Probabilités et Statistiques pour l'Informatique

Akka ZEMMARI
LaBRI - Université de Bordeaux

zemmari@u-bordeaux.fr

Table des Matières

- **Introduction, définitions, ...**
- **Probabilités discrètes**
- **Probabilités continues**
- **Inférences statistiques**
- **Applications en Informatique**
- **Génération aléatoire**

Introduction

L'utilisation des probabilités et statistiques remonte à très loin ...

- premiers concepts apparus dans l'antiquité : jeux de hasard, etc
- les activités humaines sont affectées par le hasard :
  - les accidents,
  - le temps qu'il fait,
  - tirage au sort dans le sport,
  - etc

Introduction

Et de nos jours : ....

- Un vaccin est-il plus efficace qu'un autre ?
- Évolution de l'épidémie actuelle, quel modèles utiliser ? Quel est l'impact des mesures prises ?
- Que pensez-vous de l'affirmation suivante : "il y a plus de malades de COVID chez les vaccinés que chez les non vaccinés."
- ...

Introduction

Et en MIAGE :

- L3 : Probabilités et Statistiques
- Master 1 :
  — S7 : Statistiques et Marketing
  — S8 : Processus stochastiques et Simulation
- Master 2 :
  — S9 : IA & Données.

Introduction - Définitions

- Expérience aléatoire : expérience où le hasard intervient rendant le résultat imprévisible.
- Événement : assertion relative au résultat d'une expérience aléatoire.
- La probabilité est une évaluation du caractère probable d'un événement.
- Le mot probable signifie : qui peut se produire dans l'expérience aléatoire.

$\Rightarrow$ Il faut donner une définition assez rigoureuse de ces concepts ...

Probabilités Discrètes

Exemple. On lance deux dés.

$$\Omega = D \times D = \left\{ \boxed{\bullet}\ \boxed{\bullet}\ , \boxed{\bullet}\ \boxed{\bullet\,\bullet}\ , \ldots, \boxed{\bullet\,\bullet}\ \boxed{\bullet\,\bullet} \right\}$$

avec

$$D = \left\{ \boxed{\bullet}\ , \boxed{\bullet\,\bullet}\ , \boxed{\bullet\,\bullet}\ , \boxed{\bullet\,\bullet}\ , \boxed{\bullet\,\bullet}\ , \boxed{\bullet\,\bullet} \right\}$$

L'espace Ω contient $6 \times 6 = 36$ éléments.

Distribution uniforme :

tous les éléments de Ω sont de même probabilité $\frac{1}{36}$

$\Longleftrightarrow$

tous les éléments de D sont de même probabilité $\frac{1}{6}$.

****

Définitions. Un *espace probabilisé discret* est un couple $(\Omega, \mathbb{P}r)$, où Ω est un ensemble non vide au plus dénombrable* et $\mathbb{P}r$ une application de Ω dans $[0, 1]$ telle que :

$$\sum_{\omega \in \Omega} \mathbb{P}r(\omega) = 1.$$

$*$. Voir le définition au tableau.

- Ω : *espace des événements élémentaires*
- $A \subset \Omega$: *événement*
- $\mathbb{P}r$: une *loi* (ou *distribution*) *de probabilité sur* Ω
- On prolonge $\mathbb{P}r$ sur $\mathcal{P}(\Omega)$ par :

$$\mathbb{P}r(A) = \sum_{\omega \in A} \mathbb{P}r(\omega), \quad \forall A \subset \Omega$$

$\mathbb{P}r(A)$: *probabilité de* A

Proposition

- $\mathbb{P}r(\emptyset) = 0$
- $\mathbb{P}r(\bar{A}) = 1 - \mathbb{P}r(A)$
- $\mathbb{P}r(\cup_{i \in I} A_i) = \sum_{i \in I} \mathbb{P}r(A_i)$, pour toute famille au plus dénombrable $A_i, i \in I$ d'éléments de $\mathcal{P}(\Omega)$ 2-à-2 disjoints.

