Statistiques et probabilités
Cours n°2

Guillaume Postic

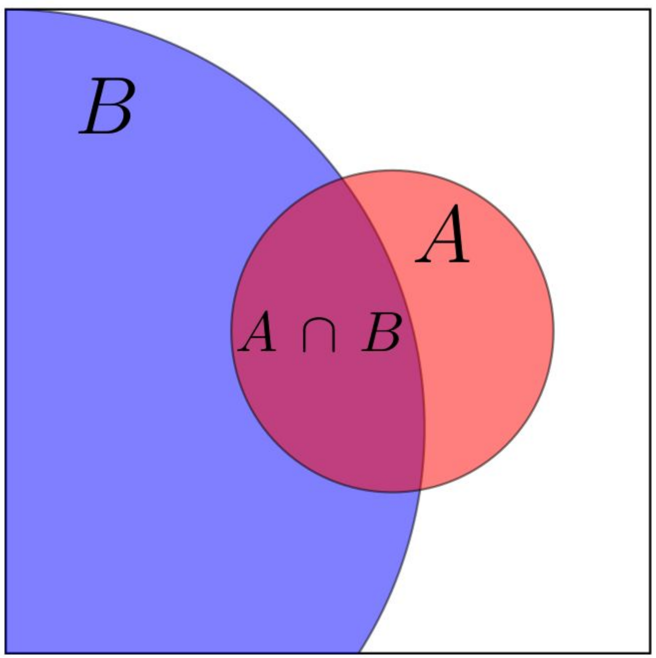
Université Paris-Saclay, Univ. Evry
Département informatique

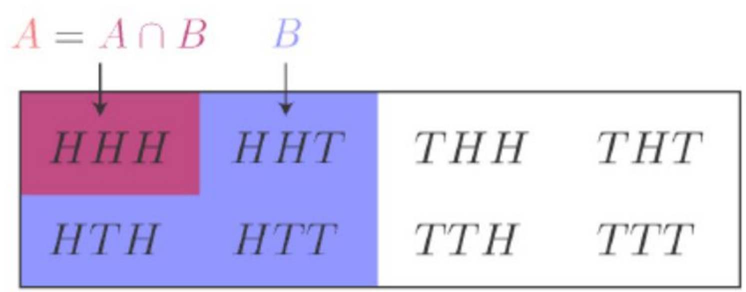
Master 1 MIAGE - 2022/2023

Probabilité conditionnelle

« La probabilité de *A* sachant *B* »

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{avec} \quad P(B) \neq 0$$





▲ Exemple : pile (*T*) ou face (*H*)

◂ Représentation abstraite

Deux principes

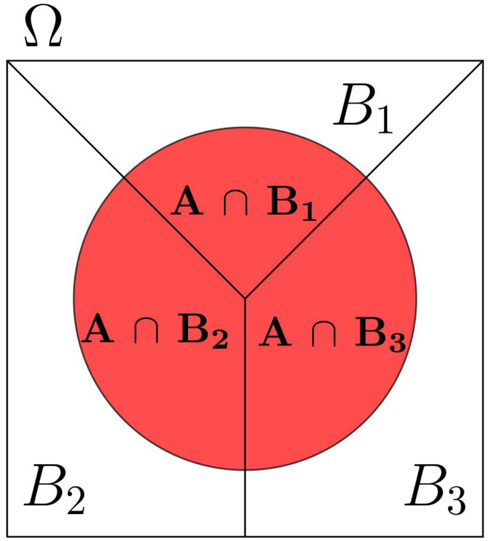
Principe de multiplication

$P(A \cap B) = P(A|B) \cdot P(B)$

Formule des probabilités totales

Si B_1, B_2 et B_3 forment un système exhaustif (ou partition) de Ω (incompatibles deux-à-deux et réunion est l'univers tout entier), alors

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$$
$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)$$

