

Probabilités et Statistiques
Licence de Mathématiques (Parcours Math-Info),
UE MA6031

Alain Yger

INSTITUT DE MATHÉMATIQUES, UNIVERSITÉ BORDEAUX 1, TALENCE 33405,
FRANCE

E-mail address: `Alain.Yger@math.u-bordeaux1.fr`

Version du 26 novembre 2012.

RÉSUMÉ. L'objectif du cours est une introduction « pratique » à la théorie des probabilités et au raisonnement statistique. S'agissant d'un cours ciblé sur le parcours Math-Info, l'accent sera mis en priorité sur tout ce qui relève plutôt des mathématiques « discrètes ». Lois normales, vecteurs gaussiens, statistique gaussienne seront aussi bien sûr envisagés. Une référence sur lequel ce cours s'appuiera (souvent de manière très « light ») sera les chapitres 11 (Probabilités) et 12 (Statistique descriptive, statistique inférentielle) de [MathAp], voire aussi les chapitres 7 (codage), 8 (Cryptographie), 10 (Signal) de ce même ouvrage pour certaines applications en relation avec d'autres champs. Ce cours sera illustré avec l'utilisation de logiciels (MATLAB, avec son clone libre Scilab, disponible sur <http://www.scilab.org/fr>), permettant en particulier tous deux la simulation de lois probabilistes. Des logiciels plus spécialement dédiés au raisonnement statistique (comme le logiciel libre R, <http://www.r-project.org/>) seront également mis à contribution dans la seconde partie du cours (dédiée au raisonnement statistique, où interviennent les lois de probabilité, « en famille » cette fois). Sont fournies en annexes les annales 2011-2012 : texte et corrigé du DS (1h30), textes et corrigés des examens de session 1 et de session 2 (3h00).

Table des matières

Chapitre 1. Univers des possibles, probabilités, indépendance et conditionnement	1
1.1. En guise d'introduction : la notion d'« épreuve » ou expérience	1
1.2. Mesure de probabilité sur un ensemble fini	3
1.3. Mesure de probabilité sur un ensemble dénombrable	9
1.4. Tribus et probabilités sur un ensemble quelconque	13
1.5. Conditionnement et indépendance	21
Chapitre 2. Variables aléatoires et théorèmes « limite »	31
2.1. Notion de VAR, espérance, variance, moments	31
2.2. Le Théorème Limite « Central » (TLC)	45
2.3. Les lois des grands nombres (LGNf) et (LGNF)	48
Chapitre 3. Initiation au raisonnement statistique	51
3.1. Indicateurs numériques d'une série de données statistiques	51
3.2. La notion d'estimateur ; des probabilités à la statistique	53
3.3. Estimation par intervalle ; gaussiennes et test de Student	54
Annexe A. Annales 2011-2012, Texte et corrigé du DS	61
Annexe B. Annales 2011-2012, Texte et corrigé de l'examen de session 1	67
Annexe C. Annales 2011-2012, Texte et corrigé de l'examen de session 2	75
Bibliographie	79
Index	81

Univers des possibles, probabilités, indépendance et conditionnement

1.1. En guise d'introduction : la notion d'« épreuve » ou expérience

Le lancer d'un dé à six faces constitue un exemple de ce que l'on appelle une *épreuve* (ou une *expérience*), l'ensemble $\{1, \dots, 6\}$ constituant la liste des résultats possibles pour cette expérience.

Si le dé n'est pas pipé, tout non-mathématicien intuirait qu'il a une chance sur six de réaliser « six » lors d'une telle épreuve. Il argumenterait pour cela empiriquement en vous faisant constater que, s'il itère M fois cette épreuve (avec M très grand), le quotient $M(6)/M$ (où $M(6)$ désigne le nombre de « six » obtenus lors de la série de M coups) se met à tendre asymptotiquement vers $1/6$ lorsque M tend vers l'infini. Pareille argumentation relève, on le verra plus loin dans ce cours, du point de vue de la *statistique empirique*.

Formulons mathématiquement le problème en considérant l'ensemble Ω de tous les résultats possibles de l'épreuve, ici

$$\Omega = \{1, \dots, 6\}.$$

Le résultat de l'épreuve est *aléatoire*. L'épreuve consiste en effet à choisir « au hasard » un élément dans l'ensemble Ω , espace que l'on qualifie ici, on le verra dans la section suivante, d'*univers des possibles*. Si le résultat aléatoire de l'épreuve relève du hasard, ce hasard n'en est pas moins « organisé » : se donner la *loi probabiliste* à laquelle obéit cette épreuve aléatoire matérialisée par le lancer du dé, c'est-à-dire la « loi » qui « pilote » ici le hasard, consiste ici à se donner une collection de 6 nombres réels positifs p_j , $j = 1, \dots, 6$ de somme 1. La quantité p_j représente ce que notre non-mathématicien interpréterait comme *le nombre de chances de réaliser le résultat j en lançant le dé*, ou encore la *probabilité de l'évènement* : « le résultat de l'épreuve est un « six » ».

Dans le cas de dés non pipés, il est naturel de travailler avec le modèle mathématique :

$$p_j = \frac{1}{6}, \quad j = 1, \dots, 6.$$

Ce modèle est un cas particulier du cas où Ω est un ensemble fini et où chaque singleton (ou encore évènement élémentaire) $\{a\}$ de Ω est « chargé » avec la masse

$$P(\{a\}) = \frac{1}{\text{card } \Omega},$$

auquel cas, une partie A de Ω (que l'on appelle un *évènement*) est « chargée » avec la « masse » (ou le « poids »)

$$P(A) = \frac{\text{card } A}{\text{card } \Omega};$$

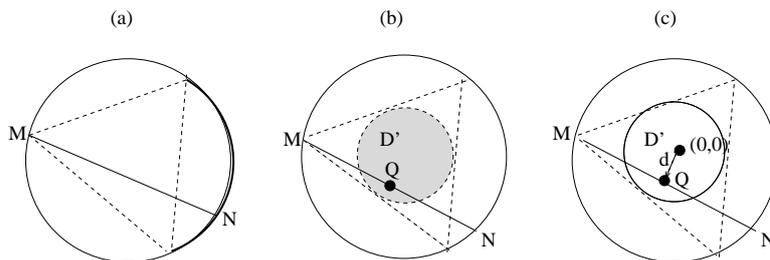


FIGURE 1.1. Le paradoxe de Bertrand.

cette distribution de probabilité est dite *distribution uniforme* sur Ω .

Dans le cas de dés pipés, on doit par contre envisager d'autres modèles, comme par exemple

$$p_1 = p_2 = p_3 = p_4 = \frac{1}{12}, \quad p_5 = p_6 = \frac{1}{3}.$$

L'univers des possibles n'est pas toujours fini, comme le montre l'exemple *géométrique* suivant : l'épreuve consiste à prendre au hasard deux points distincts sur un cercle Γ donné du plan \mathbb{R}^2 de centre l'origine et de rayon 1 et à tracer la corde qui les joint (comme sur trois schémas présentés sur la figure 1.1). On peut poser la question naive suivante : les choix de points étant équiprobables sur $\Gamma \times \Gamma$ (on va préciser ceci ultérieurement), quel « nombre de chances » ou encore « probabilité » a-t-on que la longueur de la corde joignant deux points M et N de Γ soit supérieure à celle (ici $\sqrt{3}$, car le cercle Γ est de rayon 1) du côté d'un triangle équilatéral inscrit dans Γ ? (voir la figure 1.1, schéma de gauche). Notre non-spécialiste pourra répondre en disant qu'il y a une chance sur trois que ce qu'il souhaite se réalise lorsqu'il choisit les deux points au hasard (et indépendamment l'un de l'autre) : son raisonnement heuristique s'appuie sur le fait que, dès que l'un des points est fixé, le second doit être dans un arc de cercle de longueur $\pi/3$. Ce calcul empirique de probabilité est fondé sur le fait que l'univers des possibles Ω choisi ici est le produit $\Omega = \Gamma \times \Gamma$ du cercle par lui-même, la probabilité d'un produit d'arcs de cercles $I \times J$ étant

$$\frac{\text{longueur}(I)}{\pi} \times \frac{\text{longueur}(J)}{\pi}.$$

Ceci est cependant ambigu, comme le faisait remarquer le mathématicien Joseph Bertrand en 1899. On peut en effet aussi repérer la corde par son milieu Q et considérer comme ensemble d'événements Ω l'ensemble des positions possibles de ce milieu après tracé de la corde, soit le disque D dont Γ est le bord ; cette fois, l'univers des possibles n'est plus l'ensemble des paires de points de Γ , mais l'ensemble des points du disque fermé $D(0, 1)$ (ces points repèrent le milieu de la corde, voir la figure 1.1, schéma central) ; avec ce choix, le fait que la corde ait une longueur supérieure à celle du triangle équilatéral inscrit se rephrase en « le milieu de la corde est dans le disque D' de rayon $1/2$ et de centre l'origine » ; notre non-spécialiste pourrait donc tout autant affirmer qu'il y a une chance sur 4 (puisque $1/4$ représente le quotient de la surface de D' par la surface de $D(0, 1)$) que son souhait se réalise. On pourrait aussi repérer la corde (encore une fois après l'avoir tracé) par sa distance $d \in [0, 1]$ à l'origine (voir la figure 1.1, schéma de droite) et, en prenant comme univers des

possibles $[0, 1]$, dire que l'évènement que l'on cherche est celui qui correspond au fait que cette distance soit inférieure ou égale à $1/2$, soit l'intervalle $[0, 1/2]$, qui est de longueur $1/2$; comme la longueur de $[0, 1]$ est normalisée égale à 1, la probabilité de notre évènement suivant cette troisième interprétation vaudrait $1/2$.

Cet exemple (paradoxal) illustre le fait que la mise sous forme mathématique d'un problème de nature probabiliste exige au préalable la définition rigoureuse de trois choses :

- un « *univers Ω des possibles* » (les points de Ω peuvent par exemple permettre de repérer les résultats d'une épreuve, on les appelle pour cela « *réalisations* » de l'épreuve); c'est l'ubiquité de cet univers des possibles qui explique le paradoxe de Bertrand;
- une famille de sous-ensembles de Ω candidats à pouvoir être *mesurés* (la contrainte étant que le mathématicien n'a que l'accès au dénombrable pour envisager le calcul de mesures); ce seront les « *évènements aléatoires* »;
- une règle enfin pour « *mesurer* » chacun de ces évènements aléatoires, avec la convention que la mesure de l'univers des possibles Ω vaille 1; si Ω représente (comme c'est souvent le cas) l'ensemble des résultats d'une certaine épreuve, alors la « *mesure* » de A sera interprétée par notre non-spécialiste comme le « *pourcentage de chances* » qu'a le résultat de l'épreuve de « *tomber* » dans l'évènement aléatoire A ; on l'appellera (dans le contexte où se déroule l'épreuve) « *probabilité* de A ». La règle pour mesurer les évènements aléatoires s'appelle « *loi de probabilité* » régissant l'épreuve.

Nous allons dans la suite de ce chapitre définir proprement les notions mathématiques consolidant ces idées intuitives. Le modèle mathématique rigoureux (que nous ne ferons qu'esquisser ici) est du au mathématicien soviétique Andreï Kolmogorov (1903-1997).

1.2. Mesure de probabilité sur un ensemble fini

1.2.1. Généralités. Se donner une *probabilité* sur un ensemble fini Ω (ou encore une *mesure de probabilité* sur cet ensemble), c'est par définition affecter à chacun des sous-ensembles A de Ω un nombre réel $P(A) \in [0, 1]$, ce en respectant les clauses suivantes :

$$(1.1) \quad \begin{aligned} \forall A, B \in \mathcal{P}(\Omega), A \cap B = \emptyset \implies P(A \cup B) &= P(A) + P(B) \\ P(\Omega) &= 1. \end{aligned}$$

Une fois ces deux règles posées, on dispose des règles de calcul suivantes :

$$(1.2) \quad \begin{aligned} P(\emptyset) &= 0 \\ P(\Omega \setminus A) &= 1 - P(A) \quad \forall A \in \mathcal{P}(\Omega) \\ P(A \cup B) &= \overbrace{P(A) + P(B)} - P(A \cap B) \quad \forall A, B \in \mathcal{P}(\Omega) \\ P(A \cup B \cup C) &= \overbrace{P(A) + P(B) + P(C)} - \overbrace{(P(B \cap C) + P(C \cap A) + P(A \cap B))} \\ &\quad + P(A \cap B \cap C) \quad \forall A, B, C \in \mathcal{P}(\Omega) \quad \text{etc.} \end{aligned}$$

On peut d'ailleurs établir en exercice la formule générale, dite *de Poincaré* :

$$(1.3) \quad \begin{aligned} P(A_1 \cup \dots \cup A_N) &= \\ &= \sum_{j=1}^N (-1)^{j-1} \sum_{1 \leq j_1 < \dots < j_k \leq N} P(A_{j_1} \cap \dots \cap A_{j_k}) \quad \forall A_1, \dots, A_N \in \mathcal{P}(\Omega). \end{aligned}$$

REMARQUE 1.1. La donnée d'une probabilité sur un ensemble fini à N éléments (que l'on note x_1, \dots, x_N , par conséquent $\Omega = \{x_1, \dots, x_N\}$) équivaut à celle d'un vecteur ligne

$$[p_1, \dots, p_N]$$

d'entrées toutes dans $[0, 1]$ et vérifiant de plus $\sum_{j=1}^N p_j = 1$. On pose alors, pour chaque $j = 1, \dots, N$, $P(\{x_j\}) = p_j$ et, pour tout sous-ensemble $A = \{x_{j_1}, \dots, x_{j_k}\}$ de Ω (j_1, \dots, j_k étant k numéros distincts dans $\{1, \dots, N\}$),

$$P(A) = P(\{x_{j_1}, \dots, x_{j_k}\}) = \sum_{l=1}^k P(\{x_{j_l}\}) = \sum_{l=1}^k p_{j_l}.$$

DÉFINITION 1.1. L'ensemble fini Ω est appelé *univers des possibles*. Les éléments x_1, \dots, x_N sont alors interprétés comme les résultats d'une épreuve. Toute partie A de Ω est appelée *événement aléatoire* (l'évènement étant : « le résultat de l'épreuve appartient à A »). Les singletons $\{x_j\}$, $j = 1, \dots, N$, sont appelés *événements aléatoires élémentaires*.

Se donner une *probabilité* sur un ensemble fini Ω de cardinal $N \in \mathbb{N}^*$ revient à fixer la règle conditionnant l'épreuve dont Ω recense toutes les réalisations possibles. L'application $A \in \mathcal{P}(\Omega) \mapsto P(A)$ (ou encore, ce qui revient au même ici, la donnée du vecteur ligne $[p_1, \dots, p_N]$, $p_j \in [0, 1]$, $\sum_j p_j = 1$) est dite *loi de probabilité* de l'épreuve.

EXEMPLE 1.1 (l'exemple du lancer de dé pipé ou non, cf. l'introduction de cette section). Supposons que l'expérience consiste en le jet (une fois) d'un dé à six faces. L'univers des possibles associé est $\{1, \dots, 6\}$. Dire que le dé n'est pas pipé revient à dire que tous les résultats de l'expérience sont affectés du même « poids », c'est-à-dire que la probabilité P conditionnant l'expérience est la probabilité dite *uniforme*, où tous les événements élémentaires $\{j\}$, $j = 1, \dots, N$, sont affectés du même poids $p_j = P(\{j\}) = 1/6$. Les affecter de poids p_j non tous égaux à $1/6$ revient à dire que l'expérience est faite avec un dé pipé (ou sous des conditions qui nous ramènent à pareille situation, ce qui revient au final au même).

1.2.2. La loi uniforme discrète $\mathcal{U}(N)$. La loi uniforme sur un ensemble Ω de cardinal N est celle où le vecteur ligne $[p_1, \dots, p_N]$ est le vecteur $[1/N, \dots, 1/N]$. Pour tout sous-ensemble A de Ω , on a donc $P(A) = |A|/N$. La simulation sous un logiciel de calcul scientifique tel MATLAB se fait *via* la routine `rand` : la commande

```
>> f=ceil(N.*rand(M,1));
```

permet de simuler M fois l'épreuve consistant à choisir un entier entre 1 et N , tous les choix étant équiprobables, *i.e.* affectés du poids $1/N$. Sous Scilab, la routine correspondante est :

```
function [f]= RndInt (1,M,1,N)
```

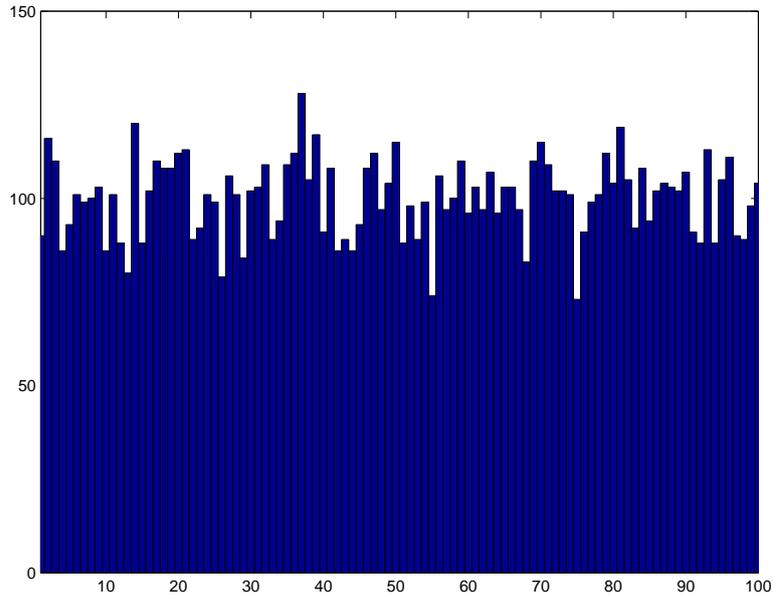


FIGURE 1.2. Histogramme de $M = 10000$ réalisations d'une épreuve suivant une loi uniforme $\mathcal{U}(100)$.

Il faut noter que la génération aléatoire d'entiers (entre 1 et N) suivant une loi uniforme passe par l'étape consistant à choisir un « réel » (exprimé en virgule flottante) de manière aléatoire dans $]0, N[$, puis à l'arrondir (routine `ceil`) à l'élément de $\{1, \dots, N\}$ immédiatement supérieur. La routine `rand` (sous MATLAB) génère en effet un nombre en virgule flottante (tous les choix étant équiprobables) dans $]0, 1]$. Sur la figure 1.2, on a représenté l'histogramme (obtenu suivant la commande `hist(f, x)`, où `x=1:N`) correspondant à la répartition de $M = 10000$ tirages aléatoires d'un nombre entier entre 1 et 100 suivant le mécanisme de génération décrit ci-dessus (loi uniforme sur $\{1, \dots, 100\}$). On notera que lorsque M devient grand, le pourcentage de tirages donnant comme résultat un entier donné x de $\{1, \dots, 100\}$ tend vers 1%, ce qui correspond au fait que $P(\{x\}) = 1/100$ pour tout x .

REMARQUE 1.2. Lorsque l'on envisage les choses sous l'angle statistique, nous verrons que la loi uniforme mérite aussi le qualificatif de « loi du risque pire ». Cela se traduit concrètement par le fait que l'histogramme de la figure 1.2 soit sensiblement « à niveau constant » (ici en l'occurrence à peu près 100).

1.2.3. Loi binomiale $\mathcal{B}(N, p)$ et approximations. Soit $N \in \mathbb{N}^*$. Le nombre de parties à k éléments ($0 \leq k \leq N$) de l'ensemble à $N + 1$ éléments $\{0, \dots, N\}$ est égal au coefficient binomial $\binom{N}{k} = N! / (k!(N - k)!)$. Supposons que l'ensemble $\Omega = \{0, \dots, N\}$ soit interprété comme l'univers des possibles dans l'épreuve suivante : on totalise le nombre S_N de « pile » obtenus lors d'une suite de N lancers d'une pièce qui peut être pipée, ces lancers étant considérés comme « indépendants » les uns des autres¹. Il existe un nombre $p \in [0, 1]$ (différent de $1/2$ si et seulement si

1. Il faut ici comprendre cette notion d'« indépendance » de manière heuristique ; ceci sera précisé ultérieurement.

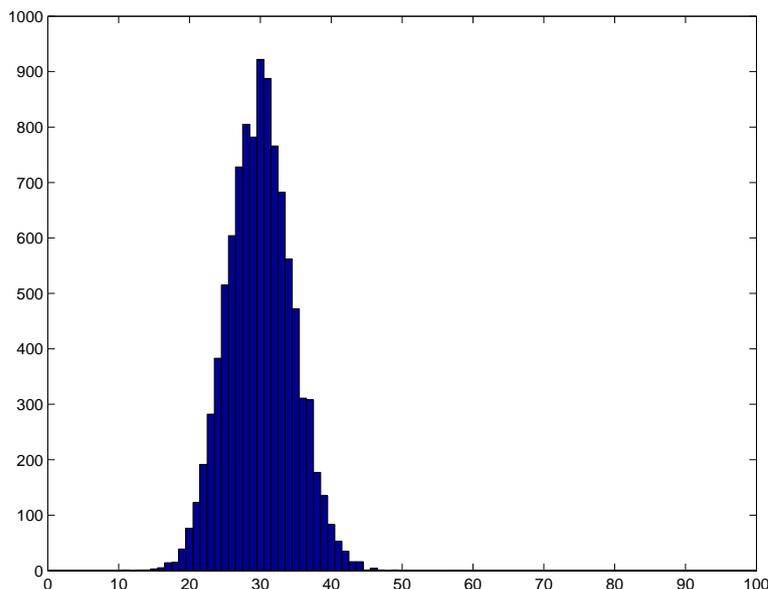


FIGURE 1.3. Histogramme de $M = 10000$ réalisations suivant une loi $\mathcal{B}(100, 0.3)$.

la pièce est pipée) tel que le pourcentage de chances de tomber sur « pile » lors d'un lancer vaille p . Le vecteur ligne $[p_0, \dots, p_N]$ correspondant à la distribution de probabilité qui régit cette épreuve dont les résultats possibles (les valeurs possibles prises par S_N) sont $\{0, \dots, N\}$ est le vecteur ligne à $N + 1$ entrées :

$$\left[(1-p)^N, \binom{N}{1} p(1-p)^{N-1}, \dots, \binom{N}{k} p^k (1-p)^{N-k}, \dots, \binom{N}{N-1} p^{N-1} (1-p), p^N \right].$$

La routine

```
>> f=binornd(N,p,1,M);
```

permet de simuler M fois l'épreuve consistant à totaliser le nombre de « pile » obtenus au terme d'une partie de pile ou face à N lancers. Sous Scilab, la même simulation s'effectue sous la routine

```
function [f] = Binomial(1,M,p,N)
```

Sur la figure 1.3, on a représenté l'histogramme (obtenu suivant la commande `hist(f,x)`, où `x=0:N`) correspondant à la répartition de $M = 10000$ bilans aléatoires S_{100} de séquences indépendantes de parties de pile ou face à $N = 100$ lancers (avec ici $p = 0.3$) suivant le mécanisme de génération décrit ci-dessus (loi binomiale sur $\{0, \dots, 100\}$). On notera que lorsque M devient grand, le pourcentage de réalisations donnant comme résultat un entier donné k de $\{0, \dots, 100\}$ tend vers $\binom{100}{k} p^k (1-p)^{100-k}$, où $p = 0.3$. Dans le cas particulier $N = 1$, la loi $\mathcal{B}(p, 1)$ s'appelle *loi de Bernoulli*; l'univers des possibles est alors un ensemble à deux éléments qui peut être incarné par $\{P, F\}$ (lancer de pile ou face), ou encore par $\{0, 1\}$ ou $\{-, +\}$. Voici deux modèles classiques où l'épreuve est régie par une loi binomiale, le premier d'inspiration « discrète », le second d'inspiration « continue » :

- Le tirage avec remise dans une urne contenant R boules rouges et B boules blanches (ces nombres R et B étant tous les deux supérieurs au égaux à un entier $N \geq 1$ donné) : le nombre de boules rouges tirées au terme de N tirages successifs (effectués de manière indépendantes les uns des autres) suit une loi binomiale $\mathcal{B}(N, p)$, avec $p = R/(B + R)$.
- Soit $[0, T]$ un intervalle temporel. L'épreuve consiste à répéter N fois (de manière indépendante) l'expérience consistant à « marquer » un point de $[0, T]$, tous les marquages étant équiprobables (on dit que le marquage d'un point est « uniforme » par analogie avec la terminologie utilisée dans le cas discret, cf. la sous-section 1.2.2, on y reviendra). Le nombre de points ainsi marqués dans l'intervalle $[t_1, t_2] \subset [0, T]$ suit une loi binomiale du type $\mathcal{B}(N, (t_2 - t_1)/T)$. La présentation de ce modèle « continu » d'épreuve régie par la loi binomiale prépare ici l'approche de la distribution de Poisson (sous-section 1.3.1).

Lorsque N devient grand, les calculs des coefficients binomiaux deviennent numériquement hors de portée, sauf à prendre le logarithme et à utiliser la formule de Stirling que nous mentionnons plus loin (en (1.7)). Il convient donc de connaître des approximations des probabilités des événements $A \subset \{0, \dots, N\}$ lorsque la loi en jeu est la loi binomiale $\mathcal{B}(N, p)$, avec N grand. Ces approximations sont de deux types :

- Lorsque $N \rightarrow \infty$, $p \rightarrow 0$ avec $Np = \lambda$ restant constant, on constate que

$$(1.4) \quad \lim_{N \rightarrow +\infty} \binom{N}{k} p^k (1-p)^{N-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

En pratique, on peut considérer une telle approximation

$$(1.5) \quad \binom{N}{k} p^k (1-p)^{N-k} \simeq \frac{(Np)^k}{k!} e^{-Np}$$

valable lorsque $N \geq 30$ et $Np \leq 5$ ou $N \geq 50$ et $p < 1/10$.

- On observe aussi (pour l'instant empiriquement, cf. par exemple l'histogramme figurant sur la figure 1.3) que, si p est fixé et N est assez grand,

$$(1.6) \quad \left| \sum_{k=0}^{[x]} \binom{N}{k} p^k (1-p)^{N-k} - \frac{1}{\sqrt{2\pi} \sqrt{Np(1-p)}} \int_{-\infty}^x \exp\left(-\frac{(t-Np)^2}{2Np(1-p)}\right) dt \right| \leq C \frac{p^2 + (1-p)^2}{\sqrt{p(1-p)}} \frac{1}{\sqrt{N}} \quad ([x] := \text{floor}[x], \quad x \in [0, N]), \quad C \leq 0.7655.$$

On peut fonder ici ce résultat (admis pour l'instant, version quantifiée du célèbre théorème de Moivre-Laplace dont nous reparlerons) pour une part sur la *formule de Stirling*

$$(1.7) \quad q! \simeq \sqrt{2\pi} q^{q+1/2} e^{-q} \quad q \rightarrow +\infty,$$

mais on le justifiera plus loin dans ce cours en énonçant le « Théorème Limite Central » (TLC) dans la section 2.2. La version quantifiée (très utile comme « garde-fou ») que nous citons pour l'instant ici en (1.6) est le *théorème de*

*Berry-Esseen*². L'approximation

$$(1.8) \quad \sum_{k=0}^{[x]} \binom{N}{k} p^k (1-p)^{N-k} \simeq \frac{1}{\sqrt{2\pi} \sqrt{Np(1-p)}} \int_{-\infty}^x \exp\left(-\frac{(t-Np)^2}{2Np(1-p)}\right) dt \\ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-Np}{\sqrt{Np(1-p)}}} e^{-u^2/2} du, \quad x \in [0, N],$$

se trouve ainsi justifiée pourvu que N soit assez grand et $p(1-p)$ pas trop petit (en pratique, dès que $N > 30$ et $\min(Np, N(1-p)) > 5$ ou $N > 30$ et $Np(1-p) > 9$). On constate que le champ de validité de cette approximation (1.8) ne recoupe pas en général celui de l'approximation (1.4) : dans un cas (pour l'approximation (1.4)), Np doit être petit, donc p petit, dans l'autre cas (pour l'approximation (1.8)), on doit éviter précisément les petites valeurs de Np .

1.2.4. La loi hypergéométrique de paramètres $R+B, R, N$ et les lois polyhypergéométriques. Considérons une urne renfermant $R+B$ boules, R de ces boules étant rouges et les B autres étant blanches. L'épreuve consiste maintenant à répéter $N \leq R+B$ fois l'opération de tirage d'une boule dans l'urne (suivant une loi uniforme $\mathcal{U}(R+B-(l-1))$, l désignant le numéro du tirage, ce qui signifie que les $R+B-(l-1)$ boules présentes dans l'urne au moment du tirage ont la même chance d'être tirées), suivie de l'exclusion de la boule tirée de l'urne (tirage sans remise), puis à compter le nombre de boules rouges tirées au terme des N tirages successifs sans remise. L'univers des possibles (pour ce nombre de boules rouges tirées) est alors l'ensemble fini

$$\Omega = \{\max(0, N-B), \dots, \min(N, R)\} \subset \mathbb{N}.$$

Chacun des entiers k de cet ensemble d'entiers Ω est affecté de la probabilité

$$p_k := P(\{k\}) = \frac{\binom{R}{k} \times \binom{B}{N-k}}{\binom{R+B}{N}}$$

et l'on peut interpréter dans ce modèle p_k comme la probabilité de l'évènement aléatoire suivant : « le nombre total de boules rouges tirées au terme des N tirages est égal à k ». La loi de probabilité ainsi définie est dite *loi hypergéométrique* $\mathcal{H}(N; R, B)$ de paramètres $R+B, R, N$. Elle est simulée, répétée M fois, sous MATLAB *via* la routine suivante :

```
>> f=hygernd(R+B,R,N,1,M);
```

Sous Scilab, il convient tout d'abord de calculer les deux quantités

$$(1.9) \quad E_{N,R,B} = N \times \frac{R}{R+B}, \quad V_{N,R,B} = N \times \frac{R}{R+B} \times \frac{B}{R+B} \times \frac{B+R-N}{B+R-1}$$

(qui représentent, on le verra plus tard, *moyenne* et *variance* de la loi) et la routine est alors :

```
function [f]=HyperGeom(1,M,E,V)
```

2. Dû à Andrew Berry et Carl-Gustav Esseen (autour de 1941-1942).

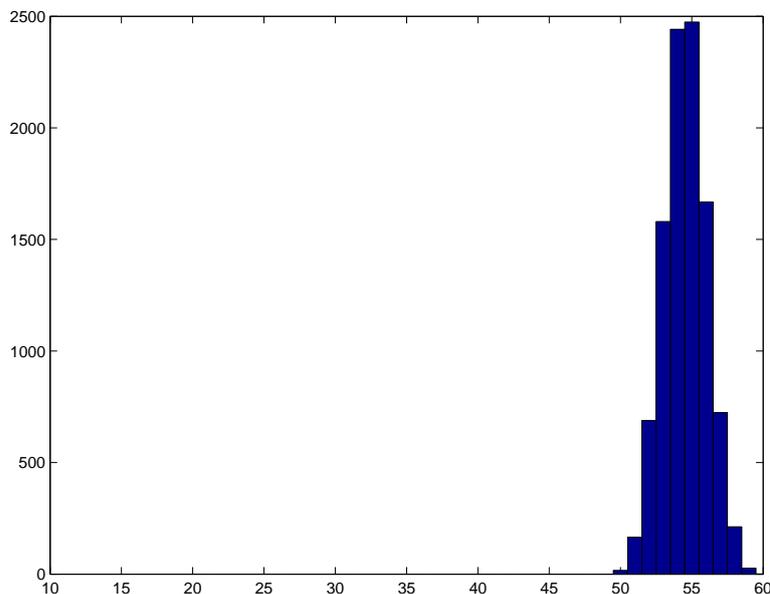


FIGURE 1.4. Histogramme de $M = 10000$ réalisations suivant une loi hypergéométrique de paramètres $N = 100, R = 60, B = 50$.

On en a représenté l'histogramme (avec $M = 10000$) avec les valeurs $R = 60, B = 40, N = 50$, l'espace des possibles étant $\{10, \dots, 50\}$, sur la figure 1.4. La loi hypergéométrique régit le modèle mathématique soutendant les enquêtes d'opinion : il convient en effet de ne pas interroger deux fois la même personne. Lorsque l'on a $N/(R + B) < 1/10$, l'approximation de la loi par la loi binomiale $\mathcal{B}(N, p)$, où $p = R/(R + B)$, devient licite.

