

Mathématiques actuarielles IARD I

Travail pratique 1

Nicholas Langevin

Alexandre Turcotte

2018-10-26

Question 1

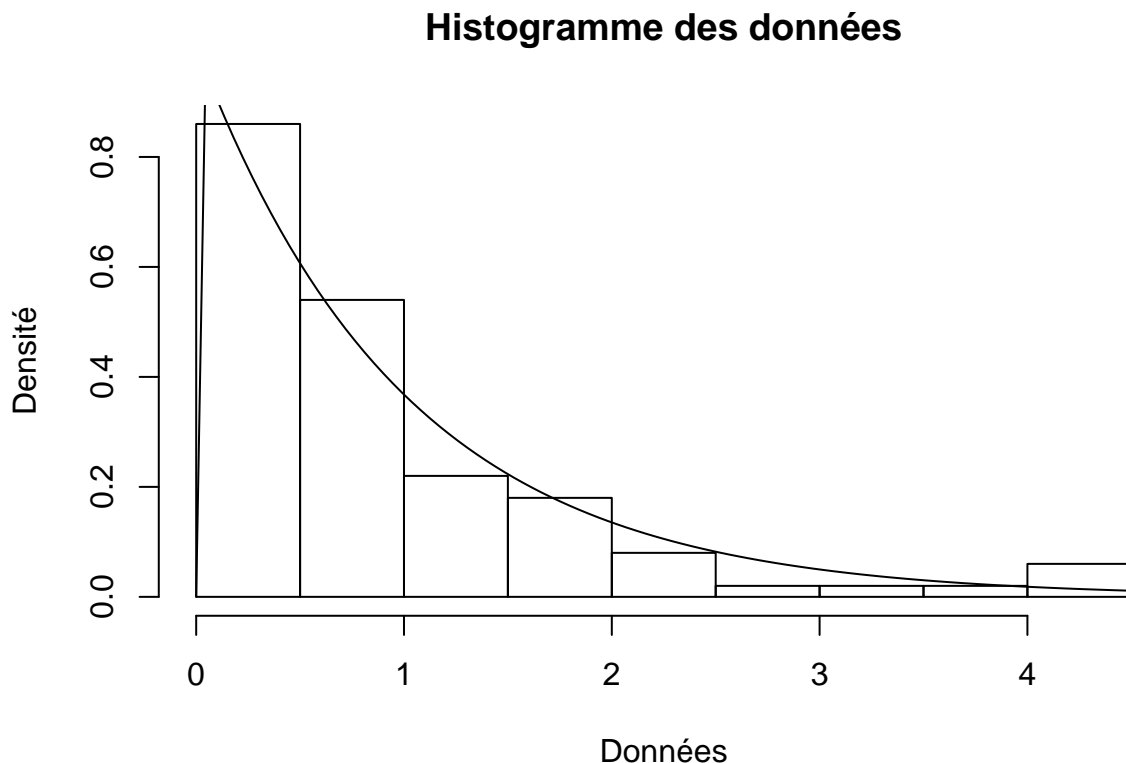
a) Estimation du coefficient d'asymétrie

```
data <- rexp(100, 1)

coef_asymetrie <- function(x){
  mu <- mean(x)
  sd <- sd(x)
  mean((x - mu)^3) / sd^3
}

estimateur_coef_asymetrie <- coef_asymetrie(data)

hist(data, probability = TRUE,
      main = "Histogramme des données",
      ylab = "Densité",
      xlab = "Données");curve(dexp(x,1), add = TRUE)
```



À partir de l'histogramme, il est possible de noter que la distribution est plus dense et concentrée à gauche et que la queue de la distribution tend vers la droite. Par conséquent, la distribution n'est pas symétrique et le coefficient d'asymétrie devrait être positif. Cela concorde effectivement avec le coefficient d'asymétrie estimé empiriquement puisqu'il est de 1.8456646 comparativement à celui de la loi normale qui est égal à 0. La loi normale a une distribution symétrique et comme le coefficient d'asymétrie des valeurs simulées est plus grand que celui de la loi normale, cela implique que la distribution est asymétrique vers la droite. Donc, elle possède une queue de distribution à droite comme il est possible d'observer sur l'histogramme précédent.

