



Statistiques Loi normale

- CCI Campus
- Philippe Schlegel

Loi normale :

- Loi uniforme
 - Espérance, variance et écart type de la loi uniforme
 - Loi normale
 - Approximation d'une loi binomiale par une loi normale
 - Théorème de la limite centrée
 - Cas concrets
 - Questions
-

Loi uniforme :

- Les lois uniformes continues forment une famille de probabilités à densité caractérisées par le fait que tous les intervalles de même longueur inclus dans le support de la loi ont la même probabilité.
- La densité de probabilité de la loi uniforme continue est une fonction qui porte sur l'intervalle $[a; b]$:
 - $F(x) =$
 - $1/(b-a)$ pour $a \leq x \leq b$
 - 0 sinon

Loi uniforme – densité de probabilité :

- Nous allons tracer la densité de probabilité représentée par une fonction qui porte sur l'intervalle $[2; 5]$.
 - Commençons par tracer un repère orthonormé et plaçons sur l'axe des abscisses les valeurs 2 et 5.
 - Nous allons tracer un rectangle d'aire 1 (100% des individus) dont l'un des côtés est le segment de l'axe des abscisses entre 2 et 5.
 - Déterminer la valeur k , qui sera l'ordonnée des 2 autres points du rectangle.
 - Vérifier que $k = 1/b-a$.
-

Loi uniforme – densité de probabilité :

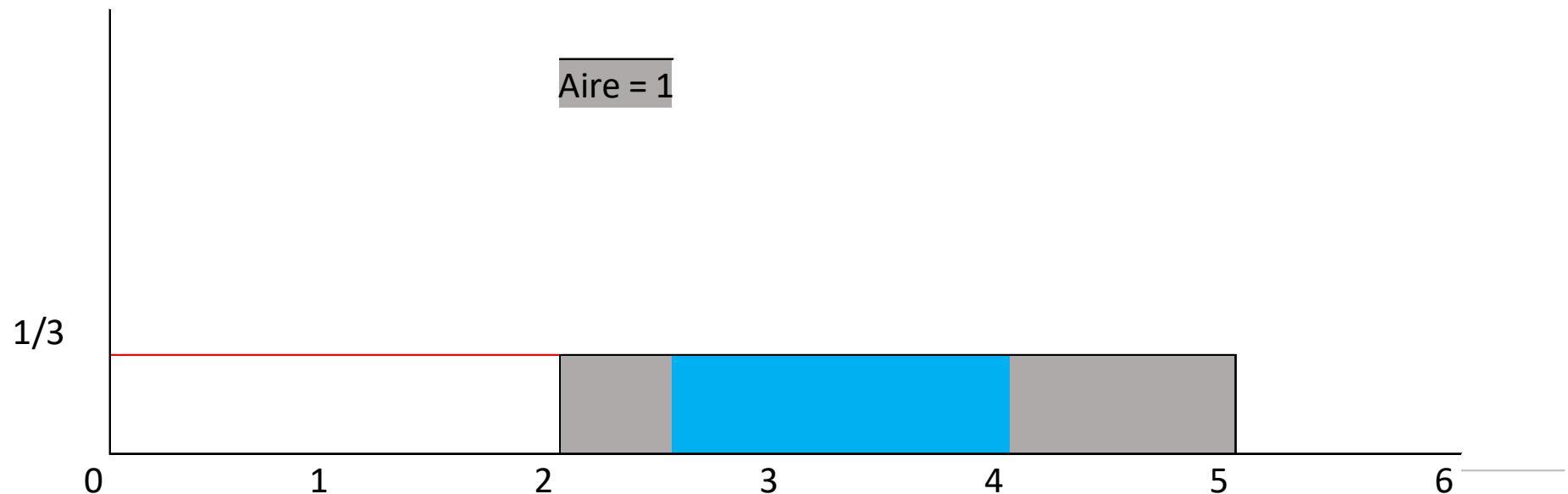
- Cette construction met bien en évidence les propriétés de la loi uniforme :



- La forme de rectangle témoigne bien que chaque individu de l'intervalle a bien la même chance de se produire.

