PREPARATION A L'AGREGATION – TPs de PROBA–STAT

Théorème Central Limite

On considère la suite $\{X_n,\, n \in \mathbb{N}\}$ de vecteurs aléatoires de dimension 2, i.i.d., t.q. $X_k = \exp(2i\pi U_k)$ (on identifie ici le plan complexe à $\mathbb{R}^2$), où les $\{U_k\}$ sont i.i.d. de loi commune la loi uniforme sur $[0,1]$.

1 Préciser la matrice de transition de la chaîne de Markov $\{X_n,\, n \in \mathbb{N}\}$.

2 Calculer l'espérance et la matrice de covariance du v.a. X_1.

3 Montrer que quand $n \to \infty$,

$$Z_n := \|n^{-1/2}(X_1 + \cdots + X_n)\|^2 \to Z$$

en loi, où Z suit la loi exponentielle de paramètre 1. On appelle fonction de survie de la v.a. Z la fonction $x \to G(x) = \mathbb{P}(Z > x)$.

4 On veut illustrer cette convergence en loi numériquement. Pour cela on va simuler m réalisations indépendantes de la suite $(X_1, \ldots, X_n)$, c'est à dire que l'on va simuler $\{(X_1^\ell, \ldots, X_n^\ell),\, \ell = 1, 2, \ldots, m\}$, donc on va devoir simuler $n \times m$ réalisations indépendantes de la loi uniforme sur $[0,1]$, et on va tracer sur un même graphique la fonction de survie empirique des $\{Z_n^\ell,\, \ell = 1, 2, \ldots, m\}$ et la fonction de survie de Z sur l'intervalle $[0,6]$, ainsi qu'un intervalle de confiance approprié (ou peut–être la différence des deux fonctions de survie, avec une normalisation convenable). Quel est le meilleur choix du couple (n,m), sachant que l'on s'autorise au total $n \times m = 10^6$ tirages ? Présenter le résultat des simulations graphiquement, de la façon qui vous paraît la plus parlante.

Attention ! Il faut choisir m assez grand par rapport à n, pour que l' "erreur de Monte Carlo" ne masque pas l'écart entre la loi de Z_n et celle de Z.

Chaînes de Markov

On considère une file d'attente en temps discret. X_n désigne le nombre de clients dans la file en attente ou en train de se faire servir à l'instant n.

Entre les instants n et $n+1$ arrivent Y_{n+1} clients, et si $X_n > 0$ partent Z_{n+1} clients.

On suppose que X_0, Y_1, Z_1, Y_2, $Z_2 \ldots$ sont indépendantes, avec les Y_n i.i.d. de Bernoulli de paramètre p (i.e. $\mathbb{P}(Y_1 = 1) = p = 1 - \mathbb{P}(Y_1 = 0)$), et les Z_n i.i.d. de Bernoulli de paramètre q (i.e. $\mathbb{P}(Z_1 = 1) = q = 1 - \mathbb{P}(Z_1 = 0)$), avec $0 < p, q < 1$. Alors la suite $\{X_n,\, n \in \mathbb{N}\}$ est donnée par la formule de récurrence

$$X_{n+1} = (X_n - Z_{n+1})^+ + Y_{n+1}.$$

C'est une chaîne de Markov. Elle est irréductible apériodique, et récurrente positive si $p < q$, de probabilité invariante π sur $\mathbb{N}$ donnée par

$$\pi_0 = \frac{q-p}{q}, \quad \pi_k = (q-p)\frac{p^k(1-q)^{k-1}}{q^{k+1}(1-p)^k},\ k \geq 1.$$

1. Simuler et tracer une trajectoire de $\{X_n,\, n \geq 0\}$ pendant un temps long ($n = 1000$ au moins), pour $p = 1/2$ et successivement $q = 3/5, 7/13, 15/29, 1/2$. Tracer les courbes correspondantes, et commenter.

2. Puisque $\{X_n\}$ est irréductible, récurrente positive et apériodique, $(P^n)_{xy} \to \pi_y$. Tracer l'histogramme empirique de $(P^n)_{x\cdot}$, pour $n = 10, 20, 50, 150, 250, 500$, pour une taille d'échantillon de 2500. On représentera l'histogramme de π sur le même graphique. On traitera les cas $p = 1/2$, $q = 3/5, 15/29$.

