Statistiques et probabilités
Cours n°3

Guillaume Postic

Université Paris-Saclay, Univ. Evry
Département informatique

Master 1 MIAGE - 2023/2024

Variable aléatoire discrète 🆚 continue

- VA discrète prends un ensemble fini de valeurs distincts
  - Exemple {1, 2, 3, 4, 5, 6} pour un dé à 6 faces

- VA continue peut prendre une infinité de valeurs possibles dans l'intervalle de définition
  - Exemples : longueur, temps, température

- VA discrètes et continues sont des **VA quantitatives**
  - Par opposition aux VA **qualitatives**...

Variable aléatoire discrète

Soit une variable aléatoire X qui assigne un nombre a pour la réalisation d'un évènement ω :

- La fonction de masse de X est donnée par

$$p_X(a) \text{ ou } p(a) = P(X = a)$$

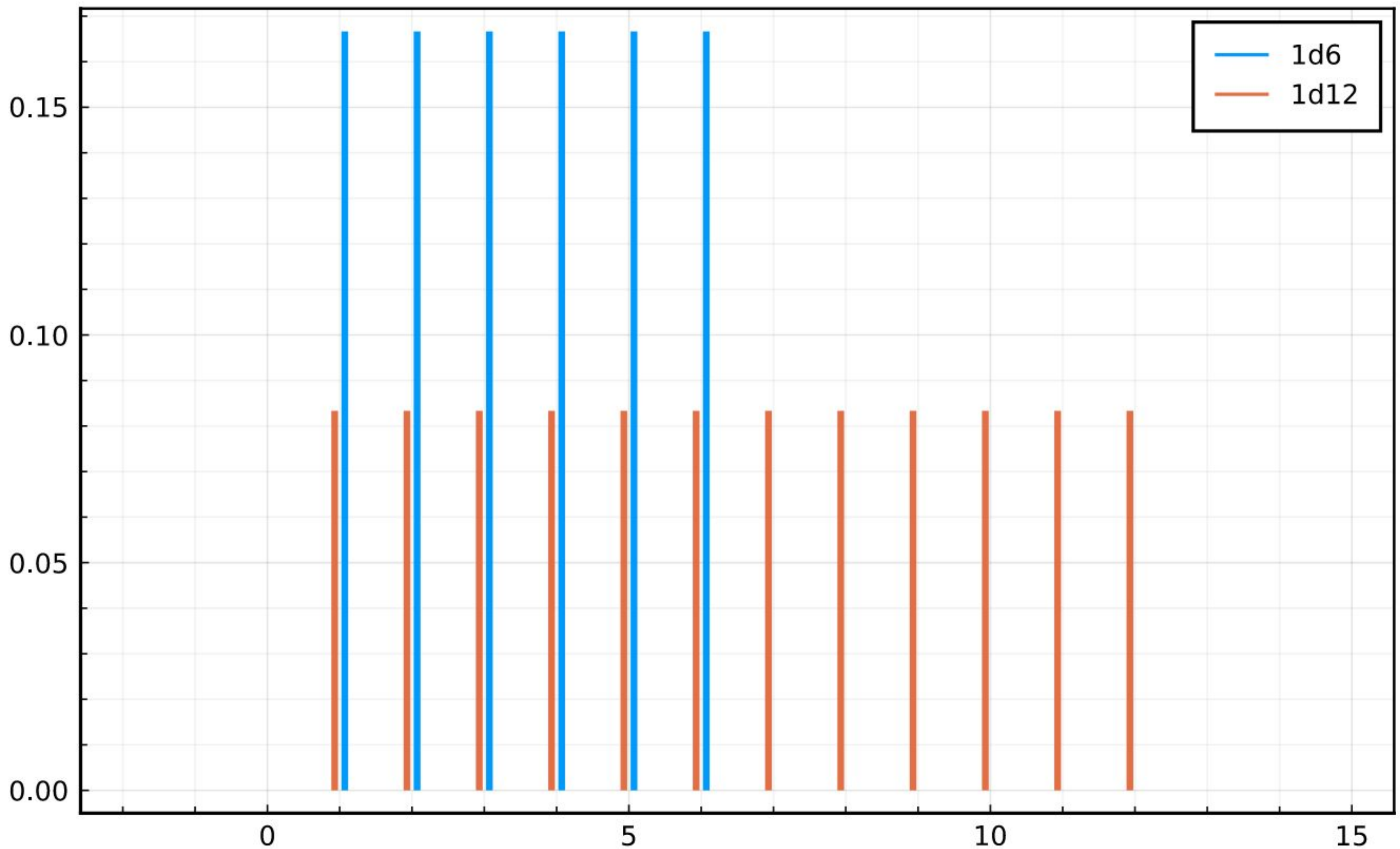
  Note : elle défini la loi de probabilité discrète suivie par X

- La fonction de répartition (fonction de distribution cumulative) est donnée par

