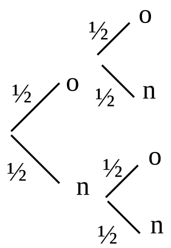
XVI. Variables aléatoires - lois de probabilités

1. Exemple introductif

Le pouvoir d'une minorité résolue
(tiré de A. Engel, l'enseignement des probabilités et de la statistique, ed. CEDIC 1975)

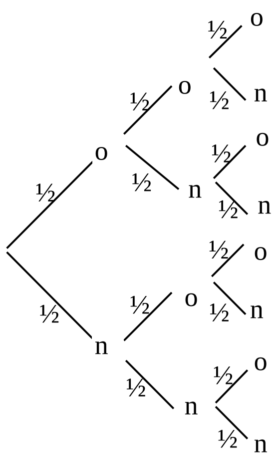
1. Un comité se compose de trois membres. Pierre sait parfaitement ce qu'il veut : il vote oui à la proposition. Les deux autres membres sont indécis et votent au hasard. Quelle est la probabilité que ce soit le oui qui l'emporte ?
2. Un autre comité se compose de cinq membres dont deux ont décidé de s'opposer à une proposition. Les trois autres membres votent au hasard. Quelle est la probabilité de rejet de la proposition ?
3. Dans un pays d'un million de votants, presque tous sont indécis. Deux mille personnes sont opposées à une loi de vote des étrangers. Quelle est la probabilité qu'on rejette cette loi ?

La première question se résout aisément à l'aide du diagramme en arbre :



Pierre gagne si au moins un des autres membres vote oui et donc
P(Pierre gagne) = $\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{3}{4}$

De même pour la seconde question :



Pour que la proposition soit bloquée, il faut qu'au moins une personne s'oppose à la proposition. Et donc :
P(proposition bloquée) = 7 . $\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{7}{8}$

Il nous serait bien difficile d'employer la même méthode pour résoudre la troisième question. C'est pourquoi, nous allons progressivement, tout au long de ce chapitre, étudier des techniques qui vont nous permettre de répondre à ce genre d'interrogation. Notamment, l'idée est de créer une fonction qui "compterait" le nombre de "non" et ensuite, de pouvoir calculer la probabilité pour que cette fonction prenne une valeur précise ou une valeur inférieure à un nombre fixe.

2. Notion de variable aléatoire

Exemple 1:
Considérons l'expérience aléatoire qui consiste à lancer simultanément 3 pièces de monnaie discernables.
La catégorie d'épreuve de cette E.A. est donc Ω = {PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF}
Considérons l'application qui à chaque issue possible de l'épreuve donne pour image le nombre de faces figurant dans cette issue. X est donc une application de Ω dans R. telle que :
X(PPP) = 0 X(PPF) = X(PFP) = X(FPP) = 1 X(PFF) = X(FPF) = X(FFP) = 2
et X(FFF) = 3
L'ensemble des valeurs possibles de X est donc {0, 1, 2, 3}

Exemple 2 :
Considérons l'E.A. qui consiste à lancer simultanément 2 dés discernables.
et sa C.E. : {(1, 1) ; (1, 2) ; (1, 3) ; (1, 4) ; (1, 5) ; (1, 6) ; (2, 1) .....(6, 6)}
Considérons l'application qui à chaque issue possible de l'épreuve donne pour image la somme des points figurant sur ces 2 dés. Nous aurons :

X (1, 1) = 2 X (1, 2) = X (2, 1) = 3
X (1, 3) = X (2, 2) = X (3, 1) = 4 X (1, 4) = X (2, 3) = X (3, 2) = X (4, 1) =5
X (1, 5) = X (2, 4) = X (3, 3) = X (4, 2) = X (5, 1) = 6
X (1, 6) = X (2, 5) = X (3, 4) = X (4, 3) = X (5, 2) = X (6, 1) = 7
X (2, 6) = X (3, 5) = X (4, 4) = X (5, 3) = X (6, 2) = 8
X (3, 6) = X (4, 5) = X (5, 4) = X (6, 3) = 9 X (4, 6) = X (5, 5) = X (6, 4) = 10
X (5, 6) = X (6, 5) = 11 X (6, 6) = 12

L'ensemble des valeurs possibles de X est {2, 3, 4, ...12}

Définition
Une variable aléatoire liée à une épreuve aléatoire est une application de la catégorie d'épreuve de cette expérience aléatoire dans R
$X : \Omega \to R \qquad A \to X(A)$

Dans les deux exemples précédents, l'ensemble des valeurs possibles est un ensemble fini. Dans ce cas, la variable aléatoire est une variable aléatoire discrète. On emploie également cette appellation pour une variable aléatoire dont l'ensemble des valeurs possibles est un ensemble infini dénombrable.
Exemple :
Considérons comme épreuve une suite indéfiniment prolongée de parties de pile ou face, et comme variable aléatoire, celle qui à tout résultat possible de l'épreuve fait correspondre le nombre de faces obtenu. L'ensemble des valeurs de cette variable aléatoire est donc N. C'est également une variable aléatoire discrète.

3. Distribution de probabilité d'une variable aléatoire discrète.

3.1 Exemple 1

A partir de l'exemple 1 du point 1 , considérons les ensembles suivants :
E_1= {PPP} E_2 ={PPF , PFP , FPP} E_3 ={PFF , FPF , FFP} E_4 = {FFF}
Nous avons :
$\forall\ x \in E_1 : X(x) = 0 \qquad \forall\ x \in E_2 : X(x) = 1 \qquad \forall\ x \in E_3 : X(x) = 2 \qquad \forall\ x \in E_4 : X(x) = 3$
Et $P(E_1) = \frac{1}{8} = P(E_4) \qquad P(E_2) = P(E_3) = \frac{3}{8}$
Ces valeurs sont également les probabilités que la variable aléatoire prenne la valeur 0, 3, 1 ou 2
$P(X = 0) = P(X = 3) = \frac{1}{8}$ et $P(X = 1) = P(X = 2) = \frac{3}{8}$
En présentant ces résultats en tableau : nous obtenons :

x_i	0	1	2	3
$P(X = x_i)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

avec : $\sum_i P(X = x_i) = 1$

3.2 Exemple 2

E_1 ={(1, 1)} E_2 = {(1, 2) ; (2, 1)}
E_3 = {(1, 3) , (2, 2) , (3, 1)} E_4 = {(1, 4) , (2, 3) , (3, 2) , (4, 1)}
E_5 = {(1, 5) , (2, 4) , (3, 3) , (4, 2) , (5, 1)} E_6 = {(1, 6) , (2, 5) , (3, 4) , (4, 3) , (5, 2) , (6, 1)}
E_7 = {(2, 6) , (3, 5) , (4, 4) , (5, 3) , (6, 2)} E_8 = {(3, 6) , X (4, 5) , (5, 4) , (6, 3)}
E_9 = {(4, 6) , (5, 5) , (6, 4)} E_{10} = {(5, 6) , (6, 5)} E_{11} = {(6, 6)}
Et $P(E_1) = P(E_{11}) = \frac{1}{36} \qquad P(E_2) = P(E_{10}) = \frac{2}{36} \qquad P(E_3) = P(E_9) = \frac{3}{36}$
$P(E_4) = P(E_8) = \frac{4}{36} \qquad P(E_5) = P(E_7) = \frac{5}{36} \qquad P(E_6) = \frac{6}{36}$
Ces valeurs sont également les probabilités que la variable aléatoire prenne la valeur 2, 3, 4, ..., 9, 10, 11 ou 12
$P(X = 2) = P(X = 12) = \frac{1}{36} \qquad P(X = 3) = P(X = 11) = \frac{2}{36}$

$P(X = 4) = P(X = 10) = \frac{3}{36}$ $P(X = 5) = P(X = 9) = \frac{4}{36}$

$P(X = 6) = P(X = 8) = \frac{5}{36}$ $P(X = 7) = \frac{6}{36}$

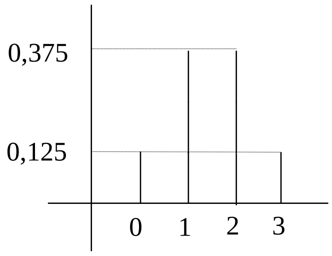
x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X = x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

<u>En général :</u>

Soit $(\Omega, \mathscr{P}(\Omega), p)$ un espace probabilisé
X une variable aléatoire
et $\{x_1, x_2, ..x_n, ...\}$ l'ensemble des valeurs possibles de cette variable aléatoire
Soit $E_k \subset \Omega : \forall x \in E_k : X(x) = x_k$: à chaque X_k est donc associé un sous-ensemble de Ω
Notons la probabilité de E_k par $P(E_k) = p_k$
C'est également la probabilité de la valeur x_k de X : nous écrivons : $p_k = P(X = x_k)$
La fonction définie par $f(x_i) = P(X = x_i)$ est la distribution ou loi de probabilité de X, donnée habituellement sous la forme d'un tableau comme dans les exemples proposés.