Exemple. On peut adopter une hypothèse non uniforme en posant :

$$\mathbb{P}r_2(\ \boxed{\cdot}\ ) = \mathbb{P}r_2(\ \boxed{\vdots}\ ) = \frac{1}{4}$$

$$\mathbb{P}r_2(\ \boxed{\cdot^{\cdot}}\ ) = \mathbb{P}r_2(\ \boxed{\cdot^{\cdot}}\ ) = \mathbb{P}r_2(\ \boxed{::}\ ) = \mathbb{P}r_2(\ \boxed{::}\ ) = \frac{1}{8}$$

Sur $\Omega = D \times D$, nous pouvons définir la distribution produite :

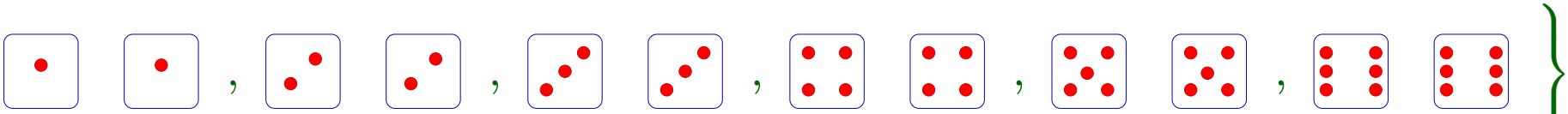
$$\mathbb{P}r_{22}(dd') = \mathbb{P}r_2(d)\mathbb{P}r_2(d')$$

Dans le cas où les deux dés sont pipés, nous avons

$$\mathbb{P}r_{22}(\ \boxed{\vdots}\ \ \boxed{\cdot^{\cdot}}\ ) = \frac{1}{4}.\frac{1}{8} = \frac{1}{32}.$$

On démontrera facilement que $\mathbb{P}r_{22}$ est bien une distribution de probabilité sur Ω.

A titre d'exemple, on peut définir l'événement "double" par :

$A =$ 

Or nous avons par définition :

$$\mathbb{P}r(A) = \sum_{\omega \in A} \mathbb{P}r(\omega)$$

Pour la distribution uniforme, cette probabilité vaut

$$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{6}.$$

Alors que, pour $\mathbb{P}r_{22}$, elle vaut

$$\frac{1}{16} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{64} + \frac{1}{16} = \frac{3}{16}.$$

Définition. Soit $(\Omega, \mathbb{P}r)$ un espace probabilisé discret et soit Ω' un ensemble non vide au plus dénombrable. Une *variable aléatoire* (v.a.) X à valeurs dans Ω' est une application de Ω dans Ω'. Nous prenons souvent pour Ω' un sous-ensemble de $\mathbb{N}$ ou de $\mathbb{R}$.
On pourra munir Ω' d'une loi de probabilité $\mathbb{P}r_X$ en posant, pour tout $\omega' \in \Omega'$:

$$\mathbb{P}r_X(\omega') = \mathbb{P}r(X^{-1}(\{\omega'\}))$$

Proposition. $\mathbb{P}r_X$ est une loi de probabilité sur Ω', i.e.

$$\sum_{\omega' \in \Omega'} \mathbb{P}r_X(\omega') = 1.$$

Nous avons de plus, pour tout $A' \subset \Omega'$:

$$\mathbb{P}r_X(A') = \mathbb{P}r(X^{-1}(A'))$$

Lorsqu'il n'y a pas le danger de confusion, $\mathbb{P}r_X$ sera désignée par $\mathbb{P}r$.