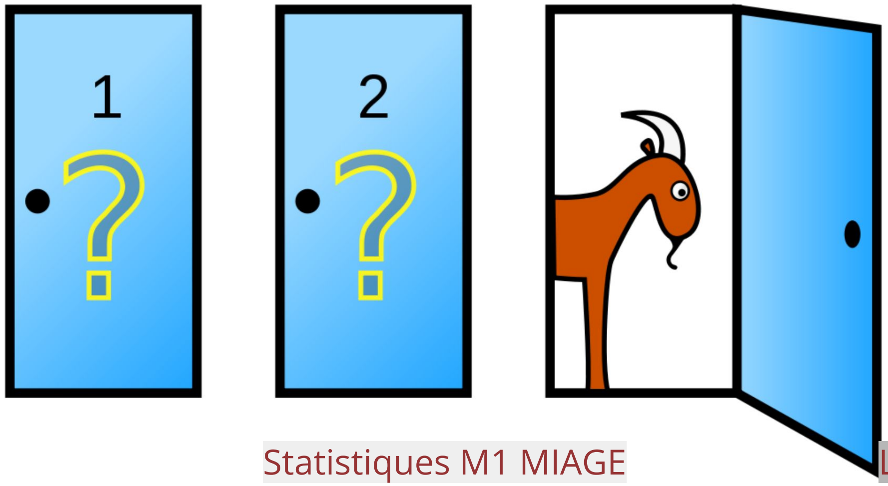


Let's Make a Deal, avec Monty Hall (1)

- Trois portes l'une cache une voiture, les deux autres une chèvre
- Le candidat choisit une porte
- Monty Hall ouvre une porte non-choisie et cachant une chèvre (il sait où est la voiture)
- Le candidat est alors autorisé à changer de porte

Quelle est la meilleure stratégie pour gagner ?

(a) Changer **(b) Ne pas changer** **(c) Peu importe**



Let's Make a Deal, avec Monty Hall (2)

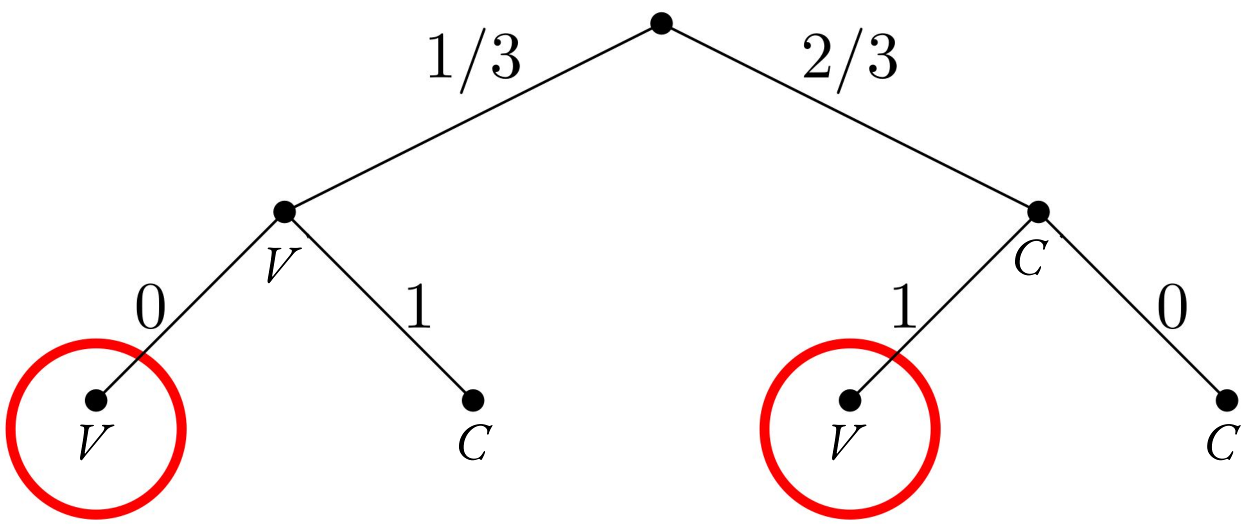
Q : Organisez le problème de Monty Hall sous la forme d'un arbre et calculez la probabilité de gagner si le candidat change de porte choisie quoi qu'il arrive.

$$P(voiture \mid change) = ?$$

Let's Make a Deal, avec Monty Hall (2)

Q : Organisez le problème de Monty Hall sous la forme d'un arbre et calculez la probabilité de gagner si le candidat change de porte choisie quoi qu'il arrive.

$$P(voiture \mid change) = ?$$



La probabilité totale de V est $P(V \mid C) = ⅓ \cdot 0 + ⅔ \cdot 1 = ⅔$

Indépendance

Des évènements A et B sont indépendants si la probabilité que l'un survienne n'est pas affectée par la probabilité que l'autre soit survenu.

