# Rapport TP PMS

YEBKA Massinissa

RANDRIAMORA Yevann

ROUSSEL Guillaume

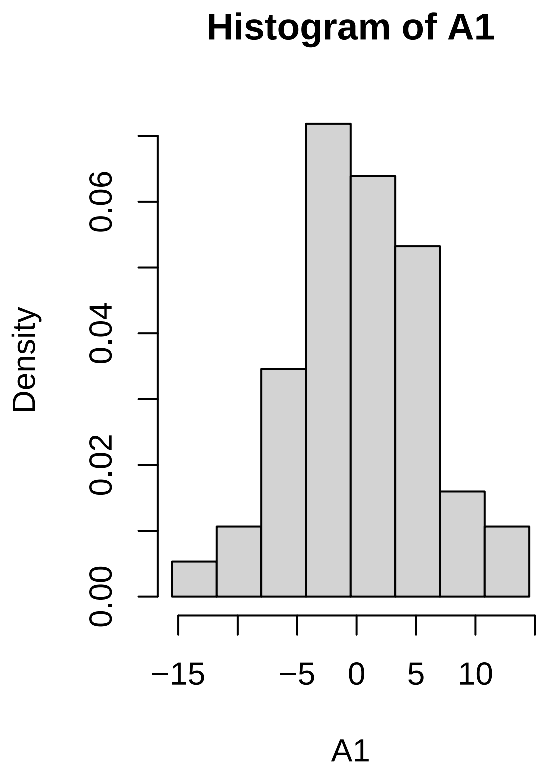
Affiliés à : ENSIMAG

Le : 16/04/2021

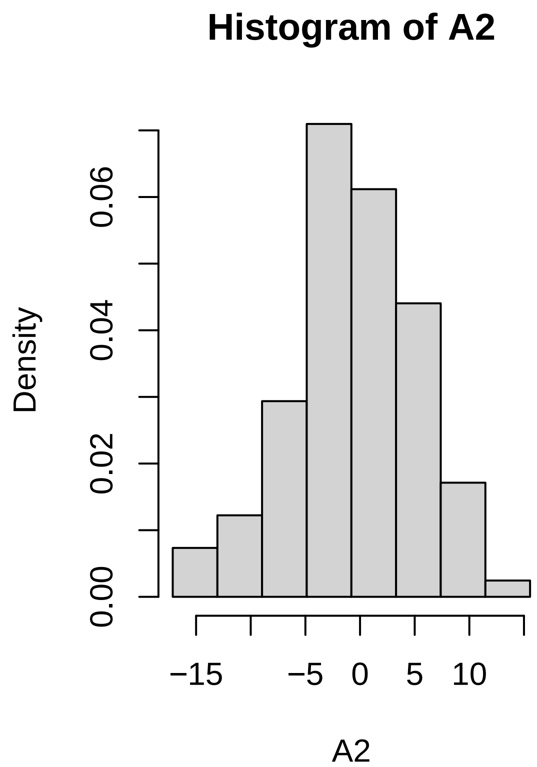
Principes et Méthodes Statistiques

## Analyse de la vitesse

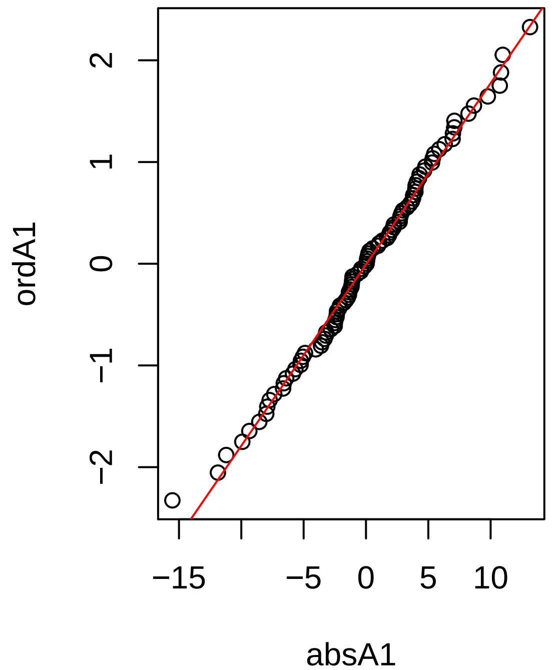
**Question 1**



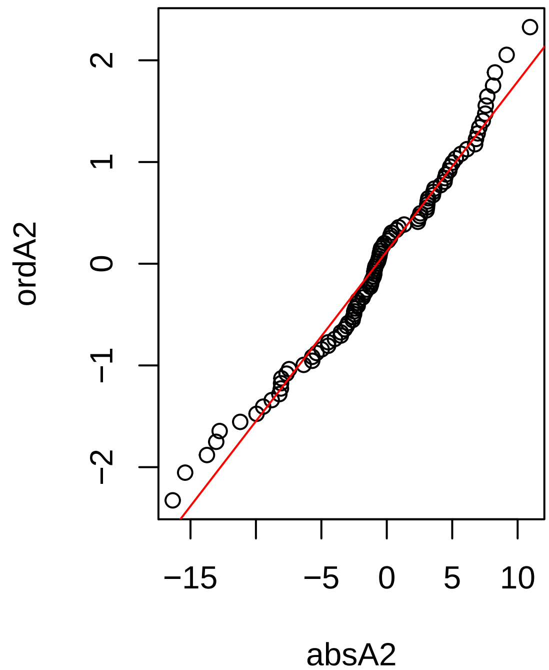
Histogram of A1
Density
0.00
0.02
0.04
0.06
−15
−5
0
5
10
A1

Histogram of A2
Density
0.00
0.02
0.04
0.06
−15
−5
0
5
10
A2

ordA1
−2
−1
0
1
2
−15
−5
0
5
10
absA1

ordA2
−2
−1
0
1
2
−15
−5
0
5
10
absA2

```
##    absA1
## 31.48487
```

On constate que les points des graphes de probabilités 1 et 2 suivent bien l'allure d'une droite et donc qu'il est raisonnable d'admettre que $A_1$ et $A_2$ sont de loi $\mathcal{N}(0, \sigma^2)$.

