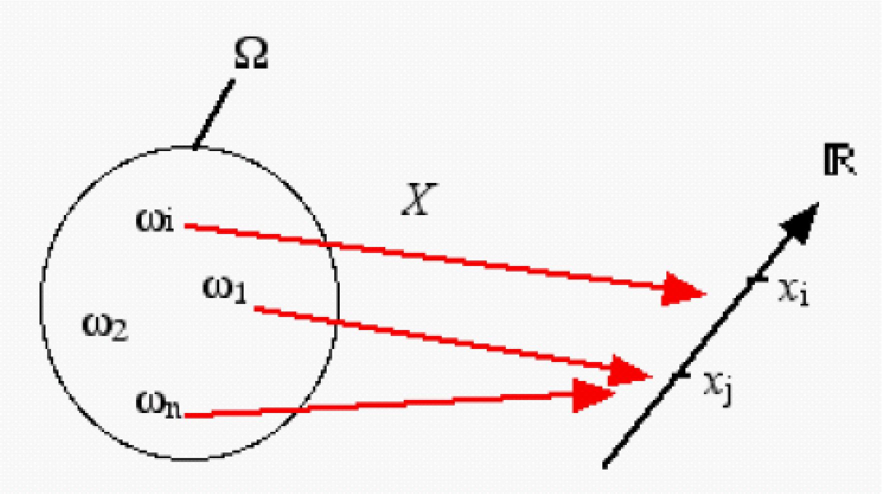
Variables Aléatoires

1. Introduction

Etant donné un **espace probabilisé** d'espace fondamental Ω et de mesure de probabilité P, on appelle **variable aléatoire** sur cet espace, toute **application** X de Ω dans $\mathbb{R}$ telle que :

$$X: \quad \varepsilon(\Omega) \rightarrow \mathbb{R}$$
$$\omega \rightarrow X(\omega)$$



A chaque évènement élémentaire ω de Ω correspond un nombre réel x associé à la variable aléatoire X. Comme l'indique le graphe, il n'y a pas obligatoirement autant de valeurs possibles prises par la variable aléatoire X que d'évènements élémentaires. La valeur x correspond à la **réalisation** de la variable X pour l'évènement élémentaire ω.

Une variable aléatoire est dite **discrète** si elle ne prend que des **valeurs discontinues** dans un intervalle donné (borné ou non borné). L'ensemble des nombres entiers est discret. En règle générale, toutes les variables qui résultent d'un **dénombrement** ou d'une **numération** sont de type discrètes.

Exemples :

