Mathématiques actuarielles IARD I
Travail pratique 1

Nicholas Langevin
Alexandre Turcotte

2018-10-26

Question 1

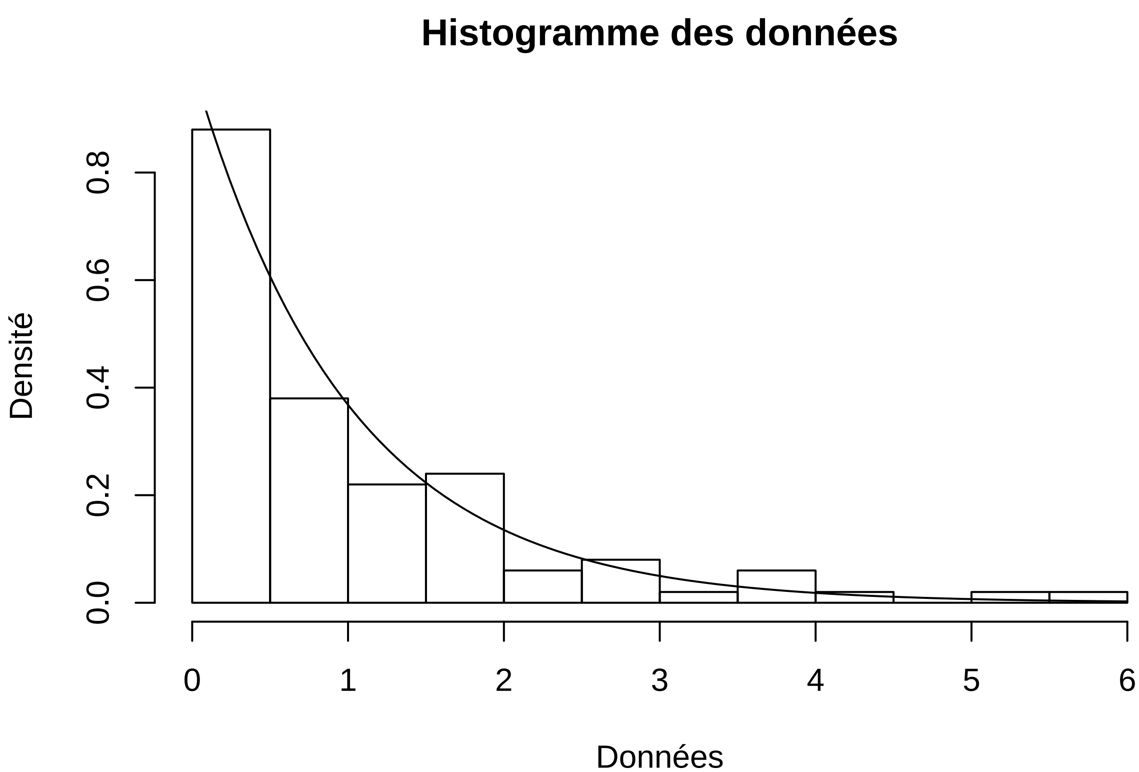
a) Estimation du coefficient d'asymétrie

```
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 1.7627689 conparativement 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)
echantillion <- lapply(1:50, function(i) rexp(100, teta))
coef_asymetrie_simul <- sapply(1:50, function(i) coef_asymetrie(echantillion[[i]]) )

moyenne_coef_asymetrie <- mean(coef_asymetrie_simul)
variance_coef_asymetrie <- sum((coef_asymetrie_simul - moyenne_coef_asymetrie)^2) / (length(coef_asymet
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 0.2435421. 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 :

$$[0.7955269, 2.7300109]$$

c) Coefficient d'asymétrie théorique

Les moments de la loi exponentielle sont donnés par:

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

Ainsi, le coeficient d'asymétrie théorique est donnée par

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

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 1.7627689, 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 0.7955269, 2.7300109. 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 :

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

	Percentile	Limite u	Espérance limitée
1	0.25	0.29	0.24
2	0.35	0.43	0.33
3	0.50	0.69	0.46
4	0.60	0.92	0.55
5	0.75	1.39	0.71
6	0.85	1.90	0.82

Table 1: Valeurs de l'espérance limitée pour chaque limite u donnée

$E[min(X, u)]$ 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 $E[min(X, u)]$

	Percentile	Limite u	Espérance limité	Variance
1	0.50	0.69	0.46	0.00
2	0.75	1.39	0.71	0.00

Table 2: Valeurs de l'espérance limitée et de sa variance pour les percentiles 0.5 et 0.75

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 :

$$[0.4134017, 0.5058479], \; pour \; u = F^{-1}(0.5)$$

$$[0.6085057, 0.8031621], \; pour \; u = F^{-1}(0.75)$$

c) $E[min(X,u)]$ théorique

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

Ainsi, les valeurs théoriques des espérances limitées sont :

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

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 0.4596248 et 0.7058339, 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 [0.4134017, 0.5058479], pour le premier, et [0.6085057, 0.8031621], 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) Détermination de la fonction de survie à l'aide de l'estimateur Kaplan-Meier

	i	yi	Si	ri	Kaplan-Meier
