

LOIS À DENSITÉ : LOIS EXPONENTIELLES

MODÉLISATION DE LA DURÉE ENTRE DEUX ÉRUPTIONS SUCCESSIVES AU PITON DE LA FOURNAISE

1° PROTOCOLE D'OBSERVATION ET DE RECUEIL DES DONNÉES

Données de 1985 à 2012, obtenues par Alain Busser auprès de l'observatoire volcanologique de La Réunion.

2° LES LIGNES DE COMMANDES R DES ÉTAPES D'UN PREMIER TRAITEMENT

```
# IMPORTATION DU FICHIER CSV. LE FORMAT D'ARRIVÉE EST UN OBJET "data.frame"
setwd("T:/DocuTPHR/MathProbaStat/CasConcrets/EruptionsA")
EruptionA <- read.csv("EruptDuree.csv", sep = ";")

# ON VÉRIFIE QUE TOUT S'EST BIEN PASSÉ
str(EruptionA)
'data.frame': 104 obs. of 2 variables:
 $ date : Factor w/ 94 levels "1985-12-29","1986-03-19",...: 1 2 2 3 4 5 5 6 7 8 ...
 $ duree: int 28 80 0 4 6 106 0 122 14 10 ...

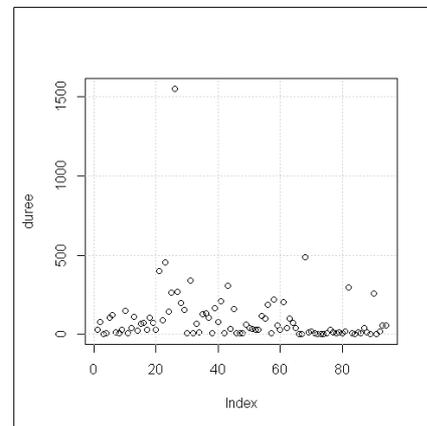
# ON CRÉE UN NOUVEAU data.frame SANS LES DURÉES 0
EruptA <- EruptionA[EruptionA$duree != 0, ]

# ON VÉRIFIE QUE TOUT S'EST BIEN PASSÉ
str(EruptA)
'data.frame': 94 obs. of 2 variables:
 $ date : Factor w/ 94 levels "1985-12-29","1986-03-19",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ duree: int 28 80 4 6 106 122 14 10 31 147 ...

# ON "ATTACHE" L'OBJET EruptA pour ne pas avoir à répéter son nom
attach(EruptA)
```

```
# ON PRÉPARE LES CLASSES POUR CONSTRUIRE L'HISTOGRAMME DE
# LA DISTRIBUTION OBSERVÉE DES DURÉES ENTRE DEUX ÉRUPTIONS SUCCESSIVES
```

```
summary(EruptA) # Pour avoir l'étendue des durée
plot(duree)     # Pour avoir une idée de la
                distribution des durées
grid()
```



```
      date      duree
1985-12-29: 1   Min.   : 1.00
1986-03-19: 1   1st Qu.:  8.25
1986-03-23: 1   Median : 38.00
1986-03-29: 1   Mean    : 97.81
1986-07-13: 1   3rd Qu.: 113.75
1986-11-12: 1   Max.    :1552.00
(Other)      :88
```

```
# Dans une loi exponentielle, l'espérance = l'écart
type, on regarde ce que valent les paramètres
d'échantillon :
```

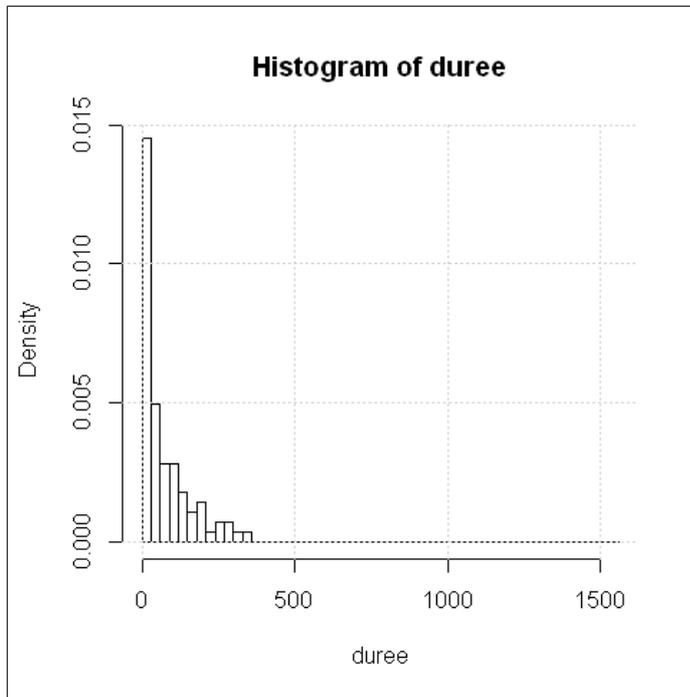
```
mean(EruptA$duree)
[1] 97.80851
```

```
sd(EruptA$duree)
[1] 184
```

```
# Ça ne plaide pas en faveur d'un modèle exponentiel !
```

```
# ON CONSTRUIT LES CLASSES EN FONCTION DE CETTE DISTRIBUTION
# DES CLASSES D'ÉTENDUE 30 DE 0 À 360 PUIS UNE SEULE CLASSE DE 360 À 1560
subdiv1 <- c(seq(0, 360, 30), 1560)
# subdiv2 <- c(seq(0, 366, 30.5), 1555.5)
subdiv1
[1] 0 30 60 90 120 150 180 210 240 270 300 330 360 1560
```

