

Modélisation des risques d'AVC avec la POT

Adama Sall

2025-06-20

METHODOLOGIE ADAPTÉE

1. Choix des seuils
2. Extraction des excès
3. Ajustement de la GPD
4. Diagnostics visuels
5. Probabilités de dépassement
6. Quantiles extrêmes

```
library(evir)
library(evd)

library(ismev)

library(readxl)
library(dplyr)

library(tidyverse)

library(extRemes)

library(ggplot2)
library(gridExtra)

donnee <- read.csv("C:/Users/Hp/Desktop/formation R/Données/base_impropre.csv")
head(donnee)

##   Genre Age Hypertension Maladie_Cardiaque Situation_Matrimoniale Type_travail
## 1     1  67           0              1              1
## 2     0  61           0              0              1
## 3     1  80           0              1              1
## 4     0  49           0              0              1
```

```

2
## 5      0  79          1          0          1
3
## 6      1  81          0          0          1
2
##   Residence Taux_glucose_moyen      IMC Statut_Fumer AVC
## 1          1          228.69 36.60000      0      1
## 2          0          202.21 28.89324      1      1
## 3          0          105.92 32.50000      1      1
## 4          1          171.23 34.40000      2      1
## 5          0          174.12 24.00000      1      1
## 6          1          186.21 29.00000      0      1

```

`attach(donnee)`

`summary(donnee)`

```

##      Genre      Age      Hypertension      Maladie_Cardiaque
## Min.   :0.0000   Min.   : 0.08   Min.   :0.00000   Min.   :0.00000
## 1st Qu.:0.0000   1st Qu.:25.00   1st Qu.:0.00000   1st Qu.:0.00000
## Median :0.0000   Median :45.00   Median :0.00000   Median :0.00000
## Mean   :0.4143   Mean   :43.23   Mean   :0.09746   Mean   :0.05401
## 3rd Qu.:1.0000   3rd Qu.:61.00   3rd Qu.:0.00000   3rd Qu.:0.00000
## Max.   :2.0000   Max.   :82.00   Max.   :1.00000   Max.   :1.00000
## Situation_Matrimoniale Type_travail      Residence      Taux_glucose_moyen
## Min.   :0.0000      Min.   :0.000   Min.   :0.000   Min.   : 55.12
## 1st Qu.:0.0000      1st Qu.:2.000   1st Qu.:0.000   1st Qu.: 77.25
## Median :1.0000      Median :2.000   Median :1.000   Median : 91.89
## Mean   :0.6562      Mean   :2.168   Mean   :0.508   Mean   :106.15
## 3rd Qu.:1.0000      3rd Qu.:3.000   3rd Qu.:1.000   3rd Qu.:114.09
## Max.   :1.0000      Max.   :4.000   Max.   :1.000   Max.   :271.74
##      IMC      Statut_Fumer      AVC
## Min.   :10.30   Min.   :0.0000   Min.   :0.00000
## 1st Qu.:23.80   1st Qu.:1.0000   1st Qu.:0.00000
## Median :28.40   Median :1.0000   Median :0.00000
## Mean   :28.89   Mean   :0.9812   Mean   :0.04873
## 3rd Qu.:32.80   3rd Qu.:1.0000   3rd Qu.:0.00000
## Max.   :97.60   Max.   :2.0000   Max.   :1.00000

```

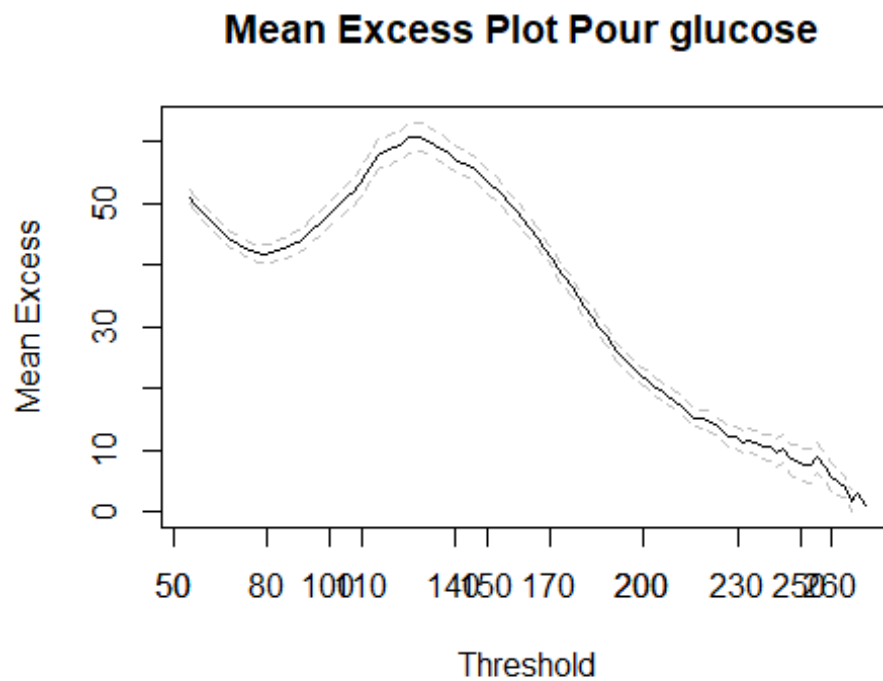
Selection de seuil pour la glucose avec les méthodes graphiques

Méthode 1: Mean Excess Plot (Approche POT classique)

```

mrlplot(donnee$Taux_glucose_moyen, main = "Mean Excess Plot Pour glucose ", x
lab = "Threshold")
axis(1, at = seq(50, 280, by = 30))

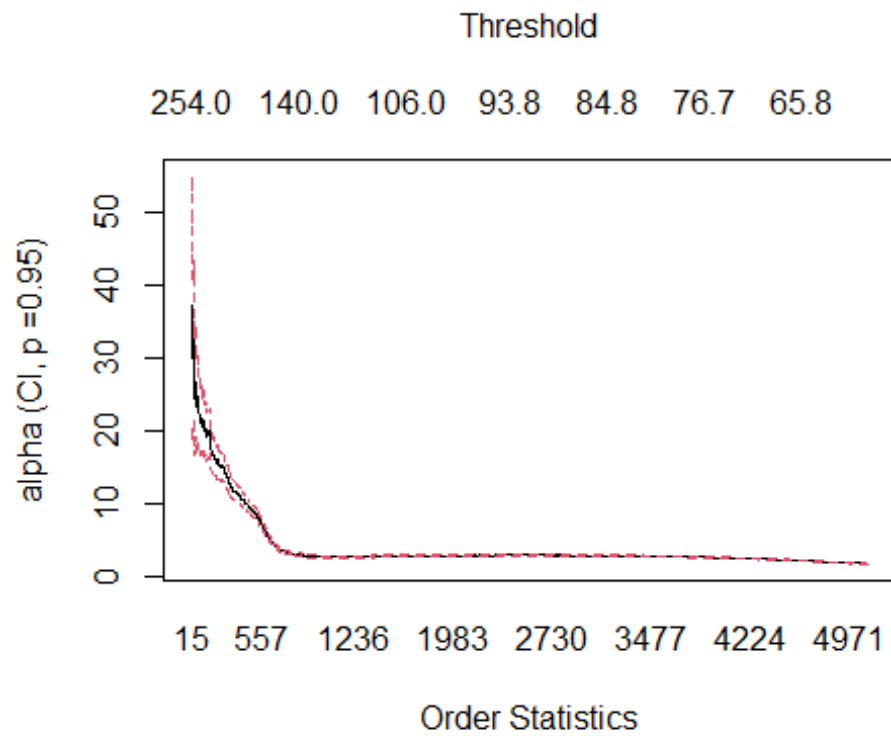
```



La courbe devient relativement linéaire à partir de ~130 jusqu'à ~180. Le seuil optimal semble être dans l'intervalle [130, 150].

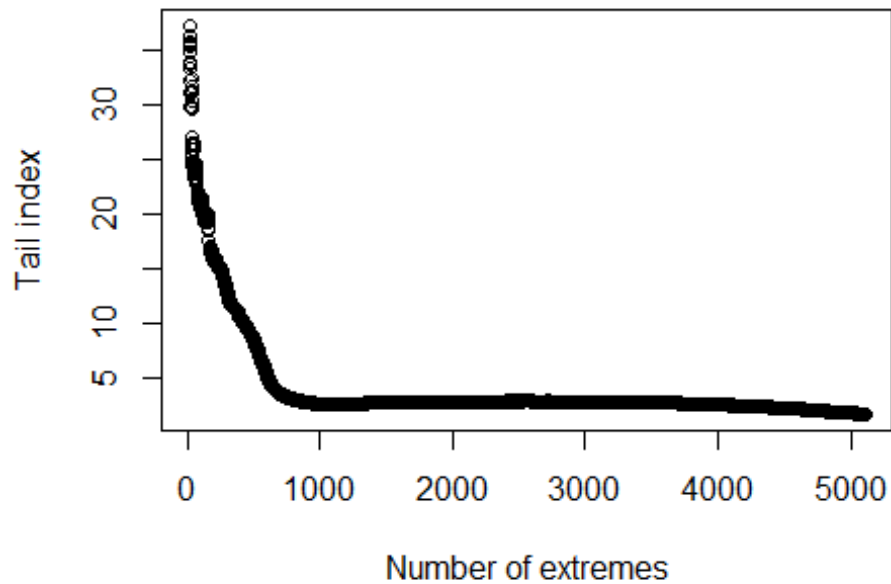
Méthode 2: Hill Plot

```
# Hill Plot  
hill_result <- hill(donnee$Taux_glucose_moyen)
```



```
plot(hill_result,  
     main = "Hill Plot - Glucose",  
     xlab = "Number of extremes",  
     ylab = "Tail index")
```

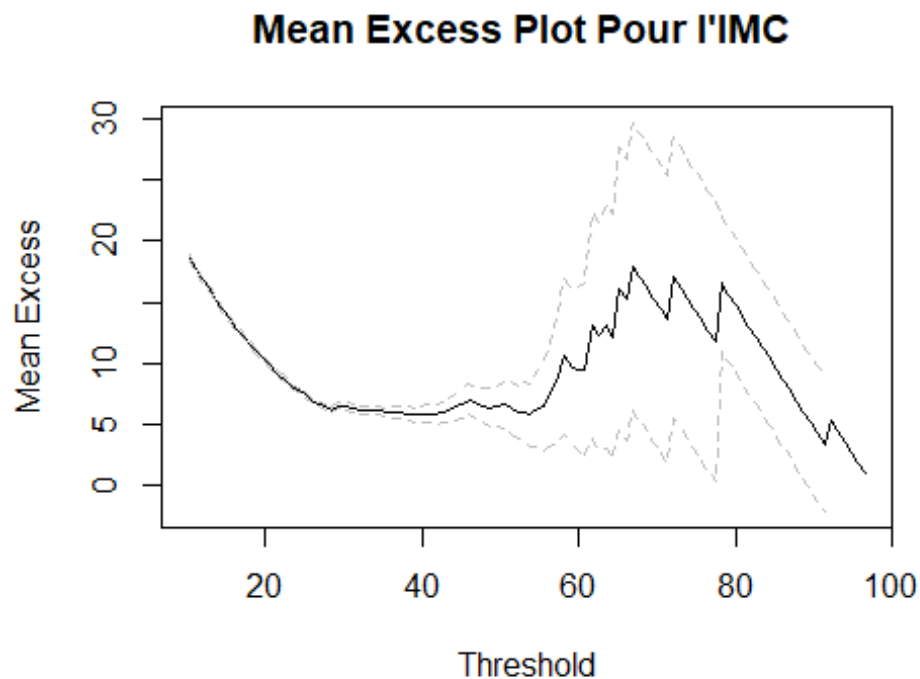
Hill Plot - Glucose



L'indice devient stable entre 4–8 après environ 715 à 1 500 observations extrêmes .
Cela correspond, selon en haut, à un seuil d'environ 109–124.

