TP2 : Résolution de MDP

Paul Crespin

Politiques

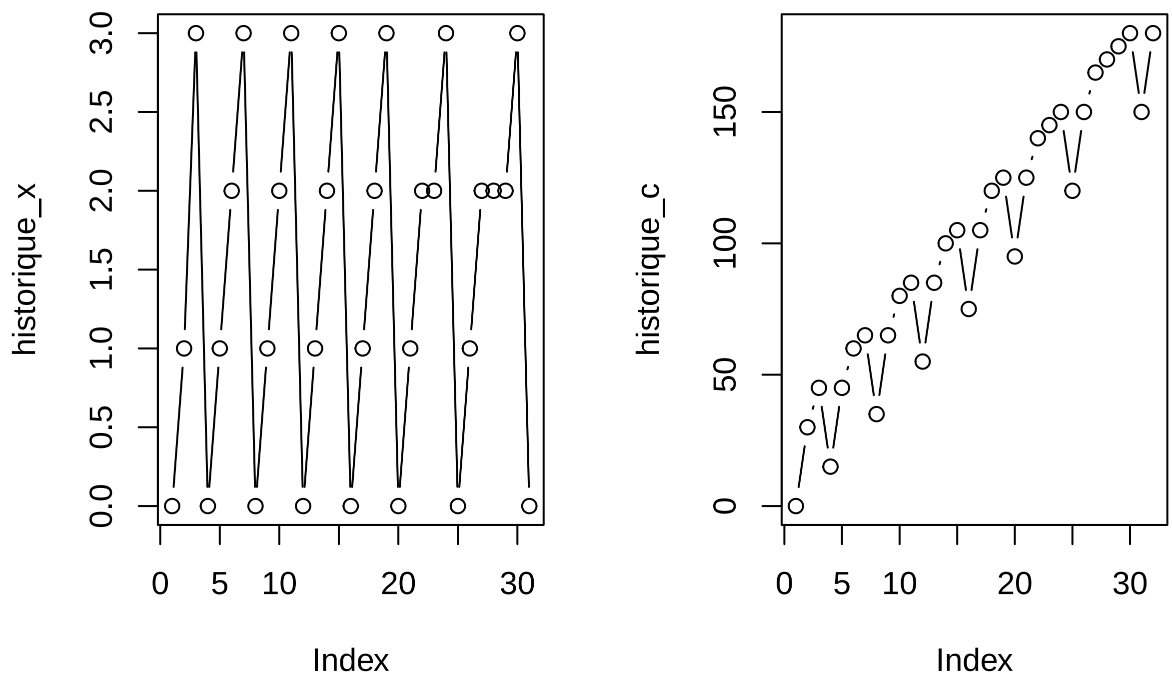
Question 1

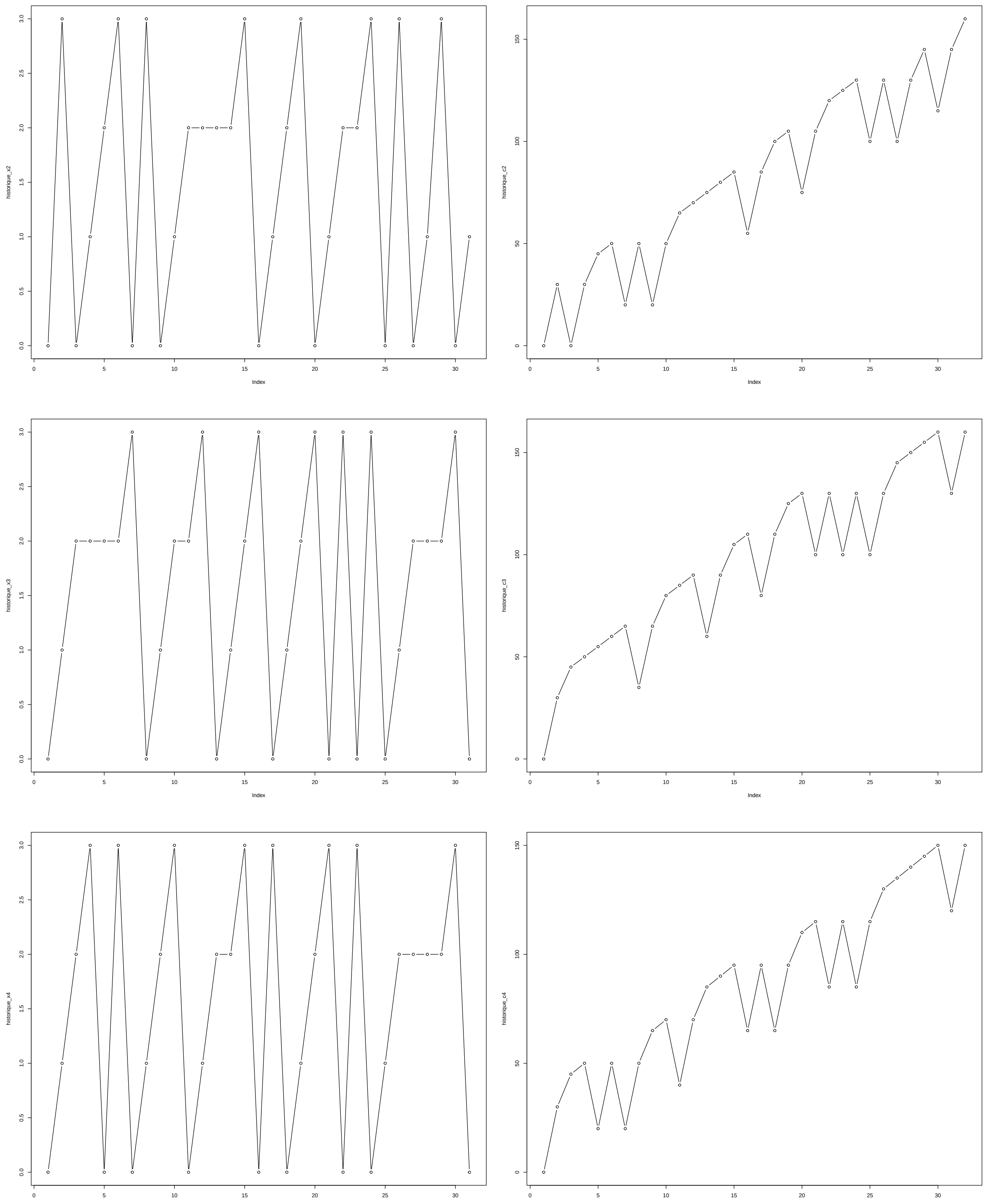
```
# On crée les variables de base
nb_jours=30
x=0
c=0
historique_x=c(x)
historique_c=c(c)

# On fait la boucle sur le nombre de jours avec une nouvelle loi uniforme chaque jour
#et une actualisation des gains/coûts selon l'état où la machine se trouve.

for (i in 2:(nb_jours+1)) {
  a=runif(1, min = 0, max = 1)

  if (x==0){if (a<0.2){x=3} else {x=1}
    c=c+30
  } else if (x==1){if (a<0.2){x=3} else {x=2}
    c=c+15
  } else if (x==2){if (a<0.5){x=3}
    c=c+5
  } else {x=0
    c=c-30}

  historique_x[i]=x
  historique_c[i]=c
}

# On ajoute la récompense finale
if (x==0){c=c+30
} else if (x==1){c=c+15
} else if (x==2){c=c+5
}

historique_c[nb_jours+2]=c
```

On fait partir le “for” de “2” car on a le gain nul du jour 0 à l’emplacement 1 de notre historique, le gain du jour 0 vers le jour 1 ira donc à la deuxième place. On va jusqu’à “jour+1” car on part de 2 et que le dernier jour, on a quand même le gain instantanée. On rajoute ensuite la récompense finale à l’emplacement “jour+2”.

Question 2

```r
par(mfrow=c(1,2))
plot(historique_x, type='b')
plot(historique_c, type='b')
```





On peut voir ici que la machine tombe souvent en panne et entraine régulièrement des pertes de 30 liés à la rénovation. Il y a également régulièrement des stagnations dans un mauvais état, ce qui entraîne de faibles gains.