On peut envisager des situations plus complexes avec une population de boules constituée de R boules rouges, de B boules blanches, de V boules vertes, *etc.* Ce sont ces situations que l'on rencontre dans les enquêtes d'opinion où les sondés sont soumis à un questionnaire à « choix multiples » (et non plus à une simple alternative). Ce sont les lois dites *polyhypergéométriques*, $\mathcal{PH}(N; N_1, \dots, N_r)$. Les entiers N_j représentent les tailles des populations de nature différente cohabitant (boules rouges, boules blanches, boules vertes, *etc.*). L'ensemble des possibles est l'ensemble fini

$$\Omega := \{(k_1, \dots, k_r) \in \mathbb{N}^r; k_j \leq N_j, j = 1, \dots, r, \text{ et } k_1 + \dots + k_r = N\},$$

et la distribution de probabilité sur cet ensemble Ω est définie par

$$P(\{(k_1, \dots, k_r)\}) = \frac{\prod_{j=1}^r \binom{N_j}{k_j}}{\binom{N_1 + \dots + N_r}{N}} \quad \forall (k_1, \dots, k_r) \in \Omega.$$

1.3. Mesure de probabilité sur un ensemble dénombrable

Il est de nombreux modèles d'épreuve dont le résultat aléatoire peut balayer tout un ensemble dénombrable (donc en bijection avec \mathbb{N}), et non plus seulement un ensemble fini. On est donc amené à définir ce qu'est une mesure de probabilité

sur un tel ensemble (que, pour simplifier, on considèrera bien souvent comme \mathbb{N}). Il suffit pour cela d'étendre les règles (1.1) de manière à intégrer le passage du fini au dénombrable. Se donner une probabilité sur \mathbb{N} revient à se donner une application

$$P : A \in \mathcal{P}(\Omega) \rightarrow [0, 1]$$

de manière à ce que

$$(1.10) \quad \begin{aligned} & \forall A_1, \dots, A_k, \dots \in \mathcal{P}(\Omega), \quad (A_k \cap A_l = \emptyset \quad \forall k \neq l) \\ & \implies P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \\ & P(\Omega) = 1. \end{aligned}$$

Aux règles de calcul (1.2) toujours valables ici, il convient d'ajouter ce que l'on appellera le *principe des probabilités totales* : si $(A_k)_{k \in \mathbb{N}^*}$ est une suite de sous-ensembles de $\Omega \simeq \mathbb{N}$ deux à deux disjoints et telle que $\bigcup_{k \geq 1} A_k = \Omega$ (on dit que les A_k , $k \in \mathbb{N}^*$, partitionnent Ω), alors

$$(1.11) \quad \forall A \in \mathcal{P}(\Omega), \quad P(A) = \sum_{k=1}^{\infty} P(A \cap A_k)$$

(ceci restant évidemment valable lorsqu'il n'y a qu'un nombre fini de A_k et non plus une infinité dénombrable).

Se donner une probabilité sur \mathbb{N} revient donc à se donner une suite $[p_0, p_1, \dots, p_k, \dots]$ de nombres dans $[0, 1]$ tels que $\sum_{k \geq 0} p_k = 1$. On pose ensuite $P(\{k\}) = p_k$ et

$$P(A) = \sum_{k \in A} p_k \quad \forall A \subset \mathbb{N}.$$

Une loi de probabilité sur un ensemble Ω fini ou dénombrable est appelée *loi (de probabilité) discrète*. Nous en donnons deux exemples importants dans les sous-sections suivantes.

1.3.1. La loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ . L'épreuve consistant à marquer N points sur un intervalle temporel $[0, T]$, envisagée dans la sous-section 1.2.3 (conduisant à l'approximation (1.5) de la loi binomiale) comme modèle d'épreuve régi par la loi binomiale, peut être envisagée aussi sous l'angle d'un *processus d'émission de particules* depuis un tissu (ici la droite), la densité d'émission étant égale à un nombre $\delta > 0$. Dans ce cas, on peut interpréter δ comme N/T , lorsque N et T sont tous les deux asymptotiquement grands, auquel cas on peut considérer que l'on se trouve dans la situation où N et T tendent simultanément vers $+\infty$, avec $N/T \simeq \delta$. Sous ces hypothèses, le nombre $N([t_1, t_2])$ de particules émises depuis un intervalle $[t_1, t_2] \subset [0, T]$ se trouve régi, on l'a vu précisément dans la sous-section 1.2.3, par la loi

$$\begin{aligned} P(\{N([t_1, t_2]) = k\}) &= \binom{N}{k} \left(\frac{t_2 - t_1}{T}\right)^k \left(1 - \frac{t_2 - t_1}{T}\right)^{N-k} \simeq \\ &\simeq \exp\left(-N \frac{t_2 - t_1}{T}\right) \frac{1}{k!} \left(\frac{(t_2 - t_1)N}{T}\right)^k = e^{-\delta(t_2 - t_1)} \frac{(\delta(t_2 - t_1))^k}{k!}. \end{aligned}$$

En posant $\lambda([t_1, t_2]) = \delta(t_2 - t_1)$, on voit que le nombre de particules émis depuis l'intervalle $[t_1, t_2]$ est régi par une loi ne dépendant que de la longueur $t_2 - t_1$ de l'intervalle, avec

$$P(\{k\}) = e^{-\lambda([t_1, t_2])} \frac{(\lambda([t_1, t_2]))^k}{k!}.$$

Ceci nous amène naturellement à introduire la définition suivante.

DÉFINITION 1.2 (loi de Poisson, ou des évènements rares). La *loi de Poisson* $\mathcal{P}(\lambda)$ de paramètre λ (ou encore loi des évènements rares d'occurrence moyenne λ) est la probabilité définie sur \mathbb{N} par

$$(1.12) \quad p_k = e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}.$$

REMARQUE 1.3 (processus de Poisson et *renouvellement*). La terminologie « *loi des évènements rares* » sera justifiée ultérieurement : si l'on dispose d'un stock infini de composants dont la durée de vie moyenne est $1/\lambda$, et que l'on admette qu'un composant défaillant soit immédiatement remplacé par un composant neuf dès qu'il tombe en panne (on a affaire à ce que l'on appelle un *processus de renouvellement*), le nombre de pannes précédant l'instant T suit une loi de Poisson de paramètre λT . De même, dans une population d'atomes, le nombre d'atomes désintégrés pendant le laps de temps $[0, T]$ suit une loi de Poisson de paramètre λT , où $1/\lambda$ figure la durée de vie moyenne d'un atome de la population.

Sous l'environnement **MATLAB**, la simulation de M réalisations d'une épreuve obéissant à une loi de Poisson de paramètre `lambda` s'obtient *via* la routine :

```
>> f=poissrnd(lambda,1,M);
```

Sous **Scilab**, cette simulation s'effectue sous la routine :

```
function [f] = Poisson(1,M,lambda)
```

Nous avons figuré (figure 1.5) l'histogramme correspondant à $M = 10000$ réalisations d'une épreuve suivant une loi de Poisson de paramètre $\lambda = 33$ (nombre moyen d'occurrences égal à 33). On pourra le comparer avec l'histogramme de $M = 10000$ réalisations de la loi binomiale $\mathcal{B}(100, 0.3)$ à laquelle nous le confrontons ici (voir la figure 1.3).

1.3.2. La loi géométrique (ou de Pascal) de paramètre p . L'ensemble $\{0, 1\}^{\mathbb{N}}$ des suites de 0 ou de 1 est en bijection avec $\mathcal{P}(\mathbb{N})$ ou avec \mathbb{R} et n'est donc plus un ensemble dénombrable. L'univers des possibles correspondant aux résultats d'un jeu infini de « pile ou face » (tous les lancers étant heuristiquement indépendants) n'est pas dénombrable. Cependant, pour l'épreuve consistant en le « numéro du premier « pile » » (ou encore « épreuve du premier succès »), l'univers des possibles est \mathbb{N} . La loi de probabilité régissant cette épreuve est donnée par

$$(1.13) \quad p_k = P(\{k\}) = p(1-p)^k, \quad k \in \mathbb{N}.$$

C'est la loi géométrique $\mathcal{G}(p)$ (ou de Pascal) de paramètre p , p représentant ici la probabilité d'obtenir « pile » lors d'un lancer. Sous **MATLAB**, cette loi est simulée sous la routine :

```
>> f=geornd(p,1,M);
```

Sous **Scilab**, la même simulation s'effectue sous la routine :

```
function [f] = Geom(1,M,p)
```

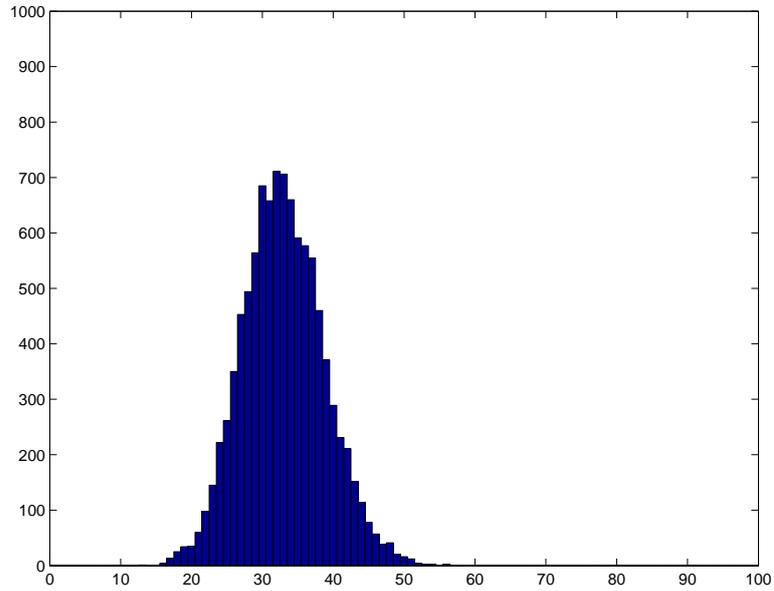


FIGURE 1.5. Histogramme de $M = 10000$ réalisations suivant une loi de Poisson $\mathcal{P}(33)$.

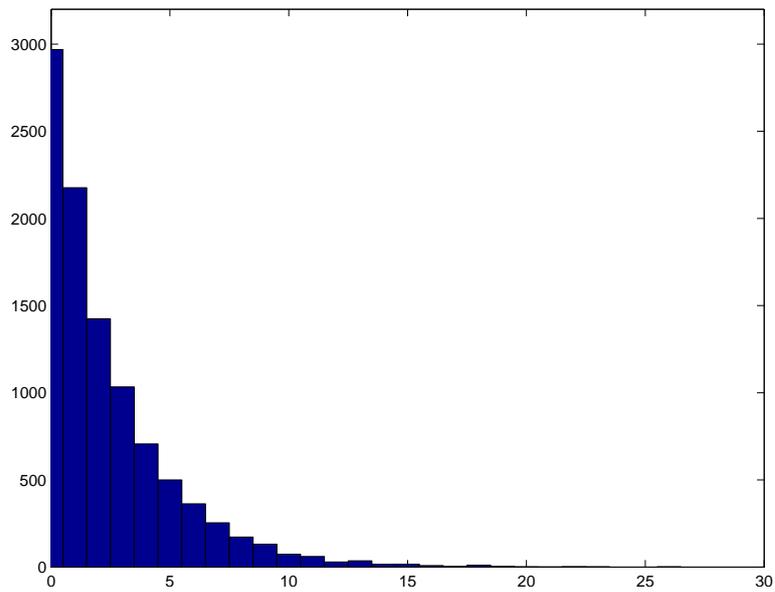


FIGURE 1.6. Histogramme de $M = 10000$ réalisations suivant une loi géométrique de paramètre $p = 0.3$.

Sur la figure 1.6, on a figuré l'histogramme correspondant à $M = 10000$ réalisations d'une épreuve régie par cette loi géométrique $\mathcal{G}(p)$, avec $p = 0.3$.

1.4. Tribus et probabilités sur un ensemble quelconque

1.4.1. Du cadre dénombrable au cadre non dénombrable. La notion de mesure est intimement liée au concept de dénombrable. On ne saurait en effet mesurer n'importe quoi, car pour mesurer, il faut savoir compter ! Or \mathbb{R} , \mathbb{R}^n , $\{0, 1\}^{\mathbb{N}}$ ne sont pas dénombrables (bien que ce soient des candidats à être « univers des possibles » pour des épreuves), d'où une difficulté pour parler (comme dans les sections précédentes) de probabilité, de loi d'une épreuve, *etc.* On contourne ainsi cette difficulté.

DÉFINITION 1.3 (notion de tribu). Une *tribu* (sur un ensemble Ω) est un sous-ensemble de l'ensemble de toutes les parties de Ω , qui

- d'une part contient \emptyset ;
- d'autre part est stable par union dénombrable et passage au complémentaire.

Le couple (Ω, \mathcal{F}) constitué d'un ensemble et d'une tribu sur cet ensemble est ce que l'on appelle un *univers probabilisable* ou *espace probabilisable*.

EXEMPLE 1.2 (tribu de Borel sur \mathbb{R}^n). La plus petite tribu sur \mathbb{R}^n contenant tous les ensembles du type $]a_1, b_1] \times \cdots \times]a_n, b_n]$, avec $-\infty < a_j < b_j < +\infty$ est strictement contenue dans $\mathcal{P}(\mathbb{R}^n)$. On l'appelle *tribu borélienne* sur \mathbb{R}^n et on la note $\mathcal{B}(\mathbb{R}^n)$. Si on lui adjoint tous les ensembles *négligeables*, c'est-à-dire tous les sous-ensembles des boréliens de volume nul, on obtient la *tribu de Lebesgue* $\mathcal{L}(\mathbb{R}^n)$. Cette tribu, même plus grosse que $\mathcal{B}(\mathbb{R}^n)$, n'en est pas moins différente de la tribu $\mathcal{P}(\mathbb{R}^n)$ de toutes les parties de \mathbb{R}^n : il existe des sous-ensembles « pathologiques » de \mathbb{R}^n auxquels on ne saurait attribuer de volume n -dimensionnel (voir le célèbre et curieux paradoxe de Banach-Tarski³).

EXEMPLE 1.3 (tribu produit sur $\{0, 1\}^{\mathbb{N}}$ ou sur $E^{\mathbb{N}}$, E fini). La plus petite tribu sur l'ensemble $\{0, 1\}^{\mathbb{N}}$ contenant tous les produits $A_1 \times \cdots \times A_k \times \cdots$, où tous les $A_k \subset \{0, 1\}$, sauf un nombre fini, sont égaux à $\{0, 1\}$, est une tribu sur $\{0, 1\}^{\mathbb{N}}$ qui est strictement contenue dans la tribu de toutes les parties de $\{0, 1\}^{\mathbb{N}}$. On l'appelle *tribu produit sur $\{0, 1\}^{\mathbb{N}}$* . Ceci se généralise au cas où $\{0, 1\}$ est remplacé par un ensemble fini E de cardinal $N > 2$.

Un tel ensemble Ω s'interprète comme l'univers des possibles d'une épreuve. Reste bien sûr à préciser l'épreuve et à associer à l'épreuve un univers des possibles, ce qui n'est pas toujours chose facile (*cf.* le paradoxe de Bertrand dans la section 1.1). Les modèles $\Omega = \mathbb{R}^n$ (épreuve prenant ses valeurs dans \mathbb{R} ou \mathbb{R}^n), $\Omega = \{0, 1\}^{\mathbb{N}}$ (épreuve : résultats d'une partie infinie de « pile ou face »), ou bien $\Omega = E^{\mathbb{N}}$ (épreuve : marche aléatoire sur un graphe orienté à N sommets) sont les incarnations les plus fréquentes de cette association *épreuve* \rightarrow *univers des possibles*.

Dans chaque cas, il convient de choisir la tribu : la tribu $\mathcal{P}(\Omega)$ si Ω est fini ou dénombrable, la tribu $\mathcal{B}(\mathbb{R}^n)$ ou $\mathcal{L}(\mathbb{R}^n)$ si $\Omega = \mathbb{R}^n$, la tribu produit si $\Omega = \{0, 1\}^{\mathbb{N}}$ ou $\Omega = E^{\mathbb{N}}$ avec E fini. Les éléments de la tribu sont les *événements aléatoires* attachés à l'épreuve.

Reste enfin à fixer la règle qui « préside » au hasard, c'est-à-dire à fixer une *loi de probabilité* sur cette tribu. Ceci suppose, si Ω n'est pas fini ou dénombrable (par exemple dans les cas $\Omega = \mathbb{R}^n$, $\Omega = \{0, 1\}^{\mathbb{N}}$, $\Omega = E^{\mathbb{N}}$, *etc.*) que cette tribu ne saurait être la tribu de toutes les parties, mais au pire la tribu de Lebesgue $\mathcal{L}(\mathbb{R}^n)$, la

3. Voir, si vous êtes curieux, http://fr.wikipedia.org/wiki/Paradoxe_de_Banach-Tarski

tribu produit, *etc.*) Les singletons (dont on assure en général qu'ils font partie de la tribu, c'est le cas pour la tribu borélienne sur \mathbb{R}^n , pour la tribu produit sur $\{0, 1\}^{\mathbb{N}}$ ou $E^{\mathbb{N}}$, *etc.*) sont les *événements aléatoires élémentaires*.

DÉFINITION 1.4 (espace probabilisé). Etant donné un ensemble Ω et une tribu \mathcal{T} sur Ω , se donner une probabilité sur (Ω, \mathcal{T}) revient à se donner une application

$$A \in \mathcal{T} \rightarrow P(A)$$

de manière à ce que les clauses (1.10) soient remplies. La donnée de Ω, \mathcal{T}, P est ce que l'on appelle la donnée d'un *univers probabilisé* (ou *espace probabilisé*) (Ω, \mathcal{T}, P) .

REMARQUE 1.4. Toutes les règles de calcul (1.2) restent valables, ainsi que le principe des probabilités totales (1.11) lorsque $\bigcup_{k=1}^{\infty} A_k = \Omega$.

1.4.2. Le cas de $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$; notion de loi à densité sur \mathbb{R}^n . Une probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ définit ce que l'on appelle une *loi (de probabilité) continue*.

Dans le cas particulier de la droite réelle ($n = 1$), bien souvent vue comme incarnation de l'axe des temps t , il est important de définir deux fonctions importantes, la première utile en théorie des probabilités, la seconde en statistique. Nous donnons ici ces deux définitions.

DÉFINITION 1.5 (fonction de répartition d'une loi de probabilité continue sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). La *fonction de répartition* d'une loi de probabilité P sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est par définition la fonction croissante

$$F_P : T \in \mathbb{R} \mapsto P(]-\infty, T]).$$

Du fait des règles (1.10), cette fonction croissante tend vers 0 par valeurs supérieures lorsque t tend vers $-\infty$ et vers 1 par valeurs inférieures lorsque t tend vers $+\infty$.

DÉFINITION 1.6 (fonction de quantile d'une loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). La *fonction de quantile* d'une loi de probabilité P sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est par définition la fonction

$$F_P^{\leftarrow} : u \in]0, 1[\mapsto \inf \{T \in \mathbb{R}; F_P(T) \geq u\},$$

Parmi les lois continues, nous allons dégager ici une classe importante, celles des lois (de probabilité) à *densité*.

À tout borélien de \mathbb{R}^n , on sait associer (de manière unique) un volume n -dimensionnel

$$A \mapsto \text{vol}_n(A) := \int_{\mathbb{R}^n} dx_1 \dots dx_n \in [0, \infty]$$

de manière à ce que $\text{vol}_n([0, 1]^n) = 1$ (condition de normalisation), que l'on ait $\text{vol}_n(A) = \text{vol}_n(x + A)$ pour tout borélien de \mathbb{R}^n , et que la clause (*cf.* (1.10))

(1.14)

$$\forall A_1, \dots, A_k, \dots \in \mathcal{B}(\mathbb{R}^n), \left(A_k \cap A_l = \emptyset \quad \forall k \neq l \right)$$

$$\implies \int_{\bigcup_{k=1}^{\infty} A_k} dx_1 \dots dx_n = \text{vol}_n\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \int_{A_k} dx_1 \dots dx_n = \sum_{k=1}^{\infty} \text{vol}_n(A_k)$$

soit remplie.

Étant donnée une fonction $f : \mathbb{R}^n \rightarrow [0, +\infty[$ telle que $f^{-1}([0, x]) \in \mathcal{B}(\mathbb{R}^n)$ pour tout $x \geq 0$, on peut lui associer une intégrale en posant

$$(1.15) \quad \int_{\mathbb{R}^n} f(x) dx_1 \cdots dx_n := \sup_{h>0} \left(\sum_{k=0}^{\infty} kh \int_{f^{-1}([kh, (k+1)h])} dx_1 \cdots dx_n \right) \in [0, \infty].$$

DÉFINITION 1.7 (densités et loi à densité sur \mathbb{R}^n). Une fonction $f : \mathbb{R}^n \rightarrow [0, \infty[$ telle que $f^{-1}([0, x]) \in \mathcal{B}(\mathbb{R}^n)$ pour tout $x \geq 0$ et que $\int_{\mathbb{R}^n} f(x) dx_1 \cdots dx_n = 1$ est appelée *densité de probabilité* sur \mathbb{R}^n . À toute densité de probabilité f sur \mathbb{R}^n , on peut associer une probabilité P_f sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ en posant

$$(1.16) \quad \forall A \in \mathcal{B}(\mathbb{R}^n), \quad P_f(A) := \int_{\mathbb{R}^n} f(x) dx_1 \cdots dx_n.$$

Cette probabilité P_f est dite *loi de probabilité à densité f* sur \mathbb{R}^n . Si P est une probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ telle que $P = P_f$ pour une densité $f : \mathbb{R}^n \rightarrow [0, \infty[$ (nécessairement unique), on dit que la probabilité P_f est une *loi à densité*.

REMARQUE 1.5. Si $P = P_f$ est une loi à densité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, alors on a $P_f(\{x_0\}) = 0$ pour tout $x_0 \in \mathbb{R}^n$, puisque le volume n -dimensionnel de tout singleton $\{x_0\}$ est nul. Il faut donc prendre garde au fait que $f(x_0)$ ne représente en aucune manière $P_f(\{x_0\})$ (la probabilité de ce singleton valant toujours 0). Par contre, si f est continue en un point $x_0 \in \mathbb{R}^n$, on a

$$P_f(x_0 + [-h/2, h/2]^n) \sim h^n f(x_0) \quad \text{lorsque } h \rightarrow 0_+.$$

Si $n = 1$ et $P = P_f$, la fonction de répartition de $P = P_f$ est

$$F_{P_f} : T \in \mathbb{R} \mapsto \int_{-\infty}^T f(t) dt.$$

C'est, fait important, une fonction continue⁴ de T en vertu de ce que l'on appelle le *théorème de convergence monotone en théorie de l'intégration* (admis ici). Si la densité $t \mapsto f(t)$ est continue sur \mathbb{R} , la fonction de répartition F_{P_f} est dérivable et l'on a $F'_{P_f} \equiv f$ sur \mathbb{R} (ceci résulte du *théorème fondamental de l'analyse* vu en Terminale). Si l'ensemble $f^{-1}(]0, +\infty[)$ est un intervalle ouvert I de \mathbb{R} sur lequel f est continue, la fonction de répartition F_{P_f} est continue et strictement croissante sur cet intervalle et la fonction de quantile $F_{P_f}^{\leftarrow}$ est définie par

$$(1.17) \quad \left(u \in]0, 1[\text{ et } T = F_{P_f}^{\leftarrow}(u) \right) \iff \left(T \in I \text{ et } u = F_{P_f}(T) \right),$$

autrement dit $F_{P_f}^{\leftarrow}$ est l'inverse de la restriction de la fonction F_{P_f} à l'intervalle I . Son graphe s'obtient donc en prenant le symétrique du graphe de F_{P_f} (au dessus de l'intervalle I) par rapport à la première bissectrice $u = T$ dans le plan des (T, u) (voir la figure 1.7).

1.4.3. La loi uniforme et la routine rand. La loi uniforme discrète $\mathcal{U}(N)$ (« loi des pires risques », cf. la section 1.2.2) existe aussi dans sa version continue.

4. Ceci n'est bien sûr pas le cas pour toute loi de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, prendre une loi discrète par exemple.

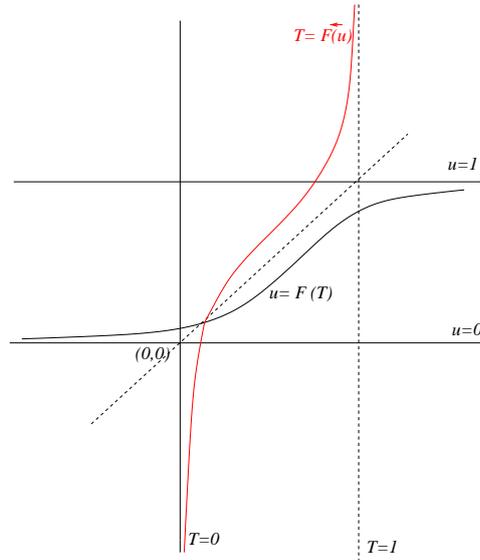


FIGURE 1.7. Fonction de répartition et fonction de quantile

DÉFINITION 1.8 (loi uniforme continue sur un ouvert borné de \mathbb{R}^n). Soit U un ouvert borné de \mathbb{R}^n . La loi uniforme continue sur U est par définition la loi de probabilité à densité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, ayant pour densité la fonction

$$(x_1, \dots, x_n) \mapsto \begin{cases} 1 & \text{si } x \in U \\ 0 & \text{si } x \notin U. \end{cases}$$

On a, pour cette loi $P_{U,\text{unif}}$

$$P_{U,\text{unif}}(A) = \frac{\int_{A \cap U} dx_1 \dots dx_n}{\int_U dx_1 \dots dx_n} = \frac{\text{vol}_n(A \cap U)}{\text{vol}_n(U)} \quad \forall U \in \mathcal{B}(\mathbb{R}^n).$$

EXEMPLE 1.4 (la routine `rand`). Lorsque $n = 1$ et $U =]0, 1[$, la routine `rand` simule sous MATLAB ou Scilab la loi uniforme continue sur $]0, 1[$. Par exemple

```
>> f = rand(1,M);
```

génère (en ligne), M réalisations successives de l'épreuve consistant à prendre au hasard un nombre dans $]0, 1[$, tous les choix étant en un certain sens équiprobables. Bien sûr, il s'agit du point de vue informatique de réalisations d'une loi uniforme discrète sur l'ensemble des nombres exprimables en virgule flottante dans la machine entre 0 (exclus) et 1 (inclus). Grâce à la routine

```
>> f = rand (n,M);
```

on génère M réalisations (sous forme chacune d'un vecteur colonne à n lignes) de la loi uniforme sur $]0, 1[^n$. Pour simuler la loi uniforme sur le disque ouvert $D(0, 1)$ (lorsque $n = 2$) ou sur la boule ouverte $B(0, 1)$ (lorsque $n = 3$), il faut se ramener à une simulation de loi uniforme sur $]0, 1[\times]0, 2\pi[$ ou $]0, 1[\times]0, 2\pi[\times]0, \pi[$ en utilisant les coordonnées polaires (lorsque $n = 2$) ou sphériques (lorsque $n = 3$). Notons aussi que l'on ne change rien à la loi lorsque l'on remplace l'ouvert borné U par

le compact $K = \overline{U}$ puisque la frontière de tout ouvert borné de \mathbb{R}^n est de volume n -dimensionnel nul.

REMARQUE 1.6 (loi uniforme sur arcs de courbes de longueur finie ou morceaux de surface d'aire finie). En paramétrant un arc de courbe plane de longueur finie (par exemple le cercle unité) ou un morceau de surface d'aire finie dans \mathbb{R}^3 (par exemple la sphère unité), on introduit et on simule grâce à **rand** des lois uniformes continues sur un arc de courbe de longueur finie, un morceau de surface d'aire finie, etc.

Pour les lois de probabilité sur \mathbb{R} , on a l'intéressant résultat suivant :

PROPOSITION 1.1. *Soit X une épreuve aléatoire à valeurs réelles, dont le résultat se plie à une loi de probabilité à densité P_f sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On suppose que $f^{-1}(]0, \infty[)$ est un intervalle ouvert I de \mathbb{R} sur lequel f est continue. Alors on réalise une simulation de l'épreuve X en prenant $F_{P_f}^{\leftarrow}(Y)$, où Y est une simulation de la loi uniforme sur $]0, 1[$.*

REMARQUE 1.7. C'est par le biais de ce résultat (en particulier de la dernière assertion) que l'on sera capable de simuler une loi à densité P_f sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ en simulant une loi uniforme sur $]0, 1[$ (routine **rand**), puis en composant le résultat de cette simulation avec la fonction de quantile $F_{P_f}^{\leftarrow}$ de P_f (si tant est que cette fonction de quantile soit aisément explicitable).

DÉMONSTRATION. On sait que la fonction de quantile $F_{P_f}^{\leftarrow}$ est, sous les hypothèses faites ici sur la densité f , l'inverse de la restriction de F_{P_f} à l'intervalle I (cf. (1.17)). On a donc, pour tout $T \in]0, 1[$,

$$P_f(\{F_{P_f}(X) \leq T\}) = P_f(\{X \leq F_{P_f}^{\leftarrow}(T)\}) = F_{P_f}(F_{P_f}^{\leftarrow}(T)) = T.$$

La fonction de répartition de la loi commandant le résultat de l'épreuve $F_{P_f}(X)$ est bien celle de la loi uniforme sur $]0, 1[$. \square

1.4.4. Les lois normales $\mathcal{N}(\mu, \sigma^2)$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. L'approximation (1.8) de la loi binomiale $\mathcal{B}(N, p)$ lorsque N est grand et $Np(1-p)$ pas trop petit (par exemple $N > 30$ et $Np(1-p) > 9$ ou $N > 30$ et $\min(Np, N(1-p)) > 5$) fait apparaître une loi à densité particulièrement importante sur \mathbb{R} , la loi à densité $\mathcal{N}(\mu, \sigma^2)$ (où $\mu \in \mathbb{R}$ et $\sigma^2 > 0$), de densité

$$g_{\mu, \sigma^2}(t) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

Cette loi est appelée *loi normale de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$* ⁵. Dans le cas particulier $\mu = 0$ et $\sigma^2 = 1$, cette loi à densité s'appelle *loi normale centrée réduite*.

La *fonction de Gauß*

$$t \in \mathbb{R} \mapsto g_{0,1}(t) := \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

(densité de la loi $\mathcal{N}(0, 1)$) joue un rôle très important tant en mathématiques qu'en physique pour plusieurs raisons : une explication « physique » tient au fait que cette fonction soit vecteur propre (de valeur propre $\sqrt{2\pi}$) de la transformation qui à une

5. Nous avons adopté la convention (anglosaxonne) maintenant la plus communément répandue en dénotant la loi $\mathcal{N}(\mu, \sigma^2)$ et non, comme c'était parfois le cas en France, $\mathcal{N}(\mu, \sigma)$.

fonction $\varphi : \mathbb{R} \mapsto \mathbb{C}$ définie dans l'espace des temps associe sa transformée de Fourier (ou spectre) définie dans l'espace \mathbb{R} des fréquences cette fois par

$$\forall \omega \in \mathbb{R}, \hat{\varphi}(\omega) = \int_{\mathbb{R}} e^{-i\omega t} \varphi(t) dt.$$

À ce titre, on retrouve la fonction de Gauß au cœur de la modélisation mathématique en physique théorique (mécanique quantique, modélisation des particules), en traitement de l'information, en théorie du signal ou de l'image (modélisation des impulsions ou des pixels). Le fait que la loi $\mathcal{N}(\mu, \sigma^2)$, avec $\mu > 0$, apparaisse comme « loi limite » (sous des conditions à préciser) de la loi binomiale de paramètres N et p tels que $Np \simeq \mu$ et $\sigma^2 \simeq Np(1-p)$, *i.e.* $1-p = \sigma^2/\mu$ et $N \simeq \mu/p$ n'est pas étranger à cette propriété physique importante concernant la transformation de Fourier.

Le graphe de la fonction g_{μ, σ^2} présente une symétrie par rapport à la droite verticale $t = m$. Il se déduit du graphe de la fonction $g_{0,1}$ de la manière suivante :

- translation de μ , suivie d'une homothétie de rapport σ suivant l'axe des temps (axe réel) ;
- homothétie de rapport $1/\sigma$ suivant l'axe vertical (la surface comprise entre l'axe réel et le graphe devant rester constante et égale à 1).

Plus σ augmente, plus le lobe se voit « étiré » suivant la direction horizontale et donc « affaissé » ; plus σ diminue, plus ce lobe se voit en revanche « comprimé » suivant la direction horizontale, ce qui a pour conséquence que le maximum (au point μ) augmente. On constate que la fonction g_{μ, σ^2} tend de manière certes très rapide (en $(1/(\sigma\sqrt{2\pi})) \exp(-t^2/2\sigma^2)$) vers 0_+ lorsque t tend vers $\pm\infty$, mais on ne peut en tirer profit que que si σ n'est pas trop petit.

La simulation (répétée M fois) sous MATLAB de la loi normale de paramètres μ et σ est fournie par

```
>> f = normrnd (mu, sigma, 1, M);
```

Sous Scilab, la routine `grand (1, M, 'norm', mu, sigma)` génère (comme vecteur ligne) M réalisations suivant une loi $\mathcal{N}(\mu, \sigma^2)$. Sur les figures 1.8 et 1.9, nous avons représenté (en haut) le graphe de la densité de la loi et (en bas) l'histogramme des $M = 10000$ résultats obtenus, classés sur $[-15, 15]$ avec un pas $h = 0.1$.

