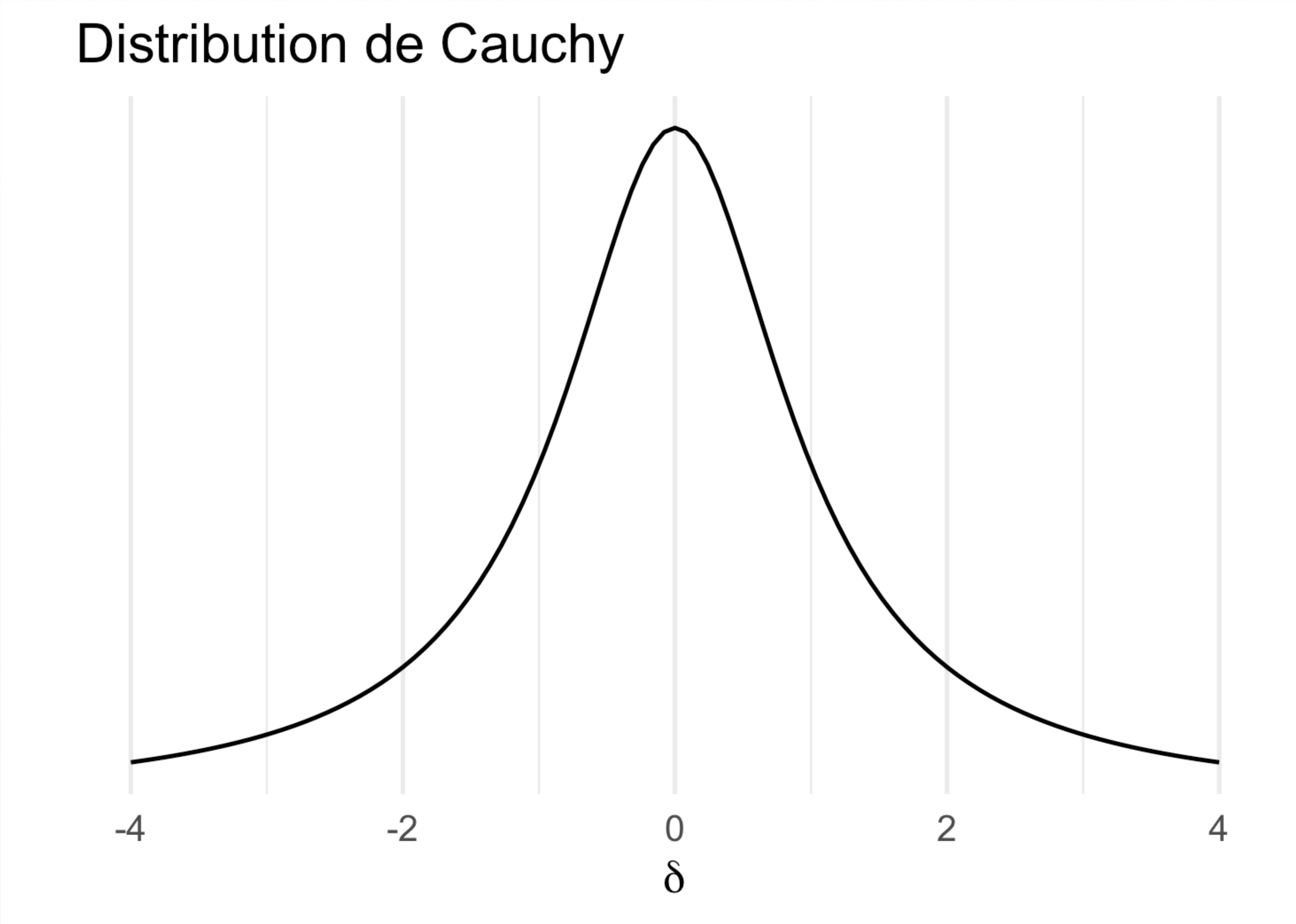
Bayes factor BF_{12}			Interpretation
	$>$	100	Extreme evidence for $\mathcal{M}_1$
30	–	100	Very strong evidence for $\mathcal{M}_1$
10	–	30	Strong evidence for $\mathcal{M}_1$
3	–	10	Moderate evidence for $\mathcal{M}_1$
1	–	3	Anecdotal evidence for $\mathcal{M}_1$
	1		No evidence
1/3	–	1	Anecdotal evidence for $\mathcal{M}_2$
1/10	–	1/3	Moderate evidence for $\mathcal{M}_2$
1/30	–	1/10	Strong evidence for $\mathcal{M}_2$
1/100	–	1/30	Very strong evidence for $\mathcal{M}_2$
	$<$	1/100	Extreme evidence for $\mathcal{M}_2$

