VARIABLES ALÉATOIRES SUR UN ESPACE PROBABILISÉ FINI

Dans tout ce chapitre, (Ω, P) est un espace probabilisé FINI et les lettres E, F, G désignent des ensembles quelconques.

1 INTRODUCTION AUX VARIABLES ALÉATOIRES

Définition (Variable aléatoire)

- On appelle *variable aléatoire sur Ω à valeurs dans E* toute application X de Ω dans E. Dans la plupart des situations, E est l'ensemble $\mathbb{R}$ et on dit que X est une *variable aléatoire réelle*.
- Pour toute partie A de E, l'événement $X^{-1}(A)$ est noté $\{X \in A\}$ ou $(X \in A)$. Dans le cas où A est un singleton $\{x\}$, on emploie plutôt les notations $\{X = x\}$ ou $(X = x)$. Enfin, dans le cas où $E = \mathbb{R}$ et où $A =]-\infty, x]$ pour un certain $x \in \mathbb{R}$, on emploie plutôt les notations $\{X \leqslant x\}$ et $(X \leqslant x)$ — même chose avec toutes les autres formes d'intervalles.

Exemple

- Pour modéliser le lancer d'un dé à 6 faces 2 fois, on peut choisir pour univers Ω l'ensemble $[\![1,6]\!]^2$ des 2-listes de $[\![1,6]\!]$ et pour probabilité P sur Ω la probabilité uniforme. La fonction X_1 (resp. X_2) « valeur obtenue au premier (resp. deuxième) lancer » est une variable aléatoire réelle sur Ω d'image $[\![1,6]\!]$. La fonction $S = X_1 + X_2$ « somme des deux faces obtenues » en est une autre, d'image $[\![2,12]\!]$.
- Pour modéliser le tirage simultané de k entiers entre 1 et n, on peut choisir pour univers Ω l'ensemble $\mathscr{P}_k\big([\![1,n]\!]\big)$ des k-combinaisons de $[\![1,n]\!]$ et pour probabilité P sur Ω la probabilité uniforme. La fonction « plus petit entier tiré » est une variable aléatoire réelle sur Ω d'image $[\![1,n-k+1]\!]$ — comme on tire k entiers, il est impossible que le plus petit tiré soit l'un des entiers $n-k+2, \ldots, n$.

1.1 LOI D'UNE VARIABLE ALÉATOIRE

Définition-théorème (Système complet d'événements associé à une variable aléatoire) Soit X une variable aléatoire sur Ω. Les événements $\{X = x\}$, x décrivant $X(\Omega)$, forment un système complet d'événements de Ω appelé le *système complet d'événements associé à X*.

Démonstration Tout élément ω de Ω appartient à un et un seul événement de la forme $\{X = x\}$ — en l'occurrence pour $x = X(\omega)$. En d'autres termes : $\Omega = \bigsqcup_{x \in X(\Omega)} \{X = x\}$. ■

Définition (Loi d'une variable aléatoire) Soit $X : \Omega \longrightarrow E$ une variable aléatoire. On appelle *loi de X* l'application $x \overset{P_X}{\longmapsto} P(X = x)$ de $X(\Omega)$ dans $[0,1]$. Bien remarquer que : $\displaystyle\sum_{x \in X(\Omega)} P(X = x) = 1$.

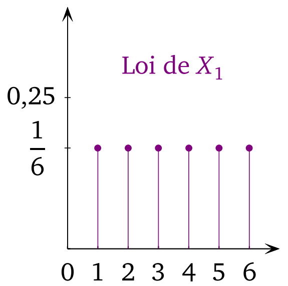
Explication

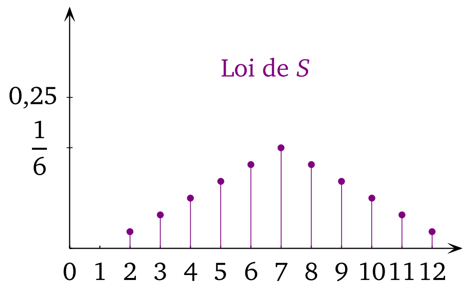
- La loi de X mesure la vraisemblance des valeurs de X. Les valeurs possibles de X, tout en étant possibles, n'ont en effet pas forcément les mêmes chances de réalisation et c'est précisément cette information que P_X mesure.
- La loi de X est définie sur $X(\Omega)$, mais on peut calculer $P(X = x)$ pour tout $x \in E \setminus X(\Omega)$: $P(X = x) = P(\varnothing) = 0$.

Exemple On reprend l'exemple du lancer d'un dé à 6 faces 2 fois où nous avons noté X_1 (resp. X_2) la valeur obtenue au premier (resp. deuxième) lancer et $S = X_1 + X_2$ la somme des deux faces obtenues. Que sont les lois de X_1 et S ?

Démonstration Rappelons qu'ici : $\Omega = [\![1,6]\!]^2$ et que P est la probabilité uniforme sur Ω.

- **Loi de X_1 :** X_1 a pour image $[\![1,6]\!]$.
  Pour tout $i \in [\![1,6]\!]$: $\{X_1 = i\} = \{i\} \times [\![1,6]\!]$, donc : $P(X_1 = i) = \dfrac{\left|\{X_1 = i\}\right|}{|\Omega|} = \dfrac{6}{36} = \dfrac{1}{6}$.
- **Loi de S :** S a pour image $[\![2,12]\!]$. Soit $s \in [\![2,12]\!]$. Les éléments de $\{S = s\}$ sont tous les couples de la forme $(i, s-i)$ pour lesquels : $1 \leqslant i \leqslant 6$ et $1 \leqslant s-i \leqslant 6$, i.e. : $1 \leqslant i \leqslant 6$ et $s-6 \leqslant i \leqslant s-1$.
  Du coup, pour tout $s \in [\![2,7]\!]$: $\{S = s\} = \{(i, s-i)\}_{1 \leqslant i \leqslant s-1}$, donc : $P(S = s) = \dfrac{\left|\{S = s\}\right|}{|\Omega|} = \dfrac{s-1}{36}$,
  et pour tout $s \in [\![8,12]\!]$: $\{S = s\} = \{(i, s-i)\}_{s-6 \leqslant i \leqslant 6}$, donc : $P(S = s) = \dfrac{\left|\{S = s\}\right|}{|\Omega|} = \dfrac{13-s}{36}$.





Définition (Lois conditionnelles d'une variable aléatoire) Soient X une variable aléatoire sur Ω et $A \in \mathscr{P}(\Omega)$ tel que : $P(A) > 0$. On appelle *loi conditionnelle de X sachant A* l'application $x \longmapsto P_A(X = x)$ de $X(\Omega)$ dans $[0,1]$.

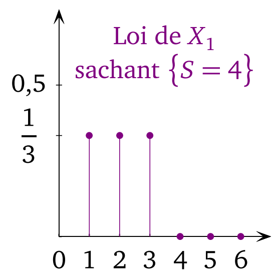
Exemple On reprend l'exemple du lancer d'un dé à 6 faces 2 fois où nous avons noté X_1 (resp. X_2) la valeur obtenue au premier (resp. deuxième) lancer et $S = X_1 + X_2$ la somme des deux faces obtenues. Quelle est la loi conditionnelle de X_1 sachant $\{S = 4\}$?

Démonstration Rappelons qu'ici : $\Omega = [\![1,6]\!]^2$ et que P est la probabilité uniforme sur Ω.

Pour tout $k \in [\![1,6]\!]$: $P(X_1 = k \text{ et } S = 4) = P(X_1 = k \text{ et } X_2 = 4-k) = \begin{cases} \dfrac{\left|\{(k, 4-k)\}\right|}{|\Omega|} = \dfrac{1}{36} & \text{si } k \in [\![1,3]\!] \\ 0 & \text{sinon,} \end{cases}$

et donc comme : $P(S = 4) = \dfrac{4-1}{36} = \dfrac{1}{12}$ d'après l'exemple précédent :

$$P_{\{S=4\}}(X_1 = k) = \frac{P(X_1 = k \text{ et } S = 4)}{P(S = 4)} = \begin{cases} \dfrac{1}{3} & \text{si } k \in [\![1,3]\!] \\ 0 & \text{sinon.} \end{cases}$$



Définition-théorème (Image d'une variable aléatoire par une fonction) Soient X une variable aléatoire sur Ω et $f : X(\Omega) \longrightarrow F$ une fonction. La variable aléatoire $f \circ X : \Omega \longrightarrow F$ est notée simplement $f(X)$.

La loi $P_{f(X)}$ de $f(X)$ est entièrement déterminée par f et la loi de X. Précisément, pour tout $y \in f\big(X(\Omega)\big)$:

$$P\big(f(X) = y\big) = \sum_{\substack{x \in X(\Omega) \\ y = f(x)}} P(X = x).$$

Démonstration Pour tout $y \in f\big(X(\Omega)\big)$, remarquer simplement que : $\{f(X) = y\} = \displaystyle\bigsqcup_{\substack{x \in X(\Omega) \\ y = f(x)}} \{X = x\}$. ■

Exemple On choisit un entier X au hasard entre 1 et $2n$. On pose : $Y = \left\lfloor \frac{X}{2} \right\rfloor$. Quelle est la loi de Y ?

Démonstration On choisit pour univers Ω l'ensemble $[\![1,2n]\!]$ et pour probabilité P sur Ω la probabilité uniforme. La variable X a pour image $[\![1,2n]\!]$ et pour tout $i \in [\![1,2n]\!]$: $\{X=i\} = \{i\}$, donc : $P(X=i) = \frac{1}{2n}$.
Quant à Y, son image est l'ensemble $[\![0,n]\!]$ et pour tout $j \in [\![0,n]\!]$: $P(Y=j) = \sum_{1 \leqslant i \leqslant 2n,\ j = \left\lfloor \frac{i}{2} \right\rfloor} P(X=i)$.

Conclusion : $P(Y=0) = P(X=1) = \frac{1}{2n}$, $P(Y=n) = P(X=2n) = \frac{1}{2n}$ et pour tout $j \in [\![1,n-1]\!]$:

$$P(Y=j) = P(X=2j) + P(X=2j+1) = \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$

1.2 Espérance d'une variable aléatoire réelle

Définition (Espérance d'une variable aléatoire réelle) Soit X une variable aléatoire réelle sur Ω.

