APPROXIMATION BINOMIALE - NORMALE

UNE ILLUSTRATION DU THEOREME DE MOIVRE - LAPLACE

PICHARD JEAN FRANÇOIS
IREM DE ROUEN

On utilise souvent l'approximation de la loi binomiale par la loi normale (nom maintenant consacré donné à la loi de Laplace-Gauss, étudiée pourtant en premier par Daniel Bernoulli). Ceci est justifié par le théorème de Moivre-Laplace, qui est un cas particulier du théorème central-limit.

Cependant, quelques difficultés demeurent pour les étudiants, et quelquefois pour les utilisateurs de la statistique, qui se posent en particulier les questions suivantes :

- comment établir le lien entre la loi binomiale, qui est discrète, et la loi normale qui est continue ?
- à partir de quelle taille d'échantillon peut-on valablement utiliser cette approximation ?

Posée sous cette forme, cette dernière question est imprécise puisque la loi binomiale dépend de 2 paramètres, notés habituellement n et p, qui vont influer tous les deux sur les valeurs de probabilité. Il faut de plus préciser ce qu'on entend par approximation valable. La réponse sera différente suivant l'utilisation que l'on en fait : probabilité d'événement, test, intervalle de confiance.

L'application du théorème de Moivre-Laplace à une v.a. binomiale $X \in \mathcal{B}(n, p)$ permet d'obtenir des valeurs approchées pour les probabilités :

$$P[X = k] \approx P\left[\frac{k - 1/2}{\sqrt{npq}} \leq U \leq \frac{k + 1/2}{\sqrt{npq}}\right], \text{ pour n grand, } k \in \mathbb{N}, \text{ et où } U \in \mathcal{N}(0,1).$$

Pour établir le lien entre loi binomiale et loi normale, on va représenter la distribution de la v.a. X de loi $\mathcal{B}(n,p)$ par un histogramme. Pour cela, on associe à chaque valeur entière k de X un intervalle $[k-1/2, k+1/2]$ de largeur 1 centré en cette valeur, la hauteur du rectangle correspondant étant $P[X=k]$. L'aire de tous ces rectangles est alors égale à 1, puisque c'est la somme des probabilités. L'association d'un intervalle de largeur 1 à une valeur entière est appelée correction de continuité.

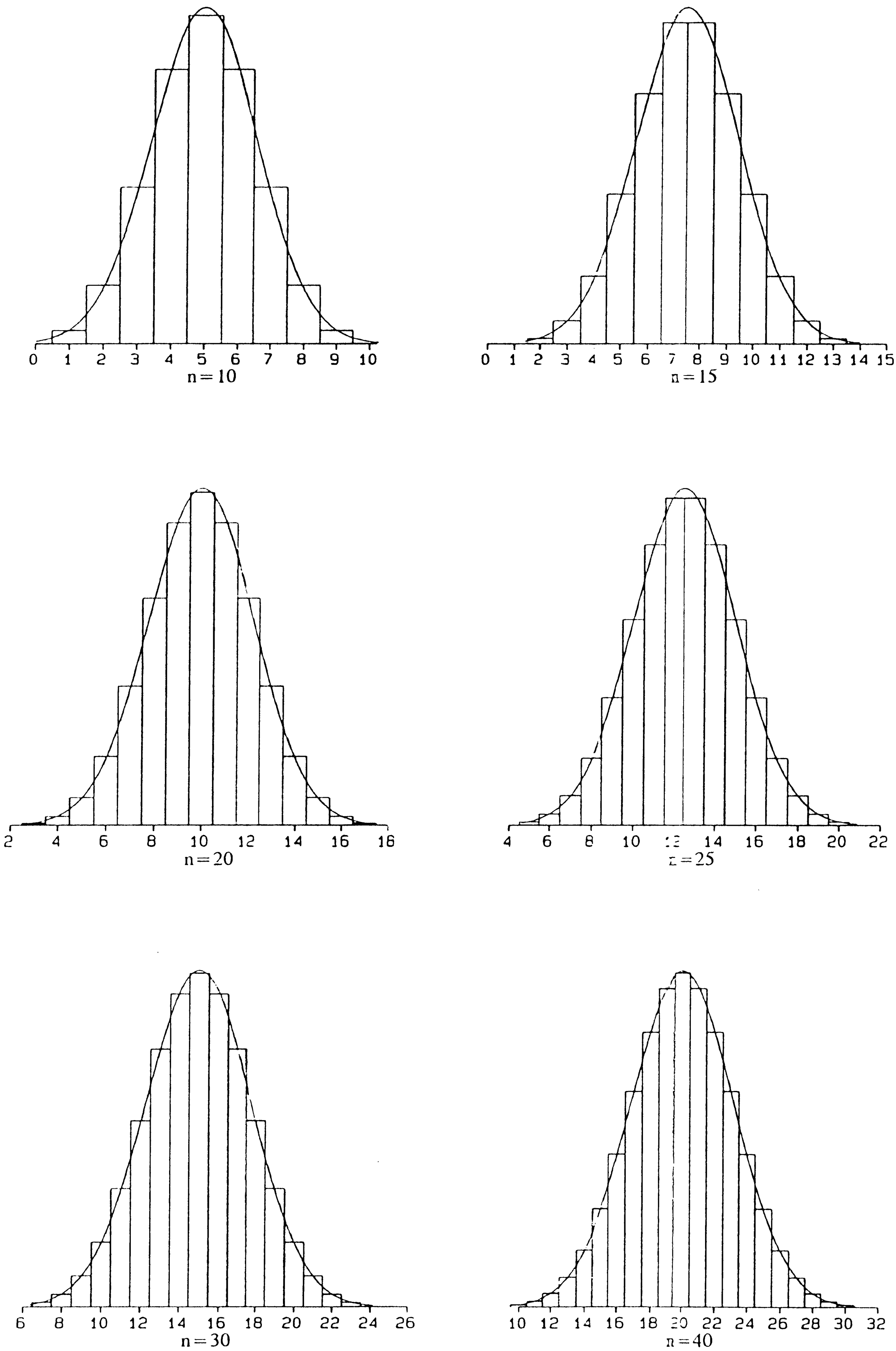
La comparaison de ces deux lois se fera en superposant le graphe de la densité de la loi normale à l'histogramme ainsi construit. L'aire totale sous cette courbe est aussi égale à 1. Les probabilités associées à une valeur de X sont d'une part l'aire du rectangle correspondant, et pour l'approximation normale d'autre part, l'aire de la surface sous la courbe densité ayant pour base l'intervalle considéré.