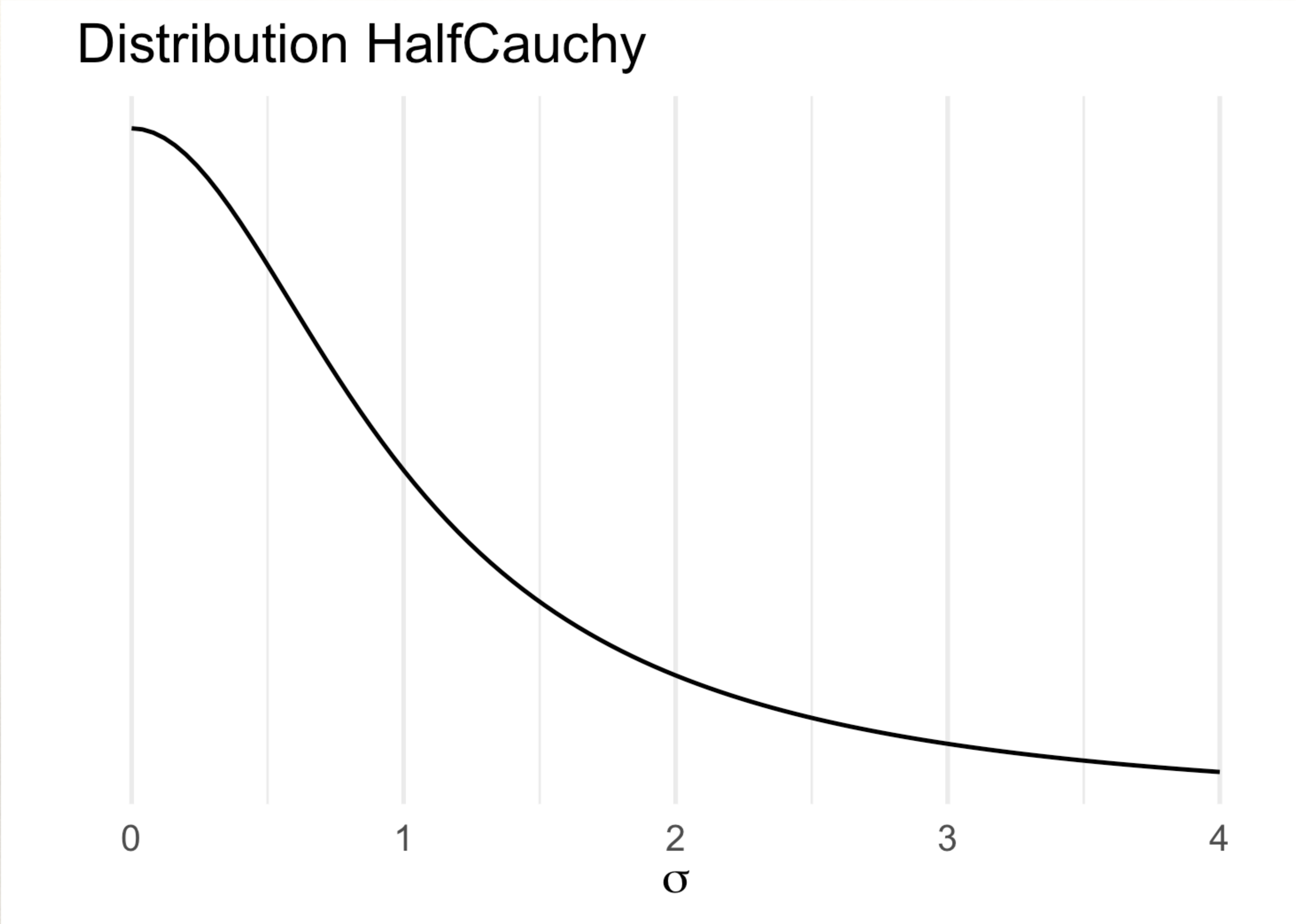
Retour aux hypothèses classiques de tests