Exemple. Soit $S(\omega)$ la v.a. qui est la somme des points des deux dés, par exemple $S($ $) = 3 + 6 = 9$. Le tableau ci-dessous caractérise la loi de probabilité pour la v.a. S par rapport à chacune des distributions (uniforme et biaisée) :

s	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}r_{11}(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\mathbb{P}r_{22}(S=s)$	$\frac{4}{64}$	$\frac{4}{64}$	$\frac{5}{64}$	$\frac{6}{64}$	$\frac{7}{64}$	$\frac{12}{64}$	$\frac{7}{64}$	$\frac{6}{64}$	$\frac{5}{64}$	$\frac{4}{64}$	$\frac{4}{64}$

Espérance Mathématique et Variance

Une variable aléatoire admet un certain nombre de valeurs typiques. Nous considérons dans la suite les v.a. à valeurs dans $\mathbb{R}$.

Définition. En arithmétique usuelle la valeur moyenne de n nombres est définie comme leur somme divisée par n. En calcul des probabilités, l'*espérance* d'une v.a. est définie comme la somme des valeurs prises pondérées par les probabilités respectives, c'est-à-dire :

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} x.\mathbb{P}r(X = x)$$

lorsque cette somme converge absolument* . ($X(\Omega)$ est l'ensemble des valeurs prises par la v.a. X). Sinon, on dit que X *n'admet pas d'espérance*.

$*$. Voir le définition au tableau.

Exemple. Revenons au cas de la v.a. S, désignant la somme des deux dés, et calculons son espérance pour chacune des distributions $\mathbb{P}r_{11}$ et $\mathbb{P}r_{22}$:

- $\mathbb{E}_{11}(S) = 2.\frac{1}{36} + 3.\frac{2}{36} + ... + 12.\frac{1}{36} = 7.$
- $\mathbb{E}_{22}(S) = 2.\frac{4}{64} + 3.\frac{4}{64} + 4.\frac{5}{64} + ... + 12.\frac{4}{64} = 7.$

Etant donné que l'espérance de la v.a. X, désignant le point obtenu par un seul dé, vaut :

- $\mathbb{E}_1(X) = \frac{1}{6}.(1 + 2 + ... + 6) = 3.5$ (pour le dé uniforme)
- $\mathbb{E}_2(X) = \frac{1}{4}.(1 + 6) + \frac{1}{8}.(2 + 3 + 4 + 5) = 3.5$ (pour le dé pipé)

on peut se demander s'il y a un lien entre l'espérance d'une somme de v.a. et la somme des espérances.

Linéarité de l'espérance

Proposition. Soient X_1 et X_2 deux v.a. définies sur le même espace probabilisé discret $(\Omega, \mathbb{P}r)$ et admettant toutes deux une espérance. Soit $\alpha \in \mathbb{R}$. Alors :

$$\begin{aligned} \mathbb{E}(\alpha X_1) &= \alpha \mathbb{E}(X_1) \\ \mathbb{E}(X_1 + X_2) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) \end{aligned}$$

Que peut-on dire de l'espérance d'un produit de v.a. ?

Définitions. On peut considérer plusieurs v.a. sur le même espace probabilisé discret. Soient les v.a. X et Y définies sur $(\Omega, \mathbb{P}r)$. La loi (ou distribution) *conjointe* est la donnée de :

$$\mathbb{P}r_{XY}(X = x, Y = y) = \mathbb{P}r(\{\omega \in \Omega | X(\omega) = x, Y(\omega) = y\})$$

pour tout x et tout y possibles.
X et Y sont dites *indépendantes*, si pour tout x et tout y possibles, on a $\mathbb{P}r_{XY}(X = x, Y = y) = \mathbb{P}r_X(X = x)\mathbb{P}r_Y(Y = y)$.

Proposition. Si X et Y sont deux v.a. indépendantes admettant une espérance, alors la v.a. produite XY admet une espérance et

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

.

Exemple. Soient X et Y les points obtenus en lançant deux dés (nous considérons les cas des dés uniformes et celui des dés pipés). Ces deux v.a. sont indépendantes (par la définition de $\mathbb{P}r_{11}$ ou de $\mathbb{P}r_{22}$). Soient $S = X + Y$ et $P = XY$. Nous avons :

- $\mathbb{E}(S) = \frac{7}{2} + \frac{7}{2} = 7, \mathbb{E}(P) = \frac{7}{2}.\frac{7}{2} = \frac{49}{4}$.
  (un calcul direct fondé sur les probabilités des différentes valeurs de P aboutit au même résultat dans les deux cas.)