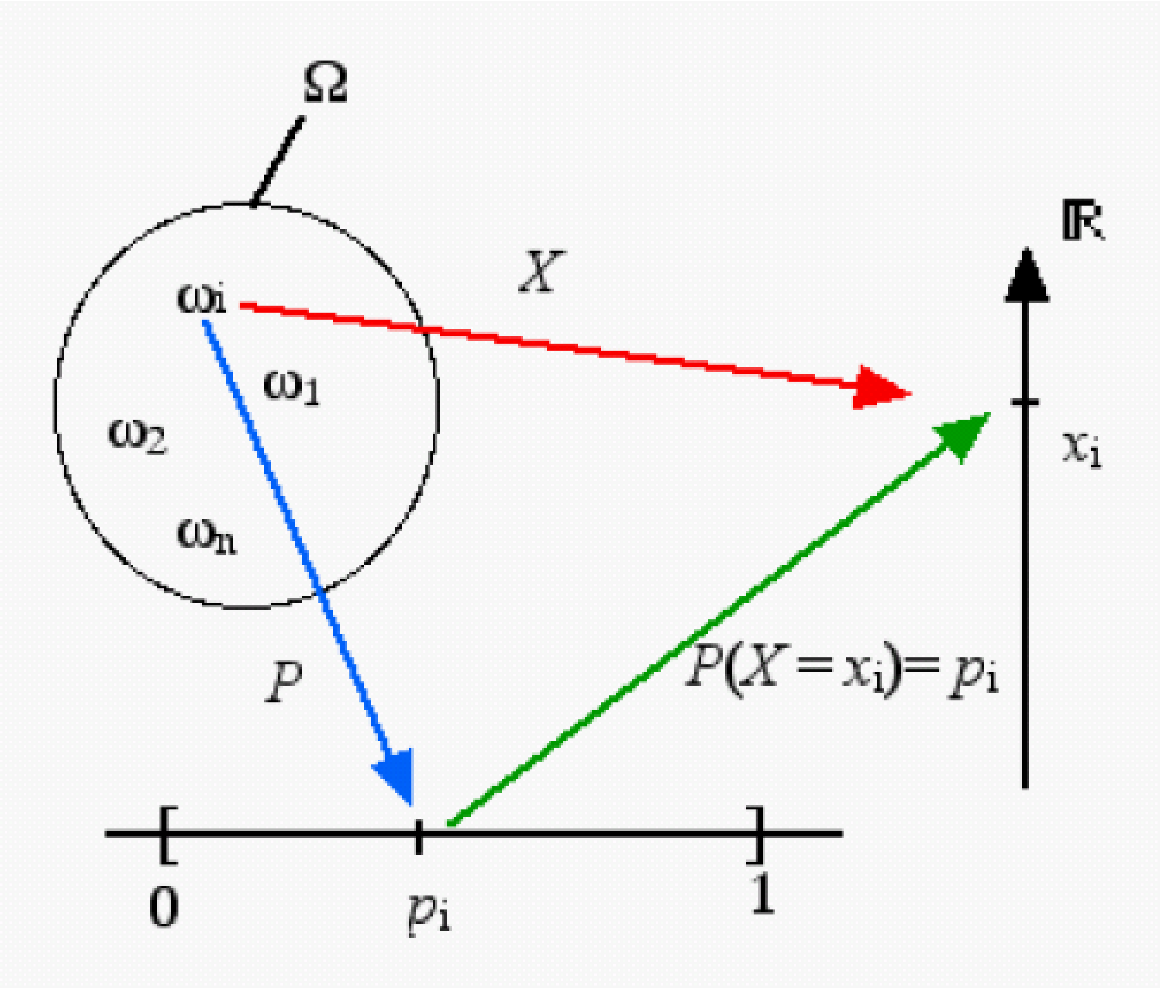
Les variables aléatoires,

- le nombre de petits par porté pour une espèce animale donnée (chat, marmotte, etc),
- le nombre de bactéries dans 100 ml de préparation,
- le nombre de mutations dans une séquence d'ADN de 10 kb,

etc…

sont des **variables aléatoires discrètes**.

Une variable aléatoire est caractérisée par l'ensemble des valeurs qu'elle peut prendre et par l'expression mathématique de la probabilité de ces valeurs. Cette expression s'appelle **la loi de probabilité** (ou **distribution de probabilité**) de la variable aléatoire.



La loi de probabilité d'une **variable aléatoire discrète** est entièrement déterminée par les probabilités p_i des évènements $\{X = x_i\}$, x_i parcourant l'univers image $X(\Omega)$. La **loi de probabilité** est donnée par les $(x_i, p_i)_i$.

Remarque : Afin de simplifier l'écriture, nous noterons pour la suite du cours :

$P(\{X = x_i\})$ équivalent à $P(X=x_i)$ ou p_i

On appelle **fonction de répartition** d'une variable aléatoire X, la fonction F_X telle que :

$$F_X : \mathbb{R} \to \mathbb{R}$$

$$t \to F_X(t) = P(X < t)$$

Concrètement la fonction de répartition correspond à la **distribution des probabilités cumulées**. Le plateau atteint par la fonction de répartition correspond à la valeur de probabilité 1 car $\sum_i p_i = 1$.

L'importance pratique de la fonction de répartition est qu'elle permet de calculer la probabilité de tout intervalle dans $\mathbb{R}$.

Les **propriétés** associées à la fonction de répartition sont les suivantes :

Soit F_X la fonction de répartition d'une **variable aléatoire discrète** X alors :

(P$_1$) $\forall t \in \mathbb{R} \quad 0 \leq F_X(t) \leq 1$

(P$_2$) F_X est croissante sur $\mathbb{R}$

(P$_3$) $\lim_{t \to -\infty} F_X(t) = 0$ et $\lim_{t \to +\infty} F_X(t) = 1$

(P$_4$) si $a \leq b \quad P(a \leq X \leq b) = F_X(b) - F_X(a)$

Exemple :

