1.7 Fonction de répartition et diagramme cumulatif

On appelle fonction de répartition d'une variable statistique quantitative toute application définie par :

$$F : \mathbb{R} \rightarrow [\,0\,,1\,]$$
$$x \rightarrow F(x) = P(X \leq x)$$

$F(x)$ proportion des individus dont la valeur de la variable est strictement inférieure ou égale à x, c'est-à-dire $X \leq x$.

1.7.1 Cas de la variable statistique quantitative discrète

$F(x)$ = fréquence de $(X \leq x) = f_1 + f_2 + \ldots + f_p = F_p$ tel que : $f_1, f_2, \ldots, f_p$ sont les fréquences des valeurs de la variable $\leq x$, si non $F(x) = 0$. Donc

$$F(x) = \begin{cases} 0 & si \quad x < x_1 \\ F_i & si \quad x_i \leq x < x_{i+1} \\ 1 & si \quad x_r \leq x \end{cases}$$

tel que r désigne l'ordre de la **<u>dernière valeur (modalité)</u>**.

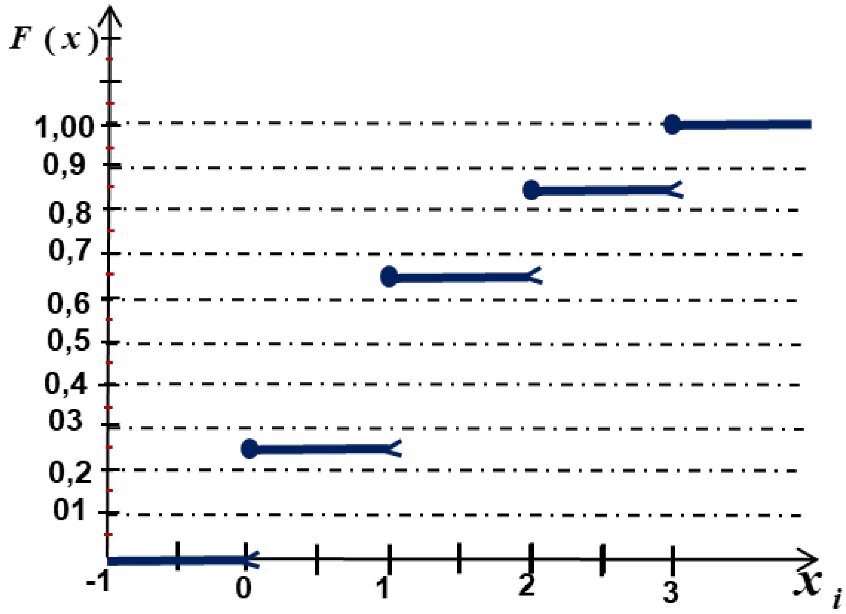
Exemple 1.7.1 Le tableau suivant, donne le nombre d'absences des étudiants au module d'analyse.

Nbre d'absences x_i	*Effectifs* n_i	f_i	F_i
0	5	0,25	0,25
1	8	0,40	0,65
2	4	0,20	0,85
3	3	0,15	1,00
Total	20	1,00	

Donc la fonction de répartition correspondante est

$$F(x) = \begin{cases} 0 & si & x < 0 \\ 0{,}25 & si & 0 \leq x < 1 \\ 0{,}65 & si & 1 \leq x < 2 \\ 0{,}85 & si & 2 \leq x < 3 \\ 1 & si & 3 \leq x \end{cases}$$

Ainsi on obtient la représentation de la fonction de répartition, appelée **diagramme cumulatif** ou **diagramme intégral**



Remarque 1.7.1 Dans le cas discret on a une fonction en escalier.

1.7.2 Cas de la variable statistique quantitative continu

Dans ce cas on commence par la technique d'obtention de la courbe de la fonction de répartition qui est appelée **courbe cumulative** qui est une courbe approximative, en utilisant l'interpolation linéaire entre deux points dans le plan.

En fait : L'**interpolation linéaire** est la méthode la plus simple pour estimer la valeur prise par une fonction continue entre deux points déterminés (interpolation). Elle consiste à utiliser pour cela la fonction affine (de la forme $f(x) = m.x + b$) passant par les deux points déterminés.

