CONVERGENCE DES SUITES DE VARIABLES ALÉATOIRES

THÉORÈME CENTRAL LIMITE:

École Supérieure d'Informatique, SBA
2ième année CP

22 Février 2018

Plan de cours

INTRODUCTION

- Ce cours est consacré aux deux résultats fondamentaux des probabilités: la loi faible des grands nombres et le théorème central limite qui précise la vitesse de cette convergence.
- L'énoncé de ces résultats nécessite l'introduction de plusieurs modes de convergence pour les variables aléatoires: la convergence en probabilité et la convergence en loi.
- La loi faible des grands nombres est énoncée. On donne des applications de ce résultat.
- Le théorème central limite est énoncé. D'autres résultats qui complètent le théorème central limite sont présentés.

- Popur extraire un certain nombre de caractéristique (moyenne, écart-type, ...) d'une suite d'observation, on s'intéresse au comportement des ces observations lorsquela taille de l'échantillon grandit.
- Les variables aléatoires X_n, X utilisées dans ce cours sont toutes définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \Pr)$, à valeurs réelles, de fonctions de répartition respectives F_n et F.

INÉGALITÉ DE MARKOV

- Soient X une variable aléatoire réelle et ϕ une application borélienne (continu) de $\mathbb{R}$ dans $[0, \infty]$,
- Alors, pour tout réel $a > 0$:

$$\Pr[\phi(X) \geq a] \leq \frac{\mathbb{E}[\phi(X)]}{a} .$$

- En particulier, si X est une v.a.r. positive d'espérance finie, alors pour tout réel $a > 0$:

$$\Pr[X \geq a] \leq \frac{\mathbb{E}(X)}{a} .$$

Inégalité de Markov: démonstration

- Posons $A = \{\phi(X) \geq a\}$.
- Puisque ϕ est positive, on peut écrire:

$$\begin{aligned} \mathbb{E}[\phi(X)] &= \mathbb{E}[\phi(X)\mathbb{1}_A] + \mathbb{E}[\phi(X)\mathbb{1}_{A^c}] \\ &\geq \mathbb{E}[\phi(X)\mathbb{1}_A] \geq a\mathbb{E}[\mathbb{1}_A] = a\,\Pr(A) \end{aligned} .$$

Inégalité de Markov: Corollaire

- Soit $m \geq 1$ et Soit X une variable aléatoire réelle telle que $\mathbb{E}(|X|^m) < \infty$;
- Alors, pour tout réel $a > 0$:

$$\Pr[|X| \geq a] \leq \frac{\mathbb{E}(|X|^m)}{a^m}.$$

INÉGALITÉ DE BIENAYMÉ-TCHEBYSCHEV

- Pour toute variable aléatoire réelle X de carré intégrable $(\mathbb{E}(X^2) < \infty)$ et pour tout réel quelconque k strictement positif, on a:

$$\Pr(|X - \mathbb{E}(X)| > k\sigma_X) \leq \frac{1}{k^2}.$$

- Autres formes de cette inégalité:

$$\Pr(|X - \mathbb{E}(X)| < k\sigma_X) \geq 1 - \frac{1}{k^2}.$$

$$\Pr(|X - \mathbb{E}(X)| > k) \leq \frac{\sigma_X^2}{k^2},$$

- Ces différentes formes permettent de comprendre la signification de l'écart-type.

- *L'écart-type caractérise la dispersion de la distribution autour de l'espérance mathématique*.
  - Pour $k = 10$, l'événement: $|X - \mathbb{E}(X)| > 10\sigma_X$ a peu de chances de se réaliser, en effet:

$$\Pr(|X - \mathbb{E}(X)| > 10\sigma_X) \leq \frac{1}{100}\,.$$

  - Supposons $\sigma = 0$, alors: $\Pr(|X - \mathbb{E}(X)| > k) = 0$. X est presque sûrement égale à $\mathbb{E}(X)$.