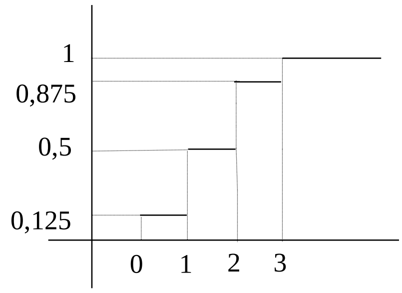
3.3 Représentation

Sur un graphique, nous portons en abscisse les valeurs possibles de la variable aléatoire et en ordonnée leurs probabilités respectives.
Pour le premier exemple traité ci-dessus, nous obtenons le graphique ci-contre.
La représentation de la distribution de probabilité du second exemple se fait de façon similaire.
Remarquons la ressemblance entre *distribution de probabilité* d'une variable aléatoire et *distribution des fréquences* d'une variable statistique.



4. Fonction de répartition d'une variable aléatoire discrète

Etant donné une variable aléatoire X, on définit la fonction de répartition F de X par la fonction
$F : R \to R : F(a) = P(X \le a)$
Nous avons donc $F(a) = \sum_{x_i \le a} P(X = x_i)$
En reprenant le premier exemple traité ci-dessus, nous obtenons le graphique suivant :
(Le second cas est également obtenu de manière similaire)
Comme la distribution de probabilité correspond au diagramme des fréquences en statistiques, la fonction de répartition correspond au diagramme cumulatif.



5. Espérance mathématique d'une variable aléatoire discrète.

<u>Exemple 1</u>
On jette trois pièces de monnaie bien équilibrées. Combien de faces obtient-on en moyenne ?
Cette question prend tout son sens lorsqu'on réalise un grand nombre de fois l'expérience. Supposons qu'on la réalise 1000 fois. On s'attend alors à observer 125 fois le résultat PPP, 125 fois le résultat PPF ....
et donc en moyenne, nous aurons : $0 . \frac{125}{1000} + 1 . \frac{375}{1000} + 2 . \frac{375}{1000} + 3 . \frac{125}{1000} = \frac{1500}{1000}$
ou : $0 . \frac{1}{8} + 1 . \frac{3}{8} + 2 . \frac{3}{8} + 3 . \frac{1}{8} = \frac{12}{8} = \frac{3}{2}$

<u>Exemple 2</u>
Lorsqu'on lance 2 dés, quelle est la somme des points obtenus en moyenne ?
En procédant comme dans le premier exemple, nous obtenons :
$2 . \frac{1}{36} + 3 . \frac{2}{36} + 4 . \frac{3}{36} + 5 . \frac{4}{36} + 6 . \frac{5}{36} + 7 . \frac{6}{36} + 8 . \frac{5}{36} + 9 . \frac{4}{36} + 10 . \frac{3}{36} + 11 . \frac{2}{36} + 12 . \frac{1}{36} = \frac{252}{36} = 7$

En général

L'espérance mathématique d'une variable aléatoire discrète X s'obtient en réalisant la somme des produits de chacune des valeurs possibles de X par la probabilité associée à cette valeur c'est à dire :

$$\boxed{E(X) = \mu_x = \sum_k p_k x_k}$$

Remarquons à nouveau l'analogie avec la recherche de la moyenne en statistique : $\overline{X} = \sum_k f_k x_k$

C'est pourquoi l'espérance mathématique d'une variable aléatoire est également appelée moyenne de celle-ci ou plus rarement valeur du jeu.

6. Variance et écart-type d'une variable aléatoire discrète

La variance d'une variable aléatoire discrète X est définie par : $V(X) = \sigma_x^2 = \sum_k p_k (x_k - \mu_x)^2$

Son écart-type $\sigma_x = \sqrt{V(X)}$ et nous retrouvons à nouveau l'analogie avec la statistique.

6.1 Propriétés

1. $V(X) = \sum_k p_k (x_k - \mu_x)^2 = \sum_k p_k x_k^2 - 2\,\mu_x \sum_k p_k x_k + \sum_k p_k (\mu_x)^2 = \sum_k p_k x_k^2 - 2\,\mu_x.\mu_x + \mu_x^2 = \sum_k p_k x_k^2 - \mu_x^2$
2. Etant donné une v.a. X, on peut en déduire d'autres. Si Y = aX + b, on peut montrer :
   a) E(Y) = a E(X) + b
   b) Var (Y) = a^2 Var (X)

Reprenons nos deux exemples :

6.2 Exemple 1

x_k	0	1	2	3
$x_k-\mu_x$	-1.5	-0.5	0.5	1.5
$(x_k-\mu_x)^2$	2.25	0.25	0.25	2.25
p_k	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$V(X) = \frac{1}{8} . 2.25 + \frac{3}{8} . 0.25 + \frac{3}{8} . 0.25 + \frac{1}{8} . 2.25 = \frac{3}{4}$ et $\sigma_x = \frac{\sqrt{3}}{2}$

6.3 Exemple 2

x_k	2	3	4	5	6	7	8	9	10	11	12
$x_k-\mu_x$	-5	-4	-3	-2	-1	0	1	2	3	4	5
$(x_k-\mu_x)^2$	25	16	9	4	1	0	1	4	9	16	25
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$V(X) = 25 . \frac{1}{36} + 16 . \frac{2}{36} + 9 . \frac{3}{36} + 4 . \frac{4}{36} + 1 . \frac{5}{36} + 0 . \frac{6}{36} + 1 . \frac{5}{36} + 4 . \frac{4}{36} + 9 . \frac{3}{36} + 16 . \frac{2}{36} + 25 . \frac{1}{36} = \frac{210}{36} = \frac{35}{6}$

Et $\sigma_x = \sqrt{\frac{35}{6}}$

6.4 Notations

Paramètre	Echantillon	Population
Moyenne	$\overline{X}$	$E(X) = m = \mu$
Variance	s^2	σ^2
Ecart-type	s	σ
Proportion	Fréquence relative	Probabilité

6.5 Exercices

1. On propose à un joueur de choisir parmi les deux jeux suivants :
   a) On lance un dé bien équilibré. Si le dé montre la face "1" ou "6", le joueur gagne 12 points, sinon il perd 6 points.
   b) On lance le dé. S'il montre 6, le joueur gagne 120 points, sinon, il perd 25 points.
   Quel jeu est le plus intéressant et selon quel critère ?
   Pour répondre à cette question, imaginez que le même jeu est répété un grand nombre de fois.

2. Envisagez les trois jeux repris ci-dessous. Parmi eux, lequel est le plus intéressant pour un joueur et selon quels critères ?
   a) On lance un dé à 6 faces. Si c'est un 6, le joueur empoche 20 €; si ce n'est pas un 6, il perd 1 €.
   b) On lance un dé à 6 faces. Si on obtient 1 ou 6, le joueur empoche 30 €, sinon il doit payer 10 €.
   c) Le joueur tire un billet dans une urne qui en contient 100 dont un rouge et 99 bleus. S'il tire le rouge, la banque lui donne 1 000 000 €, sinon il doit payer 1 000 €.

3. On extrait 4 cartes d'un jeu de 52 cartes.
   a) Le nombre de cartes de cœur ainsi extraites est une variable aléatoire. Quelle est sa distribution de probabilités ? Calculer sa moyenne et son écart-type.
   b) Même question si on considère la variable aléatoire "nombre d'as"

7. Distribution binomiale

7.1 Epreuve à deux issues

Une épreuve à deux issues ou épreuve de Bernouilli est une expérience aléatoire dans laquelle on ne s'intéresse qu'à la réalisation ou la non réalisation d'un événement A. Dans ce type d'épreuves, deux résultats sont possibles. L'un d'eux sera noté succès (S) et l'autre échec (E).

