

In [1]:

Ce cours a été régénéré le 2018-09-06 00:48:31.460155

## Représenter une information

In [2]:

### Activité : Échange d'information

Proposez plusieurs façons de représenter un nombre, dont au moins une qui n'est pas la représentation décimale habituelle. Donnez la représentation par exemple pour le nombre 9. Est-ce que votre représentation a des inconvénients ?

Je choisis la représentation suivante :

---

Le nombre 9 se représente :

---

Les inconvénients de cette représentation sont par exemple :

---

In [3]:

### Activité : Devinette ?

Travaillez en paires (ou triplètes). Proposez une méthode pour transmettre d'une personne à une autre le résultat d'un lancer de dé (lancer caché par la première personne, la deuxième doit pouvoir énoncer le résultat). Votre méthode fonctionne-t-elle si le dé comporte 20 faces ? Et si le dé est à six faces mais étiqueté par des couleurs ?

Choix : 1, 2, 3, 4, 5, 6

Essayez de transmettre : 6

Essayez de transmettre : 3

Essayez de transmettre : 5


Choix : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

Essayez de transmettre : 15

Essayez de transmettre : 6

Essayez de transmettre : 17

Choix : , , , , , 

Et ensuite, essayez de transmettre : 

**Information** Une information est une donnée que l'on peut interpréter pour se construire une représentation du monde sur laquelle on peut agir.

Claude Shannon a été l'un des premiers à définir l'information comme une quantité mesurable.

△ L'information de Shannon n'est pas associé au sens ou à la cognition.

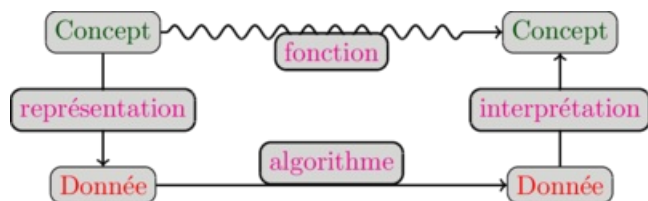
- Il s'est intéressé à quantifier des sources aléatoires
- L'information diminue l'incertitude sur une source aléatoire
- On mesure donc la quantité d'information relative à un événement
- Par exemple, si on a six possibilités pour un dé, l'information permet de savoir quelle face ; ou au moins d'éliminer des possibilités

Il est possible de représenter différents concepts sur un seul schéma.

Ce que l'on aime pouvoir faire, c'est appliquer une *fonction* à des *concepts*.

Pour cela, il faudra choisir une représentation du concept de départ, transformer le concept en *données* utilisables, appliquer un *algorithme* sur les *données*, et faire l'*interprétation* des données résultantes (le contraire de la représentation initiale).

In [4]:



On ne sait pas manipuler le *sens* des choses.

Un même concept peut avoir plusieurs représentations très distinctes. Par exemple, le nombre 14 ([https://fr.wikipedia.org/wiki/Syst%C3%A8me\\_d%C3%A9cimal](https://fr.wikipedia.org/wiki/Syst%C3%A8me_d%C3%A9cimal)), le nombre XIV ([https://fr.wikipedia.org/wiki/Num%C3%A9ration\\_romaine](https://fr.wikipedia.org/wiki/Num%C3%A9ration_romaine)), le nombre ||||| ([https://fr.wikipedia.org/wiki/Syst%C3%A8me\\_unaire](https://fr.wikipedia.org/wiki/Syst%C3%A8me_unaire)) sont toutes des représentations du même concept.

L'informaticien doit connaître les principales représentations/interprétations des concepts classiques.

Nous verrons dans ce cours :

- les nombres
- les valeurs analogiques (concepts inexploitable) et comment les traduire en nombre
- les textes
- quelques éléments sur les média structurés

Mais l'essentiel est de comprendre que l'ordinateur (comme toute machine) n'est capable que de manipuler des choses dans l'univers des *données* et non pas des concepts.

## Ce qu'on ne peut pas représenter

Les systèmes d'informations (et les ordinateurs en particulier) ne sont pas équipés pour traiter n'importe quel type de données. Toutes les informations sont représentées sous forme de nombres (entiers ou à précision limitée) pour être traitées par les ordinateurs. Le monde réel est *analogique*, la représentation des ordinateurs est *numérique* (ou digitale).

In [5]:

Répondez sur chacun de ces éléments par :

- **n** si la donnée est numérique
- **a** si elle est intrinsèquement analogique
- **p** si elle est analogique mais que la précision des mesures fait que c'est de fait numérique

Le fait d'avoir un rendez-vous un certain jour à une certaine heure [anp] >

\_\_\_\_\_

La pression de l'air [anp] > \_\_\_\_\_

Le résultat (stable) d'un dé [anp] > \_\_\_\_\_

Votre nom de famille [anp] > \_\_\_\_\_

Votre nombre de frères et sœurs [anp] > \_\_\_\_\_

Votre taille [anp] > \_\_\_\_\_

La couleur de vos yeux [anp] > \_\_\_\_\_

Les sentiments que vous nourrissez à l'égard de vos parents [anp] >

\_\_\_\_\_

**\*\*Vous avez répondu correctement à 0 questions sur 8**

# Mesure de l'information

L'information est une quantité qui se mesure.