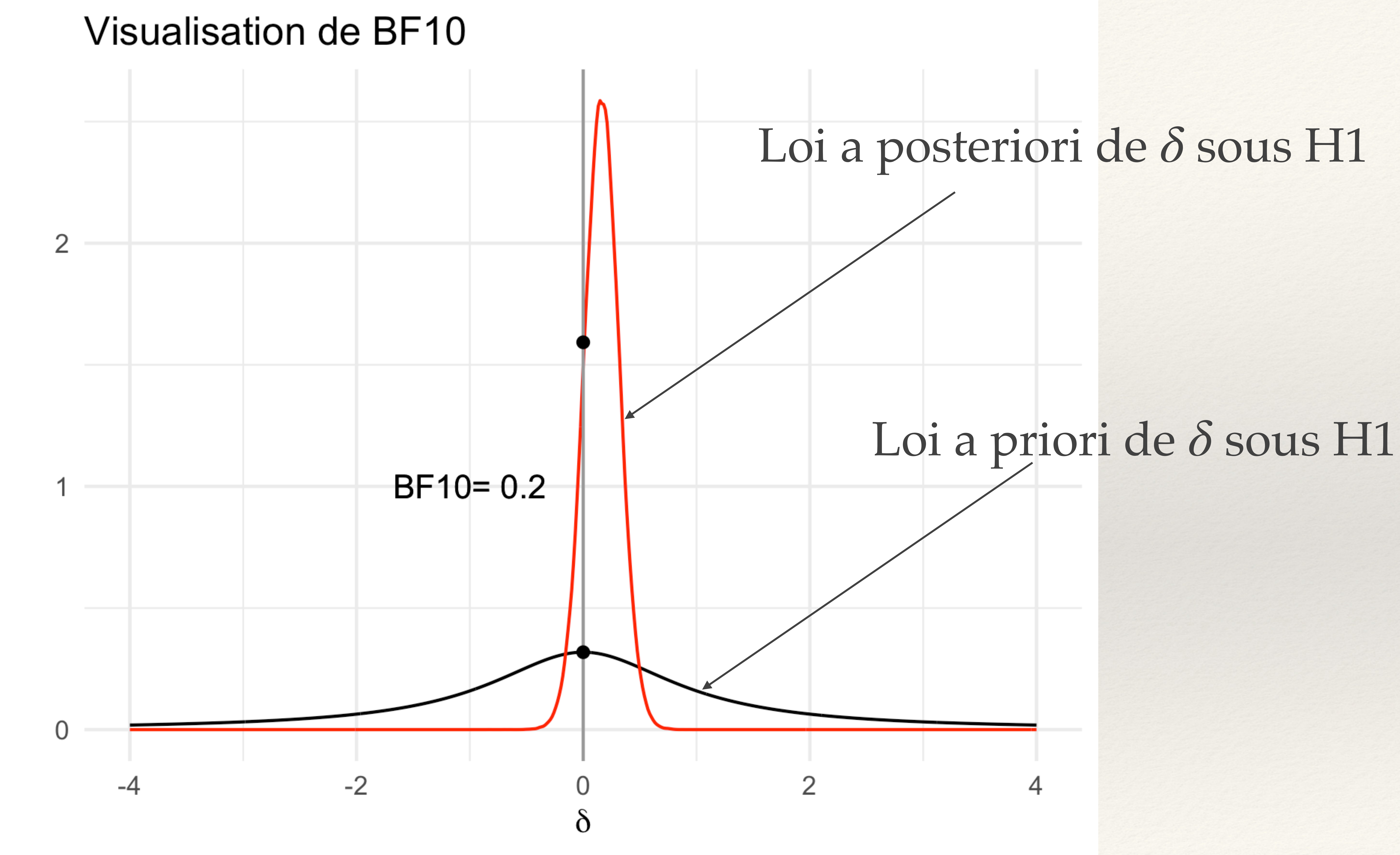
- On veut tester $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$
- Savage-Dickey formula : $BF_{10} = \dfrac{P(\theta = \theta_0 \mid H_1)}{P(\theta = \theta_0 \mid Obs, H_1)}$.

Test de student univarié : Exemple $\mu = 0$

- On considère le modèle $X \sim \mathcal{N}(\mu, \sigma)$ de paramètres inconnus.
- On veut tester $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$.
- Pour ce faire on va tester $H_0 : \delta = 0$ versus $H_1 : \delta \neq 0$ où $\delta = \dfrac{\mu}{\sigma}$. Avantage : on a plus de problème d'unité, valable dans de très nombreux contextes.
- Pour H1 on choisit $\delta \sim \mathscr{C}(0,1)$.