Indépendance $\Leftrightarrow P(A \mid B) = P(A)$ (avec $P(B) \neq 0$)

$\Leftrightarrow P(B \mid A) = P(B)$ (avec $P(A) \neq 0$)

(pour n'importe quel A et B)

$\Leftrightarrow P(A \cap B) = P(A)\, P(B)$

Indépendance

Pour deux jets de dés, considérons les évènements suivants :

- A = « le premier jet fait 3 »
- B = « la somme fait 6 »
- C = « la somme fait 7 »

Question : A est indépendant de

(a) B et C (b) B seulement

(c) C seulement (d) Ni B, ni C

Indépendance

Solution (dé n°1 et dé n°2)

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(A) = 1/6$, $\boldsymbol{P(A|B)} = 1/5$. Pas égales, donc pas indépendantes

$P(A) = 1/6$, $\boldsymbol{P(A|C)} = 1/6$. Égales, donc indépendantes

Autre façon :

$P(A \cap B) = 1/36 \neq P(A)\ P(B) = 1/6 \times 5/36$. Pas indépendantes

$P(A \cap C) = 1/36 = P(A)\ P(C) = 1/6 \times 6/36$. Indépendantes

Théorème de Bayes (1)

- Trouver $P(A|B)$ à partir de $P(B|A)$
- $P(B)$ souvent calculée avec la formule des probabilités totales

Théorème de Bayes (2)

Mise à jour Bayésienne de la probabilité d'une hypothèse *H* grâce à une donnée (observation) *d* :

Vraisemblance

Probabilité a posteriori

$$p(H \mid d) = \frac{p(d \mid H)}{p(d)} \, p(H)$$

Probabilité a priori

Vraisemblance marginale

$$p(d) = \Sigma_i \, p(d \mid H_i) \, p(H_i) = \Sigma_i \, p(d \cap H_i)$$

Probabilités totales :
somme pour toutes les hypothèses

$$p(d_1, d_2, \ldots, d_n \mid H) = \Pi_i \, p(d_i \mid H)$$

Si indépendance mutuelle des évènements (observations)

Notation

- $p(A,B) = p(A \cap B)$
- $p(A | B,C) = p(A | B \cap C) = p(A \cap B \cap C) / p(B \cap C)$
- $p(A,B | C) = p(A \cap B | C) = p(A \cap B \cap C) / p(C)$

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

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

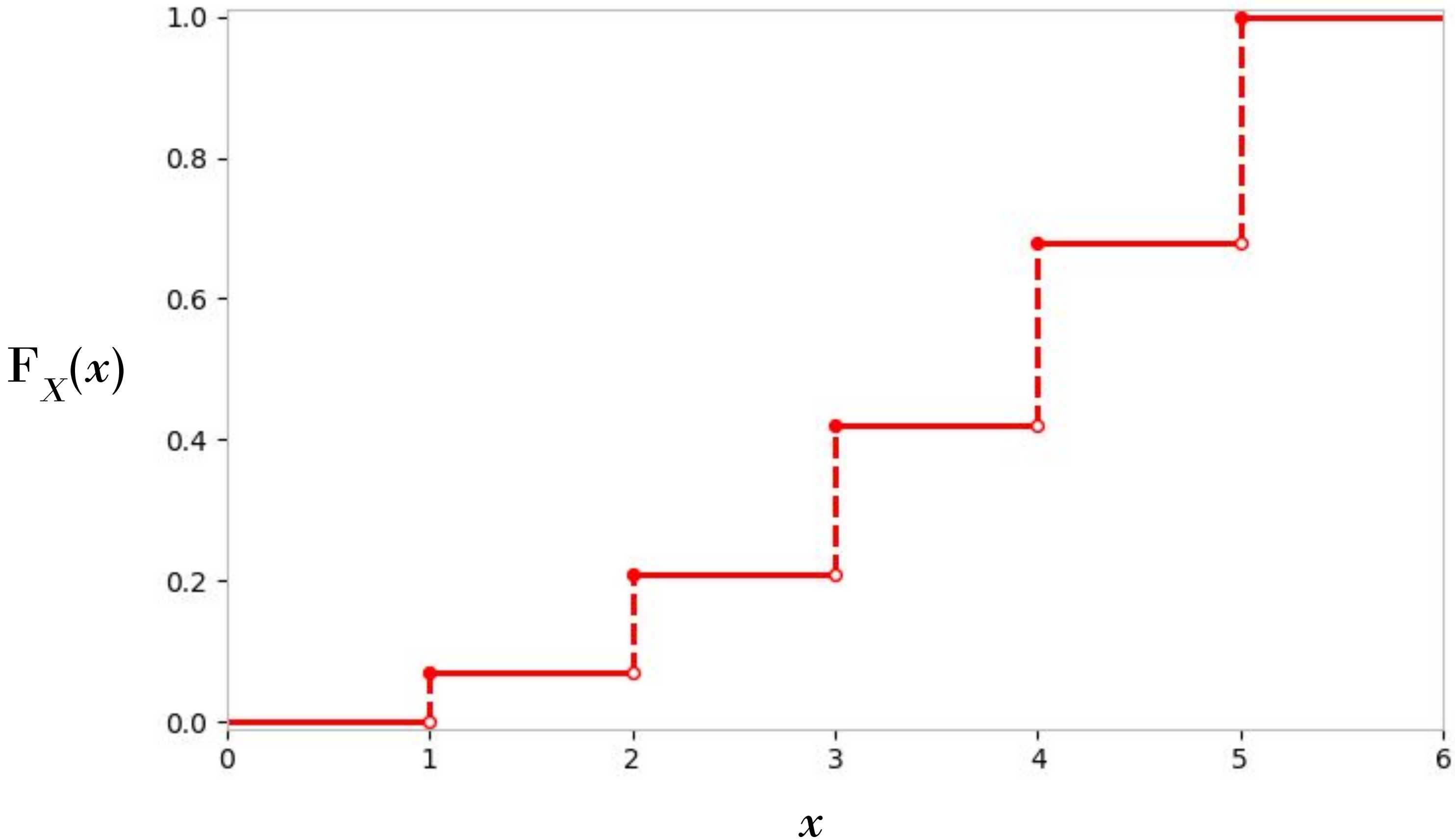
  Note : elle défini la loi de probabilité discrète suivie par X