$$F_X(a) \text{ ou } F(a) = P(X \leq a)$$

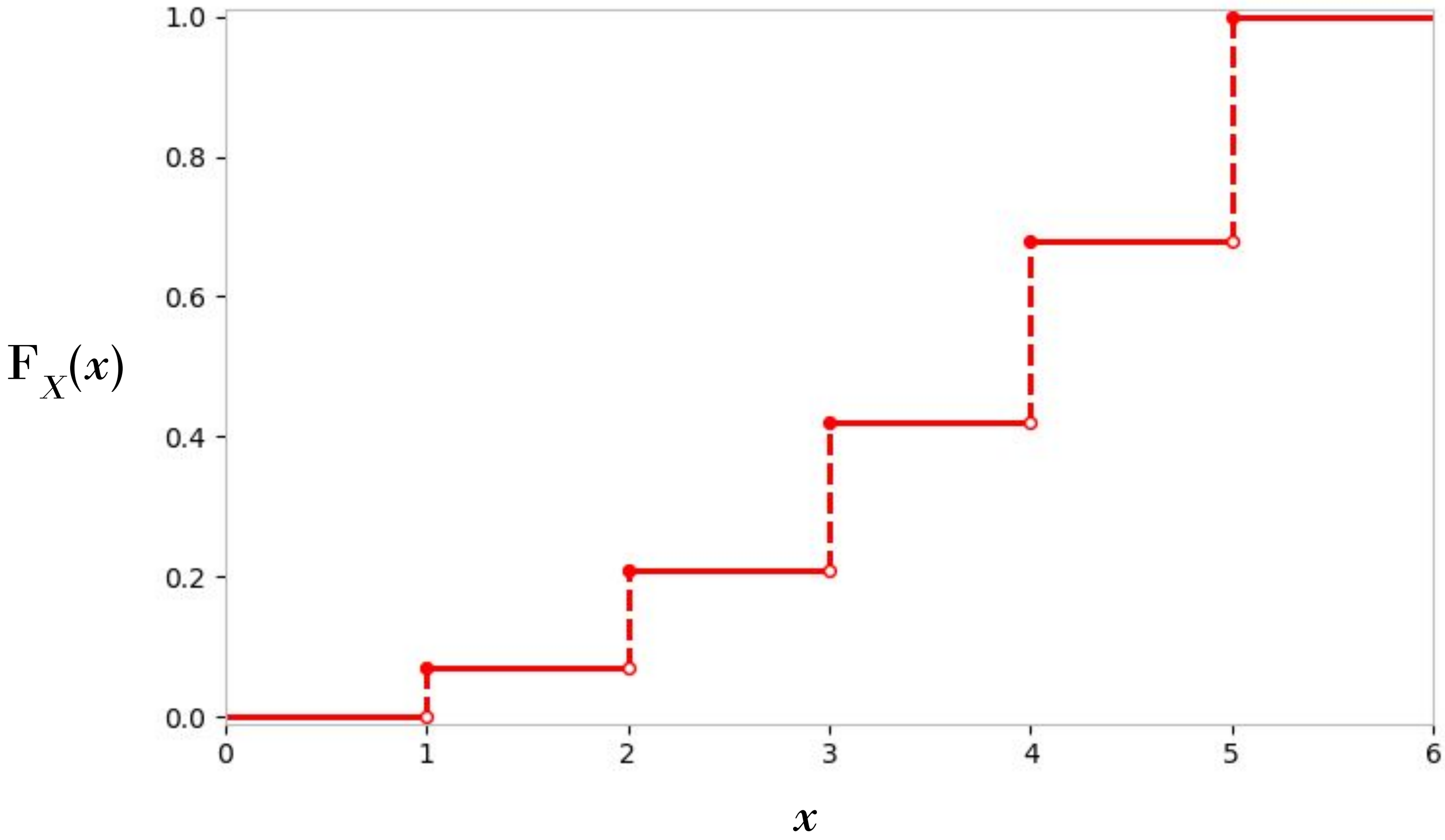
Variable aléatoire discrète

Exemple : fonction de masse du résultat d'un jet de dé à [**six**|**douze**] faces



Variable aléatoire discrète

Exemple : fonction de répartition du résultat d'un jet de dé à six faces

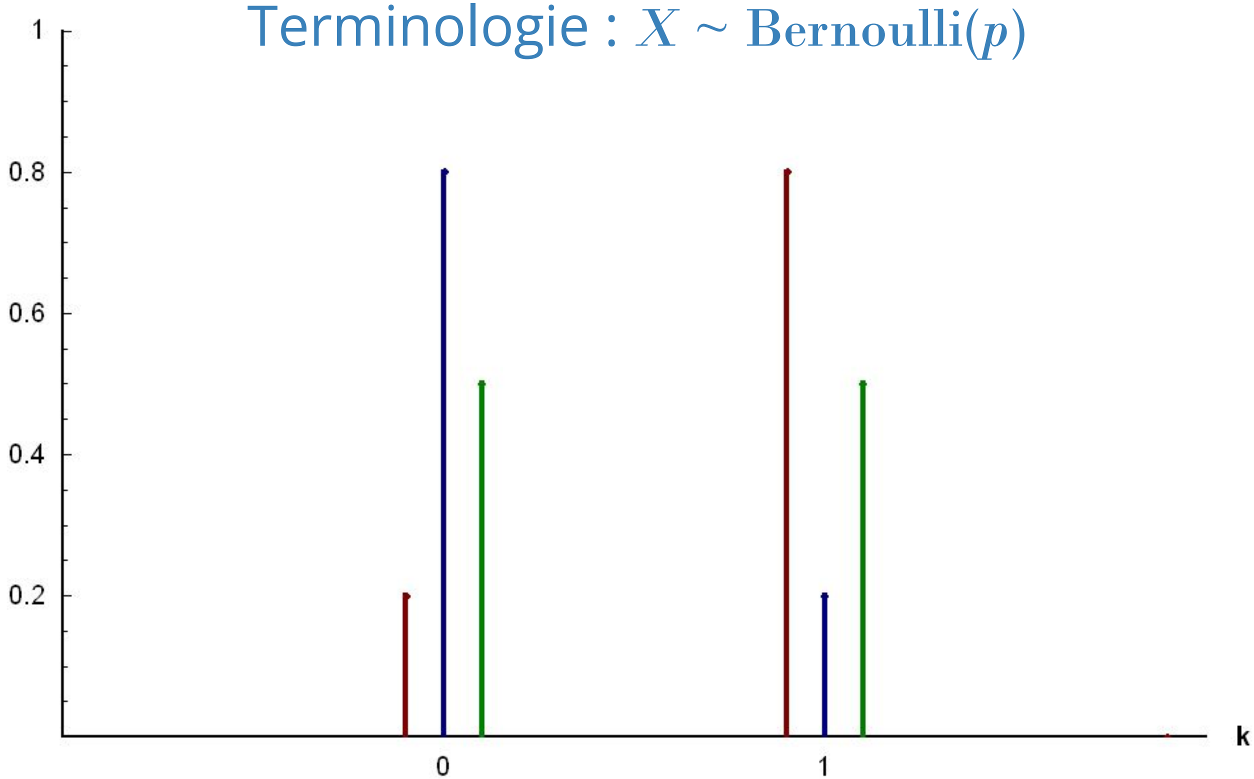


Distributions discrètes (1)

Loi de Bernoulli

Loi de probabilité d'une variable aléatoire discrète qui prend la valeur 1 avec la probabilité p et 0 avec la probabilité q = 1 – p.