On considère l'évènement ω « lancer de 3 pièces ». On introduit une variable aléatoire X définie par $X(\omega)$ « nombre de piles de l'évènement ω». La loi de probabilité de X est :

Nombre de piles	$P(X = x_i)$	F_X
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

Dans le cas d'une variable aléatoire discrète, on utilise un **diagramme en bâtons** pour visualiser la **distribution de probabilités** et une fonction en escalier pour **la fonction de répartition.**

Exemple :

On considère l'évènement ω « lancer de 3 pièces ». On introduit une variable aléatoire X définie par $X(\omega)$ « nombre de piles de l'évènement ω». La loi de probabilité de X est :

Nombre de piles	$P(X = x_i)$	F_X
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

Dans le cas d'une variable aléatoire discrète, on utilise un **diagramme en bâtons** pour visualiser la **distribution de probabilités** et une fonction en escalier pour **la fonction de répartition.**

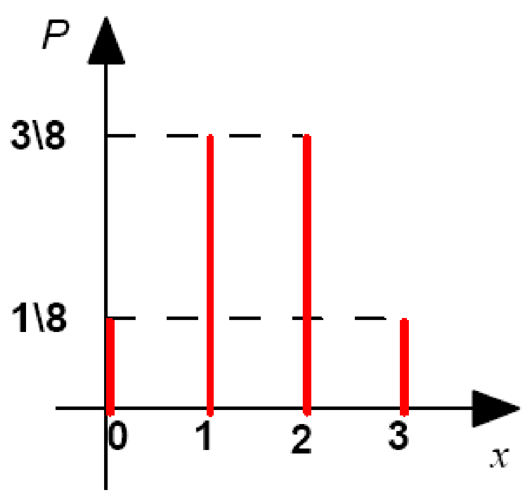
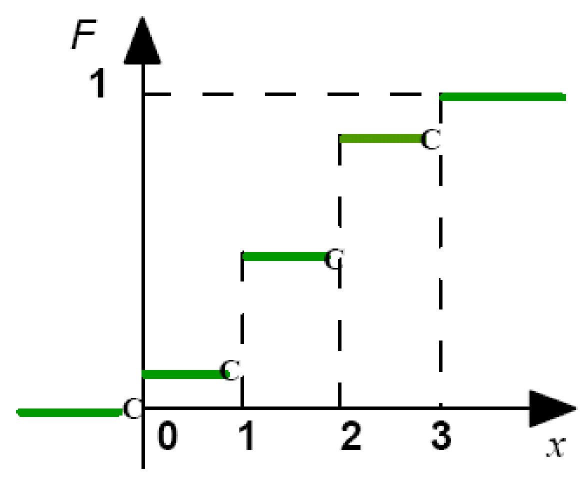


Diagramme en bâtons



Fonction de répartition

Une variable aléatoire est dite **continue** si elle peut prendre **toutes les valeurs** dans un intervalle donné (borné ou non borné). En règle générale, toutes les variables qui résultent d'une **mesure** sont de type continu.

Exemples :

Les variables aléatoires,

- le masse corporelle des individus pour une espèce animale donnée,
- taux de glucose dans le sang,
- etc.

sont des **variables aléatoires continues**.

Dans le cas d'une **variable aléatoire continue**, la loi de probabilité associe une probabilité à chaque ensemble de valeurs définies **dans un intervalle donné**. En effet, pour une variable aléatoire continue, la probabilité associée à l'évènement $\{X = a\}$ est nulle, car il est impossible d'observer exactement cette valeur.
On considère alors la probabilité que la variable aléatoire X prenne des valeurs comprises dans un intervalle $[a,b]$ tel que $P(a \leq X \leq b)$.

Lorsque cet intervalle tend vers 0, la valeur prise par X tend alors vers une fonction que l'on appelle **fonction densité de probabilité** ou **densité de probabilité**.