b) Intervalle de confiance pour le coefficient d'asymétrie

```
teta <- mean(data)
echantillon <- lapply(1:50, function(i) rexp(100, teta))
coef_asymetrie_simul <- sapply(1:50, function(i) coef_asymetrie(echantillon[[i]]))

moyenne_coef_asymetrie <- mean(coef_asymetrie_simul)
variance_coef_asymetrie <- sum((coef_asymetrie_simul - moyenne_coef_asymetrie)^2) / (length(coef_asymetrie) - 1)
variance_coef_asymetrie_modifie <- var(coef_asymetrie_simul)

IC_coef_asymetrie <- cbind(estimateur_coef_asymetrie - qnorm(0.975) * sqrt(variance_coef_asymetrie),
                           estimateur_coef_asymetrie + qnorm(0.975) * sqrt(variance_coef_asymetrie))
```

Avec la méthode de ré-échantillonnage, la variance estimée pour le coefficient d'asymétrie est de 0.1358482. À partir de l'estimateur ponctuel du coefficient d'asymétrie calculé en a) et de cette variance estimée, il est possible d'obtenir l'intervalle de confiance suivant pour le coefficient d'asymétrie :

[1.1232692, 2.56806]

c) Coefficient d'asymétrie théorique

Les moments de la loi exponentielle sont donnés par:

$$\begin{aligned}
 E[x] &= M'_x(t) \Big|_{t=0} & E[x^2] &= M''_x(t) \Big|_{t=0} & E[x^3] &= M'''_x(t) \Big|_{t=0} \\
 &= \frac{d}{dt} \left(\frac{\theta}{\theta - t} \right) \Big|_{t=0} & &= \frac{d}{dt} \left(\frac{\theta}{(\theta - t)^2} \right) \Big|_{t=0} & &= \frac{d}{dt} \left(\frac{2\theta}{(\theta - t)^3} \right) \Big|_{t=0} \\
 &= \left(\frac{\theta}{(\theta - t)^2} \right) \Big|_{t=0} & &= \left(\frac{2\theta}{(\theta - t)^3} \right) \Big|_{t=0} & &= \left(\frac{6\theta}{(\theta - t)^4} \right) \Big|_{t=0} \\
 &= \frac{1}{\theta} & &= \frac{2}{\theta^2} & &= \frac{6}{\theta^3}
 \end{aligned}$$

Ainsi, le coefficient d'asymétrie théorique est donnée par

$$\begin{aligned}
 \gamma &= \frac{E[(x - \mu)^3]}{\sigma^3} \\
 &= \frac{1}{\sigma^3} (E[x^3 - 3x^2\mu + 3x\mu^2 - \mu^3]) \\
 &= \frac{1}{\sigma^3} (E[x^3] - 3\mu E[x^2] + 3\mu^2 E[x] - \mu^3) \\
 &= \theta^3 \left[\frac{6}{\theta^3} - 3 \left(\frac{1}{\theta} \right) \left(\frac{2}{\theta^2} \right) + 3 \left(\frac{1}{\theta} \right)^3 - \left(\frac{1}{\theta} \right)^3 \right] \\
 &= \theta^3 \left[\frac{6}{\theta^3} - \frac{6}{\theta^3} + \frac{2}{\theta^3} \right] \\
 &= 6 - 6 + 2 \\
 &= 2
 \end{aligned}$$

Le coefficient d'asymétrie théorique de la loi exponentielle de moyenne 1 est de 2. Cela s'avère compatible avec l'estimé ponctuel obtenu en a) puisque celui-ci est de 1.8456646, ce qui est assez près de 2. Pour ce qui est de l'intervalle de confiance obtenu en b), il est possible d'observer que la valeur théorique est incluse dans cette intervalle qui est 1.1232692, 2.56806. Par conséquent, les estimateurs obtenus sont compatibles avec la valeur théorique. Toutefois, ceux-ci ne sont pas très précis en raison du nombre d'observations qui est assez faible, soit 100 observations. En augmentant le nombre d'observation, l'intervalle de confiance serait plus petit et l'estimateur serait plus précis.

