

Correction du devoir maison

Partie I : la loi hypergéométrique.

1. On effectue k tirages successifs et sans remise dans une urne contenant a boules blanches et b boules noires, avec $k \leq \min(a, b)$. Au mieux, on tire donc k boules blanches (possible car $k \leq a$), et au pire on tire k boules noires (possible car $k \leq b$). Tous les intermédiaires étant possibles, on en déduit que $S_k(\Omega) = \llbracket 0, k \rrbracket$.

2. Rappelons qu'on peut assimiler le tirage sans remise de k boules au tirage simultané de k boules dans l'urne.

Dénombrons pour commencer le nombre de tirages possibles : il s'agit de choisir une partie à k éléments (les k boules tirées) dans un ensemble à $a + b$ éléments. Il y a donc $\binom{a+b}{k}$ manières de tirer simultanément k boules dans l'urne.

Fixons $i \in S_k(\Omega)$. On cherche à présent à dénombrer le nombre de tirages simultanés composés de i boules blanches, et donc de $k - i$ boules noires. Il s'agit pour cela de choisir les boules blanches parmi les a boules blanches, soit $\binom{a}{i}$ manières possibles. Et pour chaque choix de i boules blanches, il y a $\binom{b}{k-i}$ manières de choisir les $k - i$ boules noires parmi les b boules noires de l'urne. Il y a donc en tout $\binom{a}{i} \binom{b}{k-i}$ tirages comportant i boules blanches.

Les tirages étant tous équiprobables, on en déduit la formule :

$$P(S_k = i) = \frac{\binom{a}{i} \binom{b}{k-i}}{\binom{a+b}{k}}.$$

3. Si $k = 1$, on fait donc un tirage d'une boule dans une urne contenant a boules blanches et b boules noires. Dans ce cas, $S_1(\Omega) = \{0, 1\}$, et S_1 suit donc une loi de Bernoulli de paramètre $P(S_1 = 1) = \frac{a}{a+b}$.

Ainsi, $S_1 \hookrightarrow \mathcal{B}\left(\frac{a}{a+b}\right)$ et $E(S_1) = \frac{a}{a+b}$.

4. S_k est une variable aléatoire discrète finie. Elle admet donc une espérance.

5. Le but de cette question est de montrer par récurrence sur $k \in \mathbb{N}^*$ la propriété suivante, notée P_k : « quels que soient $(n_1, n_2) \in (\mathbb{N}^*)^2$ tels que $k \leq \min(n_1, n_2)$, si $X \hookrightarrow \mathcal{H}(k, n_1, n_2)$, alors $E(X) = k \frac{n_1}{n_1 + n_2}$ ».

(a) Pour $k = 1$, on a vu à la question 3. qu'alors $E(X) = \frac{a}{a+b}$. Donc la propriété

P_1 est vraie.

(b) On suppose à présent que P_k est vraie pour un certain entier $k \in \mathbb{N}^*$.

Soient n_1, n_2 tels que $k + 1 \leq \min(n_1, n_2)$. On considère une urne contenant n_1 boules blanches et n_2 boules noires. On effectue une suite de $k + 1$ tirages sans remise dans cette urne, et on note X_i la variable aléatoire qui vaut 1 si la i -ème boule tirée est blanche, et 0 sinon.

i. $X_1 + \dots + X_{k+1}$ représente le nombre de boules blanches obtenues au cours des $k + 1$ tirages sans remise dans l'urne.

ii. X_1 suit la loi de Bernoulli $\mathcal{B}\left(\frac{a}{a+b}\right)$.

D'après la question précédente, $X_1 + \dots + X_{k+1}$ représente le nombre de boules blanches obtenues au cours des $k + 1$ tirages sans remise dans l'urne. On reconnaît la situation de la question 1., de sorte que :

$$\boxed{X_1 + \dots + X_{k+1} \hookrightarrow \mathcal{H}(k + 1, n_1, n_2)}$$

- iii. On suppose l'évènement $[X_1 = 1]$ réalisé. On a donc obtenu une boule blanche au premier tirage. À l'issue de ce premier tirage, l'urne contient donc $n_1 - 1$ boules blanches, et n_2 boules noires. $X_2 + \dots + X_{k+1}$ représente le nombre de boules blanches obtenues lors de k tirages successifs sans remise dans une urne à $n_1 - 1$ boules blanches et n_2 boules noires. On reconnaît ici la situation de la question 1., ce qui permet de conclure que la loi de $X_2 + \dots + X_{k+1}$ sachant $[X_1 = 1]$ est la $\boxed{\text{loi hypergéométrique } \mathcal{H}(k, n_1 - 1, n_2)}$.

Supposons à présent l'évènement $[X_1 = 0]$ réalisé. On a cette fois obtenu une boule noire au premier tirage. À l'issue de ce tirage, l'urne contient donc n_1 boules blanches et $n_2 - 1$ boules noires. On peut donc conclure que la loi de $X_2 + \dots + X_{k+1}$ sachant $[X_1 = 0]$ est la $\boxed{\text{loi hypergéométrique } \mathcal{H}(k, n_1, n_2 - 1)}$.