Afin de visualiser l'évolution de l'écart entre la loi binomiale et la loi normale, lorsque la taille de l'échantillon augmente, à valeur de p constante, les graphiques suivants sont tracés par rapport à la variable centrée réduite ; cela élimine l'effet taille.

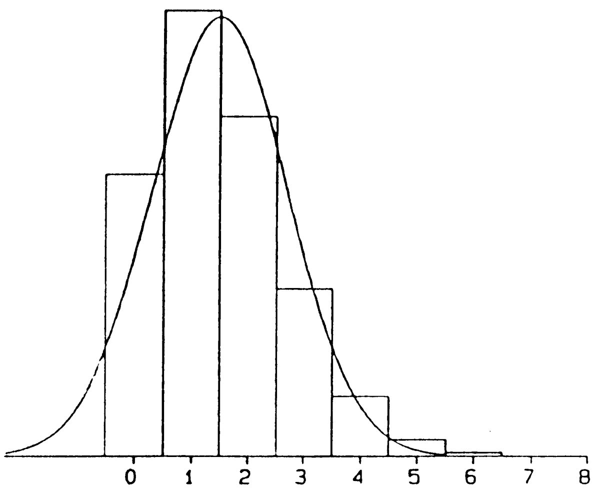
Pour caractériser le degré d'approximation, on a indiqué l'écart maximum entre les probabilités pour les valeurs de X et celles correspondantes de U, ainsi que la valeur de X où ce maximum est atteint. Cependant, pour le calcul approché de la fonction de répartition de X (les probabilités cumulées) à l'aide de celle de U, l'écart est plus grand pour certaines valeurs que l'écart maximum sur les probabilités individuelles.

On voit ainsi que l'approximation est très bonne lorsque la loi est symétrique ($p=1/2$), même pour de petites valeurs de n ; alors que pour une loi fortement dissymétrique (p petit ou p voisin de 1), la taille doit être grande afin d'obtenir le même degré d'approximation.

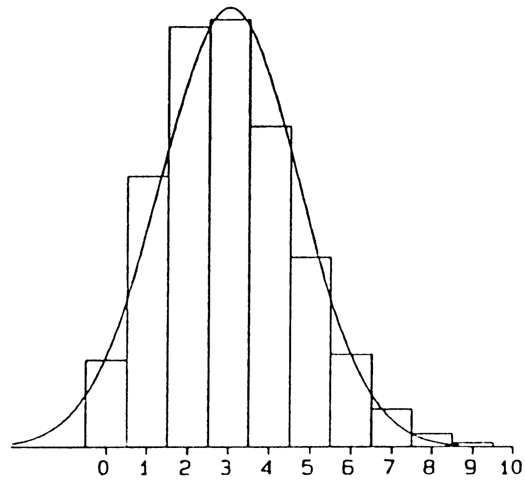
La convergence (en loi) d'une v.a. suivant une loi de Poisson $\mathcal{P}(\lambda)$, centrée réduite, vers la loi normale quand $\lambda \to \infty$ se visualise de la même façon en superposant l'histogramme correspondant à la loi de Poisson et le graphe de la densité de la loi normale.