Terminologie : $X \sim \mathbf{Bernoulli}(p)$



Distributions discrètes (2)

Loi binomiale

Soit X_1, X_2, … , X_n *n* variables aléatoires de Bernoulli de paramètre *p*, indépendantes et identiquement distribuées, alors leur somme *N* est une variable aléatoire, qui suit la loi binomiale :

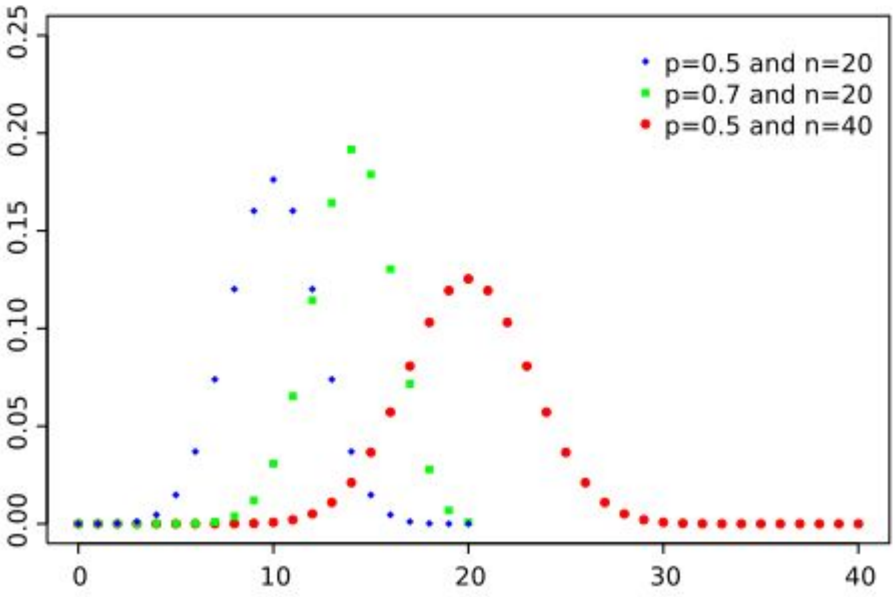
$$N = \sum_{k=1}^{n} X_k \sim \mathcal{B}(n, p)$$

Sa fonction de masse donne la probabilité d'obtenir *k* succès après *n* épreuves de Bernoulli :

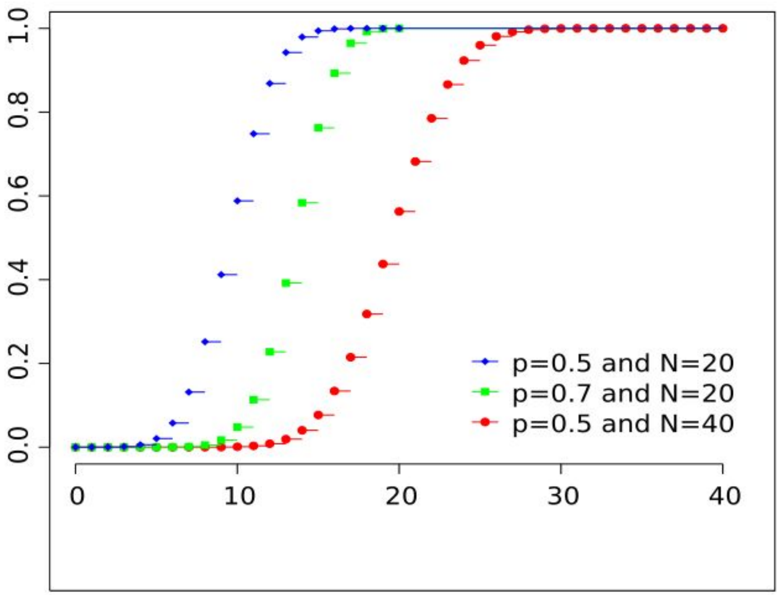
$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Distributions discrètes (2)

Loi binomiale



Fonction de masse



Fonction de répartition

Distributions discrètes (2)

Loi binomiale

$$\mathrm{E}(\mathrm{X}) = \boldsymbol{np}$$

$$\mathrm{V}(\mathrm{X}) = \boldsymbol{npq}$$

Distributions discrètes (3)

Loi géométrique

Selon la convention choisie :

- la loi du nombre *X* d'épreuves de Bernoulli indépendantes de probabilité de succès $p \in]0,1[$ (ou $q = 1 - p$ d'échec) nécessaire pour obtenir le premier succès. *X* est la variable aléatoire donnant le rang du premier succès. Le support de la loi est alors {1, 2, 3, ...}.

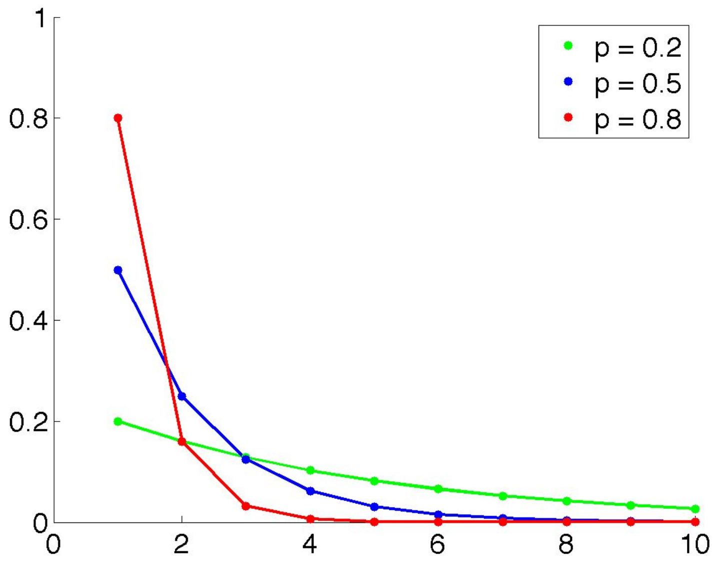
$$\mathbb{P}(X = k) = q^{k-1}p.$$

- La loi du nombre $Y = X - 1$ d'échecs avant le premier succès. Le support de la loi est alors {0, 1, 2, 3, ...}.

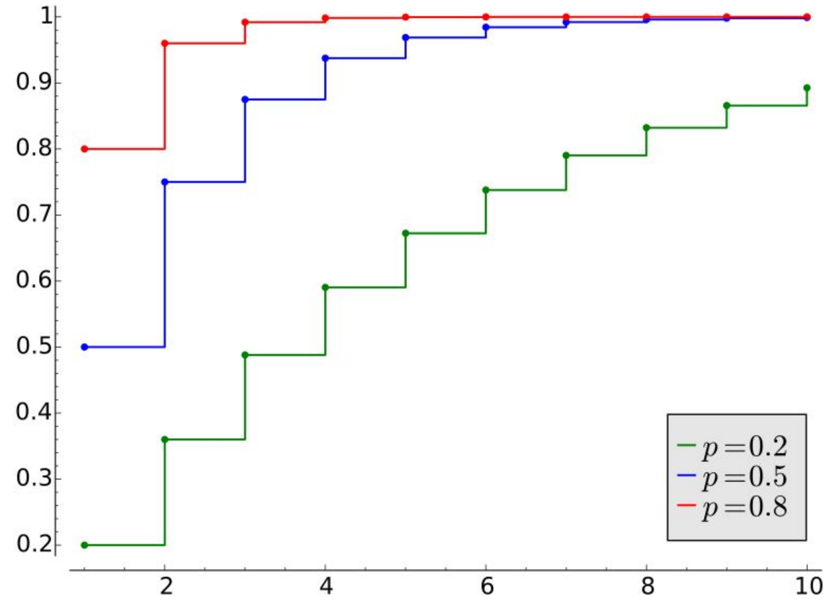
$$\mathbb{P}(Y = k) = q^{k}p.$$

Distributions discrètes (3)

