PROBABILITES DISCRETES

ENONCE DE L'EXERCICE

ENONCE–32

Une urne contient des boules numérotées de 1 à N, où N est un entier supérieur ou égal à 2. Elles sont indiscernables au toucher et l'on effectue n tirages successifs d'une boule avec remise. Pour tout $j \in [\![1, N]\!]$, on note Z_j le numéro obtenu au $j^{\text{ème}}$ tirage. On suppose N inconnu et on cherche à l'estimer.

1) On considère la moyenne empirique :

$$M_n = \frac{1}{n}(Z_1 + Z_2 + \cdots + Z_n)$$

Calculer l'espérance $E(M_n)$ de M_n et en déduire un estimateur sans biais de N.

2) On considère maintenant $X_n = \sup(Z_1, \ldots, Z_n)$ et on cherche à montrer que X_n est un estimateur asymptotiquement sans biais de N, c'est-à-dire que :

$$\lim_{n \to +\infty} E(X_n) = N$$

a) Donner la fonction de répartition de X_n.

b) Montrer que, pour toute variable aléatoire Y à valeurs dans $[\![1, N]\!]$, on a :

$$\sum_{k=1}^{N} P(Y \geq k) = E(Y)$$

c) En déduire que $E(X_n) \geq N - \dfrac{N}{n+1}$ et conclure.

© EDUKLUB SA
Tous droits de l'auteur des oeuvres réservés. Sauf autorisation, la reproduction ainsi que toute utilisation des oeuvres autre que la consultation individuelle et privée sont interdites.

Fichier généré pour Visiteur (), le 31/05/2021

CORRIGE DE L'EXERCICE NUMERO 32

1)

Par linéarité de l'espérance, $E(M_n) = E(\frac{1}{n}\sum_{j=1}^{n} Z_j) = \frac{1}{n}\sum_{j=1}^{n} E(Z_j)$.

Or les variables Z_j suivent la loi uniforme sur $[\![1, N]\!]$ puisqu'à chaque tirage, la situation est la même qu'initialement. Donc $E(Z_j) = \frac{N+1}{2}$.

Il s'ensuit que

$$E(M_n) = \frac{1}{n} n \frac{N+1}{2} = \frac{N+1}{2}.$$

On en déduit qu'un estimateur sans biais de N est $T_n = 2M_n - 1$ puisque

$$E(T_n) = 2E(M_n) - 1 = N + 1 - 1 = N.$$

2–a)

La variable X_n est discrète puisque, par indépendance des variables Z_j, on a : $X_n(\Omega) = [\![1, N]\!]$. En effet, pour toute valeur $k \in [\![1, N]\!]$, l'événement $(X_n = k)$ est réalisé lorsque toutes les variables Z_j prennent la valeur k, ce qui n'est pas impossible.

Donc, puisque X_n est discrète, sa fonction de répartition est entièrement déterminée par les valeurs $F_{X_n}(k)$ pour $k \in [\![1, N]\!]$.

Soit $k \in [\![1, N]\!]$. On a :

$F_{X_n}(k) = P(X_n \leq k) = P(Z_1 \leq k) \cap Z_2 \leq k \cap \ldots \cap Z_n \leq k)$, car dire que le plus grand des n nombres $Z_1, \ldots, Z_n$ est inférieur ou égal à k veut dire que chacun d'entre eux est inférieur ou égal à k.

Par indépendance (implicite) des variables Z_j, on peut écrire

$$F_{X_n}(k) = P(Z_1 \leq k)P(Z_2 \leq k)\ldots P(Z_n \leq k) = \left(\frac{k}{N}\right)^n$$

2–b)

Il n'y a pas de problème d'existence de l'espérance puisque $Y(\Omega)$ est un ensemble fini.

$\sum_{k=1}^{N} P(Y \geq k) = \sum_{k=1}^{N}\sum_{j=k}^{N} P(Y = j)$; en effet l'événement $(Y \leq k) = \bigcup_{j=k}^{N}(Y = j)$ et les événements $(Y = j)$ sont deux à deux incompatibles.