- 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 répartition du résultat d'un jet de dé à six faces

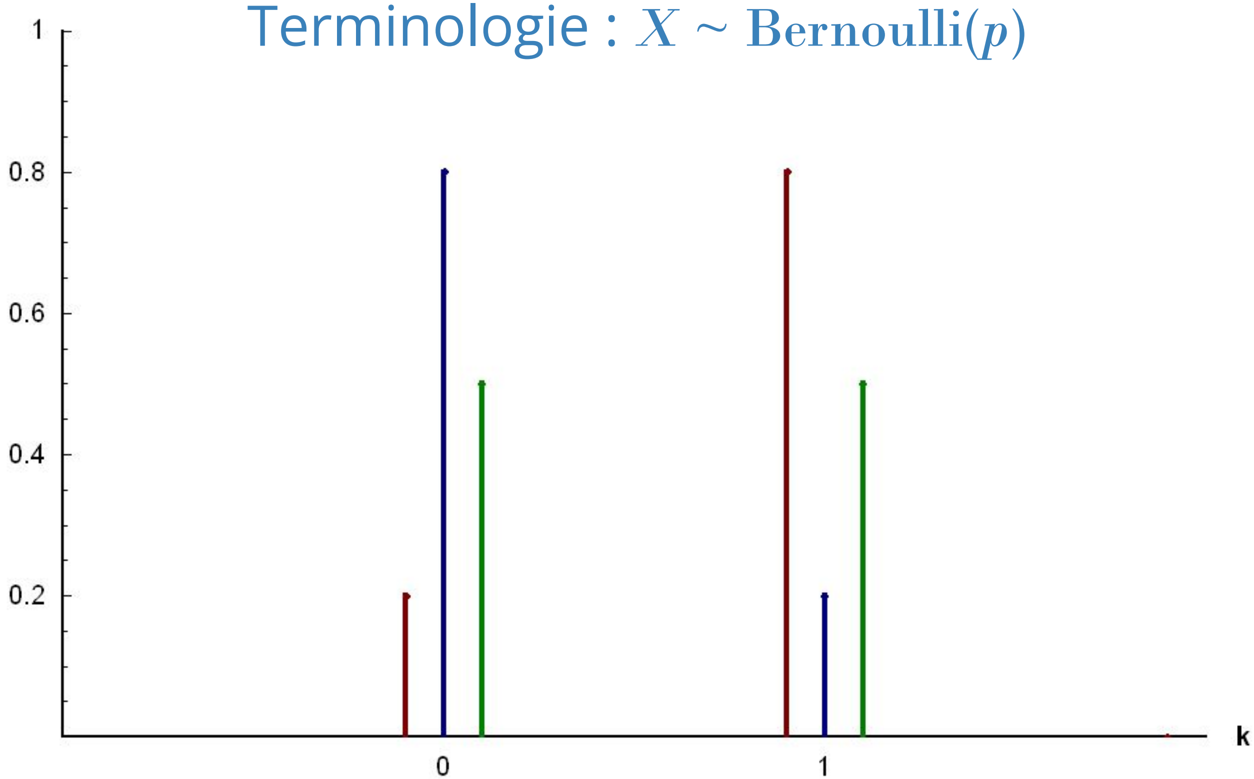


Distributions discrètes (1)

