Loi forte des grands nombres et théorème central limite

L'objectif de ce TP est de mettre en pratique la loi fortes des grands nombres (LFGN) et le théorème central limite (TCL) rappelées ci-dessous.

LFGN Soient (X_n) une suite de variables aléatoires indépendantes et de même loi. Alors si $E[|X_1|] < \infty$,

$$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow[n\to\infty]{p.s} E[X_1]$$

TCL Si de plus $E[X_1^2] < \infty$, alors

$$\sqrt{n}\left(\bar{X}_n - E[X_1]\right) \xrightarrow[n\to\infty]{Loi} \mathcal{N}(0, Var(X_1))$$

Loi fortes des grands nombres

Une entreprise de vêtement américaine *MadeInC* souhaite réaliser une enquête sur la taille des françaises afin d'ajuster la taille de ces vêtements aux clientes de l'hexagone. Pour cela elle recrute 50 enquêteurs dans toute la France afin de mesurer la taille des françaises. On sait, via une enquête secrètement réalisée par un concurrent, que la taille des clientes suit une loi normale de moyenne $\mu = 162.5$ cm et de variance $\sigma^2 = 2$ cm.

1. Simuler une population de $n = 1000$ clientes et tracer sa densité.
2. Calculer la moyenne empirique $\bar{x}_n$ et l'écart type empirique s de votre population. Pourquoi ne retrouvez-vous pas exactement 162.5 et $\sqrt{2}$?
3. Que représentent $\bar{x}_n$ et s^2 par rapport à μ et σ^2. Justifier votre réponse.
4. Sachant que $\bar{x}_n$ est la réalisation de la variable aléatoire $\bar{X}_n$. Donner la loi, l'espérance et la variance de $\bar{X}_n$.
5. Chaque enquêteurs mesure $n = 30$ femmes. Simuler les résultats des 50 enquêteurs. Calculer la taille moyenne empirique $\bar{x}_n$ obtenue par chaque enquêteur. Vous venez de créer un échantillon de 50 tailles moyennes.
6. Tracer la distribution des moyennes et commenter les résultats en les comparant avec le graphique précédent.
7. Quelle est la probabilité qu'une femme mesure (voir la fonction `norm.cdf` du module `scipy.stats`) :
   - exactement 168 cm ?
   - moins de 160 cm ?
   - plus de 165 cm ?
   - entre 158 cm et 164 cm ?

8. Un des enquêteurs de *MacInC*, un peu étourdi (ou mal organisé), a réalisé des enquêtes dans différents pays, mais il a oublié d'inscrire le pays associé à chaque enquête. Il trouve un fichier avec 50 mesures. La moyenne des données de ce fichier est 164. On suppose que la variance de la taille est la même dans tous les pays et vaut 2. Cette enquête a-t-elle été réalisée en France ? Justifier votre réponse

TCL

> L'entreprise d'ampoules *Neaugreen* souhaite estimer la durée de vie de nouvelles ampoules grand luxe en cristal à 45 euros l'unité. Pour cela, elle demande à 200 de ses succursales de réaliser des tests d'usure.
> Dans chaque succursale, les employés doivent maintenir $n = 5$ ampoules allumées en permanence et enregistrer la durée de vie de chaque ampoule. A l'issue de l'enquête, le chargé détudes statistiques analyse les résultats. Cette enquête a montré que la durée de vie des ampoules suivent une loi exponentielle de paramètre $\lambda = 5$ (notée par $\mathcal{E}(5)$). Ceci correspond à une durée de vie moyenne de $1/\lambda = 0,2$ année (2,4 mois) et une variance variance $1/\lambda^2$ de 0,04 années.

1. Simuler un échantillon de 1000 ampoules suivant une loi $\mathcal{E}(5)$ et tracer son histogramme.
2. Simuler les résultats des 200 succursales (simulation n = 5 ampoules/succursale). Calculer la durée de vie moyenne de chaque succursale $(\bar{x}_n^1, ..., \bar{x}_n^{200})$.
3. Afficher l'histogramme de l'échantillon des moyennes. Calculer la moyenne et l'écart-type de la moyenne. Commenter.
4. Le chargé détude statistiques est un peu étonné par la forme de l'histogramme obtenu et demande aux 200 succursales de refaire les tests, mais cette fois avec 1000 ampoules. Répétez l'expérience et visualiser les résultats à l'aide d'un histogramme.
5. Interprêter les résultats de simulation à l'aide du théorème central limite.
6. Quelle est la probabilite qu'une ampoule Neaugreen ait une durée de vie moyenne :
   - inférieure à 1 mois ?
   - superieure à 12 mois

Bonus

1. Quelle est la loi de la variable $\sqrt{n}\left(\frac{\bar{X}_n-\mu}{S}\right)$?
2. Tracer sa fonction de densité. Que devient la loi de cette variable lorsque $n \to \infty$?
   Rappels :
   - estimateur de la variance : $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$
   - $(n-1)\frac{S^2}{\sigma^2}$ suit une loi du Chi deux à $n-1$ degré de liberté $\mathcal{X}^2(n-1)$
   - Soit Z une variable gaussienne $\mathcal{N}(0,1)$ et U une variable $\mathcal{X}^2(k)$, Z et U indépendantes. Alors $\frac{Z}{\sqrt{\frac{U}{k}}}$ suit une loi de student à k degrés de liberté.