### Question 2

On suppose que $A_1$ et $A_2$ suivent une loi normale centrée et variance $\sigma^2$.

Ainsi $A_1 \hookrightarrow \mathcal{N}(0, \sigma^2)$ et de même pour $A_2$. Alors d'après l'annexe du cours

$$\frac{A_1}{\sigma} \hookrightarrow \mathcal{N}(0,1) \Longrightarrow \frac{A_1^2}{\sigma^2} \hookrightarrow \mathcal{G}(\frac{1}{2}, \frac{1}{2})$$

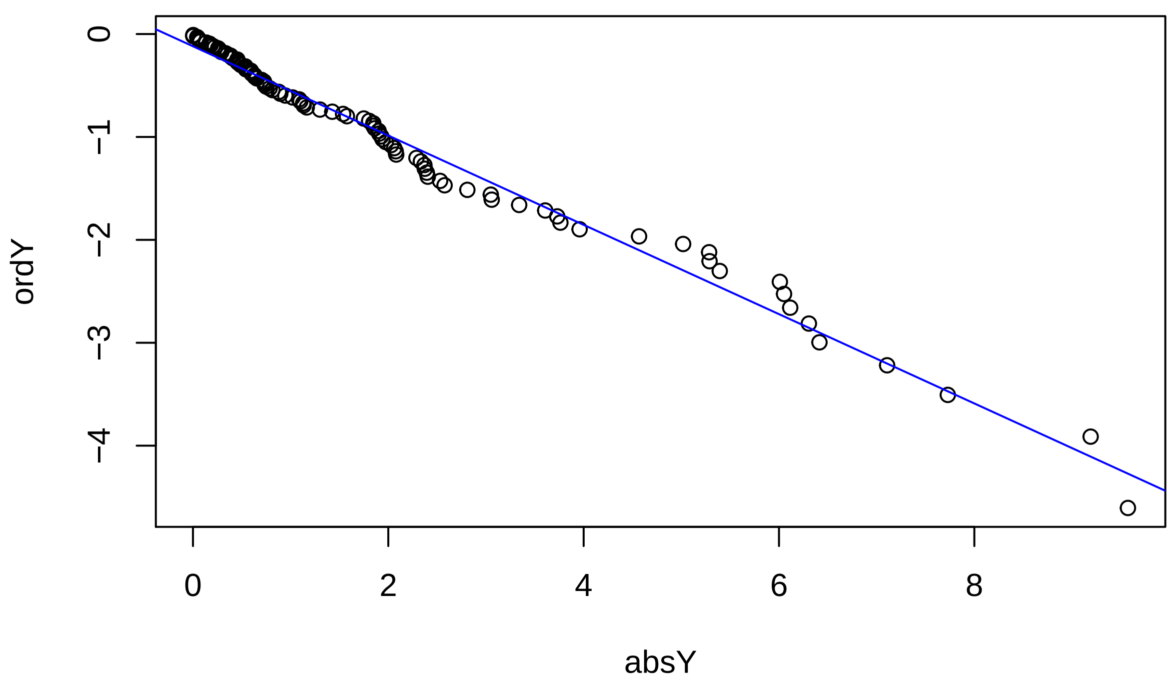
Et

$$\frac{A_2}{\sigma} \hookrightarrow \mathcal{N}(0,1) \Longrightarrow \frac{A_2^2}{\sigma^2} \hookrightarrow \mathcal{G}(\frac{1}{2}, \frac{1}{2})$$

Donc d'après une propriété du cours de la loi gamma

$$\boxed{\frac{A_1^2 + A_2^2}{\sigma^2} \hookrightarrow \mathcal{G}(1, \frac{1}{2})} \tag{1}$$

Et donc d'après le cours, $\frac{X^2}{\sigma^2}$ suit une loi exponentielle de paramètre $\frac{1}{2}$. On vérifie cette hypothèse par une régres-




sion linéaire.

```
##      absY
## 0.4338142
```

## Question 3

On a $\forall t \in \mathbb{R}^+$

$\mathbb{P}(\frac{X^2}{\sigma^2} \leq t) = \mathbb{P}(X^2 \leq \sigma^2 t) = \mathbb{P}(X \leq \sigma\sqrt{t}) = 1 - e^{-\frac{t}{2}}$.

On a donc $F_X(\sigma\sqrt{t}) = 1 - e^{-\frac{t}{2}}$ et donc par un changement de variable, $F_X(u) = 1 - e^{-\frac{u^2}{2\sigma^2}}$.

Ainsi on a $\forall x \in \mathbb{R}^+, f_X(x) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$. Or $\forall t < 0, \mathbb{P}(\frac{X^2}{\sigma^2} \leq t) = 0$ donc finalement on a

$$\boxed{\forall x \in \mathbb{R}, f_X(x) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}\chi_{\mathbb{R}^+(x)}} \tag{2}$$

$X$ suit bien une loi de Rayleigh de paramètre $\sigma^2$.

