Tutorat N.2: somme de variables aléatoires et théorème de la limite centrale

Le but de ce tutorat est d'étudier le comportement limite de la somme d'un grand nombre de variables aléatoires. La densité de probabilité d'une variable aléatoire X à valuers réelles sera notée $P_X(x)$. Par simplicité, nous supposerons la loi de X symétrique, $P_X(x) = P_X(-x)$, de sorte que le $n$ème moment $M_n(X) = \langle X^n \rangle$ soit nul pour n impair.

A. Fonction génératrice et théorème de la limite centrale (TLC)

La fonction génératrice est définie de la facon suivante:

$$G_X(t) = \left\langle e^{itX} \right\rangle = \int_{-\infty}^{\infty} dx P_X(x) e^{itx} \ . \tag{1}$$

On définit le $n$ème cumulant de X, $M_n^c(X)$, par

$$\log G_X(t) = \sum_{n=1}^{\infty} M_n^c(X) \frac{(it)^n}{n!} \tag{2}$$

1. Montrer que $G_X(t)$ existe et que $|G_X(t)| \leq 1$.
2. Comment obtient-on $P_X(x)$ à partir de $G_X(t)$?
3. Les cumulants ne sont pas toujours bien définis. Donner la condition sur $P_X(x)$ pour que les prémiers n cumulants soient bien définis.
4. Comment s'expriment les deux prémiers cumulants non nuls en fonction des moments de X?
5. Montrer que les cumulants d'une variable gaussiene ($P_X(x) = \mathcal{G}_\sigma(x) = \exp(-x^2/(2\sigma^2))/\sqrt{2\pi\sigma^2}$) sont nuls pour $n > 2$.
6. On considère $S_n = X_1 + \cdots + X_N$, où les X_i sont N variables aléatoires indépendantes distribuées avec la même loi que X. Que vaut la fonction génératrice de S_N en fonction de celle de X ? Que valent les cumulants de S_N en fonction de ceux de X ?
7. *Théorème de la limite centrale (TLC)* - On définit $s_N = S_N/\sqrt{N}$. Calculer la fonction génératrice de s_N quand N est très grand. En déduire que la distribution de s_N converge vers une distribution gaussienne de variance $\sigma^2 = M_2^c(X) = \left\langle X^2 \right\rangle$, qu'on denotera $\mathcal{G}_\sigma(x)$.
8. A travers d'une expansion de la fonction génératrice de s_N à l'ordre 4 en t, monter que

$$P_{s_N}(x) = \mathcal{G}_\sigma(x) \left[1 + \frac{M_4^c(X)}{24 N M_2^c(X)^2} \left(3 - \frac{6x^2}{M_2^c(X)} + \frac{x^4}{M_2^c(X)^2} \right) + o(1/N) \right] \tag{3}$$

   En déduire que $P_{S_N}(x)$ peut être approximée par une loi gaussienne de largeur $\sigma\sqrt{N}$ dans un intervalle de largeur d'ordre N^α et donner la valeur de α.

9. Est-il nécessaire que tous les X_i aient la même distribution? Si possible, généraliser le TLC au cas où chaque X_i a une distribution differénte.

B. Grandes déviations

Nous venons de voir qu'il existe un intervalle de taille N^α (avec $1/2 < \alpha < 1$) dans laquelle le TLC s'applique pour N très grand. La taille de cet intervalle est donc petite devant N, et on peut se poser la question de quelle est la distribution limite de S_N quand ses valeurs sont d'ordre N. On s'interesse donc à la variable $u_N = S_N/N$ et on cherche la distribution de u_N quand ses valeurs sont finis par rapport à N qui est supposé très grand. On définit par commodité $g_X(t) = \log G_X(t)$.

1. En utilisant la méthode du col, montrer que $P_{u_N}(u) \sim C_N \exp[-N f_X(u)]$, où la "fonction des grandes déviations" $f_X(u)$ est definie par
$$f_X(u) = iz^* u - g_X(z^*) \ , \tag{4}$$
z^* étant la solution de $g'_X(z) = iu$.
2. On suppose que la distribution de X soit une gaussienne de largeur 1, $P_X(x) = \mathcal{G}_1(x)$. Calculer $g_X(t)$ et, en utilisant l'equation (4), calculer $f_X(u)$. Est-ce que dans ce cas on peut calculer $f_X(u)$ d'une facon plus simple?
3. Montrer, sous l'hypothèse que $P_X(x) = P_X(-x)$ et que tous les cumulants de X existent, que la solution z^* est un nombre imaginaire pur, $z^* = i\zeta^*$ avec ζ^* réel. Il apparaît donc clairement que $f_X(u)$ et $g_X(i\zeta)$ sont reliées par transformation de Legendre. En déduire que $f_X(u)$ est une fonction convexe.

C. Rélation avec les marches aléatoires

Considérons une marche aléatoire à une dimension, c'est-à-dire une particule qui, à chaque pas, saute d'une distance a vers la droite ou vers la gauche avec la même probabilité. On appelle X_n la variable aléatoire qui décrit la distance parcourue au $n$ème pas.

1. On note que toutes les X_n ont la même distribution $P_X(x)$. Quelle est la distribution de probabilité $P_X(x)$ de X, *vue comme une variable aléatoire continue* ? Quelle est sa fonction génératrice $G_X(t)$?
2. Quelle est la distribution de probabilité de s_N ?
3. Quelle est la fonction de grande déviations $f_X(u)$?