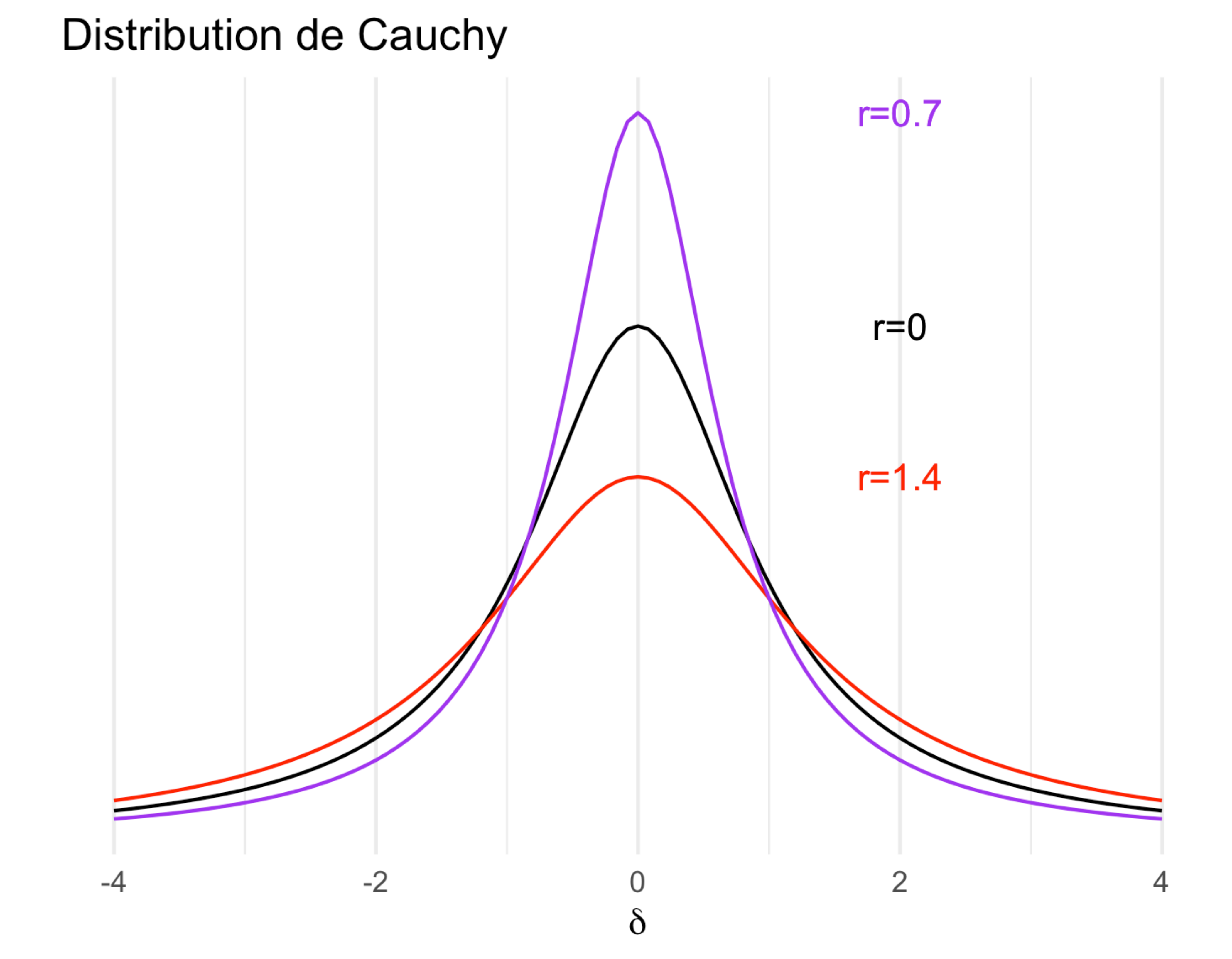

Distribution de Cauchy
-4
-2
0
2
4
δ


Distribution HalfCauchy
0
1
2
3
4
σ


Visualisation de BF10
Loi a posteriori de δ sous H1
Loi a priori de δ sous H1
BF10= 0.2
2
1
0
-4
-2
0
2
4
δ