Donc la fonction de répartition dans le cas quantitative continu est définit de la même façon que dans le cas quantitatif discret :

$$\begin{aligned} F: \ & \mathbb{R} \to [\,0\,,1\,] \\ & x \to F(x) = P(X \leq x) \end{aligned}$$

$F(x)$ proportion des individus dont la valeur de la variable est strictement inférieure ou égale à x, c'est-à-dire $X \leq x$.

La courbe cumulative, est une ligne brisée obtenue en joignant :

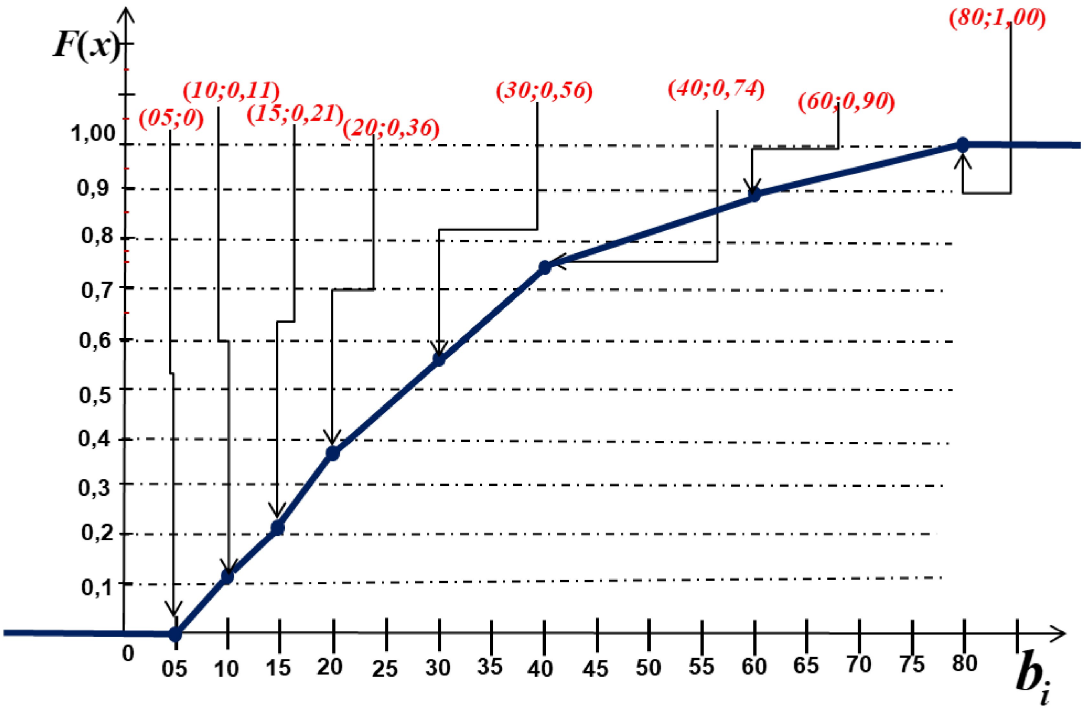
1. **Les différents points de coordonnés (b_i, F_i) dans l'ordre croissant avec $F_0 = 0$.**
2. Et en joignant du côté gauche du point $(b_0\,,\,F_0)$ la ½ droite $y = 0$ et du coté droit du

point $(b_r\ , F_r)$ la ½ droite $y = 1$.

Exemple 1.7.2 On reprend l'exemple de la page 16.

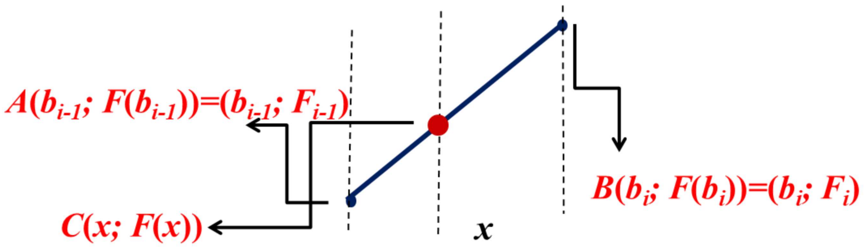
$[b_{i-1}\ , b_i[$	n_i	N_i	f_i	F_i
[5 , 10[	11	11	0 ,11	0,11
[10 , 15[	10	21	0 ,10	0,21
[15, 20[	15	36	0 ,15	0,36
[20, 30[	20	56	0 ,20	0,56
[30 , 40[	18	74	0 ,18	0,74
[40 , 60[	16	90	0 ,16	0,90
[60, 80[	10	100	0 ,10	1,00

