PROBLÈME 2

Objet : Etude de tirages successifs dans une population, et estimation asymptotique.

On effectue des tirages successifs avec remise dans une population comportant N individus.

Chaque individu a la même probabilité d'être tiré.

Ces individus sont notés a_1, a_2,..., a_N. Parmi eux, on distingue un échantillon de r individus a_1, a_2,..., a_r fixés.

On se propose d'étudier la v.a.r. X_r égale au nombre de tirages nécessaires pour que les r individus de l'échantillon aient été tirés.

Partie I

a) On suppose dans cette question que $r = 1$.

Déterminer la loi de X_1, et donner sans calcul $E(X_1)$ et $V(X_1)$.

b) (i) Déterminer l'ensemble des valeurs $X_r(\Omega)$ prises par X_r.

(ii) Calculer $P(X_r = r)$.

c) On se propose de calculer $P(X_r = r + 1)$.

On pose alors $A = (X_r = r + 1)$, et on note A_0, l'évènement : "les individus a_1, a_2,..., a_r de l'échantillon sont obtenus en $r + 1$ tirages, et la 1ère apparition de a_i précède la 1ère apparition de a_{i+1}, $1 \leq i \leq r - 1$".

(i) Exprimer $P(A)$ en fonction de $P(A_0)$.

(ii) A_0 est réalisé lorsque les $r+1$ premiers tirages ont amené a_1, $a_2,\ldots, a_r$ et un autre individu b.

Montrer que b peut être l'un quelconque des individus $a_1,\ldots a_N$, à l'exception de a_r.

(iii) Soit $A_{0,1}$, l'évènement : "A_0 est réalisé, et b est l'un des individus $a_{r+1},\ldots,a_N$".

Calculer $P(A_{0,1})$.

(iv) Soit maintenant $A_{0,2}$, l'évènement : "A_0 est réalisé, et b est l'un des individus $a_1,\ldots, a_{r-1}$".

Calculer $P(A_{0,2})$.

(v) Déduire de ce qui précède que : $P(E) = \dfrac{r!}{2.N^{r+1}}.r.(2.N-1-r)$.

Partie II

On cherche à estimer dans cette partie l'espérance $E(X_r)$ et la variance $V(X_r)$ de la v.a.r. X_r.

On note T_i la v.a.r. égale au nombre de tirages effectués, lorsque, pour la 1ère fois, i des individus $a_1, a_2,\ldots, a_r$ sont tirés.

On pose $U_1 = T_1$ et $U_i = T_i - T_{i-1}$, $2 \leq i \leq r$.

a) Interpréter les v.a.r. U_i, et justifier leur indépendance.

b) (i) Reconnaitre la loi de U_i.

(ii) Donner les valeurs de $E(U_i)$ et $V(U_i)$.

(iii) Déduire de ce qui précède l'expression de $E(X_r)$ et $V(X_r)$ en fonction des deux sommes suivantes :

$$S_{1,r} = \sum_{i=1}^{r} \frac{1}{i}\,;\; S_{2,r} = \sum_{i=1}^{r} \frac{1}{i^2}.$$

c) On se propose de déterminer un équivalent asymptotique de $E(X_r)$ et $V(X_r)$ lorsque r tend vers $+\infty$.

(i) Montrer que $\forall k \geq 1 : \dfrac{1}{k+1} \leq ln(k+1) - ln(k) \leq \dfrac{1}{k}$.

En déduire que $ln(r+1) + 1 - ln(2) \leq \displaystyle\sum_{i=1}^{r} \frac{1}{i} \leq ln(r)$.

(ii) Conclure que $S_{1,r} \underset{+\infty}{\approx} \ln(r)$.

Quel est le comportement asymptotique de $S_{2,r}$ lorsque r tend vers $+\infty$?

(iii) Donner un équivalent de $E(X_r)$ et de $V(X_r)$ lorsque r tend vers $+\infty$.

Partie III

On fait ici une étude asymptotique de la loi des v.a.r. X_r, lorsque le nombre N des individus de la population totale augmente indéfiniment.

a) Cette question a pour but de déterminer la fonction de répartition de X_r.

(i) Soit $n \in \mathbb{N}^*$. Soit B_k, l'évènement défini par : " k des individus de l'échantillon a_1, a_2,..., a_r n'ont pas été obtenus au bout de n tirages".

Montrer que $P(B_k) = C_r^k.(\frac{N-k}{N})^n$.

(ii) Soit C, l'évènement : "Tous les individus a_1, a_2, ..., a_r n'ont pas été obtenus à l'issue des n premiers tirages".

En appliquant convenablement la formule de Poincaré, montrer que :

$$P(C) = \sum_{k=1}^{r}(-1)^{k-1}.C_r^k.(1-\frac{k}{N})^n.$$

(iii) En déduire que $P(X_r \leq n) = \sum_{k=0}^{r}(-1)^k.C_r^k.(1-\frac{k}{N})^n$.

b) On suppose que N et n tendent vers $+\infty$ de sorte que $\frac{n}{N}$ tende vers p, $p > 0$.

Calculer la limite de $P(X_r \leq n)$ lorsque n tend vers $+\infty$.

c) Soit F, la fonction $\mathbb{R} \longrightarrow \mathbb{R}$ définie par :

$$\begin{cases} F(x) = K.(1-e^{-x})^r & si \quad x > 0 \\ F(x) = 0 & si \quad x \leq 0 \end{cases}.$$

(i) Montrer que, pour une valeur convenable de K que l'on déterminera, F est la fonction de répartition d'une v.a.r. continue Y.

(ii) Déterminer une densité f de Y, et construire les courbes de f et F pour $r \geqslant 2$.

(iii) Soient Z_1, Z_2,..., Z_r, r v.a.r. indépendantes suivant une loi exponentielle de paramètre $\lambda = 1$

(la condition d'indépendance se traduit par :

$\forall (x_1, x_2, \ldots, x_r) \in \mathbb{R}^r$:

$P(Z_1 \leq x_1, Z_2 \leq x_2, \ldots, Z_r \leq x_r) = P(Z_1 \leq x_1).$
$P(Z_2 \leq x_2) \cdots P(Z_r \leq x_r)$).

Montrer que la v.a.r. Y précédente a même loi que la v.a.r.
$Z = max(Z_1, Z_2, \cdots, Z_r)$.

d) Montrer l'existence de $E(Y)$, et calculer sa valeur.

Indication : Pour calculer $E(Y)$, on intègrera par parties, en prenant comme primitive de $t \mapsto r.e^{-t}.\left(1-e^{-t}\right)^{r-1}$ la fonction $t \mapsto \left(1-e^{-t}\right)^{r}-1$.

e) (i) Montrer les deux égalités suivantes :

$$(1) \qquad S_{1,q} = \sum_{k=1}^{q} \frac{1}{k} = \sum_{k=1}^{q} \frac{(-1)^{k-1}}{k}.C_q^k.$$

$$(2) \qquad \frac{S_{1,q+1}}{q+1} = \sum_{k=0}^{q} \frac{(-1)^k.C_q^k}{(k+1)^2}.$$

(ii) En déduire une relation entre $E(Y)$ et $S_{1,r}$.