# TP - PMS

Hagenburg Arthur Vu Germain Samake Habibata

---

**1 - Analyse du vent**

**1.**

On commence par créer des histogrammes de même classe pour les deux jeux de données A1 et A2 avec le code :

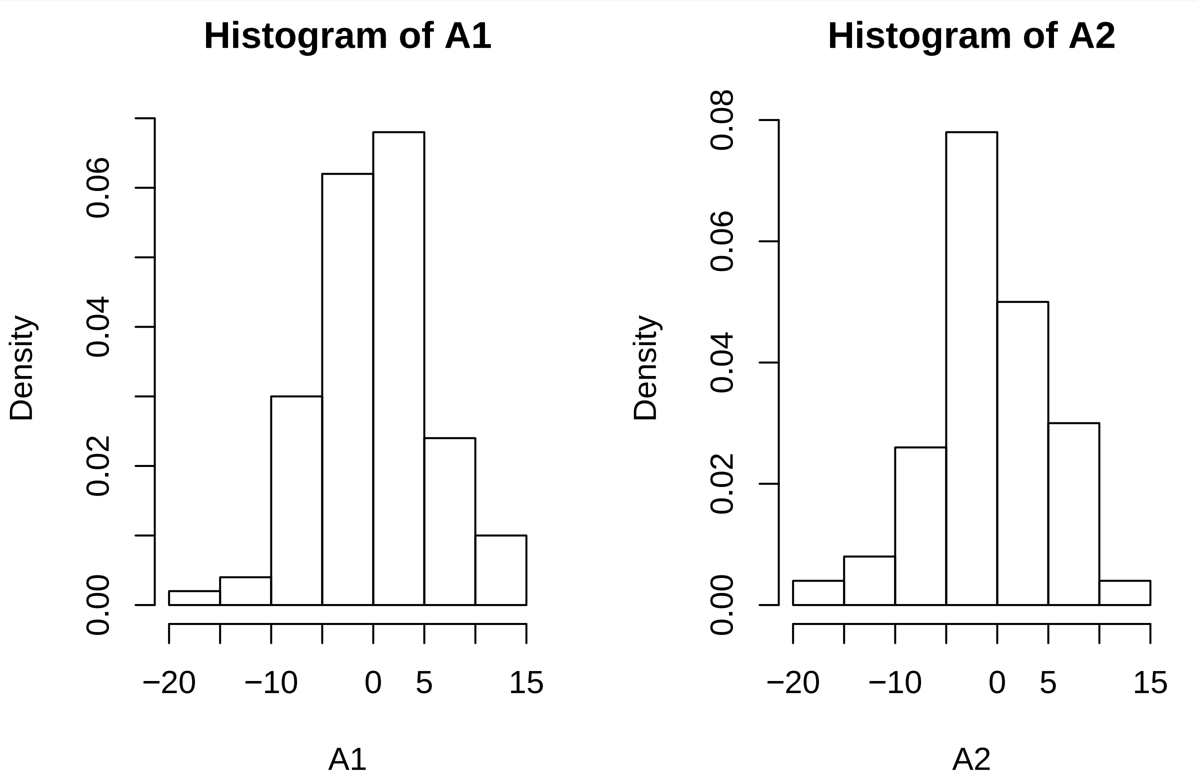
```
vent <- read.table("~/Documents/Cours/S2/PMS/TP/vent.csv", sep=";", header=T)
attach(vent)

n = length(A1)

split.screen((1:2))
```

```
## [1] 1 2
```

```
screen(1);hist(A1, prob=T)
screen(2);hist(A2, prob=T)
```




On constate que A1 et A2 semble effectivement suivrent une loi normale centrée et semble avoir la même variance.

Il est donc raisonnable d'admettre que A1 et A2 sont des variables aléatoires de loi normale centrée et de même variance $\sigma^2$

Avec le code :

```
# Première estimation :
v1 = var(A1)
v2 = var(A2)

sigma_carre_naif = (v1+v2)/2
sigma_carre_naif
```

```
## [1] 32.80702
```