Loi géométrique



Fonction de masse



Fonction de répartition

Distributions discrètes (3)

Loi géométrique

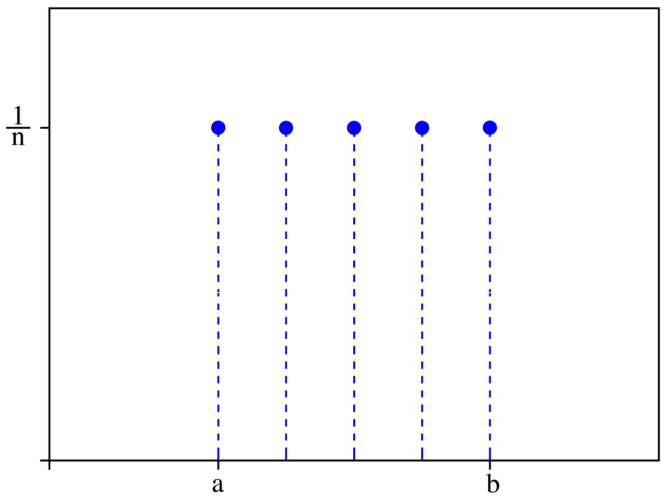
$$E(X) = 1 \,/\, p$$

$$V(X) = q \,/\, p^2$$

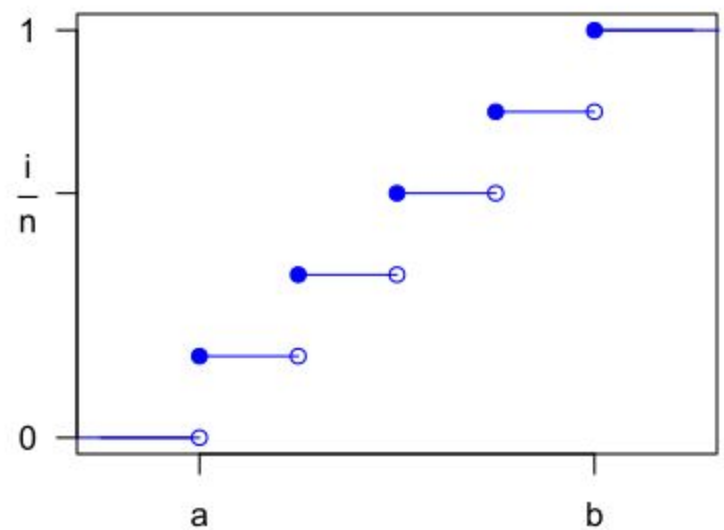
Distributions discrètes (4)

Loi uniforme

Loi de probabilité discrète indiquant une probabilité de se réaliser identique (équiprobabilité) à chaque valeur d'un ensemble fini de valeurs possibles.



Fonction de masse



Fonction de répartition

Distributions discrètes (4)

Loi uniforme

$$E(X) = (a + b) / 2$$

$$V(X) = (n^2 - 1) / 12$$

Distributions discrètes (5)

Loi de Poisson

Loi de probabilité discrète qui décrit le comportement du **nombre d'événements se produisant dans un intervalle de temps fixé**, si ces événements se produisent avec une fréquence moyenne ou espérance connue, et indépen damment du temps écoulé depuis l'événement précédent.

Si le nombre moyen d'occurrences dans un intervalle de temps fixé est λ, alors la probabilité qu'il existe exactement *k* occurrences (*k* étant un entier naturel, *k* = 0, 1, 2...) est

$$p(k) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

Distributions discrètes (5)

Loi de Poisson

Si l'intervalle de temps n'est pas fixé, alors

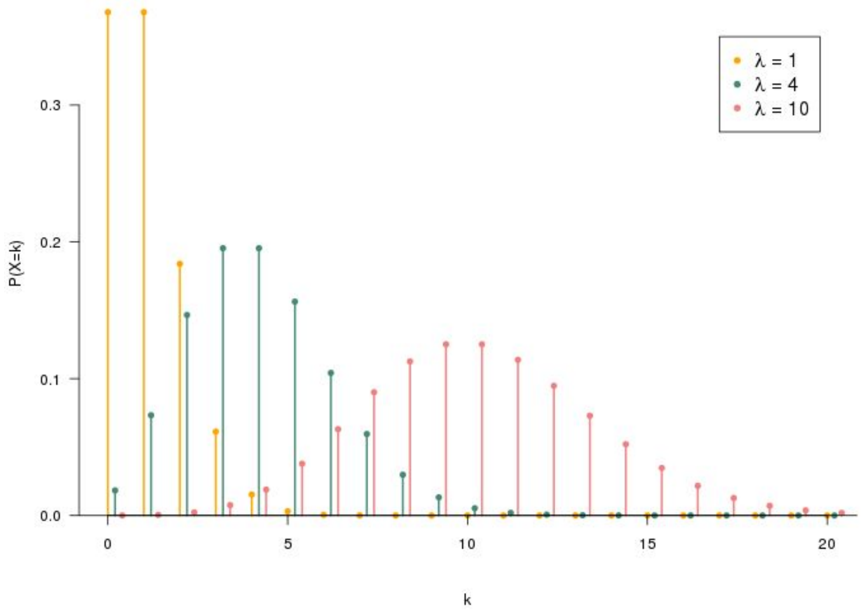
- **λ** est la **fréquence** à laquelle l'évènement survient
- le nombre N_t d'occurrences dans un intervalle de longueur t suit une loi de Poisson « d'intensité » λt :

