Probabilités discrètes

Bruno Grenet

1 Vocabulaire et définitions de base

Vocabulaire. Un *espace probabilisé* est une modélisation d'une expérience probabiliste : il est constitué d'un ensemble d'*évènements primitifs*, appelé l'*univers* et noté Ω, que l'on suppose fini ou dénombrable, et de leurs *probabilités* associées $p_\omega \in [0,1]$, telles que la somme des probabilités des évènements primitifs vaut 1 : $\sum_{\omega\in\Omega} p_\omega = 1$. Un *évènement* est tout sous-ensemble $E \subset \Omega$ de l'univers.

Une *variable aléatoire (discrète)* est une fonction $X : \Omega \to V$ de l'ensemble des évènements primitifs dans un ensemble de valeurs V. Si l'ensemble de valeurs est un sous-ensemble de $\mathbb{R}$, on parle de *variable aléatoire réelle*.

Remarque. Dans ce cours, toutes les variables aléatoires rencontrées sont réelles et discrètes.

Définition 1.1. Si $X : \Omega \to V$ est une variable aléatoire à valeur dans V, on définit pour tout $v \in V$ l'évènement « $X = v$ » comme l'ensemble $\{\omega \in \Omega : X(\omega) = v\}$, c'est-à-dire qu'il s'agit de l'ensemble des évènements primitifs de valeur v. On définit de manière analogue les évènements « $X \neq v$ », « $X \geq v$ », « $X > v$ », etc.

Soit Ω un univers et pour tout évènement primitif $\omega \in \Omega$, soit $p_\omega \in [0,1]$ sa probabilité associée. On définit la fonction $\Pr : \mathscr{P}(\Omega) \to [0,1]$ par $\Pr[E] = \sum_{\omega\in E} p_\omega$ pour tout évènement E. En particulier,

$$\Pr[X = v] = \sum_{\omega\in\Omega : X(\omega)=v} p_\omega$$

et de même pour $\Pr[X \geq v]$, etc.

Remarque. L'application de la fonction $\Pr : \Omega \to V$ à un évènement est notée avec des crochets plutôt que des parenthèses.

Exemple. On considère le lancer d'un dé à 6 faces. Les évènements primitifs sont « obtenir un 1 », « obtenir un 2 », …, « obtenir un 6 ». On note cet ensemble $\Omega = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$. Si le dé est non biaisé, chaque évènement a probabilité $1/6$. On peut alors définir différentes variables aléatoires sur cet espace probabilisé.

- La variable aléatoire qui décrit le *nombre de points* obtenus avec le dé est la fonction $X : \Omega \to V = \{1,2,3,4,5,6\}$ définie par $X :$ ⚀ $\mapsto 1, \ldots,$ ⚅ $\mapsto 6$. Ainsi, pour tout $v \in V$, $\Pr[X = v] = 1/6$ et $\Pr[X \geq 3] = 2/3$ par exemple.
- La variable aléatoire qui détermine la parité du nombre de point obtenu est la fonction $Y : \Omega \to \{0,1\}$ définie par $Y :$ ⚀ $\mapsto 0$, $2 \mapsto 1, \ldots,$ ⚅ $\mapsto 1$ (1 représente « vrai » et 0 « faux »), et $\Pr[Y = 1] = 1/2$.

Remarque. Dans le cas de la variable aléatoire X, on peut assimiler un évènement primitif et sa valeur (par ex. ⚂ et sa valeur 3), et la probabilité associée à un évènement ($p_{⚂} = 1/6$) avec la probabilité de la valeur de cet évènement ($\Pr[X = 3] = 1/6$). L'exemple de la variable aléatoire Y montre qu'on ne peut pas toujours faire cette assimilation, donc prudence !

Notations. Si E et F sont deux évènements, on note $E \wedge F$ leur intersection, c'est-à-dire l'ensemble $\{\omega \in \Omega : \omega \in E \wedge \omega \in F\}$ et $E \vee F$ leur union. On note de plus $\neg E$ le complémentaire de l'ensemble E, c'est-à-dire l'ensemble $\{\omega \in \Omega : \omega \notin E\}$.