- Mais S et P ne sont pas indépendantes.
  Nous avons toutefois $\mathbb{E}(S + P) = \mathbb{E}(S) + \mathbb{E}(P) = 7 + \frac{49}{4}$.
  Un calcul des probabilités des valeurs de SP permet d'obtenir son espérance :
  $\mathbb{E}(SP) = \frac{637}{6}$ dans le cas des dés uniformes et 112 dans le cas des pipés ; qui sont différentes, alors que dans les deux cas :

$$\mathbb{E}(S)\mathbb{E}(P) = 7.\frac{49}{4} = \frac{343}{4}.$$

Définitions. Un paramètre important qui vient tout de suite après l'espérance est la variance. Elle mesure la dispersion d'une v.a. Si X est une v.a. définie sur $(\Omega, \mathbb{P}r)$, sa *variance* est définie par :

$$\mathbb{V}ar(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

lorsque celle-ci existe.

Un exemple ***.**

Dans une classe de 25 élèves, à un devoir, on observe les notes suivantes :

5 notes à 8, 5 notes à 9, 5 notes à 10, 5 notes à 11 et 5 notes à 12.

- L'espérance (ou moyenne des notes) est 10.

- La variance vaut $4/5 + 1/5 + 0/5 + 1/5 + 4/5 = 2$. Elle est assez faible, les notes sont donc très centrées autour de la moyenne.

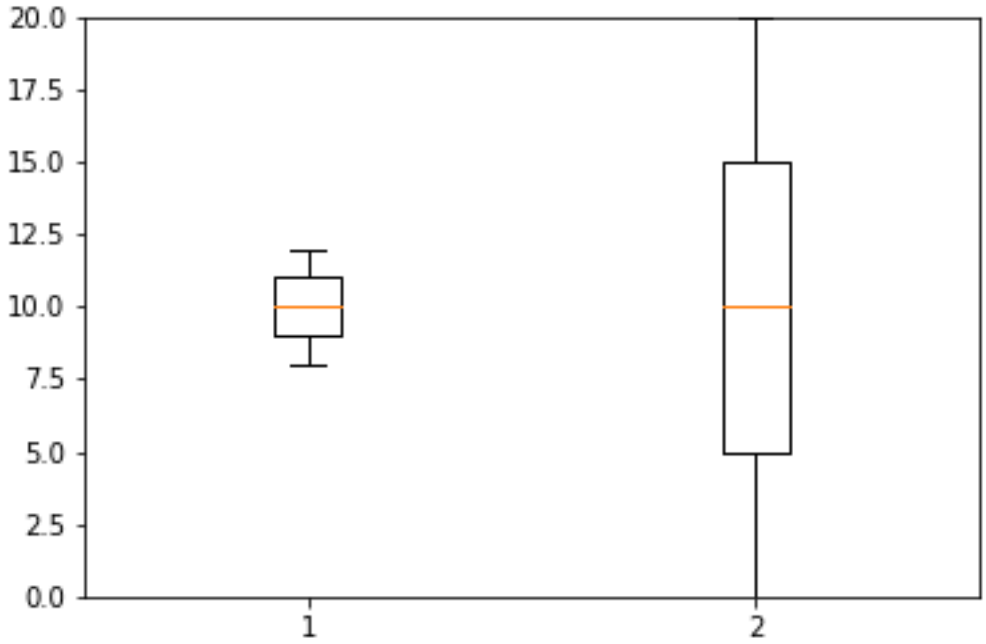
*. https://www.bibmath.net/dico/index.php?action=affiche&quoi=./e/esperance.html

Dans une autre classe de 25 élèves, au même devoir, les élèves ont obtenu : 5 notes à 0, 5 notes à 5, 5 notes à 10, 5 notes à 15 et 5 notes à 20.

