Statistique : étude de cas

CC1 : 8 octobre 2018

Corrigé

Exercice 1 :

1. Les variables aléatoires $X_i, i \in \{1, ..., 5\}$ sont indépendantes, donc :

$$\begin{aligned} L(\lambda) = & \quad \mathbb{P}_\lambda(X_1 = 0, X_2 = 1, X_3 = 0, X_4 = 0, X_5 = 0) \\ = & \quad \mathbb{P}(X_1 = 0) \times \mathbb{P}(X_2 = 1) \times \mathbb{P}(X_3 = 0) \times \mathbb{P}(X_4 = 0) \times \mathbb{P}(X_5 = 0) \\ = & \quad \frac{\lambda^0}{0!}e^{-\lambda}\frac{\lambda^1}{1!}e^{-\lambda}\frac{\lambda^0}{0!}e^{-\lambda}\frac{\lambda^0}{0!}e^{-\lambda}\frac{\lambda^0}{0!}e^{-\lambda} \\ = & \quad \lambda e^{-5\lambda} \end{aligned}$$

2.

λ	$L(\lambda)$
0.5	0.0410425
0.2	0.07357589
0.1	0.06065307

3. Relativement à un échantillon $(X_i)_{i=1,...,n}$, la vraisemblance d'un paramètre λ est :

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{X_i}}{X_i!}e^{-\lambda}$$

Sa log-vraisemblance s'écrit alors :

$$l(\lambda) = ln(L(\lambda)) = -n\lambda + ln(\lambda)\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} ln(X_i!)$$

et sa dérivée est :

$$l'(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i$$

Dressons son tableau de variations :

λ		$\frac{1}{n}\sum_{i=1}^{n} X_i$	
l'	$+$	0	$-$
$l(\lambda)$	$\nearrow$	$\|$	$\searrow$

L'estimateur du maximum de vraisemblance de λ est donc $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Sa réalisation sur l'échantillon de la question 1. est $\overline{X} = 0.2$, ce qui est cohérent avec les calculs de la question 2.

4.

	μ	σ^2
$\overline{X}^2$	$3.1628\ 10^{-2}$	$4.352905\ 10^{-3}$
$\overline{X}^2 - \frac{S_c^2}{n}$	$1.092\ 10^{-2}$	$1.912945\ 10^{-3}$
$\overline{X}^2 - \frac{\overline{X}}{n}$	$1.1024\ 10^{-2}$	$1.758967\ 10^{-3}$

Le premier estimateur semble biaisé tandis que les deux autres produisent des estimations de qualités comparables. Le dernier semble avoir une variance plus faible, ce qui dénote une incertitude plus faible. Celui-ci est donc préférable aux deux autres.

Le code utilisé est :

```
M = matrix(NA,10000,5)
mu = rep(NA,10000)
sigma2 = rep(NA,10000)

for (i in 1:10000) {
M[i,] = rpois(5,0.1)
mu[i] = mean(M[i,])
sigma2[i] = var(M[i,])
}

e1 = mu**2
e2 = mu**2 - sigma2/5
e3 = mu**2 - mu/5
```

Exercice 2 :