La fonction de répartition de la loi normale centrée ne s'exprime pas par une formule simple. On doit donc se référer à une table (voir par exemple l'Annexe 3 de [MathAp], tables 1,2,3) pour disposer d'une connaissance de $F_{\mathcal{N}(0,1)}$ et de $F_{\mathcal{N}(0,1)}^{\leftarrow}$ et être (par exemple) à même de réaliser une simulation à partir de la commande `rand` (comme suggéré dans la remarque 1.7). Signalons que cette simulation est précisément celle que réalise la *planche de Galton*, que vous avez peut être pu voir au Palais de la Découverte : un stock de billes tombe dans un « flipper » vertical (comme sur la figure 1.10), la règle étant que, lorsqu'elle bute sur un plot, la bille tombe à gauche ou à droite avec la probabilité $1/2$. Le nombre de bandes horizontales du flipper est N et l'on modélise ainsi une épreuve régie par une loi binomiale $\mathcal{B}(N, 1/2)$ (recentrée entre $-[N/2]$ et $[N/2]$ plutôt qu'entre 0 et N). Dès que N est grand, les conditions d'approximation (1.8) sont remplies et l'histogramme figuré par le stock de billes organisé en colonnes verticales correspond à celui d'une épreuve régie par la loi $\mathcal{N}(0, N/4)$.

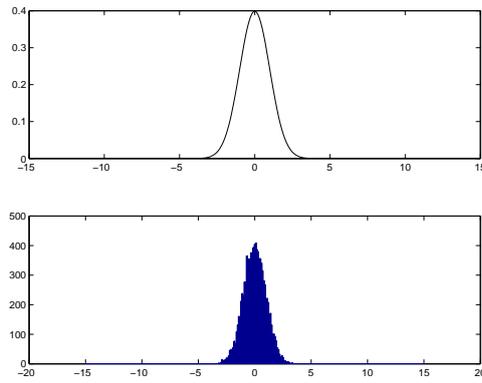


FIGURE 1.8. Loi normale $\mathcal{N}(0, 1)$

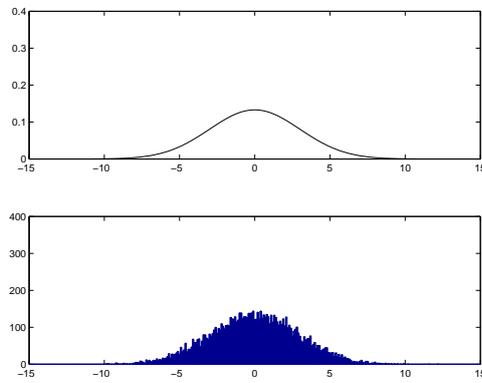


FIGURE 1.9. Loi normale $\mathcal{N}(0, 3)$

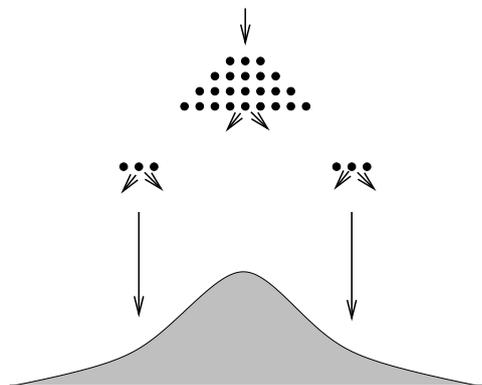


FIGURE 1.10. Planche de Galton

1.4.5. Les lois normales en dimension supérieure. Dans ce cours, nous mentionnerons peu les exemples de lois à densité en dimension $n > 1$. Cependant, comme les *vecteurs gaussiens* jouent un rôle important en traitement (simultané et comparé) de données, en analyse statistique, en traitement d'image (pixels), *etc.*, nous mentionnerons ici l'exemple des lois normales en dimension supérieure.

Soit Σ^2 une matrice réelle symétrique définie positive de taille (n, n) et μ un vecteur de \mathbb{R}^n . On définit la loi $\mathcal{N}(\mu, \Sigma^2)$ sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ comme la loi à densité :

$$f_{\mu, \Sigma} : x = (x_1, \dots, x_n) \mapsto (2\pi)^{-n/2} \frac{1}{\sqrt{\det \Sigma^2}} \exp\left(-\frac{1}{2} {}^t(x - \mu) \cdot [\Sigma^2]^{-1} \cdot (x - \mu)\right).$$

La matrice Σ^2 , en temps que matrice réelle symétrique, se diagonalise sous la forme

$$\Sigma^2 = P \cdot \text{diag}[\sigma_1^2, \dots, \sigma_n^2] \cdot {}^t P,$$

où P est une matrice orthogonale réelle (*i.e.* dont les vecteurs colonne forment une base orthonormée). Cette décomposition est donnée sous MATLAB (par exemple) sous la forme

```
>> [U,D,V] = svd (Sigma2);
```

Elle fournit $U = P$ et $V = {}^t P$, les σ_j^2 , $j = 1, \dots, n$ (valeurs propres de Σ^2 rangés dans l'ordre décroissant) étant les termes diagonaux de la matrice diagonale D ainsi obtenue. Une fois cette matrice P connue, voici comment il est possible de simuler une loi normale $\mathcal{N}(\mu, \Sigma^2)$. Si $\mu = (\mu_1, \dots, \mu_n)$ et $(\sigma_1, \dots, \sigma_n)$ sont les racines des valeurs propres de D (donc de Σ^2), on a la routine :

```
>> X = normrnd (mu, [sigma_1 sigma_2 ... sigma_n], 1, M);
>> XX = P*X;
```

Dans le cas $n = 2$, nous avons figuré sur la figure 1.11 l'histogramme correspondant à 10000 réalisations suivant la loi $\mathcal{N}(\mu, \Sigma^2)$ lorsque $\mu = (20, 10)$ et Σ^2 est la matrice `diag[64 9]`.

1.4.6. Loi exponentielle $\mathcal{E}(\lambda)$ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. La loi exponentielle $\mathcal{E}(\lambda)$, où $\lambda > 0$, est la version continue de la loi géométrique (ou de Pascal) $\mathcal{G}(p)$ vue à sous-section 1.3.2 (penser à la relation $1 - p = e^{-\lambda}$, *i.e.* $\lambda = -\log(1 - p) \simeq p$ au voisinage de $p = 0$). Cette loi est en effet la loi à densité

$$f_{\mathcal{E}(\lambda)} : t \in \mathbb{R} \mapsto \lambda \chi_{]0, \infty[}(t) e^{-\lambda t}.$$

Si I est un intervalle de \mathbb{R} , on a donc

$$P_{\mathcal{E}(\lambda)}(I) = \lambda \int_{I \cap]0, \infty[} e^{-\lambda t} dt.$$

La fonction de répartition est ici aisée à calculer. On trouve

$$(1.18) \quad F_{\mathcal{E}(\lambda)}(T) = - \int_{-\infty}^T d[\exp(-\lambda t)] = \chi_{]0, \infty[}(T) (1 - \exp(-\lambda T)).$$

La fonction de quantile est donc la fonction inverse de la restriction de cette fonction à $I =]0, \infty[$, soit :

$$F_{\mathcal{E}(\lambda)}^{\leftarrow} : u \in]0, 1[\mapsto -\frac{\log(1 - u)}{\lambda}.$$

Pour simuler une loi exponentielle $\mathcal{E}(\lambda)$, la remarque 1.7 nous indique comment procéder. Sous MATLAB ou Scilab, la routine

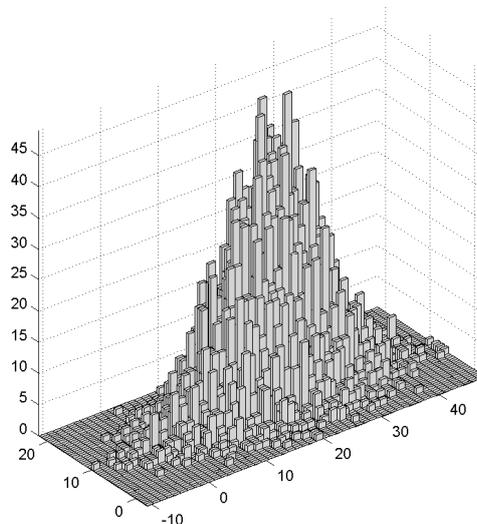


FIGURE 1.11. Histogramme des résultats d'une épreuve (X, Y) suivant une loi $\mathcal{N}((20, 10), \text{diag}[64, 9])$.

```
>> f_aux = rand(1,M);
>> f = - (log (1-f_aux))/lambda ;
```

fournit (en ligne) M réalisations d'une épreuve régie par la loi $\mathcal{E}(\lambda)$. Sur la figure 1.12, nous avons représenté l'histogramme synthétisant les résultats de $M = 10000$ réalisations d'une épreuve régie par une loi $\mathcal{E}(1/5)$. Le pas utilisé pour le tracé de l'histogramme est ici 0.5 (on note que $0.5/5 \times 10000 = 1000$, ce qui est à peu près en accord avec la valeur maximale⁶ constatée sur cet histogramme, le choix d'un pas $h < 0.5$ permettant d'affiner le résultat).

La loi exponentielle modélise le hasard présidant à la durée de vie de composants et joue donc un rôle important, en relation d'ailleurs avec la loi de Poisson ou des évènements rares $\mathcal{P}(\lambda)$, dans l'étude de questions de *fiabilité*. On y reviendra une fois introduite la notion de conditionnement.

1.5. Conditionnement et indépendance

1.5.1. Probabilité trace et probabilité conditionnelle. Soit (Ω, \mathcal{T}, P) un univers probabilisé et $A \in \mathcal{T}$ un évènement aléatoire de probabilité non nulle. Nous allons attacher deux espaces probabilisés à cette configuration (Ω, A) , déduits de l'univers probabilisé (Ω, \mathcal{T}, P) .

DÉFINITION 1.9 (trace d'une tribu sur un évènement et probabilité trace). Soit (Ω, \mathcal{T}, P) un univers probabilisé et $A \in \mathcal{T}$ un évènement aléatoire. La collection des $A \cap E$, où $E \in \mathcal{T}$, définit une tribu \mathcal{T}_A sur A , dite *trace de la tribu \mathcal{T} sur*

⁶. Elle aurait du valoir 1000, en accord avec le fait que la densité tend vers $1/5$ lorsque l'on tend vers 0_+ .

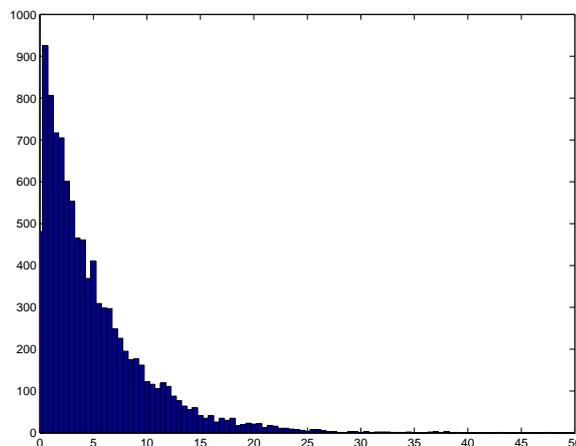


FIGURE 1.12. Histogramme de $M = 10000$ réalisations sous loi géométrique $\mathcal{E}(1/5)$.

l'évènement aléatoire A . Si de plus $P(A) > 0$, l'application

$$B \in \mathcal{T}_A \mapsto P_A(B) := \frac{P(B)}{P(A)}$$

définit une probabilité sur (A, \mathcal{T}_A) , dite *probabilité induite par P sur A* ou *probabilité trace de P sur A* .

La probabilité trace de P sur A dépend bien sûr de A au sens où c'est une probabilité sur A . Il est important de définir une probabilité qui dépend encore de A , mais soit cette fois une probabilité sur l'univers probabilisé (Ω, \mathcal{T}) . De fait, la restriction à \mathcal{T}_A de la probabilité sur (Ω, \mathcal{T}) que nous allons construire sera précisément la probabilité trace P_A .

DÉFINITION 1.10 (conditionnement par un évènement aléatoire de probabilité non nulle). Soit (Ω, \mathcal{T}, P) un univers probabilisé et $A \in \mathcal{T}$ un évènement aléatoire de probabilité non nulle. L'application

$$(1.19) \quad B \in \mathcal{T} \mapsto \frac{P(A \cap B)}{P(A)} = P(B | A)$$

définit une nouvelle probabilité sur Ω , dite *probabilité conditionnelle (conditionnée⁷ ici par l'évènement aléatoire A)*.

EXEMPLE 1.5 (un exemple sur un ensemble fini (cf. l'approche des probabilités au lycée)). Supposons que l'univers des possibles soit l'ensemble de toutes les possibilités de choisir 8 cartes dans un jeu de 32 cartes. Il s'agit donc d'un ensemble de cardinal $\binom{32}{8}$ que nous supposons ici équipé de la loi uniforme (tous les tirages sont équiprobables). L'évènement B : « on tire exactement deux valets » a pour probabilité $P(B) = \binom{4}{2} \times \binom{28}{6}$ (au numérateur, on a dénombré les choix possibles : choisir les deux valets parmi 4 dans un premier temps, puis les 6 cartes restantes

7. On dit aussi parfois « probabilité de B quand A » pour $P(B | A)$.

parmi les 28 autres, une fois les 4 valets retirés). L'évènement $A \supset B$: « on tire au moins un valet » a pour probabilité $P(A) = 1 - \binom{28}{8} / \binom{32}{8}$ (il est en effet plus aisé de calculer la probabilité de $\Omega \setminus A$). On a donc ici

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)}{P(A)} = P_A(B) = \frac{\binom{4}{2} \binom{28}{6}}{\binom{32}{8} - \binom{28}{8}} \simeq 0.30504.$$

Si A_1, \dots, A_N sont N évènements d'un univers probabilisé (Ω, \mathcal{F}, P) tels que l'on ait $P(A_1 \cap \dots \cap A_N) > 0$, on prouve par récurrence la formule suivante :

$$(1.20) \quad P(A_1 \cap \dots \cap A_N) = P(A_1)P(A_2 | A_1) \cdots P(A_N | A_1 \cap \dots \cap A_{N-1}).$$

L'autre règle de calcul importante impliquant le conditionnement est la manière dont il intervient lorsque couplé à la formule de probabilités totales (1.11). C'est la *formule de Bayes*.

PROPOSITION 1.2 (formule de Bayes). *Soient (Ω, \mathcal{F}, P) un espace probabilisé et $(A_k)_{k \geq 1}$ une suite d'évènements aléatoires ($A_k \in \mathcal{F}$ pour tout k) deux à deux disjoints, tous de probabilité strictement positive, et tels que*

$$P(\Omega \setminus \bigcup_{k=1}^{\infty} A_k) = 0.$$

Alors, pour tout évènement aléatoire $B \in \mathcal{F}$, on a

$$(1.21) \quad P(B) = \sum_{k=1}^{\infty} P(B | A_k) P(A_k).$$

REMARQUE 1.8 (système complet d'évènements aléatoires). Une suite (finie ou infinie) d'évènements aléatoires $(A_k)_{k \geq 1}$ dont l'union est de probabilité 1 est dite *système complet d'évènements* dans l'univers probabilisé (Ω, \mathcal{F}, P) .

DÉMONSTRATION. On applique la formule des probabilités totales (1.11) qui assure ici

$$P(B) = \sum_{k=1}^{\infty} P(B \cap A_k) = \sum_{k=1}^{\infty} P(B | A_k) P(A_k)$$

d'après la définition (1.19) des $P(B | A_k)$. □

Une autre manière de tourner les choses est la formule dite *formule des causes*, plus importante encore dans la pratique que la formule de Bayes (1.21).

PROPOSITION 1.3 (formule des causes ou des preuves). *Soit (Ω, \mathcal{F}, P) un univers probabilisé et $(B_k)_{k \geq 1}$ un système complet d'évènements aléatoires. On suppose tous les B_k de probabilité strictement positive. Alors, pour tout évènement aléatoire A de probabilité strictement positive, pour tout entier $k = 1, 2, \dots$, on a*

$$(1.22) \quad P(B_k | A) = \frac{P(A | B_k)P(B_k)}{\sum_{j=1}^{\infty} P(A | B_j)P(B_j)} \quad \forall k = 1, 2, \dots$$

DÉMONSTRATION. C'est une conséquence immédiate de la Proposition 1.2 : le dénominateur au second membre de (1.22) est en effet égal à $P(A)$ d'après cette formule. □

Voici l'interprétation de la formule des causes (ou des preuves), ce qui justifie au passage l'une ou l'autre des terminologies utilisées, et indique comment on utilise pareille formule concrètement. Si A est un événement aléatoire de probabilité strictement positive et si les B_k , $k = 1, \dots, N$, constituent un jeu d'hypothèses, le problème que l'on se pose est celui de savoir si la réalisation de A confirme ou infirme chacune des hypothèses B_k , $k = 1, \dots, N$. On connaît les probabilités *a priori* des B_k , $k = 1, \dots, N$, ainsi que les probabilités conditionnelles $P(A | B_k)$, $k = 1, \dots, N$, et l'on souhaite donc calculer les probabilités *a posteriori* $P(B_k | A)$, $k = 1, \dots, N$. C'est ce que permet précisément la formule des causes (1.22). Le fait que $P(B_k | A)$ soit voisin de 1 constitue alors une indication (donc une « preuve » à un certain seuil d'erreur près) de ce que la réalisation de A confirme l'hypothèse B_k ; en revanche, le fait que $P(B_k | A)$ soit voisin de 0 constitue une indication de ce que la réalisation de A infirme l'hypothèse B_k .

EXEMPLE 1.6 (cf. [MathAp], exemple 11.9, chapitre 11). Dans une population, on suppose qu'avec une probabilité p , un individu donné est malade, tandis qu'avec la probabilité $1 - p$ il ne l'est pas (loi de Bernoulli). On soumet chaque individu de la population à un test de dépistage (non totalement fiable) : quand l'individu est malade, le test le confirme seulement avec une probabilité q_1 ; quant l'individu est bien portant, le test l'infirme avec une probabilité q_2 , évidemment très petite par rapport à q_1 . L'univers Ω est ici l'ensemble de la population. Si B_1 désigne l'évènement « un individu donné est malade », on a $P(B_1) = p$ et $P(B_2) = 1 - p$ si $B_2 = \Omega \setminus B_1$. Si l'on désigne par A l'évènement « le test de dépistage s'est révélé positif pour un individu donné », on sait que $P(A | B_1) = q_1$ et que $P(A | B_2) = q_2$. On a donc, grâce à la formule des causes (1.22) :

$$P(B_1 | A) = \frac{q_1 p}{q_1 p + q_2 (1 - p)} = \frac{1}{1 + \frac{q_2 (1 - p)}{q_1 p}}$$

$$P(B_2 | A) = \frac{q_2 (1 - p)}{q_1 p + q_2 (1 - p)} = \frac{1}{1 + \frac{q_1 p}{q_2 (1 - p)}}.$$

C'est la position du nombre

$$\epsilon = \frac{q_2}{q_1} \times \frac{1 - p}{p}$$

dans $]0, \infty[$ qui indique quelles sont les chances d'effrayer inutilement un individu ayant réagi au test en lui annonçant qu'il est effectivement malade. En effet, on a dans ce cas $P(B_1 | A) = 1/(1 + \epsilon)$ tandis que $P(B_2 | A) = \epsilon/(1 + \epsilon)$.

1.5.2. Conditionnement et marches aléatoires sur les graphes orientés.

Comme vous l'avez vu certainement au lycée (à l'occasion d'exercices, sans doute aussi au bac), la notion de conditionnement est souvent en étroite relation avec la théorie des graphes.

DÉFINITION 1.11 (graphe orienté, matrice d'adjacence, matrice d'adjacence pondérée). On appelle *graphe orienté* $\Gamma = \{E, V\}$ la donnée d'un ensemble fini E de cardinal N (dit ensemble des *sommets* du graphe) et d'un sous-ensemble V de $E \times E$, dit ensembles de *flèches* ou *liens* (le couple (i, j) est dit *flèche* de i vers j , on dit aussi *arête orientée* $i \rightarrow j$). La *matrice d'adjacence* \mathbb{A}^Γ du graphe Γ est la

matrice de taille (N, N) dont le terme général est

$$(1.23) \quad a_{ij}^\Gamma = \begin{cases} 1 & \text{si } (i, j) \in V \\ 0 & \text{si } (i, j) \notin V. \end{cases}$$

La *matrice d'adjacence pondérée* \mathbb{G}^Γ du graphe⁸ Γ est la matrice de taille (N, N) dont le terme général est

$$(1.24) \quad g_{ij}^\Gamma = \begin{cases} \frac{1}{L_i} & \text{si } (i, j) \in V, \text{ où } L_i = \text{card} \{j \in E; (i, j) \in V\} > 0 \\ 0 & \text{si } (i, j) \notin V. \end{cases}$$

EXEMPLE 1.7 (la toile internet). L'exemple de la toile internet est un bon exemple de graphe orienté. Les sommets sont les pages **web**, les flèches sont les liens (une page peut créer un lien sur elle-même). On estime $N = \text{card}(E)$ à environ une trentaine de milliards! C'est donc en référence à **Google** et au célèbre algorithme **Pagerank** introduit par Serguey Brin et Larry Page à Stanford vers 1997 que nous avons noté \mathbb{G}^Γ la matrice d'adjacence pondérée d'un graphe orienté.

REMARQUE 1.9 (matrices stochastiques). Une propriété importante de la matrice d'adjacence pondérée \mathbb{G}^Γ (à condition toutefois que de tout sommet i du graphe parte au moins une arête orientée $i \rightarrow j$, ce que l'on peut toujours supposer en décidant que tous les couples (i, i) sont dans V) est que les entrées sont toutes entre 0 et 1 et que la somme des entrées sur chaque ligne est égale à 1. Chaque ligne $[g_{i1}, \dots, g_{iN}]$ peut donc être interprétée comme une loi de probabilité sur l'ensemble $E \simeq \{1, \dots, N\}$. Nous verrons plus loin que cette loi de probabilité peut être interprétée comme une loi conditionnelle. Une telle matrice est dite *stochastique*.

Soit $\Gamma = \{E, V\}$ un graphe orienté, tel que tout élément de E soit extrémité d'au moins une flèche. Les sommets du graphe sont numérotés de 1 à N . L'ensemble $\Omega = \{1, \dots, N\}^{\mathbb{N}}$ (équipé de la tribu produit, cf. l'exemple 1.3) figure l'univers des possibles de l'épreuve suivante, dite *marche aléatoire sur le graphe orienté* Γ : un robot stupide se déplace sur le graphe Γ , ce depuis sa position initiale $i(0) = 1$; à chaque instant k , ce robot passe (si cela est possible) de $i(k)$ (sa position à cet instant) à $i(k+1) \in \{j; (i(k), j) \in V\}$ (tous les choix étant équiprobables puisque le robot est stupide); s'il n'existe aucun j tel que $(i, j) \in V$, il va indifféremment (tous les choix étant équiprobables) vers l'un des éléments $i(k+1) \in \{1, \dots, N\}$. Les résultats possibles de l'épreuve sont toutes les listes $\{i(0), i(1), i(2), \dots\}$ des positions successives du robot dans l'ensemble E des sommets du graphe aux instants $k = 0, k = 1, \dots$. L'univers des possibles est bien l'ensemble $\Omega = \{1, \dots, N\}^{\mathbb{N}}$, équipé de la tribu produit \mathcal{F} . Reste à définir la loi de probabilité régissant ici le hasard.

Supposons dans un premier temps que pour chaque sommet du graphe, il existe au moins une arête orientée issue de ce sommet, *i.e.*, pour tout $i \in \{1, \dots, N\}$, $\{j; (i, j) \in V\} \neq \emptyset$. Soit P une probabilité sur (Ω, \mathcal{F}) , en accord avec les règles régissant le hasard et précisées ci dessus. Soit $k \in \mathbb{N}$. Si i et j sont deux sommets, notons A_i^{k+1} l'évènement aléatoire $\ll \{X_{k+1} = i\} \gg$ et A_j^k l'évènement aléatoire $\ll \{X_k = j\} \gg$. Du fait des hypothèses, l'évènement A_j^k est de probabilité strictement positive et l'on a, suivant la règle régissant ici le hasard :

$$(1.25) \quad P(A_i^{k+1} | A_j^k) = 1/L_j = g_{ji}^\Gamma \quad \forall i, j \in \{1, \dots, N\}.$$

8. On comprendra pourquoi cette notation \mathbb{G}^Γ plus loin.

Pour chaque indice $j = 1, \dots, N$, la j -ième ligne $[g_{j1}, \dots, g_{jn}]$ de la matrice \mathbb{G}^Γ s'interprète donc comme une distribution de probabilité (conditionnelle) sur l'ensemble E , considéré ici comme univers des possibles pour l'épreuve X_{k+1} . Soit $\mu_0 = [1, 0, \dots, 0]$ et, pour $k \in \mathbb{N}^*$, μ_k le vecteur ligne dont les entrées sont les $P(A_i^k)$. La formule de Bayes ((1.21), Proposition 1.2) implique

$$P(A_i^{k+1}) = \sum_{j=1}^N P(A_i^{k+1} | A_j^k) = \sum_{j=1}^N P(A_i^{k+1} | A_j^k) P(A_j^k),$$

soit

$$(1.26) \quad \mu_{k+1} = \mu_k \cdot \mathbb{G}^\Gamma \quad \forall k \in \mathbb{N}.$$

Si l'on avait supposé le robot non plus stupide, mais aveugle, la règle régissant le hasard s'en trouverait modifiée : lorsqu'il est positionné en $i(k)$ à l'instant k , le robot irait indifféremment sur $i(k+1) \in \{1, \dots, N\}$. Il faut dans ce cas remplacer la probabilité P sur Ω par une probabilité \tilde{P} adaptée à ce nouveau modèle et l'on a maintenant, à la place de (1.25),

$$(1.27) \quad \tilde{P}(A_i^{k+1} | A_j^k) = \frac{1}{N} \quad \forall i, j \in \{1, \dots, N\}.$$

Ceci conduit, si $\tilde{\mu}_0 = \mu_0$ et $\tilde{\mu}_k$ désigne, pour $k \in \mathbb{N}^*$, le vecteur ligne dont les entrées sont les $\tilde{P}(A_i^k)$, $i = 1, \dots, k$, aux relations

$$(1.28) \quad \tilde{\mu}_{k+1} = \tilde{\mu}_k \cdot \frac{1}{N} \mathbf{ones}(N, N) \quad \forall k \in \mathbb{N}$$

en place de (1.26).

Si le graphe Γ est maintenant quelconque (des sommets peuvent n'être le point de départ d'aucune flèche), il est naturel d'introduire un *facteur de risque* $1 - \kappa$ choisi *a priori* dans $]0, 1[$ (petit si possible) et de considérer que la probabilité P_κ sur (Ω, \mathcal{S}) régissant maintenant la marche aléatoire du robot est telle que

$$(1.29) \quad P_\kappa(A_i^{k+1} | A_j^k) = (1 - \kappa) \tilde{P}(A_i^{k+1} | A_j^k) + \kappa P(A_i^{k+1} | A_j^k) \quad \forall i, j \in \{1, \dots, N\}.$$

Auquel cas, si $\mu_0^\kappa = \mu_0$ et μ_k^κ désigne, pour $k \in \mathbb{N}^*$, le vecteur ligne dont les entrées sont les $P_\kappa(A_i^k)$, on a les relations suivantes :

$$(1.30) \quad \begin{aligned} \mu_{k+1}^\kappa &= \mu_k^\kappa \cdot \left[\frac{(1 - \kappa)}{N} \mathbf{ones}(N, N) + \kappa \mathbb{G}^\Gamma \right] \\ &= \frac{1 - \kappa}{N} \mathbf{ones}(1, N) + \kappa \mu_k^\kappa \cdot \mathbb{G}^\Gamma \quad \forall k \in \mathbb{N}. \end{aligned}$$

La suite de vecteurs ligne μ_k^κ converge en effet, lorsque k tend vers l'infini, vers l'unique point fixe de l'application strictement contractante

$$\mu \mapsto \frac{1 - \kappa}{N} \mathbf{ones}(1, N) + \kappa \mu \cdot \mathbb{G}^\Gamma,$$

point fixe constituant une « probabilité d'équilibre » permettant ainsi d'associer une « pertinence » à chaque sommet. Le modèle que nous venons d'explicitier ici à titre d'exemple est un modèle de *chaîne de Markov homogène* à N états, modèle sur lequel nous reviendrons.

EXEMPLE 1.8 (**Google** et **Pagerank**). Dans le cas de la marche aléatoire sur le **web**, on s'accorde sur le choix de $\kappa \simeq 0.85$. On obtient ainsi une distribution de pertinence μ_∞^κ pour les N pages **web** de la toile. Nous venons de décrire ici le principe (simplifié) de l'algorithme **Pagerank** sur lequel se fonde **Google**. On pourra consulter également [**Eiser**] pour plus de détails.

1.5.3. Conditionnement et loi exponentielle. Supposons que les atomes d'un corps radioactif se désintègrent de manière aléatoire. On considère un espace probabilisé (Ω, \mathcal{F}, P) adapté aux règles imposées par le physicien, à savoir la règle suivante : la probabilité que, si un atome n'est pas désintégré à l'instant t_0 , il ne se désintègre pas dans l'intervalle de temps $[t_0, t_0 + t]$, est une fonction qui ne dépend que de t (fonction que l'on appellera $F(t)$). Examinons alors l'évolution d'un atome de la population, intact à l'instant $t = 0$. Appelons pour cela A_s l'évènement « l'atome est encore intact à l'instant $t \geq 0$ ». On a, pour chaque instant $s \geq 0$, $P(A_s) = 1 - F(s)$ et

$$P(A_{s+t}) = P(A_{s+t} \mid A_s)P(A_s), \quad \forall s, t \geq 0.$$

En tenant compte de la règle qui préside au hasard qui régit le processus, il vient donc :

$$1 - F(s+t) = (1 - F(s))(1 - F(t)) \quad \forall s, t \geq 0.$$

En résolvant cette équation fonctionnelle classique (et en supposant F continue en $t = 0$), on voit qu'il existe un réel (forcément positif) λ tel que $F(t) = e^{-\lambda t}$, d'où

$$1 - F(t) = \exp(-\lambda t) \quad \forall t \geq 0.$$

On a donc $F(t) = 1 - e^{-\lambda t}$ pour tout $t \geq 0$. La durée de vie d'un atome donné de la population suit donc la loi exponentielle de paramètre λ (cf. la sous-section 1.4.6). On retrouve en effet ici (avec F) la fonction de répartition (1.18) de cette loi exponentielle (il est en effet naturel de prolonger F par 0 sur $] -\infty, 0[$).

On pourrait plus prosaïquement remplacer les atomes par les composants d'un système. La durée de vie d'un composant (le fait que le composant tombe en panne remplaçant maintenant la désintégration atomique) suit une loi exponentielle de paramètre $\lambda > 0$. On interprétera ultérieurement ce paramètre λ en terme de « durée de vie moyenne » (λ représentera en fait l'inverse de la durée de vie moyenne).

1.5.4. Événements aléatoires indépendants ; épreuves indépendantes.

La notion d'« indépendance » est une notion capitale en probabilités. On dispose d'une notion rigoureuse d'*indépendance entre événements aléatoires* (d'un même espace probabilisé, associé à une épreuve). Mais on parle aussi (cette fois de manière plus heuristique) d'« *indépendance entre épreuves aléatoires* ». Ceci signifie qu'il convient de choisir, si les épreuves sont considérés « en famille » (et non plus individuellement), une probabilité sur l'univers des possibles de ces épreuves en famille, de manière à ce que ce choix rende compte de l'hypothèse d'indépendance heuristique entre les épreuves que nous souhaitons envisager dans le modèle. Nous allons donc préciser ici ce double visage de la notion d'indépendance en probabilités.

DÉFINITION 1.12 (indépendance de deux événements aléatoires, mutuelle indépendance). Soit (Ω, \mathcal{F}, P) un espace probabilisé (Ω étant l'univers des possibles attaché à une épreuve). Deux événements aléatoires A et B de \mathcal{F} sont dits *indépendants* si et seulement si on a la condition d'indépendance

$$(1.31) \quad P(A \cap B) = P(A) \times P(B).$$

Plus généralement, si $\{A_k; k \in \mathbb{N}^*\}$ désigne une famille d'évènements aléatoires ($A_k \in \mathcal{F}$ pour tout $k \in \mathbb{N}^*$), on dit que ces évènements sont *mutuellement indépendants* si et seulement si, pour tout $L \in \mathbb{N}^*$, pour tout sous-ensemble à L éléments $\{k_1, \dots, k_L\}$ (distincts ou non) de \mathbb{N}^* , on a

$$(1.32) \quad P\left(\bigcap_{l=1}^L A_{k_l}\right) = \prod_{l=1}^L P(A_{k_l}).$$

REMARQUE 1.10 (Attention à ne pas confondre « indépendance » et « incompatibilité »!). Deux évènements aléatoires A et B sont dits *incompatibles* (ou encore *disjoints*) si $A \cap B = \emptyset$. Cette notion (ensembliste) d'incompatibilité n'a rien à voir avec la notion (elle, probabiliste, car elle implique la probabilité P) d'indépendance entre A et B (relativement à une probabilité P donnée sur l'espace probabilisable (Ω, \mathcal{F})).

REMARQUE 1.11 (Attention encore!). Il faut aussi prendre garde au fait que l'*indépendance d'une famille d'évènements aléatoires « deux-à-deux »* ou encore « *par paires* » (ce qui constitue une hypothèse fréquente dans la pratique, car commode à manier) n'implique pas en général la mutuelle indépendance de ces évènements. Les conditions de mutuelle indépendance (1.32) (pour tout L , pour toute partie $\{k_1, \dots, k_L\} \subset \mathbb{N}^*$) sont en général bien plus lourdes à exprimer, donc beaucoup moins maniables en pratique. Il est donc souvent prudent de les éviter si toutefois cela s'avère possible et de leur substituer l'hypothèse plus faible d'« indépendance deux-à-deux » (ou encore par paires).