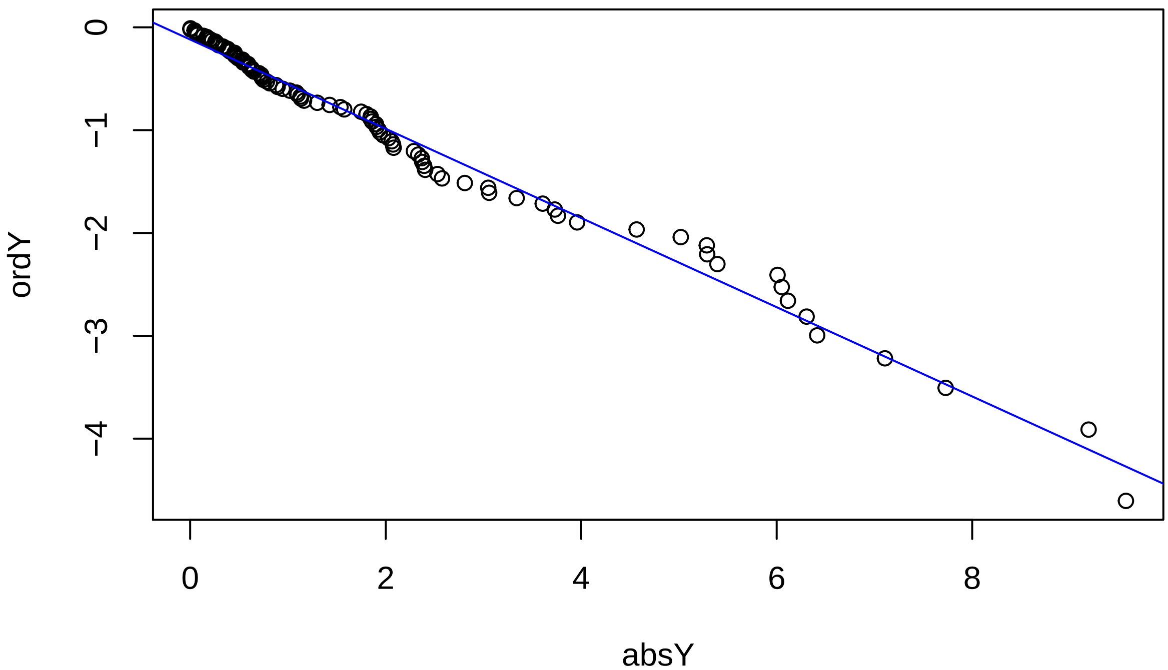
## Question 4

On a

$$F_X(x) = 1 - e^{-\frac{x^2}{2\sigma^2}}$$
$$\Downarrow$$
$$ln(1 - F_X(x)) = -\frac{1}{2\sigma^2}x^2$$

Donc le graphe de probabilités pour la loi de Rayleigh est $((x_i^*)^2, ln(1 - \frac{i}{n}))_{1 \leq i \leq n-1}$.

```
## sigma^2 =  31.48487
```




### Question 5

On a $E[X] = \int_{\mathbb{R}^+} x \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx$

On fait une intégration par partie avec

$$\begin{cases} u = -e^{-\frac{x^2}{2\sigma^2}} \\ u' = \frac{x}{2\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \\ v = x \\ v' = 1 \end{cases}$$

ainsi on a

$$E[X] = [-xe^{-\frac{x^2}{2\sigma^2}}]_0^{+\infty} + \int_{\mathbb{R}^+} e^{-\frac{x^2}{2\sigma^2}} = 0 + \sigma\sqrt{\frac{\pi}{2}}$$

D'où

$$\boxed{E[X] = \sigma\sqrt{\frac{\pi}{2}}} \tag{3}$$

Puis on a

$$E[X^2] = \int_{\mathbb{R}^+} x^2 \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx$$

On fait l'intégration par partie suivante

$$\begin{cases} u = -e^{-\frac{x^2}{2\sigma^2}} \\ u' = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \\ v = x^2 \\ v' = 2x \end{cases}$$

D'où

$$E[X^2] = [-x^2 e^{-\frac{x^2}{2\sigma^2}}]_0^{+\infty} + \int_{\mathbb{R}^+} 2xe^{-\frac{x^2}{2\sigma^2}} = 0 + 2[-\sigma^2 e^{-\frac{x^2}{2\sigma^2}}]_0^{+\infty}$$

D'où

$$E[X^2] = 2\sigma^2$$

Or d'après la formule de Huygens

$$Var[X] = E[X^2] - E[X]^2 = 2\sigma^2 - \frac{\sigma^2\pi}{2}$$

D'où

$$\boxed{Var[X] = \frac{4-\pi}{2}\sigma^2}$$

On a $E[X] = \sigma\sqrt{\frac{\pi}{2}}$, d'où $\sigma^2 = \frac{2}{\pi}E[X]^2$.

On pose $\forall i \in \mathbb{N}, i \in [1,n], X_i = \sqrt{A_{1i}^2 + A_{2i}^2}$

d'où par la méthode des moments

$$\boxed{\tilde{\sigma_n}^2 = \frac{2}{\pi n^2}(\sum_1^n X_i^2)} \quad (4)$$

Ainsi

$$E[\tilde{\sigma_n}^2] = \frac{2}{\pi n^2}E[(\sum_{i=1}^n X_i^2)]$$

$$= \frac{2}{\pi n^2}E[\sum X_i^2 + 2\sum_{1\leq i<j\leq n} X_i X_j]$$

$$= \frac{2}{\pi n^2}[nE[X_i^2] + 2E[\sum_{1\leq i<j\leq n} X_i X_j]]$$

Or $\forall i, E[X_i^2] = 2\sigma^2$ et $2E[\sum_{1\leq i<j\leq n} X_i X_j] = (n^2-n)E[X_1 X_2] = (n^2-n)E[X_1]E[X_2]$