Ces notations dans le langage de la logique correspondent aux notions intuitives : $\Pr[E \wedge F]$ est la probabilité que l'évènement E *et* l'évènement F se produisent ; $\Pr[E \vee F]$ est la probabilité que l'évènement E *ou* l'évènement F se produise ; $\Pr[\neg E]$ est la probabilité que l'évènement E *ne* se produise *pas*.

2 Espérance, variance, écart-type

Définition 2.1. L'*espérance* d'une variable aléatoire réelle discrète $X : \Omega \to V$ est définie par

$$\mathbb{E}[X] = \sum_{v \in V} v \times \Pr[X = v] = \sum_{\omega \in \Omega} X(\omega) p_\omega.$$

Intuitivement, il s'agit de la moyenne pondérée des valeurs de tous les évènements possibles.

Remarque. Dans la définition précédente, si V est infini, il se peut que la somme soit elle-même infinie et l'espérance non définie. Pour que l'espérance soit définie, il faut en fait que la somme soit *absolument convergente*. Ce sera toujours le cas dans les exemples considérés, et on pourra donc ignorer ce point technique.

Exemple. Pour les variables aléatoires X et Y définies précédemment, $\mathbb{E}[X] = \sum_{v=1}^{6} v \times 1/6 = 7/2$ et $\mathbb{E}[Y] = 1 \times \Pr[Y = 1] + 0 \times \Pr[Y = 0] = 1/2$. On utilise régulièrement la remarque suivante : si une variable aléatoire Y est à valeurs dans $\{0, 1\}$, $\mathbb{E}[Y] = \Pr[Y = 1]$.

Définition 2.2. La *variance* d'une variable aléatoire réelle discrète X est

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2.$$

Son *écart-type* $\sigma(X)$ est la racine carrée de la variance. Intuitivement, l'écart-type mesure la *moyenne (quadratique) des écarts à la moyenne*.

Exemple. La variance de la variable aléatoire X est $\mathrm{Var}(X) = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2$. On calcule $\mathbb{E}\big[X^2\big] = \sum_{v=1}^{6} v^2 \Pr[X = v] = 1/6 \times 91$. Donc $\mathrm{Var}(X) = 91/6 - (7/2)^2 = 35/12$ et $\sigma(X) = \sqrt{35/12} \simeq 1,71$. La variance de la variable Y est $\mathrm{Var}(Y) = \mathbb{E}\big[Y^2\big] - \mathbb{E}[Y]^2 = 1/2 - 1/4 = 1/4$ et son écart-type est $\sigma(Y) = 1/2$.

3 Probabilités conditionnelles et indépendance

Définition 3.1. La probabilité d'un évènement *E conditionnée à un évènement F*, ou *probabilité de E sachant F*, est

$$\Pr[E|F] = \frac{\Pr[E \wedge F]}{\Pr[F]}.$$

L'espérance d'une variable aléatoire $X : \Omega \to V$ *conditionnée à un évènement* F, ou *espérance de* X *sachant* F, est

$$\mathbb{E}[X|F] = \sum_{v \in V} v \times \Pr[X = v|F].$$

Remarque. Les probabilité et espérance conditionnelles ne sont définies que pour un évènement F dont la probabilité est non nulle. Intuitivement, elles consistent à *changer d'univers*, c'est-à-dire à considérer F comme le nouvel univers.

Remarque. Si F est l'évènement Ω, on trouve $\Pr[E|\Omega] = \Pr[E]$ et $\mathbb{E}[X|\Omega] = \mathbb{E}[X]$.

Exemple. Si F est l'évènement « $Y = 1$ » (le dé est pair), alors $\Pr[X = 1|Y = 1] = 0$ et $\Pr[X = 2|Y = 1] = 1/3$. De plus, $\mathbb{E}[X|Y = 1] = 4$.

La *formule de Bayes* permet d'inverser les probabilités conditionnelles.

Théorème 3.2. *Si* E *et* F *sont deux évènements de probabilités non nulles,*

$$\Pr[E|F] = \frac{\Pr[F|E]\Pr[E]}{\Pr[F]}.$$

Démonstration. On applique simplement la définition : $\Pr[E|F] = \Pr[E \wedge F]/\Pr[F]$. Puisque $E \wedge F = F \wedge E$, $\Pr[E \wedge F] = \Pr[F|E]\Pr[E]$. □

Deux évènements sont *indépendants* s'ils n'influent pas l'un sur l'autre. En particulier, il le sont si $\Pr[E|F] = \Pr[E]$ (le fait que l'évènement F se produise n'a pas d'impact sur la probabilité de E de se produire). La définition formelle suit.