1	1	30.00	1.00	10.00	0.90
2	2	40.00	1.00	9.00	0.80
3	3	57.00	1.00	8.00	0.70
4	4	65.00	1.00	7.00	0.60
5	5	84.00	1.00	5.00	0.48
6	6	90.00	1.00	4.00	0.36
7	7	98.00	1.00	2.00	0.18
8	8	101.00	1.00	1.00	0.00

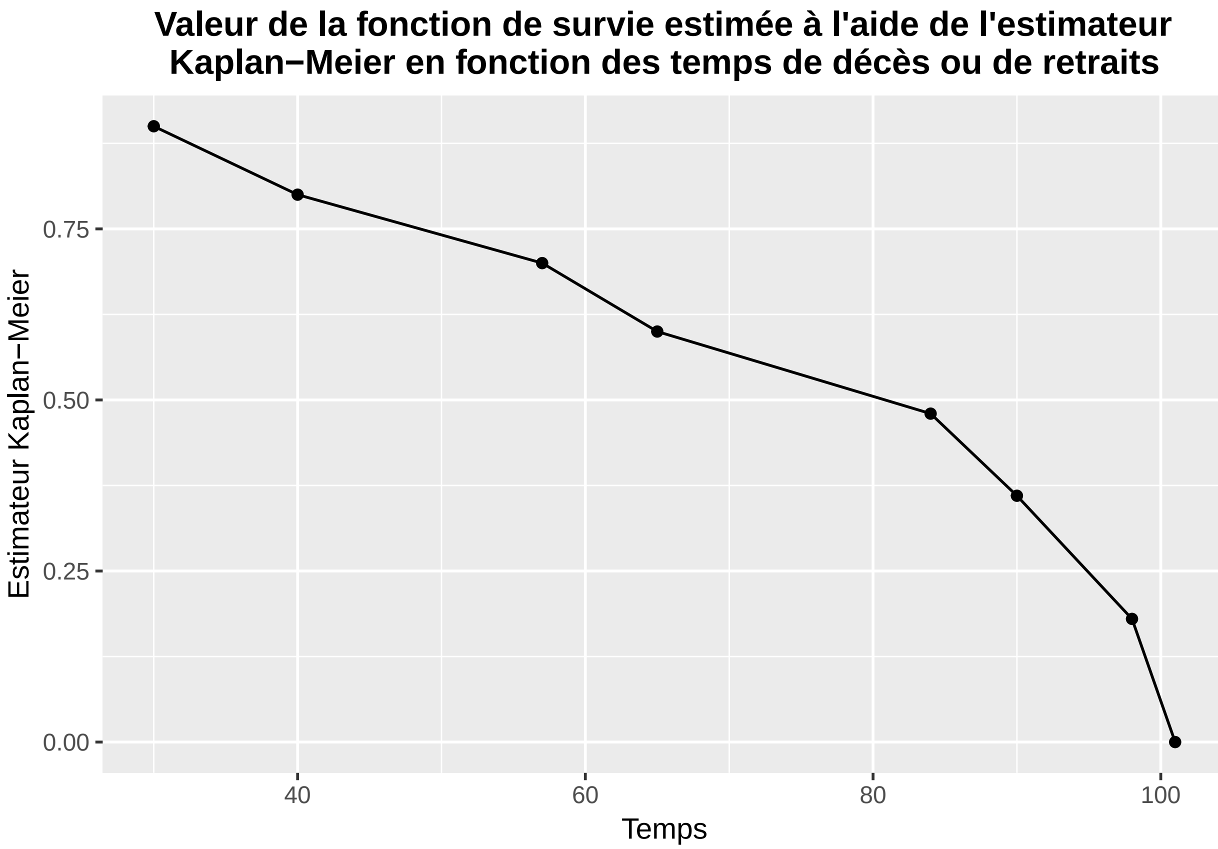
Table 3: Données permettant de trouver l'estimateur Kaplan-Meier

La table précèdente présente toutes les données utiles permettant de déterminer l'estimateur Kaplan-Meier. Par conséquent, il est possible de calculer cet estimateur à partir de sa définition :

$$\hat{S_n}(t) = \begin{cases} 1,\ 0 \leqslant t < y, \\ \\ \prod_{i=1}^{j-1}(\frac{r_i - S_i}{r_i}),\ y_{j-1} \leqslant t < y_i, \\ \\ \prod_{i=1}^{k}(\frac{r_i - S_i}{r_i}),\ t > y_k \end{cases}$$

$$\hat{S_n}(i) = \begin{cases} 1,\ 0 \leqslant i < 1, \\ 0.9,\ 1 \leqslant i < 2, \\ 0.8,\ 2 \leqslant i < 3, \\ 0.7,\ 3 \leqslant i < 4, \\ 0.6,\ 4 \leqslant i < 5, \\ 0.48,\ 5 \leqslant i < 6, \\ 0.36,\ 6 \leqslant i < 7, \\ 0.18,\ 7 \leqslant i < 8, \\ 0,\ i \geqslant 8 \end{cases}$$

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



Calculer un estimateur pour la variance de la fonction de survie à l'aide de la formule de Greenwood :

$$Var(\hat{S_n(t)}) = (S_n(\hat{y_j}))^2 \sum_{i=1}^{j} \frac{S_i}{r_i(S_i - r_i)}$$

$$Var(\hat{S_n(y_j)}) = \begin{cases} 0.009, & i = 1, \\ 0.016, & i = 2, \\ 0.021, & i = 3, \\ 0.024, & i = 4, \\ 0.02688, & i = 5, \\ 0.02592, & i = 6, \\ 0.02268, & i = 7 \end{cases}$$

Alors, en ayant les valeurs de l'estimateur Kaplan-Meier et celles de Greenwood, il est possible de déduire les intervalles de confiance au moment des décès. Toutefois, il faut déterminer l'intervalle de confiance au niveau 95% pour S(50), qui est entre deux moments de décès. Alors, il faut faire une interpolation linéaire (OGIVE) entre $y_i = 40$ et $y_i = 57$.

La fonction de survie sous OGIVE est donnée par:

$$\begin{aligned}
F_n^{OGIVE}(x) &= (\frac{c_j - x}{c_j - c_{j-1}})F_n(c_{j-1}) + (\frac{x - c_{j-1}}{c_j - c_{j-1}})F_n(c_j) \\
S_n^{OGIVE}(x) &= 1 - F_n^{OGIVE}(x) \\
&= 1 - (\frac{c_j - x}{c_j - c_{j-1}})F_n(c_{j-1}) - (\frac{x - c_{j-1}}{c_j - c_{j-1}})F_n(c_j) \\
&= 1 - (\frac{c_j - x}{c_j - c_{j-1}}) + (\frac{c_j - x}{c_j - c_{j-1}})S_n(c_{j-1}) - (\frac{x - c_{j-1}}{c_j - c_{j-1}}) + (\frac{x - c_{j-1}}{c_j - c_{j-1}})S_n(c_j) \\
&= 1 - (\frac{(c_j - x) + (x - c_{j-1})}{c_j - c_{j-1}}) + (\frac{c_j - x}{c_j - c_{j-1}})S_n(c_{j-1}) - (\frac{x - c_{j-1}}{c_j - c_{j-1}})S_n(c_j) \\
&= 1 - 1 + (\frac{c_j - x}{c_j - c_{j-1}})S_n(c_{j-1}) - (\frac{x - c_{j-1}}{c_j - c_{j-1}})S_n(c_j) \\
