Simulations stochastiques - TD 7

Derniers exercices sur les tests et Révisions

Exercice 1 (Test d'adéquation de Kolmogorov)**.** Soit $(X_1,\dots,X_n)$ un n-échantillon de fonction de répartition F. On note $\mathbb{F}_n$ la fonction de répartition empirique associée :

$$\forall t\in\mathbb{R},\ \mathbb{F}_n(t)=\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i\leqslant t}\,.$$

Soit $F_{\text{réf}}$ une fonction de répartition continue. On souhaite tester $H_0 : F = F_{\text{réf}}$ contre $H_1 : F \neq F_{\text{réf}}$ sur la base de l'observation de $(X_1,\dots,X_n)$. On rappelle que le test d'adéquation de Kolmogorov est donné par $\mathbb{1}_{D_n>d_{n,1-\alpha}}$, où $d_{n,1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi de D_n sous H_0, et où

$$D_n=\sup_{t\in\mathbb{R}}\left|\mathbb{F}_n(t)-F_{\text{réf}}(t)\right|\,.$$

1. Rappeler pourquoi, sous H_0, la loi de D_n ne dépend en fait pas de la fonction de répartition $F_{\text{réf}}$. Quelle conséquence intéressante cela a-t-il sur le quantile $d_{n,1-\alpha}$?
2. Nous allons montrer que le test est consistant, i.e., si H_1 est vraie, alors $\mathbb{P}(D_n > d_{n,1-\alpha})\to 1$ quand $n\to+\infty$.
   (a) En utilisant le théorème de Glivenko-Cantelli, montrer que $\lim\limits_{n\to+\infty} d_{n,1-\alpha}=0$.
   (b) On suppose maintenant que H_1 est vraie, et on fixe $t\in\mathbb{R}$ tel que $F(t)\neq F_{\text{réf}}(t)$. Montrer que, presque sûrement, $\liminf\limits_{n\to+\infty} D_n \geqslant \left|F(t)-F_{\text{réf}}(t)\right|$.
   (c) En déduire que le test est consistant. Qu'est-ce que cela signifie en pratique ?
3. On souhaite tester si une variable aléatoire X suit une loi normale $\mathcal{N}(2,1)$. On observe échantillon de taille 21 :

   0.3 0.7 0.9 1.2 1.4 1.4 1.5 1.5 1.6 1.9 2.0 2.1 2.1 2.3 2.5 2.6 2.7 3.0 3.8 3.9 4.0

   Que vaut $D_n(\omega)$ sur cet exemple ? Donner une expression simple à calculer sur un ordinateur.

Exercice 2 (Test de Kolmogorov-Smirnov pour deux échantillons indépendants)**.** Soit $(X_1,\dots,X_n)$ un n-échantillon de fonction de répartition F, et $(Y_1,\dots,Y_m)$ un m-échantillon de fonction de répartition G et indépendant de $(X_1,\dots,X_n)$. On suppose que F et G sont continues sur $\mathbb{R}$. On note $\mathbb{F}_n$ la fonction de répartition empirique de $(X_1\dots,X_n)$, $\mathbb{G}_m$ la fonction de répartition empirique de $(Y_1,\dots,Y_m)$ et on définit

$$D_{n,m}=\sup_{t\in\mathbb{R}}|\mathbb{F}_n(t)-\mathbb{G}_m(t)|.$$

On s'intéresse aux hypothèses H_0 : "$F=G$" et H_1 : "$F\neq G$".

1. Que traduit l'égalité $F = G$ sur les lois des X_i et des Y_j ?
2. Soit $U_1, \dots, U_n, V_1, \dots, V_m$ des variables aléatoires i.i.d. uniformes sur $[0, 1]$. On notera F^{-1} et G^{-1} les inverses généralisées des fonctions F et G. Montrer que $\big(F^{-1}(U_1), \dots, F^{-1}(U_n)\big)$ est un n-échantillon de fonction de répartition F, que $\big(G^{-1}(V_1), \dots, G^{-1}(V_m)\big)$ est un m-échantillon de fonction de répartition G, et que ces deux échantillons sont indépendants.
3. En déduire que, si $F = G$, alors la loi de $D_{n,m}$ ne dépend pas de F.
4. En utilisant le fait que la loi de $D_{n,m}$ est connue sous H_0 (cette loi est tabulée), construire un test de H_0 contre H_1.
5. On souhaite comparer deux médicaments censés soulager la douleur post-opératoire. Sur 16 patients dont 8 ont pris un médicament A habituel et les 8 autres un médicament B expérimental, on a observé les temps de soulagement suivants (en heures). Y a-t-il une différence significative au niveau 5% entre A et B ?

A	6.8	3.1	5.8	4.5	3.3	4.7	4.2	4.9
B	4.4	2.5	2.8	2.1	6.6	0.0	4.8	2.3

Exercice 3. On notera Φ la fonction de répartition d'une loi gaussienne centré réduite. Dans la suite, on considère une famille de lois de probabilités $(\mathbb{P}_\theta, \theta \in]0, 1[)$ ayant pour fonction de répartition

$$F_\theta(x) = (1-\theta)\Phi\left(\frac{x-m}{\sigma}\right) + \theta\Phi\left(\frac{x-m}{5\sigma}\right), \quad x \in \mathbb{R}$$

Dans la suite, on suppose que l'on dispose d'un échantillon $X_1, \dots, X_n$ de variables aléatoires i.i.d. distribuées selon $\mathbb{P}_\theta$ pour un θ que l'on cherche à estimer. On supposera les paramètres m et $\sigma > 0$ connus.

