RMS 2022 [648, 649, 651]

> *On considère un espace probabilisé* $(\Omega, \mathcal{A}, \mathbf{P})$.
> *Soient* $m < n$*, deux entiers strictement positifs. On note* E*, l'ensemble* $\mathfrak{P}_m(\llbracket 1, n\rrbracket)$ *des parties de* $\llbracket 1, n\rrbracket$ *dont le cardinal est égal à* m *et on considère une variable aléatoire* $A\ :\ \Omega \to E$ *:*
>
> $$\forall\, F \in \mathfrak{P}_m(\llbracket 1, n\rrbracket), \quad [A = F] \in \mathcal{A}$$
>
> *dont la loi est uniforme :*
>
> $$\forall\, F,\ G \in \mathfrak{P}_m(\llbracket 1, n\rrbracket), \quad \mathbf{P}(A = F) = \mathbf{P}(A = G).$$
>
> *Déterminer la loi de la variable aléatoire* $X = \max A - \min A$ *et son espérance.*

✍ ***Première variante***

Une urne contient n *boules numérotées de* 1 *à* n*. On extrait* m *boules de l'urne et on note* M*, le plus grand des numéros tirés. Calculer l'espérance et la variance de* M*.*

❧ ***Deuxième variante***

On répartit aléatoirement m *boules parmi* n *cases numérotées de* 1 *à* n *(chaque case pouvant contenir au plus une boule). Soit* Δ*, la différence entre le plus grand numéro et le plus petit numéro des cases occupées. Calculer la loi et l'espérance de* Δ*.*

Évidemment, $M = \max A$ *et* $\Delta = X$ *et je privilégie l'énoncé qui précise le modèle probabiliste à utiliser...*

❧ Comme $\#(E) = \binom{n}{m}$, on a $\mathbf{P}(A = F) = \frac{1}{\binom{n}{m}}$ pour toute partie $F \in E$.

❧ Pour toute partie $F \in E$, on a

$$1 \leqslant \min F \leqslant n - m + 1 \quad \text{et} \quad m \leqslant \max F \leqslant n.$$

Par conséquent, $\min A$ et $M = \max A$ sont des variables aléatoires discrètes à valeurs dans $\llbracket 1, n - m + 1\rrbracket$ et $\llbracket m, n\rrbracket$ respectivement.

✍ *Il est fastidieux, mais pas difficile de prouver que* $\min A$ *et* $\max A$ *sont effectivement des variables aléatoires discrètes : on peut écrire* $[\min A = k]$ *et* $[\max A = k]$ *comme des unions finies disjointes d'ensembles de la forme* $[A = F] \in \mathcal{A}$*.*

❧ Soit $m \leqslant k \leqslant n$. Le maximum de $F \in E$ est égal à k si, et seulement si, l'entier k appartient à F et si les $(m - 1)$ autres éléments de F sont choisis dans le sous-intervalle $\llbracket 1, k - 1\rrbracket$.

Comme la loi de A est uniforme sur E,

$$\forall\, m \leqslant k \leqslant n, \quad \mathbf{P}(M = k) = \mathbf{P}(\max A = k) = \frac{\binom{k-1}{m-1}}{\binom{n}{m}}.$$

❧ Soit $1 \leqslant k \leqslant n - m + 1$. Le minimum de $F \in E$ est égal à k si, et seulement si, l'entier k appartient à F et si les $(m - 1)$ autres éléments de F sont choisis dans le sous-intervalle $\llbracket k + 1, n\rrbracket$. Par conséquent,

$$\forall\, 1 \leqslant k \leqslant n - m + 1, \quad \mathbf{P}(\min A = k) = \frac{\binom{n-k}{m-1}}{\binom{n}{m}}.$$

✍ ***Remarque en passant***

Comme $[\max A = m] = \big[A = \{1, \dots, m\}\big] \subset [\min A = 1]$*, alors*

$$[\max A = m] \cap [\min A = 1] = [\max A = m]$$

et par conséquent

$$\mathbf{P}(\min A = 1, \max A = m) = \mathbf{P}(\max A = m) = \frac{1}{\binom{n}{m}} > 0.$$

Or la loi de $\min A$ *est connue et comme* $1 \leqslant m < n$*,*

$$\mathbf{P}(\min A = 1) = \frac{\binom{n-1}{m-1}}{\binom{n}{m}} = \frac{m}{n} < 1.$$

Donc $\mathbf{P}(\min A = 1, \max A = m) \neq \mathbf{P}(\min A = 1)\,\mathbf{P}(\max A = m)$ *et les deux variables* $\min A$ *et* $\max A$ *ne sont pas indépendantes.*

❧ Passons à la loi du couple $(\min A, \max A)$.

► Comme la partie $A(\omega)$ compte m éléments, alors nécessairement

$$\max A(\omega) - \min A(\omega) \geqslant m - 1.$$

D'autre part, $\max A(\omega) \leqslant n$ et $\min A(\omega) \geqslant 1$, donc

$$\max A(\omega) - \min A(\omega) \leqslant n - 1.$$

Ainsi X est une variable aléatoire discrète (en tant que différence de deux variables aléatoires discrètes) à valeurs dans $[\![m-1, n-1]\!]$.