Exemple1 : On lance une pièce de monnaie. Les deux issues possibles sont bien entendu P et F . Dans ce cas : $P(S) = \frac{1}{2}$ et $P(E) = \frac{1}{2}$

Exemple 2 :On lance un dé et on considère les événements :

S : le chiffre obtenu est > 4 $\quad P(S) = \frac{2}{6} = \frac{1}{3}$

E : le chiffre obtenu est ≤ 4 $\quad P(E) = \frac{4}{6} = \frac{2}{3}$

7.2 Répétition d'épreuves à 2 issues

Certaines expériences aléatoires sont constituées par la répétition d'une même épreuve à 2 issues.
Lorsque ces épreuves successives sont indépendantes, une telle expérience est un schéma de Bernouilli.
Nous obtenons alors des suites composées de n termes telles que SSESE…
Dans le cas où n vaut 3 l'ensemble des possibilités sera donc {SSS,SSE,SES,ESS,SEE,ESE, EES,EEE}
Pour le 1er exemple $P(SES) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

Pour le 2ème exemple : $P(SES) = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{27}$

De façon générale, nous noterons P(S) = p et P(E) = q avec p + q = 1
Dans ce cas : $P(SES) = p \cdot q \cdot p = p^2q$

7.3 Variable aléatoire nombre de succès

Lors d'un schéma de Bernouilli, on s'intéresse à la variable aléatoire X prenant pour valeur le nombre de succès obtenus durant les n épreuves. L'ensemble des valeurs possibles de cette variable aléatoire est {0, 1, 2, 3, ...n}
Dans les exemples précédents, si n = 3, l'événement X = 2 = {SSE, SES, ESS}

7.4 Distribution binomiale

Dans un schéma de Bernouilli, calculons la probabilité P(X = k)
Chaque issue donnant X = k a pour probabilité $p^k q^{n-k}$
Le nombre de ces issues est le nombre de permutations de n éléments parmi lesquels on trouve k fois l'élément S et (n - k) fois l'élément E (en effet, il s'agit du nombre de listes du type SESESSE…comprenant k fois S et n – k fois E) c. à d. $P_n^{k,n-k} = \dfrac{n!}{k!(n-k)!} = C_n^k$

Et nous avons donc

$$\boxed{P(X = k) = p_k = C_n^k\ p^k q^{n-k}}$$

Une telle distribution de probabilité est appelée distribution binomiale.

N.B. : $\sum_{k=0}^{n} p_k = (q + p)^n = 1^n = 1$ (par la formule du binôme de Newton)

Ceci explique le nom de cette loi de probabilité : toutes les valeurs des p_k sont les termes du binôme de Newton.

Notation : $\boxed{B_{n,p}(k) = C_n^k\ p^k (1 - p)^{n-k}}$

<u>Exemple</u> : On jette un dé 5 fois de suite, quelle est la probabilité que le chiffre obtenu soit exactement 2 fois strictement supérieur à 4

S : le chiffre est > 4 : P (S) = p = $\frac{1}{3}$

E : le chiffre est ≤ 4 : P (E) = q = $\frac{2}{3}$ n = 5

$P(X = 2) = C_5^2 \left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^3 = \dfrac{5!}{2!3!}\cdot\dfrac{1}{9}\cdot\dfrac{8}{27} = \dfrac{80}{243} \cong 0.329$

7.5 Exercices :

1. Un questionnaire à choix multiple comporte 10 questions. Pour chaque question, 2 réponses sont possibles. Un élève répond au hasard à chacune de ces 10
   a) Quelle est la probabilité qu'il obtienne 5 bonnes réponses ? sol : 0.246
   b) Quelle est la probabilité qu'il n'obtienne pas la moitié des bonnes réponses ? sol : 0.377
   c) Quelle est la probabilité qu'il obtienne une cote strictement supérieure à 7 ? sol : 0.0547
   d) Reprendre les 3 questions précédentes dans le cas où il y a trois réponses proposées pour chaque question.
   Sol : a) 0.1366 b) 0.7869 c) $3.4039 \cdot 10^{-3}$

2. On tire une carte dans un jeu de 52 cartes et on remet ensuite la carte dans le jeu. On répète 20 fois l'opération et on note le nombre de carreaux obtenus. Déterminer le type de distribution de probabilité de cette variable aléatoire ? Déterminer a) P(x = 5) ; b) P(x = 7) ; c) P(x = 10) ; d) P(x ≤ 3) ; e) P(x ≥ 17)
   sol : a) 0.202 b) 0.1124 c) 0.00992 d) 0.225 e) $2.9605 \cdot 10^{-8}$

3. On jette simultanément 2 dés et on note la somme des points. On procède de la sorte 5 fois consécutivement. Quelle est la probabilité que la somme soit strictement supérieure à 10 plus d'une fois (2 fois ou plus)?
   sol : 0.05858

4. On suppose que 1% des pièces fabriquées par une machine sont défectueuses. Calculer la probabilité pour que dans un échantillon de 100 pièces, 3 pièces ou plus soient défectueuses. Solution : 0.08

5. On répond par vrai ou faux à une série de 10 questions. Quelle est la probabilité d'obtenir un résultat d'au moins 5 sur 10 (5 ou plus) en ayant répondu complètement au hasard ? Sol : 0.6230

6. A chaque balle un tireur a une chance sur 3 de toucher la cible. Il tire 6 coups de feu. On demande
   a) quelle est la probabilité qu'il touche exactement 2 fois la cible ? sol : 0.329
   b) Quelle est la probabilité qu'il touche au moins 2 fois la cible ? sol : 0.6488

7. Une machine fabrique des pièces. On accepte qu'en moyenne 2% d'entre elles devront être mises au rebut. On surveille la fabrication en prélevant à intervalles réguliers des échantillons de 20 pièces qui viennent d'être fabriquées et dont on contrôle la qualité. Si toutes les pièces de l'échantillon sont bonnes ou si une seule est mauvaise, on laisse continuer la production. Si on trouve plus d'une pièce mauvaise, on soupçonne la production de fournir plus de 2% de rebuts, on arrête la machine et on procède à une inspection et un réglage.
   a) Calculer la probabilité de demander un réglage alors que la machine n'en a aucun besoin.

b) Calculer la probabilité que cette procédure de contrôle ne décèle pas un incident de production qui aurait fait passer la proportion des pièces défectueuses à une valeur de 5% sol : a) 0.05989 b) 0.7358

7.6 Moyenne, variance et écart-type d'une distribution binomiale

$$E(X) = \sum_{k=0}^{n} k.p_k = \sum_{k=0}^{n} k.C_n^k p^k q^{n-k}$$

$$\text{or } k\ C_n^k = \frac{k.n!}{k!(n-k)!} = n \cdot \frac{(n-1)!}{(k-1)!(n-k)!} = n\ C_{n-1}^{k-1}$$

$$\Rightarrow E(X) = \sum_{k=1}^{n} n.C_{n-1}^{k-1} p^k q^{n-k} = np \sum_{k=1}^{n} C_{n-1}^{k-1} p^{k-1} q^{(n-1)-(k-1)}$$

En posant k - 1 = j : $E(X) = np \sum_{j=0}^{n-1} C_{n-1}^{j} p^j q^{(n-1)-j} = np\,(p+q)^{n-1} = np$

Et nous avons donc obtenu l'espérance mathématique d'une loi binomiale : $E(X) = np$

$$Var(X) = E(X^2) - (np)^2 = \sum_{k=0}^{n} k^2.p_k - n^2 p^2 = \sum_{k=0}^{n} k^2.\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} - n^2 p^2$$

$$= np \sum_{k=1}^{n} k.\frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} - n^2 p^2$$

$$= np \left\{ \sum_{k=1}^{n} (k-1).\frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} + \sum_{k=1}^{n} .\frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1}(1-p)^{n-k} \right\} - n^2p^2$$

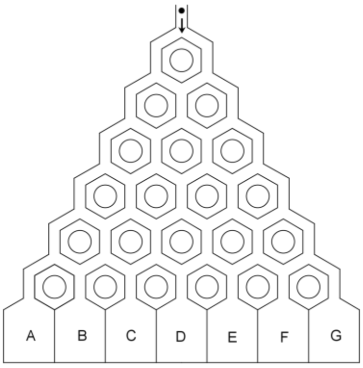
$$= np\,\left\{E(B_{n-1;p}) + (p + (1-p))^{n-1}\right\} = np\,\{(n-1)p + 1\} - n^2 p^2 = n^2 p^2 - np^2 + np - n^2 p^2 = np\,(1-p) = npq$$

<u>Conclusion</u> : $\boxed{X = B_{n,p} \Rightarrow E(X) = np \text{ et } Var(X) = npq}$

7.7 Illustration de la loi binomiale : la planche de Galton

La planche de Galton, imaginée en 1889, permet de visualiser la loi binomiale.

Considérons une planche sur laquelle sont disposés des clous (ou écrous) en quinconce comme sur la figure ci-contre. Au-dessus, se trouve un réservoir comportant un certain nombre de billes. Lorsqu'on ouvre le réservoir, les billes descendent une par une, vers les urnes A, B, C ...Supposons en plus qu'une bille rebondit au-dessus de chaque clou avec la même probabilité vers la gauche ou vers la droite.

