Théorèmes limites

Rappels

- n-échantillon d'une loi : n VAR $X_1, \dots, X_n$ mutuellement indépendantes et de même loi.
- Moyenne empirique d'un n échantillon : $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n}$. C'est une VAR et pas un nombre !
- Variance empirique d'un n échantillon (c'est aussi une VAR) :

$$S_n^2 = \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n}.$$

I Résultats fondamentaux

a) Inégalité de Markov

[Inégalité de Markov] Thm.1
Si X est une variable aléatoire positive admettant un moment d'ordre 1, pour tout réel $\lambda > 0$:

$$P(X \geq \lambda) \leq \frac{E(X)}{\lambda}$$

Rem.1 | Permet de majorer la probabilité qu'une VAR prenne de très grandes valeurs.

b) Inégalité de Tchebychev

[Inégalité de Tchebychev] Thm.2
Si X est une VAR admettant un moment d'ordre 2, alors pour tout réel $\epsilon > 0$:

$$P\left(|X - E(X)| \geq \varepsilon\right) \leq \frac{V(X)}{\varepsilon^2}$$

Rem.2 | Quantifie le risque que la valeur observée de VAR dévie de sa moyenne.

c) Loi faible des grands nombres

[Loi faible des grands nombres] Thm.3
Si $X_1 \dots X_n$ sont n VAR admettant un moment d'ordre 2, et de variance σ^2, alors $\overline{X}$ admet un moment d'ordre 2 et :

a. (*Version qualitative*)
$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left(|\overline{X} - E(\overline{X})| \geq \varepsilon\right) = 0.$$
b. (*Version quantitative*)
$$\forall \varepsilon > 0 \quad P(|\overline{X} - E(\overline{X})| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

II Lois approximativement égales

[Convergence en loi - pas à savoir] Déf.1
Soit X une VAR de fonction de répartition F_X On dit qu'une suite $(X_n)_{n\geq 0}$ a une loi approximativement égale à la loi de X si **en tout point** x **de continuité de** F_X : $\lim_{n\to+\infty} F_{X_n}(x) = F_X(x)$. On dit alors que (X_n) converge en loi vers X.

[Utilisation de l'approximation] Thm.4
Si (X_n) a une loi approximativement égale à la loi de X, alors pour tous réels $a \leq b$ en lesquels F_X est continue, et n assez grand [a]

$$P(X_n \in]a, b]) \simeq P(X \in]a, b]).$$

a. le seuil étant fixé par dans les cas classiques par les théorèmes limites.

Rem.3 |

a. En effet, comme $P(X_n \in]a, b]) = F_{X_n}(b) - F_{X_n}(a)$, cela donne :
$$P(X_n \in]a, b]) \underset{n\to\infty}{=} P(X \in]a, b]) + o(1),$$
ce qui traduit bien que les lois sont approximativement égales.
b. Dans le cas particulier où X,et X_n sont des VAR prenant un nombre fini de valeurs, il suffit d'avoir :
$\forall x \in X(\Omega)\ P(X_n = x) = P(X = x) + o(1)$ puisque : $P(X = x) = P(X \in]a, b])$ avec $b = x + \varepsilon$ et $a = x - \varepsilon$ pour un $\varepsilon > 0$ suffisamment petit bien choisi).

III Théorèmes limites

Usage : remplacer une loi donnée par une loi approximativement égale pour les calculs

a) Loi hypergéométrique

Rem.4 | Des commentaires et remarques.

a. Loi géométrique paramétrée par N, M, n :
— N est la taille de la population totale étudiée (exemple : une urne contenant N boules).
— M est la taille de la sous-population qui nous intéresse (exemple : l'urne contient exactement M boules noires).
— n est la la taille de l'échantillon prélevé (exemple : on extrait une poignée de n boules de l'urne).
La loi géométrique est la loi du nombre de boules noires dans l'échantillon

[Approximation binomiale de la loi hypergéométrique] Thm.5
Dès que $N \geq 10n$, la loi $\mathscr{H}(N, M, n)$ est approximativement la loi binomiale $\mathscr{B}\left(n, \dfrac{M}{N}\right)$.

Exple.1 |
Pour `N,M,n=50,9,5` (seuil de validité de l'approximation) :

```
k |Loi  B(n,p)  |Loi H(N,M,n)
____________________________
0 |0.37         |0.353
1 |0.406        |0.43
2 |0.178        |0.181
3 |0.039        |0.032
4 |0.004        |0.002
5 |0.0          |0.0
```

b) Approximation poissonienne

[Approximation poissonienne de la loi binomiale] Thm.6
Dès que N est grand et p petit, de sorte que Np est de l'ordre de quelques unités, la loi $\mathscr{B}(N, p)$ est approximativement la loi $\mathscr{P}(Np)$.

Exple.2 |
Pour `N,p = 900,1/365` (On a bien N grand, p petit, et $Np \simeq 2$ est de l'ordre de quelques unités) :

k	B(N,p)	P(Np)
0	0.085	0.085
1	0.209	0.209
2	0.258	0.258
3	0.213	0.212
4	0.131	0.131
5	0.064	0.065
6	0.026	0.027
7	0.009	0.01
8	0.003	0.003
9	0.001	0.001

c) Théorème central-limite

[Central-limite forme 1 : variance connue] **Thm.7**

Si $n \geq 30$ et si (X_n) un n-échantillon de variables aléatoires d'espérance μ et de variance $\sigma^2 > 0$, alors la loi de l'erreur normalisée $X^\star = \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ est approximativement une loi $\mathcal{N}(0,1)$.

