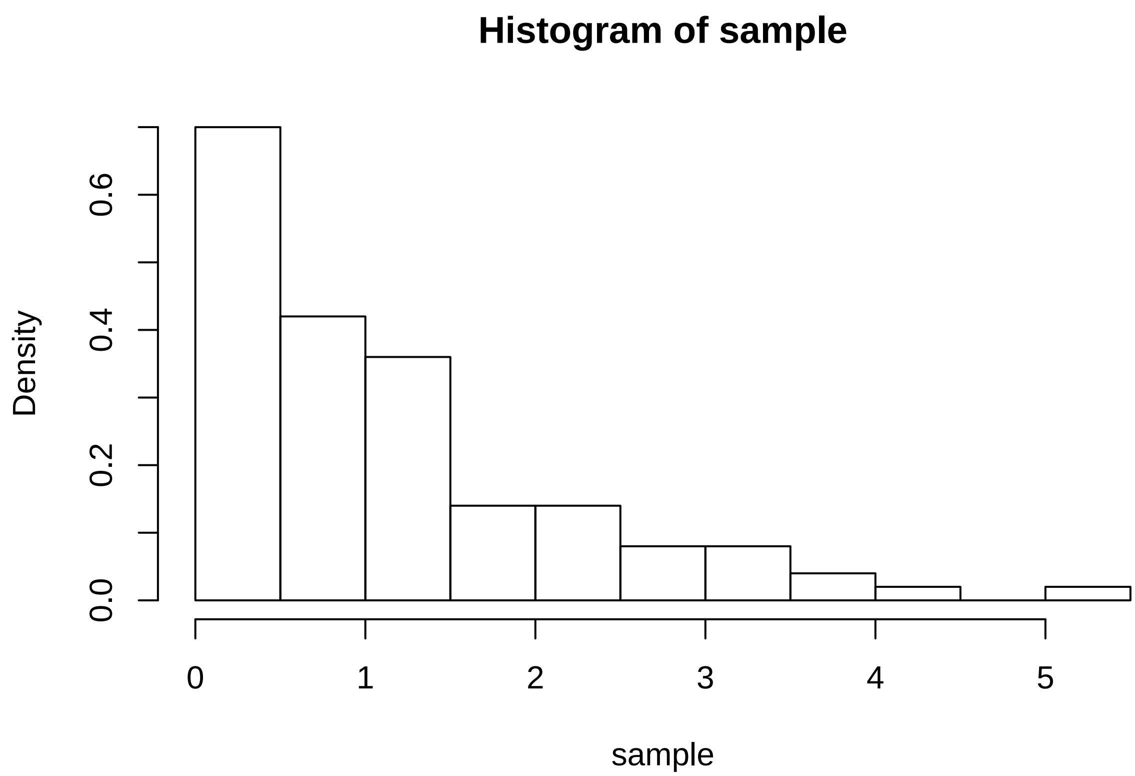
Question 1

a)

À l'aide de la simulation de 100 données, un coefficient d'asymétrie, $\gamma = 1.4012841$, a pu être estimé en utilisant la fonction empirique. On constate que $\gamma > 0$ et donc que la distribution est asymétrique vers la droite. L'histogramme de l'échantillons vient appuyer le fait ci-dessus:



b)

Afin d'estimer la variance de γ, le paramètre de la distribution a été estimé à l'aide de la moyenne empirique selon la formule suivante : $1/\hat{\Theta} = \bar{X}$, ce qui implique que $\hat{\Theta} = 0.8774698$.

Grâce a cet estimé, il a été possible d'effectuer la méthode "bootstrap" afin d'estimer la variance. Pour ce faire, 50 échantillons de tailles 100 d'une loi exponentielle de paramètre $\hat{\Theta}$ ont été générés. Par la suite, sur chacun des 50 échantillons, le coefficient a été calculé puis finalement la variance des 50 échantillons a été mesurés, donnant ainsi le résultat suivant : $Var(\hat{\gamma}) = 0.2036475$.

