Mini-Projet Statistique

ZHAO Fubang(Group 4)

19:45:28 25 oct 2016

Je vous prie par avance de me pardonner si je commets des erreurs en français.

Exercice 1

```
set.seed(42,kind = "Marsaglia-Multicarry")
```

1.1

Le $g(\theta)$ pour $0 < \theta < 1$ est

$$\begin{aligned} g(\theta) &= \sum_{k=1}^{+\infty} k\theta(1-\theta)^{k-1} \\ &= \theta \sum_{k=1}^{+\infty} -\frac{\partial(1-\theta)^k}{\partial x_1} \\ &= -\theta \frac{1}{\partial\theta}(\frac{1}{1-(1-\theta)}) \\ &= \frac{1}{\theta} \end{aligned}$$

1.2

Par le mesure de comptage sur $\mathbb{N}^* = \{1, 2, 3, ..., \}$ le modèle est dominé. Parce que $\mathbb{N}^* = \{1, 2, 3, ..., \}$ est dénombrable.

1.3

Car le modèle $\{P\theta, \theta \in]0, 1[\}$ est régulier, on a

$$var_\theta[T(x)] \geq \frac{g'(\theta)^2}{I(\theta)}$$

Et la variance de $T(x)$ est

$$\begin{aligned} var_\theta[T(x)] &= var_\theta(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n} var_\theta(x_1) \\ &= \frac{1-\theta}{n\theta^2} \end{aligned}$$

Car $(X_1, ..., X_n)$ sont n *v.a. i.i.d.* au sens des hypothèses du théorème de Cramér-Rao, la quantité d'information de Fisher est

$$
\begin{aligned}
I(\theta) = nI_1(\theta) &= n\mathbb{E}_\theta[(\frac{\partial log\theta(1-\theta)^{k-1}}{\partial\theta})^2] \\
&= n\mathbb{E}_\theta[(\frac{1}{\theta} - \frac{k-1}{1-\theta})^2] \\
&= nvar(\frac{k-1}{1-\theta}) \\
&= \frac{nvar(x)}{(1-\theta)^2} \\
&= \frac{n}{(1-\theta)\theta^2}
\end{aligned}
$$

Et le paramètre d'intérêt $g(\theta)$ est

$$g(\theta) = \mathbb{E}_\theta(T(x)) = \mathbb{E}_\theta(X_1) = \frac{1}{\theta}$$

On peut conclure: l'estimateur $T(X)$ de $g(\theta)$ est non biaisé et vérifie

$$var_\theta[T(x)] = \frac{g'(\theta)^2}{I(\theta)}$$

DoncT(X) atteint la borne de Cramér-Rao, C'est à dire, $T(x)$est un estimateur UVMB.

1.4

Le risque quadratique de S_h est

$$
\begin{aligned}
R(\theta, S_h) &= \mathbb{E}_\theta[(g(\theta) - S_h(X))^2] \\
&= \mathbb{E}_\theta[(\frac{1}{\theta} - hT(X))^2] \\
&= \frac{1}{\theta^2} - \frac{2h}{\theta^2} + h^2(\frac{1}{\theta^2} + var(T(X))) \\
&< R(\theta, T) = var(T(X))
\end{aligned}
$$

C'est à dire:

$$\theta^2(h^2-1)var(T(X) + (h-1)^2 < 0$$

Si $h > 1$, on a

$$h\, n'existe\, pas$$

Si $h < 1$, on a

$$\frac{n-1+\theta}{n+1-\theta} < h < 1$$

Alors quand on a $\frac{n-1+\theta}{n+1-\theta} < h < 1$, on va avoir $R(\theta, S_h) < R(\theta, T)$. Parce que$R = \frac{1}{\theta^2} - \frac{2h}{\theta^2} + h^2(\frac{1}{\theta^2} + var(T(X)))$ est une fonction quadratique. Donc evidemment,