On appelle **densité de probabilité** toute application continue par morceaux :

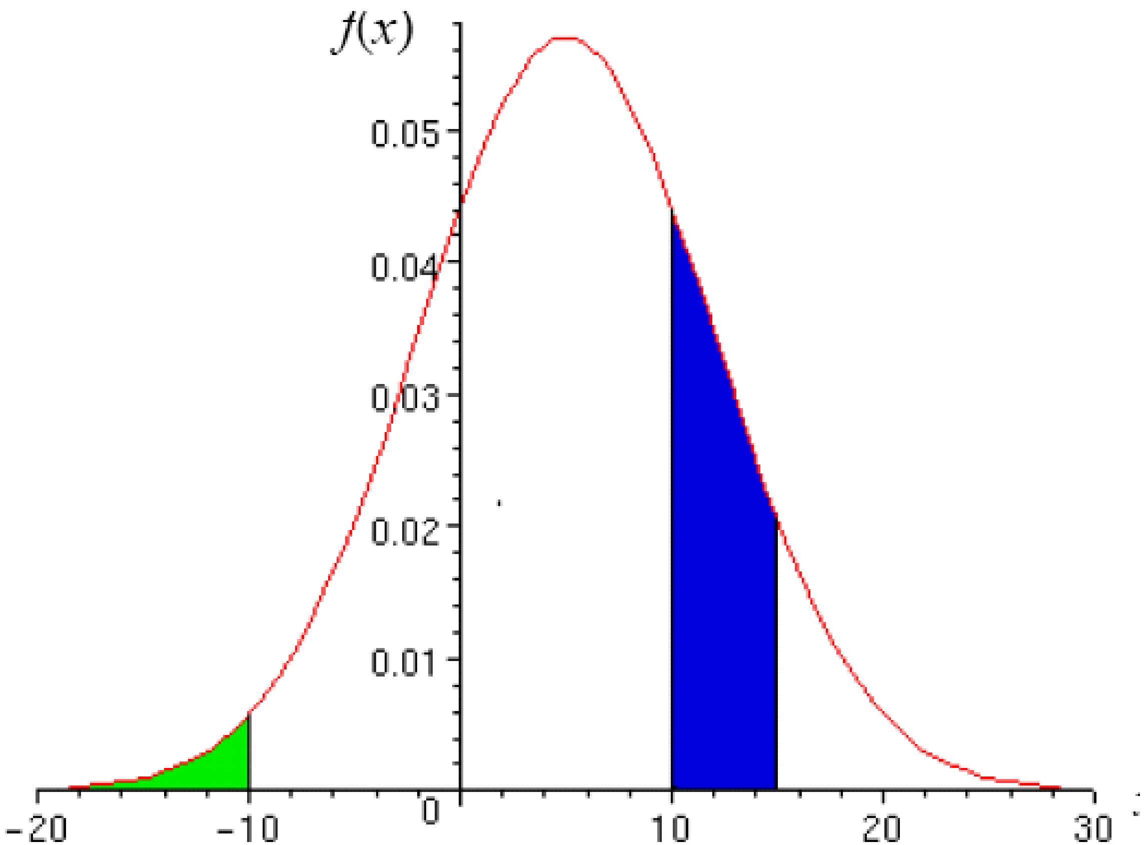
$$f : \mathbb{R} \to \mathbb{R}$$

$$x \to f(x)$$

telle que :

(P$_1$) $\forall x \in \mathbb{R}$ $\quad f(x) \geq 0$

(P$_2$) $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ (en supposant que $\int_{-\infty}^{+\infty} f(x)\,dx$ existe)



Soit une **fonction densité de probabilité** $f(x)$:

(1) l'aire hachurée **en vert** correspond à la probabilité

$$P(X < -10)$$

(2) l'aire hachurée **en bleu** correspond à la probabilité

$$P(+10 < X < +15)$$

Remarque : Cette fonction densité de probabilité est une loi de probabilité car l'aire sous la courbe est égale à 1 pour toutes les valeurs de x définies.

Si comme pour les variables aléatoires discrètes, on définit la **fonction de répartition** de X par :

$$\begin{aligned} F_X : \quad & \mathbb{R} \to \mathbb{R} \\ & t \mapsto F_X(t) = P(X < t) \end{aligned}$$

alors la relation entre la **fonction de répartition** F_X et la fonction **densité de probabilité** $f(x)$ est la suivante :