1. Quelle est la loi d'une variable aléatoire X distribuée selon $\mathbb{P}_0$? selon $\mathbb{P}_1$?
2. Expliquer pourquoi la loi de X distribuée selon P_θ admet une densité et la calculer.
3. Calculer $\mathbb{E}_\theta(X_1)$.
4. Calculer $\mathrm{Var}_\theta(X_1)$.
5. Au vu des questions précédentes, proposer un estimateur fortement consistent de θ à partir de l'échantillon $X_1, \dots, X_n$.
6. Proposer une méthode de simulation de l'échantillon $X_1, \dots, X_n$.

Les exercices suivants correspondent à l'examen de l'année 2016

Exercice 4. Soit U une variable aléatoire uniforme sur $[0, 1]$. La variable $2U - 1$ est alors uniforme sur $[-1, 1]$. De la même manière, pour chacune des lois de probabilités suivantes, déterminer une variable aléatoire uniquement fonction de U et ayant cette loi. Justifier les réponses.

1. La loi uniforme sur $[-3, 12]$.
2. La loi d'une variable égale à 0 avec probabilité $1/2$, 1 avec probabilité $1/3$, -1 avec probabilité $1/6$.

3. La loi d'une variable aléatoire réelle dont la fonction de répartition en x vaut $1/2 + (\arctan(x)/\pi)$ (on admettra que cette fonction est une bijection de $\mathbb{R}$ vers $(0,1)$, par ailleurs la loi de probabilité définie ici s'appelle *loi de Cauchy*).

Exercice 5. Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires i.i.d. uniformes sur $[-1,1]^2$. On note K le plus petit indice k tel que $X_k \in [0,1]^2$.

1. Déterminer la probabilité que $K = k$ pour tout $k \geqslant 1$: comment appelle-t-on la loi de K ?
2. Quelle est la loi de X_K ? (pas de démonstration demandée)

Exercice 6. On considère l'intégrale suivante :

$$I = \int_0^1 \int_0^1 \int_0^1 f(x,y,z)dxdydz$$

avec

$$f(x,y,z) = \sin(e^{3x-4y+5z} + e^{x-3y+7z} - 4xy).$$

1. Comment voir, sans calcul, que f est bien intégrable sur $[0,1] \times [0,1] \times [0,1]$?
2. On se donne des variables aléatoires i.i.d. uniformes sur $[0,1]$, notées $(X_k)_{k\geqslant 1}$, $(Y_k)_{k\geqslant 1}$, $(Z_k)_{k\geqslant 1}$. On pose

$$T_k := f(X_k, Y_k, Z_k).$$

   Déterminer une suite de variables aléatoires ne dépendant que de $(T_k)_{k\geqslant 1}$ et convergeant presque sûrement vers I.
3. Comment voir, à nouveau sans calcul, que la variance de T_k est inférieure ou égale à 1 ?
4. Déterminer une suite $(I_n)_{n\geqslant 1}$ de variables aléatoires, ne dépendant que de $(T_k)_{k\geqslant 1}$, telle que pour $n \geqslant 1$,

$$\mathbb{P}[|I - I_n| \leqslant 5/\sqrt{n}] > 0.95$$

   et pour tout n suffisamment grand,

$$\mathbb{P}[|I - I_n| \leqslant 2/\sqrt{n}] > 0.95.$$

   On justifiera bien sûr ces deux inégalités.

Exercice 7. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires i.i.d., de loi exponentielle de paramètre λ (avec $\lambda > 0$). On notera $\mathbb{E}_\lambda$, Var_λ et $\mathbb{P}_\lambda$ les espérances, variances et probabilités calculées avec le paramètre λ.

Partie 1

1. Montrer que $\mathbb{E}_\lambda[X_1] = 1/\lambda$ et $\mathrm{Var}_\lambda[X_1] = 1/\lambda^2$.
2. On considère $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Montrer que, pour tout $n \geqslant 1$, $\mathbb{P}(\bar{X}_n > 0) = 1$.
3. Soit $\alpha \in]0,1[$. Montrer que, pour tout $n \geqslant 1$,

$$\mathbb{P}_\lambda\left(\left|\bar{X}_n - \frac{1}{\lambda}\right| > \frac{1}{\lambda\sqrt{n\alpha}}\right) \leqslant \alpha \ .$$

4. En déduire un intervalle de confiance de niveau α pour λ (valable pour tout $n \geqslant 1$).

Partie 2
On s'intéresse maintenant au test d'hypothèses suivant :

$$H_0 : \lambda \leqslant \lambda_0$$
$$H_1 : \lambda > \lambda_0$$

5. Soit φ_n un test de l'hypothèse H_0 contre l'hypothèse H_1. Comment est défini le risque de première espèce de φ_n ?
6. Soit $t \in \mathbb{R}$ un réel quelconque.
   (a) Rappeler comment on peut simuler une v.a. de loi exponentielle de paramètre λ à partir d'un v.a. de loi $\mathcal{U}([0,1])$.
   (b) Montrer par couplage que la fonction $\lambda \mapsto \mathbb{P}_\lambda(\bar{X}_n < t)$ est une fonction croissante sur $\mathbb{R}_+^*$. On admettra qu'elle est aussi continue.
7. Soit $\alpha \in]0,1[$. Monter que le test φ_n défini par

$$\varphi_n(X_1, \ldots, X_n) = \mathbb{1}_{\left\{\bar{X}_n < \frac{1}{\lambda_0} - \frac{1}{\lambda_0\sqrt{n\alpha}}\right\}}$$

   est de niveau α.
8. Application numérique : on souhaite tester si $\lambda \leqslant 1$ ou $\lambda > 1$. On observe un échantillon de taille $n = 10000$, de moyenne empirique égale à 0,6. Que peut-on en déduire au niveau $\alpha = 1\%$? Par ailleurs, que signifie ce 1% concrètement ?