On trouve une première estimation de $\sigma^2$ en faisant la moyenne des variances de nos deux échantillons. On trouve

$$\boxed{\sigma^2_{naïf} = 32.81}$$

**2.**

On a :

$$\frac{X^2}{\sigma^2} = \frac{A_1^2}{\sigma^2} + \frac{A_2^2}{\sigma^2}$$

On sait que $\frac{A_1}{\sigma}$ suit une loi normale $\mathcal{N}(0,1)$, de même pour $A_2$
Donc $\frac{A_1^1}{\sigma^2}$ et $\frac{A_2^2}{\sigma^2}$ suivent la loi du Chi2 à 1 degrès de liberté : $\chi^2_1$
Or une somme de variable aléatoire suivant la loi du Chi2 à 1 degrès de liberté suit une loi Gamma $\mathcal{G}(1, 1/2)$ qui est aussi une loi exponentielle $\mathcal{E}xp(1/2)$

Donc

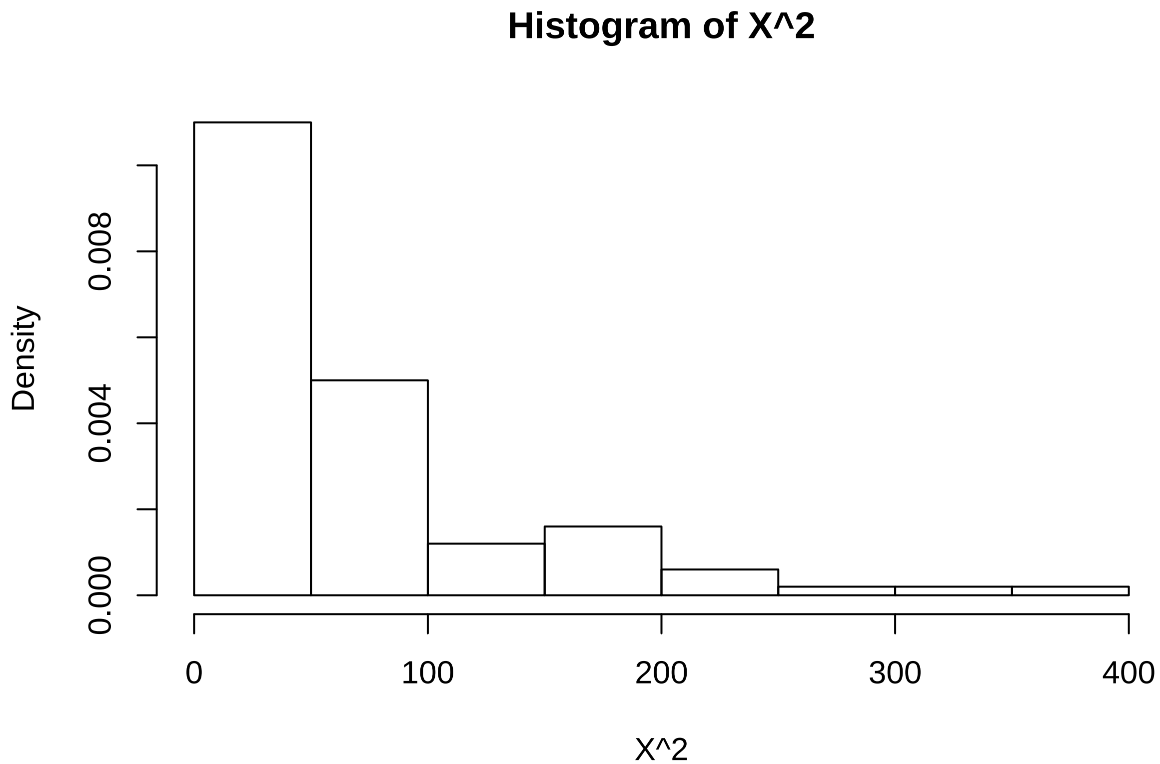
$$\boxed{\frac{X^2}{\sigma^2} \hookrightarrow \mathcal{E}xp(\frac{1}{2})}$$

.

Pour le vérifier sur notre jeu de données, on trace l'histogramme de $X^2$ pour vérifier que son allure est proche de celle d'une loi exponentielle.
De plus, pour vérifier la valeur du paramètre $\lambda$, on calcule la moyenne de $\frac{X^2}{\sigma^2}$ en prenant l'estimation $\sigma^2_{naïf}$ pour $\sigma^2$. Si tout est bon, comme $\mathbb{E}[\frac{X^2}{\sigma^2}] = \frac{1}{\lambda}$, on doit trouver $lambda = \frac{1}{2}$

```
X = sqrt(A1**2+A2**2)

hist(X**2, prob=T)
```




```
E = mean(X**2/32.81)
lambda = 1/E
lambda
```

```
## [1] 0.502295
```

On trouve : $lambda = 0.5023$ qui est très proche de $\frac{1}{2}$ et l'allure de l'histogramme est bien celle d'une loi exponentielle.