3. Comparer graphiquement les quantités

$$n^{-1}\sum_{k=1}^{n} \mathbf{1}_{\{X_k = x\}}, \quad x \in \mathbb{N}$$

et l'histogramme de π, pour $n = 500, 1000, 2000, 3000, 5000$. On traitera les cas $p = 1/2$, $q = 3/5, 15/29$.

Remarque Les valeurs de q, et de n aux questions **2** et **3** sont données à titre indicatif. On cherchera éventuellement les bonnes plages de variation des valeurs, pour illustrer au mieux les phénomènes que l'on veut montrer.

Monte–Carlo, Formule de Black–Scholes

Le but de ce **TP** est d'appliquer la méthode de Monte–Carlo au calcul du prix d'une option d'achat (*call*) européenne portant sur un sous–jaçant dont le cours à l'instant 0 est de 105 Euros, au prix d'exercice $K = 110$ Euros, à échéance d'un an, avec un taux bancaire à 5% (i.e. $rT = 0,05$), et une volatilité telle que $\sigma\sqrt{T} = 0,3$.

1 Calculer le prix du call en appliquant la méthode de Monte–Carlo à la formule

$$C_0 = \mathbb{E}^* \left[\left(S_T - Ke^{-rT} \right)_+ \right],$$

avec 1 000 tirages. On prendra soin d'évaluer (de façon approchée) la variance de la v.a. dont on cherche à estimer l'espérance, et on donnera un intervalle de confiance pour la quantité cherchée.

2 Faire le même calcul (y compris l'intervalle de confiance) en combinant la même méthode appliquée à la formule pour le prix du put :

$$P_0 = \mathbb{E}^* \left[\left(Ke^{-rT} - S_T \right)_+ \right],$$

avec le même nombre de tirages, et la formule de parité call–put.

3 Calculer une troisième fois le prix de la même option d'achat, en utilisant cette fois la formule

$$C_0 = S_0 F(d_1) - Ke^{-rT} F(d_2),$$

avec

$$d_1 = \frac{1}{\sigma\sqrt{T}} \log\left(\frac{S_0}{K}\right) + \frac{r\sqrt{T}}{\sigma} + \frac{\sigma\sqrt{T}}{2},$$

$$d_2 = \frac{1}{\sigma\sqrt{T}} \log\left(\frac{S_0}{K}\right) + \frac{r\sqrt{T}}{\sigma} - \frac{\sigma\sqrt{T}}{2},$$

et la fonction de répartition F de la $N(0,1)$ fournie par Matlab.

Processus de Galton–Watson

Le but de ce **TP** est d'illustrer graphiquement certains comportements énoncés dans le cours du 20 janvier sur l'évolution de la taille d'une population.

On rappelle ces résultats. On suppose que les nombres d'enfants de chaque individu de chaque génération sont i.i.d., de loi celle de la v.a. ξ à valeurs dans $\mathbb{N}$. On note Z_n le nombre d'individus de la génération n. On suppose que $Z_0 = 1$.

Si $\mathbb{E}\xi < 1$, alors la poulation s'éteint p.s., et la vitesse d'extinction est contrôlée par la majoration

$$\mathbb{P}(Z_n > 0) \leq [\mathbb{E}\xi]^n.$$

Si $\mathbb{E}\xi = 1$, on a encore p.s. extinction, mais la convergence de $\mathbb{P}(Z_n > 0)$ vers 0 est beaucoup plus lente, puisque si $\mathbb{E}(\xi^2) < \infty$, un équivalent de $\mathbb{P}(Z_n > 0)$ quand $n \to \infty$ est donné par

$$\frac{2}{n\mathbb{E}[\xi(\xi-1)]}.$$

Si $\mathbb{E}\xi > 1$, et $\mathbb{E}(\xi^2) < \infty$, alors la probabilité d'extinction est strictement plus petite que 1, et lorqu'il n'y a pas extinction la population explose à vitesse exponentielle. Plus précisément,

$$\frac{Z_n}{(\mathbb{E}\xi)^n} \to W, \text{ quand } n \to \infty,$$

avec $W > 0$ sur l'événement “non extinction”.

