

Simulations des phénomènes aléatoires
LM 346 : PARTIE 1.

Raphaël KRIKORIAN
Université Paris 6

Remis à jour pour l'année 2014-2015 par Irina KOURKOVA

Remerciements : L’auteur tient à remercier vivement Valentin KONAKOV pour sa lecture attentive et critique d’une version précédente de ce polycopié. De nombreuses erreurs et inexactitudes ont pu ainsi être corrigées. Celles qui subsistent sont bien évidemment sous la responsabilité de l’auteur.

Ceci est la première partie du polycopié du cours LM346 "Processus et Simulations" remise à jour pour l’année 2014-15 par I.Kourkova. Elle s’accompagne des feuilles d’exercices N1 et N2. La deuxième partie du cours (Chaines de Markov sur un espace fini et leurs simulations) constitue un document séparé et s’accompagnera de la feuille d’exercices N3.

Calcul de la note finale. La note finale sera calculée de la manière suivante :

$$\text{Total}/100 = \text{Examen}/60 + \text{TP}/20 + \max \left(\text{CC}/20, (1/3)\text{Examen}/60 \right).$$

La note de CC/20 est la **Moyenne** (et non le max!) de deux contrôles sur 20 points dans l’amphi. En effet, un contrôle sera organisé au milieu du semestre (il portera sur la première partie du cours : simulations et initiation aux statistiques) et un contrôle sera organisé a la fin (sur la deuxième partie – chaines de Markov.) Vous serez prevenus par le secrétariat sur les dates.

La note de TP vous sera donnée en contrôle de TP. En cas d’absence à un de ces contrôles (en TP ou en amphi), la note attribuée pour ce contrôle sera 0.

Pour les étudiants en télé-enseignement : Vous serez invités à passer les contrôles en amphi qui seront organisés deux fois courant ce semestre pour obtenir une note de CC et vous entraîner sur le cours. Les présence à ces contrôles n’est en aucun cas obligatoire! En cas d’absence, c’est un tiers de la note de l’examen qui va emporter dans le maximum, dans le calcul de la note finale.

Comme les étudiants du télé-enseignement ne peuvent pas se rendre aux TP, leur note de TP sera calculée comme un tiers de la note de l’examen. Ceci se fera **uniquement** pour les étudiants inscrits en télé-enseignement! Vous allez recevoir :

- 1er envoi : Première partie du polycopié du cours, feuille N1.
- 2ème envoi : Correction de la feuille N1. Feuille N2.
- 3ème envoi : Correction de la feuille N2. Deuxième partie du polycopié. Feuille N3 et sa correction en même temps.

Dans chaque feuille de TD, il vous est demandé de rédiger la correction d’un exercice et de me l’envoyer scannée par E-mail ou par la poste, pour pouvoir vous donner des conseils personnels sur la rédaction. Ces devoirs non-obligatoires ne peuvent en aucun cas porter une influence quelconque sur la note finale du module.

Table des matières

1	Rappels de théorie des probabilités	5
2	Simulations de variables aléatoires	11
2.1	Simulation via la fonction de répartition	12
2.1.1	Lois discrètes	12
2.1.2	Application : simulation de lois binomiales	13
2.1.3	Lois données par des fonctions de répartition	14
2.1.4	Application : simulation de la loi exponentielle	15
2.1.5	Application ; simulation d'une loi de Poisson	16
2.2	La méthode du rejet	17
2.2.1	Simulation de v.a.r de densité donnée	17
2.2.2	Variante	21
2.3	Simulations de lois normales	22
2.4	Loi conditionnelle	24
3	Méthode de Monte Carlo	27
3.1	Introduction	27
3.2	Description de la méthode	28
3.3	Variante	32
4	Rudiments de Statistiques	33
4.1	Sondages	33
4.2	Statistiques gaussiennes	34
4.3	Test du chi-deux	38

Chapitre 1

Rappels de théorie des probabilités

Dans ce cours nous supposons connues les bases du calcul des probabilités comme elles sont présentées dans le polycopié du cours LM345.

Espaces probabilisés, variables aléatoires, lois Un espace probabilisé est la donnée

- d'un espace Ω que l'on appelle *l'espace des états*. Quand on modélise une situation physique Ω est l'ensemble des états du système que l'on considère. Bien souvent cet espace gigantesque est inaccessible à l'expérience ;
- d'un sous-ensemble \mathcal{B} de $\mathcal{P}(\Omega)$ qui est *l'ensemble des événements*. Dans une situation concrète c'est l'ensemble de tous les résultats d'expériences que l'on peut effectuer sur le système. En théorie des probabilités cet ensemble \mathcal{B} est une *tribu* (on dit aussi une σ -algèbre) : c'est un sous ensemble de l'ensemble des parties de Ω qui contient l'ensemble vide, est stable par passage au complémentaire et par unions (donc intersections) dénombrables ;
- d'une *probabilité* \mathbf{P} : pour tout événement $A \in \mathcal{B}$ le réel $\mathbf{P}(A)$ est le degré de vraisemblance de l'événement A ; c'est un nombre compris entre 0 et 1. Mathématiquement, une probabilité est une application $\mathbf{P} : \mathcal{B} \rightarrow [0, 1]$ telle que $\mathbf{P}(\Omega) = 1$ et pour toute famille dénombrable d'événements *disjoints deux à deux* $A_i \in \mathcal{B}$, $\mathbf{P}(\cup_{i \in \mathbf{N}} A_i) = \sum_{i \in \mathbf{N}} \mathbf{P}(A_i)$.

On dit qu'une propriété P est vraie \mathbf{P} -presque partout sur Ω si elle est vraie pour un ensemble de $\omega \in \Omega$ dont le complémentaire est (inclus dans) un ensemble de probabilité nulle.

Mentionnons le lemme utile suivant dit de Borel Cantelli : *Si A_1, \dots, A_n, \dots est une suite d'événements telle que $\sum_{k \geq 0} \mathbf{P}(A_k) < \infty$ alors \mathbf{P} -presque tout point de Ω n'appartient qu'à un nombre fini de A_k .*

Une *variable aléatoire réelle* (resp. un *vecteur aléatoire*)¹ sur Ω est une application $X : \Omega \rightarrow \mathbf{R}$ (resp. $X : \Omega \rightarrow \mathbf{R}^m$) telle que pour tout intervalle I ouvert de \mathbf{R} (resp. pavé ouvert de \mathbf{R}^m) $X^{-1}(I)$ est un événement (donc dans \mathcal{B})².

La *loi* d'une variable aléatoire (resp. un vecteur aléatoire) est la mesure de probabilité μ_X définie sur \mathbf{R} (resp. \mathbf{R}^m) muni de sa tribu borélienne (c'est-à-dire la plus petite *tribu engendrée* par les ouverts de \mathbf{R} (resp. de \mathbf{R}^m)). De façon générale, étant donnée une partie \mathcal{S} constituée de sous-ensembles de Ω , il existe une unique tribu $\mathcal{B}(\mathcal{S})$ qui contient les éléments de \mathcal{S} et qui est minimale au sens de l'inclusion pour cette propriété; on l'appelle la tribu engendrée par \mathcal{S} . définie par $\mu_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A)$. De façon concrète la loi d'une v.a est la donnée de toutes les probabilités $\mathbf{P}(X \in I)$ pour tous les intervalles I de \mathbf{R} (resp. pavés de \mathbf{R}^m).

La *fonction de répartition* d'une v.a X est la fonction $F_X : \mathbf{R} \rightarrow [0, 1]$ définie par $F_X(t) = \mathbf{P}(X \leq t)$. C'est une fonction croissante admettant 0 et 1 pour limites respectives en $-\infty$ et ∞ et continue à droite en tout point. Réciproquement, étant donnée une fonction vérifiant ces propriétés, il existe une v.a X dont c'est la fonction de répartition.

Espérance, variance Pour toute v.a.r. positive X il est possible de définir l'*espérance* $\mathbf{E}(X)$; c'est soit un nombre positif ou nul soit $+\infty$. Quand X est une v.a.r. de signe quelconque (ou un vecteur aléatoire) il est possible de définir l'espérance $\mathbf{E}(X) \in \mathbf{R}$ ($\mathbf{E}(X) \in \mathbf{R}^m$ pour les v.a.) toutes les fois où $\mathbf{E}(|X|) < \infty$. On dit dans ce cas que X est *intégrable* ou L^1 . L'espérance est une application linéaire de l'espace vectoriel des fonctions intégrables sur \mathbf{R} qui vérifie le *théorème de convergence dominée* : Soit (X_n) une suite de v.a. intégrables telles que a) $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ pour tout $\omega \in \Omega$; b) il existe une v.a.r. Z intégrable telle que pour tout n et tout ω on ait $|X_n(\omega)| \leq Z(\omega)$; alors i) X est une v.a ii) on a $\lim_{n \rightarrow \infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.

On dit qu'une v.a.r. est L^2 ou de carré intégrable si $\mathbf{E}(X^2) < \infty$. On définit dans ce cas la *variance* de X : $\mathbf{Var}(X) = \mathbf{E}\left(|X - \mathbf{E}(X)|^2\right)$. On a aussi $\mathbf{Var}(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$.

Mentionnons le théorèmes de Markov : si X est une v.a.r. intégrable et $\lambda > 0$ on a $\mathbf{P}(|X| > \lambda) \leq \mathbf{E}(|X|)/\lambda$ et celui de Bienaymé-Tchebychev : si X est une v.a.r. de carré intégrable et $\lambda > 0$ on a $\mathbf{P}(|X - \mathbf{E}(X)| > \lambda) \leq \mathbf{Var}(|X|)/\lambda^2$.

¹nous écrirons souvent v.a.r. pour variable aléatoire réelle et v.a. pour vecteur aléatoire ou variable aléatoire

²On note encore $\{X \in I\}$ l'événement $X^{-1}(I)$ qui par définition est $\{\omega \in \Omega : X(\omega) \in I\}$

Formule de transfert, densité On dit qu'un v.a. admet une *densité* s'il existe une fonction $\rho_X : \mathbf{R}^m \rightarrow [0, \infty[$ telle que pour tout pavé $I \subset \mathbf{R}^m$ on a $\mathbf{P}(X \in I) = \int_I \rho_X(x_1, \dots, x_m) dx_1 \cdots dx_m$. Il est équivalent de dire que pour toute fonction continue bornée $\phi : \mathbf{R}^m \rightarrow \mathbf{R}$ on a

$$\mathbf{E}(\phi(X)) = \int_{\mathbf{R}} \phi(x) \rho_X(x) dx.$$

C'est ce que l'on appelle dans ce cours la *formule de transfert*. Dans le cas où X prend ses valeurs dans un ensemble dénombrable $E \subset \mathbf{R}$

$$\mathbf{E}(\phi(X)) = \sum_{e \in E} \phi(e) \mathbf{P}(X = e).$$

Ces deux formules sont vraies toutes les fois où $\phi(X)$ est L^1 (ce qui est équivalent au fait que $x \mapsto \phi(x) \rho_X(x)$ est L^1 ou dans le deuxième cas que la série de terme général $\phi(e) \mathbf{P}(X = e)$ est absolument convergente.

Combinée à la formule du changement de variable, cette formule permet de calculer les densités de nombreuses v.a ou vecteurs aléatoires.

Quand un vecteur aléatoire $X = (X_1, \dots, X_m)$ admet une densité $\rho_X(x_1, \dots, x_m)$ on peut dire que les v.a X_i admettent chacune une densité ρ_i (appelée *i*-ième densité marginale) donnée par

$$\rho_i(x_i) = \int_{\mathbf{R}} \cdots \int_{\mathbf{R}} \rho_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_m) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_m.$$

Fonctions génératrices, fonctions caractéristiques Si $X : \Omega \rightarrow E \subset \mathbf{N}$ est une v.a à valeurs dans un ensemble dénombrable $E \subset \mathbf{N}$ on appelle *fonction génératrice* de X la série définie pour $0 \leq t < 1$ par

$$\Phi_X(t) = \sum_{k=0}^{\infty} \mathbf{P}(X = k) t^k.$$

Cette fonction permet dans certains cas de calculer les moments de X (c'est-à-dire les $\mathbf{E}(X^k)$), en particulier l'espérance et la variance. La v.a X est intégrable (resp. de carré intégrable) si et seulement si la dérivée (resp. la dérivée seconde) par rapport à t de $\Phi_X(t)$ admet une limite quand $t \rightarrow 1^-$ et dans ce cas $\lim_{t \rightarrow 1^-} \Phi_X'(t) = \mathbf{E}(X)$ (resp. $\lim_{t \rightarrow 1^-} \Phi_X''(t) = \mathbf{E}(X(X-1))$).

De façon générale, pour toute v.a $X : \Omega \rightarrow \mathbf{R}$ on définit la *fonction caractéristique* $\Phi_X(t)$ de X par

$$\Phi_X(t) = \mathbf{E}(e^{itX})$$

($i = \sqrt{-1}$). Quand X admet une densité ρ_X , $\Phi_X(t)$ est la transformée de Fourier de $\rho_X : \Phi_X(t) = \int_{\mathbf{R}} \rho_X(x) e^{itx} dx$. La fonction caractéristique de X caractérise la loi de X : deux v.a ont même loi si et seulement si leurs fonctions caractéristiques sont égales.