Selection de seuil pour IMC avec les méthodes graphiques

```
mrlplot(donnee$IMC, main = "Mean Excess Plot Pour l'IMC ", xlab = "Threshold")
```



Analyse du graphe

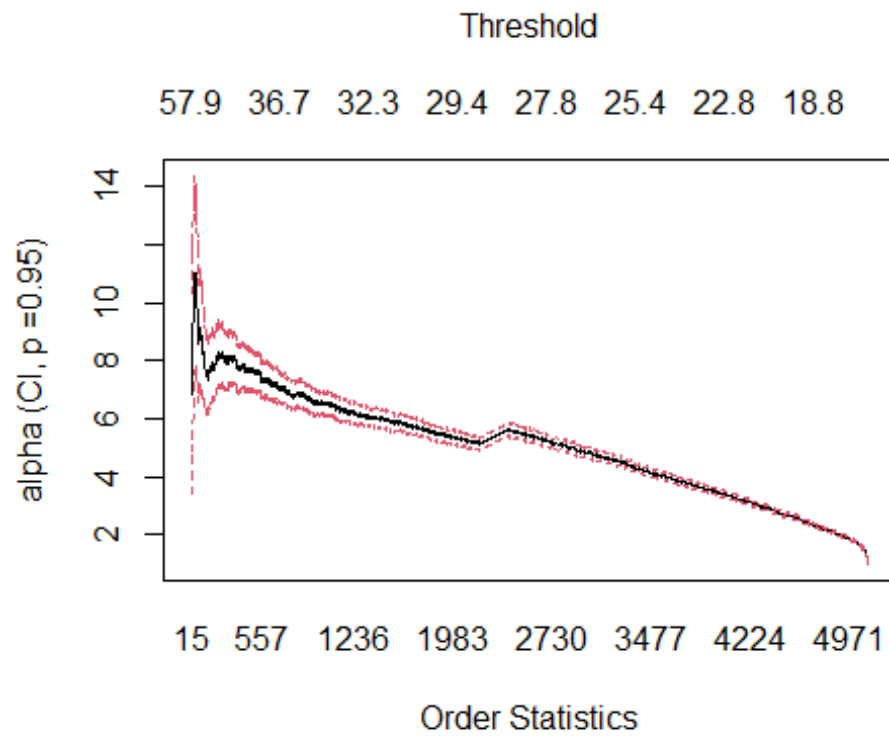
Pour des seuils faibles entre 10 et 30, le mean excess diminue rapidement, ce qui suggère que les valeurs au-dessus de ces seuils ne sont pas particulièrement extrêmes et que la queue de la distribution n'est pas lourde dans cette région.

Entre les seuils 30 et 60, le mean excess devient stable et faible, indiquant une absence d'accumulation de valeurs extrêmes dans cette région.

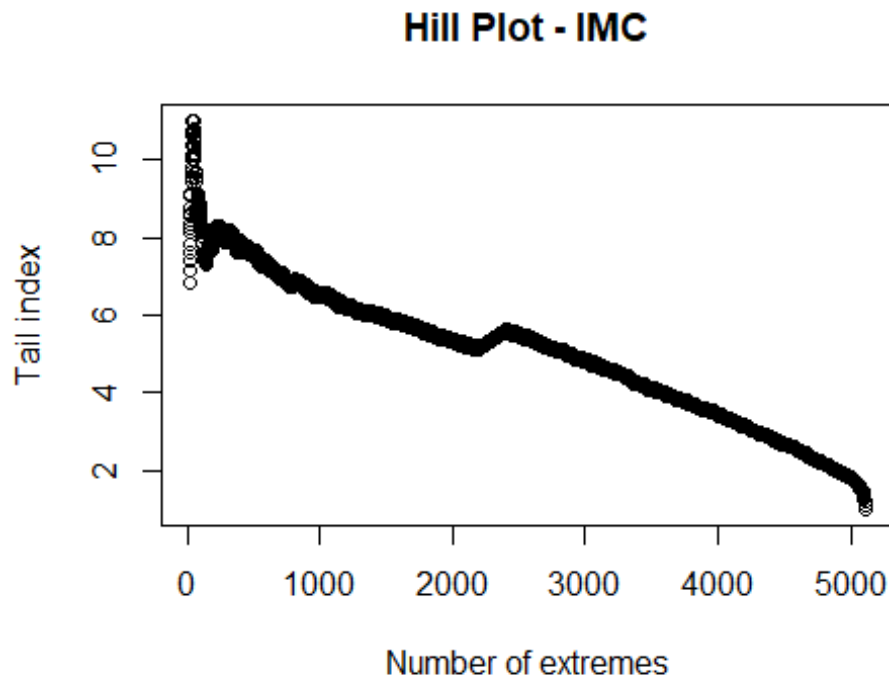
À partir d'un seuil d'environ 60, le mean excess commence à augmenter de manière marquée, atteignant un pic autour de 70-80, puis redescend. Cette augmentation indique la présence de valeurs extrêmes au-delà de ces seuils, ce qui peut signaler un changement de comportement dans la distribution de l'IMC.

D'après les résultats on peut retenir un seuil autour de 60

```
# Hill Plot  
hill <- hill(donnee$IMC)
```



```
plot(hill,  
     main = "Hill Plot - IMC",  
     xlab = "Number of extremes",  
     ylab = "Tail index")
```



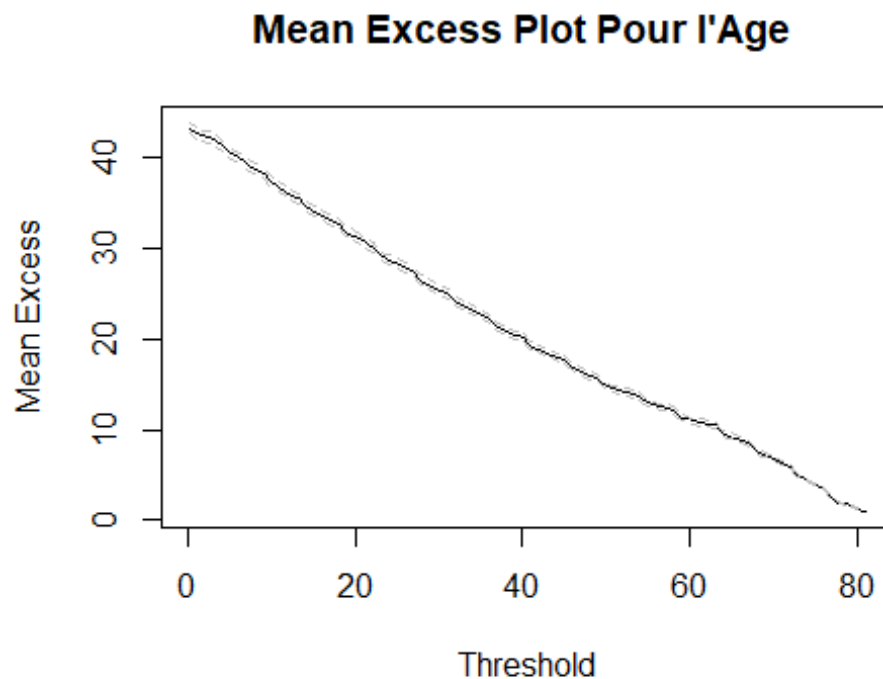
Pour des seuils plus faibles (< 27), la courbe décroît, ce qui indique que l'on inclut des valeurs moins extrêmes.

Entre les seuils d'environ 27 à 32 (ordre statistique ~ 1000 à 2000), la courbe de alpha se stabilise autour d'une valeur comprise entre 5 et 7, avec des intervalles de confiance plus étroits. Cette zone de stabilité pour choisir un seuil.

Pour des seuils élevés (seuil > 30), l'estimation de alpha est instable et les intervalles de confiance sont larges,

Selection de seuil pour l'age avec les méthodes graphiques

```
mrlplot(donnee$Age, main = "Mean Excess Plot Pour l'Age ", xlab = "Threshold"
)
```

La courbe du mean excess montre une décroissance lineaire sans rupture ni queue lourde .Ce comportement indique que la distribution de l'âge ne présente pas de queue lourde ni de rupture brutale : il n'y a pas de seuil à partir duquel les excès moyens augmentent ou se stabilisent.

