Variables aléatoires, lois de probabilité

M.R. Amini, C. Amblard

AMA-Analyse de données, Modélisation et Apprentissage automatique,
LIG-Laboratoire d'Informatique de Grenoble, UJF

Intuition

C'est quoi une *variable aléatoire*?

Lorsqu'on est devant une épreuve aléatoire, on s'inéteresse plus souvent à la *valeur* attribuée à l'épreuve qu'à son résultat.

Exemple

Par exemple, lorsque l'on joue à un jeu de hasard on s'intéresse plus au gain qu'on aura avec notre jeu qu'à son résultat.

Définition

Une variable aléatoire ($v.a.$) X est un nombre réel que l'on associe à chaque élément e de l'ensemble Ω. X est donc une application définie comme:

$$\begin{array}{rcl} X:\Omega & \rightarrow & \mathbb{R} \\ e & \mapsto & X(e)=x \end{array}$$

X est la variable aléatoire réelle et x est une réalisation de cette $v.a.$

Exemple

Exemple du lancer de dé

On lance un dé. Si le résultat est un 5 ou un 6 on gagne 1€, sinon on perd 1€.
Question: Comment calculer la probabilité de gagner 1€?

- $\Omega = \{1,2,3,4,5,6\}$
- Soit e_i la réaisation de l'evenement : la valeur du dé lancé est i
- Card $\Omega = 6$ et Card $\mathcal{P}(\Omega) = 2^6$
- Soit X la $v.a.$ qui associe à tout résultat du dé un *gain*:
  $X(e_1) = X(e_2) = X(e_3) = X(e_4) = -1$ et $X(e_5) = X(e_6) = 1$
  L'ensemble des valeurs possibles de X est: $\mathcal{X} = \{-1,1\} \subset \mathbb{R}$,
  X *est une variable discrète, binaire*.

Réponse: Définir une probabilité $\mathbb{P}_X$ sur $\mathcal{X}$, en retournant dans (Ω, C, P) .

$$\begin{aligned}
\mathbb{P}_X(X=1) &= P(X^{-1}(1)) \\
&= P(e_5 \cup e_6), \\
&= P(e_5) + P(e_6) \text{ car } e_5 \cap e_6 = \emptyset, \\
&= 1/6 + 1/6 = 1/3. \\
\mathbb{P}_X(X=-1) &= P(e_1 \cup e_2 \cup e_3 \cup e_4) = 2/3
\end{aligned}$$

Exemple

Simulation de mots de passe valides

- Un mot de passe = 8 signes extraits du dictionnaire $D = \cup_{j=1}^{4}(D_j)$ avec $D_1 = \{A, \cdots, Z\}$, $D_2 = \{a, \cdots, z\}$,
$D_3 = \{0, \cdots, 9\}$, $D_4 = \{"?", "!", ".", ",", ":", ""\}$
- Soit X_i le i^{eme} signe du mot de passe. $X_i \in D$,
- Soit Y le mot de passe. $Y = (X_1, \cdots, X_8)$, $Y \in D^8$.
- On peut générer automatiquement ces mots de passe suivant les lois:
  1. Les signes $\{X_i\}$ sont indépendants et identiquement distribués (i.i.d). Tous les symboles de D sont equiprobables (la distribution est uniforme).
  2. Les signes sont i.i.d. A l'intéreur de chaque dictionnaire, la distribution est uniforme et le choix du dictionnaire suit aussi une loi uniforme.
  3. Les signes ne sont pas indépendants. Si X_i est une lettre, alors X_{i+1} est un chiffre ou un symbole de D_4 (et inversement). La distribution est uniforme au sein de chaque dictionnaire.

Exemple

Simulation de mots de passe valides

1. Les signes $\{X_i\}$ sont i.i.d. Tous les symboles de D sont equiprobables.

$$\forall x \in D, \forall i \in \{1, \cdots, 8\}, P(X_i = x) = \frac{1}{68},$$

$$\forall y = (x_1, \cdots, x_8) \in D^8, P(Y = y) = \frac{1}{68^8},$$

$$\forall i \in \{1, \cdots, 8\}, P(X_i \in D_4) = \frac{6}{68} = \frac{3}{34} \sim \frac{1}{11}$$

$$P(\exists i \in \{1, \cdots, 8\}, P(X_i \in D_4) = 1 - \left(\frac{62}{68}\right)^8$$