Question 2

a) Estimation de $E[\min(X, u)]$

La fonction quantile théorique d'une loi exponentielle ($\theta = 1$) est déterminée de cette façon :

$$\begin{aligned}X &\sim \text{Exp}(\theta = 1) \\F(x) &= 1 - e^{-\frac{x}{\theta}} \\&= 1 - e^{-x} \\k &= 1 - e^{-x} \\1 - k &= e^{-x} \\F_x^{-1}(k) &= x = -\ln(1 - k)\end{aligned}$$

```
k <- c(0.25, 0.35, 0.5, 0.6, 0.75, 0.85)
fonction_percentile <- function(x) -log(1-x)
limite_u <- sapply(k, function(i) fonction_percentile(i))

Estimateur_Esperance_limite <- sapply(1:length(limite_u), function(u) mean(sapply(1:100, function(i) min(
##  Percentile  Limite.u  Espérance.limitée
## 1      0.25 0.2876821      0.2421759
## 2      0.35 0.4307829      0.3391848
## 3      0.50 0.6931472      0.4828666
## 4      0.60 0.9162907      0.5733623
## 5      0.75 1.3862944      0.6955198
## 6      0.85 1.8971200      0.7787864
```

$E[\min(X, u)]$ estimé est une fonction croissante en fonction du u . En effet, le tableau présente une augmentation de la valeur de l'espérance lorsque u augmente. Cela s'avère tout à fait logique puisque lorsque u est petit, la fonction $\min(X, u)$ prend davantage en considération les valeurs de u . Par conséquent, les valeurs supérieures de x sont réduites, ce qui réduit donc la moyenne, car elle ne tient compte que des valeurs inférieures ou égales à u . Alors, si u augmente, l'espérance prendra en compte des valeurs plus grande de x , car u aura augmenté.

b) Intervalle de confiance pour $E[\min(X, u)]$

```
teta <- mean(data)
echantillon <- lapply(1:50, function(i) rexp(100, teta))
Esperance_limite_simul <- sapply(c(3,5), function(u)
  sapply(1:50, function(j)
    mean(sapply(1:100, function(i)
      min(echantillon[[j]][i], limite_u[u])))))

moyenne_Esperance_limite_simul <- sapply(1:2, function(i) mean(Esperance_limite_simul[,i]))
variance_Esperance_limite_simul <- sapply(1:2, function(i) sum((Esperance_limite_simul[,i] - moyenne_Esperance_limite_simul[i])^2))
variance_Esperance_limite_simul_modifie <- sapply(1:2, function(i) var(Esperance_limite_simul[,i]))

resultats_b <- data.frame("Percentile" = c(k[3], k[5]), "Limite u" = c(limite_u[3], limite_u[5]), "Espérance" = c(moyenne_Esperance_limite_simul[3], moyenne_Esperance_limite_simul[5]))

Est_Esperance_limite <- rep(0,2)
Est_Esperance_limite[1] <- Estimateur_Esperance_limite[3]
Est_Esperance_limite[2] <- Estimateur_Esperance_limite[5]

IC_Esperance_limite_simul <- lapply(1:2, function(i)
  cbind(Est_Esperance_limite[i] - qnorm(0.975) * sqrt(variance_Esperance_limite_simul[i]),
    Est_Esperance_limite[i] + qnorm(0.975) * sqrt(variance_Esperance_limite_simul[i])))
```

Avec la méthode de ré-échantillonnage, la variance estimée pour $E[\min(X, F^{-1}(0.5))]$ est de ??????, alors que celle pour $E[\min(X, F^{-1}(0.75))]$ est de ??????. À partir des estimateurs de $E[\min(X, u)]$ calculés en a) pour $u = F^{-1}(0.5)$ et $u = F^{-1}(0.75)$ et de ces variances estimées, il est possible d'obtenir les intervalles de confiance suivants :

$$[0.4366736, 0.5290597], \text{ pour } u = F^{-1}(0.5)$$

$$[0.5964305, 0.794609], \text{ pour } u = F^{-1}(0.75)$$

c) $E[\min(X, u)]$ théorique

$$\begin{aligned}
X &\sim \text{Exp}(\theta = 1) \\
S(x) &= e^{-x}, \text{ où } x > 0 \\
u &= F_x^{-1}(k) = -\ln(1 - k) \\
E[\min(X, u)] &= E[X \wedge u] \\
&= \int_0^u x f_x(x) dx + \int_0^u u f_x(x) dx \\
&= \int_0^u S_x(x) dx \\
&= \int_0^u e^{-x} dx \\
&= [-e^{-x}]_0^u \\
&= 1 - e^{-u} \\
&= 1 - e^{-(-\ln(1-k))} \\
&= 1 - (1 - k) \\
&= k
\end{aligned}$$