**3.**

$X \geq 0$ donc

$$F_X(t) = P(X \leq t) = P\left(\frac{X^2}{\sigma^2} \leq \frac{t^2}{\sigma^2}\right) = F_{\frac{X^2}{\sigma^2}}\left(\frac{t^2}{\sigma^2}\right)$$

Donc

$$F_X(t) = \left(1 - \exp\left(-\frac{t^2}{2\sigma^2}\right)\right) 1_{\mathbb{R}^+}(t)$$

Donc en dérivant :

$$f_X(t) = \frac{t}{\sigma^2} \exp\left(-\frac{t^2}{2\sigma^2}\right) 1_{\mathbb{R}^+}(t)$$

Donc

$$\boxed{X \hookrightarrow \mathcal{R}(\sigma^2)}$$

**4.**

On trouve facilement que :

$$\ln(1 - F_X(t)) = -\frac{t^2}{2} \frac{1}{\sigma^2}$$

Le graphe de probabilité est donc de la forme

$$\left(\frac{x_i^2}{2}, \ln\left(1-\frac{i}{n}\right)\right)_{i\in\{1,..,n\}}$$

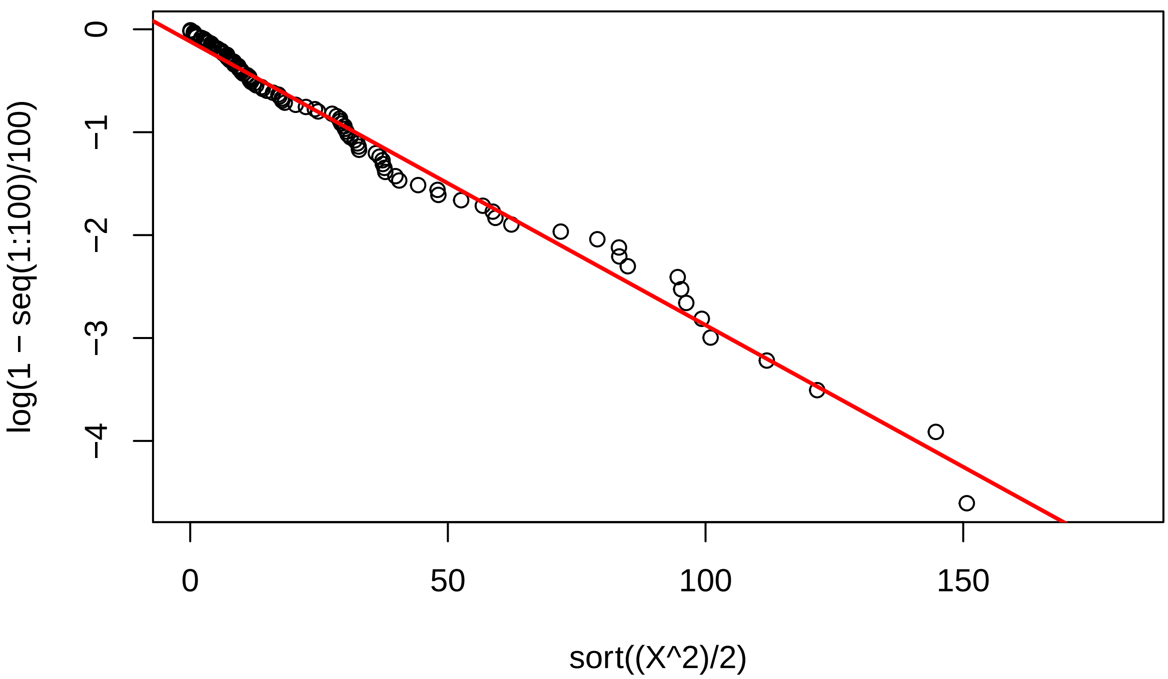
On peut donc tracer le graphe de probabilité sur R et faire une regression linéaire pour estimer graphiquement le paramètre $\sigma^2$:

```
plot(sort((X**2)/2), log(1-seq(1:100)/100))

abs <- sort((X**2)/2)
ord <- log(1-seq(1:100)/100)

reg <- lm(ord[1:99]~abs[1:99])

abline(reg, col="red", lwd=2)
```




```
sigma_carre_g = -1/(coef(reg)[2])
sigma_carre_g
```

```
## abs[1:99]
##  36.28843
```