- On appelle *espérance de X* le réel : $\mathrm{E}(X) = \sum_{x \in X(\Omega)} P(X=x)\, x$.
- On dit que X est *centrée* si : $\mathrm{E}(X) = 0$.

✌ **Explication** ✌ L'espérance de X est une moyenne des valeurs de X, donc un *indicateur de position*. Précisément, chaque valeur x, x décrivant $X(\Omega)$, s'y trouve comptabilisée en proportion de sa probabilité d'occurrence $P(X=x)$. Ainsi, plus $P(X=x)$ est proche de 1, plus x a d'importance dans le calcul.

Exemple On reprend l'exemple du lancer d'un dé à 6 faces 2 fois où nous avons noté X_1 la valeur obtenue au premier lancer. Que vaut l'espérance de X_1 ?

Démonstration Pour tout $i \in [\![1,6]\!]$: $P(X_1 = i) = \frac{1}{6}$, donc : $\mathrm{E}(X_1) = \sum_{i=1}^{6} P(X_1=i)\, i = \frac{1}{6}\sum_{i=1}^{6} i = \frac{7}{2}$.

Théorème (Propriétés de l'espérance) Soient X et Y deux variables aléatoires réelles sur Ω.

(i) **Une autre expression :** $\mathrm{E}(X) = \sum_{\omega \in \Omega} P\big(\{\omega\}\big)\, X(\omega)$.

(ii) **Linéarité :** Pour tous $\lambda, \mu \in \mathbb{R}$: $\mathrm{E}(\lambda X + \mu Y) = \lambda \mathrm{E}(X) + \mu \mathrm{E}(Y)$.

(iii) **Positivité :** Si $X \geqslant 0$, alors $\mathrm{E}(X) \geqslant 0$. **Croissance :** Si $X \leqslant Y$, alors $\mathrm{E}(X) \leqslant \mathrm{E}(Y)$.

(iv) **Inégalité triangulaire :** $\big|\mathrm{E}(X)\big| \leqslant \mathrm{E}\big(|X|\big)$.

Démonstration

(i) Les événements $\{X=x\}$, x décrivant $X(\Omega)$, forment un système complet d'événements, donc :

$$\sum_{\omega \in \Omega} P\big(\{\omega\}\big) X(\omega) = \sum_{x \in X(\Omega)} \sum_{\omega \in \{X=x\}} P\big(\{\omega\}\big) X(\omega) = \sum_{x \in X(\Omega)} \left(\sum_{\omega \in \{X=x\}} P\big(\{\omega\}\big) \right) x = \sum_{x \in X(\Omega)} P(X=x)\, x = \mathrm{E}(X).$$

(ii) $$\mathrm{E}(\lambda X + \mu Y) = \sum_{\omega \in \Omega} P\big(\{\omega\}\big)\, (\lambda X + \mu Y)(\omega) = \lambda \sum_{\omega \in \Omega} P\big(\{\omega\}\big)\, X(\omega) + \mu \sum_{\omega \in \Omega} P\big(\{\omega\}\big)\, Y(\omega) = \lambda \mathrm{E}(X) + \mu \mathrm{E}(Y).$$

(iii) Positivité évidente par définition de l'espérance. Pour la croissance, si $X \leqslant Y$, alors : $Y - X \geqslant 0$, donc par linéarité : $\mathrm{E}(Y) - \mathrm{E}(X) = \mathrm{E}(Y-X) \geqslant 0$.

(iv) Conséquence de la croissance, sachant que : $-|X| \leqslant X \leqslant |X|$. ■

Exemple Soit $m \in \mathbb{R}$. Si on considère m comme une variable aléatoire constante sur Ω, alors : $\mathrm{E}(m) = m$.

Démonstration Afin d'y voir clair, notons X la variable aléatoire constante de valeur m sur Ω. Dans ces conditions : $X(\Omega) = \{m\}$ et $P(X=m) = P(\Omega) = 1$, donc : $\mathrm{E}(X) = m \times 1 = m$.

Exemple On reprend l'exemple du lancer d'un dé à 6 faces 2 fois où nous avons noté X_1 (resp. X_2) la valeur obtenue au premier (resp. deuxième) lancer et $S = X_1 + X_2$ la somme des deux faces obtenues. Que vaut l'espérance de S ?

Démonstration Pour tout $s \in [\![2,7]\!]$: $P(S=s) = \dfrac{s-1}{36}$ et pour tout $s \in [\![8,12]\!]$: $P(S=s) = \dfrac{13-s}{36}$, donc : $\mathrm{E}(S) = \displaystyle\sum_{s=2}^{12} P(S=s)\, s = \sum_{s=2}^{7} \frac{s-1}{36} \times s + \sum_{s=8}^{12} \frac{13-s}{36} \times s$, mais c'est compliqué ! Or n'oublions pas que : $S = X_1 + X_2$. Nous avons calculé plus haut la loi puis l'espérance de X_1 et il n'est pas dur de se convaincre qu'on aurait trouvé le même résultat pour X_2. Conclusion : $\mathrm{E}(S) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = \dfrac{7}{2} + \dfrac{7}{2} = 7$.

Théorème (Formule de transfert) Soient X une variable aléatoire sur Ω et $f : X(\Omega) \longrightarrow \mathbb{R}$ une fonction. L'espérance de $f(X)$ est entièrement déterminée par f et la loi de X :

$$\mathrm{E}\big(f(X)\big) = \sum_{x \in X(\Omega)} P(X=x)\, f(x).$$

Démonstration Les événements $\{X = x\}$, x décrivant $X(\Omega)$, forment un système complet d'événements, donc :

$$\mathrm{E}\big(f(X)\big) = \sum_{\omega\in\Omega} P\big(\{\omega\}\big)\, f\big(X(\omega)\big) = \sum_{x\in X(\Omega)} \sum_{\omega\in\{X=x\}} P\big(\{\omega\}\big)\, f\big(X(\omega)\big) = \sum_{x\in X(\Omega)} \left(\sum_{\omega\in\{X=x\}} P\big(\{\omega\}\big)\right) f(x)$$
$$= \sum_{x\in X(\Omega)} P(X=x)\, f(x).$$ ■

Exemple On reprend l'exemple du lancer d'un dé à 6 faces 2 fois où nous avons noté X_1 la valeur obtenue au premier lancer. Que vaut l'espérance de X_1^2 ?

Démonstration $\mathrm{E}\big(X_1^2\big) = \displaystyle\sum_{i=1}^{6} P(X_1 = i)\, i^2 = \frac{1}{6}\sum_{i=1}^{6} i^2 = \frac{1}{6} \times \frac{6\times 7\times 13}{6} = \frac{91}{6}$.

2 Lois usuelles

Pour toute variable aléatoire X sur un espace probabilisé fini (Ω, P) : $\displaystyle\sum_{x\in X(\Omega)} P(X=x) = 1$. Réciproquement, si on se donne un ensemble fini $E = \{x_1, \dots, x_n\}$ et des réels $p_1, \dots, p_n \in [0,1]$ pour lesquels : $\displaystyle\sum_{i=1}^{n} p_i = 1$, existe-t-il toujours une variable aléatoire X sur un certain espace probabilisé (Ω, P) telle que pour tout $i \in [\![1,n]\!]$: $P(X = x_i) = p_i$? La réponse est oui, on peut par exemple poser : $\Omega = E$ et $X = \mathrm{Id}_E$ et définir une probabilité P sur Ω en posant pour tout $i \in [\![1,n]\!]$: $P\big(\{x_i\}\big) = p_i$ — on obtient bien ainsi, comme voulu : $P(X = x_i) = P(\mathrm{Id}_E = x_i) = P\big(\{x_i\}\big) = p_i$. Cette remarque justifie la consistance de tous les énoncés de théorèmes et d'exercices qui, pour définir une variable aléatoire, nous en donnent seulement l'image et les probabilités des événements élémentaires.

Définition-théorème (Loi uniforme) Soient X une variable aléatoire et E un ensemble FINI non vide. On dit que X *suit la loi uniforme sur E*, ce qu'on note : $X \hookrightarrow \mathscr{U}(E)$, si pour tout $x \in E$: $P(X=x) = \dfrac{1}{|E|}$.

L'espérance de X est alors la moyenne « naturelle » des valeurs $x_1, \dots, x_n$ de X : $\mathrm{E}(X) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} x_k$.

✌ **Explication** ✌ La loi uniforme est la loi pour laquelle les événements $\{X = x\}$, x décrivant E, sont équiprobables.

Démonstration Pour tout $k \in [\![1,n]\!]$: $P(X = x_k) = \dfrac{1}{n}$, donc : $\mathrm{E}(X) = \displaystyle\sum_{k=1}^{n} \frac{x_k}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k$. ■

Exemple

- Lorsqu'on choisit au hasard un entier X entre 1 et n, X suit la loi uniforme $\mathscr{U}\big([\![1,n]\!]\big)$.
- Lorsqu'on lance simultanément un dé bleu et un dé rouge à 6 faces, le couple (B,R) des valeurs obtenues avec ces dés — avec B pour le dé bleu et R pour le rouge — suit la loi uniforme $\mathscr{U}\big([\![1,6]\!]^2\big)$.

Exemple On choisit un entier X au hasard entre 1 et $2n$. Quelle est la loi de X sachant $\{X \leqslant n\}$? Quelle est loi de $(-1)^X$?

Démonstration Clairement, X suit la loi uniforme $\mathscr{U}\big([\![1,2n]\!]\big)$.