INÉGALITÉ DE BIENAYMÉ-TCHEBYSCHEV: REMARQUE

- L'inégalité de Bienaymé-Tchebyschev impose seulement l'existence des moments d'ordre 1 et 2. Comme elle ne fait pas intervenir la loi de probabilité suivie par la variable aléatoire considérée, elle peut s'appliquer à de nombreuses lois mais elle donne une majoration de la probabilité beaucoup trop grande.
- On peut comparer les majorations qu'elle donne avec celles obtenues en considérant la loi exacte suivie par la variable aléatoire.

INÉGALITÉ DE BIENAYMÉ-TCHEBYSCHEV: EXEMPLE 1

- Si $X \rightsquigarrow \mathcal{B}(10, \frac{1}{2})$,alors $\mathbb{E}(X) = 5$ et $\sigma_X^2 = \frac{5}{2}$.
- Un calcul direct donne l'estimation

$$\Pr(|X - 5| \geq 4) = \Pr(X \leq 1 \cup X \geq 9) \simeq 0.021\,,$$

- Alors que l'inégalité de Bienaymé-Tchebyschev donne la majoration:

$$\Pr(|X - 5| \geq 4) \leq \frac{2.5}{4^2} \simeq 0.156\,.$$

Inégalité de Bienaymé-Tchebyschev: exemple 2

- Un échantillon d'effectif n doit être extrait d'une distribution de moyenne m et d'écart-type σ.
- On cherche le plus petit entier n vérifiant la condition:

$$\Pr(|\overline{X} - m| < \frac{\sigma}{4}) \geq 0.99\,.$$

- On applique l'inégalité de Bienaymé-Tchebyschev à la variable $\overline{X}$ qui a pour moyenne m et pour variance σ^2/n:

$$\Pr(|\overline{X} - m| < \frac{k\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}\,.$$

- C'est-à-dire: $1 - \frac{1}{k^2} = 0.99$ et $\frac{k}{\sqrt{n}} = \frac{1}{4}$, donc $n = 1600$.

- En utilisant le théorème central limite (section suivente), on détermine une autre valeur pour n, meilleure que la première:
  - La variable aléatoire $\frac{\overline{X}-m}{\sigma/\sqrt{n}}$ suit une loi normale centrée réduite quand n tend vers l'infini. D'où:

$$\Pr\left(|\frac{\overline{X}-m}{\sigma/\sqrt{n}}| < 2.3263\right) \geq 0.99 \Leftrightarrow \frac{\sigma}{\sqrt{n}}2.3263 = \frac{\sigma}{4} \Rightarrow n \cong 86\,.$$

CONVERGENCE DES SUITES DE VARIABLES ALÉATOIRES: INTRODUCTION

- Grâce à ces théorèmes de convergence, le calcul des probabilités trouve sa justification dans l'étude des phénomènes mettant en jeu des populations ou des observations nombreuses (méthodes de sondage, théorie de l'estimation, des tests...).

Convergence des suites de variables aléatoires: introduction

- Les modes de convergence les plus utilisés en probabilité sont les suivants:
  - la convergence en probabilité notée p,
  - la convergence presque sûre notée $p.s.$,
  - la convergence en loi notée $\mathcal{L}$,
  - la convergence en moyenne quadratique notée $m.q$.
- La plus forte des convergences de suites de variables aléatoires est la convergence presque sûre. Ce concept nécessite d'avoir défini une variable aléatoire comme une application mesurable d'un espace probabilisé dans un autre.
- Ces différents modes de convergence ne sont pas indépendants et satisfont aux implications suivantes:
  $p.s. \rightarrow p \rightarrow \mathcal{L}$
  $m.q. \rightarrow p \rightarrow \mathcal{L}$

CONVERGENCE EN PROBABILITÉ

- Une suite de variables aléatoires $(X_n)_{n\geq 1}$, non nécessairement indépendantes, converge en probabilité vers la variable aléatoire X si et seulement si:

$$\forall \epsilon > 0 \; ; \lim_{n\to\infty} \Pr\left(|X_n - X)| > \epsilon\right) = 0 \, .$$