&= (\frac{c_j - x}{c_j - c_{j-1}})S_n(c_{j-1}) - (\frac{x - c_{j-1}}{c_j - c_{j-1}})S_n(c_j)
\end{aligned}$$

Donc, dans le cas de $S_n(50)$, la fonction de survie sous OGIVE est :

$$\begin{aligned}
S_n^{OGIVE}(50) &= (\frac{57 - 50}{57 - 40})S_n(y_{40}) - (\frac{50 - 40}{57 - 40})S_n(y_{57}) \\
&= 0.7411765
\end{aligned}$$

L'estimer de la variance sous OGIVE est donnée par:

$$Var(S_n^{OGIVE}(x)) = \frac{(c_j - c_{j-1})^2 Var(Y) + (x - c_{j-1})^2 Var(Z) + 2(c_j - c_{j-1})(x - c_{j-1})Cov(Y, Z)}{(n(c_j - c_{j-1}))^2}$$

où,

$$\begin{aligned}
Var(Y) &= nS_n(c_{j-1})(1 - S_n(c_{j-1})) \\
Var(Z) &= n(S_n(c_{j-1}) - S_n(c_j)(1 - S_n(c_{j-1}) + S_n(c_j))) \\
Cov(Y, Z) &= -n(1 - S_n(c_{j-1}))(S_n(c_{j-1}) - S_n(c_j))
\end{aligned}$$

Donc, dans le cas de $S_n(50)$, la variance de la fonction de survie sous OGIVE est :

$$\begin{aligned}
Var(Y) &= 8S_n(y_{40})(1 - S_n(y_{40})) = 1.28 \\
Var(Z) &= 8(S_n(y_{40}) - S_n(y_{57})(1 - S_n(y_{40}) + S_n(y_{57}))) = 1.36 \\
Cov(Y, Z) &= -8(1 - S_n(y_{40}))(S_n(y_{40}) - S_n(y_{57})) = -0.16
\end{aligned}$$

$$\begin{aligned}
Var(S_n^{OGIVE}(50)) &= \frac{(57 - 40)^2 Var(Y) + (50 - 40)^2 Var(Z) + 2(57 - 40)(50 - 40)Cov(Y, Z)}{(8(57 - 40))^2} \\
&= 0.0244118
\end{aligned}$$

Ainsi, en ayant $S_n(50)$ et $Var(S_n(50))$, il est possible d'évaluer l'intervalle de confiance au niveau de confiance 95%.

$$\begin{aligned} IC\ pour\ S_n(y_{50}) :&\ S_n(\hat{y_{50}})) \pm z_{0.975}\sqrt{Var(\hat{S_n}(\hat{y_{50}}))} \\ :&\ 0.7411765 \pm z_{0.975}\sqrt{0.0244118} \\ :&\ [0.4349465, 1.0474064] \end{aligned}$$

Mettre un commentaire ici

c) Intervalle de confiance pour $S_n(50)$ avec la transformation log (-log)

La valeur estimée de la fonction de survie et la valeur estimée de la variance sont toujours les mêmes que celles de la section b). Par conséquent, seulement l'intervalle de confiance est modifié. Alors, l'intervalle de confiance log (-log) au niveau de confiance 95% est déterminé de cette façon :

$$\begin{aligned} IC\ log\ transformed\ pour\ S_n(y_{50}) :&\ [S_n(\hat{y_{50}}))^{1/u}, S_n(\hat{y_{50}}))^{u}] \\ \text{où } u =&\ e^{\frac{z_{0.975}\sqrt{Var(\hat{S_n}(\hat{y_{50}}))}}{S_n(\hat{y_{50}}))ln(S_n(\hat{y_{50}})))}} \\ =&\ e^{\frac{z_{0.975}\sqrt{0.0244118}}{0.7411765ln(0.7411765)}} \end{aligned}$$

$$IC\ log\ transformed\ pour\ S_n(y_{50}) :\ [0.3042543, 0.9273784]$$

Mettre commentaire ici

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

Annexe

Question 1

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