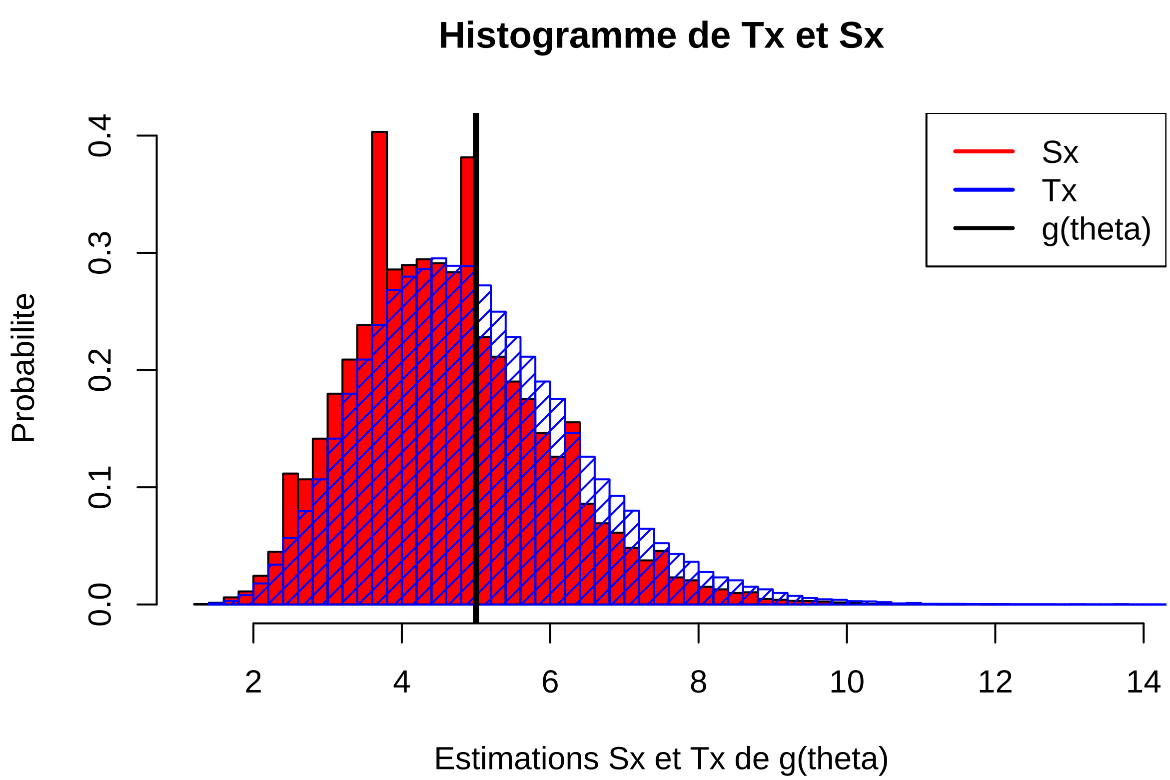
$$h^*(\theta) = \frac{1}{\theta^2 var(T(X)) + 1} = \frac{n}{n+1-\theta}$$

1.5

Car $h^*(\theta) = \frac{1}{\theta^2 var(T(X))+1} = \frac{n}{n+1-\theta}$, c'est à dire h^* est une fonction de θ, donc il n'exsite pas h^* qui minimse le risque quadratique uniformément en θ.

1.6

```
#Exercice1
#a
h = 10/(11-0.2)
theta_0 = 0.2
n = 10
M = 10**5
Z = matrix(nrow = n, ncol = M)
for(i in 1:M)
{
  v= rgeom(10, theta_0)
  Z[,i] = matrix(v, nrow=10)+1
}
#b
Tx = colMeans(Z)
Sx = h*Tx
#c
hist(Sx, col = "red", probability = TRUE,
     main = "Histogramme de Tx et Sx", breaks = 50,
     xlab = "Estimations Sx et Tx de g(theta)",
     ylab = "Probabilite")
hist(Tx, col = "blue", probability = TRUE, add = TRUE,
     density = 15, breaks = 50)
abline(v = 1/theta_0, lwd = 3)
legend("topright", lwd = 2,col = c("red", "blue", "black"),
       legend = c("Sx", "Tx","g(theta)"))
```



```r
mean(Sx)
```

```
## [1] 4.627897
```

```r
mean(Tx)
```

```
## [1] 4.998129
```

```r
#d
L1 = (Tx-mean(Tx))**2
Lh = (Sx-mean(Tx))**2
```

```r
#e
#R(theta_0,T)
(1-theta_0)/(n*(theta_0**2))
```

```
## [1] 2
```

```r
#R(theta_0,S)
(1/theta_0**2+(1-theta_0)/(n*(theta_0**2)))*h**2-(2/theta_0**2)*h+1/theta_0**2
```

```
## [1] 1.851852
```

```r
mean(L1)#l'approximation de R(theta_0,T)
```