- On note $X_n \xrightarrow{\Pr} X$.
- La convergence en probabilité de la suite $(X_n)_{n\geq 1}$ vers la variable aléatoire X est la convergence de la suite de variables aléatoires $(X_n - X)$ vers 0:

$$\forall \varepsilon, \eta > 0 \, , \exists N(\varepsilon, \eta) \text{ tel que } n > N \Rightarrow \Pr\left(|X_n - X| > \epsilon\right) < \eta$$

CONVERGENCE EN PROBABILITÉ: EXEMPLE 1

- Considérons la suite de variables aléatoires (X_n) définie par:

$$\Pr(X_n = 0) = 1 - \frac{1}{n^2} \qquad \Pr(X_n = 1) = \frac{1}{n^2}\,.$$

- Montrer que (X_n) converge en probabilité vers 0 quand n tend vers l'infini.

CRITÈRE DE CONVERGENCE EN PROBABILITÉ

- Si $\mathbb{E}(X_n) \to a$, il suffit de montrer que $Var(X_n) \to 0$, pour établir la convergence en probabilité de la suite (X_n) vers a (application de l'inégalité de Bienayme-Tchebycheff).
- Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires; s'il existe $r > 0$ tel que la suite de terme général $\mathbb{E}[|X_n|^r]$ tende vers 0, alors la suite $(X_n)_{n\geq 1}$ converge en probabilité vers 0.
- En effet, on a, d'après l'inégalité de Markov (corollaire), pour tout $\varepsilon > 0$
$$\Pr(|X_n| \geq \varepsilon) \leq \frac{\mathbb{E}[|X_n|^r]}{\varepsilon^r} \to 0\,.$$

Critère de convergence en probabilité: Exemple

- Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. de Bernoulli de paramètre p_n tel que $\lim_{n\to\infty} p_n = 0$. Alors la suite converge en probabilité vers 0.
- En effet, pour $\varepsilon \in]0,1[$, on a:

$$\Pr(|X_n| > \varepsilon) = p_n \quad \text{et} \lim_{n\to\infty} p_n = 0\,.$$

Convergence en probabilité: Remarque

- La convergence en probabilité n'entraîne pas celle des moments; en particulier, elle n'entraîne pas celle des espérances mathématiques.
- Considérons la suite de variables aléatoires (X_n), chaque variable prenant deux valeurs 0 et α_n avec les probabilités suivantes:

$$\Pr(X_n = 0) = 1 - \frac{1}{n} \qquad \Pr(X_n = \alpha_n) = \frac{1}{n}\,.$$

- La suite (X_n) converge en probabilité vers 0 quand n tend vers l'infini.
- $\mathbb{E}(X_n) = \frac{\alpha_n}{n}$.
- Selon les valeurs de α_n, on obtient pour $\mathbb{E}(X_n)$ différentes limites, finies ou non, ou pas de limite:
  $\alpha_n = \sqrt{n}\,, \quad \mathbb{E}(X_n) \to 0$; $\alpha_n = n\,, \quad \mathbb{E}(X_n) = 1\,;$
  $\alpha_n = (-1)^n n\,, \quad \mathbb{E}(X_n) = (-1)^n\,; \alpha_n = n^r\,, \quad \mathbb{E}(X_n) \to \infty\,.$
- tandis que l'espérance mathématique de la limite de (X_n) est nulle puisque cette limite est égale à 0.

- Soient $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ deux suites de variables aléatoires qui convergent en probabilité vers les variables aléatoires X et Y respectivement, et soit $h : \mathbb{R}^2 \to \mathbb{R}$ une fonction continue en tout point (x, y) de $\mathbb{R}^2$. Alors, la suite aléatoire $h(X_n, Y_n)_{n\geq 1}$ converge en probabilité vers la variables aléatoire $h(X, Y)$.
- En particulier, si $X_n \xrightarrow{P} X$ et si f est une fonction réelle, continue en tout point de $\mathbb{R}$, alors $f \circ X_n \xrightarrow{P} f \circ X$.