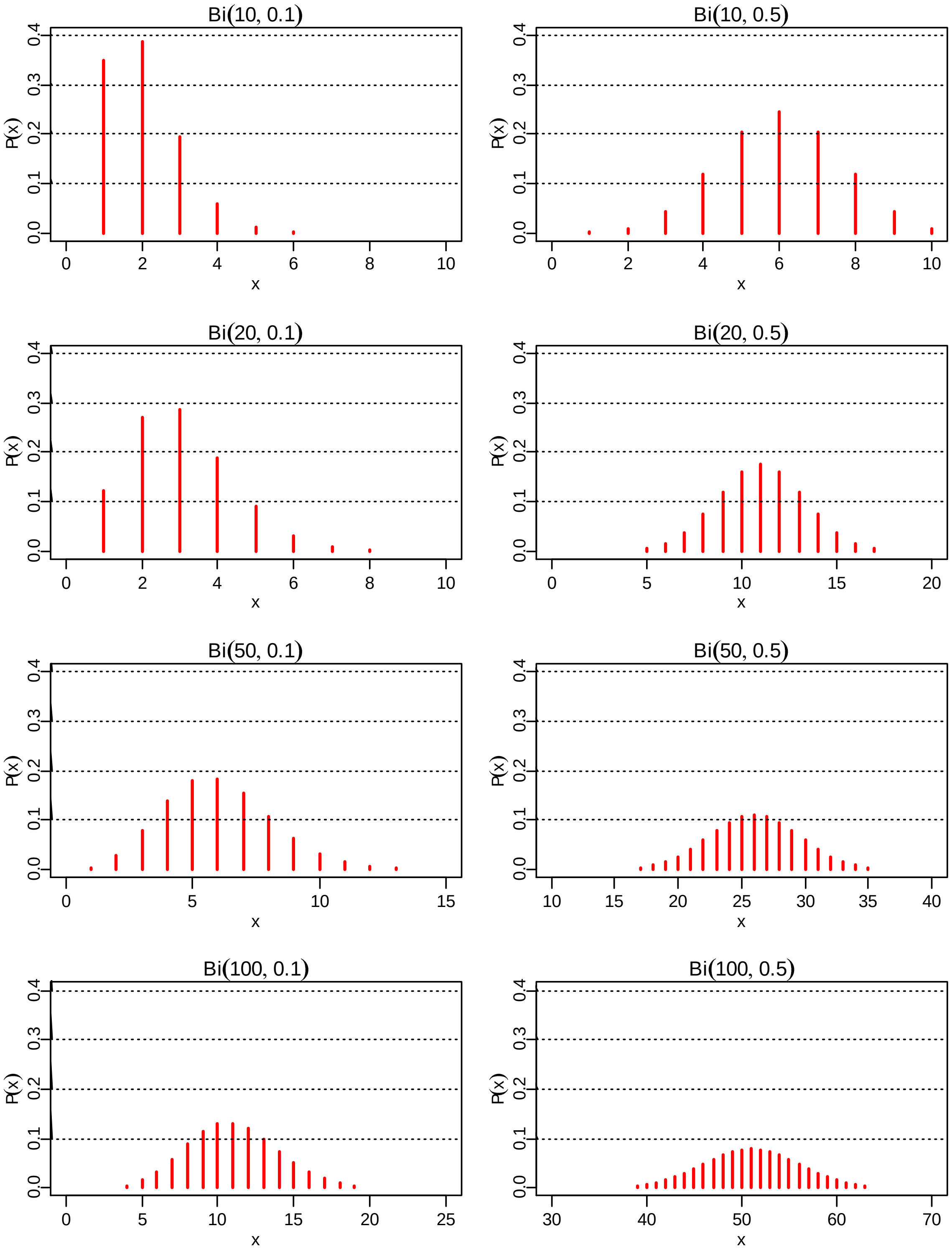


1. Dans quelle urne y aura-t-il le plus de billes ? Pourquoi ?
2. Avec quelle probabilité une bille se trouvera-t-elle dans une urne ?
3. Imaginons un jeu qui consiste à introduire des billes dans la planche.
   Le joueur gagne le nombre d'euros égal au numéro de l'urne dans laquelle tombe une boule. Quelle somme gagne-t-on en moyenne à ce jeu ? Réaliser les calculs pour des planches ayant 2 urnes, 3 urnes, 4 urnes, 5 urnes, 6 urnes.
4. <u>*Première généralisation de la planche de Galton*</u> : considérons une planche de Galton sur laquelle on a disposé n rangées de clous en quinconce et donc n+1 urnes numérotées de 0 à n. Supposons toujours qu'une bille rebondit au-dessus de chaque clou avec la même probabilité vers la gauche ou vers la droite.
   a) Combien y a-t-il de chemins au total ?
   b) Combien y a-t-il de chemins menant à une urne donnée ?
   c) Avec quelle probabilité une bille se trouvera-t-elle dans une urne donnée ?
5. <u>*Seconde généralisation de la planche de Galton*</u> : considérons une planche de Galton sur laquelle on a disposé n rangées de clous en quinconce et donc n + 1 urnes numérotées de 0 à n + 1. Supposons cette fois-ci qu'une bille rebondit au-dessus de chaque clou avec une probabilité différente : p vers la gauche et (p - 1) vers la droite. Avec quelle probabilité une bille se trouvera-t-elle dans une urne donnée ?

7.8 Comparaison de quelques binomiales

En observant les distributions de probabilité représentées ci-dessous, nous constatons

- lorsque p = 0.5, la distribution est toujours symétrique
- lorsque p est petit, elle est asymétrique pour n petit et devient symétrique lorsque n grandit. On considère que la distribution est symétrique lorsque l'espérance mathématique (np) est supérieure à 10 (dernier graphe représenté pour p ≠ 0.5)



8. La loi de Poisson

8.1 *Notion*

La loi de Poisson est une loi de distribution infiniment dénombrable. Elle apparaît dans de nombreux phénomènes naturels, comme le nombre d'appels téléphoniques par minute dans un central téléphonique, le nombre de fautes d'impression par page dans un livre, le nombre de particules α émises par une substance radioactive...

Dans le modèle de Poisson, le nombre attendu de réalisations est proportionnel à la dimension de l'échantillon. Il s'agit donc d'une répartition uniforme. Le cas du nombre d'habitants au km^2 ne correspond pas au modèle de Poisson, car les habitants ne sont pas répartis au hasard dans un pays.

La variable aléatoire X, nombre de réalisations dans un intervalle déterminé, est appelée variable aléatoire de Poisson. Elle est caractérisée par un seul paramètre appelé μ

Cette loi se définit par $p_k = P(X = k) = \dfrac{e^{-\mu}\mu^k}{k!}$

On peut montrer que l'espérance mathématique d'une telle variable aléatoire est égale à sa variance et vaut le paramètre μ

8.2 *Application :*

On suppose que dans un livre de 200 pages, 220 erreurs d'impression sont distribuées au hasard. Calculer la probabilité pour qu'une page donnée contienne
a) aucune erreur
b) une erreur
c) 2 erreurs
d) 2 erreurs ou plus
Solutions : a) 0.333 b) 0.366 c) 0.201 d) 0.301

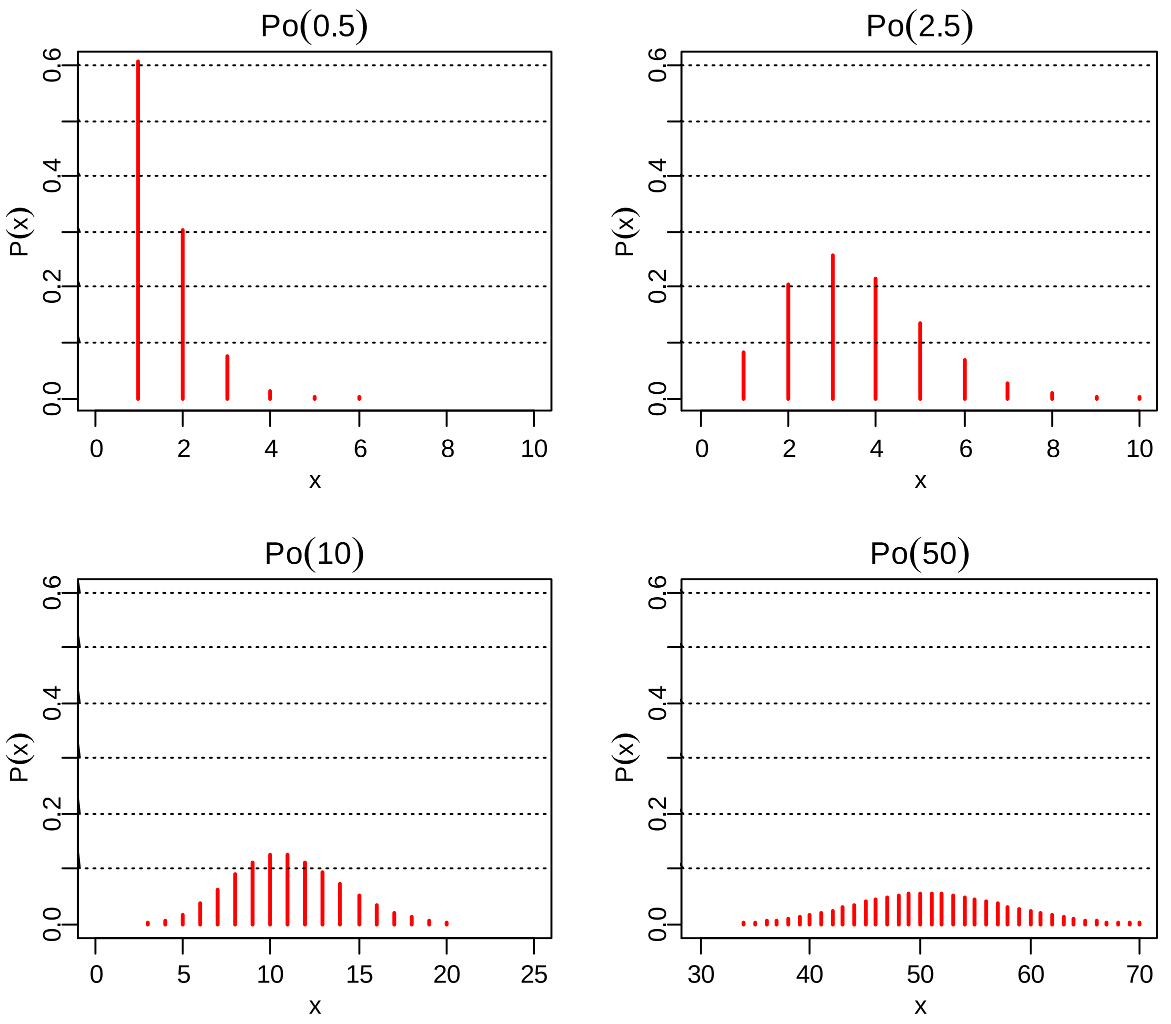
Dans de nombreux cas, pour la facilité des calculs, on se sert de tables de probabilités