Le mean excess plot ne montre pas de zone de stabilité qui justifierait le choix d'un seuil spécifique..

```
age_ex <- donnee[donnee$Age >= 60,]$Age
glucose_ex <- donnee[donnee$Taux_glucose_moyen >= 126,]$Taux_glucose_moyen
IMC_ex <- donnee[donnee$IMC >= 30,]$IMC
```

Visualisation des excès

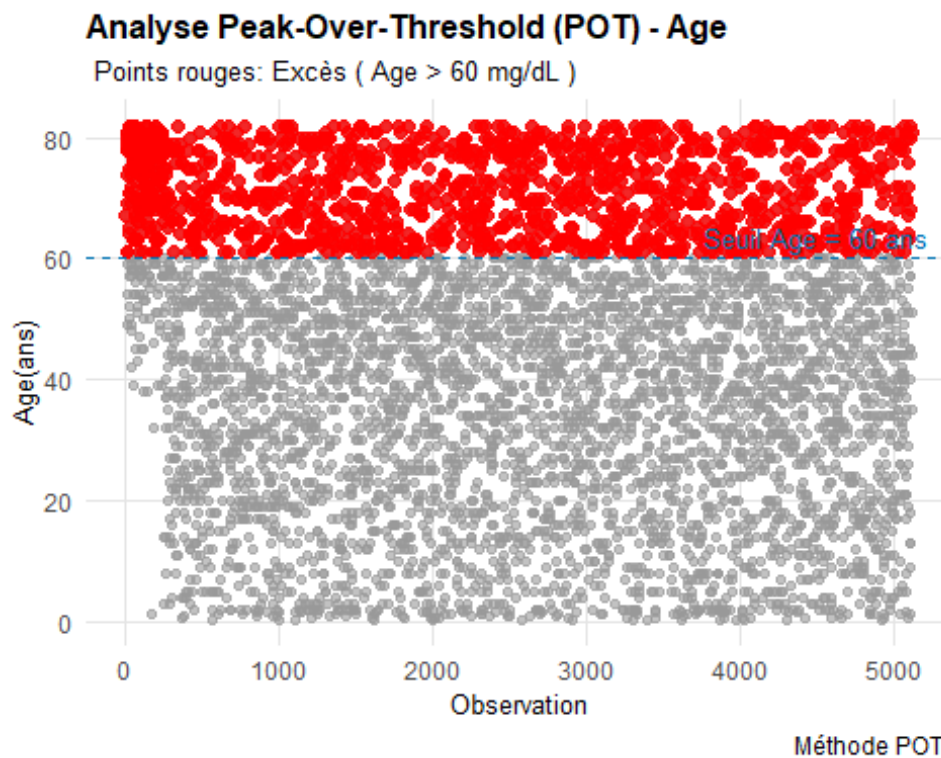
```
donnee <- donnee %>%
  mutate(observation = seq_along(Age))

#Observation vs Age avec seuil à 60 ans
plot_age <- ggplot(donnee, aes(x = observation, y = Age)) +
  geom_point(color = "gray60", alpha = 0.6, size = 1.5) +
  geom_point(data = subset(donnee, Age > 60), color = "red", size = 2, alpha
= 0.8) +
```

```

geom_hline(yintercept = 60, linetype = "dashed", color = "#0072B2", linewidth
th = 0.7) +
  annotate("text", x = Inf, y = 60, label = "Seuil Age = 60 ans",
    hjust = 1.1, vjust = -0.5, color = "#0072B2", size = 3.5) +
  labs(title = "Analyse Peak-Over-Threshold (POT) - Age",
    subtitle = "Points rouges: Excès ( Age > 60 mg/dL )",
    x = "Observation",
    y = "Age(ans)",
    caption = "Méthode POT ") +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 12),
    plot.subtitle = element_text(size = 10),
    axis.title = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90", linewidth = 0.2),
    panel.grid.minor = element_blank()
  )
grid.arrange(plot_age, ncol = 1)

```



```

#Observation vs Taux de glucose moyen avec seuil à 126 mg/dL
plot_glucose <- ggplot(donnee, aes(x = observation, y = Taux_glucose_moyen))
+
  geom_point(color = "gray60", alpha = 0.6, size = 1.5) +
  geom_point(data = subset(donnee, Taux_glucose_moyen > 126), color = "red",
    size = 2, alpha = 0.8) +
  geom_hline(yintercept = 126, linetype = "dashed", color = "#D55E00", linewidth
dth = 0.7) +

```

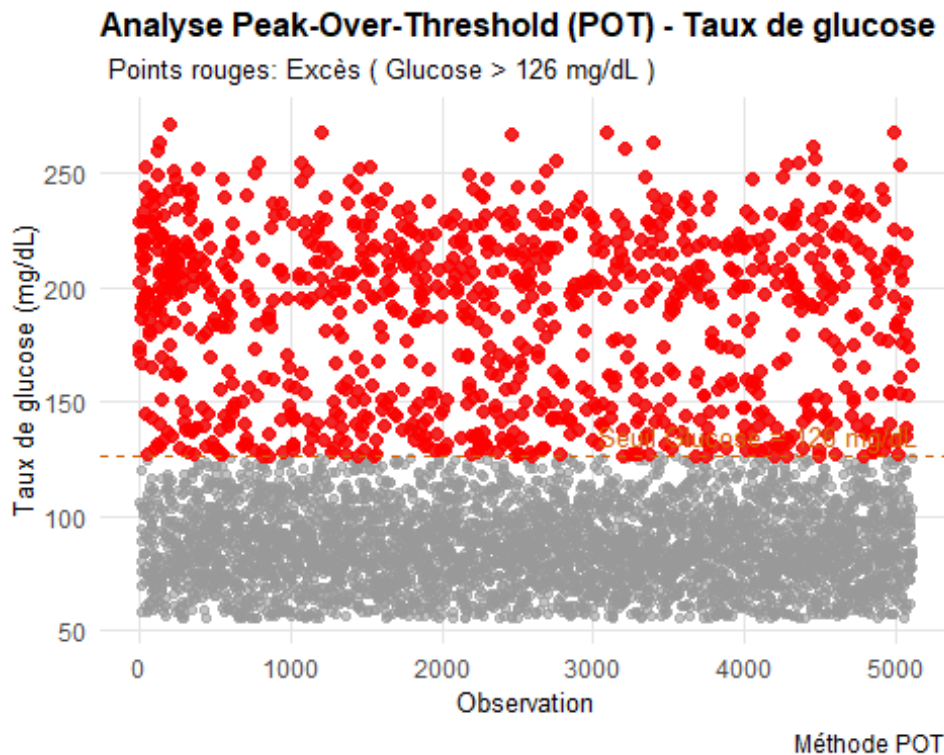
```

annotate("text", x = Inf, y = 126, label = "Seuil Glucose = 126 mg/dL",
        hjust = 1.1, vjust = -0.5, color = "#D55E00", size = 3.5) +
labs(title = "Analyse Peak-Over-Threshold (POT) - Taux de glucose",
     subtitle = " Points rouges: Excès ( Glucose > 126 mg/dL )",
     x = "Observation",
     y = "Taux de glucose (mg/dL)",
     caption = "Méthode POT ") +
theme_minimal() +
theme(
  plot.title = element_text(face = "bold", size = 12),
  plot.subtitle = element_text(size = 10),
  axis.title = element_text(size = 10),
  panel.grid.major = element_line(color = "gray90", linewidth = 0.2),
  panel.grid.minor = element_blank()
)

```

Affichage

```
grid.arrange(plot_glucose, ncol = 1)
```



```

donnee <- donnee %>%
  mutate(observation = seq_along(IMC))

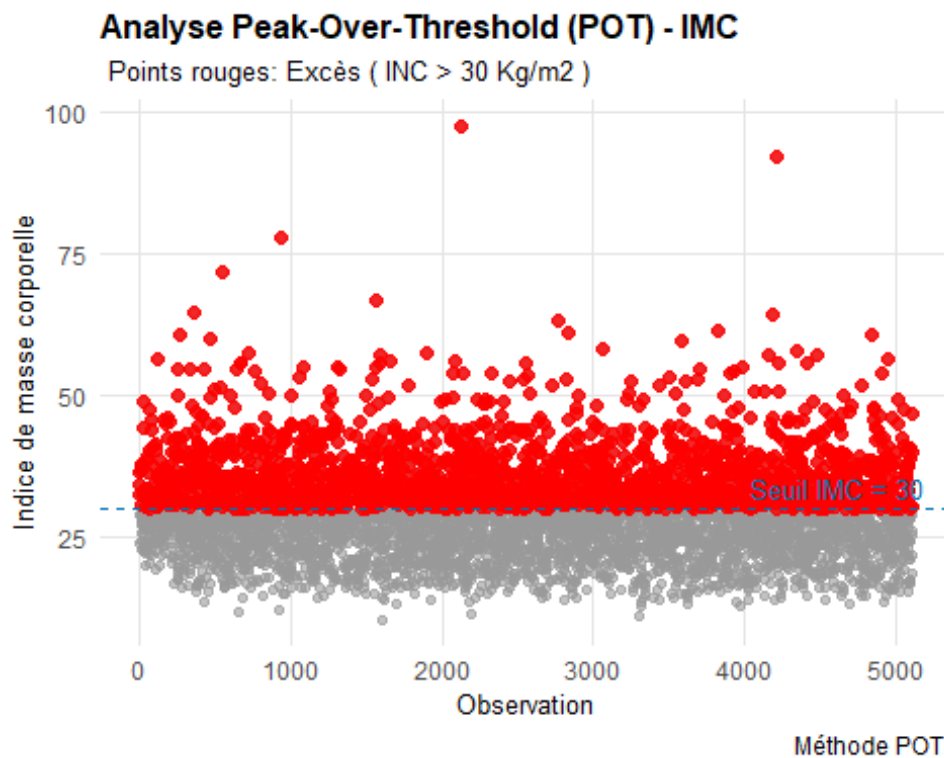
#Observation vs Age avec seuil à 60 ans
plot_imc <- ggplot(donnee, aes(x = observation, y = IMC)) +
  geom_point(color = "gray60", alpha = 0.6, size = 1.5) +
  geom_point(data = subset(donnee, IMC > 30), color = "red", size = 2, alpha
= 0.8) +

```

```

geom_hline(yintercept = 30, linetype = "dashed", color = "#0072B2", linewidth
th = 0.7) +
  annotate("text", x = Inf, y = 30, label = "Seuil IMC = 30 ",
    hjust = 1.1, vjust = -0.5, color = "#0072B2", size = 3.5) +
  labs(title = "Analyse Peak-Over-Threshold (POT) - IMC",
    subtitle = "Points rouges: Excès ( INC > 30 Kg/m2 )",
    x = "Observation",
    y = "Indice de masse corporelle",
    caption = "Méthode POT ") +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 12),
    plot.subtitle = element_text(size = 10),
    axis.title = element_text(size = 10),
    panel.grid.major = element_line(color = "gray90", linewidth = 0.2),
    panel.grid.minor = element_blank()
  )
grid.arrange(plot_imc, ncol = 1)

```



Calcul des excès

```

# Seuils
seuil_age <- 60
seuil_glucose <- 126
seuil_imc <- 30

# Excès

