Université Pierre et Marie Curie — **Année 2018–2019**
Licence - Mathématiques - 3LM246 — Processus et simulations

Série d'exercices N°2
Méthode de Monte Carlo, rudiments de statistiques

Exercice 1 : Quantiles. Soit X une va de fonction de répartition continue telle que $-X$ a la même loi que X. Soit, pour $x \in]0,1[$, $q_x \in \mathbb{R}$ tel que $P(X \leq q_x) = 1 - x$, (les q_x sont appelées les *quantiles* de la loi de X). Montrer que pour tout $x \in]0, 1/2[$,

$$P(|X| \leq q_x) = 1 - 2x.$$

Exercice 2 : Intervalle de confiance pour la méthode de Monte-Carlo (exercice non-obligatoire). Soit (X_i) une suite de v.a.i.i.d de loi uniforme sur $[0,1]^d$ et f une fonction continue sur $[0,1]^d$ vers $\mathbb{R}$. Soit, pour $x \in]0,1[$, $q_x \in \mathbb{R}$ tel que $P(N \leq q_x) = 1-x$, où N est une va gaussienne centrée réduite sur $\mathbb{R}$ (les q_x sont appelées les *quantiles* de la loi gaussienne standard). On pose, pour tout n, $S_n(f) = \sum_{i=1}^n f(X_i)$.

En utilisant un théorème cours, donner, en fonction de $q_{0,005}$, de $||f||_\infty$, et de $Var(f(X_1))$, un entier n_0 à partir duquel la probabilité que $S_n(f)/n$ soit à moins de $1/1000$ de l'intégrale de f contre la mesure de Lebesgue soit de au moins $0,98$.

Exercice 3 : Intervalle de confiance non asymptotique pour le paramètre de la loi de Bernouilli.

Soit, pour $\theta \in]0,1[$, P_θ la loi de Bernouilli de paramètre θ.

1. Soit $\theta \in]0,1[$. Soit $X_1, \ldots, X_n$ une famille de v.a.i.i.d. de loi P_θ. Montrer, en utilisant l'inégalité de Chebyshev, que pour tout $\varepsilon > 0$,

$$\mathrm{P}\Big(\big|\frac{1}{n}\sum_{i=1}^n X_i - \theta\big| \geq \varepsilon\Big) \leq \frac{1}{4n\varepsilon^2}.$$

2. Déduisez-en que pour $\alpha \in]0,1[$, pour tout $n \geq 1$, en prenant $\varepsilon > 0$ tel que $\alpha = \frac{1}{4n\varepsilon^2}$,

$$[\frac{1}{n}\sum_{i=1}^n X_i - \varepsilon, \frac{1}{n}\sum_{i=1}^n X_i + \varepsilon]$$

définit un intervalle de confiance de fiabilité au moins $100.(1-\alpha)\%$ pour estimer θ.

3. On considère un sondage dans lequel deux réponses sont possibles, "oui" et "non". Combien faut-il interroger de personnes pour avoir une estimation à 0.03 près du pourcentage de gens en faveur de "oui", et ce au moins 19 fois sur 20.

Exercice 4 : Intervalles de confiance non asymptotiques pour les paramètres de lois gaussiennes. On considère $Y_1, \ldots, Y_n$ des v.a.i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$. On pose

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i, \quad V_n = \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y}_n)^2.$$

1. Rappeler la loi de

$$\frac{\sqrt{n}}{\sigma}(\bar{Y}_n - \mu).$$

2. En déduire, en utilisant les quantiles de la loi gaussienne centrée réduite, pour $\alpha \in]0,1[$, un intervalle de confiance de fiabilité $100(1-\alpha)\%$ pour estimer μ lorsque σ est connu.

3. Déduire de 1., en utilisant les quantiles de la loi gaussienne centrée réduite, pour $\alpha \in]0,1[$, un intervalle de confiance de fiabilité $100(1-\alpha)\%$ pour estimer σ lorsque μ est connu.

4. Rappeler la loi de

$$\frac{\bar{Y}_n - \mu}{\sqrt{V_n/(n-1)}}.$$

5. En déduire, en utilisant les quantiles de la loi de Student à $n-1$ degrés de liberté, pour $\alpha \in]0,1[$, un intervalle de confiance de fiabilité $100(1-\alpha)\%$ pour estimer μ lorsque σ est inconnu.

6. Rappeler la loi de

$$\frac{nV_n}{\sigma^2}.$$

7. En déduire, en utilisant les quantiles de la loi du chi-deux à $n-1$ degrés de liberté, pour $\alpha \in]0,1[$, un intervalle de confiance de fiabilité $100.(1-\alpha)\%$ pour estimer σ lorsque μ est inconnu.

Exercice 5 : Intervalles de confiance asymptotiques pour une suite de v.a.i.i.d. L^2. On considère $(Y_i)_{i\geq 1}$ des v.a. i.i.d. L^2 de moyenne μ et de variance σ^2. On pose

$$\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad V_n = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2.$$

1. Rappeler la limite presque sûre de $\bar{Y}_n$ et la limite en loi de

$$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sigma}.$$

En déduire , en utilisant les quantiles de la loi gaussienne centrée réduite, pour $\alpha \in]0,1[$, un intervalle de confiance asymptotique de fiabilité $100(1-\alpha)\%$ pour estimer μ lorsque σ est connu.

2. Montrer que

$$V_n = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2,$$

en déduire la limite presque sûre de V_n.