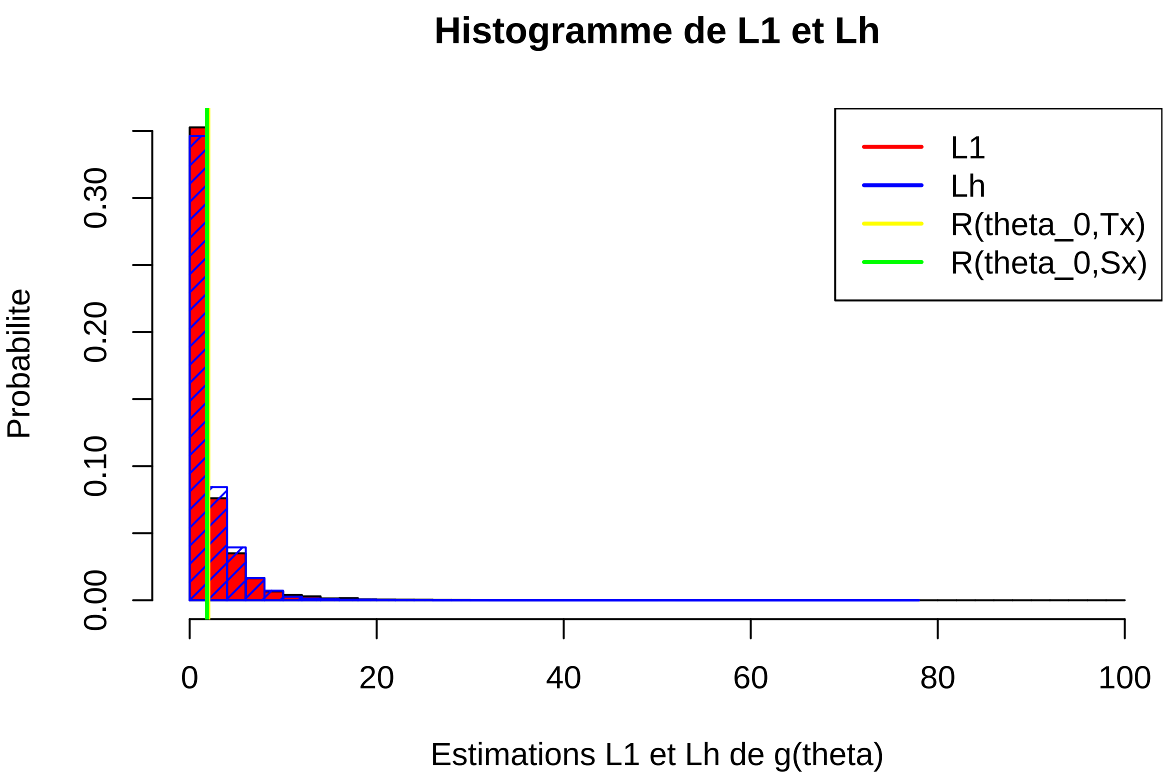
```
## [1] 1.992093
```

```r
mean(Lh)#l'approximation de R(theta_0,S)
```

```
## [1] 1.84497
```

Avec $h = \frac{10}{11-0.2}$, qui est inclus dans $\frac{n-1+\theta}{n+1-\theta} < h < 1$, on a toujours $R_S < R_T$. Cela, même pour les estimations, est conformé à la conclusion de la quesiton 4.

```r
#f
hist(L1, col = "red",
     probability = TRUE,
     main = "Histogramme de L1 et Lh",
     breaks = 50,
     xlab = "Estimations L1 et Lh de g(theta)",
     ylab = "Probabilite")
hist(Lh, col = "blue", probability = TRUE, add = TRUE,
     density = 15, breaks = 50)
abline(v = mean(L1), lwd = 2, col = "yellow")
abline(v = mean(Lh),lwd = 2, col ="green")
legend("topright", lwd = 2,col = c("red", "blue", "yellow", "green"),
       legend = c("L1", "Lh", "R(theta_0,Tx)", "R(theta_0,Sx)"))
```



Exercice 2

2.1

Car le résultat est à nouveau une loi Beta, dont on dois préciser les paramètres

$$\begin{aligned}\pi(\theta|x) &= \frac{p_\theta(x)\pi(\theta)}{m(x)} \\ &\propto p_\theta(x)\pi(\theta) \\ &= \prod_{i=1}^{n} p_\theta(x_i)\pi(\theta) \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^n(1-\theta)^{\Sigma_{i=1}^n x_i - n} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+n-1}(1-\theta)^{\beta+\Sigma_{i=1}^n x_i - n-1}\end{aligned}$$

Donc $\pi(\theta|x)$ est une loi Beta: $Beta(\alpha+n, \beta+\Sigma_{i=1}^n x_i - n)$.