Finalement, un intervalle de confiance de a été calculé à l'aide de la statistique normale suivante :

$\hat{\gamma} \sim N(\gamma, Var(\hat{\gamma}))$

Ainsi l'intervalle de confiance de niveau 95% pour γ a été créé selon la formule suivante: $\hat{\gamma} \pm Z_{0.975} * Var(\hat{\gamma})$ donnant ainsi $\gamma \, \epsilon \, [0.5168048, 2.2857633]$.

c)

$$E[(X-\mu)^3]/\sigma^3 = (E[X^3] - 3E[X^2]\mu + 2\mu^3)/\sigma^3$$
$$\text{Or,dans le cas d'une exponentielle,} E[X^3] = 6/\lambda^3,\ E[X^2] = 2/\lambda^2,\ E[X] = \mu = 1/\lambda$$
$$\rightarrow E[(X-\mu)^3] = 2/\lambda^3$$
$$\text{Avec } \sigma^3 = 1/\lambda^3$$
$$\rightarrow E[(X-\mu)^3]/\sigma^3 = 2 = \gamma$$

Par rapport à la théorie, notre estimé ponctuel a une différence de 0.5987159, une différence très mineure pour un échantille de taille 100. De plus, la valeur théorique de 2 se trouve dans l'intervalle de confiance construit précédemment.

Question 2

a)

À partir de la distribution empirique de notre échantillon initial, les valeurs de $E[min(X,u)]$ pour les valeurs de $u\,\epsilon[F^{-1}(0.25), F^{-1}(0.35), F^{-1}(0.50), F^{-1}(0.60), F^{-1}(0.75), F^{-1}(0.85)]$,ou $F^{-1}()$ représente la fonction quantile de la loi exponentielle théorique, ont été estimés: 0.2482293, 0.3531438, 0.5156485, 0.6263346, 0.8088922, 0.9350867

De ces résultats il est possible d'observer que plus la valeur de u est élevé, plus la valeur estimée de $E[min(X,u)]$ est élevée. Cette variation est tout à fait logique, étant donnée que lorsque u augmente, certaines valeurs prises dans le calcul de l'espérance sont plus élevés dû à l'augmentation du seuil maximal des valeurs de X.

b)

Afin d'estimer la variance de $E[min(X,u)]$, pour $u\,\epsilon[F^{-1}(0.5), F^{-1}(0.75)]$, la méthode "bootstrap" a été utilisé.

Tel qu'au numéro 1 b), 50 échantillons de tailles 100 d'une loi exponentielle de paramètre $\hat{\Theta}$ ont été générés. Par la suite, sur chacun des 50 échantillons, le limited-loss a été calculé puis finalement la variance des 50 échantillons a été mesurés.

Ainsi l'intervalle de confiance de niveau 95% pour $E[min(X,u)]$ a été créé selon la formule suivante: $E[min\hat{(}X,u)] \pm Z_{0.975} * Var(E[min(X,u)])$ donnant ainsi.

$E[min(X, F^{-1}(0.5))]\,\epsilon\,[0.4707925, 0.5605046]$.

$E[min(X, F^{-1}(0.75))]\,\epsilon\,[0.7171035, 0.9006808]$.

c)

Afin de pouvoir comparer les valeurs expérimentales, il est nécessaire d'avoir les valeurs théoriques. Celles-ci se calculent de la façon suivante:

$$
\begin{aligned}
E[min(X,u)] &= \int_0^u e^{-\lambda x} dx \\
&= \frac{1-e^{-\lambda u}}{\lambda} \\
&= F_X(u) * E[X]
\end{aligned}
$$

Dans notre cas, les valeurs théoriques du limited-loss en a) sont respectivement de 0.25, 0.35, 0.50, 0.60, 0.75, 0.85. En rappelant que les valeurs sont expérimentales sont:
0.2482293, 0.3531438, 0.5156485, 0.6263346, 0.8088922, 0.9350867. On constate dès lors que les valeurs expérimentales sont près des valeurs théoriques. En effet, les valeurs de 0.5 et 0.75 sont inclus dans les intervalles de confiance trouvé en b). De ce fait, il est possible de constater que les estimateurs sont plutôt performants.