```

```
age_exces <- donnee$Age[donnee$Age >= seuil_age] - seuil_age
glucose_exces <- donnee$Taux_glucose_moyen[donnee$Taux_glucose_moyen >= seuil_glucose] - seuil_glucose
imc_exces <- donnee$IMC[donnee$IMC >= seuil_imc] - seuil_imc
```

Visualisation des distributions des excès d'âge

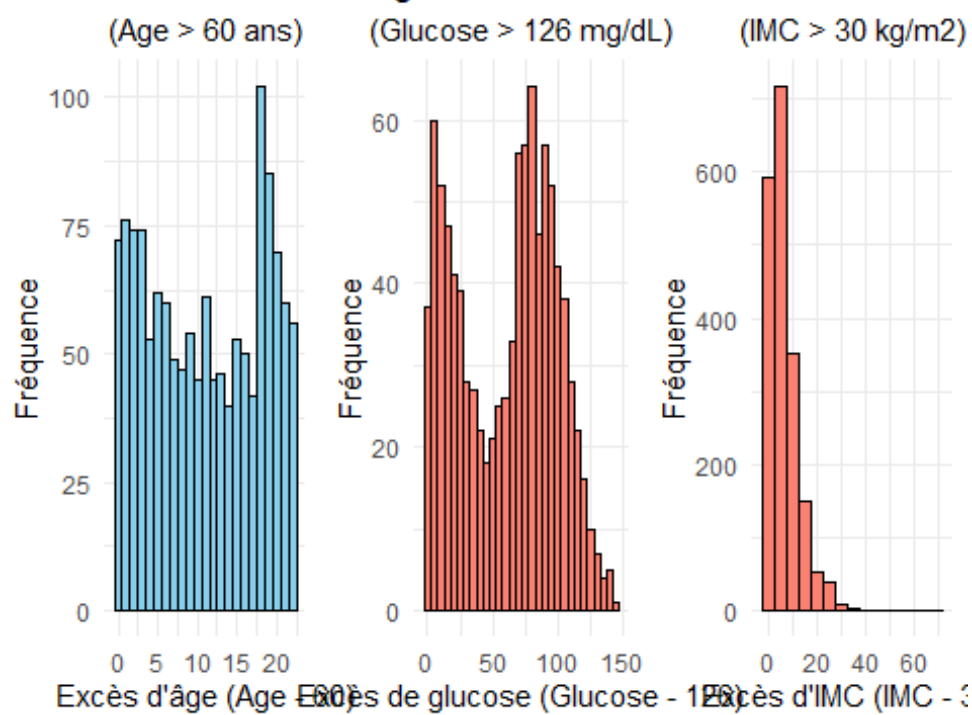
```
plot_dist_exces_age <- ggplot(data.frame(x = age_exces), aes(x = x)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black", na.rm = TRUE) +
  labs(title = "Distribution des Excès d'Âge",
       subtitle = "(Age > 60 ans)",
       x = "Excès d'âge (Age - 60)", y = "Fréquence") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

# Histogramme des excès de glucose avec fréquence
plot_dist_exces_glucose <- ggplot(data.frame(x = glucose_exces), aes(x = x)) +
  geom_histogram(binwidth = 5, fill = "salmon", color = "black", na.rm = TRUE) +
  labs(title = "Distribution des Excès de Glucose",
       subtitle = "(Glucose > 126 mg/dL)",
       x = "Excès de glucose (Glucose - 126)", y = "Fréquence") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

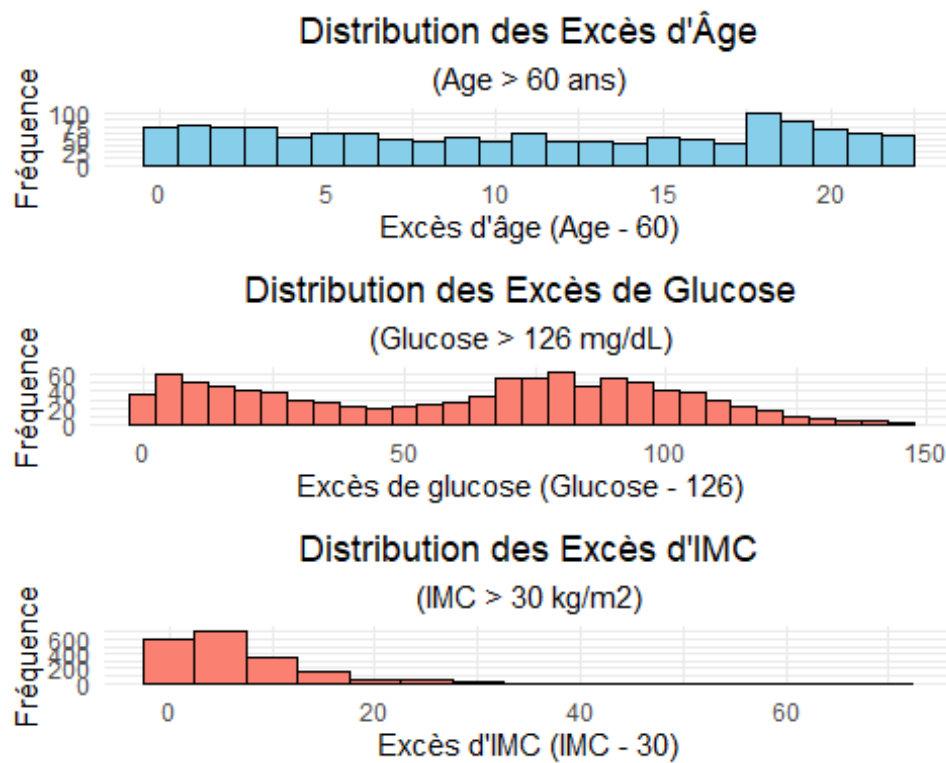
# Histogramme des excès de IMC avec fréquence
plot_dist_exces_imc <- ggplot(data.frame(x = imc_exces), aes(x = x)) +
  geom_histogram(binwidth = 5, fill = "salmon", color = "black", na.rm = TRUE) +
  labs(title = "Distribution des Excès d'IMC",
       subtitle = "(IMC > 30 kg/m2)",
       x = "Excès d'IMC (IMC - 30)", y = "Fréquence") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

# Affichage côte à côte
grid.arrange(plot_dist_exces_age, plot_dist_exces_glucose, plot_dist_exces_imc,
             ncol = 3)
```

Distribution des Excès d'Âge, de Glucose et d'IMC



```
grid.arrange(plot_dist_excès_age, plot_dist_excès_glucose, plot_dist_excès_imc, ncol = 1)
```



Distribution des excès d'âge

On constate probablement une asymétrie à droite. Cela reflète une queue modérément lourde : les valeurs très élevées d'âge sont moins fréquentes, mais existent. Les excès sont concentrés entre 15 et 20 ans, avec un pic après 15, suggérant que la majorité des individus à risque extrême ont entre 75 et 80 ans

Distribution des excès de glucose

On constate aussi que la forme de distribution pour la glucose est plus étalée que celle de l'âge, avec un pic vers 50 mg/dL d'excès, mais des observations allant jusqu'à 150.

Ajustement d'un modèle avec la méthode POT sur le variable Age

```
# Ajustement GPD pour Les excès d'âge (methode MLE)
modele_age <- fevd(age_excès, threshold = 0, type = "GP", method = "MLE")

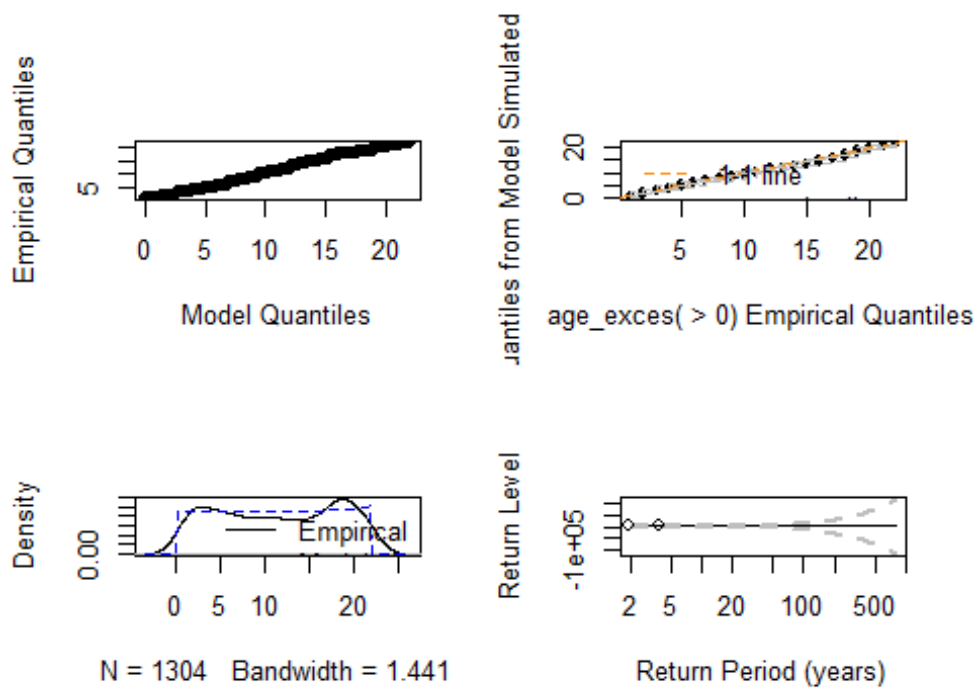
## Warning in log(z): Production de NaN
## Warning in log(z): Production de NaN

summary(modele_age)

##
## fevd(x = age_excès, threshold = 0, type = "GP", method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
##
## Negative Log-Likelihood Value: 3976.847
##
##
## Estimated parameters:
## scale shape
## 22.706579 -1.032117
##
## Standard Error Estimates:
## scale shape
## 2e-08 2e-08
##
## Estimated parameter covariance matrix.
## scale shape
## scale 4.000002e-16 -1.113724e-26
## shape -1.113724e-26 4.000002e-16
##
## AIC = 7957.695
##
## BIC = 7968.041

plot(modele_age)
```

```
fevd(x = age_exces, threshold = 0, type = "GP", method = "MLE")
```



```
cat("=====\n")
## =====
## =====

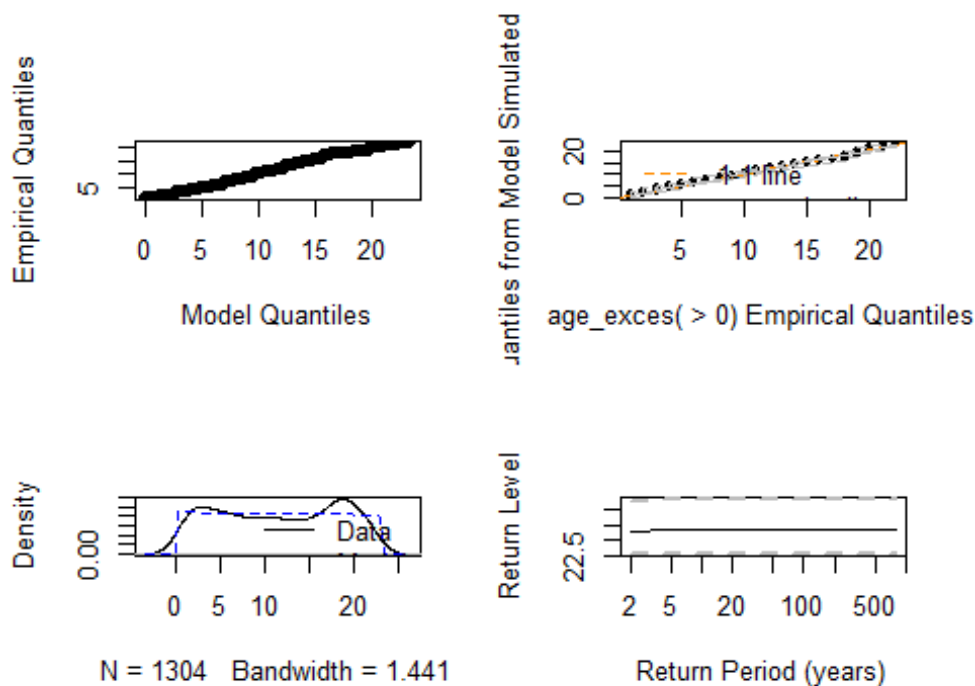
# Ajustement GPD (methode L-moment)
modele_age_Lmom <- fevd(age_exces, threshold = 0, type = "GP", method = "Lmoments")
summary(modele_age_Lmom)

## fevd(x = age_exces, threshold = 0, type = "GP", method = "Lmoments")
## [1] "GP Fitted to age_exces using L-moments estimation."
##      scale      shape
## 22.7091729 -0.9736578

plot(modele_age_Lmom)
```



```
evd(x = age_exces, threshold = 0, type = "GP", method = "Lmoments")
```