2.2

$$\mathbb{E}_\pi(\theta|x) = \frac{\alpha+n}{\alpha+\beta+\Sigma_{i=1}^n x_i}$$

Car $\mathbb{E}_\pi(\theta|x)$ est une fonction de X, on peut considérer $M(X) = E_\pi(\theta|X)$ comme un estimateur de θ_0.

2.3

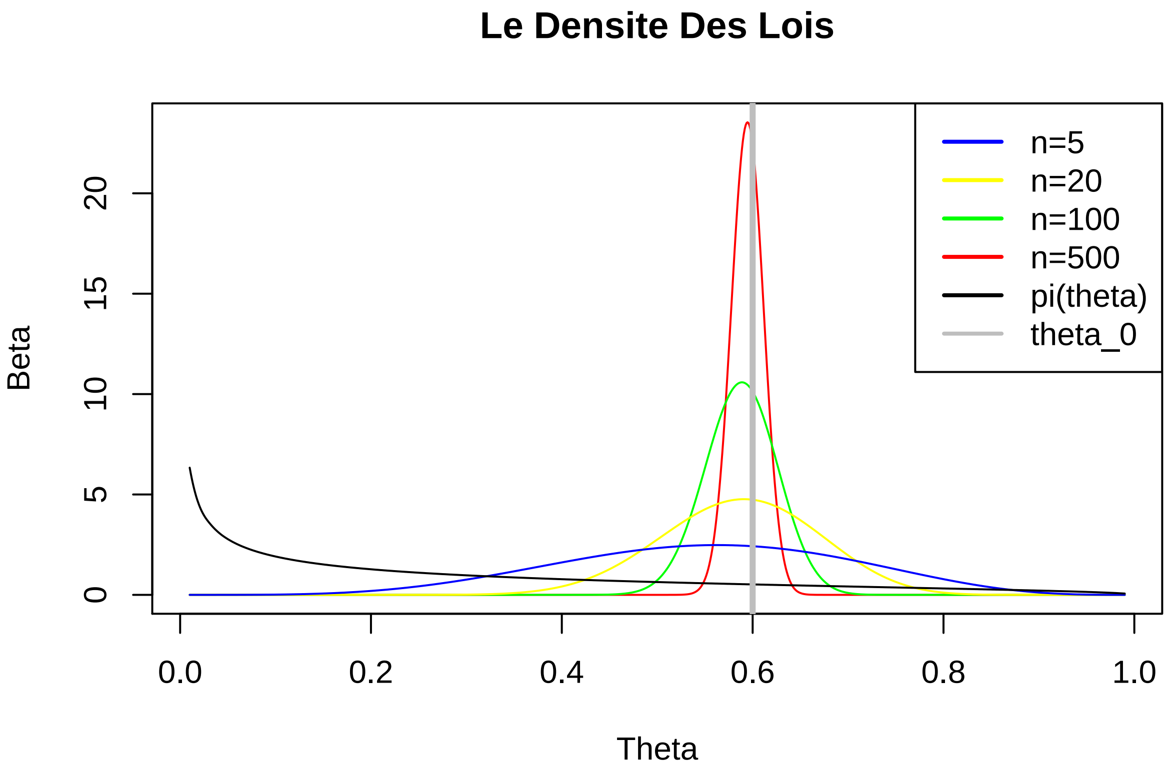
$$\mathbb{E}_\pi(\theta|x) = \frac{\frac{\alpha}{n}+1}{\frac{\alpha+\beta}{n}+\frac{1}{n}\Sigma_{i=1}^n x_i}$$

Car la loi des grands nombres, lorsque $n \to +\infty$, $\mathbb{E}_\pi(\theta|x) \to \frac{1}{\bar{X}} = \frac{1}{\frac{1}{\theta}} = \theta$.

Donc $M_n = \mathbb{E}_\pi(\theta|X^n)$ converge $\mathbb{P}_{\theta_0} - \textit{presque}$ sûrement vers θ_0.