$$\mathbb{P}(N_t = k) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

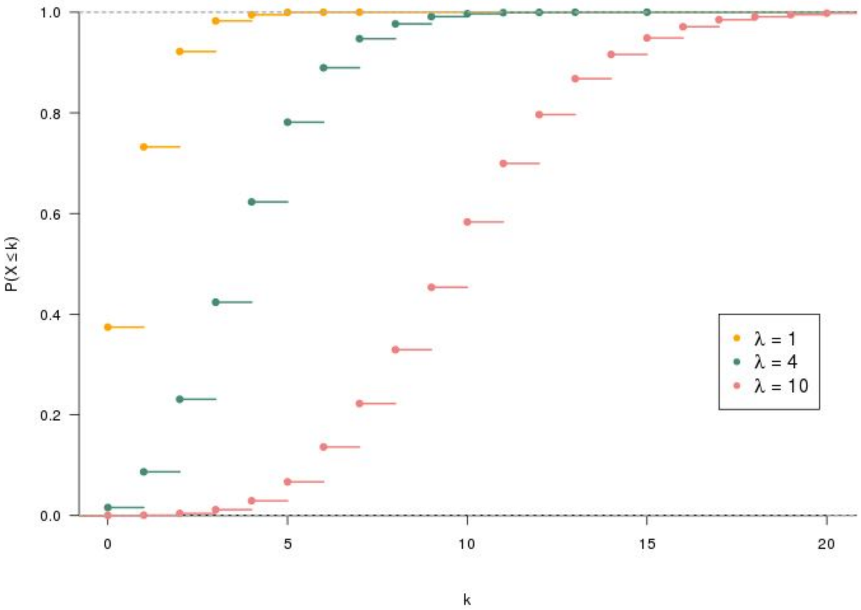
Distributions discrètes (5)

Loi de Poisson

$$X \sim \text{Pois}(\lambda)$$



Fonction de masse



Fonction de répartition

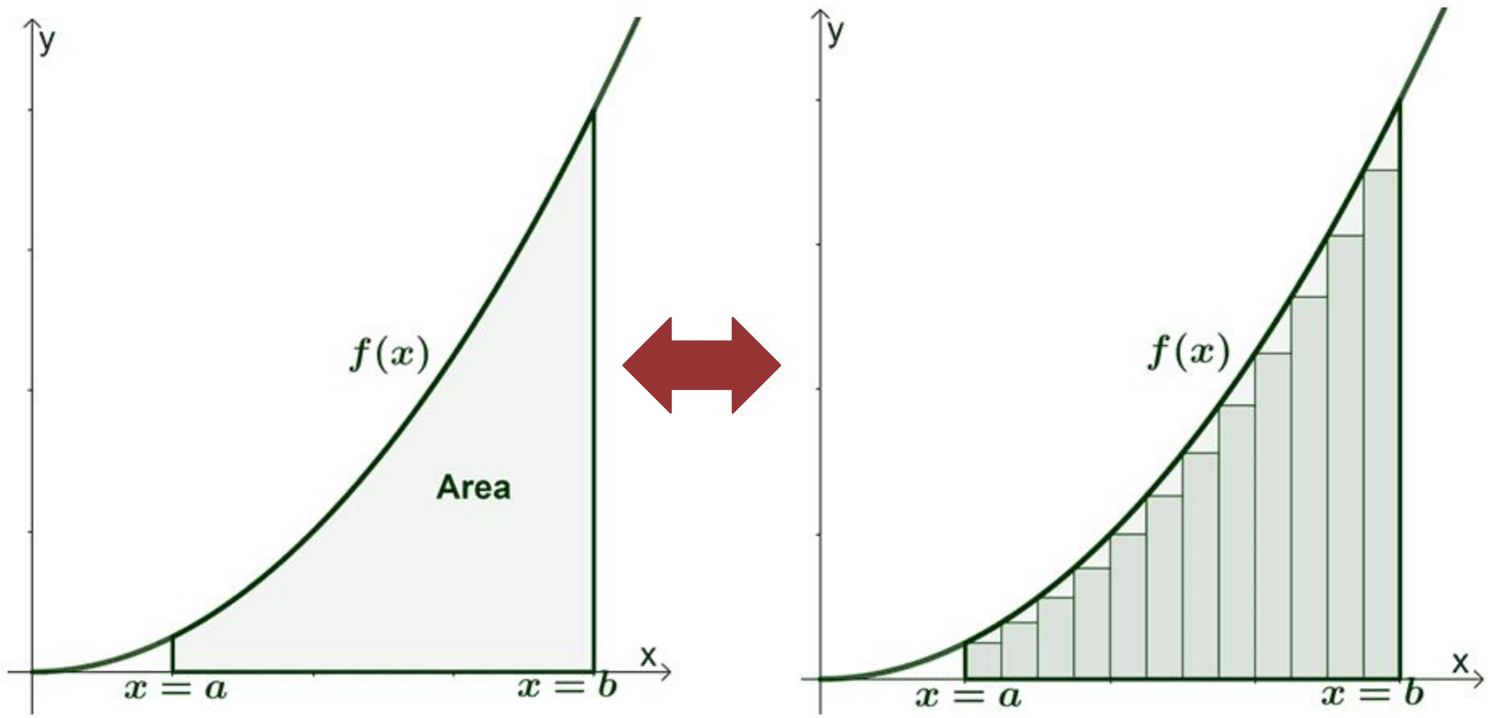
Distributions discrètes (5)

Loi de Poisson

$$E(X) = \lambda$$

$$V(X) = \lambda$$

Rappels



Rappels

- La somme de Riemann S de f sur $[a, b]$ est :

$$S = \sum_{i=1}^{n} f(x_i^*)\,\Delta x_i$$

- Si le **pas de la subdivision** tend vers zéro (*i.e.* si n tend vers $+\infty$), alors la somme converge vers :

$$\int_a^b f(t)\,\mathrm{d}t$$

Variable aléatoire continue

- Intervalle de valeurs continu :
  - e.g. [0, 1], [a, b], [0, ∞), (−∞, ∞)