Ainsi, les valeurs théoriques des espérances limitées sont :

$$\begin{aligned}
E[\min(X, F_x^{-1}(0.5))] &= 0.5 \\
E[\min(X, F_x^{-1}(0.75))] &= 0.75
\end{aligned}$$

L'espérance limitée théorique de la loi exponentielle de moyenne 1 est égale à son quantile. Par conséquent, $E[\min(X, F^{-1}(0.5))] = 0.5$ et $E[\min(X, F^{-1}(0.75))] = 0.75$. Cela s'avère compatible avec les valeurs estimées obtenues en a) puisque ceux-ci sont de 0.4828666 et 0.6955198, ce qui est assez près des valeurs théoriques. Pour ce qui est des intervalles de confiance obtenu en b), il est possible d'observer que les valeurs théoriques sont incluses dans chacun des intervalles respectifs qui sont $[0.4366736, 0.5290597]$, pour le premier, et $[0.5964305, 0.794609]$, pour le second. Par conséquent, les estimateurs obtenus sont compatibles avec les valeurs théoriques. Il est donc possible d'affirmer que ces estimateurs non-paramétriques sont assez performants puisque les valeurs obtenues sont très près de la valeur théorique.

Question 3

a) Estimation de la fonction de survie avec Kaplan-Meier

```
Temps <- c(30, 40, 57, 65, 65, 84, 90, 92, 98, 101)
Cens <- c(1, 1, 1, 1, 0, 1, 1, 0, 1, 1)

donnees <- data.frame(Temps, Cens)

yi <- unique(donnees[which(donnees[,2] != 0),1])
i <- 1:length(yi)
Si <- rep(1,length(yi))
ri <- c(10,9,8,7,5,4,2,1)

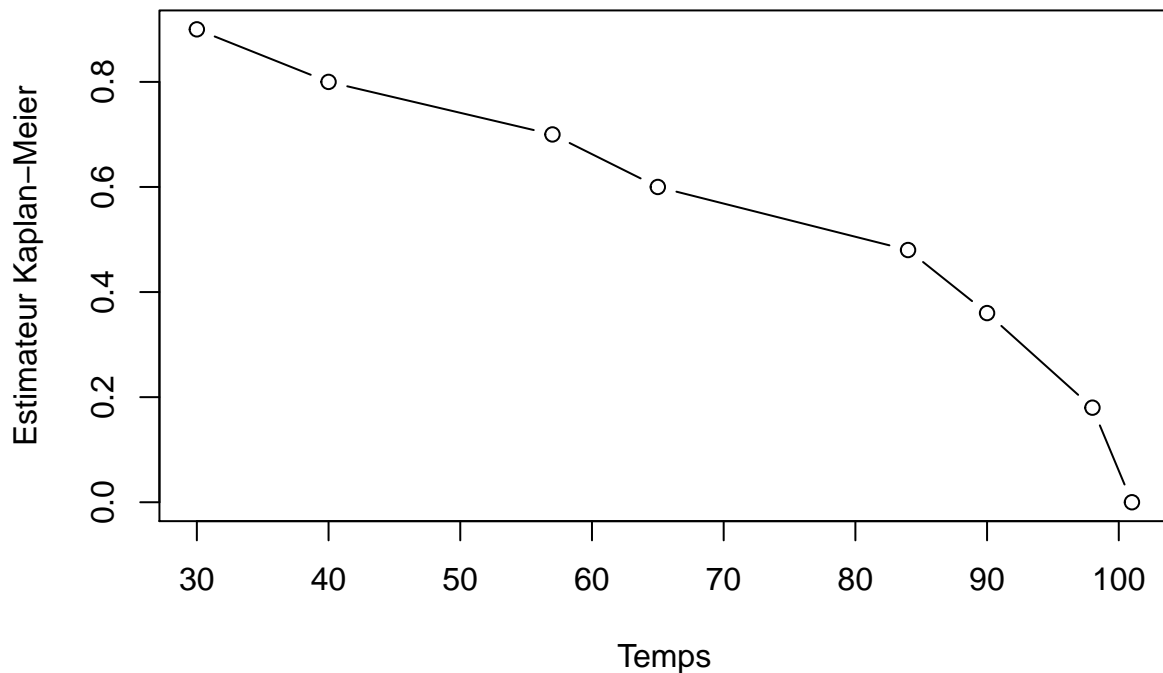
Estimateur_Kaplan_Meier <- cumprod(1-Si/ri)
jeu_de_donnees <- data.frame(i, yi, Si, ri, "Kaplan-Meier" = cumprod(1-Si/ri))
```