- **Loi conditionnelle de X sachant $\{X \leqslant n\}$:** $P(X \leqslant n) = \displaystyle\sum_{k=1}^{n} P(X=k) = \sum_{k=1}^{n} \frac{1}{2n} = \frac{1}{2}$, donc pour tout $k \in [\![1,n]\!]$: $P_{\{X\leqslant n\}}(X=k) = \dfrac{P(X=k)}{P(X\leqslant n)} = \dfrac{\frac{1}{2n}}{\frac{1}{2}} = \dfrac{1}{n}$ et pour tout $k \in [\![n+1,2n]\!]$: $P_{\{X\leqslant n\}}(X=k) = 0$.
  La loi conditionnelle de X sachant $\{X \leqslant n\}$ est ainsi la loi uniforme $\mathscr{U}\big([\![1,n]\!]\big)$.
- **Loi de $(-1)^X$:** La variable aléatoire $(-1)^X$ suit la loi uniforme sur $\{-1,1\}$ car :
$$P\Big((-1)^X = 1\Big) = \sum_{i=1}^{n} P(X=2i) = \sum_{i=1}^{n} \frac{1}{2n} = \frac{1}{2} \qquad \text{et} \qquad P\Big((-1)^X = -1\Big) = \sum_{i=0}^{n-1} P(X=2i+1) = \sum_{i=0}^{n-1} \frac{1}{2n} = \frac{1}{2}.$$

Définition-théorème (Loi de Bernoulli)

- Soient X une variable aléatoire et $p \in [0,1]$. On dit que X *suit la loi de Bernoulli de paramètre p*, ce qu'on note $X \hookrightarrow \mathscr{B}(p)$, si : $P(X=1) = p$ et $P(X=0) = 1-p$.

  L'espérance de X vaut alors : $\mathrm{E}(X) = p$.
- **Exemple fondamental :** Pour tout événement $A \in \mathscr{P}(\Omega)$ de probabilité p : $\mathbb{1}_A \hookrightarrow \mathscr{B}(p)$.
  En particulier : $\mathrm{E}(\mathbb{1}_A) = P(A)$.

Démonstration

- **Espérance :** $X(\Omega) = \{0,1\}$, donc : $\mathrm{E}(X) = P(X=1)\times 1 + P(X=0)\times 0 = p$.
- **Exemple fondamental :** $P(\mathbb{1}_A = 1) = P(A) = p$ et $P(\mathbb{1}_A = 0) = P\big(\overline{A}\big) = 1-p$. ■

Exemple Pour tous $A_1,\ldots,A_n \in \mathscr{P}(\Omega)$: $P\left(\displaystyle\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} P\big(A_{i_1} \cap \ldots \cap A_{i_k}\big)$ (formule du crible).

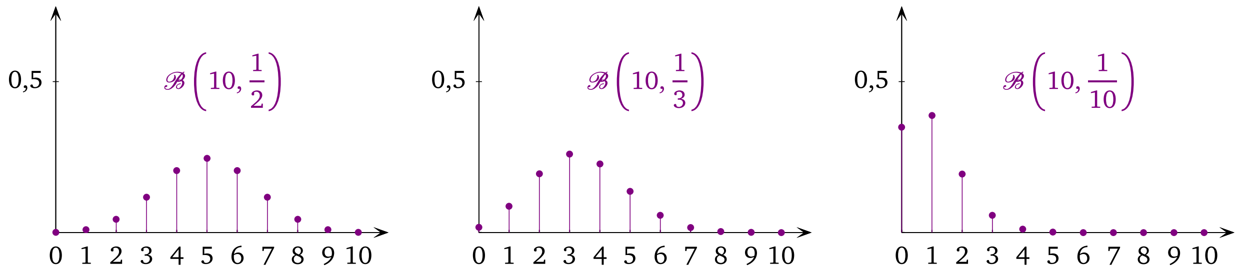
Démonstration Cette formule hors programme a été énoncée sans preuve au chapitre « Probabilités sur un espace probabilisé fini ». On commence par un calcul d'indicatrices :
$$\mathbb{1}_{A_1\cup\ldots\cup A_n} = 1 - \mathbb{1}_{\overline{A_1\cup\ldots\cup A_n}} = 1 - \mathbb{1}_{\overline{A_1}\cap\ldots\cap\overline{A_n}} = 1 - \prod_{i=1}^{n} \mathbb{1}_{\overline{A_i}} = 1 - \prod_{i=1}^{n} \big(1 - \mathbb{1}_{A_i}\big) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathbb{1}_{A_{i_1}\cap\ldots\cap A_{i_k}},$$
puis par linéarité de l'espérance :
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \mathrm{E}\big(\mathbb{1}_{A_1\cup\ldots\cup A_n}\big) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathrm{E}\big(\mathbb{1}_{A_{i_1}\cap\ldots\cap A_{i_k}}\big)$$
$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} P\big(A_{i_1} \cap \ldots \cap A_{i_k}\big).$$

Intéressons-nous à présent à la répétition n fois indépendamment d'une expérience aléatoire à deux issues, disons « favorable » et « défavorable », de probabilité p pour l'issue favorable. Quelle est loi du nombre X d'issues favorables ? La variable aléatoire X a pour image $[\![0,n]\!]$ et pour tout $k \in [\![0,n]\!]$, l'événement $\{X=k\}$ est réalisé si et seulement si on a obtenu k issues favorables et $n-k$ issues défavorables. Or il y a $\dbinom{n}{k}$ manières de distribuer ces k issues favorables et chacune de ces possibilités a la même probabilité $p^k(1-p)^{n-k}$ d'arriver, donc : $P(X=k) = \dbinom{n}{k} p^k (1-p)^{n-k}$.

Définition-théorème **(Loi binomiale ou loi des tirages AVEC remise)** Soient X une variable aléatoire et $p \in [0,1]$. On dit que X *suit la loi binomiale de paramètre* (n,p), ce qu'on note : $X \hookrightarrow \mathscr{B}(n,p)$, si pour tout $k \in [\![0,n]\!]$:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

L'espérance de X vaut alors : $\mathrm{E}(X) = np$.



✌ **Explication** ✌ Lorsqu'on répète n fois indépendamment une expérience aléatoire à deux issues de probabilité p pour l'issue favorable, le **NOMBRE D'ISSUES FAVORABLES** suit la loi binomiale $\mathscr{B}(n,p)$.

- L'expression « loi des tirages AVEC remise » par laquelle on décrit parfois la loi binomiale trouve son explication dans l'exemple suivant. Lorsqu'on tire AVEC REMISE — donc indépendamment — n boules dans une urne contenant des boules blanches en proportion p et des boules noires en proportion $1-p$, la variable aléatoire « nombre de boules blanches tirées » suit la loi binomiale $\mathscr{B}(n,p)$.
- Clairement, la loi binomiale $\mathscr{B}(1,p)$ n'est autre que la loi de Bernoulli $\mathscr{B}(p)$!

Démonstration

- **Cohérence de la définition :** D'après la formule du binôme : $\displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \big(p+(1-p)\big)^n = 1.$
- **Espérance :** Dans $\mathbb{R}[T]$: $\displaystyle (T+1-p)^n = \sum_{k=0}^{n} \binom{n}{k} T^k (1-p)^{n-k}.$ Dérivons puis multiplions par T :

$$nT(T+1-p)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} kT^k (1-p)^{n-k}. \quad \text{Évaluons en } p : \quad np = \sum_{k=0}^{n} \binom{n}{k} kp^k (1-p)^{n-k}.$$

Finalement : $\displaystyle \mathrm{E}(X) = \sum_{k=0}^{n} P(X=k)\, k = \sum_{k=0}^{n} \binom{n}{k} kp^k (1-p)^{n-k} = np.$ ■

Exemple On lance 5 fois un dé équilibré à 6 faces dont 2 blanches et 4 noires. Avec quelle probabilité obtient-on exactement 3 fois une face noire ?

Démonstration La face noire a une probabilité $\dfrac{4}{6} = \dfrac{2}{3}$ d'apparaître quand on lance le dé une fois. Par indépendance des lancers, le nombre N de faces noires obtenues à l'issue des 5 lancers suit donc la loi binomiale $\mathscr{B}\left(5, \dfrac{2}{3}\right)$, et donc enfin : $\displaystyle P(N=3) = \binom{5}{3} \left(\frac{2}{3}\right)^3 \left(1-\frac{2}{3}\right)^{5-3} = \frac{80}{243} \approx 0,33.$

3 Famille de variables aléatoires et indépendance

3.1 Couples de variables aléatoires

Nous ne parlerons que de **COUPLES** de variables aléatoires dans ce paragraphe, mais les définitions et résultats présentés se généralisent sans difficulté aux familles d'un nombre fini quelconque de variables aléatoires.

Remarquons pour commencer que les couples de variables aléatoires ne sont pas une notion nouvelle. Nous avons autorisé nos variables aléatoires à prendre leurs valeurs dans un ensemble quelconque, un couple (X,Y) de variables aléatoires n'est donc jamais qu'une variable aléatoire à valeurs dans un ensemble produit.

Définition-théorème (Couple de variables aléatoires, loi conjointe et lois marginales) Soient X et Y deux variables aléatoires sur Ω.

- La variable aléatoire $\left\{\begin{array}{ccc}\Omega & \longrightarrow & X(\Omega)\times Y(\Omega)\\ \omega & \longmapsto & \big(X(\omega),Y(\omega)\big)\end{array}\right.$ est notée simplement (X,Y) et sa loi $P_{(X,Y)}$ est souvent appelée la *loi conjointe de* X *et* Y.
- La loi P_X de X est appelée la *première loi marginale de* (X,Y) et la loi P_Y de Y sa *deuxième loi marginale*.
- Les événements $\big\{X=x \text{ et } Y=y\big\}$, (x,y) décrivant $(X,Y)(\Omega)$ ou $X(\Omega)\times Y(\Omega)$, forment un système complet d'événements appelé le *système complet d'événements associé à* (X,Y).

Exemple Une urne contient 3 boules blanches et 1 noire et on en tire successivement 2 boules sans remise. Le couple de couleurs ainsi obtenu est noté (C_1,C_2) — avec B pour « blanc » et N pour « noir ». Quelle est la loi de (C_1,C_2) ?