- iv. On applique la formule de l'espérance totale à $X_2 + \dots + X_{k+1}$ avec pour SCE ($[X_1 = 0]$, $[X_1 = 1]$). On notera pour commencer qu'on n'a pas ici de problème de convergence puisque $X_2 + \dots + X_{k+1}$ est une variable finie. On peut donc appliquer la formule de l'espérance totale :

$$E(X_2 + \dots + X_{k+1}) = E(X_2 + \dots + X_{k+1} \mid [X_1 = 1])P(X_1 = 1) + E(X_2 + \dots + X_{k+1} \mid [X_1 = 0])P(X_1 = 0).$$

Par hypothèse de récurrence, puisque $k \leq \min(n_1 - 1, n_2)$:

$$E(X_2 + \dots + X_{k+1} \mid [X_1 = 1]) = k \frac{(n_1 - 1)}{(n_1 - 1) + n_2},$$

et de même :

$$E(X_2 + \dots + X_{k+1} \mid [X_1 = 0]) = k \frac{n_1}{n_1 + (n_2 - 1)}.$$

On en déduit que :

$$\begin{aligned} E(X_2 + \dots + X_{k+1}) &= k \frac{(n_1 - 1)}{(n_1 - 1) + n_2} \frac{n_1}{n_1 + n_2} + k \frac{n_1}{n_1 + (n_2 - 1)} \frac{n_2}{n_1 + n_2} \\ &= k \frac{(n_1 - 1)n_1 + n_1 n_2}{(n_1 + n_2 - 1)(n_1 + n_2)} = k \frac{(n_1 - 1 + n_2)n_1}{(n_1 + n_2 - 1)(n_1 + n_2)} = \boxed{k \frac{n_1}{n_1 + n_2}} \end{aligned}$$

- v. Par linéarité de l'espérance :

$$E(X_1 + X_2 + \dots + X_{k+1}) = E(X_1) + E(X_2 + \dots + X_{k+1}) = \frac{n_1}{n_1 + n_2} + k \frac{n_1}{n_1 + n_2}$$

$$\boxed{= (k + 1) \frac{n_1}{n_1 + n_2}}.$$

- (c) Puisque $X_1 + \dots + X_{k+1} \hookrightarrow \mathcal{H}(k + 1, n_1, n_2)$, nous venons de prouver que si P_k est vraie, alors P_{k+1} est vraie.

Par principe de récurrence, on peut donc conclure que pour tout $k \in \mathbb{N}^*$, pour tout $n_1, n_2 \in \mathbb{N}^*$ tels que $k \leq \min(n_1, n_2)$, et si $X \hookrightarrow \mathcal{H}(k, n_1, n_2)$, alors :

$$\boxed{E(X) = k \frac{n_1}{n_1 + n_2}}.$$

Partie II : variance de la loi hypergéométrique.

Soit $k \in \mathbb{N}^*$, et soient n_1, n_2 deux entiers tels que $k \leq \min(n_1, n_2)$ et soit X une variable aléatoire suivant la loi hypergéométrique $\mathcal{H}(k, n_1, n_2)$.

6. X étant une variable aléatoire à support fini, elle admet une variance.

7. $([X = i])_{i \in \llbracket 0, k \rrbracket}$ étant un système complet d'évènements, il suit :

$$\sum_{i=0}^k P(X = i) = 1,$$

soit encore :

$$\sum_{i=0}^k \frac{\binom{n_1}{i} \binom{n_2}{k-i}}{\binom{n_1+n_2}{k}} = 1.$$

D'où la formule suivante, appelée *identité de Vandermonde* :

$$\boxed{\binom{n_1 + n_2}{k} = \sum_{i=0}^k \binom{n_1}{i} \binom{n_2}{k-i}}$$

vraie pour tout $k \in \llbracket 0, \min(n_1, n_2) \rrbracket$.

8. On ne connaît pas la loi de $X(X - 1)$, il nous faut donc appliquer le théorème de transfert ici, possible car X est à support fini (pas de problème de convergence). On obtient :

$$\begin{aligned} E(X(X - 1)) &= \sum_{i=0}^k i(i - 1)P(X = i) \\ &= \frac{1}{\binom{n_1+n_2}{k}} \sum_{i=2}^k i(i - 1) \binom{n_1}{i} \binom{n_2}{k-i} \quad \text{puisque les termes en } i = 0, 1 \text{ sont nuls} \\ &= \frac{1}{\binom{n_1+n_2}{k}} \sum_{i=2}^k n_1(n_1 - 1) \binom{n_1 - 2}{i - 2} \binom{n_2}{k-i} \quad \text{car } i(i - 1) \binom{n_1}{i} = n_1(n_1 - 1) \binom{n_1 - 2}{i - 2} \\ &= \frac{n_1(n_1 - 1)}{\binom{n_1+n_2}{k}} \sum_{j=0}^{k-2} \binom{n_1 - 2}{j} \binom{n_2}{k-j-2} \quad \text{en posant } j = i - 2 \\ &= \frac{n_1(n_1 - 1)}{\binom{n_1+n_2}{k}} \binom{n_1 + n_2 - 2}{k - 2} \quad \text{par l'identité de Vandermonde} \\ &= n_1(n_1 - 1) \frac{k!(n_1 + n_2 - k)!}{(n_1 + n_2)!} \frac{(n_1 + n_2 - 2)!}{(k - 2)!(n_1 + n_2 - k)!} \\ &= \boxed{\frac{n_1(n_1 - 1)k(k - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)}} \end{aligned}$$