Facteur d'échelle dans la loi a priori

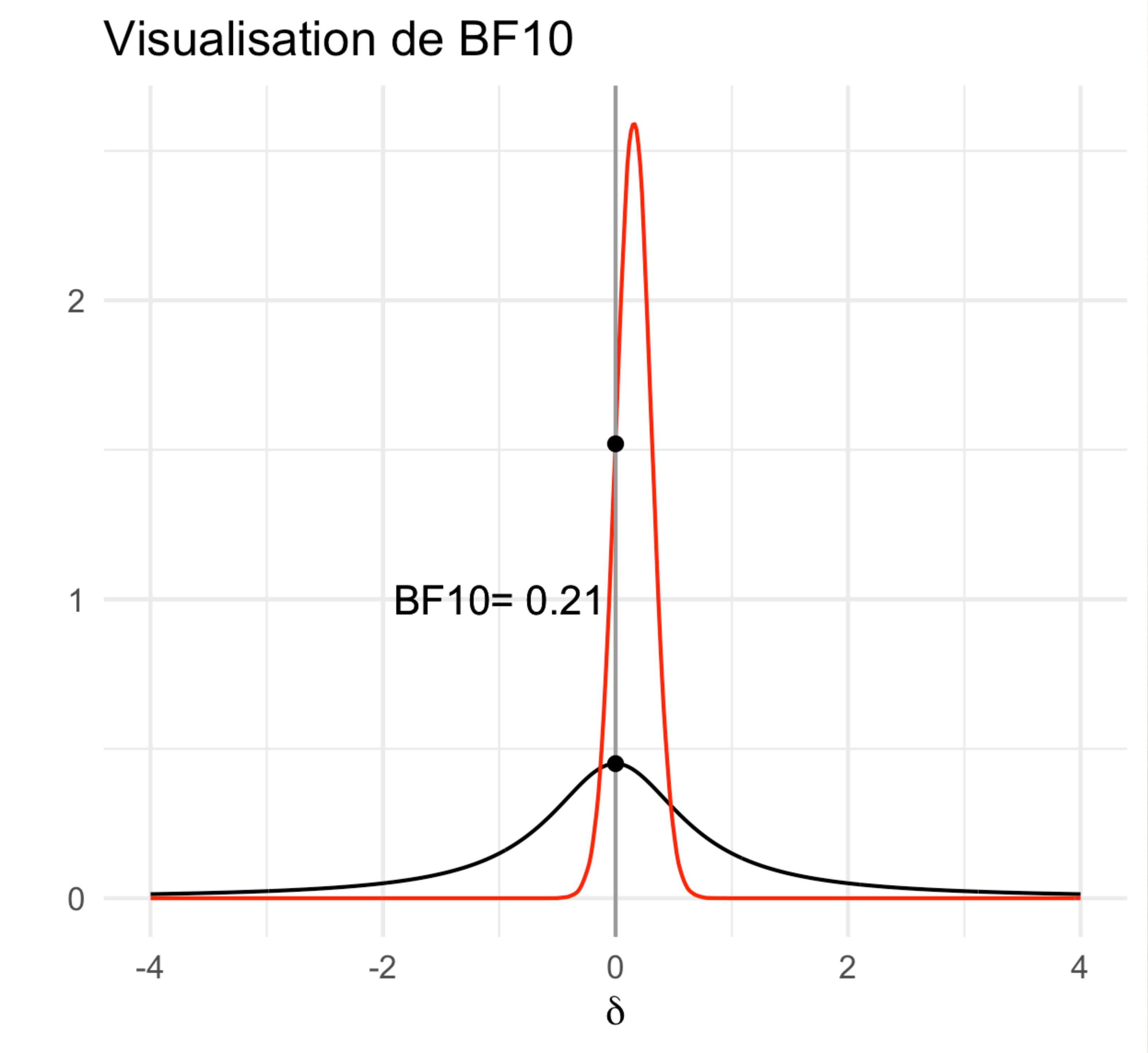
- On regarde fréquemment la variation du facteur de Bayes par rapport à la loi a priori qui a été choisie pour H1.
- Pour ce faire on considère $H_1 : \delta \sim \mathscr{C}(0,r)$ avec $r = \sqrt{2}$ (medium), $r = 1$ (wide) et $r = \dfrac{\sqrt{2}}{2}$ (ultrawide)

