VARIABLES ALÉATOIRES FINIES CORRECTION

Exercice 1

Une population de N individus est répartie en n groupes de k personnes chacun (donc $N = nk$). On suppose que dans cette population, il y a une proportion $p \in]0;1[$ de porteurs de Plasmodium falciparum (le parasite responsable du paludisme). On dispose d'un test permettant d'établir de façon certaine qu'un échantillon de sang contient ou non le parasite. Pour éviter d'effectuer N tests, on procède ainsi : pour chacun des N individus, on prélève un échantillon de sang et l'on répartit ces N prélévements selon n groupes prédéfinis G_1, G_2, ..., G_n (chaque groupe contenant k prélévements) ; pour chaque groupe G_i, on extrait du sang de chacun des k prélévements que l'on mélange pour obtenir un échantillon H_i ; on teste tous les échantillons H_1, H_2, ..., H_n : si le test de H_i est négatif, aucun des individus du groupe G_i n'est porteur du parasite et si le test H_i est positif, on teste un à un les k prélévements du groupe G_i pour détecter les individus infectés. On note X le nombre de groupes G_i pour lesquels le test de H_i est positif et T le nombre total de tests effectués par cette méthode.

1. *Déterminer la loi de X, son espérance et sa variance.*

   Le test de l'échantillon H_i est négatif si, et seulement si, les k individus du groupe G_i sont sains. Or chacun d'entre eux n'est pas malade avec probabilité $q = 1 - p$ et les individus sont indépendants vis-à-vis de la maladie, donc la probabilité qu'aucun des individus de G_i ne soient porteurs est égale à q^k. Par passage à l'événement contraire, on en déduit que la probabilité que le test de H_i soit positif est égale à $1 - q^k$.
   La variable aléatoire X compte alors le nombre de tests positifs dans une succession de n tests indépendants tels que pour chaque groupe, le test est positif avec probabilité $1-q^k$ et négatif avec la probabilité contraire. On reconnait donc un schéma binomial de paramètres n et $1 - q^k$, c'est-à-dire

   $$\boxed{X \hookrightarrow \mathscr{B}(n; 1 - q^k).}$$

   Par suite, on a, d'après le cours,

   $$\boxed{E(X) = n(1 - q^k) \qquad \text{et} \qquad V(X) = nq^k(1 - q^k).}$$

2. *Donner l'espérance et la variance de T en fonction de n, k et p.*

   Chaque groupe dont le test est négatif génère 1 test alors que chaque groupe dont le test est positif génère $k + 1$ tests (1 test pour le groupe et k tests pour les individus qui le constituent). Par suite, on a $T = (k + 1)X + 1(n - X)$, d'où
   $$T = kX + n.$$
   On en déduit que
   $$E(T) = E(kX + n) = kE(X) + n = kn(1 - q^k) + n$$
   et
   $$V(T) = V(kX + n) = k^2V(X) = k^2nq^k(1 - q^k),$$
   donc
   $$\boxed{E(T) = kn(1 - q^k) + n \qquad \text{et} \qquad V(T) = k^2nq^k(1 - q^k).}$$

3. *Que représente la variable aléatoire $nk - T$? Déterminer $E(nk - T)$.*

   La variable aléatoire $nk - T$ est la différence entre le nombre de tests qu'il aurait fallu faire avec une méthode de tests individu par individu et le nombre de tests effectués dans la méthode étudiée ici. Plus cette variable est grande, plus la méthode est efficace (en particulier, si elle prend une valeur négative, c'est que la méthode par groupes est moins efficace que la méthode exhaustive). Autrement dit,

   $$\boxed{nk - T \text{ mesure le nombre (algébrique) de tests économisés.}}$$

On a $E(nk-T) = nk - E(T) = nk - kn(1-q^k) - n$ d'après la question 2. b), d'où

$$\boxed{E(nk-T) = n(kq^k - 1)}$$

4. *Dans la province d'Isfahan en Iran, la proportion d'individus infectés vaut $p = 0,092$. Pour des raisons techniques, le nombre de groupes n est préalablement fixé à la valeur n_0. Quelle est la valeur optimale de k (c'est-à-dire la valeur de k pour laquelle $n_0k - T$ est maximale en moyenne) ?*