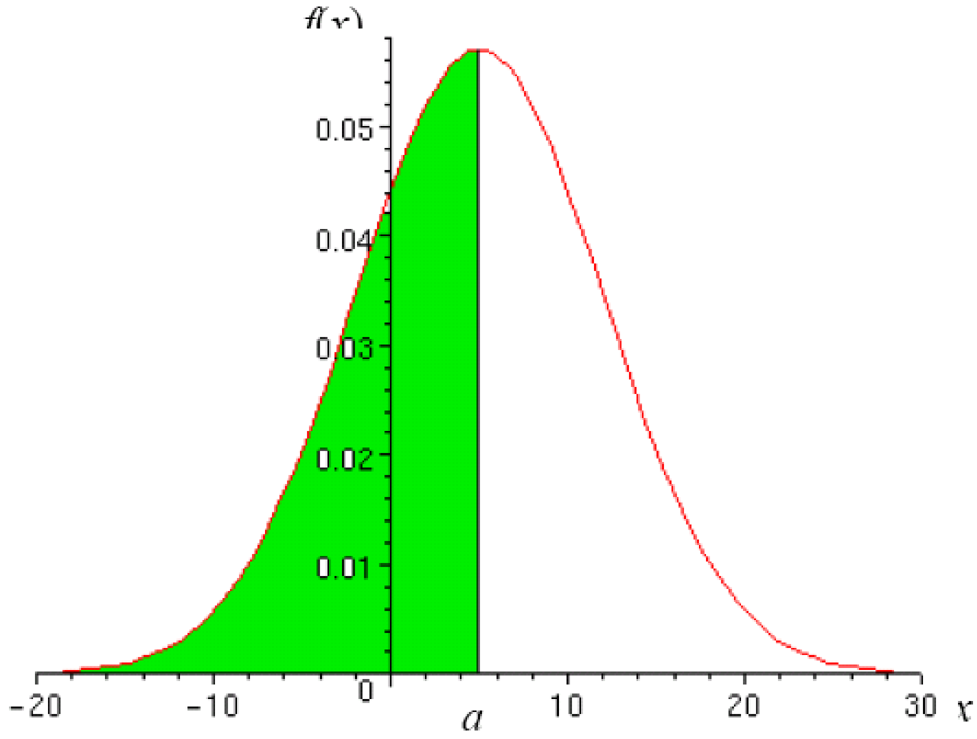
$$\forall\, t \in \mathbb{R} \quad F_X(t) = P(X < t) = \int_{-\infty}^{t} f(x)\, dx$$

Soit X une **variable aléatoire absolument continue** de densité f et de fonction de répartition F_X, alors :

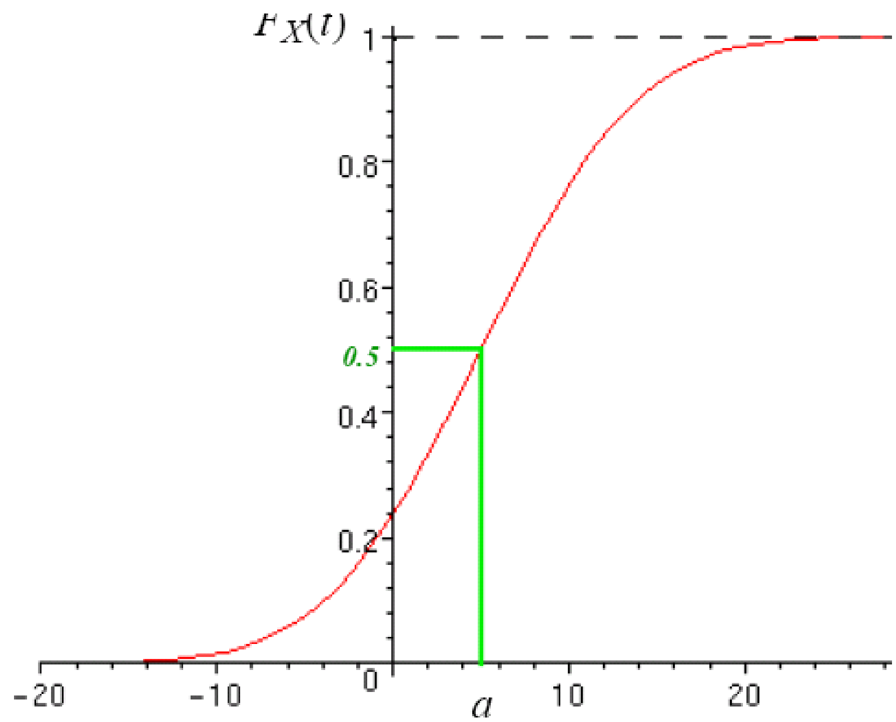
(P_1) $P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f(x)dx$ avec $a < b$

(P_2) $\forall\, a \in \mathbb{R}$ $P(x = a) = 0$ si f est continue à droite du point a.