Démonstration $P(C_1=B \text{ et } C_2=B)=P(C_1=B)\,P_{\{C_1=B\}}(C_2=B)=\dfrac{3}{4}\times\dfrac{2}{3}=\dfrac{1}{2}$

et de même : $P(C_1=B \text{ et } C_2=N)=\dfrac{3}{4}\times\dfrac{1}{3}=\dfrac{1}{4}$,

$P(C_1=N \text{ et } C_2=B)=\dfrac{1}{4}\times\dfrac{3}{3}=\dfrac{1}{4}$ et $P(C_1=N \text{ et } C_2=N)=\dfrac{1}{4}\times\dfrac{0}{3}=0$.

$C_1 \backslash C_2$	B	N
B	$\frac{1}{2}$	$\frac{1}{4}$
N	$\frac{1}{4}$	0

Théorème (Calcul des lois marginales à partir de la loi conjointe) Soient X et Y deux variables aléatoires sur Ω. La loi (conjointe) $P_{(X,Y)}$ du couple (X,Y) détermine entièrement ses lois marginales P_X et P_Y. Précisément, pour tout $x\in X(\Omega)$: $P(X=x)=\displaystyle\sum_{y\in Y(\Omega)}P(X=x \text{ et } Y=y)$ et pour tout $y\in Y(\Omega)$: $P(Y=y)=\displaystyle\sum_{x\in X(\Omega)}P(X=x \text{ et } Y=y)$.

Démonstration Pour tout $x\in X(\Omega)$:

$$P(X=x)=P\left(\{X=x\}\ \cap \bigsqcup_{y\in Y(\Omega)}\{Y=y\}\right)=P\left(\bigsqcup_{y\in Y(\Omega)}\{X=x \text{ et } Y=y\}\right)=\sum_{y\in Y(\Omega)}P(X=x \text{ et } Y=y).$$ ■

✌ **Explication** ✌ On représente parfois la loi de (X,Y) sous la forme d'un tableau à deux entrées donnant $P(X=x \text{ et } Y=y)$ en fonction des valeurs possibles de $x\in X(\Omega)$ en lignes et de $y\in Y(\Omega)$ en colonnes. La figure ci-contre illustre alors le calcul de P_X et P_Y à partir de $P_{(X,Y)}$. On a noté : $X(\Omega)=\big\{x_1,\ldots,x_m\big\}$, $Y(\Omega)=\big\{y_1,\ldots,y_n\big\}$ et $p_{ij}=P(X=x_i \text{ et } Y=y_j)$ pour tout $(i,j)\in[\![1,m]\!]\times[\![1,n]\!]$.

$X \backslash Y$	y_1	y_2	$\cdots$	y_n	
x_1	p_{11}	p_{12}	$\cdots$	p_{1n}	$\longrightarrow \displaystyle\sum_{j=1}^n p_{1j}=P(X=x_1)$
x_2	p_{21}	p_{22}	$\cdots$	p_{2n}	$\longrightarrow \displaystyle\sum_{j=1}^n p_{2j}=P(X=x_2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_m	p_{m1}	p_{m2}	$\cdots$	p_{mn}	$\longrightarrow \displaystyle\sum_{j=1}^n p_{mj}=P(X=x_m)$

Première loi marginale P_X de (X,Y)

$$P(Y=y_1)=\sum_{i=1}^m p_{i1}\qquad \cdots \qquad \sum_{i=1}^m p_{in}=P(Y=y_n)$$

Deuxième loi marginale P_Y de (X,Y)

Théorème (Image d'un couple de variables aléatoires par une fonction et formule de transfert associée) Soient X et Y deux variables aléatoires sur Ω et $f:(X,Y)(\Omega)\longrightarrow F$ une fonction.

- **Loi :** Pour tout $z\in f(X,Y)(\Omega)$: $$P\big(f(X,Y)=z\big)=\sum_{\substack{(x,y)\in(X,Y)(\Omega)\\ z=f(x,y)}}P(X=x \text{ et } Y=y).$$
- **Formule de transfert :** $$\mathrm{E}\big(f(X,Y)\big)=\sum_{(x,y)\in(X,Y)(\Omega)}P(X=x \text{ et } Y=y)\,f(x,y).$$

Démonstration Ce ne sont pas là de nouveaux théorèmes. Il suffit de poser : $Z = (X, Y)$ pour s'en rendre compte, car un couple de variables aléatoires n'est jamais qu'une variable aléatoire de plein droit à valeurs dans un ensemble produit. ■

✌ **Explication** ✌ En particulier :

$$\boxed{P(X+Y=z) = \sum_{\substack{(x,y)\in(X,Y)(\Omega)\\ z=x+y}} P(X=x \text{ et } Y=y)} \quad \text{et} \quad \boxed{P(XY=z) = \sum_{\substack{(x,y)\in(X,Y)(\Omega)\\ z=xy}} P(X=x \text{ et } Y=y).}$$

Exemple On reprend encore l'exemple du lancer d'un dé à 6 faces 2 fois où nous avons noté X_1 (resp. X_2) la valeur obtenue au premier (resp. deuxième) lancer et $S = X_1 + X_2$ la somme des deux faces obtenues. Quelle est la loi conjointe du couple (X_1, X_2) ? Et la loi de $|X_2 - X_1|$?

Démonstration Rappelons qu'ici : $\Omega = [\![1,6]\!]^2$ et que P est la probabilité uniforme sur Ω.

- **Loi de** (X_1, X_2) **:** Le couple (X_1, X_2) a pour image $[\![1,6]\!]^2$ et pour tout $(i,j) \in [\![1,6]\!]^2$:

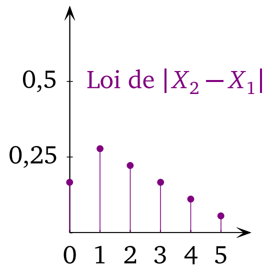
$$P(X_1 = i \text{ et } X_2 = j) = \frac{\big|\{(i,j)\}\big|}{|\Omega|} = \frac{1}{36}, \qquad \text{donc :} \quad (X_1, X_2) \hookrightarrow \mathscr{U}\big([\![1,6]\!]^2\big).$$

- **Loi de** $|X_2 - X_1|$ **:** $|X_2 - X_1|$ a pour image $[\![0,5]\!]$ et pour tout $d \in [\![0,5]\!]$:

$$P\big(|X_2 - X_1| = d\big) = P\big(X_2 - X_1 = d \text{ ou } X_1 - X_2 = d\big) = \begin{cases} P(X_1 = X_2) & \text{si } d = 0 \\ P(X_2 - X_1 = d) + P(X_1 - X_2 = d) & \text{sinon.} \end{cases}$$

Or : $P(X_1 = X_2) = \sum_{k=1}^{6} P(X_1 = k \text{ et } X_2 = k) = \sum_{k=1}^{6} \frac{1}{36} = \frac{1}{6}$ et pour tout $d \in [\![1,5]\!]$:

$$\begin{aligned} P\big(|X_2 - X_1| = d\big) &= P(X_2 - X_1 = d) + P(X_1 - X_2 = d) \\ &= \sum_{i=1}^{6-d} P(X_1 = i \text{ et } X_2 = d+i) + \sum_{j=1}^{6-d} P(X_1 = d+j \text{ et } X_2 = j) \\ &= \sum_{i=1}^{6-d} \frac{1}{36} + \sum_{j=1}^{6-d} \frac{1}{36} = 2 \times \frac{6-d}{36} = \frac{6-d}{18}. \end{aligned}$$



Exemple Dans un centre d'appels, un employé effectue n appels téléphoniques vers n correspondants distincts dont chacun décroche avec probabilité p.

- On note N_1 le nombre de correspondants qui ont décroché. Quelle est donc la loi de N_1 ? Réponse SANS CALCUL : $N_1 \hookrightarrow \mathscr{B}(n,p)$, car les appels sont indépendants et la probabilité d'obtenir un correspondant ne dépend pas du correspondant choisi.
- L'employé rappelle un peu plus tard les $n - N_1$ correspondants qui n'ont pas décroché lors de sa première série d'appels. On note N_2 le nombre de ces correspondants qui décrochent cette fois et N le nombre total des correspondants qui ont décroché. Quelle est la loi de N ?

Démonstration On cherche la loi de $N = N_1 + N_2$.

- Pour tout $k \in [\![0,n]\!]$, sous l'hypothèse que $N_1 = k$, les $n-k$ appels de la deuxième série d'appels satisfont les mêmes hypothèses que ceux de la première série, donc la loi conditionnelle de N_2 sachant $\{N_1 = k\}$ est la loi binomiale $\mathscr{B}(n-k,p)$.
- La variable aléatoire N a pour image $[\![0,n]\!]$ et pour tout $s \in [\![0,n]\!]$:

$$\begin{aligned} P(N=s) = P(N_1 + N_2 = s) &= \sum_{\substack{0 \leqslant i,j \leqslant n \\ i+j=s}} P(N_1 = i \text{ et } N_2 = j) = \sum_{i=0}^{s} P(N_1 = i \text{ et } N_2 = s-i) \\ &= \sum_{i=0}^{s} P(N_1 = i)\, P_{\{N_1 = i\}}(N_2 = s-i) = \sum_{i=0}^{s} \binom{n}{i} p^i (1-p)^{n-i} \binom{n-i}{s-i} p^{s-i} (1-p)^{(n-i)-(s-i)}. \end{aligned}$$

Or avec des notations évidentes :

$$\binom{n}{i}\binom{n-i}{s-i}=\frac{n!}{i!(n-i)!}\times\frac{(n-i)!}{(s-i)!(n-s)!}=\frac{n!}{(n-s)!(s-i)!i!}=\frac{n!}{s!(n-s)!}\times\frac{s!}{i!(s-i)!}=\binom{n}{s}\binom{s}{i},$$

donc : $P(N=s)=\sum_{i=0}^{s}\binom{n}{s}\binom{s}{i}p^s(1-p)^{2n-s-i}=\binom{n}{s}p^s(1-p)^{2n-s}\sum_{i=0}^{s}\binom{s}{i}(1-p)^{-i}$

$$\overset{\text{Binôme}}{=}\binom{n}{s}p^s(1-p)^{2n-s}\left(1+\frac{1}{1-p}\right)^s=\binom{n}{s}p^s(2-p)^s(1-p)^{2n-2s}=\binom{n}{s}p^s(2-p)^s\big((1-p)^2\big)^{n-s}$$
$$=\binom{n}{s}\big(2p-p^2\big)^s\Big(1-\big(2p-p^2\big)\Big)^{n-s}.$$

Finalement : $N\hookrightarrow\mathscr{B}\big(n,2p-p^2\big)$.