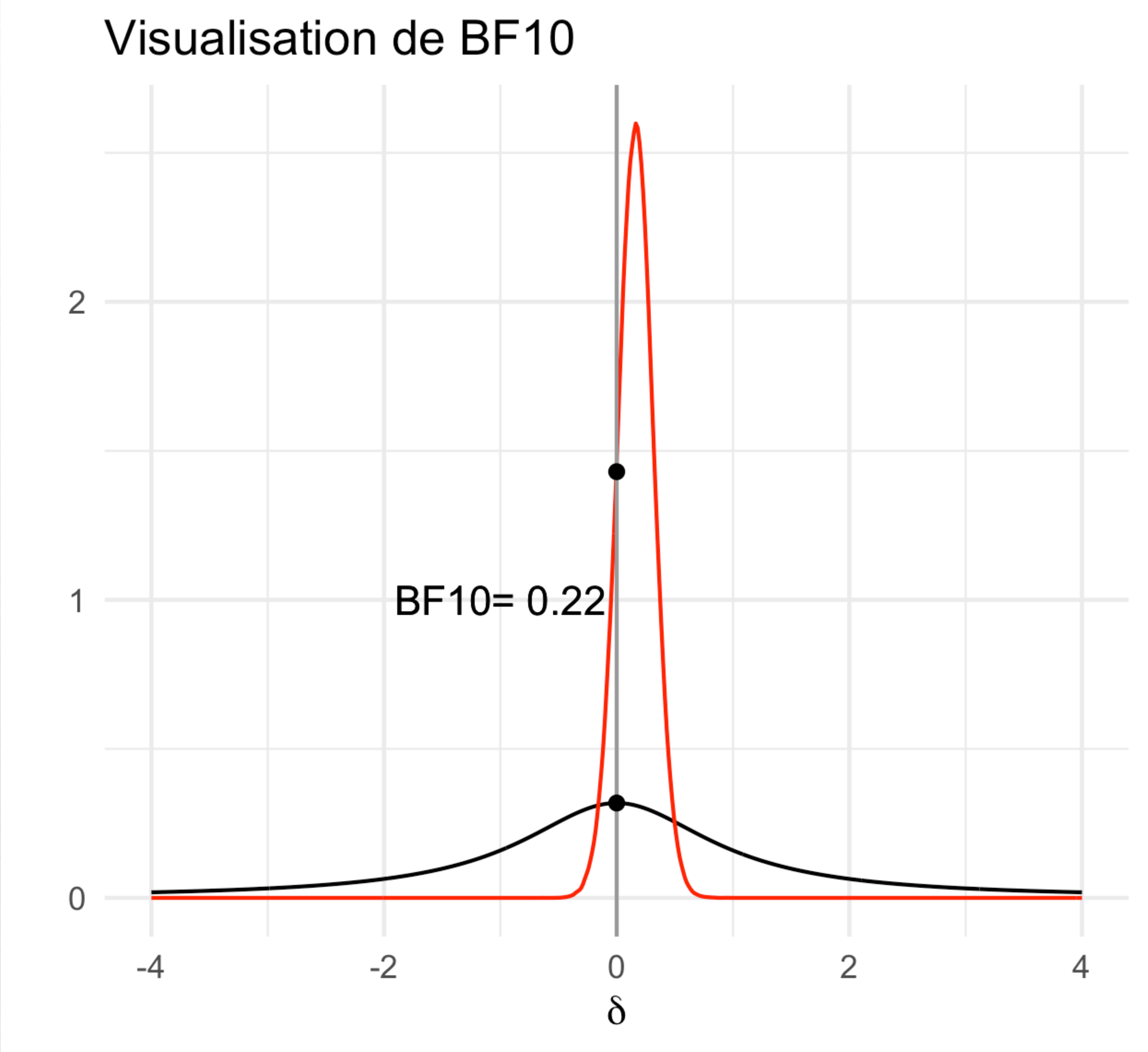


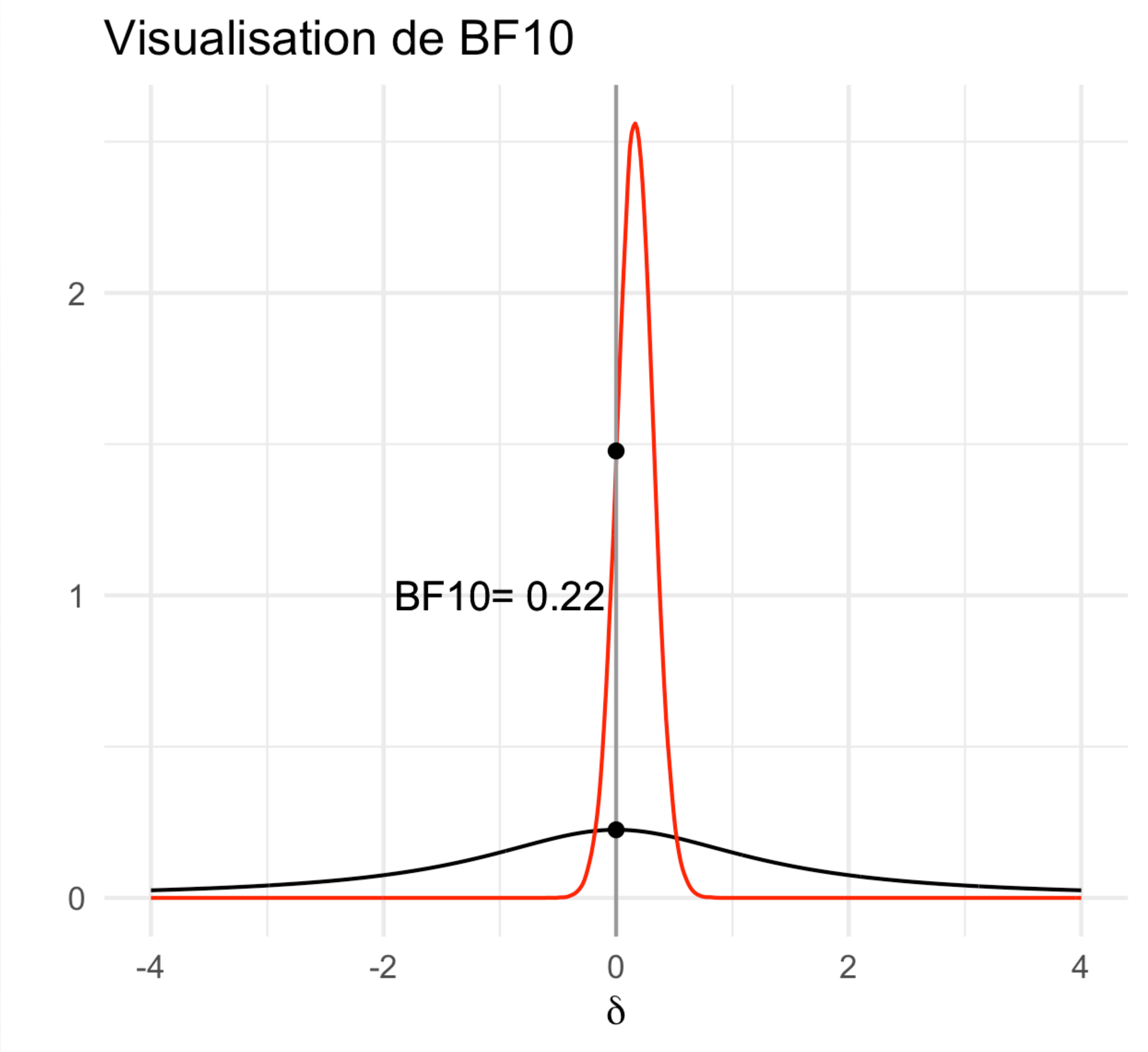
$$r = \frac{\sqrt{2}}{2}$$

$$r = 1$$

$$r = \sqrt{2}$$



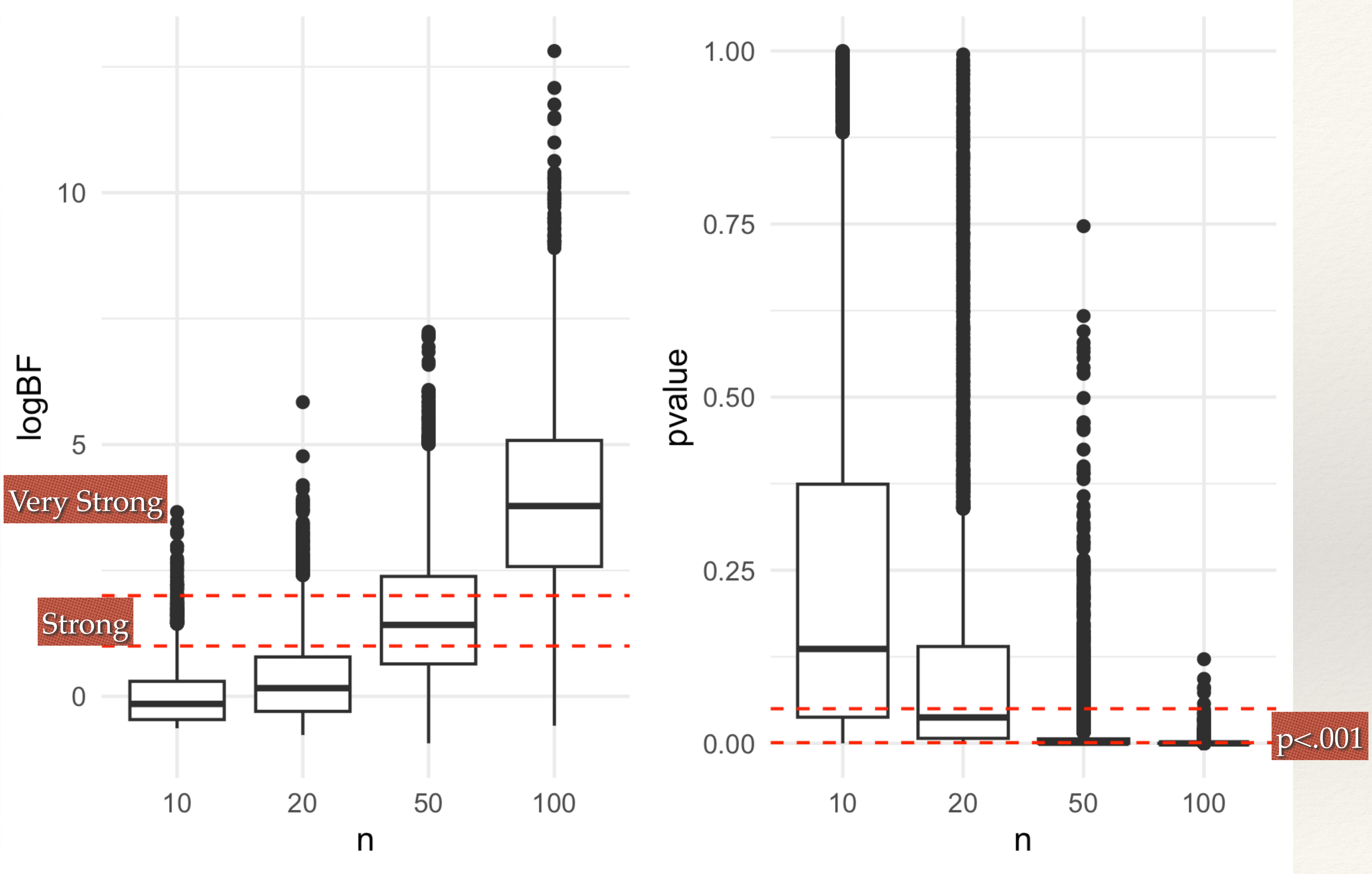




On peut ainsi optimiser la valeur du BF en fonction de r

(Voir TP)

Facteur de Bayes versus p-value



Anova et Modèles linéaires

- On peut réaliser des tests Bayésiens sur tous les modèles linéaires vus précédemment, c'est à dire les modèles de type :
  - Régression linéaire,
  - ANOVA
  - ANCOVA
  - Régression linéaire à effets aléatoires.

Exemple de régression linéaire

30 employés ont répondus à une enquête : data(attitude)

7 variables : on veut expliquer la note globale

(Rating) par le traitement des plaintes (complainte), Learning (opportunité d'apprendre), …

```
Bayes factor analysis
--------------
[1] complaints                        : 417938.6 ±0.01%
[2] complaints + learning             : 207271.9 ±0%
[3] complaints + learning + advance : 88041.54 ±0%
[4] complaints + raises               : 77498.99 ±0%
[5] complaints + privileges           : 75015.23 ±0%