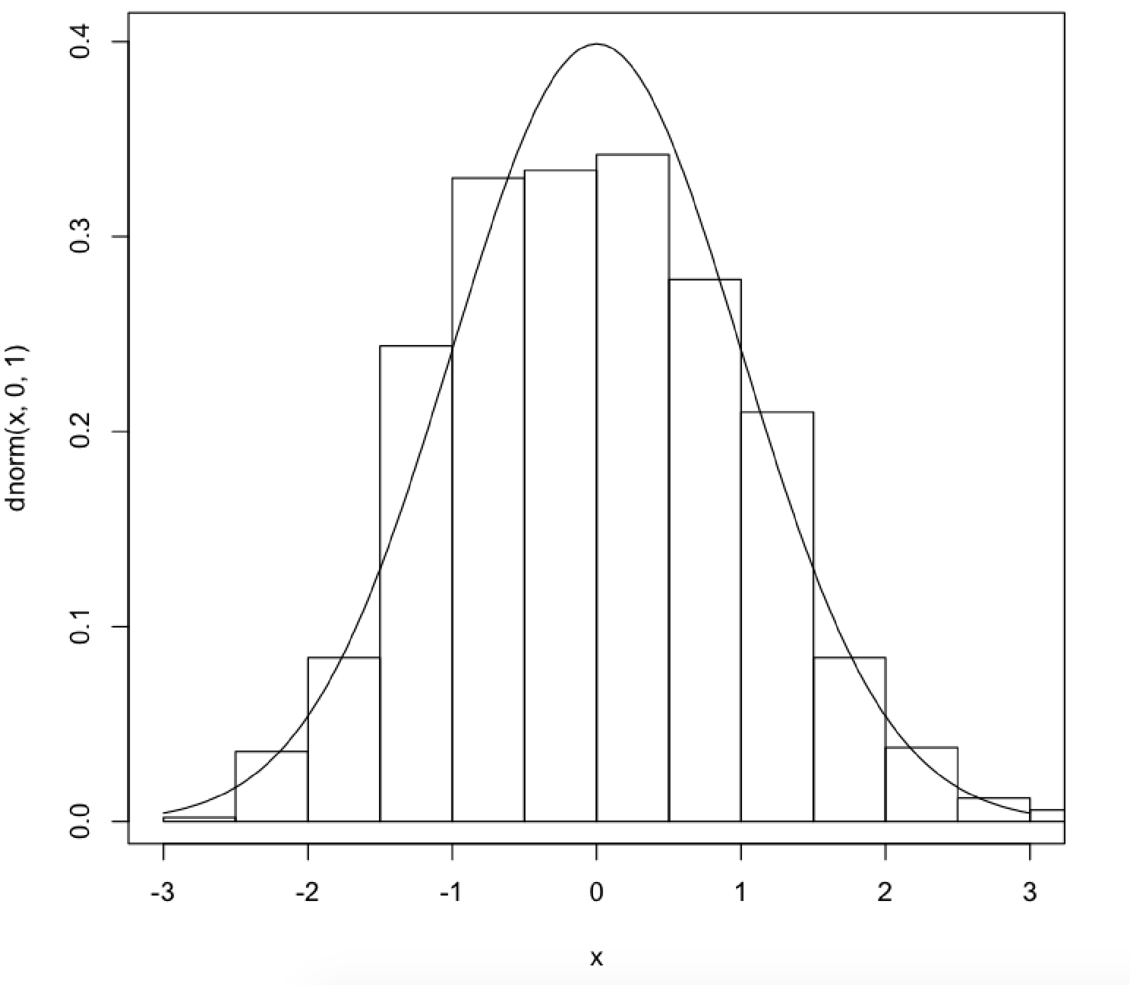
1. Calculons :

$$\begin{aligned} F_\lambda(x) = & \ \mathbb{P}(X \leqslant x) \\ = & \ \mathbb{P}\left(-\frac{ln(U)}{\lambda} \leqslant x\right) \\ = & \ \mathbb{P}\left(U \geqslant \exp(-\lambda x)\right) \\ = & \ 1 - \mathbb{P}\left(U \leqslant \exp(-\lambda x)\right) \\ = & \ 1 - \exp(-\lambda x)\mathbb{1}_{\{\exp(-\lambda x)<1\}} = 1 - \exp(-\lambda x)\mathbb{1}_{\{x>0\}} \end{aligned}$$

   Dérivons :

$$f_\lambda(x) = \frac{d}{dx}F_\lambda(x) = \lambda\exp(-\lambda x)\mathbb{1}_{\{x>0\}}$$

2. 3. 4.



5. Le théorème illustré ici est le théorème central limite.
   Notons F la fonction de répartition associées à la loi normale centrée réduite $\mathcal{N}(0,1)$.

   Théorème : Si $(X_i)_i$ est une suite de variables aléatoires iid de moyenne μ et d'écart-type $\sigma \neq 0$, et si $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$, alors :
$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{(n)}} \leqslant x\right) = F(x)$$

6. On obtient les estimations :

$\hat{\mu}$	2.979461
$\hat{\sigma^2}$	8.774494

   qui sont cohérentes avec les valeurs théoriques $\mu = 3$ et $\sigma^2 = 9$.

Dans cet exercice, on a utilisé le code suivant :

```
M=matrix(NA,1000,50)
mu = rep(NA,1000)
sigma2 = rep(NA,1000)
Z = rep(NA,1000)

for (i in 1:1000) {
        for (j in 1:50) {
                M[i,j] = -3*log(runif(1))
        }
        mu[i] = mean(M[i,])
        sigma2[i] = var(M[i,])
        Z[i] = sqrt(50)*(mu[i]-3)/3
}

curve(dnorm(x,0,1),-3,3)
hist(Z,prob=T,add=T)

print(mean(mu))
print(mean(sigma2))
```