Loi uniforme – Calcul de probabilité :

- Maintenant que nous avons compris comment se définit la loi uniforme, nous allons nous intéresser à des calculs de probabilités.
- Je voudrais calculer $P(2,5 \leq x \leq 4)$ autrement dit quelle est la probabilité que x soit compris entre 2,5 et 4.
- Il suffit pour cela de calculer l'aire du rectangle ci-dessous :

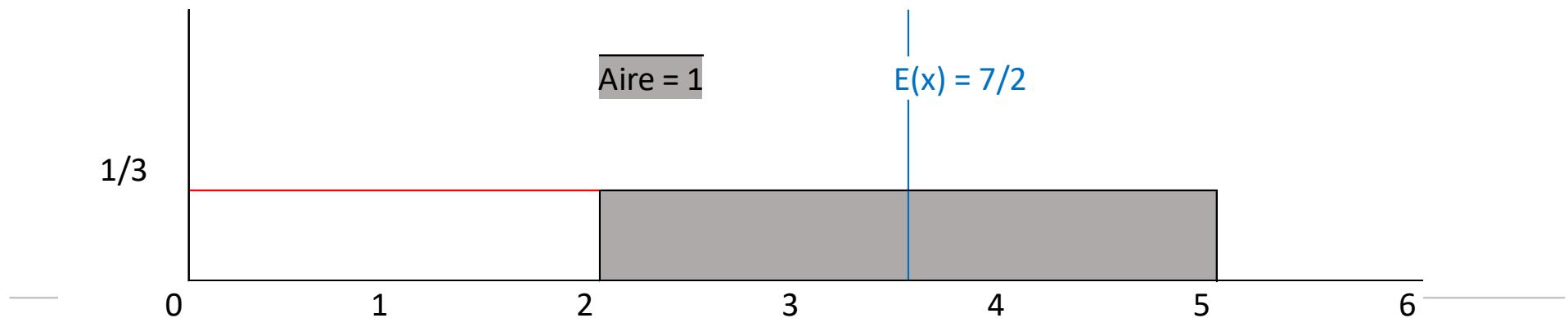


Loi uniforme – Calcul de probabilité :

- Pour calculer l'aire, il y a plusieurs possibilités :
 - Soit on utilise l'intégrale : $\int_{2,5}^4 f(x)dx = \int_{2,5}^4 \frac{1}{3} dx$
 - Soit on calcule simplement l'aire du rectangle puisque nous connaissons sa hauteur ($1/3$) et sa longueur ($4 - 2,5$)
 - Dans les deux cas, nous arrivons à la conclusion que $P(2,5 \leq x \leq 4) = 1/3 * 1,5 = 0,5$
 - On peut noter qu'il s'agit également du ratio entre la longueur du rectangle de base et celle de la probabilité :
 - $(5 - 2) * 0,5 = (4 - 2,5)$. Ceci est normal puisque la hauteur du rectangle est la même dans les deux cas.
 - On peut donc généraliser : $P(c \leq x \leq d) = (d-c)/(b-a)$
-

Loi uniforme – Espérance :

- Comme dans toutes les loi de probabilité, l'espérance est la valeur que l'on s'attend à trouver en moyenne.
- Cette définition prend tout son sens ici puisque le calcul de l'espérance est équivalent à la moyenne de l'intervalle.
 - Pour une loi uniforme $U([a;b])$ l'espérance $E(X) = (a+b) / 2$
- Calculons donc l'espérance dans notre loi uniforme $U([2;5])$
- Graphiquement l'espérance est la valeur qui divise le rectangle représentant notre loi uniforme en 2 :



Loi uniforme – Calcul de l'espérance :

- Pour une loi à valeurs discrètes, nous avons vu que nous faisions la sommes des produits valeur * probabilité ($x_i * p_i$).
- Ici nous avons des valeurs continues mais le principe reste le même, nous allons juste utiliser une intégrale au lieu d'une somme :

- $E(X) = \int_a^b x \cdot f(x) dx$

- $f(x)$ est une fonction constante $\frac{1}{b-a}$ et ne contient pas x :

- $$E(X) = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}$$

Loi uniforme – Variance :

- La variance exprime la dispersion des valeurs d'un échantillon.
 - $V(X) = (b-a)^2/12$
- La formule de la variance correspond à sa définition. Elle provient de la simplification de l'expression :
 - $V(X) = \int_a^b x^2 f(x) dx - (E(x))^2$
 - $f(x)$ est une fonction constante $\frac{1}{b-a}$ et ne contient pas x :
 - $V(X) = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b - \left(\frac{a+b}{2} \right)^2 = \frac{1}{b-a} \frac{b^3 - a^3}{3} - \frac{(a+b)^2}{4}$
 - $V(X) = \frac{1}{b-a} \frac{(b-a)(a^2+ab+b^2)}{3} - \frac{(a+b)^2}{4}$
 - $V(X) = \frac{(a^2+ab+b^2)}{3} - \frac{(a^2+2ab+b^2)}{4} = \frac{(a^2-2ab+b^2)}{12} = \frac{(b-a)^2}{12}$