La mesure de base est le *bit*, abréviation de *binary digit* (chiffre binaire). C'est la quantité d'information qui permet de déterminer exactement la réponse à une question binaire (oui/non).

Pour exprimer  $k$  choix possibles distincts, il faut  $\lceil \log_2(k) \rceil$  bits distincts. ( $\lceil x \rceil$  est l'arrondi par dessus).

Dit autrement,  $k$  bits d'information permettent de distinguer  $2^k$  choix.

Cette nature multiplicative de la mesure en bits est très importante. Elle dérive du fait que quand on a plusieurs choix indépendants, on *multiplie* les possibilités, on ne les additionne pas.

Les fractions de bits sont utilisés dans certains contextes, mais en général on ne s'intéresse qu'à des bits entiers.

In [6]:

## Activité : Le jeu du fakir

Je peux deviner n'importe quel nombre entre 0 et 127.

### Question 1

Est-ce que votre nombre est dans la liste suivante :

1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [o/N]> o

### Question 2

Est-ce que votre nombre est dans la liste suivante :

2, 3, 6, 7, 10, 11, 14, 15, 18, 19, 22, 23, 26, 27, 30, 31, 34, 35, 38, 39, 42, 43, 46, 47, 50, 51, 54, 55, 58, 59, 62, 63, 66, 67, 70, 71, 74, 75, 78, 79, 82, 83, 86, 87, 90, 91, 94, 95, 98, 99, 102, 103, 106, 107, 110, 111, 114, 115, 118, 119, 122, 123, 126, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> n

### Question 3

Est-ce que votre nombre est dans la liste suivante :

4, 5, 6, 7, 12, 13, 14, 15, 20, 21, 22, 23, 28, 29, 30, 31, 36, 37, 38, 39, 44, 45, 46, 47, 52, 53, 54, 55, 60, 61, 62, 63, 68, 69, 70, 71, 76, 77, 78, 79, 84, 85, 86, 87, 92, 93, 94, 95, 100, 101, 102, 103, 108, 109, 110, 111, 116, 117, 118, 119, 124, 125, 126, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [o/N]> o

### Question 4

Est-ce que votre nombre est dans la liste suivante :

8, 9, 10, 11, 12, 13, 14, 15, 24, 25, 26, 27, 28, 29, 30, 31, 40, 41, 42, 43, 44, 45, 46, 47, 56, 57, 58, 59, 60, 61, 62, 63, 72, 73, 74, 75, 76, 77, 78, 79, 88, 89, 90, 91, 92, 93, 94, 95, 104, 105, 106, 107, 108, 109, 110, 111, 120, 121, 122, 123, 124, 125, 126, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [o/N]> o

### Question 5

Est-ce que votre nombre est dans la liste suivante :

16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [o/N]> o

### Question 6

Est-ce que votre nombre est dans la liste suivante :

32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [o/N]> o

### Question 7

Est-ce que votre nombre est dans la liste suivante :

64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [o/N]> o

Votre nombre est 125 !

In [7]:

### Activité : Le jeu du fakir, deuxième version

Je peux deviner n'importe quel nombre entre 0 et 127.

#### Question 1

Est-ce que votre nombre est dans la liste suivante :

1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> n

#### Question 2

Est-ce que votre nombre est dans la liste suivante :

2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> n

#### Question 3

Est-ce que votre nombre est dans la liste suivante :

4, 12, 20, 28, 36, 44, 52, 60, 68, 76, 84, 92, 100, 108, 116, 124

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> n

#### Question 4

Est-ce que votre nombre est dans la liste suivante :

8, 24, 40, 56, 72, 88, 104, 120

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> o

#### Question 5

Est-ce que votre nombre est dans la liste suivante :

24, 56, 88, 120

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> n

#### Question 6

Est-ce que votre nombre est dans la liste suivante :

40, 104

Est-ce que votre nombre est dans la liste ci-dessus ? [0/n]> n

#### Question 7

Est-ce que votre nombre est dans la liste suivante :

72

Est-ce que votre nombre est dans la liste ci-dessus ? [0/N]> o

Votre nombre est 72 !

La première version de ce jeu du fakir fonctionne avec des listes qui sont tout le temps les mêmes, qui peuvent donc être préimprimées.

Il n'y a pas de magie : 7 questions binaires permettent de trouver n'importe quel nombre entre 0 et 127, parce que ces 7 questions, bien posées, permettent de distinguer 128 possibilités ( $2^7$ ).

On pourrait de la même façon faire deviner une couleur parmi 128, un mot parmi 128, etc.