[Central-limite forme 2 : variance inconnue] **Thm.8**

Si $n \geq 30$ et si (X_n) un n-échantillon de variables aléatoires d'espérance μ et de variance inconnue, alors la loi de l'erreur normalisée empirique $X^\star = \dfrac{\overline{X}-\mu}{\frac{S_n}{\sqrt{n}}}$ est approximativement une loi $\mathcal{N}(0,1)$.

Les résultats qui suivent sont des cas particuliers du TCL forme 1

d) Approximation gaussienne de la loi de Poisson

[Approximation gaussienne de la loi de Poisson] **Thm.9**

Si $\lambda \geq 18$, la loi de Poisson $\mathcal{P}(\lambda)$ est approximativement une loi $\mathcal{N}(\lambda,\lambda)$.

Rem.5

a. On a donc une loi normale de paramètres ceux de la loi de Poisson qu'elle approche.

b. La somme de n variables indépendantes de loi $\mathcal{P}(\alpha)$ suit une loi de Poisson de paramètre $n\alpha$, donc le théorème s'applique aussi avec $\lambda = n\alpha$ dès lors que $n\alpha \geq 18$.

e) Approximation gaussienne de la loi binomiale

[de Moivre-Laplace] **Thm.10**

Si $p > 0$, $q = 1-p$, $n \in \mathbf{N}$ sont tels que $n \geq 30$ et $np, nq \geq 5$, alors la loi $\mathcal{B}(n,p)$ est approximativement une loi $\mathcal{N}(np, npq)$.

Rem.6 On a donc une loi normale de paramètres ceux de la loi binomiale qu'elle approche.

Rem.7

a. $T_1>$ En pratique, on utilise ces théorèmes en se ramenant à une **loi normale centrée réduite** par centrage-réduction de la loi de Poisson/binomiale :

1) Si $X \rightsquigarrow \mathcal{P}(n\alpha, n\alpha)$ on déduit que la loi de $X^\star = \dfrac{X - \overbrace{n\alpha}^{E(X)}}{\underbrace{\sqrt{n\alpha}}_{\sigma_X}}$ suit approximativement une loi $\mathcal{N}(0,1)$.

2) Si $X \rightsquigarrow \mathcal{B}(n,p)$ on déduit que la loi de $X^\star = \dfrac{X - \overbrace{np}^{E(X)}}{\underbrace{\sqrt{npq}}_{\sigma_X}}$ suit approximativement une loi $\mathcal{N}(0,1)$.

b. $T_2>$ **Correction de continuité.** Comme les lois de Poisson et binomiale sont discrètes à valeurs entières, pour avoir une approximation de $P(X=k)$, on utilise ces théorèmes en effectuant une *correction de continuité*, qui consiste à dire que :

1) si Z suit la loi normale qui approche celle de X :

$$P(X=k) = P\left(k-\frac{1}{2} \leq X \leq k+\frac{1}{2}\right)$$

$$\simeq P(Z \in \left[k-\frac{1}{2}, k+\frac{1}{2}\right])$$

2) Ensuite, on calcule cette dernière probabilité par centrage-réduction de la loi de Z, ce qui permet de se reporter aux valeurs des quantiles de la loi de $Z^\star \rightsquigarrow \mathcal{N}(0,1)$.

Exple.3 *Estimer $P(X=3)$ pour $X \rightsquigarrow \mathcal{P}(20)$.* Comme $20 \geq 18$, on peut considérer l'approximation de X par une variable Z de loi $\mathcal{N}(m,m)$ où $m=20$ valide (mais on est quasiment au seuil de validité).

a. On écrit d'abord que

$$P(X=3) = P(X \in [2,5;3,5]) \overset{\text{Thm. 9}}{\simeq} P(Z \in [2,5;3,5])$$

b. Ensuite on centre-réduit : la variable $Z^\star = \dfrac{Z-m}{\sqrt{m}}$ suit une loi $\mathcal{N}(0,1)$ et de plus :

$$Z \in [2,5;3,5] = Z^\star \in \left[\underbrace{\frac{2,5-m}{\sqrt{m}}}_{a};\underbrace{\frac{3,5-m}{\sqrt{m}}}_{b}\right].$$

Par déf. de la loi $\mathcal{N}(0,1)$: $P(Z^\star \in [a,b]) = \Phi(b)-\Phi(a)$, et la calculatrice donne $0,048 \simeq \Phi(b)-\Phi(a) \simeq P(X=3)$, la valeur réelle de $P(X=3)$ étant environ $0,052$: approximation acceptable.

Exple.4 *Combien de fois N faut-il lancer une pièce de monnaie pour être sûr à 95% que le nombre de pile observés soit dans l'intervalle $I = \left[\frac{9N}{20}, \frac{11N}{20}\right]$?*

La loi de la VAR X_N qui compte le nombre de piles est une loi $\mathcal{B}\left(N, \frac{1}{2}\right)$. Comme $E(X_N) = N/2$, on cherche N tel que $P\left(X_N \in \left[E(X) - \frac{N}{20}, E(X)+\frac{N}{20}\right]\right) \geq 95\%$. En supposant $N \geq 30$, et comme $p = 1/2$, les conditions du théorème de Moivre-Laplace sont vérifiées : X_N suit approximativement une loi normale de centre $E(X_N)$ et de variance $V(X_n) = N/4$. On centre-réduit :

$$X \in I = \overbrace{\frac{X-E(X)}{\frac{\sqrt{N}}{2}}}^{X^\star} \in \left[-\frac{\sqrt{N}}{10}, \frac{\sqrt{N}}{10}\right].$$

Il suffit donc de choisir N de sorte que $\sqrt{N}/10 \geq 2,58$ (quantile d'ordre $0,975$ de loi normale centré réduite). L'inéquation donne : $N \geq 666$.