Loi uniforme – Ecart type :

- L'écart-type se calcule comme d'habitude en prenant la racine carré de la variance :

$$\bullet \quad \sigma = \sqrt{V(X)} = \frac{\sqrt{(b-a)^2}}{\sqrt{12}} = \frac{b-a}{2\sqrt{3}}$$

- Terminons donc cette exemple en calculant la variance et l'écart type de notre exemple de loi uniforme $U([2;5])$

$$\bullet \quad V(X) = \frac{(b-a)^2}{12} = \frac{(5-2)^2}{12} = \frac{9}{12} = \frac{3}{4} = 0,75$$

$$\bullet \quad \sigma = \frac{b-a}{2\sqrt{3}} = \frac{5-2}{2\sqrt{3}} = \frac{3}{2\sqrt{3}} = 0,866 = \sqrt{V(X)}$$

Loi normale – Définition :

- La loi normale est une des lois de probabilités les plus utilisée pour modéliser les phénomènes naturelles issues de plusieurs événements aléatoires.
- Elle est également appelée loi gaussienne, loi de Gauss ou loi de Laplace (1749-1827) – Gauss (1777-1855).
- La loi normale est une loi de probabilité absolument continue qui dépend :
 - De son espérance : μ
 - De son écart-type : σ
- Densité de probabilité : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$

Loi normale – Intervalles :

- La loi normale s'intéresse à des valeurs continues ($P(a < x < b)$) alors que la loi binomiale s'intéresse à des valeurs discrètes ($P(X=2)$).
 - Elle est donc préférable lorsque les problèmes deviennent plus compliqués.
 - Elle travaille sur des intervalles plutôt que sur des valeurs discrètes.
 - En loi normale : $P(X=k) = 0$ on a donc égalité entre
 - $P(x < a) = P(x \leq a)$
 - $P(x > a) = P(x \geq a)$
-

Loi normale – Représentation :

- Nous allons représenter une loi normale sous Excel.
 - Commençons par donner des valeurs de x de 0 à 10 par intervalle de $1/10^{\text{ième}}$
 - Utilisons la fonction Excel loi.normale et étudions ses paramètres
 - X : la valeur du paramètre
 - Espérance : sommet de la courbe
 - Ecart-type : Dispersion de la courbe
 - Cumulative : Est-ce que la courbe doit être cumulative ?
 - Modifions ces paramètres et voyons leur influence sur la courbe et notons l'aspect symétrique de la courbe.
-

Loi normale – Représentation :

- La probabilité de la densité de la courbe est de 1, c'est la surface sous la courbe.
 - Comme la courbe est symétrique nous savons que :
 - $P(x < \text{espérance}) = P(x > \text{espérance}) = 0,5$
 - Ceci va nous aider à généraliser pour n'importe quelle valeur :
 - $P(x < a) = 1 - P(x > a)$
 - $P(x > a) = 1 - P(x < a)$
 - Reprenons la courbe pour comprendre ces égalités.
 - Représentions également sur la courbe $P(a < x < b)$.
 - Voyons que $P(a < x < b) = 1 - P(x < a) - P(x > b)$
-

Loi normale – Exemples :

- La loi normale permet de représenter des événements qui ont de fortes chances de s'approcher de l'espérance et des chances réduites de s'en éloigner significativement :
 - Marche au hasard sur les 4 points cardinaux
 - 1000 lancés de 100 pièces avec comptage des piles et des faces
 - Usure d'un pneu
 - Taille des individus
 - Poids d'une balle de tennis
 - Dimension d'une pièce en sortie d'usine

Loi normale – Espérance, variance, écart type :

- La loi normale est définie par son espérance et son écart type.
- Ces deux informations sont donc initialement données
- Pour la variance, on sait qu'il s'agit du carré de l'écart type
 - $V(X) = \sigma(x)^2$
- Représenter sous Excel la loi normale : N(10,2)

Loi centrée réduite :