<u>Exemple</u>
Supposons qu'une maternité enregistre en moyenne 6 entrées par jour. Quelle est la probabilité d'en enregistrer 10 ou plus en 1 jour?
Nous avons ici une v. a. Po(6)
$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9161 = 0.0839$ (la valeur 0.9161 est lue dans les tables de Poisson)

8.3 *<u>Exercices</u>*

1. A l'aérodrome de Zaventem, le trafic aérien obéit à une loi de Poisson de moyenne $\mu = 20$ avions par heure.
   a) on considère deux intervalles successifs de 15 minutes; quelle est la probabilité des événements suivants :
   E_1 : "le nombre total d'avions observés pendant les 30 minutes est égal à 8"
   E_2 : " 4 avions sont observés pendant le 1er intervalle et 4 pendant le second"
   E_3 : "le nombre total d'avions observés pendant les 30 minutes est strictement inférieur à 10
   b) on observe le ciel pendant 10 intervalles de 6 minutes ; quelle est la probabilité que pendant 3 des ces intervalles on ait observé aucun avion ?
   <u>Solutions</u> : a) $P(E_1) = 0.1126$ $P(E_2) = 0.0308$ $P(E_3)= 0.4579$ b)$P(E_4) = 0.107$

On trouvera ci-après les distributions de Poisson pour quelques valeurs du paramètre allant de 0.5 à 50. On constate, comme dans le cas de la distribution binomiale, que la distribution de Poisson tend vers une distribution symétrique en forme de cloche lorsque la moyenne augmente.



8.4 Approximation d'une binomiale par une loi de Poisson

On peut montrer : Si X = Bi (n; p) et $n \to \infty$; $p \to 0$ alors $P(X = k) = C_n^k\, p^k\,(1 - p)^{n-k} \to \frac{1}{k!}\mu^k e^{-\mu}$ où $\mu = np$

C'est pourquoi lorsqu'on se rapproche de ces conditions, on utilisera la loi de Poisson pour calculer des probabilités relatives à une variable aléatoire binomiale, car cela permet de se limiter à l'emploi d'un paramètre (μ) au lieu de deux (n et p)

En pratique, les statisticiens considèrent qu'il faut avoir $n > 50$ et $np < 5$ (c. à d. $p \leq 0.1$)

En quelque sorte, la loi de Poisson est un cas particulier de la binomiale lorsque $n > 50$ et $np < 5$

Exemple 1 :

X : v. a. Bi (50, 0.1) sa moyenne : $\mu = 5$ et sa variance $\sigma^2 = 4.5$

Soit à calculer la $P(X \leq 4)$ En utilisant les tables pour une Bi (50, 0.1), nous obtenons : $P(X \leq 4) = 0.4312$

Comme les valeurs de la moyenne et de la variance sont assez proches, on va approximer la binomiale par une Poisson de moyenne $\mu = 5$, et par les tables, nous obtenons :pour une v. a. Po (5) : $P(X \leq 4) = 0.4405$

L'approximation sera bien sûr d'autant meilleure que la moyenne et la variance de la variable aléatoire binomiale sont proches.

Exemple 2 : La fréquence d'une maladie dans une population donnée égale 1 pour 100 000. Quelle est la probabilité que sur 10 000 individus prélevés au hasard, il y en ait au moins un atteint de cette maladie ?

Le nombre d'individus atteints de psoriasis est une v.a. Bi (10 000; 0.00001) avec $\mu = np = 0.1$ et $npq = 0.099999 \cong 0{,}1$ La variance étant quasi égale à la moyenne, cette variable binomiale peut être approximée par une v. a. de Poisson de paramètre $\mu = 0.1$

Par les tables : P[1 ≤ X ≤10 000] = 1 - P(X ≤ 0) = 0.09516
Par le calcul exact de la probabilité cherchée, en considérant que X est une v.a. Bi(10 000; 0.00001)
$P[1 \leq X \leq 10\,000] = 1 - P(X = 0) = 1 - C^0_{10\,000}p^0q^{10\,000} = 1 - 0.99999^{10\,000} = 0.095163$

Et on constate que l'approximation par la loi de Poisson est excellente.
Exercice :
Un marchand de graines constate qu'en moyenne, 5% des graines de la variété ne germent pas. Il vend des paquets de 100 graines et garantit 90% de germination.
a) On prend un paquet au hasard. Quelle est la probabilité que la garantie soit violée ?
b) On prend 10 paquets au hasard. Quelle est la probabilité que la garantie soit violée au moins une fois ?
Solution : a) 0.0137 b) 0.1289

9. Variables aléatoires continues.

De nombreux exemples tels que la mesure de la taille d'une personne, un relevé de températures, le taux de cholestérol d'une population, des mesures de temps, de masse, de longueur, …sont des variables aléatoires continues. Or les distributions binomiales et de Poisson sont des distributions de probabilités discrètes. Pour bien décrire le comportement d'une variable aléatoire continue, nous allons devoir considérer les distributions de probabilités continues.

Si X est une variable aléatoire dont l'ensemble des valeurs X(S) est un intervalle (ou R), alors la distribution de probabilité devient une courbe f(x) définie sur l'ensemble des points de cet intervalle ou sur R. Mais comment tracer cette courbe ?

Exemple :

1. *On tire un nombre au hasard dans l'intervalle [2;4]. Quelle est la probabilité que ce nombre soit π ?*

2. *On reprend l'exemple précédent et on se demande : quelle est la probabilité que ce nombre soit > 3.6 ?*

L'intuition nous pousse à répondre 0 dans le premier cas et 0.2 dans le second.

Ici, la catégorie d'épreuve $\Omega = [2, 4]$ et $X : \Omega \to R : \omega \to X(\omega) = 0$

En reprenant la notion de base de probabilité : $P(A) = \dfrac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \dfrac{\# A}{\# \Omega}$, la probabilité d'obtenir un nombre déterminé vaut $\dfrac{1}{\infty} = 0$ tandis que toutes les valeurs de l'intervalle [2, 4] étant équiprobables, la probabilité d'obtenir un nombre compris entre 3,6 et 4 vaut la longueur de l'intervalle [3,6 ; 4] divisée par la longueur de l'intervalle [2, 4] c. à d. $\frac{0.4}{2} = 0.2$

Si la probabilité d'obtenir une valeur déterminée est nulle, comment établir la distribution de probabilité ?

Pour une variable aléatoire discrète, nous avons observé la similitude entre cette distribution et le diagramme en bâtonnets des statistiques en remarquant $\sum_i P(X = x_i) = 1$ qui correspond à $\sum_i f(x_i) = 1$.

De même, la loi de probabilité d'une variable aléatoire continue correspond à une extension de la notion d'histogramme employé lors d'études de séries statistiques groupées en classes. Les surfaces des rectangles construits sur chaque classe étaient proportionnelles aux fréquences de ces classes. La somme des surfaces de ces rectangles valant 1.

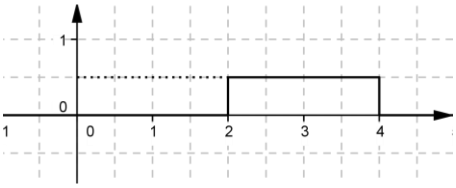
Dans notre exemple, il faut que la loi de probabilité soit telle que la surface limitée par l'axe des x, les droites d'équation x = 2 et x = 4 soit égale à 1.

Cette loi de probabilité sera donc une fonction nulle pour des valeurs de x plus petites que 2 ou pour des valeurs de x supérieures à 4.