A partir de là, intervertissons l'ordre des sommations en remarquant que l'on a : $1 \leq k \leq j \leq N$. Donc j prend toutes les valeurs de $[\![1, N]\!]$ et lorsque j est fixé, k varie entre 1 et j ; d'où la formule

$$\sum_{k=1}^{N}\sum_{j=k}^{N} P(Y = j) = \sum_{j=1}^{N}\sum_{k=1}^{j} P(Y = j) = \sum_{j=1}^{N} jP(Y = j) = E(Y).$$

$$\boxed{E(Y) = \sum_{k=1}^{N} P(Y \geq k)}$$

© EDUKLUB SA
Tous droits de l'auteur des oeuvres réservés. Sauf autorisation, la reproduction ainsi que toute utilisation des oeuvres autre que la consultation individuelle et privée sont interdites.

Fichier généré pour Visiteur (), le 31/05/2021

2–c)

On a ici :

$$\begin{aligned} E(X_n) &= \sum_{k=1}^{N} P(X_n \geq k) \\ &= \sum_{k=1}^{N} (1 - P(X_n < k)) \\ &= \sum_{k=1}^{N} (1 - F_{X_n}(k-1)) \\ &= \sum_{k=1}^{N} (1 - (\frac{k-1}{N})^n) \end{aligned}$$

formule valable pour $k = 1$ car $F_{X_n}(0) = 0$

On obtient donc $E(X_n) = N - \sum_{k=1}^{N} (\frac{k-1}{N})^n = N - \sum_{j=0}^{N-1} (\frac{j}{N})^n$ après avoir effectué dans la somme le changement d'indice $j = k - 1$.

Or :

$\forall t \in \left[\frac{j}{N}, \frac{j+1}{N}\right]$, $0 \leq \frac{j}{N} \leq t$; on peut donc élever cette inégalité à la puissance n et puisque cette fonction puissance est croissante sur $\mathbb{R}_+$, on obtient $(\frac{j}{N})^n \leq t^n$.

Intégrons cette inégalité entre $\frac{j}{N}$ et $\frac{j+1}{N}$; les bornes sont dans l'ordre croissant, cela donne :

$$\int_{\frac{j}{N}}^{\frac{j+1}{N}} (\frac{j}{N})^n dt \leq \int_{\frac{j}{N}}^{\frac{j+1}{N}} t^n dt \text{ ou encore } \frac{1}{N}(\frac{j}{N})^n \leq \int_{\frac{j}{N}}^{\frac{j+1}{N}} t^n dt$$

Sommons ces dernières inégalités pour j variant de 0 à $N-1$ et pour l'intégrale utilisons la relation de Chasles, il vient :

$$\begin{aligned} \sum_{j=0}^{N-1} \frac{1}{N}(\frac{j}{N})^n &\leq \sum_{j=0}^{N-1} \int_{\frac{j}{N}}^{\frac{j+1}{N}} t^n dt \\ \sum_{j=0}^{N-1} \frac{1}{N}(\frac{j}{N})^n &\leq \int_0^1 t^n dt \\ \sum_{j=0}^{N-1} \frac{1}{N}(\frac{j}{N})^n &\leq \frac{1}{n+1} \end{aligned}$$

$$\text{Donc finalement , } \sum_{j=0}^{N-1} (\frac{j}{N})^n \leq \frac{N}{n+1}.$$

Or on sait que $E(X_n) = N - \sum_{j=0}^{N-1} (\frac{j}{N})^n$; l'inégalité précédente permet d'écrire

$$E(X_n) \geq N - \frac{N}{n+1}.$$

Il est évident que $E(X_n) \leq N$ puisque N est la plus grande des valeurs possibles de X_n. On a donc l'encadrement

$$N - \frac{N}{n+1} \leq E(X_n) \leq N.$$

$$\boxed{\text{Par encadrement : } \lim_{n \to +\infty} E(X_n) = N}$$

Tous droits de l'auteur des oeuvres réservés. Sauf autorisation, la reproduction ainsi que toute utilisation des oeuvres autre que la consultation individuelle et privée sont interdites.

Fichier généré pour Visiteur (), le 31/05/2021