Définition 3.3. Deux évènements E et F sont dits indépendants si $\Pr[E \wedge F] = \Pr[E]\Pr[F]$.

Deux variables aléatoires $X : \Omega \to V$ et $Y : \Omega \to W$ sont dites indépendantes si pour tout $v \in V$ et $w \in W$, $\Pr[X = v \wedge Y = w] = \Pr[X = v]\Pr[Y = w]$.

En appliquant la définition de probabilité conditionnelle, on retrouve la caractérisation intuitive d'indépendance.

Proposition 3.4. *Si* E *et* F *sont deux évènements de probabilités différentes de* 0 *et* 1*,* E *et* F *sont indépendants si et seulement si* $\Pr[E|F] = \Pr[E]$.

Exemple. Soit E l'évènement « $X \geq 3$ » et F l'évènement « $Y = 1$ ». Alors $\Pr[E] = 2/3$, $\Pr[F] = 1/2$ et $\Pr[E \wedge F] = 1/3$ car $X \geq 3$ et $Y = 1$ si le dé tombe sur ⚃ ou ⚅.Ces deux évènements sont donc indépendants. À l'inverse, les évènements « $X = 3$ » et « $Y = 1$ » ne sont pas indépendants puisque la probabilité de l'intersection est nulle.

Remarque. On utilise souvent la définition *dans l'autre sens*. Par exemple, si on considère deux lancers de dés qui n'ont rien à voir, on peut dire que ces deux lancers sont indépendants (l'un n'influence pas l'autre) et donc si on note X_1 la variable aléatoire correspondant aux points obtenus sur le premier dé et X_2 celle pour le second dé, les variables aléatoires sont indépendantes. On peut en déduire que $\Pr[X_1 = 6 \wedge X_2 \leq 3] = \Pr[X_1 = 6]\Pr[X_2 \leq 3]$.

Remarque. Deux évènements dont l'intersection est vide sont dits *disjoints*. La probabilité de l'intersection est alors nulle. *Attention à ne pas confondre disjoint et indépendant : deux évènements disjoints ne sont pas indépendants !*

4 Quelques propriétés

Toutes les propriétés énoncées dans cette partie reste vraie dans le cas de probabilités et espérances conditionnelles. On les énonce sans conditionnement par simplicité d'écriture.

Proposition 4.1. *Soit $X : \Omega \to V$ une variable aléatoire réelle discrète, et E et F deux évènements. Alors*

i. $\sum_{v\in V} \Pr[X = v] = 1$*;*
ii. $\Pr[\neg E] = 1 - \Pr[E]$*;*
iii. Si E et F sont disjoints, $\Pr[E \wedge F] = 0$*;*
iv. $\Pr[E \vee F] = \Pr[E] + \Pr[F] - \Pr[E \wedge F]$*;*
v. $\Pr[E \vee F] \le \Pr[E] + \Pr[F]$ ***(inégalité de Boole)****.*

Démonstration. i. Par définition, $\sum_v \Pr[X = v] = \sum_v \sum_{\omega:X(\omega)=v} p_\omega = \sum_{\omega\in\Omega} p_\omega = 1$.
ii. $\Pr[\neg E] = \sum_{\omega\notin E} p_\omega = \sum_{\omega\in\Omega} p_\omega - \sum_{\omega\in E} p_\omega = 1 - \Pr[E]$.
iii. Si E et F n'ont aucun évènement primitif en commun, $\sum_{\omega\in E\wedge F} p_\omega$ est une somme vide, qui vaut donc 0.
iv. Par définition, $\Pr[E \vee F] = \sum_{\omega\in E\vee F} p_\omega$. Or $\Pr[E] + \Pr[F] = \sum_{\omega\in E} p_\omega + \sum_{\omega\in F} p_\omega$ et dans cette somme, les éléments qui sont à la fois dans E et dans F sont comptés deux fois. On trouve donc $\Pr[E \vee F]$ en retranchant les éléments comptés deux fois, c'est-à-dire $\sum_{\omega\in E\wedge F} p_\omega = \Pr[E \wedge F]$.
v. C'est une conséquence triviale du point précédent.