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 $X_1, X_2, \ldots, X_n$ *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 :

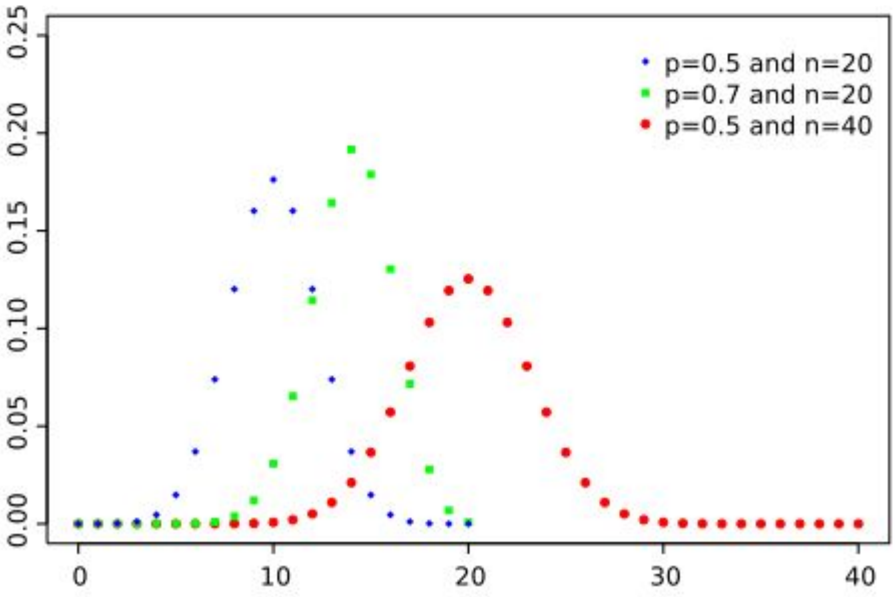
$$N = \sum_{k=1}^{n} X_k \sim \mathcal{B}(n, p)$$

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

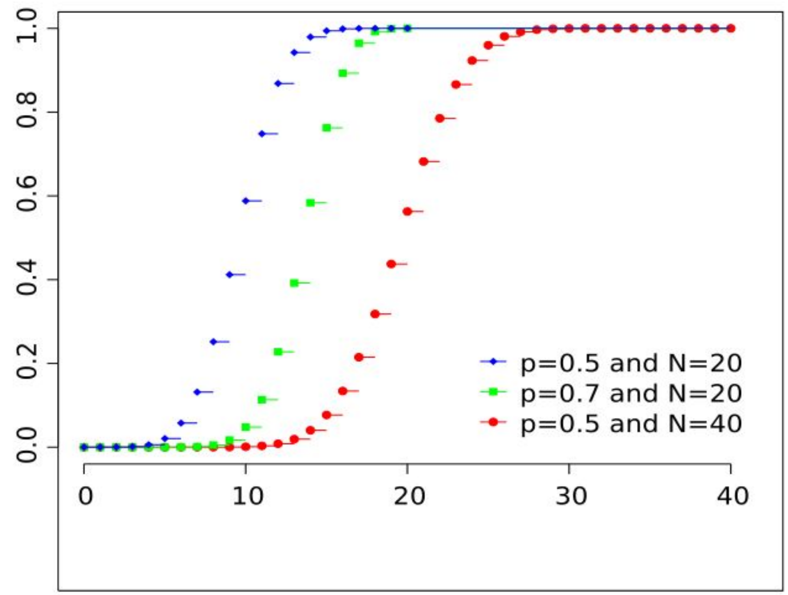
$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Distributions discrètes (2)

Loi binomiale



Fonction de masse



Fonction de répartition

Distributions discrètes (3)

Loi géométrique

Selon la convention choisie :

- la loi du nombre *X* d'épreuves de Bernoulli indépendantes de probabilité de succès *p* ∈]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, ...}.