In [8]:

### Activité : Dénombrement d'objets

Combien de bits sont nécessaires pour distinguer un élève parmi 637 ? >

Combien de bits sont nécessaires pour distinguer un article parmi 1251 ? >

Combien de bits sont nécessaires pour distinguer un dessin parmi 4 ? >

Les bits sont regroupés en diverses unités. Le principal multiple du bit est l'octet, qui est un groupe de 8 bits. On utilise aussi les préfixes SI des unités (kilo, méga, giga, etc.).

**Attention :** Au début, le mot *byte* désignait le nombre de bits qui était référencé par une *adresse*. Il désigne maintenant un octet (le mot *octet* existe en anglais britannique, mais aux US il est remplacé par *byte* et signifie un mot de 8 bits). Son abréviation est *B*, à ne pas confondre avec l'abréviation de bit qui est *b* et qui vaut donc 8 fois moins.

Préfixe	Valeur	Ab.	Préfixe	Valeur	Ab.
kilo	$10^3$	k	milli	$10^{-3}$	m
mega	$10^6$	M	micro	$10^{-6}$	$\mu$
giga	$10^9$	G	nano	$10^{-9}$	n
tera	$10^{12}$	T	pico	$10^{-12}$	p
peta	$10^{15}$	P	exa	$10^{18}$	E

Toutefois, pour des raisons historiques, il a été décidé d'utiliser aussi les préfixes SI pour d'autres groupements, donnant naissance à une deuxième classification, nommée « échelle binaire » (la classification habituelle en puissances de 10 étant du coup « l'échelle décimale »). Comme montré par le calcul, ces unités sont proches de leur équivalent décimal (SI) (en effet,  $2^{10} = 1024 \approx 1000 = 10^3$ ).

Préfixe binaire	Valeur	Ab.	Ab. SI	Approximation
kilo	$2^{10}$	k	ki	$\sim 2,5\%$ (=2,4%)
mega	$2^{20}$	M	Mi	$\sim 5\%$
giga	$2^{30}$	G	Gi	$\sim 7,5\%$
tera	$2^{40}$	T	Ti	$\sim 10\%$
peta	$2^{50}$	P	Pi	$\sim 12,5\%$
exa	$2^{60}$	E	Ei	$\sim 15\%$

Le domaine où l'échelle binaire est la plus utilisée est l'expression de la capacité d'éléments mémoires (souvent conjugués aux octets, et non aux bits). D'autres domaines, comme le débit d'information (quantité d'information divisé par un temps), est beaucoup plus souvent exprimé dans l'échelle décimale, de même que dès qu'on utilise des préfixes négatifs.

### Quelques ordres de grandeur

- $10^3$  bits: carte à bande magnétique
- $10^6$  bits: un fax d'une page
- $10^9$  bits: Capacité d'un CD ou du génome humain
- $10^{12}$  bits: Un disque dur moyen en 2008
- $10^{15}$  bits: 1/10\ieme\ taille des serveurs de Google
- $10^{18}$  bits: Tout ce qui est imprimé dans le monde.

Pour le débit, on a plutôt :

- 1 b/s: vieille sonde spatiale (9 b/s), morse (40 b/s)
- $10^3$  b/s: 2G (9,6 kb/s), modems (56 kb/s)
- $10^6$  b/s: ADSL (20 Mb/s)
- $10^9$  b/s: Réseau local Gigabit (1 Gb/s), USB (0,48 Gb/s), Infiniband (60 Gb/s)
- $10^{12}$  b/s: Trafic total USA cumulé sur internet (12 Tb/s)
- $10^{15}$  b/s: Trafic total international sur internet (0.5 Pb/s)

**Note :** le trafic sur 2008 par seconde dépasse le total de 1993 (année)

In [9] :

### Activité : Conversion d'unités

Transformez  $1\,000 \times 2^{12}$  b en ko

Votre proposition > \_\_\_\_\_

Transformez 2 o en b

Votre proposition > \_\_\_\_\_

Transformez  $30 \times 2^{10}$  o en Mio

Votre proposition > \_\_\_\_\_

Transformez 16 000 Mio en b

Votre proposition > \_\_\_\_\_

Transformez 5 120 kio en Mio

Votre proposition > \_\_\_\_\_

Transformez  $60 \times 2^{14}$  Mo en Gib

Votre proposition > \_\_\_\_\_

Au delà des mesures simples d'information, on peut composer les bits avec d'autres unités.

Par exemple, les unités les plus utilisées au-delà de celles déjà vues sont les unités d'information divisées par un temps (débit), et les unités d'information divisées par une surface ou un volume (densité surfacique ou volumique d'information).

### Exercice

Un élément d'ordinateur est capable d'émettre 1024 bits en 0,5 nanosecondes. Quel est le débit (quantité d'information divisée par le temps) de cet élément en bits par secondes ? Quelle est la bonne unité pour ce débit ?