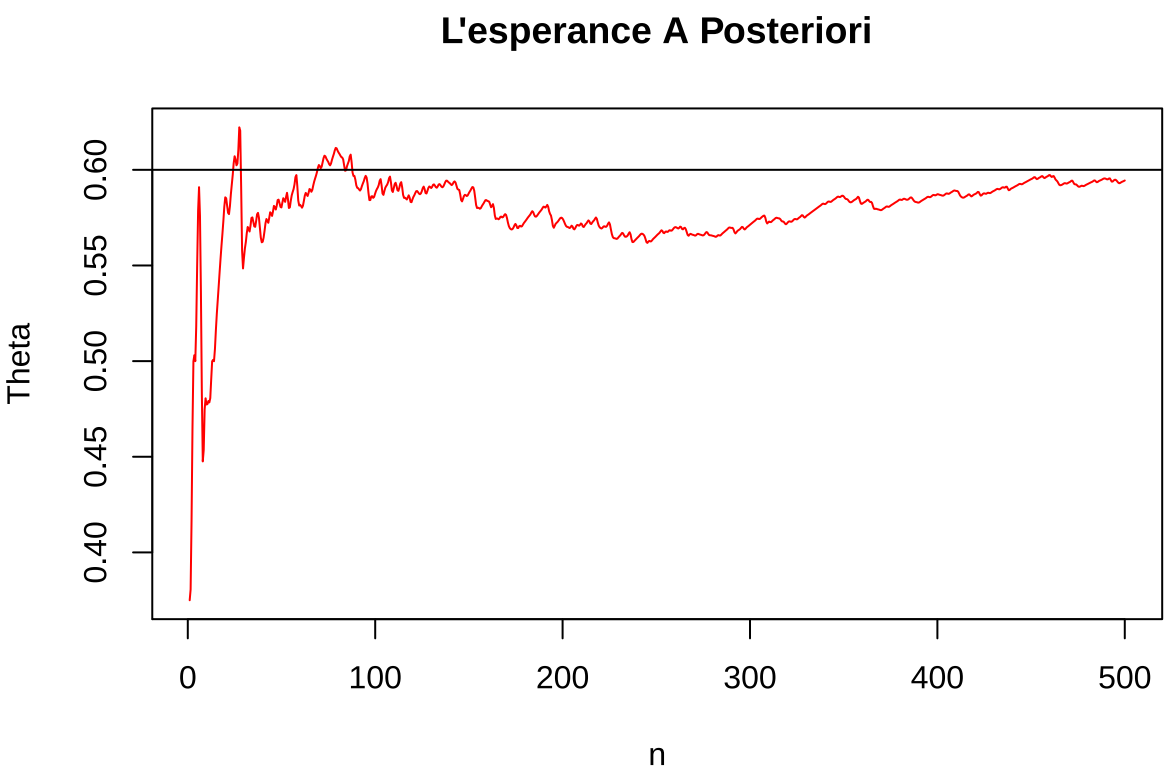
2.4

```r
#Excercice2
#a
grille = seq(0, 1, by = 0.01)
L = length(grille)
grille = grille[-c(1,L)]
x= rgeom(500, 0.6)+1
#b
n_pi = 500
v_beta = dbeta(grille, 1/2+n_pi, 3/2-n_pi+sum(x[1:n_pi]))
sp = spline(grille, v_beta, n=1000)
plot(sp,type="l", col="red", xlab = "Theta", ylab = "Beta",
     main = "Le Densite Des Lois")
n_pi = 100
v_beta = dbeta(grille, 1/2+n_pi, 3/2-n_pi+sum(x[1:n_pi]))
sp = spline(grille, v_beta, n=1000)
lines(sp, type="l", col="green")
n_pi = 20
v_beta = dbeta(grille, 1/2+n_pi, 3/2-n_pi+sum(x[1:n_pi]))
sp = spline(grille, v_beta, n=1000)
lines(sp, type="l", col="yellow")
n_pi = 5
v_beta = dbeta(grille, 1/2+n_pi, 3/2-n_pi+sum(x[1:n_pi]))
sp = spline(grille, v_beta, n=1000)
lines(sp, type="l", col="blue")

v_pi = dbeta(grille, 1/2, 3/2)
sp = spline(grille, v_pi, n=1000)
lines(sp, type="l", col="black")

abline(v=0.6, col="grey", lwd = 3)
legend("topright", lwd = 2,col = c("blue", "yellow", "green", "red", "black", "grey"),
       legend = c("n=5", "n=20", "n=100", "n=500", "pi(theta)","theta_0"))
```