4 Variables aléatoires indépendantes

Définition (Paire de variables aléatoires indépendantes) Soient X et Y deux variables aléatoires sur Ω. On dit que X et Y sont *indépendantes* si pour tous $x\in X(\Omega)$ et $y\in Y(\Omega)$, les événements $\{X=x\}$ et $\{Y=y\}$ sont indépendants, i.e. :

$$P(X=x \text{ et } Y=y)=P(X=x)\,P(Y=y).$$

✌ **Explication** ✌ En général, les lois marginales du couple (X,Y) ne déterminent pas sa loi conjointe, mais c'est le cas tout de même lorsque X et Y sont indépendantes.

Théorème (Une propriété des paires de variables aléatoires indépendantes) Soient X et Y deux variables aléatoires sur Ω. Si X et Y sont indépendantes, alors pour toutes parties A de $X(\Omega)$ et B de $Y(\Omega)$, les événements $\{X\in A\}$ et $\{Y\in B\}$ sont indépendants, i.e. :

$$P(X\in A \text{ et } Y\in B)=P(X\in A)\,P(Y\in B).$$

Démonstration Pour tous $A\in\mathscr{P}\big(X(\Omega)\big)$ et $B\in\mathscr{P}\big(Y(\Omega)\big)$.

$$P(X\in A)\,P(Y\in B)=P\left(\bigsqcup_{x\in A}\{X=x\}\right)P\left(\bigsqcup_{y\in B}\{Y=y\}\right)=\left(\sum_{x\in A}P(X=x)\right)\left(\sum_{y\in B}P(Y=y)\right)=\sum_{(x,y)\in A\times B}P(X=x)\,P(Y=y)$$
$$\overset{\text{Indépen-dance}}{=}\sum_{(x,y)\in A\times B}P(X=x \text{ et } Y=y)=P\left(\bigsqcup_{(x,y)\in A\times B}\{X=x \text{ et } Y=y\}\right)=P(X\in A \text{ et } Y\in B).$$

■

Définition-théorème (Ensemble fini de variables aléatoires mutuellement indépendantes) Soient $X_1,\dots,X_n$ des variables aléatoires sur Ω.

- On dit que $X_1,\dots,X_n$ sont (*mutuellement*) *indépendantes* si pour tous $x_1\in X_1(\Omega),\dots,x_n\in X_n(\Omega)$:

$$P(X_1=x_1 \text{ et } \dots \text{ et } X_n=x_n)=P(X_1=x_1)\ \dots\ P(X_n=x_n).$$

  On peut montrer que cette définition est équivalente à la suivante : pour tous $x_1\in X_1(\Omega),\dots,x_n\in X_n(\Omega)$, les événements $\{X=x_1\},\dots,\{X_n=x_n\}$ sont (mutuellement) indépendants.

- On peut montrer également que si $X_1,\dots,X_n$ sont (mutuellement) indépendantes, alors pour toutes parties A_1 de $X_1(\Omega)$… et A_n de $X_n(\Omega)$, les événements $\{X_1\in A_1\},\dots,\{X_n\in A_n\}$ sont (mutuellement) indépendants. En particulier :

$$P(X_1\in A_1 \text{ et } \dots \text{ et } X_n\in A_n)=P(X_1\in A_1)\ \dots\ P(X_n\in A_n).$$

✖ **Attention !** ✖ (Mutuellement) indépendantes $\Longrightarrow$ **Deux à deux** indépendantes mais **la réciproque est fausse !**

Exemple Soient $p_1,\ldots,p_n \in [0,1]$ et $X_1,\ldots,X_n$ des variables aléatoires indépendantes de lois respectives $\mathscr{B}(p_1),\ldots,\mathscr{B}(p_n)$. Alors : $X_1\ldots X_n \hookrightarrow \mathscr{B}\big(p_1\ldots p_n\big)$.

Démonstration La variable aléatoire $X_1\ldots X_n$ a pour image $\big\{0,1\big\}$ et :

$$P(X_1\ldots X_n=1)=P(X_1=1 \text{ et } \ldots \text{ et } X_n=1) \underset{\text{dance}}{\overset{\text{Indépen-}}{=}} P(X_1=1)\ \ldots\ P(X_n=1)=p_1\ldots p_n$$

et : $P(X_1\ldots X_n=0)=1-P(X_1\ldots X_n=1)=1-p_1\ldots p_n$. ■

Théorème (Sommes de variables aléatoires indépendantes de lois binomiales) Soient $p\in[0,1]$, $m,n\in\mathbb{N}^*$ et $X,Y,X_1,\ldots,X_n$ des variables aléatoires sur Ω.

(i) Si $X\hookrightarrow\mathscr{B}(m,p)$ et $Y\hookrightarrow\mathscr{B}(n,p)$ et si X et Y sont **indépendantes**, alors : $X+Y\hookrightarrow\mathscr{B}(m+n,p)$.

(ii) Si $X_i\hookrightarrow\mathscr{B}(p)$ pour tout $i\in[\![1,n]\!]$ et si $X_1,\ldots,X_n$ sont **indépendantes**, alors : $\displaystyle\sum_{i=1}^n X_i\hookrightarrow\mathscr{B}(n,p)$.

Démonstration Il suffit bien sûr d'établir l'assertion (i). La variable aléatoire $X+Y$ est à valeurs dans $[\![0,m+n]\!]$ et pour tout $k\in[\![0,m+n]\!]$: $\displaystyle P(X+Y=k)=\sum_{i=0}^k P(X=i \text{ et } Y=k-i) \underset{\text{dance}}{\overset{\text{Indépen-}}{=}} \sum_{i=0}^k P(X=i)\ P(Y=k-i)$

$$=\sum_{i=0}^k\binom{m}{i}p^i(1-p)^{m-i}\times\binom{n}{k-i}p^{k-i}(1-p)^{n-k+i}=p^k(1-p)^{m+n-k}\sum_{i=0}^k\binom{m}{i}\binom{n}{k-i}.$$

Or dans l'anneau des polynômes $\mathbb{R}[T]$: $(T+1)^m(T+1)^n=(T+1)^{m+n}$, donc après identification des coefficients de degré k : $\displaystyle\sum_{i=0}^k\binom{m}{i}\binom{n}{k-i}=\binom{m+n}{k}$ — généralisation de la formule de Vandermonde.

Conclusion : $\displaystyle P(X+Y=k)=\binom{m+n}{k}p^k(1-p)^{m+n-k}$, donc comme voulu : $X\hookrightarrow\mathscr{B}(m+n,p)$. ■

Exemple Soient $p\in[0,1]$ et $X_1,\ldots,X_n$ des variables aléatoires indépendantes sur Ω de même loi de Bernoulli $\mathscr{B}(p)$. Nous venons de voir que : $\displaystyle\sum_{i=1}^n X_i\hookrightarrow\mathscr{B}(n,p)$. En particulier : $\displaystyle \mathrm{E}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n\mathrm{E}(X_i)=\sum_{i=1}^n p=np$ — autre manière d'envisager le calcul de l'espérance de la loi binomiale !

Théorème (Indépendance des images de variables aléatoires indépendantes par des fonctions) Soient $X_1,\ldots,X_n$ des variables aléatoires sur Ω et $f:(X_1,\ldots,X_m)(\Omega)\longrightarrow F$ et $g:(X_{m+1},\ldots,X_n)(\Omega)\longrightarrow G$ deux fonctions. Si $X_1,\ldots,X_n$ sont indépendantes, alors $f(X_1,\ldots,X_m)$ et $g(X_{m+1},\ldots,X_n)$ le sont aussi.

Démonstration Nous nous contenterons d'établir un cas particulier de ce résultat. Soient X et Y deux variables aléatoires indépendantes sur Ω et $f:X(\Omega)\longrightarrow F$ et $g:Y(\Omega)\longrightarrow G$ deux fonctions. Montrons que $f(X)$ et $g(Y)$ sont également indépendantes. Or pour tous $a\in f(X)(\Omega)$ et $b\in g(Y)(\Omega)$:

$$\begin{aligned}P\big(f(X)=a \text{ et } g(Y)=b\big)&=P\Big(X\in f^{-1}\big(\{a\}\big) \text{ et } Y\in g^{-1}\big(\{b\}\big)\Big) \underset{\text{dance}}{\overset{\text{Indépen-}}{=}} P\Big(X\in f^{-1}\big(\{a\}\big)\Big)\ P\Big(Y\in g^{-1}\big(\{b\}\big)\Big)\\&=P\big(f(X)=a\big)\ P\big(g(Y)=b\big).\end{aligned}$$ ■

Exemple Soient X, Y et Z trois variables aléatoires indépendantes sur Ω. Alors $X+Y$ et Z sont indépendantes.

Théorème **(Espérance du produit de deux variables aléatoires réelles indépendantes)** Soient X et Y deux variables aléatoires réelles sur Ω. Si X et Y sont INDÉPENDANTES, alors : $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.

Ce résultat s'étend naturellement à un nombre fini quelconque de variables aléatoires indépendantes.

Démonstration $$\mathrm{E}(X)\,\mathrm{E}(Y) = \left(\sum_{x\in X(\Omega)} P(X=x)\,x\right)\left(\sum_{y\in Y(\Omega)} P(Y=y)\,y\right)$$
$$= \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} P(X=x)\,P(Y=y)\,xy = \sum_{(x,y)\in (X,Y)(\Omega)} P(X=x \text{ et } Y=y)\,xy = \mathrm{E}(XY).$$

La dernière égalité est un cas particulier de la formule de transfert — appliquée à la variable aléatoire (X,Y) et la fonction $(x,y) \longmapsto xy$. ■

✖ **Attention !** ✖ L'identité : $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$ est fausse en général. Par exemple, pour toute variable aléatoire X de loi uniforme sur $\{-1,1\}$: $\mathrm{E}(X) = \dfrac{1}{2}\times 1 + \dfrac{1}{2}\times(-1) = 0$ donc : $\mathrm{E}(X)^2 = 0$, mais : $\mathrm{E}\big(X^2\big) = \mathrm{E}(1) = 1$.