Convergence en probabilité: Corollaire

- Soit $(X_n)_{n\geq 1}$ et $(Y_n)_{n\geq 1}$ deux suites de variables aléatoires qui convergent en probabilité vers les variables aléatoires X et Y respectivement, alors:
  - $X_n + Y_n \xrightarrow{\quad P \quad} X + Y$;
  - $\lambda X_n \xrightarrow{\quad P \quad} \lambda X$;
  - $X_n Y_n \xrightarrow{\quad P \quad} XY$;
  - $X_n / Y_n \xrightarrow{\quad P \quad} X / Y$.

CONVERGENCE EN LOI: DÉFINITION

- Souvent, en statistique, on ne connait pas les lois: on observe seulement une certaine distribution.
- Les théorèmes de convergence en loi permettent de justifier certaines approximations de distributions observées par des lois théoriques connues.
- On introduit un nouveau concept de convergence pour une suite de v.a., qui ne dépend cette fois que de la loi de chacune de ces v.a.

Convergence en loi: définition

- Une suite de variables aléatoires $(X_n)_{n\geq 1}$, non nécessairement indépendantes, de fonction de répartition F_n, converge en loi vers la variable aléatoire X, de fonction de répartition F, quand n tend vers l'infini; si et seulement si $\lim_{n\to\infty} F_n(x) = F(x)$ en tout point de continuité x de F.
- Cela signifie que, quand n est grand, la loi de probabilité de X_n est approximativement la loi de fonction de répartition F.

Convergence en loi: remarque

- La convergence est réalisée aux points de continuité de la fonction F. C'est la convergence ponctuelle de la suite des fonctions de répartition.
- En un point de discontinuité x_0 de F, différentes situations sont possibles, soit $F_n(x_0)$ n'a pas de limite ou a une limite différente de $F(x_0)$.

Convergence en loi: exemple

- Soit X_n la variable aléatoire suivant la loi normal $\mathcal{N}(0, \frac{1}{n})$.
- La suite des variables (X_n) converge en loi vers 0 quand $n \to \infty$.
- Soit F_n la fonction de répartition de X_n, c'est-à-dire la fonction définie par:

$$F_n(x) = \Pr(X_n < x)$$

- Quand $n \to \infty$, $F_n(x)$ a pour limite 0 si $x \leq 0$ et pour limite 1 si $x > 0$.
- La suite $F_n(x)$ converge donc, pour toutes les valeurs de x différentes de 0, vers la fonction de répartition $F(x)$, définie par:

$$F(x) = 0 \text{ si } x \leq 0 \;\text{ et }\; F(x) = 1 \;\text{ si }\; x > 0\,.$$

- Or $\forall n$, $F_n(0) = 0,5$ donc $F(0) = 0 \neq F_n(0)$.
- Au point de discontinuité 0, la limite de $F_n(x)$ est différente de $F(0)$.

Convergence en loi: Théorème

- Si X_n suit la loi binomiale $\mathcal{B}(n, \frac{\lambda}{n})$ et si X suit la loi de Poisson $\mathcal{P}(\lambda)$, alors $X_n \xrightarrow{\mathcal{L}} X$.

- Le théorèmes suivant donne la relation entre la convergence en loi d'une suite de variables aléatoires (X_n) et la convergence de la suite (ϕ_n) des fonctions caractéristiques.
- **Théorème 1:**

Si la suite de variables aleatoires (X_n) converge en loi vers la variable aléatoire X quand $n \to \infty$, alors la suite (ϕ_n) des fonctions caractéristiques converge vers la fonction caractéristique ϕ de la variable aléatoire X, la convergence étant uniforme dans tout intervalle fini de $\mathbb{R}$.