La **fonction de répartition** correspond aux *probabilités cumulées* associées à la variable aléatoire continue sur l'intervalle d'étude (graphe ci-dessous).



Fonction densité de probabilité $f(x)$



Fonction de répartition F_X

L'aire **hachurée en vert** sous la courbe de la fonction densité de probabilité correspond à la probabilité $P(X < a)$ et vaut **0,5** car ceci correspond exactement à la moitié de l'aire totale sous la courbe. Cette probabilité correspond à la valeur de la fonction de répartition au **point** d'**inflexion de la courbe** (voir cours analyse).

Soit F_X la fonction de répartition d'une variable aléatoire **absolument continue** X alors :

(P₁) F_X est continue sur $\mathbb{R}$, dérivable en tout point où f est continue et alors $F_X' = f$

(P₂) F_X est croissante sur $\mathbb{R}$

(P₃) F_X est à valeurs dans $[0,1]$

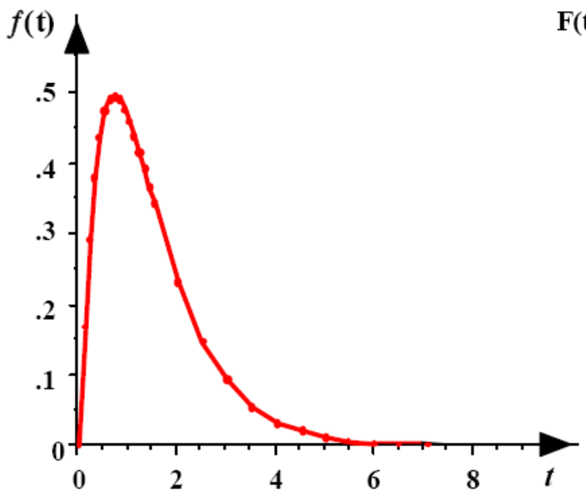
(P₄) $\lim\limits_{t \to -\infty} F_X(t) = 0$ et $\lim\limits_{t \to +\infty} F_X(t) = 1$

Exemple :

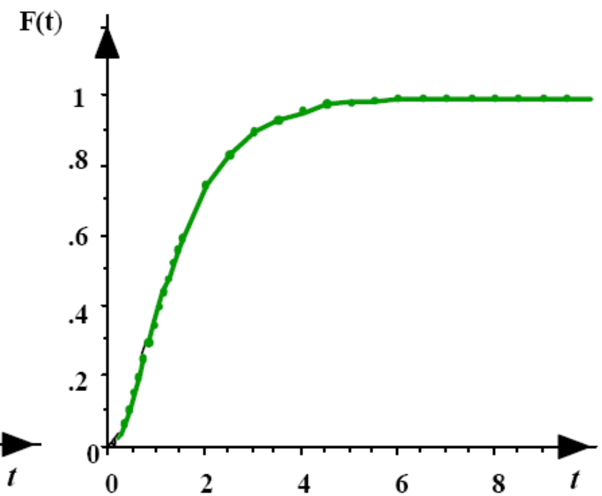
Dans une population de **canards colverts**, lors d'une alerte, l'ensemble des individus quittent leur lieu de repos. Ainsi à $t = 0$, la surface de l'étang est déserte et la probabilité qu'un canard regagne l'étang entre les temps t_1 et t_2 (en minutes) est donnée par :

$\int_{t_1}^{t_2} f(t)dt$ avec $f(t) = 2e^{-t} - 2e^{-2t}$ qui représente la fonction **densité de probabilité**.

La primitive de $f(t)$, $F_T(t)$, **fonction de répartition** est de la forme :



Fonction de densité de probabilité



Fonction de répartition

L'évolution de la recolonisation de l'étang par les *canards colverts* en fonction du temps est donnée par la **courbe rouge**. On observe ainsi que plus de 50 % des canards se posent sur l'étang au cours des 2 premières minutes qui suivent l'alerte. Au bout de 7 minutes, tous les canards ont regagné l'étang. La distribution des probabilités cumulées est donnée sur la **courbe verte.**

Si X est une **variable aléatoire discrète** définie sur un univers probabilisé Ω, on appelle espérance de X, le réel défini par : $E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega)$

Remarque : Si $X(\Omega)$ est infini, on n'est pas sûr que l'espérance existe. L'espérance mathématique est également notée $\mu(X)$, μ_X ou encore μ si aucune confusion n'est à craindre.