Indépendance Une suite de v.a (resp. vecteurs aléatoires) X_1, \dots, X_n, \dots est dite *indépendante* si pour tout n et tous intervalles de \mathbf{R} (resp. pavés de \mathbf{R}^m) I_1, \dots, I_n on a

$$\mathbf{P}(X_1 \in I_1, \dots, X_n \in I_n) = \mathbf{P}(X_1 \in I_1) \cdots \mathbf{P}(X_n \in I_n).$$

On peut dans la définition précédente remplacer "intervalles" (resp. "pavés") par boréliens.

Dans le cadre des v.a admettant des densités, on a le résultat suivant : *Le vecteur aléatoire $X = (X_1, \dots, X_m)$ admet une densité et cette densité est le produit de ses densités marginales (densités de chaque X_i) si et seulement si les v.a X_1, \dots, X_m sont indépendantes.*

Une proposition importante pour vérifier que des v.a. sont indépendantes et la suivante : *Si l'on dispose d'une suite Y_1, Y_2, \dots de v.a. indépendantes et si X_1 s'écrit comme une fonction $f_1(Y_1, \dots, Y_{i_1})$, X_2 comme une fonction $f_2(Y_{i_1+1}, \dots, Y_{i_2})$ etc. X_n comme une fonction $f_n(Y_{i_{n-1}+1}, \dots, Y_{i_n})$ ($i_1 < i_2 < \dots < i_n < \dots$) alors les v.a. X_1, \dots, X_n sont indépendantes.*

Une application de l'indépendance est : Si X_1, \dots, X_n sont indépendantes et intégrables alors

$$\mathbf{E}(X_1 \cdots X_n) = \mathbf{E}(X_1) \cdots \mathbf{E}(X_n).$$

Si en outre elles sont de carrés intégrables

$$\mathbf{Var}(X_1 + \cdots + X_n) = \mathbf{Var}(X_1) + \cdots + \mathbf{Var}(X_n).$$

Un autre critère d'indépendance est le suivant : *X_1, \dots, X_n sont indépendantes, si et seulement si pour tous réels t_1, \dots, t_n*

$$\mathbf{E}(e^{it \cdot X}) = \mathbf{E}(e^{it_1 X_1}) \cdots \mathbf{E}(e^{it_n X_n})$$

où on a noté $t \cdot X = t_1 X_1 + \cdots + t_n X_n$.

Convergence des v.a On dit qu'une suite de v.a X_1, \dots, X_n, \dots converge *presque-sûrement* (et on abrège p.s) vers une v.a X si et seulement si pour \mathbf{P} -presque tout $\omega \in \Omega$ $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$.

Elle converge en *loi* si par définition pour toute fonction continue bornée $\phi : \mathbf{R} \rightarrow \mathbf{R}$ on a $\mathbf{E}(\phi(X_n)) = \mathbf{E}(\phi(X))$. Il y a deux autres définitions équivalentes. La première repose sur la fonction de répartition : *La suite de v.a (X_n) converge en loi vers la v.a X si et seulement si pour tout $t \in \mathbf{R}$ qui est point de continuité de la fonction de répartition $F_X(t) = \mathbf{P}(X \leq t)$ de X on a $\lim_{n \rightarrow \infty} F_{X_n}(t) = F_X(t)$.* La deuxième se formule en termes de fonctions caractéristiques : *La suite de v.a (X_n) converge en loi vers la v.a X si et seulement si pour tout $t \in \mathbf{R}$, la suite de fonctions caractéristiques $\Phi_{X_n}(t)$ converge vers la fonction caractéristique $\Phi_X(t)$ quand $n \rightarrow \infty$.* Mentionnons enfin que la convergence presque-sûre entraîne la convergence en loi.

Exemples classiques *Loi de Bernoulli de paramètre p* : C'est la loi d'une v.a à valeurs dans $\{0, 1\}$. Cette loi est déterminée par $p = \mathbf{P}(X = 1)$. Concrètement elle modélise le lancer d'une pièce de monnaie (pile a la probabilité p de sortir). On a $\mathbf{E}(X) = p$, $\mathbf{Var}(X) = p(1 - p)$.

Loi binomiale de paramètre (n, p) : C'est la loi d'une v.a à valeurs dans $\{0, \dots, n\}$ telle que

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

Elle modélise par exemple la probabilité d'obtenir k pile quand on joue n fois à pile/face (où la probabilité d'obtenir pile vaut p). C'est aussi la loi d'une somme de n v.a indépendantes suivant une loi de Bernoulli de paramètre p . On a $\mathbf{E}(X) = np$, $\mathbf{Var}(X) = np(1 - p)$.

Loi géométrique de paramètre $0 \leq a < 1$: C'est la loi d'une v.a à valeurs dans \mathbf{N} telle que

$$\mathbf{P}(X = k) = (1 - a)a^k.$$

On a $\mathbf{E}(X) = \frac{a}{1-a}$, $\mathbf{Var}(X) = \frac{a}{(1-a)^2}$.

Loi de Poisson de paramètre $\lambda > 0$: C'est la loi d'une v.a à valeurs dans \mathbf{N} telle que

$$\mathbf{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

On a $\mathbf{E}(X) = \lambda$, $\mathbf{Var}(X) = \lambda$. Si $np_n \rightarrow \lambda$ alors la loi binomiale de paramètres (n, p_n) converge en loi vers la loi de Poisson de paramètre λ .

Loi uniforme sur $[a, b]$ C'est la loi d'une v.a à valeurs dans $[a, b]$ de densité

$$\rho(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

On a $\mathbf{E}(X) = (a+b)/2$, $\mathbf{Var}(X) = (b-a)^2/12$.

Loi exponentielle de paramètre θ : C'est la loi d'une v.a à valeurs dans $[0, \infty[$ de densité

$$\rho(x) = \theta e^{-\theta x} \mathbf{1}_{[0, \infty[}(x).$$

On a $\mathbf{E}(X) = 1/\theta$, $\mathbf{Var}(X) = 1/\theta^2$.

Loi normale $\mathcal{N}(\mu, \sigma)$: C'est la loi d'une v.a à valeurs dans \mathbf{R} de densité

$$\rho(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

On a $\mathbf{E}(X) = \mu$, $\mathbf{Var}(X) = \sigma^2$. On dit qu'elle est centrée réduite quand $\mu = 0$, $\sigma = 1$. Une v.a X suit une loi normale centrée réduite si et seulement si $\sigma X + \mu$ suit une loi normale $\mathcal{N}(\mu, \sigma^2)$.

Loi forte des grands nombres Si (X_n) est une suite de v.a intégrables, indépendantes et de même loi, alors la suite des moyennes $(X_1 + \cdots + X_n)/n$ converge $\mathbf{P} - p.s$ vers la v.a constante $\mathbf{E}(X_1)$.

Théorème Central Limit Si (X_n) est une suite de v.a de carrés intégrables, indépendantes et de même loi, alors la suite

$$\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \cdots + X_n}{n} - \mathbf{E}(\mathbf{X}_1) \right)$$

où $\sigma^2 = \mathbf{Var}(X_1)$, converge en loi vers une loi normale centrée réduite.

Chapitre 2

Simulations de variables aléatoires

Nous présentons dans ce chapitre diverses méthodes de simulations de variables aléatoires. Le contexte est le suivant : on suppose que l'on dispose d'un ordinateur qui, toutes les fois que l'on fait appel à la fonction `Random`, produit un nombre réel compris entre 0 et 1 suivant une loi uniforme (dans la pratique ce nombre ne peut être que rationnel). On se propose alors de simuler une variable aléatoire X de loi donnée. De façon plus mathématique, on modélisera le k -ième appel à la fonction `Random` par une v.a.r $U_k(\omega)$ suivant une loi *uniforme* dans l'intervalle $[0, 1]$ et on supposera que la suite de v.a.r U_1, \dots, U_n, \dots est *indépendante*. À présent, supposons que $X : \Omega \rightarrow \mathbf{R}$ soit une v.a. ayant une loi donnée. Si X est à valeurs dans \mathbf{Z} cela signifie que l'on connaît les réels $p_i = \mathbf{P}(X = i)$ et plus généralement si X est à valeurs réelles cela signifie entre autres choses que pour tout intervalle $I \subset \mathbf{R}$ on connaît $\mathbf{P}(X \in I)$. Comment construire X à partir de la suite U_1, \dots, U_n, \dots ?

Un problème plus général et plus intéressant sera de construire une suite de v.a.r X_k , *indépendantes* et de loi *donnée*.

Illustrons sur deux exemples les deux principales méthodes de simulation dont nous parlerons.

Premier exemple. On veut simuler un jeu de pile/face (1/0) avec $\mathbf{P}(X = 1) = p$; en d'autres termes, on veut simuler une variable de Bernoulli à valeurs dans $\{0, 1\}$. Une façon simple de procéder est la suivante : découpons l'intervalle $[0, 1]$ en deux intervalles I_0 et I_1 de longueur respectives $1 - p$ et p : par exemple $I_0 = [0, 1 - p]$ et $I_1 = [1 - p, 1]$ mais on pourrait également choisir $I_0 = [p, 1]$ et $I_1 = [0, p]$. Comme la v.a.r U_1 suit une loi uniforme sur $[0, 1]$ on a

$$\mathbf{P}(U_1 \in I_0) = 1 - p \quad \mathbf{P}(U_1 \in I_1) = p.$$

Si on définit

$$\begin{cases} X_1(\omega) = 1 \text{ si } U_1 \in I_1 \\ X_1(\omega) = 0 \text{ si } U_1 \in I_0 \end{cases}$$

il est facile de voir que X_1 est une v.a.r à valeurs 0 ou 1 et que $\mathbf{P}(X_1 = 1) = \mathbf{P}(U_1 \in I_1) = p$. On a donc démontré que

$$X_1 = \mathbf{1}_{I_1} \circ U_1$$

est une v.a suivant la loi de Bernoulli de paramètre p .

Si l'on veut simuler une *suite* de v.a *indépendantes* suivant une même loi de Bernoulli, il suffit de poser $X_k = \mathbf{1}_{I_1} \circ U_k$.

Second exemple : Essayons de simuler une v.a X qui suit une loi uniforme sur $[0, a]$, $0 \leq a \leq 1$. On cherche donc X telle que

$$\mathbf{P}(X \in I) = \frac{|I \cap [0, a]|}{a},$$

où $|I|$ désigne la longueur de I si I est un intervalle (plus généralement la mesure de Lebesgue de I si I est un borélien de \mathbf{R}). Il est très facile de vérifier que la v.a $X = aU_1$ suit une loi uniforme sur $[0, a]$ et par conséquent $X_k = aU_k$ est une suite de v.a *indépendantes* suivant une loi uniforme sur $[0, a]$.

Il est possible de simuler une loi uniforme sur $[0, a]$ d'une autre façon. Définissons $Y(\omega)$ comme étant la valeur du premier $U_k(\omega)$ qui tombe dans $[0, a] \subset [0, 1]$. De façon plus précise : soit

$$Y(\omega) = U_{\nu(\omega)}(\omega), \text{ avec } \nu(\omega) = \inf\{k \geq 1 : U_k(\omega) \in [0, a]\}.$$

Il n'est pas difficile de voir que Y suit une loi uniforme sur $[0, a]$ (**exercice** ou voir la section sur la méthode du rejet). Si on veut produire une suite de v.a *indépendante* de loi uniforme sur $[0, a]$ on procèdera de la manière suivante : on pose comme précédemment $\nu_1(\omega) = \nu(\omega) = \inf\{k : U_k(\omega) \in [0, a]\}$ et par récurrence $\nu_r(\omega) = \inf\{k > \nu_{r-1}(\omega) : U_k(\omega) \in [0, a]\}$. Puis on pose $Y_k(\omega) = U_{\nu_k(\omega)}(\omega)$. La suite de v.a Y_k est indépendante et chaque Y_k suit une loi uniforme sur $[0, a]$.

2.1 Simulation via la fonction de répartition

2.1.1 Lois discrètes

Soit $X : \Omega \rightarrow \{a_1, \dots, a_r\}$ une v.a ne prenant qu'un nombre fini de valeurs. La loi de X est déterminée par la connaissance des réels $p_i := \mathbf{P}(X =$

a_i). Remarquons que $p_i \in [0, 1]$ et $p_1 + \dots + p_r = 1$. Soit I_1, \dots, I_r une partition de $[0, 1]$ en intervalles de longueurs respectives p_1, \dots, p_r . On peut, par exemple, choisir $I_1 = [0, p_1[$, $I_2 = [p_1, p_1 + p_2[$, $I_s = [p_1 + \dots + p_{s-1}, p_1 + \dots + p_s[$ ($1 \leq s < r$), $I_r = [p_1 + \dots + p_{r-1}, 1[$. Si on pose,

$$X_1(\omega) = a_i \text{ si } U_1 \in I_i$$

c'est-à-dire

$$X_1 = \sum_{i=1}^r a_i \mathbf{1}_{I_i} \circ U_1$$

on voit que $\mathbf{P}(X = a_i) = \mathbf{P}(U_1 \in I_i) = p_i$.