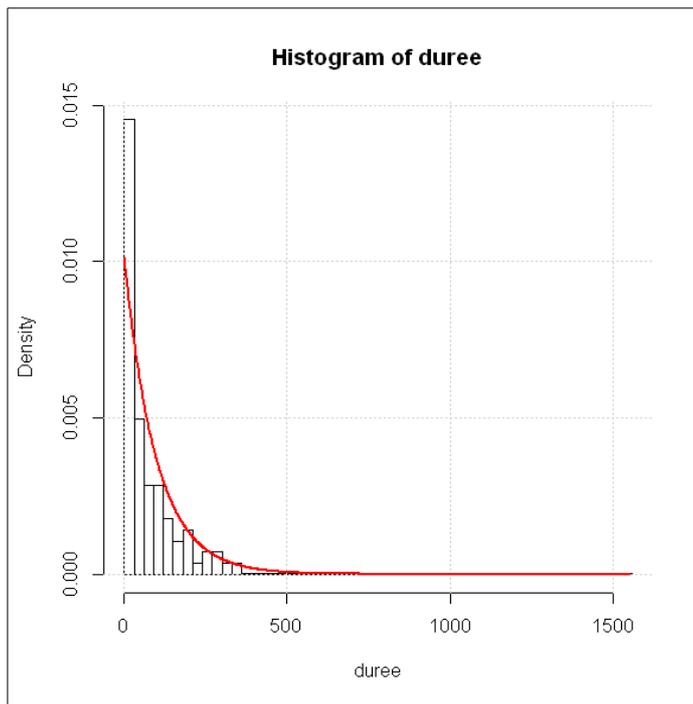
```
# ON CONSTRUIT L'HISTOGRAMME
# freq = FALSE pour avoir les densités en ordonnée
# right = FALSE pour avoir des classes du type [a ; b[
histo <- hist(duree, breaks = subdiv1, freq = FALSE, right = FALSE)
grid()
```



```
# L'OBJET histo CONTIENT, ENTRES AUTRES, LES DEUX LISTES SUIVANTES
histo$breaks # Les bornes des classes utilisées dans notre exemple c'est biensûr subdiv1.
[1] 0 30 60 90 120 150 180 210 240 270 300 330 360 1560

histo$counts # Les effectifs observés des classes, sur la série étudiée
[1] 41 14 8 8 5 3 4 1 2 2 1 1 1 4
```

```
# ON AJUSTE LA DISTRIBUTION THÉORIQUE EXPONENTIELLE PRISE COMME MODÈLE
lambda <- 1 / mean(duree)
abscissesX <- seq(min(duree), max(duree), length = 1000)
lines(abscissesX, dexp(abscissesX, lambda), col = "red", lwd = 2)
```



```
# CALCUL DES EFFECTIFS THÉORIQUES POUR LE KHI-DEUX
# pexp(x, lambda) calcule P(X < x)
OrdreSubdiv1 <- length(subdiv1)
EffTheo <- NULL
for (i in 1:(OrdreSubdiv1 - 1)) {
  EffTheo <- c(EffTheo,
    (pexp(histo$breaks[i + 1], lambda) - pexp(histo$breaks[i], lambda)) *
    length(duree))
}
```

```
EffTheo # Affichage des résultats pour contrôle
[1] 24.8296013 18.2709938 13.4448077 9.8934331 7.2801353 5.3571262
[7] 3.9420697 2.9007929 2.1345639 1.5707301 1.1558301 0.8505237
[13] 2.3693811
```

Certains effectifs théoriques sont inférieurs à 5, et même à 1, on ne peut pas effectuer le test du khi 2 dans ces conditions, il faut construire d'autres classes afin d'avoir des effectifs théoriques supérieurs à 5. On effectue quand même le test dans ces conditions pour voir les conséquences du non respect des conditions d'application du test.

```
# CALCUL DE LA VALEUR OBSERVÉE DU KHI DEUX
(khi2Observe <- sum((histo$counts - EffTheo)^2 / EffTheo))
[1] 18.38974

# CALCUL DE LA PROBABILITÉ CRITIQUE ("P-VALUE")
# lower.tail = FALSE pour avoir P(KHI > khiobs)
pchisq(khi2Observe, length(histo$counts) - 2, lower.tail = FALSE)
[1] 0.0729669
```

```
# CONCLUSION PROVISoire :
# On obtient une  $P \approx 0,07297$  ("P-value"). Au seuil de 5% l'échantillon de durées observées
# ne nous permet pas de rejeter l'hypothèse du modèle exponentiel comme
# distribution théorique pour la variable durée entre deux éruptions successives
# du volcan A.
# Les conditions d'application du test n'étant pas remplies, nous allons recommencer
# avec d'autres classes.
```

```
# ATTENTION AUX OBSERVATIONS "EXTRÊMES" AVEC LA LOI EXPONENTIELLE !!!
# Que se passe-t-il si on élimine l'unique observation à 1552 jours ?
# Essayez-donc. Il n'y a que deux lignes et quelques caractères à
# changer dans la procédure.
```