Exemple

Simulation de mots de passe valides

2. $\{X_i\} i.i.d.$ Choix du dictionnaire equiprobable et loi uniforme au sein de chaque dictionnaire.

$$P(X_i \in D_4) = \frac{1}{4},$$

$$\begin{aligned}\forall x \in D_4, P(X_i = x) &= \sum_{j=1}^{4} P(X_i = x | X_i \in D_j) P(X_i \in D_j) \\ &= P(X_i = x | X_i \in D_4) P(X_i \in D_4) \\ &= \frac{1}{6} \times \frac{1}{4}\end{aligned}$$

$$\forall y = (x_1, \cdots, x_8) \in \prod_{i=1}^{8} D_{j_i} \text{ avec } D_{j_i} \in \{1, \cdots, 4\},$$

$$P(Y = y) = \prod_{i=1}^{8} P(X_i = x_i | X_i \in D_{j_i}) P(X_i \in D_{j_i})$$

$$P(\exists i \in \{1, \cdots, 8\} | P(X_i \in D_4) = 1 - \left(\frac{3}{4}\right)^8$$

Exemple

Simulation de mots de passe valides

3. Les signes ne sont pas indépendants. Si X_i est une lettre, alors X_{i+1} est un chiffre ou un symbole de D_4 (et inversement). On a alors:

$$P(X_{i+I} \in D_1 \cup D_2 | X_i \in D_1 \cup D_2) = 0 \qquad P(X_{i+I} \in D_3 \cup D_4 | X_i \in D_3 \cup D_4) = 0$$
$$P(X_{i+I} \in D_1 \cup D_2 | X_i \in D_3 \cup D_4) = 0.5 \quad P(X_{i+I} \in D_3 \cup D_4 | X_i \in D_1 \cup D_2) = 0.5$$

Soient: $p_{jk} = P(X_{i+1} \in D_k | X_i \in D_j)$. La matrice des $\{p_{jk}\}_{j,k \in \{1,\cdots,4\}}$ s'appelle la matrice de transition. On a alors:

$$p_{jj} = 0,\ p_{1,2} = p_{2,1} = p_{3,4} = p_{4,3} = 0$$
$$p_{1,3} = p_{3,1} = p_{1,4} = p_{4,1} = 0.5 \cdots$$
$$P(Y \in D_1 \times D_4 \times D_3 \times D_2 \times D_4 \times D_3 \times D_1 \times D_2) = P(D_1) p_{14} p_{43} p_{32} p_{24} p_{43} p_{31} p_{12}$$

Définition formelle d'une *v.a.*

Variable aléatoire

Soient:

- un espace probabilisé (Ω, C, P),
- une application $X : \Omega \to \mathbb{R}$.

On définit l'espace probabilisé $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ par:

- **La tribu** *des boréliens* $\mathcal{B}(\mathbb{R})$, engendrée par les intervalles $]-\infty, a[$,
- **La loi** $\mathbb{P}$ de X:

$$\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{P}(B) = P(X^{-1}(B)) = P(\{e_i \in \Omega \mid X(e_i) \in B\})$$

- Remarque: Cette définition n'a de sens que si X est *mesurable*:

$$\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) \in \Omega$$

Une variable aléatoire réelle est une *application mesurable* d'un espace probabilisable quelconque dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Schématiquement parlant ...

$$\begin{array}{c}(\Omega, C, P)\\ \Big\downarrow X\\ (\mathbb{R}, \mathcal{B}, \mathbb{P})\end{array}$$

$$C \xrightarrow{P} [0, 1]$$

$$X^{-1}(B) = A \longmapsto P(A)$$

$$\mathcal{B}(\mathbb{R}) \xrightarrow{\mathbb{P}} [0, 1]$$

$$B \longmapsto \mathbb{P}(B)$$

Fonction de répartition

Notations

- L'événement $X^{-1}(B) = \{e \in \Omega \mid X(e) \in B\}$ sera noté par $\{X \in B\}$,
- L'événement $X \in]-\infty, a]$ sera noté par $\{X < a\}$,
- L'événement $X \in [a, b[$ sera noté par $\{a \leq X < b\}$,