Analyse des résultats

Le paramètre shape estimé par MLE est proche de -1 (negatif), avec des erreurs standards très faibles, ce qui peut signifier une distribution bornée et une instabilité de l'estimation.

la GPD estime qu'il y a une borne supérieure, donc un âge au-delà duquel aucun excès n'existe dans les données

Avec la méthode des L-moments, le paramètre shape est -0,97 (negatif) mais l'estimation plus stable, ce qui confirme aussi une distribution à queue bornée,

densité GPD sur Age

```
# Histogramme des excès d'âge
hist(age_exces, prob = TRUE, breaks = 15,
     main = "Ajustement GPD aux excès d'âge : MLE vs L-moments",
     xlab = "Excès d'âge (Age - 60)",
     col = "lightblue", border = "white")

# Courbe GPD ajustée par L-moments
curve(dgpd(x, loc = 0, scale = 22.71, shape = -0.97),
     from = 0, to = max(age_exces),
     add = TRUE, col = "red", lwd = 2)

# Courbe GPD ajustée par MLE
curve(dgpd(x, loc = 0, scale = 22.71, shape = -1.03),
     from = 0, to = max(age_exces),
```

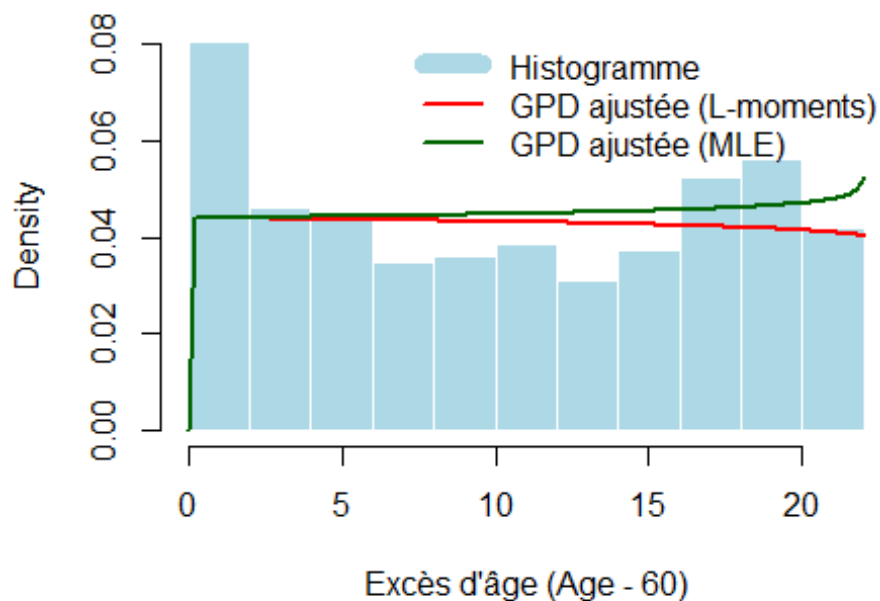
```

add = TRUE, col = "darkgreen", lwd = 2)

# Légende des deux courbes + histogramme
legend("topright",
      legend = c("Histogramme", "GPD ajustée (L-moments)", "GPD ajustée (MLE)"),
      col = c("lightblue", "red", "darkgreen"),
      lwd = c(10, 2, 2), # épaisseur différente pour l'histogramme
      bty = "n")

```

Ajustement GPD aux excès d'âge : MLE vs L-mome



Methode L-moment

La courbe est plus plate et colle mieux à l'histogramme (plus de stabilité).

Elle suppose une queue bornée mais moins fortement ($\text{shape} = -0.97$)

Methode MLE

La courbe monte légèrement dans les extrêmes indique que le modèle suppose une queue plus lourde.

La densité estimée augmente légèrement en fin de distribution : cela correspond au shape négatif mais plus proche de 0 ($\text{shape} = -1.03$). Conclusion : l'ajustement par L-moments semble plus stable et cohérent avec les données, bien que les deux donnent une forme bornée pour la queue de la distribution des excès d'âge.

fonction de répartition cumulée (CDF)

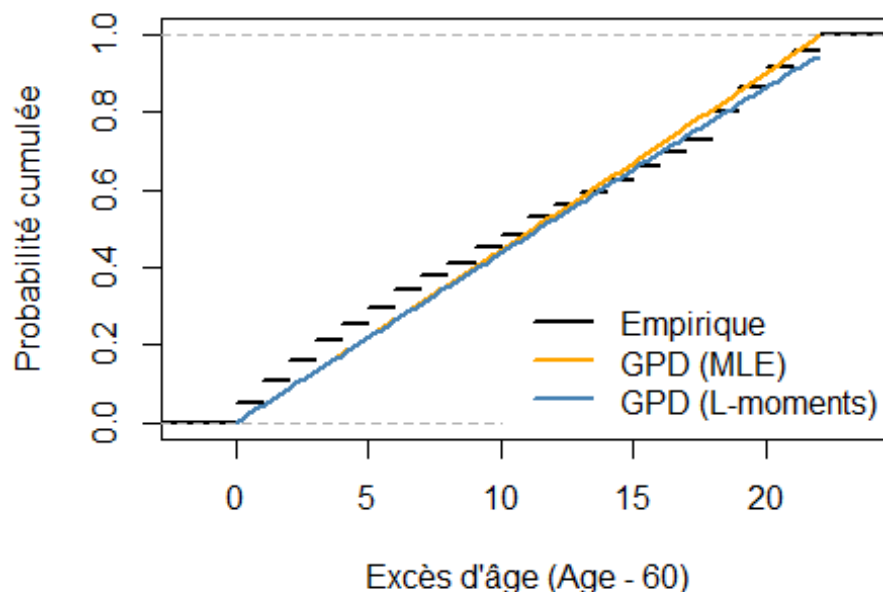
```
plot(ecdf(age_exces),
     main = "CDF des excès d'âge : Empirique vs GPD (MLE & L-moments)",
     xlab = "Excès d'âge (Age - 60)",
     ylab = "Probabilité cumulée",
     col = "black", lwd = 2,
     do.points = FALSE)

# Ajouter la courbe GPD (MLE)
curve(pgpd(x, loc = 0, scale = 22.71, shape = -1.03),
      from = 0, to = max(age_exces),
      add = TRUE, col = "orange", lwd = 2)

# Ajouter la courbe GPD (L-moments)
curve(pgpd(x, loc = 0, scale = 22.71, shape = -0.97),
      from = 0, to = max(age_exces),
      add = TRUE, col = "steelblue", lwd = 2)

# La légende
legend("bottomright",
      legend = c("Empirique", "GPD (MLE)", "GPD (L-moments)"),
      col = c("black", "orange", "steelblue"),
      lwd = 2,
      bg = "white",
      box.lty = 0)
```