Question 3

On va reprendre le même code qu'à la question 1 en rajoutant une boucle pour refaire x fois la simulation sur 30 jours et un historique du gain final:

```
nb_jours=30
nb_boucles=50000
p=matrix(nrow = 1, ncol = nb_boucles)

# On crée une matrice à 1 ligne et autant de colonnes que nécessaire.

for (ii in 1:nb_boucles){

  x=0
  c=0

  # Ces variables ont besoin d'être remises à 0 à chaque nouvelle boucle et on retire les historiques.

  for (i in 2:(nb_jours+1)) {
    a=runif(1, min = 0, max = 1)

    if (x==0){if (a<0.2){x=3} else {x=1}
      c=c+30
    } else if (x==1){if (a<0.2){x=3} else {x=2}
      c=c+15
    } else if (x==2){if (a<0.5){x=3}
      c=c+5
    } else {x=0
      c=c-30}
  }

  if (x==0){c=c+30
  } else if (x==1){c=c+15
  } else if (x==2){c=c+5
  }

  p[1,ii]=c

  # On rajoute le ii-ème gain final à la ii-ème place de la matrice.

}
```

On finit par faire la moyenne et la variance de toute la ligne.

```
## [1] "Estimation du gain:"
```

```
## [1] 170.6171
```

```
v=(mean(p^2)-mean(p)^2)/(nb_boucles-1)
# On ne multiplie pas par "nb_boucles" car on le divise par cette même quantité dans le calcul de l'IC
```

```
## [1] "IC"
```

```
## [1] 170.498
```

```
## [1] 170.7362
```