## De l'analogique au digital

Pour tout ce qui est mesurable, mais dont la mesure est analogique, avant de le traiter dans un ordinateur, il faut *acquérir* les données puis les transformer en *nombres* (nombres à précision limitée, ou entiers). Cette opération s'appelle la *numérisation* (on dit aussi la *digitalisation*, mais c'est un anglicisme).

Elle se fait presque toujours de la même façon :

- Filtrage perceptuel physique
- Découpage (volumique) (pour les phénomènes multidimensionnels)
- Échantillonnage (pour les phénomènes temporels)
- Quantification (réduction à un nombre d'états finis)
- Filtrage perceptuel numérique

Les deux premières étapes forment la discrétisation (spatiale ou temporelle), et la troisième la quantification.

### Filtrage perceptuel physique

La première chose à comprendre c'est qu'obtenir une empreinte exacte d'une chose n'est pas possible. Nos instruments de mesure ont toujours des limites.

Lorsque ces limites sont calculées de façon à être en dessous du seuil de perception de l'être humain, mais ne sont pas maximales, on opère un *filtrage perceptuel* sur les données acquises. La connaissance des mécanismes de perception est importante pour établir ces limites. Par exemple, la vitesse optimale de succession des images dans un film ou les fréquences audibles par l'oreille humaine vont limiter l'intérêt de données qui vont au-delà de ces limites.

Toutefois, un filtrage plus faible (donc une mesure plus précise) permet un travail sur des choses au-delà de la perception humaine, ce qui est parfois important (possibilité de zoom sur une image, de vue dans des gammes d'onde lumineuses hors de la vision humaine, d'isolation de son, etc.).

Ce filtrage est considéré comme en dehors de notre domaine informatique, même s'il a des conséquences.

## Découpage volumique

Pour pouvoir manipuler une information qui est continue dans une dimension spatiale, on va prendre des échantillons le long d'un découpage de l'espace (surface, volume, selon les cas). On utilise la plupart du temps une grille régulière (carrée ou cubique) pour faire de l'échantillonnage. À chacun de ces éléments on attribue une valeur unique, découpant ainsi l'image en pixels (ou voxels).

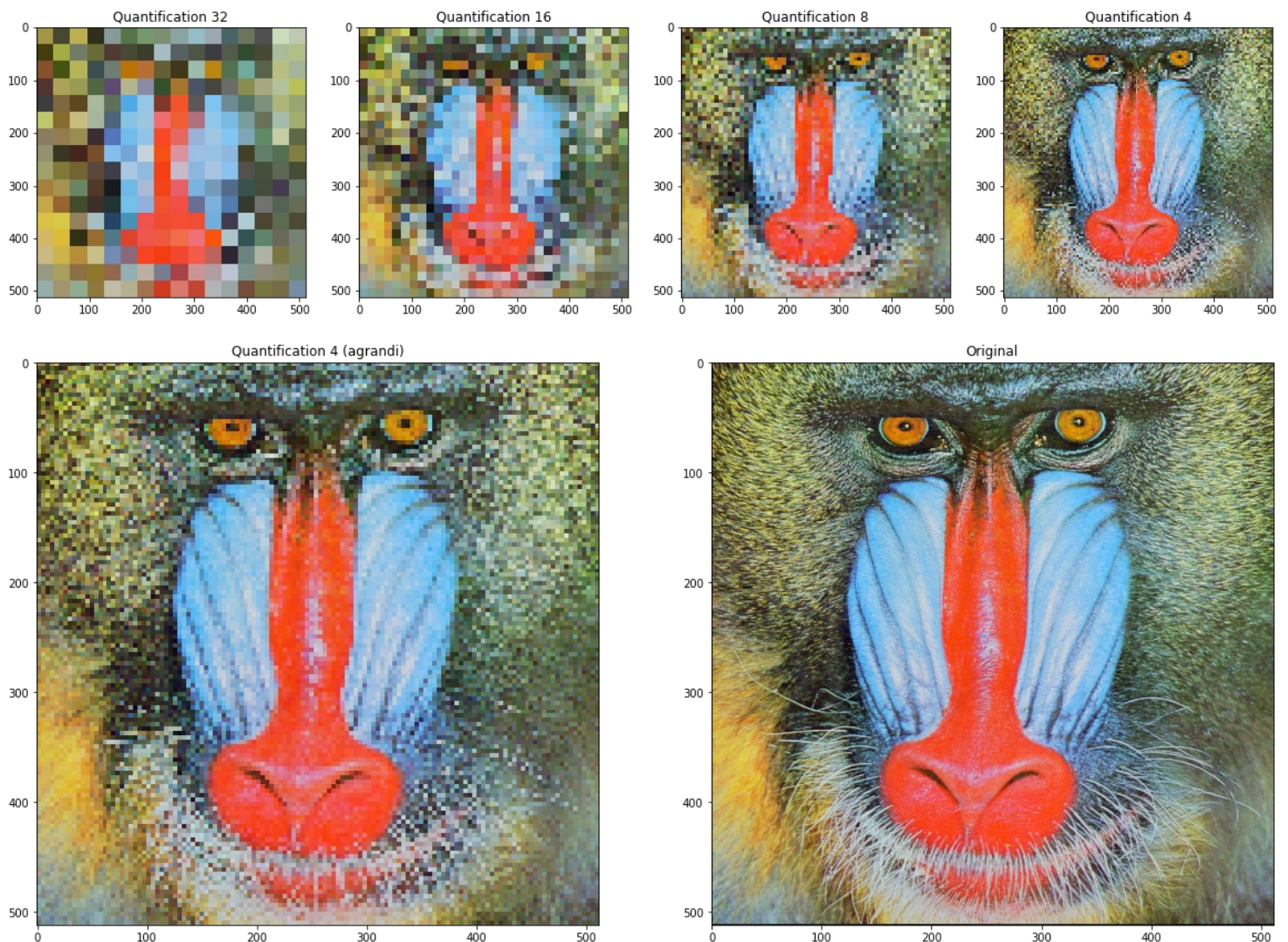
La grille carrée n'est pas universelle : on utilise aussi des grilles rectangulaires, triangulaires. La technique d'origine de la photographie consiste en l'utilisation d'une grille aléatoire (la répartition des grains d'argent au sein du film photographique).