On trouve donc $\boxed{\sigma_g^2 = 36.288}$ comme estimation graphique de $\sigma^2$ qui est relativement proche de $\sigma^2_{naïf} = 32.81$

**5.**

$X > 0$, on a donc :

$$\mathbb{E}[X] \; = \int_0^{+\infty} (1 - F_X(x))dx = \int_0^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx$$

On reconnait une intégrale de Gauss dont la valeur est :

$$\boxed{\mathbb{E}[X] \; = \sigma\sqrt{\frac{\pi}{2}}}$$

De plus, comme $\frac{X^2}{\sigma^2} \hookrightarrow \mathcal{E}xp(\frac{1}{2})$ on a :

$$\mathbb{E}[X^2] = \frac{1}{\sigma^2}\mathbb{E}\left[\frac{X^2}{\sigma^2}\right] = 2\sigma^2$$

Donc

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2\left(\frac{4-\pi}{2}\right)$$

D'où :

$$\boxed{Var(X) = \sigma^2\left(\frac{4-\pi}{2}\right)}$$

On a donc directement l'estimateur

$$\boxed{\tilde{\sigma}_n^2 = \frac{2}{4-\pi}S_n^{'2}}$$

et

$$\mathbb{E}[\tilde{\sigma}_n^2] = \frac{2}{4-\pi}Var(X) = \frac{2}{4-\pi}\sigma^2$$

Donc $\tilde{\sigma}_n^2$ est biaisé et

$$\boxed{\tilde{\sigma}_n^{'2} = \frac{4-\pi}{2}\tilde{\sigma}_n^2}$$

où $\tilde{\sigma}_n^{'2}$ est un estimateur sans biais de $\sigma^2$

**6.**

On calcule la log-vraisemblance :

$$\begin{aligned}\ln\left(\mathcal{L}(\sigma^2, x_i)\right) &= \ln\left(\prod_{i=1}^n f_X(x_i, \sigma^2)\right) = \sum_{i=1}^n \ln\left(f_X(x_i, \sigma^2)\right) \\ &= \sum_{i=1}^n \ln\left(\frac{x_i}{\sigma^2}e^{-\frac{x_i^2}{2\sigma^2}}\right) = \sum_{i=1}^n\left[\ln(x_i) - 2\ln(\sigma) - \frac{x_i^2}{2\sigma^2}\right]\end{aligned}$$

On trouve :

$$\frac{\partial}{\partial\sigma^2}\left(\ln\left(\mathcal{L}(\sigma^2, x_i)\right)\right) = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n x_i^2$$

Or

$$-\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n x_i^2 = 0 \Leftrightarrow \sigma^2 = \frac{1}{2n}\sum_{i=1}^n x_i^2$$

Donc

$$\boxed{\hat{\sigma}_n^2 = \frac{1}{2n}\sum_{i=1}^n X_i^2}$$

De plus,

$$\mathbb{E}[\hat{\sigma}_n^2] = \frac{1}{2n}\sum_{i=1}^n[X_i^2] = \frac{1}{2n}\sum_{i=1}^n 2\sigma^2 = \sigma^2$$

Donc $\hat{\sigma}_n^2$ est sans biais.

Calculons l'efficacité de cet estimateur :

- $\frac{\partial}{\partial \sigma^2}\mathbb{E}[\hat{\sigma}_n^2] = 1$
- $I_n(\sigma^2) = -\mathbb{E}\left[\frac{\partial}{\partial \sigma^2}\left(-\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n x_i^2\right)\right] = -\mathbb{E}\left[-\frac{n}{\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^n x_i^2\right] = \frac{n}{\sigma^4}$
- $Var(\hat{\sigma}_n^2) = \frac{1}{4n^2}Var(\sum_{i=1}^n X_i^2 \overset{i.i.d}{=} \frac{1}{4n^2}\sum_{i=1}^n Var(X_1^2) = \frac{\sigma^4}{4n}Var(\frac{X_1^2}{\sigma^2}) = \frac{\sigma^4}{n}$ car $Var(\frac{X_1^2}{\sigma^2}) = 4$

Or

$$Eff(\hat{\sigma}_n^2) = \frac{\frac{\partial}{\partial \sigma^2}\mathbb{E}[\hat{\sigma}_n^2]}{I_n(\sigma^2)Var(\hat{\sigma}_n^2)}$$

Donc

$$\boxed{Eff(\hat{\sigma}_n^2) = 1}$$

et $\hat{\sigma}_n^2$ est efficace.