► Soient donc deux entiers k et ℓ tels que

$$1 \leqslant k \leqslant n - m + 1, \qquad m \leqslant \ell \leqslant n \qquad \text{et} \qquad m - 1 \leqslant \ell - k \leqslant n - 1.$$

Le minimum de $F \in E$ est égal à k et son maximum est égal à ℓ si, et seulement si, la partie F contient k et ℓ et si les $(m-2)$ éléments restants de F sont choisis dans le sous-intervalle $[\![k+1, \ell-1]\!]$. L'astuce taupinale nous permet de dénombrer en toute sérénité :

$$\ell - 1 = k + (\ell - k - 1).$$

Il y a donc $(\ell - k - 1)$ entiers dans $[\![k+1, \ell-1]\!]$ et par conséquent,

$$\mathbf{P}(\min A = k, \max A = \ell) = \frac{\binom{\ell-k-1}{m-2}}{\binom{n}{m}}.$$

❧ On peut en déduire la loi de X.
Soit $m - 1 \leqslant q \leqslant n - 1$. D'après la Formule des probabilités totales,

$$\begin{aligned}
\mathbf{P}(X = q) &= \sum_{k=1}^{n-m+1} \mathbf{P}(X = q, \min A = k) \\
&= \sum_{\substack{1 \leqslant k \leqslant n-m+1 \\ k+q \leqslant n}} \mathbf{P}(\min A = k, \max A = k + q) \\
&= \sum_{k=1}^{n-q} \frac{\binom{q-1}{m-2}}{\binom{n}{m}} = (n - q)\frac{\binom{q-1}{m-2}}{\binom{n}{m}}
\end{aligned}$$

(puisque les termes de la somme sont tous égaux).

✍ *Pour calculer les espérances de* M *et de* X*, il faut se souvenir que*

$$\forall\, 1 \leqslant m < n, \qquad \sum_{k=m}^{n} \binom{k}{m} = \binom{n+1}{m+1}.$$

Cette propriété peut se démontrer
→ par récurrence sur n *(à partir de la formule dite du triangle de Pascal) ;*
→ de manière combinatoire : choisir $(m+1)$ *entiers dans* $[\![1, n+1]\!]$ *consiste à choisir d'abord le plus grand de ces entiers (on choisit* K *entre* $(m+1)$ *et* $(n+1)$*) et à choisir ensuite* m *entiers dans le sous-intervalle* $[\![1, K-1]\!]$ *— on prend* $k = K - 1$ *pour écrire la somme) ;*
→ en se fondant sur le fait que M *est une variable aléatoire à valeurs dans* $[\![m, n]\!]$ *et donc que*

$$\sum_{k=m}^{n} \mathbf{P}(M = k) = 1.$$

(Il convient ici de décaler les valeurs de m *et* n*.)*

❧ L'espérance de M est, par définition, égale à

$$\sum_{k=m}^{n} k\,\mathbf{P}(M=k) = \sum_{k=m}^{n} k\frac{\binom{k-1}{m-1}}{\binom{n}{m}} = \sum_{k=m}^{n} m\frac{\binom{k}{m}}{\binom{n}{m}} = \frac{m(n+1)}{m+1}.$$

❧ On peut calculer de manière analogue

$$\mathbf{E}(\min A) = \sum_{k=1}^{n-m+1} k\frac{\binom{n-k}{m-1}}{\binom{n}{m}} = (n+1) - \mathbf{E}(M).$$

(Il y a quelques astuces classiques en route.) Mais un tel résultat est surprenant et mérite d'être expliqué.

Comme l'application $\sigma : E \to E$ définie par

$$\forall\, F \in E, \quad \sigma(F) = \{n - x + 1,\ x \in F\}$$

est une bijection de E dans E (c'est même une sorte de symétrie, puisque $\sigma\circ\sigma = I_E$) et comme la loi de A est uniforme sur E, on en déduit que la loi de $B = \sigma(A)$ est aussi la loi uniforme sur E :

$$\forall\, F \in E, \quad \mathbf{P}(B = F) = \mathbf{P}\big(\sigma(A) = F\big) = \mathbf{P}\big(A = \sigma^{-1}(F)\big) = \frac{1}{\#(E)}.$$

Comme A et B ont même loi, les variables aléatoires $\min A$ et $\min B$ ont également même loi et en particulier

$$\mathbf{E}(\min A) = \mathbf{E}(\min B).$$

Par définition de σ,

$$\forall\, \omega \in \Omega, \quad y \in B(\omega) \iff (n+1) - y \in A(\omega)$$

donc

$$\forall\, \omega \in \Omega, \quad \min B(\omega) = (n+1) - \max A(\omega).$$

Donc, par linéarité de l'espérance,

$$\mathbf{E}(\min A) = \mathbf{E}(\min B) = (n+1) - \mathbf{E}(\max A) = \frac{n+1}{m+1}.$$

✍ *Il est intéressant de noter que*

$$\min A \overset{\text{loi}}{=} (n+1) - \max A$$

alors que ces deux variables ne sont certainement pas égales en tant que fonctions de Ω dans $[\![1, n]\!]$.

❧ Par linéarité de l'espérance, on trouve enfin

$$\mathbf{E}(X) = \mathbf{E}(\max A) - \mathbf{E}(\min A) = \frac{(m-1)(n+1)}{m+1}.$$