D'où

$$E[\tilde{\sigma_n}^2] = \frac{2}{\pi n^2}[2n\sigma^2 + (n^2-n)\frac{\sigma^2\pi}{2}]$$

D'où enfin

$$E[\tilde{\sigma_n}^2] = \frac{\sigma^2}{\pi n}[4 + (n-1)\pi]$$

On pose donc l'estimateur débiaisé

$$\boxed{\tilde{\sigma_n}'^2 = \frac{\pi n}{4+(n-1)\pi}\tilde{\sigma_n}^2} \quad (5)$$

## Question 6

On a

$$f_X(x) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}\chi_{\mathbb{R}_+}(x)$$

D'où

$$\prod_1^n f_X(x_i) = \frac{\prod x_i}{\sigma^{2n}}e^{-\frac{1}{2\sigma^2}\sum_1^n x_i^2}$$

D'où, en appliquant $ln$ à l'expression précédente strictement positive,

$$ln(\prod f_X(x_i)) = \sum_1^n ln(x_i) - 2nln(\sigma) - \frac{\sum_1^n x_i^2}{2\sigma^2}$$

D'où en dérivant par rapport à $\theta$,

$$\frac{\partial}{\partial\theta}ln(\prod_i^n f_{X_i}(x_i)) = 0 - \frac{2n}{\sigma} + \frac{\sum x_i^2}{2}(\frac{2}{\sigma^3})$$

Donc

$$\frac{\partial}{\partial\theta}ln(\prod_1^n f_{X_i}(x_i)) = 0$$

$$\Updownarrow$$

$$2n = \frac{\sum_1^n x_i^2}{2n}$$

$$\Updownarrow$$

$$\sigma^2 = \frac{\sum_1^n X_i^2}{2n}$$

Donc on a la valeur de l'estimateur

$$\boxed{\tilde{\sigma_n}^2 = \frac{\sum_1^n X_i^2}{2n}} \tag{6}$$

D'où

$$\boxed{E[\tilde{\sigma_n}^2] = \frac{1}{2n}\sum_1^n E(X_i^2) = \frac{1}{2}2\sigma^2 = \sigma^2} \tag{7}$$

Donc l'estimateur est sans biais. Ensuite

$$\frac{\partial}{\partial\theta}[E[\tilde{\sigma_n}^2]] = 2\sigma$$

Donc

$$(\frac{\partial}{\partial\theta}[E[\tilde{\sigma_n}^2]])^2 = 4\sigma^2$$

$$\frac{\partial}{\partial\theta}(ln(\mathcal{L}(\sigma, X_1, ..., X_n))) = -\frac{2n}{\sigma} + \frac{\sum X_i^2}{\sigma^3}$$

D'où

$$E[\frac{\partial^2}{\partial\theta^2}(ln(\mathcal{L}(\sigma, X_1, ..., X_n)))] = -\frac{2n}{\sigma^2} + \frac{6n}{\sigma^2} = \frac{4n}{\sigma^2}$$

Ensuite on a

$$Var(\tilde{\sigma_n}^2) = Var(\frac{\sum_1^n X_i^2}{2n}) = \frac{1}{4n^2}\sum_1^n X_i^2$$

Or $\frac{X^2}{\sigma^2} \hookrightarrow \mathcal{E}(\frac{1}{2})$ d'où $\frac{X^2}{\sigma^2} \hookrightarrow \mathcal{G}(1, \frac{1}{2})$

Donc d'après le cours on a

$$\sigma^2\frac{X^2}{\sigma^2} = X^2 \hookrightarrow \mathcal{G}(1, \frac{1}{2\sigma^2})$$

$$\Downarrow$$

$$Var(X^2) = 4\sigma^4$$

$$\Downarrow$$

$$Var(\tilde{\sigma_n}^2) = \frac{\sigma^4}{n}$$

Et on a donc $In(\tilde{\sigma_n}^2)Var(\tilde{\sigma_n}^2) = \frac{4n}{\sigma^2}\frac{\sigma^4}{n} = 4\sigma^2$.

D'où

$$\boxed{Eff(\tilde{\sigma_n}^2) = \frac{4\sigma^2}{4\sigma^2} = 1} \tag{8}$$

Donc l'estimateur est efficace.

## Question 7

```
## sigma_g2 =  31.48487
## sigma2_EMM =  30.96835
## sigma2p_EMM =  30.88397
## sigma2_EMV =  32.66009
```

## Question 8

On a $\forall i,\ X_i \hookrightarrow \mathcal{R}(\delta^2)$.

Or $\frac{X_i}{\sigma^2} \hookrightarrow \mathcal{E}(\frac{1}{2})$ donc selon l'annexe 7.3.4 du cours, on sait que $\sum_1^n X_i$ suit une loi gamma de paramètres $(n, \frac{1}{2})$.

On a alors

$$\begin{aligned}\mathbb{P}(a \leq \frac{\sum_1^n X_i^2}{\delta^2} \leq b) &= \mathbb{P}(\frac{1}{a} \geq \frac{\delta^2}{\sum_1^n X_i^2} \geq \frac{1}{b}) \\ &= \mathbb{P}(\frac{\sum_1^n X_i^2}{a} \geq \delta^2 \geq \frac{\sum_1^n X_i^2}{b}) \\ &= F_{\chi^2_{2n}}(b) - F_{\chi^2_{2n}}(a) = 1 - \alpha = 1 - \frac{\alpha}{2} - \frac{\alpha}{2}\end{aligned}$$

On pose alors $a = \mathcal{Z}_{2n,\ 1-\frac{\alpha}{2}}$ et $b = \mathcal{Z}_{2n,\ \frac{\alpha}{2}}$.