□

Exemple. On reprend l'exemple du dé avec la variable aléatoire X qui compte le nombre de points et Y la parité. La probabilité que $X = 3$ ou $Y = 1$ est $\Pr[X = 3 \vee Y = 1] = \Pr[X = 3] + \Pr[Y = 1] - \Pr[X = 3 \wedge Y = 1]$. Comme Y vaut 0 quand X vaut 3, on en déduit que cette probabilité vaut $1/6 + 1/2 = 2/3$, ce qu'on retrouve bien directement : les résultats possibles sont ⚁, ⚂, ⚃, ⚅.

Proposition 4.2. *Soit $X, Y : \Omega \to V$ des variables aléatoires réelles discrètes, et $\lambda, \mu \in \mathbb{R}$. Alors*

i. $\mathbb{E}[\lambda X + \mu Y] = \lambda\mathbb{E}[X] + \mu\mathbb{E}[Y]$ ***(linéarité de l'espérance)****;*
ii. pour $\lambda, \mu \in \mathbb{R}$, $\mathrm{Var}(\lambda X + \mu) = \lambda^2 \mathrm{Var}(X)$ *et* $\sigma(\lambda X + \mu) = \lambda\sigma(X)$*;*
iii. si X et Y sont indépendantes, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ *et* $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$*;*
iv. si $V \subset \mathbb{N}$, $\mathbb{E}[X] = \sum_{v\ge 1} \Pr[X \ge v]$.

Démonstration.

i. $\mathbb{E}[\lambda X + \mu Y] = \sum_\omega (\lambda X(\omega) + \mu Y(\omega))p_\omega = \lambda \sum_\omega X(\omega)p_\omega + \mu \sum_\omega Y(\omega)p_\omega = \lambda\mathbb{E}[X] + \mu\mathbb{E}[Y]$.
ii. $\mathbb{E}[\lambda X + \mu] = \lambda\mathbb{E}[X] + \mu$ donc $\mathbb{E}[\lambda X + \mu]^2 = \lambda^2\mathbb{E}[X]^2 + 2\lambda\mu\mathbb{E}[X] + \mu^2$. De plus, $\mathbb{E}\big[(\lambda X + \mu)^2\big] = \sum_\omega (\lambda X(\omega) + \mu)^2 p_\omega = \lambda^2\mathbb{E}\big[X^2\big] + 2\lambda\mu\mathbb{E}[X] + \mu^2$. En faisant la différence, on obtient $\lambda^2(\mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2) = \lambda^2 \mathrm{Var}(X)$. Et $\sigma(\lambda X + \mu) = \sqrt{\lambda^2 \mathrm{Var}(X)} = \lambda\sigma(X)$.
iii. $\mathbb{E}[XY] = \sum_x \sum_y xy \Pr[X = x \wedge Y = y] = \sum_x \sum_y (x \Pr[X = x])(y \Pr[Y = y])$. Pour conclure, il faut utiliser la propriété d'absolue convergence évoquée précédemment, qui permet de *sortir* $x \Pr[X = x]$ de la somme sur y. Pour la variance, l'indépendance implique que $\mathbb{E}\big[(X + Y)^2\big] = \mathbb{E}\big[X^2 + 2XY + Y^2\big] = \mathbb{E}\big[X^2\big] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}\big[Y^2\big]$. On conclut facilement.

iv. Puisque $V \subset \mathbb{N}$, $v \Pr[X=v]$ peut s'écrire $\Pr[X=v]+\cdots+\Pr[X=v]$ (v fois). Alors

$$\begin{aligned}\mathbb{E}[X] = \Pr[X=1]&\\ &+\Pr[X=2]+\Pr[X=2]\\ &+\Pr[X=3]+\Pr[X=3]+\Pr[X=3]\\ &+\quad\cdots\end{aligned}$$

En sommant verticalement, il vient

$$\begin{aligned}\mathbb{E}[X] &= \sum_{v\geq 1}\Pr[X=v]+\sum_{v\geq 2}\Pr[X=v]+\sum_{v\geq 3}\Pr[X=v]+\cdots\\ &= \Pr[X\geq 1]+\Pr[X\geq 2]+\Pr[X\geq 3]+\cdots\\ &= \sum_{v\geq 1}\Pr[X\geq v].\end{aligned}$$