- L'espérance vaut toujours 10, mais pourtant il est clair que les deux classes sont très différentes ! Calculons la variance : elle vaut $100/5 + 25/5 + 0/5 + 25/5 + 100/5 = 50$.

La variance est beaucoup plus grande que dans le premier cas, ce qui signifie que les notes sont très espacées !

Ce que l'on peut observer visuellement :



Définition. La racine carrée de la variance est appelée *écart-type* et est notée par σ :

$$\sigma(X) = \sqrt{\mathbb{V}ar(X)}.$$

Proposition. Nous avons :

$$\mathbb{V}ar(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Proposition. Si X et Y sont deux v.a. indépendantes admettant chacune une variance, alors la v.a. $X + Y$ admet une variance qui est la somme des deux variances.

Quelques Distributions Discrètes

Loi de Bernoulli. La v.a. X prend deux valeurs : 1 avec la probabilité p ou 0 avec la probabilité q ; on suppose que $p, q \in [0, 1]$ et $p + q = 1$. Nous avons :

- $\mathbb{E}X = p$
- $\mathbb{V}X = pq$
- $\sigma X = \sqrt{pq}$

Utilisation. Cette loi intervient souvent de faÃ§on implicite lorsqu'on veut traiter une probabilité comme une espérance. En effet c'est la loi de la v.a. qui est la fonction indicatrice d'un événement A de probabilité p :

$$Pr(A) = \mathbb{E}\mathbf{1}_A.$$

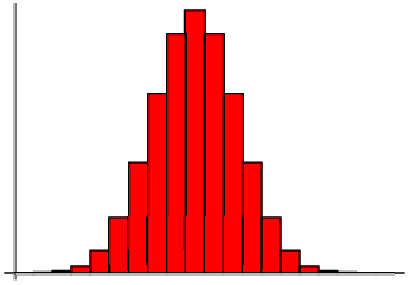
Loi binomiale. On effectue n épreuves identiques et indépendantes ; la probabilitié de succès dans chacune étant supposée égale à p et celle d'échec $q = 1 - p$.
Posons $X =$ le nombre total de succès obtenus dans les épreuves. X est une v.a. qui peut prendre la valeur k dans l'intervalle $[0, n]$ avec la probabilité :

$$p_k = Pr(X = k) = \binom{n}{k} p^k q^{n-k}. \text{ (pourquoi ?)}$$

(**Rappel :** $\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!}$.)

On dit alors que X *suit une loi binomiale de paramètres* n et p.

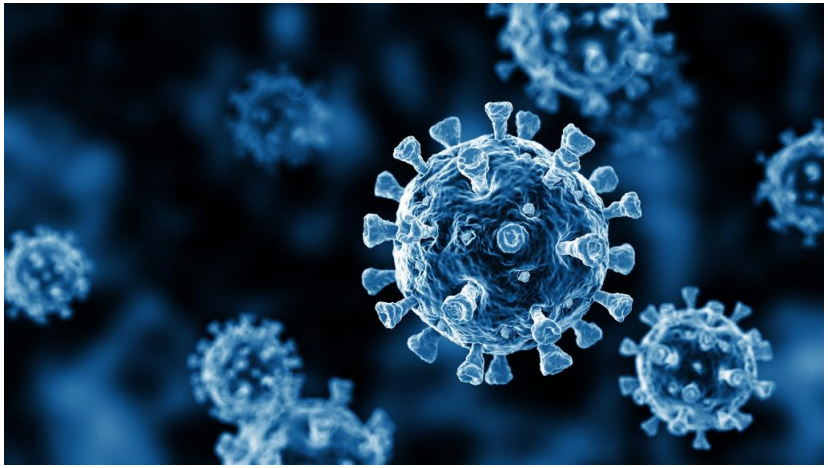
Il est clair qu'on peut considérer la v.a. binomiale X comme la somme de n v.a. de Bernoulli identiques et indépendantes de même paramètre p. Nous avons donc $\mathbb{E}X = np$, $\mathbb{V}X = npq$ et $\sigma X = \sqrt{npq}$.