CONVERGENCE EN LOI DE VARIABLES ALÉATOIRES DISCRÈTE

- Soit $(X_n)_{n\geq 1}$ la variable aléatoire à valeurs dans $\mathbb{Z}$ et X une variable aléatoire à valeurs dans $\mathbb{Z}$.
- Notons $(p_{n,k}, k \in \mathbb{Z})$ la loi de $(X_n)_{n\geq 1}$ et $(\alpha_k, k \in \mathbb{Z})$ la loi de X. Alors les deux propriétés suivantes sont équivalentes:
  - pour tout $k \in \mathbb{Z}$ on a : $\lim\limits_{n\to\infty} p_{n,k} = \alpha_k$;
  - $X_n \xrightarrow{\mathcal{L}} X \quad (n \to \infty)$.

CONVERGENCE EN LOI DE VARIABLES ALÉATOIRES DISCRÈTE

- Soit $(X_n)_{n\geq 1}$ la variable aléatoire à valeurs dans $\mathbb{Z}$ et X une variable aléatoire à valeurs dans $\mathbb{Z}$.
- Notons $(p_{n,k}, k \in \mathbb{Z})$ la loi de $(X_n)_{n\geq 1}$ et $G_n(u) = \mathbb{E}(u^{X_n})$ sa fonction génératrice; enfin $(\alpha_k, k \in \mathbb{Z})$ la loi de X $G(u) = \mathbb{E}(u^X)$ sa fonction génératrice. Alors les deux propriétés suivantes sont équivalentes:
  - pour tout $k \geq 0$ on a : $\lim\limits_{n\to\infty} p_{n,k} = \alpha_k$ ($X_n \xrightarrow{\mathcal{L}} X$ ($n \to \infty$));
  - Pour tout $u \in]0,1[$ on a: $\lim\limits_{n\to\infty} G_n(u) = G(u)$.
- Ce théorème fournit le critère de convergence en loi d'une suite de variables aléatoires à valeurs entières positives.

- **Théoreme 2:** (Levy-Cramer-Dugue)

Si la suite (ϕ_n) de fonctions caractéristiques converge simplement, quand $n \to \infty$, vers une fonction ϕ et si la partie réelle de ϕ est continue a l'origine, alors:

- ϕ est une fonction caractéristique,
- la suite de fonctions de répartition F_n de X_n converge simplement vers la fonction de répartition F dont ϕ est la transformée de Fourier-Stieltjes, la convergence a lieu en tous les points de continuité de F.

- Si F est continue, la convergence est uniforme.

- Le fait que $X_n \xrightarrow{\mathcal{L}} X$ n'implique pas que pour tout réel x on ait $f_n(x) \to f(x)$ lorsque n tend vers l'infini.

Convergence en loi: Applications

- La convergence en loi est la plus faible mais elle est la plus utilisée car elle permet d'approcher la fonction de répartition de X_n par celle de X.
- C'est ainsi que l'on justifie la convergence de la loi binomiale vers la loi de Poisson, de la loi binomiale et de la loi de Poisson vers la loi normale (section: approximation).

- La convergence en probabilité entraîne la convergence en loi.
- La réciproque n'est pas exacte:
  - Soit X une variable aléatoire de loi $\frac{1}{2}(1\!\!1_0 + 1\!\!1_1)$ et $Y = 1 - X$. Alors X et Y ont même loi et $|X - Y| = 1$.
  - Considérons la suite $(X_n)_{n\geq 1}$ définie par $X_n = Y$ pour tout $n \geq 1$. Alors la suite $(X_n)_{n\geq 1}$ converge en loi vers X (trivialement), mais elle ne converge pas vers X en probabilité, puisque $|X_n - X| = |Y - X| = 1$.

- La réciproque est exacte dans le cas particulier où la limite X se réduit à une variable aléatoire presque certaine.
- Nous allons montrer que si la suite (X_n) converge en loi vers 0, alors elle converge en probabilité vers 0.
- Supposons donc que l'on ait:

$$F_n(x) \to \begin{cases} 1 & \text{si } x < 0 \\ 0 & \text{si } x > 0 \end{cases}$$

- On peut écrire, pour tout $\varepsilon > 0$ et tout $\eta > 0$ tel que $\varepsilon - \eta > 0$:

$$\begin{aligned} \Pr(|X_n| > \varepsilon) &= \Pr(X_n > \varepsilon) + \Pr(X_n < -\varepsilon) \\ &\leq \Pr(X_n > \varepsilon - \eta) + \Pr(X_n < -\varepsilon) \end{aligned}.$$

- D'où $\Pr(|X_n| > \varepsilon) \leq F_n(\varepsilon - \eta) + F_n(-\varepsilon) \to 0$.