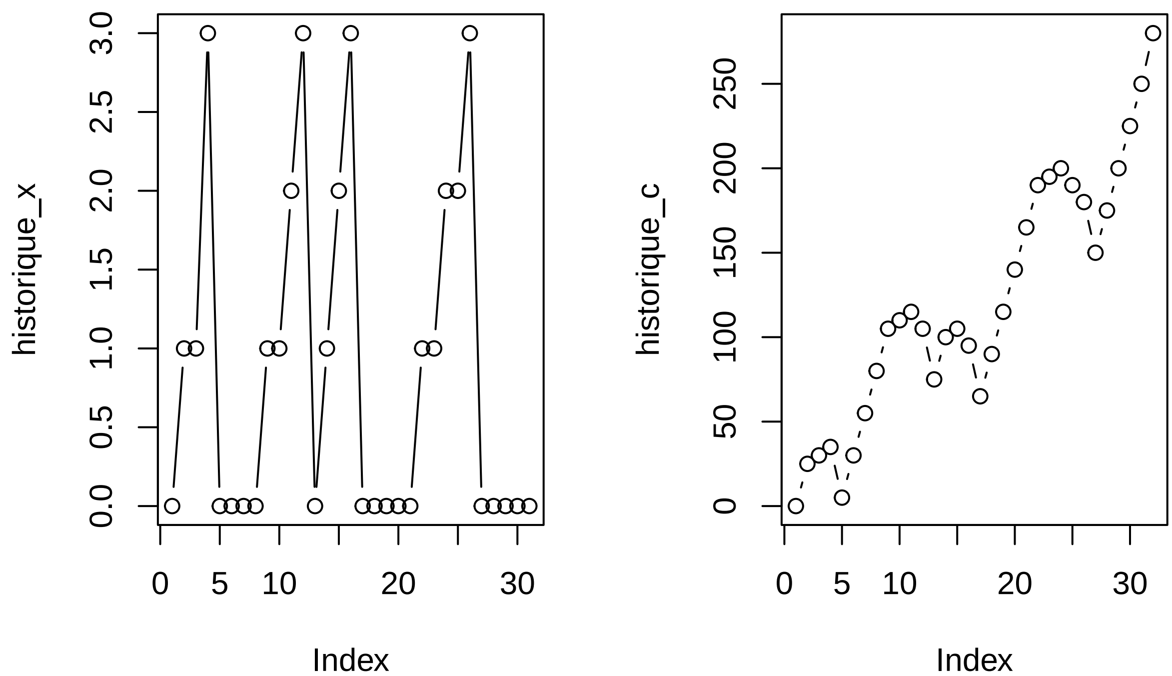
Question 4

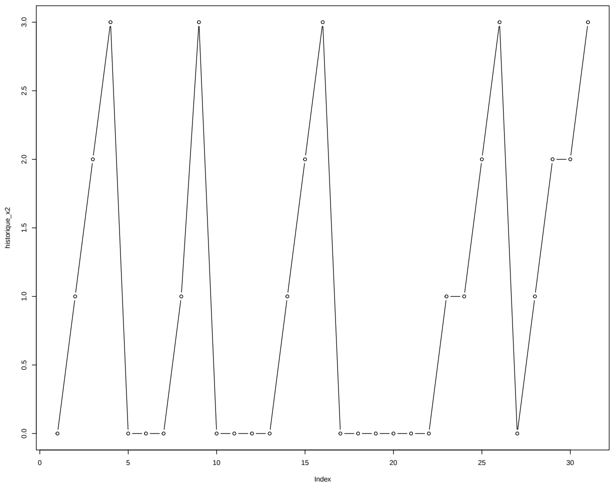
On refait le même code qu'à la question 1 en ajustant la politique:

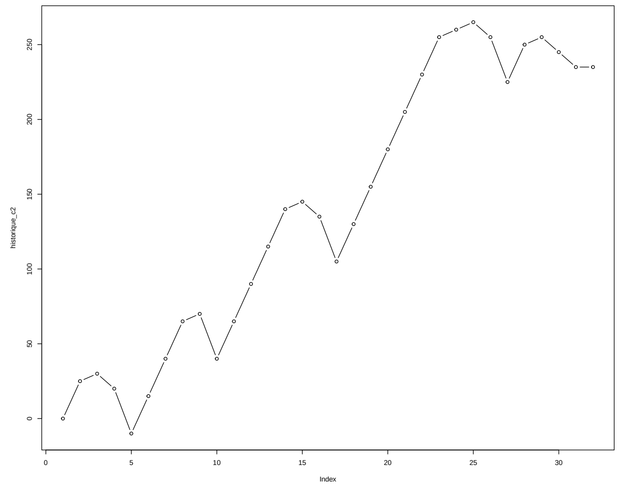
```
nb_jours=30
x=0
c=0
historique_c=c(c)
historique_x=c(x)

for (i in 2:(nb_jours+1)) {

  a=runif(1, min = 0, max = 1)

  if (x==0){

    if (a<0.25){x=1}
    c=c+25

  } else if (x==1){

    if (a<0.1){x=3} else if (a>0.6) {x=1} else {x=2}
    c=c+5

  } else if (x==2){

    if (a<0.75){x=3}
    c=c-10

  } else {x=0
    c=c-30}

  historique_x[i]=x
  historique_c[i]=c
}

if (x==0){
  c=c+30

} else if (x==1){
  c=c+15

} else if (x==2){
  c=c+5

}

historique_c[nb_jours+2]=c
```