Distribution binomiale

Loi binomiale : exemple *

COVID-19 et risque de contact



Voir le tableau

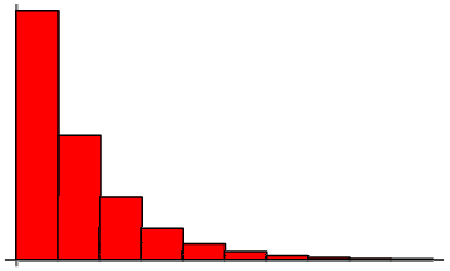
*. https://www.maths-et-tiques.fr/telech/covid-bino.pdf

Loi géométrique. Soit $p > 0$ la probabilité de succès dans une épreuve. Posons $q = 1 - p$. Nous répétons la même épreuve indépendamment jusqu'à l'obtention d'un *premier* succès.
Soit X la v.a. désignant le nombre d'épreuves effectuées.
C'est une v.a. qui peut prendre la valeur $k \in \mathbb{N}^*$ avec la probabilité :

$$p_k = Pr(X = k) = q^{k-1}p. \text{ (pourquoi ?)}$$

On dit alors que X suit une *loi géométrique de paramètre* p. On calcule son espérance et sa variance soit par une méthode directe soit en dérivant sa fonction génératrice :

- $\mathbb{E}X = \frac{1}{p}$
- $\mathbb{V}X = \frac{q}{p^2}$
- $\sigma X = \frac{\sqrt{q}}{p}$



Distribution géométrique

Exemple.

On fait tourner 4 fois une roue numérotée de 0 à 9 .

- Quelle est la probabilité que le 0 ne sorte qu'à la 4ème tentative ?

On fait tourner la roue jusqu'à ce que le 0 sorte.

- Quelle est la probabilité de voir sortir le 0 pour la première fois à la $n$ème tentative ? Quelle est l'espérance mathématique du nombre de tentatives ?

Loi de Poisson. Cette distribution intervient dans l'étude du nombre d'événements intervenant dans un intervalle de temps (files d'attente). Elle propose aussi des modèles statistiques pour des événements dits "rares" lorsqu'un grand nombre de cas a été étudié. La variable aléatoire X *suit une loi de Poisson de paramère* λ, si X peut prendre la valeur $k \in \mathbb{N}$ avec la probabilité :