De fa con plus générale si on pose

$$X_k = \sum_{i=1}^r a_i \mathbf{1}_{I_i} \circ U_k$$

on voit que les X_k forment une suite de v.a indépendantes car de la forme $f(U_k)$ où $f = \sum_{i=1}^r a_i \mathbf{1}_{I_i}$ (les U_k sont indépendantes) et de même loi $\mathbf{P}(X_k = a_i) = p_i$.

Exercice : Démontrer que le résultat précédent se généralise au cas d'une v.a prenant un nombre dénombrable de valeurs

2.1.2 Application : simulation de lois binomiales

Par définition une v.a X suit une la loi binomiale (n, p) si elle est à valeurs dans $\{0, 1, \dots, n\}$ et

$$\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

Une fa con simple de simuler une loi binomiale est d'utiliser le fait suivant : Si Z_1, \dots, Z_n est une suite de v.a indépendantes suivant une même loi de Bernoulli de paramètre p ($\mathbf{P}(Z_k = 1) = p$), alors la v.a $X = Z_1 + \dots + Z_n$ suit une loi binomiale (n, p) .

Exercice : Démontrer l'assertion précédente.

Ainsi, pour simuler une suite de v.a indépendantes $(X_k)_{k \geq 1}$ suivant une même loi binomiale (n, p) il suffira de poser

$$X_k = \sum_{i=kn}^{kn+n-1} \mathbf{1}_{[0,p]}(U_i).$$

Exercice : Pourquoi ?

2.1.3 Lois données par des fonctions de répartition

Soit F une *fonction de répartition* c'est-à-dire une fonction $F : \mathbf{R} \rightarrow [0, 1]$ croissante, continue à droite en tout point et admettant 0 comme limite en $-\infty$, 1 comme limite en $+\infty$. Cette fonction F détermine une loi : X admet F pour fonction de répartition si pour tout t , $\mathbf{P}(X \leq t) = F(t)$.

Supposons pour simplifier que F soit strictement croissante et continue en tout point. Il est facile de voir que F est un homéomorphisme croissant de \mathbf{R} sur $]0, 1[$ (c'est-à-dire une bijection continue et croissante dans les deux sens). On peut donc définir $G : \mathbf{R} \rightarrow]0, 1[$ par $G = F^{-1} : G(x) = t$ si et seulement si $F(t) = x$.

Remarquons à présent que si U suit une loi uniforme sur $[0, 1]$ on a

$$F(t) = \mathbf{P}(U \in [0, F(t)])$$

car $F(t) \in [0, 1]$ et par conséquent

$$F(t) = \mathbf{P}(U \leq F(t)).$$

Mais $\mathbf{P}(U \leq F(t)) = \mathbf{P}(F^{-1}(U) \leq t)$ car F est strictement croissante. On a donc démontré que la v.a $X = F^{-1}(U)$ admet F pour fonction de répartition.

Le résultat précédent se généralise au cas où la fonction de répartition F n'est que *croissante et continue à droite* (et n'est donc pas nécessairement un homéomorphisme). Le lemme suivant indique ce qu'il faut faire :

Lemme 2.1.1 Soient $F : \mathbf{R} \rightarrow [0, 1]$ une fonction de répartition et $G :]0, 1[\rightarrow \mathbf{R}$ la fonction définie par

$$G(u) = \inf\{s \in \mathbf{R} : u \leq F(s)\}.$$

Alors,

i) *l'inf est en fait un min ;*

ii) *pour tous $t \in \mathbf{R}$, $u \in]0, 1[$ on a*

$$u \leq F(t) \iff G(u) \leq t.$$

Démonstration. — Démontrons i). Il suffit pour cela de démontrer que pour $u \in]0, 1[$ l'ensemble $E_u = \{s \in \mathbf{R} : u \leq F(s)\}$ est un ensemble fermé, non vide borné inférieurement (car dans ce cas l'inf est un min). Que E_u soit non vide est clair et qu'il soit borné inférieurement résulte du fait que si ce n'était pas le cas, on pourrait trouver une suite s_n tendant vers $-\infty$ et telle que $F(s_n) \geq u$. Mais comme la limite de F en $-\infty$ vaut 0, on aurait $u = 0$ ce qui est contraire à l'hypothèse $u \in]0, 1[$. Pour démontrer que c'est un fermé,

il suffit de démontrer, u et s étant fixés, que si une suite de $s_n \in E_u$ converge vers un réel s on a toujours $s \in E_u$, c'est-à-dire $u \leq F(s)$: s'il existe n tel que $s_n \leq s$ c'est clair car comme F est croissante, $u \leq F(s_n) \leq F(s)$; sinon, on a pour tout n , $s_n > s$ et on peut extraire de la suite s_n une sous-suite s_{n_k} qui converge vers s en décroissant. Comme F est continue à droite $F(s_{n_k}) \rightarrow F(s)$ et en passant à la limite dans l'inégalité $u \leq F(s_{n_k})$ on obtient encore $u \leq F(s)$.

ii) Si $u \leq F(t)$, alors $t \in E_u$ et par définition $t \geq \inf E_u = G(u)$. Réciproquement, si $G(u) \leq t$, comme on a vu que $\inf E_u$ est un min (E_u est fermé) on peut dire que $G(u) \in E_u$. Mais par définition de E_u cela signifie que $F(G(u)) \geq u$. Comme F est croissante $F(t) \geq F(G(u)) \geq u$.

□

Une conséquence du lemme précédent est le théorème suivant :

Théorème 2.1.1 *Si F est une fonction de répartition et U_1, \dots, U_n, \dots une suite de v.a.r indépendantes suivant une loi uniforme sur $[0, 1]$, alors la suite de v.a.r $(X_k)_{k \geq 1}$ où $X_k = G(U_k)$ est indépendante et chaque X_k suit une loi dont la fonction de répartition est F .*

Démonstration.— Comme $F(t)$ est dans $[0, 1]$ et que U_k suit une loi uniforme sur $[0, 1]$ on a $F(t) = \mathbf{P}(U_k \leq F(t))$; mais d'après le lemme précédent $\{U_k \leq F(t)\} = \{G(U_k) \leq t\}$ si bien que $F(t) = \mathbf{P}(X_k \leq t)$ ce qui signifie que chaque v.a X_k suit une loi dont la fonction de répartition est F . Les v.a.r X_k sont bien évidemment indépendantes (puisque de la forme $G(U_k)$ où les U_k sont indépendantes).

□

2.1.4 Application : simulation de la loi exponentielle

La loi exponentielle de paramètre $\lambda > 0$ est la loi de densité $\lambda \mathbf{1}_{t > 0}(t)e^{-\lambda t}$. Sa fonction de répartition $F(t)$ vaut 0 si $t < 0$ et si $t \geq 0$

$$F(t) = \int_0^t \lambda e^{-\lambda s} ds = 1 - e^{-\lambda t}.$$

Cette fonction est croissante et continue et sa fonction de répartition qui est un homéomorphisme de $]0, \infty[$ sur $]0, 1[$ s'inverse facilement :

$$F(t) = u \iff t = -\frac{1}{\lambda} \log(1 - u).$$

On voit que $G(u) = -\frac{1}{\lambda} \log(1 - u)$. Ainsi, $X = -\frac{1}{\lambda} \log(1 - U)$ est une v.a. suivant une loi exponentielle de paramètre λ . Puisque $1 - U$ suit également une loi uniforme sur $[0, 1]$ on voit que $Y = -\frac{1}{\lambda} \log(U)$ suit également une loi exponentielle de paramètre λ .

2.1.5 Application ; simulation d'une loi de Poisson

Il est facile de simuler une loi de Poisson de paramètre $\lambda > 0$ ($\mathbf{P}(X = k) = e^{-\lambda}(\lambda^k/k!)$) en utilisant le résultat de la section précédente et le théorème suivant.

Théorème 2.1.2 *Soit $(T_k)_{k \geq 1}$ une suite de v.a.r indépendantes suivant une même loi exponentielle de paramètre λ . Alors, la v.a*

$$\nu_{t,\lambda} = \max\{k \geq 1 : T_1 + \dots + T_k \leq t\}$$

suit une loi de Poisson de paramètre $t\lambda$.

Démonstration. — Notons pour $k \geq 1$, $S_k = T_1 + \dots + T_k$. Comme la suite $(T_i)_{i \geq 1}$ est indépendante et que chaque T_i suit une loi de densité $\lambda e^{-\lambda} \mathbf{1}_{[0, \infty)}(\cdot)$, il n'est pas difficile de voir que S_k suit une loi de densité $\rho_{S_k}(t)$, nulle pour $u \leq 0$, et donnée par

$$\rho_{S_k}(u) := \lambda^k \exp(-\lambda u) \int_{0 \leq u_1 \leq u_2 \leq \dots \leq u_{k-1} \leq u} du_1 \dots du_{k-1}$$

pour $u \geq 0$. En effet, pour toute f continue bornée

$$\mathcal{E}f(T_1 + \dots + T_k) = \int_{t_1 \geq 0, t_2 \geq 0, \dots, t_k \geq 0} f(t_1 + \dots + t_k) \lambda^k \exp(-\lambda(t_1 + \dots + t_k)) dt_1 \dots dt_k$$

Par le changement de variables $t_1 = u_1, t_1 + t_2 = u_2, \dots, t_1 + \dots + t_{k-1} = u_{k-1}, t_1 + \dots + t_k = u$, on a $t_1 = u_1, t_2 = u_2 - u_1, \dots, t_{k-1} = u_{k-1} - u_{k-2}, t_k = u - u_{k-1}$ de Jacobien égal à 1 on a

$$\mathcal{E}f(T_1 + \dots + T_k) = \int_{u \geq 0} f(u) \lambda^k \exp(-\lambda u) \int_{0 \leq u_1 \leq \dots \leq u_{k-1} \leq u} du_1 \dots du_{k-1}$$

Comme l'expression sous l'intégrale est invariante par permutation des indices $1, \dots, k-1$ et puisque $[0, u]^{k-1}$ est l'union disjointe (aux bords près) des ensembles $\{0 \leq u_{\sigma(1)} \leq u_{\sigma(2)} \leq \dots \leq u_{\sigma(k-1)} \leq u\}$ (σ décrivant toutes les $(k-1)!$ permutations de $\{1, \dots, k-1\}$) on déduit que

$$\rho_{S_k}(u) = \frac{u^{k-1}}{(k-1)!} \lambda^k e^{-\lambda u}.$$

L'événement $\{\nu_{t,\lambda} = k\}$ égale $\{S_k \leq t \text{ et } S_k + T_{k+1} > t\}$ et sa probabilité vaut $\mathbf{E}(\mathbf{1}_{S_k \leq t} \mathbf{1}_{S_k + T_{k+1} > t})$. On a

$$\begin{aligned} \mathbf{E}(\mathbf{1}_{S_k \leq t} \mathbf{1}_{S_k + T_{k+1} > t}) &= \int_{S \leq t} \int_{T > t-S} \frac{S^{k-1}}{(k-1)!} \lambda^k \exp(-\lambda S) \lambda \exp(-\lambda T) dS dT \\ &= \int_0^t \frac{S^{k-1}}{(k-1)!} \lambda^k \exp(-\lambda S) \exp(-\lambda(t-S)) dS \\ &= \exp(-\lambda t) ((k-1)!)^{-1} \lambda^k \int_0^t S^{k-1} dS = \exp(-\lambda t) (k!)^{-1} (\lambda t)^k \end{aligned}$$

□

2.2 La méthode du rejet

2.2.1 Simulation de v.a.r de densité donnée

La méthode de la section précédente ne permet de simuler que des v.a dont on connaît de façon plus ou moins explicite la fonction de répartition. Dans la pratique la description de la densité d'une v.a sera plus explicite que celle de sa fonction de répartition et il est souhaitable d'avoir à notre disposition une méthode faisant appel à la densité de la v.a plutôt qu'à sa fonction de répartition.

Décrivons la méthode en faisant les hypothèses suivantes sur la densité ρ de la v.a X que l'on veut simuler :

Hypothèses : i) la densité ρ est à support compact dans un intervalle $[a, b]$: cela signifie que pour tout $x \in \mathbf{R} - [a, b]$ on a $\rho(x) = 0$.

ii) la densité ρ est continue (ou continue par morceaux) sur $[a, b]$ et majorée par une constante c : pour tout $x \in [a, b]$, $0 < \rho(x) \leq c$.

Méthode : a) On simule une suite indépendante de *vecteurs* aléatoires $Z_k := (V_k, V'_k)$ suivant une loi uniforme dans le rectangle $R = [a, b] \times [0, c]$.