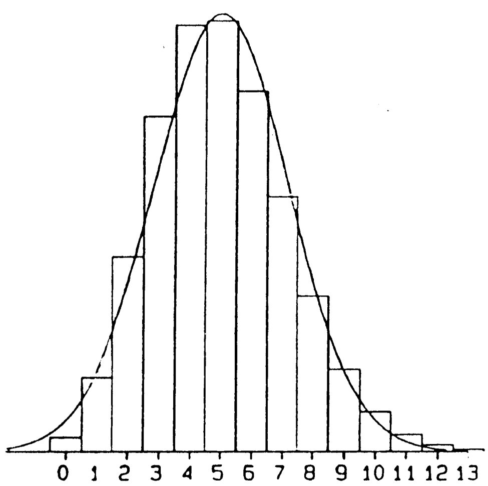
Approximation Binomiale-Normale p=0.5



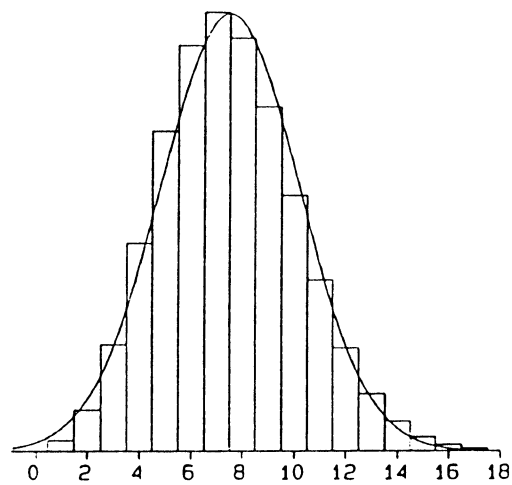
n=30, ecart max=6.05% pour k=0



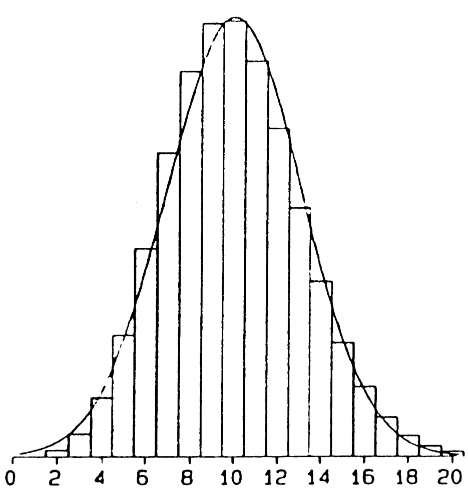
n=60, ecart max=2.95% pour k=2



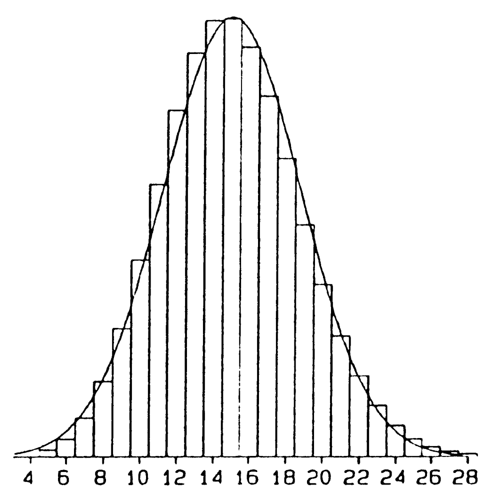
n=100, ecart max=1.96% pour k=3



n=150, ecart max=1.25% pour k=5

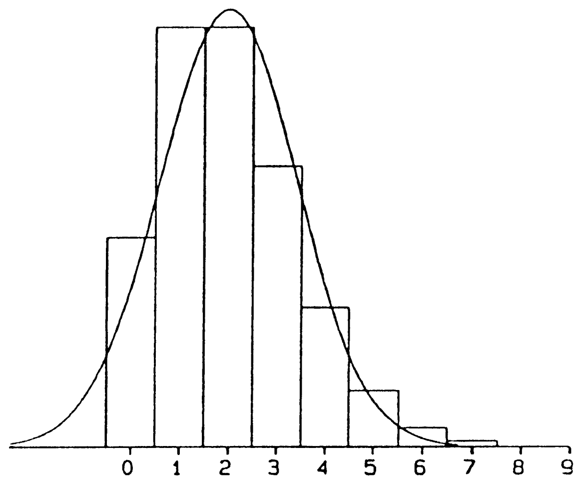


n=200, ecart max=0.91% pour k=8

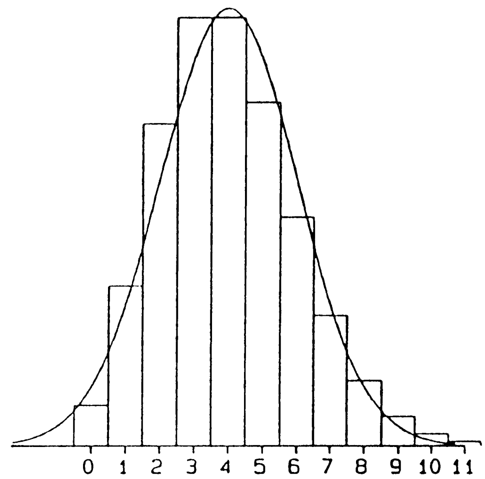


n=300, ecart max=0.63% pour k=12