La source est altérée par ce choix. Selon le cas, on peut prendre la valeur comme étant celle mesurée au début de l'intervalle, à la fin, au milieu. Le choix n'a qu'une importance relative : si on ne sait pas à l'avance ce qu'on a mesuré, les propriétés ne changent pas radicalement.

L'information est perdue : une fois qu'on a échantillonné, on ne peut pas retrouver l'information. On peut essayer de la deviner, mais ce n'est pas de la magie ([https://www.youtube.com/watch?v=I\\_8ZH1Ggjk0](https://www.youtube.com/watch?v=I_8ZH1Ggjk0)) et donc on ne pourra pas recréer des détails inhabituels perdus.

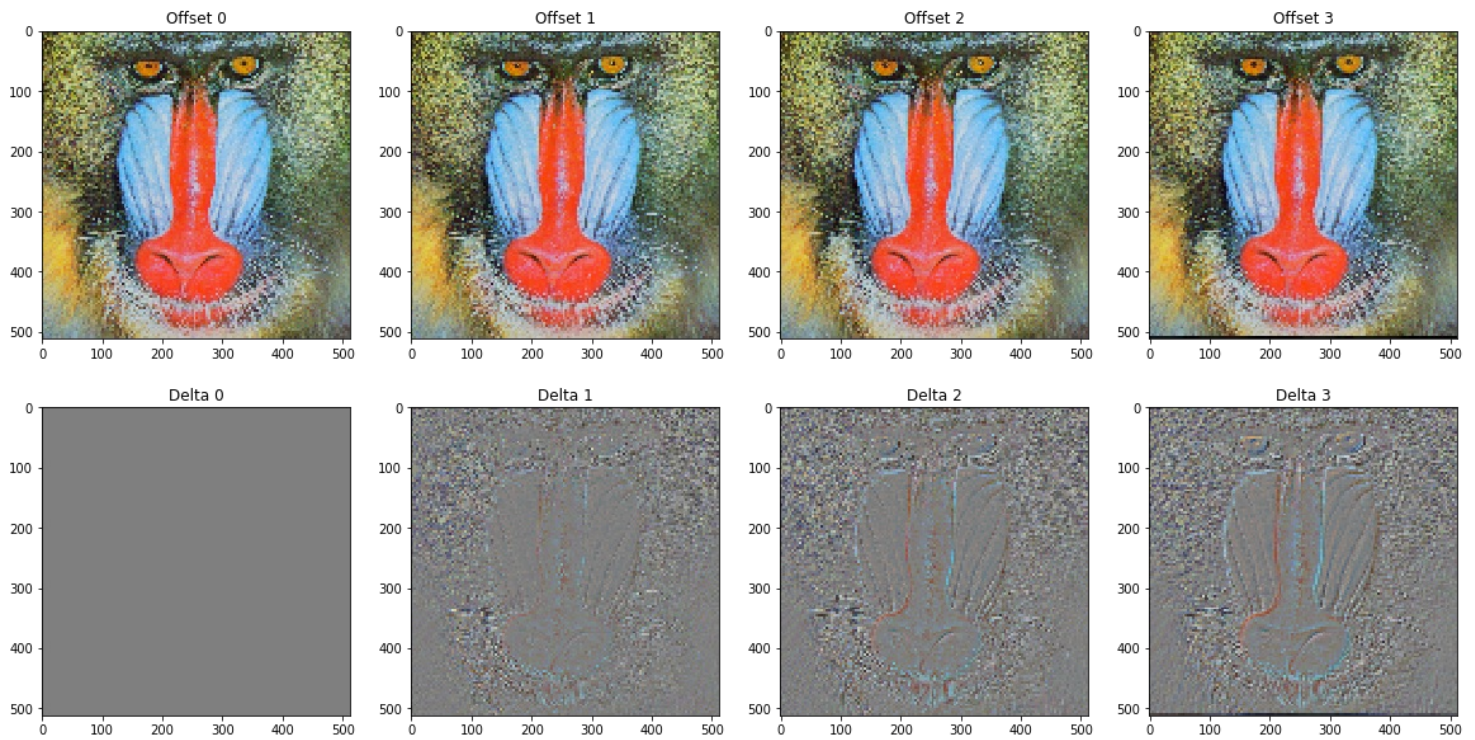
In [10]:

In [11]:





In [12]:



## Échantillonnage temporel

Rappel: une fréquence, c'est une quantité d'événements par unité de temps ( $1\text{Hz} = 1\text{s}^{-1}$ ).

Un signal à un moment donné peut toujours se représenter comme une somme de signaux sinusoïdaux. Si le signal est périodique, les coefficients sont alors constants, et les fréquences à utiliser uniquement des multiples de la fréquence de base.

Un grand miracle du traitement du signal est que ce qui est valable pour un signal périodique peut s'étendre à un signal non-périodique (mais c'est en dehors de la portée de ce cours).

Pour la plupart des signaux, on peut considérer uniquement les contributions à ce signal jusqu'à une *fréquence maximale*, les contributions au-delà étant négligeables. Les instruments font ce filtre naturellement (ou rajoutent du bruit au-delà de certaines fréquences) !

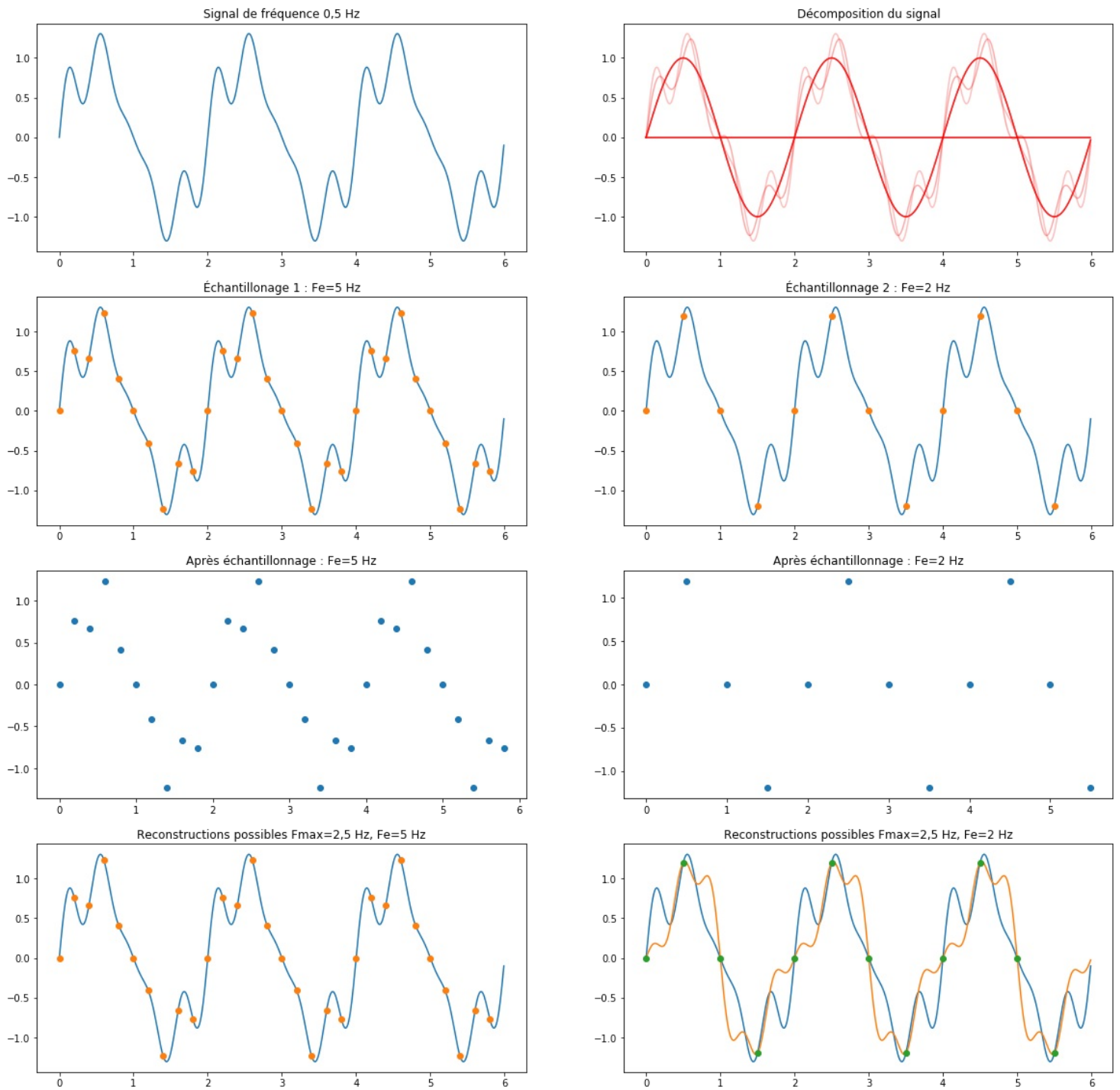
Pour pouvoir reconstruire exactement un signal périodique qui peut être décomposé avec une fréquence maximale, sa fréquence d'échantillonnage  $f_e$  doit vérifier (*Théorème d'échantillonnage de Nyquist-Shannon*):