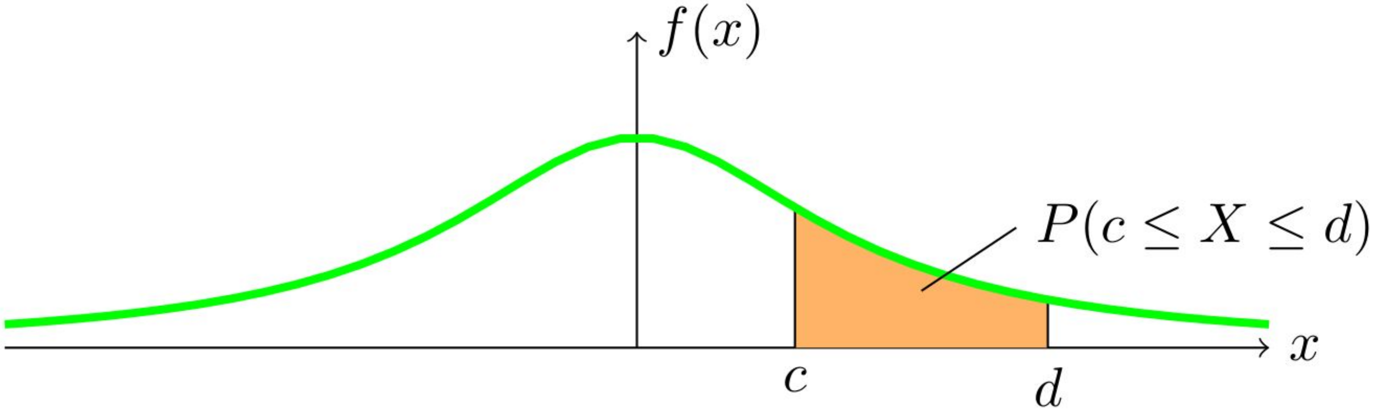
- Fonction de densité ou densité de probabilité

$$f(x) \geq 0; \quad P(c \leq x \leq d) = \int_c^d f(x)\,dx$$

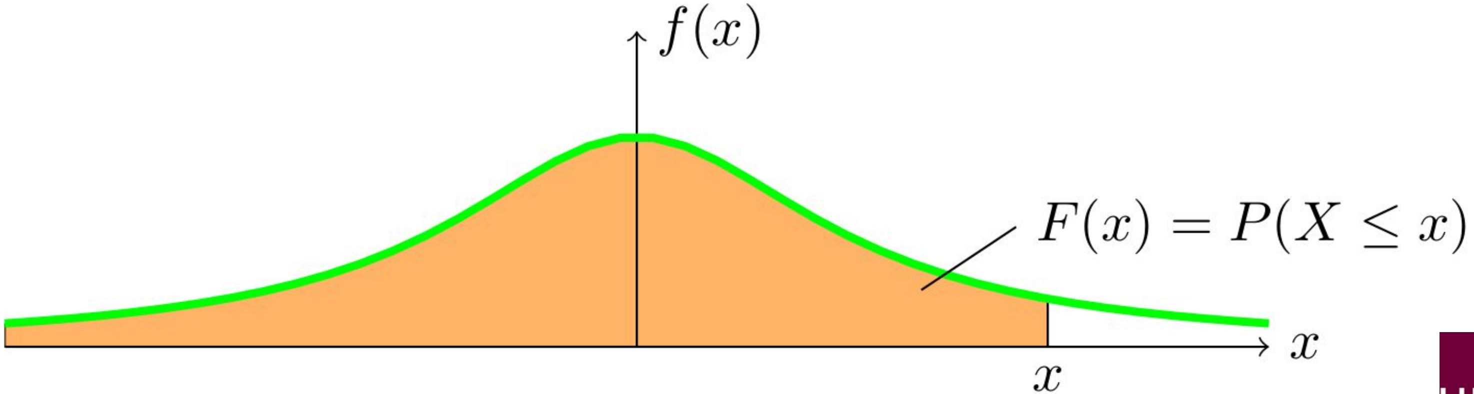
- Fonction de répartition

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$$

Variable aléatoire continue



Fonction de densité et probabilité



Fonction de densité et fonction de répartition

Propriétés de la fonction de répartition

(également valables pour les variables aléatoires discrètes)

- (Définition) $F(x) = P(X \leq x)$
- $0 \leq F(x) \leq 1$
- Jamais décroissante
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to +\infty} F(x) = 1$
- $P(c < X \leq d) = F(d) - F(c)$
- $F'(x) = f(x)$

Espérance

X continue sur $[a, b]$, avec la fonction de densité $f(x)$:

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

X discrète, de valeurs $x_1, \ldots, x_n$, avec la fonction de masse $p(x_i)$:

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

Variance et écart-type

Pour toute variable aléatoire X de moyenne μ

$$\mathrm{Var}(X) = E((X-\mu)^2), \qquad \sigma = \sqrt{\mathrm{Var}(X)}$$

X continue sur $[a, b]$, avec la fonction de densité $f(x)$:

$$\mathrm{Var}(X) = \int_a^b (x-\mu)^2 f(x)\, dx.$$

X discrète, de valeurs $x_1, \ldots, x_n$, avec la fonction de masse $p(x_i)$:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} (x_i-\mu)^2 p(x_i).$$

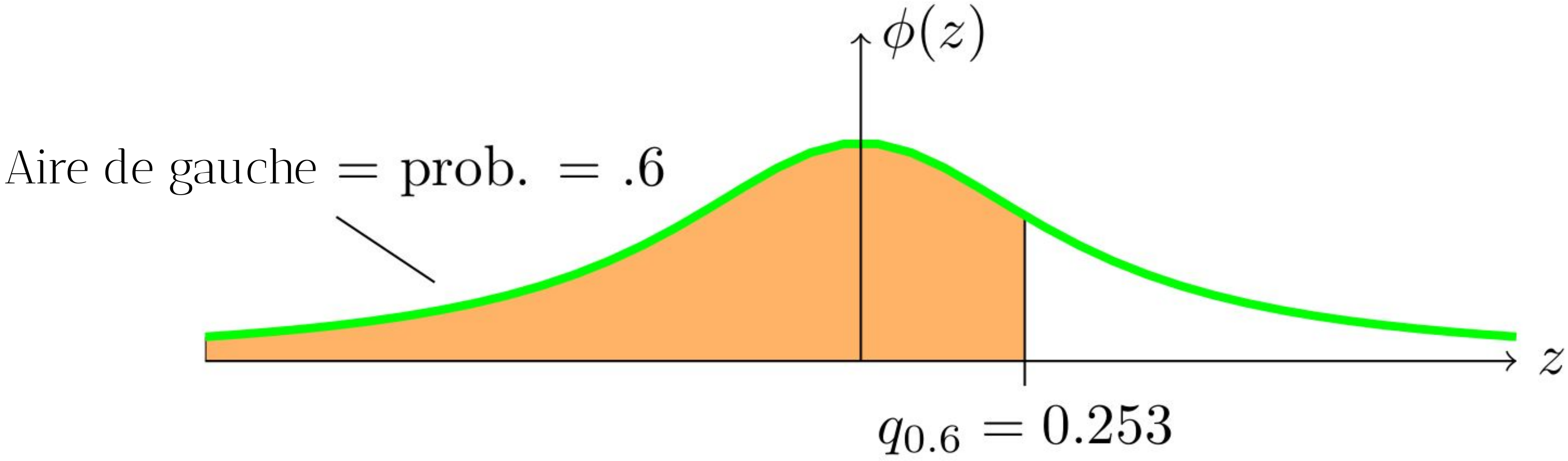
Quantiles (1)