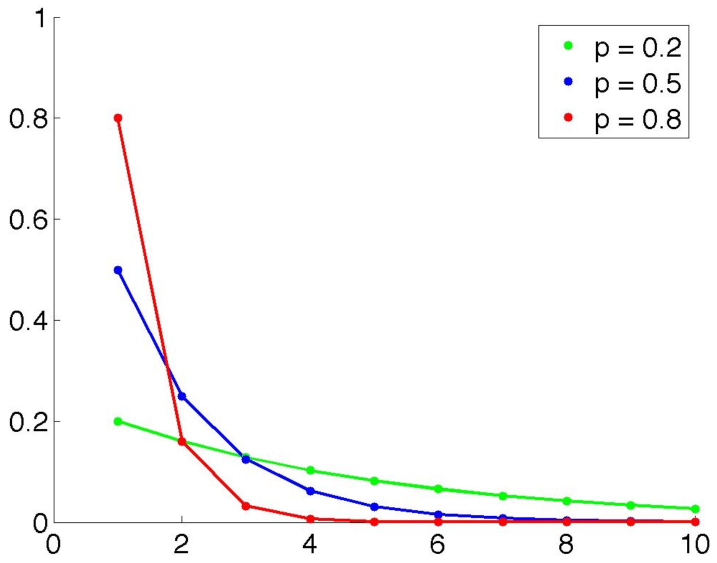
$$\mathbb{P}(X = k) = q^{k-1}p.$$

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

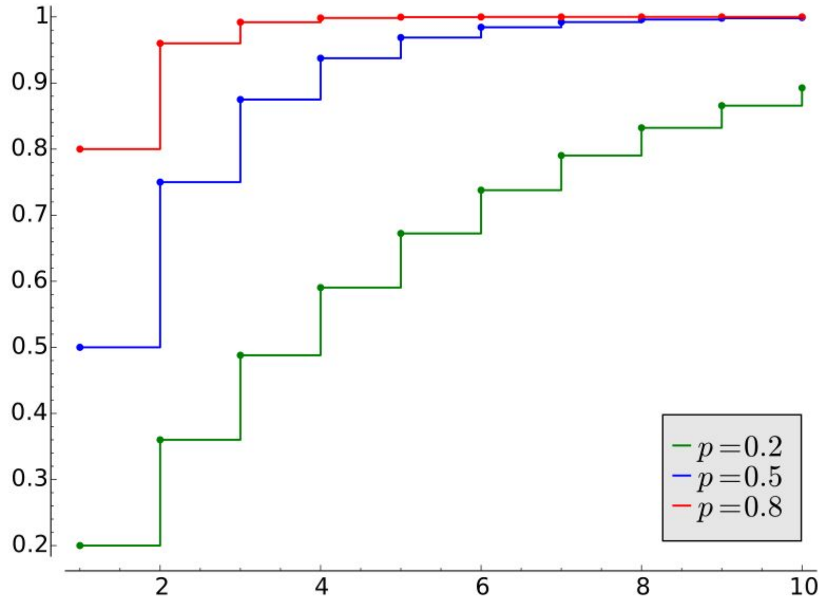
$$\mathbb{P}(Y = k) = q^{k}p.$$

Distributions discrètes (3)

Loi géométrique



Fonction de masse

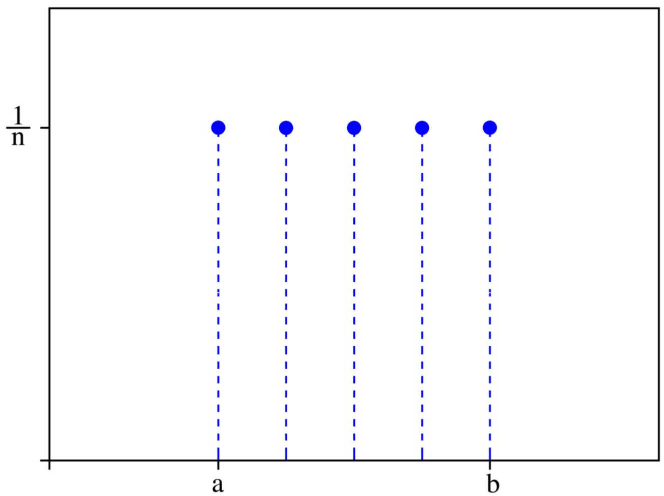


Fonction de répartition

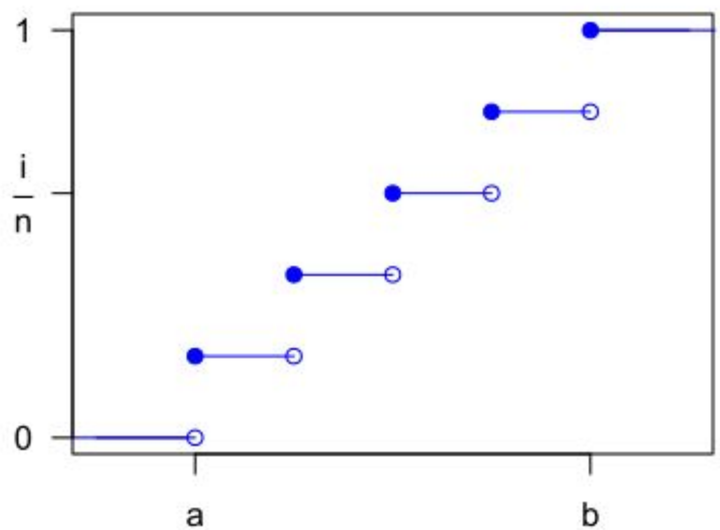
Distributions discrètes (4)