Donc

$$\mathbb{P}(\frac{\sum_1^n X_i^2}{\mathcal{Z}_{2n,\ 1-\frac{\alpha}{2}}} \geq \delta^2 \geq \frac{\sum_1^n X_i^2}{\mathcal{Z}_{2n,\ \frac{\alpha}{2}}}) = 1 - \alpha$$

D'où finalement

$$\boxed{I = [\frac{\sum_1^n X_i^2}{\mathcal{Z}_{2n,\ 1-\frac{\alpha}{2}}}, \frac{\sum_1^n X_i^2}{\mathcal{Z}_{2n,\ \frac{\alpha}{2}}}]} \tag{9}$$

```
## [1] 27.0973
## [1] 40.14071
```

## Question 9

On note $m_0 = 9$ m/s et $m$ la moyenne.

On pose $H_0 : m \geq m_0$ et $H_1 : m < m_0$. En effet il est grave de rejeter $H_0$ à tord, car on construirait des éoliennes dont les roulements s'useraient trop rapidement

On a l'équivalence

$$\begin{gathered} m \geq m_0 \\ \Updownarrow \\ \sigma\sqrt{\frac{\pi}{2}} \geq \sigma_0\sqrt{\frac{\pi}{2}} \\ \Updownarrow \\ \sigma \geq \sigma_0 \\ \Updownarrow \\ \boxed{\sigma^2 \geq \sigma_0^2} \end{gathered}$$

On a cette dernière équivalence car $\sigma$ est une valeur toujours positive.

Ensuite on a

$$\alpha = \sup_{H_0}(\mathbb{P}(X_1, ..., X_n \in W : \delta^2)) = \sup_{\delta^2 > \delta_0^2}(\mathbb{P}(\frac{\sum_1^n X_i^2}{2n} < l_\alpha))$$

$$= \sup_{H_0}(\mathbb{P}(\frac{\sum_1^n X_i^2}{\delta^2} < \frac{2nl_\alpha}{\delta^2}))$$

D'où d'après la question précédente,

$$\alpha = \sup_{\delta^2 > \delta_0^2}(F_{\chi^2_{2n}}(\frac{2nl_\alpha}{\delta^2})) = F_{\chi^2_{2n}}(\frac{2nl_\alpha}{\delta_0^2})$$

D'où

$$\boxed{\alpha = F_{\chi^2_{2n}}(\frac{2nl_\alpha}{\delta_0^2})}$$

Donc

$$\boxed{\frac{\mathcal{Z}_{2n,1-\alpha}\delta_0^2}{2n} = l_\alpha}$$

On a donc la région critique $(W = (x_1, ..., x_n) : \tilde{\delta_n}^2 < l_\alpha)$.

Le $\alpha_c$ de la p_valeur vérifie

$$\sum_1^n X_i = \mathcal{Z}_{2n,1-\alpha_c}\delta_0^2$$

Donc $\mathcal{Z}_{2n,1-\alpha_c} = \frac{\sum_1^n X_i^2}{\delta_0^2} \Longrightarrow \alpha_c = F_{\chi^2_{2n}}(\frac{\sum_1^n X_i^2}{\delta_0^2})$

```
## [1] 13.52552
```

```
## p_valeur =  1.365938e-78
```

La p_valeur est très faible et on a donc tendance à ne pas rejeter $H_0$, on ne peut donc pas se permettre de constuire des éoliennes dans cette zone.

# Vérifications expérimentales à base de simulations

## Question 1

```r
# Pour simuler un échantillon de taille n de la loi R(sigma²) :
sigma <- 1
n <- 100

ech_rayleigh <- sqrt(rnorm(n, mean = 0, sigma^2)^2 + rnorm(n, mean = 0, sigma^2)^2)
ech_rayleigh
```

```
##   [1] 0.69262620 1.49501718 1.41426251 2.00238771 1.75548357 1.75072523
##   [7] 2.00868579 1.97358243 1.88767607 1.50832154 1.91903604 0.34412058
##  [13] 1.40213024 0.86002681 1.26509596 1.73661223 1.21529937 1.81260675
##  [19] 1.21420979 1.21256093 1.51784427 0.65538271 1.67808873 0.44167812
##  [25] 1.33517688 0.18106446 0.62271979 0.65645857 0.61733658 2.12433829
##  [31] 0.96976301 1.75233240 1.19471913 1.17985594 0.76186372 1.28900589
##  [37] 0.32180645 1.49286072 0.32860836 1.17174186 0.09873914 1.91675851
##  [43] 1.14997609 1.90957517 0.99441528 0.36454108 0.20654319 0.71786770
##  [49] 0.58872226 1.27564466 0.40115345 2.11470596 2.04070245 0.20571307
##  [55] 1.38410704 1.08209053 1.43069005 1.36881013 0.76418118 1.19031002
```

```
##  [61] 1.29331301 1.72219961 1.06527237 1.52823746 3.19176270 1.80119054
##  [67] 0.60295589 1.62765239 0.78612644 2.47557323 0.39864594 0.88158434
##  [73] 1.97923486 0.61441028 1.89325084 2.00476870 1.68190047 2.69211999
##  [79] 1.85875903 1.60214894 1.04449446 2.25144408 1.58084100 0.85496612
##  [85] 0.76223358 1.11691411 1.38449424 1.95786205 0.65592188 1.00775770
##  [91] 0.66454224 0.63528016 2.09098258 0.95015559 1.28415430 2.61982461
##  [97] 0.68375719 3.18911654 1.94111043 1.22833864
```