Pour cela on procède de la façon suivante : on définit pour $k \geq 1$ les suites

$$V_k = a + (b - a)U_{2k-1}, \quad V'_k = cU_{2k}.$$

Exercice : i) Démontrer que la suite $V_1, V'_1, V_2, V'_2, \dots$ est indépendante et que pour $k \geq 1$ V_k, V'_k suivent respectivement une loi uniforme sur $[a, b]$ et $[0, c]$.

ii) Démontrer que le vecteur (V_k, V'_k) suit une loi uniforme sur le rectangle $[a, b] \times [0, c]$. ([La densité du vecteur (V_k, V'_k) est $(1/(b-a))\mathbf{1}_{[a,b]}(x)(1/c)\mathbf{1}_{[0,c]}(y)$ car V_k et V'_k sont indépendants.]

Si i on pose $\mathcal{F}_I = \{(v, v') \in [a, b] \times [0, c] : v \in I, v' \leq \rho(v)\}$, alors on obtient le corollaire de l'exercice :

$$\mathbf{P}((V_k, V'_k) \in \mathcal{F}_I) = \frac{\text{aire}(\mathcal{F}_I)}{\text{aire}(R)} = \frac{\int_{x \in I} \int_0^{\rho(x)} dy dx}{c(a-b)} = \frac{\int_I \rho(x) dx}{(b-a)c}.$$

b) On définit $\nu(\omega)$ comme étant le premier temps k où $(V_k(\omega), V'_k(\omega))$ tombe dans le sous-graphe de $\rho : \mathcal{E} = \{(v, v') \in [a, b] \times [0, c] : v' \leq \rho(v)\}$:

$$\nu(\omega) = \inf\{k \geq 1 : V'_k(\omega) \leq \rho(V_k(\omega))\}.$$

La fonction $\nu : \Omega \rightarrow \mathbf{N} \cup \{\infty\}$ est une variable aléatoire puisque pour tout n on a en posant $Z_k := (V_k, V'_k)$

$$\{\nu = n\} = (Z_1 \in \mathcal{E}^c) \cap \dots \cap (Z_{n-1} \in \mathcal{E}^c) \cap (Z_n \in \mathcal{E}).$$

Comme les vecteurs (Z_k) sont indépendants on a

$$\begin{aligned} \mathbf{P}(\nu = n) &= \mathbf{P}(Z_1 \in \mathcal{E}^c) \dots \mathbf{P}(Z_{n-1} \in \mathcal{E}^c) \mathbf{P}(Z_n \in \mathcal{E}) \\ &= (1 - \alpha)^{n-1} \alpha \end{aligned}$$

où $\alpha = \mathbf{P}(Z_k \in \mathcal{E})$. Puisque Z_k suit une loi uniforme sur R , on a

$$\alpha =: \frac{1}{(b-a)c} \int_a^b \rho(x) dx = \frac{1}{(b-a)c}.$$

On a

$$\begin{aligned} \mathbf{P}(\nu < \infty) &= \sum_{n=1}^{\infty} \mathbf{P}(\nu = n) \\ &= \sum_{n=1}^{\infty} \alpha (1 - \alpha)^{n-1} \\ &= \alpha \frac{1}{1 - (1 - \alpha)} \\ &= 1. \end{aligned}$$

Par conséquent ν est presque toujours finie.

c) On projette le vecteur (V_ν, V'_ν) sur l'axe des x ; on pose $X = V_\nu$ ou de facon plus précise

$$X(\omega) = V_{\nu(\omega)}(\omega).$$

Théorème 2.2.1 *La v.a.r X suit une loi de densité ρ .*

Démonstration.— Evaluons la loi de X : X est clairement à valeurs dans $[a, b]$ et pour tout intervalle $I \in [a, b]$, comme Ω est union disjointe des $\{\nu = n\}$ pour n décrivant \mathbf{N}^*

$$\begin{aligned} \mathbf{P}(X \in I) &= \mathbf{P}\left(X \in I \cap \bigcup_{n \geq 1} \{\nu = n\}\right) \\ &= \sum_{n \geq 1} \mathbf{P}(X \in I \text{ et } \nu = n) \\ &= \sum_{n \geq 1} \mathbf{P}(V_n \in I \text{ et } \nu = n) \\ &= \sum_{n \geq 1} \mathbf{P}(Z_1 \in \mathcal{E}^c) \cap \cdots \cap (Z_{n-1} \in \mathcal{E}^c) \cap (Z_n \in \mathcal{F}_I)) \end{aligned}$$

où $\mathcal{F}_I = \{(v, v') \in [a, b] \times [0, c] : v \in I, v' \leq \rho(v)\}$. Comme les vecteurs Z_k sont indépendants et puisque (voir ci-dessus, le corollaire de l'exercice précédent)

$$\mathbf{P}(Z_n \in \mathcal{F}_I) = \frac{\text{aire}(\mathcal{F}_I)}{\text{aire}(R)} = \frac{\int_I \rho(x) dx}{(b-a)c}$$

on obtient

$$\begin{aligned} \mathbf{P}(X \in I) &= \sum_{n \geq 1} (1 - \alpha)^{n-1} \frac{\int_I \rho(x) dx}{(b-a)c} \\ &= \frac{1}{\alpha} \frac{\int_I \rho(x) dx}{(b-a)c} \\ &= \int_I \rho(x) dx \end{aligned}$$

Cette dernière égalité, vraie pour tout intervalle I , prouve que X admet une densité et que celle-ci égale ρ . □

On peut également obtenir une suite de v.a.r indépendantes de densité ρ .

Théorème 2.2.2 *Si on pose $\nu_1 = \nu$ et par récurrence*

$$\nu_r = \inf\{k \geq \nu_{r-1} + 1 : V'_k \leq \rho(V_k)\},$$

alors la suite X_k définie par $X_k(\omega) = V_{\nu_k(\omega)}(\omega)$ est une suite de v.a.r indépendantes de densité ρ .

Démonstration.— Calculons pour tout r et tous intervalles I_1, \dots, I_r , $\mathbf{P}(X_1 \in I_1, \dots, X_r \in I_r)$. A cet effet, notons pour $k_1 < k_2 < \dots < k_r$, $\mathcal{F}_{I_1, \dots, I_r}(k_1, \dots, k_r)$, l'événement

$$\mathcal{F}_{I_1, \dots, I_r}(k_1, \dots, k_r) = \left(\bigcap_{j \notin \{k_1, \dots, k_r\}} (Z_j \in \mathcal{E}^c) \right) \cap \left(\bigcap_{i=1}^r (Z_{k_i} \in \mathcal{F}_{I_{k_i}}) \right). \quad (2.1)$$

L'événement $(X_1 \in I_1) \cap \dots \cap (X_r \in I_r)$ est l'union disjointe de tous les événements de la forme

$$\bigcap_{i=1}^r (X_i \in I_i) \cap \bigcap_{i=1}^r (\nu_i = k_i)$$

pour tous les r -uplet (k_1, \dots, k_r) tels que $1 \leq k_1 < \dots < k_r$. Comme

$$\bigcap_{i=1}^r (X_i \in I_i) \cap \bigcap_{i=1}^r (\nu_i = k_i) = \mathcal{F}_{I_1, \dots, I_r}(k_1, \dots, k_r),$$

on a

$$\mathbf{P}((X_1 \in I_1) \cap \dots \cap (X_r \in I_r)) = \sum_{k_1 < \dots < k_r} \mathbf{P}(\mathcal{F}_{I_1, \dots, I_r}(k_1, \dots, k_r)),$$

et puisque les Z_i sont indépendants, l'équation (2.1) montre que

$$\begin{aligned} \mathbf{P}((X_1 \in I_1) \cap \dots \cap (X_r \in I_r)) = \\ \sum_{1 \leq k_1 < \dots < k_r} \left((1 - \alpha)^{k_1 - 1} \frac{\int_{I_1} \rho(x) dx}{(b - a)c} (1 - \alpha)^{k_2 - k_1 - 1} \frac{\int_{I_2} \rho(x) dx}{(b - a)c} \right. \\ \left. \dots (1 - \alpha)^{k_r - k_{r-1} - 1} \frac{\int_{I_r} \rho(x) dx}{(b - a)c} \right). \end{aligned}$$

Si on fait le changement de variables $l_1 = k_1 - 1, l_2 = k_2 - k_1 - 1, \dots, l_r = k_r - k_{r-1} - 1$ dans la somme précédente on obtient

$$\begin{aligned} \mathbf{P}((X_1 \in I_1) \cap \dots \cap (X_r \in I_r)) = \\ = \sum_{l_1=0}^{\infty} (1 - \alpha)^{l_1} \dots \sum_{l_r=0}^{\infty} (1 - \alpha)^{l_r} \frac{\int_{I_1} \rho(x) dx}{(b - a)c} \dots \frac{\int_{I_r} \rho(x) dx}{(b - a)c} \\ = \left(\frac{1}{\alpha} \right)^r \frac{\int_{I_1} \rho(x) dx}{(b - a)c} \dots \frac{\int_{I_r} \rho(x) dx}{(b - a)c} \\ = \int_{I_1} \rho(x) dx \dots \int_{I_r} \rho(x) dx \end{aligned}$$

puisque $1/\alpha = (b-a)c$. Ceci démontre que le vecteur (X_1, \dots, X_r) admet une densité $\rho(x_1, \dots, x_r) = \rho(x_1) \cdots \rho(x_r)$ et la forme de cette densité montre simultanément que les v.a.r X_1, \dots, X_r sont indépendantes et admettent ρ pour densité.

□

Quel est le temps moyen de calcul de X_1 ? Celui-ci fait intervenir ν appels à la fonction Random. Il s'agit donc de calculer $\mathbf{E}(\nu)$:

$$\begin{aligned} \mathbf{E}(\nu) &= \sum_{k=0}^{\infty} k \mathbf{P}(\nu = k) \\ &= \sum_{k=1}^{\infty} k (1 - \alpha)^{k-1} \alpha. \end{aligned}$$

Le calcul de cette série s'effectue de la façon suivante : on reconnaît α que multiplie la dérivée par rapport à α de $-\sum_{k=0}^{\infty} (1 - \alpha)^k$:

$$\sum_{k=1}^{\infty} k (1 - \alpha)^{k-1} \alpha = -\alpha \frac{d}{d\alpha} \left(\frac{1}{\alpha} \right) = 1/\alpha = (b - a)c.$$

Ainsi, $\mathbf{E}(\nu) = \alpha^{-1} = (b - a)c$ est l'aire du rectangle contenant le graphe de la densité. On aura donc intérêt à choisir ce rectangle le plus petit possible.

2.2.2 Variantes

Signalons les variantes suivantes de la méthode du rejet. On se propose de simuler la fonction caractéristique d'un ouvert U de \mathbf{R}^d , par exemple du disque de centre 0 et de rayon 1 dans le plan. Une première méthode consiste à utiliser les coordonnées polaires.

Exercice : Soit U et V deux v.a indépendantes suivant une loi uniforme sur $[0, 1]$ et posons $R = \sqrt{U}$, $\Theta = 2\pi V$. Soient

$$\begin{cases} X &= R \cos(\Theta) \\ Y &= R \sin(\Theta) \end{cases}$$

Démontrer que la loi du vecteur aléatoire (X, Y) suit une loi uniforme sur le disque $D(0, 1)$.

Une deuxième méthode est d'utiliser la méthode du rejet

Exercice : Soit U_1, \dots, U_n, \dots une suite de v.a indépendantes suivant une même loi uniforme sur $[0, 1]$.

a) Démontrer que le vecteur $Z_i = (2U_{2i-1} - 1, 2U_{2i} - 1)$ ($i \geq 1$) suit une loi uniforme sur le carré $R = [-1, 1] \times [-1, 1]$.

b) Le disque $D(0, 1)$ de centre 0 et de rayon 1 est inclus dans le carré R . Posons

$$\nu = \inf\{k \geq 1 : Z_k \in D(0, 1)\},$$

et $W(\omega) = Z_{\nu(\omega)}(\omega)$. Montrer que W suit une loi uniforme sur $D(0, 1)$.

c) Posons, $\nu_0 = 0$ et par récurrence

$$\nu_r = \inf\{k \geq \nu_{r-1} + 1 : Z_k \in D(0, 1)\}.$$

Montrer que les vecteurs aléatoires $W_r(\omega) = Z_{\nu_r(\omega)}(\omega)$ suit une loi uniforme sur $D(0, 1)$.

d) Quel est le temps moyen de calcul de W ?

L'intérêt de la deuxième méthode est qu'elle fonctionne en toute dimension. Une conséquence est que l'on est capable de simuler des v.a admettant une densité uniforme sur la *sphère* de dimension quelconque : précisons ce que cela signifie. L'aire d'un ouvert U de la sphère \mathbf{S}^d est le volume du cône C_U s'appuyant sur U : $\text{aire}(U) = \text{vol}(C_U)$ où

$$C_U = \{tx : t \in [0, 1], x \in U\}.$$

Exercice : Simuler par la méthode du rejet une suite de vecteurs aléatoires indépendants W_k admettant une densité uniforme sur la boule euclidienne de dimension $d + 1$ et de rayon 1.