EXEMPLE 1.9 (indépendance *versus* non-conditionnement). Si A et B sont deux évènements aléatoires tels que $P(A) > 0$, les évènements A et B sont indépendants si et seulement si $P(B | A) = P(B)$, autrement dit le fait que A soit réalisé ou non n'influe en rien sur la probabilité de B . Ceci résulte de la définition (1.19) de $P(B | A)$. Si A est tel que $P(A) = 0$, on a $P(A \cap B) = 0$ pour tout autre évènement aléatoire B , autrement dit A est dans ce cas indépendant de tout évènement aléatoire $B \in \mathcal{F}$.

On pourra vérifier en exercice la Proposition suivante.

PROPOSITION 1.4. *Si A et B sont deux évènements aléatoires indépendants, il en est de même de leurs complémentaires $\Omega \setminus A$ et $\Omega \setminus B$.*

Considérons maintenant deux épreuves pour lesquelles le hasard est « gouverné ». Pour la première, l'univers des possibles est Ω_1 , les évènements aléatoires composent la tribu \mathcal{T}_1 , et le hasard est « piloté » par le choix de la probabilité $P_1 : \mathcal{T}_1 \rightarrow [0, 1]$. Pour la seconde, l'univers des possibles est Ω_2 , les évènements aléatoires composent la tribu \mathcal{T}_2 , et le hasard est « piloté » cette fois par le choix de la probabilité $P_2 : \mathcal{T}_2 \rightarrow [0, 1]$. Voici le modèle probabiliste qu'il faut prendre lorsque l'on a affaire aux deux épreuves ensemble et que l'on souhaite un pilotage du hasard assurant que ces deux épreuves sont « indépendantes » (heuristiquement) : l'univers des possibles (pour le couple d'épreuves) est $\Omega_1 \times \Omega_2$, la tribu est la tribu produit (la plus petite tribu contenant tous les $A \times B$, $A \in \mathcal{T}_1$, $B \in \mathcal{T}_2$, voir l'exemple 1.3), la probabilité P à choisir sur cette tribu produit est l'unique probabilité telle que

$$P(A \times B) = P_1(A) \times P_2(B) \quad \forall A \in \mathcal{T}_1, \forall B \in \mathcal{T}_2.$$

Cette probabilité est dite *probabilité produit* de P_1 et P_2 .

Ceci vaut aussi lorsque l'on affine à une liste dénombrable d'épreuves (indexées par $l \in \mathbb{N}^*$), qualifiées de « mutuellement indépendantes » (heuristiquement), par exemple une suite infinie de lancers d'une pièce, *i.e.* ce que l'on appelle un « jeu de Pile ou Face ». L'univers des possibles est alors le produit

$$\prod_{l \in \mathbb{N}^*} \Omega_l$$

des univers des possibles Ω_l , $l = 1, 2, \dots$, des diverses épreuves (par exemple $\Omega_l = \{0, 1\}$ dans le cas du jeu de Pile ou Face). La tribu à prendre est la tribu produit sur cet ensemble (*i.e.* la plus petite tribu contenant tous les $A_1 \times \dots \times A_L \times \Omega_{L+1} \times \dots$, voir l'exemple 1.3). On y met l'unique probabilité telle que, pour tout $L \in \mathbb{N}^*$, pour toute famille A_1, \dots, A_L d'évènements aléatoires ($A_l \in \mathcal{F}_l$ pour tout $l = 1, \dots, L$), on ait

$$(1.33) \quad P(A_1 \times \dots \times A_L \times \Omega_{L+1} \times \dots) = \prod_{l=1}^L P_l(A_l).$$

Cette probabilité est dite *probabilité produit* des probabilités P_l , $l \in \mathbb{N}^*$.

EXEMPLE 1.10 (lois à densité indépendantes). Si X_1, \dots, X_N correspondent à N épreuves indépendantes (à valeurs dans \mathbb{R}), toutes régies par la même loi à densité f sur \mathbb{R} , le vecteur (X_1, \dots, X_N) est une épreuve (à valeurs dans \mathbb{R}^N), régie par la loi à densité

$$(x_1, \dots, x_N) \in \mathbb{R}^N \mapsto f(x_1) f(x_2) \dots f(x_{N-1}) f(x_N).$$

Variables aléatoires et théorèmes « limite »

2.1. Notion de VAR, espérance, variance, moments

2.1.1. Épreuves et variables aléatoires. Soit Ω l'univers des possibles attaché à une épreuve aléatoire (*cf.* le chapitre 1, sous-sections 1.1 et 1.4), \mathcal{T} la tribu des événements aléatoires attachés à cette épreuve, et P la loi de probabilité sur (Ω, \mathcal{T}) « présidant » au hasard régissant la dite épreuve.

DÉFINITION 2.1 (variable aléatoire réelle (VAR), vecteur de variables aléatoires réelles). On appelle *variable aléatoire réelle* (en abrégé VAR) sur l'espace probabilisé (Ω, \mathcal{T}, P) attaché à une épreuve aléatoire toute fonction $X : \Omega \rightarrow \mathbb{R}$ telle que, pour tout $T \in \mathbb{R}$, $\{\omega \in \Omega; X \leq T\}$ soit un événement aléatoire (pour cette épreuve), *i.e.* un élément de la tribu \mathcal{T} . La fonction

$$T \in \mathbb{R} \longmapsto P(\{\omega \in \Omega; X(\omega) \leq T\})$$

est alors appelée *fonction de répartition* de la VAR X . Si $X(\Omega)$ est un ensemble fini ou dénombrable, la VAR X est dite *discrète*. Dans le cas contraire (en particulier si $X(\Omega)$ est un intervalle I de \mathbb{R}), la VAR X est dite *continue*¹. Si X_1, \dots, X_n sont n VAR sur (Ω, \mathcal{T}, P) , on dit que (X_1, \dots, X_n) est un *n -vecteur de variables aléatoires*.

EXEMPLE 2.1. Le modèle de VAR attaché à une épreuve se présente comme une fonction (constructible à partir de la connaissance de la tribu \mathcal{T}) du résultat de cette épreuve : par exemple, si l'épreuve est une partie de « Pile ou Face » (avec univers des possibles $\Omega = \{0, 1\}^{\mathbb{N}}$ équipé de la tribu produit), le numéro du jeu où l'on obtient le premier « Pile », le numéro du jeu où l'on obtient deux « Pile » successifs, le nombre de « Face » obtenus lors des N premiers jeux, *etc.* sont des VAR discrètes. Dans un processus de désintégration atomique (ou dans à un appareillage impliquant une famille de composants se comportant « indépendamment », amenés à être remplacés instantanément dès l'instant où ils tombent en panne), le nombre d'atomes se désintégrant pendant l'intervalle temporel $[0, T]$ (ou le nombre de pannes affectant le système pendant le laps de temps $[0, T]$), sont aussi des VAR discrètes. Si l'épreuve est un concours de saut en hauteur (impliquant N athlètes, chacun ayant droit à M essais), le *record* de hauteur franchi est une VAR continue².

DÉFINITION 2.2 (loi d'une VAR, d'un vecteur de VAR). Soit $X : \Omega \rightarrow \mathbb{R}$ une VAR sur un espace probabilisé (Ω, \mathcal{T}, P) attaché à une épreuve aléatoire. Il existe

1. Le mot « continue » n'est pas à entendre ici au sens où vous l'entendrez habituellement : c'est l'ensemble image $X(\Omega) \subset \mathbb{R}$ qui se trouve être en un certain sens un « continuum ».

2. Il convient bien sûr de bien préciser comment est « piloté » le hasard du concours : quelle règle régit la hauteur franchie par un athlète lors d'un saut donné, les divers sauts de chacun des athlètes à leurs divers essais sont-ils indépendants ? indépendants athlète par athlète ? *etc.*

unique probabilité P_X sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que

$$(2.1) \quad P_X(]-\infty, T]) = P(\{\omega \in \Omega; X(\omega) \leq T\}) \quad \forall T \in \mathbb{R}.$$

On appelle cette probabilité P_X la *loi* de la VAR X .

Pour un vecteur $\vec{X} = (X_1, \dots, X_n)$ de VAR, on appelle *loi* de \vec{X} l'unique probabilité $P_{\vec{X}}$ sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ telle que

$$(2.2) \quad P_{\vec{X}}(]-\infty, T_1] \times \dots \times]-\infty, T_n]) = P(\{\omega \in \mathbb{R}^n; \vec{X}(\omega) \in]-\infty, T_1] \times \dots \times]-\infty, T_n])\})$$

pour tout T_1, \dots, T_n dans \mathbb{R} .

2.1.2. Notion d'espérance d'une VAR ou d'un vecteur de VAR. La démarche présentée ici pour introduire l'espérance d'une VAR est plus une démarche de statisticien : en statistique, on dispose de variables aléatoires réalisées empiriquement par simulation et dont on désire prouver qu'elles sont assujetties à des lois de probabilité particulières : *espérance* et *variance* apparaissent ainsi comme des paramètres essentiels de telles lois de probabilité « candidates » à modéliser la loi de X , et il est donc naturel de chercher à estimer ces paramètres à partir de simulations répétées indépendamment du phénomène aléatoire observé X . Esquissons la démarche conduisant à l'estimation empirique (et, au delà, à la définition rigoureuse) de l'*espérance*³ d'une VAR. Nous envisagerons ultérieurement (exemple 2.11) l'estimation empirique de la variance, ainsi que la définition rigoureuse de ce concept.

Soit X est une variable aléatoire réelle supposée dans un premier temps positive sur l'espace probabilisable (Ω, \mathcal{T}, P) attaché à une épreuve aléatoire. Si l'on simule l'épreuve aléatoire M fois (de manière à ce que les M simulations répétées soient indépendantes entre elles, comme expliqué dans la section 1.5.4⁴), on peut, en prenant un pas $h > 0$, dresser l'histogramme de la répartition des valeurs

$$X_1(\omega), \dots, X_M(\omega)$$

obtenues par la VAR X lors des M simulations successives de l'épreuve aléatoire : pour chaque $k \in \mathbb{N}$, on reporte au dessus du segment $[kh, (k+1)h[$ le nombre $N(M, h, k)$ de valeurs $j \in \{1, \dots, M\}$ telles que $X_j(\omega)$ « tombe » dans l'intervalle $[kh, (n+1)k[$. On note ensuite

$$(2.3) \quad \frac{1}{M} \sum_{j=1}^M X_j(\omega) - h \leq x_{M,h} := \sum_{k \in \mathbb{N}} kh \frac{N(M, h, k)}{M} \leq \frac{1}{M} \sum_{j=1}^M X_j(\omega)$$

et l'on interprète le nombre $x_{M,h}$ comme le barycentre des points kh pondérés par les « poids » que sont les $N(M, h, k)/M$. On observe que ce nombre $x_{M,h}$ tend, lorsque h tend vers 0 et M vers l'infini, vers une valeur positive $E[X]$ (finie ou infinie) que l'on appellera *espérance* de la VAR positive X . La variable aléatoire

$$(2.4) \quad X_{[M]}^{\text{emp}} := \frac{1}{M} \sum_{j=1}^M X_j$$

3. Ce qualificatif est plus propre au langage probabiliste. Le statisticien parlera plutôt de *moyenne*.

4. Ceci revient donc à travailler sur Ω^M , équipé de la tribu produit \mathcal{T} par elle-même M fois et de la probabilité produit P par elle-même M fois.

définie sur Ω^N , équipé de la tribu produit de \mathcal{T} par elle-même M fois et de la probabilité produit (1.33) P_M de P par elle-même M fois (cela pour rendre compte de l'indépendance des M simulations effectuées pour obtenir une valeur approchée de $E(X)$) s'appelle *espérance empirique* de X ou encore *estimateur sans biais*⁵ de la *moyenne de l'espérance* de X . L'espérance $E(X)$ est en théorie définie⁶ par

$$(2.5) \quad \begin{aligned} E[X] &= \sup_{h>0} \left(\sum_{k \in \mathbb{N}} kh P_X([kh, (k+1)h]) \right) = \\ &= \sup_{h>0} \left(\sum_{k \in \mathbb{N}} kh P(\{\omega; X(\omega) \in [kh, (k+1)h]\}) \right). \end{aligned}$$

Lorsque $E[X]$ est fini, on dit que X est *d'espérance finie* ou encore *de moyenne finie* et l'on appelle aussi $E[X]$ dans ce cas la *moyenne* de la VAR positive X .

L'estimation approchée de $E(X)$ (lorsque X est une VAR positive) par son espérance empirique conduit à deux observations capitales :

- la prise d'espérance est linéaire : si X et Y sont deux VAR positives définies sur le même espace probabilisé (Ω, \mathcal{T}, P) , et si λ, μ sont deux constantes positives, on a

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$$

(par convention $0 \times \infty = 0$).

- la prise d'espérance est monotone : Si X et Y sont deux VAR positives définies sur le même espace probabilisé (Ω, \mathcal{T}, P) et telles que $X \leq Y$ sur un évènement aléatoire A tel que $P(A) = 1$, on a

$$E(X) \leq E(Y).$$

Cette notion d'espérance s'étend aux VAR de signe quelconque, et même aux vecteurs de VAR (en particulier aux VAR à valeurs complexes).

DÉFINITION 2.3 (espérance d'une VAR ou d'un vecteur de VAR). Une VAR $X : \Omega \rightarrow \mathbb{R}$ sur un espace probabilisé (Ω, \mathcal{T}, P) (attaché à une certaine épreuve aléatoire dont Ω figure l'univers des possibles, P représentant la « gouvernance » du hasard) est dite *avoir une espérance* si et seulement si $E(|X|) < +\infty$. On définit alors l'*espérance* ou *moyenne* de la VAR X par

$$(2.6) \quad \begin{aligned} E(X) &= E(\sup(X, 0)) - E(\sup(-X, 0)) = \\ &= \lim_{k \rightarrow +\infty} \lim_{M \rightarrow +\infty} \left(\sum_{l=-M}^{M-1} \frac{l}{2^k} P_X([l/2^k, (l+1)/2^k]) \right) \end{aligned}$$

(la limite au second membre de (2.6) existe bien du fait que $E(|X|) < +\infty$ et que par conséquent l'indétermination $\infty - \infty$ n'apparaît pas dans ce calcul d'espérance où l'on soustrait les deux nombres réels positifs finis $E(\sup(X, 0))$ et $E(\sup(-X, 0))$). Un vecteur $\vec{X} : \Omega \rightarrow \mathbb{R}^n$ de VAR sur (Ω, \mathcal{T}, P) est dit avoir une espérance si $E(\|\vec{X}\|) < +\infty$. On définit alors $E(\vec{X})$ comme le vecteur $(E(X_1), \dots, E(X_n)) \in \mathbb{R}^n$. Si $Z = X + iY$ est une VAR à valeurs complexes telle que $E(|Z|) < +\infty$, on définit en particulier l'espérance de Z comme $E(Z) := E(X) + iE(Y)$.

5. La terminologie tient ici au fait que l'on divise la somme des $X_j(\omega)$ exactement par le nombre M de valeurs « indépendantes » $X_j(\omega)$.

6. Son approximation par les $x_{M,h}$ le suggère au vu de l'interprétation en termes de probabilités des $N(M, h, k)/M$ lorsque M tend vers l'infini.

La prise d'espérance entre VAR ou vecteurs de VAR ayant une espérance est une opération \mathbb{R} -linéaire : si X et Y sont deux telles VAR (définies sur le même espace probabilisé (Ω, \mathcal{F}, P)) et si λ, μ sont deux nombres réels, $\lambda X + \mu Y$ admet aussi une espérance et l'on a

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y).$$

La même chose vaut si \vec{X} et \vec{Y} sont deux vecteurs de VAR définis sur le même espace probabilisé (Ω, \mathcal{F}, P) et ayant tous deux une espérance (ceci inclus en particulier le cas où X et Y sont des VAR complexes ayant toutes deux une espérance, auquel cas on peut prendre les scalaires λ et μ complexes et non plus réels).

REMARQUE 2.1. Si X est une VAR ayant une espérance, il en est de même (pour tout entier $M \in \mathbb{N}^*$) de l'espérance empirique, que l'on définit naturellement par $X_{[M]}^{\text{emp}} := (X_1 + \dots + X_M)/M$, X_1, \dots, X_M désignant M simulations indépendantes de X , et l'on a

$$E(X_{[M]}^{\text{emp}}) = E(X) \quad \forall M \in \mathbb{N}^*.$$

Le qualificatif « espérance empirique » pour $X_{[M]}^{\text{emp}}$ est ainsi justifié.

REMARQUE 2.2 (inégalité triangulaire, inégalité de Jensen). Si \vec{X} est un vecteur de VAR à densité ayant une espérance, on a toujours l'*inégalité triangulaire* $\|E(X)\| \leq E(\|X\|)$. En particulier, si $Z = X + iY$ est une VA complexe ayant une espérance, on a $|E(Z)| \leq E(|Z|)$. Si X est une VAR, à valeurs dans un intervalle ouvert $I \subset \mathbb{R}$, et si $\varphi : I \rightarrow \mathbb{R}$ est une fonction convexe, alors, si X admet une espérance, celle ci appartient nécessairement à l'intervalle I (car on peut l'interpréter comme un barycentre de points de l'intervalle I) et l'on a

$$\varphi(E(X)) \leq E(\varphi(X)) := E(\sup(\varphi(X), 0)) - E(\sup(-\varphi(X), 0)) \in]-\infty, \infty]$$

(inégalité importante en pratique, dite de Jensen : penser par exemple au cas où φ est l'exponentielle où $I = \mathbb{R}$ et φ est une fonction puissance $|t|^\alpha$, avec $\alpha \in [1, \infty[$).

2.1.3. Exemples de calcul d'espérance. Reprenons le catalogue des lois introduites au chapitre 1.

Commençons par le cas des VAR discrètes, où l'ensemble des valeurs est un sous-ensemble fini ou dénombrable de \mathbb{R} . Dans ce cas, si $E(X)$ existe, on a

$$E(X) = \sum_{\omega \in \Omega} \omega P(\{X = \omega\}).$$

On peut même tolérer que Ω contienne $-\infty$ ou $+\infty$, mais dans ce cas, X n'admet une espérance que si $P(\{X = \pm\infty\}) = 0$.

EXEMPLE 2.2 (VAR suivant loi uniforme $\mathcal{U}(N)$ sur un ensemble Ω à N éléments). Si Ω est un ensemble de nombres réels $\{\omega_1, \dots, \omega_N\}$, une VAR de loi uniforme sur cet ensemble (cf. la sous-section 1.2.2) a pour espérance la moyenne

$$E(X) = \frac{\sum_{j=1}^N \omega_j}{N}.$$

Si par exemple $\Omega = \{1, \dots, N\}$, on trouve

$$E(X) = \frac{1}{N} \times (1 + 2 + \dots + N) = \frac{N(N+1)}{2N} = \frac{N+1}{2}.$$

EXEMPLE 2.3 (VAR suivant une loi binomiale $\mathcal{B}(N, p)$ (cf. la sous-section 1.2.3)). Il s'agit (penser à sa réalisation dans un jeu de « Pile ou Face » à N coups, X comptant le nombre de « Pile » obtenus au final, le jet de la pièce donnant « Pile » avec la probabilité p et « Face » avec la probabilité $1 - p$) de la somme de N variables X_j de Bernoulli, valant chacune 1 avec la probabilité p et 0 avec la probabilité $1 - p$. Chacune de ces variables de Bernoulli a pour espérance $0 \times (1 - p) + 1 \times p$ et l'espérance de X suivant une loi $\mathcal{B}(N, p)$ sur $\{0, \dots, N\}$ est donc $E(X) = Np$.

EXEMPLE 2.4 (VAR suivant une loi hypergéométrique $\mathcal{H}(N, R, B)$ (cf. la sous-section 1.2.4)). On utilise ici la réalisation de X comme le nombre de boules rouges tirées au final au bout de N tirages successifs sans remise (les boules de l'urne étant indifférenciées au tirage) hors d'une urne contenant initialement R boules rouges et B boules blanches ($N \leq R + B$). Notons, pour $j = 1, \dots, N$, X_j la VAR valant 1 si la boule tirée au j -ième tirage est rouge, 0 si elle est blanche. Pour $k \in \{\max(0, N - B), \dots, \min(N, R)\}$ (univers des possibles Ω de l'épreuve), on a $X = X_1 + \dots + X_N$. Montrons par récurrence que toutes les variables X_1, \dots, X_N ont même espérance $R/(R + B) = E(X_1)$. Supposons que ce soit vrai jusqu'au cran $k < N$ et posons $S_k = X_1 + \dots + X_k$. On a

$$\begin{aligned} E(X_{k+1}) &= P(\{X_{k+1} = 1\}) = \sum_{l \in \Omega} P(\{X_{k+1} = 1\} \mid S_k = l) P(S_k = l) = \\ &= \sum_{l \in \Omega} \frac{R - l}{B + R - k} P(S_k = l) = \frac{R - E(S_k)}{B + R - k} \end{aligned}$$

d'après la formule (1.21) (théorème de Bayes 1.2). Or

$$E(S_k) = E(X_1) + \dots + E(X_k) = k \frac{R}{B + R}$$

d'après l'hypothèse de récurrence. On déduit finalement

$$E(X_{k+1}) = \frac{R - \frac{Rk}{B + R}}{B + R - Rk} = \frac{R}{B + R} = E(X_1).$$

L'espérance d'une VAR suivant une loi hypergéométrique $\mathcal{H}(N, R, B)$ vaut donc $N R/(R + B)$. Si l'on pose $p = R/(R + B)$, il n'y a donc pas, au niveau du calcul d'espérance, de distinction entre les tirages avec ou sans remise hors d'une urne (régis le premier par la loi binomiale, le second par la loi hypergéométrique). Ceci peut paraître à première vue surprenant.

Il est parfois commode, lorsque la VAR est à valeurs dans \mathbb{N} , d'introduire le concept de *fonction génératrice*. Cette fonction est définie comme l'espérance (elle existe) de la VA à valeurs complexes $\omega \mapsto s^{X(\omega)}$, soit :

$$(2.7) \quad s \in D(0, 1) \mapsto G_X(s) = E(s^X) = \sum_{k \in \mathbb{N}} P(\{X = k\}) s^k.$$

La fonction $s \in D(0, 1) \mapsto G_X(s)$ représente, dans le disque ouvert $D(0, 1)$, la somme d'une série entière de rayon de convergence R au moins égal à 1. On a

$$\forall z \in D(0, R), \quad G'_X(s) = \sum_{k=1}^{\infty} k P(\{X = k\}) s^{k-1}.$$

Si $R > 1$, la fonction G_X se prolonge analytiquement à $D(0, R)$ et l'on a par conséquent $G'_X(1) = E(X)$.

EXEMPLE 2.5 (VAR suivant une loi de Poisson $\mathcal{P}(\lambda)$ (cf. la sous-section 1.3.1)). On calcule précisément dans ce cas la fonction génératrice, qui vaut

$$\forall s \in D(0, 1), G_X(z) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda s)^k}{k!} = e^{\lambda s - \lambda} = e^{\lambda(s-1)}$$

et se prolonge en la somme d'une série entière de rayon de convergence $+\infty$. On a donc

$$E(X) = \lambda \left[e^{\lambda(s-1)} \right]_{s=1} = \lambda.$$

Une VAR suivant une loi de Poisson $\mathcal{P}(\lambda)$ a donc pour espérance le paramètre λ . Ce calcul peut être effectué d'ailleurs ici sans le recours à la fonction génératrice : en effet

$$E(X) = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

EXEMPLE 2.6 (loi géométrique (cf. la sous-section 1.3.2)). On utilise encore la fonction génératrice

$$s \in D(0, 1) \mapsto \sum_{k=0}^{\infty} P(\{X = k\}) s^k = \sum_{k=0}^{\infty} p(1-p)^k s^k = \frac{p}{1 - (1-p)s}$$

si $p \in]0, 1[$. Le rayon de convergence vaut $R = 1/(1-p) > 1$ et l'on a

$$E(X) = G'_X(1) = p \times \frac{1-p}{p^2} = (1-p) \times \frac{1}{p} = \frac{1}{p} - 1.$$

Si $p = 0$, on a bien sûr $E(X) = 0$. Si l'on définit la loi de Pascal sur \mathbb{N}^* en posant $P(\{X = k\}) = p(1-p)^{k-1}$ (ce qui est souvent le cas), l'espérance est alors $1/p$.

Pour ce qui concerne le cas des VAR continues (à valeurs réelles ou complexes, ou même à valeurs dans \mathbb{R}^n), il faut d'abord envisager le cas des variables aléatoires à densité f .

– Si X est une VAR à densité f , X admet une espérance si et seulement si

$$\int_{\mathbb{R}} |t| f(t) dt < +\infty.$$

L'espérance de X est alors définie comme

$$E(X) = \int_{\mathbb{R}} t f(t) dt = \int_{]0, \infty[} t f(t) dt - \int_{]-\infty, 0[} |t| f(t) dt$$

(cf. la définition (1.15) avec $n = 1$ pour l'intégrale sur \mathbb{R} d'une fonction positive).

– Si \vec{X} est un vecteur de VAR à densité $f(x_1, \dots, x_n)$, \vec{X} admet une espérance si et seulement si

$$\int \dots \int_{\mathbb{R}^n} \|x\| f(x) dx_1 \dots dx_n < +\infty.$$

L'intégrale de f est alors définie comme le vecteur de coordonnées

$$\begin{aligned} (E(X))_j &= \int \cdots \int_{\mathbb{R}^n} x_j f(x) dx_1 \dots dx_n = \\ &= \int \cdots \int_{\{x_j \geq 0\}} x_j f(x) dx_1 \dots dx_n - \int \cdots \int_{\{x_j < 0\}} |x_j| f(x) dx_1 \dots dx_n. \end{aligned}$$

(cf. la définition (1.15) pour l'intégrale sur \mathbb{R}^n d'une fonction positive).

- Si Z est une VA à valeurs complexes à densité $f(z) \geq 0$ sur \mathbb{C} , Z admet une espérance si

$$\int \cdots \int_{\mathbb{R}^2} |x + iy| f(x + iy) dx dy < +\infty.$$

Ceci équivaut à dire que les VAR $X := \operatorname{Re} Z$ et $Y := \operatorname{Im} Z$ ont toutes deux une espérance. L'espérance de X est alors

$$E(X) = E(\operatorname{Re} Z) + i E(\operatorname{Im} Z) = E(X) + i E(Y).$$

REMARQUE 2.3 (le cas des vecteurs de VAR fonctions de vecteurs de VAR à densité). Si X est une VAR à densité f et si Y est une autre VAR (sur le même espace probabilisé) s'exprimant comme une fonction de X ($Y = \Phi(X)$), Y est à densité si et seulement si

$$\int_{\mathbb{R}} |\Phi(t)| f(t) dt < +\infty.$$

On a dans ce cas

$$E(Y) = E(\Phi(X)) = \int_{\mathbb{R}} \phi(t) f(t) dt = \int_{\{\Phi \geq 0\}} \Phi(t) f(t) dt - \int_{\{\Phi < 0\}} |\Phi(t)| f(t) dt.$$

Ceci vaut aussi si \vec{X} est un n -vecteur de VAR à densité $f(x_1, \dots, x_n)$ et \vec{Y} un m -vecteur de VAR s'exprimant comme une fonction $\vec{Y} = \Phi(\vec{X})$. Si $m = 1$ (Y est alors une VAR), Y admet une espérance si et seulement si

$$\int \cdots \int_{\mathbb{R}^n} |\Phi(x)| f(x) dx_1 \dots dx_n < +\infty.$$

L'espérance de $Y = \Phi(\vec{X})$ vaut alors

$$E(\Phi(\vec{X})) = \int \cdots \int_{\{\Phi \geq 0\}} \Phi(x) f(x) dx_1 \dots dx_n - \int \cdots \int_{\{\Phi < 0\}} |\Phi(x)| f(x) dx_1 \dots dx_n.$$

EXEMPLE 2.7 (VAR ou des vecteurs de VAR suivant une loi normale (cf. les sous-sections 1.4.4 et 1.4.5)). En utilisant la remarque 2.3 et le fait qu'une VAR suivant une loi $\mathcal{N}(0, 1)$ admet comme espérance

$$E(X) = \frac{1}{\sqrt{t}} \int_{\mathbb{R}} t e^{-t^2/2} dt = 0$$

on voit qu'une VAR suivant une loi $\mathcal{N}(\mu, \sigma^2)$ (déduite de la précédente comme $Y = \Phi(X)$, où $\Phi(t) = \mu + \sigma t$) a pour espérance μ . De même un vecteur de VAR suivant une loi $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$ admet pour espérance le vecteur $\boldsymbol{\mu}$.

EXEMPLE 2.8 (VAR suivant une loi exponentielle $\mathcal{E}(\lambda)$ (cf. les sous-sections 1.4.6 et 1.5.3)). Si X suit une loi exponentielle de paramètre $\lambda > 0$, elle admet une espérance et l'on a

$$E(X) = \lambda \int_0^\infty t e^{-\lambda t} dt = \frac{1}{\lambda}.$$

$1/\lambda$ est interprété ainsi comme durée de vie moyenne des composants dans un système (cf. la sous-section 1.5.3).

À une VAR quelconque, il est souvent utile d'associer une VA à valeur complexes (même de module 1), dite *fonction caractéristique*⁷ de X . Il s'agit de la fonction

$$(2.8) \quad \omega \in \mathbb{R} \mapsto E(e^{i\omega X}).$$

Si \vec{X} est un n -vecteur de VAR, il s'agit de la fonction

$$\omega \in \mathbb{R}^n \mapsto E(e^{i\langle \omega, X \rangle}).$$

Si \vec{X} est à densité, cette fonction est, d'après la remarque 2.3, la fonction

$$\omega \in \mathbb{R}^n \mapsto \int \cdots \int_{\mathbb{R}^n} f(x) e^{i\langle \omega, x \rangle} dx = \widehat{f}(-\omega),$$

où \widehat{f} est la *transformée de Fourier* (ou *spectre*) de la densité f .

2.1.4. Variance d'une VAR ou d'une VA complexe ; matrice de covariance d'un vecteur de VAR. Soit $\vec{X} = (X_1, \dots, X_n)$ un vecteur de VAR sur un espace probabilisé (Ω, \mathcal{F}, P) (correspondant à une épreuve aléatoire). Si $m \in \mathbb{N}^*$, on dit que \vec{X} admet des *moments d'ordre m* si $E[\|\vec{X}\|^m] < +\infty$. Si tel est le cas, on peut définir, pour chaque sous-ensemble $\{i_1, \dots, i_k\}$ de $\{1, \dots, n\}$, pour chaque vecteur d'entiers positifs $(\nu_{i_1}, \dots, \nu_{i_k})$ de somme m , la liste de toutes les espérances (bien définies)

$$E(X_{i_1}^{\nu_{i_1}} \cdots X_{i_k}^{\nu_{i_k}})$$

(liste des *moments d'ordre m du vecteur de VAR \vec{X} , non centrés*) ou, mieux, celle de toutes les espérances (encore bien définies)

$$E\left(\prod_{\kappa=1}^k (X_{i_\kappa} - E(X_{i_\kappa}))^{\nu_{i_\kappa}}\right)$$

(liste des *moments d'ordre m du vecteur de VAR \vec{X} , cette fois centrés*).

Dans le cas particulier où $m = 2$, la donnée de tous les moments d'ordre 2 centrés équivaut à celle de la matrice symétrique réelle

$$(2.9) \quad \text{cov}(\vec{X}) := \left[E((X_i - E(X_i))(X_j - E(X_j))) \right]_{1 \leq i, j \leq n},$$

dite *matrice de covariance* du vecteur de VAR \vec{X} . Du fait de la linéarité de la prise d'espérance (lorsqu'elle est, comme ici, licite), cette matrice de covariance $\text{cov}(\vec{X})$ s'exprime aussi

$$\text{cov}(\vec{X}) = \left[E(X_i X_j - E(X_i)E(X_j)) \right]_{1 \leq i, j \leq n}.$$

Cette matrice est une matrice réelle symétrique positive (*i.e.* de valeurs propres positives) puisque l'on a (toujours en invoquant la linéarité de la prise d'espérance,

7. Attention à ne pas toutefois confondre cette notion avec celle de fonction caractéristique d'un événement aléatoire, à savoir la fonction valant 1 sur cet événement, 0 ailleurs.

lorsqu'elle est, comme ici, licite), pour tout vecteur (colonne) \vec{V} de \mathbb{R}^n ,

$${}^t\vec{V} \cdot \text{cov}(\vec{X}) \cdot \vec{V} = E\left(\left(\sum_{j=1}^n v_j X_j - E\left(\sum_{j=1}^n v_j X_j\right)\right)^2\right) \geq 0.$$

Lorsque \vec{Z} est un vecteur (Z_1, \dots, Z_n) de variables aléatoires complexes ayant des moments d'ordre 2 (*i.e.* $E(\|Z\|^2) < +\infty$), on convient d'appeler *matrice de covariance de Z* la matrice (cette fois complexe, mais hermitienne)

$$(2.10) \quad \begin{aligned} \text{cov}(\vec{Z}) &:= \\ &= \left[E((Z_i - E(Z_i))(\overline{Z_j} - E(\overline{Z_j}))) \right]_{1 \leq i, j \leq n} = \left[E(Z_i \overline{Z_j} - E(Z_i) \overline{E(Z_j)}) \right]_{1 \leq i, j \leq n}. \end{aligned}$$

Il s'agit, ici encore, d'une matrice hermitienne positive (*i.e.* de valeurs propres réelles positives ou nulles).