De plus, entre 2 et 4, toutes les valeurs sont équiprobables : la fonction de probabilité sera donc constante entre 2 et 4

Il ne reste plus qu'à déterminer la valeur de cette constante en tenant compte du fait que la surface du rectangle ainsi obtenu vaut 1 : il faut que cette constante soit égale à 0.5

La variable aléatoire ainsi construite est appelée variable aléatoire uniforme car sa densité de probabilité est la fonction constante f(x) = 0.5 sur l'intervalle [2, 4]. C'est bien une fonction continue sur cet intervalle.

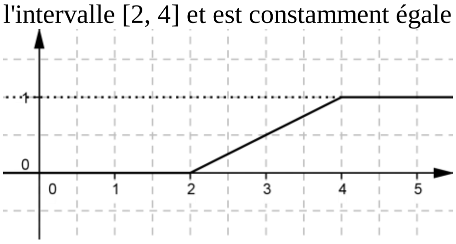


Loi de probabilité (ou fonction de densité):
si x < 2 : f(x) = 0
si 2 < x < 4 : f(x) : 0.5
si x > 2 : f(x) = 0

Pour une variable aléatoire discrète, la fonction de répartition est définie par : F(a) = P(X ≤ a)= $\sum_{x_i \le a} P(X = x_i)$

De même, dans le cas continu : F(a) = P(X < a) est représenté par la surface comprise entre la courbe, densité de probabilité, l'axe des x et la droite x = a ⇔ F(a) = P(X < a) = $\int_{-\infty}^{a} f(x)dx$

Dans notre exemple, la fonction de répartition est la fonction nulle jusque x = 2, elle croit uniformément sur l'intervalle [2, 4] et est constamment égale à 1 ensuite.



Elle est bien telle que ∀a ∈ R : F(a) = $\int_{-\infty}^{a} f(x)dx$

si a < 2 : F(a) = 0

si 2 < a < 4 : F(a) = $\int_{2}^{a} f(x)dx = \int_{2}^{a} 0.5dx = \frac{1}{2}[x]_2^a = \frac{1}{2}(a - 2)$

si a > 4 : F(a) = 1

<u>Conclusions :</u>
De manière générale : pour une variable aléatoire continue, la probabilité P(X = a) est nulle. En effet, la probabilité que X prenne exactement la valeur a est égale à la surface du rectangle de base dx infinitésimale et de hauteur f(x). Cette surface est elle aussi infinitésimale. On a donc : P (X = a) = 0 et P (X ≤ a) = P (X < a)

Les paramètres étudiés pour une variable aléatoire discrète se transforment :

– $\mu_x = E(X) = \sum_k p_k x_k$ devient : $\mu_x = E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx$

– $\sigma_x^2 = \text{Var}(X) = \sum_k p_k (x_k - \mu_x)^2$ devient : $\sigma_x^2 = \text{Var}(X) = E[(X - \mu_x)^2] = \int_{-\infty}^{+\infty} (x - \mu_x)^2 f(x)dx$

– La propriété : $\sum_i P(X = x_i) = 1$ devient : $P(-\infty < X < \infty) = 1 = \int_{-\infty}^{+\infty} f(x)dx$ (événement certain)

– $P(X < a) = F(a) = \sum_{x_i \le a} P(X = x_i)$ devient : $P(X < a) = \int_{-\infty}^{a} f(x)dx$ et de même : $P(a < X < b) = \int_{a}^{b} f(x)dx$

10. La loi normale.

10.1 Notions générales

L'observation de nombreuses séries statistiques montre que beaucoup d'entre elles ont un comportement semblable : une grande proportion des résultats groupés autour de la moyenne et de moins en moins lorsqu'on s'en éloigne.
Lorsque l'échantillon est divisé en un grand nombre de classes, l'histogramme des fréquences ressemble à une courbe en cloche symétrique par rapport à la moyenne. Les mathématiciens ayant étudié ces répartitions, ont déterminé l'équation d'une fonction qui s'en approche le mieux possible:

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ où μ et σ sont respectivement la moyenne et l'écart-type de l'échantillon. Cette fonction est appelée loi normale
On pouvait prévoir une fonction exponentielle vu les asymptotes horizontales (avec un exposant du second degré justifiant ainsi la symétrie).

Une étude de fonction classique nous donne les caractéristiques principales de cette fonction
dom f : R
La fonction n'admet pas de racines.
Elle n'a pas d'asymptote verticale.
$\lim_{x\to\pm\infty} f(x) = 0$: l'axe des abscisses : y = 0 est asymptote horizontale du graphe en $\pm\infty$

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \left(-\frac{1}{2}\cdot 2\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\right)\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{-x+\mu}{\sigma^3\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$f''(x) = \frac{1}{\sigma^3\sqrt{2\pi}}\left[-\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] + (-x+\mu)\left(-\frac{1}{2}\cdot 2\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}\right)\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\right]$$

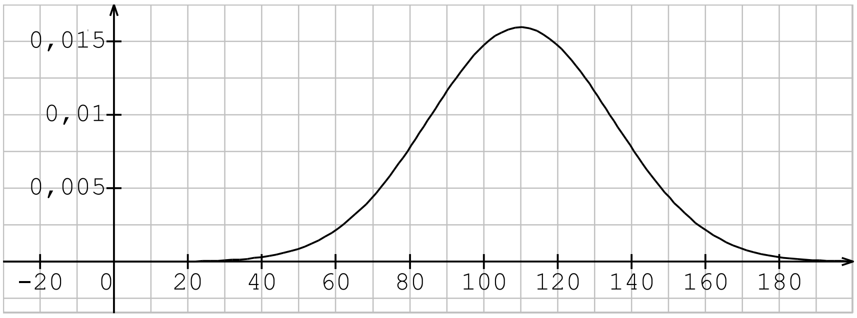
$$= \frac{1}{\sigma^3\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\left[-1+\frac{(x-\mu)^2}{\sigma^2}\right]$$

Nous en déduisons le tableau de signes :

		$\mu - \sigma$		μ		$\mu + \sigma$	
f '	+	+	+	0	-	-	-
f ''	+	0	-	-	-	0	+
f	↗	↗	↗	max	↘	↘	↘
	∪	point d'inflexion	∩	∩	∩	point d'inflexion	∪

Le maximum : $\left(\mu, \frac{1}{\sigma\sqrt{2\pi}}\right)$ Les points d'inflexion : $\left(\mu\pm\sigma, \frac{1}{\sigma\sqrt{2\pi}}\exp\left(\frac{-1}{2}\right)\right)$

Dans le cas où $\mu = 110$ et $\sigma = 25$, nous obtenons le graphique ci-contre qui illustre l'exemple :
"la force de résistance d'un matériau est distribué normalement avec une moyenne de 110kg et un écart type de 25 kg"
Cette courbe est aussi appelée : courbe de Gauss-Laplace

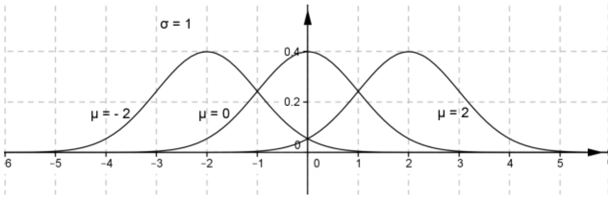


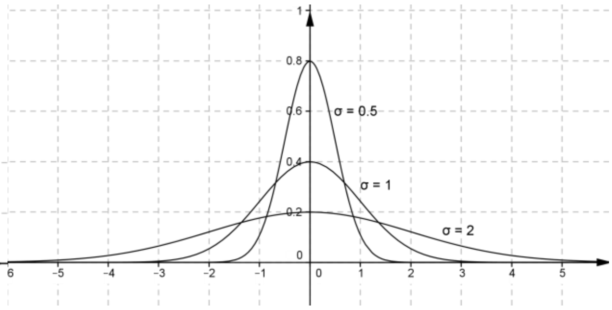
En résumé :

- La courbe admet une asymptote horizontale y = 0 en $+\infty$ et en $-\infty$
- Cette fonction a un maximum égal à $\frac{1}{\sigma\sqrt{2\pi}}$ pour $x = \mu$. Ce maximum est inversement proportionnel à σ ; il est donc d'autant plus grand que l'écart type est petit.
  L'aire comprise entre la courbe et l'axe des abscisses étant toujours égale à 1 quelles que soient les valeurs des paramètres μ_x et σ_x, il est logique que le maximum soit plus petit lorsque la courbe est plus étalée.
- La courbe présente deux points d'inflexion situés symétriquement de part et d'autre de la moyenne en des valeurs de x égales à $\mu_x - \sigma_x$ et $\mu_x + \sigma_x$.

Le premier graphique ci-dessous présente les graphes des courbes normales d'écarts types identiques et de moyennes respectivement égales à -2, 0 et 2

Le second graphique illustre la situation de courbes normales de même moyenne et d'écarts-types différents.