On peut voir que, lorque que n est plus grand, l'ésperance s'approche plus du vrai paramètre θ_0 et la courbe est plus proche de la line de θ_0 .

```
#c
E_theta_x = (1/2+(1:500))/(2+cumsum(x[1:500]))
sp = spline(1:500, E_theta_x, n=1000)
plot(sp,type="l", col="red", xlab = "n", ylab = "Theta",
     main = "L'esperance A Posteriori")
abline(h=0.6)
```



On peut voir que, lorsque n augmente, l'ésperance a posterior converge vers le vrai $\theta_0 = 0.6$.

Exercice 3

3.1

L"espace des paramètresΘ du modèle est $\Theta = \{\mu \geq 0\}$, $\Theta_0 = \{\mu = 0\}$, $\Theta_1 = \{\mu > 0\}$.

3.2

On a

$$S_n(X) \sim \mathcal{N}(\mu, \delta^2)$$

On a $\mathbb{E}_{H_0}[S_n(X)] = 0$, $var_{H_0}(S_n(X)) = \frac{1}{n}$, donc la loi de $S_n(X)$ est

$$S_n(X) \sim \mathcal{N}(0, \frac{1}{n})$$

On a

$$\sqrt{n}S_n(X) \sim \mathcal{N}(\mu, \delta^2)$$

On a $\mathbb{E}_{H_0}[\sqrt{n}S_n(X)] = 0$, $var_{H_0}(\sqrt{n}S_n(X)) = 1$, donc la loi de $\sqrt{n}S_n(X)$ est

$$\sqrt{n}S_n(x) \sim \mathcal{N}(0, 1)$$

Pour que la probabilité de rejeter à tort H_0 soit inférieure ou égale à $\alpha = 5/100$, on a

$$\int_{-\infty}^{\sqrt{n}A} p(\sqrt{n}S_n(x))\,\mathrm{d}(\sqrt{n}S_n(x)) \geq 1 - \frac{0.05}{2} = 0.975$$

Donc

$$A = \frac{F^{\leftarrow}(0.975)}{\sqrt{n}}$$

3.3

```r
#Exercice3
#3
a = 0.05
p_a = 1-(a/2)
A_10 = qnorm(p_a,sd = 1/10**0.5)
A_100 = qnorm(p_a,sd = 1/100**0.5)
A_1000 = qnorm(p_a, sd = 1/1000**0.5)
A_10;A_100;A_1000
```

```
## [1] 0.619795
```

```
## [1] 0.1959964
```

```
## [1] 0.0619795
```

3.4

L'espérance de X est

$$\mathbb{E}(x) = \mathbb{E}(e^{logx}) = \int_0^{+\infty} e^{logx} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(logx-\mu)^2}{2\sigma^2})\mathrm{d}(logx)$$

On utilise y à replacer $logx$, donc

$$\begin{aligned}
\mathbb{E}(x) &= \int_0^{+\infty} e^{logx} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(logx-\mu)^2}{2\sigma^2})\mathrm{d}(logx) \\
&= \int_{-\infty}^{+\infty} e^{y} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(y-\mu)^2}{2\sigma^2})\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(y-(\sigma^2+2\mu))^2}{2\sigma^2} + \frac{\sigma^2}{2} + \mu)\mathrm{d}y \\
&= e^{\mu+\frac{\sigma^2}{2}}
\end{aligned}$$

Car $\sigma^2 = 1$, on a:

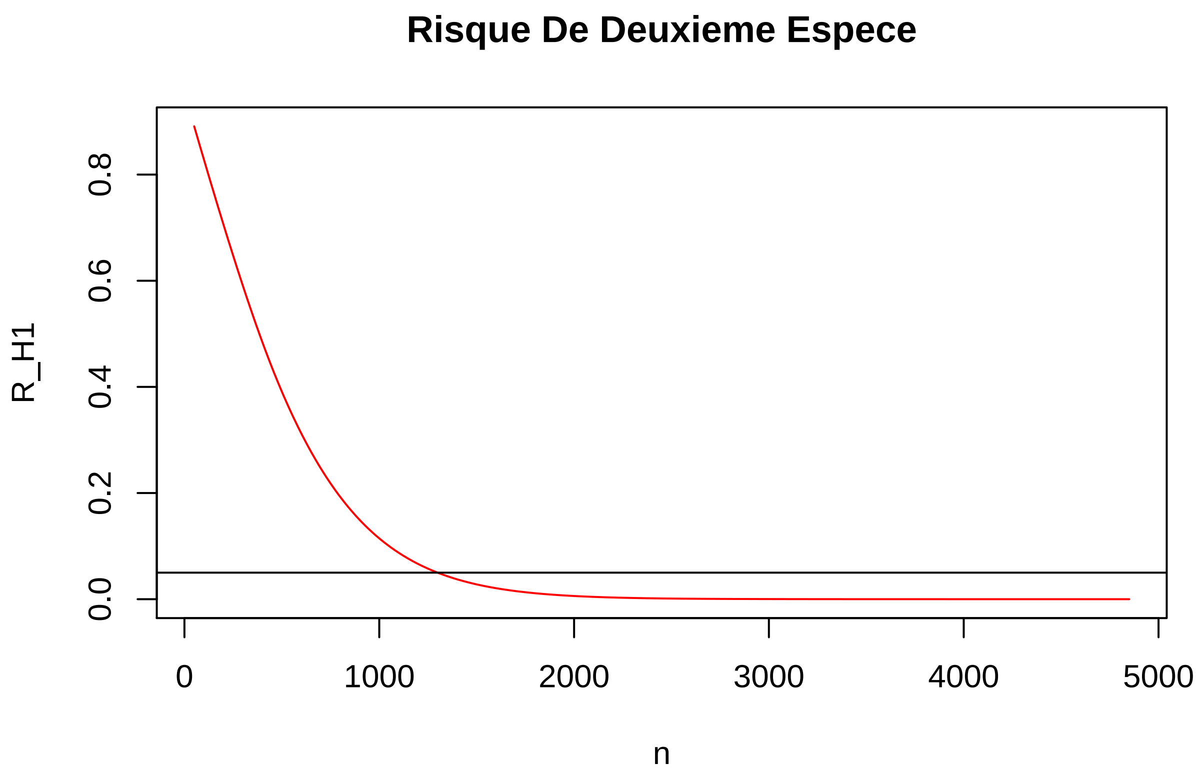
$$\mathbb{E}(x) = e^{\mu+\frac{1}{2}}$$

```r
#4
exp(0.1+0.5)
```

```
## [1] 1.822119
```

3.5

```r
#5
grille_risque = seq(50,5000,200)
R_H1 = c()
for (n_3 in grille_risque)
{
  s = pnorm(qnorm(p_a,sd = 1/n_3**0.5), mean = 0.1, sd=1/n_3**0.5)
  b = pnorm(-qnorm(p_a,sd = 1/n_3**0.5), mean = 0.1, sd=1/n_3**0.5)
  v= s-b
  R_H1 = append(R_H1, v)
}
sp = spline(grille_risque, R_H1, n=1000)
plot(sp,type="l", col="red", xlab = "n", ylab = "R_H1",
     main = "Risque De Deuxieme Espece")
abline(h=0.05)
```




```r
n_0 = 0
for (n_3 in grille_risque) {
  s = pnorm(qnorm(p_a,sd = 1/n_3**0.5), mean = 0.1, sd=1/n_3**0.5)
  b = pnorm(-qnorm(p_a,sd = 1/n_3**0.5), mean = 0.1, sd=1/n_3**0.5)
  v= s-b
  if(v<=0.05)
  {
    n_0 = n_3
    break
  }
}
n_0
```

```
## [1] 1450
```

Donc pour n variant de 50 à 5000, sur une grille de pas $h = 200$, on a la première approximation de la plus petite valeur de $n_0 = 1450$.

3.6

```
#6
for (n_3 in 1:5000) {
  s = pnorm(qnorm(p_a,sd = 1/n_3**0.5), mean = 0.1, sd=1/n_3**0.5)
  b = pnorm(-qnorm(p_a,sd = 1/n_3**0.5), mean = 0.1, sd=1/n_3**0.5)
  v= s-b
  if(v<=0.05)
  {
    n_0 = n_3
    break
  }
}
n_0
```

```
## [1] 1300
```

Donc la valeur exacte de n_0 est 1300.