Question 3

a)

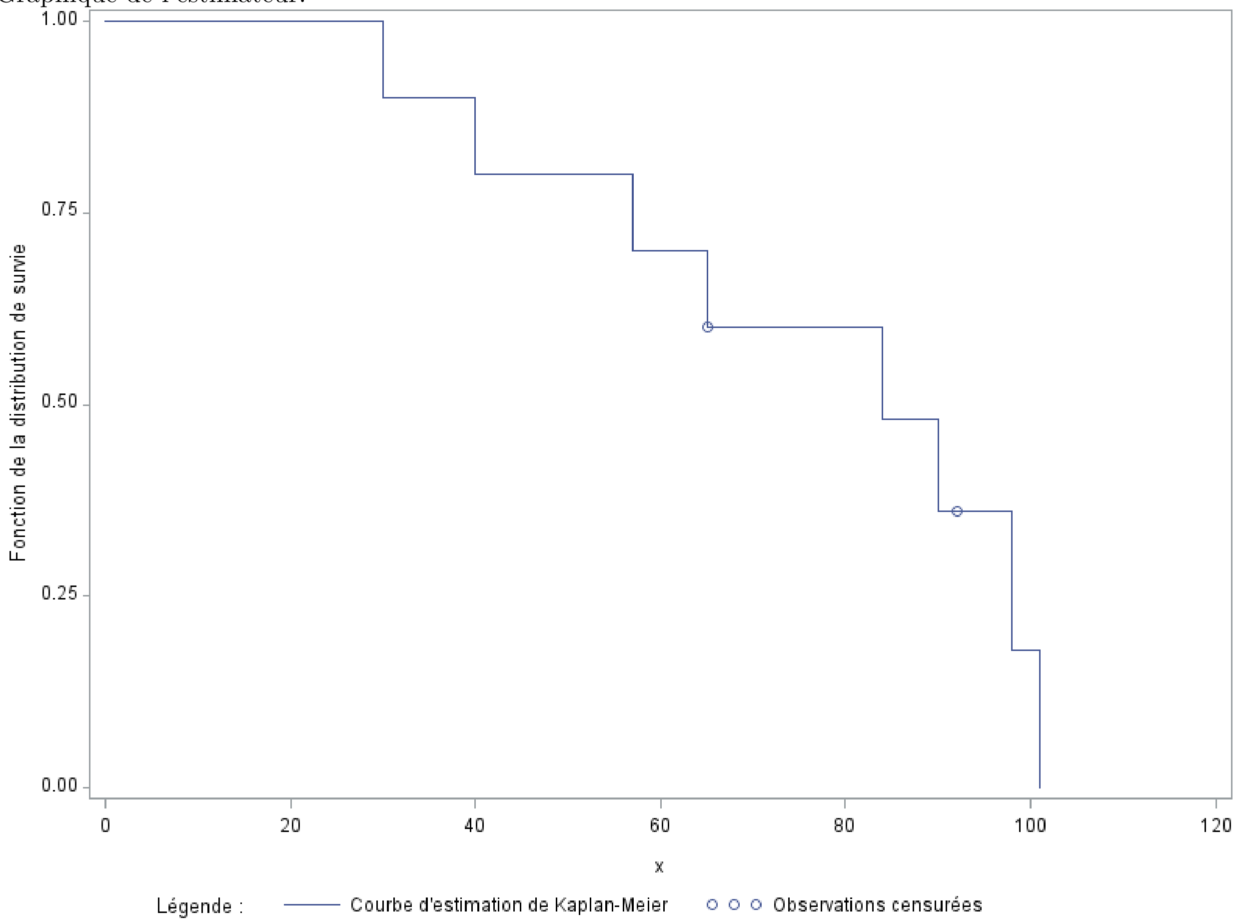
Estimateur de Kaplan-Meier:

Valeurs estimées de survie de Kaplan-Meier

x		Survie	Echec	Erreur type de survie	Nombre d'échecs	Nombre restant
0.000		1.0000	0	0	0	10
30.000		0.9000	0.1000	0.0949	1	9
40.000		0.8000	0.2000	0.1265	2	8
57.000		0.7000	0.3000	0.1449	3	7
65.000		0.6000	0.4000	0.1549	4	6
65.000	*	.	.	.	4	5
84.000		0.4800	0.5200	0.1640	5	4
90.000		0.3600	0.6400	0.1610	6	3
92.000	*	.	.	.	6	2
98.000		0.1800	0.8200	0.1506	7	1
101.000		0	1.0000	.	8	0

b)

Graphique de l'estimateur:



Intervalle de confiance, au niveau 95%, pour $S(50)$ est :

$$S(50) \in [S(\hat{50}) \pm z_{0.975}\sqrt{V\hat{a}r(S(\hat{50}))}]$$

où $V\hat{a}r(S(\hat{50}))$ est selon Greenwood $S(\hat{40})^2 * \sum_{i \leq 40} \frac{s_i}{r_i(r_i - s_i)}$, car $S(50) = S(40) = 0.8$