SUITE I.I.D: DÉFINITION

- Une suite $(X_n)_{\mathbb{N}}$ de variables aléatoires réelles est dite identiquement-distribuée , en abrégé i.d., si toutes les variables aléatoires réelles de la suite ont la même loi.
- Une suite i.i.d. de variables aléatoires réelles est une suite indépendante et identiquement distribuée de v.a.r..

MOYENNE EMPIRIQUE: DÉFINITION

- Si $(X_n)_{\mathbb{N}}$ est une suite de variables aléatoires réelles. Pour tout entier $n \geq 1$, on appelle moyenne empirique d'ordre n associée à la suite $(X_n)_{\mathbb{N}}$, et on note $\overline{X}_{(n)}$ ou plus simplement $\overline{X}$, la v.a.r. définie par

$$\overline{X}_{(n)} = \frac{X_1 + \ldots + X_n}{n}.$$

- Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable ($\mathbb{E}(X^2) < \infty$), d'espérance m et de variance σ^2, alors

$$\mathbb{E}(\overline{X}_{(n)}) = m \quad \text{et } Var(\overline{X}_{(n)}) = \frac{\sigma^2}{n}.$$

LOI FAIBLE DES GRANDS NOMBRES

- On lance N fois une pièce. Soit $N(P)$ le nombre d'apparitions du côté pile. On observe que la fréquence empirique d'apparitions du côté pile $\frac{N(P)}{N}$ "converge" vers la probabilité d'obtenir pile. Le résultat suivant justifie cette intuition.
- On modélise les lancers par une suite $(X_n)_{n\geq 1}$ de v.a.d. indépendantes de loi de Bernoulli de paramètre p. Si $X_n = 1$, cela signifie que l'on a obtenu un pile au n-ième lancer.
  $S_n = \sum_{i=1}^{n} X_i$ représente le nombre de fois où pile est apparu en n lancers, et S_n/n représente la fréquence empirique ou la moyenne empirique.
- La proposition suivante assure que la probabilité pour que $|\frac{S_n}{n} - p|$ soit plus grand que ε est proche de 0 pour n grand. Ce résultat est valable pour tout $\varepsilon > 0$. On dit que S_n/n converge en probabilité vers p.

- Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d de carré intégrable et de même loi (d'espérance m). Alors,

LOI FAIBLE DES GRANDS NOMBRES.

Pour tout réel $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr(|\overline{X}_n - \mathbb{E}[X_1]| > \varepsilon) = 0\,.$$

- C-à-d, que la suite $(\overline{X}_n)_{\mathbb{N}^*}$ des moyennes empiriques converge en probabilité vers la v.a.r constante m.

Théoreme central-limite:

- Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et de même loi d'espérance m et de variance σ finies. Alors:

La suite de variables aléatoires $\dfrac{\displaystyle\sum_{i=1}^{n} X_i - nm}{\sigma\sqrt{n}}$ converge en loi vers la loi normale centrée réduite $\mathcal{N}(0,1)$.

- Concrètement, cela signifie que la loi de toute variable aléatoire égale à la somme d'un nombre "suffisamment grand" de variables aléatoires indépendantes et de même loi est approximativement une loi normale.

THÉOREME CENTRAL-LIMITE:

- Plus précisément, pour n grand, $\sum_{i=1}^{n} X_i$ est approximativement de loi $\mathcal{N}(nm, \sigma\sqrt{n})$.
- Ce qui est remarquable, c'est que ce résultat est vrai quelle que soit la loi des X_i.
- De très nombreux phénomènes naturels sont la résultante d'un grand nombre de phénomènes élémentaires identiques, indépendants et additifs ce qui justifie l'importance (et le nom) de la loi normale.