CDF des excès d'âge : Empirique vs GPD (MLE & L-mo

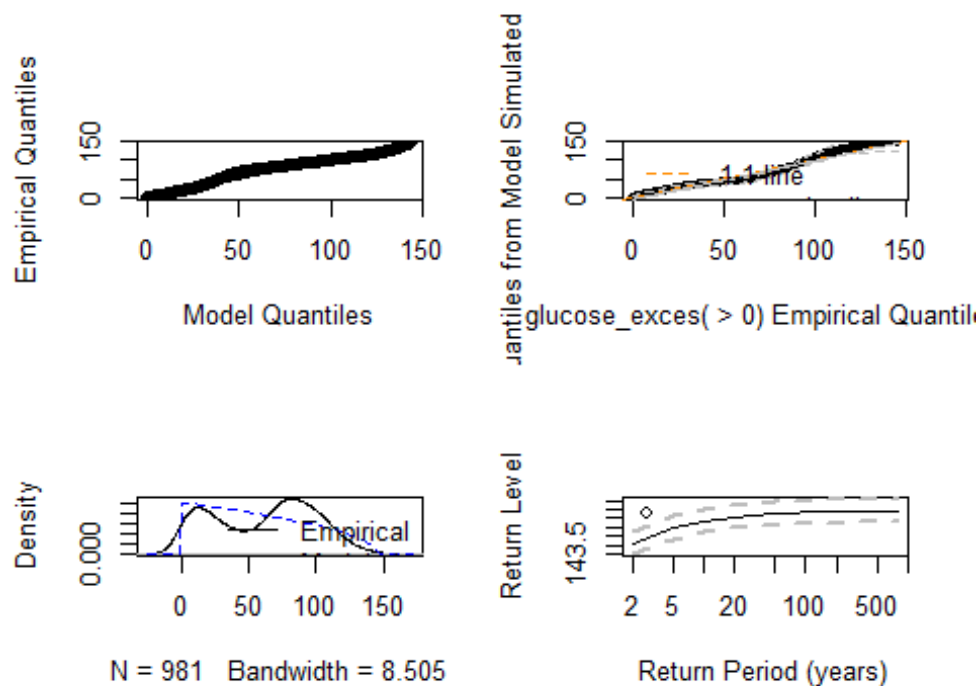


Ajustement d'un modèle avec la methode POT sur le variable glucose

```
# Ajustement GPD pour les excès de glucose (methode des MLE)
modele_glucose <- fevd(glucose_exces, threshold = 0,,type = "GP", method = "MLE")
summary(modele_glucose)

##
## fevd(x = glucose_exces, threshold = 0, type = "GP", method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
## Negative Log-Likelihood Value: 4816.235
##
## Estimated parameters:
##      scale      shape
## 96.7727478 -0.6629417
##
## Standard Error Estimates:
##      scale      shape
## 0.21592004 0.00000002
##
## Estimated parameter covariance matrix.
##           scale      shape
## scale 4.662146e-02 -4.640859e-13
## shape -4.640859e-13 4.000002e-16
##
## AIC = 9636.47
##
## BIC = 9646.247
plot(modele_glucose)
```

```
fevd(x = glucose_exces, threshold = 0, type = "GP", method = "MLE")
```



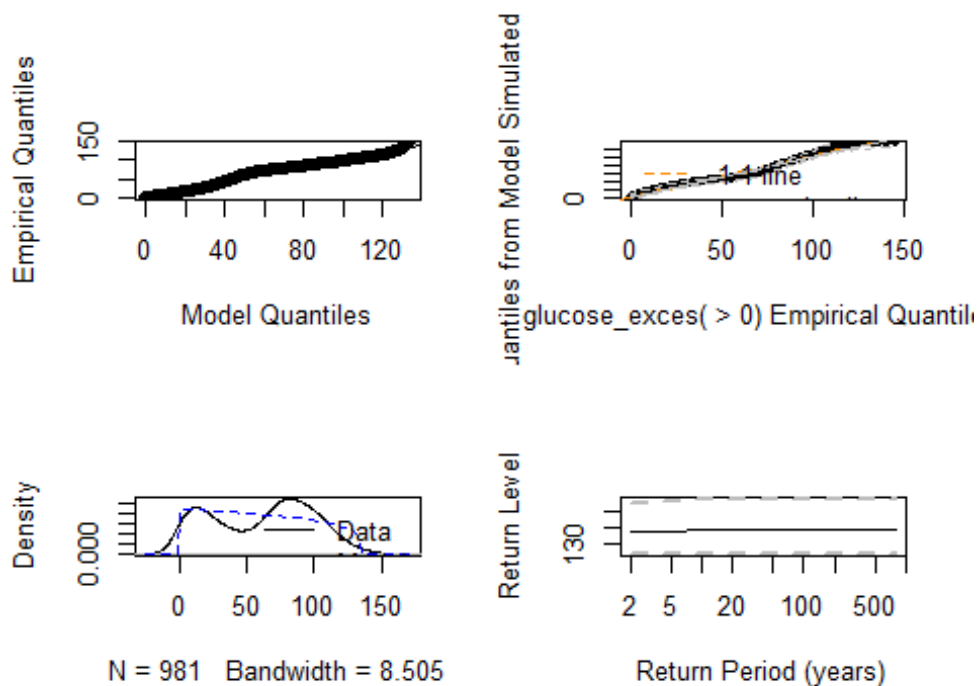
```
cat("=====\n")
## =====
## =====

# Ajustement GPD pour Les excès de glucose (methode des moments)
modele_glucose_lmom <- fevd(glucose_exces, threshold = 0, type = "GP", method
= "Lmoments")
summary(modele_glucose_lmom)

## fevd(x = glucose_exces, threshold = 0, type = "GP", method = "Lmoments")
## [1] "GP Fitted to glucose_exces using L-moments estimation."
##      scale      shape
## 110.85350  -0.82598

plot(modele_glucose_lmom)
```

```
d(x = glucose_exces, threshold = 0, type = "GP", method = "Lmoment
```



Le paramètre shape négatif dans les deux cas confirme une queue finie et modérément courte pour les excès de glucose. La méthode des L-moments donne un paramètre shape plus négatif (≈ -0.83 vs -0.66), suggérant une queue encore plus courte. Cela signifie que les excès de glucose ont une valeur maximale théorique finie.

La comparaison entre MLE et L-moments montre que cette dernière méthode est préférable.

densité GPD pour taux de glucose

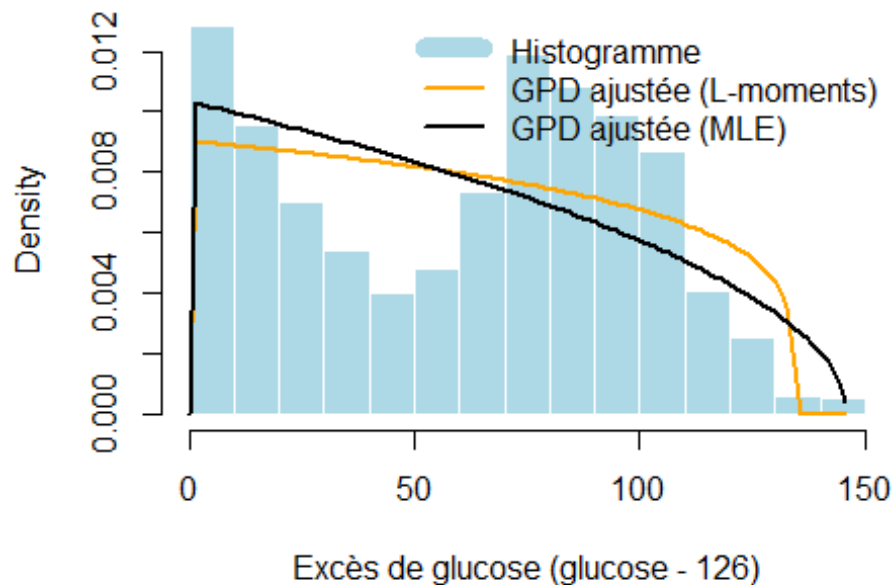
```
# Histogramme des excès de glucose
hist(glucose_exces, prob = TRUE, breaks = 15,
     main = "Ajustement GPD aux excès de glucose : MLE vs L-moments",
     xlab = "Excès de glucose (glucose - 126)",
     col = "lightblue", border = "white")

# Courbe GPD - L-moments (en rouge)
curve(dgpd(x, loc = 0, scale = 110.85350, shape = -0.82598),
      from = 0, to = max(glucose_exces),
      add = TRUE, col = "orange", lwd = 2)

# Courbe GPD - MLE (en vert foncé)
curve(dgpd(x, loc = 0, scale = 96.77275, shape = -0.66294),
      from = 0, to = max(glucose_exces),
      add = TRUE, col = "black", lwd = 2)
```

```
# Légende
legend("topright",
      legend = c("Histogramme", "GPD ajustée (L-moments)", "GPD ajustée (MLE)"),
      col = c("lightblue", "orange", "black"),
      lwd = c(10, 2, 2), bty = "n")
```

Ajustement GPD aux excès de glucose : MLE vs L-moments



fonction de répartition cumulée (CDF)

```
# Courbe ECDF des excès de glucose
plot(ecdf(glucose_exces),
     main = "CDF des excès de glucose : Empirique vs GPD (MLE & L-moments)",
     xlab = "Excès de glucose (glucose - 126)",
     ylab = "Probabilité cumulée",
     col = "black", lwd = 2)

# GPD - L-moments (rouge)
curve(pgpdp(x, loc = 0, scale = 110.85350, shape = -0.82598),
      from = 0, to = max(glucose_exces),
      add = TRUE, col = "orange", lwd = 2)

# GPD - MLE (vert foncé)
curve(pgpdp(x, loc = 0, scale = 96.77275, shape = -0.66294),
      from = 0, to = max(glucose_exces),
      add = TRUE, col = "red", lwd = 2)

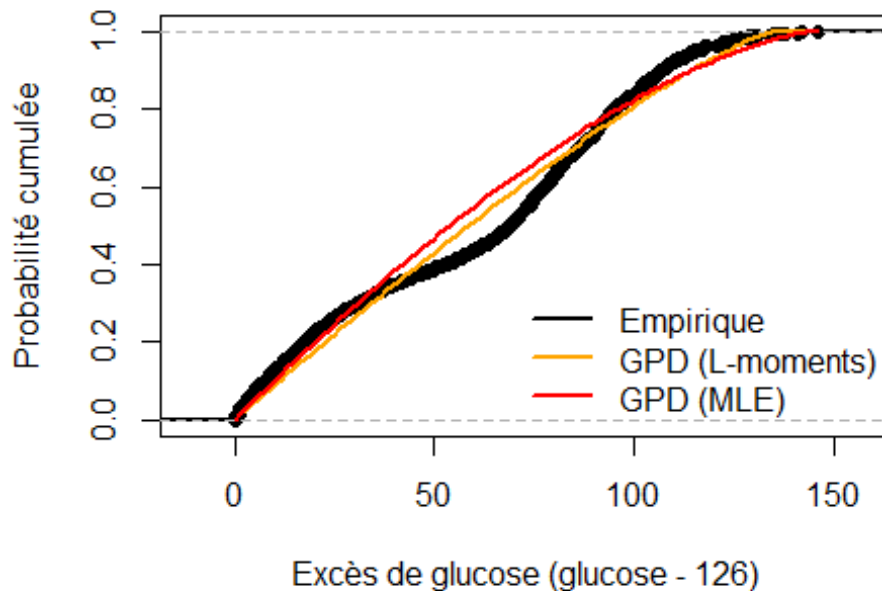
# Légende
```

```

legend("bottomright",
      legend = c("Empirique", "GPD (L-moments)", "GPD (MLE)"),
      col = c("black", "orange", "red"),
      lwd = 2, bty = "n")