Définition

- Il suffit de connaitre les probabilités $\mathbb{P}(]-\infty, a]) = P(X < a)$ pour probabiliser $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *par* $\mathbb{P}$.
- On appelle **fonction de répartition** de la $v.a.$ X l'application:

$$\begin{array}{rcl} F : \mathbb{R} & \to & [0, 1] \\ x & \mapsto & F(x) = P(X < x) \end{array}$$

Propriétés

- $P(a \leq X \leq b) = F(b) - F(a)$
- F est croissante bornée:
  - $F(-\infty) = \lim\limits_{x \to -\infty} F(x) = 0, \quad F(+\infty) = \lim\limits_{x \to +\infty} F(x) = 1$

Fonction de répartition

Exemple: Lancer de dé

- $\Omega = \{1, 2, 3, 4, 5, 6\}$,
- La $v.a.$ X:

$$\begin{aligned} X : \Omega &\to \mathcal{X} = \{-1, 1\} \\ e &\mapsto -1 \text{ si } e \in \{1, 2, 3, 4\} \\ e &\mapsto 1 \text{ si } e \in \{5, 6\} \end{aligned}$$

- La fonction de répartition F : $F(x) = P(X < x)$
  - si $x < -1, \mathcal{X} \cap (]-\infty, x[) = \emptyset \Rightarrow F(x) = 0$
  - si $x \in [-1, 1[, \mathcal{X} \cap (]-\infty, x[) = \{-1\} \Rightarrow F(x) = \frac{2}{3}$
  - si $x \in [1, \infty], \mathcal{X} \cap (]-\infty, x[) = \{-1, 1\} \Rightarrow F(x) = \frac{2}{3} + \frac{1}{3} = 1$

$v.a.$ discrète

Définition

- Toute $v.a.$ dont la fonction de répartition est une fonction en escalier, est une $v.a.$ **discrète**
- X est définie par un ensemble $\mathcal{X}$ de valeurs possibles, au plus dénombrable, et la famille $(p(x))_{x\in\mathcal{X}}$ telle que:

$$\forall x \in \mathcal{X}, p(x) \geq 0 \text{ et } \sum_{x\in\mathcal{X}} p(x) = 1$$

$v.a.$ continue

Définition

- Toute $v.a.$ dont la fonction de répartition F est une fonction continue est appelée variable **absolument continue**
- On appelle **densité** d'une $v.a.$ X la dérivée p de sa fonction de répartition

$$F(x) = \int_{-\infty}^{x} p(u)du$$

- On a ainsi $P(a \leq X < b) = F(b) - F(a) = \int_a^b p(x)dx$
- X est définie par sa densité de probabilité $p(x)$ telle que

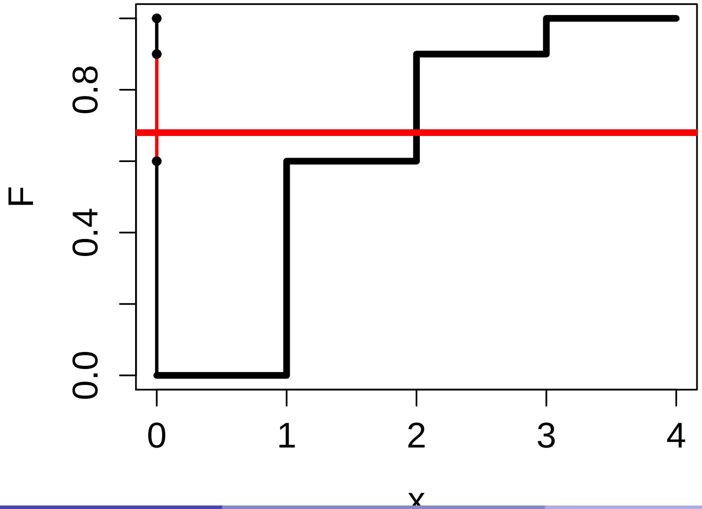
$$\forall x \in \mathbb{R},\ p(x) \geq 0 \text{ et } \int_{-\infty}^{\infty} p(x)dx = 1$$

Algorithme de simulation pour un $v.a$ discrète

Méthode par découpage d'intervalles

- Soient $X \in \{x_1, \cdots, x_n\}$, $p_i = P(X = x_i)$, $F_i = P(X \leq p_i)$,
- $[0, 1]$ est découpé en n intervalles de longueur $p_1, p_2, \cdots, p_n$
- Soit U une variable aléatoire uniforme sur $[0, 1]$ (*rand, random,runif...*)
- $X = x_i$ si $F_{i-1} < U \leq F_i$,
- Difficulté: Recherche de l'intervalle dans lequel tombe U.
- Accéleration: On peut prétraiter la loi à simuler en placant tout d'abord les intervalles les plus larges.