TCL

- Si X est de loi $\mathcal{N}(0,1)$, alors X^2 est de loi χ^2_1.
- Si $X_1,\dots,X_n$ sont indépendantes et de même loi $\mathcal{N}(m,\sigma)$, alors:
  - $\displaystyle\sum_{i=1}^{n} X_i$ est de loi $\mathcal{N}(nm,\sqrt{n}\sigma)$.
  - $\overline{X}_n = \frac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ est de loi $\mathcal{N}(m,\frac{\sigma}{\sqrt{n}})$.
  - $\frac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i-m)^2$ est de loi χ^2_n.
  - $\frac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$ est de loi χ^2_{n-1}.
- Les lois de Student et de Fisher-Snedecor sont toujours utilisées par l'intermédiaire de tables ou à l'aide d'un logiciel de statistique. Il n'est donc pas nécessaire de donner l'expression de leur densité.

DE LA LOI HYPERGÉOMÉTRIQUE PAR LA LOI BINOMIALE

- Soit E un ensemble de N éléments, dont une proportion p est de type "A". On effectue dans E une série de n tirages successifs sans remise. Soit X le nombre d'éléments de type "A" obtenus: X suit la loi hypergéométrique $\mathcal{H}(N,n,p)$.
- Si n et p sont fixés, pour tout $k \in \{0,\ldots,n\}$, alors

$$\sum_{N\to\infty} \frac{C_{N_p}^{n} C_{N_q}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}.$$

De la loi hypergéométrique par la loi binomiale: Application

- Quand N est grand, c'est-à-dire en pratique lorsque $N \geq 10n$, on peut approximer la loi hypergéométrique $\mathcal{H}(N,n,p)$ par la loi binomiale $\mathcal{B}(n,p)$.
- Dans le modèle précédent, ceci revient à considérer un tirage avec remise, car on a peu de chances de tirer 2 fois le même élément).

DE LA LOI BINOMIALE PAR LA LOI DE POISSON

- Si $(p_n)_n$ est une suite de réels de $[0, 1]$ telle que $\underset{n\to+\infty}{lim} n, p_n = \lambda$, alors, pour tout k fixé

$$\sum_{N\to\infty} C_n^k p_n^k (1 - p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}.$$

- En pratique, lorsque $p \leq 0.1$, $n \geq 30$ et $np < 15$ (ou < 10), on peut approximer la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(np)$.

De la loi de Poisson par la loi normale

- Si $(X_n)_n$ est une suite de variables aléatoires indépendantes, de même loi de Poisson $\mathcal{P}(\lambda)$ et si $S_n = X_1 + \ldots + X_n$, alors S_n suit la loi de Poisson $\mathcal{P}(n\lambda)$.
- S_n suit la loi de Poisson $\mathcal{P}(n\lambda)$ donc $\mathbb{E}(S_n) = n\lambda$ et $var(S_n) = n\lambda$.
- D'après le théorème central limite, la loi de S_n peut être approximée par la loi normale $\mathcal{N}(\mathbb{E}(S_n), var(S_n))$, c'est-à-dire $\mathcal{N}(n\lambda, \sqrt{n\lambda})$.
- En pratique, lorsque $\lambda \leq 15$, on peut approximer la loi de Poisson $\mathcal{P}(\lambda)$ par la loi $\mathcal{N}(\lambda, \sqrt{\lambda})$.