   On a
   $$E(n_0k - T) = n_0 f(k) \qquad \text{où} \qquad f(x) = xq^x - 1.$$
   La fonction f est de classe $\mathscr{C}^\infty$ sur $\mathbb{R}_+^*$ d'après les théorèmes généraux de dérivabilité et $\forall x > 0,\ f'(x) = (1 + x\ln q)q^x$. Alors $f'(x) \geqslant 0$ si, et seulement si, $x \leqslant 1/\ln(1/q)$. Par suite,

x	0		$1/\ln(1/q)$		$+\infty$
$f'(x)$	$\Vert$	$+$	0	$-$	
$f(x)$	$\Vert$ -1	$\nearrow$	α	$\searrow$	-1

   où $\alpha = \dfrac{1}{\mathrm{e}\ln(1/q)} - 1.$

   Donc $n_0 f(k)$ atteint son maximum en $\lfloor 1/\ln(1/q) \rfloor = 10$ ou $\lfloor 1/\ln(1/q) \rfloor + 1 = 11$. Or $f(10) \approx 2,809\,4$ et $f(11) \approx 2,804\,8$, donc

   $$\boxed{\text{la valeur optimale de } k \text{ est } 10.}$$

Exercice 2

On dispose de deux dés. L'un est pipé, l'autre non mais on ignore lequel (et on ne cherchera pas à le savoir). Comment faire pour simuler un dé non pipé avec ces deux dés ?

Quitte à peindre l'un des deux dés en rouge et l'autre en vert, on peut les supposer discernable. On note alors X_1 le résultat donné par le dé rouge et X_2 le résultat donné par le dé vert.

Déterminons la loi de $X_2 - X_1$. On a $(X_2 - X_1)(\Omega) = [\![-5;5]\!]$

On suppose que $X_1 \hookrightarrow \mathscr{U}([\![1;6]\!])$. Le cas où c'est la variable X_2 qui suit la loi uniforme sur $[\![1;6]\!]$ est semblable.

Pour tout $k \in [\![1;6]\!]$, on pose $p_k = P(X_2 = k)$.

On a

$$\begin{aligned}
P(X_2 - X_1 = 0) &= \sum_{k=1}^{6} P\big((X_1 = k) \cap (X_2 = k)\big) \\
&= \sum_{k=1}^{6} P(X_1 = k)P(X_2 = k) && \text{les lancers des dés sont indépendants} \\
&= \sum_{k=1}^{6} \frac{1}{6} p_k \\
&= \frac{1}{6} \sum_{k=1}^{6} p_k.
\end{aligned}$$

On a

$$\begin{aligned}
P(X_2 - X_1 = 1) &= \sum_{k=1}^{6} P\big((X_1 = k) \cap (X_2 = k+1)\big) \\
&= \sum_{k=1}^{6} P(X_1 = k)P(X_2 = k+1) && \text{les lancers des dés sont indépendants} \\
&= \sum_{k=1}^{5} \frac{1}{6} p_{k+1} \\
&= \frac{1}{6} \sum_{\ell=2}^{6} p_\ell.
\end{aligned}$$

Etc.

En définitive, on trouve que

$$\forall d \in [\![-5;5]\!]\,, \qquad P(X_2 - X_1 = d) = \begin{cases} \dfrac{1}{6} \displaystyle\sum_{k=d+1}^{6} p_k & \text{si } d \in [\![0;5]\!] \\ \dfrac{1}{6} \displaystyle\sum_{k=1}^{6+d} p_k & \text{si } d \in [\![-5;0]\!] \end{cases}$$

Dès lors, on constate que

$$P(X_2 - X_1 = 0) = \frac{1}{6},$$

et, pour tout $d \in [\![1;5]\!]$,

$$P(X_2 - X_1 = d-6) + P(X_2 - X_1 = d) = \frac{1}{6}\sum_{k=1}^{d} p_k + \frac{1}{6}\sum_{k=d+1}^{6} p_k = \frac{1}{6}\sum_{k=1}^{6} p_k = \frac{1}{6}\cdot$$

Donc

$$(X_2 - X_1 \mod 6) \hookrightarrow \mathscr{U}([\![1;6]\!]).$$

En conclusion,

$$\boxed{\text{on simule un dé non pipé en prenant la différence du dé rouge et du dé vert modulo 6.}}$$