Dans le cas des VAR ou des VA complexes ($n = 1$), nous avons naturellement la définition suivante :

DÉFINITION 2.4 (variance et écart type d'une VAR ou d'une VA complexe). Soit X une VAR sur un espace probabilisé (Ω, \mathcal{F}, P) , telle que $E(X^2) < +\infty$. La *variance* de X est définie comme le nombre positif

$$(2.11) \quad \text{Var}(X) := E\left(\left(X - E(X)\right)^2\right) = E(X^2) - (E(X))^2.$$

Le nombre $\sigma(X) := \sqrt{\text{Var}(X)}$ est alors appelé *écart type* de la VAR X . Si $Z = X + iY$ est une variable aléatoire complexe sur (Ω, \mathcal{F}, P) , telle que $E(|Z|^2) < +\infty$, on définit la *variance* de Z comme

$$(2.12) \quad \text{Var}(Z) := E\left(|Z - E(Z)|^2\right) = E(|Z|^2) - |E(Z)|^2.$$

L'écart type $\sigma(Z)$ de Z est encore la racine carrée de ce nombre $\text{Var}(Z)$.

REMARQUE 2.4 (Attention! la prise de variance est une opération quadratique et non linéaire!). La prise de variance de VAR (ou VA complexes), lorsqu'elle est licite, n'est pas une opération linéaire, mais une opération quadratique. Si X est une VAR telle que $E(X^2) < \infty$ et λ un scalaire réel, on a donc $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$. Si $Z = X + iY$ est une VA complexe telle que $E(|Z|^2) < \infty$ et λ un scalaire complexe, on a $\text{Var}(Z) = |\lambda|^2 \text{Var}(Z)$. Quant à la variance d'une somme de VAR (ou de VA complexes), ce n'est nullement en général la somme des variances de ces VAR en jeu puisqu'interviennent, comme dans toute expression quadratique, des « termes croisés » dits « termes d'interférence ». Ces termes ne disparaissent que sous une hypothèse d'indépendance deux-à-deux portant sur les VAR ou les VA complexes impliquées dans la somme (*cf.* la Proposition 2.4 plus loin).

EXEMPLE 2.9 (quelques exemples). La variance d'une loi d'une VAR suivant une loi de Bernoulli de paramètre $p \in [0, 1]$ (*i.e.* $P(\{X = 1\}) = p$ et $P(\{X = 0\}) = 1 - p$ vaut $p(1 - p)$). La variance d'une VAR discrète (à valeurs dans \mathbb{N}) se calcule

souvent à partir de fonction génératrice G_X (cf. (2.7)) de la loi : on remarque en effet que

$$G_X''(s) = \sum_{k \geq 2} k(k-1) P(\{X = k\}) s^{k-2}.$$

On a donc, au moins formellement,

$$G_X''(1) = E(X^2) - E(X) = E(X^2) - G_X'(1),$$

et ainsi

$$\text{Var}(X) = E(X^2) - (E(X))^2 = G_X''(1) + G_X'(1) - (G_X'(1))^2.$$

En particulier, pour une VAR suivant une loi de Poisson de paramètre λ , la fonction génératrice est $s \mapsto e^{\lambda(s-1)}$, et la variance vaut alors $\text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$. Pour une loi uniforme sur $[a, b]$, la variance vaut

$$\frac{1}{b-a} \int_a^b \left(t - \frac{b+a}{2}\right)^2 dt = \frac{1}{b-a} \int_{-(b-a)/2}^{(b-a)/2} u^2 du = \frac{(b-a)^2}{12}.$$

Plus généralement, si X est une VAR de densité f , on a

$$\text{Var}(X) = \int_{\mathbb{R}} \left(t - E(X)\right)^2 f(t) dt.$$

Par exemple, si X suit une loi géométrique $\mathcal{E}(\lambda)$, la variance vaut

$$\text{Var}(\mathcal{E}(\lambda)) = \lambda \int_0^\infty t^2 e^{-\lambda t} dt - \left(\lambda \int_0^\infty t e^{-\lambda t} dt\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Si X suit une loi normale $\mathcal{N}(\mu, \sigma^2)$, la variance de X vaut $\text{Var}(\mathcal{N}(\mu, \sigma^2)) = \sigma^2$.

L'importance de la variance en probabilités et en statistique tient à une inégalité capitale, dite de *inégalité de Bienaymé-Tchebychev* :

PROPOSITION 2.1 (inégalité de Bienaymé-Tchebychev). *Soit X une VAR définie sur un espace probabilisé (Ω, \mathcal{F}, P) et telle que $E(X^2) < +\infty$. Alors on a*

$$(2.13) \quad \forall \epsilon > 0, P\left(\{\omega \in \Omega; |X(\omega) - E(X)| \geq \epsilon\}\right) \leq \frac{\text{Var}(X)}{\epsilon^2} = \left(\frac{\sigma(X)}{\epsilon}\right)^2.$$

Ce résultat vaut encore si X est remplacée par une variable aléatoire complexe Z sur (Ω, \mathcal{F}, P) , telle que $E(|Z|^2) < +\infty$. Autrement dit : une information quantitative sur la variance de X permet de contrôler (en termes de probabilités d'évènements) la manière dont la VAR X prend une valeur « ϵ -éloignée » de son espérance⁸.

DÉMONSTRATION. Qui peut le plus peut le moins. On prouve donc ici cette inégalité dans le cas où Z est une VA complexe. Notons, pour $\epsilon > 0$, A_ϵ l'évènement aléatoire défini par

$$A_\epsilon := \{\omega \in \Omega; |Z(\omega) - E(Z)|^2 \geq \epsilon\}.$$

L'inégalité entre variables aléatoires sur (Ω, \mathcal{F}, P) :

$$|Z - E(Z)|^2 - \epsilon^2 \chi_{A_\epsilon} \geq 0$$

8. Pareille inégalité (2.13) ne devient bien sûr intéressante que si $\epsilon > \sigma(X)$ (car une probabilité d'évènement aléatoire est toujours majorée par 1).

conduit, lorsque l'on prend l'espérance, et que l'espérance d'une VAR positive est un nombre positif (propriété de monotonie de la prise d'espérance), à l'inégalité

$$\epsilon^2 P(A_\epsilon) \leq E(|Z - E(Z)|^2) = \text{Var}(Z),$$

qui est l'inégalité voulue. \square

2.1.5. Indépendance entre VAR et effet sur la covariance. Nous introduisons ici le concept d'indépendance mutuelle entre n VAR X_1, \dots, X_n toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) . Nous dégagerons ensuite comment se répercute pareille hypothèse d'indépendance sur la matrice de covariance $\text{cov}(\vec{X})$ du vecteur $\vec{X} = (X_1, \dots, X_n)$ de ces VAR.

DÉFINITION 2.5 (indépendance mutuelle entre VAR). Soient X_1, \dots, X_n , n VAR définies sur le même espace probabilisé (Ω, \mathcal{F}, P) et P_{X_1}, \dots, P_{X_n} leurs lois respectives sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (cf. (2.1), Définition 2.2). Soit $P_{\vec{X}}$ la loi du vecteur de VAR $\vec{X} := (X_1, \dots, X_n)$ sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ (cf. (2.2)). On dit que les VAR X_1, \dots, X_n sont *mutuellement indépendantes* (en tant que VAR sur l'espace probabilisé⁹ (Ω, \mathcal{F}, P)) si et seulement si

$$(2.14) \quad \forall T_1, \dots, T_n \in \mathbb{R}, \quad P_{\vec{X}}([-\infty, T_1] \times \dots \times]-\infty, T_n]) = \prod_{j=1}^n P_{X_j}([-\infty, T_j]).$$

REMARQUE 2.5 (indépendance mutuelle entre VA complexes). Si $Z_1 = X_1 + iY_1, \dots, Z_n = X_n + iY_n$ sont n VA complexes, on dit que Z_1, \dots, Z_n sont mutuellement indépendantes si les n VAR : X_1 ou Y_1, X_2 ou Y_2, \dots, X_n ou Y_n (il y a ainsi 2^n n -uplets possibles) sont indépendantes.

REMARQUE 2.6 (indépendance entre VAR elles-mêmes fonctions de VAR indépendantes). Si X_1, \dots, X_n sont des VAR indépendantes et Y_1, \dots, Y_n des VAR telles que, pour chaque $j = 1, \dots, n$, Y_j soit fonction uniquement de X_j ($Y_j = \Phi_j(X_j)$ pour $j = 1, \dots, n$), alors les VAR Y_1, \dots, Y_n , sont encore indépendantes. Il en est de même si $k \leq n$ et si Y_1, \dots, Y_k sont chacune fonction d'un « bloc » de variables X_j , ces divers « blocs » étant disjoints pour $j = 1, \dots, k$: les VAR Y_1, \dots, Y_k sont alors encore indépendantes. Cette remarque heuristique est vraiment importante du point de vue pratique. Cette remarque vaut encore si Z_1, \dots, Z_n sont, à la place de X_1, \dots, X_n , n VA complexes mutuellement indépendantes au sens de la Remarque 2.5.

L'*indépendance deux-à-deux* entre VAR X_1, \dots, X_n (ou VA complexes Z_1, \dots, Z_n) est évidemment une contrainte plus faible que celle d'indépendance mutuelle entre ces mêmes VAR ou ces VA complexes (comme l'est la notion d'indépendance deux-à-deux entre événements aléatoires vis-à-vis de celle d'indépendance mutuelle entre ces mêmes événements, cf. la Remarque 1.11). Mais cette contrainte d'indépendance deux-à-deux est beaucoup moins lourde à exprimer que celle d'indépendance mutuelle et, de plus, s'articule très bien avec la Proposition 2.2 ci-dessous. Elle est donc très souvent invoquée plutôt que la contrainte d'indépendance mutuelle souvent trop lourde à exprimer (et donc à gérer).

9. Cette notion de mutuelle indépendance dépend de manière essentielle de la probabilité P qui régit le hasard lors de l'épreuve aléatoire dont les X_j sont fonctions du résultat.

PROPOSITION 2.2 (espérance d'un produit de deux VAR indépendantes ou de deux VA complexes indépendantes). Soient X_1 et X_2 deux VAR définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , telles que $E(X_j^2) < +\infty$ pour $j = 1, 2$. Alors $X_1 X_2$ admet une espérance et, si de plus les deux variables X_1 et X_2 sont indépendantes, on a en prime

$$(2.15) \quad E(X_1 X_2) = E(X_1) \times E(X_2).$$

Soient Z_1 et Z_2 deux VA complexes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , telles que $E(|Z_j|^2) < +\infty$ pour $j = 1, 2$. Alors $Z_1 Z_2$ admet une espérance (complexe) et, si de plus les deux variables Z_1 et Z_2 sont indépendantes (comme VA complexes), on a en prime

$$(2.16) \quad E(Z_1 Z_2) = E(Z_1) \times E(Z_2).$$

DÉMONSTRATION. On donne d'abord une esquisse de preuve « empirique » dans le cas où X_1 et X_2 sont deux VAR. Le fait que $X_1 X_2$ ait une espérance (*i.e.* que $E(|X_1 X_2|) < +\infty$) résulte de l'inégalité

$$|X_1 X_2| \leq \frac{X_1^2 + X_2^2}{2}$$

et de la propriété de monotonie de l'espérance. Pour calculer l'espérance de $X_1 X_2$ (lorsque X_1 et X_2 sont supposées indépendantes), on reprend l'approche (empirique, cf. la sous-section 2.1.2) consistant :

- à calculer $E(X_1)$ en simulant M_1 fois X_1 de manière indépendante (ce qui aboutit aux résultats $X_{1,1}(\omega), \dots, X_{1,M_1}(\omega)$), à former l'histogramme des valeurs obtenues (le pas étant h) et enfin à calculer le barycentre des points kh affectés des coefficients $N(M_1, h, k)/M_1$;
- à calculer $E(X_2)$ en simulant M_2 fois X_2 de manière indépendante (ce qui aboutit aux résultats $X_{1,1}(\omega), \dots, X_{1,M_1}(\omega)$), à former l'histogramme des valeurs obtenues (le pas étant h) et enfin à calculer le barycentre des points kh affectés des coefficients $N(M_2, h, k)/M_2$.

On observe que les $X_{1,k_1}(\omega) X_{2,k_2}(\omega)$, $k_1 = 1, \dots, M_1$, $k_2 = 1, \dots, M_2$, peuvent être interprétés comme les résultats de $M_1 M_2$ simulations indépendantes de $X_1 X_2$ (l'indépendance venant ici du fait que X_1 et X_2 sont supposées indépendantes, et que, d'autre part, pour $j = 1, 2$, tous les $X_{j,k_j}(\omega)$ sont les résultats de M_j simulations indépendantes de la VAR X_j). On observe que l'on peut écrire

$$\frac{\sum_{k_1=1}^{M_1} \sum_{k_2=1}^{M_2} X_{1,k_1}(\omega) X_{2,k_2}(\omega)}{M_1 M_2} = \frac{\sum_{k_1=1}^{M_1} X_{1,k_1}(\omega)}{M_1} \times \frac{\sum_{k_2=1}^{M_2} X_{2,k_2}(\omega)}{M_2}.$$

En construisant les histogrammes de pas h (correspondant respectivement aux $M_1 M_2$ résultats des simulations $X_{1,k_1} X_{2,k_2}$, aux M_1 résultats des simulations X_{1,k_1} , aux M_2 résultats des simulations X_{2,k_2}), puis en calculant les barycentres des kh affectés des coefficients respectifs $N(M_1 M_2, h, k)$, $N(M_1, h, k)$, $N(M_2, h, k)$, enfin en passant à la limite lorsque M_1 et M_2 tendent vers l'infini, on aboutit bien à la formule (2.15). Le cas de deux VA complexes indépendantes Z_1 et Z_2 (aboutissant cette fois à la formule (2.16)) se traite à partir du cas des VAR en utilisant de plus la linéarité de la prise d'espérance. \square

Cette proposition 2.2 s'étend au cas de plusieurs variables aléatoires ($n > 2$) : en règle générale, si X_1, \dots, X_n sont n VAR (*resp.* si Z_1, \dots, Z_n sont n variables aléatoires

complexes, *i.e.* VAC)¹⁰, mutuellement indépendantes et ayant toutes des moments d'ordre n , alors la VAR (*resp.* VAC) $X_1 \cdots X_n$ (*resp.* $Z_1 \cdots Z_n$) a une espérance et l'on a :

$$(2.17) \quad \begin{aligned} E(X_1 \times \cdots \times X_n) &= E(X_1) \times \cdots \times E(X_n) \\ \left(\text{resp. } E(Z_1 \times \cdots \times Z_n) &= E(Z_1) \times \cdots \times E(Z_n) \right). \end{aligned}$$

Plus généralement si X_1, \dots, X_n sont encore n VAR (*resp.* Z_1, \dots, Z_n sont n VAC) mutuellement indépendantes, et que $\Phi_1(X_1), \dots, \Phi_n(X_n)$ (*resp.* $\Phi_1(Z_1), \dots, \Phi_n(Z_n)$) sont des fonctions réelles (*resp.* complexes) de ces VAR (ou VAC) telles que toutes les VAR $\Phi_j(X_j)$ (*resp.* les VAC $\Phi_j(Z_j)$) aient des moments d'ordre n , alors (suivant la Remarque 2.6 qui assure l'indépendance mutuelle des $\Phi_j(X_j)$), on a :

$$(2.18) \quad \begin{aligned} E(\Phi_1(X_1) \times \cdots \times \Phi_n(X_n)) &= E(\Phi_1(X_1)) \times \cdots \times E(\Phi_n(X_n)) \\ \left(\text{resp. } E(\Phi_1(Z_1) \times \cdots \times \Phi_n(Z_n)) &= E(\Phi_1(Z_1)) \times \cdots \times E(\Phi_n(Z_n)) \right). \end{aligned}$$

Mieux encore. Si I_1, \dots, I_K sont des sous-ensembles disjoints de $\{1, \dots, n\}$ et que $\Phi_1(X_{I_1}), \dots, \Phi_K(X_{I_K})$ sont chacune des fonctions de paquets de variables disjoints, *resp.* X_{I_1}, \dots, X_{I_K} , de telle sorte que les VAR

$$\Phi_k(X_{I_k}) = \Phi_k(X_j; j \in I_k), k = 1, \dots, K,$$

aient toutes des moments d'ordre K , alors la VAR

$$\prod_{k=1}^K \Phi_k(X_{I_k}) = \prod_{k=1}^K \Phi_k(X_j; j \in I_k)$$

a une espérance et de plus, toujours si X_1, \dots, X_n sont mutuellement indépendantes, toujours suivant la Remarque 2.6 qui assure l'indépendance mutuelle cette fois des $\Phi_k(X_{I_k})$, on a :

$$(2.19) \quad E \left[\prod_{k=1}^K \Phi_k(X_j; j \in I_k) \right] = \prod_{k=1}^K E(\Phi_k(X_j; j \in I_k)).$$

Le même résultat vaut si les VAR X_1, \dots, X_n sont remplacées par des VAC Z_1, \dots, Z_n .

Voici deux conséquences majeures immédiates de la Proposition 2.2, l'une exprimée en termes de covariance d'un vecteur de VA (réelles ou complexes) indépendantes deux-à-deux, l'autre en termes du calcul de la variance d'une somme de VA (réelles ou complexes) indépendantes deux-à-deux.

PROPOSITION 2.3 (matrice de covariance d'un vecteur de VAR ou de VA complexes indépendantes deux-à-deux). *Soient X_1, \dots, X_n , n VAR toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , supposées indépendantes deux à deux. On a, si \vec{X} désigne le vecteur de VAR (X_1, \dots, X_n)*

$$(2.20) \quad \text{cov}(\vec{X}) = \text{diag}(\text{Var}(X_1), \dots, \text{Var}(X_n)) = \text{diag}(\sigma^2(X_1), \dots, \sigma^2(X_n)).$$

Soient $Z_1 = X_1 + iY_1, \dots, Z_n = X_n + iY_n$, n VA complexes toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , supposées indépendantes deux à deux. On a, si

10. Toutes définies sur même espace probabilisé (Ω, \mathcal{F}, P) correspondant à une épreuve aléatoire dont le hasard est régi par P .

\vec{Z} désigne le vecteur de VA complexes (Z_1, \dots, Z_n)

$$(2.21) \quad \text{cov}(\vec{Z}) = \text{diag}(\text{Var}(Z_1), \dots, \text{Var}(Z_n)) = \text{diag}(\sigma^2(Z_1), \dots, \sigma^2(Z_n)).$$

PROPOSITION 2.4 (variance d'une somme de VA indépendantes deux-à-deux). Soient X_1, \dots, X_n , n VAR toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , supposées indépendantes deux à deux. On a

$$(2.22) \quad \text{Var}(X_1 + \dots + X_n) = \sum_{j=1}^n \text{Var}(X_j).$$

Soient $Z_1 = X_1 + iY_1, \dots, Z_n = X_n + iY_n$, n VA complexes toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , supposées indépendantes deux à deux. On a

$$(2.23) \quad \text{Var}(Z_1 + \dots + Z_n) = \sum_{j=1}^n \text{Var}(Z_j).$$

DÉMONSTRATION. Vérifions juste en exercice la formule (2.22) pour n VAR indépendantes deux-à-deux en utilisant pour commencer la définition de la variance, puis la linéarité de la prise d'espérance, enfin le fait que la matrice de covariance du vecteur (X_1, \dots, X_n) soit diagonale lorsque les VAR X_j sont supposées indépendantes deux-à-deux (c.f. la Proposition 2.20). Cela donne :

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= E[(X_1 + \dots + X_n)^2] - (E[X_1 + \dots + X_n])^2 \\ &= E\left[\sum_{j=1}^n X_j^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j\right] - \sum_{j=1}^n (EX_j)^2 - 2 \sum_{1 \leq i < j \leq n} EX_i EX_j \\ &= \sum_{j=1}^n E(X_j^2) + 2 \sum_{1 \leq i < j \leq n} E(X_i X_j) - \sum_{j=1}^n (EX_j)^2 - 2 \sum_{1 \leq i < j \leq n} EX_i EX_j \\ &= \sum_{j=1}^n E(X_j^2) + 2 \sum_{1 \leq i < j \leq n} EX_i EX_j - \sum_{j=1}^n (EX_j)^2 - 2 \sum_{1 \leq i < j \leq n} EX_i EX_j \\ &= \sum_{j=1}^n E(X_j^2) - \sum_{j=1}^n (EX_j)^2 = \sum_{j=1}^n \text{Var}(X_j). \end{aligned}$$

□

EXEMPLE 2.10 (variance de la loi $\mathcal{B}(N, p)$). Comme une VAR suivant une loi $\mathcal{B}(N, p)$ est somme de N VAR suivant une loi de Bernoulli de paramètre p , la variance de la loi $\mathcal{B}(N, p)$ vaut N fois la variance de la loi $\mathcal{B}(1, p)$, soit $N \times p \times (1-p)$.

EXEMPLE 2.11 (variance empirique d'une VAR). Si X_1, \dots, X_M désignent M simulations indépendantes d'une même VAR X telle que $E(X^2) < +\infty$ (toutes définies, comme X , sur le même espace probabilisé (Ω, \mathcal{F}, P) , on rappelle que la VAR :

$$\frac{X_1 + \dots + X_M}{M}$$

est notée $X_{[M]}^{\text{emp}}$ et appelée *espérance (ou moyenne) empirique* de X . L'espérance de la moyenne empirique est égale à l'espérance de X ; c'est la raison pour laquelle

la moyenne empirique $X_{[M]}^{\text{emp}}$ est appelée *estimateur sans biais* de l'espérance de X . Les variables

$$Y_j := X_j - \frac{X_1 + \cdots + X_M}{M} = X_j - X_{[M]}^{\text{emp}}, \quad j = 1, \dots, M,$$

ne sont pas indépendantes puisque leur somme est nulle. Mais on a

$$\sum_{j=1}^M Y_j^2 = \sum_{j=1}^M X_j^2 - M(X_{[M]}^{\text{emp}})^2.$$

En appliquant la Proposition 2.4 après avoir développé

$$M(X_{[M]}^{\text{emp}})^2 = \frac{\sum_{j=1}^M X_j^2}{M} + 2 \frac{\sum_{1 \leq i < j \leq n} X_i X_j}{M},$$

on constate que

$$(2.24) \quad E\left(\sum_{j=1}^M Y_j^2\right) = (M-1)\sigma^2(X).$$

La variable

$$\text{Var}_{[M]}(X)^{\text{emp}} := \frac{1}{M-1} \sum_{j=1}^M (X_j - X_{[M]}^{\text{emp}})^2$$

est appelée *variance empirique* de la VAR X . Cette terminologie est justifiée par le fait que la formule (2.11) s'écrive aussi :

$$(2.25) \quad E\left[\frac{1}{M-1} \sum_{j=1}^M (X_j - X_{[M]}^{\text{emp}})^2\right] = E\left[\frac{1}{M-1} \sum_{j=1}^M Y_j^2\right] = \text{Var}(X).$$

Cette VAR réalise un *estimateur sans biais* de la variance de la VAR X . Notez bien la division par $M-1$ et non par M (en effet $M-1$, et non M , représente ici le nombre de degrés de liberté indépendants¹¹).

2.2. Le Théorème Limite « Central » (TLC)

Considérons une suite $(X_j)_{j \geq 1}$ de VAR, toutes définies sur le même espace probabilité (Ω, \mathcal{F}, P) , univers des possibles d'une certaine épreuve aléatoire dont le hasard se trouve régi par le choix de la probabilité P .

On suppose que les variables $(X_j)_{j \geq 1}$ sont mutuellement indépendantes, ce qui signifie que pour tout entier $n \in \mathbb{N}^*$, les variables X_1, \dots, X_n sont indépendantes au sens de la Définition 2.5.

On suppose d'autre part que toutes les VAR X_j ont même loi, et que cette loi est une loi admettant des moments d'ordre 2, ce qui implique que les X_j , $j \in \mathbb{N}$, ont toutes une espérance commune $EX = m$ et une variance commune $\text{Var}(X) = \sigma^2$.

Introduisons la suite de variables $(\tilde{X}_j)_{j \geq 1}$, où

$$\forall n \in \mathbb{N}^*, \quad \tilde{X}_n = \frac{X_n - m}{\sigma}.$$

11. Il n'y a pas de « biais » car on divise précisément par le nombre $M-1$ de degrés de liberté, soit le nombre maximal (ici $M-1$) de variables indépendantes parmi Y_1^2, \dots, Y_M^2 .

Les variables $(\tilde{X}_j)_{j \geq 1}$ sont, comme les variables $(X_j)_{j \geq 1}$, mutuellement indépendantes, et toutes de même loi \tilde{X} , cette fois d'espérance $E\tilde{X} = 0$ et $\text{Var}(\tilde{X}) = 1$. La fonction caractéristique de \tilde{X} (introduite pour une VAR en (2.8)) est par définition la fonction

$$\omega \in \mathbb{R} \mapsto E[e^{i\omega X}] = E\left[1 - i\omega X - \frac{\omega^2}{2}X^2 - i\frac{\omega^3}{6}X^3 + \dots\right] = 1 - \frac{\omega^2}{2} + o(\omega^2).$$

Introduisons la suite de variables $(\tilde{Y}_j)_{j \geq 1}$ définie par

$$\tilde{Y}_n = \tilde{X}_{[n]}^{\text{emp}} = \frac{\tilde{X}_1 + \dots + \tilde{X}_n}{n} \quad \forall n \in \mathbb{N}^*$$

(il s'agit bien effet de l'espérance empirique $\tilde{X}_{[n]}^{\text{emp}}$ telle qu'elle a été définie précédemment, cf. la remarque 2.1). On a

$$E[\tilde{Y}_n] = E[\tilde{X}] = 0 \quad \text{et} \quad \sigma^2(\tilde{Y}_n) = \frac{1}{n^2}(\text{Var}(X_1) + \dots + \text{Var}(X_n)) = \frac{n}{n^2} = \frac{1}{n}.$$

Calculons maintenant la fonction caractéristique de $\sqrt{n}\tilde{Y}_n$. Comme

$$\sqrt{n}\tilde{Y}_n = \frac{1}{\sqrt{n}}(\tilde{X}_1 + \dots + \tilde{X}_n)$$

et que les variables $\tilde{X}_1, \dots, \tilde{X}_n$, donc aussi les variables $e^{i\omega\tilde{X}_1/\sqrt{n}}, \dots, e^{i\omega\tilde{X}_n/\sqrt{n}}$ sont mutuellement indépendantes, on a

$$\begin{aligned} \forall \omega \in \mathbb{R}, E[e^{i\omega\tilde{Y}_n}] &= \prod_{j=1}^n E\left[e^{i\omega X/\sqrt{n}}\right]^n = \left(1 - \frac{\omega^2}{2n} + o(\omega^2/n)\right)^n \\ &= \exp\left(-\frac{\omega^2}{2} + o(1)\right) \text{ lorsque } n \rightarrow +\infty. \end{aligned}$$

On constate que la suite des fonctions caractéristiques des VAR $\sqrt{n}\tilde{Y}_n$ converge simplement, lorsque n tend vers l'infini, vers la fonction caractéristique de la loi $\mathcal{N}(0, 1)$, à savoir la fonction $\omega \mapsto e^{-\omega^2/2}$ (qui est bien, en effet, la transformée de Fourier de la densité de la loi normale $\mathcal{N}(0, 1)$). D'après un résultat dû à Paul Lévy, ceci suffit à assurer le résultat suivant :

THEORÈME 2.1 (Théorème Limite « Central¹² » (TLC)). *Soient $(X_j)_{j \geq 1}$ une suite de VAR mutuellement indépendantes et de même loi (ayant des moments d'ordre 2, c'est-à-dire ayant une variance). Pour tout intervalle (T_1, T_2) de \mathbb{R} (les parenthèses signifiant ici que les bornes peuvent être ou non incluses, ces bornes pouvant être finies ou infinies) :*

$$(2.26) \quad \lim_{n \rightarrow +\infty} P\left(\left\{\omega \in \Omega; X_{[n]}^{\text{emp}}(\omega) - m \in \left(T_1 \frac{\sigma}{\sqrt{n}}, T_2 \frac{\sigma}{\sqrt{n}}\right)\right\}\right) = \frac{1}{\sqrt{2\pi}} \int_{(T_1, T_2)} e^{-t^2/2} dt.$$

On dit que la suite de VAR $(\tilde{Y}_n)_{n \geq 1}$ converge en loi vers une VAR suivant une loi $\mathcal{N}(0, 1)$, ce qui signifie que la suite des fonctions de répartition $(F_{\tilde{Y}_n})_{n \geq 1}$ converge vers la fonction de répartition $F_{\mathcal{N}(0,1)}$ d'une loi normale réduite centrée.

12. Parce qu'il est vraiment au cœur de la théorie. C'est aussi la pierre d'angle de l'articulation entre probabilités et statistique, une fois le résultat combiné avec la loi des grands nombres (LGN), que nous énoncerons dans la section suivante (loi qui, sans le théorème limite central, serait de peu d'intérêt en statistique).

REMARQUE 2.7 (retour à la formule de Berry-Esseen (1.6)). Si les X_j , $j \in \mathbb{N}^*$, sont des variables de Bernoulli mutuellement indépendantes, toutes de paramètre p (c'est-à-dire ne prenant que les deux valeurs 0 et 1, avec $P(\{X = 1\}) = p$, $P(\{X = 0\}) = 1 - p$), la variable $NX_{[N]}^{\text{emp}}$ suit, on la vu dans la sous-section 1.2.3, une loi binomiale $\mathcal{B}(N, p)$. L'inégalité de Berry-Esseen (1.6) donne en fait une version quantitative du Théorème Limite Central dans ce cas particulier.

Passons maintenant, une fois armés de ce théorème, à des considérations, qui, on le verra, soutendront ultérieurement la démarche statistique. Soit $\alpha \in]0, 1[$. Puisque la gaussienne

$$t \mapsto \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$$

(densité de la loi $\mathcal{N}(0, 1)$) est une fonction positive, paire, d'intégrale égale à 1, il existe un unique nombre¹³ $x_\alpha > 0$ tel que :

$$\frac{1}{\sqrt{2\pi}} \int_{|t| \geq x_\alpha} e^{-t^2/2} dt = 1 - \alpha.$$

On a

$$\lim_{\alpha \rightarrow 0_+} x_\alpha = +\infty.$$

Si l'on introduit la fonction de quantile $F_{\mathcal{N}(0,1)}^{\leftarrow}$ de la distribution de probabilité associée à la loi normale $\mathcal{N}(0, 1)$ (cf. la Définition 1.6), on constate d'ailleurs (pour des raisons de parité) que

$$x_\alpha = F_{\mathcal{N}(0,1)}^{\leftarrow}(1 - \alpha/2).$$

Le théorème limite central assure donc que, si $x \geq x_\alpha$,

$$(2.27) \quad \begin{aligned} & \lim_{n \rightarrow +\infty} P\left(\left\{\omega \in \Omega; m \in \left]X_{[n]}^{\text{emp}}(\omega) - x \frac{\sigma}{\sqrt{n}}, X_{[n]}^{\text{emp}}(\omega) + x \frac{\sigma}{\sqrt{n}}\right[\right\}\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-x}^x e^{-t^2/2} dt \geq 1 - \alpha. \end{aligned}$$

Si l'on sait *a priori* que la loi des VAR X_j , $j \in \mathbb{N}^*$ a une variance σ connue¹⁴ (mais par contre une espérance m inconnue), alors, pour le choix d'un tel $x > x_\alpha$, les deux variables aléatoires :

$$\begin{aligned} T_{x,n,\min} &:= \frac{X_1 + \cdots + X_n}{n} = X_{[n]}^{\text{emp}} - x \frac{\sigma}{\sqrt{n}} \\ T_{x,n,\max} &:= \frac{X_1 + \cdots + X_n}{n} = X_{[n]}^{\text{emp}} + x \frac{\sigma}{\sqrt{n}} \end{aligned}$$

délimitent un intervalle aléatoire $]T_{x,n,\min}, T_{x,n,\max}[$ tel que, pour n suffisamment grand,

$$(2.28) \quad P\left(\left\{\omega \in \Omega; m \in \left]T_{x,n,\min}(\omega), T_{x,n,\max}(\omega)\right[\right\}\right) \geq 1 - \alpha.$$

13. Dont on trouve un encadrement en utilisant une table numérique pour la fonction de répartition $T \mapsto \int_{-\infty}^T e^{-t^2/2} dt / \sqrt{2\pi}$ de la loi normale $\mathcal{N}(0, 1)$.

14. Bien que ceci soit assez utopiste, faisons avec pour l'instant. Nous verrons ultérieurement ce qu'il faut envisager de faire si moyenne et variance sont toutes deux *a priori* inconnues.