On va chercher à vérifier les deux premiers résultats par un calcul de type Monte Carlo, et le troisième par des simulations. On choisira des tailles d'échantillon égales à 2000, et un nombre de générations égal à 250.

On traitera dans chaque exemple deux cas de loi de probabilité : un cas d'une loi de support “petit”, et une loi de Poisson. La première est très facile à simuler. Pour simuler la seconde, on utilisera le fait que si $\{N_t,\, t \geq 0\}$ est un processus de Poisson d'intensité λ, la v.a. N_1 suit la loi de Poisson de paramètre λ. Pour simuler N_1, on simule des v.a. exponentielles de paramètre λ $\{\eta_k,\, k \geq 1\}$ (ce qui est très facile à réaliser), et on pose

$$N_1 = \text{ le nombre } \ell \text{ tel que } \eta_1 + \cdots + \eta_\ell \leq 1 < \eta_1 + \cdots + \eta_{\ell+1}.$$

1 Dans les deux cas où

$$\mathbb{P}(\xi = 0) = \frac{1}{3}, \quad \mathbb{P}(\xi = 1) = \frac{1}{2}, \quad \mathbb{P}(\xi = 2) = \frac{1}{6},$$

et ξ suit une loi de Poisson d'intensité $\frac{3}{4}$, tracer sur un même graphique les évolutions de $n^{-1} \log \mathbb{P}(Z_n > 0)$ et de $\log(\mathbb{E}\xi)$, de $n = 1$ à $n = 250$.

Pour simuler la loi discrète ci–dessus, on découpe l'intervalle $[0, 1]$ en les intervalles $I_0 = [0, \frac{1}{3}[$, $I_1 = [\frac{1}{3}, \frac{5}{6}[$ et $I_2 = [\frac{5}{6}, 1]$. Etant donné une v.a. η de loi $U(0, 1)$, on pose $\xi = i$ ssi $\eta \in I_i$, $i = 0, 1, 2$.

2 Dans les deux cas où

$$\mathbb{P}(\xi = 0) = \frac{1}{4}, \quad \mathbb{P}(\xi = 1) = \frac{1}{2}, \quad \mathbb{P}(\xi = 2) = \frac{1}{4},$$

et ξ suit une loi de Poisson d'intensité 1, tracer sur un même graphique les évolutions de $n\mathbb{P}(Z_n > 0)$ et de $\frac{2}{\mathbb{E}[\xi(\xi-1)]}$, de $n = 1$ à $n = 250$.

Pour simuler la loi discrète ci–dessus, on découpe l'intervalle $[0, 1]$ en les intervalles $I_0 = [0, \frac{1}{4}[$, $I_1 = [\frac{1}{4}, \frac{3}{4}[$ et $I_2 = [\frac{3}{4}, 1]$. Etant donné une v.a. η de loi $U(0, 1)$, on pose $\xi = i$ ssi $\eta \in I_i$, $i = 0, 1, 2$.

3 Dans les deux cas où

$$\mathbb{P}(\xi = 0) = \frac{1}{5}, \quad \mathbb{P}(\xi = 1) = \frac{1}{2}, \quad \mathbb{P}(\xi = 2) = \frac{1}{5}, \quad \mathbb{P}(\xi = 3) = \frac{1}{10},$$

et ξ suit une loi de Poisson d'intensité 2, tracer quinze évolutions du rapport

$$\frac{Z_n}{(\mathbb{E}\xi)^n},$$

de $n = 1$ à $n = 250$.

Pour simuler la loi discrète ci–dessus, on découpe l'intervalle $[0, 1]$ en les intervalles $I_0 = [0, \frac{1}{5}[$, $I_1 = [\frac{1}{5}, \frac{7}{10}[$ et $I_2 = [\frac{7}{10}, \frac{9}{10}[$, $I_3 = [\frac{9}{10}, 1]$. Etant donné une v.a. η de loi $U(0, 1)$, on pose $\xi = i$ ssi $\eta \in I_i$, $i = 0, 1, 2, 3$.

L'urne de Polya

Une urne contient des boules de deux couleurs différentes, des blanches et des noires. Aux instants $n = 1, 2, \ldots$, on tire une boule de l'urne, on la remet dans l'urne et on ajoute une boule supplémentaire, de la même couleur.