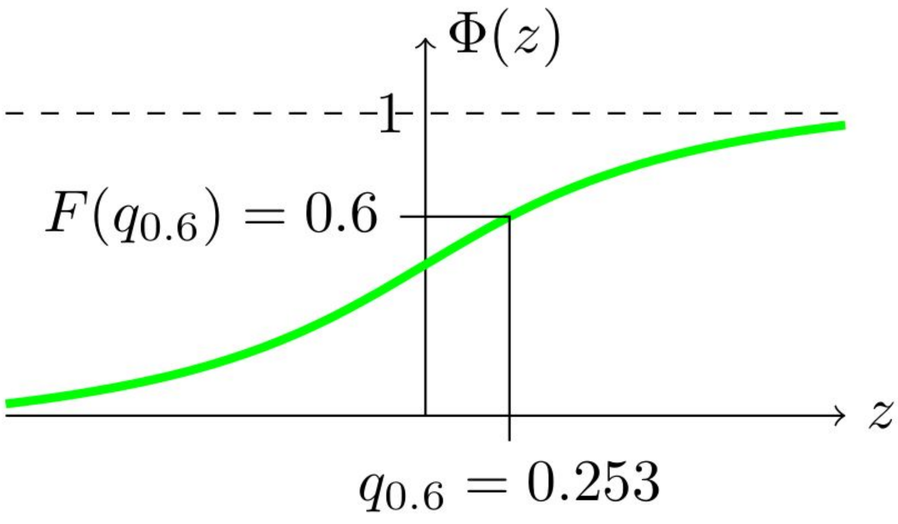
Les quantiles sont les valeurs qui divisent un jeu de données en intervalles de probabilités égales.

Il y a donc un quantile de moins que le nombre de groupes créés. Par exemple, les quartiles sont les trois quantiles qui divisent un ensemble de données en quatre groupes de même probabilité.

La **médiane**, quant à elle, est le quantile qui sépare le jeu de données en deux groupes de même probabilité.

Quantiles (2)

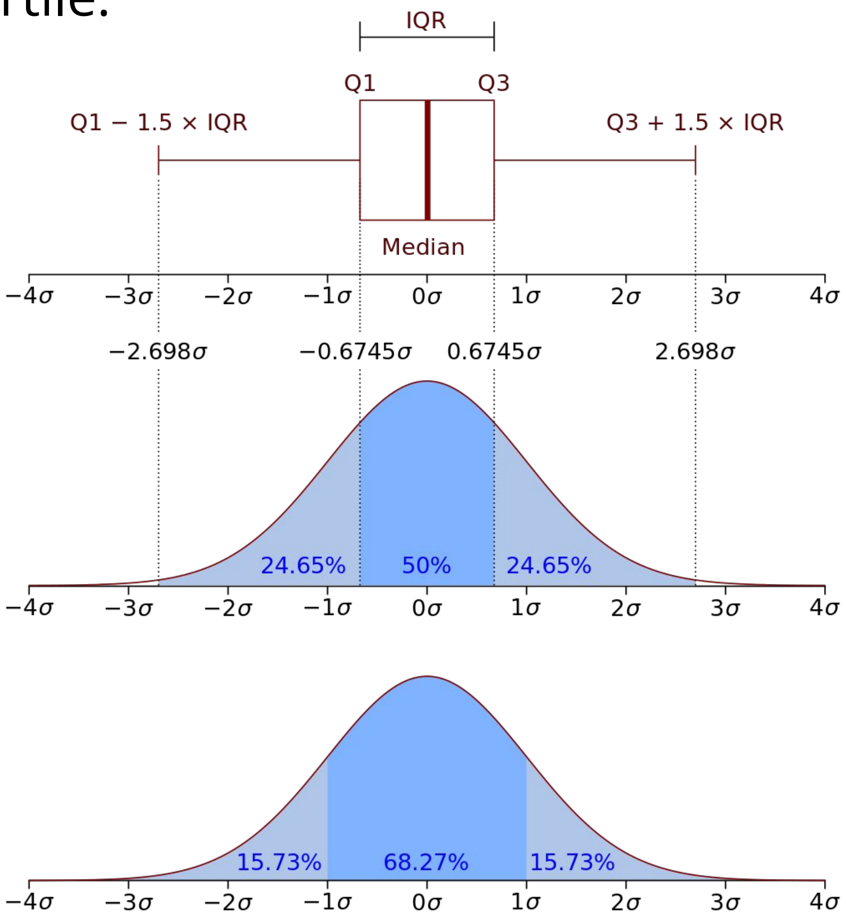




Quantiles (3) : écart interquartile

Indicateur de dispersion qui s'obtient en faisant la différence entre le troisième et le premier quartile.

Représentation en « boîte à moustaches » (*box plot*)



Distributions continues (1)

Loi exponentielle

Loi de probabilité continue qui modélise le temps ***x*** s'écoulant entre deux occurrences d'un **processus de Poisson** de paramètre **λ**, nombre moyen d'occurrences dans un intervalle de temps fixé.

Fonction de densité :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Fonction de répartition :