**7.**

```
cat("Première estimation : ", sigma_carre_naif)
```

```
## Première estimation :  32.80702
```

```
cat("Méthode graphique : ", sigma_carre_g)
```

```
## Méthode graphique :  36.28843
```

```
sigma_carre_tilde = 2/(4-pi) * var(X)
cat("Méthode des moments : ", sigma_carre_tilde)
```

```
## Méthode des moments :  39.24392
```

```
sigma_carre_chapeau = 1/(2*100) * sum(X**2)
cat("Méthode du maximum de vraissemblance : ", sigma_carre_chapeau)
```

```
## Méthode du maximum de vraissemblance :  32.66009
```

On choisit l'estimateur par la méthode du maximum de vraisemblance car il est sans biais et efficace. On retient donc $\boxed{\sigma^2 \simeq 32.66009}$

**8.**

Prenons $Y = \sum_{i=1}^n \left(\frac{X_i}{\sigma}\right)^2$ comme fonction pivotale.

Comme $\forall i \left(\frac{X_i}{\sigma}\right)^2 \hookrightarrow \mathcal{E}xp(\frac{1}{2})$ alors $Y \hookrightarrow \mathcal{G}(n, 1/2) = \chi^2_{2n}$

On a donc directement que :

$$P(z_{2n,1-\frac{\alpha}{2}} \leq Y \leq z_{2n,\frac{\alpha}{2}}) = 1 - \alpha$$

Et donc :

$$P\left(\frac{\sum_{i=1}^n X_i^2}{z_{2n,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{\sum_{i=1}^n X_i^2}{z_{2n,\frac{\alpha}{2}}}\right) = 1 - \alpha$$

On peut donc calculer cet intervalle avec R :

```
alpha = 0.05
b_inf = sum(X**2)/qchisq(1-alpha/2, 2*100)
b_sup = sum(X**2)/qchisq(alpha/2, 2*100)
cat("Intervalle de confiance pour sigma_carré au seuil de 5% : [",b_inf,";",b_sup,"]")
```

```
## Intervalle de confiance pour sigma_carré au seuil de 5% : [ 27.0973 ; 40.14071 ]
```

9.

On pose le test suivant :

- $H_0$ : Le terrain est construtible : $\mathbb{E}[X] \leq 9$ m/s $\Leftrightarrow \sigma^2 \leq \sigma_0^2 = 81\frac{2}{\pi}$ m/s
- $H_1$ : Le terrain n'est pas construtible : $\mathbb{E}[X] > 9$ m/s $\Leftrightarrow \sigma^2 > \sigma_0^2 = 81\frac{2}{\pi}$ m/s

On calcul donc

$$\alpha = \sup_{\sigma^2 \leq \sigma_0^2} P(\hat{\sigma}_n^2 > l_\alpha) = \sup_{\sigma^2 \leq \sigma_0^2} P\left(\frac{2n}{\sigma^2}\hat{\sigma}_n^2 > \frac{2n}{\sigma^2} l_\alpha\right)$$

Or

$$\frac{2n}{\sigma^2}\hat{\sigma}_n^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 = Y \hookrightarrow \chi^2_{2n}$$

Donc

$$\alpha = \sup_{\sigma^2 \leq \sigma_0^2}\left(1 - F_Y\left(\frac{2n}{\sigma^2} l_\alpha\right)\right)$$

Or $1 - F_Y\left(\frac{2n}{\sigma^2} l_\alpha\right)$ est une fonction croissante de $\sigma$, le sup est donc atteint en $\sigma_0$

Donc

$$\alpha = 1 - F_Y\left(\frac{2n}{\sigma_0^2} l_\alpha\right)$$

Finalement on a :

$$l_\alpha = \frac{\sigma_0^2}{2n} z_{2n,\alpha}$$

La zone critique de ce test est donc :

$$\boxed{W = \left\{\frac{2n}{\sigma_0^2}\hat{\sigma}_n^2 > z_{2n,\alpha}\right\}}$$

Calculons la valeur de la statistique :

```
sigma_0_carre = 2 / pi * 81
stat = 2*100 / sigma_0_carre * sigma_carre_chapeau

quant = qchisq(0.95, 2*100)
cat("stat = ", stat, "quantile = ", quant)
```

```
## stat =  126.6725 quantile =  233.9943
```