$$S_n(t) = \prod_{i=1}^t \left(1 - \frac{S_i}{r_i}\right)$$
$$Var(\hat{S}_n(t)) = (S_n(t))^2 \sum_{i=1}^t \frac{S_i}{r_i(S_i - r_i)}$$

b) Graphique de l'estimateur Kaplan-Meier et intervalle de confiance pour S(50)

```
plot(yi, Estimateur_Kaplan_Meier,
     main = "Valeur de la fonction de survie estimée à l'aide de l'estimateur  
Kaplan-Meier en fonction des temps de décès ou de retraits",
     ylab = "Estimateur Kaplan-Meier",
     xlab = "Temps",
     type = "b")
```


Valeur de la fonction de survie estimée à l'aide de l'estimateur Kaplan-Meier en fonction des temps de décès ou de retraits



```
formule_Greenwood <- Estimateur_Kaplan_Meier^2 * cumsum(Si/ri/(ri-Si))

# Estimer S(50)
x <- 50
ci <- 40
cj <- 57
alpha <- (cj-x)/(cj-ci)
Sn_50 <- 1 - (alpha)*(1 - Estimateur_Kaplan_Meier[2]) - (1-alpha)*(1 - Estimateur_Kaplan_Meier[3])

IC_KM <- cbind(Estimateur_Kaplan_Meier - qnorm(0.975) * sqrt(formule_Greenwood),
               Estimateur_Kaplan_Meier + qnorm(0.975) * sqrt(formule_Greenwood))
```

c) Intervalle de confiance pour S(50) avec la transformation log

```
u <- exp(qnorm(0.975) * sqrt(formule_Greenwood) / Estimateur_Kaplan_Meier / log(Estimateur_Kaplan_Meier))
IC_log_KM <- cbind(Estimateur_Kaplan_Meier^(1/u),
                  Estimateur_Kaplan_Meier^u)
```

L'intervalle de confiance devrait être plus petit et précis dans le cas de la transformation log.

Annexe

Échantillon de 100 données simulées à partir d'une loi exponentielle ($\theta = 1$) :

```
## [1] 0.919650415 0.717212707 0.282810109 0.032352982 2.190912565
## [6] 0.270882983 1.143504559 1.103791472 0.170448880 4.228297643
## [11] 0.634231566 0.475893708 0.345811443 0.901130986 4.380071480
## [16] 0.480748731 3.769968664 1.102820972 0.049940266 0.329412986
## [21] 0.921090073 0.009259043 0.174303661 0.036363083 0.361134369
## [26] 0.053954623 0.970069817 0.579550514 2.345057947 0.886534595
## [31] 1.528126329 0.612656378 0.111276283 0.570605947 0.452282778
## [36] 1.916396424 0.099906598 0.300634415 1.628666380 0.707144582
## [41] 2.443099664 0.098792118 0.181754319 0.458591885 0.752830435
## [46] 0.495158303 0.887041624 0.726390219 0.300775925 0.525140396
## [51] 0.457845444 1.878089780 1.017944084 0.112189366 0.027473693
## [56] 0.347687219 1.330603951 0.351387809 0.710667171 1.262885997
## [61] 0.941583196 0.016906905 0.680453659 0.926439410 4.050266197
## [66] 2.529784506 1.014500598 1.338349316 0.703833825 0.274192515
## [71] 0.043761580 1.905824954 0.760207554 0.422156895 0.548729830
## [76] 0.613149600 0.284077761 0.400143021 2.098360884 0.056349838
## [81] 3.109720215 0.965048956 0.223169311 0.002000260 1.923377812
## [86] 0.035635612 1.357862946 0.403212744 0.235059064 0.888023276
## [91] 1.160179813 0.061677486 1.834976535 1.522228185 1.521472242
## [96] 0.662886842 0.681665088 0.215569843 1.476431503 0.056690745
```