□

Exemple. On prend deux dés indépendants, on note Z la variable aléatoire qui vaut le produit des résultats. On peut calculer l'espérance de Z en définissant les variables aléatoires X_1 et X_2 correspondant à chaque dé, qui sont indépendantes. Alors $Z = X_1X_2$ et $\mathbb{E}[Z] = \mathbb{E}[X_1]\mathbb{E}[X_2] = {}^{49}/_{144}$. Par contre, imaginons un jeu à un dé où on gagne le nombre de point du dé s'il est pair et 0 sinon. Alors le nombre de points gagnés est bien la variable aléatoire $W = XY$ où X et Y sont définis comme dans les exemples précédents. Mais ces deux variables ne sont pas indépendantes, et on remarque en effet que $\mathbb{E}[W] = {}^1/_6(2+4+6) = 2 \neq \mathbb{E}[X]\mathbb{E}[Y] = {}^7/_4$.

Les propriétés suivantes permettent d'introduire des probabilités conditionnelles quand elles sont absentes.

Proposition 4.3. *Soit Ω un espace probabilisé, et $E_1 \sqcup E_2 \sqcup \cdots \sqcup E_n$ une partition de Ω. Alors*

i. $\Pr[F] = \sum_{i=1}^{n}\Pr[F|E_i]\Pr[E_i]$ *pour tout évènement F (**formule des probabilités totales**) ;*

ii. $\mathbb{E}[X] = \sum_{i=1}^{n}\mathbb{E}[X|E_i]\Pr[E_i]$ *pour toute variable aléatoire X (**formule de l'espérance totale**).*

Démonstration. Par définition des probabilités conditionnelles, $\Pr[F|E_i]\Pr[E_i] = \Pr[F\wedge E_i]$. De plus, E_i et E_j étant disjoints pour $i\neq j$, $\Pr\big[(F\wedge E_i)\wedge(F\wedge E_j)\big] = 0$. D'après la proposition 4.1 iv.,

$$\sum_{i=1}^{n}\Pr[F|E_i]\Pr[E_i] = \Pr\left[\bigvee_{i=1}^{n} F\wedge E_i\right].$$

Comme $\bigvee_i E_i = \Omega$ et $F\subset\Omega$, $\bigvee_i F\vee E_i = F$. D'où le premier point.

Le second s'obtient directement du premier et de la définition de l'espérance. □

5 Distributions classiques

Définition 5.1. Une variable aléatoire X suit la

— **loi uniforme** si $X : \Omega \to V$ avec $\Pr[X = v] = 1/|V|$ pour tout $v \in V$;
— **loi de Bernoulli** (de paramètre p) si $X : \Omega \to \{0,1\}$ avec $\Pr[X = 1] = p$;
— **loi binomiale** (de paramètres p et n) si $X : \Omega \to \mathbb{N}$ avec $\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$;
— **loi géométrique** (de paramètre p) si $X : \Omega \to \mathbb{N}$ avec $\Pr[X = n] = p(1-p)^{n-1}$;
— **loi de Poisson** (de paramètre $\lambda > 0$) si $X : \Omega \to \mathbb{N}$ avec $\Pr[X = n] = e^{-\lambda}\lambda^k/k!$.

Intuitivement, la loi uniforme consiste à choisir un objet *uniformément* dans un ensemble V. La loi de Bernoulli est la loi d'une pièce biaisée avec probabilité p d'obtenir pile. La loi binomiale compte le nombre de piles pour n lancers d'une pièce biaisée. La loi géométrique compte le nombre de lancers nécessaires pour obtenir pile avec une pièce biaisée.

Lemme 5.2. *Soit $X_1, X_2, \ldots$ des variables aléatoires* indépendantes *suivant la loi de Bernoulli de paramètre p. Alors $X = X_1 + \cdots + X_n$ suit la loi binomiale de paramètres p et n. Et $Y = \min\{i : X_i = 1\}$ suit la loi géométrique de paramètre p.*

Démonstration. Pour que $X = k$, il faut que k des n variables aléatoires X_i valent 1, et $n-k$ valent 0. Il y a par définition $\binom{n}{k}$ façons de choisir les k variables aléatoires qui valent 1. Lorsque celles-ci sont fixées, la probabilité qu'elles vaillent 1 et les autres 0 est $p^k(1-p)^{n-k}$ puisqu'elles sont indépendantes.