```
##   [1] 0.034444681 1.976524802 1.597077945 1.721201496 1.967654057
##   [6] 5.148512171 3.836438035 2.102438243 0.430976172 1.017117288
##  [11] 2.578632874 0.713034804 0.985868526 1.790859364 1.035203516
##  [16] 0.136501027 2.641098155 0.696359569 1.355542085 0.254338868
##  [21] 1.671215070 1.017665065 0.527502543 0.674778048 2.377384411
##  [26] 0.545435634 1.707276596 0.300845366 3.652644577 0.011978765
##  [31] 0.093485447 1.802843811 0.181916061 0.353865854 0.131523122
##  [36] 0.041245298 3.662635436 0.059740326 0.700280771 1.616546682
##  [41] 0.250908713 0.782823294 0.061610715 0.300786967 0.952689390
##  [46] 0.535118755 0.251606925 0.577088892 1.236221440 0.265392981
##  [51] 0.124299291 0.099341116 0.049229451 1.008292694 0.789435583
##  [56] 0.069175366 4.045770040 0.160322735 0.387075583 1.022671052
##  [61] 0.041863370 5.562570676 0.376142044 0.369259422 0.325416113
##  [66] 0.069469030 1.481762383 1.896725114 0.856860921 1.296253844
##  [71] 1.347900188 2.857434546 0.241491352 0.500151911 0.304792375
##  [76] 0.099664086 3.484890215 0.282573399 1.315402822 0.391282904
##  [81] 0.508448064 0.525883565 2.081874916 0.284247235 0.091068743
##  [86] 1.517801506 0.111707081 0.002547933 0.745828135 0.593040601
##  [91] 0.179772019 0.115526127 0.893432197 0.256077890 0.128036648
##  [96] 0.324072997 0.289789488 0.061700472 1.629747102 2.510973671
```

Question 2

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

