Exercice : Banque PT 2023 A (extrait)

Cet exercice comporte trois parties. Les parties 2 et 3 sont indépendantes.

Un jeu consiste à lancer un ballon dans un panier. On suppose que la probabilité de réussir le panier est $p \in]0,1[$ et que les lancers sont indépendants. On note $q = 1 - p$.

Partie 1 : Etude du jeu de lancer

1. On note T le nombre de lancers nécessaires pour réussir un panier pour la première fois. Quelle est la loi suivie par la variable aléatoire T ?
   On explicitera la loi sans démonstration.
2. On effectue une infinité de lancers. Calculer la probabilité de réussir au moins un panier.
3. L'organisateur du jeu ne connait pas la valeur de p et souhaite en connaître une valeur approchée. Pour cela il observe $N \in \mathbb{N}^*$ lancers et note le nombre S_N de paniers réussis.
   (a) Quelle est la loi de S_N ? On explicitera la loi en justifiant brièvement la réponse.
   (b) Montrer que $p(1-p) \leqslant \frac{1}{4}$.
   (c) Montrer que pour $\varepsilon > 0$, $\mathbf{P}\left(\left|\frac{S_N}{N} - p\right| \geqslant \varepsilon\right) \leqslant \frac{1}{4N\varepsilon^2}$.

Partie 3 : Une autre estimation du paramètre

La partie 1 donne un résultat permettant d'approcher la valeur de p par $\frac{S_N}{N}$. On prouve dans cette partie un résultat similaire en utilisant une méthode dite de grandes déviations. On suppose que $p \in]0,\frac{1}{2}[$. Pour $\lambda \geqslant 0$, on pose

$$h(\lambda) = (1-p)e^{-\lambda p} + pe^{\lambda(1-p)} \quad \text{et} \quad g(\lambda) = \ln h(\lambda).$$

On admet que pour tout $\lambda \geqslant 0$, on a $g(\lambda) \leqslant \frac{\lambda^2}{2}$.

Soit $N \in \mathbb{N}^*$. On considère des variables aléatoires $X_1, \ldots, X_N$ indépendantes et de même loi de Bernoulli $\mathcal{B}(p)$, et on pose $S_N = X_1 + \ldots + X_N$ et $\lambda \geqslant 0$.

4. (a) Montrer que
   $$\mathbf{E}\left(e^{\lambda(X_1 - p)}\right) = e^{g(\lambda)}.$$
   On justifiera bien le calcul.
   (b) En déduire que
   $$\mathbf{E}\left(e^{\lambda(S_N - Np)}\right) \leqslant e^{\frac{N}{2}\lambda^2}$$
   (c) En utilisant l'inégalité de Markov, déduire des questions précédentes que pour $\varepsilon > 0$, et $\lambda > 0$
   $$\mathbf{P}\left(\frac{S_N}{N} - p \geqslant \varepsilon\right) \leqslant \exp\left(-N\lambda\varepsilon + \frac{N}{2}\lambda^2\right)$$
   (d) En choisissant bien λ, en déduire que
   $$\mathbf{P}\left(\frac{S_N}{N} - p \geqslant \varepsilon\right) \leqslant e^{-\frac{N}{2}\varepsilon^2}$$
5. On admet que l'on pourrait montrer de manière similaire que
   $$\mathbf{P}\left(\left|\frac{S_N}{N} - p\right| \geqslant \varepsilon\right) \leqslant e^{-\frac{N}{2}\varepsilon^2}$$

On souhaite évaluer l'erreur faite en prenant $\frac{S_N}{N}$ comme valeur approchée de p. Expliquer en quoi la majoration ci-dessus est utile pour faire cette évaluation. A quelle condition sur ε et N la méthode de grandes déviations de la partie 3 est-elle plus intéressante que celle de la partie 1?

Un raisonnement, même incomplet, interprétant les résultats obtenus sera valorisé.

CORRECTION

Partie 1

1. $T \sim \mathcal{G}(p)$ car T est le rang du premier succès lors de la succession d'épreuves de Bernouilli mutuellement indépendantes, de même paramètre de succès p.
2. Il s'agit de déterminer la probabilité de l'événement $\bigcup_{n\geqslant 0}(T=n)$.

   Comme il s'agit d'une union d'événements disjoints 2 à 2, on a

$$P(\bigcup_{n\geqslant 1}(T=n)) = \sum_{n=1}^{\infty} P(T=n) = \sum_{n=1}^{\infty} q^{n-1}.p = p.\sum_{j=0}^{\infty} q^j = p.\frac{1}{1-q} = 1$$

3. (a) $S_N \sim \mathcal{B}(N,p)$ car S_N est le nombre de succès lorsque que l'on répète N épreuves de Bernoulli indépendantes et identiques, de même paramètre de succès p

   (b) Il suffit d'étudier la fonction $p \mapsto p(1-p)$ sur l'intervalle [0,1].