Formellement,

$$\Pr[X = k] = \Pr\left[\bigvee_{\substack{S \subset \{1,\ldots,n\} \\ |S| = k}} \left(\bigwedge_{i \in S} X_i = 1 \wedge \bigwedge_{i \notin S} X_i = 0\right)\right].$$

Or pour deux sous-ensembles $S_1 \neq S_2$, les deux évènements $\bigwedge_{i \in S} X_i = 1 \wedge \bigwedge_{i \notin S} X_i = 0$ sont disjoints. D'autre part, les variables aléatoires X_i sont indépendantes. Donc

$$\Pr[X = k] = \sum_{\substack{S \subset \{1,\ldots,n\} \\ |S| = k}} \prod_{i \in S} \Pr[X_i = 1] \prod_{i \notin S} \Pr[X_i = 0] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Pour la variable aléatoire Y, on note que $Y = i$ si $X_j = 0$ pour $j < i$ et $X_i = 1$. Or la probabilité de cette conjonction d'évènements indépendants est $p(1-p)^{i-1}$. □

Proposition 5.3. *Soit X une variable aléatoire. Si X suit la loi*
— ***uniforme*** *et si $V = \{1, \ldots, n\}$, $\mathbb{E}[X] = (n+1)/2$ et $\sigma^2(X) = (n^2-1)/12$;*
— ***de Bernoulli*** *de paramètre p, $\mathbb{E}[X] = p$ et $\sigma^2(X) = p(1-p)$;*
— ***binomiale*** *de paramètres p et n, $\mathbb{E}[X] = np$ et $\sigma^2(X) = np(1-p)$;*
— ***géométrique*** *de paramètre p, $\mathbb{E}[X] = 1/p$ et $\sigma^2(X) = (1-p)/p^2$;*
— ***de Poisson*** *de paramètre λ, $\mathbb{E}[X] = \sigma^2(X) = \lambda$.*

Démonstration.
— Uniforme : $\mathbb{E}[X] = \sum_{v=1}^{n} v/n = \frac{1}{n}\frac{n(n+1)}{2}$ et $\sigma^2(X) = \frac{1}{n}(\sum_v v^2 - (\sum_v v)^2) = \frac{1}{n}(n(n+1)(2n+1)/6 - (n(n+1)/2)^2) = (n^2-1)/12$.
— Bernoulli : exercice.
— Binomiale : On considère n variables aléatoires indépendantes $X_1, \ldots, X_n$ suivant la loi de Bernoulli et $X = X_1 \cdots + X_n$. D'après le lemme 5.2, X suit la loi binomiale. Donc par linéarité de l'espérance, $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$. De plus, puisque les X_i sont indépendantes, $\sigma^2(X) = \sum_{i=1}^{n} \sigma^2(X_i) = np(1-p)$.

— $\mathbb{E}[X] = \sum_{n\geq 0} np(1-p)^{n-1} = p/(1-(1-p))^2 = 1/p$ et $\sigma^2(X) = \sum_{n\geq 0} n^2 p(1-p)^{n-1} - 1/p^2 = (1-p)/p^2$ car la dérivée seconde de $\sum_{n\geq 0} x^n$ est $\sum_{n\geq 0} n^2 x^{n-1} + \sum_{n\geq 0} nx^{n-1}$.

□

6 Inégalités

Les bornes énoncées dans cette partie permettent d'obtenir des informations sur une variable aléatoire à partir de son espérance ou de sa variance.

Inégalité de Markov. *Soit X une variable aléatoire discrète à valeur positives ou nulles. Alors*

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t} \text{ et } \Pr[X \geq \lambda \mathbb{E}[X]] \leq \frac{1}{\lambda}.$$

Démonstration. On remarque que les deux formulations sont équivalentes, on posant $t = \lambda \mathbb{E}[X]$. La première s'obtient avec les deux inégalités

$$\mathbb{E}[X] = \sum_{v \in V} v \Pr[X = v] \geq \sum_{v \geq t} v \Pr[X = v] \geq \sum_{v \geq t} t \Pr[X = v] = t \Pr[X \geq t].$$

□

Exercice. Pourquoi doit-on supposer que X est à valeurs positives ou nulles ?