5 Variance, écart-type, covariance

Définition **(Moments, variance et écart-type d'une variable aléatoire réelle)** Soit X une variable aléatoire réelle.

- Pour tout $k \in \mathbb{N}$, le réel $\mathrm{E}\big(X^k\big)$ est appelé *moment d'ordre k de X* et le réel $\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^k\Big)$ son *moment d'ordre k centré*.
- On appelle *variance de X* le réel positif : $\mathrm{V}(X) = \mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)$ et *écart-type de X* le réel : $\sigma(X) = \sqrt{\mathrm{V}(X)}$.
- On dit que X est *réduite* si : $\mathrm{V}(X) = 1$.

✌ **Explication** ✌

- Comme on l'a vu, l'espérance de X est un indicateur de position, mais les valeurs de X sont-elles plutôt proches de cette valeur moyenne ou plutôt éloignées ? Toute mesure de cette proximité à la moyenne est appelée un *indicateur de dispersion*. À proprement parler, l'écart naturel entre X et son espérance est $\big|X-\mathrm{E}(X)\big|$, donc l'écart moyen de X à sa moyenne est $\mathrm{E}\Big(\big|X-\mathrm{E}(X)\big|\Big)$, mais ce n'est pas là l'indicateur de dispersion que les mathématiciens ont choisi d'utiliser couramment, ils ont préféré s'intéresser à la variance, c'est-à-dire l'écart QUADRATIQUE moyen à la moyenne — « quadratique » parce qu'on passe au carré. Pourquoi ce choix moins naturel au premier abord ? Parce que la variance est plus facile à manipuler comme on va bientôt pouvoir le constater.
- Et l'écart-type, quelle différence avec la variance ? Si par exemple X représente une longueur, $\mathrm{V}(X)$ représente une longueur AU CARRÉ, donc il n'est pas possible de comparer directement la position moyenne $\mathrm{E}(X)$ et sa dispersion moyenne $\mathrm{V}(X)$ — un physicien dirait que l'espérance et la variance ne sont pas « homogènes ». L'écart-type, au contraire, est homogène à une longueur, donc comparable à l'espérance — d'où son intérêt.

Théorème **(Propriétés de la variance)** Soit X une variable aléatoire réelle.

(i) **Formule de Huygens :** $\mathrm{V}(X) = \mathrm{E}\big(X^2\big) - \mathrm{E}(X)^2$.

(ii) **Nullité :** $\mathrm{V}(X) = 0$ si et seulement si : $P\big(X = \mathrm{E}(X)\big) = 1$, i.e. si et seulement si X est « presque sûrement » constante.

(iii) **Effet d'une transformation affine :** Pour tous $a, b \in \mathbb{R}$: $\mathrm{V}(aX+b) = a^2\mathrm{V}(X)$.

En particulier, si : $\sigma(X) > 0$, la variable aléatoire $\dfrac{X-\mathrm{E}(X)}{\sigma(X)}$ est centrée réduite.

Démonstration

(i) $\mathrm{V}(X) = \mathrm{E}\Big(X^2 - 2\overbrace{\mathrm{E}(X)}^{\text{constante}}\,X + \overbrace{\mathrm{E}(X)^2}^{\text{constante}}\Big) = \mathrm{E}\big(X^2\big) - 2\mathrm{E}(X)\,\mathrm{E}(X) + \mathrm{E}(X)^2 = \mathrm{E}\big(X^2\big) - \mathrm{E}(X)^2.$

(ii) D'après la formule de transfert : $\mathrm{V}(X) = \sum_{x\in X(\Omega)} \underbrace{P(X=x)\,\big(x-\mathrm{E}(X)\big)^2}_{\geqslant 0}$, donc :

$$\mathrm{V}(X)=0 \iff \forall x \in X(\Omega),\quad P(X=x)=0 \quad\text{ou}\quad x=\mathrm{E}(X) \iff \forall x\in X(\Omega)\setminus\big\{\mathrm{E}(X)\big\},\quad P(X=x)=0$$
$$\iff P\big(X=\mathrm{E}(X)\big)=1 \qquad \text{car} \quad \sum_{x\in X(\Omega)} P(X=x)=1.$$

(iii) Comme $\mathrm{E}(aX+b)=a\mathrm{E}(X)+b$: $\mathrm{V}(aX+b)=\mathrm{E}\Big(\big(aX+b-a\mathrm{E}(X)-b\big)^2\Big)=a^2\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)$.

Enfin, si $\sigma(X)>0$: $\mathrm{E}\left(\dfrac{X-\mathrm{E}(X)}{\sigma(X)}\right)=\dfrac{\mathrm{E}(X)-\mathrm{E}(X)}{\sigma(X)}=0$ et $\mathrm{V}\left(\dfrac{X-\mathrm{E}(X)}{\sigma(X)}\right)=\dfrac{\mathrm{V}(X)}{\sigma(X)^2}=1.$ ■

Définition-théorème **(Covariance de deux variables aléatoires réelles)** Soient $X, Y, X_1, \ldots, X_n$ des variables aléatoires réelles sur Ω.

(i) **Définition :** On appelle *covariance de X et Y* le réel : $\mathrm{cov}(X,Y)=\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\Big)$.

Clairement : $\mathrm{V}(X)=\mathrm{cov}(X,X)$ et $\mathrm{cov}(Y,X)=\mathrm{cov}(X,Y)$.

(ii) $\mathrm{cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$.

(iii) $\mathrm{V}(X+Y)=\mathrm{V}(X)+2\,\mathrm{cov}(X,Y)+\mathrm{V}(Y)$.

Plus généralement : $\mathrm{V}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{V}(X_i)+2\sum_{1\leqslant i<j\leqslant n}\mathrm{cov}(X_i,X_j)$.

(iv) Si X et Y sont indépendantes : $\mathrm{cov}(X,Y)=0$ et $\mathrm{V}(X+Y)=\mathrm{V}(X)+\mathrm{V}(Y)$.

Plus généralement, si $X_1,\ldots,X_n$ sont (seulement) DEUX À DEUX indépendantes : $\mathrm{V}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{V}(X_i)$.

✖ **ATTENTION !** ✖ La réciproque de (iv) est fausse, la covariance de X et Y peut être nulle sans que X et Y soient indépendantes — cf. contre-exemple après la preuve de la relation : $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$.

Démonstration

(ii) $\mathrm{cov}(X,Y)=\mathrm{E}\Big(XY-X\overbrace{\mathrm{E}(Y)}^{\text{constante}}-\overbrace{\mathrm{E}(X)}^{\text{constante}}Y+\overbrace{\mathrm{E}(X)\,\mathrm{E}(Y)}^{\text{constante}}\Big)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)-\mathrm{E}(X)\,\mathrm{E}(Y)+\mathrm{E}(X)\,\mathrm{E}(Y)$
$=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$.

(iii) $\mathrm{V}(X+Y)=\mathrm{E}\Big(\big(X-\mathrm{E}(X)+Y-\mathrm{E}(Y)\big)^2\Big)$
$=\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2+2\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)+\big(Y-\mathrm{E}(Y)\big)^2\Big)=\mathrm{V}(X)+2\,\mathrm{cov}(X,Y)+\mathrm{V}(Y)$.

(iv) Si X et Y sont indépendantes : $\mathrm{cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)=0$. On conclut grâce à (iii). ■

Théorème **(Variance des lois usuelles)** Soient X une variable aléatoire réelle et $p\in[0,1]$.

(i) Si $X\hookrightarrow\mathscr{B}(p)$: $\mathrm{V}(X)=p(1-p)$. (ii) Si $X\hookrightarrow\mathscr{B}(n,p)$: $\mathrm{V}(X)=np(1-p)$.

Démonstration

(i) $\mathrm{V}(X)=\mathrm{E}\big(X^2\big)-\mathrm{E}(X)^2=\big(P(X=1)\times 1^2+P(X=0)\times 0^2\big)-p^2=p-p^2=p(1-p)$.

(ii) Soient $X_1,\ldots,X_n$ des variables aléatoires indépendantes sur Ω de même loi de Bernoulli $\mathscr{B}(p)$. Comme : $\sum_{i=1}^n X_i\hookrightarrow\mathscr{B}(n,p)$, alors par indépendance : $\mathrm{V}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n\mathrm{V}(X_i)=\sum_{i=1}^n p(1-p)=np(1-p)$. Ce calcul ne concerne a priori pas la variable aléatoire X de l'énoncé, mais l'espérance d'une variable aléatoire ne dépend que de sa loi, donc le calcul que nous venons de faire dans le cas particulier de la somme $\sum_{i=1}^n X_i$ est en fait emblématique. ■

6 Un premier pas vers les grands nombres

Théorème **(Inégalité de Markov)** Soit X une variable aléatoire réelle. Pour tout $a > 0$: $P\big(|X| \geqslant a\big) \leqslant \dfrac{\mathrm{E}\big(|X|\big)}{a}$.

✌ **Explication** ✌ Comme : $\big\{|X| > a\big\} \subset \big\{|X| \geqslant a\big\}$, on peut rendre l'inégalité stricte : $P\big(|X| > a\big) \leqslant \dfrac{\mathrm{E}\big(|X|\big)}{a}$.