3. On admet que la limite en loi de

$$\frac{\sqrt{n}(\bar{Y}_n - \mu)}{\sqrt{V_n}}$$

est $\mathcal{N}(0,1)$. En déduire, en utilisant les quantiles de la loi gaussienne centrée réduite, pour $\alpha \in]0,1[$, un intervalle de confiance asymptotique de fiabilité $100(1-\alpha)\%$ pour estimer μ lorsque σ est inconnu.

4. On veut estimer la taille moyenne des hommes adultes dans une polulation. Un échantillon de 100 individus a donné une moyenne de

$$\frac{1}{100}\sum_{\text{individus testés}} \text{taille} = 175\text{cm},$$

$$\frac{1}{100}\sum_{\text{individus testés}} (\text{taille} - 175)^2 = 49\text{cm}^2.$$

Donner un intervalle de confiance à 95% pour la taille moyenne de la population. On pourra utiliser le fait que pour $q = 1,96$, $\mathrm{P}(-q \leq \mathcal{N}(0,1) \leq q) = 0,95$.

Exercice 6 : Intervalles de confiance asymptotiques pour une suite de v.a.i.i.d. de Bernouilli.

Vous pouvez me rendre la correction de cet exercice (ou sa partie) d scannée par E-mail à Irina.Kourkovaupmc.fr si vous voulez avoir des commentaires sur votre rédaction. Je vous prie aussi de me donner dans votre E-mail **vos coordonnées téléphoniques** en cas où j'ai des commentaires assez longs à faire.

On reprend les notations de l'exercice précédent, mais on suppose que les Y_i sont de loi de Bernouilli de paramètre $p \in]0,1[$. On pourra utiliser le fait que pour $q = 1,96$, $\mathrm{P}(-q \leq \mathcal{N}(0,1) \leq q) = 0,95$ et que pour $q = 2,4$, $\mathrm{P}(-q \leq \mathcal{N}(0,1) \leq q) = 0,99$.

1. Rappeler la limite en loi de

$$\frac{\sqrt{n}(\bar{Y}_n - p)}{\sqrt{p(1-p)}}.$$

Soit $\alpha \in]0, 1/2[$. Pour quel q a-t-on

$$\mathrm{P}\left(p \in \left[\bar{Y}_n - \frac{q\sqrt{\bar{Y}_n(1-\bar{Y}_n)}}{\sqrt{n}}, \bar{Y}_n + \frac{q\sqrt{\bar{Y}_n(1-\bar{Y}_n)}}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} 1-\alpha?$$

2. En remarquant que pour tout $x \in [0,1]$, $x(1-x) \leq 1/4$, montrer que

$$\left[\bar{Y}_n - \frac{q}{2\sqrt{n}}, \bar{Y}_n + \frac{q}{2\sqrt{n}}\right]$$

est un intervalle de confiance asymptotique de fiabilité au moins $100.(1-\alpha)\%$ pour estimer p.

3. Refaire la question 3 de l'exercice 3 à la lumière de ce qu'on vient de faire, et comparer les réponses. Commenter.

4. Sur le site http ://pollingreport.com/religion2.htm, on peut lire que lors d'un sondage réalisé aux USA en 2003 pour la chaine de télévision nationale FOX News que sur 900 citoyens américains votants choisis au hasard dans les Etats Unis, 639 ont répondu oui à la question "Croyez-vous personnellement au diable" (aussi impressionnante que soit la proprotion, ce sondage est très sérieux). Donner un intervalle de confiance à 95% pour la proportion des citoyen américains croyant au diable. Avec quelle probabilité peut-on dire que cette proportion est située entre 67% et 75% ?

Exercice 7 : Test de χ-deux.

a) Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes et de même loi à valeurs $\{1, \ldots, r\}$ avec probabilités $p_1, \ldots, p_r$ respectivement, $p_1 + \cdots + p_r = 1$.

Soient $N_i^n = \sum_{k=1}^n 1_{\{X_k = i\}}$ pour $i = 1, \ldots, r$. Soit $T_n = \sum_{i=1}^r \frac{(N_i^n - np_i)^2}{np_i}$. La suite $(T_n)_{n\geq 0}$, converge-t-elle en loi ? Si oui, quelle est la loi limite ?

b) Un opérateur téléphonique envisage proposer à ces clients cinq nouveaux forfaits différents en fonction de leur consommation. Il veut tester l'hypothèse que 10% de clients choisiront le premier forfait, 20% – le deuxième, 30% — le troixième, 30% – le quatrième et 10% — le cinquième. Pour cela, il fait un sondage de 1000 clients, dont le résultat est le suivant : 80 sondés veulent choisir le premier forfait, 220 sondés – le deuxième, 330 sondés – le troixième, 280 sondés – le quatrième et 90 sondés sont pour le cinquième forfait. Comment tester cette hypothèse au seuil de fiabilité 0.05 ?

c) Une chaine de consommation s'aprête à offrir un cadeau au choix parmi 6 cadeaux à chacun de ses 100000 acheteurs ayant une carte de fidélité. Elle veut tester une hypothèse que le 1er cadeau sera chosi par 15% d'acheteurs, le 2ème – par 20%, le 3ème – par 5%, le 4ème – par 10%, le 5ème – par 30%, le 6ème – par 20%. Elle offre les cadeaux à 1000 premiers acheteurs : 149 acheteurs choisissent le 1er cadeau, 202 – le deuxième, 53– le 3ème, 96 – le 4ème, 301 – le 5ème, 199 – le 6ème cadeau. Comment tester cette hypothèse au seuil de fiabilité 0.01 ?

Vous pouvez vous servir de la table sur la page web pour les trouver quatiles de la loi χ-2 à n degrés de liberté : http ://www.stat.ucl.ac.be/ISpersonnel/lambert/vete1362/tables/tquantchicarre.pdf