10.2 La normale réduite

Comme nous l'avons montré précédemment, les probabilités dans le cas des distributions de probabilités continues sont représentées par l'aire sous la courbe densité de probabilité : cela veut dire que la probabilité qu'une variable prenne une valeur entre a et b est égale à l'aire comprise sous cette courbe entre les droites verticales élevées aux points d'abscisses a et b.

Nous avons donc $P(a < X < b) = \int_a^b f(x)dx$

Dans le cas de la loi normale, on devrait donc calculer $\int_a^b f(x)dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$

Cette intégrale étant complexe, on utilisera une table d'aires.

Il est évidemment impossible de construire une table d'aire pour chaque courbe normale. C'est pourquoi nous allons essayer d'établir des liens avec une courbe de référence.

Parmi toutes les courbes normales, celle qui nous semble la plus simple est la courbe normale de moyenne nulle et d'écart type égal à 1. Ce sera la courbe de référence : elle est appelée "*normale réduite*"

La fonction de densité de probabilité f(z) de la v.a. normale réduite Z est : $f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$ notée φ (z)

Il y a une infinité de densités de probabilité de type normal, de formes différentes. Mais les calculs de probabilités correspondants peuvent être réalisés à partir d'une seule d'entre elles : celle de type normal centré réduit dont on utilise la table.

10.3 Table de la distribution normale centrée réduite.

Vu la symétrie de cette distribution, la table ne donne ces probabilités que pour des valeurs positives de Z. On remarquera donc $P(Z < -z) = P(Z > z) = 1 - P(Z < z)$

<u>Exemples</u>

1. $P(Z < 1.12) = 0.86864$
2. $P(Z > 0.6) = 1 - P(Z < 0.6) = 1 - 0.72575 = 0.27425$
3. $P(0.6 < Z < 1.3) = P(Z < 1.3) - P(Z < 0.6) = 0.90320 - 0.72575 = 0.17745$
4. $P(Z < -1.4) = P(Z > 1.4) = 1 - P(Z < 1.4) = 1 - 0.91924 = 0.08076$
5. $P(-1 < Z < 2) = P(Z<2) - P(Z<-1) = P(Z < 2) - P(Z > 1) = P(Z < 2) - [1 - P(Z < 1)] = 0.97725 - 1 + 0.84135 = 0.81860$

10.4 Probabilité pour une v.a. normale de moyenne μ et de variance σ^2

Pour une loi normale quelconque, on se ramènera toujours à une loi normale centrée réduite.

En effet, on sait que l'aire hachurée égale $P(a \le X \le b) = \int_a^b f(x)dx = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]dx$

En effectuant le changement de variable : $t = \frac{x-\mu}{\sigma}$ alors $dt = \frac{1}{\sigma}dx$

et $\int_a^b f(x)dx = \int_{a'}^{b'} \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}t^2\right]dt = \int_{a'}^{b'}\varphi(t)dt$ où $a' = \frac{a-\mu}{\sigma}$, $b' = \frac{b-\mu}{\sigma}$ et $\varphi(t) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}t^2)$

où φ (t) est une loi normale centrée, réduite.

Le calcul de $P(a \leq X \leq b)$ est donc ainsi remplacé par celui de $P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$ que l'on détermine par les tables.

Exemples :

1. On suppose que la taille de 2000 élèves est distribuée normalement avec une moyenne de 160 cm et un écart-type de 7 cm. Calculer le nombre d'élèves ayant des tailles
   a) Inférieures ou égales à 140 cm.
   b) Comprises entre 140 et 150 cm
   c) Comprises entre 150 et 175 cm
   d) Supérieures ou égales à 184 cm

   Solutions : a) $P(X < 140) = P\left(Z < \frac{140-160}{7}\right) = P(Z < -2,86) = P(Z > 2,86) = 1 - P(Z < 2,86) = 1 - 0.99788 = 0.00212 \Rightarrow$ 4,2 (4 à 5) élèves sur 2000
   Et de même b) P = 0.07424 ⇒ 148 el. c) P = 0.90746 ⇒ 1815 el. d) P = 0.0003 ⇒ 0.6 el.
2. On suppose que les diamètres des vis manufacturées par une compagnie sont distribuées normalement avec une moyenne de 0.25 cm et un écart-type de 0.02 cm. On considère qu'une vis est défectueuse si son diamètre est ≤ 0.20 cm ou ≥ 0.28 cm. Calculer le pourcentage de vis défectueuses fabriquées par la compagnie
   Sol : 7.3 %

10.5 Cas particulier

$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1)\ P(Z \leq 1) - P(Z \geq 1) = P(Z \leq 1) - (1 - P(Z \leq 1) = 2\ P(Z \leq 1) - 1 = 2 . (0.84135) - 1 = 0.6827$

L'aire comprise sous la courbe y = f(x) et les droites d'équation x = μ - σ et x = μ + σ (et donc la probabilité d'avoir un résultat compris entre ces valeurs) vaut donc 68,29 % de l'aire totale sous la courbe (qui vaut 1)

De même entre x = μ - 2 σ et x = μ + 2 σ, on trouverait 95,44%

Et entre x = μ - 3 σ et x = μ + 3 σ, on trouverait 99,74 %

Ce sont ces observations qui montrent l'intérêt de l'étude de l'écart type dans une série statistique.

10.6 Liens entre la loi binomiale et la loi normale.

Considérons l'expérience aléatoire qui consiste à jeter 8 fois de suite une pièce de monnaie bien équilibrée et à observer le nombre de faces.

On sait que $P(X = k) = C_8^k\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{8-k}$.

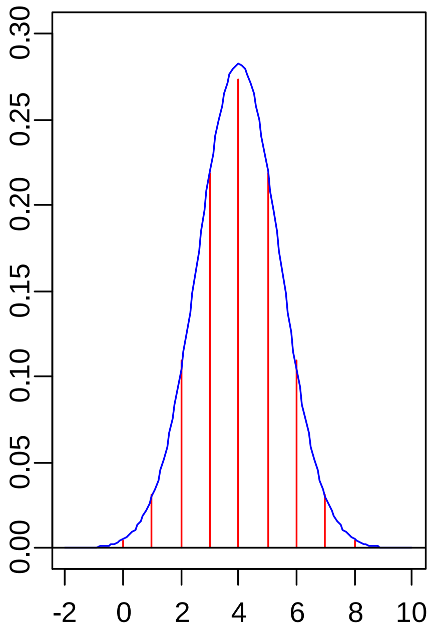
On obtient ainsi le premier tableau de valeurs ci-dessous et sa représentation graphique par le diagramme en bâtons ($\mu = n\,p = 4$ et $\sigma^2 = 2 \Rightarrow \sigma = \sqrt{2}$)

En considérant ensuite la fonction normale de moyenne μ = 4 et $\sigma = \sqrt{2}$ c. à d. : $f(x) = \frac{1}{2\sqrt{\pi}}\exp\left[-\frac{(x-4)^2}{4}\right]$

nous obtenons le second tableau de valeurs ci-après :

x_i	p_i
0	1/256 = 0.0039
1	8/256 = 0.031
2	28/256 = 0.109
3	56/256 = 0.218
4	70/256 = 0.273
5	56/256 = 0.218
6	28/256 = 0.109
7	8/256 = 0.031
8	1/256 = 0.0039

x_i	f(x)
0	0.005
1	0.029
2	0.103
3	0.219
4	0.282
5	0.219
6	0.103
7	0.029
8	0.005



Le graphique ci-contre reprend à la fois le diagramme en bâtons de la loi binomiale et le graphe de la fonction f(x).
Et nous observons : $p(X = x_i) \cong f(x_i)$

Plus généralement, on peut montrer que
Si X est une v.a. Bi (n,p)

et $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right]$

avec $\mu = np$ et $\sigma = \sqrt{npq}$. Alors, $P(X = x_i)$ et $f(x_i)$ sont d'autant plus voisins que n est grand et que p est proche de $\frac{1}{2}$

10.7 Liens entre Binomiale, Poisson et Normale

L'observation des distributions de probabilité de Poisson et Binomiale nous a montré que celles-ci tendent vers une courbe en cloche dans certaines conditions. On peut résumer ces conditions par le schéma suivant :

v.a. binomiale :
X = Bi (n,p)
$P(X = k) = C_n^k\, p^k\, (1 - p)^{n-k}$
E (X) = np   Var (X) = npq

Vers Poisson : $n \to \infty$, $p \to 0$, $np = \mu$; n>50, np<5 ; $\mu = np$

Vers normale : n ≥50, np>5, nq>5, p ≅0.5 ; $n \to \infty$, $p \cong 0.5$; $\mu = np$, $\sigma^2 = npq$

v.a. de Poisson
$X = P(\mu)$
$P(X = k) = \dfrac{e^{-\mu}\mu^k}{k!}$
$E(X) = \mu$ $Var(X) = \mu$

Poisson vers normale : $\mu > 5$, $\sigma = \mu$

v.a. normale
$X = N(\mu, \sigma^2)$
$P(a \le X \le b) = \dfrac{1}{\sigma\sqrt{2\pi}} \displaystyle\int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$
$E(X) = \mu$ $Var(X) = \sigma^2$

Les conditions théoriques telles que $n \to \infty$ sont remplacées par des conditions pratiques telles que n >50 par les statisticiens.
Pour illustrer cette possibilité de remplacer une binomiale par une loi normale sous certaines conditions nous allons maintenant répondre à la dernière partie du problème "le pouvoir d'une minorité" proposé au début de ce chapitre. Dans un pays d'un million de votants, presque tous sont indécis. Deux mille personnes sont opposées à une loi de vote des étrangers. Quelle est la probabilité qu'on rejette cette loi ?