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases}$$

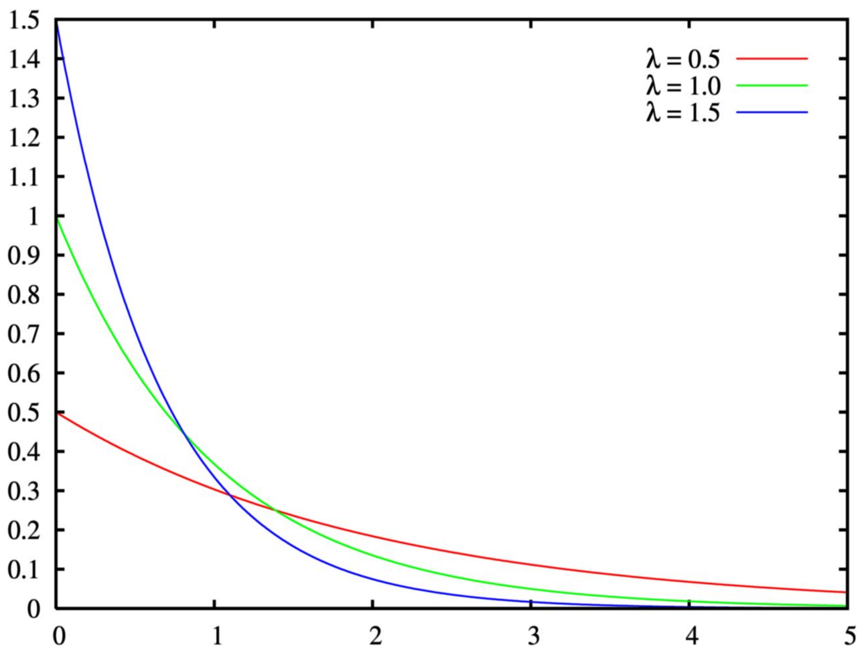
Espérance : $1/\lambda$

Variance : $1/\lambda^2$

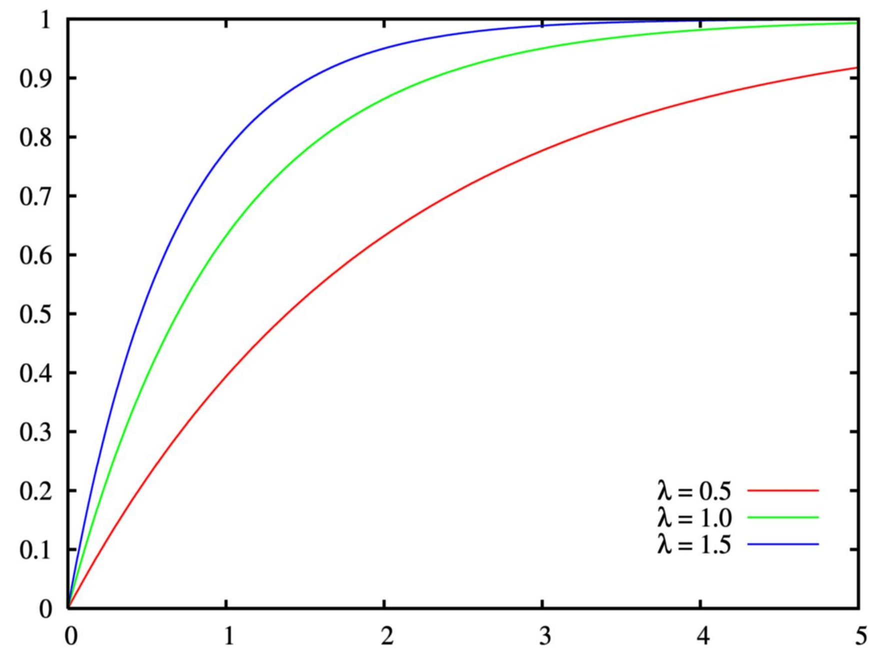
Distributions continues (1)

Loi exponentielle

$$X \sim \exp(\lambda)$$



Fonction de densité



Fonction de répartition

Distributions continues (2)

Lois uniformes continues

Les lois uniformes continues forment une famille de lois de probabilité à densité caractérisées par la propriété suivante : tous les intervalles de même longueur inclus dans le support de la loi ont la même probabilité.

Fonction de densité :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pour } a \le x \le b, \\ 0 & \text{sinon.} \end{cases}$$

Fonction de répartition :

$$F(x) = \begin{cases} 0 & \text{pour } x < a \\ \dfrac{x-a}{b-a} & \text{pour } a \le x < b \\ 1 & \text{pour } x \ge b \end{cases}$$

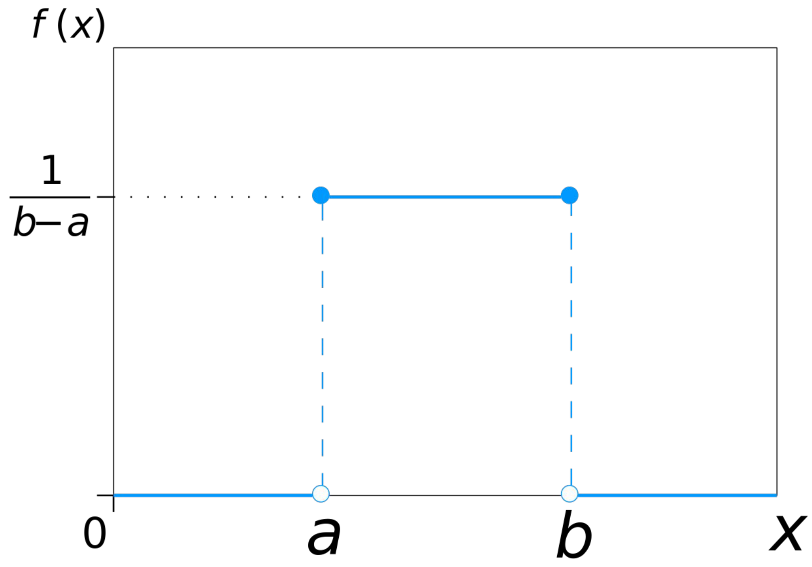
Espérance : (a+b)/2

Variance : $(b-a)^2/12$

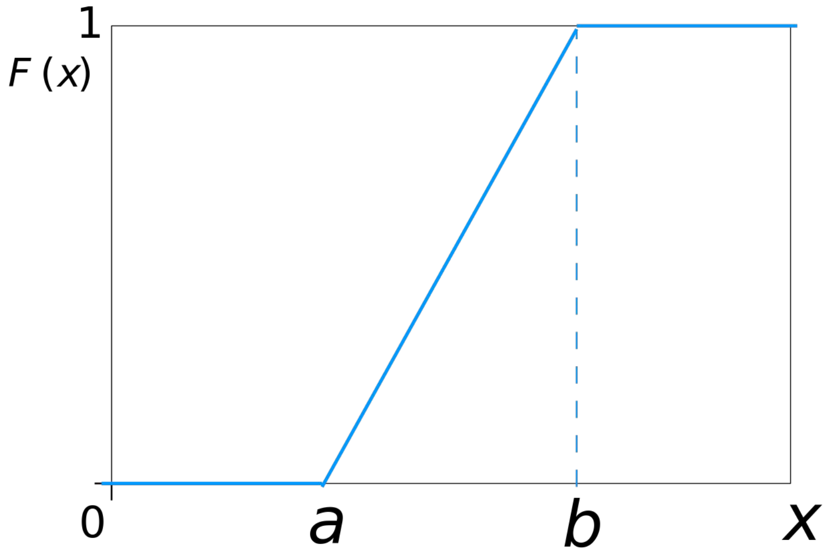
Distributions continues (2)

Lois uniformes continues

$$X \sim U(a, b)$$



Fonction de densité



Fonction de répartition