3° AMÉLIORATIONS DU TRAITEMENT : LES CONCLUSIONS SONT DIFFÉRENTES !

► On garde toutes les observations mais on fait des classes d'effectifs plus réguliers et on évite les effectifs théoriques inférieurs à 1. Un autre essai peut être fait avec des effectifs théoriques tous supérieurs à 5.

```
(subdiv1 <- c(seq(0, 210, 30), 360, 1560))
[1] 0 30 60 90 120 150 180 210 360 1560

histo <- hist(duree, breaks = subdiv1, freq = FALSE, right = FALSE)

histo$breaks # Les bornes des classes utilisées dans notre exemple c'est biensur subdiv1.
[1] 0 30 60 90 120 150 180 210 360 1560

histo$counts # Les effectifs observés des classes, sur la série étudiée
[1] 41 14 8 8 5 3 4 7 4

OrdreSubdiv1 <- length(subdiv1)
EffTheo <- NULL
for (i in 1:(OrdreSubdiv1 - 1)) {
  EffTheo <- c(EffTheo,
    (pexp(histo$breaks[i + 1], lambda) - pexp(histo$breaks[i], lambda)) * length(duree))
}

EffTheo
[1] 24.829601 18.270994 13.444808 9.893433 7.280135 5.357126 3.942070
[8] 8.612441 2.369381

(khi2Observe <- sum((histo$counts - EffTheo)^2 / EffTheo))
[1] 17.27302

pchisq(khi2Observe, length(histo$counts) - 2, lower.tail = FALSE)
[1] 0.01571806
```

Conclusion : l'hypothèse du modèle exponentiel est rejetée avec un risque de 1,57 %.

```
# ON DÉTACHE EruptA
detach(EruptA)
```

► Qu'est-ce que ça donne en éliminant l'observation 1552 ?

```
EruptA_1 <- EruptionA[EruptionA$duree != 0 & EruptionA$duree != 1552, ]
attach(EruptA_1)
(subdiv1 <- c(seq(0, 180, 30), 510))
[1] 0 30 60 90 120 150 180 510

histo <- hist(duree, breaks = subdiv1, freq = FALSE, right = FALSE)

histo$breaks # Les bornes des classes utilisées dans notre exemple c'est biensur subdiv1.
[1] 0 30 60 90 120 150 180 510

histo$counts # Les effectifs observés des classes, sur la série étudiée
[1] 41 14 8 8 5 3 14

OrdreSubdiv1 <- length(subdiv1)
EffTheo <- NULL
for(i in 1:(OrdreSubdiv1 - 1)){
  EffTheo <- c(EffTheo,
    (pexp(histo$breaks[i + 1], lambda) - pexp(histo$breaks[i], lambda)) * length(duree))
}
EffTheo
[1] 28.445371 19.744950 13.705677 9.513601 6.603731 4.583886 10.215268

(khi2Observe <- sum((histo$counts - EffTheo)^2 / EffTheo))
[1] 12.16772

pchisq(khi2Observe, length(histo$counts) - 2, lower.tail = FALSE)
[1] 0.03256056
```

Conclusion : l'hypothèse du modèle exponentiel est rejetée avec un risque de 3,26 %.

```
# ON DÉTACHE EruptA_1
detach(EruptA_1)
```

4° ET POUR DÉPARTAGER TOUT LE MONDE

► Avec les données de EruptA, on ne peut pas utiliser le test de Kolmogorof-Smirnov pour lequel il ne doit pas y avoir d'ex æquo dans les séries. Une solution provisoire peut consister à rajouter et retrancher quelques dixièmes voire centièmes aux ex æquo pour qu'il n'y en ait plus et voir ce que ça donne. À essayer...On a obtenu :

```
ks.test(EruptA_2$duree, "pexp", 1 / mean(EruptA_2$duree))
One-sample Kolmogorov-Smirnov test
data: EruptA_2$duree
D = 0.2128, p-value = 0.0003254
alternative hypothesis: two-sided
```

Conclusion : l'hypothèse du modèle exponentiel est rejetée avec un risque de 0,03254 %. C'est sans appel ! Ça confirme bien l'avis que l'on avait eu après avoir observé une moyenne très différente de l'écart type.

► Le modèle exponentiel étant le modèle des phénomènes sans mémoire, peut peut-être penser qu'il y a de la mémoire dans les éruptions volcaniques.