DÉFINITION 2.6 (intervalle de confiance au niveau de confiance $1 - \alpha$ (ou au risque α) pour une moyenne). Soit (Ω, \mathcal{T}, P) un espace probabilisé (correspondant à une épreuve aléatoire, le hasard étant régi par la distribution de probabilité P sur l'univers des possibles) et $X : \Omega \rightarrow \mathbb{R}$ une VAR d'espérance m . Soit $\alpha \in]0, 1[$. Un intervalle aléatoire $I(\omega)$ tel que,

$$P(\{\omega \in \Omega; m \in I(\omega)\}) \geq 1 - \alpha$$

est appelé *intervalle de confiance* pour la moyenne de la VAR X au *niveau de confiance* $1 - \alpha$ (ou encore au *risque* α , ou aussi au *facteur de risque* α près).

On déduit alors du théorème limite central le corollaire suivant :

COROLLAIRE 2.2 (intervalle de confiance asymptotique pour la moyenne, à variance connue). Soit $X : \Omega \rightarrow \mathbb{R}$ une VAR sur un espace probabilisé (Ω, \mathcal{T}, P) correspondant à une certaine épreuve aléatoire dont le hasard est régi par la distribution de probabilité P . Soit $\alpha \in]0, 1[$ et $x > x_\alpha = F_{\mathcal{N}(0,1)}^{-1}(1 - \alpha/2)$. Alors, pour n suffisamment grand, l'intervalle aléatoire

$$I_{x,n} := \left] X_{[n]}^{\text{emp}} - x \frac{\sigma}{\sqrt{n}}, X_{[n]}^{\text{emp}} + x \frac{\sigma}{\sqrt{n}} \right],$$

où l'espérance empirique $X_{[n]}^{\text{emp}}$ est la VAR

$$X_{[n]}^{\text{emp}} := \frac{X_1 + \cdots + X_n}{n},$$

les $X_j, j = 1, \dots, n$, figurant n simulations indépendantes de X , est un intervalle de confiance pour l'espérance $E(X)$ au niveau de confiance $1 - \alpha$ (ou encore au risque α près). L'intervalle limite

$$\overline{I_{x_\alpha, n}} = \left[X_{[n]}^{\text{emp}} - x_\alpha \frac{\sigma}{\sqrt{n}}, X_{[n]}^{\text{emp}} + x_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

est dit *intervalle de confiance asymptotique* pour m au niveau de confiance $1 - \alpha$. Les deux VAR correspondant aux bornes inférieure et supérieure de $\overline{I_{x_\alpha, n}}$ sont dites *estimateurs*¹⁵ respectivement *inférieur* ou *supérieur* de la moyenne m .

2.3. Les lois des grands nombres (LGNf) et (LGNF)

2.3.1. Version faible (LGNf). Soient $(X_j)_{j \geq 1}$ une suite de VAR, toutes définies sur un même espace probabilisé (Ω, \mathcal{T}, P) correspondant à une certaine expérience aléatoire dont le hasard est régi par P .

On suppose cette fois les VAR $(X_j)_{j \geq 1}$ seulement deux à deux indépendantes (on aura dans un second temps besoin de l'indépendance mutuelle, plus contraignante).

On suppose d'autre part que toutes les VAR X_j ont même loi, et que cette loi (on note X une VAR la suivant) est une loi admettant des moments d'ordre 2, ce qui implique que les $X_j, j \in \mathbb{N}$, ont toutes une espérance commune $EX = m$ et une variance commune $\text{Var}(X) = \sigma^2$.

On a, du fait de l'indépendance deux-à-deux des X_j :

$$\text{Var}[X_{[n]}^{\text{emp}}] = \text{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n^2} \times n \text{Var}(X) = \frac{\text{Var}(X)}{n}.$$

¹⁵. Nous reviendrons plus précisément ultérieurement sur cette terminologie dans le contexte statistique.

L'inégalité de Bienaymé-Tchebychev (Proposition 2.1) assure alors que, pour tout $\epsilon > 0$, on a :

$$(2.29) \quad P\left(\left\{\omega \in \Omega; |X_{[n]}^{\text{emp}}(\omega) - EX| \geq \epsilon\right\}\right) \leq \frac{\sigma^2(X)}{n\epsilon^2}.$$

On peut donc énoncer le résultat suivant :

THEORÈME 2.3 (loi faible des grands nombres (LGNf), version avec variance). *Soient $(X_j)_{j \geq 1}$ une suite de VAR, toutes définies sur le même espace probabilisé (Ω, \mathcal{P}) attaché à une épreuve aléatoire dont P régit le hasard, supposées deux-à-deux indépendantes et de même loi (ayant des moments d'ordre 2, c'est-à-dire ayant une variance). On a :*

$$(2.30) \quad \forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} P\left(\left\{\omega \in \Omega; |X_{[n]}^{\text{emp}}(\omega) - EX| \geq \epsilon\right\}\right) = 0.$$

On dit aussi que la suite de VAR $(X_{[n]}^{\text{emp}})_{n \geq 1}$, où

$$X_{[n]}^{\text{emp}} := \frac{X_1 + \dots + X_n}{n}$$

désigne l'espérance empirique, converge en probabilité vers la VAR constante égale à l'espérance EX de X .

REMARQUE 2.8 (la convergence en probabilité est plus forte que la convergence en loi). La convergence en probabilité d'une suite de VAR $(U_n)_{n \geq 1}$ vers une VAR limite U implique la convergence en loi (i.e. la suite $(F_{U_n})_{n \geq 1}$ des fonctions de répartition des U_n converge simplement vers la fonction de répartition F_U de U). La réciproque est en fait fautive. La loi faible des grands nombres est donc de fait un résultat puissant. C'est elle qui est de fait utile en statistique, une fois avantageusement couplée, comme on le verra, avec le théorème limite central (cf. le théorème de Slutsky dans la section suivante).

2.3.2. Version forte. Lorsque la suite des VAR $(X_j)_{j \geq 1}$ est une suite de VAR (toujours de même loi, dite loi d'une VAR X , et admettant une variance) non plus deux-à-deux indépendantes, mais cette fois mutuellement indépendantes, on peut donner une version plus forte de l'inégalité de Bienaymé-Tchebychev (2.13).

PROPOSITION 2.5 (inégalité de Kolmogorov). *Soit $(X_j)_{j \geq 1}$ une suite de VAR, toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) attaché à une épreuve aléatoire dont P régit le hasard, supposées mutuellement indépendantes et de même loi (ayant des moments d'ordre 2, c'est-à-dire une variance). On a :*

$$(2.31) \quad \forall n \in \mathbb{N}^*, P\left(\left\{\omega \in \Omega; \max_{1 \leq k \leq n} |k(X_{[k]}^{\text{emp}} - EX)| \geq \epsilon\right\}\right) \leq \frac{nV(X)}{\epsilon^2}.$$

Cette inégalité¹⁶ implique la *loi forte des grands nombres* :

THEORÈME 2.4 (loi forte des grands nombres (LGNF)). *Soit $(X_j)_{j \geq 1}$ une suite de VAR, toutes définies sur le même espace probabilisé (Ω, \mathcal{F}, P) attaché à une*

¹⁶. Nous l'admettrons ici. C'est un joli exercice sur le conditionnement et la formule des probabilités totales, mais un peu au delà du niveau de ce cours. Pour une preuve, on pourra par exemple la section 2.9.3 de [Ypin].

épreuve aléatoire dont P régit le hasard, supposées mutuellement indépendantes et de même loi (ayant des moments d'ordre 2, c'est-à-dire une variance). Alors

$$(2.32) \quad P\left(\left\{\omega \in \Omega; \lim_{n \rightarrow +\infty} X_{[n]}^{\text{emp}}(\omega) = EX\right\}\right) = 1.$$

On dit aussi que la suite de VAR $(X_{[n]}^{\text{emp}})_{n \geq 1}$ converge presque sûrement vers la VAR constante égale à EX .

REMARQUE 2.9. Une suite $(U_n)_{n \geq 1}$ de VAR sur un espace probabilisé (Ω, \mathcal{F}, P) converge presque sûrement vers une VAR U si et seulement si l'ensemble des $\omega \in \Omega$ tels que la suite $(U_n(\omega))_{n \geq 1}$ ne converge pas vers $U(\omega)$ est un évènement aléatoire A de probabilité nulle ($P(A) = 0$). Il est facile de voir que la convergence presque sûre implique la convergence en probabilité qui, elle-même, on l'a vu (Remarque 2.8), implique la convergence en loi. La loi forte des grands nombres (LGNF) implique bien la loi faible (LGNf), ce qui est donc cohérent.

REMARQUE 2.10. Le fait que X ait des moments d'ordre deux (*i.e.* une variance) n'est pas en fait nécessaire. Il suffit juste que X ait une espérance (ce qui est indispensable bien sûr pour formuler le théorème).

REMARQUE 2.11. La loi des grands nombres, qui semble à première vue, un résultat spectaculaire, n'est cependant qu'assez peu utilisée en statistique. Au contraire de la loi faible (qui, elle, repose, dans le cas où X a une variance, sur un résultat quantitatif, à savoir l'inégalité (2.29)), la loi forte n'est en effet pas aussi facilement quantifiable (on y parvient toutefois partiellement grâce à l'inégalité de Kolmogorov, comme indiqué dans la preuve que nous indiquons ici précisément à cet effet).

DÉMONSTRATION. On admet ici l'inégalité de Kolmogorov. Soit $\epsilon > 0$. Pour $n \geq 1$, on introduit l'évènement aléatoire :

$$B_n(\epsilon) := \left\{\omega \in \Omega; \sup_{1 \leq k \leq n} |X_{[k]}^{\text{emp}}(\omega) - EX| \geq \epsilon\right\}.$$

Pour tout $j \in \mathbb{N}$, on remarque en découpant convenablement cet évènement aléatoire, que :

$$\begin{aligned} P(B_{2^j}(\epsilon)) &\leq \sum_{l=l}^{\infty} P\left(\left\{\omega \in \Omega; \sup_{2^l \leq k \leq 2^{l+1}} |X_{[k]}^{\text{emp}}(\omega) - EX| \geq \epsilon\right\}\right) \\ &\leq \sum_{l=l}^{\infty} P\left(\left\{\omega \in \Omega; \sup_{1 \leq k \leq 2^{l+1}} |X_{[k]}^{\text{emp}}(\omega) - EX| \geq 2^l \epsilon\right\}\right) \\ &\leq \sum_{l=j}^{\infty} \frac{2^{l+1} \text{Var}(X)}{2^{2l} \epsilon^2} = 4 \frac{\text{Var}(X)}{2^j}. \end{aligned}$$

On en déduit donc

$$\lim_{j \rightarrow +\infty} P(B_{2^j}(\epsilon)) = 0.$$

Ceci démontre la loi forte des grands nombres sans toutefois donner de résultat tout à fait satisfaisant du point de vue quantitatif. \square

Initiation au raisonnement statistique

L'objet principal de la statistique est l'analyse de données, au travers de la recherche de lois générales de probabilité qui en régissent la répartition (du point de vue stochastique). Les lois de probabilité sont ainsi considérées « en famille », c'est-à-dire dépendant de paramètres (famille de lois binomiales indexées par un paramètre p , famille de lois de Poisson de paramètres indexées par un paramètre λ , famille de lois gaussiennes indexées par des paramètres μ et σ^2 , *etc.*). Les données dont on dispose sont confrontées au « dictionnaire » des lois de probabilité envisagées (dictionnaire précisément constitué par cette famille), puis les paramètres sont ensuite « ajustés » à ces données. L'analyse des données consiste en ce que l'on appelle la *statistique descriptive*, tandis que la confrontation de ces données à des modèles probabilistes considérés en famille (et indexés par des paramètres) et l'ajustement optimum des paramètres consiste en ce que l'on appelle la *statistique inférentielle*.

Du fait que la statistique implique l'étude des probabilités « en famille », il est logique de ne parler de raisonnement statistique qu'une fois assis, comme nous l'avons fait dans ce cours, le cadre rigoureux de la théorie des probabilités. Nous ne ferons dans ce bref chapitre qu'effleurer quelques aspects inhérents au raisonnement statistique.

3.1. Indicateurs numériques d'une série de données statistiques

Étant donnée une série de données $\{x_1, \dots, x_N\}$, on peut la présenter sous forme d'un unique tableau (on parle alors de *série brute de données*) ou l'organiser suivant les valeurs prises par un certain caractère ; lorsque les valeurs prises par ce caractère décrivent un ensemble discret, on parle alors de *tableau d'effectifs par valeur*¹, tandis que lorsque les valeurs prises par ce caractère décrivent un ensemble continu, on parle de *tableau d'effectifs par classe*².

DÉFINITION 3.1 (moyenne d'une série de données statistiques). Soit

$$x = \{x_1, \dots, x_M\}$$

une série de données statistiques. On appelle *moyenne* de cette série la quantité

$$(3.1) \quad \bar{x} = \frac{1}{M} \sum_{j=1}^M x_j.$$

1. Il s'agit des valeurs prises par le caractère (dans un ensemble de nombres discret, tel \mathbb{N} , \mathbb{Z} , *etc.*).

2. Les classes sont ici les intervalles en lesquelles se trouve découpé l'ensemble des valeurs prises par le caractère (par exemple en général un intervalle de \mathbb{R}).

si les x_j sont des données chiffrées constituant une série brute. Si les x_j sont des données organisées suivant un tableau d'effectifs par valeur ξ_j prise par un certain caractère, la *moyenne* est définie par :

$$(3.2) \quad \bar{x} = \frac{1}{M} \sum_{j=1}^M M_j \xi_j, \quad M_j := \text{card} \{j; \text{caract}(x_j) = \xi_j\}.$$

Si les x_j sont des données organisées suivant un tableau d'effectifs par classe de valeurs prises par un certain caractère, la *moyenne* est définie par :

$$(3.3) \quad \bar{x} = \frac{1}{M} \sum_{j=1}^M M_j \bar{\xi}_j,$$

$$M_j := \text{card} \{j; \text{caract}(x_j) \in \text{classe } \xi_j\}, \quad \bar{\xi}_j = \text{milieu}(\text{classe}(\xi_j)).$$

REMARQUE 3.1 (moyenne et espérance empirique). La notion de moyenne en statistique constitue le pendant de la notion d'espérance empirique d'une VAR en probabilité, telle qu'elle a été introduite au chapitre 2 (*cf.* (2.3), sous-section 2.1.2).

DÉFINITION 3.2 (écart type d'une série de données statistique). Soit

$$x = \{x_1, \dots, x_M\}$$

une série de données statistique. On appelle *écart type* de cette série la quantité s définie par

$$(3.4) \quad s^2 = \frac{1}{M-1} \sum_{j=1}^M (x_j - \bar{x})^2,$$

où \bar{x} est défini par (3.1), si les x_j sont des données chiffrées constituant une série brute. Si les x_j sont des données organisées suivant un tableau d'effectifs par valeur ξ_j prise par un certain caractère, l'*écart type* de la série est défini par :

$$(3.5) \quad s^2 = \frac{1}{M-1} \sum_{j=1}^M M_j (\xi_j - \bar{x})^2, \quad M_j := \text{card} \{j; \text{caract}(x_j) = \xi_j\},$$

où \bar{x} est cette fois défini par (3.2). Si les x_j sont des données organisées suivant un tableau d'effectifs par classe de valeurs prises par un certain caractère, l'*écart type* de la série est défini par :

$$(3.6) \quad \bar{x} = \frac{1}{M} \sum_{j=1}^M M_j (\bar{\xi}_j - \bar{x})^2,$$

$$M_j := \text{card} \{j; \text{caract}(x_j) \in \text{classe } \xi_j\}, \quad \bar{\xi}_j = \text{milieu}(\text{classe}(\xi_j)),$$

où \bar{x} est défini par (3.3).

REMARQUE 3.2 (écart type et racine de la variance empirique). La notion d'écart type en statistique constitue le pendant de la notion d'écart type empirique (*i.e.* racine de la *variance empirique*) d'une VAR en probabilité, telle qu'elle a été introduite au chapitre 2 (*cf.* exemple 2.11).

DÉFINITION 3.3 (médiane d'une série brute organisée par ordre croissant). Si $\{x_1, \dots, x_M\}$ est une série brute de données numériques organisée de manière croissante (*i.e.* $x_j \leq x_{j+1}$ pour $j = 1, \dots, M - 1$), on appelle *médiane* de cette série la quantité :

$$(3.7) \quad m = \begin{cases} \frac{x_p + x_{p+1}}{2} & \text{si } M = 2p \\ x_{p+1} & \text{si } M = 2p + 1. \end{cases}$$

3.2. La notion d'estimateur ; des probabilités à la statistique

Dans cette section, nous revenons sur le point de vue des probabilités développé aux chapitres 1 et 2.

3.2.1. Estimateur d'un paramètre et notions afférentes. Soit X une VAR (sur un espace probabilité (Ω, \mathcal{F}, P)) dont la loi (inconnue, mais de type connu) dépend d'un paramètre θ ; par exemple, comme c'est le cas dans les populations sondées à la veille d'un scrutin à deux choix, X peut être une variable de Bernoulli de paramètre $p = \theta$; ce peut être aussi, comme c'est le cas dans l'analyse du rayonnement d'un tissu organique, une loi de Poisson de paramètre la densité du tissu $\lambda = \theta$; on pourrait multiplier ainsi les exemples.

Sous l'angle de la théorie des probabilités, un *échantillon* de taille M de X est par définition la donnée de M variables aléatoires X_1, \dots, X_M , mutuellement indépendantes et de même loi, à savoir la loi de la variable (inconnue) X .

Un *estimateur* du paramètre θ est par définition la donnée, pour chaque M , d'une VAR T_M s'écrivant comme une fonction (déterministe) des variables X_1, \dots, X_M ($T_M = f_M(X_1, \dots, X_M)$). Une réalisation $(x_1, \dots, x_M) = (X_1(\omega), \dots, X_M(\omega))$ de (X_1, \dots, X_M) (dite aussi *échantillon expérimental* ou *échantillon observé*) fournit, par le biais de $t_M := f_M(x_1, \dots, x_M)$, ce que l'on appelle une *estimation à l'ordre M de θ* .

Un tel estimateur est dit *converger vers θ* si la suite de VAR $(T_M)_{M \geq 1}$ converge en probabilité vers la constante θ lorsque M tend vers $+\infty$, ce qui signifie :

$$\forall \epsilon > 0, \quad \lim_{M \rightarrow +\infty} P(|T_M - \theta| \geq \epsilon) = 0.$$

L'estimateur $(T_M)_{M \geq 1}$ de θ est dit *sans biais* si $E(T_M) = \theta$ pour tout $M \geq 1$. On dit qu'un estimateur $(T_M)_{M \geq 1}$ de θ est *absolument correct* s'il est sans biais et converge vers θ . Un estimateur $(T_N)_{M \geq 1}$ de θ est dit *efficace* s'il est absolument correct et de plus tel que la quantité $E(|T_N - \theta|^2)$ (dite aussi *risque quadratique*) soit minimale parmi toutes les quantités

$$E(|S_M - \theta|^2),$$

où $(S_M)_{M \geq 1}$ est un estimateur quelconque absolument correct de θ . Parmi les estimateurs absolument corrects du paramètre θ , les estimateurs efficaces sont donc ceux qui minimisent le risque quadratique.

3.2.2. Exemples classiques d'estimateurs.

EXEMPLE 3.1 (estimation de l'espérance). Supposons que le paramètre à estimer de la variable aléatoire inconnue X soit son espérance $E(X)$ (X étant supposée

avoir une variance). Dans ce cas, on dispose d'un estimateur absolument correct pour $E(X)$, à savoir l'espérance empirique

$$X_{[M]}^{\text{emp}} ::= \frac{X_1 + \cdots + X_M}{M}.$$

Le fait que cet estimateur converge vers $E(X)$ résulte, on l'a vu, de la loi faible des grands nombres (LGNf, Théorème 2.3).

EXEMPLE 3.2 (estimation de la variance lorsque l'espérance $m = E(X)$ est connue). Supposons que le paramètre à estimer de la variable aléatoire inconnue X soit sa variance $\sigma^2(X)$ (X étant supposée avoir une variance), l'espérance $E(X) = m$ étant une quantité connue. Un estimateur sans biais de la variance est fourni (si X admet des moments d'ordre 4, *i.e.* $E(X^4) < \infty$) par :

$$T_M := [(X - m)^2]_{[M]}^{\text{emp}} = \frac{\sum_{j=1}^M (X_j - m)^2}{M}$$

puisque la variance de X est l'espérance de $(X - m)^2$. Ceci résulte du fait que l'espérance empirique réalise un estimateur absolument correct de l'espérance du fait de la loi faible des grands nombres (LGNf, Théorème 2.3).

EXEMPLE 3.3 (estimation de la variance lorsque l'espérance est inconnue). Si X est une VAR admettant un moment d'ordre 2, la variance empirique

$$\text{Var}_{[M]}(X)^{\text{emp}} := \frac{1}{M-1} \sum_{j=1}^M (X_j - X_{[M]}^{\text{emp}})^2$$

fournit (*cf.* l'exemple 2.11) un estimateur absolument correct de la variance de X . En effet, il s'agit d'une part d'un estimateur sans biais (*cf.* les calculs faits dans l'exemple 2.11). La loi forte des grands nombres (LGNF, Théorème 2.4) implique d'autre part la convergence presque sûre (donc la convergence en probabilité) de la suite $(\text{Var}_{[M]}(X)^{\text{emp}})_{M \geq 1}$ vers $\sigma^2(X)$.

EXEMPLE 3.4 (estimation du coefficient de corrélation entre deux variables X et Y). Si X et Y sont deux variables aléatoires réelles définies sur le même espace probabilisé et ayant toutes les deux une variance, le coefficient de régression (on dit aussi *de corrélation*) de X et de Y est défini par

$$\rho(X, Y) = \text{corr}(X, Y) := \frac{E[(X - E(X))(Y - E(Y))]}{\sigma(X)\sigma(Y)}.$$

Un estimateur de ce coefficient de corrélation est donné par

$$C_{X,Y,M} := \frac{1}{M-1} \frac{\sum_{j=1}^M (X_j - X_{[M]}^{\text{emp}})(Y_j - Y_{[M]}^{\text{emp}})}{\sqrt{\text{Var}_{[M]}(X)^{\text{emp}}}\sqrt{\text{Var}_{[M]}(Y)^{\text{emp}}}}.$$

3.3. Estimation par intervalle ; gaussiennes et test de Student

3.3.1. Intervalle d'acceptation au risque α , intervalle de confiance.

Supposons que X soit une variable aléatoire réelle dépendant d'un paramètre θ et que $(T_M)_{M \geq 1}$ définisse un estimateur convergeant vers θ . La loi de T_M , $M \geq 1$, dépend du paramètre θ (que l'on ne connaît pas). Pour cerner ce paramètre, on introduit la notion *intervalle de confiance au niveau de confiance* α (ou encore

facteur de risque α). Cette notion généralise celle introduite à la Définition 2.6 dans le cas particulier où $\theta = E(X)$.

DÉFINITION 3.4 (intervalle de confiance). On appelle *intervalle de confiance au niveau de confiance* $1 - \alpha$ ($\alpha \in]0, 1[$) pour un paramètre θ attaché à une VAR définie sur un espace probabilisé (Ω, \mathcal{F}, P) tout intervalle aléatoire $I(\omega)$ tel que

$$P(\{\omega \in \Omega; \theta \in I(\omega)\}) \geq 1 - \alpha.$$

Le fait de disposer d'un estimateur convergeant $(T_M)_{M \geq 1}$ vers le paramètre θ permet d'introduire également une notion déterministe cette fois, la notion d'*intervalle d'acceptation de T_M au risque α* (α étant un seuil entre 0 et 1); un tel intervalle (déterministe) $I_{M,\alpha}$ ou encore *au facteur de risque α* près, est un intervalle contenant θ et sujet à la condition :

$$P(\{T_M \in I_{M,\alpha}\}) = 1 - \alpha.$$

Un tel intervalle d'acceptation de T_M au risque α n'est pas unique et de plus dépend de θ . Par contre, si $\beta \in [0, 1]$, l'intervalle $[t_{1,\alpha,\beta}(\theta), t_{2,\alpha,\beta}(\theta)]$ défini par les deux conditions

$$\begin{aligned} P(\{T_M < t_{1,\alpha,\beta}\}) &= \alpha\beta \\ P(\{T_M > t_{2,\alpha,\beta}\}) &= (1 - \alpha)\beta \end{aligned}$$

est, pour M assez grand, un intervalle d'acceptation de T_M au risque α parfaitement déterminé (mais dépendant toujours bien sûr de M et de θ). On se limitera souvent au cas $\beta = 1/2$ en disant que ceci revient à choisir l'intervalle d'acceptation de T_M au risque α dans une configuration où les risques sont également partagés.

Si $I_M = [\theta - \rho_{M,\alpha,1}, \theta + \rho_{M,\alpha,2}]$ est un intervalle d'acceptation de T_M au risque α , l'intervalle aléatoire cette fois :

$$[T_M - \rho_{M,\alpha,2}, T_M + \rho_{M,\alpha,1}]$$

est un intervalle de confiance pour θ au niveau de confiance $1 - \alpha$. En effet, on a :

$$\left(\theta \in [T_M - \rho_{M,\alpha,2}, T_M + \rho_{M,\alpha,1}]\right) \iff \left(T_M \in [\theta - \rho_{M,\alpha,1}, \theta + \rho_{M,\alpha,2}]\right).$$

Une réalisation $t_M = T_M(\omega)$ de T_M induit une réalisation de l'intervalle de confiance correspondant. Lorsque l'intervalle d'acceptation

$$I_M =]\theta - \rho_{M,\alpha,1}, \theta + \rho_{M,\alpha,2}[$$

est l'intervalle d'acceptation de T_M (au niveau de confiance $1 - \alpha$) dans la configuration où les risques sont partagés, l'intervalle aléatoire de confiance correspondant est appelé *intervalle de confiance aux risques partagés* pour θ (au niveau de confiance $1 - \alpha$).

REMARQUE 3.3 (notion de fourchette de vraisemblance). Un intervalle de confiance aux risques partagés (au niveau de confiance $1 - \alpha$) est aussi appelé *fourchette de vraisemblance*. On dit souvent (mais ce n'est qu'une formulation heuristique) pour définir une telle fourchette de vraisemblance : « le paramètre estimé θ appartient à la fourchette avec une probabilité (équilibrée) $1 - \alpha$ ».

EXEMPLE 3.5 (espérance empirique dans le cadre gaussien : un estimateur de l'espérance lorsque la variance est connue). Supposons que X suive une loi $\mathcal{N}(\mu, \sigma^2)$ et que la variance σ soit connue. Considérons l'estimateur de μ à l'ordre M :

$$X_{[M]}^{\text{emp}} = \frac{1}{M} \sum_{j=1}^M X_j.$$

La variable $X_{[M]}^{\text{emp}}$ est une somme de M variables mutuellement indépendantes de loi $\mathcal{N}(\mu/M, \sigma^2/M^2)$. La somme de deux VAR gaussiennes indépendantes de moyennes respectives μ_1 et μ_2 et d'écart type respectifs σ_1 et σ_2 est une VAR gaussienne de moyenne $\mu_1 + \mu_2$ et d'écart type $\sqrt{\sigma_1^2 + \sigma_2^2}$. La variable $X_{[M]}^{\text{emp}}$, qui est une somme de M gaussiennes indépendantes de moyenne μ/M et d'écart type σ/M , est donc une variable gaussienne de moyenne μ et d'écart type σ/\sqrt{M} . Si l'on choisit $\alpha = 5\%$, l'intervalle d'acceptation de l'estimateur $T_M = X_{[M]}^{\text{emp}}$ au risque α près (les risques étant partagés) est l'intervalle $[t_1(\mu), t_2(\mu)]$ donné par les conditions :

$$\begin{aligned} P\left(N < \frac{t_1(\mu) - \mu}{\sigma/\sqrt{M}}\right) &= 0.025 = 2.5/100 \\ P\left(N > \frac{t_2(\mu) - \mu}{\sigma/\sqrt{M}}\right) &= 0.025 = 2.5/100, \end{aligned}$$

où N désigne une variable aléatoire suivant une loi $\mathcal{N}(0, 1)$. En se référant à une table de la loi $\mathcal{N}(0, 1)$, on voit que

$$\begin{aligned} t_1(\mu) &= \mu - 1.96 \frac{\sigma}{\sqrt{M}} \\ t_2(\mu) &= \mu + 1.96 \frac{\sigma}{\sqrt{M}}. \end{aligned}$$

L'intervalle de confiance correspondant est :

$$\left[\frac{X_1 + \dots + X_M}{M} - 1.96 \frac{\sigma}{\sqrt{M}}, \frac{X_1 + \dots + X_M}{M} + 1.96 \frac{\sigma}{\sqrt{M}} \right].$$

Si M tend vers $+\infty$, σ/\sqrt{M} tend vers 0 et le diamètre de l'intervalle d'acceptation au niveau de confiance 5% et aux risques partagés (comme celui de l'intervalle de confiance correspondant) tend vers 0, ce qui montre que l'estimation de μ (à σ connu) est d'autant plus précise que M est grand.

3.3.2. Les lois du χ^2 et de Student. Nous introduisons ici deux lois de probabilité continues nouvelles, qui nous permettront d'enrichir notre vivier d'estimateurs pour la moyenne dans le cadre Gaussien.

DÉFINITION 3.5 (loi du chi-deux). Une variable aléatoire réelle suit une *loi du χ^2 (chi-deux)* à p degrés de liberté si elle se réalise comme une somme

$$\chi_p^2 = \sum_{j=1}^p Z_j^2,$$

où Z_1, \dots, Z_p sont p variables de loi $\mathcal{N}(0, 1)$ mutuellement indépendantes. Une telle variable est une variable à densité, de densité

$$f_p(t) = \frac{1}{2^{p/2} \Gamma(p/2)} \exp(-t/2) t^{p/2-1}, \quad p \in \mathbb{N},$$

où

$$\Gamma : x \in \mathbb{R}_{>0} \mapsto \int_0^\infty t^{x-1} e^{-t} dt$$

désigne la fonction Γ (interpolant les valeurs de la fonction factorielle, suivant l'équation fonctionnelle $\Gamma(x+1) = x\Gamma(x)$, avec $\Gamma(1) = 1$ et $\Gamma(1/2) = \sqrt{\pi}$). L'espérance d'une telle variable aléatoire est p , sa variance vaut $2p$.

DÉFINITION 3.6 (loi de Student). Une variable aléatoire réelle suit une *loi de Student* à p degrés de liberté si elle se réalise sous la forme

$$X = \frac{Z}{\sqrt{\frac{\chi_p^2}{p}}},$$

où Z suit une loi $\mathcal{N}(0, 1)$ et χ_p^2 une loi du chi-deux à p degrés de liberté, les variables Z et χ_p^2 étant supposées indépendantes.

EXEMPLE 3.6 (variance empirique dans le cadre gaussien). Si X est une variable suivant une loi $\mathcal{N}(\mu, \sigma^2)$, X_1, \dots, X_M M variables mutuellement indépendantes de loi la loi de X , alors l'estimateur canonique de σ (l'espérance μ n'étant pas supposée connue) est, on le sait, la racine carrée de la variance empirique :

$$\sqrt{\text{Var}(X)_{[M]}^{\text{emp}}} = \sqrt{\frac{1}{M-1} \sum_{j=1}^M (X_j - X_{[M]}^{\text{emp}})^2}.$$

La variable

$$\frac{X_{[M]}^{\text{emp}} - \mu}{\sqrt{\text{Var}_{[M]}(X)^{\text{emp}}/M}}$$

suit alors une loi de Student à $M-1$ degrés de liberté.

3.3.3. Le test de Student. Supposons que X suive une loi $\mathcal{N}(\mu, \sigma^2)$, où μ et σ sont deux paramètres inconnus ; la variable

$$S_{M-1} := \frac{X_{[M]}^{\text{emp}} - \mu}{\sqrt{\text{Var}_{[M]}(X)^{\text{emp}}/M}}$$

suit (voir l'exemple 3.6) une loi de Student à $M-1$ degrés de liberté (ce quelque soit la valeur de σ). On exploite alors ce fait crucial pour déterminer un intervalle de confiance pour le calcul de moyenne. Supposons que α soit un seuil de risque et que τ_α soit tel que

$$P(\{\omega \in \Omega; |S_{M-1}(\omega)| > \tau_\alpha\}) = \alpha$$

(on se reporte à une table de loi de Student à $M-1$ paramètres pour la détermination de τ_α). Si t_M est l'estimation

$$t_M = \frac{x_1 + \dots + x_M}{M}$$

de la moyenne de X à partir d'un échantillon observé (réalisation de la moyenne empirique de X) et

$$s_M = \sqrt{\frac{1}{M-1} \sum_{j=1}^M (x_j - t_M)^2}$$

celle de l'écart-type de X (réalisation de la racine carrée de la variance empirique de X) l'intervalle de confiance au risque α (les risques étant partagés) pour la moyenne μ est l'intervalle constitué des nombres x tels que

$$\left| \frac{t_M - x}{\frac{s_M}{\sqrt{M}}} \right| \leq \tau_\alpha,$$

c'est-à-dire l'intervalle :

$$\left[t_M - \tau_\alpha \frac{s_M}{\sqrt{M}}, t_M + \tau_\alpha \frac{s_M}{\sqrt{M}} \right].$$

3.3.4. Intervalles de confiance et théorème limite central. On reprend dans cette sous-section (mais cette fois à la lumière de ce chapitre dédié à l'estimation par intervalle) les développements de la sous-section 2.1. Soit X une variable de Bernoulli (liée par exemple à un certain évènement A par $X(\omega) = 1$ si $\omega \in A$, $X(\omega) = 0$ sinon). La variable X dépend d'un paramètre $p = P(A)$ qui est aussi $E(X)$. Un estimateur sans biais convergeant vers p est

$$X_{[M]}^{\text{emp}} = \frac{X_1 + \cdots + X_M}{M}.$$

D'après le théorème limite central (Théorème 2.1), la suite de variables

$$\frac{X_{[M]}^{\text{emp}} - p}{\sqrt{\frac{p(1-p)}{M}}}$$

converge en loi, lorsque M tend vers l'infini, vers une loi $\mathcal{N}(0, 1)$. Pour M assez grand ($M \geq 30$) et sous les conventions $Mp \geq 5$ et $M(1-p) \geq 5$, on peut d'ailleurs assimiler la loi de la variable

$$\frac{X_{[M]}^{\text{emp}} - p}{\sqrt{\frac{p(1-p)}{M}}}$$

à la loi $\mathcal{N}(0, 1)$.