On note B_n (resp. N_n) le nombres de boules blanches (resp. noires) a l'issue du n–ième coup. On pose

$$X_n = \frac{B_n}{B_n + N_n}.$$

On sait que si $B_0 = N_0 = 1$, la loi de X_n converge vers la mesure de Lebesgue sur $[0,1]$, et que si $B_0 = k$, $N_0 = \ell$, la loi de X_n converge vers la loi Béta(k, ℓ), de densité sur $[0,1]$

$$\frac{(k+\ell-1)!}{(k-1)!(\ell-1)!} x^{k-1}(1-x)^{\ell-1}.$$

1 Dans le premier cas ($B_0 = N_0 = 1$), simuler 10 000 urnes de Polya jusqu'à $n = 100$. Tracer les histogrammes des lois de X_{20}, X_{40}, X_{60}, X_{80} et X_{100}.

2 Même exercice avec $B_0 = 3$, $N_0 = 5$.

3 Même exercice avec $B_0 = 5$, $N_0 = 1$.

Martingales, Stratégies de jeux

On considère un casino où le joueur qui mise la somme m double son gain avec probabilité p, perd sa mise avec la probabilité $1-p$. Soit $\{Y_k,\ k=1,2,\ldots\}$ des variables aléatoires i.i.d., telles que $\mathbb{P}(Y_1=1)=p=1-\mathbb{P}(Y_1=-1)$. Un joueur qui mise m_k au k–ième coup, $1\leq k\leq n$, gagne au total la somme

$$\sum_{k=1}^{n} m_k Y_k.$$

On considère deux stratégies de jeu, qui toutes les deux ont pour but de gagner au total 10 Euros.

• La première "quitte ou double" consiste à s'arrêter à l'issue du premier gain, i.e. à l'instant $N=\inf\{n,\ Y_n=1\}$, et à miser $2^{k-1}\times 10$ Euros au k–ième coup, si tous les coups précédents ont été perdants.

• La seconde, plus timorée, consiste à miser 10 Euros à chaque coup, et à s'arrêter dès que le total des gains cumulés atteint 10 Euros, i.e. on s'arrête à l'instant $N=\inf\{n,\ \sum_{k=1}^{n} Y_k=1\}$.

I Théorie

1 Vérifier que pour chacune de ces deux stratégies, le gain est 10 Euros si l'événement $\{N<\infty\}$ est réalisé.

2 Montrer que dans le cas $p=1/2$, la marche aléatoire

$$X_n=\sum_{k=1}^{n} Y_k,\ n\geq 1$$

est une chaîne de Markov récurrent nulle. En déduire que pour la seconde stratégie $\mathbb{P}(N<\infty)=1$, mais que $\mathbb{E}(N)=+\infty$, donc l'espérance de la somme dépensée avant de terminer le jeu vaut $+\infty$.

Montrer que pour la première stratégie on a aussi que l'espérance de l'argent dépensé avant de terminer le jeu vaut $+\infty$.

3 Qu'en est–il si $p>1/2$? Si $p<1/2$?

II Simulations On suppose maintenant que deux joueurs, chacun adepte de l'une des deux stratégies ci–dessus, sont prêts à dépenser jusqu'à 100 Euros pour tenter de gagner 10 Euros. C'est à dire qu'il s'arrête soit quand il a gagné, soit quand il lui faudrait débourser plus de 100 Euros pour continuer.

1 Cas $p=1/2$. Pour chacune des deux stratégies, on demande de simuler 5 000 fois le jeu, d'en déduire une approximation de la probabilité de gagner, et de l'espérance mathématique du gain [$=10\times\mathbb{P}(\text{gagner})-100\times\mathbb{P}(\text{perdre})$].

2 Même question dans le cas $p = 0,4$.

PS 1 Les quantités que l'on demande d'approcher par simulation peuvent se calculer facilement. On pourra chercher à préciser leurs valeurs après avoir réalisé les simulations.

PS 2 Les cinéphiles pourront voir ou revoir "L'argent de la vieille", où une riche américaine plume aux cartes un couple de napolitains peu fortunés.