On en déduit que le vecteur $S_k = W_k / \|W_k\|$ suit une loi uniforme sur la sphère de dimension d (et de rayon 1). En effet, pour U un ouvert sur la sphère

$$P(S_k \in U) = P(W_k \in \|W_k\|U),$$

Comme W_k est de loi uniforme sur la boule, cette dernière probabilité est le volume du cône $\text{vol}(C_U)$. Ce volume est par ce qui est écrit dessus est bien $\text{aire}(U)$. Donc $P(S_k \in U) = \text{aire}(U)$.

2.3 Simulations de lois normales

Les lois normales ayant une importance prépondérantes en Probabilités et en Statistiques il est naturel de savoir les simuler facilement. Remarquons

que pour simuler une v.a Y suivant loi normale $\mathcal{N}(\mu, \sigma^2)$ de moyenne μ et de variance σ^2 il suffit de savoir simuler une v.a suivant une loi normale centrée réduite $\mathcal{N}(0, 1)$ puisque la v.a X définie par $Y = \mu + \sigma X$ suit une loi $\mathcal{N}(\mu, \sigma^2)$. Nous présentons dans ce qui suit une méthode pratique de simulation dite méthode de Box-Müller.

Théorème 2.3.1 *Soient U, V deux v.a.r indépendantes de loi uniforme sur $]0, 1]$. Alors, les variables aléatoires Z et T définies par*

$$Z = \sqrt{-2 \log U} \cos(2\pi V), \quad T = \sqrt{-2 \log U} \sin(2\pi V),$$

suivent toutes deux une loi normale centrée réduite. En outre, Z et T sont indépendantes. Par conséquent, si $U_1, V_1, U_2, V_2, \dots, U_n, V_n, \dots$ est une suite de v.a i.i.d de loi uniforme sur $[0, 1]$, la suite de v.a $Z_1, T_1, Z_2, T_2, \dots, Z_n, T_n, \dots$ (obtenue par le procédé précédent) est une suite de v.a i.i.d de loi $\mathcal{N}(0, 1)$.

Démonstration. — Soit $f(z, t)$ une fonction continue bornée de \mathbf{R}^2 dans \mathbf{R} . On a

$$\mathbf{E}(f(Z, T)) = \mathbf{E}(f(\sqrt{-2 \log U} \cos(2\pi V), \sqrt{-2 \log U} \sin(2\pi V))).$$

Comme les v.a U et V sont indépendantes et suivent une loi uniforme sur $[0, 1]$ la densité du vecteur (U, V) est $\mathbf{1}_{[-1, 1]}(u) \mathbf{1}_{[-1, 1]}(v)$; la formule de transfert permet donc d'écrire et d'après la formule de transfert

$$\mathbf{E}(f(Z, T)) = \int_{[0, 1]^2} f(\sqrt{-2 \log u} \cos(2\pi v), \sqrt{-2 \log u} \sin(2\pi v)) du dv.$$

Notons $\phi :]0, 1]^2 \rightarrow \mathbf{R}^2 - \{0\}$ l'application définie par

$$\phi(u, v) = (\sqrt{-2 \log u} \cos(2\pi v), \sqrt{-2 \log u} \sin(2\pi v)).$$

C'est un difféomorphisme dont l'inverse est donné par

$$(u, v) = \phi^{-1}(z, t) = (e^{-(z^2+t^2)/2}, \frac{1}{2\pi} \text{Arcos}(\frac{z}{\sqrt{z^2+t^2}})).$$

Le jacobien de ϕ^{-1} est $1/J(\phi) \circ \phi^{-1}$ où $J(\phi)$ est le jacobien de ϕ :

$$\begin{aligned} J(\phi) &= \det \begin{pmatrix} \frac{-1}{u} (-2 \log u)^{-1/2} \cos(2\pi v) & -2\pi (-2 \log u)^{1/2} \sin(2\pi v) \\ \frac{-1}{u} (-2 \log u)^{-1/2} \sin(2\pi v) & 2\pi (-2 \log u)^{1/2} \cos(2\pi v) \end{pmatrix} \\ &= \frac{-2\pi}{u} \end{aligned}$$

Ainsi,

$$J(\phi^{-1})(z, t) = \frac{-1}{2\pi} e^{-(z^2+t^2)/2}.$$

On a donc montré que pour toute fonction continue bornée f

$$\begin{aligned} \mathbf{E}(f(Z, T)) &= \int_{\mathbf{R}^2} f(z, t) |J(\phi^{-1})(z, t)| dz dt \\ &= \int_{\mathbf{R}^2} f(z, t) \frac{1}{2\pi} e^{-(z^2+t^2)/2} dz dt. \end{aligned}$$

Ceci démontre que le vecteur (Z, T) admet une densité $\rho(z, t)$

$$\rho(z, t) = \frac{1}{2\pi} e^{-(z^2+t^2)/2} = \frac{1}{\sqrt{2\pi}} e^{-(z^2)/2} \frac{1}{\sqrt{2\pi}} e^{-(t^2)/2}.$$

Comme $\rho(z, t) = n(z)n(t)$ où n est la densité d'une loi normale centrée réduite, on déduit d'une part que Z et T sont indépendantes et d'autre part qu'elles suivent chacune une loi normale centrée réduite.

La dernière assertion résulte du critère d'indépendance par paquets.

□

2.4 Loi conditionnelle

Soit $(\Omega, \mathcal{B}, \mathbf{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbf{R}$ une variable aléatoire. La loi de X est la mesure de probabilité μ_X définie sur $(\mathbf{R}, \text{Bor}(\mathbf{R}))$ définie par $\mu_X(\cdot) = \mathbf{P}(X^{-1}(\cdot))$. Si cette mesure μ_X admet une densité, c'est-à-dire s'il existe une fonction $\rho \in L^1(\mathbf{R}, dx)$ telle que pour tout intervalle de \mathbf{R} , $\mu_X(I) = \int_I \rho(x) dx$, on dit que la mesure μ_X , et par extension la v.a. X , admet ρ pour densité.

Supposons à présent que $A \in \mathcal{B}$ soit un événement tel que $\mathbf{P}(A) > 0$. Il est facile de définir sur \mathcal{B} la *probabilité conditionnelle par rapport à A* $\mathbf{P}_{|A} = \mathbf{P}(\cdot|A)$ c'est-à-dire telle que $\mathbf{P}(B) = \mathbf{P}(B \cap A)/\mathbf{P}(A)$ pour tout événement $B \in \mathcal{B}$. On peut à présent définir la *loi conditionnelle* $\mu_{X|A}$ de X par rapport à l'événement A en posant

$$\mu_{X|A} = \mathbf{P}_{|A}(X^{-1}(\cdot)) = \mathbf{P}(X^{-1}(\cdot)|A).$$

Si cette loi admet une densité $\rho_{|A}$ on l'appelle la loi conditionnelle par rapport à l'événement A .

Il est naturel de se poser la question suivante : étant donné un couple (X, A) de la v.a. X et de l'événement A comme précédemment, est-il possible

de construire une v.a. $Z_{X,A}$ sur $(\Omega, \mathcal{B}, \mathbf{P})$ qui admet pour loi/densité (par rapport à la probabilité \mathbf{P}) la loi/densité conditionnelle de X par rapport à A ? C'est-à-dire on souhaite simuler Z tel que

$$\mathbf{P}(Z \in I) = \frac{\mathbf{P}((X \in I) \cap A)}{\mathbf{P}(A)}$$

pour tout interval I . Et encore plus : une suite Z_1, \dots, Z_n de v.a. indépendantes de même loi que Z .

En utilisant la même construction que dans la méthode du rejet, il est facile de voir que c'est bien le cas et que l'on peut en fait construire une suite Z_1, \dots, Z_n, \dots indépendante admettant $\mu_{X|A}$ pour loi.

En effet, soit $(X_1, 1_{A_1}), \dots, (X_n, 1_{A_n}), \dots$ une suite de couples indépendants de v.a. de même loi que $(X, 1_A)$ (ici la fonction indicatrice $1_A(\omega) = 1$ si $\omega \in A$, cad A se réalise, et $1_A(\omega) = 0$ si $\omega \notin A$, cad A ne se réalise pas.)

Posons $\nu_0 = 0$ et par récurrence

$$\nu_r = \inf\{k > \nu_{r-1} : 1_{A_k} = 1\}$$

le premier instant lorsque l'événement A_k se réalise. Définissons la v.a. $Z_r(\omega) = X_{\nu_r(\omega)}(\omega)$, $\omega \in \Omega$. On a le résultat suivant :

Proposition 2.4.1 *La suite de v.a. $(Z_r)_{r \geq 1}$ est indépendante et chaque Z_r admet $\mu_{X|A}$ pour loi.*

Démonstration. Pour $I \subset A$

$$\begin{aligned} P(Z_{\nu_1} \in I) &= \sum_{n \geq 1} P(1_{A_1} = 0, \dots, 1_{A_{n-1}} = 0, 1_{A_n} = 1, X_n \in I) \\ &= \sum_{n \geq 0} (1 - P(A_1)) \cdots (1 - P(A_n)) P((X_n \in I) \cap A_n) \\ &= P((X \in I) \cap A) \sum_{n \geq 0} (1 - P(A))^{n-1} \\ &= P((X \in I) \cap A) \frac{1}{P(A)}. \end{aligned}$$

Donc Z_{ν_1} est de loi voulue.

La suite de la démonstration est triviale du fait que X_1, X_2, \dots sont de même loi et indépendantes.

Comme application de ce que nous venons de dire nous présentons un autre procédé important pour simuler des lois normales.

Théorème 2.4.1 Soient X, Y deux v.a indépendantes et de même loi exponentielle de paramètre 1. Notons A l'événement $\{Y > \frac{1}{2}(1 - X)^2\}$ et $\mu_{X|A}$ la loi conditionnelle de X par rapport à A . Alors, cette loi admet une densité égale à

$$\rho_{X|A}(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \mathbf{1}_{[0, \infty[}(x).$$

Démonstration. Pour tout $I \in]0, \infty[$

$$P(X \in I \mid Y > (1/2)(1 - X)^2) = \frac{P(X \in I, Y > (1/2)(1 - X)^2)}{P(Y > (1/2)(1 - X)^2)}.$$

Le numérateur vaut

$$\begin{aligned} & \int_{x \in I} \int_{y > (1/2)(1-x)^2} \exp(-x - y) dx dy \\ &= \int_I \exp(-x - (1/2)(1 - x)^2) dx = \exp(-1/2) \int_I \exp(-x^2/2) dx \end{aligned}$$

et le dénominateur vaut par le même calcul avec $I = [0, \infty[$

$$\exp(-1/2) \int_0^\infty \exp(-x^2/2) dx = \exp(-1/2) \sqrt{2\pi}/2.$$

Le Thm est prouvé.

Pour simuler une v.a. de loi Normale centrée réduite, on simule une suite de v.a. indépendantes de loi exponentielle de paramètre 1 (par exemple par la méthode d'inversion de la fonction de repartition) $X_1, Y_1, X_2, Y_2, \dots$. Soit $A_i = \{\omega : Y_i > (1/2)(1 - X_i)^2\}$, $i = 1, 2, \dots$. Alors on a une suite $(X_1, 1_{A_1}), (X_2, 1_{A_2}), \dots$. On pose $\nu = \min\{i \geq 1 : Y_i > (1/2)(1 - X_i)^2\} = \min\{i : 1_{A_i} = 1\}$.

Alors par la Proposition précédente X_ν est de loi conditionnelle $\mathbf{P}(X_\nu \in I) = \mathbf{P}((X \in I) \cap A) / \mathbf{P}(A)$. Par le Thm précédent cette loi est de densité $\frac{2}{\sqrt{2\pi}} e^{-x^2/2} \mathbf{1}_{[0, \infty[}(x)$.

Simulons ensuite une v.a. Z (indépendante de X_1, Y_1, \dots) de loi de Bernoulli de paramètre 1/2 comme par exemple $Z = 1_{\{U < 1/2\}} - 1_{\{U \geq 1/2\}}$, avec U une v.a. de loi uniforme sur $[0, 1]$.

Alors $Z \cdot X_\nu$ est de loi Normale centrée réduite (prouvez-le!).

Exercice Comment utiliser les résultats précédents pour simuler une suite de v.a. indépendante suivant une loi normale centrée réduite ?

