Statistics with R
AIMS Senegal

Tabea Rebafka
October 2018

Feuille de TP N° 3

Comparaison d'estimateurs

---

1 Simulation de lois de probabilité

Il est très utile en statistique de savoir générer des réalisations de diverses lois de probabilité. Ceci se fait sous `R` via les fonctions de la forme `r`*func*(`n,param`) où *func* indique la loi de probabilité, `n` est le nombre de réalisations à générer et `param` sont les paramètres de la loi. En voici quelques exemples, où `n` désigne le nombre de réalisations indépendantes de la loi :

Loi	Fonction `R`
uniforme $U[0,1]$	`runif(n)`
uniforme $U[a,b]$	`runif(n, a, b)`
normale $\mathcal{N}(0,1)$	`rnorm(n)`
normale $\mathcal{N}(m,s^2)$	`rnorm(n, m, s)`
Poisson $\mathcal{P}(\lambda)$	`rpois(n, lambda)`
binomiale $\mathrm{Bin}(k,p)$	`rbinom(n, k, p)`
exponentielle $\mathcal{E}(\theta)$	`rexp(n, theta)`
Gamma $\Gamma(a,b)$	`rgamma(n, a, b)`
Cauchy(θ)	`rcauchy(n, 0, theta)`

Les fonctions de la forme `r`*func* (avec *func*=`norm` ou `unif`...) ont toutes des petites sœurs de la forme

— `d`*func*(`x,arguments`) : renvoie la densité de probabilité en `x` quand *func* désigne une loi continue. Pour les lois discrètes, la probabilité de prendre la valeur `x` est renvoyée.

— `p`*func*(`x,arguments`) : renvoie la fonction de répartition en `x`,

— `q`*func*(`a,arguments`) : renvoie le quantile d'ordre `a`,

Exercice 1.

1. Calculez la valeur de la densité d'une loi normale centrée réduite au point 0.
2. Calculez la probabilité $\mathbb{P}(X \leq 0)$ pour $X \sim \mathcal{N}(0,5)$.
3. Générez 300 observations de la loi exponentielle $\mathcal{E}(4)$, stockez-les dans une variable nommée `data.exp`. Tracez l'histogramme de cet échantillon.
4. Utilisez la fonction `curve()` pour superposer la densité de la loi exponentielle $\mathcal{E}(4)$ à l'histogramme des données `data.exp`. Interprétez le résultat.
5. Générez 300 observations de la loi de Poisson $\mathcal{P}(2)$, stockez-les dans une variable nommée `data.pois`. Tracez le diagramme en bâtons pour cet échantillon par la commande

```
plot(table(data.pois)/300)
```

6. Évaluez les probabilités $\mathbb{P}(X = k)$ d'une variable aléatoire X de la loi de Poisson $\mathcal{P}(2)$ pour tout $k \in \{0, 1, \ldots, 10\}$. Superposez ces probabilités au diagramme en bâtons des données `data.pois`. Interprétez le résultat.

2 Programmation sous R

Concernant la programmation sous R lire le notebook `ProgrammationSousR.ipynb` (ou le pdf *Programmation_en_R.pdf*).

3 Convergence de la moyenne empirique

Le but de l'exercice est d'approcher la valeur de l'erreur quadratique moyenne théorique *MSE* d'un estimateur $\hat{\theta}$ de θ définie par

$$MSE = \mathbb{E}_\theta \left[(\hat{\theta} - \theta)^2 \right].$$

Pour cela, nous pouvons simuler s échantillons $\mathbf{x}_1 = (x_{1,1}, \ldots, x_{1,n}), \ldots, \mathbf{x}_s = (x_{s,1}, \ldots, x_{s,n})$ indépendants, de taille n chacun, de loi $\mathbb{P}_\theta$. Sur chaque échantillon, on évalue l'estimateur $\hat{\theta}$. Notons $\hat{\theta}_k = \hat{\theta}(\mathbf{x}_k)$ pour $k = 1, \ldots, s$ les valeurs de l'estimateur $\hat{\theta}$ associées aux différents échantillons. Une approximation de l'erreur quadratique moyenne théorique *MSE* est alors donnée par

$$\widehat{MSE} = \frac{1}{s} \sum_{k=1}^{s} (\hat{\theta}_k - \theta)^2.$$

Exercice 2.

1. Écrire une fonction nommée `mean_exp` qui prend en argument la taille d'échantillon `nb` et le paramètre `param` de la loi exponentielle. Cette fonction génère $s = 100$ échantillons de la loi exponentielle $\mathcal{E}(\texttt{param})$ et renvoie un vecteur composé de moyennes empiriques associées à ces échantillons.
2. Écrire une fonction nommée `mse` qui prend en argument un vecteur d'estimateurs `estimates` et la vraie valeur du paramètre `param`. Cette fonction renvoie l'erreur quadratique moyenne $\widehat{MSE}$ associée à ces estimations.
3. Appelez la fonction `mean_exp()` avec les arguments `param` $= 2$ et `nb` $= 100, 1000$ et 10000, et stockez les résultats dans des variables différentes. Calculez les valeurs de l'erreur quadratique moyenne pour les trois résultats et comparez.
4. Tracez sur un même graphique les boxplots associés aux vecteurs renvoyés par la fonction `mean_exp()`.
5. Commentez les résultats obtenus. Quel résultat de la théorie des probabilités explique ce phénomène ?

Exercice 3. Écrivez une fonction `mean_cauchy` qui fait la même chose que la fonction `mean_exp` mais pour des réalisations d'une loi de Cauchy. Comparez les valeurs de l'erreur quadratique moyenne et les boxplots des sorties de cette fonction pour différentes valeurs de `nb`. Que remarque-t-on ? Expliquez ce phénomène.