Complément sur le T.P. SAS numéro 09

Quelques précisions sur le théorème "central limit"

Théorème, dit *central limit*

Considérons un échantillon de n v.a.r. $X_1 \dots X_i \dots X_n$, i.i.d. selon la loi d'une v.a.r. X vérifiant :

$$\mathbb{E}(X) = \mu \quad ; \quad \mathrm{Var}(X) = \sigma^2 > 0.$$

Posons : $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ (moyenne empirique de l'échantillon). On sait que la statistique $\overline{X}_n$ vérifie :

$$\mathbb{E}(\overline{X}_n) = \mu \quad ; \quad \mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Désignons par Z_n la statistique centrée et réduite associée à $\overline{X}_n$. Il vient :

$$Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}.$$

Alors, lorsque n tend vers l'infini, Z_n converge en loi vers une loi normale centrée réduite ($\mathcal{N}(0, 1)$), quelle que soit la distribution initiale de X. $\triangle$

Ce résultat, très important, explique en grande partie le rôle fondamental de la loi normale en statistique.

Illustration avec la loi uniforme

La loi uniforme continue sur l'intervalle réel $[a, b]$ $(a < b)$ est très différente d'une loi normale. En particulier, la probabilité d'observer des valeurs sur un intevalle d'amplidude donnée (inclus dans $[a, b]$) ne dépend pas de la position de cet intervalle à l'intérieur de $[a, b]$. Pourtant, la convergence de la moyenne empirique centrée et réduite de lois uniformes indépendantes vers la loi $\mathcal{N}(0, 1)$ est très rapide. C'est ce qu'illustre l'exercice 09.8.

On rappelle qu'une loi uniforme X sur $[a, b]$ vérifie :

$$\mathbb{E}(X) = \frac{a + b}{2} \quad ; \quad \mathrm{Var}(X) = \frac{(b - a)^2}{12}.$$

La loi uniforme sur $[0, 1]$ vérifie donc :

$$\mathbb{E}(X) = \frac{1}{2} \quad ; \quad \mathrm{Var}(X) = \frac{1}{12}.$$

Dans ce dernier cas, la statistique Z_n s'écrit :

$$Z_n = \sqrt{n}\frac{\overline{X}_n - 1/2}{1/\sqrt{12}} = \sqrt{12\ n}\ (\overline{X}_n - 1/2).$$

L'objet de l'exercice 09.8 est de simuler, avec un n assez petit, un millier de statistiques Z_n pour regarder leur histogramme et contrôler ainsi leur normalité. On commence avec $n = 10$ (en général, c'est bon !), puis on diminue progressivement la valeur de n.