Chapitre 3

Méthode de Monte Carlo

3.1 Introduction

La simulation de v.a indépendantes suivant une loi donnée (p.ex uniforme) a une application importante : la calcul approché d'intégrales de fonctions en dimension grande. Pour fixer les idées, soit $f : [0, 1]^d \rightarrow \mathbf{R}$ une fonction "régulière" définie sur le cube d -dimensionnel. On se propose de calculer l'intégrale $I(f) = \int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \cdots dx_d$. Quand d égale 1 et que f est continue, les sommes

$$\frac{S_n(f)}{n} = \sum_{k=0}^{n-1} \frac{1}{n} f\left(\frac{k}{n}\right) \quad (3.1)$$

constituent une approximation de f dont on peut contrôler la qualité de la façon suivante : Si $\delta(\cdot)$ est le module de continuité ¹ de f on a

$$|I(f) - \frac{S_n(f)}{n}| \leq \delta\left(\frac{1}{n}\right).$$

Si f est plus régulière, on peut trouver des procédés plus efficaces. Par exemple, la méthode des trapèzes consiste à approcher le graphe de f par une fonction continue affine par morceaux : On pose

$$\frac{S_n}{n} = \frac{1}{2n} \left((f_0 + f_1) + (f_1 + f_2) + \cdots + (f_{n-1} + f_n) \right), \quad f_i = f(i/n)$$

Si f est C^2

$$|I(f) - \frac{S_n(f)}{n}| \leq \frac{C}{n^2}.$$

¹il est défini de la façon suivante : $\delta(\epsilon) = \sup_{|x-y| \leq \epsilon} (|f(x) - f(y)|)$. En particulier, $|f(x) - f(y)| \leq \delta(|x - y|)$. Quand f est C^1 , $\delta(\epsilon) \leq (\max_{[0,1]} |f'|) \cdot \epsilon$

La méthode de Simpson consiste à approcher f par une fonction continue et polynomiale de degré 2 par morceaux ; si n est pair on pose

$$\frac{S_n}{n} = \frac{1}{3n} \left((f_0 + 4f_1 + f_2) + (f_2 + 4f_3 + f_4) + \cdots + (f_{n-2} + 4f_{n-1} + f_n) \right), \quad f_i = f(i/n);$$

si f est de classe C^4

$$|I(f) - \frac{S_n(f)}{n}| \leq \frac{C}{n^4}.$$

En dimension supérieure ($d \geq 2$) la généralisation de (3.1) est

$$\frac{S_n(f)}{n} = \frac{1}{n^d} \sum_{k_1=0}^{n-1} \cdots \sum_{k_d=0}^{n-1} f\left(\frac{k_1}{n}, \dots, \frac{k_d}{n}\right)$$

et l'erreur que l'on commet dans l'évaluation de $I(f)$ est encore de la forme (si f est C^1)

$$|I(f) - \frac{S_n(f)}{n}| \leq C \frac{1}{n}.$$

Ainsi, pour obtenir une erreur de l'ordre de $C/n \approx \epsilon$ il faut calculer les valeurs de f en $n^d \approx (1/\epsilon)^d$ points. Quand d est grand (par exemple de l'ordre de 100) le temps de calcul est vite rédhibitoire. En outre, les méthodes que nous venons de décrire nécessitent de travailler avec une fonction f suffisamment différentiable.

Il est en fait possible de surmonter ces deux difficultés à condition d'accepter de travailler avec un algorithme probabiliste, la méthode de Monte Carlo.

3.2 Description de la méthode

Le principe repose sur la loi des grands nombres et le Théorème *Central Limit*. Si f est une fonction L^1 sur le pavé d -dimensionnel $[0, 1]^d$ et si X_1, \dots, X_n, \dots est une suite de *vecteurs* aléatoires uniformément distribués sur $[0, 1]^d$ alors la suite $Y_1 = f(X_1), \dots, Y_n = f(X_n), \dots$ est une suite de variables aléatoires indépendantes et de même loi ($\mathbf{P}(f(X_i) \in A) = \mathbf{P}(f(X_1) \in A)$ pour tout intervalle ou borélien A de \mathbf{R}). La loi des grands nombres nous enseigne que pour \mathbf{P} -presque tout $\omega \in \Omega$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left(f(X_1(\omega)) + \cdots + f(X_n(\omega)) \right) = \mathbf{E}(f(X_1)).$$

Calculons $\mathbf{E}(f(X_1))$. D'après la formule de transfert

$$\mathbf{E}(f(X_1)) = \int_{\mathbf{R}^d} f(x_1, \dots, x_d) \rho(x_1, \dots, x_d) dx_1 \cdots dx_d$$

où $\rho(x_1, \dots, x_d) = \mathbf{1}_{[0,1]^d}(x_1, \dots, x_d)$ est la densité de la loi uniforme sur $[0, 1]^d$. On a donc

$$\mathbf{E}(f(X_1)) = \int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \cdots dx_d = I(f)$$

et par conséquent \mathbf{P} -presque sûrement

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left(f(X_1(\omega)) + \cdots + f(X_n(\omega)) \right) = \int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

On est donc sûr qu'avec probabilité 1 la moyenne précédente

$$\frac{1}{n} S_n(f) := \frac{1}{n} \left(f(X_1(\omega)) + \cdots + f(X_n(\omega)) \right)$$

converge vers l'intégrale

$$I(f) = \int_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

La méthode du calcul approché de l'intégrale consiste donc à calculer $I(f)$ comme $\frac{1}{n} \left(f(X_1(\omega)) + \cdots + f(X_n(\omega)) \right)$.

Il est important de savoir à quelle vitesse la convergence précédente a lieu si $n \rightarrow \infty$. Supposons f de carré intégrable. Le théorème Central Limit nous fournit la réponse suivante : Puisque $Y_1 = f(X_1), \dots, Y_n = f(X_n), \dots$ est une suite indépendante de v.a.r de même loi et de carré intégrable on sait que

$$\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \left(f(X_1) + \cdots + f(X_n) \right) - \mathbf{E}(f(X_1)) \right), \quad \sigma^2 = \mathbf{Var}(f(X_1))$$

converge en loi vers une loi normale centrée réduite. En d'autres termes, si on pose

$$\begin{aligned} \sigma^2 &= \mathbf{E}(f(X_1)^2) - (\mathbf{E}(f(X_1)))^2 \\ &= I(f^2) - I(f)^2 \end{aligned}$$

on a pour tout $a > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{1}{n} S_n(f) - I(f) \right| < \frac{\sigma a}{\sqrt{n}} \right) = \int_{-a}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Observons que

$$\int_{-a}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0.95, \quad \text{pour } a = 1.96 \approx 2$$

et

$$\int_{-a}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 0.99, \quad \text{pour } a = 2.6$$

En conclusion, si on sait évaluer (même grossièrement) σ on peut dire avec probabilité grande que $S_n(f)/n$ est une approximation de $I(f)$ à $\sigma a/\sqrt{n}$ -près ($a = 2$ ou $a = 2.6$) pourvu que n soit assez grand. Par exemple, si ϵ est assez petit, $n = (2\sigma/\epsilon)^2$, $S_n(f)/n$ est une approximation de $I(f)$ à ϵ -près avec probabilité de 0.95. Remarquons que le calcul de $S_n(f)$ nécessite le calcul de $(2\sigma/\epsilon)^2$ valeurs de f et que ce nombre est *indépendant* de la dimension d de l'espace sur lequel on travaille. C'est un avantage considérable par rapport aux méthodes décrites dans la première section. En revanche, la faiblesse de la méthode réside dans le fait que :

- 1) le TCL n'est vrai qu'asymptotiquement : le "assez petit" (pour ϵ) ou "assez grand" (pour n) n'est *a priori* pas explicite ;
- 2) la méthode nécessite d'avoir une estimée raisonnable sur σ et suppose donc que l'on sache déjà calculer $I(f)$ et $I(f^2)$.

Ces deux problèmes admettent chacun une solution au moins d'un point de vue théorique. Leur solution donnée ci-dessous **ne sera pas demandée à l'examen**. C'est un complément de cours optionnel. La solution théorique du premier problème réside dans le résultat suivant :

Théorème 3.2.1 (Berry-Essen) *Soit X_1, \dots, X_n, \dots une suite de v.a. i.i.d. centrées ($\mathbf{E}(X_n) = 0$) telles que $\mathbf{E}(X_1^2) = \sigma^2$ et $\rho = \mathbf{E}(|X_1|^3) < \infty$. Si on note F_n la fonction de répartition de $\frac{S_n}{\sigma\sqrt{n}}$ on a pour tout $x \in \mathbf{R}$*

$$\left| F_n(x) - \int_{-\infty}^x \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right| \leq \frac{3\rho}{\sigma^3\sqrt{n}}.$$

On a donc

Théorème 3.2.2 *Pour tout a*

$$\left| \mathbf{P}\left(\left|\frac{1}{n}S_n(f) - I(f)\right| < \frac{\sigma a}{\sqrt{n}}\right) - \int_{-a}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right| \leq 6 \left(\frac{|f|_{C^0}}{\sigma}\right)^3 \frac{1}{\sqrt{n}}.$$

Pour résoudre le second problème il est naturel de remplacer σ^2 par

$$\Sigma_n^2 = \frac{1}{n}S_n(f^2) - \left(\frac{1}{n}S_n(f)\right)^2.$$

Théorème 3.2.3

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\left| \frac{1}{n} S_n(f) - I(f) \right| < \frac{\Sigma_n a}{\sqrt{n}} \right) = \int_{-a}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Démonstration. — Notons

$$Z_n = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - I(f) \right), \quad \tilde{Z}_n = \frac{\sqrt{n}}{\Sigma_n} \left(\frac{S_n}{n} - I(f) \right), \quad Y_n = \frac{\sigma}{\Sigma_n}$$

de façon que $\tilde{Z}_n = Y_n \cdot Z_n$. Puisque Σ_n converge p.s vers σ , il suffit de démontrer que si Z_n converge en loi vers Z et si Y_n converge p.s vers 1 alors $Z_n \cdot Y_n$ converge en loi vers Z . Démontrons donc le lemme suivant :

Lemme 3.2.1 *Si U_n converge en loi vers U et si V_n/U_n converge en loi vers 1 alors V_n converge en loi vers U*

Démonstration. — Comme

$$(V_n > t) \subset (U_n > \frac{t}{1+\epsilon}) \cup (\frac{V_n}{U_n} > 1+\epsilon), \quad (U_n > t(1+\epsilon)) \subset (V_n > t) \cup (\frac{U_n}{V_n} > 1+\epsilon)$$

on a

$$F_{U_n}(\frac{t}{1+\epsilon}) - \mathbf{P}(\frac{V_n}{U_n} > 1+\epsilon) \leq F_{V_n}(t), \quad F_{V_n}(t) \leq F_{U_n}(t(1+\epsilon)) + \mathbf{P}(\frac{U_n}{V_n} > 1+\epsilon)$$

et donc puisque $\frac{U_n}{V_n}$ et $\frac{V_n}{U_n}$ converge en loi vers 1² et que U_n converge en loi vers U on a pour tout t et tout $\epsilon > 0$ tels que $t/(1+\epsilon)$, $t(1+\epsilon)$ soient points de continuité de F_U

$$F_U(\frac{t}{1+\epsilon}) \leq \liminf_{n \rightarrow \infty} F_{V_n}(t) \leq \limsup_{n \rightarrow \infty} F_{V_n}(t) \leq F_U(t(1+\epsilon)).$$

Soit t un point de continuité de F_U ; il est possible de trouver ϵ aussi petit qu'on veut de façon que $t(1+\epsilon)^{\pm 1}$ soient points de continuité de F_U . Faisant $\epsilon \rightarrow 0$ sur ces ϵ on voit que $\lim_{n \rightarrow \infty} F_{V_n}(t)$ existe et vaut $F_U(t)$ pour tout t point de continuité de F_U : c'est la définition de la convergence en loi de V_n vers U .

□

Le théorème résulte alors du lemme et du fait que si la suite (Y_n) converge presque sûrement vers 1 alors elle converge aussi en loi vers 1.

□

²Si X_n converge en loi vers 1 alors X_n^{-1} converge aussi en loi vers 1 : en effet pour tout $t > 0$, $\mathbf{P}(X_n > t) = \mathbf{P}(0 < X_n^{-1} < t^{-1})$, c'est-à-dire $1 - F_{X_n}(t) = F_{X_n^{-1}}(1/t) - \mathbf{P}(X_n^{-1} = 1/t)$. Donc pour $s = 1/t > 0$ en dehors d'un ensemble dénombrable $F_{X_n^{-1}}(s)$ converge quand $n \rightarrow \infty$ vers $1 - \mathbf{1}_{s^{-1} > 1}(s) = \mathbf{1}_{s \geq 1}(s)$ et le même résultat est vrai pour $s < 0$ (on trouve 0). Il est facile de voir que cela signifie que la fonction de répartition de X_n^{-1} converge vers la fonction de répartition $\mathbf{1}_{t > 1}$ de la v.a 1 en tout point $t \neq 0$.

3.3 Variante

Si l'on désire à présent calculer des intégrales sur \mathbf{R}^d (et non plus sur des pavés compacts) on peut plus généralement procéder de la façon suivante.