Avec les données actuelles, $V\hat{a}r(S(\hat{50})) = 0.016$

Ce qui donne un intervalle de confiance pour $S(50)$ de :

$$S(50) \in [0.552082, 1.047918]$$

c)

De façon semblable, on obtient que l'intervalle de confiance de la transformation $\log(-\log)$ est :

$$
\begin{aligned}
\ln(-\ln S(50)) &\in \ln(-\ln S(\hat{50})) \pm z_{0.975} \frac{\sqrt{\hat{Var}(S(\hat{50}))}}{S(\hat{50})\ln(S(\hat{50}))} \\
\ln(-\ln S(50)) &\in [-0.111159, -2.888721] \\
&\rightarrow S(50) \in [0.4086908, 0.9458726]
\end{aligned}
$$

On constate que l'amplitude de l'intervalle est similaire à celle en b). Cependant, celui-ci se trouve décalé vers la gauche (plus près de 0).

Annexe

Question 1

a)

```
sample <- rexp(100)
skewness <- function(x) mean((x-mean(x))^3)/sd(x)^3
exp_skew <- skewness(sample)
sample
```

```
##   [1] 1.035472507 1.614111246 0.707058484 2.936920389 0.369850264
##   [6] 1.675986074 3.071088573 1.022617095 0.305356620 0.358829146
##  [11] 1.567814231 1.049396855 0.400344273 0.460378464 2.260191711
##  [16] 0.793216118 0.806884154 0.006600642 0.533013725 0.260310964
##  [21] 5.870724067 0.754016150 2.821081765 0.133006924 0.732720044
##  [26] 0.419268578 0.998194736 0.694270298 0.751412192 2.075108586
##  [31] 0.751775928 0.016505957 1.130158346 0.070254717 1.064166225
##  [36] 1.693254361 0.012318579 0.566510031 1.053115022 1.479752184
##  [41] 0.464981359 0.811162606 0.621679070 1.426448096 1.528942050
##  [46] 0.699566848 0.947778532 0.434120066 0.303538297 1.439478101
##  [51] 0.142697353 2.598443830 1.616997990 2.023445515 1.236268514
##  [56] 1.011108878 0.588971156 0.821659701 1.891138539 0.236028785
##  [61] 2.148030391 0.649089325 0.331407073 0.644406911 2.634584157
##  [66] 0.088368135 0.009077653 2.655442031 0.011521576 1.682111081
##  [71] 3.065280244 0.142818669 1.064043199 0.287174706 0.191861316
##  [76] 1.911137948 1.231764817 1.575683296 0.891091234 0.804902516
##  [81] 0.203895254 2.643125934 0.586892240 1.463624533 1.721423238
##  [86] 1.256304624 1.267077568 0.149796479 2.449290054 1.141572217
##  [91] 3.348711639 0.320910250 0.317439921 0.757635029 2.102088047
##  [96] 0.066331663 1.978574566 0.154123089 0.231987935 0.262279843
```

b)

```
lambda <- 1/mean(sample)
skew <- sapply(1:50,function(x) skewness(rexp(100,lambda)))
var_skew <- var(skew)
IC <- sapply(c(-1,1),function(i) exp_skew + i*qnorm(0.975)*sd(skew))
```

Question 2

a)

```
prob <- c(0.25,0.35,0.5,0.6,0.75,0.85)
u <- sapply(prob,function(x) qexp(x,1))
limited_loss <- function(data,min)
    sapply(min,function(y) 1/length(data) *
               sum(sapply(data, function(x)  min(y,x))))
lim_exp <- limited_loss(sample,u)
```

