On sait qu'une loi binomiale $X = B(n\,;p)$ peut être approchée par une loi normale $X = N(\mu\,;\sigma^2)$, sous réserves que les conditions minimales suivantes soient respectées : $n > 30$, $np > 5$ et $n(1-p) > 5$.

Remarque : *Conditions peu restrictives,* $n > 100$ *serait préférable.*

On sait qu'alors $\mu = n.p$ et $\sigma = \sqrt{n.p.(1-p)}$.

<u>*Intervalle de fluctuation asymptotique au seuil de*</u> $1-\alpha$:

Si la variable aléatoire X_n suit la loi binomiale $B(n\,;p)$, pour tout réel $\alpha \in]0\,;1[$:

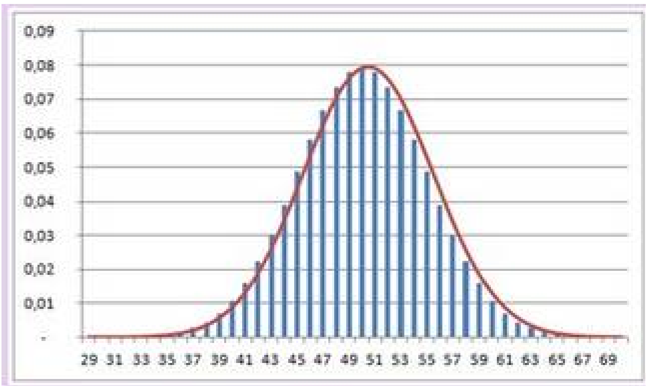
u_α désigne le réel tel que $p(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$ lorsque $Z = N(0\,;1)$.

L'intervalle $I_n = [\,p - u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\,p + u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,]$ est appelé <u>*intervalle de fluctuation asymptotique au seuil de*</u> <u>$1-\alpha$</u> . C'est-à-dire que : $\lim\limits_{n\to+\infty} p(\dfrac{X_n}{n} \in I_n) = 1-\alpha$.

<u>*On considère que la limite est atteinte dès que les conditions nécessaires sont satisfaites.*</u>

On peut alors écrire : $\boldsymbol{Prob\left(p - u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \dfrac{X_n}{n} \leq p + u_\alpha.\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha}$ **.**

Les conditions initiales ont pour objectif que la courbe représentative de la loi binomiale soit la plus équilibrée et symétrique possible, afin qu'une loi normale puisse lui être superposée.



Remarquer que les effectifs n infinis n'ont aucun sens pour une loi binomiale, et que si la loi normale n'est pas affectée par les inégalités strictes ou pas, il n'en est pas de même pour la loi binômiale.

$X = N(\mu\,;\sigma^2)$: $P(75 \leq X \leq 80) = P(75 \leq X < 80) = P(75 < X < 80)$.

$X = B(n\,;p)$: $P(75 \leq X \leq 80) = 0{,}825$, $P(75 \leq X < 80) = 0{,}707$, $P(75 < X < 80) = 0{,}653$.

L'appartenance ou non des bornes influence sensiblement le résultat pour $B(n\,;p)$.

A titre d'exemple : Soit $X = B(n\,;p) = B(20\,;0{,}25)$.

Alors : $E(X) = np = 5$. $p(X = 2) = \begin{pmatrix}20\\2\end{pmatrix}(0{,}25)^2(0{,}75)^{18} = 0{,}067$ et $p(X = 8) = \begin{pmatrix}20\\8\end{pmatrix}(0{,}25)^8(0{,}75)^{12} = 0{,}061$.

La symétrie par rapport à la moyenne 5 est très relative. $n = 20$ est trop faible, et $E(X) = np$ est trop proche de 5 .

On voit d'ailleurs dans le graphique précédent l'importance du découpage en un grand nombre de rectangles, soit n grand.

Théorème Central-Limite T.C.L.

Il est à remarquer que cette *convergence* de la loi binomiale vers une loi normale, lorsque n est grand, est valable pour toutes les lois statistiques.

A titre d'exemple :

En entreprise, pour mesurer le salaire des employés, on n'utilise jamais le *salaire moyen* $E(X) = \frac{M}{n}$, raport de la masse salariale totale M au nombre de salariés, parce que les gros salaires feront toujours que ce salaire moyen se déportera vers le haut, et que la courbe de répartition sera déformée en ce sens.

Exemple : 5 salaires, 1200 € , 1250 € , 2000 € , 7500 € , 10000 € $\Rightarrow$ $E(X) = \frac{21950}{5} = 4390$ € .

On utilise plutôt le *salaire médian Me* , tel que 50% des employés gagnent moins, et 50% gagnent plus. $Me = 2000$ € .

Ce qui est remarquable, est que si on prend 1000 entreprises, chacune ayant cette déformation de sa courbe de répartition des salaires, dont la bosse se trouve clairement à droite, la répartition statistique des salaires moyens de ces 1000 entreprises sera très symétrique par rapport à un salaire moyen global (moyenne des salaires moyens de chaque entreprise), donc lui-même non représentatif de la réalité de la répartition des salaires dans chaque entreprise.
On pourra identifier cette courbe à une loi normale.
On peut évidemment procéder avec les salaires médians, plus représentatifs de la réalité de chaque entreprise.
C'est ce qu'on appelle le T.C.L. (Théorème Central Limite).