Soit X_1, \dots, X_n, \dots une suite indépendante de vecteurs aléatoires de même loi donnée par une densité ρ qui est facile à simuler. On veut calculer $I(f) = \int_{\mathbf{R}^d} f(x) dx$

$$\begin{aligned} I(f) &= \int_{\mathbf{R}^d} f(x_1, \dots, x_d) dx_1 \cdots dx_d \\ &= \int_{\mathbf{R}^d} \frac{f(x_1, \dots, x_d)}{\rho(x_1, \dots, x_d)} \rho(x_1, \dots, x_d) dx_1 \cdots dx_d \\ &= \mathbf{E}(\phi(X)) \end{aligned}$$

où

$$\phi = \frac{f}{\rho}.$$

On supposera ρ non nulle ou plus généralement que son support contient celui de f . La loi des grands nombres nous dit que

$$\frac{S_n(\phi)}{n} = \frac{1}{n} \left(\phi(X_1) + \cdots + \phi(X_n) \right)$$

converge presque-sûrement vers $\mathbf{E}(\phi(X_1)) = I(f)$. En procédant comme dans la section précédente on peut démontrer

Théorème 3.3.1 *Si on pose*

$$\sigma_n^2 = \frac{1}{n} S_n(\phi^2) - \left(\frac{1}{n} S_n(\phi) \right)^2,$$

alors,

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(I(f) \in \left[\frac{1}{n} S_n(\phi) - \frac{\sigma_n a}{\sqrt{n}}, \frac{1}{n} S_n(\phi) + \frac{\sigma_n a}{\sqrt{n}} \right] \right) = \int_{-a}^a \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Le choix de la densité ρ est assez arbitraire mais il est naturel de vouloir minimiser la variance

$$\sigma^2 = \text{Var}(\phi(X)) = \int \frac{f^2}{\rho} dx - I(f)^2$$

la contrainte étant $\int \rho(x) dx = 1$. On pourrait utiliser la méthode des multiplicateurs de Lagrange, dont on ne donnera pas ici le détail.

Chapitre 4

Rudiments de Statistiques

4.1 Sondages

Un sondeur veut déterminer dans une population de N individus (par exemple $N = 6.10^7$) le nombre de personnes appartenant à une catégorie A (donc connaître le cardinal de l'ensemble A). Pour cela il effectue un sondage sur un échantillon de n individus (p.ex $n = 10^3$) tirés au hasard. Pour simplifier on supposera que le sondeur procède de la façon suivante. Il réalise n tirages X_1, \dots, X_n qui sont autant de variables aléatoires $X_i : \Omega \rightarrow \{1, \dots, N\}$ suivant une loi uniforme sur $\{1, \dots, N\}$. On pose alors $Y_i = \mathbf{1}_A \circ X_i$ et on définit $Z_n = Y_1 + \dots + Y_n$ le nombre de personnes interrogées qui appartiennent à la catégorie A . Il est facile de voir que les Y_i suivent une même loi de Bernoulli de paramètre p (où p est la proportion de personnes dans la population ayant l'opinion A) et sont indépendantes pourvu que les X_i le soient (**Exercice** : Prouver ces faits). Par conséquent, Z_n suit une loi de Binomiale (n, p) . Dans la pratique, on utilisera l'approximation normale que donne le Théorème Central Limit : la suite de v.a $(\sqrt{n}/\sigma)(Z_n - np)$, $\sigma^2 = p(1 - p)$, converge en loi vers une loi normale centrée réduite. On peut donc écrire

$$\mathbf{P}(\{\omega \in \Omega : p \in [\frac{Z_n(\omega)}{n} - \frac{c\sigma}{\sqrt{n}}, \frac{Z_n(\omega)}{n} + \frac{c\sigma}{\sqrt{n}}]\}) \approx \int_{-c}^c \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

[On rappelle que quand $c \approx 2$ cette intégrale est à peu près égale à 0.95 et quand $c \approx 2.6$ elle vaut à peu près 0.99.] Dans la pratique cette approximation est bonne quand par exemple $np \approx 30$. Pour simplifier nous supposons que nous sommes dans cette hypothèse asymptotique. Ainsi, si on connaît l'écart type σ on dira que l'intervalle $I_n = [\frac{Z_n}{n} - \frac{2\sigma}{\sqrt{n}}, \frac{Z_n}{n} + \frac{2\sigma}{\sqrt{n}}]$ (resp. $[\frac{Z_n}{n} - \frac{2.6\sigma}{\sqrt{n}}, \frac{Z_n}{n} + \frac{2.6\sigma}{\sqrt{n}}]$) est un *intervalle de confiance* avec une fiabilité de 95% (resp. 99%) pour la valeur de p : la probabilité pour que p se trouve dans l'intervalle ainsi

déterminé¹ est de 0.95 (resp. 0.99). Malheureusement, on ne connaît pas la valeur de σ (car sinon on connaîtrait déjà celle de p puisque $\sigma^2 = p(1-p)$). En revanche, on a toujours l'inégalité $\sigma = \sqrt{p(1-p)} \leq 1/2$ si bien que l'on peut écrire dès que n est assez grand

$$\mathbf{P}(p \in [\frac{Z_n}{n} - \frac{c}{2\sqrt{n}}, \frac{Z_n}{n} + \frac{c}{2\sqrt{n}}]) \geq \sim \int_{-c}^c \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

Un intervalle de confiance à 0.95 (resp. 0.99) est donc par exemple $[\frac{Z_n}{n} - \frac{1}{\sqrt{n}}, \frac{Z_n}{n} + \frac{1}{\sqrt{n}}]$ (resp. $[\frac{Z_n}{n} - \frac{1.3}{\sqrt{n}}, \frac{Z_n}{n} + \frac{1.3}{\sqrt{n}}]$)

4.2 Statistiques gaussiennes

Sondages gaussiens Dans l'exemple précédent, l'estimation *a priori* sur la variance des Y_i , rendue possible par le fait que les Y_i suivent une loi de Bernoulli, est l'élément clé pour obtenir un intervalle de confiance. Intéressons nous à présent au problème suivant. On sait que le diamètre d'un boulon fabriqué dans une certaine usine suit une loi normale de moyenne μ et de variance σ inconnues². On veut avoir une estimation de μ avec une certitude raisonnable (intervalle de confiance) en effectuant un nombre n de tirages (indépendants) ce que l'on modélise par n v.a Y_k , $1 \leq k \leq n$, i.i.d suivant une même loi normale $\mathcal{N}(\mu, \sigma^2)$. Puisque les Y_i sont indépendantes et gaussiennes leur somme $Z_n = Y_1 + \dots + Y_n$ est également gaussienne (**Exercice**³) et on a donc l'égalité⁴

$$\mathbf{P}(\mu \in [\frac{Z_n}{n} - \frac{c\sigma}{\sqrt{n}}, \frac{Z_n}{n} + \frac{c\sigma}{\sqrt{n}}]) = \int_{-c}^c \frac{e^{-t^2/2}}{\sqrt{2\pi}}.$$

Malheureusement on ne connaît pas σ et on ne peut donc pas préciser d'intervalle de confiance. Il faut procéder autrement. Un candidat naturel pour avoir une approximation de μ est la *moyenne empirique*

$$\bar{Y}_n = \frac{Y_1 + \dots + Y_n}{n}$$

et pour obtenir une approximation de σ^2 la variance empirique

$$V_n = \frac{(Y_1 - \bar{Y}_n)^2 + \dots + (Y_n - \bar{Y}_n)^2}{n}.$$

¹Attention, cela signifie que $p = \#A/N$ étant fixé (mais inconnu) la probabilité qu'un sondage portant sur n personnes fournisse un intervalle I_n ne contenant pas p est inférieure à 0.05

²cette loi dépend donc de deux paramètres

³On peut utiliser les fonctions caractéristiques

⁴et pas seulement une identité asymptotique

Pour formuler le théorème fondamental, on a besoin de deux définitions suivantes.

Définition. Soient X_1, \dots, X_n des v.a. indépendantes de loi Gaussienne d'espérance 0 et de variance 1. La loi de la v.a. $X_1^2 + \dots + X_n^2$ est dite *la loi de chi-deux de n degrés de liberté*. C'est aussi une loi gamma $\gamma_{1/2, n/2}$ de densité

$$\mathbf{1}_{]0, \infty[}(x) \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^x t^{(n/2)-1} e^{-t/2} dt.$$

Définition. Soient X et Y deux variables aléatoires indépendantes. La v.a. X est de loi Gaussienne d'espérance 0 et de variance 1. La v.a. Y est de loi chi-deux de $n - 1$ degrés de liberté. La loi de la v.a.

$$\frac{X\sqrt{n-1}}{\sqrt{Y}}$$

est dite la loi de Student de $n - 1$ degrés de liberté. C'est la loi de densité

$$c_{n-1} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}}, \quad c_{n-1} = \frac{\Gamma(\frac{n}{2})}{\sqrt{(n-1)\pi} \Gamma(\frac{n-1}{2})}$$

Cette densité tend vers la densité gaussienne réduite $(1/\sqrt{2\pi})e^{-t^2/2}$ quand n tend vers l'infini; on a les approximations suivantes

$P(T_n \leq a)$	= 0.95	0.99
si $n = 10$	pour $a = 2.26$	pour $a = 3.35$
20	2.09	2.86
30	2.04	2.76
∞	1.96	2.58

Théorème 4.2.1 Si les v.a Y_1, \dots, Y_n, \dots sont i.i.d. de même loi normale $\mathcal{N}(\mu, \sigma^2)$ alors

(a) la v.a

$$\frac{\sqrt{n}}{\sigma}(\bar{Y}_n - \mu)$$

suit une loi normale centrée $\frac{1}{2}e$ réduite $\mathcal{N}(0, 1)$.

(b) Les v.a. \bar{Y}_n et V_n sont indépendantes et la v.a.

$$Z_n = \frac{\bar{Y}_n - \mu}{\sqrt{V_n/(n-1)}}$$

suit une loi de Student $\mathcal{T}(n-1)$ à $n - 1$ degrés de liberté;

(c) la v.a

$$\frac{nV_n}{\sigma^2}$$

suit la loi du chi-deux é $n - 1$ degrés de liberté $\chi^2(n - 1)$.

Démonstration. (a) est évident.

Notons que $Z_n = (\sqrt{n-1}\sqrt{n}(\bar{Y}_n - \mu)/\sigma)(\sqrt{nV_n/\sigma^2})^{-1}$. Il suffit de prouver que $S = \frac{\sqrt{n}}{\sigma}(\bar{Y}_n - \mu)$ et $T = \frac{nV_n}{\sigma^2}$ sont indépendantes, et que T soit de loi $\chi^2(n-1)$. Soient $W_i = (Y_i - \mu)/\sigma$, $i = 1, \dots, n$. Les v.a. W_1, \dots, W_n sont indép., de loi Gaussienne d'espérance 0 et de variance 1. Soit $\vec{W} = (W_1, \dots, W_n)$. Prenons A une matrice orthogonale dont la première ligne se compose de \sqrt{n} . Alors le vecteur $A\vec{W}$ est aussi Gaussien, d'espérance $(0, \dots, 0)$ et de matrice de covariances $AIdA^T = AA^T = Id$ car A est orthogonale. Donc les coordonnées de $A\vec{W}$ sont des Gaussiennes indépendantes d'espérance 0 et de variance 1. Notamment $\sum_{i=2}^n ((A\vec{W})_i)^2$ est de loi de $\chi^2(n-1)$. Par le choix de la première ligne de A on a $(A\vec{W})_1 = S$. Alors S est indépendante de $(A\vec{W})_i$ pour $i = 2, \dots, n$, et donc de $\sum_{i=2}^n ((A\vec{W})_i)^2$ qui est de loi de $\chi^2(n-1)$. Mais une matrice orthogonale conserve les normes de vecteurs. Donc $S^2 + \sum_{i=2}^n ((A\vec{W})_i)^2 = \sum_{i=1}^n W_i^2$. Alors $\sum_{i=2}^n ((A\vec{W})_i)^2 = \sum_{i=1}^n W_i^2 - S^2$ ce qui est par un calcul direct $\frac{nV_n}{\sigma^2}$.

Dans l'exemple qui nous intéresse on peut donc,

(i) sans connaître la valeur de σ , obtenir un intervalle de confiance pour la moyenne μ des diamètres des boulons : la probabilité pour que l'intervalle (aléatoire)

$$[\bar{Y}_n - a\sqrt{\frac{V_n}{n-1}}, \bar{Y}_n + a\sqrt{\frac{V_n}{n-1}}]$$

contienne le réel μ (inconnu) est égale à $\mathbf{P}(|Z_n| \leq a)$ (on a utilisé le (b) du théorème précédent) ;

(ii) comparer l'écart-type empirique $\sqrt{V_n}$ et l'écart-type réel σ . Par exemple, si $n = 10$, la probabilité que $10V_{10} > 16.9\sigma^2$ est à peu près inférieure à 0.05 car si $F_{\chi^2(9)}(\cdot)$ est la fonction de répartition d'un chi-deux à 9 degrés de liberté on a $1 - F_{\chi^2(9)}(16.9) \approx 0.05$. On a utilisé le (c) du théorème précédent.