```

les excès de glucose : Empirique vs GPD (MLE & L-



```

# Ajustement GPD pour les excès de glucose (methode des MLE)
modele_imc <- fevd(imc_exces, threshold = 0,,type = "GP", method = "MLE")
summary(modele_imc)

##
## fevd(x = imc_exces, threshold = 0, type = "GP", method = "MLE")
##
## [1] "Estimation Method used: MLE"
##
##
## Negative Log-Likelihood Value: 5449.55
##
##
## Estimated parameters:
##      scale      shape
## 6.75813196 -0.03195971
##
## Standard Error Estimates:
##      scale      shape
## 0.19479051 0.01740774
##
## Estimated parameter covariance matrix.

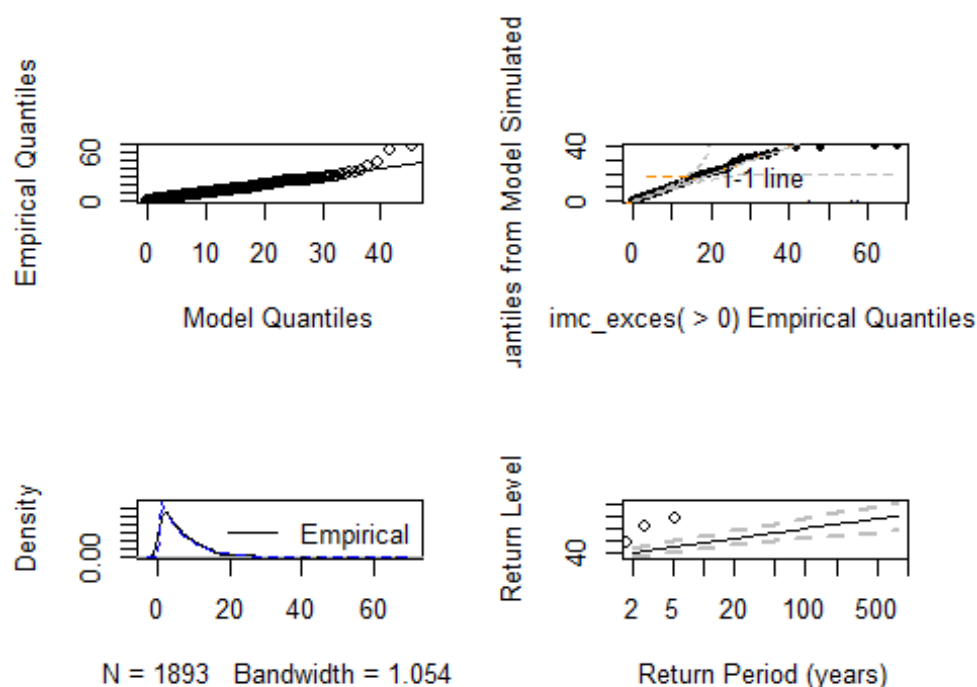
```



```
##          scale          shape
## scale  0.037943344 -0.0021584428
## shape -0.002158443  0.0003030294
##
## AIC = 10903.1
##
## BIC = 10914.19
```

```
plot(modele_imc)
```

```
fevd(x = imc_exces, threshold = 0, type = "GP", method = "MLE")
```



```
cat("=====\n")

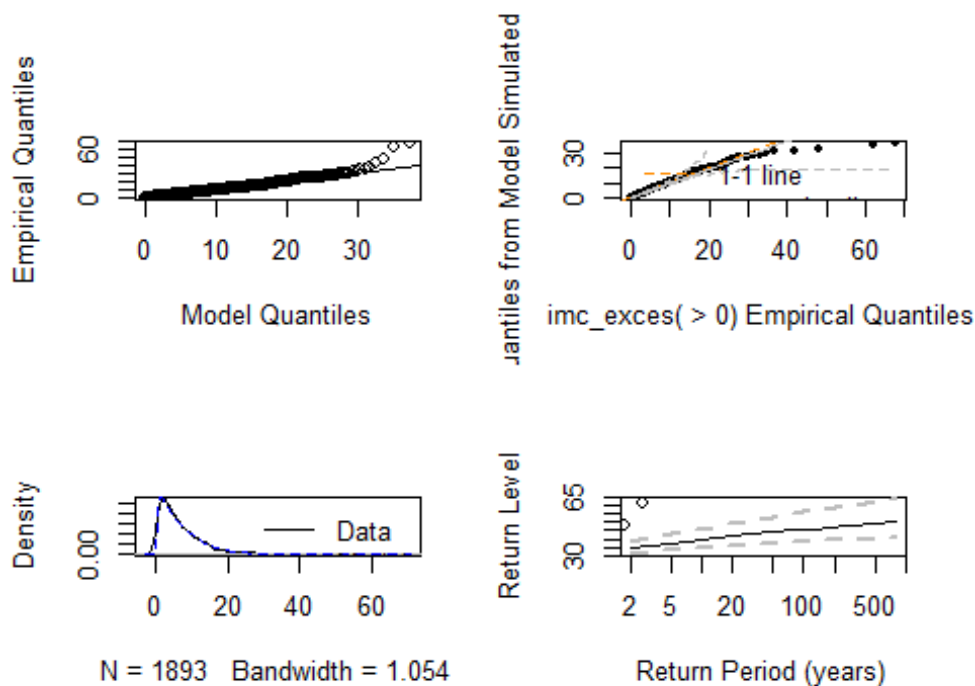
## =====
##

# Ajustement GPD pour les excès de glucose (methode des moments)
modele_imc_lmom <- fevd(imc_exces, threshold = 0, type = "GP", method = "Lmoments")
summary(modele_imc_lmom)

## fevd(x = imc_exces, threshold = 0, type = "GP", method = "Lmoments")
## [1] "GP Fitted to imc_exces using L-moments estimation."
##          scale          shape
## 7.252715 -0.107191

plot(modele_imc_lmom)
```

```
evd(x = imc_exces, threshold = 0, type = "GP", method = "Lmoments")
```



MLE donne un shape -0.032 très proche de 0 \rightarrow distribution presque exponentielle.

L-moments donne un shape -0.107 ce qui indique queue un peu plus courte (moins d'observations extrêmes prévues).

cela signifie que les valeurs extrêmes d'IMC sont peu fréquentes, mais il y a encore un risque modéré d'observer des excès.

densité GPD pour taux de glucose

```
# Histogramme des excès de imc
hist(imc_exces, prob = TRUE, breaks = 15,
     main = "Ajustement GPD aux excès d'IMC : MLE vs L-moments",
     xlab = "Excès de glucose (IMC - 30)",
     col = "lightblue", border = "white")

# Courbe GPD - L-moments
curve(dgpd(x, loc = 0, scale = 7.252715, shape = -0.107191),
      from = 0, to = max(imc_exces),
      add = TRUE, col = "orange", lwd = 2)

# Courbe GPD - MLE (en vert foncé)
curve(dgpd(x, loc = 0, scale = 6.75813196, shape = -0.03195971),
      from = 0, to = max(glucose_exces),
      add = TRUE, col = "black", lwd = 2)

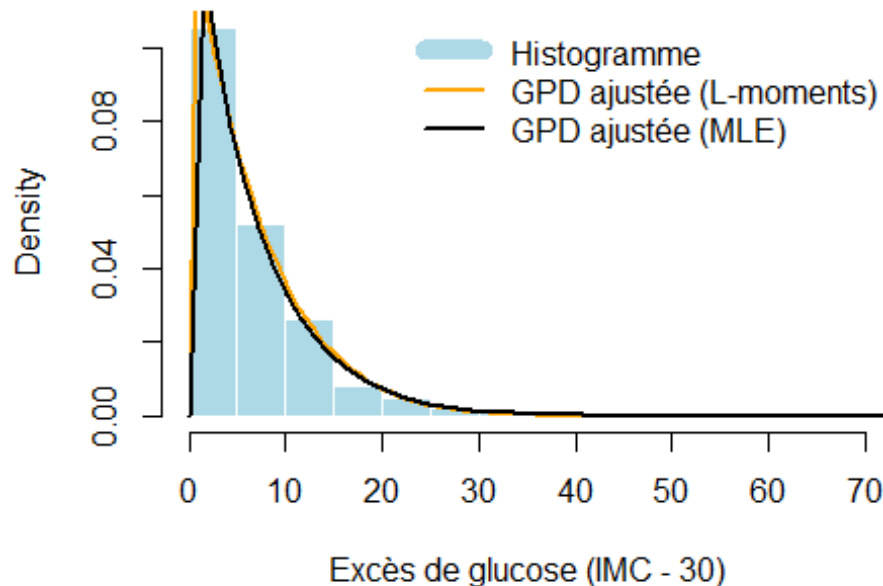
# Légende
```

```

legend("topright",
      legend = c("Histogramme", "GPD ajustée (L-moments)", "GPD ajustée (MLE)"),
      col = c("lightblue", "orange", "black"),
      lwd = c(10, 2, 2), bty = "n")

```

Ajustement GPD aux excès d'IMC : MLE vs L-mome



fonction de répartition cumulée (CDF)

```

# Courbe ECDF des excès d'imc
plot(ecdf(imc_exces),
     main = "CDF des excès d'imc : Empirique vs GPD (MLE & L-moments)",
     xlab = "Excès d'IMC (IMC - 30)",
     ylab = "Probabilité cumulée",
     col = "black", lwd = 2)

# GPD - L-moments (rouge)
curve(pgpd(x, loc = 0, scale = 7.252715, shape = -0.107191),
      from = 0, to = max(imc_exces),
      add = TRUE, col = "orange", lwd = 2)

# GPD - MLE (vert foncé)
curve(pgpd(x, loc = 0, scale = 6.75813196, shape = -0.03195971),
      from = 0, to = max(imc_exces),
      add = TRUE, col = "red", lwd = 2)

# Légende
legend("bottomright",

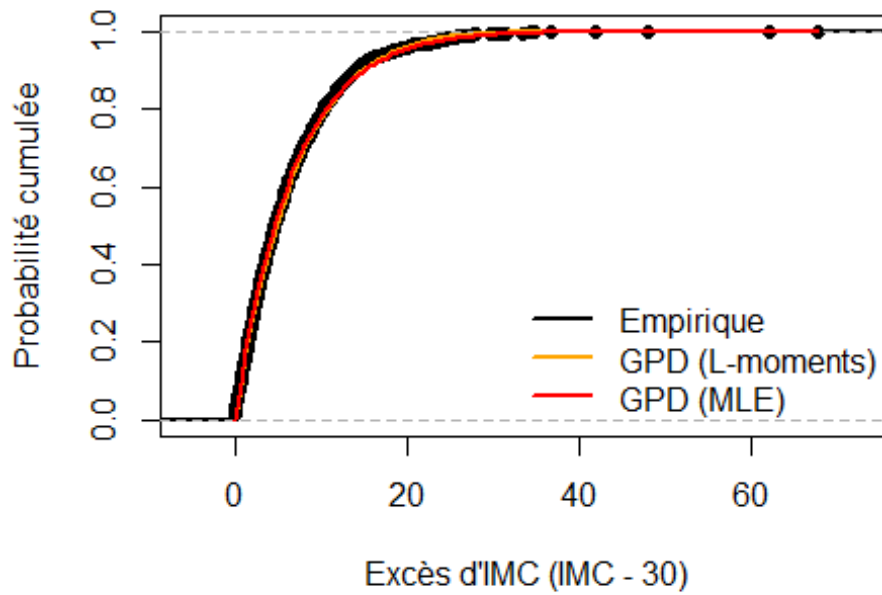
```

```

legend = c("Empirique", "GPD (L-moments)", "GPD (MLE)"),
col = c("black", "orange", "red"),
lwd = 2, bty = "n")