Simulation par découpage d'intervalles

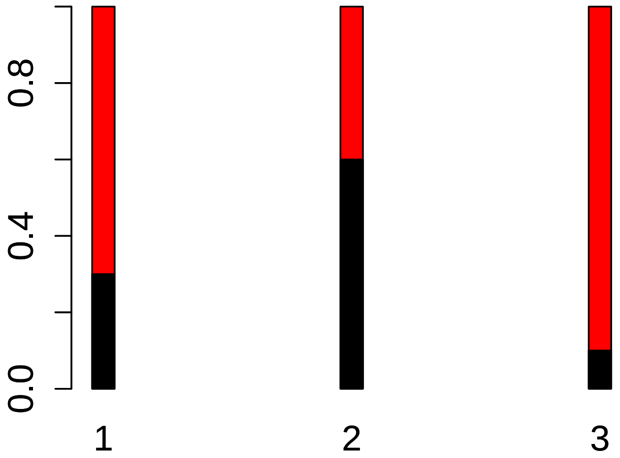


Algorithme de simulation pour un $v.a$ discrète

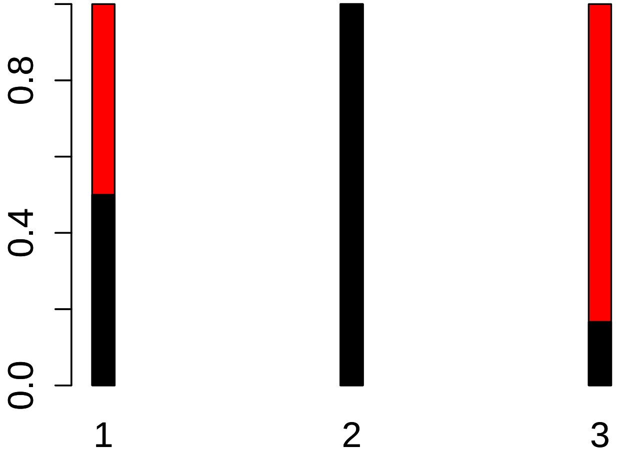
Algorithme de rejet ou de tir

- Soient $X \in \{x_1, \cdots, x_n\}$, $p_i = P(X = x_i)$,
- On note $c_i = p_i$
- **1.** Soit U une variable aléatoire uniforme sur $[0, n]$, $I = Ent[U]$, $R = U - I$
- **2.** Si $R \leq c_i$, $X = x_i$ sinon Return(1)
- Accéleration: choisir $c_i = \frac{p_i}{\max_i(p_i)}$
- Proportion de réussites par essai (facteur de tir): $\phi = \frac{\sum_i c_i}{n}$