$$f_e \geq 2f_{\text{Max}}$$

Inversement, on appelle *fréquence de Nyquist* la fréquence maximale autorisée pour un signal à fréquence d'échantillonnage donnée. C'est donc la moitié de la fréquence d'échantillonnage.

**Attention :** Des effets de bruit font qu'il est préférable d'avoir une marge pour reconstruire proprement un signal à cause de la phase suivante (quantification).

In [13]:



## Quantification

Cette opération réduit un signal à des *quanta* (singulier *quantum*) en nombre limité. Le nombre de *quanta* possibles s'appelle la *valence*.

La reconstruction exacte du signal n'est plus possible, mais reste souvent proche de l'original.

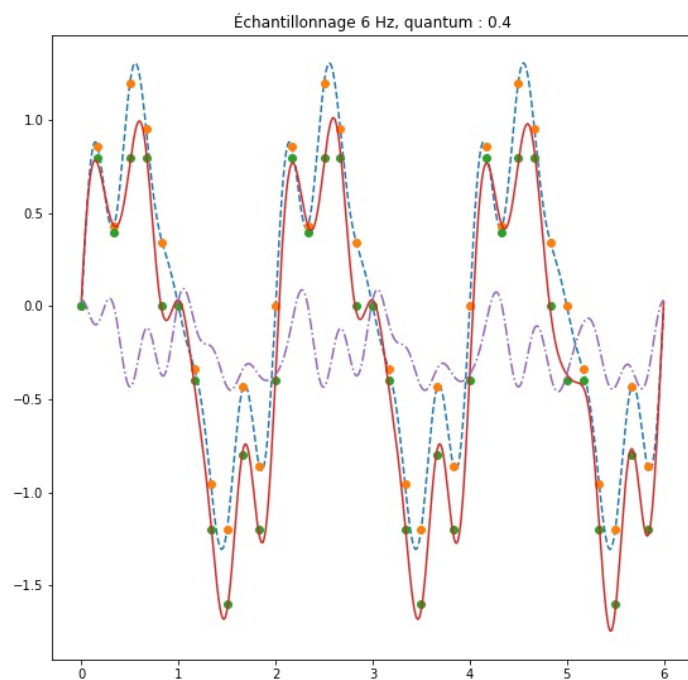
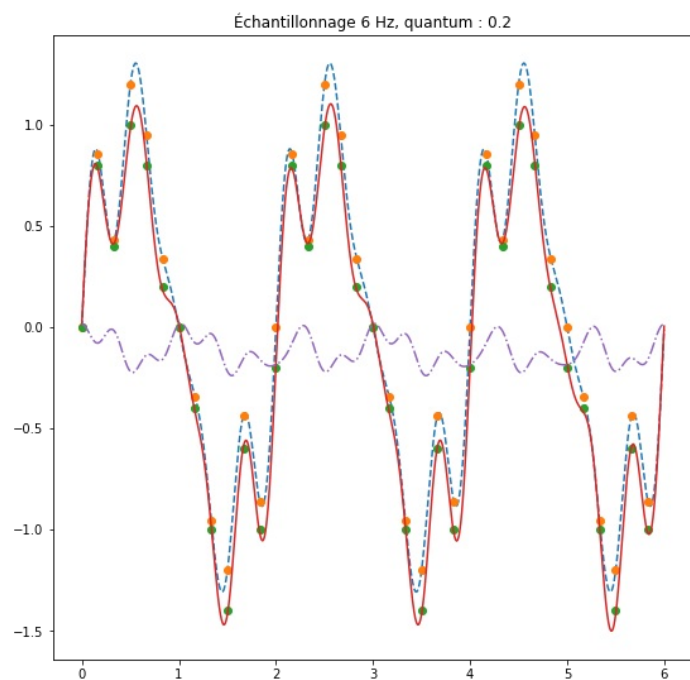
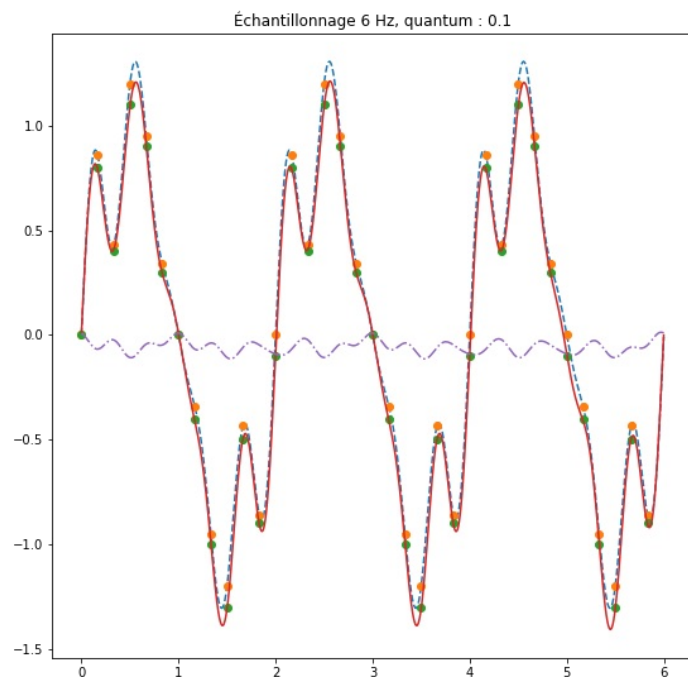
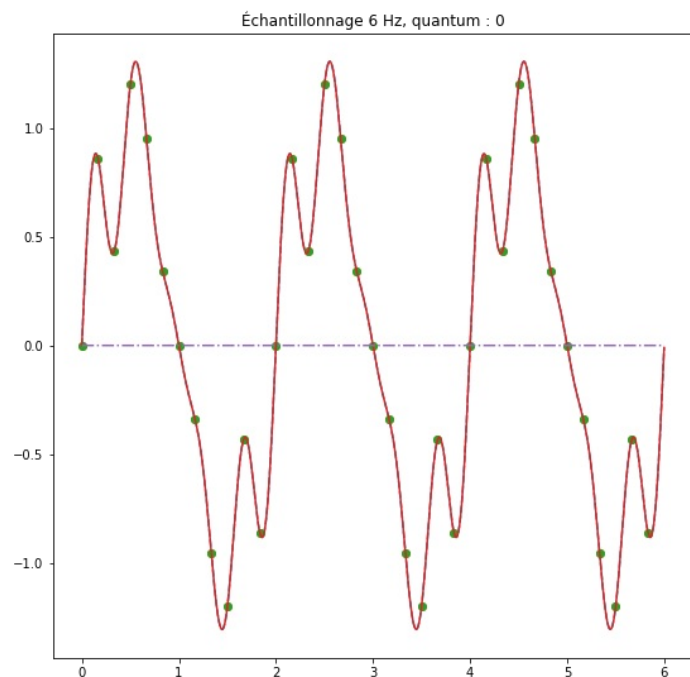
### Exemple : Signal électrique