Démonstration Tout repose sur les inégalités suivantes : $\mathbb{1}_{\{|X|\geqslant a\}} a \leqslant \mathbb{1}_{\{|X|\geqslant a\}} |X| \leqslant |X|$. Aussitôt, par croissance de l'espérance : $aP\big(|X| \geqslant a\big) = \mathrm{E}\big(\mathbb{1}_{\{|X|\geqslant a\}} a\big) \leqslant \mathrm{E}\big(|X|\big)$. ■

Théorème **(Inégalité de Bienaymé-Tchebychev)** Soit X une variable aléatoire réelle. Pour tout $a > 0$:

$$P\Big(\big|X - \mathrm{E}(X)\big| \geqslant a\Big) \leqslant \frac{\mathrm{V}(X)}{a^2}.$$

Démonstration D'après l'inégalité de Markov : $P\Big(\big(X - \mathrm{E}(X)\big)^2 \geqslant a^2\Big) \leqslant \dfrac{\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)}{a^2} = \dfrac{\mathrm{V}(X)}{a^2}$. Il reste à remarquer que les événements $\Big\{\big(X - \mathrm{E}(X)\big)^2 \geqslant a^2\Big\}$ et $\Big\{\big|X - \mathrm{E}(X)\big| \geqslant a\Big\}$ coïncident. ■

Exemple On dispose d'une pièce éventuellement truquée pour laquelle la probabilité d'apparition de PILE est notée p. Pour connaître p, on lance cette pièce n fois et on note F la fréquence d'apparition de PILE ainsi obtenue. À partir de quelle valeur de n la probabilité pour que F soit une approximation de p à 10^{-2} près est-elle supérieure à 0,9 ?

Démonstration Nous noterons N le nombre de PILE obtenus — de sorte que : $F = \dfrac{N}{n}$. Par indépendance des lancers : $N \hookrightarrow \mathscr{B}(n,p)$, donc : $\mathrm{E}(N) = np$, puis : $\mathrm{E}(F) = p$. Ce résultat justifie à lui seul qu'on choisisse F pour estimer p. D'après l'inégalité de Bienaymé-Tchebychev :

$$P\big(|F-p| \geqslant 10^{-2}\big) = P\Big(\big|N - \mathrm{E}(N)\big| \geqslant 10^{-2}n\Big) \leqslant \frac{\mathrm{V}(N)}{\big(10^{-2}n\big)^2} = \frac{np(1-p)}{10^{-4}n^2} = \frac{10^4}{n} \times p(1-p) \leqslant \frac{10^4}{n} \times \frac{1}{4} = \frac{2500}{n}.$$

Or : $P\big(|F-p| < 10^{-2}\big) \geqslant 0,9 \iff P\big(|F-p| \geqslant 10^{-2}\big) \leqslant 0,1$, donc il nous suffit de choisir n avec la condition : $\dfrac{2500}{n} \leqslant 0,1$ pour garantir que F soit une approximation de p à 10^{-2} près. Conclusion : il faut quand même lancer 25 000 fois la pièce ! — ou trouver une majoration meilleure de $P\big(|F-p| \geqslant 10^{-2}\big)$.

✌ **Explication** ✌ Le paragraphe qui suit développe une idée fondamentale et constitue l'aboutissement de nos aventures probabilistes en MPSI. On s'intéresse à une grandeur d'espérance m et d'écart-type σ. On la mesure dans un premier temps à l'occasion d'une unique expérience de résultat noté X_1. D'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$: $P\big(|X_1 - m| \geqslant \varepsilon\big) \leqslant \dfrac{\sigma^2}{\varepsilon^2}$. La mesure X_1 est donc à distance supérieure ou égale à ε de m avec probabilité au plus $\dfrac{\sigma^2}{\varepsilon^2}$.
Si l'on souhaite estimer m, l'expérience nous enseigne qu'une seule mesure a peu de chances de nous donner un résultat satisfaisant, nous obtiendrons une meilleure estimation de m en augmentant le nombre de mesures. La théorie des probabilités confirme-t-elle cette intuition ? Quel résultat quantitatif pouvons-nous en tirer ? Nous effectuons à présent n mesures $X_1, \ldots, X_n$ de la grandeur étudiée. Formellement, $X_1, \ldots, X_n$ sont des variables aléatoires réelles INDÉPENDANTES et DE MÊME LOI, d'espérance m et d'écart-type σ. Nous noterons $\overline{X}_n$ la moyenne de ces n mesures : $\overline{X}_n = \dfrac{1}{n}\sum_{k=1}^n X_k$. Par linéarité de l'espérance : $\mathrm{E}\big(\overline{X}_n\big) = \dfrac{1}{n}\sum_{k=1}^n \mathrm{E}(X_k) = \dfrac{1}{n}\sum_{k=1}^n m = m$ et par indépendance : $\mathrm{V}\big(\overline{X}_n\big) = \dfrac{1}{n^2}\sum_{k=1}^n \mathrm{V}(X_k) = \dfrac{1}{n^2}\sum_{k=1}^n \sigma^2 = \dfrac{\sigma^2}{n}$, donc : $\sigma\big(\overline{X}_n\big) = \dfrac{\sigma}{\sqrt{n}}$. En résumé :

La moyenne de n variables aléatoires **INDÉPENDANTES** et **DE MÊME LOI**, d'espérance m et d'écart-type σ, admet toujours m pour espérance, **MAIS** $\dfrac{\sigma}{\sqrt{n}}$ pour écart-type.

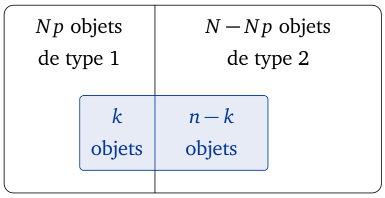
Avec n mesures, nous avons ainsi gagné un facteur $\dfrac{1}{\sqrt{n}}$ en écart-type — et $\dfrac{1}{n}$ en variance. Intuitivement, cela veut dire que chaque nouvelle mesure nous rapproche de la certitude que la moyenne empirique $\overline{X}_n$ est proche de l'espérance m. Formellement, l'inégalité de Bienaymé-Tchebychev énonce cette fois que pour tout $\varepsilon > 0$: $\boxed{P\left(\left|\overline{X}_n - m\right| \geqslant \varepsilon\right) \leqslant \dfrac{\sigma^2}{n\varepsilon^2}.}$

En particulier : $\lim\limits_{n\to+\infty} P\left(\left|\overline{X}_n - m\right| \geqslant \varepsilon\right) = 0,$ ce que conforte notre intuition. Toute limite de ce genre est appelée une *loi des grands nombres*. Il en existe de nombreuses, dont l'inégalité ci-dessus — la plus faible de toutes, mais c'est déjà ça ! Les lois des grands nombres sont fondamentales car elles justifient l'interprétation *fréquentiste* que nous avons des probabilités. Qu'avons-nous en tête quand nous disons que la probabilité d'apparition de chaque face d'un dé équilibré à 6 faces vaut $\dfrac{1}{6}$? Nous signifions par là une vérité sur le monde qui n'a rien à voir avec le fait que de tels dés existent et qu'on peut les lancer un nombre indéfini de fois. Cette probabilité vaudrait $\dfrac{1}{6}$ même si l'humanité s'interdisait rigoureusement tout lancer de dé. Cela dit, le fait est que nous avons une autre notion de probabilité en tête simultanément, nous avons aussi l'habitude de constater que lorsqu'on lance un dé de nombreuses fois, la fréquence d'apparition de chaque face tend vers $\dfrac{1}{6}$ à mesure que ce nombre augmente. Les lois des grands nombres réconcilient les deux points de vue, la fréquence d'un événement lorsqu'on répète indéfiniment une expérience aléatoire tend vers la probabilité a priori de cet événement.

7 Un complément sur la loi hypergéométrique

La *loi hypergéométrique* est une loi usuelle hors programme, mais d'utilisation courante donc à étudier sérieusement.

Avec la loi hypergéométrique, on s'intéresse au tirage simultané de n objets dans un ensemble de N objets qui sont de deux types, disons de type 1 en proportion p et de type 2 en proportion $1-p$. Parmi N objets, Np sont donc de type 1 et les $N-Np$ restants sont de type 2. Quelle est la loi du nombre X des objets de type 1 parmi les n tirés ?



On choisit pour univers Ω de cette expérience aléatoire l'ensemble $\mathscr{P}_n\big([\![1,N]\!]\big)$ des n-combinaisons de $[\![1,N]\!]$ — les objets de type 1 sont numérotés de 1 à Np et les objets de type 2 de $Np+1$ à N — et pour probabilité P sur Ω la probabilité uniforme.

À présent, quelle est donc l'image de X ? Lorsqu'on tire k objets de type 1 et $n-k$ de type 2, deux inégalités sont nécessairement vraies : $0 \leqslant k \leqslant Np$ et $0 \leqslant n-k \leqslant N-Np$, équivalentes à :

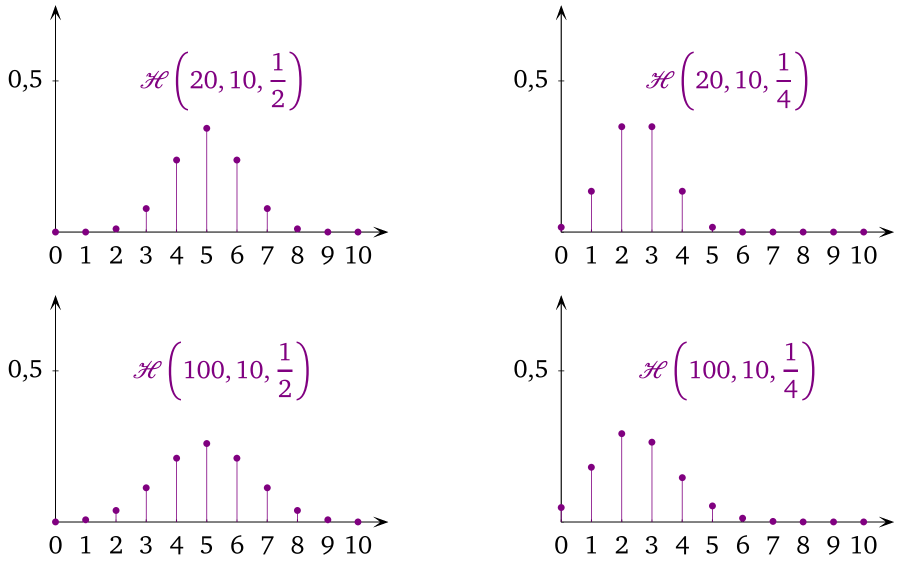
$$0 \leqslant k \leqslant Np \quad \text{et} \quad n-N+Np \leqslant k \leqslant n, \qquad \text{ou encore :} \quad k \in \Big[\!\!\Big[\max\big\{0, n-N+Np\big\}, \min\big\{Np, n\big\}\Big]\!\!\Big].$$