Algorithme de rejet



Accélération de l'algorithme de rejet

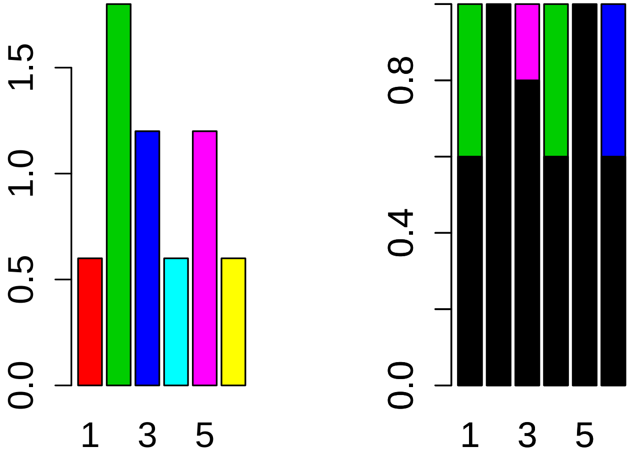


Algorithme de simulation pour un $v.a$ discrète

Algorithme de Walker

- Méthode de tir avec $\phi = 1$
- Au lieu de rejeter un tirage, on lui attribue une valeur d'aliasing.
- Chaque barre est caractérisée par une modalité x_i, une valeur seuil s_i qui permettra de choisir x_i si $U \leq s_i$ et une modalité $alias_i$ choisie si $U > s_i$.
- **Tirage:**
  - Soit U une variable aléatoire uniforme sur $[0, n]$, $W = Ent[U]$, $R = U - W$
  - Si $R \leq s_i$, alors $X = x_i$ sinon $X = alias_i$
- **Construction de la table de Walker:**
  - Supposons $p_1 \leq p_2 \leq \cdots \leq p_n$
  - $s_i = n * p_i$
  - $i = 1$
  - tant que $s_n > 1$
  - $alias_i = n$,
  - $s_n = s_n + s_i - 1$,
  - on reordonne les p_i,
  - $i = i + 1$.

Algorithme de Walker



Loi de probabilité et moments

Définition

Une **loi de probabilité** est un **modèle** qui décrit une situation.

Exemple

- La valeur obtenue par un dé suit une loi uniforme,
- l'obtention d'un 6 lors d'un lancer de dé contre l'obtension d'une autre valeur suit une loi de Bernoulli,
- le nombe de 6 obtenus lors de n lancers de dés suit une loi binomiale,
- le nombre de piques dans une main de 5 cartes tirées dans un jeu de 32 cartes suit une loi hypergéometrique,
- la taille moyenne d'un étudiant de 18 ans suit une loi normale.

Définition

Les **moments** d'une $v.a\ X$ **résument** le comportement de X.

Espérance mathématique

Définition

Soit X une $v.a.$. L' **espérance mathématique** de X, $\mathbb{E}(X)$ est:

- Si X est discrète $\mathbb{E}(X) = \sum_i x_i P(X = x_i)$
- si X est continue $\mathbb{E}(X) = \int_{-\infty}^{\infty} xp(x)dx = \int_{\mathbb{R}} xp(x)dx$

Propriétés

Pour tout réel $a \in \mathbb{R}$ et tout couple de $v.a.(X, Y)$

- $\mathbb{E}(a) = a$
- $\mathbb{E}(aX) = a\mathbb{E}(X)$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$
- Si X et Y indépendantes, $\mathbb{E}(X \times Y) = \mathbb{E}(X) \times \mathbb{E}(Y)$

Variance d'une *v.a.*

Définition

Soit X une $v.a.$ d'espérance m, La **variance** de X, $V(X)$ est:

$$V(X) = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \int_{\mathbb{R}} (x-m)^2 p(x) dx$$

Remarques

- La variance est le carré de l'écart-type, σ_X
- La variance est une mesure de la dispersion de X autour de m.

Propriétés

- $V(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$,
- $\forall a \in \mathbb{R}, V(a) = 0$,
- $\forall a \in \mathbb{R}, V(X+a) = V(X)$,
- $\forall a \in \mathbb{R}, V(aX+b) = a^2 V(X)$

Autres moments

Définition

On définit les moments centrés d'ordre k d'une $v.a.$ X d'espérance m:

$$\mu_k = \mathbb{E}\left[(X-m)^k\right]$$

Remarques

- $\mu_1 = 0$,
- $\mu_2 = V(X)$,
- μ_3 caractérise l'asymétrie de la distribution de X,
- μ_4 caractérise l'applatissement de la distribution de X

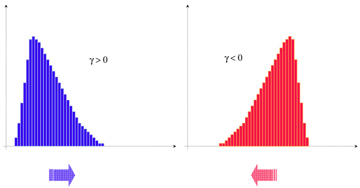
Skewness

Définition

Le coefficient d'asymétrie normalisé ou skewness est défini par

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

Un coefficient γ_3 positif, indique une queue de distribution étalée vers la droite, alors qu'un coefficient négatif indique un étalement vers la gauche.