```

IF des excès d'imc : Empirique vs GPD (MLE & L-mo



Calcul de probabilité de dépassement

```

# Probabilité de dépassement pour age
prob_dep_age <- 1 - pgpd(age_exces, loc = 0, scale = 22.7091729 , shape = -0.9736578 )
summary(prob_dep_age)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.05251 0.21931 0.51933 0.52608 0.82429 1.00000

# Probabilité de dépassement pour glucose
prob_dep_glu <- 1 - pgpd(glucose_exces, loc = 0, scale = 110.85350 , shape = -0.82598)
summary(prob_dep_glu)

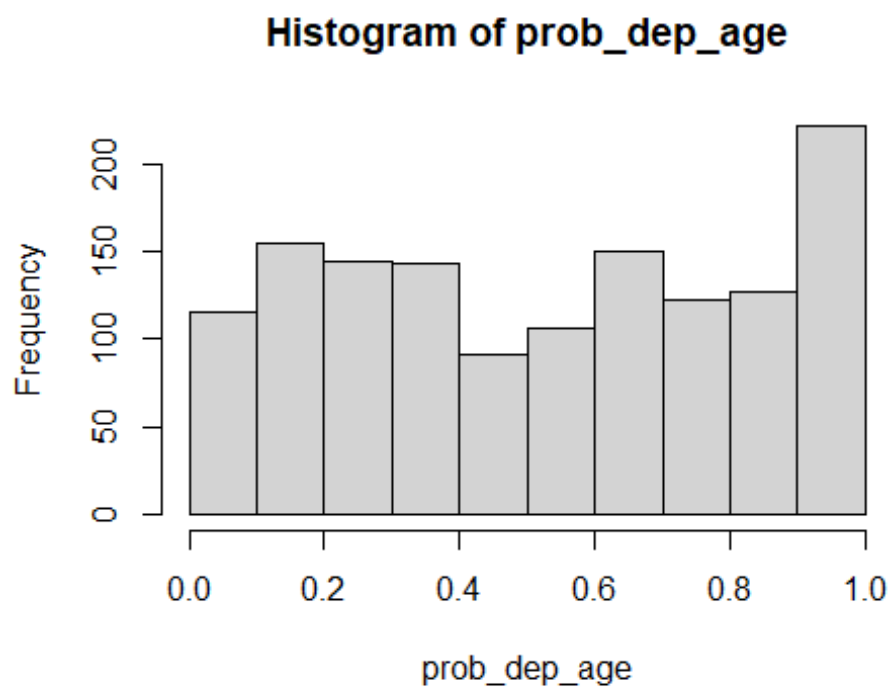
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.00000 0.25400 0.41740 0.49960 0.79060 0.99960

# Probabilité de dépassement pour imc
prob_dep_imc <- 1 - pgpd(imc_exces, loc = 0, scale = 7.252715 , shape = -0.107191)
summary(prob_dep_imc)

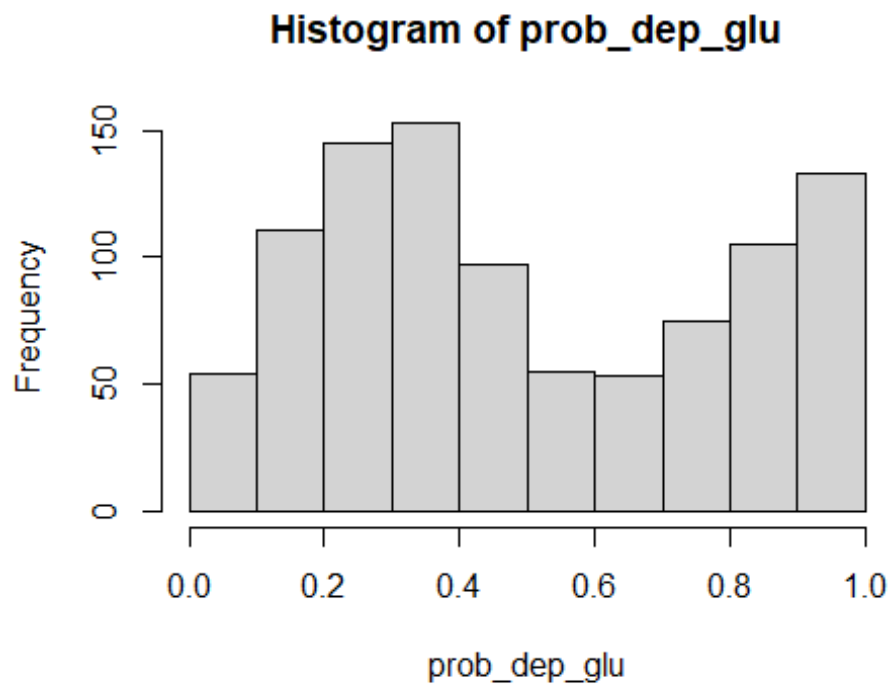
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.2599  0.5109  0.5075  0.7478  1.0000
```

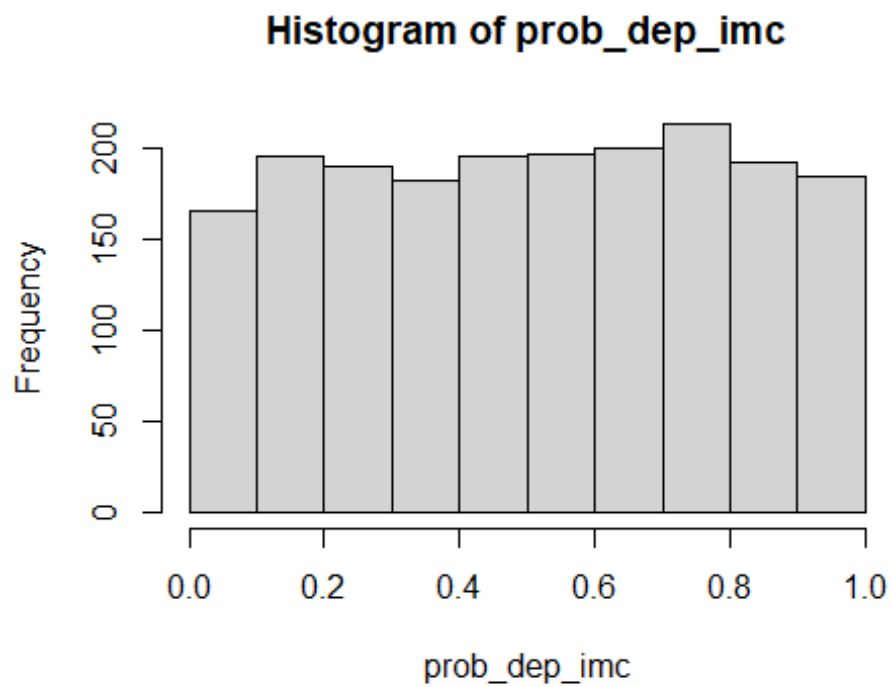
```
hist(prob_dep_age)
```



```
hist(prob_dep_glu)
```



```
hist(prob_dep_imc)
```



L'analyse des probabilités de dépassement des excès d'âge (seuil 60 ans) montre que plus de 50 % des observations ont une probabilité supérieure à 50 % d'être encore dépassées. Cela suggère une distribution relativement étalée des âges extrêmes, cohérente avec une variabilité importante dans la population âgée. En revanche, les excès de glucose (seuil 126 mg/dL) présentent une queue légèrement moins étendue : les valeurs extrêmes sont plus rares mais existent, comme en témoignent les valeurs minimales de probabilité quasi nulles (0.0000), indiquant des cas rares de glycémies très élevées.

Calcul de quantiles extrêmes

```
# Probabilités
probs <- c(0.90, 0.95, 0.99)

# Quantiles des excès
q_age_exc      <- qgpd(probs, loc = 0, scale = 22.70917, shape = -0.97366)
q_glucose_exc  <- qgpd(probs, loc = 0, scale = 110.85350, shape = -0.82598)
q_imc_exc      <- qgpd(probs, loc = 0, scale = 7.25271, shape = -0.10719)

# Ajout du seuil
q_age      <- seuil_age + q_age_exc
q_glucose  <- seuil_glucose + q_glucose_exc
q_imc      <- seuil_imc + q_imc_exc

# Affichage dans un tableau
quant_table_lmom <- data.frame(
  Quantile = paste0(probs * 100, "%"),
  Age = round(q_age, 2),
  Glucose = round(q_glucose, 2),
  IMC = round(q_imc, 2)
)

print(quant_table_lmom)

##   Quantile   Age Glucose  IMC
## 1     90% 80.85  240.17 44.80
## 2     95% 82.06  248.91 48.58
## 3     99% 83.06  257.22 56.36
```

Les quantiles extrêmes calculés à partir de la loi de Pareto généralisée ajustée par les L-moments indiquent que seuls 1 % des individus à risque dépassent un âge de 83 ans, une glycémie de 257 et un IMC de 56,36 . Ces valeurs représentent des seuils de référence au-delà desquels le risque d'AVC peut être considéré comme particulièrement élevé, notamment dans une approche de prévention ciblée sur les cas les plus extrêmes.