[6] complaints + advance              : 72759.76 ±0%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

BF décroissants

On teste H0 : tous les β_j du modèle considéré sont nuls contre au moins l'un est non nul.

Autres possibilités

```
Bayes factor analysis
--------------
[1] complaints                          : 417938.6 ±0.01%
[2] complaints + learning               : 207271.9 ±0%
[3] complaints + learning + advance     : 88041.54 ±0%
[4] complaints + raises                 : 77498.99 ±0%
[5] complaints + privileges             : 75015.23 ±0%
[6] complaints + advance                : 72759.76 ±0%

Against denominator:
  Intercept only
---
Bayes factor type: BFlinearModel, JZS
```

```
Bayes factor analysis
--------------
[1] complaints                      : 1         ±0%
[2] complaints + learning           : 0.4959386 ±0.01%
[3] complaints + learning + advance : 0.2106566 ±0.01%
[4] complaints + raises             : 0.1854315 ±0.01%
[5] complaints + privileges         : 0.1794886 ±0.01%
[6] complaints + advance            : 0.174092  ±0.01%

Against denominator:
  rating ~ complaints
```

$$BF_{C,C} = 1; \quad BF_{C+L,C} = \frac{207271.9}{417938.6} = 0.4959$$

En effet $BF_{C+L,C} = \dfrac{BF_{C+L,M_0}}{BF_{C,M_0}}$ où M_0 est le modèle avec uniquement l'intercept.

Résumé du modèle linéaire

	Parameter	Median	CI	CI_low	CI_high	pd	ROPE_CI	ROPE_low	ROPE_high	ROPE_Percentage	Rhat	ESS
1	(Intercept)	11.03	0.95	-13.49	34.40	0.82	0.95	-1.22	1.22	0.06	1	4547.21
3	complaints	0.61	0.95	0.28	0.94	1.00	0.95	-1.22	1.22	1.00	1	2754.09
6	privileges	-0.07	0.95	-0.35	0.21	0.69	0.95	-1.22	1.22	1.00	1	3788.22
5	learning	0.31	0.95	-0.04	0.66	0.96	0.95	-1.22	1.22	1.00	1	3322.57
7	raises	0.08	0.95	-0.36	0.55	0.64	0.95	-1.22	1.22	1.00	1	3101.60
4	critical	0.04	0.95	-0.27	0.33	0.60	0.95	-1.22	1.22	1.00	1	4329.12
2	advance	-0.22	0.95	-0.57	0.17	0.87	0.95	-1.22	1.22	1.00	1	2799.89

Intervalles de crédibilité

pd: une valeur élevée signifie que l'effet associé est concentré du même côté que la médiane.

ROPE: Region of Practical Equivalence, au lieu de considérer $\delta = 0$ on considère un intervalle du type $[\delta_m, \delta_M]$ (-0.1,0.1 pour les coefficients standardisés (cf Cohen 1988)

Distribution a posteriori du coefficient de complaints