Si $\alpha \in]0, 1[$ figure un seuil de risque, soit $\tau_\alpha > 0$ défini par

$$P(|\mathcal{N}(0, 1)| > \tau_\alpha) = \alpha$$

(on se reporte aux tables de la loi $\mathcal{N}(0, 1)$). Si f_M est la fréquence de réalisations de A estimée à partir de M observations, c'est-à-dire

$$f_M := \frac{x_1 + \cdots + x_M}{M},$$

l'intervalle d'acceptation (asymptotique) de $T_M = X_{[M]}^{\text{emp}}$ au niveau de confiance $1 - \alpha$ (les risques étant comme toujours partagés) est

$$A_\alpha = \left\{ t; \frac{|t - p|}{\sqrt{\frac{f_M(1-f_M)}{M}}} < \tau_\alpha \right\}$$

et l'intervalle de confiance (asymptotique) correspondant pour p (au niveau de confiance $1 - \alpha$) est donc

$$\left[f_M - \tau_\alpha \sqrt{\frac{f_M(1-f_M)}{M}}, f_M + \tau_\alpha \sqrt{\frac{f_M(1-f_M)}{M}} \right],$$

f_M représentant la fréquences des réalisations observées sur M épreuves (les contraintes $M \geq 30$ et $M \min(p, 1-p) \geq 5$ étant supposées *a priori* remplies).

Le théorème limite central s'avère un outil majeur aux fins de la construction d'estimateurs. On doit cependant souligner que lui est dans ce cas fréquemment couplé (au titre d'outil complémentaire) le théorème suivant et son corollaire.

THEORÈME 3.1 (théorème de Slutsky). *Soit $(X_M)_{M \geq 1}$ une suite de VAR convergeant en loi vers une VAR X et $(Y_M)_{M \geq 1}$ une suite de VAR convergeant en probabilité vers une VAR constante y_0 . Les suites de VAR $(X_M + y_0)_{M \geq 1}$ et $(X_M y_0)_{M \geq 1}$ convergent alors en loi respectivement vers $X + y_0$ et $X y_0$.*

DÉMONSTRATION. On admet ici ce résultat (basé sur l'utilisation des définitions et sur la formule des probabilités totales). Voir [MathAp], chapitre 12, section II.1.4 pour une preuve détaillée. \square

COROLLAIRE 3.2. *Soit $(Y_M)_{M \geq 1}$ une suite de VAR telle qu'il existe $\mu \in \mathbb{R}$ et une suite $(\sigma_M)_{M \geq 1}$ de nombres strictement positifs tendant vers 0 de manière à ce que la suite de VAR $((Y_M - \mu)/\sigma_M)_{M \geq 1}$ converge en loi vers une VAR Z . Alors, pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ continue sur \mathbb{R} et dérivable en μ , on a :*

$$\lim_{\substack{\text{loi} \\ M \rightarrow +\infty}} \frac{f(Y_M) - f(\mu)}{\sigma_M} = f'(\mu) Z.$$

DÉMONSTRATION. Il résulte du théorème de Slutsky ci dessus que la suite $(Y_M)_{M \geq 1}$ converge en loi vers la VAR constante égale à μ . On utilise ensuite la définition du nombre dérivé comme limite du taux de variation (pour plus de détails, voir [MathAp], chapitre 12, section II.1.4). \square

FIN DU COURS

Annales 2011-2012, Texte et corrigé du DS

Durée 1h30. Polycopié de cours autorisé (à l'exclusion de tout autre document).

Exercice 1.

Une urne contient R boules rouges et B boules blanches indiscernables au toucher. L'épreuve aléatoire (dans les questions **1**, **2**, **3**) consiste à répéter indéfiniment l'opération consistant à tirer une boule de l'urne, puis à la remettre dans l'urne (après avoir noté sa couleur). Il s'agit donc de tirages successifs avec remise, donc indépendants.

1. Que vaut la probabilité de tirer exactement k boules rouges lors des N premiers tirages (k désignant un entier tel que $0 \leq k \leq N$) ? [on en donnera une expression littérale] ?

Il s'agit de tirages indépendants et avec remise hors d'une urne. Le modèle probabiliste est donc celui de la section 1.2.3 du cours. Le paramètre p vaut ici $p = R/(B + R)$ (proportion de boules rouges présentes dans l'urne). Le nombre de boules rouges tirées S_N suit une loi binomiale $\mathcal{B}(N, p)$. On a donc :

$$P(\{S_N = k\}) = \binom{N}{k} \left(\frac{R}{R+B}\right)^k \left(\frac{B}{R+B}\right)^{N-k} \quad \forall k \in \{0, \dots, N\}.$$

2. On suppose (seulement dans cette question) $R = B$. On note S_N la VAR représentant le nombre de boules rouges tirées lors des N premiers tirages. Quelle loi de probabilité suit la variable S_N ? Par quelle intégrale peut on approcher la probabilité de l'évènement

$$A_N = \left\{ \omega ; \frac{N - \sqrt{N}}{2} \leq S_N(\omega) < \frac{N + \sqrt{N}}{2} \right\}$$

lorsque N devient grand ? [on précisera à partir de quelle valeur de N cette approximation est licite].

Ici $p = 1/2$, et la variable S_N suit donc une loi binomiale $\mathcal{B}(N, 1/2)$. D'après le théorème de Berry-Esseen (formule (1.8) du cours), on peut, si F_{S_N} désigne la fonction de répartition de S_N , approcher les valeurs $F_{S_N}((N + \sqrt{N})/2)$ et $F_{S_N}((N - \sqrt{N})/2)$ respectivement par :

$$F_{S_N}((N + \sqrt{N})/2) = P(\{S_N \leq (N + \sqrt{N})/2\}) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^1 e^{-u^2/2}, du$$

$$F_{S_N}((N - \sqrt{N})/2) = P(\{S_N \leq (N - \sqrt{N})/2\}) \simeq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1} e^{-u^2/2} du$$

(on applique la formule (1.8) en prenant ici comme indiqué $p = 1/2$, ce qui implique que $\sqrt{Np(1-p)} = \sqrt{N}/2$). L'erreur absolue dans chacune de ces approximations

est contrôlée (cf. l'inégalité (1.6) du cours) en :

$$|\text{erreur}| \leq \frac{.7655}{\sqrt{N}}$$

et on la considère (suivant le cours) ces approximations licites dès que $N > 30$ (on a alors en effet aussi $N/2 = Np = N(1-p) > 5$). L'approximation de $P(A_N)$ est donc :

$$P(A_N) \simeq F_{S_N}((N + \sqrt{N})/2) - F_{S_N}((N - \sqrt{N})/2) \simeq \frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-u^2/2} du.$$

3. On reprend R et B quelconques. Quelle loi suit la VAR T représentant le numéro du tirage où l'on tire une boule rouge pour la première fois ? [on donnera la valeur de la probabilité de l'évènement $\{\omega; T(\omega) = k\}$ pour $k \in \mathbb{N}^*$] Calculer en fonction de R et B l'espérance de la VAR T . Que vaut la probabilité de l'évènement $\{\omega; T(\omega) \text{ est pair}\}$?

Si $R = 0$, la VAR T est identiquement égale à $+\infty$ (on ne tire jamais de boule rouge), et son espérance vaut $+\infty$; on suppose donc toujours ensuite $R > 0$. La VAR T suit une loi géométrique de paramètre $p = R/(R+B)$. On a en effet, si $p = R/(R+B) > 0$,

$$P(\{T = k\}) = p(1-p)^{k-1} \quad \forall k \in \mathbb{N}^*$$

(attention ici à la petite différence avec le cours, où la loi géométrique est considérée sur \mathbb{N} et non, comme ici, sur \mathbb{N}^* , il faut savoir s'adapter). L'espérance de cette VAR T se calcule (par exemple) comme la valeur de la dérivée en $s = 1$ de la fonction caractéristique (cf. (2.7) dans le cours, et surtout l'exemple 2.6); cette fonction caractéristique est dans ce cas la fonction :

$$s \mapsto E[s^T] = p \sum_{k=1}^{\infty} (1-p)^{k-1} s^k = ps \sum_{k=0}^{\infty} (1-p)^k = \frac{ps}{1 - (1-p)s}.$$

On a donc $E[T] = 1 + p(1-p)/p^2 = 1/p = (R+B)/R$. La probabilité que T prenne une valeur paire vaut :

$$\begin{aligned} P(\{T \text{ pair}\}) &= \sum_{k=1}^{\infty} P(\{T = 2k\}) = p \sum_{k=1}^{\infty} (1-p)^{2k-1} \\ &= p(1-p) \sum_{k=0}^{\infty} (1-p)^{2k} = \frac{p(1-p)}{1 - (1-p)^2} = \frac{1-p}{2-p} \\ &= \frac{B}{2B+R}. \end{aligned}$$

4. On considère à nouveau une urne renfermant R boules rouges et B boules blanches indiscernables au toucher. On dispose d'autre part (hors de l'urne) d'un stock infini de boules des deux couleurs. On effectue une suite de tirages successifs suivant la règle suivante : si l'on tire une boule d'une certaine couleur (rouge ou blanche) au k -ième tirage, on observe sa couleur, puis on la remet dans l'urne, mais en y ajoutant en même temps $m \in \mathbb{N}^*$ boules de la couleur de la boule tirée (prises dans notre stock), ce avant de passer au tirage suivant. On note, pour $k \in \mathbb{N}^*$, X_k la VAR valant 1 si la boule tirée au k -ième tirage est rouge, 0 si elle est blanche.

a) Calculer les probabilités des évènements $\{\omega; X_k(\omega) = 1\}$ pour $k = 1, 2$, puis celle de l'évènement $\{\omega; X_1(\omega) = X_2(\omega) = 1\}$. Les VAR X_1 et X_2 sont-elles indépendantes ?

On a $P(\{X_1 = 1\}) = R/(B + R)$. Pour calculer $P(\{X_2 = 1\})$, on utilise la formule des probabilités totales (Proposition 1.2, formule de Bayes (1.21) du cours). On a

$$\begin{aligned} P(\{X_2 = 1\}) &= P(\{X_2 = 1\} | \{X_1 = 1\}) P(\{X_1 = 1\}) \\ &\quad + P(\{X_2 = 1\} | \{X_1 = 0\}) P(\{X_1 = 0\}) \\ &= \frac{R+m}{R+B+m} \frac{R}{R+B} + \frac{R}{R+B+m} \frac{B}{R+B} \\ &= \frac{R(R+m+B)}{(R+B)(R+B+m)} = \frac{R}{R+B}. \end{aligned}$$

On a aussi :

$$\begin{aligned} P(\{X_1 = X_2 = 1\}) &= P(\{X_2 = 1\} | \{X_1 = 1\}) P(\{X_1 = 1\}) \\ &= \frac{(R+m)R}{(R+B)(R+B+m)}. \end{aligned}$$

Si $m \neq 0$, on a

$$\frac{(R+m)R}{(R+B)(R+B+m)} \neq \left(\frac{R}{R+B}\right)^2 = P(\{X_1 = 0\}) P(\{X_2 = 0\}),$$

ce qui prouve que les VAR X_1 et X_2 ne sont pas indépendantes (la condition (2.14) de la Définition 2.5 est en effet ici en défaut). Si $m = 0$ (tirages avec remise), on sait par contre que X_1 et X_2 sont indépendantes.

b) On renote S_N la VAR $S_N := X_1 + \dots + X_N$. Que représente S_N ? Calculer les probabilités conditionnelles $P(\{\omega; X_{N+1}(\omega) = 1\} | \{\omega; S_N(\omega) = k\})$ pour $k = 0, \dots, N$. En déduire la formule

$$P(\{\omega; X_{N+1}(\omega) = 1\}) = \frac{R + mE(S_N)}{R + B + Nm}.$$

La variable S_N représente le nombre de boules rouges tirées (puis remises, chaque fois avec m autres) au terme des N tirages. Pour tout $k = 0, \dots, N$, on a :

$$P\{X_{N+1} = 1 | \{S_N = k\}\} = \frac{R + km}{R + B + mN};$$

en effet, au terme de N tirages qui ont vu sortir exactement k boules rouges, la composition de l'urne est de $R + km$ boules rouges et de $B + m(N - k)$ boules blanches. La formule des probabilités totales (formule de Bayes (1.21), Proposition 1.2 du cours) donne :

$$\begin{aligned} P(\{X_{N+1} = 1\}) &= \sum_{k=0}^N P(\{X_{N+1} = 1\} | \{S_N = k\}) P(\{S_N = k\}) \\ &= R \sum_{k=0}^N \frac{P(\{S_N = k\})}{R + B + mN} + m \frac{\sum_{k=0}^N k P(\{S_N = k\})}{R + B + mN} \\ &= \frac{R}{R + B + mN} + \frac{mE[S_N]}{R + B + mN} \end{aligned}$$

puisque l'ensemble $\{0, \dots, N\}$ représente l'ensemble de toutes les valeurs possibles que peut prendre la VAR S_N .

c) Montrer que toutes les variables X_k , $k \in \mathbb{N}^*$, ont même loi, et que cette loi ne dépend pas de m . Sont-elles mutuellement indépendantes ? deux à deux indépendantes ? En déduire la valeur de l'espérance $E(S_N)$.

On montre que toutes les variables X_k , $k \in \mathbb{N}^*$, sont des VAR de Bernoulli de même paramètre, en montrant, par récurrence sur l'entier $N \in \mathbb{N}^*$, que, pour tout $N \in \mathbb{N}^*$, les VAR X_1, \dots, X_N sont toutes des VAR de Bernoulli de paramètre $R/(B+R)$. C'est le cas de X_1 (même de X_2 , cf. le résultat de la question **a**). Supposons le résultat vrai jusqu'au rang N . On a alors

$$E[S_N] = E[X_1] + E[X_2] + \dots + E[X_N] = N \frac{R}{B+R}$$

d'après la linéarité de l'opération de prise d'espérance (cf. la sous-section 2.1.2 du cours). On a donc, en utilisant le résultat établi au **b**),

$$\begin{aligned} P(\{X_{N+1} = 1\}) &= \frac{R}{R+B+mN} + \frac{mN}{R+B+mN} \frac{R}{R+B} \\ &= \frac{R}{R+B+mN} \left(1 + \frac{mN}{R+B}\right) = \frac{R}{R+B}, \end{aligned}$$

d'où l'on conclut que les VAR X_1, \dots, X_{N+1} , ont bien même loi. Le résultat voulu est donc bien démontré par récurrence. Comme X_1 et X_2 ne sont pas indépendantes (cf. la question **a**)), les VAR X_k , $k \in \mathbb{N}^*$ ne sont pas deux-à-deux indépendantes. Elles ne sont *a fortiori* pas mutuellement indépendantes, puisque l'on sait (sous-section 2.1.5 du cours) que l'indépendance mutuelle entre VAR implique leur indépendance deux-à-deux. L'espérance de S_N vaut, on l'a vu, $E[S_N] = NE[X_1] = NR/(B+R)$.

d) On reprend la même règle, mais cette fois avec $m = -1$ (ce qui revient à effectuer des tirages successifs sans remise), en conservant les mêmes notations. Quelle loi suit la VAR S_N lorsque $N \leq R+B$? [on précisera l'univers des possibles Ω et la valeur de $P(\{\omega; S_N(\omega) = k\})$ si $k \in \Omega$]. Justifier la formule $E(S_N) = NR/(R+B)$ en vous inspirant de la démarche adoptée aux questions **a**), **b**).

La VAR S_N suit dans ce cas une loi hypergéométrique. $\mathcal{H}(N, R, B)$ (d'après la sous-section 1.2.4 du cours). L'univers des possibles est $\Omega = \{\max(0, N-B), \dots, \min(N, R)\}$, et l'on a, pour tout k dans cet ensemble :

$$P(\{S_N = k\}) = \frac{\binom{R}{k} \times \binom{B}{N-k}}{\binom{R+B}{N}}.$$

Les calculs effectués aux questions **a**) et **b**) précédentes peuvent être repris avec cette fois $m = -1$. La seule différence est que l'on n'utilise comme valeurs de k dans la question **b**) que les valeurs comprises entre $\max(0, N-B)$ et $\min(N, R)$ (correspondant à l'univers des possibles de la VAR S_N). Ce sont uniquement ces valeurs de k que l'on fait entrer en jeu dans la formule des probabilités totales utilisée aux questions **b**) et **c**) pour calculer $P(\{X_{N+1} = 1\})$ lorsque les $P(\{X_k = 1\})$ sont supposés tous égaux à $R/(R+B)$ lorsque $k = 1, \dots, N$. On a donc encore $E[S_N] = NE[X_1] = NR/(R+B)$.

Exercice 2.

Mr X. rentre chez lui le soir avec un trousseau de $k \geq 2$ clefs dont seule une ouvre sa porte. Lorsqu'il est à jeun, il essaye une des clefs au hasard, puis, si elle ne fonctionne pas, la met de côté et essaye (toujours au hasard) une des clefs restantes ; il répète cette opération après chaque échec. Lorsqu'il est ivre, il essaye une clef au hasard, puis, si elle ne fonctionne pas, recommence avec cette fois le trousseau complet ; il répète cette opération après chaque échec. On note A l'évènement « Mr. X. est à jeun » et B l'évènement « Mr. X. est ivre ».

1. Calculer, sachant que Mr. X. est à jeun, la probabilité qu'il parvienne à ouvrir la porte au terme exactement de n tentatives ($n \in \mathbb{N}^*$). On désigne par X_A la variable aléatoire représentant le nombre d'essais nécessaires à Mr. X. pour parvenir à ouvrir sa porte quand il est à jeun. Quelle loi suit la VAR X_A ? Que vaut l'espérance de X_A ?

On a :

$$P(\{X_A = 1\}) = \frac{1}{k}$$

(car une seule clef du trousseau ouvre la porte et que le choix d'une clef est fait au hasard, ce qui signifie que les k choix possibles sont équiprobables). On a aussi :

$$P(\{X_A = 2\}) = \left(1 - \frac{1}{k}\right) \times \frac{1}{k-1} = \frac{1}{k}$$

puisque le trousseau ne comprend plus que $k-1$ clefs au second essai, le premier s'étant révélé infructueux : l'évènement $\{X_A = 2\}$ est intersection des deux évènements E_1 : « le premier essai se solde par un échec » et S_2 : « le second essai se solde par un succès ». On utilise ici :

$$P(\{X_A = 2\}) = P(E_1 \cap \{X_A = 2\}) = P(E_1 \cap S_2) = P(E_1) \times P(S_2 | E_1).$$

On constate en itérant ce calcul que, pour $n \leq k$:

$$P(\{X_A = n\}) = \frac{k-1}{k} \times \frac{k-2}{k-1} \times \cdots \times \frac{k-(n-1)}{k-(n-2)} \times \frac{1}{k-(n-1)} = \frac{1}{k}.$$

Si $n > k$, on trouve $P(\{X_A = n\}) = 0$ (au terme de $k-1$ essais infructueux, il ne reste plus qu'une clef dans le trousseau, qui est forcément celle qui ouvre). La VAR X_A suit donc la loi uniforme sur $\{1, \dots, k\}$. Son espérance vaut :

$$E[X_A] = \frac{1}{k}(1 + 2 + \cdots + k) = \frac{k+1}{2}.$$

2. Calculer, sachant que Mr. X. est ivre, la probabilité qu'il parvienne à ouvrir la porte au terme exactement de n tentatives ($n \in \mathbb{N}^*$). On désigne par X_B la variable aléatoire représentant le nombre d'essais nécessaires à Mr. X. pour parvenir à ouvrir sa porte quand il est ivre. Quelle loi suit la VAR X_B ? Que vaut l'espérance de X_B ?

On a, pour tout $k \in \mathbb{N}^*$,

$$P(\{X_B = n\}) = \frac{1}{k} \left(1 - \frac{1}{k}\right)^{n-1}.$$

La VAR X_B suit donc une loi géométrique de paramètre $p = 1/k$ (cf. l'exercice I, question 3), toujours considérée sur \mathbb{N}^* et non (comme c'est le cas dans le cours) sur \mathbb{N} . L'espérance de X_B vaut donc $1/p = k$.

3. On suppose $P(A) = P(B) = 1/2$. Calculer la probabilité π_n que Mr. X soit ivre, sachant qu'il est parvenu à ouvrir sa porte au terme d'exactly n essais ($n \in \mathbb{N}^*$) [on pensera à distinguer les cas où $n \leq k$ et $n > k$].

Soit U_n l'évènement : « il a fallu à Mr. X. exactement n essais pour ouvrir la porte ». Supposons tout d'abord $n \leq k$. On sait que $P(U_n | A) = 1/k$ et que $P(U_n | B) = (1/k)(1 - 1/k)^{n-1}$. On cherche à calculer $\pi_n = P(B | U_n)$. On utilise pour cela la formule des causes (Proposition 1.3) qui stipule que :

$$P(B | U_n) = \frac{P(U_n | B)P(B)}{P(U_n | A)P(A) + P(U_n | B)P(B)}.$$

Comme ici $P(A) = P(B) = 1/2$, on a donc :

$$\begin{aligned} \pi_n &= P(B | U_n) = \frac{P(U_n | B)}{P(U_n | A) + P(U_n | B)} = \frac{\frac{1}{k} \left(1 - \frac{1}{k}\right)^{n-1}}{\frac{1}{k} + \frac{1}{k} \left(1 - \frac{1}{k}\right)^{n-1}} \\ &= \frac{k^{n-1}}{k^{n-1} + (k-1)^{n-1}}. \end{aligned}$$

Si $n > k$, on a $P(U_n | A) = 0$ et, par conséquent $\pi_n = 1$.

Annales 2011-2012, Texte et corrigé de l'examen de session 1

Durée 3h00. Polycopié de cours autorisé (à l'exclusion de tout autre document).

Exercice 1. *On compte dans une population 30% d'hommes et 70% de femmes. Un homme sur deux et une femme sur trois portent des lunettes. Quelle est la probabilité qu'une personne portant des lunettes soit une femme ?*

Supposons la population Ω (univers des possibles, de cardinal N) équipée avec la loi uniforme $\mathcal{U}(N)$. La probabilité du sous-ensemble H constitué des hommes vaut 0.3, tandis que celle du sous-ensemble F constitué des femmes vaut 0.7. Soit L le sous-ensemble constitué des individus portant des lunettes. Par hypothèses, on a $P(L | H) = 0.5$ et $P(L | F) = 0.3$. On a, d'après la formule de causes (Proposition 1.3 du cours)

$$P(F | L) = \frac{P(F \cap L)}{P(L)} = \frac{P(L | F)P(F)}{P(L | H)P(H) + P(L | F)P(F)} = 7/12 \simeq .5833.$$

Exercice 2. *Soit $c > 0$ un paramètre réel et f_c la fonction :*

$$x \in [1, 4] \mapsto f_c(x) = cx^2(4 - x),$$

prolongée par 0 hors de $[1, 4]$. À quelle condition sur c la fonction f_c est-elle une densité de probabilité sur \mathbb{R} ? On suppose cette condition remplie. Soit X une variable aléatoire réelle (VAR) de loi de probabilité la loi ainsi définie. Calculer l'espérance de X . Calculer ensuite la probabilité conditionnelle $P(\{X > 2\} | \{X \leq 3\})$.

Pour que f_c soit une densité de probabilité sur \mathbb{R} , il faut et il suffit que f_c soit une fonction positive ou nulle (c'est le cas ici puisque $c > 0$ et $4 - x \geq 0$ sur $] - \infty, 4]$) et que

$$\int_{\mathbb{R}} f_c(t) dt = c \int_1^4 t^2(4 - t) dt = c \left[4t^3/3 - t^4/4 \right]_1^4 = 81c/4 = 1.$$

On a donc $c = 4/81$. L'espérance de la VAR X est (cf. le calcul de l'espérance d'une VAR à densité, section 2.1.3 du cours) :

$$E[X] = \int_{\mathbb{R}} t f_c(t) dt = \frac{4}{81} \int_1^4 t^3(4 - t) dt = \frac{112}{45}.$$

On a

$$\begin{aligned} P(\{X > 2\} | \{X \leq 3\}) &= \frac{P(\{2 < X \leq 3\})}{P(\{X \leq 3\})} = \frac{\int_2^3 t^2(4 - t) dt}{\int_1^3 t^2(4 - t) dt} \\ &= \frac{109/12}{158/5} = \frac{545}{1896} \simeq 0.287. \end{aligned}$$

Exercice 3.

1. Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles (VAR) mutuellement indépendantes, suivant toutes la même loi de Bernoulli de paramètre $p \in]0, 1[$. Soit $M \in \mathbb{N}^*$. Quelle loi suit la VAR

$$X_{[M]}^{\text{emp}} := \frac{X_1 + \cdots + X_M}{M} ?$$

Exprimer en fonction de p et de M l'espérance et la variance de la VAR $X_{[M]}^{\text{emp}}$.

Cette variable $X_{[M]}^{\text{emp}}$ est la somme de M variables aléatoires indépendantes suivant une loi $\mathcal{B}(1, p/M)$ (loi de Bernoulli de paramètre ou espérance p/M). La variable $X_{[M]}^{\text{emp}}$ suit donc une loi binomiale $\mathcal{B}(M, p/M)$. L'espérance de $X_{[M]}^{\text{emp}}$ vaut donc $M \times p/M = p$, tandis que la variance vaut $M \times p/M \times (1 - p/M) = p(1 - p/M)$ puisque la variance d'une VAR suivant une loi de Bernoulli de paramètre p/M vaut $p/M \times (1 - p/M)$.

2. Que signifie le fait qu'une suite de VAR $(Z_n)_{n \geq 1}$ converge en loi vers une VAR Z ? Quel théorème du cours permet d'assurer que la suite de VAR $(Z_n)_{n \geq 1}$, où

$$Z_n := \sqrt{n} \frac{X_{[n]}^{\text{emp}} - p}{\sqrt{p(1-p)}} \quad \forall n \in \mathbb{N}^*$$

converge en loi vers une VAR Z ? Quelle est dans ce cas la loi de la VAR limite Z ?

Dire qu'une suite de VAR $(Z_n)_{n \geq 1}$ converge en loi vers une VAR Z signifie que la suite des fonctions de répartition $(F_{Z_n})_{n \geq 1}$ converge simplement sur \mathbb{R} vers la fonction de répartition F_Z de la VAR limite Z . C'est le Théorème Limite Central (TLC) (Théorème 2.1 du cours) qui permet d'affirmer que la suite de variables $(Z_n)_{n \geq 1}$ converge vers une VAR suivant une loi $\mathcal{N}(0, 1)$.

3. Que signifie le fait qu'une suite $(U_n)_{n \geq 1}$ de VAR converge en probabilité vers une VAR U ? Quel théorème du cours permet d'assurer que la suite de VAR $((X_{[n]}^{\text{emp}}(1 - X_{[n]}^{\text{emp}}))_{n \geq 1}$ converge en probabilité vers une VAR U ? Quelle est cette VAR U dans ce cas?

Dire qu'une suite de VAR $(U_n)_{n \geq 1}$ converge en probabilité vers une VAR U signifie :

$$\forall \epsilon > 0, \lim_{n \rightarrow +\infty} P(\{|U_n - U| \geq \epsilon\}) = 0.$$

La Loi Forte des Grands Nombres (Théorème 2.3 du cours) assure que la suite de VAR $(X_{[n]}^{\text{emp}})_{n \geq 1}$ converge presque sûrement vers la VAR constante $p = E[X_1]$. La suite de VAR $((X_{[n]}^{\text{emp}}(1 - X_{[n]}^{\text{emp}}))_{n \geq 1}$ converge donc presque sûrement vers la VAR (presque partout) constante $p(1 - p)$. Comme la convergence presque sûre implique la convergence en probabilité, la suite de VAR $((X_{[n]}^{\text{emp}}(1 - X_{[n]}^{\text{emp}}))_{n \geq 1}$ converge aussi en probabilité vers la VAR (presque partout) constante égale à $p(1 - p)$.

4. On admettra (théorème de Slutsky) que si $(Z_n)_{n \geq 1}$ est une suite de VAR convergeant en loi vers une VAR Z et $(V_n)_{n \geq 1}$ une suite de VAR convergeant en probabilité vers une VAR constante égale à $v \in \mathbb{R}$, la suite de VAR $(V_n Z_n)_{n \geq 1}$ converge en loi vers la VAR vZ . Montrer que la suite de VAR

$$\left(\sqrt{n} \frac{X_{[n]}^{\text{emp}} - p}{\sqrt{(X_{[n]}^{\text{emp}}(1 - X_{[n]}^{\text{emp}}))}} \right)_{n \geq 1}$$

converge en loi vers une VAR dont on précisera la loi.

On prend comme suite $(Z_n)_{n \geq 1}$ la suite

$$\left(\sqrt{n} \frac{X_{[n]}^{\text{emp}} - p}{\sqrt{p(1-p)}} \right)$$

(qui converge en loi vers une VAR suivant une loi $\mathcal{N}(0, 1)$ d'après la question 2) et comme suite $(V_n)_{n \geq 1}$ la suite

$$\left(\frac{\sqrt{p(1-p)}}{\sqrt{(X_{[n]}^{\text{emp}}(1-X_{[n]}^{\text{emp}}))}} \right)_{n \geq 1}$$

qui converge en probabilité (d'après la question 3, la convergence ayant même lieu en fait presque sûrement) vers la VAR (presque partout) constante égale à 1. Le Théorème (admis) de Slutsky (Théorème 3.1 du cours) implique la convergence en loi de la suite de VAR

$$V_n Z_n = \sqrt{n} \frac{X_{[n]}^{\text{emp}} - p}{\sqrt{(X_{[n]}^{\text{emp}}(1-X_{[n]}^{\text{emp}}))}}, \quad n \geq 1,$$

vers une VAR suivant une loi $\mathcal{N}(0, 1)$.

5. On suppose M assez grand. Comment faut-il choisir τ pour que

$$P\left(\left\{ X_{[M]}^{\text{emp}} - \tau \sqrt{\frac{X_{[M]}^{\text{emp}}(1-X_{[M]}^{\text{emp}})}{M}} \leq p \leq X_{[M]}^{\text{emp}} + \tau \sqrt{\frac{X_{[M]}^{\text{emp}}(1-X_{[M]}^{\text{emp}})}{M}} \right\}\right) \geq .95 ?$$

Utiliser pour cela la table fournie en annexe dans ce texte.

Comme M est supposé choisi assez grand, on peut (d'après le résultat établi à la question 4) assimiler le membre de gauche de cette inégalité à

$$P(Z \in [-\tau, \tau]),$$

où Z suit une loi $\mathcal{N}(0, 1)$: on prend en effet pour Z la limite de la suite de VAR

$$\left(\sqrt{n} \frac{X_{[n]}^{\text{emp}} - p}{\sqrt{(X_{[n]}^{\text{emp}}(1-X_{[n]}^{\text{emp}}))}} \right)_{n \geq 1}.$$

La table fournie donne la condition $\tau \geq 1.96$ pour que soit réalisée la clause $P(Z \in [-\tau, \tau]) \geq 0.95$.

6. Dans un échantillon de 350 barquettes de fraises en provenance d'Espagne, on dénombre 38 barquettes contenant des fruits avariés. Déterminer un intervalle de confiance (aux risques partagés) au facteur de risque 5% près pour la proportion p de barquettes contenant des fruits avariés, parmi les barquettes de fraises importées d'Espagne.

La réalisation de l'espérance empirique $X_{[350]}^{\text{emp}}$ est ici 38/350. La réalisation de l'intervalle de confiance (aléatoire)

$$\left[X_{[M]}^{\text{emp}} - 1.96 \sqrt{\frac{X_{[M]}^{\text{emp}}(1-X_{[M]}^{\text{emp}})}{M}}, X_{[M]}^{\text{emp}} + 1.96 \sqrt{\frac{X_{[M]}^{\text{emp}}(1-X_{[M]}^{\text{emp}})}{M}} \right]$$

cherché est donc

$$\left] \frac{38}{350} - 1.96 \frac{\sqrt{\frac{38}{350} \left(1 - \frac{38}{350}\right)}}{\sqrt{350}}, \frac{38}{350} + 1.96 \frac{\sqrt{\frac{38}{350} \left(1 - \frac{38}{350}\right)}}{\sqrt{350}} \right[\simeq]0.0760, 0.1412[.$$

Exercice 4.