Tests statistiques Il est commode à ce stade d'introduire la notion de *test statistique*. Le type de problème que l'on se propose d'étudier est le suivant : à partir d'un échantillon, c'est-à-dire une suite de v.a X_1, \dots, X_n de même loi p qui en général est inconnue, décider si une hypothèse \mathcal{H}_0 (portant sur la loi de ces v.a) est raisonnable ou pas. Si elle ne l'est pas on la rejette, sinon on l'accepte, ce qui alors signifie seulement qu'elle est plausible. Dans ce type

de discussion on introduit un *seuil* α qui est un réel compris entre 0 et 1 et quantifie le degré de certitude que l'on a lors de la procédure de rejet ou non de l'hypothèse \mathcal{H}_0 . La procédure est la suivante : on considère une *statistique* c'est-à-dire une v.a Z_{n,\mathcal{H}_0} qui est une fonction des v.a X_1, \dots, X_n et dont on connaît la loi (et donc la fonction de répartition) *pourvu que* l'hypothèse \mathcal{H}_0 soit vérifiée. On décide de rejeter ou d'accepter l'hypothèse au seuil $\alpha \in (0, 1)$ si par exemple, les données expérimentales donnent $|Z_{n,\mathcal{H}_0}| \geq t_\alpha$ où t_α est définie par $\mathbf{P}_{\mathcal{H}_0}(|Z_{n,\mathcal{H}_0}| \geq t_\alpha) = \alpha$. En d'autres termes on rejette l'hypothèse \mathcal{H}_0 si la réalisation $\omega \in \Omega$ qui donne $|Z_{n,\mathcal{H}_0}(\omega)| \geq t_\alpha$ fait partie d'un événement très peu probable.

Dans l'exemple (ii) du paragraphe précédent l'échantillon est une suite de v.a de loi $\mathcal{N}(\mu, \sigma^2)$ qui dépend des deux paramètres μ , et σ *a priori* inconnus et on veut tester l'hypothèse suivante portant sur σ : $\mathcal{H}_0 : 10V_{10} \geq 20\sigma^2$. Pour cela on construit une statistique, en l'occurrence la v.a $Z_{10} = \frac{10V_{10}}{\sigma^2}$ dont on sait qu'elle suit une loi du chi-deux à 9 degrés de liberté et pour laquelle la fonction de répartition est tabulée. Au seuil $\alpha = 0.05$ on rejette l'hypothèse car

$$\mathbf{P}(Z_{10} \geq 20) \leq \mathbf{P}(Z_{10} \geq t_\alpha) = \mathbf{P}(\chi_9^2 \geq t_\alpha)$$

où $t_\alpha = 16.9$ et $\mathbf{P}(\chi_9^2 \geq t_\alpha) = 1 - F_{\chi^2(9)}(16.9) \approx 0.05$ est plus petite que le seuil $\alpha = 0.05$ que l'on s'était fixé.

Tests de Fisher et de Student. Ces deux tests ne seront pas demandés à l'examen. On considère deux échantillons gaussiens X_1, \dots, X_{n_x} , Y_1, \dots, Y_{n_y} qui sont indépendants et suivent respectivement des lois $\mathcal{N}(\mu_x, \sigma_x^2)$ et $\mathcal{N}(\mu_y, \sigma_y^2)$. On note

$$\bar{X} = \frac{X_1 + \dots + X_{n_x}}{n_x}, \quad \bar{Y} = \frac{Y_1 + \dots + Y_{n_y}}{n_y}$$

les moyennes empiriques et

$$V_x = \frac{(X_1 - \bar{X})^2 + \dots + (X_{n_x} - \bar{X})^2}{n_x}, \quad V_y = \frac{(Y_1 - \bar{Y})^2 + \dots + (Y_{n_y} - \bar{Y})^2}{n_y}$$

les variances empiriques. Le théorème suivant permet de tester l'hypothèse $\mathcal{H}_0 : \sigma_x^2 = \sigma_y^2$.

Théorème 4.2.2 *Le rapport*

$$\frac{\frac{n_x}{n_x - 1} \frac{V_x}{\sigma_x^2}}{\frac{n_y}{n_y - 1} \frac{V_y}{\sigma_y^2}}$$

suit une loi de Fisher $\mathcal{F}(n_x-1, n_y-1)$. La densité d'une loi de Fisher $\mathcal{F}(m, n)$ est

$$f_{m,n}(t) = C_{m,n} t^{(m/2)-1} (n+tm)^{-(m+n)/2} \mathbf{1}_{t>0}, \quad C_{m,n} = \frac{m^{m/2} n^{n/2} \Gamma((m+n)/2)}{\Gamma(m/2) \Gamma(n/2)}.$$

Remarque : $n_x V_x / \sigma_x^2$ et $n_y V_y / \sigma_y^2$ suivent (d'après le Théorème 4.2.1) des lois de Student.

La statistique du test de Fisher $\mathcal{H}_0 : \sigma_x^2 = \sigma_y^2$ est donc

$$\frac{\frac{n_x}{n_x-1} V_x}{\frac{n_y}{n_y-1} V_y}$$

Sous l'hypothèse d'égalité des variances $\sigma_x^2 = \sigma_y^2$ on peut tester l'égalité des moyennes $\mathcal{H}_0 : \mu_x = \mu_y$

Théorème 4.2.3 Si $\sigma_x = \sigma_y$ la v.a

$$\frac{\sqrt{n_x + n_y - 2} (\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{n_x^{-1} + n_y^{-1}} \sqrt{n_x V_x + n_y V_y}}$$

suit la loi de Student $\mathcal{T}(n_x + n_y - 2)$.

4.3 Test du chi-deux

Ce test est souvent demandé à l'examen !

Déterminer si un dé est pipé ou non : On veut savoir si un dé à 6 faces présente chacune de ses faces de façon équiprobable. Pour cela on jette le dé n fois (n grand). Le résultat du k -ème lancer est modélisé par une v.a X_k à valeurs dans $E = \{1, \dots, 6\}$ et on suppose que les v.a X_k sont indépendantes et identiquement distribuées. Dire que le dé est pipé c'est dire que $(p_1, p_2, p_3, p_4, p_5, p_6) \neq (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$ (on a noté $p_i = \mathbf{P}(X_k = i)$). On veut donc tester l'hypothèse \mathcal{H}_0 : "la loi $(p_i)_{i \in E}$ est uniforme". Pour fixer les idées supposons qu'on réalise 600 lancers et qu'on obtienne 60 fois le 1, 108 fois le 2, 108 fois le 3, 102 fois le 4, 108 fois le 5 et 114 fois le 6. Le dé est-il pipé ou non ? Pour répondre à cette question, on utilise le théorème suivant :

Théorème 4.3.1 Soit X_1, \dots, X_n, \dots une suite de v.a. i.i.d. à valeurs dans un ensemble fini $E = \{1, \dots, r\}$. Nous noterons (p_i) leur loi commune avec $p_i = \mathbf{P}(X_n = i)$. Posons

$$N_i^{(n)} = \sum_{k=1}^n \mathbf{1}_i \circ X_k,$$

la fréquence empirique de sortie de i . Alors,

(a) le vecteur aléatoire

$$\vec{Z}_n = \left(\frac{N_1^{(n)} - np_1}{\sqrt{np_1}}, \dots, \frac{N_r^{(n)} - np_r}{\sqrt{np_r}} \right)$$

converge en loi vers un vecteur gaussien \vec{Z} d'espérance $\vec{0}$ et de matrice de covariances B , dont les éléments $b_{i,i} = (1 - p_i)$ pour $i = 1, \dots, r$, $b_{i,j} = -\sqrt{p_i p_j}$ pour $i \neq j$.

(b) la suite de v.a

$$T_n = \frac{(N_1^{(n)} - np_1)^2}{np_1} + \dots + \frac{(N_r^{(n)} - np_r)^2}{np_r},$$

converge en loi quand $n \rightarrow \infty$ vers un χ^2 à $r - 1$ degrés de liberté.

Démonstration. (a) est un corollaire direct du Thm Limite Central vectoriel appliqué aux sommes de vecteurs aléatoires indépendants $(\mathbf{1}_1 \circ X_k, \dots, \mathbf{1}_r \circ X_k)$ pour $k = 1, 2, \dots$. En effet, $E\mathbf{1}_i \circ X_k = p_i$ pour $i = 1, \dots, r$. Calculons $\text{cov}(\mathbf{1}_i \circ X_k / \sqrt{p_i}, \mathbf{1}_j \circ X_k / \sqrt{p_j})$. Pour $i = j$ c'est $\text{var}(\mathbf{1}_i \circ X_k / \sqrt{p_i}) = p_i(1 - p_i)/p_i = 1 - p_i$. Pour $i \neq j$ on a $(\mathbf{1}_i \circ X_k) \cdot (\mathbf{1}_j \circ X_k) = 0$ car X_k ne peut pas prendre les valeurs i et j en même temps. Donc c'est $-E(\mathbf{1}_i \circ X_k / \sqrt{p_i})E(\mathbf{1}_j \circ X_k / \sqrt{p_j}) = -\sqrt{p_i p_j}$.

Comme \vec{Z}_n converge en loi vers \vec{Z} , alors pour toute fonction g continue bornée $Eg(\vec{Z}_n) \rightarrow Eg(\vec{Z})$. En particulier pour toute f continue bornée $Ef(\|\vec{Z}_n\|) \rightarrow Ef(\|\vec{Z}\|)$ car $g = f \circ \|\cdot\|$ est continue bornée. Il s'en suit que $\|\vec{Z}_n\| \rightarrow \|\vec{Z}\|$ en loi.

Il reste à prouver que la loi de $\|\vec{Z}\|$ est la loi de $\chi^2(r-1)$. La matrice B est symétrique définie positive. Elle est donc diagonalisable en base orthonormée. Il existe une matrice A orthogonale telle que ABA^T est une matrice diagonale D qui se compose de valeurs propres de B . La combinaison linéaire des lignes de B avec les coefficients $\sqrt{p_1}, \dots, \sqrt{p_r}$ est 0. Donc une valeur propre est 0. La matrice $B - Id$ a toutes les lignes proportionnelles. Donc 1 est une valeur propre de B de multiplicité $r - 1$. Donc D a 1 sur la diagonale $r - 1$ fois et 0 pour le

dernier élément. Le vecteur $A\vec{Z}$ est Gaussien d'espérance $\vec{0}$ et de matrice de covariances $ABA^T = D$. Donc la loi de $\|A\vec{Z}\|$ est la loi $\chi^2(r-1)$. Mais comme A est orthogonale, elle conserve la norme, par conséquent $\|\vec{Z}\| = \|\vec{A}\vec{Z}\|$ est donc de loi $\chi^2(r-1)$. Le Thm est démontré.

Ainsi, si l'hypothèse \mathcal{H}_0 est vérifiée la v.a

$$T_{600} = \frac{(N_1^{600} - 100)^2}{100} + \dots + \frac{(N_6^{600} - 100)^2}{100},$$

doit suivre (approximativement) une loi du chi-deux à 5 ($= 6 - 1$) degrés de liberté. Or, pour une telle loi $\mathbf{P}(\mathcal{T}_5 \geq 0.412) \leq 0.005$. Dans notre expérience on a obtenu

$$T_{600}(\omega) = \frac{(60 - 100)^2}{100} + \dots + \frac{(114 - 100)^2}{100} = 3.92$$

Comme l'événement $3.92 > 0.412$ on décide au seuil $\alpha = 0.005$ de rejeter l'hypothèse \mathcal{H}_0 (le fait d'observer $\mathcal{T}_5 \geq 0.412$ est un événement rare).

De manière générale, supposons, on sait que $X_1, X_2, \dots, X_k, \dots$ sont des v.a. indépendantes et de même loi discrète prenant r valeurs a_1, \dots, a_r . On veut tester l'hypothèse que ces valeurs sont prises avec probabilités p_1, \dots, p_r . On observe la réalisation X_1, X_2, \dots, X_n pour n grand. On compose $N_i^n = \sum_{k=1}^n \mathbf{1}_{a_i} \circ X_k$ pour $i = 1, \dots, r$. On calcule la valeur de $T_n = \frac{(N_1^{(n)} - np_1)^2}{np_1} + \dots + \frac{(N_r^{(n)} - np_r)^2}{np_r}$. Sa loi est approximativement (!) de $\chi^2(r-1)$.

On se donne un seuil de confiance α et on trouve le quantile de la loi $\chi^2(r-1)$, cad $P(\chi > \chi_{\alpha, r-1}) = \alpha$. Ainsi $P(T_n > \chi_{\alpha, r-1}) \approx \alpha$.

S'il se trouve de notre calcul que $T_n > \chi_{\alpha, r-1}$, alors un événement de probabilité trop petite est réalisé, cet événement est jugé trop peu probable pour avoir lieu en réalité, par conséquent on se doute de notre hypothèse et on la rejete.

Si $T_n < \chi_{\alpha, r-1}$, ce test ne rejete pas notre hypothèse.