### Question 2

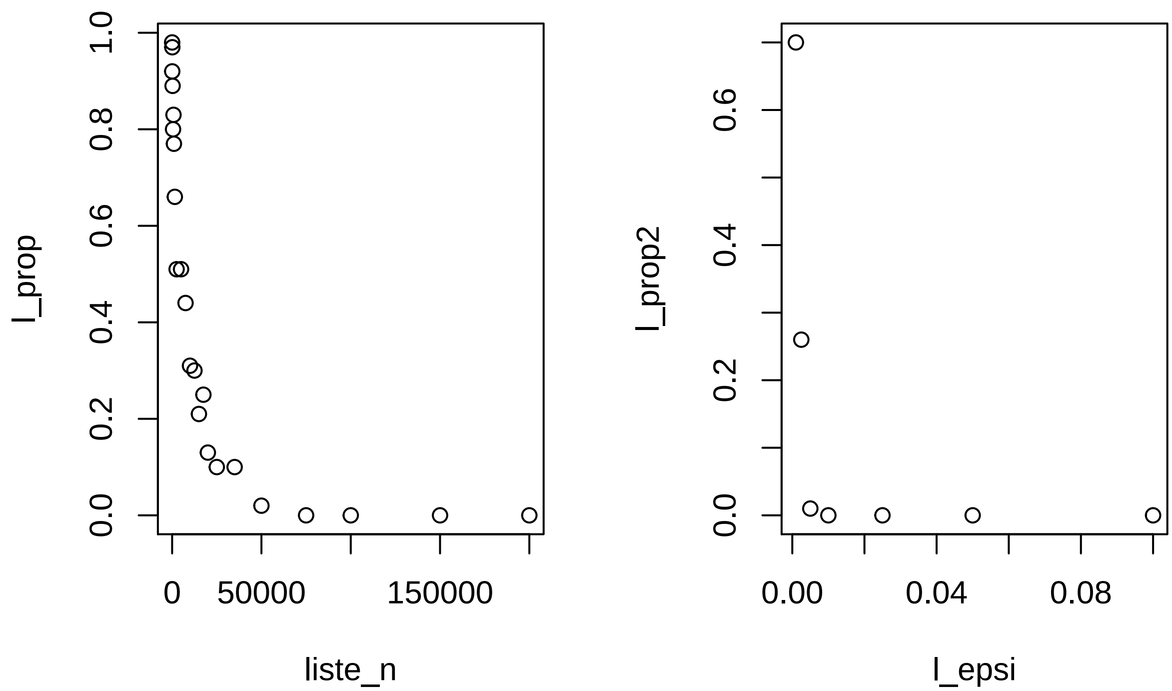
```
## proportion d'appartenance à l'intervalle de confiance =  0.955
```

### Question 3

```
## biais_g =  0.2193811
## biais_EMM =  -0.01016883
## biais_p_EMM =  -0.01286607
## biais_EMV =  -0.007499806
## rq_g =  0.04812807
## rq_EMM =  0.01104474
## rq_p_EMM =  0.01104732
## rq_EMV =  0.01042322
```

On choisit le maximum de vraisemblance car il est sans biais et efficace.

### Question 4




On a

$$\begin{aligned}\mathbb{P}(|T_n-\sigma^2|>\epsilon) &= \mathbb{P}(T_n>\sigma^2+\epsilon)+\mathbb{P}(T_n<\sigma^2-\epsilon)\\ &= 1-(\mathbb{P}(T_n<\sigma^2+\epsilon)-\mathbb{P}(T_n<\sigma^2-\epsilon))\end{aligned}$$

D'où

$$\boxed{\mathbb{P}(|T_n-\sigma^2|>\epsilon) = 1-\mathbb{P}(\sigma^2-\epsilon<T_n<\sigma^2+\epsilon)} \tag{10}$$

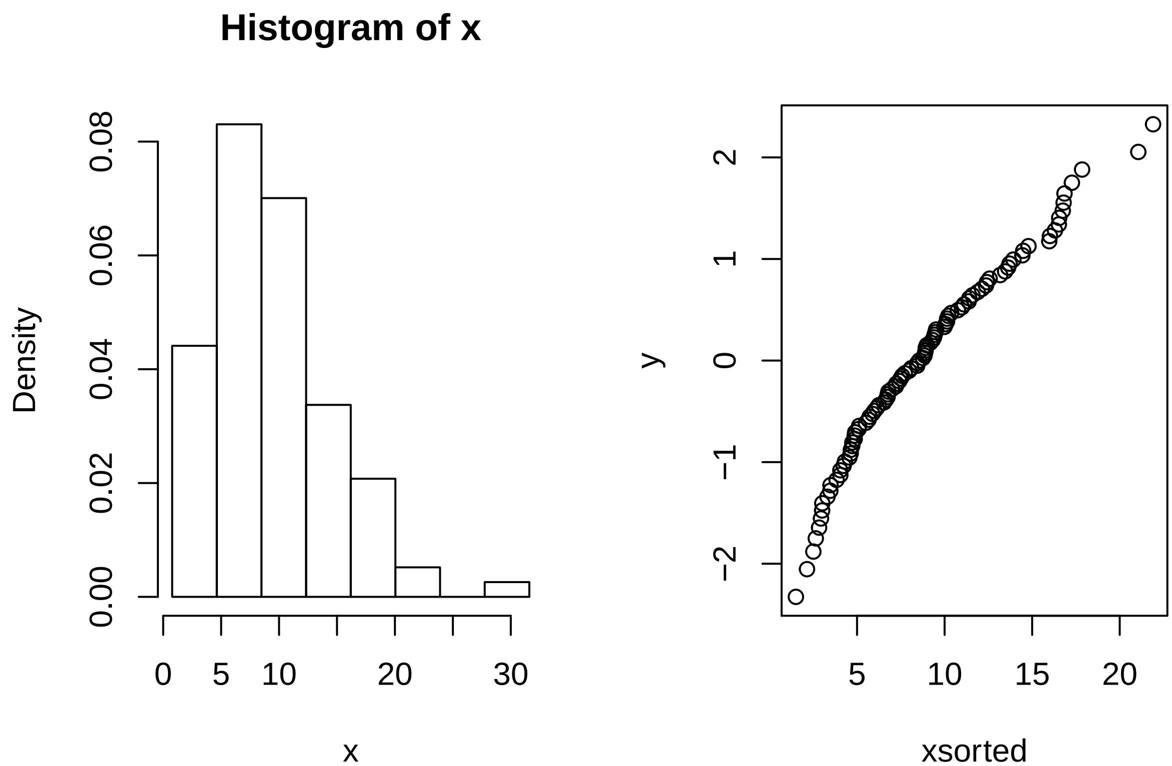
Or $T_n$ est sans biais et efficace donc $T_n \underset{n\to+\infty}{\longrightarrow} \sigma^2$.