1. Soient Y_1, \dots, Y_M M variables aléatoires réelles (VAR) mutuellement indépendantes, toutes de même loi, et ξ un nombre réel. Justifier la formule :

$$E(e^{i\xi(Y_1 + \dots + Y_M)}) = (E(e^{i\xi Y_1}))^M.$$

D'après la formule (2.17) du cours, si Z_1, \dots, Z_M sont M VAR mutuellement indépendantes et ayant toutes des moments d'ordre M , la VAR $Z_1 \dots Z_M$ a encore une espérance, qui est le produit des espérances des Z_j . Ici, on prend $Z_j = \exp(i\xi Y_j)$; ces variables Z_j sont mutuellement indépendantes car les Y_j le sont. Comme $|Z_j| \equiv 1$, Z_k a des moments d'ordre M pour tout $k = 1, \dots, M$. Toutes les espérances $E(Z_k)$, $k = 1, \dots, M$, sont égales à $E(e^{i\xi Y_1})$ puisque les Y_j ont même loi. On a donc bien la formule voulue.

2. On suppose que la loi commune des Y_j est une loi $\mathcal{N}(\mu_M, \sigma_M^2)$. Calculer, pour $\xi \in \mathbb{R}$, en fonction de ξ, μ_M, σ_M^2 , l'espérance $E(e^{i\xi Y_1})$ en admettant que

$$\forall \xi \in \mathbb{R}, \quad \int_{\mathbb{R}} e^{i\xi t} e^{-t^2/2} dt = \sqrt{2\pi} e^{-\xi^2/2}.$$

En déduire que la loi de $Y_1 + \dots + Y_M$ est une loi normale dont on donnera les deux paramètres en fonction de μ_M et de σ_M^2 .

Comme Y_1 suit une loi $\mathcal{N}(0, 1)$, alors (d'après la remarque 2.3 du cours),

$$\begin{aligned} E[e^{i\xi Y_1}] &= \frac{1}{\sqrt{2\pi\sigma_M}} \int_{\mathbb{R}} e^{i\xi t} e^{-(t-\mu_M)^2/2\sigma_M^2} dt = \frac{1}{\sqrt{2\pi}} e^{i\xi\mu_M} \int_{\mathbb{R}} e^{i\sigma_M\xi u} e^{-u^2/2} du \\ &= e^{i\mu_M\xi} e^{-\sigma_M^2\xi^2/2} \end{aligned}$$

compte-tenu de la formule admise. On a donc

$$E[e^{i\xi(Y_1 + \dots + Y_M)}] = e^{i\xi M\mu_M} e^{-M\sigma_M^2\xi^2/2}.$$

On reconnaît la fonction caractéristique d'une loi $\mathcal{N}(M\mu_M, M\sigma_M^2)$. La VAR $Y_1 + \dots + Y_M$ suit donc une loi $\mathcal{N}(M\mu_M, M\sigma_M^2)$.

3. Soient $(X_n)_{n \geq 1}$ une suite de VAR mutuellement indépendantes et suivant toutes une loi $\mathcal{N}(\mu, \sigma^2)$. Déterminer, en utilisant le résultat établi à la question 2, pour $M \in \mathbb{N}^*$, la loi de la VAR :

$$X_{[M]}^{\text{emp}} := \frac{X_1 + \dots + X_M}{M}.$$

On pose $Y_j = X_j/M$. Chaque Y_j suit une loi $\mathcal{N}(\mu/N, \sigma^2/M^2)$. La variable $X_{[M]}^{\text{emp}}$ suit donc, d'après le résultat établi à la question 2, une loi $\mathcal{N}(\mu, \sigma^2/M)$.

4. Une société fabrique des câbles d'acier pour télécabines. La société demande un audit pour déterminer la résistance moyenne à la rupture de ces câbles. Les ingénieurs s'accordent à penser que la résistance d'un câble à la rupture suit une loi normale dont l'écart type est de 25 kilos. Le cabinet d'audit teste un lot de

100 cables, pour lesquels il détermine (après essais pour chaque cable du lot) une résistance moyenne à la rupture de 2750 kilos. Donner un intervalle de confiance (aux risques partagés) au niveau de confiance 98% pour la résistance moyenne à la rupture des cables fabriqués par la société.

En raisonnant comme à la question précédente, on voit que la variable

$$Z = \frac{\sqrt{M}}{\sigma} \left(X_{[M]}^{\text{emp}} - \mu \right) = \frac{1}{\sqrt{M}\sigma} \sum_{j=1}^M (X_j - \mu)$$

suit une loi normale $\mathcal{N}(0, 1)$. La table fournie donne la condition $\tau \geq 2.326$ pour que soit réalisé $P(Z \in [-\tau, \tau]) \geq 0.98$. La réalisation de l'intervalle de confiance (aléatoire) est alors

$$\begin{aligned} & \left[2750 - 2.326 \frac{\sigma}{\sqrt{100}}, 2750 + 2.326 \frac{\sigma}{\sqrt{100}} \right] = \\ & = \left[2750 - 2.326 \frac{25}{\sqrt{100}}, 2750 + 2.326 \frac{25}{\sqrt{100}} \right] \approx]2744, 2756[. \end{aligned}$$

Exercice 5¹. Un joueur joue avec un banquier suivant le principe suivant : il paye à la banque avant de jouer une somme finie S , puis lance ensuite une pièce de monnaie non pipée en répétant ces lancers indépendamment. On dit qu'un lancer est gagnant si le résultat est « Face ». Le banquier rembourse au joueur 2^k euros (s'il en a les possibilités) si celui-ci obtient « Face » pour la première fois au k -ième lancer, le jeu s'arrêtant aussitôt après ce premier lancer gagnant ; si le banquier n'est pas en mesure de lui rembourser cette somme après ce premier lancer gagnant, il lui donne tout ce qu'il a et le jeu s'arrête également aussitôt.

1. On suppose que le banquier dispose d'un capital illimité. Montrer que l'espérance de gain du joueur est $+\infty$. Le jeu est-il plus (ou moins) favorable au joueur qu'au banquier ?

La probabilité d'obtenir « Face » pour la première fois au k -ième lancer vaut $(1/2)^{k-1} \times 1/2 = 1/2^k$. L'espérance de gain du joueur est

$$E = \sum_{k=1}^{\infty} \frac{1}{2^k} 2^k - S = +\infty.$$

Le jeu est donc défavorable au banquier puisque le joueur ne mise au départ qu'une somme finie.

2. On suppose que le banquier ne dispose que de 2^N euros ($N \in \mathbb{N}^*$). Calculer l'espérance de gain du joueur. En dessous de quelle valeur de S le jeu est-il plus favorable au joueur qu'au banquier ?

Dans ce cas, l'espérance du gain vaut

$$E = \sum_{k=1}^{N-1} \frac{1}{2^k} \times 2^k + \sum_{k=N}^{\infty} \frac{1}{2^k} 2^N - S = N + 1 - S.$$

Si $S = N + 1$, le jeu est équitable. Si $S < N + 1$, le jeu est plus favorable au joueur, tandis que si $S > N + 1$, il est plus favorable au banquier.

1. Il s'agit ici du célèbre « jeu de Saint-Petersbourg »

Exercice 6.

On considère un ensemble fini d'états $E = \{e_1, \dots, e_N\}$. Un robot se trouve positionné à l'instant $t = 0$ (le temps est ici discrétisé : $t = 0, 1, 2, \dots$) en l'état e_1 . Il se déplace ensuite d'un état à l'autre de manière stochastique suivant la règle suivante : s'il est positionné à l'instant $t = k$ sur un état e_i , la probabilité qu'il aille à l'instant $t = k + 1$ se positionner sur l'état e_j ne dépend pas de k et vaut un certain nombre $a_{i,j} \in [0, 1]$.

1. On note X_k la VAR désignant le numéro de l'état sur lequel se trouve le robot à l'instant k . Vérifier que l'on a :

$$\forall k \in \mathbb{N}, \quad P(X_{k+1} = j) = \sum_{i=1}^N a_{i,j} P(X_k = i).$$

On applique la formule de Bayes (Proposition 1.2 du cours). Le nombre $a_{i,j}$ s'interprète en effet comme la probabilité conditionnelle :

$$P(\text{le robot va en } e_j \text{ à l'instant } k + 1 \mid \text{il est positionné en } e_i \text{ à l'instant } k),$$

soit

$$P(X_{k+1} = j \mid X_k = i).$$

2. On note A la matrice $[a_{i,j}]_{1 \leq i, j \leq N}$ (i : indice de ligne, j : indice de colonne). Que peut-on dire de la somme des éléments de chaque ligne de la matrice A ? Calculer en fonction de A la loi de probabilité de la VAR X_k .

La somme des éléments de chaque ligne de la matrice A vaut 1 puisque le chaque vecteur ligne $[a_{i,1}, \dots, a_{i,N}]$ représente une distribution de probabilité (conditionnelle) sur $\{1, \dots, N\}$. La loi de probabilité de la variable X_k se calcule comme le vecteur ligne

$$X_0 * A^k = [1, 0, \dots, 0] * A^k.$$

Exercice 7. L'expérience a montré qu'un tireur à l'arc atteignait, lors d'un tir donné, sa cible avec une probabilité $p \in]0, 1[$. Il effectue une suite de tirs consécutifs t_j , $j = 1, 2, \dots$, dans des conditions d'indépendance mutuelle. Soit M un entier supérieur ou égal à 1. Soit T_M la VAR correspondant au nombre de tirs qui lui sont nécessaires pour atteindre exactement M fois la cible.

1. Calculer, pour $k \in \mathbb{N}^*$, la probabilité de l'évènement $\{T_M = k\}$.

On a, pour $k \geq M$,

$$P(\{T_M = k\}) = p^M (1-p)^{k-M} \binom{k-1}{M-1}$$

puisque cette probabilité correspond en fait à celle de l'évènement suivant : $(k-1) - (M-1) = k-M$ échecs exactement lors des $k-1$ premiers tirs, un succès au k -ième (et que les tirs sont indépendants).

2. Calculer la fonction génératrice de la loi de T_M en utilisant la formule :

$$\forall y \in]-1, 1[, \quad \frac{1}{(1-y)^M} = \sum_{k=M}^{\infty} \binom{k-1}{M-1} y^{k-M}$$

(que l'on justifiera). En déduire la valeur de $E(T_M)$.

La fonction génératrice de T_M est définie formellement par

$$\begin{aligned} s \mapsto E[s^{T_M}] &= \sum_{k=M}^{\infty} P(\{T_M = k\}) s^k \\ &= p^M s^M \sum_{k=M}^{\infty} \binom{k-1}{M-1} (1-p)^{k-M} s^{k-M} \\ &= (ps)^M \frac{1}{(1-(1-p)s)^M} = \left(\frac{ps}{1-(1-p)s} \right)^M \quad (\text{si } |s| \leq 1). \end{aligned}$$

On utilise ici la formule proposée; cette formule s'obtient en dérivant M fois les deux membres de

$$\frac{1}{1-y} = \sum_{k=0}^{\infty} y^k.$$

La dérivée de la fonction génératrice au voisinage de $[-1, 1]$ vaut

$$s \mapsto M \left(\frac{ps}{1-(1-p)s} \right)^{M-1} \times \frac{p}{(1-(1-p)s)^2}.$$

La valeur de cette dérivée en $s = 1$ vaut $E[T_M] = M/p$.

Annexe

Quantiles de dépassement de l'écart absolu de la loi normale $\mathcal{N}(0, 1)$

$$\mathbb{P}(|X| > z_\alpha) = \alpha$$

(par exemple : si $\alpha = 0.13 + 0.005$, alors $z_\alpha = 1.495$)

α	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.00	∞	3.291	3.090	2.968	2.878	2.807	2.748	2.697	2.652	2.612
0.01	2.576	2.543	2.512	2.484	2.457	2.432	2.409	2.387	2.366	2.346
0.02	2.326	2.308	2.290	2.273	2.257	2.241	2.226	2.212	2.197	2.183
0.03	2.170	2.157	2.144	2.132	2.120	2.108	2.097	2.086	2.075	2.064
0.04	2.054	2.044	2.034	2.024	2.014	2.005	1.995	1.986	1.977	1.969
0.05	1.960	1.951	1.943	1.935	1.927	1.919	1.911	1.903	1.896	1.888
0.06	1.881	1.873	1.866	1.859	1.852	1.845	1.838	1.832	1.825	1.818
0.07	1.812	1.805	1.799	1.793	1.787	1.780	1.774	1.768	1.762	1.757
0.08	1.751	1.745	1.739	1.734	1.728	1.722	1.717	1.711	1.706	1.701
0.09	1.695	1.690	1.685	1.680	1.675	1.670	1.665	1.660	1.655	1.650
0.10	1.645	1.640	1.635	1.630	1.626	1.621	1.616	1.612	1.607	1.603
0.11	1.598	1.594	1.589	1.585	1.580	1.576	1.572	1.567	1.563	1.559
0.12	1.555	1.551	1.546	1.542	1.538	1.534	1.530	1.526	1.522	1.518
0.13	1.514	1.510	1.506	1.502	1.499	1.495	1.491	1.487	1.483	1.480
0.14	1.476	1.472	1.468	1.465	1.461	1.457	1.454	1.450	1.447	1.443
0.15	1.440	1.436	1.433	1.429	1.426	1.422	1.419	1.415	1.412	1.408
0.16	1.405	1.402	1.398	1.395	1.392	1.388	1.385	1.382	1.379	1.375
0.17	1.372	1.369	1.366	1.363	1.359	1.356	1.353	1.350	1.347	1.344
0.18	1.341	1.338	1.335	1.332	1.329	1.326	1.323	1.320	1.317	1.314
0.19	1.311	1.308	1.305	1.302	1.299	1.296	1.293	1.290	1.287	1.284
0.20	1.282	1.279	1.276	1.273	1.270	1.267	1.265	1.262	1.259	1.256

Annales 2011-2012, Texte et corrigé de l'examen de session 2

Durée 3h00. Polycopié de cours autorisé (à l'exclusion de tout autre document). Consulter la table fournie en annexe dans le texte de l'examen de session 1 (Annexe B).

Exercice 1. *Sur une autoroute, les statistiques officielles révèlent qu'un conducteur sur 1000 ne respecte pas la limitation de vitesse à 130 km/h. Un radar, placé sur cet autoroute, repère l'excès de vitesse chez 99% des conducteurs fautifs (en les flashant). Cependant, du fait de défaillances ponctuelles du système, 0.3% de conducteurs en règle se trouvent aussi flashés. Calculer la probabilité pour qu'un conducteur se retrouvant flashé le soit effectivement à juste titre.*

On prend comme univers des possibles l'ensemble de tous les conducteurs, avec comme distribution de probabilité la distribution uniforme. La probabilité de l'évènement NR (un conducteur pris au hasard ne respecte pas la limitation de vitesse) est $P(NR) = 10^{-3}$. Si F désigne l'évènement « un conducteur pris au hasard est flashé », on a $P(F | NR) = 0.99$. D'autre part, si R désigne l'évènement « un conducteur pris au hasard respecte la limitation de vitesse », on a $P(F | R) = 3/100 = 3 \times 10^{-3}$. D'après la formule des causes (Proposition 1.3 du cours), on a

$$P(NR | F) = \frac{P(F | NR) P(NR)}{P(F | NR) P(NR) + P(F | R) P(R)} = \frac{0.99}{0.99 + 3 \times 0.999} \simeq 1/4.$$

Surprenant, non ?

Exercice 2.

1. Soient X_1 et X_2 deux VAR indépendantes suivant l'une une loi de Poisson de paramètre $\lambda_1 > 0$, l'autre une loi de Poisson de paramètre $\lambda_2 > 0$. Montrer que $X_1 + X_2$ suit encore une loi de Poisson dont on précisera le paramètre.

La fonction génératrice de X_1 est $s \mapsto e^{\lambda_1 s}$, celle de X_2 est $s \mapsto e^{\lambda_2 s}$. Comme X_1 et X_2 sont supposées indépendantes, la fonction génératrice de la somme est le produit des fonctions génératrices, soit $s \mapsto e^{(\lambda_1 + \lambda_2)s}$. On reconnaît là la fonction génératrice d'une loi de Poisson de paramètre $\lambda_1 + \lambda_2$. La loi de $X_1 + X_2$ est donc une loi de Poisson $\mathcal{P}(\lambda_1 + \lambda_2)$.

2. Soit $(X_k)_{k \in \mathbb{N}^*}$ une suite de VAR mutuellement indépendantes toutes de même loi, à savoir la loi de Poisson de paramètre $\lambda > 0$. Quelle est la loi de la VAR

$$X_{[n]}^{\text{emp}} : \frac{X_1 + \dots + X_n}{n} \quad ?$$

Que valent espérance et écart-type de cette VAR $X_{[n]}^{\text{emp}}$?

Comme les variables $X_j, j = 1, \dots, n$, sont mutuellement indépendantes, la fonction génératrice de la VAR $X_{[n]}^{\text{emp}}$ est le produit des fonctions génératrices des VAR X_j/n . Ces VAR sont toutes de même loi, la loi de Poisson $\mathcal{P}(\lambda/n)$, et chacune de ces fonctions génératrices est $s \mapsto e^{\lambda s/n}$. La fonction génératrice de $X_{[n]}^{\text{emp}}$ est donc $s \mapsto (e^{\lambda s/n})^n = e^{-\lambda s}$. La VAR $X_{[n]}^{\text{emp}}$ suit donc une loi de Poisson $\mathcal{P}(\lambda)$. Son espérance vaut λ , ainsi que sa variance (d'après le cours).

3. Vers quelle VAR la suite de VAR

$$\left(\sqrt{\frac{n}{\lambda}} (X_{[n]}^{\text{emp}} - \lambda) \right)_{n \geq 1}$$

converge-t-elle en loi lorsque n tend vers $+\infty$?

D'après le Théorème Limite Central (Théorème 1.1 du cours), cette suite converge en loi vers une VAR suivant une loi $\mathcal{N}(0, 1)$.

4. On suppose N grand. Déterminer un seuil x tel que

$$P\left(\left\{ \lambda \in \left] X_{[N]}^{\text{emp}} - x\sqrt{\frac{\lambda}{N}}, X_{[N]}^{\text{emp}} + x\sqrt{\frac{\lambda}{N}} \right[\right\}\right) \geq 95\%$$

(on exploitera la table donnée en annexe).

Ce seuil vaut $x = 1.96$ d'après la table donnée en annexe.

5. Dans une entreprise, on constate au final, sur une durée de deux ans, qu'il y a eu 20 accidents du travail à déplorer au sein du personnel. Pourquoi la loi de Poisson est-elle un modèle raisonnable pour modéliser la loi de la VAR X égale au nombre d'accidents susceptibles de survenir par jour ? Si l'on suppose que l'on choisit ce modèle, et donc que X suit une telle loi, déterminer (en utilisant le résultat établi à la question (4)) un intervalle de confiance pour la moyenne λ du nombre d'accidents journaliers, au niveau de confiance 95%.

Comme la loi de Poisson peut être considérée heuristiquement comme la « loi des événements rares », c'est un modèle naturel pour modéliser la loi de la VAR X . Sur une durée de $365 \times 2 = 730$ jours, la réalisation de $X_{[730]}^{\text{emp}}$ est 2/73. Les valeurs extrêmes de λ telles que

$$\lambda \in \left] 2/73 - 1.96 \sqrt{\frac{\lambda}{730}}, 2/73 + 1.96 \sqrt{\frac{\lambda}{730}} \right[$$

sont données en cherchant les racines positives des deux équations du second degré en $\sqrt{\lambda}$:

$$(2/73 - \lambda) = +1.96 \sqrt{\lambda}/\sqrt{730}, \quad (2/73 - \lambda) = -1.96 \sqrt{\lambda}/\sqrt{730}.$$

L'unique racine (positive) de la première équation est $\sqrt{\lambda} \simeq 0.1332$, tandis que celle de la seconde équation est $\sqrt{\lambda} \simeq 0.2057$. La réalisation de l'intervalle de confiance proposée est donc

$$](0.133)^2, (0.206)^2 [=]0.017, 0.043[.$$

Exercice 3.

Soit $p \in \mathbb{N}^*$, $c > 0$, et $f_{p,c}$ la fonction

$$f_{p,c} : t \in \mathbb{R} \mapsto c \chi_{[0, \infty[}(t) \exp(-t/2) t^{p-1}, \quad c > 0.$$

1. Comment faut-il choisir la constante $c = c_p$ pour que f_{p,c_p} soit la densité d'une loi de VAR ?

On a

$$\int_0^\infty t^{p-1} e^{-t/2} dt = 2^p \int_0^\infty u^{p-1} e^{-u} du = 2^p (p-1)!.$$

On doit donc, pour que l'intégrale sur \mathbb{R} de f_{p,c_p} vaille 1, choisir

$$c_p = 2^{-p}/(p-1)!.$$

2. Une fois que c_p a été choisie comme au (1), que valent espérance et variance de la VAR X_p de densité f_{p,c_p} ?

On a

$$\int_0^\infty t^p e^{-t/2} dt = 2^{p+1} \int_0^\infty u^p e^{-u} du = 2^{p+1} p!.$$

L'espérance de X_p vaut donc $c_p \times 2^{p+1} \times p! = 2p$. D'autre part

$$\int_0^\infty t^{p+1} e^{-t/2} dt = 2^{p+2} \int_0^\infty u^{p+1} e^{-u} du = 2^{p+2} (p+1)!.$$

Donc $E[X_p^2] = c_p \times 2^{p+2} (p+1)^2 = 4p(p+1)$. On en déduit donc

$$\text{Var}(X_p) = 4p(p+1) - 4p^2 = 4p.$$

3. Soit $\epsilon > 0$. Donner en fonction de p une majoration de

$$P(\{|X_p - 2p| \geq \epsilon\}).$$

Pour quelles valeurs de ϵ une telle inégalité est-elle intéressante ?

Cette majoration est donnée par l'inégalité de Bienaymé-Tchebychev (Proposition 2.1 du cours), qui donne :

$$P(\{|X_p - 2p| \geq \epsilon\}) \leq \frac{\text{Var}(X_p)}{\epsilon^2} = \frac{4p}{\epsilon^2}.$$

Cette inégalité n'est intéressante que si $\epsilon > \sigma(X_p) = 2\sqrt{p}$.

4. Soient p_1 et p_2 deux entiers strictement positifs. Vérifier la formule :

$$\forall t \in \mathbb{R}, \int_{\mathbb{R}} f_{p_1, c_{p_1}}(s) f_{p_2, c_{p_2}}(t-s) ds = f_{p_1+p_2, c_{p_1+p_2}}(t).$$

Déduire de ce calcul la densité de la loi de $X_{p_1} + X_{p_2}$ lorsque X_{p_1} et X_{p_2} sont supposées être deux VAR indépendantes, de densités respectives $f_{p_1, c_{p_1}}$ et $f_{p_2, c_{p_2}}$.

Le membre de droite de la formule vaut 0 si $t < 0$ et, si $t \geq 0$:

$$c_{p_1} c_{p_2} e^{-t} \int_0^t s^{p_1-1} (t-s)^{p_2-1} ds = c_{p_1} c_{p_2} t^{p_1+p_2-1} e^{-t} \int_0^1 u^{p_1-1} (1-u)^{p_2-1} du. (*)$$

En intégrant par parties, on trouve

$$\int_0^1 u^{p_1-1} (1-u)^{p_2-1} du = \frac{(p_1-1)! (p_2-1)!}{(p_1+p_2-1)!}.$$

En reportant dans (*), on retrouve bien la valeur de $f_{p_1+p_2, c_{p_1+p_2}}(t)$. La densité de la loi de $X_{p_1} + X_{p_2}$ est, puisque les deux variables sont indépendantes, la fonction

$$t \in \mathbb{R} \mapsto \int_{\mathbb{R}} f_{p_1, c_{p_1}}(s) f_{p_2, c_{p_2}}(t-s) ds.$$

En effet, au niveau infinitésimal

$$P(\{X_{p_1} \in [s, s+ds]\} \cap \{X_{p_2} \in [t-s, t-s+dt]\}) = f_{p_1}(s) \times f_{p_2}(t-s) ds dt$$

du fait de l'indépendance (les probabilités des deux événements se multiplient). La densité de la loi de $X_{p_1} + X_{p_2}$ est donc exactement celle de $X_{p_1+p_2}$.

Exercice 4.

Un surfeur pianote sur le net (il y a un nombre fini N de pages répertoriées, numérotées de 1 à N). S'il consulte à l'instant $t = k$ la page i , la probabilité qu'il aille à l'instant $t = k + 1$ consulter la page j ne dépend pas de k et vaut un certain nombre $a_{i,j} \in [0, 1]$.

1. On note X_k la VAR désignant le numéro de la page que consulte le surfeur à l'instant k . Vérifier que l'on a :

$$\forall k \in \mathbb{N}, \quad P(X_{k+1} = j) = \sum_{i=1}^N a_{i,j} P(X_k = i).$$

On peut appliquer la formule de Bayes (Proposition 1.2 du cours), car $a_{i,j}$ s'interprète comme la probabilité (conditionnelle) que le surfeur aille consulter la page j à l'instant $k + 1$, conditionnée par l'évènement : « le surfeur consulte la page i à l'instant k ».

2. On note A la matrice $[a_{i,j}]_{1 \leq i,j \leq N}$ (i : indice de ligne, j : indice de colonne). Que peut-on dire de la somme des éléments de chaque ligne de la matrice A ? Calculer en fonction de A la loi de probabilité de la VAR X_k (on suppose $P(\{X_1 = 1\}) = 1$).

La somme des éléments de la matrice A vaut 1 puisque l'on peut interpréter le vecteur ligne des entrées de A figurant sur la ligne i comme une distribution de probabilité (conditionnelle) sur $\{1, \dots, n\}$ (à condition toutefois que la probabilité que cette page i soit visitée par X_k ne soit pas nulle pour toutes les valeurs de k). Comme $P(\{X_1 = 1\}) = 1$, on a, en tenant compte des formules inductives établies à la question 1 :

$$L_k = [1, 0, \dots, 0] * A^{k-1}$$

si L_k désigne le vecteur ligne correspondant à la distribution de probabilité de la VAR X_k sur $\{1, \dots, N\}$.

Bibliographie

- [MathAp] A. Yger & J.A. Weil (ed.), *Mathématiques Appliquées L3*, Pearson Education, Paris, 2009.
- [Eiser] M. Eisermann, Comment fonctionne Google ?
<http://www-fourier.ujf-grenoble.fr/~eiserm/Enseignement/google.pdf>
- [Ypin] A. Yger, UE PIN401 - Initiation à l'algèbre linéaire, Probabilités-Statistique (2006). Polycopié en ligne :
http://www.math.u-bordeaux1.fr/~yger/pin401-06_1.pdf

Index

- acceptation
 - intervalle d', au niveau de confiance
 - $1 - \alpha$, 53
- adjacence
 - matrice pondérée, d'un graphe orienté, 24
 - matrice, d'un graphe orienté, 24
- aléatoire
 - évènement, 3, 13
 - évènement élémentaire, 13
 - variable réelle, 29
- arête orientée
 - d'un graphe orienté, 24
- asymptotique
 - intervalle de confiance, 46
- Banach-Tarski
 - paradoxe de, 13
- Bayes
 - formule de, 22
- Bernoulli, loi de, 6
- Berry-Esseen, théorème de, 7, 44
- Bertrand
 - Joseph, 2
 - paradoxe de, 2
- biais
 - estimateur sans, d'un paramètre, 51
 - estimateur sans, de l'espérance d'une VAR, 31
- Bienaymé-Tchebychev
 - inégalité de, 38
- binomiale $\mathcal{B}(N, p)$
 - loi, 5, 33
- borélienne, tribu, 12
- Borel, tribu de, 12
- brute
 - série de données statistique, 49
- caractéristique
 - fonction, d'une VAR, 36
- causes
 - formule des, 23
- chi-deux
 - loi du, 54
- conditionnelle
 - probabilité, 22
- conditionnement, 22
- confiance
 - intervalle de, à un niveau de confiance donné, 53
 - intervalle de, aux risques partagés, 53
 - intervalle de, d'une moyenne, 45
 - intervalle, à un niveau de confiance donné, pour une moyenne (à variance connue), 46
 - niveau de, 45
- continue
 - loi de probabilité, 13
 - VAR, 29
- convergence
 - en loi, d'une suite de VAR, 44
 - en probabilité, d'une suite de VAR, 47
 - presque sûre, 47
- correct
 - absolument, estimateur, 51
- covariance
 - d'un vecteur de VA indépendantes
 - deux-à-deux, 41
 - matrice de, d'un vecteur de VA complexes, 37
 - matrice de, d'un vecteur de VAR, 36
- covdiag, 41
- densité
 - de probabilité sur \mathbb{R}^n , 14
 - loi à, 14
- descriptive
 - statistique, 49
- discrète
 - loi de probabilité, 10
 - VAR, 29
- écart type
 - d'une série de données statistique, 50
 - d'une VAR, d'une VA complexe, 37
- échantillon, 51
- efficace
 - estimateur, 51
- empirique

- espérance, d'une VAR, 32, 52
- espérance, d'une VAR positive, 30
- espérance, dans le cadre gaussien, 54
- variance, d'une VAR, 42, 52
- variance, dans le cadre Gaussien, 55
- épreuve, 1
- espérance
 - d'un produit de VA indépendantes, 39
 - d'une VAR ou d'un vecteur de VAR, 31
 - d'une VAR positive, 30
 - empirique, d'une VAR, 32
 - empirique, d'une VAR ou d'un vecteur de VAR, 32
 - empirique, d'une VAR positive, 30
- estimateur
 - absolument correct, 51
 - convergeant, 51
 - d'un paramètre, 51
 - efficace, 51
 - inférieur, supérieur, d'une moyenne, à variance connue, 46
 - sans biais, 51
- estimateur sans biais
 - d'une espérance, 31
 - d'une variance, 43
- estimation
 - à l'ordre M , d'un paramètre, 51
- évènement aléatoire, 3
- évènements rares
 - loi des, 10
- expérience, 1
- exponentielle
 - loi $\mathcal{E}(\lambda)$, 20, 26, 35
- fiabilité, 21
- flèche
 - d'un graphe orienté, 24
- Fourier
 - transformée de, d'une densité de probabilité, 36
- génératrice
 - fonction, d'une VAR discrète, 33, 38
- géométrique, loi, 11, 34
- Galton, planche de, 19
- Gauß
 - fonction de, 17
- gaussien, vecteur, 19
- Google, 24, 26
- grands nombres
 - loi faible des (LGNf), 46
 - loi forte des (LGNF), 47
- graphe orienté, 24
- hypergéométrique, loi, 8, 33
- incompatibles, évènements aléatoires, 27
- indépendance
 - deux-à-deux de VAR ou de VA complexes, 39
 - mutuelle entre VAR, 39
- indépendantes
 - épreuves, heuristiquement, 28
- indépendants
 - évènements aléatoires, 27
 - évènements mutuellement, 27
- induite
 - probabilité, 21
- inférentielle
 - statistique, 49
- intervalle
 - de confiance, 46
- Jensen
 - inégalité de, 32
- Kolmogorov
 - Andrei, 3
 - inégalité de, 47
- Lebesgue
 - tribu de, 13
- loi
 - d'une VAR, 29
- médiane
 - d'une série brute de données statistique organisée de manière croissante, 51
- marche aléatoire
 - sur un graphe orienté, 25
- Markov
 - chaîne homogène de, 26
- moments d'ordre m
 - centrés, non centrés, 36
- moyenne
 - d'une série de données statistique, 49
 - d'une VAR, 31
- normale
 - $\mathcal{N}(\mu, \sigma^2)$, loi, 17, 35
 - loi centrée réduite, 17
 - loi en dimension $n > 1$, 19, 35
- Pagerank, algorithme, 24, 26
- Pascal, loi de, 11
- Poincaré, formule de, 4
- Poisson $\mathcal{P}(\lambda)$, loi de, 10, 34, 38
- polyhypergéométrique, loi, 8
- preuves
 - formule des, 23
- probabilisé, univers ou espace, 13
- probabilisable, univers ou espace, 12
- produit
 - probabilité (et indépendance), 28
 - tribu, 13, 28
- quantile
 - fonction de, 14

- réalisation d'une épreuve, 3
- régression
 - coefficient de, 52
- répartition
 - fonction de, 14
- renouvellement, 10
- risque
 - facteur de, 45
- risque quadratique, 51
- série de données statistique, 49
- Slutsky
 - théorème de, 57
- sommet
 - d'un graphe orienté, 24
- spectre
 - d'une densité de probabilité, 36
- statistique
 - série de données, 49
- stochastique, matrice, 24
- Student
 - loi de, 55
 - test de, 55
- système complet d'évènements aléatoires, 23
- tableau
 - d'effectifs par classe, 49
 - d'effectifs par valeur, 49
- théorème limite central, 44
- TLC, 44
- toile
 - marche aléatoire sur la, 24
- trace
 - d'une tribu, 21
 - probabilité, 21
- triangulaire, inégalité, 32
- tribu, 12
 - borélienne sur \mathbb{R}^n , 12
 - produit sur $\{0, 1\}^{\mathbb{N}}$ ou $E^{\mathbb{N}}$, E fini, 13
- uniforme
 - continue, loi sur un ouvert borné de \mathbb{R}^n , 16
- uniforme discrète
 - loi $\mathcal{Z}(N)$, 4, 32
- univers des possibles, 3
- VAR, 29
- variable aléatoire réelle, 29
- variance
 - d'une somme de VA indépendantes deux-à-deux, 42
 - d'une VAR, d'une VA complexe, 37
- variance empirique
 - d'une VAR, 42
- vraisemblance
 - fourchette de, 53