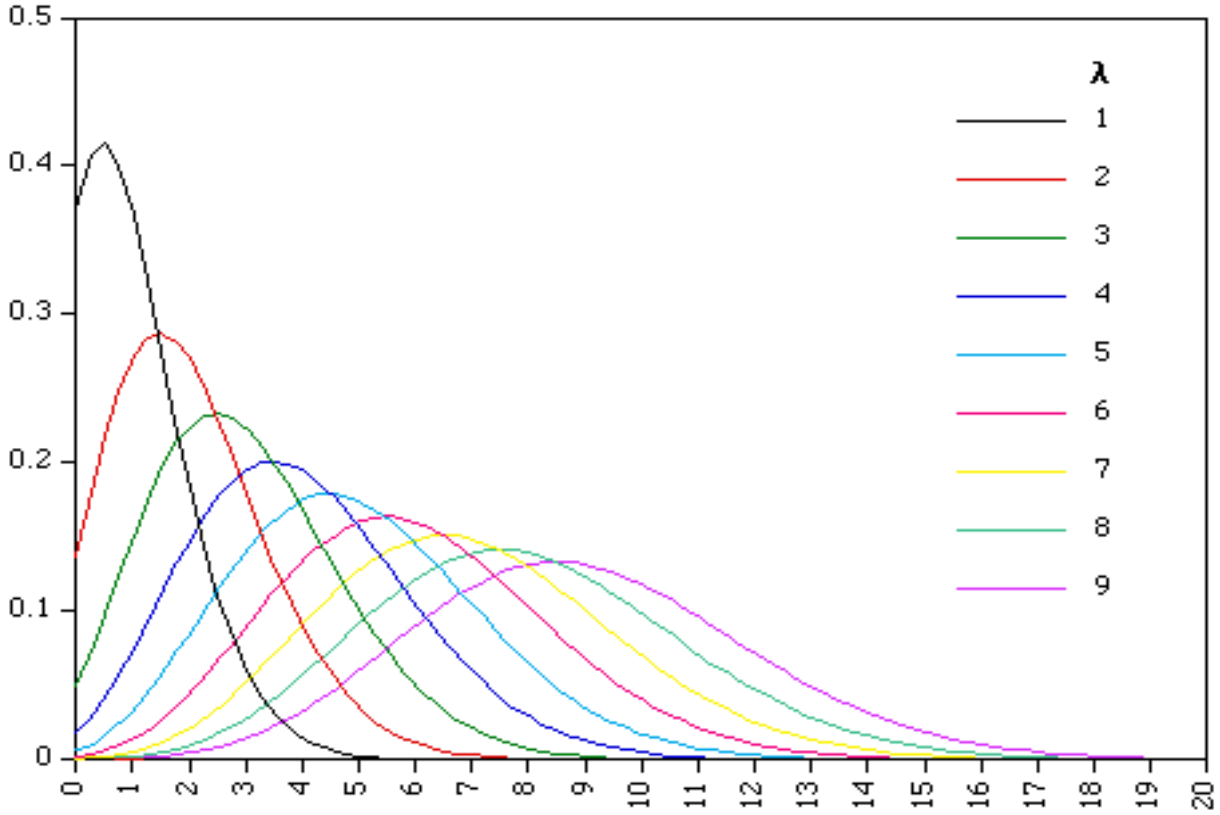
$$p_k = Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

On démontre facilement que la somme des probabilités vaut 1.

L'espérance et la variance d'une distribution de Poisson se calculent aisément et on a :

- $\mathbb{E}X = \lambda$
- $\mathbb{V}X = \lambda$
- $\sigma X = \sqrt{\lambda}$

Distribution de Poisson pour différentes valeurs de λ



Exemple. *On a répertorié dans une usine le nombre d'accidents mineurs subis par le personnel dans une journée de travail sur une période de 200 jours. Ces accidents sont survenus indépendamment les uns des autres. Les résultats sont consignés dans le tableau suivant :

Nombre d'accidents		0	1	2	3	4	5
Nombre de jours		86	82	22	7	2	1

- Quel est le nombre moyen d'accidents par jour ? Interpréter concrètement le résultat trouvé.

- On ajuste cette distribution par une loi de Poisson. Justifier cette décision et préciser cette loi ;

*. http://serge.mehl.free.fr/exos/exo_poisson1.html

- Comparer avec un ajustement par la loi binomiale.

- Quel est le nombre théorique de jours où il se produit moins de 3 accidents ? Comparer avec la réalité.

Une propriété importante de la loi de Poisson se porte sur la somme :
Proposition. La somme deux v.a. de Poisson indépendantes de paramètres λ et μ est une v.a. de Poisson de et paramètres $\lambda + \mu$.

Une modélisation par la loi de Poisson est justifiée lorsqu'on étudie les occurrences d'un événement de faible probabilité dans une population de grande taille :

Proposition (Approximation de la loi binomiale par la loi de Poisson.) Soit X_n une suite de v.a. de loi binomiale de paramètres p_n et n. On suppose que $\lim_{n\to\infty} np_n = \lambda > 0$. Alors, pour tout $k \in \mathbb{N}$:

$$\lim_{n\to\infty} Pr(X_n = k) \;=\; e^{-\lambda}\frac{\lambda^k}{k!}.$$