   (c) On utilise l'inégalité de Bienaimé Tchebychev à la variable aléatoire $Y = \dfrac{S_N}{N}$

   i) par linéarité de l'espérance $E(Y) = \dfrac{1}{N}.E(S_N) = \dfrac{1}{N}.N.p = p$

   ii) $V(Y) = V(\dfrac{1}{N}.S_N) = \dfrac{1}{N^2}.V(S_N) = \dfrac{1}{N^2}.N.p.q = \dfrac{p(1-p)}{N}$

   Ainsi

$$P(|Y - E(Y)| \geqslant \varepsilon) \leqslant \frac{V(Y)}{\varepsilon^2}$$

   donne

$$P(\left|\frac{S_N}{N} - p\right| \leq \varepsilon) \leqslant \frac{p(1-p)}{N\varepsilon^2}$$

   et d'après la question précédente , on en déduit bien que

$$P(\left|\frac{S_N}{N} - p\right| \leq \varepsilon) \leqslant \frac{1}{4N\varepsilon^2}$$

Partie 3

4. (a) La variable aléatoire $e^{\lambda(X_1-p)}$ possède une espérance car elle ne prend qu'un nombre fini de valeurs, et l'on a d'après **le théorème de transfert**

$$E(e^{\lambda(X_1-p)}) = P(X_1=0).e^{\lambda(0-p)} + P(X_1=1).e^{\lambda(1-p)} = (1-p)e^{-\lambda p} + p.e^{\lambda(1-p)} = h(\lambda) = e^{g(\lambda)}$$

   (b) Notons pour tout $i, Y_i = e^{\lambda(X_i-p)}$.
   On a

$$e^{\lambda(S_N-Np)} = Y_1.Y_2.\ldots.Y_N = \prod_{i=1}^{p} Y_i$$

   Comme les variables X_1,$X_2,\ldots,X_N$ sont indépendantes, on a aussi les Y_1,$Y_2,\ldots,Y_N$ qui sont indépendantes, et ainsi

$$E(e^{\lambda(S_N-Np)}) = \prod_{i=1}^{p} E(Y_i)$$

   et comme les Y_i suivent toutes la même loi (celle de Y_1) on a $E(Y_i) = e^{g(\lambda)}$ et donc

$$E(e^{\lambda(S_N-Np)}) = e^{Ng(\lambda)}$$

   On admet que $g(\lambda) \leqslant \dfrac{\lambda^2}{2}$, on a donc par croissance de la fonction exponentielle

$$E(e^{\lambda(S_N-Np)}) = e^{\frac{N}{2}\lambda^2}$$

(c) On commence par remarquer que les événements ci-dessous sont tous égaux (car $\lambda > 0,\ N > 0$ et la fonction exp est croissante)

$$\left(\frac{S_N}{N} - p \geqslant \varepsilon\right) = (\lambda.(S_N - N.p) \geqslant \lambda.N.\varepsilon) = \left(e^{\lambda(S_N - Np)} \geqslant e^{\lambda.N.\varepsilon}\right)$$

L'inégalité de Markov appliquée à la variable aléatoire positive $e^{\lambda(S_N - Np)}$ donne

$$P\left(e^{\lambda(S_N - Np)} \geqslant e^{\lambda.N.\varepsilon}\right) \leqslant \frac{E\left(e^{\lambda(S_N - Np)}\right)}{e^{\lambda.N.\varepsilon}}$$

avec l'inégalité montré en Q4b) cela donne bien

$$\mathbf{P}\left(\frac{S_N}{N} - p \geqslant \varepsilon\right) \leqslant \exp\left(-N\lambda\varepsilon + \frac{N}{2}\lambda^2\right)$$

(d) Il suffit de considérer $\lambda = \varepsilon > 0$)

5.
- Pour connaître une valeur approchée de p, on effectue N lancers et on considère $\dfrac{S_N}{N}$ comme approximation.
  L'inégalité précédente nous indique que si l'on souhaite obtenir une approximation de p à ε près avec une probabilité par exemple de 95%, il suffit de choisir un N (càd d'effectuer un nombre de lancers N) suffisamment grand pour avoir $e^{-\frac{N}{2}\varepsilon^2} \leqslant 0.95$.
  Grâce à l'exponentielle, on est assuré qu'en augmentant N, on obtient "rapidement" cette inégalité
- Dans la partie 1, nous avions trouvé que

$$\mathbf{P}\left(\left|\frac{S_N}{N} - p\right| \geqslant \varepsilon\right) \leqslant \frac{1}{4N\varepsilon^2}$$

La méthode des grandes variations est donc plus intéressante pour estimer p lorsque

$$e^{-\frac{N}{2}\varepsilon^2} \leqslant \frac{1}{4N\varepsilon^2}$$