b)

```
data1 <- lapply(1:50,function(x) rexp(100,lambda))
data2 <- lapply(1:50,function(x) rexp(100,lambda))
bootstrap_IC <- function(val,data){
    simul <- sapply(1:50,function(x) limited_loss(data[[x]],val))
    IC <- sapply(c(-1,1),function(i) limited_loss(sample,val) + i*qnorm(0.975)*sd(simul))
    IC
}
IC2 <- c(bootstrap_IC(qexp(0.5),data1),bootstrap_IC(qexp(0.75),data2))
data1[1]
```

```
## [[1]]
##   [1] 1.35940993 0.32018533 4.04773928 3.35459439 0.70810223 0.39999458
##   [7] 2.65927628 0.19589070 0.09083470 0.48975559 1.40720401 0.08925977
##  [13] 0.10900384 0.23275622 0.21981264 3.90835350 0.06594697 0.72652505
##  [19] 0.54330485 0.09440016 0.09369842 0.44755026 0.25157263 0.98605982
##  [25] 1.56191394 0.14002529 1.09978861 0.12967941 0.51564538 0.81753647
##  [31] 0.79934417 2.39684908 0.72365154 0.66947306 1.17891153 0.94086997
##  [37] 1.03037715 1.66142433 0.82771749 4.16234585 0.23859596 0.14785365
##  [43] 3.55269323 0.54122201 0.40614594 0.40422878 2.75098811 0.51488766
##  [49] 0.58509230 0.25243774 1.87246200 1.23659865 0.10370306 0.03333305
##  [55] 2.92408508 0.40434321 0.19752785 0.74039389 4.38712297 1.73810117
##  [61] 0.42526255 4.05504279 0.76182600 0.24788849 1.22597333 1.59904502
##  [67] 0.39781147 0.32378513 0.98956797 0.21533903 0.26975601 2.97499097
##  [73] 2.89561647 2.03513104 4.91602453 0.52112797 0.30050068 2.71908073
##  [79] 2.79144054 0.96474215 0.63389474 0.59339633 3.15207959 1.28309401
##  [85] 0.11502599 1.84224872 0.76174880 0.58328424 0.44531135 2.02413696
##  [91] 0.92564001 0.32363492 5.68035358 1.17562557 0.34758031 2.69461505
##  [97] 0.44557675 0.34659676 1.37463140 2.78662018
```

```
data2[1]
```

```
## [[1]]
##   [1] 0.0003170793 0.6006878794 0.3553665918 0.3602748089 0.4369071843
##   [6] 0.1163380771 0.7494483781 1.4503765700 0.3592394571 1.3473375118
##  [11] 0.0276334455 0.8738438314 0.7030796051 0.2719934925 1.4712713504
##  [16] 1.3915660449 2.2982083459 1.2678628597 0.0455607271 3.1743750031
##  [21] 7.4116542733 2.0663152905 0.0992615586 1.7073990569 1.2144618438
##  [26] 2.5981340520 1.2908340642 0.5128476076 0.7654689204 0.1928919242
##  [31] 6.4004419490 2.3999879449 0.5016746752 1.5447096787 1.6668120929
##  [36] 2.6042709615 3.2980284857 0.7621790565 1.6727957563 0.7770495267
```

```
## [41] 0.3898158625 0.4243805342 0.6646526662 2.4801007507 0.7675667146
## [46] 0.9561573515 0.6970333663 0.4197202475 1.1015435147 0.1674596656
## [51] 0.9503063842 0.8318097853 0.2133367996 1.1021614977 0.4486814137
## [56] 0.2868004363 0.3368568974 1.4069786684 1.8856468622 4.7792166414
## [61] 0.4547895470 0.1903092089 5.2497040123 0.7089377047 1.0273028848
## [66] 0.8673022069 0.0399893435 0.2399237427 0.0085351049 0.6700033344
## [71] 1.4588777815 0.2221090041 3.9955441286 0.0892703664 0.2739430317
## [76] 1.2844377518 2.1459683351 0.1212101479 0.2916816401 0.1535185780
## [81] 4.0841760556 0.1992413868 0.1105896074 0.3133097647 1.7101918514
## [86] 1.0638902042 1.2694212702 0.3320914767 0.4686383472 1.4468197719
## [91] 2.6858446319 0.2208113463 1.4469665067 0.8016041177 0.4511380586
## [96] 1.0942153087 0.6983743975 0.3126566270 2.4810153866 0.5063845252
```

Question 3

a)

```
data w1;*Numéro 3;

input x cens;

datalines;
        30          1
        40          1
        57          1
        65          1
        65          0
        84          1
        90          1
        92          0
        98          1
        101        1
;

proc lifetest data=w1 outsurv=wa
PLOTS=SURVIVAL;
time x*cens(0);
run;
```

Code sas:

b)

```
var <- 0.8^2 * ( 1/(10*9) + 1/(9*8))
IC <- sapply(c(-1,1),function(x) 0.8 + x*sqrt(var)*qnorm(0.975))
```

c)

```
IC2 <- sapply(c(-1,1), function(x) log(-log(0.8)) + x*qnorm(0.975)*sqrt(var)/(0.8*log(0.8)))
IC3 <- exp(-exp(IC2))
```