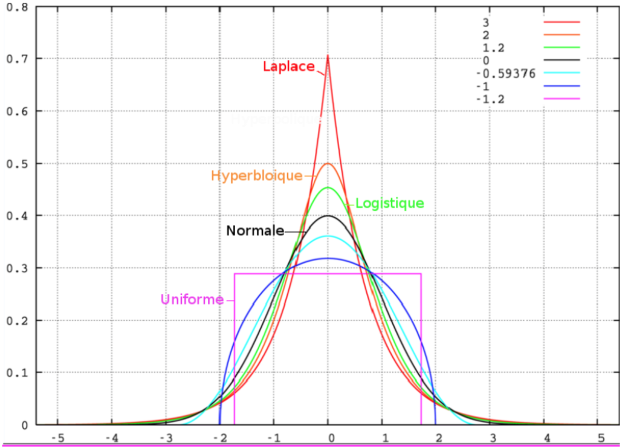


Kurtosis

D/'efinition

Le coefficient d'applatissement normalisé ou kurtosis est défini par

$$\gamma_2 = \frac{\mu_4}{\sigma^4}$$



Lois de probabilités discrètes

Loi Uniforme - définition

Soit X une $v.a.$ prenant chaque valeur de l'ensemble $\{1, ..., n\}$ la loi de X est dite uniforme sur $[\![1, n]\!]$ si

$$\forall k \in [\![1, n]\!], P(X = k) = \frac{1}{n}$$

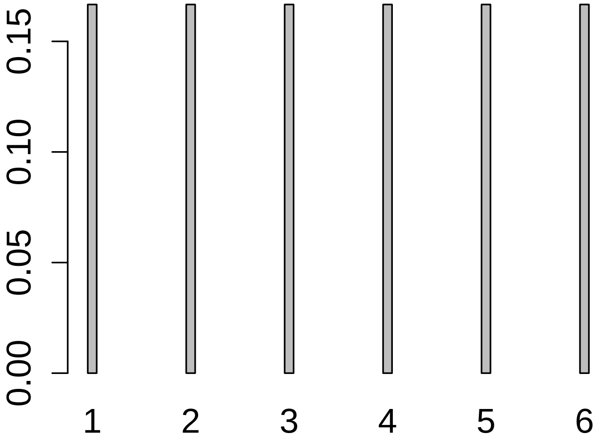
Moments

- $\mathbb{E}(X) = \frac{n+1}{2}$
- $V(X) = \frac{n^2-1}{12}$

Domaine d'application

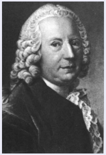
Cette loi intervient dans l'étude de divers jeux (pile ou face, dés, loterie, ..).
Loi uniforme= évenements **equiprobables**.

la platitude de la loi uniforme



Lois de probabilités discrètes

Loi de Bernoulli - définition



C'est la loi d'une $v.a.$ binaire $X \in \{0, 1\}$
Si on appelle q la probabilité que X soit égale à 1

$$\forall x \in \{0, 1\}, P(X = x) = q^x (1 - q)^{1-x}$$

Moments

$$\mathbb{E}(X) = q \quad V(X) = q(1 - q)$$

Domaine d'application

Cette loi décrit les phénomènes dont la valeur est binaire.

- en apprentissage automatique : erreur d'un modèle.
- en recherche d'information: présence ou absence d'un mot ou concept
- en général toutes les situations succés/echec ou présence/absence ou bon etat/déféctueux.

Loi de Bernoulli

Utilisation de la loi de Bernoulli en recherche d'information

Recherche d'information: Rechercher dans une collection de document ceux qui sont pertinents avec une requête.

Représentaion binaire d'un document ou d'une requête

- Vocabulaire= ensemble de mots $V = \{w_1, \cdots, w_V\}$.
- Un document texte est représenté par un vecteur d de taille V binaire,

$$d = (w_1^d, \cdots, w_V^d) \text{ où } w_i^d = 1 \text{ si } w_i \in d, \ 0 \text{ sinon.}$$

- Une requête est représentée par un vecteur q de taille V binaire,
- Soit la variable de pertinence $R_{d,q} = 1$ si le document d est pertinent avec la requête q, 0 sinon.

Problème: Donnez à l'utilisateur une liste de documents classés par ordre de pertinence.

$$\text{Score: } \Delta(d, q) = \frac{P(R_{dq} = 1|d, q)}{P(R_{dq} = 0|d, q)}$$

Loi de Bernoulli

Utilisation de la loi de Bernoulli en recherche d'information

- Règle de Bayes: $\Delta(d,q) = \dfrac{P(d|R=1,q) * P(R=1|q)}{P(d|R=0,q) * P(R=0|q)}$
- Supposons que les occurences des mots sont indépendantes, $P(d=(w_1^d,\cdots,w_V^d)|R=r,q) = \prod_{i=1}^{V} P(W_i^d = w_i^d|R=r,d)$
- Supposons que $(W_i^d|R=1,q)$ suit ue loi de Bernoulli de paramètre p_i et $(W_i^d|R=0,q)$ une loi de Bernoulli de paramètre τ_i. Alors:

$$
\begin{aligned}
\Delta(p,q) &= \Gamma(q) \prod_{i:w_i^d=w_i^q=1} \frac{p_i}{\tau_i} \prod_{i:w_i^d=0,\ w_i^q=1} \frac{1-p_i}{1-\tau_i} \\
&= \Gamma(q) \prod_{i:w_i^d=w_i^q=1} \frac{p_i}{\tau_i} \prod_{i:w_i^d=0,\ w_i^q=1} \frac{1-p_i}{1-\tau_i} \prod_{i:w_i^d=w_i^q=1} \frac{1-p_i}{1-\tau_i}\frac{1-\tau_i}{1-p_i} \\
&= \Gamma(q) \prod_{i:w_i^d=w_i^q=1} \frac{p_i(1-\tau_i)}{\tau_i(1-pi)} \prod_{i:w_i^q=1} \frac{1-p_i}{1-\tau_i}
\end{aligned}
$$

- $s(d,q) = \sum_{i:w_i^d=w_i^q=1} ln\left(\frac{p_i(1-\tau_i)}{\tau_i(1-pi)}\right)$

Lois de probabilités discrètes

Loi binomiale - définition

Soit X la somme de n variables X_i indépendantes, suivant toutes la loi de Bernoulli de paramètre q. La loi suivie par X est la loi binomiale de baramètres n et q, notée $\mathcal{B}(n, q)$

$$\forall k \in \{1, ..., n\} P(X = k) = C_n^k q^k (1 - q)^{n-k}$$

Moments

$$\mathbb{E}(X) = nq, \;\; V(X) = nq(1 - q)$$

Domaine d'application

Cette loi décrit des répétitions d'une même épreuve binaire.

- nombre d'ordinateurs en panne dans un parc de n ordinateurs indépendants,
- nombre d'occurences d'un mot dans un document,
- en général nombre de succés sur n expériences i.i.d.

Lois de probabilités discrètes

Loi de Poisson - définition

Sur une période T, un événement arrive en moyenne λ fois. La loi suivit par la $v.a.$ déterminant le nombre de fois où l'événement se produit dans la période T est:

$$\forall k \in \mathbb{N}, P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Moments

$$\mathbb{E}(X) = \lambda, \;\; V(X) = \lambda$$

La loi de Poisson est la seule loi discrète vérifiant $\mathbb{E}(X) = V(X)$

Domaine d'application

La loi de Poisson est la loi des petites probabilités, loi des événements rares. C'est la loi limite de la loi binomiale lorsque le nombre d'essais est grand et la probabilité de succés très petite.

- accidents dans un atelier de grande taille,
- réseaux informatiques (nb de requêtes en un emps T),

Loi binomiale et loi de Poisson

Utilisation des lois binomiale et de Poisson en RI

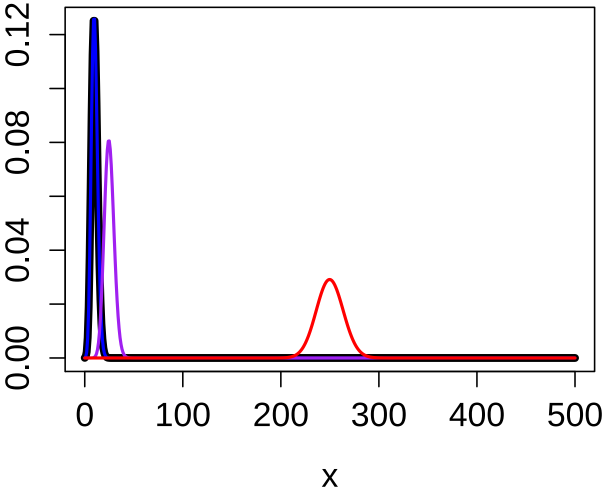
- Supposons qu'un document texte soit représenté par le vecteur $d = (x_1^d, \cdots, x_V^d)$ où x_i^d est le nombre d'occurence du mot w_i dans le document d.
- Soit N_d le nombre de mots contenus dans le document d.
- Loi binomiale: $P(X_i^d = x_i) = C_{N_d}^{x_i} q_i^{x_i} (1 - q_i)^{N_d - x_i}$
- Loi multinomiale: Chaque mot de d appartient à V et les mots sont indépendants.