Ainsi la courbe de la fonction de répartition, appelée **courbe cumulative** se dessine comme suit :



- **Technique de calcul d'une valeur de $F(x)$ pour $x \in \mathbb{R}$:**

Pour calculer $F(x)$, $\forall\ x \in \mathbb{R}$ on va réaliser **une interpolation linéaire** entre les points $A(b_{i-1}\ ; F(b_{i-1})) = (b_{i-1}\ ; F_{i-1})$ et $B(b_i\ ; F(b_i)) = (b_i\ ; F_i)$ tels que $x \in [b_{i-1}\ , b_i[$



L'équation de la droite (AB) est de la forme $y = mx + p$ tel que :

$$m = \frac{y_B - y_A}{x_B - x_A} = \frac{y_C - y_A}{x_C - x_A} = \frac{F(b_i) - F(b_{i-1})}{b_i - b_{i-1}} = \frac{F_i - F_{i-1}}{b_i - b_{i-1}}$$

$$\Rightarrow F(x) = F_{i-1} + m(x - b_{i-1}) = F_{i-1} + \frac{f_i}{b_i - b_{i-1}}(x - b_{i-1})$$

Remarque 1.7.2 Dans un cas contraire on peut calculer x si on a la valeur de $F(x)$, en utilisant toujours **une interpolation linéaire** entre les points $A(b_{i-1}\,;\,F(b_{i-1})) = (b_{i-1}\,;\,F_{i-1})$ et $B(b_i\,;\,F(b_i)) = (b_i\,;\,F_i)$ tels que $x \in [b_{i-1}, b_i[$ et $(F(x) \in [F_{i-1}, F_i[\,)$, en effet :

$$\frac{F(b_i) - F(b_{i-1})}{b_i - b_{i-1}} = \frac{F_i - F_{i-1}}{b_i - b_{i-1}} = \frac{F(x) - F_{i-1}}{x - b_{i-1}}$$

$$\Rightarrow x = (F(x) - F_{i-1})\left(\frac{b_i - b_{i-1}}{F_i - F_{i-1}}\right) + b_{i-1} \Rightarrow x = b_{i-1} + a_i\left(\frac{F(x) - F_{i-1}}{F_i - F_{i-1}}\right)$$

Ainsi on peut formuler la fonction de répartition d'une façon générale comme suit

$$F(x) = \begin{cases} 0 & si \quad x < b_0 \\ F_{i-1} + \dfrac{f_i}{b_i - b_{i-1}}(x - b_{i-1}) & si \quad b_{i-1} \le x < b_i \\ 1 & si \quad b_r \le x \end{cases} \qquad \text{avec } F_0 = 0.$$

Et la fonction de répartition pour la série statistique dans l'exemple cité ci-dessus est

$$F(x) = \begin{cases} 0 & si \quad x < 5 \\ 0 + \dfrac{0{,}11}{5}(x - 5) & si \quad 5 \le x < 10 \\ 0{,}11 + \dfrac{0{,}10}{5}(x - 10) & si \quad 10 \le x < 15 \\ 0{,}21 + \dfrac{0{,}15}{5}(x - 15) & si \quad 15 \le x < 20 \\ 0{,}36 + \dfrac{0{,}20}{10}(x - 20) & si \quad 20 \le x < 30 \end{cases} \quad \text{et} \quad F(x) = \begin{cases} 0{,}56 + \dfrac{0{,}18}{10}(x - 30) & si \quad 30 \le x < 40 \\ 0{,}74 + \dfrac{0{,}16}{20}(x - 40) & si \quad 40 \le x < 60 \\ 0{,}90 + \dfrac{0{,}10}{20}(x - 60) & si \quad 60 \le x < 80 \\ 1 & si \quad 80 \le x \end{cases}$$