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 (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

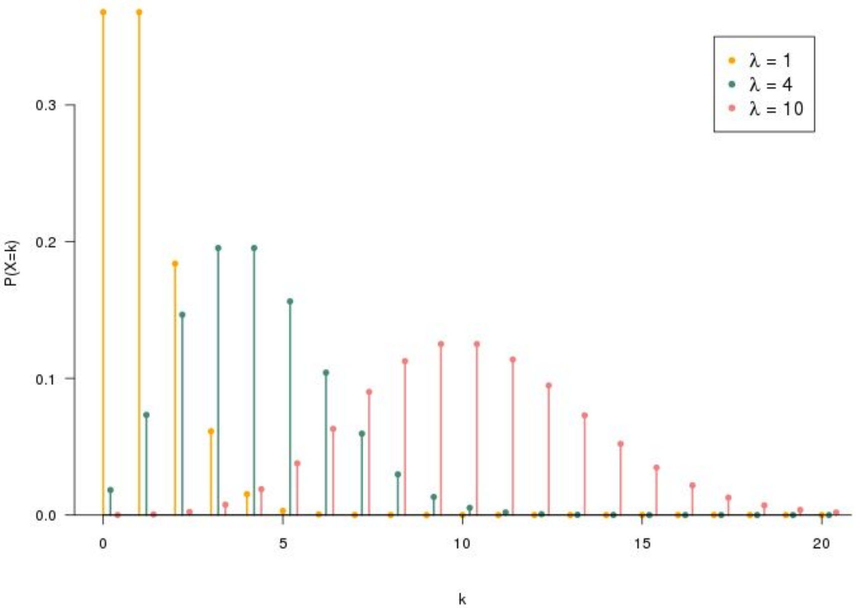
- **λ** est la **fréquence** à laquelle l'évènement survient
- le nombre N_t d'occurrences dans un intervalle de longueur *t* suit une loi de Poisson « d'intensité » λt :

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

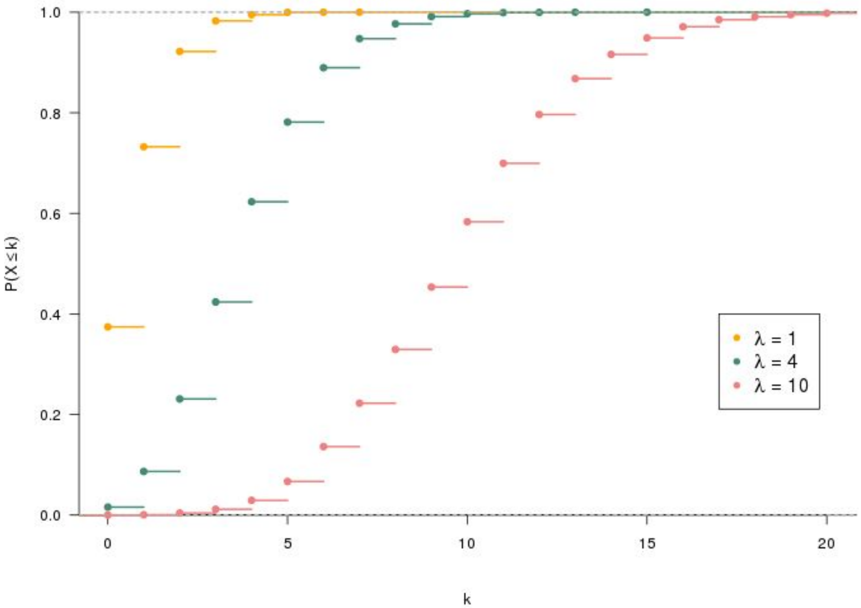
Distributions discrètes (5)

Loi de Poisson

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



Fonction de masse



Fonction de répartition

Espérance (mathématique)

Pour X une variable aléatoire prenant les valeurs x_1, x_2, …, x_n, l'espérance de X est définie par :

$$E(X) = p(x_1)x_1 + p(x_2)x_2 + \ldots + p(x_n)x_n = \sum_{i=1}^{n} p(x_i)\, x_i$$