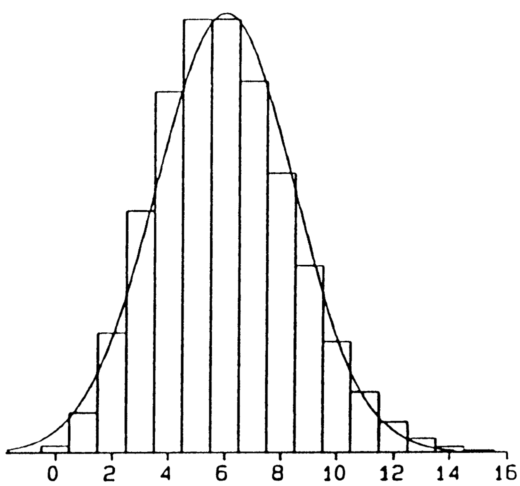
Approximation Binomiale-Normale : p=0.05



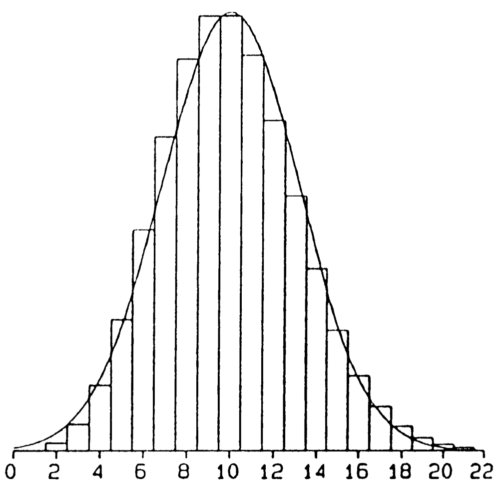
mu =2.0, ecart max=5.33% pour k=1



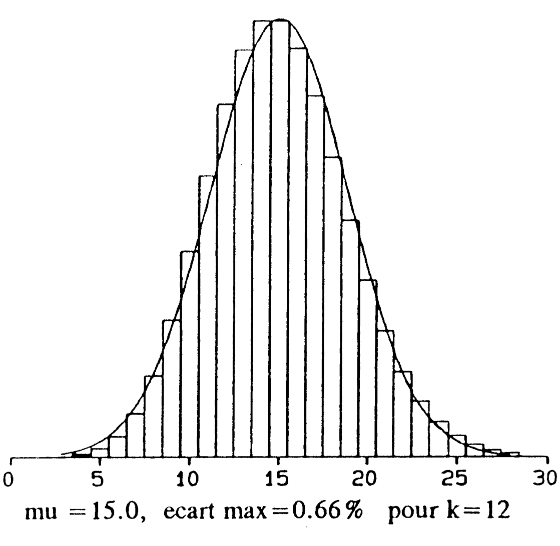
mu =4.0, ecart max=2.55% pour k=2



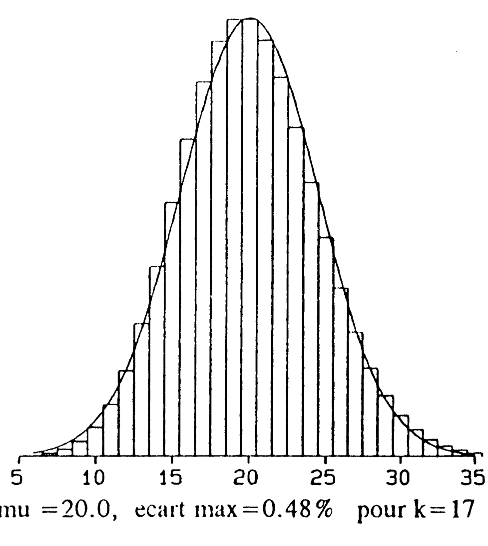
mu =6.0, ecart max=1.74% pour k=4



mu =10.0, ecart max=0.97% pour k=7



mu =15.0, ecart max=0.66% pour k=12



mu =20.0, ecart max=0.48% pour k=17

Approximation Poisson-Normale