Inégalité de Tchebycheff. *Soit X une varaiable aléatoire et t un réel strictement positif. Alors*

$$\Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\sigma(X)^2}{t^2}.$$

Démonstration. L'inégalité $|X - \mathbb{E}[X]| \geq t$ est équivalente à $(X - \mathbb{E}[X])^2 \geq t^2$. Il suffit donc d'appliquer l'inégalité de Markov à la variable aléatoire $(X - \mathbb{E}[X])^2$ (qui est bien à valeurs positives ou nulles) avec la borne t^2. On obtient $\Pr\left[(X - \mathbb{E}[X])^2 \geq t^2\right] \leq \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]/t^2$. Cela conclut la preuve par définition de l'écart-type, qui est la racine carrée de la variance. □

Les résultats suivants sont plus complexes à démontrer. Ils sont cependant très puissants et servent souvent dans l'analyse d'algorithmes probabilistes. Il existe de nombreuses formes plus ou moins équivalentes des deux inégalités énoncées ci-dessous. On se concentre sur les formulations qui nous sont le plus utile.

Inégalités de Chernoff. *Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes à valeur dans $\{0, 1\}$, telles que $\Pr[X_i = 1] = p_i \in]0, 1[$ pour tout i, et $X = X_1 + \cdots + X_n$. Alors*

i. $\Pr[X > (1 + \delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/(2+\delta)}$ *pour tout $\delta > 0$, et*
ii. $\Pr[X < (1 - \delta)\mathbb{E}[X]] \leq e^{-\mathbb{E}[X]\delta^2/2}$ *pour $0 < \delta \leq 1$.*

Démonstration. On note $\mu = \mathbb{E}[X]$ dans toute la preuve. On s'intéresse à la variable aléatoire e^{tX} pour $t > 0$. Alors

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t \sum_i X_i}\right] = \mathbb{E}\left[\prod_i e^{tX_i}\right] = \prod_i \mathbb{E}\left[e^{tX_i}\right]$$

par indépendance des X_i. Or $\mathbb{E}\left[e^{tX_i}\right] = p_i e^t + (1-p_i)$ car e^{tX_i} prend la valeur e^t avec probabilité p_i, et la valeur 1 avec probabilité $(1-p_i)$. D'autre part, comme $1+x \leq e^x$ pour tout x, $\mathbb{E}\left[e^{tX_i}\right] = 1+(e^t-1)p_i \leq e^{(e^t-1)p_i}$. Ainsi,

$$\mathbb{E}\left[e^{tX}\right] \leq \prod_i e^{(e^t-1)p_i} = e^{(e^t-1)\sum_i p_i} = e^{(e^t-1)\mu}$$

puisque $\sum_i p_i = \sum_i \mathbb{E}[X_i] = \mu$.

On peut maintenant démontrer chacun des deux points.

i. Pour tout réel strictement positif t, $X > (1+\delta)\mu$ est équivalent à $e^{tX} > e^{t(1+\delta)\mu}$, donc $\Pr[X > (1+\delta)\mu] = \Pr\left[e^{tX} > e^{t(1+\delta)\mu}\right]$. D'après l'inégalité de Markov,

$$\Pr\left[e^{tX} > e^{t(1+\delta)\mu}\right] \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}}.$$

On obtient donc

$$\Pr[X > (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}.$$

En posant $t = \ln(1+\delta)$,

$$\Pr[X > (1+\delta)\mu] \leq \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}}.$$

On peut alors montrer [1] que $(1+\delta)^{1+\delta} \geq e^{\delta\frac{1+\delta}{1+\delta/2}}$, d'où on obtient le résultat.

ii. De la même manière, on montre que

$$\Pr\left[e^{tX} < e^{t(1-\delta)\mu}\right] \leq \frac{e^{-\delta\mu}}{(1-\delta)^{(1-\delta)\mu}}$$

d'où on déduit le résultat en remarquant que pour $0 < \delta < 1$, $(1-\delta)^{1-\delta} > e^{\delta^2/2-\delta}$.

□

1. Prendre le logarithme népérien de chaque côté, et montrer que $\ln(1+x) \geq x/(1+x/2)$ pour tout $x > 0$.