9. Par linéarité de l'espérance :

$$E(X(X - 1)) = E(X^2 - X) = E(X^2) - E(X)$$

et donc $E(X^2) = E(X(X - 1)) + E(X)$. On obtient :

$$\begin{aligned} E(X^2) &= \frac{n_1(n_1 - 1)k(k - 1)}{(n_1 + n_2)(n_1 + n_2 - 1)} + k \frac{n_1}{n_1 + n_2} = kn_1 \frac{(n_1 - 1)(k - 1) + n_1 + n_2 - 1}{(n_1 + n_2)(n_1 + n_2 - 1)} \\ &= kn_1 \frac{kn_1 - k + n_2}{(n_1 + n_2)(n_1 + n_2 - 1)}. \end{aligned}$$

Enfin par la formule de Huygens, on obtient :

$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = kn_1 \frac{kn_1 - k + n_2}{(n_1 + n_2)(n_1 + n_2 - 1)} - \frac{k^2 n_1^2}{(n_1 + n_2)^2} \\ &= kn_1 \frac{(kn_1 - k + n_2)(n_1 + n_2) - kn_1(n_1 + n_2 - 1)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \\ &= kn_1 \frac{kn_1^2 + kn_1 n_2 - kn_1 - kn_2 + n_1 n_2 + n_2^2 - kn_1^2 - kn_1 n_2 + kn_1}{(n_1 + n_2)^2(n_1 + n_2 - 1)} \\ &= \boxed{kn_1 n_2 \frac{n_1 + n_2 - k}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}. \end{aligned}$$

Partie III : application. (D'après oral ESCP 2014)

10. Les n questions constituant l'examen sont tirées sans remise parmi les N . Ainsi X suit une loi $\mathcal{H}(n, a, N - a)$ (les hypothèses $n \leq a$ et $n \leq N - a$ assurent qu'on est bien dans le cadre étudié dans les premières parties, car $n \leq \min(a, N - a)$).

11. Fixons $k \in \llbracket 0, n \rrbracket$, et supposons l'évènement $[X = k]$ réalisé.

$Y - X$ représente le nombre de questions pour lesquelles le candidat a bien répondu alors qu'il ne connaissait pas la réponse. Sachant que $[X = k]$, le nombre de questions dont il ne connaît pas la réponse est $n - k$. Puisqu'il a répondu au hasard pour chacune des questions, la loi de $Y - X$ sachant $[X = k]$ est une $\boxed{\text{loi binomiale } \mathcal{B}(n - k, \frac{1}{4})}$.

On en déduit donc immédiatement que $\boxed{E(Y - X | X = k) = \frac{n - k}{4}}$.

12. Du résultat de la question précédente, il vient par linéarité de l'espérance (et donc de l'espérance conditionnelle qui est une espérance elle aussi, pour la probabilité P_A , et qui possède donc des mêmes propriétés) :

$$E(Y - X | [X = k]) = E(Y | [X = k]) - E(X | [X = k]) = E(Y | [X = k]) - k.$$

On en déduit donc que $E(Y | [X = k]) = E(Y - X | [X = k]) + k = \frac{n - k}{4} + k = \frac{n + 3k}{4}$.

$([X = k])_{k \in \llbracket 0, n \rrbracket}$ est un système complet d'évènements. On applique la formule de l'espérance totale pour Y (pas de souci de convergence ici non plus, les variables sont finies) :

$$\begin{aligned} E(Y) &= \sum_{k=0}^n E(Y | [X = k]) P(X = k) \\ &= \frac{n}{4} \sum_{k=0}^n P(X = k) + \frac{3}{4} \sum_{k=0}^n k P(X = k) \\ &= \frac{n}{4} + \frac{3}{4} E(X) = \frac{n}{4} + \frac{3na}{4N} \\ &= \boxed{\frac{n}{4} \left(1 + \frac{3a}{N} \right)}. \end{aligned}$$

Pour que l'étudiant espère avoir la moyenne, il faut que :

$$\frac{n}{4} \left(1 + \frac{3a}{N} \right) \geq \frac{n}{2} \Leftrightarrow 1 + \frac{3a}{N} \geq 2 \Leftrightarrow a \geq \frac{N}{3}.$$

Ainsi $\boxed{\text{un étudiant qui aura révisé un tiers des questions peut espérer avoir la moyenne}}$ (mais rien ne garantit qu'il l'aura !).