Pour un tel k, l'événement $\big\{X=k\big\}$ est réalisé quand on tire k objets de type 1 parmi les Np possibles ($\dbinom{Np}{k}$ possibilités) et $n-k$ de type 2 parmi les $N-Np$ restantes ($\dbinom{N-Np}{n-k}$ possibilités), donc : $P(X=k) = \dfrac{\big|\{X=k\}\big|}{|\Omega|} = \dfrac{\dbinom{Np}{k}\dbinom{N-Np}{n-k}}{\dbinom{N}{n}}.$

Définition **(Loi hypergéométrique ou loi des tirages SANS remise)** Soient X une variable aléatoire et $N, n \in \mathbb{N}^*$ et $p \in [0,1]$ tels que $Np \in \mathbb{N}$.
On dit que X *suit la loi hypergéométrique de paramètre* (N,n,p), ce qu'on note : $X \hookrightarrow \mathscr{H}(N,n,p)$, si pour tout $k \in \mathbb{N}$:

$$P(X=k) = \frac{\dbinom{Np}{k}\dbinom{N-Np}{n-k}}{\dbinom{N}{n}}.$$

En réalité : $P(X=k) \neq 0$ si et seulement si k appartient à $\left[\!\left[\max\left\{0, n-N+Np\right\}, \min\left\{Np, n\right\} \right]\!\right]$.



Explication

Lorsqu'on tire n objets simultanément dans un ensemble de N objets dont Np sont de type 1 et $N-Np$ de type 2, le **NOMBRE D'OBJETS DE TYPE 1 TIRÉS** suit la loi hypergéométrique $\mathscr{H}(N,n,p)$.

- En pratique, la distinction de deux types distincts dans un même ensemble peut recouvrir des oppositions très variées telles que sain/malade, défectueux/non défectueux, fille/garçon, blanc/rouge, personnes ayant voté « oui »/personnes ayant voté « non », etc.
- Tout ceci définit bien une loi car : $\displaystyle\sum_{k=0}^{+\infty} \frac{\dbinom{Np}{k}\dbinom{N-Np}{n-k}}{\dbinom{N}{n}} = 1$ — somme presque nulle. En effet, dans l'anneau de polynômes $\mathbb{R}[T]$: $(T+1)^{Np}(T+1)^{N-Np} = (T+1)^N$, ensuite il suffit ensuite d'identifier les coefficients de degré n dans cette identité. Encore une généralisation de la formule de Vandermonde !
- L'expression « loi des tirages SANS remise » par laquelle on décrit parfois la loi hypergéométrique se comprend facilement, le tirage de n objets simultanément est parfaitement équivalent au tirage de ces n objets sans remise.

Théorème **(Espérance et variance de la loi hypergéométrique)** Soient $N, n \in \mathbb{N}^*$ et $p \in [0,1]$ tels que $Np \in \mathbb{N}$ et X une variable aléatoire réelle de loi $\mathscr{H}(N,n,p)$. Alors : $\mathrm{E}(X) = np$ et $\mathrm{V}(X) = np(1-p)\,\dfrac{N-n}{N-1}$.

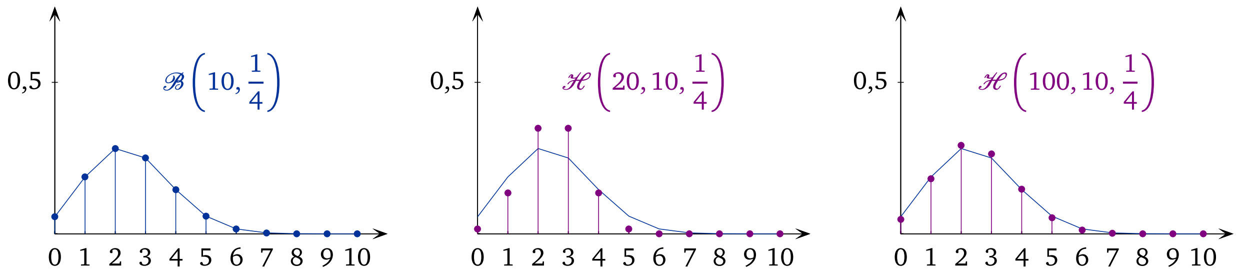
Démonstration Dans le calcul qui suit, toutes les sommes sont presque nulles.

$$\mathrm{E}(X) = \sum_{k=0}^{+\infty} P(X=k)\,k = \sum_{k=0}^{+\infty} \frac{\dbinom{Np}{k}\dbinom{N-Np}{n-k}}{\dbinom{N}{n}}\,k = \sum_{k=1}^{+\infty} \frac{\dfrac{Np}{k}\dbinom{Np-1}{k-1}\dbinom{N-Np}{n-k}}{\dbinom{N}{n}}\,k$$

$$\overset{i=k-1}{=} \frac{Np}{\dbinom{N}{n}} \sum_{i=0}^{+\infty} \binom{Np-1}{i}\binom{N-Np}{n-1-i} = \frac{Np}{\dbinom{N}{n}}\binom{N-1}{n-1} = Np \times \frac{n}{N} = np.$$

Nous admettrons le résultat pour la variance. ∎

Exemple **(Approximation de la loi hypergéométrique par la loi binomiale)** Soit X une variable aléatoire de loi hypergéométrique $\mathscr{H}(N,n,p)$ avec $N,n \in \mathbb{N}^*$ et $p \in]0,1[$. Plus N est grand, plus la loi de X est « proche » — en un sens que je laisse volontairement flou — de la loi binomiale $\mathscr{B}(n,p)$. Les figures ci-dessous sont à ce titre assez convaincantes.



Démonstration Tout d'abord, pour tout $k \in [\![0,n]\!]$: $\displaystyle\binom{Np}{k} = \frac{(Np)(Np-1)\ldots(Np-k+1)}{k!} \underset{N\to+\infty}{\sim} \frac{(Np)^k}{k!}$,

et de même : $\displaystyle\binom{N-Np}{n-k} \underset{N\to+\infty}{\sim} \frac{(N-Np)^{n-k}}{(n-k)!} \underset{N\to+\infty}{\sim} \frac{\big(N(1-p)\big)^{n-k}}{(n-k)!}$ et $\displaystyle\binom{N}{n} \underset{N\to+\infty}{\sim} \frac{N^n}{n!}$.

En retour, par quotient : $\displaystyle\frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}} \underset{N\to+\infty}{\sim} \binom{n}{k}p^k(1-p)^{n-k} \underset{N\to+\infty}{\longrightarrow} \binom{n}{k}p^k(1-p)^{n-k}$.

Comme annoncé, $P(X=k)$ tend donc vers $\binom{n}{k}p^k(1-p)^{n-k}$ pour tout $k \in [\![0,n]\!]$ lorsque N tend vers $+\infty$, ce qui signifie que la loi hypergéométrique $\mathscr{H}(N,n,p)$ « tend » vers la loi binomiale $\mathscr{B}(n,p)$. On parle ici d'une *approximation de la loi hypergéométrique par la loi binomiale* et non le contraire, car la loi hypergéométrique est plus « compliquée » que la loi binomiale.

Quelle interprétation ? La loi hypergéométrique $\mathscr{H}(N,n,p)$ est la loi du nombre d'éléments de type 1 qu'on obtient en choisissant successivement **SANS** remise n élements dans un ensemble de cardinal N dont la proportion des éléments de type 1 est p. Or quand N est grand devant n, le fait que les tirages soient effectués **SANS** remise compte assez peu car on a alors peu de chances de tirer plusieurs fois le même élément, c'est presque comme si les tirages étaient effectués **AVEC** remise selon la loi binomiale $\mathscr{B}(n,p)$.

Exemple Un institut de sondage cherche à anticiper le résultat d'un référendum en interrogeant un échantillon de n personnes. Au moment du sondage, la proportion des « oui » dans la population complète est p — valeur inconnue que l'institut essaie justement d'estimer. Si on note $\overline{p}$ la proportion des « oui » dans l'échantillon des n personnes sondées, comment faut-il choisir n pour que la probabilité de l'événement $\big\{|\overline{p}-p| \geqslant 0,05\big\}$ soit inférieure ou égale à 0,1 ?

Démonstration Notons N le nombre total d'individus dans la population étudiée. À proprement parler, le nombre $n\overline{p}$ des « oui » collectés au cours du sondage suit la loi hypergéométrique $\mathscr{H}(N,n,p)$ ou loi des tirages **SANS** remise, mais quand on fait un sondage, l'échantillon n est très petit devant la population totale N, si petit qu'en réalité la distinction **AVEC** ou **SANS** remise perd tout intérêt. Profitons donc de l'approximation au forceps de l'exemple précédent et faisons comme si la loi du nombre $n\overline{p}$ des « oui » était binomiale $\mathscr{B}(n,p)$.

Intérêt de la manœuvre : la loi binomiale $\mathscr{B}(n,p)$ est plus simple d'utilisation, nous connaissons son espérance np et sa variance $np(1-p)$. L'inégalité de Bienaymé-Tchebychev donne aussitôt ceci :

$$P\big(|\overline{p}-p| \geqslant 0,05\big) = P\Big(\big|n\overline{p} - \mathrm{E}(n\overline{p})\big| \geqslant 0,05n\Big) \leqslant \frac{\mathrm{V}(n\overline{p})}{(0,05n)^2} = \frac{400p(1-p)}{n} \leqslant \frac{100}{n},$$

car pour tout $x \in \mathbb{R}$: $x(1-x) \leqslant \dfrac{1}{4}$. À quelle condition a-t-on donc : $P\big(|\overline{p}-p| \geqslant 0,05\big) \leqslant 0,1$? Condition suffisante : $\dfrac{100}{n} \leqslant 0,1$, i.e. : $n \geqslant 1000$.