Donc $|T_n-\sigma^2|\leq\epsilon$ est un évenement quasi certain lorsque $n$ tend vers $+\infty$ et $\forall\epsilon>0,\ \mathbb{P}(|T_n-\sigma^2|\leq\epsilon) \underset{n\to+\infty}{\longrightarrow} 1$ et donc $\mathbb{P}(|T_n-\sigma^2|) \underset{n\to+\infty}{\longrightarrow} 0$.

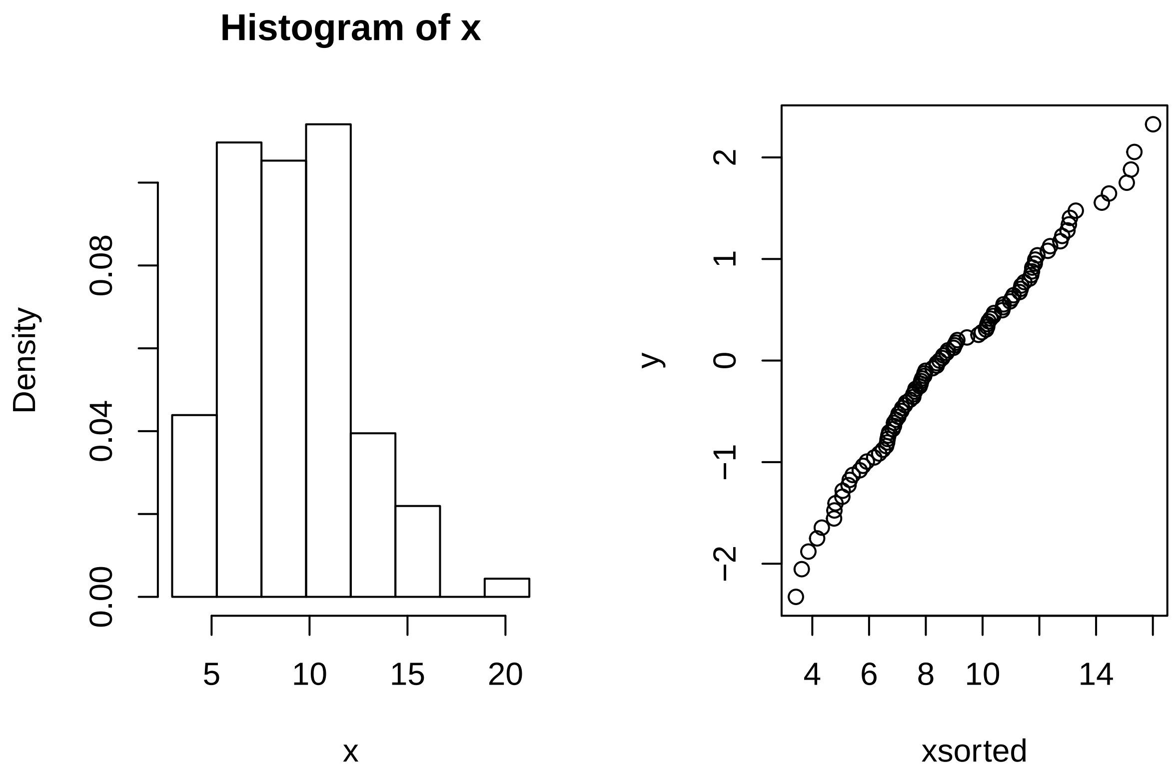
On peut clairement voir que plus on augmente $n$ et $\epsilon$, plus l'estimation de la probabilité tend rapidement vers 0. Mais plus $\epsilon$ est petit, plus la valeur 0 sera difficile à atteindre.