```r
# Déterminer les limites u à l'aide de la fonction quantile
k <- c(0.25, 0.35, 0.5, 0.6, 0.75, 0.85)
fonction_quantile <- function(x) -log(1-x)
limite_u <- sapply(k, function(i) fonction_quantile(i))

# Déterminer l'espérance limitée pour chacune des limites u
Estimateur_Esperance_limite <- sapply(1:length(limite_u), function(u)
                                    mean(sapply(1:100, function(i)
                                        min(data[i], limite_u[u]))))

# Publier les résultats pour chaque u
resultats <- data.frame(k, limite_u, Estimateur_Esperance_limite)
colnames(resultats) <- c('Percentile',"Limite u","Espérance limitée")

library(xtable)
options(xtable.comment = FALSE)
xtable(resultats, caption = "Valeurs de l'espérance limitée pour chaque limite u donnée",
                 align = c("c", "c", "c", "c"))
```

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

```
# Estimer le paramètre teta de la loi exponentielle
teta <- mean(data)

# Créer 50 échantillons de 100 données
echantillion <- lapply(1:50, function(i) rexp(100, teta))

# Déterminer l'espérance limitée pour chacun des 50 échantillons et pour
#les percentiles 0.5 et 0.75
Esperance_limite_simul <- sapply(c(3,5), function(u)
                                sapply(1:50, function(j)
                                    mean(sapply(1:100, function(i)
                                        min(echantillion[[j]][i], limite_u[u])))))

# Déterminer la variance de l'espérance limitée pour les deux percentiles donnés
variance_Esperance_limite_simul <- sapply(1:2,function(i) var(Esperance_limite_simul[,i]))

# Déterminer l'intervalle de confiance
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])))

# Publier les résultats
resultats_2 <- data.frame(c(k[3], k[5]), c(limite_u[3], limite_u[5]),
                          Est_Esperance_limite, variance_Esperance_limite_simul)
colnames(resultats_2) <- c("Percentile","Limite u","Espérance limité", "Variance")

library(xtable)
options(xtable.comment = FALSE)
xtable(resultats_2, caption = "Valeurs de l'espérance limitée et de sa variance
                 pour les percentiles 0.5 et 0.75",
                 align = c("c", "c", "c", "c", "c"))
```

c) $E[min(X, u)]$ théorique

Aucun calcul R n'a été fait dans cette section.

Question 3

a) Détermination de la fonction de survie à l'aide de l'estimateur Kaplan-Meier