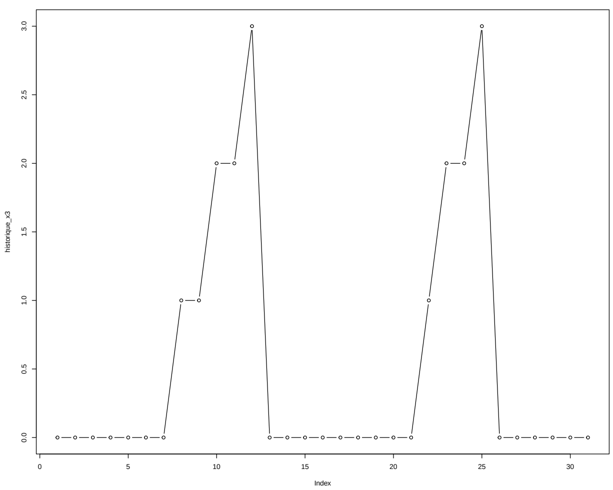
Quesion 5

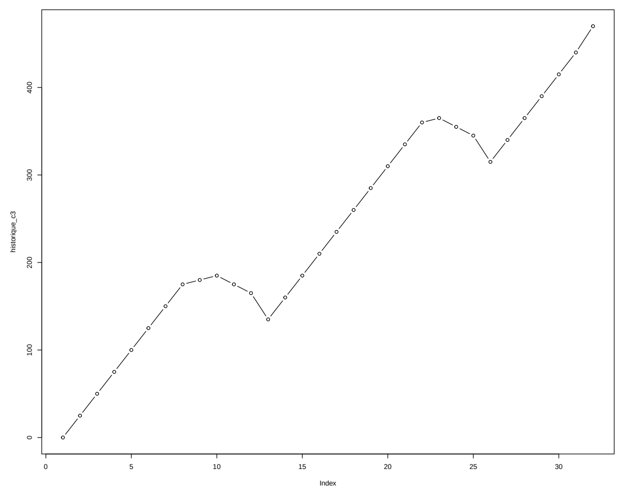
```
par(mfrow=c(1,2))
plot(historique_x, type='b')
plot(historique_c, type='b')
```

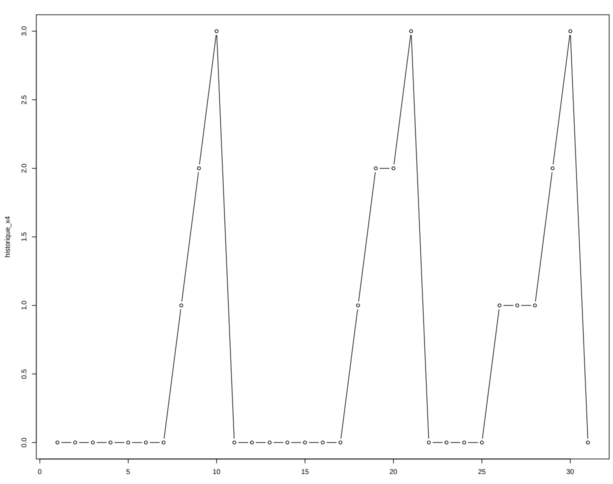


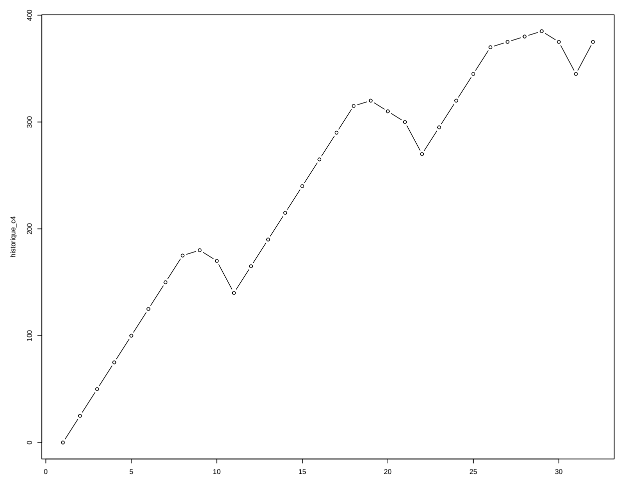












On peut voir que le gain “à l’air” d’être supérieur à la politique précédente. La machine tombe moins souvent en panne, mais l’entretien pendant le non et mauvais état ne semble pas être une bonne idée car on a des chance de maintenir plus longtemps un petit gain ou une petite perte de 5, ce qui est faible.

Question 6

Même principe que pour la question 3:

```
nb_jours=30
nb_boucles=50000
p=matrix(nrow = 1, ncol = nb_boucles)

for (ii in 1:nb_boucles){

  x=0
  c=0

  for (i in 2:(nb_jours+1)) {
    a=runif(1, min = 0, max = 1)

    if (x==0){if (a<0.25){x=1}
     c=c+25
    } else if (x==1){if (a<0.1){x=3} else if (a>0.6) {x=1} else {x=2}
      c=c+5
    } else if (x==2){if (a<0.75){x=3}
     c=c-10
    } else {x=0
      c=c-30}
  }

  if (x==0){c=c+30
  } else if (x==1){c=c+15
  } else if (x==2){c=c+5
  }

  p[1,ii]=c

}
```

```
## [1] "Estimation du gain:"
```

```
## [1] 306.9966
```

```
v=(mean(p^2)-mean(p)^2)/(nb_boucles-1)
```

```
## [1] "IC"
```

```
## [1] 305.9826
```

```
## [1] 308.0106
```

Question 7

On peut voir que la politique 1 est de manière quasiment certaine presque 1.8 plus rentable sur 30 jours que la politique 0.