Le nombre de votes négatifs suit une loi binomiale de paramètres n = 998000 et $p = \frac{1}{2}$

Comme n est très grand, et p = 0.5, npq > 10 ,nous sommes bien dans les conditions pour pouvoir approximer la binomiale par une loi normale de paramètre μ = 499000 (np) et σ^2 = 249500 (npq)

Pour que la loi soit rejetée, il faut que le nombre de votes négatifs soit ≥ 498000 (puisque 2000 sont opposés)

Pour pouvoir calculer cette probabilité, on utilisera les tables de la loi normale réduite

$$P(X \geq 498000) = P\left(Z \geq \frac{498000 - 499000}{\sqrt{249500}}\right) = P(Z \geq -2{,}002) = P(Z \leq 2{,}002) = 0{,}97725$$

Il y a donc 97,8 % de chances que la loi de vote des étrangers soit refusée.

11. Exercices Généraux

1. On joue 4 parties consécutives de pile ou face et on considère comme variable aléatoire X la longueur de la plus longue série ininterrompue de faces dans la série de 4 résultats. Etablir la distribution de X et l'illustrer par un diagramme en bâtons. Calculer la moyenne et l'écart-type de X
sol : $\overline{X}$ = 1.6875 σ = 0.9823

2. On jette ensemble un dé et une pièce de monnaie. On observe le nombre de points du dé et on l'augmente de 1 si la pièce montre pile, on le diminue de 1 si elle montre face. Le résultat ainsi obtenu est une variable aléatoire X. On en demande la distribution, la moyenne et l'écart-type. Que vaut P (1 < X < 5) ?
sol : $\overline{X}$ = 3.5 σ = 1.979 P (1 < X < 5) = 0.5

3. Une urne contient deux boules vertes et deux boules rouges. On tire les boules au hasard, une par une, sans remise, jusqu'à obtenir une boule verte. Combien doit-on tirer de boules en moyenne ? sol : 1,67

4. On lance trois fois de suite une pièce de monnaie bien équilibrée. On gagne 2€ pour chaque résultat "pile" et on perd 1€ pour chaque "face". Quelle est l'espérance de gain d'un joueur ? sol : 1,5€

5. Une épidémie de grippe répartie au hasard affecte 10% de la population scolaire d'une ville.
a) Quelle est la probabilité que dans une classe de 20 élèves il y ait 5 malades ? sol : 0.0319
b) Quelle est la probabilité qu'il y ait 10 malades dans une classe de 40 élèves ? sol : 0.0036

6. Lancez 15 fois de suite une pièce de monnaie. Si vous obtenez 15 faces, vous gagnez 20 000 000 €, sinon vous perdez 25 €. Montrez qu'à ce jeu, l'espérance de gain est supérieure à 550 €. Jouerez-vous ?
sol : E(X) = 585,35€

7. Une famille a 6 enfants. Calculer la probabilité pour qu'il y ait a) 3 garçons et 3 filles, b) moins de garçons que de filles. (On suppose que la probabilité pour qu'un enfant particulier soit un garçon est ½.)
sol : a) 0.3125 b) 0.34375

8. On lance un dé non pipé 5 fois de suite et on appelle X la variable aléatoire correspondant au nombre de fois que le point 6 est apparu lors des 5 lancers. Déterminer la loi de probabilité, l'espérance mathématique, la variance et l'écart type de cette variable aléatoire. sol : $B_{5,\frac{1}{6}}$ $\mu = \frac{5}{6}$ $\sigma = \frac{\sqrt{5}}{6}$

9. Une machine en mauvais état produit en moyenne 1 pièce défectueuse sur 3. Sur 5 pièces produites par cette machine, quelle est la probabilité d'en avoir deux défectueuses ? sol : 0.329

10. En conditions habituelles de fonctionnement, une machine produit des pièces défectueuses dans une proportion constante égale à 1/10000. Un client reçoit un lot tiré au hasard de 30000 pièces usinées par cette machine.
a) Quelle est la probabilité qu'il trouve 3 pièces défectueuses dans le lot ? sol : 0.22405
b) Quelle est la probabilité qu'il trouve strictement moins de 3 pièces défectueuses dans le lot ? sol : 0.4232
c) Quelle est la probabilité qu'il y trouve strictement plus de 7 pièces défectueuses? sol : 0.0119

11. Un livre de 400 pages contient 100 erreurs distribuées au hasard. On ouvre le livre à une page quelconque. Le nombre d'erreurs rencontrées sur cette page est une variable aléatoire X.
a) Quelle est la probabilité que le nombre d'erreurs constaté sur cette page soit 3 ? sol : 0.002028
b) Quelle est la probabilité que ce nombre d'erreurs soit strictement supérieur à 2 ? sol : 0.002

12. L'opératrice d'un standard téléphonique reçoit en moyenne 2 appels par minute. Ceux-ci sont répartis au hasard dans le temps. Quelle est la probabilité que le nombre d'appels reçus en 5 minutes dépasse 15 ?
sol : 0.0487

13. A une entrée d'autoroute arrivent, en temps normal (hors week-end, vacances et périodes de pointe), 600 véhicules par heure en moyenne, répartis au hasard dans le temps. Soit X le nombre de véhicules qui entrent sur l'autoroute en 30 secondes. Quelle est la probabilité d'avoir 10 entrées en 30 secondes ?
sol : 0.018

14. Une machine a confectionné 1000 paquets dont les masses sont distribuées normalement. La masse moyenne des sachets est de 500 gr avec un écart-type de 25 gr.
a) Combien de paquets ont une masse comprise entre 480 et 550 gr. ? sol : 765
b) Combien de paquets ont une masse comprise entre 480 et 490 gr. ? sol : 133
c) Combien ont une masse strictement supérieure à 450 gr. ? sol : 977
d) Combien ont une masse strictement inférieure à 475 gr. ? sol : 159
e) Entre quelles limites, symétriques de la moyenne, sont compris les 50% de l'échantillon ?
sol : entre 483 g et 517 g

15. Un complexe sidérurgique fabrique des pipe-lines d'un diamètre intérieur moyen égal à 50 cm avec un écart-type de 0.25 mm. Le cahier des charges alloue une tolérance sur le diamètre comprise entre 49.95 cm et 50.05 cm, sinon les conditions de vente sont refusées. En admettant que les diamètres sont distribués normalement, quel sera le pourcentage des tubes défectueux ? sol : 4,6 %

16. Supposons que la probabilité de naître fille ou garçon soit égale.
a) Calculer la probabilité que sur les 780 000 naissances d'une année en France, il y ait entre 49.9% et 50.1% de filles sol : 0.923
b)Calculer la probabilité qu'il y ait plus de 51% de filles sol : $\cong 0$

17. Des pièces circulaires fabriquées par une machine ont un diamètre dont la distribution est normale de moyenne $\mu = 10.4$ mm et d'écart-type $\sigma = 0.8$ mm.
Seules les pièces ayant un diamètre compris entre 9 et 11 mm sont utilisables.
a) Calculer le pourcentage de pièces utilisables fabriquées par cette machine.
b) Sachant qu'une pièce est utilisable, quelle est la probabilité que son diamètre soit supérieur à 10 mm ?
c) On examine 10 pièces. Quelle est la probabilité que 6 d'entre elles exactement soient utilisables ?
Sol : a) 73,33% b) 0.6339 c) 0.1652

18. *Un fabricant de soierie décide de sélectionner 10% de sa production pour confectionner des articles de qualité. Sachant que la résistance des fils de soie est distribuée selon une loi normale de paramètres $\mu = 80$ grf et $\sigma = 4$ grf, quelle est la valeur critique de la résistance au-dessus de laquelle les fils de soie seront sélectionnés ? sol : 85,12 grf

19. **Une entreprise fabrique des machines. La variable aléatoire "nombre de machines vendues chaque jour " obéit à une loi normale. Au cours des deux dernières années (700 jours ouvrables) le nombre de jours pendant lesquels elle a vendu plus de 30 machines s'est élevé à 589. Le nombre de jours pendant lesquels elle a vendu plus de 90 machines s'est élevé à 16. Quelles sont la vente journalière et l'écart type des ventes ?
sol : $\mu = 50$ $\sigma = 20$