- Calcul d'une moyenne arithmétique pondérée
- Indicateur de tendance centrale

Propriétés

$$E(X + Y) = E(X) + E(Y)$$
$$E(aX + b) = aE(X) + b$$
$$E(h(X)) = \sum_{i} h(x_i)\, p(x_i)$$

Variance (1)

Indicateur de dispersion des valeurs d'un échantillon ou d'une distribution de probabilité. Elle exprime la moyenne des carrés des écarts à la moyenne (écart quadratique moyen) :

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Sa racine carrée définit l'**écart type σ** : $\sigma^2 = V(X)$

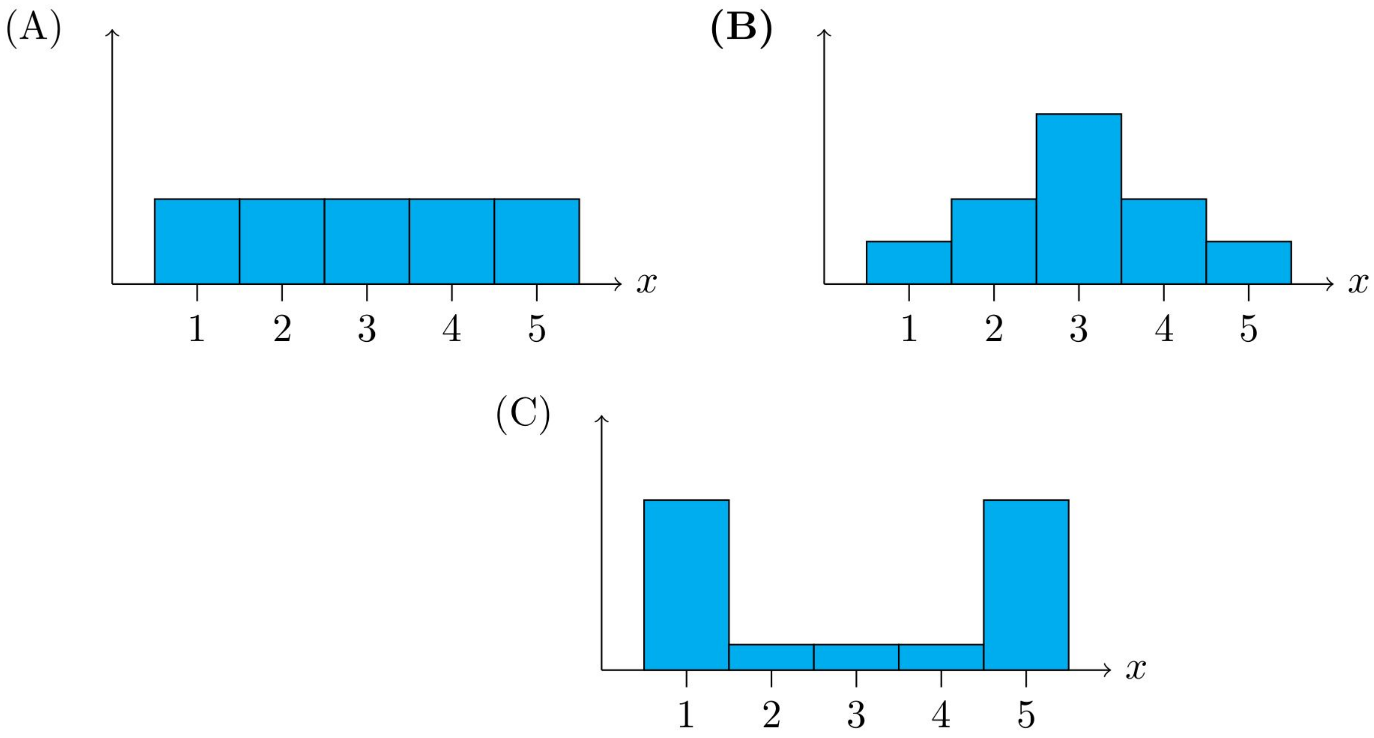
Elle est aussi égale à la différence entre la moyenne des carrés des valeurs de la variable et le carré de la moyenne : $V(X) = E(X^2) - E(X)^2$

Propriétés :

- $V(a+bX) = b^2\ V(X)$
- $V(X+Y) = V(X-Y) = V(X) + V(Y)$, si X et Y indépendants

Variance (2)

La figure ci-dessous donne les fonctions de masse de 3 variables aléatoires. Classez-les par valeur d'écart-type, du plus grand au plus petit (on supposera que *x* a toujours la même unité).



Variance (3)

Remarque : si la variance d'une variable X est nulle ($V(X) = 0$), alors X est une constante.