$$P(X_1^d = x_1^d, \cdots, X_V^d = x_V^d) = \frac{N_d!}{x_1^d! \cdots x_V^d!} {q_1}^{x_i^d} \cdots {q_V}^{x_V^d}$$

- Loi de Poisson *Le mot est rare*:

$$P(X_i^d = x_i) = \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!}$$

Loi binomiale et loi de Poisson



Lois de probabilités continues

Loi uniforme

Une $v.a.$ réelle X suit une loi uniforme sur l'intervalle $[0, a]$ si sa loi de probabilité admet pour densité

$$p(x) = \begin{cases} \frac{1}{a} & \text{si } x \in [0, a], \\ 0 & \text{si } x \in]-\infty, 0[\cup]a, \infty[\end{cases}$$

Moments

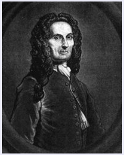
- $\mathbb{E}(X) = \frac{a}{2}$
- $V(X) = \frac{a^2}{12}$

Domaine d'application

La loi uniforme traduit l'hypothèse d'équipartition sur l'intervalle

Lois de probabilités continues

Loi de Laplace-Gauss ou Loi normale



Une $v.a.$ réelle X suit une loi de Laplace-Gauss de paramÃ¨tres m et σ si sa densité a pour expression

$$\forall x \in \mathbb{R}, p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

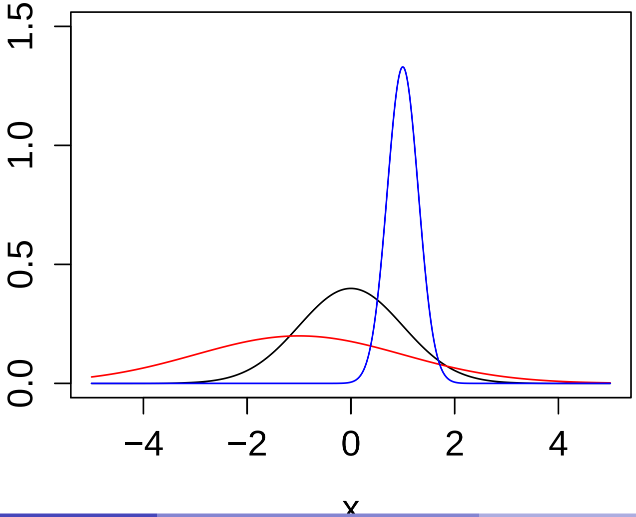
Moments

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

Domaine d'application

Cette loi joue un rôle fondamental en statistique et elle apparaît comme loi limite de certaines caractéristique d'échantillons de grandes tailles. Son domaines d'application est varié: apprentissage automatique, médecine, ...

Loi de Laplace-Gauss ou Loi normale



Inégalité de Markov

Énoncé

Soit X une $v.a.$ non négative, et a un réel strictement positif, on a alors

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Démonstration

$$aP(X \geq a) = a\int_a^{\infty} p(x)dx \leq \int_a^{\infty} xp(x)dx \leq \mathbb{E}(X)$$

Intérpretation

Si $\mathbb{E}(X)$ est petit et que la $v.a.$ X est positive alors, cette variable est proche de 0 avec une grande probabilité.

Inégalité de Chebychev

Énoncé

Soit X une $v.a.$ d'espérance $\mathbb{E}(X)$ et de variance $V(X)$ et soit $\lambda \in \mathbb{R}_+$, on a alors

$$P(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{V(X)}{\lambda^2}$$

Démonstration

Les évenement $(|X - \mathbb{E}(X)| \geq \lambda)$ et $(X - \mathbb{E}(X))^2 \geq \lambda^2)$ sont équivalents.
Soit $Y = (X - \mathbb{E}(X))^2$. $Y \geq 0$ et $E(Y) = E(X - \mathbb{E}(X))^2 = V(X)$.
On applique l'inégalité de Markov à Y en choisissant $a = \lambda^2$,

$$P(Y = (X - \mathbb{E}(X))^2 \geq \lambda^2) \leq \frac{\mathbb{E}(Y) = V(X)}{\lambda^2}$$

$$P(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{V(X)}{\lambda^2}$$

Intérpretation

Si $V(X)$ est petite, la X est proche de $\mathbb{E}(X)$ avec une forte probabilité.