Optimisation

Question 8

On reprend l'algorithme du cours et on le modifie pour nos besoins:

```
N=30

# Matrices de transition
Q <- array(0, dim=c(4 ,4 ,3 ))

# Ne rien faire : action 0
Q[ , ,1]<- matrix(c(0 ,0.8 ,0 ,0.2 ,
                    0 ,0 ,0.8 ,0.2 ,
                    0 ,0 ,0.5 ,0.5,
                    0 ,0 ,0 ,0), nrow=4, byrow=TRUE)

# Entretien : action 1
Q[ , ,2]<- matrix(c(0.75 ,0.25 ,0 ,0 ,
                    0 ,0.6 ,0.3 ,0.1 ,
                    0 ,0 ,0.75 ,0.25,
                    0 ,0 ,0 ,0), nrow=4, byrow=TRUE)

# Rénovation : action 2
Q[ , ,3]<- matrix(c(0 ,0 ,0 ,0 ,
                    1 ,0 ,0 ,0 ,
                    1 ,0 ,0 ,0 ,
                    1 ,0 ,0 ,0), nrow=4, byrow=TRUE)

#Chacune des lignes de 0 rajoutées aurait pu être n'importe quel nombre,
#cela n'a pas d'importance (pas besoin de respecter les propriétés
#de matrice de transition pour ces lignes)

# Récompense instantane
rec <- matrix(c(30 ,15 ,5 ,-Inf ,
                25 ,5 ,-10 ,-Inf ,
                -Inf ,-5 ,-25 , -30), nrow=4)

# Récompense finale
gain <- c(30, 15, 5, 0)

V <- matrix(0 ,4 ,N+1)
V[,1]=gain #On ajoute la récompense finale à la première colonne.
pi <- matrix(0 ,4 ,N)
for (n in 1:N) {
  for (x in 1:4) {
    tmp <- rec[x,]+V[, n]%*%Q[x, ,]
    pi[x, n] <- which.max(tmp)
    V[x, n+1] <- max(tmp)
  }
}

pi=pi-1

#Pour correspondre aux actions 0;1;2.
```

Ainsi, on peut voir qu'on peut espérer un gain optimal de 601.8 en lisant V[1,31], en partant d'une machine neuve.

La performance optimale V_N (0) est de 601.8.

On peut voir qu'on retrouve le même résultat en écrivant la politique optimale manuellement, par la méthode de Monte Carlo (en lisant pi):

```
## [1] "Estimation du gain:"

## [1] 601.5879

## [1] "IC"

## [1] 601.1464

## [1] 602.0294
```

Question 9

On peut voir que la politique optimale est environ 3.5 fois plus efficace que la politique 0 et environ un peu moins de 2 fois efficace de la politique 1. C'est cohérent car les deux politiques étudiées ont été prises un peu au hasard.

Question 10

La politique optimale sur 30 jours en commençant par une machine neuve consiste à rénover à chaque étape sauf dans 2 cas:

-Si la machine est dans l'état neuf (il est illégal de la rénover dans ce cas de toute façon). Ici, il réaliser l'entretien de la machine.

-Au dernier jour, si la machine est dans un mauvais état. Ici, il faut ne rien faire.

C'est cohérent car le gain moyen sera de 7.5 (5 de gain immédiat et moitié/moitié entre 5 et 0 de gain final) face à une réparation de -25 et un gain de 30 (gain sûr de 5) ou une perte nette avec un entretien. (Cela n'arrive jamais dans nos simulations de toute façon car la politique appliquée nous fait rester entre l'état neuf et bon état.)