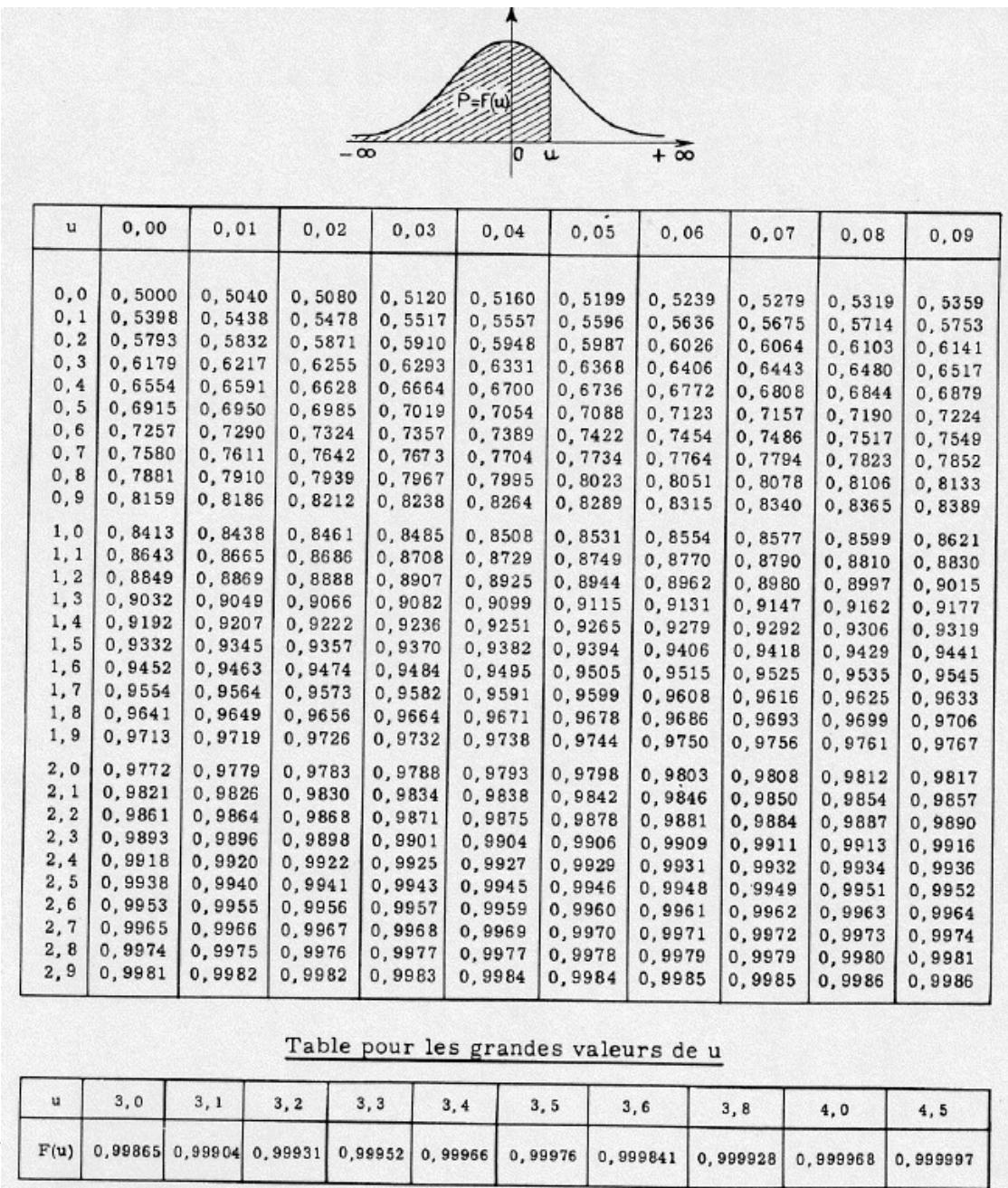
- La loi centrée réduite ou loi normale standard est une loi normale particulière : $N(0;1)$
- Elle est donc centrée sur 0 et possède un écart-type de 1.
- Densité de probabilité : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$
- La loi centrée réduite définit un standard et on peut ramener tous les cas de figure vers la loi centrée réduite à l'aide d'un changement de variable.
- Le but est de simplifier les calculs pour revenir au problème initial par la suite.

Loi centrée réduite - propriétés :

- Mettons à jour quelques propriétés de la loi centrée réduite :
 - $P(x < 0) = p(x > 0) = 0,5$
 - $P(x < -1) = p(x > 1)$
 $= 1 - P(x > -1) = 1 - P(x < 1)$
 $= 0,5 - P(-1 < x < 0) = 0,5 - P(0 < x < 1)$
- Voyons comment interpréter toutes ces égalités
- Cela nous permettra de nous familiariser d'avantage avec les propriétés de la loi normale centrée réduite.
- Il existe ensuite une table de la loi normale centrée réduite qui va nous aider à solutionner nos problèmes.