### Question 5

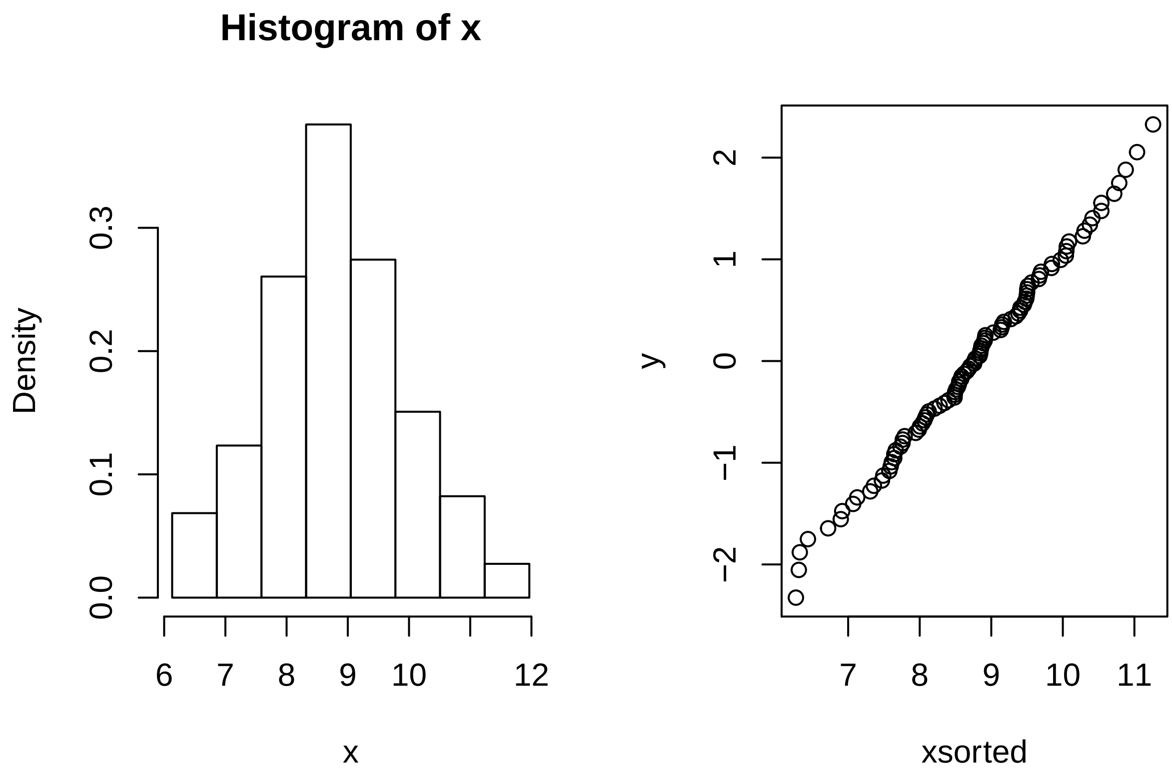
```
## n =  100moyenne =  8.968311sigma^2 =  23.15781
```



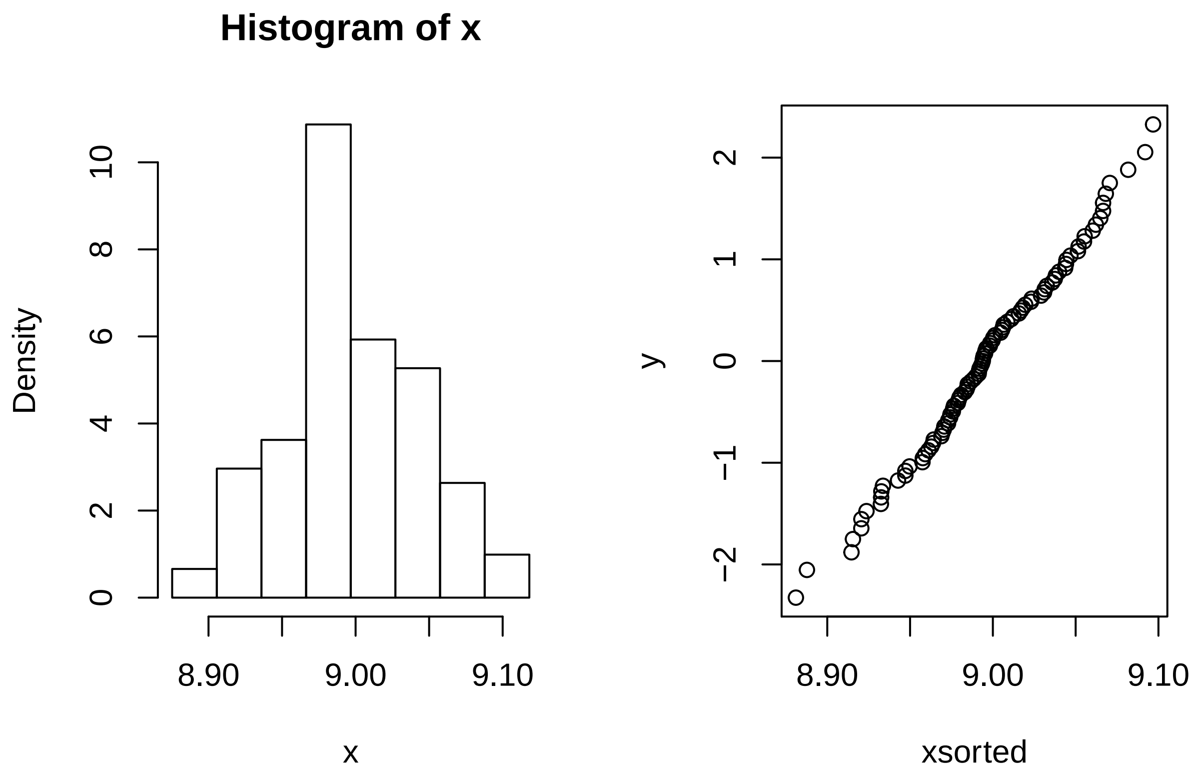

```
## n =  100moyenne =  8.935848sigma^2 =  9.406766
```




```
## n =  100moyenne =  8.749262sigma^2 =  1.375888
```




```
## n =  100moyenne =  8.996258sigma^2 =  0.002172624
```




On constate donc que l'estimateur du maximum de vraisemblance est asymptotiquement efficace.