DE LA LOI DE POISSON PAR LA LOI NORMALE: EXEMPLE

- Si X suit la loi de Poisson $\mathcal{P}(16)$, $\Pr([X = 16]) = e^{-16}\frac{16^{16}}{16!} \approx 0.0992$.
- En approximant la loi $\mathcal{P}(16)$ par la loi $\mathcal{N}(16, \sqrt{16})$ de fonction de répartition F, on obtient:

$$\Pr([X = 16]) \approx F(16.5) - F(15.5) = \phi(\frac{0.5}{4}) - \phi(-\frac{0.5}{4}) = 2\phi(0.125)$$

- Le gain de temps est surtout sensible pour le calcul des valeurs de la fonction de répartition:

$$\Pr([X \leq 20]) \approx F(20.5) = \phi(\frac{20.5 - 16}{4}) = \phi(1.125) \approx 0.8697\,;$$

alors qu'en gardant la loi de Poisson, il faudrait faire la somme de 21 termes !

- Si $(X_n)_n$ est une suite de variables aléatoires indépendantes, de même loi de Bernoulli $\mathcal{B}(p)$ et si $S_n = X_1 + \ldots + X_n$, alors S_n suit la loi binomialle $\mathcal{B}(n, p)$.
- S_n suit la loi binomiale $\mathcal{B}(n, p)$, donc $\mathbb{E}(S_n) = n\,p$ et $var(S_n) = n\,p\,q$.
- D'après le théorème central limite, la loi de S_n peut être approximée par la loi normale $\mathcal{N}(\mathbb{E}(S_n), var(S_n))$, c'est-à-dire $\mathcal{N}(n\,p, \sqrt{n\,p\,q})$.
- En pratique, lorsque $n \geq 30$, $np \geq 15$ et $npq > 5$, la loi de Poisson binomiale $\mathcal{B}(n, p)$ peut être approximée par la loi normale $\mathcal{N}(n\,p, \sqrt{n\,p\,q})$.

- Si X suit la loi $\mathcal{B}(n,p)$, et si F est la fonction de répartition de la loi $\mathcal{N}(n\,p, \sqrt{n\,p\,q})$, on remplacera donc $\Pr([X=0])$ par $F(0.5)$; $\Pr([X=k])$ par $F(k+0.5) - F(k-0.5)$ pour $1 \leq k \leq n-1$ et $P([X=n])$ par $1 - F(n-0.5)$.

- Combien de lancers d'un dé faut-il effectuer pour pouvoir affirmer avec un risque d'erreur inférieur à 5% que la fréquence d'apparition de l'as au cours de ces lancers diffère de $\frac{1}{6}$ d'au plus $\frac{1}{100}$?
- Soit S_n le nombre d'as obtenus après n lancers: S_n suit la loi binomiale $\mathcal{B}(n, \frac{1}{6})$ et la fréquence d'apparition de l'as au cours de ces lancers est $\frac{S_n}{n}$.
- On cherche n tel que $\Pr\left(|\frac{S_n}{n} - \frac{1}{6}| \geq \frac{1}{100}\right) \leq \frac{5}{100}$.
- Si $n \geq 30$, $\frac{n}{16} \geq 15$et $\frac{5n}{36} > 5$, c'est-à-dire $n \geq 90$, on peut approximer la loi de $\frac{S_n}{n}$ par la loi normale $\mathcal{N}(\frac{1}{6}, \sqrt{\frac{5n}{36}})$ et la loi de $\frac{S_n}{n} - \frac{1}{6}$ par la loi normale $\mathcal{N}(0, \sqrt{\frac{5n}{36}})$.

- On a alors

$$\begin{array}{rcl} \Pr\left(|\frac{S_n}{n}-\frac{1}{6}|\geq\frac{1}{100}\right) & = & \Pr\left(|\frac{6\sqrt{n}}{5}(\frac{S_n}{n}-\frac{1}{6})|\geq\frac{6\sqrt{n}}{5}\frac{1}{100}\right) \\ & \approx & 2\left(1-\phi\left(\frac{6\sqrt{n}}{5}\frac{1}{100}\right)\right)\approx 0.05 \end{array}.$$

- On a donc $\phi\left(\frac{6\sqrt{n}}{5}\frac{1}{100}\right)\approx 0.975\approx\phi(1.96)$ et $n\approx(1.96)^2\frac{50000}{36}\approx 5536$.
- On a bien $5536\geq 90$, ce qui légitime l'approximation.