Lecture de la table de loi normale :

- La table de la loi normale centrée réduite va nous permettre de faire des approximations des probabilités.
- La première colonne va nous donner les dixièmes et les entêtes de colonne représentent les centièmes.
- Recherchons $P(X < 0,47)$ et $P(X < 1,72)$
- Calculer x sachant que la prob. 0,8023



Lecture de la table de loi normale :

- L'aspect symétrique de la loi normale va nous permettre de trouver des équivalences pour nous ramener à un cas du tableau
- Recherchons $P(X < - 0,47)$ et $P(X < - 1,72)$
- Nous voyons qu'au delà de 3, 99% des cas sont couverts et on peut considérer qu'au-delà de 4, 100% des cas sont couverts par arrondi.

Ramener un problème de loi normale en un problème de loi normale centrée réduite :

- L'idée est de ramener une loi $N(E, \sigma)$ vers $N(0, 1)$
- Pour une probabilité de x sur N : $P(x < 2)$ par exemple il faut :
 - Ramener l'espérance à zéro : $P(x < 2) \Rightarrow P(x - E < 2 - E)$
 - Ramener l'écart type à 1 : $P(x - E < 2 - E) \Rightarrow P\left(\frac{x - E}{\sigma} < \frac{2 - E}{\sigma}\right)$
- Au terme de ces deux opérations, nous aurons donc transposé notre loi normale en loi normale centrée réduite et nous pourrons utiliser la table vue précédemment.

Loi normale par l'exemple :

- Considérons un élevage de poulet. Après 110 jours, un poulet a un poids moyen de 1500g avec un écart-type de 200g.
- La loi normale concernant le poids des poulets est donc $N(1500, 200)$. Les poulets bas de gamme sont ceux qui ont moins de 1300g, les moyens de gamme ont entre 1300 et 1750g, les poulets haut de gamme ont plus de 1750g.
- Quel est la probabilité lorsque je prends un poulet au hasard d'avoir un poulet :
 - Bas de gamme $P(X < 1300)$
 - Haut de gamme $P(1750 < X)$
 - Moyen de gamme $P(1300 < X < 1750)$

Loi normale par l'exemple :

- J'arrive au marché avec mes poulets et j'en vends 450 avec les répartitions vues ci-dessus. Quel est mon chiffre d'affaire sachant que :
 - Un poulet haut de gamme coûte 12 euros
 - Un poulet moyen coûte 10 euros
 - Un poulet bas de gamme coûte 8 euros.
- Nous pouvons résoudre cette question avec un tableau Excel ce qui va nous permettre de voir comment en modifiant les prix de chaque catégorie on peut influer sur le Chiffre d'affaire.
- Donner le poids minimum des 5% de poulets les plus lourds.

Approximation de la loi binomiale par la loi normale:

- La loi binomiale s'appuie sur des valeurs discrètes.
- La loi normale s'appuie sur des variables à valeur continue.
- Nous pouvons avoir des cas où les propriétés de la loi binomiale ne sont pas suffisantes.
- Exemple : Vous travaillez pour une entreprise qui produit des guitares et vous êtes responsable de la qualité. Vous savez que 30% des guitares produites ont un défaut. Vous décidez de porter votre analyse sur un échantillon de 100 guitares.
 - A l'aide des propriétés de la loi binomiale, il vous est possible de calculer facilement $P(X = 35) = C_n^k * p^k * q^{n-k}$

Approximation de la loi binomiale par la loi normale:

- Les choses se compliquent s'il on cherche à déterminer $P(X \leq 35) = P(x=0) + P(x=1) + \dots + P(x=35)$
 - On va donc se baser sur la loi normale pour résoudre ce genre de problème.
 - Une très bonne approximation de la loi binomiale $B(n, p)$ est la loi normale $N(np, \sqrt{np(1 - p)})$ lorsque:
 - $n \geq 30$
 - $np \geq 5$
 - $n(1-p) \geq 5$
 - Autrement dit n doit être assez grand et p ne doit pas être trop proche de 0 ou 1.
-

Approximation de la loi binomiale par la loi normale:

- Notre loi binomiale $B(100, 0,3)$ porte sur 100 guitare avec 0,3 de probabilité de défaut, vérifions si les conditions sont remplies :
 - $N \geq 30$: $n = 100$ OK
 - $Np \geq 5$: $n * p = 100 * 0,3 = 30$ OK
 - $N(1-p) \geq 5$: $n * q = 100 * 0,7 = 70$ OK
- Il est donc possible d'approcher la loi binomiale $B(100, 0,3)$ grâce à la loi normale $N(E(x), \sigma)$ avec
 - $E(x) = n * p = 100 * 0,3 = 30$
 - $\sigma = \sqrt{(n * p * q)} = \sqrt{100 * 0,3 * 0,7}$
- Donc $N(30, 4,58)$

Approximation de la loi binomiale par la loi normale:

- Nous pouvons comparer ces deux lois ($B(100, 0,3)$ et $N(30, 4,58)$) sous Excel pour se rendre compte de l'approximation.
- Revenons à notre exemple. En loi normale $P(X=35) = 0$, si je veux une approximation, il faut utiliser un intervalle :
 - $P(34,5 < x < 35,5) = p(x < 35,5) - p(x < 34,5)$ (Tracer une courbe en cas de doute).
- Je reviens à la loi centrée réduite :
 - $p(x < (35,5 - 30)/4,58) - p(x < (34,5-30)/4,58)$
 - $p(x < 1,2) - p(x < 0,98)$
 - $0,8849 - 0,8365 = 0,0484$ (4,84%) (par lecture dans la table)
- Il y a 4,84% de chance de tomber sur 35 guitares défectueuses.

Approximation de la loi binomiale par la loi normale:

- Cherchons maintenant une approximation pour $P(X \leq 35)$.
- Si je veux que 35 soit dans l'intervalle considéré, il faut que je modifie ma probabilité en $P(X < 35,5)$.
- On repart sur la loi centrée réduite
 - $P(X < (35,5 - 30) / 4,58) = 0,8849$ (nous l'avons calculé au dessus).
- Nous avons donc 88,49% de chance d'obtenir au maximum 35 guitares défectueuses sur notre échantillon.
- Pour $P(X \geq 35)$ soit on refait le calcul, soit on prend le complément $P(x \geq 35) = 1 - p(x \leq 35) = 1 - 0,8849 = 0,1151$
- Nous avons donc 11,51% de chance d'obtenir au minimum 35 guitares défectueuses sur notre échantillon.

Théorème central limite :

- Le théorème central limite établit la convergence en loi de la somme d'une suite de variables aléatoires vers la loi normale.
- Cette suite de variables doit répondre à 2 critères :
 - Les variables doivent venir de la même loi
 - Elles doivent être indépendantes entre elles.
- Cela permettrait de comprendre pourquoi la loi normale est omniprésente dans la nature.
- Il s'agit en fait de l'addition de nombreux petits phénomènes aléatoire qui conduise à la représentation gaussienne.

Théorème central limite par l'exemple :

- Modélisons quelques exemples pour voir si cela marche à tous les coups :
 - Simulons un petit jeu avec une pièce : pile je gagne un euro, face je perds un euro. Je cumule les gains et les pertes sur plusieurs lancés et je répète l'expérience.
 - Est-ce que cela change quelque chose si les gains ne sont pas équilibrés ?
 - Je lance un dé 6 et je cumule les valeurs obtenues sur plusieurs lancés et je répète l'expérience. J'observe alors comment cette somme est répartie
 - Je reprends l'expérience précédente avec la moyenne.

Théorème central limite :

- Nous nous apercevons qu'en augmentant les expériences, nous nous approchons d'une distribution normale.
 - Plus la taille de l'échantillon choisi est grande, plus nous serons sur une forme normale.
 - Si la taille de l'échantillon tend vers $+\infty$
 - Nous avons choisi de nous pencher sur la moyenne mais nous aurions eu le même résultat avec la somme des nombres ou le produit des nombres.
 - Grâce au théorème de la limite centrée, on peut donc se rapprocher d'un cas connu alors que le problème semble énoncé de façon complètement aléatoire.
-

Théorème central limite :

- C'est finalement la répétition qui fait qu'une variable aléatoire qui ne suit pas forcément une loi normale devient étudiable.
- Il est possible de cumuler plusieurs événements
 - Exemple : Lancer un dé deux fois de suite et faire la somme des résultats obtenus.
- Pour que cela soit possible, il faut que les variables aléatoires obtenues soient :
 - Indépendantes
 - De même loi

Bilan

- Résumé des notions abordées.
- Questions.
- Difficultés particulières ?