SUR L'ÉCART ENTRE LE MAXIMUM ET LE MINIMUM DE n VARIABLES ALÉATOIRES INDÉPENDANTES SUIVANT TOUTES LA MÊME LOI NORMALE RÉDUITE

Pour $n \geq 2$, nous considérons n variables aleatoires indépendantes de même loi, $X_1, \ldots, X_n$. Nous supposons que cette loi a une densité f et une fonction de répartition F. Posons

$$M_n = \max(X_1, \ldots, X_n) \quad \text{et} \quad m_n = \min(X_1, \ldots, X_n)$$

Puis

$$D_n = M_n - m_n.$$

Nous nous intéressons a la fonction de répartition de M_n et de D_n. Pour cela nous etudions la loi de M_n puis la loi du couple (m_n, M_n).

Nous avons

$$P(M_n \leq x) = P(X_1 \leq x \cap X_2 \leq x \cdots \cap X_n \leq x)$$

donc

$$P(M_n \leq x) = (F(x))^n.$$

Par dérivation, nous obtenons la densite de la variable aléatoire M_n, notée f_{M_n} :

$$f_{M_n}(x) = nf(x)(F(x))^{n-1}.$$

D'où nous déduisons l'espérance de M_n :

$$E(M_n) = n \int_{-\infty}^{+\infty} xf(x)(F(x))^{n-1}dx. \tag{1}$$

Passons a la fonction de répartition du couple (m_n, M_n). Nous avons pour $x \leq y$

$$P(m_n \leq x \cap M_n \leq y) = P(M_n \leq y) - P(m_n \geq x \cap M_n \leq y)$$

$$= F(y)^n - (F(y) - F(x))^n.$$

Donc par double dérivation par rapport à x puis par rapport à y, nous obtenons la densité du couple (m_n, M_n), notée $f_{(m_n,M_n)}$:

$$f_{(m_n,M_n)}(x, y) = n(n-1)f(x)f(y)(F(y) - F(x))^{n-2}.$$

Alors nous en déduisons la fonction de répartition de la variable D_n. Nous avons $P(D_n \leq u) = 0$ pour $u \leq 0$ et pour $u \geq 0$

$$P(D_n \leq u) = \int\int_{x \leq y \leq x+u} f_{(m_n,M_n)}(x,y)dxdy$$

$$= n(n-1)\int_{-\infty}^{+\infty} xf(x)(\int_{x}^{x+u} f(y)(F(y)-F(x))^{n-2}dy)dx$$

ou encore

$$P(D_n \leq u) = n\int_{-\infty}^{+\infty} xf(x)(F(x+u)-F(x))^{n-1}dx \qquad (2)$$

Si la loi des X_i est la loi normale centrée et réduite, alors nous avons

$$f(x) = (2\pi)^{-0.5}e^{-x^2/2} \quad \text{et} \quad F(x) = \Pi(x) = 0.5(1 + erf(2^{-0.5}x). \qquad (3)$$

Dans ce cas nous avons $m_n(X_1, \ldots, X_n) = -M_n(-X_1, \ldots, -X_n)$. Comme les variables aléatoires $-X_i$ sont indépendantes et suivent toutes la même loi normale centrée et réduite, nous avons

$$E(m_n) = -E(M_n) \quad \text{et} \quad E(D_n) = 2E(M_n).$$

Pour le calcul de $E(D_n)$ par la formule (1) et avec (3), par des methodes classiques et elementaires, on obtient $E(D_2) = 2\pi^{-0.5}$ et $E(D_3) = 3\pi^{-0.5}$. Pour les valeurs de n supérieures à 3, en utilisant un calcul approché par informatique, on voit des valeurs qui augmentent tres lentement. Maintenant nous considérons les quantités $u_{n,0.95}$ qui sont définies par $P(D_n \leq u_{n,0.95}) = 0.95$. Le cas $n = 2$ est simple car $D_2 = |X_1 - X_2|$ et $X_1 - X_2$ est une variable aléatoire suivant une loi normale, on obtient ainsi facilement $u_{2,0.95}$. Pour les valeurs de n supérieures à 2, par (2), (3) et un calcul approché effectué par ordinateur on obtient les valeurs indiquées dans la table C-iso 57256. On constate que cette suite $(u_{n,0.95})_{n\geq 2}$, liée à la precédente, croit comme cette dernière très lentement. Il n'y a pas de raisons objectives pour que ces deux suites convergent. Il est d'ailleurs possible de prouver que la suite $(E(M_n))_{n\geq 2}$ diverge. Il s'agit là d'un problème mathématique intéressant bien qu' assez éloigné des préoccupations habituelles d'un biochimiste (voir annexe maxln.pdf).