```
# Présentation des données du problème
tableau1 <- {
```

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

    data.frame(Temps, Cens)
}

# Tableau détaillé permettant de calculer l'estimateur Kaplan-Meier
tableau2 <- {
    yi <- unique(tableau1[which(tableau1[,2] != 0),1])    # moments uniques des décès
    i <- 1:length(yi)
    Si <- rep(1,length(yi))                              # nombre de décès au temps yi
    ri <- c(10,9,8,7,5,4,2,1)                            # nombre de survivants au temps yi

    data.frame(i, yi, Si, ri, "Kaplan Meier" = cumprod(1-Si/ri))
}
colnames(tableau2) <- c("i", "yi", "Si", "ri", "Kaplan-Meier")

# Déterminer l'estimateur Kaplan-Meier pour chacun des moments uniques des décès
Estimateur_KM <- cumprod(1-Si/ri)

# Publier le tableau 2
library(xtable)
options(xtable.comment = FALSE)
xtable(tableau2, caption = "Données permettant de trouver l'estimateur Kaplan-Meier",
                 align = c("c", "c", "c", "c", "c", "c"))
```

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

```
# Graphique de la fonction de survie estimée par l'estimateur Kaplan-Meier
library(ggplot2)
ggplot(data = tableau2, aes(x=unlist(tableau2["yi"]),y=unlist(tableau2["Kaplan-Meier"])))+
        geom_point() +
        geom_line() +
        ggtitle("Valeur de la fonction de survie estimée à l'aide de l'estimateur
                Kaplan-Meier en fonction des temps de décès ou de retraits") +
        theme(plot.title =  element_text(face="bold", hjust = 0.5)) +
        xlab("Temps") +
        ylab("Estimateur Kaplan-Meier")

# Calculer un estimateur pour la variance de la fonction de survie à l'aide de
# la formule de Greenwood
formule_Greenwood <- Estimateur_KM^2 * cumsum(Si/ri/(ri-Si))

# Pour ce qui est de l'intervalle de confiance au niveau 95% pour S(50), il faut faire
# une interpolation linéaire (OGIVE) entre y_i = 40 et y_i = 57.

# Estimer S(50) avec la fonction OGIVE
Sn_50 <- {
    x <- 50
    ci <- 40
    cj <- 57
    alpha <- (cj-x)/(cj-ci)
```

```r
    Estimateur_KM[2] * alpha + Estimateur_KM[3] * (1 - alpha)
}

# Estimer Var(S(50)) avec la fonction OGIVE
Var_Sn_50 <- {
    x <- 50
    ci <- 40
    cj <- 57
    alpha <- (cj-x)/(cj-ci)
    n <- length(yi)

    var_Y <- n * Estimateur_KM[2] * (1 - Estimateur_KM[2])
    var_Z <- n * (Estimateur_KM[2] - Estimateur_KM[3] * (1 - Estimateur_KM[2]
            + Estimateur_KM[3]))
    cov <- -n * (1 - Estimateur_KM[2]) * (Estimateur_KM[2] - Estimateur_KM[3])

    (var_Y*(cj-ci)^2 + var_Z*(x-ci)^2 + 2*cov*(cj-ci)*(x-ci)) / (n*(cj-ci))^2
}

# Intervalle de confiance au niveau 95% pour S(50)
IC_KM <- cbind(Sn_50 - qnorm(0.975) * sqrt(Var_Sn_50),
              Sn_50 + qnorm(0.975) * sqrt(Var_Sn_50))
```

c) Intervalle de confiance pour $S_n(50)$ avec la transformation log (-log)

```r
# La valeur estimée de la fonction de survie et la variance estimée sont
#toujours les mêmes que celles de la section b). Par conséquent,
# seulement l'intervalle de confiance est modifié :
u <- exp(qnorm(0.975) * sqrt(Var_Sn_50) / Sn_50 / log(Sn_50))

IC_log_KM <- cbind(Sn_50^(1/u),
              Sn_50^u)
```