Nous pouvons donner une autre définition de l'espérance d'une **variable aléatoire discrète** X si à $\omega \in \Omega$, on associe l'image x telle que $X(\omega) = x$.

Théorème :

Si X est une **variable aléatoire discrète** de loi de probabilité $(x_i, p_i)_i$ définit sur un nombre fini (n) d'évènements élémentaires alors :

$$E(X) = \sum_{i=1}^{n} x_i p_i$$

Si X est une **variable aléatoire absolument continue** de densité f, on appelle **espérance** de X, le réel $E(X)$, défini par : $$E(X) = \int_{-\infty}^{+\infty} x\, f(x)dx$$
si cette intégrale est **convergente**.

Exemple :

Si on reprend l'exemple de la recolonisation de l'étang par les **canards colverts**, la durée moyenne pour la recolonisation est :

$$E(T) = \int_0^{+\infty} t f(t)dt = \int_0^{+\infty} t(2e^{-t} - 2e^{-2t})dt = 3/2$$ (voir **Résultat**)

Sous ce modèle, la **durée moyenne de recolonisation** pour l'ensemble de la population de canards colverts est de **1,5 minutes.**

Remarque : Dans cet exemple, la variable étudiée t ne peut prendre que des valeurs dans $[0, +\infty[$

Les propriétés de **l'espérance** valent aussi bien pour une variable aléatoire discrète ou une variable aléatoire absolument continue.

Si X et Y sont deux variables aléatoires définies sur un même univers Ω, admettant une espérance, alors :

(P_1) $E(X+Y)=E(X)+E(Y)$

(P_2) $E(aX)=aE(X) \quad \forall a \in \mathbb{R}$

(P_3) Si $X \geq 0$ alors $E(X) \geq 0$

(P_4) Si X est un caractère constant tel que : $\forall \omega \in \Omega \quad X(\omega) = k$ alors $E(X) = k$

Remarque : Dans le cas continu, $E(X+Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x+y)\, f(xy)dxdy$. La propriété P_1 est vérifiée quelques soient les relations de dépendance ou d'indépendance statistique entre les deux variables.

5. Variance Mathématique

La **variance** d'une variable aléatoire $V(X)$ est l'espérance mathématique du carré de l'écart à l'espérance mathématique. C'est un paramètre de dispersion qui correspond au **moment centré** d'ordre 2 de la variable aléatoire X. C'est l'équivalent de la **variance observée** S^2. En effet lorsque le nombre d'épreuves n est grand, S^2 tend vers $V(X)$ (voir **estimation**).

Si X est une variable aléatoire ayant une espérance $E(X)$, on appelle **variance** de X le réel : $V(X) = E([X - E(X)]^2)$

$$\boxed{V(X) = E(X^2) - [E(X)]^2}$$

Remarque : Si $X(\Omega)$ est infini, il n'est nullement évident que $V(X)$ existe. De plus comme $[X - E(X)]^2 \geq 0$ nécessairement $V(X) \geq 0$. Par définition, une **variance est toujours positive**.
La variance est également notée σ^2 si aucune confusion n'est à craindre.

Si X est une variable aléatoire ayant une variance $V(X)$, on appelle **écart-type** de X, le réel : $$\sigma(X) = \sqrt{V(X)}$$

Remarque : L'écart-type permet de disposer d'un paramètre de dispersion qui s'exprime dans les **mêmes unités** que la variable aléatoire elle-même.
Le terme « écart-type » se traduit en anglais par le faux-ami « standard deviation ».

Si X est une variable aléatoire **discrète** de loi de probabilité $(x_i, p_i)_i$ définie sur un nombre fini (n) d'évènements élémentaires alors la variance est égale à :

$$V(X) = \sum_{i=1}^{n} (x_i - E(X))^2 p_i = \sum_{i=1}^{n} x_i^2 p_i - E(X)^2$$

Si X est une variable aléatoire **continue** donnée par sa densité de probabilité alors la variance de X est le nombre réel positif tel que :

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - E(X)^2$$