Une tension électrique comprise entre 0 (large) et 10 V (strict) peut ainsi être réduite à 10 quanta de 1 V : 0 V, 1 V, ..., 9 V.

Le nombre de bits nécessaires pour coder un état du signal peut être exprimé par  $k = \lceil \log_2 V \rceil$ .

L'erreur de reconstruction d'un signal d'origine après quantification s'appelle la *distortion*. La théorie du traitement du signal permet d'étudier précisément cette distortion, mais c'est en dehors du sujet de ce cours.

In [14]:



## Filtrage perceptuel numérique

Lorsque les informations sont destinées à être perçues par des humains directement (cas des médias, notamment), les informations peuvent subir une dernière phase d'élagage qui permet d'enlever encore de l'information.

À ce stade on est déjà dans l'information numérique (nombres stockés avec un nombre fini de bits par nombre)

Comme pour le filtrage perceptuel physique, une bonne connaissance du fonctionnement de la perception humaine est importante.

In [15]:

**Activité : Comprendre la quantification**

Un signal électrique va de 0 V à 0,999 V. Il est quantifié sur un quantum de 0,01 V.

Quel est le nombre de quanta ? > \_\_\_\_\_

Quelle quantité d'information est transportée par chaque quantum ? > \_\_\_\_\_

Ce signal est périodique, et se décompose avec des fréquences maximales qui vont jusqu'à 10 kHz.

Quel est le débit d'information nécessaire pour reconstituer ce signal à l'identique ? > \_\_\_\_\_

Quelle quantité d'information pour enregistrer ce signal pendant une heure (en bits) ? > \_\_\_\_\_

In [16]:

**Activité : Contenance d'un CD audio**

Un CD audio contient de la musique échantillonnée en stéréo sur 16 bits par piste à 44100 Hz (nombre d'échantillons par seconde). Il dure environ 80 minutes. Calculez (de tête) l'ordre de grandeur de la quantité d'information écrite dans un CD audio.

Votre proposition > \_\_\_\_\_

In [17]:

Mode interactif